**Expansión Estratégica de “BioGenesis”**

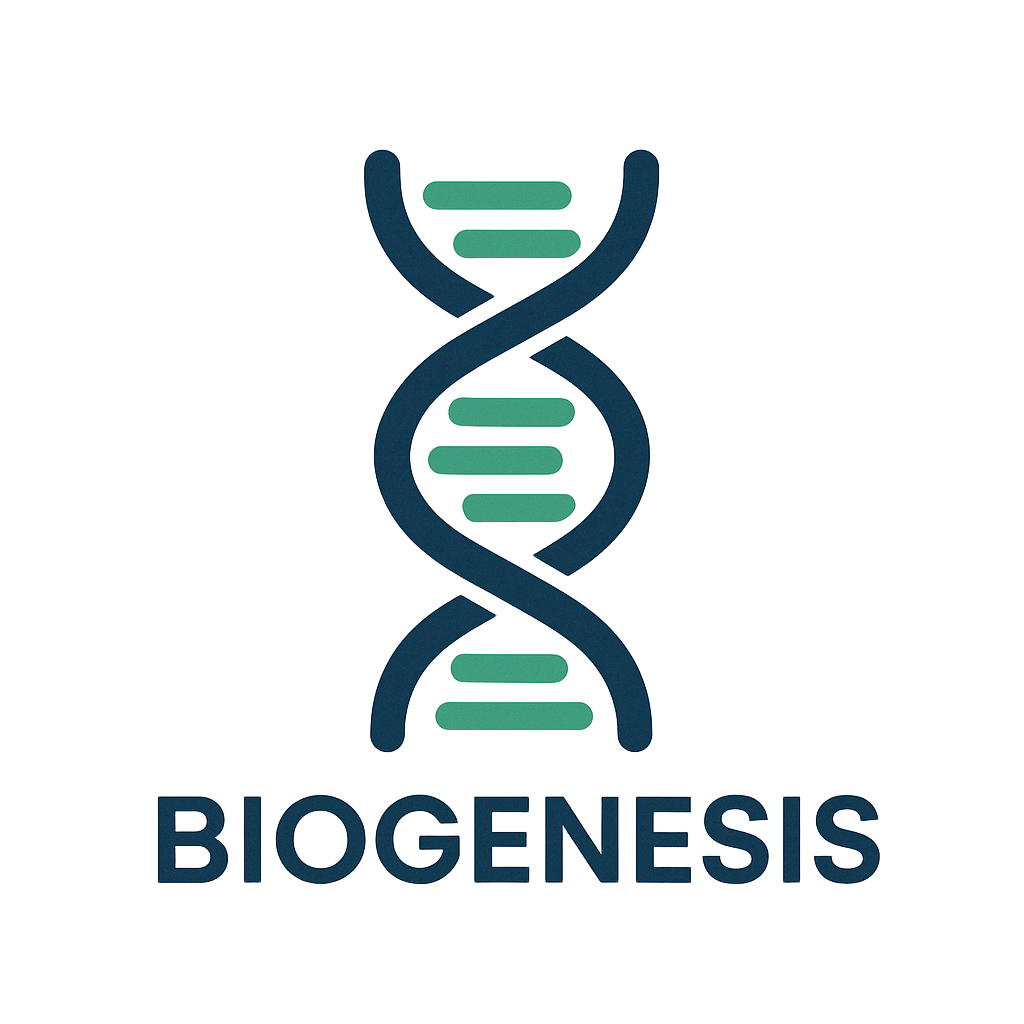
**Nombre del autor:** Portillo, Lautaro

**Email:** [lautaroportillo2115@gmail.com](mailto:lautaroportillo@gmail.com)

**Cohorte:** DAFT-14

**Fecha de entrega:** 09/06/2025

**Institución:** BioGenesis



# Introducción

Inspirada por las lecciones de la pandemia, BioGenesis, empresa líder en biotecnología, busca fortalecer la respuesta sanitaria en Latinoamérica. Su proyecto se enfoca en identificar ubicaciones estratégicas para centros de investigación y distribución en Argentina, Brasil, Chile, Colombia, México y Perú, basándose en un análisis de datos posteriores a enero de 2021 que incluye tanto el impacto del COVID-19 como factores sociales. Para lograrlo, la compañía ha decidido:

* **Mapear las vulnerabilidades** de cada región para priorizar áreas críticas.
* **Analizar la logística y conectividad** de las cadenas de suministro médico.
* **Diseñar herramientas de visualización interactivas** para evaluar y seleccionar las mejores ubicaciones posibles.

Con el enfoque logrado, BioGenesis no solo se podrá expandir, sino contribuir de manera significativa al fortalecimiento de las capacidades locales.

# Desarrollo del proyecto

# AVANCE N°1: Carga, limpieza y transformación de datos

# El presente avance documenta en detalle el proceso metodológico de Carga, Transformación y Limpieza (ETL) que se aplicó sobre el dataset data\_latinoamerica.csv. Este procedimiento fue una fase fundamental y prerrequisito indispensable para garantizar la calidad y la integridad de los análisis subsecuentes. El conjunto de datos original, caracterizado por su gran volumen con más de 12 millones de registros y una alta dimensionalidad de 50 columnas, fue sistemáticamente procesado para facilitar un análisis enfocado, relevante y reciente.

# El procedimiento técnico se estructuró de la siguiente manera:

# 1. Filtrado Estratégico y Delimitación del Alcance: La primera etapa del preprocesamiento consistió en una reducción deliberada del dataset. Se aplicó un filtro geográfico para seleccionar exclusivamente los registros pertenecientes a los países de interés del proyecto: Argentina, Brasil, Chile, Colombia, México y Perú. Adicionalmente, se implementó un filtro temporal para acotar el análisis a todos los datos registrados a partir del 1 de enero de 2021. Esta delimitación fue crucial para alinear la información con los objetivos estratégicos, permitiendo concentrar los esfuerzos en la dinámica más reciente de la pandemia y sus efectos.

# 2. Proceso Exhaustivo de Limpieza de Datos: Con el dataset ya filtrado, se procedió a una fase de limpieza rigurosa para tratar las inconsistencias y los datos faltantes.

# Gestión de Valores Nulos: Se implementaron diversas técnicas de imputación para manejar los valores ausentes. La elección de la estrategia (relleno con el promedio, la mediana, o la propagación de valores anteriores o siguientes) no fue aleatoria, sino que se determinó según el contexto y la naturaleza de cada variable para minimizar el sesgo.

# Eliminación de Datos Irrelevantes: Se realizó un escaneo para identificar y eliminar tanto filas como columnas que estaban completamente vacías, ya que no aportaban información valiosa y aumentaban la complejidad computacional.

# Transformación y Estandarización: Se ejecutó una transformación de tipos de datos en todo el conjunto. Este paso aseguró que cada columna tuviera su formato correspondiente (por ejemplo, convirtiendo textos a formatos numéricos o fechas a un tipo de dato datetime estándar), lo cual es esencial para realizar cálculos y análisis de series temporales de manera correcta.

# 3. Exploración Preliminar de Variables Clave: Una vez limpio el dataset, se realizó un Análisis Exploratorio de Datos (EDA) preliminar. El propósito de esta exploración fue doble: por un lado, validar la calidad y la coherencia de los datos después de la limpieza y, por otro, obtener una comprensión inicial de la distribución y las características de las variables más importantes. Se examinaron variables críticas como:

# Incidencia de COVID-19: Casos confirmados y dosis administradas.

# Datos Demográficos: Cobertura de la población y distribución por grupos de edad.

# Factores de Riesgo: Prevalencia de condiciones como tabaquismo y diabetes.

# Indicadores de Salud General: Tasas de mortalidad y esperanza de vida.

# 4. Consolidación y Guardado de Datos: Finalmente, el conjunto de datos resultante, ya procesado, validado y limpio, fue consolidado y exportado al archivo DatosFinalesFiltrados.csv. Este archivo representa una versión optimizada y cohesiva de los datos originales, constituyendo la base única y confiable para todas las fases subsecuentes del proyecto, incluyendo análisis más profundos, la creación de visualizaciones y el eventual desarrollo de modelos.

# Conclusión: En resumen, esta meticulosa etapa de preparación fue crítica para transformar un dataset masivo y complejo en un recurso manejable, íntegro y directamente alineado con las preguntas de investigación. El proceso no solo redujo la complejidad, sino que aseguró la robustez y fiabilidad de los datos, estableciendo una base sólida para la generación de conocimiento y conclusiones válidas en las siguientes fases del análisis.

# AVANCE N°2: Análisis Exploratorio

Tras la exitosa finalización de la fase de Carga, Transformación y Limpieza (ETL), el proyecto avanzó hacia la etapa de Análisis Exploratorio de Datos (EDA). El objetivo central de esta fase fue trascender los datos limpios para destilar información clave y conocimientos accionables, capaces de respaldar las decisiones estratégicas de expansión de los laboratorios farmacéuticos en los países seleccionados. Para ello, se combinó un análisis estadístico riguroso con un conjunto de visualizaciones de datos diseñadas para revelar patrones, tendencias, anomalías y relaciones subyacentes.

El procedimiento de esta etapa se desarrolló de la siguiente manera:

**1. Análisis Estadístico Descriptivo y Correlacional:** El primer pilar del EDA fue la aplicación de técnicas estadísticas para obtener un resumen cuantitativo del dataset. Se utilizaron medidas de tendencia central (media, mediana) y de dispersión (desviación estándar, varianza, rangos intercuartílicos) para evaluar la distribución y variabilidad de las variables clave. Este análisis incluyó:

* Un examen detallado de la **distribución de casos confirmados y fallecimientos**, identificando los valores promedio y la dispersión de estas cifras en cada país.
* Una evaluación de las **tasas de vacunación**, no solo para comparar los promedios entre países, sino también para entender su variabilidad y consistencia a lo largo del tiempo.
* Un análisis de la **relación entre indicadores demográficos** (como la densidad poblacional o la estructura de edad) y las variables de salud, sentando las bases para entender los factores contextuales de cada nación.

Adicionalmente, se calculó una **matriz de correlación** entre las variables más relevantes. El propósito de este cálculo fue cuantificar la fuerza y la dirección de las relaciones lineales, permitiendo identificar de manera objetiva qué factores estaban significativamente asociados entre sí.

**2. Estrategia de Visualización para la Interpretación de Hallazgos:** Para dar vida a los hallazgos estadísticos, se desarrolló una completa estrategia de visualización. Cada gráfico fue seleccionado con un propósito específico para facilitar la interpretación y la comunicación de resultados complejos:

* **Histogramas y gráficos de densidad:** Se utilizaron para explorar visualmente la forma de la distribución de variables continuas, como los casos confirmados y las tasas de vacunación, permitiendo identificar asimetrías, picos o comportamientos bimodales.
* **Gráficos de barras comparativas:** Fueron la herramienta principal para contrastar métricas agregadas (casos acumulados, muertes totales, tasas de vacunación) entre los diferentes países, ofreciendo una comparación directa y clara de su situación.
* **Mapas de calor (Heatmaps):** Se emplearon para representar gráficamente la matriz de correlaciones. Esta técnica permitió una interpretación rápida e intuitiva de las relaciones más fuertes (tanto positivas como negativas) a través de una escala de colores, destacando visualmente las asociaciones más significativas.
* **Diagramas de dispersión (Scatter plots):** Se crearon para investigar relaciones bivariadas, como la posible influencia de la temperatura media en la incidencia de COVID-19 y los fallecimientos, ayudando a explorar hipótesis sobre patrones estacionales o ambientales.
* **Gráficos de series temporales:** Permitieron analizar la evolución mensual de las dosis administradas, los casos y las muertes por país. Estos gráficos fueron fundamentales para observar tendencias, ciclos y la estacionalidad de la pandemia.

**3. Identificación y Documentación de Tendencias y Patrones:** La combinación de los análisis estadísticos y visuales permitió la detección de varias tendencias y patrones de alto valor estratégico:

* **Patrones Temporales:** Se identificaron y documentaron fluctuaciones significativas en los casos confirmados y en el ritmo de administración de vacunas a lo largo de los diferentes meses, sugiriendo posibles efectos de olas pandémicas, políticas de salud pública o fatiga social.
* **Factores Ambientales:** El análisis de dispersión sugirió una posible correlación entre la temperatura media y la propagación del virus, un hallazgo que, aunque no establece causalidad, aporta una variable adicional para el análisis de riesgo regional.

**4. Personalización Avanzada para la Claridad Comunicativa:** Entendiendo que la efectividad de un análisis reside en su capacidad de ser comprendido, se aplicó una personalización detallada a todas las visualizaciones. Este esfuerzo no fue meramente estético, sino funcional:

* Se asignaron **paletas de colores consistentes y diferenciadas** por país para facilitar el seguimiento a través de múltiples gráficos.
* Se redactaron **etiquetas descriptivas, leyendas claras y títulos informativos** que resumen el hallazgo principal de cada gráfico.
* Se ajustaron los **tamaños, resoluciones y fuentes** para garantizar la máxima legibilidad y profesionalismo en la presentación de los resultados.

**Conclusión General:** Esta fase de Análisis Exploratorio de Datos fue exitosa en transformar el dataset procesado en inteligencia estratégica. El análisis estadístico proveyó la base cuantitativa, mientras que las visualizaciones personalizadas permitieron una comprensión profunda y multidimensional de las dinámicas de la pandemia en los países de interés. Los hallazgos no solo revelaron diferencias significativas entre las naciones, sino que también proporcionaron los insights necesarios para comenzar a priorizar áreas y formular hipótesis estratégicas para la expansión de los laboratorios farmacéuticos, mejorando drásticamente la comunicación y el impacto de los resultados.

# AVANCE N°3: EDA con Numpy y Pandas

En esta tercera etapa del proyecto, la investigación trascendió el análisis exploratorio inicial para adentrarse en un examen más profundo y especializado de los datos, empleando herramientas avanzadas de las librerías Pandas y Numpy. El objetivo principal fue desentrañar patrones temporales complejos, tendencias subyacentes y correlaciones multivariadas que pudieran ofrecer una base empírica sólida para la toma de decisiones estratégicas. Enfocándose en la dinámica de la incidencia de COVID-19, las tasas de vacunación y las características demográficas, esta fase buscó preparar el terreno para la generación de visualizaciones avanzadas y recomendaciones de alta precisión.

El enfoque metodológico se centró en tres pilares: el análisis de series temporales, la investigación de correlaciones y la generación de gráficos avanzados.

**1. Análisis Exploratorio Avanzado y de Series Temporales:** Esta fase se distinguió por la aplicación de técnicas analíticas sofisticadas para estudiar la evolución de las métricas clave a lo largo del tiempo.

* **Identificación de Tendencias a Largo Plazo:** Se examinó la evolución mensual y anual de los casos confirmados, recuperados y fallecimientos para cada país. Esto permitió identificar la trayectoria general de la pandemia en cada región, más allá de las fluctuaciones diarias.
* **Análisis de Estacionalidad y Patrones Cíclicos:** Se aplicaron métodos para evaluar la posible periodicidad en el incremento de casos, investigando si existían patrones estacionales (por ejemplo, picos en ciertos meses del año) que pudieran ser predecibles.
* **Autocorrelación y Descomposición Temporal:** Se utilizaron técnicas avanzadas como el análisis de autocorrelación (ACF) para determinar la dependencia entre los datos pasados y actuales de la serie temporal. Además, se aplicó la descomposición temporal para separar las series de datos en sus tres componentes fundamentales: tendencia (el comportamiento a largo plazo), estacionalidad (patrones periódicos) y residuo (ruido aleatorio), logrando una comprensión mucho más profunda de su estructura.

**2. Visualización de Hallazgos y Dinámicas Clave:** Los resultados del análisis fueron representados a través de visualizaciones específicas, diseñadas para comunicar los hallazgos más destacados de manera clara y efectiva:

* **Evolución de Casos Activos vs. Recuperados:** Un análisis gráfico comparativo mostró la dinámica de recuperación frente a los casos activos. Se detectó una influencia particularmente notable en el periodo comprendido entre enero y marzo de 2022, una fase crítica para entender la capacidad de respuesta del sistema de salud.
* **Relación entre Cobertura de Vacunación y Casos:** Se observó que Brasil presentaba un caso de estudio particular, manteniendo una alta incidencia de casos a pesar de alcanzar un considerable porcentaje de población vacunada, lo que sugiere la interacción de otros factores como la aparición de nuevas variantes o la heterogeneidad en la distribución de vacunas.
* **Correlación entre Nuevos Casos y Temperatura Promedio:** La visualización de esta relación sugirió una correlación positiva, donde a mayor temperatura se observaba una mayor cantidad de casos en ciertas regiones, un hallazgo que desafía algunas hipótesis iniciales sobre la estacionalidad del virus.
* **Evolución Semanal de Nuevos Casos:** El análisis de series temporales a nivel semanal permitió identificar con gran precisión el pico de contagios más alto del periodo estudiado, localizado entre las últimas semanas de enero y las primeras de febrero de 2022.

**3. Investigación Profunda de Correlaciones Multivariadas:** Se condujo un estudio exhaustivo para cuantificar las relaciones entre un conjunto diverso de variables, con el fin de validar hipótesis y descubrir nuevos factores de influencia:

* Se investigó la **relación entre las tasas de vacunación y la disminución de la mortalidad** para confirmar y cuantificar el impacto positivo de las campañas de inmunización.
* Se analizó el **impacto de la temperatura media** en la propagación del virus, buscando patrones consistentes a través de los diferentes países.
* Se examinó la **influencia de la densidad poblacional y el grado de urbanización** en la incidencia de casos, confirmando el rol de los centros urbanos como focos de transmisión.
* Se exploraron las **correlaciones entre variables demográficas** (como la edad promedio de la población o la prevalencia de comorbilidades) y las tasas de mortalidad, para perfilar mejor los grupos de riesgo.

**Conclusión General:** Esta tercera etapa de análisis avanzado proporcionó una comprensión más rica y matizada del panorama sanitario. Las **tendencias identificadas** a través del análisis temporal revelaron patrones claros de estacionalidad y variabilidad, destacando que países como Brasil y México enfrentaron desafíos particulares y persistentes en la contención del virus. A su vez, las **relaciones significativas** validadas por el análisis de correlación proveyeron evidencia cuantitativa sobre el impacto positivo de la vacunación y la innegable influencia de factores demográficos y ambientales. En conjunto, estos hallazgos constituyen una base de conocimiento robusta, esencial para afinar las estrategias de expansión y asignación de recursos en la región.

# AVANCE N°4: Aplicaciones Prácticas – Integración en Power BI

En esta etapa final y de capitalización del proyecto, se procedió a la integración de todo el análisis realizado en las fases previas dentro de una plataforma visual e interactiva, utilizando Microsoft Power BI como herramienta central de inteligencia de negocios. Este proceso fue diseñado para traducir los complejos hallazgos analíticos en dashboards ejecutivos e intuitivos, permitiendo a los directivos y tomadores de decisiones explorar información clave de manera autónoma y dinámica. El objetivo final es facilitar la priorización de áreas geográficas para la expansión de laboratorios y la optimización de centros de vacunación, basándose en evidencia sólida extraída de los indicadores de incidencia de COVID-19 y la cobertura de inmunización.

La metodología de esta fase se centró en la conexión de los datos preparados, el diseño de los dashboards y una reflexión sobre el valor comparativo de las diferentes estrategias de visualización.

**1. Integración de Datos y Conexión de Ecosistemas (Python y Power BI):** El primer paso crítico fue establecer un puente de datos robusto entre el entorno analítico (donde se generó el dataset procesado) y la plataforma de Power BI. Esto se logró mediante la importación directa del archivo DatosFinalesFiltrados.csv. Esta integración fue fundamental, ya que aseguró que todas las visualizaciones y análisis subsecuentes se construyeran sobre la base de datos ya validada, limpia y enriquecida en las etapas anteriores, garantizando así la coherencia y la integridad de la información desde el análisis inicial hasta el reporte final. Este enfoque permite combinar la potencia del análisis técnico avanzado de Python con la capacidad de visualización interactiva de Power BI.

**2. Diseño y Construcción de Dashboards Interactivos para el Apoyo Estratégico:** Se procedió al diseño y construcción de un dashboard central en Power BI, concebido para sintetizar los hallazgos más relevantes y ofrecer una experiencia de usuario exploratoria. Los componentes clave del dashboard incluyen:

* **Indicadores Clave de Rendimiento (KPIs):** Se dispusieron en un lugar prominente para ofrecer una visión instantánea del estado de situación, mostrando métricas de alto nivel como las tasas de vacunación completas por país, la incidencia acumulada de casos por cada 100,000 habitantes y la mortalidad general asociada al COVID-19.
* **Mapa Interactivo (Cartografía Georreferenciada):** Se implementó un mapa dinámico que visualiza la distribución geográfica de la incidencia y las tasas de vacunación. Esta herramienta permite a los usuarios identificar visualmente clústeres, puntos críticos (hotspots) y áreas con oportunidades de mejora en la cobertura de vacunación.
* **Gráficos Interactivos Multidimensionales:** Se incluyó una batería de gráficos que permiten un análisis profundo:
  + **Series Temporales:** Para analizar la evolución de casos y vacunación, permitiendo identificar estacionalidad e impacto de intervenciones.
  + **Gráficos de Dispersión:** Para explorar visualmente la relación entre la cobertura de vacunación y la reducción de casos o mortalidad.
  + **Histogramas y Boxplots:** Para comprender la distribución y variabilidad de la incidencia entre países y su correlación con factores demográficos.
  + **Mapas de Calor:** Para representar de forma intuitiva las correlaciones entre variables, facilitando la identificación de relaciones complejas.
* **Filtros Dinámicos:** Se integraron filtros que otorgan al usuario el control total para segmentar la información por país, rangos de tiempo específicos (año, mes, semana) y métricas particulares, transformando el dashboard en una herramienta de análisis ad-hoc.
* **Tablas Resumidas:** Para complementar las visualizaciones, se añadieron tablas que presentan los datos cuantitativos precisos, como porcentajes exactos de población vacunada y el recuento de casos activos.

La sinergia de estos elementos convierte el dashboard en un verdadero entorno de exploración analítica, donde los directivos pueden navegar desde una vista macro a una micro con total flexibilidad.

**3. Análisis Comparativo de Estrategias de Visualización: Estática vs. Interactiva:** Como parte de la metodología, se realizó una reflexión sobre la idoneidad de cada tipo de visualización:

* **Visualizaciones Estáticas:** Se reconoce su valor por su claridad y simplicidad, siendo ideales para comunicaciones unidireccionales como reportes impresos en formato PDF, artículos o presentaciones formales donde se requiere controlar la narrativa. Su principal limitación es la falta de profundidad, ya que presentan una única perspectiva de los datos.
* **Visualizaciones Interactivas:** Se destacan como la opción superior para contextos de toma de decisiones dinámicas. Sus ventajas radican en que permiten explorar múltiples capas de información, son adaptables a las preguntas específicas de cada usuario y fomentan un entendimiento más profundo al facilitar la desagregación y el filtrado de datos en tiempo real. Su única limitación es la necesidad de acceso a la plataforma digital.

Se concluye que la capacidad de Power BI para, además, ejecutar scripts de Python directamente, representa la sinergia perfecta, permitiendo integrar análisis predictivos o estadísticos más sofisticados dentro de los dashboards y ofreciendo así un potencial analítico sin precedentes.

**Conclusión General:** La implementación en Power BI ha demostrado ser una etapa crucial que materializa el valor de todo el análisis previo. La plataforma no solo facilita la visualización de grandes volúmenes de datos, sino que transforma la manera en que los líderes interactúan con la información. Los dashboards dinámicos se erigen como la herramienta clave del proyecto, destacando patrones y áreas prioritarias de forma intuitiva y maximizando el impacto del análisis. Si bien las visualizaciones estáticas conservan su utilidad para reportes tradicionales, es la interactividad la que proporciona un nivel superior de detalle, adaptabilidad y, en última instancia, de inteligencia estratégica para la organización.

# 

# 

# 

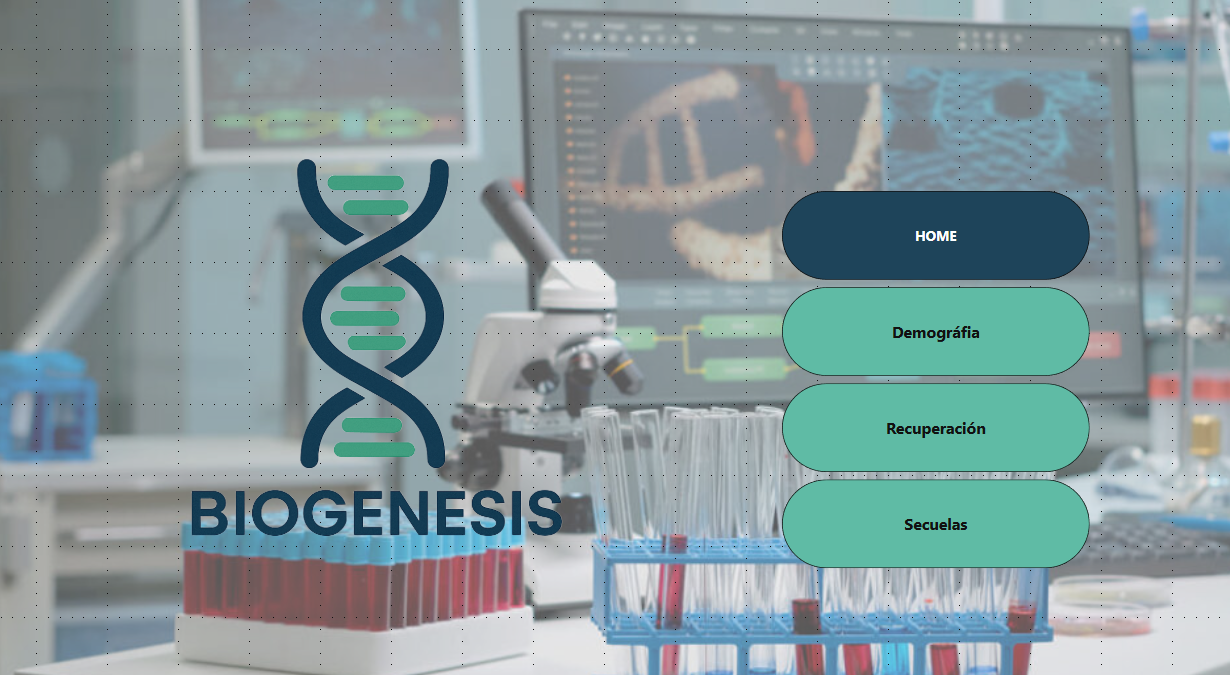
# 

# Análisis del Dashboard

**Home:**

La primera página del dashboard de BioGénesis cumple la doble función de portada y menú de navegación principal. Desde allí, los usuarios pueden acceder a los tres paneles informativos del reporte: **Demografía**, **Recuperación** y **Secuelas**.

Se incorporó un diseño profesional y temático con un menú de botones claramente identificados. Estos elementos facilitan la interacción y dirigen al usuario de forma intuitiva a las diferentes secciones de análisis.

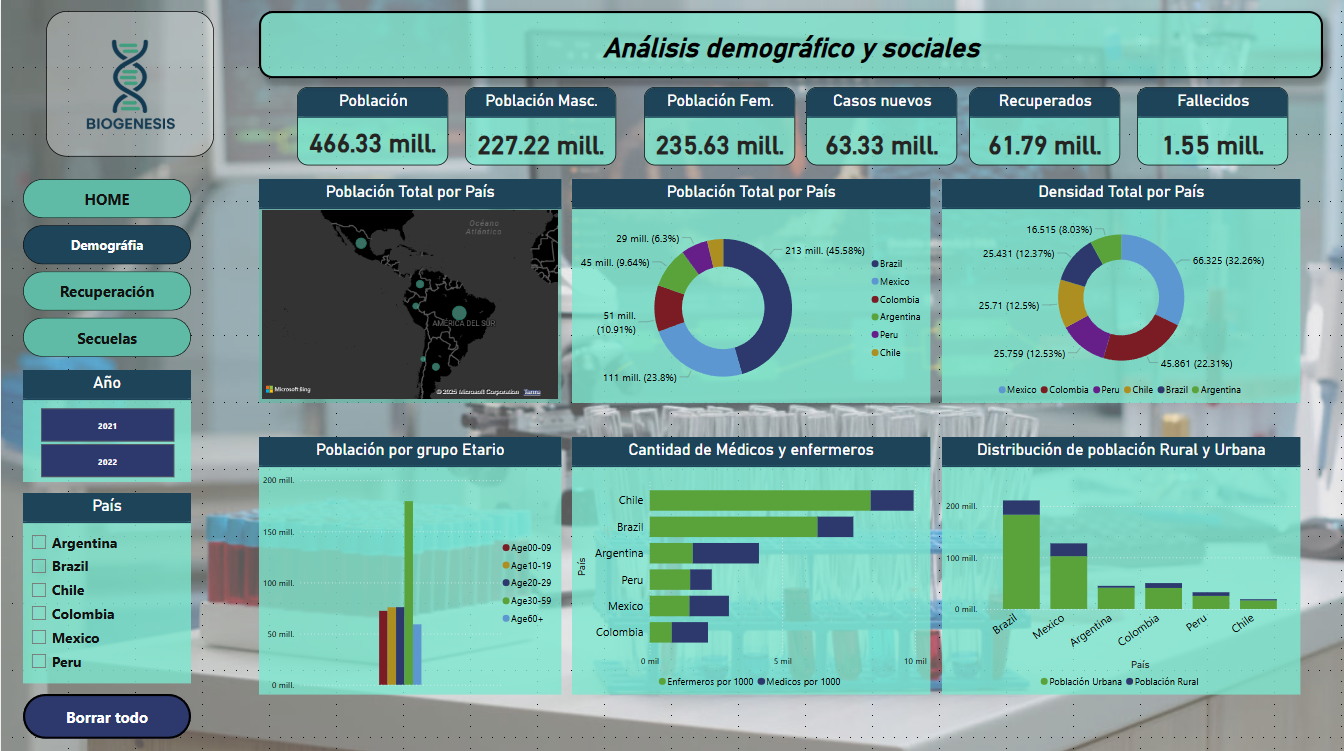


**Demografía:**

Este segundo panel se centra en el análisis demográfico y social, ofreciendo una visión comparativa de los países seleccionados. Entre sus visualizaciones principales se incluyen:

* Una fila de tarjetas (KPIs) que resumen los indicadores clave como población total, casos nuevos, recuperados y fallecidos.
* Dos gráficos de anillo para observar la distribución de la población total y la densidad poblacional por país, acompañados de un mapa para referencia geográfica.
* Un gráfico de barras que explora la estructura de la población por grupos de edad.
* Otro gráfico de barras que compara la cantidad de médicos y enfermeros, sirviendo como indicador de la capacidad sanitaria de cada nación.
* Un tercer gráfico de barras para visualizar la distribución de la población rural y urbana.

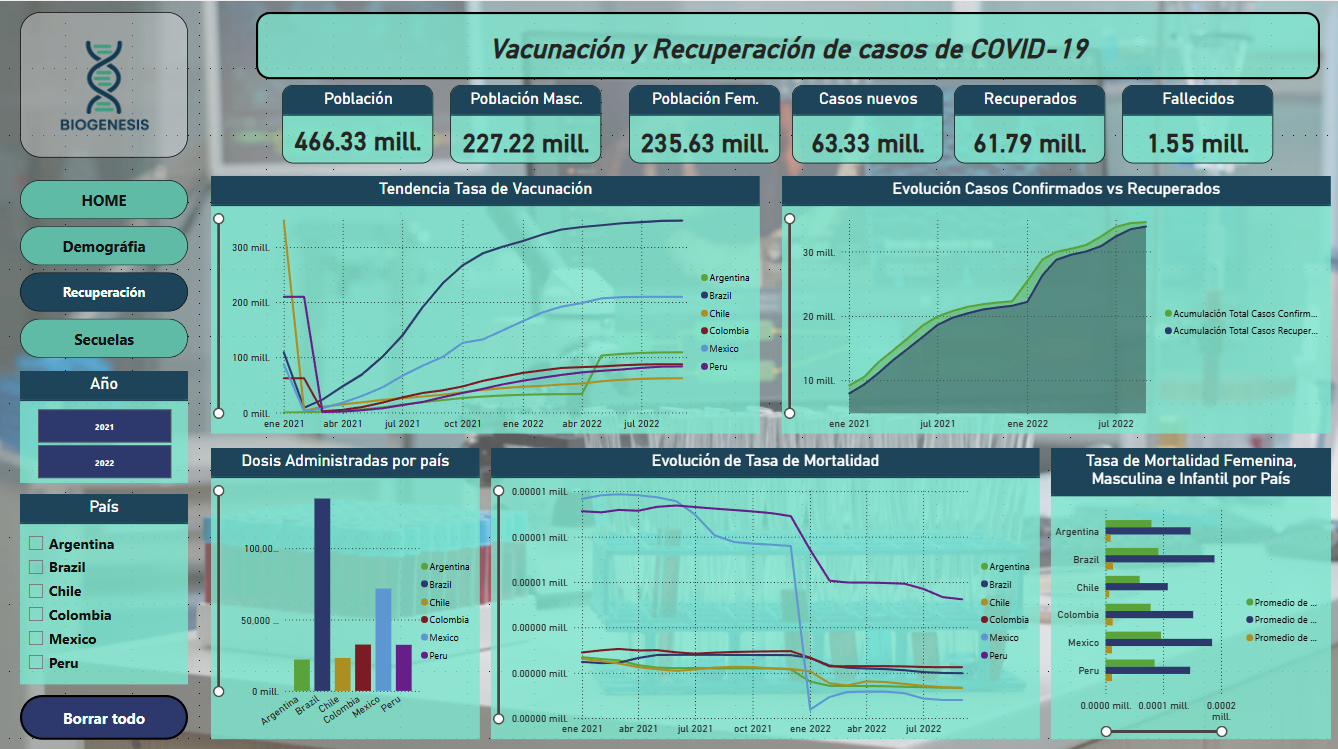
El panel también incorpora segmentadores por país y por año, lo que permite al usuario filtrar los datos y adaptar el análisis a distintas regiones y periodos temporales.

****

**Recuperación:**

Este tercer panel se enfoca en la **Vacunación y Recuperación de casos de COVID-19**, ofreciendo un análisis sobre la efectividad de la respuesta sanitaria en los países estudiados.

Entre sus visualizaciones principales se incluyen:

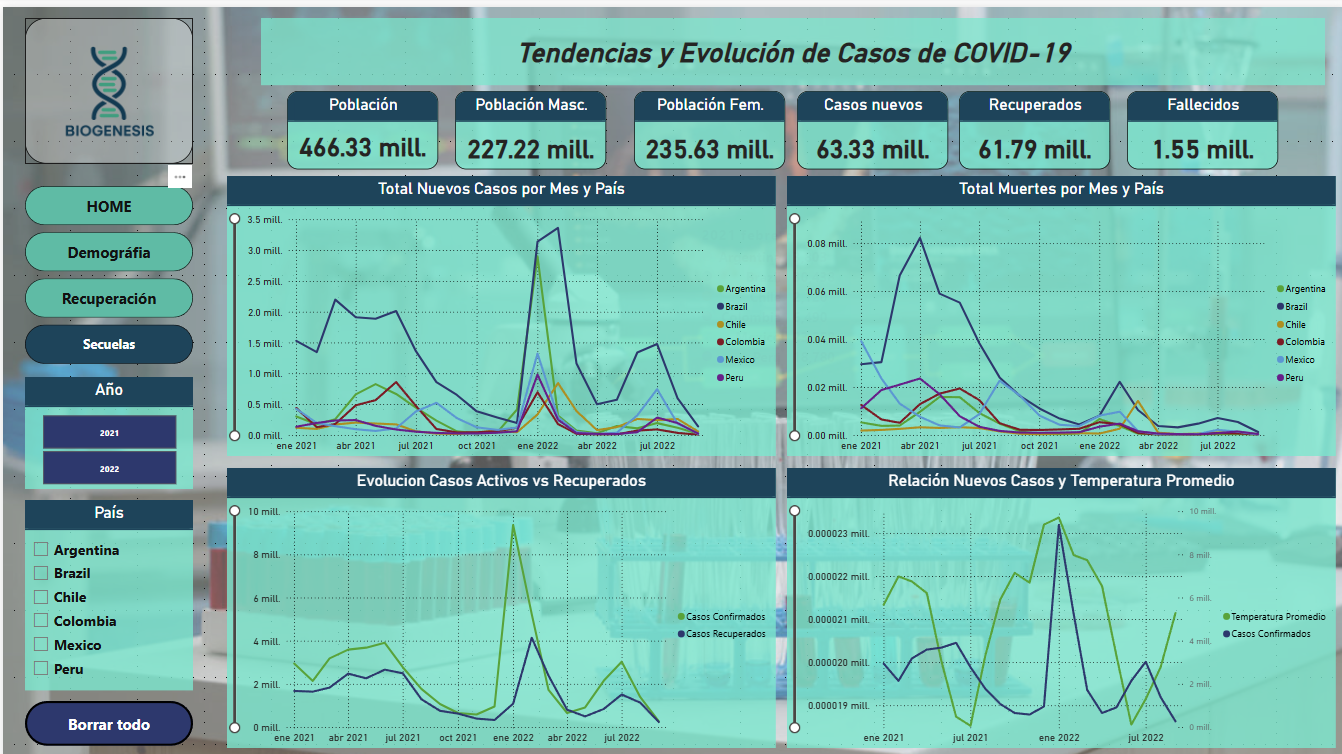
* Una fila de **tarjetas (KPIs)** en la parte superior que presentan los indicadores generales de población y de la pandemia.
* Un **gráfico de líneas** para observar la evolución de la **tasa de vacunación** a lo largo del tiempo en cada país.
* Otro gráfico de líneas (de área) que muestra la progresión de los **casos confirmados frente a los recuperados**.
* Un **gráfico de barras** que permite comparar el número total de **dosis de vacunas administradas** por cada nación.
* Un gráfico de líneas que explora la **evolución de la tasa de mortalidad** durante el periodo analizado.
* Finalmente, un gráfico de barras que desglosa y compara la **tasa de mortalidad por género (femenina, masculina) e infantil**.

**Secuelas:**

Este último panel, correspondiente a la sección **"Secuelas"**, se centra en las **Tendencias y la Evolución de casos de COVID-19** para permitir un análisis profundo de los patrones temporales y sus posibles causas.

Entre sus visualizaciones principales se incluyen:

* Una fila superior de **tarjetas (KPIs)** con los indicadores demográficos y epidemiológicos generales para dar contexto.
* Un **gráfico de líneas** que muestra la evolución mensual del **total de nuevos casos** por país, facilitando la identificación de los picos de contagio.
* Otro **gráfico de líneas** para observar el comportamiento mensual del **total de muertes** a lo largo del tiempo.
* Un **gráfico de área** que visualiza la evolución de los **casos activos frente a los recuperados**.
* Finalmente, un **gráfico de líneas combinado** que explora la **relación entre los nuevos casos y la temperatura promedio**, uno de los hallazgos clave del estudio.



### Conclusión

Tras una evaluación exhaustiva y multidimensional de los datos presentados en los paneles de control, se concluye que existe una clara y jerarquizada oportunidad para la expansión del laboratorio "Biogenesis" en la región latinoamericana. El análisis integrado de las variables demográficas, la infraestructura sanitaria, la evolución epidemiológica del COVID-19 y el progreso de las campañas de vacunación previas, converge en una recomendación estratégica inequívoca. La decisión de entrada no debe basarse únicamente en el tamaño de la población, sino en una matriz que pondere la magnitud del mercado potencial, la urgencia sanitaria demostrada y la viabilidad logística para la distribución.

En función de este análisis, **se recomienda enfocar los esfuerzos de expansión de manera prioritaria y secuencial, comenzando por el mercado de México, seguido inmediatamente por Brasil.**

**México se erige como la oportunidad estratégica de mayor prioridad.** Esta conclusión se fundamenta en la confluencia de varios factores críticos. En primer lugar, con más de 110 millones de habitantes y una de las tasas de vacunación más rezagadas del grupo analizado, México representa un mercado latente de enorme magnitud. Esta brecha en la inmunización constituye una oportunidad comercial directa y sustancial. En segundo lugar, los datos epidemiológicos revelan que el país ha sufrido picos de contagio y mortalidad extremadamente severos, lo que subraya una vulnerabilidad sistémica y genera una necesidad imperiosa de soluciones preventivas eficaces en la percepción pública y gubernamental. Finalmente, la alta concentración de su población en núcleos urbanos presenta una ventaja logística decisiva, permitiendo una implementación de campañas de vacunación más rápida y costo-efectiva. La aparente menor densidad de personal sanitario, lejos de ser un impedimento, puede ser vista como una oportunidad para que "Biogenesis" se posicione no solo como un proveedor de vacunas, sino como un socio estratégico en la administración y logística de la inmunización.

**Brasil, por su parte, representa la segunda prioridad estratégica, caracterizada por su escala inigualable.** Siendo el mercado más grande de la región con más de 210 millones de personas, el volumen potencial es máximo. Aunque su tasa de vacunación es superior a la de México, el tamaño de su población implica que aún existen millones de individuos por alcanzar. La justificación para la entrada en Brasil se ve reforzada por su historial de olas zpandémicas devastadoras, que han demostrado que incluso un sistema sanitario relativamente más robusto puede verse superado. Esto crea una demanda sostenida de vacunas y, crucialmente, de dosis de refuerzo para mantener a raya futuras crisis. La considerable población de adultos mayores (Age60+) solidifica su perfil como un mercado donde la necesidad de proteger a los más vulnerables es una prioridad de salud pública constante.

Los demás países analizados, si bien presentan ciertas oportunidades, ocupan un lugar secundario en la estrategia de expansión. **Colombia y Perú se perfilan como mercados de oportunidad secundaria.** Ambos presentan brechas de vacunación que podrían ser aprovechadas, pero su menor tamaño de mercado y los desafíos logísticos asociados a una población rural más dispersa los hacen menos atractivos para una inversión inicial a gran escala. Por último, **Chile debe considerarse un mercado maduro y de baja prioridad para una campaña de vacunación primaria.** Su éxito en alcanzar altas tasas de inmunización limita la oportunidad inmediata, relegando el potencial de este mercado a futuras campañas de refuerzo o a la introducción de vacunas de nueva generación.

### Optimización y sostenibilidad

Se implementó un proceso de optimización iterativa durante la construcción de la base de datos. Las tablas fueron estructuradas siguiendo principios de normalización para minimizar la redundancia y garantizar relaciones coherentes, lo que resultó en un acceso más rápido y una interpretación más clara de la información.

### Desafíos y soluciones

El desarrollo implicó superar desafíos comunes en Power BI y DAX. Esto incluyó el dominio de funciones clave como CALCULATE para la manipulación del contexto, la correcta aplicación de iteradores como SUMX para cálculos a nivel de fila, y la implementación de prácticas de programación defensiva, como el uso de DIVIDE para prevenir errores y la optimización de medidas para asegurar el rendimiento con grandes volúmes de datos.

### Reflexión personal

El proyecto fue una oportunidad clave para consolidar competencias en análisis de datos y adquirir habilidades prácticas en la gestión de bases de datos. Una reflexión final destaca la importancia crítica de la fase de planificación y la definición clara de objetivos. Para futuros proyectos, se identifica la oportunidad de incorporar herramientas y técnicas analíticas más avanzadas para enriquecer aún más los resultados.