**Inteligencia Artificial - MCT - 2024**

Usted es analista de datos en una cadena minorista con tiendas en varias regiones. La empresa le ha proporcionado un conjunto de datos que contiene datos de ventas del último año. Este conjunto de datos incluye información como la fecha de venta, el ID del producto, la ubicación de la tienda, el número de artículos vendidos y el importe de las ventas. Su tarea consiste en analizar este conjunto de datos para obtener información que pueda ayudar en los procesos de toma de decisiones, como la gestión de inventarios, las estrategias de ventas y la comprensión de las preferencias de los clientes.

Estructura del dataset: archivo *sales\_data.csv*

* Date: La fecha de la venta (YYYY-MM-DD).
* Product\_ID: El ID del producto vendido.
* Store\_Location: La ubicación de la tienda donde se produjo la venta.
* Units\_Sold: El número de unidades vendidas en la transacción.
* Sales\_Amount: El importe total de ventas de la transacción.

(Usando Pandas)

* Cargar el conjunto de datos en un DataFrame.
* Imprima las primeras filas del conjunto de datos para entender su estructura.
* Compruebe y resuma cualquier valor faltante o NaN en cada columna.
* Rellene los valores 0 en Units\_Sold con el valor medio de los mismos items.
* Elimine las filas en Sales\_Amount sea negativo o Store\_Location falte.
* Añada una nueva columna Sale\_Price\_Per\_Unit dividiendo Sales\_Amount por Units\_Sold.
* Convertir la columna Date a un objeto DateTime en pandas.
* Agregue los datos para encontrar el total de Units\_Sold y el Sales\_Amount por Store\_Location.
* Agrupe los datos por mes y calcule las ventas totales de cada mes para identificar tendencias a lo largo del tiempo.
* Identifique los 5 productos más vendidos (por Units\_Sold).
* Escribir una función Python para calcular la varianza de ventas de cada producto.

(usando matplotlib)

* Generar un histograma de Sales\_Amount para visualizar la distribución de los valores de ventas.
* Crear un gráfico de líneas que muestre las tendencias de ventas mensuales.

(usando Numpy)

* Crear una matriz en base a Product\_ID, Units\_Sold y Sales\_Amount.
* Extraer como vector 1D el Product\_ID y filtrar los ID duplicados.
* Calcule la media, la mediana y la desviación estándar de Sales\_Amount para comprender la distribución de las ventas.
* Normalice Sales\_Amount de modo que oscilen exactamente entre 0 y 1 para comparar el rendimiento de las ventas entre productos de diferentes escalas.
* Encontrar el ID de producto con la mayor frecuencia de ventas para centrarse en la optimización de su inventario.
* Ordene el conjunto de datos en función del importe de las ventas en orden descendente para dar prioridad a las transacciones de alto valor para su posterior análisis.
* Asegúrese de que todos los números de coma flotante del informe tengan un formato con dos decimales para garantizar la coherencia.