

Carrera: Data Analytics



Módulo 4

## Expansión Estratégica de Biogenesis con Python

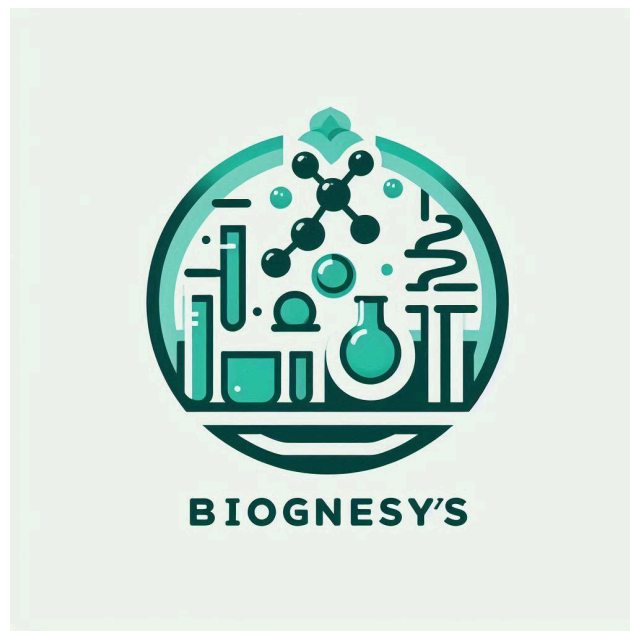
Nombre del autor: **Daniela Maria Forti Ruiz**

Email: danielafruiz@hotmail.com

Cohorte: **DAPT01**

Fecha de entrega: 29/07/2024

Institución: **BIOGENESYS**



Con este Análisis de Datos, el laboratorio farmacéutico **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 post pandemia con el fin de mejorar el acceso a las vacunas.

## Introducción

**BIOGENESYS** busca identificar **ubicaciones óptimas para la expansión de laboratorios farmacéuticos en Latinoamérica**, específicamente en **Colombia, Argentina, Chile, México, Perú y Brasil**. Para esto, se analizarán datos de incidencia de COVID-19, tasas de vacunación y la disponibilidad de infraestructuras sanitarias con el fin de optimizar la respuesta a los efectos de la pandemia y post pandemia, mejorando el acceso a las vacunas.

Este informe presenta un resumen detallado de nuestros hallazgos y recomendaciones, proporcionando una base sólida para la toma de decisiones estratégicas en BIOGÉNESYS.

## Objetivos

1. **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.**
2. **Aplicar técnicas de limpieza de datos para asegurar la calidad de los datos.**
3. **Mejorar el acceso a los datos mediante operaciones eficientes de extracción, transformación y carga (ETL).**
4. **Desarrollar dashboards interactivos con visualizaciones eficientes para la toma de decisiones informada y estratégica.**
5. **Ofrecer insights sobre posibles ubicaciones estratégicas para futuros laboratorios y centros de vacunación, abordando aspectos como la demanda de vacunas, la logística de distribución y la infraestructura sanitaria existente.**

## Desarrollo del Proyecto

### 1. Metodología de Recopilación y Selección de Datos

Para llevar a cabo este proyecto, utilizamos metodología de recopilación y selección de datos exhaustiva y precisa. Inicialmente, recibimos una base de datos extensa que contenía más de 12 millones de filas de datos relacionados con la pandemia de COVID-19 en varios países. El primer paso consistió en filtrar esta base de datos utilizando Python para reducir el conjunto de datos a solo los países latinoamericanos de interés para el laboratorio. Los países de interés son los siguientes: **Argentina, Perú, Brasil, Colombia, Chile y México**.

También se realizó un filtro por fechas de los datos desde el 1 de enero de 2021 en adelante.

Se realizaron transformaciones de datos claves para mejorar la calidad y del conjunto de datos. Esto incluyó convertir los formatos de fecha a datetime para facilitar el análisis temporal y cambiar algunos valores de objeto a numéricos para permitir cálculos precisos y análisis estadísticos adecuados.

Instalamos e importamos las siguientes librerías para el análisis de datos y la generación de gráficos:

1. **Pandas:** Para manejar y analizar datos en formato tabular, como leer archivos CSV y realizar operaciones de limpieza y transformación.
2. **NumPy:** Para realizar cálculos matemáticos y operaciones numéricas rápidas sobre arrays y matrices.
3. **Seaborn:** Para crear gráficos estadísticos atractivos y fáciles de interpretar, basados en Matplotlib.
4. **Matplotlib:** Para diseñar y personalizar gráficos, ajustando estilos, etiquetas y otros elementos visuales.

Estas herramientas nos permiten analizar datos y visualizar la información de manera efectiva.

Las técnicas de filtrado aplicadas en Python han permitido reducir considerablemente el volumen de datos a analizar, enfocándose en los países latinoamericanos clave para Biogenesys y en un periodo relevante para la pandemia actual.

Además, las transformaciones realizadas, como la conversión de formatos y la corrección de tipos de datos, han mejorado la calidad y la consistencia del conjunto de datos. Esto ha facilitado un análisis más preciso y la obtención de conclusiones valiosas.

## 2. Transformaciones y Limpieza de Datos

Las técnicas de limpieza y filtrado aplicadas en Python han permitido reducir considerablemente el volumen de datos a analizar, enfocándose en los países latinoamericanos clave para Biogénesis y en un periodo relevante para la pandemia actual.

Además, las transformaciones realizadas, como la conversión de formatos y la corrección de tipos de datos, han mejorado la calidad y la consistencia del conjunto de datos. Esto ha facilitado un análisis más preciso y la obtención de conclusiones valiosas.

Dentro de la limpieza se implementaron entre otras:

- Eliminación de valores nulos: Se eliminaron o imputaron valores faltantes.
- Normalización de datos: Ajuste de escalas y unidades para asegurar la comparabilidad.
- Creación de columnas calculadas: Tasas de incidencia, tasas de vacunación, etc.

En resumen, la metodología utilizada para recopilar, seleccionar y transformar datos ha proporcionado una base sólida para el análisis. Esto permitirá generar información clave que apoyará a Biogenesys en su estrategia de expansión y en su contribución a la lucha contra el COVID-19 en América Latina.

## **Análisis Exploratorio de Datos (EDA) y Principales Insights:**

### **Evolución de la Pandemia:**

- **Nuevos Casos Mensuales por País:** En general, se observa una tendencia descendente en el número de nuevos casos mensuales de COVID-19 en la mayoría de los países latinoamericanos. Sin embargo, Brasil presenta una excepción notable con un incremento reciente en los casos. Este patrón sugiere que, mientras la pandemia parece estar bajo control en la mayoría de las naciones de la región, Brasil podría estar enfrentando un brote o una ola prolongada.
- **Casos Activos vs. Casos Recuperados:** La cantidad de casos activos ha disminuido considerablemente en todos los países, al mismo tiempo que los casos recuperados han aumentado. Este cambio en las cifras indica una mejora en las tasas de recuperación y un avance en la gestión de la pandemia.

### **Vacunación:**

- **Número Máximo de Dosis Administradas por País:** Brasil destaca como el líder regional en la administración de dosis de vacunas, seguido por Argentina y México. Colombia y Perú se encuentran en una posición intermedia, mientras que Chile reporta el menor número de dosis administradas. Esta distribución refleja la capacidad y el ritmo de las campañas de vacunación en cada país.
- **Tendencia de Dosis Administradas por Mes:** Durante el último año, se ha registrado un aumento significativo en la administración de dosis en toda la región. Este incremento sugiere un avance positivo en la campaña de

vacunación y un esfuerzo continuo por parte de los gobiernos para ampliar la cobertura.

### **Mortalidad:**

- **Tasas de Mortalidad por Género y Edad:** Las tasas de mortalidad por COVID-19 son consistentemente más altas en hombres que en mujeres en todos los países analizados. Las tasas de mortalidad infantil son relativamente bajas en la mayoría de los países, con la excepción de Brasil, donde los datos indican una mayor mortalidad en esta categoría.

### **Aspectos a Considerar:**

- **Variantes del Virus:** Aunque los gráficos proporcionan una visión general sobre la evolución de la pandemia, no incluyen datos sobre la prevalencia de variantes del virus SARS-CoV-2 en Latinoamérica. La presencia de variantes más contagiosas o resistentes a las vacunas podría influir en la situación pandémica y en la demanda de vacunas, por lo que es crucial tener en cuenta esta información para futuras estrategias.
- **Tasas de Vacunación:** Los gráficos destacan el número máximo de dosis administradas, pero no detallan las tasas de vacunación por país o grupo etario. Analizar estas tasas permitirá identificar las áreas con menor cobertura y enfocar los esfuerzos de vacunación en esos segmentos específicos.
- **Efectividad de las Vacunas:** Los datos actuales no incluyen información sobre la efectividad de las vacunas frente a las distintas variantes del virus. Es vital seguir de cerca los estudios sobre la efectividad de las vacunas para ajustar las estrategias de vacunación y garantizar una protección adecuada contra nuevas variantes.

### **Consecuencias y Recomendaciones Estratégicas:**

#### **Vacunación:**

- **Desigualdad en el Acceso a las Vacunas:** Los datos revelan una disparidad notable en la distribución de vacunas en Latinoamérica. Mientras que países como Brasil y Argentina han administrado un número significativo de dosis, otros como Chile y Colombia presentan una cobertura mucho menor. Para abordar esta desigualdad, es crucial que la empresa desarrolle e implemente estrategias que amplíen el acceso a las vacunas en los países con menor cobertura.
- **Desafíos Logísticos en la Distribución:** La distribución de vacunas enfrenta obstáculos logísticos, especialmente en naciones con infraestructuras menos desarrolladas. Se recomienda que la empresa trabaje en la creación de

estrategias logísticas robustas para asegurar una distribución equitativa y eficiente de las vacunas en toda la región.

- **Aceptación de las Vacunas:** La aceptación de las vacunas varía significativamente entre países y grupos poblacionales. Es imperativo que la empresa lleve a cabo campañas de comunicación y educación para promover la aceptación de las vacunas y contrarrestar la desinformación que puede estar afectando la tasa de vacunación.

### **Mortalidad:**

- **Identificación de Factores de Riesgo:** Los gráficos actuales no proporcionan detalles sobre los factores que influyen en la mortalidad por COVID-19, como la edad, las comorbilidades y el acceso a la atención médica. La empresa debe investigar estos factores para identificar los grupos con mayor riesgo de mortalidad y enfocar sus esfuerzos de salud pública y vacunación en estas poblaciones vulnerables.
- **Impacto en los Sistemas de Salud:** La pandemia ha ejercido una presión significativa sobre los sistemas de salud en Latinoamérica. Es fundamental que la empresa colabore con los gobiernos locales para fortalecer las infraestructuras de salud y mejorar la atención médica para los pacientes con COVID-19.

### **Evolución de la Pandemia:**

- **Variantes del Virus:** Aunque los gráficos no incluyen información sobre las variantes del SARS-CoV-2, es crucial considerar que la aparición de variantes más contagiosas o resistentes a las vacunas podría alterar la dinámica de la pandemia. La empresa debe mantenerse actualizada sobre la evolución de estas variantes para ajustar sus estrategias de vacunación en consecuencia.
- **Tasas de Vacunación:** Los gráficos ofrecen una visión general de las dosis administradas, pero no detallan las tasas de vacunación por país o grupo de edad. Es necesario realizar un análisis más profundo para identificar las áreas y grupos con menor cobertura de vacunación, lo que permitirá focalizar mejor los esfuerzos de inmunización.
- **Efectividad de las Vacunas:** La información sobre la efectividad de las vacunas contra diversas variantes del virus no está reflejada en los gráficos actuales. La empresa debe monitorizar de cerca los estudios recientes sobre la eficacia de las vacunas para adaptar sus estrategias de vacunación y garantizar una respuesta efectiva frente a nuevas variantes.

## **Análisis de Vacunación y Consecuencias Estratégicas:**

### **Vacunación:**

- **Desigualdad en el Acceso a las Vacunas:** Los datos revelan una disparidad significativa en la distribución de vacunas en Latinoamérica. Brasil y Argentina destacan por haber administrado una mayor cantidad de dosis en comparación con países como Chile y Colombia. Esta desigualdad subraya la necesidad de implementar estrategias específicas para mejorar el acceso a las vacunas en las naciones con menor cobertura. Es crucial diseñar e implementar políticas y acciones que promuevan una distribución más equitativa de las vacunas en la región.
- **Logística de la Distribución de Vacunas:** La distribución de vacunas enfrenta desafíos logísticos considerables, especialmente en países con infraestructuras sanitarias deficientes. La empresa debe desarrollar planes logísticos robustos para asegurar una distribución eficiente y efectiva de las vacunas. Esto podría incluir la optimización de la cadena de suministro y la colaboración con organizaciones locales para superar obstáculos logísticos.
- **Aceptación de las Vacunas:** La aceptación de las vacunas muestra variabilidad entre diferentes países y grupos demográficos. Para abordar esto, la empresa debería lanzar campañas de comunicación dirigidas a aumentar la confianza en las vacunas y contrarrestar la desinformación. Las estrategias deben adaptarse a las particularidades culturales y sociales de cada país para ser efectivas.

### **Mortalidad:**

- **Factores de Riesgo:** Los gráficos actuales no proporcionan información sobre factores de riesgo clave para la mortalidad por COVID-19, como edad, comorbilidades y acceso a la atención médica. La empresa debería realizar un análisis detallado de estos factores para identificar las poblaciones de alto riesgo y ajustar las estrategias de salud pública y vacunación de manera más precisa y efectiva.
- **Impacto en los Sistemas de Salud:** La pandemia ha tenido un impacto profundo en los sistemas de salud de Latinoamérica. La empresa debe colaborar con los gobiernos locales para fortalecer las capacidades de los sistemas de salud y mejorar la calidad de atención para pacientes con COVID-19. Esto incluye el fortalecimiento de la infraestructura sanitaria y el apoyo a los profesionales de salud en la región.

## Condiciones de los Países:

- **Prevalencia de la Diabetes:** La prevalencia de diabetes es considerablemente más alta en Argentina y Brasil en comparación con Chile y Colombia. La empresa podría desarrollar e implementar estrategias de prevención y tratamiento para la diabetes en las naciones con alta prevalencia, enfocándose en la promoción de la salud y la gestión de esta enfermedad crónica.
- **Esperanza de Vida:** Se observa que la esperanza de vida es más alta en Chile y Argentina, mientras que es menor en Perú y Colombia. Para abordar esta disparidad, la empresa podría considerar estrategias de promoción de la salud y prevención de enfermedades en los países con menor esperanza de vida. Esto podría involucrar programas de salud pública y campañas educativas para mejorar la calidad de vida.
- **Índice de Desarrollo Humano:** El índice de desarrollo humano es más alto en Chile y Argentina, mientras que es más bajo en Perú y Colombia. La empresa debería enfocar sus esfuerzos en estos últimos países desarrollando estrategias de educación y capacitación que promuevan el desarrollo humano y la mejora de las condiciones socioeconómicas.

## Análisis del Dashboard

El dashboard interactivo creado en Power BI permite navegar por las diferentes métricas y visualizaciones de manera intuitiva. Los usuarios pueden:

Al acceder al dashboard, nos encontramos con una interfaz intuitiva y organizada por secciones:

- Explorar la **incidencia** de los **casos** de COVID-19 a lo largo del tiempo.
- Comparar las **tasas de vacunación** entre países.
- Evaluar la **infraestructura sanitaria** disponible en cada región.

A continuación hago una breve descripción del Dashboard:

### Página 1: Incidencia de COVID-19



En esta página, observamos una Línea Temporal que muestra la evolución de los casos por país. Podemos ver que algunos países presentan picos significativos en ciertos meses, lo cual sugiere brotes que podrían afectar nuestra estrategia de expansión.

Tenemos un Mapa Interactivo que nos ayuda a visualizar la distribución geográfica de los casos confirmados. Las regiones con mayor incidencia.

## **Página 2: Tasas de Vacunación**

En el Gráfico de Barras, comparamos las tasas de vacunación entre los países. Notamos que BRASIL lidera con la tasa más alta.

Máxima dosis administradas por país: Este gráfico de barras muestra la cantidad máxima de dosis de vacunas administradas por país. Brasil lidera la región, seguido de Argentina y México. Colombia y Perú se encuentran en una posición intermedia, mientras que Chile tiene la menor cantidad de dosis administradas.

Evolución de dosis por mes: Esta línea muestra la evolución de la cantidad de dosis de vacunas administradas en cada país. Se observa un aumento significativo en la cantidad de dosis administradas en todos los países, lo que indica que la campaña de vacunación estuvo avanzando en la región.

## **Página 3: Infraestructura Sanitaria:**

La infraestructura sanitaria es otro pilar fundamental en nuestra estrategia.

Vemos en esta pagina gráficos de:

Índice de desarrollo humano: Este gráfico de barras muestra el índice de desarrollo humano en cada país.

Camas hospitalarias, nurses y médicos c/1000 personas por país.

Identificación de Oportunidades y Riesgos: Un país con un índice de desarrollo alto puede tener mejores infraestructuras y recursos, mientras que uno con un índice bajo puede presentar mayores desafíos pero también pueden verse como potenciales oportunidades de crecimiento.

## **Conclusiones y Recomendaciones**

En conclusión, este análisis integral nos proporciona una visión clara de los factores clave para la expansión de BIOGENESYS en Latinoamérica.

- Incidencia de COVID-19: Las áreas con mayor incidencia necesitan atención prioritaria.
- Tasas de vacunación: Las áreas con bajas tasas requieren campañas intensivas.
- Infraestructura sanitaria: Las regiones con poca infraestructura necesitan inversión.

## **Conclusiones Generales:**

**Ubicaciones óptimas:** Identificación de regiones prioritarias para nuevos laboratorios y centros de vacunación.

**Políticas de vacunación:** Necesidad de campañas intensivas en áreas con bajas tasas.

**Inversión en salud:** Aumentar la infraestructura en regiones con alta incidencia y baja cobertura sanitaria.

### **Recomendaciones:**

**Nuevos laboratorios:** Priorizar regiones con alta incidencia y baja infraestructura.

**Centros de vacunación:** Establecer en áreas con bajas tasas de vacunación.

**Inversión en infraestructura:** Aumentar en regiones desatendidas para mejorar la capacidad de respuesta.

## Reflexión Personal:

Este proyecto me ha permitido desarrollar y perfeccionar varias habilidades clave en el campo del análisis de datos. Entre ellas, destacan:

### 1- Análisis de Datos:

**Exploración de Datos:** A través del análisis exploratorio de datos (EDA), aprendí a identificar patrones, tendencias y anomalías en grandes conjuntos de datos. Esencial para comprender mejor los datos antes de aplicar modelos predictivos o de visualización.

**Limpieza de Datos:** La importancia de la calidad de los datos fue reafirmada en este proyecto. Aprendí a manejar datos faltantes, eliminar duplicados y normalizar datos, asegurando que los análisis y las visualizaciones se basen en información precisa y confiable.

### 2 - Visualización de Datos:

**Herramientas de Visualización:** El uso de herramientas como Matplotlib, Seaborn y Power BI me permitió crear visualizaciones impactantes y comprensibles. Estas herramientas facilitan la comunicación de insights complejos de manera visual y más accesible.

**Dashboards Interactivos:** Desarrollé dashboards en Power BI que permiten a los usuarios interactuar con los datos, filtrando y explorando diferentes dimensiones y métricas. Esto es crucial para facilitar la toma de decisiones informadas en un entorno empresarial dinámico.

### 3 -Toma de Decisiones Estratégicas:

**Insights Accionables:** A través del análisis de datos, pude extraer insights valiosos que guían la toma de decisiones estratégicas, como la identificación de ubicaciones óptimas para nuevos laboratorios y centros de vacunación.

**Recomendaciones Informadas:** Aprendí a traducir datos en recomendaciones prácticas, considerando factores críticos como la incidencia de COVID-19, tasas de vacunación y disponibilidad de infraestructuras sanitarias.

### Integración de Python con Power BI:

**Scripts de Python en Power BI:** Implementé scripts de Python directamente en Power BI, permitiendo un análisis más avanzado y personalizado. Esta integración combina la potencia del análisis en Python con la accesibilidad y la interactividad de Power BI.

### **Lecciones Aprendidas:**

- **Exploración Inicial:** Si tuviera que comenzar este proyecto de nuevo, dedicaría más tiempo a la exploración inicial de los datos. Esta fase es crucial para identificar todas las variables relevantes y definir métricas clave desde el principio.
- **Definición de Métricas:** Una clara definición de las métricas a utilizar facilita el análisis posterior y aseguraría que todos los aspectos importantes estén cubiertos.

## Extra Credit

Para llevar este proyecto a un nivel superior, se podrían implementar las siguientes mejoras:

### 1. Análisis de Correlación Avanzado:

- **Correlación entre Variables:** Utilizar técnicas estadísticas avanzadas para explorar la correlación entre la incidencia de COVID-19, tasas de vacunación y otras variables socioeconómicas. Esto podría incluir el uso de coeficientes de correlación de Pearson o Spearman, y análisis de regresión múltiple para identificar relaciones significativas.
- **Análisis de Componentes Principales (PCA):** Implementar PCA para reducir la dimensionalidad de los datos y descubrir patrones subyacentes que no son evidentes en el análisis univariado y bivariado.

### 2. Integración de Datos Socioeconómicos:

- **Datos Demográficos y Económicos:** Incorporar datos adicionales como el nivel de ingresos, densidad de población, acceso a servicios de salud y niveles de educación. Estos factores pueden influir significativamente en la propagación del virus y en la efectividad de las campañas de vacunación.
- **Análisis de Disparidades Regionales:** Evaluar cómo las diferencias socioeconómicas entre regiones afectan la incidencia de COVID-19 y la cobertura de vacunación, proporcionando un contexto más completo para la toma de decisiones.

### 3. Enfoque en la Logística de Distribución de Vacunas:

- **Optimización de la Cadena de Suministro:** Realizar un análisis detallado de la cadena de suministro de vacunas, desde la producción hasta la entrega final. Esto podría incluir la evaluación de la capacidad de almacenamiento en frío, rutas de distribución y tiempos de entrega.
- **Modelos de Simulación:** Implementar modelos de simulación para prever posibles cuellos de botella en la distribución y diseñar estrategias para mitigarlos. Esto aseguraría una distribución más eficiente y equitativa de las vacunas.

### 4. Modelos Predictivos:

- **Predicción de Incidencia:** Desarrollar modelos predictivos utilizando técnicas de machine learning para prever la evolución de la incidencia de COVID-19 en diferentes regiones. Esto permitiría a los tomadores de decisiones anticipar y planificar mejor las necesidades futuras.
- **Eficacia de las Vacunas:** Analizar datos sobre la efectividad de diferentes vacunas en diversas poblaciones, ajustando las estrategias de vacunación en función de los resultados.

Estas mejoras no sólo enriquecerán el análisis actual, sino que también proporcionarán una base más robusta para la planificación estratégica y la toma de decisiones en BIOGENESYS, contribuyendo significativamente a la optimización de la respuesta a la pandemia y a futuras crisis sanitarias.