



# **Expansión Estratégica de Biogenesys con Python**

**Autor:**

Roman Matheus

**Correo Electrónico:**

roman.matheus.uni@gmail.com

**Cohorte:**

DAFT-13

**Fecha de Entrega:**

12/05/2025

## **Introducción**

La empresa farmacéutica BIOGENESYS busca identificar las ubicaciones óptimas para la expansión de laboratorios farmacéuticos en América Latina, centrándose en seis países clave: Colombia, Argentina, Chile, México, Perú y Brasil. El objetivo principal es mejorar la respuesta ante pandemias y optimizar el acceso a las vacunas en regiones estratégicas mediante el análisis de datos epidemiológicos, demográficos y sanitarios.

BIOGENESYS nos ha contratado como equipo de Data Analyst para realizar un estudio exhaustivo basado en datos reales, con el fin de ofrecer recomendaciones sobre dónde establecer nuevos centros de vacunación y laboratorios, con un enfoque preventivo y de desarrollo de infraestructura sanitaria.

## **Desarrollo del Proyecto**

### **Avance 1 Recopilación y Preparación de Datos**

Se recopilaron y consolidaron datos relevantes de fuentes oficiales sobre COVID-19, vacunación, salud pública y condiciones demográficas. Se seleccionaron los datos más significativos para el análisis y se creó el archivo maestro `Df_datos_finales_filtrado.csv`.

Las tareas clave incluyen: Filtrado por los seis países objetivos. Luego el filtrado por las fechas deseadas que son los años 2021-2022. Eliminación de registros totalmente en blanco y filtrado por `location_key`.

Limpieza de datos: eliminación de duplicados, nulos y datos inconsistentes, usando funciones para tratar los datos faltantes y los datos nulos.

### **Avance 2 Análisis Exploratorio y Visualización**

Utilizando Pandas, NumPy, Matplotlib y Seaborn, se llevaron a cabo los siguientes análisis:

Estadísticas descriptivas: medidas de tendencia central y dispersión.

Gráficos: histogramas, boxplots, mapas de calor, diagramas de barras y dispersión.

Correlaciones: se identificaron asociaciones entre la temperatura, densidad poblacional y el número de casos y muertes por COVID-19.

Algunos insights clave: Regiones con mayor temperatura promedio mostraron menor incidencia de casos. Alta correlación entre densidad poblacional y propagación del virus. Déficit de infraestructura sanitaria en zonas con alta mortalidad.

### Avance 3 Análisis Temporal y Tendencias

Se trabajó con series temporales para detectar: Tendencias de vacunación y contagios por mes. Estacionalidad en las tasas de incidencia. Autocorrelaciones y picos epidemiológicos.

Este análisis permitió prever áreas de alta prioridad y periodos críticos que requieren intervención rápida.

### Avance 4 Dashboard Interactivo en Power BI

Se construyó un dashboard interactivo para facilitar la exploración de los datos por parte de los directivos.

Este dashboard incluye: Filtros por país, fecha, variable y tipo de región. KPIs dinámicos de salud y vacunación. Evolución mensual de casos y dosis administradas.

## EDA e Insights

Los hallazgos más destacados del Análisis Exploratorio incluyen:

Distribución desigual de la infraestructura de salud en relación con la incidencia de casos.

Prevalencia de enfermedades preexistentes como diabetes y tabaquismo relacionada con tasas de mortalidad elevadas.

Desigualdad urbana-rural: zonas rurales presentan menor cobertura de salud y vacunación.

Edad media y distribución etaria influyen la tasa de recuperación y mortalidad.

## Análisis del Dashboard

El dashboard desarrollado en Power BI ofrece una vista integral y dinámica del impacto del COVID-19 en seis países latinoamericanos: Argentina, Brasil, Chile, Colombia, México y Perú. Su diseño facilita una navegación intuitiva mediante segmentadores por temporalidad, país, nivel de recuperación y temperatura media, lo que permite realizar análisis comparativos y específicos por contexto.

En la parte superior se resumen métricas clave como la población total, nuevos casos confirmados, nuevos recuperados, nuevas muertes y promedio de temperatura, ofreciendo un panorama inmediato del estado actual de la pandemia. Las visualizaciones principales incluyen:

- Ritmo de propagación del virus por país: Se observa un crecimiento dispar, con Brasil destacando por su volumen de casos.

- Impacto del virus por país: Representa los picos y olas de contagios, especialmente visibles entre 2021 y 2022.
- Relación entre comorbilidades y mortalidad: Chile presenta los mayores niveles combinados de prevalencia de diabetes, tabaquismo y comorbilidades, lo que podría explicar mayores tasas de mortalidad.
- Vacunación por país: Brasil lidera con claridad la cantidad de dosis administradas, seguido por México y Perú.
- Gravedad del impacto entre países: Muestra que Perú y México tienen los porcentajes más altos de impacto severo en relación a su población.
- Comparativa entre nuevos casos y temperatura promedio: El gráfico de dispersión revela que no existe una correlación clara y directa entre temperatura y nuevos casos, pero sí se identifican agrupamientos por país que pueden guiar análisis más específicos.
- Matriz de correlación: Facilita la identificación de relaciones fuertes entre variables como infraestructura sanitaria, densidad de población, tasas de vacunación, y comorbilidades.

Este dashboard no solo permite monitorear el comportamiento del virus y la vacunación, sino que también es una herramienta poderosa para tomar decisiones estratégicas. Su integración de múltiples variables lo convierte en un recurso fundamental para definir ubicaciones prioritarias para la expansión de laboratorios y centros de vacunación, permitiendo a Biogenesys optimizar sus recursos e intervenir en las regiones con mayor necesidad e impacto.

## **Conclusiones y Recomendaciones**

### **Conclusiones Generales**

Existen patrones de concentración de casos en zonas urbanas densamente pobladas.

La vacunación ha avanzado de manera desigual, con varios países mostrando cobertura insuficiente en zonas rurales.

Las tasas de mortalidad se ven afectadas por comorbilidades, acceso a servicios sanitarios y condiciones económicas.

### **Conclusiones Estratégicas**

1. Priorizar la expansión de laboratorios en zonas de alta densidad con baja infraestructura sanitaria.
2. Ampliar centros de vacunación en regiones rurales con baja cobertura actual.
3. Considerar indicadores como prevalencia de diabetes, mortalidad por contaminación y edad promedio al definir ubicaciones.

4. Diseñar estrategias diferenciadas por país considerando factores socioeconómicos.

### **Respuesta Principal del Proyecto**

Se recomienda iniciar la expansión de laboratorios y centros de vacunación en las siguientes ubicaciones estratégicas:

Brasil (alta población, baja infraestructura sanitaria).

Regiones andinas de Perú y Colombia (zonas rurales densas con alto riesgo).

Conurbano Bonaerense, Argentina (alta densidad y desigualdad en acceso a salud).

Estas ubicaciones permitirán a Biogenesys maximizar el impacto positivo de su inversión, reforzar la salud pública regional y prepararse para futuras crisis sanitarias.