Association Rules and Lift Analysis – D212

Gooden, Nina S. Gooden [ID #: 009823504]

Dr. Kesselly Kamara

Association Rules and Lift Analysis – D212

This analysis will explore the medical data of a theoretical real-world organization. By utilizing the association rules and lift analysis I’ve learned throughout the duration of this course, I will build rules and consequents to address the business’ concern about readmission in a data-first evaluation. I will also provide visualizations to support my assessments, in tandem with code for my models. As previously requested, I will also discuss the limitations and potential course of action this data supports.

## **A. The Research Question:**

1. Research Question
   1. Using market basket analysis, can we discern the top three associated prescriptions in the dataset?
2. Data Analysis Goals and Objective
   1. I will be applying market basket analysis to the medical prescription data in order to find measurable associations between data features. I will be focusing on life, then confidence, then support as my metric scale.

## **B. Justification for Technique**

1. Market Basket Analysis is a tool employed to measure “accidental transaction patterns.” (Ng, 2016)These patterns occur when the introduction of one variable statistically impacts the likelihood of another variable being triggered at the same time. MBA algorithms use association rules to count how often variables are group together in order to identify said associations.
   1. **Expected outcomes for this analysis are: I am confident that I will find many associations from this dataset, as medications are frequently paired or predictable patterns. I expect all of my select rules to have a lift higher than 1.**
2. Example of transaction in the dataset:
   1. I used .iloc to pull a transaction from the dataset after it has been cleaned. The transaction here received three prescriptions: citalopram, Benicar,

A screenshot of a computer

Description automatically generated

1. Assumption of market basket analysis:
   1. Market basket analysis assumes that any grouping of a frequent variable set must be frequent. (Kamakura, 2012) Essentially, the analysis assumes that if a pairing shows up enough times then there must be a reasonable and mathematical assertion that those items are pairable.

## **C. Data Preperation and Analysis**

1. Transform the dataset. Include cleaned version:
   1. My first step for transforming the dataset was to import my libraries and data.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. I noticed there were blank lines in the dataset, so I used .iloc to drop every other row, starting with the second row.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. Checked the data types and the shape of the DataFrame.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. In order to encode the data for .apriori(), I needed to create a list that held all of the values from the DataFrame.

Text

Description automatically generated

* 1. I transactionalize the data with TransactionEncoder() created an array of Boolean values for the data. I then passed those values back into a DataFrame.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. Next I needed to check the columns for emptiness by printing their names and looking for “nan” value in the list.

Graphical user interface, text

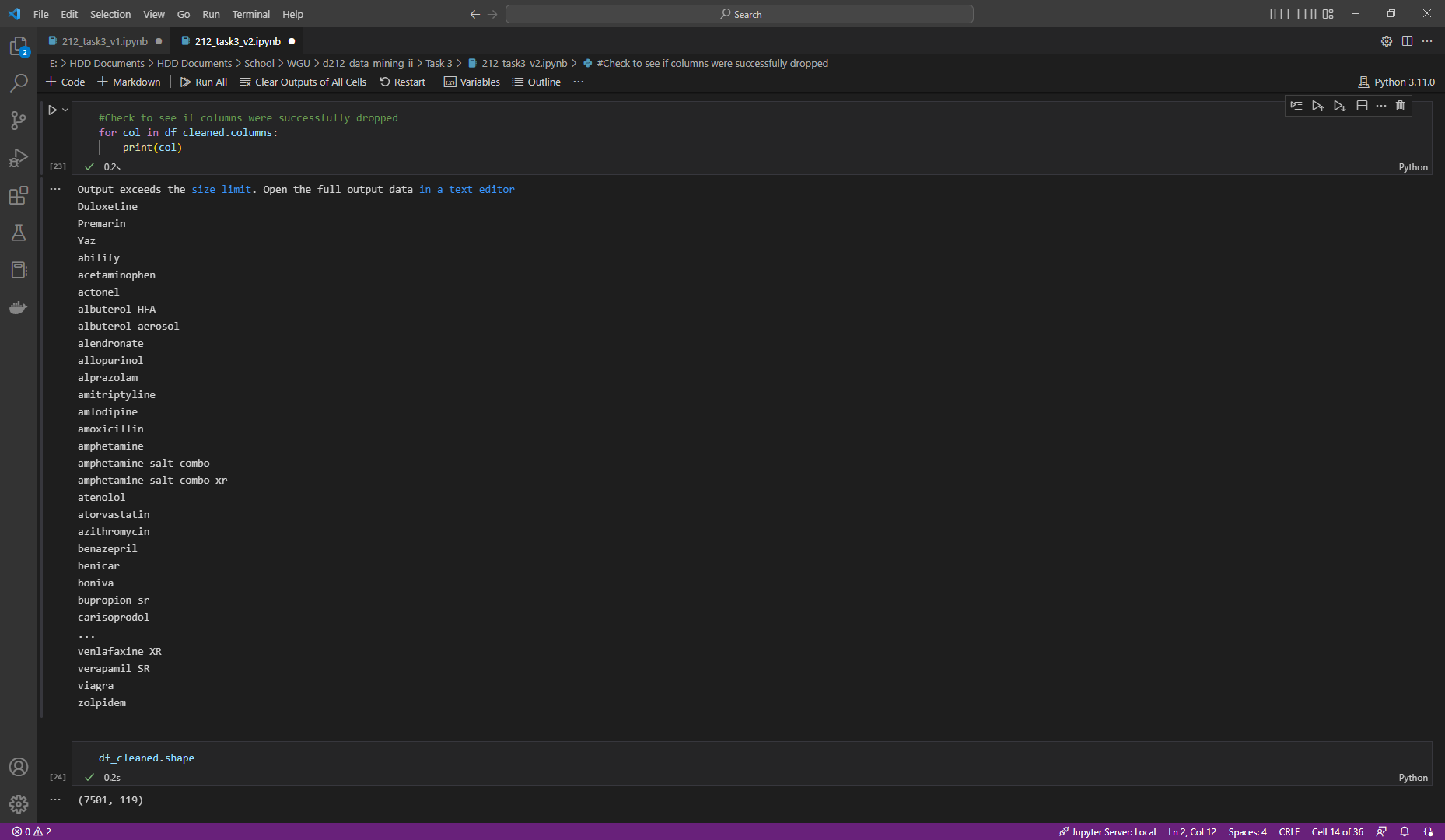
Description automatically generated

* 1. After confirming that there was a “nan” value, I dropped the resulting column from the DataFrame.

A screenshot of a computer

Description automatically generated with medium confidence

* 1. I checked the column values once more and printed the shape of the DataFrame to confirm that I had the right number of transactions.

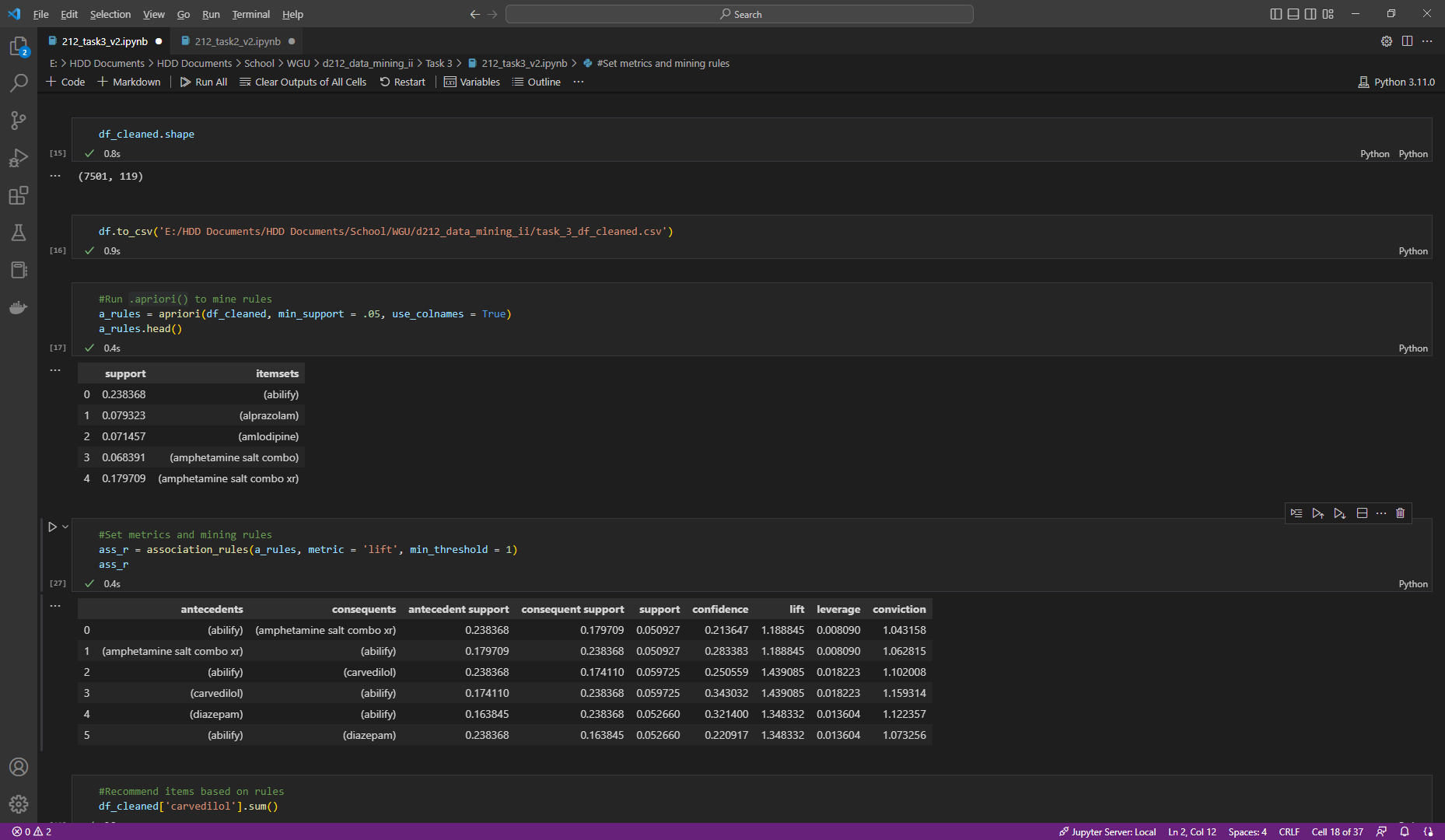


* 1. Lastly, I exported the DataFrame to a .csv file for this assignment.

A screenshot of a computer

Description automatically generated

1. Generate association rules. Provide code:
   1. I used .apriori() to establish the mine rules with a minimum support value of .05.



* 1. I set the metrics for the mining rules and set the threshold for the lift value to 1.

A screenshot of a computer

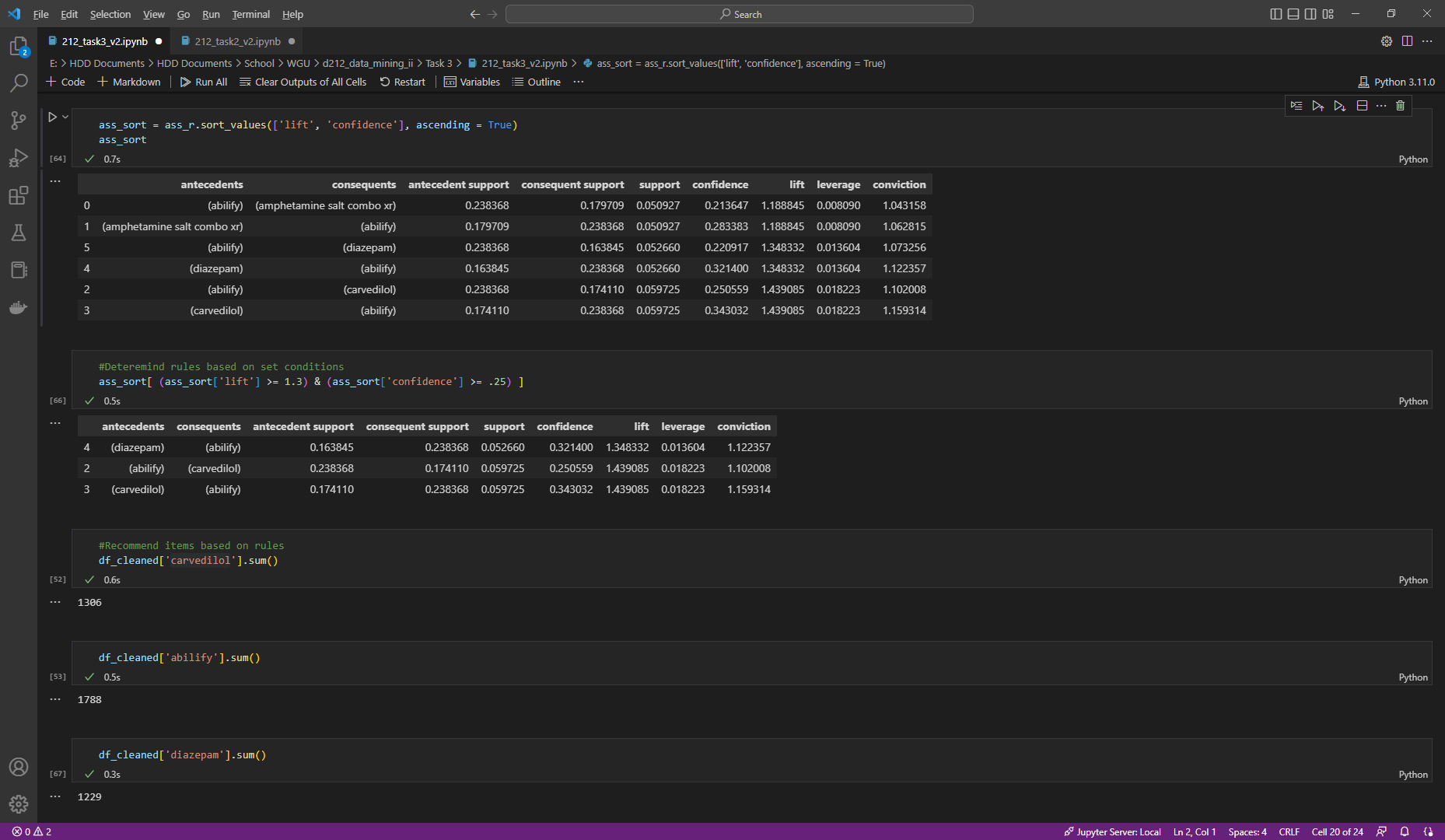
Description automatically generated with medium confidence

1. Provide values for support, lift, and confidence:
   1. Accomplished in 2b; however, I pulled the values into their own DataFrame for this assignment.

Graphical user interface

Description automatically generated

1. Identify the top three rules from the algorithm:
   1. I sorted values in the DataFrame by lift and confidence:



* 1. Then I set a median threshold for each condition in order to narrow the options down to three

A screenshot of a computer

Description automatically generated with medium confidence

* 1. Finally, I took the .sum of each antecedent for frequency

A screenshot of a computer

Description automatically generated with medium confidence

## **D. Data Summary and Implications**

1. Summarize the support, lift, and confidence results:
   1. The top result for this analysis was as follows:
      1. (carvedilol): The antecedent value is the leading predicting prescription. This following statistical values are two if this medication is prescribed first.
      2. (abilify): The consequent value is the variable most likely to be paired with the antecedent. The following statistics are true—and highest in the dataset—when the consequent is prescribed after the established antecedent.
      3. 0.059725: The support value for this rule was set at .05. As such, this value measures how popular a pairing is over the course of the DataFrame.
      4. 0.343032: The confidence for this rule is .34, which means that of occasions of the antecedent, 34% of those transactions also contain the listed consequent.
      5. 1.439085: The lift measures the likelihood of these two variables being prescribed together. At 1.44, the data states that when the antecedent is prescribed, there is an increased 1.4x likelihood that the consequent will also be prescribed.
2. Practical significance:
   1. A 34% confidence level is not uniquely alarming in a medical-focused dataset. While the correlation between the top prescriptions is statistically significant, the dataset was not large enough to draw assumptions based on this analysis alone.
3. Recommended course of action:
   1. In order to better understand whether or not this analysis is cause for action, the medical center will need to provide additional information on the prescriptions themselves. My next steps recommendation would be to conduct further analysis on whether or not these baskets also correlate with readmission factors.

## **E. Panopto**

[Data Mining II – OFM3](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=ef7536a4-78b4-400c-9dd9-af6c003aa9e2)

## **G. Reference web sources**

Kamara & Western Governors University (n.d.). *Data Mining II - D212 Theory*. Panopto. https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9541a29b-2f14-4c5d-9d86-af030005bcf6

## **H. Acknowledge sources**

Jihargifari. (2021, December 15). How To Perform Market Basket Analysis in Python - Jihargifari. Medium. https://medium.com/@jihargifari/how-to-perform-market-basket-analysis-in-python-bd00b745b106

Contributor, T. (2019, May 7). *market basket analysis*. Customer Experience. https://www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis

Kamakura, W. A. (2012). Sequential market basket analysis. *Springer Science+Business Media, LLC*, 1-3.

Ng, A. (2016). *Association Rules and the Apriori Algorithm: A Tutorial.* Algobeans: Ministry of Defence of Singapore.