D212 – DATA MINING II

Task 3: Association Rules and Lift Analysis

WGU - MSDA

Advanced Data Mining Association Rules and Lift Analysis with Churn Dataset

Richard Flores

Rflo147@wgu.edu

Student ID: 006771163

Table of Contents

Part I: Research Question

A1. Proposal of Question

From our previous Churn Dataset analysis, we have gathered a plethora of insights into customer characteristics and behavior that helps us shape the course of action in the telecommunications company to reduce customer churn and improve annual profits.

This goal of this analysis is to analyze customer transactions for items purchased from the telecom's stores across all regions of the United States using a market basket analysis. The market basket analysis will answer the question:

What are the purchasing habits of customers and which item discounts will provide interest and incentives for customers to retain service?

A2. Defined Goal

The goal of the market basket analysis is to provide Shareholders and Executives with information about customer purchasing habits. The analysis of purchasing habits can be used to determine the items that are the best in which to provide discounts and promotions in order to minimize the risk of customer churn. The information on effective discounts and promotions can be utilized by the marketing department to provide targeted offers to subscribers in conjunction with telecom subscriptions with the ultimate goal of reducing customer churn.

Part II: Market Basket Justification

B1. Explanation of Market Basket

A Market Basket Analysis, or MBA, is a great tool for increasing the effectiveness of marketing and also a great way to improve sales campaigns using collected customer sales transaction data. In short, an MBA is a modelling technique built on the assumption that if a customer purchases a certain group of items, they are more likely to purchase another group of items. MBA is also one of the key modelling techniques utilized by big-box stores to discover associations between items (Albion 2022).

A Market Basket Analysis is a Data Mining technique built upon the Mathematical concepts of *Support*, *Confidence, and Lift* which are expressed as follows:

SUPPORT	P(A U B)
CONFIDENCE	P(A U B)/P(A)
LIFT	P(A) / (A U B)

Wherein Support is the probability of A union B and Confidence is the Support of a product, previously calculated by P(A U B), divided by the second product (Albion 2022). For example, if a customer enters the telecom store and purchases a mobile phone but does not purchase a glass shield, what is the probability the customer will purchase a phone case?

In the example above, the customers transaction is an *itemset.* The selection of a mobile phone but no glass shield is the *support.* The probability of purchasing a mobile phone case is the *confidence.* This is a simple, yet effective algorithm for predicting customer purchases.

We can use the Market Basket Analysis on the provided *Market Basket (Churn)* Dataset to produce meaningful Support and Confidence algorithm calculations to assist our marketing and retention departments.

From our Market Basket Analysis, we expect the outcome will uncover insights on optimized bundles of items that can be offered at discounted prices in conjunction with telecom subscription services.

B2. Transaction Example

An excellent example of a transaction from the dataset includes all of the available 20 items purchased in a single transaction. The items available vary in category, but are all related to technology and include things such as computer accessories, memory storage, ink, and cleaning materials.

Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10
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

Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20
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

B3. Market Basket Assumption

The assumption of the Market Basket algorithm is based on Association Rules which can be used to analyze the transaction data. The Association Rules identify strong rules uncovered in the transaction data by utilizing measures of interestingness which are based on the concept of strong rules. The end result will reveal summary of quality measures including ranges of support, confidence, and lift (Li 2017).

To answer our research question in this analysis, our goal is to use the assumed association rules in the Market Basket Analysis to discover items that may be offered at discount before or during telecom subscription services.

Part III: Data Preparation and Analysis

C1. Transforming the Dataset

```
# Import DataScience Libraries for calculations, visualizations, and plots
import numpy as np
import pandas as pd

import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

import matplotlib as mpl

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

# Skip warning messages for cleaner output & display
import warnings
warnings.filterwarnings('ignore')
```

```
# Read Market Basket Dataset into Dataframe
teleco = pd.read_csv('teleco_market_basket.csv')
```

```
# Verify records present in dataset
teleco.head()
```

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

```
# Display descriptive statistics of basket dataset
teleco.describe()
```

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15
count	7501	5747	4389	3345	2529	1864	1369	981	654	395	256	154	87	47	25
unique	115	117	115	114	110	106	102	97	88	80	66	50	43	28	19
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 SB820 Cable Modem
freq	577	484	375	201	153	107	96	67	57	31	22	15	8	4	3

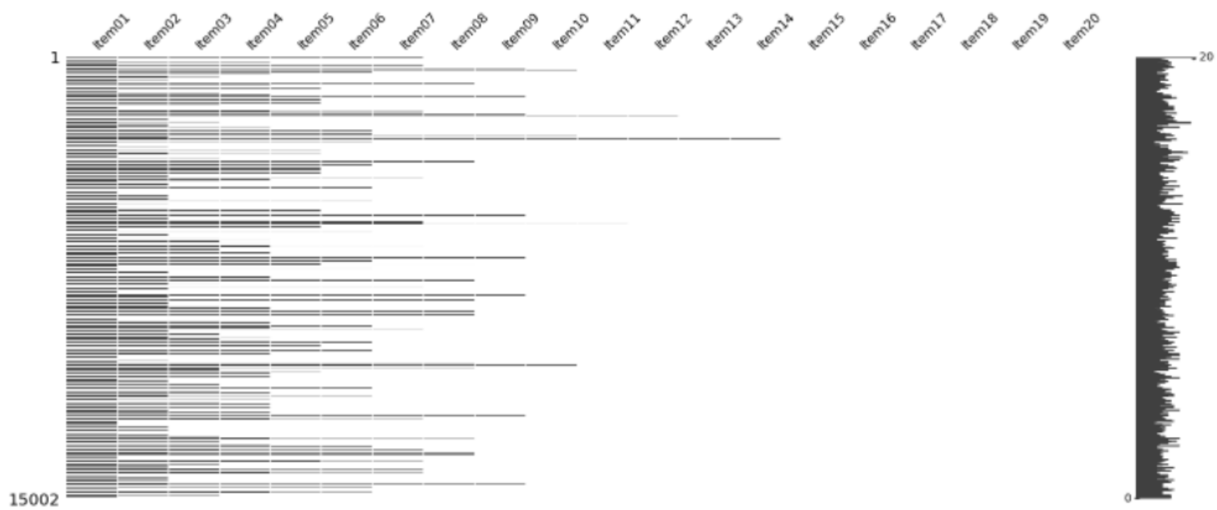
(DataCamp 2022)

```python
# Utilize missingno library to visualize missing data

!pip install missingno

import missingno as msno

# Display visualization
msno.matrix(teleco);
```




```python
# Remove records with no data contained
teleco.dropna(how='all', inplace=True)

# Verify changes
teleco.head()
```

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16
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
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
5	UNEN Mfi Certified 5-pack Lightning Cable	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
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

```python
# Changes values containing NaN to 0 for calculation
teleco.fillna(0, inplace=True)
```

```python
# Display new record size and verify features
teleco.shape
```

```
(7501, 20)
```

(DataCamp 2022)

```python
# Verify NaN values now display 0 value
teleco.head()
```

	Item01	Item02	Item03	Item04	Item05	Item06	Item07	Item08	Item09	Item10	Item11	Item12	Item13	Item14	Item15	Item16
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
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
5	UNEN Mfi Certified 5-pack Lightning Cable	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
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

```python
# Convert telco dataframe into a list for further calculation
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)
```

```python
# Verify cleaned dataframe
teleco_cleaned.head()
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
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
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
2	UNEN Mfi Certified 5-pack Lightning Cable	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
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

```python
# Extract cleaned market basket dataset
teleco_cleaned.to_csv('market_basket_prepared.csv')
```

(DataCamp 2022)

C2. Code Execution

```
# Install Apriori to utilize algorithm for association rules
!pip install apyori
from apyori import apriori

# Use algorithm to train dataset
rule_list = apriori(teleco_list, min_support = 0.003, min_confidence = 0.3, min_lift = 3, min_length = 2)
```

```
# List generated rules from Apriori algorithm
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_stat
istics=[OrderedStatistic(items_base=frozenset({'5pack Nylon Braided USB C cables'}), items_add=frozenset({'HP 63XL Ink'}), conf
idence=0.3006993006993007, lift=3.790832696715049)])
```

```
# Display amount of rules generated
print(len(rule_list))
```

```
102
```

```
# Create results variable from Apriori rules list
results = pd.DataFrame(rule_list)
```

```
# Display results
results
```

	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 Ultra 64GB card, SanDisk 128GB Ultra ...	0.015998	[((SanDisk 128GB Ultra microSDXC card), (SanDi...
...	...	...	...
97	(Dust-Off Compressed Gas 2 pack, Nylon Braided...	0.004399	[((Dust-Off Compressed Gas 2 pack, VIVO Dual L...
98	(Dust-Off Compressed Gas 2 pack, VIVO Dual LCD...	0.003200	[((Dust-Off Compressed Gas 2 pack, VIVO Dual L...
99	(Nylon Braided Lightning to USB cable, VIVO Du...	0.003066	[((Nylon Braided Lightning to USB cable, HP 61...
100	(Nylon Braided Lightning to USB cable, VIVO Du...	0.003466	[((VIVO Dual LCD Monitor Desk mount, Nylon Bra...
101	(Nylon Braided Lightning to USB cable, VIVO Du...	0.003066	[((Nylon Braided Lightning to USB cable, Scree...

102 rows × 3 columns

```
# Create support variable from results support calculation
support = results.support
```

```
# Create variables for LHS, RHS, Confidence and Lift
first_values = []
second_values = []
third_values = []
fourth_values = []
```

```
# Iterate over list using For Loop
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])
```

```
# Transfer lists into dataframes
lhs = pd.DataFrame(first_values)
rhs = pd.DataFrame(second_values)
confidence = pd.DataFrame(third_values, columns=['confidence'])
lift = pd.DataFrame(fourth_values, columns=['lift'])
```

(Kumar 2020)

```python
# Create master list from LHS, RHS, Support, Confidence and Lift results
results_final = pd.concat([lhs, rhs, support, confidence, lift], axis=1)
results_final.fillna(value=' ', inplace=True)
```

```python
# Display results
results_final
```

	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	Nylon Braided Lightning to USB cable	HP 61 ink	SanDisk Ultra 64GB card	0	VIVO Dual LCD Monitor Desk mount		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	Nylon Braided Lightning to USB cable	Screen Mom Screen Cleaner kit	SanDisk Ultra 64GB card	0	VIVO Dual LCD Monitor Desk mount		0.003066	0.534884	3.072100

102 rows × 9 columns

C3. Association Rules Table

```python
# Create column names for accessibility
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
```

	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	Nylon Braided Lightning to USB cable	0	0.003066	0.534884	3.072100
100	VIVO Dual LCD Monitor Desk mount	0	0.003466	0.440678	3.400746
101	Nylon Braided Lightning to USB cable	0	0.003066	0.534884	3.072100

102 rows × 5 columns

```python
# Display completed list of rules
results = list(rule_list)
for i in results:
    print('\n')
    print(i)
    print('**********')
```

The Association Rules Table above accurately displays the values for 102 records of transactions listing the necessary *Support*, *Confidence*, and *Lift* results.

C4. Top Three Rules

The completed Market Basket Analysis reveals the Top 3 Rules generated by the Apriori Algorithm as follows:

lhs	rhs	support	confidence	lift
5pack Nylon Braided USB C cables	HP 63XL Ink	0.005733	0.300699	3.790833
AutoFocus 1080p Webcam	SanDisk Ultra 64GB card	0.005333	0.377358	3.840659
iPhone 11 case	HP 63XL Ink	0.005866	0.372881	4.700812

Number One Rule

The number one rule calculated shows that if '5-pack Nylon Braided USB-C Cables' are purchased then 'HP 63XL Ink' is also likely to be purchased. The MBA metrics show a support value of 0.005733 signifying that just above 50% of a percentage of all transactions in the dataset show both items purchased together. From all customers who purchased the USB-C cables confidence shows 30% also purchased HP Ink. The Lift value demonstrates that customers who purchase the USB-C cables are 3.79 times more likely to purchase HP Ink in the same transaction.

Number Two Rule

The number two rule calculated shows that if 'AutoFocus 1080p Webcam' is purchased then 'SanDisk Ultra 64GB card' is also likely to be purchased. The MBA metrics show a support value of 0.005333 signifying that just above 50% of a percentage of all transactions in the dataset show both items purchased together. From all customers who purchased the AutoFocus Webcam confidence shows 37% also purchased the SanDisk memory card. The Lift value demonstrates that customers who purchase the AutoFocus Webcam are 3.84 times more likely to purchase SanDisk memory card in the same transaction.

Number Three Rule

The number three rule calculated shows that if 'iPhone 11 case' is purchased then 'HP 63XL Ink' is also likely to be purchased. The MBA metrics show a support value of 0.005866 signifying that just above 50% of a percentage of all transactions in the dataset show both items purchased together. From all customers who purchased the iPhone case confidence shows 37% also purchased HP Ink. The Lift value demonstrates that customers who purchase the iPhone case are 4.70 times more likely to purchase HP Ink in the same transaction.

Part IV: Data Summary and Implications

D1. Significance of Support, Lift, and Confidence Summary

The significance of Support, Lift, and Confidence can best be expressed by first defining the meaning of each then examining the item's relevance.

TowardsDataScience defines support as the popularity or frequency of occurrence of an item. Support can be calculated by the number of transactions containing the item to the total number of transactions or expressed as Support = Frequency (X, Y) / N (Goyal 2020).

In our Market Basket Analysis, analyzing the results of the Apriori algorithm for the Top 3 Rules shows a maximum support value of fifty percent of just one percentage point which does not show a strong case for supporting this metric as the basis of a bundled discount.

Confidence in the Apriori algorithm is the likelihood of occurrence of item Y if item X occurs or the conditional probability which can be expressed as Confidence (X => Y) = (X U Y) / X (Goyal 2020).

From the Market Basket Analysis, we see in the Top 3 Rules confidence levels of 30%, 37%, and 38%. The first rule of the Apriori algorithm shows a lackluster confidence score of only 30%.

Lift is defined as the increase in ratio of an occurrence of item Y if item X occurs which can be expressed as Lift (X => Y) = Confidence (X, Y) / Support (Y) (Goyal 2020).

The Market Basket Analysis shows some promise in the Lift metric as in the Top 3 Rules we have Lift values of 3.8, 3.8, and 4.7. From this metric we can estimate that customers who purchase a phone case are almost 5 times more likely to purchase Ink in the same transaction.

As observed in the metrics of the Top 3 Rules, the calculations from the Apriori algorithm show lackluster results in the form of Support, Confidence, and Lift. In order to provide a recommendation to shareholders and the marketing department, we should aim to achieve a confidence level in excess of 80%. Unfortunately, our highest confidence metric achieved only 38%.

D2. Practical Significance of Findings

The Market Basket Analysis, calculated using the Apriori algorithm, is designed to provide recommendations for bundles of items based on the first item selected for purchase. The metrics we observe and provide recommendations on are based on the calculated variables Support, Confidence and Lift. Unfortunately, it seems that due to a variety of possible reasons, the transaction dataset fails to provide meaningful and more importantly useful information about customer purchasing habits.

The support metric, which measures the popularity of an item, at its peak has a value of 0.5% of all purchases signifying that there is no clear commonly purchased item from the telecom store, as we would expect to see a supermarket similarly have a high support metric for selling milk or eggs. Furthermore, the confidence metric has a maximum calculated value of 37% which is not adequate for providing a recommendation to the marketing department as the value should show a measurement of about 80% or higher. Finally, the Lift metric provides some surprising results as it does show an increase of customers likely to purchase a second item if they purchase the first. In our metrics we see that if a customer purchases a Webcam, they are 3.8 times more likely to purchase a memory card as well, and this would make for a solid recommendation. However, similarly the lift metric shows that if a customer purchases an iPhone case, they are 4.7 times more likely to purchase HP ink. This indicates that there may be some issues with the provided dataset, most likely that it was randomly generated for the analysis, because the metric is implausible for a recommendation similar to saying that if a customer in a supermarket purchases bread, they are likely to purchase cat food as well.

Overall, we find that the Apriori generated metrics for the Market Basket analysis are underwhelming at best and do not provide significant insight into customer purchasing habits or recommendations for discounts to be offered by the marketing department.

D3. Course of Action

It is strongly recommended to shareholders and executives at the Telecom Company that we do not use the metrics generated in this Market Basket Analysis for consideration of customer purchasing habits or items recommended and discounted in an effort to reduce churn. The results of the Apriori metrics show an underwhelming level of insight into transaction patterns and do not show a clear, logical approach to customer purchase recommendation.

The best course of action moving forward would be to analyze how the transaction data is gathered and either correct or improve the reporting process. With an improved dataset we can revisit the Market Basket Analysis to generate insights into customer purchases and use this new information to create recommendations for items and item discounts.

Part V: Attachments

E. Panopto Recording

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=d724ee79-f809-4680-ad3d-ae73003f445e

F. Sources for Third-Party Code

Goyal, S. (2020, February 6). *Demystifying customer behavior with market basket analysis*. Medium. Retrieved April 9, 2022, from https://towardsdatascience.com/demystifying-customer-behavior-with-market-basket-analysis-87b5841def3a

DataCamp. (n.d.). *Market Basket Analysis in Python*. DataCamp. Retrieved April 10, 2022, from https://app.datacamp.com/learn/courses/market-basket-analysis-in-python

Kumar, D. V. (2020, October 24). *Hands-on guide to Market Basket Analysis with python codes*. Analytics India Magazine. Retrieved April 10, 2022, from https://analyticsindiamag.com/hands-on-guide-to-market-basket-analysis-with-python-codes/

G. Sources

Albion (n.d.). *Market basket analysis*. What is Data Mining. Retrieved April 8, 2022, from https://www.albionresearch.com/data_mining/market_basket

Li, S. (2017, September 27). *A gentle introduction on Market Basket Analysis - Association rules*. Medium. Retrieved April 9, 2022, from https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce

Goyal, S. (2020, February 6). *Demystifying customer behavior with market basket analysis*. Medium. Retrieved April 9, 2022, from https://towardsdatascience.com/demystifying-customer-behavior-with-market-basket-analysis-87b5841def3a