

DATA SCIENCE

Covid-19



**CODER HOUSE**

# ÍNDICE

## I. INTRODUCCIÓN

*Introducción y Necesidad del estudio* 3

*Metadata y Objetivos* 4

## II. MÉTODOS

*Plan analítico y de procedimiento* 5

## III ANALISIS EXPLORATORIO

*Hipótesis* 6

*Hipótesis 1* 7

*Hipótesis 2* 8

*Hipótesis 3* 10

## III. CONCLUSIONES

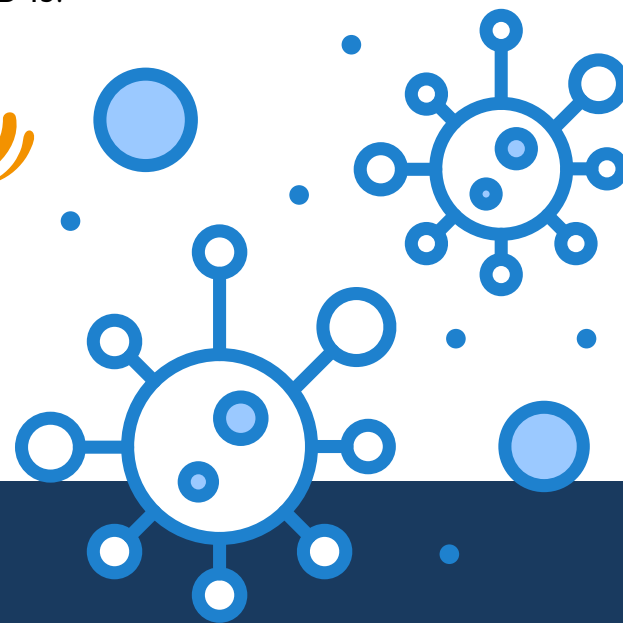
*Insight y Recomendaciones* 12

## INTRODUCCIÓN

La pandemia de COVID-19 ha tenido un impacto sin precedentes en todo el mundo. Desde su aparición en 2019, el virus se ha propagado rápidamente, afectando a millones de personas y causando una gran cantidad de muertes. La enfermedad es altamente contagiosa y puede presentarse en una variedad de formas, desde casos leves hasta casos graves y potencialmente mortales. Además, los pacientes con enfermedades preexistentes o comorbilidades tienen un mayor riesgo de sufrir complicaciones graves.

## NECESIDAD DEL ESTUDIO

En este contexto de pandemia, se propone un proyecto para abordar la pandemia de COVID-19. La necesidad de identificar los factores que estaban influyendo en la propagación del virus, como el historial médico del paciente, el cumplimiento de las órdenes de cuarentena. Utilizando esta información, pudieron desarrollar estrategias más efectivas y dirigidas para combatir el COVID-19, especialmente para las comunidades más afectadas por la pandemia. También se trabajó en colaboración con otros científicos de datos de todo el mundo, compartiendo datos y conocimientos para mejorar la respuesta global al COVID-19.



# METADATA

- 1.MEDICAL\_UNIT: saldo bancario; variable clave de interés
- 2.SEX: 1 para mujeres y 2 para hombres
- 3.PATIENT\_TYPE: tipo de atención que el paciente recibió en la unidad. 1 para regresar a casa y 2 para hospitalización.
- 4.INTUBED: si el paciente esta intubado.
- 5.PNEUMONIA: si el paciente ya tiene inflamación de los sacos de aire o no.
- 6.AGE: Edad del paciente.
- 7.PREGNANT: si la paciente está embarazada o no.
- 8.DIABETES: si el paciente tiene diabetes o no
- 9.COPD: Indica si el paciente tiene enfermedad pulmonar obstructiva crónica o no.
- 10.ASTHMA: si el paciente tiene enfermedad renal crónica o no.
- 11.INMSUPR: si el paciente tiene enfermedad renal crónica o no.
- 12.HIPERTENSION: si el paciente tiene hipertensión o no
- 13.OTHER\_DISEASE: si el paciente tiene enfermedad renal crónica o no.
- 14.CARDIOVASCULAR: si el paciente tiene una enfermedad relacionada con el corazón o los vasos sanguíneos.
- 15.OBESITY: si el paciente tiene enfermedad renal crónica o no.
- 16.RENAL\_CHRONIC: si el paciente tiene enfermedad renal crónica o no.
- 17.TOBACCO: si el paciente es consumidor de tabaco.
- 18.CLASIFFICATION\_FINAL: si el paciente estaba conectado al ventilador.
- 19.ICU: Indica si el paciente había sido ingresado en una Unidad de Cuidados Intensivos.



## OBJETIVO

El objetivo principal de este proyecto es construir un modelo de aprendizaje automático que, dado el síntoma, el estado y el historial médico actuales de un paciente con Covid-19, prediga si el paciente está en alto riesgo o no.

Se utilizarán técnicas de Machine Learning para entrenar un modelo capaz de predecir el riesgo de muerte de los pacientes con base en las variables de análisis mencionadas. Se evaluarán diferentes algoritmos de aprendizaje automático y se seleccionará el que proporcione mejores resultados.

## Mapa Analítico

### Contexto

La pandemia de COVID-19 ha afectado a millones de personas en todo el mundo y ha generado una gran cantidad de datos sobre la enfermedad y sus efectos en la salud de las personas. Este trabajo de Machine Learning se enfoca en analizar una base de datos sobre casos de COVID-19 obtenida de Kaggle.com, con el objetivo de determinar el riesgo de muerte de los pacientes.

### Objetivo

construir un modelo de aprendizaje automático que, dado el síntoma, el estado y el historial médico actuales de un paciente con Covid-19, prediga si el paciente está en alto riesgo o no.

### Herramientas

Base de datos  
Phython  
Google Colab  
API.

### Dirigido a

Personal Médico

### Contexto Comercial

Medicina

### Equipo Creativo

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua et.



## HIPÓTESIS

**¿QUÉ EDADES  
SON LAS MAS  
AFECTADAS?**

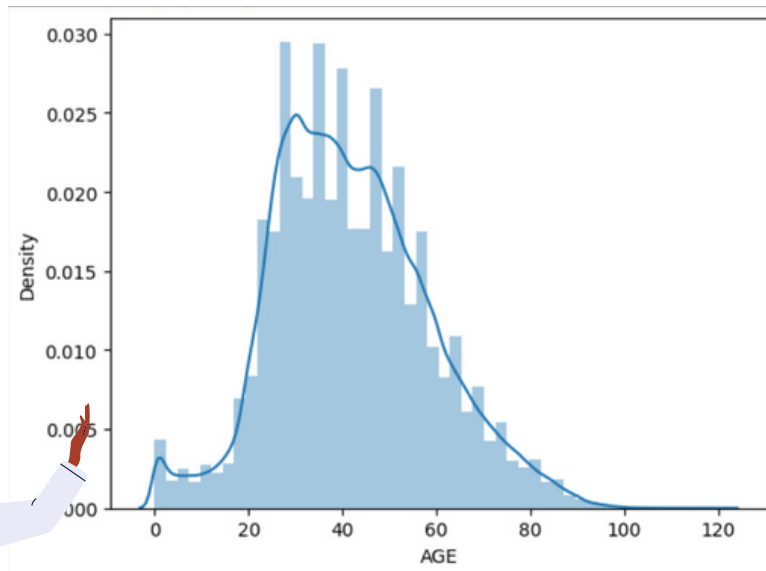
*¿El genero está relacionado con la afección  
de acuerdo a las edades estudiadas?*

**¿CUÁL EL  
PADECIMIENTO QUE  
MAYOR AFECTA AL  
PACIENTE QUE  
CONTRAE COVID?**

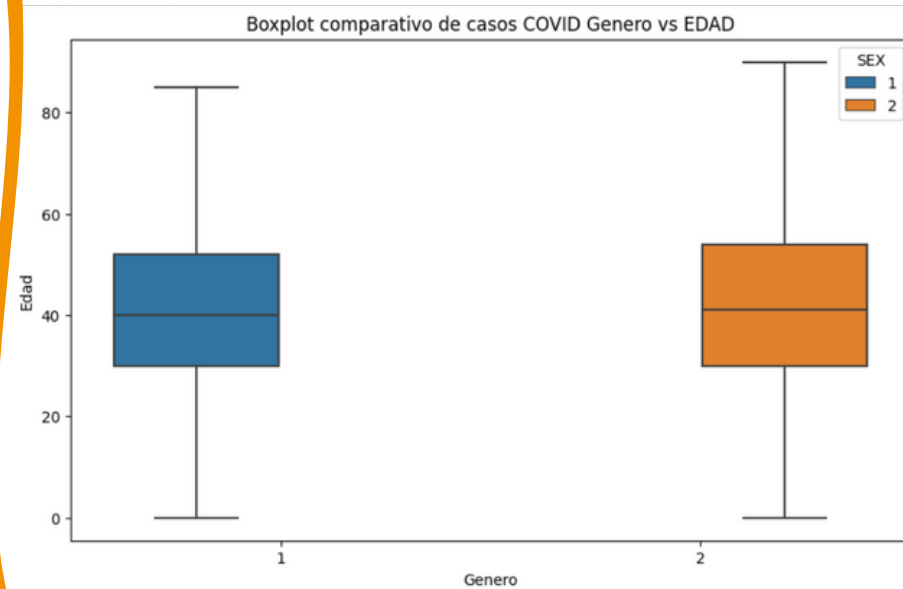
**¿CUÁL ES EL  
IMPACTO DE  
MUERTE EN  
PERSONAS  
EMBARAZADAS?**

## HIPÓTESIS 1

**¿QUÉ EDADES SON LAS MAS AFECTADAS?**

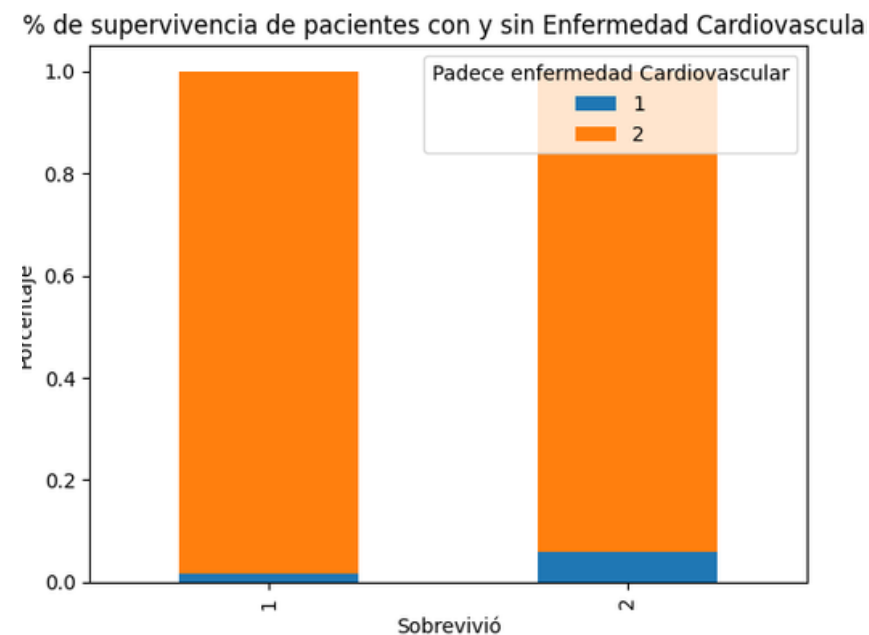
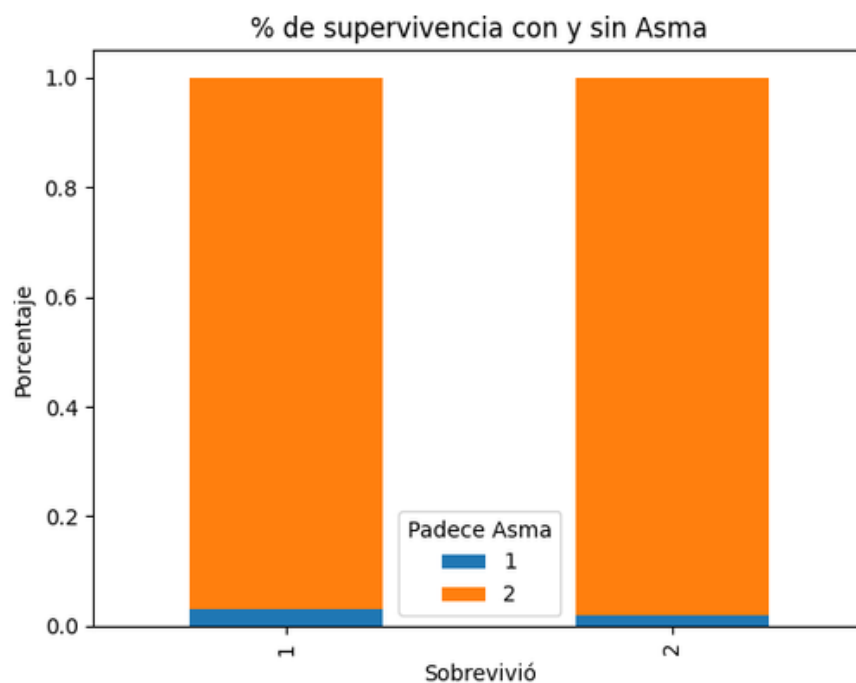


**¿EL GENERO ESTÁ REALACIONADO CON LA AFECCION DE ACUERDO A LAS EDADES ESTUDIADAS?**



## HIPÓTESIS 2

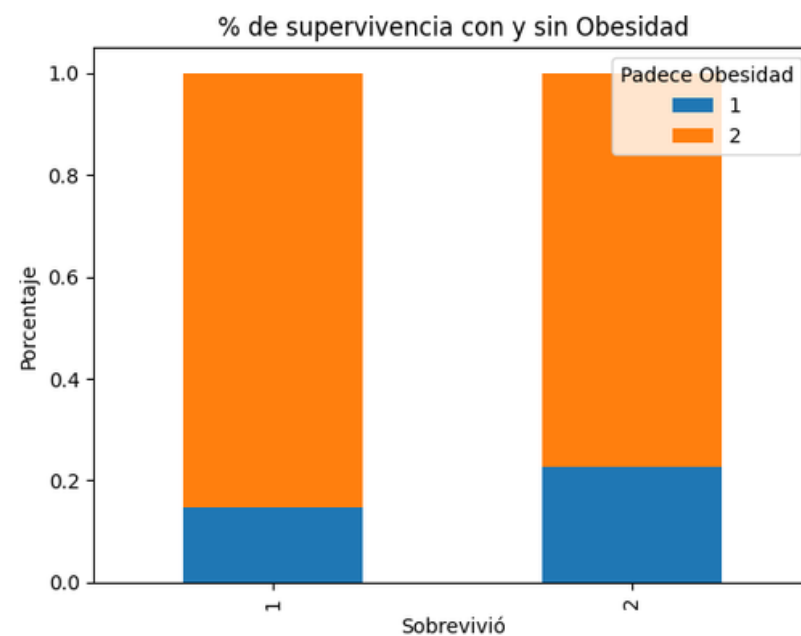
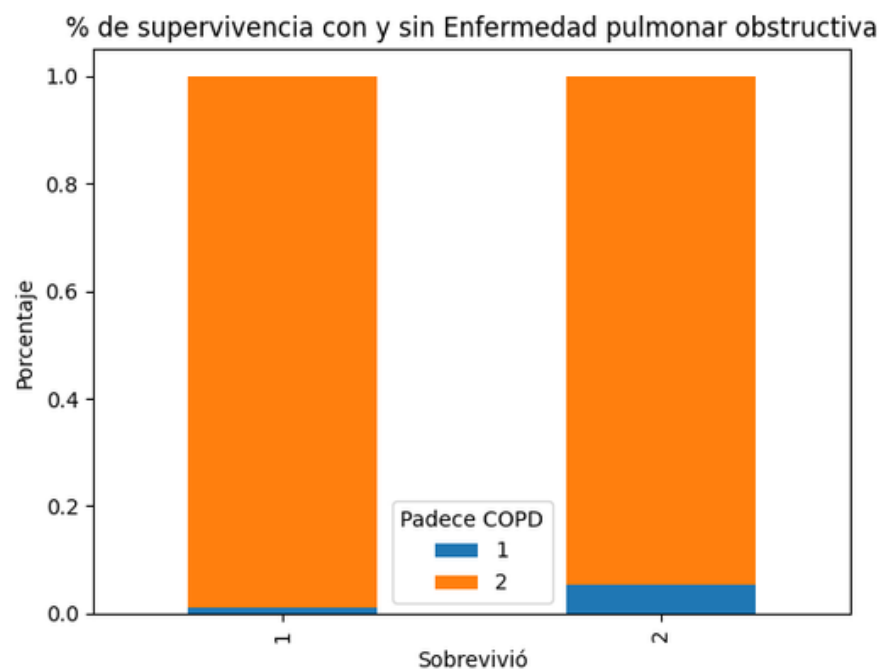
**¿CUÁL EL PADECIMIENTO QUE MAYOR AFECTA AL PACIENTE QUE CONTRAE COVID?**





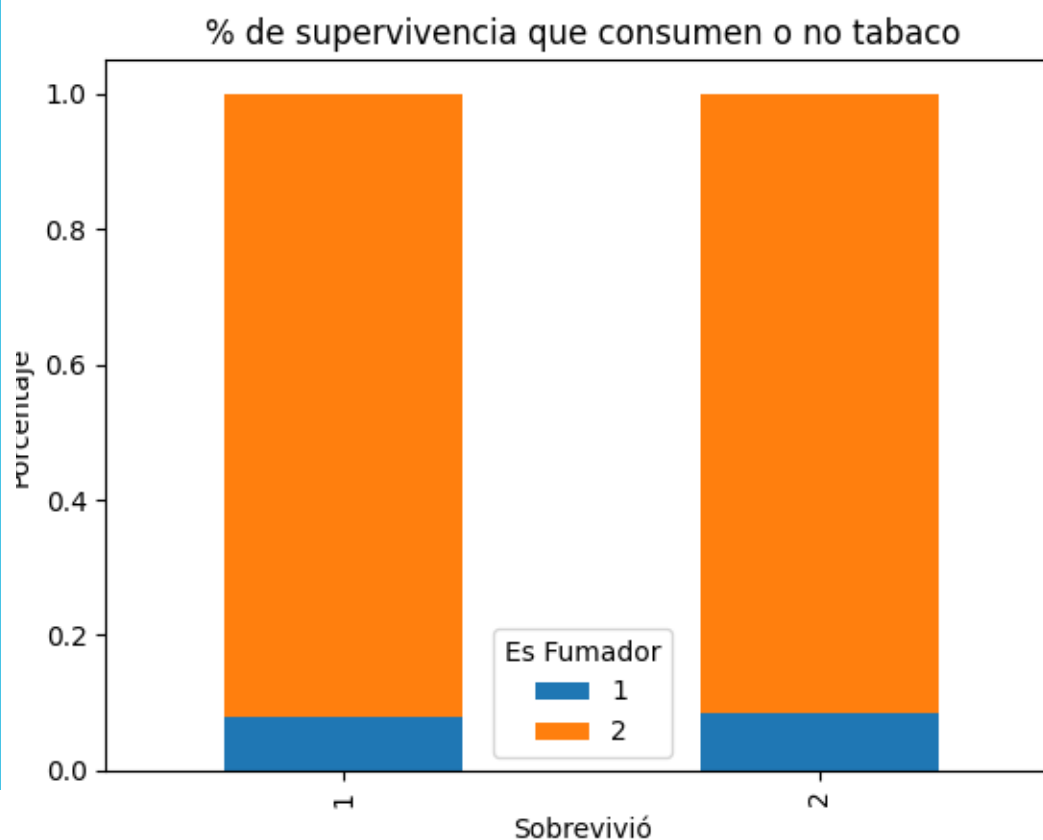
## HIPÓTESIS 2

**¿CUÁL EL PADECIMIENTO QUE MAYOR AFECTA AL PACIENTE QUE CONTRAE COVID?**



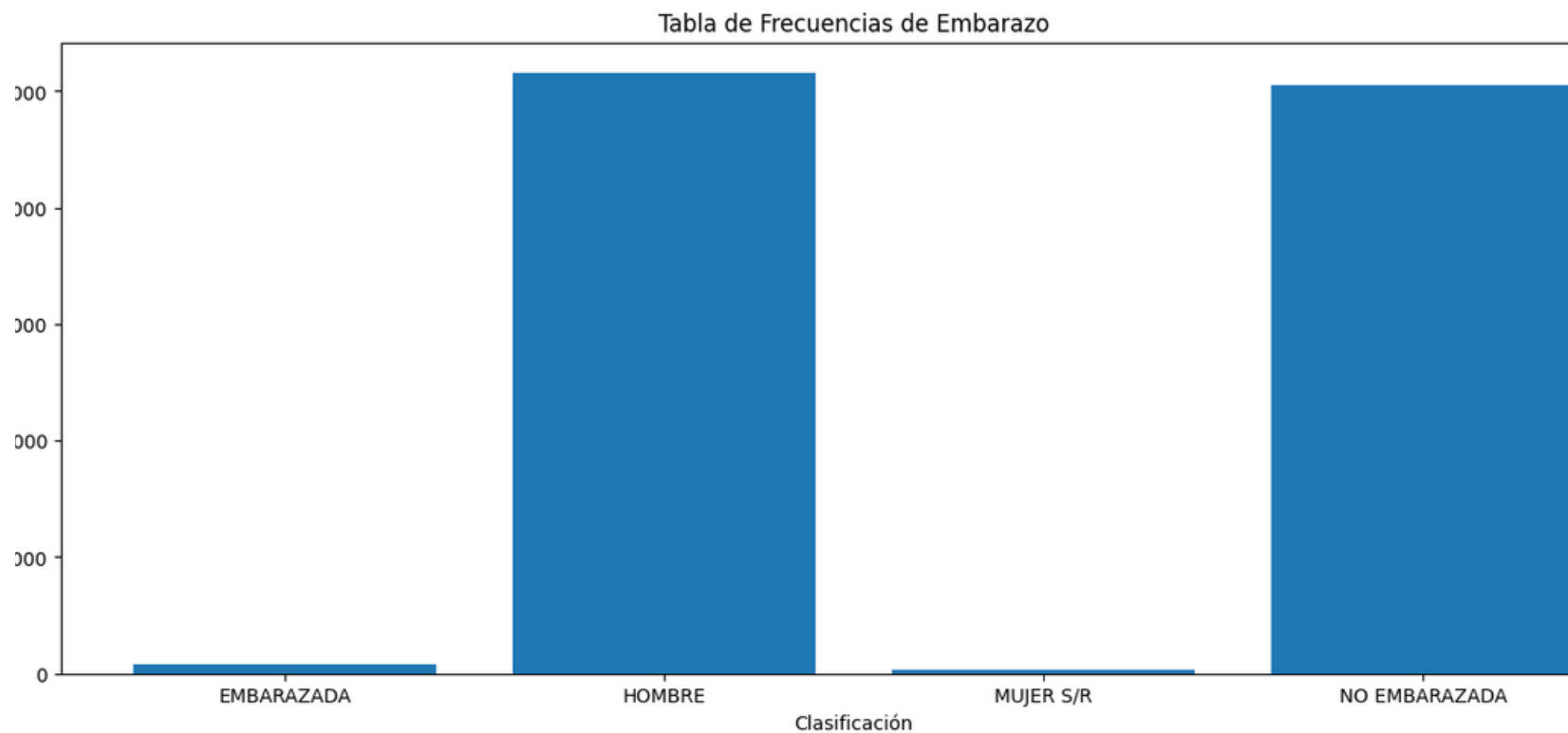
## HIPÓTESIS 2

***¿CUÁL EL PADECIMIENTO QUE MAYOR AFECTA AL PACIENTE QUE CONTRAE COVID?***



## HIPÓTESIS 3

***¿CUÁL ES EL IMPACTO DE MUERTE EN PERSONAS EMBARAZADAS?***



# INSIGHT

## HIPOTESIS 1

1. SE OBSERVA QUE LA MAYOR CANTIDAD DE N° DE CASOS POSITIVOS SE DIERON ENTRE PERSONAS DE 40 AÑOS APROXIMADAMENTE  
LA TENDENCIA NOS MUESTRA QUE MIENTRAS A MAYOR ES LA EDAD , MENOR ES LA CANTIDAD DE N° CASOS
2. LAS DISTRIBUCIONES CORRESPONDEN A LA VARIABLE EDAD Y CADA SERIE CORRESPONDE A UNA CATEGORÍA DE LA VARIABLE SEXO. PUEDE OBSERVARSE AQUÍ QUE LA DISTRIBUCIÓN ES MAYOR (ES DECIR, MAYOR DISPERSIÓN) PARA EL SEXO HOMBRES; EL QUE, POR OTRA PARTE, PRESENTA UNA EDAD PROMEDIO (MEDIANA) MENOR QUE EL DE MUJERES

## HIPOTESIS 2.

1. VEMOS CÓMO OTROS FACTORES COMO LA NEUMONÍA, HIPERTENSIÓN Y DIABETES INFLUENCIARON DE MANERA SIGNIFICATIVA LA SUPERVIVENCIA DEL PACIENTE.

## HIPOTESIS 3.

1. LA POBLACIÓN DE MUJERES EMBARAZADAS ES CASI INSIGNIFICANTE Y A SU VEZ COMPARABLE A LA CANTIDAD DE MUJERES QUE NO PUDIERON SER CATALOGADAS. AL NO SABER LA NATURALEZA DE PORQUE NO PUDIERON SER CATALOGADAS NO PODEMOS SUPONER QUE PROBABILIDAD HAY DE QUE SEAN EMBARAZADAS O NO POR LO TANTO PODRIAMOS ESPERAR QUE SEAN TODAS "EMBARAZADAS" O QUE SEAN TODAS "NO EMBARAZADAS".

## RECOMENDACIONES

1. HACER FOCO EN EL ANÁLISIS A PACIENTES DE TERCERA Y MEDIA EDAD, YA QUE SON PERSONAS DE RIESGO DE MORTALIDAD.
2. LA VARIABLE GENERO ES INDISTINTO, PERO SE REGISTRA UNA LEVE MORTALIDAD EN HOMBRES.
3. EL CONTAGIO DE UNA O MAS ENFERMEDADES PUEDE AUMENTAR EL RIESGO DE MORTALIDAD.
4. LA POBLACIÓN DE MUJERES EMBARAZADAS ES CASI INSIGNIFICANTE Y A SU VEZ COMPARABLE A LA CANTIDAD DE MUJERES QUE NO PUDIERON SER CATALOGADAS.
5. AL NO SABER LA NATURALEZA DE PORQUE NO PUDIERON SER CATALOGADAS NO PODEMOS SUPONER QUE PROBABILIDAD HAY DE QUE SEAN EMBARAZADAS O NO POR LO TANTO PODRIAMOS ESPERAR QUE SEAN TODAS "EMBARAZADAS" O QUE SEAN TODAS "NO EMBARAZADAS".
6. ANALIZAR LA RELACIÓN ENTRE ESTA VARIABLE Y LAS DEMÁS CON MÁS PROFUNDIDAD PARA DETERMINAR SI DEBE SER UTILIZADA O ELIMINADA DEL ANÁLISIS.

**LINK DEL PROYECTO:**

**NICOLASCAMINOS/DATA-SCIENCE-  
CODERHOUSE (GITHUB.COM)**