**D212 Performance Assessment Task 3**

**ASSOCIATION RULES AND LIFT ANALYSIS FOR MEDICAL DATA**

Fahim A. Akbar Student ID 001434895 Masters Data Analytics (January 1, 2021) Program Mentor: Mandy Rasmuson (385) 428-5957x5957 [fakbar3@wgu.edu](mailto:fakbar3@wgu.edu)

**Part I: Research Question**

**A. Data Mining Report Justification**

**A1. Research Question:** Which medications are concurrently being prescribed to and purchased by patients who are using Cialis?

**A2. Research Goal:** The goal of this data analysis is to find out which medications listed in the data sheet are prescribed and purchased alongside a target medicine, Cialis. Cialis is a drug used for treating erectile dysfunction (ED) and symptoms of benign prostatic hyperplasia (BPH). While this drug is doctor prescribed and sometimes used as a longer lasting alternative to Viagra, it can have minor and even severe side effects (Taguri, 2022). The less severe side effects include headache, indigestion, muscle aches, and back pain. More serious side effects can include sudden vision loss or changes in hearing (FDA, 2015). Using market basket analysis to discover the prescriptions that co-occur can help determine what the most common drugs are being prescribed with Cialis. If taking Cialis with another drug induces a side-effect, we can see if the two drugs are co-prescribed frequently. From there, preventative measures can be taken such as pharmaceutical companies disclosing the risks of taking both medications at the same time.

**Part II: Method Justification**

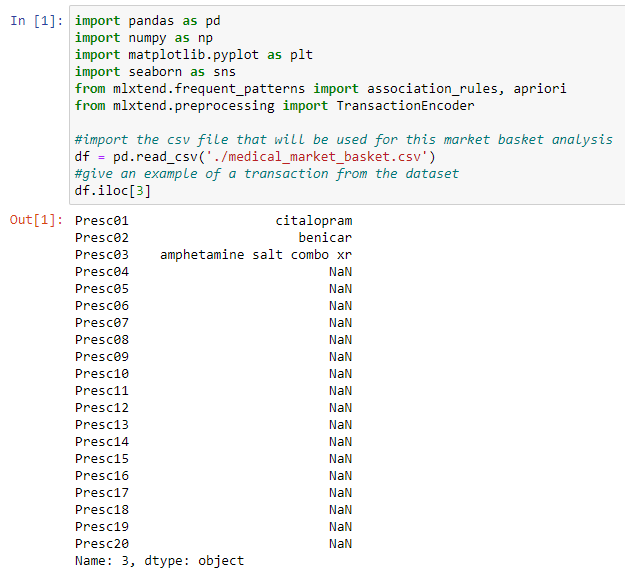
**B. Market Basket Justification**

**B1.** **Justification of using Market Basket Analyzes and what the expected outcomes are**

Market basket analysis is a data analysis technique used to identify relationships and patterns among items frequently purchased together by customers. This technique works by analyzing customer transactions and then using this data to see if there are any associations and dependencies amongst their purchases. Business can leverage the results of market basket analysis to make data-driven decisions for product placement, cross-selling, and promotional strategies. This overall helps with optimizing sales and improving customer satisfaction. (Sivek, 2020). For the medical dataset, our focus is to examine if specific medications are bought and prescribed in conjunction with Cialis. An anticipated outcome of this analysis is the generation of rules involving Cialis either as a condition or a consequence of concurrently prescribed medications.

**B2. Transaction example.**

Shown below is an example of a single transaction from the dataset.

****

The dataset can allow up to 20 different items/prescriptions (Presc01 - Presc20). The transaction demonstrated above shows 3 prescriptions: *benicar, citalopram, and amphetamine salt combo xr.* In the dataset, every row represents a distinct transaction involving the purchase of at least 1 and up to 20 prescriptions. The market basket analysis examines the relationships between the prescriptions bought together to create associations among them.

**B3. Assumptions of Market Basket Analysis**

Market basket analysis assumes that customers who buy a specific product have a higher chance of buying another specific product or group of products. By analyzing a dataset of substantial size, the existence of more persistent relationships among products can be identified.

Market basket analysis can be used to extract these relationships even when they are not readily evident. Furthermore, as the dataset encompasses all transactions, it can be utilized to track the positioning of items within a retail location. Retail stores can use this knowledge and determine whether different placements yield distinct impacts (Chaudhary, 2023). An example of this could be allergy medication sold at a drug store. Market basket analysis could show us that allergy relief products such as nasal spray has a positive correlation with allergy medications, and the two are often purchased together. The drug store could use this to strategically place these products closer together and improve the customer experience.

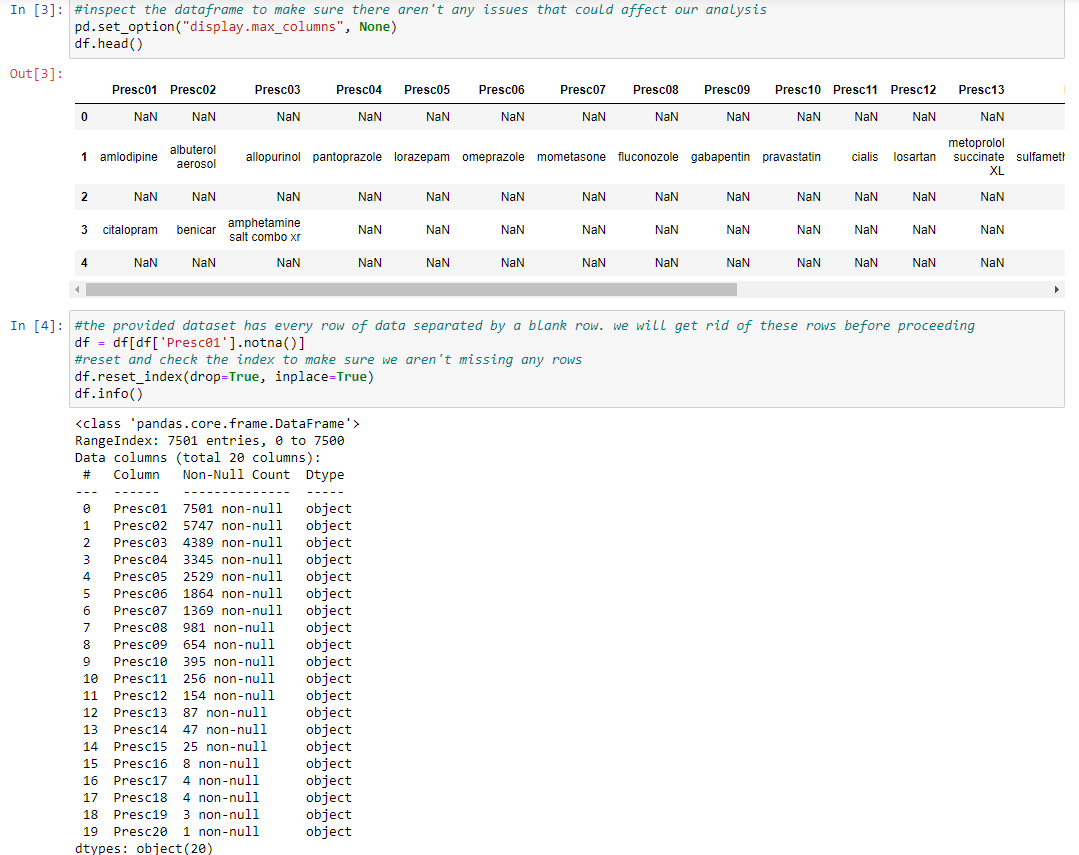
**Part III: Data Preparation and Analysis**

**C. Data Preparation**

**C1. Dataset Transformation**

With the data already loaded into a dataframe, it needs cleaned up into a format that will allow market basket analysis to be performed. The dataset is loaded into a dataframe and cleaned. It will then be formatted accordingly so market basket analysis can be conducted. This dataset lists prescriptions by name rather than by the type of medications. Since we won’t be able to compare medications by type or groups, the data will thus be prepared and analyzed by comparing medications by name against each other. After loading the data and inspecting the dataframe, it was apparent that the provided dataset has every row of data separated by a blank row. These rows were subsequently removed. Afterwards, a big list of lists was created to store the data. Once the list on lists was made, the transaction encoder was added to the list of lists, and then the data was changed and stored in a temporary array. The preparation of the dataset was performed in a Python using a Jupyter notebook environment. The Jupyter notebook file is attached to the task submission. A pdf copy of the notebook and a txt. file of code used is provided with the task submission as well. A copy of the cleaned dataset is also provided with the task submission. Lastly, the entire code used is also provided at the end of the document.

The code for preparing the dataset is showcased below:



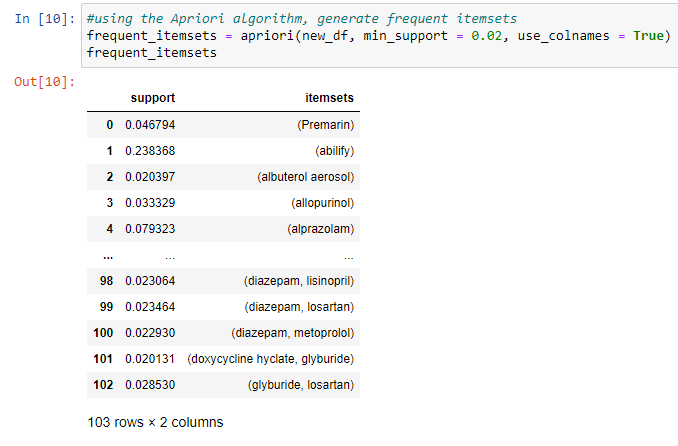


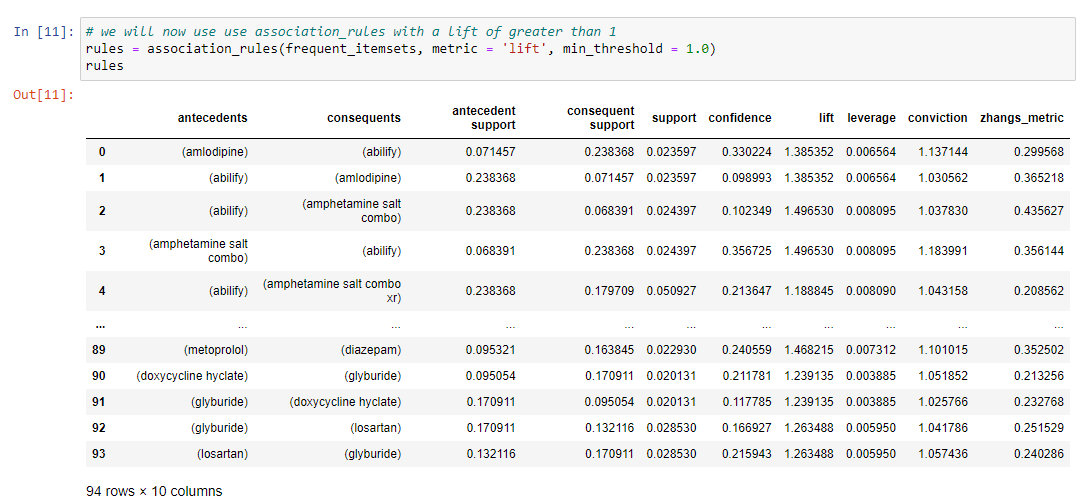
Now each medicine prescription in the overall dataset is represented as a column, with a total of 119 columns. The transactions are represented as 7501 rows, and each transaction is a list of Trues (for if the customer got the medication) and Falses (for if the customer did not get the medication).

**C2: Generation of Association Rules**

To generate the association rules, we use two functions. First, the Apriori algorithm is used to pull all associations that meet a minimum support threshold provided. For the minimum support measurement, we’ll be using 0.02 for the frequent\_itemsets table. The support score of a specific rule indicates the proportion of transactions that contain that rule. Since the dataset we’re using is not large, a lower support value could cause issues with small sample sizes. Likewise, a higher support value might be excessively exclusive, capturing only the most obvious rules in the analysis. The threshold of 0.02 is reasonable as it requires the rule to have occurred at least 151 times out of the 7501 transactions in the dataset.

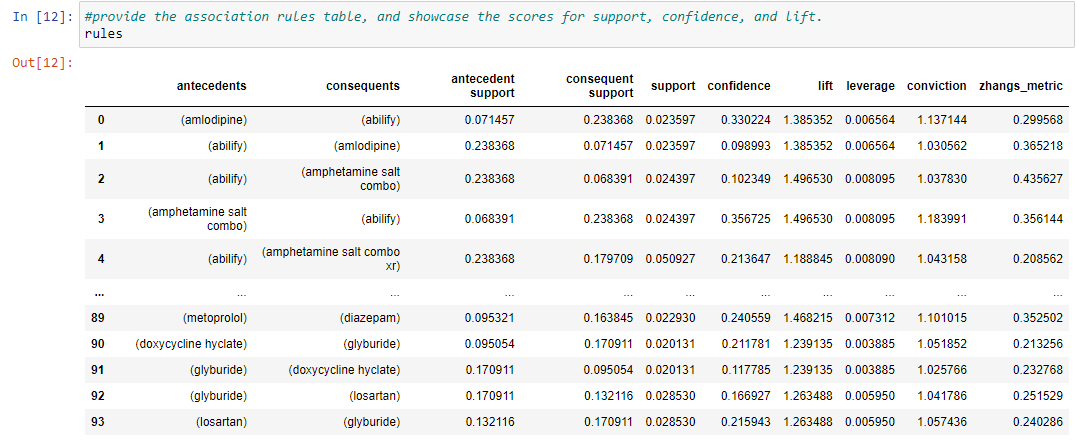
The association\_rules function is then used to expand and refine the rules. Here we incorporate an additional minimum threshold based on the lift value of each rule. Setting the minimum lift to 1.0 ensures that we only consider rules where the antecedent significantly influenced the occurrence of the consequent in the transaction.





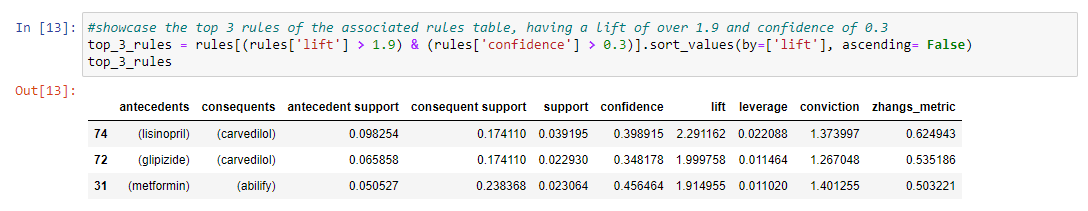
**C3: Association Rules Table**

An association rules table for the dataset was generated with the support, confidence, and lift. It is showcased below:



**C4: Top Rules**

The top three rules in the associated rules table are shown below. They have a lift of over 1.9 and confidence of 0.3.

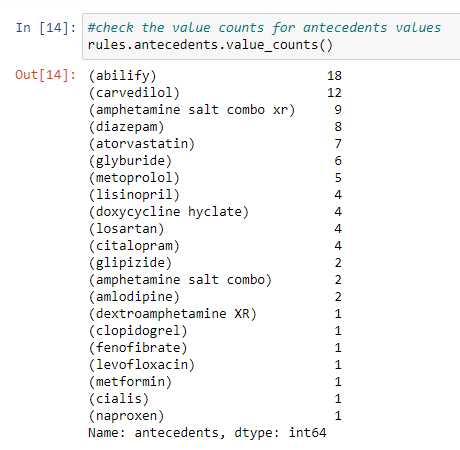


The higher lift mean that the consequent has a stronger effect being included in the transaction with the antecedent. The confidence score suggests that a proportion of all transactions, including the antecedent, featuring the consequent. Using a moderate confidence threshold requires a higher occurrence of the consequent within the antecedent transactions.

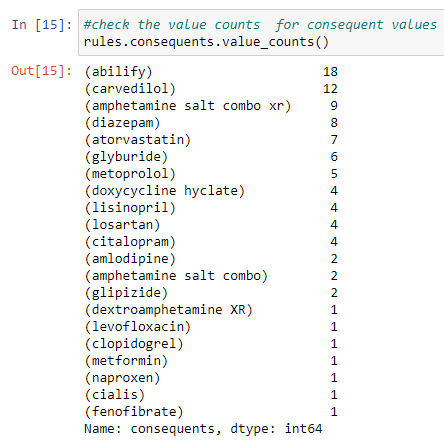
**Part IV: Data Summary and Implications**

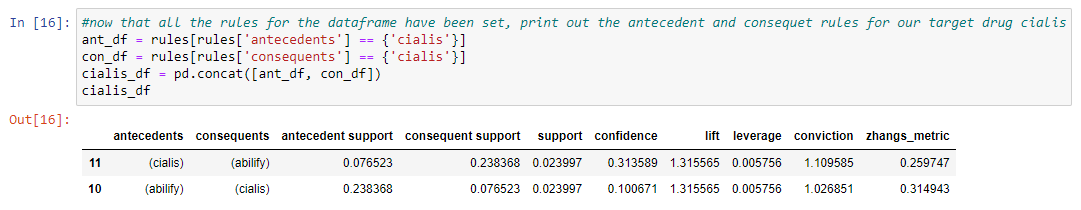
**D1: Results of Analysis**

Now that the final set of rules has been made, we will check the value counts of both the antecedents and the consequents and see if Cialis is present. The antecedent value counts are shown below.



The consequent value counts are shown below.





Two rules are generated; one indicating that "If Cialis, then Abilify" and the other indicating "If Abilify, then Cialis." Provided is an overview of what the support, confidence, and lift scores mean in relation to the above table (Chaudhary, 2023). Support is the proportion of transactions containing the rule of the total transactions in the dataset. This dataset shows that out of the 7501 transactions, approximately 2.4% contain both Abilify and Cialis. Confidence is the proportion of all transactions which include the rule over the proportion of transactions containing just the antecedent for the rule. Our analysis showed that the Cialis -> Abilify rule had a much higher confidence of about 3x more than the Abilify -> Cialis rule. The issue of which direction this rule should go (which medication is the antecedent) can largely be determined by the confidence level, which is much higher for Cialis as the antecedent, rather than the consequent. Lift is the ratio of the confidence percent to the support percent. In this study it measures the extent to which the combination of Abilify and Cialis in a transaction exceeds the expected frequency. If the lift value was less than 1, it would imply that the Abilify and Cialis combination is not bought frequently. If the lift value was equal to 1, than the purchase of one drug makes no difference on the other. Since the lift was greater than 1, it shows us that the Abilify and Cialis combination is bought frequently by customers. This proves that a positive relationship exists between the two drugs, and that the antecedent is increasing the likelihood of the consequent occurring in the transaction. Both the lift and support scores are the same for both directions that the rules could exist. Cialis is a sexual performance enhancing drug used to treat erectile dysfunction. On the other hand, Abilify is a drug used to treat several mood disorders (WebMD, 2023). Since mood disorders can apply to a broader group of patients and not just men suffering from sexual impotence, a larger use for Abilify is reasonable and expected. Furthermore, sexual impotence could be related to mood disorders, and a patient prescribed Cialis may also be prescribed drugs like Abilify to address the mood related issues. (Rosen, 2004) Taking this all into consideration, the correct rule would be "If Cialis, then Abilify". This means that if someone is purchasing or prescribed Cialis, then they are likely to purchase or be prescribed Abilify as well.

**D2: Practical Significance**

The relationship between Cialis and Abilify appears to have some practical significance. This dataset had a large amount of Cialis prescriptions, with Cialis purchases included in over 7.5% of all transactions. Since erectile dysfunction in older men is a persistent problem, it’s understandable that repeat purchases could occur increasing the amount of Cialis transactions. (WebMD, 2023). Since erectile dysfunction is caused by different factors, a prescription for Abilify isn’t always needed alongside Cialis. However, alleviating mood related issues should still be taken into account when treating erectile dysfunction.

**D3: Recommended Action**

The results of our market basket analysis shows that a relationship is present between Cialis and Abilify. We can conclude that people who need a Cialis prescription are likely to need a Abilify prescription as well. While Cialis is prescribed for erectile dysfunction and sexual enhancement for men, Abilify is a drug used to treat certain mental and mood disorders. These include bipolar disorder, schizophrenia, Tourette's syndrome, and irritability associated with autism. (WebMD, 2023). It’s possible that the co-prescriptions could be due to mood disorders along with the sexual performance issues. This relationship between the two drugs is a strong indication that patients who use Cialis may also be experiencing mood related issues. One recommendation is to include the consideration of mood related issues when prescribing Cialis, and especially when prescribing both Cialis and Abilify. This would be especially useful if no other medical causes are found for the erectile dysfunction. Furthermore, studies should be conducted by monitoring patients who use both drugs and see if there is an overall improvement in both their sexual performance and their mood. These studies can also be used to see what the side effects of taking both drugs are over a large population size.

**Part V: Attachments**

**E: Panopto Recording**

A link to the Panopto recording is provided below, and is also attached in the final submission:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=55ff9ce0-8871-4d6f-ab17-b02c0021f36b>

**F. Third Party Code Used**

<https://rasbt.github.io/mlxtend/user_guide/preprocessing/TransactionEncoder/>

<https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/>

<https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/>

<https://stackoverflow.com/questions/68020768/divide-rangelendataframecolumn-and-not-the-values>

<https://stackoverflow.com/questions/63921134/add-row-number-to-column-string-in-pandas>

<https://www.kaggle.com/code/khusheekapoor/market-basket-analysis-in-python>

<https://github.com/rasbt/mlxtend/blob/master/mlxtend/frequent_patterns/association_rules.py>

<https://www.kaggle.com/code/bbhatt001/bakery-business-model-association-rules>

<https://www.kaggle.com/code/khusheekapoor/market-basket-analysis-in-python>

<https://www.kaggle.com/code/yugagrawal95/market-basket-analysis-apriori-in-python/comments>

**G. Additional Sources**

Sivek, S. C. (2020, November 16). Market Basket Analysis 101: Key Concepts. Towards Data Science. Retrieved June 6, 2023, from <https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00>

Chaudhary, S. (2023). Market Basket Analysis. Retrieved June 6, 2023, from <https://www.turing.com/kb/market-basket-analysis>

WebMD. (2023). Cialis oral: Uses, side effects, interactions, pictures, warnings & dosing. Retrieved June 6, 2023, from <https://www.webmd.com/drugs/2/drug-77881/cialis-oral/details>

WebMD. (2023). Abilify oral: Uses, side effects, interactions, pictures, warnings & dosing. Retrieved June 6, 2023, from https://www.webmd.com/drugs/2/drug-64439/abilify-oral/details

<https://www.bloomingbiz.marketing/blog/market-basket-analysis-for-prescription-data>??

Taguri, G. (2022). Cialis vs Viagra: Which One is Better for You? Lloyd's Pharmacy Online Doctor. Retrieved June 6, 2023, from <https://onlinedoctor.lloydspharmacy.com/uk/erectile-dysfunction/cialis-vs-viagra#:~:text=The%20biggest%20difference%20between%20Viagra,hours%20after%20taking%20a%20tablet>.

U.S. Food and Drug Administration. (2015, August 13). Questions and Answers: Cialis (tadalafil). Retrieved June 6, 2023, from <https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/questions-and-answers-cialis-tadalafil>

Rosen, R C. (2004, August 16). Quality of life, mood, and sexual function: a path analytic model of treatment effects in men with erectile dysfunction and depressive symptoms. PubMed. Retrieved June 6, 2023, from <https://pubmed.ncbi.nlm.nih.gov/14961048/>

**Full code used for the project:**

**DATA CLEANING AND PREPARATION CODE**

#import packages and clean data before running the market basket analysis

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

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

from mlxtend.preprocessing import TransactionEncoder

#import the csv file that will be used for this market basket analysis

df = pd.read\_csv('./medical\_market\_basket.csv')

#give an example of a transaction from the dataset

df.iloc[3]

df = pd.read\_csv (r'C:\Users\fahim\Documents\0\_WGUDocuments\d208\1medical\_clean.csv')

df.head()

df.info()

#inspect the dataframe to make sure there aren't any issues that could affect our analysis

pd.set\_option("display.max\_columns", None)

df.head()

#the provided dataset has every row of data separated by a blank row. we will get rid of these rows before proceeding

df = df[df['Presc01'].notna()]

#reset and check the index to make sure we aren't missing any rows

df.reset\_index(drop=True, inplace=True)

df.info()

#check to make sure all the empty rows are removed

df.head()

**CODE USED FOR CREATING THE LIST OF LISTS**

#store the data in a big list of lists

temp\_big\_list = []

#iterate through each and within each row, iterate through each column

for row\_number in range(len(df)):

# Generate a temporary small list for each row

temp\_small\_list = []

for cell in range(len(df.columns)):

# check to make sure there are no null values in the cells

if not pd.isnull(df.iloc[row\_number, cell]):

#if cell contents are not null, add a string version of that cell's contents to the temporary small list

temp\_small\_list.append(str(df.values[row\_number, cell]))

#for the list of lists, add the small list to the ongoing big lists

temp\_big\_list.append(temp\_small\_list)

#print the temp\_big\_list to make sure it looks the way we want it to

print(f"list of lists... \nindex 0: {temp\_big\_list[0]}\nindex 1: {temp\_big\_list[1]}\n...\nindex7500: {temp\_big\_list[7500]}")

**CODE USED FOR CREATING A TRANSACTION ENCODER**

#create a transaction encoder

encoder = TransactionEncoder()

#add the transaction encoder to the list of lists, and then change and store the data in a temporary array

temp\_array = encoder.fit(temp\_big\_list).transform(temp\_big\_list)

#generate a new dataframe from this temporary array

new\_df = pd.DataFrame(temp\_array, columns=encoder.columns\_)

#check the new dataframe to make sure that it looks the way we want it to

new\_df

**CODE USED FOR CHECKING AND EXPORTING THE NEW DATAFRAME**

#print information about this new dataframe

new\_df.info()

#now that the new dataframe has been created, export it as a csv file

new\_df.to\_csv(r'C:\Users\fahim\Documents\0\_WGUDocuments\d212\task3\_marketbasket\_clean.csv', index=False)

**CODE USED FOR GENERATING FREQUENT ITEMSETS**

#using the Apriori algorithm, generate frequent itemsets

frequent\_itemsets = apriori(new\_df, min\_support = 0.02, use\_colnames = True)

frequent\_itemsets

**CODE USED FOR ASSOCIATION RULES**

# we will now use use association\_rules with a lift of greater than 1

rules = association\_rules(frequent\_itemsets, metric = 'lift', min\_threshold = 1.0)

rules

#provide the association rules table, and showcase the scores for support, confidence, and lift.

rules

#showcase the top 3 rules of the associated rules table, having a lift of over 1.9 and confidence of 0.3

top\_3\_rules = rules[(rules['lift'] > 1.9) & (rules['confidence'] > 0.3)].sort\_values(by=['lift'], ascending= False)

top\_3\_rules

**CODE USED FOR CHECKING ANTECDENT AND CONSEQUENT VALUES**

#check the value counts for antecedents values

rules.antecedents.value\_counts()

#check the value counts for consequent values

rules.consequents.value\_counts()

**CODE USED TO PRINT THE ANTECEDENT AND CONSEQUENT VALUES FOR THE TARGET DRUG**

#now that all the rules for the dataframe have been set, print out the antecedent and consequents rules for our target drug cialis

ant\_df = rules[rules['antecedents'] == {'cialis'}]

con\_df = rules[rules['consequents'] == {'cialis'}]

cialis\_df = pd.concat([ant\_df, con\_df])

cialis\_df