Table of Contents

[Link to the dataset 4](#_Toc98546656)

[Approach 4](#_Toc98546660)

[Exploratory Analysis Rmd and HTML 4](#_Toc98546661)

[Modelling Rmd and HTML 4](#_Toc98546662)

# Link to the dataset

The Broadband Customer Base Churn dataset is a multivariate dataset that represents monthly bill of each customer with their id, contract tenure, complaints, bandwidth, churned status, and other features. The dataset consist of 510,126 observations and 19 attributes. The dataset can be found here: [Broadband Customers Base Churn Analysis | Kaggle](https://www.kaggle.com/jitendragoyal/broadband-customers-base-churn-analysis)

# Approach

The proposed project offers customer churn prediction using data mining techniques. I will apply logistic regression and other classification models (decision tree, random forest, and naive bayes) in the R studio tool to predict the customer churn, and to come up with the essential attribute in dataset that most affects the customer decision. Moreover, I am going to perform feature engineering technique to select the most relevant variables from the dataset. What service attributes and customer characteristics have affected the churn probability of high value (revenue) customers in a particular month? How can we develop a prediction model that predicts the valuable customer (generated revenue) churn?

# Exploratory Analysis Rmd and HTML

# Modelling Rmd and HTML