Clustering NYC based on restaurant categories

Finding most and least number of restaurant types available in each neighborhood

Problem Definition and Interest

- ## Problem –
  - ➢ The problem here is to determine most and least number of restaurant types available in each neighborhood.

- ## Interest –
  - ➢ People who are planning to move to a new place, they can determine which place to choose if good food plays a great role in their life
  - ➢ Start-ups can find place to open a new restaurant where that kind of restaurant is least available.

Data Acquisition

- **NYC neighborhood dataset with locations (latitude and longitude)** : This data set contains columns Borough, Neighborhood, longitude and latitude. The dataset was downloaded from IBM skills network labs.

- **Foursquare 'Places API'** is used to pull data of the venues around different neighborhoods. Data retrieval was done based on venue category '4d4b7105d754a06374d81259' which represents venues related to food. Sending request to the api gave us data in json format. The json data then formatted into a dataframe.

Data Cleaning

- Records for which venue states are not related to 'NY' and 'New York' are dropped and all the New Yorks are replaced with NY because both represents the same. One record was removed.

- 64 Entries with city name as 'N/A' are removed.

- Data set was filtered to keep data only related to the below mentioned restaurant categories.

    'Chinese Restaurant', 'Italian Restaurant', 'Mexican Restaurant', 'American Restaurant', 'Fast Food Restaurant', 'Sushi Restaurant', 'Japanese Restaurant', 'Latin American Restaurant', 'Thai Restaurant', 'Spanish Restaurant', 'Caribbean Restaurant', 'Seafood Restaurant', 'Korean Restaurant', 'Indian Restaurant', 'French Restaurant'
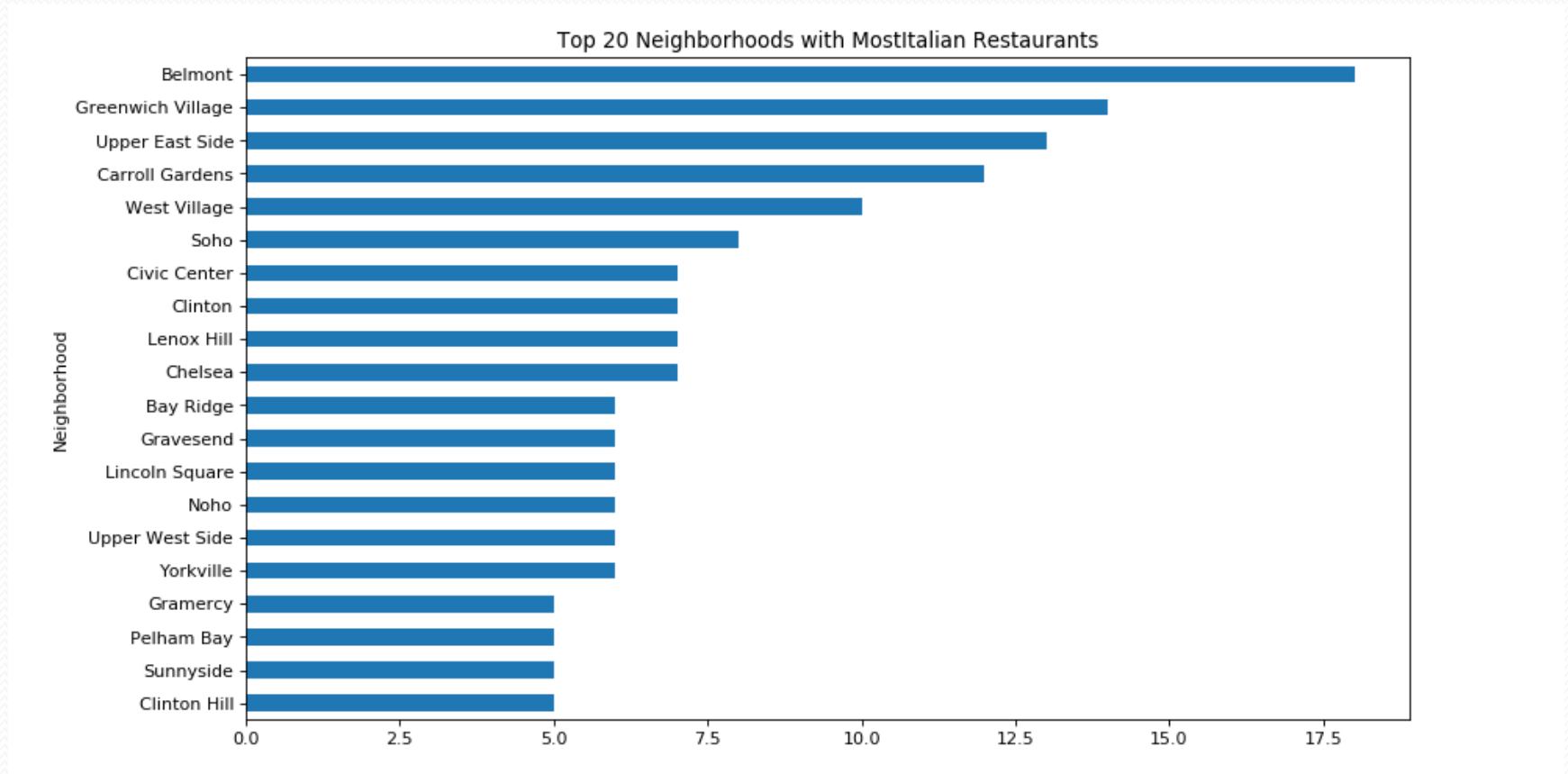
Top 20 neighborhoods with most Chinese restaurants

Top 20 Neighborhoods with MostChinese Restaurants
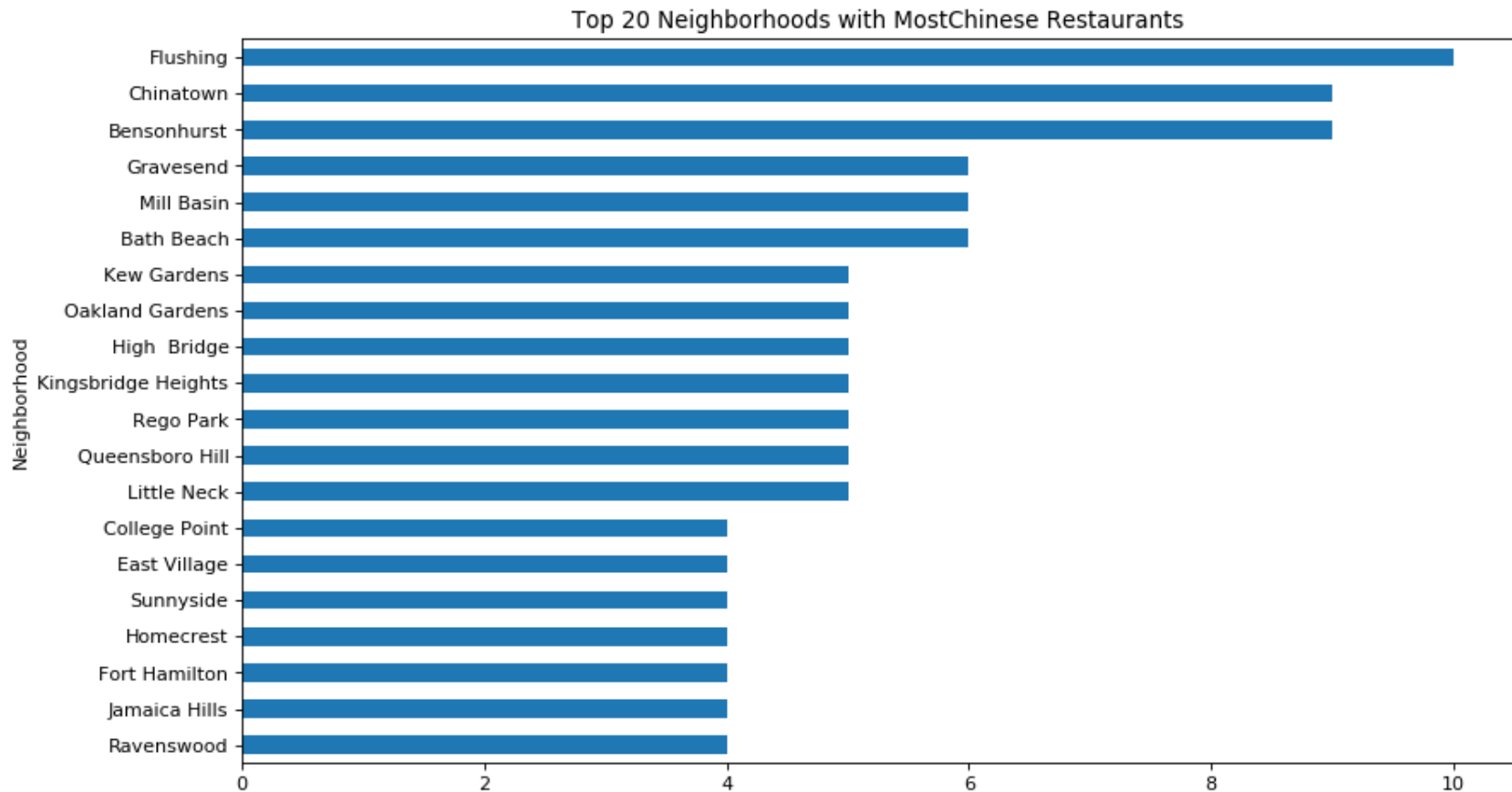
Top 20 neighborhoods with most Italian restaurants
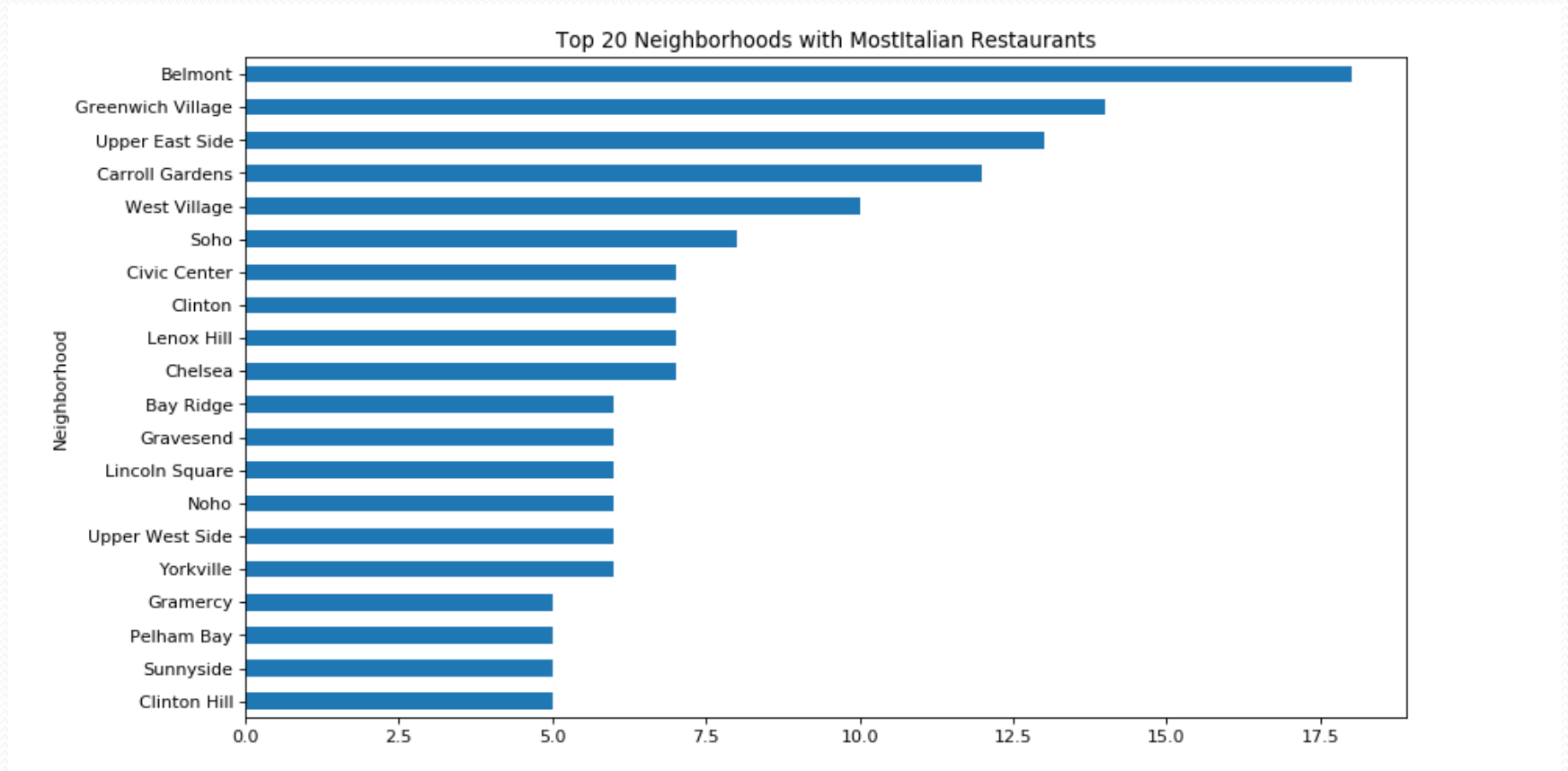
Top 20 neighborhoods with most Mexican restaurants
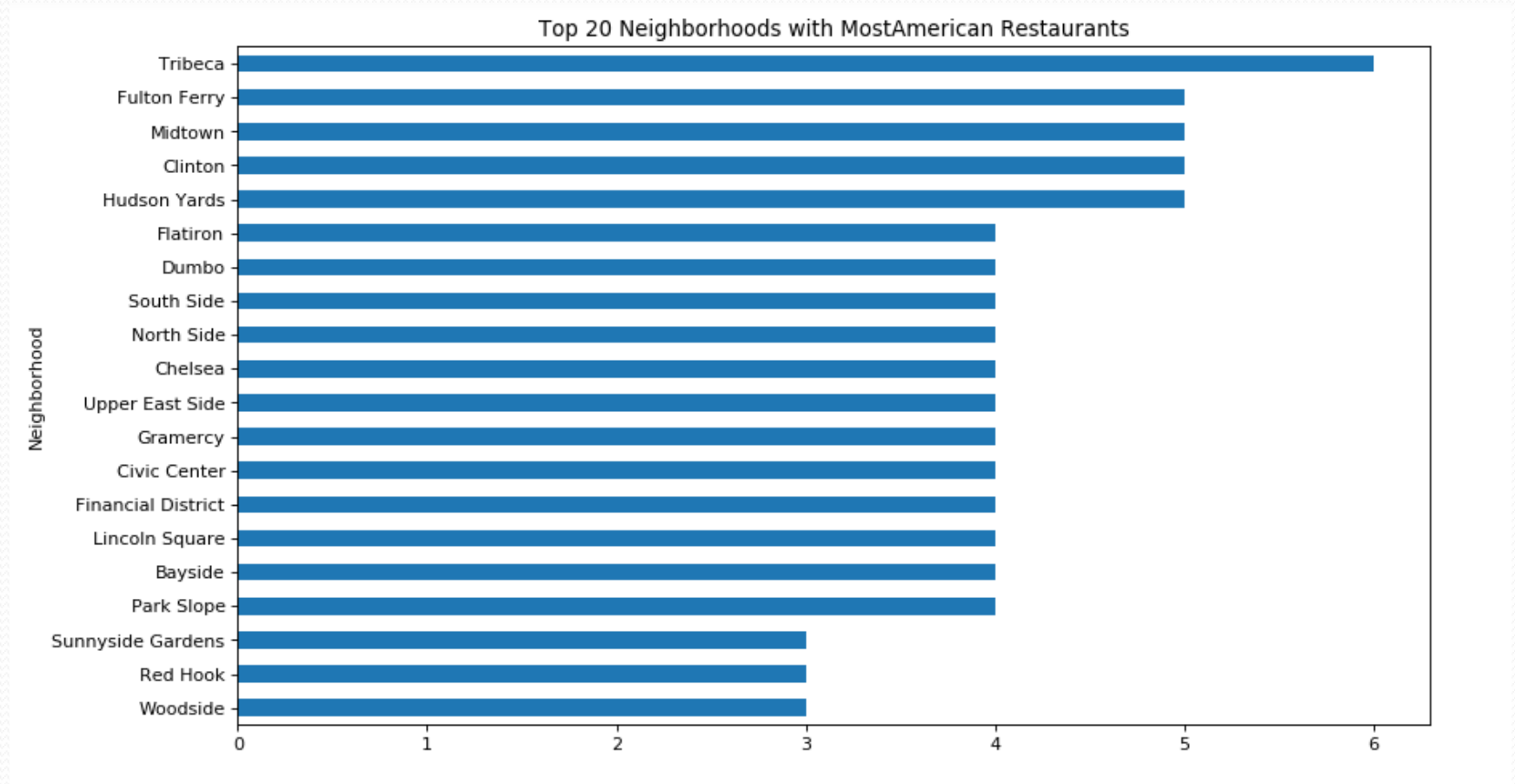

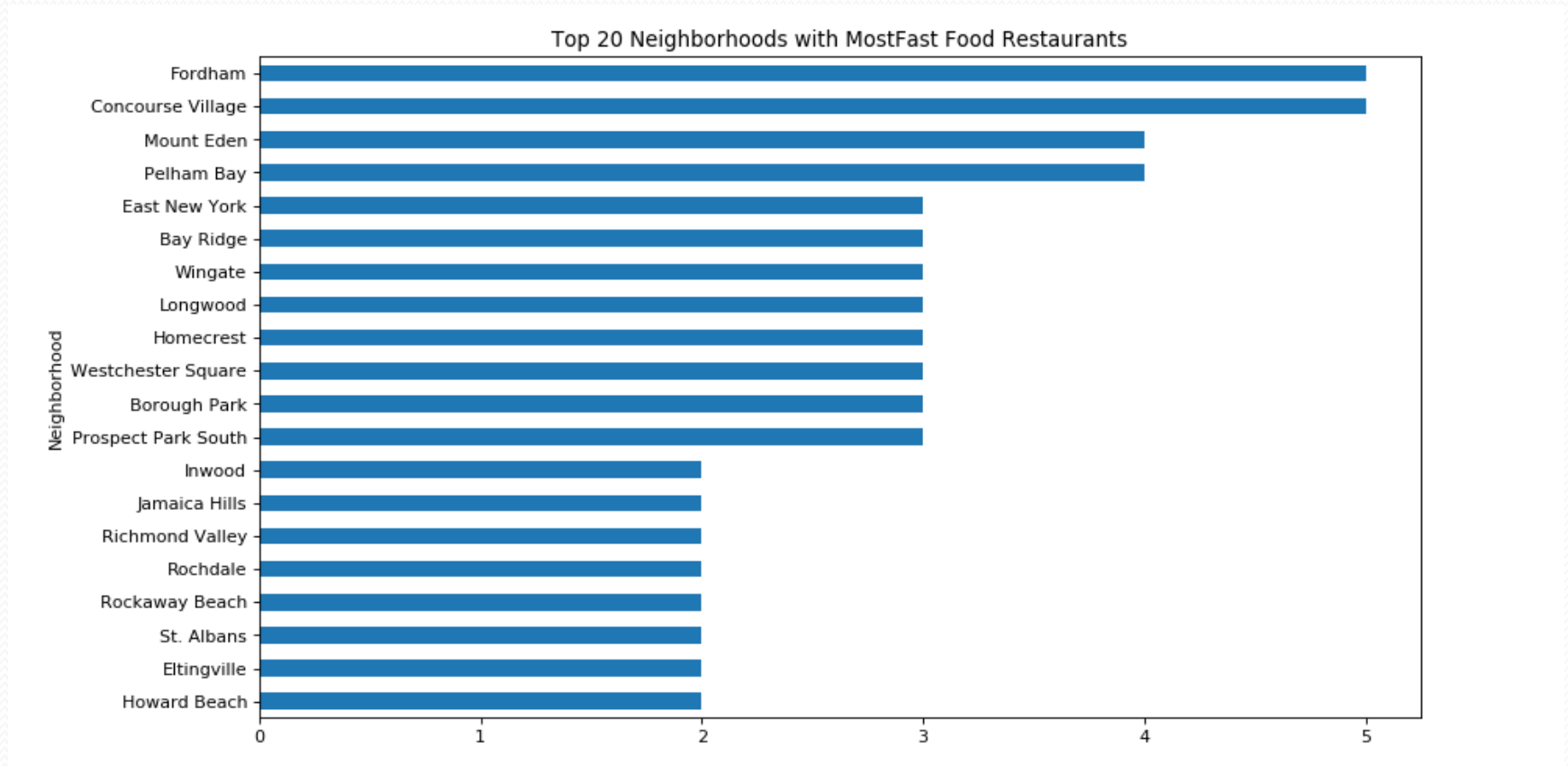Top 20 Neighborhoods with MostItalian Restaurants

Top 20 neighborhoods with most American restaurants

Top 20 Neighborhoods with MostAmerican Restaurants

Top 20 neighborhoods with most Fast Food restaurants

Top 20 Neighborhoods with MostFast Food Restaurants

Top five types of restaurants for each neighborhood

- Top categories are found by counting total number of each restaurant category for all the neighborhoods and displaying them in ascending order.

	Neighborhood	1st Top Venue Category	2nd Top Venue Category	3rd Top Venue Category	4th Top Venue Category	5th Top Venue Category
0	Allerton	Spanish Restaurant	Chinese Restaurant	Fast Food Restaurant	American Restaurant	Mexican Restaurant
1	Annadale	American Restaurant	Sushi Restaurant	Thai Restaurant	Spanish Restaurant	Seafood Restaurant
2	Arlington	Caribbean Restaurant	American Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant
3	Arrochar	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Seafood Restaurant
4	Arverne	Thai Restaurant	Sushi Restaurant	Spanish Restaurant	Seafood Restaurant	Mexican Restaurant

Bottom five types of restaurants for each neighborhood

- Bottom categories are found by counting total number of each restaurant category for all the neighborhoods and displaying them in descending order.

|	Neighborhood	1st Bottom Venue Category	2nd Bottom Venue Category	3rd Bottom Venue Category	4th Bottom Venue Category	5th Bottom Venue Category
	Allerton	Caribbean Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant
1	Annadale	Caribbean Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant	Indian Restaurant
2	Arlington	Chinese Restaurant	Fast Food Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant
3	Arrochar	American Restaurant	Caribbean Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant
4	Arverne	American Restaurant	Caribbean Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant

click to scroll output; double click to hide

Clustering

- Using Scikit learn's K means clustering algorithm 15 different clusters of neighborhoods have been created.

Cluster Label	Number of Neighborhood
1	115
13	40
5	19
8	17
0	16
9	13
12	12
14	11
6	9
7	7
4	6
2	6
11	2
10	2
3	1

Visualizing Clusters

- Neighborhood clusters are plotted on NYC's map.
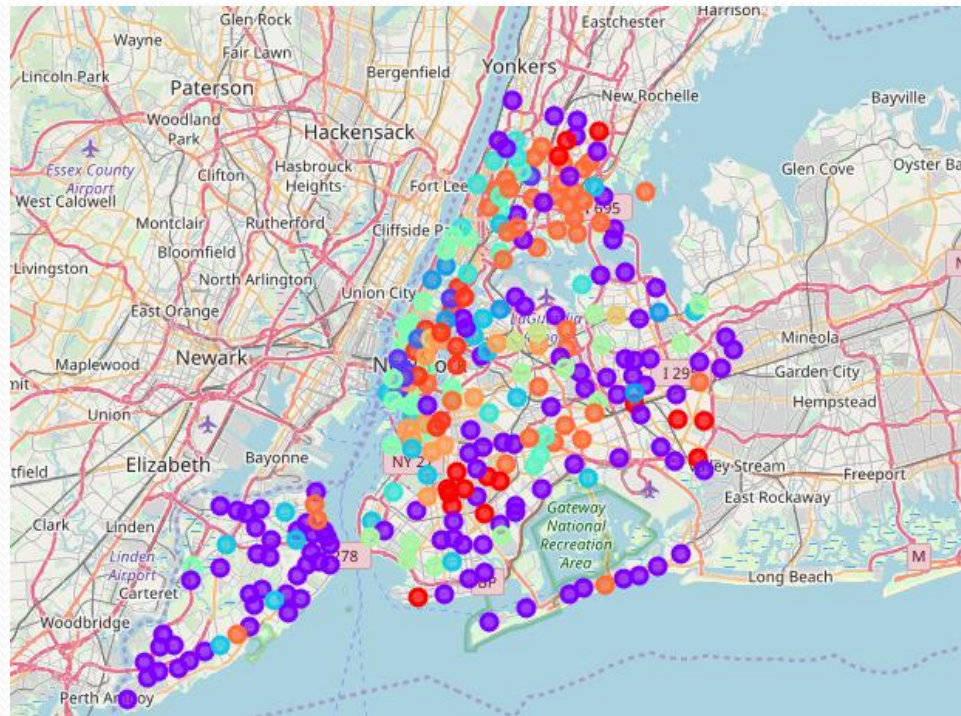
Conclusion and Future Directions

- From this analysis people interested in different cuisines can find their place of interest.

- Stakeholder how are trying to find a place to open a restaurant can find a place where a particular type of restaurant is least available.

- In future this analysis can be extended to find more precise results by collecting more data.