

## Expansión Estratégica de “Bio Genesys”

**Nombre del autor:** Beese, Ariel

**Email:** ariel2001rubio@gmail.com

**Cohorte:** DAFT-12

**Fecha de entrega:** 06/04/2025

**Institución:** BioGenesys



### Introducción

BioGenesys, empresa líder en biotecnología, se ha propuesto a fortalecer la preparación y capacidad de respuesta sanitaria en Latinoamérica frente a futuras crisis de salud pública. Con base en los aprendizajes derivados de la pandemia del COVID-19, este proyecto tiene como meta identificar regiones estratégicas para el desarrollo de centros de investigación y distribución de recursos médicos.

El enfoque de la iniciativa se centra en analizar no solo en la incidencia histórica del COVID-19, sino también factores sociales y recursos que permiten anticipar necesidades y optimizar el acceso a tratamientos, vacunas y tecnología médica. Esto le permite a BioGenesys liderar estrategias preventivas y fomentar sistemas de salud más equitativos y resilientes en países como Argentina, Brasil, Chile, Colombia, México y Perú, siendo estos los países en los que se hizo el análisis después del 1 de enero del 2021.

Se decidió:

1. Mapear vulnerabilidades regionales con el fin de identificar áreas prioritarias.
2. Analizar la conectividad y la logística de las cadenas de suministro médico.
3. Implementar estrategias de visualización interactiva, diseñando herramientas que permiten evaluar posibles ubicaciones estratégicas.

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

## Módulo 4

### Desarrollo del proyecto

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

En este avance se centró la carga, transformación y limpieza del dataset extenso proporcionado. Se hace la principal limpieza haciendo los filtros por país y fechas del 1 de enero del 2021 en adelante.

Se analizó la estructura y el contenido, con el fin de identificar los datos relevantes. Valores nulos fueron gestionados según contexto de columnas.

##### 1. Conociendo el Dataset

El archivo contaba con más de 12 millones de registros en los que había 50 columnas dentro de un csv llamado "data\_latinoamerica.csv" y se cotaba con un archivo llamado "readme.txt" el cual contenía una breve descripción del contenido de cada columna.

##### 2. Filtros

Se seleccionaron los siguientes países para el análisis:

Argentina

Brasil

Chile

Colombia

México

Perú

Y se otro filtro en donde estarían los datos registrados a partir del 1 de enero de 2021, permitiendo que se realice un análisis con un enfoque más reciente.

##### 3. Limpieza

Se realizó una limpieza de valores nulos, aplicando estrategias como el uso de valores promedio, medio, datos anteriores o siguientes, según corresponda.

Se eliminaron columnas y filas en las que existieran valores totalmente vacíos.

Se hace una transformación de datos para poder tener todos los valores con su respectivo tipo de dato.

##### 4. Variables claves

Se realizó una exploración inicial para comprender la distribución y características de las variables principales, como:

Incidencia de COVID-19: cuenta con los casos confirmados y las dosis administradas por país y periodo.

Población: donde contamos con la cobertura de la población -

Grupos de edad: aquí tenemos los diferentes rangos de edades afectados.

Grupos enfermos: donde tenemos a los fumadores y diabéticos.

Expectativas: donde tenemos variables que nos muestra tasas e índices de vida.

## Módulo 4

### 5. Guardado de los datos

Una vez finalizado toda la limpieza se procede a guardar los datos en un archivo llamado "DatosFinalesFiltrados.csv", el cual será utilizado para realizar los distintos análisis y poder comprender el contexto.

Conclusión:

La limpieza de datos fue fundamental para reducir la complejidad y mejorar el manejo del dataset. El filtrado ayudo a que el análisis se alinee con los objetivos estratégicos. Y se concreto un buen archivo para un análisis exploratorio y avanzado.

## AVANCE N°2: Análisis Exploratorio

En esta etapa, se hace un enfoque en el análisis exploratorio de los datos, para obtener información clave que respalde las decisiones estratégicas de expansión de los laboratorios farmacéuticos en los países seleccionados. Se utilizaron gráficos para identificar patrones, tendencias y anomalías.

Se utilizaron propiedades estadísticas para comprender el dataset. Se crearon visualizaciones que permitan identificar patrones y relaciones.

### 1. Análisis Estadístico

Se utilizaron medidas de tendencia central y de dispersión para evaluar la distribución de las variables claves. Como:

Distribución de los casos confirmados y fallecimientos por país.

Tasas de vacunación y su variabilidad entre países.

Relación entre indicadores demográficos y variables de salud.

Además, se calculó la correlación entre variables para identificar relaciones significativas.

### 2. Visualización de Hallazgos Clave

Las visualizaciones se realizaron para representar gráficamente las observaciones obtenidas durante el análisis estadístico. Entre las más destacadas se encuentran:

Histogramas y gráficos de densidad para explorar la distribución de los casos confirmados y las tasas de vacunación en los países seleccionados.

Gráficos de barras comparando casos confirmados, muertes acumuladas y tasas de vacunación entre países.

Mapas de calor para representar gráficamente las correlaciones entre variables, destacando relaciones significativas que podrían influir en las estrategias de expansión.

Diagramas de dispersión analizando la relación entre la temperatura media y la incidencia del COVID-19, así como entre la temperatura y los fallecimientos, lo que permite explorar posibles patrones estacionales.

Evolución temporal se analizaron las dosis de vacunas administradas, casos confirmados y muertes reportadas a lo largo del tiempo, observando variaciones mensuales por país.

### 3. Identificación de Tendencias y Patrones

El análisis exploratorio permitió detectar varias tendencias y patrones importantes:

Patrones Temporales se observaron fluctuaciones en los casos confirmados y la administración de vacunas en diferentes meses.

## Módulo 4

Factores Ambientales se identificaron posibles influencias de la temperatura media sobre la propagación del COVID -19.

### 4. Personalización de Visualizaciones

Para mejorar la presentación de los hallazgos, las visualizaciones fueron personalizadas con:

Paletas de colores diferenciadas por país.

Etiquetas descriptivas y leyendas claras para facilitar la interpretación.

Títulos informativos que destacan el propósito de cada gráfico.

Tamaños ajustados para mejorar la legibilidad.

### Conclusión:

El análisis estadístico y visual permitió identificar patrones clave en la incidencia del COVID19 y las tasas de vacunación en los países.

Las visualizaciones destacan diferencias significativas entre los países, ayudando a priorizar áreas para la expansión de los laboratorios farmacéuticos.

La personalización de los gráficos mejoró la claridad y la comunicación de los resultados.

## AVANCE N°3: EDA con Numpy y Pandas

En esta tercera etapa del proyecto, profundizamos en el análisis exploratorio de datos (EDA) utilizando herramientas avanzadas de Pandas y Numpy. Este análisis tiene como objetivo identificar patrones temporales, tendencias y correlaciones que respalden la toma de decisiones para la expansión de laboratorios farmacéuticos en América Latina. Enfocándonos en los datos de incidencia de COVID -19, tasas de vacunación y características demográficas, buscamos preparar el terreno para visualizaciones avanzadas y recomendaciones estratégicas basadas en datos sólidos.

Explorar series temporales analizar la evolución de casos, muertes y tasas de vacunación. Investigar correlaciones examinar las relaciones entre variables clave como urbanización, temperatura, incidencia de casos y vacunación. Generar gráficos avanzados representar visualmente los hallazgos clave a través de gráficos dinámicos y descriptivos.

### 1. Análisis Exploratorio

Se analizó la evolución de elementos clave del conjunto de datos, se aplicaron técnicas avanzadas como:

Identificación de tendencias para examinar la evolución mensual y anual de casos confirmados, recuperados y fallecimiento para cada país.

Estacionalidad y patrones para evaluar la periodicidad en el incremento de casos.

Análisis de auto correlación y descomposición temporal determinando relaciones entre datos pasados y actuales.

### 2. Visualización de Hallazgos

## Módulo 4

Se crearon gráficos para representar los resultados más destacados del análisis temporal y exploratorio:

Evolución de Casos Activos vs Recuperados:

Se observó el progreso en la recuperación de casos, mostrando una mayor influencia entre enero de 2022 y marzo de 2022.

Relación entre la Cobertura de Vacunación y la Reducción de Casos:

Se observa que Brasil fue el principal país que tuvo una alta cantidad de casos con un buen porcentaje de cubiertos por la vacuna.

Nuevos casos y temperatura promedio:

Se observa que a mayor temperatura mayor cantidad de casos.

Evolución Semanal de Casos Nuevos:

Se observa que entre enero y febrero del 2022 fue el pico de contagios más alto.

### 3. Investigación de Correlaciones

Se estudiaron correlaciones significativas entre las variables del dataset, destacando:

Relación entre tasas de vacunación y disminución de la mortalidad.

Impacto de la temperatura media en la propagación del virus.

Influencia de la densidad poblacional y urbanización en la incidencia de casos.

Correlaciones entre variables demográficas y tasas de mortalidad .

Conclusión:

Tendencias Identificadas , se observaron patrones claros de estacionalidad y variabilidad en los datos, con países como Brasil y México mostrando mayores desafíos en la contención del virus.

Relaciones Significativas en las correlaciones destacan el impacto positivo de la vacunación y la influencia de factores demográficos y ambientales en la incidencia del COVID19.

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

En esta etapa final del proyecto, integramos el análisis realizado en fases previas en una plataforma visual interactiva utilizando Power BI. Este proceso transforma los datos analíticos en dashboards que permiten a los directivos explorar información clave para la toma de decisiones estratégicas. El objetivo es priorizar áreas de expansión para laboratorios y centros de vacunación basados en indicadores de incidencia de COVID-19 y cobertura de vacunación.

Se importan datos preparados a Power BI utilizando los datos limpios y filtrados del archivo "DatosFinalesFiltrados.csv". Creamos el dashboard interactivo para diseñar visualizaciones que faciliten la comprensión de los hallazgos analíticos y permitan explorar los datos de forma dinámica. Se comparan visualizaciones estáticas e interactivas con el fin de destacar cómo cada enfoque aporta valor en diferentes contextos de comunicación y toma de decisiones.

## Módulo 4

### 1. Conexión de Python con Power BI

Se conectó el dataset preparado con Power BI, incluyendo:

Cargar el archivo "DatosFinalesFiltrados.csv" directamente a Power BI.

Esta integración permitió utilizar la capacidad de Power BI para combinar análisis técnico avanzado con visualización interactiva.

### 2. Creación de Dashboards Interactivos

Se diseñaron dashboards en Power BI que sintetizan los hallazgos clave del análisis. Los elementos más relevantes incluyen:

Indicadores Clave de Rendimiento (KPIs):

Tasas de vacunación por país.

Incidencia acumulada de casos por cada 100,000 habitantes.

Mortalidad relacionada con el COVID -19.

Mapa interactivo:

Muestra la distribución geográfica de la incidencia de COVID -19 y las tasas de vacunación en los países analizados.

Gráficos Interactivos:

Series temporales: Evolución de casos y vacunación por país.

Gráficos de dispersión: Relación entre cobertura de vacunación y reducción de casos.

Histogramas y boxplots: Variabilidad en la incidencia entre países y su relación con factores demográficos.

Mapas de calor: Identificación de correlaciones entre variables clave.

Filtros dinámicos:

Por país, período de tiempo y métricas específicas (por ejemplo, incidencia semanal o mensual).

Tablas resumidas:

Resúmenes de datos clave para cada país, incluyendo porcentajes de población vacunada y casos activos.

Estos elementos interactivos proporcionan a los directivos la flexibilidad para explorar áreas específicas de interés y tomar decisiones fundamentadas.

### 3. Visualizaciones Estáticas vs. Interactivas

Visualizaciones Estáticas:

Ventajas:

Claridad y simplicidad para reportes impresos.

Ideales para presentaciones en contextos formales.

Limitaciones:

Falta de dinamismo y capacidad de profundización.

Visualizaciones Interactivas:

## Módulo 4

### Ventajas:

Permiten explorar múltiples capas de información.

Adaptables a diferentes audiencias y contextos.

Favorecen una comprensión más profunda a través de filtros y desagregaciones.

### Limitaciones:

Requieren acceso a plataformas digitales y conocimientos básicos de navegación en Power BI.

Power BI también ofrece la posibilidad de ejecutar scripts de Python, lo que permite integrar análisis más sofisticados directamente en los dashboards, aprovechando la sinergia entre ambas herramientas.

### Conclusión:

Utilidad de Power BI: La plataforma facilita la visualización interactiva de grandes volúmenes de datos, permitiendo una exploración intuitiva por parte de los tomadores de decisiones.

Dashboards como herramienta clave: Las visualizaciones dinámicas destacan patrones y áreas prioritarias, maximizando el impacto del análisis previo.

Estática vs. Interactiva: Si bien las visualizaciones estáticas son útiles para reportes tradicionales, las interactivas ofrecen un nivel de detalle y adaptabilidad superior.

## Optimización y sostenibilidad

La base de datos fue optimizando paulatinamente a medida que se construyó, se estructuraron las tablas de forma tal para minimizar las redundancias y se puedan relacionar todos los datos de manera acorde. De este modo también se optimizó el acceso y la interpretación de la información.

## Desafíos y soluciones

Enfrenté desafíos comunes al trabajar con Power BI, como ser: Errores en DAX: No manejar correctamente las relaciones entre tablas, no usar CALCULATE cuando se necesita un cambio de contexto, usar SUM en lugar de SUMX en cálculos fila por fila, no manejar divisiones por cero con DIVIDE, no optimizar medidas o variables para grandes volúmenes de datos.

## Reflexión personal

Durante este proyecto, he tenido la oportunidad de consolidar mis conocimientos básicos en análisis de datos y adquirir habilidades prácticas que me han permitido entender mejor el manejo de bases de datos.

Si tuviera que volver a empezar este proyecto, probablemente enfocaría más tiempo en la planificación y en la comprensión inicial de los datos. Me aseguraría de definir claramente los objetivos del análisis desde el principio y de explorar más herramientas y técnicas avanzadas para el análisis de datos.