Taylor Cina

D212 – Data Mining II

***Part I: Research Question***

*A.  Describe the purpose of your data mining report by doing the following:*

*1.  Propose****one****question relevant to a real-world organizational situation that you will answer using market basket analysis.*

What association rules, if any, exist in the customer dataset?

*2.  Define****one****goal of the data analysis. Ensure your goal is reasonable within the scope of the selected scenario and is represented in the available data.*

The goal of this analysis is to discover association rules in the customer dataset using market basket analysis in order to better understand purchasing habits among customers.

***Part II: Market Basket Justification***

*B.  Explain the reasons for using market basket analysis by doing the following:*

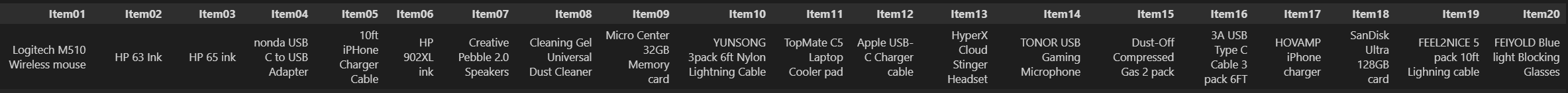
*1.  Explain how market basket analyzes the selected data set. Include expected outcomes.*

Market basket analyzes the dataset by taking a list of transactions and determining the relationship between transaction items in through different metrics. The outcome of this analysis is a list of association rules that describe the relationship of items in a transaction, such as how frequently those items appear in the dataset and how likely one or more items are to be part of a transaction given the purchase of another item.

*2.  Provide****one****example of transactions in the data set.*

Each row of the dataset represents one customer. The columns are a list of items purchased by that customer. These purchases are not necessarily all at one time.

The row below is an example of a transaction in the dataset:



*3.  Summarize****one****assumption of market basket analysis.*

Market basket analysis assumes that each transaction is an independent purchase on a specific date *(Deniran, 2023).* The dataset being aggregated over a period of time with purchases not being on specific dates could impact the efficacy of the analysis.

***Part III: Data Preparation and Analysis***

*C.  Prepare and perform market basket analysis by doing the following:*

*1.  Transform the data set to make it suitable for market basket analysis. Include a copy of the cleaned data set.*

Preprocessed dataset included as “TCina D212 T3 Onehot.csv”

*2.  Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate that the code is error free.*

The code used to run the Apriori algorithm and generate association rules is shown below:A screen shot of a computer code

Description automatically generated

*3.  Provide values for the support, lift, and confidence of the association rules table.*

The association rules table generated by the code above is displayed below:A screenshot of a computer screen

Description automatically generated

*4.  Explain the top****three****relevant rules generated by the Apriori algorithm. Include a screenshot of the top three relevant rules.* Excluding duplicate rules, the top three rules sorted by lift are as follows:

Apple Power Adapter Extension Cable -> HP 61 2 pack ink

Apple Power Adapter Extension Cable -> SanDisk Extreme 32GB 2pack card

Apple Power Adapter Extension Cable -> Apple Magic Mouse 2

The top three rules include the same item. Apple products are known for having proprietary cables and any users with Apple devices require adapter cables to use non-Apple products. The first rule is ink for an HP printer, the second rule is an SD card, and the third rule is another Apple product.

A screenshot of a black screen

Description automatically generated

***Part IV: Data Summary and Implications***

*D.  Summarize your data analysis by doing the following:*

*1.  Summarize the significance of support, lift, and confidence from the results of the analysis.*

Support is the prevalence of an item or transaction over the entire dataset. The support for the Apple Power Adapter Extension Cable is far lower than the support for any item it is paired with, indicating that it is an infrequent purchase. However, the lift value is very high for the items it’s purchased with, indicating that it is popular when paired with those items. Confidence is a measure of how frequent a transaction rule is in the pool of all transaction rules including the first item. The confidence of the rules starting with the extension cable are 0.20, indicating that there aren’t many transactions including the extension cable to begin with. Extremely high lift, low support and low confidence is indicative of the rule being supported by there only being a small number of purchases of the antecedent item. However, that antecedent item is highly associated with the items it is purchased with.

*2.  Discuss the practical significance of your findings from the analysis.*

At the very least, the top 3 rules show that the extension cable is purchased almost always with a few other specific items. Observing more rules or redefining the thresholds for the support cutoff could yield more useful rules, but the existing rule does give insight on the purchasing habits of a specific kind of customer. Namely, Apple users looking to buy other devices incompatible without an adapter cable.

*3.  Recommend a course of action for the real-world organizational situation from part A1 based on the results from part D1.*

The Apple Power Adapter Extension Cable should be placed with both other Apple peripherals as well as peripherals for other devices, such as printer ink cartridges and SD cards. Additionally, the thresholds for support could be restricted to observe only more common items to yield more generally applicable rules.

***Part V: Attachments***

*E.  Provide a Panopto video recording that includes the presenter and a vocalized demonstration showing all code used, the code being executed, and the results of all code used in the task.*

Panopto Video: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a0920c80-1292-4780-b4e1-b22a0047525e>

*F.  Record all web sources you used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.*

Deniran, O. (2023, November 27). *Boosting sales with data: The Power of Market Basket Analysis in retail*. Medium. https://medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df

*G.  Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.*

No external sources used.