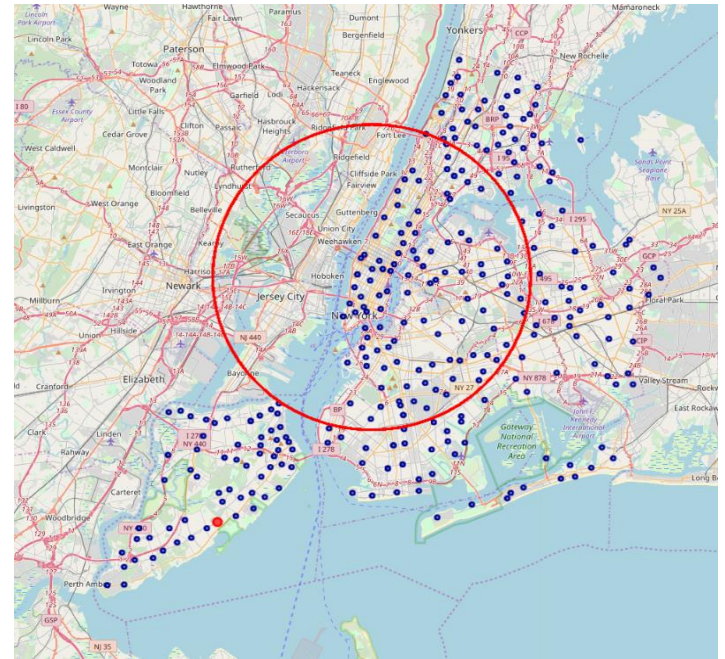


# THE BATTLE OF THE NEIGHBORHOODS

IBM DATA SCIENCE CAPSTONE PROJECT

# PROBLEM BACKGROUND

- **New York is the largest and one of the most populated cities in the United States.**
- **Every day, many individuals make decisions on moving to new York City or relocating to new neighborhood**
- **The analytic tools that would help to make decision would be useful for both individuals and rental companies**



*Map of NYC neighborhoods*

# DATA ACQUISITION AND CLEANING

1. The list of 306 NYC neighborhoods and coordinates used in the analysis was obtained from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
2. the data on the profiles of neighborhoods from the government site on [NYC Community District Profiles](#). It includes 59 observations corresponding to the number of community districts and nearly 200 variables - 8 of them were used in the analysis
3. Foursquare API is used to obtain the venue data for the neighborhoods in New York

All datasets were merged together and duplicates were removed.



# DATA - VARIABLES USED IN THE ANALYSIS

1. qcd\_full\_title - name of Community District
2. pct\_lot\_area\_open\_space - parkland and open space area (as % of total CD area)
3. crime\_per\_1000 - number of crimes per 1000 people in CD
4. pct\_hh\_rent\_burd' - Percentage of households that spend 35% or more of their income on rent
5. mean\_commute - Mean commute time to work for residents
6. unemployment\_cd - unemployment in CD
7. neighborhoods\_x - neighborhoods included in the CD
8. pct\_clean\_strts - percentage of streets rated acceptably clean in the CD
9. poverty\_rate - poverty rate in CD 10 .pct\_lot\_area\_\_\_industrial\_manufacturing - percentage of lot area in CD with industrial/manufacturing buildings

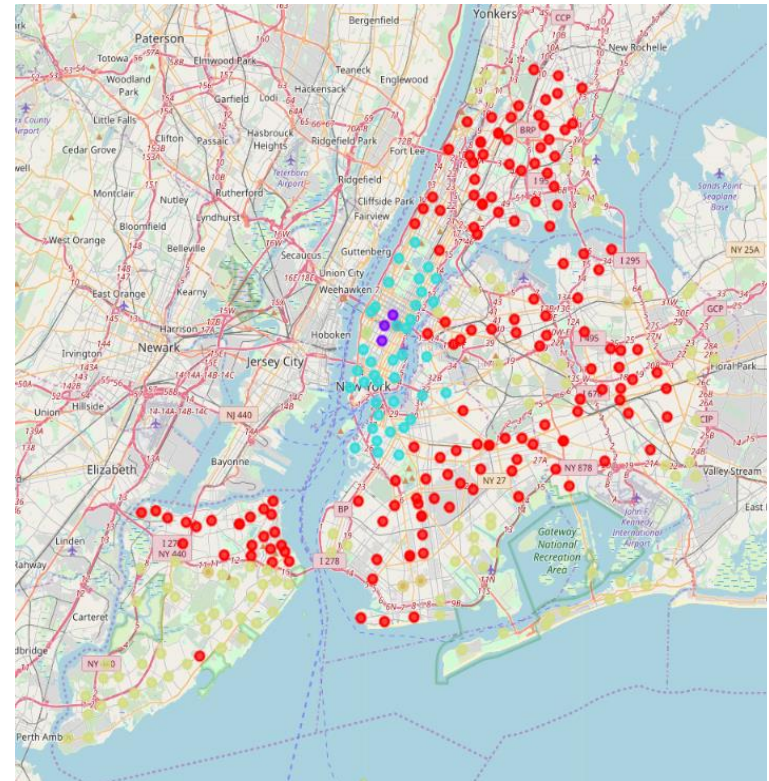
# RESULTS - EXPLANATORY DATA ANALYSIS

- Poverty is strongly and positively correlated with rent burden and unemployment and negatively correlated with street cleanliness
- Crime is higher in neighborhoods with high and low poverty rate
- Rent burden is positively correlated with poverty, unemployment and commuting time.
- The relations between size industrial and open areas with other variables are not clear


	Borough	Open_area	Crime	Street_clean	Rent_Burden	Commute_time	Unemployment	Poverty	Industrial_area	Latitude	Longitude
0	Bronx	0.124235	13.605862	95.301887	48.764151	43.830189	7.386792	24.007547	0.040005	40.850820	-73.873735
1	Brooklyn	0.136358	11.366159	93.746377	44.779710	42.710145	5.465217	20.420290	0.051886	40.646101	-73.952079
2	Manhattan	0.122914	19.313509	96.156757	36.575676	30.459459	3.975676	12.951351	0.012707	40.762799	-73.975752
3	Queens	0.167045	9.162469	96.580247	46.992593	43.864198	5.024691	19.344444	0.044659	40.707201	-73.826134
4	Staten Island	0.213780	6.764254	98.239683	45.911111	42.809524	3.676190	16.384127	0.027841	40.589165	-74.136091

# RESULTS - CLUSTERING

- **Cluster 0 (red)** - low crime rate and lots of open areas. However, rent burden, poverty and commuting time are relatively high.
- **Cluster 1 (purple)** - has the lowest poverty rate, lowest share of open area and exceptionally high crime rate.
- **Cluster 2 (blue)** - highest poverty rate and unemployment, high rent burden.
- **Cluster 3 (olive green)** - second highest crime rate. Commuting time, unemployment and rent burden are low. It is the most industrialized area.



# RECOMMENDATION FOR CLIENT

1. We recommend one of the neighborhoods in the same cluster as the client's current neighborhood (**153 neighborhoods were shortlisted**)
  2. **33 out of 153 neighborhoods** in Cluster 0 meet requirement of commute time
  3. **8 neighborhoods** meet requirement of proximity to the special school
  4. Out of these 8, **East Elmhurst** was the neighborhood with the most similar venues. Therefore, we recommended it.
- 

# CONCLUSION AND DISCUSSION

- **Clustering neighborhoods can provide valuable insights and help in generating recommendations and advertising**
- **However, the number of indicators used was limited**
- **Further analysis can include the absolute rent, maintenance and parcel prices for different neighborhoods**
- **There are serious data concerns: accessibility and foursquare API personal account limitations**

