# Part I: Research Question

## A1: Proposal of Question

This analysis aims to answer the question: What are the common prescribing trends occurring across the hospital chain? By gaining a better understanding of prescribing trends, the occurrence of drug interactions can be prevented. According to a peer-reviewed study published in 2015, this information is crucial to determining whether drug interactions are a system-wide problem (Sutherland, et al.). Preventing these types of drug interactions can help to reduce system readmission rates. According to a 2013 study on the impact of drug interactions on readmissions, “patients with potential drug interactions in a prior admission were more likely to be readmitted.” (Moura, Tavares, & Acurcio)

## A2: Defined Goal

The primary goal of this analysis is to identify the most common prescribing trends across the hospital chain. This information will help inform the hospital chain’s administrators as they work to improve system-wide outcomes like drug interactions and readmission rates. It will also serve as the answer to the proposed research question. To achieve this goal, a Market Basket Analysis will be performed.

# Part II: Market Basket Justification

## B1: Explanation of Market Basket

Market Basket Analysis is used to mine data to identify its underlying patterns (Smartbridge, 2022). Traditionally, it is used to analyze retail transactions to better understand customers and predict their purchasing behaviors. This is done by first identifying the most frequently purchased items across the transactions. Association rules are then formed between the frequently purchased items which define purchasing behavior. These rules are structured as IF-THEN statements where the items from the IF portion are known as the antecedent and items from the THEN portion are known as the consequent. To identify the relevant association rules, the Apriori algorithm is used in combination with metrics like support, lift, and confidence. In this analysis, retail transactions are being replaced with patient prescriptions. The outcomes will still be association rules. However, the associations will be between sets of prescriptions and will identify the common prescribing trends across the hospital chain.

## B2: Transaction Example

In the provided dataset, a transaction is a patient of the hospital chain represented by a row of the dataset. For this analysis, the medications have been aggregated by the drug class groups to reduce the number of items to be analyzed. The screenshot below displays a single transaction after applying the aggregation.



## B3: Market Basket Assumption

The primary assumption in Market Basket Analysis is that frequently observing two item sets in transactions implies they complement one another and that purchasing one item set will lead to the purchase of the other (Kamakura, 2012). In this analysis this means that if two sets of medications are often prescribed to the same patient, being prescribed one medication means you are likely to be prescribed the other.

# Part III: Data Preparation and Analysis

## C1: Transforming the Dataset

Several steps were taken to transform the medical\_clean dataset to make it suitable for market basket analysis. All associated code can be found in “task3.py” included with this submission. The following steps were taken:

1. The data was imported into a pandas DataFrame and the first five rows were printed to the console.
2. Empty rows were identified and removed. The remaining row count was printed to the console.
3. The numpy not a number (NaN) values were replaced with empty strings.
4. The DataFrame was converted to a list of item sets and the first five were printed to the console.
5. A set of unique prescriptions was created, the number of unique prescriptions was printed to the console, and the set was exported to a comma separated values (CSV) file.
6. Each prescription in the file was looked up on Drugs.com (Drug Index A to Z, 2022) and assigned a group based on its drug class. The specific page referenced for each of the medications can be found in the reference column of the “prescription\_map.csv” spreadsheet included with this submission.
7. The spreadsheet was read back into a pandas DataFrame and the prescriptions and group columns were zipped into a dictionary.
8. The list of item sets created in step three was mapped to the drug class groups using the dictionary. During this process, empty strings and groups duplicated within a patient were dropped. The first five records were printed to the console.
9. The mapped list of item sets was encoded using the TransactionEncoder from mlxtend’s preprocessing module and stored in a pandas DataFrame. The first five rows were printed to the console.
10. The encoded data was saved to a CSV titled “cleaned\_data.csv”.

“cleaned\_data.csv” has been included with this submission.

## C2: Code Execution

The Apriori algorithm was implemented with a minimum support threshold of 0.01. The number of item sets and the top 10 item sets with the highest support were printed to the console as shown in the following screenshot.

Text, table

Description automatically generated

The item sets were used to generate association rules with a minimum support threshold of 0.01. The number of association rules generated, and the first five association rules were printed to the console as shown in the following screenshot.

Graphical user interface, text, application

Description automatically generated

A scatterplot was generated with antecedent support on the x-axis and consequent support on the y-axis. The lift metric was encoded in the size of each point and confidence was encoded in the color of each point. There were also faint blue lines added at both x and y = 0.01 to represent the minimum support threshold that was applied.

A picture containing chart

Description automatically generated

The association rules were then filtered by a confidence value of 0.6. The number of filtered rules and their summary were printed to the console. This screenshot can be seen in the next section, C3: Association Rules Table. A heatmap of the support values was also displayed.

Table

Description automatically generated

## C3: Association Rules Table

The following screenshot displays the support, lift, and confidence of the filtered association rules table.

Table

Description automatically generated

## C4: Top Three Rules

The top three rules generated by the Apriori algorithm were:

1. IF { Diuretics and ACE Inhibitors } THEN { Beta Blockers }
2. IF { ACE Inhibitors and SSRIs } THEN { Beta Blockers }
3. IF { ACE Inhibitors and Statins } THEN { Beta Blockers }

Their summaries are displayed in the following screenshots.

Text

Description automatically generatedText

Description automatically generatedText, letter

Description automatically generated

# Part IV: Data Summary and Implications

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

The support metric for the top three association rules ranged from 0.015598 to 0.019731. Support represents the probability of the antecedent and consequent cooccurring (Kamakura, 2012). This means that of the total 7,500 patients in the dataset, the top three rules occur in 117 to 148 patients.

The lift metric for the top three association rules ranged from 2.005687 to 2.042157. Lift, represented as a ratio, is the support of the cooccurrence of the antecedent and consequent divided by the probability that the two occur independently. A lift of 1.0 indicates that purchasing one item has no impact on whether the other will be purchased. When lift is above 1.0, purchasing one item increases the likelihood that the other item will be purchased. If below 1.0, lift indicates the two items are more likely to be purchased in place of one another rather than together (Smartbridge, 2022). The lift range for the top three rules from this analysis indicate that being prescribed the antecedent medications makes a patient about twice as likely to be prescribed the consequent medication than if the two were prescribed separately.

The confidence metric for the top three association rules ranged from 0.601626 to 0.612565. Confidence is the probability that the consequent will occur conditional on the antecedent (Kamakura, 2012). This means that there is a 60.16% to 61.26% chance that the consequent will occur given the antecedent for the top three association rules.

## D2: Practical Significance of Findings

All of the filtered rules have a consequent of Beta Blockers. The top three rules have antecedents which include Diuretics, ACE Inhibitors, SSRIs, and Statins. Beta blockers, diuretics, and angiotensin-converting enzyme inhibitors (ACE Inhibitors) are all considered Antihypertensive combinations, which are used to treat high blood pressure (List of Antihypertensive combinations - Generics Only, 2022). Selective serotonin reuptake inhibitors (SSRIs) are antidepressants that are used to treat depression and anxiety (Fookes, List of Common SSRIs + Uses & Side Effects, 2022). Statins are used to reduce LDL-cholesterol (List of Statins + Uses, Types & Side Effects, 2022).

The top three association rules indicate that being prescribed a combination of diuretics and ACE inhibitors, SSRIs and ACE inhibitors, or Statins and ACE inhibitors makes a patient more likely to be prescribed a beta blocker. That is a lot of information, so let’s unpack them one by one. In the first rule, IF {Diuretics and ACE Inhibitors} THEN {Beta Blockers}, diuretics, ACE inhibitors, and beta blockers are all used to treat hypertension. This rule follows intuition. In the second rule, IF {SSRIs and ACE Inhibitors} THEN {Beta Blockers}, both ACE inhibitors and beta blockers are used to treat hypertension, but SSRIs are used to treat anxiety and depression. This combination is not as intuitive. It could imply that high anxiety or depression is associated with hypertension but requires further analysis to tease out. In the third rule, IF {Statins and ACE Inhibitors} THEN {Beta Blockers}, pairing ACE inhibitors and beta blockers again makes sense, but statins are used to reduce LDL-cholesterol. At first glance, this relationship does not appear intuitive, but it is. Both hypertension and high LDL-cholesterol are known to increase a patient’s chances of heart attack and stroke. If a more systemic approach is being taken with a patient’s treatment, it would make sense to use medications that treat multiple conditions which could increase their risk.

## D3: Course of Action

This analysis aimed to identify the common prescribing trends across the hospital chain. According to the results, those trends include treatments for hypertension, high cholesterol, anxiety, and depression. The course of action moving forward includes two steps. First, more data is needed. Market basket analysis relies heavily on large amounts of data. Unfortunately, 7,500 patients are a small sample, especially considering the chain is national. Once the additional data has been acquired, this analysis should be performed on that data to determine whether the same trends exist in the larger sample. If they do, the subset being treated for the identified conditions should be investigated further to determine if any undesirable drug-drug interactions are occurring. Focusing on this population could ultimately lead to a reduction in overall readmission rates.

# Part V: Attachments

## E. Panopto Recording

The Panopto video recording demonstrating functionality of the code used for the analysis and a summary of the programming environment has been included with this submission as “task3\_recording”.

## F. Web Sources

Drugs.com. (2022). *Drug Index A to Z*. Retrieved from Drugs.com: https://www.drugs.com/drug\_information.html

G. Sources

Drugs.com. (2022). *List of Antihypertensive combinations - Generics Only*. Retrieved from Drugs.com: https://www.drugs.com/drug-class/antihypertensive-combinations.html

Fookes, C. (2022, September 30). *List of Common SSRIs + Uses & Side Effects*. Retrieved from Drugs.com: https://www.drugs.com/drug-class/ssri-antidepressants.html

Fookes, C. (2022, September 27). *List of Statins + Uses, Types & Side Effects*. Retrieved from Drugs.com: https://www.drugs.com/drug-class/hmg-coa-reductase-inhibitors.html

Kamakura, W. A. (2012, May 22). *Sequential market basket analysis.* Retrieved from SpringerLink: https://doi.org/10.1007/s11002-012-9181-6

Moura, C. S., Tavares, L. S., & Acurcio, F. A. (2013, January 28). *Hospital readmissions related to drug interactions: a retrospective study in a hospital setting.* Retrieved from National Library of Medicine: https://doi.org/10.1590/s0034-89102013005000001

Smartbridge. (2022, March 15). *Market Basket Analysis 101: Anticipating Customer Behavior*. Retrieved from Smartbridge: https://smartbridge.com/market-basket-analysis-101/

Sutherland, J. J., Daly, T. M., Liu, X., Goldstein, K., Johnston, J. A., & Ryan, T. P. (2015, March 4). *Co-Prescription Trends in a Large Cohort of Subjects Predict Substantial Drug-Drug Interactions.* Retrieved from PLOS ONE: https://doi.org/10.1371/journal.pone.0118991