**Expansión Estratégica de Biogenesys con Python**

****

**Nombre del autor: Ana Laura Pedernera**

**Email: analaurapedernera@gmail.com**

**Cohorte: DA-F03**

**Fecha de entrega:** 1-07-2024

# Introducción

# El objetivo principal del proyecto es identificar las ubicaciones óptimas para la expansión de laboratorios farmacéuticos en Latinoamérica. Para ello, se realizará un análisis de datos de incidencia de COVID-19, tasas de vacunación y disponibilidad de infraestructuras sanitarias.

# Desarrollo del proyecto

# AVANCE 1

## Preparación y limpieza de datos

En esta sección se describe el proceso de preparación y limpieza de datos realizado para el análisis posterior. Se llevaron a cabo las siguientes tareas:

**Filtrado de datos:**

* Se filtró la información por los países y fechas de interés, eliminando registros irrelevantes para el análisis.

**Manejo de valores nulos:**

* Se analizaron los valores nulos presentes en el conjunto de datos.
* Se aplicaron diferentes técnicas de imputación de datos para completar los valores faltantes, seleccionando la técnica más adecuada en función del contexto y tipo de dato:
  + Imputación con 0: Se utilizó para columnas donde se esperaba un valor nulo o donde la ausencia de valor no tenía un impacto significativo.
  + Imputación con el último valor no nulo: Se utilizó para columnas donde era importante mantener la secuencia temporal de los datos.
  + Imputación con la mediana: Se utilizó para columnas con valores numéricos, donde la mediana era un indicador representativo de la centralidad de los datos.

Además, se transformaron los tipos de datos de las columnas a los tipos adecuados para su posterior análisis en Power BI. Se utilizaron principalmente los tipos int y float para datos numéricos.

**AVANCE 2**

Se realizó un análisis exploratorio de datos para extraer insights valiosos que sirvan de guía para la planificación estratégica de la expansión de BIOGENESYS en Latinoamérica. Se busca en esta instancia identificar patrones, tendencias y anomalías en los datos de incidencia de COVID-19, tasas de vacunación y disponibilidad de infraestructuras sanitarias, para tomar decisiones informadas sobre la ubicación de los nuevos laboratorios.

**AVANCE 3**

El objetivo central de este avance ha sido pulir y preparar los datos relacionados con la incidencia de COVID-19 para una posterior visualización avanzada. Esta visualización permitirá identificar con precisión las ubicaciones más estratégicas para la expansión de los laboratorios farmacéuticos de BIOGENESYS en Latinoamérica. Se busca obtener insights más detallados y precisos que respaldan las recomendaciones de inversión de la empresa. Se crearon visualizaciones avanzadas utilizando herramientas como Matplotlib, Seaborn o Plotly para comunicar los resultados del análisis exploratorio de manera clara y efectiva. Se utilizaron gráficos de líneas, gráficos de dispersión, mapas de calor y otras técnicas de visualización para resaltar patrones, tendencias y relaciones en los datos.

**AVANCE 4**

El objetivo de este último avance ha sido conectar Python con Power BI para diseñar un dashboard interactivo que muestre de manera efectiva los resultados del análisis de datos. Se recomienda enfocarse en mostrar visualizaciones del tablero de Power BI en lugar de las generadas en Python, ya que estas ofrecen una experiencia de usuario más interactiva y atractiva, permitiendo una mejor comprensión de los datos y la comunicación efectiva de los insights obtenidos.

Es importante aclarar que para poder leer los datos extraídos de Python, se utilizó la configuración regional “Inglés (Estados Unidos)”, esto para que el csv importado pudiera traer los datos de decimales correctamente.

# EDA e insights

Con el Gráficos de barras de países contra el resto de variables, podemos observar que los países con mayor cantidad de nuevos confirmados serían Brasil y luego Argentina, pero en cantidad de muertes y dosis administradas Argentina queda en cuarto lugar, dado que Colombia y México tienen mayores valores en estas variables. Entendemos que estos valores altos podrían deberse a que en estos países resulta haber mayor población (Brasil, Colombia y México), datos que deberíamos revisar luego en mayor medida.

Además es interesante observar que los países Chile y Argentina son los que mayor cantidad de fumadores tienen, y en México y Brasil es mayor el número de personas con diabetes, interesante dato a analizar posteriormente.

Con la Gráfica de promedio de temperatura vs nuevos casos se observa que en algunos países la temperatura no influye en la cantidad de casos confirmados, sin embargo si en otros donde hay una relación de mayor cantidad en ciertas temperaturas.

Se observa también un comportamiento similar en relación del promedio de la temperatura con las muertes, donde en Brasil se puede ver mayor cantidad de muertes en un rango de temperaturas. Además observamos un valor atípico dentro de Chile, donde deberíamos luego analizar si eliminar este valor, cambiarlo o si es un dato que es correcto, en base a lo que el cliente considere correcto por su operar diario. (En este caso no lo borraremos pero avisamos del valor encontrado)

Con la Gráfica de los casos confirmados por mes por mes por pais, Se confirma que el país con más casos es Brasil, seguido por México, Colombia, Argentina, Peru y Chile. Además, en Argentina hubo un aumento entre los meses de diciembre y febrero, podría deberse a épocas de fiestas de fin de año. A su vez en el mismo período hay una caída en México, Podría deberse el mismo al tratamiento de los nulos que se transformaron en 0.

Se concluye que de los países que mejor manejó la pandemia fue Chile, seguido por Perú, puede verse que la cantidad de muertes y casos confirmados es menor en estos países además de no ser tan desproporcionada la cantidad de vacunas administradas.

# 

# Análisis del dashboard

# Este dashboard de Power BI, compuesto por 5 hojas interactivas, ofrece una visión completa y detallada de diversos indicadores.

La hoja de Reporte General presenta indicadores totales de población, nuevos casos, muertes y dosis administradas. Permite segmentar la información por año o país para un análisis preciso. Se observan las siguientes visualizaciones: Gráfico de columnas agrupadas y líneas: muestra la evolución de la población total y las dosis totales por país. Gráfico de anillos: representa la distribución de nuevos casos por país. Gráfico de barras horizontales: ilustra la cantidad de muertes por país.

El reporte por País ofrece un panorama completo de cada país. Se observan las siguientes visualizaciones: Gráfico de columnas agrupadas: muestra la Población Total, Dosis Totales, Muertes y Nuevos Casos por país. Puede elegirse cualquiera de estos parámetros a través de una segmentación. Gráfico de dispersión: relaciona la Temperatura Promedio Celsius con los Nuevos Casos por país. Gráfico de columnas apiladas y líneas: compara la Población Rural, Población Urbana y Nuevos Casos por país.

El reporte de Población aborda indicadores de salud relevantes para la población. Se observan visualizaciones de: Gráfico de columnas agrupadas: compara la Prevalencia de Fumar y Prevalencia de Diabetes por país. Gráfico de columnas: muestra el PBI per cápita USD y las Expectativas de Vida por país. Gráfico de anillos: representa el Ratio de Mortalidad Femenina y Masculina.

Por último la Conclusión extrae conclusiones relevantes a partir del análisis de datos.

# Conclusiones y Recomendaciones

# Brasil y México en primera medida: presentan un panorama favorable para el sector salud, posicionándose como mercados con un alto potencial de crecimiento. Ambos países presentan las poblaciones más grandes de Latinoamérica. Esto se traduce en una amplia base de consumidores potenciales para productos y servicios médicos. La distribución de la población se encuentra principalmente en zonas urbanas. Este aspecto es relevante, ya que las áreas urbanas suelen tener mayor acceso a servicios de salud y mayor capacidad de pago. Brasil y México lideran en la región en cuanto a casos confirmados de COVID-19 y dosis de vacunas administradas. Esta situación refleja la necesidad de un mayor acceso a productos y servicios médicos. Si bien el PIB per cápita de ambos países se encuentra en un rango medio, este indicador no refleja necesariamente la capacidad de pago de la población. Existen segmentos de la población con mayor poder adquisitivo que pueden acceder a servicios de salud privados y productos más costosos.

# Argentina y Colombia como mercados alternativos: Si bien Chile y Perú presentan un panorama favorable en cuanto al control de la pandemia y la vacunación de su población, Argentina y Colombia se perfilan como mercados alternativos con potencial de crecimiento en el sector salud. Argentina y Colombia cuentan con poblaciones más grandes que Chile y Perú. Esto implica una base de consumidores potenciales más amplia para productos y servicios médicos. Al igual que en Chile y Perú, la población de ambos países se concentra principalmente en zonas urbanas. Este aspecto es relevante por las razones mencionadas anteriormente. La relación de población y dosis de vacunas administradas es menor en Argentina y Colombia en comparación con Chile y Perú, por lo que la demanda de productos y servicios médicos sigue siendo alta. La numerosa y creciente población, la demanda de productos y servicios médicos, y la relativa capacidad de pago de la población, son factores que sustentan esta afirmación.

# Reflexión personal

# Si bien he logrado cumplir con los objetivos principales de la tarea, se de que por el tiempo limitado no he podido aplicar todas las herramientas posibles para el análisis y la visualización de datos. El dashboard presentado cumple con su función de mostrar los resultados, pero reconozco que un diseño más interactivo y visualmente atractivo habría mejorado la experiencia del usuario y la comunicación de insights. De igual modo ha sido una experiencia sumamente enriquecedora dado que es la primera vez que tengo contacto con Python, y seguramente volveré a este proyecto más adelante para poder mejorarlo y aplicar todas las funcionalidades que creo que faltaron en esta oportunidad, en especial en la parte del dashboard en Power BI.