**Expansión estratégica de laboratorios para investigación y vacunación de COVID-19**

**Nombre del autor:** Lucas Damián Gebhardt

**Email:** lucasdamiangebhardt@gmail.com

**Cohorte:** DA-FT01

**Fecha de entrega:**  06/05/2024

**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**

El proyecto consistió en asistir al Laboratorio Biogenesys con un estudio y análisis de situación de COVID-19 para ayudar en la toma de decisiones a la hora de pensar una estrategia de expansión en Latinoamérica específicamente en Colombia, Argentina, Chile, México, Perú y Brasil.

Es fundamental poder entender el contexto y tener una noción clara de la respuesta a la pandemia de cada país y con ellos poder mejorar el acceso a vacunas en el futuro.

**Desarrollo del proyecto**El dataset original con el que cuenta el Laboratorio Biogenesys tiene mas de 22 millones de registros y más de 707 columnas, pero se nos entrega un dataset más acotado a los países de Latinoamérica, que contiene unos 12 millones de registros y 50 columnas.  
En una primera instancia, debemos cargar el dataset a un IDE donde podamos realizar la correcta manipulación y análisis de los datos.  
Se trabajó con Visual Studio Code como IDE, y luego se instaló Python , Jupyter Notebooks y librerías como Numpy, Pandas, Matplotlib y Seaborn para realizar el análisis estadístico y gráfico de los datos.

Es importante al comenzar a trabajar con los datos, corroborar la calidad de los mismos. Revisando que los tipos de datos de cada columna sean los correctos, viendo la posible existencia de valores nulos y analizando si es factible completarlos de alguna manera para poder trabajar mejor con ellos, y acotar al máximo los datos de interes.  
  
En nuestro caso particular, necesitamos analizar los datos de los siguientes paises de interes: 'Colombia','Argentina','Chile','Mexico','Peru','Brazil'

y a su vez, nos interesaban los datos registrados a partir del 01/01/2021

Con Pandas, podemos crear un dataframe que contenga solo esos datos y a su vez realizar la corrección de nulos.  
  
Una vez que ya se tiene definido el dataframe de interes, se guarda como un nuevo archivo .csv para trabajar a futuro de manera más fácil y no cargar toda la data completa (optimizamos tiempo y recursos)

**EDA e Insights**Con la ayuda de Pandas y Numpy, podemos explorar las propiedades estadísticas del dataset, mientras que con las bibliotecas Matplotlib y Seaborn, podemos encontrar de manera gráfica aquellos insights y hallazgos que nos permitan llegar a una conclusión.

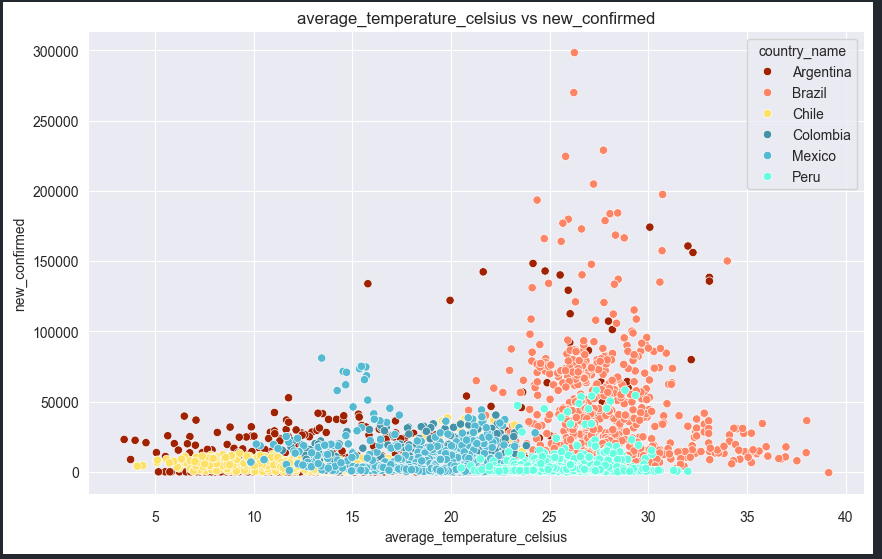
En primer lugar, se realizó una descripción estadística de todas las variables de interés del dataframe para tener un vistazo o pantallazo general, pero es bien sabido que viendo solo números, es un poco más difícil entender dichos hallazgos.

Por lo tanto pasamos a graficar para entender mejor cómo se están comportando nuestros datos.

Un primer acercamiento es crear un gráfico de correlación donde podamos ver cómo interactúan las variables entre ellas, y con ello poder ver en cuales hacer hincapié.

En un principio se hizo una rápida revisión del comportamiento de los datos de cada columna haciendo gráficos de barras, un gráfico ideal para ver comparaciones entre categorías.

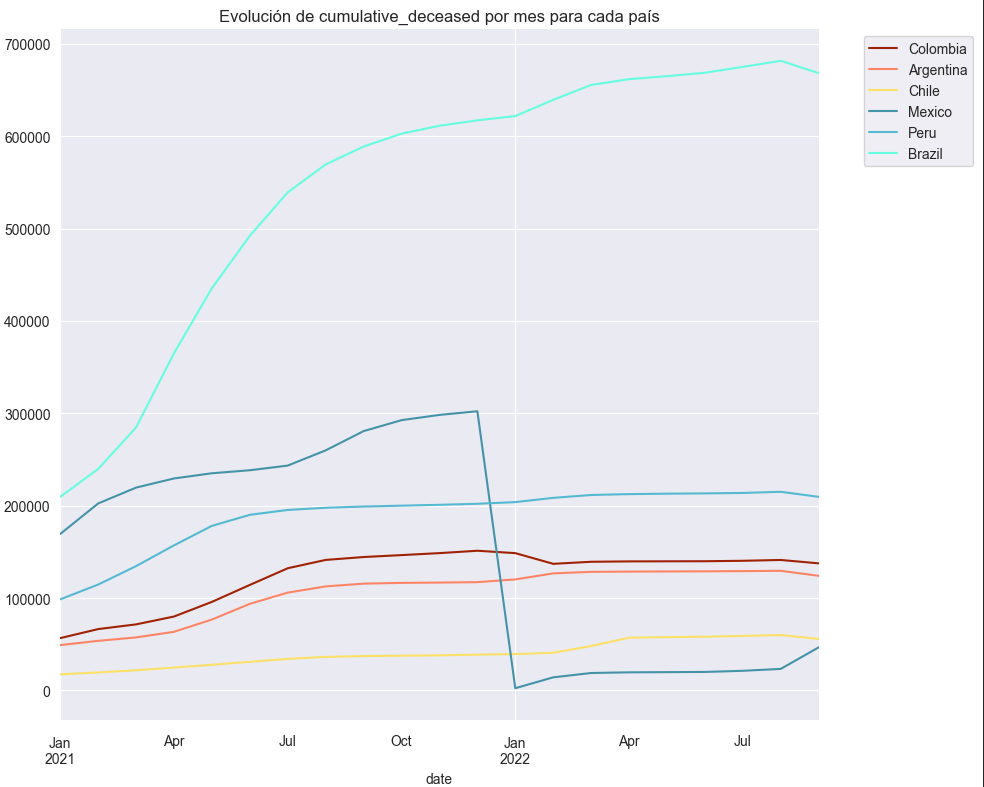
También resulto interesante ver la distribución de algunos datos y para ello realizamos los histogramas correspondientes.

Otra forma de encontrar insights es puntualizando el análisis de dos variables, por ejemplo relacionando en un gráfico de dispersión los nuevos casos confirmados de COVID-19 respecto la temperatura, esto también hace pensar que la temperatura puede tener incidencia en las muertes confirmadas, asi que se puede realizar un segundo gráfico para hallar esos valores.

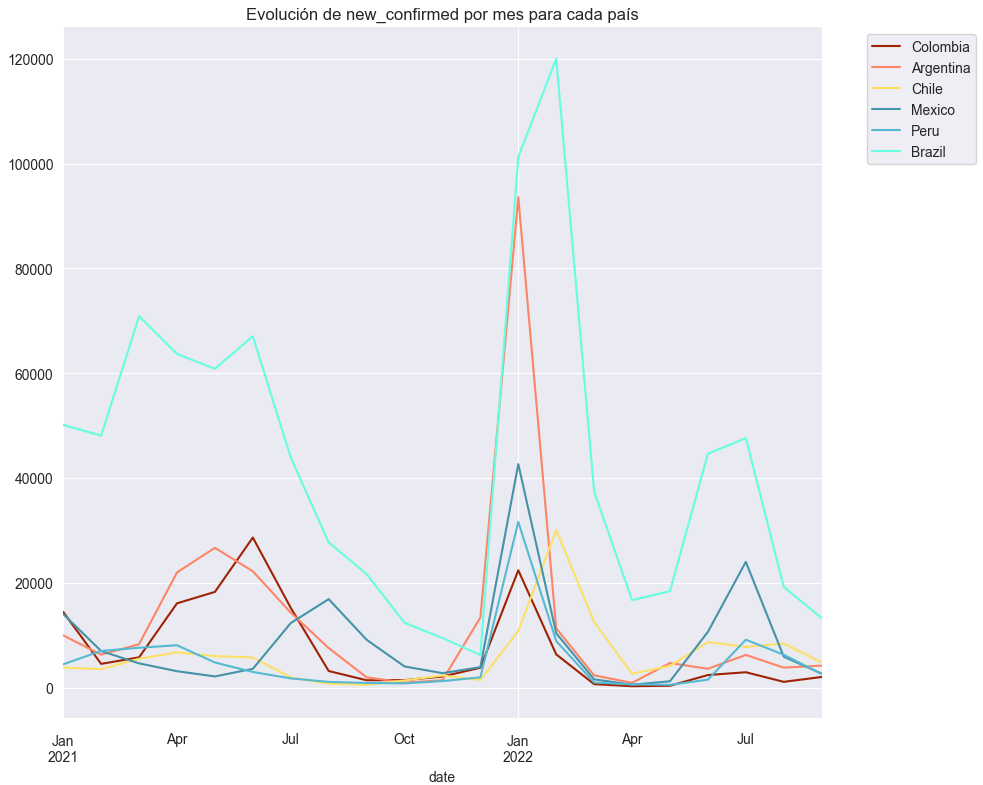
Un valor importante a conocer para nuestra toma de decisiones es conocer la cantidad de dosis de vacunas administradas por país, también podemos ver con un gráfico de línea como se comporta en el tiempo esta variable para cada país.

Otro insight importante es tener presente las muertes por pais a causa del covid-19. Se realizó la grafica correspondiente y se obtuvo que Brasil, seguido por México son los paises con mayor cantidad de muertes, seguido por Colombia, Argentina, Perú y Chile. Esto no nos está indicando que Brasil y México necesariamente manejaron mal la pandemia, sino que hay que tener en cuenta la cantidad de habitantes de cada pais, la cantidad de datos que disponemos de cada país, etc.

Por ejemplo, vemos en México un descenso brusco de la cantidad de muertos durante diciembre/enero de 2022 , y esto podría deberse bien a una buena política de acción contra la enfermedad, o puede ser que tengamos una perdidas de datos en el dataset.



Se creó un gráfico para dimensionar los nuevos casos por países a través del tiempo, y lo que se puede inferir en el mismo es que los casos venían en baja, hasta que en diciembre , por una tendencia estacional referida a las festividades (navidad/año nuevo) los casos empiezan a crecer nuevamente hasta tener un pico en enero/febrero



Es bueno notar aquí también, que todos los países reaccionaron con buenas políticas ya que los casos comienzan a descender de manera rápida para los meses siguientes.

Un valor importante que se encontró durante el EDA, es que Brasil y México son los países que manejan peor la pandemia, teniendo una gran cantidad de casos confirmados y muertes en relación al resto de países, como asi tambien se nota una cantidad de dosis de vacunas aplicadas de forma desproporcionada

Hasta este momento podríamos pensar que los países con mayores muertes y casos confirmados son los lugares ideales para poner los nuevos laboratorios. Ya que es donde se va a necesitar mayor ayuda

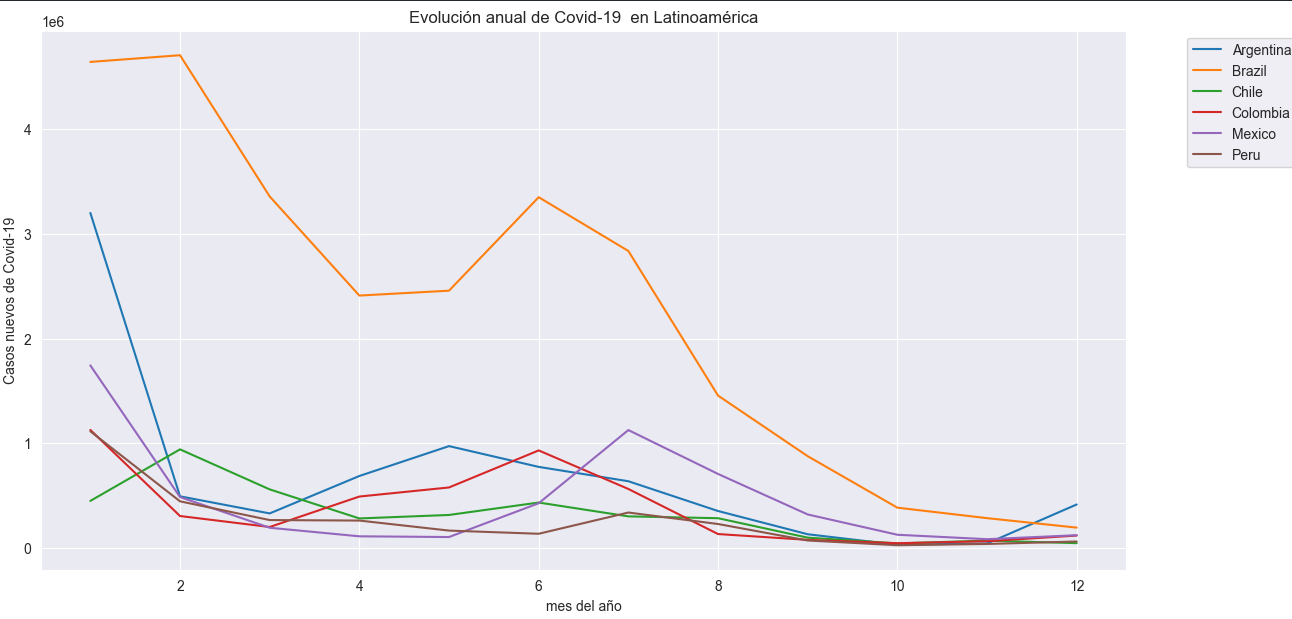
Es importante también analizar teniendo foco en la población de cada país.  
Se realizaron graficos para conocer la mortalidad femenina y masculina de cada país, y se llega a la conclusión de que los hombres tienen la mayor tasa de mortalidad, por lo tanto son los más vulnerables a la hora de enfrentarse al virus.

Podemos relacionar eso con enfermedades preexistentes o comorbilidades para ver si hay alguna enfermedad que haya causado más muertes. Con una correlación de indicadores demográficos y de salud vemos que diabetes y fumar tienen gran incidencia en la población, esto no o quita que haya otras enfermedades que también afectan.

Graficamos para ver y vemos que la diabetes es la que más muertes genera en los países seleccionados

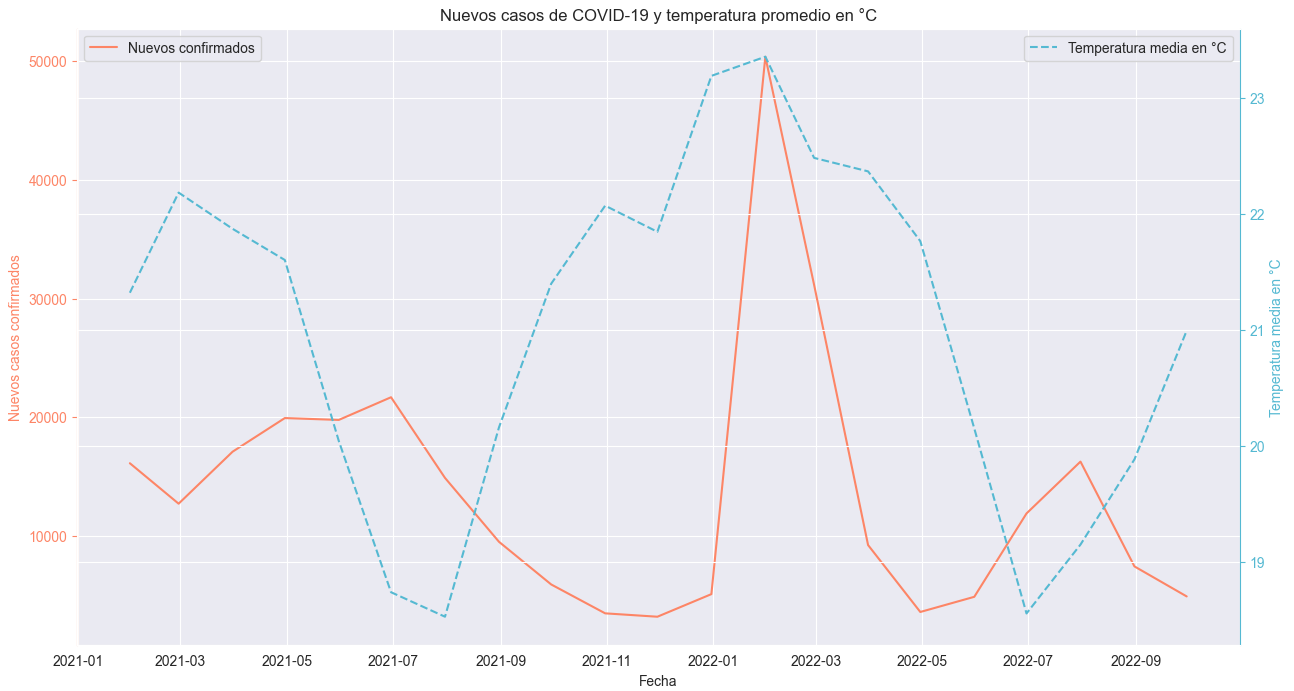
Por ahora, Brasil, México y Colombia son los países que más ayuda necesitan.

Si graficamos la evolución semanal/anual de casos nuevos por país., podemos ver que los casos nuevos y muertes nuevas han ido disminuyendo a lo largo del tiempo. Pero no indica el fin de la pandemia, porque sobre el final se ve un repunte de casos



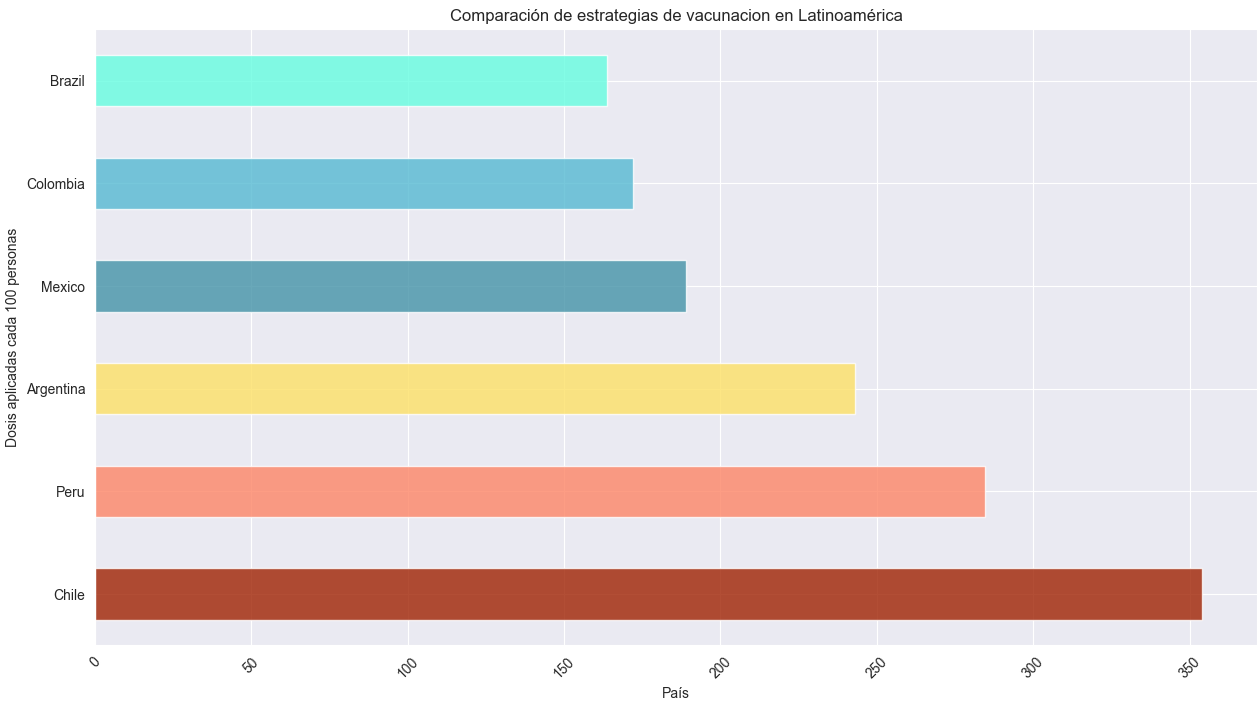
Brasil es el país con más dosis administradas, seguido de México, Argentina Colombia y Chile, pero ¿entonces hay una contradicción con lo inferido arriba de que eran los países que más necesitaban ayuda y eran los que peor estaban manejando la situación? éstos valores obtenidos acá, debemos analizarnos teniendo en cuenta el tamaño de la población de cada país y la cantidad de vacunas disponibles

Podemos también obtener insights teniendo presente factores como la temperatura, podemos ver que a medida que la temperatura promedio aumentaba los casos descienden, pero hubo un pico antes del mes de marzo pero disminuyó considerablemente debido, posiblemente, a buenas políticas de vacunación



Por último, la estrategia de vacunación de cada país nos va a ayudar a definir dónde es conveniente que el Laboratorio Biogenesys se expanda.

Vemos que Chile ha llevado la mejor estrategia de vacunación que el resto de países, y por lo contrario México, Colombia y Brasil se nota una marcada diferencia. Esto reafirma lo mencionado anteriormente, estos últimos 3 países son los que mas ayudan necesitan para enfrentar problemas futuros relacionados al Covid



Por lo tanto, luego de analizar la problemática desde varias aristas se puede llegar a la conclusión de que los países donde debe expandirse el laboratorio son Brasil, México y Colombia, ya que son los países con más casos confirmados y muertes.

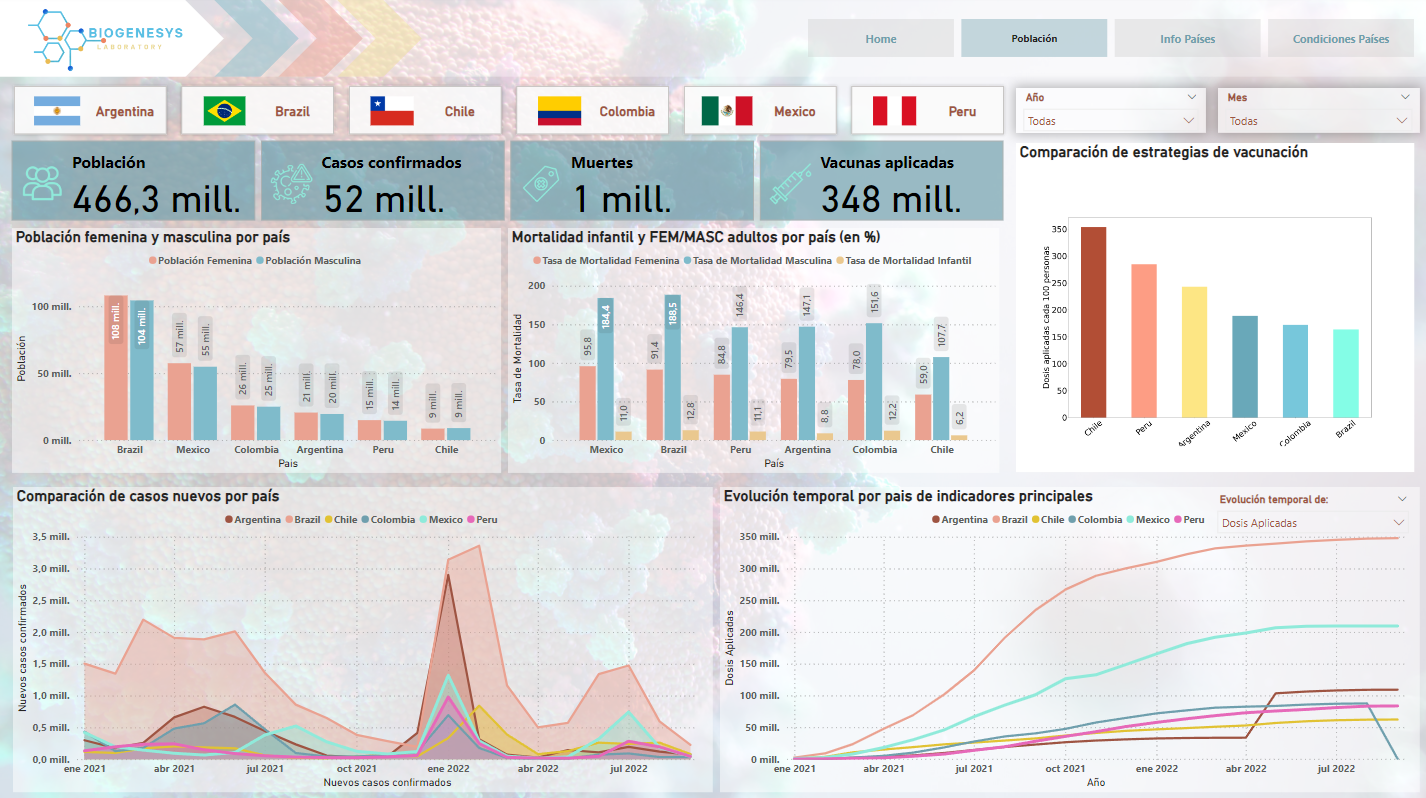
Además son los países que menos estrategia de vacunación han tenido y esto sin tener en cuenta la densidad de la población y situación socio/económica que pudiera afectar la cantidad de casos.

**Análisis del Dashboard**Tras la implementación de todas las medidas, se logra el armado del dashboard final con los gráficos que ayudan a responder a las necesidades planteadas por Laboratorios Biogenesys.

**HOME**  
En el home, vamos a poder encontrar 4 botones para moverse a la página deseada.



**PESTAÑA POBLACIÓN**

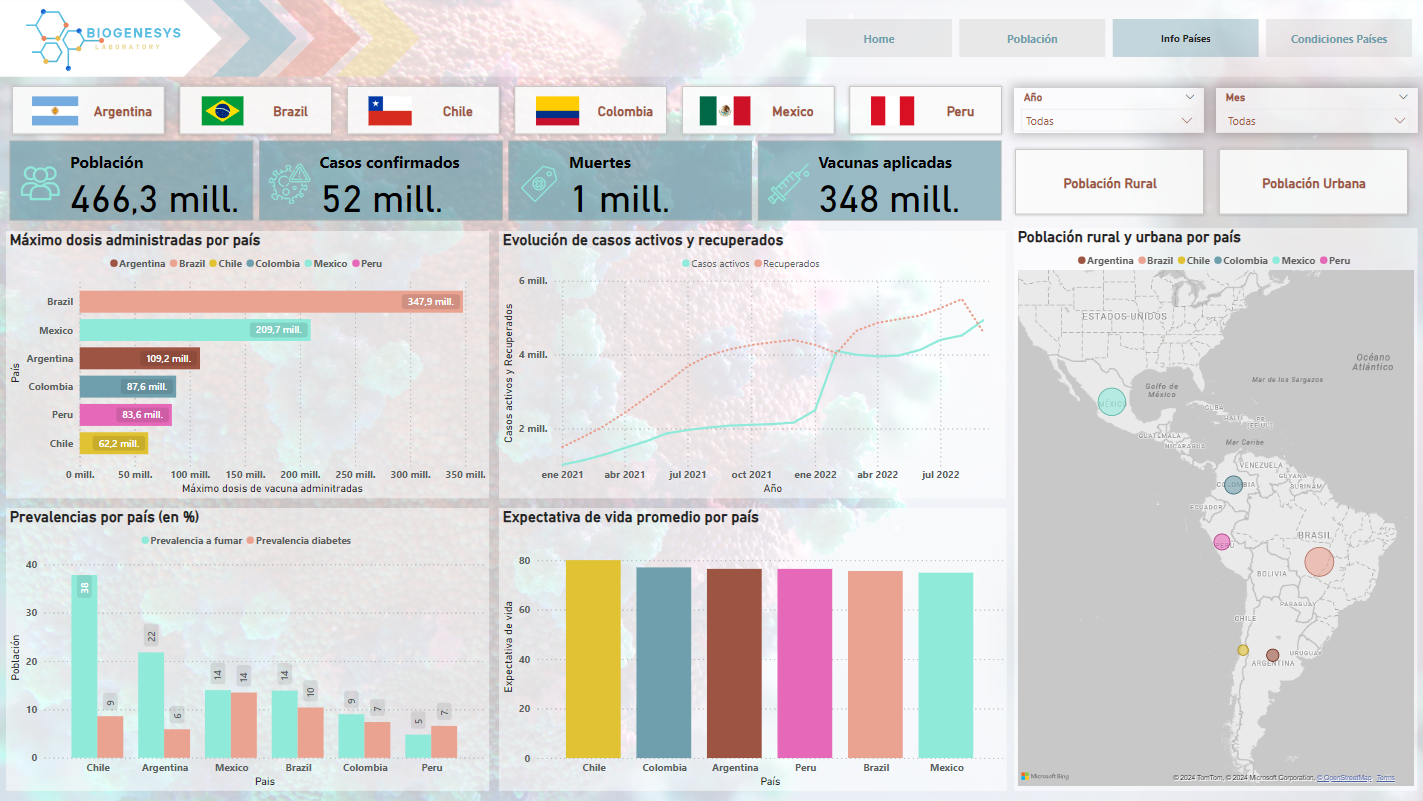


En esta página, el usuario podrá ver en un pantallazo general toda la información más relevante del COVID-19 en aspectos generales. Pudiendo filtrar la información por país, año y mes.  
  
En las tarjetas superiores, se remarcan los principales KPIs.  
  
En gráficos de columnas se puede ver la población masculina y femenina de cada país, y la mortalidad de los mismos.

En un tercer gráfico de barras, podemos ver en comparativa como fueron las estrategias de vacunación de cada país.  
  
En un gráfico de áreas se detalla de manera muy visual la comparación de nuevos casos a lo largo del tiempo por paí0s.

Por {ultimo, en el margen inferior derecho se encuentra un gráfico donde se pueden ver los indicadores principales por país (nuevos casos, nuevos recuperados, muertes, y dosis aplicadas) a lo largo del tiempo. Con el segmentador de lista desplegable se elige qué gráfico ver.

**PESTAÑA PAÍS**

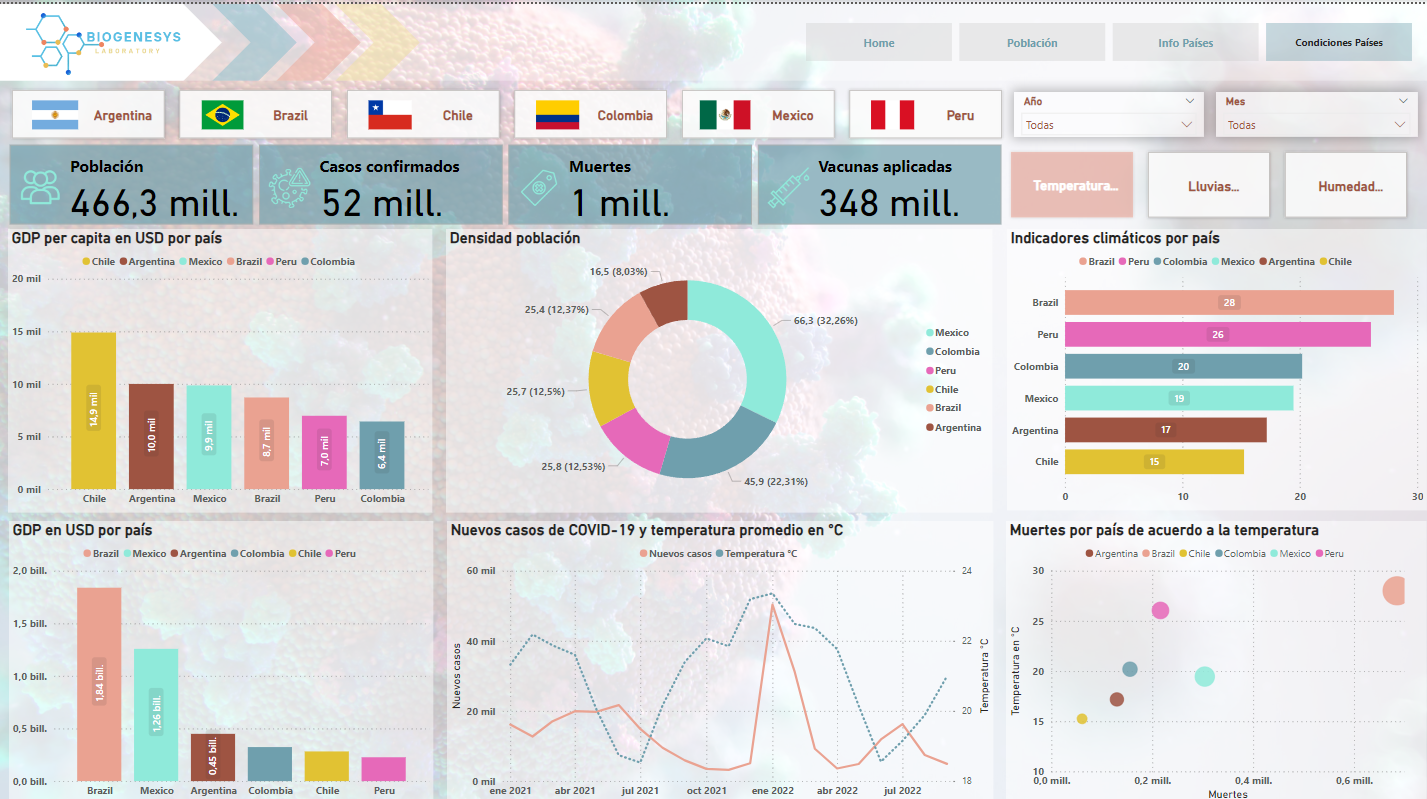


En este apartado, podemos ver datos aún más segmentados por país, o mejor dicho más descriptivo de cada uno de ellos.  
  
Se puede ver en un mapa la cantidad de población rural y población urbana.  
  
Encontramos la evolución de casos activos y casos nuevos.

Un gráfico de barra donde se muestra la expectativa de vida promedio de cada país y otro donde vemos las prevalencias de enfermedades por pais (en %)

y un gráfico muy importante es el máximo de dosis que ha aplicado cada uno de los países de interés.

Al igual que la pantalla anterior, se puede filtrar la información por país, año y mes.

**PESTAÑA CONDICIONES PAÍSES**  


En esta página, se detallan los valores más importantes para representar a las condiciones de cada país, como ser el GDP y GDP per capita, la densidad de la población.

También hay gráficos que nos ayudan a comprender la implicancia de los indicadores climáticos en el desarrollo de la pandemia, tenemos un gráfico de muertes respecto a la temperatura media por país.

Y un gráfico para ver los principales indicadores climáticos como temperatura, humedad y lluvias.

**Conclusiones y recomendaciones**Tal como se desarrolló anteriormente en el apartado de EDA E INSIGHTS , se llega a la conclusión de que Brasil, México y Colombia son los países ideales para que el Laboratorio Biogenesys se expanda.

Estos países tienen una gran población, pero no tienen una muy buena política para administrar las vacunas.

Se sugiere 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**Tras ponerme en el papel de analista de datos para resolver este informe integral para una empresa de ventas, pude entender que tan importante es comprender a los datos como “materia prima”. En el proceso de ETL, me di cuenta que no siempre vamos a obtener bases de datos brillantes, ordenadas y con una integridad de datos que nos permita arrancar rápidamente a desarrollar medidas y gráficos para responder las preguntas de negocios, sino que hay que prestar especial atención en la calidad de los datos brindados, analizarlos, transformarlos y después de eso si, volcarnos de lleno a la creación de fórmulas.

Otro punto importante que destaco, es la importancia de ser ordenado a la hora de trabajar. Si bien tenía ciertos conocimientos previos, durante la resolución de este proyecto aprendí la correcta agrupación de medidas en carpetas, y noté lo más eficiente que se vuelve el trabajo.  
  
También es importante tener tiempo suficiente para hacer un buen trabajo.