

Report on Merging and Cleaning 10 Raw CSV Files

Today, I worked on creating, merging, cleaning and creating dashboard on e-commerce data using Python and Power BI.

Creating Raw CSV Files

First, I created **10 raw CSV files** using a Python script. Each file represents e-commerce data for **10 different days**. I used the **pandas**, **os**, and **random** libraries to generate the data. Each file contains details such as **order date**, **product name**, **category**, **quantity**, and **price**.

Merging CSV Files

After creating the files, I merged all 10 CSV files into one single file. For this, I used the **glob** library to collect file paths from the folder. Then, using pandas, I merged all the files **vertically** into one dataset.

Data Cleaning

Once the data was merged, I cleaned it using pandas. I checked for **duplicate rows**, and there were **no** duplicates found. Then, I checked for **missing values**. Some columns had null values, so I handled them, for missing values in **Quantity** and **Price** columns were handled using the **median** value.

The **Category** column was not given properly for all products, so I used a **mapping technique**. Based on the product name, I mapped each product to the correct category. This helped fix incorrect and missing category values. After cleaning, the final dataset was saved as a new CSV file.

Power BI Dashboard

After cleaning the data, I imported the final CSV file into **Power BI**.

I created a dashboard to understand the data better. I also created a **DAX measure for Total Revenue**, which helped calculate total sales and helped me to understanding business performance in a simple way.