EDA Report

1. Importing Libraries and Data Preparation

The code utilizes pandas for data manipulation, matplotlib.pyplot and seaborn for visualization. Three datasets (Customers.csv, Products.csv, Transactions.csv) are imported and merged into a unified DataFrame using CustomerID and ProductID as keys. This merged dataset forms the basis for further analysis.

2. Data Cleaning and Summary

The merged dataset is examined for missing values, which are identified using isnull().sum(). Duplicate entries are removed to maintain data integrity. A summary of the dataset is generated using describe(include="all"), providing insights into its structure, including key numerical and categorical attributes.

3. Exploratory Data Analysis

a) Most Purchased Products:

The code identifies the top 10 most purchased products by summing the Quantity field, providing insights into customer preferences and high-demand items.

b) Top Customers and Regional Insights:

Customer spending is calculated by summing TotalValue, with a bar chart visualizing the top 10 spenders. Sales performance across regions is also analyzed, highlighting areas with the highest revenue contributions.

c) Sales Trends and Product Categories:

Daily sales trends are plotted to reveal temporal patterns, with TransactionDate converted into a datetime format for accurate aggregation. Product category performance is evaluated by grouping data by Category and summing TotalValue, revealing key contributors to revenue.

d) Customer Signup Patterns:

The SignupDate column is converted to datetime format, and a new column (SignupYearMonth) is created to track customer acquisition trends over time. This analysis identifies periods of significant customer growth.

e) Product Pricing Analysis:

A histogram with a KDE overlay visualizes the distribution of product prices, highlighting pricing strategies and outliers.

f) Revenue by Category:

Revenue contribution by product categories is recalculated, and a bar chart is generated to display the proportion of revenue from each category.

Key Insights

1.   Top-performing products drive a significant portion of revenue,indicating their importance in inventory and marketing decisions.
2.   High-spending customers contribute disproportionately to sales,making them valuable for targeted retention strategies.
3.   Certain regions outperform others in terms of sales, suggestingpotential for localized marketing efforts.
4.   Temporal trends in daily sales and customer signups provideactionable insights for seasonal campaigns and acquisition
strategies.

5.   Pricing analysis reveals the overall pricing strategy, includingoutliers, and informs competitive adjustments.
6.   Category-wise revenue data helps prioritize investments intop-performing segments.