
Sales Data Project Documentation

Project Overview

This project involved cleaning a messy, real-world dataset using Microsoft Excel. The raw file contained issues like inconsistent formatting, extra spaces, missing values, duplicates, incorrect quantities, and blank rows.
The goal was to transform the dataset into a clean, structured, analysis-ready file.

Data Cleaning Steps

1. Autofit Columns & Rows

Improved visibility by applying autofit to all columns and rows.

2. Removed Duplicates

Used **Data → Remove Duplicates** to eliminate repeated entries across key fields.

3. Trimmed Extra Spaces

Applied the TRIM function to correct leading, trailing, and irregular spacing.

4. Handled Blank Cells & Rows

Filtered and removed empty rows and corrected missing essential values.

5. Converted Data to a Table

Used **Ctrl + T** to create a structured Excel Table for consistent formatting and easier analysis.

6. Corrected Errors with Find & Replace

Fixed misspellings and inconsistent values like “inf”, “null”, and “0”.

7. Final Data Validation

Used helper formulas (e.g., =IF(G2=0, "Invalid Qty", "Valid")) to check for invalid quantities, incorrect data types, and any remaining blanks.

Pivot Table Analysis

Pivot tables were created to provide deeper insights into the cleaned data, including:

- **Total Sales per Region**
- **Total Sales per Product**
- **Quantity Sold per Product**
- **Sales Trend Over Time**

These pivot summaries served as the foundation for the visual reporting.

Dashboard Creation

A clean, interactive **Excel dashboard** was built using pivot tables.
It includes charts and visuals showing regional sales performance, product performance, demand trends, and overall sales progression.

Tools & Functions Used

TRIM, PROPER/UPPER/LOWER, IF, Remove Duplicates, Sorting & Filtering, Find & Replace, Excel Tables, Pivot Tables, Charts & Slicers.

Conclusion

The dataset was successfully cleaned, validated, and analyzed. Pivot tables and a dashboard were developed to provide clear insights and support data-driven decisions.
