­

Customer Churn – Data Mining II – Task 3

Vincent Sanchez

College of Information Technology, Western Governors University

Dr. Kesselly Kamara

May 25, 2023

# Part I: Research Question

##### A1. Describe the purpose of this data mining report by doing the following: Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

Are we able to use Market Basket Analysis to identify 3 association rules to help determine product placement in our stores that could yield better sales?

##### A2. Describe the purpose of this data mining report by doing the following: Define one goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

Market Basket Analysis is “a powerful tool for translating vast amounts of customer transaction and viewing data into simple rules for product promotion and recommendation” (Hull, Market Basket Analysis in Python, 2023).

This analysis goal is to utilize the association rules generated from using Market Basket Analysis to determine product placement with the organization’s stores to promote higher sales of the associated items.

# Part II: Market Basket Justification

##### B1. Explain the reasons for using market basket analysis by doing the following: Explain how market basket analyzes the selected dataset. Include expected outcomes.

Market Basket Analysis utilizes the Apriori algorithm which “is a popular algorithm for extracting frequent item sets with applications in association rule learning” (Raschka, 2023). Association rules define the relationship between items in sets, creating a pseudo-like if x then y relationship. “The first is called the antecedent and the second is called the consequent. If, for instance, we find that purchasing fiction books is associated with purchasing biographies, as the transactions … indicate, then we state it as the following rule: ‘if fiction then biography’" (Hull, What is market basket analysis?, 2023).

These associated rules are then scored using the measures of support, confidence, and lift, which are explained in section D1. It is then tasked to the analyst to determine which measures are best to evaluate in the context of the goal. This information can then be used to determine which products are best associated with each other and help determine product placement along with marketing campaigns.

Using several Python packages in the ‘mlxtend’ library, the user can leverage code to prepare, process, and analyze transactional data. Expected outcomes from the analysis will include the support, lift, and confidence measures that will be used to determine the top 3 association rules that will be used to meet the goal of this project.

##### B2. Explain the reasons for using market basket analysis by doing the following: Provide one example of transactions in the dataset.

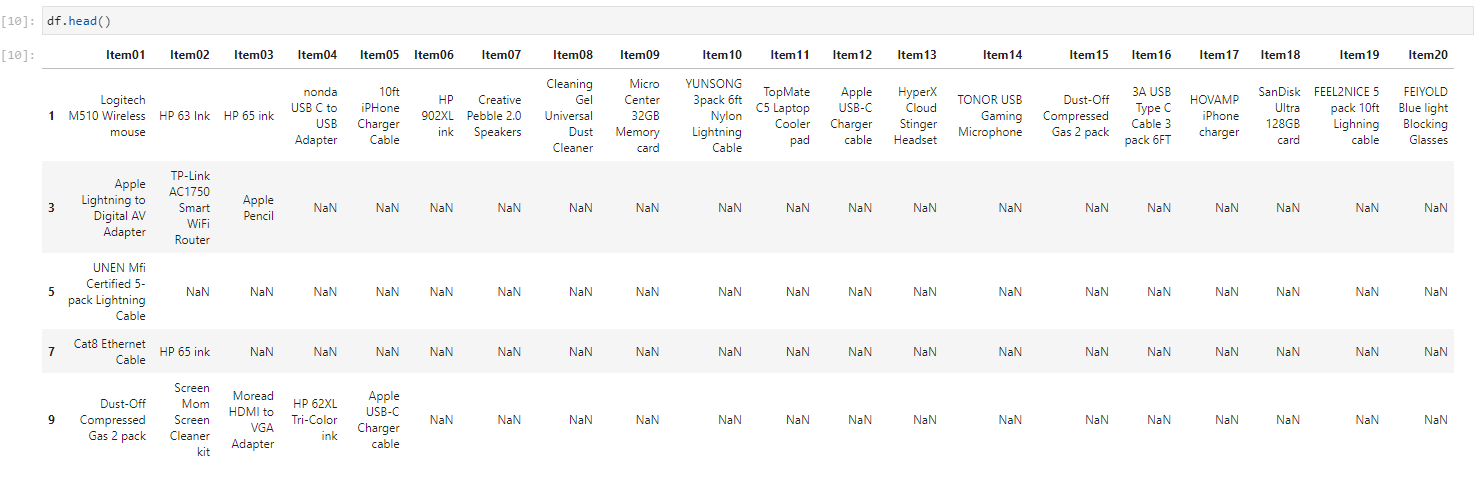


Figure 1 Transaction data set prior to encoding.

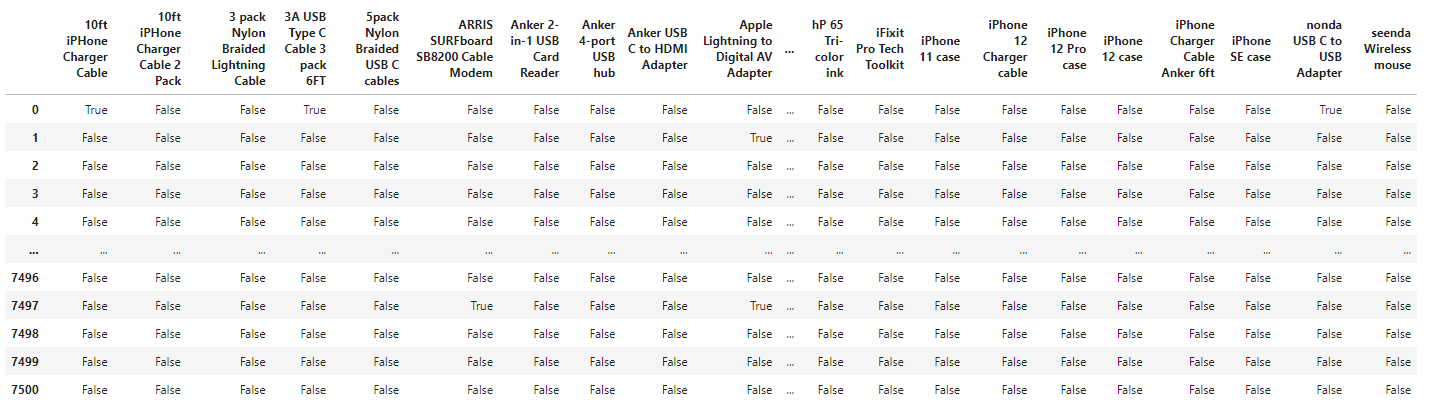


Figure 2 Transaction data after encoding.

##### B3. Explain the reasons for using market basket analysis by doing the following: Summarize one assumption of market basket analysis.

“Market basket analysis (MBA) is a data mining approach that identifies the relationship between products by examining previous transactions to identify patterns and trends. The approach relies on the assumption that customers who purchase a specific item are more likely to purchase another specific item or group of items” (Indeed Editorial Team, 2022).

Though, we can find support that if a customer purchases item x they are more likely to purchase item y alongside this purchase, the occurrence will not always be the case.

# Part III: Data Preparation and Analysis

##### C1. Prepare and perform market basket analysis by doing the following: Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.

Please refer to the comma-separated value file accompanying this submission labeled “teleco\_market\_basket\_clean.csv”.

##### C2. Prepare and perform market basket analysis by doing the following: Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

A screenshot of a computer

Description automatically generated with low confidence

Figure 3 Executed code of Apriori Function

A screenshot of a computer

Description automatically generated with low confidence

Figure 4 Executed code of association\_rules function.

A picture containing text, font, screenshot, line

Description automatically generated

Figure 5 Executed code of association\_rules with rule filtering.

##### C3. Prepare and perform market basket analysis by doing the following: Provide values for the support, lift, and confidence of the association rules table.

A picture containing text, screenshot, font, number

Description automatically generated

Figure 6 Support, Confidence, and Lift Values for association rules.

##### C4. Prepare and perform market basket analysis by doing the following: Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

A screenshot of a computer

Description automatically generated with low confidence

Figure 7 Support, Confidence, and Lift Values for filtered association rules.

The values used to filter the data were confidence and lift. The confidence value was set to any rule greater than .25 indicating that 25% percent of the time when item x was bought, item y was bought. The lift value was set to any rule greater than 1 indicating that the rule itself is completely independent of other rules.

# Part IV: Data Summary and Implications

##### D1. Summarize your data analysis by doing the following: Summarize the significance of support, lift, and confidence from the results of the analysis.

The support value is “the relative frequency that the rules show up. In many instances, you may want to look for high support in order to make sure it is a useful relationship” (Moffitt, 2017).

Confidence is the “measure of the reliability of the rule” (Moffitt, 2017). For example, if a confidence value of .5 is used for evaluation, it would mean that for 50% of the transactions where item x was bought, item y was also bought. “For product recommendation, a 50% confidence may be perfectly acceptable but in a medical situation, this level may not be high enough” (Moffitt, 2017).

Lift is “the ratio of the observed support to that expected if the two rules were independent. The basic rule of thumb is that a lift value close to 1 means the rules were completely independent. Lift values > 1 are generally more “interesting” and could be indicative of a useful rule pattern” (Moffitt, 2017).

##### D2. Summarize your data analysis by doing the following: Discuss the practical significance of the findings from the analysis.

For my analysis, I have chosen to look at association rules that have a confidence value greater than .25 and a lift value greater than 1.

A picture containing text, screenshot, font, line

Description automatically generated

Figure 8 Association rules filtered by values of both confidence and lift.

For the goal, the limit will be the top 3 association rules with the highest lift value. From the results, we can see when our customers buy the VIVO Dual LCD Monitor Desk mount typically, customers might also buy the Dust-Off compressed gas 2-pack. The next rule works with both items but with the compressed gas as the antecedent and the mount as the consequent. The last rule shows as customers buy the HP 61 ink cartridge; customers might also buy the Dust-Off compressed gas 2-pack.

##### D3. Summarize your data analysis by doing the following: Recommend a course of action for the real-world organizational situation from part A1 based on your results from part D1.

From analyzing the results of the Market Basket Analysis data mining process, we can use the data mining technique to identify 3 association rules to assist the organization in product placement. We found that the dual LCD monitor desk mount and the compressed gas (2-pack) typically get bought together. We also found that when our customers buy an HP 61 ink cartridge, they also tend to purchase compressed gas (2-pack).

My recommendation would be to place both the mount and the compressed gas close to each other as we see a high lift value in those association rules. I would also recommend placing the ink cartridge near the vicinity of the compressed gas.

If the organization is unsatisfied with these results, I would recommend perhaps extending the number of association rules further out and then experimenting with the parameters of the association rules to try to discover other patterns.

# References

Hull, I. (2023, May 25). *Market Basket Analysis in Python*. Retrieved from Data Camp: https://app.datacamp.com/learn/courses/market-basket-analysis-in-python

Hull, I. (2023, May 25). *What is market basket analysis?* Retrieved from Data Camp: https://campus.datacamp.com/courses/market-basket-analysis-in-python/introduction-to-market-basket-analysis-1?ex=1

Indeed Editorial Team. (2022, October 12). *FAQ: What Is Market Basket Analysis? (Types Plus Examples)*. Retrieved from Indeed: https://sg.indeed.com/career-advice/career-development/market-basket-analysis

Moffitt, C. (2017, July 3). *Introduction to Market Basket Analysis in Python*. Retrieved from Practical Business Python: https://pbpython.com/market-basket-analysis.html

Raschka, S. (2023, May 25). *apriori: Frequent itemsets via the Apriori algorithim*. Retrieved from mlxtend: http://rasbt.github.io/mlxtend/user\_guide/frequent\_patterns/apriori/