Instructions

Python:

1. Import modules such as pandas, matplotlib, numpy, seaborn ,os, sklearn etc.
2. Set the current directory with **os.chdir()** to read and save the csv files.
3. Install the **geopy** library with pip command (pip install geopy) in anaconda prompt.
4. Upgrade **seaborn** to latest version **0.9** (pip3 install seaborn==0.9.0) to see countplots.
5. download xgboost whl file from here (make sure to match your python version and system architecture, e.g. "xgboost-0.6-cp35-cp35m-win_amd64.whl" for python 3.5 on 64-bit machine)
6. open command prompt
7. cd to your Downloads folder (or wherever you saved the whl file)
8. pip install xgboost-0.6-cp35-cp35m-win_amd64.whl (or whatever your whl file is named).
9. The fare_amount predicted on the **test.csv** is saved to separate **result.csv** file in the current directory.

R:

1. Remove all data present in global environment → rm(list=ls(all=T))
2. For installing packages in R Studio (go to tools→Install Packages)
3. Install geosphere package
4. Install dplyr package
5. Install ggplot2 package
6. Install randomForest package
7. Install e1071 package (SVM)