

D212 Performance Assessment Task 3

July 28, 2022

Kiet Nguyen

ID: 001601720

Email: kngu179@wgu.edu

0.0.1 A. Purpose of Report

A1. Relevant Question Can we identify the items that were frequently purchased together?

A2. Analysis Goal The main goal of this analysis was to find associated items based on transactions. The relationships between associated items could help us discover the patterns of customer shopping behaviors, improve product recommendations, and increase sales through cross-selling (Hull, 2022).

0.0.2 B. Technique Justification

B1. Market Basket Explanation The market basket analysis is a process of identifying relationships between items that were frequently paired together. The analysis is built on association rules. These rules take the form of **if-then** statement between two sets of items. The first set is called the **antecedent** and the second set is the **consequent**. A simple example of an association rule could be **if milk, then cereal**. There could also be more complicated rules with multiple antecedents or consequents (Hull, 2022).

The expected outcome for this analysis would be a dataframe containing the association rules and their metrics. These metrics include:

- *Support*: the frequency of an itemset in transactions.
- *Confidence*: the probability that a customer will purchase Y, if they purchased X.
- *Lift*: the probability that two items would be purchased together more than they would individually.

B2. Transaction Example The first transaction in the dataset contained 20 items, one for each column:

- 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
- 3A USB Type C Cable 3 pack 6FT
- HOVAMP iPhone charger
- SanDisk Ultra 128GB card
- FEEL2NICE 5 pack 10ft Lightning cable
- FEIYOLD Blue light Blocking Glasses

B3. Technique Assumption The core algorithm of market basket analysis was the Apriori algorithm. This algorithm made the assumption that all subsets of a frequent itemset must be frequent. If an itemset is infrequent, all its supersets will be infrequent (Yadav, 2022).

0.0.3 C. Analysis

C1. Transform Dataset 1. Import libraries and dataset

```
[1]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

```
[2]: df = pd.read_csv('teleco_market_basket.csv')
```

```
[3]: df.shape
```

```
[3]: (15002, 20)
```

```
[4]: df.head()
```

```
[4]:
```

	Item01	Item02 \
0	NaN	NaN
1	Logitech M510 Wireless mouse	HP 63 Ink
2	NaN	NaN
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router
4	NaN	NaN

	Item03	Item04	Item05 \
0	NaN	NaN	NaN

1	HP 65 ink	nonda USB C to USB Adapter	10ft iPhone Charger Cable
2	NaN	NaN	NaN
3	Apple Pencil	NaN	NaN
4	NaN	NaN	NaN

	Item06	Item07 \
0	NaN	NaN
1	HP 902XL ink	Creative Pebble 2.0 Speakers
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item08	Item09 \
0	NaN	NaN
1	Cleaning Gel Universal Dust Cleaner	Micro Center 32GB Memory card
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item10	Item11 \
0	NaN	NaN
1	YUNSONG 3pack 6ft Nylon Lightning Cable	TopMate C5 Laptop Cooler pad
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item12	Item13 \
0	NaN	NaN
1	Apple USB-C Charger cable	HyperX Cloud Stinger Headset
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item14	Item15 \
0	NaN	NaN
1	TONOR USB Gaming Microphone	Dust-Off Compressed Gas 2 pack
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item16	Item17 \
0	NaN	NaN
1	3A USB Type C Cable 3 pack 6FT	HOVAMP iPhone charger
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item18	Item19 \
0	NaN	NaN
1	SanDisk Ultra 128GB card	FEEL2NICE 5 pack 10ft Lighning cable
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	Item20
0	NaN
1	FEIYOLD Blue light Blocking Glasses
2	NaN
3	NaN
4	NaN

2. Remove rows containing all null values

```
[5]: # Drop empty rows
df.dropna(how='all', inplace=True)
```

```
[6]: df.shape
```

```
[6]: (7501, 20)
```

3. Fill null values with zeroes

```
[7]: df.fillna(0, inplace=True)
```

4. Convert dataframe to list of lists

```
[8]: # Adapted from Association Rule Mining via Apriori Algorithm in Python (Malik, ↵
↵2022):
# https://stackabuse.com/
↵association-rule-mining-via-apriori-algorithm-in-python/

purchases = []
for row in range(len(df)):
    purchases.append([str(df.values[row, col]) for col in range(len(df.
↵columns))])
```

5. Export list as dataframe

```
[9]: teleco_clean = pd.DataFrame(purchases)

teleco_clean.head()
```

```
[9]:
```

	0 \
0	Logitech M510 Wireless mouse
1	Apple Lightning to Digital AV Adapter

2 UNEN Mfi Certified 5-pack Lightning Cable
 3 Cat8 Ethernet Cable
 4 Dust-Off Compressed Gas 2 pack

	1	2 \
0	HP 63 Ink	HP 65 ink
1	TP-Link AC1750 Smart WiFi Router	Apple Pencil
2	0	0
3	HP 65 ink	0
4	Screen Mom Screen Cleaner kit	Moread HDMI to VGA Adapter

	3	4	5 \
0	nonda USB C to USB Adapter	10ft iPhone Charger Cable	HP 902XL ink
1	0	0	0
2	0	0	0
3	0	0	0
4	HP 62XL Tri-Color ink	Apple USB-C Charger cable	0

	6	7 \
0	Creative Pebble 2.0 Speakers	Cleaning Gel Universal Dust Cleaner
1	0	0
2	0	0
3	0	0
4	0	0

	8	9 \
0	Micro Center 32GB Memory card	YUNSONG 3pack 6ft Nylon Lightning Cable
1	0	0
2	0	0
3	0	0
4	0	0

	10	11 \
0	TopMate C5 Laptop Cooler pad	Apple USB-C Charger cable
1	0	0
2	0	0
3	0	0
4	0	0

	12	13 \
0	HyperX Cloud Stinger Headset	TONOR USB Gaming Microphone
1	0	0
2	0	0
3	0	0
4	0	0

	14	15 \
--	----	------

0	Dust-Off Compressed Gas 2 pack	3A USB Type C Cable 3 pack 6FT
1	0	0
2	0	0
3	0	0
4	0	0

	16	17 \
0	HOVAMP iPhone charger	SanDisk Ultra 128GB card
1	0	0
2	0	0
3	0	0
4	0	0

	18	19
0	FEEL2NICE 5 pack 10ft Lighning cable	FEIYOLD Blue light Blocking Glasses
1	0	0
2	0	0
3	0	0
4	0	0

```
[10]: teleco_clean.to_csv('teleco_clean.csv', index=False)
```

C2. Generate Association Rules 1. Encode itemsets

```
[11]: # Adapted from Association Rules chapter of DataCamp (Hull, 2022):
# https://app.datacamp.com/learn/courses/market-basket-analysis-in-python

# One hot encode list of purchases
encoder = TransactionEncoder()
onehot = encoder.fit(purchases).transform(purchases)

# Convert encoded data to dataframe
encoded_data = pd.DataFrame(onehot, columns=encoder.columns_)

encoded_data.head()
```

```
[11]:      0  10ft iPhone Charger Cable  10ft iPhone Charger Cable 2 Pack \
0  False                             True                             False
1   True                             False                             False
2   True                             False                             False
3   True                             False                             False
4   True                             False                             False

      3 pack Nylon Braided Lightning Cable  3A USB Type C Cable 3 pack 6FT \
0                             False                             True
1                             False                             False
2                             False                             False
```

3		False	False
4		False	False
	5pack Nylon Braided USB C cables	ARRIS SURFboard SB8200 Cable Modem	\
0		False	False
1		False	False
2		False	False
3		False	False
4		False	False
	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	\
0		False	False
1		False	False
2		False	False
3		False	False
4		False	False
	Anker USB C to HDMI Adapter ...	hP 65 Tri-color ink	\
0		False ...	False
1		False ...	False
2		False ...	False
3		False ...	False
4		False ...	False
	iFixit Pro Tech Toolkit	iPhone 11 case	iPhone 12 Charger cable \
0		False	False
1		False	False
2		False	False
3		False	False
4		False	False
	iPhone 12 Pro case	iPhone 12 case	iPhone Charger Cable Anker 6ft \
0		False	False
1		False	False
2		False	False
3		False	False
4		False	False
	iPhone SE case	nonda USB C to USB Adapter	seenda Wireless mouse
0		False	True
1		False	False
2		False	False
3		False	False
4		False	False

[5 rows x 120 columns]

```
[12]: # Drop zeroes column from encoded dataframe
encoded_data.drop('0', axis=1, inplace=True)

encoded_data
```

```
[12]:      10ft iPhone Charger Cable  10ft iPhone Charger Cable 2 Pack  \
0                                True                                False
1                                False                               False
2                                False                               False
3                                False                               False
4                                False                               False
...                               ...                               ...
7496                             False                             False
7497                             False                             False
7498                             False                             False
7499                             False                             False
7500                             False                             False

      3 pack Nylon Braided Lightning Cable  3A USB Type C Cable 3 pack 6FT  \
0                                         False                                True
1                                         False                                False
2                                         False                                False
3                                         False                                False
4                                         False                                False
...                                         ...                               ...
7496                                     False                             False
7497                                     False                             False
7498                                     False                             False
7499                                     False                             False
7500                                     False                             False

      5pack Nylon Braided USB C cables  ARRIS SURFboard SB8200 Cable Modem  \
0                                         False                                False
1                                         False                                False
2                                         False                                False
3                                         False                                False
4                                         False                                False
...                                         ...                               ...
7496                                     False                             False
7497                                     False                                True
7498                                     False                             False
7499                                     False                             False
7500                                     False                             False

      Anker 2-in-1 USB Card Reader  Anker 4-port USB hub  \
0                                False                                False
1                                False                                False
```


2	False	False
3	False	False
4	False	False
...
7496	False	False
7497	False	False
7498	False	False
7499	False	False
7500	False	False

	Anker USB C to HDMI Adapter	Apple Lightning to Digital AV Adapter	...	\
0	False		False	...
1	False		True	...
2	False		False	...
3	False		False	...
4	False		False	...
...
7496	False		False	...
7497	False		True	...
7498	False		False	...
7499	False		False	...
7500	False		False	...

	hP 65 Tri-color ink	iFixit Pro Tech Toolkit	iPhone 11 case	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
...	
7496	False	False	False	
7497	False	False	False	
7498	False	False	False	
7499	False	False	False	
7500	False	False	False	

	iPhone 12 Charger cable	iPhone 12 Pro case	iPhone 12 case	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
...	
7496	False	False	False	
7497	False	False	False	
7498	False	False	False	
7499	False	False	False	

7500	False	False	False
	iPhone Charger Cable Anker 6ft	iPhone SE case \	
0	False	False	
1	False	False	
2	False	False	
3	False	False	
4	False	False	
...	
7496	False	False	
7497	False	False	
7498	False	False	
7499	False	False	
7500	False	False	

	nonda USB C to USB Adapter	seenda Wireless mouse
0	True	False
1	False	False
2	False	False
3	False	False
4	False	False
...
7496	False	False
7497	False	False
7498	False	False
7499	False	False
7500	False	False

[7501 rows x 119 columns]

2. Compute frequent itemsets

```
[13]: # Compute frequent itemsets
itemsets = apriori(encoded_data, min_support=0.005, max_len=2,
                    use_colnames=True)

itemsets
```

```
[13]:      support      itemsets
0    0.009065    (10ft iPhone Charger Cable)
1    0.050527    (10ft iPhone Charger Cable 2 Pack)
2    0.005199    (3 pack Nylon Braided Lightning Cable)
3    0.042528    (3A USB Type C Cable 3 pack 6FT)
4    0.019064    (5pack Nylon Braided USB C cables)
..      ...      ...
547  0.010265  (VicTsing Wireless mouse, VIVO Dual LCD Monito...
548  0.007466  (VIVO Dual LCD Monitor Desk mount, YUNSONG 3pa...
```

```

549 0.006666 (iPhone Charger Cable Anker 6ft, VIVO Dual LCD...
550 0.009865 (VIVO Dual LCD Monitor Desk mount, iPhone SE c...
551 0.005599 (nonda USB C to USB Adapter, VIVO Dual LCD Mon...

```

[552 rows x 2 columns]

3. Generate association rules

```

[14]: # Compute association rules from frequent itemsets
rules = association_rules(itemsets, metric='support', min_threshold=0.005)

```

C3. Association Rules Values

```

[15]: # Display association rules table
rules

```

```

[15]:
      antecedents \
0      (Anker USB C to HDMI Adapter)
1      (10ft iPhone Charger Cable 2 Pack)
2      (Apple Lightning to Digital AV Adapter)
3      (10ft iPhone Charger Cable 2 Pack)
4      (Apple Pencil)
..      ...
897     (VIVO Dual LCD Monitor Desk mount)
898     (VIVO Dual LCD Monitor Desk mount)
899     (iPhone SE case)
900     (nonda USB C to USB Adapter)
901     (VIVO Dual LCD Monitor Desk mount)

      consequents antecedent support \
0      (10ft iPhone Charger Cable 2 Pack)      0.068391
1      (Anker USB C to HDMI Adapter)      0.050527
2      (10ft iPhone Charger Cable 2 Pack)      0.087188
3      (Apple Lightning to Digital AV Adapter)      0.050527
4      (10ft iPhone Charger Cable 2 Pack)      0.179709
..      ...
897     (iPhone Charger Cable Anker 6ft)      0.174110
898     (iPhone SE case)      0.174110
899     (VIVO Dual LCD Monitor Desk mount)      0.026530
900     (VIVO Dual LCD Monitor Desk mount)      0.025730
901     (nonda USB C to USB Adapter)      0.174110

      consequent support      support      confidence      lift      leverage      conviction
0      0.050527      0.006932      0.101365      2.006162      0.003477      1.056572
1      0.068391      0.006932      0.137203      2.006162      0.003477      1.079755
2      0.050527      0.006266      0.071865      1.422329      0.001860      1.022991
3      0.087188      0.006266      0.124011      1.422329      0.001860      1.042035
4      0.050527      0.009065      0.050445      0.998387      -0.000015      0.999914

```

```

..          ...          ...          ...          ...          ...
897          0.025730  0.006666  0.038285  1.487951  0.002186  1.013055
898          0.026530  0.009865  0.056662  2.135771  0.005246  1.031942
899          0.174110  0.009865  0.371859  2.135771  0.005246  1.314817
900          0.174110  0.005599  0.217617  1.249879  0.001119  1.055608
901          0.025730  0.005599  0.032159  1.249879  0.001119  1.006643

```

[902 rows x 9 columns]

C4. Top Three Rules Support Rules

```
[16]: rules.sort_values('support', ascending=False).head(3)
```

```

[16]:          antecedents          consequents \
446  (VIVO Dual LCD Monitor Desk mount)  (Dust-Off Compressed Gas 2 pack)
447  (Dust-Off Compressed Gas 2 pack)  (VIVO Dual LCD Monitor Desk mount)
379  (Dust-Off Compressed Gas 2 pack)  (HP 61 ink)

          antecedent support  consequent support  support  confidence  lift \
446          0.174110          0.238368  0.059725  0.343032  1.439085
447          0.238368          0.174110  0.059725  0.250559  1.439085
379          0.238368          0.163845  0.052660  0.220917  1.348332

          leverage  conviction
446  0.018223  1.159314
447  0.018223  1.102008
379  0.013604  1.073256

```

Confidence Rules

```
[17]: rules.sort_values('confidence', ascending=False).head(3)
```

```

[17]:          antecedents          consequents \
423  (SanDisk Extreme 256GB card)  (Dust-Off Compressed Gas 2 pack)
366  (DisplayPort ot HDMI adapter)  (Dust-Off Compressed Gas 2 pack)
170  (Apple Lightning to USB cable)  (Dust-Off Compressed Gas 2 pack)

          antecedent support  consequent support  support  confidence  lift \
423          0.010399          0.238368  0.005066  0.487179  2.043811
366          0.011998          0.238368  0.005733  0.477778  2.004369
170          0.015598          0.238368  0.007332  0.470085  1.972098

          leverage  conviction
423  0.002587  1.485182
366  0.002873  1.458444
170  0.003614  1.437273

```

Lift Rules

```
[18]: rules.sort_values('lift', ascending=False).head(3)
```

```
[18]:
```

	antecedents	consequents	antecedent support	\
639	(iPhone 11 case)	(HP 63XL Ink)	0.015731	
638	(HP 63XL Ink)	(iPhone 11 case)	0.079323	
693	(iPhone 11 case) (Logitech M510 Wireless mouse)		0.015731	

	consequent support	support	confidence	lift	leverage	conviction
639	0.079323	0.005866	0.372881	4.700812	0.004618	1.468107
638	0.015731	0.005866	0.073950	4.700812	0.004618	1.062867
693	0.071457	0.005066	0.322034	4.506672	0.003942	1.369601

0.0.4 D. Analysis Summary

D1. Importance of Metrics *Support* is the frequency that an itemset will occur in transactions. It is calculated by dividing the total number of transactions, which gives us an idea of how popular a particular itemset is.

$$Support = \frac{freq(X \& Y)}{Total}$$

Confidence is the probability that a customer will purchase an item given a previous purchase. It tells you how often items X and Y are purchased together if you have already purchased X.

$$Confidence = \frac{Support(X \& Y)}{Support(X)}$$

Lift is the ratio of two items purchased together divided by two items purchased individually. A value greater than one tells you that two items occur in transactions together more often than you expect based on their individual support values. This means the relationship is unlikely to be explained by random chance (Hull, 2022).

$$Lift = \frac{Support(X \& Y)}{Support(X) \cdot Support(Y)}$$

D2. Practical Significance The practical significance of the analysis was to derive useful patterns that we could apply in our retail setting. Based on the findings of the analysis, we could improve sales by grouping items together based on the association rules. For example, the top rules for *support* and *confidence* showed that customers frequently purchased compressed gas together with other computer accessories. By grouping the compressed gas into the computer accessories category, we could cross-sell those related items.

D3. Recommendation We offer two suggestions for the company based on our analysis. The first suggestion is to change the store layout by placing related items close together. This would improve cross-selling as customers would be more likely to purchase both items together. Even a small increase in sales would improve our bottom line over time. The second suggestion is to build a recommendation system for our online store. This could be something simple where customers

are offered the choice to add related items to their shopping carts before checking out. Impulse purchases can become the driver for additional sales.

0.0.5 E. Panopto Recording

Link: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=8956fbea-f316-4810-a068-ae0007a8345>

0.0.6 F. Third-Party Code

Malik, U. (2022, July 21). Association Rule Mining via Apriori Algorithm in Python. Stack Abuse. Retrieved July 21, 2022, from <https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>

0.0.7 G. References

Hull, I. (2022). Market Basket Analysis in Python. DataCamp. Retrieved July 20, 2022, from <https://app.datacamp.com/learn/courses/market-basket-analysis-in-python>

Yadav, M. (2022, January 13). Apriori Algorithm. GeeksforGeeks. Retrieved July 20, 2022, from <https://www.geeksforgeeks.org/apriori-algorithm/>