

"EXPANSIÓN ESTRATÉGICA DE BIOGENESYS"



Nombre del autor: Bárbara Batlle Casas

Email: barbarabatllecasas@gmail.com

Cohorte: DA PT-03

Fecha de entrega: 03/11/2024

Institución: La empresa farmacéutica "BIOGENESYS" me contrató como Analista de Datos para estudiar y analizar la posible expansión de sus laboratorios en Latinoamérica (específicamente en Colombia, Argentina, Chile, México, Perú y Brasil), teniendo en cuenta datos relacionados al Covid-19, tasas de vacunación e infraestructura disponible en cada lugar.

Introducción

La compañía farmacéutica "BIOGENESYS" necesita identificar ubicaciones estratégicas en Latinoamérica para la expansión de sus laboratorios y centros de vacunación, principalmente en Colombia, Argentina, Chile, México, Perú y Brasil, teniendo en cuenta datos como la incidencia de COVID-19, tasas de vacunación y la infraestructura sanitaria disponible. El objetivo es mejorar la respuesta a los impactos de la pandemia y post pandemia, facilitando el acceso a las vacunas, descubriendo qué sectores de la población se deben vacunar primero y en qué ubicaciones, entre otros insights.

Desarrollo del proyecto

El dataset original que contiene los datos a analizar cuenta con 22 millones de registros y 707 columnas, por lo que, conscientes de esta magnitud, los compañeros de ingeniería de datos han preparado uno nuevo enfocado en los países de América Latina. El nuevo archivo posee 12 millones de filas aproximadamente y 50 columnas, lo que facilita su manipulación y análisis.

Como primer paso, leí el dataset con Python y apliqué un filtro por los países que nos pidió la empresa (Argentina, Chile, Colombia, México, Perú y Brasil). Esta acción fue casi insignificante ya que seguía teniendo casi 12 millones de filas. Luego, verifiqué la existencia de datos faltantes, descubriendo que había una cantidad elevada de valores nulos (de hecho, algunas columnas se encontraban casi completamente vacías). Volví a realizar un filtro para reducir el tamaño del dataset, pero esta vez por clave de ubicación (location key), obteniendo finalmente 5946 filas y 50 columnas. A su vez, apliqué un nuevo filtro por fecha mayor a 01-01-2021, obteniendo un total de 3744 filas. Volví a verificar los valores faltantes y procedí a trabajar sobre ellos. La primera tarea fue realizar un filtro for para reemplazar todos los nulos de cada columna (menos la de valores acumulados) por el promedio (mean) de los valores de la misma y segmentando por país (para que el promedio sea el de cada país y no el de todos los países). Con respecto a las columnas con valores acumulados, como no era correcto aplicar tal reemplazo ya que dichos valores van en ascenso siempre y quizás el promedio en algunos casos era menor que el valor faltante, procedí a llenar los vacíos con el valor posterior o anterior (forward fill y backward fill).

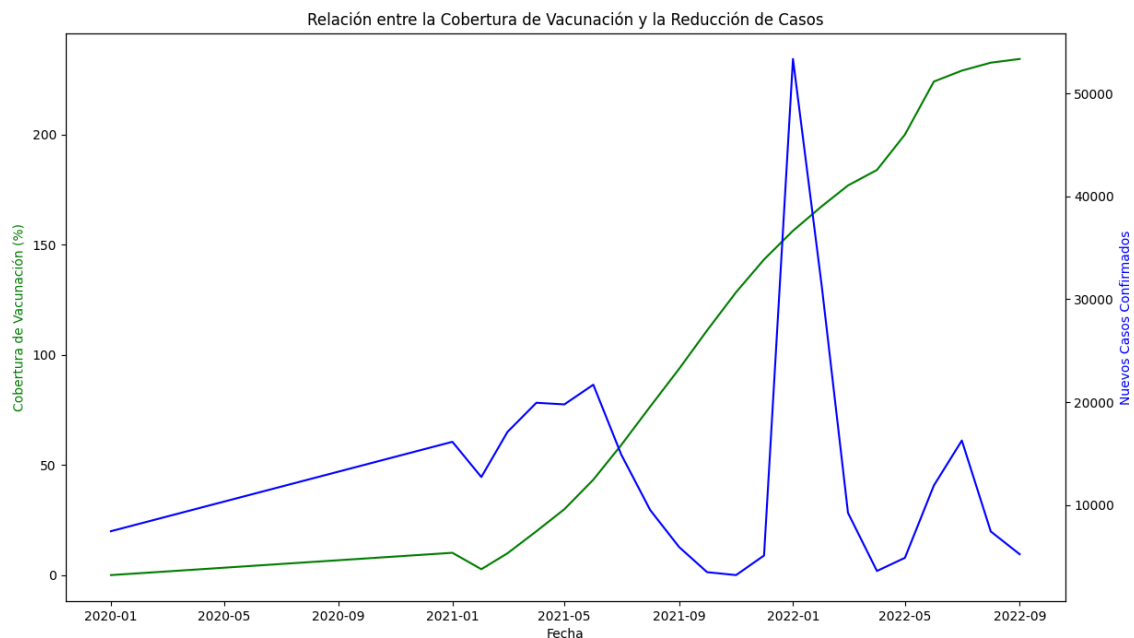
Luego de los filtros, visualicé una breve descripción de las columnas existentes, descubriendo que había valores negativos en columnas que no debería (nuevos confirmados y nuevos recuperados) y procedí a

Módulo 4

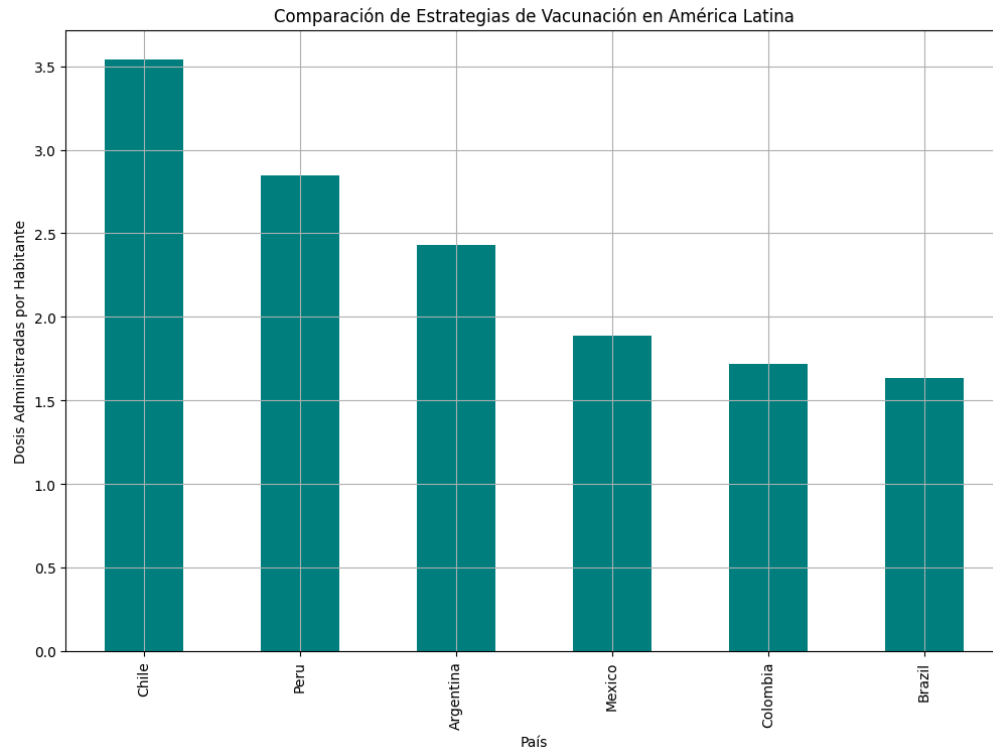
reemplazarlos por 0. Finalmente, con todas estas modificaciones y correcciones realizadas, obtuve mi dataset limpio y reducido en tamaño, listo para guardarlo como un nuevo archivo CSV. Las tareas siguientes fueron la aplicación de estadísticas descriptivas y la ejecución del análisis exploratorio de datos (EDA), descubriendo insights valiosos para presentar a la empresa a través de un dashboard interactivo con visualizaciones y gráficos.

EDA e insights

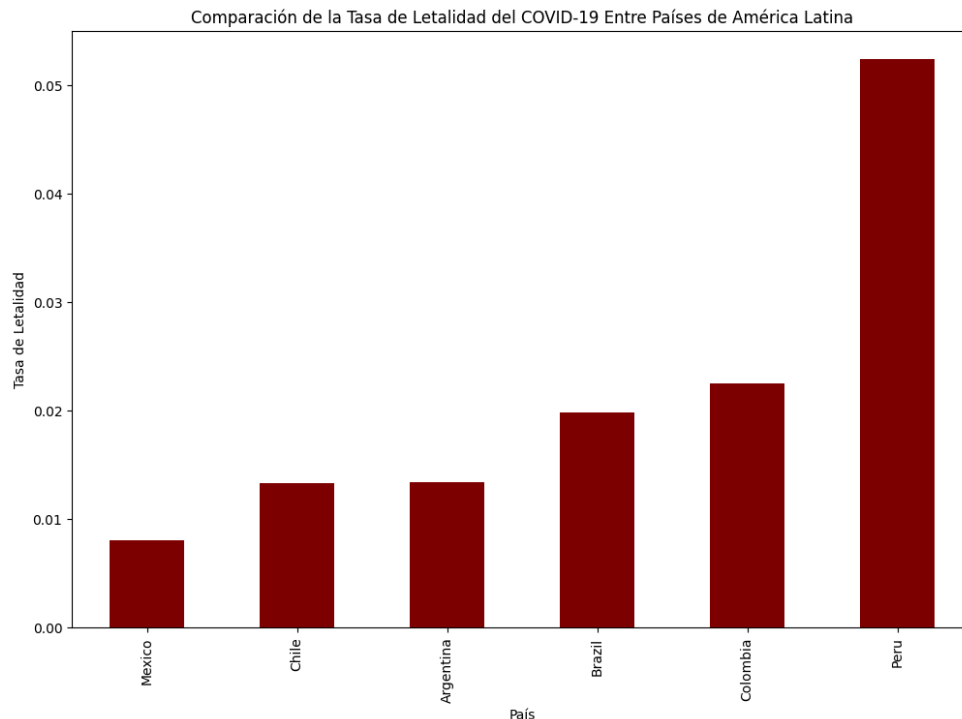
Algunos de los insights obtenidos más relevantes en esta etapa fueron:



Dicho gráfico nos muestra cómo a medida que fue avanzando la cobertura de vacunación en la población los casos de nuevos contagios fueron disminuyendo, remarcando la gran importancia de la vacunación en el manejo de la pandemia. Se puede observar un pico de contagios luego de enero del 2022, lo cual puede deberse a la concentración masiva de gente por vacaciones, época festiva, etc. Sin embargo, dicho pico disminuyó rápidamente.

Módulo 4

Este gráfico me parece clave para nuestro analisis, ya que nos muestra el promedio de dosis aplicadas por habitante segmentado por paises, reflejando que Brasil fue quién menos dosis aplicó y Chile el país que más dosis aplicó. Esto nos lleva a tener una idea de que Brasil cuenta con gran parte de su población que necesita reforzar las dosis.

Módulo 4

Este gráfico nos muestra la Tasa de Letalidad por Covid-19 en los países de nuestro estudio. El país con la tasa más alta es Perú, Brasil se encuentra en el medio y México es quien tiene la tasa más baja.

Análisis del dashboard

El dashboard cuenta con una portada, un panel de datos y análisis generales, otros con análisis económicos y demográficos, y finalmente uno con las conclusiones obtenidas.

En el panel **GENERAL** se visualizan distintos KPIs como la población total, los contagios, vacunas aplicadas, muertes y cantidad de recuperados (los cuales se pueden segmentar por países con los botones a la izquierda inferior). También se observa un gráfico de líneas con las muertes confirmadas por país a lo largo del tiempo, que con un marcador a la izquierda se pueden visualizar los mismos datos en un mapa. También hay dos gráficos de líneas que muestran los casos y muertes confirmadas en el tiempo en todos los países estudiados (con la posibilidad de segmentarlos). Finalmente este panel cuenta con dos gráficos de barras, uno con las dosis administradas por país y otro con la población total vacunada vs la sin vacunar.

En el panel **ECONOMÍA** encontramos gráficos que muestran el PBI en dólares por país, el PBI per cápita por país y el índice de desarrollo humano de cada

Módulo 4

país. A su vez, vemos un gráfico de barras que muestra la cantidad de médicos y enfermeros por cada 1000 habitantes de cada país. También vemos la cantidad de dosis administradas acumuladas por país y un gráfico de barras creado con código python que refleja el promedio de dosis aplicadas por persona por cada país (dato clave para nuestro análisis).

En el panel **DEMOGRAFÍA** vemos gráficos de torta que reflejan datos como la población rural vs urbana, la población masculina vs femenina y la tasa de mortalidad masculina vs femenina. También tenemos otro que nos muestra la distribución de la población en franjas etáreas (de 0 a 39, de 40 a 79 y mayores a 79). Para esto tuve que crear medidas adicionales en Power Bi ya que me parecía excesiva la división por edades brindada en el dataset. También hay un indicador de la esperanza de vida, el cual, al igual que todos los anteriores, puede ser segmentado por país. Finalmente vemos un gráfico de barras que muestra el % de fumadores y diabéticos por país.

Por último, en el panel **CONCLUSIONES** encontramos las sugerencias en cuanto a la ubicación estratégica para la instalación de los laboratorios y los motivos correspondientes. Dicho panel está filtrado por el país Brasil, ya que es acerca de quién queremos presentar los datos.

Conclusiones y Recomendaciones

Tras realizar un riguroso análisis he llegado a la conclusión de que el país recomendado para la Expansión Estratégica de Laboratorios para Investigación y Vacunación de COVID-19 en Latinoamérica es **BRASIL**. Los motivos son los siguientes: es el país que ha registrado el mayor número de contagios (27 millones) y muertes confirmadas (490 mil), así como la mayor cantidad de vacunas administradas. Sin embargo, estos datos están principalmente relacionados con su población, que es la más numerosa entre los países estudiados (Argentina, Chile, Colombia, México y Perú). Por lo tanto, fue necesario explorar otros indicadores. En términos de estrategias de vacunación, Brasil fue uno de los países que peor gestionó la pandemia, con un promedio de sólo 1,6 dosis aplicadas por habitante, en comparación con Chile, que aplicó 3,5. Además, cerca del 40% de su población aún no ha sido vacunada, lo que convierte al país en un lugar estratégico para establecer los laboratorios de Biogenesys. Esto se debe no solo a la gran cantidad de personas sin vacunar, sino también a la necesidad de que una parte significativa de la población reciba dosis adicionales.

Respecto a los recursos disponibles, Brasil cuenta con un promedio de 10 enfermeros y 2 médicos por cada 1,000 habitantes, además de un PBI relativamente alto, lo que sugiere oportunidades favorables para la instalación de los laboratorios.

En cuanto a las comorbilidades, los porcentajes de diabetes y tabaquismo son más bajos en comparación con otros países, por lo que su impacto en la situación actual no es decisivo. Asimismo, aunque Brasil presenta la temperatura promedio más alta, no he encontrado evidencia que sugiera que esto influya significativamente en los contagios y muertes.

Finalmente, como recomendación adicional, sugiero dirigir los esfuerzos de vacunación principalmente hacia los hombres, quienes presentan casi el doble de mortalidad que las mujeres, así como a la población joven y adulta. Dado que la esperanza de vida promedio es de 76 años, no consideraría tomar a las personas mayores de esa edad como prioridad.

Reflexión personal

El proyecto me pareció sumamente interesante y me obligó a exigirme al máximo para poder resolverlo. Fue un gran desafío personal y me genera absoluta satisfacción haberlo culminado comprendiendo cada etapa realizada. Creo que no cambiaría nada si tuviera que volverlo a empezar, ya que estoy conforme con los resultados obtenidos. Sin dudas, mi parte favorita es la de volcar los insights descubiertos en el dashboard, soy arquitecta y mi pasión es el diseño, por lo que allí puedo aplicar esta faceta que tanto me gusta.