# Project Title:

Excel Data Merging, Cleaning, and Visualization using Python, Pandas & Seaborn

# Skills Takeaway:

Python programming for data processing

Working with Excel files using pandas

Data cleaning and merging techniques using pandas

Data Visualizations using Seaborn & Matplotlib

Automating workflows using functions

# Project Domain:

Data Analytics / Automation / Reporting

# Problem Statement:

* Manually analyzing multiple Excel files with redundant or duplicate data is time-consuming and error-prone.
* There is a need to automate the merging, cleaning, and visualization of numeric data from multiple Excel sheets to make informed decisions.

# Problem Approach:

• Accept the folder path containing multiple Excel files.

• Automatically load all files, even with mixed file extensions (.xls/.xlsx).

• Merge them into one dataset and remove duplicate rows.

• Display summaries and missing value statistics.

• Generate useful visualizations such as histograms, boxplots, scatterplots, and correlation heatmaps.

• Prompt the user to name the final output Excel file after processing.

# Example Data Sample:

|  |  |
| --- | --- |
| x | y |
| 10 | 150 |
| 12 | 155 |
| 10 | 150 |
| 14 | 165 |
| 16 | 170 |

# Final Project Results:

The project successfully loaded 5 Excel files with 10 rows each (including duplicates), merged them into a single DataFrame, removed duplicate records, and generated visual insights through plots. It also provided the final cleaned dataset in a user-specified Excel file format, improving workflow efficiency and accuracy.