

Análisis estratégico para Biogenesys

Ruta de acceso al drive:

<https://drive.google.com/drive/folders/1GBPStQ9yZrbr1NCIZ6CtbEYOUWajuHQd?usp=sharing>

Nombre del autor: Agustina Luna

Email: agustinalunaux@gmail.com

Cohorte: DAFT-02

Fecha de entrega: 03/06/2024

Institución:



Introducción

BIOGENESYS necesita ubicar regiones y recolectar datos que sean de importancia y que les ayuden a tomar decisiones en su estrategia de expansión en Latinoamérica, específicamente en Colombia, Argentina, Chile, México, Perú y Brasil.

Objetivos del Proyecto

- Realizar un análisis exploratorio de datos sobre la incidencia de COVID-19 y otros factores relevantes.
- Aplicar técnicas de limpieza de datos para asegurar la calidad de los datos.
- Mejorar el acceso a los datos mediante operaciones eficientes de ETL.
- Desarrollar dashboards interactivos con visualizaciones eficientes.

1- Análisis Exploratorio de Datos:

- **Objetivo Alcanzado:** se realizó un análisis exploratorio de datos sobre la incidencia de COVID-19 y otros factores relevantes, identificando tendencias y oportunidades mediante estadísticas,

mediciones y visualizaciones. Este análisis permite a BIOGENESYS comprender mejor el entorno de los países de interés (Colombia, Argentina, Chile, México, Perú y Brasil) y las necesidades específicas de cada región.

2- Limpieza y Preparación de Datos:

- **Objetivo Alcanzado:** se aplicaron técnicas de limpieza de datos para asegurar la calidad de los datos. Esto facilitó el análisis y decisiones estratégicas confiables, eliminando datos inconsistentes y asegurando que los conjuntos de datos estuvieran listos para un análisis detallado.

3-Mejora en el Acceso a Datos mediante ETL:

- **Objetivo Alcanzado:** Se mejoro el acceso a los datos mediante operaciones eficientes de extracción, transformación y carga (ETL). Esto aumentó la eficacia del análisis y la toma de decisiones, permitiendo a BIOGENESYS manejar grandes volúmenes de datos de manera más efectiva.

4-Desarrollo de Dashboards Interactivos:

- **Objetivo Alcanzado:** Se desarrollo un dashboards interactivo con visualizaciones eficientes, permitiendo explorar datos desde múltiples perspectivas. Estos dashboards facilitaron la toma de decisiones informada y estratégica, proporcionando una herramienta visual poderosa para los directivos de BIOGENESYS.
-

Desarrollo del proyecto

Metodología de Recopilación y Selección de Datos

Para abordar el desafío planteado por BIOGENESYS, se siguieron los siguientes pasos

1. Recopilación de Datos:

- Se obtuvo un dataset inicial que contenía alrededor de 22 millones de registros y 707 columnas, abarcando una variedad de métricas relacionadas con la incidencia de COVID-19, tasas de vacunación y disponibilidad de infraestructura sanitaria.
- Los datos fueron recolectados de fuentes confiables, incluyendo bases de datos gubernamentales, organizaciones internacionales de salud y estudios de investigación.

2. Selección de Datos:

- Dada la magnitud del dataset original, se trabajó en conjunto con el equipo de ingeniería de datos para crear un subset específico enfocado en Latinoamérica, reduciendo el número de registros a 12,216,057 filas y 50 columnas.
- Los países de interés seleccionados fueron Colombia, Argentina, Chile, Mexico, Peru y Brazil, asegurando que los datos fueran relevantes para las necesidades estratégicas de BIOGENESYS.

3. Información del Archivo "readme.txt":

- Un archivo adicional, "readme.txt", fue proporcionado para detallar la selección de columnas y su significado, facilitando la comprensión y el manejo del dataset.

Transformaciones y Limpieza de Datos

La calidad y relevancia de los datos son cruciales para el análisis. Por ello, se realizaron diversas transformaciones y procesos de limpieza para asegurar que el dataset final fuera apto para el análisis.

Carga de Datos:

- Se utilizó Python y librerías como Pandas para cargar el dataset desde un archivo CSV.

Transformación de Datos:

- Se eliminaron las columnas irrelevantes para el análisis, basándonos en la información proporcionada en el archivo "readme.txt".
- Se transformaron los datos de fechas y se aseguraron formatos consistentes para su correcto manejo temporal.
- Se normalizaron valores y se manejaron los datos faltantes.

Limpieza de Datos:

- Se aplicaron técnicas de limpieza para manejar valores atípicos y erróneos.

Conclusión sobre las Transformaciones y Limpieza de Datos

Las transformaciones y limpieza de datos realizadas aseguraron que el dataset utilizado en los análisis posteriores fuera de alta calidad y adecuado para extraer insights relevantes. Este proceso fue esencial para:

- **Garantizar la Precisión:** La eliminación de datos irrelevantes y la corrección de inconsistencias aseguraron que los análisis se basaran en información precisa y confiable.
- **Facilitar el Análisis:** La normalización de datos y el manejo de valores faltantes permitieron realizar análisis estadísticos y visualizaciones de manera eficiente y sin errores.
- **Mejorar la Toma de Decisiones:** Al contar con un dataset limpio y bien estructurado, los directivos de BIOGENESYS pueden confiar en los resultados del análisis para tomar decisiones informadas sobre la expansión de laboratorios y centros de vacunación en Latinoamérica.

EDA e Insights

Durante el análisis exploratorio de datos (EDA) y la visualización de los datos, se han obtenido los siguientes insights clave:

1. Muertes por COVID-19:
 - En 2021, se reportaron aproximadamente 8.83 millones de muertes.
 - En 2022, las muertes disminuyeron significativamente a 1.39 millones.
 - El país con mayor número de muertes durante ambos años fue Brasil.
2. Recuperaciones por COVID-19:

- En 2021, el número de recuperados fue de 263.41 millones.
 - En 2022, se observó una disminución en el número de recuperados a 241.61 millones.
 - Estos datos indican una mejora en la tasa de mortalidad y un aumento en la eficacia de los tratamientos y las vacunas.
3. Casos Confirmados:
- Brasil fue el país con el mayor número de casos confirmados durante 2021 y 2022.
 - La mayor incidencia de nuevos casos se reporta entre los meses de enero y mayo, lo que sugiere un patrón estacional.
4. Vacunación:
- El mes con mayor cantidad de dosis de vacunas aplicadas fue mayo de 2022.
 - Colombia tiene el mayor porcentaje de vacunación entre los países analizados, mientras que Chile tiene el menor porcentaje de vacunación.
 - Este dato es crucial para identificar áreas donde se necesita fortalecer las campañas de vacunación.
5. Población Urbana y Rural:
- Brasil y México son los países con mayor población urbana.
 - Todos los países presentan una parte significativa de población rural, aunque no tan grande como la urbana.
 - México tiene la mayor densidad poblacional, lo cual puede influir en la propagación del virus y en la logística de distribución de vacunas.
6. Densidad de Población:
- México, con su alta densidad poblacional, requiere estrategias específicas para controlar la propagación del virus y distribuir vacunas de manera eficiente.
7. Tendencias Generales:
- Se observó que los meses con mayor cantidad de nuevos casos de COVID-19 fueron enero a mayo, lo cual debe considerarse en la planificación de recursos y personal médico durante estos meses.
 - El total de muertes por COVID-19 en todos los países analizados asciende a aproximadamente 10 millones.

Navegación del Dashboard

1. Portada

- **Descripción:** La portada del dashboard sirve como una interfaz principal que incluye botones de navegación hacia las otras páginas del dashboard. Además, presenta el título del proyecto y el logo de BIOGENESYS.
- **Elementos Clave:**
 - **Botones de Navegación:** Interactivos, permiten moverse fácilmente entre las diferentes secciones del dashboard.
 - **Título y Logo:** Proporcionan un contexto visual y profesional al informe.

2. Página 1: Análisis Temporal

- **Descripción:** Esta página está dedicada a la visualización de datos relacionados con la evolución temporal de la pandemia.

3. Página 2: Condiciones por País

- **Descripción:** Esta página permite explorar los datos específicos de cada país, proporcionando información detallada sobre la incidencia de COVID-19, tasas de vacunación,

distribución por edades, y otros indicadores de salud.

4. **Página 3: Índices Demográficos y de Salud**

- **Descripción:** Esta página se centra en la comparación de diversos índices demográficos y de salud entre los países.

Funcionalidades Interactivas

- **Botones de Filtros:** Permiten filtrar los datos por diferentes criterios, como país, rango de fechas, y otros indicadores relevantes.
- **Botones de Borrar Segmentación:** Facilitan la limpieza de los filtros aplicados, permitiendo volver a la vista general sin datos segmentados.
- **Interactividad:** Las visualizaciones interactivas permiten explorar diferentes escenarios y ajustar las estrategias en tiempo real, basándose en datos actualizados.

Posibles Conclusiones

1. **Identificación de Áreas Críticas:**

- Brasil, con el mayor número de casos y muertes, representa un área crítica que necesita atención prioritaria para la expansión de laboratorios y centros de vacunación.
- México, con su alta densidad poblacional requiere estrategias específicas para controlar la propagación del virus y distribuir vacunas de manera eficiente.

2. **Optimización de Recursos:**

- La identificación de patrones estacionales y picos en nuevos casos permite una mejor planificación y asignación de recursos durante los meses críticos de enero a mayo.
- Concentrar los esfuerzos de vacunación en los meses de mayor demanda puede aumentar la eficacia de las campañas de inmunización.

3. **Fortalecimiento de Campañas de Vacunación:**

- Colombia, con el mayor porcentaje de vacunación, puede servir como modelo para implementar estrategias exitosas en otros países con menor cobertura, como Chile.
- La información demográfica ayuda a focalizar las campañas de vacunación en áreas urbanas densamente pobladas y mejorar la cobertura en áreas rurales.

Conclusiones y Recomendaciones

Conclusión General

1. **Análisis Temporal:**

- **Patrones de Incidencia:** Los datos muestran que los picos de nuevos casos de COVID-19 se registraron entre enero y mayo, lo que sugiere la necesidad de preparación adicional durante estos meses.
- **Vacunación:** El mes de mayor administración de vacunas fue mayo de 2022, indicando una respuesta robusta en ese período.

2. **Condiciones por País:**

- **Tasa de Vacunación:** Colombia lidera en porcentaje de vacunación, mientras que Chile muestra la menor cobertura.

- **Distribución por Edades:** Los datos demográficos indican la necesidad de adaptar las estrategias de vacunación según la estructura de edad de la población.
- **Muertes y Salud:** Brasil y México registran el mayor número de muertes, correlacionándose con altos niveles de urbanización y densidad poblacional.

3. Índices Demográficos y de Salud:

- **Población Urbana y Rural:** La mayor parte de la población en Brasil y México es urbana, lo que justifica una mayor inversión en infraestructura sanitaria en ciudades densamente pobladas.
- **Mortalidad:** Las diferencias en la mortalidad infantil, masculina y femenina indican variaciones en factores de riesgo y acceso a la salud entre los países.

Conclusiones Estratégicas

1. Ubicaciones Óptimas para la Expansión de Laboratorios:

- **Brasil:** Con la mayor cantidad de casos y muertes, así como una alta densidad poblacional urbana, Brasil es una ubicación prioritaria para la expansión de laboratorios.
- **México:** La alta densidad y niveles de contaminación, junto con una significativa mortalidad por COVID-19, hacen de México otro candidato clave para la expansión.
- **Colombia:** Liderando en porcentaje de vacunación, Colombia puede servir como un centro estratégico para la distribución y administración de vacunas.

2. Políticas de Vacunación:

- **Fortalecimiento de Campañas:** Basado en el éxito de Colombia, implementar estrategias de vacunación similares en Chile y otros países con menor cobertura podría mejorar significativamente la inmunización.
- **Adaptación a Grupos de Edad:** Ajustar las campañas de vacunación según la estructura de edades, enfocando recursos en grupos de alto riesgo.

3. Inversión en Salud:

- **Infraestructura Sanitaria Urbana:** Invertir en la mejora de la infraestructura sanitaria en áreas urbanas densamente pobladas, especialmente en Brasil y México, es crucial.
- **Preparación para Picos Estacionales:** Incrementar la capacidad hospitalaria y de vacunación en anticipación a los picos de enero a mayo, mejorando la preparación y respuesta ante futuras oleadas.

4. Problemas Sociales y Económicos:

- **Acceso Equitativo a la Salud:** Desarrollar políticas que aseguren el acceso equitativo a la salud, especialmente en regiones rurales y entre poblaciones vulnerables.
- **Reducción de Contaminación:** Implementar políticas ambientales para reducir la contaminación, lo cual puede tener un impacto positivo en la reducción de la mortalidad por COVID-19 y otras enfermedades.

Reflexión personal

Este proyecto ha sido una experiencia profundamente enriquecedora y significativa en mi trayectoria como Analista de Datos. A través del proceso de análisis y visualización de datos utilizando Python y sus librerías especializadas, he adquirido habilidades técnicas fundamentales y he aplicado conocimientos teóricos en un entorno práctico y relevante.

Una de las principales lecciones aprendidas ha sido la importancia de Python como una herramienta poderosa y versátil en el análisis de datos. Desde la limpieza y transformación de datos hasta la

generación de visualizaciones informativas, Python y sus librerías como Pandas, Numpy, Matplotlib y Seaborn han sido fundamentales en cada etapa del proyecto.

Además de fortalecer mis habilidades técnicas, este proyecto me ha proporcionado una comprensión más profunda del proceso completo de análisis de datos, desde la recopilación y preparación de datos hasta la interpretación y comunicación de resultados. He aprendido a enfrentar desafíos reales en la manipulación de grandes volúmenes de datos y a extraer insights significativos que respalden la toma de decisiones estratégicas.

EXTRA CREDIT

Como parte del desafío adicional, decidí agregar una capa de visualización geoespacial a mi proyecto utilizando la librería Folium en Python. Aprovechando la información de latitud y longitud disponible en mis datos, creé mapas interactivos que proporcionan una perspectiva geográfica de la distribución de los casos de COVID-19, la cobertura de vacunación y otras variables relevantes.

La inclusión de estos mapas ha enriquecido significativamente mis descubrimientos al permitirme explorar la distribución geográfica de los datos. Además, esta visualización geoespacial proporciona una comprensión más completa de la situación y permite identificar patrones y tendencias que pueden no ser evidentes en las visualizaciones tradicionales.