# OFM3 — OFM3 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS

**Data Mining II — D212**

**PRFA — OFM3**

Jason Smith  
Student ID: 293963  
[jsmi268@wgu.edu](mailto:jsmi268@wgu.edu)  
WGU – MSDA  
Program Mentor: Christiana Okhipo

**Part I: Research Question**

Are there relationships of purchased prescriptions? Are there tendencies that patients whom buy one prescription are likely to buy another prescription?

My goal is to learn the purchasing tendencies of patients to determine which products they are likely to buy based on previous purchase history.

**Part II: Market Basket Justification**

This market basket analysis would help determine the chances of a patient to purchase a medication if they previously purchased another medication. I used the following 3 metrics to measure these associations which are Support, Confidence, and Lift (Maitra, 2019).

* Support – contains details about frequently bought items
* Confidence – indicates how often items are bought together.
* Lift – “indicates the strength of a rule over random over random occurrence” (Maitra, 2019).

I expect applying the apriori algorithm to the data set will help identify purchasing relationships.

As one example, I visually identified a purchasing relationship between “bilify” and “paraxeline”. These two drugs are sometimes purchased together.

One assumption of market basket analysis is multiple medications are often prescribed for the same condition. Market basket analysis will help identify these relationships.

**Part III: Data Preparation and Analysis**

Refer to section ‘C1’ of the attached code for the data transformation steps. The cleaned and transformed datafile, ‘'mark\_data\_clean\_d212\_task3.txt’ is in CSV format and is attached as well

**Here is the code used to generate association rules with the Apriori algorithm.**

*#Imports*

*# Pandas - supports ability to create dataframes and multi-dimensional arrays*

**import** pandas **as** pd

*# Numpy - Used to create error curves which helped to determine an ideal K-value. Used in section D2*

**import** numpy **as** np

*# Data visualization*

*#import seaborn as sns*

**import** matplotlib.pyplot **as** plt

**%matplotlib** inline

*# Convert list to transaction item*

**from** mlxtend.preprocessing **import** TransactionEncoder

*#import apriori*

**from** mlxtend.frequent\_patterns **import** apriori

*# Import association rules*

**from** mlxtend.frequent\_patterns **import** association\_rules

*# Display settings*

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

pd**.**set\_option('display.max\_rows', 100)

### C2. Apriori algorithm

*# Apply apriori algorithm*

a\_rules **=** apriori(cleaned\_df, min\_support **=** 0.02, use\_colnames **=** **True**)

a\_rules**.**head()

In [48]:

a\_rules\_results **=** list(a\_rules)

a\_rules\_results

Out[48]:

['support', 'itemsets']

In [49]:

*# Association rules*

ass\_r **=** association\_rules(a\_rules, metric **=** 'lift', min\_threshold **=** 1)

ass\_r**.**head(10)

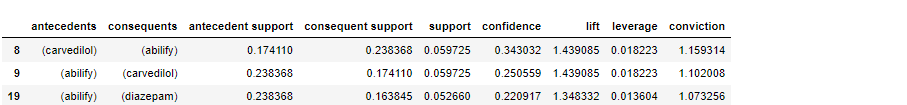
The values for the support, lift, and confidence of the association rules table follow:

Graphical user interface

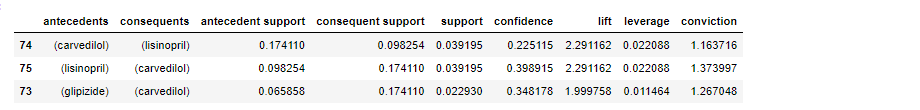
Description automatically generated

**Part IV: Data Summary and Implications**

The support metric indicates the relative concentration of a drug in the dataset. Values must be higher than zero to be considered significant. The top three support values for this analysis follow:



The lift metric measures the tendency two (or more) items are sold together. Values must be higher than one to be considered significant. The top three lift values follow:



The confidence metric measures how often the purchase of the antecedent is tied to the consequent. Confidence displays the probability the consequent will be purchased if the antecedent has been purchased. The top three confidence values follow:

A picture containing diagram

Description automatically generated

The practical significance of this analysis is we are now able associate numeric metric on lift, confidence, and support. This will allow the ability to target and create purchasing incentives for the strongest relationships.

A course of action for an organization would be to find the linking of purchase items and then try to increase sales based on this information. One such example is the drug Carvedilol is the consequent for the second and third highest lift values, for Lisinopril and Glipizide respectively. Somehow incentivizing Carvedilol may increase sales of Lisinopril and Glipizide. Incentives could include price reductions and sending marketing material and samples to doctor offices.

**Part V: Attachments**

[Link](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a8cbb956-60ee-4356-a19b-aeaf00530945) to Panopto video.

***All* web sources used to acquire data or segments of third-party code to support the application.**

1. Kesselly Kamara (April, 2022). Data Mining II.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=5aefcf2b-73cd-41f5-b4d3-aea0011efd05

(Kamara, 2022)

2. Sarit Maitra (Oct 15,2019). Association Rule Mining using Market Basket Analysis.

https://towardsdatascience.com/market-basket-analysis-knowledge-discovery-in-database-simplistic-approach-dc41659e1558

(Maitra, 2019)

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

1. Sarit Maitra (Oct 15,2019). Association Rule Mining using Market Basket Analysis.

https://towardsdatascience.com/market-basket-analysis-knowledge-discovery-in-database-simplistic-approach-dc41659e1558

(Maitra, 2019)

2. Jihargifari (Aug2, 2020). How To Perform Market Basket Analysis in Python

https://medium.com/@jihargifari/how-to-perform-market-basket-analysis-in-python-bd00b745b106

(Jihargifari, 2020)

.