

# D212 Performance Assessment Task 3

This is my performance assessment for task 3.  
012047746

## A1:PROPOSAL OF QUESTION

My research question for this task is "Is it possible to create a market basket analysis model using apriori to produce a market basket analysis to see what customers typically buy together?".

## A2:DEFINED GOAL

My goal for this analysis is to do a market basket analysis on the teleco market basket analysis data and output a result showing the most common items that are bought together. I will show this through showing top three of specific metrics: lift, support, and confidence.

## B1:EXPLANATION OF MARKET BASKET

Market basket analysis is an analysis that is performed to measure the likelihood of a group of items being bought together. This is calculated using basket data or data of what items that customers have bought in one go together. Using probabilities, it is possible to calculate the likelihood of certain items being bought with one another. From this analysis, different business actions like restructuring store layouts can be taken to further optimize profit by increasing likelihood of customers buying another item.

## B2:TRANSACTION EXAMPLE

```
In [235... # pip insall for mlxtend to encode transactions  
# pip install mlxtend
```

```
In [236... # import the libraries  
import pandas as pd  
from pandas import DataFrame  
import numpy as np  
from mlxtend.preprocessing import TransactionEn  
from mlxtend.frequent_patterns import apriori,
```

```
In [237... df = pd.read_csv('teleco_market_basket.csv')
```

```
In [238... df = df.dropna(how='all')
```

```
In [239... df
```

Out [239]:

	Item01	Item02	Item03	Item04	I
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	C
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	
5	UNEN Mfi Certified 5-pack Lightning Cable	NaN	NaN	NaN	
7	Cat8 Ethernet Cable	HP 65 ink	NaN	NaN	
9	Dust-Off Compressed Gas 2 pack	Screen Mom Screen Cleaner kit	Moread HDMI to VGA Adapter	HP 62XL Tri-Color ink	C
...	...	...	...	...	
14993	SanDisk 32GB Ultra SDHC card	Vsco 70 pack stickers	SanDisk 128GB microSDXC card	NaN	
14995	Apple Lightning to Digital AV Adapter	Nylon Braided Lightning to USB cable	Apple Pencil	USB 2.0 Printer cable	SUR S I

	Item01	Item02	Item03	Item04	I
14997	Falcon Dust Off Compressed Gas	NaN	NaN	NaN	
14999	HP 63XL Ink	Apple USB-C Charger cable	NaN	NaN	
15001	Apple Pencil	SanDisk Ultra 128GB card	RUNMUS Gaming Headset	TopMate C5 Laptop Cooler pad	

7501 rows × 20 columns

Here is one example of a transaction in the dataset below. This shows all 20 items that were purchased in this transaction by the customer together.

```
In [240... df[:1]
```

Out[240]:

	Item01	Item02	Item03	Item04	Item05	Item06	
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	10ft iPhone Charger Cable	HP 902XL ink	:

## B3:MARKET BASKET ASSUMPTION

Some assumptions of market basket analysis are that:

- Each entry in the dataset represents an independent transaction
- The sample is representative of the customerbase as a whole
- There are no significant changes in shopping behavior as a whole. Something like this that could affect the analysis is if there is a sudden toilet paper shortage that causes panic buying among customers. That may disproportionately affect the analysis in regards to toilet papers for this data set.

## C1:TRANSFORMING THE DATA SET

Here are the following steps for analysis:

1. First we need to start by converting importing the data and converting to a list. A for loop is used to iterate over each row in the dataframe to remove the index and just get the item details. The list is initiated blank and then each item is grabbed as

long as it is not null and inputted into a list within the list. These are only inputted.

2. Next we use TransactionEncoder to one hot encode our dataframe so we can input this into the apriori model. We begin this by fitting our dataframe using `.fit()`
3. After we fit the dataframe, we use transform to construct an array of one hot encoded transactions
4. We then use this array to construct a dataframe that has each item as a column and a true or false depending on whether they were included in a basket together

```
In [241... transactions = []  
for index, row in df.iterrows():  
    items = [item for item in row.values if pd.  
    transactions.append(items)
```

```
In [242... transactions[:5]
```

```
Out[242]: [['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 Lighning cable',
            'FEIYOLD Blue light Blocking Glasses'],
            ['Apple Lightning to Digital AV Adapter',
            'TP-Link AC1750 Smart WiFi Router',
            'Apple Pencil'],
            ['UNEN Mfi Certified 5-pack Lightning Cable'],
            ['Cat8 Ethernet Cable', 'HP 65 ink'],
            ['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']]
```

```
In [243... # set up encoder
encoder = TransactionEncoder().fit(transactions
```

```
In [244... # set up array
onehot = encoder.transform(transactions)
```

```
In [245... # set up df  
data = pd.DataFrame(onehot, columns = encoder.c
```

```
In [246... print(data)
```



	10ft iPhone Charger Cable	10ft iPhone Ch
arger 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 S
URFboard 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 Lightn
ing 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 Tool
kit	iPhone 11 case \	
0		False
lse	False	Fa
1		False
lse	False	Fa
2		False
lse	False	Fa
3		False
lse	False	Fa
4		False
lse	False	Fa
...		...
...	...	
7496		False
lse	False	Fa
7497		False
lse	False	Fa
7498		False
lse	False	Fa
7499		False
lse	False	Fa
7500		False
lse	False	Fa

	iPhone 12 Charger cable	iPhone 12 Pro ca
se	iPhone 12 case \	
0		False
se	False	Fal
1		False
se	False	Fal
2		False
se	False	Fal
3		False
se	False	Fal

4		False	Fal
se	False		
...		...	
...	...		
7496		False	Fal
se	False		
7497		False	Fal
se	False		
7498		False	Fal
se	False		
7499		False	Fal
se	False		
7500		False	Fal
se	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 Wirele
ss 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]

```

```

In [247... # calculate the support metric by computing the
print(onehot.mean())

0.03288973234941223

```

The cleaned dataset is a list that contain each item in the original csv file as a column header. Then for each transaction, the item is marked true if it was part of that transaction or false if it was not. This results in us knowing that there are 7501 rows or transaction and 119 items

```
In [248... data.head()
```

Out[248]:

	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	AF SURFbo SB8 C Mo
0	True	False	False	True	False	F
1	False	False	False	False	False	F
2	False	False	False	False	False	F
3	False	False	False	False	False	F
4	False	False	False	False	False	F

5 rows × 119 columns

```
In [249... df.to_csv('task3_prepared_data.csv')
```

## C2:CODE EXECUTION

Here I will run the code used to generate the association rules for the table. We will take instruction on how to develop the rules from the Data Mining II task 3 lecture.

```
In [250... rules = apriori(data, min_support=.02, use_coln
```

```
In [251... rules.head()
```

Out [251]:

	support	itemsets
<b>0</b>	0.050527	(10ft iPhone Charger Cable 2 Pack)
<b>1</b>	0.042528	(3A USB Type C Cable 3 pack 6FT)
<b>2</b>	0.029463	(Anker 2-in-1 USB Card Reader)
<b>3</b>	0.068391	(Anker USB C to HDMI Adapter)
<b>4</b>	0.087188	(Apple Lightning to Digital AV Adapter)

In [252... `rule_table = association_rules(rules, metric =`

## C3:ASSOCIATION RULES TABLE

This is the association rule table that was created for each item and their consequent

In [253... `rule_table.head()`



Out [253]:

	antecedents	consequents	antecedent support	consequent support
0	(10ft iPhone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368
1	(Dust-Off Compressed Gas 2 pack)	(10ft iPhone Charger Cable 2 Pack)	0.238368	0.050527
2	(Dust-Off Compressed Gas 2 pack)	(Anker USB C to HDMI Adapter)	0.238368	0.068391
3	(Anker USB C to HDMI Adapter)	(Dust-Off Compressed Gas 2 pack)	0.068391	0.238368
4	(VIVO Dual LCD Monitor Desk mount)	(Anker USB C to HDMI Adapter)	0.174110	0.068391

## C4:TOP THREE RULES

The three rules that we have listed here are confidence, lift, and support.

In [254...]

```
top_three_rules_conf = rule_table.sort_values('top_three_rules_conf'
```

Out [254]:

	antecedents	consequents	antecedent support	consequent support
<b>0</b>	(10ft iPhone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368
<b>36</b>	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065858	0.238368
<b>52</b>	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368

In [255]:

```
top_three_rules_lift = rule_table.sort_values('top_three_rules_lift')
```

Out [255]:

	antecedents	consequents	antecedent support	consequent support
<b>84</b>	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.174110	0.098254
<b>85</b>	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.098254	0.174110
<b>65</b>	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.065858	0.174110

In [256]:

```
top_three_rules_supp = rule_table.sort_values('top_three_rules_supp')
```

Out [256]:

	antecedents	consequents	antecedent support	consequent support
62	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368
63	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845

## D1:SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY

The definitions for each of these are:

- Confidence: Confidence gives us the probability that we'll purchase Y, given that we have purchased X. The significance of this is that it can give us an understanding of the relationships between two types of items. In regards to our analysis, we can see that there are three items that have over .41 confidence level. These items should be put in proximity to one another and be made easier to purchase with on another.
- Lift: Provides us with another metric for evaluating relationships between items. A lift value of greater than 1 tells us that two items

occur in transactions more often than we would expect based just off of their support values. The significance of this metrics is that it can give us insight into relationships between items that we may miss if we are just looking at support. In regards to our analysis, we can see that there are two items with lift over 2 and one with a near 2 lift value. This means that we should recognize those relationships as significant and take action to increase customer accessibility to both of these items together.

- Support: The support metric measures the share of transactions that contain an itemset. Its calculated by dividing the number of transactions with an item by the total number of transactions. This is significant in giving us an understanding of how often this item is being bought and the popularity of the item. In regards to our analysis, it shows us the biggest bundles of items that are bought together.

## D2:PRACTICAL SIGNIFICANCE OF FINDINGS

The finding for confidence, lift, and support were the following:

We can see that our top three antecedents and consequents involve the same consequent of "Dust-

Off Compressed Gas 2 pack". We can see that "10ft iPhone Charger Cable 2 Pack" and "Dust-Off Compressed Gas 2 pack" have the highest confidence. The "FEIYOLD Blue light Blocking Glasses" and "Dust-Off Compressed Gas 2 pack" have the second highest confidence. Finally the "SanDisk Ultra 64GB card" and "Dust-Off Compressed Gas 2 pack" have the third highest confidence. This is significant because we can consider that maybe these four items should be next to one another for shoppers to easily buy them together.

```
In [257... top_three_rules_conf[['antecedents', 'consequent
```

```
Out [257]:
```

	<b>antecedents</b>	<b>consequents</b>	<b>confidence</b>
<b>0</b>	(10ft iPhone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.456464
<b>36</b>	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.419028
<b>52</b>	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.416554

There was an item that was involved in lift for all of the three highest lift metrics. That item is "VIVO Dual LCD Monitor Desk mount". The highest lift antecedent and consequent is "VIVO Dual LCD Monitor Desk mount" and "SanDisk Ultra 64GB card".

The second highest lift is "SanDisk Ultra 64GB card" and "VIVO Dual LCD Monitor Desk mount". The third highest is "FEIYOLD Blue light Blocking Glasses" and "VIVO Dual LCD Monitor Desk mount".

```
In [258]: top_three_rules_lift[['antecedents', 'consequent
```

```
Out[258]:
```

	antecedents	consequents	lift
<b>84</b>	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	2.291162
<b>85</b>	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	2.291162
<b>65</b>	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	1.999758

The antecedent and consequent with the highest support was "VIVO Dual LCD Monitor Desk mount" and "Dust-Off Compressed Gas 2 pack". The second highest support was "Dust-Off Compressed Gas 2 pack" and "VIVO Dual LCD Monitor Desk mount". The third highest was "Dust-Off Compressed Gas 2 pack" and "HP 61 ink". This is significant because "Dust-Off Compressed Gas 2 pack" is either a antecedent or consequent in each of the top three highest support metrics. An insight that we can take from this is that the "Dust-Off Compressed Gas 2 pack" should be in an easily accessible place to customers as its a high

demand item alongside the other antecedents and consequents.

```
In [259]: top_three_rules_supp[['antecedents', 'consequent
```

	antecedents	consequents	support
62	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.059725
63	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.059725
41	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.052660

## D3: COURSE OF ACTION

This information is very useful for our stakeholders. When it comes to in-store placement of items, we can take the insights that we derive from the top support, lift, and confidence metrics and use that to assist in decision making when it comes to planning things like this. If its an online setting, we can create reccomendation systems that are built off of systems like this to encourage shoppers to buy an item that they may find useful alongside what they are buying but hadn't thought of it when they are purchasing. This could also be involved in decisions around deals for bundling, since we know that certain items are

already buying certain items together. Maybe then more customers would be willing to buy another item alongside another specific item. Overall, these insights us stakeholders in making decisions around how to frame and sell items to customers.

## G:SOURCES FOR THIRD-PARTY CODE

<https://app.datacamp.com/learn/courses/market-basket-analysis-in-python>

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=db85c4f1-0da5-4bde-a1a4-b07c0019d46d>

## H:SOURCES

<https://medium.com/@chemistry8526/boosting-sales-with-data-the-power-of-market-basket-analysis-in-retail-c79cc10a14df>

<https://www.geeksforgeeks.org/implementing-apriori-algorithm-in-python/>