International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 14-21

http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

MARKET BASKET ANALYSIS USING MACHINE LEARNING

Atul Sharma[1]
[1]Asstt. Professor
Email: atul.sharma@ipemgzb.ac.in

Dr. Mohammad Salim Hamidi[2]*
[2]Vice Chancellor
Jahan University, Kabul, Afghanistan
Email: avc@jahan.edu.af
ORCID Id: 0000-0002-1564-5149
*Main and Corresponding Author

Yousuf Hotak[3]
[3]Dean
Jahan University, Kabul, Afghanistan
Email: dean_cs@jahan.edu.af

Received: 2024-06-23 Accepted: 2024-07-17 Published online: 2024-09-03

Abstract

The test database products that are most likely to be purchased together, primarily in the retail and economic sectors, are analyzed using the "market basket analysis" method. This method works particularly well for optimization. We used the market basket dataset from the Kaggle library for our research study. Using the Python programming language, this test database was examined for the well-known management subject of "Market Basket Analysis."

Keywords: Market Basket Analysis, Machine Learning, Python.

1. INTRODUCTION

Market Basket Analysis is a valuable tool for businesses seeking to optimize their product offerings, increase cross-selling opportunities, and improve marketing strategies. Market basket analysis can be used to enhance the profitability of any business. Machine Learning is rewarding the retail industry in a unique way. It supports the retail sector in all areas, from predicting sales success to locating customers. Market basket analysis (MBA) is one such top retail application of machine learning. It helps retailers know what products people are purchasing together so that the store/website layout can be designed in the same manner.[1]

We have followed the below mentioned process for the task of Market Basket Analysis research project:

1. Collect preferably real-time transactional data from a reliable source.
2. Analyse the product sales and trends using well known algorithms like Apriori, FP growth, Decision Tree etc.
3. Interpret the results obtained as per the step 2 above.
4. Make strategy based on the Interpretation as per step 3 above.

2. REVIEW OF LITERATURE

(Chaudhary, S. (2022, February 11) has talked about the importance of Market Basket Analysis in his research; (Stevens,S. (2023, September 7) has talked critically about the Data Analysis implication using Machine Learning; (Simplilearn. (2022, November 22) has discussed about the key components of the Market Basket Analysis; (McColl, L. (2022, March 1) has discussed about the Market Basket Analysis using Python;(How to use market basket analysis for retail and marketing. (2023, December 19) talks about the analysis of Market Basket analysis for retail sector; Overview of market basket analysis. (n.d.) discusses about the overview related to the Market basket analysis; Predoiu, O. (2024, April 2) talks about customer behavior analysis; Elnahla, N. (2021) discusses about Retail lance and its Marketing Implications with reference to Market Basket Analysis.

3. RESEARCH METHODOLOGY

We have worked on the Quantitative research. The past (historical) research data has been downloaded from the Kaggle repository for analysis. Now this data has been analyzed very effectively using Python language. According to Dawson (2019), a research methodology is the primary principle that will guide your research. It becomes the general approach in conducting research on your topic and determines what research method you will use. A research methodology is different from a research method because research methods are the tools you use to gather your data (Dawson, 2019). You must consider several issues when it comes to selecting the most appropriate methodology for your topic. Issues might include research limitations and ethical dilemmas that might impact the quality of your research.[2]

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 14-21
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

4. DATA ANALYSIS AND INTERPRETATION

Even with years of professional experience working with data, the term "data analysis" still sets off a panic button in my soul. And yes, when it comes to serious data analysis for your business, you'll eventually want data scientists on your side. But if you're just getting started, no panic attacks are required.[3]

```
Python 3.12.0 (tags/v3.12.0:0fb18b0, Oct  2 2023, 13:03:3
Type "help", "copyright", "credits" or "license" for more
>>> import pandas as pd
>>> import plotly.express as px
>>> import plotly.io as pio
>>> import plotly.graph_objects as go
>>> pio.templates.default = "plotly_white"
>>> data = pd.read_csv("E:/market_basket_dataset.csv")
>>> print(data.head())
   BillNo  Itemname  Quantity  Price  CustomerID
0    1000    Apples         5   8.30       52299
1    1000    Butter         4   6.06       11752
2    1000      Eggs         4   2.66       16415
3    1000  Potatoes         4   8.10       22889
4    1004   Oranges         2   7.26       52255
>>> _
```

Figure 1. Importing utilities & reading dataset.

Figure 1 above shows us steps to import common utilities in Python which would be required for our Data analysis.

```
>>> print(data.isnull().sum())
BillNo        0
Itemname      0
Quantity      0
Price         0
CustomerID    0
dtype: int64
```

Figure 2. Verification of the consistency of data.

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 14-21
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

Figure 2 above shows that we do not have any null data in our dataset which is primary requirement for any data analysis.

Further we go ahead to check for Summary Statistics of the dataset as shown below (Figure 3).

```
>>> print(data.describe())
             BillNo    Quantity       Price    CustomerID
count    500.000000  500.000000  500.000000    500.000000
mean    1247.442000    2.978000    5.617660  54229.800000
std      144.483097    1.426038    2.572919  25672.122585
min     1000.000000    1.000000    1.040000  10504.000000
25%     1120.000000    2.000000    3.570000  32823.500000
50%     1246.500000    3.000000    5.430000  53506.500000
75%     1370.000000    4.000000    7.920000  76644.250000
max     1497.000000    5.000000    9.940000  99162.000000
>>>
```

Figure 3. Statistics for the dataset.

Figure 3 above shows the Statistical results of dataset.

Now let us look at the pictorial representation Sales Distribution of the items as shown below:

```
>>> print(rules[['antecedents', 'consequents', 'support',
  antecedents consequents   support  confidence      lift
0    (Apples)     (Bread)  0.045752    0.280000  1.862609
1     (Bread)    (Apples)  0.045752    0.304348  1.862609
2    (Apples)    (Butter)  0.026144    0.160000  0.979200
3    (Butter)    (Apples)  0.026144    0.160000  0.979200
4    (Apples)    (Cereal)  0.019608    0.120000  0.592258
5    (Cereal)    (Apples)  0.019608    0.096774  0.592258
6    (Apples)    (Cheese)  0.039216    0.240000  1.311429
7    (Cheese)    (Apples)  0.039216    0.214286  1.311429
8    (Apples)   (Chicken)  0.032680    0.200000  1.530000
9   (Chicken)    (Apples)  0.032680    0.250000  1.530000
```

Figure 4. Sales Distribution.

Antecedents: These are the things that are thought of as the association rule's "if" or beginning point. Here, the antecedents in our analysis are Bread, Butter, Cheese, and Chicken. The entities or "item sets" produced from the data are called antecedents. To put it another way, it's the IF element on the left. In the situation before, bread serves as the antecedent.[4]

Consequent: These are the things that are thought of as the association rule's "if" or beginning point. Here, the antecedents in our analysis are Bread, Butter, Cheese, and Chicken. The term "consequent" refers to an item or group of items that are encountered along with the antecedent. The THEN part of the sentence is displayed on the right-hand side. The result in the aforementioned case is butter.[5]

Support: Support quantifies the frequency with which a specific collection of items—antecedents and consequents—occurs within the dataset. It speaks to the percentage of transactions where the items are anticipated to be purchased jointly. For example, the first rule indicates that Bread and Apples are bought together in approximately 4.58% of all transactions. Support refers to the frequency or occurrence of a specific combination of items in the dataset. Thus, indicates frequency of itemset appearing in the transactions being analyzed.[6]

Confidence: The probability that the subsequent item will be bought when the preceding item is already in the basket is measured by confidence. Alternatively, it displays the likelihood of purchasing the item that comes after when the item that came before is already in the basket. Figure above shows that there is a 30.43% chance of buying Apples when Bread is already kept in the basket after purchase. The probability that a transaction that contains the items on the left-hand side of the rule (in our example, pencil and paper) also contains the item on the right-hand side (a rubber). The higher the confidence, the greater the likelihood that the item on the right-hand side will be purchased or, in other words, the greater the return rate you can expect for a given rule.[7]

Lift: Lift measures the degree of association between the antecedent and consequent items, while considering the baseline purchase probability of the consequent item. If we find a lift with a value greater than 1 then this would indicate a positive association between the antecedent and the consequent item then it would indicate that the items are most likely to be bought together rather than independently. If we obtain a value which is less than 1 then it would indicate a negative association between the two. We can find a lift of 1.86 suggests a positive association between Bread and Apples. Lift is the measure of the effect of purchasing item A on purchasing item B. It is used to determine whether the combination of items has practical value. In other words, it is used to see if the combination of items is purchased more frequently than the individual items. If the value is greater than 1, it means that the combination is effective, while if it is less than 1, it means that it is ineffective[8].

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 14-21
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

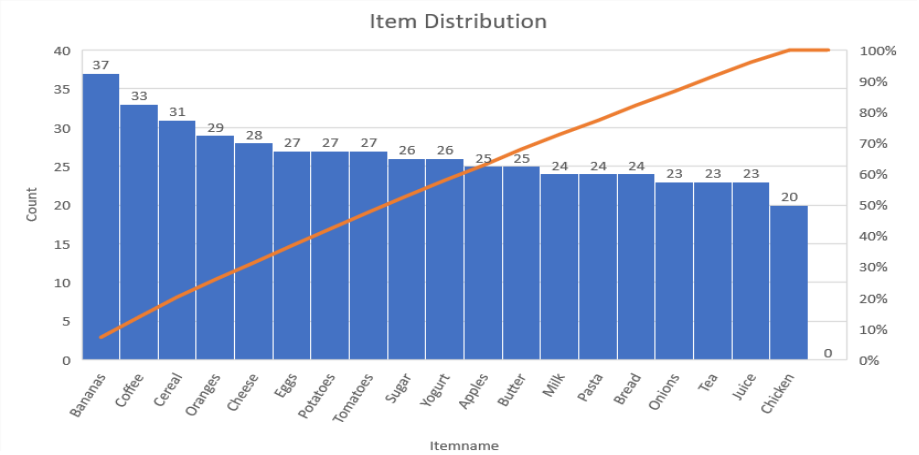

Figure 5. Item Distribution.

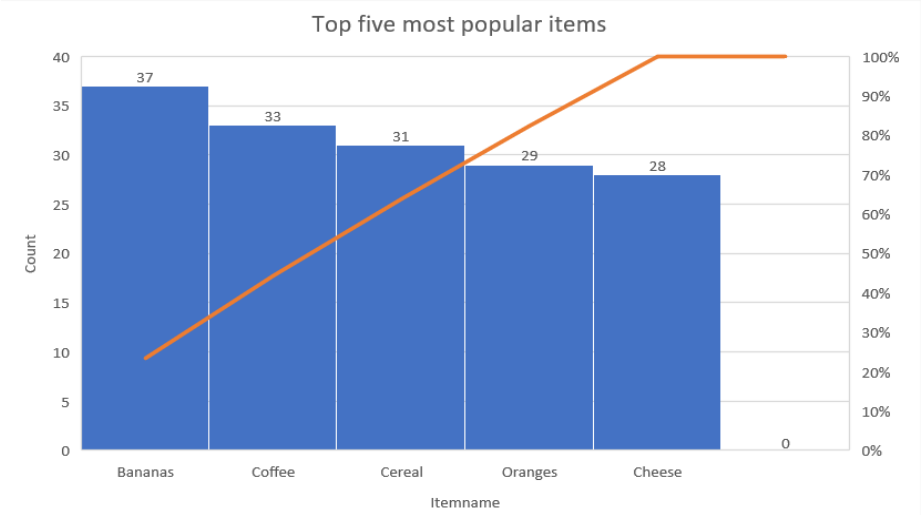

Figure 6. Top five most popular items.

It is observed that bananas are the most popular item sold in the store.

Understanding Customer Behavior

International Journal of Computer Science & Communications
Vol. 9, No. 1, 2024, pp. 14-21
http://ijcsc.ielas.org/index.php/ijcsc
ISSN: 2545-4153

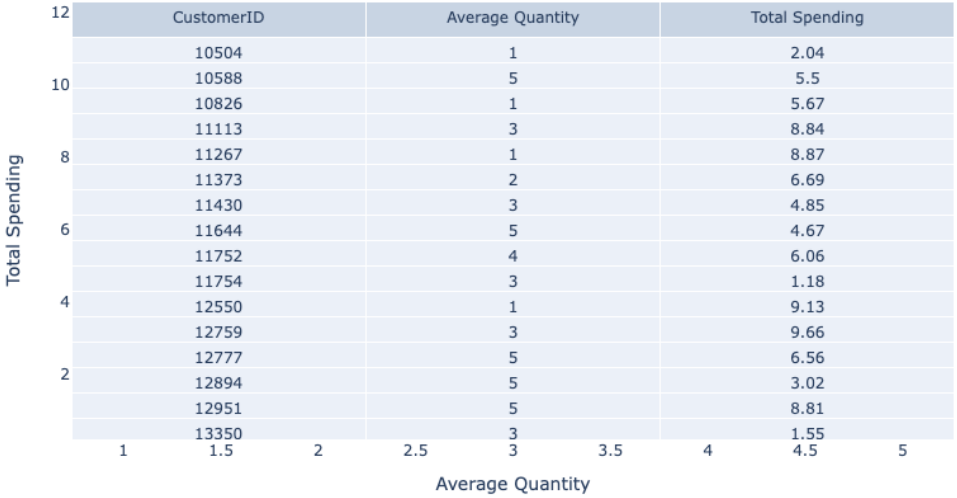

Figure 7. Understanding customer behavior.

By the term customer behavior, we understand the trends in the buying habits and factors which influence the decision to buy something else along with previous item. Here in Figure 7 above we explore customer behavior by comparing average quantity and total spending. Customer Behavior Analysis represents the study of how people make buying decisions concerning a product, service, and /or organization[9].

5. CONCLUSION, IMPLICATIONS, AND SCOPE FOR FUTURE RESEARCH

Henceforth it may be concluded that the historic data can be analyzed very effectively using Python language which is highly flexible and simple. This data analysis would be highly beneficial to end users in terms of decision-making in the future. They can very easily plan out their investment based upon the results that have been obtained with the help of this application. It would help them to have a better decision-making which would result in generating more profits. Since Market Basket Analysis is a highly productive tool to optimize the selling opportunities hence this project becomes utmost important. In the near future we would design a model wherein the predictions can be made beforehand. Artificial intelligence has revolutionized market basket analysis by automating the process of data analysis and rule discovery.[10]

REFERENCES

Chaudhary, S. (2022, February 11). Understanding market basket analysis in data mining. Hire the World's Most Deeply Vetted Developers & Teams Turing. https://www.turing.com/kb/market-basket-analysis/

Stevens, S. (2023, September 7). What is data analysis? Examples and how to get started. https://zapier.com/blog/data-analysis-example/

Simplilearn. (2022, November 22). *What is market basket analysis? Overview, uses, types, and examples* https://www.simplilearn.com/what-is-market-basket-analysis-article

Simplilearn. (2022, November 22). *What is market basket analysis? Overview, uses, types, and examples*. https://www.simplilearn.com/what-is-market-basket-analysis-article

How to use market basket analysis for retail and marketing. (2023, December 19). ThoughtSpot. https://www.thoughtspot.com/data-trends/analytics/market-basket-analysis

McColl, L. (2022, March 1). *Market basket analysis: Understanding customer behaviour*. Select Statistical Consultants. https://select-statistics.co.uk/blog/market-basket-analysis-understanding-customer-behaviour/

Overview of market basket analysis. (n.d.). Modern Big Data Analytics & BI Software - FineBI. https://intl.finebi.com/blog/market-basket-analysis

Predoiu, O. (2024, April 2). Customer Behavior analysis. Omniconvert Ecommerce Growth Blog. https://www.omniconvert.com/blog/customer-behavior-analysis/

Elnahla, N. (2021). Your Retailer Needs You: Retail lance and its Marketing Implications. https://doi.org/10.22215/etd/2021-14633

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to my adviser, Professor Mamta Bansal, for her invaluable guidance and support throughout this research. Her expertise and dedication have been a source of inspiration and motivation.