

LA PANDEMIA DE COVID-19 EN MÉXICO:

PERSPECTIVAS DESDE LA COMPUTACIÓN,
LA SALUD Y LO SOCIAL



Sergio Rafael Coria Olguín
María Guadalupe Ramírez Rojas
Hermelando Cruz Pérez
Sergio Paredes-Solís
Jesús Elizarrarás Rivas



UADY
UNIVERSIDAD
AUTÓNOMA
DE YUCATÁN



LA PANDEMIA DE COVID-19 EN MÉXICO:

PERSPECTIVAS DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Sergio Rafael Coria Olguín
María Guadalupe Ramírez Rojas
Hermelando Cruz Pérez
Sergio Paredes-Solís
Jesús Elizarrarás Rivas



UADY
UNIVERSIDAD
AUTÓNOMA
DE YUCATÁN



RA 652.2 La pandemia de COVID-19 en México : perspectivas
.D38 desde la computación, la salud y lo social / Sergio
.P35 Rafael Coria Olguín [y otros].—Mérida, Yucatán :
2025 Universidad Autónoma de Yucatán, 2025.

1 recurso electrónico (186 páginas) : ilustraciones, gráficas.

1. *Epidemiología— Procesamiento de datos*
 2. *Epidemiología— Métodos estadísticos.*
 3. *Pandemia de COVID-19, 2020—Estadísticas.*
 4. *Pandemia de COVID-19, 2020—Aspectos de salud—México.*
 5. *Pandemia de COVID-19, 2020—Aspectos sociales—México.*
I. Coria Olguín, Sergio Rafael.
-

*La pandemia de COVID-19 en México: perspectivas desde
la computación, la salud y lo social*

Primera edición: febrero de 2025

D. R. © 2024, Universidad Autónoma de Yucatán bajo el sello
de la Casa Editorial UADY Calle 60 núm. 491 A por 57, centro,
C. P. 97000, Mérida, Yucatán, México Tel. +52 (999) 923 9769
casa.editorial@correo.uady.mx www.uady.mx/casa-editorial

D. R. © 2024, Universidad Autónoma “Benito Juárez” de Oaxaca
Av. Universidad s/n, col. Cinco Señores, C.P. 68120, Oaxaca
de Juárez, Oaxaca, México. <http://www.uabjo.mx>

ISBN UADY: 978-607-8741-71-7

ISBN UABJO: 978-607-8815-40-1

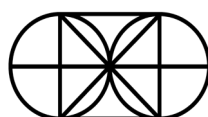


Las ideas y opiniones expresadas en esta obra son responsa-
bilidad exclusiva de los autores y no reflejan necesariamente la
postura oficial de la Universidad Autónoma de Yucatán.

Hecho en México / *Made in Mexico*



UADY
UNIVERSIDAD
AUTÓNOMA
DE YUCATÁN



CASA EDITORIAL
UADY



LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Sergio Rafael Coria Olguín,
María Guadalupe Ramírez Rojas,
Hermelando Cruz Pérez,
Sergio Paredes-Solís,
Jesús Elizarrarás Rivas

RESUMEN

Este libro presenta diversos análisis de la base de datos de casos de COVID-19 publicada por la Secretaría de Salud (SSA) del Gobierno de México. Tiene un enfoque cuantitativo y un alcance descriptivo y exploratorio, con una perspectiva de computación, ciencias de la salud y ciencias sociales. Se buscan, se descubren y se representan distintos tipos de patrones que resultan de interés para los investigadores y los practicantes de diversas disciplinas enfocadas en la salud, principalmente la salud pública. La base de datos analizada ofrece aproximadamente 40 variables, correspondientes a más de 18 millones de individuos, que incluyen casos confirmados, negativos e indeterminados de la enfermedad, del periodo de febrero de 2020 a septiembre de 2022. Debido a su gran volumen, esta base de datos requiere el uso de técnicas y herramientas de ciencia de datos. Se aplican, principalmente técnicas de estadística descriptiva y de aprendizaje automático usando software libre. Se consideran dos niveles de unidad de análisis: *individuo* y *municipio*. Para el nivel de individuo se exploran patrones de tipo clínico. Para el nivel de municipio, con base en la agregación de casos individuales, los patrones abordados son de tipo cronológico, geográfico, y sociodemográfico. Los patrones cronológicos se exploran con base en las semanas epidemiológicas definidas por la SSA y los meses. Los resultados consisten, principalmente, en análisis estadísticos, modelos de aprendizaje automático y mapas digitales. Se ofrecen al lector algunos modelos, mapas y conjuntos de datos (datasets) para su descarga gratuita en la web. El conocimiento obtenido es útil, desde un punto de vista histórico, para profundizar en la comprensión del comportamiento del COVID-19 en la población mexicana con propósitos clínicos y para el diseño o revisión de políticas públicas de prevención o atención a esta enfermedad y otras relacionadas.

Una mención aparte merece el trabajo realizado por el Cuerpo Académico de Biomedicina y Enfermedades Infecciosas y Parasitarias. Sus aportaciones al conocimiento sobre el virus del papiloma humano y sobre la influenza son relevantes, ya que am-

RESUMEN

Los problemas de salud son causantes de serios problemas de salud en nuestra sociedad. Así, la universidad aporta respuestas a los problemas de salud de nuestra entidad acorde con su responsabilidad social.

THE COVID-19 PANDEMIC IN MEXICO: PERSPECTIVES FROM COMPUTING, HEALTHCARE AND SOCIAL SCIENCES

Sergio Rafael Coria Olguín,
María Guadalupe Ramírez Rojas,
Hermelando Cruz Pérez,
Sergio Paredes-Solís,
Jesús Elizarrarás Rivas

ABSTRACT

This book presents a series of analyses on the COVID-19 database published by the Health Ministry of the Mexican Government. Its approach is quantitative with a descriptive and exploratory scope, and a perspective from computer, healthcare and social sciences. Several types of patterns are sought, discovered and represented. These are interesting to researchers and practitioners in a number of disciplines focused on public health, mainly. The database contains, approximately, 40 attributes corresponding to over 18 million individuals, including confirmed, negative, and other types of COVID-19 cases in the period from February 2020 to September 2022. Techniques and tools in the data science field are needed to analyze this database due to its relatively large size. Descriptive statistics, and machine learning techniques are applied by using free software. Two levels of analysis unit are considered: individual, and municipality. In the individual level, clinical type patterns are explored. In turn, in the municipal level, the search for patterns is based on the aggregation of individual cases, and the explored patterns are of chronological, geographic, and socio-demographic nature. Chronological patterns are explored by using Health Ministry's epidemiological weeks and months as analysis units. The overall results consist, mainly, in descriptions and explanations of statistical analyses, machine-learning models, and digital maps. A series of models, maps, and datasets are available to readers for free downloading from the Web. The obtained knowledge is useful from a historical view to deepen in the understanding of the COVID-19 pandemic in the Mexican population for clinical purposes and for designing or reviewing public policies aimed at prevention or managing of this disease and other related.

A separate mention deserves the work carried out by the Academic Body of Biomedicine and Infectious and Parasitic Diseases. His contributions to knowledge about the human papillomavirus and influenza are relevant, since both are causes of serious health problems in our society. Thus, the university provides answers to the health problems of our entity in accordance with its social responsibility.

ÍNDICE

Agradecimientos	16
Prefacio	17
Glosario	19
Introducción	21
Datos abiertos referentes a la pandemia en México	23
Del porqué de nuestro libro y un breve recuento de este	25
Capítulo 1. Diseño metodológico	29
1.1 La problemática	29
1.2 Objeto de estudio	29
1.3 Objetivos de investigación	30
1.4 Necesidad y utilidad de nuestro libro	31
1.5 Enfoque y alcance	31
1.6 Supuestos	32
1.7 Diseño de la investigación	32
1.7.1 Recolección de la información	33
1.7.2 Dos tipos de unidad de análisis	33
1.7.3 Las variables tiempo y territorio	34
1.7.4 Dos técnicas de análisis y modelación	34
1.7.5 Estadística descriptiva	35
1.7.6 Aprendizaje automático	38
1.7.7 Patrones geográficos y mapas digitales	41
1.8 Sobre árboles clasificadores e inferencia estadística	42
Capítulo 2. Panorama estadístico descriptivo general del dataset Master-3	44
2.1 Características del paciente	45
2.2 Historia clínica y comorbilidades	51
2.3 Curso clínico y resultados	54
2.4 Contexto epidemiológico	59

Capítulo 3. Panorama estadístico descriptivo de casos positivos del dataset Master-3	66
3.1 Características del paciente	66
3.2 Historia clínica y comorbilidades	72
3.3 Curso clínico y resultados	74
3.4 Contexto epidemiológico	79
Capítulo 4. Patrones de correlación multivariable hallados con aprendizaje automático en casos individuales	86
4.1 Características del paciente	90
4.1.1 Estado de mortalidad: falleció o no (atributo falleció)	91
4.1.2 Estado de intubación (atributo intubado)	92
4.1.3 Tipo de paciente: ambulatorio u hospitalizado (atributo tipo_paciente)	93
4.1.4 Clasificación final del caso: COVID-19 confirmado, negativo, inválido, etc. (atributo clasificación_final)	94
4.1.5 Entidad federativa de residencia (atributo entidad_res)	97
4.1.6 Estatus indígena (atributo indígena)	97
4.1.7 Lengua indígena (atributo habla_lengua_indig)	98
4.1.8 Sexo (atributo sexo)	99
4.1.9 Estado de embarazo (atributo embarazo)	99
4.1.10 Intervalo de edad (atributo edad_intervalo)	100
4.2 Historia clínica y comorbilidades	101
4.2.1 Diabetes (atributo diabetes)	101
4.2.2 Inmunosupresión (atributo inmunopr)	102
4.2.3 Obesidad (atributo obesidad)	102
4.2.4 Enfermedad renal crónica (atributo renal_crónica)	103
4.3 Curso clínico y resultados	104
4.3.1 Ingreso en UCI (atributo UCI)	104
4.3.2 Resultados de laboratorio (atributo resultado_lab)	105
4.3.3 Resultados de la prueba de antígeno (atributo resultado_antígeno)	106
4.4 Contexto epidemiológico	107
4.4.1 Ola epidémica o periodo interepidémico (atributo ola_o_per_interepi)	107
4.4.2 Mes y año de ingreso (atributo año_mes_ingreso)	108
4.5 Discusión de los resultados	108
Capítulo 5. Patrones cronológicos en el nivel de agregación municipal	112
5.1 El dataset MuniCovid-2022	112
5.2 Análisis por frecuencia absoluta	115
5.2.1 Frecuencias absolutas de casos positivos	115
5.2.2 Frecuencias absolutas de defunciones	116
5.3 Análisis de tasas municipales por cada 100 mil habitantes	135
5.3.1 Tasas de casos positivos	136
5.3.2 Tasas de defunción	141
5.4 Discusión de los resultados	148

Capítulo 6. Patrones geográficos en mapas digitales	149
6.1 Acmé o pico de una etapa de intensificación de transmisión del SARS-CoV-2	150
6.2 Mapas de calor	151
6.3 Mapas coropléticos	157
6.4 Mapas de tipo gif animado	163
Capítulo 7. Niveles de ocurrencia de COVID-19 con relación a los perfiles sociodemográficos de los municipios	164
7.1 Tasas de positivos en 2020: posit_2020_tasa_100mil_INTERV	167
7.2 Tasas de positivos en 2021: posit_2021_tasa_100mil_INTERV	168
7.3 Tasas de positivos en 2022 (solo hasta 17 de septiembre): posit_2022_a_sep_17_tasa_100mil_INTERV	169
7.4 Tasas de positivos desde 2020 hasta 17/Sep/2022: posit_MPAL_TOTAL_tasa_100mil_INTERV	170
7.5 Tasas de defunción en 2020: def_2020_tasa_100mil_INTERV	172
7.6 Tasas de defunción en 2021: def_2021_tasa_100mil_INTERV	173
7.7 Tasas de defunción en 2022 (solo hasta 17 de septiembre): def_2022_a_sep_17_tasa_100mil_INTERV	173
7.8 Tasas de defunción desde 2020 hasta 17/Sep/2022: def_MPAL_TOTAL_tasa_100mil_INTERV	175
7.9 Discusión de los resultados	176
Capítulo 8. Comentarios finales, recomendaciones y trabajo de investigación a futuro	178
Desafíos y retos pendientes en torno al subregistro y manejo de la información	180
Desafíos en torno a la diversidad geográfica	
Sobre el empleo de análisis de patrones de correlación multivariable	181
REFERENCIAS	183

Índice de tablas

Tabla 1. Fases de la dinámica de la transmisión de la pandemia de la COVID-19: comparativo según la OMS versus la SSA.	22
Tabla 2. Selección de variables disponibles en la base de datos de (SSA, 2022) a las que realizamos análisis de estadística descriptiva con granularidad de individuo.	35
Tabla 3. Variables de la base de datos COVID-19 de interés especial para analizar tendencias por municipio y semana epidemiológica.	37
Tabla 4. Valores de la variable clasificacion_final que se toman como casos positivos en esta investigación.	38
Tabla 5. Falleció.	45
Tabla 6. Intubado.	46
Tabla 7. Tipo de paciente.	46
Tabla 8. Clasificación final del caso.	46
Tabla 9. Entidad federativa de nacimiento.	47

Tabla 10. Entidad federativa de residencia.	48
Tabla 11. Nacionalidad.	49
Tabla 12. Indígena.	49
Tabla 13. Habla alguna lengua indígena.	49
Tabla 14. Sexo.	49
Tabla 15. Embarazo.	50
Tabla 16. Paciente pediátrico (entre 0 y 18 años de edad).	50
Tabla 17. Paciente pediátrico de 0 (cero) años de edad.	50
Tabla 18. Paciente de 65 años o más.	50
Tabla 19. Intervalo de edad.	51
Tabla 20. Diabetes.	51
Tabla 21. Inmunosupresión.	52
Tabla 22. Obesidad.	52
Tabla 23. Insuficiencia renal crónica.	52
Tabla 24. Neumonía.	52
Tabla 25. EPOC.	53
Tabla 26. Asma.	53
Tabla 27. Hipertensión.	53
Tabla 28. Enfermedad cardiovascular.	53
Tabla 29. Tabaquismo.	54
Tabla 30. Alguna otra comorbilidad.	54
Tabla 31. Individuo en unidad de cuidados intensivos (UCI).	56
Tabla 32. Se tomó muestra de laboratorio.	56
Tabla 33. Resultado de laboratorio.	57
Tabla 34. Se tomó muestra de antígeno.	57
Tabla 35. Resultado de prueba de antígeno.	57
Tabla 36. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso al establecimiento de salud.	57
Tabla 37. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción.	58
Tabla 38. Cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción.	59
Tabla 39. Origen.	60
Tabla 40. Sector.	60
Tabla 41. Entidad federativa de la unidad médica de atención.	61
Tabla 42. Entidad y municipio de residencia: solo el 79% más alto de los casos (parte 1 de 2).	62
Tabla 43. Número de ola epidémica o de periodo interepidémico.	63
Tabla 44. Año y mes de ingreso al establecimiento de salud.	64
Tabla 45. Semana epidemiológica: solo el 80.4% más alto de los casos.	65
Tabla 46. Estuvo en contacto con otro caso de COVID-19.	65
Tabla 47. Falleció: solo en casos positivos de COVID.	67
Tabla 48. Intubado: solo en casos positivos de COVID.	67
Tabla 49. Tipo de paciente: solo en casos positivos de COVID.	67
Tabla 50. Clasificación final del caso: solo en casos positivos según las categorías 1, 2 y 3 de esta variable.	67

Tabla 51. Entidad federativa de nacimiento: solo en casos positivos de COVID.	68
Tabla 52. Entidad federativa de residencia: solo en casos positivos de COVID.	69
Tabla 53. Nacionalidad: solo en casos positivos de COVID.	69
Tabla 54. Indígena: solo en casos positivos de COVID.	70
Tabla 55. Habla alguna lengua indígena: solo en casos positivos de COVID.	70
Tabla 56. Sexo: solo en casos positivos de COVID.	70
Tabla 57. Embarazo: solo en casos positivos de COVID.	70
Tabla 58. Paciente pediátrico (entre 0 y 18 años de edad): solo en casos positivos de COVID.	71
Tabla 59. Casos pediátricos de 0 (cero) años de edad: solo en casos positivos de COVID.	71
Tabla 60. Individuos de 65 años o más: solo en casos positivos de COVID.	71
Tabla 61. Intervalos de edad: solo en casos positivos de COVID.	72
Tabla 62. Diabetes: solo en casos positivos de COVID.	72
Tabla 63. Inmunosupresión: solo en casos positivos de COVID.	72
Tabla 64. Obesidad: solo en casos positivos de COVID.	72
Tabla 65. Insuficiencia renal crónica: solo en casos positivos de COVID.	73
Tabla 66. Neumonía: solo en casos positivos de COVID.	73
Tabla 67. EPOC: solo en casos positivos de COVID.	73
Tabla 68. Asma: solo en casos positivos de COVID.	73
Tabla 69. Hipertensión: solo en casos positivos de COVID.	73
Tabla 70. Enfermedad cardiovascular: solo en casos positivos de COVID.	74
Tabla 71. Tabaquismo: solo en casos positivos de COVID.	74
Tabla 72. alguna otra comorbilidad: solo en casos positivos de COVID.	74
Tabla 73. Individuo en unidad de cuidados intensivos (UCI): solo en casos positivos de COVID.	76
Tabla 74. Se tomó muestra de laboratorio: solo en casos positivos de COVID.	76
Tabla 75. Resultado de laboratorio: solo en casos positivos de COVID.	77
Tabla 76. Se tomó muestra de antígeno: solo en casos positivos de COVID.	77
Tabla 77. Resultado de prueba de antígeno: solo en casos positivos de COVID.	77
Tabla 78. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso: solo en casos positivos de COVID.	78
Tabla 79. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción: solo en casos positivos de COVID.	78
Tabla 80. Cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción: solo en casos positivos de COVID.	79
Tabla 81. Origen: solo en casos positivos de COVID.	79
Tabla 82. Sector: solo en casos positivos de COVID.	80

Tabla 83. Entidad federativa de la unidad médica de atención: solo en casos positivos de COVID.	80
Tabla 84. Entidad y municipio de residencia: solo el 78.3% más alto de casos positivos de COVID (parte 1 de 2).	82
Tabla 85. Número de ola epidémica o de periodo interepidémico: solo en casos positivos de COVID.	83
Tabla 86. Año y mes de ingreso: solo en casos positivos de COVID.	84
Tabla 87. Semana epidemiológica: solo el 80% más alto de los casos positivos de COVID.	85
Tabla 88. Estuvo en contacto con otro caso de COVID: solo en casos positivos de COVID.	85
Tabla 89. Resumen de los modelos de árboles clasificadores producidos con casos individuales.	88
Tabla 90. Análisis de Pareto del atributo target clasificacion_final	95
Tabla 91. Análisis de Pareto del atributo target clasificacion_final en el dataset Mujeres.	96
Tabla 92. Olas epidémicas y periodos interepidémicos con sus semanas y fechas de inicio y fin.	107
Tabla 93. Diccionario de datos del <i>dataset</i> MuniCovid-2022 (versión resumida).	114
Tabla 94. Conteos de casos positivos por año y mes de ingreso.	116
Tabla 95. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020.	117
Tabla 96. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020: solo el segmento de 0 a 3,500.	118
Tabla 97. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020: solo el segmento de 0 a 260.	118
Tabla 98. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021.	119
Tabla 99. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021: solo el segmento de entre 0 y 7,000.	120
Tabla 100. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021: solo el segmento de entre 0 y 500.	120
Tabla 101. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep).	121
Tabla 102. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep): solo el segmento de 0 a 7,000.	121
Tabla 103. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep): solo el segmento de 0 a 500.	122
Tabla 104. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022.	123
Tabla 105. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022: solo el segmento de 0 a 15,000.	124
Tabla 106. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022: solo el segmento de 0 a 2,000.	125

Tabla 107. Conteos de defunciones por año y mes.	126
Tabla 108. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020.	127
Tabla 109. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020: solo el segmento entre 0 y 300 defunciones.	128
Tabla 110. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020: solo el segmento entre 0 y 20.	128
Tabla 111. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021.	129
Tabla 112. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021: solo el segmento entre 0 y 255 defunciones.	130
Tabla 113. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021: solo el segmento entre 0 y 45 defunciones.	130
Tabla 114. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep).	131
Tabla 115. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep): solo el segmento entre 0 y 51.	132
Tabla 116. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep): solo el segmento entre 0 y 6.	132
Tabla 117. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022.	133
Tabla 118. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022: solo el segmento entre 0 y 510.	134
Tabla 119. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022: solo el segmento entre 0 y 90.	135
Tabla 120. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2020.	136
Tabla 121. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2020: solo el segmento de tasas entre 0 y 900.	137
Tabla 122. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2021.	137
Tabla 123. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2021: solo el segmento de tasas entre 0 y 2,000.	138
Tabla 124. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2022 (hasta 17/Sep).	139
Tabla 125. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2022 (hasta 17/Sep): solo el segmento de tasas entre 0 y 2,100.	139
Tabla 126. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio desde 2020 hasta 17/Sep/2022.	140

Tabla 127. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio desde 2020 hasta 17/Sep/2022: solo el segmento de tasas entre 0 y 6,000.	141
Tabla 128. Tasas de defunción por COVID-19 por municipio en el año 2020.	142
Tabla 129. Tasas de defunción por COVID-19 por municipio en el año 2020: solo el segmento de tasas de entre 0 y 200.	143
Tabla 130. Tasas de defunción por COVID-19 por municipio en el año 2021.	14
Tabla 131. Tasas de defunción por COVID-19 por municipio en el año 2021: solo el segmento de tasas de entre 0 y 150.	14
Tabla 132. Tasas de defunción por COVID-19 por municipio en el año 2022 (hasta 17/Sep).	145
Tabla 133. Tasas de defunción por COVID-19 por municipio en el año 2022 (hasta 17/Sep): solo el segmento entre 0 y 30.	146
Tabla 134. Tasas de defunción por COVID-19 por municipio desde 2020 hasta 17/Sep/2022.	147
Tabla 135. Tasas de defunción por COVID-19 por municipio desde 2020 hasta 17/Sep/2022: solo el segmento de tasas entre 0 y 6,000.	147
Tabla 136. Resumen de los modelos de árboles clasificadores producidos con datos de COVID-19 y datos sociodemográficos con nivel de agregación municipal, usando conjuntamente los datasets MuniCOVID-2022 y MuniMex-2020 (<i>MuniCovid_2022_y_munici_mex_2020_2023_06_20_b2</i>).	166

Índice de figuras

Figura 1. Ejemplo de árbol clasificador producido con un algoritmo de aprendizaje automático supervisado.	40
Figura 2. Fecha de ingreso al establecimiento de salud.	55
Figura 3. Fecha de inicio de síntomas.	55
Figura 4. Fecha de defunción.	56
Figura 5. Fecha de ingreso al establecimiento de salud: solo en casos positivos de COVID.	75
Figura 6. Fecha de inicio de síntomas: solo en casos positivos de COVID.	75
Figura 7. Fecha de defunción: solo en casos positivos de COVID.	76
Figura 8. Casos positivos por mes de ingreso.	117
Figura 9. Conteos de defunciones por año y mes.	127
Figura 10. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).	151
Figura 11. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).	152
Figura 12. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).	152
Figura 13. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).	153

Figura 14. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).	154
Figura 15. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).	154
Figura 16. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).	155
Figura 17. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).	155
Figura 18. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).	156
Figura 19. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).	156
Figura 20. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).	157
Figura 21. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).	158
Figura 22. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).	158
Figura 23. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).	159
Figura 24. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).	159
Figura 25. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).	160
Figura 26. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).	161
Figura 27. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).	161
Figura 28. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).	162
Figura 29. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).	163

AGRADECIMIENTOS

Sergio Rafael Coria Olguín agradece a la Universidad de la Sierra Sur (UN SIS), ubicada en Miahuatlán de Porfirio Díaz, Oaxaca, México, perteneciente al Sistema de Universidades del Estado de Oaxaca (SUNEO); al Sistema Nacional de Investigadoras e Investigadores (SNI), perteneciente al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), y a la Universidad Autónoma Benito Juárez de Oaxaca (UABJO) por el apoyo y las facilidades otorgadas para la realización de este libro.

PREFACIO

Para evitar falsas expectativas en el lector, hacemos las siguientes puntualizaciones:

El volumen relativamente grande de datos analizados (más de 18 millones de registros individuales, con aproximadamente 40 atributos) constituye, al mismo tiempo, una oportunidad científica para descubrir conocimiento y un reto tecnológico para analizar la información. Con esto en mente, los autores hemos hecho un considerable esfuerzo por identificar y destacar los patrones de tipo multivariable más relevantes que fueron descubiertos y representados como reglas clasificadoras. Sin embargo, humildemente reconocemos que quizá en algunos casos la cantidad y complejidad de los patrones hallados podría parecer abrumadora para algunos lectores. Por ello, ofrecemos nuestros *datasets* y modelos de aprendizaje automático completos para su descarga gratuita de la Web con el propósito de estimular y facilitar un mayor y más profundo análisis y entendimiento de estos patrones.

No se plantea ninguna hipótesis respecto a los fenómenos implícitos en los datos para ser demostrada o refutada. Esto puede resultar confuso o decepcionante para algunos lectores; no obstante, este modo de hacer investigación obedece al alcance descriptivo y exploratorio que adoptamos, ya que consideramos que la Base de Datos de COVID-19 de la Secretaría de Salud (SSA), del Gobierno de México, no ha sido suficientemente estudiada y tenemos la posibilidad de atraer la atención de quienes se interesan en este tema hacia patrones poco conocidos de este fenómeno.

Nuestra aproximación exploratoria y descriptiva está guiada por la metodología llamada Descubrimiento de conocimiento en bases de datos (Knowledge discovery in databases, KDD), de Fayyad et al. (1996), basada en el uso de técnicas de estadística y aprendizaje automático sobre grandes conjuntos de datos para descubrir patrones no triviales que tengan alguna utilidad teórica o práctica. Deseamos que nuestros hallazgos ofrezcan algún conocimiento novedoso al lector acerca de la pandemia

PREFACIO

de COVID-19 en México o, al menos, convaliden los resultados obtenidos por otros autores o por fuentes oficiales y que sirvan para inspirar nuevas hipótesis originadas en la evidencia cuantitativa para desarrollar nuevas investigaciones con diversos enfoques, alcances y metodologías.

Respecto a las aplicaciones prácticas del contenido y de los subproductos de este libro, ofrecemos al lector diversos recursos: datasets, tanto de casos individuales como de municipios; *scripts* (código fuente) en lenguaje de programación R que constituyen en sí mismos (y pueden volver a producir) modelos de árbol clasificador; e imágenes de mapas digitales. Todo ello puede ser aprovechado por los lectores para propósitos tales como: producir *software* de escritorio, aplicaciones móviles (*apps*) o sitios web y nuevos proyectos de investigación básica o aplicada. Un uso muy accesible es el repetir nuestros experimentos de ciencia de datos usando estos datasets y estos modelos para propósitos de verificación y para la enseñanza en estudios de licenciatura o posgrado en las áreas de computación, bioestadística, epidemiología, etc.

Los autores.

GLOSARIO

Casos negativos: casos registrados en la base de datos de COVID-19, de la Secretaría de Salud (SSA), del Gobierno Federal de México, que no cumplen con los criterios de caso confirmado de COVID-19. En esa base de datos, los casos negativos están designados con el identificador 7 en el atributo de *Clasificación final*. Sin embargo, existen además los identificadores 4, 5 y 6 que significan *inválido por laboratorio, no realizado por laboratorio, y caso sospechoso*, respectivamente.

Casos positivos: casos registrados en la base de datos de COVID-19, de la SSA, que cumplen con los criterios de caso confirmado de COVID-19 y que son elegibles para realizar análisis y estimar modelos en nuestra investigación. En esa base de datos, los casos positivos están designados con los identificadores 1, 2 y 3 en el atributo denominado *Clasificación final*, cuyos significados son, respectivamente, *confirmado por asociación clínica epidemiológica, confirmado por comité de dictaminación, y confirmado con muestra de laboratorio o prueba antigénica*.

COVID-19: enfermedad sistémica ocasionada por el virus SAR-CoV-2, que se manifiesta por infección respiratoria aguda.

Dataset: conjunto de datos en un archivo digital, que generalmente se encuentran organizados en forma de tabla; por ejemplo, en tablas de archivos de tipo CSV (*comma separated values*) o de tipo Excel, entre otros.

Endemia: ocurrencia de casos de una enfermedad dentro de los límites de cantidad de casos esperados de esa enfermedad en un periodo y territorio determinados.

Epidemia: ocurrencia de casos de enfermedad mayor al número de casos esperados de esa enfermedad.

Granularidad: nivel de detalle en la representación de información cuantitativa; por ejemplo, la cantidad de defunciones por COVID-19 puede tener nivel de detalle por municipio, o bien, por entidad federativa, o bien, nacional.

GLOSARIO

Pandemia: epidemia que abarca grandes extensiones geográficas.

SARS-CoV-2: acrónimo de *Severe Acute Respiratory Syndrome Coronavirus 2*, nombre del virus que causa la enfermedad por coronavirus de 2019 (COVID-19).

Sindemia: coexistencia, durante un periodo y dentro de un territorio determinados, de dos o más epidemias que comparten factores sociales, interactúan y causan secuelas complejas.

INTRODUCCIÓN

El año 2020 puede considerarse un punto de inflexión en la posmodernidad del siglo XXI, marcó un claro antes y después a raíz de la pandemia de la COVID-19¹. La noción de pandemia alude a su definición clásica que remite a la propagación mundial de una enfermedad (OMS, 2010). Dicha acepción está cimentada desde una perspectiva epidemiológica, considerada como evento socio-natural, ya que su ocurrencia está relacionada con daños al medio natural (Bergmann, 2023; Lusk & Chandra, 2021; Alcántara-Ayala, 2021) y al ser humano, alterando su entorno inmediato y trastocando su cotidianeidad.

Es importante destacar que la pandemia de la COVID-19 fungió como disruptor, con afectación de todas las dimensiones sociales y procesos asociados a satisfacer las necesidades e intereses de las poblaciones. Su efecto incluyó las esferas económica, política, y por supuesto, de la salud humana, tanto a nivel individual como colectivo. El fenómeno se replicó a diversas escalas, desde el ámbito local, regional y estatal, con repercusiones nacionales y globales.

Con el afán de desarrollar un breve recuento previo al inicio de la pandemia de la COVID, enfatizamos algunos eventos de importancia. A finales de 2019, el Gobierno de China reportó un brote de neumonía causado por el nuevo coronavirus SARS-CoV-2, causante de la enfermedad denominada COVID-19, que hace alusión al año de su descripción ante la comunidad internacional (Wu and McGoogan, 2020; Zhang, 2020).

¹ En adelante emplearemos pandemia de la COVID-19, pandemia de la COVID, o bien pandemia, de manera indistinta, con el fin de hacer más fluida la lectura del documento. Entiéndase que los términos anteriores hacen referencia a la pandemia causada por la enfermedad por coronavirus (COVID-19), cuyo agente causal se identifica como el virus SARS-CoV-2.

Posteriormente, a finales de enero de 2020, la Organización Mundial de la Salud (OMS) declaró que debido a los diversos brotes de neumonía² por COVID-19 asociados con un alto índice de contagios, deberían considerarse dichos eventos en su conjunto como una emergencia de salud pública de importancia internacional (ES-PHI). Se estimó que una persona infectada con el virus SARS-CoV-2 podría infectar a otras personas con solo establecer algún tipo de intercambio social con una temporalidad de al menos diez minutos, o que dicha interacción se desarrollase a una distancia corta (menor a dos metros); por la convivencia en espacios cerrados –lo cual favorecía mayores concentraciones del virus al interior–, o por el contacto físico directo, como por medio de un saludo u otro tipo de interacciones sociales.

Se determinó que los aerosoles presentes durante la interacción social son un factor común en la transmisión del virus. Por ello, se subrayó la importancia de adoptar medidas básicas de higiene, que incluyen: el uso de mascarillas faciales y cubrebocas, mantener el distanciamiento físico, evitar saludos y cualquier contacto físico directo, lavado de manos con regularidad y aislar a la persona al presentar síntomas relacionados con la enfermedad. Estas medidas, recomendadas por organismos internacionales, fueron adoptadas y promovidas por las autoridades sanitarias alrededor del mundo para contener la propagación del virus (OMS, 2020 a; OMS, 2020b). Además de lo anterior, durante el primer año de la pandemia en 2020 se promovió el confinamiento de población considerada vulnerable³ de contraer COVID-19 (Gobierno de México, 2021).

Los vuelos comerciales favorecieron la diseminación a nivel global a través de la importación de casos de viajeros internacionales (Lau *et al.*, 2020). El 11 de marzo de 2020 la OMS declaró la pandemia de la COVID-19 (OMS, 2020c), y definió además cuatro fases (o escenarios) para describir la dinámica de su transmisión al interior de los países (OMS, 2020d). En el caso de México, la Secretaría de Salud Federal (SSA) consideró tres fases de transmisión (Gobierno de México, 2020). La Tabla 1 muestra el comparativo; la diferencia radica en que la OMS contempla como primera fase la ausencia de casos.

Tabla 1. Fases de la dinámica de la transmisión de la pandemia de la COVID-19: comparativo según la OMS versus la SSA.

Fases según la OMS	Fases según la SSA
Fase 1. Nulidad de casos	Fase 1. Presencia de casos importados
Fase 2. Presencia de casos importados	Fase 2. Transmisión comunitaria en zonas geográficas.
Fase 3. Presencia de casos agrupados a una zona geográfica y/o exposición común	Fase 3. Etapa epidemiológica, con presencia en todo el territorio nacional
Fase 4. Presencia de grandes brotes por transmisión comunitaria	

Fuente: Elaboración propia.

2 Brotes en 19 países, 30 de enero, 2020.

3 Se consideraron como criterios de vulnerabilidad: el embarazo, lactancia materna, obesidad, población mayor de 60 años, enfermedades crónico-degenerativas, así como la combinación de comorbilidades.

INTRODUCCIÓN

En el contexto mexicano, a finales de marzo de 2020 se decretó el denominado inicio de la fase 2 de la pandemia de la COVID-19 por la dispersión comunitaria existente. Las autoridades federales establecieron como primera estrategia la denominada “Jornada de Sana Distancia”⁴, con promoción del confinamiento físico de la población, suspensión de las actividades escolares y laborales de los sectores público, social y privado con el fin de evitar la concentración física de personas. Solo continuaron laborando aquellos establecimientos que proveyeran servicios básicos, como el caso de hospitales, farmacias, telecomunicaciones, servicios financieros, gasolineras, supermercados, siempre y cuando se evitasen aglomeraciones y se fomentara la ventilación de los espacios físicos. Además, se promovió las medidas básicas de higiene señaladas por organismos internacionales replicándose en el territorio nacional, tanto en el ámbito estatal, municipal y local, con apoyo del sector público, encabezado por la totalidad de instituciones que integran al sistema sanitario, así como el sector educativo y el resto de las entidades gubernamentales, incluyendo al sector privado, así como la sociedad civil en general de las diversas medidas (Gobierno de México, 2020).

Datos abiertos referentes a la pandemia en México

Según cifras oficiales del Gobierno de México, durante el primer año (2020) de la pandemia de la COVID-19 se confirmaron 1,426,094 casos y 125,807 defunciones, con una tasa de incidencia de casos acumulados de 1,115.9 por cada 100,000 habitantes (Secretaría de Salud, 2020a)⁵. En 2021, al segundo año de la pandemia, se confirmaron 3,979,723 casos acumulados y 299,428 defunciones acumuladas, con una tasa de incidencia de casos acumulados de 3,085.7 por cada 100,000 habitantes (Secretaría de Salud, 2021)⁶. Al tercer año, se confirmaron 7,222,611 casos totales y 331,030 defunciones totales, con una tasa de incidencia de casos acumulados de 5,550.8 por cada 100,000 habitantes (Secretaría de Salud, 2022c)⁷.

La Secretaría de Salud Federal, a través de la Dirección General de Epidemiología, puso a disposición, a través de su plataforma de datos abiertos, la información relativa a los casos asociados a COVID-19⁸ provista por el Sistema de Vigilancia Epidemiológica de la Enfermedad Respiratoria Viral (Sisver). El Sisver instrumenta la detección oportuna de casos, toma y manejo de muestras de casos sospechosos, así como acciones de monitoreo en torno a la dinámica epidemiológica de enfermedades respiratorias virales a nivel nacional e internacional (Secretaría de Salud, 2020b).

Las acciones de vigilancia de la enfermedad respiratoria viral se sustentan en: a) Vigilancia centinela, que es desarrollada en 475 unidades de salud monitoras de enfermedad respiratoria viral (denominadas USMER) dentro del territorio nacional, donde realizan procesos muestreo y procesamiento de las mismas de aquellos casos sospechosos; b) Al resto de unidades médicas que no toman muestras se les conoce como unidades no USMER. Las unidades no USMER también proveen información

4 La Jornada Nacional de Sana Distancia tuvo una duración de 69 días de confinamiento

5 Cifras al corte del 31 de diciembre de 2020 a las 19:00hs.

6 Cifras al corte del 29 de diciembre de 2021, (informe del 31 de diciembre, 2021 a las 17:00 hs).

7 Cifras al corte del 26 de diciembre de 2021, (informe del 27 de diciembre, 2022 a las 17:00 hs).

8 Disponible a partir del 14 de abril del 2020, considerando información desde el 1 de enero del mismo año

sobre vigilancia de casos sospechosos y los refieren a unidades USMER para su confirmación. En ambos casos, tanto las unidades USMER y no USMER reportan aquellos casos de sospecha de enfermedad respiratoria viral⁹ en la plataforma del Sisver. Además de la confirmación por laboratorio, en los registros también se incluyeron aquellos casos confirmados por asociación clínica epidemiológica o bien por confirmación realizada por un Comité de dictaminación. En el caso de las defunciones a causa de COVID-19, estas son reportadas en el Sistema Epidemiológico y Estadístico de las Defunciones (SEED).

En los primeros meses de la pandemia en México, Carrillo-Vega et al. (2020) ofrecen una estimación temprana de los factores de riesgo para hospitalización y mortalidad por COVID-19 en este país. Su estudio encontró que las personas de 50 a 74 años y mayores de 75 tenían más probabilidades de morir por COVID-19 que aquellas de 25 a 49 años, ya que la edad jugó un papel significativo entre quienes enfermaron de COVID-19, siendo las personas mayores quienes enfrentaron un mayor riesgo de complicaciones graves, hospitalización y mortalidad en comparación con las personas más jóvenes.

Los datos provistos por el Sisver y el SEED son el sustrato principal de nuestro análisis. La Dirección General de Epidemiología de nivel federal dispuso el acceso libre a las bases de datos abiertos de COVID-19 en ambas plataformas. La información de los registros de dicha base solo contempla aquellos casos de quienes acudieron a solicitar atención médica dentro del sector salud. A raíz de lo anterior, se evidencia un subregistro en los casos no atendidos por el sistema sanitario público ni del sistema de medicina privada. Un claro ejemplo de este subregistro sería por aquellos individuos que enfermaron, pero nunca obtuvieron confirmación diagnóstica oficial dentro del sistema de salud. Esto podría deberse a que optaron por pruebas de diagnóstico rápido de bajo costo o simplemente decidieron autoatenderse en sus hogares sin buscar atención médica formal. Otra fracción del subregistro lo conforman los casos leves de la enfermedad. Este fenómeno de subregistro no es exclusivo de nuestro país; se observó en diversas naciones. Por ello, es probable que nunca tengamos una estimación precisa de la verdadera magnitud de la pandemia a nivel nacional e internacional. Por lo anterior, algunas incursiones de análisis han considerado la estimación del exceso de *mortalidad* que considera el número absoluto de muertes ocurridas y corresponde al acumulado de muertes¹⁰ para un determinado periodo de tiempo, como una posible aproximación más realista (Añorve, 2021; Fuentes et al., 2022).

De acuerdo con el análisis desarrollado por la Comisión Independiente de Investigación sobre la Pandemia de COVID-19 en México (Sepúlveda, 2024), la esperanza de vida en México en el periodo 2019-2021 se redujo en 4.6 años, lo que implica un retroceso equivalente a tres décadas. Esto se explica por la alta mortalidad y muertes prematuras entre adultos jóvenes, considerando que entre los grupos de 35-44 se incrementó a 47% y, en el de 45-54 años, en 70%, respectivamente. Si bien existieron políticas públicas que promovieron el distanciamiento físico y el confinamiento,

9 Considerada influenza, COVID-19 y otras infecciones respiratorias.

10 Definido por el número de defunciones observadas menos el número de defunciones esperadas según padecimiento.

INTRODUCCIÓN

la economía de las familias mexicanas promedio limitó el confinamiento total. La Comisión destaca que la política promocionada por autoridades gubernamentales denominada “Quédate en casa”, promoviendo el confinamiento, favoreció que una gran mayoría de los casos de COVID-19 demorasen su atención inicial, acudiendo solo los casos de mayor gravedad. Inclusive, se promocionó que solo se acudiera a los servicios de urgencias en caso de datos de alarma respiratoria; por lo cual, se privilegió la autoatención domiciliaria en los hogares y aproximadamente un 60% de los casos fue manejado en consultorios anexos a farmacias privadas. Como se mencionó anteriormente, todos estos casos no están contabilizados dentro de la estadística oficial de COVID-19, por lo que en México (y el resto del mundo), muy probablemente nunca se logre estimar con precisión la totalidad de casos ocurridos.

Del porqué de nuestro libro y un breve recuento de este

En el capítulo 1 describimos el porqué del presente libro. Destacamos en primer lugar que la investigación desarrollada en la presente obra persigue identificar patrones estadísticos a partir de la información oficial en torno a los casos reportados de COVID-19 en México, desde el inicio de la pandemia hasta septiembre de 2022. El estudio se apoyó del análisis de la base de datos de COVID-19 de la Secretaría de Salud del Gobierno Federal, explorando también las posibles asociaciones con variables sociodemográficas y de vivienda, utilizando información del Censo Nacional de Población y Vivienda 2020.

El objetivo principal se centró en explorar patrones estadísticos significativos en relación con los casos positivos de COVID-19, considerando la granularidad individual y municipal. Para ello, se identifican y se describen tanto comportamientos estadísticos, como aquellos posibles patrones de asociación tomando en consideración ambos niveles de granularidad, todo ello con apoyo de algoritmos de aprendizaje automático. Se utilizó aprendizaje automático supervisado para descubrir y representar patrones de correlación en la base de datos de la SSA, empleando algoritmos de árboles clasificadores. Con el fin de que la presente investigación sirva de precedente para futuras incursiones, el material correspondiente a los *datasets* utilizados en la presente obra están a disposición del público en general y el público especializado, con acceso gratuito.

En el capítulo 2 se realiza un exhaustivo estudio de la estadística descriptiva aplicada a los casos confirmados de COVID-19 en México. Se utiliza principalmente el enfoque univariado, que busca analizar las características de una sola variable, por cada uno de los análisis desarrollados, a considerar 56 variables provenientes de la base de datos de COVID-19 de la Secretaría de Salud Federal. La técnica de Pareto es empleada para identificar los principales factores que están influyendo en la propagación del virus. Este análisis proporciona información valiosa para comprender el comportamiento general de la pandemia en el país. Nos permite identificar patrones y tendencias en la evolución de los casos confirmados, así como determinar las áreas geográficas más afectadas y las edades y grupos de población más vulnerables.

El capítulo 3 se enfoca únicamente al estudio de los casos positivos de COVID-19, considerando aproximadamente 7,000,000 de casos y 56 variables relacionadas. Si-

guiendo la metodología anterior, se utilizó la técnica de Pareto (80:20) para el análisis. El propósito principal de este análisis descriptivo es identificar patrones generales de los datos y proporcionar una base sólida para futuras evaluaciones y la creación de modelos utilizando técnicas de aprendizaje automático.

El capítulo 4 se dedica a identificar patrones de correlación multivariable en registros individuales utilizando técnicas de aprendizaje automático. Los patrones se observan en cuatro conjuntos de datos derivados del conjunto de datos principal llamado “*Master-3*”, que contiene más de 18,000,000 de registros de casos de COVID-19, positivos y negativos, provenientes de la base de datos de la SSA. Se analizan cuatro subconjuntos específicos:

- i. Solo casos positivos de COVID-19 (dataset “Posit”).
- ii. Registros de mujeres que incluyen tanto casos positivos como negativos (dataset “Mujeres”).
- iii. Registros de hombres (dataset “Hombres”).
- iv. Casos positivos de mujeres (dataset “PositMujeres”).

Consideramos importante garantizar el acceso universal al conocimiento, por lo que estos conjuntos de datos, junto con el conjunto de datos principal, están disponibles para su descarga. La técnica de aprendizaje automático utilizada es la de árboles clasificadores, utilizando el algoritmo C5.0 implementado en el lenguaje de programación R. La elección de trabajar con subconjuntos específicos del conjunto de datos *Master-3* se debió a limitaciones en la infraestructura informática y a la intención de resaltar patrones clínicos y sociodemográficos específicos. Este capítulo proporciona una visión a profundidad de los patrones de correlación multivariable en registros individuales de COVID-19 utilizando técnicas de aprendizaje automático. Aunque algunos patrones pueden ser ya conocidos por los profesionales de la salud, otros podrían ofrecer nuevos conocimientos y ser la base para futuras investigaciones.

El capítulo 5 se adentra en el análisis detallado de la pandemia de COVID-19 en México desde una perspectiva municipal. A diferencia de los enfoques más generalizados, este capítulo se sumerge en el nivel micro, permitiendo una comprensión más precisa y localizada de la situación. La estructura del capítulo se organiza en tres secciones esenciales:

- i. En la primera sección, se introduce el conjunto de datos principal utilizado en el capítulo llamado MuniCovid-2022. Este conjunto de datos ha sido específicamente curado para este análisis y se deriva de la base de datos principal de la Secretaría de Salud (SSA) del año 2022. Además, se enriqueció con proyecciones poblacionales de la Consejo Nacional de Población (Conapo) del año 2019. Esta combinación de datos proporciona una visión rica y multidimensional de la situación a nivel municipal en México. Es importante destacar que este conjunto de datos está disponible públicamente y aquellos que deseen utilizarlo deben citar este libro como referencia.
- ii. La segunda sección se centra en el análisis de los conteos municipales. A través de la estadística descriptiva básica, se obtiene una visión directa y cuantitativa de la prevalencia y distribución de COVID-19 en diferentes municipios.

INTRODUCCIÓN

Se realizan conteos o frecuencias absolutas de los casos en cada municipio, lo que permite tener una idea clara de la situación.

- iii. La tercera sección ofrece un enfoque más normalizado, presenta tasas municipales por cada 100,000 habitantes. Este análisis ajusta los datos por la población, lo que proporciona una imagen más equitativa y comparable de la prevalencia del virus en diferentes regiones, independientemente de su tamaño o densidad poblacional.

El capítulo 5 se destaca por su enfoque completamente cuantitativo y descriptivo, proporcionando una base sólida y detallada sobre la situación de COVID-19 a nivel municipal en México. Aunque puede parecer técnico y detallado, esta profundidad es esencial, ya que no solo ofrece una comprensión más rica y localizada del impacto de la pandemia, sino que también sienta las bases para modelos de aprendizaje automático que se explorarán en capítulos posteriores. En esencia, este capítulo proporciona el fundamento estadístico sobre el cual se pueden construir análisis más avanzados y predictivos en el futuro.

El capítulo 6 se adentra en la viralización específica de la pandemia de COVID-19 en México, proporcionando una perspectiva espacial y temporal de la enfermedad desde principios de 2020 hasta el 17 de septiembre de 2022. A través de mapas detallados, con una granularidad enfocada a nivel municipal, el capítulo ilustra un panorama completo de la propagación y el impacto de la enfermedad. Se utilizó información precisa para trazar estos mapas: las tasas de casos positivos y las tasas de defunción, ambas expresadas por cada 100,000 habitantes de cada municipio mexicano. Esta información se extrajo de la base de datos oficial de la Secretaría de Salud (SSA, 2022) y se complementó con estimaciones de población municipal de los años 2020 a 2022, facilitadas por el Consejo Nacional de Población (Conapo, 2019). Los datos precisan el avance mensual de la pandemia y están disponibles para ser descargados. Este capítulo introduce dos tipos de mapas principales para representar estos datos: mapas coropléticos y mapas de calor. Los primeros, mapas coropléticos, utilizan diferentes colores para representar áreas geográficas basadas en el número total de casos o defunciones. Estos son particularmente útiles para identificar y resaltar las regiones con mayor incidencia de la enfermedad. Por otro lado, los mapas de calor, utilizando una gradación de colores o tonos de gris, muestran la intensidad de los casos, revelando áreas de alta concentración y cómo estas áreas evolucionaron a lo largo del tiempo.

El capítulo 6 considera la importancia de elegir el tipo de mapa adecuado según el propósito del análisis. Mientras que los mapas coropléticos son excelentes para mostrar diferencias proporcionales entre áreas, los mapas de calor son ideales para resaltar la intensidad o concentración de casos en regiones específicas. Una observación clave presentada es la preferencia por utilizar tasas en lugar de frecuencias absolutas al representar datos de COVID-19. Las tasas ofrecen una comparación equitativa de la incidencia de la enfermedad entre diferentes áreas, desvinculando el tamaño de la población del número de casos. Esto permite identificar de manera efectiva las áreas más afectadas en relación con su población, proporcionan una visión más ajustada y realista de la incidencia. En conclusión, este capítulo ofrece una poderosa herramienta virtual para comprender y analizar la pandemia de COVID-19

en México, resalta áreas críticas y brinda mayor comprensión sobre la propagación de la enfermedad a lo largo del tiempo y espacio.

El capítulo 7 se sumerge en una profunda exploración de los patrones multivariantes relacionados con la presencia de COVID-19 en los más de 2,400 municipios de México. Este análisis conecta las características sociodemográficas de cada municipio con la prevalencia de la enfermedad, proporcionando ideas sobre cómo diversos factores sociodemográficos pueden influir en la propagación y el impacto de la pandemia. Aunque este capítulo comparte similitudes con el capítulo 4 en términos de metodología, la gran diferencia radica en el nivel de análisis: mientras que el capítulo 4 se centra en casos individuales, el capítulo 7 se concentra en el nivel municipal.

Los datos para este análisis provienen de dos conjuntos de datos principales: MuniCOVID-2022 (detallado en el capítulo 5) y MuniMex-2020, que es un reflejo del Censo Nacional de Población y Vivienda 2020. Este último conjunto de datos, MuniMex-2020, es especialmente valioso ya que proporciona porcentajes detallados de habitantes y viviendas con relación a una amplia gama de características, desde demográficas hasta infraestructura y servicios básicos. Estos conjuntos de datos se combinaron para crear un archivo de datos unificado, que está disponible para el público en general. El análisis central de este capítulo se presenta en la tabla 136 (*Resumen de los modelos de árboles clasificadores producidos con datos de COVID-19 y datos sociodemográficos con nivel de agregación municipal*), que muestra una visión general de los modelos producidos. Estos modelos se centran en intervalos de tasas municipales por cada 100 mil habitantes y, de este modo, permiten comparaciones equitativas entre municipios de diferentes tamaños. Estos intervalos de tasa, denominados “t_0_a_100”, representan el número de casos (ya sean positivos o defunciones) por cada 100,000 habitantes. Es crucial mencionar que, aunque el capítulo 5 presenta algunos de estos intervalos, en este capítulo se realizaron ajustes para ofrecer una distinción más clara entre las características sociodemográficas de los municipios en diferentes subintervalos.

El enfoque en representar las tasas de forma nominal es fundamental para este análisis. Esto permite la creación de modelos de árbol clasificador, donde se puede identificar qué características comparten los municipios con tasas similares. Estos modelos, por consiguiente, ofrecen una perspectiva detallada sobre cómo determinados factores sociodemográficos pueden estar correlacionados con tasas altas, medias o bajas de COVID-19 en diferentes regiones. En resumen, el capítulo 7 proporciona una visión integral de cómo las características sociodemográficas de los municipios de México están interconectadas con la prevalencia y el impacto de la COVID-19. Estos hallazgos son esenciales no solo para comprender mejor la propagación y las implicaciones de la enfermedad, sino también para orientar futuras políticas y estrategias dirigidas a áreas específicas basadas en sus perfiles sociodemográficos.

Para concluir, el capítulo 8 presenta comentarios finales, recomendaciones y trabajo de investigación a futuro. Con todo lo anterior, aspiramos a que el presente libro proporcione un análisis inédito sobre la pandemia en México, aportando evidencia para el diseño e implementación de políticas públicas y mejoras en la atención sanitaria.

CAPÍTULO I

Diseño metodológico

1.1 La problemática

El problema de investigación que se aborda en este libro surge de la necesidad de reconocer aquellos patrones estadísticos relevantes que pudieran existir en la información disponible en las bases de datos públicas de los casos registrados de COVID-19 en México. Incluye la información registrada por la Secretaría de Salud desde el inicio de la pandemia, durante los primeros meses del año 2020, hasta septiembre de 2022.

Consideramos de importancia explorar algunos aspectos sociodemográficos, así como de índole económica y del contexto político; por ejemplo, las características de la población y de sus viviendas, los sectores de actividades económicas predominantes en las diversas áreas geográficas y los partidos políticos de adscripción de los gobernantes locales.

Mientras se elaboraba este libro, durante el primer cuatrimestre de 2023, las autoridades de salud en México señalaban ya que la pandemia se encontraba en su etapa final. Por ello, esta investigación solamente analizó los datos de casos de COVID-19 correspondientes al periodo de 2020, 2021 y hasta el 17 de septiembre de 2022 (fecha en que realizamos la descarga de la base de datos COVID-19 de la Secretaría de Salud, del Gobierno Federal).

1.2 Objeto de estudio

En sentido amplio, el objeto de estudio de esta investigación es la pandemia de COVID-19 en México y sus posibles interacciones con fenómenos económicos y sociodemográficos en el periodo de enero de 2020 a septiembre de 2022. En forma específica, el objeto de estudio es la información disponible en la base de datos de casos de COVID-19 publicada por la Secretaría de Salud (SSA) del Gobierno Federal de México desde febrero de 2020 (SSA, 2022).

Es importante precisar que nuestros análisis se centran en los casos registrados por la estadística oficial. Los casos no reportados por esta –por ejemplo, aquellos de quienes enfermaron y pudieran haberse realizado una prueba de autodiagnóstico rápido en su propio domicilio, correspondientes a casos positivos– resultan incommensurables para el objeto de nuestro estudio. Tanto en el contexto mexicano, como a nivel mundial, resulta imposible su cálculo aproximado.

Como complemento a ello, nos permitimos también estudiar las potenciales asociaciones o interacciones existentes con diversas variables de tipo sociodemográfico de los habitantes y de sus viviendas en municipios donde existieron casos de COVID-19; por ejemplo: presencia de servicios públicos en la vivienda, sexo y edad de los habitantes, niveles educativos, etc. Obtuvimos la información de las siguientes fuentes:

-Censo Nacional de Población y Vivienda 2020 (INEGI, 2021),

-Proyecciones de la población de los municipios de México 2015-2030 (Conapo, 2019).

1.3 Objetivos de investigación

El principal objetivo de esta investigación radica en explorar aquellos patrones estadísticos relevantes respecto a los casos positivos (con clasificación final de tipos 1, 2 o 3 en la base de datos de la SSA) de COVID-19 en México, registrados en materia de salud pública, en nuestro periodo de interés (de los primeros meses de 2020 a septiembre de 2022). Se consideran dos niveles de granularidad de los datos: individual, y municipal. Los patrones individuales son aquellos existentes en el *dataset* de la SSA que pueden descubrirse sin realizar ningún tipo de agregación del mismo, donde cada renglón representa a un individuo. En cambio, los patrones municipales se descubren una vez que se han agregado conteos de algunas variables de los municipios, de modo que cada renglón del *dataset* representa a un municipio de México.

Por otra parte, también nos enfocamos en la búsqueda de patrones que pudiesen existir en el nivel municipal respecto a los distintos datos sociodemográficos, económicos y políticos de los propios municipios. Es importante aclarar que no se persigue la comparabilidad entre el nivel individual y la agregación municipal, ya que incurriríamos en un grave error. Solo nos permitimos la comparabilidad de individuos con individuos y de municipios con otros municipios.

Para lograr nuestro objetivo principal, nos propusimos alcanzar los objetivos específicos que enumeramos a continuación, respecto a cada uno de los dos niveles de granularidad de los datos:

- i. Granularidad individual
 - Identificar y describir el comportamiento estadístico univariado de subconjuntos de variables de interés especial de la base de datos de registros de casos de COVID-19 de la SSA.
 - Identificar y describir patrones de asociación entre variables múltiples, a nivel individual, dentro de la base de datos de registros de casos de COVID-19 de la SSA mediante algoritmos de aprendizaje automático.

ii. Granularidad municipal

- Identificar y describir patrones cronológicos y tendencias de la pandemia en los municipios mexicanos, con información de la base de datos de registros de casos de COVID-19 de la SSA, con nivel de agregación mensual.
- Identificar y describir patrones de asociación entre variables múltiples de la base de datos COVID-19 de la SSA, con nivel de agregación municipal mensual vinculados a datos sociodemográficos, económicos y políticos, obtenidos de otras fuentes gubernamentales.
- Representar patrones geográficos y temporales de la pandemia por COVID-19 entre municipios mediante mapas digitales.

1.4 Necesidad y utilidad de nuestro libro

Consideramos que la investigación que respalda la presente obra es necesaria porque, al momento de escribir este libro, no existía un análisis similar acerca de la pandemia de COVID-19 en México. Este ejercicio sienta un precedente para futuras incursiones. Consideramos que, dada la magnitud de la pandemia de la COVID-19 y sus efectos a distintos niveles, no solo en materia de salud, existe la necesidad de conocer, con propósitos preventivos, con una mayor profundidad y detalle, los patrones existentes en este fenómeno.

La utilidad de esta investigación radicó en generar evidencia que sirviera de sustento para el diseño e implementación de políticas públicas, y planeación de programas en el sector salud y, en consecuencia, en mejoras en la atención de la salud de enfermos de COVID-19. Podemos afirmar que la presente obra es original por las razones siguientes:

- i. Se enfocó en el análisis de la información correspondiente al ámbito nacional.
- ii. Se consideraron dos unidades de análisis estadístico y modelación apoyadas con aprendizaje automático:
 - individuos (casos registrados de COVID-19), y
 - agregación de casos a nivel de municipios (o alcaldías, en el caso de la Ciudad de México),
- iii. Entre las técnicas de análisis y modelación, se utilizaron algunas de aprendizaje automático; específicamente, los árboles clasificadores, con el propósito de descubrir y representar patrones de asociación entre conjuntos de variables.

Este estudio fue factible porque tomó como materia prima la información gratuita disponible en modalidad de datos gubernamentales abiertos; también, porque los análisis y modelos de los datos se producen con herramientas de software libre. La factibilidad también fue posible por la participación de profesionales en salud pública y en ciencias de computación.

1.5 Enfoque y alcance

El enfoque de esta investigación fue de tipo cuantitativo, con un alcance descriptivo, exploratorio y correlacional, con el uso de técnicas de estadística descriptiva y de aprendizaje automático. Los alcances se centran en describir y analizar el fenómeno

de la pandemia de COVID-19 en México a partir de la búsqueda de posibles correlaciones de interés entre diversas variables, explorando algunas de sus interacciones con cuestiones de índole sociodemográfica, económica y política.

El alcance descriptivo y exploratorio de esta investigación obedeció a que la pandemia de la COVID-19 en México es un fenómeno relativamente reciente con estudio insuficiente de sus interacciones sociales. Además, la aplicación de técnicas de aprendizaje automático a la base de datos de COVID-19 de la SSA nos permitió descubrir patrones correlacionales de comportamiento entre las variables involucradas en la pandemia que pudieran ser relativa o completamente nuevos para los especialistas del tema.

1.6 Supuestos

Si bien, consideramos no plantearnos ninguna hipótesis, partimos de los siguientes supuestos:

- i. Que existen diversos patrones de tipo univariado y multivariado en los niveles de agregación individual y municipal dentro de la base de datos de COVID-19 de la SSA.
- ii. También, suponemos la existencia de posibles patrones de correlación de estos datos con respecto a variables sociodemográficas, económicas y políticas disponibles en otras fuentes de datos gubernamentales.
- iii. Todos estos patrones pudieran ser factibles de ser identificados y representados mediante técnicas de estadística, aprendizaje automático; en particular, los mapas digitales sirven para representar patrones del nivel de agregación municipal.

1.7 Diseño de la investigación

El diseño del estudio es Análisis Secundario de Datos (Clarke and Cosette, 2000). La base de datos inicial permitió la perspectiva longitudinal porque se analizaron los datos en diferentes momentos de la pandemia, y el análisis de algunos momentos específicos fue transeccional porque se midió simultáneamente la variable de interés y las variables predictoras de todo el conjunto de casos de COVID-19 del periodo señalado para descubrir y representar patrones de asociación entre esas variables. El Análisis Secundario de Datos fue organizado en las etapas que se enumeran a continuación:

- i. Elaboración del protocolo de investigación
- ii. Aprobación del protocolo de investigación por el Comité de Investigación de la Universidad de la Sierra Sur (UNSI)
- iii. Identificación de la base de datos de casos registrados de COVID-19 de la SSA, base de datos gubernamental abierta en la web
- iv. Identificación de bases de datos de fuentes gubernamentales con información agregada a nivel municipal
- v. Realización de análisis estadísticos descriptivos de la base de datos de casos registrados de COVID-19, a nivel individual

- vi. Descubrimiento de patrones de asociación entre diversas variables de la base de datos de casos de COVID-19, a nivel individual
- vii. Realización de análisis estadísticos descriptivos de la base de datos de casos registrados de COVID-19, con agregación municipal
- viii. Descubrimiento de patrones de asociación entre diversas variables de la base de datos de casos de COVID-19, a nivel municipal
- ix. Elaboración de mapas digitales

1.7.1 Recolección de la información

Debido a los objetivos y las características de esta investigación, no hubo trabajo de campo para recolección de información; fueron descargados los conjuntos de datos gubernamentales de los temas de interés y disponibles en websites. Tampoco se hizo muestreo, fueron analizados todos los registros de las bases de datos. A continuación, se enumeran los datasets requeridos y las instituciones que los ofrecen mediante sus respectivos sitios web:

- i. Base de datos de casos de casos registrados de COVID-19, de la Secretaría de Salud, del Gobierno Federal (SSA, 2022)
- ii. Censo Nacional de Población y Vivienda 2020, del Instituto Nacional de Estadística y Geografía (INEGI, 2021)
- iii. Proyecciones de la población de los municipios de México 2015-2030 (Conapo, 2019)

1.7.2 Dos tipos de unidad de análisis

En esta investigación, se realizaron análisis estadísticos y modelos de aprendizaje automático en dos tipos de unidades de análisis: individuos y municipios, o alcaldías en el caso de Ciudad de México. El primer grupo de unidad de análisis y modelos descubre y representa patrones en un *dataset* donde cada renglón es un individuo. En cambio, el segundo grupo, se enfoca en descubrir y representar patrones en un dataset donde cada renglón es un municipio, con datos agregados a partir del nivel individual. Para el nivel individual, el *dataset* utilizado es prácticamente el mismo que se publica en la base de datos de casos registrados de COVID-19 de la SSA, con algunos atributos añadidos a partir de los ya disponibles. Para el nivel municipal, se suman separadamente los diversos tipos de casos de COVID-19: positivos, negativos, intubados, hospitalizados, ambulatorios, etcétera, para obtener totales municipales. En esta investigación, se denominan casos positivos aquellos que en la base de datos de la SSA tienen asignados en el atributo de *CLASIFICACIÓN_FINAL* los valores 1, 2 o 3.

Son consideradas dos unidades de análisis para descubrir y representar patrones que tienen utilidad para propósitos distintos, aunque complementarios. La unidad de análisis individual sirve para identificar patrones de tipo clínico. La unidad de análisis municipal tuvo combinaciones consistentes con otras bases de datos de origen gubernamental que también manejan como unidad de análisis el municipio, por ejemplo, con datos sociodemográficos de los habitantes y sus viviendas, con datos de actividades económicas locales o con datos que describen las características de

los gobiernos municipales. El municipio como unidad de análisis descubrió patrones de correlación entre algunos atributos, en modo agregado, de la base de datos de casos registrados de COVID-19 de la SSA y atributos descriptivos de los municipios en aspectos sociodemográficos, económicos o políticos.

1.7.3 Las variables tiempo y territorio

Tomando en cuenta que la base de datos de COVID-19 de la SSA constituye una serie de tiempo, esta variable está representada en nuestros *datasets* mediante varios atributos: fecha de ingreso del paciente a la unidad de atención, fecha de inicio de síntomas, fecha de fallecimiento, y el número de semana epidemiológica correspondiente a cada una de estas fechas. Usando estas fechas y números de semana epidemiológica, se realizan análisis simples de estadística descriptiva (conteos, porcentajes) y de tendencias. Además, la variable tiempo (en algunas de sus diversas representaciones) se usa también en nuestros modelos de aprendizaje automático para descubrir y representar diversos tipos de correlaciones.

Además de la variable *tiempo*, el análisis de datos de una pandemia necesita tomar en cuenta variables de tipo geográfico. Estas se encuentran disponibles en la base de datos de SSA en atributos tales como el municipio y la entidad federativa de residencia del paciente, su entidad federativa de nacimiento y la entidad federativa donde se ubica la unidad médica que le dio atención. Estos atributos se analizan con estadística descriptiva (conteos y porcentajes) y se usan para producir modelos con aprendizaje automático; adicionalmente, con base en los análisis y modelos, se producen mapas digitales para ilustrar los patrones de mayor interés teórico o práctico.

1.7.4 Dos técnicas de análisis y modelación

El nombre genérico del método elegido para analizar las bases de datos mencionadas es *Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases, KDD)* (Fayyad et al., 1996). Este aplica dos tipos de técnicas de análisis y modelación: estadística descriptiva (DeSanto et al., 2015) y aprendizaje automático (Russell y Norvig, 2020; Han et al., 2011), que se aplican a grandes conjuntos de datos, denominados en inglés *data sets* o *datasets*. La razón para aplicar los dos tipos de técnicas es que la primera permite descubrir patrones generales del fenómeno y la segunda puede descubrir y representar patrones específicos con un mayor nivel de detalle, sobre todo con una perspectiva multivariable o multivariada. Estas técnicas son aplicadas por analistas con perfil profesional de computación, matemáticas o similar, colaborando en forma inter, multi o transdisciplinaria con expertos del área de conocimiento que corresponda; en este caso, con médicos.

Además del uso de la estadística descriptiva, la componente más sustanciosa de KDD consiste en usar técnicas, algoritmos y herramientas automáticas (*software*) de la rama de la inteligencia artificial denominada *aprendizaje automático (machine learning)*. Este tiene como principal objeto de estudio los procesos cognitivos del ser humano que se encargan de realizar la abstracción; con ello, se producen algoritmos y modos de representación de patrones implícitos que se descubren dentro de conjuntos de datos. La representación de estos patrones sirve, primordialmente,

para comprender mejor las interacciones entre las variables que constituyen a esos conjuntos de datos. También, sirve para producir modelos computacionales capaces de realizar la clasificación automática de casos nuevos (que no estaban dentro del *dataset* analizado) y la regresión automática (estimación) del valor desconocido de alguna variable cuantitativa a partir de datos conocidos. La clasificación y la regresión automáticas pueden aplicarse al diagnóstico y al pronóstico clínicos y a otros procesos dentro de las ciencias de la salud. A continuación, se explica cómo se producen los análisis y modelos en esta investigación con base en las técnicas mencionadas.

1.7.5 Estadística descriptiva

Los análisis estadísticos descriptivos que se realizan a la base de datos COVID-19 de la SSA consisten en conteos simples y cálculos de porcentajes y de tasas con diversos criterios. Esto se complementa con algunas gráficas, principalmente para analizar tendencias. La base de datos contiene cuarenta variables (llamadas también *atributos o descriptores*) (SSA, 2022). En términos generales, en esta investigación, la mayoría de los atributos son analizados con estadística descriptiva y, en algunos casos, modelados con técnicas de aprendizaje automático. La tabla 2 presenta las variables que resultan de interés para realizarles análisis de estadística descriptiva en el nivel de granularidad individual. Adicionalmente, la gran mayoría de estas variables son después designadas como atributo *target* para producir modelos usando un algoritmo de aprendizaje automático que genera árboles clasificadores.

Tabla 2. Selección de variables disponibles en la base de datos de (SSA, 2022) a las que realizamos análisis de estadística descriptiva con granularidad de individuo.

Núm.	VARIABLE	DESCRIPCIÓN
1	ORIGEN	El paciente fue atendido en alguna unidad de salud monitorea de enfermedades respiratorias (USMER), o bien, por alguna otra institución.
2	SECTOR	Identifica el tipo de institución del Sistema Nacional de Salud que brindó la atención.
3	ENTIDAD_UM	Identifica la entidad donde se ubica la unidad médica que brindó la atención.
4	SEXO	Identifica al sexo del paciente.
5	ENTIDAD_NAC	Identifica la entidad de nacimiento del paciente.
6	ENTIDAD_RES	Identifica la entidad de residencia del paciente.
7	MUNICIPIO_RES	Identifica el municipio de residencia del paciente.
8	TIPO_PACIENTE	Identifica el tipo de atención que recibió el paciente en la unidad. Se denomina como ambulatorio si regresó a su casa o se denomina como hospitalizado si fue ingresado a hospitalización.
9	FECHA_INGRESO	Identifica la fecha de ingreso del paciente a la unidad de atención.
10	FECHA_SÍNTOMAS	Identifica la fecha en que inició la sintomatología del paciente.
11	FECHA_DEF	Identifica la fecha en que el paciente falleció.
12	INTUBADO	Identifica si el paciente requirió de intubación.
13	NEUMONÍA	Identifica si al paciente se le diagnosticó con neumonía.
14	EDAD	Identifica la edad del paciente.
15	NACIONALIDAD	Identifica si el paciente es mexicano o extranjero.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

16	EMBARAZO	Identifica si la paciente está embarazada.
17	HABLA LENGUA IN- DIG	Identifica si el paciente habla lengua indígena.
18	INDÍGENA	Identifica si el paciente se autoidentifica como una persona indígena.
19	DIABETES	Identifica si el paciente tiene un diagnóstico de diabetes.
20	EPOC	Identifica si el paciente tiene un diagnóstico de EPOC.
21	ASMA	Identifica si el paciente tiene un diagnóstico de asma.
22	INMUSUPR	Identifica si el paciente presenta inmunosupresión.
23	HIPERTENSIÓN	Identifica si el paciente tiene un diagnóstico de hipertensión.
24	OTRA_COM	Identifica si el paciente tiene diagnóstico de otras enfermedades.
25	CARDIOVASCULAR	Identifica si el paciente tiene un diagnóstico de enfermedades cardiovasculares.
26	OBESIDAD	Identifica si el paciente tiene diagnóstico de obesidad.
27	RENAL_CRÓNICA	Identifica si el paciente tiene diagnóstico de insuficiencia renal crónica.
28	TABAQUISMO	Identifica si el paciente tiene hábito de tabaquismo.
29	OTRO_CASO	Identifica si el paciente tuvo contacto con algún otro caso diagnosticado con SARS CoV-2.
30	TOMA_MUESTRA_ LAB	Identifica si al paciente se le tomó muestra de laboratorio.
31	RESULTADO_LAB	Identifica el resultado del análisis de la muestra reportado por el laboratorio de la Red Nacional de Laboratorios de Vigilancia Epidemiológica (InDRE, LESP y LAVE) y laboratorios privados avalados por el InDRE cuyos resultados son registrados en Sisver. (Catálogo de resultados diagnósticos anexo).
32	TOMA_MUESTRA_ ANTIGENO	Identifica si al paciente se le tomó muestra de antígeno para SARS-CoV-2.
33	RESULTADO_ANTÍ- GENO	Identifica el resultado del análisis de la muestra de antígeno tomada al paciente.
34	CLASIFICACIÓN_FI- NAL	Identifica si el paciente es un caso de COVID-19 según el catálogo "CLASIFICACION_FINAL".
35	UCI	Identifica si el paciente requirió ingresar a una Unidad de Cuidados Intensivos.
36	ENT_MPIO_RES	Concatenación de ENTIDAD_RES y MUNICIPIO_RES; por ejemplo: _1_1, _32_13, _20_570.
37	SEMANA_EPIDEM	Semana epidemiológica basada en los calendarios oficiales de la SSA; se asigna en nuestro dataset con base en la FECHA_INGRESO.
38	FALLECIÓ	Dato derivado de FECHA_DEF. Si esta es igual a "9999-99-99", entonces no falleció. De otro modo, sí.
39	PEDIA_0_18	El paciente es pediátrico, con edad entre 0 y 18 años.
40	PEDIA_0	El paciente es pediátrico, con edad menor de un año.
41	PEDIA_1_4	El paciente es pediátrico, con edad entre 1 y 4 años.
42	PEDIA_5_9	El paciente es pediátrico, con edad entre 5 y 9 años.
43	PEDIA_10_14	El paciente es pediátrico, con edad entre 10 y 14 años.
44	PEDIA_15_19	El paciente es pediátrico, con edad entre 15 y 19 años.
45	EDAD_20_24	El paciente es adulto, con edad entre 20 y 24 años.
46	EDAD_25_44	El paciente es adulto, con edad entre 25 y 44 años.

CAPITULO 1. DISEÑO METODOLÓGICO

47	EDAD_45_49	El paciente es adulto, con edad entre 45 y 49 años.
48	EDAD_50_59	El paciente es adulto, con edad entre 50 y 59 años.
49	EDAD_60_64	El paciente es adulto, con edad entre 60 y 64 años.
50	EDAD_65YMAS	El paciente es adulto, con edad de 65 años o más.
51	ANIO_MES_INGRESO	Año y mes de ingreso del paciente; se asigna en nuestro dataset con base en la FECHA_INGRESO.
52	EDAD_INTERVALO	Indica el intervalo de edad del paciente con base en los grupos de edades especificados por la SSA.
53	DÍAS_SÍNTOMAS_A_INGRESO	Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso al establecimiento de salud.
54	DÍAS_SÍNTOMAS_A_DEF	Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción (valor nulo: 9999).
55	DÍAS_INGRESO_A_DEF	Cantidad de días transcurridos entre la fecha de ingreso al establecimiento de salud y la fecha de defunción (valor nulo: 9999).
56	OLA_O_PER_INTE-REPI	Número de ola o de periodo interepidémico.

Fuente: Elaboración propia con información de SSA, 2022.

Por otra parte, algunos de los atributos resultan de interés especial para el análisis de tendencias con un nivel de agregación municipal y por semanas epidemiológicas. Estas variables se presentan en la tabla 3.

Tabla 3. Variables de la base de datos COVID-19 de interés especial para analizar tendencias por municipio y semana epidemiológica.

Nombre de variable	Descripción	Formato o fuente
entidad_res	Identifica la entidad de residencia del paciente.	Catálogo: Entidades
municipio_res	Identifica el municipio de residencia del paciente.	Catálogo: Municipios
fecha_ingreso	Identifica la fecha de ingreso del paciente a la unidad de atención.	aaaa-mm-dd
fecha_síntomas	Identifica la fecha en que inició la sintomatología del paciente.	aaaa-mm-dd
fecha_def	Identifica la fecha en que el paciente falleció.	aaaa-mm-dd
clasificación_final	Identifica si el paciente es un caso de COVID-19 según el catálogo "Clasificación_final".	Catálogo: Clasificación_final

Fuente: Elaboración propia con base en el diccionario de datos de la base de datos de COVID-19 SSA, 2022.

Las variables *entidad_res* y *municipio_res* (entidad federativa de residencia y municipio de residencia, respectivamente) se usaron para producir la agregación de casos, de los diversos tipos, por municipio. Las semanas epidemiológicas se manejan tomando en cuenta la *fecha_ingreso*, que se usa para contar los diversos tipos de casos de COVID-19, con base en la variable *clasificación_final*. Al analizar las defunciones, se toma en cuenta la variable *fecha_def*. La variable *fecha_síntomas* (fecha de inicio de síntomas) se usó en algunos análisis y modelos para calcular la cantidad de días transcurridos entre el inicio de síntomas y la fecha de ingreso, o bien, entre el inicio de síntomas y la fecha de defunción. Para ciertos análisis y modelos, se usó una variable dicotómica derivada (producida por los investigadores) denominada *fa-*

llecio, con códigos sí o no, según el hecho de que una persona haya fallecido o no. El valor de esta variable se asignó automáticamente a partir del valor de la variable *fecha_def*. Si este es nulo, significa que la persona no falleció; si está registrada una fecha, entonces la persona sí falleció.

La variable *clasificación_final* puede tener uno de siete valores categóricos posibles (SSA, 2022). De estos, los que se consideran casos positivos para realizar análisis y estimar modelos en esta investigación son los tres en los que se confirma el diagnóstico de COVID-19 mediante asociación epidemiológica, por dictamen, o por examen de laboratorio positivo (ver tabla 4). Los cuatro valores restantes que se manejan en la variable *clasificación_final* son: 4) inválido por laboratorio, 5) no realizado por laboratorio, 6) caso sospechoso y 7) negativo a SARS-CoV-2. En términos generales, se supone que, además de la clase 6, las clasificaciones 4 y 5 también serían sospechosos; aunque estas dos podrían, eventualmente, tener una confirmación por asociación clínica, o bien, inválido por laboratorio e igualmente tener esta confirmación por asociación.

Tabla 4. Valores de la variable *clasificación_final* que se toman como casos positivos en esta investigación.

Clave	Clasificación	Descripción
1	Caso de COVID-19 confirmado por asociación clínica epidemiológica	Aplica cuando el caso informó ser contacto de un positivo a COVID-19 (y este se encuentra registrado en el Sis-ver) y al caso no se le tomó muestra o la muestra resultó no válida.
2	Caso de COVID-19 confirmado por comité de dictaminación	Solo aplica para defunciones bajo las siguientes condiciones: al caso no se le tomó muestra o sí se tomó muestra, pero esta resultó no válida.
3	Caso de SARS-CoV-2 confirmado	Aplica cuando: el caso tiene muestra de laboratorio o prueba antigénica y resultó positiva a SARS-CoV-2, sin importar si el caso tiene asociación clínica epidemiológica.

Fuente: Elaboración propia con información del Catálogo de la base de datos COVID-19 SSA, 2022.

1.7.6 Aprendizaje automático

Como se mencionó anteriormente, el aprendizaje automático es una rama de la inteligencia artificial que tiene el propósito de modelar e imitar la capacidad humana de descubrir y reconocer patrones en conjuntos de información (numérica, textual, visual, auditiva, etcétera). En general, el aprendizaje automático se implementa computacionalmente mediante algoritmos y representaciones de información que se materializan concretamente como *software*, estructuras de datos y archivos digitales. Los algoritmos de aprendizaje automático se agrupan en dos grandes categorías: aprendizaje automático supervisado y aprendizaje automático no supervisado. La diferencia entre ambos es que el aprendizaje supervisado descubre automáticamente asociaciones o correlaciones entre atributos de un conjunto de datos, donde un atributo se designa como variable dependiente y otros atributos se designan como variables independientes. La variable dependiente representa la clase o tipo a la cual pertenece cada uno de los casos del conjunto analizado. En contraste, los algoritmos de aprendizaje no

supervisado descubren automáticamente subconjuntos de casos que pudieran tener alguna similitud entre ellos dentro de un conjunto de información en el que no se tienen identificadas previamente las clases o tipos a los cuales pertenecen los casos. Para una mayor comprensión de los fundamentos del aprendizaje automático, se recomiendan fuentes como: (Russell and Norvig, 2020), (Han *et al.*, 2011) y (Witten *et al.*, 2016), entre otros.

El descubrimiento de patrones de correlación entre diversos atributos de la base de datos de casos registrados de COVID-19 de la SSA consiste en detectar y representar correlaciones estadísticas entre alguna variable determinada que se designa como atributo *target* (variable de interés o variable dependiente) y un conjunto de otras variables que se asumen como variables predictoras o independientes. El atributo designado como *target* (variable dependiente) corresponde a lo que en la disciplina estadística de Diseño de Experimentos generalmente se denomina variable y (minúscula, por ser un solo atributo). Por su parte, los atributos seleccionados como predictores generalmente se denotan por la variable X (mayúscula, por ser un conjunto de dos o más atributos). Los algoritmos de aprendizaje automático supervisado son de gran utilidad para descubrir, representar y explorar patrones de correlación multivariable.

En esta investigación, algunos de los atributos que resultan de mayor interés para designarlos individualmente como *target* al buscar correlaciones mediante algoritmos de aprendizaje automático supervisado son, entre otros, si el paciente falleció o no, si fue intubado o no, si fue hospitalizado o no. El propósito de los modelos así producidos es descubrir y representar computacionalmente en qué se parecen los pacientes que fallecieron y en qué se parecen los que no fallecieron; o bien, en qué se parecen los que fueron intubados y en qué se parecen los que no lo fueron; en qué se parecen los hospitalizados y en qué se parecen los ambulatorios, etcétera.

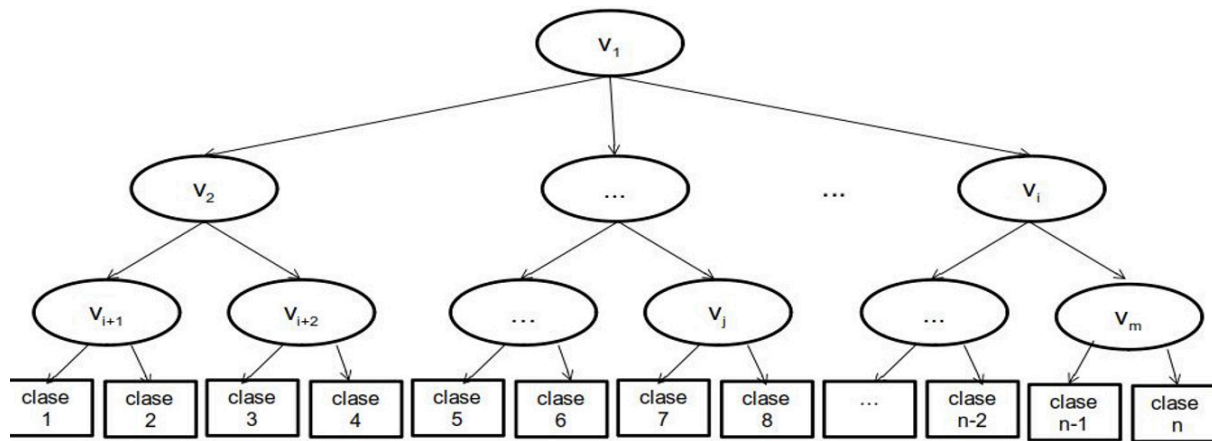
Las correlaciones mencionadas se buscan en dos modalidades: primeramente, solo entre los atributos que están dentro de la base de datos de casos de COVID-19. Posteriormente, esta base de datos es combinada con otras bases gubernamentales que resultan de interés científico; por ejemplo, la información sociodemográfica disponible en los censos de población y vivienda. Al combinar las bases de datos se producen nuevos *datasets*, en los que se buscan otras correlaciones interesantes. Las bases que se combinan con la de COVID-19 son: la del Censo Nacional de Población y Vivienda 2020 y la de Proyecciones de Población Municipal 2015-2030. Con base en ello, se descubrieron correlaciones, por ejemplo, entre las cantidades de pacientes fallecidos, intubados, hospitalizados o casos positivos ambulatorios en relación con las características sociodemográficas de los habitantes y las viviendas de los municipios. De los diversos algoritmos disponibles en el aprendizaje automático supervisado para descubrir y representar patrones, en esta investigación se usaron los denominados árboles clasificadores.

Árbol clasificador

En la disciplina del aprendizaje automático, un árbol clasificador es un diagrama de nodos y líneas que representa los patrones de correlaciones entre una variable dependiente, de tipo nominal (denominada atributo *target* o atributo de clase) y un conjunto

de variables cuantitativas o nominales que se denominan *atributos predictores* (ver Figura 1). Pueden encontrarse sólidas explicaciones de los fundamentos de los árboles clasificadores en (Russell y Norvig, 2020:697-707) y (Han et al., 2012:330-350).

Figura 1. Ejemplo de árbol clasificador producido con un algoritmo de aprendizaje automático supervisado.



Fuente: Elaboración propia.

Esencialmente, un árbol clasificador es un conjunto de reglas que describen las combinaciones de atributos y de sus valores que están asociadas a un determinado valor del atributo de clase. El árbol es un conjunto de inferencias lógicas, de la forma *Si P, entonces Q*, donde cada una de estas describe las correlaciones estadísticas entre un conjunto de atributos predictores y un atributo *target*. Es decir, el árbol muestra en qué se parecen los casos que corresponden a un determinado valor de clase del atributo *target*. Este tipo de árbol es producido **automáticamente** por un algoritmo implementado en un *software* que analiza un conjunto de datos (*dataset*). Es importante remarcar que una de las grandes ventajas de este tipo de modelo es la posibilidad de analizar con relativa facilidad conjuntos de datos que pueden contener cientos o miles de atributos (columnas) y miles o millones de casos (renglones). El analista prepara el *dataset* que desea analizar, lo alimenta a un *software* que tenga implementado algún algoritmo generador de árbol clasificador, designa al atributo que desea manejar como *target* y a los atributos que desea manejar como predictores y produce el árbol automáticamente.

En la figura 1, cada nodo elíptico del árbol representa a un atributo existente en el *dataset*. Los atributos mostrados en el diagrama son seleccionados y colocados en un orden determinado por el algoritmo generador del árbol. Algunos atributos existentes en el *dataset* podrían no ser seleccionados por el algoritmo para ser incluidos en el árbol porque el poder discriminante de esos atributos no contribuye suficientemente para distinguir entre casos correspondientes a clases distintas.

La altura que tiene un nodo dentro del diagrama de árbol representa su poder discriminante o poder clasificador. El nodo que se ubica en la parte más alta del diagrama se denomina *raíz* y constituye el atributo con mayor poder clasificador. De ahí hacia abajo, los nodos tienen cada vez menor poder clasificador. Si un atributo del *dataset* no está en el árbol, significa que el algoritmo considera que tiene muy poco poder clasificador.

Los nodos rectangulares ubicados en el extremo inferior del árbol, llamados *hojas*, representan los valores nominales de las clases a las cuales pertenecen las instancias del *dataset*. Cada línea (arista) que une a un par de nodos en el diagrama representa dos cosas: 1) una comparación del atributo ubicado en el nodo del extremo superior de la arista con el valor de ese atributo correspondiente a una instancia del *dataset* que se esté procesando, y 2) una operación lógica *AND* con el atributo ubicado en el extremo inferior de la arista. Las aristas que terminan en los nodos hoja no representan una operación *AND*, sino el final de una implicación lógica (*Si P, entonces Q*). Aunque en la Figura 1 las hojas muestran nombres de clase que son todos distintos entre sí, en una situación determinada puede ocurrir que el nombre de una clase aparezca en varios nodos y no solamente en uno; esto significa que una misma clase de instancias puede estar descrita por diferentes patrones que incluyen a distintas variables con distintos valores.

Uno de los algoritmos de aprendizaje automático supervisado más reconocidos para producir árboles clasificadores es el llamado C4.5 (Quinlan, 1993), que después fue mejorado por su autor para producir el algoritmo C5.0. Estos algoritmos se encuentran implementados en diversas herramientas de *software* libre que facilitan su uso y aprovechamiento. Al menos, existen tres implementaciones que son ampliamente usadas internacionalmente: una variante de C4.5, llamada J4.8 (Frank and Witten, 1998), está disponible en el *software* libre llamado *Waikato Environment for Knowledge Analysis* (WEKA)¹¹. Esta herramienta ofrece una interfaz gráfica altamente amigable para usuarios con conocimientos limitados en lenguajes de programación.

Por su parte, el algoritmo C5.0 tiene implementaciones para usuarios que tienen conocimientos en lenguajes de programación Python y R. En Python, existe un conjunto de bibliotecas de tipo *open source* (de código abierto), principalmente la biblioteca llamada Scikit-learn, con sus módulos de árboles clasificadores. Sin embargo, esto está sujeto a algunas limitaciones, ya que, hasta el momento de escribir este libro, esa implementación solamente tiene la capacidad para procesar atributos predictores de tipo cuantitativo, no de tipo nominal como lo son algunos atributos de la base de datos COVID-19 de SSA.

Por lo anterior, para producir los árboles clasificadores en nuestra investigación, se eligió el algoritmo C5.0 en su implementación de lenguaje R, específicamente la biblioteca llamada C50 (Kuhn and Quinlan, 2022). Los lectores interesados en profundizar en el *software* de aprendizaje automático podrían explorar posibilidades en el ámbito del *software* comercial, como el de las marcas SPSS y SAS, entre otras.

1.7.7 Patrones geográficos y mapas digitales

Los mapas digitales, producidos a partir de datos primarios o de cálculos, análisis y modelos producidos a partir de estos, facilitan la observación de patrones geográficos y cronológicos de diversos fenómenos. Por ello, en esta investigación se producen mapas digitales para observar el comportamiento de la pandemia de COVID-19 en el conjunto de los municipios mexicanos. A partir de diversos análisis estadísticos y modelos de aprendizaje automático (principalmente, estadística descriptiva, análi-

11 Disponible en <https://www.cs.waikato.ac.nz/ml/weka/>

sis de tendencias y árboles clasificadores), se producen mapas digitales usando los *softwares* libres QGIS (<https://www.qgis.org/>) y algunas bibliotecas del lenguaje de programación Python, entre otros.

A partir de un conjunto de mapas digitales correspondientes a diversos momentos ubicados dentro de un periodo de tiempo, puede producirse una animación que muestre la evolución de algún fenómeno de interés. Este tipo de animación se denomina *GIF animado*. En esta investigación, se producen algunas animaciones de mapas que muestran el comportamiento territorial y temporal de la pandemia de COVID-19 en los municipios del territorio nacional.

1.8 Sobre árboles clasificadores e inferencia estadística

Por sí mismo, el algoritmo C5.0 para generar árboles clasificadores no se considera un método formal de inferencia estadística; principalmente, porque no proporciona herramientas para realizar pruebas de hipótesis sobre los parámetros del modelo o la población objetivo y porque tampoco permite calcular intervalos de confianza para los parámetros del modelo o las predicciones realizadas. Sin embargo, a pesar de ello, es una herramienta muy útil para la clasificación de datos y la construcción de modelos descriptivos y predictivos.

Aunque en esta investigación no se plantea ninguna hipótesis y, por tanto, no se realiza su prueba, se usan algunos métodos formales de inferencia para discutir los hallazgos de los árboles clasificadores. Los métodos son:

- i. Validación usando un conjunto de prueba independiente: después de entrenar el modelo, se evalúa su desempeño sobre un conjunto de datos independiente que no fue utilizado durante el entrenamiento. Esto proporciona una medida objetiva del rendimiento del modelo. Estas medidas son dos: precisión global (porcentaje de casos clasificados correctamente) y Kappa, de Cohen (1960).
- ii. Medición de desempeño: sensibilidad (recall), especificidad, medida F1 (es la media armónica de la precisión y la sensibilidad).
- iii. Análisis de importancia de atributos: se hace con base en la frecuencia y el nivel en el que los atributos predictores aparecen en los árboles. De este modo, se detecta cuáles son los atributos con mayor poder predictivo (o discriminante) en relación con alguna variable objetivo.
- iv. Interpretación de los árboles clasificadores: se interpretan las reglas clasificadoras que constituyen a los árboles. Esto puede revelar relaciones importantes y patrones en los datos que son críticos para entender la dinámica de la pandemia.
- v. Análisis de errores de clasificación: se analizan los casos en los que un modelo de árbol falla (falsos positivos y falsos negativos), lo cual puede proporcionar información valiosa sobre subgrupos específicos o condiciones particulares en los datos. Esto incluye que, para cada regla del árbol, se calculan sus valores de support y confidence.

Con base en lo anterior, los capítulos 4 (sobre los patrones hallados en casos individuales) y 7 (a nivel municipal), los modelos de árboles clasificadores más desta-

CAPITULO 1. DISEÑO METODOLÓGICO

cables (aquellos con mayores valores de precisión global y *Kappa* de Cohen) son comentados y discutidos tomando en cuenta estas dos medidas. Además, los modelos completos que hemos puesto en la web para ser descargados gratuitamente incluyen su correspondiente matriz de confusión, junto con medidas estadísticas más finas de cada regla clasificadora, tales como la sensibilidad (*recall*), la especificidad y F1.

CAPÍTULO 2

PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

El propósito de este capítulo es presentar los resultados de los análisis de estadística descriptiva de las variables disponibles en nuestro dataset denominado *Master-3*. Este proviene de la base de datos de casos de COVID-19 de la Secretaría de Salud, del Gobierno Federal de México (SSA, 2022). *Master-3* contiene aproximadamente 18.4 millones de registros, incluyendo casos positivos para nuestro análisis, casos negativos y de otras clasificaciones, teniendo disponibles 56 variables. Puede descargarse de la nube¹² y puede usarse, citando este libro. Este panorama estadístico se denomina general porque abarca todas las clasificaciones de casos incluidos en el *dataset*. El principal tipo de análisis realizado es univariable (una sola variable en cada análisis), con la técnica de Pareto (80:20). Es decir, se realizan los conteos de cada categoría o subconjunto casos de la variable, se ordena la tabla con base en sus conteos en forma descendente, se calcula el porcentaje correspondiente a cada conteo, se van sumando los porcentajes y, finalmente, se identifican los renglones de la tabla que agrupan al 80% de los casos. Por ello, en algunas tablas se marca con color gris a los renglones que acumulan, aproximadamente, 80% de los casos de la variable.

De algunas variables del *dataset* no se realiza análisis porque sus datos forman parte de alguna otra variable y su análisis sería innecesario y redundante. Por ejemplo, el análisis de la variable *EDAD_INTERVALO* aborda a todos los intervalos de edades incluidos en el dataset, por lo que solamente se presentan análisis para algunas pocas de las variables correspondientes a los diversos intervalos de edades (p. ej. *PE-DIA_0_18* y *PEDIA_0*). De las variables que son fechas (p. ej. *FECHA_SÍNTOMAS*), se realizan conteos simples presentados en gráficas de columnas con el propósito de visualizar las tendencias.

12 Disponible en <https://1drv.ms/f/s!At6TY8A0mTpNgtZJpVP9VfhgKGGsfA?e=AZ56eP>

En otros capítulos, se presentan análisis estadísticos del dataset, pero enfocándose solamente en los casos que son considerados positivos por la variable *CLASIFICACIÓN_FINAL*. Otros análisis se enfocan solamente en pacientes mujeres y/o en ciertos segmentos de edades. En términos generales, los análisis de estadística descriptiva sirven para descubrir los patrones más generales y fáciles de detectar en el dataset y también tienen la valiosa utilidad de servir para evaluar los modelos de árboles clasificadores (o reglas clasificadoras) que se producen usando algoritmos de aprendizaje automático.

A continuación, se presenta una amplia serie de varias decenas de tablas de conteos y porcentajes (análisis de Pareto). Es recomendable poner especial atención en la columna de *Porcentaje acumulado*, ya que permite percibir el volumen estadísticamente más sustancioso de los casos de la variable correspondiente. También es recomendable comparar estos resultados con los presentados en el capítulo 3, con relación a los casos positivos de COVID-19 (con base en la variable *CLASIFICACIÓN_FINAL*). Todos los análisis de estadística descriptiva presentados son útiles también para la posterior evaluación de los modelos de aprendizaje automático que se producen.

Aunque la cantidad de estas tablas es relativamente grande, es pertinente observarlas porque, de este modo, se puede tener una visión general del comportamiento de la pandemia en México en el periodo analizado. Para facilitar la asimilación de tanta información, las tablas están agrupadas en cuatro secciones: 1) características del paciente, 2) historia clínica y comorbilidades, 3) curso clínico y resultados, y 4) contexto epidemiológico.

2.1 Características del paciente

Las características del paciente que se abordan en esta sección son quince: 1) falleció (o no), 2) intubado (o no), 3) tipo de paciente (ambulatorio u hospitalizado), 4) clasificación final del caso (COVID-19 confirmado, negativo, etc.), 5) entidad federativa de nacimiento, 6) entidad federativa de residencia, 7) nacionalidad, 8) indígena (o no), 9) habla alguna lengua indígena (o no), 10) sexo, 11) embarazo (o no), 12) paciente pediátrico (entre 0 y 18 años de edad), 13) paciente pediátrico de 0 (cero) años de edad, 14) paciente de 65 años o más, y 15) intervalo de edad.

La tabla 5 presenta el análisis de fallecidos y no fallecidos (variable núm. 38: *FALLECIÓ*). Los fallecidos constituyen el 2.3% del total. Es importante recordar que estos fallecidos pudieron ser casos positivos, o no, con base en la variable *CLASIFICACIÓN_FINAL*; es decir, pudieron haber fallecido por otras causas distintas a COVID-19.

Tabla 5. Falleció.

	Conteo	%	% acumulado
No	17,899,449	97.7	97.7
Sí	428,926	2.3	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La tabla 6 presenta el análisis de la variable INTUBADO (núm. 12), indicando si una persona fue, o no, intubada. En la gran mayoría de los casos (92.9%), la intubación no era aplicable; en el 6.4% no se aplicó; en el 0.6% sí se aplicó, y en el 0.1% la información no está especificada.

Tabla 6. Intubado.

	Conteo	%	% acumulado
No aplica	17,024,699	92.9	92.9
No	1,171,223	6.4	99.3
Sí	118,143	0.6	99.9
No especificado	14,310	0.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 7 analiza la variable TIPO_PACIENTE, donde 92.9% de los casos recibieron atención de tipo ambulatorio (identificador _1) y 7.1% fueron hospitalizados (identificador _2).

Tabla 7. Tipo de paciente.

	Conteo	%	% acumulado
Ambulatorio	17,024,699	92.9	92.9
Hospitalizado	1,303,676	7.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 8 presenta la variable CLASIFICACIÓN_FINAL (núm. 34) mostrando que más de la mitad (57.3%) de los casos fueron negativos a SARS-CoV-2, mientras que más de la tercera parte (36.5%) fueron positivos (confirmados).

Tabla 8. Clasificación final del caso.

	Conteo	%	% acumulado
Negativo a SARS-CoV-2	10,493,385	57.3	57.3
Caso de SARS-CoV-2 confirmado	6,694,337	36.5	93.8
Caso sospechoso	660,132	3.6	97.4
Caso de COVID-19 confirmado por asociación clínica epidemiológica	359,294	2.0	99.3
No realizado por laboratorio	93,552	0.5	99.8
Caso de COVID-19 confirmado por comité de dictaminación	15,912	0.1	99.9
Inválido por laboratorio	11,763	0.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 9 muestra los resultados de la entidad federativa de nacimiento (ENTIDAD_NAC) de las personas registradas. La mayoría (80.8%) se concentra solamente en 16 entidades. Las cinco entidades con mayores porcentajes son: Ciudad de México (30.7%), Estado de México (10.5%), Guanajuato (4.4%), Nuevo León (3.9%) y Vera-

CAPITULO 2. PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

cruz (3.4%). Por otra parte, las cinco entidades con menores porcentajes son: Nayarit (0.8%), Tlaxcala (0.8%), Quintana Roo (0.6%), Colima (0.6%) y (0.6%). Sin embargo, existe un 0.7% de casos en que no se ha especificado (identificador _99) la entidad de nacimiento de la persona. Este porcentaje es superior al de cada uno de los estados de Quintana Roo, Colima y

Tabla 9. Entidad federativa de nacimiento

	Conteo	%	% acumulado
Ciudad de México	5,623,492	30.7	30.7
Estado de México	1,919,290	10.5	41.2
Guanajuato	803,922	4.4	45.5
Nuevo León	708,216	3.9	49.4
Veracruz	615,945	3.4	52.8
San Luis Potosí	607,662	3.3	56.1
Jalisco	606,683	3.3	59.4
Tabasco	605,088	3.3	62.7
Puebla	520,798	2.8	65.5
Sinaloa	469,029	2.6	68.1
Coahuila	454,576	2.5	70.6
Michoacán	415,683	2.3	72.8
Tamaulipas	381,970	2.1	74.9
Sonora	373,885	2.0	77.0
Guerrero	363,803	2.0	78.9
Oaxaca	347,660	1.9	80.8
Yucatán	320,100	1.7	82.6
Morelos	316,880	1.7	84.3
Chiapas	294,915	1.6	85.9
Chihuahua	273,970	1.5	87.4
Hidalgo	247,996	1.4	88.8
Querétaro	240,355	1.3	90.1
Baja California	234,338	1.3	91.4
Aguascalientes	216,076	1.2	92.5
Durango	212,477	1.2	93.7
Baja California Sur	203,597	1.1	94.8
Zacatecas	186,794	1.0	95.8
Nayarit	147,193	0.8	96.6
Tlaxcala	145,450	0.8	97.4
No especificado	131,229	0.7	98.1
Quintana Roo	117,554	0.6	98.8
Colima	111,233	0.6	99.4
Campeche	110,516	0.6	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 10 presenta el análisis de la entidad federativa de residencia (ENTIDAD_RES) del individuo. La mayoría (80.5%) reside en dieciséis estados. Las cinco entidades

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

con más casos son: Ciudad de México (31.7%), Estado de México (10.3%), Nuevo León (4.5%), Guanajuato (4.4%) y San Luis Potosí (3.4%). En contraste, las que tienen menos casos son: Zacatecas (0.9%), Nayarit (0.8%), Tlaxcala (0.8%) Colima (0.7%) y Campeche (0.6%).

Tabla 10. Entidad federativa de residencia.

	Conteo	%	% acumulado
Ciudad de México	5,810,029	31.7	31.7
Estado de México	1,896,132	10.3	42.0
Nuevo León	824,317	4.5	46.5
Guanajuato	801,352	4.4	50.9
San Luis Potosí	628,350	3.4	54.3
Tabasco	595,944	3.3	57.6
Jalisco	595,047	3.2	60.8
Puebla	451,025	2.5	63.3
Coahuila	448,497	2.4	65.7
Veracruz	420,783	2.3	68.0
Sinaloa	407,956	2.2	70.3
Baja California	396,595	2.2	72.4
Tamaulipas	390,783	2.1	74.6
Sonora	385,571	2.1	76.7
Morelos	362,084	2.0	78.6
Querétaro	344,844	1.9	80.5
Michoacán	341,402	1.9	82.4
Baja California Sur	312,127	1.7	84.1
Yucatán	305,914	1.7	85.8
Chihuahua	288,982	1.6	87.3
Oaxaca	259,769	1.4	88.8
Guerrero	256,459	1.4	90.2
Chiapas	235,146	1.3	91.4
Hidalgo	231,009	1.3	92.7
Aguascalientes	227,358	1.2	93.9
Quintana Roo	219,710	1.2	95.1
Durango	191,091	1.0	96.2
Zacatecas	169,984	0.9	97.1
Nayarit	144,210	0.8	97.9
Tlaxcala	141,768	0.8	98.7
Colima	131,489	0.7	99.4
Campeche	112,648	0.6	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 11 muestra el análisis de la variable núm. 15 (NACIONALIDAD), donde se observa que casi la totalidad de los casos (99.3%) son mexicanos y menos del 1% son extranjeros.

Tabla 11. Nacionalidad.

	Conteo	%	% acumulado
Mexicana	18,197,434	99.3	99.3
Extranjera	130,941	0.7	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 12 analiza si la persona se autoadscribía como indígena (variable núm. 18: INDÍGENA). Solo el 0.8% del total se autoadscribió como tal.

Tabla 12. Indígena

	Conteo	%	% acumulado
No	16,958,522	92.5	92.5
No especificado	1,227,309	6.7	99.2
Sí	142,544	0.8	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 13 analiza si la persona habla, o no, alguna lengua indígena mexicana (variable núm. 17: *HABLA LENGUA INDIG*). Solamente el 0.5% de las personas registradas hablan alguna de esas lenguas.

Tabla 13. Habla alguna lengua indígena.

	Conteo	%	% acumulado
No	16,981,689	92.7	92.7
No especificado	1,246,777	6.8	99.5
Sí	99,909	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 14 presenta el análisis de la variable Sexo, mostrando que, de las personas registradas, poco más de la mitad (53.9%) son mujeres (identificador _1).

Tabla 14. Sexo.

	Conteo	%	% acumulado
Mujer	9,873,929	53.9	53.9
Hombre	8,454,446	46.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 15 muestra los resultados de la variable EMBARAZO (núm. 16), indicando que solamente el 1.1% de los casos corresponden a embarazadas.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 15. Embarazo.

	Conteo	%	% acumulado
No	9,602,848	52.4	52.4
No aplica	8,454,446	46.1	98.5
Sí	192,607	1.1	99.6
Se ignora	78,456	0.4	100.0
No especificado	18	0.0	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

El análisis de la variable Edad (núm. 14) es abordado en el correspondiente a la variable *EDAD_INTERVALO* (núm. 52), en la tabla 19. Los análisis de otras variables de edad existentes en nuestro *dataset Master-3* también están incluidos en el de *EDAD_INTERVALO*. Esas variables son: *PEDIA_1_4*, *PEDIA_5_9*, *PEDIA_10_14*, *PEDIA_15_19*, *EDAD_20_24*, *EDAD_25_44*, *EDAD_45_49*, *EDAD_50_59* y *EDAD_60_64*.

La tabla 16 muestra el análisis de los casos que se encuentran, o no, en edad pediátrica (entre 0 y 18 años), correspondientes a la variable *PEDIA_0_18* (núm. 39). Los casos pediátricos constituyen el 9.6%.

Tabla 16. Paciente pediátrico (entre 0 y 18 años de edad).

	Conteo	%	% acumulado
<i>PEDIA_0_18</i>			
No	16,575,730	90.4	90.4
Sí	1,752,645	9.6	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 17 muestra el análisis de la variable *PEDIA_0* (núm. 40), refiriéndose a los individuos en edad de cero años, que constituyen el 0.3% del total.

Tabla 17. Paciente pediátrico de 0 (cero) años de edad.

	Conteo	%	% acumulado
<i>EDAD_65YMAS</i>			
No	16,923,987	92.3	92.3
Sí	1,404,388	7.7	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 18 presenta el análisis de la variable *EDAD_65YMAS* (núm. 50), que refiere a los individuos que tienen 65 o más años de edad. Estos constituyen el 7.7% del total.

Tabla 18. Paciente de 65 años o más.

	Conteo	%	% acumulado
<i>EDAD_65YMAS</i>			
No	16,923,987	92.3	92.3
Sí	1,404,388	7.7	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 19 muestra el análisis de las edades de las personas (variable núm. 52: *EDAD_INTERVALO*). La mayoría de los casos (85.0%) corresponde a edades entre 20 y 59 años y 65 y más años. Curiosamente, el intervalo de 60 a 64 años (con 4.1%) no se ubica dentro del 85.0% mencionado. Casi la mitad (44.6%) corresponde al intervalo de 25 a 44 años.

Tabla 19. Intervalo de edad.

	Conteo	%	% acumulado
_25_44	8,176,272	44.6	44.6
_50_59	2,476,517	13.5	58.1
_20_24	1,849,808	10.1	68.2
_45_49	1,665,506	9.1	77.3
_65ymas	1,404,388	7.7	85.0
_15_19	829,399	4.5	89.5
_60_64	756,452	4.1	93.6
_10_14	491,490	2.7	96.3
_5_9	342,619	1.9	98.2
_1_4	271,869	1.5	99.7
_0	64,055	0.3	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

2.2 Historia clínica y comorbilidades

En relación con la historia clínica y las comorbilidades del individuo, a continuación, se analizan once atributos: 1) diabetes, 2) inmunosupresión, 3) obesidad, 4) insuficiencia renal crónica, 5) neumonía, 6) EPOC, 7) asma, 8) hipertensión, 9) enfermedad cardiovascular, 10) tabaquismo y 11) alguna otra comorbilidad.

La tabla 20 analiza si la persona padece, o no, DIABETES (variable núm. 19). El 7.5% de los casos era diabética (valor _1).

Tabla 20. Diabetes.

	Conteo	%	% acumulado
No	16,859,159	92.0	92.0
Sí	1,375,564	7.5	99.5
Se ignora	93,652	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 21 muestra el análisis de la variable INMUSUPR (núm. 22), donde el 0.6% de los casos presenta inmunosupresión.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 21. Inmunosupresión.

	Conteo	%	% acumulado
No	18,129,701	98.9	98.9
Sí	107,358	0.6	99.5
Se ignora	91,316	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 22 presenta los resultados de la variable OBESIDAD (núm. 26). Se observa que esta condición estaba presente en el 8.3% de los casos.

Tabla 22. Obesidad.

	Conteo	%	% acumulado
No	16,709,603	91.2	91.2
Sí	1,529,503	8.3	99.5
Se ignora	89,269	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

En la tabla 23 se muestra el análisis de la variable RENAL_CRÓNICA (núm. 27), refiriendo que el 0.9% de los casos presenta esta condición.

Tabla 23. Insuficiencia renal crónica.

	Conteo	%	% acumulado
No	18,066,958	98.6	98.6
Sí	170,635	0.9	99.5
Se ignora	90,782	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 24 presenta los resultados de la variable núm. 13 (NEUMONÍA). La presencia de neumonía (identificador con valor _2) solamente se observa en el 4.3% de los casos.

Tabla 24. Neumonía.

	Conteo	%	% acumulado
No	17,345,895	94.6	94.6
Sí	788,303	4.3	98.9
No especificado	194,177	1.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 25 muestra el análisis de presencia de Enfermedad Pulmonar Obstructiva Crónica (EPOC, variable núm. 20). El 0.7% de los casos refirió EPOC y en 0.5% se ignoraba si la tenía.

Tabla 25. EPOC.

	Conteo	%	% acumulado
No	18,116,108	98.8	98.8
Sí	121,070	0.7	99.5
Se ignora	91,197	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 26 muestra el análisis de la variable ASMA (núm. 21), indicando que el 2.0% de los casos presentaban antecedentes de asma y en el 0.5% se ignoraba si la tenía o no.

Tabla 26. Asma.

	Conteo	%	% acumulado
No	17,871,819	97.5	97.5
Sí	365,839	2.0	99.5
Se ignora	90,717	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 27 muestra el análisis de la variable HIPERTENSIÓN (núm. 23), donde se observa que el 10.3% de los casos presentaba esta condición.

Tabla 27. Hipertensión.

	Conteo	%	% acumulado
No	16,351,698	89.2	89.2
Sí	1,885,375	10.3	99.5
Se ignora	91,302	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 28 muestra los resultados de la variable CARDIOVASCULAR (núm. 25), donde se observa que 1.0% de los casos presentaba esta condición.

Tabla 28. Enfermedad cardiovascular.

	Conteo	%	% acumulado
No	18,058,248	98.5	98.5
Sí	179,019	1.0	99.5
Se ignora	91,108	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 29 presenta el análisis de TABAQUISMO (variable núm. 28), donde el 6.0% presentaba esta situación.

Tabla 29. Tabaquismo.

	Conteo	%	% acumulado
No	17,129,702	93.5	93.5
Sí	1,106,462	6.0	99.5
Se ignora	92,211	0.5	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 30 presenta los resultados de la variable denominada OTRA_COM (núm. 24), refiriéndose a alguna comorbilidad diferente de las referidas con anterioridad, agrupándolas en esta categoría.

Tabla 30. Alguna otra comorbilidad.

	Conteo	%	% acumulado
No	17,792,009	97.1	97.1
Se ignora	294,131	1.6	98.7
Sí	242,235	1.3	100.0
Total	18,328,375		

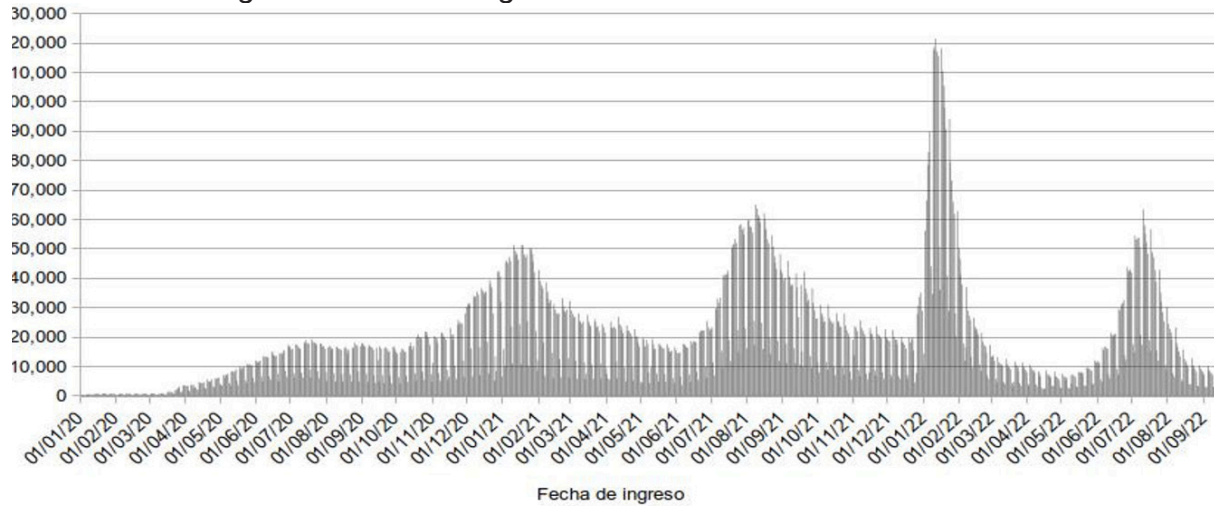
Fuente: Elaboración propia con datos de SSA, 2022.

2.3 Curso clínico y resultados

Respecto al curso clínico y los resultados de los individuos, se analizan once atributos: 1) fecha de ingreso al establecimiento de salud, 2) fecha de inicio de síntomas, 3) fecha de defunción (en su caso), 4) el individuo estuvo, o no, en la unidad de cuidados intensivos (UCI), 5) se tomó, o no, muestra de laboratorio, 6) resultado de laboratorio, 7) se tomó, o no, muestra de antígeno, 8) resultado de prueba de antígeno, 9) cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso al establecimiento de salud, 10) cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción, y 11) cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción.

La figura 2 presenta el análisis de las fechas de ingreso de las personas a la unidad de atención (variable núm. 9: *FECHA_INGRESO*). Los *picos* en la gráfica representan las cinco olas de la pandemia en México. Con base en (SSA, 2022b), las olas ocurren en los períodos que se enumeran a continuación: primera, del 16 de febrero al 26 de septiembre de 2020; segunda, del 27 de septiembre de 2020 al 17 de abril de 2021; tercera, del 06 de junio al 23 de octubre de 2021; cuarta, del 19 de diciembre de 2021 al 05 de marzo de 2022, y quinta, del 29 de mayo al 20 de agosto de 2022. Los cuatro periodos que se ubican intercalados con las cinco olas se denominan periodos interepidémicos.

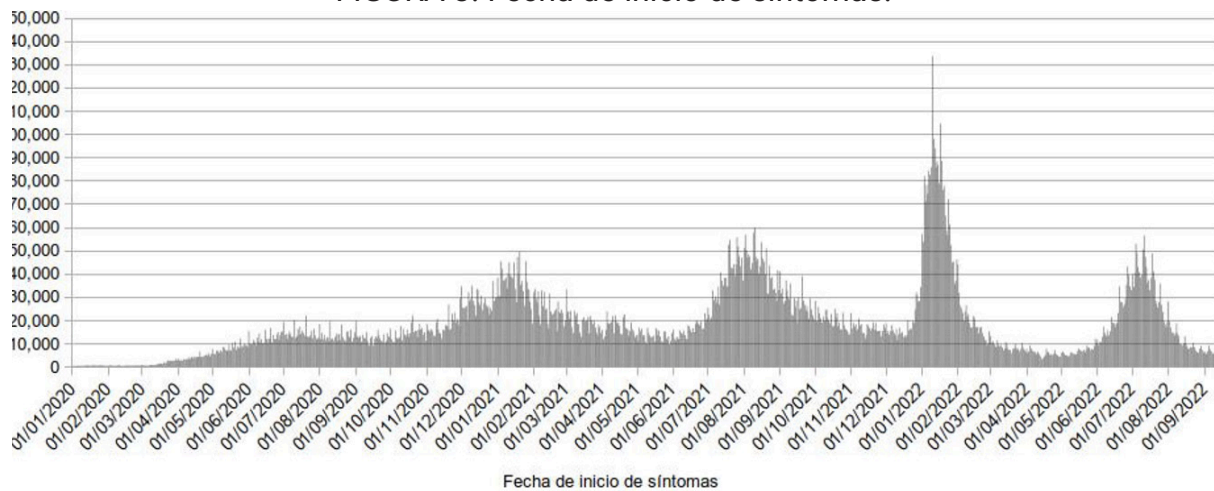
Figura 2.Fecha de ingreso al establecimiento de la salud.



Fuente: Elaboración propia con datos de SSA, 2022.

La figura 3 presenta el análisis de la cantidad de casos por fecha de inicio de síntomas (variable núm. 10: FECHA_SÍNTOMAS). Al igual que en la figura 2, los picos representan las cinco olas de la pandemia. Acá también, la cuarta ola fue la más alta.

FIGURA 3. Fecha de inicio de síntomas.

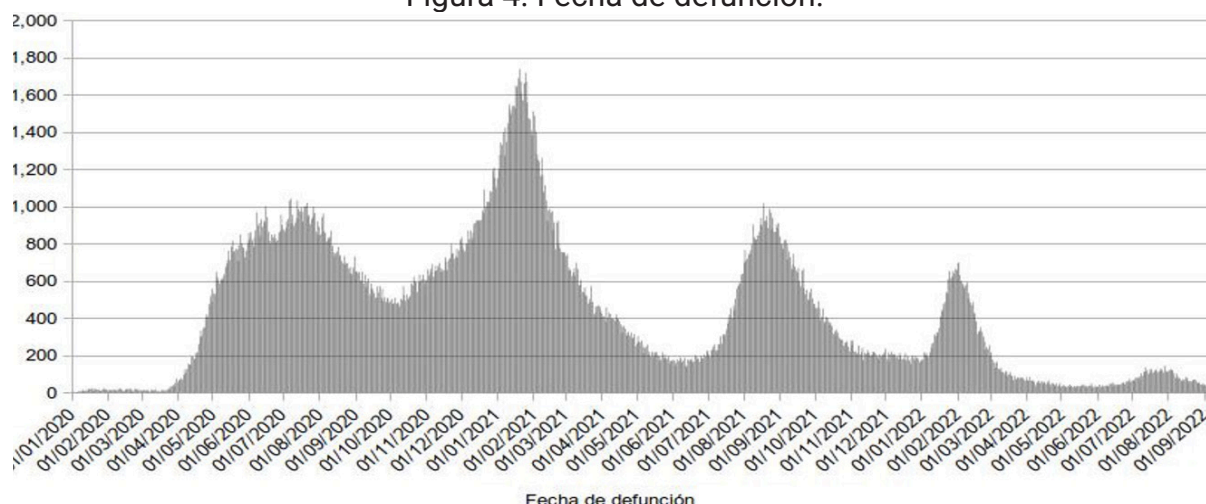


Fuente: Elaboración propia con datos de SSA, 2022.

La figura 4 presenta el análisis de la cantidad de defunciones por fecha (variable núm. 11: FECHA_DEF). Al igual que en las figuras 2 y 3, se perciben las cinco olas. Sin embargo, los picos de las olas se ubican en periodos distintos a los observados en las otras dos figuras. En las defunciones, el pico más alto ocurre en la segunda ola. En orden de mayor a menor altura, las olas son: segunda, primera, tercera, cuarta y quinta. Es evidente la disminución en la cantidad de defunciones después de la segunda ola.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Figura 4. Fecha de defunción.



Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 31 presenta el análisis de la variable UCI (unidad de cuidados intensivos) (núm. 35), mostrando que el 0.5% de los casos sí estuvieron en la UCI.

Tabla 31. Individuo en unidad de cuidados intensivos (UCI).

	Conteo	%	% acumulado
No aplica	17,024,699	92.9	92.9
No	1,204,971	6.6	99.5
Sí	84,347	0.5	99.9
No especificado	14,358	0.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 32 presenta el análisis de la variable TOMA_MUESTRA_LAB (núm. 30), indicando que en el 32.8% de los casos sí se tomó muestra de laboratorio.

Tabla 32. Se tomó muestra de laboratorio.

	Conteo	%	% acumulado
No	12,323,683	67.2	67.2
Sí	6,004,692	32.8	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 33 muestra el análisis del resultado de la prueba de laboratorio, con la variable RESULTADO_LAB (núm. 31), indicando que en el 67.2% no hubo muestra de laboratorio, por lo cual no hay resultado. El 13.1% fue positivo a SARS-CoV-2 y el 18.0% no fue positivo.

Tabla 33. Resultado de laboratorio.

	Conteo	%	% acumulado
No aplica (caso sin muestra)	12,323,683	67.2	67.2
No positivo a SARS-CoV-2	3,305,691	18.0	85.3
Positivo a SARS-CoV-2	2,408,661	13.1	98.4
Resultado no adecuado	287,804	1.6	100.0
Resultado pendiente	2,536	0.0	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 34 muestra la variable TOMA_MUESTRA_ANTÍGENO (núm. 32), indicando que en el 68.8% de los casos sí se tomó muestra.

Tabla 34. Se tomó muestra de antígeno.

	Conteo	%	% acumulado
Sí	12,615,618	68.8	68.8
No	5,712,757	31.2	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 35 muestra la variable RESULTADO_ANTÍGENO (núm. 33), en la que se observa que el 24.2% de los casos fueron positivos a SARS-CoV-2.

Tabla 35. Resultado de prueba de antígeno.

	Conteo	%	% acumulado
Negativo a SARS-CoV-2	8,174,748	44.6	44.6
No aplica (caso sin muestra)	5,712,757	31.2	75.8
Positivo a SARS-CoV-2	4,440,870	24.2	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 36 presenta el análisis de la cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso de la persona (variable núm. 53: DÍAS_SÍNTOMAS_A_INGRESO). La mayoría de los casos (80.9%) presentan entre 0 y 4 días entre ambas fechas y poco más de la cuarta parte (26.4%) fueron ingresados en la misma fecha de inicio de síntomas.

Tabla 36. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso al establecimiento de salud.

Cantidad de días	Conteo	%	% acumulado
0	4,837,158	26.4	26.4
2	2,890,530	15.8	42.2
1	2,842,877	15.5	57.7
3	2,529,657	13.8	71.5

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

4	1,721,050	9.4	80.9
5	1,205,049	6.6	87.4
6	667,100	3.6	91.1
7	613,132	3.3	94.4
8	297,603	1.6	96.0
9	186,844	1.0	97.1
10	170,214	0.9	98.0
11	95,887	0.5	98.5
14	74,012	0.4	98.9
12	68,092	0.4	99.3
13	58,485	0.3	99.6
15	39,503	0.2	99.8
Subtotal	18,297,193	99.8	*****
Otras cantidades de días	31,182	0.2	100.0
Total	18,328,375	100.0	

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 37 muestra el análisis de la cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción (variable núm. 54: *DÍAS_SÍNTOMAS_A_DEF*). Los fallecidos son 428,926 (2.3% del total). De estos, la mayoría (81.3%) falleció al haber transcurrido entre uno y veinte días. Poco más de la tercera parte (36.3%) falleció al tener entre cinco y once días transcurridos.

Tabla 37. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción.

	Conteo	%	% acumulado
7	23,853	5.6%	5.6%
8	23,557	5.5%	11.1%
9	22,614	5.3%	16.3%
6	22,000	5.1%	21.5%
10	21,884	5.1%	26.6%
5	21,277	5.0%	31.5%
11	20,450	4.8%	36.3%
12	19,888	4.6%	40.9%
4	19,302	4.5%	45.4%
13	18,328	4.3%	49.7%
14	17,871	4.2%	53.9%
15	16,696	3.9%	57.8%
3	16,419	3.8%	61.6%
16	15,359	3.6%	65.2%
17	13,748	3.2%	68.4%
2	12,674	3.0%	71.3%
18	12,327	2.9%	74.2%
19	10,984	2.6%	76.8%
20	10,015	2.3%	79.1%

CAPITULO 2. PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

1	9,374	2.2%	81.3%
Subtotal	348,620	81.3%	
Otras cantidades de días	80,306	18.7%	
Fallecimientos	428,926	100.0% (2.3% del total)	
Casos que no fallecieron	17,899,449	97.7%	
Total	18,328,375	100.0%	

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 38 presenta un análisis de la cantidad de días transcurridos entre la fecha de ingreso al establecimiento de salud y la fecha de defunción (variable núm. 55: *DÍAS_INGRESO_A_DEF*). De los fallecidos (2.3% de más de 18.3 millones de casos registrados), el 80.4% fallecieron entre 0 y 13 días después del ingreso, y el 36.0% falleció entre los 0 y 3 días.

Tabla 38. Cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción.

	Conteo	%	% acumulado
1	44,066	10.3%	10.3%
0	41,674	9.7%	20.0%
2	36,705	8.6%	28.5%
3	32,024	7.5%	36.0%
4	28,609	6.7%	42.7%
5	25,705	6.0%	48.7%
6	23,104	5.4%	54.1%
7	21,462	5.0%	59.1%
8	19,197	4.5%	63.5%
9	17,637	4.1%	67.7%
10	15,844	3.7%	71.3%
11	14,364	3.3%	74.7%
12	13,097	3.1%	77.7%
13	11,548	2.7%	80.4%
Subtotal	345,036	80.4%	
Otras cantidades de días	83,890	19.6%	
Fallecimientos	428,926	100.0% (2.3% del total)	
Casos que no fallecieron	17,899,449	97.7%	
Total	18,328,375	100.0	

Fuente: Elaboración propia con datos de SSA, 2022.

2.4 Contexto epidemiológico

En relación con el contexto epidemiológico del individuo, se analizan ocho atributos: 1) origen (alguna de las USMER o fuera de estas), 2) sector de servicios de salud que le atendió, 3) entidad federativa de la unidad médica que le atendió, 4) entidad federativa y municipio de residencia, 5) número de ola epidémica o de periodo interepidémico en que ocurrió el caso, 6) año y mes de ingreso al establecimiento de salud

que le atendió, 7) semana epidemiológica correspondiente al caso, y 8) si el individuo estuvo en contacto, o no, con otro caso de COVID-19.

La tabla 39 presenta los resultados de la variable ORIGEN, que se refiere a si la persona fue atendida dentro (valor _1) o fuera (valor _2) de alguna unidad de salud monitora de enfermedades respiratorias (USMER). Las USMER son establecimientos de salud que realizan vigilancia centinela.

Tabla 39. Origen.

Origen	Conteo	%	% acumulado
Fuera de USMER	14,045,250	76.6	76.6
USMER	4,283,125	23.4	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 40 muestra el análisis del tipo de institución del Sistema Nacional de Salud que dio atención a la persona (variable SECTOR). El 88.9% de los casos corresponde a los sectores SSA (identificador _12) e IMSS (identificador _4). El 11.1% fue atendido por alguno de los otros doce sectores, o bien, no fue especificado (código _99). El 53.3% fue atendido por la SSA y 35.7% recibió la atención del IMSS.

Tabla 40. Sector.

	Conteo	%	% acumulado
SSA	9,765,738	53.3	53.3
IMSS	6,536,591	35.7	88.9
Institución privada	1,128,471	6.2	95.1
ISSSTE	446,184	2.4	97.5
Estatad	186,227	1.0	98.6
IMSS-Bienestar	85,928	0.5	99.0
Pemex	77,316	0.4	99.4
Sedena	44,731	0.2	99.7
Semar	24,834	0.1	99.8
Municipal	14,870	0.1	99.9
Universitario	9,794	0.1	100.0
Cruz Roja	3,557	0.0	100.0
DIF	2,837	0.0	100.0
No especificado	1,261	0.0	100.0
_14 (erróneo)	36	0.0	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 41 muestra los resultados de la variable ENTIDAD_UM; es decir, la entidad federativa donde se ubica el establecimiento de salud que atendió a la persona. La mayoría (80.8%) de los casos fueron atendidos en 16 entidades. Las cinco entidades que atendieron más casos son: Ciudad de México (36.2%), Estado de México (6.2%), Nuevo León (4.5%), Guanajuato (4.4%) y San Luis Potosí (3.5%). Las cinco entidades que atendieron menos casos son: Zacatecas (0.9%), Nayarit (0.8%), Tlaxcala (0.7%),

CAPITULO 2. PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

Colima (0.7%) y Campeche (0.6%). Es destacable que la Ciudad de México atendió a más de la tercera parte de los casos. Esto se explica por la gran concentración de centros hospitalarios ubicados en la capital del país.

Tabla 41. Entidad federativa de la unidad médica de atención.

	Conteo	%	% acumulado
Ciudad de México	6,629,469	36.2	36.2
Estado de México	1,129,510	6.2	42.3
Nuevo León	823,095	4.5	46.8
Guanajuato	799,048	4.4	51.2
San Luis Potosí	639,189	3.5	54.7
Tabasco	602,088	3.3	58.0
Jalisco	591,170	3.2	61.2
Puebla	455,124	2.5	63.7
Coahuila	448,033	2.4	66.1
Sinaloa	411,503	2.2	68.4
Veracruz	405,409	2.2	70.6
Baja California	392,831	2.1	72.7
Tamaulipas	392,281	2.1	74.8
Sonora	385,033	2.1	77.0
Morelos	359,924	2.0	78.9
Querétaro	342,355	1.9	80.8
Michoacán	337,072	1.8	82.6
Baja California Sur	312,081	1.7	84.3
Yucatán	305,749	1.7	86.0
Chihuahua	287,837	1.6	87.6
Oaxaca	255,362	1.4	89.0
Guerrero	250,383	1.4	90.3
Chiapas	228,208	1.2	91.6
Aguascalientes	226,209	1.2	92.8
Hidalgo	219,862	1.2	94.0
Quintana Roo	217,295	1.2	95.2
Durango	188,893	1.0	96.2
Zacatecas	169,528	0.9	97.1
Nayarit	143,381	0.8	97.9
Tlaxcala	133,706	0.7	98.7
Colima	132,361	0.7	99.4
Campeche	114,386	0.6	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 42 presenta el análisis de la entidad federativa y municipio de residencia del individuo (variable núm. 36: ENT_MPIO_RES). En total, México tiene más de 2,400 municipios, por lo que la tabla muestra solamente el 79.0% más alto de los casos. En color gris, se marca el 33.4% más alto, que corresponden a solamente 15 municipios (o alcaldías de la Ciudad de México). De estos 15, 11 son alcaldías de la Ciudad de

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

México (CDMX). Los cinco máximos son: Iztapalapa, CDMX (_9_7), Gustavo A. Madero, CDMX (_9_5), Álvaro Obregón, CDMX (_9_10), Tlalpan, CDMX (_9_12) y San Luis Potosí, SLP (_24_28). Son destacables dos hallazgos: 1) solamente Iztapalapa tiene el 5.1% de los casos registrados y, 2) el municipio de San Luis Potosí tiene un porcentaje considerable (2.2%) de casos. Este último hallazgo podría explicarse con base en dos hechos: la población de este municipio en 2020 es de 911,908 habitantes (ranking 19 de todos los municipios de México), y 2) Knaul et al. (2021) sugieren que el estado de San Luis Potosí presentó el más bajo desempeño en políticas públicas diseñadas para informar al público, promover el distanciamiento y usar cubrebocas.

De la variable (núm. 7) municipio_res no se realiza análisis de Pareto, ya que esta es solamente un identificador de municipio dentro de cada entidad federativa, por lo que un mismo identificador puede ser igual para municipios de distintas entidades. Sin embargo, la variable ent_mpio_res, que constituye la concatenación del identificador de entidad federativa de residencia y del identificador de municipio de residencia, es analizada en la tabla 42.

Tabla 42. Entidad y municipio de residencia: solo el 79% más alto de los casos (parte 1 de 2).

Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado
1	_9_7	943,191	5.1	5.1	27	_2_4	172,134	0.9	47.2	53	_30_193	75,377	0.4	62.9
2	_9_5	730,468	4.0	9.1	28	_26_30	153,482	0.8	48.1	54	_27_2	74,848	0.4	63.3
3	_9_10	585,027	3.2	12.3	29	_15_104	148,940	0.8	48.9	55	_28_22	74,784	0.4	63.7
4	_9_12	443,073	2.4	14.7	30	_3_3	135,954	0.7	49.6	56	_11_7	70,812	0.4	64.1
5	_24_28	401,372	2.2	16.9	31	_2_2	135,816	0.7	50.4	57	_26_18	68,456	0.4	64.4
6	_9_3	385,462	2.1	19.0	32	_15_57	131,687	0.7	51.1	58	_2_1	64,906	0.4	64.8
7	_9_15	357,989	2.0	21.0	33	_9_9	129,930	0.7	51.8	59	_20_67	64,073	0.3	65.1
8	_9_17	345,755	1.9	22.9	34	_23_5	125,002	0.7	52.5	60	_25_12	63,910	0.3	65.5
9	_9_13	319,462	1.7	24.6	35	_5_30	123,034	0.7	53.2	61	_15_31	63,185	0.3	65.8
10	_27_4	279,320	1.5	26.1	36	_3_8	120,311	0.7	53.8	62	_25_1	63,154	0.3	66.2
11	_21_114	273,876	1.5	27.6	37	_9_4	116,424	0.6	54.5	63	_4_2	63,054	0.3	66.5
12	_11_20	270,686	1.5	29.1	38	_16_53	115,342	0.6	55.1	64	_19_46	62,738	0.3	66.9
13	_9_11	270,344	1.5	30.6	39	_19_26	110,584	0.6	55.7	65	_28_41	60,775	0.3	67.2
14	_9_2	267,669	1.5	32.0	40	_14_120	109,949	0.6	56.3	66	_28_32	59,888	0.3	67.5
15	_9_6	248,394	1.4	33.4	41	_10_5	109,709	0.6	56.9	67	_15_39	56,716	0.3	67.8
16	_9_14	247,151	1.3	34.8	42	_8_19	106,895	0.6	57.5	68	_13_48	53,997	0.3	68.1
17	_15_33	238,900	1.3	36.1	43	_12_1	102,026	0.6	58.0	69	_30_87	51,647	0.3	68.4
18	_19_39	234,137	1.3	37.3	44	_15_999	100,752	0.5	58.6	70	_15_109	51,280	0.3	68.7
19	_22_14	233,009	1.3	38.6	45	_15_106	99,570	0.5	59.1	71	_15_13	51,118	0.3	69.0
20	_9_8	218,696	1.2	39.8	46	_8_37	98,694	0.5	59.7	72	_15_121	50,771	0.3	69.2
21	_15_58	209,033	1.1	40.9	47	_5_35	96,012	0.5	60.2	73	_28_38	48,010	0.3	69.5
22	_9_16	200,919	1.1	42.0	48	_17_7	91,458	0.5	60.7	74	_14_67	47,917	0.3	69.8
23	_1_1	199,467	1.1	43.1	49	_19_6	85,309	0.5	61.1	75	_15_81	47,563	0.3	70.0
24	_31_50	198,070	1.1	44.2	50	_18_17	83,789	0.5	61.6	76	_19_21	47,484	0.3	70.3
25	_25_6	193,418	1.1	45.3	51	_15_25	78,747	0.4	62.0	77	_32_56	47,345	0.3	70.5
26	_14_39	193,108	1.1	46.3	52	_11_17	76,834	0.4	62.5	78	_27_13	46,963	0.3	70.8

Fuente: Elaboración propia con datos de (SSA, 2022).

CAPITULO 2. PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

Tabla 42 (cont.). Entidad y municipio de residencia: solo el 79% más alto de los casos (parte 2 de 2).

Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado
79	_19_48	46,106	0.3	71.1	100	23_8	34,775	0.2	75.7
80	_15_60	45,897	0.3	71.3	101	23_4	34,683	0.2	75.9
81	_7_101	45,729	0.2	71.6	102	_15_70	34,681	0.2	76.1
82	_12_29	45,125	0.2	71.8	103	_22_16	34,372	0.2	76.3
83	_32_17	44,944	0.2	72.0	104	_15_122	33,702	0.2	76.5
84	_14_98	44,706	0.2	72.3	105	_15_20	33,551	0.2	76.7
85	_6_2	44,327	0.2	72.5	106	_30_39	32,984	0.2	76.9
86	_11_27	43,204	0.2	72.8	107	_14_101	32,644	0.2	77.0
87	_19_31	43,161	0.2	73.0	108	_14_97	31,649	0.2	77.2
88	_16_52	42,851	0.2	73.2	109	_6_10	31,447	0.2	77.4
89	_24_35	41,591	0.2	73.5	110	_19_18	31,340	0.2	77.5
90	_5_18	41,572	0.2	73.7	111	_26_43	30,135	0.2	77.7
91	_10_7	41,265	0.2	73.9	112	_28_9	30,127	0.2	77.9
92	_17_11	40,871	0.2	74.1	113	_21_156	30,011	0.2	78.0
93	_24_13	38,965	0.2	74.4	114	_27_12	29,832	0.2	78.2
94	_15_37	37,940	0.2	74.6	115	_32_10	29,712	0.2	78.4
95	_17_6	37,525	0.2	74.8	116	_30_44	28,195	0.2	78.5
96	_25_11	37,497	0.2	75.0	117	_28_27	28,152	0.2	78.7
97	_5_25	35,707	0.2	75.2	118	_15_24	27,658	0.2	78.8
98	_7_89	35,688	0.2	75.4	119	_16_102	27,656	0.2	79.0
99	_11_15	35,531	0.2	75.6					

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 43 analiza las olas epidemiológicas y los periodos interepidémicos (variable núm. 56: OLA_O_PER_INTEREPI) en que se registraron los casos, incluyendo positivos, negativos y otros. Esto muestra que la segunda y la tercera olas por sí solas constituyen casi la mitad (48.2%) de los casos, la primera con el 25.2% y la segunda con el 23.0%.

Tabla 43. Número de ola epidémica o de periodo interepidémico.

	Conteo	%	% acumulado
Segunda ola	4,609,928	25.2	25.2
Tercera ola	4,216,972	23.0	48.2
Cuarta ola	3,161,030	17.2	65.4
Quinta ola	2,052,320	11.2	76.6
Primera ola	1,915,615	10.5	87.1
Segundo periodo interep.	916,286	5.0	92.1
Primer periodo interep.	693,510	3.8	95.8
Tercer periodo interep.	562,789	3.1	98.9
Cuarto periodo interep.	182,549	1.0	99.9
Otro tipo de semana	17,376	0.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 44 presenta el análisis del año y mes de ingreso del paciente al establecimiento de salud que le dio atención (variable núm. 51: AÑO_MES_INGRESO), indi-

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

cando que el 80.9% de los casos ingresaron en dieciocho meses (aunque no meses consecutivos). Los tres meses que tuvieron más ingresos fueron: enero de 2022, agosto de 2021 y julio de 2022.

Tabla 44. Año y mes de ingreso al establecimiento de salud.

AÑO_MES_INGRESO	Conteo	%	% acumulado
2022_1	2,225,910	12.1	12.1
2021_8	1,374,880	7.5	19.6
2022_7	1,128,135	6.2	25.8
2021_1	1,118,332	6.1	31.9
2021_7	1,083,126	5.9	37.8
2021_9	875,454	4.8	42.6
2020_12	866,708	4.7	47.3
2021_2	713,097	3.9	51.2
2021_3	631,249	3.4	54.7
2021_10	618,818	3.4	58.0
2022_2	613,272	3.3	61.4
2022_6	607,228	3.3	64.7
2021_12	539,587	2.9	67.6
2021_4	514,847	2.8	70.4
2021_11	506,440	2.8	73.2
2020_11	495,214	2.7	75.9
2021_6	464,303	2.5	78.4
2020_7	452,124	2.5	80.9
2020_10	443,498	2.4	83.3
2021_5	402,933	2.2	85.5
2020_8	402,369	2.2	87.7
2020_9	388,148	2.1	89.8
2022_8	377,314	2.1	91.9
2020_6	349,339	1.9	93.8
2022_3	275,341	1.5	95.3
2020_5	222,141	1.2	96.5
2022_5	197,104	1.1	97.6
2022_4	174,743	1.0	98.5
2020_4	114,145	0.6	99.2
2022_9	95,134	0.5	99.7
2020_3	34,280	0.2	99.9
2020_1	11,613	0.1	99.9
2020_2	11,549	0.1	100.0
Total	18,328,375		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 45 muestra el análisis de casos por semana epidemiológica, aunque solamente presentando el 80.4% más alto de los casos. Un patrón observable es que las semanas con mayores porcentajes son las cuatro primeras del año 2022, en este orden: 02, 03, 01 y 04. La semana 02 alcanzó un porcentaje de 3.7%.

CAPITULO 2. PANORAMA ESTADÍSTICO DESCRIPTIVO GENERAL DEL DATASET MASTER-3

Tabla 45. Semana epidemiológica: solo el 80.4% más alto de los casos.

Ranking	Semana	Conteo	%	% acum.	Ranking	Semana	Conteo	%	% acum.	Ranking	Semana	Conteo	%	% acum.
1	2022_02	674,173	3.7	3.7	27	2021_06	191,239	1.0	44.6	53	2021_45	124,697	0.7	66.6
2	2022_03	598,348	3.3	6.9	28	2020_53	187,032	1.0	45.7	54	2021_25	124,575	0.7	67.3
3	2022_01	430,896	2.4	9.3	29	2022_30	185,039	1.0	46.7	55	2021_16	123,666	0.7	68.0
4	2022_04	430,319	2.3	11.6	30	2021_37	178,363	1.0	47.6	56	2021_11	122,960	0.7	68.6
5	2021_32	350,599	1.9	13.6	31	2021_27	178,182	1.0	48.6	57	2022_24	118,254	0.6	69.3
6	2021_31	330,396	1.8	15.4	32	2022_25	174,185	1.0	49.6	58	2021_47	118,082	0.6	69.9
7	2021_30	321,639	1.8	17.1	33	2020_49	173,542	0.9	50.5	59	2020_46	115,136	0.6	70.6
8	2021_33	318,599	1.7	18.9	34	2021_52	172,599	0.9	51.5	60	2021_44	114,988	0.6	71.2
9	2022_28	311,889	1.7	20.6	35	2021_39	169,782	0.9	52.4	61	2020_44	113,516	0.6	71.8
10	2022_27	299,769	1.6	22.2	36	2021_08	168,552	0.9	53.3	62	2020_43	112,355	0.6	72.4
11	2021_29	291,556	1.6	23.8	37	2020_52	167,264	0.9	54.2	63	2021_48	111,467	0.6	73.0
12	2021_03	280,564	1.5	25.3	38	2021_07	161,883	0.9	55.1	64	2021_49	110,027	0.6	73.6
13	2021_02	277,370	1.5	26.8	39	2021_09	159,696	0.9	56.0	65	2021_17	106,647	0.6	74.2
14	2022_05	275,492	1.5	28.3	40	2022_06	159,520	0.9	56.8	66	2021_46	106,065	0.6	74.8
15	2021_34	272,886	1.5	29.8	41	2021_40	154,821	0.8	57.7	67	2020_47	104,917	0.6	75.4
16	2021_04	267,305	1.5	31.3	42	2021_41	150,249	0.8	58.5	68	2020_29	104,687	0.6	75.9
17	2022_29	263,550	1.4	32.7	43	2021_10	144,405	0.8	59.3	69	2020_30	104,575	0.6	76.5
18	2021_01	262,592	1.4	34.1	44	2021_42	141,508	0.8	60.1	70	2021_18	103,783	0.6	77.1
19	2022_26	242,368	1.3	35.5	45	2020_48	140,773	0.8	60.8	71	2021_24	103,566	0.6	77.6
20	2021_35	240,677	1.3	36.8	46	2021_12	135,762	0.7	61.6	72	2022_08	103,140	0.6	78.2
21	2021_28	234,626	1.3	38.1	47	2022_07	134,918	0.7	62.3	73	2021_50	101,573	0.6	78.7
22	2021_36	225,762	1.2	39.3	48	2021_14	133,949	0.7	63.0	74	2021_51	101,445	0.6	79.3
23	2020_51	201,825	1.1	40.4	49	2021_26	133,715	0.7	63.8	75	2020_45	100,285	0.5	79.8
24	2021_38	200,744	1.1	41.5	50	2022_31	133,587	0.7	64.5	76	2020_31	98,604	0.5	80.4
25	2021_05	193,746	1.1	42.5	51	2021_15	132,739	0.7	65.2					
26	2020_50	193,259	1.1	43.6	52	2021_43	129,387	0.7	65.9					

Fuente: Elaboración propia con datos de (SSA, 2022).

En la Tabla 46 se observa el análisis de la variable *OTRO_CASO* (núm. 29); es decir, si el paciente tuvo, o no, contacto con algún otro caso diagnosticado con SARS CoV-2. El 32.4% de los pacientes sí lo tuvo.

Tabla 46. Estuvo en contacto con otro caso de COVID-19.

	Conteo	%	% acumulado
No	11,770,474	64.2	64.2
Sí	5,932,849	32.4	96.6
No especificado	625,052	3.4	100
Total	18,328,375		

Fuente: Elaboración propia con datos de (SSA, 2022).

CAPÍTULO 3

PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS POSITIVOS DEL DATASET MASTER-3

Este capítulo se enfoca en presentar los resultados de estadística descriptiva solamente de los casos positivos, de las variables disponibles en el *dataset Master-3* (que está basado en SSA, 2022), con base en la variable *CLASIFICACIÓN_FINAL*. Los casos positivos de Master-3 son poco más de 7 millones, con las mismas 56 variables que se tomaron en el capítulo 2, donde se analizaron todos los registros de ese dataset. En el capítulo actual, el principal tipo de análisis realizado también es el de Pareto (80:20).

Los casos positivos se están analizando por separado porque generan un interés especialmente importante por razones obvias. En términos generales, los análisis de estadística descriptiva sirven para descubrir los patrones más generales y más fáciles de detectar en el *dataset Master-3* y también tienen la valiosa utilidad de servir para evaluar los modelos de árboles clasificadores (o reglas clasificadoras) que se producen usando algoritmos de aprendizaje automático. Del mismo modo que en el capítulo 2, se presenta a continuación una amplia serie de varias decenas de tablas de conteos y porcentajes. Estas también se agrupan en cuatro categorías: características del paciente, historia clínica y comorbilidades, curso clínico y resultados, y contexto epidemiológico.

3.1 Características del paciente

A continuación, se analizan quince atributos de los pacientes que resultaron positivos a SARS-CoV-2: 1) falleció o no, 2) fue intubado o no, 3) tipo de paciente (ambulatorio u hospitalizado), 4) clasificación final del caso (solo considerando las categorías 1, 2 y 3 de esta variable), 5) entidad federativa de nacimiento, 6) entidad de residencia, 7) nacionalidad, 8) estatus de indígena, 9) habla alguna lengua indígena o no, 10) sexo, 11) embarazo, 12) es paciente pediátrico de 0 a 18 años (sí o no), 13) es paciente pediátrico de 0 años (sí o no), 14) tiene 65 años o más de edad (sí o no), y 15) intervalo de edad.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

La tabla 47 presenta el análisis de la variable *FALLECIÓ* (núm. 38), mostrando que los fallecidos (329,896) constituyen el 4.7% de los casos positivos.

Tabla 47. Falleció: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,739,647	95.3	95.3
Sí	329,896	4.7	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 48 muestra el análisis de la variable Intubado, indicando que el 1.2% de los casos positivos fueron intubados.

Tabla 48. Intubado: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No aplica	6,360,846	90.0	90.0
No	618,471	8.7	98.7
Sí	84,558	1.2	99.9
No especificado	5,668	0.1	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

Por su parte, la tabla 49 presenta el análisis del tipo de paciente, indicando que la gran mayoría (90.0%) fueron ambulatorios.

Tabla 49. Tipo de paciente: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Ambulatorio	6,360,846	90.0	90.0
Hospitalizado	708,697	10.0	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

De la variable CLASIFICACIÓN_FINAL (núm. 34), la tTabla 50 presenta un análisis de los 7,069,543 casos positivos que corresponden a los identificadores 1, 2 y 3 de esta. La tTabla muestra que casi todos (94.1%) los casos corresponden al tipo 1, denominado Caso de SARS-CoV-2 confirmado.

Tabla 50. Clasificación final del caso: solo en casos positivos según las categorías 1, 2 y 3 de esta variable.

	Conteo	%	% acumulado
Caso de SARS-CoV-2 confirmado (identificador _1)	6,694,337	94.7	94.7
Caso de COVID-19 confirmado por asociación clínica epidemiológica (identificador _2)	359,294	5.1	99.8
Caso de COVID-19 confirmado por comité de dictaminación (identificador _3)	15,912	0.2	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La tabla 51 presenta el análisis de la entidad federativa de nacimiento (variable ENTIDAD_NAC) de las personas registradas, indicando que la mayoría (81.8%) de los casos corresponden a solamente 17 entidades. Una cuarta parte (25.1%) de todo el país corresponde solamente a Ciudad de México, seguida por el Estado de México (9.3%) y Guanajuato (5.0%).

Tabla 51. Entidad federativa de nacimiento: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Ciudad de México	1,775,415	25.1	25.1
Estado de México	655,591	9.3	34.4
Guanajuato	351,106	5.0	39.4
Nuevo León	342,352	4.8	44.2
Veracruz	303,816	4.3	48.5
Jalisco	285,316	4.0	52.5
San Luis Potosí	240,781	3.4	55.9
Puebla	231,940	3.3	59.2
Tabasco	219,052	3.1	62.3
Sinaloa	201,282	2.8	65.2
Sonora	187,439	2.7	67.8
Coahuila	183,688	2.6	70.4
Oaxaca	179,191	2.5	72.9
Tamaulipas	171,284	2.4	75.4
Guerrero	157,249	2.2	77.6
Chihuahua	151,499	2.1	79.7
Yucatán	144,413	2.0	81.8
Michoacán	140,341	2.0	83.8
Querétaro	121,556	1.7	85.5
Hidalgo	118,638	1.7	87.2
Baja California	97,324	1.4	88.5
Durango	89,381	1.3	89.8
Zacatecas	86,868	1.2	91.0
Morelos	83,630	1.2	92.2
Baja California Sur	82,971	1.2	93.4
Aguascalientes	81,519	1.2	94.5
Chiapas	76,167	1.1	95.6
Nayarit	70,788	1.0	96.6
Tlaxcala	58,413	0.8	97.4
Quintana Roo	57,569	0.8	98.3
Colima	56,349	0.8	99.1
Campeche	41,767	0.6	99.6
No especificado	24,848	0.4	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 52 muestra los resultados de la entidad federativa de residencia (variable ENTIDAD_RES), con patrones estadísticos muy similares a los de la entidad federativa de nacimiento.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Tabla 52. Entidad federativa de residencia: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Ciudad de México	1,734,072	24.5	24.5
Estado de México	700,218	9.9	34.4
Nuevo León	399,986	5.7	40.1
Guanajuato	351,110	5.0	45.1
Jalisco	283,503	4.0	49.1
San Luis Potosí	238,125	3.4	52.4
Veracruz	223,831	3.2	55.6
Tabasco	213,226	3.0	58.6
Puebla	207,617	2.9	61.6
Sonora	194,746	2.8	64.3
Coahuila	180,009	2.5	66.9
Querétaro	176,603	2.5	69.4
Tamaulipas	175,890	2.5	71.8
Sinaloa	175,037	2.5	74.3
Chihuahua	164,385	2.3	76.6
Baja California	163,871	2.3	79.0
Oaxaca	149,139	2.1	81.1
Yucatán	136,562	1.9	83.0
Baja California Sur	124,337	1.8	84.8
Hidalgo	119,135	1.7	86.4
Guerrero	115,607	1.6	88.1
Quintana Roo	111,585	1.6	89.7
Michoacán	109,675	1.6	91.2
Morelos	90,541	1.3	92.5
Aguascalientes	84,329	1.2	93.7
Zacatecas	79,514	1.1	94.8
Durango	79,240	1.1	95.9
Nayarit	71,926	1.0	96.9
Colima	67,182	1.0	97.9
Tlaxcala	56,379	0.8	98.7
Chiapas	49,333	0.7	99.4
Campeche	42,830	0.6	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 53 muestra que, respecto a nacionalidad, las personas extranjeras constituyen solamente el 0.4% de los positivos.

Tabla 53. Nacionalidad: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Mexicana	7,044,753	99.6	99.6
Extranjera	24,790	0.4	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La tabla 54 muestra que solamente el 0.7% de los casos positivos de COVID eran de población indígena, cifra similar a la de la tabla posterior.

Tabla 54. Indígena: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,716,037	95.0	95.0
No especificado	302,817	4.3	99.3
Sí	50,689	0.7	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 55 presenta una cifra de 0.5% de población que habla una lengua indígena en los casos positivos de COVID. Este porcentaje es similar al de población indígena, mostrado en la tabla precedente.

Tabla 55. Habla alguna lengua indígena: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,724,547	95.1	95.1
No especificado	309,184	4.4	99.5
Sí	35,812	0.5	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 56 presenta el análisis de la variable *SEXO* (núm. 4), indicando que poco más de la mitad (53.2%) de los individuos son mujeres.

Tabla 56. Sexo: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Mujer	3,760,702	53.2	53.2
Hombre	3,308,841	46.8	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 57. Embarazo: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	3,671,804	51.9	51.9
No aplica	3,308,841	46.8	98.7
Sí	61,236	0.9	99.6
Se ignora	27,657	0.4	100.0
No especificado	5	0.0	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 58 identifica los casos que se encuentran en edad pediátrica (entre 0 y 18 años), mediante el análisis de la variable *PEDIA_0_18* (núm. 39), mostrando que estos constituyen el 7.3% de los positivos.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Tabla 58. Paciente pediátrico (entre 0 y 18 años de edad): solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,549,949	92.7	92.7
Sí	519,594	7.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La tabla 59 identifica los casos pediátricos de menos de un año de edad (variable núm. 40: *PEDIA_0*), lo cual muestra que estos constituyen el 0.2% de los casos positivos.

Tabla 59. Casos pediátricos de 0 (cero) años de edad: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	7,058,333	99.8	99.8
Sí	11,210	0.2	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

Debido al interés especial que genera el segmento de los adultos mayores, la tTabla 60 muestra los casos positivos con edad de 65 años o más (variable núm. 50: *EDAD_65YMAS*). Estos constituyen el 8.7% del total de los positivos.

Tabla 60. Individuos de 65 años o más: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,455,097	91.3	91.3
Sí	614,446	8.7	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

El análisis de la variable *EDAD* (núm. 14) es abordado en el de la variable *EDAD_INTERVALO* (núm. 52) en la tabla 61. También esa tabla aborda implícitamente las variables de los intervalos de edades *PEDIA_1_4*, *PEDIA_5_9*, *PEDIA_10_14*, *PEDIA_15_19*, *EDAD_20_24*, *EDAD_25_44*, *EDAD_45_49*, *EDAD_50_59* y *EDAD_60_64*. Estos intervalos son los establecidos por la Secretaría de Salud (SSA) y utilizados para sus análisis estadísticos. La mayoría (87.0%) de los casos pertenece a cinco de los once intervalos. Los cinco corresponden a los segmentos de 20 a 59 años y de 65 y más; curiosamente, el intervalo de 60 a 64 años no está dentro de estos cinco intervalos. El intervalo de 25 a 44 años constituye casi la mitad (45.2%) de los casos.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 61. Intervalos de edad: solo en casos positivos de COVID.

	Conteo	%	% acumulado
25 a 44	3,194,467	45.2	45.2
50 a 59	1,010,448	14.3	59.5
45 a 49	681,450	9.6	69.1
20 a 24	652,167	9.2	78.3
65 o más	614,446	8.7	87.0
60 a 64	315,629	4.5	91.5
15 a 19	270,897	3.8	95.3
10 a 14	164,686	2.3	97.7
5 a 9	95,623	1.4	99.0
1 a 4	58,520	0.8	99.8
0	11,210	0.2	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

3.2 Historia clínica y comorbilidades

En cuanto a la historia clínica y las comorbilidades de los casos positivos, se analizan once variables: 1) diabetes, 2) inmunosupresión, 3) obesidad, 4) insuficiencia renal crónica, 5) neumonía, 6) EPOC, 7) asma, 8) hipertensión, 9) enfermedad cardiovascular, 10) tabaquismo, y 11) alguna otra comorbilidad. Las tablas 62 a 72 presentan estos análisis.

Tabla 62. Diabetes: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,425,928	90.9	90.9
Sí	622,041	8.8	99.7
Se ignora	21,574	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 63. Inmunosupresión: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	7,009,118	99.1	99.1
Sí	39,747	0.6	99.7
Se ignora	20,678	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 64. Obesidad: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,366,193	90.1	90.1
Sí	683,870	9.7	99.7
Se ignora	19,480	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Tabla 65. Insuficiencia renal crónica: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,981,805	98.8	98.8
Sí	67,177	1.0	99.7
Se ignora	20,561	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 66 presenta el análisis de la variable Neumonía, mostrando que el 7.3% de los casos tuvieron este padecimiento.

Tabla 66. Neumonía: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,525,547	92.3	92.3
Sí	516,890	7.3	99.6
No especificado	27,106	0.4	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

Tabla 67. EPOC: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	7,000,645	99.0	99.0
Sí	48,251	0.7	99.7
Se ignora	20,647	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

Tabla 68. Asma: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,916,978	97.8	97.8
Sí	132,109	1.9	99.7
Se ignora	20,456	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 69. Hipertensión: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,208,160	87.8	87.8
Sí	840,752	11.9	99.7
Se ignora	20,631	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 70. Enfermedad cardiovascular: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,978,597	98.71	98.71
Sí	70,279	0.99	99.71
Se ignora	20,667	0.29	100.00
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La tabla 71 analiza si el paciente presenta tabaquismo (variable núm. 28: TABAQUISMO), indicando su presencia en el 5.5% de casos.

Tabla 71. Tabaquismo: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,659,039	94.2	94.2
Sí	389,457	5.5	99.7
Se ignora	21,047	0.3	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 72. Alguna otra comorbilidad: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	6,907,416	97.7	97.7
Sí	104,194	1.5	99.2
Se ignora	57,933	0.8	100.0
Total	7,069,543		

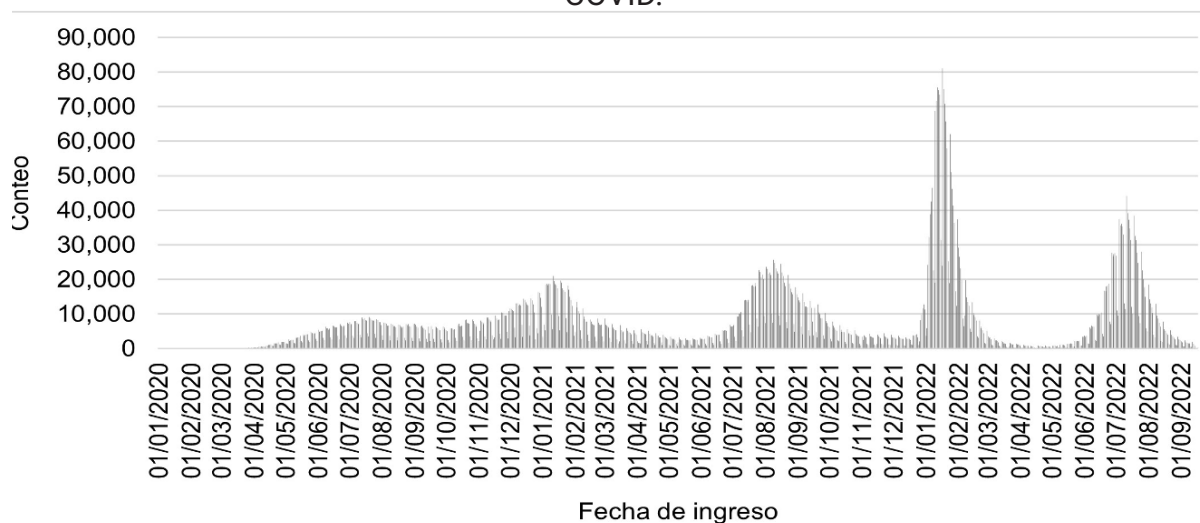
Fuente: Elaboración propia con datos de SSA, 2022.

3.3 Curso clínico y resultados

La figura 5 presenta los conteos de casos positivos por fecha de ingreso. Los cinco picos, correspondientes a las cinco olas epidémicas se observan en este orden, de mayor a menor frecuencia y sus periodos aproximados: cuarta ola, entre enero y febrero de 2022; quinta ola, entre julio y septiembre de 2022; tercera ola, entre julio y octubre de 2021; segunda ola, entre noviembre de 2020 y marzo de 2021, y finalmente, primera ola, de junio a septiembre de 2020.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

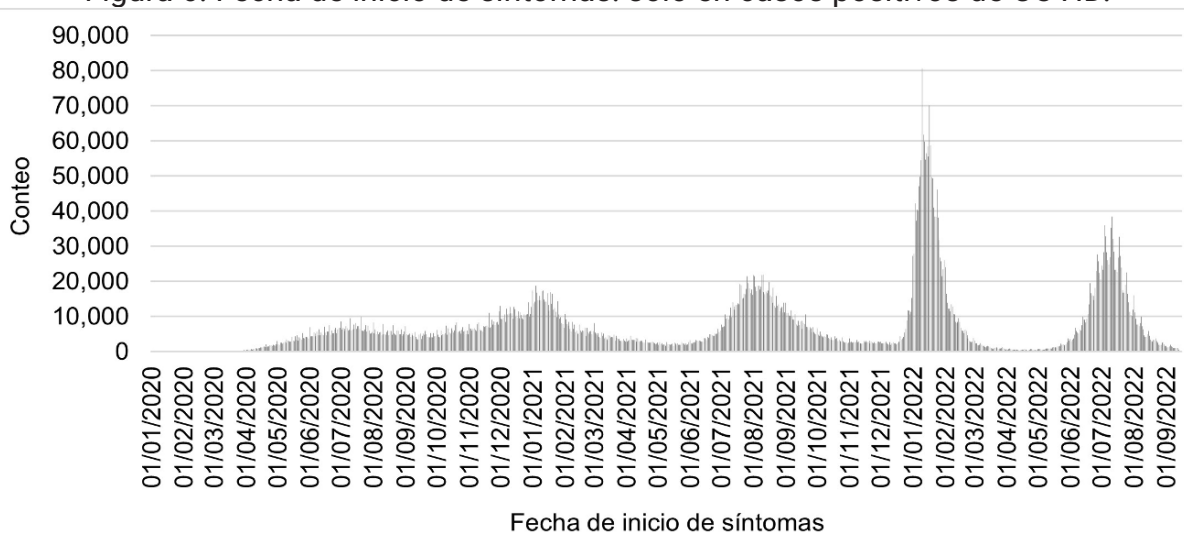
Figura 5. Fecha de ingreso al establecimiento de salud: solo en casos positivos de COVID.



Fuente: Elaboración propia con datos de (SSA, 2022).

La Figura 6 presenta conteos de casos por fecha de inicio de síntomas, mostrando patrones de comportamiento similares a los de la Figura 5.

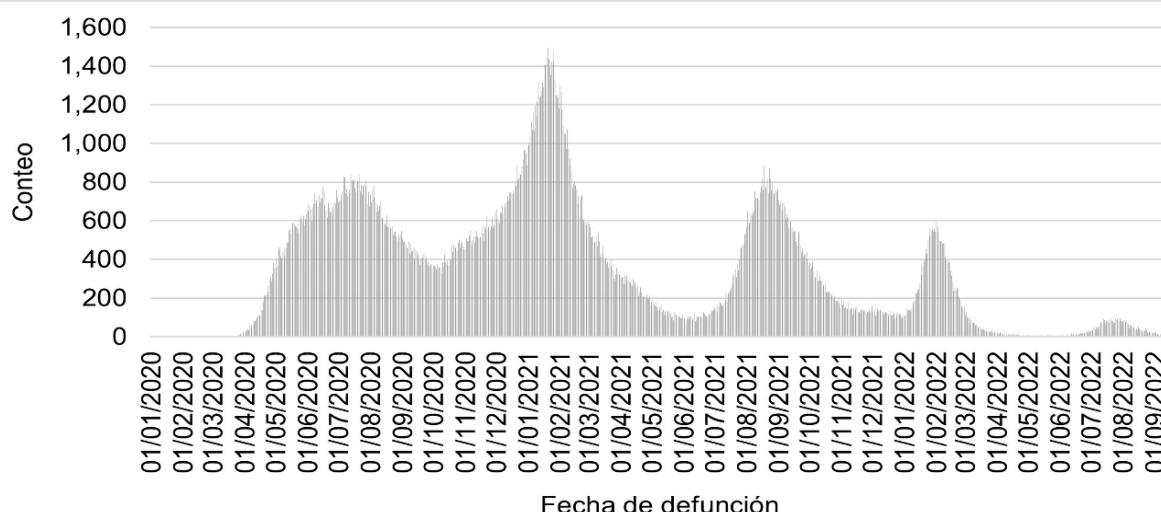
Figura 6. Fecha de inicio de síntomas: solo en casos positivos de COVID.



Fuente: Elaboración propia con datos de (SSA, 2022).

La Figura 7 presenta los conteos de defunciones por fecha (variable FECHA_DEF). Este análisis permite especular sobre la virulencia y la letalidad de las variantes del SARS-CoV-2. Durante la segunda fase intensiva de transmisión del virus (segunda ola), con circulación de la variante Delta, hubo el mayor número de registros de defunciones entre octubre de 2020 a marzo de 2021. La gráfica también muestra un descenso sostenido del registro de defunciones hasta septiembre de 2022. El resultado de la reducción de la mortalidad, teorizamos, es resultado de la combinación de diversos factores; entre ellos: el agotamiento de población susceptible, variantes menos virulentas, mejora en el conocimiento del manejo de casos de COVID-19 complicados y, principalmente, por la vacunación específica.

Figura 7. Fecha de defunción: solo en casos positivos de COVID.



Fuente: Elaboración propia con datos de SSA, 2022).

La tabla 73 analiza si se usó, o no, la unidad de cuidados intensivos (variable núm. 35: UCI), mostrando que en la gran mayoría (90.0%) de los casos no se usó y solamente en el 0.8%, sí.

Tabla 73. Individuo en unidad de cuidados intensivos (UCI): solo en casos positivos de COVID.

	Conteo	%	% acumulado
No aplica	6,360,846	89.9	89.9
No	648,961	9.2	99.1
Sí	54,072	0.8	99.9
No especificado	5,664	0.1	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 74 analiza si se realizó toma de muestra de laboratorio (variable núm. 30: TOMA_MUESTRA_LAB), indicando que en 65.1% de los casos no se realizó.

Tabla 74. Se tomó muestra de laboratorio: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	4,600,735	65.1	65.1
Sí	2,468,808	34.9	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 75 presenta el resultado de laboratorio (variable núm. 31: RESULTADO_LAB). Esto muestra que en casi dos terceras partes (65.1%) de los casos no hubo muestra y, en poco más de la tercera parte (34.1%) fue positivo a SARS-CoV-2. La clasificación final de algunos casos puede ser “positivo” aunque el resultado de laboratorio no lo sea, ya que la clasificación final pudo haberse definido con base en otros criterios.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Tabla 75. Resultado de laboratorio: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No aplica (caso sin muestra)	4,600,735	65.1	65.1
Positivo a SARS-CoV-2	2,408,661	34.1	99.1
Resultado no adecuado	44,428	0.6	99.8
No positivo a SARS-CoV-2	15,591	0.2	100.0
Resultado pendiente	128	0.0	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La Tabla 76 analiza si se tomó, o no, muestra de antígeno al paciente (variable núm. 32: *TOMA_MUESTRA_ANTÍGENO*), indicando que en más de las dos terceras partes (68.4%) sí se tomó.

Tabla 76. Se tomó muestra de antígeno: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Sí	4,838,505	68.4	68.4
No	2,231,038	31.6	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 77 muestra la variable *RESULTADO_ANTÍGENO* (núm. 33), en la que el 62.8% de los casos fue positivo a SARS-CoV-2. La clasificación final de algunos casos puede ser “positivo” aunque la prueba de antígeno no lo sea; esto se debe a que la clasificación final pudo haberse definido con base en otros criterios.

Tabla 77. Resultado de prueba de antígeno: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Positivo a SARS-CoV-2	4,436,545	62.8	62.8
No aplica (caso sin muestra)	2,231,038	31.6	94.3
Negativo a SARS-CoV-2	401,960	5.7	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La Tabla 78 muestra el análisis de cantidades de casos con base en la cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso al establecimiento de atención (variable núm. 53: *DÍAS_SÍNTOMAS_A_INGRESO*). Se observa que la gran mayoría (84.9%) ocurrió entre los 0 y los 5 días; la mitad (50.8%), entre los 0 y los 2 días.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 78. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso: solo en casos positivos de COVID.

	Conteo	%	% acumulado
2	1,263,090	17.9	17.9
1	1,169,407	16.5	34.4
0	1,159,042	16.4	50.8
3	1,117,571	15.8	66.6
4	766,673	10.8	77.5
5	524,476	7.4	84.9
Subtotal	6,000,259	84.9	
Otras cantidades de días	1,069,284	15.1	
Total	7,069,543	100.0	

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 79 presenta la cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción de los casos positivos (variable núm. 54: DÍAS_SÍNTOMAS_A_DEF). La gran mayoría (81.4%) de los 329,896 fallecimientos ocurrió entre los 2 y los 21 días después de la fecha de inicio de síntomas. La mitad (50.0%) ocurrió entre los 5 y los 14 días.

Tabla 79. Cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de defunción: solo en casos positivos de COVID.

	Conteo	%	% acumulado
1	30,265	9.2%	9.2%
0	27,619	8.4%	17.5%
2	26,688	8.1%	25.6%
3	24,403	7.4%	33.0%
4	22,255	6.7%	39.8%
5	20,248	6.1%	45.9%
6	18,503	5.6%	51.5%
7	17,390	5.3%	56.8%
8	15,689	4.8%	61.6%
9	14,514	4.4%	66.0%
10	13,154	4.0%	69.9%
11	11,929	3.6%	73.6%
12	10,887	3.3%	76.9%
13	9,609	2.9%	79.8%
14	8,331	2.5%	82.3%
Subtotal	271,484	82.3%	
Otras cantidades de días	58,412	17.7%	
Fallecimientos	329,896	100.0%	
Casos que no fallecieron	6,739,647	95.3%	
Total	7,069,543	100.0%	

Fuente: Elaboración propia con datos de SSA, 2022.

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

La tabla 80 presenta la cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción de los casos positivos (variable núm. 55:DÍAS_INGRESO_A_DEF). La gran mayoría (82.3%) de los 329,896 fallecimientos ocurrieron entre los 0 y los 14 días después de la fecha de ingreso de la persona al establecimiento de salud. Poco más de la mitad (51.5%) ocurrió entre los 0 y los 6 días.

Tabla 80. Cantidad de días transcurridos entre la fecha de ingreso y la fecha de defunción: solo en casos positivos de COVID.

	Conteo	%	% acumulado
1	30,265	9.2%	9.2%
0	27,619	8.4%	17.5%
2	26,688	8.1%	25.6%
3	24,403	7.4%	33.0%
4	22,255	6.7%	39.8%
5	20,248	6.1%	45.9%
6	18,503	5.6%	51.5%
7	17,390	5.3%	56.8%
8	15,689	4.8%	61.6%
9	14,514	4.4%	66.0%
10	13,154	4.0%	69.9%
11	11,929	3.6%	73.6%
12	10,887	3.3%	76.9%
13	9,609	2.9%	79.8%
14	8,331	2.5%	82.3%
Subtotal	271,484	82.3%	
Otras cantidades de días	58,412	17.7%	
Fallecimientos	329,896	100.0%	
Casos que no fallecieron	6,739,647	95.3%	
Total	7,069,543	100.0%	

3.4 Contexto epidemiológico

La tabla 81 analiza el origen de los casos, refiriéndose a si provienen, o no, de alguna Unidad de Salud Monitora de Enfermedades Respiratorias (USMER) . La mayoría (76.9%) de los casos no provienen de las USMER.

Tabla 81. Origen: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Fuera de USMER	5,439,446	76.9	76.9
USMER	1,630,097	23.1	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La Tabla 82 analiza la cantidad de casos positivos atendidos por cada uno de los diversos sectores de los servicios de salud. La mayoría (89.2%) corresponden a solamente dos sectores: IMSS (47.0%) y SSA (42.2%). El resto está disperso entre doce

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

sectores; de estos, los tres más significativos son: privado (4.8%), ISSSTE (2.9%) y estatal (1.1%).

Tabla 82. Sector: solo en casos positivos de COVID.

	Conteo	%	% acumulado
IMSS	3,325,546	47.0	47.0
SSA	2,983,344	42.2	89.2
Institución privada	341,437	4.8	94.1
ISSSTE	204,094	2.9	97.0
Estatad	78,658	1.1	98.1
Pemex	40,616	0.6	98.6
IMSS-Bienestar	40,499	0.6	99.2
Sedena	25,848	0.4	99.6
Semar	14,545	0.2	99.8
Municipal	7,105	0.1	99.9
Universitario	4,392	0.1	100.0
Cruz Roja	1,446	0.0	100.0
DIF	1,290	0.0	100.0
No especificado	722	0.0	100.0
_14 (erróneo)	1	0.0	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 83 analiza la cantidad de casos positivos con base en la entidad federativa de ubicación de la unidad médica de atención (variable núm. 3: *ENTIDAD_UM*). La mayoría (81.2%) de los casos fueron atendidos en 17 entidades. Las cinco entidades con más casos son: Ciudad de México, Estado de México, Nuevo León, Guanajuato y Jalisco. Por otra parte, las cinco con menos son: Nayarit, Colima, Tlaxcala, Chiapas y Campeche.

Tabla 83. Entidad federativa de la unidad médica de atención: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Ciudad de México	1,955,667	27.7	27.7
Estado de México	490,906	6.9	34.6
Nuevo León	399,850	5.7	40.3
Guanajuato	350,528	5.0	45.2
Jalisco	282,257	4.0	49.2
San Luis Potosí	240,273	3.4	52.6
Veracruz	219,043	3.1	55.7
Tabasco	214,967	3.0	58.8
Puebla	209,610	3.0	61.7
Sonora	194,909	2.8	64.5
Coahuila	180,089	2.5	67.0
Tamaulipas	177,159	2.5	69.5
Querétaro	177,048	2.5	72.0

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Sinaloa	175,309	2.5	74.5
Chihuahua	164,295	2.3	76.8
Baja California	162,507	2.3	79.1
Oaxaca	147,968	2.1	81.2
Yucatán	136,581	1.9	83.2
Baja California Sur	124,348	1.8	84.9
Hidalgo	116,760	1.7	86.6
Guerrero	114,143	1.6	88.2
Quintana Roo	110,898	1.6	89.8
Michoacán	108,142	1.5	91.3
Morelos	89,149	1.3	92.5
Aguascalientes	84,291	1.2	93.7
Zacatecas	79,816	1.1	94.9
Durango	78,288	1.1	96.0
Nayarit	71,841	1.0	97.0
Colima	67,925	1.0	97.9
Tlaxcala	53,483	0.8	98.7
Chiapas	47,238	0.7	99.4
Campeche	44,255	0.6	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

Al igual que en el capítulo 2, no se produce un análisis de Pareto de la variable específica del municipio de residencia (MUNICIPIO_RES). Sin embargo, sí se analiza esta variable en combinación (concatenación) con el identificador de entidad federativa de residencia (variable ENT_MPIO_RES). Este análisis se presenta en la tabla 84.

Tabla 84. Entidad y municipio de residencia: solo el 78.3%
más alto de casos positivos de COVID (parte 1 de 2)

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado
1	_9_7	254,586	3.6	3.6	27	_8_19	66,681	0.9	42.6	53	_16_53	31,886	0.5	59.3
2	_9_10	224,122	3.2	6.8	28	_23_5	63,603	0.9	43.5	54	_30_87	31,162	0.4	59.8
3	_9_5	214,390	3.0	9.8	29	_9_16	59,975	0.8	44.3	55	_2_1	30,254	0.4	60.2
4	_9_12	167,718	2.4	12.2	30	_3_3	58,261	0.8	45.2	56	_19_46	30,241	0.4	60.6
5	_24_28	143,909	2.0	14.2	31	_2_2	56,467	0.8	46.0	57	_25_1	28,906	0.4	61.0
6	_21_114	132,607	1.9	16.1	32	_15_104	54,704	0.8	46.7	58	_13_48	28,875	0.4	61.4
7	_11_20	129,914	1.8	17.9	33	_8_37	53,415	0.8	47.5	59	_9_4	28,432	0.4	61.8
8	_22_14	117,314	1.7	19.6	34	_15_57	53,077	0.8	48.2	60	_15_25	27,820	0.4	62.2
9	_19_39	112,212	1.6	21.2	35	_14_120	52,898	0.7	49.0	61	_17_7	27,639	0.4	62.6
10	_9_3	104,481	1.5	22.7	36	_19_26	51,213	0.7	49.7	62	_28_32	27,626	0.4	63.0
11	_27_4	104,425	1.5	24.1	37	_5_30	49,379	0.7	50.4	63	_19_48	26,525	0.4	63.4
12	_9_15	96,499	1.4	25.5	38	_9_8	48,854	0.7	51.1	64	_25_12	26,354	0.4	63.8
13	_9_2	96,438	1.4	26.9	39	_10_5	45,426	0.6	51.7	65	_28_38	26,167	0.4	64.1
14	_14_39	92,404	1.3	28.2	40	_3_8	44,652	0.6	52.4	66	_15_121	24,291	0.3	64.5
15	_9_13	92,205	1.3	29.5	41	_12_1	43,750	0.6	53.0	67	_19_21	24,035	0.3	64.8
16	_31_50	90,688	1.3	30.8	42	_19_6	42,318	0.6	53.6	68	_15_39	23,971	0.3	65.2
17	_15_33	88,765	1.3	32.0	43	_20_67	42,304	0.6	54.2	69	_6_2	23,481	0.3	65.5
18	_9_11	86,153	1.2	33.2	44	_30_193	40,473	0.6	54.8	70	_12_29	22,574	0.3	65.8
19	_25_6	80,471	1.1	34.4	45	_15_106	40,228	0.6	55.3	71	_4_2	22,109	0.3	66.1
20	_9_17	79,703	1.1	35.5	46	_18_17	39,530	0.6	55.9	72	_32_56	21,953	0.3	66.4
21	_26_30	78,980	1.1	36.6	47	_9_9	37,851	0.5	56.4	73	_28_41	21,512	0.3	66.7
22	_1_1	74,659	1.1	37.7	48	_11_17	37,175	0.5	57.0	74	_14_67	21,324	0.3	67.0
23	_9_6	72,667	1.0	38.7	49	_5_35	36,527	0.5	57.5	75	_15_31	21,115	0.3	67.3
24	_15_58	72,031	1.0	39.7	50	_26_18	34,360	0.5	58.0	76	_15_109	20,946	0.3	67.6
25	_9_14	69,966	1.0	40.7	51	_11_7	32,615	0.5	58.4	77	_19_31	20,559	0.3	67.9
26	_2_4	66,864	0.9	41.6	52	_28_22	31,981	0.5	58.9	78	_23_8	20,350	0.3	68.2

Fuente: Elaboración propia con datos de (SSA, 2022).

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

Tabla 84 (cont.). Entidad y municipio de residencia: solo el 78.3% más alto de casos positivos de COVID (parte 2 de 2).

Ranking	Entidad y municipio	Conteo	%	% acumulado	Ranking	Entidad y municipio	Conteo	%	% acumulado
79	_15_60	20,183	0.3	68.5	103	_14_97	14,589	0.2	74.2
80	_14_98	20,158	0.3	68.8	104	_28_3	14,432	0.2	74.4
81	_15_13	20,138	0.3	69.1	105	_30_39	14,254	0.2	74.6
82	_32_17	19,490	0.3	69.3	106	_6_7	13,957	0.2	74.8
83	_22_16	18,979	0.3	69.6	107	_13_51	13,906	0.2	75.0
84	_27_13	18,763	0.3	69.9	108	_11_15	13,568	0.2	75.2
85	_24_13	18,390	0.3	70.1	109	_26_43	13,537	0.2	75.4
86	_5_18	17,890	0.3	70.4	110	_30_44	13,337	0.2	75.6
87	_27_2	17,004	0.2	70.6	111	_28_27	12,681	0.2	75.8
88	_19_18	16,937	0.2	70.9	112	_32_10	12,642	0.2	76.0
89	_24_35	16,881	0.2	71.1	113	_15_122	12,459	0.2	76.1
90	_14_101	16,872	0.2	71.3	114	_15_70	12,416	0.2	76.3
91	_23_4	16,843	0.2	71.6	115	_17_6	12,401	0.2	76.5
92	_15_81	16,580	0.2	71.8	116	_16_102	12,320	0.2	76.7
93	_7_101	16,246	0.2	72.0	117	_18_20	12,317	0.2	76.8
94	_11_27	16,197	0.2	72.3	118	_30_131	12,204	0.2	77.0
95	_15_999	16,104	0.2	72.5	119	_30_118	11,945	0.2	77.2
96	_10_7	16,077	0.2	72.7	120	_19_9	11,938	0.2	77.3
97	_16_52	15,450	0.2	72.9	121	_4_3	11,891	0.2	77.5
98	_25_11	15,360	0.2	73.2	122	_17_11	11,674	0.2	77.7
99	_28_9	15,273	0.2	73.4	123	_15_37	11,219	0.2	77.8
100	_6_10	15,072	0.2	73.6	124	_15_24	11,140	0.2	78.0
101	_15_20	15,057	0.2	73.8	125	_22_6	11,090	0.2	78.1
102	_5_25	14,826	0.2	74.0	126	_26_55	10,883	0.2	78.3

Fuente: Elaboración propia con datos de (SSA, 2022).

La Tabla 85 presenta el análisis de cantidades de casos positivos por cada ola epidémica o periodo interepidémico (variable núm. 56: *OLA_O_PER_INTEREPI*), mostrando que la mayoría (83.1%) de los casos ocurrieron entre la segunda y la quinta olas. La primera ola solamente constituye el 11.2% y la cuarta ola es la que tiene más casos (24.5%).

Tabla 85. Número de ola epidémica o de periodo interepidémico: solo en casos positivos de COVID.

	Conteo	%	% acumulado
Cuarta ola	1,733,760	24.5	24.5
Segunda ola	1,545,999	21.9	46.4
Tercera ola	1,374,222	19.4	65.8
Quinta ola	1,223,355	17.3	83.1
Primera ola	795,031	11.2	94.4
Segundo periodo interepid.	150,240	2.1	96.5
Primer periodo interepid.	116,640	1.6	98.2
Tercer periodo interepid.	80,768	1.1	99.3
Cuarto periodo interepid.	49,528	0.7	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La Tabla 86 muestra el análisis de cantidad de casos positivos por mes (y año) de ingreso (variable ANIOAÑO_MES_INGRESO (núm. 51). La mayoría (81.7%) de estos ocurrieron en quince meses (no consecutivos). Los meses con más casos son: enero de 2022 (2022_1), con 19.1%, seguido por julio de ese mismo año (2022_7), con el 10.5% y agosto de 2021 (2021_8) con 7.4%. Estos tres meses constituyen más de la tercera parte (37.1%) del total.

Tabla 86. Año y mes de ingreso: solo en casos positivos de COVID.

No. Progr.		Conteo	%	% acumulado
1	2022_1	1,353,213	19.1	19.1
2	2022_7	743,875	10.5	29.7
3	2021_8	523,222	7.4	37.1
4	2021_1	424,394	6.0	43.1
5	2021_7	381,611	5.4	48.5
6	2020_12	329,996	4.7	53.1
7	2022_6	319,049	4.5	57.6
8	2022_2	290,058	4.1	61.7
9	2021_9	273,774	3.9	65.6
10	2020_7	212,274	3.0	68.6
11	2020_11	206,038	2.9	71.5
12	2021_2	195,234	2.8	74.3
13	2022_8	183,011	2.6	76.9
14	2020_10	172,147	2.4	79.3
15	2020_8	170,819	2.4	81.7
16	2020_6	158,886	2.2	84.0
17	2020_9	147,468	2.1	86.1
18	2021_3	143,254	2.0	88.1
19	2021_10	124,864	1.8	89.9
20	2021_12	120,519	1.7	91.6
21	2021_6	107,135	1.5	93.1
22	2021_4	93,647	1.3	94.4
23	2020_5	92,470	1.3	95.7
24	2021_11	82,496	1.2	96.9
25	2021_5	65,871	0.9	97.8
26	2022_3	48,608	0.7	98.5
27	2022_5	37,220	0.5	99.0
28	2020_4	29,650	0.4	99.5
29	2022_9	19,245	0.3	99.7
30	2022_4	16,365	0.2	100.0
31	2020_3	3,122	0.0	100.0
32	2020_2	8	0.0	100.0
	Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

La Tabla 87 presenta los conteos de casos por semana epidemiológica (variable núm. 37: SEMANA_EPIDEM), mostrando solamente el 80.3% más alto. Este porcen-

CAPITULO 3. PANORAMA ESTADÍSTICO DESCRIPTIVO DE CASOS
POSITIVOS DEL DATASET MASTER-3

taje está constituido por 60 semanas no consecutivas. De estas, las tres con mayores porcentajes, superando al 3.9% cada una son: 2022_02 (5.9%), 2022_03 (5.7%) y 2022_04 (3.9%).

Tabla 87. Semana epidemiológica: solo el 80% más alto de los casos positivos de COVID.

Ranking	Semana	Conteo	%	% acumulado	Ranking	Semana	Conteo	%	% acumulado	Ranking	Semana	Conteo	%	% acumulado
1	2022_02	414,689	5.9	5.9	27	2021_36	75,150	1.1	56.8	53	2020_44	42,195	0.6	76.4
2	2022_03	399,564	5.7	11.5	28	2020_50	73,954	1.0	57.8	54	2020_32	41,662	0.6	77.0
3	2022_04	272,552	3.9	15.4	29	2020_53	73,892	1.0	58.9	55	2020_45	41,267	0.6	77.5
4	2022_01	212,859	3.0	18.4	30	2020_49	65,178	0.9	59.8	56	2021_09	39,749	0.6	78.1
5	2022_28	210,243	3.0	21.4	31	2020_52	64,133	0.9	60.7	57	2020_26	39,349	0.6	78.7
6	2022_27	200,045	2.8	24.2	32	2021_05	60,131	0.9	61.5	58	2020_35	39,234	0.6	79.2
7	2022_29	174,422	2.5	26.7	33	2021_52	59,550	0.8	62.4	59	2020_42	39,091	0.6	79.8
8	2022_05	157,290	2.2	28.9	34	2022_07	58,873	0.8	63.2	60	2020_33	39,025	0.6	80.3
9	2022_26	154,144	2.2	31.1	35	2021_38	57,940	0.8	64.0					
10	2021_32	134,576	1.9	33.0	36	2021_37	57,450	0.8	64.8					
11	2021_31	128,986	1.8	34.8	37	2020_48	57,133	0.8	65.7					
12	2021_30	122,967	1.7	36.5	38	2021_27	56,963	0.8	66.5					
13	2021_33	119,026	1.7	38.2	39	2022_24	55,889	0.8	67.2					
14	2022_30	116,393	1.6	39.9	40	2021_06	51,848	0.7	68.0					
15	2021_02	110,187	1.6	41.4	41	2022_32	50,490	0.7	68.7					
16	2021_01	108,269	1.5	42.9	42	2020_29	50,301	0.7	69.4					
17	2021_29	104,820	1.5	44.4	43	2020_30	49,935	0.7	70.1					
18	2021_03	103,736	1.5	45.9	44	2020_46	48,563	0.7	70.8					
19	2021_34	102,409	1.4	47.3	45	2020_31	45,622	0.6	71.4					
20	2022_25	99,512	1.4	48.8	46	2020_47	45,128	0.6	72.1					
21	2021_04	88,132	1.2	50.0	47	2020_28	44,573	0.6	72.7					
22	2021_35	86,647	1.2	51.2	48	2020_43	44,450	0.6	73.3					
23	2022_06	82,334	1.2	52.4	49	2021_39	43,783	0.6	74.0					
24	2021_28	80,097	1.1	53.5	50	2020_27	43,534	0.6	74.6					
25	2020_51	77,147	1.1	54.6	51	2021_07	42,568	0.6	75.2					
26	2022_31	76,884	1.1	55.7	52	2021_08	42,280	0.6	75.8					

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 88 analiza si el paciente estuvo en contacto con otra persona que haya sido diagnosticada con SARS CoV-2 (variable núm. 29: OTRO_CASO). Casi dos terceras partes (63.7%) no lo estuvieron y una tercera parte (33.4%) sí.

Tabla 88. Estuvo en contacto con otro caso de COVID: solo en casos positivos de COVID.

	Conteo	%	% acumulado
No	4,501,441	63.7	63.7
Sí	2,359,700	33.4	97.1
No especificado	208,402	2.9	100.0
Total	7,069,543		

Fuente: Elaboración propia con datos de (SSA, 2022).

CAPÍTULO 4

PATRONES DE CORRELACIÓN MULTI-VARIABLE HALLADOS CON APRENDIZAJE AUTOMÁTICO EN CASOS INDIVIDUALES

Este amplio capítulo presenta los patrones más frecuentes, de tipo multivariable, que se observan en cuatro *datasets* por separado, todos provenientes del *dataset Master-3*, que contiene 18,328,375 de registros, incluyendo hombres y mujeres, de casos positivos, negativos, etc. A su vez, este *dataset* proviene de la base de datos de COVID-19 de la SSA (2022). Uno de los cuatro *datasets*, llamado *Posit*, contiene solo los casos positivos de COVID-19 (con base en el atributo *CLASIFICACIÓN_FINAL*, valores 1, 2 y 3), que son 7,069,543 registros, incluyendo hombres y mujeres. Un segundo *dataset*, llamado *Mujeres*, contiene solamente mujeres (9,873,929 registros), pero incluye casos positivos, negativos, etc. Un tercer *dataset*, llamado *Hombres*, contiene solamente hombres (8,454,446 registros), incluyendo casos positivos, negativos, etc. Finalmente, un cuarto *dataset*, llamado *PositMujeres* contiene solo casos positivos de mujeres (3,760,702 registros).

Tanto el *dataset* principal *Master-3* (ver el capítulo 2, del panorama estadístico descriptivo general del *dataset*), como los otros cuatro provenientes de este, pueden ser descargados gratuitamente de la Nube¹³. Los patrones se detectaron aplicando la técnica de aprendizaje automático denominada *árbol clasificador*, descrita en el capítulo de diseño metodológico. Para aplicarla, se usó específicamente el algoritmo llamado C5.0 (Quinlan, 1993), en su implementación para el lenguaje de programación R (R Core Team, 2019) de la biblioteca (*package*) llamada C50 (Kuhn and Quinlan, 2022).

Las razones para producir modelos de aprendizaje automático usando estos cuatro *datasets* derivados del *dataset* principal *Master-3* y no directamente de este último son las siguientes. La infraestructura de cómputo disponible para producir los modelos tiene memoria RAM de solamente 64 GB. Aunque *Master-3* tiene un tamaño

13 Disponible en <https://1drv.ms/f/s!At6TY8A0mTpNgtp5PU8kGlwSxnIriA?e=IH7rBI>

aproximado de 3.9 GB, al generar modelos de árboles clasificadores con el algoritmo C5.0, la memoria RAM disponible resulta insuficiente. Por otra parte, al separar los datos de Master-3 en varios subconjuntos con base en los valores de alguna variable determinada, se obtienen modelos que facilitan la observación de patrones de mayor interés desde los puntos de vista clínico y sociodemográfico. Para la etapa de entrenamiento (*train*) de estos modelos se usa el 75% de los casos de su respectivo dataset, y para la etapa de prueba (*test*), el 25% restante. Por ejemplo, los modelos producidos con el dataset *Posit*, que contiene más de 7 millones de casos, se producen usando solamente 5,302,157 renglones para entrenamiento, y el resto (1,767,386) para prueba. Por ello, en los resultados de cada uno de los modelos producidos, la cantidad de casos reportados es menor que la cantidad total de casos de su dataset, aunque los resultados son descriptivos de todo el conjunto.

La Tabla 89 presenta el panorama general de los 29 modelos de árboles clasificadores producidos. Todos los modelos presentados en esa tabla (y otros adicionales producidos con el dataset *Master-3*) están disponibles para ser descargados libre y gratuitamente de la Nube¹⁴. Quizá podría parecer que son demasiados; sin embargo, esta cantidad es justificada porque, al usar distintos atributos como *target*, se obtiene una visión más amplia y, al mismo tiempo, más detallada, de las interacciones estadísticas que pudieran existir entre el conjunto de todos los atributos de un dataset, lo cual permite una mayor comprensión del fenómeno analizado. Algunos de los modelos producidos tienen una gran cantidad de reglas (por ejemplo, el núm. 29 para el *target año_mes_ingreso* tiene casi 38,000). La lectura de modelos clasificadores que contienen tantas reglas resulta casi imposible; pero, generalmente, las reglas más “sustanciosas” son las que describen a las mayores cantidades de casos. Por ello, de varios modelos presentados a continuación, se presentan y se explican sus reglas más abarcativas. Estos patrones más “sustanciosos” son los que pueden constituir conocimiento novedoso para los profesionales de la salud. Estos también podrían propiciar el surgimiento de nuevas hipótesis o de preguntas de investigación para generar más conocimiento. Por ello, se invita al lector a descargar y explorar los modelos que resulten de su interés, buscando las ramas (reglas) que contengan más casos. De los modelos que hemos puesto en la Nube para ser descargados, los que contienen las reglas numeradas son aquellos que en su nombre incluyen la palabra *rules*.

La columna denominada *Atributo target* se refiere al atributo del dataset que, manualmente, fue designado como atributo de clase para producir un árbol y descubrir los patrones que lo involucran. Para la mayoría de los *targets*, se produjo más de un modelo; sus diversos modelos difieren en el dataset usado o en el subconjunto de atributos que fueron seleccionados manualmente como predictores. De algunos *targets* se produjeron modelos que no se mencionan en esta tabla porque presentan patrones triviales, que resultan obvios a los profesionales de la salud y que no contienen conocimiento innovador. Por ejemplo, del *target falleció*, su modelo *01_falleció_2023_01_29_c* (no incluido en la tabla) indica que el hecho de que un paciente haya fallecido o no, está muy correlacionado con el hecho de si fue intubado o no. Los profesionales de la salud ya saben que el fallecimiento fue más común en pacientes intubados que en no intubados; por ello, el conocimiento de la correlación entre estos dos atributos tiene poca novedad. La columna *Dataset* indica cuál

14 Disponible en <https://1drv.ms/f/s!At6TY8A0mTpNgtZKsnhAHInxcEITsA?e=7fenA7>

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 89. Resumen de los modelos de árboles clasificadores producidos con casos individuales.

No.	Atributo target	Nombre del modelo	Dataset	Cant. de predictores	Accuracy	Kappa	Cant. de reglas	Atributo en raíz
1	falleció	01_falleció_2023_01_29_d	Posit*	34	0.9670	0.6107	272	UCI
2	intubado	04_intubado_2023_01_31_b	Posit*	36	0.9886	0.9374	191	tipo_paciente
3	tipo_paciente	05_tipo_paciente_2023_02_01_c	Posit*	35	0.9697	0.8228	650	falleció
4	clasificación_final	06_clasifinal_HOMBRES_2023_02_12_c	Hombres**	34	0.7332	0.4885	7,239	obesidad
5	clasificación_final	06_clasifinal_MUJERES_2023_02_13_c_2	Mujeres**	35	0.7300	0.4645	8,717	días_ingreso_a_def
6	clasificación_final	06_clasifinal_MUJERES_2023_07_07	Mujeres**	33	0.7303	0.4653	8,602	falleció
7	entidad_res	33_entidad_res_2023_02_03_c	Posit*	33	0.3390	0.2299	13,004	otra_com
8	indígena	08_indígena_2023_02_05_b	Posit*	36	0.9575	0.3362	764	neumonía
9	habla_lengua_indig	09_habla_leng_ind_2023_02_05_b	Posit*	36	0.9587	0.3330	757	neumonía
10	Diabetes	10_diabetes_2023_02_07_b	Posit*	36	0.9137	0.1496	24	asma
11	Inmusupr	11_inmusupr_2023_02_20	Posit*	37	0.9943	0.5000	7	asma
12	Obesidad	26_obesidad_2023_07_06	Posit*	37	0.9035	0.0685	28	renal_crónica
13	renal_crónica	27_renal_crónica_2023_07_07	Posit*	37	0.9905	0.9905	19	cardiovascular
14	UCI	14_uci_2023_02_06	Posit*	37	0.9947	0.9710	146	intubado
15	UCI	14_uci_2023_02_06_b	Posit*	36	0.9930	0.9614	139	tipo_paciente
16	UCI	14_uci_2023_02_06_f	Posit*	32	0.9574	0.7514	1,078	neumonía
17	UCI	14_uci_2023_02_06_g	Posit*	31	0.9382	0.6243	1,713	renal_crónica
18	resultado_lab	31_resultado_lab_HOMBRES_2023_07_08_c	Hombres**	32	0.8070	0.5948	6,716	semana_epidem
19	resultado_lab	31_resultado_lab_MUJERES_2023_07_12	Mujeres**	32	0.8084	0.5746	7,675	semana_epidem
20	resultado_antígeno	33_resultado_antígeno_HOMBRES_2023_07_10_b	Hombres**	32	0.7731	0.6419	7,714	semana_epidem
21	resultado_antígeno	33_resultado_antígeno_MUJERES_2023_07_10_b	Mujeres**	32	0.7703	0.6331	8,272	semana_epidem
22	sexo	17_sexo_2023_02_06_b	Posit*	36	0.5701	0.1134	3,112	tabaquismo
23	embarazo	18_MUJ_EMBARAZO_2023_02_19	Mujeres**	37	0.9739	0.1440	344	obesidad
24	embarazo	18_MUJ_EMBARAZO_2023_02_19_b	Mujeres**	36	0.9740	0.1555	403	asma
25	embarazo	18_POSIT_MUJ_EMBARAZO_2023_02_19	PositMujeres*	36	0.9774	0.1469	85	asma
26	embarazo	18_POSIT_MUJ_EMBARAZO_2023_02_19_b	PositMujeres*	35	0.9774	0.1437	76	inmusupr
27	edad_intervalo	32_edad_intervalo_2023_07_06_b	Posit*	31	0.4879	0.1502	1,347	hipertensión
28	ola_o_per_interepi	34_ola_o_período_2023_02_05	Posit*	35	0.4979	0.3703	22,976	resultado_antígeno
29	año_mes_ingreso	51_año_mes_ingreso_2023_02_06	Posit*	47	0.2832	0.1898	37,959	resultado_antígeno

*Con base en clasificación_final 1, 2 o 3. **Incluye casos positivos, negativos, etc.

Fuente: Elaboración propia con datos de (SSA, 2022), usando el software R (R Core Team, 2019) con el package C50 (Kuhn and Quinlan, 2022).

fue el *dataset* usado para producir cada modelo. *Cant. de predictores* refiere a la cantidad de atributos del *dataset* que fueron seleccionados manualmente como predictores para descubrir interacciones con el *target* especificado. Para producir los modelos de algunos *targets* (por ejemplo clasificación_final, UCI, embarazo, etc.), se designaron manualmente diferentes subconjuntos de predictores. El propósito es observar posibles cambios en el predictor que es colocado automáticamente por el algoritmo C5.0 (Quinlan, 1993) como raíz del árbol. Por ejemplo, en el modelo núm. 14, con *target* UCI, al designar manualmente 37 atributos del *dataset* Posit como predictores, el predictor designado como raíz por el algoritmo es intubado. En cambio, en el modelo núm. 15, se quitó manualmente este atributo, quedando solamente 36 predictores, con lo que el algoritmo C5.0 descubrió que el predictor más poderoso, designado como raíz del árbol, es *tipo_paciente*. Algo similar, con otros atributos, ocurre en los modelos núm. 16 y 17.

La columna *Accuracy* (precisión) es una medida estadística fundamental para evaluar los modelos producidos. Está dada como el porcentaje de aciertos logrados por un modelo al ser probado con un subconjunto de datos del mismo dataset que le dio origen. Los valores posibles para *accuracy* son entre 0.0 y 1.0 (o entre 0% y 100%).

La columna *Kappa* es una medida similar a *accuracy*, pero resta el posible sesgo que podría estar causado aleatoriamente por la cantidad de distintos valores nominales del *target*, sus proporciones en la cantidad de casos del *dataset*, etc. Los valores posibles de *Kappa* son entre -1.0 y +1.0. La columna *Cantidad de reglas* indica cuántas ramas (o reglas clasificadoras) contiene el árbol. La columna *Atributo en raíz* indica cuál es el atributo que ocupa el nodo raíz del árbol y, por tanto, es el que tiene la mayor correlación estadística con el *target*.

Considerando que esta tabla contiene mucha información, una forma de facilitar su interpretación es observando, para el *target* de cada modelo, cuál es el atributo ubicado en la raíz de este. Por ejemplo: *intubado* está muy asociado con *tipo_paciente*; *indígena* está muy asociado con neumonía; diabetes, con asma; sexo, con tabaquismo, etc. Algunos de los pares *target/raíz* podrían resultar novedosos para los profesionales de la salud, sugiriéndoles nuevas preguntas de investigación para profundizar en el estudio del fenómeno.

Con base en la teoría del aprendizaje automático supervisado, los modelos que podrían considerarse más precisos son los que tienen los mayores valores de *accuracy* y *Kappa*; por ejemplo, el modelo núm. 2, para el *target intubado*, con *accuracy*=0.9886 y *Kappa*=0.9374. Sin embargo, si estos dos valores son muy cercanos a 1.0, es probable que los patrones detectados sean triviales; es decir, relativamente obvios ante los ojos de un experto del área de conocimiento correspondiente. Esto ocurre en este modelo, donde su raíz es el atributo *tipo_paciente*, sugiriendo que, el hecho de que un paciente sea intubado o no, está muy asociado a si es un paciente hospitalizado o es ambulatorio.

No existe un criterio único, estricto, para determinar si un modelo producido con algún algoritmo de aprendizaje automático supervisado (como un árbol clasificador) es confiable o no. El criterio a aplicar depende del propósito que se haya designado para el modelo; este propósito podría ser simplemente descriptivo y exploratorio, o bien, producir un *software* que realice clasificaciones o pronósticos automáticamente que requieran alta precisión, o bien, demostrar o refutar una hipótesis acerca de un proceso o fenómeno representado en el *dataset* analizado. En términos generales, un criterio ampliamente aceptado en la literatura especializada consiste en observar los valores de *accuracy* y *Kappa* del modelo y la cantidad de clases mayoritarias del atributo *target* de su *dataset*. Los primeros dos parámetros fueron comentados anteriormente; por su parte, la cantidad de clases mayoritarias se refiere a la cantidad de valores nominales distintos del atributo *target* que corresponden a, aproximadamente, el 80% de los casos del *dataset*. Para conocer la cantidad de clases mayoritarias del atributo *target* en un *dataset*, se necesita hacer un análisis de Pareto, como los que se muestran en los Capítulos 2 y 3 de este libro.

El capítulo 2 muestra los análisis de Pareto correspondientes a varios atributos de los más de aproximadamente 18.4 millones de casos del *dataset Master-3* (incluyendo positivos, negativos, etc.). El capítulo 3 presenta los análisis de Pareto solamente de los aproximadamente 7 millones de casos que son positivos. En general, un modelo aceptable con bajos niveles de exigencia tendría: *accuracy* ≥ 0.7 , *Kappa* ≥ 0.67 , y cantidad de clases mayoritarias ≥ 2 . Por ello, se recomienda al lector observar en

los capítulos 2 y 3 las cantidades de clases mayoritarias correspondientes al atributo *target* de los modelos de la tabla 89 que resulten de su interés. Por ejemplo, el *target falleció*, en el *dataset Master-3* (incluyendo casos positivos, negativos, etc.) tiene solamente una clase mayoritaria (la clase “No”, con 97.7% de los casos), y en el *dataset* que contiene solamente casos positivos, tiene también solo una clase mayoritaria (también la clase “No”, con 95.3%). Dado lo anterior, el modelo núm. 1, del *target falleció*, con *accuracy* = 0.9670 y *Kappa* = 0.6107, no cumple estrictamente los tres criterios mencionados. Sin embargo, a pesar de ello, el modelo es útil para propósitos descriptivos y explicativos porque muestra la cantidad de casos correspondientes a cada una de diversas combinaciones de valores de los atributos predictores asociados a un determinado valor del *target*.

Por otra parte, otra forma menos usada, pero potencialmente útil, de evaluar un modelo de aprendizaje automático supervisado es comparar su nivel de *accuracy* con el que podría lograr una persona experta en esa área de conocimiento. Por ejemplo: ¿qué nivel de *accuracy* podría obtener un médico experto en COVID-19 para “adivinar” si cada una de las personas de un conjunto de casos del *dataset* analizado fallecieron o no, basándose en los datos predictores disponibles? ¿Su *accuracy* sería menor, igual, o mayor que la respuesta proporcionada por el modelo clasificador producido automáticamente?

Un recurso muy valioso que se encuentra en los árboles clasificadores producidos en esta investigación son las reglas clasificadoras que los constituyen (ver el capítulo de diseño metodológico). Por ello, en la tabla 89, cada modelo tiene un hiperenlace que permite descargar de la web un archivo PDF que contiene al modelo completo, producido en lenguaje de programación R, junto con nuestro código fuente para producir el modelo, la evaluación estadística que se le realiza automáticamente y una lista del porcentaje de casos que son descritos por cada atributo predictor que forma parte del modelo (*attribute usage*). Esta lista es útil para saber cuáles son los predictores que están más asociados estadísticamente al *target*, sugiriendo (sin asegurar) una tentativa y potencial relación causal entre esos predictores y el *target*. Por todo lo anterior, cada uno de los modelos producidos podría dar origen a nuevas hipótesis o preguntas para generar nuevas investigaciones, con enfoque cuantitativo, cualitativo, o mixto.

A continuación, se presentan los patrones de interacción multivariable más destacables correspondientes a cada uno de los atributos *target* enumerados en la tabla 89. Para facilitar al lector la asimilación de tantos modelos producidos, estos se presentan a continuación organizados en cuatro secciones: 1) características del paciente, 2) historia clínica y comorbilidades, 3) curso clínico y resultados, y 4) contexto epidemiológico. Cada una de las reglas clasificatorias que se presentan puede observarse detalladamente en nuestros modelos de aprendizaje automático puestos en la Nube para su descarga gratuita.

4.1 Características del paciente

Estas características son diez: 1) estado de mortalidad (falleció o no), 2) estado de intubación (fue intubado o no), 3) tipo de paciente (ambulatorio u hospitalizado), 4) clasificación final del caso (confirmado de COVID-19, negativo, etc.), 5) entidad fede-

rativa de residencia, 6) estatus indígena (el paciente se autoadscribe como tal o no), 7) lengua indígena (habla alguna lengua indígena o no), 8) sexo, 9) estado de embarazo (en paciente mujer), 10) intervalo de edad.

4.1.1 Estado de mortalidad: falleció o no (atributo falleció)

En relación con el *target* *falleció*, el modelo núm. 1 (y otros más que no se presentan en la tabla), proveniente del *dataset* *Posit*, contiene los perfiles (patrones o reglas clasificadoras) de las personas que sí fallecieron y de los que no fallecieron. En otras palabras, este modelo identifica las variables predictores (y sus correspondientes valores) que tienen la mayor influencia en el hecho de que el paciente haya fallecido o no. En este *target*, la clase mayoritaria es solamente una: la clase *no falleció* (95.3%). Los cinco predictores que tienen mayor importancia en sus reglas clasificadoras con base en el porcentaje de casos que incluyen son: UCI (100.0%), *neumonía* (99.2%), *edad* (10.1%), *embarazo* (10.0%) y *sector* (9.3%).

Las tres reglas descriptivas multivariable que contienen más individuos son:

- i. *Si la persona no está embarazada o esta situación no es aplicable; entonces no falleció* (soporte = 5,235,596 individuos, confianza = 95.2%).
- ii. *Si la entidad federativa de residencia del individuo es alguna de: Guerrero, Hidalgo, Estado de México, Michoacán, Baja California, Sinaloa, Sonora, Tabasco, Zacatecas, Coahuila, Colima, Ciudad de México, y la presencia de asma está confirmada o descartada, y no aplica la unidad de cuidados intensivos, entonces el individuo no falleció* (soporte = 2,606,365 individuos, confianza = 99.8%).
- iii. *Si el individuo no tiene neumonía, y el resultado de laboratorio es no adecuado o no aplica por ser caso sin muestra, y se hizo toma de muestra de antígeno, y el municipio de residencia es alguno de los muchos mencionados por la regla (ver modelo 1 online, de tabla 89), entonces no falleció* (soporte = 2,237,309 individuos, confianza = 99.5%).

En las tres reglas presentadas (que contienen a las mayores cantidades de individuos en el modelo Núm. 1), las tres corresponden a casos en que el individuo no falleció. Por otra parte, las tres reglas que contienen a los mayores números de casos en que el individuo sí falleció son:

- i. *Si la edad del individuo es mayor que 52 años, y no tiene embarazo o esta situación no aplica, y la situación de uso de UCI es "sí", o "no" o "no especificado"; entonces falleció* (soporte = 332,967 individuos, confianza = 55.2%).
- ii. *Si el sector de servicios de salud que atendió al individuo es el IMSS, y su entidad federativa de residencia es cualquiera excepto Nayarit, Nuevo León, Puebla, Querétaro y Quintana Roo, y el resultado de la prueba de antígeno es positivo a SARS-CoV-2 o no aplica por ser caso sin muestra, y el uso de la unidad de cuidados intensivos es "sí", "no", o "no especificado", y su intervalo de edad es de 65 años o más, y la ola epidémica o periodo interepidémico es la primera, la*

segunda o la tercera ola o el segundo periodo interepidémico; entonces falleció (soporte = 69,724 individuos, confianza = 69.7%).

- iii. *Si el sector de servicios de salud que atendió al individuo es el IMSS, y su entidad federativa de residencia es cualquiera excepto Nayarit, Nuevo León, Puebla, Querétaro y Quintana Roo, y el uso de la unidad de cuidados intensivos es "sí", "no", o "no especificado", y el intervalo de edad del individuo es de 65 años o más, y la ola epidémica o periodo interepidémico es la primera, la segunda o el segundo periodo; entonces el individuo sí falleció (soporte = 62,402 individuos, confianza = 69.4%).*

4.1.2 Estado de intubación (atributo intubado)

El modelo núm. 2 aborda el atributo *target intubado*. Los patrones descubiertos son un poco obvios (triviales), aunque esto permite validar (convalidar) nuestros resultados con la experiencia práctica de los profesionales de la salud. Principalmente, esto se debe a que este atributo solamente tiene una clase mayoritaria: "no aplica" (valor 97), abarcando al 90% de los individuos del *dataset*. Sus cinco predictores más importantes y sus respectivos porcentajes de casos incluidos son: *tipo_paciente* (100.0%), *falleció* (10.0%), *neumonía* (7.0%), *semana_epidem* (7.0%), y *entidad_res* (5.9%)

Las tres reglas de este modelo que contienen más individuos son las siguientes:

- i. *Si el paciente es de tipo ambulatorio; entonces no fue intubado (soporte = 4,770,559 individuos, confianza = 100.0%)*

Nota: este es un patrón obvio, trivial, muy conocido por los profesionales de la salud.

- ii. *Si el individuo es de tipo hospitalizado, y no falleció, y su entidad federativa de residencia es cualquiera excepto Sinaloa y Sonora, y tiene neumonía, o bien, esto último no está especificado en sus registros; entonces, no fue intubado (soporte = 148,450 individuos, confianza = 94.5%).*

- iii. *Si el paciente es de tipo hospitalizado, y sí falleció, y fue registrado durante alguna de las (muchas) semanas epidemiológicas especificadas por la regla (ver el modelo núm. 2 online, de tabla 89), y el sector de servicios que lo atendió fue el IMSS o el IMSS-Bienestar, y la entidad federativa de ubicación del establecimiento de salud es cualquiera excepto Baja California, Campeche, Colima, Chiapas, Oaxaca, Querétaro, Quintana Roo, Veracruz y Yucatán; entonces no fue intubado (soporte = 86,666 individuos, confianza = 92.8%).*

Las tres reglas presentadas corresponden a pacientes que no fueron intubados. Por otra parte, las dos reglas que contienen más individuos que sí fueron intubados son:

- i. *Si el individuo es de tipo hospitalizado, y falleció, y el caso es registrado durante alguna de las muchas semanas epidemiológicas especificadas por la regla (ver modelo núm. 2 online, de tabla 89), y no se hizo toma de muestra de antígeno, y la cantidad de días transcurridos desde la fecha de ingreso hasta la defunción es mayor de 5, y la entidad federativa de ubicación del establecimiento de salud*

es alguna de: Aguascalientes, Durango, Guerrero, Jalisco, Nayarit, Oaxaca, Querétaro, Quintana Roo, Baja California Sur, Zacatecas, Coahuila, Colima, Chiapas, Chihuahua, o Ciudad de México, y el sector de servicios de salud que le atendió es alguno de: Semar, SSA, estatal, IMSS, ISSSTE, municipal, Pemex, o privada, y se hizo toma de muestra de laboratorio; entonces fue intubado (soporte = 9,478 individuos, confianza = 61.2%).

- ii. Si el paciente es de tipo hospitalizado, y falleció, y el caso se registró en alguna de las muchas semanas epidemiológicas especificadas por la regla (ver modelo núm. 2 online, de tabla 89), y tuvo neumonía, o bien, esto último no está especificado en los registros, y la entidad federativa de ubicación del establecimiento de salud es alguna de: Durango, Oaxaca, Quintana Roo, Tabasco, Baja California Sur, Colima, Chihuahua, o Ciudad de México, y la cantidad de días entre el ingreso y la defunción es mayor que 10, y el sector de servicios de salud es alguno de: Sedena, SSA, estatal o ISSSTE, y la entidad y municipio de residencia es alguno de los muchos especificados por la regla (ver modelo núm. 2 online, de tabla 89); entonces fue intubado (soporte = 3,583 individuos, confianza = 63.8%).

4.1.3 Tipo de paciente: ambulatorio u hospitalizado (atributo tipo_paciente)

El target *tipo_paciente*, en el modelo núm. 3, especifica si el paciente fue: ambulatorio (representado con el valor _1), hospitalizado (valor 2), o bien, no especificado (valor 99). En el dataset *Posit*, su clase mayoritaria es solo una: _1 (89.9%); es decir, la gran mayoría de los casos fueron ambulatorios. Sus predictores más importantes, con base en el porcentaje de casos que incluyen, son: *falleció* (100.0%), *neumonía* (95.7%), *sector* (92.9%), *resultado_lab* (92.7%) y *renal_crónica* (91.9%). En este modelo, las tres reglas que contienen a más individuos corresponden a pacientes de tipo ambulatorio (valor _1) y son:

- i. Si el individuo no falleció; entonces es de tipo ambulatorio (soporte = 5,054,668 individuos, confianza = 94.1%).
- ii. Si el sector de servicios de salud que atendió a la persona es el IMSS o una institución privada, y no tiene neumonía, y su edad es menor o igual que 60 años, y está determinado si es indígena o no, y no tiene diabetes, y no tiene insuficiencia renal crónica, o bien, se ignora esto último, y no falleció, y la cantidad de días transcurridos entre el inicio de síntomas y el ingreso es mayor que 0 y menor o igual que 3; entonces el individuo es de tipo ambulatorio (soporte = 1,319,376 individuos, confianza = 99.0%).
- iii. Si el origen del caso es fuera de una USMER, y el sector de servicios de salud que atendió a la persona es alguno de: Semar, SSA, universitario, estatal, IMSS, IMSS-Bienestar, o Pemex, y no tiene diabetes, y no tiene EPOC, y tuvo contacto con otro caso diagnosticado con SARS CoV-2, y no falleció, y su edad es entre 5 y 44 años, y la cantidad de días transcurridos entre el inicio de síntomas y el ingreso al establecimiento es menor o igual a 5; entonces es de tipo ambulatorio (soporte = 681,424 individuos, confianza = 99.2%).

Por otra parte, las tres reglas con más individuos que fueron pacientes de tipo hospitalizado son:

- i. Si el individuo tuvo neumonía; entonces fue de tipo hospitalizado (soporte = 388,058 individuos, confianza = 84.5%).
- ii. Si el paciente falleció; entonces fue de tipo hospitalizado (soporte = 247,488 individuos, confianza = 95.7%).
- iii. Si la entidad federativa de residencia de la persona fue cualquiera excepto Tabasco, y tuvo neumonía, y está determinado si tenía asma o no, y el resultado de laboratorio es alguno de: positivo, no positivo, pendiente o no adecuado, y la entidad federativa y municipio de residencia es alguno de los muchos especificados por la regla (ver modelo núm. 3 online, de tabla 89), y el intervalo de edad es entre 0 y 4 años, o bien, mayor o igual a 45 años; entonces es de tipo hospitalizado (soporte = 196,938 individuos, confianza = 93.9%).

4.1.4 Clasificación final del caso: COVID-19 confirmado, negativo, inválido, etc. (atributo clasificación_final)

Este es uno de los atributos más importantes de la base de datos de la SSA 2022 porque indica si el paciente es, o no, un caso de COVID-19 según el catálogo llamado *CLASIFICACIÓN_FINAL*. Puede tener alguno de siete valores de tipo categórico, los cuales reproducimos a continuación, debido a su importancia para nuestra investigación:

- 1: Caso de COVID-19 confirmado por asociación clínica epidemiológica
- 2: Caso de COVID-19 confirmado por comité de dictaminación
- 3: Caso de SARS-CoV-2 confirmado, que tiene muestra de laboratorio o prueba antigénica y resultó positiva a SARS-CoV-2, sin importar si el caso tiene asociación clínica epidemiológica.
- 4: Inválido por laboratorio
- 5: No realizado por laboratorio
- 6: Caso sospechoso
- 7: Negativo a SARS-CoV-2

La tabla 90 presenta un análisis de Pareto de los valores de *clasificación_final* en el *dataset Hombres*, donde existen dos clases mayoritarias: 7 y 3, acumulando el 93.4% de los individuos. Este *dataset* fue usado para crear el modelo núm. 4. Los cinco predictores más importantes con base en el porcentaje de individuos incluidos son: *obesidad* (100.0%), *falleció* (99.2%), *toma_muestra_antígeno* (99.1%), *sector* (97.7%) y *semana_epidem* (96.5%).

Tabla 90. Análisis de Pareto del atributo target clasificación_final en el dataset *Hombres*.

Clasificación final (según SSA, 2022)	Conteo	%	% acum.
Negativo a SARS-CoV-2 (clase 7)	4,766,194	56.4%	56.4%
SARS-CoV-2 confirmado, que tiene muestra de laboratorio o prueba antigénica y resultó positiva a SARS-CoV-2, sin importar si el caso tiene asociación clínica epidemiológica (clase 3)	3,133,956	37.1%	93.4%
Caso sospechoso (clase 6)	328,357	3.9%	97.3%
COVID-19 confirmado por asociación clínica epidemiológica (clase 1)	164,754	1.9%	99.3%
No realizado por laboratorio (clase 5)	45,415	0.5%	99.8%
COVID-19 confirmado por comité de dictaminación (clase 2)	10,131	0.1%	99.9%
Inválido por laboratorio (clase 4)	5,639	0.1%	100.0%
Total	8,454,446		

Fuente: Elaboración propia, con información de (SSA, 2022).

En este modelo, las cinco reglas clasificadoras que contienen a más individuos de casos positivos, que corresponden solamente a las clases 1 y 3 (sin haber reglas de la clase 2 en este conjunto) son:

- i. Si la entidad federativa donde se ubica el establecimiento que atendió al individuo es cualquiera excepto: Baja California Sur, Campeche, Guerrero, Morelos, Nayarit, San Luis Potosí, Tabasco, y Tlaxcala; entonces la clasificación final es COVID-19 confirmado de clase 3 (soporte = 5,446,607 individuos, confianza = 37.0%).*
- ii. Si la entidad federativa donde se ubica el establecimiento que atendió al individuo es alguna de: Aguascalientes, Tabasco, Tlaxcala, Baja California Sur, Yucatán, Campeche, Chiapas, o Ciudad de México; entonces la clasificación final es COVID-19 confirmado de clase 1 (soporte = 3,005,421 individuos, confianza = 2.7%).*
- iii. Si la entidad federativa donde se ubica el establecimiento que atendió al individuo es la Ciudad de México; entonces la clasificación final es COVID-19 confirmado de clase 1 (soporte = 2,325,068 individuos, confianza = 3.4%).*
- iv. Si el origen del caso es fuera de una USMER, y el sector de servicios de salud es SEMARSemar, IMSS o PEMEXPemex, y está determinado si tenía obesidad o no, y se hizo toma de muestra de antígeno, y la semana epidemiológica del caso individual es alguna entre la 2022_01 y la 2022_06, y entre 2022_24 y 2022_33, y el individuo no falleció; entonces la clasificación final es COVID-19 confirmado de clase 3 (soporte = 477,862 individuos, confianza = 85.5%).*
- v. Si el individuo es ambulatorio, y su edad es mayor de 13 años, y no habla lengua indígena, y su autoadscripción como indígena es negativa o no especificada, y no tiene asma, y puede tener, o no, obesidad, y no tiene enfermedad renal crónica, y no tuvo contacto con otro caso diagnosticado con SARS CoV-2, y no se hizo toma de muestra de laboratorio, pero sí se hizo toma de muestra de antígeno.*

no, y su semana epidemiológica fue entre 2022_02 y 2022_04, o entre 2022_26 y 2022_31, y no falleció, y la cantidad de días transcurridos entre el inicio de síntomas y el ingreso es menor o igual a 4; entonces la clasificación final es COVID-19 confirmado de clase 3 (soporte = 354,522, confianza = 77.4%).

Por su parte, la Tabla 91 presenta el análisis de Pareto del dataset Mujeres, donde las clases mayoritarias también son dos y estas coinciden con las del dataset *Hombres* en que son la 7 y la 3. Aquí, el porcentaje acumulado es de 94.1%.

Tabla 91. Análisis de Pareto del atributo target clasificación_final en el dataset Mujeres.

Clasificación final (según SSA, 2022)	Conteo	%	% acum.
Negativo a SARS-CoV-2 (clase 7)	5,727,191	58.0%	58.0%
SARS-CoV-2 confirmado, que tiene muestra de laboratorio o prueba antigénica y resultó positiva a SARS-CoV-2, sin importar si el caso tiene asociación clínica epidemiológica (clase 3)	3,560,381	36.1%	94.1%
Caso sospechoso (clase 6)	331,775	3.4%	97.4%
COVID-19 confirmado por asociación clínica epidemiológica (clase 1)	194,540	2.0%	99.4%
No realizado por laboratorio (clase 5)	48,137	0.5%	99.9%
Inválido por laboratorio (clase 4)	6,124	0.1%	99.9%
COVID-19 confirmado por comité de dictaminación (clase 2)	5,781	0.1%	100.0%
Total	9,873,929		

Fuente: Elaboración propia, con información de (SSA, 2022).

Con el *dataset Mujeres*, se produjeron los modelos núm. 5 y 6. Enfocándonos en el modelo 5 (muy similar al modelo 6 en sus medidas de *Aaccuracy* y *Kappa*), sus cinco atributos predictores con mayores porcentajes de individuos incluidos son: *días_ingreso_a_def* (100.0%), *toma_muestra_antígeno* (98.7%), *sector* (98.5%), *toma_muestra_lab* (98.3%), y *semana_epidem* (96.5%). Las cinco reglas clasificadoras en el modelo 5 que abarcan a las mayores cantidades de individuos son:

- Si el individuo tiene confirmado si tiene o no otras enfermedades además de COVID; entonces la clasificación final es COVID-19 confirmado de clase 3 (soporte = 7,297,560 individuos, confianza = 36.2%).*
- Si el sector de servicios de salud es alguno de: SSA, universitario, estatal, IMSS-Bienestar, ISSSTE, municipal, privado, o no especificado; entonces la clasificación final es COVID-19 confirmado de clase 1 (soporte = 4,699,730 individuos, confianza = 2.6%).*
- Si la entidad federativa de ubicación del establecimiento de salud es Morelos, San Luis Potosí, Sinaloa, Sonora, Tabasco, Tamaulipas, Baja California Sur, Coahuila, o Ciudad de México; entonces la clasificación final es COVID-19 confirmado de clase 1 (soporte = 4,095,216 individuos, confianza = 2.6%).*
- Si la ola epidémica o el periodo interepidémico del caso es alguno de: el primero, segundo, tercero o cuarto periodos; entonces la clasificación final es COVID-19 confirmado de clase 1 (soporte = 959,172 individuos, confianza = 1.9%).*

- v. Si en el individuo no era aplicable la intubación, y su edad era mayor de 33 años, y no tenía EPOC, y no se hizo toma de muestra de laboratorio, y se hizo toma de muestra de antígeno, y la semana epidemiológica fue entre 2022_02 y 2022_04, o bien, entre 2022_25 y 2022_32, y no falleció, y la cantidad de días transcurridos entre el inicio de síntomas y el ingreso al establecimiento de salud fue entre 1 y 5; entonces la clasificación final es COVID-19 confirmado de clase 3 (soporte = 509,607 individuos, confianza = 72.4%).

4.1.5 Entidad federativa de residencia (atributo entidad_res)

El modelo núm. 7 tiene como *target* a la entidad federativa de residencia del individuo. El *dataset* que le dio origen (Posit) presenta 17 clases mayoritarias, que agrupan al 81.1% de los individuos. De esas clases, las cinco primeras son: Ciudad de México (24.5%), Estado. de México (9.9%), Nuevo León (5.7%), Guanajuato (5.0%) y Jalisco (4.0%). Es notorio que la Ciudad de México por sí sola constituye casi la cuarta parte de todos los casos de este *dataset*.

En el modelo mencionado, los cinco atributos con mayores porcentajes de uso son: *otra_com* (100.0%), *sector* (99.6%), *origen* (96.4%), *resultado_antígeno* (93.8%), e *indígena* (88.3%). *Otra_com* indica si el individuo tuvo o no alguna otra enfermedad además de COVID al ser diagnosticado. *Sector* indica el sector de servicios de salud al que pertenece el establecimiento de salud que atendió al individuo. *Origen* indica si el individuo fue atendido o no por alguna de las unidades de salud monitoras de enfermedades respiratorias (USMER). *Resultado_antígeno* indica si se realizó o no la prueba de antígeno. *Indígena* indica si el individuo se autoadscribe o no como indígena. Los valores de *accuracy* = 0.3390 y *Kappa* = 0.2299 son considerablemente bajos, lo que sugiere que no existen patrones sólidos que caractericen a las distintas entidades federativas de residencia, por lo que no se percibe una necesidad de profundizar más en este modelo.

4.1.6 Estatus indígena (atributo indígena)

El modelo núm. 8 tiene como atributo *target* a la variable indígena. Esta indica si la persona se autoadscribe, o no, como tal. Este atributo difiere del atributo *habla_lengua_indig* en el sentido de que una persona que se autoadscribe como indígena, no necesariamente habla una lengua indígena. En algunos casos, ocurre que estas dos características están asociadas estadísticamente y tienen una relación causal. En el *dataset* Posit existe solamente una clase mayoritaria en este atributo, que es la 2, que significa *no* (95.0%); es decir, solamente el 5.0% de los individuos se autoadscriben como indígenas.

En el modelo núm. 8, los cinco atributos predictores con mayores porcentajes de uso son: *neumonía* (100.0%), *sector* (99.6%), *ent_mpio_res* (99.6%), *embarazo* (51.4%), y *entidad_um* (43.1%). El atributo *neumonía* funciona en forma interesante en este modelo, ya que el hecho de que se tenga confirmado si un individuo tiene o no *neumonía* es altamente influyente para determinar si es indígena o no. Esto sugiere que, en un porcentaje importante de individuos indígenas, no está especificado si el individuo tiene o no *neumonía*. El predictor *sector* se refiere al sector de servicios de salud al

que pertenece el establecimiento de salud que atendió a la persona. *Ent_mpio_res* es la entidad federativa y municipio de residencia del individuo. El significado del atributo embarazo es evidente y, finalmente *entidad_um* es la entidad federativa de ubicación del establecimiento de salud que atendió a la persona. Los valores de *accuracy* = 0.9575 y *Kappa* = 0.3362 del modelo 8 indican que es preciso en su descripción, aunque principalmente por casualidad, ya que la gran mayoría (95.0%) de los individuos del *dataset* son no-indígenas. Una de las reglas clasificadoras que contiene a más individuos es esta:

Si en el individuo está confirmada la presencia o ausencia de neumonía, y el sector de servicios de salud que le atendió es el IMSS, y su entidad federativa y municipio de residencia es alguno de los muchos especificados por esta regla (ver modelo núm. 8 online, de la Tabla 89); entonces no es indígena (soporte = 2,489,570 individuos, confianza = 99.7%).

4.1.7 Lengua indígena (atributo *habla_lengua_indig*)

En el modelo núm. 9, el target *habla_lengua_indig* indica si el individuo habla, o no, alguna lengua indígena. Es distinto al target denominado indígena en el sentido de que este último refiere a la adscripción que la persona hace, o no, de sí misma. Es decir, *habla_lengua_indig* es independiente de que la persona se autoadscriba o no como indígena, aunque en muchos casos estos dos hechos están asociados.

La clase mayoritaria en este target es: *_2* (92.7%), indicando que la gran mayoría de los individuos no hablan una lengua indígena. Los cinco predictores más útiles en el *dataset Posit* para determinar si una persona habla o no una lengua indígena son: *neumonía* (100.0%), *sector* (99.6%), *ent_mpio_res* (99.6%), *embarazo* (51.5%) y *entidad_nac* (44.5%). La presencia de los predictores *ent_mpio_res* (entidad federativa y municipio de residencia) y *entidad_nac* (entidad federativa de nacimiento) parecen esperables para determinar si una persona en México habla, o no, una lengua indígena. Sin embargo, los predictores *neumonía*, *sector* (sector de servicios de salud) y *embarazo* quizá podrían sugerir interacciones estadísticas más interesantes; por ejemplo, las siguientes. El comportamiento del predictor *neumonía* es similar al observado en el modelo núm. 8 para el target indígena. Respecto a los diversos sectores de servicios de salud, quizá alguno de estos atendió a más hablantes de lengua indígena que otros sectores. En relación con el predictor *embarazo*, esto podría sugerir que el estado embarazado es más frecuente en individuos que hablan lengua indígena.

Una de las reglas clasificadoras que abarcan a más casos en este modelo es esta:

Si en el individuo está confirmada la presencia o ausencia de neumonía, y el sector de servicios de salud que le atendió es el IMSS, y su entidad federativa y municipio de residencia es alguno de los muchos especificados por esta regla (ver modelo núm. 8 online, de la Tabla 89); entonces no habla lengua indígena (soporte = 2,493,513 individuos, confianza = 99.8%).

Nota: esta regla es muy similar a la presentada en la sección anterior, correspondiente al target *indigenaindígena*. Esta gran similitud sugiere que este target y el

de *habla_lengua_indig* están muy correlacionados estadísticamente. Ambas reglas tienen un soporte mayor a 2.4 millones de individuos, con una confianza de casi 100.0%. Por ello, los patrones representados por ambas reglas son muy sólidos.

4.1.8 Sexo (atributo sexo)

En el *dataset Posit*, el porcentaje de mujeres (valor *_1*) es 53.2% y el de hombres (valor *_2*) es 46.8%. En el modelo núm. 22, el análisis del *target* sexo tiene el propósito de identificar si existen patrones diferentes entre los hombres y las mujeres en los casos positivos (*dataset Posit*). Los resultados muestran que los predictores más poderosos (con base en el porcentaje de casos que incluyen) para distinguir entre hombres y mujeres son: *tabaquismo* (100.0%), *falleció* (100.0%), *asma* (99.6%), *sector* (91.5%), *otra_com* (88.5%) y *edad* (87.5%).

Las tres reglas que contienen a más individuos en este modelo son:

- i. Si el individuo no padece otra enfermedad además de COVID-19; entonces su sexo es femenino (soporte = 5,180,568 individuos, confianza = 53.0%).
- ii. Si la entidad federativa del establecimiento de salud que atendió al individuo es cualquiera excepto Querétaro; entonces su sexo es masculino (soporte = 5,169,248 individuos, confianza = 46.8%).
- iii. Si el sector de servicios de salud que atendió a la persona es alguno de: estatal, IMSS, IMSS-Bienestar, ISSSTE, privada, SSA, universitario, o no especificado, y su entidad federativa de nacimiento es cualquiera excepto Yucatán, y no se aplica la intubación, y la edad es mayor de 15 años, y la nacionalidad es mexicana, y no tiene tabaquismo, y estuvo en contacto con otro caso de COVID-19 o bien esto último no está especificado, y no falleció, y la ola epidémica o el periodo interepidémico es la cuarta o la quinta olas o el primero, segundo, tercero o cuarto periodos, entonces el sexo es femenino (soporte = 585,830 individuos, confianza = 61.0%).

Nota: esta regla es relativamente compleja, involucrando a nueve atributos predictores.

4.1.9 Estado de embarazo (atributo embarazo)

El atributo *target* embarazo representa si la persona se encuentra, o no, en esta situación. De este *target* se produjeron cuatro modelos, del Nonúm. 23 al 26. Los modelos 23 y 24 se produjeron con el *dataset* llamado *Mujeres*, que contiene casos solamente de mujeres, tanto positivos, como negativos, etc. Por su parte, los modelos 25 y 26 se produjeron con el *dataset PositMujeres*, que también contiene casos de mujeres, pero solamente positivos.

En el *dataset Mujeres*, solo existe una abrumadora clase mayoritaria del *target* embarazo: *_2* (la persona no está embarazada), con 97.3% de individuos. Los casos en que sí hay embarazo (valor *_1*) son solamente el 2.0%. En particular, en su modelo núm. 24 los cinco atributos predictores más poderosos son: *asma* (100.0%), *edad_intervalo* (99.7%), *sector* (96.3%), *intubado* (59.4%) y *ent_mpio_res* (59.1%). La regla clasificadora que contiene más casos de embarazo es:

Si está confirmada la presencia o ausencia de asma en el individuo, y la edad es entre 15 y 44 años, y el estado de intubación es "sí", "no" o "no especificado", y la entidad federativa de ubicación del establecimiento de atención es alguna de: Colima, Chiapas, Guanajuato, Michoacán, Morelos, Nayarit, Nuevo León, Quintana Roo, San Luis Potosí, Sinaloa, Sonora, Tabasco, Tamaulipas, Veracruz o Yucatán, y la edad es menor o igual que 33 años, y no falleció, y el sector de servicios de salud es alguno de: estatal, IMSS, IMSS-Bienestar, Pemex, privada, SSA, o universitario, y no hay toma de muestra de laboratorio, y hay toma de muestra de antígeno, y la entidad federativa y municipio de residencia es alguno de los muchos especificados por esta regla (ver modelo núm. 24 online, de la Tabla 89), entonces sí hay embarazo (soporte = 1,589 individuos, confianza = 64.3%).

Nota: esta regla es relativamente compleja porque contiene diez atributos predictores.

Igualmente, en el *dataset PositMujeres*, la única clase mayoritaria del *target* es *_2* (mujer no embarazada), con 97.6%. Los casos en que sí hay embarazo (*valor_1*) son solamente 1.6%. En el modelo núm. 25, producido con este *dataset*, los cinco predictores más poderosos son: *inmusupr* (100.0%), *edad_intervalo* (99.7%), *sector* (96.1%), *intubado* (58.6%) y *ent_mpio_res* (54.3%). En las tres reglas clasificadoras que contienen más casos coincide que no hay embarazo y son las siguientes:

- i. Si está determinado si la persona tiene o no inmunosupresión; entonces no hay embarazo (soporte = 2,813,123 individuos, confianza = 63.7%).*
- iii. Si está determinado si la persona fue intubada o no; entonces no hay embarazo (soporte = 222,432 individuos, confianza = 96.8%).*

4.1.10 Intervalo de edad (atributo *edad_intervalo*)

El atributo *edad_intervalo* es el intervalo en el que se ubica la edad del individuo. No es de tipo numérico, sino categórico. Algunos de sus valores son, por ejemplo: *_0*, *_1_4*, *_5_9*, etc., representando los intervalos de 0 años, entre 1 y 4, entre 5 y 9, etc. Los intervalos definidos en estos análisis y modelos se basan en los definidos por la Secretaría de Salud (SSA) de México para producir sus análisis estadísticos.

El modelo núm. 27 de la tTabla 89, producido para este *target*, proviene del *dataset Posit*. Se presenta un análisis de Pareto de este *target* en el Cap. capítulo 3 (tTabla 61), donde se identifican cinco clases mayoritarias, que agrupan al 87.0% de individuos. Estas son: *_25_44* (45.2%), *_50_59* (14.3%), *_45_49* (9.6%), *_20_24* (9.2%) y *_65ymas* (8.7%). En este modelo, los predictores más poderosos para determinar el valor de *edad_intervalo* son: *hipertensión* (100.0%), *EPOC* (99.9%), *intubado* (99.6%), *diabetes* (96.0%) y *embarazo* (21.0%).

Las tres reglas clasificadoras que contienen más casos tienen niveles de confianza muy bajos y son las siguientes:

- i. Si la persona no tiene diabetes; entonces su intervalo de edad es de 15 a 19 años (soporte = 4,819,696 individuos, confianza = 4.1%).*

- ii. Si la persona no tiene diabetes; entonces su edad es de 20 a 24 años (soporte = 4,819,696 individuos, confianza = 10.0%).
- iii. Si la persona no tiene diabetes; entonces su edad es de 25 a 44 años (soporte = 4,819,696 individuos, confianza = 48.0%).

4.2 Historia clínica y comorbilidades

Los modelos acerca de la historia clínica y las comorbilidades del paciente se enfocan en cuatro aspectos: 1) diabetes, 2) inmunosupresión, 3) obesidad, y 4) enfermedad renal crónica.

4.2.1 Diabetes (atributo diabetes)

En el modelo númNo. 10, el atributo target es *diabetes*. Su clase mayoritaria en el *dataset Posit* es "no" (valor _2), conteniendo al 90.9% de los individuos. En este modelo, los cinco predictores con mayores porcentajes de uso son: *asma* (100.0%), *renal_crónica* (99.7%), *falleció* (98.7%), *edad* (5.2%) y *edad_intervalo* (4.4%).

Las tres reglas que contienen más individuos en este modelo son de casos en que no existe diabetes. Estas son:

- i. Si está determinado si el individuo tiene asma o no, y no tiene enfermedad renal crónica; entonces no tiene diabetes (soporte = 5,235,688 individuos, confianza = 91.5%).
- ii. Si el paciente es de tipo ambulatorio, y está determinado si tiene asma o no; entonces no tiene diabetes (soporte = 4,757,148 individuos, confianza = 93.6%).
- iii. Si la edad es menor o igual a 42 años, y la presencia de EPOC es negativa o se ignora, y está determinado si tiene asma o no; entonces no tiene diabetes (soporte = 3,120,586 individuos, confianza = 97.9%).

Por otra parte, las tres reglas con mayor soporte en las que sí hay presencia de diabetes son:

- i. Si la persona tiene enfermedad renal crónica; entonces tiene diabetes (soporte = 50,421 individuos, confianza = 52.2%).
- ii. Si el individuo tiene edad mayor de 55 y menor o igual a 84, y está determinado si tiene asma o no, y tiene enfermedad renal crónica; entonces tiene diabetes (soporte = 25,223 individuos, confianza = 69.4%).
- iii. Si la persona tiene edad mayor de 43 y menor o igual a 82 años, y está determinado si tiene asma o no, y tiene enfermedad cardiovascular, y falleció; entonces tenía diabetes (soporte = 9,673 individuos, confianza = 55.3%).

4.2.2 Inmunosupresión (atributo inmusupr)

En el modelo núm 11, el *target inmusupr* indica si el individuo tiene, o no, inmunosupresión, en los casos que fueron positivos a COVID-19 (*dataset Posit*). Su clase mayoritaria es solo una: *_2* (98.9%), la cual representa que, prácticamente, todos los individuos no tenían inmunosupresión. Sus tres predictores más importantes y los porcentajes de casos que incluyen son: *asma* (100.0%), *tabaquismo* (99.7%) y *renal_crónica* (99.6%). Por otra parte, en porcentajes extremadamente bajos, se observan otros dos predictores: *cardiovascular* (0.29%) y *otra_com* (0.03%).

Este modelo es pequeño porque solamente contiene siete reglas clasificadoras. De estas, las tres que incluyen a más casos son:

- i. *Si está determinado si el individuo tiene asma o no, y está determinado si existe tabaquismo o no, y no tiene enfermedad renal crónica o bien se desconoce esto último; entonces no tiene inmunosupresión* (soporte = 5,235,288 individuos, confianza = 99.4%).
- ii. *Si está determinado si la persona si tiene asma o no, y está determinado si tiene tabaquismo o no, y tiene enfermedad renal crónica, y no tiene asma; entonces no tiene inmunosupresión* (soporte = 48,890 individuos, confianza = 93.3%).
- iii. *Si se desconoce si el individuo tiene asma o no, y tiene tiene enfermedad cardiovascular o bien se desconoce si la tiene; entonces, se desconoce si tiene o no inmunosupresión* (soporte = 14,696 individuos, confianza = 99.2%).

Por otra parte, solamente hay una regla que describe a los casos que sí tienen inmunosupresión:

Si está determinado si el individuo tiene asma o no, y está determinado si tiene tabaquismo o no, y tiene enfermedad renal crónica, y tiene asma, y tiene otra enfermedad además de COVID-19 o bien se desconoce si la tiene o no; entonces, tiene inmunosupresión (soporte = 476 individuos, confianza = 84.0%).

4.2.3 Obesidad (atributo obesidad)

El *target obesidad* (ver modelo númNo. 12, proveniente del *dataset Posit*, en la Tabla 89) indica si el individuo tiene obesidad o no. Su única clase mayoritaria es *_2* (91.2%), correspondiente al valor *no*. El predictor en la raíz del árbol es *renal_crónica* (insuficiencia), lo cual sugiere que, en los casos positivos de COVID-19, la obesidad y la insuficiencia renal están asociados estadísticamente.

Los tres predictores más importantes en el modelo, con base en el porcentaje de casos que incluyen, son: *renal_crónica* (100.0%), *sector* (99.7%) y *asma* (99.1%). Por otra parte, con muy pequeños porcentajes, hay otros dos predictores: *cardiovascular* (2.2%), y *entidad_um* (0.8%).

Las tres reglas clasificadoras que abarcan a más individuos en este modelo son:

- i. Si en el individuo está determinado si tiene o no enfermedad cardiovascular; entonces no tiene obesidad (soporte = 5,286,684 individuos, confianza = 90.3%).
- ii. Si el sector de servicios de salud que atendió al individuo es cualquiera excepto PEMEX Pemex o "no especificado", y no tiene asma, y está determinado si tiene o no insuficiencia renal crónica; entonces no tiene obesidad (soporte = 5,157,264 individuos, confianza = 90.7%).
- iii. Si el sector de servicios de salud que atendió al individuo es cualquiera excepto PEMEX Pemex o "no especificado", y no tiene hipertensión o bien se desconoce si la tiene, y no tiene inmunosupresión, y está determinado si tiene o no insuficiencia renal crónica; entonces no tiene obesidad (soporte = 4,612,407.2 individuos, confianza = 92.4%).

Por otra parte, las tres reglas en las que sí existe obesidad y contienen a más individuos son las siguientes:

- i. Si el sector de servicios de salud que atendió a la persona es PEMEX Pemex, y la entidad federativa de ubicación del establecimiento de salud es alguna de: Jalisco, Nuevo León, Oaxaca, Tabasco, Veracruz, o Campeche, y está determinado si tiene o no insuficiencia renal crónica, y la edad es de 20 años o más; entonces, tiene obesidad (soporte = 15,913 individuos, confianza = 52.4%).
- ii. Si el sector de servicios de salud que atendió al individuo es Pemex, y la entidad federativa de ubicación del establecimiento de salud es Oaxaca, Tabasco o Veracruz, y la semana epidemiológica del caso es alguna de estas 44: 2020_16, 2020_17, 2020_20, 2020_21, 2020_22, 2020_23, 2020_24, 2020_25, 2020_26, 2020_27, 2020_28, 2020_29, 2020_30, 2020_31, 2020_32, 2020_33, 2020_34, 2020_35, 2020_40, 2020_41, 2020_42, 2020_43, 2020_44, 2020_46, 2020_48, 2020_53, 2021_03, 2021_04, 2021_06, 2021_12, 2021_14, 2021_15, 2021_17, 2021_18, 2021_20, 2021_25, 2021_26, 2021_27, 2021_44, 2021_49, 2022_10, 2022_13, 2022_16, 2022_17, y la edad es de 20 años o más; entonces tiene obesidad (soporte = 3,756 personas, confianza = 66.1%).
- iii. Si el sector de servicios de salud que atendió al individuo es Pemex, y la entidad federativa de ubicación del establecimiento de salud es Nuevo León o Campeche, y la edad es de 20 años o más; entonces tiene obesidad (soporte = 3,459 individuos, confianza = 68.0%).

4.2.4 Enfermedad renal crónica (atributo renal_crónica)

Como su nombre lo indica, el *target* llamado *renal_crónica*, en el modelo núm. 13, representa si un individuo tiene, o no, insuficiencia renal crónica. Tiene solamente una clase mayoritaria: *_2* (98.6%), correspondiente a individuos que no tienen este padecimiento. Al respecto, en el modelo núm. 13, proveniente del *dataset Posit*, los cinco principales predictores (pudiendo cada uno de estos tener valor de *sí* o de *no*) y los porcentajes de casos que contienen son: *cardiovascular* (100.0%), *inmusupr* (99.9%), *asma* (99.7%), *intubado* (99.1%) e *hipertensionhipertensión* (10.2%).

Las reglas descriptivas que engloban a más individuos en este modelo corresponden a la clase “no” (valor _2) y son las siguientes:

- i. *Si está determinado si el individuo tiene o no una enfermedad cardiovascular, y está determinado si tiene o no asma, y no tiene inmunosupresión, y no se aplicó intubación; entonces no tiene enfermedad renal crónica (soporte = 4,738,002 individuos, confianza = 99.4%).*
- ii. *Si está determinado si el individuo tiene o no una enfermedad cardiovascular, y está determinado si tiene o no asma, y no tiene inmunosupresión, y está determinado si al individuo se le aplicó o no intubación o bien esta información no está especificada en registros, y no tiene hipertensión o bien se desconoce si la tiene; entonces no tiene enfermedad renal crónica (soporte = 328,388 individuos, confianza = 97.9%).*
- iii. *Si está determinado si el individuo tiene o no una enfermedad cardiovascular, y está determinado si tiene o no asma, y no tiene inmunosupresión, y está determinado si se le aplicó o no la intubación o bien esta información no está especificada en registros, y tiene hipertensión y su edad es mayor de 34 años; entonces no tiene enfermedad renal crónica (soporte = 185,072 individuos, confianza = 89.2%).*

Por otra parte, la regla descriptiva con la mayor cantidad de casos en que sí hubo insuficiencia renal crónica es:

Si está determinado si el individuo tiene o no una enfermedad cardiovascular, y está determinado si tiene o no asma, y no tiene inmunosupresión, y está determinado si se le aplicó o no la intubación o bien esta información no está especificada en registros, y tiene hipertensión, y su edad es menor o igual que 34, y no tiene obesidad, y no hay embarazo o esto último no es aplicable o se desconoce, y la entidad federativa de nacimiento es alguna de estas 15: Aguascalientes, Colima, Durango, Guanajuato, Hidalgo, Jalisco, México, Michoacán, Morelos, Nayarit, Puebla, Nuevo León, San Luis Potosí, Tlaxcala, Zacatecas, y no tiene asma, y no tiene diabetes o bien esto último se desconoce; entonces tiene enfermedad renal crónica (soporte = 1,315 individuos, confianza = 62.6%).

4.3 Curso clínico y resultados

Son tres los modelos acerca del curso clínico y los resultados del paciente: 1) ingreso en la unidad de cuidados intensivos (UCI), 2) resultados de laboratorio, y 3) resultados de la prueba de antígeno.

4.3.1 Ingreso en UCI (atributo UCI)

El *target UCI* representa si un individuo estuvo, o no, en la Unidad de Cuidados Intensivos de un establecimiento de salud. En los *datasets Master-3* y *Posit*, casi todos (más del 99% de) los individuos no estuvieron en la UCI (ver capítulos 2 y 3). En la Tabla 89, los modelos núm. 14 a 17 (provenientes del *dataset Posit*) muestran que al-

gunos de los predictores más asociados estadísticamente al *target UCI* y que aparecen como raíz en sus respectivos árboles, son: *intubado*, *tipo_paciente*, *neumonía* y *renal_crónica*. Estas asociaciones estadísticas hacia el *target* podrían considerarse poco novedosas; sin embargo, las descripciones cuantitativas detalladas en las reglas clasificadoras de estos modelos podrían tener alguna utilidad teórica o práctica para los expertos.

Específicamente, en el modelo mod. 17, los cinco predictores más importantes y sus respectivos porcentajes de inclusión de casos son: *renal crónica* (100.0%), *resultado_lab* (100.0%), *edad_intervalo* (99.0%), *origen* (94.4%) y *sector* (93.3%). Este modelo de árbol tiene *accuracy* = 93.8% y *Kappa* = 0.6243, que son relativamente satisfactorios. Tiene 1,713 reglas clasificadoras, por lo que puede considerarse un modelo muy especializado, ya que en cada una de sus hojas contiene cantidades muy pequeñas de individuos; es decir, es un árbol muy disperso. Una de las hojas que contienen a más individuos tiene solamente 936, pero hay que restarle 194 excepciones. Por ello, no tiene mucha utilidad profundizar en estas reglas.

4.3.2 Resultados de laboratorio (atributo *resultado_lab*)

El atributo *target* denominado *resultado_lab* indica el resultado del análisis de la muestra reportado por el laboratorio de la Red Nacional de Laboratorios de Vigilancia Epidemiológica (InNDRE, LESP y LAVE) y laboratorios privados avalados por el INNDRE cuyos resultados son registrados en Sisver. Sus valores en la base de datos de (SSA, 2022) son:

- 1: Positivo a SARS-CoV-2
- 2: No positivo a SARS-CoV-2
- 3: Resultado pendiente
- 4: Resultado no adecuado
- 97: No aplica (caso sin muestra)

Este *target* es analizado en los modelos núm. 18 (usando el *dataset Hombres*) y 19 (usando el *dataset Mujeres*) de la Tabla 89. Se analizan por separado los casos de hombres y los de mujeres por dos razones principales: primeramente, porque la infraestructura de cómputo disponible para esta investigación tiene una memoria RAM de solamente 64 GB, lo que impide producir modelos con el algoritmo C5.0 usando el *dataset* original completo, que contiene más de 18.3 millones de registros. En segundo término, porque al analizar por separado a hombres y mujeres, se podrían descubrir patrones que son más comunes en alguno de los dos sexos.

En el *dataset Hombres*, las clases mayoritarias de *resultado_lab* son dos: *_97* (no aplica, por ser un caso sin muestra), con 66.3% y *_2* (no positivo a SARS-CoV-2), con 17.9%. En el *dataset* de Mujeres, también son estas dos clases, y con porcentajes similares a los mencionados: 68.0% y 18.1%, respectivamente. Del modelo 18 (de hombres), las tres reglas que contienen a más individuos son:

- i. Si el individuo no tiene inmunosupresión; entonces el resultado de laboratorio es positivo a SARS-CoV-2 (soporte = 6,256,374 individuos, confianza = 14.0%).

- ii. Si la semana epidemiológica del caso individual es alguna de las 57 especificadas por la regla (ver modelo núm. 18 online, de la tTabla 89); entonces, se trata de un caso sin muestra, por lo cual no hay resultado de laboratorio (soporte = 5,423,487, confianza = 76.3%).
- iii. Si la entidad federativa de ubicación del establecimiento de salud que atendió al individuo es alguna de: Aguascalientes, Baja California, Campeche, Colima, Chihuahua, Guanajuato, Guerrero, Hidalgo, Jalisco, México, Morelos, Oaxaca, Querétaro, Quintana Roo, Tlaxcala, Veracruz, o Zacatecas; entonces el resultado de laboratorio es positivo a SARS-CoV-2 (soporte = 2,059,914 individuos, confianza = 17.3%).

Por su parte, las tres reglas del modelo 18 que contienen más casos con resultado de laboratorio positivo tienen valores de soporte considerablemente altos, pero confianza muy baja y son estas:

- i. Si el individuo no tiene inmunosupresión; entonces el resultado de laboratorio es positivo a SARS-CoV-2 (soporte = 6,256,374 individuos, confianza=14.0%).
Nota: esta es la misma regla del inciso i anterior.
- ii. Si la entidad federativa de ubicación del establecimiento de salud que atendió al individuo es alguna de: Aguascalientes, Baja California, Campeche, Colima, Chihuahua, Guanajuato, Guerrero, Hidalgo, Jalisco, México, Morelos, Oaxaca, Querétaro, Quintana Roo, Tlaxcala, Veracruz, o Zacatecas; entonces el resultado de laboratorio es positivo a SARS-CoV-2 (soporte = 2,059,914 individuos, confianza = 17.3%).
Nota: esta es la misma regla del inciso iii anterior.
- iii. Si la entidad federativa y el municipio de residencia del individuo son alguna de los muchos especificados por esta regla (ver modelo núm. 18 online, de la tabla 89); entonces el resultado de laboratorio es positivo a SARS-CoV-2 (soporte = 1,632,709 individuos, confianza = 19.3%).

En el modelo núm. 19 (creado con el *dataset Mujeres*), los cinco predictores más importantes son: *semana_epidem* (100.0%), *sector* (99.1%), *ent_mpio_res* (98.0%), *intubado* (88.0%) y *entidad_res* (85.4%). *Ent_mpio_res* es la entidad federativa y municipio de residencia del individuo. *Entidad_res* es la entidad federativa de residencia, pero sin tomar en cuenta el municipio. Este modelo tiene 7,675 reglas y en la mayoría de sus hojas tiene muy pocos casos (menos de 300). Por ello, no hay patrones sólidos que explorar.

4.3.3 Resultados de la prueba de antígeno (atributo resultado_antígeno)

El target resultado_antígeno es el resultado del análisis de la muestra de antígeno tomada al paciente. En el dataset Hombres, sus tres valores y, al mismo tiempo, clases mayoritarias son:

- Positivo a SARS-CoV-2 (clase 1): 23.8%.
- Negativo a SARS-CoV-2 (clase 2): 43.7%
- No aplica por ser un caso sin muestra (clase 97): 32.5%

Por su parte, en el *dataset Mujeres*, las clases mayoritarias del *target* tienen un comportamiento similar al descrito: *_2* (45.4%), *_97* (30.1%), y *_1* (24.5%). El modelo núm. 20 tiene como *target* al *resultado_antígeno* y fue creado con el *dataset Hombres*. Por su parte, el modelo 21 fue creado con el *dataset Mujeres* para el mismo *target*. En el modelo 20, los valores de *accuracy* y *Kappa* son 0.7731 y 0.6419, respectivamente, que resultan relativamente satisfactorios. Sin embargo, existen 7,714 reglas clasificadoras, cada una conteniendo muy pocos individuos. Por ello, no hay reglas que contengan suficientes casos para ser consideradas interesantes. Algo similar ocurre en el modelo 21 (de mujeres), con *accuracy* = 0.7703, *Kappa* = 0.6331 y 8,272 reglas; por ello, tampoco hay reglas destacables para comentar en este.

4.4 Contexto epidemiológico

Los modelos acerca del contexto epidemiológico abordan dos aspectos: la ola epidémica o el periodo interepidémico en que ocurrió el caso del paciente, y el año y mes de ingreso al establecimiento de salud donde se atendió.

4.4.1 Ola epidémica o periodo interepidémico (atributo *ola_o_per_interepi*)

El atributo *target ola_o_per_interepi* en el modelo núm. 28 se refiere a la ola epidémica o al periodo interepidémico en que ocurrió el caso COVID-19 de un individuo. Con base en (SSA, 2022b), en esta investigación se toman en cuenta cinco olas epidémicas y cuatro periodos interepidémicos, como se muestra en la Tabla 92.

Tabla 92. Olas epidémicas y periodos interepidémicos con sus semanas y fechas de inicio y fin.

Ola o periodo	Inicio		Fin	
	Semana epidemiológica	Fecha	Semana epidemiológica	Fecha
Primera ola	08 de 2020	16/Feb/2020	39 de 2020	26/Sep/2020
Segunda ola	40 de 2020	27/Sep/2020	15 de 2021	17/Abr/2021
Primer periodo interepid.	16 de 2021	18/Abr/2021	22 de 2021	05/Jun/2021
Tercera ola	23 de 2021	06/Jun/2021	42 de 2021	23/Oct/2021
Segundo periodo interepid.	43 de 2021	24/Oct/2021	50 de 2021	18/Dic/2021
Cuarta ola	51 de 2021	19/Dic/2021	09 de 2022	05/Mar/2022
Tercer periodo interepid.	10 de 2022	06/Mar/2022	21 de 2022	28/May/2022
Quinta ola	22 de 2022	29/May/2022	33 de 2022	20/Ago/2022
Cuarto periodo interepid.	34 de 2022	21/Ago/2022	40 de 2022	08/Oct/2022

Fuente: Elaboración propia con datos de (SSA, 2022b).

El *target ola_o_per_interepi* fue tomado del *dataset Posit*. Con base en un análisis de Pareto (ver Cap. 3, Tabla 88), las clases mayoritarias de este *target* son: cuarta ola (24.5%), segunda ola (21.9%), tercera ola (19.4%) y quinta ola (17.3%), acumulando un 83.1%. El resto de los casos corresponden a la primera ola y a los cuatro periodos interepidémicos.

Los cinco atributos predictores más poderosos en el modelo 28 son: *resultado_antígeno* (100.0%), *sector* (97.9%), *ent_mpio_res* (95.9%), *entidad_um* (92.1%) y *días_síntomas_a_ingreso* (85.6%). Con sus valores relativamente bajos de *accuracy* = 0.4979 y *Kappa* = 0.3703 y su gran cantidad de 22,976 reglas clasificadoras, cada una de estas agrupa a pocos individuos. Por ello, no se observan patrones sustanciosos que abarquen a subconjuntos de población de tamaño suficientemente interesante para profundizar en su análisis.

4.4.2 Mes y año de ingreso (atributo *año_mes_ingreso*)

El atributo *target año_mes_ingreso* es el identificador del mes y año (en representación alfanumérica) en que el individuo ingresó al establecimiento de salud donde se le dio la atención, pudiendo ser paciente hospitalizado o ambulatorio. Ejemplos de valores del *target* son: 2020_8, 2021_10, 2022_1, etc. En el modelo núm. 29, con el dataset *Posit*, las clases mayoritarias (81.7%) de este *target* son 15. De estas, las tres con mayores porcentajes son: 2022_1 (19.1%), 2022_7 (10.5%) y 2021_8 (7.4%), y las tres con menores porcentajes: 2022_8 (2.6%), 2020_10 (2.4%) y 2020_8 (2.4%).

En el modelo mencionado, los cinco predictores más importantes asociados al *año_mes_ingreso*, con los respectivos porcentajes de casos que incluyen son: *resultado_antígeno* (100.0%), *sector* (97.2%), *ent_mpio_res* (96.0%), *entidad_um* (89.0%) y *días_síntomas_a_ingreso* (84.3%). Recordar que: *ent_mpio_res* es el municipio y entidad federativa de residencia del individuo, *entidad_um* es la entidad federativa de ubicación del establecimiento de salud donde se atendió y *días_síntomas_a_ingreso* es la cantidad de días transcurridos entre la fecha de inicio de síntomas y la fecha de ingreso. Este modelo es considerablemente grande, ya que contiene 37,959 reglas clasificadoras, contando con muy pocos individuos en cada una. Además, sus valores de *accuracy* = 0.2832 y *Kappa* = 0.1898 son considerablemente bajos; por ello, no es tan útil intentar explorar las reglas clasificadoras que contienen más casos.

4.5 Discusión de los resultados

Los árboles clasificadores presentados en este capítulo han descrito algunos de los patrones de correlación estadística de tipo multi-variable que podrían considerarse más contundentes, con base en el número de individuos involucrados, que se encontraron en diversos datasets provenientes de la base de datos de COVID-19 de la SSA (2022).

Los hallazgos obtenidos podrían dividirse en dos partes: primera, la identificación de los atributos predictores más poderosos en relación con diversos atributos *target* y, segunda, la obtención de un conjunto de reglas descriptivas o clasificadoras que representan las interacciones estadísticas entre diversos subconjuntos de predictores y esos *targets*.

Algunos de los atributos *target* de mayor interés en los datasets analizados fueron: *falleció*, *intubado*, *tipo_paciente*, *uci*, *clasificación_final*, *resultado_lab* y *resultado_antígeno*. De cada uno de estos, se discuten los hallazgos más destacables a continuación. Del *target falleció*, sus tres predictores más importantes son *UCI*, *neumonía* y *edad*. De *intubado*, son *tipo_paciente*, *falleció* y *neumonía*. De *tipo_paciente*: *falle-*

ció, neumonía y sector. De uci: intubado, tipo_paciente, neumonía y renal_crónica. Del target clasificación_final, sus predictores más importantes en el dataset Hombres son: obesidad, falleció y toma_muestra_antígeno; en el dataset Mujeres: días_ingreso_a_def, toma_muestra_antígeno y sector. De resultado_lab: tanto del dataset Hombres como del de Mujeres, semana_epidem, sector y ent_mpio_res. Igualmente, de resultado_antígeno, tanto en Hombres, como en Mujeres, semana_epidem, sector y ent_mpio_res. Como puede verse, varios de estos targets y predictores están intercalados en los diversos modelos. Esto sugiere la existencia de correlaciones estadísticas entre ellos y, potencialmente, algunas posibles y tentativas relaciones causales, aunque esto último tendría que verificarse por otras técnicas.

Por otra parte, las reglas clasificadoras más destacables en los modelos producidos son aquellas que contienen a más individuos y que además tienen un valor de confianza razonablemente alto (convencionalmente, mayor al 60%). Los atributos target presentados en la Tabla 89 que tienen modelos con algunas reglas que cumplen estas condiciones son, principalmente, estos trece: *falleció, intubado, tipo_paciente, clasificación_final, indígena, habla_lengua_indig, diabetes, inmusupr, obesidad, renal_crónica, resultado_lab, sexo, y embarazo*. Entre esas reglas, algunas representan patrones obvios, triviales, que ya son conocidos por los profesionales de la salud, por ejemplo: en los casos positivos, *Si la persona es mujer no-embarazada o es un individuo masculino, entonces no falleció* (más de 5.2 millones de casos, 95.2% de confianza) (ver sección 4.1 *Falleció*, regla *i* de las que tienen más individuos). Sin embargo, aún esas reglas que representan patrones obvios pueden ofrecer información útil a los expertos porque proporcionan una cuantificación del comportamiento de esos fenómenos y, además, pueden convalidar los resultados de otros autores.

En los casos positivos, otras reglas destacables por el número de individuos que abordan, aunque constituyen patrones obvios son, por ejemplo, las cinco siguientes:

Si el paciente es ambulatorio, entonces, no es aplicable la intubación (más de 4.7 millones de casos, confianza de 100.0%) (sección 4.1.2 *Estado de intubación*, regla *i* de las que tienen más individuos).

Si el individuo no falleció, entonces, es paciente ambulatorio (más de 5 millones de casos, confianza de 94.1%) (sección 4.1.3 *Tipo_paciente*, regla *i* de las que tienen más individuos).

Si el individuo proviene de fuera de una USMER, el sector de servicios que le atendió es Semar, IMSS o PEMEX Pemex, está determinado si tenía obesidad o no, se hizo toma de muestra de antígeno, la semana epidemiológica es entre la 2022_01 y la 2022_06, y entre 2022_24 y 2022_33, y el individuo no falleció, entonces su clasificación final es 3 (confirmado a SARS-CoOV-2 mediante prueba de laboratorio o de antígeno) (más de 477,000 casos, confianza de 85.5%) (sección 4.1.4 *Clasificación_final*, regla *iv* de las que tienen más individuos).

Está determinado si el individuo tenía neumonía o no, el sector de servicios de salud que le atendió es el IMSS, y la entidad federativa y municipio de residencia es alguno de los especificados por la regla. Entonces, el individuo no es indíge-

na (más de 2.4 millones de casos, confianza del 99.7%,) (sección 4.1.6 *Estatus indígena*).

Está determinado si el individuo tiene asma o no, y no tiene enfermedad renal crónica, entonces no tiene diabetes (más de 5.2 millones de individuos, confianza de 91.5%) (sección 4.2.1 *Diabetes*, regla i de las que tienen más individuos).

Las reglas que describen patrones que no son tan obvios ni tan triviales contienen, generalmente, pocos casos y, eventualmente, incluyen un cierto número de excepciones. No obstante, pueden ser de interés para los profesionales de la salud porque pueden inspirar la elaboración de nuevas hipótesis o preguntas de investigación específicas, que se respondan con diversos enfoques y técnicas.

Esta investigación se puede convalidar mediante comparación con, al menos, cuatro estudios similares que se comentan a continuación:

Lorenzo *et al.* (2020) encontraron, mediante el algoritmo J4.8, que: “los patrones hallados por las diferentes vías empleadas coinciden en la relación fuerte entre varios rasgos de los pacientes en casos de letalidad, entre los cuales se destacan la edad, la obesidad, la hipertensión [arterial], la inmunosupresión, la diabetes, y los problemas renales y cardíacos” (p. 2).

Por otra parte, Pineda, Cruz, & Páez (2023) examinaron 68,750 registros de datos relacionados con defunciones en la Ciudad de México (CDMX) durante el año 2020, utilizando algoritmos como J4.8, reglas de decisión con el *software WEKA* de código abierto, siguiendo la metodología de minería de datos llamada KDD (*Knowledge Discovery in Databases*). Se enfocó específicamente en las comorbilidades relacionadas con enfermedades renales, cáncer, hipertensión y diabetes. Los resultados indicaron que las comorbilidades más significativamente asociadas con un mayor número de defunciones atribuidas a COVID-19 fueron la diabetes y el cáncer, particularmente entre hombres con edad mayor a 64 años.

De igual manera, el estudio de Méndez Sánchez (2021) se basó en un conjunto de 30,265 registros de pacientes. Utilizó el algoritmo J4.8 con modificaciones, y creó un árbol que clasifica a los pacientes según su tipo: ambulatorios, o bien, hospitalizados. Para los ambulatorios, se consideraron factores como la edad y la presencia de neumonía para determinar su probabilidad de recuperación o fallecimiento. En el caso de pacientes hospitalizados, se analizaron según si pertenecían a unidades médicas de enfermedades respiratorias (USMER) o no, y nuevamente se consideró la neumonía y la edad para evaluar su probabilidad de recuperación o fallecimiento.

De forma análoga, Luna-Ramírez *et al.* (2022) analizaron la evolución de la pandemia de COVID-19 en el estado de Baja California. Utilizaron datos oficiales del gobierno federal de México. Para llevar a cabo esta investigación, aplicaron la metodología KDD y la herramienta WEKA para crear modelos y descubrir patrones de comportamiento. Como resultado, detectaron relaciones significativas entre los contagios y las muertes por COVID-19 en la región. Identificaron meses de transición clave entre las diferentes olas de la pandemia, que fueron noviembre de 2020, mayo de 2021,

diciembre de 2021, mayo de 2022 y octubre de 2022. Finalmente, se sugiere a los profesionales de la salud descargar nuestros modelos y explorar los hallazgos aquí presentados para contrastarlos con los obtenidos por otros autores y para generar investigaciones innovadoras.

CAPÍTULO 5.

PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Este muy largo capítulo aborda el análisis exhaustivo de la información de la pandemia de COVID-19 en México con una agregación de datos a nivel de municipios. Es un capítulo completamente cuantitativo y descriptivo. Se presentan los hallazgos obtenidos mediante el uso de técnicas de estadística descriptiva; específicamente, simples conteos y porcentajes. El capítulo presenta, en su primera sección, una descripción de nuestro *dataset MuniCovid-2022*, la fuente de datos usada para los análisis producidos; se descarga de la Nube¹⁵ y se puede usar, citando este libro. Este dataset fue generado a partir de la base de datos de la SSA (2022) y de las proyecciones de población municipal de (Conapo, 2019). En la segunda sección del capítulo, se presentan análisis que consisten en conteos (frecuencias absolutas) de municipios; finalmente, la tercera sección presenta análisis basados en tasas municipales por cada 100,000 habitantes. Como un sub-producto complementario de investigación, se ofrecen al lector una serie de rankings de municipios con base en los diversos conteos generados (los hemos puesto en la Nube¹⁶ para ser descargados y usados, citando este libro). El capítulo tiene el propósito de dar al lector un panorama detallado de prácticamente todas las variables disponibles en la base de datos de la SSA (2022) al ser agregadas al nivel municipal. La especificidad de estos análisis estadísticos es necesaria para obtener un conocimiento detallado del fenómeno estudiado y para dar fundamento y soporte a los modelos de aprendizaje automático que se presentan en capítulos subsecuentes.

5.1 El dataset MuniCovid-2022

Nuestro dataset *MuniCovid-2022* (donde “Muni” hace referencia al nivel de agregación por municipios) contiene dos tipos de información estadística: primeramente,

15 Disponible en <https://1drv.ms/f/s!At6TY8A0mTpNgtZJpVP9VfhgKGGsfA?e=AZ56eP>

16 Disponible en https://1drv.ms/f/s!At6TY8A0mTpNheUytZNZtA6iaX6d_w?e=ohvXXN

conteos de casos positivos (atributo CLASIFICACIÓN_FINAL con valores de 1, 2 ó 3) y defunciones por COVID-19 en los municipios de México, desde el inicio de la pandemia en este país (primeros meses de 2020) hasta el 17 de septiembre de 2022. Adicionalmente, el *dataset* contiene proyecciones estadísticas de población municipal por segmentos de edades para los años 2020, 2021 y 2022. También, contiene tasas municipales de positivos y de defunciones de esos años. De los datos de COVID-19, la fuente es la base de datos COVID-19, de (SSA, 2022) y, de los datos poblacionales, la fuente son las Proyecciones de la Población de los Municipios de México 2015-2030, producidas por el Consejo Nacional de Población (Conapo, 2019). Para el año 2020, no se usan datos del Censo Poblacional 2020 porque se intentó mantener consistencia entre los años 2020, 2021 y 2022 tomando la información de una misma fuente (Conapo).

En *MuniCOVID-2022*, la cantidad total de municipios de México es de 2,465. La cantidad total de atributos (columnas) es de 451. La mayoría de estos son conteos municipales de casos positivos, defunciones y tasas por periodos semanales y mensuales, así como proyecciones de población municipal por sexo e intervalos de edad. La tabla 93 presenta una versión resumida del diccionario de datos del *dataset*. La versión completa y detallada del diccionario puede descargarse de la Nube (ver la liga de descarga del *dataset*).

Para crear *MuniCovid-2022*, se realizó un proceso de sumarización por municipio a partir de los casos de individuos, disponibles en nuestro *dataset Master-3* (ver Capítulo. 2). Este último fue producido previamente a partir de la base COVID-19 de la SSA. Después de la sumarización municipal, se realiza una mezcla (join o merge) con el *dataset* de proyecciones de población municipal por sexo y segmentos de edades, proveniente de Conapo. Esa mezcla se basa en los datos identificadores de entidad federativa y de municipio, establecidos por INEGI, que están disponibles en las dos fuentes.

**LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL**

Tabla 93. Diccionario de datos del dataset MuniCovid-2022 (versión resumida).

No.	Atributo	Descripción	Tipo de dato
1	ENT_MPIO_RES	Identificador de entidad federativa y municipio de residencia	Alfanumérico
2	ENTIDAD	Identificador de entidad federativa según INEGI	Número entero
3	NOM_ENT_ABREV	Abreviatura de entidad federativa	Alfanumérico
4	MUN	Identificador de municipio según INEGI	Número entero
5	NOM_MUN	Nombre completo de municipio	Alfanumérico
6	CLASIFICACIÓN_FINAL_1	Cantidad de casos positivos de clasificación final tipo 1: casos de COVID-19 confirmados por asociación clínica epidemiológica. Confirmado por asociación aplica cuando el caso informó ser contacto de un positivo a COVID-19 (y este se encuentra registrado en el Sisver) y al caso no se le tomó muestra o la muestra resultó no válida.	Número entero
7	CLASIFICACIÓN_FINAL_2	Cantidad de casos positivos de clasificación final tipo 2: casos de COVID-19 confirmados por comité de dictaminación. Confirmado por dictaminación solo aplica para defunciones bajo las siguientes condiciones: al caso no se le tomó muestra o si se tomó muestra, pero esta resultó no válida.	Número entero
8	CLASIFICACIÓN_FINAL_3	Cantidad de casos positivos de clasificación final tipo 3: casos de SARS-CoV-2 confirmados. Confirmado aplica cuando: el caso tiene muestra de laboratorio o prueba antigénica y resultó positiva a SARS-CoV-2, sin importar si el caso tiene asociación clínica epidemiológica.	Número entero
9	posit_AÑO_MES_INGRESO_2020_2 a posit_AÑO_MES_INGRESO_2022_9_17 (atributos 9 a 40)	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), por mes de ingreso al establecimiento de salud, del periodo de febrero/2020 a 17/Sep/2022.	Número entero
41	posit_FALLECIÓ_no	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3) que NO fallecieron en el periodo de febrero de 2020 a 17 de septiembre de 2022.	Número entero
42	posit_FALLECIÓ_si	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3) que SÍ fallecieron en el periodo de febrero de 2020 a 17 de septiembre de 2022.	Número entero
43	posit_SEMANA_EPIDEM_2020_09 a posit_SEMANA_EPIDEM_2022_37 (atributos 43 a 176)	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), por semana epidemiológica de ingreso al establecimiento de salud, en las semanas de 2020_09 a 2022_37.	Número entero
177	posit_MPAL_2020	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), de febrero a diciembre de 2020.	Número entero
178	posit_MPAL_2021	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), del año 2021.	Número entero
179	posit_MPAL_2022_AL_09_17	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), del año 2022 (solamente hasta 17/Sep/2022).	Número entero
180	posit_MPAL_TOTAL	Cantidad de casos positivos (según clasificación final con valor 1, 2 o 3), desde febrero de 2020 hasta 17/Sep/2022.	Número entero
181	def_AÑO_MES_INGRESO_2020_2 a def_AÑO_MES_INGRESO_2022_9_17 (atributos 181 a 212)	Cantidad de defunciones por COVID-19 por mes de ingreso al establecimiento de salud en el periodo febrero/2020 a 17/Sep/2022.	Número entero
213	def_SEMANA_EPIDEM_2020_09 a def_SEMANA_EPIDEM_2022_37 (atributos 213 a 346)	Cantidad de defunciones por COVID-19 por semana epidemiológica de ingreso al establecimiento de salud en las semanas de 2020_09 a 2022_37.	Número entero
347	def_MPAL_2020	Cantidad de defunciones por COVID-19 de febrero a diciembre de 2020.	Número entero
348	def_MPAL_2021	Cantidad de defunciones por COVID-19 en el año 2021.	Número entero
349	def_MPAL_2022_al_09_17	Cantidad de defunciones por COVID-19 en el año 2022, solamente hasta 17/Sep/2022.	Número entero
350	def_MPAL_TOTAL	Cantidad de defunciones por COVID-19 en el periodo de febrero de 2020 a 17/Sep/2022.	Número entero
351	pobm_2020_homb_00_04 a pobm_2022_muj_65_mm (atributos 351 a 434)	Cantidades de habitantes hombres y mujeres por intervalos de edades en 2020, 2021 y 2022 (proyección Conapo en año 2019).	Número entero
435	pobm_2020_homb	Cantidad de habitantes hombres en 2020	Número entero
436	pobm_2020_muj	Cantidad de habitantes mujeres en 2020	Número entero
437	pobm_2020_total	Población total del municipio en 2020	Número entero
438	pobm_2021_homb	Cantidad de habitantes hombres en 2021	Número entero
439	pobm_2021_muj	Cantidad de habitantes mujeres en 2021	Número entero
440	pobm_2021_total	Población total del municipio en el año 2021	Número entero
441	pobm_2022_homb	Cantidad de habitantes hombres en 2022	Número entero
442	pobm_2022_muj	Cantidad de habitantes mujeres en 2022	Número entero
443	pobm_2022_total	Población total del municipio en el año 2022	Número entero
444	posit_2020_tasa_10 mil	Tasa de casos positivos (según clasificación final con valor 1, 2 o 3) por cada diez mil habitantes en 2020.	Número con punto decimal
445	posit_2020_tasa_100 mil	Tasa de casos positivos (según clasificación final con valor 1, 2 o 3) por cada cien mil habitantes en 2020.	Número con punto decimal
446	posit_2021_tasa_10 mil	Tasa de casos positivos (según clasificación final con valor 1, 2 o 3) por cada diez mil habitantes en 2021.	Número con punto decimal
447	posit_2021_tasa_100 mil	Tasa de casos positivos (según clasificación final con valor 1, 2 o 3) por cada cien mil habitantes en 2021.	Número con punto decimal
448	def_2020_tasa_10 mil	Tasa de defunciones por COVID-19 por cada diez mil habitantes en 2020.	Número con punto decimal
449	def_2020_tasa_100 mil	Tasa de defunciones por COVID-19 por cada cien mil habitantes en 2020.	Número con punto decimal
450	def_2021_tasa_10 mil	Tasa de defunciones por COVID-19 por cada diez mil habitantes en 2021.	Número con punto decimal
451	def_2021_tasa_100 mil	Tasa de defunciones por COVID-19 por cada cien mil habitantes en 2021.	Número con punto decimal

Fuente: Elaboración propia con información de SSA, 2022 y Conapo, 2019.

5.2 Análisis por frecuencia absoluta

En los análisis estadísticos de frecuencias absolutas de casos positivos y de defunciones por municipio, la cantidad total de municipios es de 2,465, que es la cantidad total existente en la base de datos COVID-19 de (SSA, 2022). En cambio, en los análisis de las tasas por cada 100 mil habitantes, la cantidad de municipios es de 2,457 porque esta es la cantidad de municipios en las proyecciones de población municipal total, por sexo y por intervalos de edades en la base de datos de Conapo, que se usaron para calcular las tasas de los años 2020, 2021 y 2022. A continuación, se presentan más de treinta tablas con estos análisis.

La exhaustividad de estos análisis es necesaria, abordando diferentes periodos y profundizando en algunos subconjuntos de intervalos numéricos, porque así se logra una exploración profunda del comportamiento del fenómeno de interés. Tanto de los análisis de casos positivos, como de defunciones, se presentan tablas y gráficas bajo diferentes criterios de conteo, como se enumera a continuación:

- i. Por año y mes de ingreso, desde 2020 hasta 17/Sep/2022*
- ii. Solo del año 2020*
- iii. Solo del año 2021*
- iv. Solo del año 2022 (hasta 17/Sep)*
- v. De todo el periodo de interés: 2020 a 17/Sep/2022*

5.2.1 Frecuencias absolutas de casos positivos

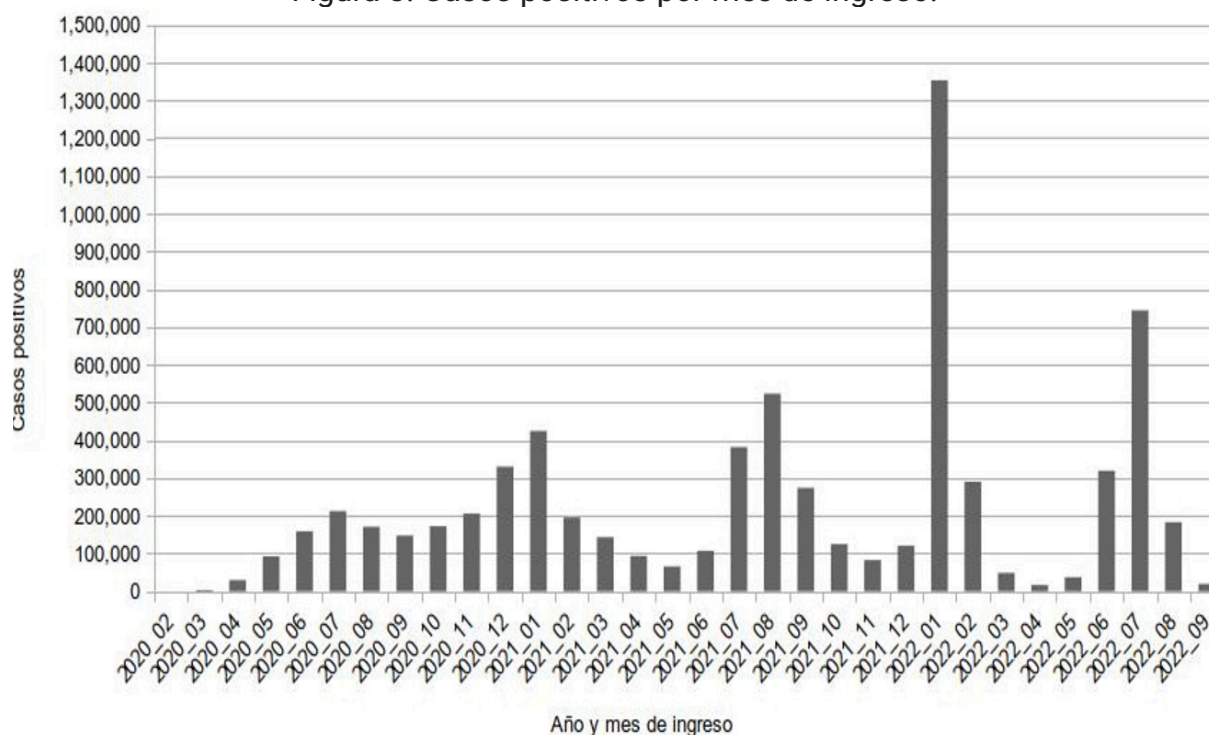
Tabla 94. Conteos de casos positivos por año y mes de ingreso.

Año y mes de ingreso	Conteo
2020_02	8
2020_03	3,122
2020_04	29,650
2020_05	92,470
2020_06	158,886
2020_07	212,274
2020_08	170,819
2020_09	147,468
2020_10	172,147
2020_11	206,038
2020_12	329,996
2021_01	424,394
2021_02	195,234
2021_03	143,254
2021_04	93,647
2021_05	65,871
2021_06	107,135
2021_07	381,611
2021_08	523,222
2021_09	273,774
2021_10	124,864
2021_11	82,496
2021_12	120,519
2022_01	1,353,213
2022_02	290,058
2022_03	48,608
2022_04	16,365
2022_05	37,220
2022_06	319,049
2022_07	743,875
2022_08	183,011
2022_09 (hasta 17/Sep)	19,245
Total	7,069,543

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 94 y la Figura 8 muestran que los cinco meses con mayores números de casos positivos fueron: enero y julio de 2022 y agosto, enero y julio de 2021.

Figura 8. Casos positivos por mes de ingreso.



Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 95. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020.

Cant. de casos positivos	Cant. de municipios	% de 2,465
De 0 a 3,500	2,375	96.3%
De 3,500 a 7,000	38	1.5%
De 7,000 a 10,500	16	0.6%
De 10,500 a 14,000	8	0.3%
De 14,000 a 17,500	8	0.3%
De 17,500 a 21,000	6	0.2%
De 21,000 a 24,500	4	0.2%
De 24,500 a 28,000	4	0.2%
De 28,000 a 31,500	1	0.0%
De 31,500 a 35,000	2	0.1%
De 35,000 a 38,500	0	0.0%
De 38,500 a 42,000	1	0.0%
De 42,000 a 45,500	1	0.0%
De 45,500 a 49,000	0	0.0%
De 49,000 a 52,500	0	0.0%
De 52,500 a 56,000	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 95 muestra que, en el año 2020, el 96.3% de los municipios tuvieron menos de 3,500 casos positivos cada uno. Debido al interés que generan estos municipios, a continuación se muestra otro análisis que se enfoca solamente en

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

ellos. Los cinco municipios con mayores conteos en este periodo fueron: cuatro alcaldías de la Ciudad de México: Iztapalapa (52,953), Gustavo A. Madero (42,016), Álvaro Obregón (39,808), y Tlalpan (34,563) y el municipio de Puebla, Puebla. (32,570).

Tabla 96. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020: solo el segmento de 0 a 3,500.

Cant. de casos positivos	Cant. de municipios	% de 2,375	% de 2,465
De 0 a 250	1,993	83.9%	81.1%
De 250 a 500	155	6.5%	6.3%
De 500 a 750	74	3.1%	3.0%
De 750 a 1,000	31	1.3%	1.3%
De 1,000 a 1,250	27	1.1%	1.1%
De 1,250 a 1,500	23	1.0%	0.9%
De 1,500 a 1,750	20	0.8%	0.8%
De 1,750 a 2,000	13	0.5%	0.5%
De 2,000 a 2,250	6	0.3%	0.2%
De 2,250 a 2,500	10	0.4%	0.4%
De 2,500 a 2,750	6	0.3%	0.2%
De 2,750 a 3,000	6	0.3%	0.2%
De 3,000 a 3,250	4	0.2%	0.2%
De 3,250 a 3,500	7	0.3%	0.3%
TOTAL	2,375	100.0%	96.7%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 96 muestra que, en el año 2020, el 81.1% de los municipios tuvo entre 0 y 250 casos positivos cada uno. Debido a que estos municipios resultan de interés, se necesita analizarlos más profundamente, lo cual se muestra en la siguiente tabla.

Tabla 97. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2020: solo el segmento de 0 a 260.

Cant. de casos positivos	Cant. de municipios	% de 2,002 municipios	% de 2,465 municipios
De 0 a 20	1,012	50.5%	41.1%
De 20 a 40	282	14.1%	11.4%
De 40 a 60	184	9.2%	7.5%
De 60 a 80	129	6.4%	5.2%
De 80 a 100	84	4.2%	3.4%
De 100 a 120	79	3.9%	3.2%
De 120 a 140	47	2.3%	1.9%
De 140 a 160	43	2.1%	1.7%
De 160 a 180	40	2.0%	1.6%
De 180 a 200	37	1.8%	1.5%
De 200 a 220	22	1.1%	0.9%
De 220 a 240	20	1.0%	0.8%
De 240 a 260	23	1.1%	0.9%
TOTAL	2,002	100.0%	81.2%

Fuente: Elaboración propia con datos de SSA, 2022.

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

La Tabla 97 muestra que, en el año 2020, el 41.1% de los municipios tenían solamente entre 0 y 20 casos positivos cada uno.

Tabla 98. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021.

Cant. de casos positivos	Cant. de municipios	Porcentaje % de 2,465 municipios
De 0 a 7,000	2,384	96.7%
De 7,000 a 14,000	39	1.6%
De 14,000 a 21,000	17	0.7%
De 21,000 a 28,000	6	0.2%
De 28,000 a 35,000	10	0.4%
De 35,000 a 42,000	4	0.2%
De 42,000 a 49,000	1	0.0%
De 49,000 a 56,000	0	0.0%
De 56,000 a 63,000	0	0.0%
De 63,000 a 70,000	1	0.0%
De 70,000 a 77,000	1	0.0%
De 77,000 a 84,000	0	0.0%
De 84,000 a 91,000	0	0.0%
De 91,000 a 98,000	0	0.0%
De 98,000 a 105,000	0	0.0%
De 105,000 a 112,000	2	0.1%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 98 muestra que en el año 2021 casi todos los municipios (96.7%) tuvieron menos de 7,000 casos positivos cada uno. Esto es razonable porque la mayoría de los municipios tienen poblaciones menores a los 50,000 habitantes. Por ello, se presenta a continuación, un análisis específico de los municipios correspondientes al intervalo mencionado. Al igual que en el año 2020, los cinco municipios con mayores conteos de positivos en 2021 fueron cuatro alcaldías de la Ciudad de México: Álvaro Obregón (107,596), Iztapalapa (105,629), Gustavo A. Madero (74,593), Tlalpan (68,791) y el municipio de Puebla, Puebla. (47,999).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 99. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021: solo el segmento de entre 0 y 7,000.

Cant. de casos positivos	Cant. de municipios	% de 2,384 municipios	% de 2,465 municipios
De 0 a 500	2,019	84.7%	82.2%
De 500 a 1,000	153	6.4%	6.2%
De 1,000 a 1,500	68	2.9%	2.8%
De 1,500 a 2,000	39	1.6%	1.6%
De 2,000 a 2,500	30	1.3%	1.2%
De 2,500 a 3,000	11	0.5%	0.4%
De 3,000 a 3,500	11	0.5%	0.4%
De 3,500 a 4,000	11	0.5%	0.4%
De 4,000 a 4,500	13	0.5%	0.5%
De 4,500 a 5,000	8	0.3%	0.3%
De 5,000 a 5,500	8	0.3%	0.3%
De 5,500 a 6,000	2	0.1%	0.1%
De 6,000 a 6,500	7	0.3%	0.3%
De 6,500 a 7,000	4	0.2%	0.2%
TOTAL	2,384	100.0%	97.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 99 muestra que, en 2021, el 82.2% de los municipios tuvo menos de 500 casos positivos cada uno. Debido a este considerable porcentaje, se muestra a continuación un análisis que se enfoca en ese intervalo de cantidades de casos.

Tabla 100. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2021: solo el segmento de entre 0 y 500.

Cant. de casos positivos	Cant. de municipios	% de 2,019 municipios	% de 2,465 municipios
De 0 a 50	1,151	57.0%	46.7%
De 50 a 100	304	15.1%	12.3%
De 100 a 150	160	7.9%	6.5%
De 150 a 200	100	5.0%	4.1%
De 200 a 250	91	4.5%	3.7%
De 250 a 300	59	2.9%	2.4%
De 300 a 350	65	3.2%	2.6%
De 350 a 400	35	1.7%	1.4%
De 400 a 450	33	1.6%	1.3%
De 450 a 500	21	1.0%	0.9%
TOTAL	2,019	100.0%	81.9%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 100 muestra que, en el año 2021, casi la mitad (46.7%) de los municipios tuvieron menos de 50 casos positivos cada uno.

Tabla 101. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep).

Cant. de casos positivos	Cant. de municipios	% de 2,465 municipios
De 0 a 7,000	2,375	96.3%
7,000 a 14,000	34	1.4%
14,000 a 21,000	18	0.7%
21,000 a 28,000	10	0.4%
28,000 a 35,000	9	0.4%
35,000 a 42,000	5	0.2%
42,000 a 49,000	3	0.1%
49,000 a 56,000	5	0.2%
56,000 a 63,000	1	0.0%
63,000 a 70,000	1	0.0%
70,000 a 77,000	1	0.0%
77,000 a 84,000	0	0.0%
84,000 a 91,000	1	0.0%
91,000 a 98,000	2	0.1%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 101 muestra que, en el año 2022 (hasta el 17 de septiembre), la gran mayoría (96.3%) de los municipios tuvo menos de 7,000 casos positivos cada uno. Por ello, la siguiente tabla se enfoca en ese intervalo de números de casos. Los cinco municipios con mayores conteos en este periodo fueron: Gustavo A. Madero, Ciudad de México (97,781); Iztapalapa, CDMX (96,004); San Luis Potosí, SLP (85,257); Álvaro Obregón, CDMX (76,718) y Tlalpan, CDMX (64,364).

Tabla 102. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep): solo el segmento de 0 a 7,000.

Cant. de casos positivos	Cant. de municipios	% de 2,375 municipios	% de 2,465 municipios
De 0 a 500	2,031	85.5%	82.7%
De 500 a 1,000	147	6.2%	6.0%
De 1,000 a 1,500	59	2.5%	2.4%
De 1,500 a 2,000	37	1.6%	1.5%
De 2,000 a 2,500	26	1.1%	1.1%
De 2,500 a 3,000	14	0.6%	0.6%
De 3,000 a 3,500	9	0.4%	0.4%
De 3,500 a 4,000	13	0.5%	0.5%
De 4,000 a 4,500	6	0.3%	0.2%
De 4,500 a 5,000	8	0.3%	0.3%
De 5,000 a 5,500	8	0.3%	0.3%
De 5,500 a 6,000	5	0.2%	0.2%
De 6,000 a 6,500	9	0.4%	0.4%
De 6,500 a 7,000	3	0.1%	0.1%
TOTAL	2,375	100.0%	96.7%

Fuente: Elaboración propia con datos de SSA, 2022.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La Tabla 102 muestra que, en 2022 (hasta el 17 de septiembre), la mayoría (82.7%) de los municipios tuvo menos de 500 casos cada uno. Por ello, la siguiente tabla se enfoca en ese segmento.

Tabla 103. Cantidad de municipios por intervalo de casos positivos de COVID-19 en el año 2022 (hasta 17/Sep): solo el segmento de 0 a 500.

Cant. de casos positivos	Cant. de municipios	% de 2,031 municipios	% de 2,465 municipios
De 0 a 25	1,044	51.4%	42.4%
De 25 a 50	248	12.2%	10.1%
De 50 a 75	157	7.7%	6.4%
De 75 a 100	118	5.8%	4.8%
De 100 a 125	73	3.6%	3.0%
De 125 a 150	60	3.0%	2.4%
De 150 a 175	42	2.1%	1.7%
De 175 a 200	47	2.3%	1.9%
De 200 a 225	49	2.4%	2.0%
De 225 a 250	33	1.6%	1.3%
De 250 a 275	27	1.3%	1.1%
De 275 a 300	22	1.1%	0.9%
De 300 a 325	15	0.7%	0.6%
De 325 a 350	12	0.6%	0.5%
De 350 a 375	10	0.5%	0.4%
De 375 a 400	21	1.0%	0.9%
De 400 a 425	12	0.6%	0.5%
De 425 a 450	14	0.7%	0.6%
De 450 a 475	14	0.7%	0.6%
De 475 a 500	13	0.6%	0.5%
TOTAL	2,031	100.0%	82.4%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 103 muestra que en 2022 (hasta el 17 de septiembre) el 42.4% de los municipios tuvieron menos de 25 casos positivos cada uno.

Tabla 104. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022.

Cant. de casos positivos	Cant. de municipios	% de 2,465 municipios
De 0 a 15,000	2,365	95.9%
De 15,000 a 30,000	44	1.8%
De 30,000 a 45,000	16	0.6%
De 45,000 a 60,000	12	0.5%
De 60,000 a 75,000	7	0.3%
De 75,000 a 90,000	5	0.2%
De 90,000 a 105,000	7	0.3%
De 105,000 a 120,000	2	0.1%
De 120,000 a 135,000	2	0.1%
De 135,000 a 150,000	1	0.0%
De 150,000 a 165,000	0	0.0%
De 165,000 a 180,000	1	0.0%
De 180,000 a 195,000	0	0.0%
De 195,000 a 210,000	0	0.0%
De 210,000 a 225,000	2	0.1%
De 225,000 a 240,000	0	0.0%
De 240,000 a 255,000	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 104 muestra que, en todo el periodo analizado, desde 2020 hasta el 17 de septiembre de 2022, casi todos los municipios (95.9%) tuvieron menos de 15,000 casos positivos cada uno. Debido a este gran porcentaje, la siguiente tabla se enfoca en los municipios de ese segmento. Los cinco municipios con mayores conteos en todo el periodo analizado fueron cuatro alcaldías de Ciudad de México: Iztapalapa (254,586), Álvaro Obregón (224,122), Gustavo A. Madero (214,390) y Tlalpan (167,718) y el municipio de San Luis Potosí, SLP (143,912).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 105. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022: solo el segmento de 0 a 15,000.

Cant. de casos positivos	Cant. de municipios	% de 2,365 municipios	% de 2,465 municipios
De 0 a 1,000	1,955	82.7%	79.3%
De 1,000 a 2,000	174	7.4%	7.1%
De 2,000 a 3,000	63	2.7%	2.6%
De 3,000 a 4,000	56	2.4%	2.3%
De 4,000 a 5,000	24	1.0%	1.0%
De 5,000 a 6,000	29	1.2%	1.2%
De 6,000 a 7,000	9	0.4%	0.4%
De 7,000 a 8,000	8	0.3%	0.3%
De 8,000 a 9,000	12	0.5%	0.5%
De 9,000 a 10,000	6	0.3%	0.2%
De 10,000 a 11,000	5	0.2%	0.2%
De 11,000 a 12,000	7	0.3%	0.3%
De 12,000 a 13,000	8	0.3%	0.3%
De 13,000 a 14,000	5	0.2%	0.2%
De 14,000 a 15,000	4	0.2%	0.2%
TOTAL	2,365	100.0%	95.9%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 105 muestra que en el periodo analizado (2020 a 17/Sep/2022), el 86.4% (79.3% + 7.1%) de los municipios tuvieron menos de 2,000 casos positivos cada uno. Por ello, la siguiente tabla se enfoca en ese segmento.

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Tabla 106. Cantidad de municipios por intervalo de casos positivos de COVID-19 desde el año 2020 hasta 17/Sep/2022: solo el segmento de 0 a 2,000.

Cant. de casos positivos	Cant. de municipios	% de 2,129 municipios	% de 2,465 municipios
De 0 a 100	1,061	49.8%	43.0%
De 100 a 200	314	14.7%	12.7%
De 200 a 300	175	8.2%	7.1%
De 300 a 400	111	5.2%	4.5%
De 400 a 500	72	3.4%	2.9%
De 500 a 600	68	3.2%	2.8%
De 600 a 700	52	2.4%	2.1%
De 700 a 800	43	2.0%	1.7%
De 800 a 900	39	1.8%	1.6%
De 900 a 1,000	20	0.9%	0.8%
De 1,000 a 1,100	22	1.0%	0.9%
De 1,100 a 1,200	27	1.3%	1.1%
De 1,200 a 1,300	24	1.1%	1.0%
De 1,300 a 1,400	19	0.9%	0.8%
De 1,400 a 1,500	15	0.7%	0.6%
De 1,500 a 1,600	19	0.9%	0.8%
De 1,600 a 1,700	8	0.4%	0.3%
De 1,700 a 1,800	11	0.5%	0.4%
De 1,800 a 1,900	19	0.9%	0.8%
De 1,900 a 2,000	10	0.5%	0.4%
TOTAL	2,129	100.0%	86.4%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 106 muestra que, en todo el periodo analizado (2020 a 17/Sep/2022), el 43.0% de los municipios tuvieron menos de 100 casos cada uno.

5.2.2 Frecuencias absolutas de defunciones

Tabla 107: Conteos de defunciones por año y mes.

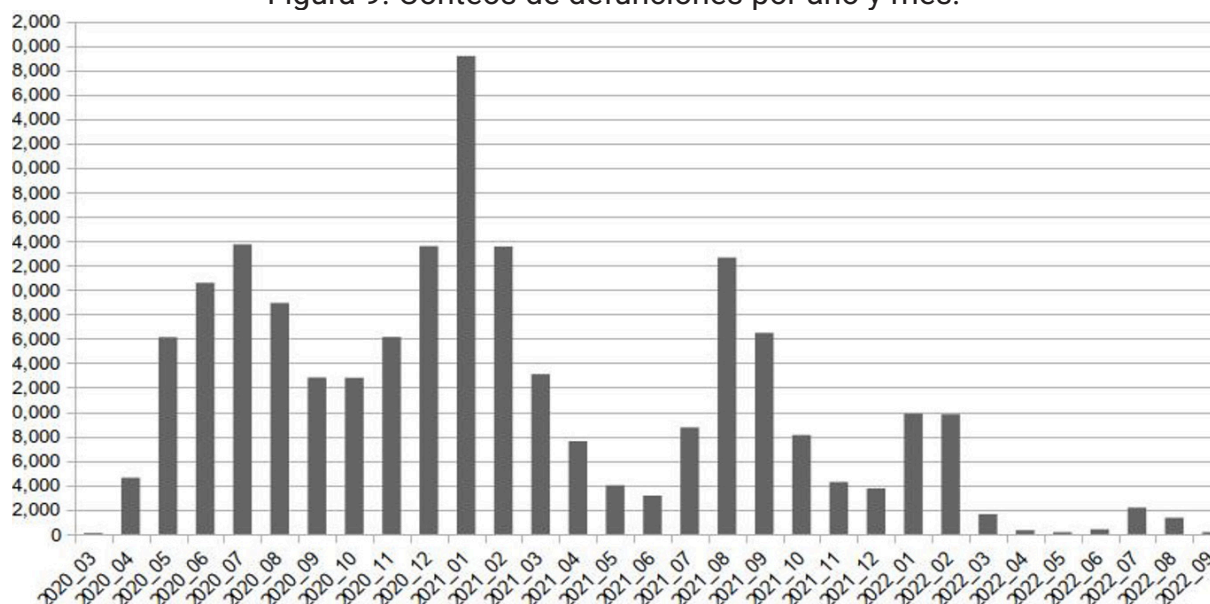
Año y mes de defunción	Conteo
2020_03	92
2020_04	4,617
2020_05	16,132
2020_06	20,584
2020_07	23,732
2020_08	18,932
2020_09	12,826
2020_10	12,802
2020_11	16,147
2020_12	23,591
2021_01	39,156
2021_02	23,558
2021_03	13,099
2021_04	7,606
2021_05	4,004
2021_06	3,152
2021_07	8,735
2021_08	22,656
2021_09	16,481
2021_10	8,110
2021_11	4,268
2021_12	3,745
2022_01	9,869
2022_02	9,810
2022_03	1,634
2022_04	319
2022_05	160
2022_06	393
2022_07	2,175
2022_08	1,345
2022_09 (hasta 17/Sep)	166
TOTAL	329,896

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 107 y la figura 9 muestran que los cinco meses con mayores cantidades de defunciones fueron: enero de 2021, julio y diciembre de 2020, y febrero y agosto de 2021.

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Figura 9. Conteos de defunciones por año y mes.



Fuente: Elaboración propia con datos de SSA, 2022.

Tabla 108. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020.

Cantidad de defunciones	Cant. de municipios	% de 2,465 municipios
De 0 a 300	2,363	95.9%
De 300 a 600	41	1.7%
De 600 a 900	21	0.9%
De 900 a 1,200	9	0.4%
De 1,200 a 1,500	9	0.4%
De 1,500 a 1,800	8	0.3%
De 1,800 a 2,100	3	0.1%
De 2,100 a 2,400	1	0.0%
De 2,400 a 2,700	2	0.1%
De 2,700 a 3,000	2	0.1%
De 3,000 a 3,300	1	0.0%
De 3,300 a 3,600	3	0.1%
De 3,600 a 3,900	1	0.0%
De 3,900 a 4,200	0	0.0%
De 4,200 a 4,500	0	0.0%
De 4,500 a 4,800	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 108 muestra que, en el año 2020, la gran mayoría (95.9%) de los municipios tuvieron menos de 300 defunciones cada uno. Por ello, la siguiente tabla se enfoca en este intervalo. Los cinco municipios con mayores conteos en este periodo fueron: Iztapalapa, CDMX (4,520); Puebla, Puebla. (3,722); Ecatepec de Morelos, Edo. de México. (3,590); Gustavo A. Madero, CDMX (3,573) y Ciudad Juárez, Chihuahua (3,411).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 109. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020: solo el segmento entre 0 y 300 defunciones.

Cantidad de defunciones	Cant. de municipios	% de 2,363 municipios	% de 2,465 municipios
De 0 a 20	1,795	76.0%	73.1%
De 20 a 40	244	10.3%	9.9%
De 40 a 60	95	4.0%	3.9%
De 60 a 80	67	2.8%	2.7%
De 80 a 100	35	1.5%	1.4%
De 100 a 120	37	1.6%	1.5%
De 120 a 140	22	0.9%	0.9%
De 140 a 160	11	0.5%	0.4%
De 160 a 180	9	0.4%	0.4%
De 180 a 200	5	0.2%	0.2%
De 200 a 220	10	0.4%	0.4%
De 220 a 240	7	0.3%	0.3%
De 240 a 260	11	0.5%	0.4%
De 260 a 280	7	0.3%	0.3%
De 280 a 300	8	0.3%	0.3%
TOTAL	2,363	100.0%	96.2%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 109 muestra que, en 2020, el 73.1% de los municipios tuvo menos de 20 defunciones cada uno. Por ello, la siguiente tabla se enfoca en ese segmento para ofrecer un mayor entendimiento del fenómeno.

Tabla 110. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2020: solo el segmento entre 0 y 20.

Cantidad de defunciones	Cant. de municipios	% de 1,787 municipios	% de 2,457 municipios
De 0 a 2	711	39.8%	28.9%
De 2 a 4	316	17.7%	12.9%
De 4 a 6	214	12.0%	8.7%
De 6 a 8	129	7.2%	5.3%
De 8 a 10	126	7.1%	5.1%
De 10 a 12	83	4.6%	3.4%
De 12 a 14	67	3.7%	2.7%
De 14 a 16	54	3.0%	2.2%
De 16 a 18	45	2.5%	1.8%
De 18 a 20	42	2.4%	1.7%
TOTAL	1,787	100.0%	72.7%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 110 muestra que, en 2020, casi la tercera parte (28.9%) de los municipios tuvo menos de dos defunciones por COVID-19 cada uno.

Tabla 111. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021.

Cantidad de defunciones	Cant. de municipios	% de 2,465 municipios
De 0 a 250	2,350	95.3%
De 250 a 500	49	2.0%
De 500 a 750	23	0.9%
De 750 a 1,000	10	0.4%
De 1,000 a 1,250	11	0.4%
De 1,250 a 1,500	7	0.3%
De 1,500 a 1,750	4	0.2%
De 1,750 a 2,000	2	0.1%
De 2,000 a 2,250	2	0.1%
De 2,250 a 2,500	2	0.1%
De 2,500 a 2,750	1	0.0%
De 2,750 a 3,000	0	0.0%
De 3,000 a 3,250	2	0.1%
De 3,250 a 3,500	0	0.0%
De 3,500 a 3,750	1	0.0%
De 3,750 a 4,000	0	0.0%
De 4,000 a 4,250	0	0.0%
De 4,250 a 4,500	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 111 muestra que, en 2021, casi todos los municipios (95.3%) tuvieron menos de 250 defunciones cada uno. Por ello, la siguiente tabla se enfoca en ese segmento. Los cinco municipios con mayores conteos en este periodo fueron: Puebla, Puebla. (4,342); Iztapalapa, CDMX (3,611); Guadalajara, Jalisco. (3,236); Ecatepec de Morelos, Estado de México (3,049) y Gustavo A. Madero, CDMX (2,739).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 112. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021: solo el segmento entre 0 y 255 defunciones.

Cant. de defunciones	Cant. de municipios	% de 2,351 municipios	% de 2,465 municipios
De 0 a 15	1,567	66.7%	63.6%
De 15 a 30	329	14.0%	13.3%
De 30 a 45	154	6.6%	6.2%
De 45 a 60	78	3.3%	3.2%
De 60 a 75	47	2.0%	1.9%
De 75 a 90	40	1.7%	1.6%
De 90 a 105	26	1.1%	1.1%
De 105 a 120	19	0.8%	0.8%
De 120 a 135	14	0.6%	0.6%
De 135 a 150	13	0.6%	0.5%
De 150 a 165	16	0.7%	0.6%
De 165 a 180	7	0.3%	0.3%
De 180 a 195	12	0.5%	0.5%
De 195 a 210	11	0.5%	0.4%
De 210 a 225	6	0.3%	0.2%
De 225 a 240	6	0.3%	0.2%
De 240 a 255	6	0.3%	0.2%
TOTAL	2,351	100.0%	95.4%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 112 muestra que el 83.1% (63.6% + 13.3% + 6.2%) de los municipios tuvieron menos de 45 defunciones cada uno. Por ello, la siguiente tabla se enfoca en ese segmento.

Tabla 113. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2021: solo el segmento entre 0 y 45 defunciones.

Cant. de defunciones	Cant. de municipios	% de 2,050 municipios	% de 2,465 municipios
De 0 a 3	755	36.8%	30.6%
De 3 a 6	317	15.5%	12.9%
De 6 a 9	216	10.5%	8.8%
De 9 a 12	175	8.5%	7.1%
De 12 a 15	104	5.1%	4.2%
De 15 a 18	101	4.9%	4.1%
De 18 a 21	64	3.1%	2.6%
De 21 a 24	58	2.8%	2.4%
De 24 a 27	61	3.0%	2.5%
De 27 a 30	45	2.2%	1.8%
De 30 a 33	38	1.9%	1.5%
De 33 a 36	25	1.2%	1.0%
De 36 a 39	33	1.6%	1.3%
De 39 a 42	29	1.4%	1.2%
De 42 a 45	29	1.4%	1.2%
TOTAL	2,050	87.2%	83.2%

Fuente: Elaboración propia con datos de SSA, 2022.

La Tabla 113 muestra que, en 2021, casi la tercera parte (30.6%) de los municipios tuvieron menos de 3 defunciones cada uno.

Tabla 114. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep).

Cant. de defunciones	Cant. de municipios	% de 2,465 municipios
De 0 a 50	2,358	95.7%
De 50 a 100	49	2.0%
De 100 a 150	19	0.8%
De 150 a 200	13	0.5%
De 200 a 250	9	0.4%
De 250 a 300	4	0.2%
De 300 a 350	3	0.1%
De 350 a 400	2	0.1%
De 400 a 450	3	0.1%
De 450 a 500	1	0.0%
De 500 a 550	3	0.1%
De 550 a 600	0	0.0%
De 600 a 650	0	0.0%
De 650 a 700	0	0.0%
De 700 a 750	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 114 muestra que, en 2022 (hasta el 17 de septiembre) casi todos los municipios (95.7%) tuvieron menos de 50 defunciones cada uno. Por ello, la siguiente tabla se enfoca en el intervalo de entre 0 y 51 defunciones. Los cinco municipios con mayores conteos en este periodo fueron: Guadalajara, Jalisco. (740); Monterrey, Nuevo León (546); León, Guanajuato. (540); Tijuana, Baja California (510) y Gustavo A. Madero, CDMX (483).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 115. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep): solo el segmento entre 0 y 51

Cant. de defunciones	Cant. de municipios	Porcentaje % de 2,361 municipios	Porcentaje % de 2,465 municipios
De 0 a 3	1,633	69.2%	66.2%
De 3 a 6	314	13.3%	12.7%
De 6 a 9	141	6.0%	5.7%
De 9 a 12	81	3.4%	3.3%
De 12 a 15	53	2.2%	2.2%
De 15 a 18	32	1.4%	1.3%
De 18 a 21	25	1.1%	1.0%
De 21 a 24	15	0.6%	0.6%
De 24 a 27	14	0.6%	0.6%
De 27 a 30	13	0.6%	0.5%
De 30 a 33	12	0.5%	0.5%
De 33 a 36	9	0.4%	0.4%
De 36 a 39	1	0.0%	0.0%
De 39 a 42	8	0.3%	0.3%
De 42 a 45	5	0.2%	0.2%
De 45 a 48	2	0.1%	0.1%
De 48 a 51	3	0.1%	0.1%
TOTAL	2,361	100.0%	95.8%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 115 muestra que en 2022 (hasta septiembre 17), las dos terceras partes (66.2%) de los municipios tenían menos de 3 defunciones cada uno.

Tabla 116. Cantidad de municipios por intervalo de defunciones de COVID-19 en el año 2022 (solo hasta 17/Sep): solo el segmento entre 0 y 6.

Cant. de defunciones	Cant. de municipios	% de 2,008 municipios	% de 2,465 municipios
0	968	48.2%	39.3%
1	427	21.3%	17.3%
2	238	11.9%	9.7%
3	129	6.4%	5.2%
4	112	5.6%	4.5%
5	73	3.6%	3.0%
6	61	3.0%	2.5%
TOTAL	2,008	100.0%	81.5%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 116 muestra que, en 2022 (hasta el 17 de septiembre), más de la tercera parte (39.3%) de los municipios tenían cero defunciones.

Tabla 117. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022.

Cant. de defunciones	Cant. de municipios	% de 2,465 municipios
De 0 a 500	2,334	94.7%
De 500 a 1,000	55	2.2%
De 1,000 a 1,500	24	1.0%
De 1,500 a 2,000	15	0.6%
De 2,000 a 2,500	7	0.3%
De 2,500 a 3,000	10	0.4%
De 3,000 a 3,500	4	0.2%
De 3,500 a 4,000	4	0.2%
De 4,000 a 4,500	3	0.1%
De 4,500 a 5,000	1	0.0%
De 5,000 a 5,500	3	0.1%
De 5,500 a 6,000	0	0.0%
De 6,000 a 6,500	0	0.0%
De 6,500 a 7,000	3	0.1%
De 7,000 a 7,500	0	0.0%
De 7,500 a 8,000	0	0.0%
De 8,000 a 8,500	1	0.0%
De 8,500 a 9,000	1	0.0%
TOTAL	2,465	100.0%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 117 muestra que en todo el periodo analizado (2020 a 17/Sep/2022), casi todos los municipios (94.7%) tenían menos de 500 defunciones cada uno. Por ello, la siguiente tabla se enfoca en el segmento de entre 0 y 510. Los cinco municipios con mayores conteos en este periodo fueron: Iztapalapa, CDMX (8,562); Puebla, Puebla (8,417); Ecatepec de Morelos, Estado de México (6,932); Gustavo A. Madero, CDMX (6,795) y Guadalajara, Jalisco (6,733).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 118. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022: solo el segmento entre 0 y 510.

Cant. de defunciones	Cant. de municipios	% de 2,337 municipios	% de 2,465 municipios
De 0 a 30	1,581	67.7%	64.1%
De 30 a 60	312	13.4%	12.7%
De 60 a 90	150	6.4%	6.1%
De 90 a 120	70	3.0%	2.8%
De 120 a 150	49	2.1%	2.0%
De 150 a 180	41	1.8%	1.7%
De 180 a 210	26	1.1%	1.1%
De 210 a 240	23	1.0%	0.9%
De 240 a 270	17	0.7%	0.7%
De 270 a 300	18	0.8%	0.7%
De 300 a 330	14	0.6%	0.6%
De 330 a 360	9	0.4%	0.4%
De 360 a 390	8	0.3%	0.3%
De 390 a 420	6	0.3%	0.2%
De 420 a 450	4	0.2%	0.2%
De 450 a 480	4	0.2%	0.2%
De 480 a 510	5	0.2%	0.2%
TOTAL	2,337	100.0%	94.8%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 118 muestra que en el periodo analizado (2020 a 17/Sep/2022), el 82.9% (64.1% + 12.7% + 6.1%) de los municipios tuvieron menos de 90 defunciones cada uno. Por ello, la siguiente tabla se enfoca en ese intervalo.

Tabla 119. Cantidad de municipios por intervalo de defunciones de COVID-19 en el periodo de 2020 a 17/Sep/2022: solo el segmento entre 0 y 90.

Cant. de defunciones	Cant. de municipios	Porcentaje % de 2,043 municipios	Porcentaje % de 2,465 municipios
De 0 a 6	762	37.3%	30.9%
De 6 a 12	343	16.8%	13.9%
De 12 a 18	221	10.8%	9.0%
De 18 a 24	155	7.6%	6.3%
De 24 a 30	100	4.9%	4.1%
De 30 a 36	92	4.5%	3.7%
De 36 a 42	65	3.2%	2.6%
De 42 a 48	57	2.8%	2.3%
De 48 a 54	62	3.0%	2.5%
De 54 a 60	36	1.8%	1.5%
De 60 a 66	33	1.6%	1.3%
De 66 a 72	33	1.6%	1.3%
De 72 a 78	34	1.7%	1.4%
De 78 a 84	26	1.3%	1.1%
De 84 a 90	24	1.2%	1.0%
TOTAL	2,043	100.0%	82.9%

Fuente: Elaboración propia con datos de SSA, 2022.

La tabla 119 muestra que en todo el periodo analizado (2020 a 17/Sep/2022) casi la tercera parte (30.9%) de los municipios tuvo menos de 6 seis defunciones.

5.3 Análisis de tasas municipales por cada 100 mil habitantes

En esta sección se presentan análisis de las tasas municipales de casos positivos y de defunciones por cada 100,000 habitantes. Se calculan con base en las siguientes fórmulas:

Tasa de casos positivos = Cantidad de casos positivos / Población municipal X 100,000

Tasa de defunción = Cantidad de defunciones / Población municipal X 100,000

A continuación, se presenta una serie de tablas que muestran análisis de estas tasas.

5.3.1 Tasas de casos positivos

Tabla 120. Tasas de casos positivos (clasificación final 1, 2 ó 3) de COVID-19 por municipio en el año 2020.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 300	1,237	50.3%
De 300 a 600	566	23.0%
De 600 a 900	277	11.3%
De 900 a 1,200	148	6.0%
De 1,200 a 1,500	83	3.4%
De 1,500 a 1,800	57	2.3%
De 1,800 a 2,100	28	1.1%
De 2,100 a 2,400	14	0.6%
De 2,400 a 2,700	10	0.4%
De 2,700 a 3,000	9	0.4%
De 3,000 a 3,300	5	0.2%
De 3,300 a 3,600	4	0.2%
De 3,600 a 3,900	4	0.2%
De 3,900 a 4,200	3	0.1%
De 4,200 a 4,500	1	0.0%
De 4,500 a 4,800	2	0.1%
De 4,800 a 5,100	5	0.2%
De 5,100 a 5,400	3	0.1%
De 5,400 a 5,700	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 120 muestra que, en el año 2020, el 84.6% (50.3% + 23.0% + 11.3%) de los municipios tuvieron tasas que no superaron los 900 casos por cada 100 mil habitantes. Por ello, la tabla siguiente se enfoca en el segmento de esos municipios. Los cinco municipios con mayores tasas en este periodo fueron: Milpa Alta, CDMX (5,467.4); Sacramento, Coahuila (5,279.3); Álvaro Obregón, CDMX (5,268.8); Tláhuac, CDMX (5,174.5) y La Magdalena Contreras, CDMX (5,094.1).

Tabla 121. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2020: solo el segmento de tasas entre 0 y 900.

Tasa por cada 100 mil habitantes	Cant. de municipios	Porcentaje % de 2,080 municipios.	Porcentaje % de 2,457 municipios.
De 0 a 60	408	19.6%	16.6%
De 60 a 120	250	12.0%	10.2%
De 120 a 180	225	10.8%	9.2%
De 180 a 240	181	8.7%	7.4%
De 240 a 300	173	8.3%	7.0%
De 300 a 360	150	7.2%	6.1%
De 360 a 420	120	5.8%	4.9%
De 420 a 480	127	6.1%	5.2%
De 480 a 540	93	4.5%	3.8%
De 540 a 600	76	3.7%	3.1%
De 600 a 660	65	3.1%	2.6%
De 660 a 720	62	3.0%	2.5%
De 720 a 780	63	3.0%	2.6%
De 780 a 840	35	1.7%	1.4%
De 840 a 900	52	2.5%	2.1%
TOTAL	2,080	100.0%	84.7%

Fuente: Elaboración propia con datos de (SSA, 2022,) y (Conapo, 2019).

La tabla 121 muestra que, en 2020, la cuarta parte de los municipios (26.8%) tuvo menos de 120 casos positivos por cada 100 mil habitantes.

Tabla 122. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2021.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 1,000	1,689	68.7%
De 1,000 a 2,000	510	20.8%
De 2,000 a 3,000	165	6.7%
De 3,000 a 4,000	47	1.9%
De 4,000 a 5,000	20	0.8%
De 5,000 a 6,000	13	0.5%
De 6,000 a 7,000	5	0.2%
De 7,000 a 8,000	2	0.1%
De 8,000 a 9,000	1	0.0%
De 9,000 a 10,000	3	0.1%
De 10,000 a 11,000	1	0.0%
De 11,000 a 12,000	0	0.0%
De 12,000 a 13,000	0	0.0%
De 13,000 a 14,000	0	0.0%
De 14,000 a 15,000	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La tabla 122 muestra que, en 2021, el 89.5% (68.7% + 20.8%) de los municipios tuvieron tasas de casos positivos menores a 2,000 casos por cada 100 mil habitantes. Debido a este alto porcentaje, la siguiente tabla se enfoca en ese segmento de municipios. Los cinco municipios con mayores tasas en este periodo fueron: Álvaro Obregón, CDMX (14,250.3); Tlalpan, CDMX (10,084.9); Milpa Alta, CDMX (9,657.1); Santa Magdalena Jicotlán, Oaxaca (9,278.4) y Tláhuac, CDMX (9,154.7). La tasa sorprendentemente alta de Santa Magdalena Jicotlán se debe a que es un municipio con población muy escasa (97 habitantes.) y a que tuvo un número relativamente grande de casos (9).

Tabla 123. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2021: solo el segmento de tasas entre 0 y 2,000.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,199 municipios	Porcentaje % de 2,457 municipios
De 0 a 100	278	12.6%	11.3%
De 100 a 200	269	12.2%	10.9%
De 200 a 300	230	10.5%	9.4%
De 300 a 400	202	9.2%	8.2%
De 400 a 500	161	7.3%	6.6%
De 500 a 600	159	7.2%	6.5%
De 600 a 700	124	5.6%	5.0%
De 700 a 800	109	5.0%	4.4%
De 800 a 900	75	3.4%	3.1%
De 900 a 1,000	82	3.7%	3.3%
De 1,000 a 1,100	76	3.5%	3.1%
De 1,100 a 1,200	70	3.2%	2.8%
De 1,200 a 1,300	68	3.1%	2.8%
De 1,300 a 1,400	58	2.6%	2.4%
De 1,400 a 1,500	47	2.1%	1.9%
De 1,500 a 1,600	47	2.1%	1.9%
De 1,600 a 1,700	41	1.9%	1.7%
De 1,700 a 1,800	34	1.5%	1.4%
De 1,800 a 1,900	39	1.8%	1.6%
De 1,900 a 2,000	30	1.4%	1.2%
TOTAL	2,199	100.0%	89.5%

Fuente: Elaboración propia, con datos de SSA, 2022 y Conapo, 2019.

La tabla 123 muestra que, en 2021, el 31.6% (11.3% + 10.9% + 9.4%) de los municipios tuvieron tasas menores a los 300 casos positivos por cada 100 mil habitantes.

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Tabla 124. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2022 (hasta 17/Sep).

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 1,000	1,826	74.3%
De 1,000 a 2,000	340	13.8%
De 2,000 a 3,000	156	6.3%
De 3,000 a 4,000	67	2.7%
De 4,000 a 5,000	24	1.0%
De 5,000 a 6,000	18	0.7%
De 6,000 a 7,000	4	0.2%
De 7,000 a 8,000	9	0.4%
De 8,000 a 9,000	5	0.2%
De 9,000 a 10,000	4	0.2%
De 10,000 a 11,000	3	0.1%
De 11,000 a 12,000	0	0.0%
De 12,000 a 13,000	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia, con datos de SSA, 2022 y Conapo, 2019.

La tabla 124 muestra que, en 2022 (hasta septiembre 17), la mayoría (74.3% + 13.8% = 88.1%) de los municipios tuvieron tasas de positivos menores a los 2,000 casos por cada 100 mil habitantes. Por ello, en la siguiente tabla se presenta un análisis enfocado en los municipios que tienen tasas menores a 2,100. Los cinco municipios con mayores tasas en este periodo fueron: Milpa Alta, CDMX (12,040.3); Azcapotzalco, CDMX (10,889.1); Álvaro Obregón, CDMX (10,172.9); La Paz, Baja California Sur (10,114.8) y San Luis Potosí, San Luis Potosí (9,675.7).

Tabla 125. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio en el año 2022 (hasta 17/Sep): solo el segmento de tasas entre 0 y 2,100.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,188 municipios	% de 2,457 municipios
De 0 a 150	762	34.8%	31.0%
De 150 a 300	375	17.1%	15.3%
De 300 a 450	250	11.4%	10.2%
De 450 a 600	177	8.1%	7.2%
De 600 a 750	115	5.3%	4.7%
De 750 a 900	99	4.5%	4.0%
De 900 a 1050	74	3.4%	3.0%
De 1050 a 1200	81	3.7%	3.3%
De 1200 a 1350	49	2.2%	2.0%
De 1350 a 1500	48	2.2%	2.0%
De 1500 a 1650	48	2.2%	2.0%
De 1650 a 1800	45	2.1%	1.8%
De 1800 a 1950	35	1.6%	1.4%
De 1950 a 2100	30	1.4%	1.2%
TOTAL	2,188	100.0%	89.1%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

La tabla 125 muestra que, en 2022 (hasta el 17 de septiembre), casi la tercera parte de los municipios (31.0%) tuvo tasas de positivos por debajo de 150 por cada 100 mil habitantes.

Tabla 126. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio desde 2020 hasta 17/Sep/2022.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 2,000	1,583	64.4%
De 2,000 a 4,000	468	19.0%
De 4,000 a 6,000	208	8.5%
De 6,000 a 8,000	106	4.3%
De 8,000 a 10,000	45	1.8%
De 10,000 a 12,000	13	0.5%
De 12,000 a 14,000	9	0.4%
De 14,000 a 16,000	6	0.2%
De 16,000 a 18,000	7	0.3%
De 18,000 a 20,000	6	0.2%
De 20,000 a 22,000	0	0.0%
De 22,000 a 24,000	3	0.1%
De 24,000 a 26,000	1	0.0%
De 26,000 a 28,000	1	0.0%
De 28,000 a 30,000	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 126 muestra que, en todo el periodo analizado, de 2020 hasta el 17 de septiembre de 2022, la gran mayoría (64.4% + 19.0% + 8.5% = 91.9%) de los municipios tuvieron tasas menores a 6,000 casos por cada 100 mil habitantes. Por ello, a continuación, se presenta un análisis enfocado en ese subconjunto de municipios. Los cinco municipios con mayores tasas en este periodo fueron alcaldías de la Ciudad de México: Álvaro Obregón (29,663.9), Milpa Alta (27,158.4), Tlalpan (24,583.6), Azcapotzalco (23,611.2) y Tláhuac (23,501.4).

Tabla 127. Tasas de casos positivos (clasificación final 1, 2 o 3) de COVID-19 por municipio desde 2020 hasta 17/Sep/2022: solo el segmento de tasas entre 0 y 6,000.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,259 municipios	% de 2,457 municipios
De 0 a 400	454	20.1%	18.5%
De 400 a 800	415	18.4%	16.9%
De 800 a 1,200	312	13.8%	12.7%
De 1,200 a 1,600	232	10.3%	9.4%
De 1,600 a 2,000	170	7.5%	6.9%
De 2,000 a 2,400	137	6.1%	5.6%
De 2,400 a 2,800	97	4.3%	3.9%
De 2,800 a 3,200	88	3.9%	3.6%
De 3,200 a 3,600	86	3.8%	3.5%
De 3,600 a 4,000	60	2.7%	2.4%
De 4,000 a 4,400	59	2.6%	2.4%
De 4,400 a 4,800	44	1.9%	1.8%
De 4,800 a 5,200	40	1.8%	1.6%
De 5,200 a 5,600	34	1.5%	1.4%
De 5,600 a 6,000	31	1.4%	1.3%
TOTAL	2,259	100.0%	91.9%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 127 muestra que en todo el periodo analizado (2020 a 17 de septiembre de 2022), poco más de la tercera parte (35.4%) de los municipios tuvieron tasas de positivos por debajo de los 800 casos por cada 100 mil habitantes.

5.3.2 Tasas de defunción

La tabla 128 muestra que, en 2020, la gran mayoría (78.4% + 18.5% = 96.9%) de los municipios tuvieron tasas menores a 200 defunciones por cada 100 mil habitantes. Por ello, la tabla que le sigue muestra el análisis del segmento de 0 a 200. Los cinco municipios con mayores tasas en este periodo fueron: Atizapán, Estado de México (1,462.7); San Vicente Nuñú, Oaxaca. (447.4); Iztacalco, CDMX (377.1); San Miguel Ixitlán, Puebla. (372.4) y Azcapotzalco, CDMX (361.9).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 128. Tasas de defunción por COVID-19 por municipio en el año 2020.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 100	1,927	78.4%
De 100 a 200	454	18.5%
De 200 a 300	63	2.6%
De 300 a 400	11	0.4%
De 400 a 500	1	0.0%
De 500 a 600	0	0.0%
De 600 a 700	0	0.0%
De 700 a 800	0	0.0%
De 800 a 900	0	0.0%
De 900 a 1,000	0	0.0%
De 1,000 a 1,100	0	0.0%
De 1,100 a 1,200	0	0.0%
De 1,200 a 1,300	0	0.0%
De 1,300 a 1,400	0	0.0%
De 1,400 a 1,500	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia, con datos de SSA, 2022 y Conapo, 2019.

Tabla 129. Tasas de defunción por COVID-19 por municipio en el año 2020: solo el segmento de tasas de entre 0 y 200.

Tasa por cada 100 mil habitantes	Cant. de municipios	Porcentaje % de 2,381 municipios.	Porcentaje % de 2,457 municipios.
De 0 a 10	560	23.5%	22.8%
De 10 a 20	176	7.4%	7.2%
De 20 a 300	186	7.8%	7.6%
De 30 a 40	191	8.0%	7.8%
De 40 a 50	172	7.2%	7.0%
De 50 a 60	171	7.2%	7.0%
De 60 a 70	148	6.2%	6.0%
De 70 a 80	126	5.3%	5.1%
De 80 a 90	101	4.2%	4.1%
De 90 a 100	96	4.0%	3.9%
De 100 a 110	78	3.3%	3.2%
De 110 a 120	90	3.8%	3.7%
De 120 a 130	61	2.6%	2.5%
De 130 a 140	46	1.9%	1.9%
De 1400 a 150	38	1.6%	1.5%
De 150 a 160	35	1.5%	1.4%
De 160 a 170	29	1.2%	1.2%
De 170 a 180	34	1.4%	1.4%
De 180 a 190	25	1.0%	1.0%
De 190 a 200	18	0.8%	0.7%
TOTAL	2,381	100.0%	96.9%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 129 muestra que, en 2020, casi la tercera parte ($22.8\% + 7.2\% = 30.0\%$) de los municipios tuvieron tasas menores a 20 defunciones por cada 100 mil habitantes.

La tabla 130 muestra que, en 2021, la gran mayoría (90.1%) de los municipios tuvieron tasas menores a 150 defunciones por cada 100 mil habitantes. Por ello, la tabla subsecuente presenta un análisis de ese subconjunto de municipios. Los cinco municipios con mayores tasas en este periodo fueron: Santa Magdalena Jicotlán, Oaxaca (2,061.9); Atizapán, Estado de México (987.0); Santo Domingo Tlatayápan, Oaxaca (704.2); La Trinidad Vista Hermosa, Oaxaca (383.1) y Manuel Benavides, Chihuahua (379.0).

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 130. Tasas de defunción por COVID-19 por municipio en el año 2021.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 150	2,213	90.1%
De 150 a 300	230	9.4%
De 300 a 450	11	0.4%
De 450 a 600	0	0.0%
De 600 a 750	1	0.0%
De 750 a 900	0	0.0%
De 900 a 1,050	1	0.0%
De 1,050 a 1,200	0	0.0%
De 1,200 a 1,350	0	0.0%
De 1,350 a 1,500	0	0.0%
De 1,500 a 1,650	0	0.0%
De 1,650 a 1,800	0	0.0%
De 1,800 a 1,950	0	0.0%
De 1,950 a 2100	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 131 muestra que, en 2021, la cuarta parte de los municipios (16.2% + 4.3% + 5.3% = 25.8%) tuvieron tasas menores a 30 defunciones por cada 100 mil habitantes.

Tabla 131. Tasas de defunción por COVID-19 por municipio en el año 2021: solo el segmento de tasas de entre 0 y 150.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,213 municipios	% de 2,457 municipios
De 0 a 10	397	17.9%	16.2%
De 10 a 20	105	4.7%	4.3%
De 20 a 30	131	5.9%	5.3%
De 30 a 40	154	7.0%	6.3%
De 40 a 50	179	8.1%	7.3%
De 50 a 60	182	8.2%	7.4%
De 60 a 70	181	8.2%	7.4%
De 70 a 80	155	7.0%	6.3%
De 80 a 90	137	6.2%	5.6%
De 90 a 100	140	6.3%	5.7%
De 100 a 110	143	6.5%	5.8%
De 110 a 120	104	4.7%	4.2%
De 120 a 130	83	3.8%	3.4%
De 130 a 140	70	3.2%	2.8%
De 140 a 150	52	2.3%	2.1%
TOTAL	2,213	100.0%	90.1%

Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019.

La tabla 132 muestra que, en 2022 (hasta el 17 de septiembre), la mayoría (89.3%) de los municipios tuvieron tasas menores a 30 defunciones por cada 100 mil habitantes.

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Por ello, su tabla posterior presenta un análisis enfocado en esos municipios. Los cinco municipios con mayores tasas en este periodo fueron: San Pedro Yucunama, Oaxaca (420.2); Abasolo, Coahuila (257.5); Magdalena Zahuatlán, Oaxaca (233.6); Santiago Tillo, Oaxaca (175.1) y Coronado, Chihuahua (164.1).

Tabla 132. Tasas de defunción por COVID-19 por municipio en el año 2022 (hasta 17/Sep).

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 30	2,193	89.3%
De 30 a 60	215	8.8%
De 60 a 90	30	1.2%
De 90 a 120	8	0.3%
De 120 a 150	3	0.1%
De 150 a 180	5	0.2%
De 180 a 210	0	0.0%
De 210 a 240	1	0.0%
De 240 a 270	1	0.0%
De 270 a 300	0	0.0%
De 300 a 330	0	0.0%
De 330 a 360	0	0.0%
De 360 a 390	0	0.0%
De 390 a 420	0	0.0%
De 420 a 450	1	0.0%
TOTAL	2,457	100.0%

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 133 muestra que, en 2022 (hasta septiembre 17), el 39.6% de los municipios tuvieron tasas menores a 2 defunciones por cada 100 mil habitantes.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 133. Tasas de defunción por COVID-19 por municipio en el año 2022 (hasta 17/Sep): solo el segmento entre 0 y 30.

Tasa por cada 100 mil habitantes	Cant. de municipios	Porcentaje % de 2,193 municipios	Porcentaje % de 2,457 municipios
De 0 a 2	974	44.4%	39.6%
De 2 a 4	78	3.6%	3.2%
De 4 a 6	122	5.6%	5.0%
De 6 a 8	118	5.4%	4.8%
De 8 a 10	115	5.2%	4.7%
De 10 a 12	123	5.6%	5.0%
De 12 a 14	98	4.5%	4.0%
De 14 a 16	109	5.0%	4.4%
De 16 a 18	90	4.1%	3.7%
De 18 a 20	95	4.3%	3.9%
De 20 a 22	67	3.1%	2.7%
De 22 a 24	79	3.6%	3.2%
De 24 a 26	48	2.2%	2.0%
De 26 a 28	38	1.7%	1.5%
De 28 a 30	39	1.8%	1.6%
TOTAL	2,193	100.0%	89.3%

Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019.

La tabla 134 muestra que, en todo el periodo analizado, desde 2020 hasta el 17 de septiembre de 2022, la gran mayoría de los municipios ($64.4\% + 19.0\% + 8.5\% = 91.9\%$) tuvieron tasas menores a las 6,000 defunciones por cada 100 mil habitantes. Por ello, la tabla 135 se enfoca en esos municipios. Los cinco municipios con mayores tasas en todo este periodo fueron: Atizapán, Estado de México (2,581.7); Santa Magdalena Jicotlán, Oaxaca (2,105.3); Azcapotzalco, CDMX (714.7); Iztacalco, CDMX (704.4) y Santo Domingo Tlatayápan, Oaxaca (704.2).

CAPITULO 5. PATRONES CRONOLÓGICOS EN EL NIVEL DE AGREGACIÓN MUNICIPAL

Tabla 134. Tasas de defunción por COVID-19 por municipio desde 2020 hasta 17/Sep/2022.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,457 municipios
De 0 a 2,000	1,583	64.4%
De 2,000 a 4,000	468	19.0%
De 4,000 a 6,000	208	8.5%
De 6,000 a 8,000	106	4.3%
De 8,000 a 10,000	45	1.8%
De 10,000 a 12,000	13	0.5%
De 12,000 a 14,000	9	0.4%
De 14,000 a 16,000	6	0.2%
De 16,000 a 18,000	7	0.3%
De 18,000 a 20,000	6	0.2%
De 20,000 a 22,000	0	0.0%
De 22,000 a 24,000	3	0.1%
De 24,000 a 26,000	1	0.0%
De 26,000 a 28,000	1	0.0%
De 28,000 a 30,000	1	0.0%
TOTAL	2,457	

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

Tabla 135. Tasas de defunción por COVID-19 por municipio desde 2020 hasta 17/Sep/2022: solo el segmento de tasas entre 0 y 6,000.

Tasa por cada 100 mil habitantes	Cant. de municipios	% de 2,259 municipios	% de 2,457 municipios
De 0 a 400	454	20.1%	18.5%
De 400 a 800	415	18.4%	16.9%
De 800 a 1,200	312	13.8%	12.7%
De 1,200 a 1,600	232	10.3%	9.4%
De 1,600 a 2,000	170	7.5%	6.9%
De 2,000 a 2,400	137	6.1%	5.6%
De 2,400 a 2,800	97	4.3%	3.9%
De 2,800 a 3,200	88	3.9%	3.6%
De 3,200 a 3,600	86	3.8%	3.5%
De 3,600 a 4,000	60	2.7%	2.4%
De 4,000 a 4,400	59	2.6%	2.4%
De 4,400 a 4,800	44	1.9%	1.8%
De 4,800 a 5,200	40	1.8%	1.6%
De 5,200 a 5,600	34	1.5%	1.4%
De 5,600 a 6,000	31	1.4%	1.3%
TOTAL		2,259	

Fuente: Elaboración propia con datos de SSA, 2022 y Conapo, 2019.

La tabla 135 muestra que, en todo el periodo de 2020 al 17 de septiembre de 2022, más de la tercera parte ($18.5\% + 16.9\% = 35.4\%$) de los municipios tuvieron tasas menores a las 800 defunciones por cada 100 mil habitantes.

5.4 Discusión de los resultados

Desde un punto de vista cronológico, los meses con más casos positivos dentro del periodo analizado fueron, en ese orden, enero y julio de 2022 y agosto de 2021; por otra parte, los meses con más defunciones fueron, en ese orden, enero de 2021 y julio y diciembre de 2020. De todo el periodo analizado (2020 a 17 de septiembre de 2022), el patrón más destacable en cuanto a los casos positivos es que el 43.0% de los municipios tuvieron menos de 100 casos cada uno. Cuatro alcaldías de la Ciudad de México (Iztapalapa, Álvaro Obregón, Gustavo A. Madero, y Tlalpan) y el municipio de San Luis Potosí, SLP en el estado homónimo, tuvieron las mayores cantidades de positivos del país. Respecto a tasas de positivos, el 35.4% de los municipios tuvieron tasas por debajo de los 800 casos por cada 100 mil habitantes. Los cinco municipios con las mayores tasas fueron alcaldías de la Ciudad de México: Álvaro Obregón, Milpa Alta, Tlalpan, Azcapotzalco y Tláhuac. Respecto a defunciones, el 30.9% de los municipios tuvieron menos de 6 defunciones cada uno. Los municipios con más defunciones fueron: Iztapalapa, CDMX; Puebla, Puebla; Ecatepec de Morelos, Estado de México; Gustavo A. Madero, CDMX y Guadalajara, Jalisco. En cuanto a tasas de defunción, lo más destacable es que 35.4% de los municipios tuvieron tasas menores a las 800 defunciones por cada 100 mil habitantes. Los cinco municipios con mayores tasas fueron: Atizapán, Estado de México; Santa Magdalena Jicotlán, Oaxaca; Azcapotzalco, CDMX; Iztacalco, CDMX y Santo Domingo Tlatayápam, Oaxaca. La mayoría de los resultados presentados en este capítulo son consistentes con los publicados por las fuentes gubernamentales de México (por ejemplo, SSA, 2022c) y diversas fuentes independientes, lo cual los convalida entre sí.

CAPÍTULO 6.

PATRONES GEOGRÁFICOS EN MAPAS DIGITALES

Este capítulo presenta una serie de mapas para la visualización de los patrones de tiempo durante la pandemia de COVID-19 en México. Los mapas abarcan desde inicios del año 2020 al 17 de septiembre de 2022 y se presentan con nivel de granularidad municipal. Se calcularon las tasas de casos positivos (clasificaciones tipo 1, 2 y 3, según SSA) y las tasas de defunción por cada 100 mil habitantes de cada municipio del país, con la base de datos de casos de COVID-19 de la Secretaría de Salud (SSA, 2022), del Gobierno Federal, y las estimaciones de población municipal para los años 2020, 2021 y 2022, producidas por el Consejo Nacional de Población (Conapo, 2019). Las tasas se calcularon para cada mes dentro del periodo mencionado. Los datos usados para producir estos mapas provienen de nuestro *dataset MuniCOVID-2022*, disponible para descarga gratuita en la web (ver la liga de descarga a inicios del capítulo. 5).

Considerando que en México esta pandemia presentó varias etapas de intensificación en su transmisión, se comenta más adelante el concepto de acmé y se abordan las denominadas *olas epidémicas* reconocidas por la SSA. Para facilitar la observación de los patrones geográficos de la pandemia a lo largo del tiempo, en las secciones siguientes se presentan dos tipos de mapas: de calor, y coropléticos. Los mapas de calor indican, mediante colores (o tonos de gris), la intensidad de casos en un área específica, identifican las áreas de mayor concentración y su evolución en el tiempo. Por otra parte, los mapas coropléticos muestran áreas coloreadas según el número total de casos o defunciones para identificar regiones con mayor incidencia. Ambos tipos de mapas son válidos y útiles dependiendo del contexto y objetivo del mapa. Si se intenta comparar la cantidad absoluta del fenómeno entre diferentes territorios, las frecuencias absolutas son más apropiadas. Si se busca resaltar diferencias proporcionales entre áreas, la comparación de tasas será más adecuada.

En los mapas de COVID-19 que se muestran a continuación, es conveniente utilizar tasas en lugar de frecuencias absolutas, ya que las tasas permiten comparar de manera

más equitativa la incidencia de la enfermedad entre diferentes regiones, independientemente de su tamaño de población. Al representar las tasas de casos positivos o de muertes por COVID-19 por cada 100,000 habitantes, se pueden identificar de manera más clara los puntos calientes o áreas con mayor impacto en relación con el tamaño de su población.

6.1 Acmé o pico de una etapa de intensificación de transmisión del SARS-CoV-2

La etapa de intensificación de transmisión del SARS-CoV-2 (conocida popularmente como “ola”) se identifica por el alto número de casos de COVID-19 de aparición rápida. El término técnico para describir el punto de máxima incidencia de la ola epidémica es *acmé* o *pico*. El pico de la epidemia constituye el periodo en el que el número de casos alcanza el punto máximo para después disminuir gradualmente. La propagación de la transmisión del virus precede al pico de la ola por aproximadamente 10 a 14 días (Zayet *et al.*, 2020). Esto se debe principalmente a dos situaciones: después de infectarse una persona, pasarán varios días para el inicio de los síntomas y, si la enfermedad es moderada o severa, pasarán otros días para que busque atención médica y ocurra el registro del caso.

Durante una pandemia, la intensificación de transmisión del SARS-CoV-2 en un país, estado, región y localidad es asincrónica. A nivel nacional, el pico de una ola representa el cúmulo de los picos que ocurren a nivel estatal o regional. En este sentido, cobra importancia representar en los mapas estas etapas de intensificación de la transmisión del SARS-CoV-2 a nivel municipal. Hasta septiembre de 2022, en México se observan cinco olas epidémicas. El acmé o pico de una ola epidémica podría delimitarse con base en alguna fecha específica, alguna semana epidemiológica, o bien, algún mes en particular. En nuestro caso, decidimos hacerlo con base en meses específicos, como se enumera a continuación:

- i. Primera ola: julio de 2020
- ii. Segunda ola: enero de 2021
- iii. Tercera ola: agosto de 2021
- iv. Cuarta ola: enero de 2022
- v. Quinta ola: julio de 2022

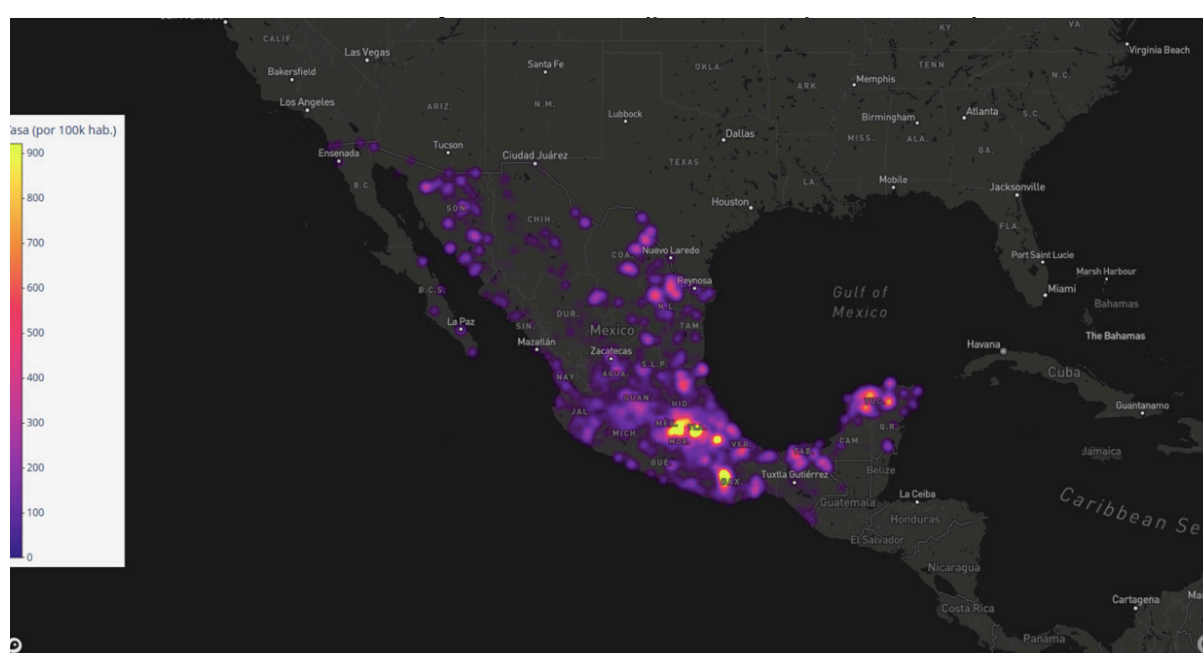
En las siguientes secciones, se presentan mapas para cada una de las cinco olas epidémicas observadas en México entre inicios del año 2020 y el 17 de septiembre de 2022 (fecha de descarga de la base de datos de SSA). Primeramente, se muestran cinco mapas de calor y, posteriormente, cinco coropléticos. Todos los mapas presentados en este capítulo fueron producidos con datos de nuestro *dataset MuniCOVID-2022*, creado a partir de (SSA, 2022). Se usó el *package geopandas*, en el lenguaje de programación Python. Hemos puesto estos mapas en la web¹⁷ para descargarse gratuitamente y usarse en trabajos de investigación o cualquier otro propósito sin fines de lucro (citando este libro). Están disponibles en archivos de tipo *jpg* y también *tiff*, con resolución de 300 puntos por pulgada (ppp).

17 Disponibles en <https://1drv.ms/f/s!At6TY8A0mTpNhF0ZfV52JHNonn4Cdw?e=chfrHq>

6.2 Mapas de calor

El análisis de los mapas de calor de casos positivos y de defunciones por COVID-19 durante diferentes periodos revela patrones significativos en la concentración de contagios y fallecimientos (representada mediante tasas por cada 100,000 habitantes) en diversos estados. Esto ofrece valiosa información para comprender la dinámica de la pandemia. A continuación, se presentan diez mapas de calor en las figuras 10 a 19. La figura 10 muestra que, durante julio de 2020, se identificó una alta concentración de casos positivos en los estados de México, Ciudad de México, Oaxaca, Tlaxcala, Veracruz y Yucatán.

Figura 10. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).

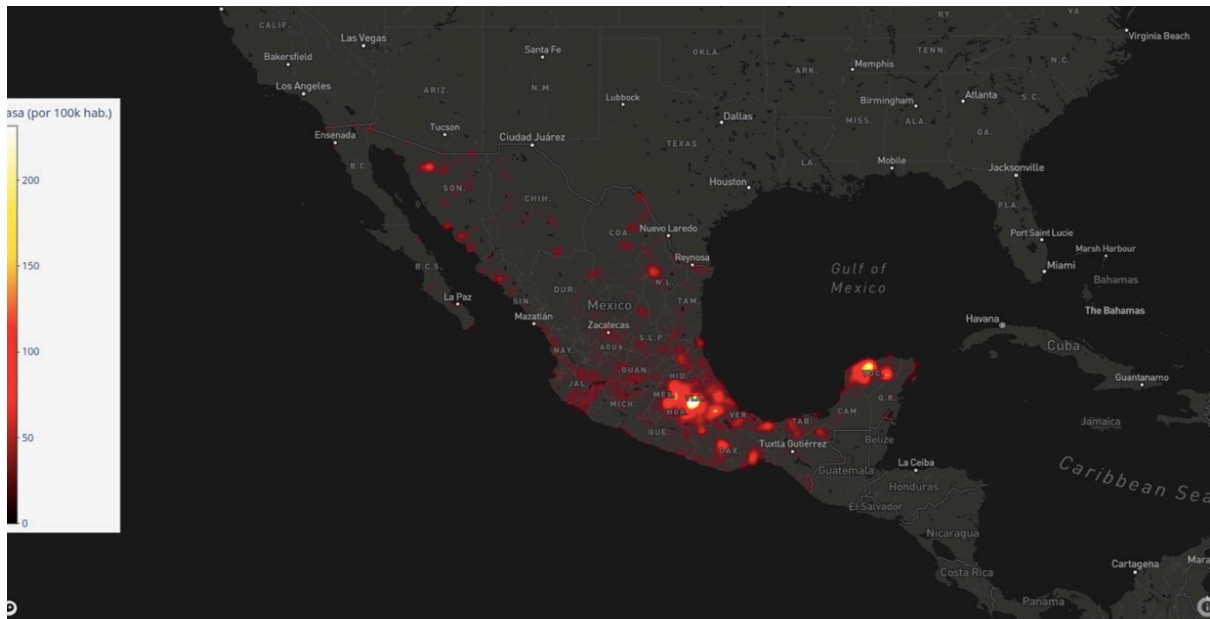


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019 usando la API de Mapbox en lenguaje de programación Python.

En términos de defunciones (ver figura 11), los estados federativos con mayor concentración fueron Estado de México, Ciudad de México, Yucatán y Tlaxcala.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

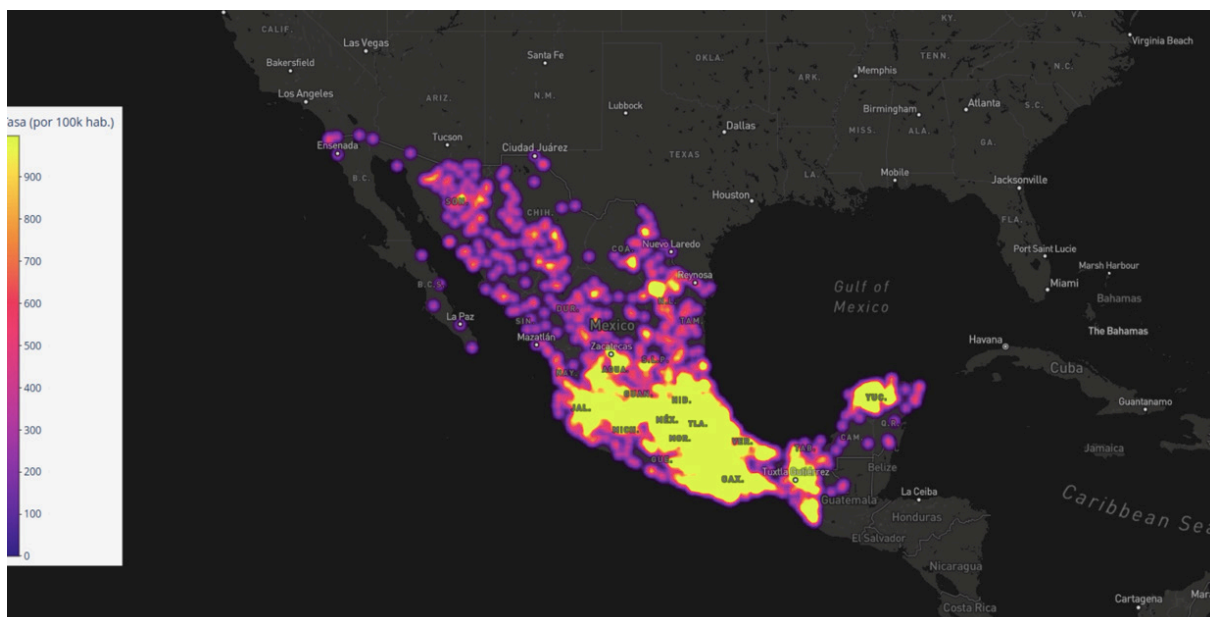
Figura 11. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

En enero de 2021, se observó un notorio incremento en la concentración de casos positivos (ver figura 12), extendiéndose además de los estados mencionados anteriormente, hacia Chiapas, Sonora, Hidalgo, Morelos, Michoacán, Colima, Jalisco, Querétaro, Guanajuato, Aguascalientes y Zacatecas.

Figura 12. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).

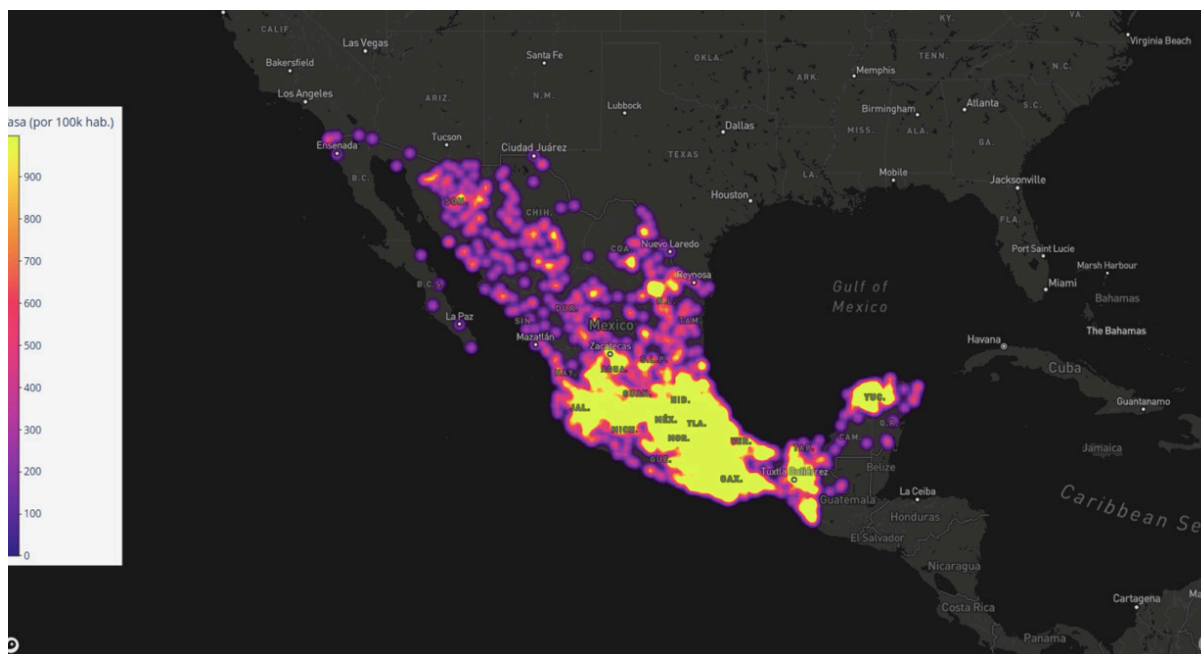


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

CAPITULO 6. PATRONES GEOGRÁFICOS EN MAPAS DIGITALES

En cuanto a defunciones (ver figura 13), se registró un aumento en la concentración en los estados de México, Ciudad de México, Veracruz, Oaxaca, Jalisco, Hidalgo y Yucatán.

Figura 13. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).

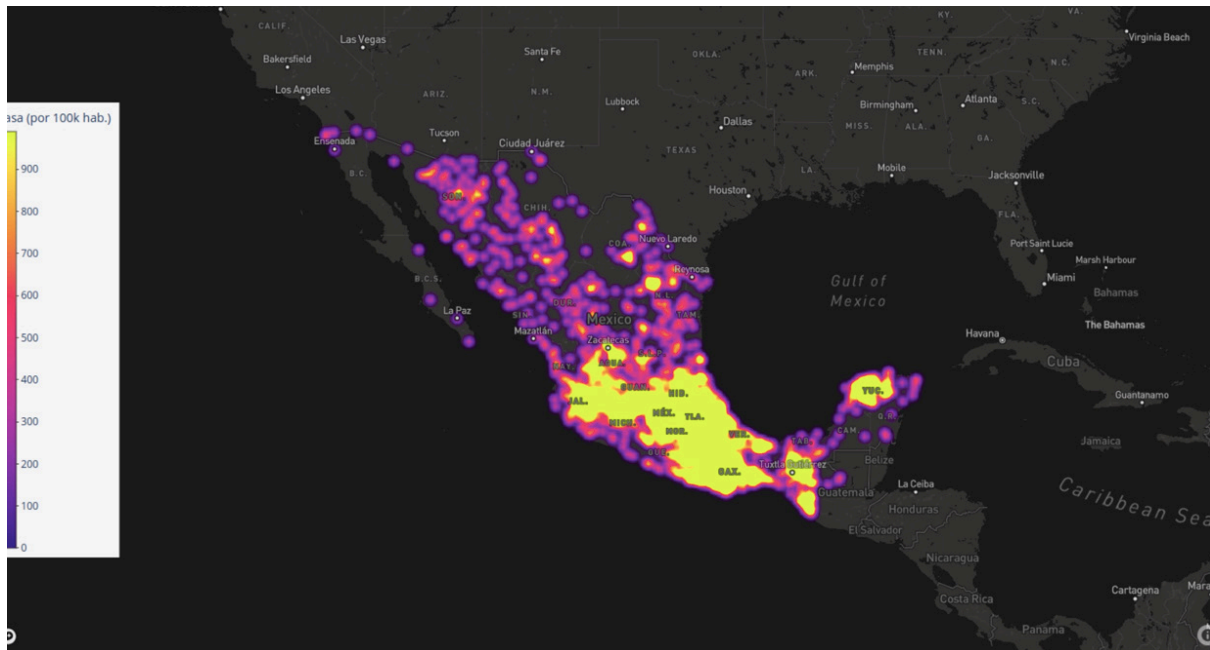


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

También, los mapas reflejan la virulencia de las diferentes cepas de SARS-CoV-2, como es el caso de la variante Delta, como ocurrió que se propagó en México alrededor de agosto de 2021. En ese mes, se observó una ligera reducción en la concentración de casos positivos (ver figura 14) y defunciones (ver figura 15) en ciertos municipios de los estados de Nuevo León, Colima, Guerrero y Tabasco. A pesar de esta tendencia, aún existían áreas con una mayor concentración de defunciones en los estados de México, Ciudad de México, Veracruz, Yucatán y Oaxaca.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Figura 14. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

Figura 15. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).



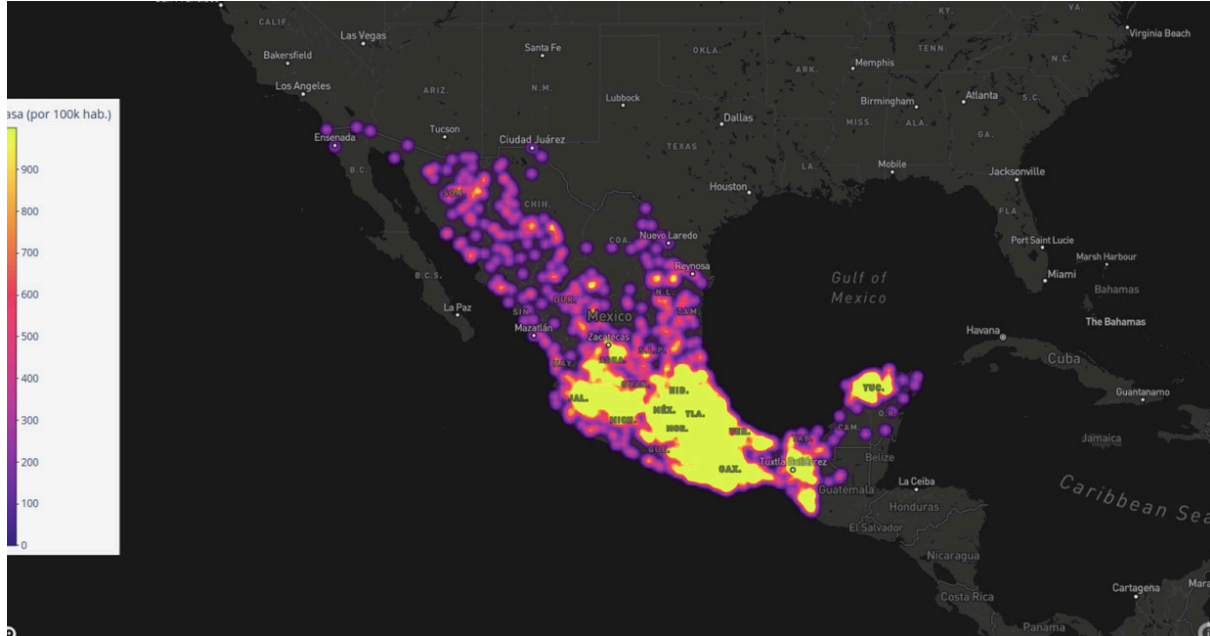
Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

En la cuarta etapa de intensificación de transmisión del SARS-CoV-2, predominó la variante *Ómicron*, una cepa poco virulenta. En enero de 2022, en general, se observó una

CAPITULO 6. PATRONES GEOGRÁFICOS EN MAPAS DIGITALES

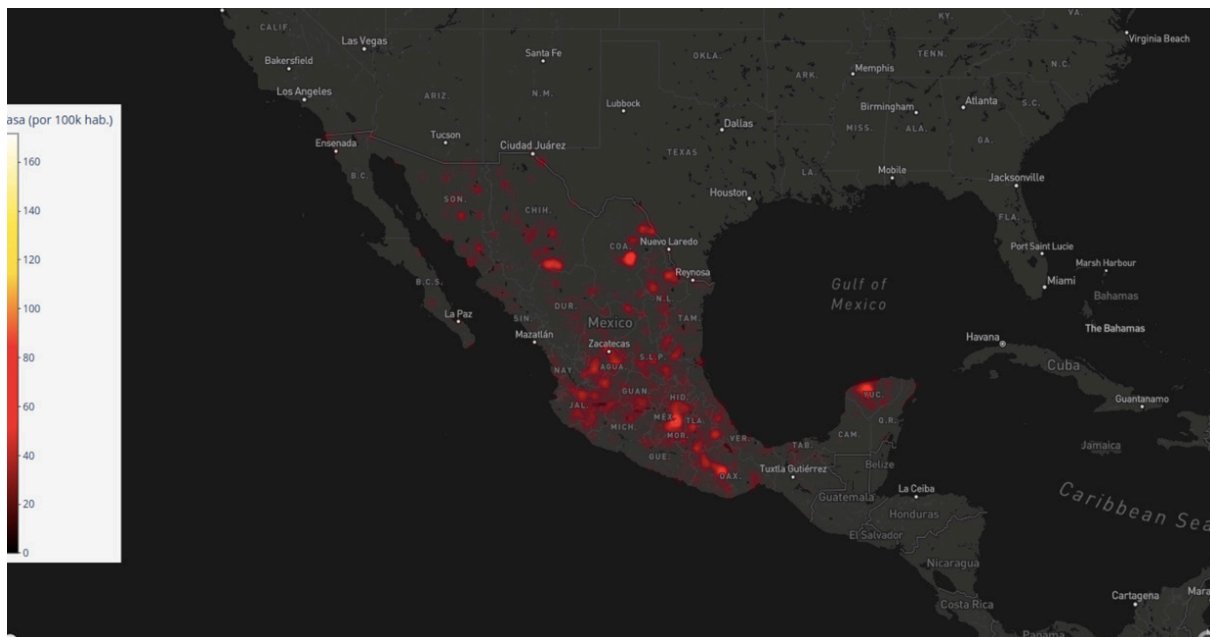
reducción en la concentración de casos positivos (ver figura 16) y defunciones (ver figura 17) en los estados de Coahuila, Sonora, Jalisco, Guanajuato, Tabasco y Oaxaca.

Figura 16. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

Figura 17. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).

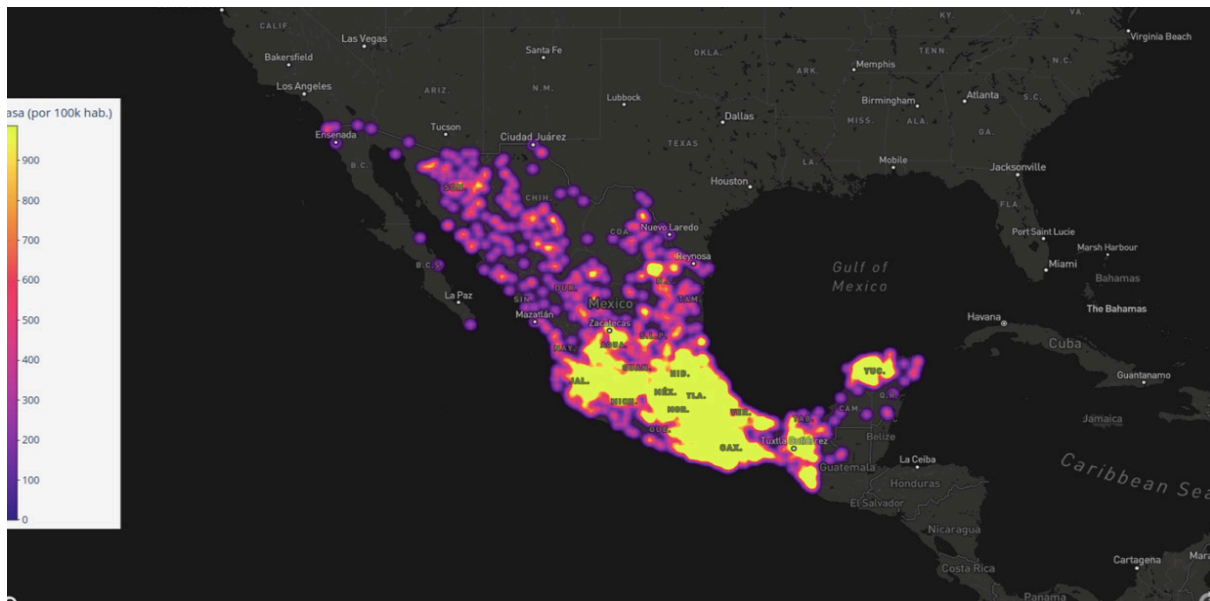


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

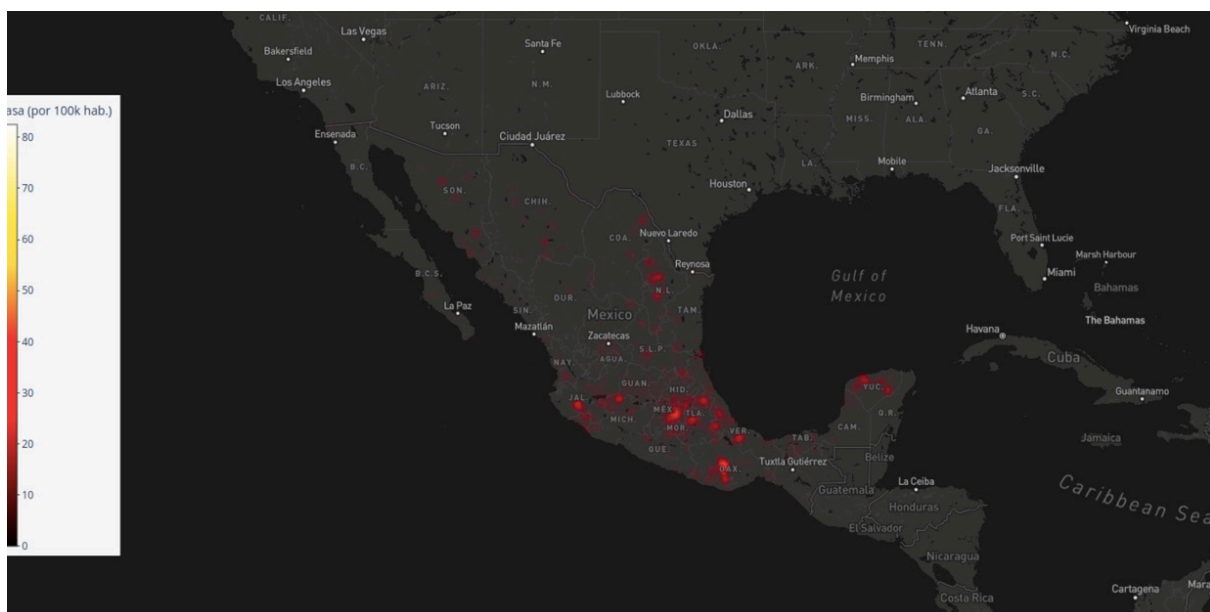
En julio de 2022, se destacó una significativa disminución en la concentración de casos positivos (ver figura 18) y defunciones (ver figura 19) en general. No obstante, las entidades Estado de México, Ciudad de México, Veracruz, Oaxaca, Jalisco, Nuevo León y Yucatán fueron identificadas como áreas con mayor concentración de casos y defunciones durante este periodo.

Figura 18. Mapa de calor de tasas municipales de positivos por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

Figura 19. Mapa de calor de tasas municipales de defunción por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).

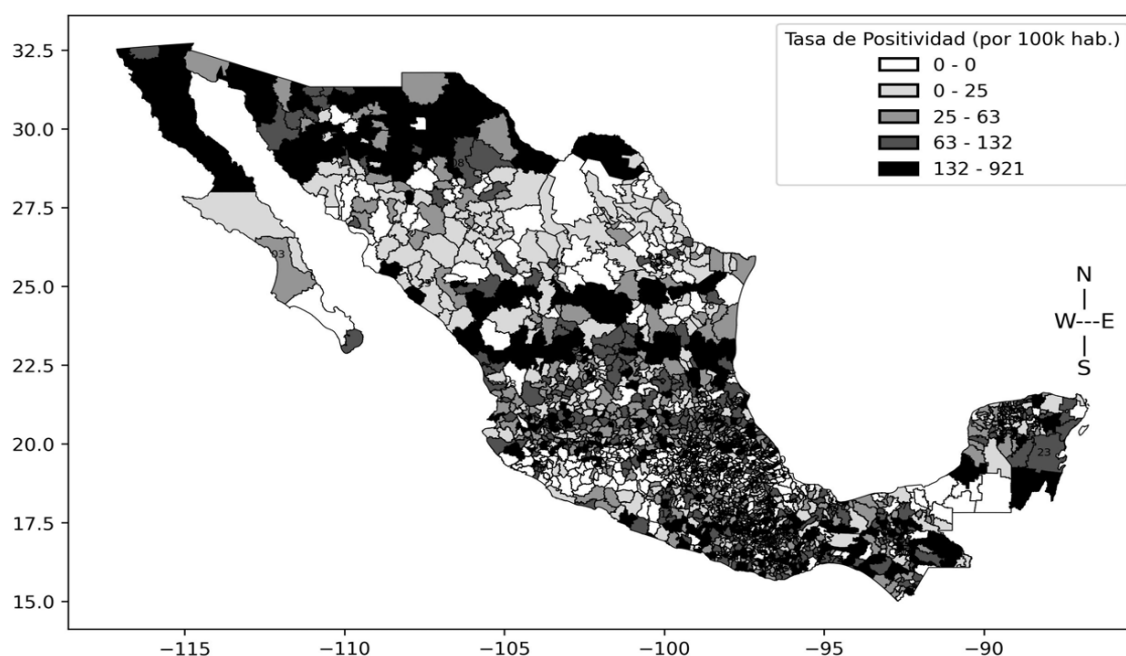


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la API de Mapbox en lenguaje de programación Python.

6.3 Mapas coropléticos

En esta sección, los mapas coropléticos muestran las tasas de positivos y de defunciones por cada 100,000 habitantes. Los resultados revelan importantes patrones y tendencias que son potencialmente útiles para la toma de decisiones en salud pública. Se presentan a continuación diez mapas coropléticos: las figuras 20 a 29 corresponden a este tipo de mapas. Iniciando con la figura 20, la evaluación de los datos de julio de 2020 muestra una marcada variabilidad en la tasa de positivos, reflejando que 568 municipios no reportaron casos positivos. Sin embargo, es destacable que 493 municipios presentaron cantidades similares de positivos en los intervalos más altos, lo cual podría indicar una propagación generalizada del contagio en estas áreas. Posiblemente debido al gran número de población susceptible, sin contacto previo a SARS-CoV-2.

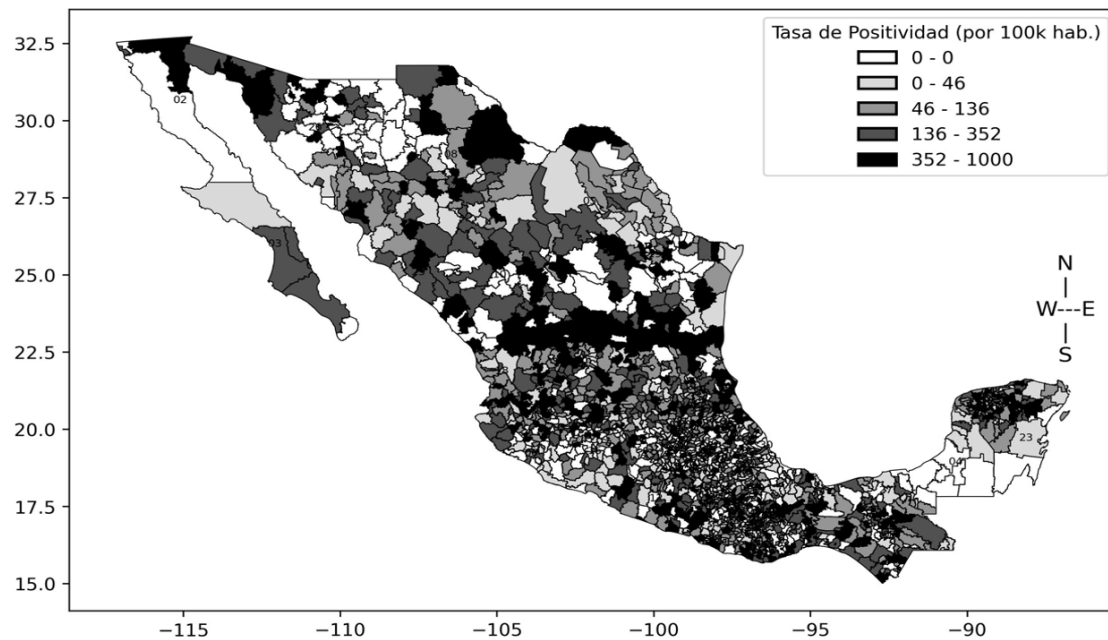
Figura 20. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

Al observar los datos de enero de 2021 (ver figura 21), es evidente que la situación cambió en comparación con el año anterior. Se registró un aumento en la cantidad de municipios sin positivos y una disminución en aquellos con tasas elevadas, señalando un posible impacto de las medidas de control implementadas a nivel nacional o regional. No obstante, se mantiene una presencia significativa de positivos en un grupo de municipios.

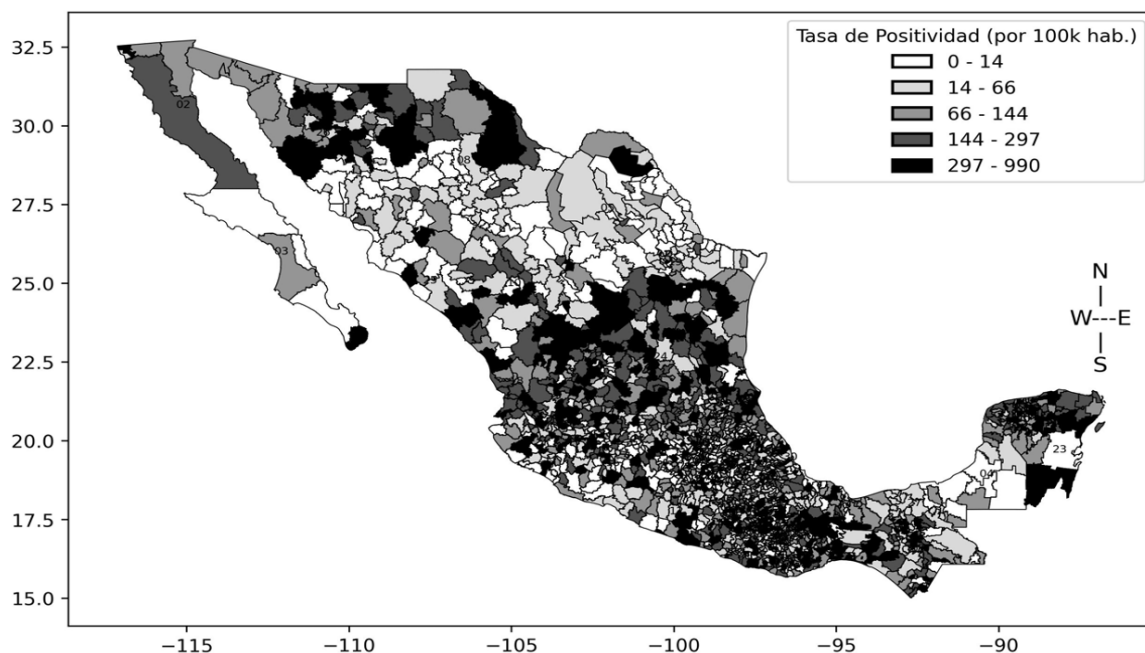
Figura 21. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

El análisis de agosto de 2021 (ver figura 22) revela una tendencia similar a la observada en enero del mismo año: aunque una mayor cantidad de municipios no reportaron positivos, se mantuvieron áreas con tasas considerables.

Figura 22. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).

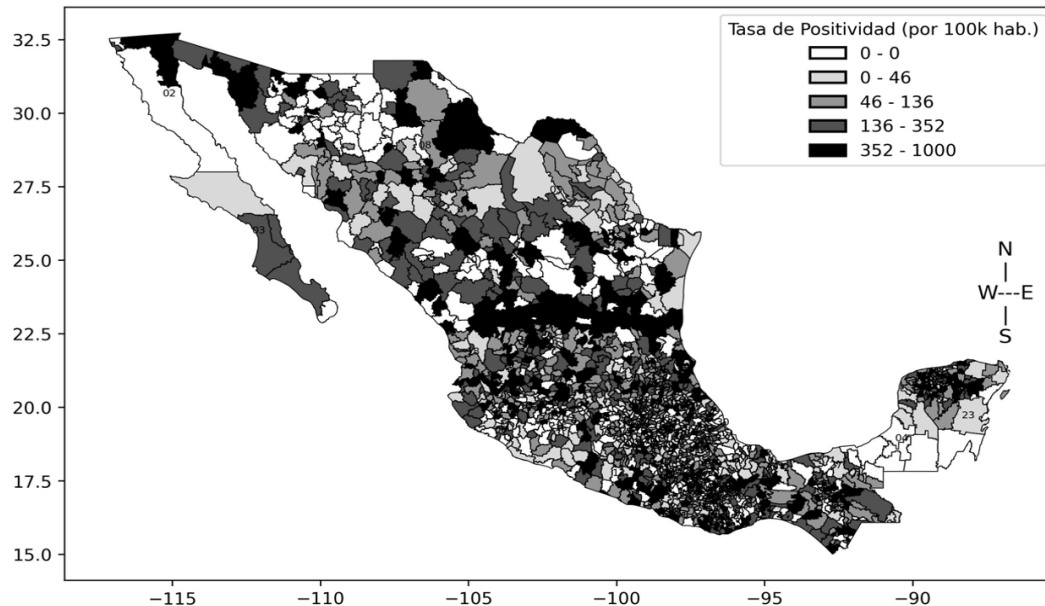


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

CAPITULO 6. PATRONES GEOGRÁFICOS EN MAPAS DIGITALES

Al acercarnos a enero de 2022 (ver figura 23), los resultados muestran una vuelta a la presencia de positivos en una mayor cantidad de municipios, lo que puede ser una señal de un posible rebrote, relajación de medidas preventivas, o mayor registro de casos por más búsqueda de atención médica.

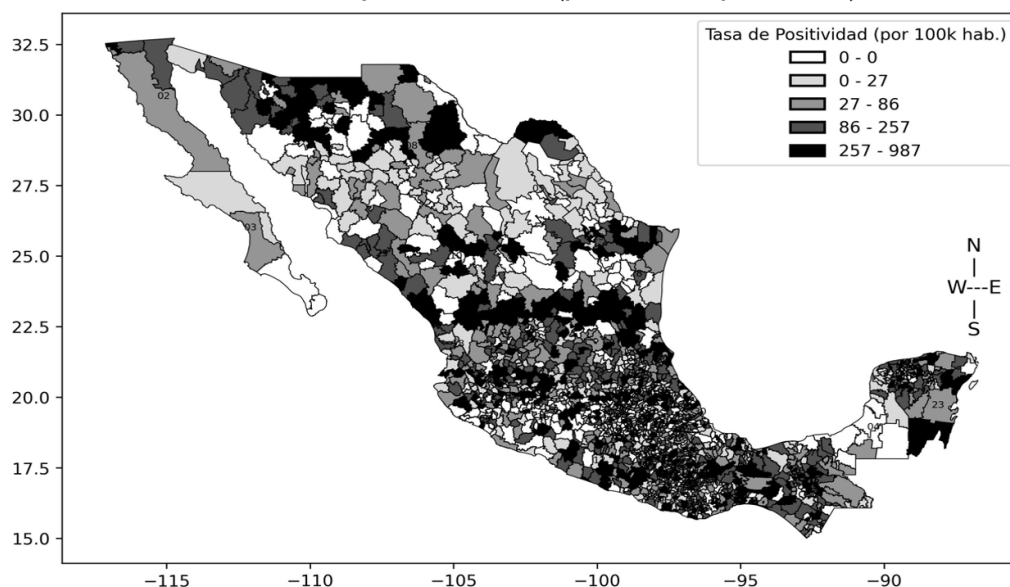
Figura 23. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

El análisis de julio de 2022 (ver figura 24) resalta un retorno a la presencia de positivos en un grupo importante de municipios.

Figura 24. Mapa coroplético de tasas municipales de positivos por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).

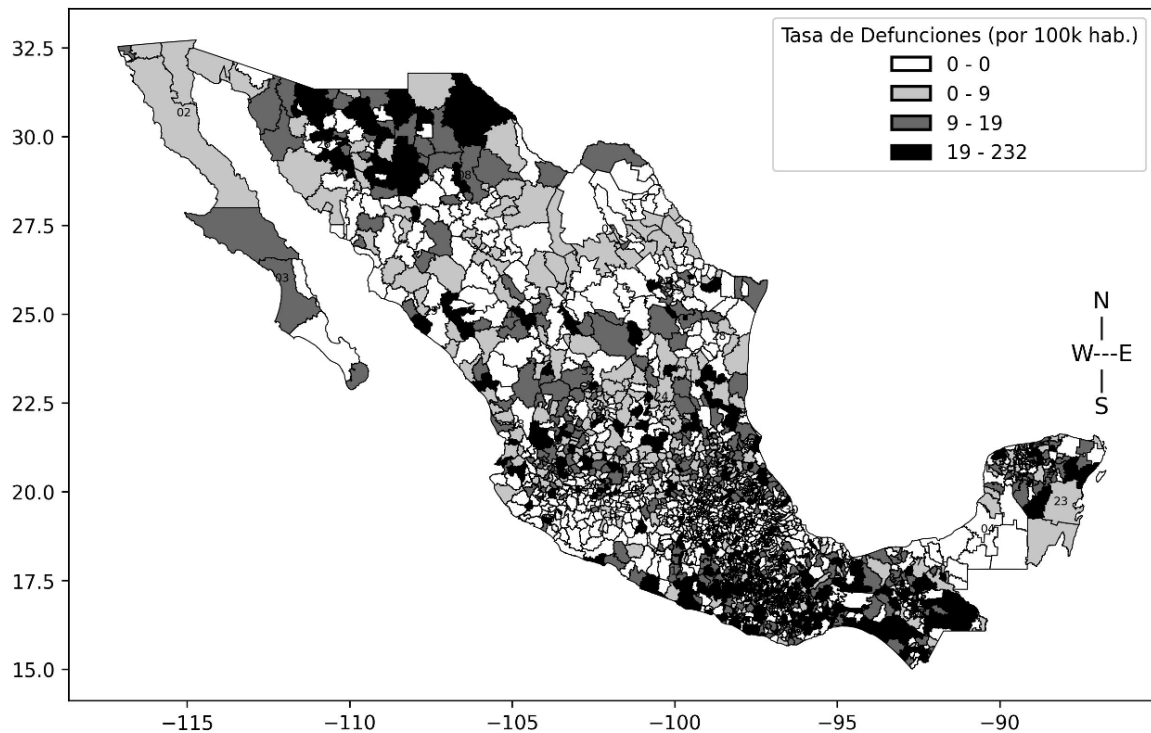


Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Por su parte, las tasas de defunción (ver las figuras 25 a 29) son esenciales para comprender la evolución de la mortalidad y su impacto en la salud pública. En julio de 2020 (figura 25) se observa el pico de la primera ola; por ello, despierta un interés especial. Se detectan 499 municipios con tasa mayor a 19.0 defunciones por cada 100,000 habitantes. En contraste, 1,095 municipios presentan una tasa igual a 0.0.

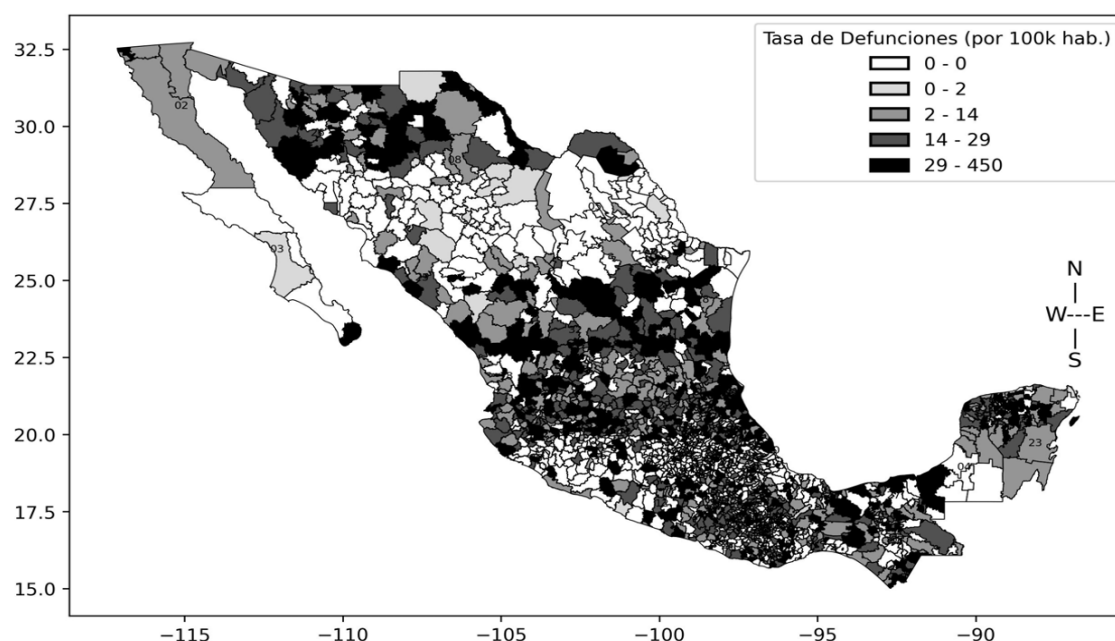
Figura 25. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en julio de 2020 (pico de la primera ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

Enero de 2021 (ver figura 26) despierta un interés especial porque forma parte del pico de la segunda ola, cuando ocurrieron más defunciones en México. Se observan 498 municipios con tasas superiores a 29 defunciones por cada 100,000 habitantes; por otra parte, se observan 954 municipios que tienen tasa 0 (cero).

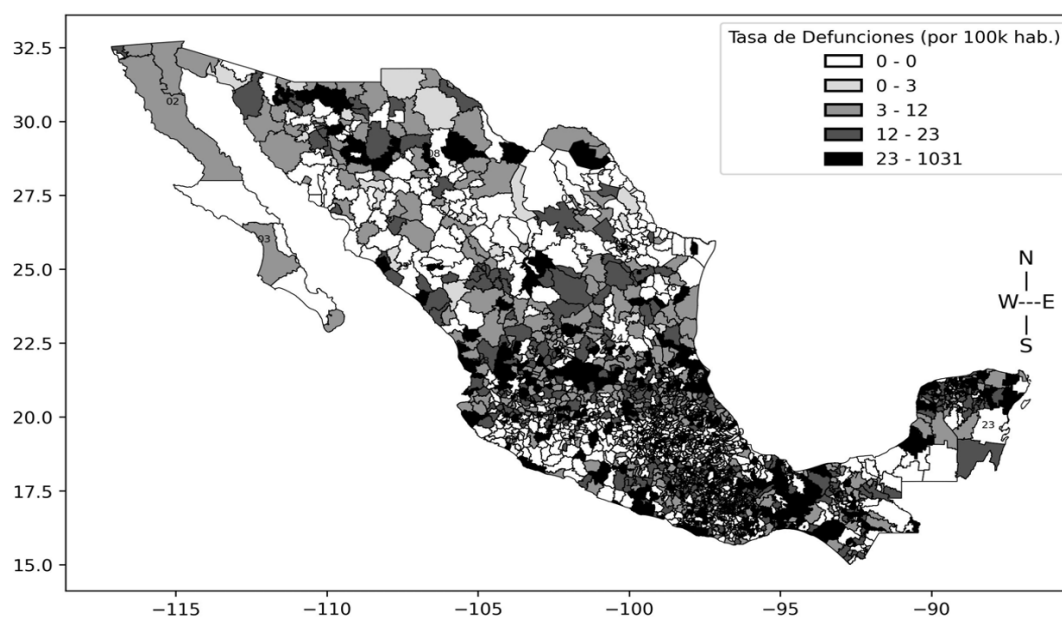
Figura 26. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en enero de 2021 (pico de la segunda ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

En agosto de 2021 (ver figura 27), como parte del pico de la tercera ola, se detectan 495 municipios con tasa de más de 23 defunciones por cada 100,000 habitantes. En cambio, se observan 951 con tasa igual a 0 (cero).

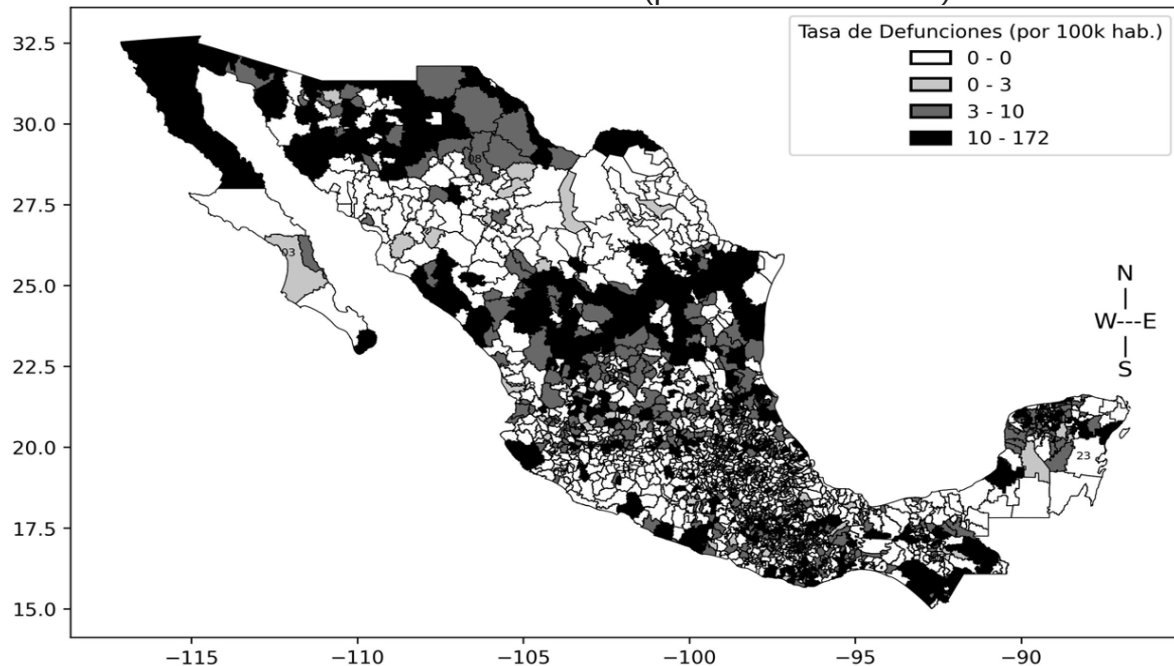
Figura 27. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en agosto de 2021 (pico de la tercera ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

En enero de 2022 (ver. figura 28), los datos revelaron que hubo un aumento significativo en la cantidad de municipios sin reporte de defunciones, alcanzando un total de 1,376. Este aumento puede indicar una mejora en la situación de mortalidad en estas áreas. No obstante, es preocupante que 109 municipios aún presentaran tasas de defunciones en un rango bajo, mientras que 488 municipios reflejaron tasas en los intervalos de 2.61 a 10.10, y de 10.11 a 171.70. Estos hallazgos resaltan la persistencia de desigualdades en la mortalidad.

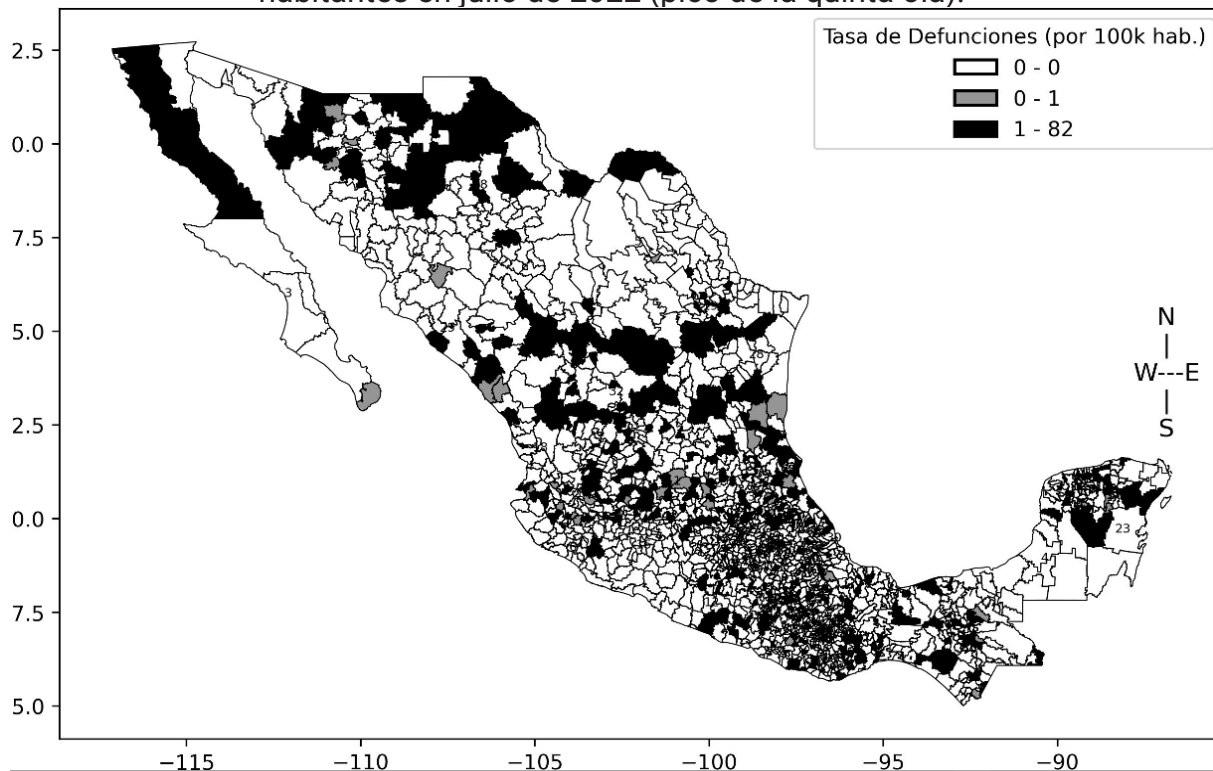
Figura 28. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en enero de 2022 (pico de la cuarta ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

A medida que se avanza hacia julio de 2022 (ver figura 29), el análisis muestra que la cantidad de municipios sin reporte de defunciones se redujo a 1,924. Mientras tanto, 50 municipios presentaron tasas de defunciones en un rango bajo, y 491 municipios reflejaron tasas en los intervalos de 1.01 a 82.50. Estos resultados demuestran que, a pesar de la disminución general de la mortalidad, aún existen áreas que enfrentan desafíos en el control de la propagación del coronavirus y la mortalidad asociada. El año 2022 parece estar caracterizado por un patrón de mejora general en la situación de mortalidad en varios municipios, con una disminución en la cantidad de defunciones reportadas y una mayor cantidad de municipios sin fallecimientos registrados. Sin embargo, es importante destacar que persisten áreas con tasas de defunciones más elevadas. La reflexión sobre estos datos pasados es de vital importancia para aprender de las experiencias y mejorar la toma de decisiones en salud pública ante situaciones futuras. La comprensión de las disparidades en la mortalidad entre diferentes regiones nos insta a abordar de manera urgente la equidad en el acceso a la atención médica, asegurando que todas las comunidades reciban una atención adecuada y oportuna.

Figura 29. Mapa coroplético de tasas municipales de defunción por cada 100,000 habitantes en julio de 2022 (pico de la quinta ola).



Fuente: Elaboración propia con datos de SSA, 2022, y Conapo, 2019, usando la biblioteca geopandas en lenguaje de programación Python.

6.4 Mapas de tipo gif animado

Para facilitar la visualización de la evolución de la pandemia sobre el territorio de México, se ofrecen al lector cuatro mapas de tipo gif animado, con granularidad geográfica municipal y granularidad cronológica por mes. Los hemos colocado en la web¹⁸ para descargarse gratuitamente (pueden usarse, citando este libro). Son dos mapas de calor y dos coropléticos. En cada uno de estos dos pares, un mapa muestra tasas de positivos y otro de defunciones por cada 100,000 habitantes. Cada uno de los cuatro mapas muestra los picos de las cinco olas epidémicas

18 Disponible en <https://1drv.ms/f/s!At6TY8A0mTpNhf0cy7zi-AVm-DYxRg?e=j91NQr>

CAPÍTULO 7. NIVELES DE OCURRENCIA DE COVID-19 CON RELACIÓN A LOS PERFILES SOCIODEMOGRÁFICOS DE LOS MUNICIPIOS

Este capítulo presenta los patrones más destacables, de tipo multivariable, que se observan en los más de 2,400 municipios de México, respecto a sus niveles de presencia de COVID-19 y su asociación estadística con las características sociodemográficas. Este capítulo es similar al capítulo 4 en el sentido de que los dos presentan resultados de modelos de árboles clasificadores producidos con el algoritmo C5.0 (Quinlan, 1993), con uso del *package* C50 (Kuhn and Quinlan, 2022) en lenguaje de programación R (R Core Team, 2019). El capítulo 4 aborda la modelación de casos individuales; en contraste, este capítulo se enfoca en el nivel de agregación municipal. Se sugiere al lector explorar el capítulo 4 para lograr mayor comprensión de este capítulo y revisar el capítulo 1 (diseño metodológico).

Los datos usados para producir los modelos presentados en este capítulo provienen de dos datasets independientes: *MuniCOVID-2022* (ver capítulo 5) y *MuniciMex-2020* (Coria et al., 2022, págs. 182-183). El *dataset* *MuniciMex-2020* es resultado del procesamiento de los datos del Censo Nacional de Población y Vivienda 2020 (INEGI, 2021), con agregación municipal para calcular porcentajes de habitantes y porcentajes de viviendas en relación con diversas características. Por ejemplo, contiene porcentajes de habitantes de cada municipio por sexo, intervalos de edad, nivel educativo, entre otros y porcentajes de viviendas con drenaje, electricidad, internet, etcétera. Los archivos *MuniCOVID-2022* y *MuniciMex-2020* fueron combinados con una operación *join* de base de datos aplicada a un atributo que concatena el identificador de entidad federativa y el identificador de municipio, ambos establecidos por el INEGI. El *dataset* resultante está disponible para descargarse de la Nube¹⁹. La producción de los árboles clasificadores de este capítulo usa datos de porcentajes de habitantes y de viviendas y descarta el uso de frecuencias absolutas, para hacer comparables las características sociodemográficas de los municipios que tienen distinto tamaño poblacional.

19 Disponible en <https://1drv.ms/u/s!At6TY8A0mTpNgtsM3Cy57R7QxnkpLg?e=QESfJL>

La tabla 136 presenta el panorama general de los modelos producidos. La columna denominada *Atributo target* se refiere al atributo del *dataset* que fue designado como atributo de clase para descubrir los patrones. Todos estos modelos están disponibles para ser descargados gratuitamente de la Nube²⁰. Todos los *targets* en la tabla son nombres de intervalos de tasas municipales por cada 100 mil habitantes; por ejemplo, el nombre *t_0_a_100*, representa al intervalo de valores de tasas entre 0 y 100 casos (ya sea de casos positivos o de defunciones) por cada 100 mil habitantes. Respecto a los intervalos de valores de las diversas tasas, se sugiere ver el capítulo 5 (patrones cronológicos por municipio), que presenta una serie de análisis de Pareto de estas. Estos análisis muestran los conjuntos de intervalos definidos para analizar cada una de las tasas y permiten identificar cuáles y cuántas son las clases mayoritarias (con porcentaje acumulado mayor o igual a 80.0%) de cada una de las tasas.

Es necesario aclarar que, para producir los modelos de árboles clasificadores en el presente capítulo, en algunos de ellos se manejaron intervalos de valores de tasas distintos a los mostrados en el capítulo 5. El motivo es que algunos de los intervalos usados en aquel capítulo contenían muchos municipios y esto no permitía distinguir las diferencias sociodemográficas correspondientes a los municipios incluidos en los subintervalos contenidos en aquellos. Por ello, esos intervalos fueron partidos en dos o más subintervalos. Entonces, las cantidades de clases mayoritarias de las tasas de los *targets* que se mencionan en el presente capítulo refieren a los intervalos de valores usados en este mismo, y no al capítulo 5.

La representación de las tasas como valores nominales permite la producción de modelos de árbol clasificador, donde el *target* tiene que ser, necesariamente, de tipo nominal. Al manejar la tasa en forma nominal, los modelos producidos permiten observar en qué se parecen los municipios que tienen tasas altas, medias o bajas, ya sea de casos positivos o de defunciones.

Los nombres de los *targets* obedecen a la nomenclatura que se describe a continuación. Todos estos nombres contienen el sufijo *tasa_100mil_INTERV*, indicando que son tasas municipales por cada 100,000 habitantes. El prefijo *posit* indica que el *target* es una tasa de casos positivos. El prefijo *def* indica que el *target* es una tasa de defunciones. Los valores 2020, 2021 y 2022 indican el año al que pertenece cada tasa. En la tabla 136 se muestran 8 *targets*, de los que se presentan 14 modelos (de algunos *targets* se presenta más de un modelo). Cada uno de los *targets* maneja un conjunto de valores que son nombres de intervalos. La cantidad y nomenclatura de estos nombres están determinados, principalmente, por el valor numérico máximo de la tasa correspondiente. Por ejemplo, en el *target* llamado *posit_2020_tasa_100mil_INTERV*, el valor numérico máximo de la tasa que le dio origen es 5,467.4, por lo que el nombre de su intervalo más alto es *t_2100_a_5700*. En contraste, en el *target* *posit_2021_tasa_100mil_INTERV*, el valor máximo de la tasa que lo originó es 14,250.2, por lo que su intervalo más alto se llama *t_4000_a_15000*. Para cada uno de los *targets* se produjo más de un modelo (no se muestran todos los modelos producidos); sus diversos modelos difieren en el *dataset* usado o en el subconjunto de atributos de un mismo *dataset* que fueron seleccionados como predictores.

20 Disponibles en <https://1drv.ms/f/s!At6TY8A0mTpNgtZLLIB9UhwxbahDvQ?e=orCKgr>

LA PANDEMIA DE COVID-19 EN MÉXICO: PERSPECTIVAS
DESDE LA COMPUTACIÓN, LA SALUD Y LO SOCIAL

Tabla 136. Resumen de los modelos de árboles clasificadores producidos con datos de COVID-19 y datos sociodemográficos con nivel de agregación municipal, usando conjuntamente los *datasets* *MuniCOVID-2022* y *MuniciMex-2020* (MuniCovid_2022_y_municimex_2020_2023_06_20_b2).

No.	Atributo target	Nombre del modelo	Predictores importantes descartados manualmente	Cant. de predictores	Accu-racy	Kappa	Cant. de reglas	Atributo predictor ubicado en la raíz
1	posit_2020_tasa_100mil_INTERV	posit_2020_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.2403	0.1697	769	vph_pc_percent
2	posit_2021_tasa_100mil_INTERV	posit_2021_tasa_100k_2023_06_27_b	Tasas COVID de 2021, 2022 y global	155	0.1981	0.1399	878	posit_2020_tasa_100mil
3	posit_2021_tasa_100mil_INTERV	posit_2021_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.1964	0.1375	852	vph_pc_percent
4	posit_2022_a_sep_17_tasa_100mil_INTERV	posit_2022_a_sep_17_tasa_100k_2023_07_04	Tasas COVID de 2022 y global	157	0.3052	0.1923	676	posit_2021_tasa_100mil
5	posit_2022_a_sep_17_tasa_100mil_INTERV	posit_2022_a_sep_17_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.2955	0.1811	744	grapoes_f_un_dec
6	posit_MPAL_TOTAL_tasa_100mil_INTERV	posit_MPAL_TOTAL_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.2549	0.1699	735	vph_pc_percent
7	def_2020_tasa_100mil_INTERV	def_2020_tasa_100k_2023_06_26	Tasas COVID de 2021, 2022 y global	160	0.2679	0.1961	824	posit_2020_tasa_100mil
8	def_2020_tasa_100mil_INTERV	def_2020_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.2208	0.1449	865	grapoes_un_dec
9	def_2021_tasa_100mil_INTERV	def_2021_tasa_100k_2023_06_26_g	Tasas COVID de 2022 y global	160	0.1948	0.1302	855	posit_2021_tasa_100mil
10	def_2021_tasa_100mil_INTERV	def_2021_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.1834	0.1184	865	vph_pc_percent
11	def_2022_a_sep_17_tasa_100mil_INTERV	def_2022_a_sep_17_tasa_100k_2023_06_26	Tasas COVID de 2022 y global	159	0.3409	0.1973	662	def_2020_tasa_100mil
12	def_2022_a_sep_17_tasa_100mil_INTERV	def_2022_a_sep_17_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.2744	0.1347	710	vph_cvj_percent
13	def_MPAL_TOTAL_tasa_100mil_INTERV	def_MPAL_TOTAL_tasa_100k_2023_06_26	*****	161	0.5568	0.5181	334	def_2021_tasa_100mil
14	def_MPAL_TOTAL_tasa_100mil_INTERV	def_MPAL_TOTAL_tasa_100k_2023_06_27	Todas las tasas de COVID	153	0.1867	0.1197	883	vph_pc_percent

Fuente: Elaboración propia con datos de SSA, 2022, y de INEGI, 2021, usando el software R (R Core Team, 2019) con el package C50 (Kuhn and Quinlan, 2022).

La columna llamada *Predictores importantes descartados manualmente*, existente en casi todos los modelos producidos, contiene algunos atributos predictores descartados intencionalmente de modo manual antes de generar los modelos. Esta acción tiene el propósito de observar y contrastar el efecto de la ausencia o presencia de los atributos mencionados en la generación de los modelos. Algunos de ellos pueden producir modelos con mayores o menores valores de *accuracy*, *Kappa* o cantidad de reglas clasificadoras. Estas variaciones resultan, en algunos casos, de posibles relaciones causales entre el *target* y algunos de sus predictores. *Cant. de predictores* se refiere a la cantidad de atributos del *dataset* que fueron seleccionados como predictores para el *target* especificado.

La columna *Accuracy (precisión)* es una medida estadística básica para determinar el porcentaje de aciertos logrados por el modelo al ser probado con un subconjunto de datos del mismo *dataset* con que fue creado. La columna *Kappa* (Cohen, 1960) muestra una estimación ajustada de *accuracy*; toma en cuenta el posible sesgo generado por la influencia de los diversos porcentajes de valores del *target*. Es decir, si el *target* en el *dataset* para producir un modelo tiene un alto porcentaje de un solo valor (por ejemplo 80% de casos con valor de *t_0_a_100*), el modelo producido no tiene mucho mérito al hacer clasificaciones automáticas. La columna *Cantidad de reglas* indica cuántas ramas (o reglas clasificadoras) contiene el árbol. La columna *Raíz* indica cuál es el atributo que ocupa el nodo raíz del árbol y, por tanto, es el que tiene la mayor correlación estadística con el *target*, sugiriendo una potencial (preliminar y sujeta a verificación) relación de causa-efecto.

Al igual que en los modelos presentados en el capítulo que abordó los casos individuales de COVID-19, los criterios de aceptación de los modelos presentados en la tabla 136, abordando el nivel de agregación municipal, también son: *accuracy* $\geq 75\%$, *Kappa* ≥ 0.67 y cantidad de clases mayoritarias mayor o igual que 2. Todos los modelos presentados en esta tabla tienen valores de *accuracy* y de *Kappa* menores a 0.6. Sin embargo, son suficientemente útiles para identificar los predictores socio-demográficos que tienen mayor asociación estadística con las tasas municipales de casos positivos y de defunción. También describen cuantitativamente las diferentes combinaciones de predictores y sus valores que están asociados a los diversos niveles de las tasas. Con estos modelos se compara en qué se parecen los municipios con tasas altas, medias o bajas de casos positivos o de defunción.

Los modelos de árboles clasificadores producidos con el *package* C50 (Kuhn and Quinlan, 2022) en lenguaje R (R Core Team, 2019) incluyen una estimación de la *importancia* o *peso* de cada predictor en relación con el *target* especificado. Esta importancia es denominada *attribute usage* (uso de atributo) y es el porcentaje de casos del *dataset* de entrenamiento (*training set*) que corresponden a los nodos terminales (llamados hojas del árbol) usando ese atributo. De este modo, el atributo predictor que es colocado automáticamente por C50 como raíz del árbol clasificador tiene la máxima estimación de importancia (100.0%). Los predictores que son colocados por C50 en los demás nodos del árbol, tienen porcentajes cada vez menores, conforme se desciende desde la raíz hacia las hojas; por ello, los atributos más cercanos a estas podrían tener valores cercanos a 0.0%. Los valores de *attribute usage* de nuestros modelos están disponibles dentro de los archivos que pusimos en la web para ser descargados.

A continuación, se describen los patrones de interacción multivariable (reglas clasificadoras) más destacables correspondientes a cada uno de los atributos *target* (categóricos) de intervalos de tasas municipales por cada 100,000 habitantes, enumerados en la tabla 136, incluyendo casos positivos y defunciones, en diversos periodos de la pandemia. De los modelos que hemos puesto en la Nube para ser descargados, los que contienen las reglas numeradas son aquellos que en su nombre incluyen la palabra *rules*.

7.1 Tasas de positivos en 2020: posit_2020_tasa_100mil_INTERV

En el modelo núm. 1 de la tabla 136, el *target* de tasas municipales de casos positivos en 2020 tiene diez clases mayoritarias, las que constituyen el 82.9% de los municipios mexicanos. De estas, las clases con mayores porcentajes son: la primera, entre 0 y 60 casos positivos por cada 100 mil habitantes, que representa el 16.6% de los municipios; segunda, la clase entre 900 y 2,100 casos positivos, con 12.8% de los municipios y, tercera, entre 60 y 120 casos, con 10.1% de los municipios.

Con base en el *attribute usage*, los cinco predictores más asociados a la tasa de casos positivos del año 2020 son:

- i. Porcentaje de viviendas que tienen computadora (*vph_pc_percent*, 100.0%)
- ii. Entidad federativa de ubicación del municipio (*nom_ent_abrev*, 97.8%)

- iii. Porcentaje de viviendas que tienen servicio de películas, música o videos de paga por Internet (*vph_spmvpi_percent*, 65.6%)
- iv. Porcentaje de viviendas con piso de tierra (*vph_pisoti_percent*, 40.0%) y
- v. Porcentaje de viviendas sin línea telefónica ni teléfono celular (*vph_sinltc_percent*, 33.8%).

Es interesante que tres de estos cinco predictores se refieren a productos y servicios de tecnologías de información y comunicación. Por otra parte, la regla descriptiva que describe a más municipios en este modelo es:

Si las viviendas con servicio de películas, música o videos de paga por Internet son menos o igual que 14.3%; entonces la tasa de positivos en 2020 es de entre 300 y 360 casos por cada 100,000 habitantes (soporte, support = 1,671; confianza, confidence = 7.0%).

Esto significa que, aunque este patrón ocurre en 1,671 municipios, no es tan relevante porque solamente en el 7.0% de ellos se cumple que la tasa es entre 300 y 360; mientras que en el 93.0% de esos municipios la tasa pertenece a otros intervalos. El intervalo de tasa más alto en este modelo es *t_2100_a_5700*; sin embargo, no se observa un patrón sociodemográfico sólido a este respecto.

7.2 Tasas de positivos en 2021: *posit_2021_tasa_100mil_INTERV*

En los modelos 2 y 3 de la tabla en cuestión, el *target* de tasas municipales de casos positivos en 2021 tiene doce clases mayoritarias, que contienen al 80.2% de los municipios. Las tres clases con porcentajes más altos son: entre 0 y 100 casos positivos, entre 100 y 200 y entre 200 y 300, con porcentajes de 11.3%, 10.9% y 9.3%, respectivamente.

De estos dos modelos, el núm. 2 no es tan interesante porque indica que la tasa de casos positivos en 2021 está influida principalmente por la tasa de positivos en 2020 (su raíz es el predictor *posit_2020_tasa_100mil*). Este patrón podría considerarse obvio; en cambio, el modelo 3 es más interesante porque tiene como raíz al predictor llamado *vph_pc_percent* (porcentaje de viviendas con computadora). En este modelo, los cinco predictores con mayor asociación estadística hacia el *target* son:

- i. Porcentaje de viviendas con computadora: *vph_pc_percent* (100.0%)
- ii. Entidad federativa de ubicación del municipio: *nom_ent_abrev* (99.0%)
- iii. Latitud geográfica del municipio: *lat_decimal* (58.1%)
- iv. Porcentaje de viviendas con internet: *vph_inter_percent* (42.6%)
- v. Porcentaje de viviendas que tienen automóvil o camioneta: *vph_autom_percent* (38.9%)

Nótese que los dos primeros de estos cinco predictores son los mismos que en el modelo 1. Además, la presencia de internet (cuarto predictor) también es un servicio de tecnologías de información y comunicación.

La regla clasificadora que describe a más casos es:

Si las viviendas con computadora personal son menos o igual que 33.1%; entonces la tasa de positivos en 2021 es de entre 200 y 300 por cada 100,000 habitantes (soporte = 1,631 municipios, confianza = 10.0%).

La regla no es tan relevante porque solo el 10.0% de estos municipios tiene tasa entre 200 y 300 positivos por cada 100 mil habitantes; el 90.0% de los municipios restantes tiene otras tasas. El intervalo de tasas más alto en este modelo es *t_4000_a_15000* y no se observa un patrón sólido al respecto.

7.3 Tasas de positivos en 2022 (solo hasta 17 de septiembre): *posit_2022_a_sep_17_tasa_100mil_INTERV*

En los modelos 4 y 5, el *target* de tasas municipales de casos positivos en 2022 (que incluye casos solamente hasta el 17 de septiembre de ese año) tiene 8 clases mayoritarias, abarcando el 81.3% de los municipios. Las tres clases con los porcentajes más altos son: entre 0 y 150 casos positivos por cada 100 mil habitantes, entre 150 y 300, y entre 300 y 450, con porcentajes de 30.9%, 15.2% y 10.1% de los municipios, respectivamente.

El modelo 4 tiene como raíz a la tasa de casos positivos del año 2021 (*posit_2021_tasa_100mil*), lo cual podría considerarse un patrón obvio. En contraste, el modelo 5 tiene como raíz al grado promedio de estudios de la población femenina (*graproes_f_un_dec*), lo cual resulta más interesante. Curiosamente, en este último modelo existe un empate con valor de 100.0% en el uso de sus dos predictores más importantes; por ello, en lugar de presentar los cinco predictores más asociados al *target*. A continuación se presentan seis, que son:

- i. Entidad federativa de ubicación del municipio: *nom_ent_abrev* (100.0%)
- ii. Grado promedio de estudios de la población femenina: *graproes_f_un_dec* (100.0%)
- iii. Porcentaje de viviendas con computadora: *vph_pc_percent* (60.6%)
- iv. Grado promedio de estudios, incluyendo a hombres y mujeres: *graproes_un_dec* (47.4%)
- v. Porcentaje de población de 12 a 14 años de edad: *p_12a14_percent* (36.1%)
- vi. Porcentaje de población de 6 a 11 años de edad: *p_6a11_percent* (35.8%)

La regla clasificadora que contiene a más municipios (el mayor soporte) en este modelo es:

Si las viviendas con computadora personal son menos o igual que 34.6%; entonces la tasa de positivos en 2022 es de entre 150 y 300 por cada 100 mil habitantes (soporte = 1,662 municipios, confianza = 16.7%).

Aunque su cantidad de municipios es considerable, el nivel de confianza es bajo. Su intervalo de tasas más altas es *t_5000_a_13000* y presenta algunos patrones considerablemente sólidos; los de mayor soporte y confianza son:

- i. Si la entidad federativa del municipio es la Ciudad de México, y su población con limitación (discapacidad) para caminar, subir y bajar es mayor al 3.1%; entonces la tasa de positivos en 2022 es entre 5,000 y 13,000 por cada 100,000 habitantes (soporte = 9 municipios, confianza = 90.9%).
- ii. Si la entidad federativa del municipio es Baja California Sur, Colima, o Tabasco, y las viviendas particulares habitadas con computadora personal son más del 34.6%; entonces la tasa de positivos en 2022 es entre 5,000 y 13,000 por cada 100,000 habitantes (soporte = 6 municipios, confianza = 87.5%).
- iii. Si la entidad federativa del municipio es Baja California Sur; entonces la tasa de positivos en 2022 es entre 5,000 y 13,000 por cada 100,000 habitantes (soporte = 4 municipios, confianza = 83.3%).

7.4 Tasas de positivos desde 2020 hasta 17/Sep/2022: posit_MPAL_TOTAL_tasa_100mil_INTERV

En el modelo 6, el *target* de tasas municipales de casos positivos desde 2020 hasta el 17 de septiembre 2022 tiene ocho clases mayoritarias, conteniendo el 80.2% de los municipios. Sus tres clases con mayores porcentajes son: entre 0 y 400, entre 400 y 800, y entre 800 y 1,200 casos positivos por cada 100 mil habitantes, con porcentajes de 18.6%, 17.0% y 12.7% de los municipios, respectivamente.

Este modelo es interesante porque describe los patrones municipales de casos positivos de todo el periodo analizado. También resulta destacable que su atributo raíz es el porcentaje de viviendas con computadora (*vph_pc_percent*). Los cinco predictores más importantes son:

- i. Porcentaje de viviendas con computadora: *vph_pc_percent* (100.0%)
- ii. Entidad federativa: *nom_ent_abrev* (96.7%)
- iii. Porcentaje de población de 18 a 24 años de edad y que asiste a la escuela: *p18a24a_percent* (51.3%)
- iv. Porcentaje de viviendas con horno de microondas: *vph_hmicro_percent* (41.5%)
- v. Porcentaje de viviendas con servicio de televisión de paga: *vph_stvp_percent* (38.4%)

Otra vez, entre los predictores más importantes se encuentran dos relacionados con las TIC (computadora y televisión de paga). Curiosamente, el sexto predictor más importante está relacionado con los servicios de salud: *pder_segp_percent*, el porcentaje de habitantes afiliados al Instituto de Salud para el Bienestar (*attribute usage*: 32.1%). En este modelo, la regla clasificadora que abarca a más municipios es:

Si la población que tiene 18 años y más con educación postbásica es menor o igual que 26.7%; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es de entre 1,200 y 1,600 por cada 100,000 habitantes (soporte = 1,526 municipios, confianza = 10.7%).

Dado que este patrón solo se cumple en el 10.7% de esos municipios, en el resto de los municipios la tasa no es de entre 1,200 y 1,600. La clase de intervalo de tasas de positivos más altas es *t_10000_a_30000* (entre 10,000 y 30,000 positivos por cada

100,000 habitantes). La regla con mayor soporte (mayor cantidad de municipios) para esta clase es:

Si la entidad federativa es Nuevo León; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es de entre 10,000 y 30,000 por cada 100,000 habitantes (soporte = 38 municipios, confianza = 12.5%).

Sin embargo, existe otra regla perteneciente a esta clase que, aunque tiene menor soporte, tiene un mayor nivel de confianza:

Si la entidad federativa es Baja California Sur o Ciudad de México; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es de entre 10,000 y 30,000 por cada 100,000 habitantes (soporte = 14 municipios, confianza = 93.7%).

Esto último es relevante porque sugiere que existen casi 14 municipios (o alcaldías) en los estados de Baja California Sur y en Ciudad de México que, en el periodo 2020-2022 tuvieron las más altas tasas de positivos del país.

En cambio, la clase de intervalo con las tasas de positivos más bajas (entre 0 y 400 positivos por cada 100,000 habitantes) se caracteriza por un conjunto de reglas descriptivas de las que se muestran a continuación solamente las cuatro con mayor soporte (más número de municipios).

- i. Si la entidad federativa del municipio es Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, o Veracruz, y la población femenina es menor del 57.7%, y la población de entre 0 y 2 años es menor o igual a 8.8%, y la población de 18 años y más con educación postbásica es menor o igual al 14.8%, y las viviendas con horno de microondas son menos o igual que 6.2%, y las viviendas con automóvil son menos o igual que 7.8%; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es entre 0 y 400 por cada 100,000 habitantes (soporte = 103 municipios, confianza = 98.1%).*
- ii. Si la población del municipio que tiene limitaciones (discapacidad) para hablar o comunicarse es mayor a 1.2%; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es entre 0 y 400 por cada 100,000 habitantes (soporte = 102 municipios, confianza = 48.0%).*
- iii. Si la población del municipio que tiene entre 0 y 2 años de edad es menor o igual al 8.8%, y la población de 18 años y más con educación postbásica es menor o igual al 14.8%, y las viviendas sin electricidad son menos o igual que el 20.6%, y las viviendas que tienen agua entubada y se abastecen del servicio público de agua son menos o igual que el 73.3%, y las viviendas con excusado son menos o igual que el 83.0%, y las viviendas con letrina son menos o igual que el 79.8%, y las viviendas con horno de microondas son menos o igual que el 6.2%, y las viviendas con automóvil son menos o igual que 26.6%, y las viviendas con motocicleta son menos o igual que 18.1%, y las viviendas con bicicleta son menos o igual que 21.6%, y las viviendas con servicio de televisión de paga son menos o igual que 47.5%, y las viviendas con servicio de películas, música o videos de paga por Internet son menos o igual que 1.1%; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es entre 0 y 400 por cada 100 mil habitantes (soporte = 88 municipios, confianza = 96.6%).*

iv. Si la entidad federativa del municipio es alguna de: Chiapas, Guerrero, Oaxaca, Puebla, o Veracruz, y la población femenina es menor o igual al 57.7%, y la población de 18 años y más con educación postbásica es menor o igual al 14.8%, y la población con religión distinta a las diversas denominaciones del cristianismo es menor o igual al 0.3%, y la población que no profesa ninguna religión es mayor que 0.6%, y las viviendas con agua entubada y que se abastecen del servicio público de agua son más del 73.3%, y las viviendas con refrigerador con menos o igual que 64.6%, y las viviendas con horno de microondas son menos o igual que 6.2%; entonces la tasa de positivos desde 2020 hasta 17/Sep/2022 es entre 0 y 400 por cada 100 mil habitantes (soporte = 61 municipios, confianza = 98.4%).

7.5 Tasas de defunción en 2020: def_2020_tasa_100mil_INTERV

En los modelos 7 y 8, el target de tasas municipales de defunción en 2020 tiene once clases mayoritarias, que abarcan al 81.8% de los municipios. Las tres clases con mayores porcentajes de municipios son: entre 0 y 10 casos positivos por cada 100 mil habitantes, entre 30 y 40 casos, y entre 20 y 30 casos, con 22.7%, 7.7% y 7.5% de los municipios, respectivamente.

El modelo 7 resulta poco interesante porque tiene como raíz al atributo de tasa municipal de casos positivos del año 2020 (*posit_2020_tasa_100mil*). En otras palabras, el predictor más poderoso respecto a la tasa de defunciones de 2020 es la tasa de casos positivos de ese mismo año, lo cual constituye un patrón obvio. En cambio, el modelo 8 tiene como raíz el grado promedio de estudios de los habitantes municipales (*graproes_un_dec*), lo que resulta más interesante. Los cinco predictores más poderosos en este modelo son:

- i. Grado promedio de estudios: *graproes_un_dec* (100.0%)
- ii. Entidad federativa de ubicación del municipio: *nom_ent_abrev* (97.5%)
- iii. Porcentaje de viviendas con computadora: *vph_pc_percent* (61.0%)
- iv. Porcentaje de viviendas con automóvil: *vph_autom_percent* (44.0%)
- v. Coordenada de longitud geográfica: *lon_decimal* (40.1%)

Del modelo 8, la regla clasificadora que describe a la mayor cantidad de municipios es:

Si la población con discapacidad auditiva es menor o igual al 3.5%; entonces la tasa de defunción en 2020 es de entre 20 y 30 por cada 100,000 habitantes (soporte = 1,773 municipios, confianza = 8.2%).

La clase del intervalo de tasas de defunción más altas en este modelo es *t_300_a_1500*. De esta clase, su regla con mayor soporte (mayor número de municipios) es:

Si la población con discapacidad es mayor del 9.8%, y la población femenina de 15 años y más con primaria completa es mayor que 8.7%; entonces la tasa de defunción en 2020 es de entre 300 y 1,500 por cada 100,000 habitantes (soporte = 49 municipios, confianza = 7.8%).

7.6 Tasas de defunción en 2021: `def_2021_tasa_100mil_INTERV`

En los modelos 9 y 10, el *target* de tasas municipales de defunción en 2021 tiene once clases mayoritarias, abarcando al 82.3% de los municipios. Las tres clases con los más altos porcentajes de municipios son: entre 0 y 10 defunciones por cada 100 mil habitantes, entre 150 y 300 defunciones, y entre 50 y 60 defunciones; conteniendo al 16.1%, 9.3% y 7.4% de los municipios.

El modelo 9 presenta un patrón obvio respecto al atributo que ocupa su raíz (*posit_2021_tasa_100mil*); es decir, las tasas de defunción en 2021 dependen de las tasas de casos positivos en ese mismo año. Por su parte, en el modelo 10, los cinco predictores más poderosos y sus respectivos porcentajes de uso son:

- i. *vph_pc_percent* (100.0%): mencionado anteriormente
- ii. *nom_ent_abrev* (84.3%): mencionado anteriormente
- iii. *pclim_pmen_percent* (56.0%): porcentaje de población con algún problema o condición mental
- iv. *p12ym_solt_percent* (53.0%): porcentaje de población que tiene 12 años o más de edad y es soltera
- v. *p6a11_noa_percent* (49.8%): porcentaje de población que tiene de 6 a 11 años de edad y no asiste a la escuela

Del modelo 9, la regla clasificadora que describe a más municipios es:

Si la población con edad entre 8 y 14 años es menor o igual de 14.8%; entonces la tasa de defunción en 2021 es de entre 60 y 70 por cada 100,000 habitantes (soporte = 1,517 municipios, confianza = 8.6%).

La clase correspondiente a las tasas más altas en este modelo es *t_300_a_2100*. De esta, la regla descriptiva que aborda a más municipios es:

Si el promedio municipal de ocupantes por vivienda es menor o igual de 2.9; entonces la tasa de defunción en 2021 es de entre 300 y 2,100 por cada 100,000 habitantes (soporte = 94 municipios, confianza = 5.2%).

7.7 Tasas de defunción en 2022 (solo hasta 17 de septiembre): `def_2022_a_sep_17_tasa_100mil_INTERV`

En los modelos 11 y 12, el *target* de tasas municipales de defunción en 2022 tiene nueve clases mayoritarias, que abarcan al 81.1% de los municipios. De estas clases, las tres que tienen mayores porcentajes de municipios son: entre 0 y 2 defunciones por cada 100 mil habitantes, entre 30 y 90 defunciones, y entre 10 y 12 defunciones, abarcando al 39.5%, el 9.9% y el 5.0% de los municipios, respectivamente.

El modelo 11 es de poco interés por que su raíz es la tasa de defunciones del año 2020 (*def_2020_tasa_100mil*), que está altamente asociada a la correspondiente tasa en el año 2022 (hasta el 17 de septiembre). En contraste, el modelo 12 es más interesante porque su raíz es el porcentaje de viviendas con consola de videojuegos

(*vph_cvj_porcent*), un patrón completamente inesperado. En el modelo 12, sus cinco atributos predictores más importantes, con base en sus porcentajes de uso, son:

- i. Porcentaje de viviendas con consola de videojuegos: *vph_cvj_percent* (100.0%)
- ii. Entidad federativa de ubicación del municipio: *nom_ent_abrev* (97.4%)
- iii. Porcentaje de población de 18 a 24 años de edad: *p_18a24_percent* (62.9%)
- iv. Porcentaje de población con alguna discapacidad: *pcon_limi_percent* (62.0%)
- v. Altitud geográfica del municipio: *altitud* (25.2%)

Del modelo 12, la regla clasificadora que describe a más municipios es:

Si la población con edad entre 15 y 64 años es menor o igual de 68.0%; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 0 y 2 por cada 100,000 habitantes (soporte = 1,747 municipios, confianza = 42.4%).

La clase con las tasas más altas en este modelo es *t_90_a_450*. Esta tiene cinco reglas; cada una de estas tiene soporte = 2 (municipios) y valor de confianza = 75.0%. Por lo tanto, en total, estas reglas describen a diez municipios. Debido a su interés, se muestran a continuación:

- i. *Si la entidad federativa es Coahuila, y las viviendas con piso de tierra son menos o igual que 0.2%, y las viviendas con televisor son más de 92.5%; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 90 y 450 por cada 100,000 habitantes (soporte = 2, confianza = 75.0%).*
- ii. *Si la entidad federativa es Oaxaca, y la altitud geográfica del municipio es menor de 677 MSNM, y el grado promedio de estudios de la población femenina es mayor a 7.6 años, y las viviendas servicio de agua dentro de la vivienda son menos o igual que 83.4%, y las viviendas con computadora personal son menos o igual que 17.9%; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 90 y 450 por cada 100,000 habitantes (soporte = 2, confianza = 75.0%).*
- iii. *Si la entidad federativa del municipio es Oaxaca, y el grado promedio de estudios de la población femenina es menor de 7.6 años, y las viviendas con electricidad son menos o igual de 94.9%, y las viviendas con consola de videojuegos son menos o igual de 1.1%; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 90 y 450 por cada 100,000 habitantes (soporte = 2, confianza = 75.0%).*
- iv. *Si la entidad federativa del municipio es Chihuahua, y la población con alguna discapacidad ligera es mayor al 17.6%, y la población con limitación para hablar o comunicarse es menor o igual de 0.7%, y el grado promedio de estudios de la población es mayor a 7.7 años; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 90 y 450 por cada 100,000 habitantes (soporte = 2, confianza = 75.0%).*
- v. *Si la entidad federativa del municipio es Oaxaca, y la población con alguna discapacidad ligera es mayor a 17.6%, y la población católica es mayor a 92.5%, y las viviendas con consola de videojuegos son más de 1.1%; entonces la tasa de defunción en 2022 (solo hasta 17 de septiembre) es de entre 90 y 450 por cada 100,000 habitantes (soporte = 2, confianza = 75.0%).*

7.8 Tasas de defunción desde 2020 hasta 17/Sep/2022: def_MPAL_TOTAL_tasa_100mil_INTERV

En los modelos 13 y 14, el *target* de tasas municipales de defunción desde 2020 hasta el 17 de septiembre de 2022 tiene diez clases mayoritarias, que constituyen el 83.1% de los municipios. Las tres clases con mayores porcentajes de municipios son: entre 0 y 25 defunciones por cada 100,000 habitantes, entre 100 y 125 defunciones, y entre 75 y 100 defunciones, con 13.5%, 9.5% y 9.1% de los municipios, respectivamente.

El modelo 13 resulta poco interesante porque en su raíz tiene al predictor *def_2021_tasa_100mil* (tasa de defunciones del año 2021); es decir, esta tasa influye en la correspondiente tasa del periodo 2020-2022, lo cual es obvio. En cambio, el modelo 14 tiene en la raíz al porcentaje de viviendas con computadora (*vph_pc_percent*). Quizá este patrón podría considerarse poco esperado por los profesionales de la salud. Los cinco predictores más importantes en el modelo 14 son:

- i. Porcentaje de viviendas con computadora: *vph_pc_percent* (100.0%)
- ii. Entidad federativa de ubicación del municipio: *nom_ent_abrev* (83.2%)
- iii. Porcentaje de población afiliada a servicios de salud del Instituto Mexicano del Seguro Social (IMSS): *pder_imss_percent* (72.0%)
- iv. Porcentaje de población de 12 años de edad y más soltera o nunca unida: *p12ym_solt_percent* (61.1%)
- v. Coordenada geográfica de longitud del municipio: *lon_decimal* (45.9%)

La regla descriptiva que abarca a la mayor cantidad de municipios en este modelo es:

Si la población de 12 años y más que está soltera es menor o igual a 27.7%; entonces la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 100 y 125 (soporte = 1,630 municipios, confianza = 9.9%).

La clase de intervalo con las tasas más altas es *t_600_a_2600* (entre 600 y 2,600 defunciones por cada 100,000 habitantes), que posee dos reglas descriptivas. Cada una de estas dos tiene un valor de soporte de 1 (municipio) y valor de confianza de 0.6667. Debido a su interés, las mostramos a continuación.

- i. *Si la entidad federativa del municipio es Ciudad de México, y la población afiliada a servicios de salud en Pemex, Sedena o Semar es mayor a 1.9%; entonces la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 600 y 2,600 por cada 100,000 habitantes (soporte = 1, confianza = 66.6%).*
- ii. *Si la población de 12 años y más que es soltera es mayor al 30.2%, y las viviendas que no tienen drenaje son más del 9.3%, y las viviendas con computadora personal son más del 12.3%; entonces la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 600 y 2,600 por cada 100,000 habitantes (soporte = 1, confianza = 66.6%).*

Por otra parte, la clase de intervalo con las tasas de defunción más bajas es *t_0_a_25* (entre 0 y 25 defunciones por cada 100,000 habitantes). De estas reglas, las tres que tienen soporte (número de municipios) más alto son:

- i. Si la población de 15 años y más residente en otra entidad en marzo de 2015 (los migrantes interestatales) es menor o igual de 1.2%, y la población de 15 a 17 años que asiste a la escuela es menor o igual de 4.9%, y las viviendas con horno de microondas son menos de o igual a 5.8%; entonces la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 0 y 25 por cada 100,000 habitantes (soporte = 86 municipios, confianza = 72.7%).
- ii. Si la población de 15 años y más con primaria completa es menor o igual a 24.3%, y la población afiliada a servicios de salud del IMSS Bienestar es menor o igual a 23.4%, y las viviendas sin tienen automóvil, camioneta, motocicleta ni motoneta son más del 94.4%, y las viviendas con horno de microondas son menos o igual a 5.8%; entonces, la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 0 y 25 por cada 100,000 habitantes (soporte = 44, confianza = 95.6%).
- iii. Si la entidad federativa del municipio es Chiapas, y la población de 3 años y más es mayor al 87.6%, y la población con discapacidad para caminar, subir o bajar es mayor al 1.2%, y la población de 12 años y más que estuvo casada o unida es menor o igual a 7.9%, y las viviendas con lavadora son menos o igual a 57.3%; entonces la tasa de defunción desde 2020 hasta 17/Sep/2022 es de entre 0 y 25 por cada 100,000 habitantes (soporte = 38, confianza = 97.5%).

7.9 Discusión de los resultados

Algo de lo más destacable de los catorce modelos presentados, siendo seis para tasas municipales de casos positivos y ocho para tasas de defunción, es que algunos de los atributos predictores más importantes son los mismos en varios de los modelos; por ejemplo, en todos los modelos para tasas de casos positivos están presentes los predictores: porcentaje de viviendas con computadora personal (*vph_pc_percent*) y entidad federativa correspondiente al municipio (*nom_ent_abrev*). Por su parte, en los modelos para defunción, los predictores más importantes son: porcentaje de viviendas con consola de videojuegos (*vph_cvj_porcent*), *vph_pc_percent* y grado promedio de estudios del municipio. Es interesante que entre los predictores más poderosos se encuentren dos del ámbito de las tecnologías de información y comunicación: computadora, y consola de videojuegos.

En otras palabras, los resultados presentados muestran, obviamente, que la mortalidad por COVID-19 está correlacionada con la morbilidad. En este sentido, se supone que donde hay morbilidad alta también hay mortalidad alta. Hasta aquí, estas asociaciones estadísticas no presentan nada novedoso. Sin embargo, la interacción de las variables de tasas de defunción con las de presencia de computadora y de consola de videojuegos en las viviendas, parecería poco esperada. Una potencial explicación preliminar y tentativa, que requeriría una validación por otros medios, sería que la tenencia de computadora y consola de videojuegos está vinculada a:

- i. Situación socioeconómica de la familia, ya que las personas con más ingresos económicos tienden a adquirir estos dispositivos.
- ii. Acceso a servicio de energía eléctrica, pues en localidades remotas no electrificadas hay escasos dispositivos de este tipo.
- iii. Acceso a servicio de internet, estrechamente ligado a disponibilidad de electricidad.

- iv. Nivel de escolaridad: cuanto más alto este nivel, mayor probabilidad de uso de los dispositivos mencionados.
- v. Los adultos que usan estos dispositivos para realizar su trabajo tendrán empleos más individualistas y con poca demanda de esfuerzo físico para desempeñarlo.
- vi. Las y los adultos mayores, más afectados por la COVID-19, usan muy poco o no usan estos dispositivos.

Dos vínculos estrechos entre la presencia de dispositivos de TIC en las viviendas y el COVID-19 que se resaltan son los siguientes. Primeramente, el acceso a la información digital: las personas que usan estos dispositivos tuvieron acceso a diversas fuentes de información (noticias, reportajes, entrevistas, documentales, infomerciales, etcétera) teóricamente tuvieron mayor oportunidad de estar al día sobre los acontecimientos importantes de prevención al contagio de SARS-CoV-2. En segundo término, el aislamiento que provoca el uso de estos dispositivos. En general el uso de la computadora es unipersonal, a diferencia de las consolas de juegos donde pueden participar varias personas jóvenes o niños. Además del confinamiento social indicado por las autoridades sanitaria aislamiento, y que muchas familias lo hicieron durante el periodo álgido de la pandemia, el uso de estos dispositivos, al ser de uso unipersonal, provocó aislamiento dentro de las familias y posiblemente esto contribuyó a menor contacto interpersonal. La presencia de estos dispositivos provocó una *sana distancia* entre los integrantes de familia y, en casos donde hubo transmisión del SARS-CoV-2, las cargas virales de menor intensidad. Para concluir, la relación entre las computadoras y consolas de videojuegos con la mortalidad por COVID-19 podrían explicarse por las condiciones socioeconómicas más prevalentes en los diversos municipios de México y por el acceso a información y la baja transmisión del SARS-CoV-2 derivada del uso de estos dispositivos.

Respecto a casos positivos, los municipios (o alcaldías) con tasas más altas (entre 10,000 y 30,000 por cada 100,000 habitantes) se parecen entre sí en que varios de ellos están ubicados en Baja California Sur y Ciudad de México. En cambio, los municipios con las tasas de positivos más bajas (entre 0 y 400 por cada 100,000 habitantes) se parecen en que: se ubican en entidades del sur y sureste del país, tales como Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, y Veracruz. También, tienen una baja presencia de servicios básicos, como electricidad, agua entubada, excusado, letrina y de bienes de consumo, tales como horno de microondas, automóvil, motocicleta, bicicleta, servicio de TV de paga, y de *streaming* por internet.

Finalmente, los municipios (o alcaldías) que tienen las tasas de defunción más altas (entre 600 y 2,600 defunciones por cada 100,000 habitantes) no tienen un patrón muy frecuente que permita caracterizarlos. En contraste, los municipios que tienen las tasas de defunción más bajas (entre 0 y 25 defunciones por cada 100,000 habitantes), se parecen en que: tienen baja recepción de migrantes interestatales; tienen baja presencia de bienes de consumo, tales como horno de microondas, automóvil, camioneta, motocicleta o lavadora; bajos porcentajes de población con primaria completa, y bajo porcentaje con afiliación al IMSS-Bienestar.

CAPÍTULO 8. COMENTARIOS FINALES, RECOMENDACIONES Y TRABAJO DE INVESTIGACIÓN A FUTURO

El presente libro presenta una contribución significativa al entendimiento y manejo de la pandemia de COVID-19 en México, destacándose por su enfoque en el análisis cuantitativo y el uso de técnicas avanzadas como el aprendizaje automático. Los resultados del análisis profundizan y aportan evidencia crucial para la formulación de políticas públicas en varios frentes. Por ejemplo, el fortalecimiento del monitoreo epidemiológico podría facilitar la detección temprana de brotes y una respuesta más efectiva, dada la factibilidad y disponibilidad de datos precisos y en tiempo real. Además, la adopción de metodologías robustas para el análisis de datos puede mejorar la comprensión de la pandemia de la COVID-19 (y cualquier otro evento epidemiológico), permitiendo intervenciones más dirigidas y efectivas. La correlación entre diversas variables y la incidencia de la enfermedad podría orientar políticas que se enfoquen en grupos de alto riesgo y regiones específicas.

Asimismo, entender los impactos diferenciados de la pandemia a nivel municipal y en grupos vulnerables es fundamental para desarrollar políticas públicas más equitativas y eficientes. Estos hallazgos podrían mejorar la respuesta inmediata y reforzar los sistemas de salud pública, haciéndolos más resilientes ante futuras emergencias sanitarias. Estos análisis son también clave para la optimización de recursos, asegurando que se utilicen de manera efectiva en la atención sanitaria y medidas preventivas. Por lo anterior, lo desarrollado en el libro es solo un ejemplo de la vasta información que podría generarse para la toma de decisiones, guiada por la evidencia basada en técnicas de estadística descriptiva y aprendizaje automático.

Iniciamos con la exposición de nuestro diseño metodológico, seguido de un panorama estadístico descriptivo del *dataset Master-3*, que es el producto de un procesamiento minucioso de los datos originales de la Secretaría de Salud. Delimitamos nuestro análisis a los casos positivos y destacamos patrones de correlación multivariante encontrados mediante el uso de árboles clasificadores, específicamente

con el algoritmo C5.0 de R. Quinlan. Además, enfatizamos patrones cronológicos a nivel municipal, cubriendo prácticamente todos los municipios mexicanos. Mediante mapas digitales, ilustramos la dispersión geográfica de la pandemia y concluimos con hallazgos sobre la correlación entre características sociodemográficas de los municipios y la prevalencia de COVID-19.

Desde una perspectiva de la metodología, se destaca la aplicación del aprendizaje automático, aplicando el algoritmo C5.0 usando *software* libre, lo que permite abordar un conjunto de alrededor de 40 variables que representan datos de más de 18 millones de individuos. Además, se proporciona el código fuente en lenguaje de programación R para generar modelos, lo que facilita la replicación de la metodología utilizada.

Entre los resultados más destacados de esta investigación, se encuentra la creación de 29 modelos de árboles clasificadores con reglas explicativas, que proporcionan una visión detallada de las interacciones estadísticas entre los atributos del conjunto de datos de casos individuales de COVID-19. Entre otros modelos, se presentan algunos que identifican las variables más influyentes en las probabilidades de hospitalización, intubamiento y fallecimiento de pacientes, lo que facilita la predicción de riesgos, la adaptación de tratamientos y la mejora de la atención médica. Además, la representación de patrones geográficos a partir de datos digitales permite la visualización de casos relacionados con el COVID-19 a lo largo del tiempo.

De este estudio, resaltamos cinco aspectos fundamentales:

- i. Propósito del estudio: buscamos identificar patrones en los casos reportados de COVID-19 en México, utilizando la base de datos oficial de la Secretaría de Salud y asociando con variables sociodemográficas del Censo Nacional de Población y Vivienda 2020.
- ii. Aproximación metodológica: comenzamos con un análisis univariado utilizando estadística descriptiva y la técnica de Pareto. Posteriormente, empleamos aprendizaje automático para identificar correlaciones multivariadas.
- iii. Principales hallazgos: detectamos tendencias en la evolución de casos confirmados, áreas geográficas más afectadas, y grupos poblacionales más vulnerables. Estudiamos la situación municipal de COVID-19 y su relación con características sociodemográficas.
- iv. Acceso universal: todos los *datasets* y modelos generados en esta investigación están disponibles en línea, promoviendo la transparencia y reproducibilidad.
- v. Implicaciones: nuestros hallazgos ofrecen una base empírica para futuras investigaciones y para el diseño de políticas y estrategias específicas.

Desafíos y retos pendientes en torno al subregistro y manejo de la información

El vínculo entre la información oficial y la identificación de patrones estadísticos es crucial. A pesar de que nos basamos en datos oficiales, estos a menudo presentan

limitantes de subregistro. Frente a este desafío, se destaca la necesidad de actualización constante de la información y la consideración de la variabilidad de datos.

Para mejorar la precisión y actualidad de la información, se recomienda establecer un comité especializado que colabore con la Dirección General de Epidemiología y la Dirección General de Información en Salud. La homogenización de los datos a nivel nacional es esencial. Reconociendo estos patrones, se pueden prever brotes, zonas críticas y ajustar las estrategias de intervención adecuadamente.

En la actualidad, con la creciente ola de digitalización, el análisis de datos avanzado se ha consolidado como una herramienta esencial en múltiples disciplinas, desde investigaciones científicas y médicas hasta economía y estudios sociales. Analizar y deducir patrones de grandes volúmenes de datos es crucial para decisiones respaldadas por pruebas y para concebir soluciones efectivas. No obstante, para maximizar el potencial de estas técnicas sofisticadas, es imperativo considerar algunos elementos esenciales, que se comentan a continuación.

Los métodos de análisis de vanguardia, como la inteligencia artificial y, específicamente, el aprendizaje automático, exigen alto poder computacional con recursos de infraestructura tecnológica actualizada en términos de *hardware* y *software*. Esto se traduce en la necesidad de equipos de alto rendimiento y *software* específico. Las bases de datos deben ser versátiles al tratar con enormes cantidades de información, mientras que las conexiones de red deben garantizar una transmisión rápida y segura de los datos.

Además de los requerimientos tecnológicos ya descritos, es esencial contar con recursos humanos capaces, tanto de manejar herramientas relacionadas, como de interpretar y aplicar los datos en escenarios reales. Esto implica desde expertos en datos y desarrolladores, hasta profesionales en áreas de aplicación específicas; por ejemplo, epidemiólogos, ante situaciones de pandemia. De ahí, la necesidad imperativa de capital humano especializado.

El mundo del análisis de datos está en perpetua transformación. Lo que hoy es una norma, mañana resultaría obsoleto. Por ello, es vital promover una formación constante, asegurando que el equipo esté actualizado en cuanto a innovaciones y prácticas emergentes por medio de programas de actualización continua.

Desafíos en torno a la diversidad geográfica

La diversidad geográfica de México implica que diferentes regiones enfrentan desafíos únicos. Por ejemplo, las regiones montañosas podrían enfrentar desafíos relacionados con el acceso a servicios básicos debido a su topografía. Las zonas urbanas, con alta densidad de población, pueden enfrentar problemas de contaminación y salud pública, mientras que las áreas rurales podrían lidiar con la falta de infraestructura y servicios de salud. Esta variedad representa los desafíos existentes en términos de políticas públicas y su aplicación. Es esencial diseñar estrategias específicas para áreas o poblaciones vulnerables, optimizando la distribución de recursos y atención médica, en razón de la experiencia desarrollada durante la pandemia.

No todas las poblaciones y regiones tienen las mismas necesidades ni enfrentan las mismas amenazas. Las comunidades indígenas, por ejemplo, pueden tener requisitos culturales y lingüísticos específicos que deben ser considerados en cualquier intervención. Las zonas rurales, alejadas de los centros urbanos, podrían requerir soluciones innovadoras para garantizar el acceso a la atención médica. Al reconocer las particularidades de cada región y población, es posible optimizar la distribución de recursos, asegurando que estos lleguen a donde más se necesitan.

La uniformidad en la aplicación de políticas es un enfoque que rara vez tiene éxito en un país tan diverso como México. Para garantizar el bienestar y progreso de todas sus regiones y poblaciones, es esencial adoptar un enfoque adaptativo y específico. Al hacerlo, se toman medidas concretas para abordar los desafíos pendientes de manera efectiva. También, se requiere fomentar la colaboración y la coordinación entre las autoridades estatales y las municipales para asegurar la respuesta eficiente y coherente. Esto incluye compartir datos epidemiológicos, estrategias de prevención y asignación de recursos de manera equitativa. Finalmente, se debe reconocer la importancia de la salud mental y ofrecer apoyo psicológico y social a las comunidades afectadas por la pandemia, especialmente en áreas con altas tasas de defunciones.

Sobre el empleo de análisis de patrones de correlación multivariable

Cuando nos referimos a correlaciones multivariadas, aludimos a las interconexiones y relaciones existentes entre múltiples variables o factores, y cómo su interacción conjunta puede influir en un resultado específico, como lo es la propagación de un virus. Estas variables pueden ser de naturaleza biológica, como la predisposición genética o la existencia de condiciones médicas previas. También, pueden ser de índole social, tales como la concentración demográfica, el movimiento constante de individuos o la facilidad de acceso a servicios de salud. Además, no podemos olvidar las variables ambientales, como las características climáticas o las prácticas de limpieza en un área determinada. Por ello, es importante considerar estas correlaciones multivariadas al momento de planificar y ejecutar intervenciones, abordando la situación desde una perspectiva sistémica.

En el contexto del COVID-19, esta correlación se convierte en una herramienta crucial para comprender la complejidad de la propagación del virus, ya que implica una multiplicidad de factores que pueden influir en su transmisión, severidad y prevención. Aplicar técnicas de aprendizaje automático en el análisis de datos relacionados con el COVID-19 puede ofrecer descubrimientos valiosos, por ejemplo, en el caso de predicción de brotes utilizando datos históricos de propagación del virus, puede prever futuros brotes basándose en patrones previamente identificados. Otra utilidad sería por ejemplo la optimización de recursos: mediante el análisis de datos, se puede prever qué áreas u hospitales podrían tener un aumento en la demanda de recursos, permitiendo una distribución más eficiente de los mismos.

El análisis de información es esencial para cualquier estrategia de combate contra una pandemia. El COVID-19 generó una avalancha de datos, desde la cantidad de casos diarios hasta investigaciones científicas sobre tratamientos. Procesar y compren-

der esta información es básico para otros tipos de análisis y toma de decisiones. Aquí es donde las correlaciones multivariantes permiten interpretar estos datos de manera más profunda y precisa, ofreciendo una visión más clara de la situación y potenciales soluciones.

Finalmente, este libro proporciona una visión detallada de la situación de México entre 2020 y 2022, sirviendo como guía para futuras investigaciones. Se plantean áreas de estudio futuras, como las desigualdades en atención médica, las repercusiones económicas y sociales de la pandemia, y cómo estas afectan diversos aspectos de la sociedad. Por último, para futuros trabajos de investigación que pudieran surgir a partir de este estudio, se proponen las áreas de enfoque siguientes:

- i. ¿Cuáles son las desigualdades en el acceso a la atención médica y su impacto en la concentración de defunciones por COVID-19 en áreas específicas?
- ii. ¿Cuáles son y cómo son las implicaciones económicas y sociales derivadas de la concentración de casos y defunciones por COVID-19 en ciertos estados de México?
- iii. ¿En qué forma estas concentraciones afectan a la economía local, la salud mental de la población, el acceso a la educación y otros aspectos sociales?

REFERENCIAS

- Alcántara-Ayala, I. (2021). "COVID-19, más allá del virus: una aproximación a la anatomía de un pandesastre sindémico". *Investigaciones geográficas*, (104). <https://doi.org/10.14350/rig.60218>.
- Añorve, A. F.C. (2021). "El exceso de Mortalidad por COVID en México: breve descripción de tres metodologías". *Boletín sobre COVID-19*, núm. 19-20-04, Facultad de Medicina, UNAM. Recuperado de <https://dsp.facmed.unam.mx/wp-content/uploads/2022/03/COVID-19-No.19-20-04-El-Exceso-de-Mortalidad-por-COVID-en-Mexico.pdf>
- Bergmann, L. C. (2023). Dominant Modes of Agricultural Production Helped Structure Initial COVID-19 Spread in the US Midwest. *ISPRS International Journal of Geo-Information*, 195.
- Carrillo-Vega, M. F., Salinas-Escudero, G., García-Peña, C., Gutiérrez-Robledo, L. M., & Parra-Rodríguez, L. (2020). Early estimation of the risk factors for hospitalization and mortality by COVID-19 in Mexico. *PloS one*, 15(9), e0238905. <https://doi.org/10.1371/journal.pone.0238905>
- Clarke S. P., and Cossette S. (2000) Secondary analysis: theoretical, methodological, and practical considerations. *Canadian Journal of Nursing Research*. Dec; 32(3):109-129. PMID: 11928128.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- Consejo Nacional de Población [Conapo]. (2019). Proyecciones de la población de los municipios de México 2015-2030. Ciudad de México, 2 de septiembre de 2019. Recuperado de la página de internet del organismo: <https://www.gob.mx/conapo/articulos/proyecciones-de-la-poblacion-de-los-municipios-de-mexico-2015-2030-215756.%20Proyecciones%20de%20la%20Poblacin%20de%20los%20municipios%20de%20Mxico>

- Coria Olguín, S. R., Cruz Pérez, H., Cruz Meléndez, C. A. (2022). *Sitios web municipales de México: Perspectiva interdisciplinaria de computación y administración pública*. Editorial Universidad Autónoma de Aguascalientes. <https://doi.org/10.33064/UAA/978-607-8834-39-6>
- DeSanto, C., Moscatelli, R., Rojas, R., Totoro, M. (2015). *Introduction to Statistics* (10th Edition). Boston, Estados Unidos: Pearson Learning Solutions.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), pp. 37-54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Frank, E. and Witten, I. H. (1998). Generating accurate rule sets without global optimization. (Working paper 98/2). Hamilton, New Zealand: University of Waikato, Department of Computer Science.
- Fuentes, N., González, G., Brugués, A., y Ramírez, A. (2022). Niveles de Confianza del Exceso de Mortalidad en los Estados Mexicanos en 2020. *Economía, sociedad y territorio*, 22(69), 603-630. Epub 16 de octubre de 2022. <https://doi.org/10.22136/est20221788>
- Gobierno de México (2020). Fases o escenarios de contingencia y nivel de propagación del COVID-19. Recuperado de <http://educacionensalud.imss.gob.mx/es/system/files/Fases-COVID19.pdf>
- Gobierno de México (2021). Criterios para las poblaciones en situación de vulnerabilidad que tienen mayor riesgo de desarrollar una complicación o morir por COVID-19 en la reapertura de actividades económicas en los centros de trabajo. Recuperado de https://coronavirus.gob.mx/wp-content/uploads/2021/06/Criterios_Vulnerabilidad_01Jun21.pdf
- Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques*. (3rd. ed.). The Morgan Kaufmann Series in Data Management Systems. Waltham: Elsevier. <https://doi.org/10.1016/B978-0-12-381479-1.00027-7>
- Instituto Nacional de Estadística y Geografía (INEGI). [2021]. Censo de Población y Vivienda 2020. Recuperado de <https://www.inegi.org.mx/programas/ccpv/2020/>
- Knaul, F., Arreola-Ornelas, H., Porteny, T., Touchton M, Sánchez-Talanquer M, Méndez Ó, Chertorivski, S., Ortega, S., Chudnovsky, M., Kuri, P., The Group from the Observatory for the Containment of COVID-19 in the Americas. (2021). *Not far enough: Public health policies to combat COVID-19 in Mexico's states*. PLoS ONE 16(6): e0251722. <https://doi.org/10.1371/journal.pone.0251722>
- Kuhn, M., and Quinlan, R. (2022). C50: C5.0 *Decision Trees and Rule-Based Models*. R package version 0.1.8. Recuperado de <https://topepo.github.io/C5.0/>
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Ichii, H., Zacharski, M. and Khosrawipour, T. (2020). The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. *Journal of Microbiology, Immunology and Infection*, 53(3), 467-472. doi: 10.1016/j.jmii.2020.03.026. Epub 2020 Mar 28. PMID: 32299783; PMCID: PMC7271256.
- Lorenzo, M. M. G., Rodríguez, Y., Hernández, A. R., García, B. B., Filiberto, Y., Rosete, A. y Bello, R. (2020). Adquisición de conocimiento sobre la letalidad de la COVID-19 mediante técnicas de inteligencia artificial. *Anales de la Academia de Ciencias de Cuba*, 10(3), 891.
- Luna-Ramírez, E., Soria-Cruz, J., Ramírez-Báez, R. F., and Díaz de León-Moreno, A. C. (2022). Characterization of SARS-CoV-2 cases and COVID-19 deaths in the State of Baja California through five waves using machine learning. *ECORFAN Journal-Spain*, 9-17. DOI: [10.35429/EJS.2022.17.9.15.20](https://doi.org/10.35429/EJS.2022.17.9.15.20).

REFERENCIAS

- Lusk, J., and Chandra, R. (2021). *Farmer and farm worker illnesses and deaths from COVID-19 and impacts on agricultural output*. PloS ONE, 16. <https://doi.org/10.1371/journal.pone.0250621>
- Méndez Sánchez, L. Y. (2021). *Metodología de aprendizaje semiautomático para el tratamiento de datos biomédicos: caso de estudio COVID-19* (tesis de maestría). Recuperado de <https://hdl.handle.net/20.500.12371/15029>
- Organización Mundial de la Salud [OMS] (2020a). Criterios de salud pública para ajustar las medidas de salud pública y sociales en el contexto de la COVID-19: anexo del documento *Consideraciones relativas a los ajustes de las medidas de salud pública y sociales en el contexto de la COVID-19*. Ginebra, Suiza. Recuperado de https://iris.who.int/bitstream/handle/10665/332169/WHO-2019-nCoV-Adjusting_PH_measures-Criteria-2020.1-spa.pdf
- Organización Mundial de la Salud [OMS] (2020b). *Coronavirus disease (COVID-19) technical guidance: Essential resource planning*. Ginebra, Suiza. Recuperado de <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/covid-19-criticalitems>
- Organización Mundial de la Salud [OMS] (2020c). Alocución del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de mayo de 2020. Recuperada de <https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-may-2020>
- Organización Mundial de la Salud [OMS] (2020d). *Medidas decisivas de preparación, disposición a la acción y respuesta frente a la COVID-19*. Orientaciones provisionales, 4 de noviembre de 2020. Recuperado de https://apps.who.int/iris/bitstream/handle/10665/336884/WHO-COVID-19-Community_Actions-2020.5-spa.pdf
- Pineda, L. S. L., Cruz, D. G. y Páez, C. L. R. (2023). Minería de datos en el estudio de comorbilidades relacionadas a las defunciones ocurridas por COVID-19 en la CDMX durante el 2020. *Revista Aristas* 10(18), 151-156. Recuperado de http://revistaaristas.tij.uabc.mx/index.php/revista_aristas/article/view/289
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <https://www.R-project.org/>.
- Russell, P., and Norvig, S. (2020). *Artificial intelligence: a modern approach* (4th edition). New Jersey, Estados Unidos: Pearson Education, Inc.
- Secretaría de Salud [SSA] (2020a). *Informe Técnico Diario COVID-19*, 31 de diciembre, 2020, 19:00 hs. Recuperado de https://www.gob.mx/cms/uploads/attachment/file/604659/Comunicado_Tecnico_Diario_COVID-19_2020.12.31.pdf
- Secretaría de Salud [SSA] (2020b). *Lineamiento Estandarizado para la Vigilancia Epidemiológica y por Laboratorio de la Enfermedad Respiratoria Viral*, agosto de 2020. Recuperado de https://coronavirus.gob.mx/wp-content/uploads/2020/09/Lineamiento_VE_y_Lab_Enf_Viral_Ago-2020.pdf
- Secretaría de Salud [SSA] (2021). *Informe Técnico Semanal COVID-19*, 31 de diciembre, 2021, 17:00 hs. Recuperado de https://www.gob.mx/cms/uploads/attachment/file/689960/Comunicado_Tecnico_Diario_COVID-19_2021.12.31.pdf

- Secretaría de Salud [SSA] (2022). *Datos abiertos de COVID-19 de México*. Recuperado de <https://www.gob.mx/salud/documentos/datos-abiertos-152127>
- Secretaría de Salud [SSA] (2022b). *Informe Integral de COVID-19 en México*, núm. 11-2022, 05 de octubre de 2022. Recuperado de https://epidemiologia.salud.gob.mx/gobmx/salud/documentos/covid19/Info-11-22-Int_COVID-19.pdf
- Secretaría de Salud [SSA] (2022c). *Informe Técnico Semanal COVID-19*, 27 de diciembre, 2022, 17:00 hs. Recuperado de (https://www.gob.mx/cms/uploads/attachment/file/788069/Informe_Tecnico_Semanal_COVID-19_2022.12.27.pdf)
- Sepúlveda, J. (Coord.). (2024). *Informe de la Comisión Independiente de Investigación sobre la Pandemia de covid-19 en México: Aprender para no repetir*. Recuperado de <https://www.comisioncovid.mx/>
- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. (2017). *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems 4th Edition). Cambridge, Reino Unido: Elsevier.
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13), 1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Zayet, S., Gendrin, V., Klopfenstein, T. (2020). Natural history of COVID-19: back to basics. *New Microbes and New Infections* 2020; 38: e100815. <https://doi.org/10.1016/j.nmni.2020.100815>
- Zhang, Y. (2020). The novel coronavirus pneumonia emergency response epidemiology team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly*, 2(113), 22. [doi:10.46234/ccdcw2020.032](https://doi.org/10.46234/ccdcw2020.032)