EXCEL

Introduction

This project focuses on analyzing a dataset using Excel to uncover key trends, relationships, and insights. The process involves cleaning the dataset to remove duplicates or missing values, organizing it into a usable format, and applying various Excel functions and formulas to support data-driven decision making.

Objectives:

The main objectives of this project are to:

1. Clean and organize raw data to ensure accuracy and consistency.
2. Use Excel functions and formulas to analyze the dataset and generate insights.
3. Summarize findings and provide actionable insights based on the analysis.

Excel Data Cleaning Report

1. I imported and reviewed the raw dataset in Microsoft Excel.
2. I used the **Autofit** feature to adjust row heights and column widths, making all data visible.
3. I removed duplicate rows using the **Remove Duplicates** tool under the Data tab.
4. I applied the **TRIM** function to all text columns to remove unnecessary spaces.
5. I cleaned numeric columns by identifying and removing non-numeric values. For example, I replaced the invalid *Price per Unit* value 'inf' with 0 and ensured all numeric fields contained valid numbers.
6. I replaced blank cells in the **Region** column with the **mode** value. Invalid entries such as *Asgard* were first erased (made blank) and then replaced with the mode. The mode formula used was: =INDEX(D2:D25, MODE.SNGL(MATCH(D2:D25, D2:D25, 0))).
7. I cleaned the *Rating* column by removing rows containing non-standard values such as *Spy, Mischief, Leader,* and *Worthy*. This was necessary because, after removing these invalid entries, the column contained four categories (*Average, Excellent, Good,* and *Poor*) with **equal frequencies (6 each)**. Since Excel's MODE function cannot resolve ties, and there was **no clear numeric scale** to convert the ratings into ordered values, it was not possible

to impute missing or invalid ratings. Therefore, the best approach was to remove the non-standard rows entirely.

8. I reformatted the **Date** column, which originally contained both date and time, to display only the date in **MM/DD/YYYY** format.
9. I converted the cleaned dataset into an **Excel Table** (Insert → Table) with headers included.
10. I validated the dataset for accuracy and consistency by applying filters, checking for blanks, and using data validation rules where necessary.

Conclusion

The dataset was successfully cleaned in Microsoft Excel. All issues with duplicates, spacing, invalid entries, and formatting were resolved. Special attention was given to the *Rating* column, where ties and lack of a numeric scale made imputation impossible, leading to removal of non-standard values. The dataset is now ready for analysis.

The cleaned dataset used for this analysis is saved as “Cleaned_Excel_Dataset.xlsx and can be found in this folder.