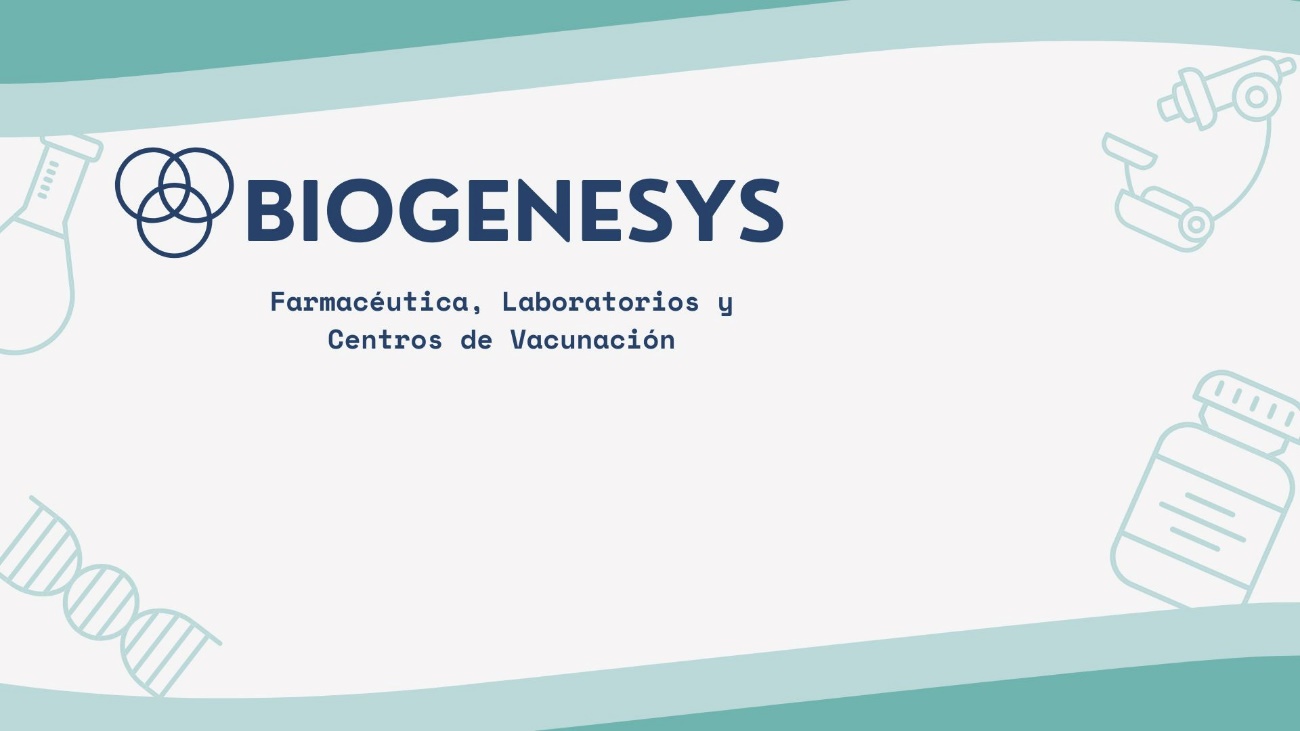
**Expansión Estratégica de BIOGENESYS**

**Nombre del autor:** Angela Viviana Gaviria Cepeda

**Email:** nanela02@hotmail.com

**Cohorte:** DA-FT10

**Fecha de entrega:** 2025-01-09

****

# 

# Introducción

Este informe debe permitir a la empresa farmacéutica BIOGENESYS identificar las ubicaciones óptimas para la expansión de laboratorios farmacéuticos y centros de vacunación, basándose en el análisis de datos de incidencia de COVID-19, tasas de vacunación, y la disponibilidad de infraestructuras sanitarias.

# Desarrollo del proyecto y EDA e insights

# 1.Carga de librerías

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import geopandas as gpd

2.Carga de archivo

Se implemento parquet parna minimizar el peso

3.Limpieza de datos

Verificamos que la cantidad de la información de los datos sea la correcta

\* columnas a trabajar

\* tipo de dato (si ay que cambiar el datatime)

4. filtrado

El área de ingeniería de datos nos pide enfocarnos en 6 países para la expansión propuesta para el análisis.

- Argentina

- Chile

- Colombia

- México

- Perú

- Brasil

5. Antes de proceder a filtrar verificamos como están escritos los nombres de los países a expandirse

6. Verificaremos la cantidad de datos faltantes por columna y si es necesario haremos una limpieza de los datos donde se presenten valores mayores a 6 millones de datos faltantes.

7. Después de haber hecho un análisis de la calidad de los datos, podemos ver que tenemos datos faltantes en muchas columnas, por lo que después realizaremos un nuevo filtro por la columna "Location\_key” pero solo vamos a utilizar las instancias que correspondan a los países que hemos seleccionado y no a sus regiones o subregiones AR, CL, CO, MX, PE, BR.

8. Filtra los datos en fechas mayores a 2021-01-01

Los datos relevantes para este análisis comprenden desde el año 2021

9. NaN

Con el nuevo filtro verificamos la cantidad de elementos vacíos.

Completar NaN por la media excepto en la columna new\_recovered, cumulative\_recovered, cumulative\_vaccine\_doses\_administered que se manejaran con el valor de cero

verificamos la cantidad de NaN y las dimensiones del nuevo dataset.

10. Guardamos el dataset en un archivo CSV

Guardamos el dataset filtrado por fecha en un archivo csv para poderlo utilizar en el análisis y que no sea necesario cargar el archivo original, recordemos que ahora con las nuevas dimensiones del dataset, el archivo es más pequeño y más fácil de manejar.

11. Inspección de la nueva data

Describimos la data con estadística descriptiva

Nos brinda información relevante de una forma resumida y entendible

12. Se crea una función que nos entregue valores que el método describe no nos entrega, como la moda, mediana, el Rango, la varianza.

13. avance 2

Acercamiento a las estadísticas básicas de los datos.

Usamos un ciclo for para recorrer las columnas y obtener la media, mediana, desviación estándar, mínimo y máximo de cada columna.

14. Análisis a lo largo del tiempo

15. Análisis de la relación entre las variables

Realizamos una matriz de correlación para ver si existe alguna relación entre las variables para tener en cuenta para un análisis posterior.

1. la matriz de correlación evidencia las variables con mayor correlación, para identificarlas más rápido, se definió un umbral cuyo valor absoluto sea mayor a 0.5 y se graficaron en un heatmap.

2. establecida las variables que tienen mayor correlación, nos ayuda en el análisis detallado posterior para lograr unas decisiones acertadas para la creación de un modelo ML.

3. Sin embargo podemos proceder a realizar una rápida revisión gráfica de la gran mayoría de las columnas para ver cómo se comportan los datos, para esta primera aproximación vamos a utilizar barplots utilizando un ciclo for y la librería seaborn.

16. Paleta de colores acorde a la empresa

17. Análisis Invariable

18. Análisis Bivariable

En esta parte nos enfocaremos en algunas variables específicas con las que podamos establecer algún tipo de relación directa

En este caso vamos a analizar la variable average\_temperature y la variable new\_confirmed, con el objetivo de entender los nuevos confirmados respecto a la temperatura.

Se evidencia que en algunos países la temperatura no influye en la cantidad de casos confirmados, sin embargo, en otros países si se puede observar una relación entre la temperatura y la cantidad de casos confirmados.

19 ahora realizaremos un análisis de la temperatura promedio con la cantidad de muertes.

La grafica anterior nos muestra que en algunos países la temperatura no influye en la cantidad de muertes, sin embargo, en otros países si se puede observar una relación entre la temperatura y la cantidad de muertes.

Procedemos a crear una matriz de correlación de las dos columnas anteriores para ver más a detalle su relación

20. Estudio longitudinal

Comportamiento de las variables a lo largo del tiempo.

Cambiamos la frecuencia temporal del DataFram y procedemos a calcular la media de los datos en cada nuevo intervalo muestreado.

#numeric\_only=True asegura que solo se consideren columnas numéricas para calcular la media, ignorando cualquier columna no numérica.

Promedio mensual, semanal

21. Nuevas columnas de Tiempo

Columna día semana, día mes y año

Procedemos a graficar concluyendo

El país que más casos confirmados ha tenido es Brasil, seguido de México, Argentina, Colombia, Chile y Perú. Esto no quiere decir que manejaran peor la pandemia, ya que hay que tener en cuenta la cantidad de habitantes de cada país además de otros factores y cantidad de datos que tenemos.

El país que más muertes ha tenido es Brasil, seguido de México, Colombia, Argentina, Perú y Chile. Sin embargo, esto no quiere decir que estos países sean los que peor han manejado la pandemia, ya que hay que tener en cuenta la cantidad de habitantes de cada país además de otros factores y cantidad de datos que tenemos.

El país que ha tenido más recuperados es Brasil, sin embargo, esto como lo dijimos viene relacionado a la cantidad de personas y vacunas disponibles.

Aunque chile tiene rangos diferentes de temperatura y eso podría haber afectado la cantidad de casos confirmados y muertes, se puede observar que chile ha manejado mejor la pandemia que los demás países.

Esto poco a poco nos ha llevado a concluir que los países donde se podría crear los nuevos centros serían donde hay más casos confirmados y muertes, ya que es donde se necesita más ayuda.

Debemos realizar un análisis detallado de los datos y su estudio de datos antes de tomar decisiones precipitadas

22. Distribución poblacional por grupos de edad

Procedamos a analizar la distribución de la población por grupos de edad en los países seleccionados, para tratar de entender cómo se comporta la población en cada país.

se evidencia la población por grupos de edad en cada país, lo que nos puede ayudar a entender cómo se comporta la población y también nos genera la idea de que podríamos graficar la cantidad de vacunas por cada grupo etario.

23. Grupo Etario más Vacunado

24. Mortalidad por Genero

Objetivo identificar la población más vulnerable en cada país, con esto ya podríamos tener una idea de cuál serían la población a vacunar primero.

La gráfica de mortalidad nos muestra que en todos los países la cantidad de muertes es mayor en los hombres que en las mujeres, lo cual nos da una ventana de análisis para evaluar qué factores provocan estos resultados

25.  Correlación entre posibles variables que pueden afectar a la población

Se pretende establecer si existe una correlación entre las variables.

26. De esta matriz podemos observar que el fumar y tener diabetes son las enfermedades preexistentes que más afectan a la población en los países seleccionados, por lo que se podría decir que estas enfermedades son las que más han causado muertes en los países seleccionados, sin embargo, esto no quiere decir que sean las únicas enfermedades que han causado muertes, sino que son las más comunes.

27. Mortalidad por países.

Analizaremos el comportamiento de la mortalidad en los países seleccionados por enfermedades preexistentes o comorbilidades.

Las dos enfermedades muestran un comportamiento similar en los países seleccionados, sin embargo, se puede observar que la diabetes es la enfermedad que más ha causado muertes en los países seleccionados, seguido del fumar.

Como conclusión de este avance podemos decir que los países que más necesitan ayuda son Brasil, México y Colombia, ya que son los países que más casos confirmados y muertes han tenido, además de ser los países que menos vacunas han administrado.

Las enfermedades preexistentes que más han afectado a la población en los países seleccionados son el fumar y la diabetes, por lo que se podría decir que estas enfermedades son las que más han causado muertes en los países seleccionados (Al ser enfermedad de base la probabilidad que una persona que ya tiene al menos una de las dos enfermedades muera es mayor).

28. Avance 3

En primera instancia se realizará un análisis del comportamiento los casos nuevos a lo largo del tiempo de manera semanal, mensual y anual.

29. Comportamiento Semanal, mensual y anual

Se evidencia que a lo largo del tiempo han ido disminuyendo los nuevos casos, es pertinente hacer un análisis más detallado para determinar que causas han propicios dicha disminución y no llegar a conclusiones precipitadas como por ejemplo que la pandemia se esté acabando.

Es por ello que queremos ver si la vacunación se asocia con la disminución de casos nuevos

30. El país con mayor dosis administradas es Brasil, seguido de México, Argentina, Colombia, Perú y Chile. Se contradice con los hallazgos anteriores sin embargo recordemos que se debe tener en cuenta la población de cada país y la cantidad de vacunas disponibles.

Comportamiento de la temperatura respecto a los nuevos casos por país

Podemos observar que la temperatura promedio ha ido aumentando y los casos confirmados han ido disminuyendo sin embargo hubo un pico antes del mes de marzo, pero disminuyó rápidamente, esto puede ser por la cantidad de vacunas administradas o por la cantidad de personas que se han vacunado.

31. Densidad poblacional

Cada país tiene un impacto debido a la urbanización y la densidad de población, por lo que vamos a analizar la densidad de población en cada país.

El país con mayor impacto en la propagación fue chile, sin embargo, esto puede ser debido a la densidad de población en chile.

32. Densidad de población por grupos de edad

El gráfico muestra claramente la desigualdad en cantidades de población por grupos de edad en los países seleccionados.

33. Vacunación

Análisis de las estrategias de vacunación empleadas para mitigar la mortalidad y los casos confirmados.

Chile es el país que mejor emplea estrategias para el cumplimiento de metas para la administración de vacunas, Brasil, Colombia y México deben evaluar la baja dosificación, ya sea si es por negación a la aplicación de vacunas o por la falta de laboratorios y centros de vacunación para mejorar el acceso a las vacunas.

34. Condiciones Preexistentes

Prevalencia de Condiciones Preexistentes en Países con Altas y Bajas Tasas de Mortalidad por COVID-19

35. Prevalencia de diabetes y Fumar por país

en casi todos los países exceptuando Peri la prevalencia de fumar es más alta por ende tiene una mayor tasa de mortalidad Ahora identificamos que la diabetes afecta mucho más a México y Brasil, por lo que se podría decir que estos países son los que más han sufrido decesos por la diabetes. Respecto a fumar se puede observar que afecta más a Chile y Argentina, por lo que se podría decir que estos países han tenido mayores decesos asociados a fumar.

36. Fatalidad por país

Los casos nuevos y las muestras nuevas han variado debido a la vacunación y a las medidas tomadas por los países, además de que se ha disminuido la cantidad de personas que se han contagiado y fallecido.

37. Análisis Temporal de la Mortalidad por COVID-19 por país

38 evolución de los casos activos y recuperados por país

39. Expansión de la empresa farmacéutica BIOGENESYS

Los países a expandirse son:

\*Brasil

\*México

\*Colombia

ya que son los países que más casos confirmados y muertes han tenido, además de ser los países que menos vacunas han administrado, tal vez las políticas de vacunación no han sido las mejores o no han tenido la cantidad de vacunas necesarias, esto sin tener en cuenta la densidad de población, el pub precipita, la cantidad de habitantes de cada país y problemas socioeconómicos que puedan afectar la cantidad de casos confirmados y muertes.

40. EXTRA CREDIT

Para un análisis más detallado graficaremos (mapas):

+ Casos confirmados

+ Cantidad de vacunas administradas

+ Muertes

+ Recuperados

# Análisis del dashboard

1. Logo de la empresa y pequeña descripción a que se dedican
2. Titulo de lo que se quiere visualizar
3. Un llamado a iniciar la navegación
4. Se creo una portada que contiene el menú con botones accionables que permiten moverse por todo el dashbord

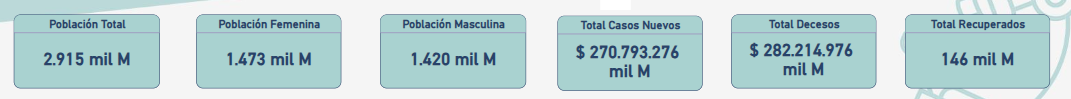


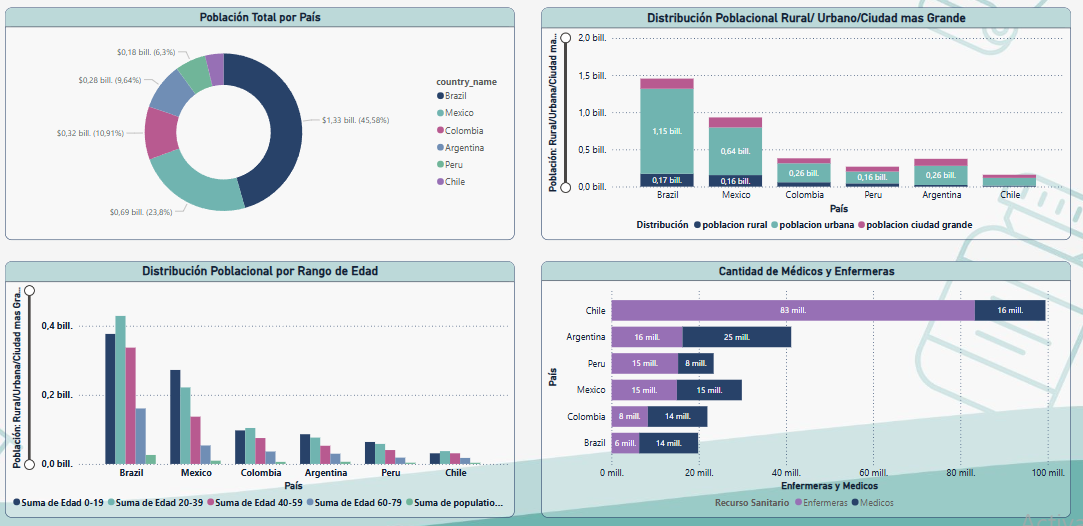
# Creación de filtros por año y país para que puedan elegir que visualizar y un botón que permite borrar todos los filtros que se pusieron además cada grafica permite hacer un filtro adicional, los filtros solo afectan a la página seleccionada

# 

Cada página tiene logo de la empresa, su respectivo título que indica en cual estamos y botones accionables que permiten moverse por todo el dashbord

# 

Cada pagina contiene tarjetas con información total relevante

Cada página contiene gráficos relevantes a lo que se quiere mostrar en cada página 

# Conclusiones y Recomendaciones

**Brasil y México (Prioridad Alta)** Con una alta densidad poblacional y elevados casos confirmados y mortalidad por el COVID-19, debe enfocar todos sus recursos para el fortalecimiento de su infraestructura sanitaria

**Argentina (Prioridad Media) A**unque la cantidad de médicos es alta, necesita mejorar la distribución de vacunas para tener un mayor cubrimiento de la población

**Perú y Colombia (Prioridad Media-Baja) R**equiere fortalecer la cantidad de médicos y enfermeras

**Chile (Prioridad baja) Cuenta** con buenos recursos sanitarios lo cual refleja una menor incidencia de los casos confirmados

Las ubicaciones óptimas para la expansión de laboratorios farmacéuticos y mejora del acceso a las vacunas, laboratorios y centros de vacunación se deben realizar en Brasil y México debido a sus bajas tasas de vacunación y de disponibilidad de infraestructuras sanitarias.

Deben ser la principal prioridad para la expansión, ya que ambos países no solo tienen la mayor cantidad de densidad poblacional de la región seleccionada, sino que también los niveles más altos de incidencia por el COVID-19 y de mortalidad.

Lo que nos lleva a concluir que requieren de una infraestructura sanitaria más potente para optimizar la respuesta a los efectos de la pandemia y post pandemia que permita mitigar los casos nuevos y bajar los índices de mortalidad.

# Reflexión personal

La importancia de tener una base sólida del conocimiento adquirido en Henry me permitió desenvolverme de la mejor forma al momento de aplicarlas las habilidades adquiridas como analista de datos

En el momento de hacer un código en Python y tener claro que se quiere obtener de esta (escogerlas estratégicamente) y saber cómo interpretar los resultados para dar una apropiada recomendación a la empresa. El tener el contacto desde cero de Python me permitió estar más abierta al conocimiento; tener la posibilidad de crear las gráficas desde cero fue muy enriquecedor en mi proceso de formación.

Si tuviera que volver a empezar este proyecto, no lo haría exactamente de la misma manera, aplicaría muchas cosas que debí mejorar en el transcurso mientras las aprendía.