Exploratory Data Analysis (EDA) Report

Name: Amisha Patel

21BCE7602

Introduction:

Using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, this report presents the results of an exploratory data analysis (EDA) that was carried out on three datasets: Products.csv, Transactions.csv, and Customers.csv. The analysis's goal was to find insights into product performance, customer behavior, and transaction patterns. The workflow was centered on cleaning, exploring, and visualizing the data in order to support informed decision-making.

Dataset Overview:

Products Dataset: Columns: ProductID, ProductName, Category, Price. No missing values were found in any of the columns.

Transactions Dataset: TransactionID, ProductID, CustomerID, Quantity, Date. No missing values were found, guaranteeing completeness for analysis.

Customers Dataset: Columns: CustomerID, Name, Gender, Age, Location. Missing data analysis revealed some gaps in demographic fields, which were addressed during preprocessing.

Key Steps in the Analysis:

Data Cleaning: Examined the datasets for inconsistencies, duplicate rows, and missing values to confirm their integrity.
used the proper imputation techniques to handle missing demographic data in the Customers dataset.
standardized column names to ensure uniformity between datasets.

Integration of Data: To produce a thorough picture of transactions and the clients and goods they are linked to, the databases were combined using pertinent keys.

Characteristic Statistics: computed summary statistics for numerical columns, such as quantity sold and average product price.
determined the best-selling items, the busiest clients, and the times with the highest transaction volume.
Visualization
Product Details: Product distribution across categories is shown in bar charts.Plots of price distribution are used to illustrate pricing trends.

Analysis of Transactions: Plots of time series that display patterns in transaction volume. Sales density heatmaps by region.

Consumer Conduct: distributions of gender and age using histograms. Age vs transaction frequency scatter plots.

Important Takeaways:

Best-Selling Items: Sales of some product categories were higher than others, indicating possible opportunities for marketing and inventory optimization.

Consumer demographics: The bulk of transactions were made by customers in the 25–40 age range, indicating a crucial target market. Participation was found to be fairly equal by gender, with minor differences by product category.

Trends by Season: Certain months saw high transaction volumes, suggesting seasonal demand trends that may direct marketing tactics.

Sales Based on Location: Opportunities for regional marketing campaigns were created by the concentration of sales in particular areas.

Conclusion and Suggestions: The study provided insightful information on consumer preferences, transaction dynamics, and product performance. These results suggest

courses of action:

1 Concentrate marketing efforts on important demographics (ages 25–40) and high-performing product categories.
2 By scheduling promotions around times of strong demand, you may take advantage of seasonal patterns.
3 Improve inventory control for well-liked items to prevent shortages.
4 Create regional marketing initiatives to take advantage of sales patterns based on geography.

Next Actions:

1 To anticipate future sales and consumer behavior, use predictive modeling.
2 Put in place real-time dashboards to continuously track transactions and sales.
3 Examine cutting-edge segmentation techniques to tailor your client service.