

Link de la carpeta de drive

https://drive.google.com/drive/folders/1sSFoN8iWseBC-gEX_eJGFddmgGemK_q5?usp=sharing

Nombre del autor: Lucas Alamanni

Email: lucas.m.alamanni@gmail.com

Cohorte: DA-PT-10

Fecha de entrega: 15/1/2016



Introducción

El siguiente informe es un análisis de la información obtenida en el período de Enero del 2021 a Julio del 2022 concerniente a la pandemia de COVID-19 en la región de América Latina. Este trabajo se enmarca en la campaña de expansión de la empresa farmacéutica BIOGENESYS y tiene como objetivo principal el acercar información valiosa y accionable con respecto a las ubicaciones óptimas para expandir nuevos laboratorios farmacéuticos en la región, teniendo en cuenta la maximización del impacto positivo sobre la salud pública y la mejora de la respuesta de vacunación ante posibles eventualidades futuras de la pandemia.

Se alcanzaron los siguientes **objetivos** organizacionales:

- 1) La información fue centralizada desde diversas fuentes en un dataset unificado, abordando los problemas concernientes a la integridad de la información y a los datos faltantes.
- 2) Se realizaron estudios exploratorios de la información recabada obteniendo métricas estadísticas de valor descriptivo de la situación de la pandemia en América Latina.
- 3) La información obtenida fue resumida en un Dashboard interactivo dirigido a los directivos de BIOGENESYS. El mismo permite explorar distintos aspectos de la información desde un enfoque accionable.
- 4) Se otorgan recomendaciones con respecto a las áreas óptimas de expansión de la empresa, siguiendo los lineamientos del análisis solicitado.

Desarrollo del proyecto

En términos generales, el objetivo del análisis fue comprender el entorno de los mercados de interés con el fin de determinar cuales son las ubicaciones óptimas para expandir los laboratorios y centros de vacunación de la empresa.

Para ello, se tienen en cuenta 3 áreas prioritarias relativas a la expansión estratégica:

- 1) Conocimiento sobre la incidencia de Covid 19 en la región de América Latina. Lo mismo implica:

- a) Estudiar factores generales relacionados con la transmisión del virus (por ejemplo correlaciones con variables como la temperatura, edad de la población y concentración demográfica). Entender las variaciones estacionales de los casos confirmados.
- b) Conocer el estado de cada país de la región con respecto a la pandemia (sus casos nuevos, sus números de muertes).

Esta información es esencial para comprender qué países tienen una necesidad prioritaria de vacunas y para planificar con un enfoque preventivo en qué épocas del año es conveniente realizar la instalación de las nuevas sedes para estar en condiciones de responder a los aumentos estacionales de los casos.

2) Diagnóstico de la situación de infraestructura sanitaria en la región.

- a) Disponibilidad de profesionales (médicos y enfermeras). Consideraciones demográficas sobre un mayor impacto de las vacunación en áreas de mayor concentración urbana.

Este aspecto trae información sobre las posibilidades locales de servirse de recursos de vacunación. Es necesario considerar que la expansión de la empresa sea viable considerando las posibilidades de infraestructura de cada país.

3) Conocimiento sobre las estrategias de vacunación actuales de la región

- a) Conocer las métricas de las distintas áreas relativas a las campañas de vacunación
- b) Entender variaciones estacionales.

Este aspecto del análisis trae luz sobre en qué regiones hay mayor o menor retraso con respecto a las campañas de prevención de Covid 19. Es sabido que una población con mayor cobertura tendrá una menor demanda del producto de la empresa.

Metodología de recopilación y selección de los datos

El análisis está centrado en su totalidad en el archivo .csv otorgado por la empresa. Sobre el mismo se realizó un primer filtro de información para quedarse con información relativa únicamente a los países solicitados y enmarcada en el período posterior al 1-1-2021. Para favorecer la consistencia de los datos se hizo un trabajo relativo al tratamiento de la información nula. Para ello, se agrupó la información de cada país y se realizaron remplazos de los valores faltantes de acuerdo a la lógica de cada variable en cuestión. Para las variables de frecuencia acumulada (las cumulative del dataset) cada valor faltante fue remplazado por su inmediato anterior. Para variables como que reflejan datos extremos (población de la ciudad más grande o temperatura mínima, por ejemplo), los valores faltantes se remplazaron con la función max o min respectivamente. Una gran parte de las variables con varios valores posibles (como population) vieron sus valores faltantes remplazados por su media. El tratamiento específico que recibió cada una se encuentra en el Script de Python.

En tanto a la creación de columnas, se necesitó crear una columna nueva para abordar el análisis posterior. Esta columna se llama dosis y se obtiene mediante la columna de dosis de vacunación acumuladas. Mediante el método .diff de Python se crea una nueva columna

restando a cada valor de esta frecuencia acumulada su inmediato anterior. El resultado es homologable a los de las columnas de nuevos casos detectados o nuevas muertes y permite luego, de manera mucho más clara, obtener gráficos e información relativa a la evolución en el tiempo de las campañas de vacunación.

Luego de estas modificaciones el dataset es guardado en un csv nuevo, el cual fue importado a Power BI para el desarrollo del Dashboard.

No se recortaron variables del dataset original ya que todas fueron objeto de un estudio de correlación mediante una matriz de correlación de tipo HeatMap.

EDA e insights

Para orientarse con respecto a las múltiples variables del dataset se realizaron múltiples cruces de correlación. En cuanto a las variables directamente relacionadas con la incidencia de Covid 19 (`new_confirmed`, `cumulative_confirmed`, `new_deceased`, `cumulative_deceased`) no registraron puntajes altos (se consideró un índice de Spearman mayor a 0.6 o menor a -0.6 para esta afirmación) en los índices de correlación con las variables consideradas del data set. En cuanto a las dosis de vacunas administradas se registró una correlación del 0.6 con la densidad poblacional. En esta misma línea de análisis, el índice más interesante es el que vincula a `gdp_per_capita_usd` con la tasa de enfermeras cada 1000 habitantes (0.87). Como conclusiones principales del análisis de correlación podemos conjeturar que los mercados más atractivos serán aquellos países que registren alta densidad poblacional y que dispongan de mayores recursos económicos. Esto se debe a que la confluencia de ambos factores podría estar relacionada con una mayor disponibilidad de recursos sanitarios (como enfermeras) y una mayor accesibilidad al sistema de salud (cuando más concentrada está la población tiende a haber mayor accesibilidad al sistema sanitario).

Para reflejar esta información en insights accionables, se construyó una matriz haciendo uso de medidas de tendencia central y de dispersión. En ella se calcula por país: la media de densidad poblacional y de área urbana; el gdp per cápita; la tasa de enfermeras y médicas por 1000 habitantes. Además, para dar una medida de dispersión comparativa se calcula para cada una de estas medidas su coeficiente de variación. Lo anterior debería ser un insumo de valor para entender la situación de infraestructura sanitaria de cada país teniendo en cuenta la correlación obtenida con las tasas de vacunación.

Debido a que todas las variables consideradas en esta matriz tienen un único no hay dispersión de los datos y el coeficiente de variación, en consecuencia, es de 0. Se optó por no reflejar este conjunto de 0 en la tabla.

En relación a la información de esta tabla es importante hacer una comparación entre Brazil y México en cuanto a su adecuación estrategia como potencial candidato a alojar a la empresa. Por un lado México presenta la mayor densidad poblacional de toda la región. Sin embargo presenta asimismo una de las dos tasas más bajas de enfermeras por 1000 habitantes. Es de esperar que esta combinación resulte en una sobresaturación del sistema de salud y por lo tanto no sea la mejor opción para insertar el producto de la empresa. Por otro lado, Brazil presenta la mayor tasa de producto bruto per cápita y una de las dos mayores tasas de enfermeras por habitantes. Si bien su densidad poblacional no es tan alta como México (puede deberse a la altísima extensión en el territorio de la selva Amazónica) sí cuenta con ciudades grandes y céntricas y con un sistema de salud desarrollado como para la inserción del producto.

Variaciones temporales

Para establecer insights con respecto a las tendencias de evolución de las tasas de vacunación y datos de incidencia se realizaron gráficos de línea utilizando como eje temporal la variable "date" del dataset.

En cuanto a las variaciones estacionales se conjetura lo siguiente: la información de nuestro dataset que se extiende solamente durante dos años (2021 y 2022) y, sobretodo, las particularidades de la evolución de una pandemia de un virus novedoso como el COVID-19 obligan a algunas conjeturas preliminares antes de analizar los datos. En términos generales los casos registrados, las vacunas realizadas, las muertes y los recuperados muestran una tendencia a decrecer con el paso del tiempo.

Esta evolución al descenso es claramente comprensible si se tiene en cuenta el desarrollo de inmunidad al virus: la exposición prolongada al mismo y las campañas de vacunación existentes hacen que la población, progresivamente sea más resistente al mismo. Esto genera menos muertos, menos vacunación y menos casos registrados (las reincidencias del virus generalmente son menos virulentas y al no llegar a consulta médica tienden a perderse los registros).

Cuando se toma una perspectiva de semanas o de días, es claro que hubo un recrudecimiento de la pandemia entre los meses enero-marzo del año 2022. Hasta aquel momento la evolución de las muertes venía en caída, sin embargo en esos meses tiene una remontada que solo se regulariza hacia junio-julio del mismo año. El hecho de que se haya descartado la correlación entre la temperatura y las variables de incidencia nos invita a pensar que en meses de verano la gente tiende a congregarse más (ya sea por vacaciones o motivos festivos). Esto, al ser un patrón estacional puede indicar una oportunidad de inserción para la empresa: pensar la instalación de la infraestructura de vacunación en otros momentos del año para estar listos a dar respuesta a un posible incremento estacional recurrente de casos registrados durante estos meses.

Estado global de la situación de salud con respecto a COVID-19

El estado de salud de cada país con respecto a la pandemia fue enfocado desde la siguiente óptica: si bien se registraron los números totales de muertos, casos confirmados y vacunaciones realizadas, hay que considerar sobre todo el impacto de la pandemia en términos relativos en cada zona, es decir, sobre el total de la población de cada región. En estos términos fue Perú el país más azotado por las muertes perdiendo al 0.74% de su población. Por otro lado, siguiendo esta lógica porcentual, es Chile el país que tuvo tanto la mayor tasa de registros nuevos como la mayor tasa de vacunación llegando a tres dosis y media (353%) por cada habitante. Esta campaña de vacunación resulta efectiva a la luz de los datos ya que si bien registra la segunda tasa de muertes por Covid en su población, no difiere tanto de la de los demás y sí muestra una brecha significativa con respecto a Perú. Esta información es relevante en términos de comparar los impactos entre distintas regiones.

Sin embargo, para la expansión de la empresa interesan los números absolutos de demanda del producto. En estos términos Brazil es el país que, debido a su gran población demandó más vacunas en el período considerado con 347 millones de vacunas consumidas. Esta es otra de las razones para considerar la pertinencia de la expansión de la empresa sobre este territorio.

Limitaciones y posibles mejoras en análisis siguientes

Por último hay que dejar asentado una de las mayores limitaciones que encontramos en la información. El hecho de que no se sepa la edad de los occisos por Covid 19 impidió trazar

correlaciones con respecto a la mayor tasa de incidencia sobre un rango etario. Esto es uno de los posibles aspectos a mejorar para próximos análisis.

Infraestructura sanitaria

El análisis arribó a las siguientes conclusiones con respecto al estado de la infraestructura de salud de la región. Argentina es el país con mayor tasa de médicos por cada 1000 personas, con amplia diferencia sobre el resto de la región. Chile, por su parte, es el país con mayores tasas de enfermeras por cada 1000 personas. Si bien estos datos son de suma importancia para decisiones de políticas públicas, nuestra evaluación de los sistemas sanitarios de la región se focalizó en los aspectos relativos a los intereses de la empresa. En estos términos se utilizaron medidas de desarrollo económico como el G.P.D per cápita como estimador de la accesibilidad de la población a los recursos de salud. En estos términos, como ya fue adelantado en el punto anterior, es Brazil el país con las condiciones más atractivas para realizar la expansión de la empresa ya que presenta el mayor índice de ingreso y el segundo mayor índice regional de disposición de enfermeras (que se considera, por encima de los médicos, el recurso fundamental para la puesta en circulación de las vacunas)

Campañas de vacunación

Para hacer una evaluación de las campañas de vacunación actuales de los países se tuvieron en cuenta el porcentaje de la población cubierto y las tendencias más actuales sobre la cantidad de vacunas otorgadas en los últimos meses.

En términos absolutos Chile y Perú son los países de la región con mayor tasa de vacunación y Brazil el último.

Con respecto a las tendencias temporales de las campañas de vacunación, cada país presenta su lógica particular que, a nuestro entender, tienen que tener un análisis centrado en la evolución de las dosis totales de vacunación según cada país. Si bien Chile y México tienen un marcado descenso de la vacunación en los 3 últimos meses del año, este comportamiento no es homologable al de los otros países de la región. Colombia, Argentina y Brazil tienen un estancamiento de su línea temporal, indicando que las campañas de vacunación se mantienen al mismo nivel de intensidad desde mediados de año.

En cuanto a la decisión estratégica de la empresa, se considera que un país con menor desarrollo de su vacunación es, potencialmente, un país con más oportunidades para desarrollar una campaña. Es decir, países como Chile parecen haber llegado a una saturación de su expansión mientras que otros, sobre todo Brazil, parecen tener terreno para seguir expandiéndose. Esta afirmación se correlaciona con el descenso de la curva de vacunación en Chile (que indica una tendencia a la baja) mientras que en Brazil, el aplanamiento puede indicar que los recursos con los que se cuenta llegaron a una meseta en términos gráficos y que, potencialmente, con más vacunas y apoyo sanitario se puedan aumentar las vacunas dadas en este país.

Análisis del dashboard

El Dashboard cuenta con una portada que redirige a las 5 pestañas del informe. En la primera cuenta con información relativa a la incidencia de COVID-19 (muertes y nuevos casos) segmentada por país y por mes. Los gráficos de líneas arrojan información sobre la evolución temporal mensual de los casos y su tendencia estacional al decrecimiento. Por

otro lado, el gráfico de la derecha muestra la evolución diaria, mostrando el aumento estacional de enero a marzo desde una óptica más granular hacia los días. El gráfico cuenta con filtros para país y mes.

La siguiente pestaña cuenta con información relativa a los valores totales y a la evolución temporal de las campañas de vacunación segmentadas por país.

Con respecto al sistema de salud, se ofrece una matriz gráfica comparativa de los valores absolutos con los que cuentan las variables seleccionadas. Al mismo tiempo, hay un gráfico de barras que muestra según el selector de variables la distribución de estas mismas por país.

Las últimas dos pestañas tienen información demográfica y climática. Las mismas se ofrecen como complemento del Dashboard ya que, si bien no mostraron correlación con las variables directamente asociadas al impacto del COVID, pueden ser interesantes para análisis específicos que necesite la empresa sobre las características geográficas y de segmentación de la población consumidora.

Conclusiones y Recomendaciones

En primer lugar, es necesario aclarar desde qué lugar nos posicionamos para dar las siguientes recomendaciones. En nuestro entendimiento, BIOGENESYS es una empresa en expansión que necesita de un mercado que se adapte a su oferta de productos pero, al mismo tiempo, esta lógica de expansión tiene que ser mediada por la filosofía altruista de la empresa la cual busca ayudar a la región más abatida por la pandemia de COVID-19. Es por eso que nuestras recomendaciones versan sobre estos dos puntos: el mercado más estratégico y el sector con más necesidades.

País más vulnerable

En términos de mortalidad, **Perú** es el país más afectado. El 0,74 % de su población murió por la pandemia de COVID. Al mismo tiempo es el segundo país más pobre de la región luego de Colombia. La diferencia de muertes en la población peruana es significativamente mayor que en la del resto de los países y, aunque su tasa de vacunación no es la menor de la región, sí creemos que es el país que más necesidades tiene en cuanto a asistencia.

Nuevamente, sabemos que la cantidad total de las ventas en un país con el anteúltimo número total de vacunas consumidas no es la mejor opción en términos de expansión estratégica pero sí es, con objetividad, la zona más afectada por el virus.

Al mismo tiempo, es un país con la mayor porcentaje de población rural y una de las menores tasas de densidad poblacional por lo que se intuye que su población tiene un escaso acceso al recurso del sistema sanitario.

Las tendencias actuales con respecto a las campañas de vacunación muestran una fuerte subida entre junio y julio y unas variaciones posteriores que podrían ser entendidas como escaso suministro o recursos de vacunación (los altibajos de agosto a diciembre pueden interpretarse como una imposibilidad de crecer en su campaña de vacunación)

Mercado más estratégico

En cuanto al mercado más atractivo para la empresa, **Brazil** se posiciona como la opción más adecuada. En primer lugar tiene los mayores ingresos y la mayor población de la zona estudiada. En términos absolutos, es el país con mayor consumo de vacunas superando por casi el doble a México. Sin embargo, es tanta su población que su tasa de vacunación por

habitante es la menor de la región. Esto nos habla de un país con muchísima demanda y, además, con una larga trayectoria para recorrer con respecto a la evolución de las campañas de vacunación. La expansión de las mismas se ve estancada desde julio, lo cual puede ser indicativo (como todavía no decrecieron los números) de que se sigue vacunando gente pero que no se tienen los insumos suficientes para una expansión más agresiva. En términos del sistema de salud, es destacable que tenga la segunda mayor tasa de disponibilidad de enfermeras. Sin embargo, su densidad poblacional es de las más bajas de la región. Este dato se puede comparar con la población rural de este país (13,18%), que no es de las más grandes de la región. Se estipula que Brazil cuenta con una extensísima área cubierta por la selva amazónica, que puede estar influyendo en la escasa densidad de población. Sin embargo, creemos que debido a la importante población urbana y el recurso de las enfermeras, este país tiene un sistema de salud adaptable a la expansión de la empresa.

Mejor momento para la expansión

Nuestro trabajo de time intelligence de las campañas de vacunación de la región muestran que de Abril a Septiembre es el mejor momento para planificar la apertura de nuevos laboratorios y vacunatorios. Esto se debe a que en este período la demanda de vacunas es sustancialmente menor. Siguiendo esta recomendación se puede estar listo para el rebrote estacional que se da entre enero y marzo que es el momento de mayor demanda de vacunas.

Reflexión personal

El uso de Python como lenguaje aplicada a ETL es uno de las mayores herramientas adquiridas en este módulo. Se hizo notorio en la creación de una columna a partir de una frecuencia acumuladas (objetivo que no pude lograr en Power Query y que con el módulo .diff se logra inmediatamente). La potencia de un script tiene así mismo el ángulo de que, modificando uno de sus pasos, se pueden lograr modificar en cadena todo el trabajo posterior. Esto permite ir avanzando en el proceso de análisis de datos y, de manera iterativa, volver sobre el principio para realizar modificaciones necesarias.

En términos de posibilidades de visualización, Power Bi prevalece en su comodidad sobre Python, ya que su interfaz es mucho más intuitiva que la sintaxis de Pandas o Seaborn. Sin embargo el entendimiento de estas librerías es necesaria para crear gráficas muy específicas y potentes (sobre todo las matrices de correlación).

En proyectos siguientes tomaré la nota mental de crear antes de pasar los datos a Power BI todas las columnas necesarias en Python ya que esto me significó mucho tiempo perdido al corregir el proceso de ETL.