

# Voice Call Quality Analysis For Superior Customer Experience & Retention

Mentor : S Sourab Reddy | Batch: BLR - PGP-DSE June 24

ANJALI S

# Introduction To Problem Statement

## Problem Statement

The telecom industry struggles with customer retention due to poor voice call quality, leading to dissatisfaction and churn. The goal of this study is to develop predictive models that identify and address factors affecting call quality, aiming to improve service, enhance customer satisfaction, reduce churn, and optimize network resources

## Objective

1. Address poor voice call quality on the basis of customer rating.
2. Develop predictive models to identify and mitigate factors affecting call quality.

## Value Added

1. Improve customer satisfaction .
2. Optimize network resources.



# Navigating Through the Project Life Cycle



## Data Collection

[data.gov](https://data.gov)  
Data captured using TRAI MyCall App



## Data Exploration

Inspection & overview



## Data Cleaning

Identifying and correcting errors, handling missing values



## Variable Analysis

Understand its distribution, relationships with other variables



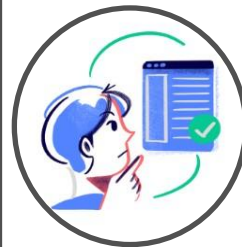
## Feature Engineering

Enhance model performance by creating meaningful features.



## Model Training

We selected Decision Tree as base models.



## Model Evaluation

Assessing a trained machine learning model's performance





# Data Collection

Data Source : Feedback from TRAI MyCall App ([data.gov](https://data.gov))

Dataset Link : [July 2018](#), [August 2018](#), [September 2018](#)

We consolidated datasets obtained from all three sources and merged them to create the final dataset for our project. Below are the finalized details after combining and refining these datasets.

Shape of dataset :

**No.of Rows :170783**

**No.of Columns : 9**

Feature Name :

Categorical Columns : 'Operator', 'In Out Travelling', 'Network Type', 'Call Drop Category', 'State Name', 'Month'

Numerical Columns : 'Rating', 'Latitude', 'Longitude'.

## The Story Behind Data Collection: My Call App

The app gathers crowdsourced data on voice quality and submits it to TRAI, contributing to the generation of a QoS report based on user feedback.



# Data Cleaning

Operator	0	Operator	0	Operator	0	Operator	0
In Out Travelling	0	In Out Travelling	0	In Out Travelling	0	In Out Travelling	0
Network Type	0	Network Type	0	Network Type	0	Network Type	0
Rating	0	Rating	0	Rating	0	Rating	0
Call Drop Category	0	Call Drop Category	0	Call Drop Category	0	Call Drop Category	0
Latitude	0	Latitude	0	Latitude	32677	Latitude	0
Longitude	0	Longitude	0	Longitude	32677	Longitude	0
State Name	17600	State Name	32678	State Name	34	State Name	0

Step 1: Checking for Null Values

Step2: We replaced "N/A-1" with null in the 'State Name' column to address non-standard missing values.

Step 3: We use logical imputation to resolve null values during data cleaning.

Step 4: We logically imputed missing values in the 'State Name' column, removed the remaining 34 nulls, and handled non-standard missing values in 'Latitude' and 'Longitude' with logical imputation.

## Inconsistency Check

Value counts of each variable will be analyzed to identify and fix inconsistencies.

## Duplicate Records

Duplicates are retained as they may represent unique user experiences, ensuring data granularity and accurate analysis.



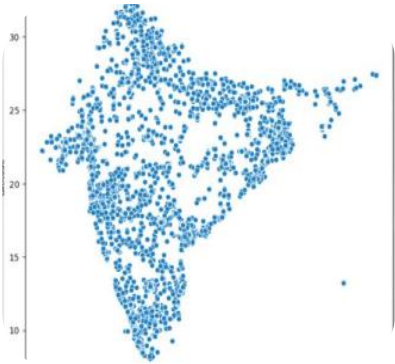
# Variable Analysis

## Descriptive Analysis - Numeric Columns

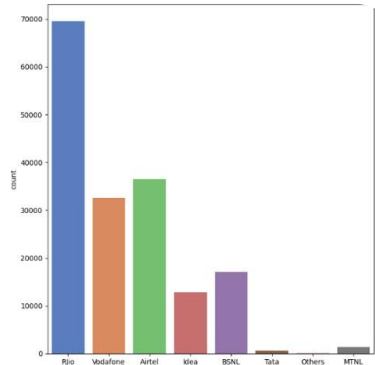
	count	mean	std	min	25%	50%	75%	max
Rating	170717.0	3.457506	1.522354	1.000000	2.000000	4.000000	5.000000	5.000000
Latitude	170717.0	21.010879	5.410369	8.084712	18.447802	20.310055	25.795465	32.987526
Longitude	170717.0	77.768154	4.791633	68.965040	73.794660	77.157185	79.096350	95.629638

## Descriptive Analysis - Categorical Columns

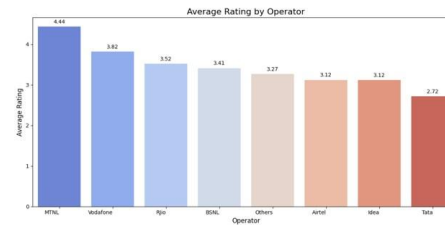
	Operator	In Out Travelling	Network Type	Call Drop Category	State Name	Month	Region
count	170717	170717	170717	170717	170717	170717	170717
unique	8	3	4	3	29	3	6
top	RJio	Indoor	4G	Satisfactory	Maharashtra	July	West
freq	69563	116820	97401	113459	45526	63646	70050



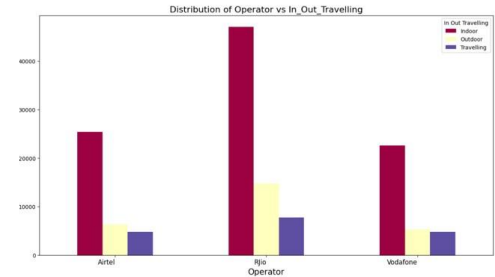
**Numeric Plot Univariate**  
Latitude & Longitude



**Categoric Plot Univariate**  
Operator



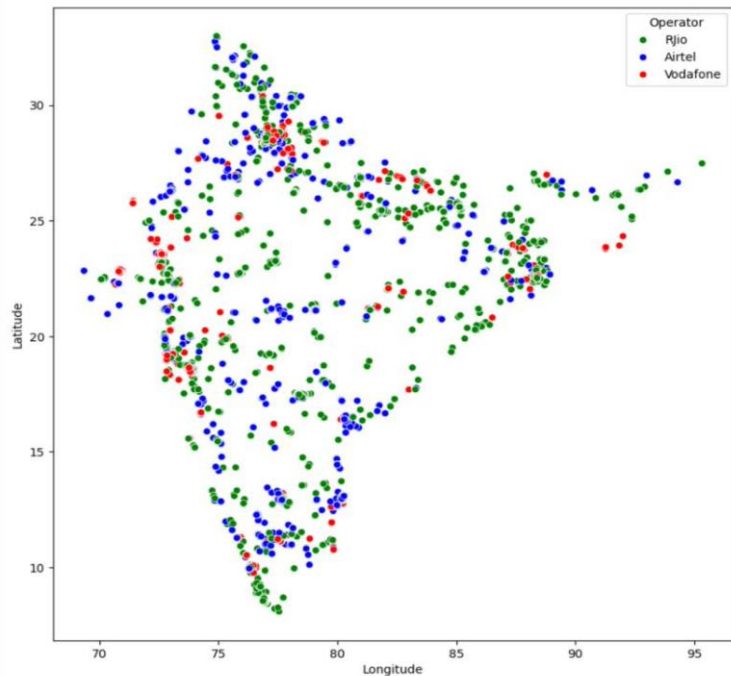
**Bi-Variate Analysis -1**  
Operator vs Avg.Rating



**Bi-Variate Analysis-2**  
Operator vs In Out Travelling



# Variable Analysis

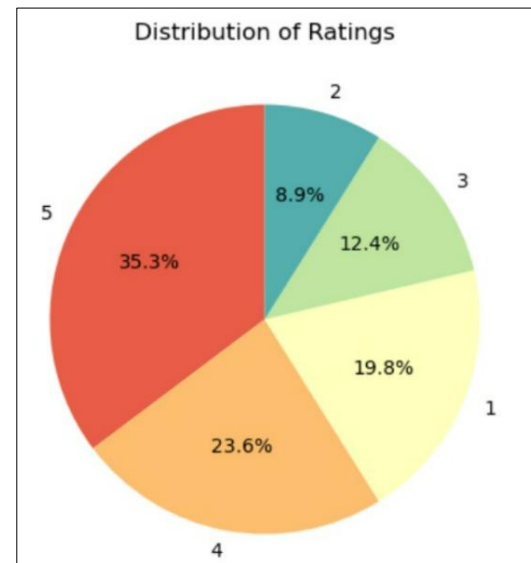


**Multi-Variate Analysis**

Leading Operators & Geo-Coordinates with Low Rating

	Features	VIF
0	Operator	3.321250
1	In Out Travelling	1.391362
2	Network Type	6.642759
3	Call Drop Category	7.129689
4	State Name	8.346654
5	Month	2.923588
6	Region	6.349714
7	Latitude	6.123720
8	Longitude	6.812539

**Multicollinearity with VIF**



**Target Class Imbalance**



# Feature Engineering

**Creating a new feature :** Added a Region column to group states into North, South, East, West, and Northeast regions for better trend analysis. Encoded categorical variables for compatibility with machine learning models.

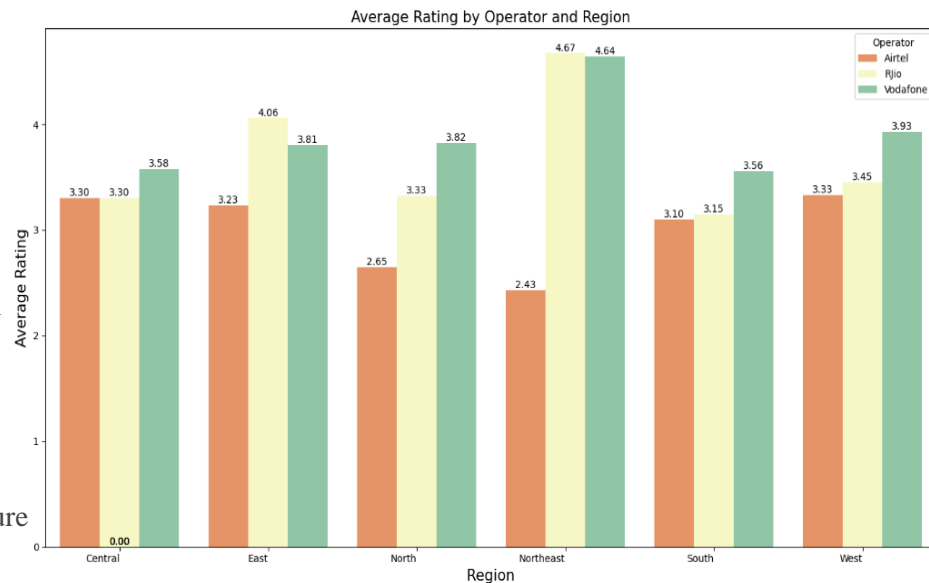
**Categorical Variables:** Convert variables like Operator, In/Out/Traveling, Network Type, Call Drop Category, State Name, Region, and Month to numerical format using Label Encoding.

**Outliers :** Outliers in the longitude feature represent valid geographic locations and are not treated, as they provide valuable insights into regional network behaviors.

**Feature Selection:** In feature selection, no columns are removed as multicollinearity is not present in the dataset.

**Scaling :** applied to numerical features like Latitude and Longitude to ensure better performance of models like k-NN model

**Dimensionality Reduction:** Dimensionality reduction is not necessary because the dataset has a small number of features, with no significant redundancy or weak correlations. Reducing dimensions could lead to the loss of valuable information.







# Model Training

- Target Variable:** Call quality rating (1 to 5).
- Task:** Multi-class classification.
- Goal:** Predict call quality using features like operator, network type, and location.

Model	Train Accuracy	Test Accuracy	Accuracy Diff (%)	Inference
Decision Tree Classifier	96.33%	83.76%	12.57%	High over-fitting, poor generalization.
k-NN Classifier	87.24%	82.79%	4.45%	Moderate gap, better generalization than some models.
Naive Bayes Classifier	56.63%	56.62%	0.01%	Almost no over-fitting; poor performance overall.
Random Forest Classifier	96.33%	85.41%	10.92%	High over-fitting,
Gradient Boosting Classifier	73.98%	73.36%	0.62%	Small gap, but low accuracy
AdaBoost Classifier	96.33%	83.99%	12.34%	High over-fitting, similar to Decision Tree.
XGBoost Classifier	84.27%	82.20%	2.07%	Small gap, better performance compared to others.



# Model Evaluation

Model	Train Accuracy	Test Accuracy	Accuracy Gap (%)	Overfitting / Generalization	Recommended Model
Random Forest Classifier	96.33 %	85.41%	10.92%	Gap difference = -2.66%	High overfitting; performance drops after tuning Not recommended due to overfitting.
Random Forest Classifier – HT & CW	93.64%	85.38%	8.26%		
Gradient Boosting Classifier	73.98%	73.36%	0.62%	Gap difference = +10.36%	Significant overfitting, but some improvement in generalization when compared to Random Forest. Not recommended due to overfitting.
Gradient Boosting Classifier -HT & CW	95.63%	84.65%	10.98%		
XGBoost Classifier	84.27%	82.2%	2.07%	Gap difference = +3.92%	Balanced overfitting, better performance and generalization. Recommended for balanced performance
XGBoost Classifier -HT & CW	88.62%	83.63%	5.99%		



## Business Insights

**Impact on Retention and Profits:** Poor voice call quality drives customer churn. A 5% improvement in retention can increase profits by 25–95% (Gartner).

**Feedback Utilization:** Platforms like the MyCall app provide actionable feedback, enabling telecom companies to address pain points and improve service quality.

**Boost Customer Retention:** Enhanced call quality directly reduces churn and fosters customer loyalty.

**Targeted Investments:** Focus infrastructure upgrades in regions with poor call quality ratings for maximum impact.

**Customer-Centric Approach:** Align network improvements with actual customer feedback to build trust and satisfaction.

**Efficient Resource Allocation:** Optimize budget by directing resources to underperforming areas.

**Proactive Monitoring:** Implement real-time network monitoring to identify and resolve issues before they affect users.

### Operator-Specific Insights:

**RJio:** Market leader with strong 4G adoption but faces isolated performance issues in Kerala, Gujarat, and Odisha.

**Vodafone:** Excels in the North East with the highest 4G user ratings.

**Airtel:** Shows variability in service quality, particularly needing improvements in the North and Northeast regions.

**THANK YOU**

