

MARKET BASKET ANALYSIS

Outline

Introduction

Data Exploration

Analysis

Recommendation

Start Page



Introduction

Context

Market basket analysis is a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

The project aims to introduce the concept of market basket analysis, which is a crucial aspect of data science in retail and e-commerce.

Goals for the EDA

To investigate:

Which are the top 20 items purchased by customers?

The distribution of customers based on their identification numbers ("Member_number").

The count or frequency of item descriptions over time

Summary of steps

The Steps taken are:

Data preparation

Exploratory Data Analysis

Market Basket Analysis

Data Visualization

Data Interpretation

Data Exploration

Dataset Overview

The dataset has 38765 rows (for observations) and 3 columns (for variables).

The variables are of different data types, including strings and numerics.

The dataset did not contain null values/ missing values and duplicate values.

Column Titles

The column titles include;
Member_Number, ItemDescription and Date.

Analysis

Exploratory Data Analysis

Exploratory Data Analysis (EDA) and cleaning processes using the Python programming language. The entire analysis was documented comprehensively in a Python notebook. Leveraging the versatility and efficiency of Python, the EDA included tasks such as data exploration, statistical summaries, and visualization techniques. This approach facilitated a thorough understanding of the dataset's structure, trends, and characteristics, ensuring a solid foundation for subsequent analyses and insights.

Feature Engineering and Association Rule Exploration:

In the market data analysis, we conducted feature engineering by grouping items per customer per day into a 'CustomerTransaction' basket. The dataset was enriched by converting relevant columns to strings for consistency. Subsequently, Apriori algorithm was applied to identify frequent itemsets and generate association rules. Focusing on positive associations using Zhang's metric, we visualized the top rules through a heatmap, offering insights into frequently co-purchased items. The analysis provides actionable insights for targeted marketing strategies, revealing customer behavior patterns.

Recommendation

Strategically, market efforts should concentrate on promoting Whole milk, given its consistently high demand. Additionally, a targeted approach to boost sales of other vegetables and rolls/buns could capitalize on their close popularity ranking.

Considering seasonal variations, leveraging the observed surge in September 2015 with targeted promotions can be a key strategy. Simultaneously, acknowledging the historical decline in October and implementing strategic adjustments will be crucial to mitigate potential drops in future years.

In the context of association rules, caution is advised when promoting items with negative association values. A detailed analysis of specific antecedent-consequent pairs with negative associations will provide insights into customer behavior patterns, aiding in informed decision-making for marketing strategies. This comprehensive summary and recommendation aim to empower decision-makers with actionable insights for effective market basket analysis strategies.