# D209 Data Mining 2

# Professor Keiona Middleton

# Mackenzie Simon

# Part 1 Research Question:

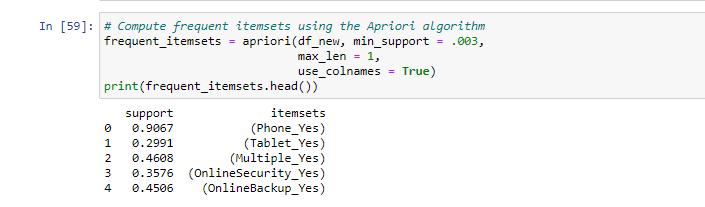
A1: For our project, we are provided customer data which gives us purchased product information. Using market basket analysis, we can predict which products will be most likely purchased after a customer buys a new phone. Through using market basket analysis, we can optimize store layout, increase sales, and hopefully retain customers longer through our premium products.

A2: The goal of our analysis is to use market basket analysis to predict which product a customer will purchase after buying a new phone.

# Part 2 Method Justification

B1. Market basket analysis works by analyzing the frequency in which products are bought together. These relationships are then used to build profiles containing If-Then rules of the items purchased. When multiple items are bought together this is known as a co-occurrence. Our analysis looks at the frequency of individual items in our data set. Frequent item sets determined by the Apriori algorithm can be used to determine association rules. Association rules consist of support, lift, and confidence.

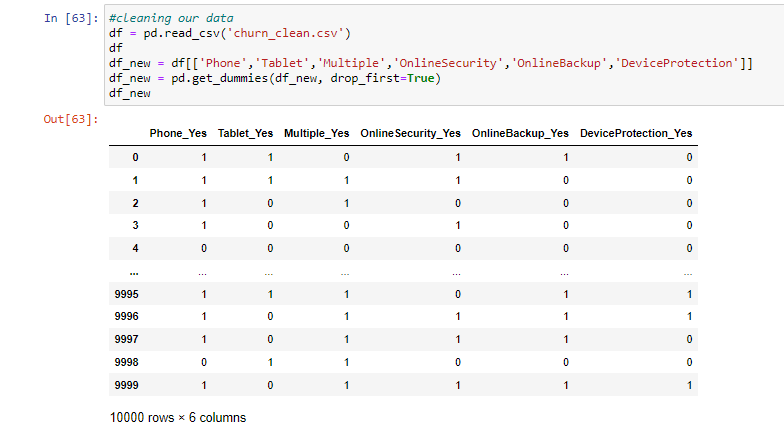
B2. One example of transaction we can use is support frequency. This allows us to see how often a phone is purchased out of all the sales. Based on the data, we can see a phone is purchased 90 percent of the time.



B3. One assumption of Market Basket Analysis is that the analysis is an unsupervised learning technique. We are not targeting a predictor variable. Jabeen (2018) describes market basket analysis as “an [unsupervised data](https://www.sciencedirect.com/topics/computer-science/unsupervised-data) science technique where there is no target variable to predict”.

# Part 3 Data Preparation and Analysis:

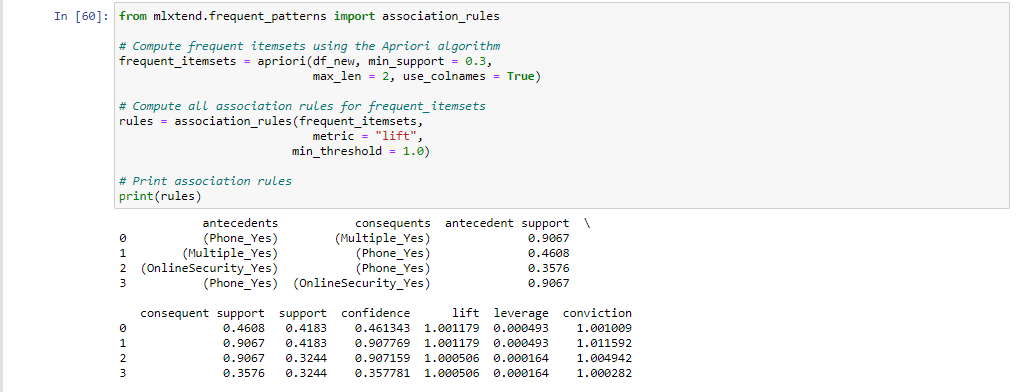
C1: In order to transform our data, we had to use dummy variables for each purchased product. For our model I focused on five key purchase items: phone, tablet, multiple phones, online security, online backup, and device protection. I dropped the columns that weren’t the five purchased items and transformed them with pandas.get\_dummies.



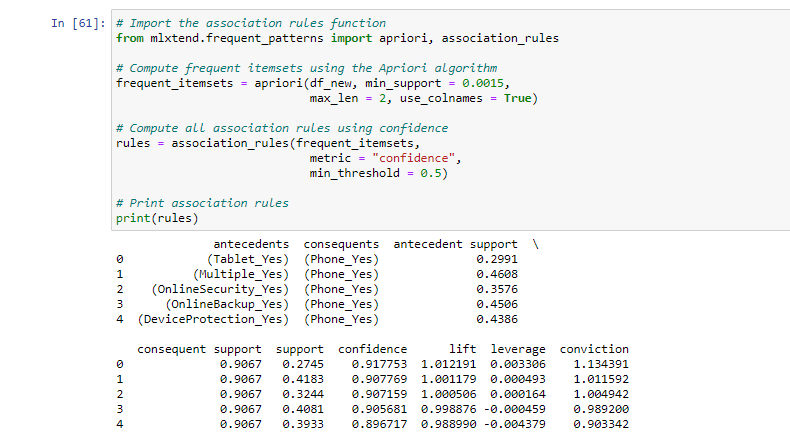
We used the Apriori algorithm to get our association rules. I ran the algorithm three times, each round emphasizing a new association characteristic of lift, confidence, and support. Ang (2016) describes how there are three common ways to measure association: support, confidence, and lift.

C2/C3:

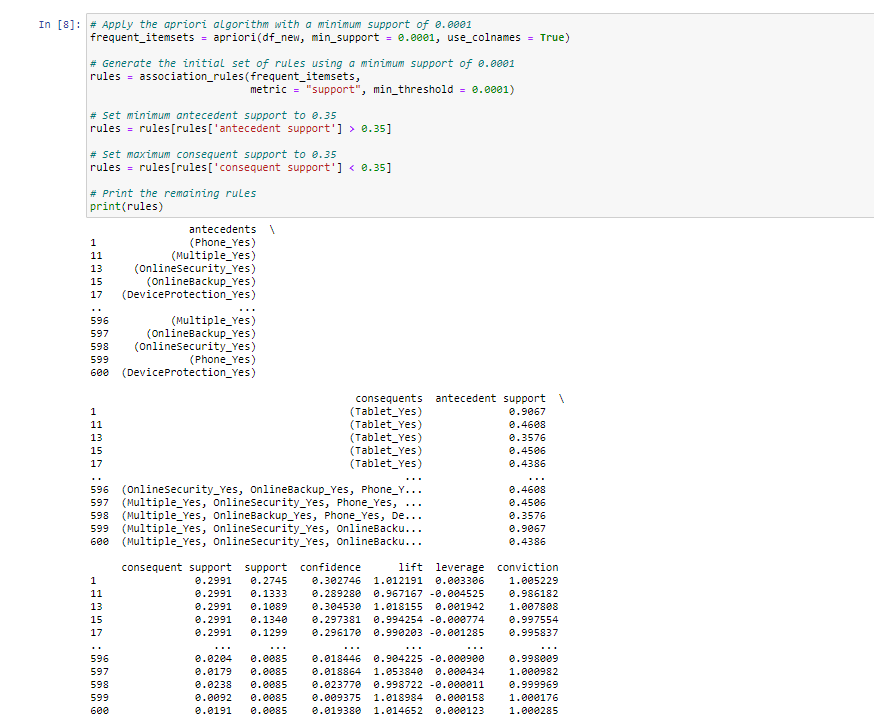
Emphasis on Lift:



Emphasis on Confidence:

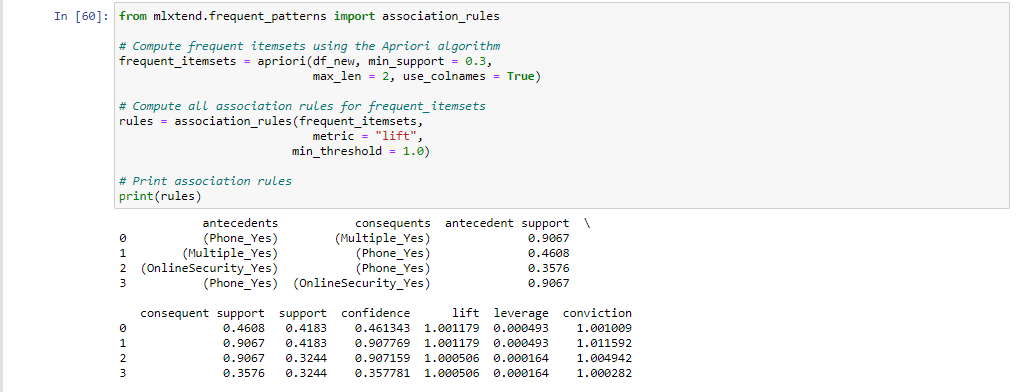


Emphasis on Support:



D4:

Based on my model with an emphasis on lift (listed below), the three top rules are multiple\_yes->phone\_yes with confidence .90, onlinesecurity\_yes->phone\_Yes with confidence of .90, and phone\_yes->multiple\_yes with confidence of .46.



# Part 4 Analysis:

D1: Overall, the emphasis for our model is to find purchases with phone\_yes as the antecedent with lift sitting over 1. We see high antecedent support for phone\_yes when paired with multiple\_yes and online security, both sitting at around 90 percent. However, the consequent for phone\_yes paired with multiple\_yes and online security are lower at .46788 and .3576. Our consequent confidence rests at .461 and .357 for each of these categories. All variables have a leverage under .001 and conviction over 1.

D2: Our model is not practically significant due to our low confidence intervals for consequent support. Essentially, we are seeing multiple phones being purchased 46 percent of the time after initially having one phone. For online security, we get to see a confidence of 35 percent of the time for when someone initially gets a phone. These aren’t very high and don’t support conclusive findings when phone\_yes is the antecedent.

D3: Based on our findings, I would continue to work on our algorithm in order to find better sales pairings with phone\_yes. I would ideally like to have a support confidence level over 90 percent when the antecedent is phone\_yes. However, we do see a trend of people purchasing multiple phones and online security after an initial phone at a low confidence. I would still suggest attempting to pair multiple phones when purchasing a new phone and continue offering online security plans with new phone plans even if the consequent support is low.

# References

Hafsa Jabeen. (2018, August 21). *R Market Basket Analysis using Apriori Examples*. DataCamp Community. https://www.datacamp.com/community/tutorials/market-basket-analysis-r.

Ng, A. (2016). *Association Rules and the Apriori Algorithm: A Tutorial*. KDnuggets. https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html.

**Resources for Python Libraries:**

https://matplotlib.org/

https://numpy.org/

<https://pandas.pydata.org/>

https://scikit-learn.org/stable/

https://seaborn.pydata.org/