

Part 1 Research Question

The question I want to be able to answer is if I can predict which services or items will be purchased that could lead to additional items being purchased as well.

The goal of the analysis is to be able to determine which services or items are most likely to be sold together.

Part 2 Market Basket Justification

Market basket analysis works by looking for combinations of items that occur together frequently in transactions. It is an easy way to see relationships between items or services that people buy. The expected outcome is rules generated by the Apriori algorithm that show the relationships between services purchased.

One example of a transaction in the dataset is

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 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

One assumption of market basket analysis is that there are no missing or null values in the dataset.

Part 3 Data Preparation and Analysis

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

```
In [2]: df_mba = pd.read_csv('telco_mba.csv')
```

In [3]: df_mba

Out[3]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09
0	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
2	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
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...

In [4]: df_mba.shape

Out[4]: (15002, 20)

In [5]: df_mba.dtypes

Out[5]: Item01 object
Item02 object
Item03 object
Item04 object
Item05 object
Item06 object
Item07 object
Item08 object
Item09 object
Item10 object
Item11 object
Item12 object
Item13 object
Item14 object
Item15 object
Item16 object
Item17 object
Item18 object
Item19 object
Item20 object
dtype: object

In [6]: records = []
for i in range(0,15002):
records.append([str(df_mba.values[i,j]) for j in range(0, 20)])

```
In [7]: TE = TransactionEncoder()  
array = TE.fit(records).transform(records)
```

```
In [8]: transf_df = pd.DataFrame(array, columns = TE.columns_  
transf_df
```

Out[8]:

	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 SB8200 Cable Modem	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	App Lightnir to Digit A Adapt
0	False	False	False	False	False	False	False	False	False	Fals
1	True	False	False	True	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	False	Tru
4	False	False	False	False	False	False	False	False	False	Fals
...	
14997	False	False	False	False	False	False	False	False	False	Fals
14998	False	False	False	False	False	False	False	False	False	Fals
14999	False	False	False	False	False	False	False	False	False	Fals
15000	False	False	False	False	False	False	False	False	False	Fals
15001	False	False	False	False	False	False	False	False	False	Fals

15002 rows × 120 columns

```
In [9]: for col in transf_df.columns:
        print(col)
```

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 128GB Ultra microSDXC card
SanDisk 128GB card
SanDisk 128GB microSDXC card
SanDisk 32GB Ultra SDHC card
SanDisk 32GB card
SanDisk Extreme 128GB card
SanDisk Extreme 256GB card
SanDisk Extreme 32GB 2pack card
SanDisk Extreme Pro 128GB card
SanDisk Extreme Pro 64GB card
SanDisk Ultra 128GB card
SanDisk Ultra 256GB card
SanDisk Ultra 400GB card
SanDisk Ultra 64GB card

```
In [10]: clean_mba = transf_df.drop(['nan'], axis = 1)
        clean_mba
```

Out[10]:

	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 SB8200 Cable Modem	Anker 2-in-1 USB Card Reader	Anker 4-port USB hub	Anker USB C to HDMI Adapter	A Light to D Adap
0	False	False	False	False	False	False	False	False	False	I
1	True	False	False	True	False	False	False	False	False	I
2	False	False	False	False	False	False	False	False	False	I
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	I
...	
14997	False	False	False	False	False	False	False	False	False	I

```
In [11]: a_rules = apriori(clean_mba, min_support = 0.04, use_colnames = True)
a_rules.head()
```

Out[11]:

	support	itemsets
0	0.043594	(Apple Lightning to Digital AV Adapter)
1	0.089855	(Apple Pencil)
2	0.066058	(Apple USB-C Charger cable)
3	0.119184	(Dust-Off Compressed Gas 2 pack)
4	0.081922	(HP 61 ink)

```
In [12]: a_rules.shape
```

Out[12]: (13, 2)

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [23]: frequent_itemset_plus = apriori(clean_mba, min_support = 0.01, use_colnames=True)
```

```
In [24]: frequent_itemset_plus['length'] = frequent_itemset_plus['itemsets'].apply(lambda
```

```
In [25]: frequent_itemset_plus
```

Out[25]:

	support	itemsets	length
0	0.119184	(Dust-Off Compressed Gas 2 pack)	1
1	0.089855	(Apple Pencil)	1
2	0.087055	(VIVO Dual LCD Monitor Desk mount)	1
3	0.085455	(USB 2.0 Printer cable)	1
4	0.081922	(HP 61 ink)	1
...
98	0.010199	(HP 63 Ink)	1
99	0.010132	(Dust-Off Compressed Gas 2 pack, SanDisk Ultra...	2
100	0.010065	(Dust-Off Compressed Gas 2 pack, Premium Nylon...	2
101	0.010065	(Dust-Off Compressed Gas 2 pack, HP 62XL Tri-C...	2
102	0.010065	(Stylus Pen for iPad, USB 2.0 Printer cable)	2

103 rows × 3 columns

In [26]: `association_rules(frequent_itemset_plus, metric = 'lift', min_threshold=0.00).sort`

Out[26]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage
0	(VIVO Dual LCD Monitor Desk mount)	(SanDisk Ultra 64GB card)	0.087055	0.049127	0.019597	0.225115	4.582324	0.015321
1	(SanDisk Ultra 64GB card)	(VIVO Dual LCD Monitor Desk mount)	0.049127	0.087055	0.019597	0.398915	4.582324	0.015321
2	(FEIYOLD Blue light Blocking Glasses)	(VIVO Dual LCD Monitor Desk mount)	0.032929	0.087055	0.011465	0.348178	3.999516	0.008599
3	(VIVO Dual LCD Monitor Desk mount)	(FEIYOLD Blue light Blocking Glasses)	0.087055	0.032929	0.011465	0.131700	3.999516	0.008599
4	(10ft iPhone Charger Cable 2 Pack)	(Dust-Off Compressed Gas 2 pack)	0.025263	0.119184	0.011532	0.456464	3.829910	0.008521
...
95	(Dust-Off Compressed Gas 2 pack)	(Apple USB-C Charger cable)	0.119184	0.066058	0.015531	0.130313	1.972713	0.007658
96	(VIVO Dual LCD Monitor Desk mount)	(USB 2.0 Printer cable)	0.087055	0.085455	0.013798	0.158499	1.854762	0.006359
97	(USB 2.0 Printer cable)	(VIVO Dual LCD Monitor Desk mount)	0.085455	0.087055	0.013798	0.161466	1.854762	0.006359
98	(USB 2.0 Printer cable)	(Dust-Off Compressed Gas 2 pack)	0.085455	0.119184	0.016864	0.197348	1.655824	0.006680
99	(Dust-Off Compressed Gas 2 pack)	(USB 2.0 Printer cable)	0.119184	0.085455	0.016864	0.141499	1.655824	0.006680

100 rows × 9 columns

The first rule is antecedent is vivo dual lcd monitor desk mount then the consequence is sandisk ultra 64gb card with a support of 0.0195997. The second rule is when feiyold blue light blocking glasses then vivo dual lcd monitor desk mount with a support of 0.011465. The third rule is that 10ft iphone charger cable 2 pack then dust off compressed gas 2 pack will be bought with a support of 0.11532.

Part IV Data Summary and Implications

The support, lift, and confidence from the results are important because it tells us out of all transactions the percentage amount that those two items are purchased together. The lift tells us that x is blank times are likely to be purchased by a customer who purchases y.

The practical significance is we can see what products are purchased most oftent together by consumers in order to pair them together.

One recommended course of action would be to run a promotion pairing the vivo monitor with the 64gb disk as those two items are purchased the most often together.

Online Sources

<https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce> (<https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce>)

<https://medium.com/@jihargifari/how-to-perform-market-basket-analysis-in-python-bd00b745b106> (<https://medium.com/@jihargifari/how-to-perform-market-basket-analysis-in-python-bd00b745b106>)

<https://medium.com/@sarahkmair7/market-basket-analysis-8dc699b7e27> (<https://medium.com/@sarahkmair7/market-basket-analysis-8dc699b7e27>)