



Tecnológico de Monterrey

Instituto tecnológico y de Estudio Superiores de Monterrey

Inteligencia artificial avanzada para la ciencia de datos II Gpo 501

M1. Actividad T5: Final Dashboard

Kathia Bejarano Zamora A01378316

Antonio Oviedo Paredes A01752114

Raul Alejandro Olivares Maldonado A01752057

Rodrigo Mejía Jiménez A01752113

Alex Federico Núñez Escobar A01751559

Natalia Duarte Corzo A01745482

Diego Fernando Landeros Austria A01751654

Sergio Manuel González Vargas A01745446

08 de Noviembre de 2023

I. Introducción

En el año 2019 surgió una enfermedad respiratoria en China, que se propagó rápidamente, convirtiéndose en una pandemia global, que hasta el día de hoy ha tomado más de 6 millones de vidas alrededor del mundo. Dado este evento, surgió la necesidad de poder monitorear el avance de la enfermedad, con el fin de identificar los síntomas del COVID-19, cuántas personas la han contraído, cuántas muertes han surgido, entre otras.

Tomando esto en cuenta y, debido a que este evento ha formado parte de nuestras vidas, nuestro equipo decidió realizar un dashboard interactivo, capaz de proporcionar esta información de manera intuitiva y práctica.

II. Base de datos

Una vez decidido el tema del que se quería realizar al dashboard, se obtuvo una base de datos de kaggle, que consta de los casos de covid en Estados Unidos. La base de datos cuenta con información cómo el número de pruebas realizadas, tanto negativas como positivas, el número de muertes, recuperaciones y personas con ventilador; todo esto clasificado por los estados del país y, de todo el año 2020. A continuación, se describen las columnas del dataset, para tener una mejor comprensión.

Columna	Significado
date	Fecha del informe de casos
state	Estado del informe de casos
positive	Número de casos positivos de COVID-19 reportados en el estado el día especificado
negative	Número de casos negativos de COVID-19 reportados en el estado el día especificado
pending	Número de casos pendientes de resultados de pruebas el día especificado
totalTestResults	Número total de pruebas de COVID-19 realizadas en el estado el día especificado
hospitalizedCurrently	Número de personas hospitalizadas con COVID-19 el día especificado
hospitalizedCumulative	Número total de personas hospitalizadas con COVID-19 desde el inicio de la pandemia
inIcuCurrently	Número de personas en la unidad de cuidados intensivos con COVID-19 el día especificado
inIcuCumulative	Número total de personas en la unidad de cuidados intensivos con COVID-19 desde el inicio de la pandemia
onVentilatorCurrently	Número de personas en ventilación mecánica con COVID-19 el día especificado
onVentilatorCumulative	Número total de personas en ventilación mecánica con COVID-19 desde el inicio de la pandemia

recovered	Número de personas que se han recuperado de COVID-19 el día especificado
death	Número de muertes por COVID-19 el día especificado
deathCumulative	Número total de muertes por COVID-19 desde el inicio de la pandemia
dataQualityGrade	Una calificación de la calidad de los datos
lastUpdateEt	La fecha y hora de la última actualización de los datos
dateModified	La fecha y hora en que se modificó el archivo de datos
checkTimeEt	La fecha y hora en que se verificó la última vez los datos
dateChecked	La fecha y hora en que se verificó la última vez los datos
fips	El código FIPS del estado
totalTestResultsSource	La fuente de los datos de las pruebas totales
hash	Un hash de los datos
grade	Un indicador de la calidad de los datos

III. Descripción de Gráficas

La primera gráfica del dashboard, muestra en un mapa a las personas hospitalizadas y fallecidas en los estados del país. El tamaño de las burbujas representa el número de muertes por estado, con lo cual se puede identificar que Nueva Jersey es el que tuvo más muertes en 2020. Al mismo tiempo, la intensidad del color de las burbujas nos indica el número de personas hospitalizadas, con lo que podemos concluir que Florida fue el estado con más pacientes hospitalizados.

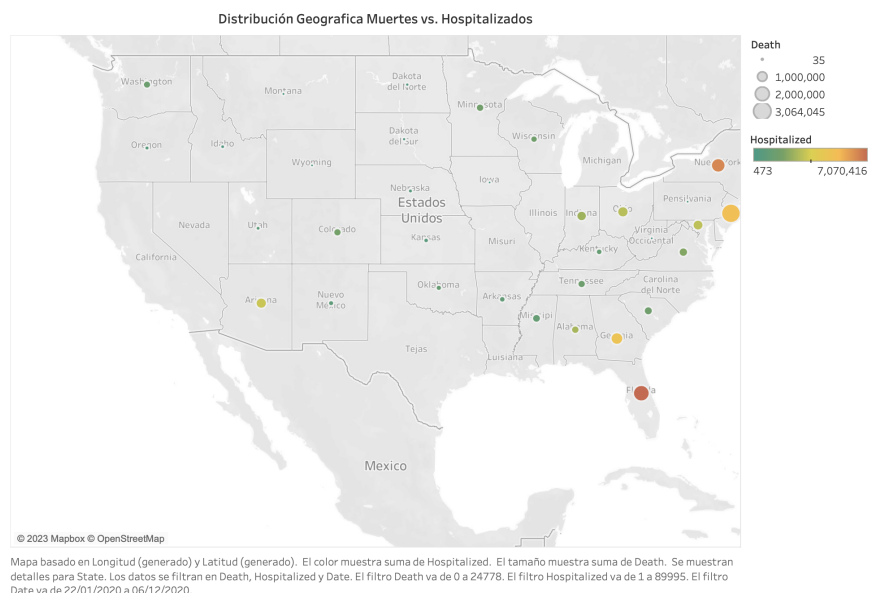


Imagen 1. Distribución geográfica muertes vs. hospitalizados.

La siguiente gráfica busca comparar los casos negativos, contra los casos positivos a lo largo de los 12 meses del año. Como se puede observar, en los primeros meses hubo casos

positivos, sin ninguno negativo y, a partir de abril, cuando las pruebas se realizaban más, comenzaron a surgir las pruebas negativas. Los resultados de las pruebas, tanto negativas como positivas fueron incrementando hasta noviembre, siendo siempre mayor la cantidad de personas con una prueba negativa que positiva. Finalmente en diciembre ambos resultados aparecen con una frecuencia menor.

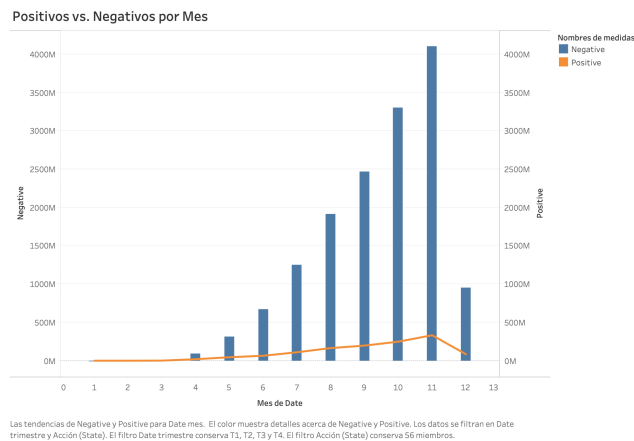


Imagen 2. Positivos vs. Negativos por Mes.

Similar a la gráfica anterior, se realizó el mismo análisis para comparar las pruebas positivas contra los pacientes que se recuperaron. Se puede identificar que los pacientes se recuperaban a partir del mes de marzo, pues a principios de 2020 la enfermedad no era muy conocida y se presentaron menos casos de personas recuperadas, las cuales probablemente no llegaron a ser registradas. Ambas medidas incrementan hasta el mes de noviembre y, disminuyen en diciembre, por lo que podemos concluir que van ligadas entre sí, pues mientras más personas con resultados positivos se presentan, más personas se pueden llegar a recuperar.

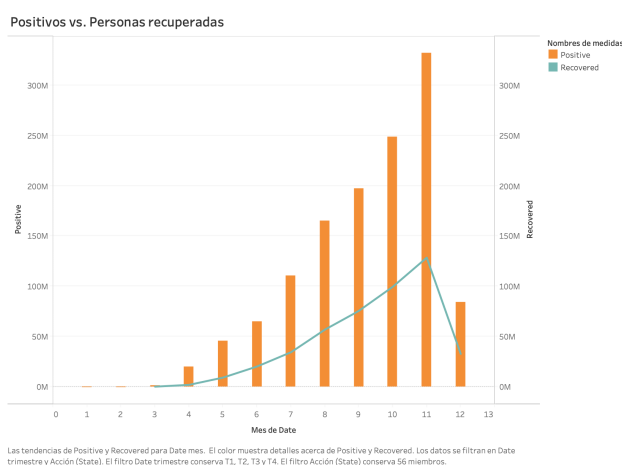


Imagen 3. Cantidad de positivos contra personas recuperadas por mes

En la siguiente gráfica se puede observar la comparación entre todas las personas que resultaron positivas, las personas hospitalizadas y las personas en riesgo con muerte probable en el 2020. De la tabla lo importante es comparar de todos los que son positivos, cuántas personas están hospitalizadas donde se observar que una mínima

parte está hospitalizada en comparación a los positivos, por ejemplo en NC (Carolina del Norte), se observa una diferencia muy grande con 38 Millones de personas positivas y ninguna persona hospitalizada pero sin ninguna persona con muerte probable.

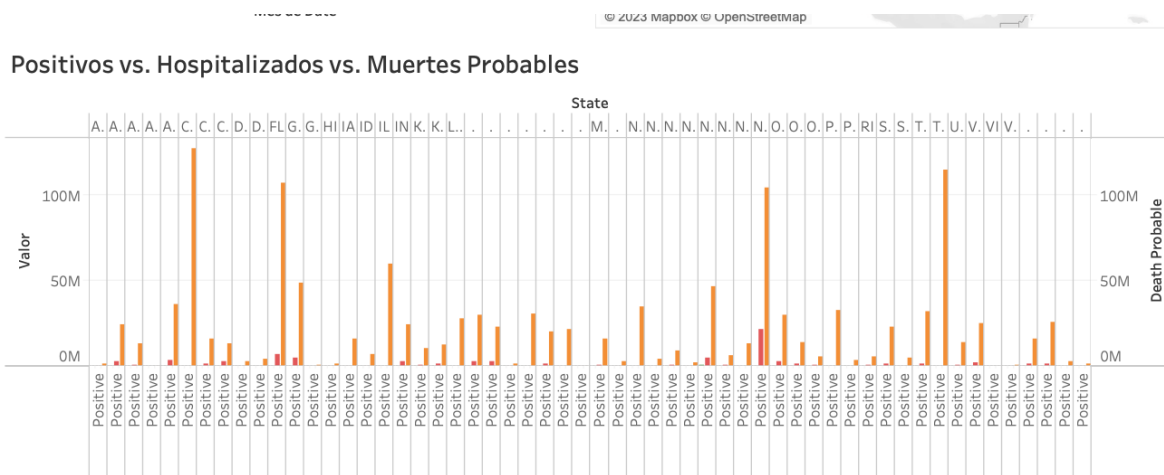


Imagen 4. Gráfica de casos positivos, hospitalizados y mortalidad probable (estimado) por estado

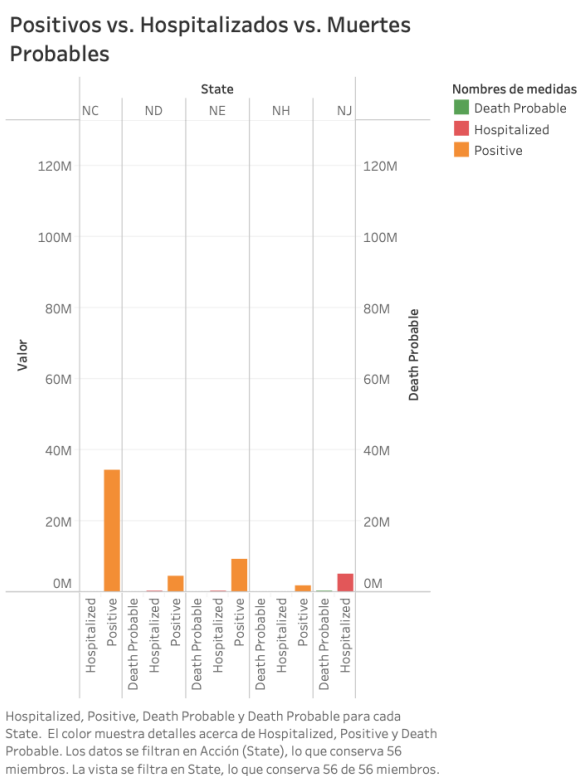


Imagen 5. Zoom-in en la gráfica de casos positivos, hospitalizados y mortalidad probable (estimado) por estado.

La siguiente tabla muestra un desglose de los casos de covid, para poder identificar a lo largo del año 2020, cuántas personas tuvieron resultados positivos, cuántas

estuvieron hospitalizadas, en ventilación, cuántas se recuperaron y, cuál fue el total de muertes.

Desglose

Positive	46,389,594
Hospitalized Currently	532,013
On Ventilator Currently	87,626
Recovered	7,156,832
Death	3,474,435

Positive, Hospitalized Currently, On Ventilator Currently, Recovered y Death. Los datos se filtran en Acción (State), lo que conserva 1 miembro.

Imagen 6. Tabla resumen del acumulado de positivos, hospitalizados, con respirador, recuperados, muertes.

La gráfica del pastel lleva a cabo un análisis distinto a las anteriores pues, no analiza a profundidad el desarrollo del Covid-19, sino que se enfoca en la evaluación de la calidad de los datos. Como podemos observar, la mayoría de los datos tiene una calidad con calificación A +, lo cual quiere decir que gran parte de los datos tienen una buena calidad. Viendo este gráfico se puede poner como objetivo enriquecer los datos que tienen una calificación de B o menor, para poder llegar a qué un 90% de los datos tengan una evaluación de calidad mayor a A.

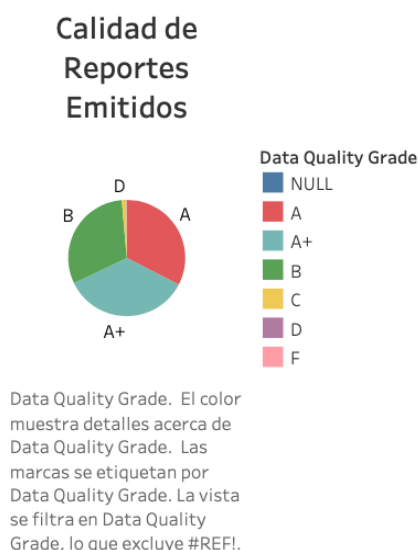


Imagen 7. Tabla resumen del acumulado de positivos, hospitalizados, con respirador, recuperados, muertes

Esta gráfica de burbujas es utilizada para dimensionar el modo de aplicación de las pruebas en los diferentes estado y dimensionar la cantidad de población que resultó positiva. Así se puede observar y realizar seguimiento del ritmo de aplicación de las pruebas en diferentes estado y complementado con gráficas anteriores se pueden identificar deficiencias en la aplicación de pruebas en diferentes estados.

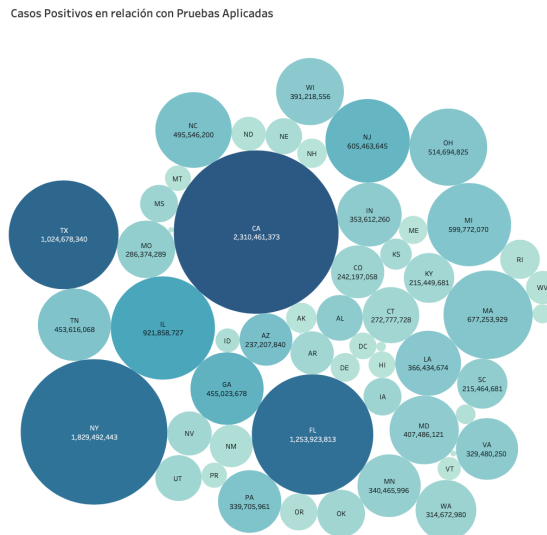


Imagen 8. Gráfica de burbujas de los casos positivos en relación a las pruebas aplicadas

IV. Conclusiones

En el transcurso de este proyecto, tuvimos la oportunidad de explorar cómo la creación de dashboards tiene un papel fundamental en el análisis de datos. Tener la capacidad de contar con una herramienta de visualización de datos se ha vuelto cada vez más relevante para las empresas. Con la información a la mano y de fácil interpretación se pueden tomar decisiones informadas de manera oportuna, mejorando la eficiencia y teniendo un impacto positivo en la capacidad de respuesta a la crisis.

En resumen, a lo largo de este proyecto, hemos identificado que la creación de dashboards va más allá de una simple visualización de datos. Es una herramienta poderosa que impulsa la eficiencia, la toma de decisiones basadas en datos y la comunicación efectiva. En situaciones como la pandemia de COVID-19, un dashboard bien diseñado puede marcar la diferencia al brindar acceso a información actualizada, personalizada y fácil de comprender.