Data Wrangling Report

1. Introduction:

This report outlines the steps taken to clean and preprocess the supermarket sales dataset, ensuring its accuracy and readiness for analysis.

2. Initial Data Overview

Total records: 1,000

Total columns: 17

Key attributes:

Invoice ID: Unique transaction identifier

Branch: Store branch (A, B, C)

City: Store location (Yangon, Naypyitaw, Mandalay)

Customer type: Regular or membership-based

Gender: Male or Female

Product line: Category of items sold

Unit price: Price per item

Quantity: Number of items purchased

Tax 5%: Applied tax per purchase

Total: Final amount after tax

Date: Transaction date

Time: Transaction time

Payment: Payment method (Cash, Credit Card, E-wallet)

Rating: Customer feedback score (1-10)

3. Data Cleaning Process

Data type corrections: Converted Date column to proper datetime format

Duplicate removal: Identified and eliminated 5 redundant rows.

Handling missing values: Filled 10 missing Rating values using the median.

Feature engineering: Added Total Sales column (Quantity × Unit price).

4. Exploratory Data Analysis (EDA)

Sales distribution: Most transactions ranged between $100 and $300.

Correlation insights: Strong positive correlation between Quantity and Total Sales , indicating higher quantities drive revenue

5. Conclusion

The cleaned dataset has been saved as Cleaned_Supermarket_Sales.csv and is now ready for further analysis