**Carrera:Data Analytics**

**Modulo 4**



**Proyecto Integrador**

**Farmaceutica Biogenesis**

**Expansion**

**Nombre del autor: Yanina Liliana Soto**

**Email: yana\_mta\_2010@hotmail.com**

**Cohorte: DAFT\_06**

**Fecha de entrega: Domingo 22 de Septiembre del año 2024**

**Institución:**

La empresa farmacéutica BIOGENESYS busca identificar las ubicaciones óptimas para la expansión de laboratorios farmacéuticos, basándose en el análisis de datos de incidencia de COVID-19, tasas de vacunación, y la disponibilidad de infraestructuras sanitarias. La meta es optimizar la respuesta a los efectos de la pandemia y postpandemia con el fin de mejorar el acceso a las vacunas.

**Introducción:**

BIOGENESYS necesita la estrategia de expansión en Latinoamérica específicamente en Colombia, Argentina, Chile, México, Perú y Brasil. Este proyecto es vital para que estén preparados y puedan tener una respuesta rápida ante cualquier situación futura que pueda surgir, la directiva ha propuesto que para poder hacer esta inversión necesitan ubicar regiones y recolectar datos que sean de importancia y que les ayuden a tomar decisiones.

**Desarrollo del proyecto:**

El desarrollo del proyecto se realizo en cuatro partes;

En el PRIMER AVANCE se realizo la carga y transformacion de los datos,el dataset asignado de formato csv, contaba con un poco mas de 12 millones de filas y 50 columnas, se creo un notebook en python y se cargo el archivo

Se revisaron los conjuntos de datos en busca de inconsistencias, duplicados y valores nulos. Los datos incompletos fueron completados utilizando técnicas de interpolación o imputación cuando fue posible, se han realizado filtros y luego se ha guardado en dataset filtrado para su posterior uso en el AVANCE DOS,ya con los datos limpios, se continuo con las visualizaciones para profundizar en el analisis exploratorio para extraer insights valiosos que no puedan orientar en la planificacion, atraves de graficos seleccionados.

En el AVANCE TRES, realizamos un analisis mas profundo y sofisticado con el objetivo de pulir y preparar los datos para visualizaciones mas avanzadas que nos permitan identificar con precision las ubicaciones mas estrategicas de expansion de los laboratorios

Finalmente en el AVANCE CUATRO, realizamos la integracion con POWER BI, presentando los hallazgos analiticos utilizando sus visualizaciones, teniendo como objetivo sintetizar el analisis realizado en dashboards y reportes, integrando tambien graficos generados en python.

**Conclusión:**

En conclusion; los paises donde debe expandirse la empresa farmaceutica son, en primer lugar , Brazil ,Mexico y Colombia, ya que son los paises que mas casos confirmados y muertes han tenido, ademas de ser los paises que menos vacunas han administrado, quizas no se han manejado de la mejor manera en cuanto a la vacunacion o no han tenido la cantidad necesarias de vacunas,cual fuere el motivo, tras estos analisis de los datos, pudimos llegar a esta valiosa conclusion

**EDA e insights:**

Realizando diversos analisis , pudimos concluir los siguientes insights;

Se ha podido obtener visualizaciones dando a revelar que en algunos paises la temperatura no influye en la cantidad de casos confirmados ,sin embargo en otros paises si se puede observar una relacion entre la temperatura y la cantidad de casos confirmados de covid.

En algunos paises la temperatura no influye en la cantidad de muertes , sin embargo en otros paises si se puede observar una relacion entre la temperatura y la cantidad de muertes.

Se puede observar que la cantidad de dosis administradas en los paises seleccionados ha ido aumentado a lo largo del tiempo

Realizando un analisis mas detallado de las dosis administradas por mes en cada pais , podemos observar que en algunos paises la cantidad de dosis administradas ha ido aumentando a lo largo del tiempo, sin embargo en otros paises la cantidad de dosis administradas ha ido disminuyendo , esto puede ser debido a muchos factores, como la cantidad de vacunas disponibles,la cantidad de personas que se han vacunado,entre otros.

El pais que mas muertes ha tenido es Brasil, seguido de Mexico,Colombia,Argentina,Peru y Chile. Sin embargo, esto no quiere decir que estos paises sean los que peor han manejado la pandemia , ya que hay que tener en cuenta la cantidad de habitantes de cada pais ademas de otros factores y cantidad de datos que tenemos.

La disminucion de casos en Mexico puede ser una perdida de datos o una disminucion real de casos

El pais que mejor se recupero ha sido Brasil, sin embargo esto viene relacionado a la cantidad de personas y vacunas disponibles.

De estos paises, el que mejor manejo la pandemia fue Chile, seguido de Peru,Argentina,Colombia,Mexico y Brasil, se puede observar que la cantidad de muertes y casos confirmados es menor en estos paises ademas de no ser tan desproporcionada la cantidad de vacunas administradas.

De los paises que peor manejaron la pandemia fue Brasil, seguido de Mexico,Colombia,Argentina,Peru y Chile, se puede observar que la cantidad de muertes y casos confirmados es mayor en estos paises ademas de ser desproporcionada la cantidad de vacunas administradas.

Analizamos la distribucion de la poblacion por grupos de edad en los paises seleccionados,para tratar de entender como se comporta la poblacion en cada pais y en un futuro poder tomar decisiones mas acertadas.

Se ha podido observar que la cantidad de casos confirmados y muertes ha ido disminuyendo a lo largo del tiempo,sin embargo esto no quiere decir que la pandemia haya terminado, ya que la cantidad de casos confirmados y muertes sigue siendo alta ademas se nota un pico en las ultimas fechas.

El pais con mayor dosis administradas es Brasil, seguido de Mexico,Argentina,Colombia,Chile y Peru. Se contradice con los hallazgos anteriores sin embargo recordemos que se debe tener en cuenta la poblacion de cada pais y la cantidad de vacunas disponibles.

comportamiento de la temperatura en los paises seleccionados:

El pais con mayor impacto en la propagacion fue Chile, aunque es extraño ya que Chile fue uno de los paises que mejor manejo la pandemia,sin embargo esto puede ser debido a la densidad de poblacion en Chile

Nuevamente vemos el buen manejo de Chile (se dieron 350 dosis),como no asi de Colombia y Brazil.

Todos los paises tuvieron un comportamiento similar en cuanto a la fatalidad, sin embargo, viendo las otras graficas observamos que han disminuido.

Los casos nuevos y las muertes nuevas han variado , esto debido a la vacunacion y a las medidas tomadas por estos paises , han disminuido la cantidad de personas que se han contagiado y fallecido.

**Analisis del dashboard:**

Al abrir el dashboard de Power BI desarrollado para la expansión de BIOGENESYS en Latinoamérica, te encuentras con una interfaz amigable y organizada que facilita la comprensión de los datos y permite extraer información clave de manera eficiente.sis del dashboard

Se realizaron cuatro solapas, la primera de portada, la segunda de forma general, la tercera de Fecha y la ultima de Pais

Para comenzar con la portada,de manera sencilla se inserto una imagen alusiva al tema con botoneras de navegacion de paginas y titulo, luego el tablero de datos de latinoamerica de visualizacion en general, se incorporaron filtros de paises, tarjetas ,graficos de barras, algunos confeccionados con graficos de POWER BI y otros realizamos con scripts de Python, a continuacion adjunto foto captura.



Luego en el siguiente tablero de Fechas se realizaron visualizaciones de POWER BI y tambien de Python (total de muertes por mes y recuperados por mes), se utilizaron los siguientes Scripts

#librerias

import matplotlib.pyplot as plt

import pandas as pd

import seaborn as sns

#datos

DatosFinalesFiltrado=pd.read\_csv(r"C:\Users\yana\Downloads\DatosFinalesFiltrado.csv")

#Como evoluciono la mortalidad:

DatosFinalesFiltrado['mortality\_rate'] = DatosFinalesFiltrado['cumulative\_deceased'] /DatosFinalesFiltrado['cumulative\_confirmed']

#grafico

plt.figure(figsize=(14,7))

sns.lineplot(data=DatosFinalesFiltrado, x ='date', y='mortality\_rate',estimator='mean',ci=None)

plt.title('Analisis Temporal de la Mortalidad por Covid-19 en America Latina')

plt.xlabel('Fecha')

plt.ylabel('Tasa de Mortalidad')

plt.show()

---------------------------------------------------------------------------------------

#importacion de librerias necesarias

import seaborn as sns

import matplotlib.pyplot as plt

import pandas as pd

#importacion de datos

DatosFinalesFiltrado=pd.read\_csv(r"C:\Users\yana\Downloads\DatosFinalesFiltrado.csv")

#grafico

plt.figure(figsize=(14,7))

sns.lineplot(data=DatosFinalesFiltrado, x ='date', y='cumulative\_recovered',estimator='mean',ci=None)

plt.title('Recuperaciones por mes')#titulo

plt.xlabel('Fecha')

plt.ylabel('Recuperados')

plt.show()

A continuacion adjunto imagen del tablero:



Finalmente se realizo un tablero a nivel Pais, uno de columnas apiladas de personas recuperadas a nivel pais, otro de un mapa donde se puede observar los casos confimados de COVID-19 a nivel Pais y otros dos realizados en python , uno de ellos donde se puede observar la comparacion de victimas fatales femeninas vs masculinas y otro donde se observa el total de dosis administradas tambien a nivel pais.

Adjunto Scripts utilizados :

#Grafica de mortalidad por genero y por PAIS

#importacion de librerias necesarias

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# seleccionamos la los datos filtrados

DatosFinalesFiltrado=pd.read\_csv(r"C:\Users\yana\Downloads\DatosFinalesFiltrado.csv")

#Generamos el grafico de barras

plt.figure(figsize=(10,6))

sns.barplot(DatosFinalesFiltrado, x= 'country\_name',y='adult\_male\_mortality\_rate',color='green',label='Male Mortality Rate') # seleccionamos los datos

sns.barplot(DatosFinalesFiltrado, x= 'country\_name',y='adult\_female\_mortality\_rate',color='yellow',alpha=0.6, label='Female Mortality Rate')#seleccionamos los datos

plt.xticks(rotation=45) # rotacion de datos

plt.ylabel('Mortality Rate')

plt.title('Comparacion de la tasa de mortalidad masculina vs femenina por pais')#titulo

plt.legend(bbox\_to\_anchor=(1.05,1), loc='upper left')# informacion del grafico al costado

plt.show()#mostrar el grafico

--------------------------------------------------------------------------------------

#Dosis administradas por PAIS

#importacion de librerias

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

#llamamos a la base de datos

DatosFinalesFiltrado=pd.read\_csv(r"C:\Users\yana\Downloads\DatosFinalesFiltrado.csv")

# Grafico

plt.figure(figsize=(5,5)) # medidas del grafico

sns.lineplot(data=DatosFinalesFiltrado, x ='country\_name', y='cumulative\_vaccine\_doses\_administered',estimator='mean',ci=None)#seleccionamos datos que queremos utilizar

plt.xlabel('Pais')

plt.ylabel('Dosis Administradas')

plt.show()

A continuacion adjunto imagen;



**Conclusión General y recomendaciones:**

El análisis detallado realizado a través de las visualizaciones y el dashboard en Power BI ha permitido identificar de manera efectiva las regiones óptimas para la expansión de BIOGENESYS en Latinoamérica. Al integrar datos socioeconómicos, demográficos, de salud, y de infraestructura tecnológica, se logró obtener un panorama completo que guía las decisiones estratégicas para una inversión rentable y sostenible en los países objetivo: Colombia, Argentina, Chile, México, Perú y Brasil. Las visualizaciones facilitaron la comparación de factores clave como políticas de vacunación, inversión en salud, densidad poblacional, y problemas sociales y económicos, proporcionando un enfoque claro y centrado en la identificación de las áreas con mayor potencial para el crecimiento de la empresa.

En conclusion y como recomendacion se considera que los paises donde debe expandirse la empresa farmaceutica son, en primer lugar , Brazil ,Mexico y Colombia, ya que son los paises que mas casos confirmados y muertes han tenido, ademas de ser los paises que menos vacunas han administrado.

**Reflexión personal:**

Este proyecto ha sido una experiencia enriquecedora que me ha permitido consolidar y ampliar mis habilidades como Analista de Datos, especialmente en el uso de Power BI para la visualización y análisis de datos complejos. Durante el proceso, aprendí a manejar grandes volúmenes de información y a integrarlos de manera efectiva para proporcionar insights claros y accionables. También perfeccioné mi capacidad para seleccionar indicadores clave, realizar limpiezas de datos y transformar conjuntos de datos en visualizaciones que facilitan la toma de decisiones estratégicas,tambien me parecio increible como pude aprender a leer los errores que iban apareciendo en cada proceso ,de todas formas creo que quedaron analisis un poco inconclusos a mi parecer.

En resumen, el proyecto fue una excelente oportunidad para aplicar y mejorar mis habilidades en análisis de datos, y aunque se lograron en general los objetivos, siempre hay oportunidades para optimizar y mejorar el proceso de trabajo.

En respuesta a si tuviera que volver a empezar este proyecto, haría algunas cosas de manera diferente, invertiría más tiempo en la etapa de planificación y definición de objetivos, para asegurar que todos los indicadores y variables relevantes estuvieran claramente identificados desde el inicio,como asi tambien mas tiempo en realizar mas medidas, ya que creo que no genere todas las que podian haber sido necesarias.