Exploratory Data Analysis Report

Name: Sahil Gupta

Assignment: Data Science Assignment - eCommerce Transactions Dataset

Overview of the Dataset: -

- **Customers.csv:**

  o Contains customer information, including unique IDs, names, regions, and signup dates.

- **Products.csv:**

  o Includes product details such as unique product IDs, names, categories, and prices.

- **Transactions.csv:**

  o Records transaction details such as transaction IDs, customer IDs, product IDs, transaction dates, quantities purchased, prices, and total transaction values.

Insights: -

- **Total Transactions by Region:**

  o Analysis revealed that certain regions (e.g., Region A) dominate sales, contributing the highest revenue. This indicates significant customer activity and sales opportunities in these regions.

- **Total Sales by Product Category:**

  o Product categories such as Electronics and Appliances emerged as the most popular, generating the highest revenue. These categories should be prioritized for marketing and inventory management.

- **Monthly Sales Trend:**

  o The monthly sales trend highlighted peaks during certain months, likely due to seasonal promotions or holiday periods. This emphasizes the need for targeted campaigns during these high-activity periods.

- **Top 10 Customers by Total Spending:**

  o A small subset of customers contributes disproportionately to revenue. For example, the top 10 customers account for a significant portion of the total sales, suggesting a need for personalized offers and loyalty programs.

- **Correlation Analysis:**

  o The correlation heatmap revealed:

- A strong positive correlation between Price and TotalValue, indicating that higher-priced products generate greater revenue.
- A moderate positive correlation between Quantity and TotalValue, emphasizing the impact of bulk purchases on sales.

Key Visualizations: -

- **Bar Plot - Total Transactions by Region:**
  - Showcased the regional distribution of total sales.

- **Bar Plot - Total Sales by Product Category:**
  - Illustrated revenue contribution by each product category.

- **Line Plot - Monthly Sales Trend:**
  - Visualized the sales trend over time to identify seasonal patterns.

- **Bar Plot - Top 10 Customers by Total Spending:**
  - Highlighted the most valuable customers contributing to revenue.

- **Heatmap - Correlation Analysis:**
  - Showed the relationships between key numeric variables such as Quantity, Price, and Total Value.

Recommendations: -

- Focus marketing efforts on high-performing regions and categories to maximize revenue potential.

- Leverage the insights from the sales trend to plan targeted promotional campaigns during high-sales periods.

- Develop loyalty programs to retain top customers and encourage repeat purchases.

- Optimize inventory management by prioritizing popular product categories.

- Encourage bulk purchasing through discounts or offers to boost total revenue.

Deliverables: -

- This report accompanies the code file Sahil_Gupta_EDA.ipynb, which contains all analysis steps and visualizations.