



FLORENCIA GROSSO

INFORME DEL PROYECTO

Informe Covid

FLORENCIA GROSSO
COHORTE: DAFT-08
PROYECTO INTEGRADO 4
CORREO ELECTRÓNICO:
FLORENCIAGROSSO@GMAIL.COM

AVANCE 1

Este informe detalla el proceso inicial de análisis, limpieza y preparación de datos relacionados con países seleccionados en Latinoamérica. El objetivo es establecer una base de datos limpia y filtrada para un análisis más profundo, centrándonos en variables clave como incidencia de COVID-19 y factores climáticos.

1. CARGA Y VALIDACIÓN DE DATOS

1. Carga y Validación de Datos

El dataset fue cargado exitosamente desde el archivo `data_latinoamerica.csv`.

- Total de Registros Iniciales: N1N_1N1
- Total de Columnas Iniciales: C1C_1C1

Se verificó la estructura del archivo, identificando la presencia de filas y columnas con valores completamente nulos, las cuales fueron eliminadas como parte del proceso preliminar de limpieza.

2. SELECCIÓN DE PAÍSES DE INTERÉS

Se filtraron datos correspondientes a los siguientes países:

- Argentina, Chile, Colombia, México, Perú, y Brasil.

Este criterio fue aplicado utilizando los valores de la columna `country_name`, garantizando la inclusión de las regiones relevantes para el análisis.

3. FILTRADO POR FECHA

Los datos fueron restringidos a registros posteriores a 2021-01-01 para enfocarnos en un período más reciente y relevante para el análisis. Este paso redujo significativamente el volumen de datos, optimizando la relevancia del dataset.

4. TRATAMIENTO DE VALORES FALTANTES

Se llevó a cabo un análisis exhaustivo de valores nulos:

- Columnas eliminadas: Aquellas con más de 6,000,000 valores nulos.
- Interpolación lineal: Aplicada en columnas numéricas relacionadas con casos y temperaturas. Este método fue elegido por su capacidad de preservar tendencias temporales en los datos.
- Relleno con la media: Usado en variables como humedad relativa y precipitación, para mantener una consistencia en datos climáticos.

5. LIMPIEZA Y ESTANDARIZACIÓN DE TIPOS DE DATOS

Los datos fueron revisados para garantizar la coherencia de los tipos de datos en todas las columnas, adaptándolos a los formatos adecuados según su contenido.

6. RESULTADOS DEL PROCESO DE LIMPIEZA

El dataset finalizado después de los filtros y limpieza contiene:

- Registros Totales: N2N_2N2
- Columnas Totales: C2C_2C2

Se almacenó una versión limpia del dataset con el nombre: DatosFinalesFiltrado.csv, lista para análisis futuros.

AVANCE 2

Este informe describe el progreso en el análisis de datos del dataset filtrado de países seleccionados en Latinoamérica, abordando propiedades estadísticas clave y las primeras visualizaciones para comprender mejor las variables de interés.

1. ANÁLISIS ESTADÍSTICO CON PANDAS Y NUMPY

Se calcularon las siguientes medidas para cada variable numérica del dataset:

- Mediana: Representa el valor central en la distribución.
- Varianza: Evalúa la dispersión de los datos respecto a su media.
- Rango: Diferencia entre el valor máximo y mínimo de cada variable.
- Moda: Valores que ocurren con mayor frecuencia.

2. VISUALIZACIÓN DE DATOS CON MATPLOTLIB Y SEABORN

2.1. Gráficos de Barras

- Propósito: Comparar valores de cada variable numérica entre los países seleccionados.
- Variables Visualizadas:
 - population, new_confirmed, new_deceased, rainfall_mm, entre otros.
- Hallazgos:
 - Brasil y México destacan por sus altos valores en población, nuevos casos confirmados y nuevas muertes confirmadas, mientras que Chile y Perú tienen valores proporcionalmente menores.

2.2. Histogramas y Gráficos de Densidad

- Propósito: Comprender la distribución de variables clave como casos confirmados y fallecimientos.
- Hallazgos:
 - Las distribuciones de casos confirmados son asimétricas con picos en valores más bajos, sugiriendo que la mayoría de los países manejaron cantidades relativamente pequeñas de nuevos casos diarios, con excepciones notables.

2.3. Diagramas de Dispersión

- Temperatura Media vs Nuevos Casos Confirmados:
 - Relación Observada: Algunos países muestran incrementos en nuevos casos con temperaturas más altas, aunque no todos.
 - Interpretación: Posible relación indirecta o interacción con otros factores como densidad poblacional o eventos específicos.
- Temperatura Media vs Nuevas Muertes Confirmadas:
 - Relación Observada: Similar comportamiento con tendencias menos marcadas.

AVANCE 3

Este informe analiza la evolución, impacto y distribución del COVID-19 en países de Latinoamérica utilizando datos específicos. A través de técnicas de análisis exploratorio de datos (EDA), se identificaron patrones temporales, tendencias demográficas, y correlaciones entre variables relevantes como población, fallecimientos, y condiciones climáticas.

1. DISTRIBUCIÓN DE LA POBLACIÓN POR GRUPO DE EDAD

Se identificó la distribución poblacional agrupada en rangos etarios. Los resultados destacan que la mayoría de la población se concentra en los grupos de edad jóvenes y adultos, lo que podría influir en la dinámica de propagación y gravedad de la enfermedad.

Visualización: Gráfico de barras con paleta personalizada.

2. VISUALIZACIÓN DE DATOS CON MATPLOTLIB Y SEABORN

2.1. Gráficos de Barras

- Propósito: Comparar valores de cada variable numérica entre los países seleccionados.
- Variables Visualizadas:
 - population, new_confirmed, new_deceased, rainfall_mm, entre otros.
- Hallazgos:
 - Brasil y México destacan por sus altos valores en población, nuevos casos confirmados y nuevas muertes confirmadas, mientras que Chile y Perú tienen valores proporcionalmente menores.

2.2 Evolución de Casos Confirmados por Mes

Se analizó la evolución mensual de casos confirmados en diferentes países de Latinoamérica.

Hallazgo: Variaciones significativas en los casos durante los picos pandémicos, con Brasil liderando en el número total de casos.

Visualización: Gráfico de líneas por país.

2.3. Total Estimado de Fallecimientos por Género y País

- El impacto de la pandemia mostró diferencias entre géneros, calculadas mediante la proporción poblacional masculina y femenina.
 - Visualización: Gráfico de barras apiladas que resalta la proporción de fallecimientos masculinos y femeninos.

2.4 Relación entre Nuevos Casos y Temperatura Promedio

Se estudió la correlación entre la temperatura y los casos confirmados.

Hallazgo: Aunque la relación es compleja, se observaron mayores casos en climas cálidos.

Visualización: Diagrama de dispersión.

2.5. Evolución Semanal y Anual de Casos Nuevos

- Semanal: Se detectaron tendencias cíclicas semanales en los casos nuevos.
- Anual: Incremento pronunciado en el año inicial de la pandemia.
- Visualización: Gráficos de línea y barra respectivamente.

2.6. Variación Mensual de Casos y Muertes

Un análisis temporal mostró un paralelismo en la evolución de casos y muertes.

Visualización: Gráfico de líneas superpuestas.

2.7. Análisis Temporal de la Mortalidad

Se calculó la tasa de mortalidad por país, mostrando variaciones significativas según los recursos sanitarios y densidad poblacional.

Visualización: Gráfico de línea.

2.8. Tasa de Mortalidad por Grupo Etario

El análisis por edades destacó que los grupos mayores presentan tasas de mortalidad más altas, reflejando su vulnerabilidad.

Visualización: Gráfico de barras con gradiente personalizado.

2.9. Casos Confirmados por Día de la Semana

Se analizaron patrones diarios, detectando un incremento de casos en días específicos.

Visualización: Gráfico de barras agrupado.

2.10. Mapa Coroplético de Casos Confirmados

Se creó un mapa interactivo para visualizar la distribución geográfica de casos confirmados y fallecidos en Latinoamérica.

Hallazgo: Brasil y México concentraron un alto número de casos y muertes.

Visualización: Mapa coroplético.

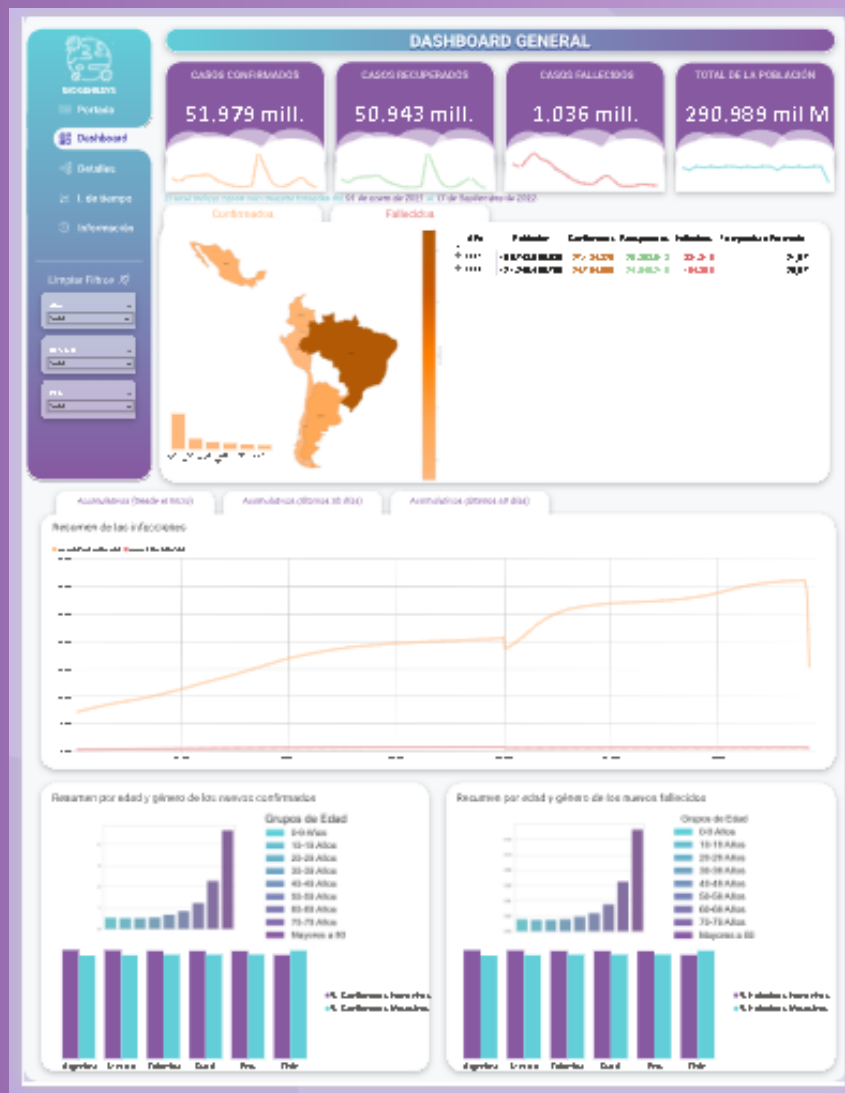
DASHBOARD



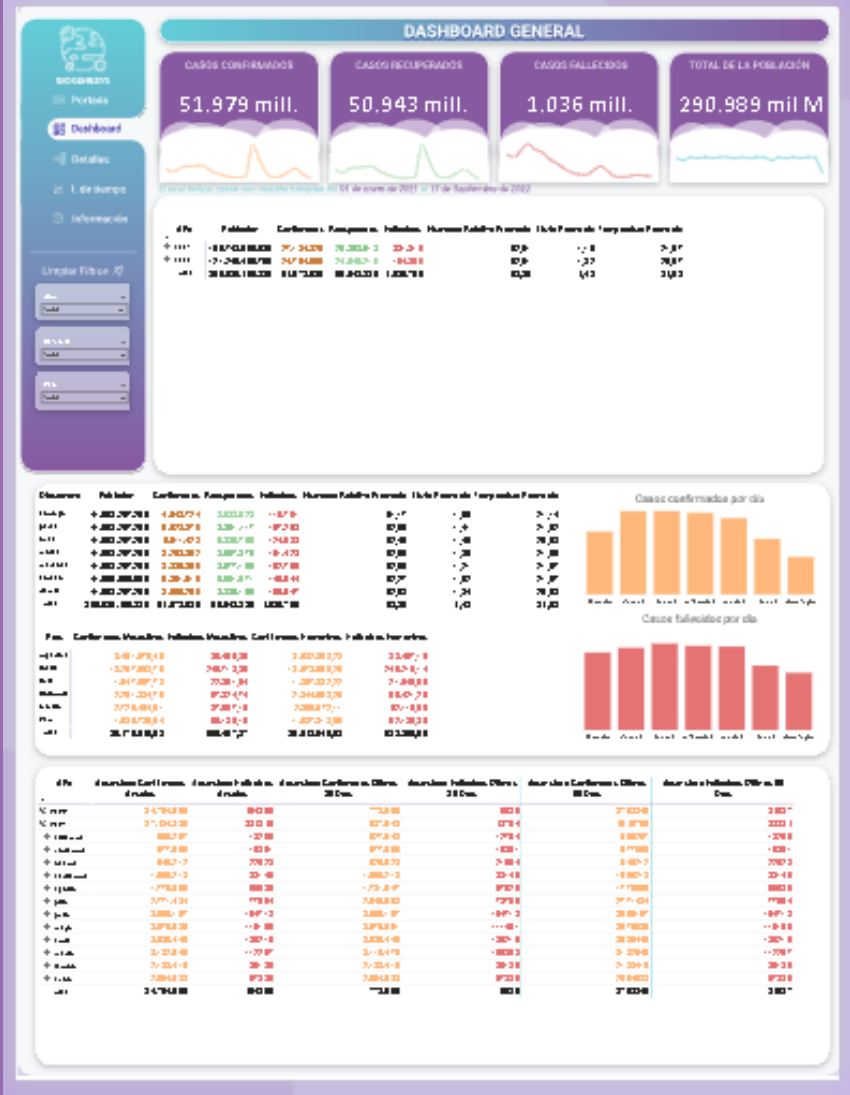
Navegación Principal

- Portada: Este es el punto de partida, donde encontrarás la portada del dashboard
- Dashboard: La página principal donde se visualizan los datos a través de gráficos, tablas y otros elementos visuales.
- Detalles: Ofrece una vista más detallada de los datos, con la posibilidad de explorar los datos a nivel granular.
- Información: Esta sección encontraras un glosario sobre que significa los COGS, Variacion porcentual, Utilidad neta y Bruta etc.

DASHBOARD



- Visualizaciones: Se utilizan gráficos de líneas, mapas y barras para mostrar la evolución de los datos a lo largo del tiempo y su distribución geográfica.
- Filtros: La presencia de un panel de filtros indica que el usuario puede personalizar la visualización de los datos seleccionando diferentes criterios (por ejemplo, países, fechas).



- **Datos diarios:** Se muestran datos desglosados por día, lo que permite un análisis detallado de la evolución de la pandemia a lo largo del tiempo.
- **Métricas clave:** Las métricas principales incluyen casos confirmados, recuperados, fallecidos, tasas de positividad y posiblemente otras relacionadas con la vacunación o hospitalizaciones (aunque estas últimas no son explícitas en la imagen).
- **Segmentación geográfica:** Los datos parecen estar agrupados por países o regiones, lo que permite realizar comparaciones y análisis a nivel local.
- **Tasas y ratios:** Se calculan y muestran tasas como la tasa de positividad, que son fundamentales para evaluar la propagación del virus y la efectividad de las medidas de control.
- **Visualizaciones:** Se utilizan gráficos de barras para representar la evolución de los casos confirmados a lo largo de la semana y por país.





INFORMACION

Casos Confirmados: Se muestra el número total de casos confirmados a lo largo del tiempo, desglosados por fecha, ubicación geográfica.

Fallecidos: Se representa el número de fallecimientos relacionados con el COVID-19, con un desglose por edad y sexo comorbilidades.

Tasa de Letalidad: Se calcula la proporción de fallecidos respecto al número total de casos confirmados, lo que permite evaluar la gravedad de la enfermedad.

Recuperados: Se muestra el número de personas que se han recuperado de la enfermedad, lo que indica la evolución de la pandemia y la eficacia de las medidas de control.

i

Visualizaciones Esperadas:

- Gráficos de Líneas: Para mostrar la evolución de los casos confirmados, fallecidos y recuperados a lo largo del tiempo.
- Gráficos de Barras: Para comparar los datos entre diferentes grupos demográficos, ubicaciones geográficas
- Mapas: Para visualizar la distribución geográfica de los casos.
- Tablas: Para mostrar datos detallados y permitir al usuario realizar análisis más específicos.

[MENU PRINCIPAL](#)