Telecom Churn Analysis

Derive valuable insights using python

Mohammed Yahya Ansari

Almabetter – EDA Capstone – 1

Telecom Churn Analysis

Mr. Prafull Korde, Mr. Mohammad Yahya Ansari

Data science trainees, Alma better, Bangalore

Abstract

The telecommunications industry is responsible for the transmission of information over long distances, typically through the use of electronic signals or electromagnetic waves. This includes the use of telephone and television networks. Here in this project, we have used Orange S.A., formerly France Telecom S. A's telecommunication company data and done the Exploratory Data Analysis. This study was done to produce and observe meaningful insights and derive helpful conclusions like what are the parameters which causing the customer churn. This study will help to observe the data with various visuals aiding to make the data crisp and easy to understand.

Introduction

Orange S.A., formerly known as France Telecom S.A., is a French multinational telecommunications corporation. It is the seventh largest telecommunications company in the world, with a presence in 26 countries. It has 266 million customers worldwide and employs 89,000 people in France, and 59,000 elsewhere. In 2021, the group had revenue of €42.4 billion. The company's head office is located in the 15th arrondissement of Paris.

Orange is the brand used by the group for its mobile, broadband, and pay-tv services, as well as for its corporate and public sector services. The company was founded in 1994 and is headquartered in Paris, France. It is a publicly traded company, with shares listed on the Euronext Paris stock exchange. In addition to its telecommunications services, Orange also provides a range of other products and services, including cloud computing, cybersecurity, and internet of things (IoT) solutions. Thus, studying this large amount of data makes it important to understand trends and get other valuable Insights from a business point of view.

This is the logo of Orange S.A. telecom company.

Business Understanding & Problem statement

This initial phase of data analysis focuses on understanding the objectives of the project and requirements from a business point of view, and then converting this knowledge into a data analysis problem definition. Customer retention consists of “Identifying which customers are likely to Churn, determining which customers should retain and developing strategies to retain profitable customers”. The main thing in retention process is identifying Churn ratio which is a very meaningful and vital determination for many companies. Determination of Churn ratio indicators is also very important. By using those indicators, firms can make prediction on future behavior of new customers and can develop new strategies much before customers start to think about churn. Thus, it is vital to build a very successful and accurate Churn model during the retention studies.

Orange S.A., formerly France Telecom 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 canceled 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.

Methodology

We collected the data of play store and user reviews from website of alma better. Our basic approach was to make a copy of original clean the data and make it ready for data analysis and data visualization. Steps will be explained in details below.

Step 1:- Data Acquisition & Overview

Obtain a representative sample of the data from the telecom company, including customer demographic information, usage patterns, and churn status.

Necessary copies of data were made to save the parent data from any permanent modifications. General layout of data was studied using .info(), .describe() and .head() methods.

Variables: - The description of the dataset

Variable	Description	Data type	Unique Values	Min	Max
State	Categorical for 50 states and capital DC	object	51	NaN	NaN
Account Length	Number of days account has been active	int64	212	1	243
Area Code	Code Number of Area having some States included in each area code.	int64	3	NaN	NaN
International Plan	Activated International plan or not	object	2	NaN	NaN
Voicemail plan	Activated voicemail plan or not	object	2	NaN	NaN
Number vmail messages	Count of vmail messages sent	int64	46	0	51
Total day minutes	Total minutes used during day time	float64	1667	0	350.8
Total day calls	Total number of calls during day time	int64	119	0	165.00
Total day charge	Total charge during day time	float64	1667	0	59.64
Total eve minutes	Total minutes used during evening time	float64	1611	0	363.70
Total eve calls	Total number of calls during evening time	int64	123	0	170.00

Total eve charge	Total charge during evening time	float64	1440	0	30.91
Total night minutes	Total minutes used during night time	float64	1591	23.20	395.00
Total night calls	Total number of calls during night time	int64	120	33.0	175.00
Total night charge	Total charge during night time	float64	933	1.04	17.77
Total intl minutes	Total minutes used of international call	float64	162	0	20.00
Total intl calls	Total number of international calls	int64	21	0	20.00
Total intl charge	Total charge of international calls	float64	162	0	5.40
Customer Service Calls	Number of calls to customer service	int64	10	0	9
Churn	Customer churn	bool	2	NaN	NaN

Step 2:- Data cleaning

Function “df.info()” was made to review the dataset & check the basic information of dataset & process to determine the null values ,duplicates and data types. And we found that there are no missing values and duplicate values in the dataset.

Step 3:- Data Wrangling

In data wrangling the first thing we did is to find out how customers have churn and how many are not and their rate. We know that plotting charts will give us more visual representation. So, from this we got to know that there are ~14.5 % churn customers are there.

After that we did some manipulations on basis of area code which we tabularized it.

Area Code	Churn	Churn Rate (%)			Day Charge	Eve Charge	Night Charge	International Charge
408	**False**	21.48		**Mean**	30.12	17.11	8.96	2.73
	True	3.66		**Median**	29.98	17.21	8.98	2.73
415	**False**	42.57		**Mean**	30.87	17.06	9.09	2.79
	True	7.08		**Median**	30.72	17.06	9.14	2.79
510	**False**	21.45		**Mean**	30.39	17.11	9.01	2.74
	True	3.75		**Median**	30.51	17.16	8.91	2.75

From above table we can see that there is more customer are from area code 415 but if we check the distribution of churn, it almost same in all area code.

Churn on basis of international plan

In this we found out that those who has international plan their churn rate is higher almost 42.41 % customers are churned.

Churn on basis of International & Voice mail plan

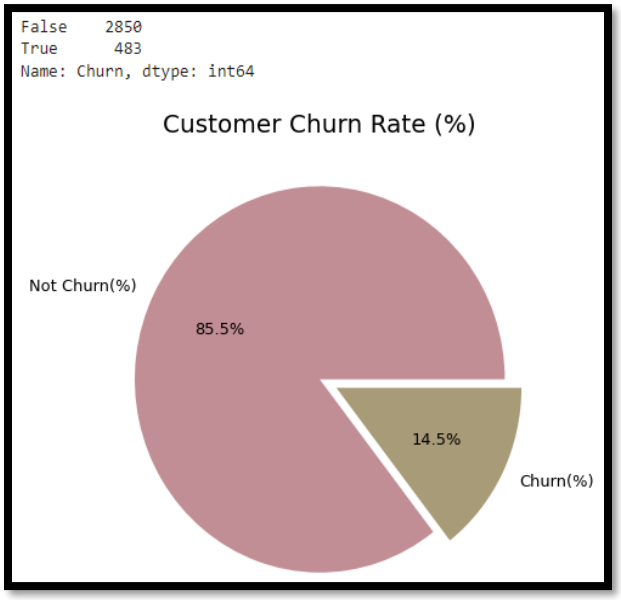
Total Customer churn are: - 483

		International Plan	
		Yes	**No**
Voice mail plan	**Yes**	36	44
	No	101	302

So, from this table we can say that those who has neither international plan nor having v-mail plan are churned high, almost around 62 %.

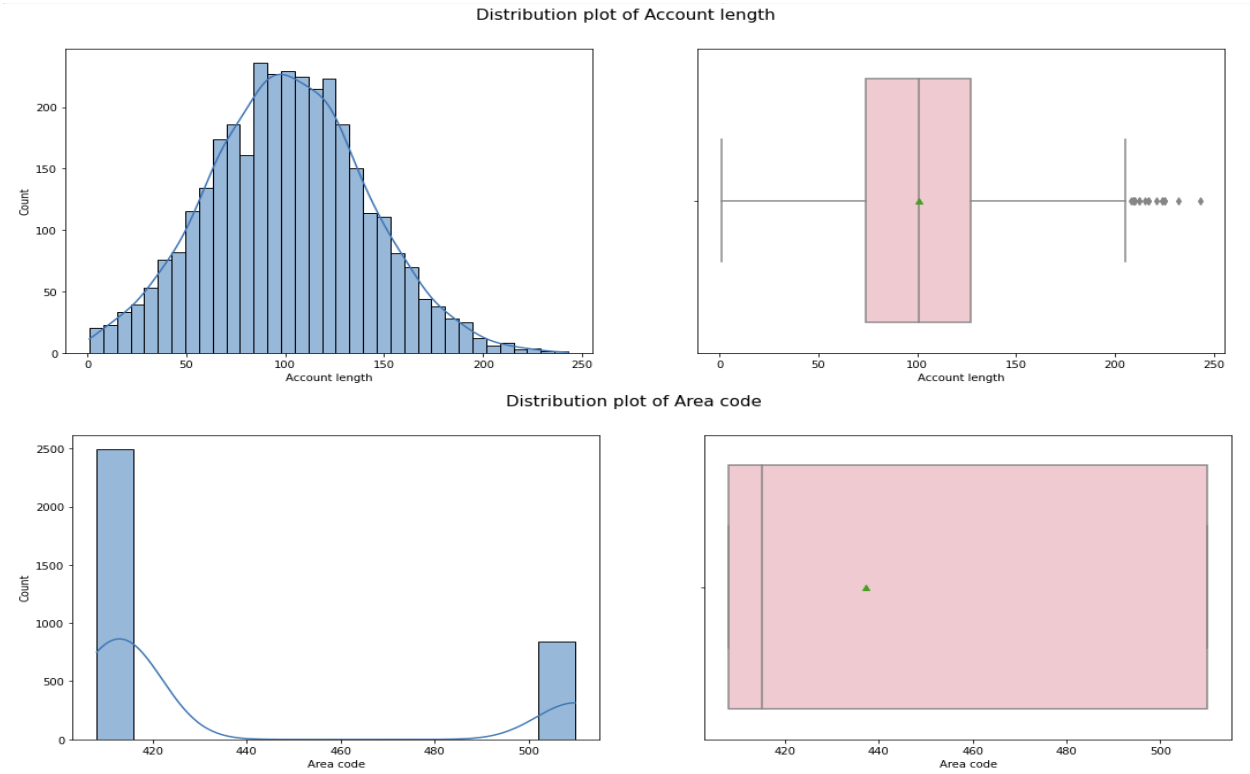
Step 4:- Data Visualisation

Visualization of dependent variable

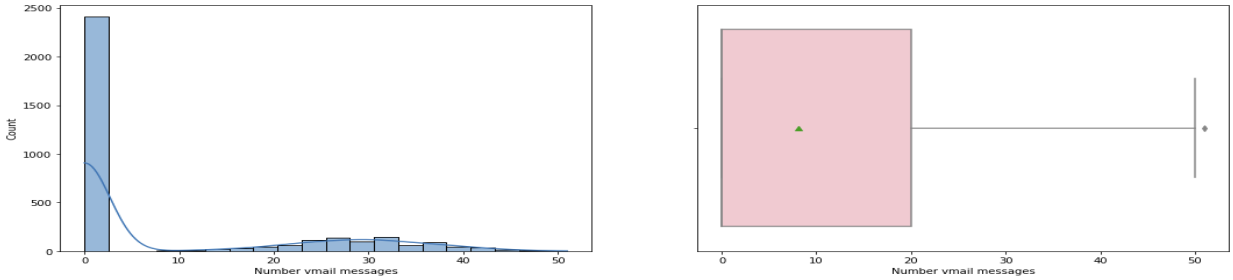


From the above chart we got to know that, there are 2850 customers which are not churned which is 85.5% of the whole customers data given in the dataset. In other hand, 483 customers are churned which is 14.5 % of the whold customers data given in the dataset.14.5% customers are churned which might look like a small number, but once upon a time the 14.5% was 1.45% which has grown upto 14.5%. So, immediate action should be taken.

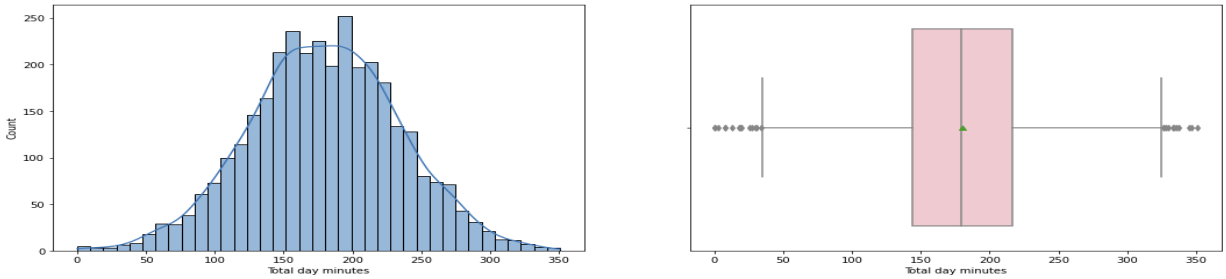
Column wise Histogram and Box Plot (Univariate)



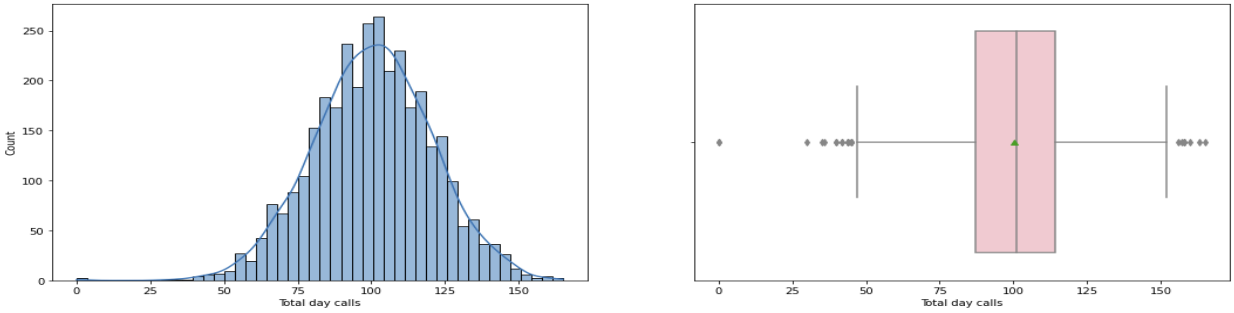
Distribution plot of Number vmail messages



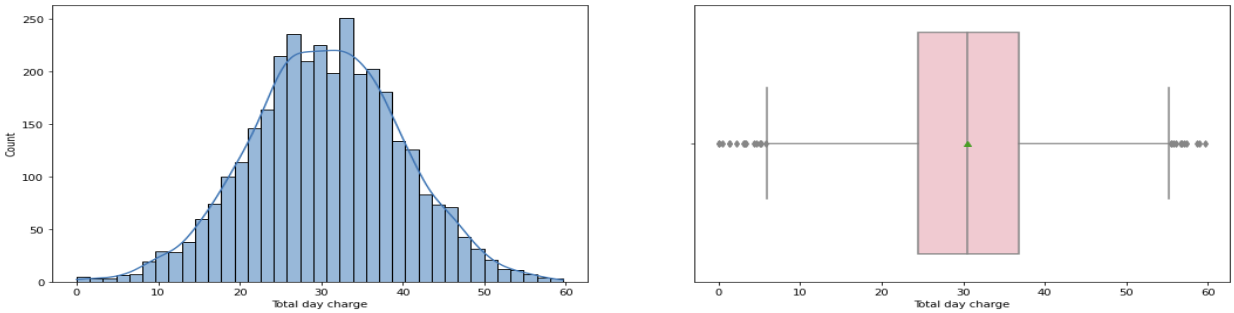
Distribution plot of Total day minutes



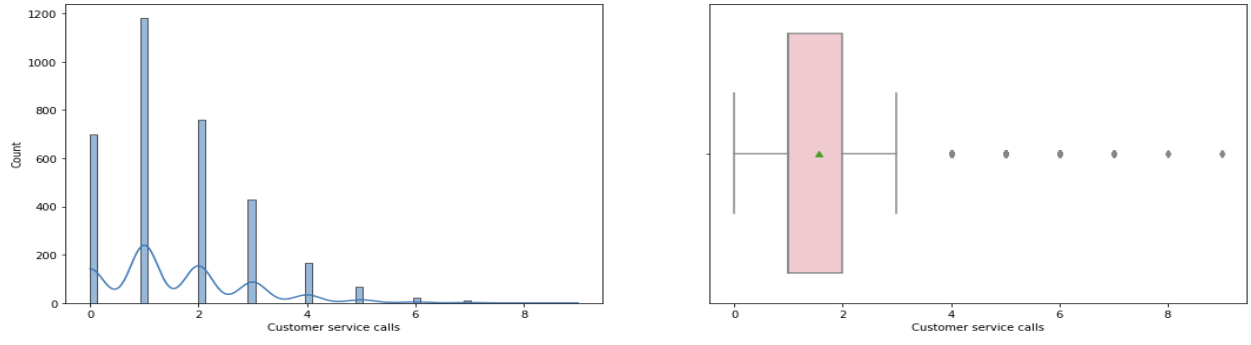
Distribution plot of Total day calls



Distribution plot of Total day charge



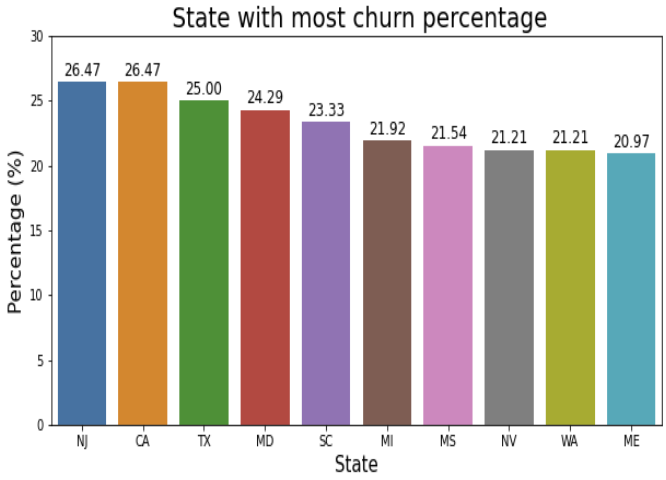
Distribution plot of Customer service calls

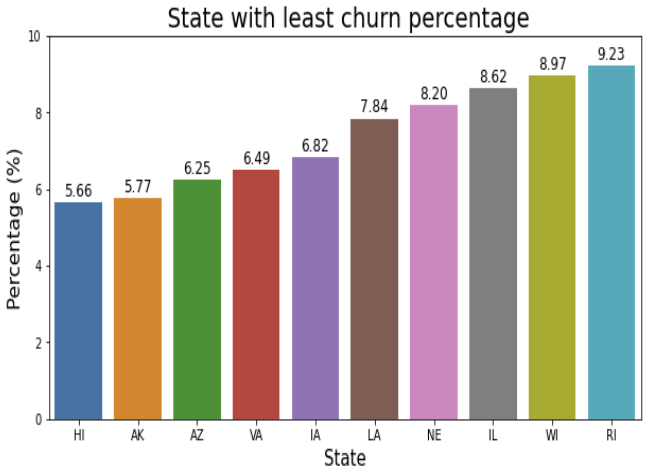


Histplot and boxplot are used to display the distribution of the data. It can help detect unusual observations and also it shows the data is symmetric or not.

From above distribution charts we can see that, all columns are symmetric distributed and mean is nearly same with median for numerical columns. Here Area code will be treated as text values as there are only 3 values in the particular column.

State vs Average true churn percentage





There are 51 states having different churn rates.

CA, NJ ,TX , MD ,SC ,MI, MS, NV, WA, ME are the ones who have higher churn rate more then 20% which is more than 50% of average churn rate.

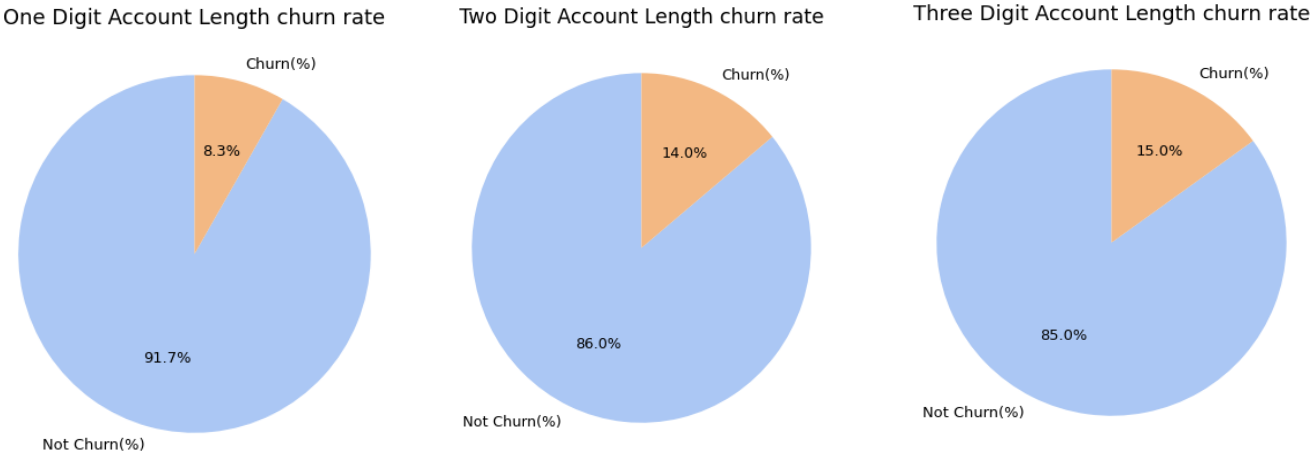
And HI, AK, AZ, VA, IA, LA, NE, IL, WI, RI,are the ones who have lower churn rate which is less than 10%.

Account length with churn

```
Churn
False    22
True      2
dtype: int64
```

```
Churn
False    1378
True      225
dtype: int64
```

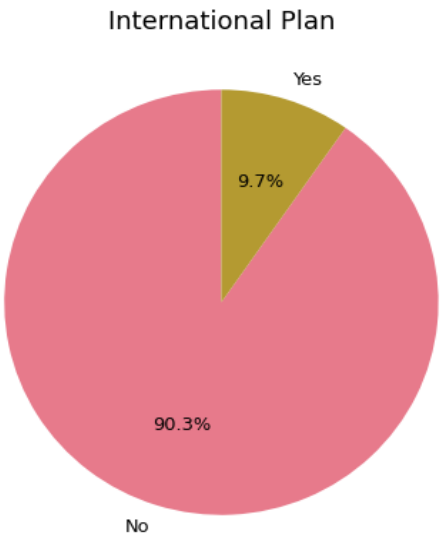
```
Churn
False    1450
True      256
dtype: int64
```

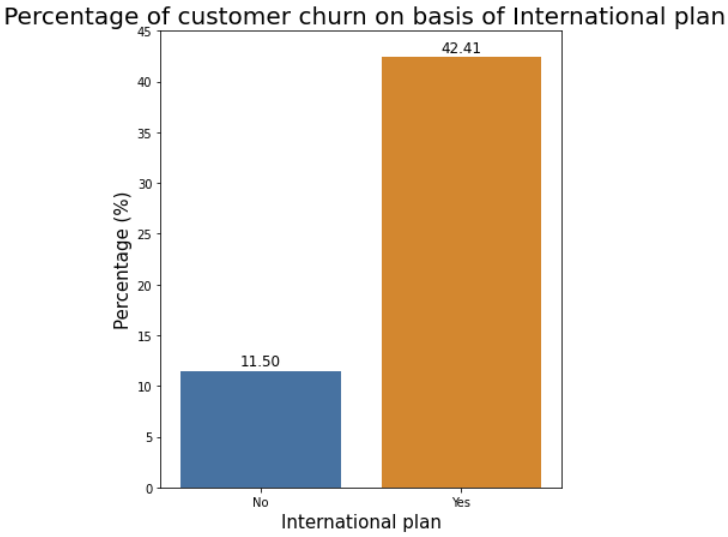


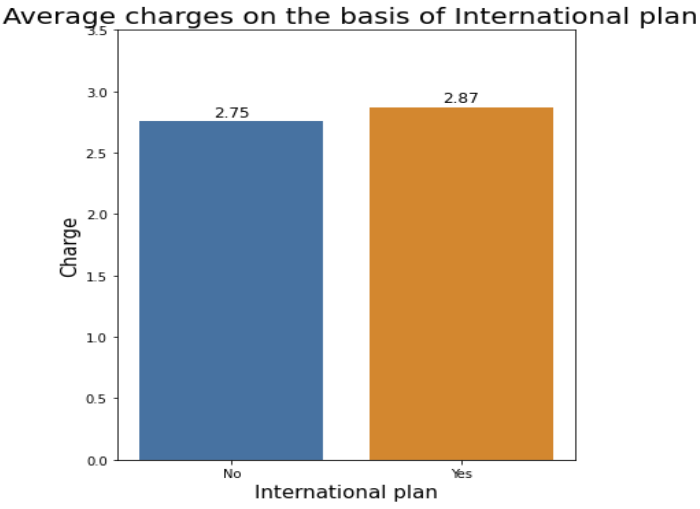
As from the above record, we can see that two-digit account length customers are churning with a number of 225.

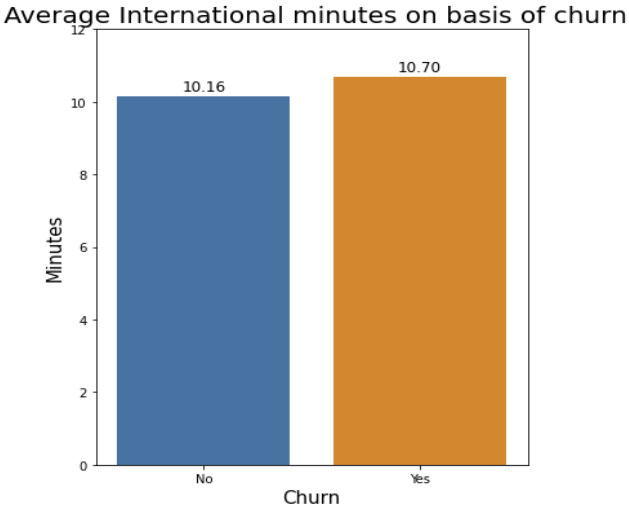
And three-digit account length customers are churning with a number of 256, so their churning rate is slightly higher.

International Plan









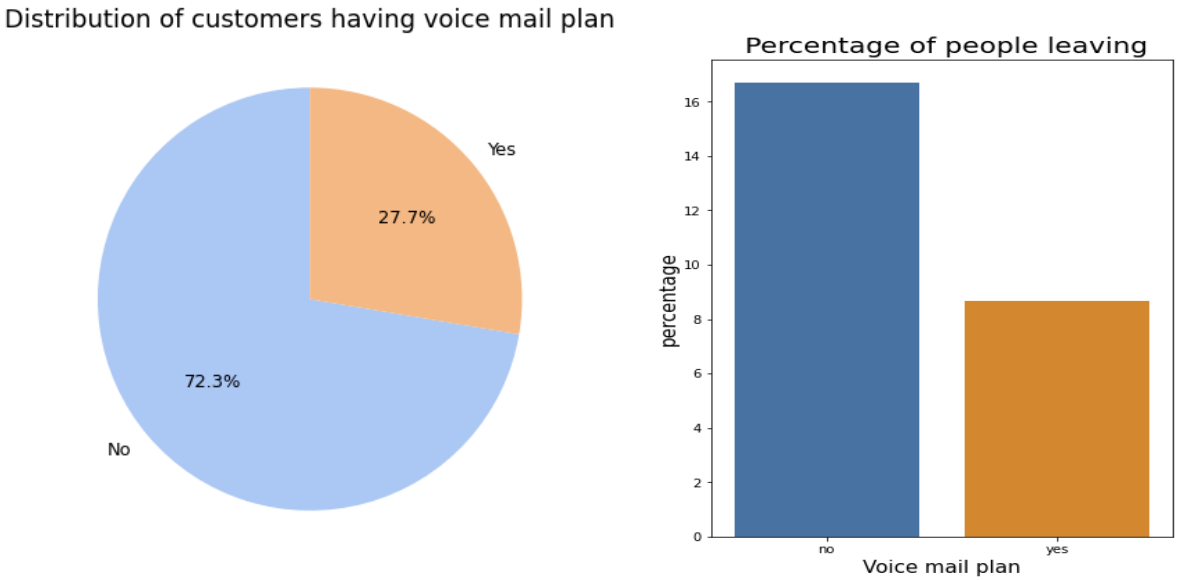
In this analysis, 3010 don't' have an international plan, 323 have an international plan Among those who have an international plan 42.4 % people churn. Whereas among those who don't have an international plan only 11.4 % people churn.

Among those who have an international plan their average charge is 2.87 and they talk for 10.7 minutes average. Whereas among those who don't have an international plan their average charge is 2.75 and they talk for 10.16 minutes average.

The reason why people having international plan might be leaving is that they are paying same amount of money for international calls as for those customers who don't have an international

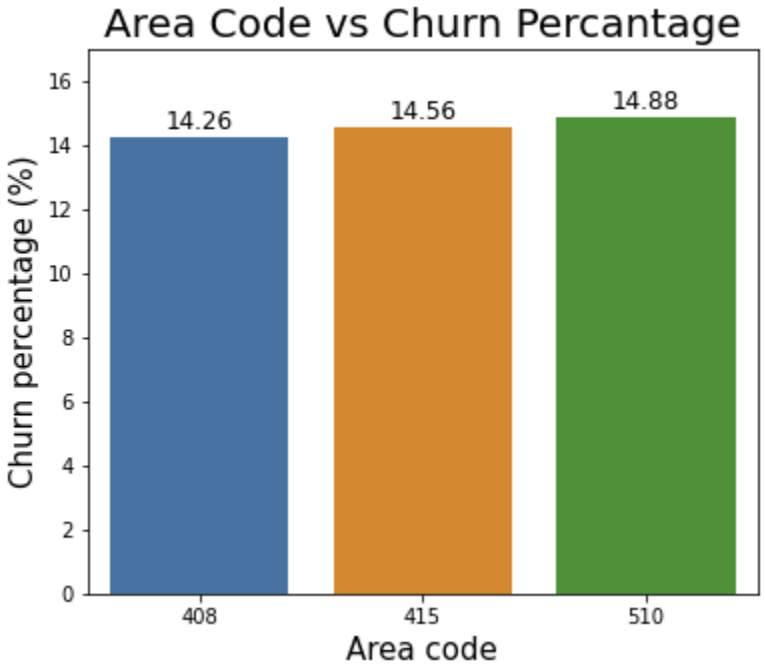
plan. Hence, they aren't getting any benefits for having an international plan so they might be unhappy.

Voice Mail plan



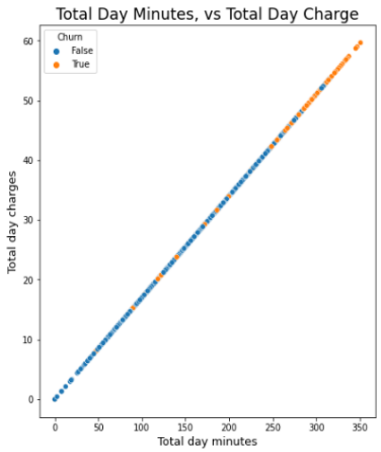
There are 2411 customers who don't have voice mail plan and 922 have voice mail plan, among those who don't have voice mail 16.7 % people churn. Whereas among those who have a voice mail plan 8.7 % people churn.

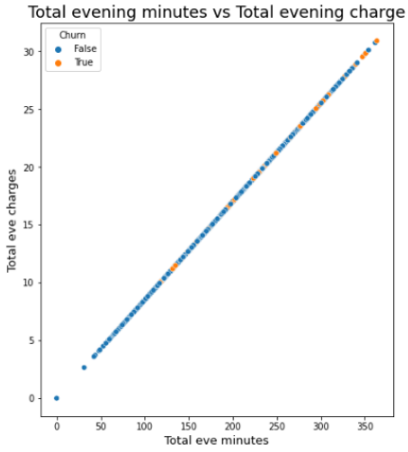
Area code

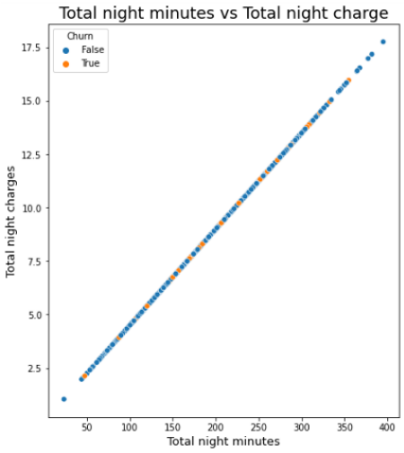


All Area Code have around 14% Churn rate. So, Area Code doesn't matter.

Overall Calls

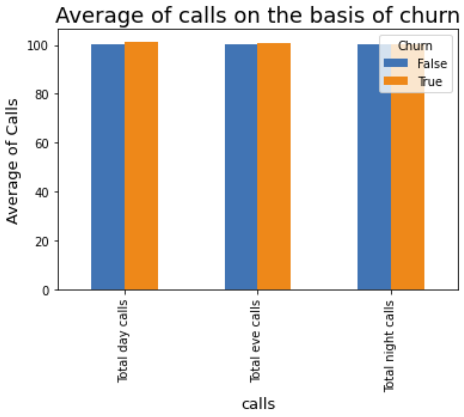


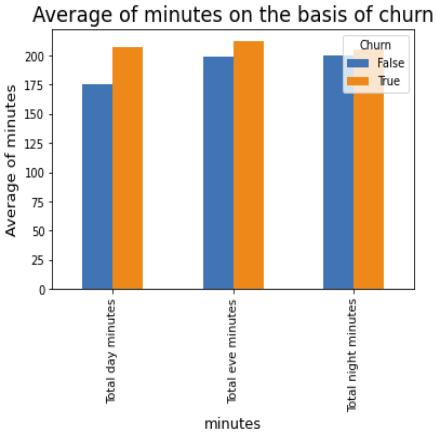


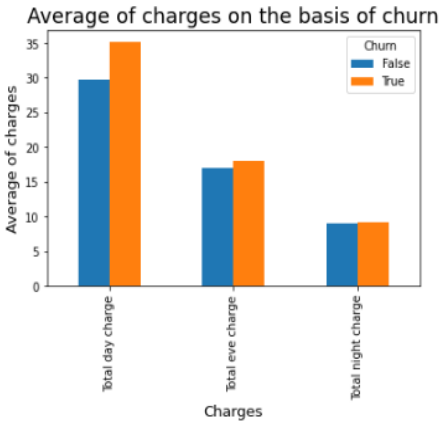


Churn customers speak more minutes that non-churn customers at day, evening and night. Hence, they pay more charge that non-churn customers. We can retain churn customers if we include master plan.

In Master plan if a customer is talking more minutes, then we can charge a little less amount from him or he can get discount or additional few minutes to talk. This will make customers who are going to churn happy and they will not leave the company.



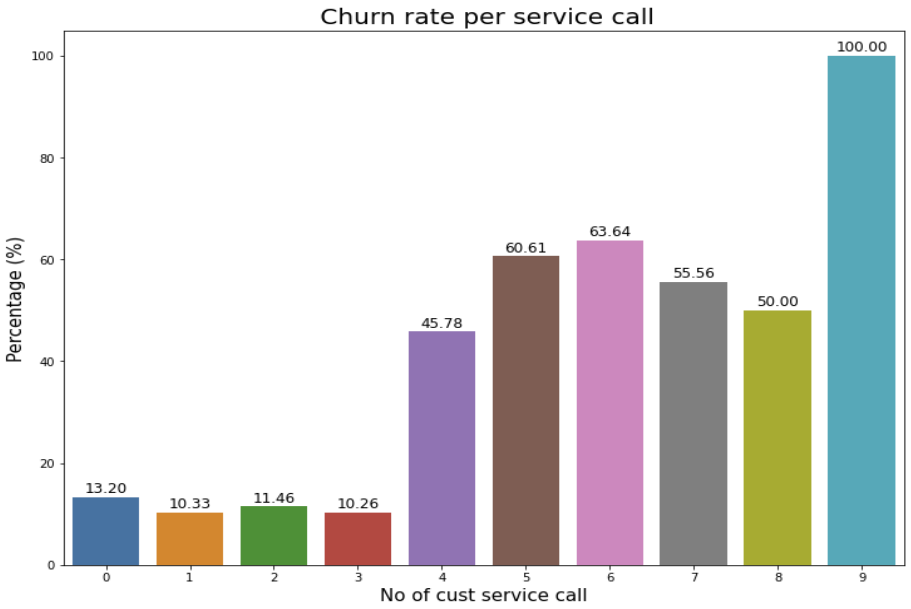




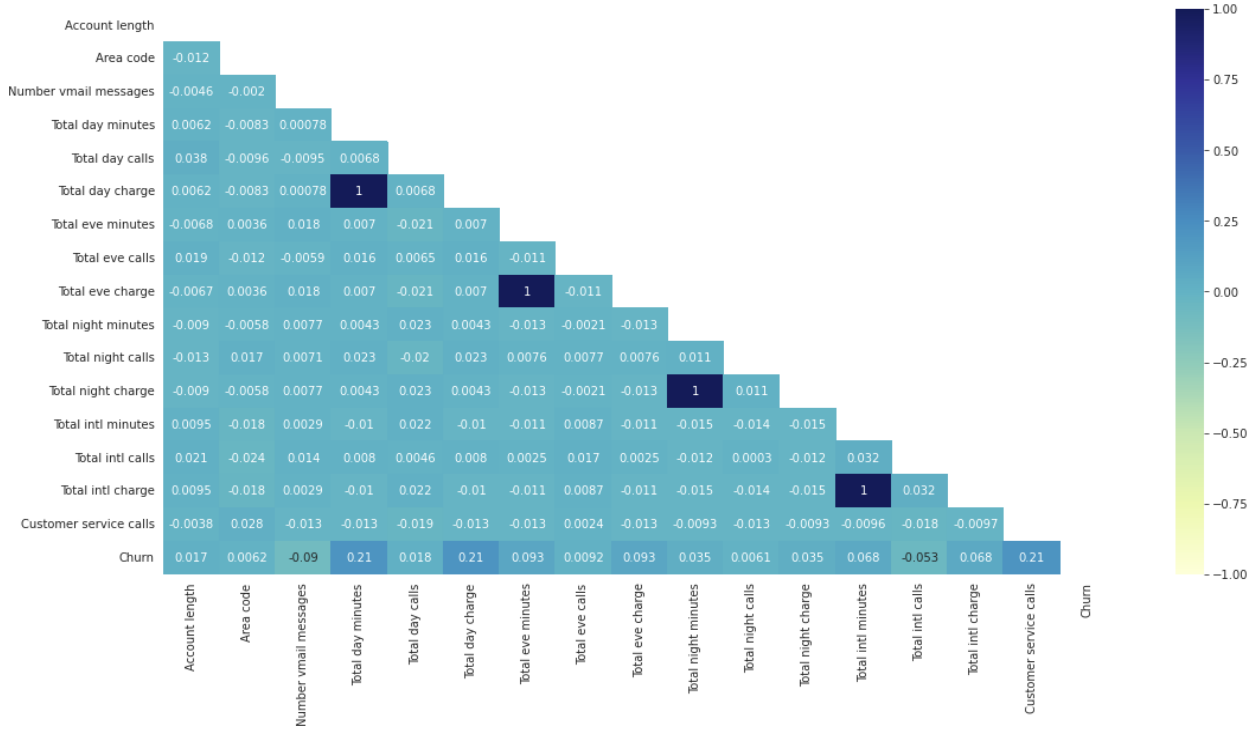
Customer Service Calls

The service calls of customers varies from 0 to 9. Those customers who make more service calls they have a high probability of leaving.

As we can see from graph, customers with more than 5, their churning rate is more. Hence customers who make more than 5 service calls, their queries should be solved immediately and they should be given better service so that they don't leave the company.



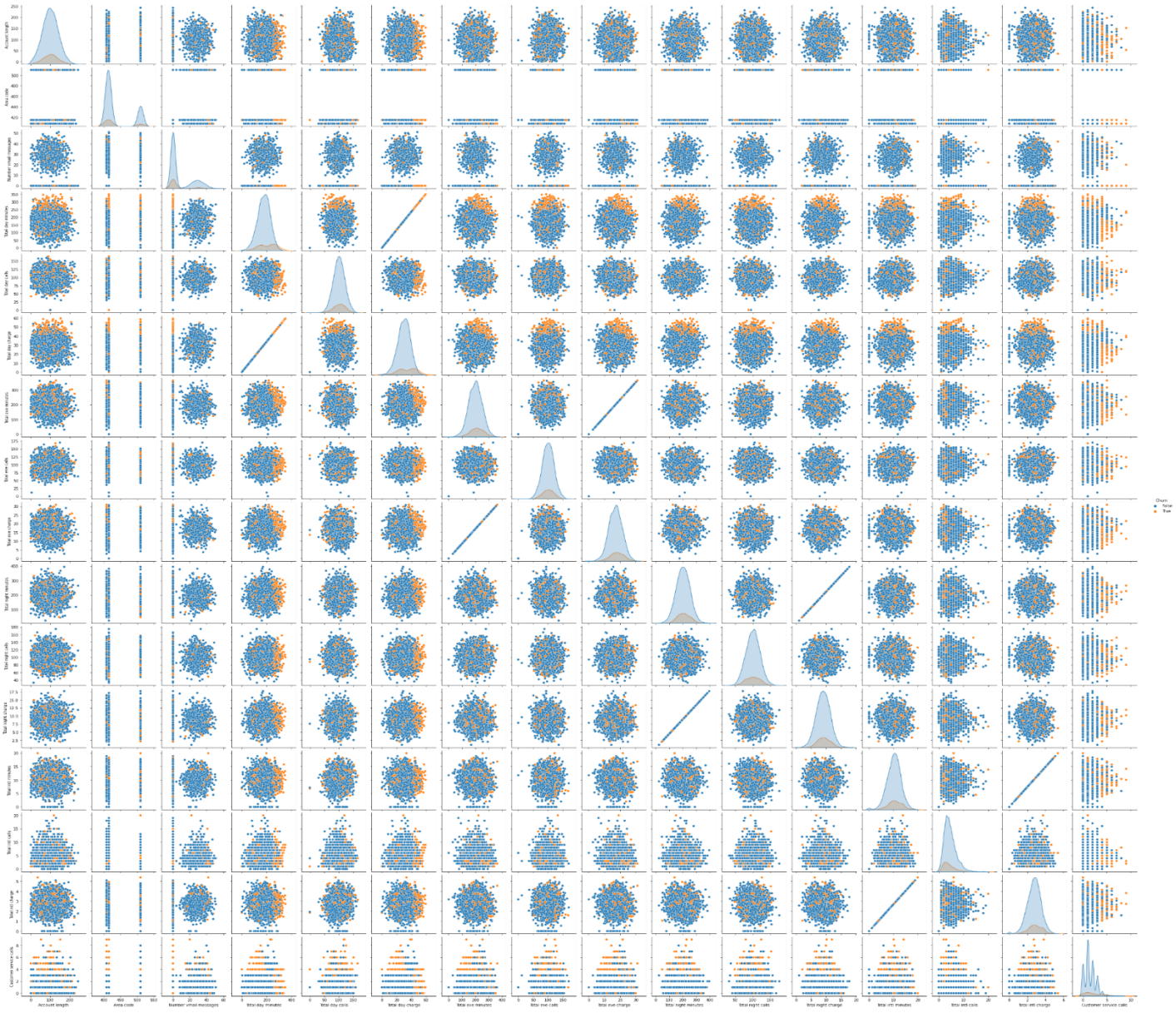
Correlation Heatmap



From the above correlation heatmap, we can see total day charge & total day minute, total evening charge & total evening minute, total night charge & total night minute are positively highly correlated with a value of 1.

Customer service call is positively correlated only with area code and negatively correlated with rest variables.

Pair plot



From the above chart we got know, that there is less linear relationship between variables and data points aren't linearly separable. Churned customers data is clustered and overlapped each other. Non churn data are quite symmetrical in nature and churn customer data are quite non-symmetric in nature. In this whole pair plot the importance of area code can be seen and no. of churn with respect to different features are really insightful.

Solution to Business Problem

- Modify International Plan as the charge is same as normal one.
- Be proactive with communication.
- Ask for feedback often.
- Periodically throw Offers to retain customers.
- Look at the customers facing problem in the most churning states.
- Lean into best customers.
- Regular Server Maintenance.
- Solving Poor Network Connectivity Issue.
- Define a roadmap for new customers.
- Analyze churn when it happens.
- Stay competitive.

Conclusion

- The four charge fields are directly related to the minute fields.
- The area code and state fields may not be relevant and can be excluded.
- Customers with the International Plan tend to churn more often.
- Customers who have had four or more customer service calls churn significantly more than other customers.
- Customers with high day and evening minute usage tend to churn at a higher rate.
- There is no clear relationship between churn and the variables such as day calls, evening calls, night calls, international calls, night minutes, international minutes, account length, or voice mail messages.

References

1. https://www.orange.com/en/finance/investors/consolidated-results
2. https://towardsdatascience.com/customer-churn-in-telecom-segment-5e49356f39e5
3. https://www.comviva.com/blog/references/efficient-ways-for-customer-churn-analysis-in-telecom-sector/