Capstone Project - The Battle of the Neighbourhoods

By AGO MAGAI ATOSHI

INTRODUCION

- This project work seeks to provide prospective investors with an analysis and recommendation of the most promising neighbourhood to start a restaurant business in New york City.

- When an individual wants to start a new restaurant business, he or she will try to locate a neighbourhood that is conducive enough for the business to prosper. This will guarantee profit to the entrepreneur.

- To perform the analysis, data about New york will be collected and using regression, similar neighbourhoods with high restaurant patronage will be selected. The neighbourhood with the highest population among the selected neighbourhoods will be deemed to be the most promising.

DATA

- The data to be used for this project will be from four(4) sources;

- First, the basic New York data from the link 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json' that contains neighbourhood coordinates and other data.

- Second, Neighbourhood venue data using foursquare API to obtain different venues including restaurants, gymns, yoga centers, meuseums, etc. This API will also be used to obtain data about venue rankings which will be used to rank the venues accourding to their level of importance.

- Thirdly, data about New York's Neighbourhood populations will be collected from the link 'https://data.cityofnewyork.us/api/views/swpk-hqdp/rows.csv?accessType=DOWNLOAD' which can be found on the websited 'https://catalog.data.gov/dataset/new-york-city-population-by-neighborhood-tabulation-areas-f2447'. A neighbourhood with an optimum population is good for business.

- Lastly, data about Grosss Domestic Product (GDP) and Per-Capita Income oe the Boroughs in New York will also be collected from the website 'https://en.wikipedia.org/wiki/Demographics_of_New_York_City'. Good GDP and per-capita income are good indictors of the economic stregth and purchasing power of the citezens of the respective Boroughs and Neighbourhoods.

	Borough	PerCapita	Neighborhood	Population	nLatitude	nLongitude	Venue	vLatitude	vLongitude
0	Bronx	30100	Wakefield	31078	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898083	-73.850259
1	Bronx	30100	Wakefield	31078	40.894705	-73.847201	Lammy's Restaurant	40.890846	-73.848778
2	Bronx	30100	Wakefield	31078	40.894705	-73.847201	Cool Running Restaurant	40.898399	-73.848810
3	Bronx	30100	Wakefield	31078	40.894705	-73.847201	Bay restaurant	40.890850	-73.848860
4	Bronx	30100	Wakefield	31078	40.894705	-73.847201	Fort Negril Jamaican Restaurant	40.898087	-73.850266

METHODOLOGY

- As mensioned above the decision criteria for the best neighbourhood is based on the clustering of restaurants present in the neighborhood.

- The high presence or cluster will indicate favourable conditions for such businesses hence Kmeans clustering technique was used to cluster similar neigbourhoods.

```
[4]: array([[-0.78553518, -0.69620616,  2.09120657,  1.20743439,  2.13984596,
          1.16321424],
         [-0.78553518, -0.69620616,  2.09120657,  1.20743439,  2.04238151,
          1.18300522],
         [-0.78553518, -0.69620616,  2.09120657,  1.20743439,  2.14410185,
          1.1825746 ],
         ...,
         [-0.78303194, -0.97541947, -1.64003994, -1.92913223, -1.66762245,
          -1.97261593],
         [-0.78303194, -0.97541947, -1.64003994, -1.92913223, -1.67829076,
          -1.98440474],
         [-0.78303194, -0.97541947, -1.64003994, -1.92913223, -1.68291463,
          -1.96575014]])
```

```python
[5]: num_clusters = 25
k_means = KMeans(init="k-means++", n_clusters=num_clusters, n_init=12)
k_means.fit(cluster_ds)
labels = k_means.labels_
labels
```

```
[5]: array([ 3,  3,  3, ..., 10, 10, 10], dtype=int32)
```

RESULT

Labels	PerCapita	Population	nLatitude	nLongitude	vLatitude	vLongitude	Count
0	368500	46746	40.807501	-73.952379	40.807704	-73.952400	159
1	35800	22887	40.686752	-73.983049	40.686708	-73.982976	227
2	41400	53751	40.766997	-73.825502	40.765278	-73.825163	111
3	30100	31078	40.871226	-73.887739	40.870871	-73.888387	236
4	368500	46746	40.748952	-73.985027	40.749402	-73.984915	469
5	35800	22887	40.596874	-73.975677	40.597751	-73.975019	135
6	30500	27770	40.557860	-74.179898	40.558518	-74.180986	48
7	41400	53751	40.743123	-73.889017	40.743132	-73.888692	187
8	35800	22887	40.671488	-73.886769	40.671641	-73.885941	108
9	41400	53751	40.691473	-73.841430	40.692192	-73.841688	80
10	30500	27770	40.613662	-74.094443	40.613731	-74.094209	108
11	35800	22887	40.650238	-73.942488	40.650047	-73.942551	206
12	41400	53751	40.712945	-73.792883	40.713229	-73.791780	74

13	368500	46746	40.773469	-73.960827	40.773361	-73.960542	267
14	41400	53751	40.755030	-73.924468	40.754283	-73.923496	150
15	368500	46746	40.720615	-73.999173	40.720312	-73.999103	514
16	35800	22887	40.635833	-74.011215	40.635645	-74.011213	113
17	35800	22887	40.709703	-73.948361	40.709995	-73.948102	148
18	41400	53751	40.678529	-73.750661	40.678461	-73.749613	34
19	30100	31078	40.840527	-73.847795	40.840765	-73.848347	174
20	368500	46746	40.862097	-73.926422	40.861630	-73.924779	71
21	30100	31078	40.834685	-73.909002	40.834880	-73.908688	292
22	41400	53751	40.746431	-73.741481	40.745329	-73.740819	83
23	41400	53751	40.593965	-73.787220	40.593939	-73.785529	27
24	41400	53751	40.736268	-73.854508	40.735967	-73.855203	87

- The KMeans clustering algorithm was used for this captstone project with a k value of 25 bacause of the high number of restaurants returned by the foursquare API to provide a sufficient spread of the data.

- On a close observation cluster fifteen(15) had the highest value of restaurants of Five Hundred and Fourteen(514) restarants within its neighbourhoods which were clustered together. This cluster also had a high level of population of 46,740 persons and also a high per-capita income of 368,500(USD) on the average.

- Based on the aforemensioned cluster-15 was chosen for further analysis.

DISCUSSION

[37]:   <AxesSubplot:xlabel='Neighborhood'>
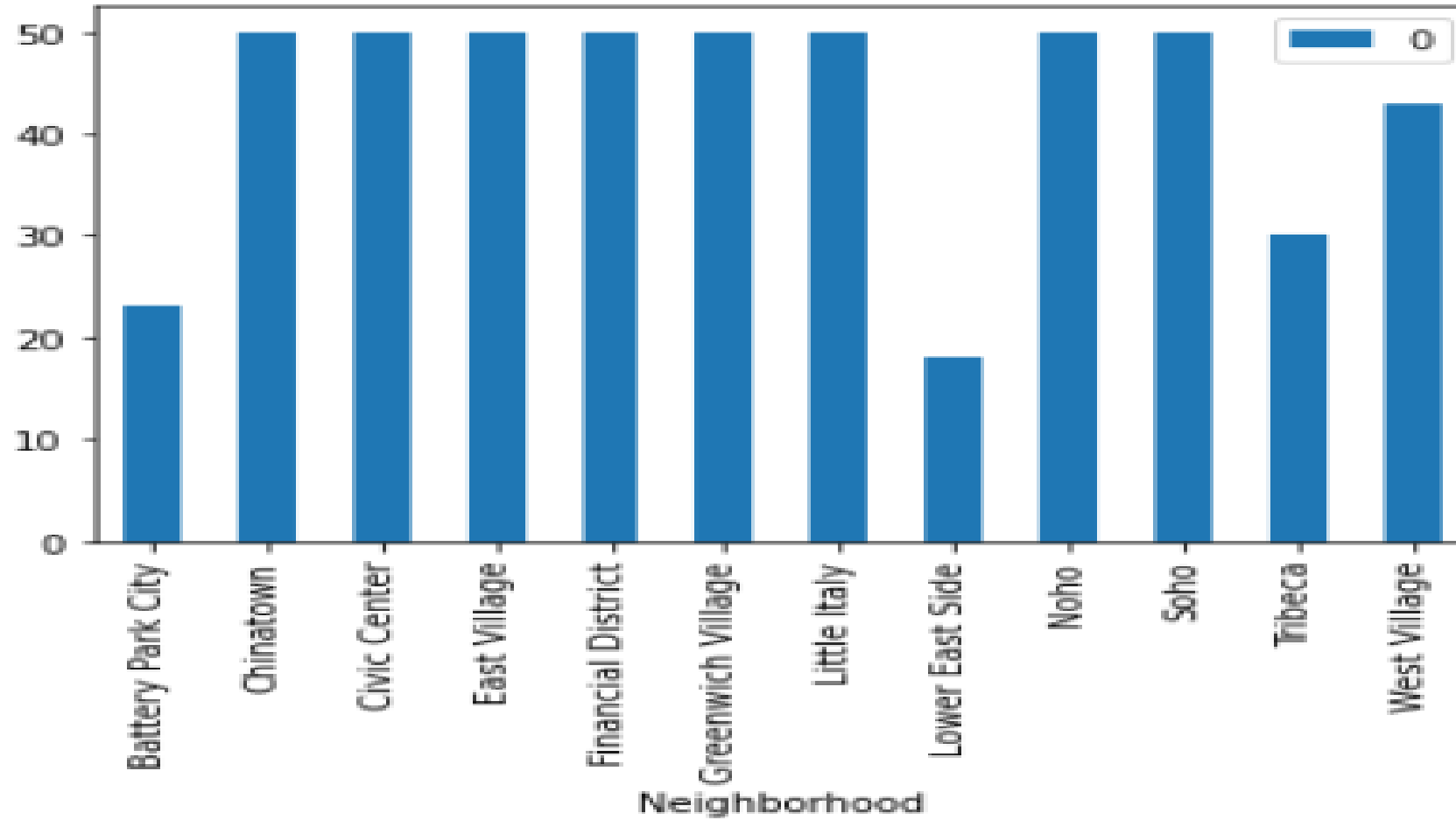
- The break down of cluster 15 reveals that Chinatown along with other Neighborhoods of Civic Center, East Village, Financial District, Greenwich Village, Little Italy, Noho and Soho all the the highest number of restaurants of 50 each in their respective neighborhoods.

- This implies the potential investor can invest in any of these towns because they all have high presence of restaurants, high level of population reflecting good market and also highl level of per-capita income refecting good purchasing power.

CONCLUSION

- Conclusively, the battle of the neighbourhoods have been a fierce one. However, Civic Center, East Village, Financial District, Greenwich Village, Little Italy, Noho and Soho proved to be the most prospective neighborhoods for citing a restaurant business.