

# How will customers churn?

Zhou Zhou

## Purpose

The project focuses on applying the tools in Google Cloud Platform to analyze how the companies can make more customers churn and building the neural network model to predict whether the customers will churn or not.

The potential audiences are expected to be the decision makers in the companies.

## Dataset

Telco Customer Churn Dataset ([https://www.kaggle.com/blatchar/telco-customer-churn#WA\\_Fn-UseC\\_-Telco-Customer-Churn.csv](https://www.kaggle.com/blatchar/telco-customer-churn#WA_Fn-UseC_-Telco-Customer-Churn.csv)) is employed in the project. The dataset includes the information of customers' churn and 20 attributes which may influence their churn such as gender, total charges, monthly charge, payment method, contract, streaming TV, tech support, etc.

## Plan

Data will be loaded into Big Query. Dataproc will be used to create a Hadoop cluster and Pyspark is expected to be used. Also, Datalab will be used for more interactive exploration and DataStudio will be used to create dashboard to visualize our products to users.

Principle analysis is expected to be employed (PCA analysis has been done in "FinalKBS\_zhou.ipynb"), by which the dominant components will be obtained. Based on them, my plan is to take the dominant components as X values for the neural network and employ Tensorflow to train the model. Plus, random forest classification is expected to be employed to analyze from another angle.