Programming for Data Science DS-GA 1007

Prof. Greg Watson

DS-GA-1007

Fall 2015
December 15, 2015

<u>**Project's Team:**</u>
- Rafael Garcia Cano da Costa (rgc292)
- Kristen Kwan (kk3175)

<u>Project User Guide</u>

<u>Index:</u>

Introduction:

      This documentation intends to give a general idea to a final user why this software was conceived and how it can be implemented.

Purpose:

      This program was developed for offering visual analyses for use in understanding data collected from housing complaints by the NYC Department of Housing Preservation and Development.

Software Dependencies:

      This program assumes that the final user has Python 2.7 installed as well as its respective libraries such as Numpy, Pandas, Matplotlib, Scipy, Pylab, Pyplot, Pydot, and Scikit-learn.

Required Files:

      In order to run this program, the following file is required:
- complaint_problems.csv

If this file is not available, an updated file can be obtained from this link:
- https://data.cityofnewyork.us/Housing-Development/Complaint-Problems/a2nx-4u46

However, it is important to download the file as a .csv file, and save the file under the "term_project" directory having the name "complaint_problems.csv". The link for downloading is available under the link called "Export" where another link under "Download as" called "CSV" is available. The respective link is as follow:
- https://data.cityofnewyork.us/api/views/a2nx-4u46/rows.csv?accessType=DOWNLOAD

The directory tree can be seen below:
- project_1007/term_project

Having all aforementioned set, the program is ready for running.

User Interaction:

      The user can run the program on commandline. In general most Operational Systems can run commandline under the name of "Terminal". After opening the Terminal, the code "python project.py" needs to be typed on the commandline. Once this is accomplished, the program will request three inputs from the user.

      The first input will be the identification number(ID) of the complaint type the user wants to analyze. A table containing all available IDs and their corresponding complaint types will be displayed guiding the user in choosing which one he wants. The user needs to type a number. For example, if among the column of numbers under "Complaint Type" the number "55" is an option and the user wants to choose this number, he needs to type: 55.

      For the second and third inputs, the user will be asked to enter the start date and then the end date for his analysis. The available date range of the housing data will be provided to let the user know what dates are available for his analysis. The user will be asked to enter their desired dates in the format of YYYYMMDD. Therefore, if the user wants to enter the date January 29th, 2014, he needs to type: 20140129.