Capstone Project

Telecom Churn EDA

Team Members:

Mohammed Yahya Ansari

Prafull Korde

Business Overview

- Orange S.A., formerly France Telecom S.A.
- Presence in 26 countries.
- It has 266 million customers worldwide.
- Revenue : € 42.5 billion (FY 2021)
- Business : Mobile, Broadband, Pay-TV services, Cloud computing, Cybersecurity, & IoT
- This dataset consists of cleaned customer activity data (features), along with a churn label specifying whether a customer cancelled the subscription.

Problem Statement & Business Objective

Telecom churn analysis is the process of identifying customers who are likely to cancel their service or switch to a different service provider. This is an important problem for telecom companies, as churn can have a significant impact on their revenue and profitability.

Business Objective

- Identifying the key cause of the customer churn
- Provide steps to retain the valuable customer

Methodology

Dataset First look

```
[ ] # Dataset First Look
    data.head()
```

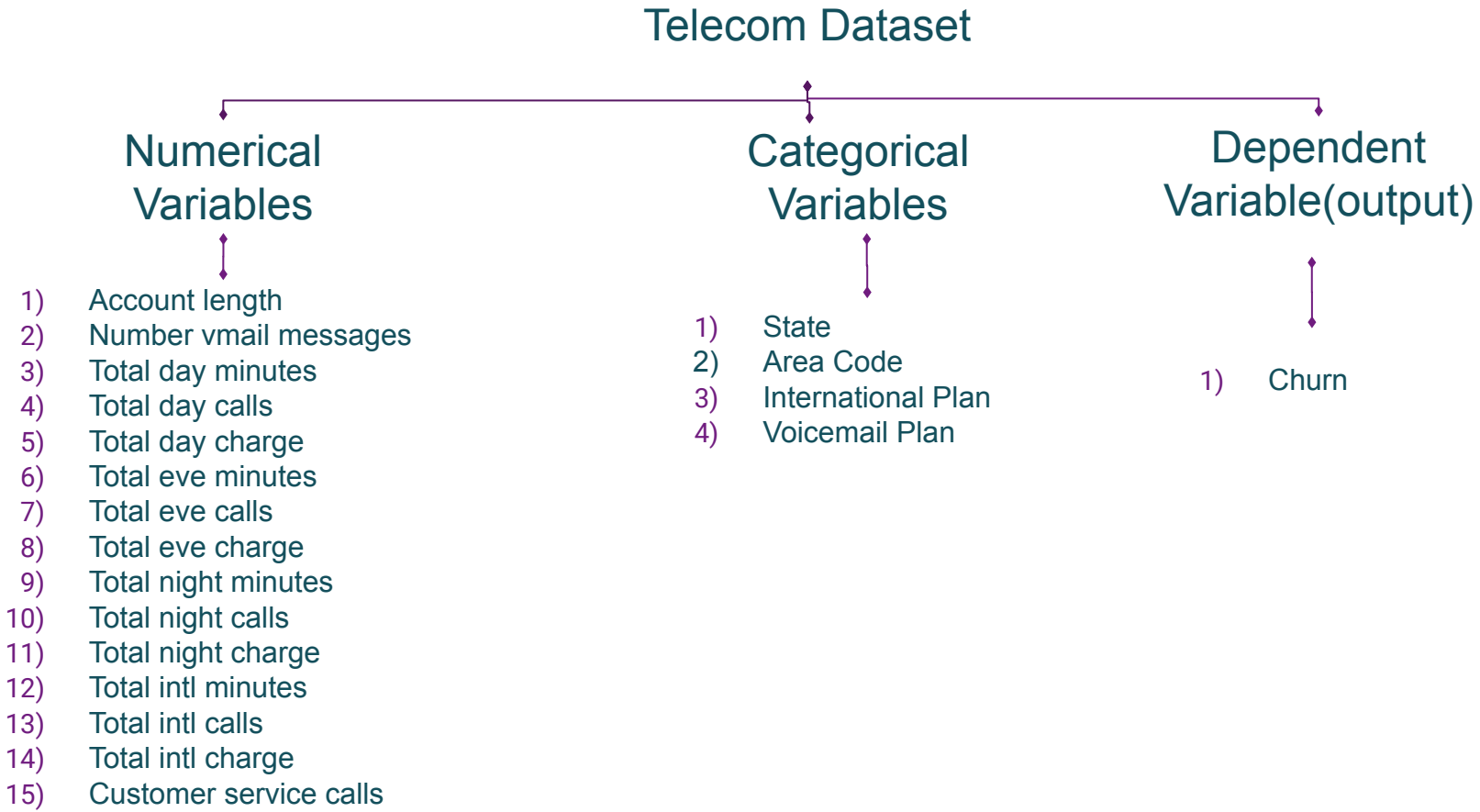
	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge	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

```
[ ] data.tail()
```

	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
3328	AZ	192	415	No	Yes	36	156.2	77	26.55	215.5	126	18.32	279.1	83	12.56	9.9	6	2.67	2	False
3329	WV	68	415	No	No	0	231.1	57	39.29	153.4	55	13.04	191.3	123	8.61	9.6	4	2.59	3	False
3330	RI	28	510	No	No	0	180.8	109	30.74	288.8	58	24.55	191.9	91	8.64	14.1	6	3.81	2	False
3331	CT	184	510	Yes	No	0	213.8	105	36.35	159.6	84	13.57	139.2	137	6.26	5.0	10	1.35	2	False
3332	TN	74	415	No	Yes	25	234.4	113	39.85	265.9	82	22.60	241.4	77	10.86	13.7	4	3.70	0	False

3333 Rows and 20 Columns

Data Summary



Variables

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 (408,415,510)	NaN	NaN
International Plan	Activated International plan or not	object	2 (Yes, No)	NaN	NaN
Voicemail plan	Activated voicemail plan or not	object	2 (Yes, No)	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

Variables(Contd.)

Variable	Description	Data type	Unique Values	Min	Max
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 (True, False)	NaN	NaN

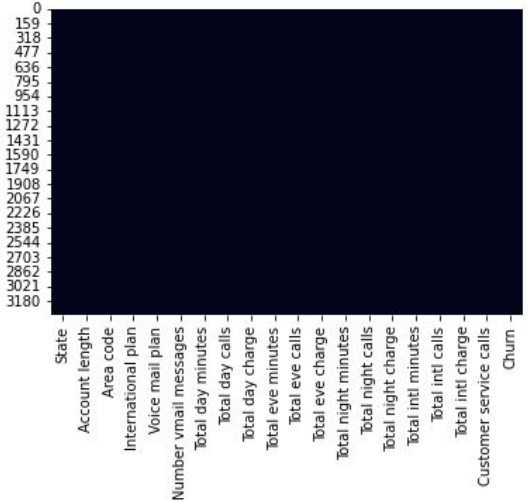
Data Cleaning

```
[ ]  # Checking duplicated rows count
     data.duplicated().sum()

     0
```

```
# Visualizing the missing values
# Checking Null Value by plotting Heatmap
sns.heatmap(data.isnull(), cbar=False)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f438d8ba160>
```



There are no missing values and duplicate values in dataset.

Data Wrangling

Customer Churn on Whole Dataset

	Churn	Churn Rate
False	2850	85.51
True	483	14.49

Customer Churn on the basis of Area code

		Churn	churn rate
Area code	**Churn**		
408	**False**	716	21.48
	True	122	3.66
415	**False**	1419	42.57
	True	236	7.08
510	**False**	715	21.45
	True	125	3.75

Data Wrangling(Contd.)

Area code 408

	mean	Median
Account length	101.88	102.00
Area code	408.00	408.00
Number vmail messages	7.67	0.00
Total day minutes	177.18	176.35
Total day calls	100.50	100.00
Total day charge	30.12	29.98
Total eve minutes	201.28	202.50
Total eve calls	99.79	99.00
Total eve charge	17.11	17.21
Total night minutes	199.21	199.55
Total night calls	99.04	100.00
Total night charge	8.96	8.98
Total intl minutes	10.13	10.10
Total intl calls	4.43	4.00
Total intl charge	2.73	2.73
Customer service calls	1.52	1.00

Area code 415

	mean	Median
Account length	101.07	100.00
Area code	415.00	415.00
Number vmail messages	8.36	0.00
Total day minutes	181.59	180.70
Total day calls	100.58	101.00
Total day charge	30.87	30.72
Total eve minutes	200.65	200.70
Total eve calls	100.50	101.00
Total eve charge	17.06	17.06
Total night minutes	202.04	203.00
Total night calls	100.40	101.00
Total night charge	9.09	9.14
Total intl minutes	10.34	10.50
Total intl calls	4.56	4.00
Total intl charge	2.79	2.84
Customer service calls	1.55	1.00

Area code 510

	mean	Median
Account length	100.25	100.00
Area code	510.00	510.00
Number vmail messages	8.02	0.00
Total day minutes	178.79	179.45
Total day calls	100.10	100.00
Total day charge	30.39	30.51
Total eve minutes	201.32	201.95
Total eve calls	99.67	100.50
Total eve charge	17.11	17.16
Total night minutes	200.23	197.95
Total night calls	100.60	100.00
Total night charge	9.01	8.91
Total intl minutes	10.14	10.20
Total intl calls	4.37	4.00
Total intl charge	2.74	2.75
Customer service calls	1.62	1.00

Data Wrangling(Contd.)

International plan	Churn	Churn	churn rate
No	False	2664	79.93
	True	346	10.38
Yes	False	186	5.58
	True	137	4.11

Voice mail plan	Churn	Churn	churn rate
No	False	2008	60.25
	True	403	12.09
Yes	False	842	25.26
	True	80	2.40

- Those who has international plan their churn rate is high.
- Around 42 % customers are churned.

- No such difference

Data Wrangling(Contd.)

International + Voicemail

Churn	Churn_count	rate
False	56	60.87
True	36	39.13

Only Voicemail

Churn	Churn_count	rate
False	786	94.7
True	44	5.3

Only International

Churn	Churn_count	rate
False	130	56.28
True	101	43.72

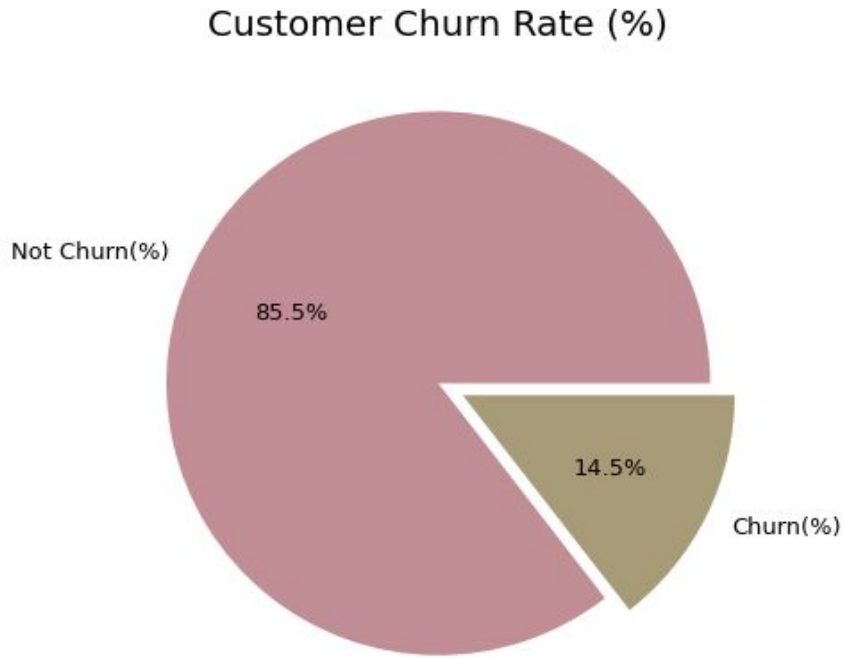
No International No Voicemail

Churn	Churn_count	rate
False	1878	86.15
True	302	13.85

Data Visualization

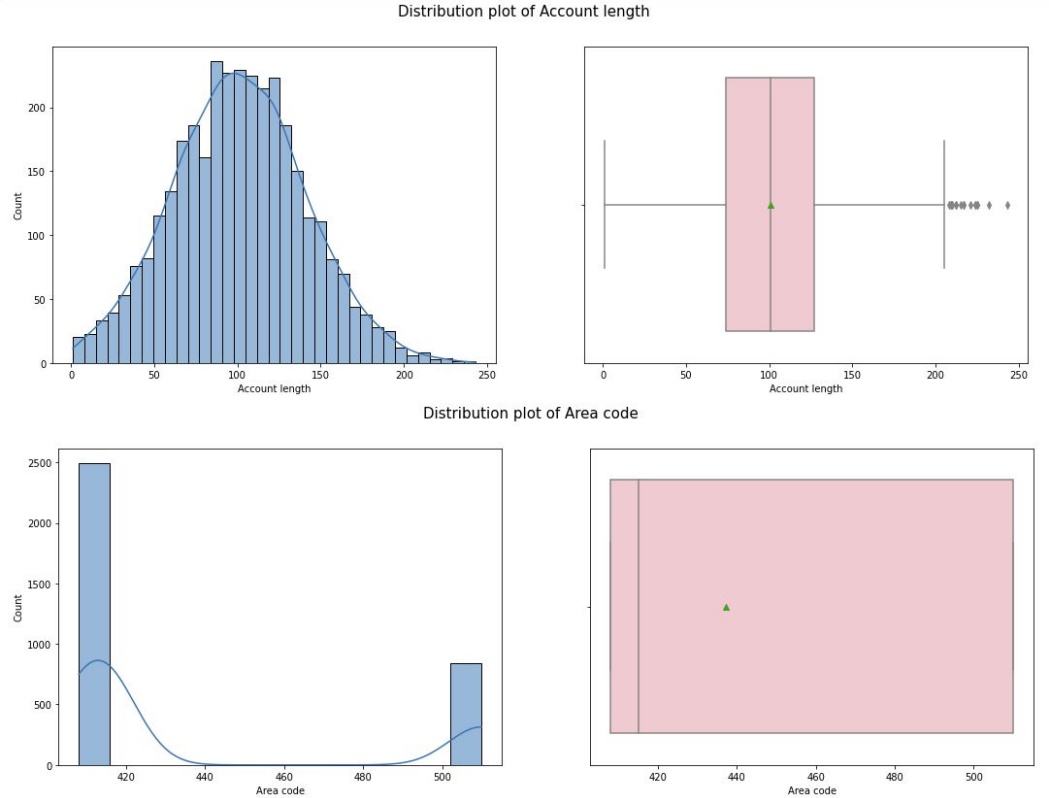
Dependent Variable

```
False    2850
True      483
Name: Churn, dtype: int64
```



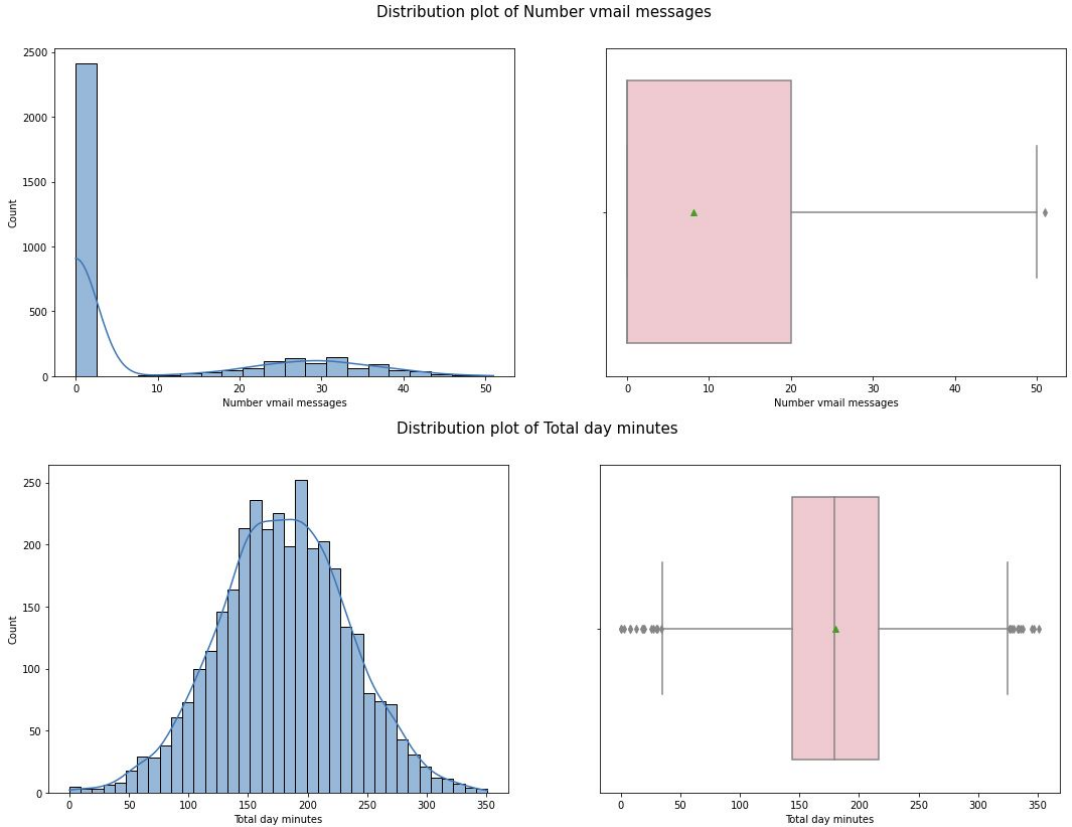
Data Visualization

Distribution Plot and Box plot (Univariate Analysis)



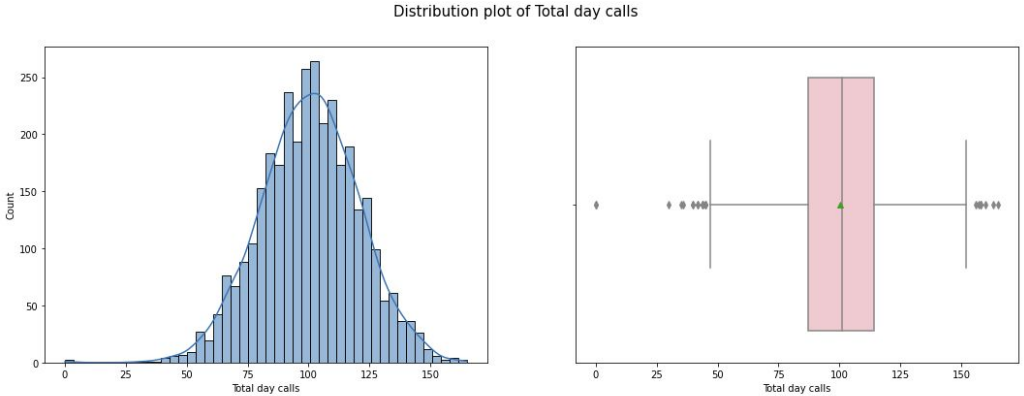
Data Visualization

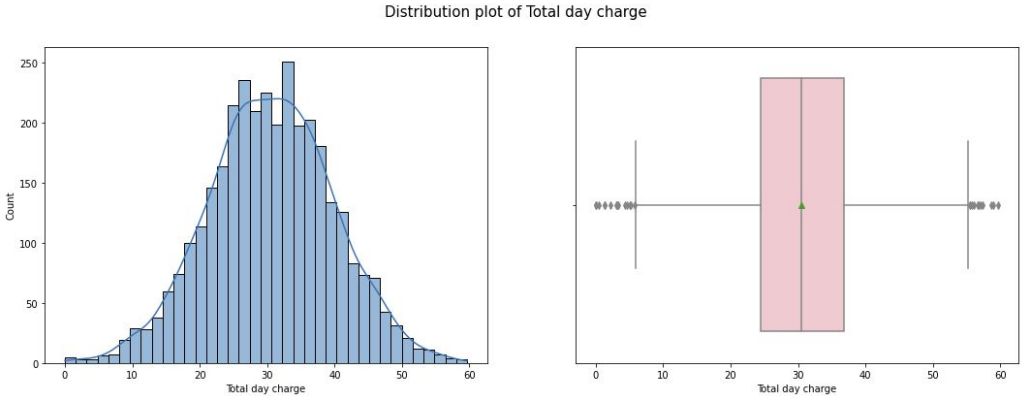
Distribution Plot and Box plot(Contd.)



Data Visualization

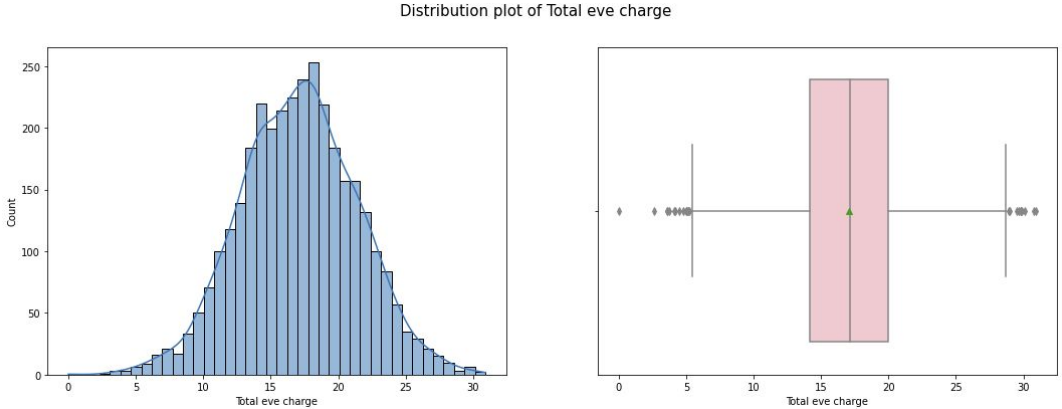
Distribution Plot and Box plot(Contd.)





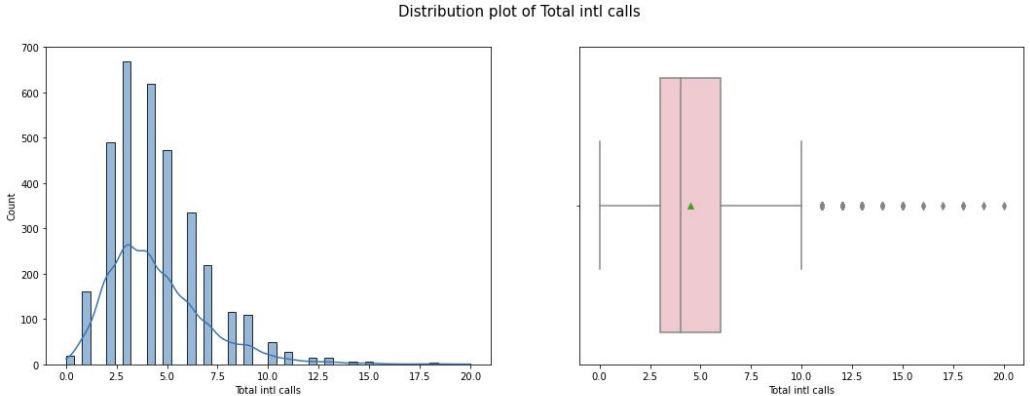
Data Visualization

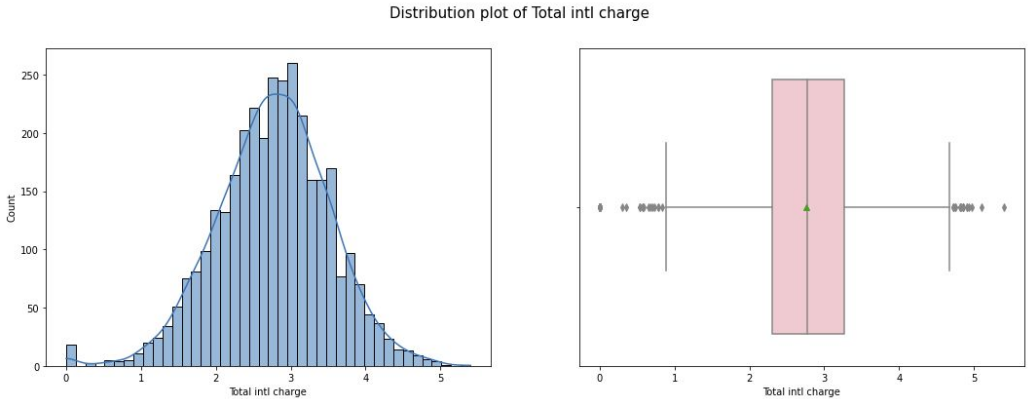
Distribution Plot and Box plot(Contd.)



Data Visualization

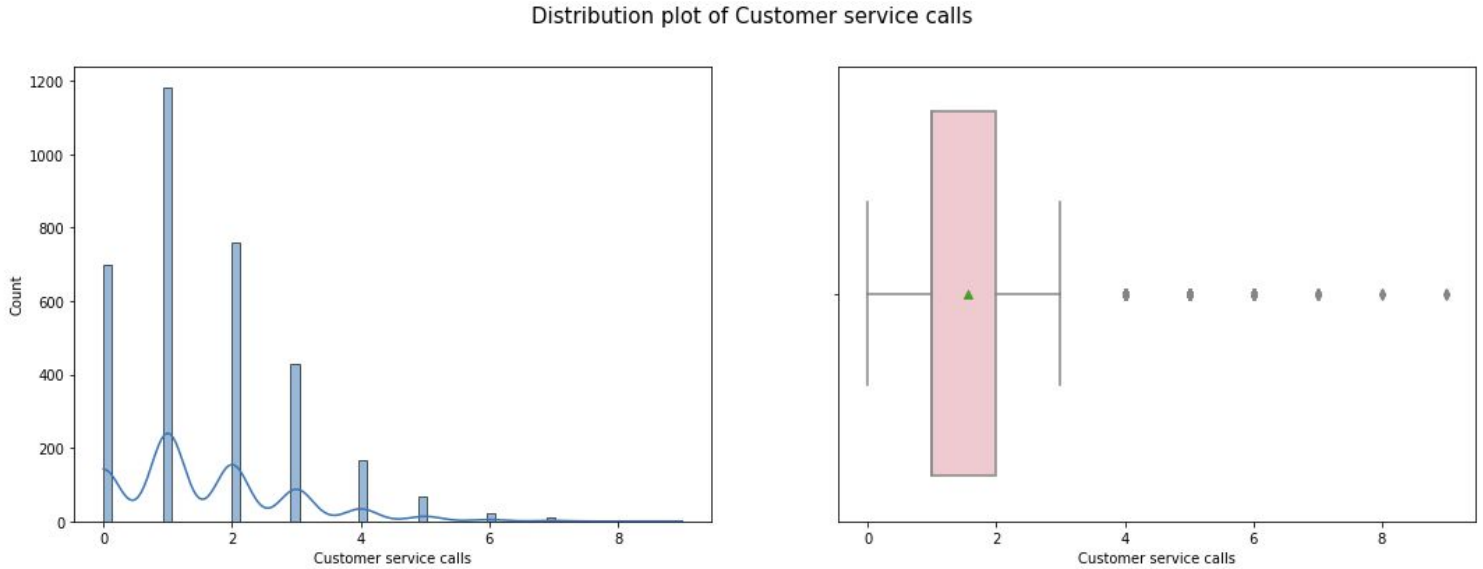
Distribution Plot and Box plot(Contd.)





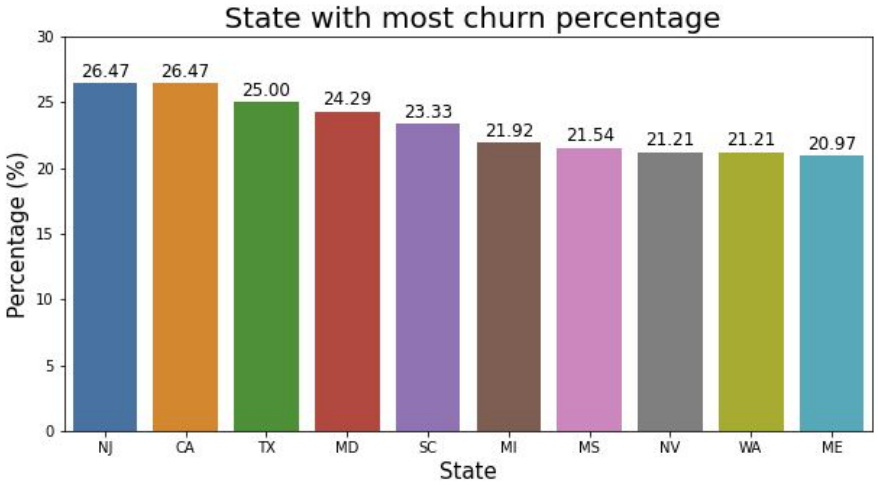
Data Visualization

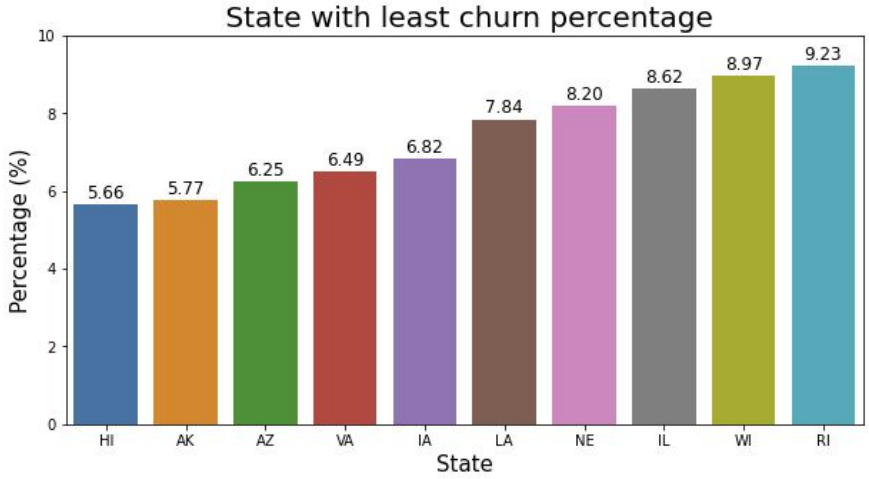
Distribution Plot and Box plot(Contd.)



Data Visualization

State vs Average Churn Percentage (Bivariate Analysis)

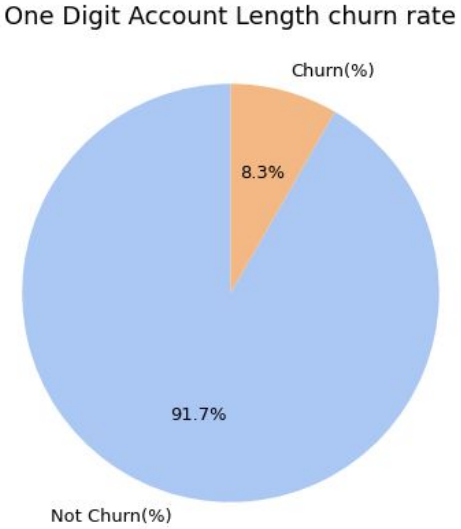




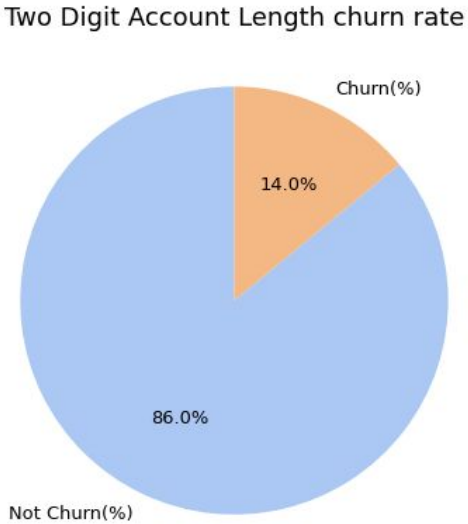
Data Visualization

Account Length Pie Chart

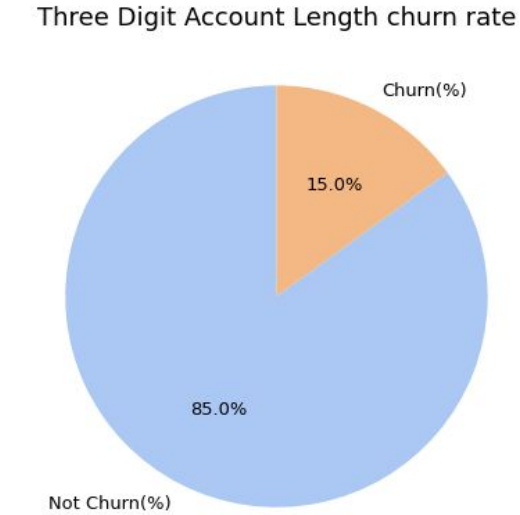
```
Churn
False    22
True      2
dtype: int64
```




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

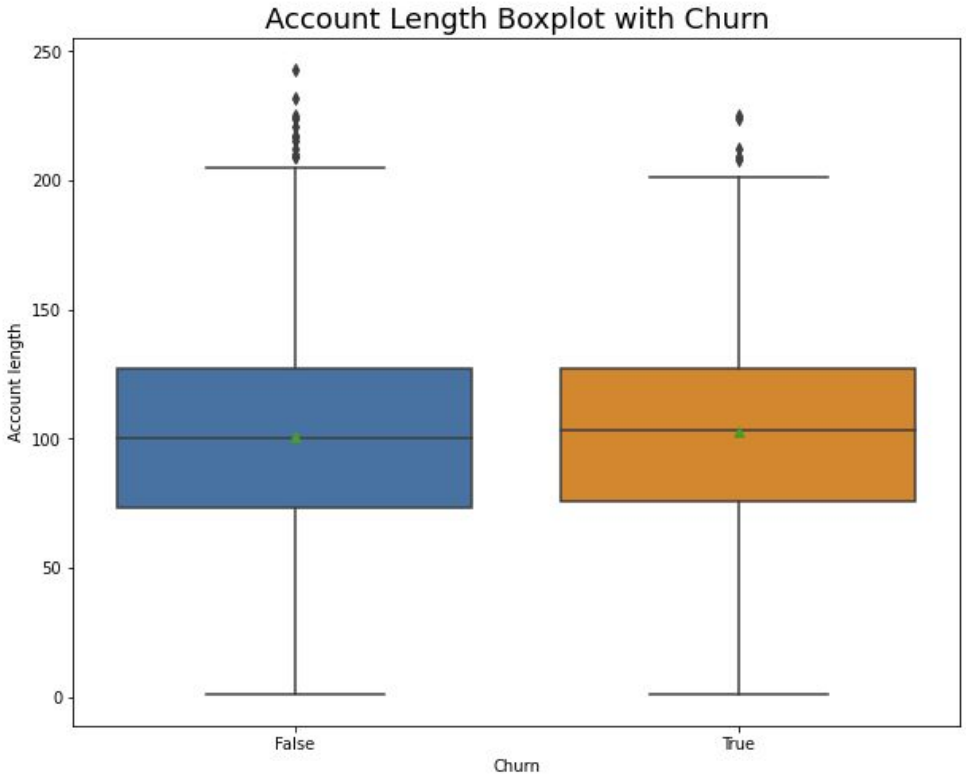



```
Churn
False    1450
True      256
dtype: int64
```



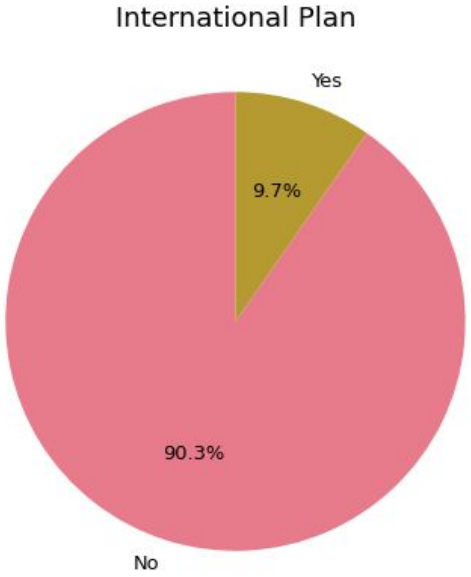
Data Visualization

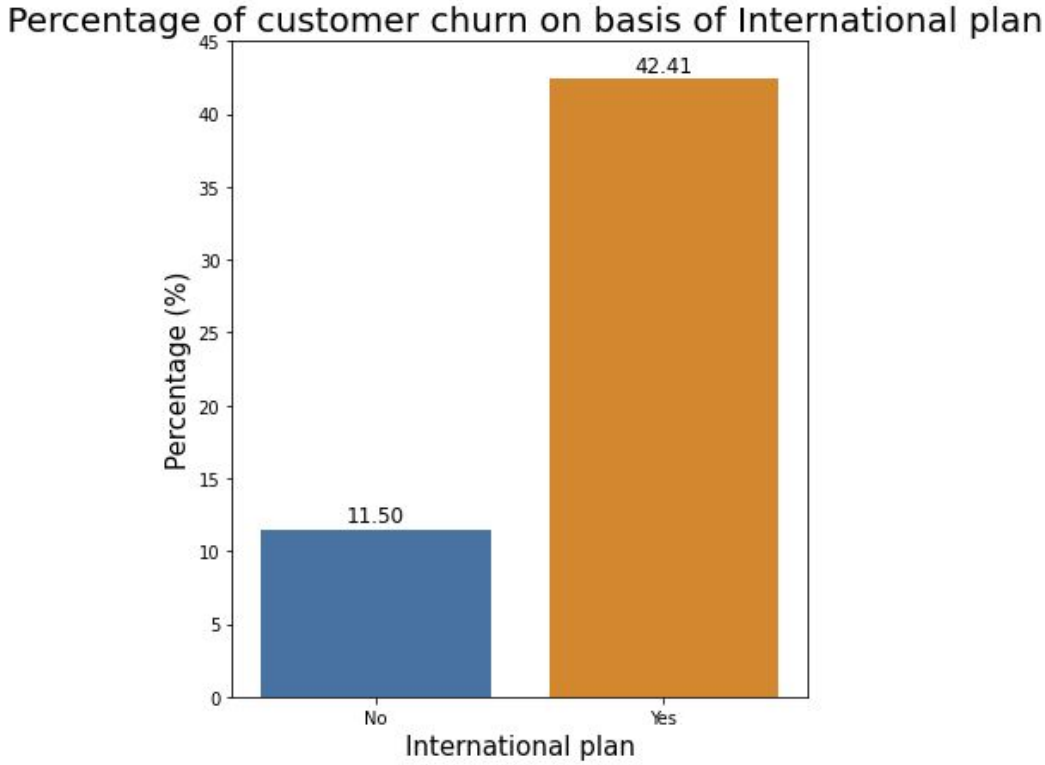
Account Length with Churn Box Plot



Data Visualization

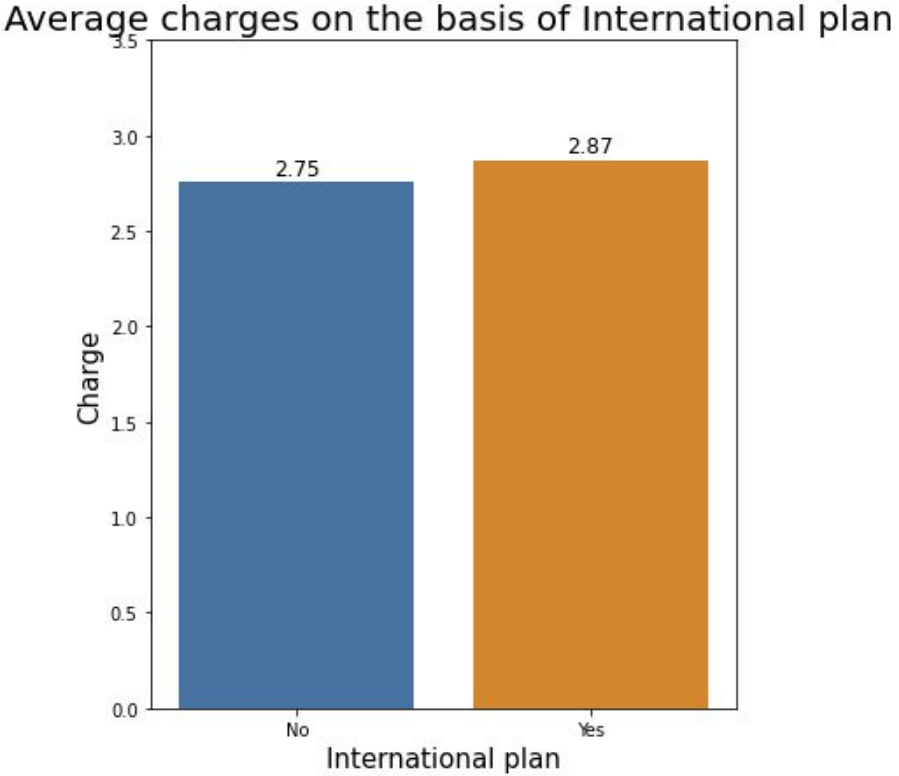
International Plan

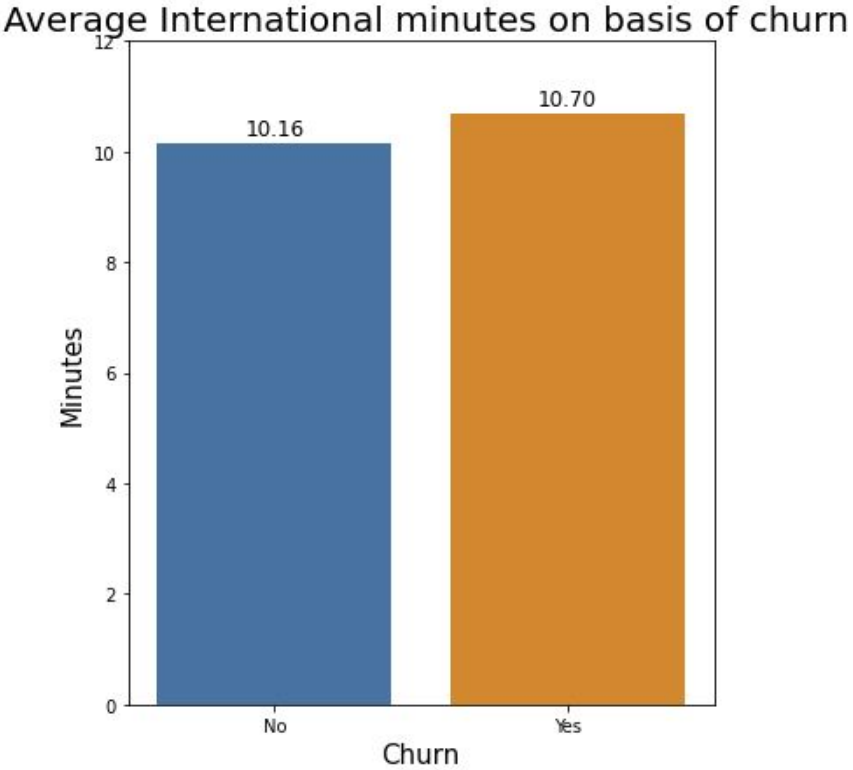




Data Visualization

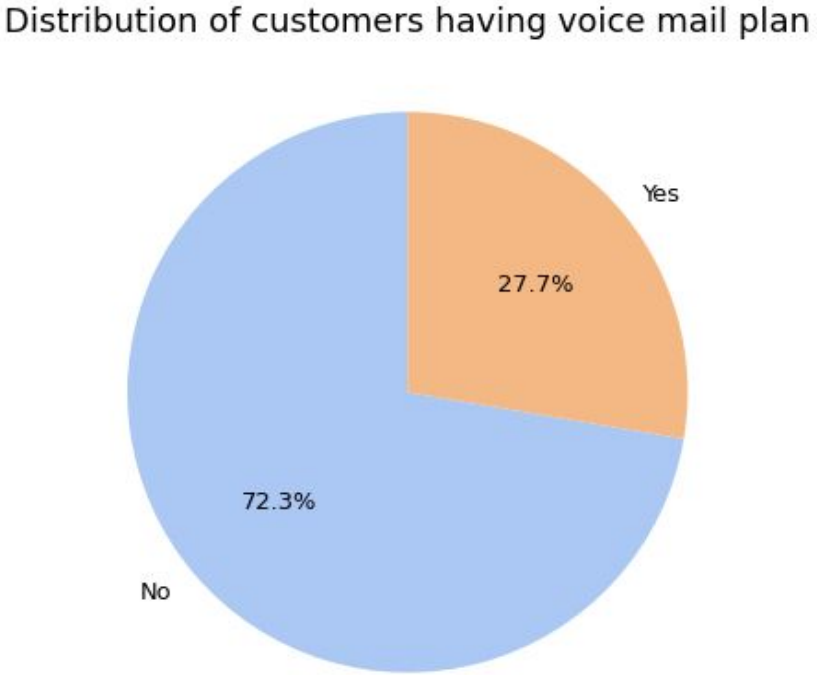
International Plan(Contd.)

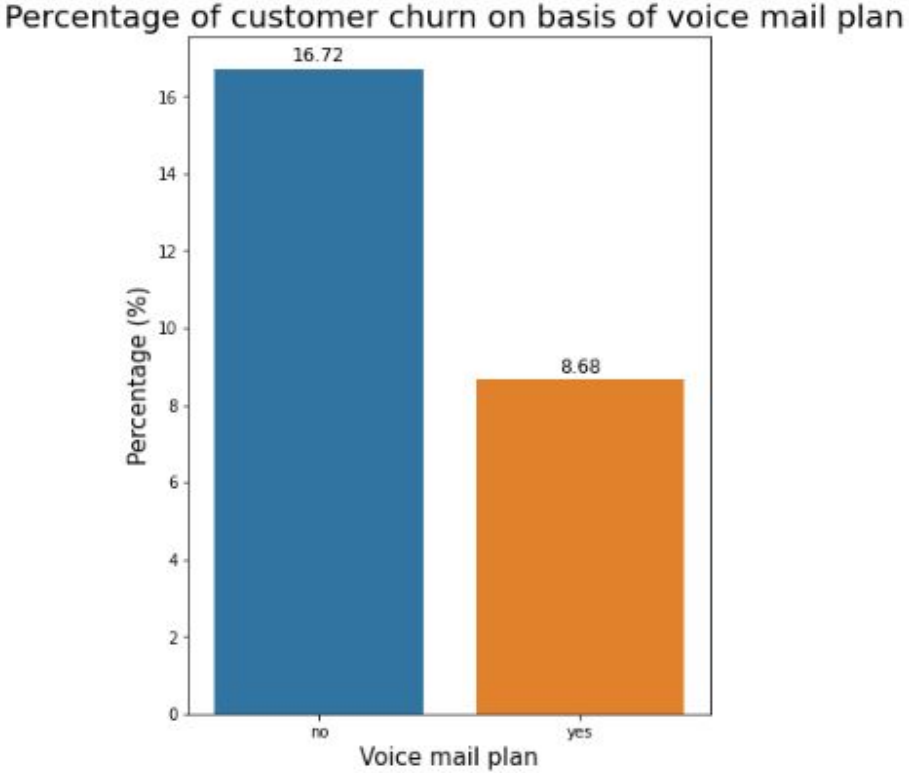




Data Visualization

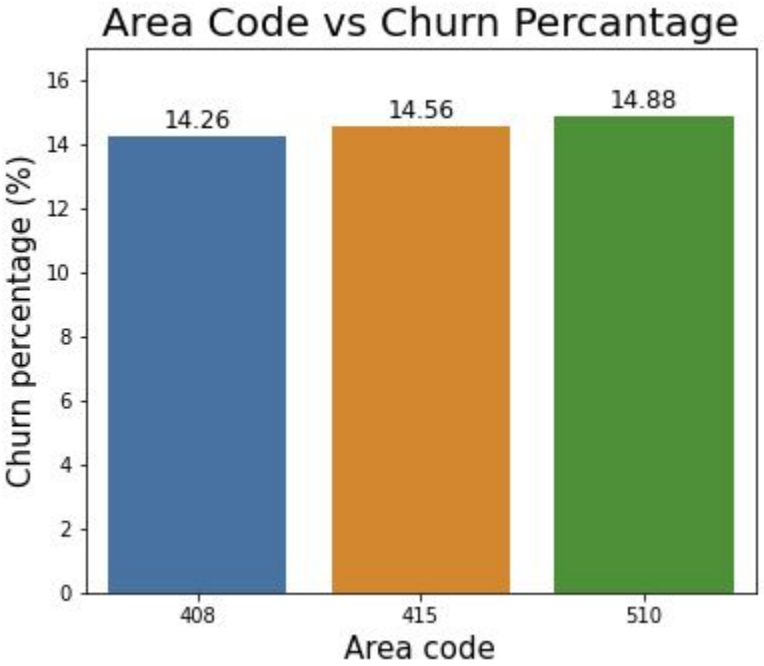
Voicemail Plan





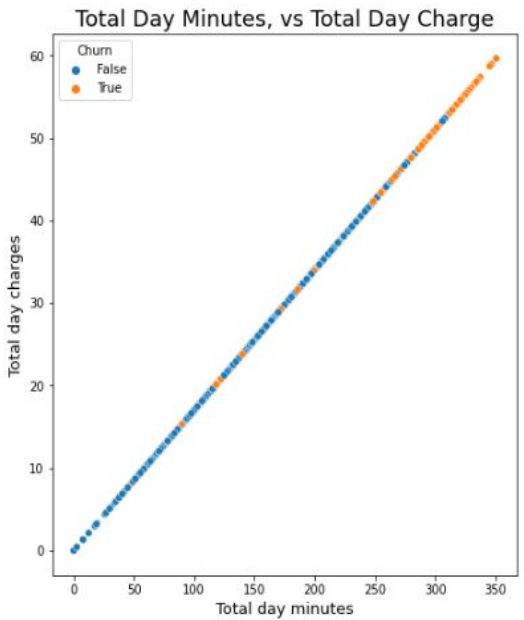
Data Visualization

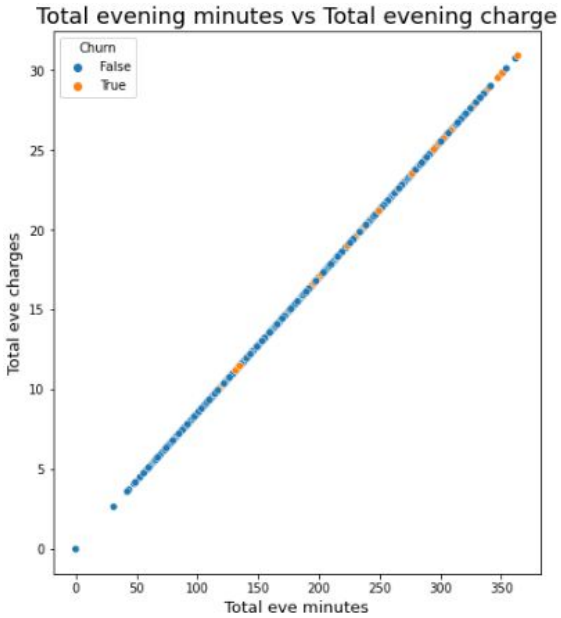
Area code vs Churn Percentage

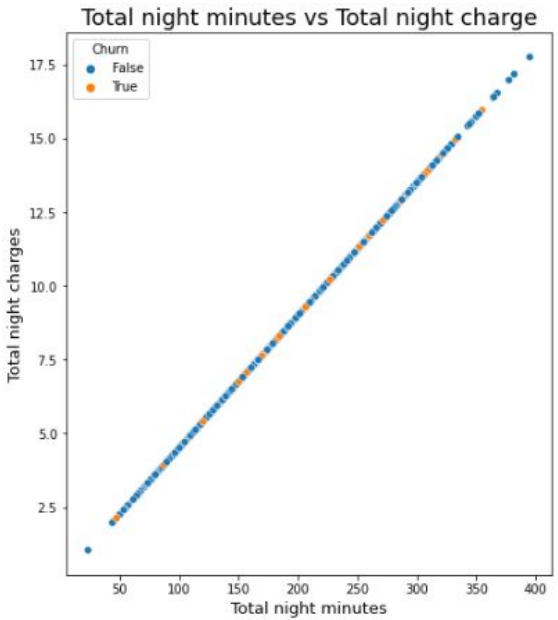


Data Visualization

Minute vs charge

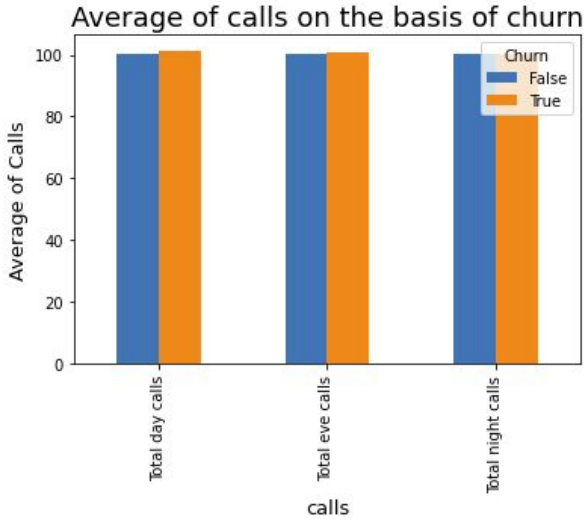


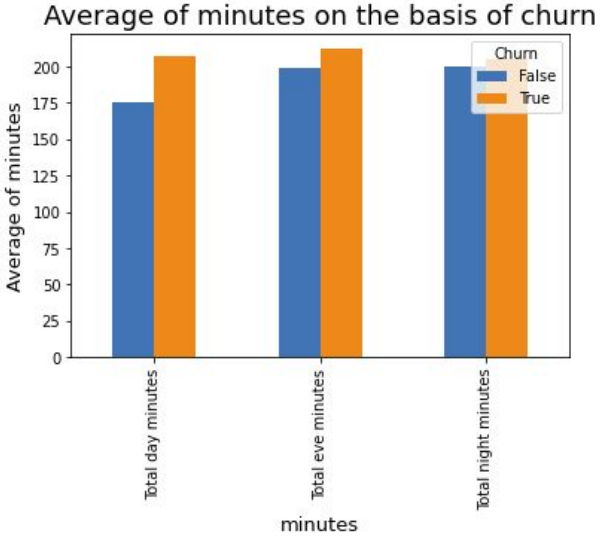


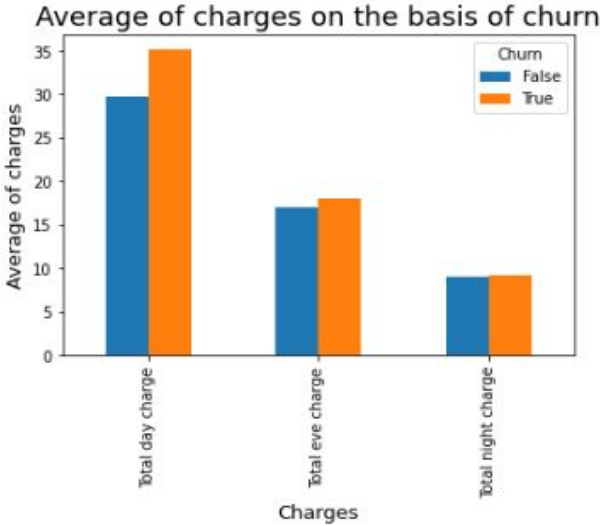


Data Visualization

Average of calls on the basis of churn

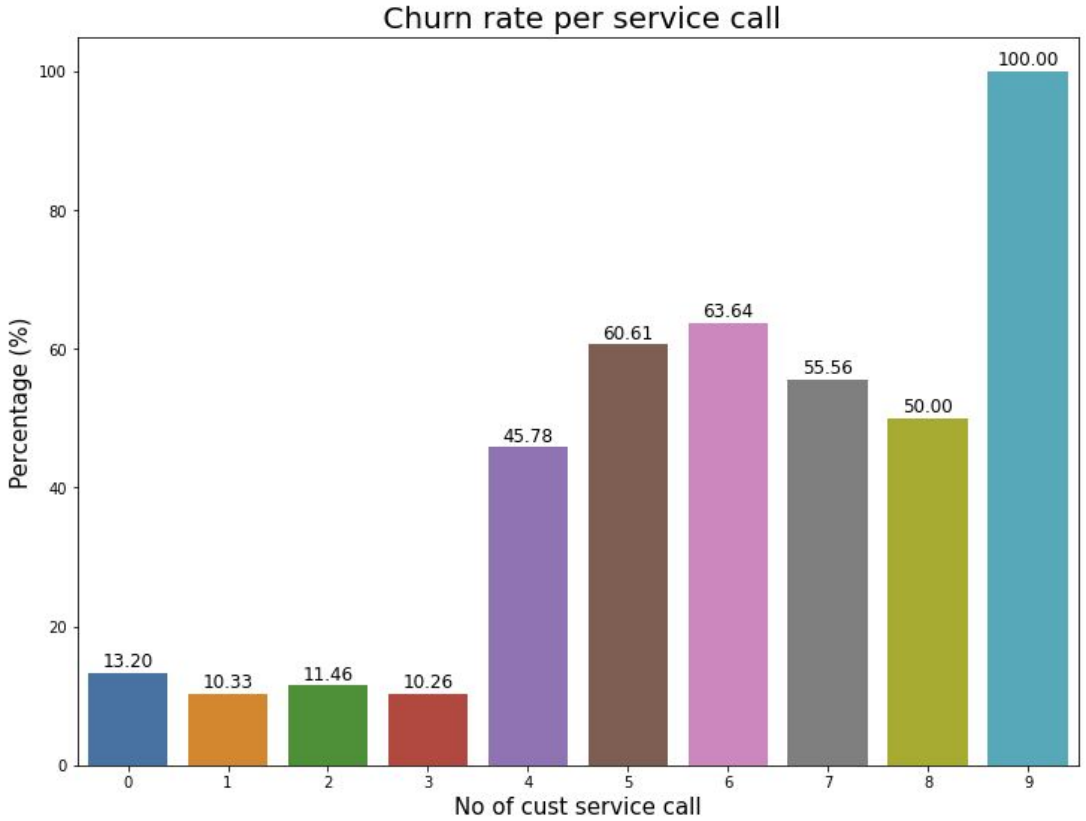






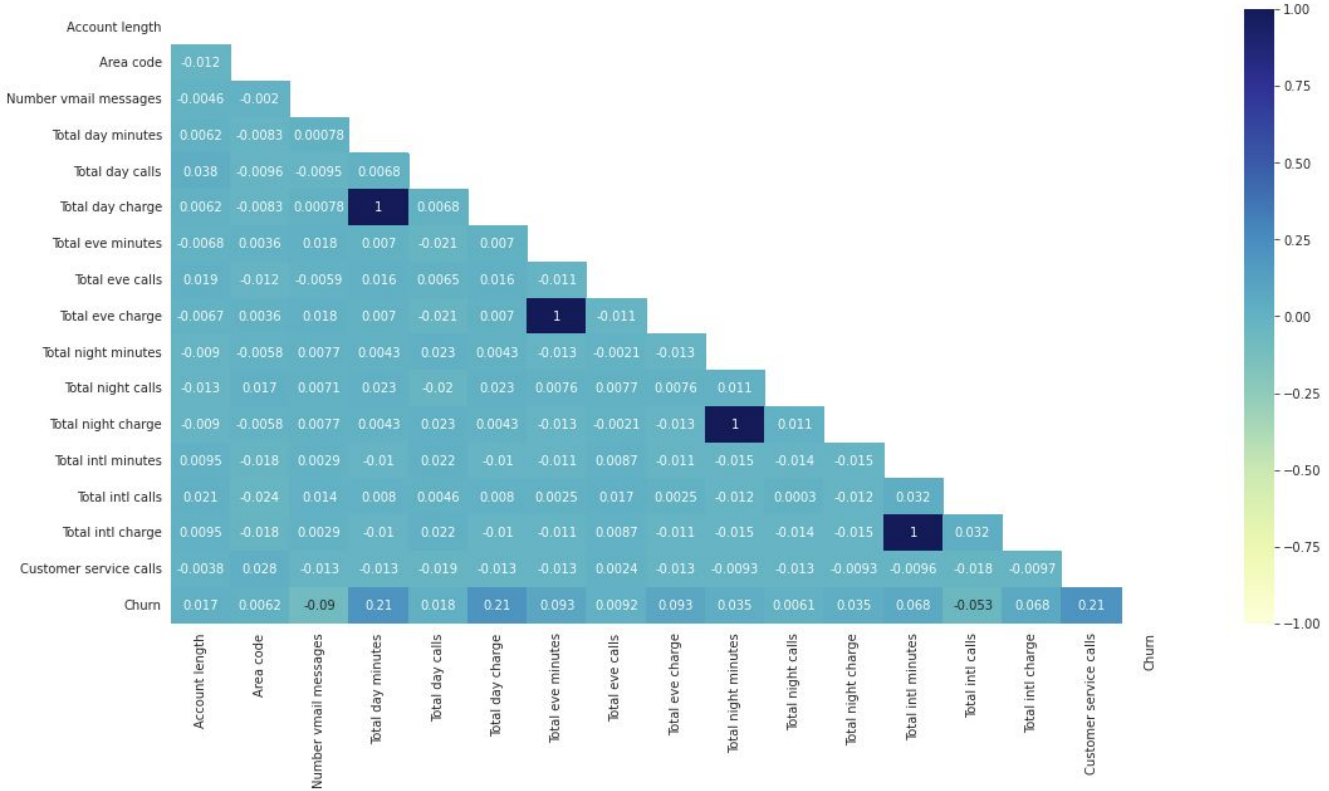
Data Visualization

Service Call vs Churn Percentage



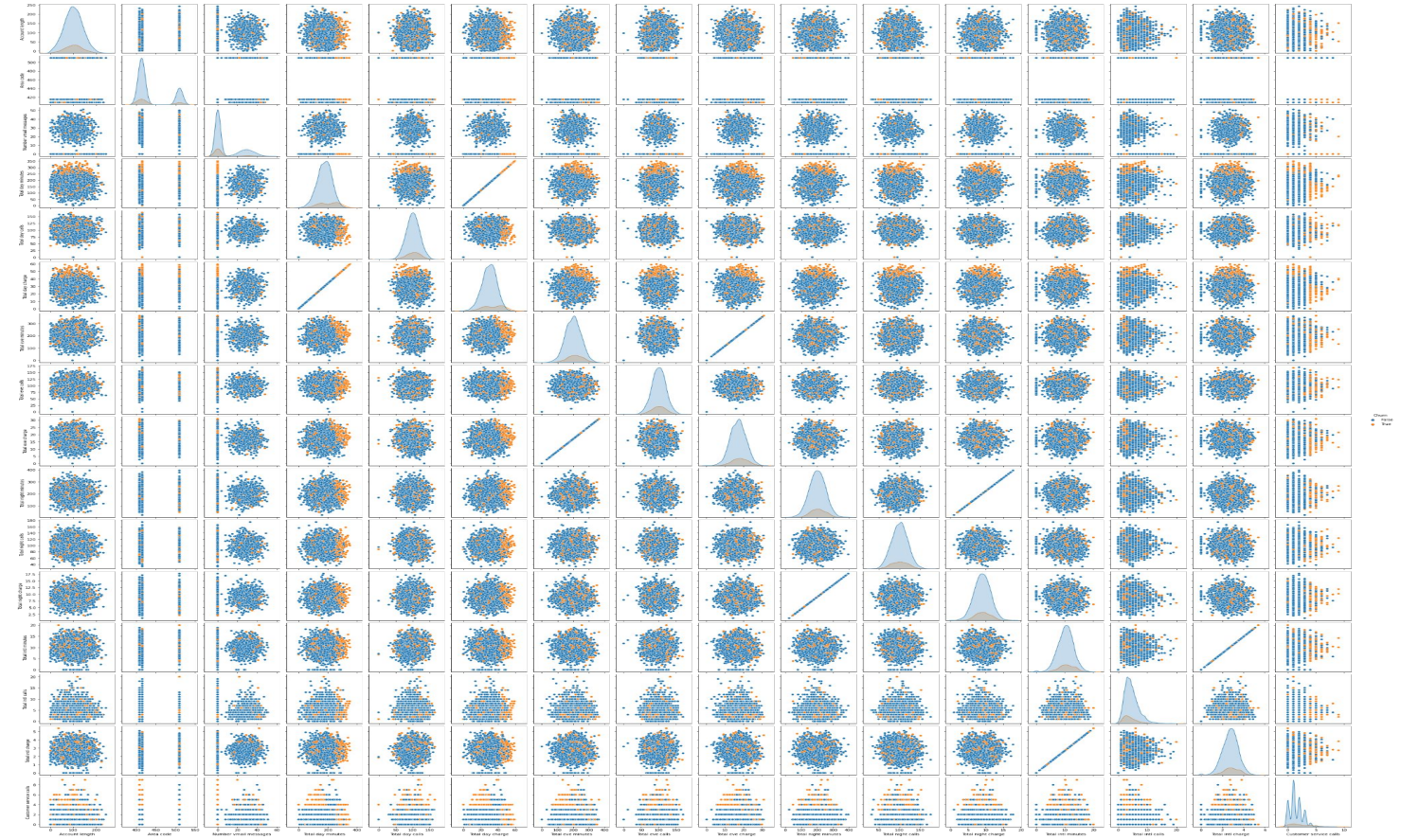
Data Visualization

Correlation Heatmap



Data Visualization

Pair Plot



Conclusion

- The four charge fields are directly related to the minute fields.
- The area code 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.

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.

The End

Thank You