Capstone Proposal

Problem statement:

To predict the churn of users of a telecommunication company.

Context:

Hypothetically, a telecommunication, or any company that relies on users subscriptions/ return, would like to either reduce the churn rate of its users or identify sources of its churn rate. Given the metadata on its users, the team would develop insights and potentially implement changes on its product/service.

Constraint:

There is only one instance of user data. In reality, the data on each user would be a time series. Our analysis won't be able to capture

Scope of solution:

The scope of our solution only considers the data on our users when the churn of users can be influenced by factors such as competitors and the health of the industry as a whole.

Data Sources: Data is obtained from Kaggle.

Outline of Solution:

Data Cleaning: Check for missing values, inconsistency, and outliers
Exploration: Maybe create more columns for analysis. Explore distribution of the features.
Modeling: Logistic Regression, Random Forest, Naive Bayes, SVM
Deliveries: Project report and Slide deck