**Expansión Estratégica de Biogenesys con Python**

**Nombre del autor: Gonzalo Nicolás Arratia Vargas**

**Email: gonzaloarratia76@gmail.com**

**Cohorte: DAFT-12**

**Fecha de entrega:** 07-04-2025

**Institución: Farmacia Biogenesys**



# Introducción

# Este proyecto fue desarrollado para BIOGENESYS, empresa farmacéutica líder que busca optimizar su expansión en Latinoamérica mediante análisis de datos de COVID-19. El objetivo principal fue identificar ubicaciones estratégicas para nuevos laboratorios y centros de vacunación, analizando 12+ millones de registros de 6 países (Argentina, Brasil, Chile, Colombia, México y Perú) desde 2021.

# Los objetivos específicos del proyecto incluyen:

# Realizar un análisis exploratorio de datos (EDA) para identificar tendencias y oportunidades.

# Aplicar técnicas de limpieza y transformación de datos para garantizar la calidad del análisis.

# Desarrollar visualizaciones interactivas en Power BI que faciliten la toma de decisiones estratégicas.

# Identificar áreas prioritarias para la expansión de laboratorios y centros de vacunación.

# Desarrollo del proyecto

#### Metodología de recopilación y selección de datos

# Para este proyecto se utilizó una base de datos pública sobre COVID-19, la cual contiene información detallada sobre nuevos casos confirmados, fallecimientos, dosis administradas de vacunas, entre otros indicadores relevantes. La fuente original fue seleccionada por su confiabilidad, frecuencia de actualización y cobertura geográfica (especialmente países de América Latina).

# Durante la recopilación, se filtraron los registros exclusivamente relacionados con países latinoamericanos, ya que el enfoque del análisis está orientado a la región. Se priorizó el período comprendido entre enero y diciembre del año 2021 para asegurar una ventana temporal homogénea que permita comparaciones mensuales y semanales estandarizadas.

#### Transformaciones y limpieza de datos

# Se realizaron diversas transformaciones para preparar el dataset y facilitar un análisis temporal y comparativo eficaz. Entre ellas:

# Conversión del campo date a formato de fecha (datetime) para facilitar operaciones con índices temporales.

# Establecimiento del índice temporal en el DataFrame para agrupar y analizar por semana (.isocalendar().week) y por mes (.dt.month).

# Agrupación de datos por país y periodo temporal (semanal/mensual), utilizando groupby() y Grouper() para calcular sumas y promedios.

# Reemplazo de valores nulos o inconsistentes en columnas críticas para evitar errores de cálculo.

# Cálculo de métricas derivadas como porcentaje de crecimiento mensual y acumulado de vacunación, necesarias para visualizar tendencias.

# Creación de nuevas columnas (week, month) y normalización de unidades para gráficos comparativos.

# Estas transformaciones permitieron no solo limpiar y homogeneizar los datos, sino también derivar variables útiles para detectar patrones temporales, identificar correlaciones y evaluar el impacto de la vacunación o las políticas sanitarias en la evolución de los casos.

# 

# EDA e insights

# Durante el análisis exploratorio de datos (EDA), descubrimos varios puntos clave que nos ayudaron a entender mejor cómo fue avanzando la pandemia en los seis países seleccionados: Argentina, Brasil, Chile, Colombia, México y Perú. Acá te contamos los hallazgos más importantes:

# No hay una relación directa entre el clima y los casos: Aunque al principio se pensaba que el virus afectaba más en climas fríos, vimos que países como Brasil, con temperaturas altas, también tuvieron muchos contagios. O sea, el clima solo no explica los brotes.

# La vacunación ayudó, pero no fue suficiente por sí sola: En general, los casos empezaron a bajar cuando aumentaron las vacunas, pero esto no fue igual en todos lados. En algunos países, los casos seguían altos incluso con campañas de vacunación fuertes, lo que muestra que también influyen otros factores como el comportamiento social o la aparición de nuevas variantes.

# Brasil tuvo el mayor impacto en todos los indicadores: Fue el país con más casos, más muertes y también el que aplicó más vacunas. Esto puede estar relacionado con su gran población, pero también con decisiones políticas y la manera en que se gestionó la crisis.

# Los países con menor población presentaron menos casos y muertes: No es una gran sorpresa, pero se confirmó que hay una relación directa entre el tamaño de la población y el número de contagios y fallecimientos, aunque siempre hay excepciones por cómo se manejó la pandemia en cada país.

# Hay una alta correlación entre casos y muertes: A mayor cantidad de contagios, también hubo más fallecimientos, lo que era esperable. Pero este dato sirve para confirmar que, en la mayoría de los casos, las muertes aumentaban casi en paralelo con los contagios.

# Análisis del dashboard

#### Navegación Estratégica

# El dashboard está organizado en 3 vistas principales e interactivas, pensadas para que el usuario pueda navegar de forma simple y al mismo tiempo sacar conclusiones rápidas y útiles sobre la evolución del COVID-19 en los seis países analizados. Cada página tiene su propio enfoque:

# Visión General Acá se puede ver un resumen rápido de los casos, muertes y vacunación por país. Sirve como un pantallazo general para comparar el impacto del virus entre naciones y entender el panorama de un vistazo.

# Análisis Detallado por País En esta sección se puede elegir un país y ver sus datos específicos a lo largo del tiempo. Hay gráficos que muestran cómo fueron creciendo los contagios, cuántas vacunas se aplicaron y cómo evolucionaron las muertes. Ideal para profundizar en un solo caso.

# Factores Relacionados Esta vista conecta las variables del COVID-19 con otras como temperatura, densidad poblacional, esperanza de vida, etc. Ayuda a entender si hay alguna relación o patrón entre el contexto de cada país y cómo se vivió la pandemia.

# Todas las páginas están pensadas para que el usuario explore de manera intuitiva, pueda filtrar por país o por año, y descubra insights sin perderse entre tantos datos. El diseño busca claridad visual, interacción fluida y toma de decisiones basadas en evidencia.

# Conclusiones y Recomendaciones

#### Conclusiones principales

# Después de explorar el dashboard, se pueden sacar varias ideas clave:

# La temperatura no es determinante: Aunque se analizaron los casos y muertes en relación con la temperatura promedio, no se encontró una relación directa. Países con climas muy distintos (como Brasil y Chile) tuvieron cifras altas, lo que muestra que hay otros factores más importantes en juego.

# Brasil fue el caso más crítico: En casi todos los gráficos, Brasil se destaca por tener las cifras más altas de casos, muertes y vacunación. Esto se relaciona con su gran población, pero también indica que el impacto fue mucho mayor.

# La vacunación ayudó, pero no fue suficiente sola: A medida que aumentaron las dosis aplicadas, en general los casos comenzaron a bajar. Sin embargo, se notan picos incluso con altos niveles de vacunación, lo que indica que también influyeron cosas como nuevas variantes y el relajamiento de las medidas sanitarias.

# Los factores sociales y de salud son clave: La matriz de correlación muestra que variables como la densidad poblacional, la mortalidad adulta, el desarrollo humano y las condiciones preexistentes (como diabetes) tienen un fuerte impacto en la dinámica de la pandemia.

# En resumen, este dashboard nos muestra que la pandemia fue (y es) un fenómeno muy complejo. No se puede entender mirando solo un dato o una sola causa. Para poder hacer frente a una situación como esta, hay que mirar muchos aspectos a la vez: la salud, la economía, cómo vive la gente, qué recursos hay, qué decisiones se toman, etc.

# Así que, más allá de los gráficos y los números, este trabajo nos deja una idea muy clara: cada país vivió su propia realidad durante la pandemia, y para mejorar la respuesta ante futuras crisis, es fundamental tener una mirada amplia, que tenga en cuenta no solo la medicina, sino también lo social, lo económico, lo ambiental y lo humano.

#### Recomendaciones

# Basado en los hallazgos, se pueden proponer algunas ideas para futuras estrategias:

# Políticas adaptadas a cada contexto: Dado que cada país vivió la pandemia de forma distinta, es clave diseñar respuestas que se ajusten a las realidades sociales, demográficas y sanitarias de cada lugar.

# No confiar solo en la vacunación: Si bien es una herramienta fundamental, debe ir acompañada de medidas como el control de movilidad, comunicación efectiva y preparación del sistema de salud para evitar picos críticos.

# Monitorear más allá del clima: Es importante no centrarse solo en el factor climático. La combinación de factores como densidad, calidad del sistema sanitario y comportamiento social es la que realmente explica la evolución de los contagios y muertes.

# Mejorar la recolección de datos: Una base sólida de datos bien organizados y detallados permite un análisis más profundo y decisiones más acertadas en el futuro.

# Reflexión personal

# En general, estoy satisfecho con el proceso y los resultados obtenidos, y creo que este proyecto ha sentado las bases, pero si tendría que afinar unos detalles más en Estadísticas. Tuve muchos problemas para instalar las extensiones pero lo pude solucionar creo que fue donde me tomó mucho más tiempo.

# Este proyecto ha sido una experiencia muy desafiante que me permitió desarrollar habilidades claves como por ejemplo: Limpieza y transformación de datos, Análisis exploratorio de datos (EDA), Visualización de datos con Python y Power BI..

# En una próxima vez utilizaría técnicas de machine learning para predecir áreas de riesgo futuro. Hubiera creado una nueva columna para los recuperados ya que no era muy cómodo usar la columna existente ya que había muchos valores nulos así que no era muy confiable. Mi idea era que luego de 2 semanas los casos confirmados pasaban a ser recuperados restando obviamente los decesos