**BIOGENESYS**

**Nombre del autor: Ezequiel Caballero**

**Email: caballeroezze@gmail.com**

**Cohorte: DA-FT07**

**Fecha de entrega:** 21/10/2024

**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

# BIOGENESYS te ha contratado como Data Analyst para que realices un estudio que ayudará en su estrategia de expansión en Latinoamérica específicamente en Colombia, Argentina, Chile, México, Perú y Brasil. Este proyecto es vital para que estén preparados y puedan tener una respuesta rápida ante cualquier situación futura que pueda surgir, la directiva ha propuesto que para poder hacer esta inversión necesitan ubicar regiones y recolectar datos que sean de importancia y que les ayuden a tomar decisiones. Por lo tanto, tú deberás utilizar distintas herramientas vistas en la carrera para poder ayudar a los directivos a tomar decisiones en este proceso.

# Los objetivos son concretos y medibles:

# Realizar un análisis exploratorio de datos sobre la incidencia de COVID-19 y otros factores relevantes, identificando tendencias y oportunidades mediante estadísticas, mediciones y visualizaciones.

# Aplicar técnicas de limpieza de datos para asegurar la calidad de los datos, facilitando análisis y decisiones estratégicas confiables.

# Mejorar el acceso a los datos mediante operaciones eficientes de extracción, transformación y carga (ETL), aumentando la eficacia del análisis y la toma de decisiones.

# Desarrollar dashboards interactivos con visualizaciones eficientes, permitiendo explorar datos desde múltiples perspectivas para una toma de decisiones informada y estratégica.

# Se espera que utilices los datos proporcionados para ofrecer insights valiosos sobre posibles ubicaciones estratégicas para futuros laboratorios y centros de vacunación. Tu análisis debe abordar aspectos como la demanda de vacunas, la logística de distribución y la infraestructura sanitaria existente. Este análisis de tendencias y demografía es fundamental para comprender el entorno del mercado en los países de interés.

# Desarrollo del proyecto

# 1. Carga de datos: Se cargaron datos desde un archivo CSV llamado `data\_latinoamerica.csv`.

# 2. Dimensiones del dataset: El dataset contiene 12,216,057 filas y 50 columnas.

# 3. Distribución de países: Se revisó el conteo de registros por país, mostrando algunos países como Brasil, México, Perú, Colombia y Argentina entre otros.

# 4. Primeras filas del dataset: Las columnas incluyen datos sobre:

# - Confirmados y fallecidos nuevos y acumulados por día.

# - Dosis de vacunas administradas.

# - Información demográfica y de salud como población, esperanza de vida, mortalidad por comorbilidades, etc.

# Metodología de Recopilación y Selección de Datos

# 1. Origen de los datos: Los datos han sido cargados desde un archivo CSV (`data\_latinoamerica.csv`), que contiene información de diferentes países de América Latina. Las columnas presentes sugieren que los datos abarcan aspectos relacionados con la pandemia de COVID-19, tales como:

# - Nuevos casos confirmados y fallecidos.

# - Acumulados de casos confirmados, fallecidos y recuperados.

# - Dosis de vacunas administradas.

# - Variables demográficas como población, área urbana/rural, esperanza de vida, tasas de mortalidad, etc.

# 2. Descripción de los datos: El dataset tiene más de 12 millones de registros (12,216,057 filas) y 50 columnas. La granularidad de los datos parece ser diaria, con entradas para cada país a lo largo de un período de tiempo.

# 3. Selección de datos:

# - Se seleccionaron países latinoamericanos para realizar el análisis, dado que las filas contienen información desagregada por país.

# - Se incluyeron variables demográficas, de salud pública y epidemiológicas.

# Limpieza y Transformación de Datos

# 1. Tratamiento de valores nulos:

# - En el conjunto de datos, se observan columnas con valores faltantes. Por ejemplo, hay muchos valores faltantes en las columnas de "cumulative\_recovered" y "new\_recovered". Esto puede deberse a la falta de disponibilidad o reporte incompleto de información en ciertos países.

# - Para abordar los valores nulos, se puede optar por:

# - Rellenar valores nulos con interpolaciones (si se espera continuidad en los datos).

# - Eliminar las filas o columnas si hay demasiados valores faltantes que afecten la calidad del análisis.

# 2. Transformación de los tipos de datos:

# - Algunas columnas como las fechas o los números pueden requerir conversiones a los tipos de datos correctos (por ejemplo, fechas deben ser de tipo `datetime`).

# - Se aplican transformaciones como la normalización de datos (en caso de que se vayan a usar métodos estadísticos) y la conversión de valores categóricos a numéricos si es necesario.

# 3. Filtrado de columnas y registros irrelevantes:

# - Es posible que haya columnas que no sean directamente relevantes para el análisis, como algunas variables altamente correlacionadas o redundantes.

# - Los registros con errores de captura o datos que no contribuyan significativamente al análisis podrían eliminarse.

# EDA e insights

1. Distribución de Casos y Muertes por País

* Insight: Los gráficos mostraron una distribución desigual de los casos confirmados y muertes entre los países analizados (Argentina, Brasil, Chile, Colombia, México y Perú). Brasil y México experimentaron un número significativamente mayor de casos confirmados y muertes en comparación con otros países, lo que sugiere una mayor severidad en el impacto del COVID-19.
* Análisis: Los histogramas y gráficos de barras indicaron una mayor concentración de casos y fallecimientos en ciertos períodos, alineados con las olas de la pandemia. Esto es clave para entender los momentos críticos en cada país y cómo respondieron ante picos de contagios.

2. Tendencias Temporales en Casos Confirmados y Muertes

* Insight: Los gráficos de líneas mostraron una clara estacionalidad en el número de nuevos casos confirmados y fallecimientos en todos los países. Los picos más pronunciados se registraron entre mediados de 2021 y principios de 2022, coincidiendo con las olas más severas de la pandemia en la región.
* Análisis: La visualización temporal permitió identificar las fases de aumento y reducción de casos, sugiriendo una relación con la implementación de políticas de confinamiento, distanciamiento social, y el despliegue de campañas de vacunación.

3. Impacto de las Políticas de Vacunación

* Insight: El análisis de la vacunación mostró que países como Chile y Argentina lograron avances significativos en sus programas de vacunación en comparación con Brasil y Perú, donde la distribución de vacunas fue más lenta. Esto se correlacionó con una reducción más rápida en los casos confirmados en los primeros países.
* Análisis: Los gráficos comparativos de nuevos casos confirmados y el porcentaje de vacunación revelaron una relación inversa en países con políticas de vacunación más exitosas. A medida que la vacunación avanzaba, los casos nuevos tendían a disminuir, confirmando la efectividad de la vacunación en controlar la pandemia.

4. Relación entre Indicadores Demográficos y Pandemia

* Insight: Los datos demográficos mostraron una relación entre la densidad de población, esperanza de vida, y las tasas de mortalidad por comorbilidades con la propagación del virus y las muertes. Brasil y México, con grandes áreas urbanas densamente pobladas, sufrieron un mayor impacto en términos de contagios.
* Análisis: El cruce de estos factores demográficos con los datos de casos y muertes indica que áreas urbanas y densamente pobladas son más propensas a una mayor propagación del virus, mientras que la baja esperanza de vida y comorbilidades contribuyeron a mayores tasas de mortalidad.

5. Correlación entre Enfermedades Preexistentes y Mortalidad

* Insight: Los países con mayores tasas de comorbilidades, como la diabetes y el tabaquismo, también presentaron mayores tasas de mortalidad por COVID-19. Perú y Colombia destacaron por tener una mayor mortalidad relacionada con comorbilidades, lo que puede haber agravado el impacto del virus en sus poblaciones.
* Análisis: La correlación entre estos factores sugiere la importancia de las condiciones preexistentes en la vulnerabilidad frente al COVID-19, lo que refuerza la necesidad de focalizar campañas de vacunación y atención médica en poblaciones con mayores riesgos de salud.

6. Comparación de la Infraestructura Sanitaria

* Insight: El análisis de la infraestructura sanitaria mostró una disparidad significativa entre los países. Chile y Argentina tienen un mayor número de médicos y enfermeros por cada 1,000 habitantes en comparación con Perú y Brasil, lo que facilitó una mejor respuesta ante la pandemia.
* Análisis: Esta disparidad en la infraestructura médica refleja cómo los países con más recursos pudieron manejar mejor las olas de contagios y garantizar la atención adecuada a sus poblaciones. Esto es crucial para la toma de decisiones sobre la expansión de laboratorios y centros médicos en regiones con infraestructura deficiente.

7. Tasa de Mortalidad Infantil y Necesidades Sanitarias

* Insight: Las tasas de mortalidad infantil y los niveles de pobreza fueron más altos en países como Perú y Colombia, lo que sugiere que estas naciones enfrentan mayores desafíos en su sistema de salud pública. Estos factores aumentan la necesidad de una mayor inversión en infraestructura sanitaria y programas de atención médica.
* Análisis: Estas variables son clave para identificar ubicaciones óptimas para la expansión de laboratorios farmacéuticos, ya que indican una mayor necesidad de servicios médicos especializados y acceso a medicamentos y vacunas.

# ANÁLISIS DEL DASHBOARD

# El dashboard presentado en el archivo se enfoca en el impacto del COVID-19 y la vacunación en varios países de Latinoamérica, proporcionando una visión clara de la situación a lo largo del tiempo. A continuación, detallo la explicación del dashboard y algunos insights clave que puedes extraer de él:

# Impacto del COVID-19 por País y Fecha

# Gráfico: Gráficos de líneas que muestran la evolución de los nuevos casos confirmados y los fallecidos a lo largo del tiempo en países como Argentina, Brasil, Chile, Colombia, México y Perú.

# Insights:

# Las curvas permiten identificar picos importantes en los contagios y las muertes en distintos momentos entre 2021 y 2022, coincidiendo con las diferentes olas de la pandemia.

# Brasil y México parecen tener un mayor número de fallecidos y casos confirmados en comparación con otros países.

# Impacto de la Vacunación

# Gráfico: Comparación entre el número de nuevos casos confirmados y el porcentaje de población vacunada a lo largo del tiempo.

# Insights:

# Se observa una tendencia decreciente en los casos confirmados conforme aumenta el porcentaje de personas vacunadas en países como Chile y Argentina.

# El gráfico muestra cómo la campaña de vacunación ha sido efectiva para disminuir el número de casos en la región.

# Distribución de la Población por Rangos de Edad

# Gráfico: Barras que muestran la población distribuida por grupos de edad (0 a 9, 10 a 19, etc.).

# Insights:

# La población se concentra en los grupos de edad entre 20 y 49 años, lo cual puede ser útil para orientar estrategias de vacunación o atención sanitaria a estas edades más vulnerables.

# Los grupos de mayores de 60 años, aunque más reducidos, pueden ser priorizados por su vulnerabilidad frente a la enfermedad.

# Indicadores Socioeconómicos

# Datos clave:

# PBI (Producto Bruto Interno) y PBI per cápita: Muestran la situación económica de los países (ej. Brasil tiene un PBI significativo).

# Índice de Desarrollo Humano (IDH): Valores como 0.747 destacan el nivel de desarrollo de estos países, que influye en cómo gestionan la pandemia.

# Insights:

# El IDH puede correlacionarse con una mejor infraestructura de salud y capacidad de respuesta ante el COVID-19, lo que podría explicar las diferencias en el número de casos y fallecidos.

# Factores de Salud y Riesgos

# Gráfico: Comparación de indicadores como el porcentaje de fumadores (17%) y la prevalencia de diabetes (9%).

# Insights:

# Estos factores pueden aumentar la mortalidad por comorbilidades. Países con tasas más altas de fumadores o diabetes podrían tener mayor vulnerabilidad frente al COVID-19.

# Distribución de la Vacunación

# Gráfico: Suma de las dosis de vacunas administradas en cada país.

# Insights:

# Chile y Argentina aparecen como líderes en términos de vacunación en relación con su población, mientras que otros países como México y Perú parecen haber tenido un ritmo más lento.

# Este tipo de análisis puede ayudar a correlacionar la cantidad de vacunas administradas con la reducción de casos confirmados.

# Conclusiones y Recomendaciones

# El análisis del impacto de COVID-19 y las políticas de vacunación en Latinoamérica ha permitido identificar factores clave que influyen en la expansión de laboratorios farmacéuticos. Con base en los datos visualizados en el dashboard y en el análisis del notebook, las conclusiones generales del proyecto son las siguientes:

# Conclusión General

# El análisis ha revelado disparidades significativas entre países de Latinoamérica en cuanto a la respuesta al COVID-19, las políticas de vacunación, y la infraestructura sanitaria. Los países con mayor inversión en salud y programas de vacunación eficaces lograron contener mejor la pandemia, mientras que aquellos con una infraestructura médica deficiente y altos niveles de pobreza enfrentaron mayores desafíos.

# Conclusiones Estratégicas:

# 1. Políticas de Vacunación:

# Países como Chile y Argentina mostraron una rápida respuesta de vacunación, lo que ayudó a reducir significativamente los casos confirmados y fallecimientos. Estos países pueden servir como modelos a seguir en términos de eficiencia en las políticas de vacunación.

# Brasil y México presentaron un alto número de casos y fallecimientos, pero con una menor efectividad en la vacunación en comparación con otros países, lo que sugiere un área de oportunidad para mejorar la distribución de vacunas.

# 2. Inversión en Salud:

# La disponibilidad de personal médico y de enfermería por cada 1,000 habitantes muestra que los países con mejores infraestructuras sanitarias, como Chile, tienen mejores resultados en el manejo de la pandemia.

# En países con menor inversión en salud y mayor prevalencia de comorbilidades, como Perú y Colombia, existe una mayor necesidad de refuerzo en la infraestructura médica.

# 3. Densidad de Población y Factores Socioeconómicos:

# Países con una alta densidad de población, como Brasil y México, enfrentaron mayores dificultades en la implementación de políticas de distanciamiento social y control de la pandemia. Además, los problemas socioeconómicos como la pobreza y el acceso limitado a servicios de salud complican aún más la situación.

# Perú y México, con sus desafíos económicos, requieren políticas integradas que incluyan educación sanitaria y mejoras en la infraestructura.

# 4. Identificación de Ubicaciones Óptimas para la Expansión de Laboratorios:

# Las regiones con alta prevalencia de enfermedades crónicas, baja cobertura de vacunación y sistemas de salud deficientes (como Perú y Colombia) representan áreas estratégicas para la expansión de laboratorios farmacéuticos.

# Países como Brasil y México, que presentan una gran demanda de servicios de salud debido a sus altas tasas de mortalidad y casos confirmados, también son candidatos para la expansión de laboratorios.

# Recomendaciones:

# 1. Fortalecer la infraestructura sanitaria en áreas críticas:

# Invertir en regiones con baja cobertura de servicios médicos, como áreas rurales de Perú y Colombia, puede ser clave para mejorar el acceso a medicamentos y tratamientos.

# 

# 2. Ampliar los programas de vacunación y distribución eficiente de medicamentos:

# Mejorar la logística y distribución de vacunas en países como México y Brasil, donde las disparidades en la cobertura de vacunación afectaron la respuesta a la pandemia.

# 3. Inversión en Salud Pública y Educación:

# Implementar programas educativos que aborden temas de salud pública, como la importancia de la vacunación y la prevención de enfermedades crónicas, puede reducir la vulnerabilidad de la población en futuras crisis sanitarias.

# 4. Fomentar alianzas público-privadas:

# Trabajar en conjunto con gobiernos locales para identificar las necesidades médicas y coordinar esfuerzos para la expansión de laboratorios en países con grandes brechas en su infraestructura sanitaria.

# Reflexión personal

# Durante este proyecto, he aprendido a gestionar y analizar grandes volúmenes de datos relacionados con la salud pública, desarrollando habilidades en el análisis exploratorio de datos, limpieza, y la visualización de información compleja a través de dashboards interactivos. Uno de los aprendizajes más importantes ha sido cómo los datos pueden influir directamente en la toma de decisiones estratégicas en el sector salud.

# Si tuviera que empezar el proyecto de nuevo, pondría mayor énfasis en recopilar datos adicionales relacionados con factores sociales y económicos, como los niveles de pobreza, acceso a la educación, y condiciones de vida en las áreas más afectadas por la pandemia. Además, incluiría análisis predictivos para anticipar cómo podrían evolucionar las necesidades de expansión de laboratorios farmacéuticos a medida que las condiciones sociales y de salud cambian en la región. También automatizaría parte del análisis para generar reportes actualizados y dinámicos.

# La experiencia me ha brindado una comprensión integral de cómo los datos pueden transformar las estrategias de negocio, y me ha permitido refinar mis habilidades como analista de datos.

# 