Customer churn analysis on SyriaTel company

- 23.4.2023
- Amos kipkirui

Overview

- SyriaTel prides itself in provision of a wide range of services, including mobile, landline, and internet connectivity, to its customers. However, the company faces the menace of customer churn which is a problem that is witnessed across several other telecommunication companies.
- Customer churn is where the customers stop or discontinue using the services of the company.
  We tried to detect whether there is a noticeable pattern for this, study the customer behavior and key factors that influences churn.
- Customer churn not only impacts the revenue and the market share of the company but also it tests the customer loyalty to the service or the product that a company offers them.
  We utilized data and machine learning algorithms to detect and prevent churn.

Business understanding

- To effectively address the issue of customer churn, we ought to have a good understanding of our business context;
- **Business context**
- SyriaTel is one of the top mobile network providers in Syria. The company offers services including calls, news, message, GSM and internet services. The industry currently experiences stiff competition and therefore retaining customers is crucial for sustaining company's revenue growth. Doing customer churn analysis will help us identify features that are contributing to customer churn. This will enable the implementation of measures that will help reduce the customers stoppage from using company's services.
- The stakeholder audience here include; Data analytics team, sales team, senior management and the marketing executives.

- **Objectives**
- Includes:
- 1. To develop a machine learning model that can accurately predict customer churn based on historical customer data and behavioral patterns
- 2. To identify key features that significantly contribute to customer churn. This will enable for drawing of insights into the underlying causes.
- 3. To Achieve a high level of accuracy and recall in the churn prediction model to minimize both false negatives (customers who churn but were not predicted) and false positives (customers predicted to churn but do not).
- 4. To Utilize the prediction model to come up with mitigation measures to help curb the customer churn problem.
- 5. To come up with recommendations to the company on how to retain their customers as this will help maintain the company's market share.

Data understanding

- SyriaTel company is rich in customer data which will be useful in the customer churn analysis.
- The data contains customer demographics, billing information and service usage records
- We are going to analyze this dataset to help us uncover patterns, correlations, and trends that can help us identify the key factors influencing churn
- Key data variables that we are going to analyze here include; the state where the customer resides, the customers area code, international plan, voice mail plan, billing information which includes;

Data understanding continuation

total day charge, total eve charge, total night charge and total international charge.

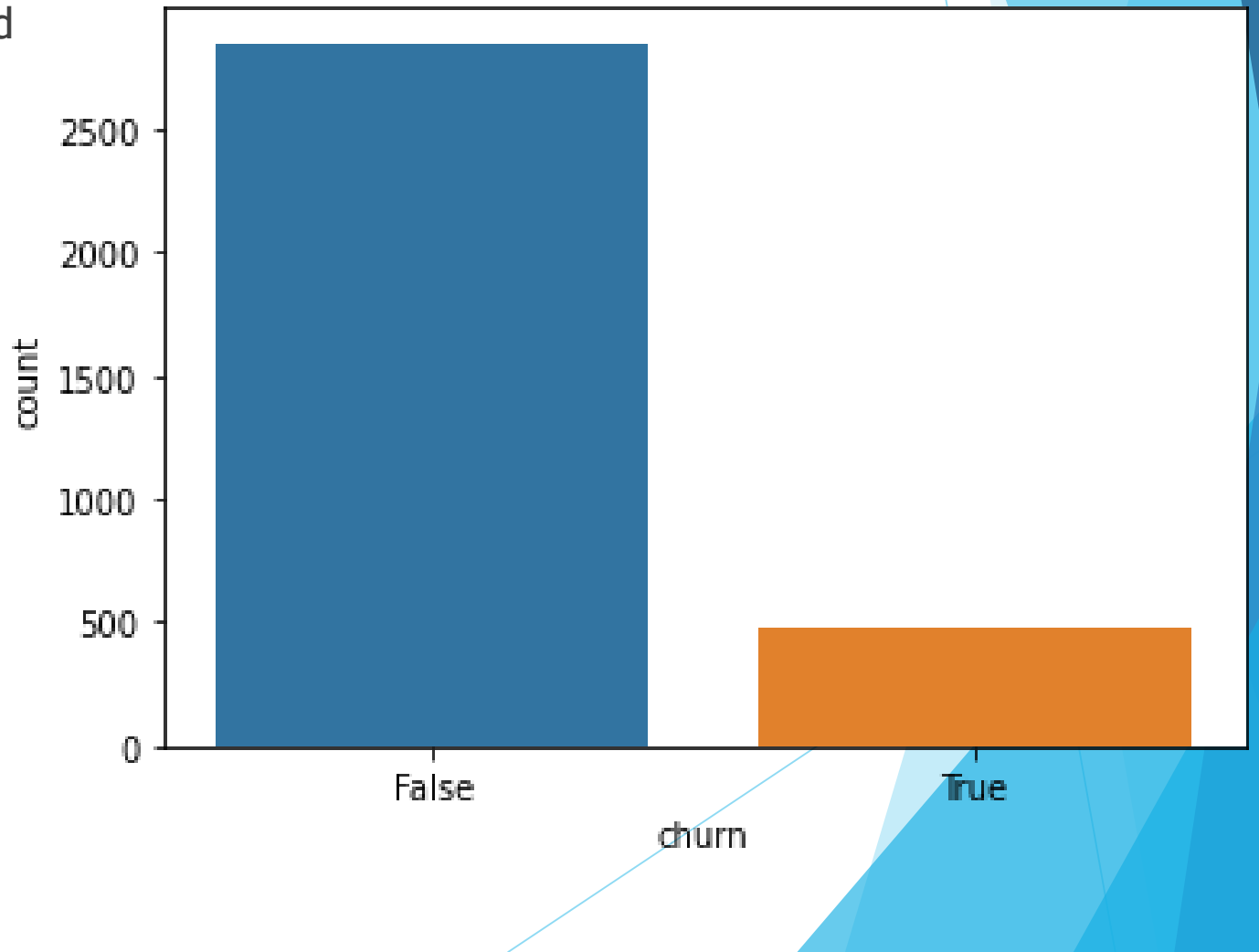
We will also look at customer service calls.

Exploratory data analysis

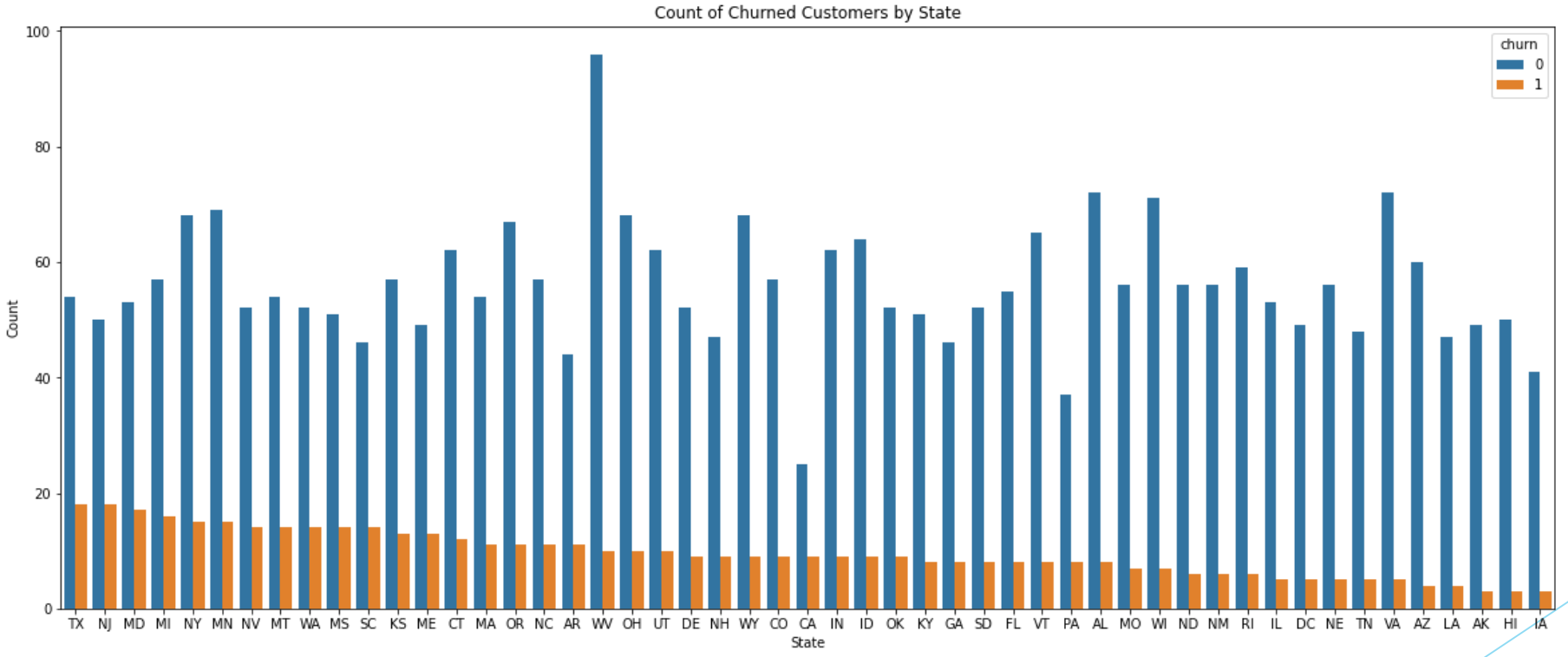
- We performed an exploratory data analysis on the preprocessed data making use of the key variables indicated on the previous slides.
- We will look at their relationships with churn.

Customers distribution based on whether they churn or not.

- There are more customers who did not churned as compared to those who did churn.



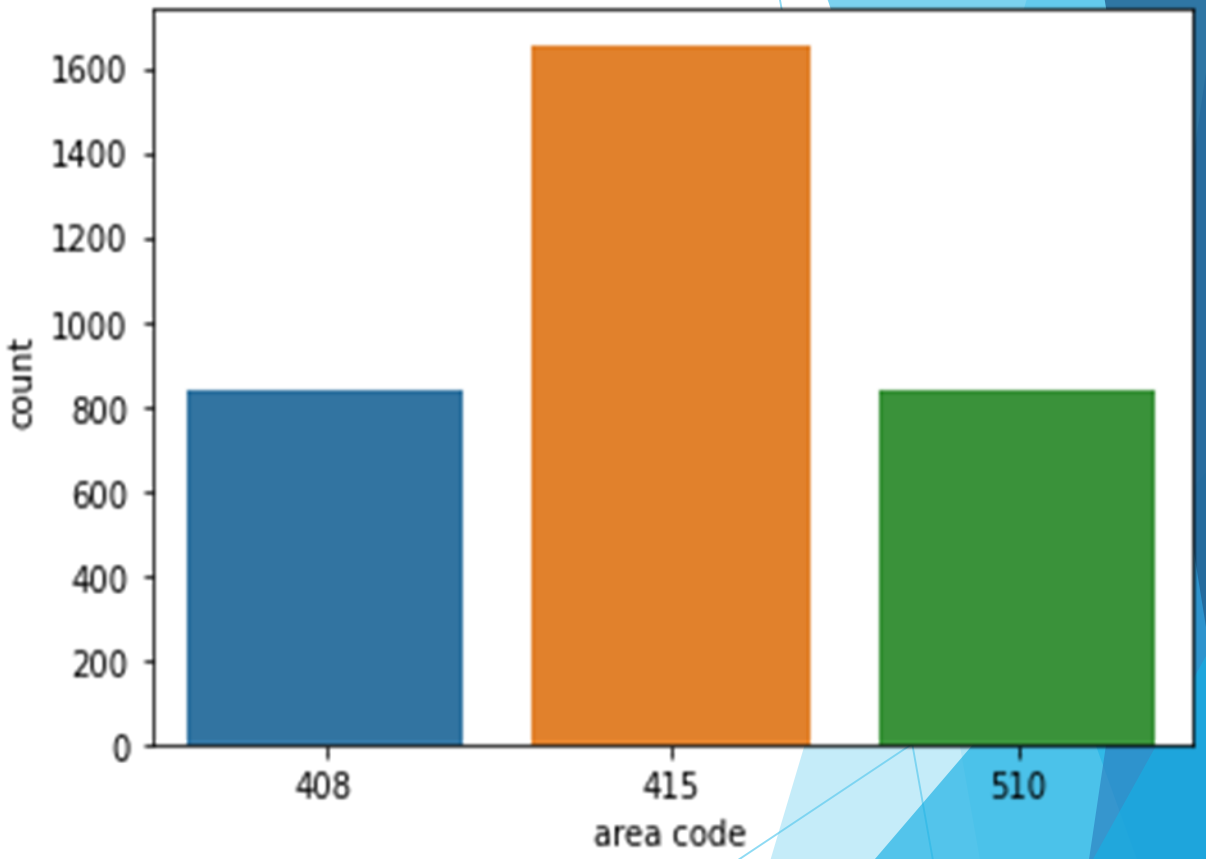
Relationship between the state in which the customer resides with customer churn



- MD, TX, NJ. MI and NY have the highest number of churned customers. These states stand out with significantly higher bar heights compared to other states.

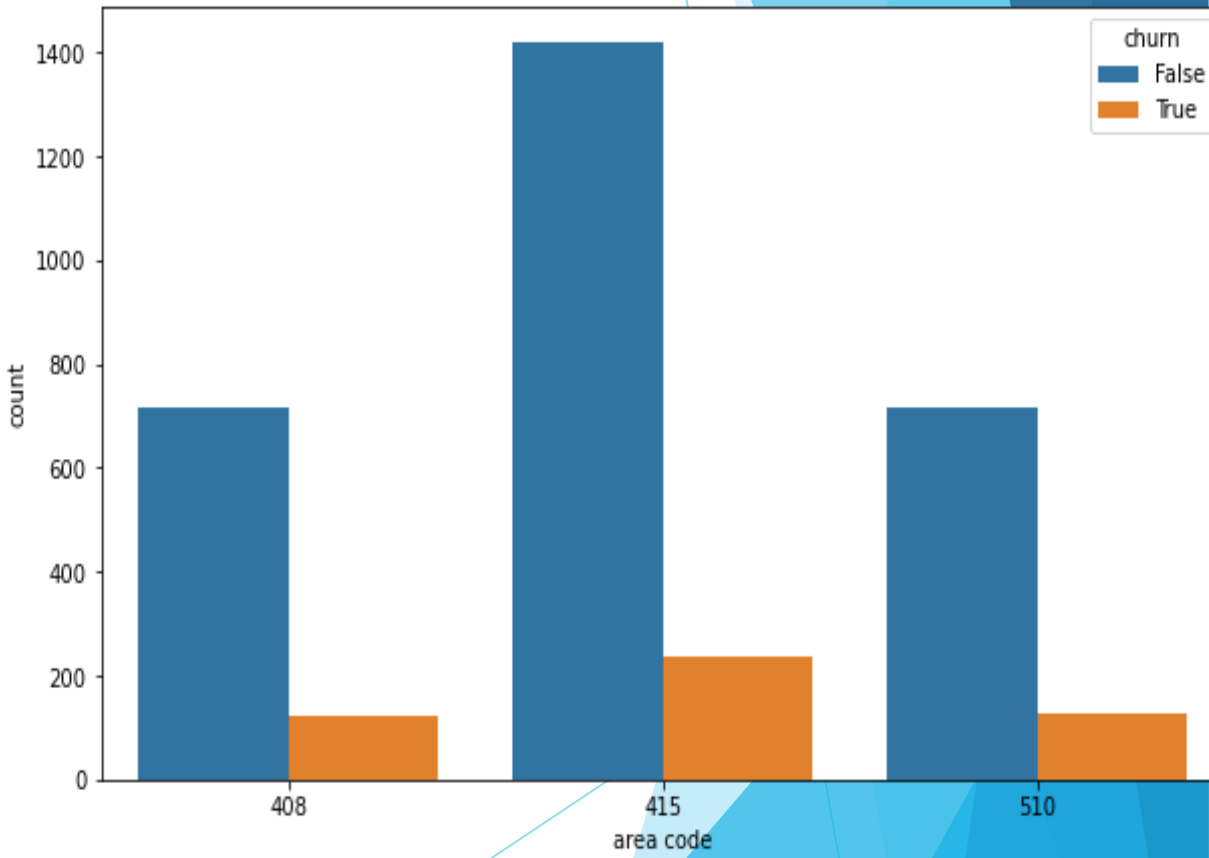
Relationship between the customers area code with customer churn

- Relationship between the customers area code with customer churn
- From the result, we can see that there are 1655 customers from area code 415. There are around a quarter of the total customers from area codes 510 and 408.



What is the relationship between area codes and customer churn

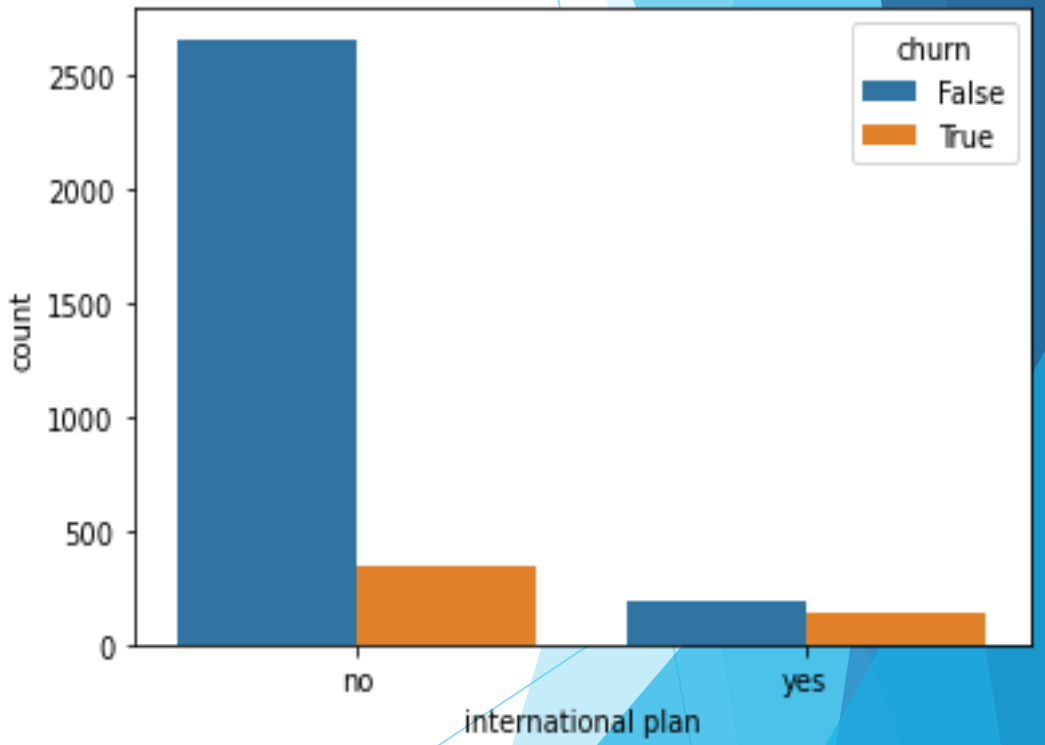
- The combination of a higher churn rate and a larger customer base in the 415 area code should raise concerns for the company. It indicates that there may be specific factors or issues within this area code that contribute to a higher likelihood of churn. Customers from area codes 408 and 510 seems more satisfied with the services of the company.



Next, is the analysis on international plan

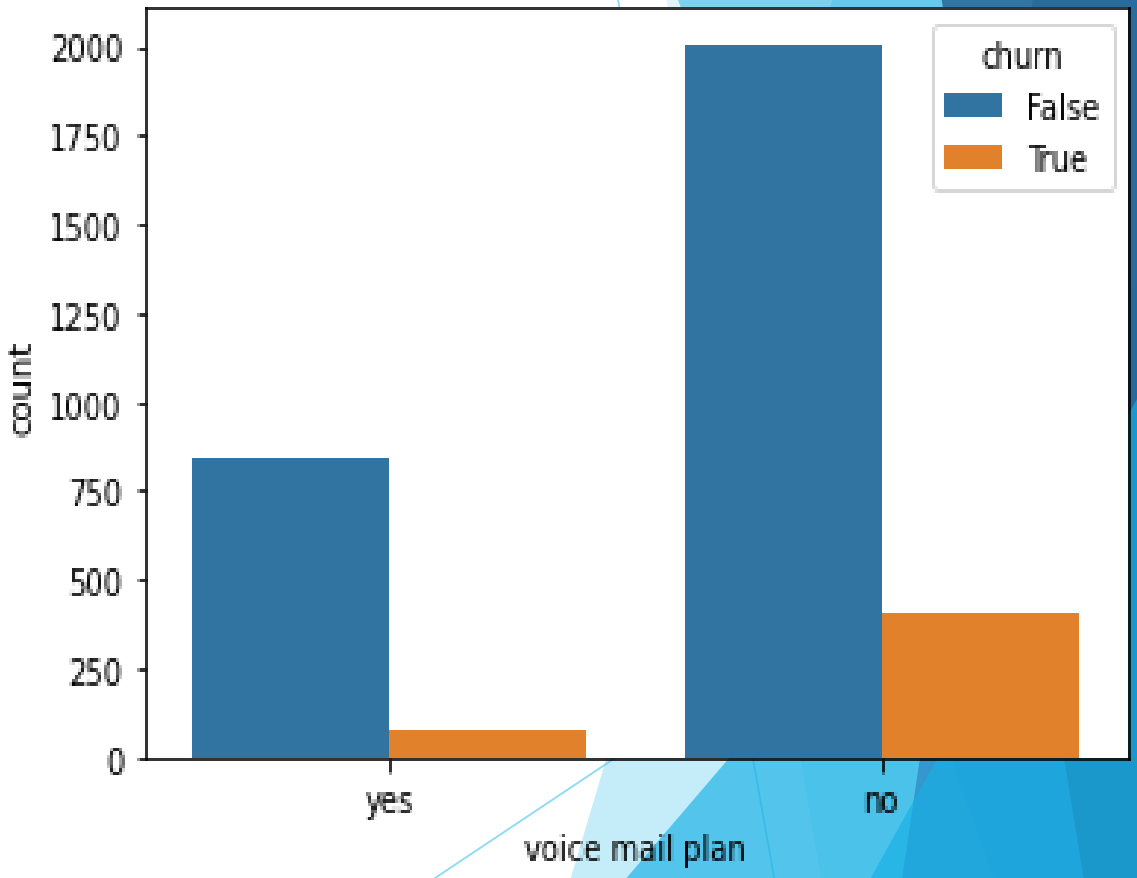
- Customers without an international plan have a higher churn rate compared to those with an international plan. The bar representing customers without an international plan is noticeably higher, indicating a larger proportion of churned customers in this group.

- This suggests that having an international plan may contribute to higher customer loyalty or retention. Customers with an international plan may find value in the services or benefits provided, leading to a lower likelihood of churning.



Analysis on voice mail plan relationship with churn

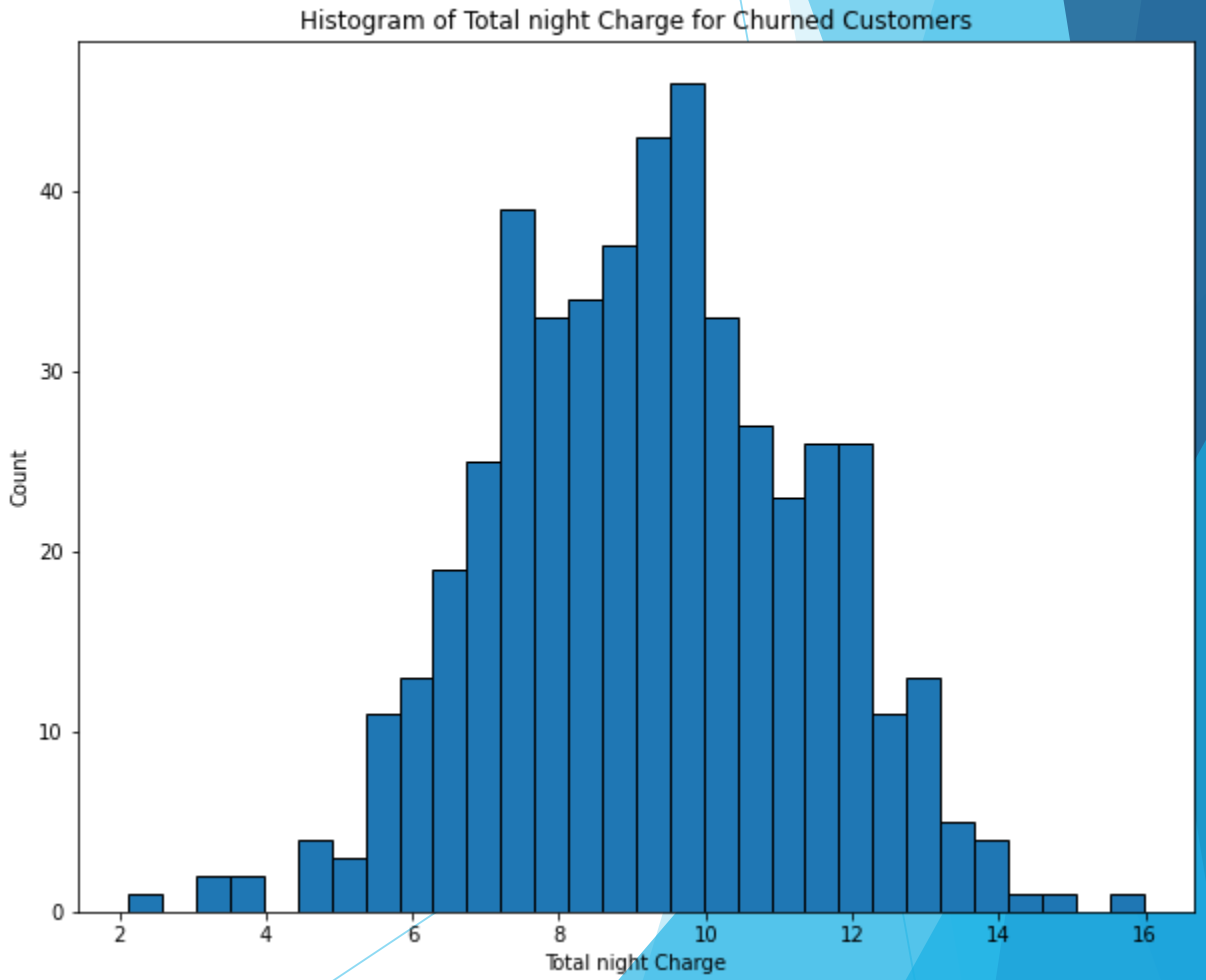
- From the analysis, we can see that a lot of customers are not subscribed to voice mail plan.
- Customers without a voice mail plan have a higher churn rate compared to those with a voice mail plan. The bar representing customers without a voice mail plan is noticeably higher, indicating a larger proportion of churned customers in this group.
- This suggests that the presence or absence of a voice mail plan can be a contributing factor to customer churn. Customers without a voice mail plan may perceive less value in the service or may have alternative communication preferences, leading to a higher likelihood of churning.



Relationship of churn with charges that the customers are subjected for the services they are given.

1. Churn versus total night charge

- The histogram is normally distributed, it implies that the charges during the night for churned customers follow a predictable pattern. This predictability can be valuable in understanding customer behavior and making informed decisions.



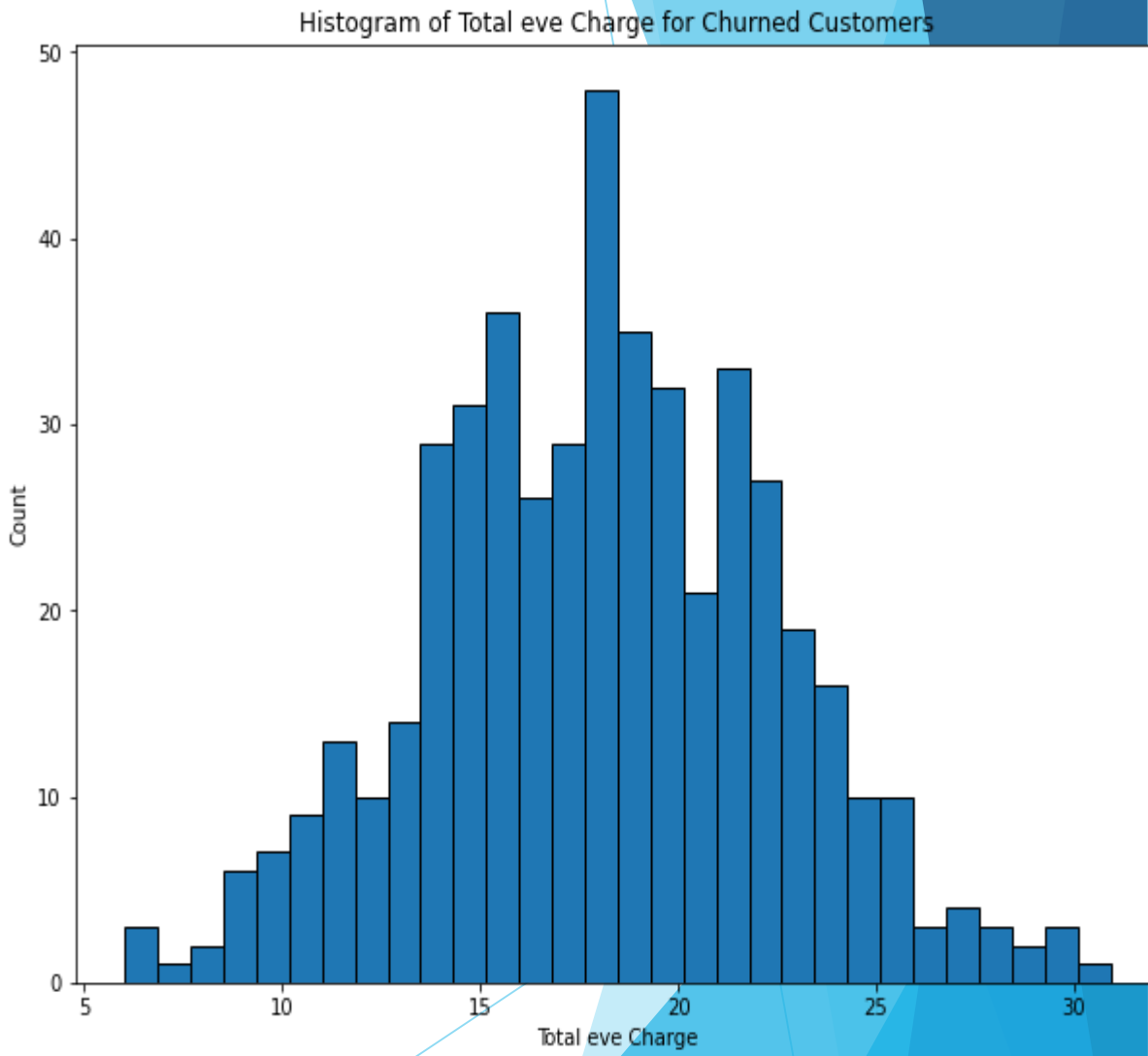
2. Churn versus total night charge

- The distribution of total evening charges for churned customers appears to be approximately normal. The histogram follows a bell-shaped curve, indicating a relatively balanced distribution of charges.

- The majority of churned customers had total evening charges in the range of approximately $10 to $20. The histogram bars are highest in this range, suggesting that a significant number of churned customers had evening charges within this range.

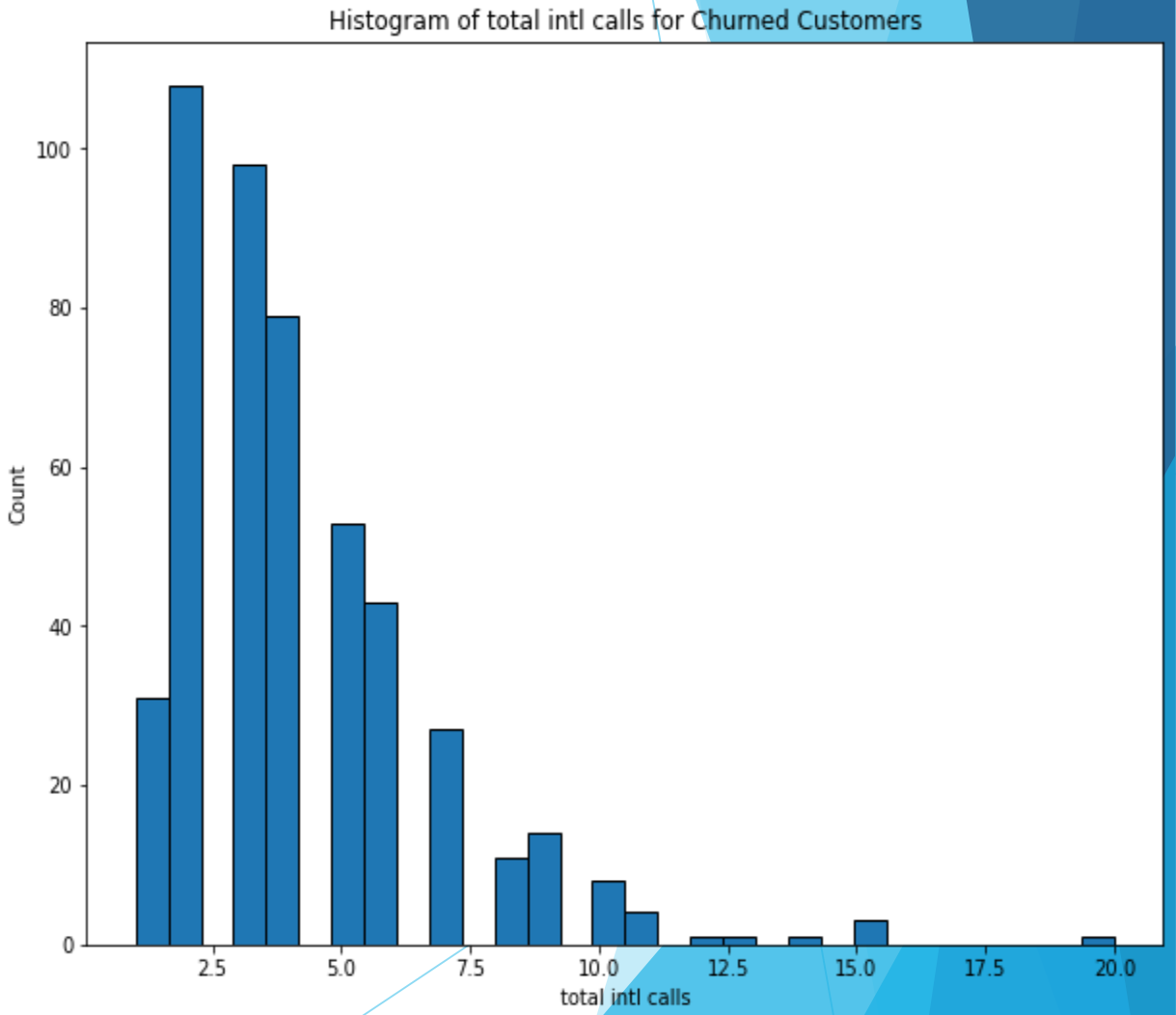
- There is a noticeable decrease in the count of churned customers as the total evening charges increase beyond $20. The bars gradually decrease in height, indicating that fewer churned customers had higher evening charges.

- The peak of the histogram appears to be around $15, indicating that a significant number of churned customers had total evening charges close to this value.



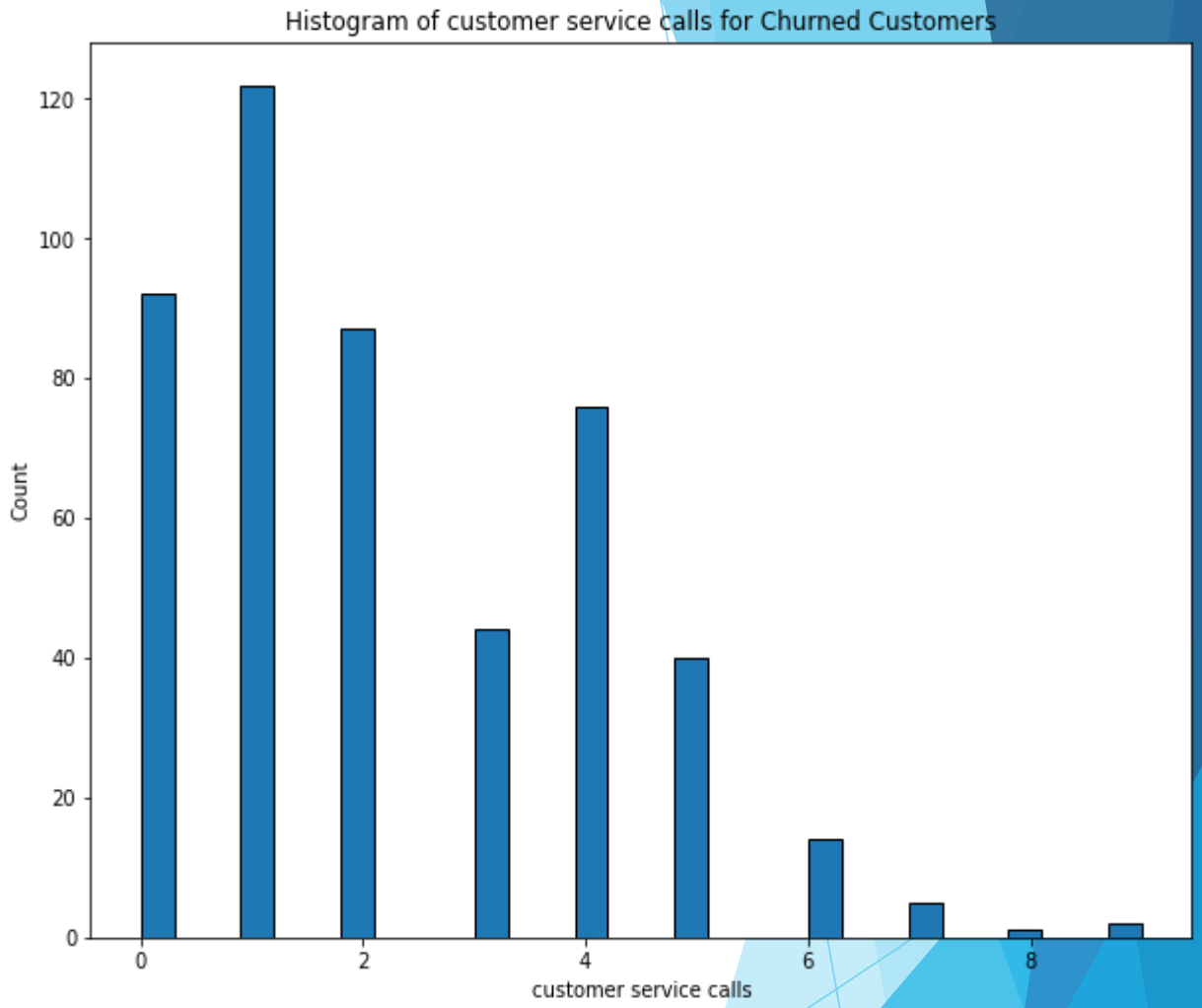
Relationship between total international calls with churn

- The majority of churned customers made a relatively low number of international calls. The histogram bars are concentrated towards the left side of the x-axis, indicating that most churned customers had minimal international call activity.
- The peak of the histogram appears to be around 2-4 international calls. This indicates that the majority of churned customers had either minimal international call activity or made a few calls before deciding to churn.
- The number of international calls does not seem to be a significant contributing factor to customer churn. Most churned customers had minimal international call activity, and even those who made a higher number of calls represent a smaller portion of the overall churned customer population.



Customer service calls relationship with churn

- Most churned customers have made a relatively low number of customer service calls. The majority of the bars in the histogram are concentrated towards the lower end of the x-axis, indicating that a significant portion of churned customers had fewer interactions with customer service.



Modeling for Customer Churn Prediction

Introduction

- Customer churn is among the challenges that are faced by companies in day to day business operations. It directly impact the business revenues and its growth.
- We will discuss on the different models that were used in the prediction of the customer churn and to provide effective measures to mitigate this challenge.

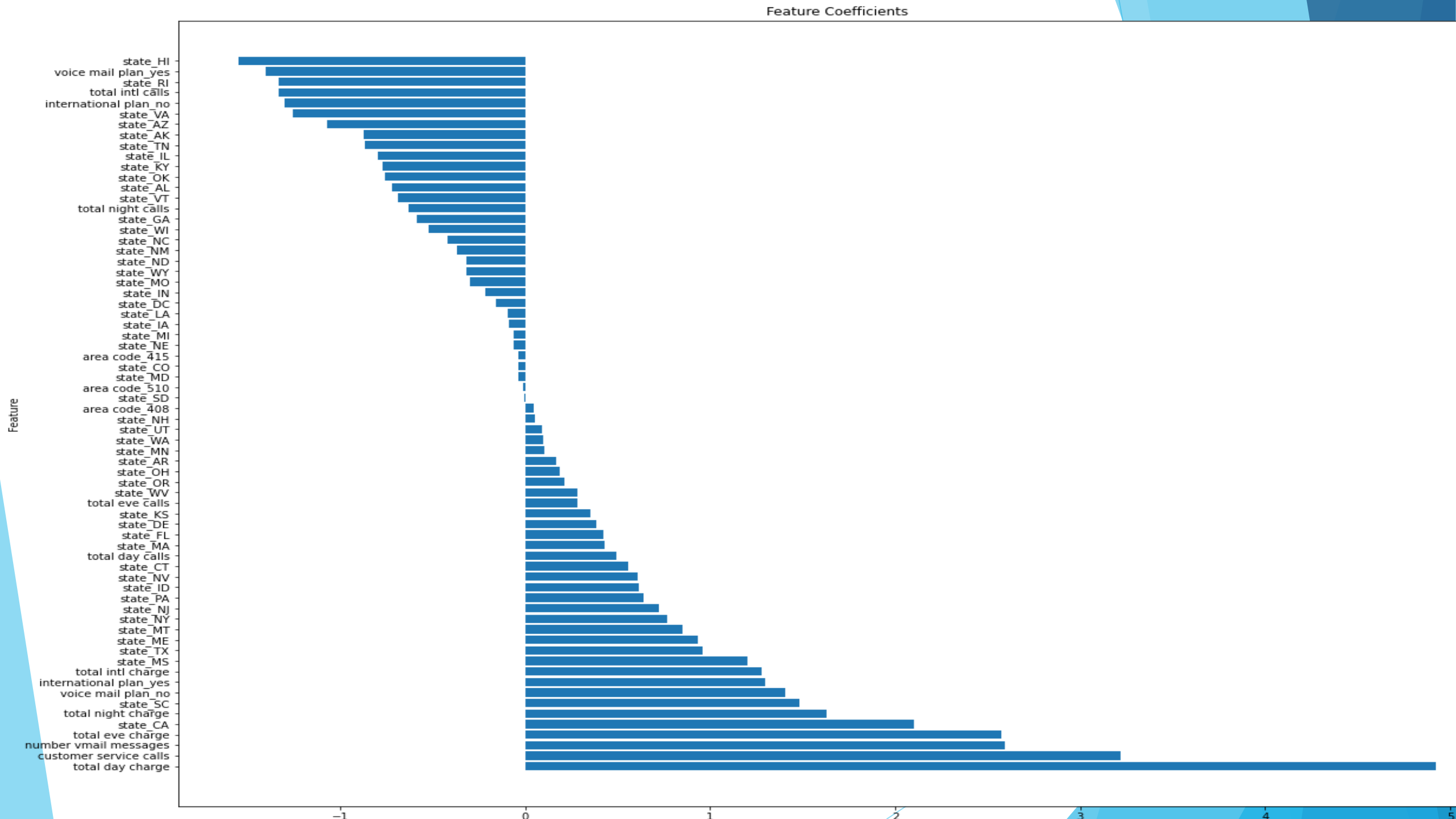
Logistic Regression - Baseline Model

- Logistic regression was used as the baseline model since this is a simple and easy to interpret model.
- It had the following accuracies;

a) Accuracy on training data of about 79%

b) Accuracy on test data of about 76%

Also we looked at the important features in the model and they were as follows;


Feature Coefficients
Feature
state_HI
voice mail plan_yes
state_RI
total intl calls
international plan_no
state_VA
state_AZ
state_AK
state_TN
state_IL
state_KY
state_OK
state_AL
state_VT
total night calls
state_GA
state_WI
state_NC
state_NM
state_ND
state_WY
state_MO
state_IN
state_DC
state_LA
state_IA
state_MI
state_NE
area code_415
state_CO
state_MD
area code_510
state_SD
area code_408
state_NH
state_UT
state_WA
state_MN
state_AR
state_OH
state_OR
state_WV
total eve calls
state_KS
state_DE
state_FL
state_MA
total day calls
state_CT
state_NV
state_ID
state_PA
state_NJ
state_NY
state_MT
state_ME
state_TX
state_MS
total intl charge
international plan_yes
voice mail plan_no
state_SC
total night charge
state_CA
total eve charge
number vmail messages
customer service calls
total day charge
−1
0
1
2
3
4
5

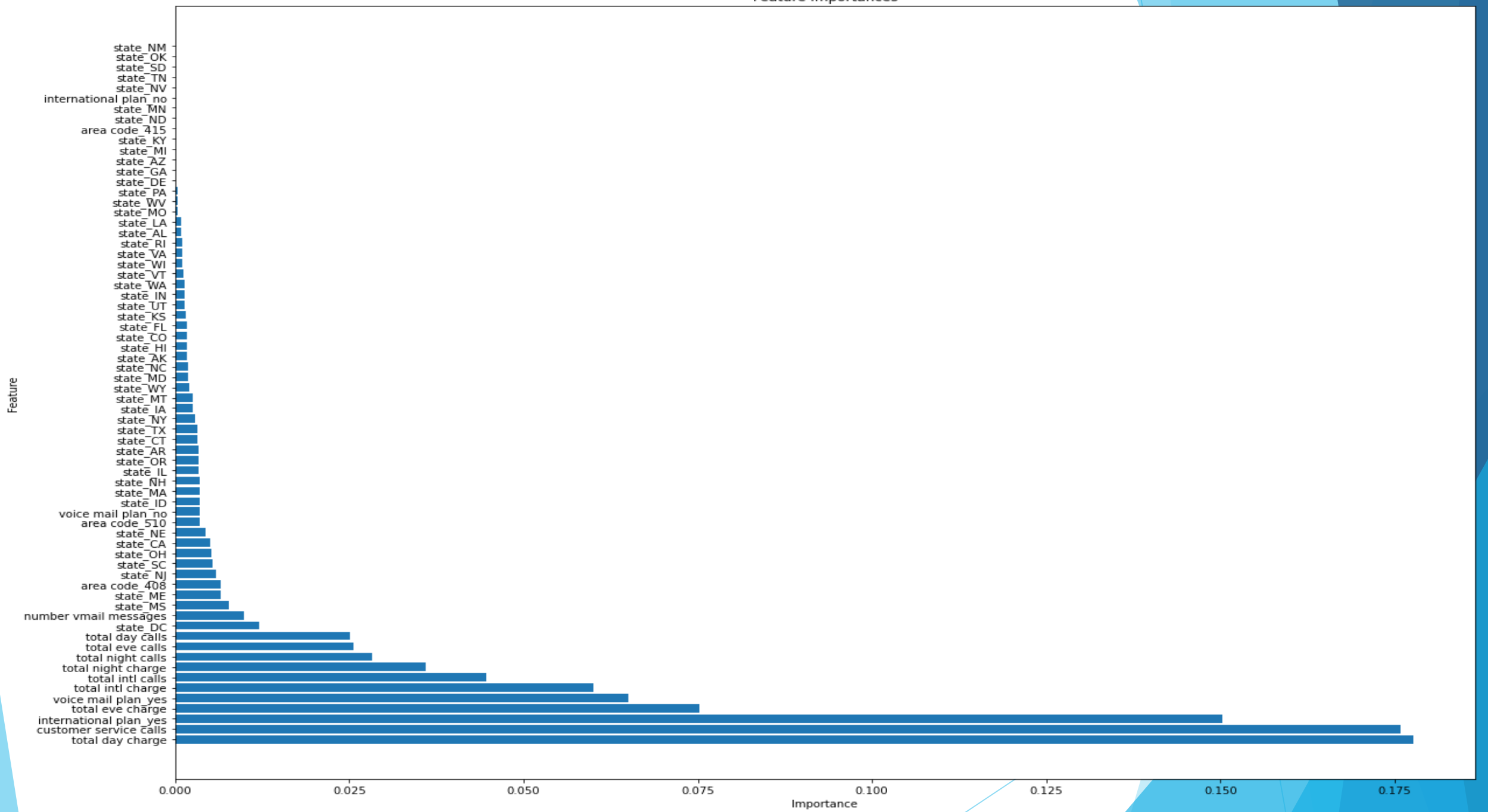
- From the above important features graph, we can see that total day charge had the most significant impact on the target variable followed by customer service calls and then in the third most significant feature is the total eve charge. We also see that state_HI has the most -ve relationship with the target variable; at state_HI, the likelihood of churn increases, followed by voice mail plan_yes. When the customer has the voice mail plan yes, the likelihood of the customer churning, increases.
- We further build other models and also included a complex model; random forest classifier

Decision Tree Classifier

- We went ahead and build a decision tree classifier. It is was used to capture the relationships in the data. It depicts the important features and which are helpful in the design of strategies to mitigate the customer churn problem.
- It had an accuracy of about 100 % on the training data and on testing data about 88%.
- This model showed a overfitting and therefore couldn’t generalize well on the testing data or the unseen data.

- The following were the important features in the decision tree classifier;

Total day charge, customer service calls, customers with the international plan, total evening charge and voice mail plan as yes together with total international charge have the most significant impact on the target variable.


Feature Importances
state_NM
state_OK
state_SD
state_TN
state_NV
international plan_no
state_MN
state_ND
area code_415
state_KY
state_MI
state_AZ
state_GA
state_DE
state_PA
state_WV
state_MO
state_LA
state_AL
state_RI
state_VA
state_WI
state_VT
state_WA
state_IN
state_UT
state_KS
state_FL
state_CO
state_HI
state_AK
state_NC
state_MD
state_WY
state_MT
state_IA
state_NY
state_TX
state_CT
state_AR
state_OR
state_IL
state_NH
state_MA
state_ID
voice mail plan_no
area code_510
state_NE
state_CA
state_OH
state_SC
state_NJ
area code_408
state_ME
state_MS
number vmail messages
state_DC
total day calls
total eve calls
total night calls
total night charge
total intl calls
total intl charge
voice mail plan_yes
total eve charge
international plan_yes
customer service calls
total day charge
Feature
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Importance

Random forest classifier

- We went ahead and performed a more complex model
- We performed an ensemble model combining multiple decision trees, to address overfitting and improve predictive performance, Random forest classifier.
- The Random forest classifier had an accuracy of about 100% on the training classifiers and on test data about 92%.
- This classifier had a better generalization ability as compared to decision tree classifier.
- Despite its generalizability, the Random forest classifier showed overfitting.
- We then analyzed the significant features on the target features. They were as follows; The total day charge, customer service calls, total international charge, international plan_yes and total night charge

Model evaluation

- Evaluation is the best practice in determining the performance of the models.
- There are various evaluation metrics which includes; accuracy, precision, recall and cross-validation.
- We evaluated our models to choose the best performing one. We will use it as our final model.
- We used accuracy and cross validation in our evaluation process.
- Logistic Regression baseline model:
  - Accuracy (Train): 79%
  - Accuracy (Test): 76%
- Decision Tree Classifier:
  - Accuracy (Train): about 100%
  - Accuracy (Test): about 88%
- Random Forest:
  - Accuracy (Train): about 100%
  - Accuracy (Test): about 92%

- From our evaluation of the models, we found out that Decision tree classifiers and Random forest showed overfitting. Logistic regression model showed a good balance in both the training and the test data.
- We performed a hyper parameter tuning on the baseline model to improve on its accuracy.

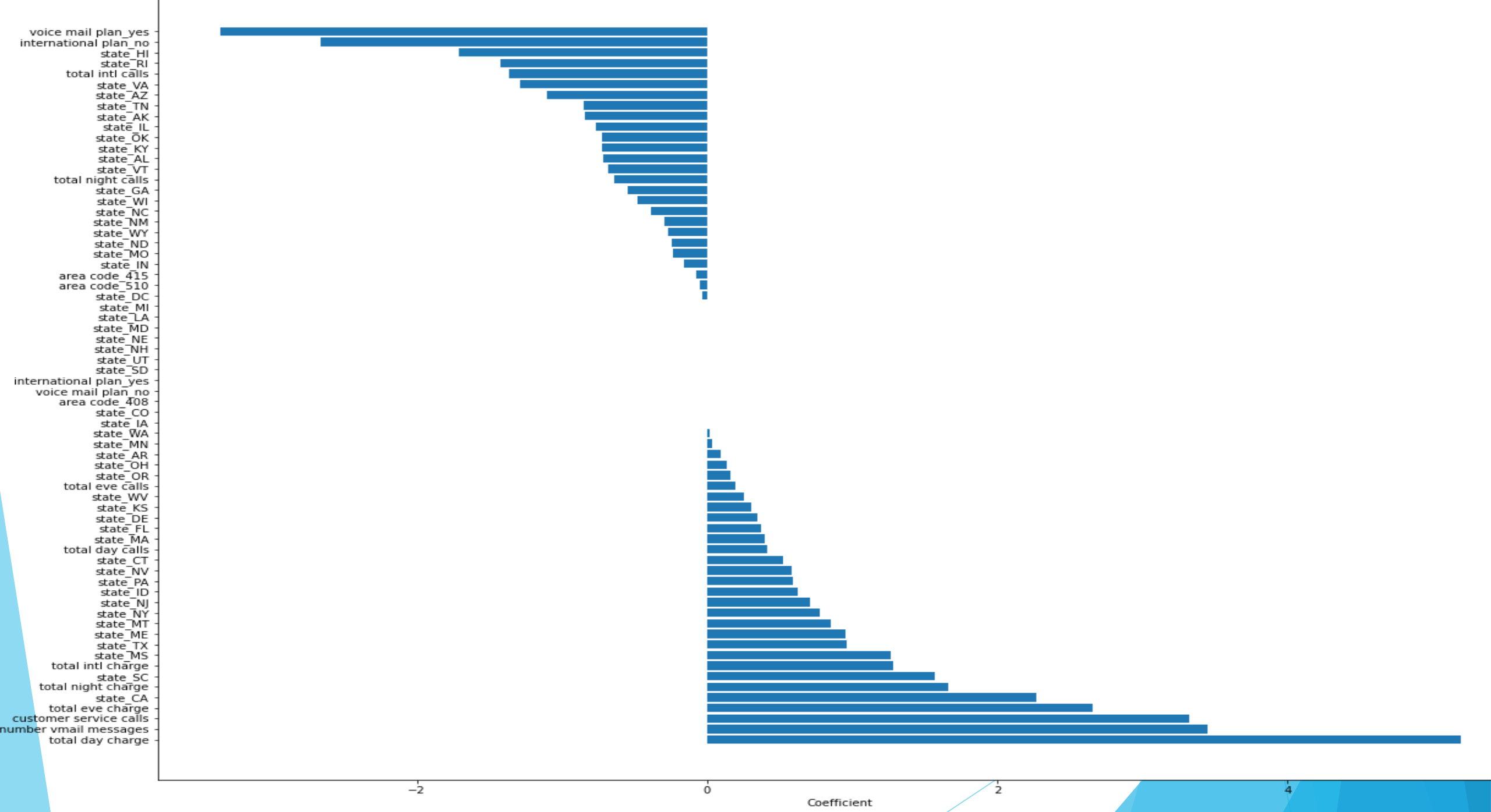
Hyper parameter Tuning

- Hyper parameter tuning helps optimize model performance by finding the best combination of hyper parameters.
- We tuned the hyper parameters of the best performing model, the logistic regression model. This was aimed at improving its accuracy and also generalizability.

Logistic regression hyper parameter tuning

- The tuning of the logistic regression‘s hyper parameters, improved its general performance though slightly.
- Its accuracy as well as its generalizability were slightly improved with the new accuracy about 81 % on the training data and on the test data about 79%.
- It showed a good balance between simplicity and predictive power.
- The following features were found to be important in the model;

total charge, number of voice mail messages, customer service calls, total evening charge, state of CA


voice mail plan_yes
international plan_no
state_HI
state_RI
total intl calls
state_VA
state_AZ
state_TN
state_AK
state_IL
state_OK
state_KY
state_AL
state_VT
total night calls
state_GA
state_WI
state_NC
state_NM
state_WY
state_ND
state_MO
state_IN
area code_415
area code_510
state_DC
state_MI
state_LA
state_MD
state_NE
state_NH
state_UT
state_SD
international plan_yes
voice mail plan_no
area code_408
state_CO
state_IA
state_WA
state_MN
state_AR
state_OH
state_OR
total eve calls
state_WV
state_KS
state_DE
state_FL
state_MA
total day calls
state_CT
state_NV
state_PA
state_ID
state_NJ
state_NY
state_MT
state_ME
state_TX
state_MS
total intl charge
state_SC
total night charge
state_CA
total eve charge
customer service calls
number vmail messages
total day charge
−2
0
2
4
Coefficient

Recommendations

- Based on the insights that were drawn in the exploratory data analysis, I would recommend the following ;

1. Improve on the customer engagement strategies. This will ensure that all customers are heard and there issues resolved before they churn.
2. Identify customers who do not have an international plan and offer them a personalized plan like lowering rates for using the services, giving them discounted bundles and other benefits. This will help prevent them from churning
3. Also we should make the activation of international plans easy for the customers. This will see a rise in the number of customers with the international plan.
4. There is a higher churn rate in area code 415. To curb this, there should be a targeted customer engagement in this area to get to know the underlying causes of the high churn rate. This will help develop strategies to mitigate the churn rate.
5. Analyze the pricing and offers related to evening charges. Reevaluate on the pricing and the offers by offering better pricing and offers that are more friendlier to the customers.

6. Use the predictability of the customer charges in the night to make personalized recommendations. This could include offering better bundles deals, increasing their talk minutes at night. This is possible with the predicted customer behavior. This can improve their loyalty to the company

7. Focus on customer retention strategies on the states with high churn rate. They include providing them with better deals like giving them bundle offers, increasing their talk minutes. The high churn rate could also be due to poor network coverage in these states. Therefore, there should be a mechanism to ensure a good network coverage on these states. Look at the underlying reasons for the customer churn in these states and devise ways to curb them.

8. Educate the customers on the benefits of the voice mail plans.
9. Use the customers usage patterns to tailor voice mail plan that suits their individual needs.

10. Focus more on retaining the existing satisfied customers to build a good customer base.

Conclusion

- In conclusion, the analysis of the customer churn in SyriaTel has provided clear knowledge on the factors that leads to churning of the customers as well as valuable insights into the customer behaviors.
- The models have provided a clear predictive power on customers churn as well as identifying important features that greatly influence customer retention.

Next steps

- The company can go ahead and develop specific retention strategies such as ensuring certain areas that had high churn rate are given quality network coverage and offers like increased bundles at a cheaper rate
- Monitor and evaluate the ever changing customers behavior in order to stay up to date with the customer churn mitigation strategies
- Stay up to day date the market trends which are bound to change every other single day.
- Focus on the customer satisfaction. This will ensure that the customers who have not churned are always enjoying the services of the company and that at no point will they discontinue with the usage of the company's services and products.

You can reach me via :

Name: Amos Kipkirui

Email : emohkipkirui756@gmail.com

Phone number : +254707286055

Thank you !