**Expansión Estratégica de Laboratorios Biogénesys para la investigación y vacunación de COVID-19**

**Nombre del autor:** Gabriel Lozano Lasso

**Email:** lozanogabo5800@gmail.com

**Cohorte:** DA-Ft10

**Fecha de entrega:** lunes, 27 de Enero de 2025

**Institución:** Biogenesys es una compañía farmacéutica comprometida con mejorar la salud pública global. A través de la investigación y el desarrollo de nuevos tratamientos, buscan llevar esperanza a las comunidades más vulnerables. Su enfoque en la responsabilidad social los impulsa a trabajar para que todos tengan acceso a los medicamentos que necesitan.

# Introducción

**Propósito del Proyecto:**

Biogenesys quiere encontrar los mejores lugares en Latinoamérica para construir nuevos laboratorios. Van a usar datos sobre el COVID-19, las vacunas y los hospitales para decidir dónde es más necesario y así poder ayudar a más personas.

Utilizando análisis de datos sobre la incidencia de COVID-19, tasas de vacunación y disponibilidad de infraestructuras sanitarias, se busca optimizar la respuesta ante la pandemia y postpandemia, mejorando el acceso a las vacunas en las regiones seleccionadas.

# Desarrollo del proyecto

# Primer Avance:

**Proceso de Análisis de Datos:**

Recibí una base de datos muy extensa, con 12,216,057 filas y 50 columnas, en formato CSV. Esta contenía numerosos datos faltantes, nulos y atípicos.

Se realiza un proceso de limpieza y normalización de los datos, desde su extracción, transformación y carga

1. **Limpieza de Datos:**
   * Corrección de datos faltantes y nulos, así como la gestión de valores atípicos, asegurando que los datos fueran precisos y completos.
   * Filtrado por países específicos como: Colombia, Argentina, Chile, México, Perú y Brasil, y por fechas mayores a 2021-01-01, reduciendo significativamente la cantidad de datos.
2. **Normalización de Datos:**
   * Se normalizaron los datos para facilitar su análisis.
   * Verificación de las columnas a trabajar con su respectiva cantidad de datos reducida
3. **Análisis de Mediana, Varianza y Rango:**

* **¿Qué representa la mediana?**
  + La mediana es el valor central de un conjunto de datos ordenado.
* **¿Cómo varía la dispersión de los datos en el conjunto de datos analizado, en términos de la varianza y el rango?**
  + La varianza mide la dispersión de los datos en torno a la media, indicando qué tan dispersos están los valores. Un valor alto de varianza indica que los datos están muy dispersos. El rango es la diferencia entre el valor máximo y el mínimo, proporcionando una medida de la amplitud de los datos. Un rango amplio indica una gran dispersión.
* **¿Qué nos puede indicar esto sobre la consistencia o la variabilidad de los datos en relación con la mediana?**
  + Si la varianza y el rango son altos, significa que hay una alta variabilidad en los datos, lo que puede indicar inconsistencias. Si son bajos, los datos son más consistentes y están más agrupados alrededor de la mediana. Esto puede ser crucial para determinar la fiabilidad de los datos y su interpretación.

# Segundo Avance:

Se realizaron gráficos a partir de las diversas herramientas proporcionadas por Python, descritas así:

* Pandas y Numpy: Propiedades estadísticas de los datos
* Matplotlib y Seaborn: Para la visualización de datos.

Algunos de los gráficos que se realizaron fueron:

* Histogramas y gráficos de densidad: Para entender la distribución de la incidencia de COVID-19 y las tasas de vacunación.
* Gráficos de barras: Para comparar diferentes regiones.
* Mapas de calor: Para identificar correlaciones entre diferentes variables.
* Gráficos de dispersión: Para explorar posibles relaciones entre las variables.

# Tercer Avance:

Se realiza análisis exploratorio de datos. Enfocándose principalmente en los datos relacionados del Covid-19 a partir de series temporales para comprender la evolución de los diversos países y como afrontaron la pandemia. Algunas de las gráficas generadas fueron

Se generaron las siguientes gráficas, entre otras:

* Evolución de Casos semanales a lo largo del tiempo: Evalúa el desempeño del covid-19 en el tiempo
* Progreso de vacunación en el tiempo por países: Permite conocer que países tuvieron mejor desempeño en el control del covid-19
* Relación entre la temperatura y los casos de covid-19: permite encontrar relación entre a mayor temperatura los casos de covid-19 incrementan
* Impacto de la Urbanización en la Propagación del COVID-19: Para analizar cómo la densidad poblacional influye en la propagación del virus.
* Evolución Anual de Casos Nuevos: Para observar la evolución de la pandemia a lo largo del tiempo.
* Distribución de la Población por Edad en Países: Para comprender cómo las características demográficas pueden afectar la vulnerabilidad ante el COVID-19.
* Comparación de Estrategias de Vacunación en América Latina: Para analizar y comparar las diferentes políticas de vacunación implementadas.
* Evolución del Número de Muertes Diarias: Evalúa la mortalidad diaria a lo largo del tiempo.
* Prevalencia de Condiciones Preexistentes en Países con Altas y Bajas Tasas de Mortalidad: Para identificar posibles factores de riesgo.
* Variación Mensual de Casos y Muertes: Para observar fluctuaciones mensuales en los datos.
* Comparación entre casos activos y recuperados por país: Evalúa el desempeño de los países ante los nuevos casos registrados

Se determinan conclusiones que son encontradas a partir de los gráficos lo cual permite una mayor visual sobre el desarrollo del virus covid-19 y el accionar de los diversos países antes dicha pandemia.

# Cuarto Avance:

Una vez generado los gráficos, se integra toda la información y es trasladada a Power Bi para su respectivo análisis. Esto permite:

1. Crear Paneles Interactivos: Que facilitan la visualización dinámica de los datos y permiten a los usuarios explorar los resultados de manera más intuitiva.
2. Análisis Comparativo en Tiempo Real: Compara escenarios y métricas diferentes de manera ágil.

El dashboard realizado en Power BI incluye las siguientes secciones:

1. Portada: Introducción al proyecto y a sus diferentes secciones divididas por botones para su mayor eficiencia
2. Vacunación: Análisis poblacional en áreas rurales y urbanas. Tasa de mortalidad infantil
3. Enfermedades: Relación entre enfermedades prevalentes y su influencia en el covid-19.
4. Países: Reporte de casos por países en el tiempo y su distribución en mapas.
5. Interfaz de usuario gráfica, Texto

   Descripción generada automáticamenteGráfico, Aplicación

   Descripción generada automáticamenteImagen que contiene Interfaz de usuario gráfica

   Descripción generada automáticamenteConclusiones: Análisis sobre los gráficos y respuesta a la pregunta inicial de ¿Dónde debería BYOGENESIS expandir sus servicios?

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza mediaGráfico, Gráfico de barras

Descripción generada automáticamente

**Conclusiones y Recomendaciones**

Las conclusiones encontradas a partir de los resultados son los siguientes:

1. **Identificación de Ubicaciones Óptimas**: Brasil, México y Argentina se identificaron como los países con mayor necesidad y potencial para la expansión de laboratorios debido a su alta incidencia de COVID-19 y menor cobertura de vacunación.
2. **Prioridad:** Brasil es el país con mas casos reportados, muertes y uso de las vacunas. Es una prioridad atender dada su densidad poblacional
3. **Estrategias de Vacunación**: En términos de administración de vacunas, si bien Brasil lidera con el mayor número de dosis administradas promedio, su performance por persona fue de la peor de la región. Le siguieron Colombia y México, con lo cual se detecta una necesidad de mejorar la infraestructura de salud que permita ampliar el marco de cobertura.

**Reflexión Personal**

A lo largo de este proyecto he adquirido habilidades técnicas relacionadas al análisis de datos a partir de código en Python, desde la carga de los datos, pasando por limpieza y normalización hasta el uso de herramientas avanzadas como matplotlib o seaborn para la generación de gráficos. En la interpretación de hizo uso de Power bi para la toma de decisiones estratégicas. Si tuviera que volver a empezar este proyecto, me enfocaría aún más en la automatización de procesos de limpieza de datos y en la implementación de funciones de orden superior para optimizar el análisis. Esto podría acelerar el proceso y mejorar la eficiencia del análisis de grandes volúmenes de datos.

**EXTRA CREDIT**

Aunque no implementé todas las mejoras en esta etapa, investigué sobre funciones de orden superior que pueden ser muy útiles en futuros proyectos. Algunas de estas funciones incluyen:

1. Funciones de Map: Para aplicar operaciones a todos los elementos de un iterable.
2. Funciones de Reduce: Para realizar agregaciones en una lista.
3. Funciones Lambda: Para definiciones rápidas y simples de funciones.
4. Funciones de Iteración (Itertools): Para manipulaciones eficientes de iterables.

Estas mejoras potenciales contribuirían a hacer el análisis de datos aún más robusto y eficiente.