**Research Question**

*Research Question:* Which prescription combinations are most common? Can we predict which prescriptions will be needed in the future by patients based on their existing or past prescriptions? We can answer this question using market basket analysis.

Our goal of this analysis is to learn which combinations of prescriptions are frequently needed by patients. This can help us make predictions about the future prescription needs of patients. Thus, we can theoretically predict the conditions and afflictions that cause the need for the prescriptions. Of course, predicting medical needs is not so simple, but we may be able to keep an eye out for certain conditions if our past patients have shown co-occurrences between them.

**Market Basket Justification**

Market basket analysis is a pattern prediction tool that allows organizations to identify products or services that are frequently purchased together (as part of a “basket”). Market basket analysis uses “association rules” to determine co-occurrences. Association rules are essentially if-then “statements that connect an ‘antecedent’ item to a ‘consequent’ item” (Sivek, 2020), and they do not imply causal relationships between items. These rules are extrapolated using the Apriori algorithm, which operates on the assumption that “all subsets of a frequent itemset to be frequent. Similarly, for any infrequent itemset, all its supersets must also be infrequent” (*Data Science – Apriori Algorithm in Python - Market Basket Analysis, 2021*).

In this scenario, the items in a single row make up “transactions”, which are the various prescriptions each patient has been given. For example, the first row in our dataset represents a patient whose prescriptions (at least once) in the prior two years were abilify, albuterol HFA, albuterol aerosol, allopurinol, amlodipine, Cialis, fluconazole, gabapentin, glipizide, levofloxacin, lorazepam, losartan, metoprolol succinate XL, mometasone, omeprazole, pantoprazole, pravastatin, promethazine, spironolactone, and sulfamethoxazole. This patient’s 20-prescription history is the longest in our dataset, which is named “MBA\_dataset.xlsx” and attached separately as part of this submission.

This analysis will identify prescriptions commonly given to the same patient in this two-year historical period. The code used to prepare and analyze the data is named “Task\_3\_Code.pdf” and is attached separately as part of this submission.

**Data Analysis Summary and Implications**

The screenshot below shows the top three rules generated by the Apriori algorithm:

Table

Description automatically generated

This analysis was guided by three key association rules: support, lift, and confidence. These metrics and their calculations, retrieved from <https://towardsdatascience.com/association-rules-2-aa9a77241654>, are below.

Text

Description automatically generated with low confidence

A picture containing letter

Description automatically generated

Letter

Description automatically generated with low confidence

Support measures frequency of itemsets, and the minimum value for my association rules table for support is 0.03. In other words, any given prescription must have accounted for at least 3% of all prescriptions to be included in the association rules table. Confidence, which measures the likelihood that one prescription is needed given another is also needed, was set at a minimum threshold of 0.20 (or at least a 20% chance). Finally, the minimum lift for my association rules table was set at the standard level of 1.0. This implies that “the antecedent is in fact increasing the likelihood of the consequent also appearing in a transaction” (*Sivek, 2020*), and here, means that the antecedent prescription increases the likelihood of the consequent prescription. Ultimately, 26 prescription combinations met these criteria, and they are represented as dots in the below scatterplot, which shows the relationship between confidence, support, and lift.

Chart, scatter chart

Description automatically generated

In a practical sense, these findings are extremely interesting, but likely lead to more questions than answers. If we are looking strictly at the top three rules from the earlier screenshot, we can see that the first two rules suggest patients often require both Abilify *and* amphetamine salt combo XR. According to Abilify’s website, the drug is used to treat various mood disorders such as bipolar disorder and schizophrenia, particularly in children (*Abilify® (Aripiprazole): Official Site*). Meanwhile, amphetamine salt combo XR, also known as Adderall, is used to treat attention-deficit hyperactivity disorder (ADHD), for which approximately 9.4% of children have received a diagnosis (*Data and statistics about ADHD, 2021*). This raises several questions about the link between these two drugs as well as their underlying causal conditions. This link should be further explored, as it’s possible these conditions are being misdiagnosed, confused with each other, or one drug leads to the need for the other. The third rule in the above screenshot shows Atorvastatin as an antecedent and Abilify as a consequent. Atorvastatin is used to treat high cholesterol, and as a preventative for heart disease (*Atorvastatin, 2018*). The link between these two should, again, be further explored. Abilify is frequently listed on our association rules table, and this frequency could indicate that it is either overprescribed or has side effects that lead to the need for other prescriptions.

My overall recommendation is to further explore all rules in our association table and study the links between the drugs that appear together. I would also be interested in seeing demographic information on these patients. Furthermore, we should become weary of prescribing Abilify unless we have a great deal of confidence that it is not causing unnecessary adverse side effects and is also being properly prescribed.

**References**

*Abilify® (Aripiprazole): Official Site*. ABILIFY® (aripiprazole) | Official Site. (n.d.). Retrieved December 13, 2021, from https://www.abilify.com/.

Centers for Disease Control and Prevention. (2021, September 23). *Data and statistics about ADHD*. Centers for Disease Control and Prevention. Retrieved December 13, 2021, from https://www.cdc.gov/ncbddd/adhd/data.html.

*Data science - Apriori algorithm in Python - Market Basket Analysis*. Intellipaat. (2021, August 31). Retrieved December 13, 2021, from https://intellipaat.com/blog/data-science-apriori-algorithm/.

National Health Service. (2018, December 19). *Atorvastatin*. NHS. Retrieved December 13, 2021, from https://www.nhs.uk/medicines/atorvastatin/.

Sivek, S. C. (2020, November 16). *Market basket analysis 101: Key concepts*. Medium. Retrieved December 13, 2021, from https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00.