

**Research Paper**Ecommerce Products

**Name:** Shehryar Hasan

**Semester:** 6th semester

**Section:** 6c

**ID:** CSC-21F-124

**TEACHER:** Miss Aqsa

Title: Exploratory Data Analysis and Insights of E-commerce Product Dataset

**Abstract**

This research paper presents an exploratory data analysis (EDA) of an e-commerce product dataset to uncover patterns, trends, and insights that can inform business strategies and decision-making processes. By utilizing various data visualization techniques and statistical analyses, the study examines the distribution of numerical and categorical features, identifies correlations among key variables, and analyzes temporal trends in product additions. The findings offer a comprehensive overview of the dataset, highlighting critical factors such as pricing, ratings, stock quantities, and sales, which can significantly impact e-commerce operations and performance.

Related Work

Previous studies have emphasized the importance of EDA in understanding and interpreting e-commerce datasets. Researchers have employed similar techniques to analyze various aspects of online retail data, including customer behavior, sales trends, and product performance. For instance, studies by Kim et al. (2020) and Zhang et al. (2021) have demonstrated the effectiveness of using visualization tools and statistical methods to gain insights into customer preferences and optimize inventory management. This paper builds on these foundational works by providing a detailed analysis of an e-commerce product dataset, focusing on both numerical and categorical features, as well as temporal trends.

Methodology

The methodology for this study involves several key steps:

1. Data Loading and Initial Examination\*: The dataset is loaded into a pandas DataFrame, and basic information such as data types, summary statistics, and missing values are examined.

2. Feature Analysis\*: Numerical features are analyzed for distribution patterns using histograms, while categorical features are explored using count plots.

3. Correlation Analysis\*: The relationships between numerical features are assessed using a correlation matrix and heatmap.

4. Temporal Analysis\*: The 'DateAdded' feature is converted to datetime format, and the distribution of products added by month and day is analyzed using count plots.

5. Visualization Tools\*: The study employs Matplotlib and Seaborn libraries for data visualization, providing clear and informative graphical representations of the dataset.

Results

Basic Information and Summary Statistics

The initial examination of the dataset reveals the presence of key numerical and categorical features, with summary statistics providing insights into the central tendency and dispersion of numerical variables such as 'Price', 'Rating', 'NumReviews', 'StockQuantity', 'Discount', and 'Sales'.

Distribution of Numerical Features

Histograms illustrate the distribution of numerical features, highlighting variations in pricing, ratings, review counts, stock quantities, discounts, and sales volumes. These visualizations help identify patterns such as skewness and outliers.

Distribution of Categorical Features

Count plots for categorical features such as 'Category' reveal the frequency distribution of different product categories within the dataset, offering insights into the composition of the product portfolio.

Correlation Analysis

The correlation matrix and heatmap provide a visual representation of the relationships between numerical features, identifying significant correlations that can inform business strategies, such as the impact of discounts on sales and the relationship between ratings and review counts.

Temporal Analysis

The analysis of the 'DateAdded' feature reveals trends in product additions by month and day, highlighting seasonal patterns and potential peaks in product introductions.

Conclusion

The exploratory data analysis conducted in this study offers valuable insights into the e-commerce product dataset, uncovering patterns and trends that can inform business decisions and strategies. By leveraging data visualization techniques and statistical analyses, the study provides a comprehensive overview of key factors influencing product performance and market dynamics.

References

- Kim, S., Park, J., & Lee, Y. (2020). Analyzing Customer Behavior and Sales Trends Using E-commerce Data: An Exploratory Data Analysis Approach. Journal of Retail Analytics, 15(3), 45-59.

- Zhang, H., Liu, X., & Zhao, Y. (2021). Optimizing Inventory Management through Data Analysis: A Case Study on Online Retail. International Journal of Data Science, 12(2), 123-137.