**Expansión estratégica del Laboratorio Biogenesys**

**Nombre del autor: Castillo Antonella**

**Email:** cpn.castilloantonella@gmail.com

**Cohorte:** DA-FT03

**Fecha de entrega:** 01 de julio del 2024

**Institución:** La compañía farmacéutica BIOGENESYS está investigando las mejores ubicaciones para expandir sus laboratorios. Para ello, están analizando datos sobre la incidencia de COVID-19, las tasas de vacunación y la disponibilidad de infraestructuras de salud. El objetivo es mejorar la respuesta a los efectos de la pandemia y postpandemia, facilitando el acceso a las vacunas.

****

# Introducción

El proyecto implicó apoyar al Laboratorio Biogenesys mediante un estudio y análisis de la situación del COVID-19 para facilitar la toma de decisiones en la estrategia de expansión en Latinoamérica, enfocándose en Colombia, Argentina, Chile, México, Perú y Brasil. Es crucial comprender el contexto y la respuesta de cada país a la pandemia para mejorar el acceso a las vacunas en el futuro.

# Desarrollo del proyecto

El dataset original utilizado por el Laboratorio Biogenesys tiene más de 22 millones de registros y más de 707 columnas. Sin embargo, nos proporcionan un dataset más reducido que abarca países de Latinoamérica, con alrededor de 12 millones de registros y 50 columnas.

Primero, cargamos el dataset en un IDE adecuado para su correcta manipulación y análisis. Utilizamos Visual Studio Code como IDE, junto con Python, Jupyter Notebooks y bibliotecas como Numpy, Pandas, Matplotlib y Seaborn para realizar análisis estadísticos y gráficos.

Es fundamental verificar la calidad de los datos al empezar a trabajar con ellos. Esto implica revisar que los tipos de datos de cada columna sean correctos, identificar valores nulos y evaluar si es posible completarlos para mejorar la calidad del análisis, así como limitar los datos a los que son de interés.

En nuestro caso, necesitamos analizar datos de los siguientes países: 'Colombia', 'Argentina', 'Chile', 'Mexico', 'Peru' y 'Brazil', y que los datos sean posteriores al 01/01/2021.

Con Pandas, podemos crear un dataframe que contenga solo estos datos y corregir los valores nulos. Una vez definido el dataframe de interés, lo guardamos en un nuevo archivo .csv para facilitar futuros análisis y optimizar tiempo y recursos al no cargar todos los datos completos.

# EDA e insights

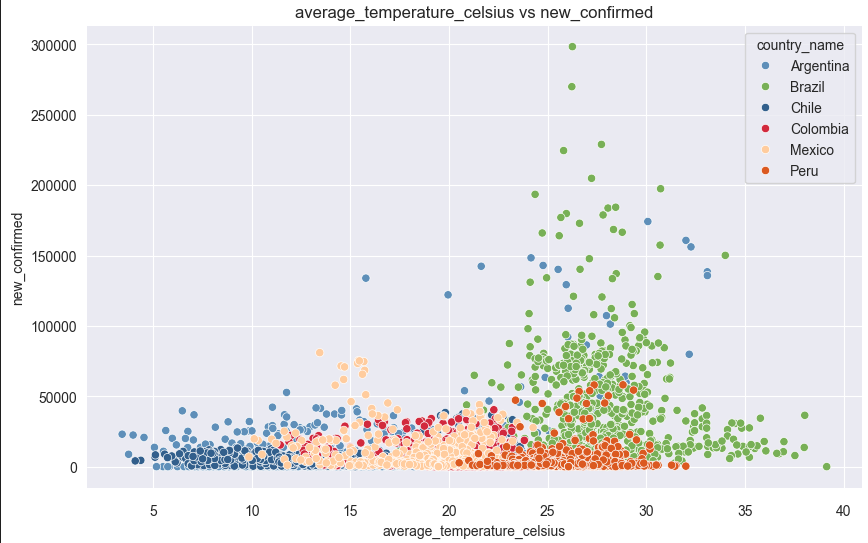
Utilizando Pandas y Numpy, podemos explorar las propiedades estadísticas del dataset. Con las bibliotecas Matplotlib y Seaborn, podemos visualizar gráficamente aquellos insights y hallazgos que nos permitan llegar a conclusiones significativas.

En primer lugar, realizamos una descripción estadística de todas las variables de interés del dataframe para obtener una visión general. Sin embargo, solo mirando números, puede ser difícil comprender estos hallazgos. Por lo tanto, pasamos a la visualización gráfica para entender mejor el comportamiento de nuestros datos.

Un primer enfoque es crear un gráfico de correlación para observar cómo interactúan las variables entre sí y determinar en cuáles debemos centrarnos.

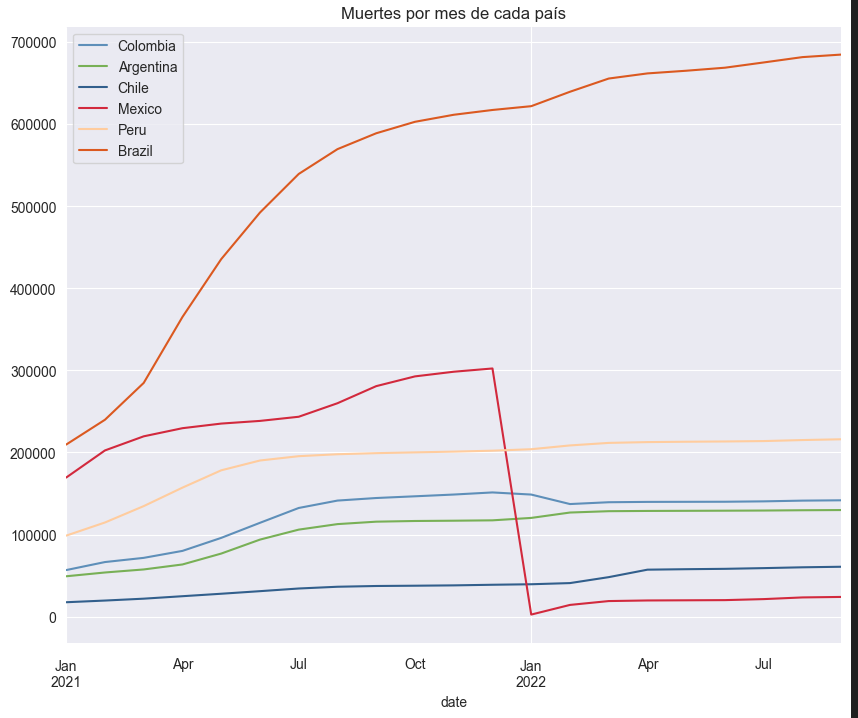
Inicialmente, realizamos una revisión rápida del comportamiento de los datos en cada columna mediante gráficos de barras, ideales para comparar categorías. También resultó interesante observar la distribución de algunos datos, para lo cual creamos los histogramas correspondientes.

Otra forma de descubrir insights es enfocando el análisis en dos variables específicas. Por ejemplo, al relacionar en un gráfico de dispersión los nuevos casos confirmados de COVID-19 con la temperatura, podemos investigar si la temperatura afecta el número de muertes confirmadas. Para esto, se puede crear un segundo gráfico que examine esta relación.

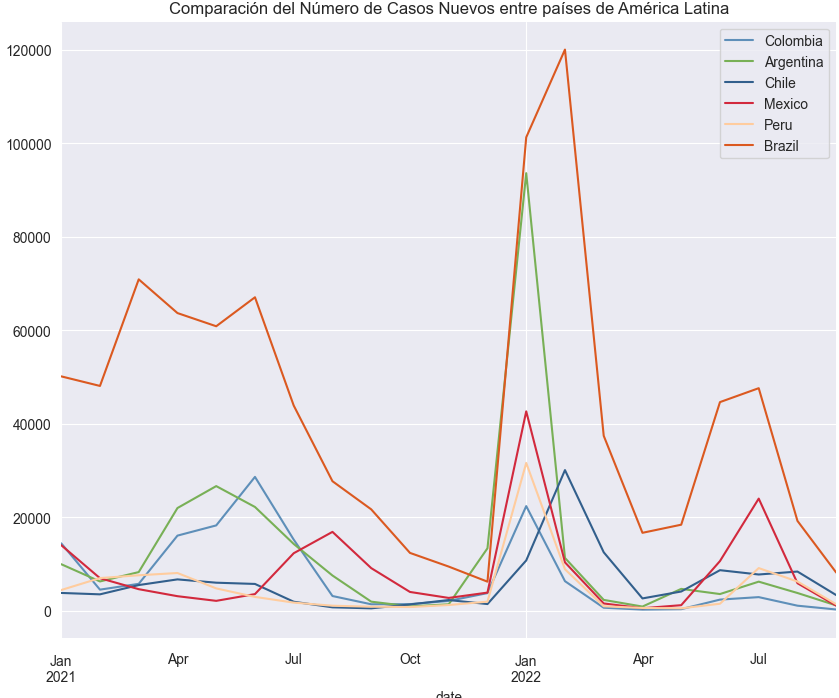


Un factor crítico a tener en cuenta en nuestro proceso de toma de decisiones es comprender la cantidad de dosis de vacunas administradas por país. Además, un gráfico de líneas puede ayudarnos a observar cómo cambia esta variable a lo largo del tiempo para cada país.

Otro dato importante es tener en cuenta el número de muertes causadas por COVID-19 por país. Se creó un gráfico correspondiente y se reveló que Brasil y México tienen el mayor número de muertes, seguidos de Colombia, Argentina, Perú y Chile. Sin embargo, esto no significa necesariamente que Brasil y México hayan gestionado mal la pandemia. En cambio, es esencial considerar el tamaño de la población de cada país, los datos disponibles y otros factores relevantes.



Se creó un gráfico para rastrear los nuevos casos por país a lo largo del tiempo, y lo que se puede inferir de él es que los casos fueron disminuyendo hasta diciembre, cuando una tendencia estacional relacionada con las festividades (Navidad/Año Nuevo) hizo que los casos comenzaran a aumentar nuevamente, alcanzando un máximo. en enero/febrero.



Vale la pena señalar que todos los países respondieron eficazmente con buenas políticas, ya que los números de casos comenzaron a disminuir rápidamente en los meses siguientes.

Un hallazgo significativo durante el Análisis Exploratorio de Datos (EDA) fue que Brasil y México son los países que peor han manejado la pandemia, con un alto número de casos confirmados y muertes en comparación con otros países. Además, se administra una cantidad desproporcionada de dosis de vacuna.

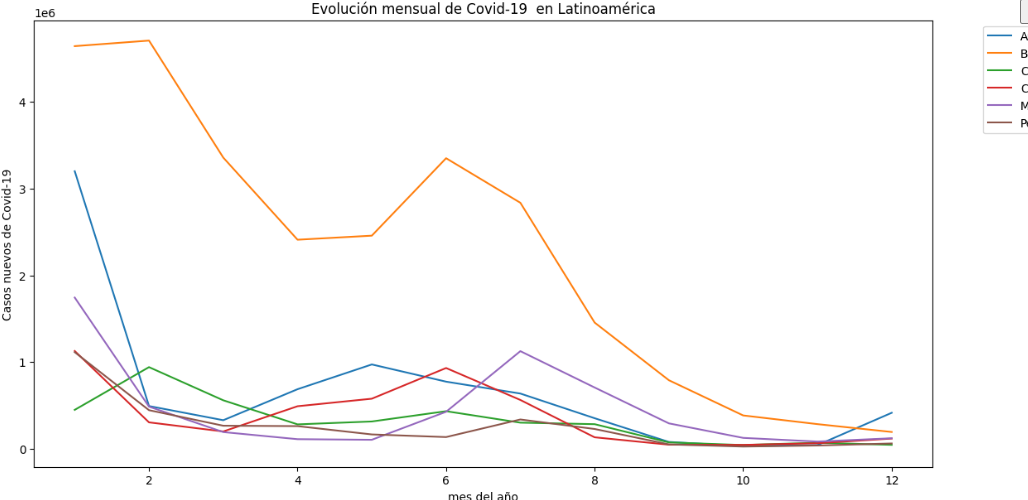
En este punto, podría parecer lógico suponer que los países con el mayor número de muertes y casos confirmados son los lugares ideales para nuevos laboratorios, ya que requerirían la mayor asistencia.

Sin embargo, también es fundamental analizar la población de cada país. Se crearon gráficos para examinar las tasas de mortalidad femenina y masculina en cada país, lo que llevó a la conclusión de que los hombres tienen una tasa de mortalidad más alta, lo que los hace más vulnerables al virus.

Esto podría estar relacionado con condiciones preexistentes o comorbilidades para ver si alguna enfermedad ha causado más muertes. Al correlacionar indicadores demográficos y de salud, encontramos que la diabetes y el tabaquismo tienen una incidencia significativa en la población, aunque esto no descarta el impacto de otras enfermedades.

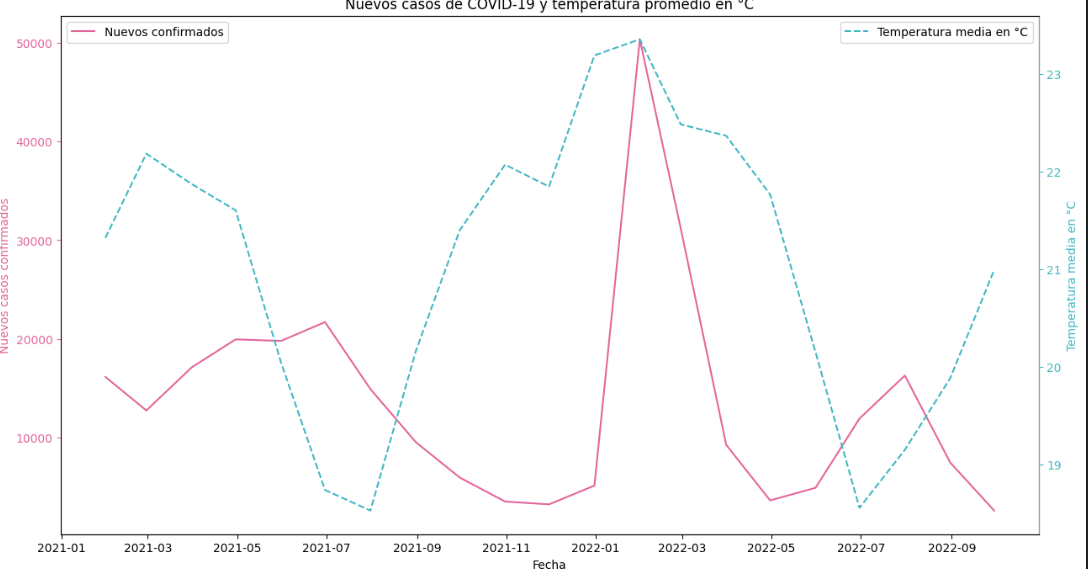
Cuando graficamos los datos, vimos que la diabetes es la principal causa de muerte en los países seleccionados.

Por ahora, Brasil, México y Colombia son los países que necesitan más ayuda. Si graficamos la evolución semanal/anual de nuevos casos por país, podemos ver que los nuevos casos y las muertes han ido disminuyendo con el tiempo. Sin embargo, esto no indica el fin de la pandemia, ya que se observa un resurgimiento de los casos hacia el final.



Brasil es el país con más dosis de vacunas administradas, seguido de México, Argentina, Colombia y Chile. Sin embargo, esto parece contradecir la inferencia anterior de que estos países eran los que necesitaban más ayuda y estaban manejando peor la situación. Para conciliar esto, debemos considerar el tamaño de la población de cada país y la cantidad de vacunas disponibles.

También podemos obtener información teniendo en cuenta factores como la temperatura. Observamos que a medida que aumentó la temperatura promedio, los casos disminuyeron, pero hubo un pico antes de marzo, que posteriormente disminuyó significativamente, posiblemente debido a políticas efectivas de vacunación.



Finalmente, la estrategia de vacunación de cada país nos ayudará a determinar dónde es conveniente que el Laboratorio Biogenesys se expanda.

Vemos que Chile ha implementado la mejor estrategia de vacunación entre los países, mientras que México, Colombia y Brasil muestran una marcada diferencia. Esto refuerza la conclusión anterior de que estos tres países son los que más ayuda necesitan para abordar los problemas futuros relacionados con el Covid.

Por lo tanto, luego de analizar el tema desde múltiples ángulos, podemos concluir que los países donde el Laboratorio Biogenesys debería expandirse son Brasil, México y Colombia, ya que son los países con más casos confirmados y muertes.

Además, son los países con estrategias de vacunación menos efectivas, sin considerar la densidad poblacional y la situación socioeconómica que podría incidir en el número de casos.

# Análisis del dashboard

Luego de implementar todas las medidas, logramos armar con éxito el tablero final con gráficos que ayudan a dar respuesta a las necesidades planteadas por Biogenesys Laboratorios.

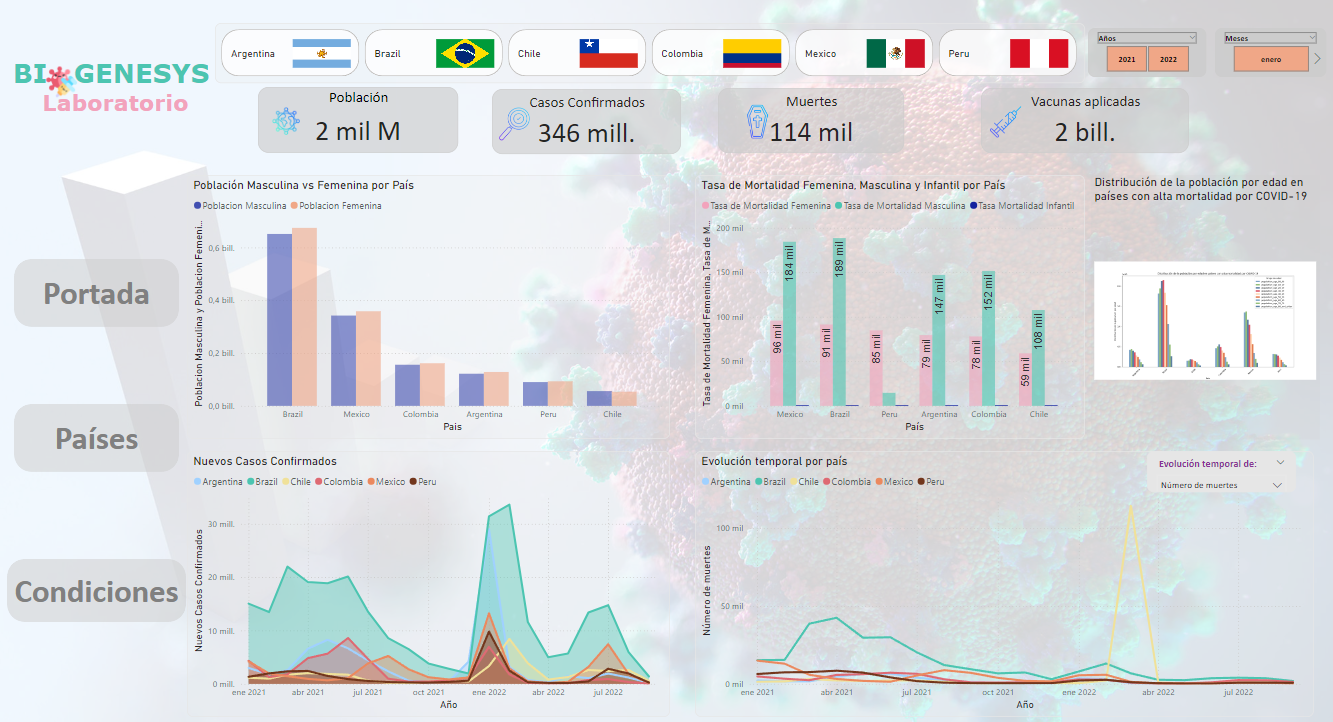
Este panel proporciona una visualización completa de los datos, lo que permite a Biogenesys tomar decisiones informadas sobre su estrategia de expansión y abordar eficazmente la pandemia de Covid-19 en los países objetivo.

**Portada:**



En la portada nos vamos a encontrar con cuatro botones que nos ayudan a navegar por las páginas del tablero.

**Población:**



En esta página, el usuario puede consultar de un vistazo una descripción general completa de la información más relevante sobre la COVID-19. Los datos se pueden filtrar por país, año y mes, lo que proporciona una visión personalizada del impacto de la pandemia.

Las tarjetas superiores resaltan los indicadores clave de rendimiento (KPI), proporcionando una descripción rápida de la situación.

Los gráficos de columnas muestran la población masculina y femenina de cada país, así como las tasas de mortalidad (este último muestra también la tasa de mortalidad de los niños)

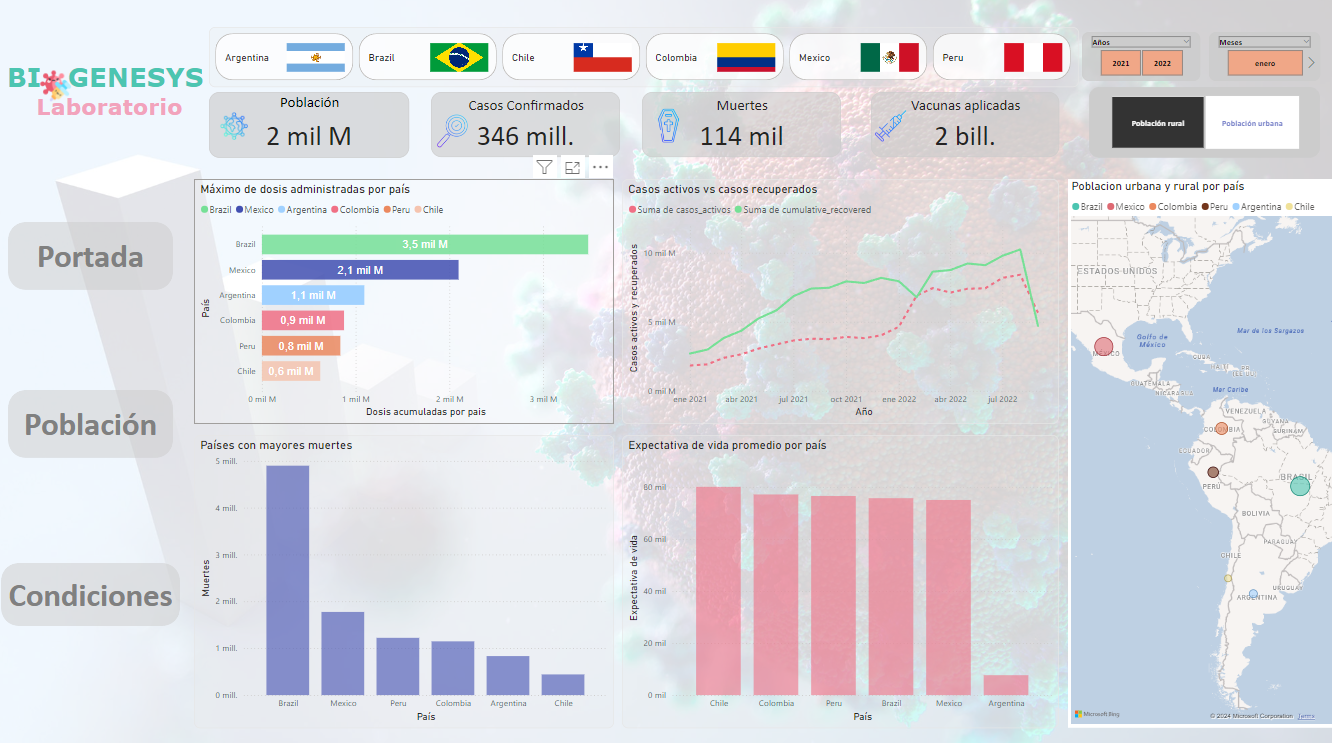
Un tercer gráfico de barras donde muestra el rango de edad con relación a las muertes ocurridas. Ente gráfico fue exportada de Python y los filtros que se realicen no influyen en esta figura-

Un gráfico de áreas proporciona una comparación visual y detallada de los casos nuevos a lo largo del tiempo para cada país.

Finalmente, en la esquina inferior derecha, un gráfico muestra los principales indicadores de cada país (nuevos casos, nuevas recuperaciones, muertes y dosis administradas) a lo largo del tiempo. El selector de lista desplegable permite a los usuarios elegir qué gráfico ver.

Este panel proporciona una forma interactiva y fácil de usar para explorar los datos de COVID-19, lo que permite a los Laboratorios Biogenesys tomar decisiones informadas sobre su estrategia de expansión y respuesta a la pandemia.

**Países:**



En esta sección profundizamos en los datos específicos de cada país, proporcionando una descripción más detallada de la situación de cada uno.

Un mapa muestra la distribución de la población rural y urbana de cada país, ofreciendo una representación visual de la demografía de la población.

Podemos ver la evolución de los casos activos y los casos nuevos a lo largo del tiempo, lo que proporciona información sobre la progresión de la pandemia en cada país.

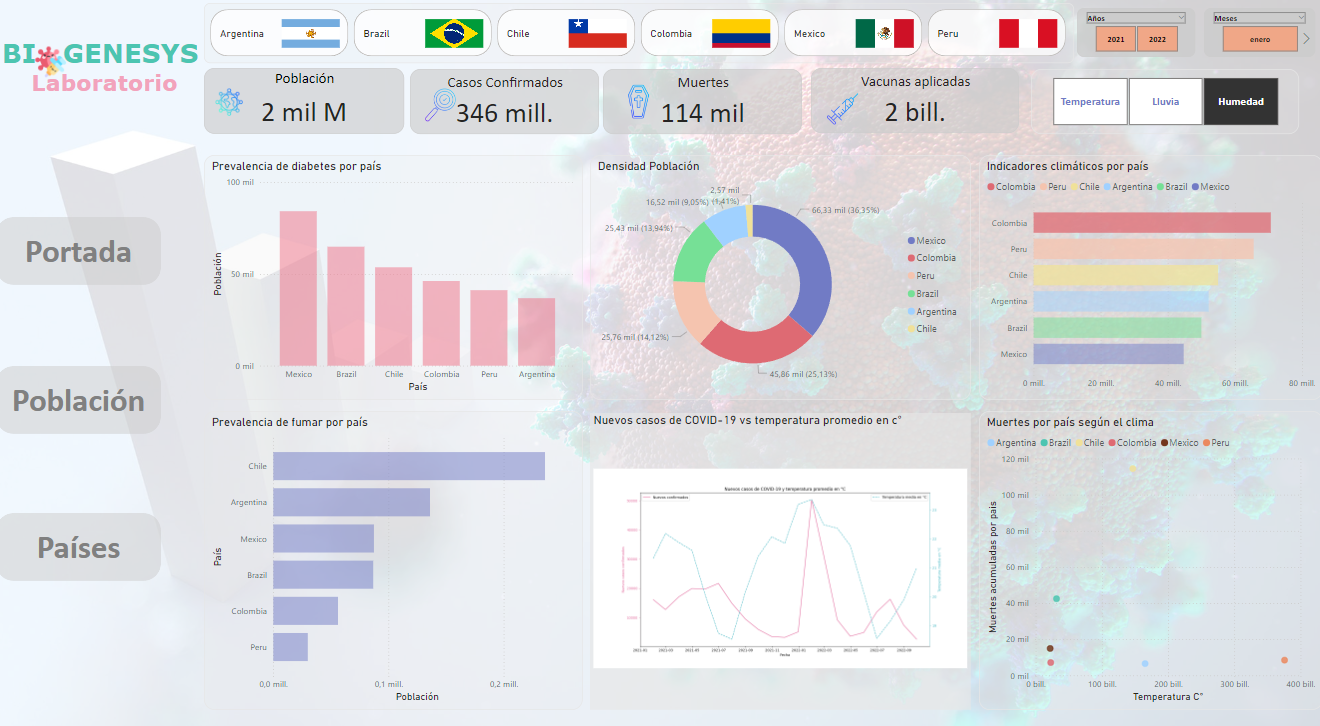
También podemos visualizar un gráfico de barras donde muestra cual es el país con mas muertes.

Un gráfico de barras muestra la esperanza de vida promedio para cada país, mientras que otro gráfico muestra la prevalencia de enfermedades en cada país como porcentaje.

Un gráfico crítico destaca el número máximo de dosis administradas por cada país de interés, proporcionando una métrica clave para evaluar las estrategias de vacunación.

Al igual que en la pantalla anterior, los usuarios pueden filtrar la información por país, año y mes para adaptar su análisis a necesidades específicas.

**Condiciones Países:**



Esta página proporciona una descripción detallada de los valores clave que representan las condiciones de cada país, incluyendo:

Gráficos de barras donde muestra la prevalencia de las enfermedades como la diabetes y fumar.

Un grafico circular, con la densidad de la población.

Un grafico que muestra las muertes versus la temperatura por país.

Un gráfico que muestra los principales indicadores climáticos, incluida la temperatura, la humedad y las precipitaciones.

A modo de aclaración cuando trabaje el tema de las tarjetas KPI, el n° de población no me un número confiable, todo lo contrario pasa cuando se realiza un filtrado por país.

# Conclusiones y Recomendaciones

Como se ha venido trabajando en el apartado del EDA y INSIGHTS, llego a la conclusión que con los conocimientos adquiridos durante la fase EDA (Análisis exploratorio de datos) Brasil, México y Colombia son los países ideales para que Biogenesys Laboratorios se expanda.

Estos países poseen una gran población, pero carecen de una política sólida de administración de vacunas. Esto presenta una oportunidad para que los Laboratorios Biogenesys no sólo establezcan nuevos laboratorios en estos países sino también desarrollen una nueva estrategia de vacunación que apoye y ayude a los gobiernos locales.

Una recomendación que puedo sugerir es que Biogenesys Laboratorios adopte un enfoque doble, es decir, que además de instalar los nuevos laboratorios en esos países, se cree una estrategia nueva de vacunación para dar soporte/ayuda a los gobiernos.

# Reflexión personal

# Como analista de datos, tuve la oportunidad de trabajar en este informe integral para una empresa de ventas. A lo largo de esta experiencia, me di cuenta de la importancia de comprender los datos como la "materia prima" que alimenta nuestro análisis.

Durante el proceso de extracción, transformación y carga (ETL), aprendí que no siempre recibimos conjuntos de datos impecables y bien organizados con una integridad perfecta. En cambio, debemos prestar mucha atención a la calidad de los datos proporcionados, analizarlos, transformarlos y solo entonces podremos sumergirnos en la creación de fórmulas y visualizaciones para responder preguntas comerciales.

Otra de las principales enseñanzas de esta experiencia es la importancia de estar organizado mientras se trabaja. Esto incluye: seguimiento de los datos, documentar suposiciones y estructurar código y archivos de manera lógica y consistente.

Como para ir concluyendo, es la primera vez que veía un programa como Python por el cual pasé muchos momentos de frustración a no poder comprenderlo bien, por suerte pude superarlo para seguir adelante con el proyecto. Se que aun me queda mucho mas por aprender y practicar para lograr ser un buen DA