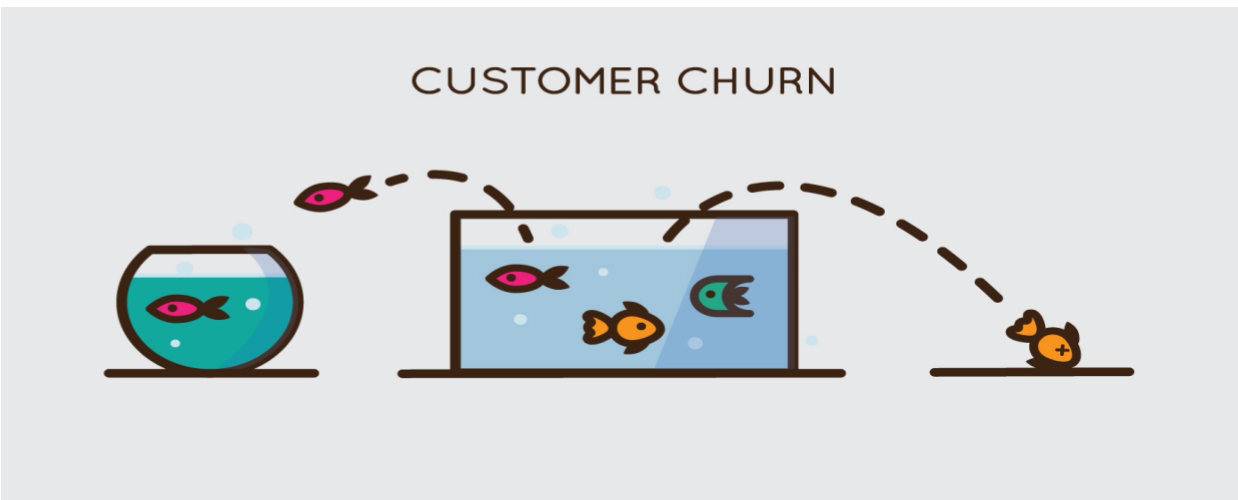
Understanding and Predicting Customer Churn:

A Data Analytics and Machine Learning Project

By

Sigli Mumuni



PROJECT REPORT
(dated April 2021)

Table of Contents

Executive summary

Customer churn refers to a situation in business where customers withdraw from a service or cease to carry on a business relationship with a company. Customer churn can be detrimental to a business's bottom line as it takes away valuable income streams and may contribute to negative word of mouth.

Using data collected from over 2600 customers, I performed in-depth analysis to understand the issue of customer churn as it relates to the company in question, and built machine learning models to predict future customer churn. This helped to provide guidance in answering three main questions:

1. What factors contribute the most to customer churn?
2. Which groups of customers are more likely to churn?
3. What can the business do to prevent customers from churning?

In the proceeding pages, I undertook an analysis of the churn dataset using a wide range of data analytical tools including Python, R and Tableau. First, the dataset was preprocessed to ensure it was clean and consistent, after which an exploratory data analysis was conducted to find relationships among the different attributes and how they can help us further understand the issue of customer churn. Finally, I undertook the task of building several machine learning models using three classification algorithms: Naive Bayes, Decision Tree, and Random Forests. The goal of these classification models was to help the company more reliably predict future customer churn before it occurs in order to take the necessary steps to prevent it from happening.

Based on the significantly better results generated by the Random Forest algorithm and the classification model it generated, I found high reliability and high predictability of the following top-5 factors that contribute most to customer churn:

- Day Charge, which alone contributes close to 23.5% of the explained variance, and;
- International Plan with a percent importance of 21.4%.

These two variables alone contribute close to half of the variance in the model, as well as:

- Evening Charge
- Customer Service Calls
- International Charge

From the results, it can be concluded that customers with a high call volume have on average higher churn rates. Customers with an international plan also have on average higher churn rate than those on a domestic call plan.

Recommendations

We recommend doing a price and rate plan comparison with our competitors. If price sensitivity is a concern, perhaps discount plans can be offered to these high valued customers with higher-than-average call volumes.

In addition, a customer survey can be conducted to find out if there are any specific pain points in the customer experience that the company should be aware of. This may help to better understand why customers with an international plan have a higher churn than this on a domestic call pan.

Also, our call centers should system track the frequency of a specific customer calling our service centers, as a customer service call could be an indication that the customer is more likely to churn within a certain time frame. Pending the customer's complaint, a more competitive offer can be made to this specific group of customers and potentially mitigate the number of customers churning.

Next steps:

1. A plan for stakeholder engagement based on the preliminary results described in this report is highly recommended.
2. Develop an action plan with the business to start tracking the attributes mostly linked to churn.
3. Develop a mitigation or response plan to help manage churn and bring down overall churn rates.
4. Test the effectiveness of the response plan by measuring and comparing actual churn month-over-month.
5. Link the customer churn data set to other sources of information external to the company such as income, education, homeownership, etc. to gain increased knowledge of our current customer base as well as potential future customers.
6. Socialize the concept and interpretation of classification models within the organization as perhaps other managers may be interested in a similar analysis to help grow their business.

A final recommendation is to do a periodic review of the classification model and the resulting attributes that are mostly linked to churn to assure continuous validity of the conclusions reached and churn mitigation strategies in the near future.

Introduction

Clarifying the business problem

The company would like to know in advance which customers have a high risk of churning in the near future. With this knowledge additional mitigating measures may be put in place to reduce future churn rates.

Identifying the stakeholders

Management has initiated the request as they have shown interest to be able to anticipate and reduce future customer churn rates. At the initial stage, stakeholders involved are executive management and the data science team.

Mapping the business problem to a data science problem

The initial task is to characterize customer churn through data analytics methods. As management would like to understand and be able to predict future churn, I have identified this as a classification problem which allows for predictive modelling.

Describing the analytical approach[1]

Classification is the process of identifying and describing different classes of data. The classes are predetermined based on the dataset used. This type of activity is also called supervised learning. Once the model is built, it can be used to classify new data.

The first step is training the model which is accomplished by using a training set of data that has already been classified. Each record in the training data contains an attribute called the class label, which indicates which class the record belongs to.

Some of the important issues with regards to the model and the algorithm that produces the model include:

- the model's ability to predict the correct class of new data;
- the computational cost associated with the algorithm, and;
- the scalability of the algorithm.

The recommendations in this report will concentrate on the model's ability to predict the correct class of new data. Prior to discussing the specific algorithms used in this report, I will discuss and describe the characteristics of the churn dataset that will be used in the classification process.

[1] Source: Fundamentals of database systems. Elmasri and Navathe, 7th Edition 2016.

Data Preparation

Description of churn dataset

The churn dataset has 21 attributes with 3,333 observations including a binary class attribute about churn. The dataset consists of the company's client information such as the customer's phone plan with the company and phone number, as well as call activity measured in number of calls and time measured in minutes during daytime, nighttime, and international calls including voicemail activity.

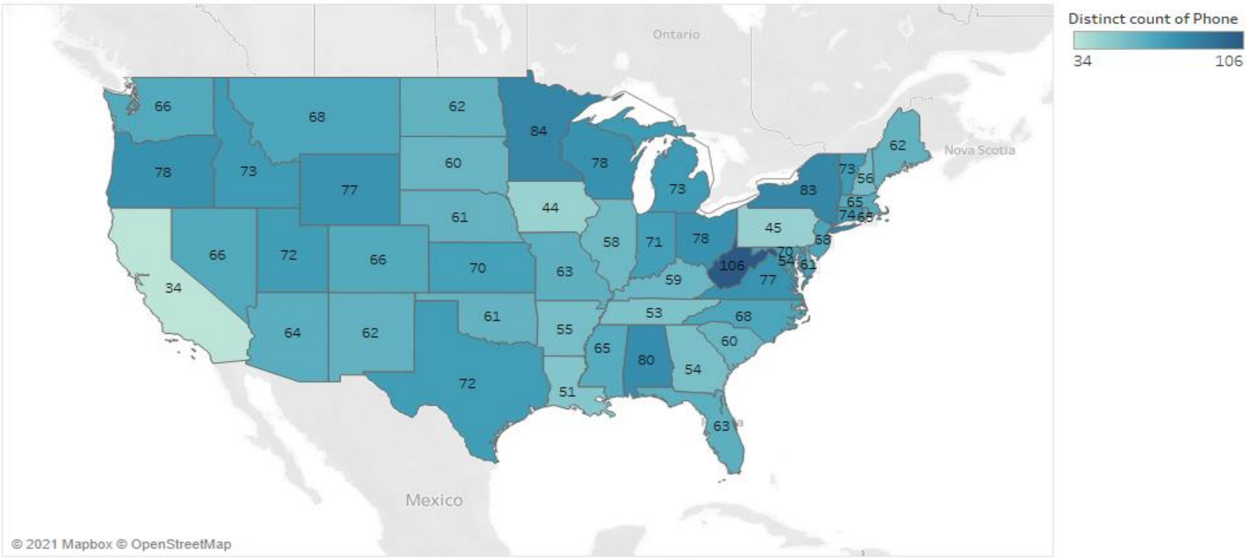
A detailed description of the attributes in the dataset used are given below:

- **State:** Customer's State of residence.
- **Account Length:** Integer number showing the duration of activity for customer account.
- **Area Code:** Area code of customer.
- **Phone Number:** Phone number of customer.
- **Inter Plan:** Binary indicator showing whether the customer has an international calling plan.
- **VoiceMail Plan:** Indicator of voice mail plan.
- **No of Vmail Mesgs:** The number of voicemail messages.
- **Total Day Min:** The number of minutes the customer used the service during day time
- **Total Day Calls:** Discrete attribute indicating the total number of calls during day time.
- **Total Day Charge:** Charges for using the service during day time (continuous data type).
- **Total Evening Min:** The number of minutes the customer used the service during evening time.
- **Total Evening Calls:** The number of calls during evening time.
- **Total Evening Charge:** Charges for using the service during evening time.
- **Total Night Min:** Number of minutes the customer used the service during night time.
- **Total Night Calls:** The number of calls during night time.
- **Total Night Charge:** Charges for using the service during night time.
- **Total Int Min:** Number of minutes the customer used the service to make international calls.
- **Total Int Calls:** The number of international calls.
- **Total Int Charge:** Charges for international calls.
- **No of Calls Customer Service:** The number of calls to customer support service.
- **Churn:** Class attribute with binary values (True for churn and False for not churn).

Customer concentration and churn by State

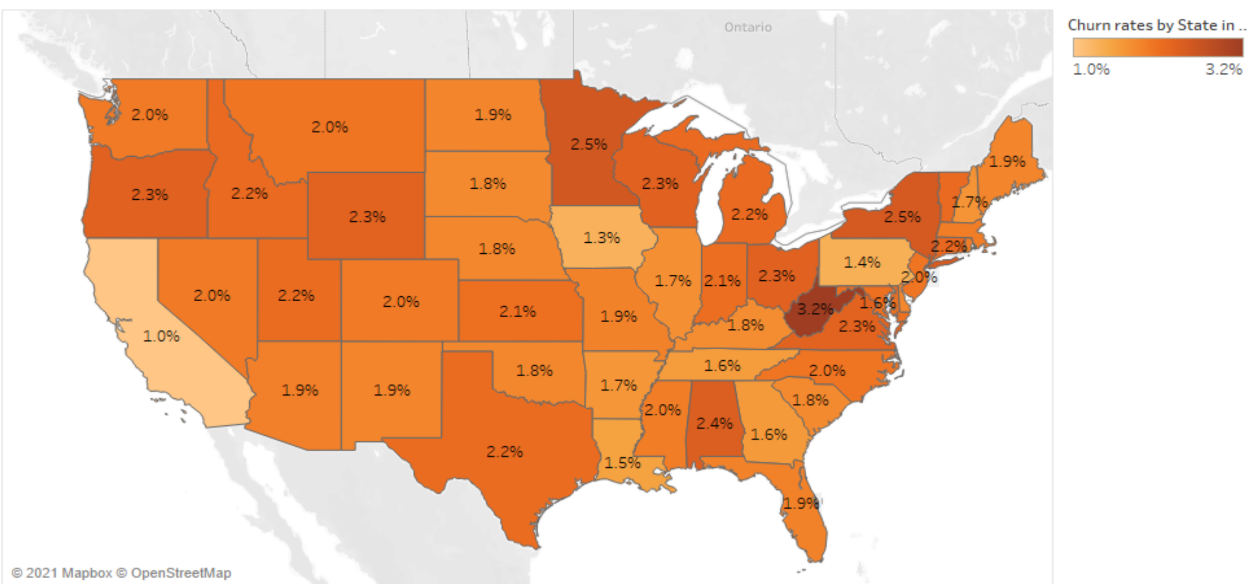
Using the customer phone data and churn data, I present the company's client base and churn rates geographically for all states in the US, as shown in the two graphs below. It appears that there is no significant or high geographical concentration risk within the dataset other than perhaps for the state of West-Virginia.

Customers by US States represented in churn dataset



Map based on Longitude (generated) and Latitude (generated). Color shows distinct count of Phone. Details are shown for State. The data is filtered on Phone, which keeps multiple members.

Churn rates by State in the US



Map based on Longitude (generated) and Latitude (generated). Color shows % of Total Count of Churn?. Details are shown for State.

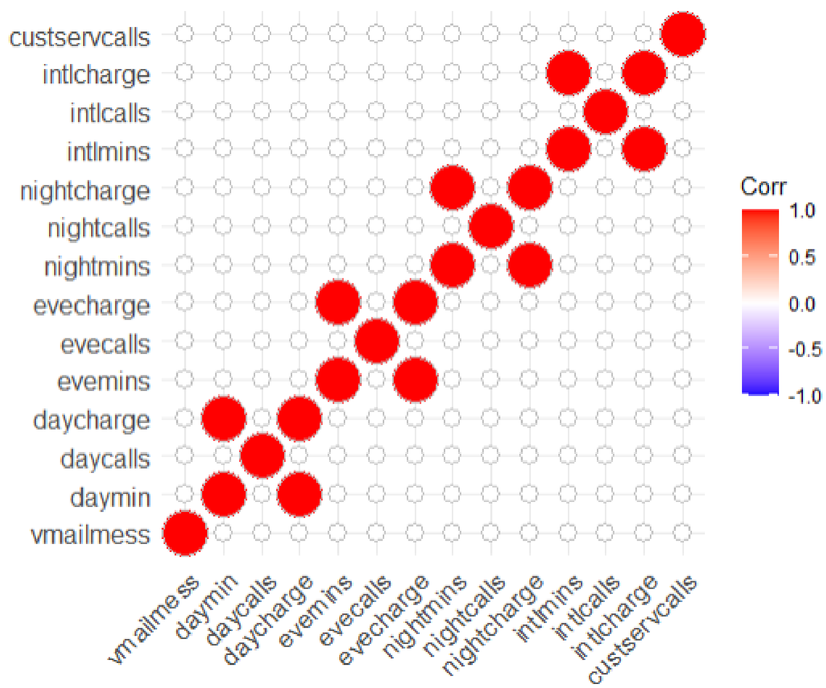
Cleaning the dataset

Correlation

The dataset has been cleaned by removing attributes that show a high correlation (0.999) with the attribute "**Total Charge**" which because of this high correlation are not be necessary for further analysis and have been removed from the dataset. These attributes are:

- **Total Evening Min**
- **Total Night Min**
- **Total Int Min**
- **Total Night Min**

The following graph is a visual of the correlation between the different attributes. As it appears most attributes have low or no correlation between them.



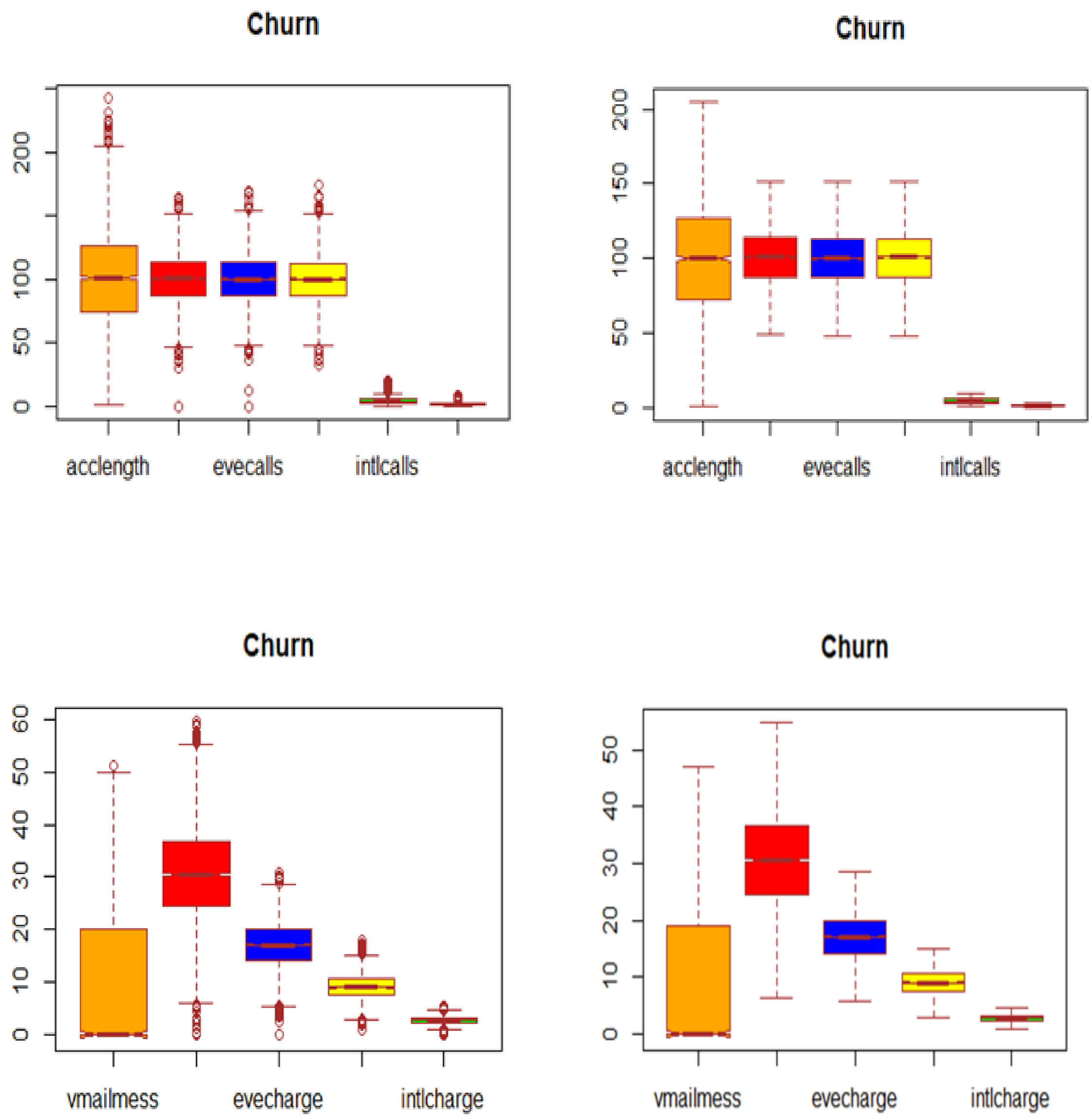
Outliers

The dataset has been further cleaned by identifying and removing so-called outliers.

Outliers are observations that are numerically distant from the rest of the data. When reviewing the two box plots, as shown below, there are data points located outside of the so-called "whiskers" of the boxplot i.e. outside 1.5 times the interquartile range above the upper quantile and below the lower quantile.

I recommend removing the outliers as they could be an indication of incorrectly collected information, but more so I believe that the sample size is not materially impacted by dropping of these questionable outliers or that the interpretation of results is critical to its outcome[2].

The box plot shows the results before (left) and after removing (right) of the outliers.



2 When Should You Delete Outliers from a Data Set? - Atlan | Humans of Data

Characteristics of dataset before and after cleaning

With the use of R, the dataset has been summarized for each attribute in the table below, before and after cleaning of the dataset, for its more numerical and statistical characteristics, as follows:

- Minimum value
- 1st quantile (threshold of first 25% of observations)
- Median
- Mean
- 3rd quantile (threshold of 75% of observations)
- Maximum value
- Standard deviation

	acclength		vmailmess		daycalls		daycharge		evecalls		evecharge		nightcalls		nightcharge		intlcalls		intlcharge		custservcalls	
	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned	original	cleaned
Min	1	1	0	0	0	49	0	6	0	48	0	5.6	33	48	1.04	3	0	1	0	1	0	0
1st Qu.	74	73	0	0	87	87	24	24	87	87	14.2	14	87	87	7.52	7.5	3	3	2.3	2	1	1
Median	101	100	0	0	101	101	31	31	100	100	17.1	17	100	101	9.05	9.1	4	4	2.8	3	1	1
Mean	101	100	8.1	8.1	100	101	31	31	100	100	17.1	17	100	100	9.04	9.1	4.5	4.3	2.8	3	1.6	1.3
3rd Qu.	127	127	20	19	114	114	37	37	114	113	20	20	113	113	10.6	11	6	6	3.3	3	2	2
Max.	243	205	51	47	165	152	60	55	170	152	30.9	29	175	152	17.8	15	20	10	5.4	5	9	3
Sd.	40	39	14	14	20	19	9.3	9	20	19	4.31	4.2	20	19	2.28	2.2	2.5	2.1	0.8	1	1.3	1

When comparing the before cleaning and after cleaning datasets for each attribute, it is noticeable that by removing the outliers, the minimum values are seeing an increase and the maximum values a decrease, bringing both minimum and maximum closer to the mean, and more so reducing the standard deviation as one would expect when removing outliers.

Other observations

The dataset does not have any missing values as indicated by "NA".

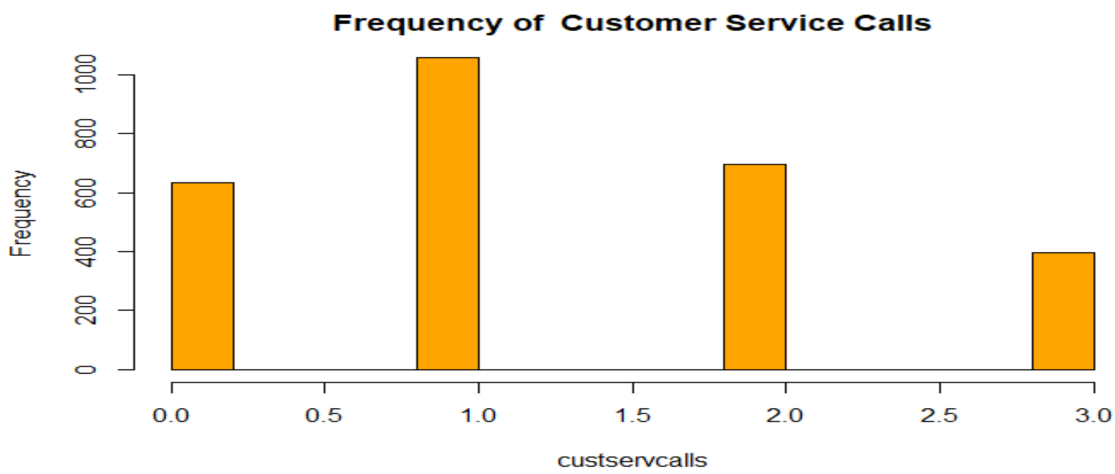
For ease of reference, all "True" and "False" values within the churn attribute have been replaced by "yes" and "no".

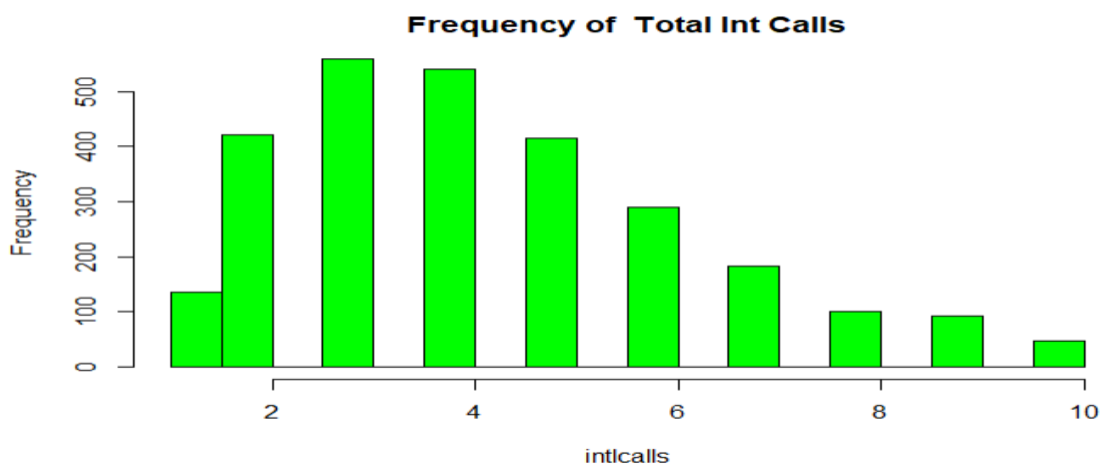
Before cleaning	After cleaning
churn -> "yes" 483" observations	churn cleaned-> "yes" 302 observations
churn -> "no" 2,850 observations	churn cleaned -> "no" 2,484 observations

As a result of the data cleaning process, the dataset, which will be used for further presentation, has now been reduced to 16 attributes consisting of a total of 2,786 observations.

Distribution of numerical attributes

When summarizing and characterizing the data set, it can be concluded that most attributes have a normal distribution and as a result the dataset is centered around the mean. However, for **No of Calls Customer Service** and **Total Int Calls** the dataset is slightly skewed to the right, as per the graphs presented below.





Skewness to the right is an important observation as it shows that the mean may not be the best indicator to reflect the "centerness" of the dataset and as a result the median or mode is perhaps a better indicator.

Although there are various ways to create a better fitting distribution[3], I have concluded that the skewness of both attributes is not significant to justify a correction as the values for both mean, median, and mode lie within each other's proximity:

- No of Calls Customer Service: mean 1.3, median 1, mode 1.

[3] 1.3.3.14.6. Histogram Interpretation: Skewed (Non-Normal) Right (nist.gov)

- Total Int Calls: mean 4.3, median 4, mode 3.

Imbalanced class distribution

From initial observation, it has been established that the dataset has an imbalanced class distribution by a ratio of almost 1:6, with the majority class being the "no" class[4].

Working with imbalanced datasets can be problematic if there are too few examples of the minority class to incorporate into the decision boundary. Subsequently, the model becomes extremely good at predicting the majority class but does not do so well with the minority class.[5]

To improve the models' performance. I proceeded to balance the dataset, so it has equal numbers of both classes. To do this, I implemented the so-called "oversampling" technique. As the name suggests, this is achieved by oversampling the minority class in the training dataset. Examples are drawn from the minority class and duplicated to match their occurrence with the majority class. This only serves to balance the dataset without amending or including any additional information.[6]

While many approaches exist to accomplish a more balanced dataset, I chose the most widely used algorithm, called the Synthetic Minority Oversampling Technique, or SMOTE algorithm.

Summary

As part of the data preparation process, I have scrubbed the dataset as follows:

- Removed attributes with a strong correlation to the attribute "Total Day Charge".
- Reviewed and removed any outliers.
- Established that there are no missing values in the dataset.
- Reviewed the dataset for its distribution and determined that two attributes are slightly, but not significantly skewed to the right.

After completion of the data cleaning process, I have established two datasets which will be used to run the algorithms. They are as follows:

- Filtered and imbalanced dataset
- Filtered and balanced dataset

[4] Out of a total of 2,699 records, a total of 2,307 customers are classified as "no" while the remaining 392 are classified as "yes".

[5] Brownlee, J (2020) SMOTE for Imbalanced Classification with Python (Last accessed at: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/ on 12th April, 2021)

[6]Chawla, N et al (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002) 321–357

Classification algorithms

For the purpose of comparison, I have decided on the implementation of three classification algorithms to predict future customer churn. They are as follows:

- Decision Tree;
- Naïve Bayes, and;
- Random Forest.

I will briefly describe all three algorithms and the interpretation of their outputs in the following section.

Decision Tree

The decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor is called the root node. Decision trees can handle both categorical and numerical data. [7]

Naïve Bayes

Naive Bayes is a machine learning algorithm that is mostly used to solve classification problems. It is based on Bayes Theorem and is one of the simplest, yet powerful algorithms in use with applications in many industries. Naive Bayes is a classification technique based on an assumption of independence between predictors which is known as Bayes' theorem. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. One of the biggest advantages of Naive Bayes is that it requires relatively small amounts of data to train the model. In this report, I implemented the Gaussian Naïve Bayes algorithm for classification. [8]

Random Forests

The Random forest algorithm is an ensemble method used mainly for classification and regression. Unlike decision trees, random forests grow not one, but a multitude of decision trees. Each tree gives a classification, and "votes" for that class, after which the classification with the most votes is selected from all the trees within the "forest".[9]

I made the decision to include the random forest algorithm as it is well established that growing an ensemble of trees (as opposed to a single tree) and allowing them to vote for the most popular class will significantly improve classification accuracy. But beyond that, unlike decision trees, random forests do not overfit, are able to balance error in classification caused by imbalanced data sets, give estimates of what variables are important, and several other advantages.[10]

---

[7] https://www.saedsayad.com/decision_tree.htm

[8] Naive Bayes Explained: Function, Advantages & Disadvantages, Applications in 2021 | upGrad blog.
What Is Naive Bayes?. Before we build a classifier, let's… | by Navjot Singh | The Startup | Medium.

[9] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). The Elements of Statistical Learning (2nd ed.). Springer.

[10] Leo Breiman (2001) Random Forests. Machine Learning, 45, 5–32, 2001

Approach

To compare the outputs of machine learning models that have different features, I used cross validation as a way to analyze how well the supervised learning models are performing on a dataset that was not part of the data utilized to train the model i.e. how well the model generalizes[11].

First, I created a baseline model by creating a training set using the cleaned, imbalanced dataset and evaluated its performance using selected metrics, which I will discuss in more detail in the section below, on the test set.

Secondly, I trained the same classification algorithm on the cleaned, balanced dataset and evaluated its performance on the test set using the same selected metrics.

Finally, I compared the performance of the different models by using selected performance metrics.

Determine performance measures

When measuring how well the machine learning algorithm is performing on the test set, I will apply the evaluation metric of the so-called 2x2 confusion matrix, also known as error matrix[12].

The confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in an actual class. The name stems from the fact that it makes it easy to see whether the system is confusing two classes i.e. commonly mislabeling one as another.

As follows:

		Predicted Class	
		+	**-**
Actual Class	**+**	f_{++}	f_{+-}
	-	f_{-+}	f_{--}

f_{++} : measure all "True" positives (TP) which indicates that the outcome of the model or predicted value matches that of the actual value. Vice versa f_{--} (TN): measures all "True" negatives which also indicates how well the model is performing. On the other hand, f_{-+} (FP) and f_{+-} (FN) measures a mismatch between the actual value and the predicted value by the model.

Based on these measurements, the confusion matrix constructs three different evaluation metrics as follows:

[11] Padmanabhan & Jenkins, 2014.

[12] Confusion matrix - Wikipedia

- Accuracy is measured by: $\frac{f_{++}+f_{--}}{T}$
- Recall is measured by: $\frac{f_{++}}{f_{++}+f_{\pm}}$
- Precision is measured by: $\frac{f_{++}}{f_{++}+f_{+-}}$

The evaluation method of the confusion matrix allows for a more detailed analysis than mere proportion of correct classifications, measured by accuracy. As discussed, Accuracy only will yield misleading results if the data set is imbalanced; that is, when the numbers of observations in different classes vary greatly.

The F-score is another evaluation metric which represents the harmonic mean of the precision (p) and recall (r) values. That is:

$$F = \frac{2pr}{p+r}$$

High precision is achieved almost always at the expense of recall and vice versa. It is a matter of the application's context whether to tune the system for high precision or high recall. F-score is typically used as a single measure that combines precision and recall comparing different result sets.

One of the properties of harmonic mean is that the harmonic mean of two numbers tends to be closer to the smaller of the two. Thus, F is automatically biased toward the smaller of the precision and recall values. Therefore, for a high F-score, both precision and recall must be high.

Later in this report, I will be using the above evaluation metrics to make comparisons between the classification models.

Determining the right strategy for the data split

To determine the baseline model, I have applied a simple train-test set to split the dataset into a training set and a test set. I chose this method as the dataset is deemed to be sufficiently large enough such that samples from the original dataset can be split randomly into subsets without a negative impact on the representativeness of the original dataset. Using Sklearn, I applied a split percentage of 80% and 20% between the training set and test set respectively.

Training the models

I built a total of 6 models, using two datasets (balanced and imbalanced) and 3 algorithms. As our baseline models for all three algorithms, I used the cleaned dataset with imbalanced classes. The reason behind my choice of dataset for the baseline model is to more easily compare model performance between that of the imbalanced dataset and the balanced dataset. I decided to retain all features selected during the cleaning process (14 in total) since all features were deemed to contribute substantially to the explanation of the variance in the models as described in the section below on feature importance.

We run the three selected algorithms, first on the imbalanced dataset, and then on the balanced dataset, bringing the total number of models to six. For each of the six models, I performed an iterative process where I utilized hyper parameter tuning between the different iterations to arrive at the optimal or best performing model. With regards to the tree models for example, optimization of the decision tree classifier was undertaken by pre-pruning techniques such as tweaking the maximum depth as a control variable for the expansion of nodes. The “splitter” parameter also enabled us to alternate between “best” which chooses the best split and “random” which selects the best random split. Aside from pre-pruning parameters, “criterion”, which is the function that measures the quality of a split, was used to alternate between “gini” for Gini impurity and “entropy” for information gain. Unique to the random forest algorithm, is the “n-estimators” parameter, which allowed for the specification of the number of trees in the forest.

At the conclusion of the model building process and having derived the best performing model for each of the 6 variations, I proceeded to run a comparison using selected performance metrics as explained in the next section.

Comparing results of algorithms

In this section, I will compare the results using the evaluation metrics described earlier.

After running the algorithms, I collected all the measurements for the imbalanced (baseline model) and balanced dataset including true positives, true negatives, as well as false positives and false negatives and calculated the evaluation measures for accuracy, recall, and precision, as presented in the tables below.

Decision Tree - imbalanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	57	33
- (/no)	30	716
Observations	836	
	Ouput	
Accuracy	0.92	
Recall	0.63	
Precision	0.66	

Random Forest - imbalanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	51	2
- (/no)	36	747
Observations	836	
	Ouput	
Accuracy	0.95	
Recall	0.96	
Precision	0.59	

Naïve Bayes - imbalanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	33	45
- (/no)	54	704
Observations	836	
	Ouput	
Accuracy	0.88	
Recall	0.42	
Precision	0.38	

Decision Tree - balanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	693	59
- (/no)	55	684
Observations	1491	
	Ouput	
Accuracy	0.92	
Recall	0.92	
Precision	0.93	

Random Forest - balanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	707	27
- (/no)	41	716
Observations	1491	
	Ouput	
Accuracy	0.95	
Recall	0.96	
Precision	0.95	

Naïve Bayes - balanced		
	Predicted class	
Actual class	+ (/yes)	- (/no)
+ (/yes)	655	195
- (/no)	93	548
Observations	1491	
	Ouput	
Accuracy	0.81	
Recall	0.77	
Precision	0.88	

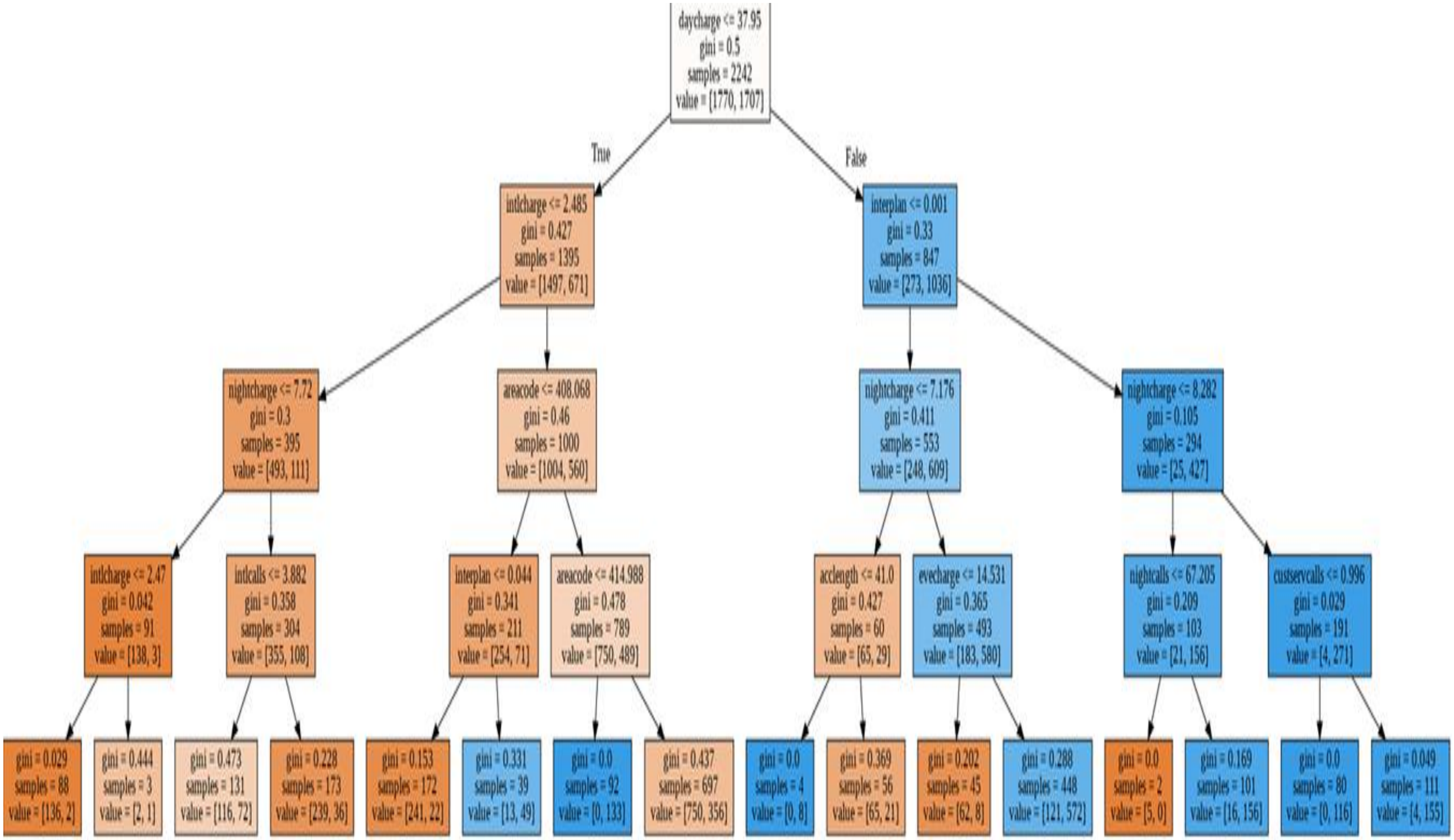
As per table below, looking at the results of our evaluation metrics for Accuracy, Recall, and Precision, we can conclude that the balanced dataset shows a significant improvement over the use of the imbalanced dataset especially for the decision tree and Naïve Bayes algorithm. However, we can also conclude that the Random Forest algorithm shows consistently high results for all three-evaluation metrics of the confusion matrix as it relates to the balanced datasets which indicates the high predictive capability of the Random Forest model.

Classification Models	Imbalanced dataset				Balanced dataset			
	Accuracy	Recall	Precision	F-score	Accuracy	Recall	Precision	F-score
Decision Tree	0.92	0.63	0.66	0.64	0.92	0.92	0.93	0.92
Naïve Bayes	0.88	0.42	0.38	0.40	0.81	0.92	0.88	0.90
Random Forest	0.95	0.96	0.59	0.73	**0.95**	**0.96**	**0.95**	**0.95**
Based on 836 observations					***Based on 1,491 observations***			

The same for the Random Forest classification model can be concluded for the evaluation metric of the F-score, as its high F-score indicates that both precision and recall must be high, as it also appears from the results in the table above.

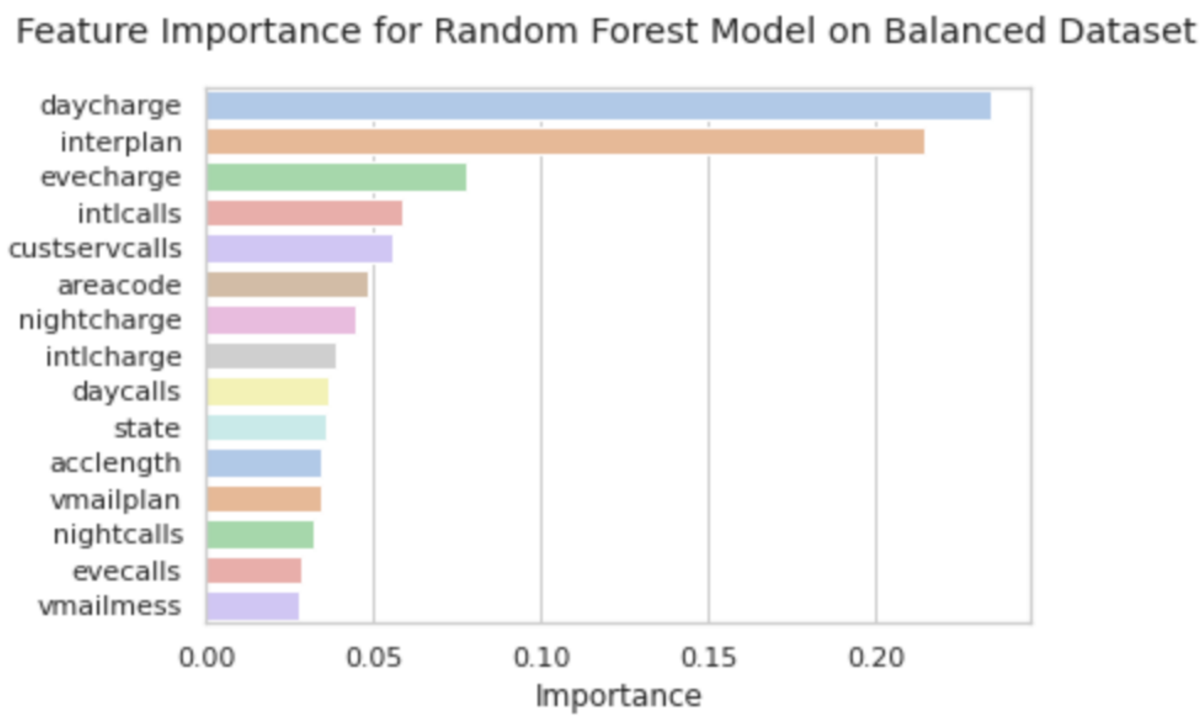
One of the decision trees generated by the Random Forest Classification is presented on the next page.

Random Forest - Classification Tree



Feature importance

The bar chart below shows all the features used in building our random forest model with the balanced dataset, ranked according to importance. This ranking is based on a numeric representation of how much of the explained variance in the model is contributed by the feature in question. These computed values describe how important the features are for the machine learning model and can shed more light on how important those features are in the overall approximation of the relationship between the predictor variables and target variable. The scores on the x-axis of the bar chart represent relative importance, meaning the percent importance of each feature.



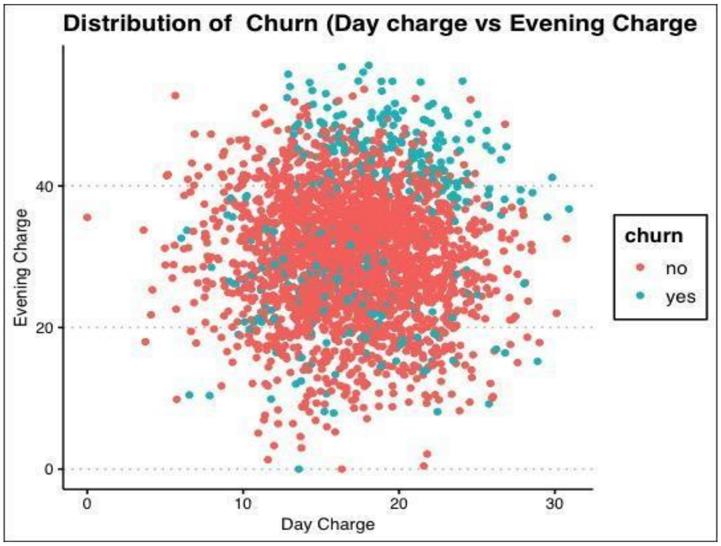
In the bar chart, it can be seen that the most important feature in our decision tree to explain and predict churn is Day Charge, which alone contributes close to 23.5% of the explained variance, followed by International Plan with a percent importance of 21.4%. These two variables alone contribute close to half of the variance in the model, after which Evening Charge, Customer Service Calls and International Charge make up the rest of the top 5 performing features.

In the next section, I will conduct additional analysis on these five attributes in an attempt to shed more light on how they relate to the target variable "churn".

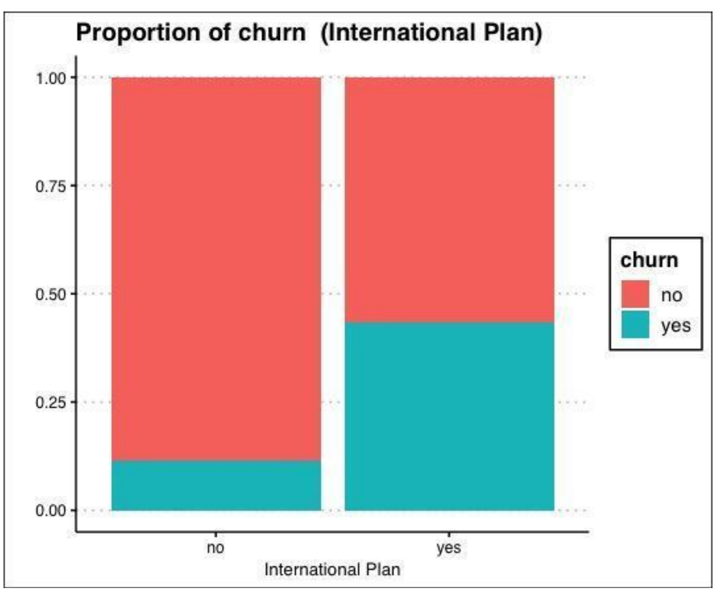
Factors that contribute to churn

One of the key objectives of this project is to identify the most important factors that contribute to customer churn. This is key to revealing which groups of customers are likely to churn, so as to implement the necessary interventions to prevent this from happening.

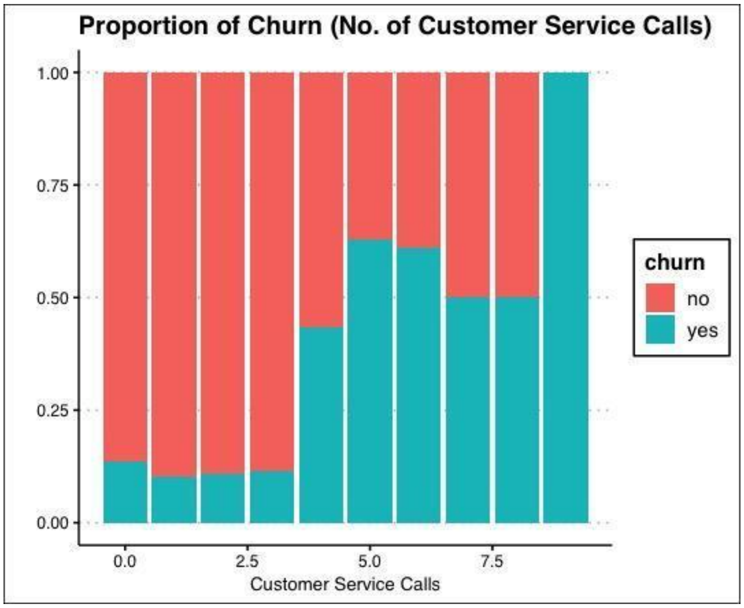
Having uncovered the most important attributes in the best performing classification model, the next goal is to attempt to draw linkages between these attributes and the target variable ‘churn’.



Two important factors that contribute to churn are Day Charge and Evening Charge, which in turn are highly correlated with Day Calls and Evening Calls. Looking at the scatter plot above, it can be seen clearly how these attributes relate with the target variable. Churning customers have a higher Day and Evening charge on average than non-churning customers. This could perhaps be an indication that our rate plans become less competitive, or our customers become more price sensitive as their overall phone charges increase due to increased call volume or call duration.



Customers on an international plan have a higher churn rate on average than customers who aren't. It is not immediately clear why this is the case, as it is an issue that will require additional investigation. Perhaps surveys can be conducted to find out if there are any pain points in the customer experience that the company should be aware of. Nevertheless, we can conclude from the bar charts above that customers on an international plan are more likely to churn compared to customers who aren't.



Quite expectedly, the number of customer service calls is highly correlated with the likelihood of a customer churning. Frequent customer service calls is indicative of frustration with a product or service, and is usually a precursor to churn. This is confirmed from the above chart. Here, we see a sudden spike in churn rates with an increase in the number of customer service calls.

Any strategy that hopes to tackle the issue of customer churn as it relates to the company, will have to include the four main drivers of churn we have outlined above. In the next section, I propose concrete steps the company can take to improve customer satisfaction and reduce the rate of churn.

Recommendations

The telecom industry is highly saturated, with low customer growth rates. Any effort to retain and increase valuable market share must focus heavily on customer retention. Based on our findings, we propose the following strategies to control the rate of churn in the company:

1. As customers with a high call volume and high call duration have on average higher churn rates, we recommend doing a price and rate plan comparison with our competitors. If price sensitivity is a concern, perhaps discount plans can be offered to these high valued customers with high call volumes. Also, instituting a customer reward/loyalty program that targets frequent users with special discounts and rewards will contribute towards creating more value for customers.
2. In addition, a customer survey can be conducted to find out if there are any specific pain points in the customer experience that the company should be aware of. This may help to better understand why customers with an international plan have a higher churn than those on a domestic call pan.
3. Also, the company should system track the frequency of a specific customer calling, as a customer service call could be an indication that the customer is more likely to churn within a certain time frame. These customers can be specifically targeted and given focused attention to address issues of concern to them. Further to that, the company can probe these clients further to understand what issues may be causing pain points in the customer experience.

Next steps:

1. A plan for stakeholder engagement based on the preliminary results described in this report is highly recommended.
2. Develop an action plan with the business to start tracking the attributes mostly linked to churn.
3. Develop a mitigation or response plan to help manage churn and bring down overall churn rates.
4. Test the effectiveness of the response plan by measuring and comparing actual churn month-over-month.

5. Link the customer churn data set to other sources of information external to the company such as income, education, homeownership, etc. to gain increased knowledge of our current customer base as well as potential future customers.
6. Socialize the concept and interpretation of classification models within the organization as perhaps other managers may be interested in a similar analysis to help grow their business.

A final recommendation is to do a periodic review of the classification model and the resulting attributes that are mostly linked to churn to assure validity of the conclusions drawn in order to continuously mitigate churn in the future.

End