Capstone Project - The Battle of Neighborhoods - Final Project Report

By Ratnesh Mehrotra

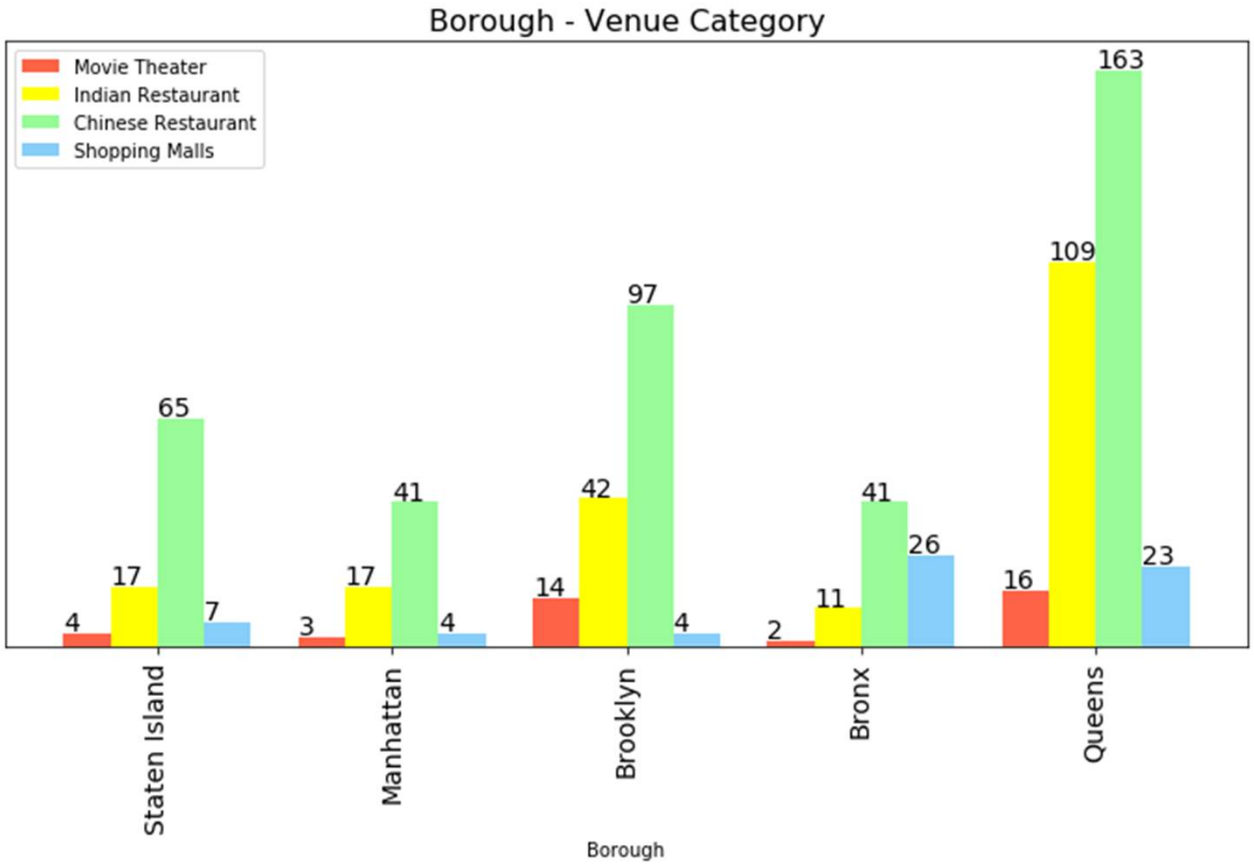
Objective

Come up with suggestions on locations of theatres which can be shortlisted for screening of the Indian movie in New York City

Data Set Analysed

- Boroughs and their Neighbourhoods, with their GeoCodes
- Following popular venues in each neighbourhood:
  - Movie Theater
  - Indian Restaurant
  - Chinese Restaurant
  - Shopping Mall
- Data Source
  - NYU Spatial Data Repository
  - FourSquare APIs

Understanding the Data Set



Number of each category in each Borough

Understanding the Data Set

```
neighborhoods_venues_sorted
```

:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Arverne	Chinese Restaurant	Shopping Mall	Movie Theater	Indian Restaurant
1	Astoria	Indian Restaurant	Shopping Mall	Movie Theater	Chinese Restaurant
2	Astoria Heights	Chinese Restaurant	Indian Restaurant	Shopping Mall	Movie Theater
3	Auburndale	Indian Restaurant	Shopping Mall	Movie Theater	Chinese Restaurant
4	Bay Terrace	Shopping Mall	Indian Restaurant	Chinese Restaurant	Movie Theater
5	Bayside	Indian Restaurant	Shopping Mall	Movie Theater	Chinese Restaurant
6	Bayswater	Chinese Restaurant	Shopping Mall	Movie Theater	Indian Restaurant
7	Beechhurst	Chinese Restaurant	Shopping Mall	Movie Theater	Indian Restaurant
8	Bellaire	Indian Restaurant	Chinese Restaurant	Shopping Mall	Movie Theater
9	Belle Harbor	Chinese Restaurant	Shopping Mall	Movie Theater	Indian Restaurant
10	Bellerose	Indian Restaurant	Chinese Restaurant	Shopping Mall	Movie Theater

Common Venues in each neighbourhood, sorted in descending order of popularity

Preparing the Data Set

Hot encoded to convert the categorical values into integer values

]:

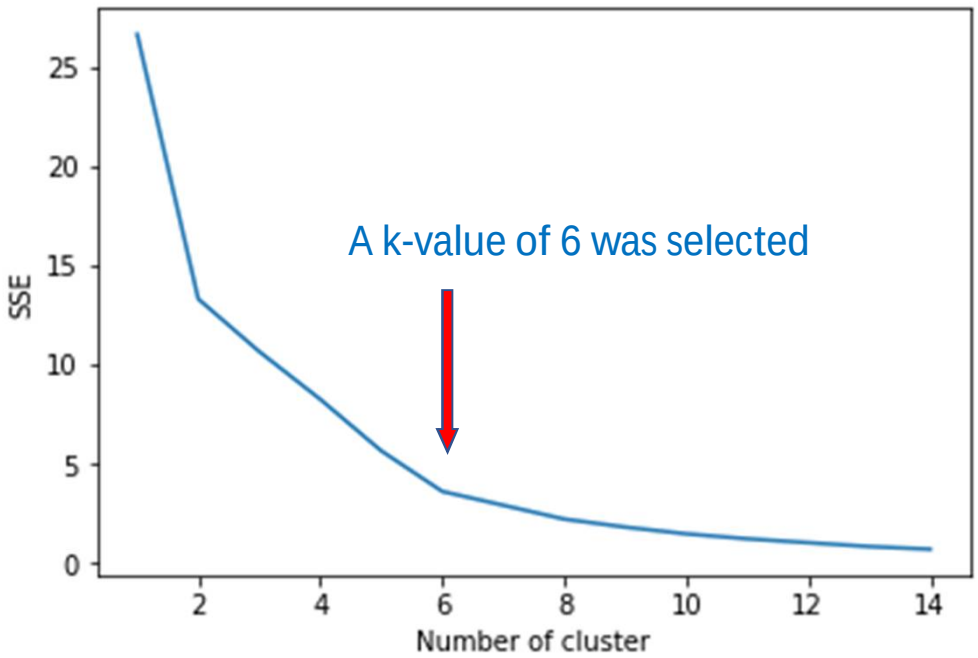
	Neighborhood	Chinese Restaurant	Indian Restaurant	Movie Theater	Shopping Mall
60	Laurelton	0	0	0	1
107	Lefrak City	0	0	0	1
142	Lefrak City	1	0	0	0
171	Lefrak City	0	0	0	1
236	Belle Harbor	1	0	0	0

Mean of the frequency of occurrence of each category

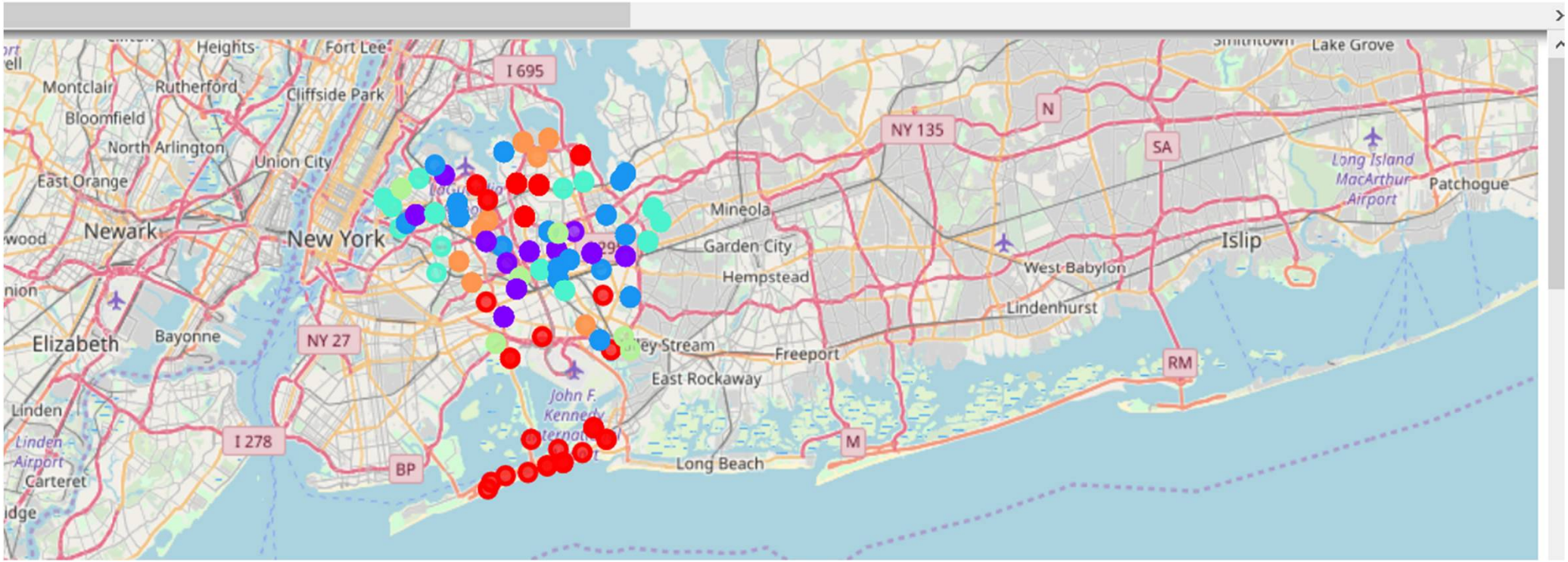
	Neighborhood	Chinese Restaurant	Indian Restaurant	Movie Theater	Shopping Mall
0	Arverne	1.000000	0.000000	0.0	0.0
1	Astoria	0.000000	1.000000	0.0	0.0
2	Astoria Heights	0.666667	0.333333	0.0	0.0
3	Auburndale	0.000000	1.000000	0.0	0.0
4	Bay Terrace	0.250000	0.250000	0.0	0.5

Modelling Approach

- Since the solution to problem lies in realm of identification of groups of neighbourhoods in a borough having Theatres, Restaurants and Shopping Malls, clustering modelling technique was selected to be most appropriate.
- k-means algorithm was selected as the clustering algorithm.
- Graph of SSE versus K-count was plotted to determine optimum value of K

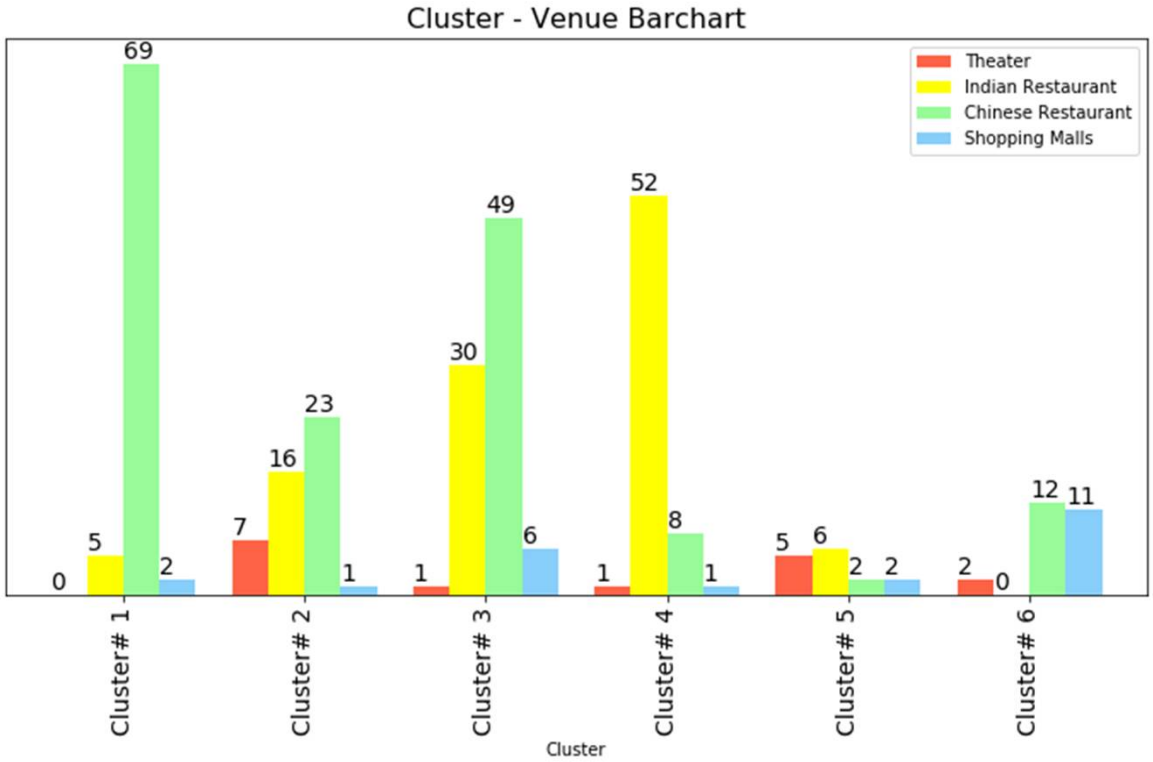


Output of Model



Clusters in Queens Borough plotted on NYC Map

Results



Composition of each cluster

1.Cluster #1 does not show any Movie Theater and hence can be discarded from consideration

2.Clusters #2 and # 4 do have good number of Indian and Chinese Restaurant but only one shopping mall each. If the movie screening happens to be during festival seasons, when malls offer good promotional offers, viewers may want to explore more than one mall and may prefer theaters located in vicinity of many shopping malls

3.Cluster #3 and #5 have all the venue categories of interest. These clusters include more than one restaurants and malls, and can be shortlisted

4.Cluster #6 does not show any Indian restaurant and hence theatres located in that region may not be popular with the movie goers

Result

Cluster # 3:
Movie Theater: AMC Fresh Meadows 7
Neighbourhood: Pomonok

Popularity CR > MT> SM > IR

Legend

MT: Movie Theater
CR: Chinese Restaurant
IR: Indian Restaurant
SM: Shopping Mall

Cluster # 5:
Movie Theater: Cinemart Cinemas, Kew Gardens Cinema
Neighbourhood: Kew Gardens

Popularity : IR > MT > SM > CR

Cluster # 5:
Movie Theater: Linden Boulevard Multiplex Cinemas
Neighbourhood: Lindenwood

Popularity : MT > CR> SM > IR

Cluster # 5:
Movie Theater: AMC Fresh Meadows 7
Neighbourhood: Utopia

Popularity : MT > CR > SM > IR

Cluster # 5:
Movie Theater: United Artists Kaufman Astoria 14
Neighbourhood: Ravenswood

Popularity : IR > MT > SM > CR

Popularity is based on output in "Table: Common Venues – sorted" (refer Notebook)

Conclusion

- This report can be used to get first level of indications where screening can be held. It is premised on the fact that an Indian movie goer will want to dine out, and do shopping when going to theatre.
- What it does not factor is the distance between selected venues – for e.g. between Theater and a restaurant
- This analysis can further be enhanced by taking weighted considerations of factors like upscale factor (for e.g. Manhattan may be preferred over Queens), approachability, ethnic diversity in the neighborhood, time of the year (festive season versus normal time of the year) etc.
- The analysis can also include venues which can be relevant - for example whether there are daycare centers nearby. Couples with young children would prefer going to theaters which have daycares nearby. Presence of such venues will be a benefit. Absence of this can lead to a new business opportunity