WGU D212 Data Mining II

Task 3 - Association Rules and List Analysis

Ednalyn C. De Dios

August 25, 2023

Environment

- Python: 3.9.9
- Jupyter: 7.0.2

Part I - Research Question

A1. Propose one question relevant to a real-world organizational situation that you will answer using market basket analysis.

> ***Can we use market basket analysis to determine which items should be group together in a shelf or booth?***

A2. Define one goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

The ultimate goal of this data analysis is to increase profitability by increasing the efficiency of the store's shelving practices. We will use association rules and lift analysis to identify which group of items go are most frequently bought together. This will inform the decisions of stakeholders in matters where product placement is involved, for example.

Part II - Market Basket Justification

B1. Explain how market basket analyzes the selected dataset. Include expected outcomes.

We are using association rules and lift analysis to conduct a market basket analysis. From all items that were bought, it tries to infer which items are going to be purchased individually and together. Included outcomes involve metrics of probability and reliability in terms of support, confidence, and lift.

B2. Provide one example of transactions in the dataset.

Within a minimum threshold of 0.05 for support, Apple Pencil tops the charts with a support of 0.179709.

Out[12]:

	support	itemsets
0	0.050527	(10ft iPHone Charger Cable 2 Pack)
1	0.068391	(Anker USB C to HDMI Adapter)
2	0.087188	(Apple Lightning to Digital AV Adapter)
3	0.179709	(Apple Pencil)
4	0.132116	[illegible]
5	0.062525	(Cat8 Ethernet Cable)
6	0.238368	(Dust-Off Compressed Gas 2 pack)
7	0.065858	(FEIYOLD Blue light Blocking Glasses)

B3. Summarize one assumption of market basket analysis.

Market basket analysis assumes that all subsets of frequent itemsets are frequent. Similarly, it also assumes that infrequent subsets has infrequent parents too (Roshan, 2020).

Part III - Data Preparation and Analysis

C1. Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.

In [1]:
```
# setting the random seed for reproducibility
import random
random.seed(493)

# for manipulating dataframes
import pandas as pd
import numpy as np

# for visualizations
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")
from IPython.display import Image

# for market basket analysis
import mlxtend
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import association_rules, apriori

# to print out all the outputs of the cell
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

```
# set display options
import warnings
warnings.filterwarnings('ignore')
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
pd.set_option('display.max_colwidth', None)
```

In [2]:
```
# read the csv file
df = pd.read_csv('teleco_market_basket.csv')
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15002 entries, 0 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: 2.3+ MB
```

Out[2]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	It
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	YUN 3pa Ligh
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 [3]:
```
df.shape
```

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

In [4]:
```
# remove records with all null values
df = df.dropna(how = 'all')
df.shape
```

Out[4]:
```
(7501, 20)
```

In [5]:
```
# create a list of lists for encoding
trans = []
for i in range(df.shape[0]):
    trans.append([str(df.values[i,j]) for j in range(0, df.shape[1])])
```

In [6]:
```
# transactionalize dataset to prepare for apriori
TE = TransactionEncoder()
array = TE.fit(trans).transform(trans)
```

In [7]:
```
cleaned_df = pd.DataFrame(array, columns = TE.columns_)
cleaned_df.head()
cleaned_df.tail()
```

Out[7]:

	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	Ligl to I Ac
0	True	False	False	True	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	

Out[7]:

	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
7496	False	False	False	False	False	False	False	False	False
7497	False	False	False	False	False	True	False	False	False
7498	False	False	False	False	False	False	False	False	False
7499	False	False	False	False	False	False	False	False	False
7500	False	False	False	False	False	False	False	False	False

In [8]:
```
# list items as columns
for col in cleaned_df.columns:
    print(col)
```

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
Apple Lightning to Digital AV Adapter
Apple Lightning to USB cable
Apple Magic Mouse 2
Apple Pencil
Apple Pencil 2nd Gen
Apple Power Adapter Extension Cable
Apple USB-C Charger cable
AutoFocus 1080p Webcam
BENGOO G90000 headset
Blue Light Blocking Glasses
Blue Light Blocking Glasses 2pack
Brother Genuine High Yield Toner Cartridge
Cat 6 Ethernet Cable 50ft
Cat8 Ethernet Cable
CicTsing MM057 2.4G Wireless Mouse
Cleaning Gel Universal Dust Cleaner
Creative Pebble 2.0 Speakers
DisplayPort ot HDMI adapter
Dust-Off Compressed Gas
Dust-Off Compressed Gas 2 pack
FEEL2NICE 5 pack 10ft Lighning cable
FEIYOLD Blue light Blocking Glasses
Falcon Dust Off Compressed Gas
HOVAMP Mfi 6pack Lightning Cable
HOVAMP iPhone charger
HP 61 2 pack ink
HP 61 Tri-color ink
HP 61 ink
HP 62XL Tri-Color ink
HP 62XL ink
HP 63 Ink
HP 63 Tri-color ink
HP 63XL Ink
HP 63XL Tri-color ink
HP 64 Tri-Color ink
HP 64 ink
HP 65 ink
HP 902XL ink
HP 952 ink
HP ENVY 5055 printer
HP952XL ink
HooToo USB C Hub
HyperX Cloud Stinger Headset
Jelly Comb 2.4G Slim Wireless mouse
Leader Desk Pad Protector
Logitech M510 Wireless mouse
Logitech MK270 Wireless Keyboard/Mouse

Logitech MK345 Wireless combo
Logitech USB H390 headset
M.2 Screw kit
Mfi-Certified Lightning to USB A Cable
Micro Center 32GB Memory card
Microsot Surface Dock 2
Moread HDMI to VGA Adapter
Mpow HC6 USB Headset
NETGEAR CM500 Cable Modem
NETGEAR Nighthawk WiFi Router
NETGEAR Orbi Home Mesh WiFi System
Nylon Braided Lightning to USB cable
PS4 Headset
Premium Nylon USB Cable
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
Screen Mom Screen Cleaner kit
Stylus Pen for iPad
Syntech USB C to USB Adapter
TONOR USB Gaming Microphone
TP-Link AC1750 Smart WiFi Router
TP-Link AC4000 WiFi router
TopMate C5 Laptop Cooler pad
UNEN Mfi Certified 5-pack Lightning Cable
USB 2.0 Printer cable
USB C to USB Male Adapter
USB Type C Cable
USB Type C to USB-A Charger cable
VIVO Dual LCD Monitor Desk mount
VicTsing Mouse Pad
VicTsing Wireless mouse
Vsco 70 pack stickers
Webcam with Microphone
XPOWER A-2 Air Pump blower
YUNSONG 3pack 6ft Nylon Lightning Cable
hP 65 Tri-color ink
iFixit Pro Tech Toolkit
iPhone 11 case

```
iPhone 12 Charger cable
iPhone 12 Pro case
iPhone 12 case
iPhone Charger Cable Anker 6ft
iPhone SE case
nan
nonda USB C to USB Adapter
seenda Wireless mouse
```

In [9]:
```python
cleaned_df = cleaned_df.drop(['nan'], axis=1)
```

In [10]:
```python
# list items as columns
for col in cleaned_df.columns:
    print(col)
```

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
Apple Lightning to Digital AV Adapter
Apple Lightning to USB cable
Apple Magic Mouse 2
Apple Pencil
Apple Pencil 2nd Gen
Apple Power Adapter Extension Cable
Apple USB-C Charger cable
AutoFocus 1080p Webcam
BENGOO G90000 headset
Blue Light Blocking Glasses
Blue Light Blocking Glasses 2pack
Brother Genuine High Yield Toner Cartridge
Cat 6 Ethernet Cable 50ft
Cat8 Ethernet Cable
CicTsing MM057 2.4G Wireless Mouse
Cleaning Gel Universal Dust Cleaner
Creative Pebble 2.0 Speakers
DisplayPort ot HDMI adapter
Dust-Off Compressed Gas
Dust-Off Compressed Gas 2 pack
FEEL2NICE 5 pack 10ft Lighning cable
FEIYOLD Blue light Blocking Glasses
Falcon Dust Off Compressed Gas
HOVAMP Mfi 6pack Lightning Cable
HOVAMP iPhone charger
HP 61 2 pack ink
HP 61 Tri-color ink
HP 61 ink
HP 62XL Tri-Color ink
HP 62XL ink
HP 63 Ink
HP 63 Tri-color ink
HP 63XL Ink
HP 63XL Tri-color ink
HP 64 Tri-Color ink
HP 64 ink
HP 65 ink
HP 902XL ink
HP 952 ink
HP ENVY 5055 printer
HP952XL ink
HooToo USB C Hub
HyperX Cloud Stinger Headset
Jelly Comb 2.4G Slim Wireless mouse
Leader Desk Pad Protector
Logitech M510 Wireless mouse
Logitech MK270 Wireless Keyboard/Mouse

Logitech MK345 Wireless combo
Logitech USB H390 headset
M.2 Screw kit
Mfi-Certified Lightning to USB A Cable
Micro Center 32GB Memory card
Microsot Surface Dock 2
Moread HDMI to VGA Adapter
Mpow HC6 USB Headset
NETGEAR CM500 Cable Modem
NETGEAR Nighthawk WiFi Router
NETGEAR Orbi Home Mesh WiFi System
Nylon Braided Lightning to USB cable
PS4 Headset
Premium Nylon USB Cable
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
Screen Mom Screen Cleaner kit
Stylus Pen for iPad
Syntech USB C to USB Adapter
TONOR USB Gaming Microphone
TP-Link AC1750 Smart WiFi Router
TP-Link AC4000 WiFi router
TopMate C5 Laptop Cooler pad
UNEN Mfi Certified 5-pack Lightning Cable
USB 2.0 Printer cable
USB C to USB Male Adapter
USB Type C Cable
USB Type C to USB-A Charger cable
VIVO Dual LCD Monitor Desk mount
VicTsing Mouse Pad
VicTsing Wireless mouse
Vsco 70 pack stickers
Webcam with Microphone
XPOWER A-2 Air Pump blower
YUNSONG 3pack 6ft Nylon Lightning Cable
hP 65 Tri-color ink
iFixit Pro Tech Toolkit
iPhone 11 case

```
iPhone 12 Charger cable
iPhone 12 Pro case
iPhone 12 case
iPhone Charger Cable Anker 6ft
iPhone SE case
nonda USB C to USB Adapter
seenda Wireless mouse
```

In [11]:
```
# save the prepared data set
cleaned_df.to_csv('teleco_prepared3.csv', index=False)
```

C2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

In [12]:
```
# apriori function
a_rules = apriori(cleaned_df, min_support=0.05, use_colnames=True)
a_rules
```

Out[12]:

	support	itemsets
0	0.050527	(10ft iPHone Charger Cable 2 Pack)
1	0.068391	(Anker USB C to HDMI Adapter)
2	0.087188	(Apple Lightning to Digital AV Adapter)
3	0.179709	(Apple Pencil)
4	0.132116	(Apple USB-C Charger cable)
5	0.062525	(Cat8 Ethernet Cable)
6	0.238368	(Dust-Off Compressed Gas 2 pack)
7	0.065858	(FEIYOLD Blue light Blocking Glasses)
8	0.059992	(Falcon Dust Off Compressed Gas)
9	0.163845	(HP 61 ink)
10	0.058526	(HP 62XL Tri-Color ink)
11	0.079323	(HP 63XL Ink)
12	0.071457	(Logitech M510 Wireless mouse)
13	0.095321	(Nylon Braided Lightning to USB cable)
14	0.051060	(Premium Nylon USB Cable)
15	0.052393	(SAMSUNG EVO 32GB card)
16	0.063325	(SanDisk Ultra 128GB card)
17	0.098254	(SanDisk Ultra 64GB card)
18	0.129583	(Screen Mom Screen Cleaner kit)
19	0.095054	(Stylus Pen for iPad)
20	0.081056	(Syntech USB C to USB Adapter)
21	0.076523	(TopMate C5 Laptop Cooler pad)
22	0.170911	(USB 2.0 Printer cable)
23	0.080389	(USB Type C to USB-A Charger cable)
24	0.174110	(VIVO Dual LCD Monitor Desk mount)
25	0.050927	(Apple Pencil, Dust-Off Compressed Gas 2 pack)
26	0.052660	(HP 61 ink, Dust-Off Compressed Gas 2 pack)
27	0.059725	(VIVO Dual LCD Monitor Desk mount, Dust-Off Compressed Gas 2 pack)

C3. Provide values for the support, lift, and confidence of the association rules table.

In [13]:
```
# association rules
a_rules = association_rules(a_rules, metric = 'lift', min_threshold = 1)
a_rules
```

Out[13]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leve
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.179709	0.238368	0.050927	0.283383	1.188845	0.0
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.238368	0.179709	0.050927	0.213647	1.188845	0.0
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.01
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.01
4	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.01
5	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.01

C4. Identify the top three rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

In [14]:
```
a_rules[(a_rules['lift'] >= 1.0) & (a_rules['confidence'] >= 0.2)].sort_values('lif
```

Out[14]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leve
4	(VIVO Dual LCD Monitor Desk mount)	(Dust-Off Compressed Gas 2 pack)	0.174110	0.238368	0.059725	0.343032	1.439085	0.01
5	(Dust-Off Compressed Gas 2 pack)	(VIVO Dual LCD Monitor Desk mount)	0.238368	0.174110	0.059725	0.250559	1.439085	0.01
2	(HP 61 ink)	(Dust-Off Compressed Gas 2 pack)	0.163845	0.238368	0.052660	0.321400	1.348332	0.01
3	(Dust-Off Compressed Gas 2 pack)	(HP 61 ink)	0.238368	0.163845	0.052660	0.220917	1.348332	0.01
0	(Apple Pencil)	(Dust-Off Compressed Gas 2 pack)	0.179709	0.238368	0.050927	0.283383	1.188845	0.00
1	(Dust-Off Compressed Gas 2 pack)	(Apple Pencil)	0.238368	0.179709	0.050927	0.213647	1.188845	0.00

In [15]:
```
print( 'Number of transactions with Dust-Off Compressed Gas 2 pack: ' + str(cleaned
print( 'Number of transactions with VIVO Dual LCD Monitor Desk mount: ' + str(clean
```

```
Number of transactions with Dust-Off Compressed Gas 2 pack: 1788
Number of transactions with VIVO Dual LCD Monitor Desk mount: 1306
```

In [16]:
```
print( 'Number of transactions with Dust-Off Compressed Gas 2 pack: ' + str(cleaned
print( 'Number of transactions with HP 61 ink: ' + str(cleaned_df['HP 61 ink'].sum(
```

```
Number of transactions with Dust-Off Compressed Gas 2 pack: 1788
Number of transactions with HP 61 ink: 1229
```

In [17]:
```
print( 'Number of transactions with Dust-Off Compressed Gas 2 pack: ' + str(cleaned
print( 'Number of transactions with Apple Pencil: ' + str(cleaned_df['Apple Pencil'
```

```
Number of transactions with Dust-Off Compressed Gas 2 pack: 1788
Number of transactions with Apple Pencil: 1348
```

Part IV. Data Summary and Implications

D1. Summarize the significance of support, lift, and confidence from the results of the analysis.

Support is "the proportion of orders that include the item set" (Nguyen, 2022). In the analysis, the support of 0.05 signifies that the item sets in the association rules table appear

5% of the time. Also in this analysis, confidence, which expresses the percentage of times a particular item occurs after another item (Nguyen, 2022), is 21-34%. But these metric does not necessarily indicate a relationship between the items rather than by chance. for this, we have to use lift- an indication of association between items (Nguyen, 2022).

A lift of exactly 1 suggests pure random chance. A lift of less than 1, suggests that items are brought together less regulasrly than random while a lift of more than 1 suggests that items are brought together more regularly than random. The analysis features lift ranging from 1.1 to 1.4, suggesting that the itemsets occur more often than by chance alone. In this others, they have an actualk association.

D2. Discuss the practical significance of the findings from the analysis.

The practical significance of the findings is the availability of empirical data that supports the association of three items sets. Knowing the contents of these item sets can inform stakeholders how to place their products for maximum visibility.

D3. Recommend a course of action for the real-world organizational situation from part A1 based on your results from part D1.

Based on the item sets surfaced by the market basket analysis, I would recommend to stakeholders the importance of grouping the contents of the item sets together in a shelf or display area. The easier the customers can spot these items, the faster they can give the organization money in exchange for those goods.

Part V. Attachments

E. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

URL: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=919a3afc-e593-452b-b9fd-b06901714fcb

F. Record all web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

- https://github.com/ecdedios/code-snippets/blob/main/notebooks/master.ipynb
- https://towardsdatascience.com/introduction-to-simple-association-rules-mining-for-market-basket-analysis-ef8f2d613d87
- https://medium.com/mlearning-ai/if-i-buy-a-diaper-i-will-surely-pick-up-a-beer-e692895a0c65

G. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

- https://medium.com/analytics-vidhya/market-basket-analysis-association-rule-mining-with-visualizations-cda24d537019
- https://towardsdatascience.com/introduction-to-simple-association-rules-mining-for-market-basket-analysis-ef8f2d613d87

In [18]:
```
print('Successful run!')
```

```
Successful run!
```