Telco Customer Churn: R Classification

Business Context & Challenge

Voluntary customer churn is a major concern across a multitude of industries. On average, it costs 5 to 25 times more to acquire a new customer than to retain an existing one so understanding the threshold at which a company can discount its products in order to keep a customer is of extreme importance to maintaining revenue growth. Not only is the churn metric critical to marketing teams, it can also be used as one of the key indicators to interpret a company's overall health. Since retaining customers is so integral, analyzing customer behavior using machine learning techniques in order to detect patterns leading to churn can be incredibly powerful. When a company knows in advance which customers are likely to leave, it can offer incentives and apply additional marketing techniques to convince the customer to stay and retain revenues that would otherwise have been lost. With such tremendous potential for preventing revenue losses , churn prediction models have become one of the most widely used machine learning applications across many industries.

This project will use the Telco customer data set hosted on '**Kaggle.com**'. It contains 7,043 unique customer records for a telecom company that provides details about customer accounts, types of services and other demographic information. Additional features were added to derive insights which are discussed in the feature engineering component of the project.

A series of classification models were developed to predict customer churn with machine learning algorithms such as Random Forest, RPART, XGBoost and Neural network. A comparative analysis of the results was conducted upon which **Stepwise Regression** was determined to be the best churn detection model.

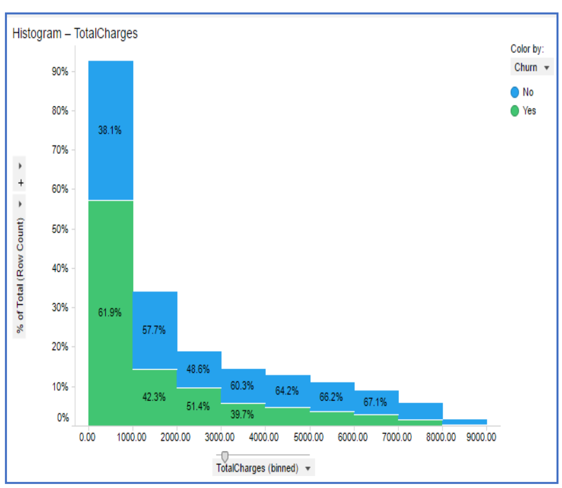
Exploratory Data Analysis

Data Exploration & Visualization

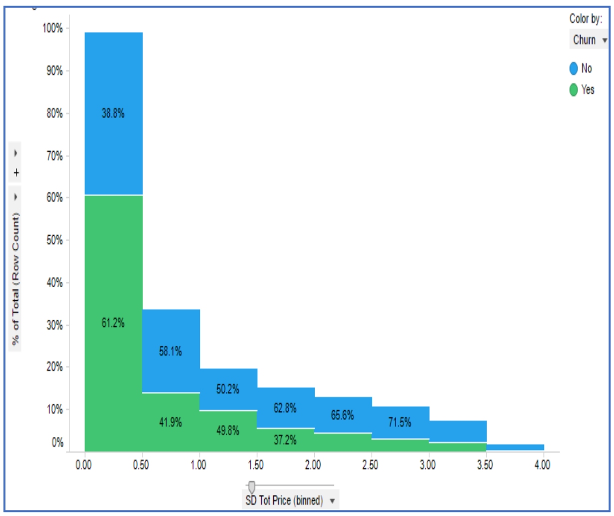
Before any kind of model development, it is imperative to explore the data and draw meaningful explanations on customer behavior. Hence, an extensive amount of exploratory data analysis (EDA) was performed on different features, such as total charges, tenure, monthly charges, contract type, payment methods etc. A series of visualizations were constructed to provide insights on the data set which assist with building hypotheses around what the effect of a feature is likely to be in the modeling process.

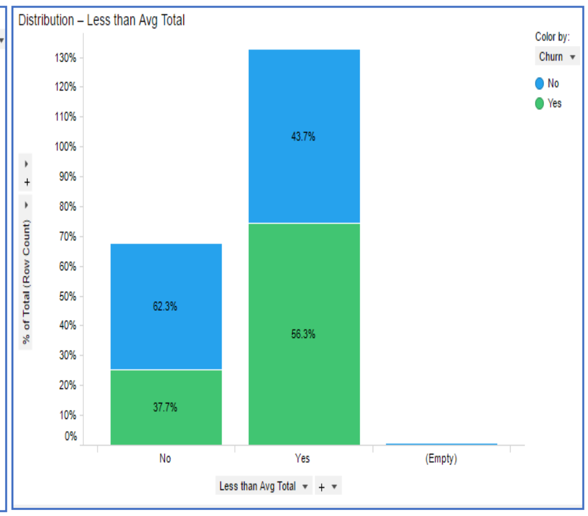
Total Charges

The Total Charges feature provides information on the total charges incurred by the customer. The histogram below shows that the Total Charges feature is positively skewed where customers on the lower end of the total charge spectrum seem to churn more frequently. For example, 61.9% of customers in the 0-1001-dollar total charge bucket end up churning compared with 39.7% for customers in the 3,000 to 4,000-dollar bucket.



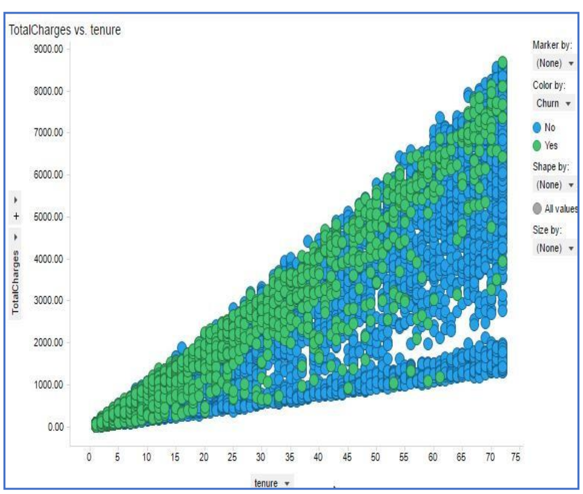
This is also evident in the visualization below which shows the proportion of customers that churned, against the price a customer paid over average total price (binned). Both visuals reinforce the earlier observation that customers with low total charges tend to churn more frequently: the higher the customers' total charges compared with the average total charge, the less likely they will churn. The bar chart highlights this as well by comparing the proportion of customers that churned when their Total Charges were more than the Average Total Charges.

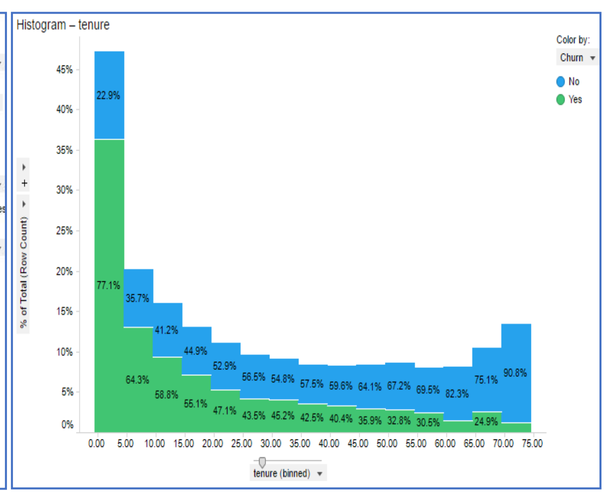




Tenure

The tenure feature that provide customer's tenure information (months) is also positively skewed, which supports the observations made with the Total Charges feature.



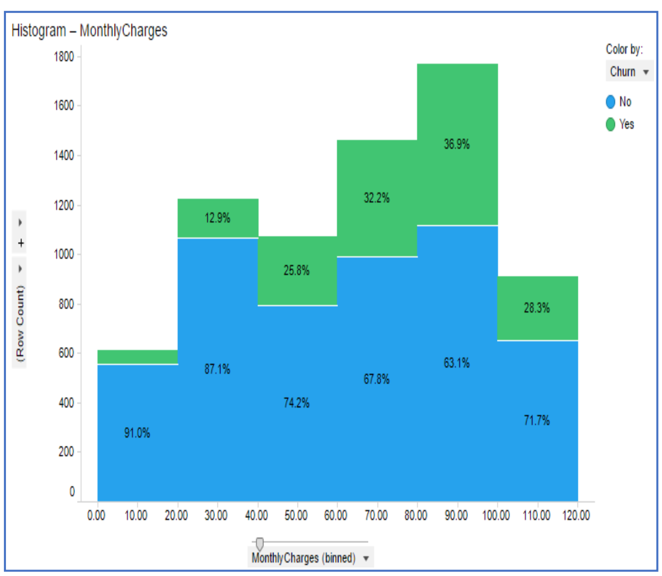


The last set of visuals in the previous page indicate that longer-tenured clients tend to attrite significantly less. As an example, 77.1% of the clients tenured less than or equal to 6 months end up churning while that number is less than 10% for customers tenured for 70 months or more.

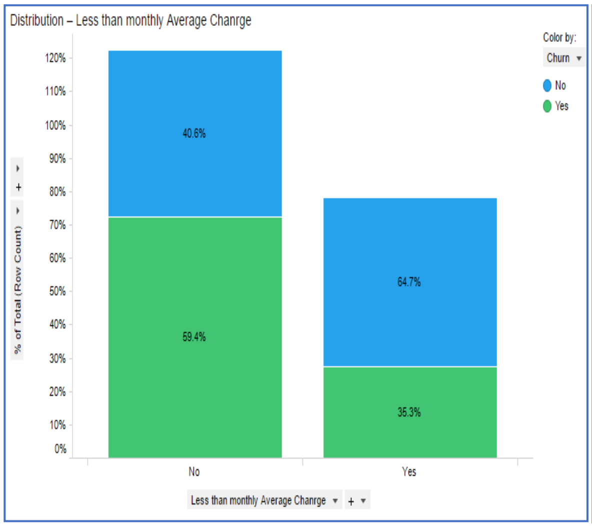
The positive relationship between tenure and total charges can be explained by the fact that these two variables are highly (positively) correlated. The correlation coefficient between them is 0.83.

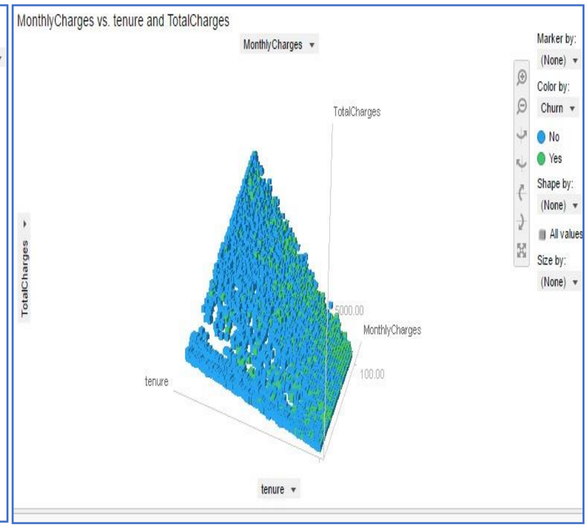
Monthly Charges

The story with Monthly Charges is quite different from Total Charges. With higher monthly charges, customers tend to exhibit a rate of attrition. In the diagram below, as we move from the 0-20 bin to the 80-100 bin, the percentage of customers churning increases from 9% to approximately 37%.



This behavior is summarized in the bar chart (below): if the monthly bill of a customer is below the average monthly bill in our data set, they are less likely to churn - only 35.3% of customers in this bucket churn - compared with the much higher 59.4% when the monthly bill is higher than the higher monthly bill. This potentially highlights the fact that telecom customers are price sensitive and are more prone to attrition if they feel that they can get a lower priced plan elsewhere.

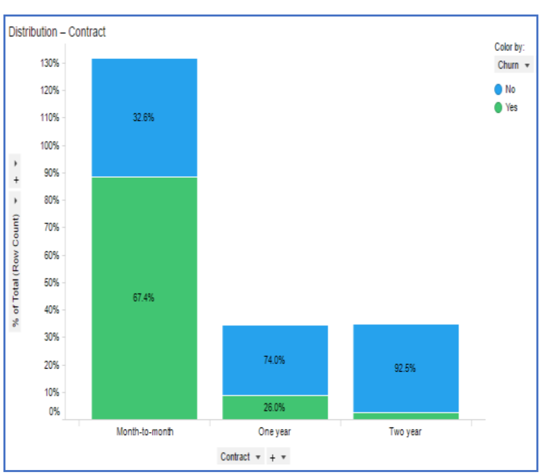


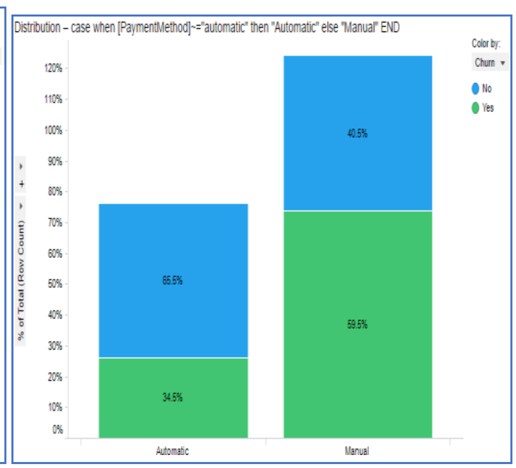


The contrasting relationship can be explained by the 3-d scatterplot above, which shows the relationships between Total Charges, Monthly Charges and Tenure. If total charges are higher due to tenure, then customers churn less often, while if the increase in total charges is more dependent on increased monthly charges, then customers churn more often. This was the basis of one of our most important engineered feature: **Monthly charges as a Percentage of Total Charges**.

Contract and Payment Method

The contract feature appears to be a highly relevant factor which contributes to customer attrition behavior.





This is highlighted in the bar chart above where Contracted customers churn less frequently than customers on a Month-to-month contract (26% for a 1-year contract versus 67.4% for month to month). On the other hand, less than 8% of customers on a 2-year contract ended up churning. This highlights the importance of getting customers to sign contracts. The Payment Method feature offers additional insight into customer churn behavior. The bar chart on the top right shows that automatic payment method customers churn at a rate of 34.5% compared with the manual bucket whose churn rate is 59.5.

Device Protection, Online Backup and Online Security

The pie chart visualizations in the appendix highlight the fact that a smaller fraction of customers churn with additional features activated, for example, it can be seen that of the customers that churned, only 15.8% had Online Security compared to 33.3%, where customers did not churn. Lastly, customers who churnedhad a much smaller proportion of Device Protection compared with those who did not churn (29.2% vs 36.3%). It seems as if, having an enhanced feature set with your phone plan (upgraded plans) reduce the chance of churn.

Streaming Movies, Streaming TV and Multiple Lines

The feature to churn relationship (mentioned above) does not hold in the case of multiple services.

It seems as customers that have signed up for multiple services like Movie Streaming, TV Streaming and Multiple Lines are more prone to churn. As per the visualizations that are shown in the appendix (same heading as above), the churning customer group has a higher proportion of users with these services activated (compared with the group of customers that did not churn).

Customer Demographics

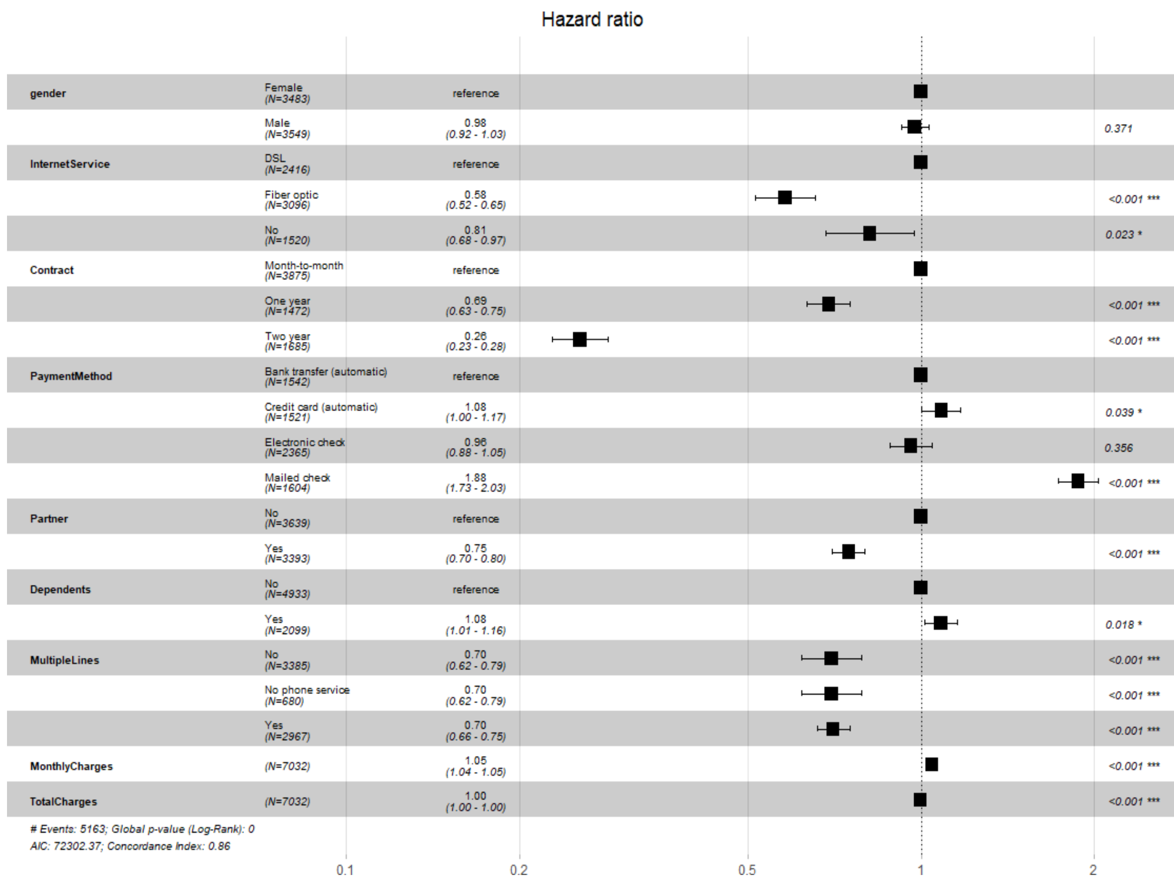
Lastly, from a demographic standpoint, we see that customers with no dependents and no partners are more likely to churn. This might be because of complexity associated with moving an entire family's accounts to a new service provider. This highlights the potential that adding family plans provides towards having reduced customer churn.

Survival Analysis

Next, a preliminary survival analysis was conducted to understand the expected timeline on when the customers could churn. In this analysis, the tenure feature that holds the timeline information for churn is also the response feature. As a next step, we created the binary flag for censored. When churn is no, we categorized them as 1 else 0 indicating that 1 is surviving, 0 is not. To avoid leakage, we drop the original churn feature from the data. This indicates 1 as the customers being surviving, 0s are the ones who are not. Below are the stratified plots for contract, multiple lines and internet services.

The Kaplan-Meier survival curves in the appendix provide us with the proportion of customers that survived over time. Each curve represents the survival probability with that feature activated, holding everything else constant. As an example, the last visual gives us the survival probability of customers based on the 3 levels of the internet service feature. We can see in the figure, that having internet service is beneficial for the firm, since a higher proportion of customers survive at each time interval. From among the internet options, having Fiber Optic is preferable (green curve is higher, thus showing a higher survival proportion). Similar observations can be made by looking at the other two visuals. When comparing the 3 Kaplan-Meier survival curve visualizations, we can see that contract type seems to be the best predictor of survival, as it has the flattest slope (especially in earlier time periods).

We use the Cox proportional hazard model to investigate the association between churning and some of the predictor features in the data set. From the results, we use the forest plot that shows the hazard ratios that are derived from the model for the covariates.



Each of the plotted features can only be compared against the reference row for that feature set. The reference row is the plotted on the vertical line marked by ‘1’ and represents ‘normal’ risk. If a feature plots on the left of this line, then it has a lower risk than the reference point (holding everything else constant). On the other hand, if it plots to the right, that means that it has more risk than normal. The magnitude of this risk is determined by how far the point plots from the reference

('1') line. The risk in question here is of a customer churning. To illustrate, let's look at the contract type feature. Contracted customers are less risky than month-to-month customers and customers on a 2-year contract are the least at risk of churning, as its plotted point is close to the '0.2' line. This means that their churn risk is only 20% of the month-to-month customers (holding everything else constant).

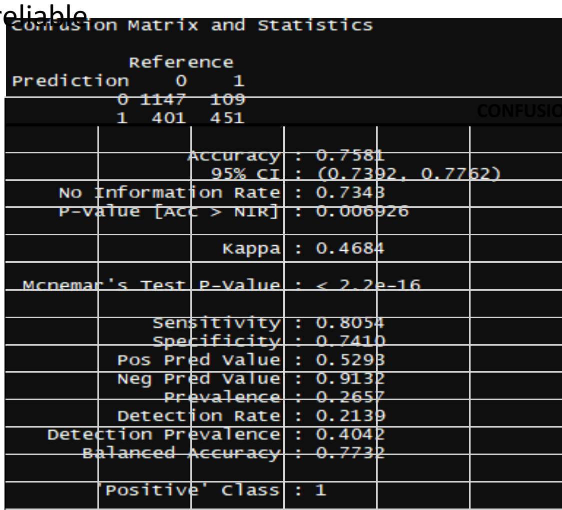
Model Development

After pre-processing the data for model development, it contained 27 features and 7,032 instances. Since the data is slightly imbalanced, stratified random sampling was used for the training and testing (holdout) split. Approximately 70% of the data is used to train the models and 30% of the data is used for testing. Predictions are made on the testing data set which is then used to validate the results. Since predicting customer behaviour is a binary feature, classification models were used to train and predict. The resulting confusion matrix was used to calibrate the output of our models and the outcomes of the prediction (true positive, true negative, false positive, false negative). In the data set, 26% of the customer are churning. Hence, the prediction threshold was set to 26%, i.e. anything less than 26% will be classified as 0, meaning the customer is not churning. Anything above 26% will be classified as 1. The models that were utilized include: Stepwise Logistic Regression, CTREE Classification, RPART Classification, Random Forest, XGBoost and Neural Network.

Stepwise Logistic Regression

As logistic regression doesn't deal well with too many inputs, it is often a challenge to select the optimal number of features to use in logistic regression. This is when step wise regression is useful. There are three algorithms, forward, backward and both. The forward method keeps adding and dropping features one by one, until it gets the best significance. The backward model includes all the features in the model first and then starts dropping depending on their significance. The third method is a mix of both.. Stepwise regression selected some features that have overlap with the previous logistic regression including senior citizen, dependents, multiple lines, device protection, online streaming, tenure, monthly charges as a percentage of total charges etc. but this also has a rank deficiency problem, for which the predictions may be misleading and are not reliable.

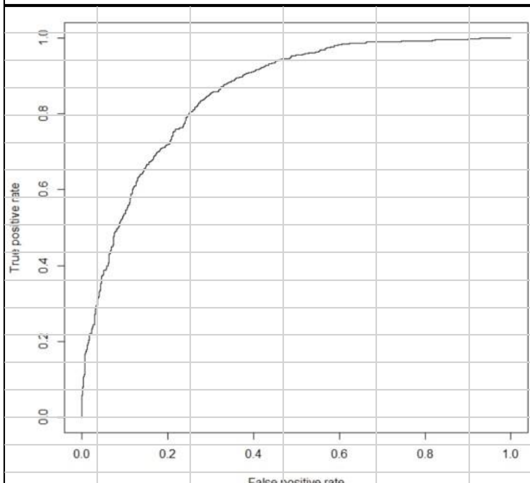
CONFUSION MATRIX



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1147  109
         1  401  451

               Accuracy : 0.7581
                 95% CI : (0.7392, 0.7762)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 0.006926

                  Kappa : 0.4684

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8054
            Specificity : 0.7410
         Pos Pred Value : 0.5293
         Neg Pred Value : 0.9132
             Prevalence : 0.2657
         Detection Rate : 0.2139
   Detection Prevalence : 0.4042
      Balanced Accuracy : 0.7732

       'Positive' Class : 1
```

With only the selected features by the algorithm, overall results are good, but predictions can be misleading due to rank definiciencyof the data. As stated earlier, this is a drawback of logistic regression where it's unable to provide reliable results when there are too many inputs.

ROC CURVE



The ROC curve shows the plot between true positive rates and true negative rates. The area under the AUC curve is 85%. This indicates that this is a good model. But since, the predictions aren't reliable due to rank deficiency, this could possibly mean overfitting.

While stepwise logistic regression selected the most important features, the predictions may not reliable due to overfitting.

CTREE Classification

Decision trees was the next classification model that was used. Decision tree builds regression models in the shape of a tree structure. They are based on an algorithm called ID3 that uses entropy and information gain methods. Data is broken down into smaller subsets and illustrates the most important features in Nodes and Leaves. Nodes represent a decision test, examining a single feature and move to the next outcome. Leaves represent the outcome of the decision from the Nodes. Then, to get better results we start pruning the tree by removing the duplicates or use business logic.

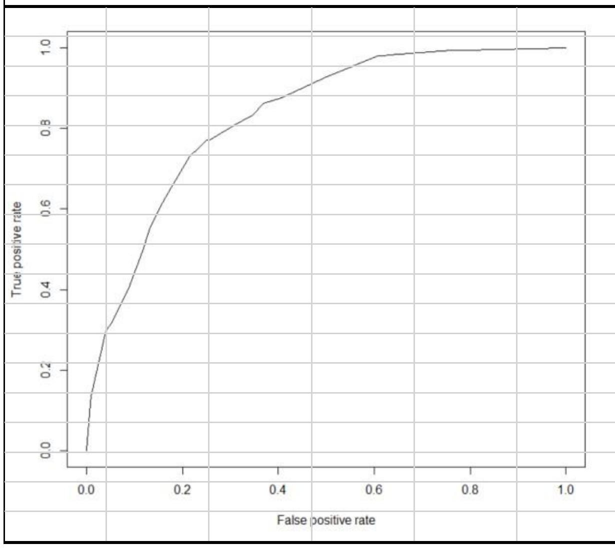
For modelling, the CTREE classification model was used. A snippet of the plot generated by the CTREE model is provided in the appendix. The plot shows that the root note is the Contract feature which means that Contract is the most important indicator in predicting customer churn. From contract, the branches are month to month and one year, two year, that leads to the second Node. After contract, the second node is internet service, then depending on the type of internet service, the next important features are monthly charges, tenure and so on. This is how decision trees come up with the most important features that have predicting power in determining customer behaviour. Then, we use our trained model to make predictions on our holdout set. The confusion matrixand the ROC curve are shown in the next page.

CONFUSION MATRIX

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1160  128
         1  388  432

               Accuracy : 0.7552
                 95% CI : (0.7363, 0.7734)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 0.0154

                  Kappa : 0.4536

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7714
            Specificity : 0.7494
         Pos Pred Value : 0.5268
         Neg Pred Value : 0.9006
             Prevalence : 0.2657
         Detection Rate : 0.2049
   Detection Prevalence : 0.3890
      Balanced Accuracy : 0.7604

       'Positive' Class : 1
```

From the confusion matrix, the sensitivity is 77% and the specificity is 75%, The difference between sensitivity and specificity is 2%. This indicates that this is a good model to predict customer behaviour. Accuracy of the model is 76%, however, accuracy itself is not a good measure for classification.

ROC CURVE



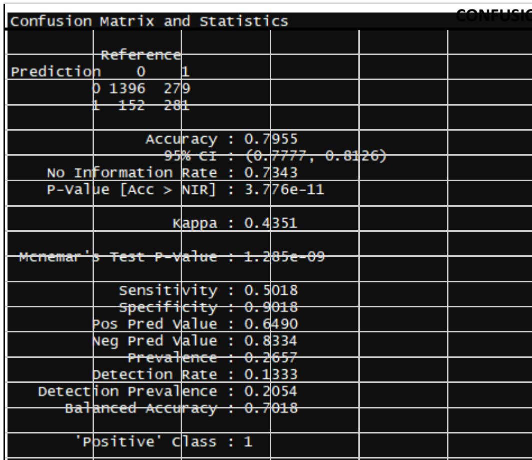
The ROC curve shows the plot between true positive rates and true negative rates. The area under the AUC curve is 83% which indicates that it is a good model.

77% of the time the model was successful to predict true positives and 75% of the time the model was successful to predict true negatives. With an AUC of 83% and a lower difference between sensitivity and specificity, this model can be used for prediction purposes.

RPART Classification

Recursive partitioning is another algorithm used to build tree-based model that has a complexity parameter. The complexity parameter is the minimum improvement in the model needed at each node. Like CTREE, the algorithm splits the data into smaller chunks and then the process is repeated on the smaller chunks. The complexity parameter given by the algorithm for the model is 0.002. The tree is pruned at this level and shows similarities where contract is the node root, then, branches was month to month, one year and two year and so on. A snippet of the pruned decision tree with recursive partitioning is attached in the appendix.

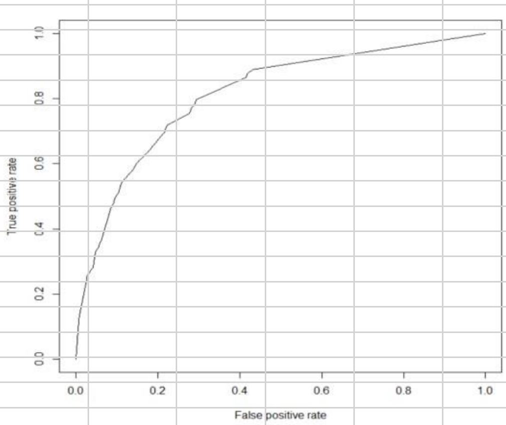
CONFUSION MATRIX



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1396  279
         1  152  281

               Accuracy : 0.7955
                 95% CI : (0.7777, 0.8126)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 3.776e-11

                  Kappa : 0.4351

 Mcnemar's Test P-Value : 1.285e-09

            Sensitivity : 0.5018
            Specificity : 0.9018
         Pos Pred Value : 0.6490
         Neg Pred Value : 0.8334
             Prevalence : 0.2657
         Detection Rate : 0.1333
   Detection Prevalence : 0.2054
      Balanced Accuracy : 0.7018

       'Positive' Class : 1
```

From the confusion matrix, the sensitivity is 50% and the specificity is 90%, The difference between sensitivity and specificity is 40%. This indicates that this is not a good model to predict customer behaviour. Accuracy of the model is 80%, however, accuracy itself is not a good measure for classification.

ROC CURVE



The ROC curve shows the plot between true positive rates and true negative rates. The area under the AUC curve is 81%.

RPART has a high difference between true positives and true negative. The model is predicting true negatives better, but not doing a great job in predicting true positives as it is missing a lot of the customers that are churning which is very costly to the company. Hence, this is not a good model and the predictions are not reliable.

Random Forest

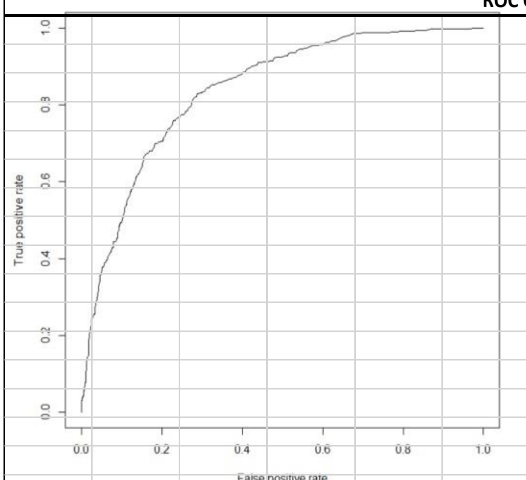
The random Forest algorithm combines large number of decision trees and provides an ensemble model. When large number of uncorrelated trees operate together as a group, they tend to perform better. The first plot in the random forest appendix shows different classes (colored) and out of bag samples (black). From the plot the error seems to be lowered at around 150 trees. One of the drawbacks of random forest is that it's not easy to interpret. The confusion matrix for the random forest model is provided below.

CONFUSION MATRIX

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1160  126
         1  388  434

               Accuracy : 0.7562
                 95% CI : (0.7373, 0.7744)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 0.01191

                  Kappa : 0.4562

 Mcnemar's Test P-Value : < 2e-16

            Sensitivity : 0.7750
            Specificity : 0.7494
         Pos Pred Value : 0.5280
         Neg Pred Value : 0.9020
             Prevalence : 0.2657
         Detection Rate : 0.2059
   Detection Prevalence : 0.3899
      Balanced Accuracy : 0.7622

       'Positive' Class : 1
```

From the confusion matrix, the sensitivity is 78% and the specificity is 75%, The difference between sensitivity and specificityis3%.Thisindicatesthatthisisnotagoodmodelto predict customer behaviour. Accuracy of the model is 76%, however, accuracy itself is not a good measure for classification.

ROC CURVE



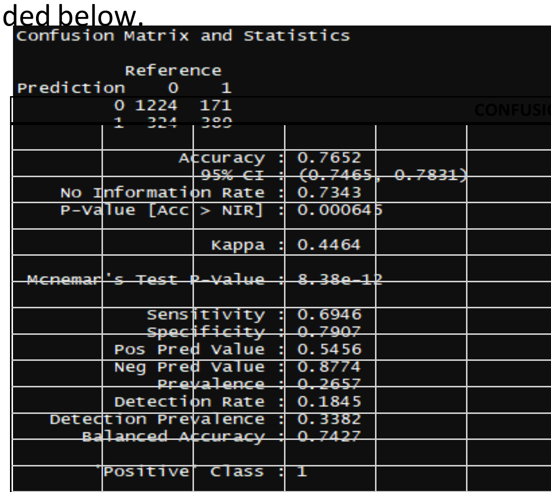
The ROC curve shows the plot between true positive rates and true negativerates.TheareaundertheAUCcurveis84%.

With high sensitivity and AUC along with lower difference between sensitivity and specificity, this is a reliable model that can be used to predict churn.

XGBoost

Next, we execute a gradient boosting framework called XGBoost which stands for Extreme Gradient Boosting. This is an ensemble technique where new models are repeatedly added to correct errors made by existing models until no further improvements can be made. Like Random forest, interpret-ability is not easy for XGBoost models. The confusion matrix is provided below.

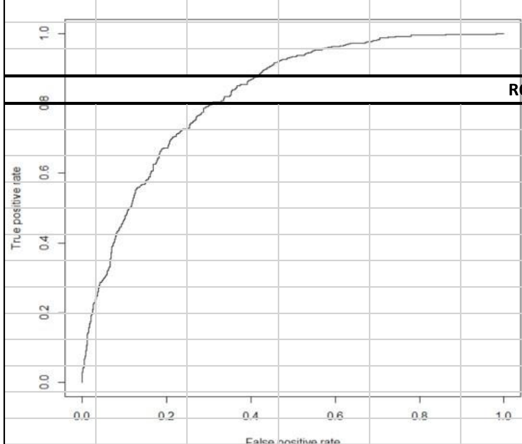
CONFUSION MATRIX



```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1224  171
         1  324  380

               Accuracy : 0.7652
                 95% CI : (0.7465, 0.7831)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 0.000645

                  Kappa : 0.4464

 Mcnemar's Test P-Value : 8.38e-12

            Sensitivity : 0.6946
            Specificity : 0.7907
         Pos Pred Value : 0.5456
         Neg Pred Value : 0.8774
             Prevalence : 0.2657
         Detection Rate : 0.1845
   Detection Prevalence : 0.3382
      Balanced Accuracy : 0.7427

       'Positive' Class : 1
```

From the confusion matrix, the sensitivity is 69% and the specificity is 79%, The difference between sensitivity and specificity is 10%. This indicates that this is not a good model to predict customer behaviour. Accuracy of the model is 77%, however, accuracy itself is not a good measure for classification.

ROC CURVE



The ROC curve shows the plot between true positive rates and true negative rates. The area under the AUC curve is 82%.

XGBoost has a high difference between true positives and true negatives. The model is predicting true negatives better, but not doing a great job in predicting true positives. Hence, it is not recommended to use this model for churn prediction as a lot of the churning customers are being missed.

Neural Network

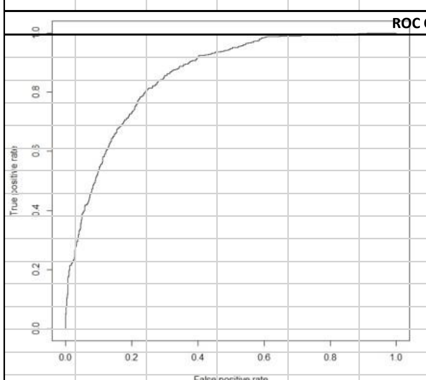
Neural Network is a series of algorithms that can adapt to changing input and recognize the relationship in the data through a similar process that mimics the way human brain operates. The model has the criteria to provide the best results without the need to redesign output criteria. The neural network plot is provided in the appendix. The confusion matrix and ROC curve is provided below.

CONFUSION MATRIX

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1400  269
         1  148  291

               Accuracy : 0.8022
                 95% CI : (0.7845, 0.819)
    No Information Rate : 0.7343
    P-Value [Acc > NIR] : 2.183e-13

                  Kappa : 0.4554

 Mcnemar's Test P-Value : 4.192e-09

            Sensitivity : 0.5196
            Specificity : 0.9044
         Pos Pred Value : 0.6629
         Neg Pred Value : 0.8388
             Prevalence : 0.2657
         Detection Rate : 0.1380
   Detection Prevalence : 0.2083
      Balanced Accuracy : 0.7120

       'Positive' Class : 1
```

From the confusion matrix, the sensitivity is 52% and the specificity is 91%, The difference between sensitivity and specificity is 10%. This indicates that this is not a good model to predict customer behaviour. Accuracy of the model is 80%, however, accuracy itself is not a good measure for classification.

ROC CURVE



The ROC curve shows the plot between true positive rates and true negative rates. The area under the AUC curve is 86%.

Neural Network has a high difference between true positives and true negative. The model is predicting true negatives better, but not doing a great job in predicting true positives. Hence, it is not recommended to use this model to predict churn (using the same logic that was used in XGBoost (above).

Model Results

To summarize, the results of each of the models are provided below.

	Accuracy	AUC	Sensitivity	Specificity
Stepwise Regression	76%	86%	81%	74%
CTREE	76%	83%	77%	75%
RPART	80%	81%	50%	90%
Random Forest	76%	84%	78%	75%
XGBoost	82%	77%	69%	79%
Neural Network	80%	86%	52%	90%

To predict churn, the aim is to have high true positives, i.e. high sensitivity and high AUC. Following this logic, the best model is Random Forest with a Sensitivity of 78% and an AUC of 84%. The second-best model is CTREE classification with a sensitivity of 77% and an AUC of 83%. Note that stepwise regression gives a higher sensitivity (81%) and a higher AUC (86%). However, during this prediction, the result comes with a warning of misleading prediction due to rank deficiency. The misleading prediction associated with higher sensitivity could be symptoms of over fitting. Hence, this was not chosen as the best model. Further analysis is required before using this model for churn prediction.

Cost Matrix

We also used a cost-savings based approach to rank our model performance. The cost savings table below was constructed by using the cost of customer churn calculation number and some assumptions. We assumed that the company would send a $50 coupon to customers that the model predicts would churn and further assumed that 20% of these customers would choose to remain with the firm (because of the $50 coupon). We then calculated the cost savings generated by the firm for each of the models by comparing the cost after deploying the model (coupon cost plus churn cost) with the cost of the no-model scenario. The results of this exercise supported the conclusion made earlier. Stepwise regression had the greatest cost savings ($72, 253) on the test data, while Random Forest came in at a close second ($ 69, 438). Using the same logic as above, we choose to deploy the Random Forest Model, due to the **Rank Deficiency** issues with Stepwise Regression.

Model	Accurac	AUC	Sensitivit	Specificit	Number of Customers (Test set)	FP	FN	TP	TN	CostofChur	Coupon Co	Cost Without Mod	CostwithModel	Cost Saving	Rank
StepwiseRegression	76%	86 %	81%	74%	2,108	401	109	451	1,147	$ 1,269	$ 50	$ 711,029	$ 638,776	$ 72,253	1
CTREE	76%	83%	77%	75%	2,108	388	128	432	1,160	$ 1,269	$ 50	$ 711,029	$ 641,998	$ 69,030	3
RPART	80%	81 %	50%	90%	2,108	152	279	281	1,396	$ 1,269	$ 50	$ 711,029	$ 660,972	$ 50,057	6
Random Forest	76%	84%	78%	75%	2,108	388	126	434	1,160	$ 1,269	$ 50	$ 711,029	$ 641,591	$ 69,438	2
XGBoost	82%	77 %	69%	79%	2,108	324	171	389	1,224	$ 1,269	$ 50	$ 711,029	$ 647,562	$ 63,467	4
Neural Network	80%	86%	53%	90%	2,108	154	262	298	1,394	$ 1,269	$ 50	$ 711,029	$ 657,608	$ 53,421	5

* Assuming that 20% of customers given a coupon choose to stay with the Telco

The confusion matrix for the random forest model is shown below.

	0	**1**
0	**1160**	**126**
1	**388**	**434**

Out of the 560 churns in the hold out set, the model was able to perfectly predict 432 (true positives) of them, while predicting 1160 (true negatives) no churns perfectly from 1548 no churns.

From the data set we computed that the cost of churn is $1,269 per customer (see churn cost calculation table in appendix)? By being able to predict churn, the company should be able to take actions, such as sending coupons and offering promotions in order to retain some of these customers.

The matrix below shows the percentage of true positives, true negatives, false positives and false negatives.

TRUE NEGATIVES 55%	FALSE NEGATIVES 6%
FALSE POSITIVES 18%	TRUE POSITIVES 21%

For true positives, the customer churns and the model predicts churn. In this case, the company can take steps to retain the customer by offering promotions or sending coupons. For false positives, the customer doesn't churn, but the model predicts churn. In this case, the company is suggested to send coupons or offer deals to clients who are unlikely to churn in the first place. For true negatives, the customers don't churn, and the model also predicts no churn. There are no costs associated with this (no action required for the firm). For false negatives, the customer churns, but the model predicts no churn. In this case the cost to the company is high as the company loses the opportunity to retain the client by offering promotions ($1,269). By using the cost matrix from the random forest model, the company can likely retain more customers.

Recommendations

Based on the results of the modelling exercise, we believe that Telco should implement the following three recommendations to reduce the direct and indirect costs associated with customer churn. These recommendations were made under the assumption that acquiring a new customer is costlier for a company than keeping an existing one. For telco, the cost of customer churn was calculated through an arithmetic (see appendix) and determined to be $1,269 per customer. Additionally, it is assumed that it would be possible to convince a (churning) customer to stay by offering them a $50 coupon.

The simplest recommendation would be to use the output (predictions) of the model and offer a coupon to customers to stay with the firm. While this might give them a short-term incentive (to stay), once the benefits of the coupon expire, their risk of churning would increase (again). The aim of our recommendations is to change the long-term risk profile of the customers to make them more loyal to Telco:

1) Provide financial incentives to consumers to sign 2-year contracts: Based on the output of the model, the firm should provide financial incentives to customers – those that the model predicts will churn – to switch into 2-year contracts. Since customers on long-term (2-year) contracts have a lower churn rate (7.5%), it would make sense for Telco to offer an incentive to customers to achieve this.
2) Create a bundle for additional features (Online Backup, Device Protection, Online Security) and incentivize (financial) customers at a risk of churning to add these bundles to their account. This recommendation came out of our observation that a significantly higher proportion of non-churning customers had signed up for one or more of these features.
3) Telco should create a new Phone Service plan geared towards customers with families (dependents and/or partners). It was observed that customers with partners and dependents have a lower probability of churning. Marketing these plans to such customers at a lower cost (than purchasing individual lines) will be beneficial to both the firm and customer: It

will generate cost savings for the firm through a lower churn rate and provide the customers with increases convenience and lower overall cost.

For each of the recommendations above, that the firm offer these to customers before they are 20 months into their tenure. This is because of a sharp decrease in survival probability (based on Kaplan-Meier survival curve) around this time.

Conclusion

In summary, by implementing the random forest model that was developed in this project, Teleco will likely be able to generate significant cost savings – we were able to show savings of $69, 438 on 2,108 test records - through the deploying the offer incentive program on the churning customers. Besides the insights we learned from the analytics perspective, Telco is encouraged to send out customer exit surveys or customized emails to find out more regarding the unexpected churn. Meanwhile, increasing customer engagement by focusing on the customer satisfaction and retention throughout the product life cycle can potentially assist in lowering the churn rate as well.

Appendix

Model Performance based on cost savings

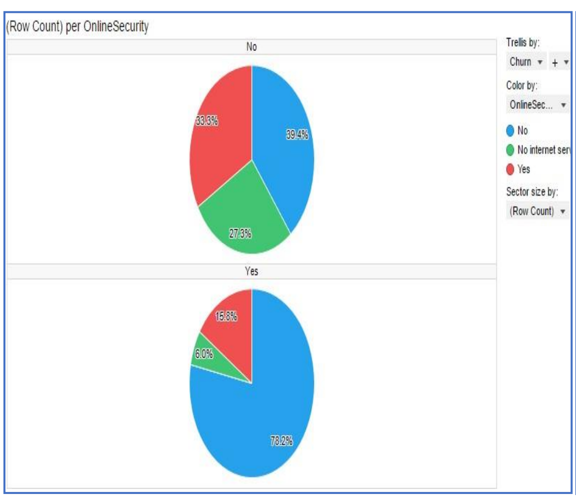
Model	Accurac	AUC	Sensitivit	Specificit	NumberofCustomers(Testset)	FP	FN	TP	TN	Costof	CouponCo	Cost Without Mod	CostwithModel	Cost Saving	Rank
Stepwise Regression	76%	86%	81%	74%	2,108	401	109	451	1,147	$ 1,269	$ 50	$ 711,029	$ 638,776	$ 72,253	1
CTREE	76%	83%	77%	75%	2,108	388	128	432	1,160	$ 1,269	$ 50	$ 711,029	$ 641,998	$ 69,030	3
RPART	80%	81%	50%	90%	2,108	152	279	281	1,396	$ 1,269	$ 50	$ 711,029	$ 660,972	$ 50,057	6
Random Forest	76%	84%	78%	75%	2,108	388	126	434	1,160	$ 1,269	$ 50	$ 711,029	$ 641,591	$ 69,438	2
XGBoost	82%	77%	69%	79%	2,108	324	171	389	1,224	$ 1,269	$ 50	$ 711,029	$ 647,562	$ 63,467	4
Neural Network	80%	86%	53%	90%	2,108	154	262	298	1,394	$ 1,269	$ 50	$ 711,029	$ 657,608	$ 53,421	5

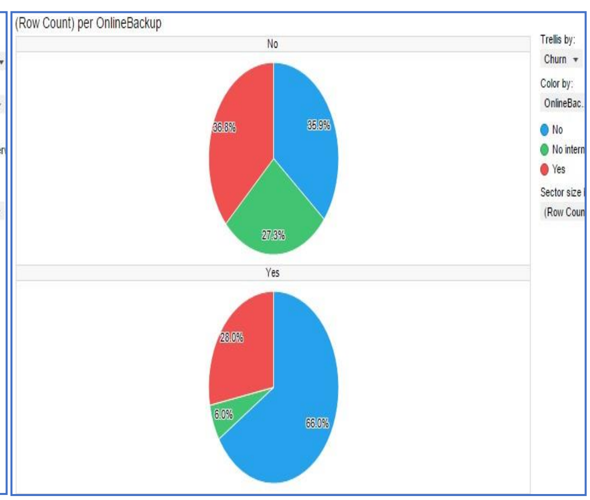
* Assuming that 20% of customers given a coupon choose to stay with the Telco

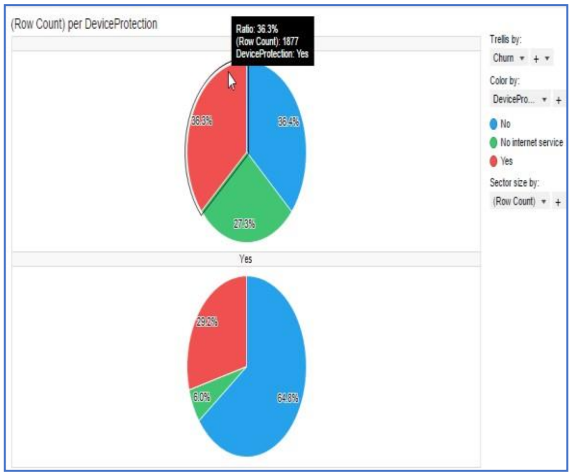
Cost of Churn Calculation

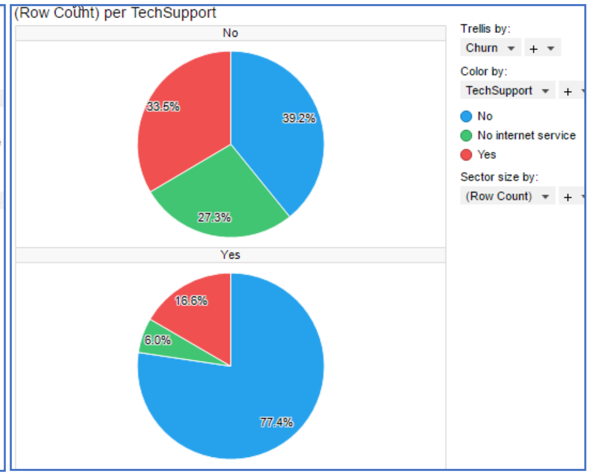
Average Tenure No Churn (Months)	37.57
Average Tenure Churn (Months)	17.98
Difference (Months)	19.59
Average Monthly Charges ($)	$ 64.77
Cost of Churn ($)	$ 1,269

Device Protection, Online Backup and Online Security Pie Charts

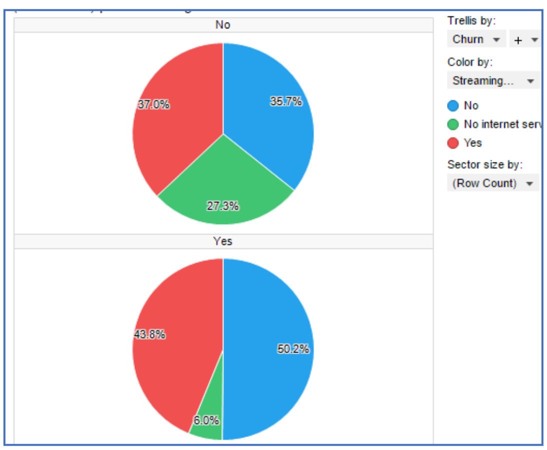


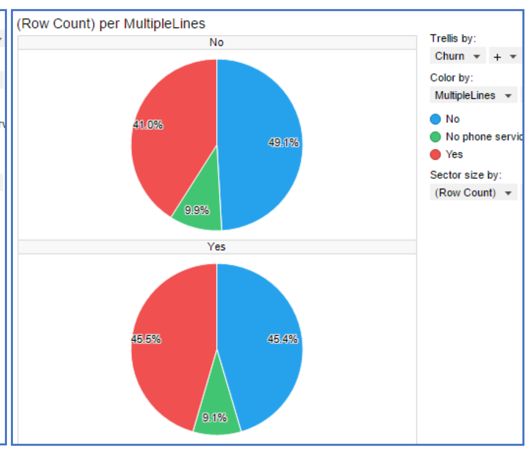




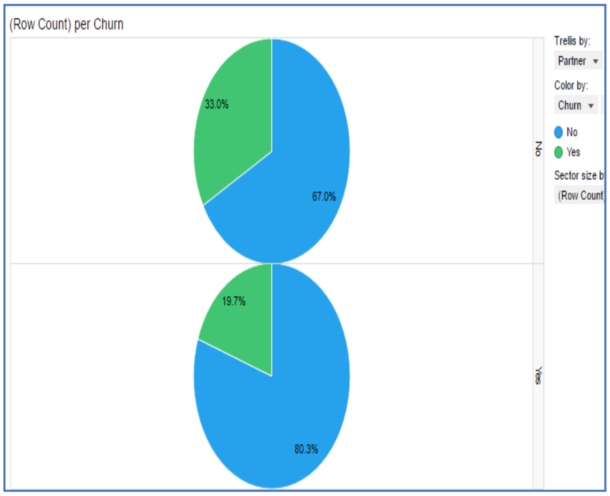


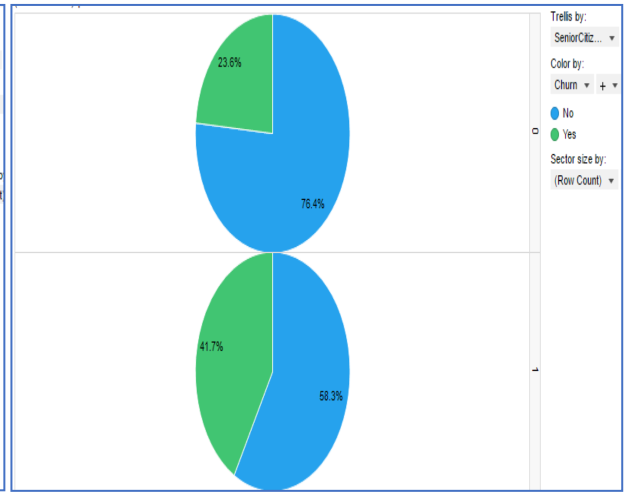
Streaming Movies, Streaming TV and Multiple Lines





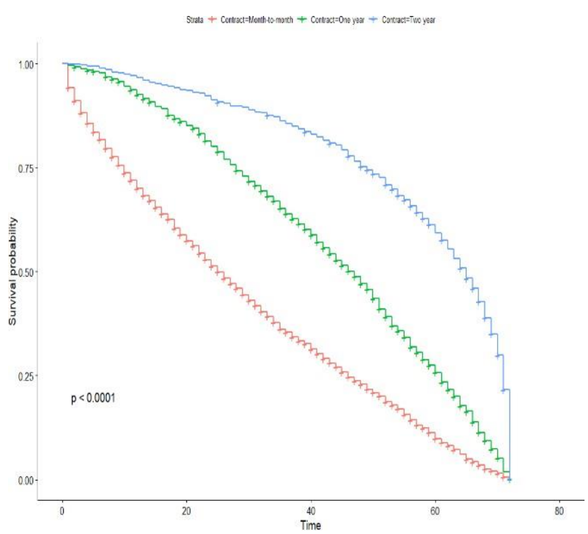
Customer Demographics

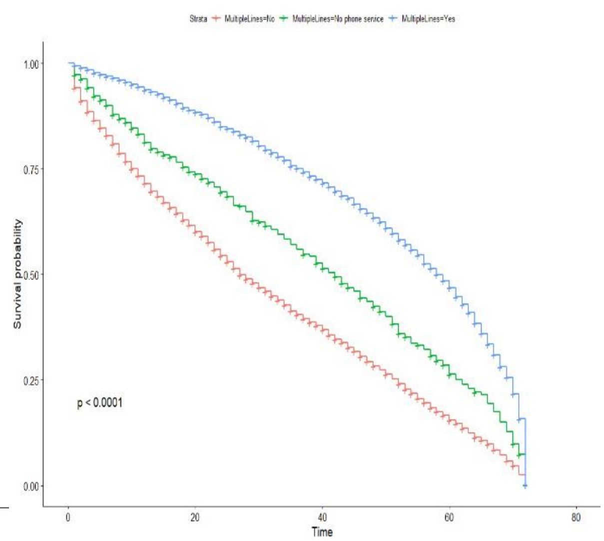


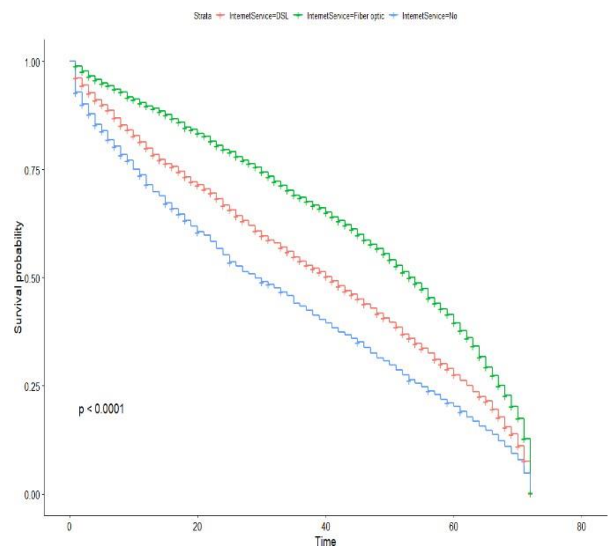


Survival Analysis

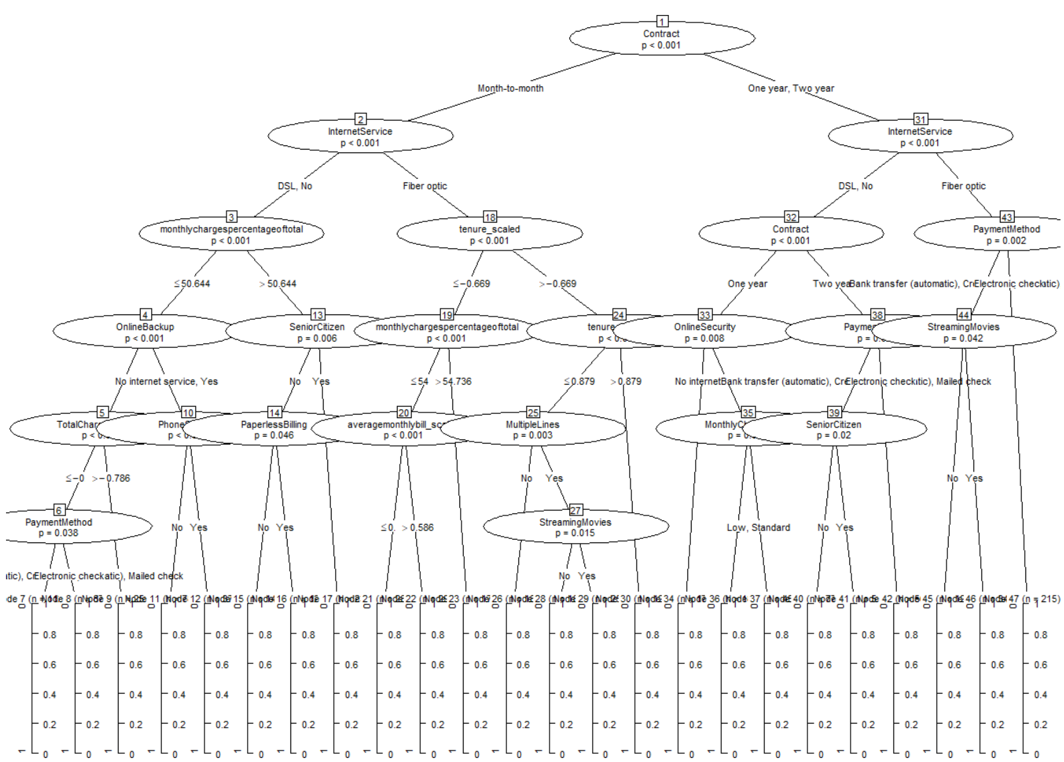
Kaplan Mier Plots



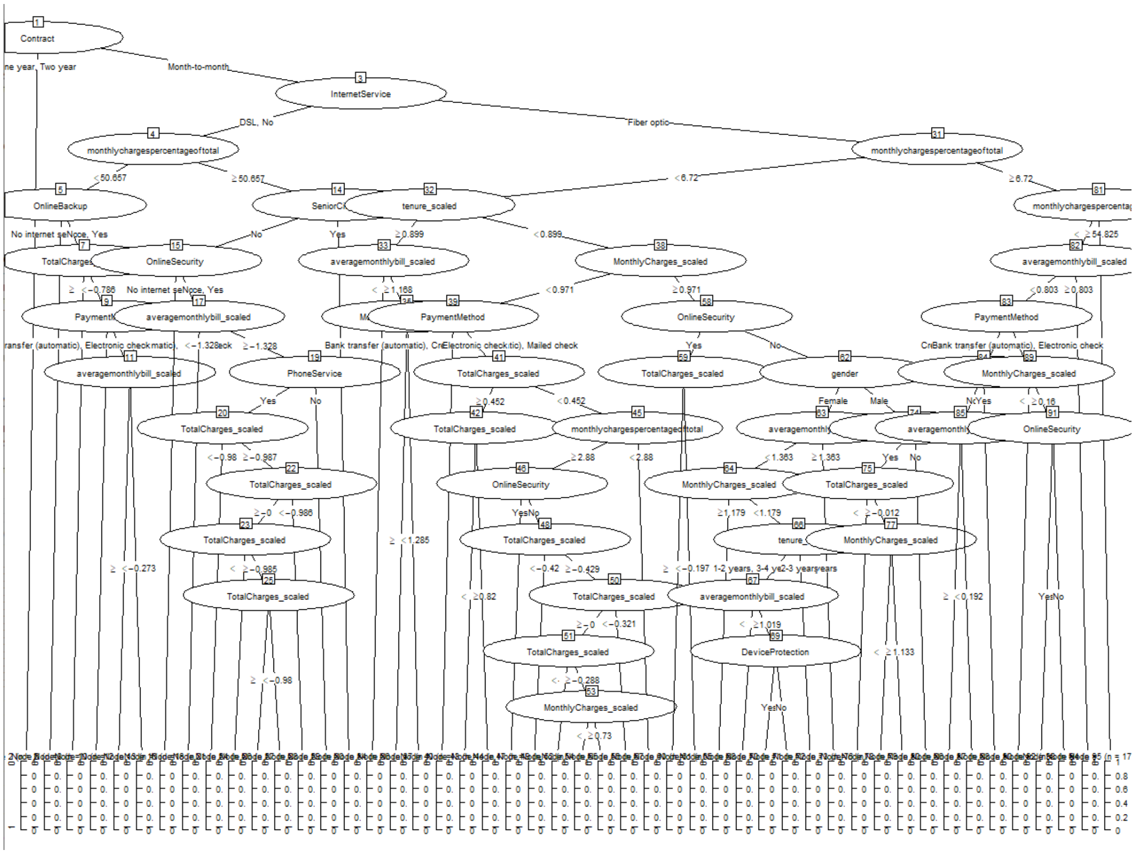




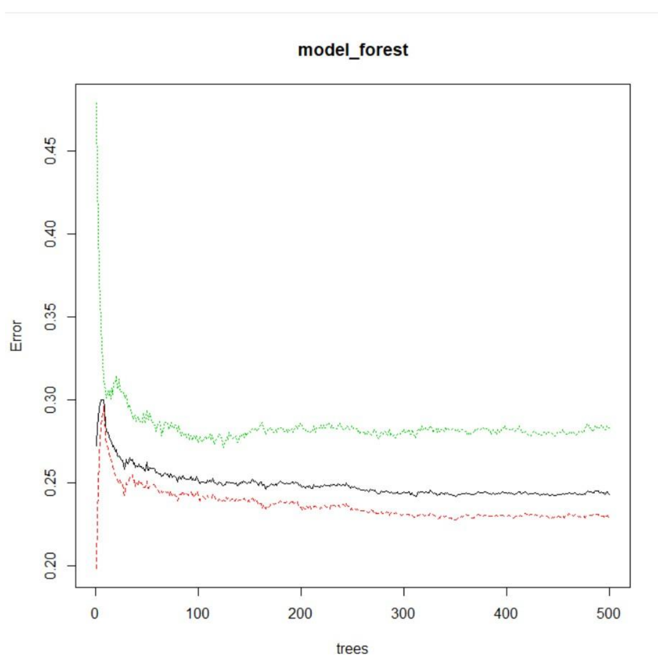
CTREE SNIPPET



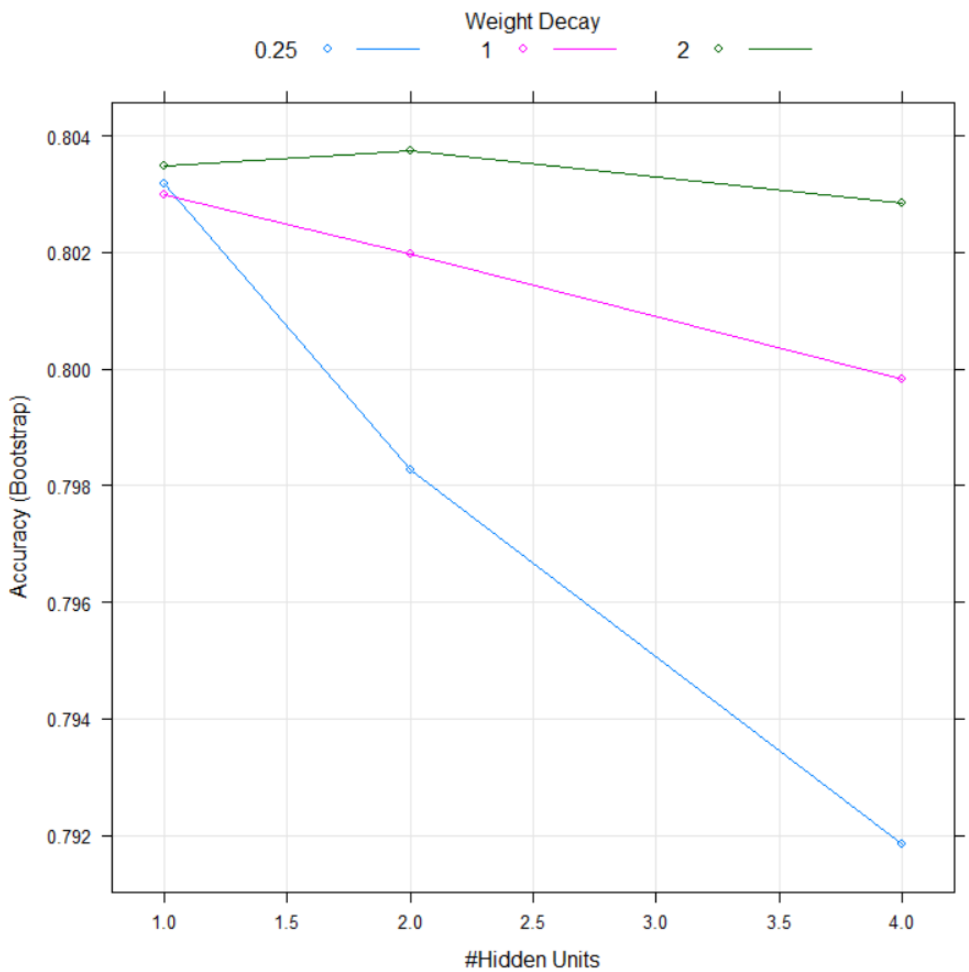
RPART SNIPPET



Random Forest Snippet



Neural Network Snippet



Reference

https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new- customers-nurture-the-existing-ones/#793f68bc5a8e

https://acquire.io/blog/reduce-customer-churn-rate/