Call quality analysis is crucial for both internet service providers and customers. ISPs could analyze the market and understand their competitors' merits and flaws to enhance their own customer base. Whereas for customers, this would help them to choose the best option available. So it is a win-win situation for both the parties. The analysis can be divided into three major parts, pre-processing, understanding the dataset through visualizations, and data analysis algorithms. Each aspect is equally significant and none of them can work alone. Pre Processing includes transforming raw data into a useful and efficient format. Data visualizations include various plots and graphs to understand the dataset better. The algorithm techniques enable in applying various models on the dataset and get concrete results. In the paper, K means clustering is used to predict the best network operator for a particular location. This would directly help the customers to choose the best network in their city. For a particular city, the latitude and longitude are found, which are used as parameters for clustering. Multiple Regression is used to find the relationship between upload and download data speed, signal strength, and location. This analysis is done for the complete dataset as a whole followed by analysis for each operator. Similar models are used to find relationships between data speed and signal strength. Logistic Regression is used to find the relationship between the call drop category and other features. Chi-square tests will be used to find if literacy rate and network technology, state rainfall and rating, population, and rating are independent of each other or not.

The dataset consists of details of numerous callers located in different parts of India. For each caller, different attributes like the network service provider, the network technology (4G or 3G) that the caller was using, in-out mode he was in that is whether he was traveling or not, what rating he gave to the call, etc were collected. The dataset is obtained by combining a few datasets of September 2018 from data.gov.in. The dataset is a good mix of qualitative and quantitative features. The features are as follows: Service Provider(RJio, Vodafone, BSNL, Airtel, Idea), In-out traveling communication(Indoor, Outdoor, Travelling), Network technology (3G, 4G), Rating (1-5), Call drop status(satisfactory, poor voice quality, call dropped), Latitude, Longitude, State(Tamil Nadu, Punjab, Haryana, etc), average upload and download data speed, average upload and download signal strength, population, rainfall, literacy rate and area of the state.