

Capstone Project - The Battle of Neighborhoods - Best Indian Restaurants in NY

Introduction

Business problem

- Where are all the best Indian restaurants in New York?
- Which Boroughs have the best Indian restaurants?
- A local travel agency wants to take tourists to popular Boroughs that have the best Indian restaurants
- Are some of the best rated Indian restaurants in Manhattan or other Boroughs?
- I want to open a new Indian Restaurant and want to know which boroughs have the best rated ones – find a borough that may not have much competition

Key Goal of the business problem

- Identify the boroughs/neighborhoods in New York that have the best rated Indian restaurant (Rated 8.0 & higher)

Methodology

- Identify the data set – Foursquare
- Understand the data
- Collect the data
- Preprocess the data
- Prepare the data
- Modelling
- Analyze the output
- Prepare final summary

Approach

- Collect FourSquare Location data for NY – key elements includes name, neighborhood, borough, longitude, latitude etc
- Prepare neighborhood list, map to borough, identify restaurants rated 8.0 & higher,
- Review the data, output at each step, visualize
- Pre-process the data
- Cluster the data (K-Means, K=6)
- Review output
- Visualize the output

Detailed Approach/Data analysis

- Collect data
- Identify neighborhoods with restaurants rated 8.0 & higher

```
[47]: ny_neighborhood_stats=ny_neighborhood_stats[ny_neighborhood_stats['Average Rating']>=8.0]
ny_neighborhood_stats
```

```
[47]:
```

	Neighborhood	Average Rating
1	Bayside	8.2
5	Civic Center	8.2
7	Flatiron	8.4
8	Fort Greene	8.7
11	Greenwich Village	8.6
18	New Dorp	8.1
19	Noho	8.8
20	North Side	8.5
22	Prospect Lefferts Gardens	8.6
25	Sutton Place	8.4
26	Tribeca	9.1
29	West Village	8.5
30	Woodlawn	8.0

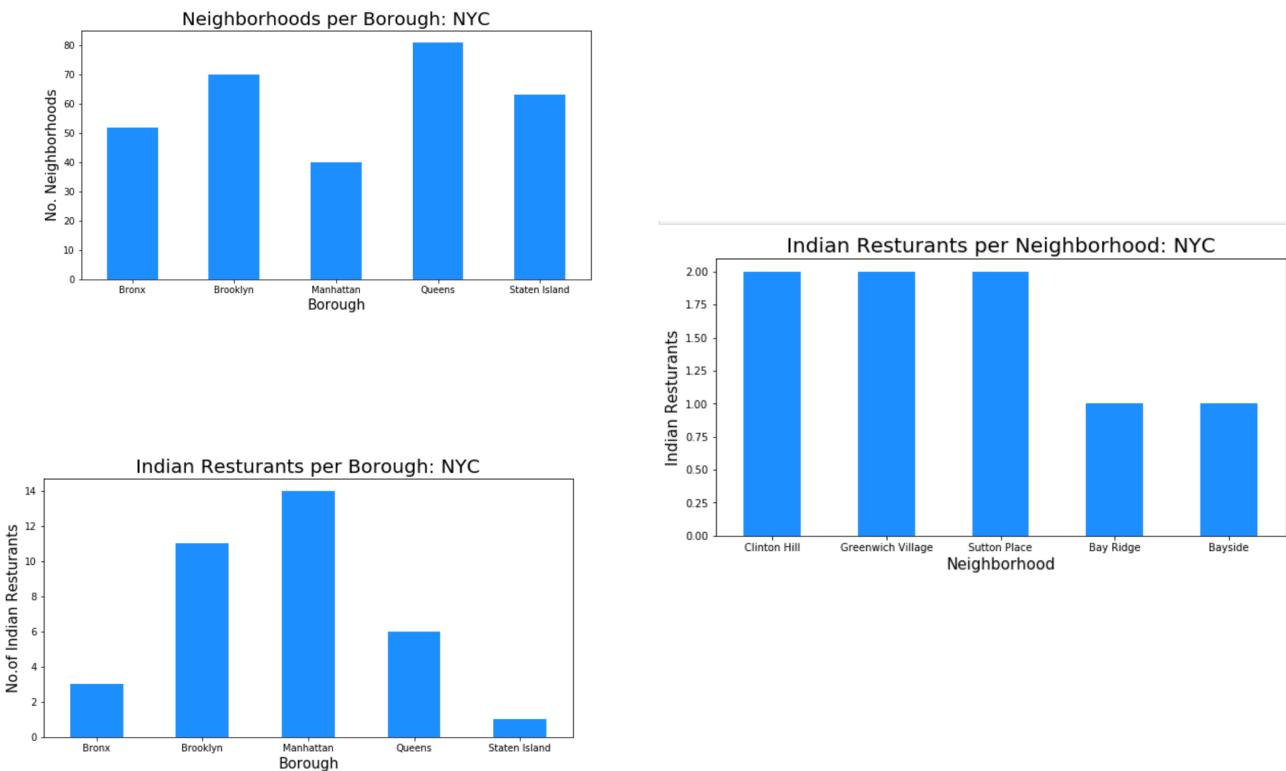
- Map data to NY borough', get other key data

```
[48]: ny_neighborhood_stats=pd.merge(ny_neighborhood_stats,ny_data, on='Neighborhood')
ny_neighborhood_stats=ny_neighborhood_stats[['Borough','Neighborhood','Latitude','Longitude','Average Rating']]
ny_neighborhood_stats
```

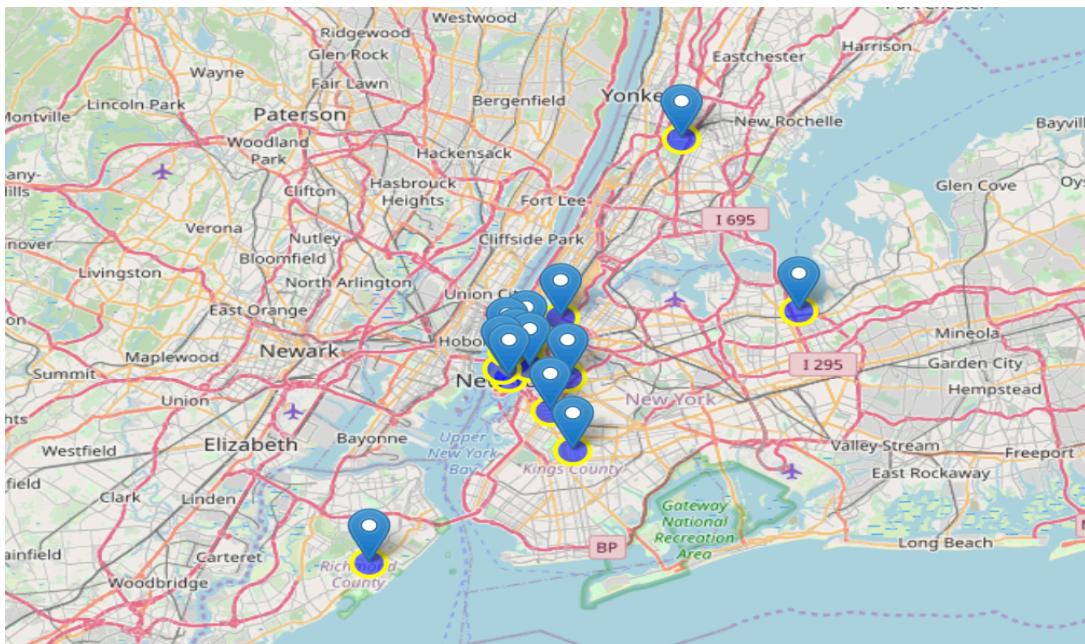
```
[48]:
```

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Bayside	40.766041	-73.774274	8.2
1	Manhattan	Civic Center	40.715229	-74.005415	8.2
2	Manhattan	Flatiron	40.739673	-73.990947	8.4
3	Brooklyn	Fort Greene	40.688527	-73.972906	8.7
4	Manhattan	Greenwich Village	40.726933	-73.999914	8.6
5	Staten Island	New Dorp	40.572572	-74.116479	8.1
6	Manhattan	Noho	40.723259	-73.988434	8.8
7	Brooklyn	North Side	40.714823	-73.958809	8.5
8	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	8.6
9	Manhattan	Sutton Place	40.760280	-73.963556	8.4
10	Manhattan	Tribeca	40.721522	-74.010683	9.1
11	Manhattan	West Village	40.734434	-74.006180	8.5
12	Bronx	Woodlawn	40.898273	-73.867315	8.0

- Visualize



- Map the data



- Prepare the data

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Borough	Average Rating	Latitude	Longitude
0	0	0	0	1	0	Queens	8.2	40.766041	-73.774274
1	0	0	1	0	0	Manhattan	8.2	40.715229	-74.005415
2	0	0	1	0	0	Manhattan	8.4	40.739673	-73.990947
3	0	1	0	0	0	Brooklyn	8.7	40.688527	-73.972906
4	0	0	1	0	0	Manhattan	8.6	40.726933	-73.999914
5	0	0	0	0	1	Staten Island	8.1	40.572572	-74.116479
6	0	0	1	0	0	Manhattan	8.8	40.723259	-73.988434
7	0	1	0	0	0	Brooklyn	8.5	40.714823	-73.958809
8	0	1	0	0	0	Brooklyn	8.6	40.658420	-73.954899
9	0	0	1	0	0	Manhattan	8.4	40.760280	-73.963556
10	0	0	1	0	0	Manhattan	9.1	40.721522	-74.010683
11	0	0	1	0	0	Manhattan	8.5	40.734434	-74.006180
12	1	0	0	0	0	Bronx	8.0	40.898273	-73.867315

- Cluster the data

```
[188]: from sklearn.cluster import KMeans
#Cluster Neighborhoods
# set number of clusters
kclusters = 6

ny_neighborhood_stats_clustering = ny_neighborhood_stats_onehot.drop('Borough', 1)
ny_neighborhood_stats_clustering
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(ny_neighborhood_stats_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
# to change use .astype()
```

[188]: array([3, 5, 5, 2, 5, 4, 1, 2, 2, 5, 1, 5, 0], dtype=int32)

- Review Output

	Cluster_Labels	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
0	3	0	0	0	1	0	8.2	40.766041	-73.774274
1	5	0	0	1	0	0	8.2	40.715229	-74.005415
2	5	0	0	1	0	0	8.4	40.739673	-73.990947
3	2	0	1	0	0	0	8.7	40.688527	-73.972906
4	5	0	0	1	0	0	8.6	40.726933	-73.999914
5	4	0	0	0	0	1	8.1	40.572572	-74.116479
6	1	0	0	1	0	0	8.8	40.723259	-73.988434
7	2	0	1	0	0	0	8.5	40.714823	-73.958809
8	2	0	1	0	0	0	8.6	40.658420	-73.954899
9	5	0	0	1	0	0	8.4	40.760280	-73.963556
10	1	0	0	1	0	0	9.1	40.721522	-74.010683
11	5	0	0	1	0	0	8.5	40.734434	-74.006180
12	0	1	0	0	0	0	8.0	40.898273	-73.867315

- Examine Cluster #1

```
[236]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 0, :]
```

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
12	1	0	0	0	0	8.0	40.898273	-73.867315

- Examine Cluster #2

```
[237]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 1, :]
```

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
6	0	0	1	0	0	8.8	40.723259	-73.988434
10	0	0	1	0	0	9.1	40.721522	-74.010683

- Examine Cluster #3

```
[213]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 2, :]
```

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
3	0	1	0	0	0	8.7	40.688527	-73.972906
7	0	1	0	0	0	8.5	40.714823	-73.958809
8	0	1	0	0	0	8.6	40.658420	-73.954899

- Examine Cluster #4

```
[214]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 3, :]
```

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
0	0	0	0	1	0	8.2	40.766041	-73.774274

- Examine Cluster #5

```
[215]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 4,
```

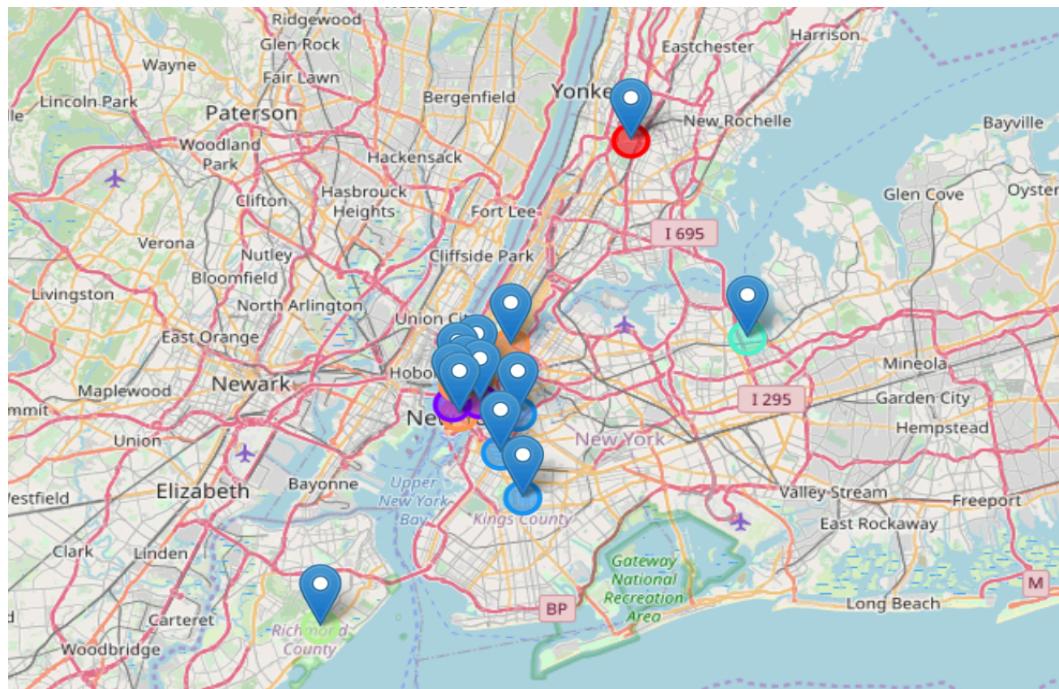
	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
5	0	0	0	0	1	8.1	40.572572	-74.116479

- Examine Cluster #6

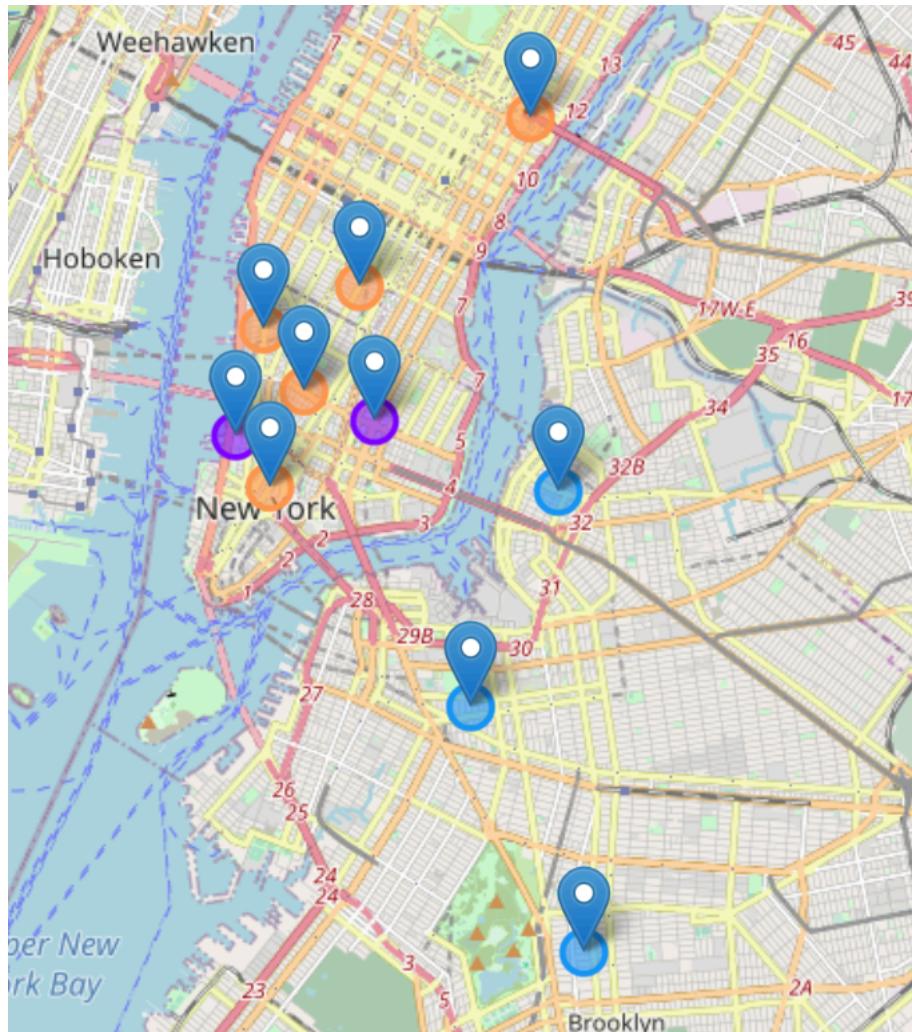
```
[216]: #Examine cluster
ny_neighborhood_stats_clustering.loc[ny_neighborhood_stats_clustering['Cluster_Labels'] == 5,
```

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Average Rating	Latitude	Longitude
1	0	0	1	0	0	8.2	40.715229	-74.005415
2	0	0	1	0	0	8.4	40.739673	-73.990947
4	0	0	1	0	0	8.6	40.726933	-73.999914
9	0	0	1	0	0	8.4	40.760280	-73.963556
11	0	0	1	0	0	8.5	40.734434	-74.006180

Results



Results (Contd)



Summary & Conclusion

- Manhattan has restaurants rated 8.8 & higher
- Manhattan also has restaurants rated ≥ 8.2 & < 8.8
- Brooklyn has the next best rated restaurants ≥ 8.5 & ≤ 8.7
- Other clusters include Bronx, Queens, Staten Island