

Report on supermarket

Student name	Student ID	Group
Sara Abed Alzulfie	443002749	2
Raghad Adel Alzulfie	444003159	1

Instructor name: Umaima Falath Date: 25-10-2024





Table of Contents

3	INTRODUCTION:
4	DATASET CHARACTERISTICS:
5	DATA EXPLORATION:
6	DATA PREPARATION
7	APRIORI MODELLING
8	RESULT
11	CONCLUSION:
12	RECOURSES:





Introduction:

The retailer wants to target customers with suggestions on itemset that a customer is most likely to purchase. I was given dataset contains data of a retailer; the transaction data provides data around all the transactions that have happened over a period. Retailer will use result to grove in his industry and provide for customer suggestions on itemset, we be able increase customer engagement and improve customer experience and identify customer behavior. I will solve this problem with use Association Rules type of unsupervised learning technique that checks for the dependency of one data item on another data item.





Dataset Characteristics:

1. File name: market data.csv 2

2. Number of Attributes: 7

• Bill No: 6-digit number assigned to each transaction. Nominal.

• Item name: 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.

• Customer ID: 5-digit number assigned to each customer. Nominal.

• Country: Name of the country where each customer resides. Nominal.

3. Number of Row: 5220654. Number of Attributes: 7

Data link:

https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kaggle.com%2Fdatasets%2Faslanahmedov%2Fmarket-basket-analysis&data=05%7C02%7Cs444001268%40uqu.edu.sa%7Cb29e4ed0f82b4dc4d19b08dced1178b9%7C79a057fbb0d544dd8f900bf7151f5c3f%7C0%7C0%7C638645906502953348%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=%2FMwPuZF0PXWW134ym5n36cnBhXmxpdKHvDeHIYWsOWU%3D&reserved=0





Data Exploration:

Data Loading and Preprocessing:

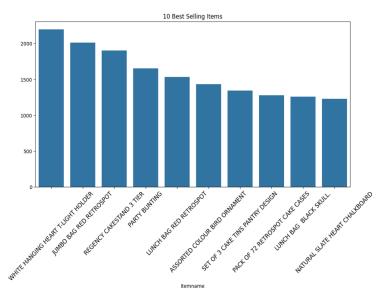
We began by loading the dataset from Google Drive into a pandas Data Frame. The dataset, named marketdata.csv, after loading the data, we examined its structure by checking the shape and displaying the first few rows.

Next, we examined the presence of missing values and to ensure the integrity of our analysis, any duplicate entries in the Data Frame were removed.

The 'Date' column was converted to a datetime format to facilitate time-based analyses. Then we add new columns for year, month, day, and day name for further analysis.

Visualization of Best-Selling Items

To visualize the data, we created a bar plot to display the top 10 best-selling items:



The image shows a bar chart displaying the 10 best-selling items.

The x-axis lists the item names, and the y-axis shows the count or quantity of each item sold. The top-selling item appears to be "WHITE HANGING HEART T-LIGHT HOLDER", followed by other popular items such as "JUMBO BAG RED ."RETROSPOT", "REGENCY CAKE STAND 3 TIER", and "PARTY BUNTING





Data Preparation

Before we carry out modeling processing using an a priori algorithm, we must first prepare the data. At this stage, we group the data and store it in the 'item_count' variable. So that the data can provide information about member number, date, item Description, and number of items purchased.

Next, we will create a transaction column that informs that if CustomerID buys an item on one day, then the transaction is worth 1, and if the next day he buys again then the transaction is worth 2. and if a CustomerID buys then the transaction is worth 3. and so on. the transaction value will always increase.

	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

Next, we will create 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.





Apriori Modelling

First, we import the required libraries, namely 'association_rule' and 'apriori'.

Next, we create a basket variable to hold the grouped data in list form and holds the dataset that will be analyzed using a priori.

then use the Apriori algorithm to create association rules. The Apriori algorithm is used to discover frequent item sets in large transactional datasets. It aims to identify items that are frequently purchased together in transactional data. It helps uncover patterns in customer behavior, allowing businesses to make informed decisions about product placement, promotions, and marketing.





Result

In this analysis, we examined the relationships between various products using three key metrics: Support, Confidence, and Lift.

1. Lift Analysis

Antecedent	Consequent	Lift
(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	4.451266
(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	9.251914
(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	9.251914
(ALARM CLOCK BAKELIKE PINK)	(ALARM CLOCK BAKELIKE RED)	8.759637
(BAKING SET SPACEBOY DESIGN)	(BAKING SET 9 PIECE RETROSPOT)	6.152928

• Highest Lift:

The highest lift value is 9.251914 for both the rules (ALARM CLOCK BAKELIKE RED) \rightarrow (ALARM CLOCK BAKELIKE GREEN) and (ALARM CLOCK BAKELIKE GREEN) \rightarrow (ALARM CLOCK BAKELIKE RED). This indicates a very strong positive association, suggesting that purchasing one color significantly increases the likelihood of purchasing the other color.

• Lowest Lift:

The lift for the rule (60 TEATIME FAIRY CAKE CASES) → (PACK OF 72 RETROSPOT CAKE CASES) is 4.451266, which is still significant, indicating a strong positive association





2. Support Analysis

Antecedent	Consequent	Support
(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	0.094950
(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.088434
(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	0.074238
(ALARM CLOCK BAKELIKE PINK)	(ALARM CLOCK BAKELIKE RED)	0.066093
(BAKING SET SPACEBOY DESIGN)	(BAKING SET 9 PIECE RETROSPOT)	0.067954

• Highest Support:

• The rule with the highest support is between (60 TEATIME FAIRY CAKE CASES) and (PACK OF 72 RETROSPOT CAKE CASES) with a support of 0.094950. This indicates that about 9.5% of the transactions contain both of these items.

• Lowest Support:

• The rule with the lowest support is between (ALARM CLOCK BAKELIKE PINK) and (ALARM CLOCK BAKELIKE RED), with a support of 0.066093, meaning about 6.6% of transactions include both.





3. Confidence Analysis

Antecedent	Consequent	Confidence
(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	0.649510
(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.686842
(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	0.818182
(ALARM CLOCK BAKELIKE PINK)	(ALARM CLOCK BAKELIKE RED)	0.774648
(BAKING SET SPACEBOY DESIGN)	(BAKING SET 9 PIECE RETROSPOT)	0.821918

Highest Confidence:

• The highest confidence value is **0.821918** for the rule (**BAKING SET SPACEBOY DESIGN**) → (**BAKING SET 9 PIECE RETROSPOT**). This indicates that when the baking set is purchased, there is an 82.2% chance that the customer also purchases the retrospot baking set.

• Lowest Confidence:

• The lowest confidence is **0.649510** for the rule **(60 TEATIME FAIRY CAKE CASES)** → **(PACK OF 72 RETROSPOT CAKE CASES)**, indicating a 64.9% chance that if a customer buys the teacake cases, they will also buy the retrospot cases.





Conclusion:

In this analysis, we used the A-Priori algorithm to create association rules. A-Priori algorithms are effective tools for identifying frequently occurring items, allowing us to uncover which products are often purchased together. This helps in revealing customer behavior patterns and assists businesses in making informed decisions regarding product placement, promotions, and marketing strategies.

Overall, the insights gained from this analysis can help businesses better align their offerings with customer preferences, ultimately driving sales and improving customer satisfaction. By understanding purchasing patterns, companies can tailor their strategies to enhance the shopping experience and foster customer loyalty.





Recourses:

https://muhammadfadlisyukur.medium.com/grocery-market-basket-analysis-with-apriori-algorithm-f491aab7f755

https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kag gle.com%2Fdatasets%2Faslanahmedov%2Fmarket-basket-analysis&data=05%7C02%7Cs444001268%40uqu.edu.sa%7Cb29e4ed0f82b4d c4d19b08dced1178b9%7C79a057fbb0d544dd8f900bf7151f5c3f%7C0%7C0%7C638645906502953348%7CUnknown%7CTWFpbGZsb3d8eyJWljoiMC4wLjAwM DAiLCJQljoiV2luMzliLCJBTil6lk1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&s data=%2FMwPuZF0PXWW134ym5n36cnBhXmxpdKHvDeHIYWsOWU%3D&rese rved=0







