**Project Purpose**

Telecommunications organizations often suffer from a loss of revenue due to customers choosing to terminate their services. According to the data consideration provided with the dataset, telecommunications companies experience customer churn at a rate of approximately 25 percent per year. This results in a loss of revenue as it cost approximately ten times more to acquire a new customer than to keep an existing customer.

The organization is hoping to retain customers by offering discount products and services. The question that I plan to answer with this analysis is "What products do customers tend to purchase together?"

One goal that I have for this analysis is to identify groupings of products that customers tend to purchase together. I plan to accomplish this by using market basket analysis.

**Explanation of Method**

Market basket analysis is a technique that attempts to find patterns in large data sets (Moffitt, 2017). It does this by grouping items that frequently occur in transactions together. The Apriori algorithm, which is used in this analysis, can quickly scan through the data creating all possible combinations of items, and provide the relative frequency (or support metric) that the combination might occur.

From this association rules can be created using an antecedent and consequent. Each grouping in an association table can be interpreted as follows: "If (antecedent) then (consequent)". This means that if the item or items in the antecedent occurs in a transaction, then the consequent occurs in the same transaction.

Metrics can be used to measure the significance of these groupings, and curate results based on chosen thresholds. One metric that was mentioned earlier is support, or the relative frequency that the combination of items might occur. Another metric is confidence, which describes the reliability of the association rule. Finally, the lift is the ratio of the support to the expectation that the rules are independent (Hull, n.d.)

**Assumptions**

One assumption of market basket analysis is that when there is a "joint occurrence of two or more products in most baskets, it implies that the products are complements in the purchase, therefore, purchase of one will lead to the purchase of others" (Tian, 2015). In this analysis, I will be looking for the top three complimentary items or groups of items that customers tend to purchase together.

**Preprocessing**

In [1]:
```
# Import necessary libraries
import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

import warnings
warnings.filterwarnings('ignore') # Ignore warning messages for readability
```

In [3]:
```
# Read in data set and view head
df = pd.read_csv('teleco_market_basket.csv')
pd.options.display.max_columns = None
df.head()
```

Out[3]:

| | Item01 | Item02 | Item03 | Item04 | Item05 | Item06 | Item07 | Item08 | Item09 | Item10 | Item11 | Item12 | Item13 | Item14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 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 |
| **2** | 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 |
| **4** | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

In [4]:
```
# View number of rows and null values
df.info()
```

```
<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
```

- It appears that some rows are made up entirely of null values, therefore those rows will be dropped.

In [5]:
```
# Drop all rows that are completely nan values
df= df.dropna(axis = 0, how = 'all')
```

In [6]:
```
# View df info to ensure dropped rows have been removed
df.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  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: 1.2+ MB
```

- We are left with about half of the rows remaining. Now we create a list of lists for the remaining rows, then delete any remaining "nan" values.

In [7]:
```
# Create a list of lists for each row in df
transactions = []
transactions = df.values
```

In [8]:
```
# Remove nan values frrom lists
nan = float('nan')
transactions = [[i for i in j if i == i] for j in transactions]
```

In [9]:
```
#Print sample transaction
print(transactions[20])
```

```
['VicTsing Wireless mouse', 'Logitech M510 Wireless mouse', 'iPhone 11 case', 'iPhone SE case', 'Apple Penci
l', 'HP 61 ink', 'SanDisk Extreme 32GB 2pack card']
```

- Transactions will now be encoded in preparation for analysis.

In [10]:
```
# Instantiate transaction encoder and identify unique items (ref F2)
encoder = TransactionEncoder().fit(transactions)

# One-hot encode transactions
onehot = encoder.transform(transactions)

# Convert one-hot encoded data to DataFrame
onehot = pd.DataFrame(onehot, columns = encoder.columns_)
onehot.head(1)
```

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 | Apple Lightning to Digital AV Adapter | Apple Lightning to USB cable | Apple Magic Mouse 2 | Apple Pencil | Apple Pencil 2nd Gen | A Exte |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | True | False | False | True | False | False | False | False | False | False | False | False | False | False | |

In [11]:
```
# Save transformed data to Excel
onehot.to_excel('market_basket_transformed.xlsx', index = False, encoding = 'utf-8')
```

**Market Basket Analysis**

In [12]:
```
# Compute frequent itemsets using the Apriori algorithm
frequent_itemsets = apriori(onehot, min_support = 0.006, use_colnames = True)
frequent_itemsets.head()
```

Out[12]:

| | support | itemsets |
|---|---|---|
| 0 | 0.009065 | (10ft iPHone Charger Cable) |
| 1 | 0.050527 | (10ft iPHone Charger Cable 2 Pack) |
| 2 | 0.042528 | (3A USB Type C Cable 3 pack 6FT) |
| 3 | 0.019064 | (5pack Nylon Braided USB C cables) |
| 4 | 0.010932 | (ARRIS SURFboard SB8200 Cable Modem) |

In [13]:
```
# Print number of itemsets
print("There are", frequent_itemsets.shape[0], "sets of items.")
```

```
There are 542 sets of items.
```

In [14]:
```
# Compute all association rules using confidence metric
rules = association_rules(frequent_itemsets, metric = "confidence", min_threshold = 0.5)
rules.head()
```

Out[14]:

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|---|---|---|---|---|---|---|---|---|---|
| 0 | (10ft iPHone Charger Cable 2 Pack, Screen Mom ... | (Dust-Off Compressed Gas 2 pack) | 0.015198 | 0.238368 | 0.008532 | 0.561404 | 2.355194 | 0.004909 | 1.736520 |
| 1 | (10ft iPHone Charger Cable 2 Pack, VIVO Dual L... | (Dust-Off Compressed Gas 2 pack) | 0.014265 | 0.238368 | 0.007466 | 0.523364 | 2.195614 | 0.004065 | 1.597933 |
| 2 | (VIVO Dual LCD Monitor Desk mount, 3A USB Type... | (Dust-Off Compressed Gas 2 pack) | 0.013465 | 0.238368 | 0.006799 | 0.504950 | 2.118363 | 0.003589 | 1.538496 |
| 3 | (Apple Pencil, Premium Nylon USB Cable) | (Dust-Off Compressed Gas 2 pack) | 0.011732 | 0.238368 | 0.006399 | 0.545455 | 2.288286 | 0.003603 | 1.675590 |
| 4 | (Apple Pencil, SanDisk Ultra 64GB card) | (Dust-Off Compressed Gas 2 pack) | 0.019997 | 0.238368 | 0.010132 | 0.506667 | 2.125563 | 0.005365 | 1.543848 |

In [15]:
```
# Print the number of rules after applying the confidence metric with a threshold of 5
print("There are", rules.shape[0], "rules after applying a confidence metric with a minimum value of 0.5.")
```

```
There are 14 rules after applying a confidence metric with a minimum value of 0.5.
```

In [16]:
```
# Print association rules table
rules
```

Out[16]:

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|---|---|---|---|---|---|---|---|---|---|
| 0 | (10ft iPHone Charger Cable 2 Pack, Screen Mom ... | (Dust-Off Compressed Gas 2 pack) | 0.015198 | 0.238368 | 0.008532 | 0.561404 | 2.355194 | 0.004909 | 1.736520 |
| 1 | (10ft iPHone Charger Cable 2 Pack, VIVO Dual L... | (Dust-Off Compressed Gas 2 pack) | 0.014265 | 0.238368 | 0.007466 | 0.523364 | 2.195614 | 0.004065 | 1.597933 |
| 2 | (VIVO Dual LCD Monitor Desk mount, 3A USB Type... | (Dust-Off Compressed Gas 2 pack) | 0.013465 | 0.238368 | 0.006799 | 0.504950 | 2.118363 | 0.003589 | 1.538496 |
| 3 | (Apple Pencil, Premium Nylon USB Cable) | (Dust-Off Compressed Gas 2 pack) | 0.011732 | 0.238368 | 0.006399 | 0.545455 | 2.288286 | 0.003603 | 1.675590 |
| 4 | (Apple Pencil, SanDisk Ultra 64GB card) | (Dust-Off Compressed Gas 2 pack) | 0.019997 | 0.238368 | 0.010132 | 0.506667 | 2.125563 | 0.005365 | 1.543848 |
| 5 | (Cat8 Ethernet Cable, Screen Mom Screen Cleane... | (Dust-Off Compressed Gas 2 pack) | 0.011332 | 0.238368 | 0.006133 | 0.541176 | 2.270338 | 0.003431 | 1.659967 |
| 6 | (FEIYOLD Blue light Blocking Glasses, HP 61 ink) | (Dust-Off Compressed Gas 2 pack) | 0.016398 | 0.238368 | 0.008266 | 0.504065 | 2.114649 | 0.004357 | 1.535749 |
| 7 | (FEIYOLD Blue light Blocking Glasses, Nylon Br... | (Dust-Off Compressed Gas 2 pack) | 0.011332 | 0.238368 | 0.006532 | 0.576471 | 2.418404 | 0.003831 | 1.798297 |
| 8 | (FEIYOLD Blue light Blocking Glasses, Screen M... | (Dust-Off Compressed Gas 2 pack) | 0.017064 | 0.238368 | 0.008532 | 0.500000 | 2.097595 | 0.004465 | 1.523264 |
| 9 | (Falcon Dust Off Compressed Gas, HP 61 ink) | (Dust-Off Compressed Gas 2 pack) | 0.014665 | 0.238368 | 0.007599 | 0.518182 | 2.173871 | 0.004103 | 1.580745 |
| 10 | (Nylon Braided Lightning to USB cable, SanDisk... | (Dust-Off Compressed Gas 2 pack) | 0.016931 | 0.238368 | 0.009199 | 0.543307 | 2.279277 | 0.005163 | 1.667711 |
| 11 | (Screen Mom Screen Cleaner kit, SanDisk Ultra ... | (Dust-Off Compressed Gas 2 pack) | 0.021997 | 0.238368 | 0.011065 | 0.503030 | 2.110308 | 0.005822 | 1.532552 |
| 12 | (Stylus Pen for iPad, SanDisk Ultra 64GB card) | (Dust-Off Compressed Gas 2 pack) | 0.014531 | 0.238368 | 0.007466 | 0.513761 | 2.155327 | 0.004002 | 1.566375 |
| 13 | (Nylon Braided Lightning to USB cable, SanDisk... | (VIVO Dual LCD Monitor Desk mount) | 0.016931 | 0.174110 | 0.008666 | 0.511811 | 2.939582 | 0.005718 | 1.691742 |

In [17]:
```
# Sort rules by lift, confidence, and support to determine top 3 rules
rules.sort_values(by=['lift', 'confidence'], ascending = False)
rules.head(3)
```

Out[17]:

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|---|---|---|---|---|---|---|---|---|---|
| 0 | (10ft iPHone Charger Cable 2 Pack, Screen Mom ... | (Dust-Off Compressed Gas 2 pack) | 0.015198 | 0.238368 | 0.008532 | 0.561404 | 2.355194 | 0.004909 | 1.736520 |
| 1 | (10ft iPHone Charger Cable 2 Pack, VIVO Dual L... | (Dust-Off Compressed Gas 2 pack) | 0.014265 | 0.238368 | 0.007466 | 0.523364 | 2.195614 | 0.004065 | 1.597933 |
| 2 | (VIVO Dual LCD Monitor Desk mount, 3A USB Type... | (Dust-Off Compressed Gas 2 pack) | 0.013465 | 0.238368 | 0.006799 | 0.504950 | 2.118363 | 0.003589 | 1.538496 |

In [18]:
```
# Print top 3 rules
print("Rule #1: If", ([x for x in rules['antecedents'][0]]), "then", ([x for x in rules['consequents'][0]]))
print("Rule #2: If", ([x for x in rules['antecedents'][1]]), "then", ([x for x in rules['consequents'][1]]))
print("Rule #3: If", ([x for x in rules['antecedents'][2]]), "then", ([x for x in rules['consequents'][2]]))
```

```
Rule #1: If ['10ft iPHone Charger Cable 2 Pack', 'Screen Mom Screen Cleaner kit'] then ['Dust-Off Compressed
Gas 2 pack']
Rule #2: If ['10ft iPHone Charger Cable 2 Pack', 'VIVO Dual LCD Monitor Desk mount'] then ['Dust-Off Compress
ed Gas 2 pack']
Rule #3: If ['VIVO Dual LCD Monitor Desk mount', '3A USB Type C Cable 3 pack 6FT'] then ['Dust-Off Compressed
Gas 2 pack']
```

**Summary of Findings**

For the top three rules, the support metrics were calculated ranging between 0.007 and 0.009. This is the relative frequency that the antecedent and consequent occur divided by the total number of transactions. This metric seems low as it is a percentage, but it is used in calculating other more helpful metrics.

The confidence of the top three rules is approximately 0.5 for all. This means that approximately 50% of the time when a customer purchased the group of products in the antecedent, they also purchased the product in the consequent. This metric is calculated using the combined support of the antecedent and the consequent, divided by the support of the antecedent (Hull, n.d).

Finally, the lift for the top three rules ranges between 2.1 and 2.4. Greater lift values mean stronger associations between items. This metric is calculated as the combined support of the antecedent and the consequent, divided by the support of the antecedent multiplied by the divided by the support of the consequent. A lift value that is larger than one means that the items "occur in transactions together more often than we would expect based on their individual support values. This means that the relationship is unlikely to be explained by random chance" (Hull, n.d).

**Practical Significance of Findings**

Based on the measurements described above, there is a strong relationship between the items present in the top three association rules. Also, when looking at all of the rules created, it appears that Dust-Off Compressed Gas 2 pack is present as a consequent in all but one of the rules created using the chosen metrics with a confidence metric of approximately 0.5. The 10ft iPhone Charger Cable 2 Pack and VIVO Dual LCD Monitor Desk Mount are each present in two of the three association rules.

This gives the company three items to consider discounting. Applying a discount to those items might increase the likelihood for additional customers to buy the items, which may foster brand loyalty. This loyalty may result in a lower level of customer churn.

My recommended course of action for the stakeholders of the telecommunications company is to offer a discount for the 10ft iPhone Charger Cable 2 Pack and VIVO Dual LCD Monitor Desk Mount. Additionally, the Dust-Off Compressed Gas 2 pack should be offered as a free bonus item or with a deep discount. Compressed air is usually a fairly cheap product to purchase, so the trade-off of any revenue lost by giving it away may be made up for with the increased customer loyalty. This loyalty may decrease customer churn, which would be much better for the bottom line.

**Sources**


- Hull, I. (n.d.). Market Basket Analysis in Python. Retrieved February 24, 2021, from https://learn.datacamp.com/courses/market-basket-analysis-in-python (https://learn.datacamp.com/courses/market-basket-analysis-in-python)
- Moffitt, C. (2017, July 03). Introduction to market basket analysis in python. Retrieved February 24, 2021, from https://pbpython.com/market-basket-analysis.html (https://pbpython.com/market-basket-analysis.html)
- Tian, H. (2015, September 01). Market Basket Analysis. Retrieved February 24, 2021, from https://sarahtianhua.wordpress.com/portfolio/market-basket-analysis/ (https://sarahtianhua.wordpress.com/portfolio/market-basket-analysis/)


**Helpful Sites Used in Coding Project**

1. https://stackoverflow.com/questions/57790623/how-to-remove-nan-from-list-of-lists-with-string-entries (https://stackoverflow.com/questions/57790623/how-to-remove-nan-from-list-of-lists-with-string-entries)
2. https://campus.datacamp.com/courses/market-basket-analysis-in-python/introduction-to-market-basket-analysis (https://campus.datacamp.com/courses/market-basket-analysis-in-python/introduction-to-market-basket-analysis)