Project1: Comprehensive E-Commerce Data Analysis

Problem Statement:

An e-commerce platform is experiencing rapid growth and needs a detailed analysis of its sales, customers, and products. The data is stored in separate tables, and merging them is essential for deriving meaningful insights. You are tasked to clean, merge, and analyze the data using **NumPy, Pandas, Seaborn, and Matplotlib**.

Scenario:

The platform handles thousands of orders daily across multiple countries. The datasets provided include **Orders**, **Customers**, and **Products**, with extensive entries to simulate a real-world scenario.

Sample Datasets:

Order Data (orders.csv):

Order ID	Customer ID	Product ID	Quantity	Order Status	Order Date
1001	C001	P001	2	Delivered	2024-01-10
1002	C002	P002	3	Pending	2024-02-05
1003	C003	P003	NULL	Delivered	2024-03-12
1004	C001	P004	4	Cancelled	2024-01-15
1005	C004	P002	5	Delivered	2024-02-20
1006	C005	P005	1	Delivered	2024-01-11
1007	C006	P003	NULL	Delivered	2024-02-13
1008	C007	P006	3	Pending	2024-03-05
1009	C008	P002	7	Delivered	2024-04-08
1010	C009	P007	5	Cancelled	2024-05-01

Customer Data (customers.csv):

Customer ID	Name	Country	Age	Join Date
C001	Alice	USA	30	2023-01-10

Customer ID	Name	Country	Age	Join Date
C002	Bob	UK	45	2023-05-12
C003	Charlie	Canada	NULL	2023-03-19
C004	Diana	Australia	28	2023-06-30
C005	Eva	Germany	35	2023-04-14
C006	Frank	USA	50	2023-02-18
C007	Grace	India	27	2023-07-22
C008	Helen	Japan	42	2023-09-01
C009	Isaac	UK	39	2023-08-05

Product Data (products.csv):

Product ID	Category	Product Name	Price Per Unit	Stock Quantity
P001	Electronics	Laptop	800	50
P002	Clothing	Jacket	100	100
P003	Furniture	Office Chair	150	30
P004	Electronics	Smartphone	600	70
P005	Appliances	Washing Machine	400	20
P006	Books	Python Book	30	200
P007	Electronics	Tablet	300	25

Required Queries:

Part 1: Data Cleaning

1. Fill missing Quantity in the Orders table with the median value.
2. Replace missing Age in the Customers table with the average age of all customers.
3. Drop orders with the status Cancelled.
4. Ensure no duplicate rows exist in any dataset.

Part 2: Data Merging and Manipulation

1. Merge Orders and Products using Product ID.

2. Merge the result with Customers using Customer ID.
3. Create a new column Order Value as Quantity * Price Per Unit.
4. Filter data for orders where Order Value > $500.
5. Calculate total Order Value for each country.
6. Identify the top 5 products with the highest total sales (Order Value).

Part 3: Data Visualization

1. Create a bar chart showing total sales by Country.
2. Use a line chart to visualize sales trends over time.
3. Generate a heatmap to analyze correlations between Quantity, Price Per Unit, and Order Value.
4. Create a pie chart showing the percentage contribution of each product category to total sales.
5. Visualize stock levels for all products using a horizontal bar chart.

Part 4: Advanced Insights

1. Find the customer who contributed the most to total sales.
2. List the top 3 countries with the highest average Order Value.
3. Highlight low-stock products (Stock Quantity < 20) using conditional formatting.

Expected Outputs:

- Cleaned and merged datasets.
- Detailed insights into sales, customer behavior, and product performance.
- Professional visualizations that effectively communicate findings.
- Actionable recommendations for inventory management and customer targeting.
