**Expansión Estratégica de Biogenesys con Python**

**Nombre del autor:** Camila Michel López Herrejón

**Email:** camii.herrejon@gmail.com

**Cohorte:** DAFT - 16

**Fecha de entrega:** Lunes 04 de agosto de 2025

**Institución:** BIOGENESYS

## **1. Introducción**

Durante la pandemia de COVID-19, la necesidad de tomar decisiones rápidas y basadas en datos se convirtió en un imperativo para organizaciones sanitarias, gobiernos y empresas del sector salud. En este contexto, Biogénesis, una empresa dedicada a la expansión de soluciones médicas en América Latina, requiere identificar las regiones óptimas para invertir en infraestructura sanitaria, específicamente en laboratorios farmacéuticos.

El presente proyecto tiene como propósito analizar datos de salud pública, demografía, vacunación y factores de riesgo asociados al COVID-19 en seis países latinoamericanos: Argentina, Brasil, Chile, Colombia, México y Perú. Se pretende entregar a Biogénesis un dashboard interactivo y un informe que le permita tomar decisiones estratégicas de expansión, priorizando variables como cantidad de contagios, disponibilidad médica, porcentaje de población vulnerable y nivel de vacunación.

A lo largo del proyecto se aplicaron técnicas avanzadas de limpieza y análisis exploratorio en Python, seguidas de una visualización interactiva en Power BI. Cada paso fue documentado y diseñado para que las conclusiones puedan derivar en recomendaciones accionables, alineadas con los intereses y la misión de Biogénesis como empresa líder en innovación y salud.

## **2. Desarrollo del Proyecto**

El proceso de análisis comenzó con la selección y depuración de un extenso dataset de salud pública en América Latina, que contenía más de 12 millones de registros relacionados con casos confirmados, muertes, vacunación, variables climáticas y demográficas. Para enfocar el análisis en lo que realmente le interesa a Biogénesis, se aplicaron filtros estratégicos para conservar únicamente los datos a partir de enero de 2021 y correspondientes a seis países claves: Argentina, Brasil, Chile, Colombia, México y Perú.

Si bien inicialmente se planeaba incluir información hasta finales de 2022, durante el desarrollo de los gráficos en Power BI se detectó que el dataset disponible contenía registros solo hasta septiembre de ese año. Esta limitación fue identificada durante la validación visual, lo que evidencia la importancia del análisis iterativo y detallado.

Posteriormente, se aplicaron técnicas rigurosas de limpieza y transformación de datos, incluyendo el tratamiento de valores nulos, la verificación de tipos de datos y la imputación diferenciada según el contexto de cada variable. Por ejemplo, las columnas de clima con pocos nulos fueron rellenadas con su media general para conservar la distribución, mientras que las variables sanitarias como recuperados o dosis aplicadas fueron imputadas con ceros en los casos donde se infirió que la falta de datos representaba ausencia de reporte.

Una vez limpio el dataset, se desarrolló un análisis exploratorio profundo utilizando herramientas como Pandas, Seaborn y Matplotlib. Se calcularon medidas estadísticas por país (promedios, varianzas, medianas y rangos) y se diseñaron visualizaciones para detectar patrones en la propagación del virus, diferencias por género, edad, y relaciones con factores ambientales como temperatura o urbanización. A partir de esto, se identificaron hallazgos clave, como la alta concentración de muertes en hombres.

Con base en el dataset limpio, se exportó un archivo final (DatosFinalesFiltrado.csv) que sirvió como fuente directa para la visualización en Power BI. En este entorno se diseñó un dashboard interactivo dividido en cuatro secciones: Portada, Análisis General, Vacunación y Demografía & Riesgo. Cada sección fue pensada para responder una serie de preguntas estratégicas de negocio, con filtros sincronizados por país y año, tarjetas de indicadores clave, gráficos personalizados y medidas DAX específicas para facilitar la toma de decisiones.

## **3. EDA e Insights**

El análisis exploratorio de datos tuvo como objetivo detectar patrones relevantes, desigualdades sanitarias y tendencias de comportamiento del COVID-19 en América Latina. A través de la combinación de visualizaciones estadísticas y temporales, se logró convertir datos complejos en conocimientos accionables para la toma de decisiones estratégicas.

### **Página 1: Análisis General**

#### ▪ Tarjetas de indicadores clave (Casos Confirmados, Recuperados, Muertes, Vacunas Administradas)

Estas tarjetas resumen el impacto agregado del COVID-19 en los seis países analizados. Se observa que:

* Brasil concentra la mayor cantidad de casos y muertes, reflejando una propagación más intensa y prolongada del virus.
* México también presenta cifras altas, lo cual refuerza la hipótesis de que la densidad poblacional, la urbanización y el acceso desigual a salud pública fueron factores clave.
* Las dosis administradas muestran una correlación positiva con los países de mayor población, pero no siempre con mejores resultados sanitarios.

#### ▪ Gráfico de líneas: evolución mensual de casos confirmados por país

Este gráfico permite observar la dinámica de los contagios a lo largo del tiempo.

* Se detectan picos significativos durante el segundo y tercer trimestre de 2021, especialmente en Brasil.
* Algunos países como Perú y Chile presentan curvas más planas, lo que podría estar relacionado con respuestas sanitarias más oportunas o menor densidad urbana.

#### ▪ Mapa geográfico: muertes acumuladas por país

* El mapa confirma que Brasil y México concentran los mayores valores de muertes absolutas, seguidos por Argentina y Colombia.
* Los países como Perú y Chile muestran cifras menores, lo cual podría atribuirse a factores geográficos y demográficos.

Este mapa facilita visualizar el riesgo sanitario por país. Las regiones más afectadas requieren más inversión en atención post-pandemia, rehabilitación y capacidad hospitalaria.

#### ▪ Gráfico de columnas: comparación de muertes por país entre 2021 y 2022

* Se observa una disminución clara de muertes en casi todos los países en 2022, lo que puede estar relacionado con el avance de la vacunación.

Este descenso puede interpretarse como un indicador de efectividad de políticas sanitarias y de vacunación, útil para modelar estrategias en otros países.

### **Página 2: Vacunación**

#### ▪ Gráfico de área acumulada: dosis administradas por país (2021-2022)

* Brasil lidera la administración de vacunas a lo largo del tiempo, seguido por México y Argentina.
* Chile, pese a su menor tamaño poblacional, muestra un avance sostenido.

Chile se destaca por su eficiencia en vacunación relativa a su población, lo que posiciona al país como referente logístico regional.

#### ▪ Barras horizontales: dosis totales por país

* Este gráfico refuerza el hallazgo anterior y permite comparaciones absolutas rápidas.
* Muestra un patrón directamente relacionado con la población, pero también con el acceso al sistema de salud.

El acceso equitativo a vacunas no siempre es proporcional a la necesidad sanitaria. Esto revela oportunidades para intervenir en zonas con déficits de cobertura relativa.

#### ▪ Gráfico de barras apiladas: dosis por género

* Los datos muestran una distribución bastante equitativa entre hombres y mujeres, aunque en algunos países se observa una leve mayoría femenina.
* Esto podría deberse a diferencias en los grupos prioritarios o en el acceso temprano a campañas de vacunación.

#### ▪ Gráfico de líneas: población vacunada por grupo etario

* La mayor cobertura se concentra en personas entre 20 y 39 años, lo que corresponde a la población económicamente activa.
* La vacunación en mayores de 60 también es significativa, aunque decrece en el grupo de 80+.

La estrategia de vacunación priorizó la economía activa y los grupos de riesgo, pero aún queda margen para reforzar protección en adultos mayores con movilidad limitada.

### **Página 3: Demografía y Riesgo**

#### ▪ Gráfico de barras por sexo: tasa de mortalidad por enfermedades crónicas

* Este gráfico muestra que en todos los países la tasa de mortalidad masculina es consistentemente más alta que la femenina.
* Brasil y México lideran en mortalidad masculina, mientras que Chile presenta una brecha menor.

Este patrón puede reflejar tanto diferencias biológicas como brechas en el acceso a la salud o hábitos de vida entre hombres y mujeres. Para Biogénesis, esto sugiere que futuras campañas de concientización o prevención podrían beneficiarse si están segmentadas por género, sobre todo en regiones de alta mortalidad masculina.

#### ▪ Gráfico de dispersión: urbanización vs mortalidad por contaminación

* Este gráfico cruza el porcentaje de población urbana con la mortalidad por enfermedades respiratorias atribuibles a la contaminación del aire.
* Se observa que Brasil y México tienen altos niveles de urbanización, pero su nivel de mortalidad por contaminación no es necesariamente el más elevado.
* Por otro lado, países como Perú o Colombia presentan tasas más preocupantes en relación a su nivel de urbanización.

Este análisis permite identificar zonas urbanas con alta vulnerabilidad ambiental, incluso en países que no destacan por su densidad. Para Biogénesis, esto representa una oportunidad estratégica para instalar laboratorios o centros de salud en ciudades con baja cobertura médica pero alta exposición a factores de riesgo ambiental.

#### ▪ Gráfico de columnas: porcentaje de población de adultos mayores (60+)

* Chile lidera con el mayor porcentaje de adultos mayores, seguido de cerca por Brasil
* Perú y México tienen una población más joven, lo cual puede modificar el tipo de atención médica que requieren.

La población envejecida es más propensa a sufrir complicaciones graves por COVID-19 y otras enfermedades. Por eso, Biogénesis debería priorizar infraestructura médica geriátrica en países como Chile y Brasil, donde la demanda futura será mayor.

#### ▪ Gráfico de área: médicos por cada 1000 habitantes

* Argentina se destaca como el país con mayor disponibilidad médica, seguido por Chile.
* Brasil y Perú se ubican por debajo del promedio latinoamericano, lo que podría generar saturación en sus sistemas de salud.

La capacidad médica instalada es un indicador directo de resiliencia sanitaria. Biogénesis podría encontrar más espacio para expandir servicios o laboratorios en países con menor cobertura médica, donde su intervención tendría mayor impacto.

## **4. Análisis del Dashboard**

El dashboard final fue diseñado en Power BI con un enfoque estratégico, visualmente atractivo y funcional para usuarios no técnicos. Su estructura se divide en cuatro páginas principales, organizadas con hipervínculos, filtros interactivos y elementos visuales intuitivos que permiten al usuario explorar los datos de forma ágil y personalizada.

### **▪ Navegación general**

Desde la portada, el usuario puede acceder a cada sección mediante botones con diseño moderno y colores llamativos (azul, verde y rosa neón sobre fondo oscuro). Esta estética no solo realza la experiencia visual, sino que también facilita la diferenciación entre secciones temáticas.

Cada página contiene:

* Un segmentador sincronizado de Año, de Mes y de País
* Gráficos adaptados al tipo de dato (líneas para evolución, áreas para acumulación, columnas para comparaciones, etc.).
* Tarjetas y etiquetas que resumen los indicadores clave por país.

### **▪ Lógica por secciones**

1. **Análisis General**: resume los principales indicadores sanitarios del COVID-19 (casos, muertes, dosis administradas), su evolución en el tiempo y su distribución entre países. Permite detectar picos, tendencias y comparaciones anuales clave para evaluar la efectividad de políticas públicas.
2. **Vacunación**: profundiza en el avance y la cobertura de las campañas de vacunación. Incluye comparativas absolutas, análisis por género y edad, y evolución acumulada mensual. Esta sección permite a Biogénesis detectar posibles brechas de cobertura.
3. **Demografía y Riesgo**: cruza variables sociales, demográficas y médicas para identificar poblaciones vulnerables. Incluye visualizaciones sobre mortalidad por sexo, contaminación, envejecimiento y disponibilidad de personal médico. Esta página es clave para tomar decisiones de expansión basadas en necesidades específicas de cada país.

### **▪ Valor agregado visual**

El dashboard fue diseñado con:

* Paleta oscura elegante (#0D1B2A) con contrastes neón para mejorar la legibilidad.
* Distribución simétrica de los elementos para evitar sobrecarga visual.
* Gráficos combinados (barras, líneas, áreas, dispersión) que optimizan la lectura de datos según su naturaleza.

## **5. Conclusiones y Recomendaciones**

A lo largo de este proyecto se analizó la evolución del COVID-19 en seis países de América Latina entre enero de 2021 y septiembre de 2022. Se estudiaron variables sanitarias, demográficas, climáticas y sociales con el objetivo de identificar oportunidades estratégicas para la expansión de laboratorios farmacéuticos por parte de Biogénesis.

**Conclusiones generales**

* Brasil y México concentraron los valores más altos de casos confirmados y muertes, lo que evidencia una gran necesidad de capacidad médica y prevención continua.
* La vacunación avanzó de forma desigual: Chile mostró gran eficiencia relativa, mientras que México y Perú reflejan áreas con posibles brechas de cobertura.
* Las muertes por enfermedades crónicas y por contaminación afectan más a los hombres y a poblaciones urbanas con alta densidad, lo cual plantea retos específicos de salud pública.
* El grupo de edad predominante en contagios y vacunación corresponde a la población económicamente activa (20-39 años), pero aún persisten riesgos entre adultos mayores.
* La disponibilidad de médicos por habitante es muy variable: Argentina y Chile están bien posicionados, mientras que Brasil y Perú presentan oportunidades para reforzar infraestructura médica.

**Recomendaciones estratégicas para Biogénesis**

1. Priorizar expansión en Brasil y Perú  
   Ambos países presentan una combinación de alta carga sanitaria, baja disponibilidad médica y crecimiento urbano acelerado. Instalar laboratorios en estas zonas permitiría cubrir necesidades urgentes con alto impacto social.
2. Fortalecer atención geriátrica en Chile y Brasil  
   Estos países concentran los mayores porcentajes de población mayor a 60 años. Invertir en laboratorios enfocados en medicina geriátrica, rehabilitación y diagnóstico temprano en estas regiones permitiría anticiparse a la demanda futura.
3. Establecer alianzas logísticas en Chile  
   Dada su eficiencia relativa en vacunación, Chile representa un excelente modelo de gestión sanitaria. Biogénesis podría replicar su experiencia o establecer colaboraciones con actores locales.
4. Instalar laboratorios móviles en zonas urbanas contaminadas  
   En países como Colombia y Brasil, donde la urbanización y la contaminación se asocian a mortalidad respiratoria, se recomienda implementar laboratorios móviles o satelitales que permitan monitoreo ambiental y control temprano de enfermedades.
5. Desarrollar campañas diferenciadas por género  
   Dado que la mortalidad masculina fue sistemáticamente mayor, Biogénesis podría implementar campañas de concientización orientadas a hombres adultos en los países con mayor brecha, como Brasil y México.

## **6. Reflexión personal**

Este proyecto fue un reto desafiante y a la vez profundamente enriquecedor. Me permitió aplicar de forma práctica todo lo aprendido en el módulo, desde la limpieza de datos con Python hasta el diseño de visualizaciones interactivas con Power BI. A lo largo del proceso, no solo desarrollé habilidades técnicas, sino también la capacidad de interpretar los datos con un enfoque crítico y estratégico, pensando siempre en el impacto real que pueden tener estas decisiones para una organización como Biogénesis.

Lo más valioso para mí fue aprender a convertir datos en conocimiento accionable. En lugar de limitarme a mostrar números, entendí cómo usar gráficos y medidas para responder preguntas de negocio, identificar patrones reales y proponer soluciones concretas. El análisis visual me ayudó a ver de forma más clara las desigualdades sanitarias entre países, las brechas por edad y género, y los posibles focos de intervención médica.

En definitiva, este proyecto me ayudó a consolidar mi perfil como analista de datos, con una visión más analítica, crítica y enfocada en generar valor real a partir de los datos.