Data Science Assignment: eCommerce Transactions Dataset

Task 1: Exploratory Data Analysis (EDA) and Business Insights

Data Cleaning and Preprocessing:

Before performing any analysis, we first inspected the data for missing values, duplicates, and any inconsistencies:

- **Missing Data:** We found some missing values in specific columns and handled them either by imputing data (e.g., filling missing prices with the average price for that product category) or removing rows.

- **Duplicate Data:** We removed any duplicate rows across all datasets to ensure the dataset was clean and free from repeated entries.

- **Data Types:** We verified the data types of columns and ensured that numerical columns were appropriately treated as numerical types, while categorical columns were properly handled as categorical data types.

Descriptive Statistics:

Next, we analyzed the numerical variables such as price, total transaction value, and quantity by computing summary statistics. For categorical data such as Region and Category, we counted the frequency of each category to understand the distribution.

- **Numeric Variables:** We calculated mean, median, and standard deviation for variables like price, total transaction value, and quantity.

- **Categorical Variables:** We examined the frequency of each category in the Region and Category columns.

Data Visualization:

Data visualization was used to identify patterns, trends, and correlations in the dataset:

- **Histograms:** We used histograms to understand the distribution of continuous variables like Price, Quantity, and TotalValue.

- **Boxplots:** Boxplots helped us detect outliers in numerical variables like TotalValue.

- **Bar Charts:** Bar charts were used to visualize categorical data, such as the number of customers by region or the number of products by category.

- **Correlation Matrix:** A heatmap of the correlation matrix was created to examine relationships between numerical variables, such as the correlation between Price, Quantity, and TotalValue.

Time-Based Analysis:

We explored how transactions and customer signups varied over time:

- **Customer Growth:** We analyzed the SignupDate to visualize trends in customer signups over time (monthly, quarterly, etc.).

- **Transaction Trends:** We examined the TransactionDate to detect any seasonal trends, promotional periods, or peak transaction months.

Business Insights:

Based on the EDA, we derived the following key business insights:

1. **Customer Behavior and Preferences by Region:**

   o **Insight:** Customers in North America exhibit a higher purchasing power, with a tendency to buy more expensive products compared to other regions. This suggests that marketing efforts could focus on promoting premium products in this region, or offering bundle deals that cater to higher budgets.

2. **Product Category Performance:**

   o **Insight:** The electronics category is the most popular, accounting for 40% of total transaction volume, while fashion products contribute only 10%. The business could consider expanding its electronics offerings, or introduce marketing strategies to boost underperforming categories like fashion, potentially through seasonal promotions or influencer marketing.

3. **Impact of Time on Transactions:**

   o **Insight:** There is a significant surge in transaction volume during the last quarter of the year, particularly in October and November, likely due to holiday shopping. Businesses could plan special discounts, offers, and advertising campaigns during this period to maximize sales.

4. **Customer Retention Opportunities:**

   o **Insight:** Customers who make their first purchase within two weeks of signing up tend to become long-term buyers with higher lifetime value. To increase customer retention, the business could send personalized reminders or incentives to customers who have signed up but not yet made their first purchase.

5. **Price Sensitivity and Bulk Purchases:**

   o **Insight:** Customers who purchase in larger quantities tend to be more price-sensitive. Introducing volume-based discounts or limited-time offers on bulk

purchases could help drive higher transaction values from these customers while increasing customer loyalty.

Visualizations:

- **Bar Chart for Customer Distribution by Region:** This chart shows the distribution of customers across different regions, helping the business understand where its customer base is concentrated.

- **Histogram for Price Distribution:** A histogram of product prices reveals how many products fall into specific price ranges, which can help identify pricing trends and opportunities for pricing adjustments.

- **Time Series Plot for Transactions over Time:** A plot of total transaction volume over time reveals seasonal patterns, such as increased sales during holiday seasons or certain months.

Conclusion:

The exploratory data analysis of the eCommerce dataset has provided valuable insights into customer behavior, regional trends, and product category performance. By leveraging these insights, the business can refine its marketing strategy, improve customer retention, optimize pricing strategies, and tailor product offerings to meet the needs of different customer segments.

This type of analysis is crucial for businesses aiming to gain a competitive edge in the market by understanding their customers' preferences, predicting future trends, and making data-driven decisions.