Coursera Capstone Project

The Battle of Neighbourhoods -Report (Week 1)

2. Data Section:

Description of the data and its sources that will be used to solve the problem.

2.1 Description of the Data:

- Use Foursquare and geopy data to map top 10 venues for all Manhattan neighbourhoods and clustered in groups (as per Course LAB)
- Use foursquare and geopy data to map the location of subway metro stations, separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately
- Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations.
- create a map that depicts, for instance, the average rental price per square ft, around a radius of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the relative price per subway area.
- Addresses from rental locations will be converted to geodata (lat, long) using Geopy-distance and Nominate.
- Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc.

2.2 How data will be used to solve the following problem:

- the cost of rent (per square ft) around a mile radius from each subway metro station.
- what is the area of Manhattan with best rental pricing that meets criteria established?
- distance from workplace (Park Ave and 53 rd St) and the tentative future home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighbourhoods and around metro stations?
- Are there trade-offs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.