**Part I: Research Question**

A.  Describe the purpose of this 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:

I will use the descriptive market basket analysis to identify which items are typically bought together by the customers.

2. 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.

The stakeholders can review the data provided by the analysis and create incentives for services that are bought together to keep the customers that are likely to terminate their contracts with the company. This will lead to a lower churn rate.

**Market Basket Justification**

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

1. Explain how the market basket analyzes the selected dataset. Include expected outcomes.

The Descriptive Market Basket is an unsupervised algorithm that gathers insight from the historical data by group items bought together to identify purchasing behaviour.

Expected outcomes: Item bought together will be grouped to present multiple customer profiles.

2.  Summarize **one** assumption of the market basket analysis.

The assumption of MBA is that joint occurrence of two or more products in most baskets suggest that these products are complements in purchase. Therefore, purchase of one will lead to purchase of others (Chantal D. Larose, & Daniel T. Larose. (2019)).

3. Provide one example of transactions in the dataset.

Graphical user interface

Description automatically generated with medium confidence

**Part III: Data Preparation and Analysis**

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

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

A. D212\_cleanDataSet.csv attached

2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

A. Graphical user interface, text, application, email

Description automatically generated

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

A. Table

Description automatically generated with medium confidence

4. Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

A.  
 Graphical user interface

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.

A. I have set the support to .02 to ensure that the algorithm captures items that appear in the transactions at least 2% of the time. After applying the filter of .02 on confidence, the customers who buy ‘VIVO Dual LCD monitor’ are 2.9 times likely to buy a ‘SanDisk Ultra 64GB card’ with the confidence of 22%.

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

A. This analysis provides stakeholders with insight to place items with high lift together for walk-in customer and provide suggestion during checkout or provide them as a combo as a sale item to increase sales.

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

A. VIVO Dual LCD monitor and SanDisk Ultra 64GB card should be placed together on the same shelf in the same isle. Additionally, advertise them as combo deals to all customers.

**Part V: Attachments**

E. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment. Panopto Link:

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

https://developers.google.com/. (2019, November 29). *Clustering Algorithms | Clustering in Machine Learning |*. Google Developers. Retrieved June 18, 2022, from https://developers.google.com/machine-learning/clustering/clustering-algorithms

Li, S. (2018, June 21). *A Gentle Introduction on Market Basket Analysis — Association Rules*. Medium. https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

GeeksforGeeks. (2018, November 28). *How to select multiple columns in a pandas dataframe*. https://www.geeksforgeeks.org/how-to-select-multiple-columns-in-a-pandas-dataframe/

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

Chantal D. Larose, & Daniel T. Larose. (2019). Data Science Using Python and R. Wiley.