

Expansión Estratégica de Biogenesys con Python

Nombre del autor: Efrain Chapal

Email: chapalefrain@gmail.com

Cohorte: 13

Fecha de entrega: 14 de mayo 2025

Institución:



Introducción

Este proyecto tuvo como propósito principal realizar un análisis de datos exhaustivo sobre el impacto de la pandemia de COVID-19 y las campañas de vacunación en seis países clave de América Latina (Argentina, Brasil, Chile, Colombia, México y Perú). El objetivo organizacional para BIOGENESYS fue identificar, a partir de estos datos, las ubicaciones óptimas para una potencial expansión de sus laboratorios farmacéuticos, con el fin de mejorar el acceso a vacunas y tratamientos en la región en un contexto postpandemia.

Desarrollo del proyecto

La metodología se centró en un conjunto de datos proporcionado en formato CSV, que contenía información diaria sobre casos de COVID-19, muertes, vacunación, y variables demográficas y socioeconómicas para múltiples países. Se realizó una selección inicial de datos para enfocar el análisis en Argentina, Brasil, Chile, Colombia, México y Perú, para fechas posteriores al 1 de enero de 2021.

Las transformaciones y limpieza de datos incluyeron:

- se aplicó una propagación hacia atrás por país
- Conversión de tipos de datos, especialmente asegurando que las fechas estuvieran en formato `datetime`.
- Creación de columnas calculadas relevantes en Python para el Análisis Exploratorio de Datos (EDA), como medias móviles de 7 días para casos y muertes, tasas de crecimiento, cobertura de vacunación por 100 personas, y preparación de datos para análisis específicos como la distribución de edad.
- La conclusión principal de esta etapa es que se obtuvo un dataset enriquecido, apto para el análisis visual y la posterior carga en Power BI, aunque se identificaron algunas anomalías en los datos reportados por los países (como picos de muertes inusuales o saltos en datos de vacunación) que requieren una interpretación cautelosa.

Módulo 4

EDA e insights

El Análisis Exploratorio de Datos (EDA) realizado en Python, a través de múltiples visualizaciones, reveló insights clave:

- **Impacto Heterogéneo de la Pandemia:** Las olas de COVID-19 (casos y muertes) variaron significativamente en magnitud y temporalidad entre los países analizados. Brasil, Perú y México mostraron picos de mortalidad diaria muy altos en ciertas etapas.
- **Desacoplamiento Casos-Mortalidad:** Se observó una tendencia general hacia un desacoplamiento entre el número de casos y el número de muertes en olas posteriores (especialmente con Ómicron en 2022), probablemente debido al avance de la vacunación y cambios en la virulencia de las variantes.
- **Desempeño de Vacunación Diverso:** Chile destacó por una campaña de vacunación rápida y de alta cobertura. Perú mostró una notable capacidad de aceleración. Otros países tuvieron ritmos más graduales. Se identificaron particularidades como saltos en los datos de vacunación de Argentina.
- **Factores Demográficos y de Riesgo:**
 - La distribución de edad varía, con países como Perú y Colombia mostrando poblaciones más jóvenes, y Chile y Argentina con mayor proporción de adultos mayores.
 - Se encontró una asociación a nivel agregado entre una mayor prevalencia de tabaquismo y el grupo de países con tasas de mortalidad por COVID-19 más altas. La relación con la diabetes fue menos directa en el análisis agregado.
 - La urbanización es alta en la mayoría de los países, pero su correlación con la propagación máxima de casos presentó excepciones como México.
- **Anomalías en los Datos:** Se detectaron irregularidades en los datos, como el pico extremo de muertes en Chile en marzo de 2022 y la caída abrupta de reportes al final del periodo en varios países, lo cual fue crucial para la interpretación de la "situación actual" del dataset.

Análisis del dashboard

El dashboard interactivo desarrollado en Power BI se estructuró en tres pestañas principales para facilitar la navegación y el storytelling:

1. Pestaña 1: Panorama COVID-19 LATAM:

- **Navegación:** El usuario puede utilizar los filtros de país y fecha para explorar. Las tarjetas KPI ofrecen un resumen del impacto (casos, muertes, vacunación). Un mapa muestra geográficamente la mortalidad máxima, y gráficos de barras comparan las tasas de impacto entre países. Un gráfico de líneas permite ver la evolución de casos vs. muertes para el país seleccionado.
- **Conclusiones:** Permite identificar rápidamente los países más afectados en términos absolutos y relativos, y entender la dinámica de las olas pandémicas individuales. Por ejemplo, se puede ver el "desacoplamiento" en el gráfico de líneas.

2. Pestaña 2: Análisis de Cobertura de Vacunación:

- **Navegación:** Similar a la anterior, con filtros de país y fecha. Los KPIs se centran en la vacunación. La visualización principal es un gráfico de líneas que compara el progreso de la cobertura de vacunación (dosis/100p) de todos los países.

Módulo 4

- **Posibles:** Permite identificar líderes en vacunación (Chile, Perú en su aceleración), la velocidad de las campañas y los niveles de cobertura alcanzados, infiriendo capacidades logísticas y de respuesta.
- 3. **Pestaña 3: Factores para Expansión y Conclusiones:**
 - **Navegación:** Presenta visualizaciones comparativas de la distribución de edad (barras agrupadas), y una tabla con factores de riesgo, socioeconómicos y de infraestructura sanitaria. Incluye un cuadro de texto con las conclusiones y recomendaciones finales.
 - **Conclusiones Posibles:** Esta pestaña sintetiza la información. Se pueden cruzar los perfiles demográficos (ej. alta proporción de adultos mayores en Chile) con la capacidad de respuesta (alta vacunación en Chile) y otros factores (IDH, personal sanitario) para perfilar cada país de cara a la estrategia de expansión de BIOGENESYS. Las recomendaciones se basan en esta vista integrada.

Conclusiones y Recomendaciones

La síntesis del análisis, combinando el EDA en Python y la exploración del dashboard interactivo en Power BI, nos lleva a una comprensión multifacética del impacto de la pandemia y las capacidades de respuesta en los seis países latinoamericanos estudiados. Se confirma la heterogeneidad de la región y la importancia de la vacunación en mitigar la severidad de la COVID-19.

Conclusiones Estratégicas y Respuesta Principal del Proyecto: Considerando políticas de vacunación observadas (velocidad, cobertura), impacto de la pandemia (mortalidad, casos), perfiles demográficos (distribución de edad), factores de riesgo (tabaquismo, diabetes), indicadores socioeconómicos (IDH, PIB per cápita) y de infraestructura sanitaria (personal médico), las **ubicaciones óptimas para una potencial expansión de laboratorios farmacéuticos de BIOGENESYS** se perfilan de la siguiente manera:

1. **Prioridad Alta - Mercados con Alta Capacidad Demostrada y Necesidades Demográficas Específicas:**
 - **Chile:** Destaca por su excepcional ejecución de la campaña de vacunación y una población más envejecida, sugiriendo una demanda sostenida de productos para adultos y refuerzos. Su infraestructura sanitaria parece ser de las más robustas.
 - **Perú:** Mostró una notable capacidad de aceleración en su vacunación, alcanzando alta cobertura. Su sistema de salud, aunque presionado, y su población (más joven pero con buena disposición a la vacunación) lo hacen un candidato interesante.
2. **Consideración Estratégica - Mercados de Gran Escala con Potencial Continuo:**
 - **Brasil y Argentina:** Por su considerable tamaño poblacional, representan mercados con necesidades diversas y continuas. La evaluación de la infraestructura existente para posibles colaboraciones será clave. Argentina, similar a Chile en perfil demográfico, también presenta un mercado para adultos.
3. **Oportunidad con Análisis Adicional - Mercados con Dinámicas Particulares:**
 - **Colombia:** Con una población joven y una capacidad de vacunación que fue mejorando, representa un mercado en crecimiento potencial.

Módulo 4

- **México:** A pesar de su alta urbanización, la dinámica de casos reportados versus su CFR inicial sugiere la necesidad de un análisis más profundo (posible subregistro). Su gran población es atractiva, pero requiere una comprensión más detallada de las necesidades y capacidades reales de su sistema de salud.

Recomendaciones Adicionales:

- Realizar estudios de mercado específicos en los países priorizados.
- Evaluar el entorno regulatorio y la competencia local.
- Investigar la infraestructura de cadena de frío y distribución.
- Monitorear continuamente la evolución epidemiológica y las políticas de salud pública.

Reflexión personal

Este proyecto ha sido una experiencia de aprendizaje intensiva y muy valiosa. Fortalecí mis habilidades en el ciclo completo del análisis de datos, desde la limpieza y transformación con **Python (Pandas)**, pasando por la visualización exploratoria con **Seaborn/Matplotlib**, hasta la creación de un dashboard interactivo en **Power BI**. Cada etapa presentó desafíos que se convirtieron en lecciones; por ejemplo, el manejo de datos anómalos o la correcta configuración de los visuales de Python en Power BI me enseñaron la importancia de la verificación constante y el análisis crítico.

Si tuviera que empezar de nuevo, dedicaría aún más tiempo a la validación inicial de los datos fuente para anticipar anomalías y planificarla con mayor detalle qué cálculos realizar en Python para optimizar el flujo hacia Power BI. No obstante, los errores fueron cruciales para mi crecimiento, reforzando mi proactividad y mi capacidad para aprender de cada obstáculo, transformando datos en conclusiones accionables.

EXTRA CREDIT

Como parte del desafío extra credit, se desarrolló un mapa interactivo utilizando la librería Folium en Python. Este mapa visualiza los países de estudio en América Latina, representando cada uno con un marcador circular cuyo tamaño es proporcional al máximo de casos confirmados acumulados por 100,000 habitantes. Al hacer clic en cada marcador, se despliega una ventana emergente (popup) con el nombre del país, su tasa máxima de casos por 100k y el total de casos acumulados correspondientes a ese pico. Esta visualización geoespacial, generada como un archivo HTML independiente, permitió una perspectiva adicional sobre la magnitud del impacto de la pandemia en la región y complementó los análisis realizados en el dashboard de Power BI, demostrando la capacidad de utilizar herramientas de Python para visualizaciones geoespaciales interactivas.