D212 Task 3 PA

Shawn Wheeler

000364855

Table of Contents

[Part I: Research Question 3](#_Toc170413177)

[A1: Proposal of Question 3](#_Toc170413178)

[A2: Defined Goal 3](#_Toc170413179)

[Part II: Market Basket Justification 3](#_Toc170413180)

[B1: Explanation of Market Basket 3](#_Toc170413181)

[B2: Transaction Example 3](#_Toc170413182)

[B3: Market Basket Assumption 4](#_Toc170413183)

[Part III: Data Preparation and Analysis 4](#_Toc170413184)

[C1: Transforming the Data Set 4](#_Toc170413185)

[C2: Code Execution 8](#_Toc170413186)

[C3: Association Rules Table 9](#_Toc170413187)

[C4: Top Three Rules 10](#_Toc170413188)

[Part IV: Data Summary and Implications 11](#_Toc170413189)

[D1: Significance of Support, Lift, and Confidence Summary 11](#_Toc170413190)

[D2: Practical Significance of Findings 11](#_Toc170413191)

[D3: Course of Action 12](#_Toc170413192)

[Part V: Attachments 12](#_Toc170413193)

[E: Panopto Video of Code 12](#_Toc170413194)

[E1: Panopto Video of Programs 12](#_Toc170413195)

[F: Sources for Third-Party Code 12](#_Toc170413196)

[G: Sources 12](#_Toc170413197)

## Part I: Research Question

### A1: Proposal of Question

Using market basket analysis, which medications have frequently been prescribed together, and how can this information be used for enhanced service?

### A2: Defined Goal

The goal of the analysis is to examine associations between medications prescribed to patients historically using market basket analysis in order to determine further action.

## Part II: Market Basket Justification

### B1: Explanation of Market Basket

A market basket refers to a combination of items that a customer might purchase together as in a trip to the grocery store. Market basket analysis describes a technique that helps identify those specific relationships, or associations, over a large amount of data. Rules created based on those associations are used to then organize the data based on specific criteria using the Apriori algorithm, such as, support, lift, and confidence (Li, 2017).

The expected outcome from performing market basket analysis on the medical data set is that it will allow for identifying greater insight into combinations of medications that are often used in conjunction. In other situations, it would be used for potentially determining cross-item discounts, marketing campaigns, or other ways to increase revenue by having a greater understanding of customer patterns. (GeeksForGeeks, 2022).

### B2: Transaction Example

A transaction example is shown below for a combination of medications from the data set. The individual has prescribed five medications over time.

A screenshot of a computer

Description automatically generated

### B3: Market Basket Assumption

One assumption for market basket analysis relies on a lack of significant changes occurring throughout the duration of the data collection. There can be factors outside of the data set itself that may account for particular habits. A recent example would be data acquired during the Covid lockdowns for a retail store. Therefore, customer behavior is assumed to be consistent over time for a particular data set. (Oluwakemi, 2023).

## Part III: Data Preparation and Analysis

### C1: Transforming the Data Set

*Cleaned Data Set Attached:* d212task3clean.csv

The process listed below was used to prepare the data for analysis:

1. Opening data for manipulation and data exploration

df = pd.read\_csv("C:/Users/Owner/medical\_market\_basket.csv")

df.head()

A screenshot of a computer

Description automatically generated

df.info()

A screenshot of a computer

Description automatically generated

1. Check for missing values

df.isnull().sum()

A screenshot of a computer screen

Description automatically generated

1. Remove empty rows

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

1. Convert dataframe into a list of lists

rows = []

for i in range(0, 7501):

rows.append([str(df.values[i,j])

for j in range (0,20)])

1. Use TransactionEncoder and examine data

trans\_encoder = TransactionEncoder()

fit\_trans\_encoder = trans\_encoder.fit(rows).transform(rows)

# Return arrays to DataFrame

medications = pd.DataFrame(fit\_trans\_encoder, columns = trans\_encoder.columns\_)

# looking

medications

A screenshot of a computer

Description automatically generated

for col in medications.columns:

print(col)

A screenshot of a computer screen

Description automatically generated

1. Remove empty column and examine data

cleaned\_df = medications.drop(['nan'], axis = 1)

cleaned\_df.head()

A screenshot of a computer

Description automatically generated

cleaned\_df.info()

A screen shot of a computer

Description automatically generated

1. Write to .csv

cleaned\_df.to\_csv('d212task3clean.csv', index = False)

### C2: Code Execution

The below code was executed to create association rules using Apriori and includes screenshots as applicable:

1. Rename dataframe after having saved prepared copy to a .csv

df = cleaned\_df

1. Create Apriori object (rules)

# (Course Materials, n.d.)

rules = apriori(df, min\_support = 0.02, use\_colnames = True)

rules.head(5)

A screenshot of a computer

Description automatically generated

1. Create rules table using association\_rules and apriori object

rule\_table = association\_rules(rules, metric = 'lift', min\_threshold = 1)

rule\_table.head(20)

A screenshot of a computer

Description automatically generated

### C3: Association Rules Table

*Associated Rules Table Screenshot:*

*A screenshot of a computer

Description automatically generated*

### C4: Top Three Rules

The top three rules shown below have been sorted by lift. Lift, as described below in D1, is one indication of the performance of an association. For the first item set, for instance, one who is prescribed carvedilol (the antecedent) will be prescribed lisinopril (the consequent) with a 22.5% confidence, a lift of 2.29, and a support value of 3.9% for the combination of the two medications.

*Screenshot of Top Three Rules based on lift:*

*A screen shot of a computer

Description automatically generated*

## Part IV: Data Summary and Implications

### D1: Significance of Support, Lift, and Confidence Summary

Support refers to the proportion of all transactions that contain a set of items. The higher the support value, the higher the likelihood of them having been prescribed together with 1 representing 100% of the time. In the first rule of the top three shown ranked by lift, support is shown as being .039195 for carvedilol and lisinopril. Looking at the second set, lisinopril and carvedilol, it makes sense that it has the same value of .039195 since they are the same medications in both instances. This value converts to 3.9195% of all transactions where both of the medications occur. The third rule also includes carvedilol, but it is the consequent of glipizidel, and the support shown for that transaction is .022930 meaning that 2.293% of all transactions include those medications.

Confidence explains the “proportion of all the transactions that contain all the items in the itemset over the proportion of transactions containing just one of them.” (Ph. D., S. C. S., 2020). In other words, it’s the likelihood that the consequent is prescribed based on the original, antecedent, being prescribed. A higher value is again ideal. In the same example, the first rule shows that there is a confidence of .225115 which is a 22.5115% chance that someone who was prescribed carvedilol will also be prescribed lisinopril. For the second set, someone prescribed lisinopril has a 39.8915% chance of also having been prescribed carvedilol. For the third rule, there is a confidence of 34.8178% that someone prescribed glipizide will be prescribed carvedilol.

Finally, lift is a way of expressing the correlation between two items if a relationship is assumed not to exist. A lift above 1 shows an increased likelihood of the consequent also appearing. A lift below one reveals that the antecedent actually decreases the likelihood of the consequent appearing. A lift of 1 represents no change in the likelihood of the consequent appearing due to the antecedent. (Ph. D., S. C. S., 2020). In the example shown, the first set has a lift of 2.291162. This is a 2.291162 times increase in the likelihood of the medications being prescribed together than if prescribed separately. Since it is the same medications for the second set, the lift is the same. Lastly, the third set shows a lift of 1.99758 indicating an almost two times enhanced likelihood of the medications both being prescribed.

### D2: Practical Significance of Findings

For the first set, if an individual is prescribed carvedilol, they are decently likely to also be prescribed lisinopril at some point. There is a confidence of 22.5% and a lift of 2.29 with carvedilol as the antecedent.

Secondly, lisinopril is the antecedent with carvedilol as the consequent. This shows that being prescribed lisinopril, with 39.9% confidence and a lift of 2.29, indicates a carvedilol prescription which is a higher confidence than the first set with the same medications.

Thirdly, someone who is prescribed glipizide will be prescribed carvedilol with a 34.8% confidence and a lift of 2.

### D3: Course of Action

Based on the results shown, there is a relationship between not only the drugs investigated but others, as well. Those relationships could be considered for further research in determining root causes for combinations and perhaps even streamlining treatment based on medication history. More data could be collected to provide even more useful results over time. These insights could perhaps even assist with being able to determine issues arising for a patient in the future.

## Part V: Attachments

### E: Panopto Video of Code

*See Attached Link: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=53976112-431e-4cfe-b93d-b19e0005d86b*

### E1: Panopto Video of Programs

*See Attached Link: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=53976112-431e-4cfe-b93d-b19e0005d86b*

### F: Sources for Third-Party Code

Course Materials (n.d.)

### G: Sources

GeeksforGeeks (2022, March 13). *Market Basket Analysis in Data Mining*. (2022, March 13). GeeksforGeeks. https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/

Li, S. (2017, September 25). *A Gentle Introduction on Market Basket Analysis — Association Rules*. Medium; Towards Data Science. https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Oluwakemi Helen Deniran. (2023, November 27). *Boosting Sales with Data: The Power of Market Basket Analysis in Retail*. Medium; Medium. https://medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df

Ph.D, S. C. S. (2020, November 17). *Market Basket Analysis 101: Key Concepts*. Medium. https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00