BIOGENESYS

Expansión Estratégica en Latinoamérica

**Nombre del autor:** Franco BONAVENTO

**Email:** franbonavento@gmail.com

**Cohorte:** DAFT-04

**Fecha de entrega:** lunes, 29 de julio de 2024



# Introducción

BIOGENESYS es una empresa farmacéutica que busca la Expansión Estratégica de Laboratorios para Investigación y Vacunación de COVID-19

Fuimos contratados como analistas de datos para buscar e 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 post-pandemia con el fin de mejorar el acceso a las vacunas.

Nos solicitan que el análisis deberá incluir una evaluación sobre la demanda de vacunas, la logística de su distribución y la infraestructura sanitaria ya existente.

# 

# Desarrollo del proyecto

# El proyecto se realizó en 4 etapas:

ETAPA 1 - Carga y transformación de datos con Python:

El equipo de ingeniería de datos, nos brinda un dataset en formato csv que contiene más de 12 millones de registros y 50 columnas. Lo acompañan de un readme.txt que muestra las columnas del dataset con sus descripciones.

Comenzamos utilizando VSC con Júpiter Notebook vinculado a Python para cargar el dataset anteriormente mencionado y comenzar a familiarizarnos con este conjunto de datos.

La empresa nos solicita filtrar por países: Argentina, Brasil, Colombia, Perú, México, Chile y con una fecha mayor a 1 de enero de 2021.

Luego de este proceso nos encontramos con un dataset bastante más pequeño en tamaño, el cual lo guardamos para manipular y continuar con el análisis utilizando este último y de esta forma ser más eficientes en el uso de los recursos de nuestro equipo para agilizar la carga de los datos.

Además realizamos una limpieza puesto que el dataset contenía muchos valores nulos y calculamos medidas estadísticas descriptivas para lograr un primer acercamiento a los datos que estamos manejando.

ETAPA 2 - Análisis Exploratorio y Primeras Visualizaciones:

Luego de chequear que no estábamos trabajando con datos repetidos ni nulos, pasamos a setear como índice a nuestro campo fecha y asegurarnos que la misma sea del tipo fecha.

Como en la etapa anterior ya habíamos realizado análisis estadísticos de los datos, pasamos a visualizar el comportamiento de las variables de nuestro dataset. analizamos la correlación entre las mismas. generamos gráficas de barras, de líneas, de violín, con KDE, de dispersión, de caja, barras apiladas. Todas ellas para comparar en general y por país.

Esta segunda etapa nos reveló información valiosa que estaban en los datos, permitiendo comenzar a armar una idea del rumbo de nuestros objetivos a alcanzar.

Utilizamos varias librerías de Python: Pandas, Numpy, Matplotlib y Seaborn.

Tratamiento de nulos: damos una explicación porque el dataset que nos brindaron poseía muchos valores nulos.

>>new\_confirmed/decesed: cantidad= 21 - se rellenan con cero por ser pocos

>>cumulative\_confirmed/decesed: cantidad 21 - por ser valores que se acumulan, se comienza desde el mínimo, completando el valor siguiente con el anterior.

>>cumulative\_vaccine\_doses\_administered: cantidad 584 - mismo caso que anterior.

>>variables de temperatura, humedad y lluvias: se rellenan con la media de cada variable. no superan los 90 en cantidad de nulos por variable, es decir son pocos. no afectan a los resultados del dataset en gran medida.

>>new\_recovered: cantidad = 2.736 - luego de rellenar con cero, se pasan por el promedio.

>>cumulative\_recovered: cantidad 2.736 nulos - luego de pasarlos a ceros, se calcula el acumulado. rellenando con el dato anterior.

(los últimos solo tienen valores para Brasil y Colombia)

ETAPA 3 - EDA - Técnicas avanzadas:

Comenzamos con un análisis exploratorio más avanzado gracias a Pandas y numpy que permiten realizar esto.

Por lo que realizamos gráficas que nos permitieron analizar comportamientos, tendencias, distribuciones, correlaciones más profundas que en la etapa anterior.

Teniendo en cuenta siempre el objetivo de nuestro proyecto.

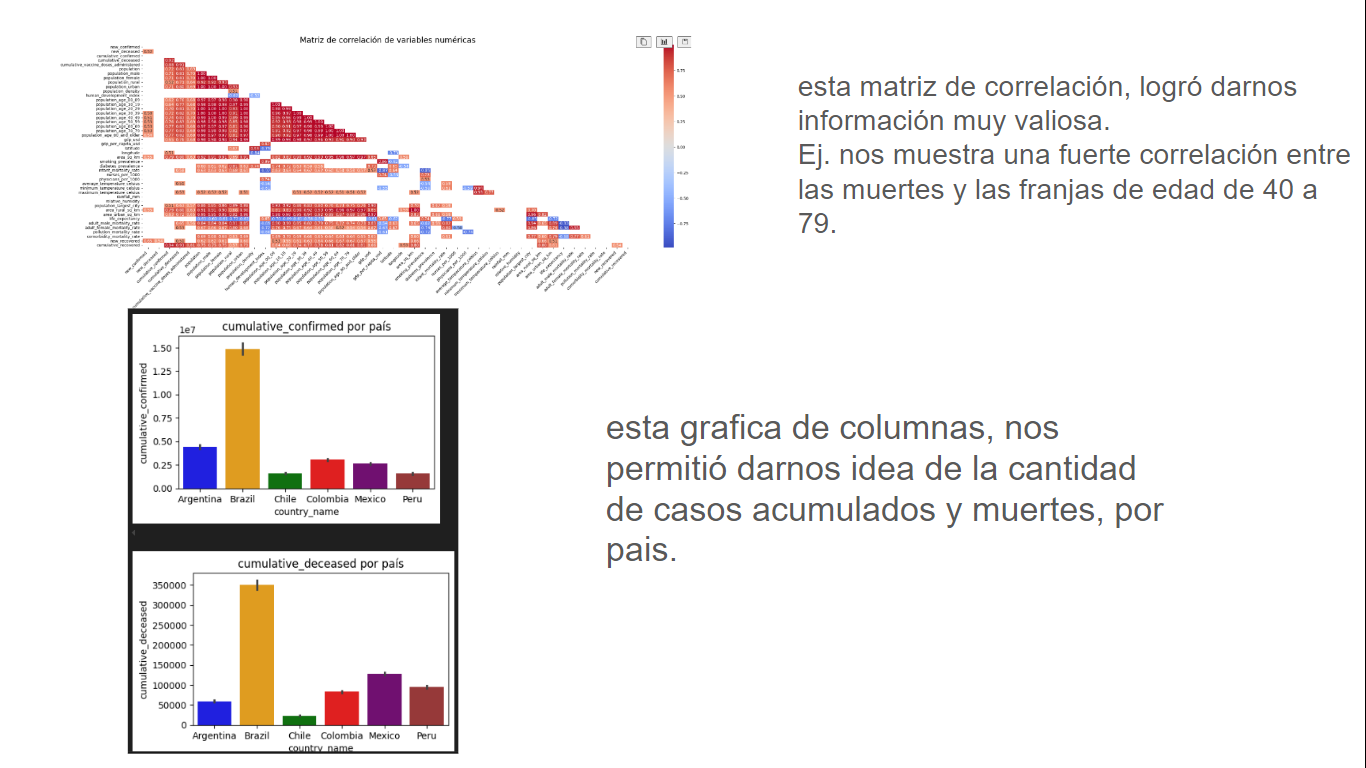
ETAPA 4 - Integración con PowerBI:

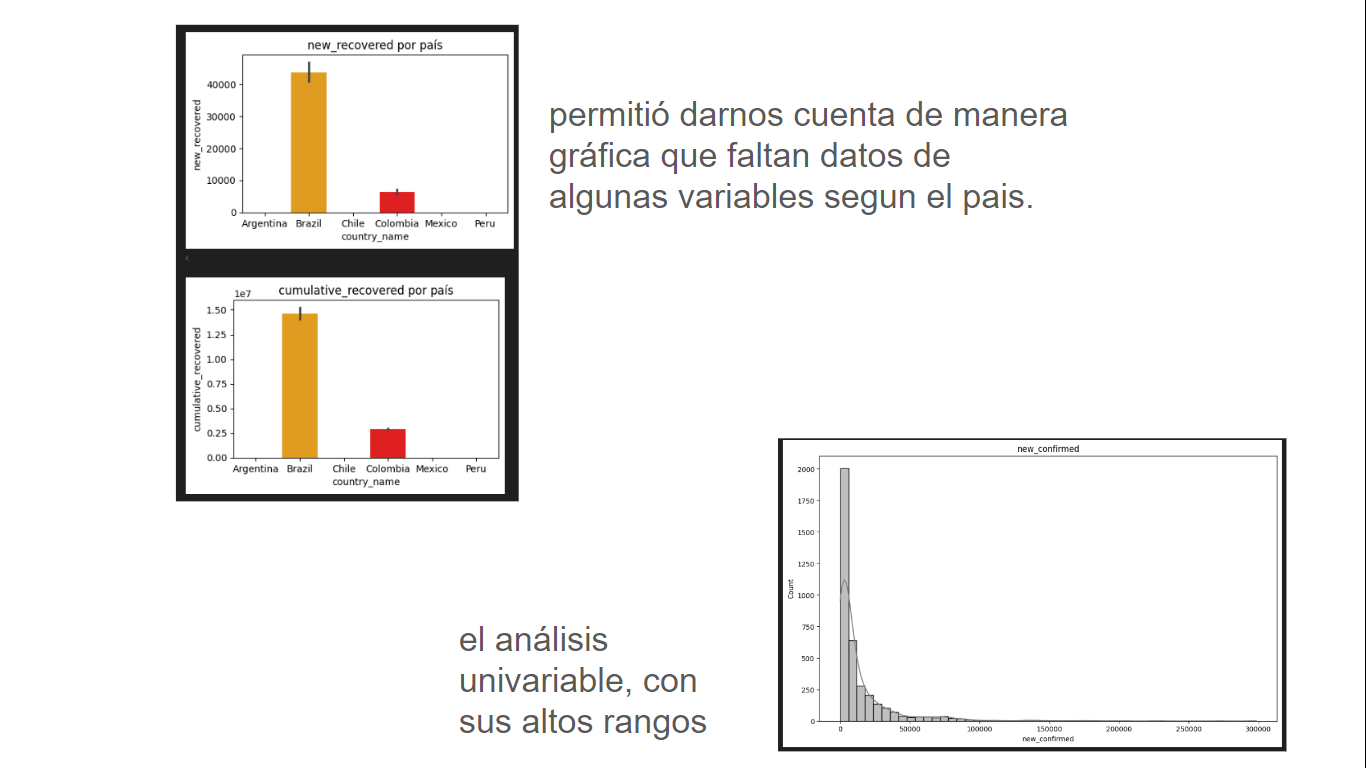
Luego de culminar con los insights hallados que no permitieron encontrar las etapas anteriores, procedimos a conectar nuestro dataset a PowerBI para lograr armar un dashboard, sabiendo que al momento de presentar a los interesados nuestro trabajo, logramos que puedan comprender mejor y de una forma más amigable e interactiva, el comportamiento de los datos y contarles nuestra propuesta estratégica para su expansión de una manera convincente.

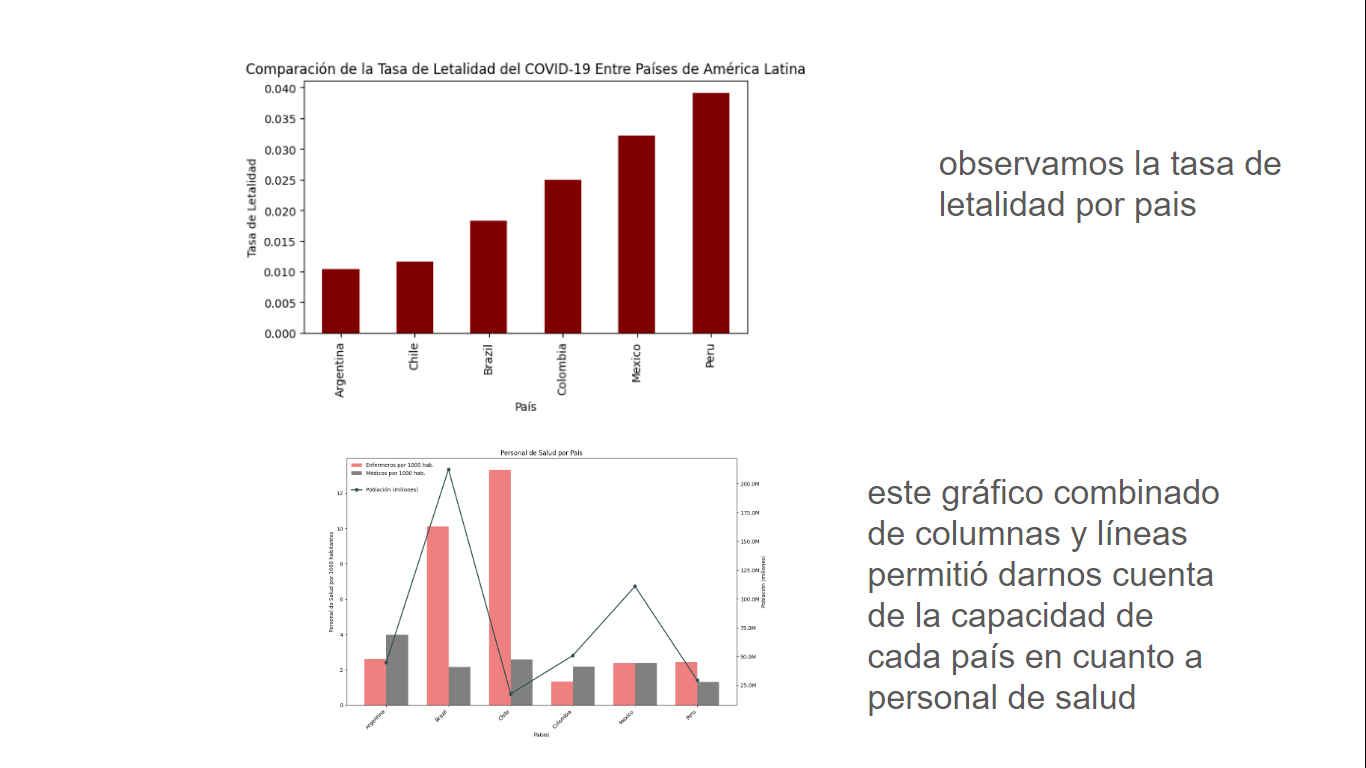
Esta se destaca en cuanto a lo visual por su dinamismo e interactividad, algo que no tiene el reporte que fuimos creando en las etapas anteriores con nuestro Yúpiter Notebook.

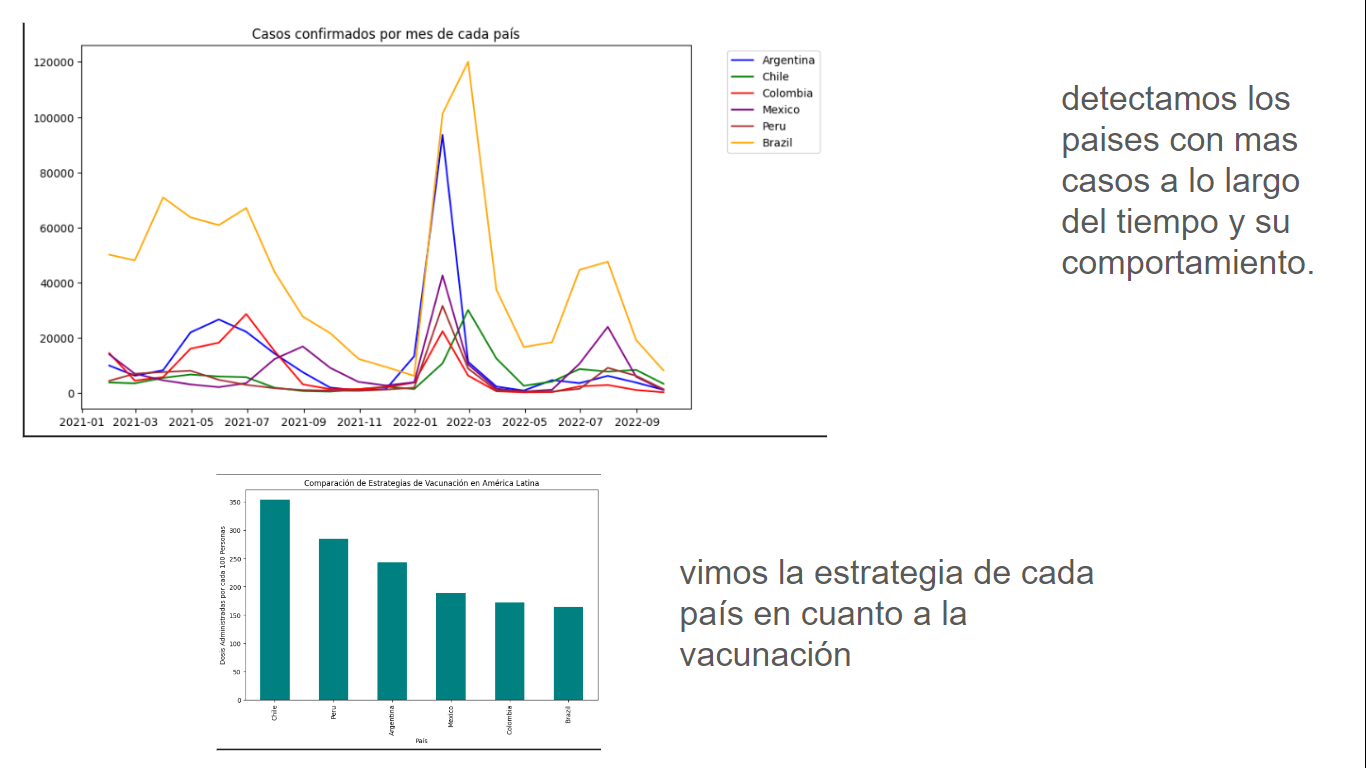
# EDA e insights

# 









# Análisis del dashboard

# El dashboard comienza con una página llamada portada, desde esta página, se navega por el resto de las páginas, las cuales son: GENERAL, INCIDENCIA y CONCLUSIONES, mediante un botón de navegación, el cual funciona haciendo control + clic. Luego en las otras paginas, el mismo botón de navegación lo van a encontrar arriba del lienzo, y permite navegar por el resto de las páginas, con control + clic.

En las paginas GENERAL, INCIDENCIA las visualizaciones son interactivas en su mayoría, permitiendo hacer clic sobre una zona en particular y automáticamente el resto de las gráficas de esa pagina van a mostrar los datos en relación a lo seleccionado.  
Además posee cada página de reportes, un botón para restablecer a los valores iniciales las visualizaciones.

# Conclusiones y Recomendaciones

# EN UNA PRIMERA ETAPA, LA EXPANSIÓN SE REALIZARÁ EN MÉXICO Y BRASIL, PUESTO QUE SON PAÍSES CON MUCHA DENSIDAD DE POBLACIÓN, POR LO QUE SE GENERAN MUCHOS CONTAGIOS, SON PAÍSES GRANDES Y POSEEN UN PIB PER CÁPITA CONSIDERABLE. TODO ESTO SABIENDO QUE TIENEN DESDE ALTAS A MEDIAS LAS TASAS DE LETALIDAD POR COVID. Y ADEMÁS ESTRATÉGICAMENTE PARA LA EMPRESA SE DISTRIBUYEN EN ZONAS DIFERENTES, PUDIENDO LUEGO HACER UNA EXPANSIÓN EN UNA SEGUNDA ETAPA. ADEMÁS TIENEN UN DÉFICIT EN VACUNAS DADAS POR CADA 100 PERSONAS, LO QUE HACE INCLINARSE POR ESTOS PAÍSES EN UNA PRIMERA ETAPA

# LA SEGUNDA ETAPA ABARCA LOS PAÍSES DE COLOMBIA, PERÚ QUE POSEEN ZONAS DE MUCHA DENSIDAD DE POBLACIÓN, PERO EL PBI PER CÁPITA ES MENOR, POR LO QUE LA EMPRESA DEBERÁ AFRONTAR NEGOCIACIONES PARA LOGRAR COMPETIR EN EL MERCADO EN CASO DE QUE LOS ESTADOS DECIDAN DESTINAR LAS INVERSIONES HACIA OTROS SECTORES.

# COMO UNA POSIBLE TERCERA ETAPA QUEDARÍA ARGENTINA Y CHILE, SIENDO ESTE ÚLTIMO EL QUE MEJOR MANEJÓ LA PANDEMIA POR LO QUE NO NOS INTERESA PARA NUESTRO OBJETIVO DE EXPANSIÓN

# 

# Reflexión personal

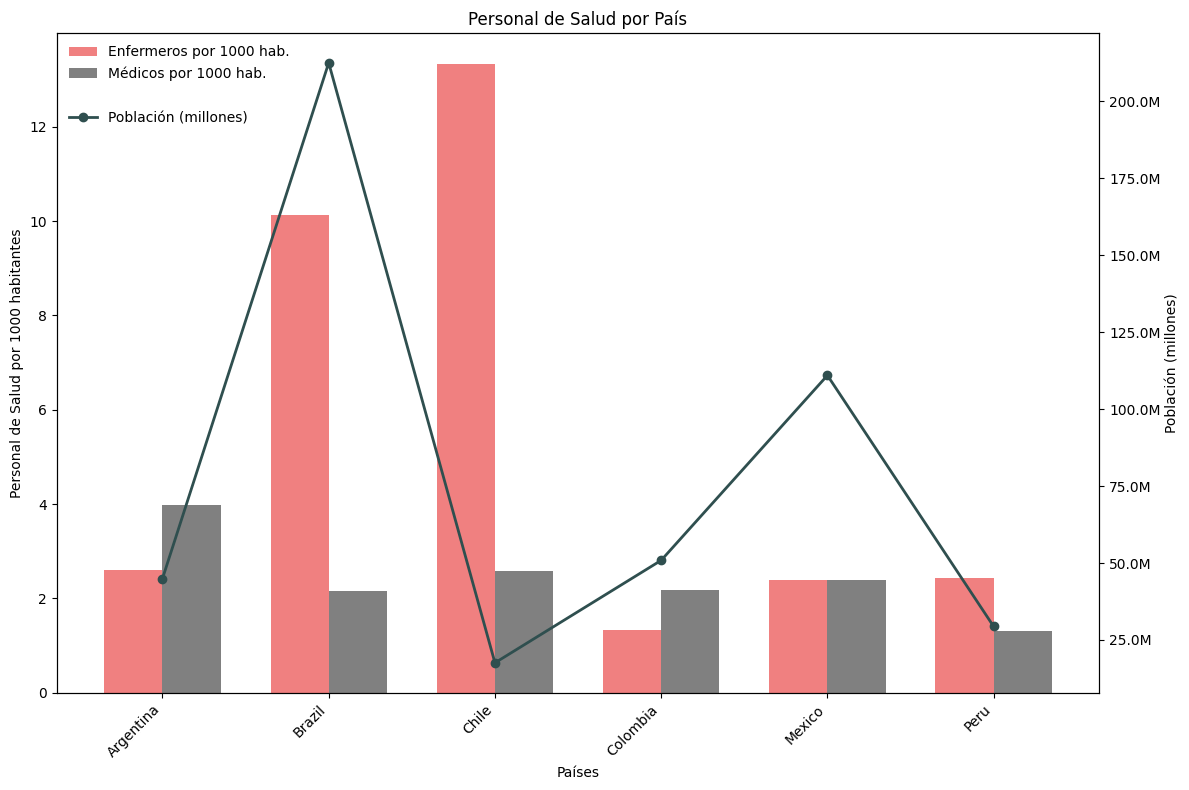
# A lo largo de este proyecto se aprendieron muchas cosas! Se pudieron integrar conceptos previos de google sheets, sql, power BI e incorporarlos con Python en VSC.

# Permitió comprender cómo se deben analizar las gráficas y cómo se crean, utilizando el lenguaje de programación python con sus librerías tan importantes para los analistas de datos. Sorprendió la capacidad que dan los gráficos de encontrar patrones, tendencias, correlaciones entre los datos que trabajamos. Sabiendo que comenzamos con un dataset muy grande y se pudo manipular con conocimiento y saber limpiarlo y normalizar. Además de guardarlo con sus modificaciones para compartir a quien lo desea para su análisis u otra herramienta para importarlo.

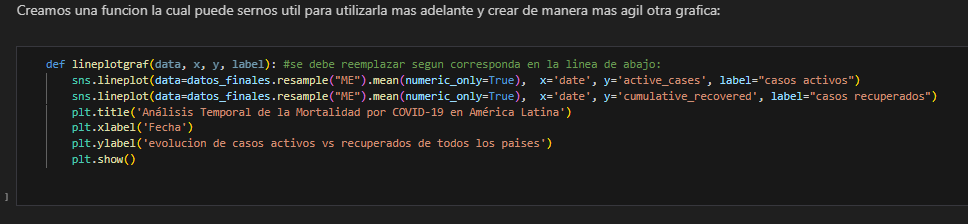
Como reflexión final, me hubiera gustado tener más momentos de práctica, para lograr un mejor resultado del proyecto.

# EXTRA CREDIT

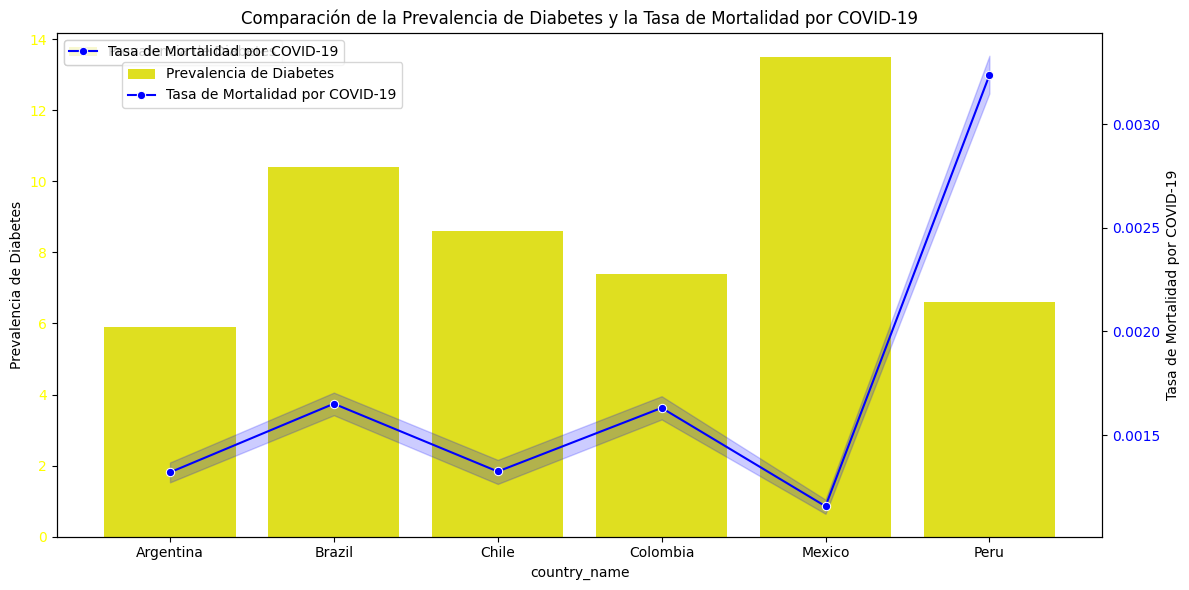
Esta gráfica es muy importante porque logra sintetizar información importante sobre personal de salud disponible y cantidad de población de cada país.



Se crea una función, para utilizar para graficar, según el caso:



Se crea en una sola gráfica combinando líneas con columnas para resumir la relación de 2 variables:



Creamos un heatmap, para ver en un mapa los países y datos de interés. casos confirmados, muertes, vacunas y recuperados:

