Capstone Project

Project Title - Telecom Churn Analysis

Team Members

1. Ankit Kumar
2. Nilanjan Chandra
3. Nitesh Singh
4. Rishabh Kumar
5. Sanjog Mishra

Index

- Introduction: Telecom Churn
- Problem Statement
- Features of the dataset
- Data pipeline
- Data summary
- Feature Selection
- EDA and Feature Engineering
- Inference
- factors responsible for Churning
- Final recommendation

Introduction: Telecom Churn

Telecom companies normally give more effort towards customer acquisition than retention. However, the cost of customer acquisition is around five times more than the new customer acquisition.

According to research done by Bain & Company, increasing customer retention rates by 5% can increase profits by 25% to 95%.

Churn is also known as customer attrition . With the help of the Churn dataset, companies try to find the specific reasons behind churn numbers and act on those factors with proper action plans.

Introduction: Telecom Churn

Also the telecom companies perform the churn analysis in order to know inb advance about the customer segments who are about to leave the service and get a chance to prevent that.

Generally multiple factors are responsible behind the customer dissatisfaction and decision of leaving the facility.

In our project we have tortured the given dataset so that it confess to our query of finding the possible factors which lead to Customer Churn.

Problem Statement

Orange S.A., formerly France Télécom S.A., is a French multinational telecommunications corporation. The Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features), along with a churn label specifying whether a customer cancelled the subscription.

Explore and analyze the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.

Features of the dataset

The dataset contains 3333 number of rows and 20 number of columns.
The important features of the dataset are

State : It shows the different states in which data is classified. Total number of unique states is 51

Account length : It shows number of days for a customer uses product

Area code : Categorical data shows different areas in which the service is used

International plan : Categorical data shows whether customer has International plan or not

Voice mail plan : Categorical data shows whether customer has voice mail plan or not

Total day calls : Total day calls in a day for a customer

Total day charge : Total day Charge for day calls in a day for a customer

Total day minutes : Total day call minutes in a day for a customer

Total eve minutes : Total evening call minutes in a day for a customer

Features of the dataset

Total eve calls : Total evening calls in a day for a customer

Total eve charge : Total evening Charge for evening calls in a day for a customer

Total night minutes : Total night call minutes in a day for a customer

Total night calls : Total night calls in a day for a customer

Total night charge : Total night Charge for night calls in a day for a customer

Total intl minutes: Total international call minutes in a day for a customer

Total intl calls : Total international calls in a day for a customer

Total intl charge : Total international Charge for international calls in a day for a customer

Customer service calls : It shows number of customer service calls for a particular customer

Churn : It shows whether customer has churned or not

Features of the dataset: Added Columns

Total Charges: An additional column ‘total_charges’ was created which has the value as sum of the total day charge, total evening charge and total night charge amount.

Total minutes: Another column ‘total_minutes’ was created which has the value as sum of the total day minutes, total evening minutes and total night minutes

Hours: Hours column has been added with the Churn True Dataset

Months: Month Column has been added with the Churn True Dataset

Data Pipeline

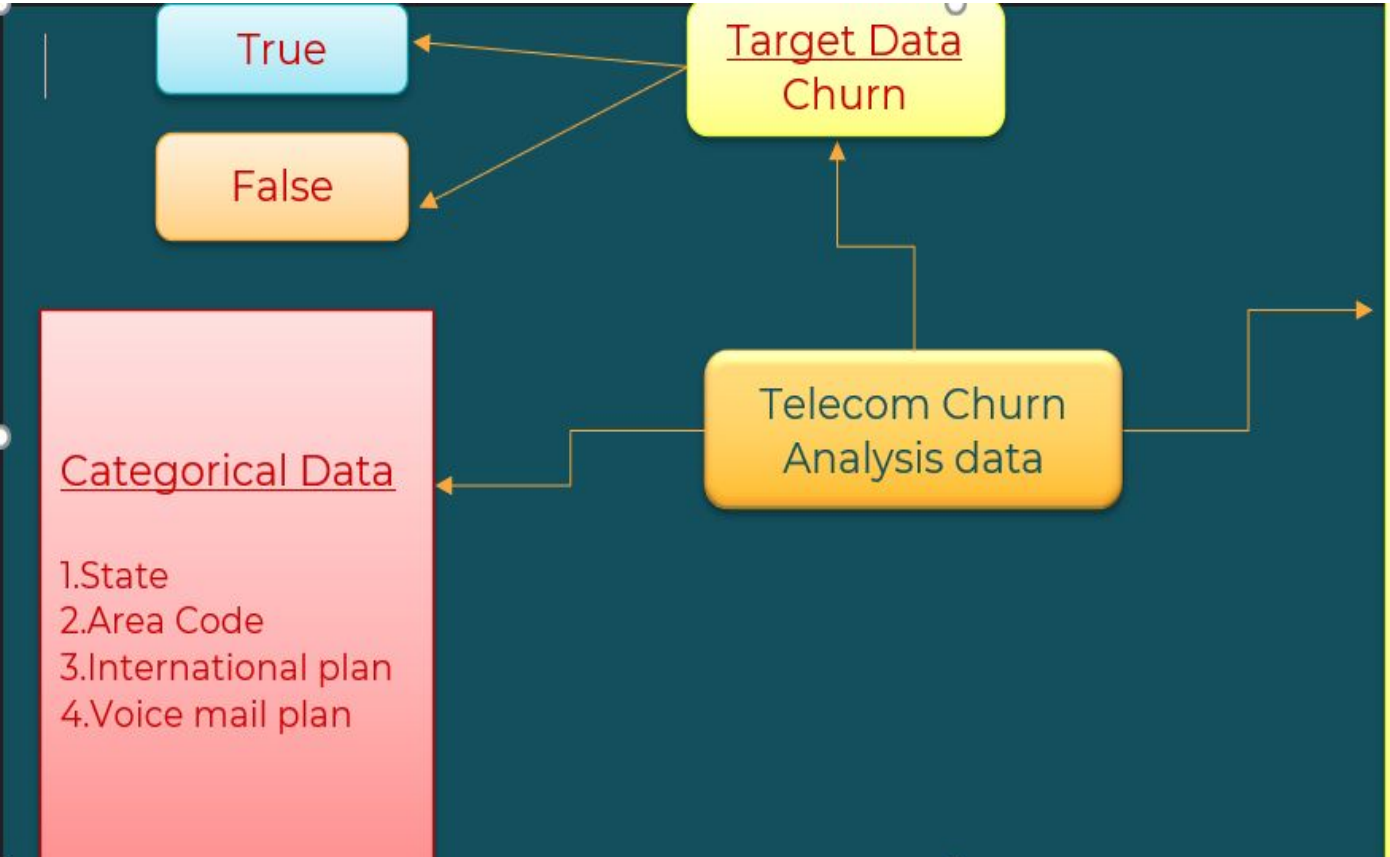
Data preprocessing-1 : In the first step we have checked for the presence of any Null value. We also analysed for the important and unnecessary features,like checking for data consistency .

Data preprocessing-2 : In this step we manually go through the each features and segregated the categorical and numeric/continuous data.Also added two extra columns for analysis named total charges and total minutes.

EDA : In this step we have done some exploratory data analysis on the selected features to find the various relations amongst the features with the targeted column.

Data Visualization : In the final step we created models using univariate, bivariate analysis and plotting to draw some conclusions or insights from the data. So that we can explain what are the important reasons behind the attrition and can recommend how to prevent that.

Data Summary



Numeric Data

1. Account length
2. Number vmail messages
3. Total day minutes
4. Total day calls
5. Total day charge
6. Total eve minutes
7. Total eve calls
8. Total eve charge
9. Total night minutes
10. Total night calls
11. Total night charge
12. Total intl minutes
13. Total intl calls
14. Total intl charge
15. Customer service calls

Checking for Null Value

#	Column	Non-Null Count	Dtype
0	State	3333 non-null	object
1	Account length	3333 non-null	int64
2	Area code	3333 non-null	int64
3	International plan	3333 non-null	object
4	Voice mail plan	3333 non-null	object
5	Number vmail messages	3333 non-null	int64
6	Total day minutes	3333 non-null	float64
7	Total day calls	3333 non-null	int64
8	Total day charge	3333 non-null	float64
9	Total eve minutes	3333 non-null	float64
10	Total eve calls	3333 non-null	int64
11	Total eve charge	3333 non-null	float64
12	Total night minutes	3333 non-null	float64
13	Total night calls	3333 non-null	int64
14	Total night charge	3333 non-null	float64
15	Total intl minutes	3333 non-null	float64
16	Total intl calls	3333 non-null	int64
17	Total intl charge	3333 non-null	float64
18	Customer service calls	3333 non-null	int64
19	Churn	3333 non-null	bool

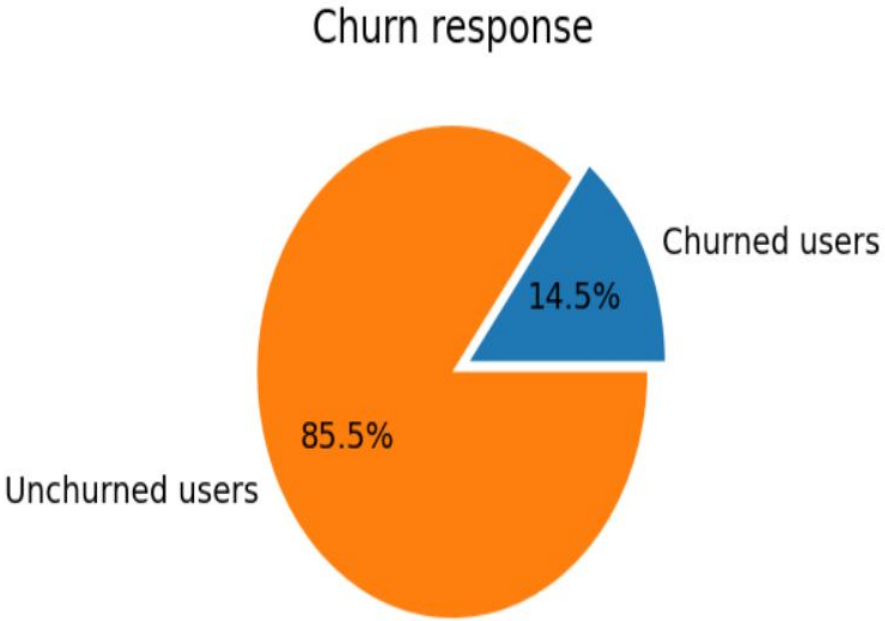
So There is no Null value present in the Dataset and data is Uniform in nature.

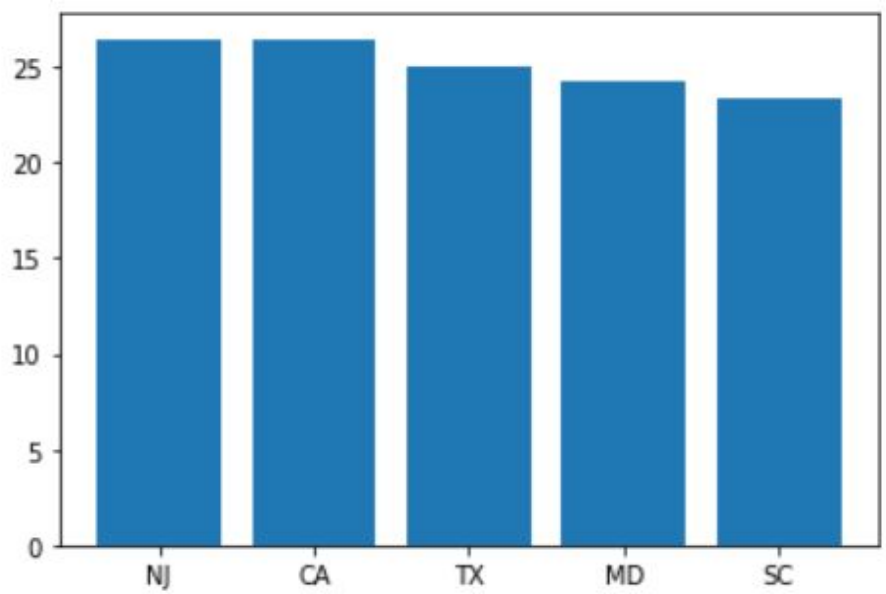
Checking Data Consistency and Missing or Garbage Values

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls
count	3333	3333	3333	3333	3333	3333	3333	3333
mean	101.0648065	437.1824182	8.099009901	179.7750975	100.4356436	30.56230723	200.980348	100.1143114
std	39.82210593	42.37129049	13.68836537	54.4673892	20.06908421	9.259434554	50.71384443	19.92262529
min	1	408	0	0	0	0	0	0
25%	74	408	0	143.7	87	24.43	166.6	87
50%	101	415	0	179.4	101	30.5	201.4	100
75%	127	510	20	216.4	114	36.79	235.3	114
max	243	510	51	350.8	165	59.64	363.7	170

count	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls
mean	3333	3333	3333	3333	3333	3333	3333	3333
std	17.08354035	200.8720372	100.1077108	9.039324932	10.23729373	4.479447945	2.764581458	1.562856286
min	4.310667643	50.57384701	19.56860935	2.275872838	2.791839548	2.461214271	0.753772613	1.315491045
25%	0	23.2	33	1.04	0	0	0	0
50%	14.16	167	87	7.52	8.5	3	2.3	1
75%	17.12	201.2	100	9.05	10.3	4	2.78	1
max	20	235.3	113	10.59	12.1	6	3.27	2
	30.91	395	175	17.77	20	20	5.4	9

Exploratory Data Analysis



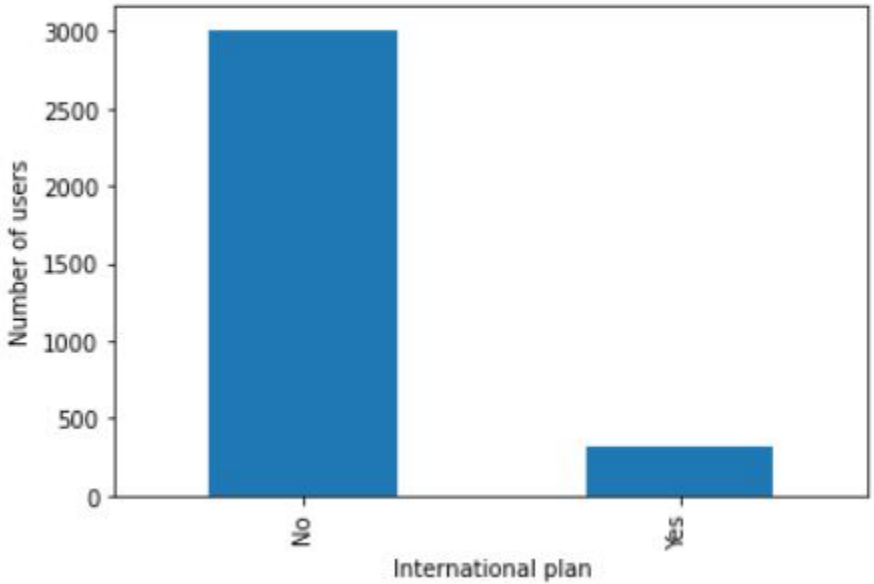


Finding total number of churned customers

Top 5 states with are higher churning rates are NJ, CA, TX, MD and SC and the churning rate is from 23 -26 %.

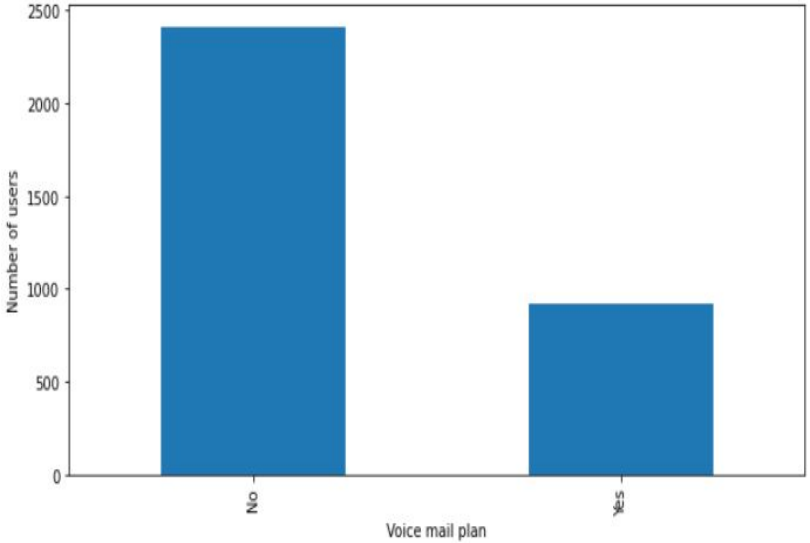
EDA: Univariate Analysis

1. International Plan users



Around 9.7 % of the customers use International plan

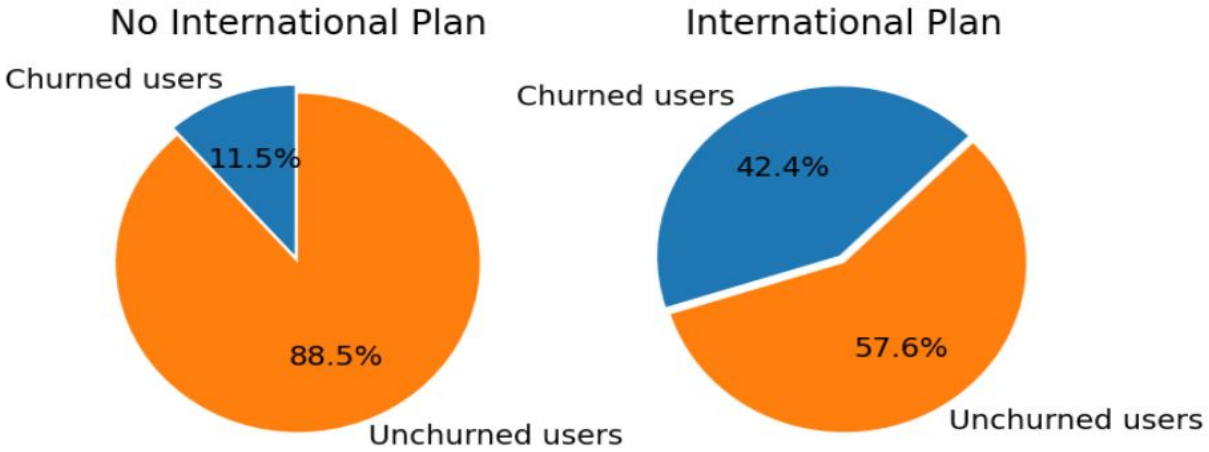
2. Voice mail users:



Around 28 % of the customers use Voice mail

EDA: Univariate Analysis

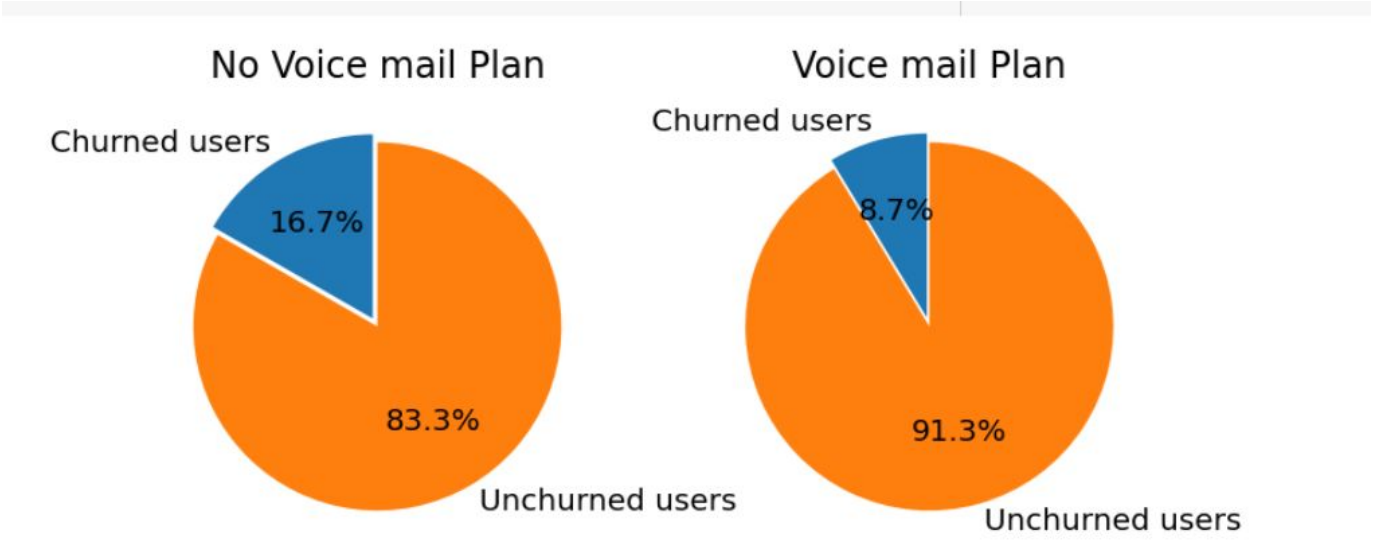
3.Dependency of Churn customer with the International Plan



The figure above shows the % of churn across the people who have intl plan is approx 4 times of those who are not on international plan

EDA: Univariate Analysis

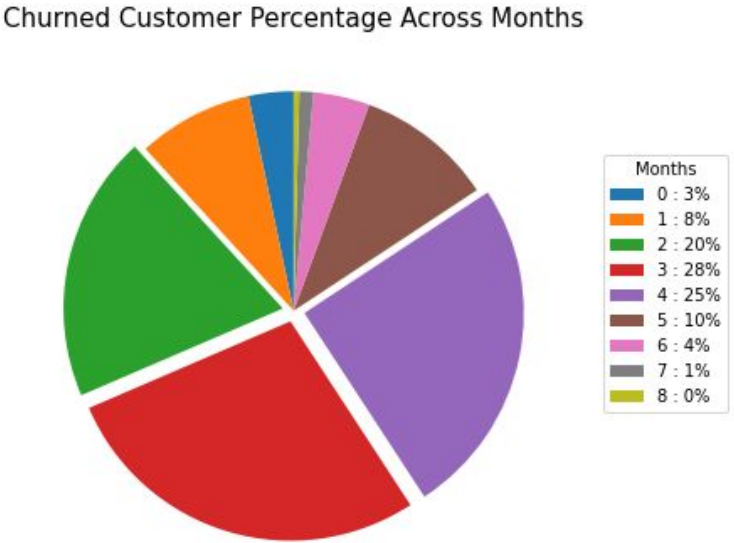
4.Dependency of Churn customer with the Voicemail Plan



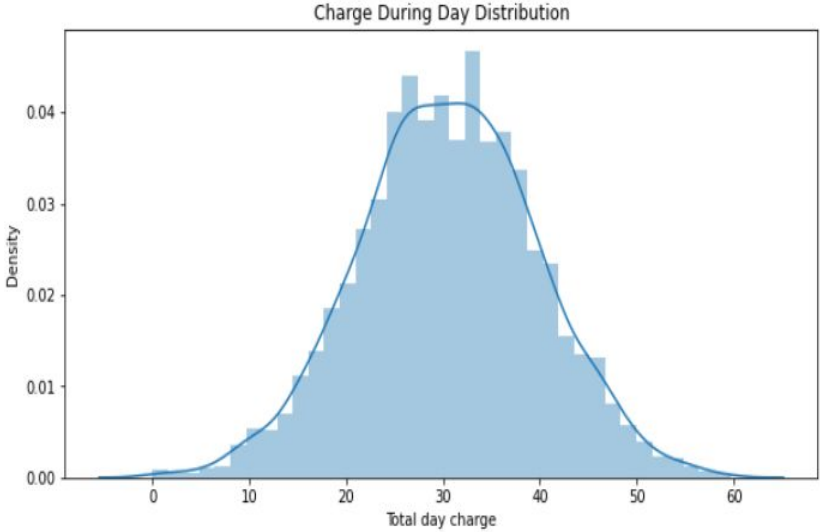
The figure above shows the % of churn across the people who have voicemail plan is approx 2 times of those who are not on voicemail plan..

EDA: Univariate Analysis

5. Churn customer percentage across months:



6. Total Day charges

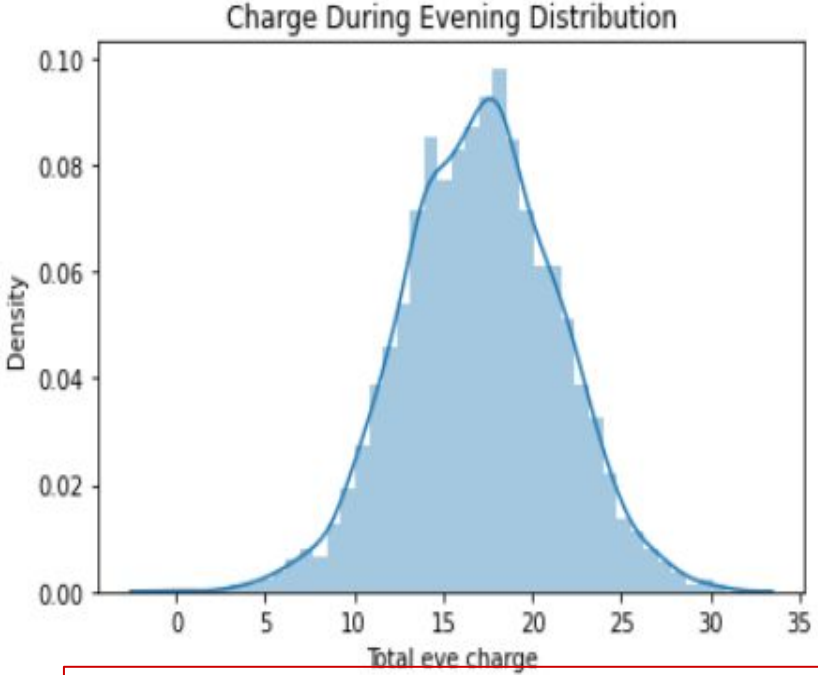


Majority of churning of customers occurs in the interval of 2 to 4 months which in terms of account length is between 84 days and 140 days.

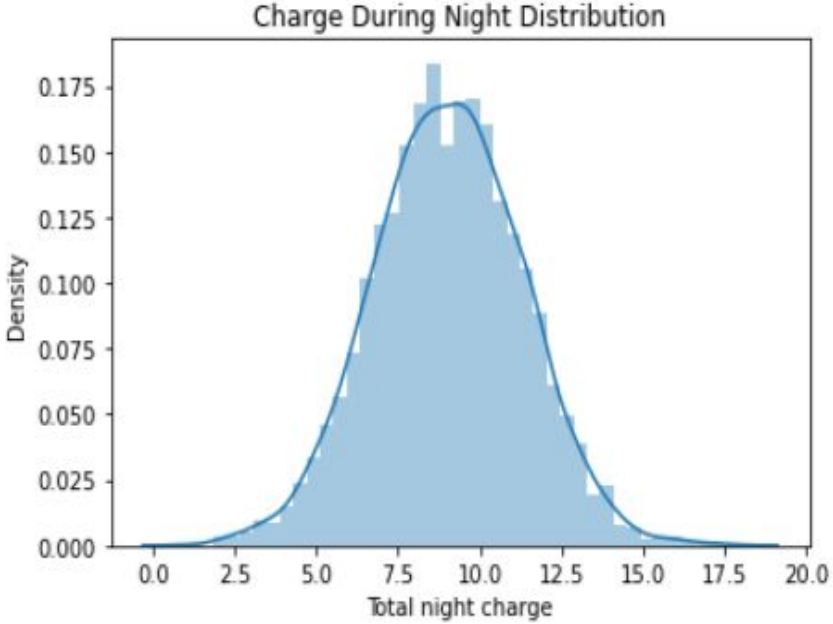
The distribution of Total Day charges, total day minutes is same as they are highly correlated

EDA: Univariate Analysis

7. Total evening charges



8. Total night charges



The call distribution and charge distribution is similar for both evening and night

EDA: Univariate Analysis

9. Relation between Total call charges and Total call duration among Day, Evening and Night hours

	Day	Evening	Night
Median Tariff in units	30.50	17.12	9.05
Median Call Duration in minutes	179	201.4	201.2

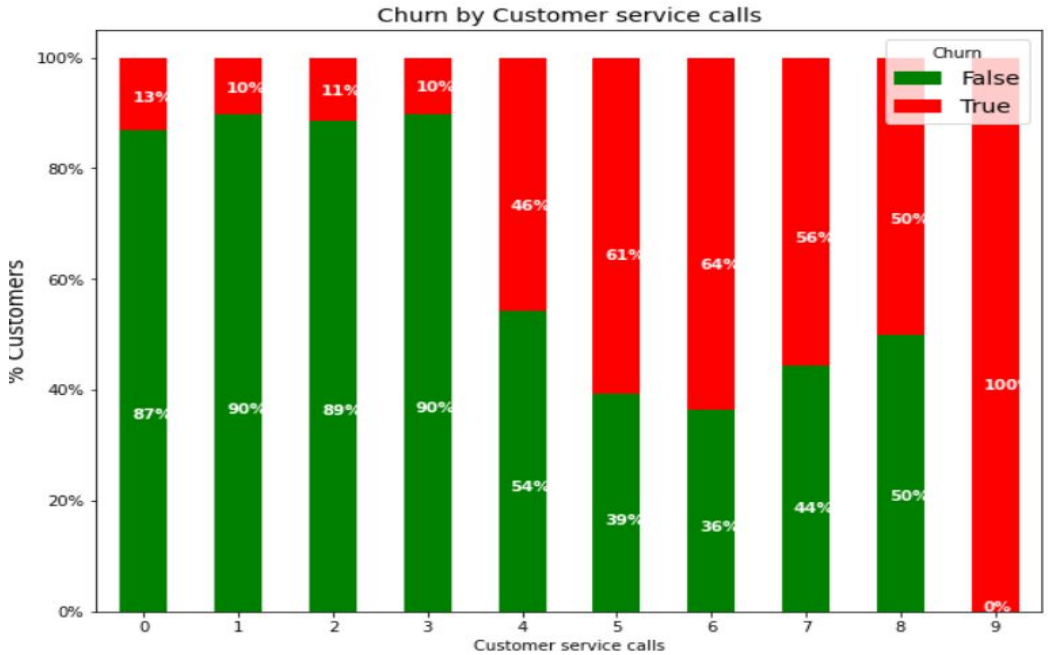
EDA: Univariate Analysis

10. Relation between Area code and Churn

Area Code	Churn False	Churn True	Percentage
408	716	122	14.56
415	1419	236	14.26
510	715	125	14.88

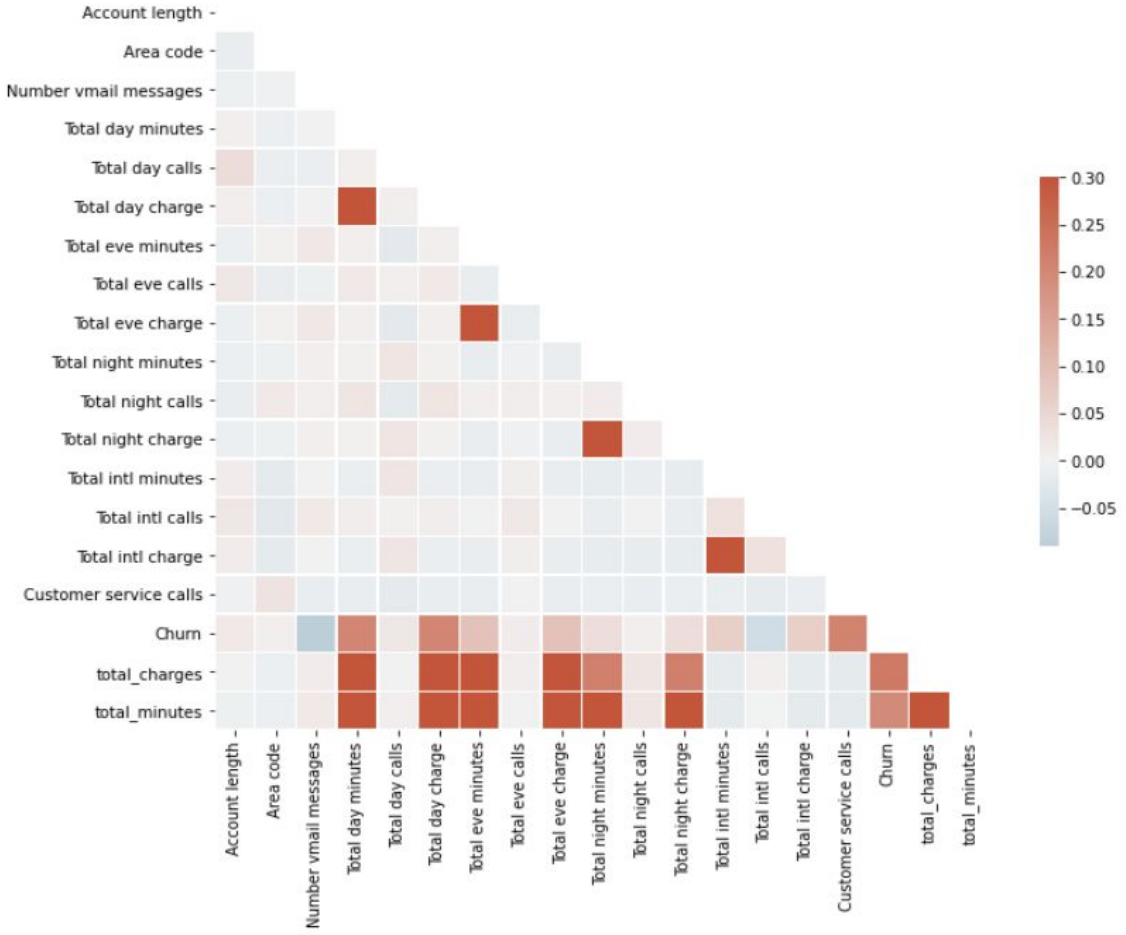
From the Churn percentage it can be deduced that no significant relation exists between Area code and Churn customer

11.Churn vs Customer Service calls



A significant spike in churn can be seen from the above figure as customer calls increases beyond 3 customer calls.

EDA: Correlation among the features

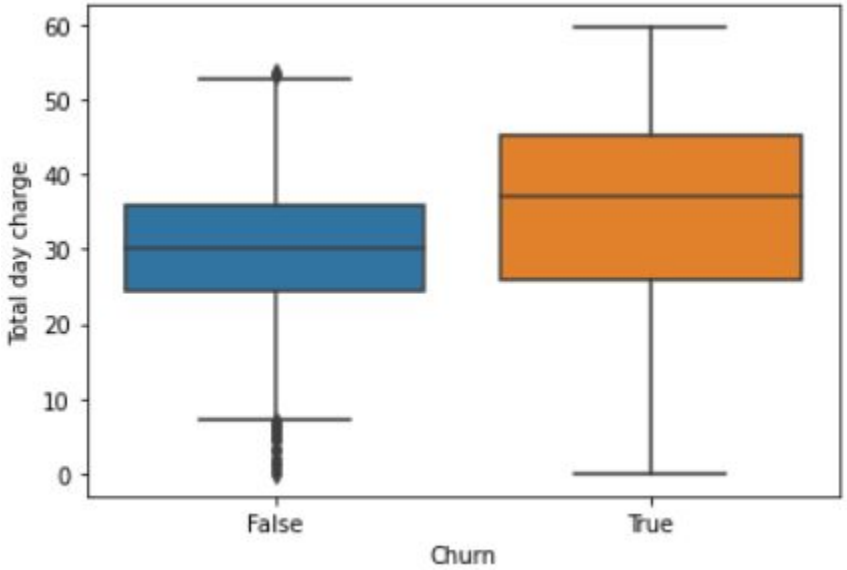


Some relations can be found as

- ➔ Total day charge to Total Day minutes is at 1
- ➔ Total evening charge to Total evening minutes is at 1
- ➔ Total night charge to Total night minutes is at 1
- ➔ Total Charge to Churn is at 0.21
- ➔ Total minutes to Churn is at 0.21
- ➔ Customer service calls to Churn is at 0.21

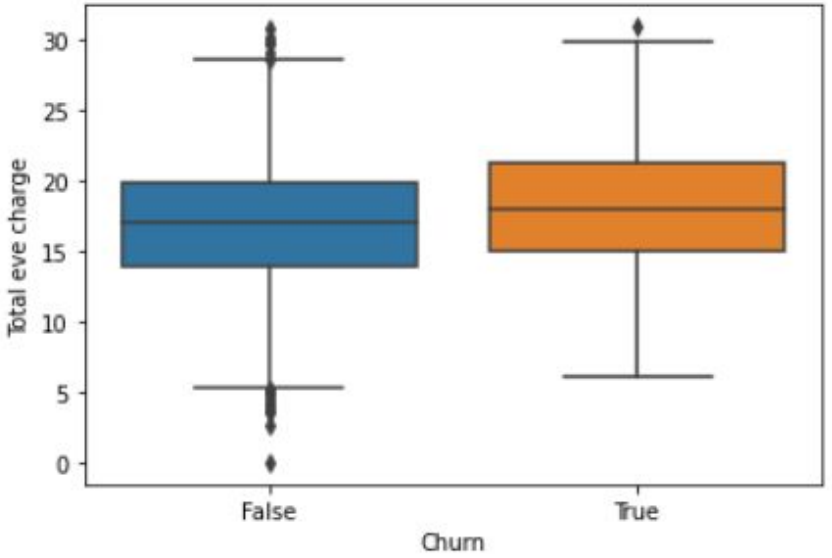
EDA: Univariate Analysis

12.Total day Charges



Total day charges are more in case of Churned customer

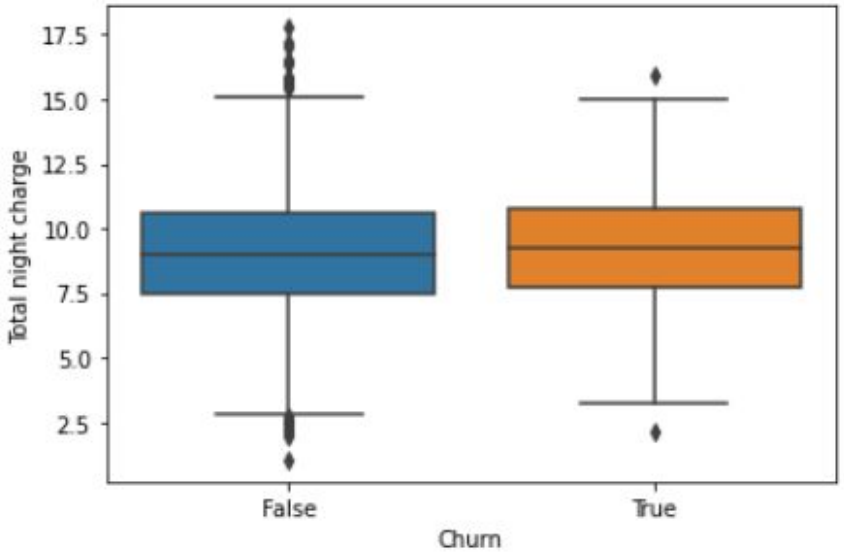
13. Total Evening Charges



Total evening charges are slightly high in case of Churned customer

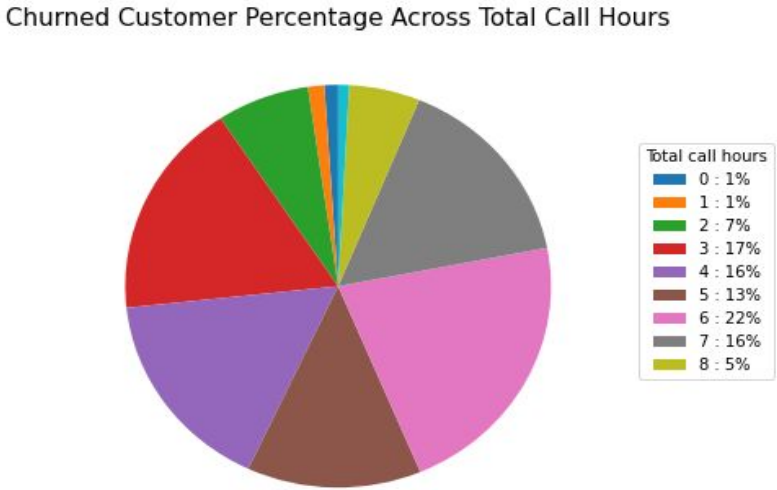
EDA: Univariate Analysis

14.Total night Charges



Total night charges are slightly more in case of Churned customer

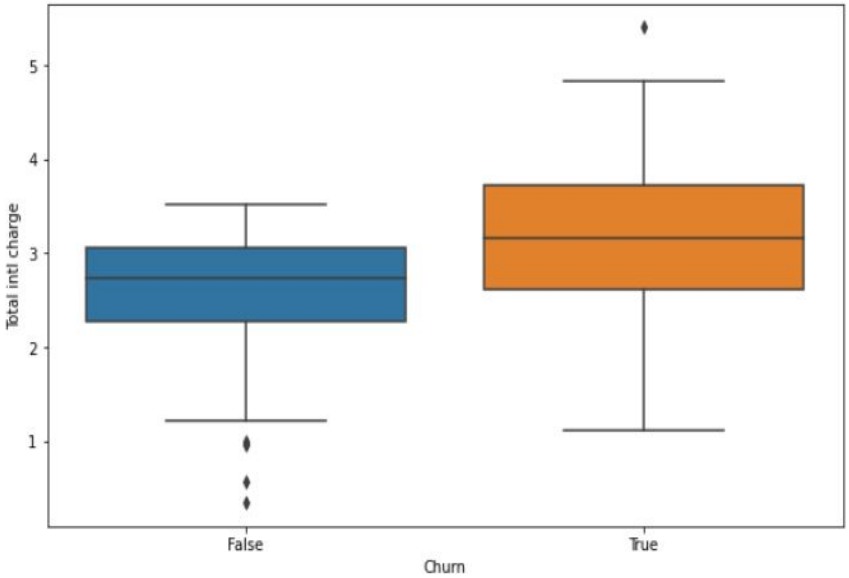
15. Percentage of Churned Users vs total call hours



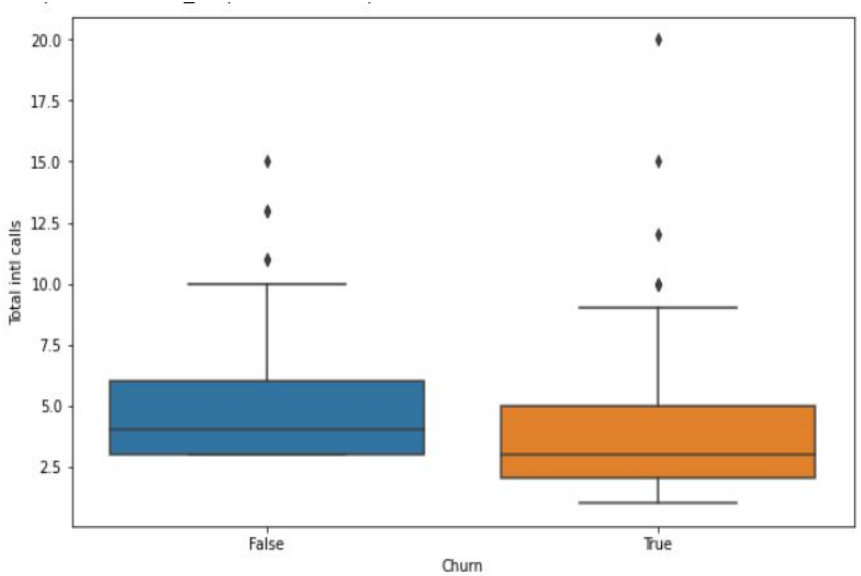
Customers who had total call hours equal to 6 hours had the highest churn percentage

EDA: Univariate Analysis

16.Total international charges



17. Total international calls

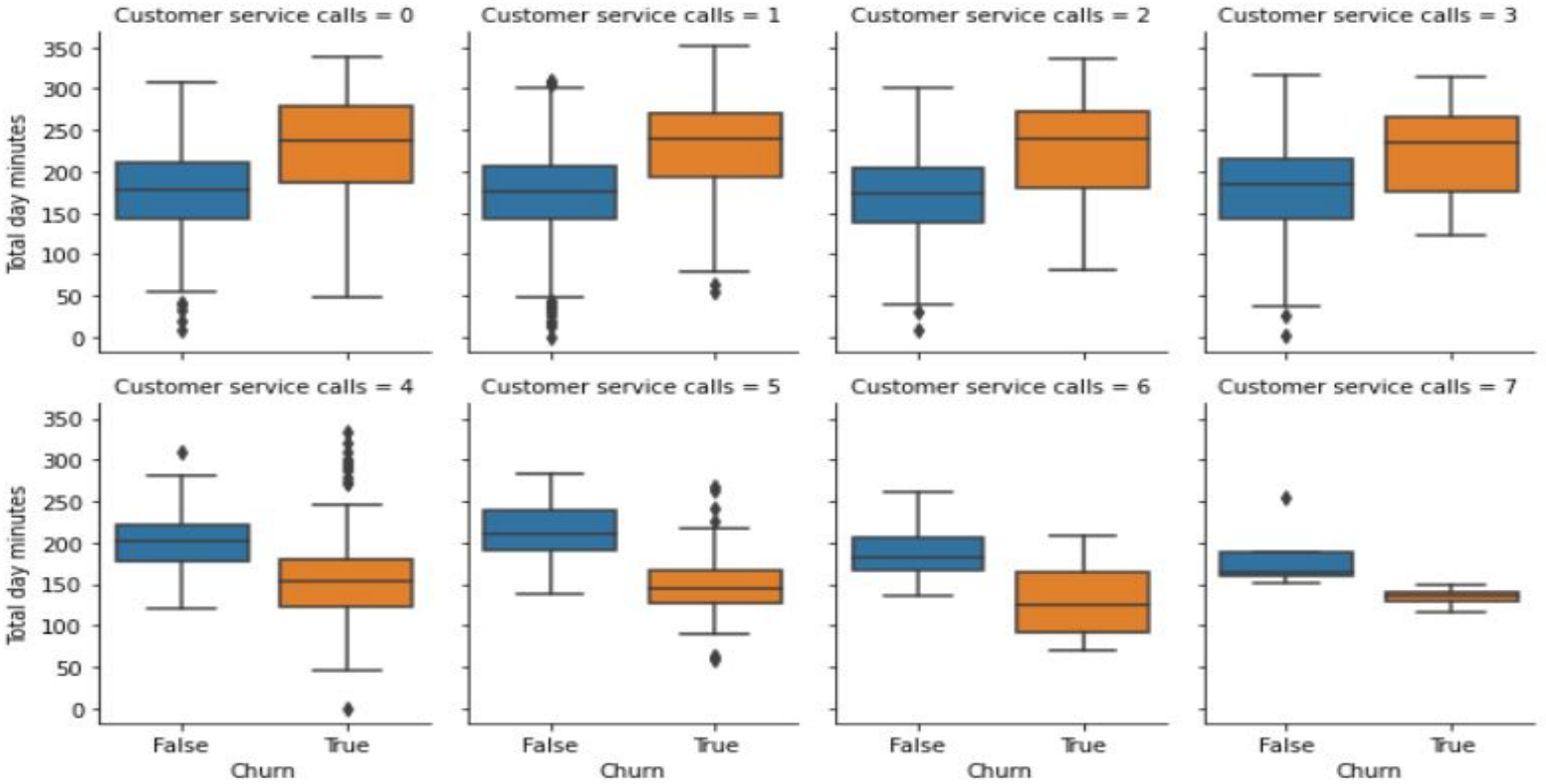


Total international charges are more in case of Churned customer

Churned customers are making less international calls

EDA: Bivariate Analysis

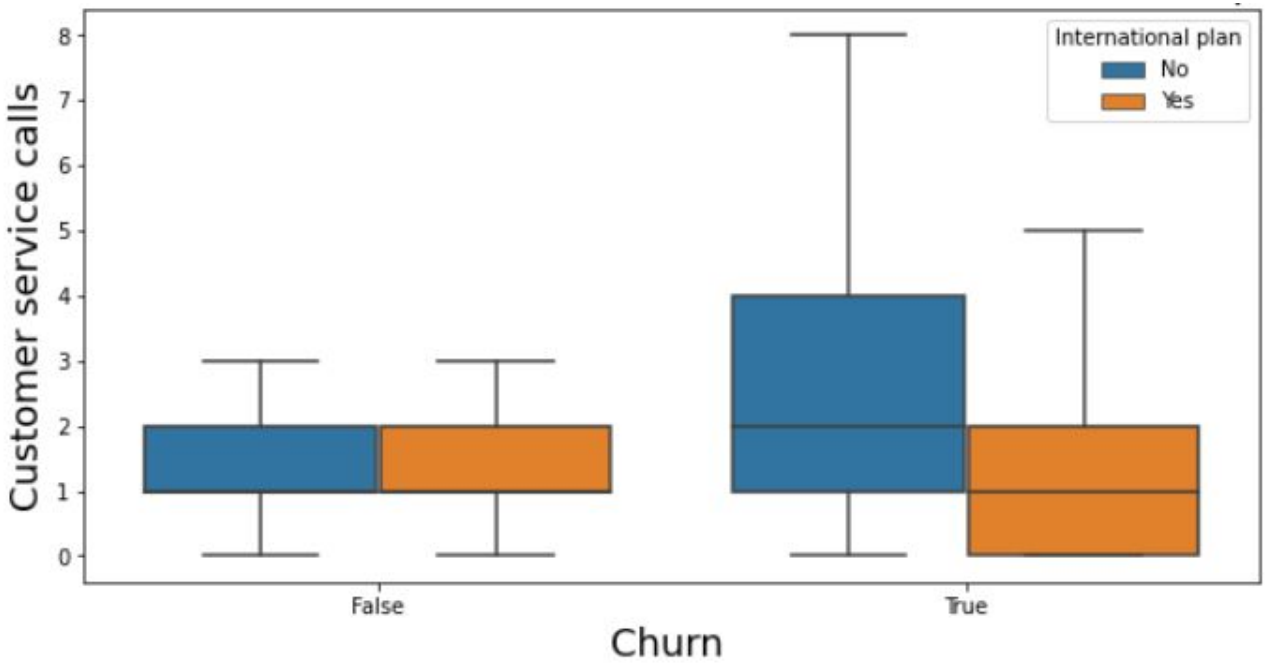
18.Analysis of Total day minutes with Churn when Customer service calls increases



EDA: Bivariate Analysis

19.Relation between customer service calls and Churn for those who have international plan

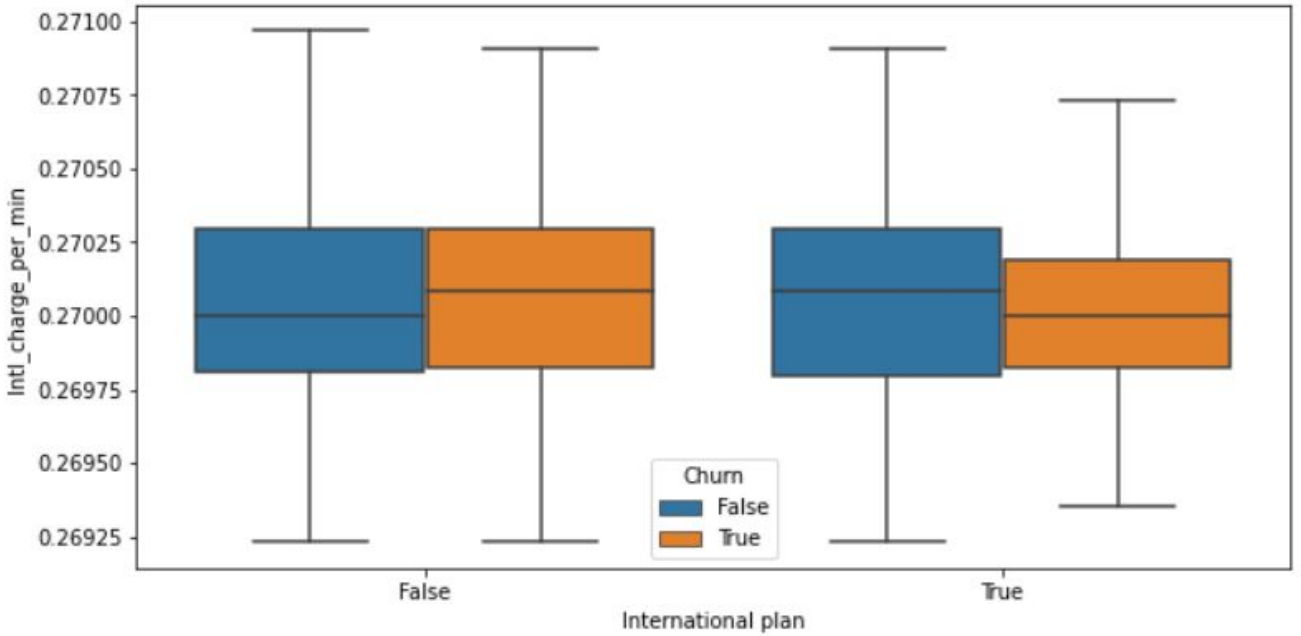
.



EDA: Bivariate Analysis

20. International charge/minute vs International plan taking churn into consideration

.



Inference

1. Based on the analysis it can be said that the Churn customers are paying significant tariff mostly during the day hours even though the duration of call minutes is the least among three.

2. Customer churn increases as the the number of Customer call increases (goes beyond 3).

3. There is no significant effect of Area code on attrition as the churn ratio is almost same for all area codes. The Churn percentage is 25-26% for top five states.

4. Churn ratio on the International plan is 4 times higher than those who have not any international plan

5. No advantage of having International plan for the customers as the International Charges/min for International plan users and non-users are almost same.

6. The median values of the Total number of day, evening, night calls are roughly the same across churn and non-churn population, however the median values of the duration of day, evening and night calls and charges for that across the churn population is slightly higher than the non-churn population.

Factors responsible for Churning

1. Higher call rate during the day hours.

2. Increase in customer service calls triggered churning of customers.

3. International plan facility is not effectively optimized or structured properly which leads to customer dissatisfaction

Final Recommendations

1. Charging rate should be managed especially for the day time and long duration call. Telecom companies have to analyze which model can be profitable for them even after reducing the tariff.
2. Telecom Companies can provide incentives for talking more number of minutes.
3. International plan needs to be improved. International call rates can be optimized.
4. Grievance redressal for users should be tackled within 2-3 Customer service calls.
5. Focus on top 5 states where churning rate is higher. An improved and versatile network facility can be set up.

Q & A

“Without data you’re just another person with an opinion”..... W. Edwards Deming