

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS  
DEPARTAMENTO DE POSTGRADO**

**PROYECTO DE TITULACIÓN**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE:**

**“MAGÍSTER EN ESTADÍSTICA APLICADA”**

**TEMA:**

**ANÁLISIS MULTIVARIANTE DE LA INFLUENCIA DE LOS FACTORES  
SOCIOECONÓMICOS Y DEMOGRÁFICOS EN LA SALUD DE  
PACIENTES CON COVID-19, EN LA CIUDAD DE GUAYAQUIL  
DURANTE EL AÑO 2020**

**AUTOR:**

**VICTOR ENRIQUE ARIAS GALÁN**

**Guayaquil - Ecuador**

**2024**

## RESUMEN

Este estudio analizó los datos de los pacientes con diagnóstico de COVID-19, en la ciudad de Guayaquil durante el año 2020 junto con datos de recursos humanos sanitarios, datos sociales de servicios básicos, tiempo de espera para el diagnóstico y acceso a servicios de salud tanto Públicos como Privados.

El análisis se lo realizó utilizando una técnica multivariante denominada HJ-Biplot, dando como resultado que los distritos 09D01, 09D03, 09D04 y 09D05 son los que más aportan a la dimensión 1 (88,29%) del HJ-Biplot, comprobando que estos distritos tienen la mayor oferta de la Red Pública Integral de Salud (RPIS) y de la Red Privada, menores tiempos de espera para el diagnóstico que conlleva a mayor atención ambulatoria y niveles altos de servicios básicos.

La dimensión 2 (11,18%) del HJ-Biplot está conformado por los distritos 09D02, 09D06, 09D07, 09D08, 09D09 y 09D10 cuyos factores de oferta de servicios tanto de la RPIS y Privada es baja, servicios básicos en niveles medio-bajo y los tiempos de espera para el diagnóstico son altos lo que conlleva a mayor necesidad de hospitalización.

En conclusión, este análisis con HJ-Biplot comprobó la utilidad de esta técnica para relacionar los factores socioeconómicos y demográficos e identificar los distritos de Guayaquil más afectados por el COVID-19.

**Palabras clave:** Recursos humanos sanitarios, RPIS, HJ-Biplot, Atención ambulatoria, Hospitalización.

# ABSTRACT

This study analyzed the data of patients diagnosed with COVID-19 in the city of Guayaquil during 2020 along with data on health human resources, social data on basic services, waiting time for diagnosis and access to health services, both Public and Private.

The analysis was carried out using a multivariate technique called HJ-Biplot, resulting in the districts 09D01, 09D03, 09D04 and 09D05 being the ones that contribute the most to dimension 1 (88.29%) of the HJ-Biplots, verifying that these Districts have the greatest offer of the Integrate Public Health Network (RPIS) and the Private Network, shorter waiting times for diagnosis that leads to greater outpatient care and high levels of basic services.

Dimension 2 (11.18%) of the HJ-Biplot is made up of districts 09D02, 09D06, 09D07, 09D08, 09D09 and 09D10 whose service supply factors for both the RPIS and Private are low, basic services at medium levels. low and waiting times for diagnosis are high, which leads to a greater need for hospitalization.

In conclusion, this analysis with HJ-Biplot verified the usefulness of this technique to relate socioeconomic and demographic factors and identify the districts of Guayaquil most affected by COVID-19.

**Keywords:** Health human resources, RPIS, HJ-Biplot, Ambulatory care, Hospitalization.

## **DEDICATORIA**

Dedico este trabajo a mi Madre por confiar en mí, a mi Esposa por la paciencia de apoyarme en mis estudios, pero especialmente está dedicada a mis amigos, compañeros y colegas del Ministerio de Salud Pública que ofrendaron sus vidas durante la pandemia de COVID-19 para mantener el Sistema de Salud Pública en funcionamiento.

## **AGRADECIMIENTO**

Al Ministerio de Salud Pública del Ecuador por brindarme las facilidades para la elaboración de esta Tesis de Fin de Master, por el acceso a sus bases de datos.

Al Señor Doctor Johny Javier Pambabay Calero PhD por su apoyo y paciencia para dirigir mi Tesis.

A la Señora Doctora María Purificación Galindo Villardón PhD por ser mi mentora y guía en este nuevo mundo que es para mí, la Estadística Aplicada.

Y finalmente a la Señora Licenciada Nury Norly Higuera Carriel MSc por su paciencia en el diseño y diagramación de este trabajo.

# DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

---

Victor Enrique Arias Galán

# TRIBUNAL DE GRADUACIÓN

---

María Purificación Galindo Villardón PhD.

PRESIDENTE

---

Johny Javier Pambabay Calero PhD

TUTOR

---

Francisco Moreira Villegas Mgtr.

DOCENTE EVALUADOR

## ABREVIATURAS O SIGLAS

COE	Comité de Operaciones de Emergencia
COVID-19	COronaVirus 2019 Disease
DDDC	Detecting Deviating Data Cells
FODA	Fortalezas, Oportunidades, Debilidades y Amenazas
IESS	Instituto Ecuatoriano de Seguridad Social
INEC	Instituto Nacional de Estadísticas y Censo
ISSFA	Instituto de Seguridad Social de las Fuerzas Armadas
ISSPOL	Instituto de Seguridad Social de la Policia
MSP	Ministerio de Salud Pública
OGC	Open Geospatial Consortium
OMS	Organización Mundial de la Salud
OPS	Organización Panamericana de la Salud
PAHO	Pan American Health Organization
QGIS	Quantum geographic information system
RPIS	Red Pública Integral de Salud
RT-PCR	Real Time - Polymerase Chain Reaction
SARS-Cov-2	Severe acute respiratory syndrome por Coronavirus tipo 2
SENPLADES	Secretaría Nacional de Planificación
VIEPI	Sistema Nacional de Vigilancia en Salud Pública
WHO	World Health Organization

# TABLA DE CONTENIDO

<b>1. INTRODUCCIÓN</b>	<b>1</b>
<b>1.1. Antecedentes</b>	<b>1</b>
<b>1.2. Descripción Del Problema</b>	<b>2</b>
<b>1.3. Objetivos</b>	<b>3</b>
<b>1.4. Hipótesis</b>	<b>3</b>
<b>1.5. Alcance</b>	<b>3</b>
<b>2. MARCO TEÓRICO</b>	<b>5</b>
<b>2.1. SARS-Cov-2</b>	<b>5</b>
<b>2.1.1. Origen</b>	<b>5</b>
<b>2.1.2. Ruta de Trasmisión</b>	<b>5</b>
<b>2.1.3. Sintomatología</b>	<b>5</b>
<b>2.2. Estado Del Arte</b>	<b>6</b>
<b>2.3. Delimitación Del Geográfica Y Poblacional Del Estudio</b>	<b>7</b>
<b>2.3.1. Delimitación Geográfica Del Estudio</b>	<b>7</b>
<b>2.3.2. Población De Estudio</b>	<b>9</b>
<b>3. METODOLOGÍA</b>	<b>11</b>
<b>3.1. Origen De Bases De Datos Por Utilizar</b>	<b>11</b>
<b>3.1.1. Base De Datos De Pacientes Con COVID-19</b>	<b>11</b>
<b>3.1.2. Base De Datos Del Registro Estadístico RAS</b>	<b>12</b>
<b>3.1.3. Base De Datos De Sectores Anonimizados Año 2020</b>	<b>13</b>
<b>3.2. Estadística Descriptiva De La BD De Pacientes Con COVID-19</b>	<b>14</b>
<b>3.2.1. Total De Pacientes Positivos A COVID-19</b>	<b>14</b>
<b>3.2.2. Distribución De Los Pacientes Por Tipo De Institución</b>	<b>16</b>
<b>3.3. Metodología Estadística</b>	<b>16</b>
<b>3.3.1. Método Detecting Deviating Data Cells</b>	<b>16</b>
<b>3.3.2. Random Forest</b>	<b>18</b>
<b>3.3.3. HJ-Biplot</b>	<b>19</b>
<b>4. RESULTADOS</b>	<b>21</b>
<b>4.1. Preprocesamiento</b>	<b>21</b>
<b>4.1.1. Consolidación Y Anonimización BD Pacientes COVID-19-2020</b>	<b>21</b>
<b>4.1.2. Extracción De Datos Sociales De Los Sectores Censales</b>	<b>22</b>
<b>4.1.3. DDDC De Las Bases De Sectores Anonimizados</b>	<b>23</b>

<b>4.1.4. Extracción De BD Del RAS INEC</b>	25
<b>4.1.5. Ejecución del Modelo Random Forest</b>	25
<b>4.2. Elaboración Y Análisis Multivariante Con El Método HJ-Biplot</b>	29
<b>4.2.1. Paquetes De R</b>	29
<b>4.2.2. Generación Del Código</b>	30
<b>4.2.3. Resultados Del Código Del HJ-Biplot</b>	30
<b>4.2.4. Gráficos HJ-Biplot</b>	32
<b>5. CONCLUSIONES Y RECOMENDACIONES</b>	34
<b>5.1. Conclusiones</b>	34
<b>5.2. Recomendaciones</b>	35
<b>5.2.1. Recomendaciones Para Replicar El Estudio</b>	35
<b>5.2.2. Recomendaciones de Salud Pública</b>	36
<b>6. REFERENCIAS</b>	37
<b>7. APÉNDICES</b>	42
<b>7.1. BD De Pacientes COVID-19</b>	42
<b>7.2. Sectores Anonimizados INEC Distritos De Guayaquil</b>	42
<b>7.3. Códigos Y Resultados De La Aplicación De La Técnica DDDC</b>	42
<b>7.4. Códigos De La Técnica Random Forest Aplicado En La BD COVID-19</b>	42
<b>7.5. Códigos De La Técnica HJ-Biplot Con Resultados</b>	42

## LISTADO DE FIGURAS

<b>Figura 2.1</b> Mapa de área de estudio	9
<b>Figura 3.1</b> Distribución de casos COVID-19 por distritos – Guayaquil 2020	14
<b>Figura 3.2</b> Tipo de Egreso de pacientes con COVID-19 por distritos de Guayaquil del año 2020	15
<b>Figura 3.3</b> Mapa de distribución de casos COVID-19 por distritos – Guayaquil 2020	15
<b>Figura 3.4</b> Segmentación de pacientes COVID-19 por institución de atención y distritos – Guayaquil 2020	16
<b>Figura 4.1</b> Herramientas de QGIS de investigación vectorial	23
<b>Figura 4.2</b> Mapa de celdas obtenida por el paquete cellWise, de celdas marcadas y perdidas	24
<b>Figura 4.3</b> Variables en relación con las dimensiones del HJ-Biplot de COVID-19 año 2020.	33
<b>Figura 4.4</b> Cluster de individuos del HJ-Biplot de los pacientes COVID-19 año 2020	33

## LISTADO DE TABLAS

<b>Tabla 2.1</b> Distritos Administrativos Zona 8	8
<b>Tabla 2.2</b> Distritos Administrativos de la ciudad de Guayaquil	8
<b>Tabla 4.1</b> Resumen por distrito de los resultados del modelo DDDC	24
<b>Tabla 4.2</b> Resumen de tasas por 10000 habitantes de médicos y enfermeras, por Distrito año 2020	25
<b>Tabla 4.3</b> Matriz de confusión de Random Forest	26
<b>Tabla 4.4</b> Estadísticas de la matriz de confusión del Random Forest	26
<b>Tabla 4.5</b> Resultados de variables más importante dadas por el algoritmo RF	27
<b>Tabla 4.6</b> Número de variables con mejores métricas	27
<b>Tabla 4.7</b> Matriz consolidada 10 filas y 18 columnas	28
<b>Tabla 4.8</b> Matriz traspuesta consolidada 18 filas y 10 columnas	29
<b>Tabla 4.9</b> Resultados generales del HJ-Biplot	30
<b>Tabla 4.10</b> Contribución relativa de las columnas	30
<b>Tabla 4.11</b> Contribución relativa de las filas	31

## CAPITULO 1

### 1. INTRODUCCIÓN

#### 1.1. Antecedentes

La pandemia de COVID-19 en Ecuador fue el resultado del contagio masivo por COVID-19 en la población ecuatoriana, inició en Wuhan, China a mediados de diciembre del año 2019 en trabajadores de un mercado de la ciudad de Wuhan que presentaban un tipo de neumonía atípica de evolución rápida y que después se relacionó a un virus de origen zoonótico conocido como SARS-Cov-2, y que luego se denominaría COVID-19

El 12 de enero 2020 se publica la secuencia genómica del virus causante del COVID-19, el 13 de enero del 2020 se reporta el primer caso de dicho virus fuera de las fronteras de China, específicamente en Tailandia (OMS, 2020)

La Secretaria General de Comunicación de la Presidencia (Presidencia, 2020) expresó en un boletín: “*Se trata de un caso importado*, sostuvo la secretaria de Estado y explicó que la paciente es una adulta mayor ecuatoriana que llegó desde España, el pasado 14 de febrero. Al momento de su arribo no presentó ningún síntoma; sin embargo, días después, sintió fiebre y malestar general por lo que sus familiares la llevaron a una casa de salud. Ante este cuadro, el 27 de febrero, se le practicaron las pruebas, dando positivo para el nuevo COVID-19”. Este boletín marca el inicio de la pandemia de COVID-19 en Ecuador.

El 17 de marzo del 2020 se decreta toque de queda en todo el Ecuador desde las 21:00 hasta las 5:00, suspensión de trabajo presencial y circulación de personas excepto los que laboren en sectores estratégicos; en la provincia del Guayas, el toque de queda se extendió desde las 16:00 (COE G. d., 2020)

El 31 de marzo del 2020, Guayaquil reportaba 1116 pacientes positivos a COVID-19 con pruebas RT-PCR según reportes del COE Nacional (COE S. N., 2020), para ese instante los

sistemas de salud de la Ciudad de Guayaquil colapsaron, no solo por falta de camas de cuidados intensivos e intermedios, sino que se sumó la escasez de insumos (en especial Oxígeno medicinal y prendas de protección personal) y medicamentos; todo esto unido a que un número importante del personal de atención directa de salud se volvieron positivos a COVID-19.

Además, que el sistema sanitario no se encontraba preparado para un evento de tal magnitud, una coordinación interinstitucional lenta, desinformación de la comunidad de la magnitud del evento epidemiológico y una población sin ingresos estables y viviendo en hacinamiento.

El 19 de mayo del 2020 por gestiones de la Alcaldía de Guayaquil y colaboración de la empresa privada se inauguró el Hospital Bicentenario en los predios de la Antigua Maternidad Enrique Sotomayor; esta infraestructura destinada para atender a pacientes con COVID-19, con una capacidad instalada de 75 camas de cuidados intermedios y capacidad de transporte medicalizado de pacientes que necesiten atención en Unidades de Cuidados Intensivos (MetroEcuador, 2020).

## **1.2. Descripción del Problema**

El 29 de febrero del año 2020 se confirmó el primer caso de COVID-19 en el Ecuador (MSP, 2020); la ciudad de Guayaquil fue la más golpeada durante la primera ola de COVID-19, llegando a su máximo pico el mes de abril.

Con este análisis se pretende identificar la correlación entre los factores socioeconómicos y demográficos que incidieron en la velocidad de transmisión del virus en Guayaquil y su alta letalidad.

La pandemia de la enfermedad por coronavirus 2019 (COVID-19) ha supuesto un gran reto para los sistemas sanitarios, económicos y políticos de todo el mundo. Por tanto, es esencial para las partes interesadas, ante una posible pandemia de características similares,

comprender las relaciones entre los factores socioeconómicos, demográficos y sanitarios que afectan a la pandemia.

### **1.3. Objetivos**

El objetivo general es analizar el efecto de los diferentes factores socioeconómicos, demográficos y sanitarios sobre la letalidad y la incidencia de infectados durante la primera oleada de COVID-19 en la ciudad de Guayaquil, poniendo de manifiesto la utilidad de las técnicas de análisis multivariante.

Los objetivos específicos son:

- Realizar un análisis exploratorio y preprocesamiento de datos e identificar y tratar datos atípicos con las técnicas estadísticas adecuadas.
- Encontrar un mapa factorial que permita visualizar las zonas de Guayaquil más afectadas y definir las estructuras de covariación entre los factores responsables de su posición en los gradientes.
- Proporcionar a las autoridades sanitarias conocimientos científicos sobre los que asentar sus decisiones ante una pandemia de similares características y así poder mitigar sus consecuencias y salvar vidas.

### **1.4. Hipótesis**

Existe una relación entre los factores socioeconómicos, sanitarios y el riesgo de infectarse/morir por COVID-19 y es posible identificar con herramientas multivariantes los sectores más afectados.

### **1.5. Alcance**

Este proyecto se focalizará en la ciudad de Guayaquil durante la primera ola de la pandemia (2020) y su población será todos los pacientes con domicilio en Guayaquil, y que fueron ingresados con diagnóstico positivo a COVID-19 con prueba PCR e ingresados en la plataforma VIEPI y COVID-19 PCR tanto de la Red Pública Integral de Salud (IESS, ISSFA, ISSPOL y MSP) como la Red complementaria.

Se consolidará la información según distrito de planificación territorial (Planificación S. N., 2012), del domicilio del paciente con diagnóstico de COVID-19.

Para concluir, la parte de fundamentos que ampliará los conceptos descritos en esta introducción, serán expuestos en el marco teórico.

## CAPITULO 2

### 2. MARCO TEÓRICO

#### 2.1. SARS-Cov-2

##### 2.1.1. Origen

De acuerdo con el reporte de la Organización Mundial de la Salud (WHO, Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020) el origen del SARS-Cov-2 es un virus zoonótico de la familia pan-Betacoronavirus cuyo huésped original son los murciélagos, pero aún no se conoce sus huéspedes intermedios.

Produciéndose el primer brote en humanos en la ciudad de Wuhan durante el mes de diciembre del año 2019, y más específicamente en el mercado de mariscos y animales vivos de dicha ciudad (WHO, WHO-convened Global Study of the Origins of SARS-CoV-2, 2020).

##### 2.1.2. Ruta De Trasmisión

El COVID-19 se transmite a través de gotitas expulsadas por la vía respiratoria durante el contacto cercano sin protección entre un individuo infectado y otro sano; también a través de fómites, que son objetos inertes que pueden contaminarse con fluidos (mucosas, saliva, etc.) de un paciente infectado con COVID-19, aun cuando esta forma indirecta de infección es poco común por lo lábil que es el virus del COVID-19 (Barrera-Núñez, 2022)

Pero los entornos de trasmisión más comunes son: lugares concurridos, situaciones de contacto cercano (menor a 2 metros) y espacios confinados y con ventilación insuficiente que permite que el virus se mantenga en suspensión en las gotitas y aerosoles expulsados por los pacientes afectados, aun cuando las distancias entre las personas sean mayores a 2 metros.

##### 2.1.3. Sintomatología.

Los síntomas de COVID-19 no son específicos y la presentación de la enfermedad puede variar desde ningún síntoma (asintomático) hasta neumonía grave y muerte. A partir del 20 de febrero de 2020 y basados en 55924 casos confirmados por laboratorio, los signos y síntomas típicos incluyen: fiebre (87,9%), tos seca (67,7%), fatiga (38,1%), producción de

esputo (33,4%), dificultad para respirar (18,6%), dolor de garganta (13,9%), dolor de cabeza (13,6%), mialgia o artralgia (14,8%), escalofríos (11,4%), náuseas o vómitos (5,0%), congestión nasal (4,8%), diarrea (3,7%) y hemoptisis (0,9%). ), y congestión conjuntival (0,8%) (WHO, Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020).

## **2.2. Estado Del Arte**

Desde el inicio de la pandemia se han realizado muchos estudios que analizan los factores socioeconómicos, demográficos y de acceso a los servicios de salud y la incidencia del COVID-19 en distintas poblaciones

Uno de los más difundidos es el Análisis multidimensional de la evolución de la pandemia de la COVID-19 en países de las Américas (Medina Hernández EJ, 2022) en donde se hace una evaluación de comparando datos de los sistemas de salud previo a la llegada del virus a la Región, frente a los casos y muertes acumuladas antes del despliegue de las estrategias de inmunización de la población, y el estado actual de la vacunación

Un estudio muy importante fue el denominado Factores de riesgo poblacional de mortalidad por COVID-19 en 93 países, en el que se analizan las enfermedades crónicas degenerativas, y factores socioeconómicos que aumentan la posibilidad de enfermar y morir por COVID-19, utilizando modelos lineales generalizados para evaluar las variables predictoras y la variable dependiente; pero no se analizaron las relaciones entre las variables predictoras entre si (Hashim, 2020)

Estudios como el denominado Determinantes socioeconómicos, demográficos y sanitarios de la pandemia de COVID-19: un estudio ecológico de España (García, 2021); se centran en la inversión en salud por región, tasa de profesionales de salud por 100000 habitantes, particulados en el ambiente e inequidades de ingreso por región y otras variables más de tipo demográfico.

También hay estudios de factores de riesgo que los adultos mayores y ancianos tengan mayor letalidad que los otros grupos etarios, como el realizado en Yakarta Indonesia (Febrianti T, 2023).

Otros estudios han relacionado las políticas de salud no terapéuticas y su incidencia en la evolución del COVID-19 como el estudio en el que se realiza una evaluación global retrospectiva de factores asociados con las Políticas de salud y la evolución del COVID-19 (Choi, 2022).

Se han realizado el monitoreo de la evolución de la pandemia del COVID-19 en el Reino Unido, y sus regiones, por medio análisis de componentes principales ponderados (Swallow Ben, 2021)

Otros análisis han utilizado índices compuestos como el Índice global de crisis en salud para medir la evolución del COVID-19 (Halder, 2023)

Hay estudios del nivel socioeconómico y el riesgo de hospitalización y fallecer por COVID-19 en áreas urbanas del sur de Europa (Fernández-Martínez, 2022)

### **2.3. Delimitación del Geográfica y Poblacional del Estudio**

#### **2.3.1. Delimitación Geográfica del Estudio**

Como se suscribe en el Acuerdo Ministerial No 557-2012 publicado en el Registro Oficial No 290 (SENPLADES S. N., 2012) en su artículo 1: “Establecer como distritos administrativos de planificación a nivel nacional los siguientes:” refiérase a la “Tabla 2.1” y la “Figura 2.1”.

**Tabla 2.1***Distritos Administrativos Zona 8*

ZONAS ADMINISTRATIVAS DE PLANIFICACIÓN	PROVINCIA GUAYAS	CÓDIGO DISTRITAL	DESCRIPCIÓN DEL DISTRITO (PARROQUIAS RURALES, URBANAS Y CANTONES)
ZONA 8	CANTONES GUAYAQUIL SAMBORONDÓN DURAN	09D01	XIMENA 1 PARROQUIA RURAL: PUNA - ESTUARIO DEL RÍO GUAYAS
		09D02	XIMENA 2
		09D03	GARCÍA MORENO - LETAMENDI AYACUCHO - OLMEDO- BOLÍVAR- SUCRE URDANETA - 9 DE OCTUBRE - ROCAFUERTE - PEDRO CARBO- ROCA
		09D04	FEBRES CORDERO
		09D05	TARQUI- 1, TENGUEL
		09D06	TARQUI-2
		09D07	PASCUALES - 1
		09D08	PASCUALES - 2
		09D09	TARQUI-3
		09D10	PARROQUIAS RURALES: PROGRESO - ELMORRO- POSORJA ÁREA DE EXPANSIÓN DE GUAYAQUIL SAMBORONDÓN
		09D23	SAMBORONDÓN
		09D24	DURAN

Fuente: Acuerdo Ministerial SENPLADES 557-2012 página 5

De los distritos administrativos de la Zona 8 serán **sólo** incluidos los que pertenecen al Cantón Guayaquil “Tabla 2.2”.

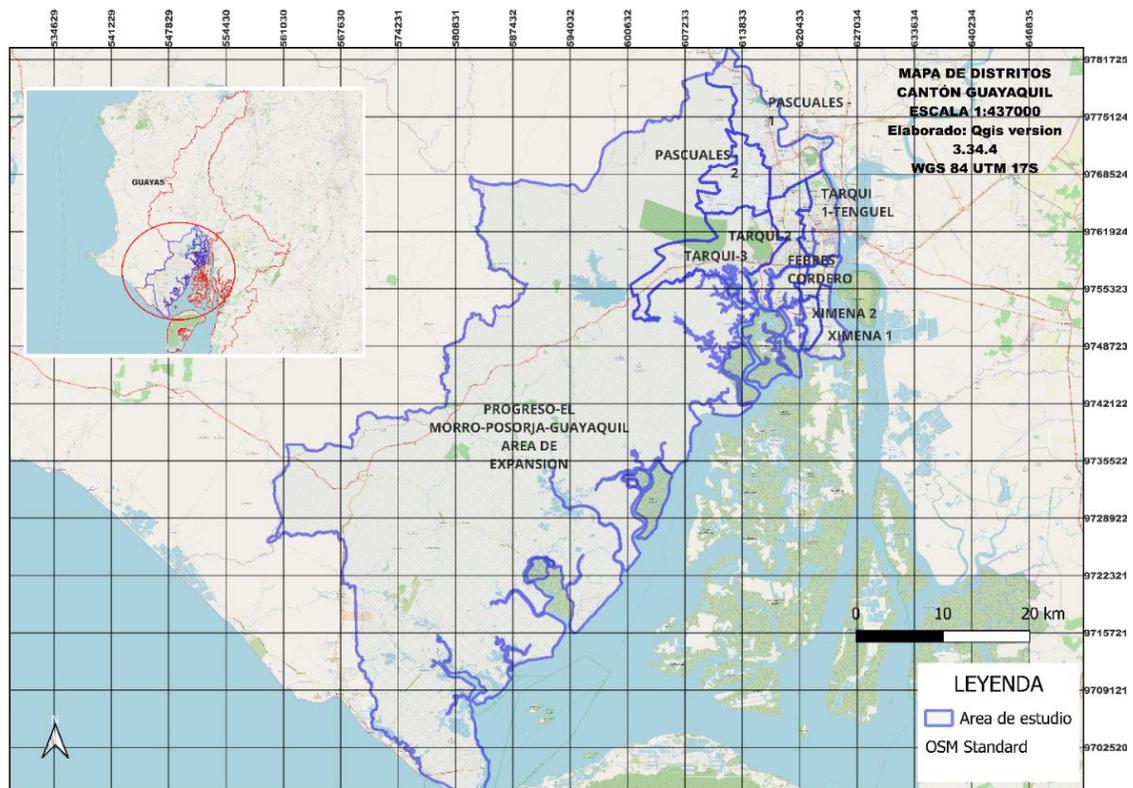
**Tabla 2.2***Distritos Administrativos de la ciudad de Guayaquil*

CÓDIGO DISTRITO	NOMBRE
09D01	Ximena 1
09D02	Ximena 2
09D03	García Moreno, Letamendi, Ayacucho, Olmedo, Sucre, Urdaneta, 9 de Octubre, Pedro Carbo, Roca
09D04	Febres Cordero
09D05	Tarqui 1 – Tenguel
09D06	Tarqui 2
09D07	Pascuales 1
09D08	Pascuales 2
09D09	Tarqui 3
09D10	Progreso, El Morro, Posorja, Zona de expansión de Guayaquil

Fuente: Acuerdo Ministerial SENPLADES 557-2012 página 5

**Figura 2.1**

*Mapa de área de estudio*



Nota: Adaptado por el autor del mapa por distritos de SENPLADES con base de uso libre de OSM con QGIS 3.34.4 (SENPLADES, 2024) WGS 84 UTM 17S

### **2.3.2. Población de Estudio.**

La población que se incluirá en el estudio tendrá los siguientes criterios de inclusión:

- Pacientes con prueba RT-PCR positiva que estén registrados en el Sistema Nacional de Vigilancia en Salud Pública – Ecuador (VIEPI) o en el aplicativo COVID-19-PCR.
- Sitio de domicilio en alguno de los Distritos del Cantón Guayaquil
- El resultado de su prueba positiva fue entre marzo y diciembre del año 2020.

Se segmentarán los pacientes por grupos etarios, que se basan en los ciclos de vida como lo indica el Manual del Modelo de Atención Integral de Salud – MAIS en su página 62 -

Tabla 1 (MSP M. d., 2018); el cual ayuda a garantizar la atención integral de manera oportuna, sin costo y en especial a los grupos prioritarios.

La segmentación por grupo etario será la siguiente:

- Menores 1 año – Lactantes.
- 1 a 4 años – Primera infancia.
- 5 a 9 años – Infancia.
- 10 a 14 años – Primera etapa de la adolescencia.
- 15 a 19 años – Segunda etapa de la adolescencia.
- 20 a 64 años – Adulto.
- 65 a 74 años – Adulto mayor.
- 75 años o más – Vejez.

También se segmentarán por otros criterios como son:

- Sexo.
- Sistema de salud en cual se atendió.
- Distrito de domicilio.
- Y otros que se consideren necesarios (tiempo de atención, etc.).

La metodología por utilizar será cuantitativa sobre la base de datos de paciente del año 2020, y será ampliada en el siguiente capítulo.

## CAPITULO 3

### 3. METODOLOGÍA

#### 3.1. Origen de Bases de Datos Por Utilizar

Las bases de datos que utilizaremos serán las siguientes:

- Exportado de pacientes ingresados en el Sistema Nacional de Vigilancia en Salud Pública Ecuador (PAHO, 2024) hasta el mes de junio del 2020.
- Exportado ingresados en el Aplicativo COVID-19 PCR desde julio 2020 hasta diciembre 2020, actualmente fuera de línea y reemplazado por el aplicativo Sistema de Vigilancia Integrada de COVID y Otros Virus Respiratorios del MSP (MSP M. d., Sistema de Vigilancia Integrada de COVID y Otros Virus Respiratorios, 2021)
- Base de datos del Registro Estadístico de Recursos y Actividades de Salud 2020 diciembre (INEC I. N., 2024)
- Base de datos de sectores censales anonimizados año 2020 (INEC C. E., 2024)

##### **3.1.1. Base de Datos de Pacientes con COVID-19.**

Esta base de datos, que es la unión de los exportados del VIEPI (hasta junio 2020) y los exportados del aplicativo COVID-19 PCR (desde julio 2020).

Los criterios de inclusión de los pacientes fueron los siguientes:

- Paciente con prueba RT-PCR positiva ingresados en alguno de los sistemas de información del MSP durante el año 2020.
- Pacientes cuyo domicilio sea alguno de los distritos de la ciudad de Guayaquil (Tabla 2.2) durante el año 2020.

La base de datos luego del filtrado según criterios anteriores cuenta con 19906 registros y 48 variables de las cuales, solo utilizaremos, para este estudio las siguientes:

- Institución. – Se refiere a que institución atendió al paciente, sea de la Red Pública Integral de Salud (RPIS) o Red complementaria (Privados), en formato de texto.
- Fecha de atención. – Fecha en que fue atendido el paciente en formato de fecha DD/MM/AAAA.
- Distrito domicilio. – Distrito en donde vivía el paciente durante su infección de COVI-19, está en formato de texto con el formato de la Secretaria Nacional de Planificación (Planificación S. N., 2024).
- Sexo. - Hombre o mujer, en formato de texto.
- Edad. – En años cumplidos, en formato numérico.
- Inicio de síntomas. – Fecha de inicio de síntomas en formato de fecha DD/MM/AAAA.
- Comorbilidad. – Si el paciente padecía de otras enfermedades antes del diagnóstico de COVID-19, es una variable binaria en formato de texto.
- Fec\_entr1. – Fecha de entrega del resultado del RT-PCR, formato de fecha DD/MM/AAAA
- Condición final. – Condición de egreso del paciente, formato texto variable binaria (VIVO o MUERTO)

Esta matriz de datos se consolidó por la variable Distrito y además las edades se segmentaron según los grupos etarios utilizados por el MSP del Ecuador (MSP M. d., 2018).

### **3.1.2. Base de datos del Registro Estadístico de Recursos y Actividades de Salud 2020 Diciembre (INEC I. N., 2024).**

Esta Base de datos, como su nombre lo indica contiene el registro de todos y cada uno de los recursos y actividades de salud a nivel nacional en el año 2020. Esta base de datos contiene 4136 registros y 1931 variables, de las se utilizaron las siguientes:

- Provincia. – Solo se utilizó la provincia del Guayas, es una variable cualitativa en formato de texto.

- Cantón. – Solo utilizamos el cantón Guayaquil, es una variable cualitativa en formato de texto.
- Parroquia. - Las pertenecientes al cantón Guayaquil, es una variable cualitativa en formato de texto.
- Clase. – Es el tipo de servicio de salud, es una variable cualitativa en formato de texto.
- Entidad. – Indica si pertenece a la Red Pública Integral de Salud (RPIS) o a la Red Complementaria (privados).
- Sector. – SI es público o privado, variable cualitativa en formato de texto
- Tipo. – Define el tipo de establecimiento de salud según su capacidad resolutive, es una variable cualitativa en formato de texto
- Bloque (1) total médicos. - Es el número total de médicos en el registro indicado, variable cuantitativa en formato de número.
- Bloque (1) Enfermeras Total. – Es el número de enfermeras en el registro indicado, variable cuantitativa en formato de número.

### **3.1.3. Base de Datos de Sectores Anonimizados Año 2020 (INEC C. E., 2024).**

Esta base de datos obtenida del INEC es un archivo de tipo geopackage, que es un formato abierto de información geoespacial construido en base a SQLite utilizado para transferir gran cantidad de datos espaciales y creado por la Open Geospatial Consortium (OGC, 2024).

Esta base de datos en este formato espacial contiene los registros de todos los sectores censales del Ecuador, de los cuales, se seleccionaron los pertenecientes a los distritos del cantón Guayaquil utilizando QGIS versión 3.34.4 con licencia GNU.

Las variables que se seleccionaron por distrito pertenecientes al cantón Guayaquil son las siguientes:

- p\_hog. – Esta variable cuantitativa indica el número de personas por hogar en ese sector censal, variable cuantitativa continua, formato numérico.

- pob\_serv\_b. – Porcentaje de población con servicios básicos en ese sector censal, variable cuantitativa continua con formato numérico (entre 0 y 1)

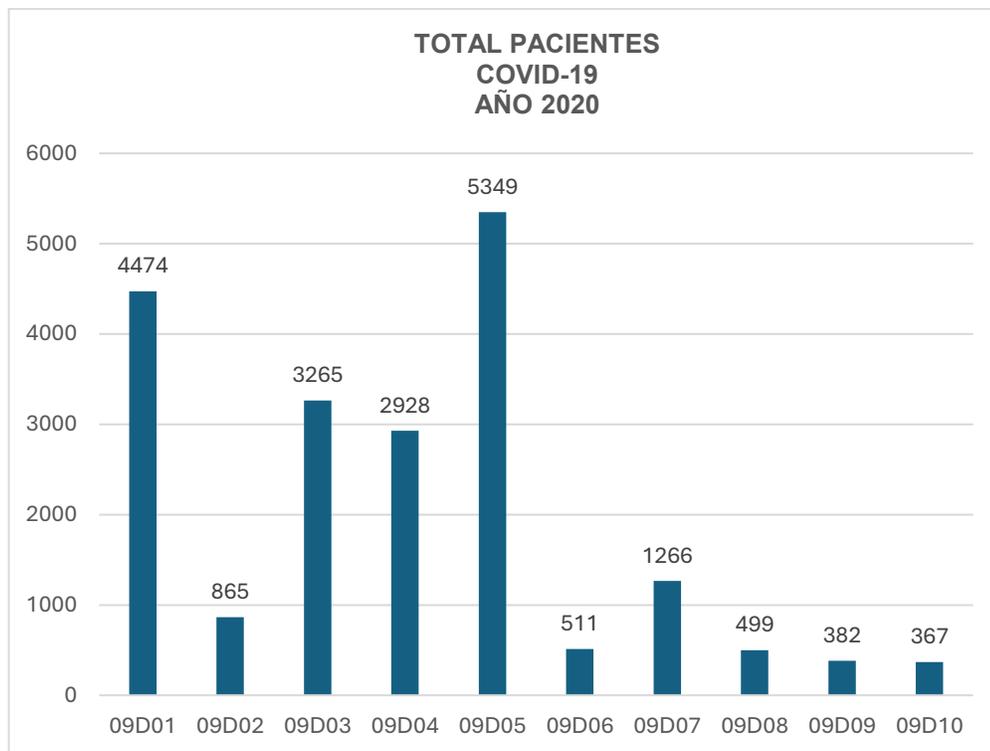
### 3.2. Estadística Descriptiva de la Base de Datos de Pacientes con COVID-19

#### 3.2.1. Total de Pacientes Positivos a COVID-19.

Durante el año 2020 en la ciudad de Guayaquil se registraron 19906 pacientes con prueba positiva RT-PCR positiva distribuidos por distritos según por egresos, mostrado en “Figura 3.2” y totales, mostrado en las” Figura 3.1” y “Figura 3.3”.

**Figura 3.1**

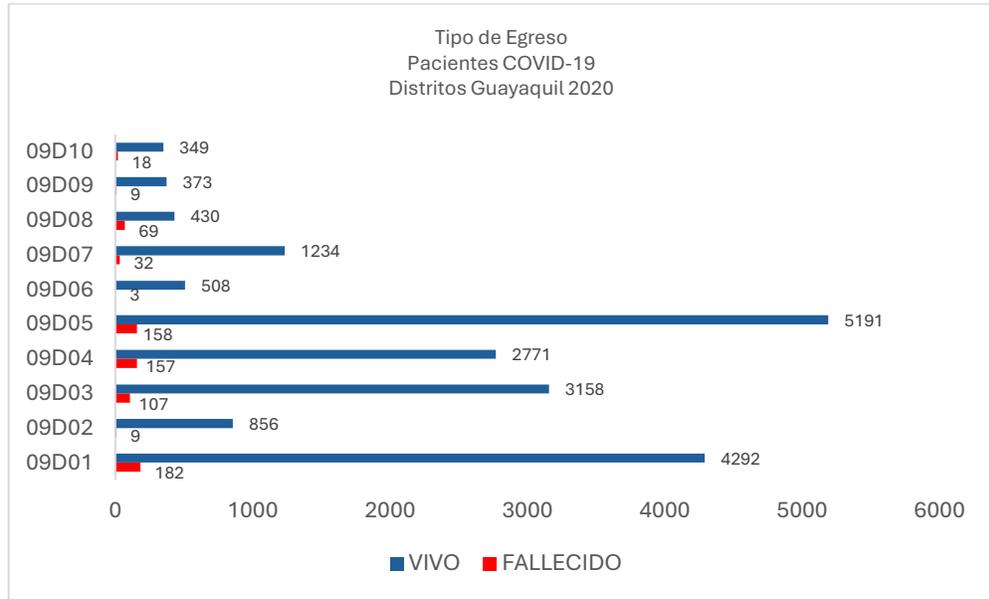
*Distribución de casos COVID-19 por distritos – Guayaquil 2020*



Nota: Para ubicar los nombres de los distritos referirse a la “Tabla 2.2”

**Figura 3.2**

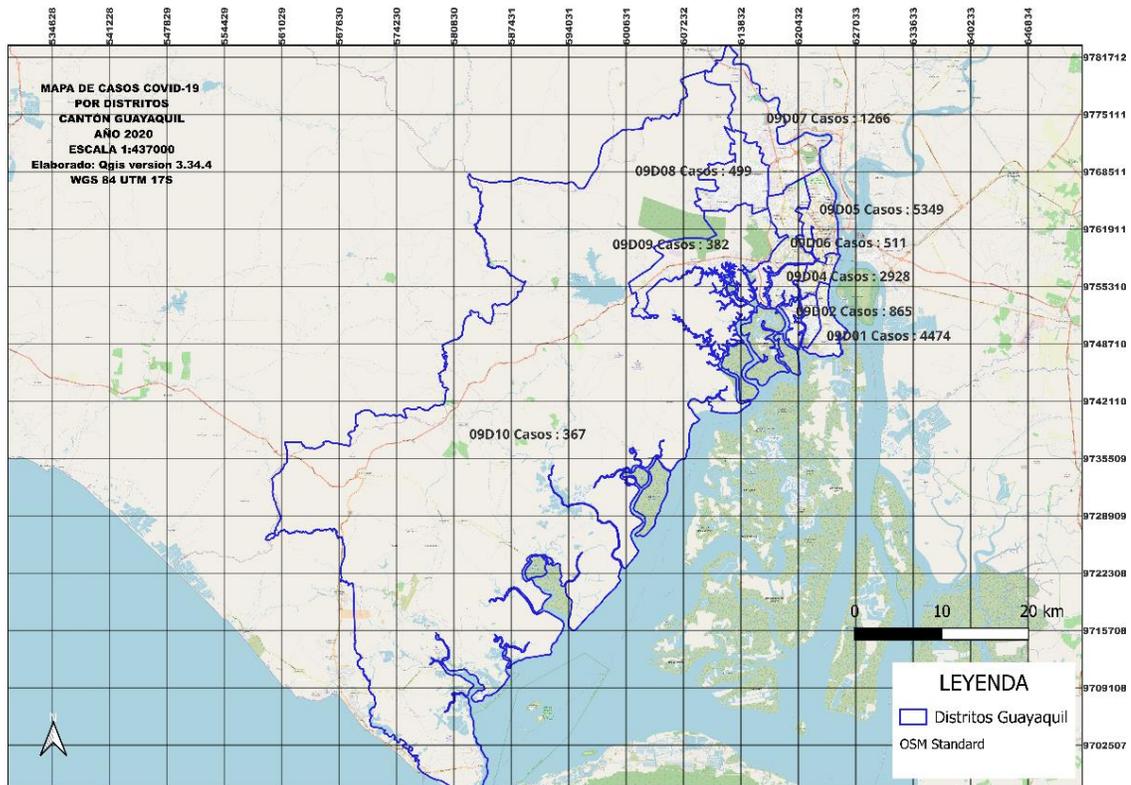
*Tipo de Egreso de pacientes COVID-19 por distritos – Guayaquil 2020*



Nota: Tipo de egreso es la condición final del paciente

**Figura 3.3**

*Mapa de distribución de casos COVID-19 por distritos – Guayaquil 2020*

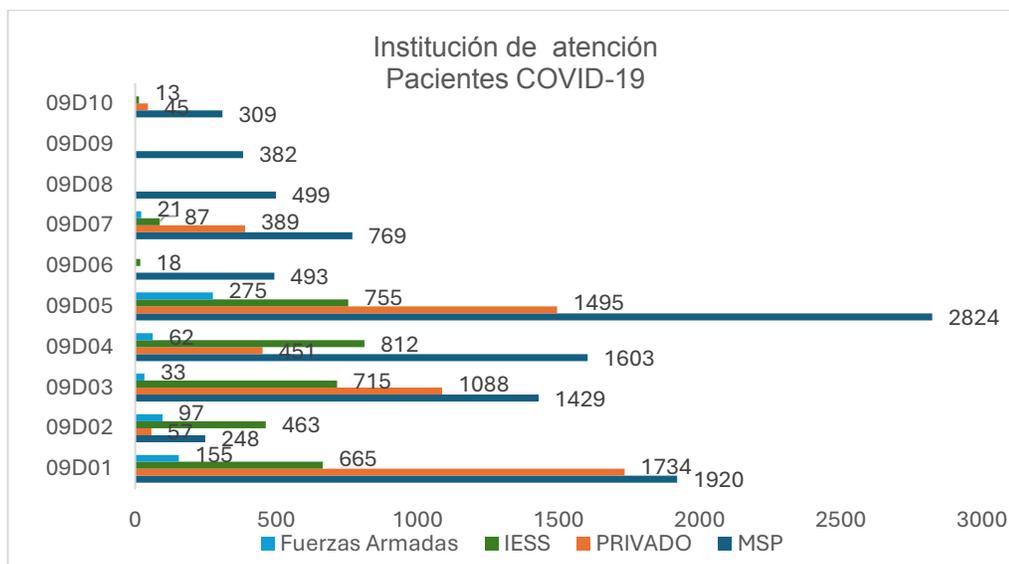


### 3.2.2. Distribución de los Pacientes por Tipo de Institución.

Se segmentan los pacientes por la institución donde se atendió, sea esta del RPIS o Privados, como se muestra en la “Figura 3.4”.

**Figura 3.4**

*Segmentación de pacientes COVID-19 por institución de atención y distritos – Guayaquil 2020.*



Nota: Para ubicar los nombres de los distritos referirse a Tabla 2.2

### 3.3. Metodología Estadística

#### 3.3.1. Método Detecting Deviating Data Cells

El método Detecting Deviating Data Cells (DDDC) es una técnica de estadística robusta que permite identificar y analizar datos atípicos o anómalos en conjuntos de datos complejos, fue descrita y explicada por Peter J. Rousseeuw (Rousseeuw, 2018).

Se basa en la importancia de detectar datos anómalos en estructuras multivariantes, que permite los siguientes beneficios en el análisis estadístico:

- Mejorar la calidad de los datos, porque, al identificar celdas anómalas nos ayuda a corregir y mantener la calidad e integridad de los datos.

- Evita sesgos en el análisis, debido a que pueden distorsionar la interpretación de los resultados estadístico.
- Facilita la identificación de problemas al evidenciar desviaciones (errores en el ingreso de datos) no previstas por el investigador y que pueden ser relevantes en el estudio.

La técnica DDDC un método para detectar celdas de datos desviadas en conjuntos de datos multivariantes que incluyen los siguientes pasos y técnicas, en resumen:

- Preprocesamiento de datos por medio de estandarización de las columnas  $j$  de la matriz, por medio de un estimador robusto de localización (como la mediana muestral) y un estimador robusto de escala (como la desviación mediana absoluta - MAD); que asume que este estimador ya está centrado; lo que nos lleva a una  $\mathbf{Z}_{ij}$  estandarizada.
- Detectamos los datos atípicos univariantes de la matriz  $\mathbf{Z}_{ij}$  y los incluimos en una nueva matriz  $\mathbf{W}$  (que sean menores o iguales al valor crítico de corte <sup>1</sup>) y, los valores en blanco.
- Relaciones bivariantes entre dos columnas (variables)  $h \neq j$ , que tengan una correlación límite  $\geq 0.5$  (variables conectadas), y a estas variables se les aplicara una medida de correlación robusta al par de variables  $h, j$ .
- Se calculan los valores de predicción (por las correlaciones robustas)  $Z_{ij}$  para todas las celdas.
- Alargamiento de las predicciones  $Z_{ij}$  ( $i$  posición en la fila y  $j$  posición en la columna) para todas las variables  $j$  conectadas.
- Marcar celdas potencialmente atípicas a partir de las celdas estandarizadas, los que nos lleva a crear una matriz  $\mathbf{Z}_{imp}$  ( $Z$  imputada).
- Marcar filas potencialmente atípicas en base de los residuos de las celdas estandarizadas

---

<sup>1</sup> El punto crítico de corte esta dado por  $c = \sqrt{\chi_{1,p}^2}$ ,

- Desestandarizar (volver a la escala original de las variables) la matriz imputada  $Z_{imp}$ , en una matriz  $X_{imp}$  deshaciendo el paso uno. Esta matriz imputada  $X_{imp}$  tendrá las celdas atípicas marcadas y de existir, también estarán marcadas las filas atípicas.
- Este método de estadística robusta se utiliza en este estudio, para detectar valores atípicos, que pudieran ser valores mal ingresados en las bases de datos o valores que siendo atípicos podrían estar evidenciando variables ocultas en el estudio.

### **3.3.2. Random Forest**

Random Forest (Breiman, 2001) es una técnica de aprendizaje supervisado ampliamente utilizada en el campo de la estadística aplicada, la técnica se basa en la construcción de múltiples árboles de decisión, cada uno de ellos entrenado utilizando un subconjunto aleatorio de variables predictivas y observaciones del conjunto de datos original.

En última instancia, se crea un modelo que combina las predicciones de estos árboles individuales, lo que da como resultado predicciones que son más precisas y sólidas que las de un único árbol de decisión. Los pasos resumidos para la aplicación de la técnica Random Forest son los siguientes:

- El primer paso es identificar las variables predictoras que son relevantes para el problema que se está resolviendo, esto puede implicar un análisis exploratorio de los datos, eliminar variables altamente correlacionadas y seleccionar las variables más significativas.
- En cada iteración del algoritmo, se selecciona aleatoriamente un subconjunto de observaciones del conjunto de datos original (comúnmente conocido como "bagging"), esto ayuda a reducir la variación en el modelo final y evita el sobreajuste.
- Se construye un árbol de decisión para cada subconjunto seleccionado de observaciones, este proceso se repite de forma recursiva hasta que se alcanza un criterio de parada, como el número máximo de nodos o la pureza mínima de los nodos de las hojas.

- Según el método para la creación de árboles de decisión, puede utilizar métricas para determinar la precisión y rendimiento del modelo, como el error cuadrático medio (MSE) o coeficiente de determinación (R cuadrado).
- Además, se realizar una validación cruzada para obtener estimaciones más sólidas del rendimiento del modelo, siendo la más comúnmente utilizada la k-fold cross validation.

### 3.3.3. HJ-Biplot

Sea  $\mathbf{X}_{n \times p}$  una matriz de datos de  $n$  filas (individuos) por  $p$  columnas (variables). Las columnas de  $\mathbf{X}$  suelen estar estandarizadas para hacerlas comparables.

El HJ-Biplot es un método de análisis de datos exploratorio que busca patrones ocultos en la matriz de datos. El método es de alguna manera similar al Análisis de Correspondencia (Benzécri, 1973), pero no está restringido solo a datos de frecuencia.

El HJ-Biplot es una representación conjunta, en un espacio vectorial de baja dimensión (normalmente un plano bidimensional), de las filas y columnas de  $\mathbf{X}$ , mediante el uso de marcadores (puntos/vectores),  $i_1, i_2, \dots, i_n$  para sus filas y  $j_1, j_2, \dots, j_p$  para sus columnas. Los marcadores se obtienen a partir de la habitual descomposición en valores singulares (SVD) de la matriz de datos  $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ , donde  $\mathbf{U}$  son los vectores propios de  $\mathbf{X}\mathbf{X}^T$ ,  $\mathbf{V}$  son los vectores propios de  $\mathbf{X}^T\mathbf{X}$  y  $\mathbf{\Sigma}$  es una matriz diagonal que contiene los valores singulares (es decir, las raíces cuadradas de los valores propios no ceros de  $\mathbf{X}\mathbf{X}^T$  y  $\mathbf{X}^T\mathbf{X}$ ), tomando como marcadores de fila las filas de  $\mathbf{I} = \mathbf{U}\mathbf{\Sigma}$  y como marcadores de columna las filas de  $\mathbf{J} = \mathbf{V}\mathbf{\Sigma}$  en la dimensión apropiada.

Galindo (GALINDO VILLARDÓN, 1986) muestra que es posible, con esta elección particular de los marcadores, representar conjuntamente filas y columnas en un mismo sistema de referencia obteniendo la más alta calidad de representación (bondad de ajuste).

Las reglas de interpretación del HJ-Biplot son una combinación de las reglas utilizadas en otras Técnicas de Escalamiento Multidimensional, Análisis de Correspondencia, Análisis Factorial y Biplots Clásicos (**GH<sup>T</sup>** biplot y **JK<sup>T</sup>** biplot) (Gabriel K. R., 1971):

- Las distancias entre marcadores de fila se interpretan como una función inversa de sus similitudes, de forma que los marcadores más cercanos (individuos) son más similares. Esta propiedad permite la identificación de clúster de individuos con perfiles similares. Se puede utilizar cualquier técnica de agrupación jerárquica o no jerárquica como ayuda para identificar los clústeres más relevantes.
- Las longitudes de los marcadores de columna (vectores) se aproximan a la desviación estándar de las variables.
- Los cosenos de los ángulos entre los vectores columna aproximan las correlaciones entre variables de tal manera que, los ángulos agudos pequeños se asocian a variables con alta correlación positiva, los ángulos obtusos cercanos al ángulo llano se asocian a variables con alta correlación negativa y; los ángulos rectos están asociados a variables no correlacionadas. De la misma manera, los cosenos de los ángulos entre los marcadores de variables y los ejes (Componentes Principales) aproximan las correlaciones entre ellos. Para datos estandarizados, estos se aproximan a las cargas factoriales en el Análisis Factorial.
- El orden de las proyecciones ortogonales de los marcadores de fila (puntos) sobre un marcador de columna (vector) se aproxima al orden de los elementos de fila (valores) en esa columna (la misma propiedad es válida para la proyección de los marcadores de columna en la dirección definida por un marcador de fila). Cuanto mayor es la proyección de un punto individual sobre un vector variable, más se desvía este individuo del promedio de esa variable. Para que esto se interprete correctamente, la relación de aspecto del gráfico hj-biplot debe ser 1:1, es decir, las escalas físicas horizontales y verticales tienen que ser las mismas.

## CAPITULO 4

### 4. RESULTADOS

#### 4.1. Preprocesamiento

##### 4.1.1. *Consolidación y Anonimización Base de Datos Pacientes COVID-19 año 2020.*

En este preprocesamiento se anonimizó cada individuo de la base de datos para que no puedan ser identificables como lo indica el documento denominado “Política para el tratamiento de datos personales en el Ministerio de Salud Pública”, en la página 14 en el acápite Bases de datos, punto tres que indica textualmente “En lo posible se utilizará pseudo-anonimización y anonimización sobre las bases de datos en producción” (MSP, 2022).

Luego de la anonimización, se corrigieron errores en la migración de las bases de datos, siendo los errores más frecuentes los siguientes:

- Datos duplicados de pacientes
- Errores en ciertos caracteres especiales (tildes y diéresis) que podría excluir datos de la base.
- Errores en sitio de domicilio de los pacientes.

Al final se obtuvo una base de datos con 19906 registros y 10 variables, cuyo link se encuentra en el “Anexo A - link A1” que son las siguientes:

- Institución. – Se refiere a que institución atendió al paciente, sea de la Red Pública Integral de Salud (RPIS) o Red complementaria (Privados), en formato de texto.
- Fecha de atención. – Fecha en que fue atendido el paciente en formato de fecha DD/MM/AAAA.

- Distrito domicilio. – Distrito en donde vivía el paciente durante su infección de COVI-19, está en formato de texto con el formato de la Secretaria Nacional de Planificación (Planificación S. N., 2024).
- Sexo. - Hombre o mujer, en formato de texto.
- Edad. – En años cumplidos, en formato numérico.
- Inicio de síntomas. – Fecha de inicio de síntomas en formato de fecha DD/MM/AAAA.
- Comorbilidad. – Si el paciente padecía de otras enfermedades antes del diagnóstico de COVID-19, es una variable binaria en formato de texto.
- Fec\_entr1. – Fecha de entrega del resultado del RT-PCR, formato de fecha DD/MM/AAAA
- Condición final. – Condición de egreso del paciente, formato texto variable binaria (VIVO o MUERTO)

#### ***4.1.2. Extracción De Datos Sociales De Los Sectores Censales Por Distrito De La Ciudad Guayaquil.***

La base de datos que contenía los datos sociales que incluiríamos en nuestro estudio se encontraban en un geopackage (OGC, 2024) de uso libre del censo del INEC del año 2020 (INEC C. E., 2024), y para extraerlos se utilizó un SIG de uso libre denominado QGIS 3.34.4, y por medio de Herramientas de investigación vectorial se extrajo todos los sectores censales de cada distrito del área de estudio (Figura 4.1).

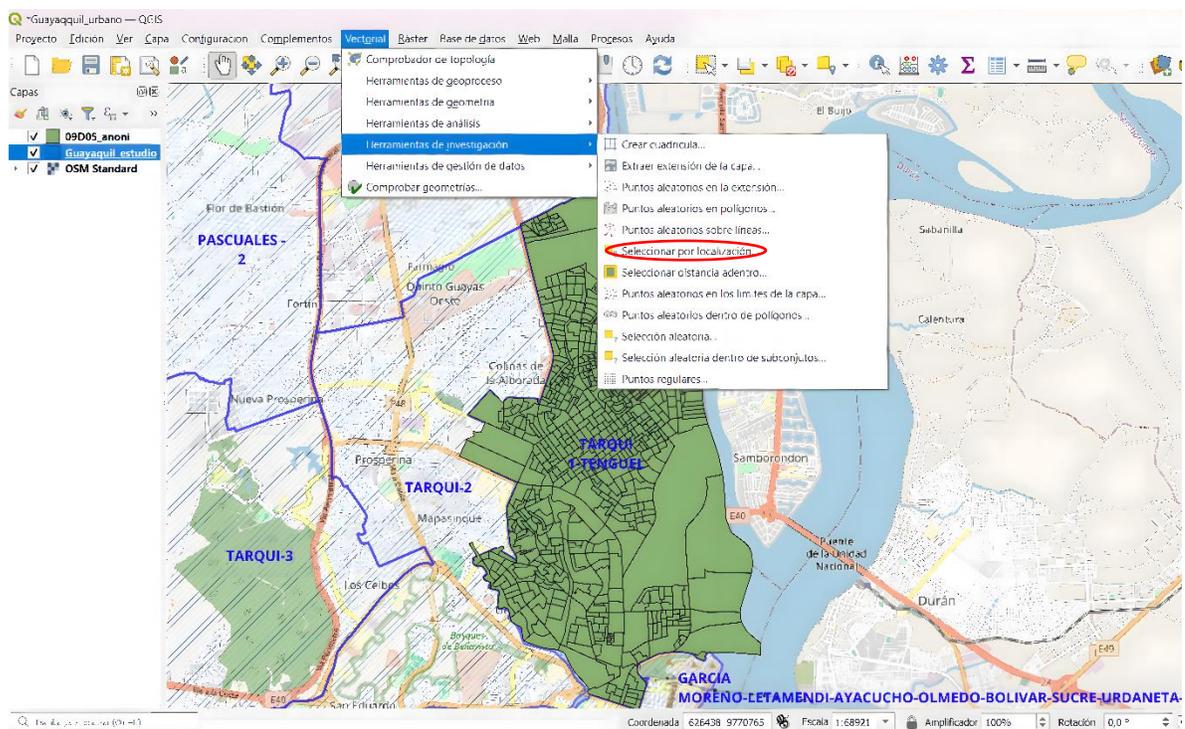
Esta Herramienta vectorial utiliza una capa base, que en nuestro caso es el distrito que nos da la base para extraer los sectores anonimizados del censo INEC año 2020, seleccionando solo los sectores que se encuentran dentro de la capa base, que para nosotros es el distrito seleccionado.

De estos sectores censales anonimizados, que está disponible en “Anexo A – link A2” , se extrae de su base de datos (csv) lo siguiente:

- fid. - Que es la identificación única del sector censal anonimizado del cantón Guayaquil.
- p\_hog. - Número de habitantes por hogar en dicho sector censal.
- pob\_ser\_b. – Población de viviendas con servicios básicos.

**Figura 4.1**

*Herramientas de QGIS de investigación vectorial*



Nota: Esta figura muestra un ejemplo del procedimiento que se utilizó para cada distrito del área de estudio y fue elaborado por el autor con QGIS 3.34.4

**4.1.3. Detección de Datos Perdidos y Datos Atípicos (DDDC) de las Bases**

***Extraídas de los Sectores Anonimizados de los Distritos.***

Luego de extraer los sectores censales de cada distrito de la ciudad de Guayaquil se utilizó el método de estadística robusta denominado Detecting Deviating Data Cells (DDDC), utilizando la librería en R *cellwise* cuyo objetivo es detectar valores atípicos en las celdas del conjunto de datos (Raymaekers, 2023).

Para cada conjunto de sectores censales por distrito se elaboró un script individual que nos marcó las celdas atípicas y ND, luego las imputó por medio del método DDDC, lo que nos dio dos nuevos conjuntos de datos denominados  $09dxx_i$  que nos da el conjunto de datos de ese distrito ya imputado, y el otro denominado  $09dxx_f$  que son las celdas marcadas como atípicas y que fueron imputadas, que están disponibles en “Anexo A – link A3”.

En la “Tabla 4.1” se resumen las imputaciones de los distritos del estudio, y en la “Figura 4.2” se incluye una imagen de las salidas graficas del modelo.

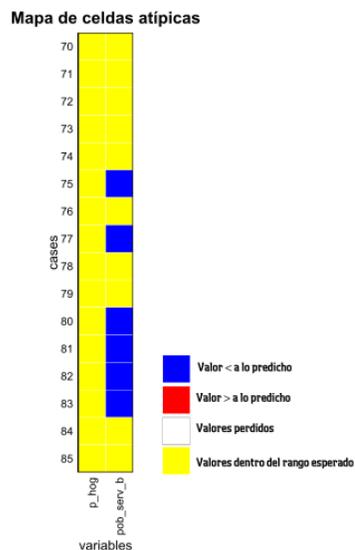
**Tabla 4.1**

*Resumen por distrito de los resultados del modelo DDDC*

<i>Distrito</i>	<i>n</i> filas	<i>p</i> columnas	celdas marcadas
09D01	696	2	83
09D02	556	2	76
09D03	596	2	74
09D04	777	2	100
09D05	731	2	145
09D06	818	2	70
09D07	860	2	230
09D08	1034	2	382
09D09	212	2	34
09D10	558	2	4

**Figura 4.2**

*Mapa de celdas obtenida por el paquete cellWise, de celdas marcadas*



#### **4.1.4. Extracción de Datos de la BD del INEC de Registro Estadístico de Recursos y Actividades de Salud 2020 (RAS).**

De esta BD obtuvimos el número de médicos y número de enfermeras por distrito de la ciudad de Guayaquil y luego se calculó la tasa por 10000 habitantes de dichos recursos humanos sanitarios, como lo resumimos en la “Tabla 4.2”.

**Tabla 4.2**

*Resumen de tasas por 10000 habitantes de médicos y enfermeras, por Distrito año 2020*

Distrito	Población 2020	med_10000 hab.	enf_10000 hab.
09D01	367822	29,14	14,87
09D02	279454	30,42	27,98
09D03	266475	43,27	26,49
09D04	405013	14,76	9,83
09D05	298463	61,72	33,57
09D06	340678	24,57	18,76
09D07	295114	15,11	7,76
09D08	368212	6,19	1,58
09D09	51911	6,93	0,96
09D10	53678	6,15	0,93

#### **4.1.5. Ejecución del Modelo Random Forest.**

En la matriz de datos de individuos con COVI-19, con todas las variables sociales, extraída de los sectores anonimizados del Censo 2020 y variables de recursos sanitarios extraídos de la base de datos RAS del INEC año 2020, se aplica la técnica de aprendizaje supervisado, utilizando el paquete en R denominado *randomForest* (Breiman L. e., 2022), para clasificar las variables más importantes.

Se corre el script, cuyos códigos y modelo se encuentran en el “Anexo A – link A4” , con todas las variables incluidas en la nueva base de datos y se obtienen los siguientes resultados presentado en la “Tabla 4.3”.

**Tabla 4.3***Matriz de confusión de Random Forest*

MATRIZ DE CONFUSIÓN		
<i>Predicción</i>	<i>Referencia</i>	
	MUERTO	VIVO
MUERTO	683	32
VIVO	61	19130

**Tabla 4.4***Estadísticas de la matriz de confusión del Random Forest*

ESTADÍSTICAS MATRIZ DE CONFUSIÓN	
<i>Métrica</i>	<i>Valor</i>
Precisión	0,9953
Kappa	0,9338
Sensibilidad	0,9180
Precisión balanceada	0,9581
Test de Mcnemar's	0,00369

La interpretación de la matriz de confusión y sus estadísticas presentadas en “Tabla 4.3” y “Tabla 4.4” nos indica una Precisión (Accuracy) de 0.9953 nos indica una buena clasificación de los casos y la kappa de 0.9338 nos indica que el modelo es casi perfecto, la Precisión Equilibrada (Balanced Accuracy) de 0.958 indica un buen rendimiento del modelo y una Sensibilidad de 0.91801 nos indica que el modelo tiene pocos falsos negativos.

Por último, el Test Mcnemar`s P- Value de 0.00391 nos indica que el resultado tiene significancia estadística.

Las variables más importantes se resumen en la “Tabla 4.5”.

**Tabla 4.5**

*Resultados de variables más importante dadas por el algoritmo RF*

<i>Variables</i>	<i>En conjunto</i>
Lugar de atención UCI	<b>100</b>
T_IS_Dx	<b>43,008</b>
T_IS_At	<b>40,057</b>
T_At_Dx	39,836
Lugar de atención Hospitalización	<b>27,437</b>
Sexo MUJER	<b>11,373</b>
p_hog	<b>8,144</b>
G_etario 75 o + años	<b>7,72</b>
G_etario 20 a 64 años	<b>7,414</b>
Institución IESS	<b>7,352</b>
enf x 10000 hab.	<b>7,344</b>
Institución MSP	<b>7,128</b>
pob_serv_b	<b>6,512</b>
med x 10000 hab.	<b>6,401</b>
Comorbilidad SI	<b>6,354</b>
G_etario 65 a 74 años	<b>5,96</b>
Distrito domicilio 09D04	2,295
Distrito domicilio 09D03	2,188
Institución JUNTA DE BENEFICENCIA	<b>1,896</b>
Institución FUERZAS ARMADAS	<b>1,715</b>

Nota: Las variables cuyos valores en conjunto están en **negritas** son las que se utilizarán

Las métricas del Random Forest nos da 20 variables como el número optimo, con una Precisión de 0.968 eso significa que los datos están bien equilibrados, pero la kappa de 0.5099 entre moderada y sustancial y se resume en la “Tabla 4.6”.

**Tabla 4.6**

*Número de variables con mejores métricas*

<i>mtry</i>	<i>Precisión</i>	<i>Kappa</i>	<i>SD Precisión</i>	<i>SD Kappa</i>
2	0,9626	0,0000	0,0002	0,0000
<b>20</b>	<b>0,9683</b>	<b>0,5099</b>	<b>0,0035</b>	<b>0,0545</b>
38	0,9675	0,5052	0,0034	0,0523

Nota: 20 variables dan la mejor Precisión y Kappa

Luego de realizar el preprocesamiento y consolidación de todas las variables por distrito obtenemos una matriz de 18 columnas y 10 filas como se grafica en la “Tabla 4.7”, con lo que se produce el efecto “frontera” del rango de las variables predictoras debido a que tienen pocos datos, por lo que los errores en la predicción se van a incrementar, lo que se denomina *la maldición de la dimensionalidad* (Bellman, 1961).

Este efecto de dimensionalidad está relacionado con el efecto vecindario, pues al aumentar las dimensiones (mayor número de variable predictoras) todos los datos estarán en un mismo vecindario más grande, lo que lleva a una pérdida de flexibilidad de los modelos.

**Tabla 4.7**

*Matriz consolidada 10 filas y 18 columnas*

Distrito	RPIS	Privado	< 20 años	20 a 64 años	65 o mas años	Hombre	Mujer	letalidad COVID-19	Comorb_NO	Comorb_SI	Ambulatorio	Hospitalización	UCI	med_10000 hab	enf_10000 hab	p_hog	T_IS_Dx	T_IS_At
09D01	2740	1734	161	3826	487	2298	2176	0,04068	4389	85	4006	330	138	29,1445	14,8713	3,4496	6	2
09D02	808	57	10	793	62	419	446	0,01040	856	9	836	28	1	30,4165	27,9831	3,4760	14	6
09D03	2180	1085	100	2718	447	1708	1557	0,03277	3200	65	2973	212	80	43,2686	26,4940	3,1088	7	3
09D04	2477	451	72	2479	377	1445	1483	0,05362	2846	82	2525	282	121	14,7650	9,8268	3,4218	10	5
09D05	3863	1486	240	4445	664	2775	2574	0,02954	5019	330	4502	755	92	61,7162	33,5720	2,9819	6	3
09D06	511	0	44	406	61	252	259	0,00587	510	1	509	1	1	24,5687	18,7567	3,4151	6	5
09D07	877	389	66	1099	101	651	615	0,02528	1216	50	1194	56	16	15,1128	7,7597	3,4445	5	3
09D08	499	0	18	377	104	275	224	0,13828	343	156	216	281	2	6,1921	1,5752	3,5358	14	7
09D09	382	0	33	296	53	220	162	0,02356	354	28	382	0	0	6,9349	0,9632	3,3276	9	5
09D10	322	45	20	305	42	196	171	0,04905	356	11	330	34	3	6,1478	0,9315	3,3425	9	4

Para disminuir este efecto vamos a trasponer la matriz obtenida para obtener una matriz de 18 filas y 10 columnas como se grafica en la “Tabla 4.8”, con lo que teóricamente disminuirá, la maldición de la dimensionalidad.

Esta matriz traspuesta tendrá como variables los distritos de la ciudad de Guayaquil y como datos la información demográfica, de tipo de atención, lugar de atención, letalidad, información socioeconómica, recursos humanos sanitarios y tiempos de atención.

**Tabla 4.8***Matriz traspuesta consolidada 18 filas y 10 columnas*

<i>ID</i>	<i>09D01</i>	<i>09D02</i>	<i>09D03</i>	<i>09D04</i>	<i>09D05</i>	<i>09D06</i>	<i>09D07</i>	<i>09D08</i>	<i>09D09</i>	<i>09D10</i>
RPIS	2740	808	2180	2477	3863	511	877	499	382	322
Privado	1734	57	1085	451	1486	0	389	0	0	45
< 20 años	161	10	100	72	240	44	66	18	33	20
20 a 64 años	3826	793	2718	2479	4445	406	1099	377	296	305
65 o más años	487	62	447	377	664	61	101	104	53	42
C_Hombre	2298	419	1708	1445	2775	252	651	275	220	196
C_Mujer	2176	446	1557	1483	2574	259	615	224	162	171
letalidad COVID-19	0,041	0,010	0,033	0,054	0,030	0,006	0,025	0,138	0,024	0,049
Comorb_NO	4389	856	3200	2846	5019	510	1216	343	354	356
Comorb_SI	85	9	65	82	330	1	50	156	28	11
Ambulatorio	4006	836	2973	2525	4502	509	1194	216	382	330
Hospitalización	330	28	212	282	755	1	56	281	0	34
UCI	138	1	80	121	92	1	16	2	0	3
med_10000 hab.	29,14	30,42	43,27	14,76	61,72	24,57	15,11	6,19	6,93	6,15
enf_10000 hab.	14,871	27,983	26,494	9,827	33,572	18,757	7,760	1,575	0,963	0,931
pob_serv_b	0,980	0,983	0,989	0,967	0,994	0,983	0,982	0,888	0,973	0,603
T_IS_Dx	6	14	7	10	6	6	5	14	9	9
T_IS_At	2	6	3	5	3	5	3	7	5	4

## **4.2. Elaboración y Análisis Multivariante con el Método HJ-Biplot.**

### **4.2.1. Paquetes de R.**

Se utilizó *MultBiplotR* que es un paquete de Análisis Multivariantes usando Biplot para el software R, que incluye Biplot clásicos, MANOVA Biplot, Análisis de Correspondencia, Biplots Logístico, HJ-Biplot entre otros.

Este paquete fue programado primero en MATLAB y luego migrado a R, por su autor José Vicente Villardón (Vicente Villardon, 2010)

Del paquete principal se utilizó principalmente el “HJ Biplot with added features” para el análisis y generar gráficos.

#### 4.2.2. Generación del Código.

El código con el que se realizó el análisis y se generaron los HJ-Biplot, las bases de datos, los scripts y resultados los encuentra en “Anexo A – link A5”

#### 4.2.3. Resultados del Código del HJ-Biplot.

Los resultados generales del HJ-Biplot de la data analizada se resume en “Tabla 4.9”

**Tabla 4.9**

*Resultados generales del HJ-Biplot*

Eigenvalues & Inercia			
Dim	Eigenvalue	Inercia	Acumulado
[1,]	171771857,9	88,290	88,290
[2,]	21757582,0	11.183	99.473
[3,]	828755,3	0,426	99.899

**Tabla 4.10**

*Contribución relativa de las columnas*

CONTRIBUCIÓN DE LAS COLUMNAS			
Columnas	Dim 1	Dim 2	Dim 3
09D01	99,28	0,12	0,59
09D02	16,01	83,5	0,12
09D03	98,97	0,75	0,22
09D04	96,04	2,51	1,24
09D05	99,08	0,71	0,18
09D06	0,15	99,7	0,00
09D07	46,38	52,73	0,80
09D08	0,94	96,08	0,82
09D09	0,90	98,97	0,06
09D10	1,30	98,39	0,24

En la “Tabla 4.9” identifica que la dimensión 1 aporta con la mayor parte de la Inercia con un 88,29%, mientras que la dimensión 2 aporta solamente con 11,18% de la inercia; con lo que a simple vista se ven los efectos de la maldición de la dimensionalidad.

En la “Tabla 4.10” se explica la contribución de las columnas a la dimensión 1 siendo las que más contribuyen son los distritos 09D01, 09D03, 09D04 y 09D05; esto se explica por la fuerte correlación entre ellas, es decir son similares en sus características sociodemográficas, de accesibilidad a servicios de salud tanto de la RPIS como de la Red Complementaria.

En la dimensión 2 se encuentran los distritos 09D02, 09D06, 09D08, 09D09, 09D10 y 09D07, de igual manera comparten dificultades de acceso a servicios de salud, así como deficiencia de acceso a servicios básicos y pocos recursos humanos sanitarios.

Las dimensiones se ven mejor reflejadas en la “Figura 4.3”

**Tabla 4.11**

*Contribución relativa de las filas*

CONTRIBUCIÓN DE LAS FILAS			
<b>Filas</b>	<b>Dim 1</b>	<b>Dim 2</b>	<b>Dim 3</b>
RPIS	98,15	0,16	1,68
Privado	48,14	43,50	7,95
< 20 años	28,92	70,91	0,00
20 a 64 años	99,61	0,37	0,00
65 o más años	1,12	97,99	0,31
C_Hombre	94,61	5,25	0,00
C_Mujer	93,44	6,50	0,03
letalidad COVID-19	41,90	57,96	0,00
Comorb_NO	99,86	0,12	0,02
Comorb_SI	28,82	69,74	0,77
Ate_Amb	99,65	0,19	0,10
Hospitalización	4,59	88,46	3,51
UCI	32,77	66,70	0,00
med_10000 hab.	39,42	60,41	0,01
enf_10000 hab.	40,62	59,19	0,01
pob_serv_b	41,89	57,97	0,00
T_IS_Dx	42,16	57,71	0,01
T_IS_At	42,08	57,78	0,00

El aporte de las filas que se muestra en la “Tabla 4.11”, nos indica que en la dimensión 1, los aportes se pueden dividir en:

- Sitio de atención, RPIS (98,15%) y Privado (48,14%)
- Grupo etario de 20 a 64 años (99,61%)
- Sin comorbilidad (99,86%)
- Atención ambulatoria (99,65%)

El aporte de las filas en la dimensión 2 se puede dividir en:

- Grupos etarios, < 20 años (70,91%) y 65 o más años (97,99%).
- Internación hospitalaria, UCI (66,70%) y Hospitalización (86,64%).
- Recursos humanos sanitarios, médicos por 10000 hab. (60,41%) y enfermeras por 10000 hab. (59,19%).
- Variable sociodemográfica población con servicios básicos (57,97%).
- Variable de tiempo entre inicio de síntomas y atención médica (57,78%).
- Variable de tiempo entre inicio de síntomas y diagnostico por RT-PCR (57,71%).

#### **4.2.4. Gráficos HJ-Biplot.**

En la “Figura 4.4” se graficaron los cluster de los individuos dando cinco cluster muy bien identificados.

Cluster 1 conformado por los pacientes sin comorbilidad, el grupo etario de 20 a 64 años y las atenciones ambulatorias.

Cluster 2 conformado solo por la Red Pública Integral de Salud.

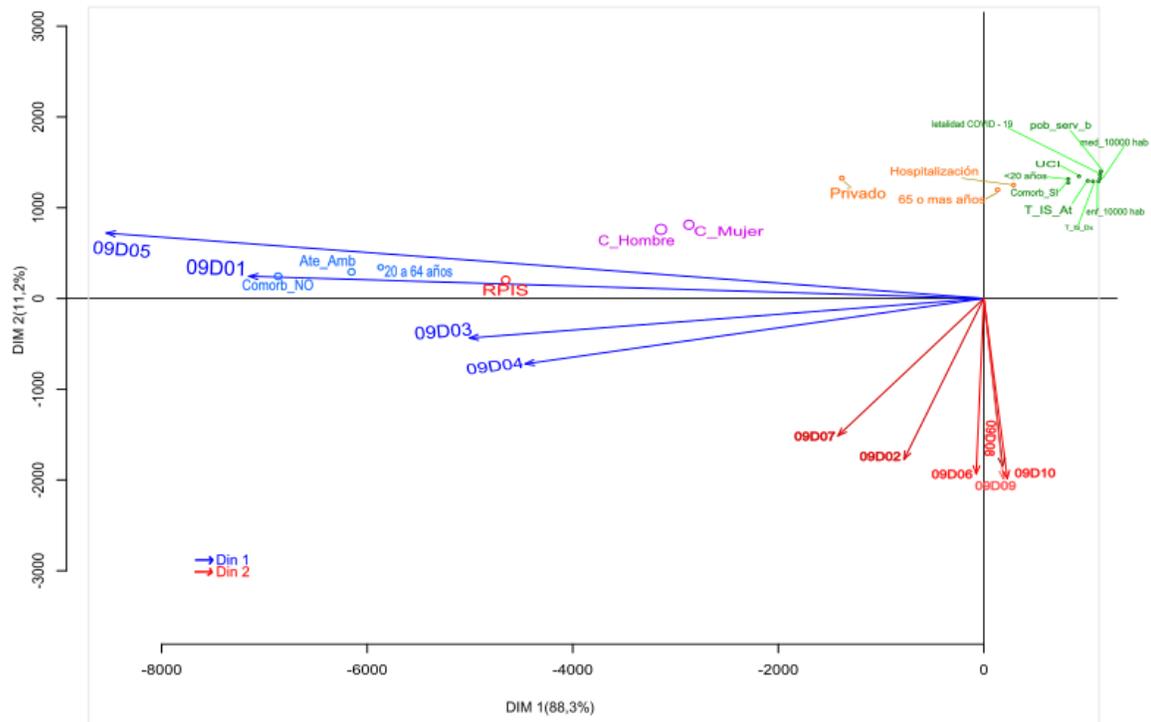
Cluster 3 conformado por Mujeres y Hombres.

Cluster 4 Conformado por el grupo etario de 65 o más años, Hospitalizados y Atenciones en la Red Privada

Cluster 5 Conformado por el grupo etario < 20 años, atención en UCI, recursos humanos sanitarios, porcentaje de servicios básicos por vivienda y los tiempos de atención.

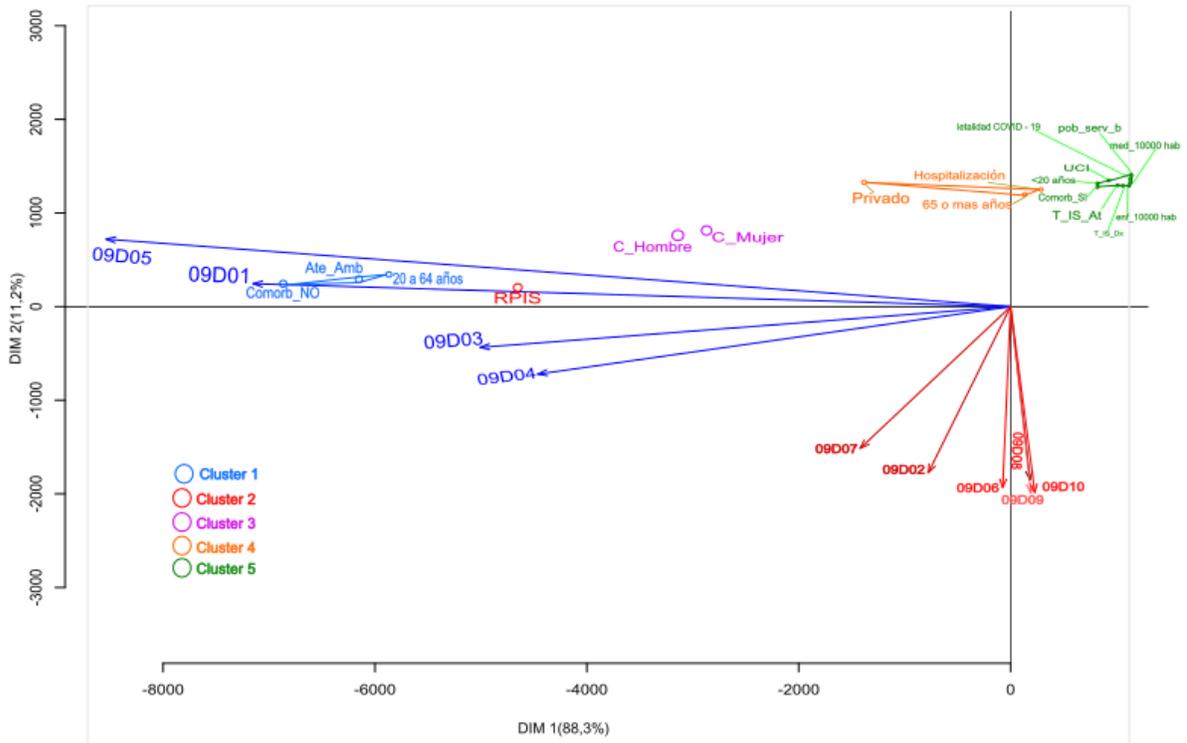
**Figura 4.3**

*Variables en relación con las dimensiones del HJ-Biplot de COVID-19 - 2020*



**Figura 4.4**

*Cluster de individuos del HJ-Biplot de los pacientes COVID-19 año 2020*



## CAPÍTULO 5

### 5. CONCLUSIONES Y RECOMENDACIONES.

#### 5.1. Conclusiones.

Una vez realizado todos los preprocesamiento y procesamiento final, se puede llegar a las siguientes conclusiones:

- Para obtener datos socioeconómicos, demográficos y de servicios de salud se tuvo que acceder a diferentes fuentes de información tanto del MSP como del INEC.
- Esta data está sin consolidar por Distritos como lo exige SENPLADES (SENPLADES S. N., 2012) y en especial el INEC utiliza en sus consolidados la división por parroquias, y para poder conseguir los datos por Distritos se utilizó software de Sistemas de información Geográfica para poder individualizar dicha información en Distritos.
- Luego de toda la extracción de información de las BD geográficas del INEC 2020 (INEC I. N., 2024) se utilizó el método de estadística robusta DDDC para poder identificar valores anómalos o faltantes, y poderlos imputar.
- Ya con una matriz con todos los pacientes con diagnóstico de COVID-19 del año 2020 y las variables sociales, recursos humanos sanitarios y de tiempo de atención, se realizó un Random Forest con validación K-fold, utilizando la como variable objetivo “egreso (vivo o muerto)”, lo que nos dio 20 variables que eran significativas para el modelo “Tabla 4.5”.
- Luego de todo este preprocesamiento obtuvimos una BD con 10 filas y 18 columnas, siendo las filas los distritos y las columnas, las variables a utilizar “Tabla 4.7”, la cual no podíamos analizar por cumplir todos los criterios de la “maldición de la dimensionalidad”, es decir pocos individuos y muchas variables.
- Para poder subsanar el problema de la “maldición de la dimensionalidad” se transpuso la matriz original, dando como una matriz de 18 filas y 10 columnas (distritos), a la cual se le aplicó el Método Multivariante HJ-Biplot, dando los siguientes resultados:

- Los distritos que más aportan a la dimensión 1 (88,29%) son los distritos 09D01, 09D03, 09D04 y 09D05, con lo que confirma que por la longitud de sus vectores en el HJ-Biplot “Figura 4.3” son los que más variabilidad tienen, y comprueba que los factores socioeconómicos y demográficos influyeron en la evolución de la pandemia de COVID-19; además estos cuatro distritos están altamente correlacionado con los individuos de los grupos etarios de 20 a 64 años, que recibieron atención ambulatoria (menor tiempo de espera para el diagnóstico), sin comorbilidades y acceso a la RPIS.
- Por otro lado, en la dimensión 2 (11,18%) están los distritos 09D02, 09D07, 09D08, 09D09 y 09D10 con menos variabilidad y que son la imagen opuesta de los distritos que conforman la dimensión 1; por tanto, tienen un menor acceso a la RPIS, menos recursos humanos sanitarios, servicios básicos de nivel medio a bajo y tiempos de espera mayores para el diagnóstico.
- Finalmente se concluyen que existen dos grupos de distritos claramente diferenciados, el primero (dimensión 1) con mayor acceso a los servicios sanitarios de la RPIS y Privada que permitió diagnósticos oportunos que influyo mucho en los niveles de Hospitalización y por ende disminuyo el riesgo de letalidad. El segundo grupo de distritos (dimensión 2) con bajo acceso a servicios sanitarios que conlleva a falta de atención o retardo de esta, que conlleva a una mayor letalidad y a un subregistro de los afectados y fallecidos por el COVID-19 en estos distritos.

## **5.2. Recomendaciones.**

### **5.2.1. Recomendaciones Para Replicar El Estudio.**

Las recomendaciones para la réplica del estudio son las siguientes:

- Se debe ampliar la base de los distritos del estudio incluyendo todos los distritos de la Zona 8 y Zona 5, lo que daría un numero de 37 individuos (distritos) con las mismas 18 variables.

- Solicitar al INEC las BD de los recursos y actividades de salud del año del estudio, en formatos SIG, para facilitar la selección de la información por distrito.
- Tener acceso a hardware de alta potencia de cálculo computacional para poder realizar análisis de valores anómalos o perdidos.

### **5.2.2. Recomendaciones de Salud Pública.**

Las recomendaciones para el Sistema Nacional de Salud serían las siguientes:

- En este tipo de pandemias lo principal es el acceso oportuno al sistema sanitario para un diagnóstico oportuno, que conlleva a menores niveles de hospitalización y letalidad.
- Las inequidades del acceso a los servicios de salud es algo que se planifica a largo plazo, por tanto, se deben tomar medidas de contingencia similares a los implementados en los desastres naturales. Como por ejemplo centros ambulatorio para toma de pruebas, unidades clínicas móviles y Hospitales de campaña.
- Recomendar medidas de mitigación de esos factores sociales que inciden en la evolución de estas crisis sanitarias, como por ejemplo paquetes de ayuda alimentaria y paquetes de prendas bioseguridad, para las familias de bajos recursos.
- Y finalmente implementar una vigilancia epidemiológica centinela con análisis estadísticos avanzados, para detectar brotes de nuevos virus con potencial pandémico.

## 6. REFERENCIAS.

Barrera-Núñez, D. T.-I.-M.-G.-C. (15 de Agosto de 2022). Revisión rápida de la transmisión del SARS-CoV-2 por contacto con objetos y superficies. *Salud Pública de México*, 63(1), 126-135. <https://doi.org/https://doi.org/10.21149/11877>

Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. New Jersey: Princeton University Press. <https://books.google.com.ec/books?id=POAmAAAAMAAJ>

Benzécri, J. (1973). L'Analyse des Correspondances. *L'Analyse des Données, II*.

Breiman, L. (2001). Random Forest. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Breiman, L. e. (23 de Mayo de 2022). <https://www.stat.berkeley.edu>. <https://doi.org/DOI:10.1023/A:1010933404324>

Choi, A. J. (09 de Mayo de 2022). A Retrospective Global Assessment of Factors Associated With COVID-19 Policies and Health Outcomes. *Frontiers in public health*, 10(843445), 1-10. <https://doi.org/10.3389/fpubh.2022.843445>

COE, G. d. (17 de Marzo de 2020). Informe de Situación No009 COE-Nacional. Quito, Plinchincha, Ecuador. <https://www.gestionderiesgos.gob.ec/wp-content/uploads/2020/03/Informe-de-Situaci%C3%B3n-No009-Casos-Coronavirus-Ecuador.pdf>

COE, S. N. (31 de Marzo de 2020). Situación Nacional del COVID 19 Infografía No 35. Quito, Plinchincha, Ecuador. <https://www.gestionderiesgos.gob.ec/wp-content/uploads/2020/03/INFOGRAFIA-NACIONALCOVI-19-COE-NACIONAL-31032020-17h00.pdf>

- Febrianti T, S. N. (31 de Mayo de 2023). Determinant Factors of Mortality in Pre-elderly and Elderly Patients With COVID-19 in Jakarta, Indonesia. *J Prev Med Public Health.*, 56(3), 231-237. <https://doi.org/10.3961/jpmph.23.008>.
- Fernández-Martínez, N. R.-M.-B. (12 de Diciembre de 2022). Socioeconomic differences in COVID-19 infection, hospitalisation and mortality in urban areas in a region in the South of Europe. *BMC Public Health*, 22(1), 1-10. <https://doi.org/10.1186/s12889-022-14774-6>
- Gabriel K. R. (Diciembre de 1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, 58(3), 453-467.  
<https://doi.org/https://doi.org/10.2307/2334381>
- GALINDO VILLARDÓN, M. P. (07 de Abril de 1986). Una alternativa de representación simultánea: HJ-Biplot. *Questiio, Quaderns d'Estadística, Sistemes, Informatica i Investigació Operativa*, 10(1), 13-23.
- García, C. (29 de Marzo de 2021). Socioeconomic, demographic and healthcare determinants of the COVID-19 pandemic: an ecological study of Spain. *BMC Public Health*, 21(1), 606. <https://doi.org/10.1186/s12889-021-10658-3>
- Halder, S. (04 de Agosto de 2023). Global health crisis index (GHCI): searching country-wise scenarios during the COVID-19 pandemic. *Journal of Public Health*, 32(6), 1-10.  
<https://doi.org/10.1007/s10389-023-02045-6>
- Hashim, M. J. (3 de Septiembre de 2020). Population Risk Factors for COVID-19 Mortality in 93 Countries. *Journal of epidemiology and global health*, 10(3), 203-208.  
<https://doi.org/https://doi.org/10.2991/jegh.k.200721.001>
- INEC, C. E. (2024). *Censo Ecuador*. (INEC, Productor) Retrieved 10 de Marzo de 2024, from <https://www.censoecuador.gob.ec/#>

- INEC, I. N. (2024). *ecuadorencifras.gob.ec*. Retrieved 23 de 04 de 2024, from <https://www.ecuadorencifras.gob.ec/actividades-y-recursos-de-salud/>
- Medina Hernández EJ, M. O. (2022). *Análisis multidimensional de la evolución de la pandemia de la COVID-19 en los países de las Americas*. PAHO. Washington: Rev Panam Salud Publica. <https://doi.org/https://doi.org/10.26633/RPSP.2022.49>
- MetroEcuador. (19 de Mayo de 2020). Cynthia Viteri entrega a Guayaquil el Hospital Municipal Bicentenario. Guayaquil, Guayas, Ecuador. Retrieved 01 de Abril de 2023, from <https://www.metroecuador.com.ec/ec/noticias/2020/05/19/cynthia-viteri-entrega-guayaquil-hospital-municipal-bicentenario.html>
- MSP. (29 de Septiembre de 2022). <https://www.salud.gob.ec>. [https://www.salud.gob.ec:https://www.salud.gob.ec/wp-content/uploads/2023/04/AC-00027-2022-SEP-29-Politica-de-tratamiento-de-datos\\_compressed.pdf](https://www.salud.gob.ec:https://www.salud.gob.ec/wp-content/uploads/2023/04/AC-00027-2022-SEP-29-Politica-de-tratamiento-de-datos_compressed.pdf)
- MSP, M. d. (01 de Marzo de 2018). Manual del Modelo de Atención Integral de Salud. *MAIS*. Quito, Pichincha, Ecuador: MSP.
- MSP, M. d. (17 de julio de 2021). *Sistema de Vigilancia Integrada de COVID y Otros Virus Respiratorios*. Retrieved 24 de 04 de 2024, from <https://sive.msp.gob.ec/Account/Login?ReturnUrl=%2F>
- OGC, O. G. (02 de Febrero de 2024). *GeoPackage*. (OGC, Editor) <https://www.geopackage.org/spec140/index.html>
- OMS, O. M. (27 de Abril de 2020). COVID-19: cronología de la actuación de la OMS. Ginebra, Lake Geneve, Suiza. <https://www.who.int/es/news/item/27-04-2020-who-timeline---covid-19>

PAHO, O. P. (24 de Abril de 2024). *Sistema Nacional de Vigilancia en Salud Publica Ecuador*.

(O. P. Salud, Productor) VIEPI: <https://viepi.msp.gob.ec/login.php>

Planificación, S. N. (28 de Mayo de 2012). <https://www.planificacion.gob.ec>. (SENPLADES, Ed.)

Retrieved 07 de Abril de 2024, from [https://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/05/REGISTRO-OFICIAL\\_DISTRITOS-Y-CIRCUITOS.pdf](https://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/05/REGISTRO-OFICIAL_DISTRITOS-Y-CIRCUITOS.pdf)

Planificación, S. N. (2024). *Secretaría Nacional de Planificación*.

<https://www.planificacion.gob.ec/6-codigos-de-identificacion-de-los-districtos-y-circuitos/>

Presidencia, S. G. (29 de Feb de 2020). Se registra el primer caso de coronavirus en Ecuador – Secretaría General de Comunicación de la Presidencia. Quito, Pichincha, Ecuador.

<https://www.comunicacion.gob.ec/se-registra-el-primer-caso-de-coronavirus-en-ecuador/>

Raymaekers, J. (25 de Octubre de 2023). cellWise 2.5.3. *Analyzing Data with Cellwise Outliers*.

Holanda, Paises Bajos. <https://cran.r-project.org/package=cellWise>

Rousseeuw, P. J. (03 de Abril de 2018). Detecting Deviating Data Cells. *Technometrics*, 60(2),

135-145. <https://doi.org/10.1080/00401706.2017.1340909>

SENPLADES. (09 de Abril de 2024). Mapa de distritos de ZONA 8. *Distritos Urbanos de*

*Guayaquil*. (V. E. Arias Galan, Recopilador) Guayaquil, Guayas, Ecuador.

SENPLADES, S. N. (16 de Febrero de 2012). Acuerdo Ministerial No 557-2012. 557-2012

*Confórmense 140 distritos administrativos de planificación*. Quito, Pichincha, Ecuador:

Registro Oficial. [https://www.planificacion.gob.ec/wp-](https://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/05/REGISTRO-OFICIAL_DISTRITOS-Y-CIRCUITOS.pdf)

[content/uploads/downloads/2013/05/REGISTRO-OFICIAL\\_DISTRITOS-Y-](https://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/05/REGISTRO-OFICIAL_DISTRITOS-Y-CIRCUITOS.pdf)

[CIRCUITOS.pdf](https://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/05/REGISTRO-OFICIAL_DISTRITOS-Y-CIRCUITOS.pdf)

Stewart, R. (1965). Formal planning: The staff planner's role at start up. *California: Stanford Research Institute*, 250.

Swallow Ben, X. W.-G. (02 de Marzo de 2021). Tracking the national and regional COVID-19 epidemic status in the UK using weighted principal component analysis. *Phil. Trans. R. Soc. A*, 380(2233), 1-15. <https://doi.org/doi:10.1098/rsta.2021.0302>

Vicente Villardon, J. L. (Enero de 2010). MULTBILOT: A package for multivariate analysis using biplots. Salamanca, España. <http://biplot.usal.es/ClassicalBiplot/index.html>

WHO, O. M. (31 de Julio de 2020). Retrieved 07 de Abril de 2024, from <https://www.who.int/publications/m/item/who-convened-global-study-of-the-origins-of-sars-cov-2>

WHO, O. M. (2020). *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. Reporte de investigación, Organización Mundial de la Salud. Retrieved 07 de Abril de 2023, from [https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19))

## 7. APÉNDICES

### 7.1. Base De Datos De Pacientes Con Diagnóstico De COVID-19 Del Año 2020 Con Domicilio En La Ciudad De Guayaquil, Debidamente Anonimizados.

Link: <https://espolec->

[my.sharepoint.com/:x/g/personal/serabauz\\_espolec/Ef8GZRaicYVER2YxIKfu4JUBd\\_WAfyqdrQQNDPdb8524PA?e=V3alqZ](https://espolec-my.sharepoint.com/:x/g/personal/serabauz_espolec/Ef8GZRaicYVER2YxIKfu4JUBd_WAfyqdrQQNDPdb8524PA?e=V3alqZ)

### 7.2. Sectores Anonimizados Del INEC Por Distritos De La Ciudad De Guayaquil En Formato Shapefile.

Link: <mailto:https://espolec->

[my.sharepoint.com/:f/g/personal/serabauz\\_espolec/EhSCHsUki5NDnzLh1pII\\_JIBk07\\_9qyyduVbV9gSR7jN-w?e=cfdkFe](https://espolec-my.sharepoint.com/:f/g/personal/serabauz_espolec/EhSCHsUki5NDnzLh1pII_JIBk07_9qyyduVbV9gSR7jN-w?e=cfdkFe)

### 7.3. Códigos Y Resultados Por Distrito De La Aplicación De La Técnica DDDC.

Link <https://espolec->

[my.sharepoint.com/:f/g/personal/serabauz\\_espolec/Es13jz2HoBPt6fISM5czq4BpP\\_SmPufxYViPrcGQBrqEg?e=AC45Rp](https://espolec-my.sharepoint.com/:f/g/personal/serabauz_espolec/Es13jz2HoBPt6fISM5czq4BpP_SmPufxYViPrcGQBrqEg?e=AC45Rp)

### 7.4. Códigos De La Técnica Random Forest Aplicado En La Base De Datos De Pacientes Con COVID-19, Con Los Resultados.

Link <https://espolec->

[my.sharepoint.com/:f/g/personal/serabauz\\_espolec/EkpaBI0o4A9GrE4hr410L9kB9T0UBWUzN86JqkHomqSevg?e=EPoqij](https://espolec-my.sharepoint.com/:f/g/personal/serabauz_espolec/EkpaBI0o4A9GrE4hr410L9kB9T0UBWUzN86JqkHomqSevg?e=EPoqij)

### 7.5. Códigos De La Técnica HJ-Biplot Aplicado A Los Datos Consolidados Y Sus Resultados.

Link: <https://espolec->

[my.sharepoint.com/:f/g/personal/serabauz\\_espolec/Em8CH9b886tBmtGz3esnTTUBgvqyzX-yEGcUZIX-tH-DGA?e=iGjUbL](https://espolec-my.sharepoint.com/:f/g/personal/serabauz_espolec/Em8CH9b886tBmtGz3esnTTUBgvqyzX-yEGcUZIX-tH-DGA?e=iGjUbL)