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.

For this analysis I will be using the apriori model. The apriori model depends on calculating support. Support is the total number of transactions for a specific item (or items together) divided by total number of transaction. Off of this support, we can start building other insights that gives us insight into the relationship between different items. For instance, lift can be calculated by the support of the antecedent and consequent together, divided by the support of the consequent and antecedent. Another example is confidence which is the frequency of the the antecedent and the consequent together over the frequency of the antecedent.

B2:TRANSACTION EXAMPLE

In [2]:
```
# pip insall for mlxtend to encode transactions
# pip install mlxtend
```

In [3]:
```
# 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 [4]:
```
df = pd.read_csv('teleco_market_basket.csv')
```

In [5]:
```
df = df.dropna(how='all')
```

In [6]:
```
df
```

Out[6]:

	Item01	Item02	Item03	Item04	It
1	Logitech M510 Wireless mouse	HP 63 Ink	HP 65 ink	nonda USB C to USB Adapter	iF Ch
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	U Ch
...	...	...	...	...	
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	SURF SE M

	Item01	Item02	Item03	Item04	It
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 [7]:
```
df[:1]
```

Out[7]:

	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	(S

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 signficant 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 [8]:
```
transactions = []
for index, row in df.iterrows():
    items = [item for item in row.values if pd.
    transactions.append(items)
```

In [9]:
```
transactions[:5]
```

Out[9]:
```
[['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 [10]:
```
# set up encoder
encoder = TransactionEncoder().fit(transactions
```

In [11]:
```
# set up array
onehot = encoder.transform(transactions)
```

In [12]:
```
# set up df
data = pd.DataFrame(onehot, columns = encoder.c
```

In [13]:
```
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-por
t 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                     Fa
lse           False
1                   False                     Fa
lse           False
2                   False                     Fa
lse           False
3                   False                     Fa
lse           False
4                   False                     Fa
lse           False
...                   ...
...             ...
7496                False                     Fa
lse           False
7497                False                     Fa
lse           False
7498                False                     Fa
lse           False
7499                False                     Fa
lse           False
7500                False                     Fa
lse           False

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

```
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 [14]:
```python
# calculate the support metric by computing the
print(onehot.mean())
```

```
0.03288973234941223
```

The cleaned dataset is a list that contain each itmem 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 [15]:
```
data.head()
```

Out[15]:

	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	ARI SURFboa SB82 Cal Mod
0	True	False	False	True	False	Fa
1	False	False	False	False	False	Fa
2	False	False	False	False	False	Fa
3	False	False	False	False	False	Fa
4	False	False	False	False	False	Fa

5 rows × 119 columns

In [16]:
```
data.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 [17]:
```
rules = apriori(data, min_support=.02, use_coln
```

In [18]:
```
rules.head()
```

Out[18]:

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

In [19]:

```
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 [20]:

```
rule_table.head()
```

Out[20]:

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

In [21]:
```
rule_table.to_csv('task3_association_table.csv'
```

C4:TOP THREE RULES

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

In [22]:
```
top_three_rules_conf = rule_table.sort_values('
top_three_rules_conf
```

Out[22]:

	antecedents	consequents	antecedent support	consequent support
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.050527	0.238368
37	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.065858	0.238368
53	(SanDisk Ultra 64GB card)	(Dust-Off Compressed Gas 2 pack)	0.098254	0.238368

In [23]:

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

Out[23]:

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

In [24]:

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

Out[24]:

	antecedents	consequents	antecedent support	consequent support
62	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110
63	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368
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 [25]:
```
top_three_rules_conf[['antecedents','consequent
```

Out[25]:

	antecedents	consequents	confidence
0	(10ft iPHone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.456464
37	(FEIYOLD Blue light Blocking Glasses)	(Dust-Off Compressed Gas 2 pack)	0.419028
53	(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 [26]:
```
top_three_rules_lift[['antecedents','consequent
```

Out[26]:

	antecedents	consequents	lift
84	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	2.291162
85	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	2.291162
65	(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 [27]:
```
top_three_rules_supp[['antecedents','consequent
```

Out[27]:

	antecedents	consequents	support
62	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.059725
63	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	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.a
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/