Data Wrangling Report - Supermarket Sales

---

1. Introduction

This report summarizes the data cleaning steps performed on the Supermarket Sales dataset to ensure accuracy, consistency, and usability for analysis.

---

2. Data Quality Issues Identified

During the initial data assessment, several issues were detected, including:

- **Missing values** in the 'Tax 5%' and 'Total' columns.
- **Negative values** in the 'Quantity' column.
- **Invalid Ratings**: Some values were outside the expected range of **1 to 10**.
- **Incorrect Data Types** in the 'Quantity' and 'Unit price' columns.
- **Duplicate records** present in the dataset.
- **Outliers** in the 'Total' column, which could affect analysis.

---

3. Data Cleaning Steps Performed

To address these issues, the following data cleaning techniques were applied:

1. **Handling Missing Values**
    - 'Tax 5%' was filled using **5% of the calculated 'Subtotal'**.
    - 'Total' was filled using **Subtotal + Tax** to ensure accuracy.
2. **Correcting Negative Values**
    - Negative values in 'Quantity' were replaced with the **average positive quantity**.
3. **Fixing Invalid Ratings**
    - Any rating **below 1 or above 10** was replaced with the **median rating** to maintain consistency.
4. **Converting Data Types**
    - Converted 'Quantity' and 'Unit price' to **numeric** format to facilitate calculations.
5. **Removing Duplicate Records**
    - Identified and **removed duplicate entries** to avoid redundancy in analysis.

6. **Handling Outliers**
    - Used the **Interquartile Range (IQR) method** to remove extreme outliers in the 'Total' column.

---

4. Final Data Status

After the cleaning process:

- All **missing values** were addressed.
- Data **consistency and accuracy** were improved.
- The dataset is **now ready for analysis** with a reliable structure.