

Exploratory Data Analysis (EDA) of eCommerce Transactions

This report presents a structured approach to conducting Exploratory Data Analysis (EDA) on a dataset comprising customers, transactions, and products in an eCommerce framework. The analysis employs Python libraries—pandas for data manipulation and seaborn & matplotlib for data visualization—to derive insights that inform business decisions.

Data Loading and Preprocessing

The initial step involves importing necessary libraries and loading the datasets using `pd.read_csv()`. The datasets include customer information, transaction details, and product specifications. Following data loading, preliminary inspections like displaying the first few rows and checking the dimensions of each dataset are performed. This step provides an overview and helps identify any immediate issues such as missing or duplicate values.

To ensure data integrity, duplicate entries are removed from both the transactions and products datasets using `drop_duplicates()`. Additionally, column names are standardized to lowercase and stripped of whitespace to prevent potential errors during analysis.

Descriptive Statistics

Descriptive statistics are generated for the numerical columns in each dataset using the `describe()` method. This analysis reveals essential statistics such as the mean, minimum, and maximum values of various fields, shedding light on customer demographics and transaction behavior. For instance, the average transaction quantity and total value provide insights into purchasing patterns.

Univariate Analysis

Univariate analysis focuses on individual variables. Histograms are created to visualize the distribution of quantity purchased and product prices. These visualizations reveal purchasing trends; for example, if most transactions involve low to moderate quantities, this may indicate consumer behavior that favors bulk purchasing incentives or discounts.

Bivariate and Multivariate Analysis

Bivariate analysis explores relationships between two variables. The datasets are merged to create a comprehensive DataFrame, which allows for deeper analysis. A scatter plot illustrates the relationship between price and quantity, revealing potential price sensitivity among customers.

Moreover, a correlation matrix is computed to evaluate relationships between quantity, total value, and unit price. Insights from this analysis may show, for

example, that as prices increase, quantities sold decrease, indicating a need for pricing strategies aligned with consumer purchasing power.

Business Insights Derived from EDA

1. **Customer Behavior:** Analysis of the average total value and quantities purchased provides insights into customer preferences. Understanding purchasing patterns can help tailor marketing strategies and product placements.
2. **Product Performance:** Identifying high-performing products based on sales volume and revenue can assist in inventory management and future product development. Products that show a high correlation between quantity sold and pricing should be prioritized for promotions.
3. **Pricing Strategy:** The scatter plot demonstrates how changes in price affect quantity sold, suggesting that careful pricing strategies could enhance sales. Promotions or discounts on high-margin products could lead to increased overall revenue.
4. **Market Segmentation:** Data on customer regions can facilitate targeted marketing campaigns. Insights derived from regional demographics can guide product offerings that resonate with specific customer segments.
5. **Sales Forecasting:** Understanding trends in purchasing behavior over time can aid in more accurate sales forecasting, ensuring that inventory meets customer demand without overstocking.

Conclusion

This comprehensive EDA effectively highlights critical trends in customer behavior, product performance, and pricing strategies in the eCommerce sector. By leveraging these insights, businesses can make data-driven decisions that enhance customer satisfaction, optimize inventory management, and ultimately drive revenue growth. The combination of rigorous data analysis and visualization leads to actionable strategies, allowing businesses to stay competitive in a dynamic marketplace.