D212 Task 3 Association Rules and Lift Analysis

Western Governs University

Table of Contents

Part I: Research Question

A1. Proposal of Question

What are the most frequently purchased items by customers of a telecommunication company, and how can this information be used to improve product placement and promotional strategies?

A2. Defined Goal

One goal of the market basket analysis would be to identify the specific services or products that are most strongly bought together for increased placement and promotional strategies.

Part II: Market Basket Justification

B1. Explanation of Market Basket

Market basket analysis is a technique used in data mining to identify patterns and between different items that customers purchase together (Saini, n.d.). It involves analyzing the data of customer transactions to determine the items that are frequently bought together and the strength of the relationship between them. The analysis will reveal item combinations that are frequently purchased together, providing insights into customer preferences, and purchasing patterns.

B2. Transaction Example

An example of a transaction from the data set:

Dust-Off Compressed Gas 2 pack	Screen Mom Screen Cleaner kit	Moread HDMI to VGA Adapter	HP 62XL Tri-Color ink	Apple USB-C Charger cable

B3. Market Basket Assumption

One assumption of market basket analysis is that there is some degree of correlation between the items being analyzed. Such that customers who purchase one item are more likely to purchase another item as well. The analysis assumes that the data is complete and accurate.

Part III: Data Preparation and Analysis

C1. Transforming the Data Set

The following steps were taken to transform the data set:

1. view contents of the data frame

```
#content of df
df
```

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	10ft iPHone Charger Cable	HP 902XL ink	Creative Pebble 2.0 Speakers	Cleaning Gel Universal Dust Cleaner	Micro Center 32GB Memory card	YUNSONG 3pack 6ft Nylon Lightning Cable	TopMate C5 Laptop Cooler pad	Apple USB-C Charger cable	HyperX Cloud Stinger Headset	TONOR USB Gaming Microphone	Dust-Off Compressed Gas 2 pack
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14997	Falcon Dust Off Compressed Gas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14998	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14999	HP 63XL Ink	Apple USB-C Charger cable	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15001	Apple Pencil	SanDisk Ultra 128GB card	RUNMUS Gaming Headset	TopMate C5 Laptop Cooler pad	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2. Show dimensions of the data frame

```
#dimensions of df
df.shape
```

```
(15002, 20)
```

3. Show statistical summary of data

```
#statistical summary
df.describe
```

```
<bound method NDFrame.describe of                                                Item01  \
1                    Logitech M510 Wireless mouse
3           Apple Lightning to Digital AV Adapter
5       UNEN Mfi Certified 5-pack Lightning Cable
7                            Cat8 Ethernet Cable
9                  Dust-Off Compressed Gas 2 pack
...                                          ...
14993                SanDisk 32GB Ultra SDHC card
14995       Apple Lightning to Digital AV Adapter
14997               Falcon Dust Off Compressed Gas
14999                                  HP 63XL Ink
15001                                Apple Pencil

                                    Item02                       Item03  \
1                                HP 63 Ink                    HP 65 ink
3          TP-Link AC1750 Smart WiFi Router                  Apple Pencil
5                                      NaN                          NaN
7                                HP 65 ink                          NaN
9             Screen Mom Screen Cleaner kit    Moread HDMI to VGA Adapter
...                                    ...                          ...
14993                 Vsco 70 pack stickers  SanDisk 128GB microSDXC card
14995  Nylon Braided Lightning to USB cable                  Apple Pencil
14997                                  NaN                          NaN
14999               Apple USB-C Charger cable                       NaN
15001                SanDisk Ultra 128GB card         RUNMUS Gaming Headset

                              Item04                           Item05  \
1         nonda USB C to USB Adapter          10ft iPHone Charger Cable
3                                NaN                              NaN
5                                NaN                              NaN
7                                NaN                              NaN
9             HP 62XL Tri-Color ink          Apple USB-C Charger cable
...                             ...                              ...
14993                            NaN                              NaN
14995         USB 2.0 Printer cable  ARRIS SURFboard SB8200 Cable Modem
14997                            NaN                              NaN
14999                            NaN                              NaN
15001  TopMate C5 Laptop Cooler pad                               NaN
```

4. Show missing values

```
#Show number of missing values
df.isnull().sum()
```

```
Item01     7501
Item02     9255
Item03    10613
Item04    11657
Item05    12473
Item06    13138
Item07    13633
Item08    14021
Item09    14348
Item10    14607
Item11    14746
Item12    14848
Item13    14915
Item14    14955
Item15    14977
Item16    14994
Item17    14998
Item18    14998
Item19    14999
Item20    15001
dtype: int64
```

5. Drop missing values

```
#Drop missing values
df.dropna(how='all', inplace=True)
```

6. show data fram again

```
df
```

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	It
1	M310 Wireless mouse	HP 65 Ink	HP 65 ink	nonda USB C to USB Adapter	10ft iPHone Charger Cable	HP 902XL ink	Creative Pebble 2.0 Speakers	Cleaning Gel Universal Dust Cleaner	Micro Center 32GB Memory card	YUNSONG 3pack 6ft Nylon Lightning Cable	TopMate C5 Laptop Cooler pad	Apple USB-C Charger cable	HyperX Cloud Stinger Headset	TONOR USB Gaming Microphone	Du Compr Gas 2
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	UNEN Mfi Certified 5-pack Lightning Cable	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
7	Cat8 Ethernet Cable	HP 65 ink	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
9	Dust-Off Compressed Gas 2 pack	Screen Mom Screen Cleaner kit	Moread HDMI to VGA Adapter	HP 62XL Tri-Color ink	Apple USB-C Charger cable	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
14993	SanDisk 32GB Ultra SDHC card	Vsco 70 pack stickers	SanDisk 128GB microSDXC card	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
14995	Apple Lightning to Digital AV Adapter	Nylon Braided Lightning to USB cable	Apple Pencil	USB 2.0 Printer cable	ARRIS SURFboard SB8200 Cable Modem	Apple USB-C Charger cable	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
14997	Falcon Dust Off Compressed Gas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

7. Show info of data frame

```
#summary of df
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7501 entries, 1 to 15001
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: 1.2+ MB
```

8. Drop rows with missing values

```
df.dropna(subset=['Item01'], inplace=True)
df_list = df.T.apply(lambda x: x.dropna().tolist()).tolist()
```

9. Save to new data frame

```
#save to new data frame
df_cleaned = pd.DataFrame(df_list)
```

10. Show contents of new data frame

```
df.cleaned
```

11. Show number of columns in data frame

```
df_cleaned.columns
```

12. Save cleaned file to csv

```
df_cleaned.to_csv('D212_telco_cleaned')
```

C2. Code Execution

1. #install mlxtend to use 'TransactionEncoder'

```
!pip install mlxtend
```

```
Collecting mlxtend
  Downloading mlxtend-0.22.0-py2.py3-none-any.whl (1.4 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.4/1.4 MB 5.8 MB/s eta 0:00:00a 0:00:01
click to scroll output; double click to hide
Requirement already satisfied: joblib>=0.13.2 in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (1.2.0)
Requirement already satisfied: pandas>=0.24.2 in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (1.5.3)
Requirement already satisfied: scikit-learn>=1.0.2 in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (1.
2.1)
Requirement already satisfied: scipy>=1.2.1 in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (1.10.0)
Requirement already satisfied: matplotlib>=3.0.0 in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (3.6.
3)
Requirement already satisfied: setuptools in ./opt/anaconda3/lib/python3.9/site-packages (from mlxtend) (65.6.3)
Requirement already satisfied: contourpy>=1.0.1 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.0.
0->mlxtend) (1.0.7)
Requirement already satisfied: fonttools>=4.22.0 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.
0.0->mlxtend) (4.38.0)
Requirement already satisfied: cycler>=0.10 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.0.0->m
lxtend) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.0.0->
mlxtend) (9.4.0)
Requirement already satisfied: pyparsing>=2.2.1 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.0.
0->mlxtend) (3.0.9)
Requirement already satisfied: packaging>=20.0 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.0.0
->mlxtend) (23.0)
Requirement already satisfied: kiwisolver>=1.0.1 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=3.
0.0->mlxtend) (1.4.4)
Requirement already satisfied: python-dateutil>=2.7 in ./opt/anaconda3/lib/python3.9/site-packages (from matplotlib>=
3.0.0->mlxtend) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in ./opt/anaconda3/lib/python3.9/site-packages (from pandas>=0.24.2->mlxt
end) (2022.7)
Requirement already satisfied: threadpoolctl>=2.0.0 in ./opt/anaconda3/lib/python3.9/site-packages (from scikit-learn
>=1.0.2->mlxtend) (3.1.0)
Requirement already satisfied: six>=1.5 in ./opt/anaconda3/lib/python3.9/site-packages (from python-dateutil>=2.7->ma
tplotlib>=3.0.0->mlxtend) (1.16.0)
Installing collected packages: mlxtend
Successfully installed mlxtend-0.22.0
```

2. #OneHot Encoding

```
encoder = TransactionEncoder()
onehot = encoder.fit(df_list).transform(df_list)
onehot = pd.DataFrame(onehot, columns=encoder.columns_)
```

3. #save Oneone-hot encoded boolean to new file

```
onehot.to_csv('onehot.csv')
```

4. # Use the Apriori algorithm to generate frequent itemsets

```
frequent_itemsets = apriori(onehot, use_colnames=True, min_support=0.001, max_len
=3)
```

5. #generate a df of association rules from the frequent itemsets generated by the Apriori algorithm

```
rules = association_rules(frequent_itemsets)
rules
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(10ft iPHone Charger Cable 2 Pack, VicTsing Wi...	(Dust-Off Compressed Gas 2 pack)	0.002000	0.238368	0.001866	0.933333	3.915511	0.001390	11.424477	0.746097
1	(iPhone 11 case, 5pack Nylon Braided USB C cab...	(HP 63XL Ink)	0.002666	0.079323	0.002533	0.950000	11.976387	0.002321	18.413545	0.918953
2	(Cat8 Ethernet Cable, Anker 2-in-1 USB Card Re...	(Dust-Off Compressed Gas 2 pack)	0.001733	0.238368	0.001466	0.846154	3.549776	0.001053	4.950607	0.719539
3	(Anker 2-in-1 USB Card Reader, TP-Link AC1750 ...	(Screen Mom Screen Cleaner kit)	0.001600	0.129583	0.001333	0.833333	6.430898	0.001126	5.222504	0.845854
4	(Brother Genuine High Yield Toner Cartridge, F...	(VIVO Dual LCD Monitor Desk mount)	0.001466	0.174110	0.001200	0.818182	4.699220	0.000945	4.542394	0.788355
5	(Cat8 Ethernet Cable, SanDisk Extreme 256GB card)	(Dust-Off Compressed Gas 2 pack)	0.001466	0.238368	0.001200	0.818182	3.432428	0.000850	4.188975	0.709702
6	(FEEL2NICE 5 pack 10ft Lighning cable, SanDisk...	(Dust-Off Compressed Gas 2 pack)	0.001200	0.238368	0.001067	0.888889	3.729058	0.000781	6.854686	0.732715
7	(Premium Nylon USB Cable, SanDisk Ultra 400GB ...	(Dust-Off Compressed Gas 2 pack)	0.001466	0.238368	0.001200	0.818182	3.432428	0.000850	4.188975	0.709702
8	(SanDisk Extreme 256GB card, SanDisk Ultra 64G...	(Dust-Off Compressed Gas 2 pack)	0.001866	0.238368	0.001600	0.857143	3.595877	0.001155	5.331422	0.723254
9	(iPhone 11 case, Screen Mom Screen Cleaner kit)	(Logitech M510 Wireless mouse)	0.001866	0.071457	0.001600	0.857143	11.995203	0.001466	6.499800	0.918347
10	(SAMSUNG 128GB card, SanDisk Ultra 128GB card)	(Screen Mom Screen Cleaner kit)	0.001466	0.129583	0.001200	0.818182	6.313973	0.001010	4.787295	0.842857
11	(Stylus Pen for iPad, SanDisk 128GB card)	(VIVO Dual LCD Monitor Desk mount)	0.002133	0.174110	0.001733	0.812500	4.666587	0.001362	4.404746	0.787390

C3. Association Rules Table

Values for the support, lift, and confidence of the association rules table:

values_1 = (rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']]) (*Selvaraj, 2023)*

values_1

	antecedents	consequents	support	confidence	lift
0	(10ft iPHone Charger Cable 2 Pack, VicTsing Wi...	(Dust-Off Compressed Gas 2 pack)	0.001866	0.933333	3.915511
1	(iPhone 11 case, 5pack Nylon Braided USB C cab...	(HP 63XL Ink)	0.002533	0.950000	11.976387
2	(Cat8 Ethernet Cable, Anker 2-in-1 USB Card Re...	(Dust-Off Compressed Gas 2 pack)	0.001466	0.846154	3.549776
3	(Anker 2-in-1 USB Card Reader, TP-Link AC1750 ...	(Screen Mom Screen Cleaner kit)	0.001333	0.833333	6.430898
4	(Brother Genuine High Yield Toner Cartridge, F...	(VIVO Dual LCD Monitor Desk mount)	0.001200	0.818182	4.699220
5	(Cat8 Ethernet Cable, SanDisk Extreme 256GB card)	(Dust-Off Compressed Gas 2 pack)	0.001200	0.818182	3.432428
6	(FEEL2NICE 5 pack 10ft Lighning cable, SanDisk...	(Dust-Off Compressed Gas 2 pack)	0.001067	0.888889	3.729058
7	(Premium Nylon USB Cable, SanDisk Ultra 400GB ...	(Dust-Off Compressed Gas 2 pack)	0.001200	0.818182	3.432428
8	(SanDisk Extreme 256GB card, SanDisk Ultra 64G...	(Dust-Off Compressed Gas 2 pack)	0.001600	0.857143	3.595877
9	(iPhone 11 case, Screen Mom Screen Cleaner kit)	(Logitech M510 Wireless mouse)	0.001600	0.857143	11.995203
10	(SAMSUNG 128GB card, SanDisk Ultra 128GB card)	(Screen Mom Screen Cleaner kit)	0.001200	0.818182	6.313973
11	(Stylus Pen for iPad, SanDisk 128GB card)	(VIVO Dual LCD Monitor Desk mount)	0.001733	0.812500	4.666587

C4. Top Three Rules

Top three rules generated by the Apriori algorithm:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
1	(iPhone 11 case, 5pack Nylon Braided USB C cab...	(HP 63XL Ink)	0.002666	0.079323	0.002533	0.950000	11.976387	0.002321	18.413545	0.918953
0	(10ft iPHone Charger Cable 2 Pack, VicTsing Wi...	(Dust-Off Compressed Gas 2 pack)	0.002000	0.238368	0.001866	0.933333	3.915511	0.001390	11.424477	0.746097
11	(Stylus Pen for iPad, SanDisk 128GB card)	(VIVO Dual LCD Monitor Desk mount)	0.002133	0.174110	0.001733	0.812500	4.666587	0.001362	4.404746	0.787390

Part IV: Data Summary and Implications

D1. Significance of Support, Lift, and Confidence Summary

In a market basket analysis, the support is the frequency of transactions in which the itemset appears. It represents the popularity of the itemset in the dataset. A higher support value indicates that the itemset is frequently bought together. For the first rule the support is 0.002666 means that the itemset appears in 0.2666% of all transactions. The first second rule the item set appears i9n0.2% of all transactions. For the third rule the itemset appears in 0.2133% of all transactions.

Confidence indicates how often the consequent items is purchased when the antecedent item in purchased (Sivek, 2021). For the first rule if a customer buys the itemset of (iPhone 11 case, 5pack Nylon Braided USB C cable, and 6ft Mfi Certified Lightning Cable, there is a 95% chance that they will also buy HP 63XL Ink. For the second rule if the customer buy the itemset then there is a 93% change that they will also buy Dust off Compressed Gas 2 pack. For

the third rule there is a 81.25% change the customer will also buy VIVO Dual LCD Monitor Desk mount if they buy the itemset.

Lift is the ratio of the observed support to the expected support if the items were independent of each other. It is the strength of the association between the antecedent and consequent item. A lift value greater than 1 indicates a positive correlation which would suggest that they are often purchased together (Sivek, 2021). For the first rule customers who buy these two sets are 11.976387 times more likely to buy both items than if the items were purchased separately. For the first rule customers who buy these two sets are 3.915511 times more likely to buy both items than if the items were purchased seperately. For the first rule customers who buy these two sets are 4.666587 times more likely to buy both items than if the items were purchased separately.

D2. Practical Significance of Findings

The practical significance of the findings from the market basket analysis:

1. If a customer purchases an iPhone 11 case and a 5-pack of nylon braided USB-C cables, they are likely to also purchase an HP 63XL ink cartridge, with a confidence of 95% and a lift of 12.

2. If a customer purchases a 10ft iPhone charger cable 2-pack and a VictSing wireless mouse, they are likely to also purchase a Dust-Off compressed gas 2-pack, with a confidence of 93% and a lift of 3.9.

3. If a customer purchases a stylus pen for an iPad and a SanDisk 128GB memory card, they are also likely to also purchase a VIVO dual LCD monitor desk mount, with a confidence of 81% and a lift of 4.7.

D3. Course of Action

Rule one shows that when iPhone 11 case and a 5-pack of nylon braided USB-C cables, they are very likely to also purchase an HP 63XL ink cartridge. This itemset has a high support, confidence, and lift. Suggesting they are frequently bought together and should be marketed together or placed next to each other in a store to encourage customers to buy them together.

For the rule 10ft ipHone Charger Cable 2 Pack, VicTsing Wireless Mouse and Dust-Off Compressed Gas 2 pack there is a high confidence but low support and lift. This suggests that while these items are commonly bought together, their purchase is not necessarily dependent on each other. The store could consider placing these items in different sections to give customers more options.

For the rule Stylus Pen for iPad, SanDisk 128GB card and VIVO Dual LCD Monitor Desk mount, there is a high confidence and low support and lift values. This suggests that while these items are commonly bought together, their purchase is not necessarily dependent on each other. The store could consider placing these items in different sections to give customers more options.

Part V. Attachments

E. Panopto Recording

Link to panopto recording:

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=16541395-3cb5-46d7-a150-affb017cbcb7

F. Web Sources

Selvaraj, N. (2023). How to Perform Market Basket Analysis in Python. *365 Data Science*. https://365datascience.com/tutorials/python-tutorials/market-basket-analysis/

G. Sources

Saini, H. (n.d.). *Market Basket Analysis: An Overview | Analytics Steps*. https://www.analyticssteps.com/blogs/market-basket-analysis-overview

Sivek, S. C., PhD. (2021, December 16). Market Basket Analysis 101: Key Concepts - Towards Data Science. *Medium*. https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00