

# Project Proposal

Dataset: [Telco customer churn: IBM dataset \(kaggle.com\)](https://www.kaggle.com/airoot/telco-customer-churn)

The problem statement relating to the data is “What are the characteristics of customers that are more likely to indicate churning”

The context is that a fictional telecom company provides internet and phone services to 7043 residents in California. Each row identifies a customer with information about them, such as their location, gender, marital status, as well as the information about their plan with the telecom company. All of this collected information is relevant to the topic of the data; which is whether or not a specific customer churned. Churning is when a customer stops using a service or product, by either cancellation or failure to renew.

The criteria for success is if I can identify factors that are more likely to indicate likelihood of churning and create a model that could potentially predict a churn.

The constraints of the dataset to consider are that the data is only from California residents in the third quarter of the given year, and the overall likelihood of Californians to churn is unknown compared to the rest of the United States, or even the world.

The stakeholders are IBM, where the dataset was pulled from, and the fictional CEO of the company as well as people in the relevant departments of the company

The key data source is IBM, where the data is sourced from.