R Code Instruction

We built OLS model in R. Within the 'Code_R.Rmd' file, We imported the training set called "train_ols.csv" and test set called "test_ols.csv". Then we used AIC to select best model. After selected the model, we calculated the r-squared and mean squared error for OLS model. Eventually, we selected the top 10 important variables in this model.

Python Code Instructions

We applied LASSO regression, Decision Tree regression and Random Forest regression by Python. Within the 'Code_Python.py' file, we imported the 'X_train_dummy.csv', 'y_train_dummy.csv', 'X_test_dummy.csv' and 'y_test_dummy.csv' and ran these three models with cross validation one by one. Then, we extracted the r-squared and mean squared error of each model. Finally, we extracted the most important 10 features within each model.