Project Objective: Explore the churn (the term “churn” represents someone ending their service) of an internet and telephone provider based on variety of customer features.

The data source in this project has the column named “Churn” as the label, which is binary (yes or no) indicating whether a certain customer churned or not. The features including the gender, monthly charges, total charges, tenure (the amount of months a customer was or has been on a customer) and contract type, etc., of a customer. Full data source is in the file with title “Telco-Customer-Churn.csv”.

The project is separated into two parts: exploratory data analysis and prediction using different tree-based model. In the first part, different aspects of the data are explored by visualization to get a general idea of what are the key features to the class label. In the second part, machine learning models including simple decision tree, random forest and AdaBoost models are explored to predict the class label.