

# Coursera Capstone : Battle of the neighborhoods



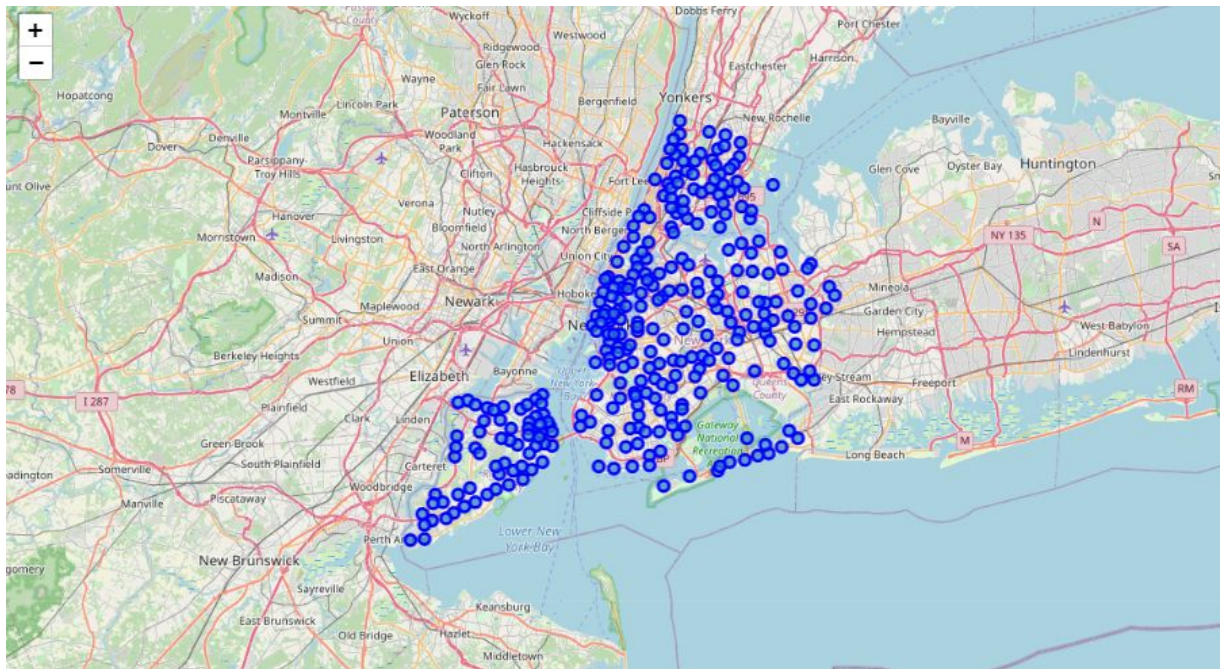
# Problem Statement

1. Finding a good place to start a restaurant in NYC
2. The Real Estate prices are too steep so there is no way to turn back once a location is fixed
3. Being one of the most developed and populated areas in the world, the competition is high

# Data

## Sources -

1. Week 3 data on  
New York City
2. Foursquare API



# Data Operations

1. Clean data from the source in W3
2. Find count/frequency of number of venues and restaurants in each neighborhood using one hot encoding
3. Create 5 clusters using K Means clustering



# Clusters

The clusters are divided based on the number of donut shops into 5 clusters.

From 0 shops to max ie. 4 shops

```
In [38]: df3 = nyc_merged.loc[nyc_merged['Cluster Labels'] == 2]  
df3
```

Out[38]:

	Neighborhood	Donut Shop	Cluster Labels	Borough	Latitude	Longitude
44	East Harlem	0.142857	2	Manhattan	40.792249	-73.944182
83	Huguenot	0.142857	2	Staten Island	40.531912	-74.191741
96	Long Island City	0.105263	2	Queens	40.750217	-73.939202

```
In [39]: df4 = nyc_merged.loc[nyc_merged['Cluster Labels'] == 3]  
df4
```

Out[39]:

	Neighborhood	Donut Shop	Cluster Labels	Borough	Latitude	Longitude
0	Allerton	0.333333	3	Bronx	40.865788	-73.859319

```
In [42]: df5 = nyc_merged.loc[nyc_merged['Cluster Labels'] == 4]  
df5
```

Out[42]:

	Neighborhood	Donut Shop	Cluster Labels	Borough	Latitude	Longitude
41	Downtown	0.033333	4	Brooklyn	40.690844	-73.983463
84	Hunters Point	0.041667	4	Queens	40.743414	-73.953868
85	Inwood	0.058824	4	Manhattan	40.867684	-73.921210



# Conclusion

Looking at the competition increasing from cluster 1 having no competition to cluster 5 containing neighborhoods with highest competition.

Thus to conclude, opening a donut shop in neighborhoods from clusters 1 and 2 seems the best solution.

Thank You