EDA – E-commerce Sales Transactions

Duration: 1.5 Hours

Dataset Provided: [click here]

Instructions

1. Use Python (Pandas, NumPy, Matplotlib/Seaborn) to perform the tasks.
2. Clearly show your code + output/plots.
3. Focus on clarity of analysis, correctness of preprocessing, and insights.
4. Submit your Jupyter Notebook/Colab file at the end.

Section A – Data Cleaning & Preprocessing

5. Load the dataset and display its shape and first 5 rows.
6. Check for missing values in each column and handle them appropriately.
7. Identify and remove duplicate rows.
8. Standardize the 'Country' column (fix inconsistent values like 'India ' and 'us').
9. Check for outliers in UnitPrice using visualization (boxplot/histogram). Briefly comment.
10. Fill or handle missing values in TotalAmount column.
11. Convert PurchaseDate into proper datetime format and extract Year, Month as new columns.

Section B – Exploratory Data Analysis

12. Show summary statistics (mean, median, min, max, std) for Quantity, UnitPrice, TotalAmount.
13. Find the top 5 product categories by total sales (TotalAmount).
14. Analyze sales trends month-wise. Plot a bar/line chart.
15. Show sales contribution by Payment Method (visualize as a pie/bar chart).
16. Find the top 3 cities with the highest number of transactions.
17. Compare average order value (AOV) across countries.
18. Identify which Product Category had the highest average discount.

Section C – Insights & Reporting

19. Write 3 key insights about customer purchasing behavior from the dataset.
20. Suggest 2 data-driven recommendations for an e-commerce company based on your analysis.
21. Summarize the main data quality issues you faced and how you fixed them.