TELECOM CHURN MANAGEMENT

NIRANJAN KUMAR	D18023
NARAYANA	D18030
SASIDHAR	D18014
ANIL P	D18035
SRAVANYA G TAYI	D18039
SATISH	D18036

Business Context – WHY?

- Customer Retention Rate has a strong impact on the customer lifetime value.
- Understanding the true value of a possible customer churn will help the company in it’s customer relationship management.
- If we are able to predict the churn customers in advance, the attributes of customers whom we are going to loose in near future one can take corrective action so that we can minimize this problem.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Business Questions

- Which factors are contributing to churn.
- How to control the churn rate.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Project Objective

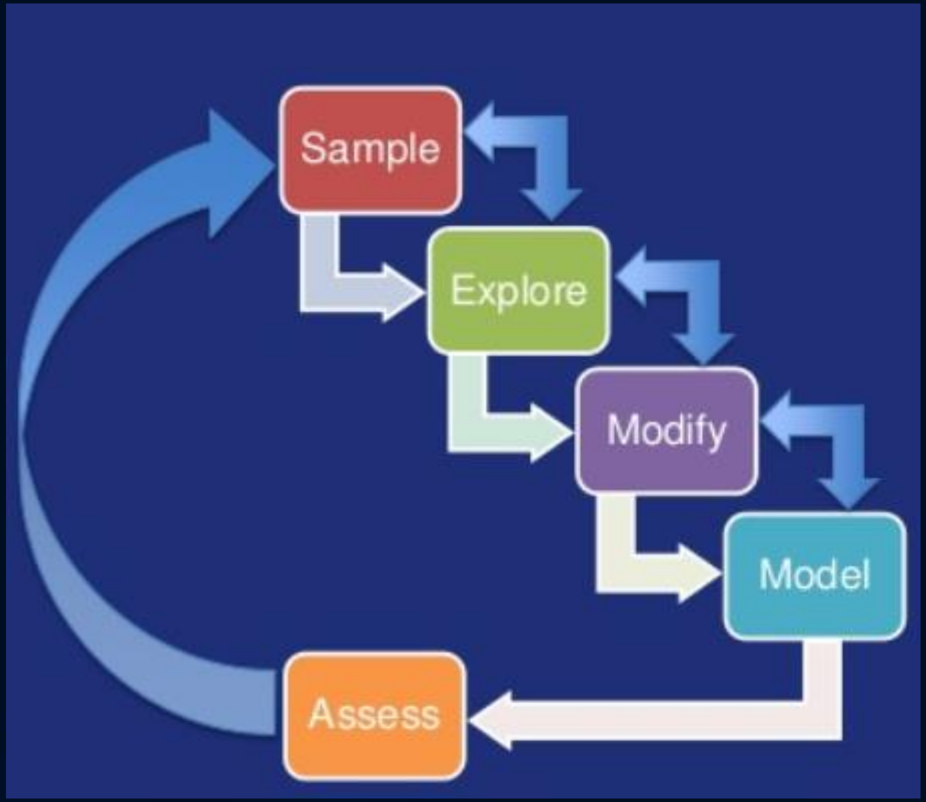
- The goal of the study is to apply analytical techniques to predict a customer churn and analyse the churning and non-churning customers.

- Develop a system through which the client could identify the customer churn rate and decide what should be the appropriate incentive for them.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

SEMMA Framework

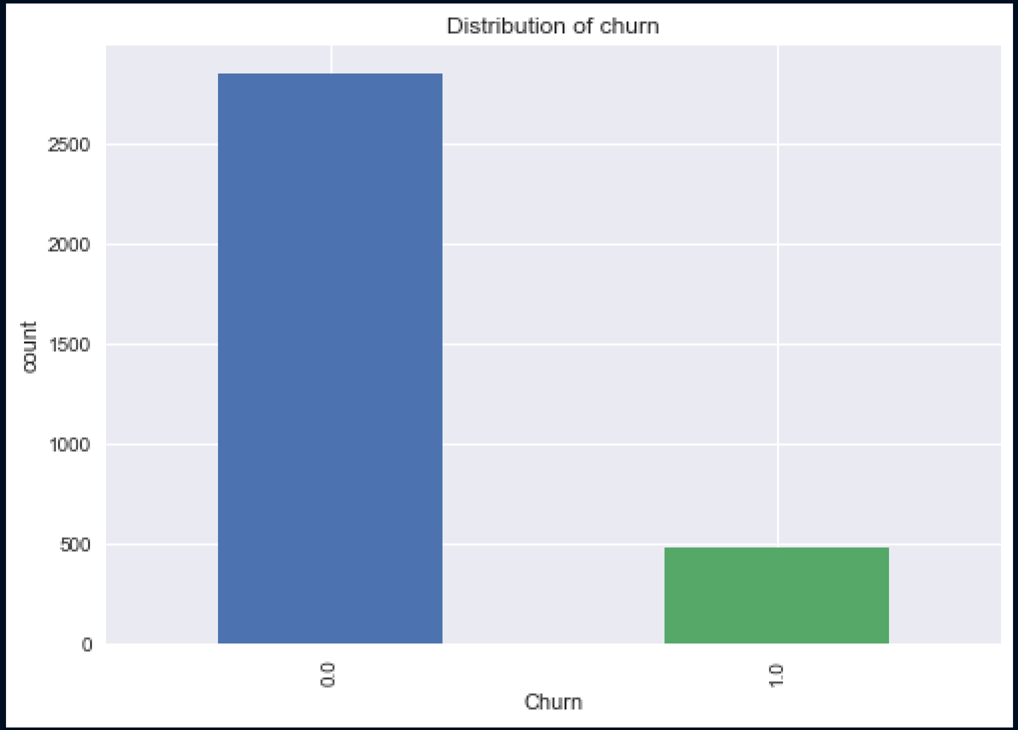
- ***SEMMA*** is an acronym that stands for *Sample*, *Explore*, *Modify*, *Model*, and *Assess.*
- SEMMA offers an easy to understand process, allowing an organized and adequate development and maintenance of DM projects.



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Challenges in Data

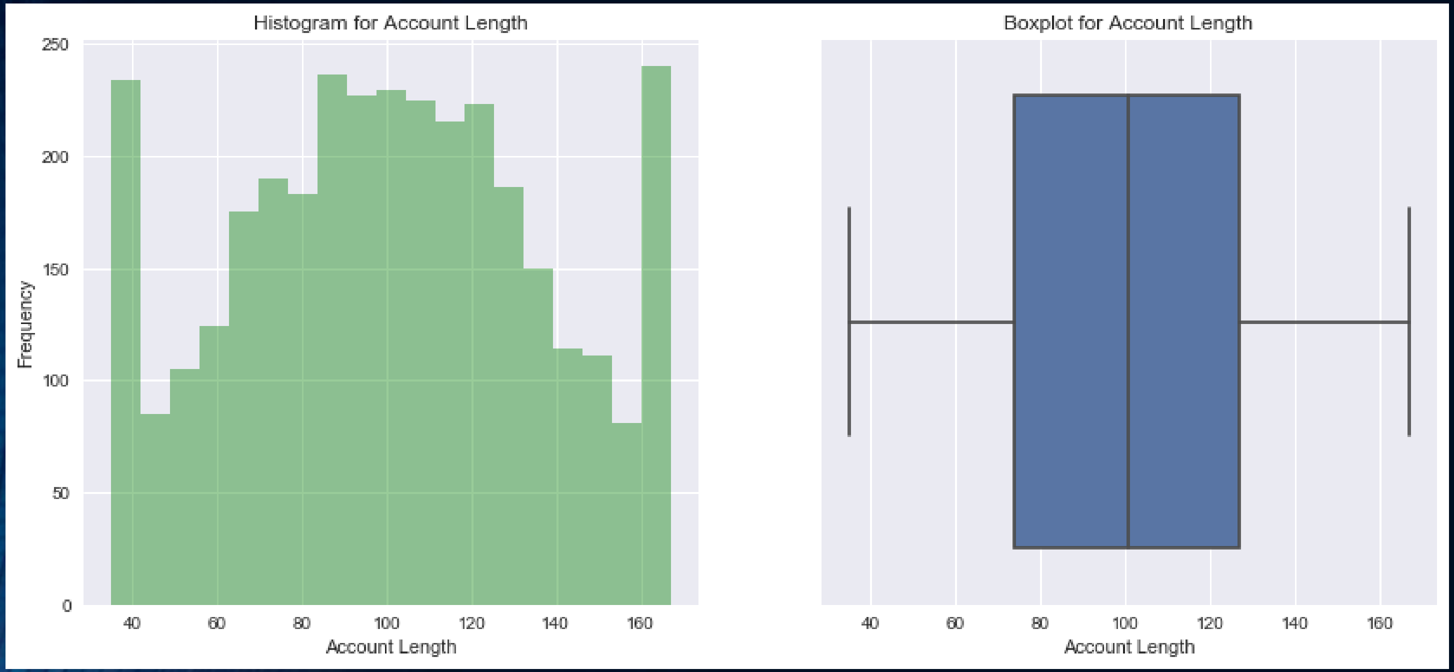
- It is the case of imbalanced data, so majority classes dominate over minority classes causing the machine learning classifiers to be more biased towards majority classes.
- This causes poor classification of minority classes.

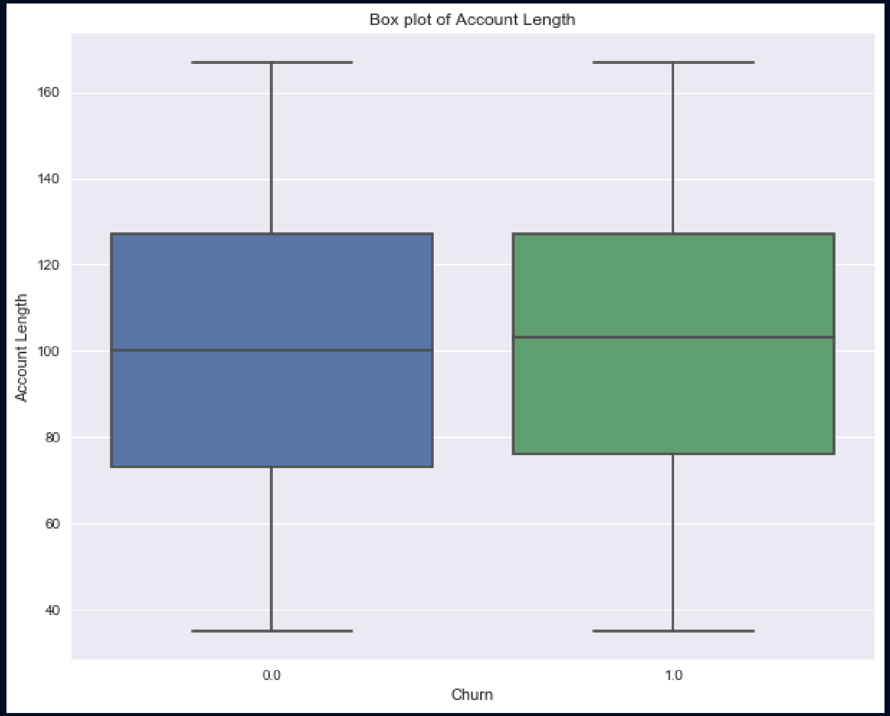


Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

EXPLORATORY DATA ANALYSIS

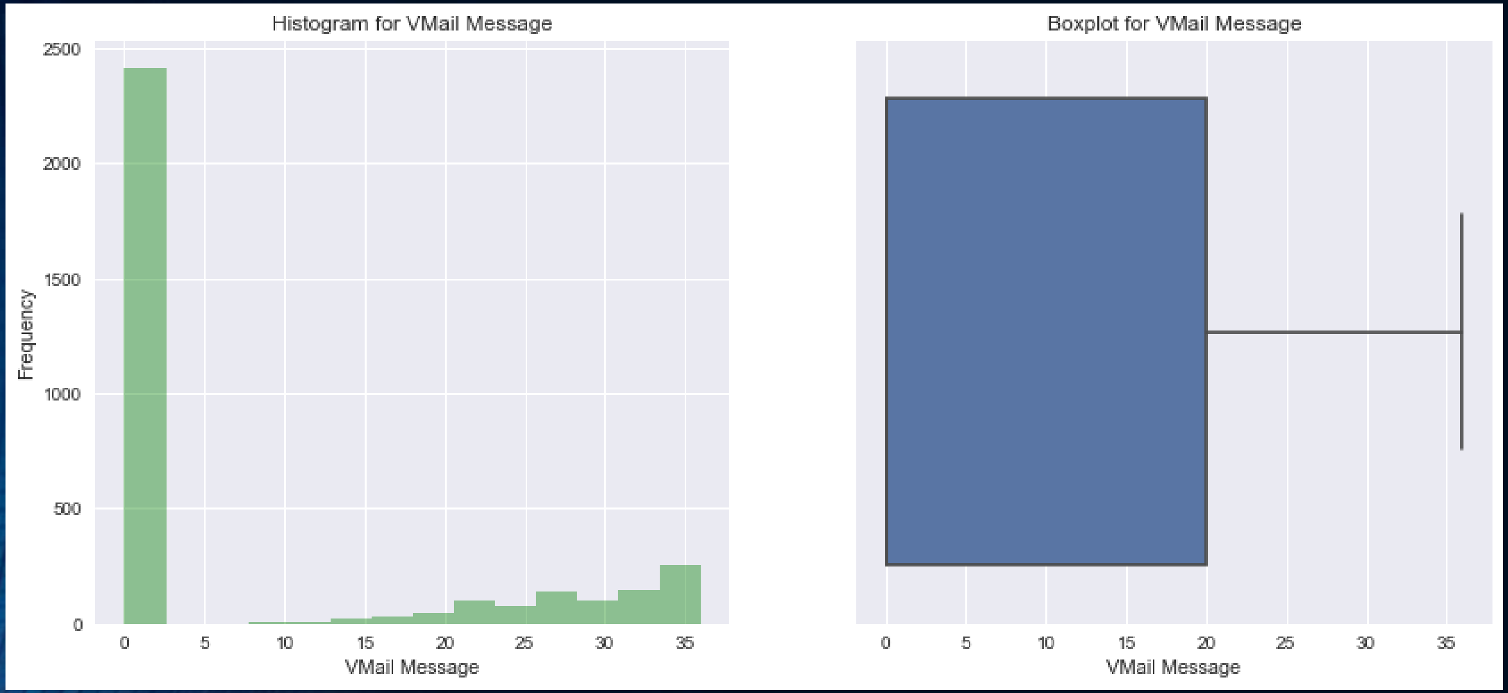
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

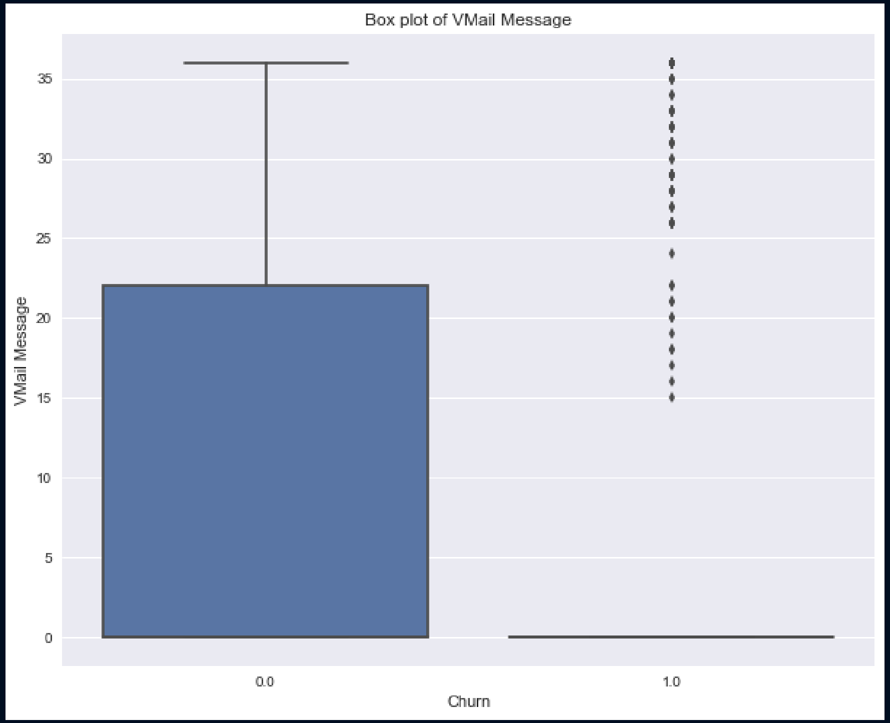




- Customer's Account duration almost follows normal distribution. It shows that the data has a mix of old and new customers.
- Account Length has no effect on the churn rate.

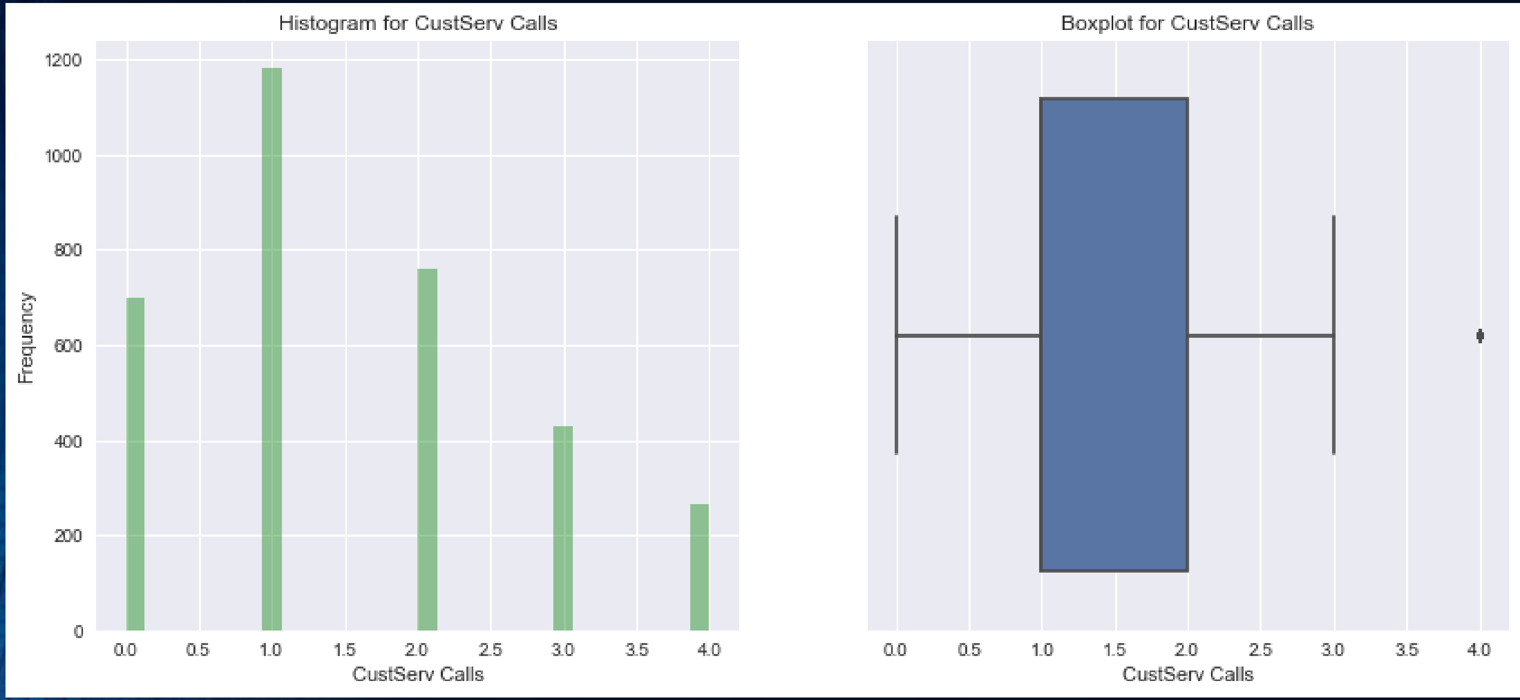
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

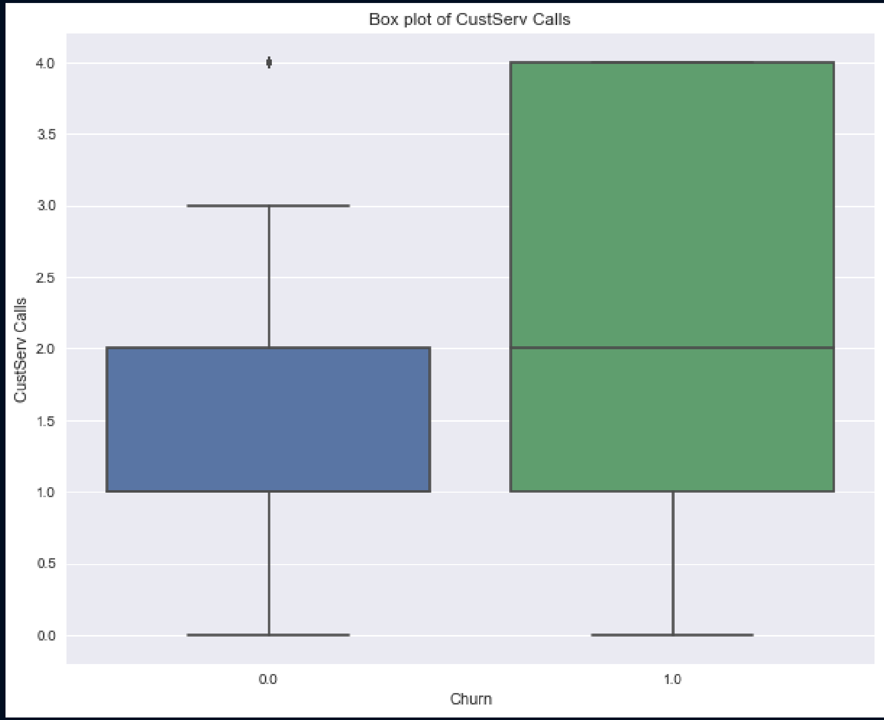




- 2411 customers out of 3333 is not at all using the Vmail Message service.
- Almost 72% of customers are not using this service.

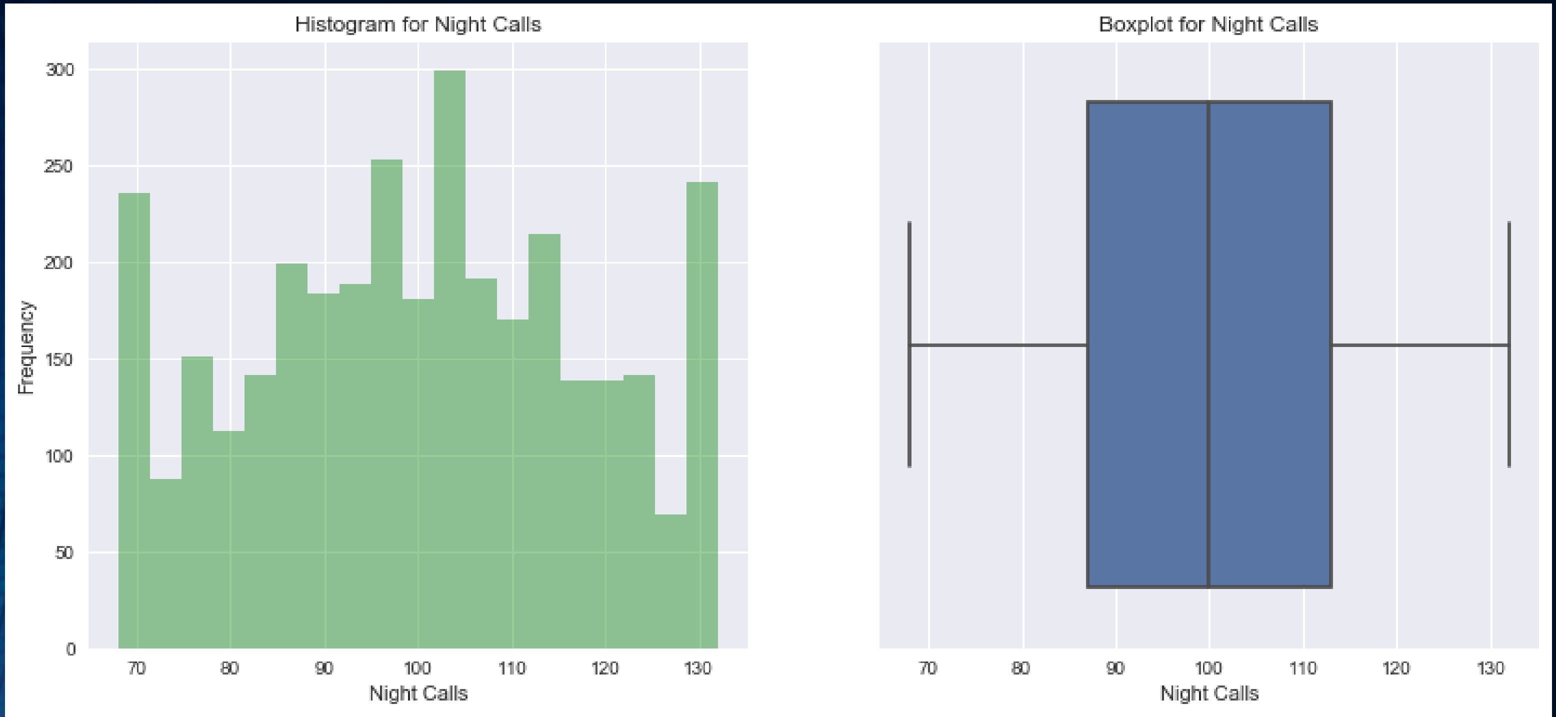
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

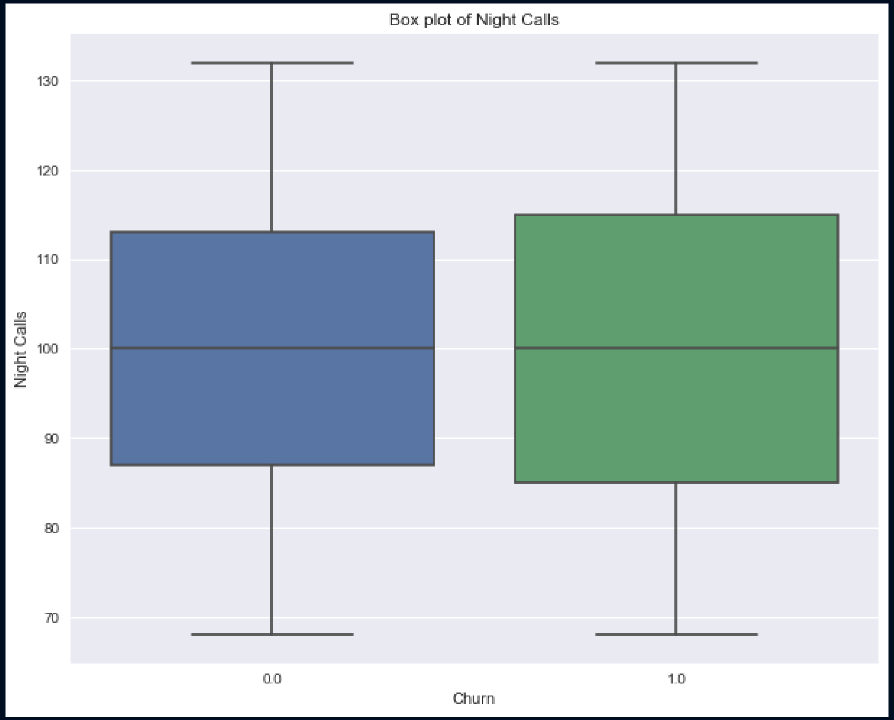




- Majority of customers (1200) has called customer service only once.
- 750 customers has called customer service twice.

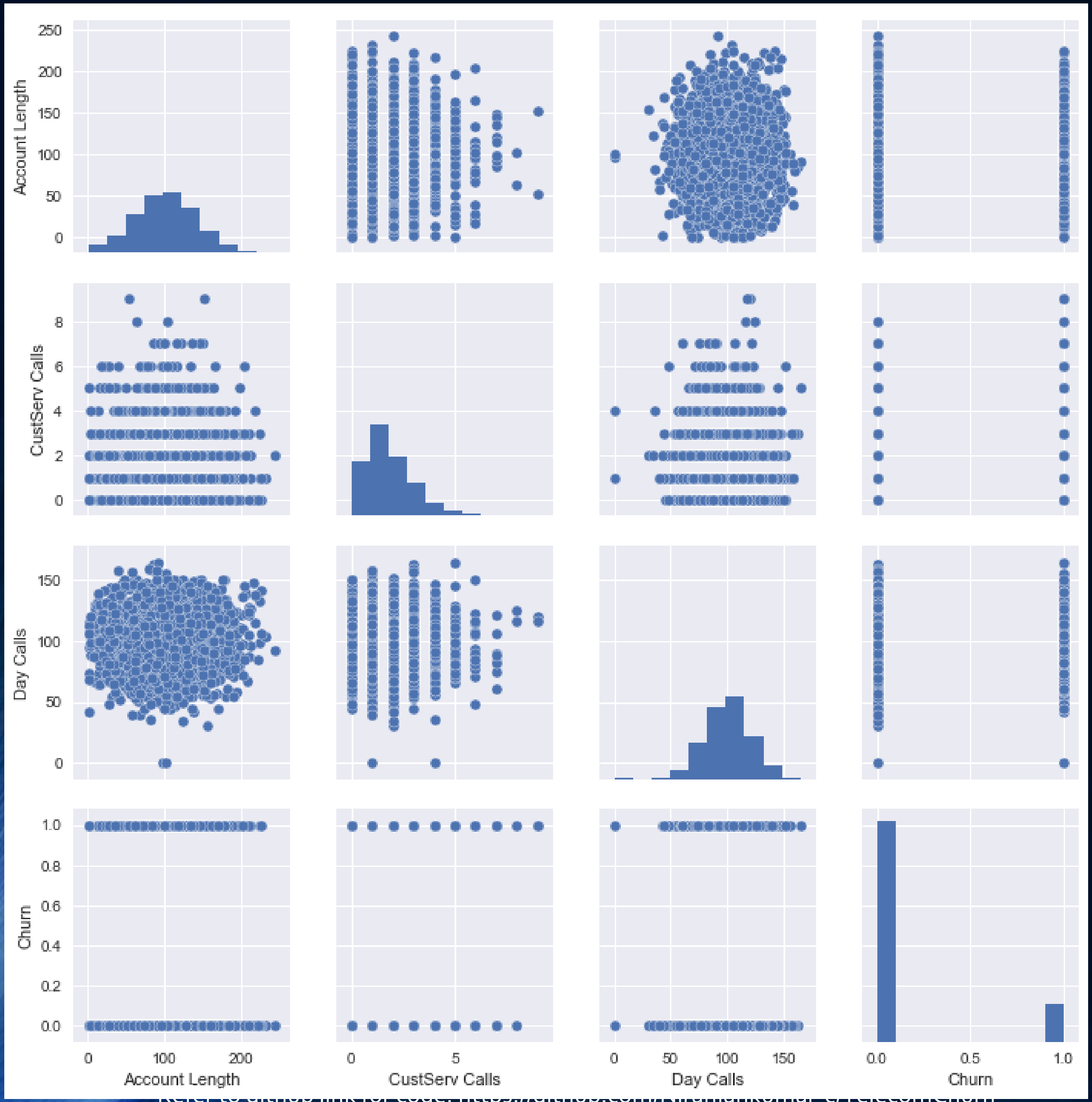
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn





- Customers who are making more calls at Night are leaving the service provider.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn



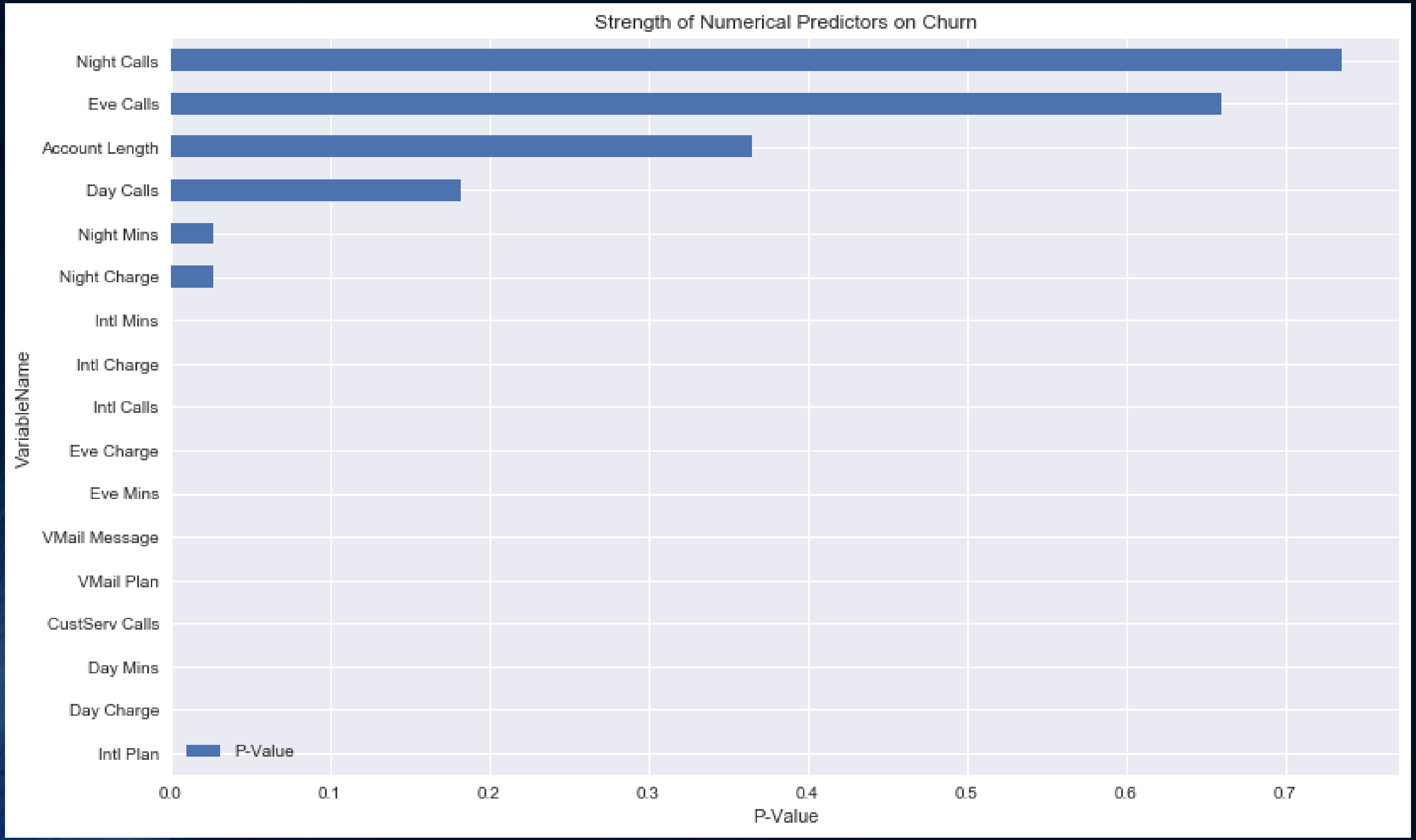
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

There is no linear relationship between the predictors and target variable.

STRENGTH OF PREDICTORS

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

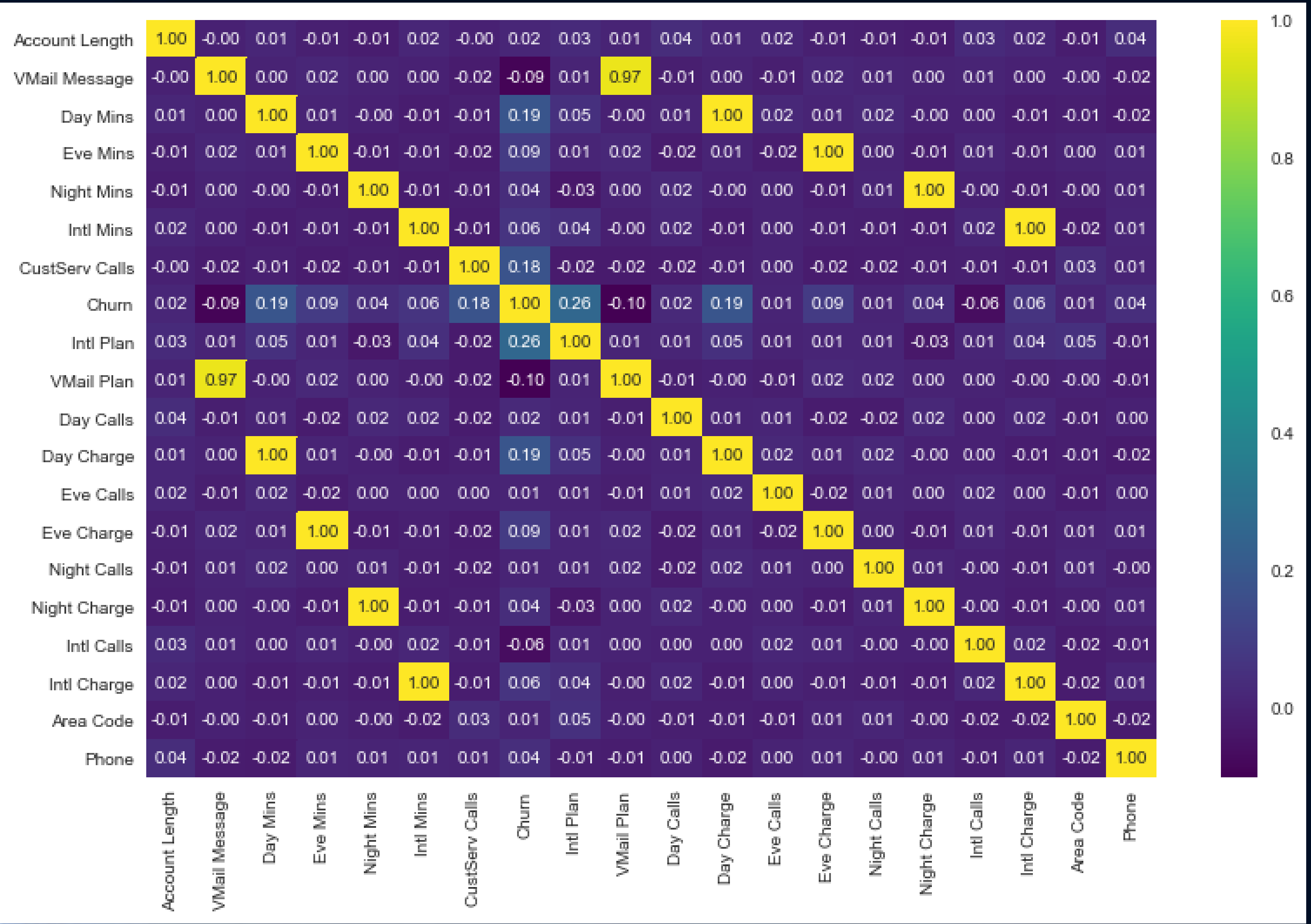
Strength of Predictors



Performed T-Test to find the strength of numerical predictors on churn.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Correlation Plot



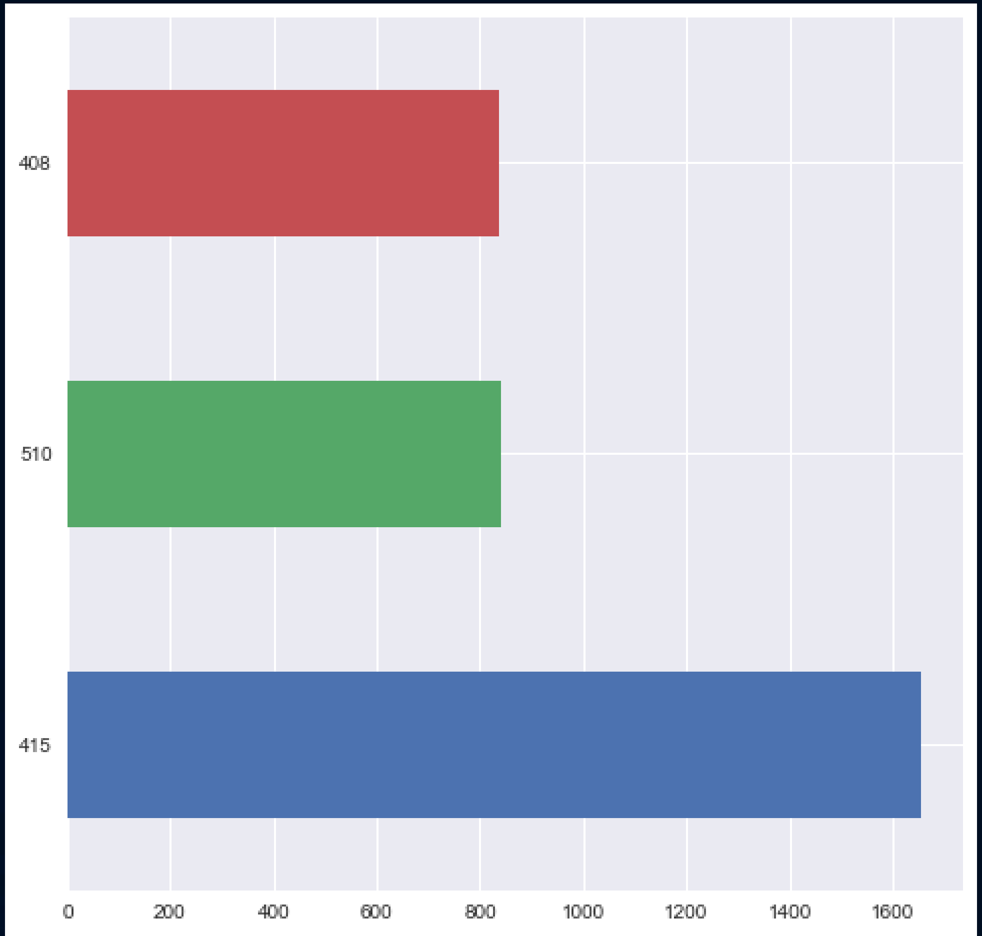
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Direct correlation between

- Day Charge & Day Mins
- Night Charge & Night Mins
- Intl Charge & Intl Mins

Feature Engineering

- This is often the most important step in applied machine learning
- Transformed the "Area Code" variable to three different variables.
- Created three different flag variables to indicate the Area code.



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Model Building

- As there is no linear relationship between dependent and independent variables, and also the number of data points is less, parametric techniques will perform very poor.

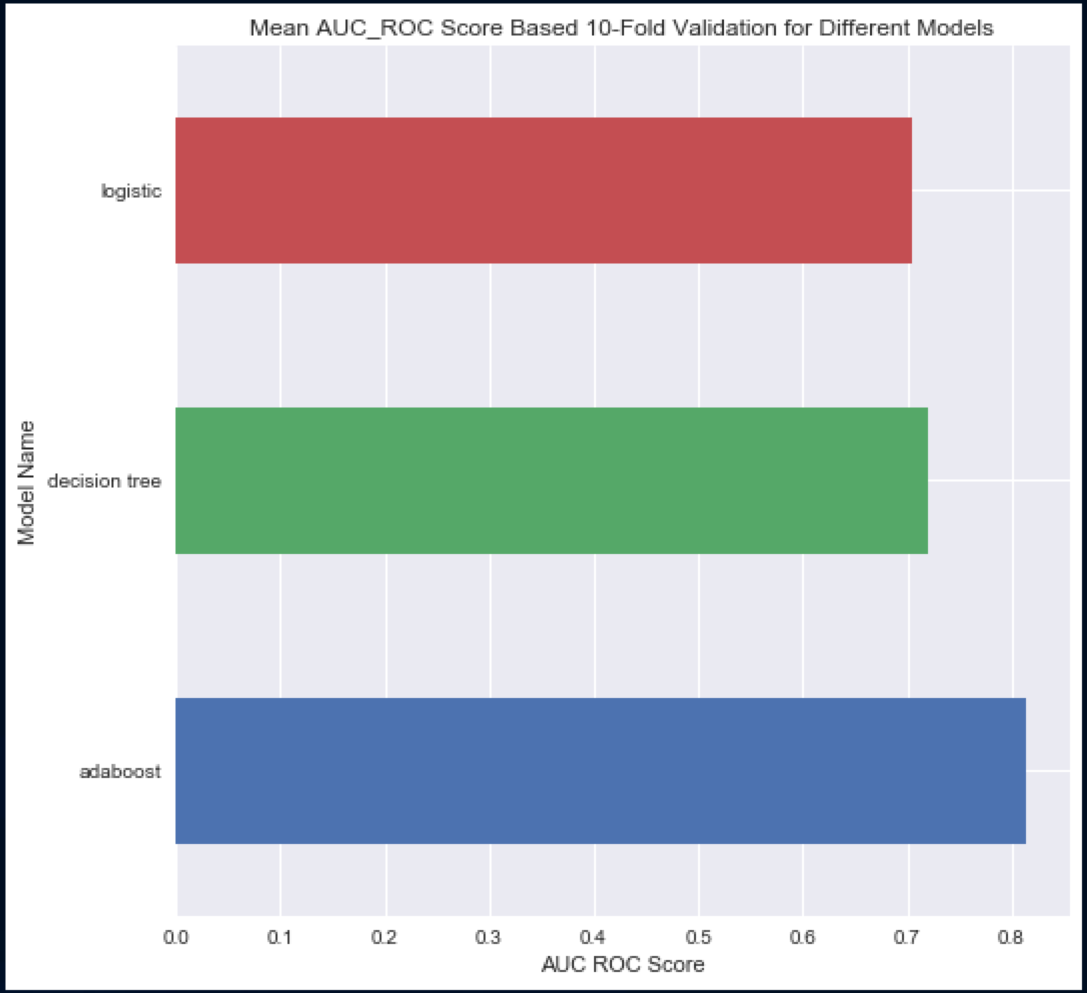
- We have tried 3 different modelling techniques with stratified sampling to reduce the affect of imbalanced dataset.
  - Logistic Regression
  - Decision Trees
  - AdaBoost

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Model Selection

- We have built the logistic regression, decision tree classifier and adaboost classifier and evaluated the models based on AUROC Score.

- Performed 10 Fold Cross Validation to identify how well model performed without any overfitting.

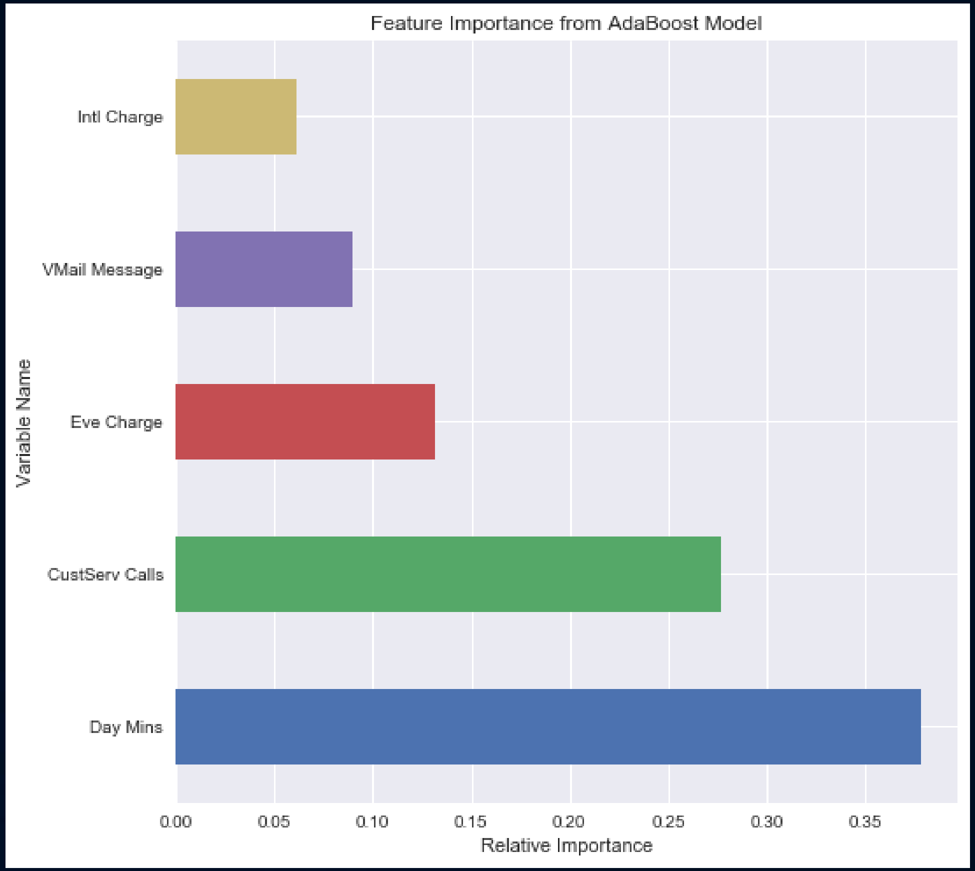


Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Feature Importance

From the feature importance plot, we can see factors for fraud detection

- Day Mins
- Customer Calls
- Evening Call Charges



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Model Validation

- Model Validation is an critical phase in SEEMA Framework before they are deployed for use in the field.

- Model Validation Metrics
  - Confusion Matrix
  - ROC Curve
  - Gain Chart
  - Lift Chart
  - KS Chart

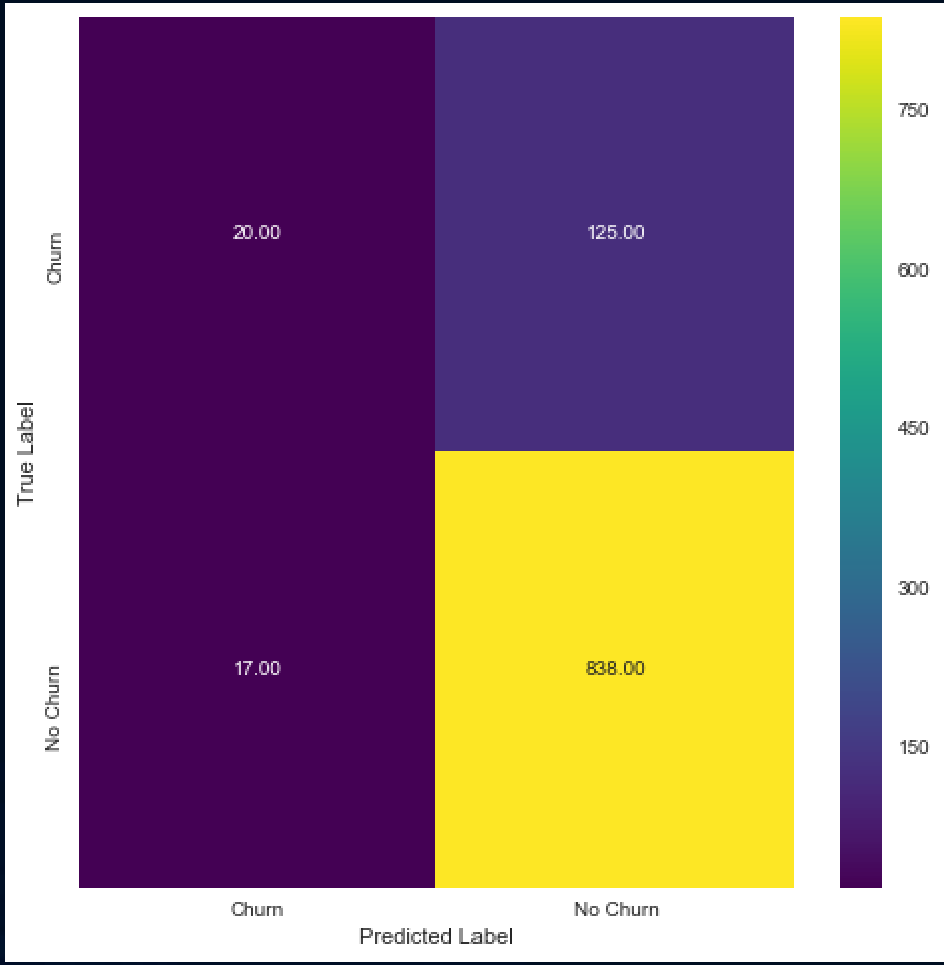
Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Confusion Matrix

- Accuracy may not be the right measure at times, especially if your Target class is not balanced.

- Specificity is high and sensitivity is low, primality driven by threshold value we have choose.

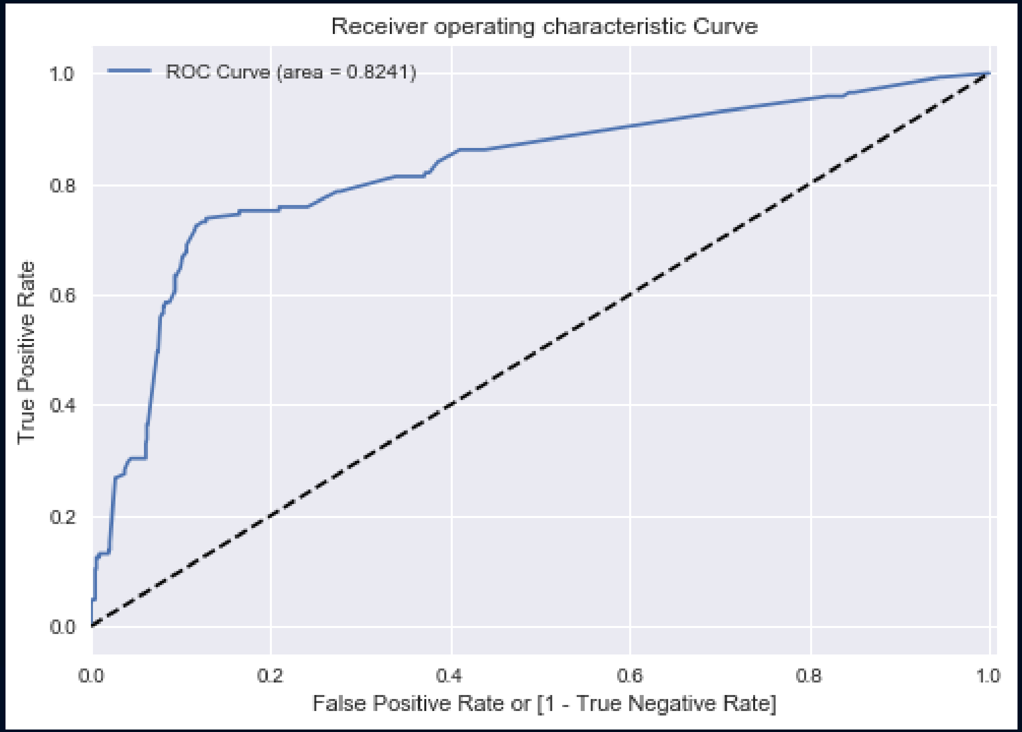
- Precision = 82%
- Recall = 86%



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Receiver Operating Characteristic curve

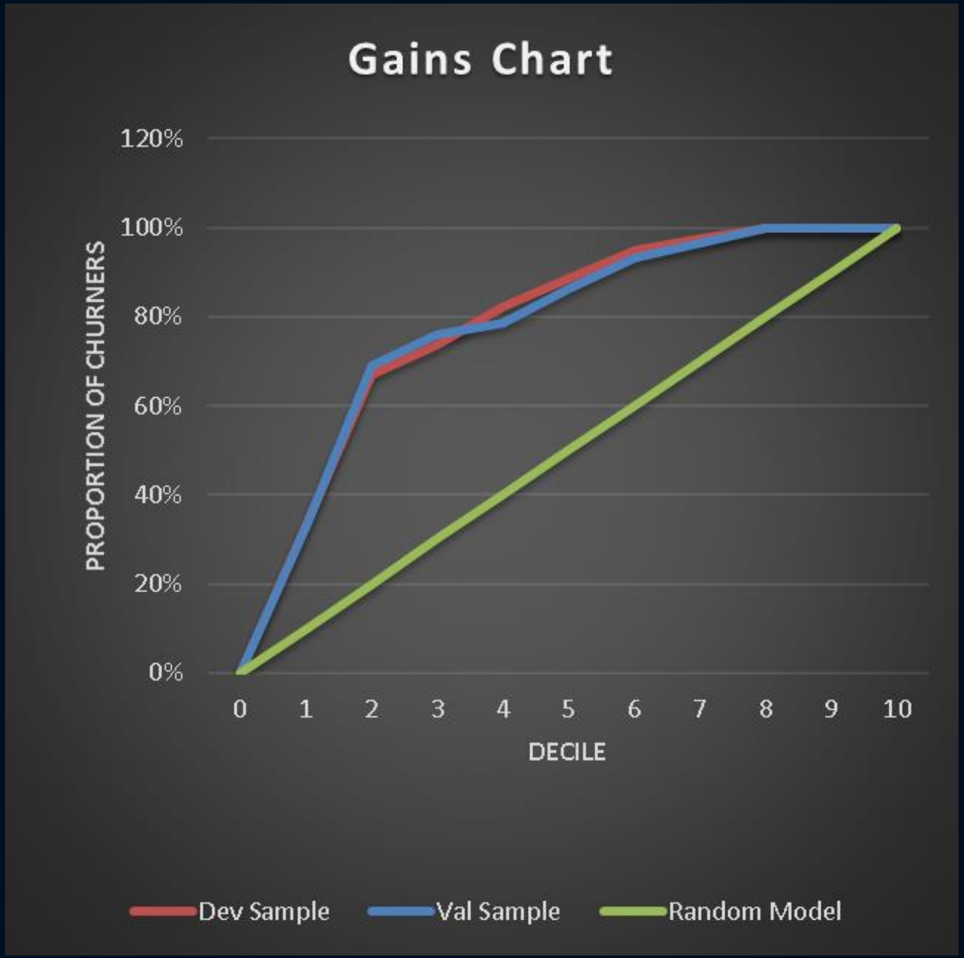
- The plot of 'True Positive Rate' (Sensitivity/Recall) against the 'False Positive Rate' (1-Specificity) at different classification thresholds.
- The area under the ROC curve (AUC) measures the entire two-dimensional area underneath the curve.
- It is a measure of how well a parameter can distinguish between two diagnostic groups.



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Gain Chart

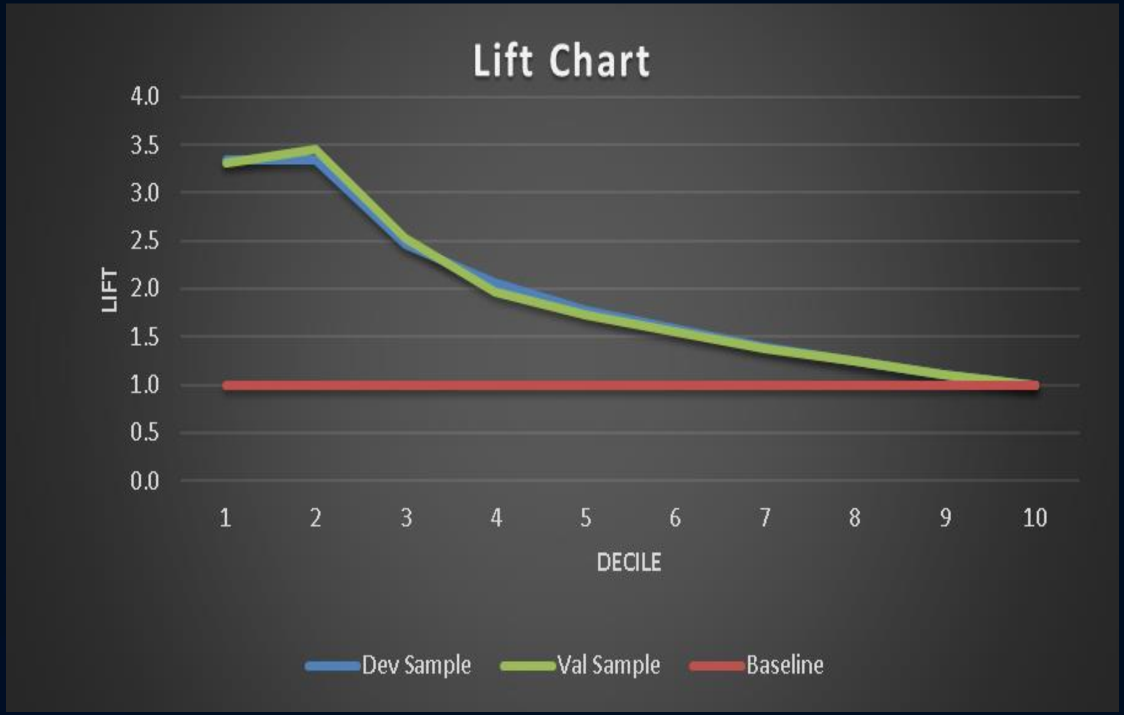
- Gain chart tells you how well is your model segregating responders from non-responders.
- The 'random' line in the corresponds to the case of capturing the responders ('Ones') by random selection.
- By targeting the first 40% of the data, the model will be able to capture 82.2% of the churners.



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Lift Chart

- Lift chart compares the response rates with and without using the classification model.
- It is used to evaluate the usefulness of the model.
- The Cumulative Lift for top two deciles is 3.45.
- we can expect 3.45 times the total number of targets (churners) to be found than by randomly selecting 20% of the data without a model.



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Kolmogorov Smirnov chart

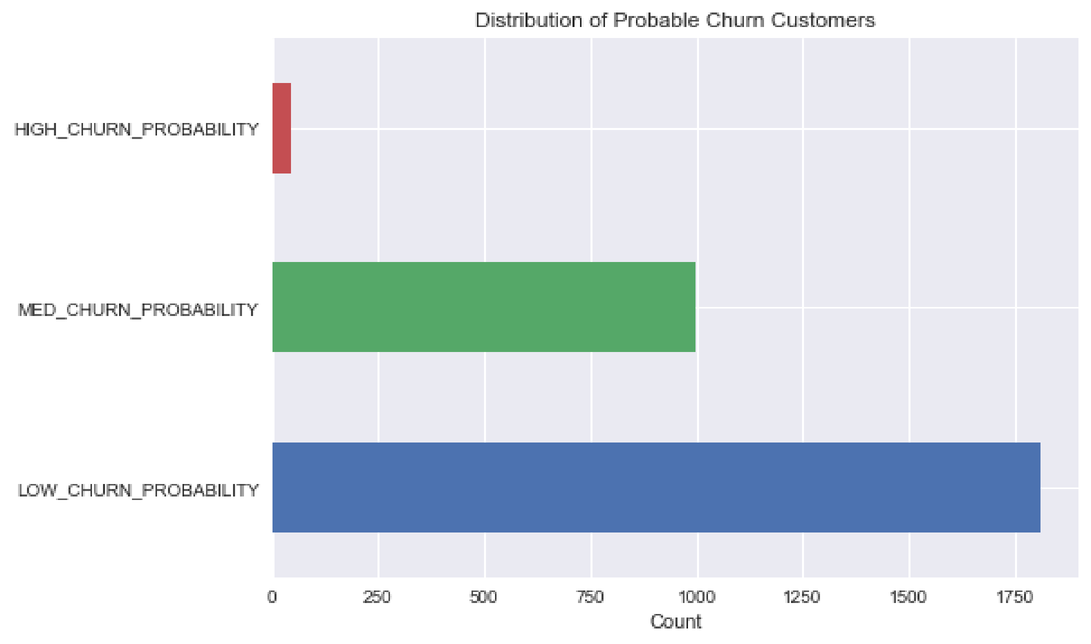
- K-S is a measure of the degree of separation between the positive and negative distributions.
- If the K-S is 100, model is able to separate groups in which one group contains all the positives and the other all the negatives.
- KS is maximum at second decile and KS score is 58.32%.

Test Sample									
Decile	Defaulters	Non Defaulters	Total	Default RATE	Default PERCENTAGE	CUMU. Default PERCENT	Non Default PERCENT	CUMU. Non Default PERCENT	KS
1	48	52	100	48.00%	33.10%	33.10%	6.08%	6.08%	27.02%
2	52	39	91	57.14%	35.86%	68.97%	4.56%	10.64%	58.32%
3	10	99	109	9.17%	6.90%	75.86%	11.58%	22.22%	53.64%
4	4	47	51	7.84%	2.76%	78.62%	5.50%	27.72%	50.90%
5	11	138	149	7.38%	7.59%	86.21%	16.14%	43.86%	42.35%
6	10	225	235	4.26%	6.90%	93.10%	26.32%	70.18%	22.93%
7	5	126	131	3.82%	3.45%	96.55%	14.74%	84.91%	11.64%
8	5	129	134	3.73%	3.45%	100.00%	15.09%	100.00%	0.00%
9	0	0	0	#DIV/0!	0.00%	100.00%	0.00%	100.00%	0.00%
10	0	0	0	#DIV/0!	0.00%	100.00%	0.00%	100.00%	0.00%
	145	**855**	**1,000**					KS	58.32%

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Actionable Business Insights

- Create Segments of customers based on probability of churn.
  - 0 – 0.4: Low churn probability
  - >0.4 – 0.5: Medium churn probability
  - >0.5: High churn probability



Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Actionable Business Insights

- For customers with High Churn probabilities will require an immediate call to understand their grievances/complains.
- May be a new relationship manager will be helpful to understand their choices in terms of payment, language and complaints.

- For customers with Medium Churn probabilities may require souvenirs for their time with the service provider.
- Keep in touch with call and do monitor their grievances and complains for better contribution

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

Actionable Business Insights

- By targeting the customers who are present in the 40% of the data, we will be able to prevent 82% of churners.

- Improve the customer service experience, so that the number of calls to the customer care will be reduced. May be introduce 24/7 chat bot to deal with customer complaints without the hassle of calling.

Refer to github link for code: https://github.com/Niranjankumar-c/TelecomChurn

THANK YOU