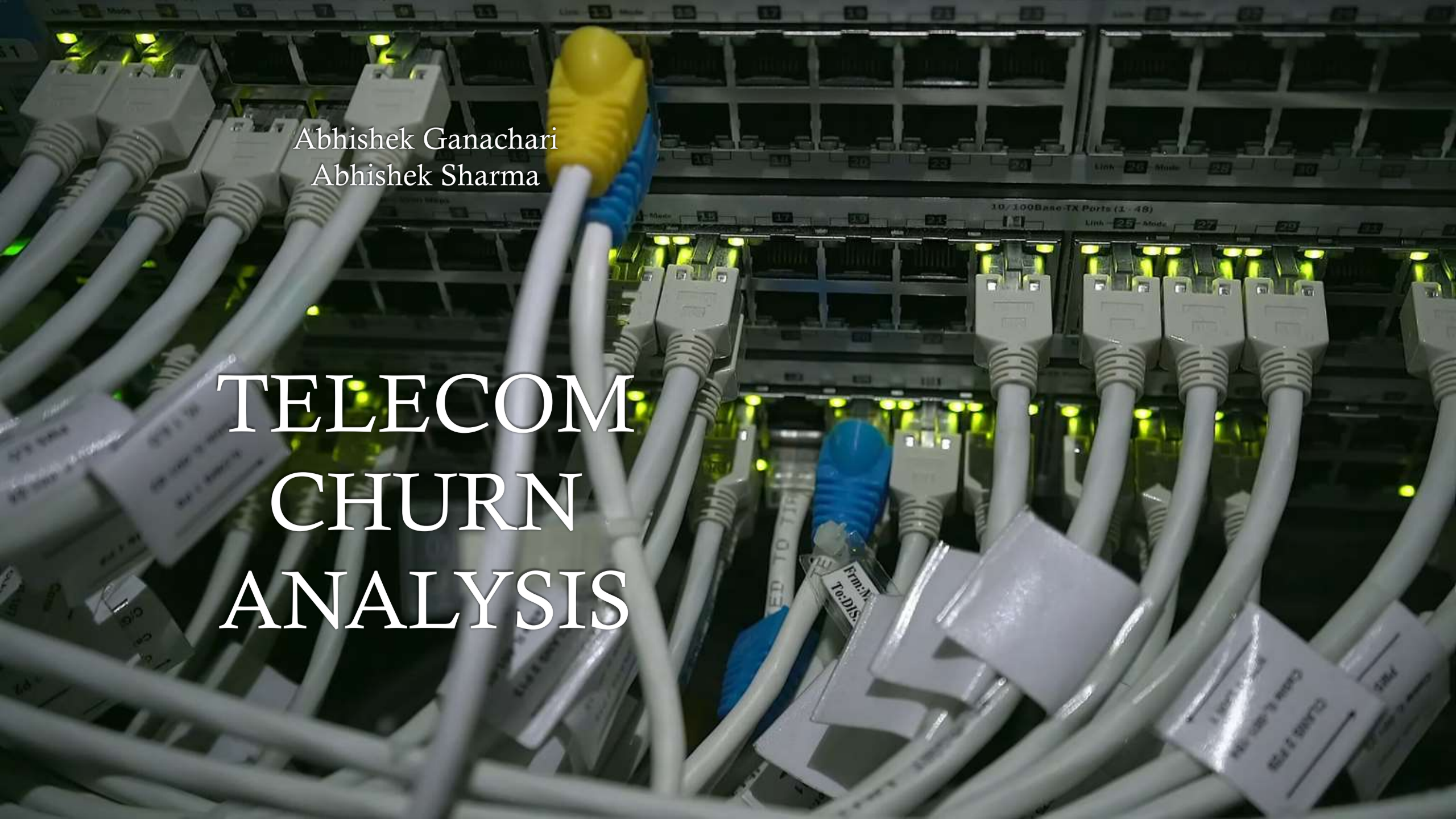


Abhishek Ganachari
Abhishek Sharma

TELECOM CHURN ANALYSIS

Content

- Problem statement
- Purpose
- Detailed data summary
- Exploratory Data Analysis (EDA)
- Proposals
- EDA Final Verdict
- Feature engineering, model construction, and deployment.

Problem statement

Customer churn is a major issue in the telecom sector.

Churn prediction is simply identifying consumers who are likely to discontinue their service subscription.

Churn is a concern in the telecom sector since it costs more to recruit a new customer than it does to prevent an existing one from leaving.

Purpose

Increase the company's earnings by keeping customers.

Reduce the Customer churn by discovering the root cause of the issue.

Provide deals and discounts while improving service quality without sacrificing profit.

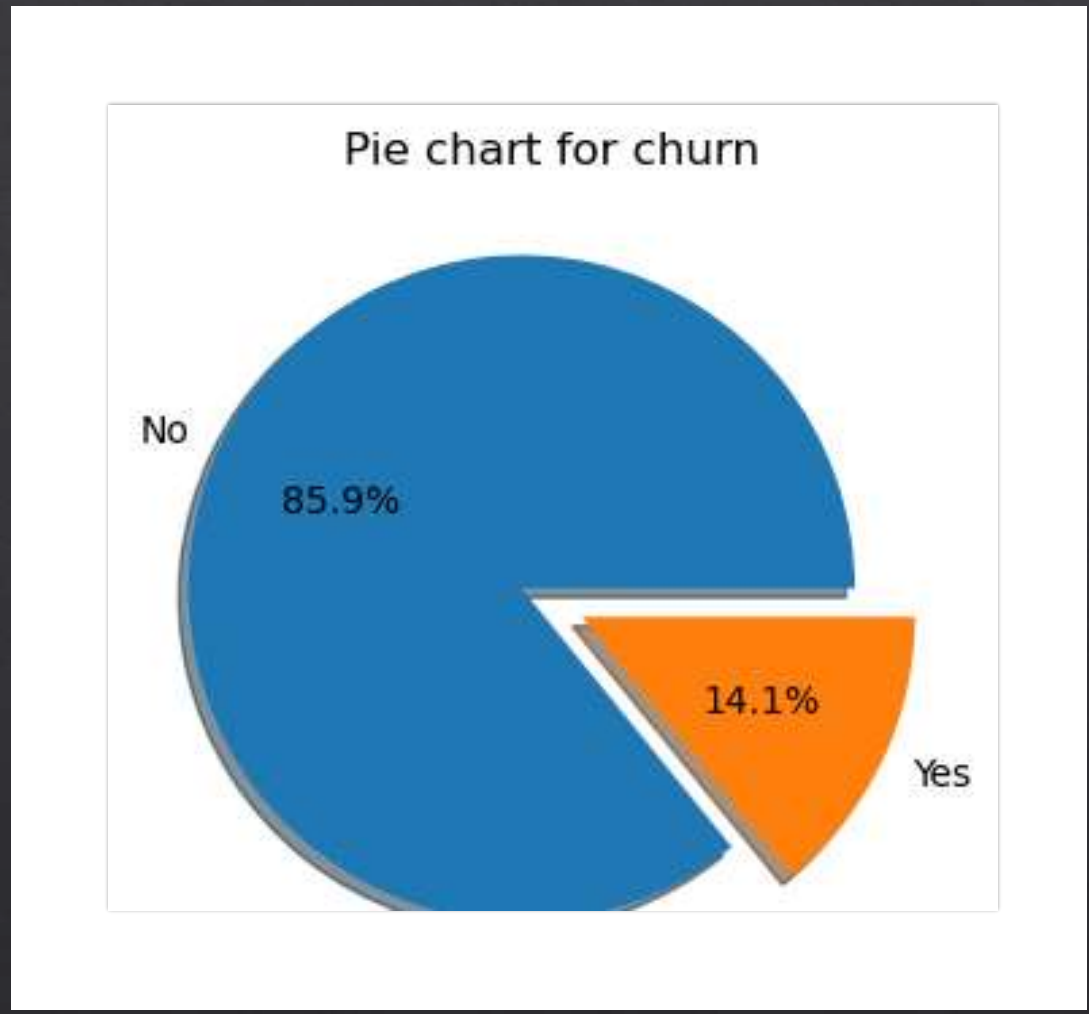
Detailed data summary

- The first five data sets
- The most recent five data sets
- Shape of Data - (5000,21)
- Checking for Data Types - We have modified the data types of some of the data sets.
- Checking for null values - we had some null values in the day charge and eve min columns that we had populated with mean values (because they were numeric columns).
- Detecting duplicate values - 0
- Checking for distinct values

Exploratory Data Analysis - Churn

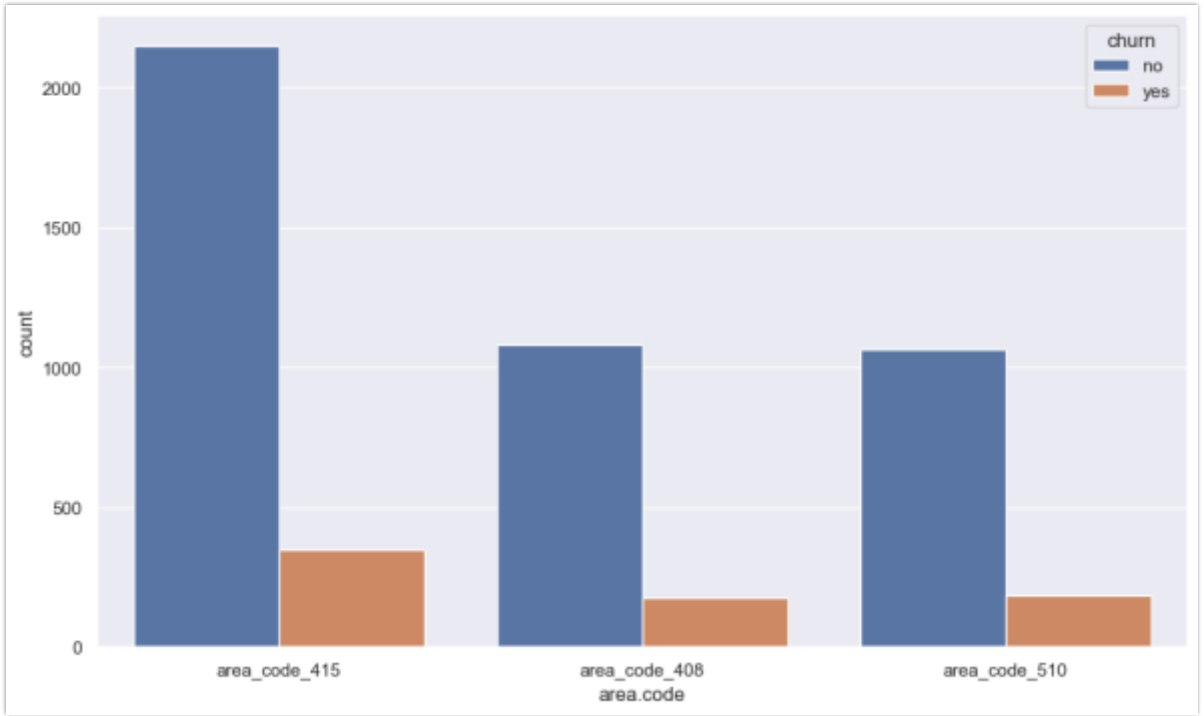
- Total number of consumers in the data set = 5000
- Total customer churn = 707 no.
- Number of unchurned consumers = 4293

- The percentage information are presented in the chart.



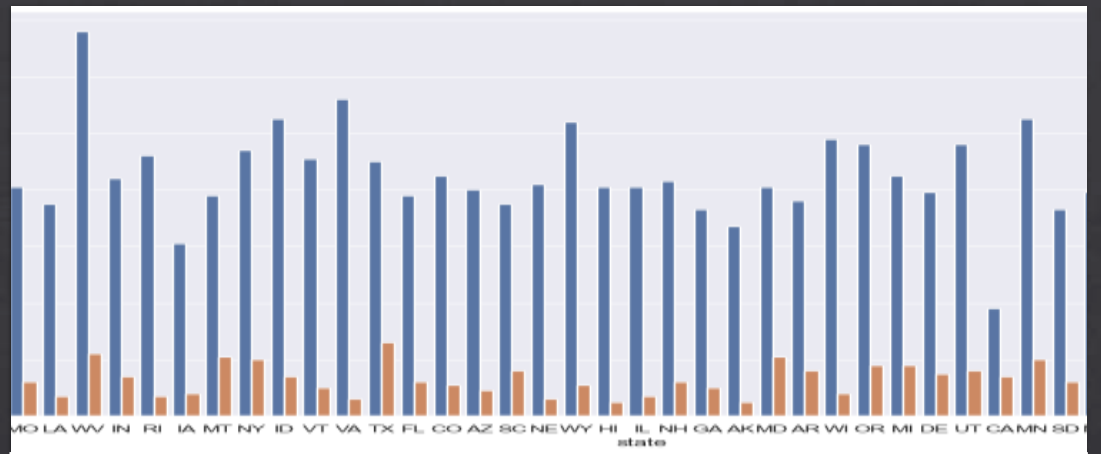
Exploratory Data Analysis – Area code

- In the given data, we can see that there are only three distinct values, namely 415,408,510, and the churn rate of these area codes is nearly same.
- We do not believe there is any relationship between the "area code" and the "churn" caused by the customer leaving the operator.



Exploratory Data analysis - State

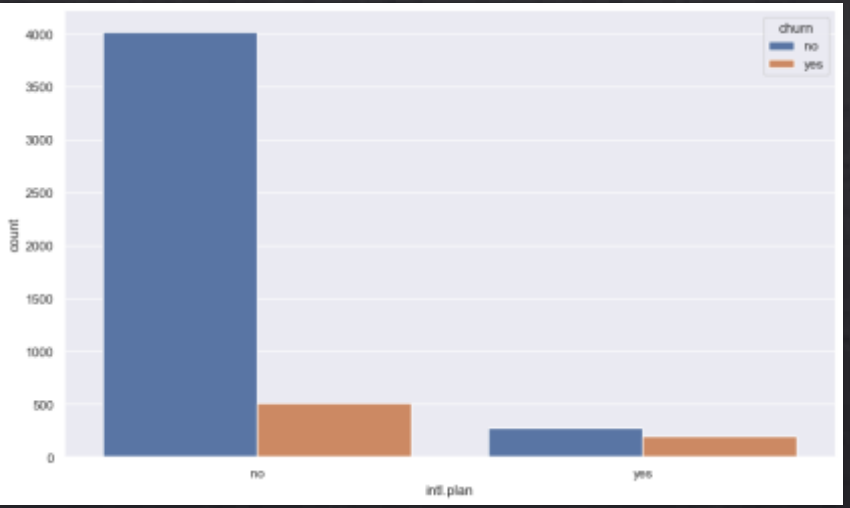
- Following analysis, it is discovered that there are a total of 51 distinct states.
- Ten of these 51 distinct states have a higher churn rate of 21%.
- This churn rate could be attributed to the cellular network's limited service region.



Sr. No	State Name	% Churn > 21%
1	CA	26.9
2	NJ	25
3	WA	24.4
4	TX	22.4
5	MT	21.21
6	MD	20.58
7	NV	18.88
8	ME	18.44
9	KS	18.18
10	OK	17.77

Exploratory Data Analysis – International plan

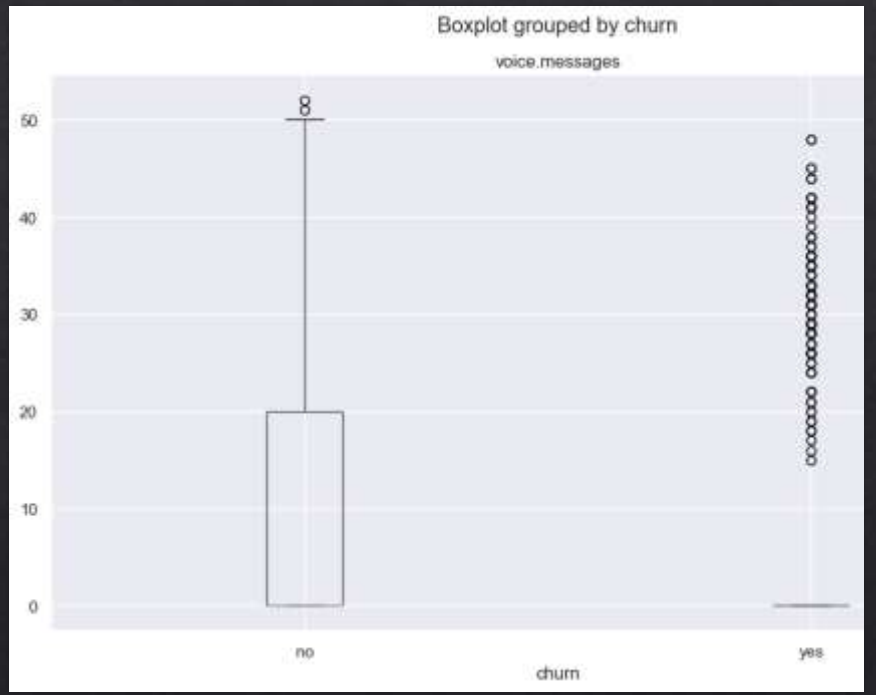
- The total number of clients without an international plan is 4527.
- There are 473 foreign plan subscribers in total.
- As a result, we see that consumers who have chosen an international plan have a 42% turnover rate.
- People who purchased International plans are essentially churning in large numbers.Most likely due to network troubles or a hefty call charge.



Sr. No	Plan Type	Not Churn	Churn	% Churn
1	No International Plan	4019	508	11.22%
2	With International Plan	274	199	42.07%

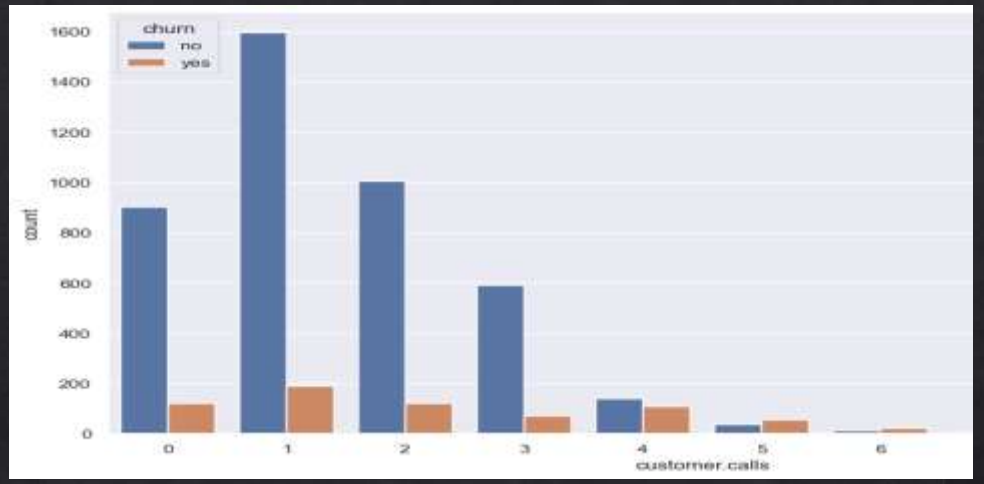
Exploratory Data Analysis – Voice message

- After analysing the aforementioned voice mail feature data, we can conclude that there is churn when there are more than 20 voice-mail messages.To do so, we must improve the quality of voice mail.



Exploratory Data Analysis – Customer Calls

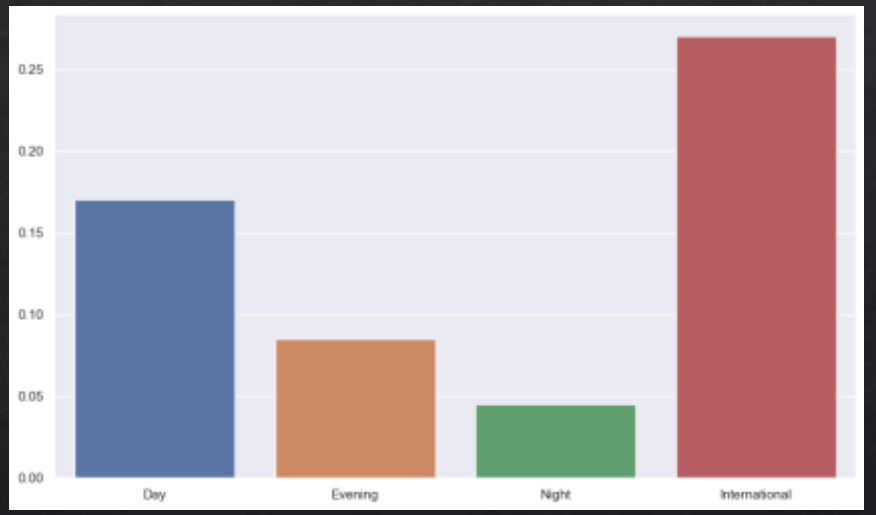
- According to the data presented above, customer vehicle services are not good at responding to customer inquiries, and customers who have phoned more than five times have a 60% attrition rate. As a result, they must upgrade their services.



Sr. No	No of Customer calls	Not Churn	Churn	% Churn
1	0	902	121	11.8
2	1	1596	190	10.63
3	2	1005	122	10.82
4	3	592	73	10.97
5	4	141	111	44.04
6	5	38	58	60.41
7	6	12	22	64.7
8	7	6	7	53.84
9	8	1	1	50
10	9	0	2	100

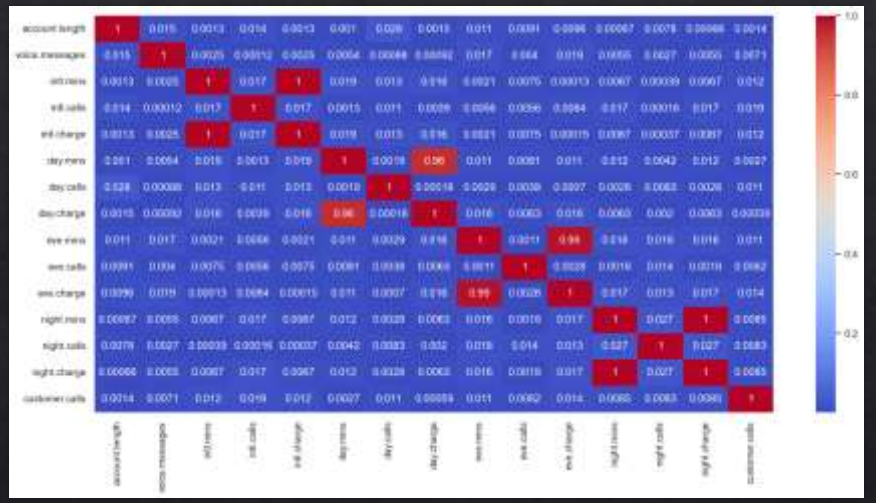
Exploratory Data Analysis – Remaining columns

- After careful consideration, Day Charge vs Day Mins
- Eve Mins vs Eve Charge
- Night Mins vs. Night Charge
- International charges vs International Mins
- We conclude that the aforementioned dataset shows that total day/night/evening minutes/calls/charges do not indicate any kind of causation for churn rate.
- However, international call charges are high in comparison to others; this is a clear fact, but it may be a reason for international plan subscribers to leave.



Proposal

- Increase network coverage in the churned state
- Customers can get certain discounts from overseas plans.
- Improve voice mail quality or solicit customer feedback
- Give a discount to customers who spend more time with you.
- Improve call Centre service by soliciting frequent input from consumers about their problems and attempting to resolve them as quickly as feasible.



Some columns, such as 'Day Charge' and 'Day Minutes,' show a strong association, and these features will be deleted during feature selection to avoid 'Multi collinearity Issues.'

The Final Verdict from EDA

Customers who used foreign plan services had a greater turnover rate.

So, in general, folks who purchased International insurance are churning in large numbers. Most likely due to network troubles or a hefty call charge.

When there are more than 20 voice mail messages in the voicemail section, there is churn, which basically implies that the voice mail quality is poor.

Customers with a high number of day call minutes and a high call fee are more likely to churn.

Feature Engineering & Model Building

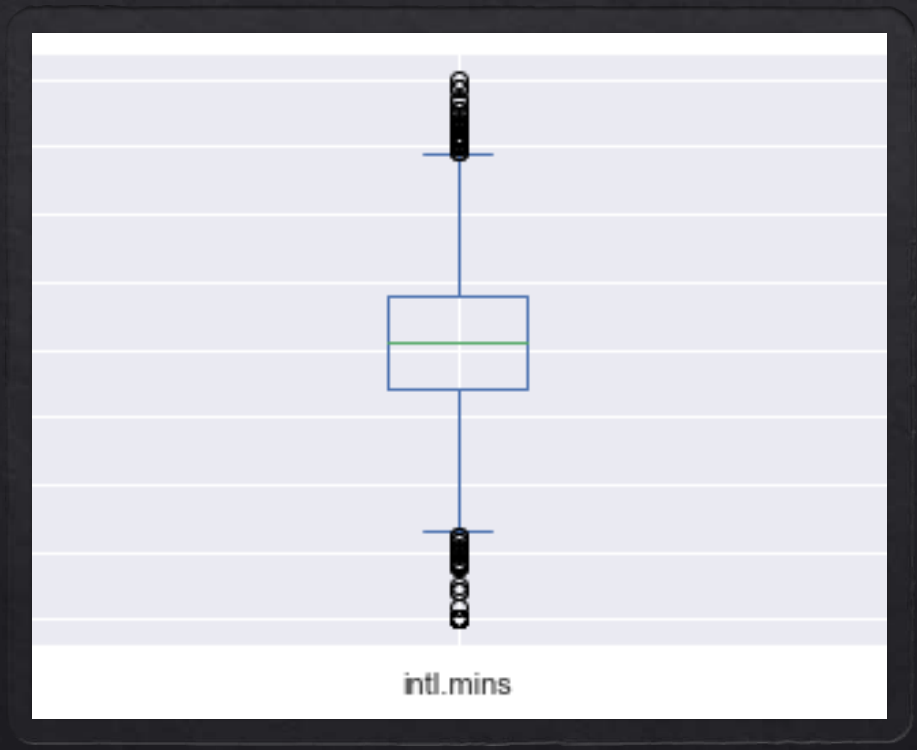
OUTLIER DETECTION AND FEATURE ENGINEERING.

THE DATA SET IS BEING BALANCED.

MODEL CONSTRUCTION WITH MANY MODELS.

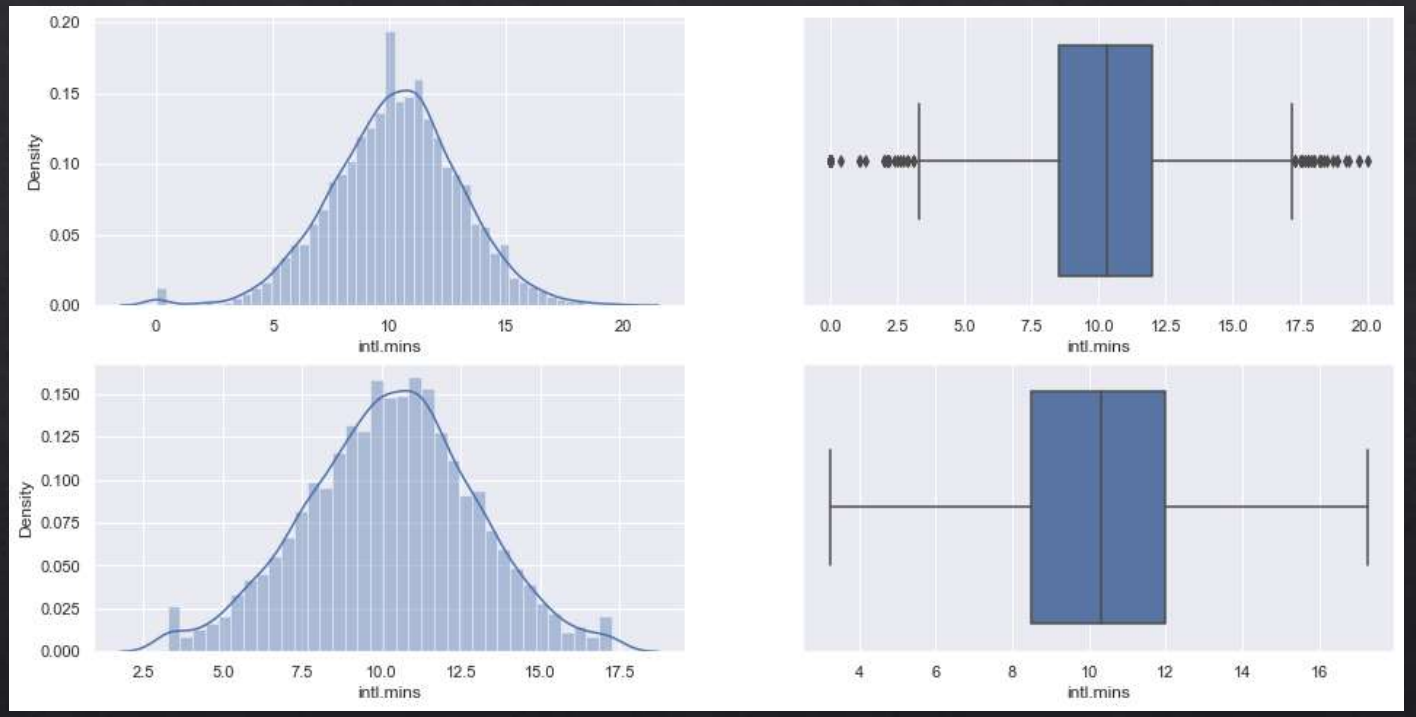
Outlier Detection Part

We used the IQR approach to detect outliers in the international minutes column first.



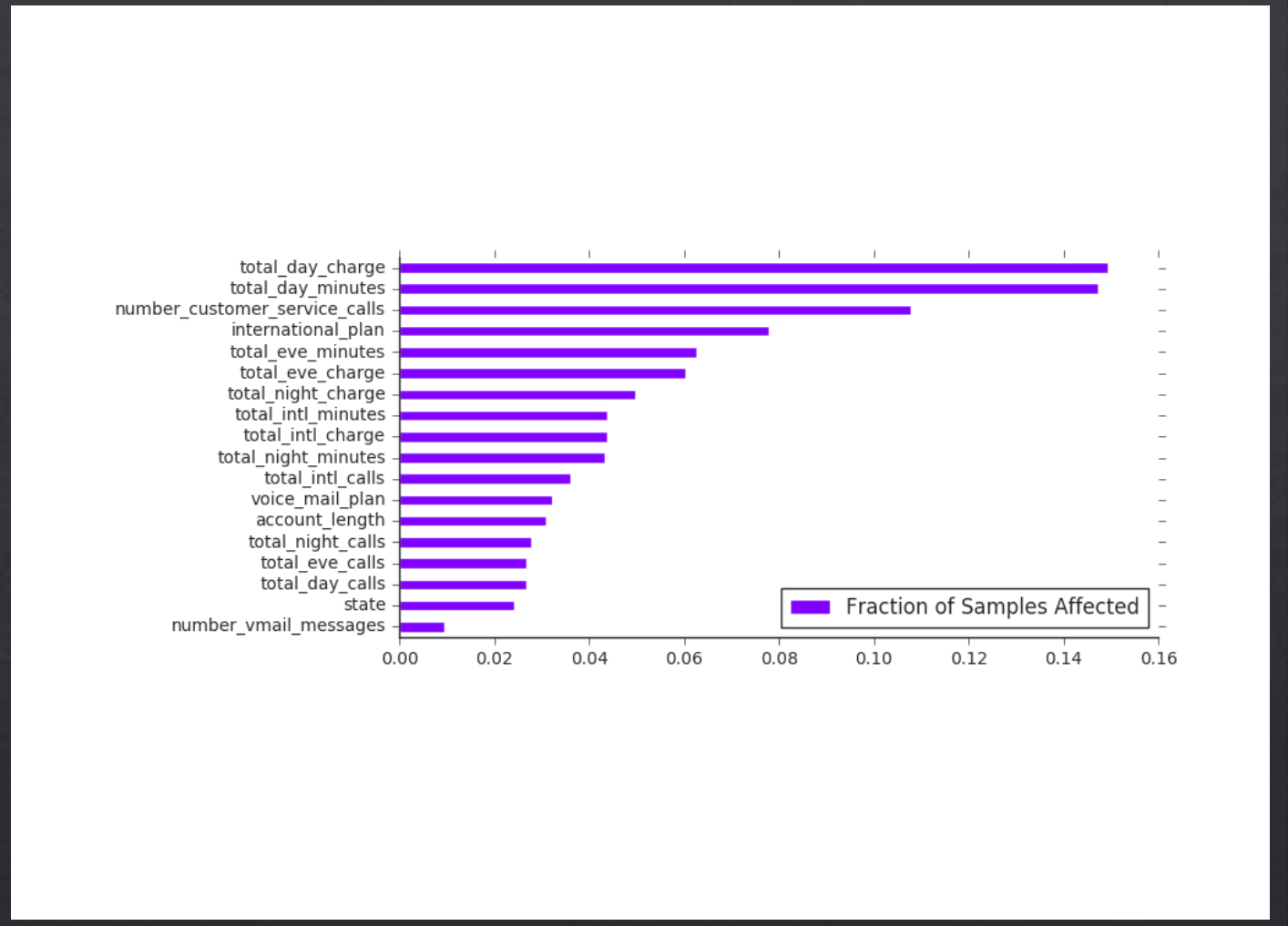
Outlier Detection Part

- The following were the outcomes of using the IQR method:
- Capping:-



Feature Engineering

- Identifying the trait with the highest predictive power.



Feature Engineering

- Identifying the trait with the highest predictive power. To find the best features, we employed the random forest algorithm.
- According to the Random Forest algorithm's feature importance study, the following features had the highest predictive power:
- total_day_minutes total_day_charge total_day_calls total_eve_minutes international_plan

- Splitting the data into train and test subsets: We divided the data into train and test subsets.

Feature Engineering

- We employed the balancing procedures listed below.

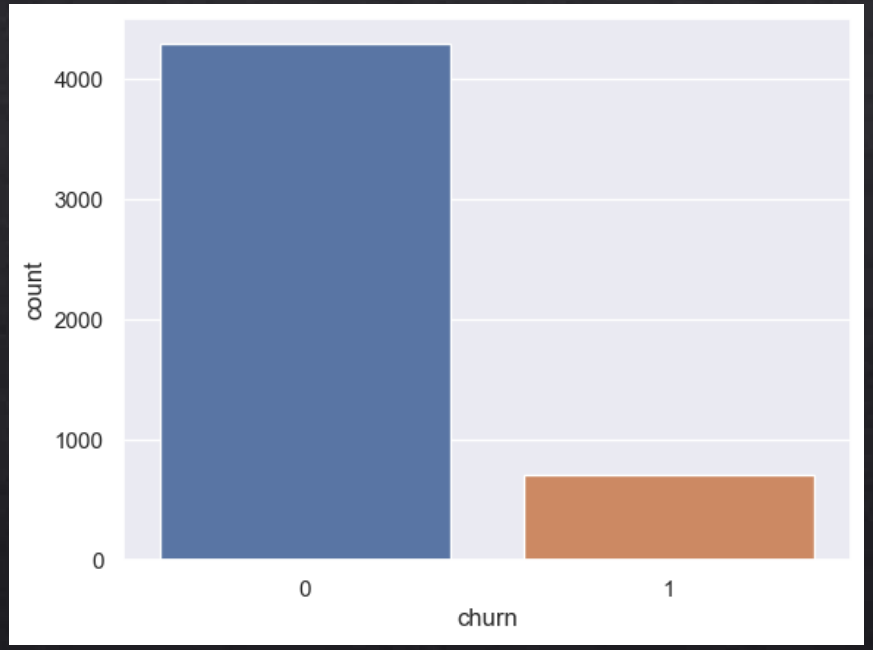
- Using smote for oversampling.
- Under the heading of Sampling.
- Oversampling is a well-known sampling technique.

- We discovered that there are 4293 non-churn customers and 707 churn consumers.

- The data is extremely skewed.

- The data is balanced with a Random Over sample and smoote.

- As a result, the old data - (4293,22), (707,1)
- New information - (4293,22) and (4293,1)



Whether the data was balanced or not

Choosing the ideal model

Sr. No	Model	Accuracy (%)	Remarks
1	Logistic Regression	76	
2	Decision Trees	82	
3	Random Forest	99	We choose RF as our final model
4	Adaboost	79	
5	Gaussian	75	

Deployement

- We used "Stream lit" to deploy the code after the Model Building and Evaluation phase.
- We chose the best model and used it in our deployment phase.