

Mini Project

You are given a Python program that reads the data from CSV file named with **Sales_Data_Analysis.csv** (CSV → Comma Separated Values).

Requirement explanation available in the below link

<https://youtu.be/2exbPNYzYwo?si=1r1wLZXpnbQo4oJf>

1. **Sales_Data_Analysis** file contains the various columns saying **InvoiceNo**, **StockCode**, **Description**, **Quantity**, **InvoiceDate**, **UnitPrice**, **CustomerID** and **Country**.
 - a. InvoiceNo represents the unique number for buyer for a bill.
 - b. StockCode represents the unique to each stock for the products.
 - c. Description represents the product name in the dataset.
 - d. Quantity column represents how many products the user has purchased.
 - e. InvoiceDate represents the invoice processed date along with time.
 - f. Unit price represents the price of the one product.
 - g. Customer id represents and points to a specific customer.
 - h. Country column represents the user belongs to which country.

Read csv file, process the data and provides various functions to analyze **Sales_Data_Analysis** information based on different criteria as given below. Refer the below questions and perform some analysis.

1. Load the dataset and clean it by handling missing values and checking for duplicates in critical columns like InvoiceNo or CustomerID.
2. Calculate basic statistics such as total quantity sold, total revenue generated, average unit price, and number of unique customers.
3. Calculate the total revenue generated by each country and add a new column called **"Revenue"** and display **"France"** country revenue.
4. Identify the top 10 most sold **Products (Description)** based on **Quantity**.
5. Find the total number of Invoices and total revenue generated by each customer and display revenue of **CustomerID "12347"**.
6. Extract the month from **InvoiceDate** and calculate the total sales per month to visualize the trend and add a new column called **"Month"** and display the month sales of month **"2011-12"**.
7. Calculate the average number of items (**Quantity**) purchased per invoice for each **CustomerID "12583"**.
8. Find total number of countries and identify the top 5 countries with the highest number of unique invoices.
9. Analyze the distribution of unit prices using NumPy to get the **mean**, **median**, and **standard deviation**.

10. Identify transactions where the **Quantity** is negative (assumed to be refunds) and calculate the total value of refunded products.
11. Calculate total price for each product in a new column "**Total Price**" and extract data to a new file.