The Battle of Neighborhoods – Part I

Applied Data Science Capstone – Data section

Data

To work with this project, following data sources will be required:

- New York City data that contain all neighborhoods and boroughs with their localization. That will be primary source in this project, and based on that other data will be gathered.
  Source: https://cocl.us/new_york_dataset

  How will look like data?
  In raw form:


```
{'type': 'FeatureCollection',
 'totalFeatures': 306,
 'features': [{'type': 'Feature',
   'id': 'nyu_2451_34572.1',
   'geometry': {'type': 'Point',
    'coordinates': [-73.84720052054902, 40.89470517661]},
   'geometry_name': 'geom',
   'properties': {'name': 'Wakefield',
    'stacked': 1,
    'annoline1': 'Wakefield',
    'annoline2': None,
    'annoline3': None,
    'annoangle': 0.0,
    'borough': 'Bronx',
    'bbox': [-73.84720052054902,
     40.89470517661,
     -73.84720052054902,
     40.89470517661]}},
```


  What we are interested there is: coordinates, name and borough. After cleaning it will look like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

- To visualize boundaries of previously mentioned boroughs – we will require geojson file that will contain information about their boundaries (to create choropleth map using folium library).

Source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm for boroughs
Source: https://www.simboli.eu/download-of-geojson-files/ for neighbors
Data for this will look like:


```
{
  "type": "FeatureCollection",
  "generator": "overpass-ide",
  "copyright": "The data included in this document is from www.openstreetmap.org. The data is made available under ODbL.",
  "timestamp": "2019-01-27T21:32:02Z",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "@id": "relation/8398069",
        "admin_level": "10",
        "boundary": "administrative",
        "name": "Marble Hill",
        "place": "neighbourhood",
        "type": "boundary",
        "wikipedia": "en:Marble_Hill,_Manhattan"
      },
      "geometry": {
        "type": "Polygon",
        "coordinates": [
          [
            [
              -73.918383,
              40.8758485
            ],
            [
              -73.9178781,
              40.8757351
            ],
            [
              -73.9177827,
              40.8756544
            ],
```


- To gather information about restaurants – Foursquare API will be required. Most interesting endpoint from our perspective: **venues/search** and with range set to 1.5km, limited to 100, with category id equal to 4bf58dd8d48988d110941735 (Italian restaurants category). As a second endpoint, I will use **venues/venue_id** (premium call) to gather elements such as: price, rating, stats, popular, like, dislike, phrases – all of them can be useful in further analyze and grouping.
  Source: https://foursquare.com/