Capstone Project-1

TELECOM CHURN ANALYSIS(EDA)

Team Members

Suhas Tantri

Raghavendra A Kulkarni

AGENDA

- Introduction
- Defining Problem Statement
- Exploratory Data Analysis
- Conclusions
- Measures

Introduction

- Churn is a problem for telecom companies because it is expensive to acquire a new customer and companies want retain the existing customers. Churn rate has strong impact on future revenue of the company.

- Consumers today go through a complex decision making process before subscribing to any one of the numerous Telecom service options

- Customer loyalty becomes an issue .Hence ,it is becoming increasingly important for telecommunications companies to proactively identify factors that have tendency to unsubscribe and take preventive measures to retain customers

PROBLEM STATEMENT

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

Exploratory Data Analysis

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

DATA SUMMARY

```
<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
```

Variable Description

1. State:- Name of the state in which customer resides.
2. Account_Length:- It's a integer variable indicating how long the account has been active.
3. International_plan:-Its a boolean variable where it denotes whether the customer has subscribed the international plan
4. Voice_mail_plan:-Its a boolean variable where it denotes whether the customer has subscribed the voice mail plan
5. Number_vmail_messages:- It's a integer variable which indicates the number of voice mail messages of the customer
6. Customer_service_calls:- It's a integer variable which denotes the number of calls made by customer to the customer care service center.
7. Churn:- It's a boolean variable which represents whether the customer has churned the service or not.

EDA PROCESS

```
# lets look at first five values
df.head()
```

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls	Churn
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70	1	False
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70	1	False
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29	0	False
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	2	False
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	3	False

EDA PROCESS

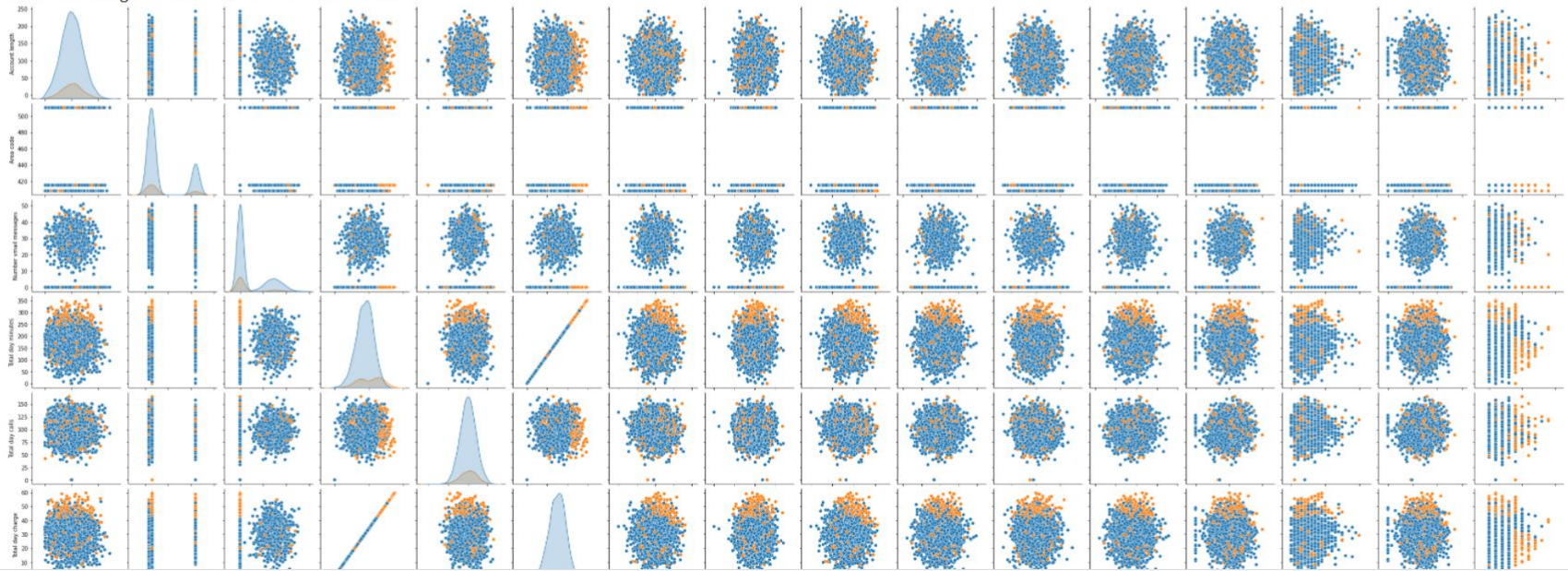
```
#Finding the corelation between independent variables from the dataset
df.corr()
```

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	Customer service calls	Churn
Account length	1.000000	-0.012463	-0.004628	0.006216	0.038470	0.006214	-0.006757	0.019260	-0.006745	-0.008955	-0.013176	-0.008960	0.009514	0.020661	0.009546	-0.003796	0.016541
Area code	-0.012463	1.000000	-0.001994	-0.008264	-0.009646	-0.008264	0.003580	-0.011886	0.003607	-0.005825	0.016522	-0.005845	-0.018288	-0.024179	-0.018395	0.027572	0.006174
Number vmail messages	-0.004628	-0.001994	1.000000	0.000778	-0.009548	0.000776	0.017562	-0.005864	0.017578	0.007681	0.007123	0.007663	0.002856	0.013957	0.002884	-0.013263	-0.089728
Total day minutes	0.006216	-0.008264	0.000778	1.000000	0.006750	1.000000	0.007043	0.015769	0.007029	0.004323	0.022972	0.004300	-0.010155	0.008033	-0.010092	-0.013423	0.205151
Total day calls	0.038470	-0.009646	-0.009548	0.006750	1.000000	0.006753	-0.021451	0.006462	-0.021449	0.022938	-0.019557	0.022927	0.021565	0.004574	0.021666	-0.018942	0.018459
Total day charge	0.006214	-0.008264	0.000776	1.000000	0.006753	1.000000	0.007050	0.015769	0.007036	0.004324	0.022972	0.004301	-0.010157	0.008032	-0.010094	-0.013427	0.205151
Total eve minutes	-0.006757	0.003580	0.017562	0.007043	-0.021451	0.007050	1.000000	-0.011430	1.000000	-0.012584	0.007586	-0.012593	-0.011035	0.002541	-0.011067	-0.012985	0.092796
Total eve calls	0.019260	-0.011886	-0.005864	0.015769	0.006462	0.015769	-0.011430	1.000000	-0.011423	-0.002093	0.007710	-0.002056	0.008703	0.017434	0.008674	0.002423	0.009233
Total eve charge	-0.006745	0.003607	0.017578	0.007029	-0.021449	0.007036	1.000000	-0.011423	1.000000	-0.012592	0.007596	-0.012601	-0.011043	0.002541	-0.011074	-0.012987	0.092786

EDA PROCESS

```
#Ananlyzing dataset using pairplots
sns.pairplot(df,hue='Churn')
```

```
<seaborn.axisgrid.PairGrid at 0x7fde08449650>
```

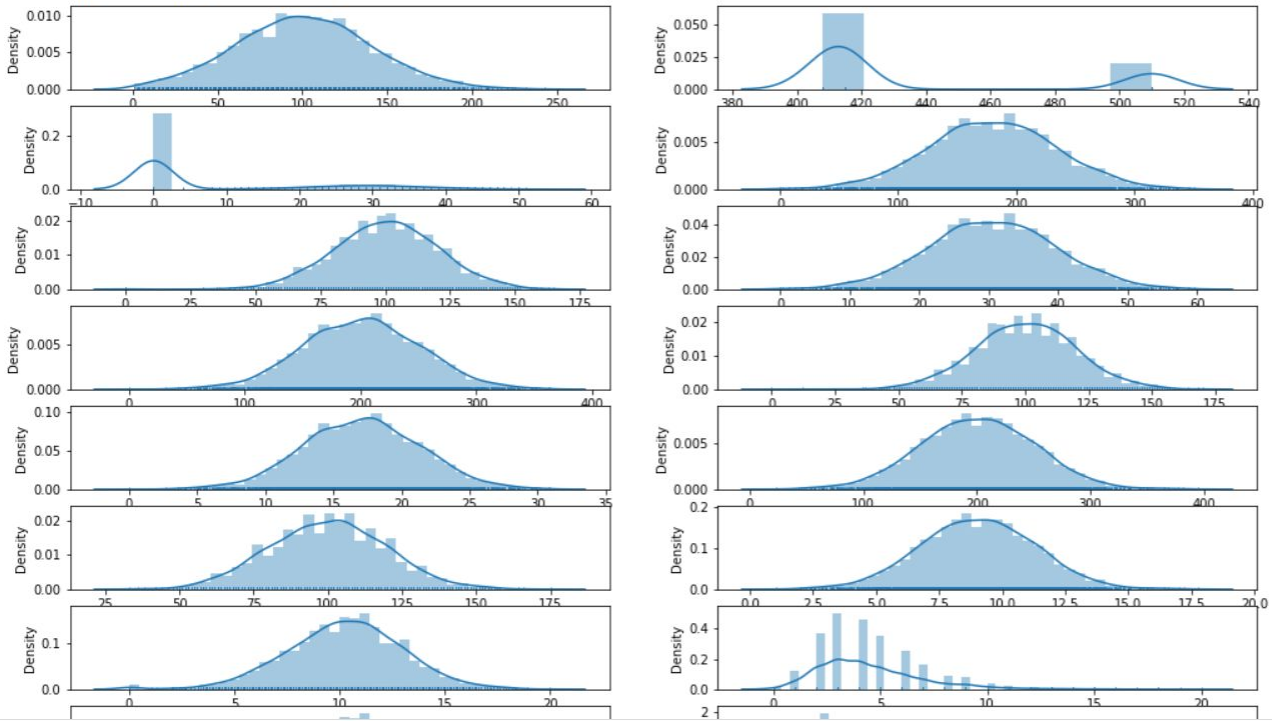


EDA PROCESS

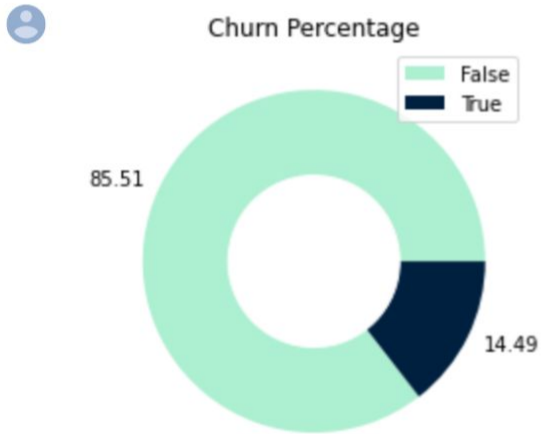
File Edit View Insert Runtime Tools Help Changes will not be saved

+ Code + Text Copy to Drive Connect Editing

```
warnings.warn(msg, FutureWarning)
```



EDA PROCESS

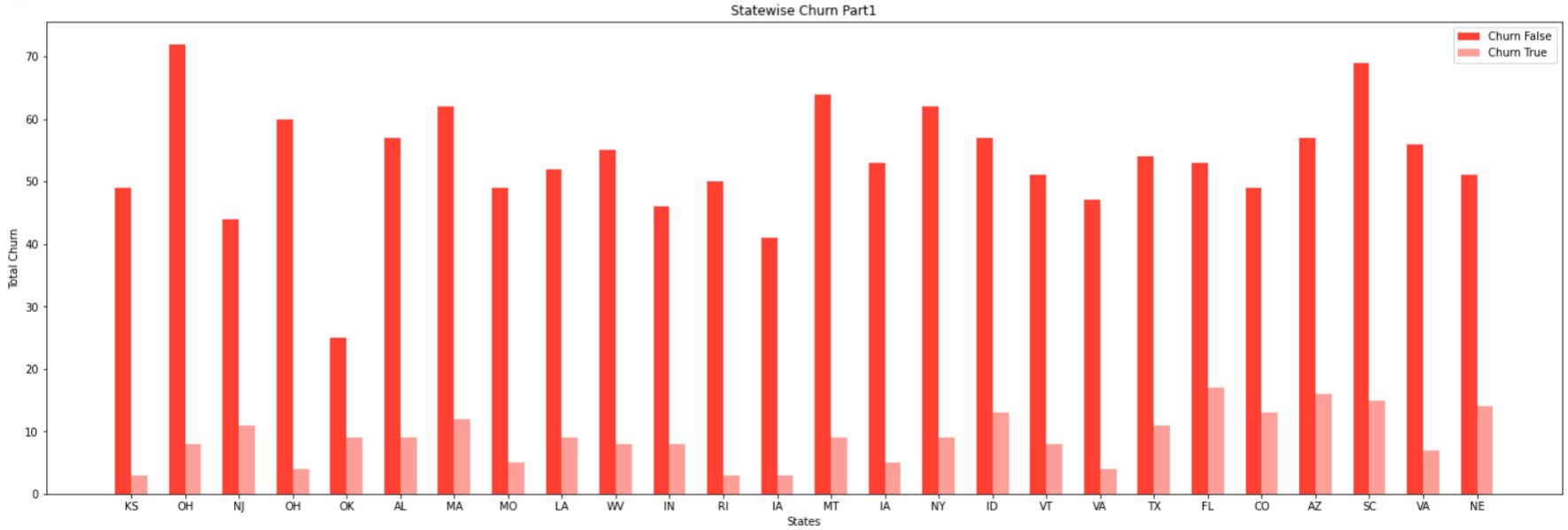


- ***From the above donut chart we can infer that (483) 14.49% of customers have left the company.***

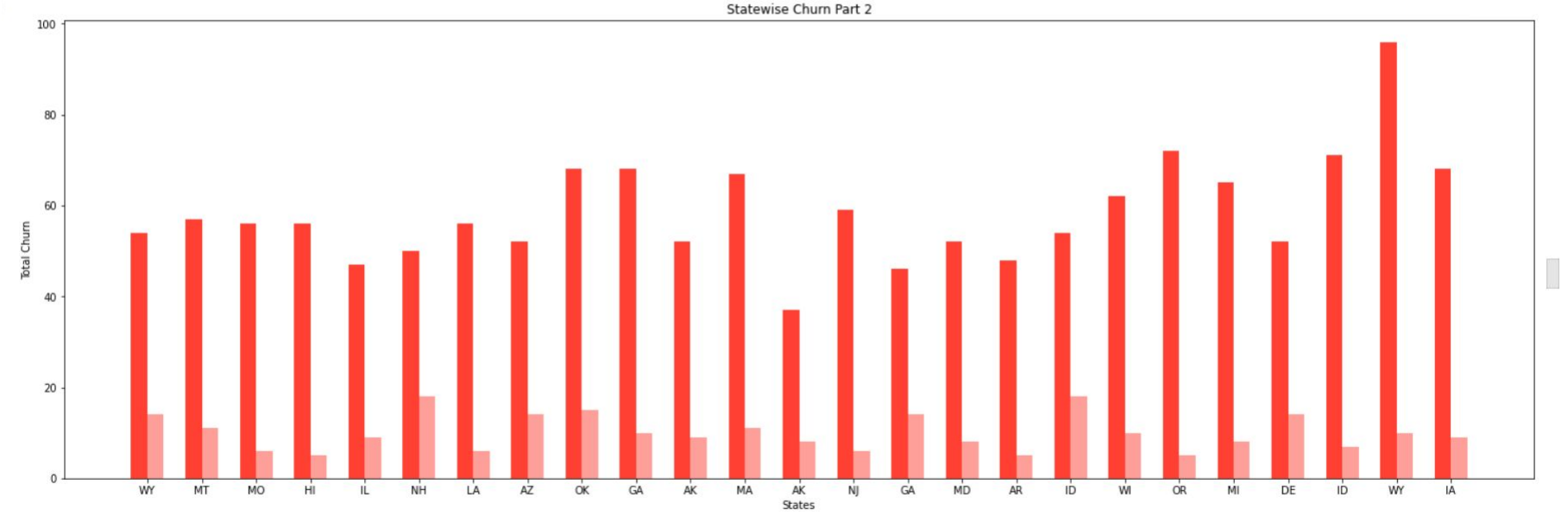
EDA PROCESS

]

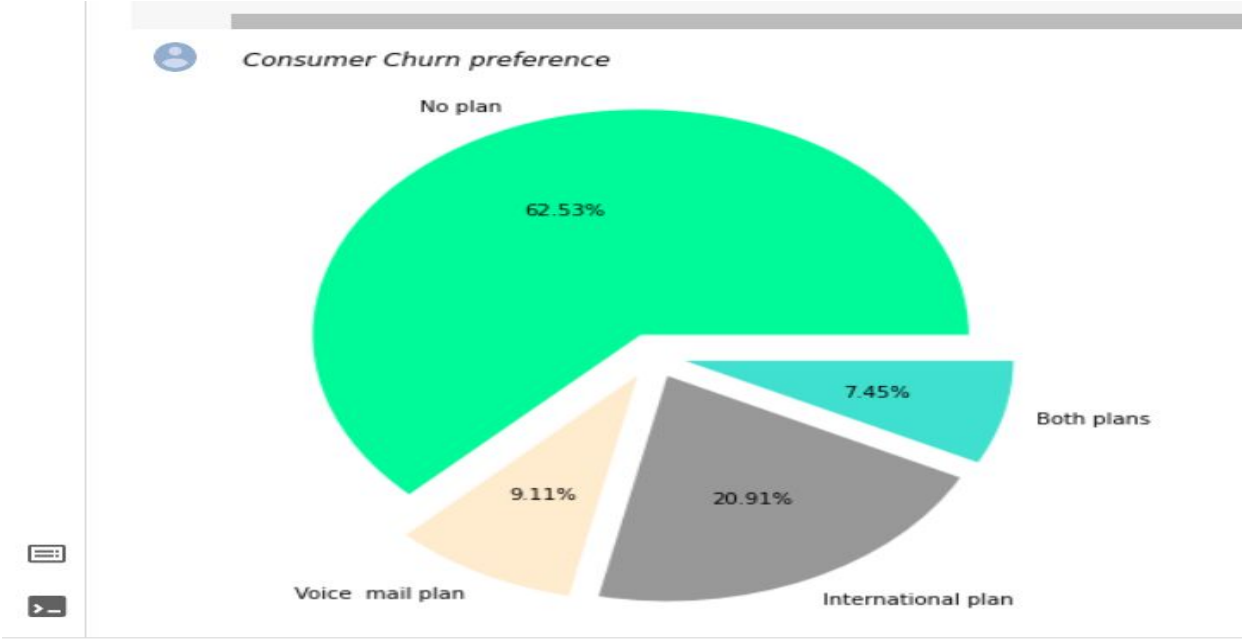
[ ]


Statewise Churn Part1
Churn False
Churn True
Total Churn
0
10
20
30
40
50
60
70
KS
OH
NJ
OH
OK
AL
MA
MO
LA
WV
IN
RI
IA
MT
IA
NY
ID
VT
VA
TX
FL
CO
AZ
SC
VA
NE
States

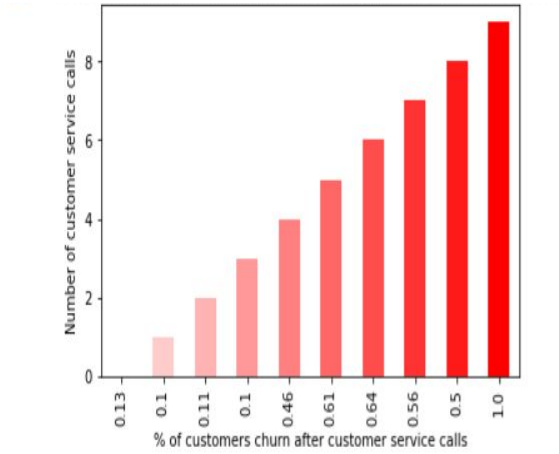
EDA PROCESS


Statewise Churn Part 2
Total Churn
States
0
20
40
60
80
100
WY
MT
MO
HI
IL
NH
LA
AZ
OK
GA
AK
MA
AK
NJ
GA
MD
AR
ID
WI
OR
MI
DE
ID
WY
IA

EDA PROCESS

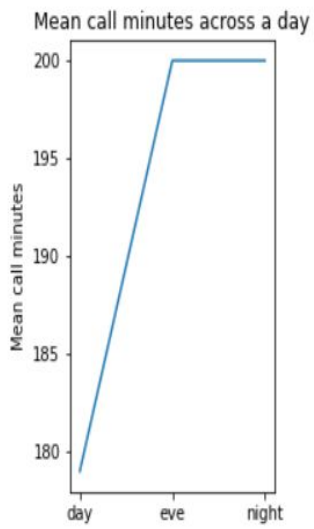

Consumer Churn preference
No plan
62.53%
7.45%
Both plans
9.11%
20.91%
Voice mail plan
International plan

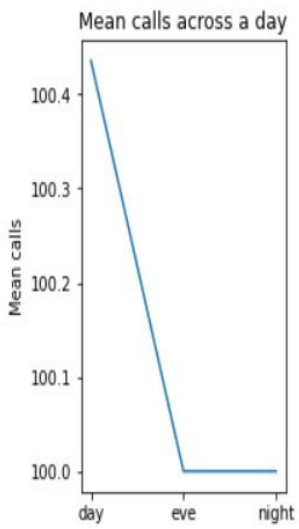
EDA PROCESS

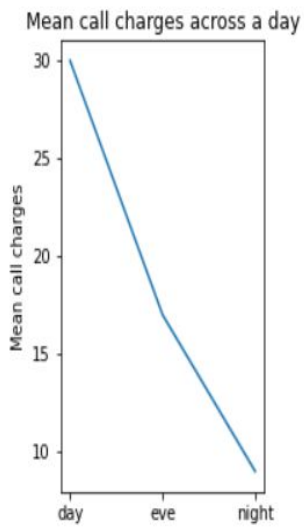


From this plot it can be observed that 62.6 % of customers churn after 4 customer service calls.

EDA PROCESS

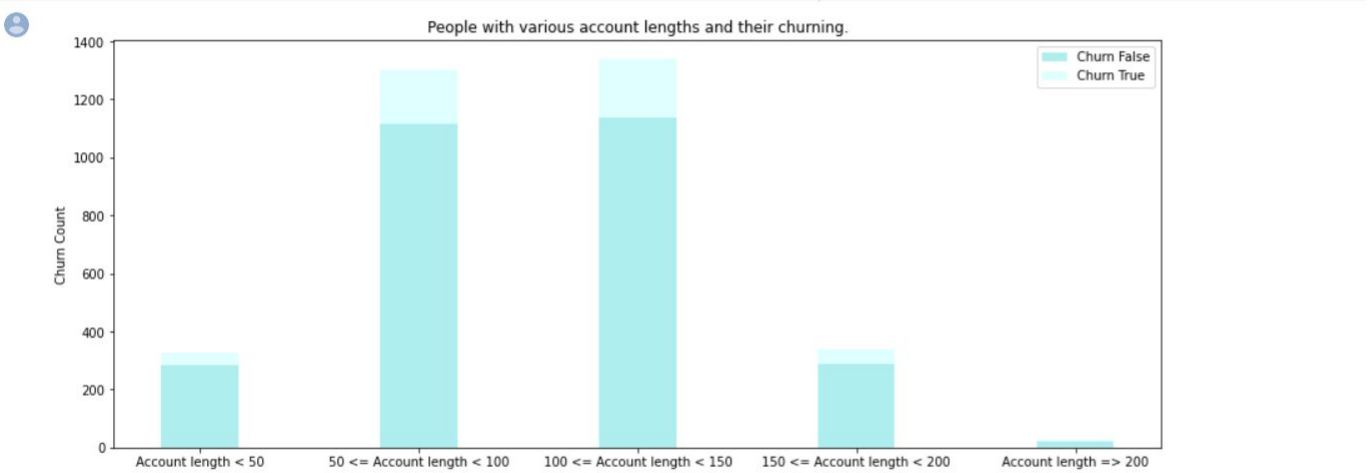






From these plots we can infer that it costs the most to make day calls

EDA PROCESS



From the graph we can infer that customers above account length of 150 are less in number compared to those below account length of 150. so, the Telecom Org has to provide favourble offers so that the customers wont churn

CONCLUSIONS

1) From the above donut chart we can infer that (483) 14.49% of customers have left the company.

2) OH,MT,SC are the states which have good retention rate

3) Fl, AZ & NH are the states which have bad retention rate

4) We can infer that people from state 'WY' are enjoying the telecom service most among all other states

5) Customers with no plan churn the most. International plan subscribers tend to churn more compared to the voice mail plan customers as the cost for international calls are more compared to voice mail.

6) Customers with more than 4 customer service calls are 62.6 % likely to churn the company.

7) The mean of the total charge is 12.3 % more for customers who churned compared to the customers who did not churn .

8) Customers above account length of 150 are less in number compared to those below account length of 150.

MEASURES

1) The company can look at reducing customer service calls , as the customers are likely to churn the company as the number of calls increase.

2) The company can also look at reducing Day call rates as there is a massive difference between day call rates and night call rates.

3) The company can also work towards reducing international call rates so that international call plan holders will not churn the company.

4) The company can also look at giving offers to their subscription plans so that non subscribers of any plan can be nudged towards purchasing a subscription , as 62.53 % of the total churning is done by non subscribers who dont have a subscrption plan.

5)The Telecom Organization has to provide favourble offers to the customers whose account length is above 150 as loyalty bonus so that the customers wont churn.

THANK YOU