

Análisis y Visualización de datos

Práctica de Análisis y Visualización de Datos con Python

Contexto:

Imaginemos que trabajas para una empresa de análisis de mercado que recopila datos de ventas de productos de una cadena de tiendas. Te han asignado la tarea de analizar y visualizar los datos para identificar patrones, tendencias y posibles problemas de calidad en los datos.

Base de datos a utilizar:

Este conjunto de datos '**dataset_practica.csv**' simula información de ventas diarias en dos tiendas (**Tienda_A y Tienda_B**) para dos productos (**Producto_1 y Producto_2**) durante tres días consecutivos en dos regiones (**Norte y Sur**). Cada fila contiene información sobre la fecha de la venta, la tienda, el producto, la cantidad vendida, el precio de venta unitario y la región.

Este dataset incluye datos faltantes en la columna "Tienda" (por ejemplo, la fila 13), datos fuera de rango (por ejemplo, la fila 13 tiene una cantidad de 20 unidades vendidas pero la tienda está en blanco) y casos comunes de ventas para que los estudiantes practiquen diferentes técnicas de análisis y limpieza de datos.

Además se posee otro dataset '**datos_extra.csv**' con las mismas características solo que contiene información sobre 2 productos más (**Producto_3 y Producto_4**) de las mismas dos tiendas, este dataset posee las mismas columnas que el anterior.

Por ultimo el dataset '**Info_Productos.csv**' posee información sobre los productos 1, 2, 3 y 4. La base de datos adicional tiene para cada producto la siguiente información, fabricante, peso, dimensiones, fecha de fabricación y fecha de vencimiento, así como información sobre transacciones asociadas, datos meteorológicos y datos económicos.

Tareas:

Importación de Datos y Chequeos Básicos:

- Importa el conjunto de datos de ventas.
- Realiza los siguientes chequeos básicos:
 - Verifica el número de filas y columnas.
 - Muestra las primeras filas para revisar la estructura de los datos.
 - Verifica los tipos de datos de cada columna.
 - Identifica si hay columnas con valores faltantes.

Filtros de Columnas y Filas, Integración de Datos:

- Filtra las columnas relevantes para el análisis (por ejemplo, fecha, producto, cantidad vendida, precio, etc.).

Análisis y Visualización de datos

- Aplica un filtro para seleccionar solo los datos de las tiendas ubicadas en una región específica.
- Si tienes datos adicionales sobre los productos o las tiendas en otros archivos, intégralos utilizando la función de unión (join) o concatenación (append).

Agregaciones:

- Calcula las ventas totales por producto para cada mes.
- Calcula el promedio de ventas diarias para cada tienda.
- Determina la cantidad total de productos vendidos en cada región.

Chequeos Básicos de Calidad:

- Verifica la consistencia de los datos, como valores negativos para cantidades vendidas o precios.
- Examina la distribución de los datos para identificar posibles problemas, como valores atípicos o extremos.

Valores Faltantes e Inconsistencias:

- Identifica las filas que contienen valores faltantes y decide cómo manejarlos (eliminar, imputar, etc.).
- Busca inconsistencias en los datos, como fechas fuera de rango o nombres de productos incorrectos.

Valores Atípicos/Extremos:

- Utiliza técnicas estadísticas para detectar valores atípicos en las ventas o los precios.
- Decide si estos valores atípicos deben ser corregidos o si indican problemas reales en los datos.