**Coursera Capstone Project**

## The Battle of Neighborhoods

**Data**

**Composed by: Ekaterina Deriabina**

# 2. Data

In this project the following data were used:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data | Name | Type | Condition | Content | Source |
| Zip Codes | List of zip codes | Wiki | N/A | List of zip codes of Seattle |  |
| Geocoding | Geocoding API | JSON | N/A | Google Maps API calls | Google Maps |
| Forsquare | Foursquare location data | JSON | N/A | Foursquare API calls | <https://foursquare.com/> |

**3. Methodology**

In my project the API Forsquare were used as data source of all the required information about the schools, shops, public places and their ratings. To get the information about the location of the required objects (latitude and longtitute) the zip codes of the required city were used.

To avoid the problems with the requests limitation in the Foursquare API the radius was selected to be 1000, and the places number is to be 500.

To make the choice of the housing easier, the five neighborhoods were compared with the following parameter: the schools ratings and the pricing of housing. Moreover the amount of the public places to visit were compared on the basis of the most common venues.

The comparation of the communities were performed with the using of API Foursquare, whereas the data was worked through with such python’s libraries as pandas, numpy, scilearn kit. The detailed description of the used libraries are given in the following table:

|  |  |
| --- | --- |
| Pandas | Library for Data Analysis |
| NumPy | Library to handle data in a vectorized manner |
| JSON | Library to handle JSON files |
| Geopy | To retrieve Location Data |
| Requests | Library to handle http requests |
| Matplotlib | Python Plotting Module |
| Sklearn | Python machine learning Library |
| Folium | Map rendering Library |