MARKET BASKET ANALYSIS REPORT

By students :

Hala faiz alharbi	444001733
Rawyah Alasmari	443005517

introduction

In this report, we want to identify purchasing patterns and relationships between products purchased together by customers, we will use Apriori algorithms to discover frequent item sets and generate association rules.

Objective

Using Association Rule Learning techniques to conduct market basket analysis aims to identify relationships between items purchased in customer transactions. This will help you understand customer behavioral patterns and offer targeted recommendations.

Data Overview

We have Supermarket data, which consists of 522064 rows and 7 columns. The columns we have are:

- BillNo: 6-digit number assigned to each transaction. Nominal.
- Itemname: Product name. Nominal.
- Quantity: The quantities of each product per transaction. Numeric.
- Date: The day and time when each transaction was generated. Numeric.
- Price: Product price. Numeric.
- CustomerID: 5-digit number assigned to each customer. Nominal.
- Country: Name of the country where each customer resides. Nominal.

Data Preprocessing

The first thing we did was load the data using the pandas library. Next, we used the head() and info() functions to understand data and we've checked if there are any missing values, we found

Missing Values: BillNo (0) ,Itemname (1455) ,Quantity (0) ,Date (0) ,Price (0) CustomerID (134041) ,Country (0) .

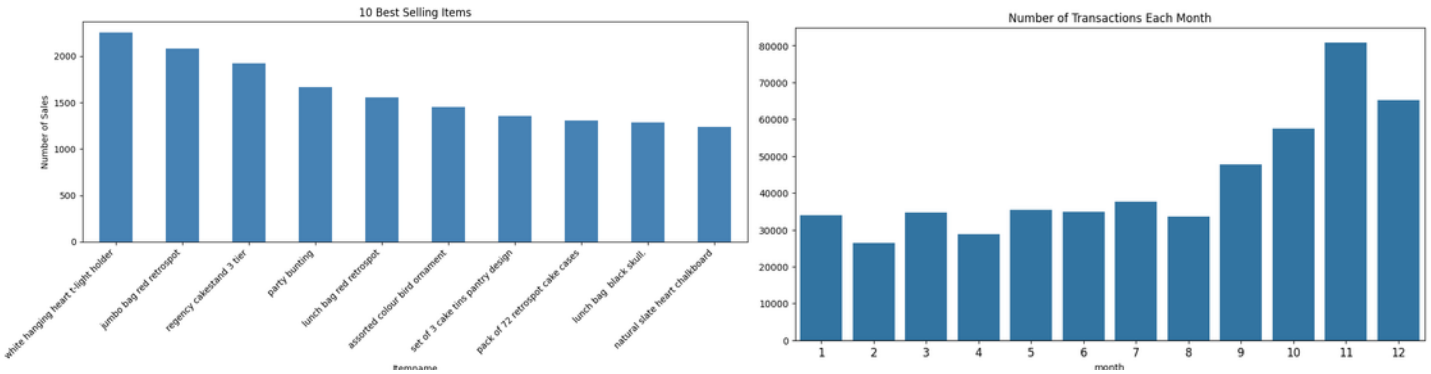
second we have dropped duplicate data , then we change the Date column data type to the DateTime data type, convert all values of Itemname column into lowercase .

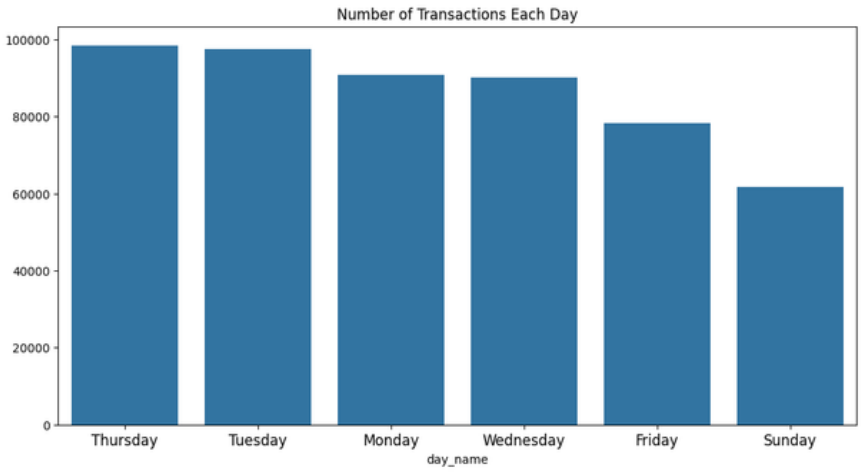
Data Exploratory

According to our research, the White Hanging Heart T-Light Holder is the top-selling product.

Most months have average sales, but December is the highest and February is the lowest.

The difference in the number of sales each day is very slight, the highest sales are on Thursday, and the lowest are on sunday.





Data preparation

we group the data and store it in the 'item_count' variable. So that the data can provide information about CustomerID , date , Itemname , and number of iItemname . Then we will create a transaction column that informs that if CustomerID buys an item on one day, then the transaction is worth 1 and the transaction value will always increase if he buys again.

Next, we create basket variable have a pivot on the data so that we can analyze the relationship between the items using an a priori algorithm. If the item is purchased using the transaction number, the value is 1, otherwise, the value is 0.

	CustomerID	Date	Itemname	Count	Transaction
0	12346.0	2011-01-18 10:01:00	medium ceramic top storage jar	1	1
1	12347.0	2010-12-07 14:57:00	3d dog picture playing cards	1	2
2	12347.0	2010-12-07 14:57:00	airline bag vintage jet set brown	1	2
3	12347.0	2010-12-07 14:57:00	alarm clock bakelike chocolate	1	2
4	12347.0	2010-12-07 14:57:00	alarm clock bakelike green	1	2

Apriori Modelling

we import tow libraries ‘association_rules’ , ‘apriori’ . we use an a priori algorithm to create association rules , So we got Top 5 Discovered Rules:

1. Rule:
    - Antecedent: 60 Teatime Fairy Cake Cases
    - Consequent: Pack of 72 Retrospect Cake Cases
    - Support: 0.0617
    - Confidence: 0.6495
    - Lift: 4.45
    - Real-World Application: Customers who purchase one type of cake cases are likely to buy another type. This suggests potential for bundling cake case sets.
2. Rule:
    - Antecedent: Alarm Clock Bakelike Red
    - Consequent: Alarm Clock Bakelike Green
    - Support: 0.0607
    - Confidence: 0.6868
    - Lift: 9.25
    - Real-World Application: Customers who buy one color of the alarm clock tend to buy another color. Offering promotions for multiple colors or "buy one, get one in a different color" could increase sales.
3. Rule:
    - Antecedent: Alarm Clock Bakelike Green
    - Consequent: Alarm Clock Bakelike Red
    - Support: 0.0607
    - Confidence: 0.8182
    - Lift: 9.25
    - Real-World Application: Similar to the previous rule, but in reverse. Cross-selling strategies for the alarm clocks in different colors would be beneficial.

4.Rule:

- Antecedent: Alarm Clock Bakelike Pink
- Consequent: Alarm Clock Bakelike Red
- Support: 0.0512
- Confidence: 0.7746
- Lift: 8.76
- Real-World Application: Pink and red alarm clocks are often purchased together. Offering a "color combo" discount could boost sales.

5.Rule:

- Antecedent: Baking Set Spaceboy Design
- Consequent: Baking Set 9 Piece Retrospect
- Support: 0.0559
- Confidence: 0.8219
- Lift: 6.15
- Real-World Application: Different baking sets are frequently bought together, indicating an opportunity for bundling similar baking products.

Conclusion

The association rules discovered in this Market Basket Analysis highlight the potential for product bundling and cross-selling. Products such as cake cases and alarm clocks in different colors are often bought together, indicating strong consumer preferences for variety. Implementing targeted marketing campaigns based on these associations, such as "buy one, get another at a discount," could significantly boost revenue and improve customer satisfaction.