**Expansión Estratégica de Biogenesys**

**Nombre del autor: Carlos Fernando Dussan Rivera**

**Email: cafedusriv@gmail.com**

**Cohorte: DAPT-01**

**Fecha de entrega: 29 de julio de 2024**

**Institución:**



# Introducción

# En el contexto actual, donde la pandemia de COVID-19 ha planteado desafíos sin precedentes para el sector de la salud, la expansión estratégica de laboratorios farmacéuticos se ha convertido en una prioridad fundamental. Biogenesys, una empresa farmacéutica comprometida con la innovación y la excelencia, ha identificado la necesidad de optimizar su presencia en Latinoamérica para mejorar el acceso a vacunas y fortalecer la infraestructura sanitaria en la región.

# El proyecto "Expansión Estratégica de Biogenesys" tiene como objetivo principal identificar las ubicaciones óptimas para la expansión de laboratorios farmacéuticos en países clave de Latinoamérica, incluyendo Colombia, Argentina, Chile, México, Perú y Brasil. Este análisis se basa en una evaluación exhaustiva de la incidencia de COVID-19, las tasas de vacunación y la disponibilidad de infraestructuras sanitarias. La meta es desarrollar una estrategia de expansión que permita a Biogenesys responder de manera efectiva a las necesidades de salud pública y estar preparados ante futuras contingencias sanitarias.

# Este proyecto no solo representa un avance significativo para Biogenesys en términos de expansión y fortalecimiento de su posición en el mercado, sino que también ha sido una oportunidad invaluable para aplicar y consolidar habilidades adquiridas en análisis de datos, visualización y toma de decisiones estratégicas. A través de este proyecto, hemos aprendido a manejar grandes volúmenes de datos, aplicar técnicas avanzadas de limpieza y transformación de datos, y utilizar herramientas de visualización como Power BI para presentar insights de manera clara y accionable.

# El proceso ha involucrado una combinación de métodos cualitativos y cuantitativos, permitiendo una comprensión profunda de las tendencias y patrones en los datos relacionados con la pandemia. Este enfoque ha facilitado la identificación de áreas prioritarias para la expansión, basándose en criterios sólidos y datos verificables. Además, el proyecto ha puesto de manifiesto la importancia de una toma de decisiones informada y basada en datos, un principio esencial en el ámbito del análisis de datos y la inteligencia empresarial.

# En resumen, este informe documenta el proceso completo del proyecto, desde la recopilación y limpieza de datos hasta la visualización y análisis, proporcionando una guía detallada y fundamentada para la expansión estratégica de Biogenesys en Latinoamérica. A través de este trabajo, no solo hemos aportado valor significativo a la empresa, sino que también hemos fortalecido nuestras competencias como analistas de datos, preparados para enfrentar desafíos complejos y contribuir al éxito de proyectos críticos en el futuro.

# 

# Desarrollo del proyecto

**El desarrollo del proyecto "Expansión Estratégica de Biogenesys" se llevó a cabo en cuatro avances principales. A continuación, se describe el proceso seguido en cada uno de estos avances, explicando de manera concreta cada paso, las decisiones tomadas y las acciones realizadas.**

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

**Paso 1: Leer Readme.txt**

* **Descripción:** Se revisó el archivo README para entender la estructura y el contenido del conjunto de datos.
* **Objetivo:** Asegurarse de que se comprendieran los datos disponibles y su formato.
* **Acciones:** Se identificaron las columnas y su significado, lo cual fue esencial para planificar la limpieza y análisis de datos posteriores.

**Paso 2: Crear el Notebook y Preparar el Entorno**

* **Descripción:** Se creó el notebook "PIDA\_M4\_Carlos\_Fernando\_Dussan.ipynb" y se preparó el entorno de trabajo.
* **Objetivo:** Configurar el entorno de trabajo para análisis de datos en Python.
* **Acciones:**
  + Instalación de librerías necesarias como Pandas, Numpy, Matplotlib y Seaborn.
  + Carga de librerías en el notebook.
  + Verificación de las versiones de las librerías para asegurar compatibilidad.
  + Configuración del entorno para evitar advertencias futuras que pudieran afectar la claridad del análisis.

**Paso 3: Leer el Archivo data\_latinoamerica.csv**

* **Descripción:** Carga del archivo CSV con los datos de Latinoamérica.
* **Objetivo:** Importar los datos al entorno de trabajo para su análisis.
* **Acciones:**
  + Lectura del archivo CSV utilizando Pandas, cargando los datos en un DataFrame.
  + Visualización rápida de las primeras filas del DataFrame para asegurar que los datos se habían cargado correctamente y confirmar la estructura esperada.

**Paso 4: Verificar la Cantidad de Registros y Columnas**

* **Descripción:** Verificación del tamaño del conjunto de datos.
* **Objetivo:** Entender la dimensión de los datos disponibles.
* **Acciones:**
  + Conteo de registros y columnas del DataFrame para confirmar la cantidad de datos cargados y asegurarse de que se correspondía con la información del README.

**Paso 5: Seleccionar Países de Interés**

* **Descripción:** Filtrado de datos para incluir solo los países relevantes.
* **Objetivo:** Enfocar el análisis en los países de interés.
* **Acciones:**
  + Utilización de la función isin() de Pandas para filtrar los datos y mantener solo los registros de Colombia, Argentina, Chile, México, Perú y Brasil.
  + Verificación de los datos filtrados revisando el número de registros y los nombres de los países en el DataFrame resultante.

**Paso 6: Filtrar por Fechas Mayores a 2021-01-01**

* **Descripción:** Filtrado de datos para incluir solo registros posteriores al 1 de enero de 2021.
* **Objetivo:** Analizar datos recientes y relevantes.
* **Acciones:**
  + Aplicación de un filtro para seleccionar solo los registros con fechas posteriores al 1 de enero de 2021. Esta decisión se tomó para enfocarnos en datos recientes y relevantes para la situación actual de la pandemia y la vacunación.

**Paso 7: Completar Valores Faltantes a Nivel de País**

* **Descripción:** Identificación y llenado de valores faltantes en el conjunto de datos.
* **Objetivo:** Asegurar la integridad de los datos.
* **Acciones:**
  + Realización de un conteo de valores nulos por columna para identificar dónde faltaban datos.
  + Relleno de los valores faltantes usando la media por país en las columnas relevantes (por ejemplo, new\_confirmed, new\_deceased). Esta decisión se tomó para mantener la consistencia de los datos sin introducir sesgos significativos.

**Paso 8: Limpieza Preliminar de Datos**

* **Descripción:** Limpieza adicional de datos basada en el análisis preliminar.
* **Objetivo:** Preparar un conjunto de datos limpio y coherente para el análisis.
* **Acciones:**
  + Uso de la media por país para completar valores faltantes.
  + Relleno de valores nulos en ciertas columnas con 0 donde tenía sentido (por ejemplo, new\_recovered, cumulative\_recovered).
  + Verificación final de valores nulos y dimensiones del DataFrame para asegurar que los datos estuvieran listos para un análisis más detallado.

**Paso 9: Examinar Características Básicas del Dataset**

* **Descripción:** Exploración inicial de las características del conjunto de datos.
* **Objetivo:** Obtener una comprensión general de los datos.
* **Acciones:**
  + Uso de funciones de Pandas para describir las columnas, calcular medidas estadísticas básicas (media, mediana, desviación estándar) y obtener información general del DataFrame. Esto proporcionó una visión clara de la distribución y variabilidad de los datos.

**Paso 10: Guardar Datos Filtrados**

* **Descripción:** Guardado del conjunto de datos limpio y filtrado.
* **Objetivo:** Tener una copia de los datos listos para el análisis.
* **Acciones:**
  + Exportación de los datos filtrados y limpios a un nuevo archivo CSV (DatosFinalesFiltrado.csv). Esta acción permitió almacenar una versión intermedia de los datos que podía ser utilizada en análisis futuros sin necesidad de repetir el proceso de limpieza.

**Paso 11: Calcular Estadísticas Descriptivas**

* **Descripción:** Cálculo de estadísticas descriptivas para las variables numéricas.
* **Objetivo:** Profundizar en la comprensión de los datos.
* **Acciones:**
  + Cálculo de medidas de tendencia central (media, mediana) y dispersión (varianza, rango) para cada columna numérica. Esto se realizó recorriendo cada columna y describiendo sus características estadísticas, lo que ayudó a identificar patrones y outliers en los datos.

**Paso 12: Crear Función para Mediana, Varianza y Rango**

* **Descripción:** Desarrollo de una función para calcular estadísticas clave.
* **Objetivo:** Facilitar el cálculo de estadísticas descriptivas.
* **Acciones:**
  + Implementación de una función que calcula la mediana, varianza y rango de cada columna numérica. Esta función se aplicó al conjunto de datos para obtener una visión detallada de la distribución de los valores y su dispersión. La decisión de crear esta función fue para estandarizar y automatizar el proceso de obtención de estas estadísticas en futuros análisis.

**Extra Credit**

* **Descripción:** Exploración del uso de funciones de orden superior para una manipulación eficiente de los datos.
* **Objetivo:** Aplicar técnicas avanzadas de programación para optimizar el procesamiento de datos.
* **Acciones:**
  + Se utilizaron funciones de orden superior como map(), filter() y reduce() para manipular los datos de manera más declarativa y concisa. Por ejemplo, se aplicó map() para transformar datos, filter() para seleccionar subconjuntos específicos y reduce() para agregar valores. Esta práctica mejoró la eficiencia y la legibilidad del código.

**Avance 2: Análisis Exploratorio - Visualización**

**Paso 1: Abrir el archivo “PIDA\_M4\_Nombre\_Apellido.ipynb” y añadir el título “Avance 2”**

* **Descripción:** Continuar el análisis en el mismo notebook utilizado en el avance 1.
* **Objetivo:** Mantener un registro continuo y organizado de todo el análisis.
* **Acciones:** Se abrió el notebook existente y se añadió una nueva sección titulada "Avance 2" para separar claramente los diferentes pasos y mantener la estructura del proyecto.

**Paso 2: Importar las librerías necesarias**

* **Descripción:** Preparación del entorno para la visualización de datos.
* **Objetivo:** Configurar el entorno para generar gráficos y visualizaciones efectivas.
* **Acciones:**
  + Se importaron librerías de visualización como Matplotlib y Seaborn, además de Pandas y Numpy, que ya estaban importadas en el avance 1.
  + Se configuraron estilos de visualización personalizados utilizando Seaborn para asegurar que los gráficos fueran claros y estéticamente agradables.

**Paso 3: Análisis Estadístico con Pandas y Numpy**

* **Descripción:** Exploración de propiedades estadísticas del dataset.
* **Objetivo:** Entender mejor la situación actual y las necesidades de las áreas en estudio.
* **Acciones:**
  + Se calcularon medidas de tendencia central (media, mediana) y dispersión (desviación estándar, varianza) utilizando Pandas y Numpy.
  + Se analizaron las correlaciones entre variables para identificar relaciones significativas que pudieran influir en la toma de decisiones. Esto incluyó el cálculo de matrices de correlación y la identificación de correlaciones fuertes (por encima de 0.5).
  + Se generaron tablas descriptivas y matrices de correlación para visualizar estas relaciones y entender mejor la estructura de los datos.

**Paso 4: Visualización de Datos con Matplotlib y Seaborn**

* **Descripción:** Representar los hallazgos por cada país o de manera general a través de gráficos y visualizaciones.
* **Objetivo:** Descubrir patrones, tendencias y anomalías en los datos.
* **Acciones:**
  + **Histogramas y Gráficos de Densidad:** Se generaron para entender la distribución de la incidencia de COVID-19 y las tasas de vacunación.
  + **Gráficos de Barras:** Se crearon para comparar diferentes regiones y visualizar diferencias en la incidencia y vacunación.
  + **Mapas de Calor:** Se usaron para identificar correlaciones entre diferentes variables y visualizar relaciones significativas.
  + **Gráficos de Dispersión:** Se utilizaron para explorar posibles relaciones entre variables como la temperatura media y los casos confirmados, y la temperatura media y las muertes confirmadas.
  + **Boxplots y Violinplots:** Se utilizaron para comparar la distribución de variables clave entre diferentes países.
  + **Evolución Temporal:** Se generaron gráficos que mostraban la evolución de las dosis administradas, muertes, casos confirmados y recuperaciones por mes de cada país.
  + **Comparaciones entre Países:** Se crearon gráficos para comparar el número de casos nuevos y dosis acumuladas entre países.
  + **Distribución de Población:** Se visualizaron la distribución de la población por grupos de edad y se compararon métricas demográficas y de salud.

**Paso 5: Identificación de Tendencias y Patrones**

* **Descripción:** Análisis detallado de los datos relacionados con la incidencia de COVID-19 y las tasas de vacunación para identificar tendencias a largo plazo y patrones estacionales o geográficos.
* **Objetivo:** Prever áreas de alta prioridad para la asignación de recursos y la implementación de medidas preventivas y de control.
* **Acciones:**
  + Se investigaron posibles influencias de factores como la temperatura en la propagación de la enfermedad.
  + Se analizaron variaciones en los índices de vacunación en diferentes regiones y en distintos momentos del año.
  + Se agruparon datos por semana y mes para visualizar tendencias temporales y patrones estacionales.
  + Se analizaron series temporales para detectar cambios en la incidencia y vacunación a lo largo del tiempo.

**Paso 6: Personalización de Visualizaciones**

* **Descripción:** Mejora de la calidad y claridad de las visualizaciones.
* **Objetivo:** Hacer las visualizaciones más informativas y atractivas.
* **Acciones:**
  + Se ajustaron paletas de colores, etiquetas, leyendas y títulos en las visualizaciones.
  + Se añadieron etiquetas y leyendas para mejorar la interpretación de los gráficos.
  + Se ajustaron los tamaños de los gráficos y la resolución (DPI) para asegurar que las visualizaciones se mantuvieran claras y legibles al exportarlas.
  + Se personalizaron los estilos de gráficos utilizando las capacidades de Seaborn para resaltar diferentes aspectos de los datos y hacer los gráficos más atractivos.

**Avance 3: Análisis Exploratorio Detallado con Numpy y Pandas**

**Paso 1: Abrir el archivo “PIDA\_M4\_Nombre\_Apellido.ipynb” y añadir el título “Avance 3”**

* **Descripción:** Continuar el análisis en el mismo notebook utilizado en los avances anteriores.
* **Objetivo:** Mantener un registro continuo y organizado de todo el análisis.
* **Acciones:** Se abrió el notebook existente y se añadió una nueva sección titulada "Avance 3" para separar claramente los diferentes pasos y mantener la estructura del proyecto.

**Paso 2: Importar las librerías necesarias**

* **Descripción:** Preparación del entorno para el análisis detallado de datos.
* **Objetivo:** Configurar el entorno para realizar un análisis exploratorio avanzado.
* **Acciones:**
  + Se importaron librerías necesarias como Pandas, Numpy, Matplotlib y Seaborn.
  + Se configuraron estilos de visualización personalizados utilizando Seaborn para asegurar que los gráficos fueran claros y estéticamente agradables.

**Paso 3: Realizar un análisis exploratorio detallado utilizando técnicas avanzadas de Pandas y Numpy**

* **Descripción:** Análisis detallado de las series temporales para comprender la evolución de elementos específicos del conjunto de datos.
* **Objetivo:** Identificar tendencias a largo plazo, estacionalidad y patrones temporales relevantes.
* **Acciones:**
  + Se agruparon los datos por semana y mes para analizar la evolución temporal de casos y muertes.
  + Se utilizaron técnicas de descomposición de series temporales para identificar componentes estacionales y tendencias a largo plazo.
  + Se realizaron análisis de autocorrelación para detectar patrones repetitivos en los datos.
  + Se generaron gráficos de línea para visualizar la evolución de casos confirmados, muertes y recuperaciones a lo largo del tiempo.

**Paso 4: Generar gráficas acordes a tu análisis**

* **Descripción:** Creación de visualizaciones avanzadas para representar hallazgos complejos.
* **Objetivo:** Proporcionar una visión clara y detallada de las relaciones y patrones en los datos.
* **Acciones:**
  + **Evolución de Casos Activos vs. Recuperados:** Gráficos de línea que muestran la evolución de casos activos y recuperados.
  + **Tasa de Crecimiento (%):** Gráficos de barras o líneas que muestran la tasa de crecimiento de casos y muertes.
  + **Relación entre la Cobertura de Vacunación y la Reducción de Casos:** Gráficos de dispersión que muestran la relación entre la tasa de vacunación y la reducción de casos.
  + **Progreso de la vacunación por país:** Gráficos de barras o líneas que muestran el progreso de la vacunación en cada país.
  + **Nuevos casos y temperatura promedio:** Gráficos de dispersión que muestran la relación entre la temperatura media y los nuevos casos.
  + **Impacto de la Urbanización en la Propagación del Covid-19:** Gráficos que comparan la incidencia de COVID-19 en áreas urbanas vs. rurales.
  + **Evolución Semanal y Anual de Casos Nuevos:** Gráficos de línea que muestran la evolución semanal y anual de nuevos casos.
  + **Distribución de la Población por Edad en Países:** Gráficos de barras que muestran la distribución de la población por grupos de edad.
  + **Comparación de Estrategias de Vacunación en América Latina:** Gráficos comparativos que muestran la efectividad de diferentes estrategias de vacunación.
  + **Evolución del Número de Muertes Diarias:** Gráficos de línea que muestran la evolución diaria de muertes.
  + **Prevalencia de Condiciones Preexistentes en Países con Altas y Bajas Tasas de Mortalidad:** Gráficos de barras o dispersión que muestran la relación entre condiciones preexistentes y tasas de mortalidad.
  + **Variación Mensual de Casos y Muertes:** Gráficos de barras o líneas que muestran la variación mensual de casos y muertes.
  + **Análisis Temporal de la Mortalidad:** Gráficos de línea que muestran la evolución temporal de la mortalidad.
  + **Comparación de la Situación Actual (Contexto de los datos):** Gráficos comparativos que muestran la situación actual en comparación con periodos anteriores.

**Paso 5: Investigar posibles correlaciones entre las distintas características**

* **Descripción:** Investigación de posibles correlaciones entre distintas características del conjunto de datos.
* **Objetivo:** Identificar relaciones significativas que puedan influir en la toma de decisiones.
* **Acciones:**
  + Se calcularon matrices de correlación utilizando Pandas para identificar correlaciones entre variables.
  + Se crearon mapas de calor para visualizar las correlaciones y resaltar relaciones fuertes (por encima de 0.5).
  + Se analizaron las correlaciones entre variables como la temperatura media, casos confirmados, muertes y tasas de vacunación.

**Paso 6: Aplicar las funciones creadas sobre cada columna de acuerdo a tus necesidades**

* **Descripción:** Aplicación de funciones personalizadas sobre las columnas del conjunto de datos según sea necesario.
* **Objetivo:** Automatizar tareas repetitivas y simplificar el análisis de datos.
* **Acciones:**
  + Se crearon y aplicaron funciones personalizadas para calcular estadísticas descriptivas (mediana, varianza, rango) sobre cada columna numérica.
  + Se utilizaron funciones para transformar y limpiar datos, asegurando que estuvieran listos para el análisis.
  + Se aplicaron funciones para realizar cálculos específicos según las necesidades del análisis, como la tasa de crecimiento y la variación porcentual.

**Extra Credit**

* **Descripción:** Creación de mapas utilizando la información de latitud y longitud disponible en los datos.
* **Objetivo:** Visualizar geoespacialmente la información y obtener una perspectiva más completa de la situación.
* **Acciones:**
  + Se utilizaron herramientas como Geopandas y Folium para crear mapas interactivos que muestran la distribución de casos de COVID-19 y la cobertura de vacunación por región.
  + Se crearon mapas de calor geoespaciales para resaltar áreas con alta incidencia de COVID-19 y baja tasa de vacunación.
  + Se exploraron diferentes plataformas y técnicas para enriquecer el análisis geoespacial y proporcionar insights adicionales sobre la distribución geográfica de los datos.

**Avance 4: Aplicaciones Prácticas - Integración en Power BI**

**Paso 1: Conexión de Python con Power BI**

* **Descripción:** Importación del conjunto de datos preparado y analizado en Power BI.
* **Objetivo:** Trasladar el análisis técnico a una plataforma que permita la visualización interactiva y la comprensión intuitiva de los datos por parte de los directivos.
* **Acciones:**
  + Exportación de datos limpios y procesados desde Python a archivos CSV, incluyendo indicadores como el número de camas por país y las posiciones óptimas de expansión.
  + Importación de estos archivos CSV en Power BI utilizando la funcionalidad de importación de datos.
  + Verificación de la correcta importación de los datos en Power BI, asegurando que todas las columnas y registros estaban presentes y en el formato correcto.

**Paso 2: Creación de Dashboards en Power BI**

* **Descripción:** Diseño de dashboards interactivos que muestren de manera efectiva los resultados del análisis de datos.
* **Objetivo:** Proporcionar una herramienta visual y interactiva para explorar los datos de incidencia de COVID-19, cobertura de vacunación y variables relacionadas.
* **Acciones:**
  + Creación de varias páginas en Power BI para organizar las visualizaciones temáticamente (incidencia de COVID-19, tasas de vacunación, infraestructura sanitaria).
  + Utilización de gráficos de líneas, barras, dispersión, mapas de calor y otras visualizaciones para representar los datos.
  + Adición de filtros y segmentadores interactivos para permitir a los usuarios explorar los datos de manera dinámica.
  + Generación de indicadores clave, como el número de camas por país y la posición óptima de expansión utilizando técnicas de escalado y ponderación de variables.

**Paso 3: Comparación de Visualizaciones Estáticas vs. Interactivas**

* **Descripción:** Evaluación de los beneficios de utilizar visualizaciones interactivas en Power BI frente a gráficos estáticos generados en Python.
* **Objetivo:** Destacar cómo Power BI permite una exploración más profunda y flexible de los datos.
* **Acciones:**
  + Comparación de visualizaciones generadas en Python (histogramas, gráficos de densidad, gráficos de dispersión) con sus equivalentes en Power BI.
  + Destacar la capacidad de Power BI para interactuar con los datos en tiempo real, ajustando filtros y viendo cambios inmediatos en las visualizaciones.
  + Evaluación de la facilidad de uso y la accesibilidad de los dashboards interactivos para los usuarios no técnicos.

**Entrega del Proyecto Integrador**

**Dashboard en Power BI**

**Requisitos:**

* **Interactividad:** Los dashboards deben permitir al usuario explorar diferentes dimensiones de los datos.
* **Visualizaciones Claras:** Las visualizaciones deben ser claras y ofrecer insights directos sobre las áreas de prioridad para la expansión.
* **Segmentaciones:** Incluir segmentaciones por ubicación geográfica, tendencias temporales de incidencia y cobertura de vacunación.

**Acciones:**

* Diseño de dashboards que incluyan visualizaciones interactivas como gráficos de líneas, barras, dispersión, mapas de calor y otros tipos de gráficos.
* Configuración de filtros y segmentadores para permitir la exploración dinámica de los datos.
* Aseguramiento de que las visualizaciones sean claras y proporcionen insights directos sobre las áreas prioritarias para la expansión.

**Informe Escrito**

**Requisitos:**

* **Documentación de Métodos:** Detallar los métodos de análisis utilizados.
* **Conclusiones y Recomendaciones:** Incluir conclusiones alcanzadas y cómo los hallazgos pueden guiar la estrategia de expansión de laboratorios y centros de vacunación.
* **Explicación de la Selección de Ubicaciones:** Explicar cómo la selección de ubicaciones estratégicas basada en los datos puede optimizar la respuesta a la pandemia y mejorar el acceso a las vacunas.

**Acciones:**

* Redacción de un informe que incluye la metodología utilizada, el análisis de datos, visualizaciones clave y conclusiones.
* Destacar los insights más importantes y proporcionar recomendaciones estratégicas para la expansión de laboratorios farmacéuticos.
* Explicar cómo se utilizaron indicadores como el número de camas por país y las posiciones óptimas de expansión para guiar las decisiones estratégicas.

**Notebook con Código Fuente**

**Requisitos:**

* **Código Fuente:** Incluir todo el código utilizado para el análisis y la visualización.
* **Comentarios:** Asegurar que el código esté bien comentado para explicar cada paso del proceso analítico.

**Acciones:**

* Recopilación de todo el código fuente utilizado en el análisis y la visualización en el notebook "PIDA\_M4\_Nombre\_Apellido.ipynb".
* Adición de comentarios detallados para explicar cada paso del proceso analítico, asegurando que fuera claro y comprensible.

**Extra Credit**

**Pasos Futuros y Mejoras Sugeridas:**

* **Incorporación de Más Fuentes de Datos:** Investigar cómo expandir el análisis incorporando más fuentes de datos.
* **Inteligencia de Tiempo en Power BI:** Explorar la implementación de inteligencia de tiempo para analizar tendencias y patrones temporales más complejos.
* **Modelos de Machine Learning:** Aplicar modelos de machine learning para predecir tendencias futuras y mejorar la toma de decisiones basada en datos.

**Acciones:**

* Incluimos el dato de las camas de hospitales disponibles.

Considero que se realice:

* Exploración de nuevas fuentes de datos que podrían ser integradas en futuros análisis para enriquecer los insights.
* Investigación de técnicas de inteligencia de tiempo en Power BI para mejorar el análisis de tendencias y patrones temporales.
* Consideración de la aplicación de modelos de machine learning para predecir la evolución de la pandemia y la efectividad de las campañas de vacunación.

# EDA e insights

**Análisis Exploratorio de Datos (EDA)**

**1. Importación y Preparación de Datos**

* **Descripción:** Los datos importados fueron sometidos a un riguroso proceso de limpieza y transformación.
* **Objetivo:** Asegurar que los datos sean consistentes, precisos y listos para el análisis.
* **Acciones:**
  + Importación de datos desde múltiples archivos CSV utilizando Pandas.
  + Eliminación de duplicados y manejo de valores nulos mediante imputación de medias y medianas según la relevancia de cada columna.
  + Normalización y escalado de variables numéricas para asegurar comparabilidad.
  + Generación de nuevas variables e indicadores clave, como la tasa de incidencia de COVID-19 y la cobertura de vacunación.

**2. Análisis Descriptivo**

* **Descripción:** Exploración inicial de las características del conjunto de datos para entender su distribución y estadísticas básicas.
* **Objetivo:** Obtener una comprensión general de los datos antes de realizar análisis más profundos.
* **Acciones:**
  + Cálculo de estadísticas descriptivas como media, mediana, desviación estándar y percentiles para cada variable.
  + Visualización de la distribución de las principales variables utilizando histogramas y gráficos de densidad.
  + Identificación de outliers mediante boxplots y análisis de su impacto en los resultados.

**3. Análisis de Correlación**

* **Descripción:** Evaluación de las relaciones entre diferentes variables del conjunto de datos.
* **Objetivo:** Identificar correlaciones significativas que puedan influir en la toma de decisiones.
* **Acciones:**
  + Cálculo de matrices de correlación utilizando el método de Pearson.
  + Visualización de las correlaciones mediante mapas de calor, destacando las relaciones más fuertes (por encima de 0.5).
  + Análisis de la relación entre variables clave, como la incidencia de COVID-19 y las tasas de vacunación, así como entre la infraestructura sanitaria (número de camas) y la tasa de mortalidad.

**4. Análisis Temporal**

* **Descripción:** Estudio de la evolución de las variables clave a lo largo del tiempo.
* **Objetivo:** Identificar tendencias, patrones estacionales y cambios significativos en los datos.
* **Acciones:**
  + Agrupación de datos por semana y mes para analizar la evolución temporal de casos confirmados, muertes y recuperaciones.
  + Uso de técnicas de descomposición de series temporales para separar componentes estacionales, tendencias y ruido.
  + Visualización de la evolución temporal mediante gráficos de líneas y áreas apiladas.

**5. Análisis Geoespacial**

* **Descripción:** Evaluación de la distribución geográfica de los datos.
* **Objetivo:** Identificar patrones espaciales y áreas prioritarias para la expansión de laboratorios.
* **Acciones:**
  + Creación de mapas interactivos utilizando Geopandas y Folium para mostrar la distribución de casos de COVID-19 y la cobertura de vacunación por región.
  + Generación de mapas de calor para resaltar áreas con alta incidencia de COVID-19 y baja tasa de vacunación.
  + Análisis de la relación entre la ubicación geográfica y la evolución de la pandemia.

**6. Análisis Comparativo entre Países**

* **Descripción:** Comparación de métricas clave entre los diferentes países de interés.
* **Objetivo:** Evaluar y comparar la evolución de la pandemia y la vacunación en los países seleccionados.
* **Acciones:**
  + Creación de gráficos de barras apiladas para comparar la tasa de mortalidad masculina vs. femenina por país.
  + Visualización de la distribución de la población por grupos de edad en cada país mediante gráficos de barras.
  + Comparación de la prevalencia de enfermedades preexistentes y la tasa de mortalidad en diferentes países utilizando gráficos de dispersión.

**7. Análisis de Tendencias y Patrones**

* **Descripción:** Identificación de tendencias a largo plazo y patrones relevantes en los datos.
* **Objetivo:** Prever áreas de alta prioridad para la asignación de recursos y la implementación de medidas preventivas y de control.
* **Acciones:**
  + Investigación de posibles influencias de factores como la temperatura en la propagación de la enfermedad.
  + Análisis de variaciones en los índices de vacunación en diferentes regiones y en distintos momentos del año.
  + Agrupación de datos por semana y mes para visualizar tendencias temporales y patrones estacionales.

**Insights**

**1. Distribución de la Incidencia de COVID-19 y Cobertura de Vacunación en Latinoamérica**

* **Insight:** La incidencia de COVID-19 muestra variaciones significativas entre los países estudiados, con Argentina y Brasil presentando las tasas más altas de casos confirmados y Chile con una cobertura de vacunación relativamente alta.
* **Acción:** Enfocar esfuerzos de expansión de laboratorios y centros de vacunación en Argentina y Brasil para controlar la propagación del virus y mejorar la cobertura de vacunación.
* **Visualización:** Mapas de calor y gráficos de barras que muestran la incidencia de COVID-19 y la cobertura de vacunación por país.

**2. Relación entre Infraestructura Sanitaria y Tasa de Mortalidad**

* **Insight:** Existe una correlación negativa moderada entre la disponibilidad de camas hospitalarias y la tasa de mortalidad por COVID-19, especialmente evidente en Perú y México.
* **Acción:** Invertir en la mejora de la infraestructura sanitaria en países con baja disponibilidad de camas, como Perú y México, para reducir la tasa de mortalidad.
* **Visualización:** Gráficos de dispersión y mapas interactivos que muestran la relación entre el número de camas hospitalarias y la tasa de mortalidad.

**3. Evolución Temporal de Casos Confirmados y Muertes**

* **Insight:** Los datos muestran picos significativos de casos confirmados y muertes en diferentes periodos.
* **Acción:** Planificar y fortalecer las medidas preventivas y campañas de vacunación durante los meses de mayor riesgo.
* **Visualización:** Gráficos de líneas que muestran la evolución temporal de casos confirmados y muertes por país.

**4. Desigualdad en la Cobertura de Vacunación**

* **Insight:** Existen desigualdades en la cobertura de vacunación entre los países, con Brasil, Colombia y México mostrando una cobertura menor en comparación con Chile, Perú y Argentina.
* **Acción:** Implementar estrategias específicas para aumentar la cobertura de vacunación en Brasil, Colombia y México, tales como unidades móviles de vacunación y campañas de sensibilización.
* **Visualización:** Gráficos de barras y mapas interactivos que muestran la cobertura de vacunación en diferentes países.

**5. Prevalencia de Enfermedades Preexistentes y Mortalidad**

* **Insight:** La prevalencia de enfermedades preexistentes, como la diabetes, está asociada con tasas más altas de mortalidad por COVID-19 en México y Brasil.
* **Acción:** Desarrollar programas de salud específicos para personas con condiciones preexistentes, incluyendo monitoreo y apoyo adicional para minimizar el riesgo de complicaciones graves.
* **Visualización:** Gráficos de barras y dispersión que muestran la relación entre enfermedades preexistentes y la tasa de mortalidad.

# 

# 

# Análisis del dashboard

**Relato sobre Cómo Navegar el Dashboard Realizado en Power BI**

El dashboard creado en Power BI para el estudio "Expansión Estratégica de Biogenesys" se compone de nueve pestañas, cada una enfocada en diferentes aspectos de los datos analizados. A continuación se describe cómo navegar por cada pestaña y qué información se puede obtener de cada una.

**1. Portada (Cover)**

* **Descripción:** La portada presenta el proyecto y proporciona un punto de entrada visualmente atractivo.
* **Componentes:**
  + Imagen representativa del estudio de vacunación.
  + Información del analista y detalles de contacto.

**2. Población (Population)**

* **Descripción:** Esta pestaña proporciona una visión general de la población en los países estudiados.
* **Componentes:**
  + **Gráfico de Barras:** Muestra la densidad de población por país.
  + **Gráfico de Líneas y Barras Combinado:** Presenta la expectativa de vida y el PIB per cápita por país.
  + **Gráfico de Barras:** Indica el índice de desarrollo humano por país.
  + **Mapa:** Visualización geográfica de la distribución de la población.
* **Navegación:** Utilice los gráficos para comparar la densidad poblacional y otros indicadores demográficos clave entre los países.

**3. Demografía (Demographics)**

* **Descripción:** Proporciona un análisis demográfico detallado de la población.
* **Componentes:**
  + **Gráfico de Barras Apiladas:** Muestra la población por rangos de edad para cada país.
  + **Gráficos de Pastel:** Presentan la distribución de la población por género y la división entre áreas urbanas y rurales.
* **Navegación:** Analice la composición demográfica para entender mejor las características de la población en cada país.

**4. Datos Acumulativos de COVID-19 (Cumulative COVID-19 Data)**

* **Descripción:** Presenta datos acumulativos sobre casos confirmados, recuperaciones y muertes por COVID-19.
* **Componentes:**
  + **Gráfico de Líneas:** Muestra la evolución de los casos confirmados acumulativos por país.
  + **Gráfico de Líneas:** Presenta la evolución de las recuperaciones acumulativas por país.
  + **Gráfico de Líneas:** Indica la evolución de las muertes acumulativas por país.
* **Navegación:** Utilice los gráficos de líneas para comparar la evolución de la pandemia en los diferentes países y analizar tendencias a lo largo del tiempo.

**5. Mortalidad y Riesgos para la Salud (Mortality and Health Risks)**

* **Descripción:** Analiza las tasas de mortalidad y los riesgos para la salud en los países estudiados.
* **Componentes:**
  + **Gráfico de Barras:** Compara la tasa de mortalidad por sexo y por país.
  + **Gráfico de Barras:** Presenta la tasa de mortalidad infantil por país.
  + **Gráfico de Barras Apiladas:** Muestra la prevalencia de diabetes, tabaquismo, y mortalidad relacionada con la contaminación y comorbilidades.
* **Navegación:** Compare los riesgos para la salud y las tasas de mortalidad entre los países para identificar áreas con mayores necesidades de intervención.

**6. Datos y Estrategia de Vacunación (Vaccination Data and Strategy)**

* **Descripción:** Proporciona una visión general de los datos y estrategias de vacunación en los países.
* **Componentes:**
  + **Gráfico de Barras:** Muestra la dosis máxima acumulada de vacunas administradas por país.
  + **Gráfico de Barras:** Presenta la estrategia de vacunación por país.
  + **Gráfico de Líneas:** Indica la evolución de las dosis acumuladas de vacunas administradas por mes y por país.
* **Navegación:** Analice los datos de vacunación para evaluar la efectividad de las estrategias y la cobertura de vacunación en cada país.

**7. Datos de Incidencia de COVID-19 (COVID-19 Incidence Data)**

* **Descripción:** Presenta datos sobre la incidencia de COVID-19 en los países estudiados.
* **Componentes:**
  + **Gráfico de Líneas:** Muestra la dosis acumulada de vacunas administradas por mes y por país.
  + **Gráficos de Líneas:** Presentan nuevos casos confirmados y nuevas muertes por COVID-19 por mes y por país.
* **Navegación:** Utilice los gráficos para monitorear la incidencia de nuevos casos y muertes, y evaluar la evolución de la pandemia en los diferentes países.

**8. Estadísticas de Salud y Vacunación (Healthcare and Vaccination Statistics)**

* **Descripción:** Proporciona datos sobre la infraestructura sanitaria y las estadísticas de vacunación.
* **Componentes:**
  + **Gráfico de Barras:** Muestra la cantidad de camas, enfermeras y médicos por 1000 habitantes en cada país.
  + **Gráfico de Barras y Líneas Combinado:** Indica la infraestructura hospitalaria, la tasa de vacunación y la suma de incidencias por país.
  + **Gráfico de Barras:** Presenta la tasa de vacunación, la incidencia y la infraestructura hospitalaria por país.
* **Navegación:** Compare la infraestructura sanitaria y las estadísticas de vacunación para identificar áreas que requieren mejoras en la capacidad sanitaria.

**9. Puntajes de Salud y Conclusiones (Country Healthcare Scores and Conclusions)**

* **Descripción:** Proporciona un resumen de los puntajes de salud y las conclusiones finales del análisis.
* **Componentes:**
  + **Tabla:** Muestra la infraestructura, incidencia, tasa de vacunación y densidad por país.
  + **Gráfico de Barras:** Presenta los puntajes de salud acumulados por país.
* **Navegación:** Utilice esta pestaña para obtener una visión general de los resultados del análisis y las recomendaciones estratégicas basadas en los puntajes de salud.

**Conclusiones del Dashboard**

1. **Priorización de Intervenciones:** El dashboard permite identificar rápidamente los países y regiones con mayor necesidad de intervención, especialmente aquellos con alta incidencia de COVID-19 y baja cobertura de vacunación.
2. **Tendencias Temporales:** La visualización temporal de los datos ayuda a entender mejor cómo evolucionan los casos y muertes a lo largo del tiempo, permitiendo una planificación más efectiva de las medidas sanitarias.
3. **Desigualdades en la Vacunación:** Se pueden observar claras desigualdades en la cobertura de vacunación entre y dentro de los países, lo que sugiere la necesidad de estrategias específicas para aumentar la vacunación en áreas desatendidas.
4. **Impacto de la Infraestructura Sanitaria:** La correlación entre la disponibilidad de camas hospitalarias y la tasa de mortalidad subraya la importancia de invertir en infraestructura sanitaria para reducir las muertes por COVID-19.
5. **Planificación Estratégica:** El análisis comparativo y geoespacial proporciona una base sólida para la planificación estratégica de la expansión de laboratorios y centros de vacunación, optimizando la asignación de recursos y maximizando el impacto positivo en la salud pública.

# 

# Conclusiones y Recomendaciones

**Conclusión General:**

El análisis exhaustivo y detallado realizado a lo largo del proyecto "Expansión Estratégica de Biogenesys" ha proporcionado una comprensión profunda de la incidencia de COVID-19, las tasas de vacunación y la infraestructura sanitaria en los países estudiados (Argentina, Brasil, Chile, Colombia, México y Perú). A través de técnicas avanzadas de análisis de datos y visualización, hemos identificado las ubicaciones óptimas para la expansión de laboratorios farmacéuticos y centros de vacunación, basándonos en datos sólidos y verificables. Este informe documenta todo el proceso, desde la limpieza y transformación de datos hasta la integración en Power BI para la visualización interactiva, destacando los insights clave y las recomendaciones estratégicas para optimizar la respuesta a la pandemia y mejorar el acceso a las vacunas en Latinoamérica.

**Conclusiones Estratégicas:**

1. **Priorización de Expansión en Brasil y Chile:**
   * **Conclusión:** Brasil y Chile presentan las tasas más altas en el puntaje combinado de infraestructura, incidencia de COVID-19 y tasa de vacunación, indicando una necesidad prioritaria de intervención.
   * **Recomendación:** Focalizar los esfuerzos de expansión de laboratorios y centros de vacunación en estos países para controlar la propagación del virus y mejorar la cobertura de vacunación.
2. **Expansión Estratégica en Argentina:**
   * **Conclusión:** Argentina muestra una alta incidencia de casos confirmados y una tasa de vacunación moderada, lo que sugiere la necesidad de fortalecer la infraestructura sanitaria.
   * **Recomendación:** Enfocar la expansión en áreas urbanas densamente pobladas y regiones con menor acceso a la infraestructura sanitaria para optimizar los recursos.
3. **Mejora de la Infraestructura Sanitaria en Perú:**
   * **Conclusión:** Perú presenta una alta tasa de mortalidad y baja infraestructura sanitaria a pesar de una tasa de vacunación relativamente alta.
   * **Recomendación:** Invertir en la mejora de la infraestructura sanitaria en Perú para reducir la tasa de mortalidad y aumentar la efectividad de las campañas de vacunación.
4. **Incrementar la Cobertura de Vacunación en México y Colombia:**
   * **Conclusión:** México y Colombia tienen infraestructuras sanitarias relativamente débiles y tasas de vacunación bajas.
   * **Recomendación:** Implementar estrategias específicas para aumentar la cobertura de vacunación, como unidades móviles de vacunación y campañas de sensibilización, y mejorar la infraestructura sanitaria para facilitar un despliegue más eficiente de las vacunas.
5. **Monitoreo y Planificación Basada en Tendencias Temporales:**
   * **Conclusión:** Los datos muestran picos significativos de casos confirmados y muertes en diferentes periodos, lo que requiere una respuesta oportuna y planificada.
   * **Recomendación:** Fortalecer las medidas preventivas y las campañas de vacunación durante los meses de mayor riesgo identificados a través de la evolución temporal de los casos y muertes.
6. **Desarrollo de Programas de Salud para Personas con Condiciones Preexistentes:**
   * **Conclusión:** La prevalencia de enfermedades preexistentes, como la diabetes, está asociada con tasas más altas de mortalidad por COVID-19 en países como México y Brasil.
   * **Recomendación:** Desarrollar programas de salud específicos para personas con condiciones preexistentes, incluyendo monitoreo y apoyo adicional, para minimizar el riesgo de complicaciones graves.

**Identificación de Ubicaciones Óptimas para la Expansión:**

Utilizando la ponderación de factores clave como la infraestructura, incidencia y tasa de vacunación, se han identificado las ubicaciones óptimas para la expansión de laboratorios farmacéuticos en Latinoamérica:

1. **Brasil:**
   * **Puntaje Total:** 0.62
   * **Motivo:** Alta incidencia de COVID-19, amplia población y necesidad de mejorar la infraestructura sanitaria.
   * **Acción:** Establecer nuevos laboratorios y centros de vacunación en las principales ciudades y regiones con alta incidencia de casos.
2. **Chile:**
   * **Puntaje Total:** 0.60
   * **Motivo:** Estrategia de vacunación efectiva y necesidad de apoyo en infraestructura para mantener y mejorar la tasa de vacunación.
   * **Acción:** Apoyar la infraestructura existente con nuevos laboratorios para mantener y mejorar la tasa de vacunación.
3. **Argentina:**
   * **Puntaje Total:** 0.47
   * **Motivo:** Alta incidencia de casos confirmados y una cobertura de vacunación que requiere expansión.
   * **Acción:** Focalizar la expansión en áreas urbanas densamente pobladas y regiones con menor acceso a la infraestructura sanitaria.
4. **Perú:**
   * **Puntaje Total:** 0.19
   * **Motivo:** Alta mortalidad y necesidad de mejorar la infraestructura sanitaria.
   * **Acción:** Expandir laboratorios en áreas con alta mortalidad y baja disponibilidad de camas hospitalarias para reducir la tasa de mortalidad.
5. **México:**
   * **Puntaje Total:** 0.05
   * **Motivo:** Necesidad de mejorar la infraestructura sanitaria y aumentar la cobertura de vacunación.
   * **Acción:** Establecer laboratorios en regiones con alta incidencia y baja cobertura de vacunación, especialmente en áreas rurales.
6. **Colombia:**
   * **Puntaje Total:** 0.04
   * **Motivo:** Necesidad de incrementar la cobertura de vacunación y mejorar la infraestructura sanitaria.
   * **Acción:** Focalizar la expansión en regiones con baja cobertura de vacunación y alta incidencia de COVID-19.

Estas recomendaciones y conclusiones proporcionan una guía estratégica para Biogenesys en su expansión en Latinoamérica, asegurando que los recursos se utilicen de manera eficiente y se maximice el impacto positivo en la salud pública.

# Reflexión personal

**Lo Aprendido Durante el Proyecto:**

Trabajar en el proyecto "Expansión Estratégica de Biogenesys" ha sido una experiencia enriquecedora que me ha permitido aplicar y expandir mis habilidades como Analista de Datos. A lo largo de este proceso, he aprendido a manejar grandes volúmenes de datos, limpiar y transformar datos para obtener conjuntos de datos utilizables y realizar análisis exploratorios detallados para extraer insights valiosos. La integración de estos análisis en Power BI para crear dashboards interactivos ha sido particularmente gratificante, ya que me ha permitido visualizar los resultados de manera clara y efectiva.

**Habilidades Adquiridas:**

1. **Análisis de Datos Avanzado:** He mejorado significativamente mi capacidad para analizar datos complejos utilizando Python y librerías como Pandas y Numpy. Esto incluye la habilidad para realizar análisis estadísticos, identificar tendencias y patrones, y explorar correlaciones entre diferentes variables.
2. **Visualización de Datos:** He desarrollado competencias avanzadas en la creación de visualizaciones efectivas utilizando herramientas como Matplotlib, Seaborn y Power BI. Aprendí a comunicar insights de manera clara y persuasiva a través de gráficos interactivos y dashboards.
3. **Limpieza y Transformación de Datos:** Adquirí experiencia práctica en la limpieza y transformación de datos, lo cual es fundamental para asegurar que los análisis sean precisos y confiables. Esto incluye el manejo de valores nulos, la normalización de datos y la creación de nuevas variables e indicadores.
4. **Integración de Python y Power BI:** Aprendí a integrar scripts de Python en Power BI, lo que me permitió combinar el poder analítico de Python con las capacidades de visualización de Power BI para crear dashboards interactivos y dinámicos.
5. **Toma de Decisiones Basada en Datos:** Este proyecto me ha ayudado a entender la importancia de utilizar datos para guiar la toma de decisiones estratégicas. Aprendí a evaluar diferentes escenarios y priorizar acciones basadas en los insights obtenidos del análisis de datos.

**Reflexión Sobre la Repetición del Proyecto:**

Si tuviera que volver a empezar este proyecto, hay varias cosas que mantendría igual debido a su efectividad, pero también hay algunas áreas donde haría cambios para mejorar el proceso y los resultados:

**Mantendría Igual:**

1. **Enfoque Metodológico:** Continuaría utilizando una metodología estructurada que abarca desde la limpieza y transformación de datos hasta el análisis exploratorio y la visualización. Este enfoque sistemático asegura que cada etapa del proyecto se realice de manera efectiva y contribuye a la obtención de resultados precisos.
2. **Uso de Herramientas:** Mantendría el uso de herramientas como Python y Power BI, ya que han demostrado ser extremadamente efectivas para el análisis y la visualización de datos.
3. **Integración de Fuentes de Datos:** Seguiría integrando múltiples fuentes de datos para obtener una visión más completa y precisa del problema. Esto permite analizar el contexto desde diferentes perspectivas y obtener insights más profundos.

**Cambiaría:**

1. **Planificación del Tiempo:** Dedicaría más tiempo a la fase inicial de planificación para asegurar que todas las etapas del proyecto tengan el tiempo adecuado asignado. Esto ayudaría a evitar contratiempos y asegurar que cada fase se complete de manera óptima.
2. **Automatización de Procesos:** Implementaría más scripts automatizados para la limpieza y transformación de datos. Esto no solo ahorra tiempo sino que también reduce el riesgo de errores humanos.
3. **Validación Cruzada de Datos:** Aumentaría el enfoque en la validación cruzada de datos para asegurar la precisión de los análisis. Esto incluye la verificación de datos desde múltiples fuentes y la utilización de técnicas de validación adicionales.
4. **Interacción con Stakeholders:** Involucraría más a los stakeholders desde el inicio del proyecto para asegurar que sus necesidades y expectativas estén claramente comprendidas y alineadas con los objetivos del proyecto.

**Conclusión Personal:**

Este proyecto ha sido una experiencia de aprendizaje increíblemente valiosa que ha fortalecido mis habilidades analíticas y mi capacidad para tomar decisiones basadas en datos. He ganado una comprensión más profunda de cómo los datos pueden guiar estrategias efectivas y cómo las visualizaciones pueden comunicar insights de manera poderosa. Si bien siempre hay oportunidades para mejorar, estoy orgulloso del trabajo realizado y confío en que las lecciones aprendidas aquí me servirán en futuros proyectos.

# EXTRA CREDIT

En el desarrollo del proyecto "Expansión Estratégica de Biogenesys", se realizaron varias mejoras y tareas adicionales en algunos de los avances, conforme a las oportunidades de extra credit que se presentaron. A continuación, se detallan las mejoras implementadas en cada uno de estos avances.

**Avance 1: Preparación de Datos**

* **Exploración de Funciones de Orden Superior:**
  + **Descripción:** Se exploró el uso de funciones de orden superior, como apply(), map(), y lambda en Python, para una manipulación más eficiente de los datos.
  + **Implementación:** Utilizamos estas funciones para transformar y limpiar los datos de manera más eficiente. Por ejemplo, se emplearon apply() y lambda para realizar operaciones complejas en columnas específicas del DataFrame, lo cual simplificó el código y mejoró la eficiencia del procesamiento de datos.

**Avance 2: Análisis Exploratorio - Visualización**

* **Extra Credit:** No se realizaron tareas adicionales en este avance específico.

**Avance 3: Análisis Temporal Avanzado**

* **Creación de Mapas Geoespaciales:**
  + **Descripción:** Utilizando Folium, se crearon mapas interactivos que visualizan la distribución geográfica de los casos confirmados de COVID-19 y las tasas de vacunación por país.
  + **Implementación:** Estos mapas permitieron explorar la distribución espacial de los datos, proporcionando una perspectiva más completa de la situación sanitaria en los países de interés. Se utilizaron coordenadas de latitud y longitud para ubicar los datos geográficamente.
* **Exploración de Machine Learning:**
  + **Descripción:** Se realizó una exploración preliminar de técnicas de Machine Learning para modelar y predecir tendencias futuras en la incidencia de COVID-19.
  + **Implementación:** Se prepararon los datos mediante la imputación de valores faltantes y la normalización de datos. Se implementaron modelos de regresión lineal simples para prever la evolución de los casos confirmados y las tasas de vacunación, proporcionando una base para la planificación proactiva.

**Avance 4: Integración en Power BI**

* **Extra Credit:** No se realizaron tareas adicionales en este avance específico.

**Entrega del Proyecto Integrador**

* **Investigación de Datos de Infraestructura Sanitaria:**
  + **Descripción:** Se investigaron y analizaron datos adicionales sobre la infraestructura sanitaria, específicamente el número de camas hospitalarias por mil habitantes en cada país.
  + **Implementación:** Estos datos se integraron en el análisis para evaluar la capacidad sanitaria de cada país y su preparación para la expansión de laboratorios farmacéuticos. Esto permitió identificar áreas con infraestructura insuficiente y priorizar intervenciones basadas en la capacidad hospitalaria.

**Resumen de Mejoras Extra Credit**

1. **Función de Orden Superior:** Implementación de funciones de orden superior en la manipulación de datos para mejorar la eficiencia y simplicidad del código.
2. **Mapas Geoespaciales con Folium:** Creación de mapas interactivos para visualizar la distribución geográfica de los casos confirmados de COVID-19 y las tasas de vacunación.
3. **Datos de Infraestructura Sanitaria:** Investigación y análisis de datos adicionales sobre la infraestructura sanitaria, específicamente el número de camas hospitalarias por mil habitantes.

Estas mejoras y tareas adicionales han enriquecido el proyecto, proporcionando insights más profundos y visualizaciones más efectivas. Han permitido una mejor comprensión de los datos y una toma de decisiones más informada, cumpliendo con los objetivos estratégicos de Biogenesys para la expansión de laboratorios farmacéuticos y centros de vacunación en Latinoamérica.