OFM3 — OFM3 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS

- Ryan L. Buchanan
- Student ID: 001826691
- Masters Data Analytics (12/01/2020)
- Program Mentor: Dan Estes
- 385-432-9281 (MST)
- rbuch49@wgu.edu

Scenario 1

One of the most critical factors in customer relationship management that directly affects a company's long-term profitability is understanding its customers. When a company can better understand its customer characteristics, it is better able to target products and marketing campaigns for customers, resulting in better profits for the company in the long term.

You are an analyst for a telecommunications company that wants to better understand the characteristics of its customers. You have been asked to perform a market basket analysis to analyze customer data to identify key associations of your customer purchases, ultimately allowing better business and strategic decision-making.

Part I: Research Question

A1. Proposal of Question:

Which are the items of interest in combination with discounts that might reduce customer churn? That is, by analyzing a list of transactions, may we be able to better understand which items will endear us to customers more if offered as discounted with our services? This question will be answered using **market basket analysis**.

A2. Defined Goal:

Stakeholders in the company will benefit by knowing, with some measure of confidence, which customers are at highest risk of churn because this will provide weight for decisions in marketing improved services to customers with these characteristics and past user experiences. The goal of this data analysis is to present items for discount purchase to company stakeholders to consider when creating customer enticements and marketing promotions. We will endeavor to help decision makers better understand which combinations of features (items in concert with telecom services) put their customers at lower risk of churning.

Part II: Market Basket Justification

B1. Explanation of Market Basket:

As pointed out by Li, "[m]arket basket analysis is one of the key techniques used . . . to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions" (Li, p. 1).

This analysis proposes to identify which combinations of telecom peripherals and ICT tools customers prefer and purchase together most often. We will try to identify those items purchased most often together and demonstrate the relationships between these different items.

We expect that we will discover an optimal combination of items to offer at discounts in coordination with our services.

Our plan for analysis includes:

- Prepare the dataset
- Discover missing values
- Run the Apriori method to identify association rules
- Check the rules with highest values for confidence, support and lift
- Recommend a course of action following the results of our analysis

B2. Transaction Example:

On quick inspection of the given dataset, transactions are easily distinguishable. The very first transactions includes a larger list of twenty items including:

- 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

These twenty items were purchased by one customer, synchronously.

B3. Market Basket Assumption:

One assumption of MBA is to make determinations by building association rules. These rules, suggests Dr. Susan Sivek, "are just statements that connect an 'antecedent' item to a 'consequent' item. Association rules also do not imply causal relationships, only co-occurrence" (Sivek, p. 1).

So, for instance in our research proposal, we would like to identify items that would purchased before subscribing to a telecom service, or, perhaps, items that would be used in coordination with telecom services.

C1. Transforming the Dataset:

In [1]:
```
# Standard data science imports
import numpy as np
import pandas as pd

# Visualization libraries
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
# Change color of Matplotlib font
import matplotlib as mpl

COLOR = 'white'
mpl.rcParams['text.color'] = COLOR
mpl.rcParams['axes.labelcolor'] = COLOR
mpl.rcParams['xtick.color'] = COLOR
mpl.rcParams['ytick.color'] = COLOR
```

In [3]:
```
# Increase Jupyter display cell-width
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:75% !important; }</style>"))
```

In [4]:
```
# Ignore Warning Code
import warnings
warnings.filterwarnings('ignore')
```

In [5]:
```
# Load data set into Pandas dataframe
teleco = pd.read_csv('data/teleco_market_basket.csv')
```

In [6]:
```
# Examine the features of the dataset
teleco.columns
```

Out[6]:
```
Index(['Item01', 'Item02', 'Item03', 'Item04', 'Item05', 'Item06', 'Item07',
       'Item08', 'Item09', 'Item10', 'Item11', 'Item12', 'Item13', 'Item14',
       'Item15', 'Item16', 'Item17', 'Item18', 'Item19', 'Item20'],
      dtype='object')
```

In [7]:
```
# Get an idea of dataset size
teleco.shape
```

Out[7]:
```
(15002, 20)
```

In [8]:
```
# Examine first few records of dataset
teleco.head()
```

Out[8]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	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	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
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	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	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [37]:
```
# View DataFrame info
# teleco.info
```

In [10]:
```
# Get an overview of descriptive statistics
teleco.describe()
```

Out[10]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18
count	7501	5747	4389	3345	2529	1864	1369	981	654	395	256	154	87	47	25	8	4	4
unique	115	117	115	114	110	106	102	97	88	80	66	50	43	28	19	8	3	3
top	Dust-Off Compressed Gas 2 pack	Dust-Off Compressed Gas 2 pack	Dust-Off Compressed Gas 2 pack	Dust-Off Compressed Gas 2 pack	Apple USB-C Charger cable	USB 2.0 Printer cable	Apple USB-C Charger cable	Apple USB-C Charger cable	Apple USB-C Charger cable	Apple USB-C Charger cable	TopMate C5 Laptop Cooler pad	Apple USB-C Charger cable	Apple USB-C Charger cable	Apple USB-C Charger cable	ARRIS SURFboard SB8200 Cable Modem	Syntech USB C to USB Adapter	SanDisk Ultra 128GB card	Brother Genuine High Yield Toner Cartridge
freq	577	484	375	201	153	107	96	67	57	31	22	15	8	4	3	1	2	2

In [11]:
```
# Get data types of features
teleco.dtypes
```

Out[11]:
```
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 [12]:
```
# Discover missing data points within dataset
data_nulls = teleco.isnull().sum()
print(data_nulls)
```

```
Item01     7501
Item02     9255
Item03    10613
Item04    11657
Item05    12473
Item06    13138
Item07    13633
Item08    14021
Item09    14348
Item10    14607
Item11    14746
Item12    14848
Item13    14915
Item14    14955
Item15    14977
Item16    14994
Item17    14998
Item18    14998
Item19    14999
Item20    15001
dtype: int64
```

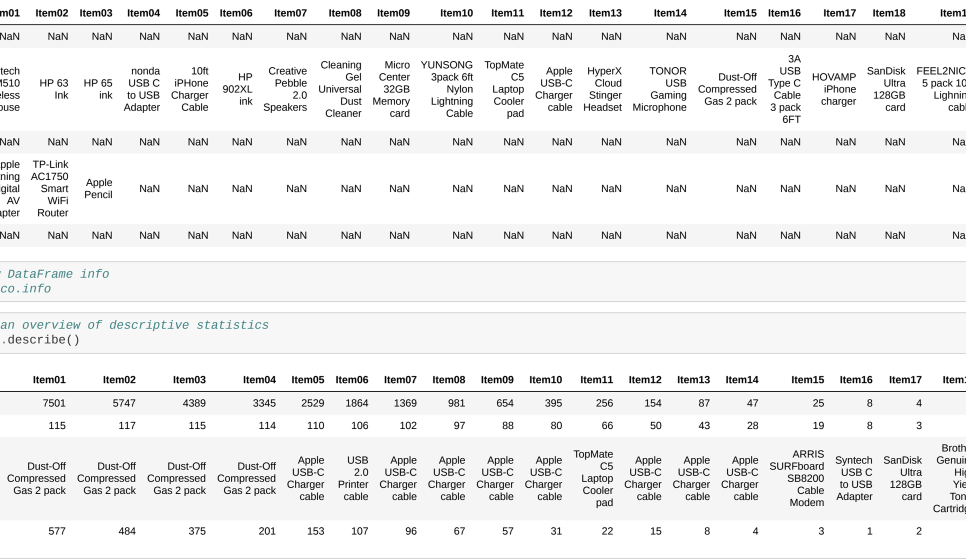
In [13]:
```
# Check for missing data & visualize missing values in dataset

# Install appropriate library
!pip install missingno

# Importing the libraries
import missingno as msno

# Visualize missing values as a matrix
msno.matrix(teleco);
"""(GeeksForGeeks, p. 1)"""
```

```
Requirement already satisfied: missingno in c:\users\vreed\anaconda3\lib\site-packages (0.5.0)
Requirement already satisfied: numpy in c:\users\vreed\anaconda3\lib\site-packages (from missingno) (1.18.1)
Requirement already satisfied: scipy in c:\users\vreed\anaconda3\lib\site-packages (from missingno) (1.4.1)
Requirement already satisfied: seaborn in c:\users\vreed\anaconda3\lib\site-packages (from missingno) (0.10.0)
Requirement already satisfied: matplotlib in c:\users\vreed\anaconda3\lib\site-packages (from missingno) (3.1.3)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\users\vreed\anaconda3\lib\site-packages (from matplotlib->missin
Requirement already satisfied: cycler>=0.10 in c:\users\vreed\anaconda3\lib\site-packages (from matplotlib->missingno) (0.10.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\vreed\anaconda3\lib\site-packages (from matplotlib->missingno) (2.8.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\vreed\anaconda3\lib\site-packages (from matplotlib->missingno) (1.1.0)
Requirement already satisfied: six in c:\users\vreed\anaconda3\lib\site-packages (from cycler>=0.10->matplotlib->missingno) (1.14.0)
Requirement already satisfied: setuptools in c:\users\vreed\anaconda3\lib\site-packages (from kiwisolver>=1.0.1->matplotlib->missingno) (45.2.
Requirement already satisfied: pandas>=0.22.0 in c:\users\vreed\anaconda3\lib\site-packages (from seaborn->missingno) (1.0.1)
Requirement already satisfied: pytz>=2017.2 in c:\users\vreed\anaconda3\lib\site-packages (from pandas>=0.22.0->seaborn->missingno) (2019.3)
```

```
WARNING: You are using pip version 21.2.4; however, version 21.3 is available.
You should consider upgrading via the 'c:\users\vreed\anaconda3\python.exe -m pip install --upgrade pip' command.
```

Out[13]:
```
'(GeeksForGeeks, p. 1)'
```

```
findfont: Font family ['sans-serif'] not found. Falling back to DejaVu Sans.
findfont: Font family ['sans-serif'] not found. Falling back to DejaVu Sans.
findfont: Font family ['sans-serif'] not found. Falling back to DejaVu Sans.
```



In [14]:
```
# Drop records with missing values
teleco.dropna(how='all', inplace=True)

# Review changes
teleco.head()
```

Out[14]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19
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
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	UNEN Mfi Certified 5-pack Lightning Cable	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	Cat8 Ethernet Cable	HP 65 ink	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	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	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [15]:
```
# Replace empty values with 0
teleco.fillna(0, inplace=True)
```

In [16]:
```
# Get an idea of dataset size after changes
teleco.shape
```

Out[16]:
```
(7501, 20)
```

In [17]:
```
# Review changes to DataFrame
teleco.head()
```

Out[17]:

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19
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
3	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	UNEN Mfi Certified 5-pack Lightning Cable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Cat8 Ethernet Cable	HP 65 ink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	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	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In [18]:
```
# Confirm no null values
teleco.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7501 entries, 1 to 15001
Data columns (total 20 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Item01  7501 non-null   object
 1   Item02  7501 non-null   object
 2   Item03  7501 non-null   object
 3   Item04  7501 non-null   object
 4   Item05  7501 non-null   object
 5   Item06  7501 non-null   object
 6   Item07  7501 non-null   object
 7   Item08  7501 non-null   object
 8   Item09  7501 non-null   object
 9   Item10  7501 non-null   object
 10  Item11  7501 non-null   object
 11  Item12  7501 non-null   object
 12  Item13  7501 non-null   object
 13  Item14  7501 non-null   object
 14  Item15  7501 non-null   object
 15  Item16  7501 non-null   object
 16  Item17  7501 non-null   object
 17  Item18  7501 non-null   object
 18  Item19  7501 non-null   object
 19  Item20  7501 non-null   object
dtypes: object(20)
memory usage: 1.2+ MB
```

In [19]:
```
# Convert dataset into list format for use with Apriori algorithm
teleco_list = []
for i in range(0, 7501):
    teleco_list.append([str(teleco.values[i, j]) for j in range(0, 20)])
teleco_cleaned = pd.DataFrame(teleco_list)
```

In [20]:
```
# Review DataFrame
teleco_cleaned.head()
```

Out[20]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	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
1	Apple Lightning to Digital AV Adapter	TP-Link AC1750 Smart WiFi Router	Apple Pencil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	UNEN Mfi Certified 5-pack Lightning Cable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Cat8 Ethernet Cable	HP 65 ink	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	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	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In [21]:
```
# Extract prepared dataset
teleco_cleaned.to_csv('data/teleco_market_basket_prepared.csv')
```

In [22]:
```
teleco_list[:1]
```

Out[22]:
```
[['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']]
```

C2. Code Execution:

In [23]:
```
# Generate association rules from Apriori algorithm
from apyori import apriori

# Train Apriori algorithm on the dataset
rule_list = apriori(teleco_list, min_support = 0.003, min_confidence = 0.3, min_lift = 3, min_length = 2)
```

In [24]:
```
# Review generate rules
rule_list = list(rule_list)
print(rule_list[0])
```

```
RelationRecord(items=frozenset({'5pack Nylon Braided USB C cables', 'HP 63XL Ink'}), support=0.005732568990801226, ordered_statistics=[Ordered
```

In [25]:
```
# Print number of rules
print(len(rule_list))
```

```
102
```

In [26]:
```
# Transform results into DataFrame structure
results = pd.DataFrame(rule_list)
```

In [27]:
```
# View results list
results
```

Out[27]:

	items	support	ordered_statistics
0	(5pack Nylon Braided USB C cables, HP 63XL Ink)	0.005733	[((5pack Nylon Braided USB C cables), (HP 63XL...
1	(AutoFocus 1080p Webcam, SanDisk Ultra 64GB card)	0.005333	[((AutoFocus 1080p Webcam), (SanDisk Ultra 64G...
2	(HP 63XL Ink, iPhone 11 case)	0.005866	[((iPhone 11 case), (HP 63XL Ink), 0.372881355...
3	(Logitech M510 Wireless mouse, iPhone 11 case)	0.005066	[((iPhone 11 case), (Logitech M510 Wireless mo...
4	(SanDisk 128GB Ultra microSDXC card, SanDisk U...	0.015998	[((SanDisk 128GB Ultra microSDXC card), (SanDi...
...	...	...	...
97	(Dust-Off Compressed Gas 2 pack, VIVO Dual LCD...	0.004399	[((Dust-Off Compressed Gas 2 pack, VIVO Dual L...
98	(Dust-Off Compressed Gas 2 pack, Screen Mom Sc...	0.003200	[((Dust-Off Compressed Gas 2 pack, VIVO Dual L...
99	(HP 61 ink, VIVO Dual LCD Monitor Desk mount, ...	0.003066	[((SanDisk Ultra 64GB card, Nylon Braided Ligh...
100	(Screen Mom Screen Cleaner kit, HP 61 ink, VIV...	0.003466	[((VIVO Dual LCD Monitor Desk mount, Nylon Bra...
101	(Screen Mom Screen Cleaner kit, VIVO Dual LCD ...	0.003066	[((Screen Mom Screen Cleaner kit, SanDisk Ultr...

102 rows × 3 columns

In [28]:
```
# Separate support to indiviual DataFrame
support = results.support
```

In [29]:
```
# Instantiate four empty lists to contain lhs, rhs, confidence and lift
first_values = []
second_values = []
third_values = []
fourth_values = []
```

In [30]:
```
# Create for loop to iterate over list
for i in range(results.shape[0]):
    single_list = results['ordered_statistics'][i][0]
    first_values.append(list(single_list[0]))
    second_values.append(list(single_list[1]))
    third_values.append(single_list[2])
    fourth_values.append(single_list[3])
```

In [31]:
```
# Convert lists into DataFrame
lhs = pd.DataFrame(first_values)
rhs = pd.DataFrame(second_values)
confidence = pd.DataFrame(third_values, columns=['confidence'])
lift = pd.DataFrame(fourth_values, columns=['lift'])
```

In [32]:
```
# Concatenate lists into single DataFrame
results_final = pd.concat([lhs, rhs, support, confidence, lift], axis=1)
results_final.fillna(value=' ', inplace=True)
```

In [33]:
```
# View final results
results_final
```

Out[33]:

	0	1	2	0	1	2	support	confidence	lift
0	5pack Nylon Braided USB C cables			HP 63XL Ink			0.005733	0.300699	3.790833
1	AutoFocus 1080p Webcam			SanDisk Ultra 64GB card			0.005333	0.377358	3.840659
2	iPhone 11 case			HP 63XL Ink			0.005866	0.372881	4.700812
3	iPhone 11 case			Logitech M510 Wireless mouse			0.005066	0.322034	4.506672
4	SanDisk 128GB Ultra microSDXC card			SanDisk Ultra 64GB card			0.015998	0.323450	3.291994
...	...	...	...	...	...	...	...	...	...
97	Dust-Off Compressed Gas 2 pack	VIVO Dual LCD Monitor Desk mount	Nylon Braided Lightning to USB cable	0	SanDisk Ultra 64GB card		0.004399	0.366667	3.731841
98	Dust-Off Compressed Gas 2 pack	VIVO Dual LCD Monitor Desk mount	SanDisk Ultra 128GB card	0	Screen Mom Screen Cleaner kit		0.003200	0.470588	3.631566
99	SanDisk Ultra 64GB card	Nylon Braided Lightning to USB cable	HP 61 ink	VIVO Dual LCD Monitor Desk mount	0		0.003066	0.534884	3.072100
100	VIVO Dual LCD Monitor Desk mount	Nylon Braided Lightning to USB cable	HP 61 ink	0	Screen Mom Screen Cleaner kit		0.003466	0.440678	3.400746
101	Screen Mom Screen Cleaner kit	SanDisk Ultra 64GB card	Nylon Braided Lightning to USB cable	VIVO Dual LCD Monitor Desk mount	0		0.003066	0.534884	3.072100

102 rows × 9 columns

C3. Association Rules Table:

In [34]:
```
# Set column names
results_final.columns = ['lhs', 1, 2, 'rhs', 1, 2, 'support', 'confidence', 'lift']
results_final_1 = results_final[['lhs', 'rhs', 'support', 'confidence', 'lift']]
results_final_1
```

Out[34]:

	lhs	rhs	support	confidence	lift
0	5pack Nylon Braided USB C cables	HP 63XL Ink	0.005733	0.300699	3.790833
1	AutoFocus 1080p Webcam	SanDisk Ultra 64GB card	0.005333	0.377358	3.840659
2	iPhone 11 case	HP 63XL Ink	0.005866	0.372881	4.700812
3	iPhone 11 case	Logitech M510 Wireless mouse	0.005066	0.322034	4.506672
4	SanDisk 128GB Ultra microSDXC card	SanDisk Ultra 64GB card	0.015998	0.323450	3.291994
...	...	...	...	...	...
97	Dust-Off Compressed Gas 2 pack	0	0.004399	0.366667	3.731841
98	Dust-Off Compressed Gas 2 pack	0	0.003200	0.470588	3.631566
99	SanDisk Ultra 64GB card	VIVO Dual LCD Monitor Desk mount	0.003066	0.534884	3.072100
100	VIVO Dual LCD Monitor Desk mount	0	0.003466	0.440678	3.400746
101	Screen Mom Screen Cleaner kit	VIVO Dual LCD Monitor Desk mount	0.003066	0.534884	3.072100

102 rows × 5 columns

Highest combination of Support, Confidence and LIft

After running the final results to create the association rules table, we can demonstate mathematically that "5pack Nylon Braided USB C cables" and "HP 63XL Ink" have the highest combination of values for our three metrics:

For "5pack Nylon Braided USB C cables" → "HP 63XL Ink"

- Support = 0.0057
- Confidence = 0.3007
- Lift = 3.7908

In [36]:
```
# Visualize the list of rules
results = list(rule_list)
for i in results:
    print('\n')
    print(i)
    print('**********')
```

C4. Top Three Rules:

The top three rules are as follow:

1. If "5pack Nylon Braided USB C cables" then "HP 63XL Ink" with:

- Support = 0.0057
- Confidence = 0.3007 = 30%
- Lift = 3.7908
  Our confidence in this rule demonstrates that out of all customers who purchased the "5pack Nylon Braided USB C cables", 30% also purchased the "HP 63XL Ink". The simplest metric of support, with a value of 0.0057, demonstrates that a little more than half a percentage of all transactions contain both items. A lift value of 3.7908 demonstrates that once a customer has purchased the "5pack Nylon Braided USB C cables", they are 3.8 times more likely to also purchase the "HP 63XL Ink".

  #2. If "AutoFocus 1080p Webcam" then "SanDisk Ultra 64GB card" with:
- Support = 0.0053
- Confidence = 0.3774 = 38% of customers also purchased consequent
- Lift = 3.8407 = 3.8 times more likely to purchase consequent following purchase of antecedent

  #3. If "iPhone 11 case" then "HP 63XL Ink" with:
- Support = 0.0051
- Confidence = 0.3729 = 37% of customers also purchased consequent
- Lift = 4.7008 = 4.7 times more likely to purchase consequent following purchase of antecedent

Part IV: Analysis

D1. Significance of Support, Lift, and Confidence Summary

Our top three rules compare the metrics:

- $Support = \frac{frequency(X,Y)}{N}$ → Giving us the number of total transactions containing this particular itemset.
- $Confidence = \frac{frequency(X,Y)}{frequency(X)}$ → Giving us a probability of the consequent given the antecedent.
- $Lift = \frac{Support}{Support(X)*Support(Y)}$ → Giving us the coefficient of likelihood given the antecedent; that is, how many more times likely is the consequent to be purchased once the antecedent has been purchased.

  The results of this analysis are not particularly compelling. None of the rules have a confidence level of greater than 40% and certainly not the greater than 80% which would be an optimal value for signifcances.
  Our highest **confidence** is in rule #2 at 38%, while the #1 rule (given its analysis in combination with our other metrics of interest) is only 30%.
  The **support** for the pairing of any of the given top three rules' itemsets does not occur in more than half a percentage point of all transactions, and, again, is not compelling.
  Finally, the **lift** ratio gives us some hope that once a customer has purchased the antecedent item they will also purchase the consequent item. Our highest lift metric at "4.7 times more likely" is demonstrated by the relationship between purchasing an "iPhone 11 case" and then purchasing some "HP 63XL Ink".

D2. Practical Significance of Findings

We do not find that these results contain very much practical significance as we really cannot be confident that any itemset will be purchased even half of the time. We have a greater chance of predicting the outcome of a coin flip, now, do we not? We can see that if one of the antecedents is selected for purchase, say a webcam, it is about 4 times more likely that the customer will also purchase the consequent, say a memory card.

So, for example, if that half a percentage point of people pick up a 5 pack of USB cables, they are nearly 4 times as likely to pickup some HP ink for the printer.

These results really do not give us much to go on. Perhaps we need together more data before suggesting any significance. And, of course, further analysis is recommended.

D3. Course of Action

Therefore, based on the previous analysis and commentary of significance, we do not recommend company decision makers move forward with the original plan of promoting our service by discounted or, even, giving away free items for subscribing to our telecom service. Not only did we not find any significance in our market basket analysis of this transaction dataset, none of the pairings suggested customers who used telecom services would like or need some consequent item.

That is, if we had found a significant relationship with say, many transactions where customers purchased two related telecommunications peripherals, we might suggest one of those items for potential customer discount and a marketing promotion. We did not find that. We found ink being purchased when, perhaps we might be looking for a relationship where both a webcam and ethernet cable were purchased at the same time.

No action is warranted at this time. More data needs to be gathered and analyzed before confident action can be recommended by our data science team.

E. Panopto Recording

link

F. Web Sources

- GeeksForGeeks. (2019, July 4). *Python | Visualize missing values (NaN) values using Missingno Library*. GeeksForGeeks. https://www.geeksforgeeks.org/python-visualize-missing-values-nan-values-using-missingno-library/
- Gupta, A. (2021). *Implementing Apriori algorithm in Python*. GeeksForGeeks. https://www.geeksforgeeks.org/implementing-apriori-algorithm-in-python/
- Kumar, V. (2020, May 11). *Hands-On Guide To Market Basket Analysis With Python Codes*. AnalyticsIndiaMag.com. https://analyticsindiamag.com/hands-on-guide-to-market-basket-analysis-with-python-codes/ Intellipaat. (2021). Introduction to Apriori Algorithm in Python. https://intellipaat.com/blog/data-science-apriori-algorithm/
- Umredkar, R. (2020, November 30). *Guide To Association Rule Mining From Scratch*. AnalyticsIndiaMag.com. https://analyticsindiamag.com/guide-to-association-rule-mining-from-scratch/
- Yogesh. (2018). *Market Basket Analysis (Apriori) in Python*. Kaggle. https://www.kaggle.com/yugagrawal95/market-basket-analysis-apriori-in-python

G. Sources

- Li, S. (2017, September 24). *A Gentle Introduction on Market Basket Analysis — Association Rules*. TowardsDataScience. https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce
- Sivek, S. (2020, November 16). *Market Basket Analysis 101: Key Concepts*. TowardsDataScience. https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00