Data Wrangling Report

Introduction

This report details the data wrangling process for the Supermarket Sales dataset. The objective was to clean, preprocess, and structure the data for analysis.

Data Overview

- **Dataset Name**: Supermarket Sales
- **Total Entries**: 1006
- **Total Columns**: 16

Issues Identified and Fixes

1. Handling Missing Values

- **Issue**: Missing values in 'Total' and 'Tax 5%' columns.
- **Solution**:
  - Dropped rows where 'Total' was missing.
  - Filled missing values in 'Tax 5%' using the median value.

2. Fixing Data Types

- **Issue**:
  - 'Unit price' was stored as an object instead of numeric.
  - 'Date' and 'Time' columns were not in proper datetime format.
- **Solution**:
  - Converted 'Unit price' to numeric.
  - Converted 'Date' and 'Time' to appropriate datetime formats.

3. Location Column Consolidation

- **Issue**: Three separate columns (Yangon, Naypyitaw, Mandalay) represented city information.
- **Solution**: Merged them into a single 'City' column and removed the original three columns.

4. Handling Anomalies

- **Issue**:
  - Duplicate records were present.
  - Some numerical values were negative.
- **Solution**:
  - Removed duplicate records.
  - Converted negative values to absolute values.

Summary Statistics

- Data cleaning resulted in a structured dataset ready for analysis.
- The cleaned dataset was saved as 'Cleaned_Supermarket_Sales.csv'.

Conclusion

The data is now properly formatted and free from inconsistencies, making it suitable for business analysis and visualization.