Day 2 – Report

Task – The Missing Data Trap

Day – Tuesday

Date – 10th Feb 2026

Time – 11:00 AM – 2:00 PM

As I started studying dataset earlier. I went through some other information of dataset and changed the column names in Excel only which will helpful for further process and it will reduce some amount of work ahead.

As in task sheet was mentioned only about to remove duplicated, missing values and clean the dataset, but I started to plot essential visualizations so it will be easier to understand the data.

- Started by uploading dataset into Jupyter lite as instructed to use only Python.
- Imported the recommended libraries for the Cleaning and Plotting.
- Read the dataset, and understood the structure, data types and summary of dataset, using Python.
- Checked for the duplicated values and null values. There are no duplicated values but there are lots of null values are present in dataset, and the datatypes are mismatched (All are objects).
- As there are many missing values are present in dataset so, calculated missing values percentage to see how percent missing values are there.
- Then Converted datatypes from object to respective types(float).
- After that, Object datatypes values are replaced and filled the missing values with NA. And float values were replaced by median.
- After removing, once again checked and print the null values query.

- After that done some calculations for columns 'Quantity' and 'Total Spent' as we changed the values to float.
- After all operations saved the cleaned dataset in csv format.
- Done some value counts of values for further visualizations.
- Plotted Bar graph for Items, which shows no of items sold.

As there are more count of missing values and data not organized well it has taken more time to make it clean that's why I able to plot only one visualization. I will try to plot remaining visualizations.