Customer Churn - Prediction Models & Analysis

A report from IESEG Consultants to TelCo

Hina Hussain
Inderpreet Rana
Mui Han Ma

Content:

I. Introduction

The objective of this report is to investigate customer churn for TelCo. TelCo is one of the largest telecommunication service providers. As part of its value system to provide quality customer experience, it wants to investigate reasons for customer churn. Currently, a simple churn prediction model is being used but TelCo has acquired the services of IESEG Consultants to propose an efficient churn prediction model.

In this report, IESEG Consultants study the company data and its existing model. The available information is then used to build more accurate models. Interpretation techniques are also outlined in the report to help TelCo seamlessly implement the proposed models.

4250 Customers

598 Churners

14% Churn Rate

II. TelCo Data & Current Scenario

The report is based on the following customer level data recorded by TelCo.

Variable	Name	Meaning
1	State	2 letter code of the US state of customer residence
2	Account_length	Number of months the customer has been with the current telco provider
3	Area_code	
4	International_plan	The customer has international plan.
5	Voice_mail_plan	The customer has voice mail plan.
6	Number_vmail_messages	Number of voice-mail messages.
7	Total_day_minutes	Total minutes of day calls.
8	Total_day_calls	Total number of day calls.
9	Total_day_charge	Total charge of day calls.
10	Total_eve_minutes	Total minutes of evening calls.
11	Total_eve_calls	Total number of evening calls.
12	Total_eve_charge	Total charge of evening calls.

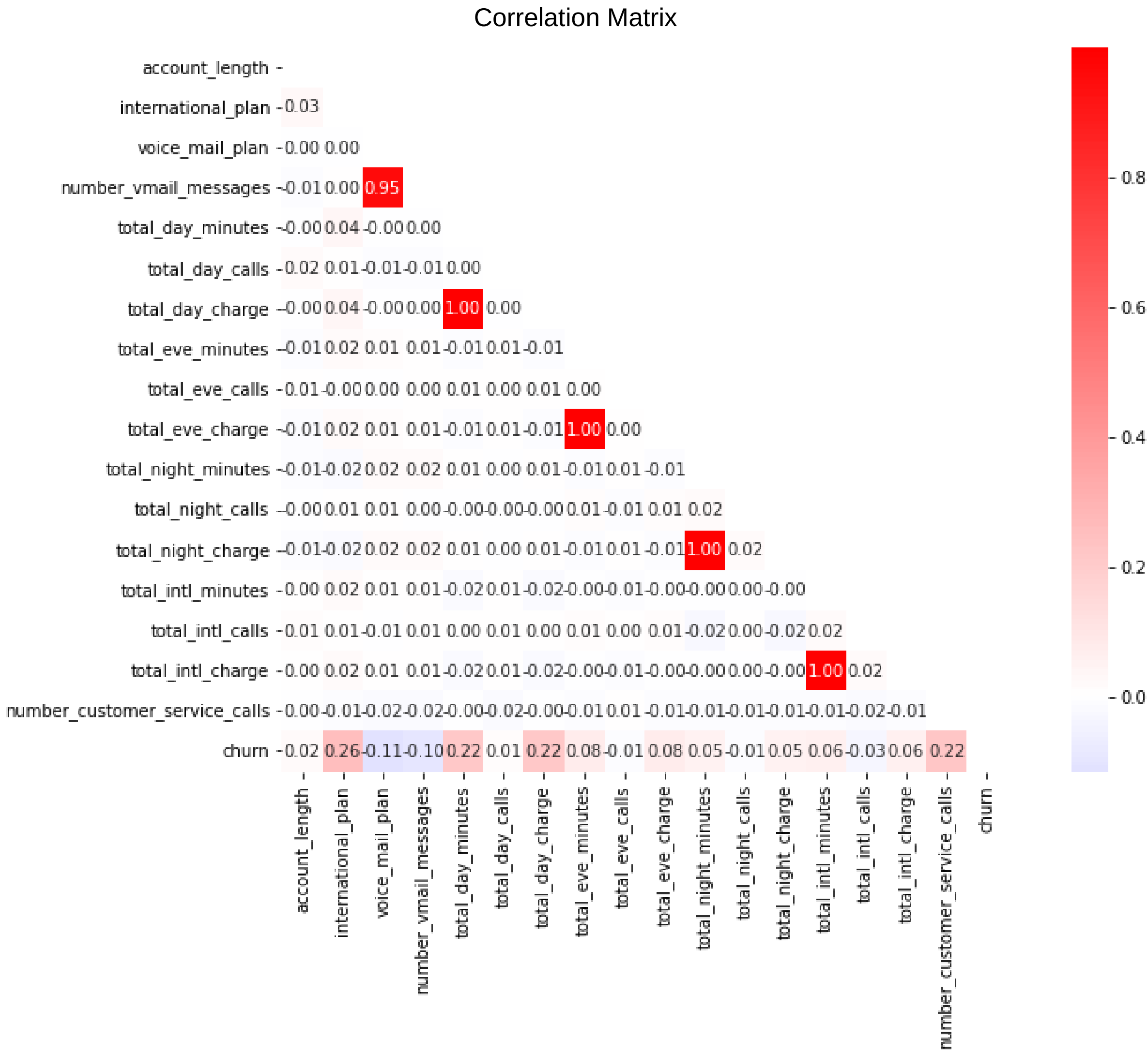
13	Total_night_minutes	Total minutes of night calls.
14	Total_night_calls	Total number of night calls.
15	Total_night_charge	Total charge of night calls.
16	Total_intl_minutes	Total minutes of international calls.
17	Total_intl_calls	Total number of international calls.
18	Total_intl_charge	Total charge of international calls.
19	Number_customer_ service_calls	Number of calls to customer service.
20	Churn	Customer churn

Currently, TelCo is using a linear model to predict churn. This method is used for its ease of interpretability. The insights provided by the coefficients of the linear model are used to continuously enhance the customer experience and take preventative measures for potential churn in the future. Additionally, customers with high churn probabilities receive personalized retention offers (e.g., discounts, free minutes).

III. Feature Selection

The first step to building a model with high accuracy is to select variables wisely. Variables can have correlations between themselves and if this is not considered, the results can be biased.

This report uses two main methods for feature selection **- Correlation Matrix** and **Feature Importance**. A correlation matrix is a table showing correlation coefficients between sets of variables. Each variable in the table is correlated with each of the other values in the table. This shows which pairs have the highest correlation. Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. They improve the efficiency and effectiveness of a predictive model on the problem.



How to read: The further away the correlation coefficient is from zero, the stronger the relationship between the two variables.

Feature Importance method was run on three models - Decision Trees, Random Forest, and Neural Network. The following tables show the weights given to each variable by the three models.

Feature	Neural network	Decision tree	Random forest
total_day_minutes	*0.5033 ± 0.1453*	*0.1032*	*0.2032*
number_customer_service_calls	*0.4499 ± 0.3130*	*0.1458*	*0.1371*
number_vmail_messages	*0.2243 ± 0.2151*	0.0790	0.0447
total_eve_minutes	*0.2144 ± 0.2345*	0.0427	0.0470
total_night_minutes	*0.1654 ± 0.2240*	0.0148	0.0177
total_eve_calls	0.0746 ± 0.1520	-	0.0070
total_intl_calls	0.0377 ± 0.0088	*0.1026*	0.0317
total_day_calls	0.0328 ± 0.0296	-	0.0101
total_eve_charge	0.0294 ± 0.0345	0.0811	*0.0505*
international_plan	0.0227 ± 0.0695	0.0773	*0.0895*
total_day_charge	0.0158 ± 0.0320	*0.2395*	*0.2260*

voice_mail_plan	0.0015 ± 0.0238	-	0.0433
account_length	0.0010 ± 0.0029	0.0036	0.0087
total_intl_minutes	0.0008 ± 0.0149	*0.1103*	0.0289
total_night_charge	-0.0001 ± 0.0004	-	0.0185
total_night_calls	-0.0005 ± 0.0145	-	0.0071
total_intl_charge	-0.0031 ± 0.0054	-	0.0290

Taking both these methods into account, 4 variables were selected to base the study on. Additionally, the variable account length is also included as it shows the customers' loyalty, and it is one of the KPIs of TelCo.

Selected variables:

International Plan
Number of Voicemail messages
Total Day Minutes
Number of Customer Service Calls
Account Length

IV. Prediction Models

This report runs the following four models on the dataset provided to predict customer churn. The available dataset is first used to train each model and then predict the accuracy of each model.

A. **Logistic regression**
   A logistic regression model is used for classification problems e.g., churner vs non-churner. This dependent variable is predicted by analysing the relationship with independent variables (the selected features in this case)

B. **Decision tree**
   Decision tree analysis involves making a tree-shaped diagram to chart out a course of action (churn or not churn) based on certain conditions (selected features).

C. **Random forest**
   Random forest is a Supervised Machine Learning Algorithm used for classification. It builds decision trees on different samples and takes their majority vote for classification

D. **Neural networks**

   A neural network is a series of algorithms that tries to recognize underlying relationships in the data by mimicking the way the human brain operates. Neural networks can adapt to changing input and are known for providing high accuracy.

V. Data Split & Model Accuracy

Data Split

For training and validating the predictive models built, the available training dataset was split by the ratio 80:20. 20% of the data (equivalent to 850 customers) was used to validate the model. The split was stratified based on the proportion of churners vs non-churners

	Original Dataset	Train Dataset	Validation Dataset
Churners	598	486	112
Non-Churners	3,652	2,914	738
Total	4,250	3,400	850
Churn rate	*14.07%*	*14.29%*	*13.17%*

Model Accuracy

The following table shows the accuracy percentage for train and validation dataset performed in each of the predictive model:

Predictive Model\Dataset	Train dataset	Validation dataset
Logistic regression	85.44%	87.06%
Decision tree	91.62%	88.82%
Random forest	91.21%	90.00%
Neural networks	91.11%	90.12%

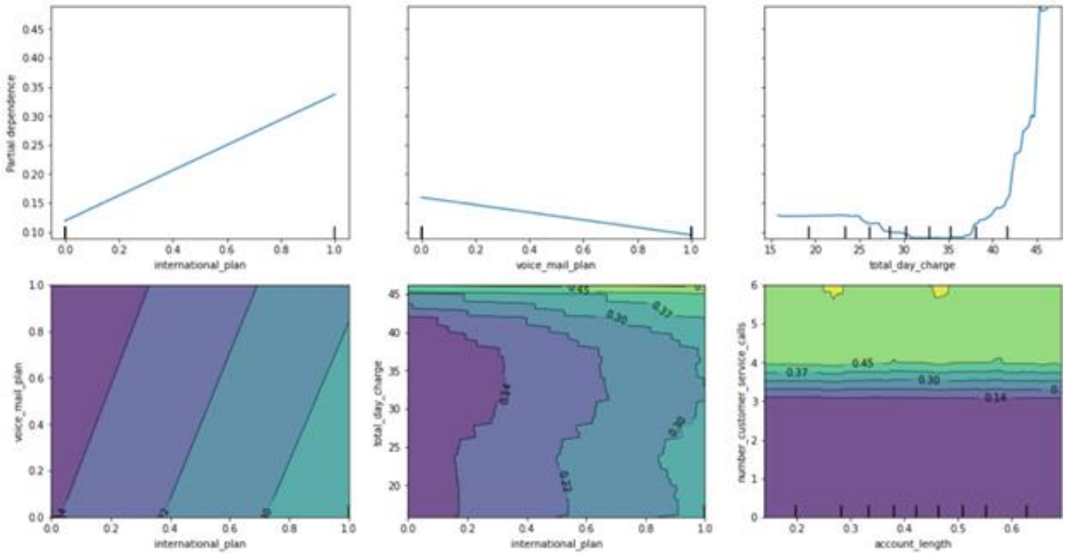
Given the high accuracy and ease of interpretability of Random Forest models, the report proposes using this model to predict customer churn. The attached Python Script provides step by step details on how to construct the Random Forest model as well as the other models. The following sections of the report provide details on how to interpret the chosen Random Forest model.

VI. Analysis

A. <u>Partial dependence plots (PDP)</u>

A partial dependence (PD) plot depicts the functional relationship between a small number of input variables and churn. They show how the predictions partially depend on values of the input variables of interest.

How to read: For 1D PDP graph: the line shows the trend on how the feature affects to the customer churn. For 2D PDP graph: the area shows how the two features affect to the customer churn, with the partial dependence marked on the outline.



The graphs above show a positive corelation between international plan and customer churn, and a negative corelation between voice mail plan and churn. There is a positive corelation between total day charge of more than USD35 and customer churn.

For the 2D PDP graphs, after taking both international plan and voice mail plan into account, the positive churn effect is only eliminated by a little by the negative corelation of voice mail plan. When international plan and total day charge are put together, they still have the similar partial dependence value like the international plan except for the total day charge more than USD45, it stays at a high partial dependence value to the customer churn.

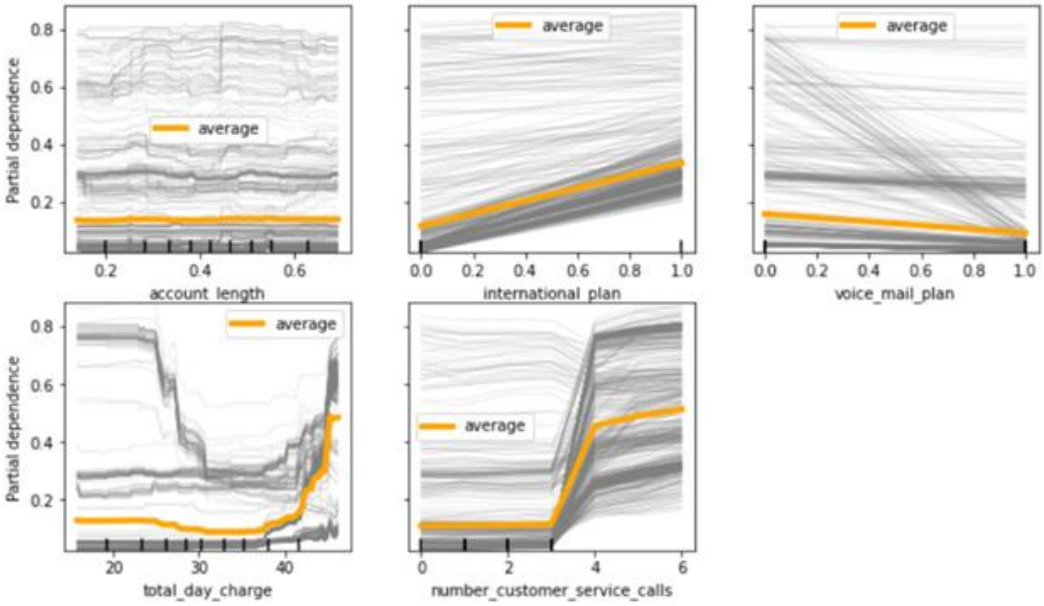
However, no relationship can be seen between account length and customer churn compared with the number of customer service calls. The number of customer service calls starts to have more effect on the customer when there are 3 calls or more.

B. <u>Individual conditional expectation (ICE)</u>

Individual Conditional Expectation (ICE) plots display one line per customer that shows how the instance's prediction changes when a feature changes.

The result is a set of points for an instance with the feature value from the grid and the respective predictions.

How to read: Each grey line stands for one customer, and the orange line stands for the average of all the customers.



Like PDP graph, no relationship can be seen between the customer churn and account length, the samples are scatter in the plot.

There is a positive relationship for international plan, and the samples are concentrated near to the average line.

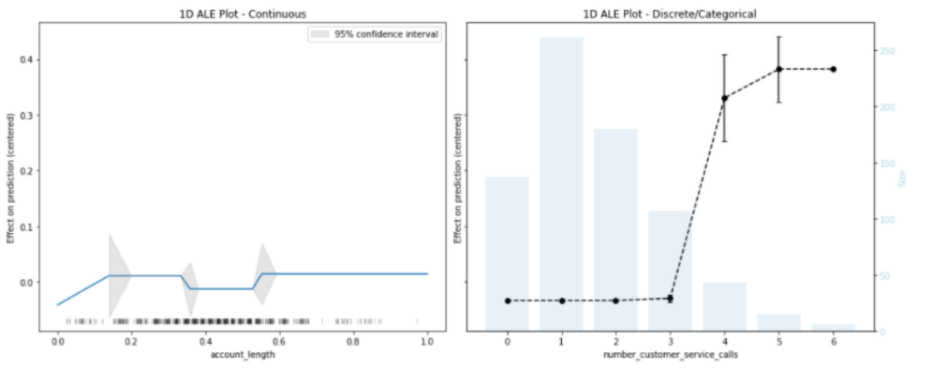
For the voice mail plan, there is a negative corelation while the samples are mainly differentiated into two groups, one is starting from 0.15 partial dependence while another group is starting from 0.3 partial dependence. But both groups show the negative relationship with the customer churn.

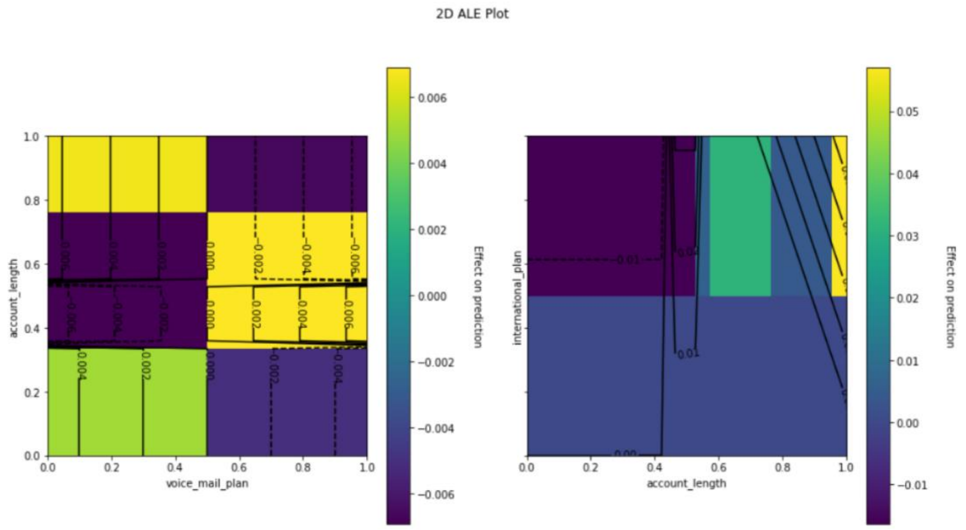
If only consider the average line, there is a sharp increase for the partial dependence to the customer churn when the total day charge is more than USD35. However, when all the grey lines are considered, there are three mainly three groups of customers having different behaviour towards customer churn with the total day charge. However, it is hard to define the reason for the different behaviour without further investigation.

For the plot of number of customer service calls, it shows there is a sharp increase for the partial dependence to the customer churn for the number of calls between 3 to 4.

C. <u>Accumulated local effects (ALE)</u>
The value of the ALE can be interpreted as the main effect of the feature at a certain value compared to the average prediction of the data. It does so by isolating the change in prediction caused by a change in a single feature. As the name implies, it does this by defining localized areas of our feature. For each local area, we take all data samples where the feature’s value falls within the area and vary the value of that feature holding all other feature values of the sample's constant. We then calculate the differences in predictions between the start and end of each area.

How to read: Like PDP graph, 1D ALE graph shows the trend on how the feature affect the prediction while 2D ALE graph shows how the two selected features affect the prediction.





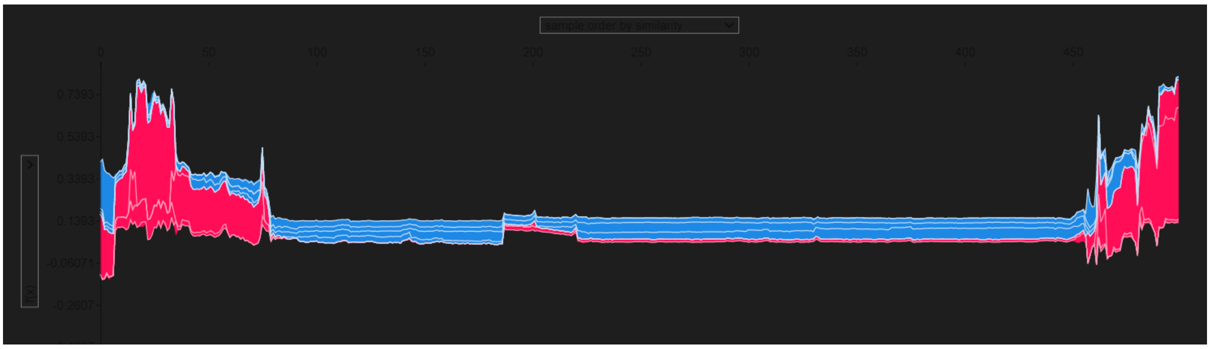
From the above 1D ALE plot, it shows that account length has a very thin effect to the customer churn. For the number of customer service calls, same as PDP & ICE graph, there is a sharp increase of the effect on prediction between 3 to 4 calls.

For 2D ALE plots, there are different groups under variables, TelCo should emphasis more on the customer with voice mail plan and with higher account length.

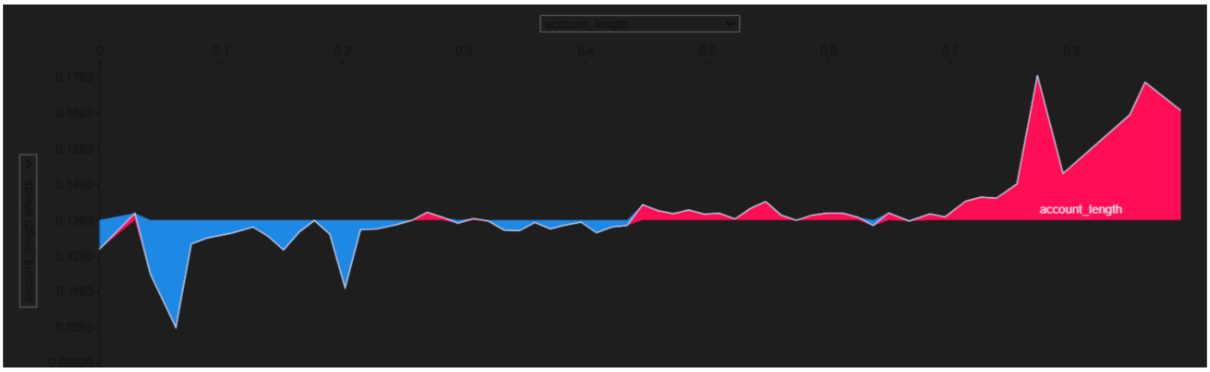
D. <u>Shapley values</u>

Shapley values help us to understand how much each feature contributed to the overall prediction. This is calculated by predicting the likelihood to churn with all the features. We then, remove one feature at a time (replacing it with a different representative value from our underlying dataset) to understand the impact of that feature – thus, understanding the contribution that it made to the overall prediction.

How to read: When the value is red, it means that the feature value was high. The more to the right the bar goes, the more a positive value contributes towards a true prediction. The more to the left the bar goes, the more a negative value contributes towards a true prediction



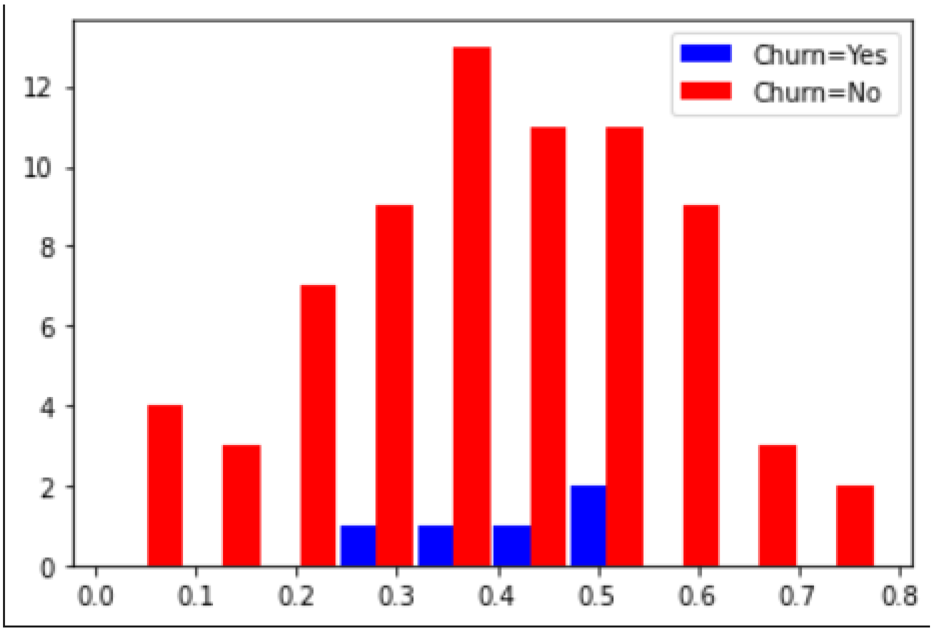
The above graph shows all the features that TelCo have, and the churners are in the two extreme groups.



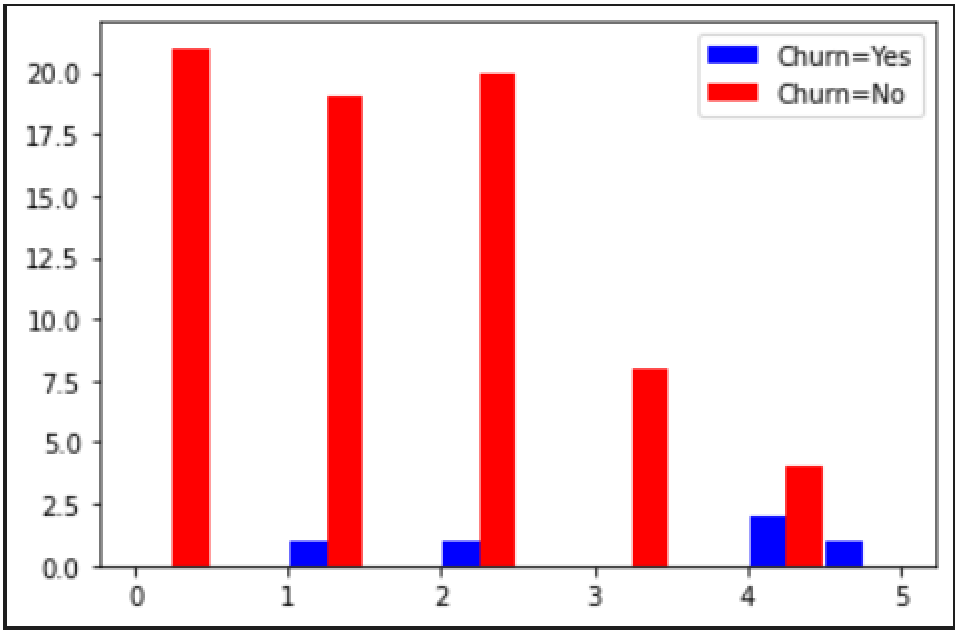
The above graph shows the customers who has higher account length is having higher chance for churn.

VII. Sub-set Data – International Plan

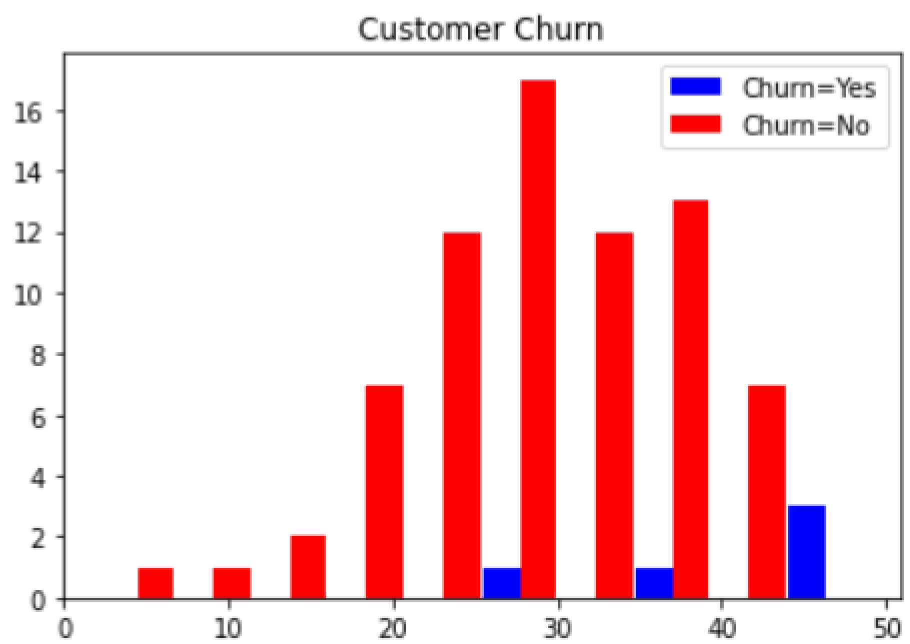
The following graphs are produced for customers who subscribe to an international plan.



This graph shows that customers are more likely to churn during mid lifecycle if the tenure of all of TelCo's customers are considered



If the customer service calls are 4 or higher, there is a higher chance for that customer to churn.



This graph shows that the higher the total day charge the more the likelihood of customer churn.

Appendix I - Reference

Feature Importance: https://machinelearningmastery.com/calculate-feature-importance-with-python/

Random Forest: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20regression.&text=It%20performs%20better%20results%20for%20classification%20problems

Neural Network: https://www.investopedia.com/terms/n/neuralnetwork.asp

Partial dependence plots: https://blogs.sas.com/content/subconsciousmusings/2018/06/12/interpret-model-predictions-with-partial-dependence-and-individual-conditional-expectation-plots/

Individual conditional expectation: https://christophm.github.io/interpretable-ml-book/ice.html

Accumulated local effects: https://www.enjine.com/blog/interpreting-machine-learning-models-accumulated-local-effects/

Shapley values: https://kodey.co.uk/2021/08/01/using-shapley-values-to-explain-your-ml-models/