

# **TELECOM CHURN ANALYSIS**

## **TEAM MEMBERS**

**Sukanya Nambiar**

**Subramanya L**

**Suhas K L**

# **CONTENTS:-**

1. Business Problem statement.
2. Objective
3. Detailed summary of data
4. Exploratory Data Analysis
5. Recommendation
6. Conclusion for EDA
7. Feature engineering, model building & Deployment part.

# **BUSINESS PROBLEM STATEMENT**

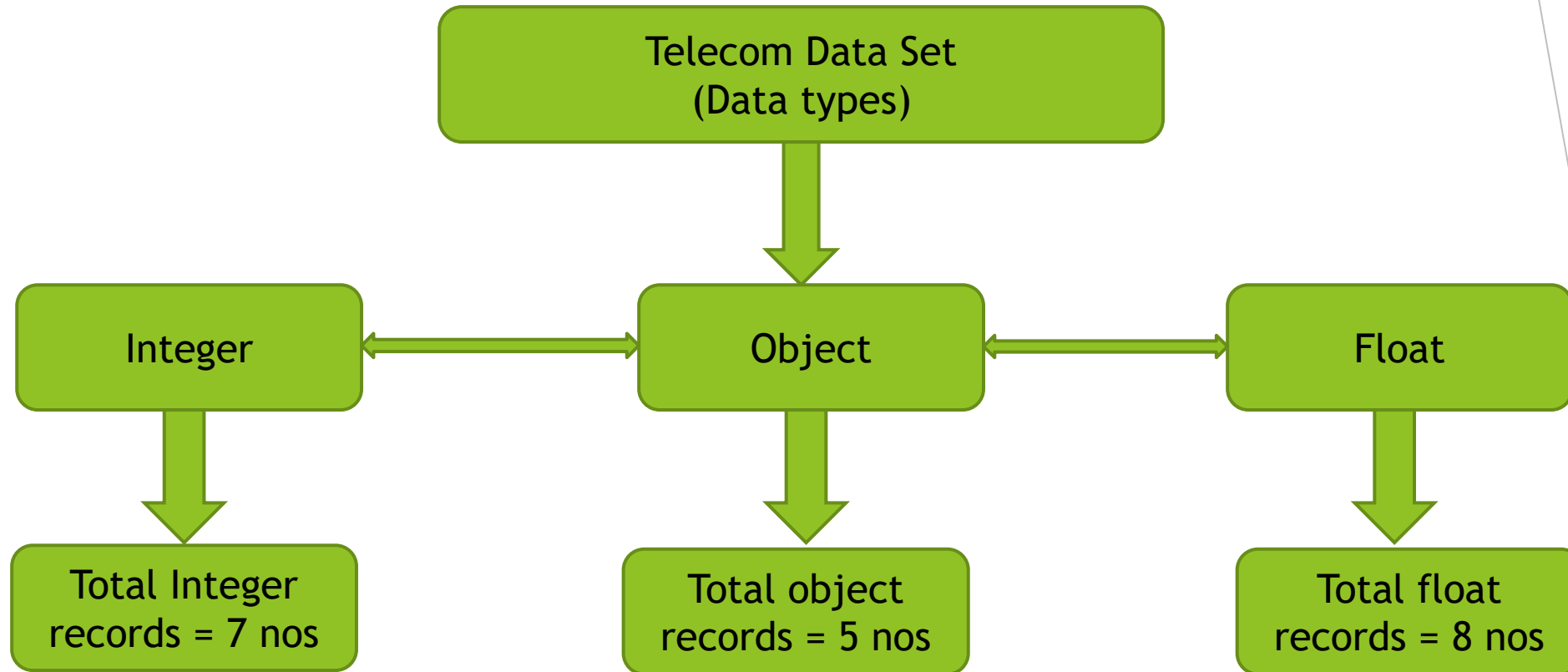
- Customer churn is a big problem in telecom industry.
- Churn prediction is nothing but detecting customers who are likely to cancel their subscription to a service.
- Churn is a problem to the telecom industry since it is more expensive to acquire a new customer than to keep your existing one from leaving.

# OBJECTIVE

- ▶ **Maximize:** Company's profit by retaining customer.
- ▶ **Minimize:** Customer churn by identifying the key cause of the problem.
- ▶ **Business Constraint:**  
Provide offers and discount and improve the service quality without compromising with profit.

*This is a classification project since the variable to be predicted is binary (churn or not churn). The objective here is to predict churn probability, conditioned on the customer features.*

# DETAIL SUMMARY OF DATA



## Important Notes:-

1. Here the decision variable is the 'churn' column - 'y'
2. Rest all the others lies In the feature space 'X'

# DETAIL SUMMARY OF DATA

## (GETTING A BIG PICTURE OF DATA SET)

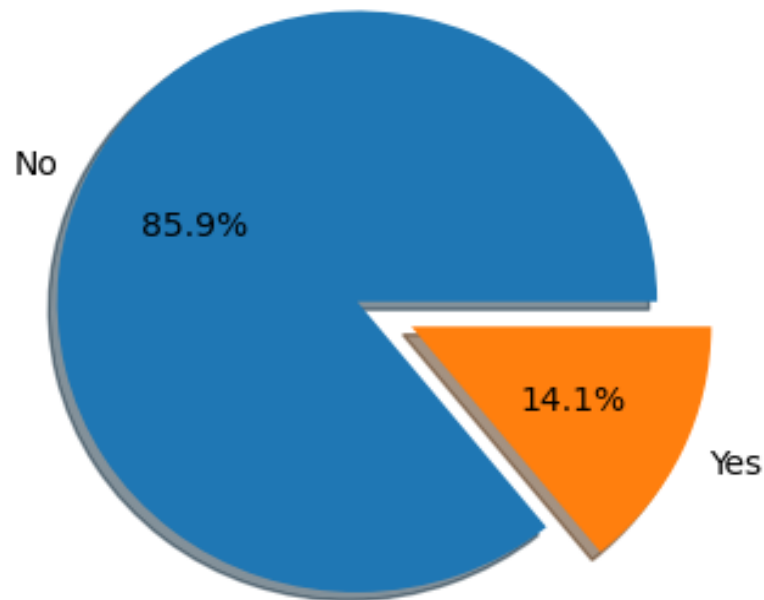
1. First 5 data sets
2. Last 5 data sets
3. Data Shape - (5000,21)
4. Checking for Data Types - For some of the data sets we have changed their data types.
5. Checking for null values - we had some null values in day charge & eve min column we had filled them up with mean values (since they were numeric columns)
6. Checking for duplicated values - 0
7. Checking for unique values

# EXPLORATORY DATA ANALYSIS

## (Analyzing dependent variable 'Churn')

### 1. Study with the Pie Chart

Pie chart for churn



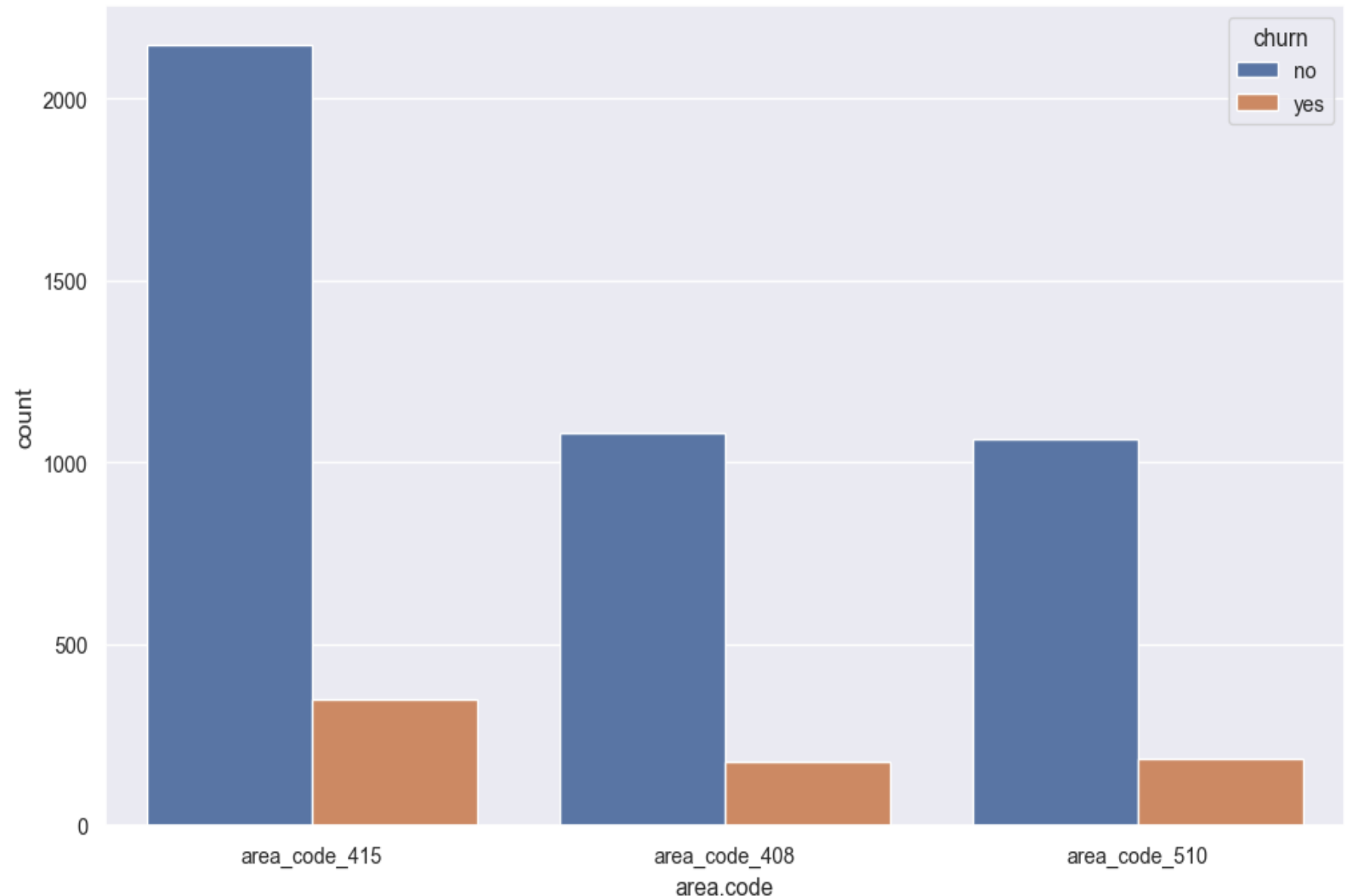
No of total customers in data set = 5000 nos  
No of total customers churn = 707 nos  
No of un-churn customers = 4293 nos

The % wise details is shown as given in the chart.

# EXPLORATORY DATA ANALYSIS

## 1. Analyzing the area code column

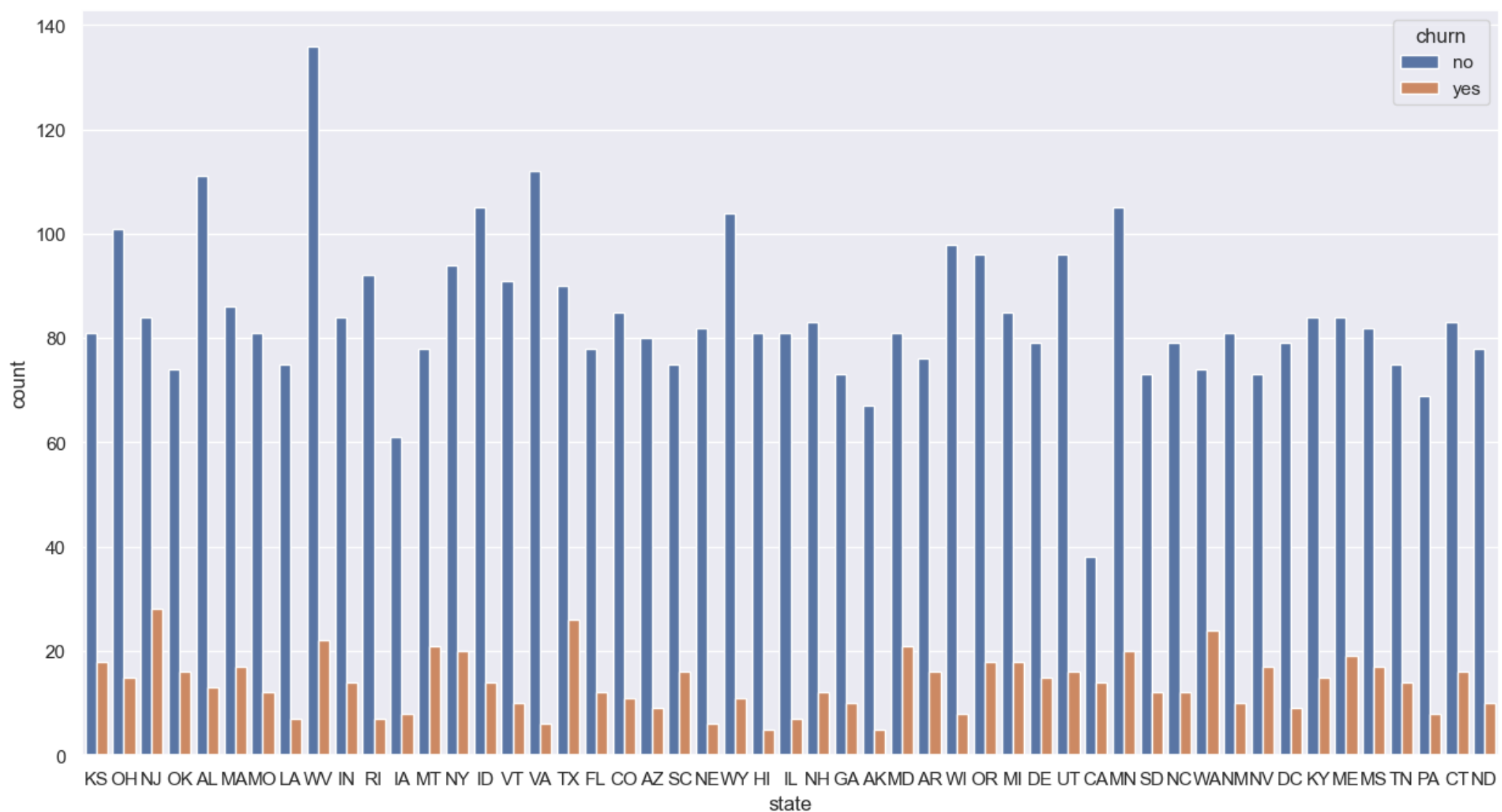
*In the above data, we notice that there is only 3 unique value are there i.e 415, 408, 510 and the churn rate of these area codes are almost same. we don't think there is any kind of relation present between the "area code" and "churn" due to which the customer leaves the operator.*





# EXPLORATORY DATA ANALYSIS

## 2. Analyzing the state column



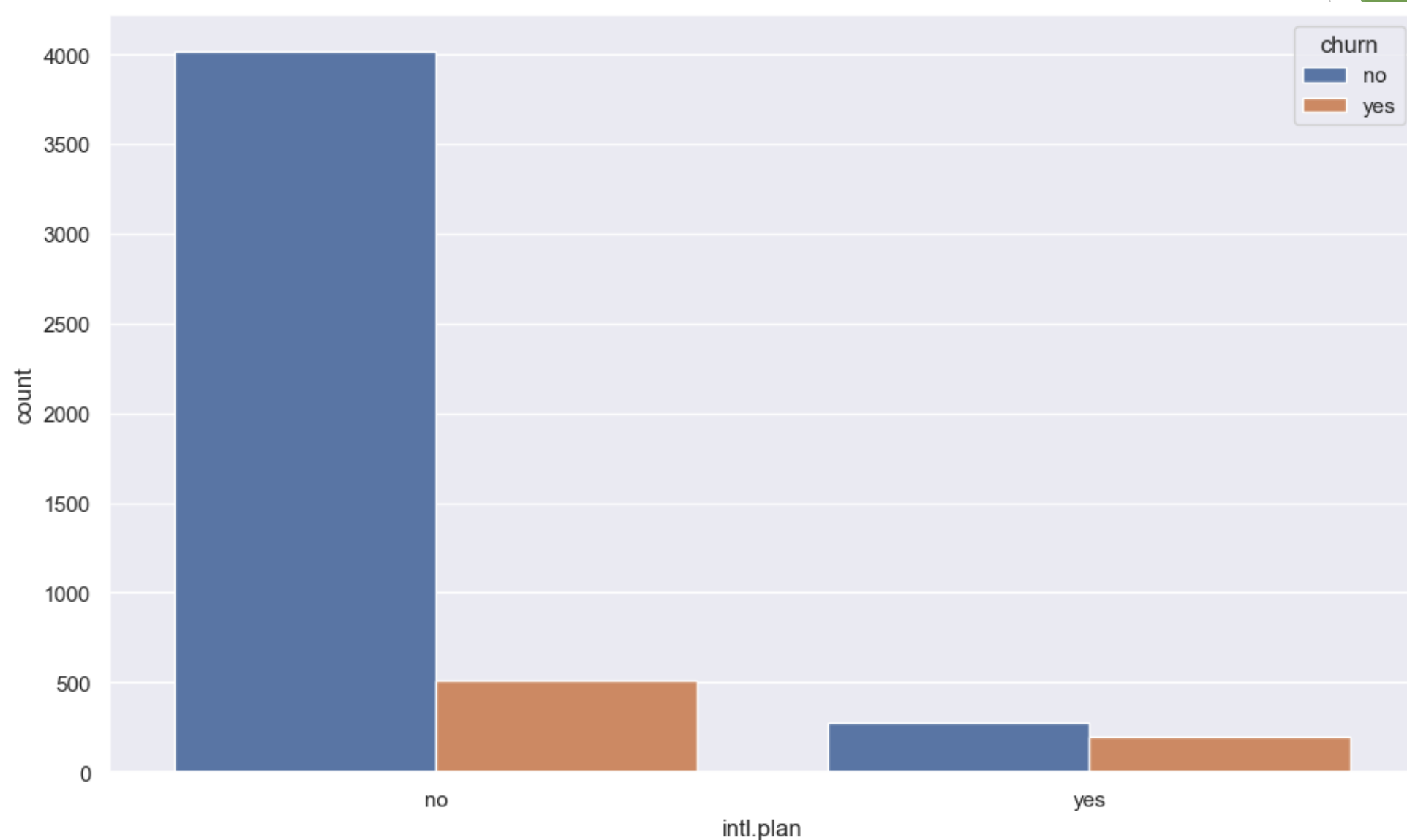
# EXPLORATORY DATA ANALYSIS

- After analyzing it is seen that, there are total 51 unique states.
- Out of these 51 unique states, 10 states have higher churn rate of 21%
- The reason for this churn rate may be due to low coverage area of the cellular network.
- The following are the top 10 states which have this problem.

Sr. No	State Name	% Churn > 21%
1	CA	26.9
2	NJ	25.00
3	WA	24.40
4	TX	22.40
5	MT	21.21
6	MD	20.58
7	NV	18.88
8	ME	18.44
9	KS	18.18
10	OK	17.77

# EXPLORATORY DATA ANALYSIS

## 3. Analyzing the international plan column



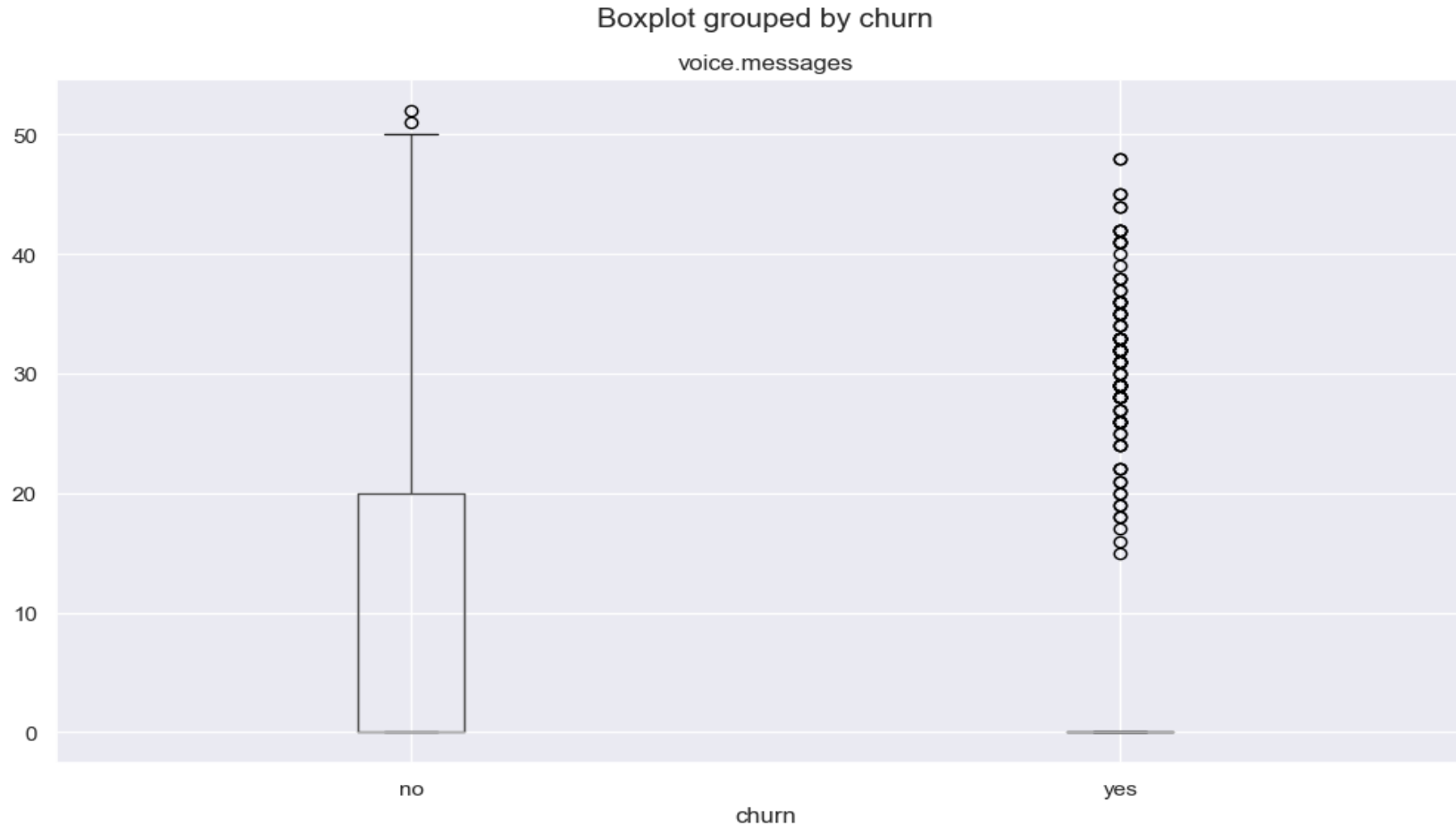
# EXPLORATORY DATA ANALYSIS

- Total customers without international plan are 4527
- Total customers with international plan are 473
- Hence we observe that, people who have taken international plan have more churn rate of 42%
- So basically the people who bought International plans are churning in big numbers. Probably because of connectivity issues or high call charge.

Sr. No	Plan Type	Not Churn	Churn	% Churn
1	No International Plan	4019	508	11.22%
2	With International Plan	274	199	42.07%

# EXPLORATORY DATA ANALYSIS

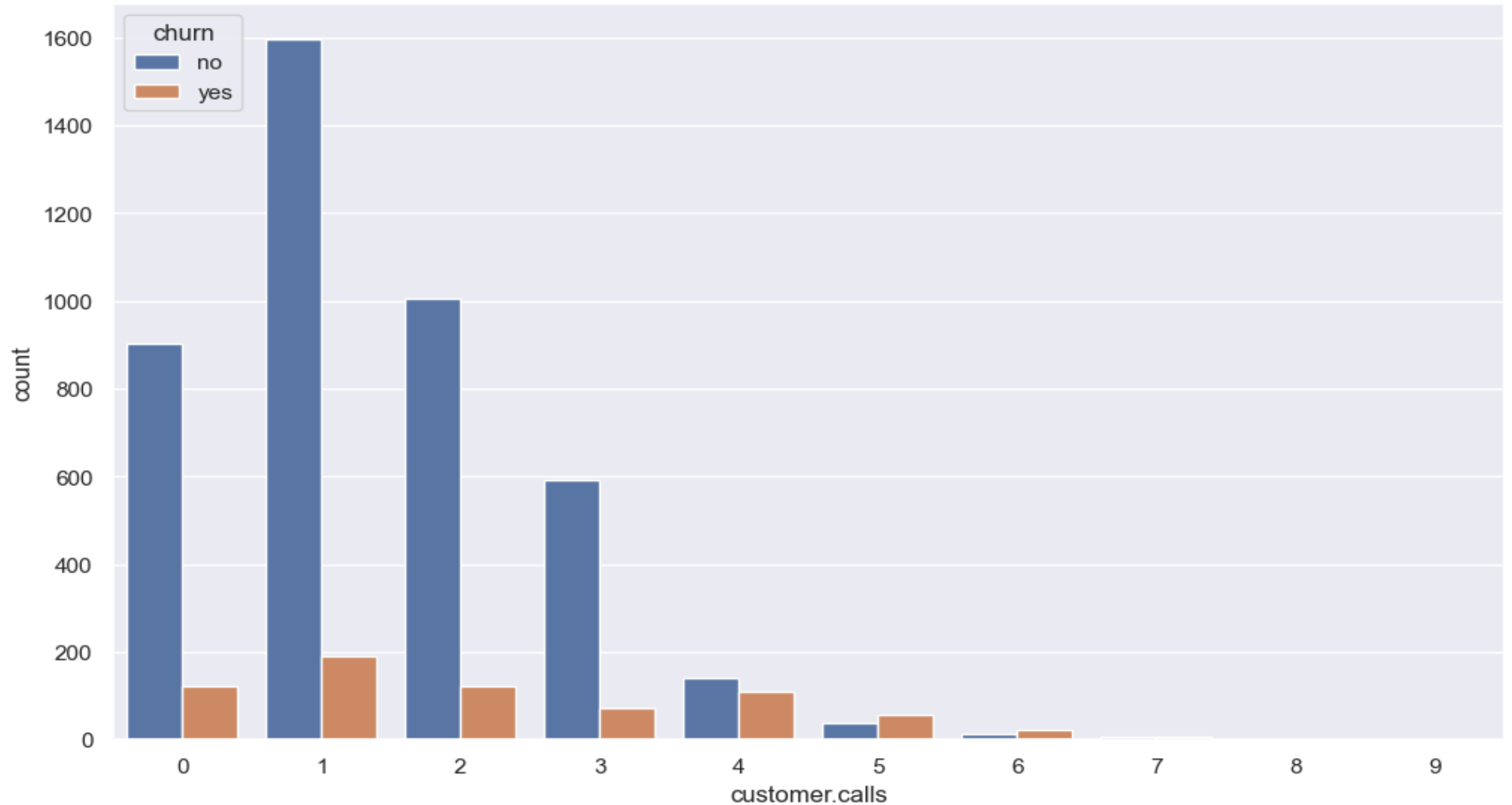
## 4. Analyzing the voice message column



After analyzing the above voice mail feature data we get an insight that when there are more than 20 voice-mail messages then there is a churn. For that, we need to improve the voice mail quality.

# EXPLORATORY DATA ANALYSIS

## 5. Analyzing the customer calls column



# EXPLORATORY DATA ANALYSIS

- The above data shows that customer car services are not good at attending the queries of the customer, & also customer who had called more than 5 times have % churn of 60%. Hence they need to improve their services

Sr. No	No of Customer calls	Not Churn	Churn	% Churn
1	0	902	121	11.8
2	1	1596	190	10.63
3	2	1005	122	10.82
4	3	592	73	10.97
5	4	141	111	44.04
6	5	38	58	60.41
7	6	12	22	64.70
8	7	6	7	53.84
9	8	1	1	50.00
10	9	0	2	100.00

# EXPLORATORY DATA ANALYSIS

## 6. Analyzing other remaining columns

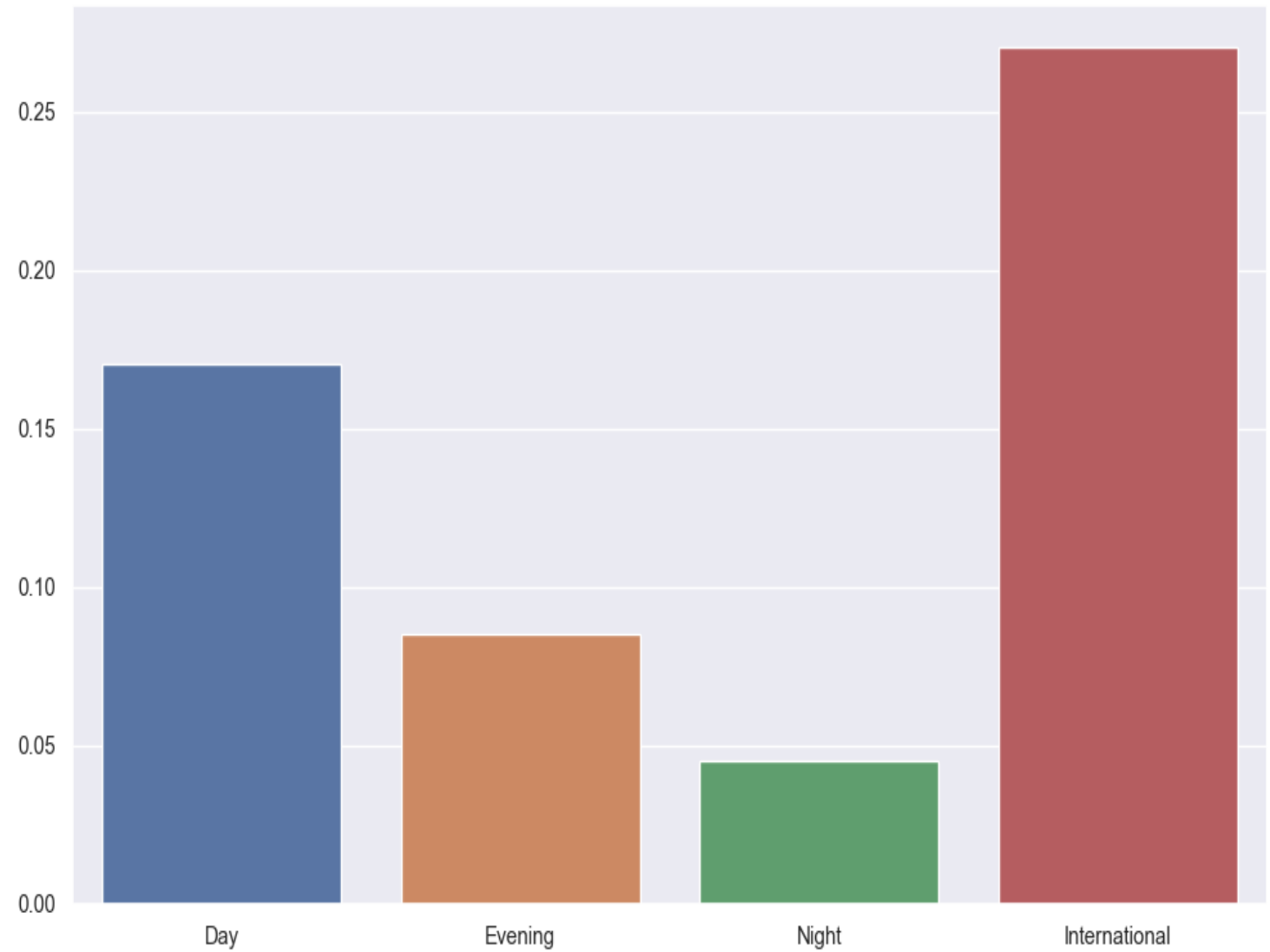
After analyzing,

- ❖ Day charge vs Day Mins
- ❖ Eve charge vs eve Mins
- ❖ Night Charge vs Night Mins
- ❖ Intl charges vs Intl Mins

We come to the conclusion that,

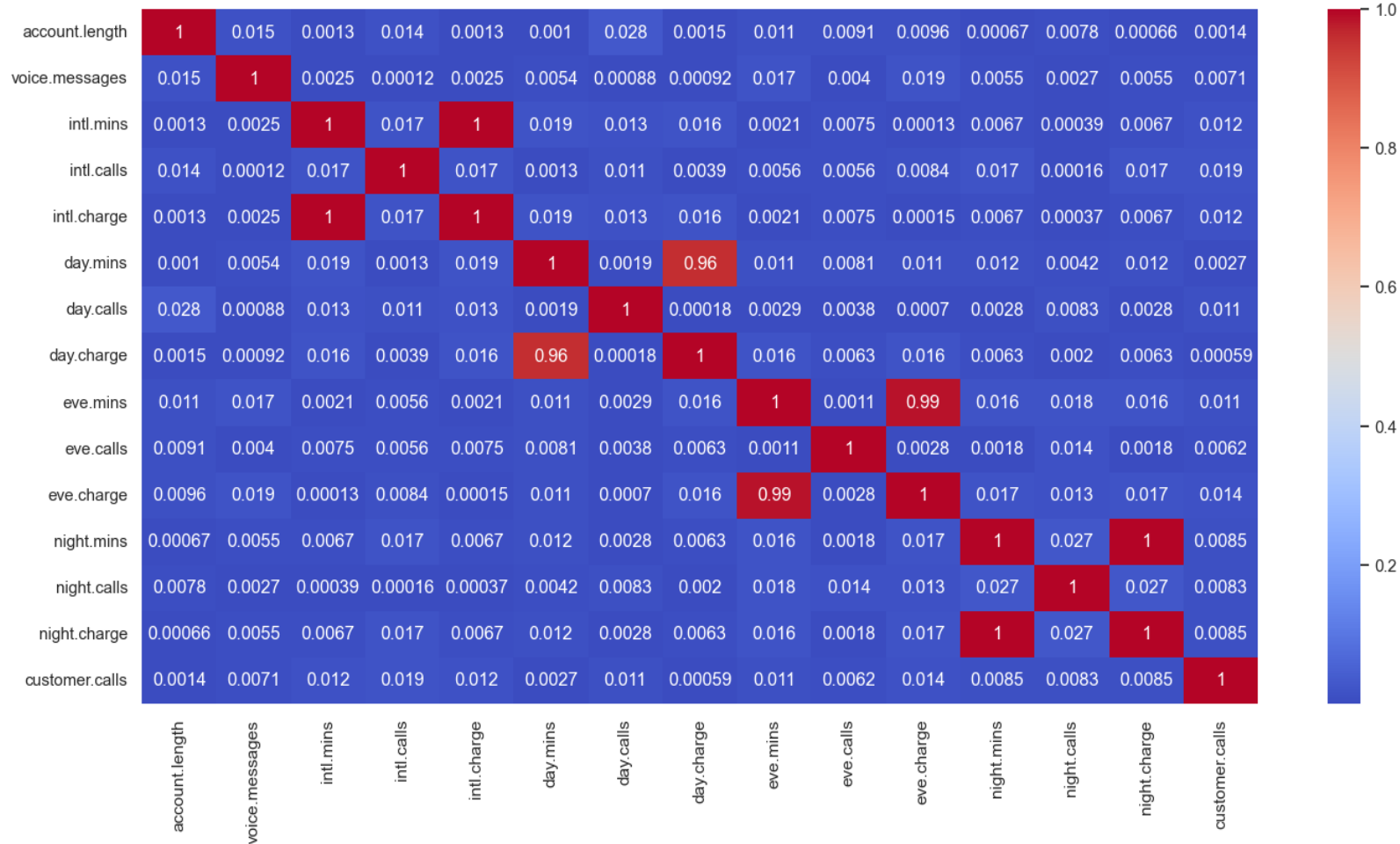
*The above dataset we have noticed that total day/night/eve minutes/call/charges do not put any kind of cause for churn rate.*

*But international call charges are high as compare to others it's an obvious thing but that may be a cause for international plan customers to churn out.*





# CORRELATION MATRIX



We see a strong correlation in some columns, Such as 'Day Charge' & 'Day Minutes' & these features will be removed during feature selection to avoid 'Multi collinearity Issues'

# RECOMMENDATIONS

1. Improve network coverage churned state
2. In international plan provide some discount plan to the customers
3. Improve voice mail quality or take feedback from the customer
4. Provide discount to those customer who spent more minutes.
5. Improve the service of call center & take frequently feedback from the customers regarding their issue & try to solve it as soon as possible.

# CONCLUSION FOR EDA

These are some of the quick insights on churn analysis from above analysis :

- ▶ The customers who used international plan services are higher in churners .  
So basically the people who bought International plans are churning in big number. Probably because of connectivity issues or high call charge.
- ▶ In voicemail section when there are more than 20 voice mail messages there is churn so it basically means that the quality of voice mail is not good.
- ▶ The customer who have high day call minutes also have high call price these customer tends to churn.

# Feature Engineering & Model Building Part

Today we will be covering the following points:-

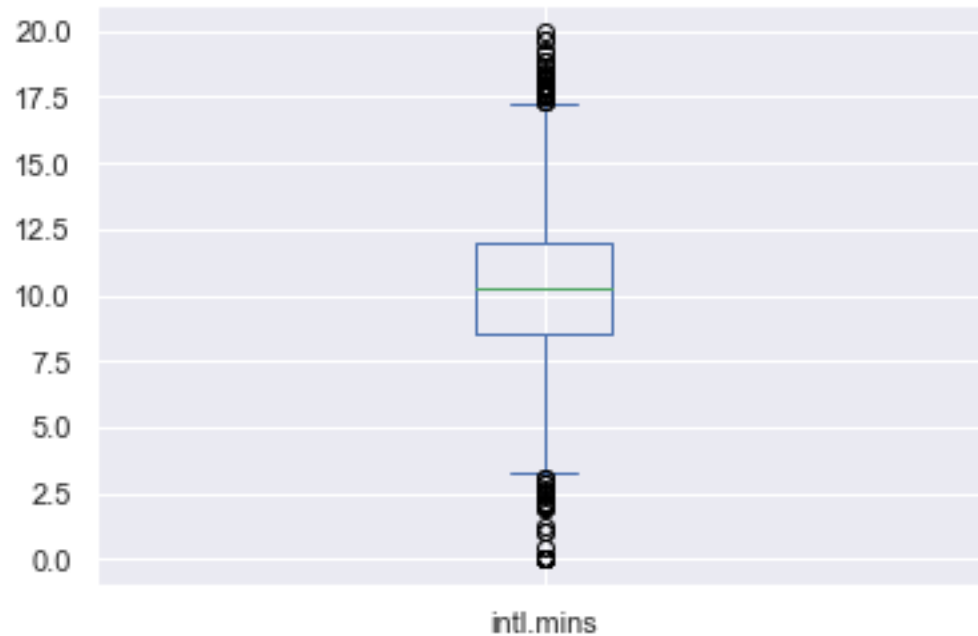
1. Outlier Detection part
2. Feature Engineering.
3. Balancing of the data set.
4. Model Building using different models.

# Outlier Detection Part

- We have removed the outliers using IQR method

Ex:-

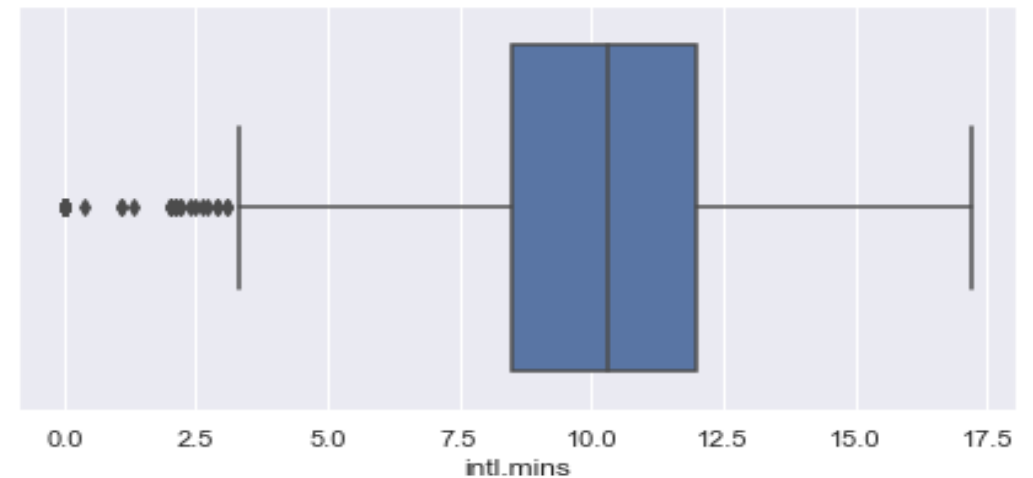
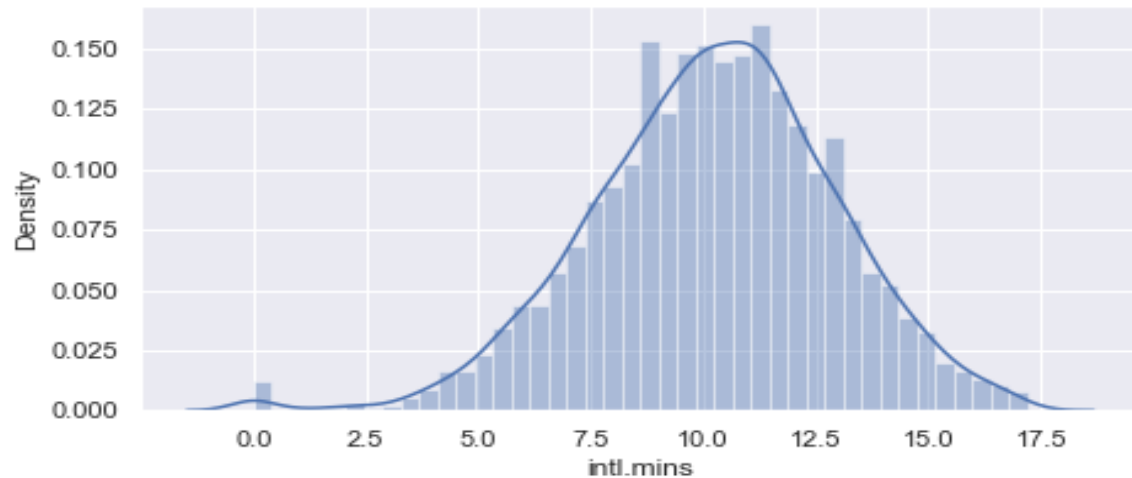
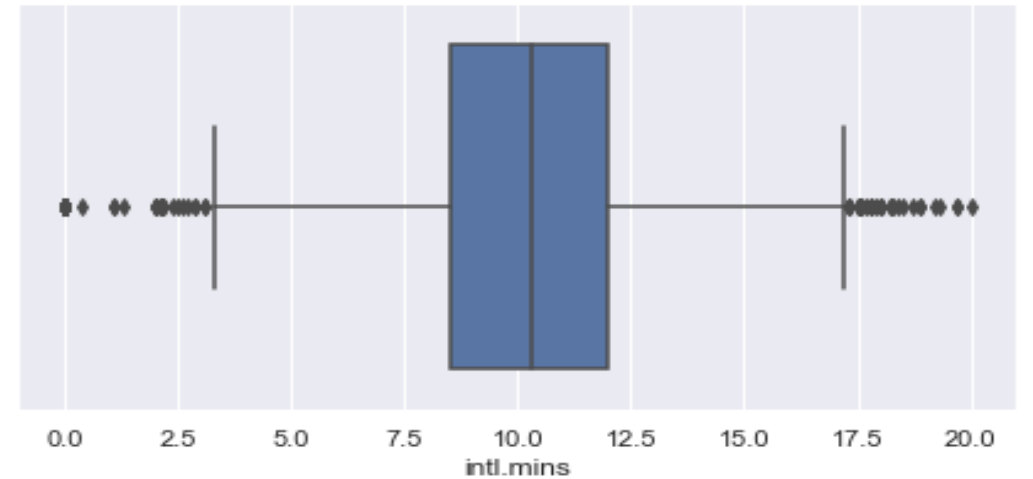
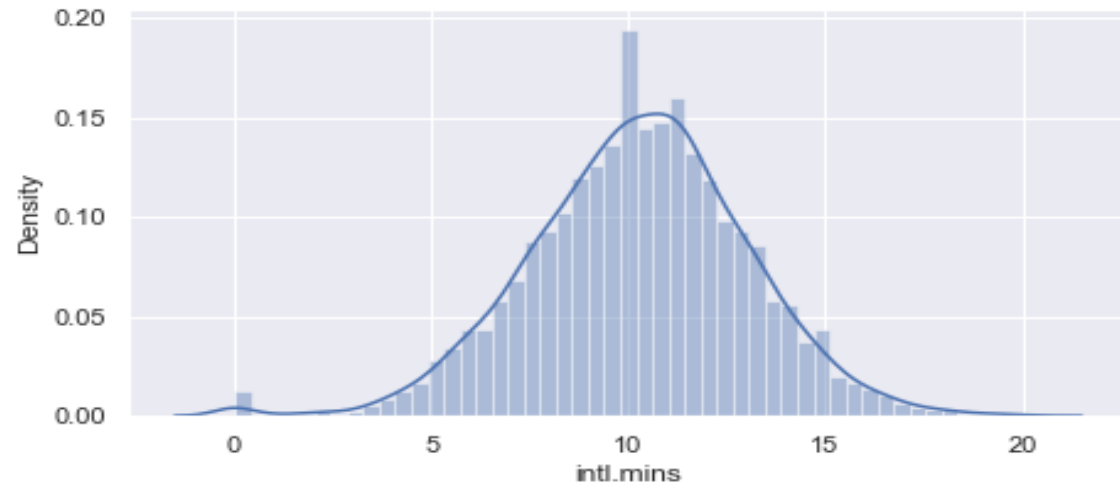
1. Detecting using Box plot first for intl. mins column.



# Outlier Detection Part

► After applying IQR method the following were the results,

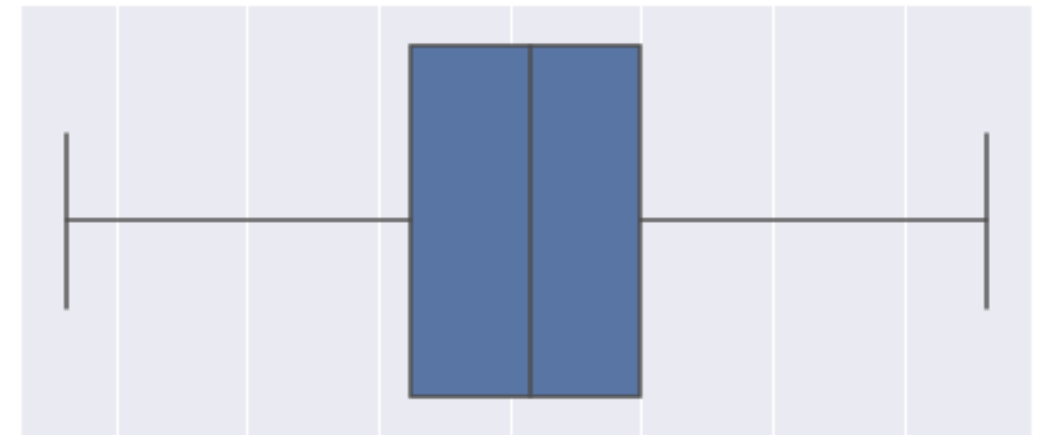
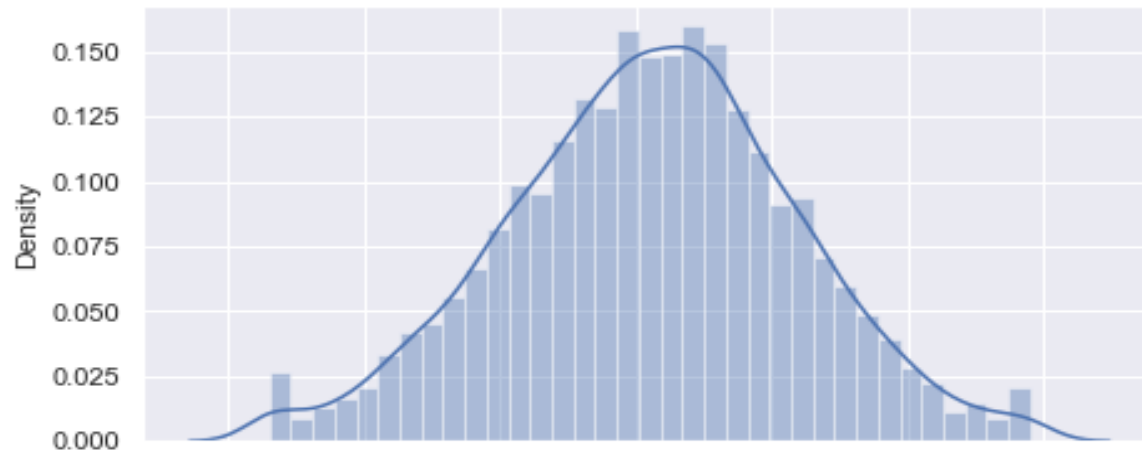
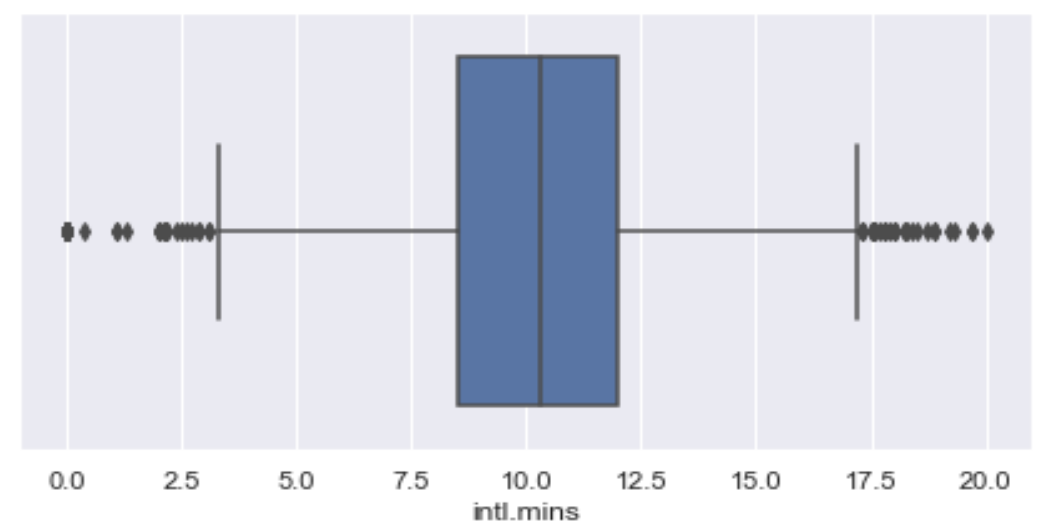
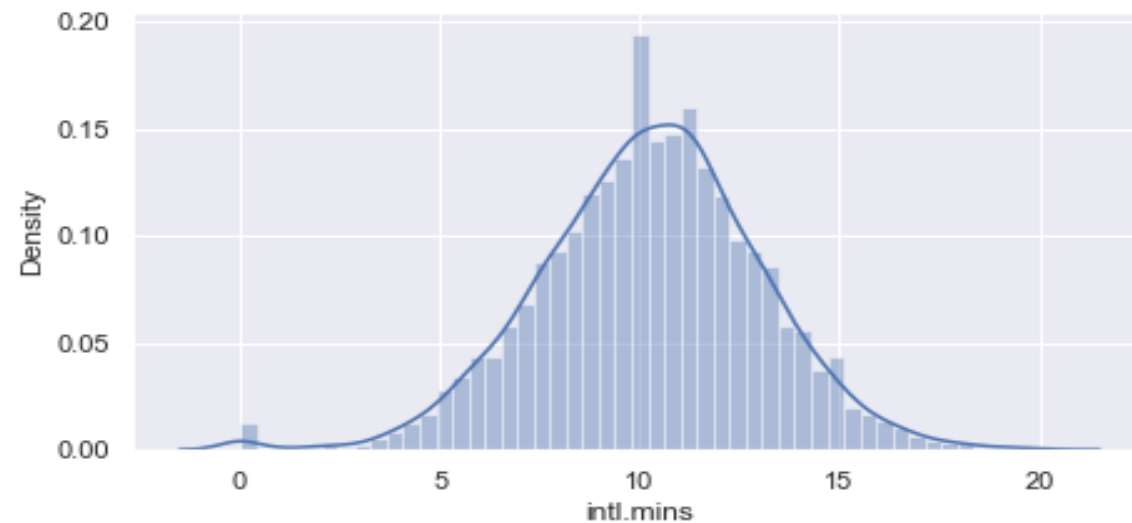
► Trimming:-



# Outlier Detection Part

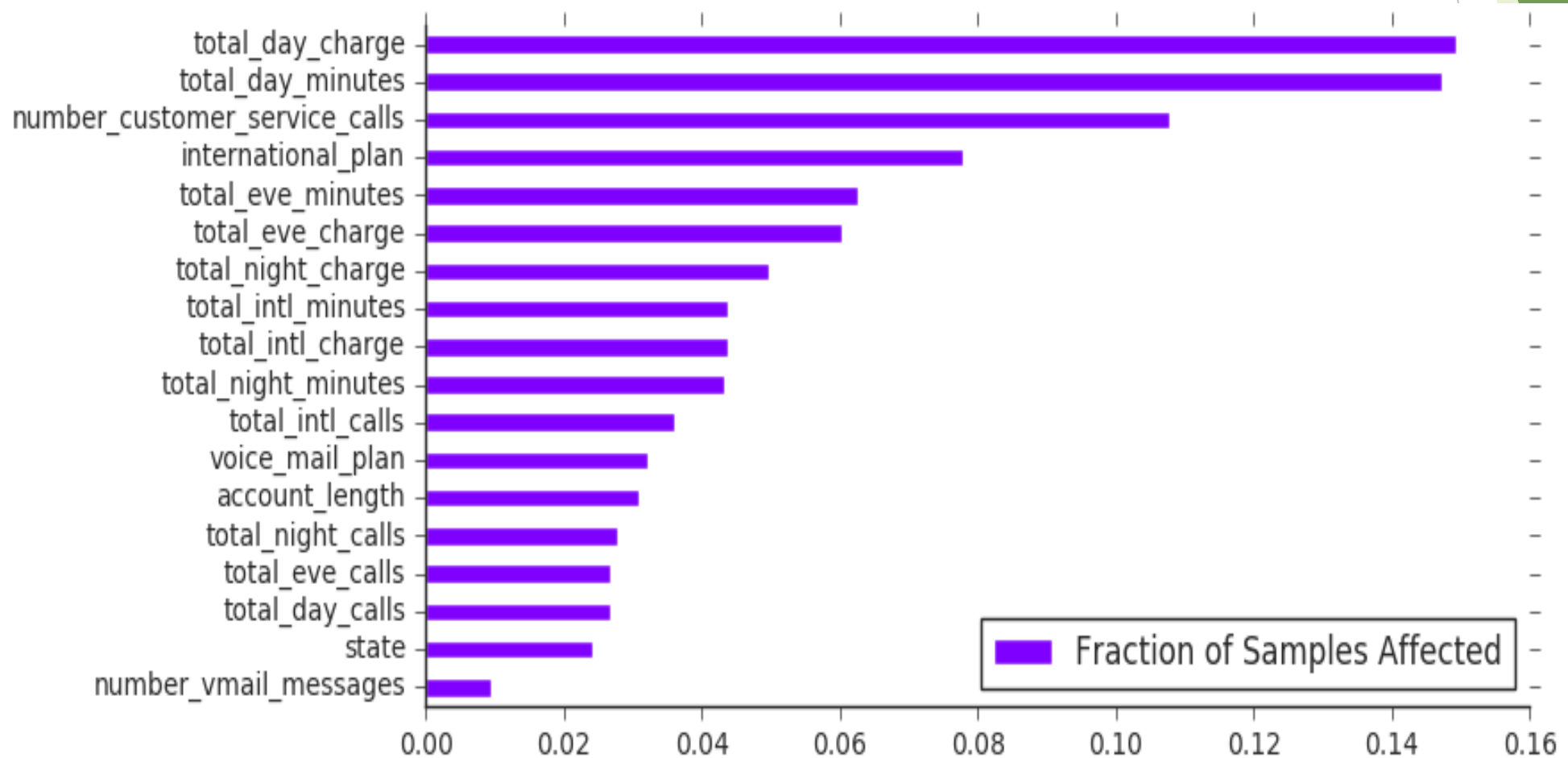
► After applying IQR method the following were the results,

► Capping:-



# Feature Engineering.

1. Finding the feature which have highest predictive power.





# Feature Engineering.

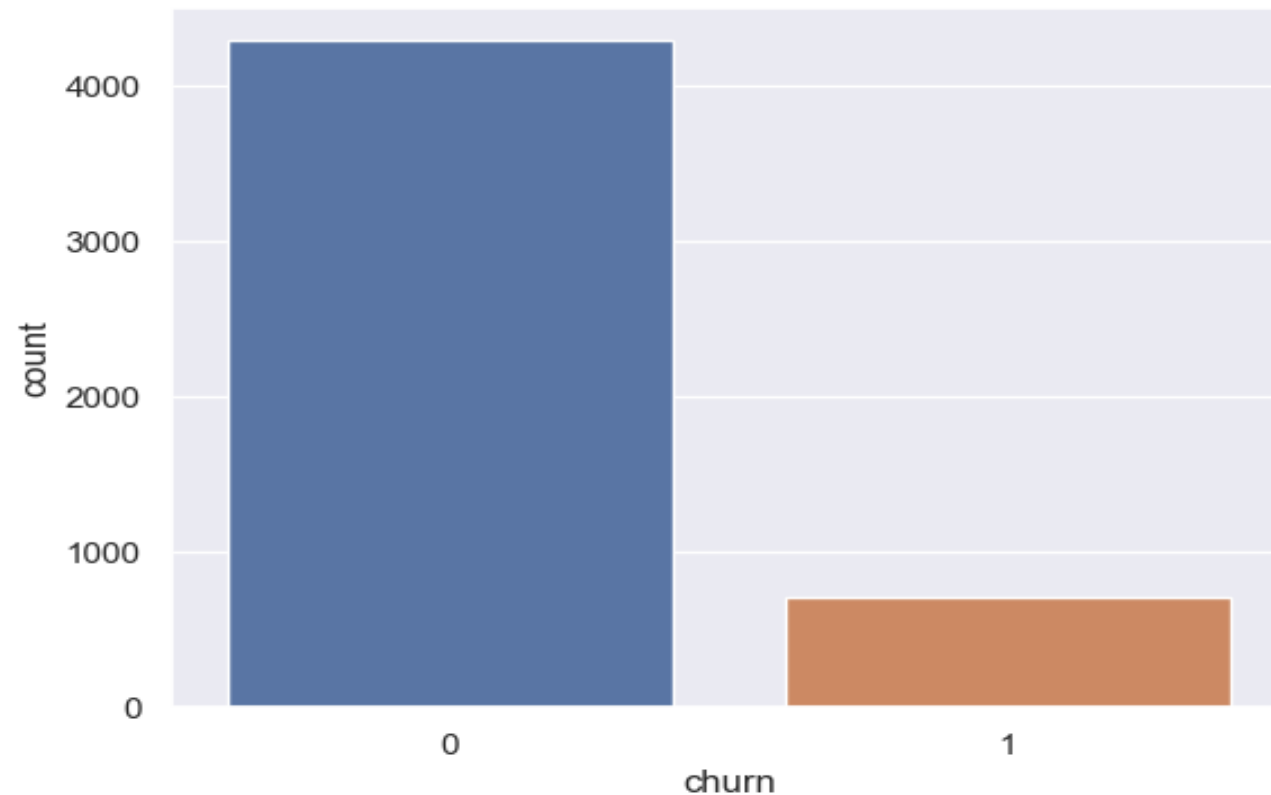
Finding the feature which have highest predictive power. We used random forest algorithm to find out the best features.

- ▶ According to the feature importance analysis produced by the Random Forest algorithm, the following features had the highest predictive power:
- ▶ total\_day\_minutes
- ▶ total\_day\_charge
- ▶ number\_customer\_service\_calls
- ▶ total\_eve\_minutes
- ▶ international\_plan

Splitting the data into train & test split :- We have split the data into train & test subsets.

# Feature Engineering.

## 1. Finding the data was balanced or no:-



We have used the following balancing techniques.

1. Over Sampling using smote.
2. Under Sampling.
3. Classic over Sampling technique

We came to know that,

1. Non churn customers - 4293
2. Churn customers - 707

3. The data is highly imbalanced.

The data is balanced using Random Over sample, using smote

Hence,

Old data - (4293,22) , (707,1)

New data - (4293,22) , (4293,1)

# Model Building

## Selecting the Best Model

Sr. No	Model	Accuracy (%)	Remarks
1	Logistic Regression	76	
2	Decision Trees	82	
3	Random Forest	99	We choose RF as our final model
4	Adaboost	79	
5	Gaussian	75	

# Model Deployment



**DEPLOYMENT**

- After Model building & Evaluation process, we have deployed the code using “Stream lit”
- We have selected the best model & used in our deployment phase.

×

Type of chart

Scatterplots

Scatterplot Settings

X axis

day\_calls

Y axis

day\_charge

account\_length

101.66

1.00

243.00

voice\_mail\_plan

0.28

0.00

1.00

voice mail messages

Created on Tue May 24 19:58:07 2022

@author: Subash S

# Customer Churn Prediction App

This app predicts the Customer Churn Prediction!

## Description of Dataset

account\_length - Account Length of the Customer

voice\_mail\_plan - Voice Mail Plan for the Customer

voice\_mail\_messages - Voice Mail Messages for the Customer

day\_mins - Call day Mins Of the Customer

evening\_mins - Call Evening Mins Of the Customer

night\_mins - Call Night Mins Of the Customer

international\_mins - Call NightInternational Mins Of the Customer

voice\_mail\_messages

8.18

0.00

51.00

day\_mins

179.78

0.00

350.00

evening\_mins

300.98

0.00

363.78

night\_mins

206.87

23.20

395.00

international\_mins

18.24

0.00

20.00

customer\_service\_calls

1.56

0.00

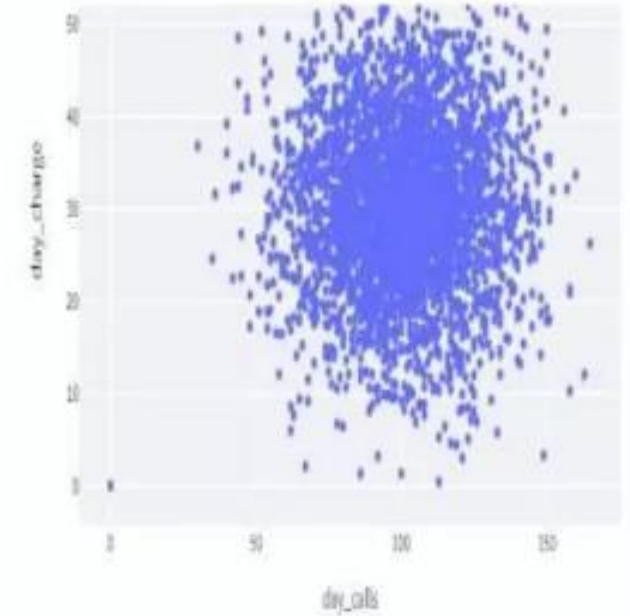
9.00

international\_plan

0.00

0.00

0.00



## Churn Prediction Web App

Enter Account Length

14

Enter Voice Mail Plan

0



# Churn Prediction Web App

Enter Account Length

14

Enter Voice Mail Plan

0

Enter Voice Mail Messages

45

Enter Day Mins

89

Enter Evening Mins

25

Enter Night Mins



20

Enter Evening Calls of the Customer

100

Enter Evening Charges of the Customer

10

Enter Night Calls of the Customer

150

Enter Night Charges of the Customer

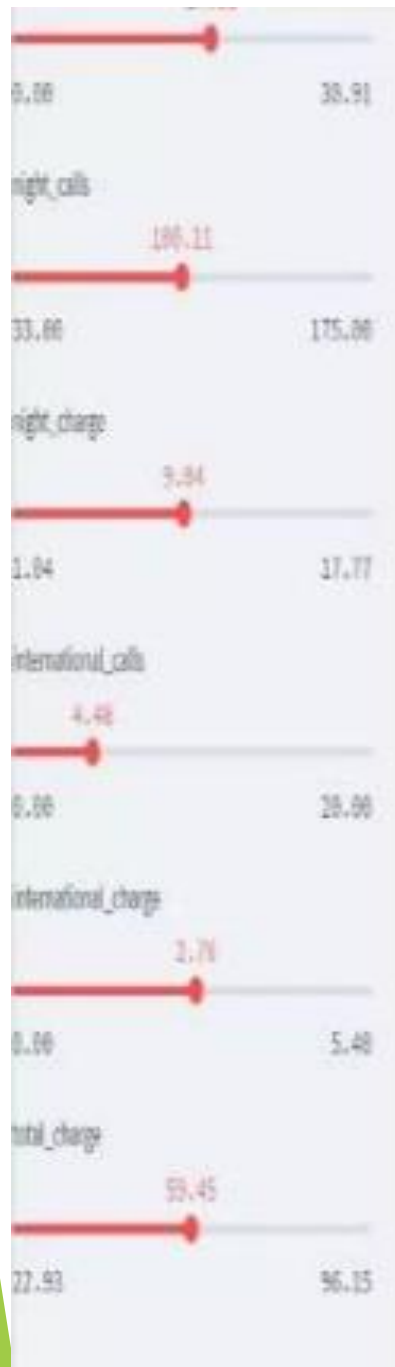
10

Enter International Calls of the Customer

52

Enter International Charges of the Customer

10



Enter International Calls of the Customer

52

Enter International Charges of the Customer

30

Enter Total Charges of the Customer

50

Customer Churn Result

The person is Churn

Show SHAP Graphs

Output

