

TECANA AMERICAN UNIVERSITY

Postdoctoral Program in Biostatistics



TESIS POSTDOCTORAL

**“ANÁLISIS DE RIESGOS DE CONTAGIO POR COVID-19
EN LA POBLACIÓN CON MAYOR VULNERABILIDAD DE
MÉXICO DEL PERÍODO ABRIL A OCTUBRE DE 2020”**

Cursante: Oralia Nolasco Jáuregui, Eng. D.

TAU Academic Direction: Dr. Jesús R. Rivas, Ph.D.

“Por la presente juro que doy fe de que soy la única autora del presente trabajo de investigación y que su contenido es consecuencia de mi labor, experiencia e investigación académica”.

Septiembre de 2021, Jalisco México.

DEDICATORIA

Al bello equipo que formamos mi esposo y yo.

A mis padres por su innato apoyo.

A mis hermanos por su compañía y constantes alegrías.

Muchas Gracias.

AGRADECIMIENTO

Agradezco al *academic-staff* de *Tecana American University* por su apoyo, guía y constancia en la corrección de puntos a evaluar.

Muchas Gracias.

ÍNDICE GENERAL

DEDICATORIA	II
AGRADECIMIENTO	III
ÍNDICE DE FIGURAS.....	VIII
ÍNDICE DE TABLAS	XI
GLOSARIO	XII
RESUMEN.....	XIV
ABSTRACT	XVI
INTRODUCCIÓN	1
CAPÍTULO 1	3
EL PROBLEMA	3
1.1 Descripción del Problema	3
1.1.1 Enunciado del Problema	3
1.1.2 Interrogante Principal.....	4
1.1.3 Interrogantes Secundarias	4
1.2 Justificación de la Investigación	4
1.3 Objetivos	5
1.3.1 Objetivo General	5
1.3.2 Objetivos Específicos.....	5
1.4 Formulación de las Hipótesis.....	5
1.4.1 Hipótesis Principal	6

1.4.2 Hipótesis Secundarias	6
1.5 Conclusiones	6
CAPÍTULO 2	8
MARCO METODOLÓGICO	8
2.1 Bases Teóricas de la Investigación	8
2.1.1 Tipo de Investigación.....	11
2.2 Alcances	13
2.3 Limitaciones	13
2.4 Matriz de Operacionalización de Variables	14
2.5 Conclusiones	15
CAPÍTULO 3	16
MARCO TEÓRICO REFERENCIAL.....	16
3.1 Antecedentes	16
3.1.1 COVID-19 en México.....	16
3.1.2 COVID-19 en el Mundo.....	17
3.1.3 Infodemia	18
3.2 Marco Referencial.....	22
3.2.1 Mexicanos	22
3.2.2 Etnias en México.....	27
3.2.3 Nivel Económico y urbanización en México	29
3.2.4 Salud en México.....	30
3.3 Índice de Marginación.....	30

3.3.1 Parámetros Índice de Marginación.....	32
3.4 Conclusiones	35
CAPÍTULO 4	37
METODOLOGÍA	37
4.1 Nivel 1 de HSP.....	38
4.1.1 Extracción de los datos.....	39
4.1.2 Interpretación de los datos	39
4.2 HSP: Nivel 2	44
4.3 HSP: Nivel 3	44
4.4 Nivel 4: HSP	47
4.4.1 Diseño de <i>Python Flask Framework</i>	48
4.4.2 Configuración del Entorno de la <i>Flask Framework</i>	49
4.4.3 Validación y Configuración de los <i>Templates</i>	49
4.4.4 Funciones de Configuración	56
4.4.5 WAF en Ejecución	57
4.5 Conclusiones	64
CAPÍTULO 5	67
ANÁLISIS E INTERPRESTACIÓN DE RESULTADOS	67
5.1 Regiones en México y Acrónimos	67
5.2 Jerarquía de las Variables.....	69
5.3 Análisis de las Edades de los Pacientes	69
5.3.1 Resultados de Análisis de las Edades de los Pacientes	70

5.4 Resultados por Comorbilidades en México	71
5.4.1 Porcentajes de Infectados con Comorbilidades en México.....	72
5.5 Análisis de Covarianzas	75_Toc82088070
5.6 Resultados representados con mapa tipo SuperHeat.....	78
5.6.1 Resultados SuperHeat y algoritmo de Machine-Learning	79
5.7 Análisis de Probabilidad de Comorbilidades en México	82
5.8 Resultados Índice Marginal.....	86
5.9 Resultados de Defunciones	87
5.10 Resultados de Hospitalizados en UCI.....	87
5.11 Conclusión	88
CAPÍTULO 6	91
CONCLUSIONES	91
BIBLIOGRAFÍA	96
WEBGRAFÍA	105
ANEXO A.....	109
ANEXO B	115
ANEXO C	116

ÍNDICE DE FIGURAS

Figura 1. Mapa de las variantes lingüistas en México y sus etnias (captura de pantalla de: https://www.gob.mx/cultura/es/articulos/lenguas-indigenas?idiom=es).....	28
Figura 2. Captura de pantalla del archivo correspondiente al día 12 de Abril de los 203 días de análisis. Aquí se observa los campos en datos crudos sin decodificar.....	38
Figura 3. Captura de pantalla la HSP en la sección con lenguaje C para la extracción de los archivos <i>csv</i> y la decodificación de datos utilizando los <i>tokens</i>	40
Figura 4. Captura de pantalla de la HPS en un primer post-procesamiento. La columna G ya tiene una extracción de las entidades federativas de la Región Noroeste, es decir: 2, 3, 5, 25 ó 26.....	41
Figura 5. Representación de las Comorbilidades donde el espacio de los NH pacientes totales es igual a la intersección de los 9 subespacios y el resto contiene el área de No-Comorbilidades-NH.....	45
Figura 6. Cálculo de la Suma de la Probabilidades de los n subconjuntos de las comorbilidades y sus intersecciones en la m región.....	46
Figura 7. Gráfica Bayesiana con gráfico tipo <i>DAG</i> y la función <i>model2network</i> para el cálculo de los valores de la red.....	47
Figura 8. Diagrama a bloques de los elementos para la WAF: 1) Diseño del FrameWork, 2) Configuración del entorno, 3) Validación del Template y 4) WAF en Ejecución.....	47
Figura 9. La flecha roja indica la activación del template que utiliza el método POST de la <i>RegistrationForm()</i> para dar de alta a un nuevo usuario.....	58

Figura 10. La flecha roja indica la activación del template de Login que utiliza el método POST de la LoginForm() para abrir una sesión al usuario registrado.....	59
Figura 11. La flecha roja indica la activación del template de configuración del usuario, como son los campos de mail, username y selección de imagen que utilizan el método POST de la UpdateAccountForm(). En verde aparece la opción de logout().....	60
Figura 12. La flecha roja indica la activación del template de home.html. Aquí la autora publica un mensaje de bienvenida al blog.....	61
Figura 13. La flecha roja indica la activación del template de postting.html que utiliza los métodos de GET y POST de la PostForm() para crear una publicación en el blog.....	62
Figura 14. La función delete_post(post_id) necesita de route tal como: @login_required y tendrá el numero del post adecuado con el método POST: @app.route("/<init:post_id>/delete", methods=['POST']).....	63
Figura 15. La template de reset_request.html la cual redirecciona a la template de login.html con la form de RequestResetForm() con los métodos GET y POST para solicitar el campo password.....	63
Figura 16. En verde se señala el territorio correspondiente a R1, en amarillo a la R2, en azul es R3, en gris R4 y en marrón el territorio de la R5.....	67
Figura 17. Red de Jerarquías en la extracción de los datos indicando con colores sus diferentes niveles en orden descendente: naranja, verde, azul, gris y amarillo (captura de pantalla extraída de Nolasco-Jáuregui, O., Quezada-Téllez, L., Rodríguez-Torres, E., y et al., 2021).....	68
Figura 18. Distribuciones de Edades (0-100 años) en la R4 por 10 días.....	70

Figura 19. Análisis de las 9 comorbilidades registradas en las 5 Regiones de México en Paciente en Cuarentena.....	73
Figura 20. Análisis de las 9 comorbilidades registradas en las 5 Regiones de México en Paciente Hospitalizados.....	73
Figura 21. Submatriz con la correlación negativa de las comorbilidades en las regiones de México.....	77
Figura 22. Submatriz con la correlación positiva en las 5 regiones de México.....	77
Figura 23. Mapa de de densidad de pacientes correlacionado con comorbilidades y regiones en México.....	84
Figura 24. Salida de la red para los pacientes Hospitalizados y sus coeficientes de probabilidades para las 5 regiones de México en 203 días de registro con COVID-19.....	85
Figura 25. Salida de la red para los pacientes No-Hospitalizados y sus coeficientes de probabilidad para las 5 regiones de México en 203 días de registro con COVID-19.....	85

ÍNDICE DE TABLAS

Tabla 1. Matriz de operacionalización de Variables.....	15
Tabla 2. Indicadores socioeconómicos para el análisis del IM por entidad federativa en México para el 2015 (CONAPO, 2016).....	34
Tabla 3. Porcentajes de los pacientes COVID-19 por cada región de México sin comorbilidades.	72
Tabla 4. Diccionario de los datos COVID-19 en México con los 35 descriptores y la interpretación de sus campos.....	114
Tabla 5. Conteos totales divididos por Regiones, tipo de pacientes y Comorbilidades en 203 días de registro con pandemia en México.....	115
Tabla 6. Estructura del diccionario que se crea cada vez que se registra un nuevo Usuario.	116
Tabla 7. Estructura del diccionario que se crea cada vez que se registra una nueva Publicación.	116

GLOSARIO

<i>Big- Data</i>	Término traducido como macrodatos (datos masivos) inteligencia de datos y datos a gran escala. Este término hace referencia a conjuntos de datos tan grandes y complejos que necesitan de aplicaciones informáticas no tradicionales.
<i>Ciencias Fáticas</i>	A diferencia de las ciencias formales, estas ciencias se fundamentan en la percepción u observación del mundo exterior, tiene una referencia empírica y se ciñe a los hechos.
<i>COVID-19</i>	El COVID-19 es un nuevo virus detectado del tipo coronavirus; el coronavirus es una familia extensa de virus que puede causar enfermedad tanto en animales como en seres humanos. En el caso de los seres humanos, hasta donde se conoce varios de estos coronavirus inducen desde un resfriado común hasta enfermedades respiratorias graves como el Síndrome Respiratorio de Oriente Medio (MERS) y el Síndrome Respiratorio Agudo Severo (SARS).
<i>Flask</i>	Es un framework minimalista escrito en Python que permite crear aplicaciones y con un mínimo número de líneas de código.
<i>Framework</i>	En software, es un entorno de trabajo que tiene un conjunto estandarizado de conceptos, prácticas, propiedades, eventos, métodos y son pensados para resolver una problemática.
<i>Machine-Learning</i>	Término traducido como aprendizaje automático, aprendizaje automatizado ó aprendizaje de máquinas.

Es un subcampo de las ciencias de la computación y una rama de la inteligencia artificial.

Meta-análisis

Es un método cuantitativo que combina los resultados de estudios independientes (usualmente extraído de ensayos ya publicados) para agruparlos según un tema en específico y evaluarlos a través de herramientas de calidad metodológica.

Python

Es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código.

Template

Es un tipo de plantilla que tiene una interfaz estándar y que se puede modificar. Su contenido y estructura permite interactuar con un sistema ó con un usuario; suele tener un tema y diseño definido.

TECANA AMERICAN UNIVERSITY

“Análisis de Riesgos de Contagio por COVID-19 en la Población con Mayor Vulnerabilidad de México del Período Abril a Octubre de 2020”

TESIS POSTDOCTORAL

Autora: Dra. Oralia Nolasco Jáuregui

Septiembre, 2021

RESUMEN

Dentro de la hipótesis que motiva a la realización de esta investigación es la vulnerabilidad y predisposición en la salud de la población mexicana es potencialmente la causa que México se encuentre en la lista de los 10 países con mayor número de contagios y es uno de los 5 países con mayor número de muertes por COVID-19 en todo el mundo.

El objetivo del presente estudio es analizar el riesgo de contagio por COVID-19 en la población con mayor vulnerabilidad de México del periodo Abril a Octubre del 2020 (1,089,852 pacientes).

Las características del método son del tipo aplicado y según su nivel de profundización en el objeto de estudio se enmarca en una investigación del tipo análisis descriptivo, correlacional y explicativo. Los datos empleados son cuantitativos y son resultado de un instrumento cuestionario constituido por 34 campos y prueba al virus. Se ha elegido un método inductivo ya que parte del razonamiento de una serie de hechos particulares en la sociedad mexicana por cada región del país y con ello se puede hacer una conclusión general.

De los resultados se obtiene que las regiones R4 con el mayor número de pacientes contagiados tanto del tipo Hospitalizados como del tipo No-Hospitalizados, seguida de R5 en hospitalizados y R3 en ambulatorios. La edad promedio en pacientes ambulatorios es alrededor de 40 años y de los hospitalizados es cercano a los 53 años. Las comorbilidades más sensibles en

los pacientes Hospitalizados son tres: la EPOC, Renal y Otras Enfermedades. Los pacientes Hospitalizados que ha requerido un respirador mecánico es alrededor de 1.4% a nivel nacional.

Como conclusión hay un mayor riesgo de perder la vida en R4 y R5. Mientras que la R2 es la que registra el menor riesgo de perder la vida de todo el territorio nacional.

PALABRAS CLAVES: Pandemia, COVID-19, México, Población Vulnerable, Probabilidad, Comorbilidades, Macrodatos, Aprendizaje Máquina, Aplicación Web, mapas SuperHeat, Infodemia.

ABSTRACT

The hypothesis that motivates the realization of this research is the vulnerability and predisposition in the health of the Mexican population is potentially the cause that Mexico is on the list of the 10 countries with the highest number of infections and is one of the 5 countries with the highest number of deaths due COVID-19.

The objective of this study is to analyze the risk of contagion by COVID-19 in the most vulnerable population of Mexico from April to October 2020 (1,089,852 patients).

The characteristics of the method are the applied type and according to its level of deepening in the object of study it is framed in an investigation of the descriptive, correlational, and explanatory analysis type. The data used are quantitative and they are the result of a questionnaire instrument consisting of 34 fields and the virus test. An inductive method has been chosen since it starts from the reasoning of series of particular facts in Mexican society for each 5 regions of the country and with this analysis a general conclusion can be made.

The results show that the R4 has the highest number of infected Hospitalized and ambulatories patients, followed by R5 in hospitalized patients and R3 with NH patients. The average age in Non-Hospitalized patients is around 40 years old, while in hospitalized patients it is close to 53 years old. The most sensitive comorbidities in hospitalized patients are three: COPD, kidney disease and other diseases. The averages of Hospitalized patients who have required a mechanical respirator was 1.4% in the country.

In conclusion, there is a greater risk of losing life in the R4 and R5 regions than in the R2 region.

KEYWORDS: Pandemic, COVID-19, Mexico, Vulnerable Population, Probability, Comorbidities, Big-Data, Machine-Learning, Web Application, SuperHeat maps, Infodemic.

INTRODUCCIÓN

Hasta el último reporte de casos de contagios de COVID-19 en el mes de octubre del año 2020 realizado por la Universidad de John Hopkins (JHU por sus siglas en inglés), los Estados Unidos Mexicanos (México) se encuentra en la lista de los 10 países con mayor número de contagios a nivel mundial y es uno de los 5 países con mayor número de muertes por COVID-19 en todo el mundo (JHU, 2021; Ritchie, H., Ortiz-Ospina, E., Beltekian, D., y et al., 2021; ONU, 2021).

La información anterior fuerza a un estudio que analice el riesgo regional y la probabilidad de adquirir el virus por la población mexicana más vulnerable. Es muy importante puntualizar que esta investigación surge por la necesidad de una personalización (Holmes, S., y Huber, W., 2018) ya que cada país tiene una infinidad de variables que no son compartidas entre ellos.

El objetivo de esta investigación es analizar el riesgo de contagio por COVID-19 en la población con mayor vulnerabilidad en las 5 regiones geográficas de México en un periodo de 203 días de pandemia.

Se tiene un detallado análisis con múltiples herramientas que permitan desde diferentes perspectivas comparar el fenómeno (Zhou, D., Zhu, S., Yu, K., y et al., 2008), su probabilidad de contagio y riesgo de la población mexicana dependiente de su vulnerabilidad (Índice de Marginación).

Por la necesidad de difusión de la información correcta y apropiada, surge la necesidad de engendrar una WAF (de sus siglas en inglés Web App Framework) con un carácter de orden científico y sólo se alimentará de fuentes oficiales y fidedignas (Stackoverflow, 2020; INER, 2021). Es una realidad que se retroalimentará de la información que los ciudadanos aporten ya que tendrá un servicio de tipo público. Pero sólo usuarios registrados y verificados por un humano podrán utilizar el servicio de *post* (proviene del inglés y se traduce

como publicación). Para el contenido malicioso se utilizan análisis de cadenas de caracteres por medios de *tokens* (proviene del inglés y se traduce como ficha) que activan alarmas al administrador de la WAF.

Este documento le brinda al lector el diagrama de flujo inédito (ver Figura 8 en la sección 4.4), dentro de la HSP (Herramienta Propia de Software) donde el pseudocódigo es descrito a detalle para que cualquier ciudadano que tenga cierto conocimiento en programación de software será capaz de reproducir, diseñar y programar esta WAF.

El lector no debe preocuparse por la sintaxis del *scripting* (proviene del inglés y se traduce como lenguajes de programación en secuencia de comandos) de Python 3 ni por el HTML ya que son funciones descritas detalladamente en este documento para una WAF. Este documento lleva de la mano al lector y muestra en un tipo diagrama de flujo los pasos que nos ayudaron para generar el producto final que es la WAF (e incursionar a aquellos que requieran un producto similar).

El escrito está organizado de la siguiente manera: en el CAPÍTULO 1 se hace uso del método científico y se enuncia el planteamiento del problema, justificación, objetivos, alcances y limitaciones. En el CAPÍTULO 2 se describe un breve el tipo de investigación. En el CAPITULO 3 se presenta el histórico del panorama a la llegada del COVID-19 a México. En el CAPÍTULO 4 se asienta un diagrama general del diseño de la herramienta HPS y la metodología de la investigación. En el CAPÍTULO 5 se presentan los análisis y resultados obtenidos de los datos. En la última sección CAPÍTULO 6 se presentan las conclusiones de esta investigación.

CAPÍTULO 1

EL PROBLEMA

1.1 Descripción del Problema

Es importante destacar la situación de riesgo previo en México ante la llegada de la pandemia. En el 2018 en México se registraron 722 611 defunciones de las cuales el 56.4% fueron hombres, el 43.5% mujeres y en 376 casos el sexo no fue especificado. Del total de defunciones el 88.4% se debieron a enfermedades relacionadas con la salud. Mientras que 11.6% fueron por otras causas, principalmente accidentes, homicidios y suicidios.

Las tres principales causas de muerte tanto para hombres como para mujeres son las enfermedades del corazón, la diabetes mellitus y los tumores malignos. Los homicidios representan la cuarta causa de muerte en hombres (INEGI, 2018).

Es necesario recordar que México se divide en 5 regiones por sus características geográficas y sobre todo culturales: la región Noroeste (R1) establecida por 5 estados, la región Noreste (R2) que agrupa a otros 5 estados, la región Centro (R3) formada por 6 estados, la región Occidente (R4) con 8 estados y la región Sureste (R5) que abarca 7 estados.

1.1.1 Enunciado del Problema

Cada región tiene condiciones climáticas, costumbres y orígenes étnicos diferentes. En México hay un registro de más de 68 pueblos indígenas además del mestizaje, es por ello que surge la pregunta: Dada la exposición, comorbilidades (Bansal, M., 2020), edades y nivel económico en la sociedad mexicana. El enunciado del problema viene planteado de la siguiente manera: ¿Cuál es la probabilidad de contagio para la población dependiente de las

condiciones actuales en la sociedad Mexicana?, ¿Es la misma relación? Y si hay variabilidad ¿Cuál es la relación cuantitativa entre las relaciones?. El periodo de estudio abarca casi 203 días de registro con más de un millón de pacientes.

1.1.2 Interrogante Principal

-¿Cuáles son los riesgos de contagio con COVID-19 en la población con mayor vulnerabilidad de México del periodo Abril a Octubre 2020?

1.1.3 Interrogantes Secundarias

Con base a lo anterior, surgen las siguientes interrogantes:

- ¿Cuál es la probabilidad de riesgo de contagio en las 5 regiones de México?,
- ¿A qué se debe la variabilidad en las probabilidades de contagio a lo largo del territorio nacional? ,
- ¿Por qué las comorbilidades de los pacientes son factor en las zonas de mayor contagio? y
- ¿Se valúa alguna relación entre las edades de los pacientes, nivel económico y las zonas de mayor contagio?.

1.2 Justificación de la Investigación

La justificación más importante de esta investigación es analizar y brindar la información *nominada* necesaria para enfrentar la terrible pandemia que azota a México por todas y cada una de sus regiones.

Dentro de las aportaciones más significativas de este estudio se encuentran el análisis de los datos con una herramienta propia que presente de manera sintetizada las probabilidades de riesgos de contagio por regiones a la población en general.

La presentación de la información por regiones es con el fin de mostrar los datos en pirámide invertida, es decir, de general a lo particular. De manera digerible, amable y de carácter científico a los lectores que consulten la información.

Se pretende hacer pública la herramienta para que cualquier civil que desee saber cuál es la probabilidad de contagio en la zona que habita, sea de manera personalizada y por cada situación que se vive en su entorno.

1.3 Objetivos

1.3.1 Objetivo General

Analizar el riesgo de contagio por COVID-19 en la población con mayor vulnerabilidad de México del periodo Abril a Octubre del 2020.

1.3.2 Objetivos Específicos

- Definir las probabilidades de contagio por cada región del país.
- Caracterizar las probabilidades de contagio.
- Relacionar la probabilidad de contagio y las comorbilidades de los pacientes por cada región de México.
- Conocer las edades y las probabilidades de contagio de los pacientes por cada región.

1.4 Formulación de las Hipótesis

La vulnerabilidad en la población mexicana es potencialmente la causa que México se encuentre en la lista de los 10 países con mayor número de contagios y es uno de los 5 países con mayor número de muertes por COVID-19 en todo el mundo.

Dada la exposición al virus, comorbilidades, edades, acceso al sector salud y nivel económico en la sociedad mexicana; es innegable una alta probabilidad de contagio al virus del COVID-19 por las complicaciones que puede traer a los

pacientes con un precedente de comorbilidad (Valente-Acosta, B., Hoyo-Ulloa, I., Espinosa-Aguilar, L., y et al., 2020).

Dependiente de la morbilidad es el riesgo de contagio y adjunto a ello la edad del paciente son factores fundamentales para indagar cómo va a ser la lucha contra el virus. Lo anterior es una candidatura a ser hospitalizado, necesidad de un respirador artificial e incluso perder la vida ante el contagio. No es la misma relación entre un paciente adulto con padecimientos precedentes al virus a un joven sano.

1.4.1 Hipótesis Principal

Las comorbilidades precedentes, edad y condición socioeconómica de los pacientes los convierte en candidatos con alto riesgo de contagio y por ende vulnerables ante el COVID-19.

1.4.2 Hipótesis Secundarias

- La posibilidad de contagio es definible e identificable.
- Con la caracterización del riesgo de contagio se comprende la relación que con la probabilidad de contagio y la vulnerabilidad.
- La valuación del riesgo de contagio se brinda conocimiento para evitar los contagios en México.

1.5 Conclusiones

Este estudio se puede tomar como una vía de información confiable apegada al método científico que pretende facilitar la vida de los habitantes de México cuya situación de pandemia ya es por ende complicada.

Esta investigación logra un aporte al estado del conocimiento de tal manera que favorece y contribuye a la difusión de la información sin fines de lucro. Esta investigación toma tecnologías actuales para su desarrollo. Tiene una sustentación cien por ciento científica, utiliza variables que son únicas del país y las enmarca un entorno humanista.

La presentación de los datos por región es una de las mejores vías de interpretación visual ya que es rápida y muy amigable para el lector. La división de regiones en México ya fue indagada y cavilada de manera estratégica por compartir características internas entre las etnias (ver Figura 16).

La formulación de las hipótesis secundarias las que permiten canalizar el enfoque hacia un fin particular. En este caso es la *caracterización, estimación y relación* de la probabilidad de contagio. Son las comorbilidades, nivel socioeconómico, acceso a la salud, origen étnico, precedentes de la población de las 5 regiones de México los estimadores de probabilidad al contagio y se pueden cuantificar. En la técnica de la metodología por su nivel de profundización en el objeto de estudio se refiere a la del tipo cuantitativo ya que comienza con el análisis exploratorio, descriptivo, correlacional y meta-análisis (ver CAPITULO 4).

CAPÍTULO 2

MARCO METODOLÓGICO

2.1 Bases Teóricas de la Investigación

A través del método científico se busca observar, describir, explicar y predecir un fenómeno (Ladrón de Guevara, L., 1981). Para ello se parte de técnicas que permitan recopilar información del fenómeno que posteriormente habrá de ser analizadas para emitir conclusiones o enunciados (Ramírez-González, A., 2006).

Kerlinger en 1982 define: “La investigación científica es una investigación sistemática, controlada, empírica y crítica; de proposiciones hipotéticas sobre las supuestas relaciones que existen entre fenómenos naturales” (Adwere-Boamah, J., Delay, D., y Jones, O., 1982).

La investigación científica responde en esencia al método científico y por tanto implementa una serie de procedimientos diversos pero ordenados que inician con una pregunta surgida generalmente dentro del campo de dominio del investigador y desde una base teórica particular (Álvarez, C., y Sierra, V., 1995). Tal pregunta lleva a la formulación de una hipótesis y de ahí da paso al diseño de un programa de toma, organización y análisis de información. Para finalizar con las conclusiones o enunciados sobre la pregunta (Gómez, M., 2006). Este proceso implica la elección de instrumentos de medición o la creación de los mismos (Ramírez-González, A., 2006).

Alberto Ramírez-Gonzales declara que “Existe poca claridad entre las clases, los métodos y las técnicas de investigación. Lo que lleva a una gran transposición conceptual entre unos y otros”. Hay varios autores que han trabajado estas diferencias López de Lerma, G. en 2015, Miller, D. y Salkind,

N. en 2015, Cuevas-Romo, A., Hernández-Sampieri, R., Leal-Pérez, et al. en 2016, Muñoz-Razo, C. en 1998 y Tamayo y Tamayo, M. en el año 1997, entre otros.

Carlos Muñoz Razo en 1998 define tres diferentes tipos de investigación: 1) *Documental o teórica*: “se centra exclusivamente en la recopilación de datos existentes en forma de documental y su propósito es profundizar en las teorías (aportes ya emitidos sobre un tópico de estudio)”;

2) *De campo o práctica*: “se basa en información primaria (que será colectada directamente de la fuente) y requiere en esencia de un instrumento para ello”;

y 3) *Documental y de campo*: “la recopilación y el tratamiento de datos se conjuga en forma más o menos equitativa a fin de generar información nueva y consolidarla con la obtenida con otros autores”.

Mario Tamayo y Tamayo en 1997 describen a las ciencias *fácticas* como aquellas que se fundamentan en la percepción u observación del mundo exterior. Su razonamiento es en esencia inductivo por lo que va de lo particular a lo general. Su propósito es la verificación, lo que lleva a la generación de un conocimiento temporal. Entonces este conocimiento se acepta como verdad hasta que nuevas técnicas, explicaciones con fenómenos más coherentes o ampliaciones del fenómeno se reconoce como un caso singular que no cumple la regla.

Las características más importantes del método científico son: 1) de *analogías*, 2) *inductivo-deductivo* y 3) *análisis-síntesis*.

López-Yepes en 1990 diferencia tres tipos de *analogías*: de *semblanza*, es cuando hay una diferencia indiscutible; de *modulación*, cuando hay diferencias con una significación distinta; de *proporcionalidad*, cuando los elementos comunes dan diferencias con intensidades varias.

La *inducción* es un modo de razonar que consiste en sacar de los hechos particulares una conclusión general y también analiza una porción de un todo.

Por lo tanto, va de lo particular a lo general. En cambio, la *deducción* es un método de razonamiento que parte de lo general aceptando como válido hacia aplicaciones particulares (Ramírez-González, A., 2006).

Ladrón de Guevara en 1981 expresa que el *método inductivo* puede ser clasificado en uno de los siguientes cinco tipos de métodos: 1) de *concordancias*: “si dos o más casos en que se produce un fenómeno sólo poseen en común una cierta circunstancia”; 2) de *diferencias*: “si dos casos en uno el fenómeno que se estudia aparece y en el otro no”; 3) *combinado* de concordancias y diferencias: “si el fenómeno que se investiga aparece en dos o más casos los cuales concuerdan por poseer circunstancias en común; y dicho fenómeno, no aparece en dos o más casos”; 4) de *variaciones concomitantes*: “si determinados cambios de un fenómeno van seguidos siempre de determinados cambios en otro fenómeno (el primero es la causa) y parte de la causa o condición necesaria del segundo”; y 5) de *residuos*: “si un fenómeno complejo se sustrae (lo que es efecto de una parte de las circunstancias) el resto de dicho fenómeno ha de ser efecto de las circunstancias restantes”.

Ladrón de Guevara en 1981 expresa que *las circunstancias* puede ser clasificadas en una de las siguientes tres tipos: 1) *condiciones contribuyentes*: “son aquellas que no asegurando la aparición de un fenómeno sino que contribuyen a que aumente la probabilidad de que ocurra”; 2) *condiciones alternativas*: “son las que surgen como alternativa a las condiciones contribuyentes”; y 3) *condiciones contingentes*: “son aquellas que permiten que un factor sea contribuyente de otro. Por ejemplo, en sitios con A y B hay mayor probabilidad de C” (Ladrón de Guevara, L., 1981).

Fernández-Collado, C., Baptista-Lucio, P. y Hernández-Sampieri, R. en 2014 definen que la investigación según el *objeto de estudio* se divide en dos: 1) investigación *pura* o *teórica* y 2) investigación *aplicada*. Además explica que de acuerdo a la profundización en el objeto de estudio una investigación puede caracterizarse como básicamente exploratoria, descriptiva, correlacional o

explicativa; pero no debe situarse únicamente una como tal. Aunque un estudio sea en esencia exploratorio contendrá elementos descriptivos; o bien un estudio correlacional incluirá componentes descriptivos y lo mismo ocurre con los demás alcances. Entonces, es posible que una investigación se inicie como exploratoria o descriptiva y después llegue a ser correlacional y explicativa.

Para Buendía-Eisman, L., Expósito-López, J. y Sánchez-Martin, M. en 2012, las situaciones en las que se hace el muestreo se deben de clasificar y ellos los dividen en: 1) *muestreo deliberado*: los sujetos de la prueba poseen características necesarias para la investigación. Los resultados son difíciles de generalizar, un ejemplo son los enfermos de SIDA; *del tipo accidental o causal*: sujetos que casualmente se encuentran en el lugar y momento decidido por el investigador, por ejemplo: pacientes hospitalizados la noche de navidad; y *del tipo voluntarios*: por decisión de los sujetos se someten de forma voluntaria a una investigación (generalmente firman un documento donde aseguran que fue por voluntad propia).

Fernández-Collado, C., Baptista-Lucio, P., y Hernández-Sampieri, R. en 2014 describen los *instrumentos del tipo secundario*. Son aquellos datos recolectados por otros investigadores, cuestionarios, revisión de documentos, registros públicos tanto en físico y electrónicos. Por ejemplo si la hipótesis fuera: “la violencia en la Ciudad de México es mayor que en Caracas”. Entonces se acudiría a la base de datos del gobierno federal en la página oficial para descargar los datos de acceso público relacionados con la violencia, asaltos, violaciones, robos a casa-habitación, etc.

2.1.1 Tipo de Investigación

Por lo anterior se asevera que esta investigación es tipo práctica y tiene una inferencia fáctica, es decir, que tiene una preferencia por un marco empírico y se ciñe en los hechos. Esta investigación es procedente de un método inductivo y por verificación. Se ha elegido un método inductivo ya que parte del

razonamiento de una serie de hechos particulares en la sociedad mexicana por cada región del país y con ello se puede hacer una conclusión general. Esta investigación es práctica de tipo secundario y tiene una inferencia fáctica procedente de un método inductivo enfatizando en las variaciones concomitantes por cada región del país y por verificación de variables cuantitativas. Las condiciones de los datos empleados son del tanto del ejemplar necesarias y contingentes. Las técnicas del método empleadas son cuantitativas y longitudinales.

La documentación del método va de lo particular a lo general, bajo un estricto apego al método científico y con una correlación por verificación de variables cuantitativas y semi-cuantitativas.

Haciendo una profundización en el método inductivo se ha seleccionado el de tipo concomitante que se focaliza en determinados cambios de un fenómeno y van seguidos siempre de una serie de cambios en otro fenómeno, es decir, el primero es la causa y condición necesaria para el segundo. Las reglas seleccionadas en este estudio para el método inductivo son condiciones tanto necesarias (sin ellas no ocurre el fenómeno) y contingentes (favorecen al fenómeno).

Por la profundización en el objeto de estudio esta investigación es en esencia un análisis exploratorio y contiene elementos descriptivos. Pero, a lo largo de la metodología se realizan estudios del tipo correlacional e incluye componentes descriptivos por cada etapa. Entonces, esta investigación se inicia como exploratoria o descriptiva y después llega a ser del tipo correlacional y explicativa.

El tipo de muestreo es deliberado ya que los sujetos no eligieron ser parte del muestreo son sus características (positivo al virus) que poseen que los hacen forzosos para la investigación. La técnica seleccionada es longitudinal ya que estudia el comportamiento del grupo de sujeto de estudio (1, 089,852 pacientes)

por un periodo que abarca desde el 12 abril 2019 hasta el 31 de octubre de 2020.

2.2 Alcances

- *Espacial:* la población objeto de estudio está constituida por los pacientes que dieron positivo a COVID-19 en las 5 regiones del territorio nacional.
- *Longitudinal:* los datos empleados pertenecen a la población objeto de Abril 2019 a Octubre del año 2020.
- *Temporal:* el tiempo de investigación de la presente tesis de posgrado será establecido según el cronograma de actividades, con una duración de 6 meses desde el mes de marzo hasta el mes de agosto del 2021.
- *Marco Institucional:* los datos utilizados son de acceso libre y se encuentran en la página web del gobierno federal en la sección de la secretaría de Salud en México (DGE, 2021).

2.3 Limitaciones

Datos: la población objeto de estudio está en constante movimiento, es decir, los avances en el conocimiento científicos sobre el COVID-19 han mejorado con respecto a los primeros meses de la pandemia.

- Debe incluirse el hecho de los medicamentos (los cuales fueron seleccionados por los médicos en su momento para contener el virus) han sido diferentes a lo largo de estos 203 días de registro. Debido a lo anterior, es que hay fenómenos inmersos en el análisis que es probable que no sean evidentes.
- También hay una serie de mutaciones del virus que dependen de la región donde fueron descubiertos y de la reacción de los habitantes (Chen, A., Altschuler, K., Zhan, S., y et al., 2021).
- Incluso si se toma en cuenta el hecho de que se esperan nuevos fenómenos relacionados con la aplicación de la vacuna, número de dosis necesarias, infectados ya vacunados y un gran porcentaje de la población que

no quiere vacunarse, entre otros (Chen, J., Gao, K., Wang, R., y Wei, G. W., 2021).

2.4 Matriz de Operacionalización de Variables

Las características de los indicadores en la matriz de operacionalización (ver Tabla 1) son en su gran mayoría de variables (tipo de pacientes, comorbilidades, edades y regiones en México) de carácter cuantitativas, seguidas de las variables de tipo descriptivas, deductivas y longitudinales.

Los datos empleados son del tipo cuantitativos y semi-cuantitativos; ya que son resultado de un instrumento cuestionario constituido por 34 campos y prueba al virus (ver Anexo A).

Entre la gama de técnicas se encuentran las correlaciones, correlogramas, dendogramas, covarianzas y el método Bayesiano.

El uso de la red bayesiana (Nagarajan, R., Scutari, M., y Lébre, S., 2013; Scutari, M., y Denis, J. B., 2014) se hace con el objetivo que sea una técnica de tal manera que la obtención de las probabilidades sea dinámica y automatizada.

Dentro de la presentación de los datos se encuentra técnicas como curvas de distribución (Cowell, F., y Flachaire, E., 2015), mapas de densidades, mapas de calor, gráficos de cajas que involucra variables de estadística descriptiva como la media y el promedio.

Problema	Objetivos	Variables	Instrumento	Indicador
	Definir las probabilidades de contagio por cada región del país.	-Pacientes Hospitalizados -Pacientes Ambulatorios -5 Regiones	-Red Bayesiana Estática -Herramienta propia de extracción de datos	-Cuantitativo -Longitudinal
	Caracterizar las probabilidades de	-Pacientes Hospitalizados	-Correlación entre	-Cuantitativo -Longitudinal

¿Es la misma probabilidad de riesgo de contagio en las 5 regiones de México?	contagio por cada región.	-Pacientes Ambulatorios -5 Regiones	Probabilidades -Gráfico de Cajas	-Descriptivo -Deductivo
	Relacionar la probabilidad de contagio y las comorbilidades de los pacientes por cada región.	-Pacientes Hospitalizados -Pacientes Ambulatorios -5 Regiones -Nivel económico	-Covarianzas -Correlograma -Gráfico de Cajas	-Cuantitativo -Longitudinal -Descriptivo -Deductivo
	Conocer las edades y probabilidades de contagio de los pacientes por cada región.	-Pacientes Hospitalizados -Pacientes Ambulatorios -5 Regiones -Edades	-Curva de distribución	-Cuantitativo -Longitudinal -Descriptivo -Deductivo

Tabla 1. Matriz de operacionalización de Variables.

2.5 Conclusiones

Esta investigación tiene una preferencia por un marco empírico y se ciñe en los hechos. Además que la selección del método inductivo del tipo concomitante, ratifica que la preexistencia de comorbilidades y características particulares en la sociedad mexicana genera cambios de un fenómeno como es la pandemia por el virus del COVID-19. Es debido a lo primero que México es vulnerable ante el virus condición necesaria para el segundo fenómeno (México encabeza las listas a nivel mundial en contagios y muertes).

Por la profundización en el objeto de estudio esta investigación es en esencia un análisis exploratorio y contiene elementos descriptivos. Pero a lo largo del análisis de los datos que se realizan estudios del tipo correlacional e incluye componentes descriptivos por cada etapa. Esta investigación se inicia como exploratoria o descriptiva (ver CAPITULO 3) y después llega a ser del tipo correlacional y explicativa (ver CAPITULO 5).

CAPÍTULO 3

MARCO TÉORICO REFERENCIAL

3.1 Antecedentes

3.1.1 COVID-19 en México

Para el 28 de febrero el Instituto de Diagnóstico y Referencia Epidemiológica "Dr. Manuel Martínez Báez" (InDRE, 2020) confirmó el primer caso de COVID-19 en México (Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R., 2020).

Dando seguimiento se encontraron otros cuatro casos de personas que habían viajado a Italia de los cuales tres de ellos habían presentado síntomas leves. Dos de estos pacientes permanecieron en la Ciudad de México y uno en Sinaloa. El cuarto paciente no desarrolló síntomas por lo que sólo quedó como portador. Quizá este fue el primer paciente asintomático registrado en México (Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R., 2020).

Los días siguientes se presentaron nuevos casos positivos al padecimiento. El 29 de febrero de 2020, se reportó un caso positivo en una paciente de sexo femenino que regresó a Torreón Coahuila también procedente de Italia. Sus síntomas son leves y se quedó en casa para cuidados domiciliarios.

El siguiente caso que fue confirmado ocurre en Tuxtla Gutiérrez Chiapas a una joven de 18 años con síntomas leves. También llegó de Italia y era contacto del caso de Torreón Coahuila. Por lo que al primero de Marzo del 2020 todos los casos en México eran importados (Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R., 2020).

3.1.2 COVID-19 en el Mundo

La ciudad de Wuhan en China fue el centro de atención mundial debido a un nuevo brote de enfermedad respiratoria denominada COVID-19 (Jordan, R., Adab, P., y Cheng, K., 2020). En diciembre del 2019, en la provincia de Hubei en el mercado mayorista de mariscos de Huanan fue el centro del desarrollo de esta nueva epidemia. La Organización Mundial de la Salud (OMS) fue notificada por las autoridades Chinas hasta el 31 de diciembre del 2019 sobre este nuevo padecimiento (Hui, D., Azhar, E., Madani, T., y et al., 2020).

El Coronavirus es de una familia extensa de virus que puede causar enfermedad tanto en animales como en los seres humanos (Fang, L., Karakiulakis, G., y Roth, M., 2020). En el caso de los seres humanos hasta donde se conoce varios de estos coronavirus inducen desde un resfriado común hasta enfermedades respiratorias graves como el Síndrome Respiratorio de Oriente Medio (MERS) y el Síndrome Respiratorio Agudo Severo (SRAS) (Hui, D., Azhar, E., Madani, T., y et al., 2020).

El 11 de marzo ante el número de contagios de 118,000 en 114 países y con 4,291 personas fallecidas la OMS declaró oficialmente como pandemia al COVID-19 (ONU, 2021). El 16 de Marzo del año 2020 en México la Secretaría de Salud anunció la implementación de la sana distancia como medida preventiva (DGE, 2021). A partir de esta medida se suspendieron temporalmente la totalidad de actividades salvo en algunos casos, las actividades prioritarias como salud, seguridad, entre otras (Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R., 2020).

Una de las principales medidas impuestas por el gobierno fue que la población se mantuviera en casa y con ello evitar el incremento de contagios. Entre los sectores prioritarios a proteger fueron las personas adultas así como aquellos con padecimientos crónicos como enfermedades cardiovasculares, la diabetes, las enfermedades respiratorias crónicas, la hipertensión y el cáncer. Estas

enfermedades generan un alto riesgo de complicación entre los pacientes que contraían el virus (Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R., 2020).

3.1.3 Infodemia

El término Infodemia ha sido utilizado para resaltar los peligros de los fenómenos de desinformación durante el manejo de una pandemia (Eysenbach, G., 2020; ONU, 2021; Mendoza, M., Poblete, B., y Castillo, C., 2010). Incluso se ha demostrado que la infodemia puede acelerar el proceso contagioso al influir y fragmentar la respuesta social (Starbird, K., Maddock, J., Orand, M., y et al., 2014).

En el caso de la epidemia de COVID-19 muestra el impacto crítico de la desinformación (Kim, L., Fast, S. M., y Markuzon, N., 2019). La difusión de información puede influir fuertemente en el comportamiento de las personas y alterar la eficacia de las contramedidas implementadas por los gobiernos (Shaman, J., Karspeck, A., Yang, W., y et al., 2013). A este respecto, los modelos para pronosticar la propagación del virus COVID-19 están comenzando a tener en cuenta la respuesta conductual de la población con respecto a las intervenciones de salud pública y la comunicación dinámica detrás del consumo de contenido (Viboud, C., y Vespignani, A., 2019).

La propagación mundial del coronavirus se ve afectada por la propagación de información errónea, esto hace que las poblaciones sean más vulnerables a la enfermedad a través de la resistencia a los esfuerzos de mitigación de los gobiernos (Buchanan, M., 2020).

M. Cinelli y colaboradores hicieron un documento de análisis del COVID-19 en las redes sociales y la infodemia. Aquí abordan un análisis masivo de la difusión de la información de las plataformas Twitter, Instagram, YouTube, Reddit y Gab. Analizaron plataformas convencionales con contenido regulado como Twitter, Instagram y YouTube. También analizaron redes sociales no

reguladas como Gab y Reddit (ésta última tiene control nulo sobre el contenido publicado). Se registraron cambios de comportamiento entorno al 20 de enero, día en que la Organización Mundial de la Salud (OMS) emitió su primer informe de situación sobre la COVID-19. Ellos descubrieron que el mayor aumento en el número de publicaciones registradas fueron para el 21 de enero en Gab, el 24 de enero para Reddit, el 30 de enero para Twitter, el 31 de enero para YouTube y el 5 de febrero para Instagram. Por lo tanto, las plataformas de redes sociales parecen tener tiempos específicos para el consumo de contenido; tales como patrones que pueden depender de la diferencia en términos de audiencia y mecanismos de interacción entre los tipos de plataformas. De los datos analizados filtrando contenidos relacionadas con: coronavirus, pandemic, etc. El conjunto de datos derivado se compone de 1,342,103 publicaciones y 7,465,721 comentarios producidos por 3,734,815 usuarios. En este trabajo ellos encontraron que la difusión de información es muy cuestionable para Gab y Reddit ya que son entornos más susceptibles a la difusión de información errónea (Cinelli, M., Quattrociochi, W., Galeazzi, A., y et al., 2020).

Según C. Cuello-García y colaboradores mencionan que hay más de 3,800 millones de personas que utilizan las redes sociales en todo el mundo. La cantidad de información recibida a través de estas plataformas afecta la forma en que se percibe y se enfrenta a la pandemia (Cuello-Garcia, C., Pérez-Gaxiola, G., y van Amelsvoort, L., 2020). Tanto los pacientes, médicos y los científicos comparten información relacionada con COVID-19 en Twitter, Facebook u otros canales de redes sociales (Jurkowitz, M., y Mitchell, A., 2020; Hitlin, P., y Olmstead, K., 2019). Los profesionales de la salud hoy en día pueden comunicarse entre sí en diferentes partes del mundo (Smailhodzic, E., Hooijsma, W., Boonstra, A., y et al., 2016). Del mismo modo los científicos de todos los campos pueden interconectarse rápidamente y difundir los resultados de sus investigaciones (potencializando así su alcance científico por el fácil

acceso a la información). Las redes sociales son ahora parte de nuestras vidas y las usamos para hacer frente al distanciamiento social (Zarocostas, J., 2020).

En Cuello-García y colaboradores documentan que hay varios proyectos de investigación en redes sociales. Estas líneas de investigación estudian la desinformación en las redes sociales que se enfocan en detectar sus fuentes y cómo contrarrestarlas de manera eficiente para disminuir cualquier posible daño (Chou, W., Oh, A., y Klein, W., 2018). Sin embargo deben explorarse ciertas lagunas en estas investigaciones incluyendo la detección de poblaciones susceptibles (OraliaNJ, 2021), características sociodemográficas e ideológicas asimetrías en la intención de difundir desinformación (Wang, Y., McKee, M., Torbica, A., y et al., 2019). Esto sin duda se beneficiará de un enfoque interdisciplinario. Hay estudios y análisis de las redes sociales con big-data (que se traduce como macrodatos), minería de datos y vigilancia que se pueden utilizar para detectar patrones en el contenido de la web (Neumann, G., y Kawaoka, Y., 2019). La inteligencia artificial puede ayudar a desarrollar algoritmos basados en la recolección de datos y métodos basados en el conocimiento empírico de las experiencias de pacientes (Chitra, U., y Musco, C., 2020; Holone, H., 2016). La inteligencia artificial puede ayudar con este y otros desarrollos como una robusta herramienta de investigación computacional. Es alentador hoy en día el uso más común de la inteligencia artificial para pacientes y personal de atención médica con análisis secundarios de los datos de las redes sociales (Lau, A. Y., y Staccini, P., 2019). Este campo se encuentra todavía en sus primeras etapas y no está libre de errores pero es un área que vale la pena explorar.

K. Yang y colaboradores analizaron la prevalencia y difusión de enlaces a contenido de baja credibilidad sobre la pandemia en dos importantes plataformas de redes sociales. Ellos caracterizaron las similitudes y diferencias en patrones de difusión de personas del tipo *influencers* (término últimamente utilizado para personas influyentes). Al comparar las dos plataformas ellos

encontraron divergencia entre la prevalencia de fuentes populares de baja credibilidad y videos sospechosos. Es una realidad que es una minoría de cuentas y páginas que ejercen una fuerte influencia y dominancia en cada plataforma. Además se descubrió que en ambas plataformas existe evidencia de un intercambio coordinado de contenido de infodermia. La naturaleza abierta de esta manipulación apunta a la necesidad de estrategias de mitigación a nivel social. Sin embargo hasta el día de hoy no hay límites por políticas de acceso a datos inconsistentes y poca capacidad de los usuarios para estudiar y manipulaciones los ecosistemas de la información. Durante los primeros meses de la pandemia se observaron oleadas similares de baja credibilidad de contenido en ambas plataformas. La fuerte correlación entre las líneas de tiempo del volumen de contenido de baja y alta credibilidad revela que estos picos probablemente fueron impulsados por la atención pública a la crisis en lugar de ráfagas de contenido malicioso (Yang, K., Pierri, F., Hui, P. y et al., 2021)

M. Islam y colaboradores extrajeron entre el 31 de diciembre de 2019 al 5 de abril de 2020 contenido de artículos noticiosos para comparar y contrastar datos recopilados de varias fuentes. Ellos encontraron 2,311 informes de rumores, estigmas y conspiración en 25 idiomas de 87 países. Los resultados se relacionaron con enfermedades, transmisión y mortalidad (24%); control medidas (21%), tratamiento y curación (19%); causa de la enfermedad e incluido el origen (15%); relacionados con violencia (1%) y diversos (20%). De los 2,276 informes para los que se disponía de valoraciones de texto 1,856 afirmaciones eran falsas (82%). Este estudio demuestra que los rumores afectan a la confianza entre los trabajadores de la salud, desconfianza en los gobiernos, los profesionales de la salud y afectar la salud de las personas. Este artículo también ha demostrado que la conspiración, teorías y rumores pueden motivar a las personas a no vacunarse o recibir sus antibióticos. Este artículo recomienda a los gobiernos y agencias internacionales de salud continuar publicando

información correcta y apropiada al contexto respaldada por evidencia científica sobre COVID-19 en sus sitios web. Las agencias nacionales e internacionales incluidas las agencias de verificación de datos no solo deben identificar rumores, teorías de conspiraciones y desacreditarlos sino que también debe involucrar a las empresas de redes sociales para difundir información correcta (Islam, M., Sarkar, T., Khan, S., y et al., 2020).

Por la necesidad de difusión de nuestros resultados regionalizados, de carácter científico, información que es verás y apropiada surge la idea de engendrar la WAF (de sus siglas en inglés Web App FrameWork). La WAF tiene un carácter de orden científico, que sólo se alimentará de fuentes oficiales y fidedignas. La WAF se retroalimentará de la información que los ciudadanos aporten ya que tendrá un servicio de tipo público. Pero sólo usuarios registrados y verificados por un humano podrán utilizar el servicio. Para el contenido malicioso se utilizan análisis de cadenas de caracteres por medios de *tokens* (alarmas) que le indicarán al administrador la presencia de contenido malicioso. La Figura 8 dentro de este documento brinda al lector un diagrama de flujo inédito donde se describe a detalle el pseudocódigo para que cualquier ciudadano que tenga cierto conocimiento en programación de software será capaz de diseñar, programar y reproducir esta WAF.

3.2 Marco Referencial

Para el marco referencial se puntualiza que México tiene una población de casi 126 millones de personas al 2020. Lo que convierte a México en el undécimo país más poblado del mundo. Actualmente México goza de un bono de población debido a que la mayor concentración de la población se encuentra en el intervalo de 10-40 años (Garza, G., y Schteingart, M., 2010).

3.2.1 Mexicanos

México tiene sus orígenes desde la antigua Tenochtitlán. La fundación de la ciudad Mexica fue en el posclásico en el año 1325. Ésta se ubicó en la cuenca

del valle de México. La capital Mexica fue una de las ciudades emblemáticas y de mayor jerarquía en Mesoamérica. Destacan los sistemas hidráulicos para aquella época, ya que aprovechaban al máximo de los recursos naturales. Ellos contenían las aguas en canales para evitar inundaciones, además que éstos abastecían de agua dulce que hacían circular al interior de la toda ciudad para sus cultivos y se desplazaban en ellos como medio de transporte mediante chinampas (Molina-Torres, R., Nolasco-Jáuregui, O., Rodríguez-Torres, E., y et al., 2021).

Una de las teorías dice que la palabra *mexicano* se deriva de *mexihco* que significa “lugar donde vive *Hutzilopchtli*” (nombre del dios de la guerra y patrón de los mexicas) (Aguilar-Moreno, M., 2007 y Rosas, F. M., 2005). El término *mexicano* aparece por primera vez impreso en el siglo XVI en Barcelona en 1566. Este documento describe la expedición que partió del puerto de Acapulco a España, refiriéndose a los *mexicanos* como la población que habitaba en los límites geográficos de la Nueva España (es decir México conquistado por los españoles). Esta línea de navegación que estuvo activa durante casi 250 años fue la ruta preferida para el comercio y se le conoció como el *galón de Manila* (Folch, D., 2013).

Desde principios del siglo XIX los *mexicanos* fueron clasificados como *mestizos* (Barbachano, F., 1964). Lo que significa que no se identifican ni como una cultura indígena ni como el patrimonio cultural heredado por los españoles (Lizcano Fernández, F., 2005).

Los *mexicanos* que se identifican como *mestizos* son principalmente de ascendencia europea, indígena y africano (el legado de la esclavitud) en la Nueva España. Sin embargo, los genetistas teorizan que en regiones de México que no tenían presencia de esclavos los rastros de la ascendencia africana podrían provenir de colonos españoles y no de los propios esclavos africanos ya que dicha ascendencia tiene su origen en el norte de África. Dependiendo de la región algunos *mestizos* pueden tener pequeños rastros de herencia asiática

debido a los miles de filipinos y chinos (esclavos asiáticos de origen diverso, no sólo chinos). La inmigración asiática (específicamente la china) puede ayudar a explicar la contribución relativamente alta de Asia en el noroeste de México (por ejemplo, Sonora) (Navarrete, F., 2005).

El proyecto del *Mapa del Genoma de los Mexicanos* realizado por el INMIGEN (Instituto Nacional de Medicina Genómica del Instituto Nacional de Salud fundado en el año 2004) constituye un logro indiscutible para nuestro país ya que poco se ha trabajado en este campo en Latinoamérica. En este proyecto intervinieron el Gobierno Federal a través de la Secretaría de Salud, gobiernos estatales (Yucatán, Zacatecas, Sonora, Veracruz, Guerrero, Guanajuato, Tamaulipas, Durango, Oaxaca, Campeche), universidades estatales, comunidades indígenas, miles de voluntarios participantes y el personal de dicho Instituto. Los fondos para lograr la conclusión de este proyecto los otorgó el Gobierno Federal, la Fundación Mexicana para la salud y la Fundación Gonzalo Río Arronte. Con el conocimiento generado sobre el *genoma mexicano*, se abren caminos como son la integración de la información genética para la toma de decisiones en materia de salud pública, la creación de legislación como son los esquemas de propiedad intelectual internacional, nacional y su relación con su soberanía. El nivel de acceso-rentabilidad tanto pública como privada y la transferencia de tecnología para llevar a cabo las futuras investigaciones en México (NMED, 2010).

El proyecto del *Mapa del Genoma de los Mexicanos* se presenta como una herramienta pública de gran utilidad que permitirá iniciar nuevas líneas de investigación derivadas de las variaciones genéticas asociadas a la predisposición de enfermedades, el mejoramiento en su manejo y terapéutico (Jimenez-Sanchez, G., 2003).

Martha Díaz en su estudio llamado el *genoma mexicano* (NMED, 2010) ha revelado la existencia de componentes ancestrales significativos en las poblaciones de diferentes regiones de México. Así como la identificación de

SNPs (polimorfismos de un solo nucleótido) únicos y que el estudio de estos grupos poblacionales será útil para abordar problemas de estratificación y diseño de estudios complejos como son los campos de la farmacogenética, vacunogenética y la toxicogenómica. Por ejemplo en la vacunogenómica es posible identificar los genes que regulan las respuestas inmunes a las proteínas virales como es el caso de la respuesta de antígeno de superficie de la hepatitis B donde aproximadamente el 10% de las personas mexicanas no responden a la vacunación con HBsAg (es el primer marcador serológico que aparece después de la infección y su persistencia por más de 6 meses indica una hepatitis B crónica). Otro de los campos importantes a desarrollar con el *genoma mexicano* es el de la nutrigenómica la cual estudia la expresión de los genes en relación con los procesos metabólicos y el desarrollo de enfermedades asociadas.

En promedio el mayor componente ancestral genético en *mexicanos* que se identifican a sí mismos como *mestizos* es el indígena. Aunque la diferencia de incidencia entre los compuestos indígenas y europeos es relativamente pequeña con ambos usualmente representando no más del 40% de su composición genética (Gorodezky, C., Alaez, C., Vázquez-García, M., y et al., 2001). Los estudios genéticos realizados sobre *indígenas mexicanos* revelan una ascendencia indígena predominante pero con variaciones más altas de lo esperado en los componentes ancestrales europeos y africanos. Las investigaciones existentes sugieren que la ubicación geográfica juega un papel más significativo en la determinación de la composición genética de la persona indígena promedio que los rasgos culturales. Un ejemplo de esto es la población indígena de Tlapa en el estado de Guerrero que a pesar de que en su mayor parte habla español y tiene las mismas costumbres culturales que los mexicanos no indígenas muestra una ascendencia indígena del 95% (Gorodezky, C., Alaez, C., Vázquez-García, M., et al., 2001). En contraste, los pueblos indígenas de habla náhua del estado de Veracruz tienen una ascendencia media europea del

42% y una ascendencia africana del 22% (Buentello-Malo, L., Peñaloza-Espinosa, R., Salamanca-Gómez, F., y et al., 2008).

Un estudio realizado en 2006 por el INMIGEN que genotipó 104 muestras informó que los *mexicanos mestizos* son 58.96% europeos, 35.05% asiáticos (principalmente amerindios) y 5.03% otros. De los seis estados que participaron en el estudio, el estado de Sonora mostró la ascendencia europea más alta de casi 70%. Mientras que el estado de Guerrero presentaba la ascendencia europea más baja de alrededor de 50% (Schwartz-Marín, E., y Silva-Zolezzi, I., 2010; Valdez-Velazquez, L. L., Mendoza-Carrera, F., Perez-Parra, S., y et al., 2011; Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., y et al., 2009).

Un estudio autosómico realizado en *mestizos* de las tres ciudades más grandes de México reportó que los *mestizos* de la ciudad de México tenían en promedio un 50% de ascendencia europea, 1% de ascendencia africana y 49% de ascendencia amerindia. Mientras que los *mestizos* de las ciudades de Monterrey y Guadalajara tenían en promedio un 60% de ascendencia europea y un 40% de ascendencia indígena. También en *mexicanos* de los estados de Nuevo León, Zacatecas y San Luis Potosí se encontró que el promedio de ascendencia indígena era del 22% mientras que el 78% fue de ascendencia genética de origen español/europeo. En la Ciudad de México se reportó que la ascendencia media de los *mestizos mexicanos* era 57% europea, 40% amerindia y 3% africana (Cerda-Flores, R., Villalobos-Torres, M., Barrera-Saldaña, H., y et al., 2002).

Canizalez-Román, A., y colaboradores en el 2018 realizaron una serie de tablas donde ilustran las frecuencias de diferentes grupos sanguíneos en varias ciudades y estados mexicanos. Debido a que la población amerindia de México posee exclusivamente el tipo sanguíneo "O" la presencia de otros grupos sanguíneos puede dar una idea aproximada de la cantidad de influencia extranjera que hay en cada estado que ha sido analizado. Los resultados de estos estudios no deben tomarse como estimaciones exactas de los porcentajes de

diferentes grupos étnicos que habitan México (por ejemplo, el porcentaje de tipos sanguíneos A, B+ es porcentaje de mexicanos blancos) debido a que un *mexicano mestizo* puede poseer sangre de tipo "A", "B" etc. Se observa que los grupos sanguíneos "A" y "B" son más comunes en voluntarios jóvenes mientras que los grupos "AB" y "O" son más comunes en voluntarios de edad avanzada. El número total de voluntarios analizados en el estudio fue de 271,164 (Canizalez-Román, A., Campos-Romero, A., Castro-Sánchez, J. A., y et al., 2018).

Un estudio realizado en hospitales de la Ciudad de México reportó que en promedio el 51.8% de los recién nacidos *mexicanos* presentaban la marca mongólica y el 48.2% de los bebés analizados estaba no presente (Domínguez, L., y Fragoso, G., 2002). La mancha mongola aparece con una frecuencia muy alta (85-95%) en niños asiáticos nativos americanos y africanos (Ayala Uribe, M. 1970). La lesión de piel casi siempre aparece en niños sudamericanos y *mexicanos* que son racialmente *mestizos*. Mientras que tiene una frecuencia muy baja (5-10%) en niños caucásicos. Según el IMSS (Instituto Mexicano del Seguro social) en el 2012 publica que a nivel nacional alrededor de la mitad de los bebés mexicanos tienen la mancha mongola.

3.2.2 Etnias en México

Según la cifra del INEGI de 2015 se hablan 364 variantes lingüísticas provenientes de 68 agrupaciones en todo México (ver Figura 1). La población étnica en México se divide en dos grandes campos: 1) el porcentaje de población que es hablante de una (o varias lenguas indígenas) y el español; 2) la monolingüe que habla su lengua madre de origen indígena.

La población hablante de lengua indígena tiene un grado de vulnerabilidad por la discriminación de la que es objeto al presentarse por un servicio de salud y no tener protección social en salud y/o un registro (Leyva-Flores, R., Infante-

Xibille, C., Gutiérrez, J. P. y et al., 2013; Vargas, R., Cobos, A. y Saa, T., 2016; Guadarrama-Olivera, R., Hualde-Alfaro, A., y López-Estrada, S., 2012).



Figura 1. Mapa de las variantes lingüistas en México y sus etnias (captura de pantalla de: <https://www.gob.mx/cultura/es/articulos/lenguas-indigenas?idiom=es>).

-Porcentaje de población que habla una lengua indígena y español: el mínimo registro de porcentaje de 0.01% población hablante indígena los tienen 59 municipios del país y el máximo porcentaje de un 100% lo tiene un municipio Theiupango de Veracruz (Lastra, M., González, C., Pérez, M., Guzmán, y et al., 2013).

-Porcentaje de población que sólo habla lengua indígena: este grupo tiene una mayor vulnerabilidad debido a la falta de un enfoque de interculturalidad en los programas de salud dirigidos particularmente la monolingüe (Gracia, M. A. y Horbath, J. E., 2019). Lo que reduce sus garantías de derecho de acceso a la

información y limitan la atención que se le presta a este sector en instalaciones de servicio de salud (Monroy, R., 2017; Figueroa, A., 2000).

El mínimo porcentaje de 0.01% registro de población monolingüe se tiene en más de 1,327 municipios (de 2,469 municipios totales en México y 16 Alcaldías en la CDMX) y el máximo porcentaje casi 64% del mismo se tiene en Chalchihuitán Chiapas (Lastra, M., González, C., Pérez, M., Guzmán, y et al., 2013).

3.2.3 Nivel Económico y urbanización en México

A inicios del siglo XXI México es uno de los países altamente urbanizado de América Latina. Para el año 2005 tiene 367 ciudades que se caracterizan por la concentración de sus habitantes en 57 zonas metropolitanas de las cuales mantienen el 83% de la población urbana nacional. Por ello el país enfrenta grandes retos sobre todo en las zonas urbanas como: servicios de infraestructura para la población en materia de electricidad, telecomunicaciones, agua, drenaje, vivienda, vialidades, educación, salud y seguridad pública entre las más sustanciales. También es en las urbes mexicanas donde se concentran las empresas de servicios comerciales e industriales que deberán disponer de un conjunto de modernas condiciones, servicios generales de producción urbanos y metropolitanos (Garza, G., y Schteingart, M., 2010).

El desarrollo económico y la urbanización están orgánicamente vinculados ya que son dos procesos inseparables que caracterizan la estructura de la sociedad. La distribución de la población en la actualidad es del 67% que vive en ciudades y los niveles de pobreza es muy alta en las zonas rurales. El 46% de la población es pobre en todo el país y con esta relación México ocupa el primer lugar entre los países de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) en cuanto a la diferencia más extensa entre el 10% del más rico y el más pobre.

3.2.4 Salud en México

En México hay más de 30 millones de personas sin protección social en salud. El alto porcentaje de gasto de bolsillo de las familias mexicanas expone a su población a gastos excesivos por motivos de salud. En el año 2000 más de tres millones de hogares mexicanos presentaron gastos catastróficos y/o empobrecedores por motivos de salud (Kassir, R., 2020).

En 2006 esta cifra se había reducido gracias a la implantación de diversos programas sociales. Estudios económicos en salud muestran que los gastos excesivos (catastróficos y empobrecedores) por motivos de salud disminuyeron de 4.1% (alrededor de 2.6 millones de hogares) a 2.7% (alrededor de 1.8 millones de hogares) entre 2004 y 2008 (Knaul, F., Arreola-Ornelas, H., Méndez, O., Lafontaine, L. y Wong, R., 2010). Estos mismos estudios indican que hay gran variación en la prevalencia de estos gastos.

Para el 2008 el 2.05% en los hogares urbanos presentaron gastos ruinosos por motivos de salud. Los hogares que cuentan con seguro de salud presentaron una prevalencia de 2.67% contra 3.53% en los hogares que no cuentan con seguro de salud. Los hogares con adultos mayores presentan una prevalencia de gastos fatales por motivos de salud del casi 6.02% y que contrasta con la prevalencia de los hogares que no cuentan ni con niños ni con adultos mayores (que es de 1.36%) (Dantés, O., Sesma, S., Becerril, V., y et al., 2011).

3.3 Índice de Marginación

El IM (Índice de Marginación) se calcula utilizando el procedimiento estadístico multivariado llamado Análisis de Componentes Principales. Conceptualmente este procedimiento es compatible con la naturaleza multidimensional de la Marginación. Técnicamente se transforma un conjunto de variables correlacionadas en otro no correlacionado (Spector, P., 2014). De manera que la primera componente principal formada explica tanta la

variabilidad de la componente principal estandarizada resultante del análisis (ver Tabla 2). Su expresión algebraica es (CONAPO, 2016):

$$Y_{i1} = \sum_{j=1}^9 c_j Z_{ij} = IM_i \quad (2.1)$$

Donde:

Y_{i1} = Valor de la unidad de análisis i en la componente principal estandarizada 1 (primera componente principal).

c_j = Ponderador del indicador estandarizado j .

Z_{ij} = Indicador estandarizado j de la unidad de análisis i .

IM_{zi} = Valor del índice de Marginación de la unidad de análisis i .

$i = 1, \dots, 32$. En el caso de que las unidades de análisis sean las entidades federativas, o bien, $i = 1, \dots$ hasta el número total de municipios que exista en cada año de referencia cuando las unidades de análisis sean los municipios.

$j = 1, \dots, 9$. Se refiere a los nueve indicadores socioeconómicos seleccionados para el análisis de componentes principales.

De esta manera la situación de cada unidad de análisis descrita inicialmente con la información de nueve variables relacionadas queda ahora descrita con un solo valor dado por el IM. El IM puede tomar valores negativos o positivos. Siendo los valores positivos los más altos niveles de marginación. Calculando el índice es posible ordenar y clasificar a las unidades de análisis para que finalmente mediante la *Técnica de Estratificación Óptima* por Dalenius y Hodges obtener en cinco categorías los *Grados de Marginación (GM)*: Muy Alto, Bajo, Medio, Alto y Muy Alto.

El IM se ha calculado en México desde el año 1990. En su edición más reciente correspondiente al año 2015 el insumo de información para calcularlo fue en la Encuesta Intercensal 2015 levantada por el INEGI. Para el IM del 2015 no se

levantó censo en poblaciones menores a 5 mil habitantes por lo que sólo se tomaron 8 indicadores socioeconómicos (CONAPO, 2016).

3.3.1 Parámetros Índice de Marginación

**Porcentaje de condición de analfabetismo (ANALF).* Se refiere a la población de 15 años (y más) que declara saber leer y escribir un recado. Esta condición se clasifica en analfabetas, analfabetas y no especificado.

** Porcentaje de 15 años o más sin primaria completa (SPRIM). Porcentaje de Nivel educativo básico.* Cada una de las etapas que conforman el Sistema Educativo Nacional.

Ocupantes en Vivienda. Espacio delimitado generalmente por paredes y techos de cualquier material. Con entrada independiente y construido para la habitación de personas. Que al momento del levantamiento censal la estuvieran usando para vivir. Las viviendas se clasifican en particulares o colectivas. Las casas particulares se dividen en: casa independiente, vivienda en edificio, vivienda en vecindad, vivienda en cuarto de azotea, cochera, vivienda móvil y refugio.

**Porcentaje de ocupantes en viviendas particulares con Energía eléctrica (OVSEE).* Distinción de las viviendas particulares habitadas según la existencia de luz eléctrica (independientemente de la fuente que provenga).

**Porcentaje de ocupantes en viviendas sin agua entubada (OVSAE).* Clasificación de las viviendas particulares habitadas que se abastecen de agua para consumo personal y doméstico. Conforme el acceso de los ocupantes de las viviendas al agua se clasifican en: agua entubada dentro de la vivienda, agua entubada fuera de la vivienda, agua entubada de llave pública, agua entubada que acarrearán de otra vivienda, agua de pipa, agua de pozo, agua de lago, agua de arroyo y no especificado.

**Porcentaje de viviendas con algún nivel de hacinamiento (VHAC).* Es el cálculo de viviendas con un solo dormitorio y número de ocupantes, con dos dormitorios y número de ocupantes, etc.

**Porcentaje de ocupantes en viviendas con piso de tierra (OVPT).* En las viviendas particulares el elemento predominante es el piso de: tierra, cemento, madera y no especificado.

**Porcentaje de población en localidades con menos de 5,000 habitantes (PL<5000).* Es la clasificación de las localidades a nivel nacional según el número de habitantes.

NOMBRE	POB	ANA LF	SPRIM	OVSDE	OVSEE	OVSAE	VHAC	OVPT	PL <5000	PO2SM	IM	GM
Aguascalientes	1312544	2.6	11.89	0.67	0.3	0.81	21.86	0.75	25.16	34.6	-0.89	Bajo
Baja California	3315766	1.96	10.46	0.26	0.47	2.82	23.03	1.15	10.35	22.85	-1.1	Muy bajo
Baja California Sur	712029	2.51	11.59	0.38	1.14	7.15	26.23	4.08	15.62	22.37	-0.6	Bajo
Campeche	899931	6.68	18.56	4.08	1.59	6.49	37.93	2.95	30.88	40.69	0.46	Alto
Coahuila de Zaragoza	2954915	1.99	9.56	0.62	0.29	1.67	23.62	0.66	12.15	27.94	-1.1	Muy bajo
Colima	711235	3.9	15.24	0.38	0.38	0.96	25.65	2.45	14.47	31.18	-0.73	Bajo
Chiapas	5217908	14.98	31.71	2.9	2.49	13.45	44.46	11.78	57.86	62.46	2.41	Muy alto
Chihuahua	3556574	2.67	12.65	1.4	1.81	2.39	22.16	1.76	17.05	34.77	-0.6	Bajo
Ciudad De México	8918653	1.49	6.62	0.04	0.04	1.1	19.19	0.47	0.67	28.26	-1.45	Muy bajo
Durango	1754754	3.17	14.5	3.68	2.63	3.26	24.58	4.33	36.19	39.16	0.05	Medio
Guanajuato	5853677	6.39	19.12	3.31	0.68	4.13	25.36	2	34.67	37.41	-0.07	Medio
Guerrero	3533251	13.73	27.25	13.03	2.4	15.64	42.11	14.86	49.68	53.29	2.56	Muy alto
Hidalgo	2858359	8.26	18.13	3.09	1.12	5.83	28.17	3.19	58.71	46.22	0.5	Alto
Jalisco	7844830	3.55	14.9	0.86	0.34	1.84	22.12	1.59	17.5	29.4	-0.82	Bajo

México	16187608	3.37	11.77	1.68	0.38	4.03	28.53	1.92	19.11	35.28	-0.57	Bajo
Michoacán de Ocampo	4584471	8.35	25.35	2.32	0.81	4.26	28.05	5.9	40.58	44.88	0.5	Alto
Morelos	1903811	4.99	15.09	1.03	0.44	5.81	27.04	3.88	24.65	42.08	-0.2	Medio
Nayarit	1181050	5.07	17.56	4.66	2.54	4.36	27.63	3.92	39.14	37.96	0.31	Medio
Nuevo León	5119504	1.64	8.38	0.16	0.12	1.36	23.09	0.8	6.7	16.15	-1.39	Muy bajo
Oaxaca	3967889	13.65	29.22	2.44	2.87	13.05	38.33	13.44	61.51	49.46	2.12	Muy alto
Puebla	6168883	8.39	21.32	1.58	0.88	6.94	35.27	5.69	38.5	52.16	0.69	Alto
Querétaro	2038372	4.57	13.01	2.93	0.67	3.23	24.21	1.52	39.07	25.92	-0.49	Bajo
Quintana Roo	1501562	3.9	13.32	2.19	1.11	2.7	36.31	2.41	14.36	28.32	-0.37	Medio
San Luis Potosí	2717820	6.33	18.87	2.34	2.29	10.68	25.21	5.56	40.08	43.64	0.58	Alto
Sinaloa	2966321	4.18	16.82	2.57	0.52	2.99	30.37	2.85	32.85	32.7	-0.24	Medio
Sonora	2850330	2.19	11.15	1.05	1.05	2.52	26.63	2.61	17.39	29.93	-0.7	Bajo
Tabasco	2395272	5.4	17.25	1.81	0.43	10.1	32.5	3.69	53.65	36.96	0.3	Medio
Tamaulipas	3441698	3.03	13.35	0.37	0.73	2.51	28.69	1.52	13.92	37.39	-0.62	Bajo
Tlaxcala	1272847	3.98	12.6	1.36	0.47	1.14	32	2	36.4	51.47	-0.2	Medio
Veracruz	8112505	9.51	25.04	1.53	1.62	13.39	32.03	6.84	46.2	49.68	1.14	Alto
Yucatán	2097175	7.47	21.17	10.1	1.08	1.64	36.42	1.67	26.27	47.6	0.51	Alto
Zacatecas	1579209	4.42	19.31	4.17	0.62	3.13	24.99	1.25	48.19	45.4	0.01	Medio

Tabla 2. Indicadores socioeconómicos para el análisis del IM por entidad federativa en México para el 2015 (CONAPO, 2016).

**Porcentaje de población ocupada con ingresos de hasta 2 salarios mínimos (PO2SM). Persona mayor de 12 años que en la semana realizó una actividad económica por lo menos una hora y recibió pago económico por ello.*

**Índice de Marginación (IM). Es una medida-resumen que permite diferenciar entidades federativas y municipios de acuerdo con las carencias que padece la población. La marginación es un fenómeno multidimensional que considera la exclusión de la población al proceso de desarrollo y al disfrute de sus beneficios. Como resultado de acceso a la educación, la residencia en viviendas*

inadecuadas, la percepción de ingresos monetarios insuficientes y las relacionadas con la residencia en localidades pequeñas (CONAPO, 2016). El cálculo del IM (ver Tabla 2) involucra la construcción de indicadores de carencias o déficit en educación, vivienda, ingresos monetarios y en la distribución de la población (Monroy, R., 2017).

3.4 Conclusiones

El arribo de la pandemia por el virus COVID-19 a México se da a conocer con el primer paciente que da positivo al virus en el mes de marzo del 2020. A la llegada del virus se encuentra con el siguiente escenario: México tiene una población de casi 126 millones de personas al 2020. Lo que lo convierte en el undécimo país más poblado del mundo. Con un factor bono en contra del virus que es que la mayor parte de su población se encuentra en el intervalo de 10-40 años.

Uno de los coeficientes de alto riesgo de dispersión del virus es la alta concentración de la población en espacios reducidos y México para el año 2005 registró a 367 ciudades que se caracterizan por la concentración de sus habitantes en 57 zonas metropolitanas. De las cuales poseen el 83% de la población urbana nacional. Conjuntamente más de 30 millones de personas no tienen protección social en salud.

En México la población hablante de lengua indígena tiene un grado de vulnerabilidad debido a la falta de interculturalidad en el país. México tiene 1,327 municipios con hablantes monolingües lo cual dificulta la difusión de la información en su lengua nativa en esta pandemia.

México ocupa el primer lugar entre los países de la OCDE en cuanto la diferencia más extensa entre el más pobre y el más rico. El IM surge de la necesidad de la caracterización de todas estas variables que implican la marginación de su población. Sobresalen entidades con un GM *Alto* a *Muy Alto*

como Campeche, Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, San Luis Potosí, Veracruz y Yucatán.

En época de pandemia la difusión de la información es fundamental y de carácter verás. Es por eso que surge la necesidad de diseñar y codificar la WAF. Su diagrama de diseño se dona y se describe a detalle dentro de la HPS (ver CAPITULO 4) siendo así parte del aporte de esta investigación.

CAPÍTULO 4

METODOLOGÍA

En este capítulo se describe a detalle la creación y diseño de la Herramienta Propia en Software (HPS). Es muy importante resaltar que la HPS es una propuesta inédita de la autora. La intensa descripción del diseño de la HPS es con el objetivo de ser reproducida por cualquier persona siguiendo la metodología descrita en esta sección. La HPS nace de la necesidad de hacer una extracción (ver Figura 2), lectura y decodificación correcta de los datos de COVID-19 en México (ya que sin un conocimiento en programación no es posible utilizarlos).

La metodología para el diseño de la HPS se divide en cuatro niveles. En el primer nivel se tiene el análisis, manipulación y extracción de los datos de los sujetos de estudio. El primer nivel une las características del lenguaje C y del lenguaje R. NOTA: en este capítulo el *pseudocódigo* se resalta con letra cursiva haciendo referencia a nombres de variables, comandos, funciones, librerías, conceptos, métodos, eventos, propiedades y configuraciones (que no se traducen y se toman como tal).

Con esta herramienta se desmenuza el instrumento tipo encuesta que ha utilizado la secretaría de epidemiología y la secretaría de salud de México para tener el registro de todos los pacientes con COVID-19.

El segundo nivel de la HSP se describe la síntesis de la técnica ya que es en este nivel es donde se interpretan los valores cuantitativos: tales como correlación de variables, tipo de paciente, regiones, comorbilidades, edades, nivel económico, etc. En el segundo nivel de la HSP se busca valuar y brindar al investigador la posibilidad de precisar las zonas de mayor riesgo de contagio en el territorio nacional.

El tercer nivel de la HSP se caracteriza por ser dinámico. Es decir, se utiliza una red Bayesiana que será capaz de hacer cálculos probabilísticos cada vez que su información de entrada sea actualizada. Si los registros de cada día cambian los cálculos serán dinámicos y se actualizarán junto con cada movimiento en la entrada. En el nivel 4 se encuentra la descripción de la creación de una WAF que una herramienta inédita de la autora quien pretende difundir información veraz y de orden científico por este medio.

4.1 Nivel 1 de HSP

La captura de pantalla de la Figura 2 corresponde al archivo *200412COVID19MEXICO.csv* del 12 de Abril 2020. Este archivo no tiene ningún tratamiento. Tal cual se descarga de la base de datos del gobierno federal (en Datos Abiertos Dirección General de Epidemiología; DGE, 2021) y es uno de los 203 días a decodificar. La Figura 2 es un ejemplo del tipo de formato que se tiene que decodificar para los 35 campos y sus diccionarios (ver Anexo A).

	A	B	C	D	E	F	G	H	I	J	
1	FECHA_ACTUALIZACION	ID_REGISTRO	ORIGEN	SECTOR	ENTIDAD_UM	SEXO	ENTIDAD_NAC	ENTIDAD_RES	MUNICIPIO_RES	TIPO_PACIENTE	
2	2020-04-21	09e8dc	2	9	15	1	15	15	37	2	
3	2020-04-21	1dd782	2	12	9	1	15	9	3	1	
4	2020-04-21	0efbaaf	2	9	28	2	16	28	32	1	
5	2020-04-21	013a6c	99	3	15	2	15	15	106	1	
6	2020-04-21	091a48	1	12	15	2	15	15	31	2	
7	2020-04-21	1a72fe	1	12	9	1	99	9	11	1	
8	2020-04-21	1a5595	2	12	9	2	99	9	8	1	
9	2020-04-21	13236c	1	4	2	2	21	2	4	2	
10	2020-04-21	0cdeb12	1	12	9	2	99	9	12	2	
11	2020-04-21	038dad	1	12	9	1	99	9	14	1	
12	2020-04-21	1d8c4e	1	12	31	2	31	31	50	2	
13	2020-04-21	08939c	1	12	12	1	12	12	56	2	
14	2020-04-21		75608	2	12	2	1	12	1	1	
15	2020-04-21	0fe0fd	1	12	12	1	99	12	59	2	
16	2020-04-21	0d7157	1	9	9	1	99	3	8	2	
17	2020-04-21		49899	1	12	12	1	12	1	2	
18	2020-04-21	113cf3	2	12	12	1	12	12	29	1	
19	2020-04-21	1b6e87	1	4	9	2	19	9	15	1	
20	2020-04-21	17016d	2	12	14	2	14	14	120	1	
21	2020-04-21	074c15	1	12	15	1	7	15	109	2	
22	2020-04-21		71659	2	4	15	1	99	15	33	2
23	2020-04-21	1beb4d	1	12	19	1	99	19	46	1	
24	2020-04-21	1e5be0	2	9	15	1	15	15	37	1	
25	2020-04-21	17d1a5	2	12	31	1	23	31	50	1	

Figura 2. Captura de pantalla del archivo correspondiente al día 12 de Abril de los 203 días de análisis. Aquí se observa los campos en datos crudos sin decodificar.

Es muy importante resaltar que cada archivo *csv* (de sus siglas en inglés Comma Separated Values File ó archivos separados por comas) tienen matrices

de longitud de cientos de miles de filas por 35 columnas. Cada fila corresponde al registro de un paciente (*ID_Registro*).

4.1.1 Extracción de los datos

El instrumento tiene dentro de sus 35 campos: 1) la fecha de la actualización, 2) ID del paciente, 3) País de Origen, 4) Institución que brindó la atención, 5) Número de la Clínica de Salud, 6) Sexo, 7) Lugar de nacimiento, 8) Lugar de residencia, 9) Municipio de residencia, 10) Tipo de paciente, 11) Fecha de ingreso, 12) Fecha de síntomas, 13) Fecha de defunción, 14) Intubado, 15) Neumonía, 16) Edad, 17) Nacionalidad, 18) Embarazo, 19) Habla lengua indígena, 20) Diabetes, 21) EPOC, 22) Asma, 23) Inmunosupresión, 24) Hipertensión, 25) Otras complicaciones, 26) Cardiovascular, 27) Obesidad, 28) Renal Crónica, 29) Tabaquismo, 30) Otra droga, 31) Resultado de COVID-19, 32) Migrante, 33) País de Nacionalidad, 34) País de Origen y 35) Entró a UCI (Unidad de Cuidados Intensivos). Cada campo tiene su propio diccionario (ver Anexo A).

El archivo de la Figura 2 se ingresa al algoritmo de búsqueda de empate de caracteres por medio de *tokens*. El algoritmo hace conteos de coincidentes y con ello se comienza en análisis estadístico. La búsqueda de empate de caracteres está escrito en lenguaje C, ver ejemplo de captura en la Figura 3.

4.1.2 Interpretación de los datos

4.1.2.1 Pacientes Hospitalizados y No-Hospitalizados por Regiones

El primer paso para la interpretación de los datos es trabajar con campo número 8 llamado *ENTIDAD_RES*. Para este campo su diccionario lista las 32 entidades federativas por orden alfabético del territorio nacional comenzando con la entidad 01 correspondiente al estado de Aguascalientes (por orden alfabético) hasta la entidad 32 perteneciente al estado de Zacatecas. Además incluye las variables 97 que decodifica el valor de *NO_APLICA*, 98 para *SE_IGNORA* y 99 para *NO_ESPECIFICADO*. El objetivo de comenzar con

este campo es crear la clasificación de los pacientes en los 5 subconjuntos referentes a las 5 regiones de México. Como se puede observar en la Figura 4 dependiendo del valor numérico del diccionario los pacientes con COVID-19 serán clasificados por Región Noroeste; por ejemplo, si su lugar de residencia es la entidad 2, 3, 5, 25 ó 26.

Pero aún no se contabiliza como paciente contagiado hasta que se conozcan los campos: número 2 y el campo número 31. El campo número 2 es muy importante ya que corresponde al *ID_REGISTRO* del paciente y concierne al número del caso o expediente. El campo número 31 identifica el resultado del examen a SARS-CoV-2; donde la clave 1 es para *Positivo*, 2 es para *No-Positivo* y 3 significa que el resultado está *Pendiente*.

```

1  #include <stdio.h>
2  #include <stdlib.h>
3  #include <string.h>
4  #include <sys/types.h>
5  #include <sys/stat.h>
6  #include <unistd.h>
7  #include <dirent.h>
8
9  #include "divide_regiones.h"
10
11 const char strRegionNoroeste[5] = {2, 3, 8, 25, 26};
12 const char strRegionNoreste[5] = {5, 10, 19, 24, 29};
13 const char strRegionOccidente[8] = {1, 6, 11, 14, 16, 18, 22, 32};
14 const char strRegionCentro[9] = {9, 12, 13, 15, 17, 21, 29, 36};
15 const char strRegionSureste[7] = {4, 7, 20, 23, 27, 30, 31};
16 const char strRegionNA[3] = {97, 98, 99};
17
18 int correDivideRegiones(void)
19 {
20     unsigned int x;
21     DIR *d;
22     struct dirent *dir;
23     d = opendir("C:/datos_covid_entrada/");
24     x = 1;
25
26     printf("\n");
27     if (d)
28     {
29         while ((dir = readdir(d)) != NULL)
30         {
31             if ((strcmp(dir->d_name, ".") == 0) ||
32                 (strcmp(dir->d_name, "..") == 0))
33             {
34                 //do nothing
35             }
36             else
37             {
38                 printf("Adding %s\n", dir->d_name);
39                 divide_regiones(dir->d_name);
40             }
41             x++;
42         }
43     }
44 }

```

Figura 3. Captura de pantalla la HSP en la sección con lenguaje C para la extracción de los archivos csv y la decodificación de datos utilizando los *tokens*.

Por medio del uso de cadenas de texto se extrae la información seleccionada de los pacientes clasificados que cumplen con las características mencionadas. Es decir, se crean 5 archivos para cada región con la información del diccionario 1

que corresponde a *FECHA_ACTUALIZACION* por cada día de análisis. La herramienta propia crea 5 archivos *FECHA_ACTUALIZACION_REG1_NO.csv* correspondientes al subconjunto de los pacientes de la Región 1 del Noroeste (ver ejemplo en la Figura 4).

Y así consecutivamente, el archivo *FECHA_ACTUALIZACION_REG2_NE.csv* para los de la Región 2 del Noreste, *FECHA_ACTUALIZACION_REG3_OC.csv* para los de la Región 3 del Occidente, *FECHA_ACTUALIZACION_REG4_CE.csv* para los de la Región 4 del Centro del país y *FECHA_ACTUALIZACION_REG5_SUR.csv* para la Región 5 del Sureste de la república Mexicana.

	A	B	C	D	E	F	G	H	I	J
1	FECHA_ACTUALIZACION	ID_REGISTRO	ORIGEN	SECTOR	ENTIDAD_UM	SEXO	ENTIDAD_NAC	ENTIDAD_RES	MUNICIPIO_RES	TIPO_PACIENTE
2	2020-04-19	0c51ea	1	4	3	1	25	3	8	1
3	2020-04-19	1af10a	1	4	2	1	25	2	4	1
4	2020-04-19	19ebfa	1	12	2	2	2	2	4	1
5	2020-04-19	03bad7	2	12	26	2	22	26	30	1
6	2020-04-19	0d5d97	2	12	3	2	3	3	8	1
7	2020-04-19	1429a2	1	4	3	1	3	3	8	1
8	2020-04-19	195a7e	2	4	2	2	7	2	4	1
9	2020-04-19	16d143	2	12	2	2	25	2	4	1
10	2020-04-19	0d4154	2	4	25	2	25	25	6	1
11	2020-04-19	117ef6	2	4	2	1	2	2	2	1
12	2020-04-19	13b93e	2	4	25	2	25	25	6	1
13	2020-04-19	0993df	2	12	25	1	25	25	6	1
14	2020-04-19	0cfa7c	2	4	2	2	2	2	4	1

Figura 4. Captura de pantalla de la HPS en un primer post-procesamiento. La columna G ya tiene una extracción de las entidades federativas de la Región Noroeste, es decir: 2, 3, 5, 25 ó 26.

Posteriormente se hace un análisis con el diccionario número 10 llamado *TIPO_PACIENTE* (ver Figura 4 en columna J). Con esta clasificación se procede a separar los pacientes del tipo 1 que se encuentran en cuarentena en casa, los que están hospitalizados son del tipo 2 y la clave 99 para los que se desconoce. Esta jerarquía se utiliza para cada archivo de la clasificación anterior y se dividen en dos subconjuntos más: *NH (No-Hospitalizados)* y *H (Hospitalizados)*. Para detallar la nueva clasificación se sitúa un ejemplo: si se tiene el archivo *FECHA_ACTUALIZACION_REG3_NE_NH.csv* significa que son los pacientes de la región Noreste que se encuentran en casa y *FECHA_ACTUALIZACION_REG3_NE_H.csv* para los pacientes de la región Noreste que están hospitalizados.

Finalmente se crea y actualizan los conteos. Si un paciente dio positivo al test, tiene un *ID* único, pertenece a la Región 1 del Noroeste y se encuentra Hospitalizado entonces se incrementa el contador *REG1_H_counter* (y así se reproduce para cada región).

4.1.2.2 Edades de los Pacientes por Regiones

El campo 13 llamado *EDAD* es la variable numérica que divide en 2 subconjuntos: *FECHA_ACTUALIZACION_REG3_OC_NH_EDAD* y *FECHA_ACTUALIZACION_REG3_OC_H_EDAD*. Lo cual significa que extraen la edad de los pacientes *No-Hospitalizados* y *Hospitalizados* de la Región 3 de Occidente para hacer un promedio de ellos (y así se implementa por cada región). Una vez que se generan todos archivos de edades por las 5 regiones del país se utiliza la herramienta de histogramas dividido en 5 *bines* o *contenedores* correspondientes a: de 0-20 años, de 21-40 años, 41-60 años, de 61-80 años y de 81-100 años.

4.1.2.3 Interpretación de datos de Comorbilidades por Regiones

La última extracción de los datos en el nivel 1 de la HSP son las 9 comorbilidades reportadas en los datos de México: 1) Diabetes, 2) Asma, 3) Cardiovascular, 4) Hipertensión, 5) Obesidad, 6) Inmunosupresión, 7) EPOC (Enfermedad Pulmonar Obstructiva Crónica), 8) Otras Enfermedades y 9) Enfermedades Renales.

El campo número 20 llamado *DIABETES* (D) tiene las siguientes claves: 1 para un paciente que tiene la comorbilidad, 2 es para los que no tienen esta comorbilidad, 97 para no aplica y 98 se ignora. Una vez que se tiene el subconjunto *FECHA_ACTUALIZACION_REG3_OC_NH* se extraen aquellos pacientes que tienen 1 en el diccionario 20 y posteriormente se crea dos nuevos subconjuntos: *FECHA_ACTUALIZACION_REG3_OC_NH_D* y *FECHA_ACTUALIZACION_REG3_OC_H_D* para cada región. En este caso son los pacientes de occidente del tipo ambulatorios (*NH*) y hospitalizados (*H*)

que cumplen con las características descritas en la sección 4.1.2.1, sección 4.1.2.2 y además tienen la comorbilidad de Diabetes. Consecutivamente se hace lo mismo para el resto de las comorbilidades. Por ejemplo el campo número 21 llamado *EPOC (EP)* tiene la clave 1 para pacientes que tienen la comorbilidad, 2 es para los que no tienen esta comorbilidad. El campo 22 llamado *ASMA (A)* y el campo 23 llamado *INMUSUPR (IN)* es dedicado a los pacientes que tienen la comorbilidad de inmunosupresión si su clave es 1. El campo 24 se llama *HIPERTENSION (HI)* y el campo 25 *OTRAS_COM (OE)* el cual se dedica a los pacientes que tienen otras comorbilidades que no están dentro de estos 9 subconjuntos. El campo 26 se llama *CARDIOVASCULAR (CA)*, el campo 27 es para la comorbilidad de *OBESIDAD (OB)* y finalmente el campo 28 es para *RENAL_CRONICA (RE)*.

4.1.2.4 Pacientes que necesitaron Respirador Artificial

Un análisis que resulta muy interesante es respecto al campo número 35 llamado *UCI* (Unidad de Cuidados Intensivos) cuya variable 1 significa que el paciente Hospitalizado necesitó de un respirador, 2 es para los Hospitalizados que no necesitaron de un respirador, 97 para no aplica y 98 se ignora. En este subconjunto se aplica únicamente a los pacientes Hospitalizados en cada región. Por ejemplo el archivo *FECHA_ACTUALIZACION_REG2_NE_H_UCI* significa que son los pacientes de la región Noreste que se encuentran hospitalizados y además requirieron la UCI; es decir, estuvieron en la unidad de cuidados intensivos y necesitaron de un respirador artificial (se hace el mismo procedimiento para cada región del país).

4.1.2.5 Pacientes que Fallecieron

El campo número 13 llamado *FECHA_DEF (DEF)* se refiere a la fecha en la que fallece un paciente por COVID-19 que junto con el campo 35 de la UCI tiene el valor de 1. Es decir, son los pacientes que necesitaron un respirador

artificial y fallecieron (se hace el mismo procedimiento para cada región del país).

Se crean los archivos de *FECHA_ACTUALIZACION_REG2_NE_H_UCI_DEF* que significa que son los pacientes de la región Noreste que se encuentran hospitalizados, estuvieron en UCI y fallecieron.

4.2 HSP: Nivel 2

En el segundo nivel de la HSP se trabajan las herramientas visuales por cada uno de los archivos antes mencionados. Esto para dar una presentación a los datos extraídos e interpretados a lo largo de la sección 4.1. La presentación de los datos se encuentra holgadamente descrita en el CAPITULO 4.

4.3 HSP: Nivel 3

En el tercer nivel de la HSP se retoman todos los archivos de la sección 4.1 y ahora se trabajan con respecto a los análisis de probabilidad. Aquí las comorbilidades se extraen por regiones y cada región dividida en 9 subconjuntos (9 comorbilidades). El objetivo es armar una red con la retroalimentación de manera dinámica; así, si cambia alguno de los índices de contagio en alguna región el cálculo sea retroalimenta de inmediato (para ello se utiliza una Red Bayesiana).

Es importante resaltar que el cálculo de las probabilidades de las comorbilidades se debe hacer con eventos independientes sin una intersección. Actualmente los datos no están con esa modalidad. Los 9 subconjuntos están inter-seccionados. En otras palabras un paciente no necesariamente tiene una sola comorbilidad; en la gran mayoría desgraciadamente, los pacientes sufren más de una comorbilidad (ver la Figura 5).

Para el tratamiento de los eventos dependientes y su intersección. La autora propone resolver el problema con teoría de subconjuntos (Ecuación 4.1 - Ecuación 4.3) como se presenta en la Figura 6 (Nolasco-Jáuregui, O., Quezada-

Téllez, L., Rodríguez-Torres, E., y et al., 2021). La teoría de conjuntos es muy utilizada en lógica matemática en diagrama de Venn y sus derivados. Hay autores que hablan de éste y otros tratamientos para iniciar con un análisis tipo bayesiano (Nagarajan, R., Scutari, M., y Lébre, S., 2013; Scutari, M., y Denis, J., 2014; Spiegel, M., Schiller, J., y Srinivasan, R., 2013).

Nota: hay dos análisis separados: para pacientes tipo H y NH. Los acrónimos para los 9 subespacios son (Figura 5): *HI=Hipertensión*, *OB=Obesidad*, *IN=Inmunosupresión*, *A=Asma*, *D=Diabetes*, *RE=Renal*, *OE=Otras Enfermedades*, *EP=EPOC* y *CA=Cardiovascular*.

Definiendo a $P(A_m)$ como la probabilidad de que un paciente con COVID-19 se encuentre en una de las $m = \{1, \dots, 5\}$ regiones. Entonces $P(\overline{A_m})$ es definida como probabilidad complementaria en la cual un paciente está en otra región.

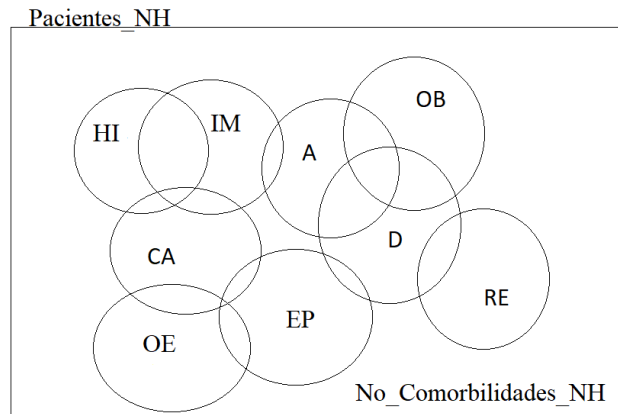


Figura 5. Representación de las Comorbilidades donde el espacio de los NH pacientes totales es igual a la intersección de los 9 subespacios y el resto contiene el área de *No-Comorbilidades-NH*.

$$P(A_m) = 1 - P(\overline{A_m}) \quad (4.1)$$

La intersección entre las regiones y sus comorbilidades se definen como:

$$P(A_m \cap B_n) = P(A_m) * P(B_n|A_m) \quad (4.2)$$

Donde $P(B_n|A_m)$ es la probabilidad de la B_n región tal que las comorbilidades $A_n = \{1, \dots, 9\}$. Como consecuencia la probabilidad de cada comorbilidad es obtenida mediante:

$$P(B_n) = P(A_m) - P(A_m \cap B_n) = P(A_m) - k * P(A_m) * P(B_n|A_m) \quad (4.3)$$

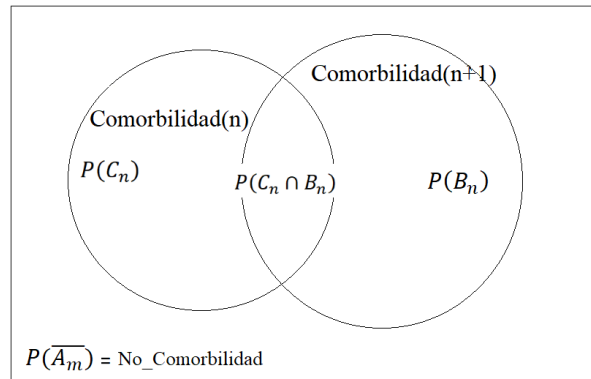


Figura 6. Cálculo de la Suma de la Probabilidades de los n subconjuntos de las comorbilidades y sus intersecciones en la m región.

Donde k es un coeficiente que es proporcional al número de intersecciones con los otros subconjuntos y tiene valores $k = \{0, \dots, 1\}$, ver Figura 6. Con base en la última ecuación 4.3 se generan unas tablas que serán la entrada a la red Bayesiana de acuerdo a los niveles jerárquicos de extracción de los datos (ver Figura 7).

La Figura 7 describe la Red Bayesiana con el gráfico tipo *DAG*, su relación entre los nodos hijos y su relación con las hojas de la red. Para la definición de la red Bayesiana se utiliza las bibliotecas *bnlearn* y *Rgraphviz* en el lenguaje R para generar las salidas de las probabilidades. Es importante notar la relación directa entre los *Pacientes* (P) y su tipo de clasificación *NH* y *H*. Seguido entonces de la clasificación jerárquica entre las 5 regiones (R1, R2, R3, R4 y R5). Subsecuentemente la dependencia entre las 9 comorbilidades es calculado en dos categorías más: *Comorbilidades_H* y *Comorbilidades_NH* (ver Figura 7).

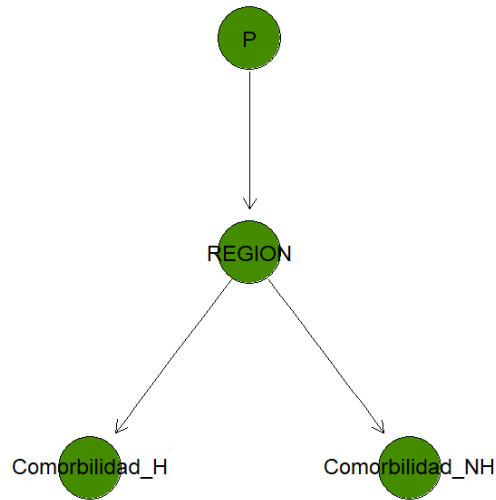


Figura 7. Gráfica Bayesiana con gráfico tipo *DAG* y la función *model2network* para el cálculo de los valores de la red.

4.4 Nivel 4: HSP

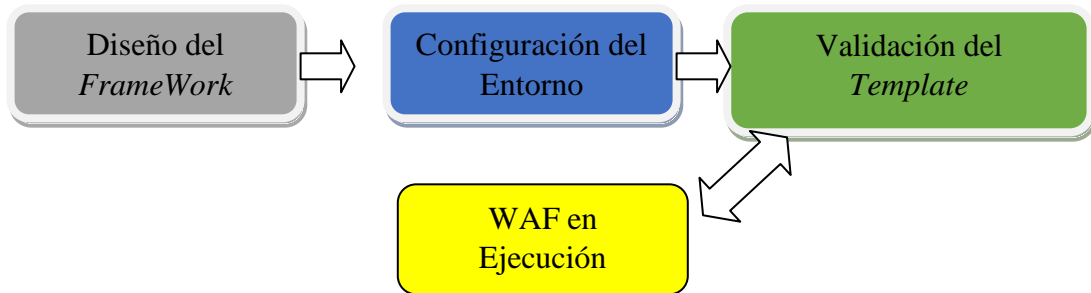


Figura 8. Diagrama a bloques de los elementos para la WAF: 1) Diseño del *FrameWork*, 2) Configuración del entorno, 3) Validación del *Template* y 4) WAF en Ejecución.

El objetivo principal del nivel 4 de la HSP es describir y engendrar una aplicación en software llamada WAF para la difusión de información sin fines de lucro que es vital para esta pandemia.

En el nivel 4 de la HSP se resume en la Figura 8 que contiene un diagrama mínimo a bloques del procedimiento para codificar la WAF. La Figura 8 es una propuesta de la autora que resulta de los aprendizajes que se necesitaron para

diseñar, desarrollar e implementar la WAF. La Figura 8 rige las divisiones de esta sección 4.4.

Esta WAF surge por la necesidad de difusión de información regionalizada, se retroalimentará de la información que los ciudadanos aporten y tiene un servicio de tipo público.

4.4.1 Diseño de *Python Flask Framework*

La selección del software para la codificación de la WAF es en *Python* (lenguaje de programación de alto nivel) y *Flask* (desarrollado por Armin Ronacher y sus librerías están escritas en Python) con una *Framework* (es una abstracción en software hecha para unir herramientas del tipo visual e interactuar con el usuario) que une *HTML* (es el lenguaje para el diseño de herramientas en la web). También tiene secciones escritas en *JavaScripts* (lenguaje de usos múltiples, tales como programación orientada a *objetos*, *clases*, funciones dinámicas, etc.) y entre otros (GoodFirms, 2020; Monnich, A., Ronacher, A., Lord, D., y et al., 2015).

El *IDE* que seleccionamos (de sus siglas en inglés Integrated Development Environment, en concreto es el ambiente donde el programador carga y desarrolla las codificaciones) es de *JetBrains* conocido como el de *PyCharm* 2020.3.2 (JetBrains, 2021) para Windows con una licencia de *ID T35185X9L2* con un costo mensual hasta el momento de casi 20 USD.

En este caso se seleccionó a *Flask* y *PEP 3333* (Python Web Server Gateway Interface v1.0.1) (Python Software Foundation, 2021). *Flask* es una excelente *framework* en el cual se pueden desarrollar aplicaciones web y aquí se utiliza el *Flask* versión 1.0.

Esta WAF es del tipo *blog*, donde múltiples usuarios pueden publicar comentarios (sólo usuarios registrados) los cuales tienen fecha y hora. El administrador de la WAF puede borrar *post* (aquí *post* se utilizan como una función tipo diccionario y está guardado en la base de datos por cada usuario

registrado, ver ANEXO C) de cualquier usuario cuyo contenido no sea real y/o de contenido peligroso.

4.4.2 Configuración del Entorno de la *Flask Framework*

El primer paso es la configuración del ambiente. En este caso el *IDE* seleccionado es *PyCharm* con *Python*, *PHP* (HTML) y *Flask* (Monnich, A., Ronacher, A., Lord, D., y et al., 2015). Por ende al crear el proyecto se debe seleccionar que el intérprete será *Python*. Con ello reconocerá los archivos con extensión *.py*; la configuración del proyecto es la parte medular ya que aquí es donde se deben instalar los paquetes que son necesarios para su correcto funcionamiento. El primer paquete que debe instalarse es *PIP* que es la base para el manejo de *Python 3*. Es decir, es una herramienta que permite la instalación, administración de las librerías y dependencias de *Python*. Una vez instalado *PIP* ya se puede instalar *Flask* y sus dependencias. Para comprobar su correcta instalación se debe ir a la *terminal* de *Python* e importar *Flask*. Si no marca error ya está correctamente instalado. La sugerencia inicial al crear el proyecto es hacerlo en una carpeta raíz dedicada al proyecto (llamada *App*). Enseguida elegir la opción de *empty Project* y finalmente seleccionar el intérprete de *Python*. Es necesaria también una carpeta llamada *static* para los archivos del tipo estático (como el *main.css*, *forms.py*, *models.py*, *run.py* y *routes.py*) y otra carpeta para los *templates* que están escritos en lenguaje HTML.

4.4.3 Validación y Configuración de los *Templates*

El archivo *main.css* es del tipo CSS (de sus siglas en inglés *Cascading Style Sheets*) que contiene el estilo de múltiples páginas web que comparten ruteos en HTML; tales como colores, estilos de letra, etc. (Wilke, C., 2018). Esta manera de escribir el código de las múltiples páginas provee más flexibilidad y control en la presentación de las características que comparten reglas en la sintaxis del formato (ver Figura 9).

En el archivo *routes.py* se definen las rutas que tendrán las múltiples páginas web y sus métodos tales como: *home.html*, *about.html*, *register.html*, *login.html*, *logout.html*, *account.html*, *create_post*, *post_id*, *post_id/delete*, *user*, *reset_password* y *reset_password/token*.

El archivo de *form.py* contiene las 6 *FlaskForms* definidas como tipo *class* que se van a utilizar en esta WAF llamadas: *RegistrationForm*, *LoginForm*, *UpdateAccountForm*, *PostForm*, *RequestResetForm* y *ResetPasswordForm*.

Hay un tutorial muy detallado que describe varios consejos que fueron muy útiles en este diseño (Shcafer, C., 2021). La diferencia de codificación más grande con respecto a Corey Shcafer son las variables de entorno (*environment configurations*). Aquí se tiene un *IDE de PyCharm* y por ende sus configuraciones son muy diferentes.

4.4.3.1 *Bootstraps*

Los *bootstraps* son los estilos de los *templates* que contienen *meta-tags*, *JavaScripts (JS)* y configuraciones en *CSS*. Los *bootstraps* son configuraciones tipo estándar donados por algunos programadores a la comunidad de *Open Source* y que se encuentran en diferentes lenguajes como *JS*, *CSS* ó *HTML*. El estilo de los *bootstraps* puede ser modificado a conveniencia ya que los autores lo permiten (pero es muy importante leer el uso y condiciones de sus códigos). Es posible modificar los contenedores, las divisiones, agregar barras de navegación, configuración de ventanas, texto interactivo, etc.

Los estilos de los *templates* es del tipo *inheritance* a la múltiple extensión de archivos *HTML* dividida en bloques. Lo que permite una mejor organización del código del archivo *layout.html* y se divide de acuerdo a sus funciones internas (hijos). Es posible utilizar las librerías de *code-snippets* de los *bootstraps* por cada extensión y modificar sus características en cada sección.

La manera de cambiarse de localidad dentro de las extensiones es agregando en el código de HTML a el método: `url_for("about.html")` de la librería de *Flask* y se incluye el nombre del archivo.

El archivo *about.html* contiene en texto la descripción de la página y una pequeña bibliografía de los autores de la WAF. El archivo *account.html* contiene las divisiones en clases de la forma de las cuentas de los usuarios y sus diccionarios (ver ANEXO C y Tabla 6). En el archivo *account.html* se visualiza el nombre del usuario, la fotografía o imagen cargada, los métodos *POST* y muestra error cuando se desea cargar un archivo diferente a un tipo imagen como *JPG*, etc.

El archivo *create_post.html* crea una función *post* si el usuario está activo es decir, que existe un registro del usuario y está abierta la sesión. El archivo *home.html* es la página principal y muestra la *form* configurada como *default*. Además contiene las funciones tipo *post* existentes y los organiza por paginación. Se crea un diccionario *posts* con la estructura de la Tabla 7, ver ANEXO C.

El archivo *layout.html* contiene el código de los métodos *head* y *body* de todas las rutas de las páginas web. Es aquí donde se definen las clases para el *container*, las barras de desplazamiento, navegadores, pestañas, etc. El archivo *login.html* contiene las *forms* tipo grupo que se utilizarán para adquirir los datos del usuario y corroborar su validez en caso de estar registrado (utiliza el método *POST*). El archivo *register.html* contiene las *forms* tipo grupo que se utilizarán para adquirir los datos para el registro del nuevo usuario (corroborar si ya existe en la base de datos) y envía error. Además que corrobora su validez por cada uno de los campos utiliza el método *POST* y al terminar envía al usuario a `url_for()` de `route_login()`. El archivo *reset_request* manipula la sección de contenido y la *form* de grupo para configurar los campos de: *password*, validez del campo de *email* en la base de datos y el método de *submit*. El archivo de

user_post.html contiene la *form* para borrar una función tipo *post* (sólo puede ser borrado por el usuario que lo creó).

4.4.3.2 Emulación

Para la emulación de la WAF y descartar errores se utiliza la *IP 127.0.0.1* que tiene un propósito especial *IPV4*. Además es muy importante la configuración del *localhost* y el puerto *4555* (Mitchel, B., 2020). Si se utiliza la configuración correcta de esta ruta se conectará automáticamente; y se comprueba mandando y recibiendo mensajes de la *TCP/IP stack*.

Para confirmar que las propiedades sean visualmente correctas se debe copiar la dirección *url* <http://127.0.01.:4555/> en el buscador *google chrome*. Cada vez que haya cambios en el código del proyecto *App* se debe refrescar las propiedades con *F5* en el buscador.

El código se puede correr desde la línea de comandos con la asignación del proyecto *App* junto con las propiedades de la librería de *Flask* codificando: *app.run(port=4555, debug=True)*. De esta manera se activará una sesión y se ingresa con un número *PIN* en modo *debugger*.

4.4.3.3 Levantamiento de los Módulos *Flask*

La manera de invocar a las funciones de *render* de los *template* que pertenecen a la librería de *Flask* es por medio de propiedades como *import* y *from*.

Dentro de las funciones de *Flask* utilizadas son: *render_template*, *flash*, *redirect*, *request*, *url_for* y *abort*; todo esto descrito en el archivo *routes.py*. En el archivo de *models.py* se utiliza de *flask_login* la librería llamada *UserMixin*. Para el archivo de inicialización se utilizan funciones como: *flask_sqlalchemy*, *flask_bcrypt* y *flask_login*. En el archivo de configuración de las *forms* se utilizan las funciones que configuran a *FlaskForm* como: *flask_wtf* y *flask_wtf.file*.

Un módulo elemental que provee *Python* es *OS*. Éste provee las funciones de interacción con el sistema operativo, por ejemplo: *os.getcwd()*; que permite al programador ubicarlo en la dirección de la carpeta en la que se encuentra el archivo del proyecto y donde está la sesión activa. El modulo de *datetime* contiene la consulta al sistema operativo en tiempo real, captura la fecha y hora local.

Es importante mencionar que si se desea traer de un archivo a otro, de una clase o de método se puede importar tal cual y se logra con los otros *módulos*. En el archivo *routes.py* se incluyen módulos definidos en el archivo *models.py* de la siguiente manera: *from App.models import User, Post*; de igual manera en el archivo *forms.py* se incluyen módulos definidos en *models.py* con la siguiente codificación: *from App.models import User*.

4.4.3.4 Configuración de las *Forms*

La esencia de crear una *form* para cada archivo es para poder recibir y enviar parámetros al usuario de manera visual. De manera interna se crea una clase y método por cada uno de las *forms* con el respectivo manejo de los objetos.

Es importante agregar las librerías que se van a utilizar para el manejo y configuración de las *forms flask_wtf* codificando: *from flask_wtf import FlaskForm*.

1) *Form* del Usuario

La *form* para el registro de usuario (*RegistrationForm*) se va a crear en una clase que recibe como argumento el parámetro llamado *FlaskForm*. En esta *form* se tienen varios campos donde se van a adquirir los datos del usuario a registrarse como: *username, email, password, confirm_password, accept_tos* y *submit*. Para poder editar estos campos es necesario incorporar algunos métodos de las bibliotecas de *wtfirms* tales como: *StringField, PasswordField, SubmitField, BooleanField, validators* y *TextAreaField*. También se necesita la *flask_wtf.file* e importar *FileField*. Hay algunas restricciones que deben

considerarse para los campos que serán recibidos por las variables tipo *StringField*. Es importante generar un error si el campo se deja limpio o vacío y mandar leyendas para el usuario (o dejar notas en el *template*) de tal manera que hay un número restringido de caracteres. Para el campo de *username* que recibe los datos con la función de *StringField* es necesaria una restricción entre 4 y 25 caracteres (es lo que resulta más conveniente). Para el campo de *email* que recibe los caracteres con *StringField* también la restricción es de mínimo 6 y máximo 35 caracteres.

Para hacer una verificación que los campos tales como: *username*, *email*, *password*, *confirm_password* y sean introducidos correctamente por el usuario antes de guardarlos en la base de datos es necesario utilizar las *funciones de validación* que trae *Flask* en sus librerías.

Se define la *form* de *LoginForm* que tiene cierto parecido a la *RegistrationForm*. Se va a crear en una clase que recibe como parámetro las características de las *FlaskForm*. En esta *form* se tienen varios campos donde se van a adquirir los datos del usuario que ya está registrado en la base de datos; datos como *email*, *password*, *remember* y *submit*. Para poder editar estos campos es necesario incorporar algunos métodos de las bibliotecas de *wtfoms* tales como: *StringField*, *PasswordField*, *SubmitField* y *BooleanField*. Hay algunas restricciones que deben considerarse para las variables tipo *StringField* en *LoginForm* (como *email*). Es importante generar un error si el campo se deja limpio (o vacío) y mandar leyendas cuando el usuario registrado no coincide con la base de datos (es necesario utilizar *funciones de validación*). Para la variable de *password* que es del tipo *PasswordField* es importante generar un error si el campo se deja limpio (o vacío) y mandar leyendas cuando el usuario registrado no coincide con la base de datos. En el caso del campo *remember* que es del tipo *BooleanField* ya que el usuario decide si quiere que la sesión se quede abierta recibiendo *True* o *False*.

2) *Form* para los diccionarios de *posts*

De la misma manera la *form* para la función de registro de *post* (*PostForm*) tiene varios campos donde se van a adquirir los datos del usuario a registrar un nuevo *post* como *title*, *content* y *submit* (ver Tabla 7). Para poder editar estos campos es necesario incorporar algunos métodos de las bibliotecas de *wtfoms* tales como *StringField*, *SubmitField* y *TextAreaField*. Hay solo una restricción a la hora de crear un *post* y es que se genera un error si alguno de los campos se deja limpio (o vacío).

3) *Form* para recuperar contraseña

Para la situación donde el usuario ya está registrado pero no recuerda su *username* o *password* de registro se crea la *form* llamada *RequestResetForm* que recibe como parámetro las propiedades de las *FlaskForm*. La *form* de *RequestResetForm* tiene dos campos donde se van a adquirir los datos del usuario a recuperar (al menos debe recordar el campo de *email* de registro). Para poder editar estos campos es necesario incorporar algunos métodos de las bibliotecas de *wtfoms* tales como *StringField* y *SubmitField*. Para el campo de *email* que recibe los caracteres con *StringField* con la restricción es de mínimo 6 y máximo 35 caracteres. Es importante mencionar que se va a generar un error si el campo se deja limpio (o vacío) y manda leyendas cuando el *email* del usuario registrado no coincide con la base de datos (ver Tabla 6).

En el caso donde el usuario hizo una solicitud de cambio de *password* las instrucciones fueron recibidas en el email del usuario y la *form* que el usuario verá al darle clic al enlace enviado a su email será la *form* de *ResetPasswordForm* (que recibe como parámetro las propiedades de las *FlaskForm*). La *ResetPasswordForm* tiene tres campos donde se van a adquirir los datos del usuario del *password* a cambiar. Para poder editar estos campos es necesario incorporar algunos métodos de las bibliotecas de *wtfoms* tales como: *PasswordField* y *SubmitField*. Hay algunas restricciones que deben

considerarse para las variables tipo *PasswordField* en *ResetPasswordForm*. Para el campo de *password* es importante generar un error si el campo se deja limpio (o vacío) y es indispensable utilizar *funciones de validación*. Para la variable de *confirm_password* que es también del tipo *PasswordField* es importante generar un error si el campo se deja limpio y mandar leyendas cuando *password* no coincide con *confirm_password*.

4.4.4 Funciones de Configuración

Para decidir el nombre del dominio es importante visitar sitios como www.namecheap.com; donde se puede saber qué nombres de dominios ya existen y seleccionar otro nombre. Es muy importante el pago de un dominio para que la WAF sea visualizadas por la web y la renta de un *manejador de la cloud* (Tatham, S., 2020). Dentro de las características más importantes al rentar es que tenga una conexión vía *SSH Access* (Linode, 2020). Hay servicios que ofrecen *node servers* donde además puedes pagar por un gran volumen de ellos. En este caso con un *nodo* (1CPU), 25G de *Storage* y 1GB de RAM es suficiente para la WAF. Otro parámetro a seleccionar en la renta es la *región*: puede ser America, Europa, Asia, etc. Es comprensible que por cada característica a seleccionar el costo de la renta mensual aumente, pero el servicio para nuestra WAF será modesto. Enseguida se configura la *password* del *administrador* con la que se puede aumentar el *performance* en cualquier momento, aumento en las carpetas de *storage*, etc. Finalmente, cuando se realiza el pago se coloca la dirección *url* en la línea de comandos y con los comandos *SSH* se tendrá comunicación con el servidor; pero es necesario instalar también un intermediario como *PuTTY* (Tatham, S., 2020).

La utilización de *PuTTY* es diferente según el sistema operativo a utilizar y no es la única opción para los esquemas que *SSH* trabaja. *PuTTY* es la opción que la autora decidió para este manejador (Tatham, S., 2020). *Putty* es un tipo de *SSH* que tiene configuraciones de tipo *telnet client* y se utiliza cuando el usuario

necesita la generación de un *llave encriptada* (se configura dentro de un archivo llamado *authorized_key*)(Linode, 2020).

1) Funciones de carga al Servidor

Hay múltiples maneras de cargar el proyecto *App* tipo *flask* en el servidor. Una de ellas es usar la librería de *PIP* y el método de *freeze* con lo que se hace un tipo de encriptado de todas las dependencias que tiene nuestro proyecto *App*. Es decir, todas las librerías y paquetes que fueron necesarios para configurar el *enviroment* se puede configurar en un archivo llamado *req.txt*. Este archivo contiene la lista de esas dependencias. El archivo *req.txt* se debe copiar en la raíz de la carpeta del proyecto. Para la configuración del *ambiente virtual* del proyecto *App* en el servidor se debe instalar: *python-env* y *python-pip*. Y con el comando: *source venv/bin/activate* se carga el *ambiente virtual* seguido de la siguiente manera: *pip install -r req.txt*.

En la emulación del proyecto *App* en el servidor local se utilizaron *llaves encriptadas* tales como: la *SECRET_KEY*, la *SQALCHEMY_DATA_URI*, *EMAIL_USER* y *EMAIL_PASS*. Que son datos del administrador y se deben de cambiar antes de dar clic a *load*. Debes de generar esta información de manera *aleatoria* en el *prompt* (por seguridad y tratarse de datos sensibles). Toda esta información se agregará a un archivo de configuración que se debe crear en el servidor con el método *JSON* en el archivo llamado *config.json* con el método *touch*.

4.4.5 WAF en Ejecución

Continuando con la configuración del *ambiente virtual* en el servidor y para cargar el proyecto *App*. Se van a utilizar los comandos de *export* quien recibe al archivo de *run.py*. Se debe correr con el número de host codificado de la siguiente manera: *flask run -host=0.0.0.0*.

Una vez que se instale en el servidor *gunicorn* de la librería *PIP* se utiliza para la configuración del alias de las *location* de las funciones *routes* e incluso se

puede cambiar el *server_name* que se está rentando por un alias ya que *unicorn* es un manejador del servidor y del código en *Python* en *CSS*.

Seguido de la instalación del método supervisor que será el encargado de hacer un *autostart* del proyecto *App*. En el caso de aparecer un error tipo *crash* se trae la codificación: *App sutorestart=true, stopasgroup=true*; al igual para matar procesos y garantizar su funcionalidad de manera correcta se utiliza: *killasgroup=true*. Para ello es necesario agregar las líneas de código en el archivo de configuración. Una vez que se tienen las configuraciones del *DNS* (Linode, 2021) en el servidor y las claves del administrador (para el control de los datos) en este momento se tiene la ¡WAF está en ejecución correctamente en la web!.

1) *Templates* en Ejecución

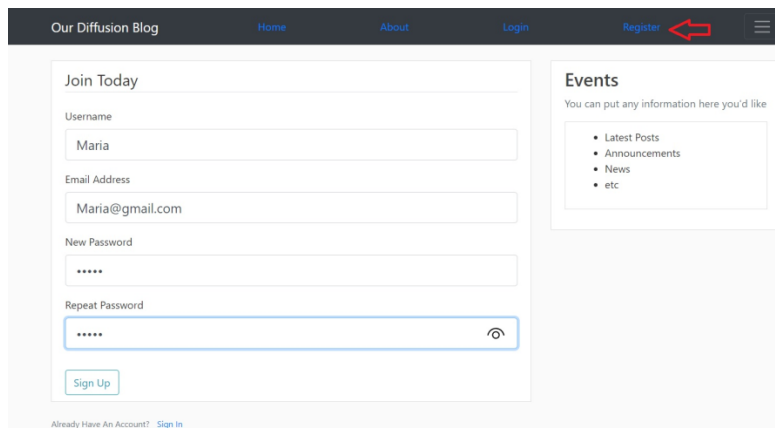


Figura 9. La flecha roja indica la activación del *template* que utiliza el método *POST* de la *RegistrationForm()* para dar de alta a un nuevo usuario.

El archivo *register.html* contiene el *template* que une la *form* de *RegisterForm* con sus campos. El archivo *register.html* se define por una propiedad *content-section* tipo *form-group* activado por un método *POST* que contiene una serie de *border-bottom* del *mb-4* con una leyenda de *label=Join Today* para su identificación. Además de estar dentro de una clase y una división del *grupo general*.

Todos forman una *form.hidden_tag()* en un *fieldset* dentro de la *form-control* donde se encuentran definidas otras *forms* tales como: *form.username.label*, *form.email.label*, *form.password.label*, *form.confirm.password* y *form.submit* del tipo *btn-outline-info*. Con texto configurado como: *muted* y *small class* con la *label= Already Have Account?*. Entonces la función de *href=/login* dirige a la localidad del *template login.html* (ver Figura 9). Lo anterior sucede cuando el usuario a registrarse activa el evento de *Sign Up* que llama a la *template* de *login.html*.

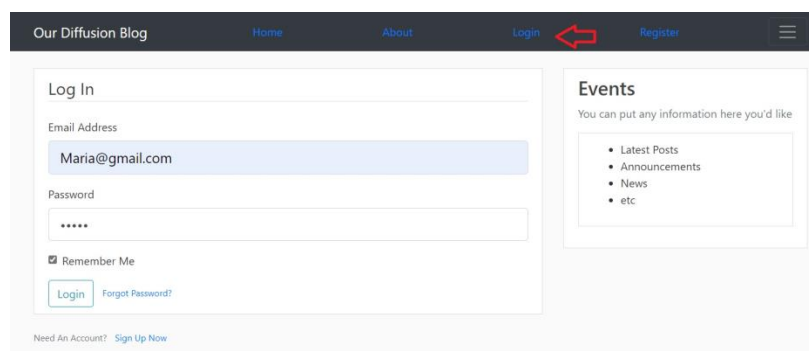


Figura 10. La flecha roja indica la activación del *template* de *Login* que utiliza el método *POST* de la *LoginForm()* para abrir una sesión al usuario registrado.

El archivo *login.html* contiene el *template* que une la *form* de *LoginForm* con sus campos (ver Figura 10). El archivo *login.html* define una *content-section* tipo *form-group* activado por un método *POST* que contiene una serie de *border-bottom* del tipo *mb-4* con la leyenda: *label=Log In* para su identificación. Además de estar dentro de una clase y una división del *grupo general*. Todos forman una *form.hidden_tag()* en un *fieldset*. Dentro de la *form-control-label* se encuentra *form.email.label*, *form.password.label*, *form.remember* y *form.submit* del tipo *btn-outline-info*. Con texto configurado como: *muted* y *small class* con la leyenda: *label=Need An Account?*. Entonces se manda invocar al evento *href=/register* de la *template register.html* (ver Figura 10). Cuando el usuario a registrarse activa el evento de *Login* se manda a la localidad de la *template home.html*. Si el usuario desea recuperar el campo

del *password* entonces la *class* tipo *text muted ml-2* llama al evento *href=/reset_request* en el *template* de *reset_request.html*.

Hay que recordar que a partir de este *template* se están verificando cada uno de los campos (ya que se trata de un usuario registrado en la base de datos). En caso de no ser coincidentes se activan mensajes llamados *flash* (definidos en la librería que lleva el mismo nombre de *flash*) y son los que se activan para una orden de manera preventiva (o de error).

La *class* del *template* de *account.html* es del tipo *content-section media* tiene características muy peculiares como una *class* del tipo *rounded-circle* para *account-imag* cuyo método *src* es del tipo *image_file*. Aquí es muy importante puntualizar que dentro del diccionario de cada usuario se agrega un campo para la imagen *current_user.image* del tipo *default*. En el caso que el usuario no tenga una selección de fotografía para su perfil (ver la Figura 11, ahí se observa la imagen *default*) se activa la imagen *default* para todos los usuarios de nuestra WAF.

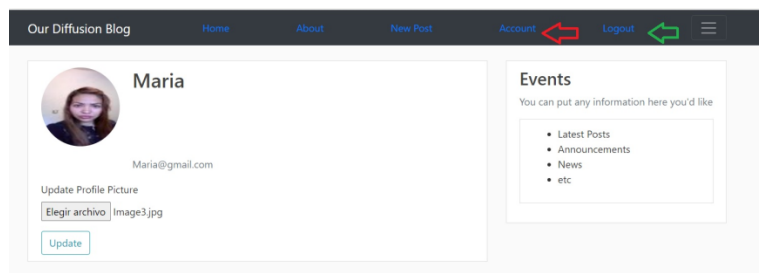


Figura 11. La flecha roja indica la activación del *template* de configuración del usuario, como son los campos de *mail*, *username* y selección de *imagen* que utilizan el método *POST* de la *UpdateAccountForm()*. En verde aparece la opción de *logout()*.

El proceso de carga de archivo para la *imagen* con la codificación: *f=form.picture.data*. Da validación a la extensión del archivo cargado con la siguiente codificación: *f.file.rfind('.pdf')*; y sólo acepta extensiones: *.pdf*, *.png*, *.jpeg* y *.gif*. Enseguida se sustituye en el campo de *imagen* del diccionario de la

base de datos del usuario con `f.save(f.filename)` y `old_image=Image.open(f.filename)`.

Hay un consejo que es muy interesante recordar y son las dimensiones de la *imagen* que el usuario tiene en el perfil. Se convierte una vez que se cargue la imagen a la *form* esto se hace con la codificación: `old_image.thumbnail(125,125)` y se remueve la imagen anterior con: `os.remove(f.filename)`. Enseguida se agregan los nuevos valores al diccionario de la base de datos con `db.session.add(current_user)` y `db.session.commit()`. Una vez que se carga la nueva imagen se regresa al mismo *template* de `account.html` visualizándose la nueva imagen seleccionada por el usuario.

El *template* de `home.html` su contenido está regido por un ciclo *For* de los `post.items`. Es decir, en este *template* se tiene todos los diccionarios tipo `post` de todos los usuarios organizados por el método de *fechas* en un historial hasta el *post* más actual (ver Figura 12).

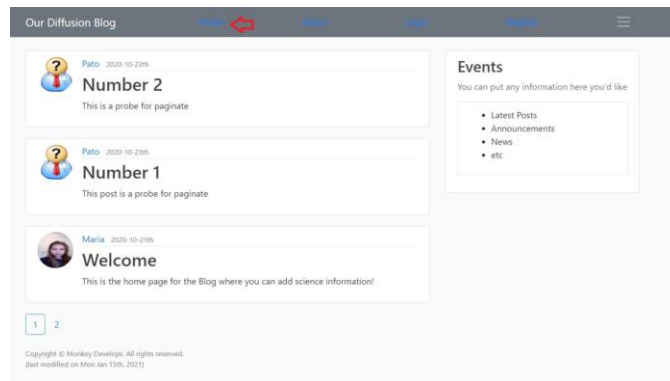


Figura 12. La flecha roja indica la activación del *template* de `home.html`. Aquí la autora publica un mensaje de bienvenida al blog.

En la Figura 12 se muestran 3 publicaciones de prueba que están en el *template* de `home.html`. Aquí son dos usuarios diferentes del historial de *posts* organizados por el método de *fecha*. Cada *post* tiene 5 campos tales como: `post.author.image`, `post.author.username`, `post.data_posted`, `post.title` y `post.content`. Para que la lista por *fecha* de los *posts* aparezcan en el *template* de

home.html es debido a que los *posts* se encuentran en la base de datos y se utiliza un ciclo *For* para organizarlos: *post.iter_page* y *page_num*.

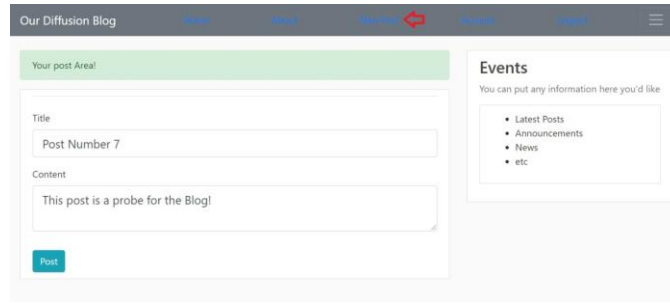


Figura 13. La flecha roja indica la activación del *template* de *posting.html* que utiliza los métodos de *GET* y *POST* de la *PostForm()* para crear una publicación en el blog.

También en la Figura 12 aparece un *text-muted* que se encuentra en la parte baja de la imagen. Es una *class* tipo *small* que depende de una propiedad tipo *footer* que informa la última modificación a la WAF y los derechos reservados a los autores del dominio llamado *OurDiffusionBlog*.

El *template* de *create_post.html* utiliza una *class* tipo *media content-section* que se activa con el método *POST* y es del tipo *form.hidden_tag()* que se conecta con función *new_post()* en la función de *location/create_post* (ver Figura 13).

Hay una verificación de que el usuario esté activo con la codificación: *current_user.is_active*. Los campos de la *form.title* y *form.content* entonces se guardan en el diccionario del usuario cuando crea un diccionario tipo *post*. Después se activa la *form.submit* lo cual permite que se cree un *nuevo post* en el diccionario del usuario. Enseguida de crear un *post* se redirecciona al usuario al *template* de *home.html* para que el usuario visualice su publicación, ver Figura 13.

En el *template* de *posting.html* se encuentra en una función tipo *block content* que describe una propiedad *media content-section* y necesita un argumento del tipo *post_id* a borrar, ver Figura 14.

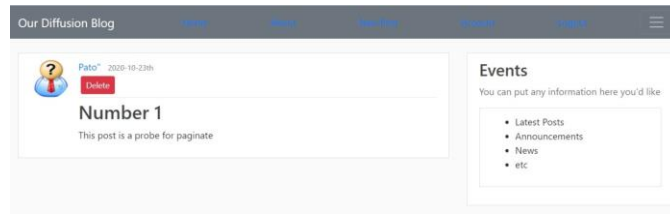


Figura 14. La función `delete_post(post_id)` necesita de `route` tal como: `@login_required` y tendrá el numero del `post` adecuado con el método POST: `@app.route("<init:post_id>/delete", methods=['POST'])`.

El argumento `post_id` se da a la función cuando el usuario hace clic sobre el `post`. En caso que el `post.author` sea diferente a `current_user` no podrá borrar el `post`. Entonces, no aparece el evento `deleteModal`. Pero en caso contrario, la codificación sería: `post.author == current_user` y la `class` tipo `modal` muestra el `template` de `posting.html`. Ya definida la ruta en la función `delete_post` se activa en el archivo llamado `routes.py` con la siguiente codificación: `post=Post.query.get_or_404(post_id)`. Es esta función la encargada de borrar en la base de datos, en el diccionario seleccionado con la siguiente codificación: `db.session.delete(post)`. Después se redirecciona al `template` de `home` con: `redirect(/home)`. Para que el usuario corrobore que fue eliminado su post (ver Figura 14).

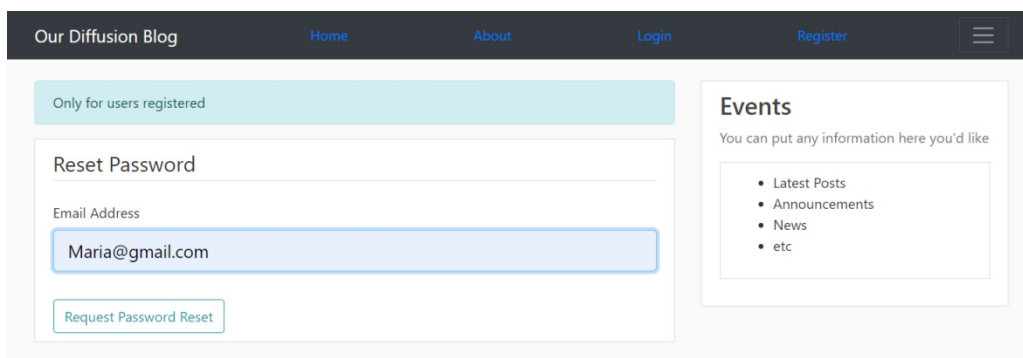


Figura 15. La `template` de `reset_request.html` la cual redirecciona a la `template` de `login.html` con la `form` de `RequestResetForm()` con los métodos `GET` y `POST` para solicitar el campo `password`.

Para el que el *template* de *reset_request.html* funcione correctamente el usuario debe solicitar el cambio del campo *password*. El cuál debe estar registrado para ello se utiliza se corrobora con: *current_user.is_authenticated*. Entonces se hace un filtrado en la base de datos con su campo *email* con la siguiente codificación: *User.query.filter_by(email=form.email.data)* (ver Figura 15). En caso que todo lo anterior sea *True* entonces se llama a la función: *send_reset_email(email)*.

4.5 Conclusiones

La HPS es el núcleo de esta investigación ya que surge de la necesidad del manejo e interpretación gráfica de los datos de COVID-19 en México (ver Figura 2). La HPS es una propuesta inédita y se divide en 4 niveles. La HPS utiliza un algoritmo de búsqueda por coincidencias del método de *tokens* y cadena de caracteres.

La HPS en su nivel 1 hace toda la extracción y separación de características por archivos. La primera división en el nivel 1 es por medio del campo 8 referente a la búsqueda de la región de residencia (ver Figura 4). Seguido del tipo de paciente *NH* ó *H* con el campo 10. La segunda división se logra con el campo 13 correspondiente a la edad de los pacientes. La tercera división en el nivel 1 de la HSP se saca con el tipo de morbilidad que padecen los pacientes contagiados por cada región. La cuarta división se obtiene con la búsqueda y extracción de los pacientes que poseen todo lo anterior, están hospitalizados y necesitaron de un respirador artificial. Finalmente se realiza el conteo de los pacientes de la cuarta división que fallecieron en la *UCI*.

El tratamiento de los eventos independientes de las 9 comorbilidades (campo 20 al campo 28) es una propuesta de la autora y lo describe en el nivel 3 de la HSP (ver sección 4.3). El cálculo de la suma de las probabilidades de padecer una comorbilidad por cada región del país se calcula por medio de los *n* subconjuntos y sus intersecciones en las *m* regiones. Con el tratamiento anterior

ya es posible introducir los datos de entrada a la Red Bayesiana descrita por *bnlearn* en lenguaje R (Nolasco-Jáuregui, O., Quezada-Téllez, L., Rodríguez-Torres, E., y et al. ,2021).

Gracias al nivel 1 y el nivel 3 de la HSP se logra el cálculo de las probabilidades de las comorbilidades (con la Red Bayesiana) por las 5 regiones de México. Este método es dinámico, ya que cada cambio en la entrada será reflejado en la salida.

En el nivel 4 de la HSP se describe el diagrama a bloques (ver Figura 8) que la autora sintetizó cuando ejecutó la WAF en *Python* y *Flask* para la difusión de la información que en esta investigación se presenta. La Figura 8 es una aportación de esta investigación; ya que son los pasos mínimos necesarios para diseñar, desarrollar e implementar la WAF (y se describen a detalle en la sección 4.4).

Hay varios consejos que se describen a lo largo de la sección 4.4, uno de ellos es la utilización de las múltiples páginas cuando se programa en HTML. Esta WAF se plantea como una solución a la difusión de la información en época de pandemia donde cada ciudadano que desee compartir información vital la utilice. La WAF se encuentra en ejecución correctamente y sus diferentes *templates* están en funcionamiento; lo cual se comprueba en las 7 capturas de pantalla de la Figura 9 – Figura 15.

Esta WAF no pretende ser competencia a ninguna *Aplicación Web* que haya en el mercado. No pretende ser una sustitución y mucho menos busca un fin comercial. Dentro de los objetivos secundarios de este documento se encuentra mostrar una propuesta de diseño que puede mejorar al paso del tiempo (es una versión inicial) e incluso se pretende ser un guía de estudio para cualquier ciudadano que desee reproducir el contenido de esta propuesta.

La WAF es una idea de nuestro grupo de investigación con el objetivo de difundir nuestros avances en la línea de investigación que se ha venido

trabajando en los últimos meses sobre los datos COVID-19. Una vez que cubrimos nuestra necesidad se decide agregar un área de *publicaciones* para los ciudadanos (que en esta situación de pandemia es imprescindible estar informados).

Nosotros rentamos un nodo (1 CPU) con 25G de almacenamiento y 1GB de RAM e incluso pagamos por la renta del dominio *Our.DifussionBlog.com.mx* en la región de México /¿(características que fueron suficientes para nuestra WAF).

Desafortunadamente se tuvo que dejar de pagar estos servicios porque esta pandemia trajo una crisis económica a nuestros hogares y no recibimos apoyo financiero de ninguna institución. Sin embargo, la WAF está activa en un servidor local y justo la autora está trabajando en la próxima versión mejorada. De hecho para verificar su funcionalidad, se utiliza el url <http://127.0.0.1.:4555/> del navegador de *Google Chrome* que es un servicio gratis.

Con el diseño de la HSP se facilita el análisis de datos COVID-19 en México en sus 35 campos. La interpretación y presentación de los resultados de manera gráfica se encuentra en el nivel 2 de la HSP y es descrito a detalle en el CAPITULO 5 de este documento.

CAPÍTULO 5

ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

El presente capítulo describe la parte medular de esta investigación. Se refiere concretamente el análisis, manejo e interpretación de la información. El primer paso es retomar el uso de las variables involucradas en el nivel 1 y nivel 3 de la HSP. Las variables están clasificadas por región (ver Figura 16) y tienen una jerarquía (ver Figura 17).

5.1 Regiones en México y Acrónimos

Primero recordemos que los Estados Unidos Mexicanos (México) son 32 entidades federativas y cada uno de ellos tiene una situación política, económica, poblacional y social particular (Holmes, S., y Huber, W., 2018; Michonneau, F., y Fournier, A., 2018).



Figura 16. En verde se señala el territorio correspondiente a R1, en amarillo a la R2, en azul es R3, en gris R4 y en marrón el territorio de la R5.

Las entidades agrupadas por regiones permiten tener un análisis más eficiente de la información estadística. Por ello las regiones están denominadas como:

región Noroeste (R1), región Noreste (R2), región Occidente (R3), región Centro (R4) y región Sureste (R5). En la Figura 16 se muestra el mapa de México dividido por colores y las 5 regiones.

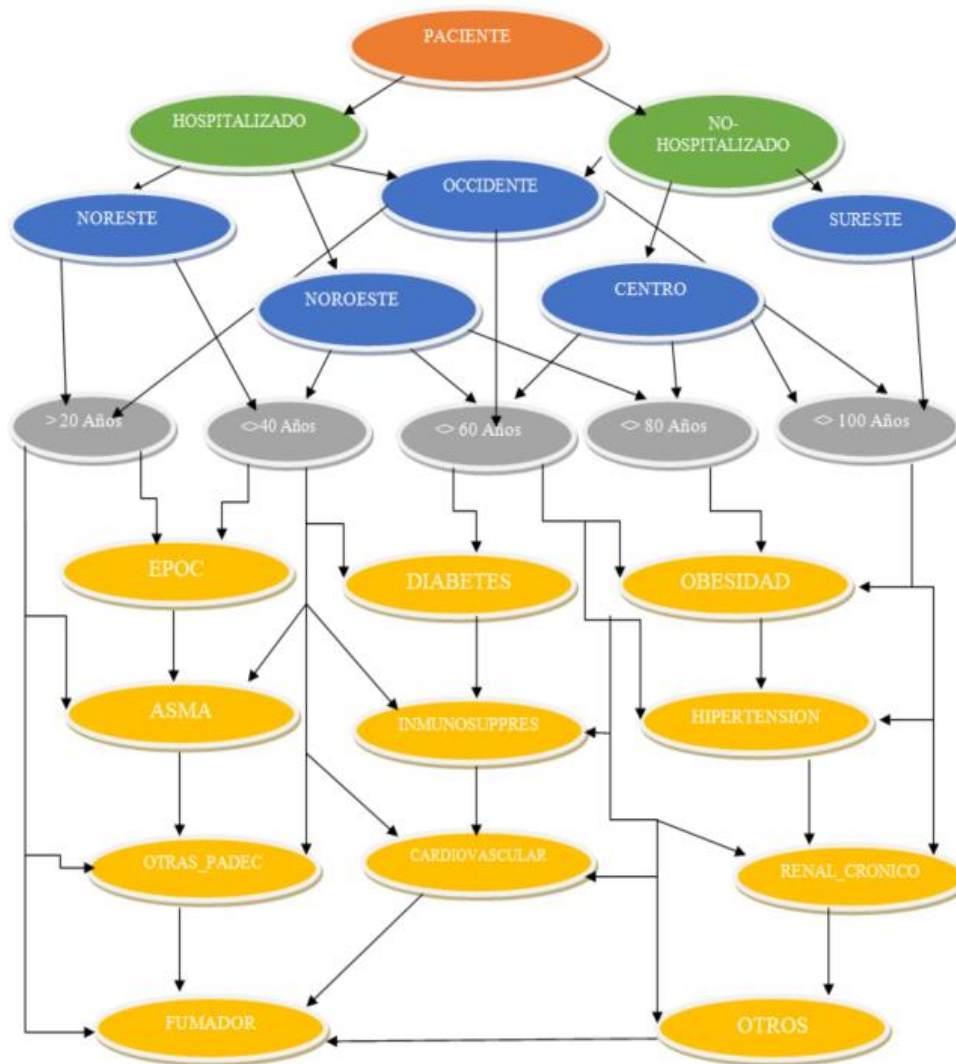


Figura 17. Red de Jerarquías en la extracción de los datos indicando con colores sus diferentes niveles en orden descendente: naranja, verde, azul, gris y amarillo (captura de pantalla extraída de Nolasco-Jáuregui, O., Quezada-Téllez, L., Rodríguez-Torres, E., y et al., 2021).

La región Noroeste de México (R1) agrupa las siguientes entidades federativas: Baja California, Baja California Sur, Chihuahua, Sinaloa y Sonora. En la región

Noreste (R2) se encuentran: Coahuila, Durango, Nuevo León, San Luis Potosí y Tamaulipas. Para la región Centro (R3) están: la Ciudad de México, Estado de México, Guerrero, Hidalgo, Morelos, Puebla y Tlaxcala. En la región Occidente (R4) se encuentra: Aguascalientes, Colima, Guanajuato, Jalisco, Michoacán, Nayarit, Querétaro y Zacatecas. Por último la región Sureste (R5) abarca los estados de Campeche, Chiapas, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

5.2 Jerarquía de las Variables

La jerarquía en las variables se visualizan en la Figura 17; la cual es también la metodología del proceso de extracción de los datos (ver sección 4.1) descrita en el nivel 1 de la HSP. Esta red de extracción asimismo indica la relación entre los niveles y dependencia que hay entre ellos. En el primer nivel de la jerarquía se encuentran todos los pacientes que fueron seleccionados de manera involuntaria al dar positivo al virus (color naranja). Le sigue otro nivel de clasificación de los pacientes: hospitalizados (H) y los que son puestos en cuarentena (NH) en casa bajo supervisión médica (en color verde). La siguiente categoría de clasificación corresponde a las 5 regiones del país descritas en la subsección anterior (color azul) R1, R2, R3, R4 y R5. La consiguiente categoría son los pacientes clasificados por edades, esto es por cada región y según el tipo de paciente (color gris). Y la última clasificación son las 9 comorbilidades registradas por el Sector Salud Federal (color amarillo). Cabe señalar que la base de datos analizada en esta investigación son a partir del 12 de abril al 31 de Octubre del 2020 (1,809,852 pacientes) dando un total de 3,654 archivos. Esto debido a que son 203 registros diarios por dos tipos de pacientes por 9 comorbilidades. Es decir, casi 56 GB de datos (más detalles en OraliaNJ, 2020).

5.3 Análisis de las Edades de los Pacientes

Para el análisis de edades se tienen 5 intervalos. El primer intervalo es de 0-20 años, el segundo de 21-40 años, el tercero de 41-60 años, el cuarto de 61-80 y el

quinto 81-100 años (ver Figura 17). El segundo nivel de la HSP tiene resultados en el análisis de edades como el que se muestra en la Figura 18. La Figura 18 presenta la distribución de las edades para pacientes de la R4 en 10 gráficos. Donde cada gráfico corresponde a un día de registro de edades de cada paciente que dio positivo al virus. En este caso, el análisis es tanto para pacientes H y NH. Por ejemplo, el primer gráfico es del 12 de Abril para R4N (esto es, pacientes No-Hospitalizados de la Región 4). El archivo 200412COVID19MEXICO R4H corresponde a la distribución de edades de los pacientes infectados el 12 de Abril en la región 4 y se encuentran hospitalizados. Así consecutivamente, se tienen análisis de edades por cada región y tipo de paciente (NH ó H) por los 203 días de registro (OraliaNJ, 2020).

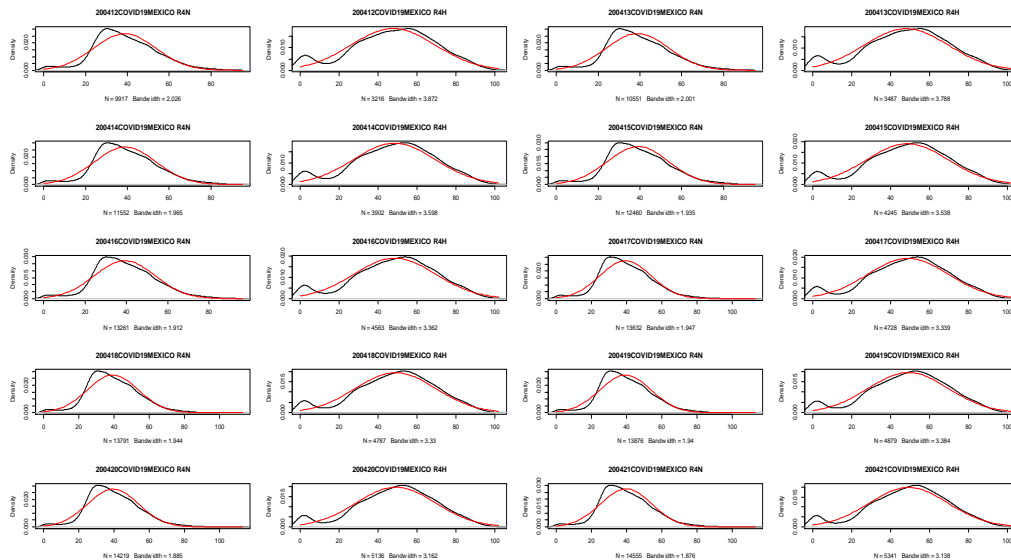


Figura 18. Distribuciones de Edades (0-100 años) en la R4 por 10 días.

5.3.1 Resultados de Análisis de las Edades de los Pacientes

Una vez obtenido la media y la desviación estándar por día. Se calcularon estos mismos estadísticos para los 203 días estudiados en las cinco regiones geográficas en el nivel 2 de la HSP.

Se puede mencionar que en la R1 los pacientes NH tienen una mediana de los promedios de la edad de 39.838 años y en el caso de los pacientes H es de 55.105 años. En la R2 los pacientes NH la mediana de los promedios de la edad de los registros diarios es de 38.9691 años y para los pacientes H es de 53.207 años. Para la R3 los pacientes NH es de 38.1323 años y para los H es de 53.4386. En R4 los pacientes NH es de 40.3153 años y los pacientes H es de 52.871. Para la R5 la edad de los pacientes NH es de 40.123 años y para los pacientes H es de 53.3629 años.

Se puede destacar que aquellas personas que han adquirido el COVID-19 se encuentran en los intervalos de edades de 20-40 y de 40-60 años. Para los casos positivos a COVID-19 en edades de 20-40 años no han requerido hospitalización, por lo que los síntomas presentados son leves y estos pueden ser tratados en casa. Ante una falta de monitoreo estricto se presume que este sector pasa a ubicarse en cuarentena para su recuperación. Por otro lado, aquellos pacientes que han sido hospitalizados se encuentran en el intervalo de edades de 40-60 años. La evidencia empírica muestra que estas distribuciones se mantienen independientemente de la región analizada, es significativo encontrar que el promedio de las edades y su desviación estándar son valores numéricos similares en todas las regiones (ver Figura 18).

5.4 Resultados por Comorbilidades en México

En la R1 se han infectado a 110,856 pacientes (NH=102,559 y H=8,297), en la R2 han sido 150,691 pacientes (NH=143,170 y H=7,521), para la R3 se tienen 180,375 (NH=170,030 y H=10,345), la R4 con 515,670 (NH=482,407 y H=33,263) y la R5 con 132,260 (NH=120,352 y H=11,908) (ver Anexo B y Tabla 5).

La Tabla 3 describe los pacientes que dieron positivo al examen de COVID-19 y reportaron no tener ninguna comorbilidades (se describe por cada región de México). En total se tienen 717,824 infectados sin tener antecedentes de

ninguna de las 9 comorbilidades con los siguientes resultados: NH=689,369 y H=28,455.

Ahora se menciona el desglose de los pacientes sin comorbilidad infectados por cada región: en la R1 se infectaron 69,606 pacientes (NH=66,762 y H=2,844), en la R2 han sido 98,979 pacientes (NH=95,886 y H=3,093), para la R3 se tienen 123,519 (NH=119,661 y H=3,858), la R4 con 348,531 (NH=333,919 y H=14,612) y la R5 con 77,189 (NH=73,141 y H=4,048). De estos, cerca de alrededor del 95% al 96% son pacientes tipo NH y del 5% al 4% requieren hospitalización respectivamente por cada región.

Región	NH (%)	H (%)
R1	95.91	4.085
R2	96.88	3.12
R3	96.88	3.12
R4	95.81	4.19
R5	94.76	5.24

Tabla 3. Porcentajes de los pacientes COVID-19 por cada región de México sin comorbilidades.

5.4.1 Porcentajes de Infectados con Comorbilidades en México

Las enfermedades crónicas degenerativas han sido un factor de riesgo en el desarrollo del COVID-19. Dependiendo del padecimiento, conlleva a presentar un mayor o menor riesgo en mostrar complicaciones a la enfermedad (e incluso, hasta la muerte). En la Figura 19 y Figura 20 se muestran los pacientes hospitalizados (H) y No-hospitalizados (NH) con diversas comorbilidades para las 5 regiones geográficas. Para todas las regiones destacan al menos tres enfermedades degenerativas más frecuentes entre los pacientes H y NH: EPOC,

Renal y otras enfermedades. Cabe señalar que en la R4 sus valores son mayores a diferencia de la R2 (ya que la densidad de población es mayor en la R4 que en la R2). Por ello, se puede observar los mayores casos de pacientes con estos padecimientos.

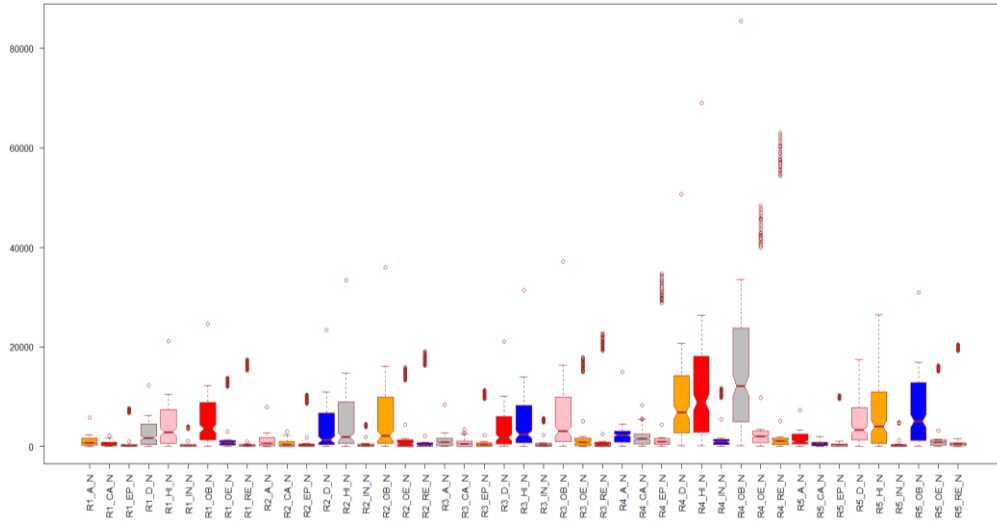


Figura 19. Análisis de las 9 comorbilidades registradas en las 5 Regiones de México en Paciente en Cuarentena.

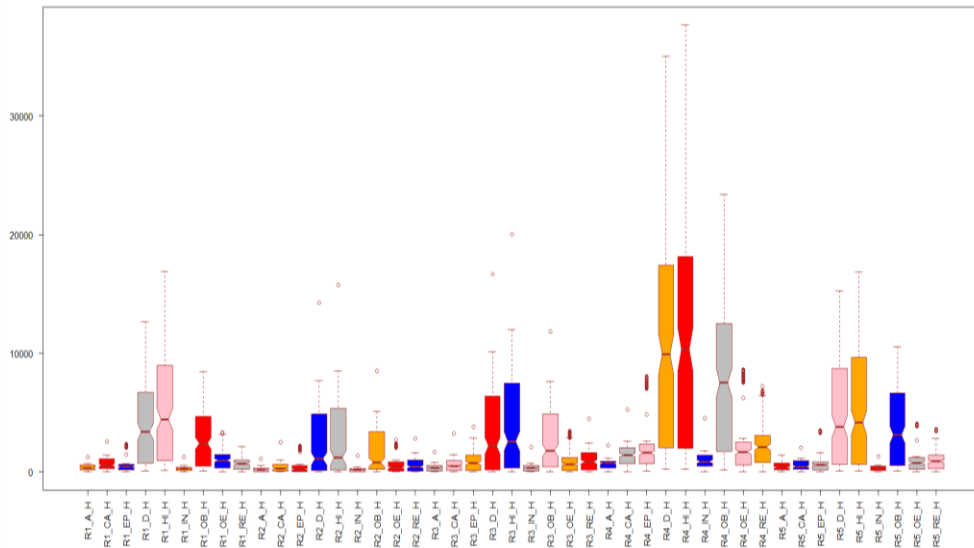


Figura 20. Análisis de las 9 comorbilidades registradas en las 5 Regiones de México en Paciente Hospitalizados.

De las 9 comorbilidades que han presentado los pacientes atendidos por COVID-19 se tiene la siguientes frecuencias relativas del total de pacientes por cada región analizada: Región={No-hospitalizado y Hospitalizado}.

1) El asma son 7,565 con (NH= 5,394 y H=2,171) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={0.5% y 2.7%}, R2={0.48% y 2.55%}, R3={0.548% y 5.094%}, R4={0.54% y 2.78%} y R5={0.5182% y 2.494%}.

2) Los problemas cardiovasculares son 17,268 con (NH= 13,817 y H=3,451) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={1.70% y 6.38%}, R2={1.55% y 4.613%}, R3={1.55% y 5.27%}, R4={1.14% y 4.75%} y R5={1.30% y 3.77%}.

3) Para el caso de hipertensión son 8,533 con (NH= 6,506 y H=2,027) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={0.629% y 2.9%}, R2={0.574% y 2.16%}, R3={0.678% y 3.093%}, R4={0.659% y 3.051%} y R5={0.584% y 2.418%}.

4) En el EPOC son 93,256 con (NH= 74,316 y H=18,940) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={7.6% y 28.55%}, R2={7.25% y 28.94%}, R3={6.684% y 27.675%}, R4={7.175% y 24.113%} y R5={8.52% y 29.475%}.

5) La Diabetes Mellitus son 9,800 con (NH= 8,569 y H=1,231) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={0.402% y 1.084%}, R2={0.664% y 0.491%}, R3={0.43% y 1.275%}, R4={0.630% y 1.415%} y R5={2.84% y 4.207%}.

6) Para el caso de la Inmunosupresión se son 32,651 con (NH= 30,787 y H=1,864) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={3.9362% y 3.4831%}, R2={3.13% y 2.79%}, R3={3.30% y 3.39%}, R4={2.429% y 1.93%} y R5={4.083% y 3.115%}.

7) La obesidad son 14,806 con (NH= 11,688 y H=3,118) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={1.34% y 6.05% }, R2={0.95% y 4.002%}, R3={1.209% y 5.21%}, R4={1.164% y 3.848% } y R5={1.025% y 4.165% }.

8) Los problemas Renales son 159,256 con (NH= 142,619 y H=16,637) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={16.93% y 25.539%}, R2={13.284% y 21.632%}, R3={13.347% y 23.6829%}, R4={13.065% y 20.6535% } y R5={17.0425% y 29.982% }.

9) Otros enfermedades son 133,867 con (NH= 112,131 y H=21,736) de ellos se ha tenido una frecuencia entre los pacientes por región de: R1={13.42% y 38.773%}, R2={11.0623% y 32.2563%}, R3={10.509% y 32.875%}, R4={10.029% y 25.914% } y R5={13.522% y 34.1954% }.

Cabe señalar que las principales comorbilidades que presentan los pacientes que han dado positivo a COVID-19 son los problemas renales en todas las regiones.

5.5 Análisis de Covarianzas

En esta subsección se presenta un análisis de la dependencia de las comorbilidades con las diferentes regiones. La red de correlaciones (Illowsky, B., y Dean, S., 2013) está construida mediante la siguiente agrupación jerárquica: en un primer nivel se encuentra el paciente, para un segundo nivel se divide en pacientes Hospitalizados y No-Hospitalizados, en un tercer nivel se encuentran las cinco regiones geográficas, en un cuarto nivel se encuentran los cinco intervalos de edades y en el quinto nivel se tiene los 9 tipos de comorbilidades (o padecimientos), ver Figura 17.

La medida de correlación de estas comorbilidades se describe mediante el coeficiente de correlación de *Pearson* (Spiegel, M., Schiller, J., and Srinivasan, R., 2013; Kassambara, A., 2020) cuyo valor se encuentra en el intervalo $\rho \in [-1,1]$ (Emerson, R., 2015). La matriz de correlación obtenida es una matriz cuadrada de 90×90 elementos. Para un mejor análisis de esta matriz se

dividió en dos submatrices (ver Figura 21 y Figura 22). Cada submatriz representa una alta correlación positiva y correlación negativa entre dichos padecimientos (Kassambara, A., 2020).

La Figura 21 describe la correlación negativa que presenta en algunas comorbilidades por regiones del país entre los pacientes NH y H. Es notable ver que la comorbilidad EPOC aparece con una correlación negativa de casi -1 en todas las regiones excepto la R5 en pacientes NH. También se tiene una correlación negativa muy alta en las regiones R4, R3 y R1 (en comorbilidades renal y otras enfermedades en pacientes NH). En los pacientes H sobresale una correlación muy alta negativa en EPOC, Inmunosupresión y Otras Enfermedades en todas las regiones. Más sin embargo, la R1 en pacientes NH con problemas cardíacos aparece con una correlación de -0.5 en todas las regiones excepto contra R4 con pacientes con Asma y en pacientes tipo H con correlación casi nula. La R3 tiene una correlación de -0.4 en pacientes con EPOC hospitalizados, contra casi todas las regiones y tipos de comorbilidad excepto contra la R4 (con pacientes con Asma en pacientes tipo H) con correlación casi nula. La R3 tiene una correlación de casi -0.3 en pacientes con Otras Enfermedades tipo hospitalizados, contra casi todas las regiones y tipos de comorbilidad excepto contra la R4 (con pacientes con Asma en pacientes tipo H) con correlación muy cercana a nula. La R5 tiene una correlación de -0.2 en pacientes con enfermedades Cardíacas en pacientes ambulatorios contra casi todas las regiones y tipos de comorbilidad excepto contra la R4 (con pacientes con Asma en pacientes tipo H) con correlación muy cercana a nula. La R2 tiene una correlación de -0.3 en pacientes con problemas Renales en hospitalizados, contra todas las regiones y tipos de comorbilidad excepto contra la R4 (con pacientes con Asma en pacientes tipo H) con correlación casi nula. La R3 tiene una correlación de -0.11 en pacientes con problemas renales hospitalizados contra todas las regiones y tipos de comorbilidad.

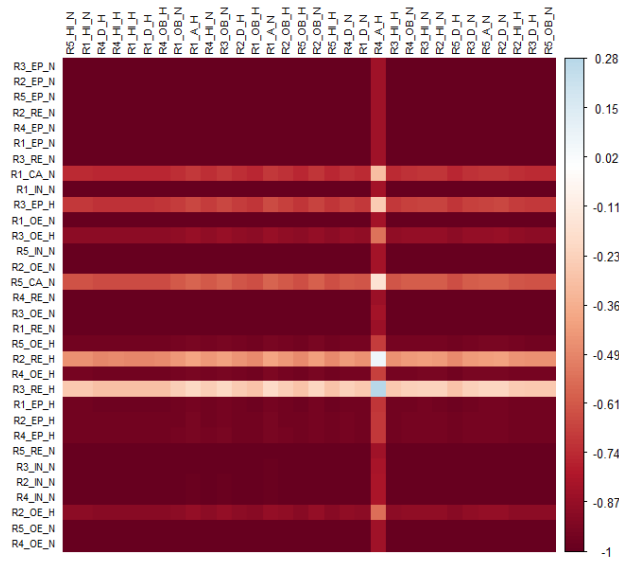


Figura 21. Submatriz con la correlación negativa de las comorbilidades en las regiones de México.

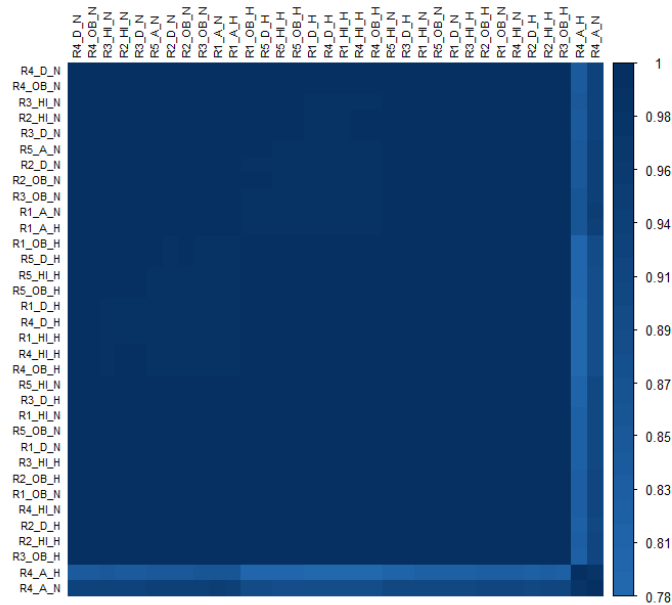


Figura 22. Submatriz con la correlación positiva en las 5 regiones de México.

La Figura 22 representa la submatriz con alta correlación positiva en los padecimientos y el crecimiento de los pacientes infectados. Donde existe una relación inversa entre los niveles jerárquicos de la red (por cada región de

México). Hay una correlación positiva muy alta de casi 0.9 tanto en pacientes ambulatorios como hospitalizados de todas las regiones que padecen Hipertensión, Diabetes, Obesidad y Asma.

Hay un cambio de correlación en la R4 de H que tiene 0.8 de correlación positiva en pacientes con Asma contra los pacientes de la misma R4 que padecen diabetes y obesidad (pero que son pacientes ambulatorios). También la R4 de H tiene 0.8 de correlación positiva en pacientes con hipertensión, contra pacientes con obesidad en ambulatorios de la R2. Al igual la R4 de H tiene 0.8 de correlación positiva en pacientes con diabetes, contra pacientes con obesidad en ambulatorios de la R3.

Continuando con la correlación positiva y la R4, ésta tiene 0.7 de correlación positiva en pacientes con asma tanto H y NH con respecto a R1. También la R4 de H tiene casi 0.7 de correlación positiva en pacientes con Obesidad y diabetes contra todas las comorbilidades de los pacientes H en R1. De igual manera, la R4 de H tiene casi 0.7 de correlación positiva en pacientes con diabetes, hipertensión y obesidad contra todas las comorbilidades de los pacientes H de la R5.

5.6 Resultados representados con mapa tipo SuperHeat

La Figura 23 presenta un mapa SuperHeat (Barter, R., and Yu, B., 2018). El mapa *SuperHeat* es una propuesta que surge en el año 2019 por el equipo de *Lumen-Learning* donde se unen elementos como el dendograma (mapa de jerarquías); que une el mapa de calor y la matriz de correlaciones. Este mapa es una herramienta de muy reciente ingreso al lenguaje R, cuyos códigos fueron donados para su uso y acceso gratuito. Este mapa es muy contundente si se usa adecuadamente. Resulta un poco complejo al principio, pero una vez que se entiende correctamente su codificación es una herramienta muy útil para presentar conjuntos de datos muy grandes y en varias dimensiones. Este mapa sintetiza toda la información analizada, por lo cual la autora la reconoce como

una herramienta primordial para presentar basta información en una sola gráfica. Para este caso, los mapas SuperHeat son el complemento final para el análisis que se ha venido describiendo en las secciones anteriores. Es el cierre perfecto de análisis día a día de las comorbilidades por regiones en México para pacientes infectados por COVID-19.

5.6.1 Resultados SuperHeat y algoritmo de Machine-Learning

La Figura 23 tiene en su eje izquierdo el dendrograma que indica en sus líneas más alejadas la representación de la jerarquía superior y hacia el centro del mapa las jerarquías de menor relación entre todos los elementos. En el mapa *SuperHeat* se pueden ver 3 grandes grupos resultantes de un algoritmo de *Aprendizaje de Máquinas* (de su traducción *Machine Learning*) llamado *K-means*. Es un tipo de algoritmo de *Agrupamiento* (cuya traducción es *clustering*) que se caracteriza por indicar qué tan semejantes son, es decir el coeficiente de similaridad. El coeficiente de similaridad se logra con una distancia llamada *Distancia Euclideana* (Barter, R., and Yu, B., 2018).

En la Figura 23 el eje derecho indica el número de día con pandemia (inicio el 12 de Abril 2019 en México hasta el 31 de Octubre del 2020) y su densidad de infectados en este caso son 203 días de análisis. La escala de colores al lado derecho del mapa indica la densidad de infectados que van desde 1000 en azul hasta 200 en rosa correspondiente a los casos diarios. El día con menor número de contagios en todas las regiones se originó el día 178.

Como se puede observar a grandes rasgos el SuperHeat tiene tres grandes *agrupamientos* ó *clusterings*. La más alta densidad de contagios se dio en los primeros días de pandemia (desde el día 1 hasta el día 150), prácticamente en todas las regiones y en todas las comorbilidades los pacientes (N y H) se contagiaron (hubo cerca de 1000 casos diarios de contagios, en color azul).

Por orden descendente en el SuperHeat de la Figura 23 se tiene:

-El primer *agrupamiento* resultante del algoritmo *k-means* selecciona a los días que tienen el mayor número de transiciones de contagios que van desde la escala mayor de contagios (1000) a la menor (200). En este primer grupo se encuentran los contagios alrededor de los días 151 a 174 de pandemia en México.

-En el segundo *agrupamiento* se tiene el menor número de contagios y se encuentran entre un rango de 400 a menos de 200 contagios diarios en todas las regiones y por el tipo de comorbilidad de los pacientes (se encuentra entre los días 175 y 203 de pandemia).

-El tercer *agrupamiento* es el que concentra la mayor densidad de contagios y va desde el día 1 hasta el día 150 de pandemia.

En el primer grupo, se hace el análisis de casos diarios que van del día 151 al día 174 y se describe a continuación: la R1 en pacientes H con Asma y Diabetes tiene una prolongación de promedio de densidad de pacientes infectados de casi 1000 casos diarios. La R2 en pacientes H con Diabetes, Hipertensión, Inmunosupresión tiene una prolongación de promedio de densidad de pacientes infectados de casi 1000 casos diarios. R3 en pacientes H con Diabetes tiene una prolongación de promedio de densidad de pacientes infectados de casi 1000 casos diarios. Seguidos de R1 en pacientes H con Hipertensión e Inmunosupresión tiene una prolongación de promedio de densidad de pacientes infectados por mayor tiempo de casi 800 casos diarios. R2 en pacientes H con enfermedades Cardiovasculares, Obesidad e Inmunosupresión tiene una prolongación de promedio de densidad de pacientes infectados de casi 800 casos diarios. R3 en pacientes H con Hipertensión tiene una prolongación de promedio de densidad de pacientes infectados de casi 800 casos diarios. R5 en pacientes H con Hipertensión y Asma tiene una prolongación de promedio de densidad de pacientes infectados de casi 800 casos diarios. Le sigue, R1 en pacientes H con Obesidad tiene una prolongación de promedio de densidad de pacientes infectados de casi 600 casos diarios que van del día 151 al día 174.

R3 en pacientes H con Asma, Obesidad, Cardiovascular hospitalizados tiene una prolongación de promedio de densidad de pacientes infectados de casi 600 casos diarios. R4 en pacientes H con Diabetes tiene una prolongación de promedio de densidad de pacientes infectados de casi 600 casos diarios. R5 en pacientes H con Diabetes, enfermedades Cardiovasculares, Inmunosupresión y Obesidad tiene una prolongación de promedio de densidad de pacientes infectados de casi 600 casos diarios. Característicamente, el resto de las comorbilidades con densidad de infectados alrededor de 200 casos diarios se encuentran los pacientes Ambulatorios.

En el segundo grupo, se tienen 4 casos que tiene una prolongación de promedio de densidad de pacientes NH infectados por casi todos los 28 días alternando entre 800 y 600 casos diarios: R1 en pacientes con EPOC y Inmunosupresión; y R2 en pacientes H con Inmunosupresión. Hay sólo 3 casos que tiene una prolongación de promedio de densidad de pacientes infectados por casi todos los 28 días de casi 600 casos diarios: R1 en pacientes H con Inmunosupresión, R5 en H con Inmunosupresión y R5 en pacientes NH con Inmunosupresión. El resto de las regiones y comorbilidades entre los pacientes hospitalizados y ambulatorios se encuentran alternando con una densidad de infectados alrededor de 400 y menos 200 casos diarios.

En el tercer grupo, en todas las regiones tanto en pacientes NH y H con tienen su mayor densidad de contagios alrededor de 1000 diarios en los días 1 a 20 aproximadamente de pandemia. Para pacientes No Hospitalizados, la R1 en pacientes con EPOC, Inmunosupresión, Renal tienen la prolongación de promedio de densidad de pacientes infectados por mayor tiempo que van desde 1000 a 800 diarios de casi 150 días de pandemia. La R2 en pacientes con enfermedades Cardiovasculares, EPOC, Renal, Otras Enfermedades e Inmunosupresión con NH tienen la prolongación de promedio de densidad de pacientes infectados por mayor tiempo que van desde 1000 a 800 diarios de casi 150 días de pandemia. La R3 en pacientes con enfermedades Cardiovasculares,

EPOC, Renal, Inmunosupresión con NH tienen la prolongación de promedio de densidad de pacientes infectados por mayor tiempo que van desde 1000 a 800 diarios de casi 150 días de pandemia. La R5 en pacientes con enfermedades Cardiovasculares, EPOC, Inmunosupresión y Renal con NH tienen la prolongación de promedio de densidad de pacientes infectados por mayor tiempo que van desde 1000 a 800 diarios de casi 150 días de pandemia. Para los pacientes Hospitalizados, la R2 es la única región que tiene la prolongación de promedio de densidad de pacientes infectados por mayor tiempo que van desde 1000 a 800 diarios de casi 150 días de pandemia, en comorbilidades como Asma, enfermedades Cardiovasculares, EPOC e Inmunosupresión.

5.7 Análisis de Probabilidad de Comorbilidades en México

Para el uso correcto de la Red Bayesiana es fundamental la definición correcta de las entradas. Para ello es imprescindible mencionar que los eventos deben de ser independientes. Situación que con estos datos no se tienen. En el nivel 3 de la HSP se describe este panorama, es decir, deben de convertirse los subconjuntos de las 9 comorbilidades a eventos independientes. Por lo tanto se deben quitar las intersecciones con las otras comorbilidades (para mayor detalle ver sección 4.3).

Una vez que se tienen las entradas como la Red Bayesiana los requiere, entonces los cálculos de la probabilidad con el uso de la biblioteca de *bnlearn* en R por cada comorbilidad se muestran en la Figura 24 y Figura 25.

La Figura 24 muestra qué probabilidad hay de ser un paciente Hospitalizado por COVID-19 en la Región 1 (R1) y tener EPOC es del 24%, Diabetes 1%, Obesidad 8%, Asma 4%, Hipertensión 4%, Inmunosupresión 4%, Cardiovascular 8%, Renal 23% y Otras enfermedades 24%. La probabilidad de ser un paciente Hospitalizado por COVID-19 en la Región 2 (R2) y tener EPOC es del 25%, Diabetes 0.83%, Obesidad 6.4%, Asma 4.1%, Hipertensión 4%,

Inmunosupresión 4.1%, Cardiovascular 7.3%, Renal 23.3% y Otras enfermedades 25%.

La probabilidad de ser un paciente Hospitalizado por COVID-19 en la Región 3 (R3) y tener EPOC es del 24%, Diabetes 1%, Obesidad 6%, Asma 6%, Hipertensión 4%, Inmunosupresión 6%, Cardiovascular 7%, Renal 22% y Otras enfermedades 24%.

La probabilidad de ser un paciente Hospitalizado por COVID-19 en la Región 4 (R4) y tener EPOC es del 24%, Diabetes 2.4%, Obesidad 6.3%, Asma 4.71%, Hipertensión 5.14%, Inmunosupresión 4.4%, Cardiovascular 7%, Renal 22% y Otras enfermedades 24%. La probabilidad de ser un paciente Hospitalizado por COVID-19 en la Región 5 (R5) y tener EPOC es del 24%, Diabetes 5.9%, Obesidad 5.9%, Asma 3.6%, Hipertensión 2.43%, Inmunosupresión 3.6%, Cardiovascular 5.2%, Renal 24% y Otras enfermedades 24.3%.

Un panorama poco diferente se observa en la Figura 25 ya que muestra que probabilidad de ser un paciente en cuarentena por COVID-19 en la Región 1 (R1) y tener EPOC es del 20%, Diabetes 10%, Obesidad 4.3%, Asma 2.9%, Hipertensión 1.9%, Inmunosupresión 2.9%, Cardiovascular 9%, Renal 24% y Otras enfermedades 25%. La probabilidad de ser un paciente NH (Ambulatorio) por COVID-19 en la Región 2 (R2) y tener EPOC es del 21%, Diabetes 8.9%, Obesidad 4%, Asma 1.7%, Hipertensión 1.8%, Inmunosupresión 1.9%, Cardiovascular 4%, Renal 25.7% y Otras enfermedades 31%. La probabilidad de ser un paciente No Hospitalizado por COVID-19 en la Región 3 (R3) y tener EPOC es del 20%, Diabetes 10%, Obesidad 4.7%, Asma 1.9%, Hipertensión 2.5%, Inmunosupresión 1.9%, Cardiovascular 5%, Renal 25% y Otras enfermedades 29%. La probabilidad de ser un paciente en Cuarentena por COVID-19 en la Región 4 (R4) y tener EPOC es del 19%, Diabetes 9%, Obesidad 4.7%, Asma 1.9%, Hipertensión 2.51%, Inmunosupresión 1.9%, Cardiovascular 3%, Renal 25% y Otras enfermedades 33%.

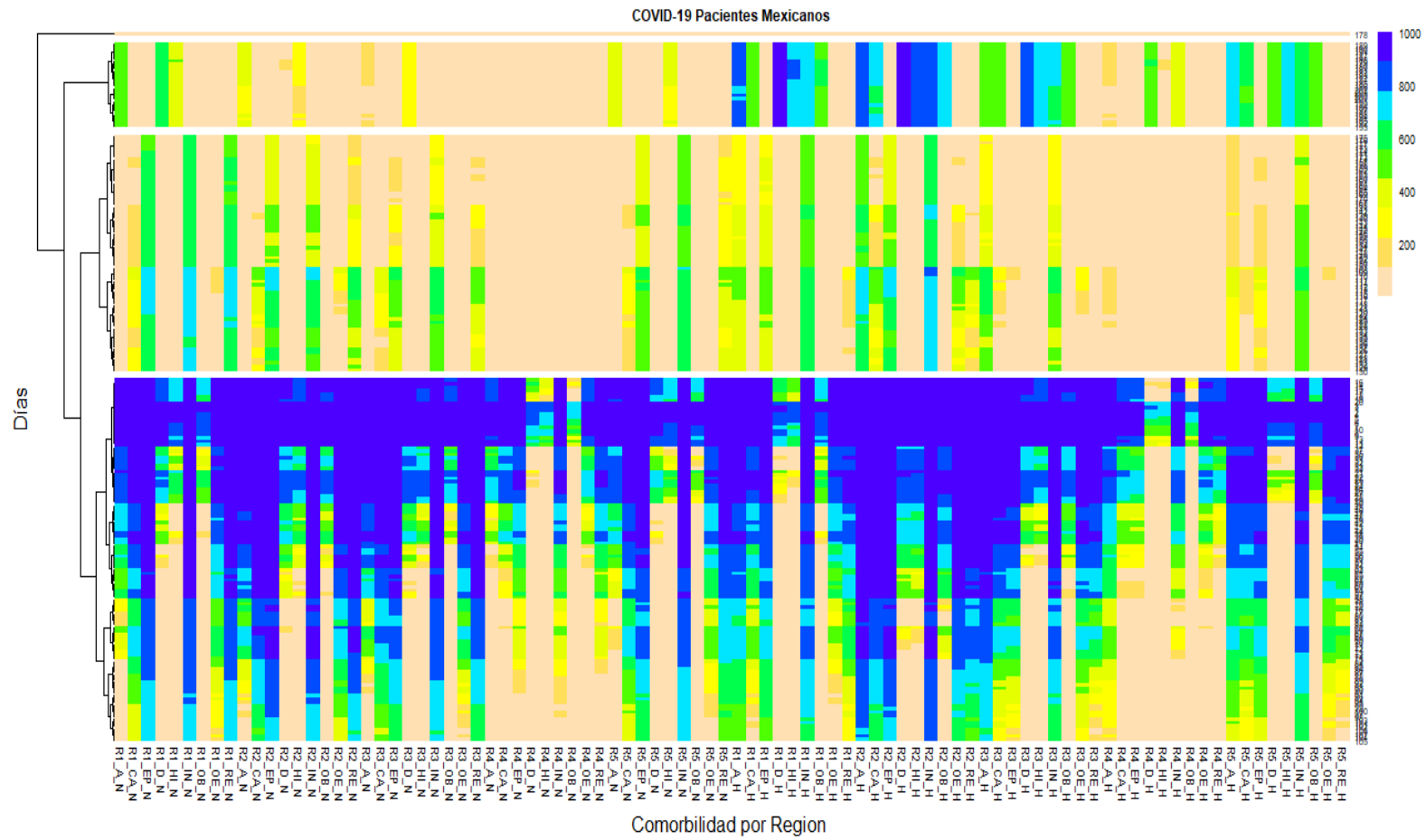


Figura 23. Mapa de de densidad de pacientes correlacionado con comorbilidades y regiones en México.

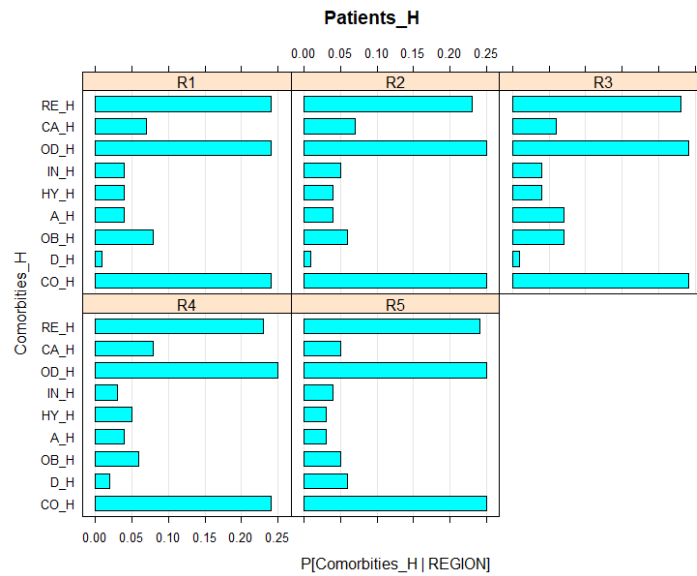


Figura 24. Salida de la red para los pacientes Hospitalizados y sus coeficientes de probabilidades para las 5 regiones de México en 203 días de registro con COVID-19.

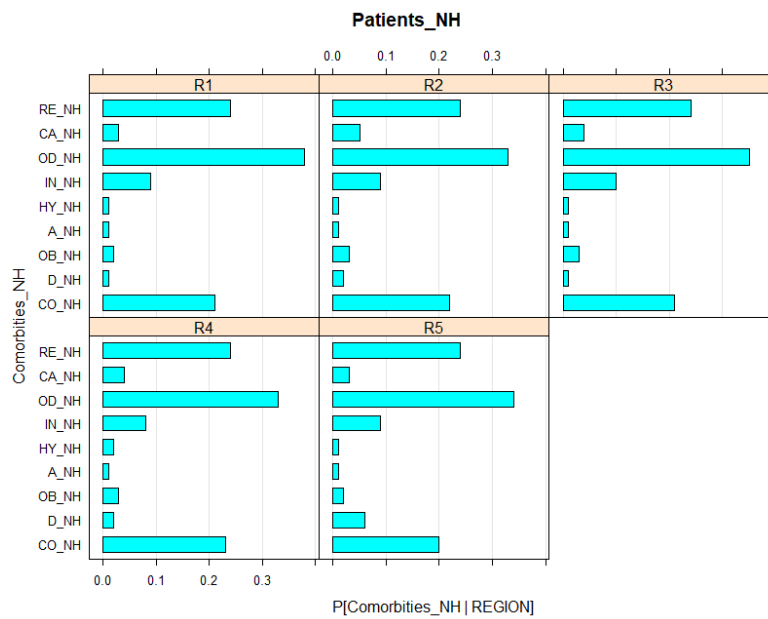


Figura 25. Salida de la red para los pacientes No-Hospitalizados y sus coeficientes de probabilidad para las 5 regiones de México en 203 días de registro con COVID-19.

La probabilidad de ser un paciente en Cuarentena por COVID-19 en la Región 5 (R5) y tener EPOC es del 19%, Diabetes 10%, Obesidad 3.9%, Asma 1.7%, Hipertensión 1.59%, Inmunosupresión 1.6%, Cardiovascular 3.9%, Renal 25.3% y Otras enfermedades 33%.

5.8 Resultados Índice Marginal

Para evaluar la situación de marginación en cada región en México se utiliza el IM. El IM puede tomar valores negativos o positivos (mediante la Técnica de Estratificación Óptima por Dalenius y Hodges). Y la clasificación del GM es en cinco categorías de *Grados de Marginación*: Muy Alto, Bajo, Medio, Alto y Muy Alto. El análisis de entidad federativa será: {IM y GM}.

Para la R1 tiene a Baja California Norte con {-1.1 y Muy Bajo}, Baja California Sur {-0.6 y Bajo}, Chihuahua {-0.6 y Bajo}, Sinaloa {-0.24 y Medio} y Sonora {-0.7 y Bajo}. La R1 en promedio tiene un IM de -0.648 por ende un GM Bajo con tendencia a *Muy Bajo*.

La R2 tiene a Coahuila con {-1.1 y Muy Bajo}, Durango {0.05 y Medio}, Nuevo León {-1.39 y Muy Bajo}, San Luis {0.58 y Alto} y Tamaulipas {-0.62 y Bajo}. La R2 en promedio tiene un IM de -0.496 por ende un GM es *Bajo*.

En la R3 la Ciudad de México tiene {-1.45 y Muy Bajo}, Estado de México {-0.57 y Bajo}, Guerrero {2.56 y Muy Alto}, Hidalgo {0.5 y Alto}, Morelos {-0.2 y Medio}, Puebla {0.69 y Alto} y Tlaxcala {-0.2 y Medio}. La R3 en promedio tiene un IM de 0.19 por ende GM es Medio con tendencia a *Alto*.

La R4 tiene a Aguascalientes con {-0.89 y Bajo}, Colima {-0.73 y Bajo}, Guanajuato {-0.07 y Medio}, Jalisco {-0.82 y Bajo}, Michoacán {0.5 y Alto}, Nayarit {0.31 y Medio}, Querétaro {-0.49 y Bajo} y Zacatecas {0.01 y Medio}. La R4 en promedio tiene un IM de -0.2725 por ende GM es Medio con tendencia a *Bajo*.

La R5 tiene a Campeche con {0.46 y Alto}, Chiapas {2.41 y Muy Alto}, Oaxaca {2.12 y Muy Alto}, Quintana Roo {-0.37 y Medio}, Tabasco {0.3 y Medio}, Veracruz {1.14 y Alto} y Yucatán {0.51 y Alto}. La R5 en promedio tiene un IM de 0.938571 por ende GM es *Alto*.

En orden ascendente por valores de IM y los GM las regiones quedan ordenadas como: R1, R2, R4, R3 y R5. Es decir, la R5 tiene un *Alto* grado de marginación y R1 tiene un grado de marginación *Muy Bajo*.

5.9 Resultados de Defunciones

Para la frecuencia relativa del número de defunciones a nivel nacional se toma primero al número de contagiados 1,089,852 y las 105,365 defunciones totales con lo que se tiene el 9.6678 % nacional.

Para la R1 se tienen 15,638 defunciones (NH=1,126 y H=14,512), R2 con 13,748 con (NH=2,204 y H=11,544), R3 con 15,884 (NH=1,827 y H=14,057), R4 con 42,826 (NH=3,727 y H=39,099) y R5 tiene 17,269 defunciones (NH=2,074 y H=15,195). Destacando R4 por tener el mayor número de defunciones tanto en H como para NH. Seguida muy de cerca por R5 también con el mayor número de defunciones (en NH y H en todo el país) y le sigue R3. Es importante mencionar que la R3 tiene la metrópoli de la Ciudad de México y abarca gran porcentaje de la población total de R3 (Molina-Torres, R., Nolasco-Jáuregui, O., Rodríguez-Torres, E., Quezada-Téllez. L., 2021).

5.10 Resultados de Hospitalizados en UCI

La frecuencia relativa del número de defunciones a nivel nacional se tienen 1,089,852 contagiados con 15,507 pacientes tipo H que entraron a UCI; lo cual es 1.4228% en el territorio nacional.

Para la R1 se tienen 8,297 pacientes hospitalizados de los cuales 2,361 necesitaron un respirador artificial. Por lo tanto tiene una frecuencia relativa de

28.4560%. De estos 2,361 fallecieron 1,194 en UCI en consecuencia el 50.57% del total.

En R2 se tienen 7,521 de los cuales 1,816 entraron a UCI entonces el 24.1457% de los hospitalizados. De estos 1,816 fallecieron 695 por lo tanto el 24.68%.

En R3 tiene 10,345 totales hospitalizados 2,545 entraron a UCI es decir el 24.60125%. De los 2,545 fallecieron 1,239 esto es el 48.68%.

La R4 tiene 33,263 de los cuales 5,882 entraron a UCI en consecuencia el 17.68331% de los hospitalizados. De estos 5,882 fallecieron 3,323 por lo tanto el 56.49%.

En R5 tiene 11,908 totales hospitalizados 2,903 entraron a UCI en pocas palabras el 24.3785%. De estos 2,903 fallecieron 1,507 es decir el 51.911%.

Por lo que el mayor riesgo de perder la vida siendo un paciente hospitalizado por COVID-19 en México es en R1, R5 y R4 (seguida muy cerca de R3). Y la R2 tiene el menor porcentaje de pacientes que fallecieron en UCI.

5.11 Conclusión

Este capítulo es la parte medular de la investigación ya que aquí se hace el análisis de los datos de salida de la HPS.

En primera instancia se tienen los resultados de la distribución del análisis de edades (para los pacientes contagiados por N y NH) que son 203 días de registro por las 5 regiones del país. De ello, se puede destacar que en edades de 20-40 años no han requerido hospitalización (pacientes ambulatorios) y los que han sido hospitalizados se encuentran en el intervalo de edades 40-60 años (ver sección 5.3.1). Todos estos resultados son constantes en todas las regiones.

Del número de contagiados se puntualiza que se tienen que de los 1,089,852 pacientes contagiados en México: 93.454% son ambulatorios (NH) y 6.54% (H) están hospitalizados. Del total de pacientes H en 12% es de la R1, R2 es el

11%, R3 el 15%, R4 el 47% y R5 el 17%. Del total de pacientes contagiados tipo NH el 10% tiene su lugar de residencia en R1, R2 el 14%, R3 el 17%, R4 el 47% y R5 el 12%. Subrayando que la R4 tiene el mayor número de pacientes contagiados tanto del tipo H como del tipo NH. R4 es seguida en orden descendente por el número de porcentaje de la R5 en pacientes tipo hospitalizados y R3 en ambulatorios.

El objetivo de esta investigación es observar e identificar la correlación que existe entre las comorbilidades como precedente a la llegada al virus y el comportamiento por cada región (tipo de etnias que habitan) y deducir si es la carga genética, hábitos y/o características de marginación de esa región que coadyuva a la dispersión del COVID-19 y sus mutaciones. Por lo anterior, se tienen varios análisis de las comorbilidades y se plantean de la siguiente manera:

1) Por comorbilidad se obtiene una frecuencia relativa y se grafica su referente en gráfica de cajas (ver Figura 19 y Figura 20) dividido por tipo de paciente NH y H respectivamente.

2) Análisis de covarianzas. Es muy interesante conocer el coeficiente de correlación de *pearson* entre todos los 90 elementos de la matriz. Es decir, por las 9 comorbilidades por 5 regiones por 2 tipos de pacientes. Como resultado, surgen dos subdivisiones entre aquellos que tienen una correlación negativa (Figura 21) y los de correlación positiva (Figura 22).

3) El análisis de los mapas de SuperHeat, es un aporte muy significativo para este análisis con herramientas de aprendizaje de máquinas en macrodatos (*big-data*). El mapa SuperHeat une todos los análisis anteriores en un gráfico, pero lo más importante es que este análisis es día por día. Mientras que en el análisis del punto 1) y 2) toma los totales de 203 días. Es fundamental saber leer este gráfico ya que de ello depende la correcta interpretación (lo cual ya se describe muy bien en la sección 5.6 de este documento).

4) Es el análisis y resultados de probabilidades de comorbilidades con la red Bayesiana. En la red Bayesiana se obtienen los análisis por morbilidad como eventos independientes a diferencia del punto 1), donde son eventos dependientes. Aquí la red permite hacer los cálculos para cualquier entrada ya que previamente fue diseñada con los requerimientos que esta investigación necesita. Por lo tanto, es un análisis regional para 203 días de pandemia dividido por paciente NH (Figura 23) y H (Figura 24).

5) En este análisis se culmina con el IM. El IM es fundamental para cuantificar la vulnerabilidad de la población debido a la tendencia del grado de marginación en la región del país que habitan. Es un factor trascendental para la dispersión del virus. Es la necesidad económica la que no permite a la población tener cuarentena en casa, las carencias de agua, números de habitantes en un hogar, número de ocupantes por dormitorio, nivel de educación, etc. Son los factores anteriores que fomentan a no cumplir con las medidas de sanidad para enfrentar el COVID-19.

Es trascendental resaltar la aportación de cada análisis (punto 1 al punto 5 de este documento) ya que no se desea perder ninguna esquina de referencia. Con ello se buscan la coincidencia de los resultados de manera cuantitativa y meta-análisis.

CAPÍTULO 6

CONCLUSIONES

La **hipótesis principal** que motivó a la realización de esta investigación es la vulnerabilidad y la predisposición en la salud de la población mexicana. Se prevé que son potencialmente la causa que México se encuentre en la lista de los 10 países con mayor número de contagios y es uno de los 5 países con mayor número de muertes por COVID-19 en todo el mundo.

Haciendo uso de la matriz de operacionalización de variables se plantea por cada objetivo secundario el instrumento a codificar y utilizar. Se garantiza que todos ellos se lograron con la HSP. En el primer de los **objetivos secundarios** se pretendía *definir* las probabilidades de contagio por cada región del país. Por lo que se codificó en el nivel 3 de la HSP con la red bayesiana (ver sección 4.3). También se codificaron en R los gráficos de cajas ver sección 4.2 y sección 5.4.1. Se puede concluir que dependiente de la morbilidad es el riesgo de contagio y adjunto a ello la edad del paciente son factores fundamentales para indagar cómo va a ser la lucha contra el virus, candidatura a ser hospitalizado, necesidad de un respirador artificial e incluso perder la vida ante el contagio. No es la misma relación entre un paciente adulto con padecimientos precedentes al virus a un joven sano.

En el segundo de los **objetivos secundarios** se pretendía *caracterizar* las probabilidades de contagio por cada región del país. Se utilizaron métodos de covarianzas, los resultados se analizan en la sección 5.5. Se puede concluir que dada la exposición, comorbilidades, edades, acceso al sector salud, nivel económico en la sociedad mexicana, entre otros. La alta probabilidad de contagio al virus del COVID-19 por las complicaciones que puede traer a los pacientes con un precedente de comorbilidad (Valente-Acosta, B., Hoyo-Ulloa, I., Espinosa-Aguilar, L., y et al., 2020).

En el tercer de los **objetivos secundarios** se pretendía *relacionar* las probabilidades de contagio por cada región del país. Se codificaron correlaciones entre los coeficientes de probabilidad día a día, ver sección 5.6. Se puede concluir que el primer *agrupamiento* resultante del algoritmo *k-means* en el mapa SuperHeat selecciona a los días que tienen el mayor número de transiciones de contagios que van desde la escala mayor de contagios (1000) a la menor (200). En este primer grupo se encuentran los contagios alrededor de los días 151 a 174 de pandemia en México de los 203 analizados.

En el tercer de los **objetivos secundarios** se pretendía *conocer* las edades y probabilidades de contagio por cada región. Se utilizaron las curvas de distribución por los 203 días de registro y 3,654 archivos totales (ver sección 5.3). Se puede concluir que en edades de 20-40 años no han requerido hospitalización (pacientes ambulatorios) y los que han sido hospitalizados se encuentran en el intervalo de edades 40-60 años (ver sección 5.3.1). Todos estos resultados son constantes en todas las regiones.

Es muy importante puntualizar que **el objetivo principal** que pretendía *analizar* el riesgo de contagio por COVID-19 en la población con mayor vulnerabilidad de México. Se logró cuantificar y dar respuesta por medio de los 5 puntos analizados en el CAPUTULO 5. Se resume que la verificación de la vulnerabilidad de la población se hace por medio de las siguientes variables cuantitativas enfatizando las variaciones concomitantes por cada región del país. Como conclusiones se tiene que:

- 1) La mayor concentración de las diferentes etnias que son monolingües y los que hablan su lengua indígena y español se concentran en la región R3, R5 y una sección de la R4, ver Figura 1.
- 2) Según el análisis de marginación calculado en la ecuación 4.1 por componentes principales resulta la Tabla 2 y conjuntamente con los 8

indicadores socioeconómicos resulta que la R5 por poseer 5 entidades federativas con grados de marginación con una tendencia a *Muy Alto* tales como: Campeche (GM=Alto), Chiapas (GM=Muy Alto), Oaxaca (GM=Muy Alto), Veracruz (GM=Alto) y Yucatán (GM=Alto). Seguido en orden descendente por la R3 por tener 3 entidades federativas con grados de marginación de *Alto a Muy Alto*. Tales como Guerrero (GM=Muy Alto), Hidalgo (GM=Alto) y Puebla (GM=Alto).

Coincidentemente la densidad de etnias en el territorio nacional y el IM tienden a mostrar a R3 y R5 como las regiones con la población más vulnerable.

De las **interrogantes secundarias surgen** las siguientes cuestiones: *¿Se valúa alguna relación entre las edades de los pacientes, nivel económico y las zonas de mayor contagio?*. Para dar respuesta a la pregunta anterior es importante puntualizar que México tiene un bono y es que su población es relativamente joven, lo cual ha sido un factor muy favorable ante el COVID-19 y se demuestra con el gran porcentaje de pacientes ambulatorios casi 93.454% (NH) contra 6.54% (H) que están hospitalizados:

1) Se concluye que en R4 tiene una mediana de los promedios de edad de los pacientes NH es de 40.3153 años y los pacientes H es de 52.871. Para la R5 la edad de los pacientes NH es de 40.123 y para los pacientes H es de 53.3629. Edades mayores que en el resto de las regiones, casualmente R4 tiene el mayor número de pacientes contagiados tanto del tipo H como del tipo NH y se encuentra seguida muy de cerca de R5 en hospitalizados. Desafortunadamente, el análisis de defunciones correlaciona que los contagiados en R4 tienen un alto riesgo a perder la vida tanto NH como H seguidos de los H en R5.

2) Como consecuencia, el riesgo más alto de perder la vida se infiere con el análisis de los pacientes hospitalizados y la relación con la UCI; esto potencializa el uso de un respirador artificial y con ello los porcentajes tienen una tendencia de fallecer por más del 50% de sus hospitalizados en R1, R4 y R5.

De las **interrogantes secundarias** surgen las siguientes cuestiones: *¿Cuál es la probabilidad de riesgo de contagio en las 5 regiones de México?, ¿A qué se debe la variabilidad en las probabilidades de contagio a lo largo del territorio nacional? y ¿Por qué las comorbilidades de los pacientes son factor en las zonas de mayor contagio?.* Para dar respuesta:

1) Como resultado, es indispensable recordar que la inferencia fáctica se hace por medio del análisis de comorbilidades que ratifica que la preexistencia a ser un paciente hospitalizado por COVID-19 y características particulares en la sociedad mexicana genera cambios de un fenómeno como es la dispersión de la pandemia. Es debido a lo primero que México es vulnerable ante el virus condición necesaria para el segundo fenómeno (México encabeza las listas a nivel mundial en contagios y muertes). Destacan tres comorbilidades en todo el territorio nacional EPOC con EPOC son 93,256 con (NH= 74,316 H=18,940), los problemas renales son 159,256 con (NH= 142,619 H=16,637) y otras enfermedades con 133,867 con (NH= 112,131 H=21,736).

2) A manera de cierre se comenta que se destaca la EPOC R5 con 8.52% en pacientes NH seguido muy de cerca con R2 y R4 con alrededor del 7% ambas. En EPOC en pacientes H tiene R5 casi de 29.475%, seguido muy de cerca por R3 con 27.675%. En problemas renales, R5 con 17.04% en pacientes NH. En pacientes H tiene R5 casi de 29.982%, seguido muy de cerca por R3 con 23.6829% y R1 con 25.539%. Otras enfermedades aparece de nuevo R5 con el mayor porcentaje en pacientes NH con 13.522% y también en pacientes H con 34.1954%, R3 con 32.875% y R1 con 38.773%.

3) En consecuencia, los pacientes con mayor porcentaje de comorbilidad tales como EPOC, enfermedades renales y otras enfermedades son los pacientes hospitalizados por lo que se deduce que la comorbilidad los hace vulnerables al virus. El riesgo a perder la vida aumenta al ser hospitalizado por COVID-19 y necesitar un respirador artificial.

Como comentario final, la HSP es producto de esta investigación la cual fue herramienta fundamental para realizar este trabajo de investigación. Por la necesidad de difusión de la información correcta, apropiada y (evitar la *Infodemia*) surge la necesidad de engendrar una WAF. Con un carácter de orden científico y sólo se alimentará de fuentes oficiales (y fidedignas). Es una realidad que se retroalimentará de la información que los ciudadanos aporten ya que tendrá un servicio de tipo público. Pero sólo usuarios registrados y verificados por un humano podrán utilizar el servicio de *publicaciones*. Para el contenido malicioso se utilizan análisis de cadenas de caracteres por medios de *alarmas* que activan alarmas al administrador de la WAF.

La HSP es producto de esta investigación la cual fue herramienta fundamental para realizar este trabajo de investigación y además los siguientes documentos:

1. Nolasco-Jáuregui, O., Quezada-Téllez, L., Rodríguez-Torres, E., y et al. (2021). **COVID-19 Patients Analysis using Superheat Map and Bayesian Network to identify Comorbidities Correlations under Different Scenarios.** MedRxiv, doi: <https://doi.org/10.1101/2021.05.11.21257055>
2. Molina-Torres, R., Nolasco-Jáuregui, O., Rodríguez-Torres, E., y et al. (2021). **A comparative analysis of urban development, economic level, and COVID-19 cases in Mexico City.** Journal of Urban Management, doi: <https://doi.org/10.1016/j.jum.2021.06.007>
3. Nolasco-Jáuregui, O. (2021). **Pandemic Information Dissemination Web Application: A Manual Design for Everyone.** En revisión.

BIBLIOGRAFÍA

Adwere-Boamah, J., Delay, D., y Jones, O. (1982). **Fundamental points of view of teachers on education: An investigation of the Kerlinger and Kaya attitudinal scale.** Educational Research Quarterly.

Aguilar-Moreno, M. (2007). **Handbook to life in the Aztec world.** Oxford University Press, USA.

Álvarez, C., y Sierra, V. (1995). **Metodología de la investigación científica.** Santiago de Cuba: Universidad de Oriente.

Aragón-Nogales, M., Vargas-Almanza, I., y Miranda-Novales, R. (2020). **COVID-19 por SARS-CoV-2: La Nueva Emergencia de Salud.** Revista Mexicana de Pediatría, 86(6), pp. 213-218.

Bansal, M. (2020). **Cardiovascular disease and COVID-19.** Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(3), 247-250.

Barbachano, f. c. (1964). **El mestizaje en México.** Revista de Indias, 24, 95.

Barter, R., and Yu, B. (2018). **Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data.** Journal of Computational and Graphical Statistics, 27(4), pp. 910-922.

Bartolomé, M. A., Baines, S. G., Ribeiro, G. L., & et al. (1996). **Pluralismo cultural y redefinición del Estado en México.** Universidade de Brasília, Departamento de Antropologia.

Buchanan, M. (2020). **Managing the infodemic,** Doctoral dissertation, Nature Publishing Group.

Buendía-Eisman, L., Expósito-López, J., y Sánchez-Martín, M. (2012). **Investigación evaluativa de programas de formación profesional para el empleo en el ámbito local.** *Revista de Investigación Educativa*, 30(1), 161-179

Buentello-Malo, L., Peñaloza-Espinosa, R. I., Salamanca-Gómez, F., & Cerda-Flores, R. M. (2008). **Genetic admixture of eight Mexican indigenous populations: Based on five polymarker, HLA-DQA1, ABO, and RH loci.** *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 20(6), 647-650.

Canizalez-Román, A., Campos-Romero, A., Castro-Sánchez, J. A., López-Martínez, M. A., Andrade-Muñoz, F. J., Cruz-Zamudio, C. K., ... & Alcántar-Fernández, J. (2018). **Blood groups distribution and gene diversity of the ABO and Rh (D) loci in the Mexican population.** *BioMed research international*, 2018.

Cerda-Flores, R. M., Villalobos-Torres, M. C., Barrera-Saldaña, H. A., Cortés-Prieto, L. M., Barajas, L. O., Rivas, F., ... & Chakraborty, R. (2002). **Genetic admixture in three Mexican Mestizo populations based on D1S80 and HLA-DQA1 loci.** *American Journal of Human Biology*, 14(2), 257-263.

Chen, A., Altschuler, K., Zhan, S., and et al. (2021). **COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest.** *Elife*, 10, p. e63409.

Chen, J., Gao, K., Wang, R., and Wei, G. W. (2021). **Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies.** *Chemical science*, 12(20), pp. 6929-6948.

Chitra, U., and Musco, C. (2020). **Analyzing the impact of filter bubbles on social network polarization,** In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 115-123.

- Chou, W., Oh, A., and Klein, W. (2018). **Addressing health-related misinformation on social media**, *Jama*, *320*(23), 2417-2418.
- Cinelli, M., Quattrocioni, W., Galeazzi, A., and et al. (2020). **The COVID-19 social media infodemic**, *Scientific Reports*, *10*(1), pp. 1-10.
- Cowell, F., and Flachaire, E. (2015). **Statistical methods for distributional analysis**. In *Handbook of income distribution* (Vol. 2, pp. 359-465). Elsevier.
- Cuello-Garcia, C., Pérez-Gaxiola, G., and van Amelsvoort, L. (2020). **Social media can have an impact on how we manage and investigate the COVID-19 pandemic**, *Journal of clinical epidemiology*, *127*, p. 198.
- Cuevas-Romo, A., Hernández-Sampieri, R., Leal-Pérez, B., et al. (2016). **Enseñanza-aprendizaje de ciencia e investigación en educación básica en México**. *Revista electrónica de investigación educativa*, *18*(3), pp. 187-200.
- Dantés, O., Sesma, S., Becerril, V., and et al. (2011). **Sistema de salud de México**. *Salud pública de México*, *53*, pp. 220-232.
- Domínguez, L. P. I., & Fragosó, G. F. (2002). **Mongolian spot frequency in newborns**. *Revista Mexicana de Pediatría*, *69*(4), 144-145.
- Emerson, R. (2015). **Causation and Pearson's correlation coefficient**. *Journal of visual impairment & blindness*, *109*(3), pp. 242-244.
- Eysenbach, G. (2020). **How to fight an infodemic: the four pillars of infodemic management**, *Journal of medical Internet research*, *22*(6), p. e21820,
- Fang, L., Karakiulakis, G., and Roth, M. (2020). **Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?**. *The Lancet. Medicina Respiratoria*, *8*(4), p. e21.
- Fernández-Collado, C., Baptista-Lucio, P., y Hernández-Sampieri, R. (2014). **Metodología de la Investigación**. Editorial McGraw Hill.

Figueroa, A. (2000). **La exclusión social como una teoría de la distribución.** Exclusión social y reducción de la pobreza en América Latina y El Caribe, pp. 13-24.

Folch, D. (2013). **El galeón de Manila.** Universitat Pompeu Fabra, Shanghai, Instituto Cervantes.

Garza, G., y Schteingart, M. (2010). **Los grandes problemas de México. Desarrollo urbano y regional.** Tomo 2. El Colegio de México AC.

Gómez, M. (2006). **Introducción a la metodología de la investigación científica.** Editorial Brujas.

Gorodezky, C., Alaez, C., Vázquez-García, M. N., de la Rosa, G., Infante, E., Balladares, S., ... & Muñoz, L. (2001). **The genetic structure of Mexican Mestizos of different locations: tracking back their origins through MHC genes, blood group systems, and microsatellites.** Human immunology, 62(9), 979-991.

Gracia, M. A. y Horbath, J. E. (2019). **Condiciones de vida y discriminación a indígenas en Mérida, Yucatán, México.** Estudios sociológicos, 37(110), pp. 277-307.

Guadarrama-Olivera, R., Hualde-Alfaro, A., y López-Estrada, S. (2012). **Precariedad laboral y heterogeneidad ocupacional: una propuesta teórico-metodológica.** Revista mexicana de sociología, 74(2), pp. 213-243.

Hitlin, P., and Olmstead, K. (2019). **The science people see on social media,** Pew Research Center.

Holmes, S., and Huber, W. (2018). **Modern statistics for modern biology.** Cambridge University Press.

Holone, H. (2016). **The filter bubble and its effect on online personal health information,** Croatian medical journal, 57(3), p. 298.

Hui, D., Azhar, E., Madani, T., and et al. (2020). **The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health -The latest 2019 novel coronavirus outbreak in Wuhan, China.** *Journal Internacional de Enfermedades Infecciosas*, 91, pp. 264-266.

Illowsky, B., and Dean, S. (2013). **Testing the Significance of the Correlation Coefficient.** *Introductory Statistics.*

Islam, M., Sarkar, T., Khan, S., and et al. (2020). **COVID-19–related infodemic and its impact on public health: A global social media analysis,** *The American journal of tropical medicine and hygiene*, 103(4), p. 1621.

Jimenez-Sanchez, G. (2003). **Developing a platform for genomic medicine in Mexico.** *Science*, 300(5617), 295-296.

Jordan, R., Adab, P., and Cheng, K.(2020). **COVID-19: risk factors for severe disease and death.**

Jurkowitz, M., and Mitchell, A. (2020). **Americans who primarily get news through social media are least likely to follow COVID-19 coverage, most likely to report seeing made-up news,** *Pew Research Center.*

Kassir, R. (2020). **Risk of COVID19 for patients with obesity.** *Obesity Reviews*, 21(6).

Kim, L., Fast, S. M., and Markuzon, N. (2019). **Incorporating media data into a model of infectious disease transmission.** *PloS one*, 14(2), p. e0197646.

Knaul, F., Arreola-Ornelas, H., Méndez, O., Lafontaine, L. y Wong, R. (2010). **Prevalencia de gasto catastrófico y empobrecedor por motivos de salud en México. México, DF: Fundación Mexicana para la Salud.** Documento de Trabajo.

Ladrón de Guevara, L. (1981). **Metodología de la investigación científica: problemas del método en ciencias sociales.** Universidad Santo Tomás, Centro de Enseñanza Desescolarizada.

Lastra, M., González, C., Pérez, M., Guzmán, y et al. (2013). **Índice de vulnerabilidad ante el COVID-19 en México.**

Lau, A. Y., and Staccini, P. (2019). **Artificial intelligence in health: new opportunities, challenges, and practical implications,** Yearbook of medical informatics, 28(01), pp. 174-178.

Leyva-Flores, R., Infante-Xibille, C., Gutiérrez, J. P. y et al. (2013). **Inequidad persistente en salud y acceso a los servicios para los pueblos indígenas de México, 2006-2012.** Salud pública de México, 55, pp. 123-128.

Lizcano Fernández, F. (2005). **Composición étnica de las tres áreas culturales del continente americano al comienzo del siglo XXI.** Convergencia, 12(38), 185-232.

López de Lerma, G. (2015). **Análisis comparativo de metodologías para la enseñanza y el aprendizaje de la lengua latina.**

López-Yepes, J. (1990). **Fundamentos de información y documentación.** Revista Española de Documentación Científica, 13(2), p. 781.

Maria Guadalupe Ayala Uribe, D. (1970). **La mancha mongolica.** International Journal of Dermatology, 9(3), 180-185.

Mendoza, M., Poblete, B., and Castillo, C. (2010). **Twitter under crisis: Can we trust what we RT?.** In Proceedings of the first workshop on social media analytics, pp. 71-79.

Miller, D., and Salkind, N. (2002). **Handbook of research design and social measurement.** Sage.

Molina-Torres, R., Nolasco-Jáuregui, O., Rodríguez-Torres, E., and et al. (2021). **A comparative analysis of urban development, economic level, and COVID-19 cases in Mexico City.** Journal of Urban Management, doi: <https://doi.org/10.1016/j.jum.2021.06.007>

Monroy, R. (2017). **La occidentalización de los asentamientos indígenas en Morelos, México.** Revista Biodiversidad Neotropical, 7(1), pp. 22-29.

Muñoz-Razo, C. (1998). **Cómo elaborar y asesorar una investigación de tesis.** Pearson Educación.

Nagarajan, R., Scutari, M., and Lébre, S. (2013). **Bayesian networks in r.** Springer, Vol.122, pp. 125-127.

Navarrete, F. (2005). **El mestizaje y las culturas regionales.** Las Relaciones Interétnicas.

Neumann, G., and Kawaoka, Y. (2019). **Predicting the next influenza pandemics,** The Journal of infectious diseases, 219, pp. S14-S20.

Nolasco-Jáuregui, O., Quezada-Téllez, L., Rodríguez-Torres, E., and et al. (2021). **COVID-19 Patients Analysis using Superheat Map and Bayesian Network to identify Comorbidities Correlations under Different Scenarios.** MedRxiv, doi: <https://doi.org/10.1101/2021.05.11.21257055>

Parra, E. J., Kittles, R. A., & Shriver, M. D. (2004). **Implications of correlations between skin color and genetic ancestry for biomedical research.** Nature genetics, 36(11), S54-S60.

Ramírez-González, A. (2006). **Ecología: Métodos de muestreo y análisis de poblaciones y comunidades.** Pontificia Universidad Javeriana.

Rosas, F. M. (2005). **México Tenochtitlan.** Ediciones Era.

Schwartz-Marín, E., & Silva-Zolezzi, I. (2010). **“The Map of the Mexican’s Genome”**: overlapping national identity, and population genomics. *Identity in the information society*, 3(3), 489-514.

Scutari, M., and Denis, J. B. (2014). *Bayesian networks: with examples in R*. CRC press.

Shaman, J., Karspeck, A., Yang, W., and et al. (2013). **Real-time influenza forecasts during the 2012–2013 season**, *Nature communications*, 4(1), pp. 1-10.

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., ... & Jimenez-Sanchez, G. (2009). **Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico**. *Proceedings of the National Academy of Sciences*, 106(21), 8611-8616.

Smailhodzic, E., Hooijsma, W., Boonstra, A., and et al. (2016). **Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals**, *BMC health services research*, 16(1), pp. 1-14.

Spiegel, M., Schiller, J., and Srinivasan, R. (2013). **Probability and statistics (Vol.2)**. McGrawHill, New York.

Starbird, K., Maddock, J., Orand, M., and et al. (2014). **Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing**. *Conference Proceedings*.

Tamayo y Tamayo, M. (1977). **Metodología formal de la investigación científica (No. H65. T35 1983)**. **Redactar correctamente las conclusiones y recomendaciones de un informe de investigación. Elaborar y organizar adecuadamente la bibliografía de una investigación.**

Valdez-Velazquez, L. L., Mendoza-Carrera, F., Perez-Parra, S. A., Rodarte-Hurtado, K., Sandoval-Ramirez, L., Montoya-Fuentes, H., ... & Rivas, F.

(2011). **Renin gene haplotype diversity and linkage disequilibrium in two Mexican and one German population samples.** *Journal of the Renin-Angiotensin-Aldosterone System*, 12(3), 231-237.

Valente-Acosta, B., Hoyo-Ulloa, I., Espinosa-Aguilar, L., and et al. (2020). **COVID-19 severe pneumonia in Mexico City-First experience in a Mexican hospital.** *MedRxiv*.

Vargas, R., Cobos, A. y Saa, T. (2016). **El contexto sí importa. Violencia de género hacia las estudiantes en escuelas secundarias y preparatorias en Guanajuato.** *La Manzana de la discordia*, 8(2), pp. 17-30.

Viboud, C., and Vespignani, A. (2019). **The future of influenza forecasts,** *Proceedings of the National Academy of Sciences*, 116(8), pp. 2802-2804.

Wang, Y., McKee, M., Torbica, A., and et al. (2019). **Systematic literature review on the spread of health-related misinformation on social media,** *Social science & medicine*, 240, p. 112552.

Yang, K., Pierri, F., Hui, P. and et al. (2021). **The COVID-19 Infodemic: Twitter versus Facebook.** *Big Data & Society*, 8(1), p. 20539517211013861.

Zarocostas, J. (2020). **How to fight an infodemic,** *the lancet*, 395 (10225), p. 676.

Zhou, D., Zhu, S., Yu, K., and et al. (2008). **Learning multiple graphs for document recommendations.** In *Proceedings of the 17th international conference on World Wide Web*, pp. 141-150.

WEBGRAFÍA

DGE (2021). **Datos Abiertos Dirección General de Epidemiología**. Bases históricas COVID-19. Consulta 23 Abril 2021, de Gobierno de México. Sitio Web:

https://www.gob.mx/cms/uploads/attachment/file/604001/Datos_abiertos_historicos_2020.pdf

InDRE (2020). **Evaluaciones de pruebas COVID-19**. Consulta 22 de Enero 2021, de Instituto de Diagnóstico y Referencia Epidemiológicos Dr. Manuel Martínez Báez. Sitio Web: <https://www.gob.mx/salud/acciones-y-programas/instituto-de-diagnostico-y-referencia-epidemiologicos-indre>

INER (2021). **Información sobre el COVID-19**. Consulta 21 de Noviembre 2020, de Instituto Nacional de Enfermedades Respiratorias. Sitio Web: <https://www.gob.mx/salud/iner>

JetBrains (2021). **The Python IDE for Professional Developers**. Consulta 8 de Enero 2021, de JetBrains S.R.O.: <https://www.jetbrains.com/pycharm/>

JHU (2021). **COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University**. Consulta 12 de Enero 2021, de Johns Hopkins University. Sitio Web: <https://coronavirus.jhu.edu/map.html>

Kassambara, A. (2020). **Correlation Test Between Two Variables in R**. Consulta 15 Febrero 2021, de Statistical tools for high-throughput data analysis. (STHDA). Sitio Web: <http://www.sthda.com/english/wiki/wiki.php?title=correlation-analyses-in-r>

Linode (2020). **Use Public Key Authentication with SSH**. Consulta Enero 19 2021, de Linode, LLC. Sitio Web: <https://www.linode.com/docs/guides/use-public-key-authentication-with-ssh/#windows>

Michonneau, F., and Fournier, A. (2018). **Data visualization with ggplot2**. Consulta 22 Enero 2021, Data Carpentry Contributors. Sitio Web: <https://datacarpentry.org/R-ecology-lesson/04-visualization-ggplot2.html>

Molina-Torres, R., Nolasco-Jáuregui, O., Rodríguez-Torres, E., Quezada-Téllez, L. (2021). **A comparative analysis of urban development, economic level, and COVID-19 cases in Mexico City**. Journal of Urban Management.

Monnich, A., Ronacher, A., Lord, D., and et al., (2015). **The Pallets Projects (Flask)**. Consulta 9 de Enero 2021, de WSGI.org: <https://palletsprojects.com/p/flask/>

ONU (2021). **La enfermedad del Coronavirus, una emergencia mundial**. Consulta 24 Febrero 2021, de Organización de las Naciones Unidas. Sitio Web: <https://www.un.org/es/coronavirus>

OraliaNJ (2021). **COVID-19 Portafolios Análisis**. Consulta 15 Febrero 2021, de GitHub, Inc. Sitio Web: <https://github.com/OraliaNJ/COVID-19-Portfolios-Analysis>

Python Software Foundation (2021). **Python Web Server Gateway Interface v1.01**. Consulta 10 de Enero de 2021, de python.org: <https://www.python.org/dev/peps/pep-3333>

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., and et al. (2021). **Statistics and Research Coronavirus Pandemic (COVID-19)**. Consulta 12 de Noviembre 2020, de Our World in Data de Oxford Martin School and University of Oxford. Sitio Web: <https://ourworldindata.org/coronavirus>

Shcafer, C. (2021). **Development, Design, DIY and more [Blog]**. Consulta el 10 de Enero 2021, de CoreyMS: <https://coreyms.com/>

Stackoverflow. (2020). **How to store vector in dataframe in R**. Consulta 24 Enero, 2021, de Stack Exchange Inc. Sitio Web: <https://stackoverflow.com/questions/51307970/how-to-store-vector-in-dataframe-in-r>

Wilke, C. (2018). **Color relative to numeric value**. Consulta 25 Febrero 2021, de the R Graph Gallery. Sitio Web: <https://www.r-graph-gallery.com/294-basic-ridgeline-plot.html#color>

CONAPO (2016). **Índices de marginación**. Consulta 25 de Enero 2021, de Consejo Nacional de Población en México. Sitio Web: http://www.conapo.gob.mx/es/CONAPO/Indices_de_Marginacion

GoodFirms (2020). **What is a Web Framework?**. Consulta 7 de Enero 2021, de GoodFirms: <https://www.goodfirms.co/glossary/web-framework/>

Tatham, S. (2020). **Download PuTTY**. Consulta Enero 18 2021, de putty.org. Sitio Web: <https://www.putty.org/>

INEGI (2018). **Encuesta Intercensal 2015**. Consulta 31 Enero 2021, de Sistema Nacional de Información Estadística y Geográfica en México. Sitio Web: <https://www.inegi.org.mx/programas/intercensal/2015/>

Spector, P. (2014). **Using t-tests in R**. Consulta 12 Febrero 2021, de University of California, department of Statistics. Sitio Web: <https://statistics.berkeley.edu/computing/r-t-tests>

Linode (2021). **DNS Manger**. Consulta Febrero 14 2021, de Linode, LLC. Sitio Web: <https://www.linode.com/docs/guides/dns-manager/>

NMED (2010). **El genoma Mexicano**. Consulta 13 Marzo 2021, de Word.Press.com. Sitio Web: <https://nutriologamarthaedithdiaz.wordpress.com/about/>

Mitchel, B. (2020). **127.0.0.1 IP Address Explained**. Consulta Enero11 2021, de LifeWire. Sitio Web: <https://www.lifewire.com/network-computer-special-ip-address-818385>

OraliaNJ (2020). **COVID-19 México Análisis**. Consulta 10 Diciembre 2020, de GitHub Inc. Sitio Web: https://github.com/OraliaNJ/COVID-19_Mex_Analysis

ANEXO A

NUM	NOMBRE DE VARIABLE	DESCRIPCIÓN DE VARIABLE	FORMATO O FUENTE	DESCRIPCIÓN	
1	FECHA_ACTUALIZACION	La base de datos se alimenta diariamente, esta variable permite identificar la fecha de la última actualización.	AAAA-MM-DD	AAAA-MM-DD	
2	ID_REGISTRO	Número identificador del caso	TEXTO	TEXTO	
3	ORIGEN	La vigilancia centinela se realiza a través del sistema de unidades de salud monitoras de enfermedades respiratorias (USMER). Las USMER incluyen unidades médicas del primer, segundo o tercer nivel de atención y también participan como USMER las unidades de tercer nivel que por sus características contribuyen a ampliar el panorama de información epidemiológica, entre ellas las que cuentan con especialidad de neumología, infectología o pediatría.	CATÁLOGO: ORIGEN		
				CLAVE	DESCRIPCIÓN
				1	USMER
				2	FUERA DE USMER
99	NO ESPECIFICADO				

4	SECTOR	Identifica el tipo de institución del Sistema Nacional de Salud que brindó la atención.	CATÁLOGO: SECTOR	<table border="1"> <thead> <tr> <th>CLAVE</th> <th>DESCRIPCIÓN</th> </tr> </thead> <tbody> <tr><td>1</td><td>CRUZ ROJA</td></tr> <tr><td>2</td><td>DIF</td></tr> <tr><td>3</td><td>ESTATAL</td></tr> <tr><td>4</td><td>IMSS</td></tr> <tr><td>5</td><td>IMSS-BIENESTAR</td></tr> <tr><td>6</td><td>ISSSTE</td></tr> <tr><td>7</td><td>MUNICIPAL</td></tr> <tr><td>8</td><td>PEMEX</td></tr> <tr><td>9</td><td>PRIVADA</td></tr> <tr><td>10</td><td>SEDENA</td></tr> <tr><td>11</td><td>SEMAR</td></tr> <tr><td>12</td><td>SSA</td></tr> <tr><td>13</td><td>UNIVERSITARIO</td></tr> <tr><td>99</td><td>NO ESPECIFICADO</td></tr> </tbody> </table>	CLAVE	DESCRIPCIÓN	1	CRUZ ROJA	2	DIF	3	ESTATAL	4	IMSS	5	IMSS-BIENESTAR	6	ISSSTE	7	MUNICIPAL	8	PEMEX	9	PRIVADA	10	SEDENA	11	SEMAR	12	SSA	13	UNIVERSITARIO	99	NO ESPECIFICADO
				CLAVE	DESCRIPCIÓN																													
				1	CRUZ ROJA																													
				2	DIF																													
				3	ESTATAL																													
				4	IMSS																													
				5	IMSS-BIENESTAR																													
				6	ISSSTE																													
				7	MUNICIPAL																													
				8	PEMEX																													
				9	PRIVADA																													
				10	SEDENA																													
				11	SEMAR																													
				12	SSA																													
13	UNIVERSITARIO																																	
99	NO ESPECIFICADO																																	
5	ENTIDAD_UM	Identifica la entidad donde se ubica la unidad médica que brindó la atención.	CATALÓGO: ENTIDADES	<table border="1"> <thead> <tr> <th>CLAVE_ENTIDAD</th> <th>ENTIDAD_FEDERATIVA</th> </tr> </thead> <tbody> <tr><td>01</td><td>AGUASCALIENTES</td></tr> <tr><td>02</td><td>BAJA CALIFORNIA</td></tr> <tr><td>..</td><td>...</td></tr> <tr><td>32</td><td>ZACATECAS</td></tr> </tbody> </table>	CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA	01	AGUASCALIENTES	02	BAJA CALIFORNIA	32	ZACATECAS																				
				CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA																													
				01	AGUASCALIENTES																													
				02	BAJA CALIFORNIA																													
..	...																																	
32	ZACATECAS																																	
6	SEXO	Identifica al sexo del paciente.	CATÁLOGO: SEXO	<table border="1"> <thead> <tr> <th>CLAVE</th> <th>DESCRIPCIÓN</th> </tr> </thead> <tbody> <tr><td>1</td><td>MUJER</td></tr> <tr><td>2</td><td>HOMBRE</td></tr> <tr><td>99</td><td>NO ESPECIFICADO</td></tr> </tbody> </table>	CLAVE	DESCRIPCIÓN	1	MUJER	2	HOMBRE	99	NO ESPECIFICADO																						
				CLAVE	DESCRIPCIÓN																													
				1	MUJER																													
2	HOMBRE																																	
99	NO ESPECIFICADO																																	
7	ENTIDAD_NAC	Identifica la entidad de nacimiento del paciente.	CATALÓGO: ENTIDADES	<table border="1"> <thead> <tr> <th>CLAVE_ENTIDAD</th> <th>ENTIDAD_FEDERATIVA</th> </tr> </thead> <tbody> <tr><td>01</td><td>AGUASCALIENTES</td></tr> <tr><td>02</td><td>BAJA CALIFORNIA</td></tr> <tr><td>..</td><td>...</td></tr> <tr><td>32</td><td>ZACATECAS</td></tr> </tbody> </table>	CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA	01	AGUASCALIENTES	02	BAJA CALIFORNIA	32	ZACATECAS																				
				CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA																													
				01	AGUASCALIENTES																													
				02	BAJA CALIFORNIA																													
..	...																																	
32	ZACATECAS																																	
8	ENTIDAD_RES	Identifica la entidad de residencia del paciente.	CATALÓGO: ENTIDADES	<table border="1"> <thead> <tr> <th>CLAVE_ENTIDAD</th> <th>ENTIDAD_FEDERATIVA</th> </tr> </thead> <tbody> <tr><td>01</td><td>AGUASCALIENTES</td></tr> <tr><td>02</td><td>BAJA CALIFORNIA</td></tr> <tr><td>..</td><td>...</td></tr> <tr><td>32</td><td>ZACATECAS</td></tr> </tbody> </table>	CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA	01	AGUASCALIENTES	02	BAJA CALIFORNIA	32	ZACATECAS																				
				CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA																													
				01	AGUASCALIENTES																													
				02	BAJA CALIFORNIA																													
..	...																																	
32	ZACATECAS																																	
9	MUNICIPIO_RES	Identifica el municipio de residencia del paciente.	CATALÓGO: MUNICIPIOS	<table border="1"> <thead> <tr> <th>CLAVE_MUNICIPIO</th> <th>MUNICIPIO</th> <th>CLAVE_ENTIDAD</th> </tr> </thead> <tbody> <tr><td>001</td><td>AGUASCALIENTES</td><td>01</td></tr> <tr><td>002</td><td>ASIENTOS</td><td>01</td></tr> <tr><td>...</td><td>..</td><td>..</td></tr> <tr><td>058</td><td>STA MARIA</td><td>32</td></tr> </tbody> </table>	CLAVE_MUNICIPIO	MUNICIPIO	CLAVE_ENTIDAD	001	AGUASCALIENTES	01	002	ASIENTOS	01	058	STA MARIA	32															
				CLAVE_MUNICIPIO	MUNICIPIO	CLAVE_ENTIDAD																												
				001	AGUASCALIENTES	01																												
				002	ASIENTOS	01																												
...																																
058	STA MARIA	32																																
10	TIPO_PACIENTE	Identifica el tipo de atención que recibió el paciente en la unidad. Se denomina como ambulatorio si regresó a	CATÁLOGO: TIPO_PACIENTE	<table border="1"> <thead> <tr> <th>CLAVE</th> <th>DESCRIPCIÓN</th> </tr> </thead> <tbody> <tr><td>1</td><td>AMBULATORIO</td></tr> <tr><td>2</td><td>HOSPITALIZADO</td></tr> </tbody> </table>	CLAVE	DESCRIPCIÓN	1	AMBULATORIO	2	HOSPITALIZADO																								
				CLAVE	DESCRIPCIÓN																													
1	AMBULATORIO																																	
2	HOSPITALIZADO																																	

		su casa o se denomina como hospitalizado si fue ingresado a hospitalización.		99	NO ESPECIFICADO
11	FECHA_INGRESO	Identifica la fecha de ingreso del paciente a la unidad de atención.	AAAA-MM-DD	AAAA-MM-DD	
12	FECHA_SINTOMAS	Identifica la fecha en que inició la sintomatología del paciente.	AAAA-MM-DD	AAAA-MM-DD	
13	FECHA_DEF	Identifica la fecha en que el paciente falleció.	AAAA-MM-DD	AAAA-MM-DD	
14	INTUBADO	Identifica si el paciente requirió de intubación.	CATÁLOGO: SI NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
15	NEUMONIA	Identifica si al paciente se le diagnosticó con neumonía.	CATÁLOGO: SI NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
16	EDAD	Identifica la edad del paciente.	NÚMERICA EN AÑOS	EDAD	
17	NACIONALIDAD	Identifica si el paciente es mexicano o extranjero.	CATÁLOGO: NACIONALIDAD	CLAVE	DESCRIPCIÓN
				1	MEXICANA
				2	EXTRANJERA
				99	NO ESPECIFICADO

18	EMBARAZO	Identifica si la paciente está embarazada.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
19	HABLA LENGUA INDIG	Identifica si el paciente habla lengua indígena.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
20	DIABETES	Identifica si el paciente tiene un diagnóstico de diabetes.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
21	EPOC	Identifica si el paciente tiene un diagnóstico de EPOC.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
22	ASMA	Identifica si el paciente tiene un diagnóstico de asma.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
23	INMUSUPR	Identifica si el paciente presenta inmunosupresión.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
24	HIPERTENSION	Identifica si el paciente tiene un diagnóstico de hipertensión.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO

25	OTRAS_COM	Identifica si el paciente tiene diagnóstico de otras enfermedades.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
26	CARDIOVASCULAR	Identifica si el paciente tiene un diagnóstico de enfermedades cardiovasculares.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
27	OBESIDAD	Identifica si el paciente tiene diagnóstico de obesidad.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
28	RENAL_CRONICA	Identifica si el paciente tiene diagnóstico de insuficiencia renal crónica.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
29	TABAQUISMO	Identifica si el paciente tiene hábito de tabaquismo.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
30	OTRO_CASO	Identifica si el paciente tuvo contacto con algún otro caso diagnosticado con SARS CoV-2	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
31	RESULTADO	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).	CATÁLOGO: RESULTADO	CLAVE	DESCRIPCIÓN
				1	Positivo SARS-CoV-2
				2	No positivo SARS-CoV-2
				3	Resultado pendiente

32	MIGRANTE	Identifica si el paciente es una persona migrante.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO
33	PAIS_NACIONALIDAD	Identifica la nacionalidad del paciente.	TEXTO, 99= SE IGNORA	TEXTO, 99= SE IGNORA	
34	PAIS_ORIGEN	Identifica el país del que partió el paciente rumbo a México.	TEXTO, 97= NO APLICA	TEXTO, 99= SE IGNORA	
35	UCI	Identifica si el paciente requirió ingresar a una Unidad de Cuidados Intensivos.	CATÁLOGO: SI_ NO	CLAVE	DESCRIPCIÓN
				1	SI
				2	NO
				97	NO APLICA
				98	SE IGNORA
				99	NO ESPECIFICADO

Tabla 4. Diccionario de los datos COVID-19 en México con los 35 descriptores y la interpretación de sus campos.

ANEXO B

P	H	NH	REG_H	#PATIENTS_H_SI	CO_H	D_H	OB_H	A_H	HY_H	IN_H	OD_H	CA_H	RE_H	NR_H
1089852	71334	1018518	8297	5453	2369	90	502	229	241	229	3217	530	2119	2844
			7521	4428	2177	37	301	192	163	192	2426	347	1627	3093
			10345	6487	2863	132	539	527	320	527	3401	546	2450	3858
			33263	18651	8021	471	1280	926	1015	926	8620	1581	6870	14612
			11908	7860	3510	501	496	297	288	297	4072	450	3571	4048
			REG_NH	#PATIENTS_NH_SI	CO_NH	D_NH	OB_NH	AS_NH	HY_NH	IN_NH	OD_NH	CA_NH	R_NH	NR_NH
			102559	35797	7696	413	1420	775	646	775	13764	1750	17366	66762
			143170	47284	10383	951	1361	698	822	698	15838	2233	19019	95886
			170030	50369	11366	744	2057	932	1154	932	17869	2719	22694	119661
			482407	148488	34614	3041	5616	2621	3181	2621	48385	5543	63029	333919
			120352	47211	10257	3420	1234	624	703	624	16275	1572	20511	73141

Tabla 5 Conteos totales divididos por Regiones, tipo de pacientes y Comorbilidades en 203 días de registro con pandemia en México.

ANEXO C

Usuario	Campo
Username	Personalizado
Imagen	Personalizado
Email	Personalizado
Password	Personalizado
Posts	Diccionarios n posts creados

Tabla 6. Estructura del diccionario que se crea cada vez que se registra un nuevo Usuario.

POSTS	Campo
post_id	Personalizado
title	Personalizado
content	Personalizado
Fecha	Personalizado

Tabla 7. Estructura del diccionario que se crea cada vez que se registra una nueva Publicación.