D212 Market Basket Analysis

The purpose of this Data Mining Report is to identify key relationships among variables in the dataset. One question that can translate to the real-world organization would be to find relationships among the variables that lead to identifying purchasing patterns that are currently happening and we are not aware of. One goal is to find these relationships to see how they can assist the company in turning some profit to present to the stakeholders.

Market Basket Analysis is very different compared to other models. Market Basket Analysis is dependent on the dataset having columns where all the values are either True or False or 1’s and 0’s. This is different than most models that allow features to have continuous or ordinal data. It works by looking for combinations of items that occur together frequently in transactions. To put it another way, it allows retailers to identify relationships between the items that people buy (Li, 2017). The expected outcomes consist of two outputs called Antecedents and Consequents. These are associated with “If this then that” scenarios or rules. One example of a transaction in the dataset would be: If a customer purchased a lightning cable, then they also purchased another lightning cable. One assumption of Market Basket Analysis is that the Apriori algorithm works with the assumption that any subset of a frequent itemset must be frequent (Vadakkanmarveettil, 2021).

The Apriori algorithm was able to process the data after it had been converted to Boolean values. Once the data was processed then the code was able to generate the association rules. Attached are screenshots of the code and the output of rules.

**Table

Description automatically generated**

Once the Apriori algorithm was able to run, specific parameters were applied to see what rules appeared to show the strongest relations. Attached is a screenshot for reference. The rules were filtered to have a support greater than .002, confidence greater than .85, and lift greater than 1.00. Once these filters were applied the top three rules that populated were if a person bought an Apple Lightning to Digital AV Adapter then they also bought a Cat8 Ethernet Cable, Mpow HC6 USB Headset then they bought the Apple Lightning to Digital AV Adapter, and if a person bought a Cat8 Ethernet Cable then they also bought an Anker USB C to HDMI Adapter.

Graphical user interface

Description automatically generated with medium confidence

In summary, the dataset provided very strong rules. Support had to be adjusted since the highest value was .009865 and it quickly lowered to .003 and .002. The significance of support was that it was positive, but the key findings were that confidence and lift show a strong relationship. Lift has a value greater 1. When lift has a value greater than 1 then this implies the relationship is not due to chance. Confidence tells us what the likelihood was that both the antecedent and the consequent would occur. The analysis shows that the confidence was above 1.

The practical significance of the findings would be that in all three rules the main components were either Cat8 Ethernet Cable or the Apple Lightning to Digital AV Adapter. Since these items are small, then a good course of action would be to put these two items next to each other at the store or within reach for the customer to take both at the same time.

# Bibliography

Li, S. (2017, 09 24). *A Gentle Introduction on Market Basket Analysis — Association Rules* . Retrieved from Towards Data Science: https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Vadakkanmarveettil, J. (2021, 04 05). *Market Basket Analysis – A Complete Overview (2021)* . Retrieved from jigsawacademy.com: https://www.jigsawacademy.com/market-basket-analysis/