actividad 1: Análisis tendencias en D3.js

Actividad del proyecto. Actividad grupal

UNIR

Herramientas de Visualización

Río Vázquez, **José Manuel**

López Gismeros, **Javier**

Avila Cuadros, **Christian Renzo**

Contenido

[Contexto del Proyecto 2](#_Toc196503965)

[Limpieza y Análisis de los datos 2](#_Toc196503966)

[Implementación 2](#_Toc196503967)

[Modulo 1: Series Temporales 2](#_Toc196503968)

[Análisis de Negocio 3](#_Toc196503969)

[Modulo 2: Representación geográfica de valor y cantidad de ventas acumuladas 4](#_Toc196503970)

# Contexto del Proyecto

El presente proyecto tiene como objetivo desarrollar un dashboard interactivo que integre múltiples módulos de visualización de datos, permitiendo el análisis de series temporales, indicadores clave (KPIs) y la visualización geoespacial de la información. Cada módulo se encarga de representar diferentes dimensiones del negocio, posibilitando una toma de decisiones informada basada en datos reales de ventas.

# Limpieza y Análisis de los datos

Se realizó un proceso sistemático de limpieza sobre el conjunto de datos originales, compuesto por archivos CSV mensuales de ventas y un archivo auxiliar de direcciones geográficas.

En primer lugar, se consolidaron los archivos de ventas en un único *DataFrame* y se verificó la consistencia de los tipos de datos, especialmente en las columnas numéricas y de fecha. Se extrajeron y normalizaron los componentes de la dirección a partir del campo “Purchase Address” para asegurar la compatibilidad en la unión de datos. Se aplicaron filtros para eliminar filas con valores faltantes (NaN) y se revisaron *outliers* mediante técnicas de análisis descriptivo, garantizando que los valores atípicos o erróneos no afectaran el análisis posterior. Finalmente, se llevó a cabo la validación de la coherencia en la variable “Revenue”, recalculando su valor en función del “Price Each” y “Quantity Ordered”, y normalizando todas las columnas relevantes para la integración posterior.

Para la implementación y visualización geoespacial del segundo módulo de nuestro dashboard, se realizó una preparación específica sobre las direcciones asociadas a cada venta registrada. Esta preparación consistió principalmente en la extracción de las localidades a partir de dichas direcciones, lo que permitió agrupar las ventas por localidad. A partir de estas agrupaciones, se calcularon agregados como el total de ventas y la cantidad de transacciones. Esta etapa previa de procesamiento nos permitió reducir la cantidad de ubicaciones geoespaciales a visualizar, facilitando así la representación en el mapa de los totales acumulados de ventas y número de transacciones por localidad, mes a mes, durante el año 2019.

Para el módulo 3: visualización de productos más vendidos, se ha realizado un preprocesamiento de los datos:

* Se seleccionaron los valores de las columnas Product' y 'Quantity Ordered’,
* Se eliminaron valores que pudieran contener nulos o vacíos.
* Se validó que fueran valores numéricos.
* Se agruparon por tipo de producto y se hizo un sumatorio para cada uno de ellos.

# Implementación

La implementación del dashboard se estructura en tres módulos independientes en el archivo “index.html”: módulo de serie temporal, mapa y peso en las ventas de cada producto. Cada módulo se desarrolla de manera autónoma, permitiendo la colaboración de distintos miembros del equipo.

## Modulo 1: Series Temporales

El módulo de serie temporal, encargado por este integrante, utiliza D3.js para la visualización de la evolución de los ingresos a través del tiempo. Se crearon funciones en “serieTemporal.js” que cargan los datos desde archivos CSV (diarios o mensuales) ubicados en la carpeta “data”, se *parsean* las fechas y valores numéricos, y se genera un gráfico de línea interactivo. La aplicación responde a la interacción del usuario mediante botones que permiten alternar entre la vista diaria y la mensual, recargando el gráfico de forma dinámica.

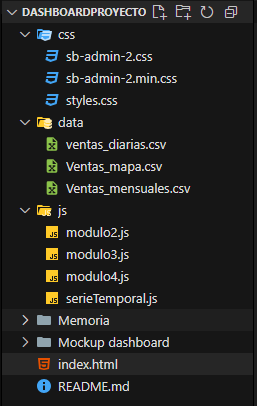


Ilustración : Estructura del Proyecto

### Análisis de Negocio

El estudio de las series temporales reveló una alta volatilidad en los ingresos diarios, observándose fluctuaciones significativas de un día a otro. Se identificaron días en los que los ingresos eran prácticamente nulos, alternando con picos pronunciados que alcanzaron valores sustancialmente mayores. Un análisis mensual permitió evidenciar que dicho comportamiento atípico se concentraba en el periodo de abril a mayo, donde se registró un incremento excepcional en los ingresos en contraste con los meses restantes, los cuales mostraron cifras más moderadas y estables. Dichos hallazgos sugieren la posible influencia de eventos o campañas de marketing puntuales que afectan la dinámica de ventas, siendo recomendada una investigación detallada en estos periodos para optimizar estrategias de negocio a futuro.013119037

## Modulo 2: Representación geográfica de valor y cantidad de ventas acumuladas

Con el objetivo de visualizar de forma clara y efectiva los datos de ventas por ciudad a lo largo del año, se implementó una solución interactiva basada en Leaflet.js y D3.js. Esta solución permite representar geográficamente los datos, utilizando puntos localizados en un mapa que corresponden a distintas ciudades.

Cada punto en el mapa representa una ciudad, y su visualización se construye con base en dos dimensiones clave:

* **Tamaño del círculo**: proporcional a la **cantidad de ventas** registradas en esa ciudad.
* **Color del círculo**: correspondiente al **valor total de ventas** acumulado.

Esta doble codificación visual facilita la identificación rápida de patrones de concentración tanto en volumen como en valor económico, permitiendo a los usuarios detectar de manera intuitiva las ciudades con mayor actividad comercial.

Además, se incluyeron **leyendas visuales** que explican el significado del tamaño y color de los círculos, y se integró una herramienta interactiva de **selección por mes**, lo que permite filtrar los datos dinámicamente a lo largo del año. Al posicionar el cursor sobre cada punto, se muestra un tooltip informativo con el detalle específico de la ciudad, el número de ventas y el monto total.

Esta solución proporciona una representación gráfica clara, interactiva y escalable para el análisis geográfico de ventas, útil tanto para monitoreo ejecutivo como para la toma de decisiones estratégicas.

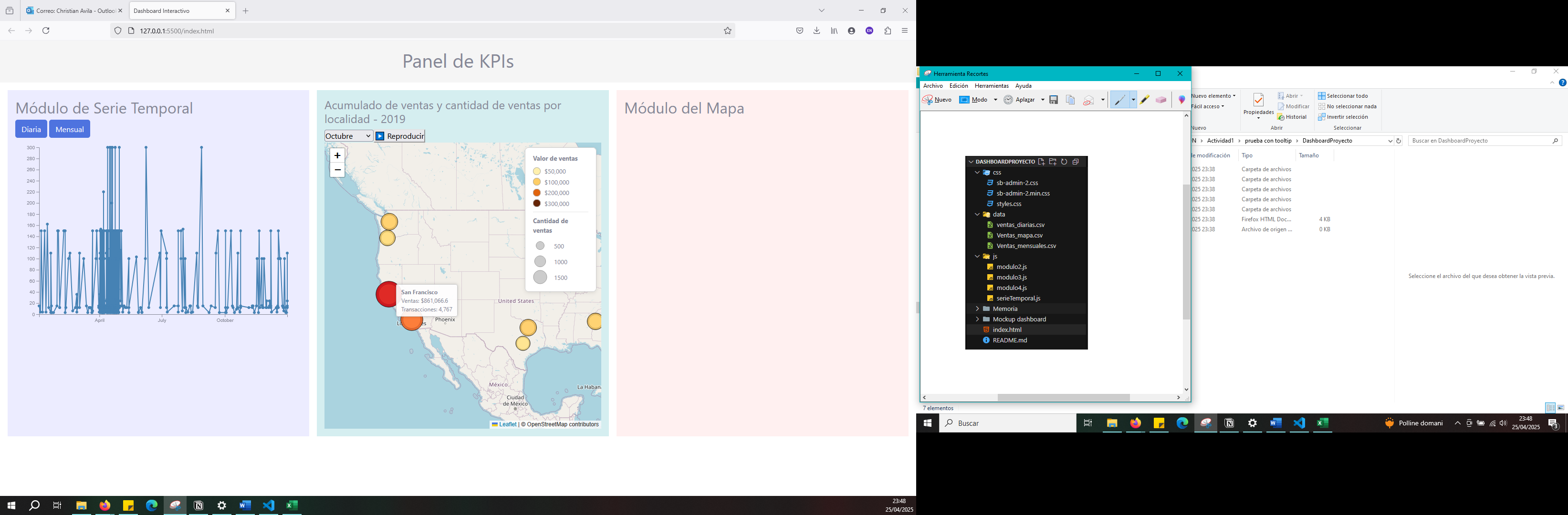


Ilustración 2: Ubicación geográfica de las ventas 2019

## Modulo 3: Visualización de productos más vendidos

Este módulo tiene como objetivo representar de manera clara y ordenada los productos más vendidos, facilitando la comparación visual de las cantidades acumuladas. La visualización, implementada con D3.js, consiste en un gráfico de barras horizontales que muestra los productos ordenados de mayor a menor según la cantidad de unidades vendidas.

El gráfico se genera a partir del archivo productos\_vendidos.csv, ubicado en la carpeta data, y que han sido limpiados como se comenta en el apartado *Limpieza y análisis de los datos.* Los datos son procesados en el archivo productosVendidos.js, donde se convierte la cantidad vendida a valores numéricos y se ordenan los productos en orden descendente. La escala horizontal representa las unidades vendidas, mientras que la escala vertical corresponde a los nombres de los productos.

Cada barra del gráfico se traza con una altura proporcional al producto correspondiente, y se acompaña de un eje izquierdo con etiquetas alineadas a la izquierda para facilitar la lectura. Además, se incorpora un sistema de tooltip que, al pasar el cursor sobre cada barra, muestra información detallada del producto y su cantidad vendida, mejorando la interactividad y usabilidad.

Para mejorar la legibilidad del eje horizontal, se configuraron los ticks en intervalos de 5.000 unidades. Este nivel de granularidad permite apreciar con mayor claridad las diferencias entre productos, incluso cuando los valores son elevados. Asimismo, se realizaron ajustes en la disposición de las etiquetas para evitar superposiciones con las barras, desplazándolas ligeramente hacia la izquierda.

Esta visualización permite identificar de forma rápida los productos con mayor demanda, siendo útil para análisis comerciales, planificación de inventarios y decisiones de abastecimiento.

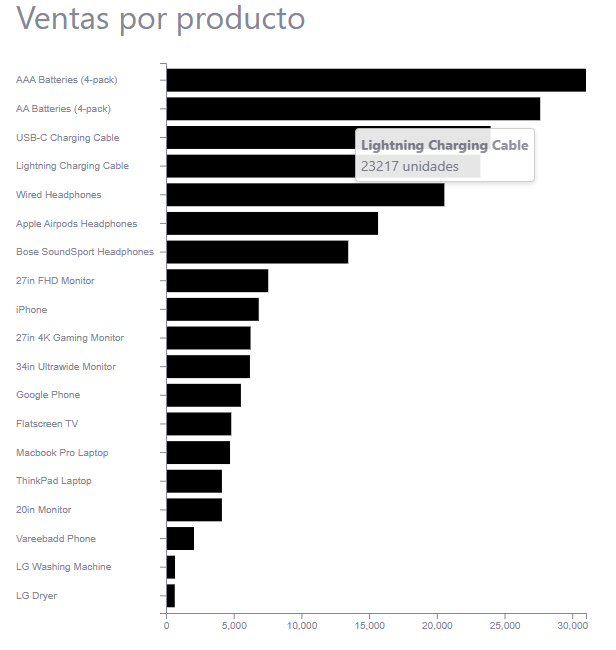


Ilustración 2: Productos más vendidos