Market Basket Analysis of Customer Transactional Data

Mike Mattinson

Western Governors University

D212: Data Mining II

Task 3: Association Rules and Lift Analysis

Dr. Kesselly Kamara

May 27, 2022

Revision 1

Abstract

Transaction data will be analyzed using Apriori algorithm also known as "Market Basket Analysis" or "Association Rules". Approximately 7,500 customer transactions will be considered. There are approx. 119 unique items used in those transactions. The top association rules will be generated and reviewed. Minimum support of 3/1000. Minimum confidence of 3/10. Rules with lift values above 1 are considered to have strong correlation between the antecedents and consequents. Specifically, blue-light reading glasses have a six (6) times likelihood of purchase when computer accessories of "Dust-Off Compressed Gas 2 pack" and "Anker 2-in-1 USB Card Reader" are purchased.

Keywords: Data Mining. Market Basket Analysis. Association Rules and Lift Analysis.

Scenario 1

One of the most critical factors in customer relationship management that directly affects a company's long-term profitability is understanding its customers. When a company can better understand its customer characteristics, it is better able to target products and marketing campaigns for customers, resulting in better profits for the company in the long term.

You are an analyst for a telecommunications company that wants to better understand the characteristics of its customers. You have been asked to perform a market basket analysis to analyze customer data to identify key associations of your customer purchases, ultimately allowing better business and strategic decision-making.

List of Tables

List of Figures

Part I: Research Question

A. Describe the purpose of this data mining report by doing the following:

A1. Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

What are the top three (3) association rules for customer purchasing data?

A2. Define one goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

Apply Market Basket Analysis (MBA) to customer purchasing data to determine the top three (3) association rules sorted by highest lift calculations. There are customarily three (3) metrics to choose from, support, confidence and/or lift, I will be using lift as the primary metric. These metrics are visually depicted in Figure 1 below.

Figure 1

Market Basket Analysis metrics, small visual example

$$Support = \frac{frq(X,Y)}{N}$$

$$Rule: X \Rightarrow Y \longrightarrow Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

Example:

Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

Notes. Figure courtesy of Susan Li (Li, 2017)

Part II: Market Basket Justification

B. Explain the reasons for using market basket analysis by doing the following:

B1. Explain how market basket analyzes the selected dataset. Include expected outcomes.

Market Basket Analysis is a type of transactional analysis that will determine which items of the transaction data are likely to be included with other items. It can be a single item with another single item, or it can be a set of items with other sets of items. The analysis will generate a series of association rules, kind of like logical "if-then" conditions, and then calculate a number of standard metrics used to determine how likely the two sets are found together.

The sets X and Y come from all possible combinations of subsets from the transaction data. These sets are generated when the apriori algorithm is executed.

The antecedent set (X) is the "if" part of the if-then condition.

The consequents set (Y) is the "then" part of the if-then condition. The antecedent and consequents are mutually exclusive sets.

The rule is named as X➔Y, or the set X implies the set Y, of "if set X occurs, then the set Y will also occur."

The support metric is basically a count of how many times the set if found in the total number of transactions, for the set X, the support (X) is defined as the frequency of X divided by the total number of transactions. If you are referring to the support of (X➔Y), then that would be calculated as the frequency of X or Y divided by the total number of transactions.

The confidence metric is the ratio of the support of (X➔Y) divided by the support (X). Since the total number of transactions is in both the numerator and denominator, it can be

cancelled out and you are left with the frequency (X→Y) divided by frequency (X). Figure 1 does a much better job of summarizing the three (3) metrics.

Lift is the final metric, it is calculated as the support (X→Y) divided by (support (X) * support (Y)). See Figure 1. There are other metrics such as leverage and conviction that are available in the apriori algorithm, but will not be used during this analysis.

The Apriori algorithm is accomplished in two (2) steps: first, the total list of all possible itemsets is generated, but because there could potentially be a very big list, the list is truncated using a minimum support value supplied by user. Secondly, once this frequency itemsets is generated, then the association rules are generated and then "pruned" to find and explore the associations until the desired number of rules is achieved.

The user is able to prune the total list of associated rules using minimum cutoff values for any of the metrics being used.

The final list of rules is then output showing the rule and metrics, this is the expected output from running the apriori algorithm.

B2. Provide one example of transactions in the dataset.

Table 1 shows one (1) example from the customer purchasing data. The raw .CSV data had a blank row after every transaction, so I elected to remove the blank rows during the initial file read using the skiprows method. The transactions dataset has a list of from one (1) item to as many as 20 items. The example in Table 1 has six (6) items.

Table 1

Show one example from the transactions dataset

```
In [29]:  ▶ # load data from .csv & show first few records
            raw = pd.read_csv('data/teleco_market_basket.csv',
                    skiprows=lambda x: x % 2) # remove every other blank row
            #raw.fillna(0,inplace=True) # replace empty value with 0
            raw.iloc[[25]].T
```

Out[29]:

	25
Item01	PS4 Headset
Item02	VIVO Dual LCD Monitor Desk mount
Item03	10ft iPHone Charger Cable 2 Pack
Item04	HP 65 ink
Item05	Screen Mom Screen Cleaner kit
Item06	SanDisk 128GB microSDXC card
Item07	NaN
Item08	NaN
Item09	NaN
Item10	NaN
Item11	NaN
Item12	NaN
Item13	NaN
Item14	NaN
Item15	NaN
Item16	NaN
Item17	NaN
Item18	NaN
Item19	NaN
Item20	NaN

Notes. Using Pandas iloc method to retrieve item 25. Notice this customer transaction purchased six (6) items.

B3. Summarize one assumption of market basket analysis.

   The apriori algorithm assumes the antecedents and consequents are dependent. The lift

ratio is a "better way to judge the strength of an association rule" (Bruce, Gedeck, Shmueli, &

Patel, 2019), referring to level of independence of two sets. There are two other metrics, leverage

and conviction, that can be used to show the measure of independence between two sets.

Part III: Data Preparation and Analysis

C. Prepare and perform market basket analysis by doing the following:

C1. Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.

Load raw data from file. Table 2 shows the structure of the dataframe after the initial file read. There are 7,501 transactions that have at least one (1) item. There are 5,747 transactions that have at least two (2) items, etc.

Remove blank rows. The original file of transactions contains empty rows, there is a blank row after every transaction. It is necessary to remove the blank rows to make the data suitable for market basket analysis. I elected to remove the blank rows, every other line of the file, during the file read using a pandas method called 'skiprows'. When the file is read, every other line will be skipped. There are 7,501 total unique transactions, so that the final number of rows should be 7,501.

Remove blank columns. It is also important to look for blank columns. After the initial raw data is created, I will use a pandas method 'dropna' to find and remove blank columns. Market basket analysis depends on not having blank rows or columns when the algorithm is executed. There are 119 total unique items, so that the final number of columns should be 119. Table 3 shows one (1) extra column, which you can see is the 'nan' item. Table 4 shows the final dataframe with the blank column removed, showing the correct number of rows and columsn.

Transform data to a "list of lists". The next step is to transform the dataframe into a "list of lists" using a double for-loop where the columns become the items and the row values are True or False depending on whether the item was part of that row's transactions or not.

Convert "list of lists" back to dataframe. After the array is created, convert the array into a new dataframe.

(Optional). At this point, it is easy to convert from Boolean data types of True and False values into integer data types of 1 or 0 values. Many analysts prefer to use 1's and 0's instead of Boolean values. It can be done by converting data types using the "array = arrary.astype(int)" method. This type conversion takes the Boolean values and converts True's and False's to 1s and 0s. If this is desired, then you would also change the package used to create the association rules. For the purpose of this assignment, I have elected to follow the course instructor's recommendations to keep the data as Boolean and to use the mlxtend package to create the association rules.

Export final dataframe. Table 5 shows the final dataframe ready to be used by the apriori algorithm. This final dataframe is exported to .CSV file.

Table 2

Dataset after removing empty rows

```
In [6]:  ▶| # dataframe structure
            print(raw.info())
            print('{} empty rows.'.format(raw.isna().all(axis=1).sum()))
            print(raw.shape)

            <class 'pandas.core.frame.DataFrame'>
            RangeIndex: 7501 entries, 0 to 7500
            Data columns (total 20 columns):
             #   Column  Non-Null Count  Dtype
            ---  ------  --------------  -----
             0   Item01  7501 non-null   object
             1   Item02  5747 non-null   object
             2   Item03  4389 non-null   object
             3   Item04  3345 non-null   object
             4   Item05  2529 non-null   object
             5   Item06  1864 non-null   object
             6   Item07  1369 non-null   object
             7   Item08  981 non-null    object
             8   Item09  654 non-null    object
             9   Item10  395 non-null    object
             10  Item11  256 non-null    object
             11  Item12  154 non-null    object
             12  Item13  87 non-null     object
             13  Item14  47 non-null     object
             14  Item15  25 non-null     object
             15  Item16  8 non-null      object
             16  Item17  4 non-null      object
             17  Item18  4 non-null      object
             18  Item19  3 non-null      object
             19  Item20  1 non-null      object
            dtypes: object(20)
            memory usage: 1.1+ MB
            None
            0 empty rows.
            (7501, 20)
```

Notes. Items purchased from previous customers based on historical data. The data set consists of 7,501 customer purchase history and 20 columns/variables. Item01-Item20: Item that customer bought. Only one transaction was for a customer who purchased 20 items. After loading the initial dataset, only reading in every other row, there are zero empty rows.

Table 3

120 items including the blank 'nan' item

```
In [8]:  ▶| print(TE.columns_)
            print(len(TE.columns_))
```

```
['10ft iPHone Charger Cable', '10ft iPHone Charger Cable 2 Pack', '3 pack Nylon Braided Lightning
Cable', '3A USB Type C Cable 3 pack 6FT', '5pack Nylon Braided USB C cables', 'ARRIS SURFboard SB8
200 Cable Modem', 'Anker 2-in-1 USB Card Reader', 'Anker 4-port USB hub', 'Anker USB C to HDMI Ada
pter', 'Apple Lightning to Digital AV Adapter', 'Apple Lightning to USB cable', 'Apple Magic Mouse
2', 'Apple Pencil', 'Apple Pencil 2nd Gen', 'Apple Power Adapter Extension Cable', 'Apple USB-C Ch
arger cable', 'AutoFocus 1080p Webcam', 'BENGOO G90000 headset', 'Blue Light Blocking Glasses', 'B
lue Light Blocking Glasses 2pack', 'Brother Genuine High Yield Toner Cartridge', 'Cat 6 Ethernet C
able 50ft', 'Cat8 Ethernet Cable', 'CicTsing MM057 2.4G Wireless Mouse', 'Cleaning Gel Universal D
ust Cleaner', 'Creative Pebble 2.0 Speakers', 'DisplayPort ot HDMI adapter', 'Dust-Off Compressed
Gas', 'Dust-Off Compressed Gas 2 pack', 'FEEL2NICE 5 pack 10ft Lighning cable', 'FEIYOLD Blue ligh
t Blocking Glasses', 'Falcon Dust Off Compressed Gas', 'HOVAMP Mfi 6pack Lightning Cable', 'HOVAMP
iPhone charger', 'HP 61 2 pack ink', 'HP 61 Tri-color ink', 'HP 61 ink', 'HP 62XL Tri-Color ink',
'HP 62XL ink', 'HP 63 Ink', 'HP 63 Tri-color ink', 'HP 63XL Ink', 'HP 63XL Tri-color ink', 'HP 64
Tri-Color ink', 'HP 64 ink', 'HP 65 ink', 'HP 902XL ink', 'HP 952 ink', 'HP ENVY 5055 printer', 'H
P952XL ink', 'HooToo USB C Hub', 'HyperX Cloud Stinger Headset', 'Jelly Comb 2.4G Slim Wireless mo
use', 'Leader Desk Pad Protector', 'Logitech M510 Wireless mouse', 'Logitech MK270 Wireless Keyboa
rd/Mouse', 'Logitech MK345 Wireless combo', 'Logitech USB H390 headset', 'M.2 Screw kit', 'Mfi-Cer
tified Lightning to USB A Cable', 'Micro Center 32GB Memory card', 'Microsot Surface Dock 2', 'Mor
ead HDMI to VGA Adapter', 'Mpow HC6 USB Headset', 'NETGEAR CM500 Cable Modem', 'NETGEAR Nighthawk
WiFi Router', 'NETGEAR Orbi Home Mesh WiFi System', 'Nylon Braided Lightning to USB cable', 'PS4 H
eadset', 'Premium Nylon USB Cable', 'RUNMUS Gaming Headset', 'SAMSUNG 128GB card', 'SAMSUNG 256 GB
card', 'SAMSUNG EVO 32GB card', 'SAMSUNG EVO 64GB card', 'Sabrent 4-port USB 3.0 hub', 'SanDisk 12
8GB Ultra microSDXC card', 'SanDisk 128GB card', 'SanDisk 128GB microSDXC card', 'SanDisk 32GB Ult
ra SDHC card', 'SanDisk 32GB card', 'SanDisk Extreme 128GB card', 'SanDisk Extreme 256GB card', 'S
anDisk Extreme 32GB 2pack card', 'SanDisk Extreme Pro 128GB card', 'SanDisk Extreme Pro 64GB car
d', 'SanDisk Ultra 128GB card', 'SanDisk Ultra 256GB card', 'SanDisk Ultra 400GB card', 'SanDisk U
ltra 64GB card', 'Screen Mom Screen Cleaner kit', 'Stylus Pen for iPad', 'Syntech USB C to USB Ada
pter', 'TONOR USB Gaming Microphone', 'TP-Link AC1750 Smart WiFi Router', 'TP-Link AC4000 WiFi rou
ter', 'TopMate C5 Laptop Cooler pad', 'UNEN Mfi Certified 5-pack Lightning Cable', 'USB 2.0 Printe
r cable', 'USB C to USB Male Adapter', 'USB Type C Cable', 'USB Type C to USB-A Charger cable', 'V
IVO Dual LCD Monitor Desk mount', 'VicTsing Mouse Pad', 'VicTsing Wireless mouse', 'Vsco 70 pack s
tickers', 'Webcam with Microphone', 'XPOWER A-2 Air Pump blower', 'YUNSONG 3pack 6ft Nylon Lightni
ng Cable', 'hP 65 Tri-color ink', 'iFixit Pro Tech Toolkit', 'iPhone 11 case', 'iPhone 12 Charger
cable', 'iPhone 12 Pro case', 'iPhone 12 case', 'iPhone Charger Cable Anker 6ft', 'iPhone SE cas
e', 'nan', 'nonda USB C to USB Adapter', 'seenda Wireless mouse']
120
```
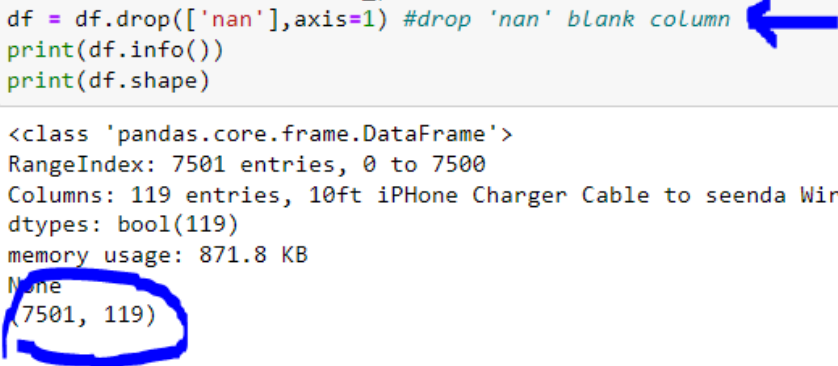
Notes. The extra 'nan' item will need to be removed before running the analysis.

Table 4

Final dataframe with correct number of rows and columsn

```
In [10]:  ▶ # create cleaned dataframe
            df = pd.DataFrame(array,
                    columns=TE.columns_)
            df = df.drop(['nan'],axis=1) #drop 'nan' blank column  ⬅
            print(df.info())
            print(df.shape)

            <class 'pandas.core.frame.DataFrame'>
            RangeIndex: 7501 entries, 0 to 7500
            Columns: 119 entries, 10ft iPHone Charger Cable to seenda Wireless mouse
            dtypes: bool(119)
            memory usage: 871.8 KB
            None
            (7501, 119)
```

Notes. The final shape of the dataframe with 7,501 rows and 119 unique transaction items for columns.

Table 5

Final dataframe ready for market basket analysis

```
In [15]: ▶| # export clean data ready for market basket analysis
            df.to_csv('tables\d212_3_1_cleaned.csv', index=False, header=True)
            print(df.info())
            print(df.shape)
            df.head(5).T

            <class 'pandas.core.frame.DataFrame'>
            RangeIndex: 7501 entries, 0 to 7500
            Columns: 119 entries, 10ft iPHone Charger Cable to seenda Wireless mouse
            dtypes: bool(119)
            memory usage: 871.8 KB
            None
            (7501, 119)
```

Out[15]:

	0	1	2	3	4
10ft iPHone Charger Cable	True	False	False	False	False
10ft iPHone Charger Cable 2 Pack	False	False	False	False	False
3 pack Nylon Braided Lightning Cable	False	False	False	False	False
3A USB Type C Cable 3 pack 6FT	True	False	False	False	False
5pack Nylon Braided USB C cables	False	False	False	False	False
...	...	...	...	...	...
iPhone 12 case	False	False	False	False	False
iPhone Charger Cable Anker 6ft	False	False	False	False	False
iPhone SE case	False	False	False	False	False
nonda USB C to USB Adapter	True	False	False	False	False
seenda Wireless mouse	False	False	False	False	False

119 rows × 5 columns

Notes. The final dataframe ready for the market basket analysis. There are 7,501 transactions and 119 columns. The columns are the unique transaction items and the rows are Boolean True or False depending on whether that item was included or not in the transaction. The file is exported to 'tables\d212_3_1_cleaned.csv' and is attached to the submission.

C2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

The basic process to execute the apriori algorithm is:

a. create the frequency itemsets using some minimal support cutoff set by user, then

b. create rules from the itemsets created above, then

c. prune the rules using additional filters, then

d. repeat steps a, b and c until you achieve desired results.

Table 6 shows an example of executing these steps.

C3. Provide values for the support, lift, and confidence of the association rules table.

Table 6 also shows the metrics calculated for each rule. Metrics include support, lift and confidence.

C4. Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

Table 7 shows a list of the top three (3) rules sorted by lift values.

Table 6

Example code to generate association rules using Apriori algorithm

```
In [66]:  ▶ # min_sup 3/1000 - min_conf 3/10 (sorted by lift)
            pd.set_option('display.max_colwidth', None)
            itemsets = apriori(df, min_support=3/1000, use_colnames=True)
            rules = association_rules(itemsets, metric='confidence', min_threshold=3/10)
            rules[output_cols].sort_values(by=['lift'], ascending=False).round(3).head(8)
```

Out[66]:

	antecedents	consequents	support	confidence	lift
556	(Nylon Braided Lightning to USB cable, 10ft iPHone Charger Cable 2 Pack)	(Screen Mom Screen Cleaner kit, Dust-Off Compressed Gas 2 pack)	0.003	0.383	7.987
135	(Dust-Off Compressed Gas 2 pack, Anker 2-in-1 USB Card Reader)	(FEIYOLD Blue light Blocking Glasses)	0.004	0.403	6.116
283	(VIVO Dual LCD Monitor Desk mount, AutoFocus 1080p Webcam)	(SanDisk Ultra 64GB card)	0.003	0.489	4.981
78	(iPhone 11 case)	(HP 63XL Ink)	0.006	0.373	4.701
530	(SanDisk 128GB Ultra microSDXC card, USB 2.0 Printer cable)	(SanDisk Ultra 64GB card)	0.003	0.462	4.697
546	(VIVO Dual LCD Monitor Desk mount, iPhone Charger Cable Anker 6ft)	(SanDisk Ultra 64GB card)	0.003	0.460	4.682
555	(Nylon Braided Lightning to USB cable, 10ft iPHone Charger Cable 2 Pack, Dust-Off Compressed Gas 2 pack)	(Screen Mom Screen Cleaner kit)	0.003	0.605	4.671
608	(Nylon Braided Lightning to USB cable, HP 61 ink, Dust-Off Compressed Gas 2 pack)	(Logitech M510 Wireless mouse)	0.003	0.329	4.601

Notes. The frequency itemsets is the first part and is used to created itemsets using a minimum support cutoff (in this case the minimum support cutoff is set at 3/1000), and the second part is used to creates the rules using a specific metric (in this case the minimum confidence cutoff is set at 3/10). .head(8) shows the first eight (8) rules meeting these criterion. The number one (1) rule has a very strong lift value of 8.

Table 7

Top three (3) rules sorted by lift

```
In [69]:  ▶| # top 3 rules sorted by lift
             pd.set_option('display.max_colwidth', None)
             itemsets = apriori(df, min_support=3/1000, use_colnames=True)
             rules = association_rules(itemsets, metric='confidence', min_threshold=4/10)
             rules[output_cols].sort_values(by=['lift'],
                       ascending=False).round(3).head(3)
```

Out[69]:

	antecedents	consequents	support	confidence	lift
39	(Dust-Off Compressed Gas 2 pack, Anker 2-in-1 USB Card Reader)	(FEIYOLD Blue light Blocking Glasses)	0.004	0.403	6.116
91	(VIVO Dual LCD Monitor Desk mount, AutoFocus 1080p Webcam)	(SanDisk Ultra 64GB card)	0.003	0.489	4.981
206	(SanDisk 128GB Ultra microSDXC card, USB 2.0 Printer cable)	(SanDisk Ultra 64GB card)	0.003	0.462	4.697

Notes. After creating association rules and filtering and sorting by lift values, the final set of rules is available. The "best" association rule has a lift of 6, indicating a strong likelihood of occurrence. This completes the original analysis goal stated in section A2 and answers the research question in section A1.

Part IV: Data Summary and Implication

D. Summarize your data analysis by doing the following:

D1. Summarize the significance of support, lift, and confidence from the results of the analysis.

Here is what the metrics look like for the top rule:

	antecedents	consequents	support	confidence	lift
39	(Dust-Off Compressed Gas 2 pack, Anker 2-in-1 USB Card Reader)	(FEIYOLD Blue light Blocking Glasses)	0.004	0.403	6.116

The support value for the first rule is 0.004. This number is calculated by dividing the number of transactions containing "Dust-Off Compressed Gas 2 pack", "Anker 2-in-1 USB Card Reader", and "FEIYOLD Blue light Blocking Glasses" by the total number of transactions.

The confidence level for the rule is 0.403, which shows that out of all the transactions that contain both "Dust-Off Compressed Gas 2 pack" and "Anker 2-in-1 USB Card Reader", 40.3 percent contain "FEIYOLD Blue light Blocking Glasses" too.

The lift of 6.116 tells us that "FEIYOLD Blue light Blocking Glasses" is 6 times more likely to be bought by the customers who buy both "Dust-Off Compressed Gas 2 pack" and "Anker 2-in-1 USB Card Reader" compared to the default likelihood sale of "FEIYOLD Blue light Blocking Glasses".

D2. Discuss the practical significance of the findings from the analysis.

The transactions data was small and the minimum support of 3/1000 for the frequency itemsets was small, the overall support values for all rules was very small. However, within the

problem space of these 7,500 transactions, there is a strong likihood of these top rules with a confidence of at least 40%.

D3. Recommend a course of action for the real-world organizational situation from part A1 based on your results from part D1.

Based on the findings, recommend positioning the "blue light" glasses, which are often used when operating a computer or other electronic displays, recommend positioning these glasses near to the computer accessories, specifically the "Dust-Off Compressed Gas 2 pack" and "Anker 2-in-1 USB Card Reader". This recommendation is based on the strong association between the glasses and these computer accessories.

Part V: Attachments

E. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=db41eb38-2833-4b2b-b513-aea20142eccb

F. Record all web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

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

H. Demonstrate professional communication in the content and presentation of your submission.

References

Bruce, P. C., Gedeck, P., Shmueli, G., & Patel, N. R. (2019). *Data Mining for Business Analytics Concepts, Techniques and Applications in Python.* Wiley & Sons, Incorporated, John.

Li, S. (2017, September). A Gentle Introduction on Market Basket Analysis — Association Rules. *A Gentle Introduction on Market Basket Analysis — Association Rules*. Retrieved May 24, 2022, from https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Maitra, S. (2021, September). Association Rule Mining using Market Basket Analysis. *Association Rule Mining using Market Basket Analysis*. Retrieved May 24, 2022, from https://towardsdatascience.com/market-basket-analysis-knowledge-discovery-in-database-simplistic-approach-dc41659e1558

Market Basket Analysis. (2015, September). *Market Basket Analysis*. Retrieved May 26, 2022, from https://sarahtianhua.wordpress.com/portfolio/market-basket-analysis/

Mazzeschi, M. (2020, June). Performing a Market Basket Analysis with Machine Learning. *Performing a Market Basket Analysis with Machine Learning*. Retrieved May 24, 2022, from https://pub.towardsai.net/performing-a-market-basket-analysis-with-machine-learning-94b5514ce1cb

Pipis, G. (2021, May). Market Basket Analysis and Association Rules from Scratch. *Market Basket Analysis and Association Rules from Scratch*. Retrieved May 24, 2022, from https://medium.com/geekculture/market-basket-analysis-and-association-rules-from-scratch-b621dd55e776

Roshan, B. (2020, December). Market Basket Analysis- Association Rule Mining with visualizations. *Market Basket Analysis- Association Rule Mining with visualizations*.

Retrieved May 26, 2022, from https://medium.com/analytics-vidhya/market-basket-analysis-association-rule-mining-with-visualizations-cda24d537019

Sivek, S. C. (2020, November). Market Basket Analysis 101: Key Concepts. *Market Basket Analysis 101: Key Concepts*. Retrieved May 26, 2022, from https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00

Tang, L., Li, J., Du, H., Li, L., Wu, J., & Wang, S. (2022, February 28). Big Data in Forecasting Research: A Literature Review. *Big Data Research, 27*, 100289. doi:10.1016/j.bdr.2021.100289

Yıldırım, S. (2020, October). Market Basket Analysis with Pandas. *Market Basket Analysis with Pandas*. Retrieved May 24, 2022, from https://towardsdatascience.com/market-basket-analysis-with-pandas-246fb8ee10a5