Capstone Project - 1

Exploratory data analysis- Telecom Churn Analysis

Akifuddin Kashif | Zeeshan Ahmed

Agenda

In this project, we will perform Exploratory Data Analysis on Telecom Churn Dataset and will get some meaningful insights.

What is Exploratory Data Analysis?

- Exploratory Data Analysis refers to the critical process of performing initial investigations on data so on discover patterns, to spot anomalies, to test hypothesis and to see assumptions with the assistance of summary statistics and graphical representations.

What is Churning?

- Customer churn (also known as customer attrition) refers to when a customer (player, subscriber, user, etc.) ceases his or her relationship with a company. Online businesses typically treat a customer as churned once a particular amount of time has elapsed since the customer’s last interaction with the site or service. The full cost of churn includes both lost revenue and the marketing costs involved with replacing those customers with new ones. Reducing churn is a key business goal of every online business.

About Our Dataset

This data set comprises of information regarding the churning of a customer for a Telecom Company. It includes information such as number of minutes spent on a call, amount charged, total number of calls, whether or not it was an international call at different times of a day namely day, evening(eve),night.It shows for how long an individual has been loyal to the company in the form of Account Length, Customer's choice on voice plan and number of customer service calls recieved. it also includes data pertaining to customer's location i.e, State and Area Code

Let’s know our data

State: Name of the State
Account length: Length of the account

Area Code:Area where customer resides

International Plan: If customer opted for International Plan

Voice mail plan: If customer opted for Voice mail Plan

Number vmail messages: Number of voicemails recieved

Total day minutes: minutes spent on call during day time
Total day calls : Number of calls made during day time

Total day charge : Amount charged for calls during day time

Total eve minutes: minutes spent on call during evening time

Total eve calls :Number of calls made during evening time

Total eve charge : Amount charged for calls during evening time

Total night minutes : minutes spent on call during night time

Total night calls : Number of calls made during night time

Total night charge : Amount charged for calls during night time

Total intl minutes : minutes spent on internatinal calls

Total intl calls : Number of calls made internationally

Total intl charge : Amount charged for international calls

Customer service calls : Number of customer service calls recieved

Churn : If customer churned

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 20 columns):
 #   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
dtypes: bool(1), float64(8), int64(8), object(3)
memory usage: 498.1+ KB
```

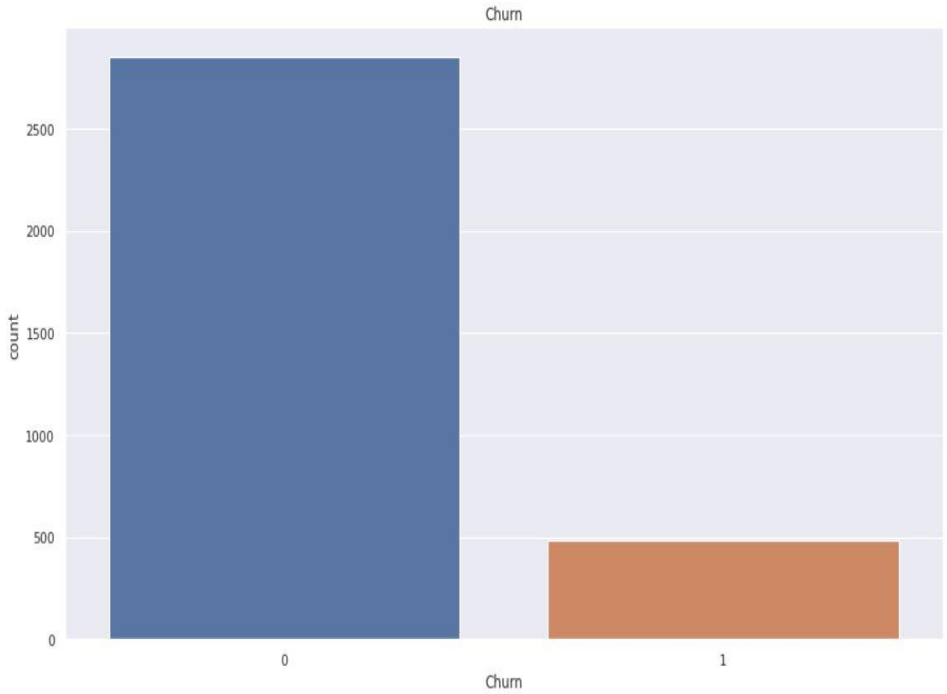
Exploratory Data Analysis

Data cleaning and Manipulation

- While checking for null values, it was found that there weren’t any among the columns of the dataset.
- There were two ways in which we handled outliers namely Visualizing using BoxPlot and IQR(Inter Quartile Range) method.
- BoxPlot : We plotted all the column boxplots using matplotlib and removed the outliers which were out of bounds.
- IQR method : We defined a function to identify outliers using lower and upper bounds of column values. We successfully handled the outliers by removing them through boxplots and verifying the presence of outliers by IQR method and noticed that there weren’t any.

Count of customers

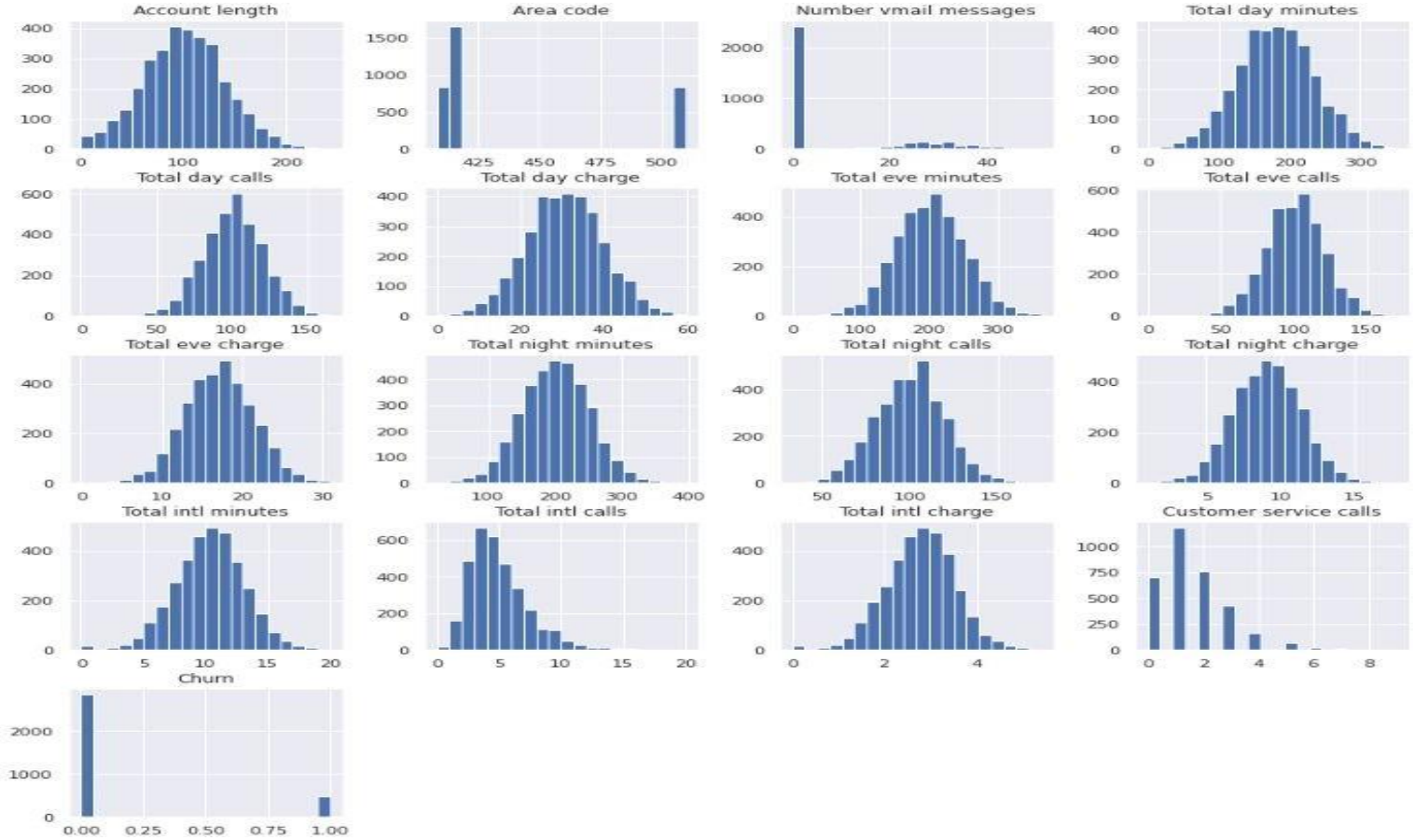
Churned= 1 , Not Churned= 0



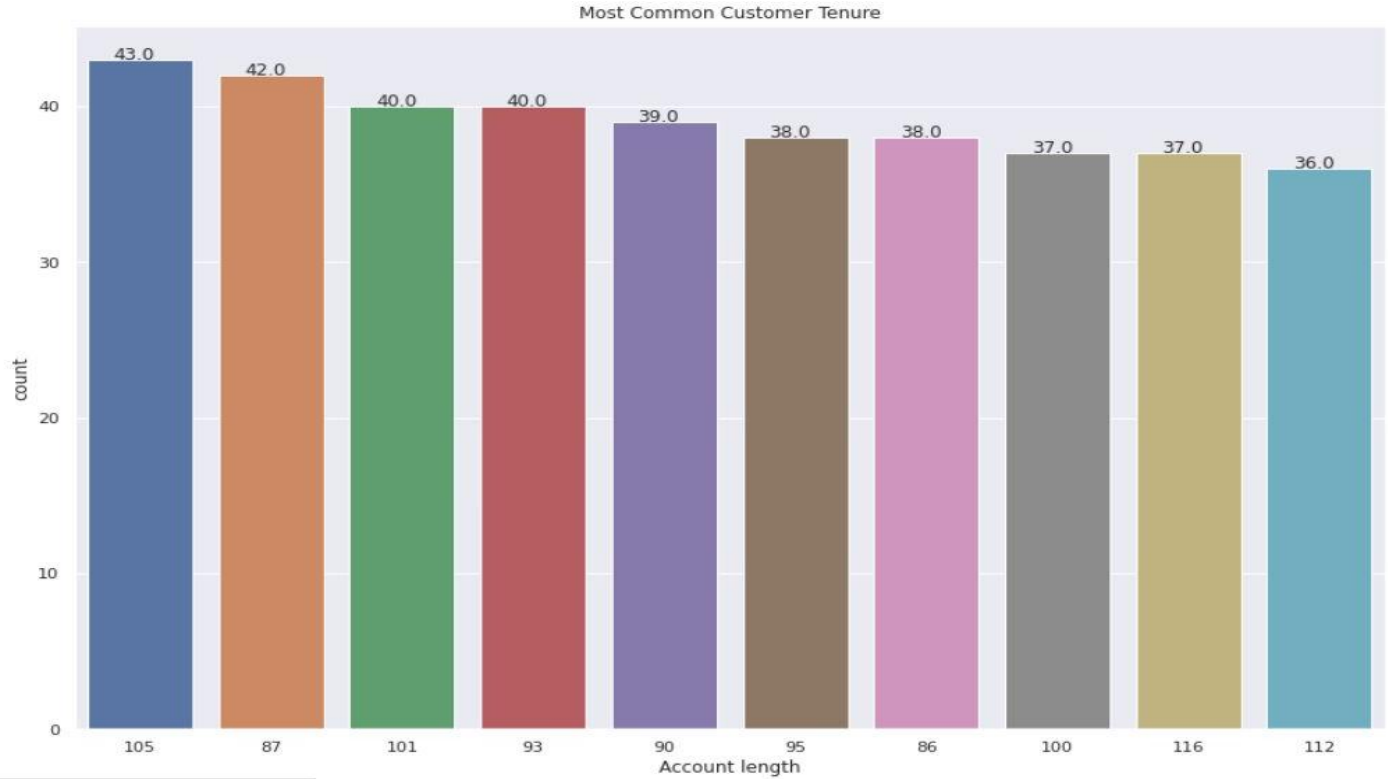
Top 5 Customer count by State

```
WV     106
MN      84
NY      83
AL      80
WI      78
Name: State, dtype: int64
```

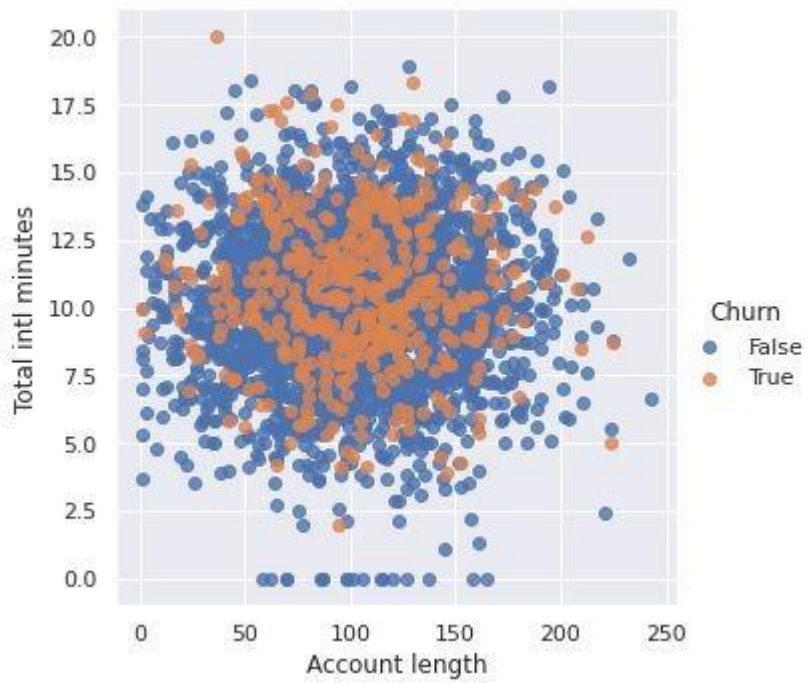
Distribution of Columns



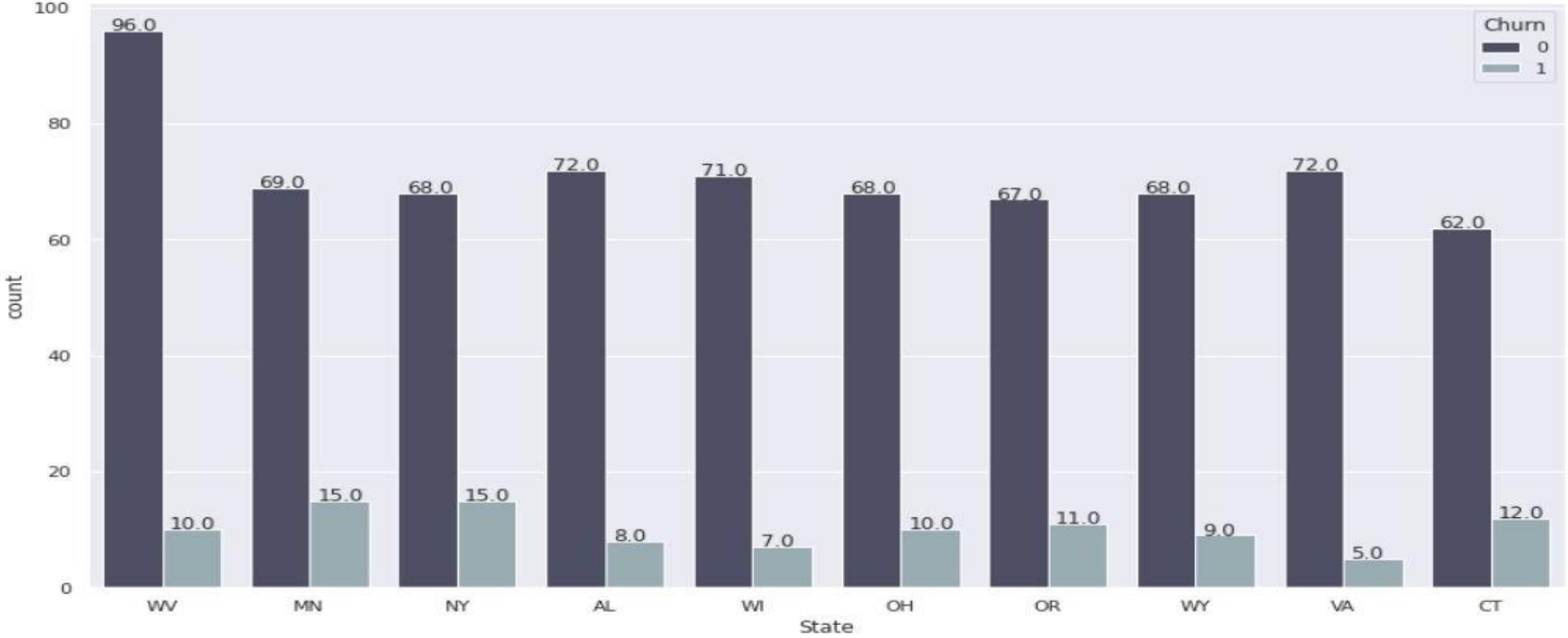
Top 10 most common Account length(in months)



Churn with International Plan and Account Length

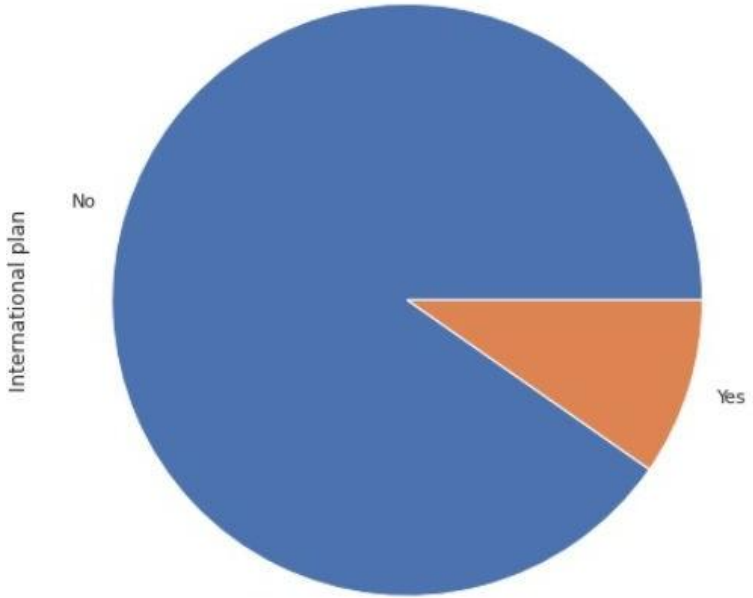


Top 10 States in terms of Churn Activity.

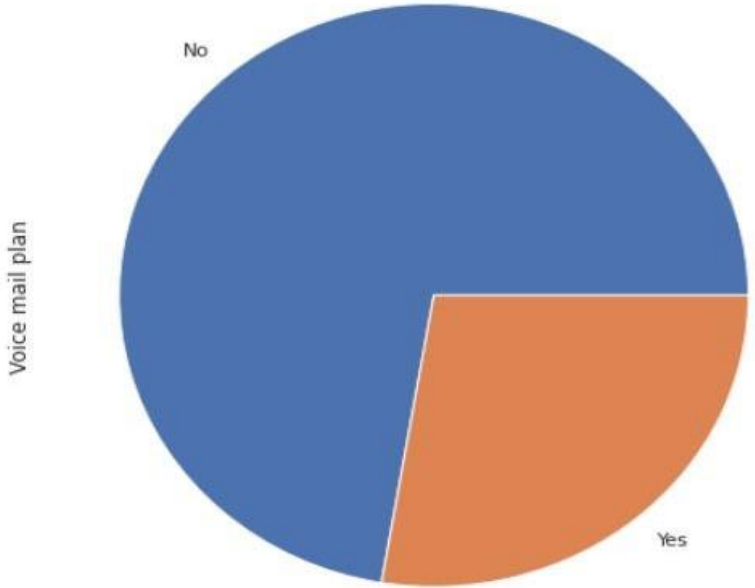


Type of Plans(International and Voice Mail) customers preferred.

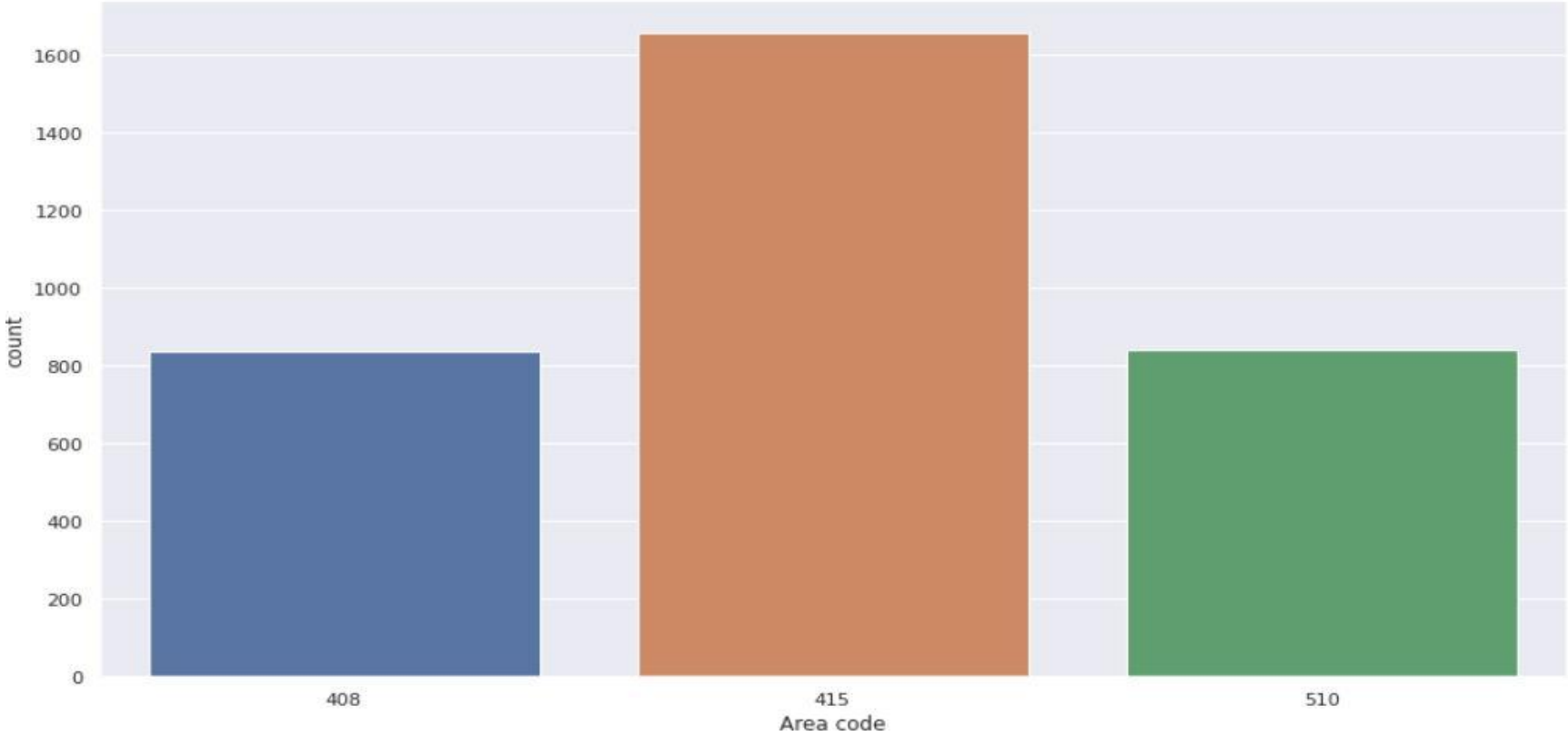
International Plan



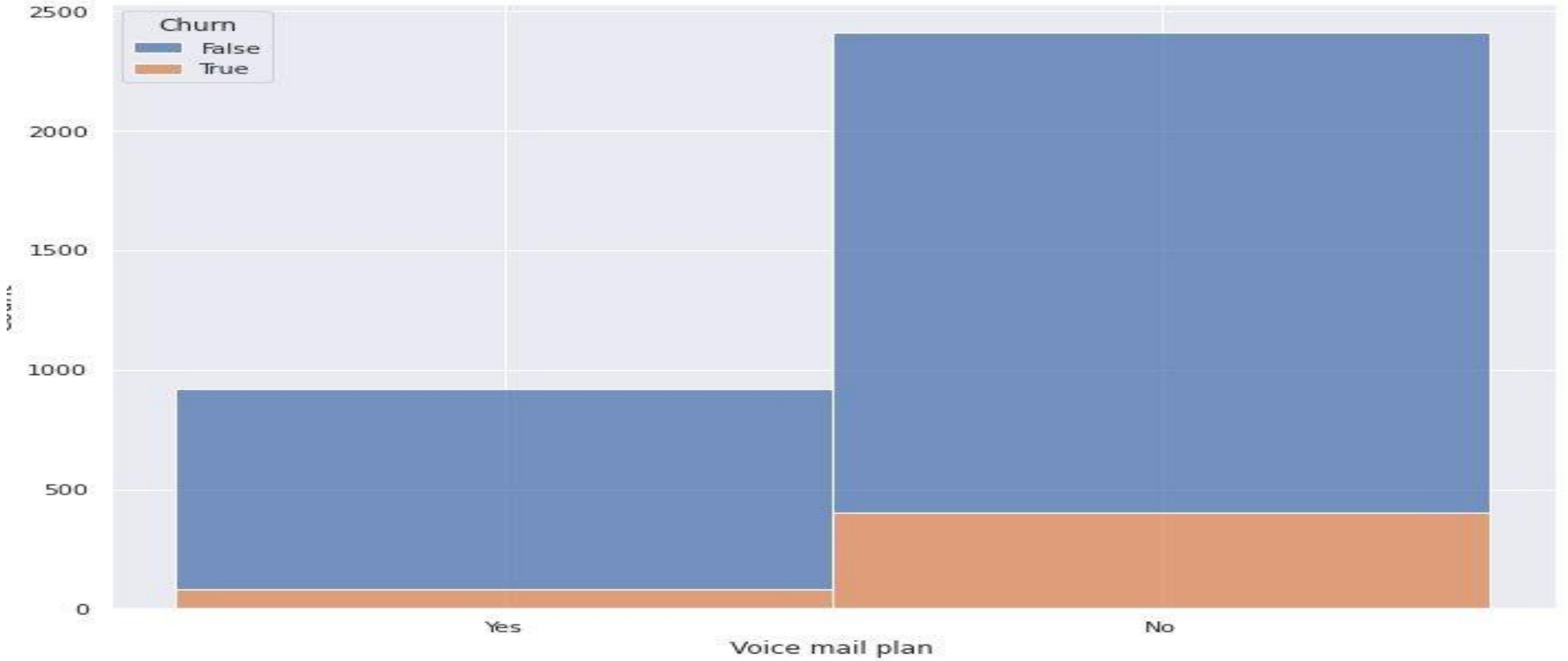
Voice Mail Plan



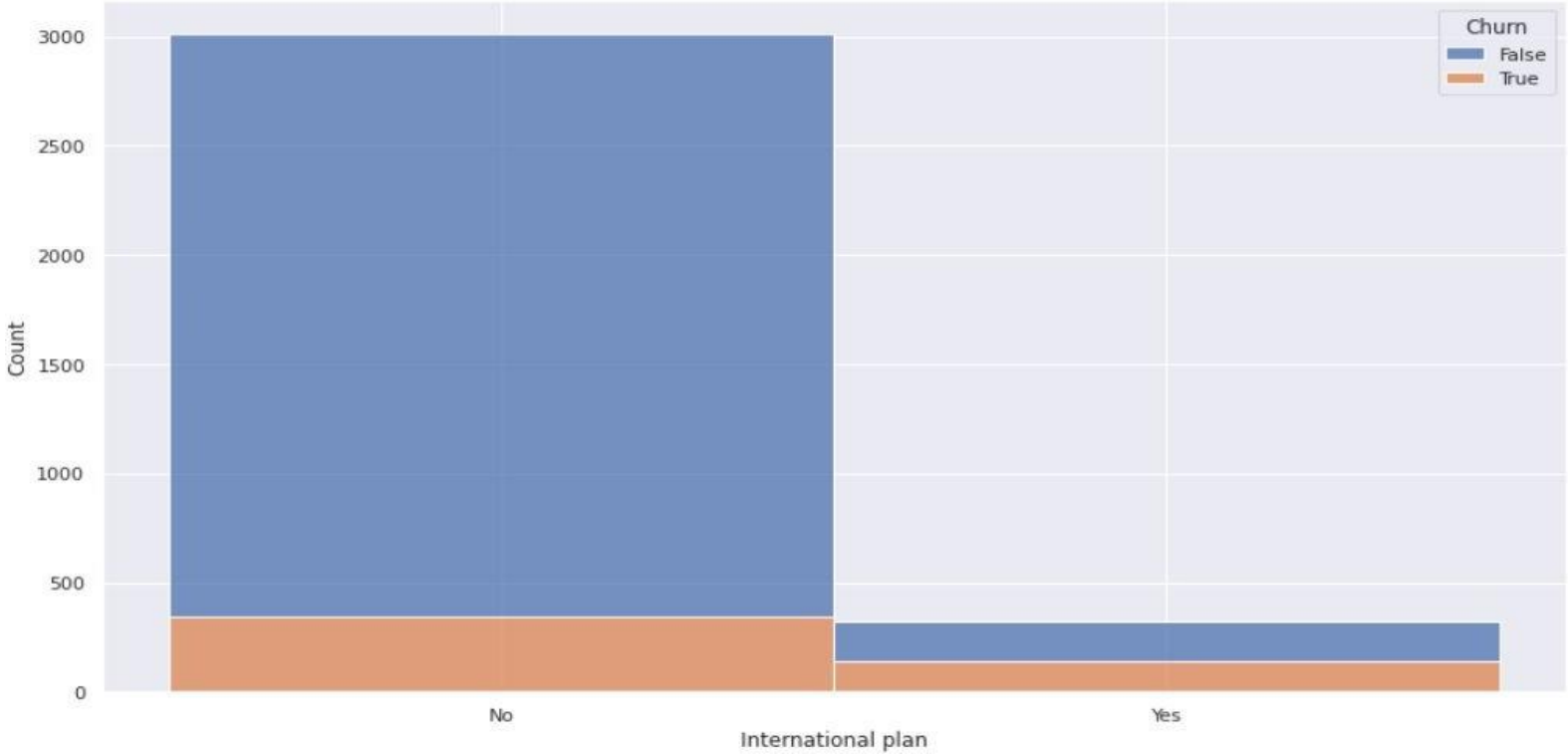
Customer distribution by Area Code



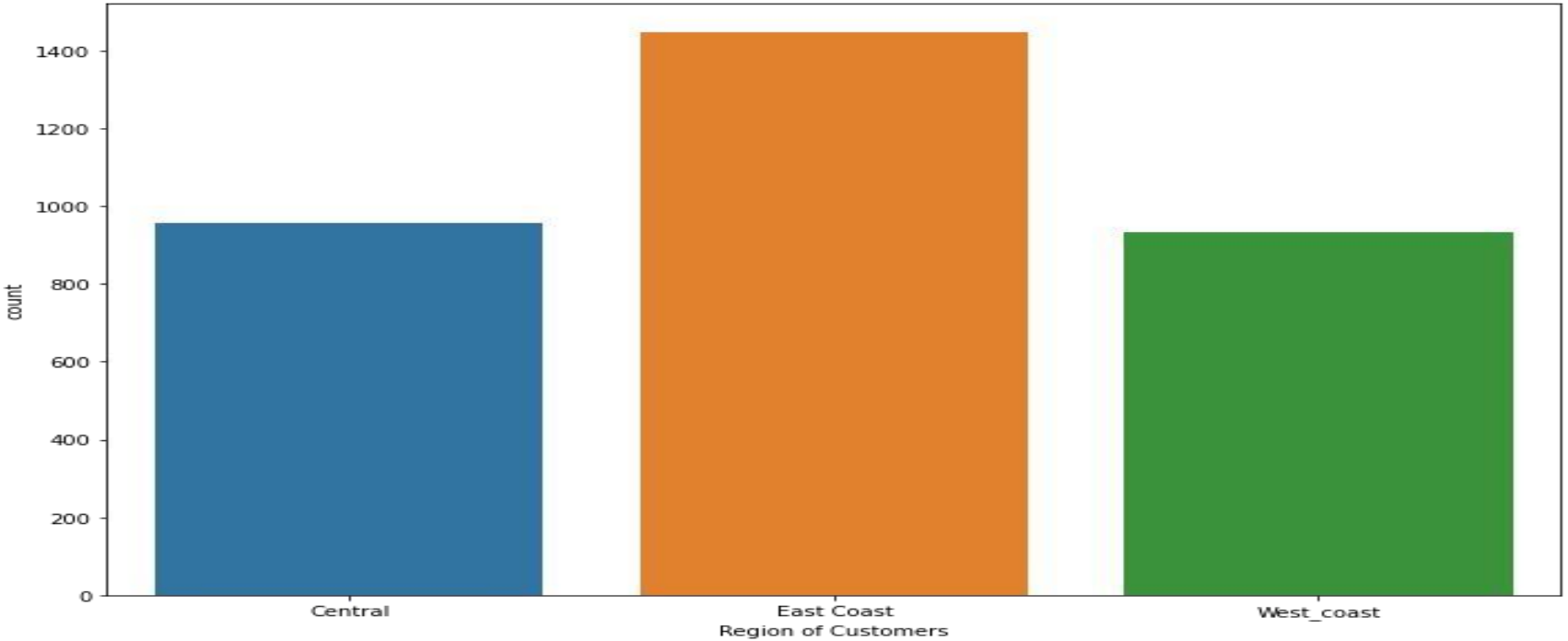
Customer Churn based on VoiceMail Plan



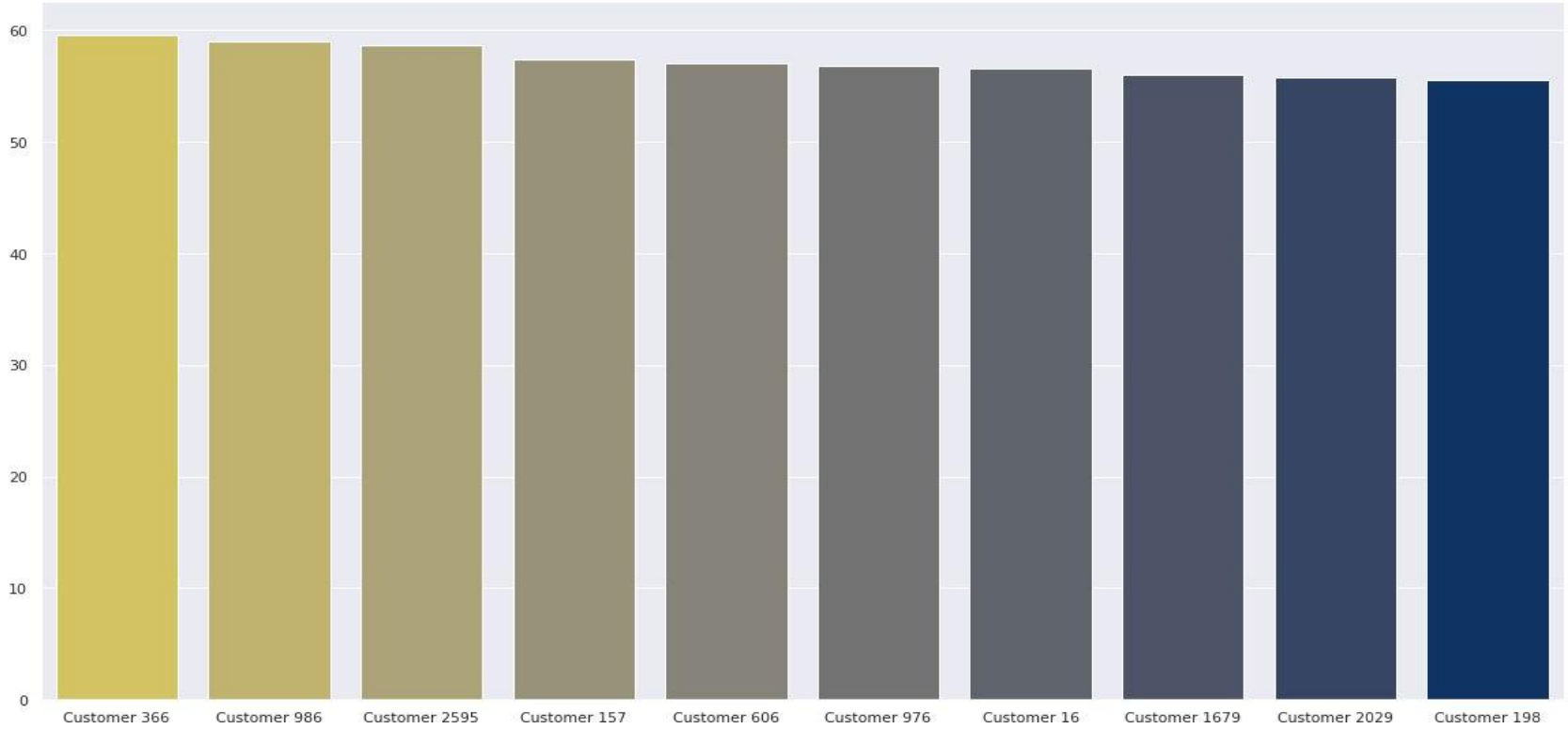
Customer Churn based on International Plan



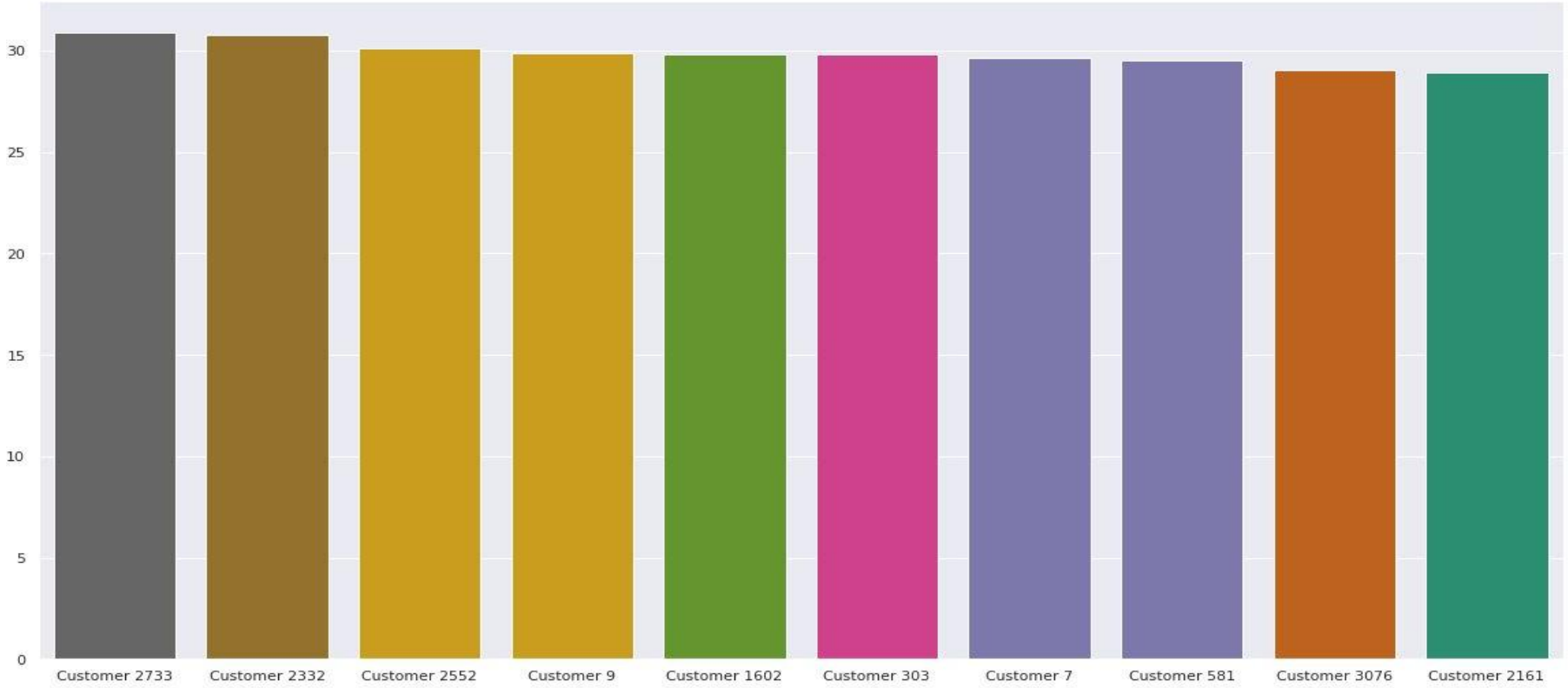
Churn Activity in different regions



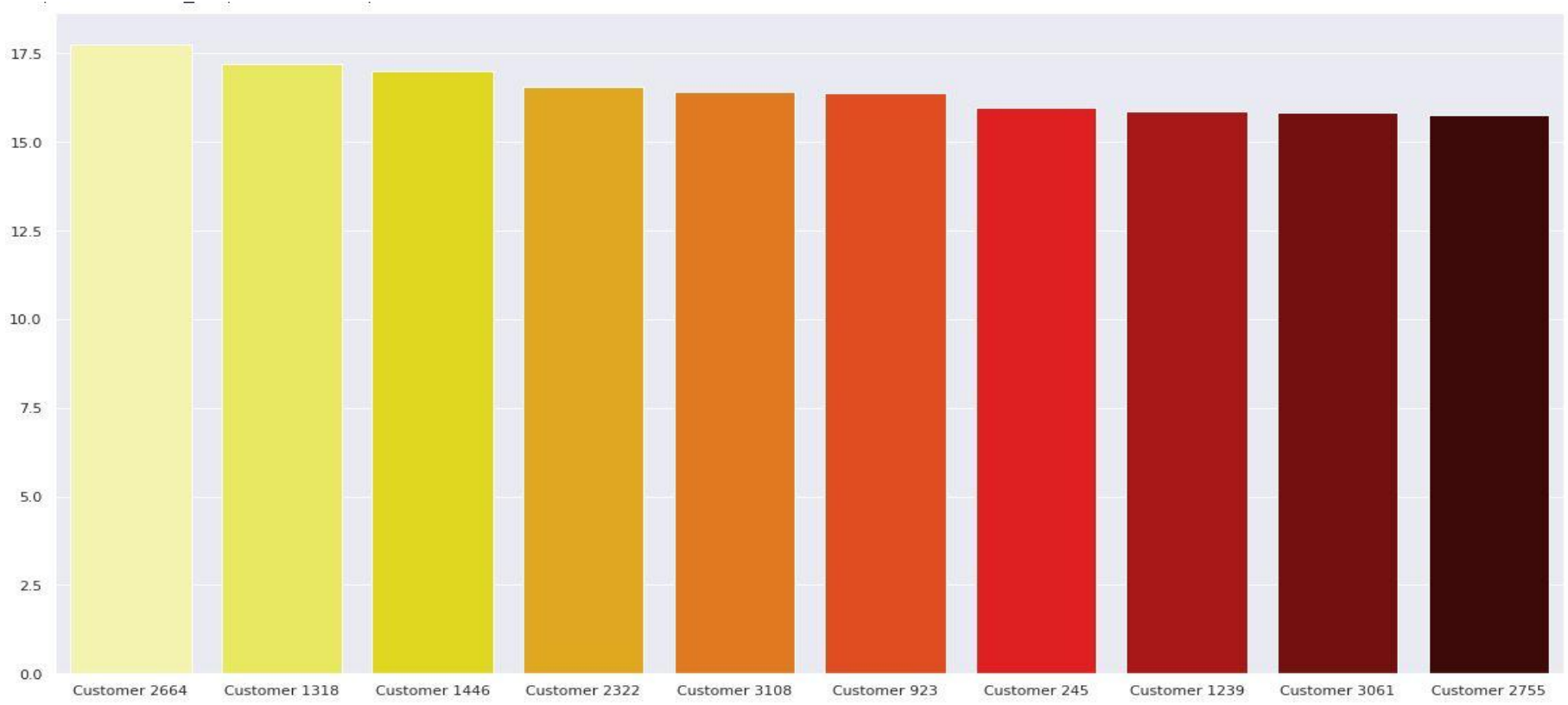
Top 10 Customers who spent the most during overall day time



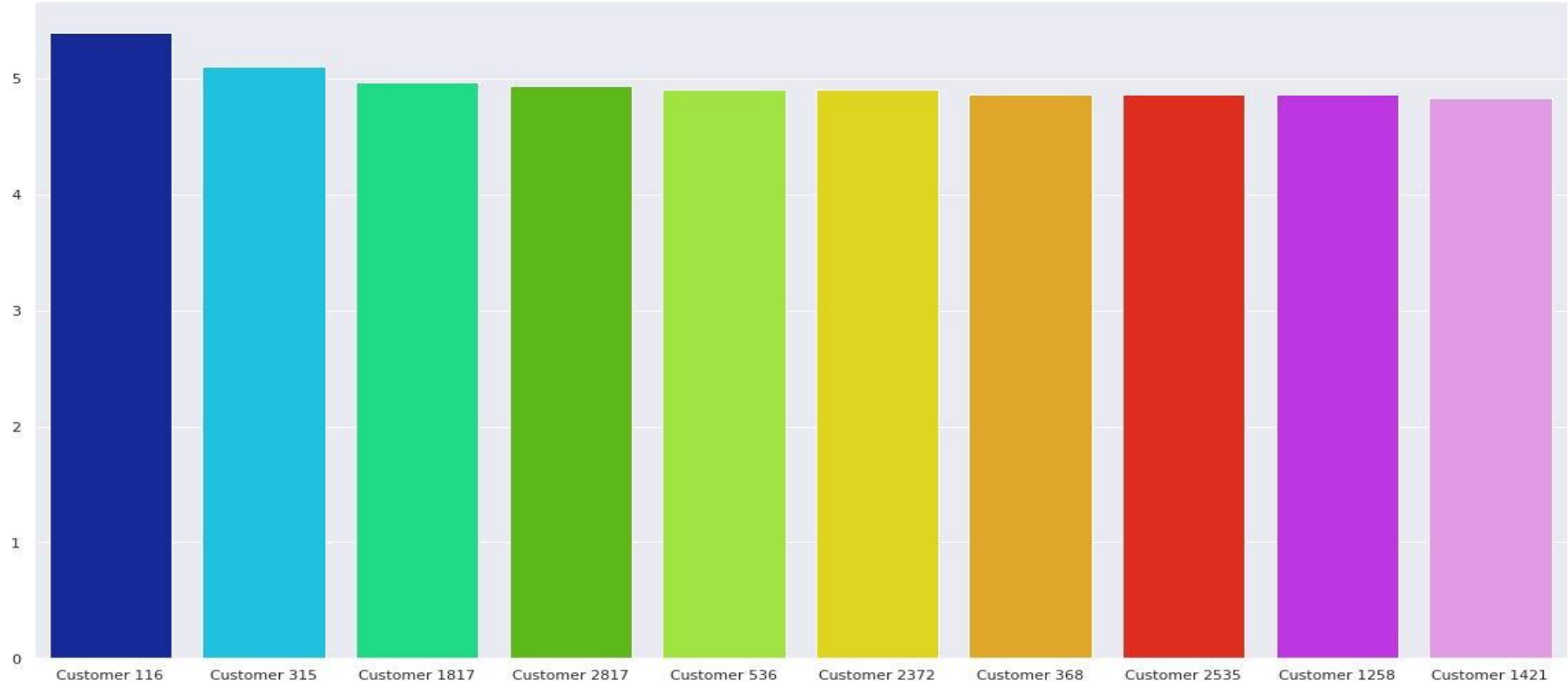
Top 10 Customers who spent the most during overall evening time



Top 10 Customers who spent the most during overall Night time

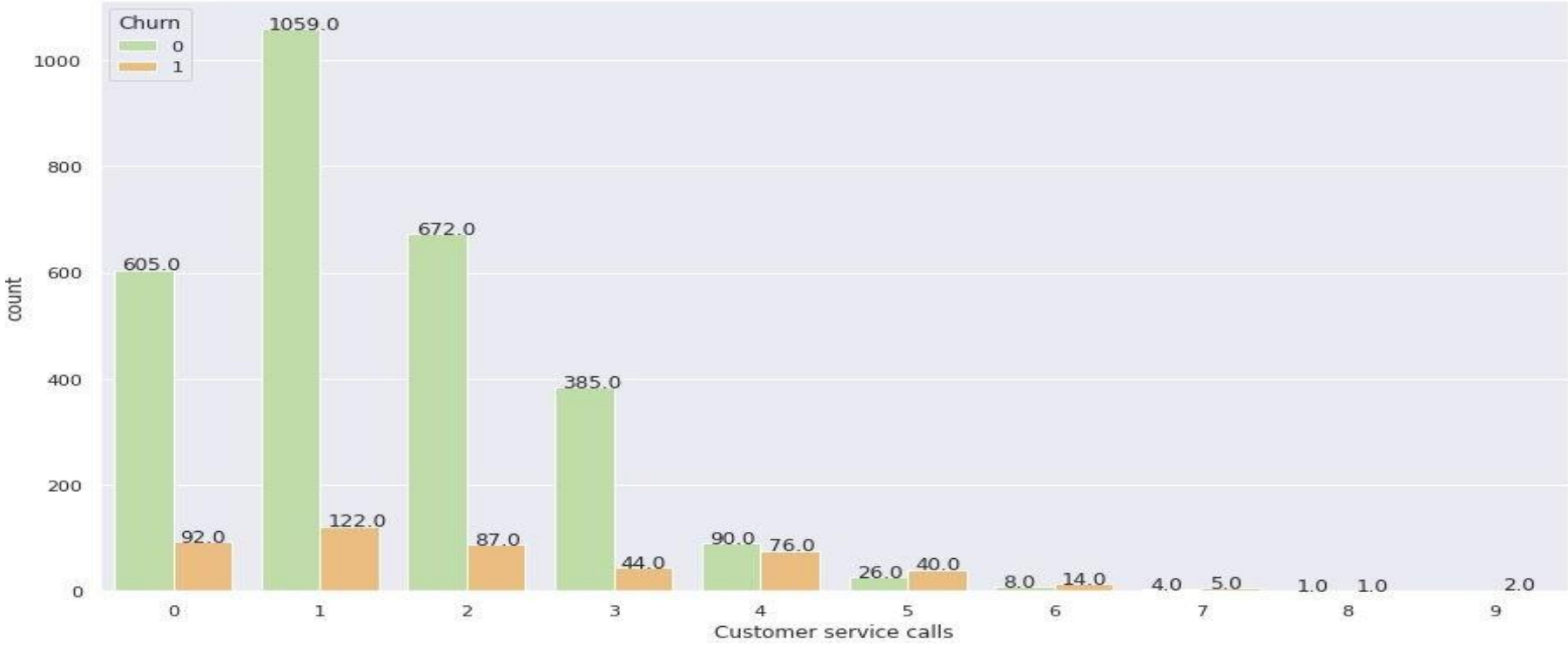


Top 10 Customers who spent the most on overall International Calls



Churn activity in different states by account length

Customer service calls by churn rate



Conclusions

- Most of the customers don’t prefer international plans and voicemail plans.
- Most of the customers spend more time on calls during day time.
- Business is booming in Area Code 415 specifically in the East Coast Region of the US.
- Customers on average are using the services for a period of around 8 years.
- Most of the customers are located in West Virginia.
- The factors that are affecting the most for customer churn are International Plan, Total day charge, Customer service calls.

To retain customers :

- The company should give incentives who spend more time on calls during day time
- The company should concentrate more on prices than features.
- The company should lower prices for international plans and its hourly call rates.
- The company should decrease the frequency of customer service calls received by customers.