**Expansión Estratégica de Biogenesys con Python**

**Nombre del autor: Herna Roman**

**Email: hernanroman25@gmial.com**

**Cohorte: DA\_PT01**

**Fecha de entrega:** 29/07/2024

**Institución:**

# Introducción

# BIOGENESYS Me ha contratado como Data Analyst para que realice un estudio que ayudará en su estrategia de expansión en Latinoamérica. El propósito del proyecto es identificar ubicaciones estratégicas para la expansión de laboratorios farmacéuticos y centros de vacunación en Latinoamérica. El objetivo principal es mejorar la respuesta a la pandemia y la post pandemia de COVID-19, optimizando el acceso a las vacunas y a las infraestructuras sanitarias. Esto se logrará mediante el análisis de datos de incidencia de COVID-19, tasas de vacunación y la disponibilidad de infraestructuras sanitarias en Colombia, Argentina, Chile, México, Perú y Brasil.

# Desarrollo del proyecto

#### Avance 1: Carga y Transformación de Datos

# Métodos y Procesos:

# Lectura y comprensión del archivo Readme.txt para entender la estructura del dataset.

# Creación de un notebook en Python para la carga del archivo CSV.

# Verificación de la cantidad de registros y columnas del dataset.

# Filtrado de datos para los países específicos (Colombia, Argentina, Chile, México, Perú y Brasil) y fechas mayores a 2021-01-01.

# Limpieza preliminar de datos eliminando filas con valores faltantes y ajustando los tipos de datos.

# Análisis de las características básicas del dataset.

# Cálculo de estadísticas descriptivas como la mediana, varianza y rango mediante funciones personalizadas.

# Exploración del uso de funciones de orden superior para una manipulación eficiente de los datos.

# Conclusión:

# Se logró una comprensión clara de la estructura del dataset y se prepararon los datos para el análisis exploratorio. La limpieza preliminar y la transformación de datos permitieron establecer una base sólida para los análisis subsecuentes.

#### Avance 2: Análisis Estadístico y Visualización de Datos

# Métodos y Procesos:

# Importación de librerías necesarias: Pandas, Numpy, Matplotlib y Seaborn.

# Exploración de las propiedades estadísticas del dataset mediante el cálculo de medidas de tendencia central, dispersión y correlaciones entre variables.

# Creación de visualizaciones:

# 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 variables.

# Gráficos de dispersión para explorar posibles relaciones entre variables.

# Conclusión:

# El análisis estadístico y las visualizaciones proporcionaron insights valiosos sobre la distribución y relación de las variables. Esto ayudó a identificar patrones y áreas prioritarias para la expansión.

#### Avance 3: Análisis Temporal

# Métodos y Procesos:

# Análisis exploratorio detallado de series temporales para comprender la evolución de elementos específicos del conjunto de datos.

# Identificación de tendencias, estacionalidad y patrones temporales relevantes mediante análisis de autocorrelación y descomposición de series temporales.

# Generación de gráficas avanzadas:

# Evolución de casos activos vs. recuperados.

# Tasa de crecimiento de casos.

# Relación entre la cobertura de vacunación y la reducción de casos.

# Conclusión:

# El análisis temporal reveló tendencias y patrones significativos, y útiles para la toma de decisiones estratégicas.

#### Avance 4: Conexión con Power BI y Creación de Dashboards

# Métodos y Procesos:

# Importación del conjunto de datos final en Power BI.

# Creación de un dashboard interactivo que muestra los resultados del análisis de datos.

# Inclusión de visualizaciones interactivas:

# Incidencia de COVID-19.

# Cobertura de vacunación.

# Correlaciones e infraestructura

# Comparación de visualizaciones estáticas y dinámicas.

# Integración de gráficos generados con Python en Power BI.

# Conclusión:

# La conexión de Python con Power BI permitió trasladar el análisis técnico a una plataforma accesible para los directivos. Los dashboards interactivos facilitan la exploración de datos y la identificación de áreas prioritarias para la expansión.

# EDA e insights

# Análisis Exploratorio de Datos (EDA)

# Se realizó un análisis estadístico utilizando Pandas y Numpy para explorar las propiedades del dataset. Se calcularon medidas de tendencia central, dispersión y correlaciones entre las variables. Los principales insights obtenidos incluyen:

# Distribución de la Incidencia de COVID-19: La incidencia de COVID-19 varía significativamente entre los países. Brasil y México mostraron las tasas más altas de casos confirmados.

# Cobertura de Vacunación: Chile y Perú lideran en términos de tasas de vacunación, lo que sugiere una fuerte infraestructura de salud pública y campañas de vacunación efectivas.

# Correlaciones: Se encontraron correlaciones significativas entre la densidad de población y la incidencia de COVID-19, así como entre la cobertura de vacunación y la reducción de casos.

# Visualizaciones

# Se utilizaron Matplotlib y Seaborn para crear visualizaciones que representan los hallazgos. Estas incluyen histogramas, gráficos de densidad, gráficos de barras y mapas de calor. Estas visualizaciones ayudaron a identificar patrones y tendencias clave en los datos.

# 

# Análisis del dashboard

# Navegación del Dashboard en Power BI

# El dashboard creado en Power BI permite una exploración interactiva de los datos. Incluye varias páginas que muestran diferentes aspectos del análisis:

# Portada: Muestra una portada, con la botonera de las siguientes hojas.

# Incidencia de COVID-19: Gráficos de líneas y mapas de calor que muestran la evolución de casos activos, recuperados y muertes por país.

# Cobertura de Vacunación: Gráficos de líneas y barras que ilustran la administración de dosis y la cobertura de vacunación por país.

# Correlaciones e infraestructura: Gráficos de dispersión y matrices de correlación que exploran las relaciones entre diferentes variables.

# El dashboard facilita la identificación de áreas prioritarias para la expansión mediante la visualización de datos de manera clara y accesible. Los usuarios pueden interactuar con los gráficos y filtros para explorar diferentes escenarios y realizar análisis detallados.

# Conclusiones y Recomendaciones

**Identificación de Ubicaciones Óptimas para la Expansión de Laboratorios Farmacéuticos**

Tras analizar la incidencia de COVID-19, la cobertura de vacunación, las correlaciones entre variables clave y la infraestructura sanitaria disponible en diferentes países de Latinoamérica, se puede concluir que:

1. **Brasil y México** son candidatos ideales para la expansión de laboratorios farmacéuticos debido a su alta incidencia de casos y la necesidad urgente de infraestructura sanitaria adicional. La alta población y las áreas densamente pobladas indican un gran mercado y una necesidad crítica de servicios médicos.
2. **Chile** destaca por su alta tasa de vacunación, lo que sugiere un sistema de salud relativamente eficiente y una buena infraestructura para la implementación de nuevas instalaciones médicas.
3. **Colombia y Argentina** presentan oportunidades significativas debido a la moderada incidencia de casos y la creciente necesidad de mejorar la infraestructura sanitaria y los recursos médicos.
4. **Perú** muestra una necesidad particular en la mejora de la capacidad hospitalaria y la disponibilidad de recursos médicos, lo que hace que sea un destino estratégico para la expansión.

La expansión en estos países no solo mejorará la capacidad de respuesta ante pandemias futuras, sino que también fortalecerá la infraestructura sanitaria regional, promoviendo un mejor acceso a servicios médicos esenciales y mejorando la salud pública en general. La implementación de laboratorios en estas áreas contribuirá significativamente a la preparación y respuesta ante emergencias sanitarias, promoviendo una mayor equidad en la atención médica en toda Latinoamérica.

# 

# Reflexión personal

# Aprendizajes y Habilidades Adquiridas

# Este proyecto ha permitido profundizar en habilidades de análisis de datos, limpieza y transformación de datos, visualización interactiva y comunicación de resultados. He aprendido a utilizar Python con las librerías Pandas y Numpy para explorar las propiedades del dataset también Matplotlib y Seaborn para crear visualizaciones que representan los hallazgos y a aplicar técnicas avanzadas de análisis de series temporales.

# Reflexión sobre el Proyecto

# Si tuviera que empezar de nuevo, enfocaría más tiempo en la etapa de recopilación de datos para asegurar una mayor variedad de fuentes y variables. También integraría más técnicas de machine learning para hacer predicciones y análisis más profundos.

# En resumen, este proyecto ha demostrado cómo un análisis basado en datos puede guiar decisiones estratégicas para la expansión de laboratorios y centros de vacunación, optimizando la respuesta a la pandemia y mejorando el acceso a las vacunas en Latinoamérica.

# 