**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 groups of items are most often purchased together?

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 the data analysis is to find out what items are most purchased together in the telco space?

**Part II: 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 data set. Include expected outcomes.

Market basket analysis analyzes data by finding associations and relationships between items purchased together during transactions. It dooes this by leveraging association rule mining techniques to uncover patterns in transaction data. It identifies which items are frequently purchased together and quantifies the strength of their relationships.In this instance I expect items from similar brands to be in similar baskets (ex: Apple products being bought together)

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

One example of a transaction shown in the table is: 10ft iPhone Charger Cable 2 Pack, Apple Lightning to USB cable, and the HP952XL ink. It is in row 17.

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

One assumption is **Transaction Independence**: It assumes that each transaction is independent of others, meaning the behavior of one customer does not directly influence another customer's transaction.

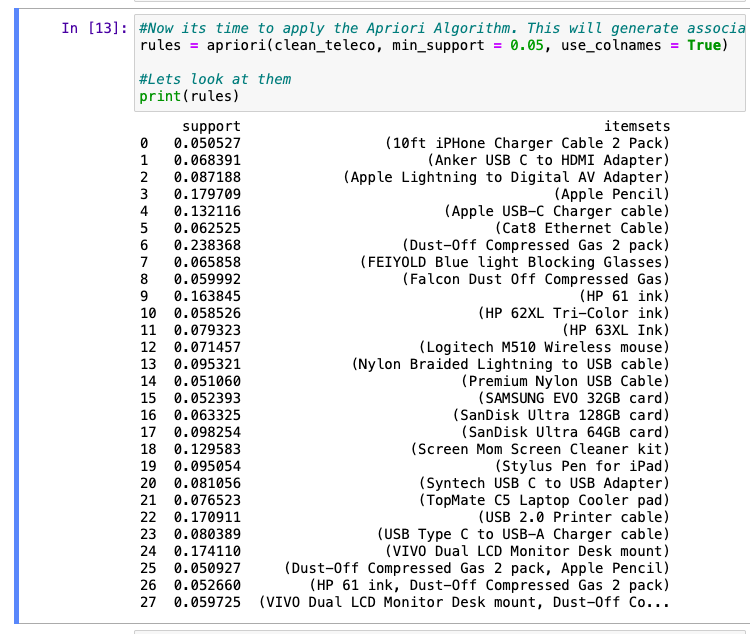
**Part III: Data Preparation and Analysis**

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.

\*Cleaned data attached with submission\*

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



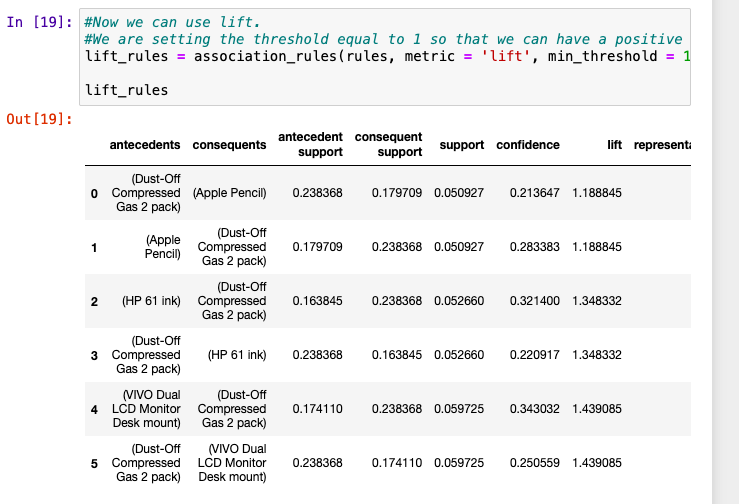
3. Provide values for the support, lift, and confidence of the association rules table.

\*It is shown in the attached notebook and screenshot below\*

4. Explain the top **three** relevant rules generated by the Apriori algorithm. Include a screenshot of the top three relevant rules.

**The top 3 relevant rules (ordered by the lift values). Please note they are listed in groups of two since they are inverse of one another. I listed the inverse pairs together:**

1. **(Lift Value of 1.4)** If a Dust-Off Compressed Gas 2 pack is bought, customers are also likely to buy a VIVO Dual LCD Monitor Desk mount. The inverse is also true and the only difference in the two when looking at the below table is the confidence. The support and lift are identical.
2. **(Lift Value of 1.3)** If HP 61 ink is bought, customers are also likely to buy a Dust-Off Compressed Gas 2 pack. The inverse is also true with the only difference being the confidence. The difference is a little larger than the above rule, however the support and lift are the same.
3. **(Lift Value of 1.2)** If a Dust-Off Compressed Gas 2 pack is bought, customers are also likely to buy an Apple Pencil. The inverse is also true and the only difference in the two when looking at the below table is the confidence. The support and lift are identical.



**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.

The **support** values are used to assess how commonly the rule applies across transactions.

Based on the results, all of the transactions and their inverse transactions happen around 5% of the time. This could indicate more of a niche purchasing behavior.

The **confidence** values are used to evaluate how reliable the relationship is.

Based on the results, this indicates more of a moderate association between the items. There is a group of people who do purchase these items together but not the majority

The **lift values** are used to identify whether the relationship is significant and actionable.

Based on the results, this indicates a high association. These items are purchased together more often than random chance.

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

What my findings tell me is that the Dust-Off Compressed Gas 2 pack is a very popular item that is bought alongside all of the other items listed in the table. If that is bought, its pretty likely at least one of the alternative items on the table are being bought alongside it

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

Based on the results of my analysis, I would suggest taking the Dust-Off Compressed Gas 2 pack and setting up some sort of bundle deal where if you buy that and one of the other items in the table, you can get some benefit or discount. This would take advantage of the clear relationship that we are seeing and increase the revenue that is received from selling these items together.

**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.

1. Include the presenter and a vocalized demonstration describing the programs used to complete this task in the Panopto video recording.

*Note: The audiovisual recording should feature you visibly presenting the material (i.e., not in voiceover or embedded video) and should simultaneously capture both you and your multimedia presentation.  
  
 Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the “WGU” option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto’s website.  
  
 To submit your recording, upload it to the Panopto drop box titled “Data Mining II – OFM4” Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.*

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.

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

H. Demonstrate professional communication in the content and presentation of your submission.