Market Basket Analysis Report

Introduction

Market Basket Analysis (MBA) is a data mining technique used to discover relationships between items purchased together in transactions. By identifying these relationships, businesses can make informed decisions regarding product placement, promotions, and cross-selling strategies. This report outlines the process of performing Market Basket Analysis on the Brazilian E-Commerce Public Dataset by Olist, resulting in the generation of association rules.

Methodology

The analysis follows these key steps:

1. Data Preparation:

- The dataset was read from CSV files, including order items, orders, products, and category translations.
- Duplicate entries in the data were removed to ensure accuracy.
- Date columns were converted to the datetime type, and new columns for year, month, and day were created to facilitate time-based analysis.

2. Merging Data:

- The order items were merged with product details and category translations to obtain English descriptions of the product categories. This allows for better readability and understanding of the data.

3. Creating the Basket:

- The basket was prepared for analysis by counting the number of items for each product category on each date. The data was grouped by date and product category, and the item counts were calculated.

4. Applying the Frequent Pattern Growth Algorithm:

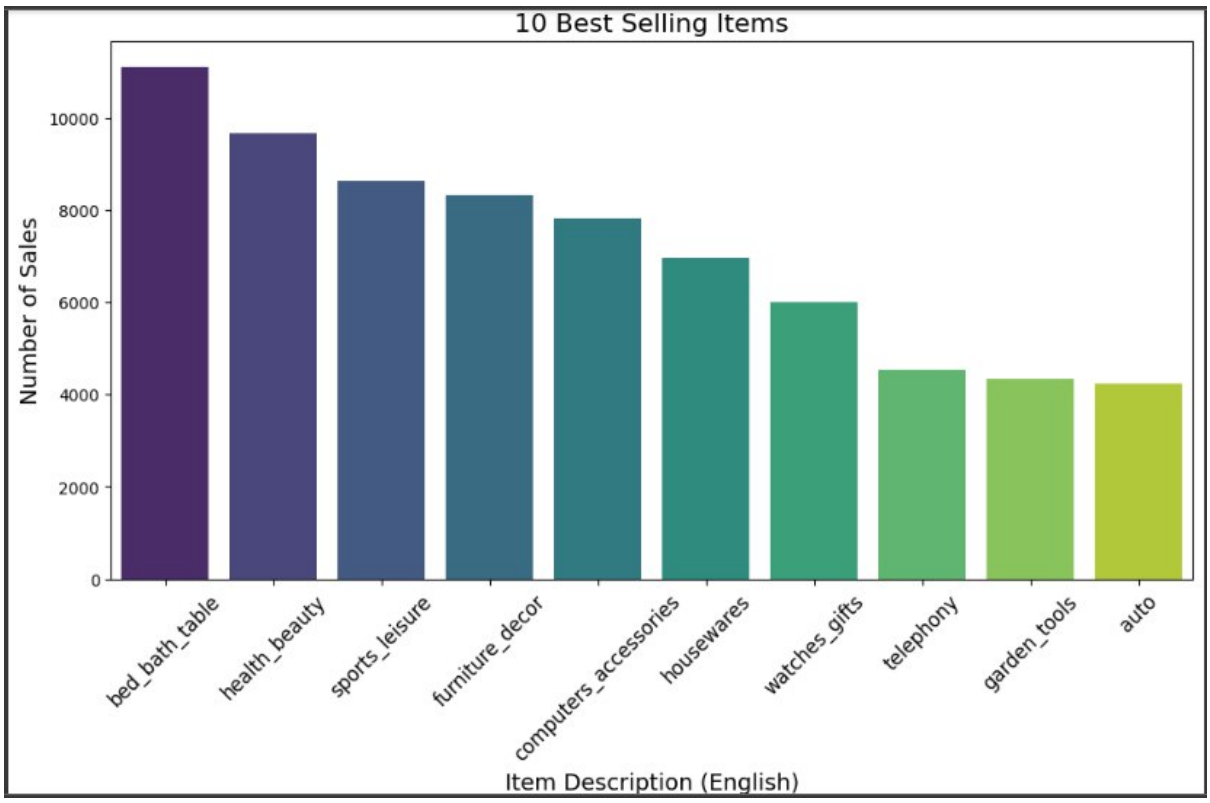
- The fpgrowth function from the mlxtend library was employed to identify frequent itemsets from the transaction data. This method helps find itemsets that appear frequently in transactions.

5. Generating Association Rules:

- After obtaining the frequent itemsets, the association_rules function was used to generate rules based on these itemsets. The metric for evaluation is set to “lift,” with a minimum threshold of 1.0.

Results

The outcome of the analysis includes a set of association rules that highlight the relationships between items purchased together. Each rule indicates:



The image above highlights the top 10 products purchased from the dataset, shedding light on customer preferences and reflecting prevalent buying patterns.

- **Antecedent:** The item(s) that, when purchased, lead to the purchase of the consequent item(s).
- **Consequent:** The item(s) that are likely to be purchased when the antecedent is purchased.
- **Support:** The proportion of transactions that contain the itemset.
- **Confidence:** The likelihood that the consequent is purchased given the antecedent.
- **Lift:** A measure of how much more likely the consequent is purchased when the antecedent is purchased, compared to the likelihood of purchasing the consequent independently.

Insights

- The generated association rules can help in formulating marketing strategies, such as product bundling or personalized recommendations based on purchasing patterns.
- For instance, if the rule indicates that customers who buy a specific category of products are likely to buy another category, a store could promote these items together to increase sales.
- Analyzing the strength of the rules through metrics like lift can help prioritize which associations are most relevant for decision-making.

By utilizing Market Basket Analysis, businesses can leverage data-driven insights to enhance customer experience and optimize sales strategies.