Three Project Idea Drafts

Draft 1: What is the potential salary someone could earn in the U.S.? What factors make the difference between people? As a broad idea, I’m planning to explore and analyze the data set from United States Census Data on income and earnings of individuals and personal characteristics associated, such as age, sex, race, education attainment, etc. to develop useful models for predicting someone’s income with given personal characteristics.

Draft 2: Will a newly married couple divorce in 5 years? What are the significant divorce predictors? A dataset of 170 couple’s questionnaire responds on 54 related questions scaled from 1-5 will be used to complete a classification-type data analysis and prediction. For instance, the data set is obtained from University of California, Irvine (UCI) Machine Learning Repository database.

Draft 3: It is crucial for merchants to identify who can be converted into repeated buyers when providing big promotions. By targeting on these potential loyal customers, merchants can greatly reduce the promotion cost and enhance the return on investment. From Alibaba’s Aliyun Cloud database, data sets of merchants and their corresponding new buyers acquired during the promotion on the "Double 11" day are provided and used for develop the prediction model.