W. P. CAREY SCHOOL of BUSINESS

ARIZONA STATE UNIVERSITY

W. P. CAREY SCHOOL OF BUSINESS HONOR CODE MASTER'S PROGRAMS

We, the students of the W.P. Carey Master's programs, are committed to maintaining the highest levels of honesty, integrity, and professional courtesy for ourselves and amongst our classmates. We believe that ethical behavior is not only the cornerstone of a good academic program, but of good business as well. As such, we are steadfast in our commitment to the following principles:

- Academic Integrity – We view the W.P. Carey Academic Integrity Policy as a living, breathing document, not just concepts on a page. As such, we are resolute in ensuring that we abide by the principles contained within, both in letter and in spirit. Furthermore, we will hold our classmates to the same high standards to which we hold ourselves.

- Consideration of Others – As W. P. Carey students, we believe that we should treat others as we wish to be treated. To this end, we will be considerate of classmates, professors, and recruiters in all aspects. Furthermore, we will constructively add to the collaborative environment by helping and encouraging others.

- Professionalism – As W. P. Carey students, we believe that the behaviors we practice today will become habit in the future. To that end, we are committed to act in a professional manner in all school settings, including classes, networking events, and interviews. We will be cognizant of the fact that we are representing the program and our classmates at all times.

Full Name (printed): Sreevatsan Agoramurthy

Signature____ Date ____________________

Full Name (printed): Yi-Wun Chen

Signature____ Date ____________________

Full Name (printed): Marlee Coolidge

Signature____ Date ____________________

Full Name (printed): Mohammadreza Aramimehr

Signature____ Date ____________________

Onsite MSBA Applied Project Report

Spring 2020

W.P. Carey, ASU

Team C-1	Sreevatsan Agoramurthy Yi-Wun Chen Mohammadreza Aramimehr Marlee Coolidge
Topic	Customer Churn Analysis
Client	Cable One 210 East Earll Drive, Phoenix - 85012

Executive Summary

Cable One, a provider of phone, video, and internet services, battles against an industry wide challenge: customer churn. It's a well known fact that acquiring a new customer costs upto 5 times more than retaining an existing customer. This is the foundation of our objective here, to gain a competitive advantage against other competitors , CableOne needs to increase its customer retention rates. Our team was tasked with exploring customer data to identify potential causes of the attrition and recommend possible solutions to deter it.

Our team used business intelligence tools like Tableau and data analysis tools like the libraries of Python to perform exploratory data analysis, correlation matrices, and build a classification model to predict if customer churn. Through these analysis, we made the following key findings:

- High attrition rates in the first 36 months of a customer's tenure
- Promotions help with customer retention, but lead to churn when promotion period ends
- Forced upgrades of a customer's plan cause increased customer churn
- The more plans a customer is on, the less likely they are to leave
- LifeTimeValue, a value used by CableOne to monitor customer churn, performs well
- Phone and Video services see significantly higher rates of churn
- Large 'One Time charges' seem to cause customer churn

Given the above findings, our team used the provided data and external business research to recommend the following solutions:

- Customer satisfaction does seem to be high, but we have to consider improving customer experience in the first 3 years of their tenure.
- An increase in number of promotions offered or reducing the price reduction when a customer is on a promotion
- Forced upgrade is a major problem
- Increase the number of bundled packages sold (ie) offer phone and video services only as a bundle with internet
- Implement zero one-time charges as much as possible
- Since customers are satisfied with low rates, this offers us some wiggle room to reduce one time charges
- Refrain from offering promotions during the first 1-2 years of a customer's tenure

The above findings and recommendations assume that the data provided to us was representative of the overall customer base for Cable One. The findings are also limited to the details provided only about demographics, usage, products, and orders.

Background

Cable One is a nationwide, American company headquartered in Phoenix, AZ who provides internet and cable services. The company provides services such as phone service, internet connection, and cable. Cable One's customers pay one time fees for any hardware they may own (like a router) and then subscription fees for the ongoing services (like a monthly phone plan).

Sparklight lists CableOne's Pros as having "no contract required, affordable prices, highest customer satisfaction ratings, and no standard price increase on internet packages after the promo period”. The website also lists Cable One’s cons being “modem lease required, activation fee ($40) and the first month’s payment required at signup, slow upload speeds, and price increases on bundle packages after promo period”. Broad Band Now also rated the company as a 2.3 out of 5 stars for overall performance.

Apart from knowledge about Cable One, it is known that the cable and internet service industry has changed very much over the past decade due to a sharp decrease in cable demand as internet streamed television has become popular. Also, the need for in-home internet services has increased. Many companies have suffered under these changes, yet Cable One has remained prominent and well liked by customers through the years as we can see by google trend reports available on Google.

A primary concern for Cable One is customer attrition. In order for the company to grow it is vital that Cable One retain all the customers it currently has while gaining new ones. In order to keep all of the current customers Cable One needs to know why customers have or are leaving and shape their business to prevent these pitfalls.

Problem Statement

Cable One’s challenge is to understand customer churn through a variety of analytics methods. They want to identify root causes for customer attrition and be provided with solutions to these issues. In order for Cable One to have a competitive advantage over their competitors they need to keep their customers as happy and content as possible. The data analytics team at Cable One can substantiate our analysis and recommendations. They can then explain these findings to the marketing, sales, and customer service teams to adjust the business processes as needed to maximize customer satisfaction and therefore maximize customer tenure.

The following analysis and solutions assume that Cable One has provided our analytics team with data representative of the business situation they are in. Our solutions also assume that the four data sources we were provided represent the customer journey accurately. Our solutions only reference the customer demographics, product information, usage information, and order information.

Methods

Our team approached the project in the following steps:

1. Cleaning the data through preprocessing to make it useful to work with
2. Exploring each dataset individually
3. Exploring all datasets together
4. Plotting key infographics – revealing possible causes of defection
5. Building a random forest classification model to predict when a customer is likely to leave
6. Reporting our final findings

By viewing the data sets individually we knew which traits were prominent and trendy to each category of data. This was efficient because when it came time to connect all of the data sets and overlay the information from different data sets we had a good idea of what to look for.

Through visualizing in Tableau and measuring correlation using Python script we were able to narrow our key findings to the most important relationships.

We also approached the project this way so that every team member could have a part and voice in the exploratory data analysis. Our key findings and recommendations were strengthened through the diversity of our educational and work experience backgrounds.

At the start of the project our team was provided with four well organized data sets from Cable One that were used to conduct the analysis. The 4 data sets are as follow:

1. Demographics:

Contains key information on customer's
- Geographical location
- Lifetime Value
- Subscriberwise score
- Clusters defined by ConneXions
- Account information
- Customer Status – Active/Disconnected

Number of customers identified = 252573 unique customers

2. Products:
   Contains product information that a customer has subscribed to/owns.
3. Usage:

Contains information on product usage by the customer –
- Amount of internet usage : Upload, Download, Overall
- Number of warnings given to customer for over-usage
- Amount of over-usage of data
- Forced upgrade information

4. Orders:

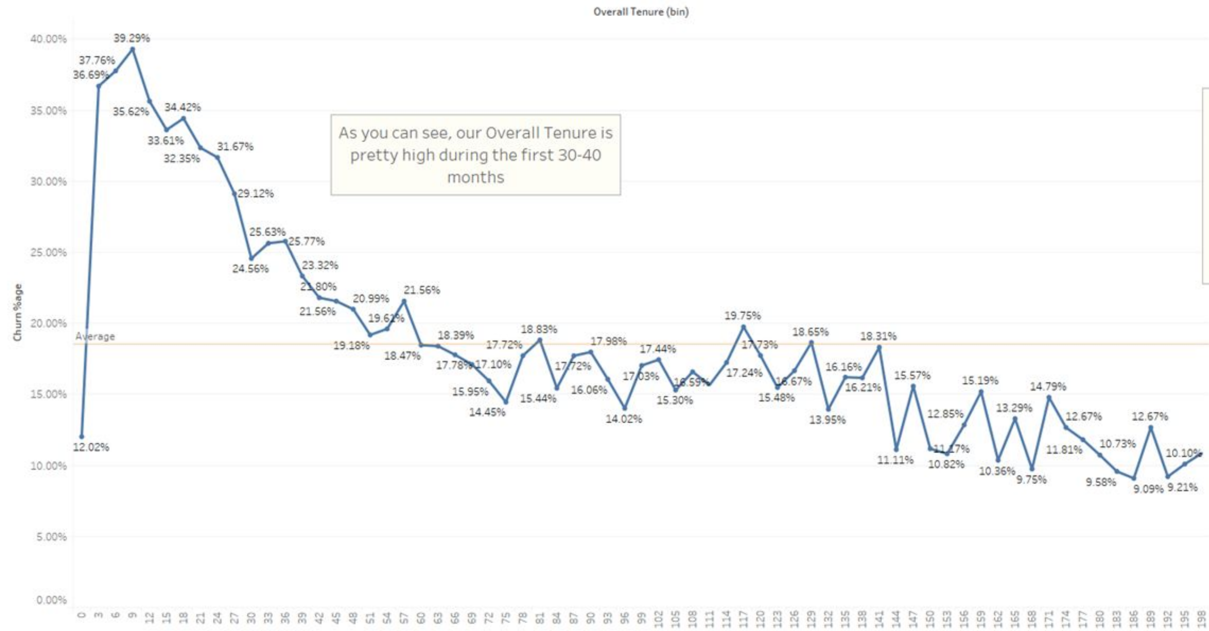
Contains information on orders placed by customer
- Addition of services
- Removal of services
- Change of services
- General service

Due to the organized and complete structure of the data upon arrival, there was no need to need to perform significant data cleaning. We used Python majorly to preprocess each dataset, impute missing values and perform a preliminary exploratory data analysis. Later, we leveraged the use of Tableau to analyze the combined dataset and build dashboards that assisted us in deriving insights and providing recommendations

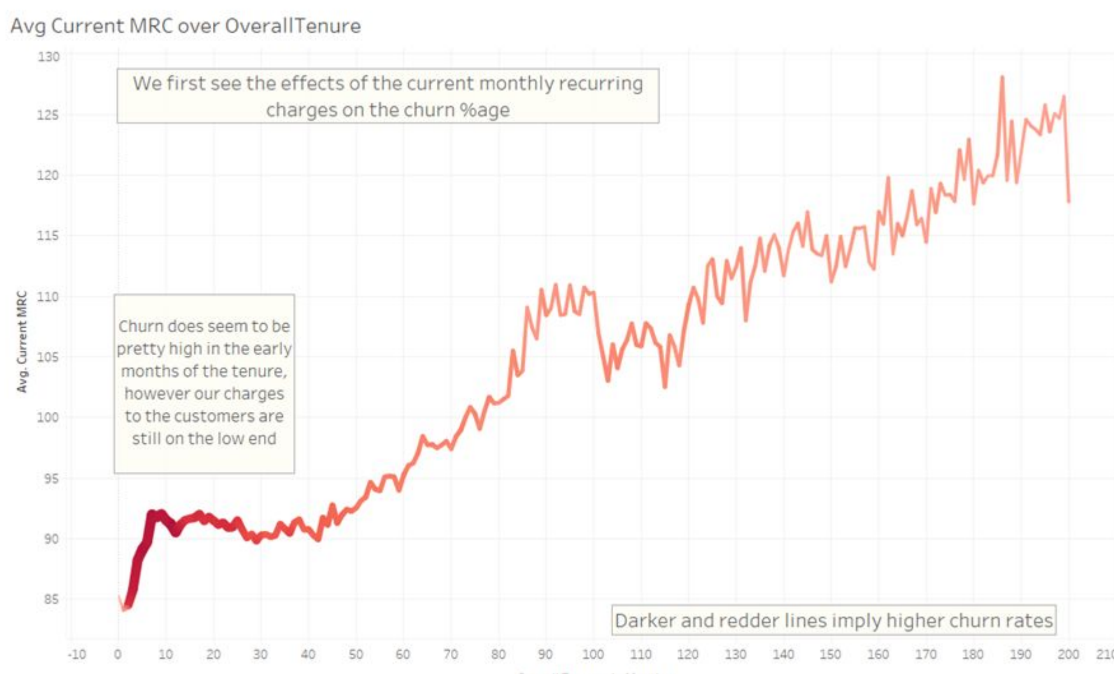
Results and Conclusions

Although many observations were made throughout the exploratory data analysis performed for this applied project our team has included here the most influential key findings that drove our analysis and final recommendations:

The first key finding is that high attrition rates exist in the first 36 months of a customer's tenure. After 3 years of being with Cable One, a customer becomes less and less likely to churn as. This trend can be seen by the high blue bar on the following graph that shows the increased churn rate that decreases significantly after the 36th month.

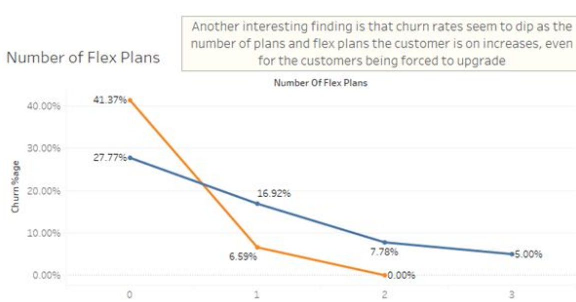


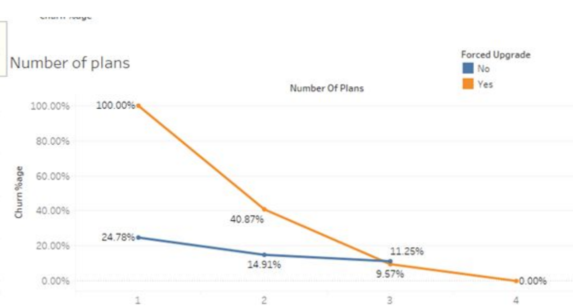
Next it can be seen that promotions help with customer retention, but lead to churn when promotion period ends. The darkest red part on the trend line indicates that more customers are leaving Cable One in the first few months or years as their promotions are ending and normal pricing and services are kicking in.



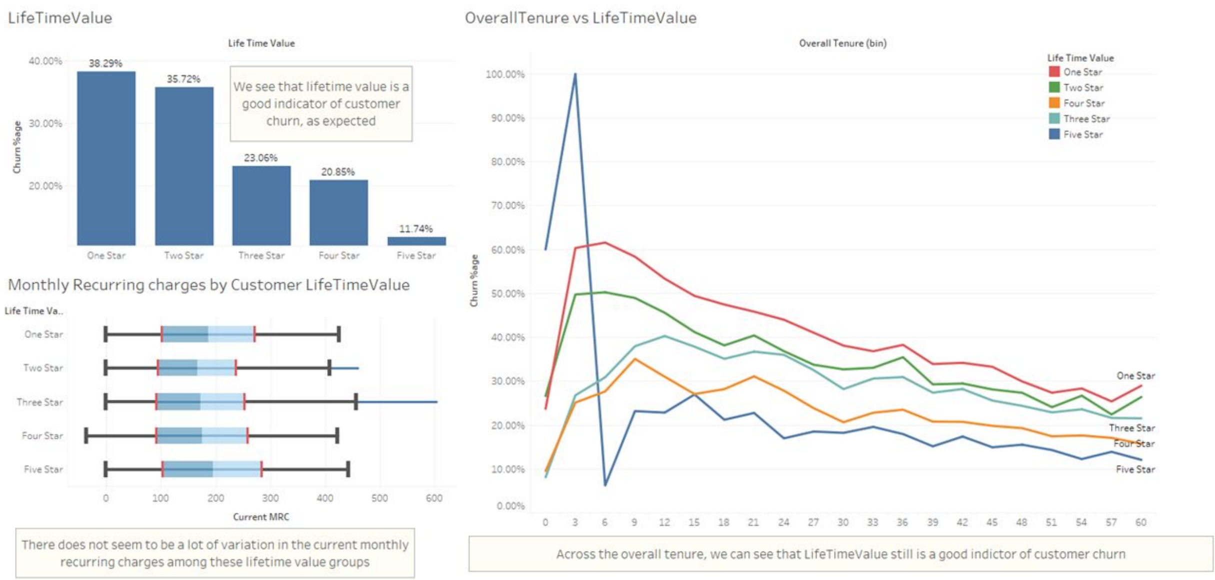
As we analyzed what other financial factors influence customer church we can see in the figure to the left that forced upgrades cause increased customer churn. When a customer over-uses their data and bandwidth too many months in a row Cable One forces them to upgrade their service. The much taller blue bar (41%) represents the church rate amongst customers who were forced to upgrade where the lower blue bar(26%) represents the lower churn rate amongst customers who were not forced to upgrade their plan. Therefore, customers are more likely to leave if their plan is forced to upgrade.

The next finding is that the more plans a customer is on, the less likely they are to leave. The below visual shows how even customers who are more likely to churn because they are forced to upgrade will not leave the company as often when they have multiple plans or services through Cable One.

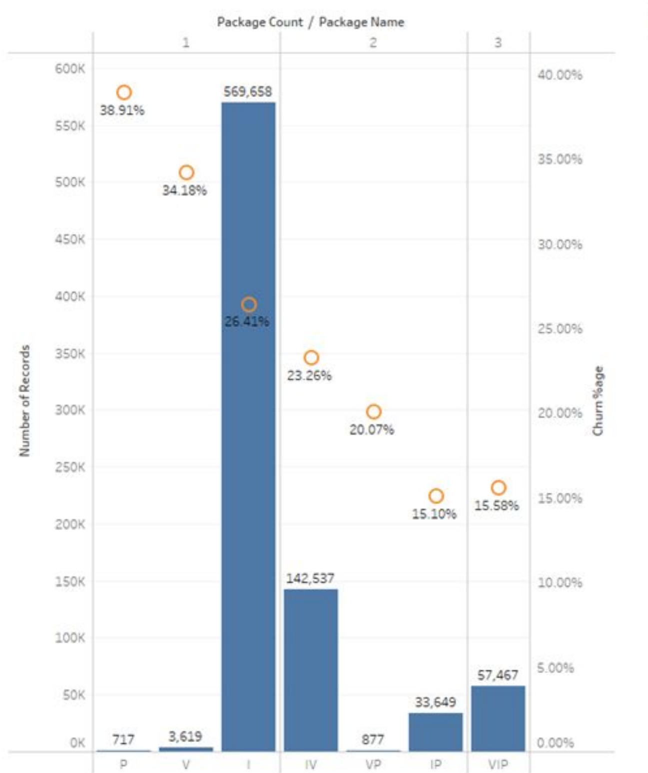




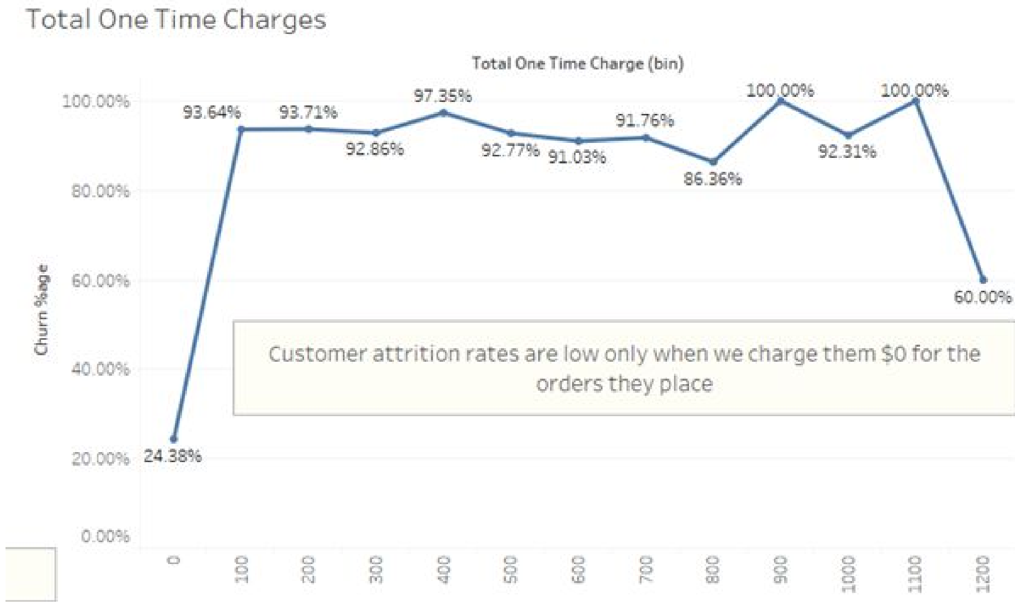
Cable One ranks their customers with a five star scale to indicate their lifetime value to the company. This LifeTimeValue is a strong indicator of customer churn. Customers with lower value tend to leave the company more frequently. This finding is important for sales and customer service teams to know how to prioritize certain groups of customers.



Apart from the financial terms or longevity of a customer's experience with Cable One, the type of product they use through the company can play a part in the likelihood of staying. Phone and Video services have significantly higher rates of churn. These are the product lines that finance or sales teams should focus on bettering the customer experience.

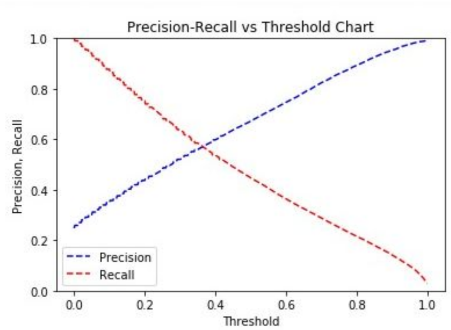


Finally, it can be seen that one time charges seem correlated to customer churn. The below chart shows that the only really low customer attrition rate exists where there are no one time charges. Any other amount of one time charges exist with a much higher attrition rate.



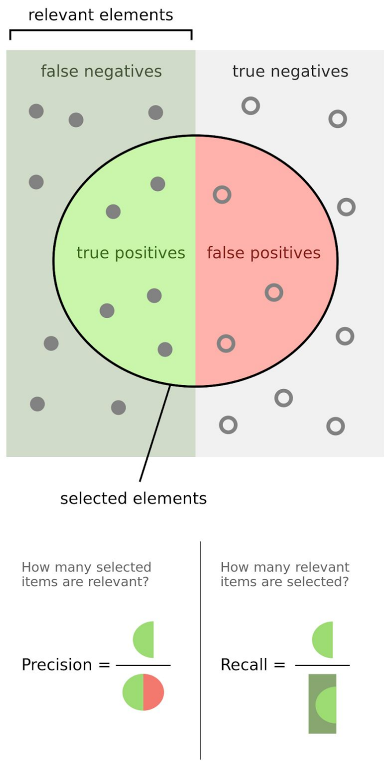
Our findings are based on the four provided data sets (demographics, usage, products, and orders) that range in time period from January of 2018 to October of 2019.

In addition to the exploratory data analysis that was documented above, our team also performed a classification model on the provided data to try and predict if a customer were going to churn or not. The random forest classifier model can be seen below:




```
In [90]: print(metrics.classification_report(y_test,preds_final.iloc[:,1]))
              precision    recall  f1-score   support

           0       0.86      0.85      0.86    119776
           1       0.56      0.59      0.57     39302

    accuracy                           0.78    159078
   macro avg       0.71      0.72      0.71    159078
weighted avg       0.79      0.78      0.79    159078

In [91]: metrics.recall_score(y_test,preds_final.iloc[:,1])
Out[91]: 0.5852373924991094

In [92]: metrics.roc_auc_score(y_test,preds_final.iloc[:,1])
Out[92]: 0.7169357547587719

In [93]: metrics.confusion_matrix(y_test,preds_final.iloc[:,1])
Out[93]: array([[101646,  18130],
               [ 16301,  23001]], dtype=int64)
```

A key point to note here would be the accuracy score of 0.78 and a class 1 recall score of 0.59. To better explain our findings , we use the below image to shine more light on precision and recall scores.



Precision denotes the number of selected items that are relevant. In this scenario it represents how many of the predicted churn customers are actually churn customers, which 56%. However, we are more interested in the recall scores. Recall is defined as the number of relevant items that are selected. In our case, 59% of the churn customers are identified correctly. This is our primary objective and hence we should be looking to increase this score. A primary law of adjusting the thresholds in a classification model is that recall and precision score are inversely proportional according to the curve given in the previous image. CableOne can take the gamble of falsely predicting a non-churn customer if it means they can identify a higher number of churn customers. Furthermore, the classification model returns us probabilities of being classified in a particular class. This information could be useful to CableOne to prioritize their customer retention efforts.

Although the above model would need to be reinforced before Cable One implemented its use, with a more accurate model the customer service team at Cable One could use the model to identify customers who might leave the company. They could specifically target these customers with special treatment or promotions to prolong the relationship with the customer and decrease overall customer attrition.

Recommendations

Given the above findings, our team used the provided data and external business research to recommend the following solutions:

- Customer satisfaction does seem to be high, but we have to consider improving customer experience in the first 3 years of their tenure. This could be accomplished through new client videos that explain the online payments, services, customer care and services available, or other important pieces of information that could make a client feel like they belong with the company even early on in their history.
- An increase in the number of promotions offered could keep clients at Cable One even after their original promotion. Alternatively, reducing the price reduction when a customer is on a promotion could make the transition from promo to non-promo less distasteful to customers.
- Forced upgrades are a major problem, therefore:
  - Educate sales representatives to better understand customer needs thereby selling appropriate plans

    - Drop plans that have a low data and bandwidth limit so that the customer isn't inevitably forced to upgrade when it is not enough
    - Provide customers with a 'Pay as you go' plan in the first 2 months to gauge their usage requirements before committing to a product plan
    - Offer promotion immediately after a customer is forced to upgrade so they are less likely to leave at the time of the price jump
- Focus on selling more bundled packages because customer attrition is low amongst customers who have bundle packages
- Offer phone and video services only as a bundle with internet to support the above solution and lower the attrition rates amongst these individual product lines
- Implement zero one-time charges as much as possible
- Customers are more satisfied with low rates which offers Cable One the flexibility to reduce one time charges
- Refrain from offering promotions during the first 1-2 years of a customer's tenure
    - Reduces cost to company when a large proportion of customers leave in the beginning anyway
    - Instead offer promotions after 24-30 months to keep the customer enticed beyond the 3 year mark where customer attrition naturally declines

Conclusion

To conclude our analysis of customer churn at Cable One our team has provided a SWOT Analysis based on the data driven observations:

Strengths ● Low Pricing ● Customer Satisfaction ● Wide range of plans ● No contracts required	Weaknesses ● High churn rates ● Marketing strategies ● Flexibility of plans
Opportunities ● COVID-19 crisis increases demand ● Large internet corporations have poor customer service ● Market is still untapped in rural areas where company already has presence	Threats ● Increase in popularity of video streaming services like Netflix, Amazon Prime Customer Satisfaction ● Competition is very intense

Future work or continuation on this project could include the following:

- Include data from customer surveys before & after the following customer experiences
    - Promotion
    - Product offers
    - Forced upgrade

    - Exit
- A time series classification model is too complex for us to attempt right now and hence out of scope. However with a deep learning model such as a LSTM, we can attempt to predict customer churn while taking into consideration seasonality trends.
- Obtaining marketing data to supplement churn analysis by geographical region

References

BlastChar. "Telco Customer Churn." *Kaggle*, 23 Feb. 2018, www.kaggle.com/blastchar/telco-customer-churn.

"Cable One Availability Map: Cable One Packages." *HighSpeedInternet.com*, www.highspeedinternet.com/providers/cable-one.

"Cable ONE Captures Top Spot in the West in J.D. Power Residential Internet Satisfaction Study." *Cable ONE Captures Top Spot in the West in J.D. Power Residential Internet Satisfaction Study*,www.sparklight.com/press/Pages/Cable-ONE-Captures-Top-Spot-in-the-West-in-J.D.-Power-Residential-Internet-Satisfaction-Study.aspx.

Google Trends, Google, trends.google.com/.

"Internet Providers in Your Area - Find ISPs Near You." *HighSpeedInternet.com*, www.highspeedinternet.com/.

Laurie, et al. "2020 Sparklight (Formerly Cable One) Customer Reviews." *Broadband Now*, 14 Sept. 2018, broadbandnow.com/Sparklight-reviews.

Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

"Tableau."

Appendix – Reproduction of Results

The images below represent the scripts and tools used to complete our analysis. The tools used to complete the analysis are Tableau and Python.

Comparing shapes of dataset and after transformation

```
In [6]: print(demo.shape)
        print(demo_1.shape)

(14911436, 21)
(14853041, 21)
```

```
In [7]: demo_1.reset_index(inplace=True)
        demo_1.head()
...
```

Sorting by PreviousActiveDays in descending and then dropping duplicates(retaining only first occurrence) to have the most up-to-date information on customer by demographics

```
In [8]: demo_1.sort_values(by='PreviousActiveDays',inplace=True,ascending=False)
        demo_1.head()
...

In [9]: demo_1.drop_duplicates(subset='CustomerNodeId',inplace=True)
        demo_1.head()
...

In [10]: demo_1.set_index('CustomerNodeId',inplace=True)

In [11]: demo_1.shape
Out[11]: (808639, 21)
```

Preprocesing the demographics dataset by first sorting using PreviousActiveDays in descending and dropping duplicates.

```
In [10]: orders = pd.read_csv('Orders/Orders.csv')
         orders.head()
...

In [11]: orders.shape
Out[11]: (14911395, 19)

In [14]: orders.sort_values(by='MonthKey',inplace=True,ascending=False)
         orders.drop_duplicates(subset='CustomerNodeId',inplace=True)
         orders.shape
Out[14]: (808632, 19)

In [15]: demo.shape
Out[15]: (808632, 5)

In [16]: orders.to_csv('Orders/Orders_Simple.csv',columns=orders.columns)

In [41]: prod = pd.read_csv('Product/Products.csv')
         prod.head()
...

In [42]: prod.drop(columns='ProductOfferName',inplace=True)

In [43]: prod.sort_values(by='MonthKey',inplace=True,ascending=False)
         prod.drop_duplicates(subset='CustomerNodeId',inplace=True)
         prod.shape
Out[43]: (808632, 16)

In [ ]: prod.to_csv('Product/Products_Simple.csv',colu)
```

Same approach is followed for the other datasets

```
In [14]: from sklearn.model_selection import train_test_split

         X_train,X_test,y_train,y_test = train_test_split(X,y,train_size=0.8,random_state=1)
         print(X_train.shape)

(636309, 29)
```

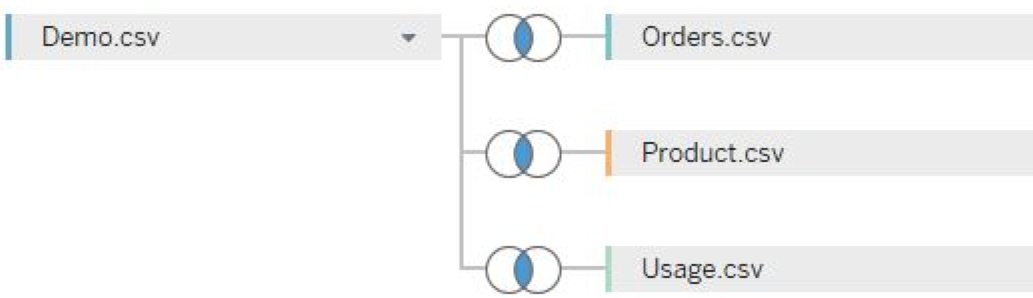
One Hot Encoding

```
In [15]: X_train.head()
```

One hot encoding using get_dummies()

```
In [16]: trainx = pd.get_dummies(X_train)
         testx = pd.get_dummies(X_test)
         testx = testx.reindex(columns = trainx.columns, fill_value=0)
```

Preparing data for training - Splitting dataset into test and validation test and then performing one-hot encoding on the train and test data



Combining the datasets using Tableau prep. All datasets are inner joined using the CustomerNodeId

```
              precision    recall  f1-score   support

           0       0.93      0.84      0.88    133687
           1       0.44      0.68      0.53     25391

    accuracy                           0.81    159078
   macro avg       0.69      0.76      0.71    159078
weighted avg       0.85      0.81      0.83    159078
```

Our vanilla random forest model with a 0.5 threshold that provides a higher accuracy & recall score but lower precision

```
metrics.roc_auc_score(y_test,preds_final.iloc[:,1])
```

```
0.7169357547587719
```

```
metrics.confusion_matrix(y_test,preds_final.iloc[:,1])
```

```
array([[101646,  18130],
       [ 16301,  23001]], dtype=int64)
```

Our final model with a 0.35 threshold that provides us better precision rates