**Association Rules**

Kelsey Thompson

Colorado State University Global

MIS 445: Statistics in Business Analytics

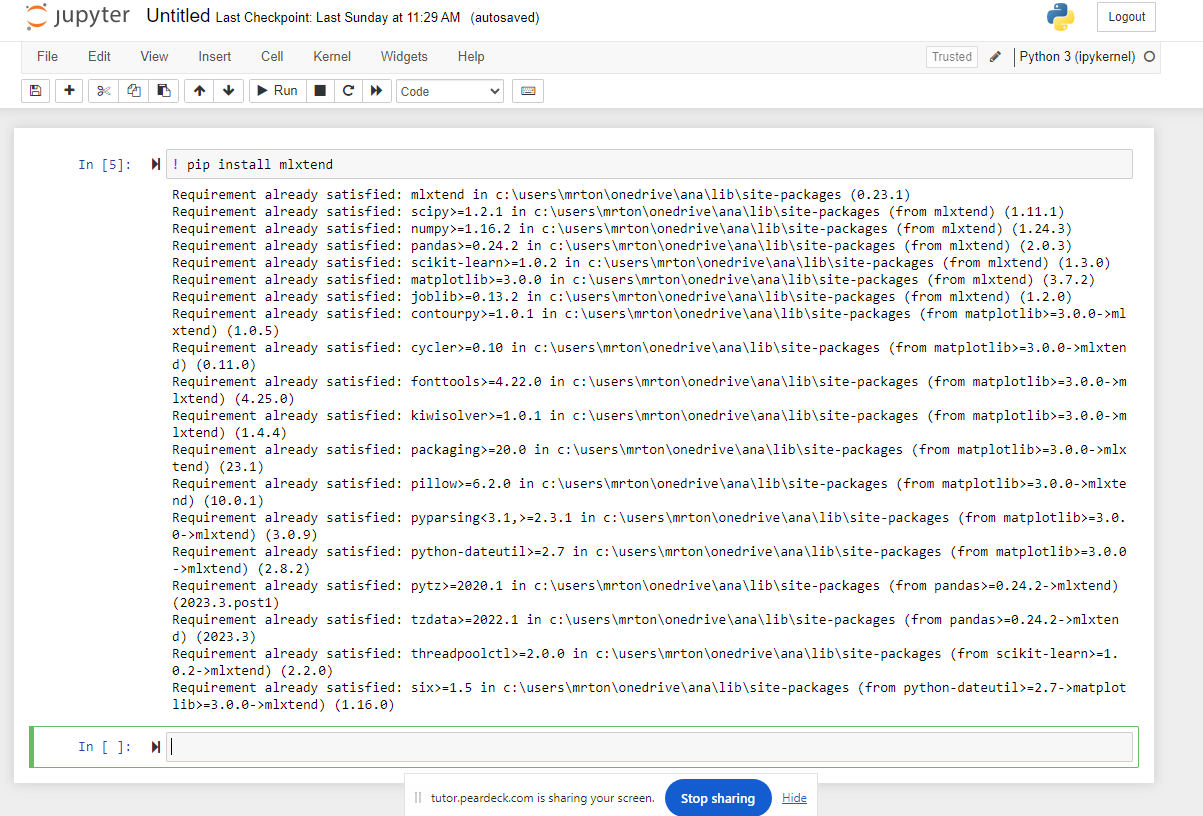
Dr. Mamdouh Babi

January 3, 2024 turned in

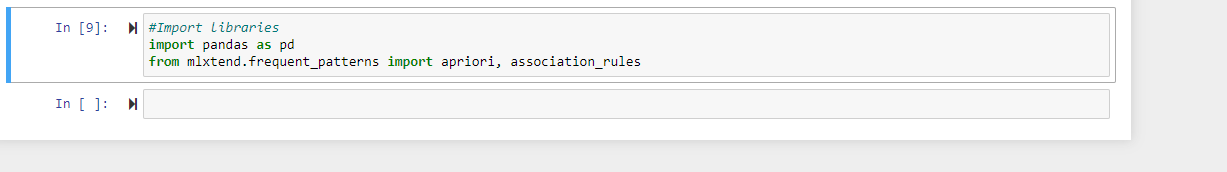
**Association Rules**

For this assignment, I utilized Jupyter Notebook to complete the association rule analysis. Using the subset, [Cosmetics dataset](https://csuglobal.instructure.com/courses/85277/files/5949330?wrap=1)[Download Cosmetics dataset](https://csuglobal.instructure.com/courses/85277/files/5949330/download?download_frd=1). This dataset consisted of cosmetic purchases at a large chain drugstore. I aim to analyze associations among purchases of these items for purposes of point-of-sale display, guidance to sales personnel in promoting cross sales, and guidance for piloting an eventual time-of-purchase electronic recommender system to boost cross sales.

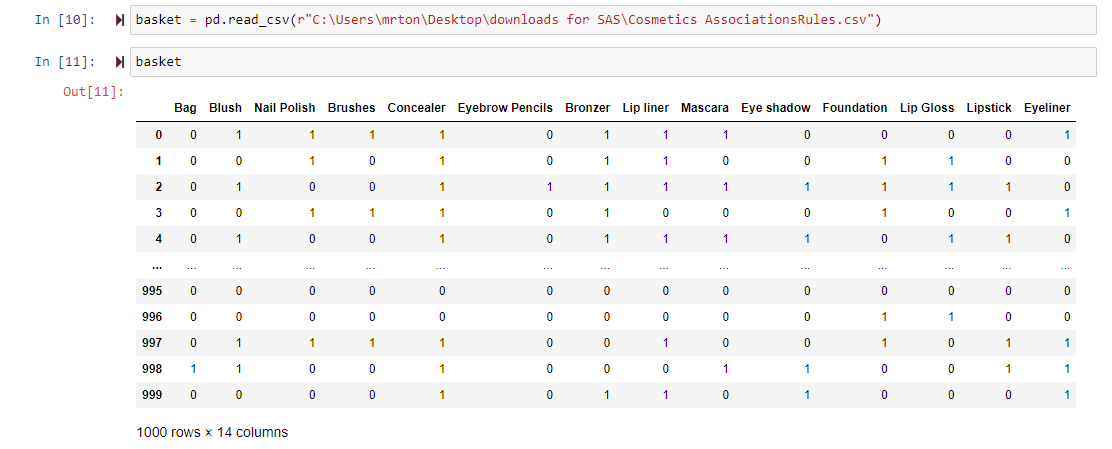
1. I followed along with the video and the instructions from the author. Provided below are the screenshots of the outputs from my code:

a) I downloaded Anaconda and Jupyter Notebook. I opened Jupyter and installed ‘mlxtend’ library.

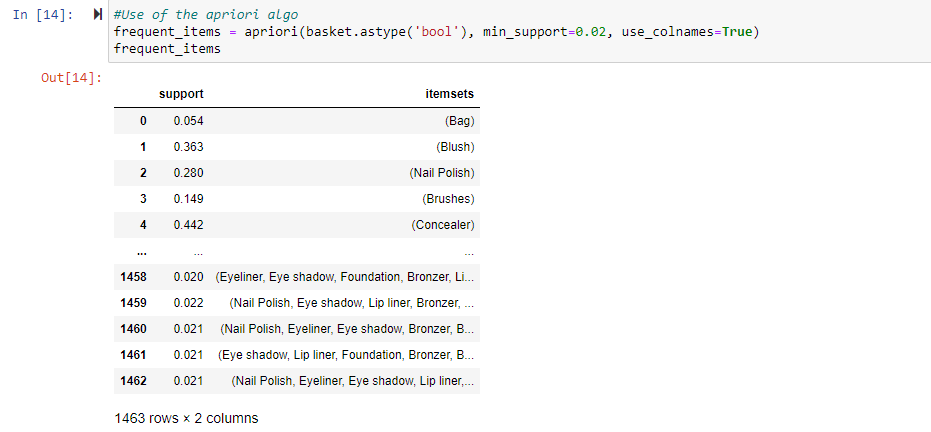
b) Next, I wrote code to make the apriori and association tools available.



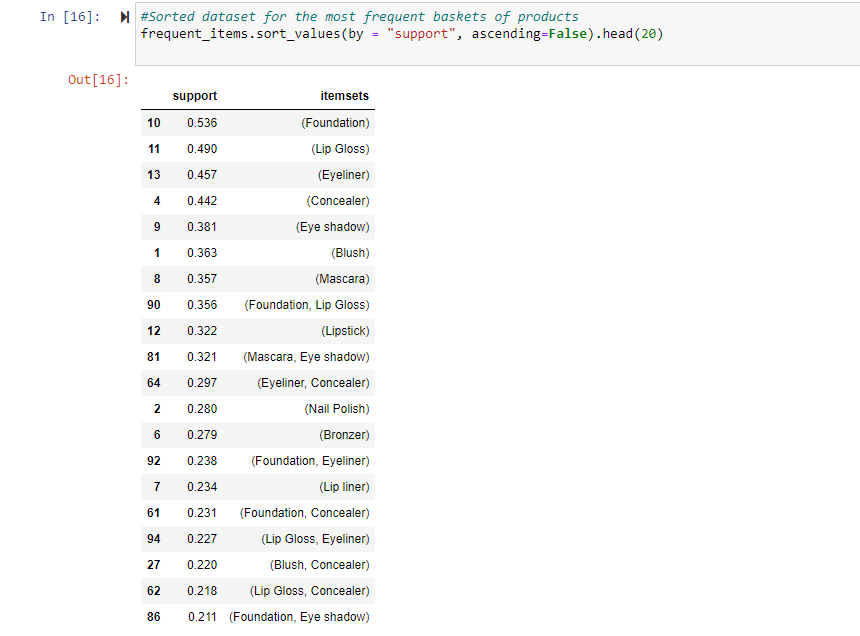
c) Then, I loaded my saved dataset and assigned it to the variable name, ‘basket,’ (indicated in this assignment):



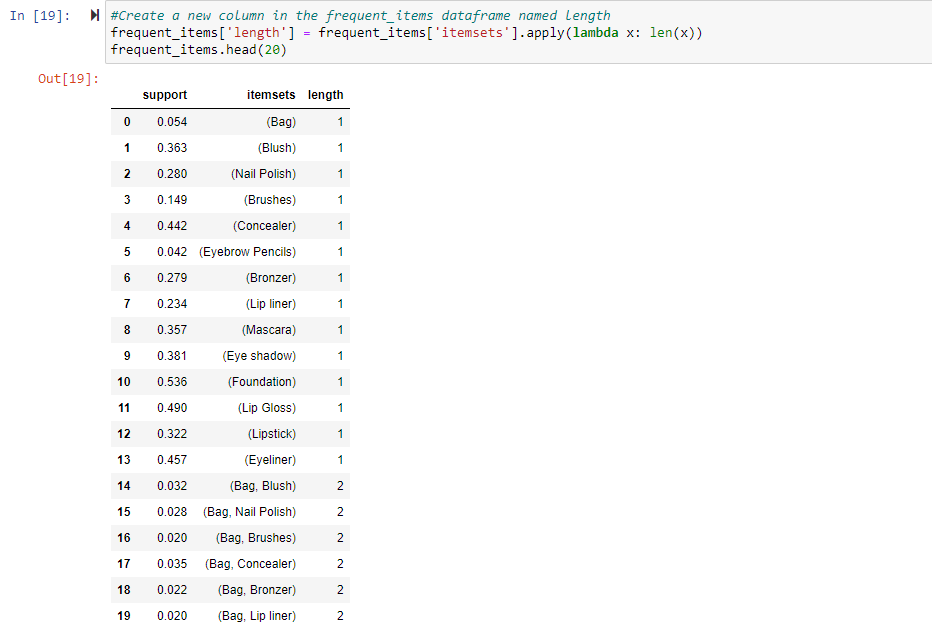
d) I run the apriority function with the minimum support set to .02, as instructed by this assignment. I also used ‘bool,’ as instructed. The data set now has 1463 rows of 2 columns.



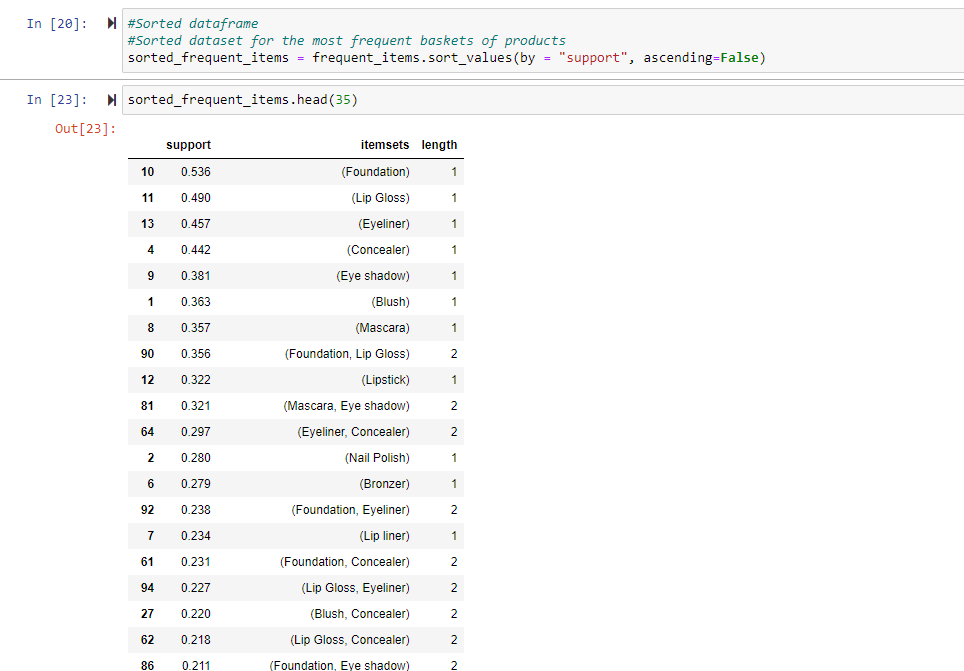
e) I assorted the frequent items table, using the support code and then I expanded it to 20



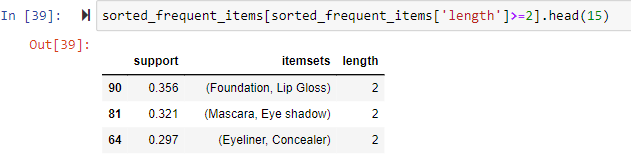
f) As in the video, I used the result from the apriori function and I created a new column, ‘length.’ This indicates how many items are in each basket. I tailored this to give me the 20 rows for viewing purposing.

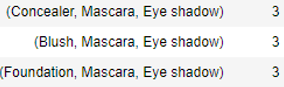


g) I created a new variable named ‘sorted frequent items.’ I may now can see the baskets that are more frequent. I tailored the table to show the top 35.

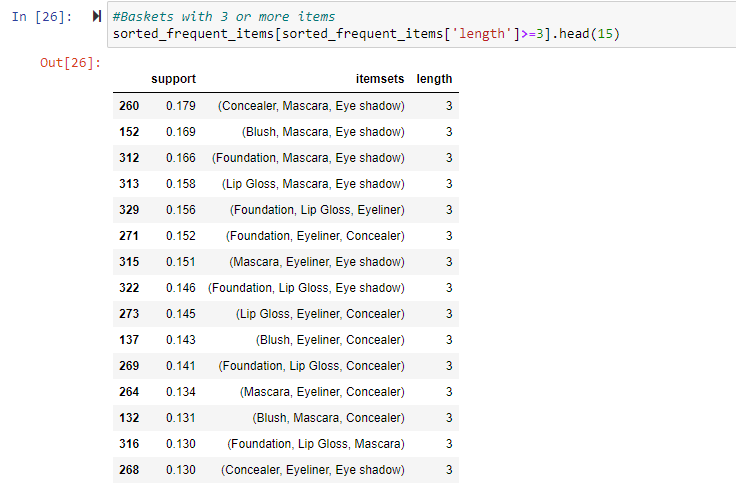


h) The following beauty items are frequently bought together:

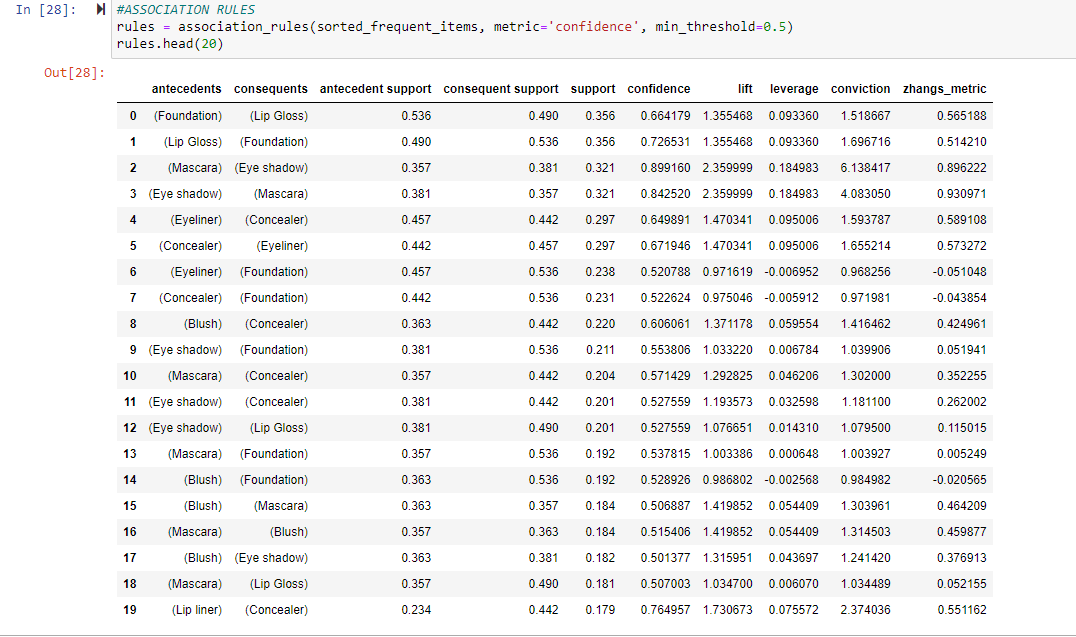
****

****

Now, I filtered the baskets to 3 or more items. The support numbers will be lower than 1 item or 2 item baskets:



1. Then, I made the association rules and analyzed the ‘confidence’ row to view the percentages for those variations:



j) Confidence:

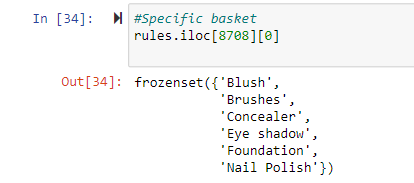
I sorted the table by the confidence to show the rules of the baskets that have higher probability. Based on the ‘antecedents,’ the ‘consequents’ will have a confidence probability (from the data):

Here are (3) rules from the result of the confidence analysis:

1. When Blush, brushes. Concealer, eye shadow, foundation, nail polish

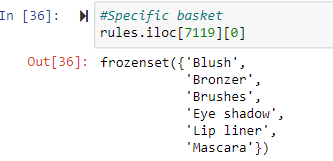
are all purchased together, the customer will also purchase mascara.





1. When eye shadow, lip liner, bronzer, blush and mascara are purchased together, there is a 100% probability that the customer will buy nail polish as well.

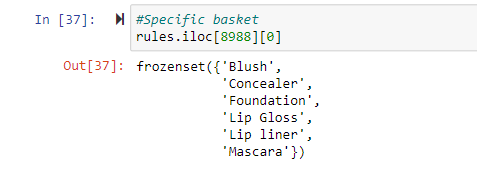




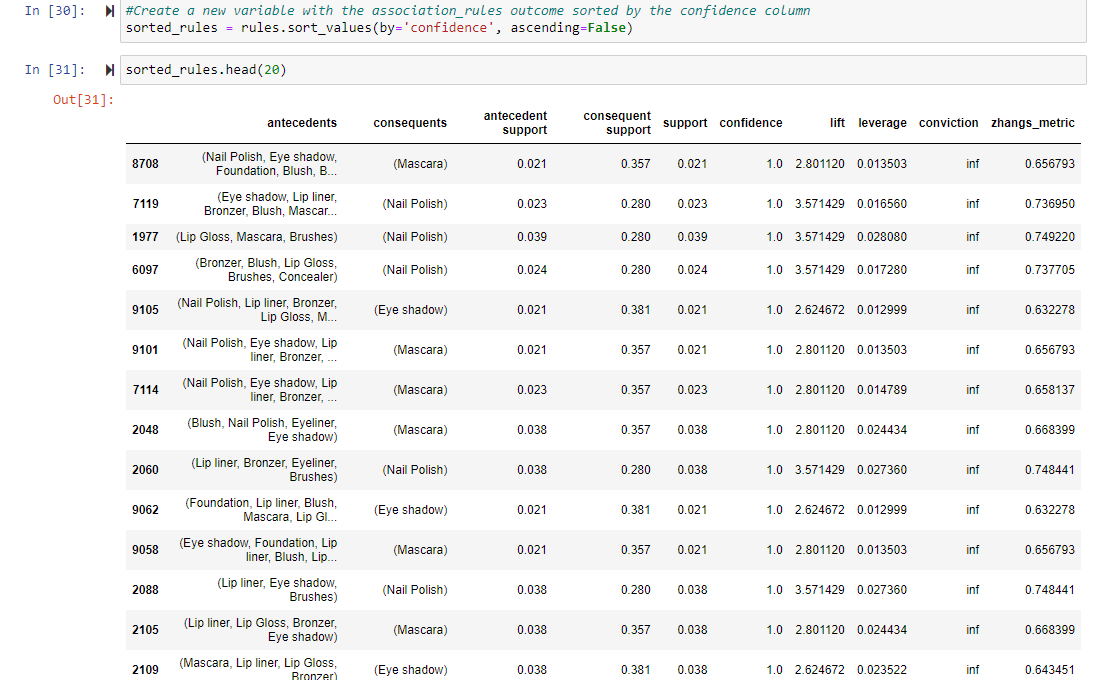
1. When foundation, lip liner, blush, mascara, lip gloss and concealer are purchased together, there is a 100% probability that the customer will also obtain bronzer

and eye shadow.





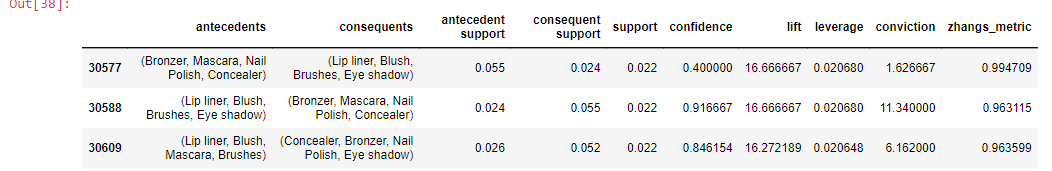
Here is the dataset’s first 20 lines for the highest confidence (step j)



k) I updated the metric (lift) and minimum threshold (set to 1):



Here are the (3) rules from the result of the lift analysis:

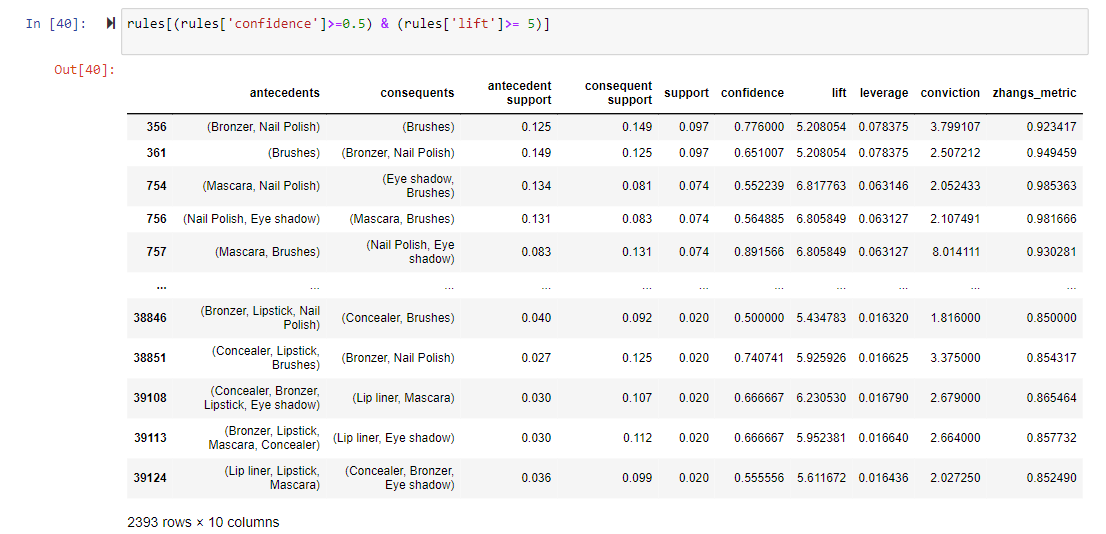


1. The lift (row 30577) is 16.666667. Meaning, this combination is 16.7 times more frequent than if the items were independent. Even though the support is low, the rule/frequency is also rare. The confidence is only .4. Resulting in this combination of antecedents. The consequents are purchased collectively, 40% of the time.

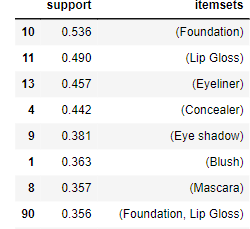
2. The lift (row 30588) is 16.666667. Consequently, this combination is 16.7 times more frequent than if the items were independent. Even though the support is low, the confidence is high. The confidence is 91.6667%. This combination of antecedent and consequents are purchased together 91.6667%. of the time.

3. The lift (row 30609) is 16.272189. This combination is 16.3 times more frequent than if the items were independent. Even through the support is low, the confidence is high (84.6154%). This combination of antecedent and consequents are purchased together 84.6154% of the time.

l) Lastly, I wished to see what had a confidence greater than or equal to 0.5, where the lift was greater than or equal to 5. Here are the results.



1. Proceeding are the provided takeaways on the analysis. These are recommendations for the drugstore to consider:
2. It's noteworthy that as the number of products purchased decreases, the confidence level tends to increase (reference table below). The most common occurrence is baskets with one or two items. Among these, foundation, lip gloss, and eyeliner emerge as the top-selling items (refer to the details below). It would be beneficial for the business to consider introducing variations of these popular products.

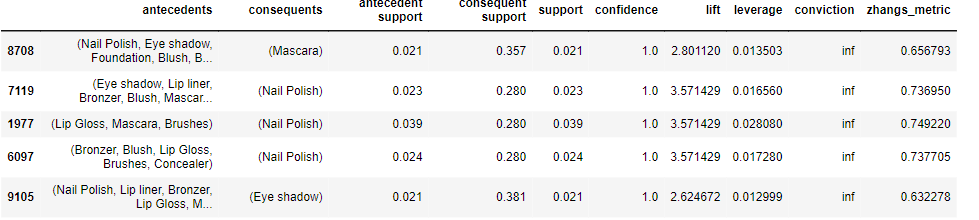


1. The combinations most seen purchased together are foundation and lip gloss, proceeding by mascara and eye shadow, then eyeliner and concealer. These products would do well to be placed near one another and/or in promotions.

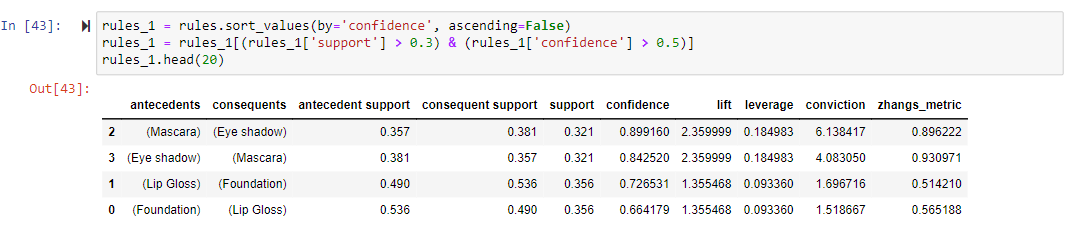


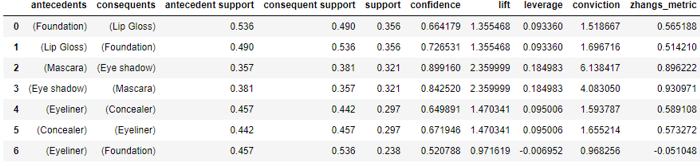
1. Also, the rules with the highest confidence are generally paired with a very low support. Signifying that specific combination is very rare. The confidence alone is not a strong explanatory variable.

Here is an example of a low support and a high confidence:



The store could tailor a filter for this data in order to visualize the antecedents with the highest support as well as the highest confidence. These rise as the best sellers to promote. I have provided this filter below:



1. Consider the association rules chart from step i. above. Here are the first 7 columns for reference. I would recommend that the store create makeup kits based upon which items are frequently bought together. On the other hand, they could focus their marketing on the best sellers/combinations: 

Personal reference

#Install the library

pip instal mlxtend

#Import libraries

import pandas as pd

from mlxtend.frequent\_patterns import apriori, association\_rules

#Import data

basket = pd.read\_csv(r"")

###########################

#APRIORI

#Use of the apriori algo

frequent\_items = apriori(basket.astype('bool'), min\_support=0.02, use\_colnames=True)

frequent\_items.head(20)

#Sorted dataset for the most frequent baskets of products

frequent\_items.sort\_values(by = "support", ascending=False)

#Create a new column in the frequent\_items dataframe named length

frequent\_items['length'] = frequent\_items['itemsets'].apply(lambda x: len(x))

frequent\_items

#Sorted dataframe

#Sorted dataset for the most frequent baskets of products

sorted\_frequent\_items = frequent\_items.sort\_values(by = "support", ascending=False)

sorted\_frequent\_items

#Baskets with 3 or more items

sorted\_frequent\_items[sorted\_frequent\_items['length']>=3].head(15)

#This new line

sorted\_frequent\_items[sorted\_frequent\_items['length']>=2].head(15)

#############################

#ASSOCIATION RULES

rules = association\_rules(sorted\_frequent\_items, metric='confidence', min\_threshold=0.5)

rules.head(20)

#Create a new variable with the association\_rules outcome sorted by the confidence column

sorted\_rules = rules.sort\_values(by='confidence', ascending=False)

sorted\_rules.head(20)

#Specific basket

rules.iloc[8708][0]

#We use Lift instead of the confidence

rules = association\_rules(sorted\_frequent\_items, metric='lift', min\_threshold=1)

rules.head(20)

#Create a new variable with the association\_rules outcome sorted by the confidence column

sorted\_rules = rules.sort\_values(by='lift', ascending=False)

sorted\_rules.head(20)

rules[(rules['confidence']>=0.5) & (rules['lift']>= 5)]

#create filter with highest confidence/support

rules\_1 = rules.sort\_values(by='confidence', ascending=False)

rules\_1 = rules\_1[(rules\_1['support'] > 0.3) & (rules\_1['confidence'] > 0.5)]

rules\_1.head(20)