

Analyze customer behavior based on transaction data

Objective

To leverage a transactional dataset and perform a comprehensive Exploratory Data Analysis (EDA) to gain deep insights into customer behaviour. This analysis will focus on uncovering customer preferences, identifying purchasing patterns, and understanding overall customer trends. These insights will be used to generate actionable recommendations for Supermarkets to improve business, ultimately enhancing profits and better customer retention.

Scope

The scope of the analysis is to gain insights from Nielsen store transaction data. The dataset includes information about monthly sales, store codes, bill IDs, bill amounts, quantities, values, prices, and categorical attributes such as group, subgroup, company name, mother brand, and brand. The analysis aims to understand sales trends, customer preferences, store performance, and the relationship between different variables.

Requirements

- Python environment with libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn.
- Access to the Nielsen store transaction dataset.
- Basic understanding of data preprocessing, exploratory data analysis (EDA), statistical analysis, and visualization techniques.

Tasks

Data Loading and Understanding

- Load the Nielsen store transaction dataset and understand its structure, features, and data types.

Data Preprocessing

- Check for missing values and handle them if any.
- Identify and remove duplicate records.
- Encode categorical variables into numerical values.

Exploratory Data Analysis (EDA)

- Analyze the distribution of numerical features.
- Visualize overall sales trends by month and store.
- Explore daily sales patterns.
- Identify top-selling products and brands.
- Perform univariate analysis on numerical and categorical variables.
- Conduct correlation analysis between numerical features.
- Investigate the relationship between categorical variables using chi-square tests.

Insights and Conclusion

- Summarize key findings and insights from the analysis.
- Provide recommendations based on insights to improve sales performance, marketing strategies, and store management.
- Conclude the analysis with implications for business decisions and potential areas for further investigation.

Insights and Findings

Monthly Sales Trends: Identify the highest and lowest sales months.

- month M2 has produced the highest revenue.

Store Performance: Analyze total sales and average sales per store to identify top-performing and underperforming stores.

- Store N8 has the highest revenue in all 3 months, whereas store N7 has the lowest revenue off all 10 stores. When it comes to average sales in each store, N3 has the highest average sales.
- Store N8: Given its consistently high revenue, store N8 may not require significant changes in strategy. However, it's essential to maintain the factors contributing to its success and possibly explore opportunities for further growth.
- Store N7: Store N7's consistently low revenue highlights the need for attention and potential intervention. Strategies could include evaluating factors such as product assortment, pricing, marketing efforts, or customer service to improve performance.
- Store N3: Understanding why store N3 has the highest average sales can inform strategies that other stores can adopt to increase their transaction values. This might involve replicating successful practices from store N3 or leveraging insights into customer preferences and purchasing behaviour unique to that store.

Product Preferences: Determine the most popular products, brands, and categories among customers.

- From the univariate analysis, the most sold product group is Biscuits, Cream is most bought subcategory of products. The most preferred products by the customers are from Hindustan Uniliver Limited and the most products are sold by TATA, DAIRY MILK and AMUL mother brands.

Sales Patterns: Understand daily and weekly sales patterns to optimize inventory management and staffing.

- Customers tend to purchase on every first 3-4 days of the month and the highest sales are also observed at last few days of the month. This implies that customers may receive their salaries at the beginning of the month, leading to increased purchasing power and higher spending on consumer goods during this time.
- Businesses can capitalize on this trend by launching promotions or advertising campaigns targeting customers during the first few days of the month to stimulate sales further.
- Towards the end of the month, customers may be more inclined to make purchases as they anticipate upcoming expenses or to utilize remaining budget allocations.

Correlation Analysis: Explore the relationship between numerical features such as bill amount, quantity, and value.

- Value and Price are highly correlated.
- Value and Bill amount are partially correlated.

Categorical Analysis: Investigate the association between categorical variables such as product groups, subgroups, and brands.

- The low p-values indicate that the variables are not independent and that there is likely a relationship between them.

Portfolio

Kaggle: <https://www.kaggle.com/code/monishashrees/eda-customer-behavior-analysis-nielsen>

Github:

https://github.com/BigDataEngineer09/EDA_Coderscave_CustomerBehaviorAnalysis_Nielson

Conclusion

The analysis of Nielsen store transaction data provides valuable insights into sales performance, customer preferences, and store operations. By understanding sales trends, product preferences, and customer behavior, businesses can make informed decisions to optimize sales strategies, improve inventory management, enhance customer experience, and drive overall growth and profitability. Further analysis and experimentation may be needed to validate findings and explore additional factors influencing sales and business outcomes.