

IBM – Coursera
Data Science Specialization

The Capstone Project
The best neighbourhood in New York to setup an
Indian Style Restaurant

Clement Emerson A (clementemerson@gmail.com)

May 2020

Contents

Part I The Problem	3
Introduction:	3
Business Problem:.....	3
Target audience of this project:.....	3
Part II Data	4
Location Data	4
Data Source	4
Part III Methodology	5
Idea behind the analysis:	5
Indian Restaurant Categories:.....	5
Configuration for 1 st Cluster:.....	5
Configuration for 2 nd Cluster:.....	6
Configuration for 3 rd Cluster:	6
Configuration for 4 th Cluster:	6
Part IV Results	7
All types of Indian Restaurant in 5 Clusters:	7
All Indian Restaurant in 3 Clusters:.....	8
All Indian Restaurants in 5 Clusters:	9
Sum of all types of Indian Restaurant in 5 Clusters:	10
Part V Discussions	11
Part VI Limitations and Suggestions for Future Works	12
Part VII Conclusion	13

Part I The Problem

Introduction:

Indian population in New York, US is always increasing. This is evident from the wiki page

https://en.wikipedia.org/wiki/Indians_in_the_New_York_City_metropolitan_region

Rank ↕	Borough ↕	City ↕	Indian Americans ↕	Density of Indian Americans per square mile ↕	Percentage of Indian Americans in municipality's population ↕
1	Queens (2014) ^[33]	New York City	144,896	1,326.5	6.2
2	Brooklyn (2012)	New York City	25,270	357.9	1.0
3	Manhattan (2012)	New York City	24,359	1,060.9	1.5
4	The Bronx (2012)	New York City	16,748	398.6	1.2
5	Staten Island (2012)	New York City	6,646	113.6	1.4
	Total (2014) ^[31]	New York City	227,994	753.4	2.7

Even the data we have is updated as of 2014, the Indian population in NY is keep on increasing. So, there is a great opportunity to open any Indian style business in NY, especially Indian style restaurants. From the above figure, we can easily say, that, we can open a restaurant in Queens borough. But exactly where? Which neighbourhood of Queens borough is the best? There may be some other neighbourhood in some other borough which is also good for the same. We need a clearly defined approach to find the best places for setting up the Indian style restaurants. Also, the approach, that we use to find the best neighbourhood, should follow proven scientific methods. Indian restaurant and Pizza place are not same. So, the approach to find the best location for an Indian style restaurant is also different from the approach to find best location to setup a Pizza place. In this report, I am carefully drafted such a method using machine learning and finally finds a few places which are best to setup an Indian style restaurant. Even this report is New York centric, this can be easily extended to other cities also, to find the best neighbourhood in that city to setup an Indian style restaurant.

Business Problem:

I wish to keep this simple and clear. I am trying to find the answer for the following question: **Where would an Indian hotelier open an Indian style restaurant in New York, USA?**

Target audience of this project:

This project is directly useful for the investors and Indian hoteliers who are interested in setting up an India style restaurant in New York, USA. Deciding the location for setting up any business is very crucial. And the first and foremost questions in the mind of investors. This project helps them in

deciding the location. This project finds all neighbourhoods which holds good for this purpose. So, if the investor or hotelier already has a business in one of the neighbourhoods, he can extend the business to other places listed in the project results.

Part II Data

Location Data

I need the below data to carry out my study on Indian restaurants.

- List of neighbourhoods in NYC
- At max, 100 venues in each neighbourhood
- Venue category

Data Source

All data related to neighbourhoods in NYC are located in the following URL,

https://geo.nyu.edu/catalog/nyu_2451_34572. The URL has a geo JSON, in which borough, neighbourhood name, its latitude and longitude and a lot of other information is there. The data in the geo JSON is helpful to draw map. The data of our interest is, borough name, neighbourhood name, latitude and longitude. The data set has 5 boroughs and 306 neighbourhoods.

For the remaining data, we can rely on foursquare location APIs. You can find the URL for the endpoints here <https://developer.foursquare.com/docs/places-api/endpoints/>. An account needs to be created in foursquare.com to use their APIs. Foursquare provides free as well as premium APIs. For this project, I am using the free APIs.

With the neighbourhood data in our hand (latitude and longitude), we can get the venue information and the venue category information from foursquare.com. The following API is used to get the venue related information.

https://api.foursquare.com/v2/venues/explore?&client_id=CLIENT_ID&client_secret=CLIENT_SECRET&v=20180605&ll=40.89470517661,-73.84720052054902&radius=500&limit=100§ion=food

The above API will return information about 100 food related venues which are in 500m radius from the latitude and longitude mentioned in the API. This data is used for the learning purposes.

For this project, I am assuming that a person can travel up to 2KM on an average, for the purpose of eating in a restaurant. Hence, I am keeping the radius as 2KM.

Part III Methodology

Idea behind the analysis:

One can setup a Pizza Place nearly anywhere in New York. But, not an Indian Style Restaurant. Hence an analysis is needed to understand the demand for Indian cuisine in the neighbourhoods. Basically, lot of this demand comes from Indians, as Indian cuisine is very unique and demand for such a cuisine from others will be less. So, one way is to identify the population density of Indians of each neighbourhood in NYC, and decide based on the information. But this does not assure me, that if such demand actually exists or not. Hence, I am taking another approach which gives me more assurance. That is, I am collecting all the restaurants and going through them to find, where most of the Indian style restaurants are established. I am taking this approach as Indian style restaurants are already established in NYC.

Based on this information, I am clustering the neighbourhoods in 4 different methods considering the no of counts of the Indian style restaurants in the neighbourhood. Clustering the dataset once, is enough to find what we need, but clustering with different configuration increases our confidence in the result. As you can see in the result, all the cluster configuration yields comparable result.

Indian Restaurant Categories:

If someone goes through the dataset carefully, they can find there are 4 different categories of Indian style restaurants in the neighbourhoods. They are, Indian Restaurant, Dosa Place, North Indian Restaurant and South Indian Restaurant. In these, Indian Restaurant is the biggest classification for all the restaurants. And others are specific restaurant types, as India is multi-cultural.

I am using kMeans clustering for this research. This state-of-the-art machine learning algorithm clusters similar neighbourhoods together. So, in my research, this approach clusters the neighbourhood based on the demand for Indian cuisines the neighbourhood has.

With this knowledge, I am making 4 different clusters. I am explaining them below.

Configuration for 1st Cluster:

In this configuration I am clustering the neighbourhoods based on all types of Indian restaurants. I am trying to cluster the neighbourhoods into 5 clusters.

Cluster Base: Mean value of All types of Indian Restaurant established in the neighbourhood, No of Clusters: 5

Configuration for 2nd Cluster:

In this configuration, I am clustering the neighbourhoods based on 'Indian Restaurant' into 3 clusters.

Cluster Base: Mean value of Indian Restaurant, No of Clusters: 3

Configuration for 3rd Cluster:

In this configuration, I am clustering the neighbourhoods based on 'Indian Restaurants' into 5 clusters, for finer results.

Cluster Base: Mean value of 'Indian Restaurant', No of Clusters: 5

Configuration for 4th Cluster:

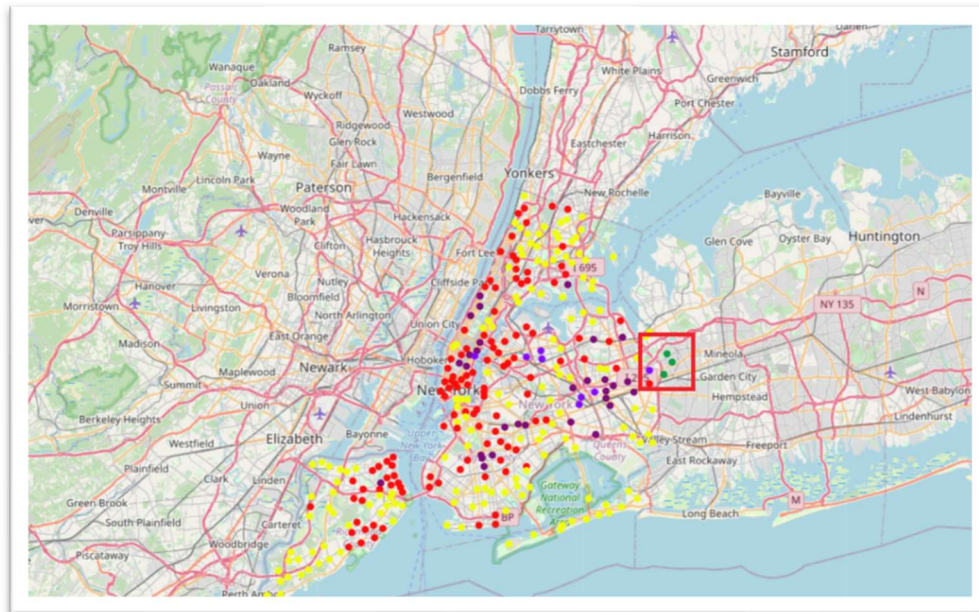
In this configuration, I am summing up all individual mean value of the Indian style restaurants and then clustering the neighbourhoods based on it into 5 clusters.

Cluster Base: Sum of all Mean value of All types of Indian Restaurant, No of Clusters: 5

Part IV Results

All types of Indian Restaurant in 5 Clusters:

Mean value of All types of Indian Restaurant established in the neighbourhood, No of Clusters: 5



Cluster 4/5 AllTypes of Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllTypes'] == 3, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

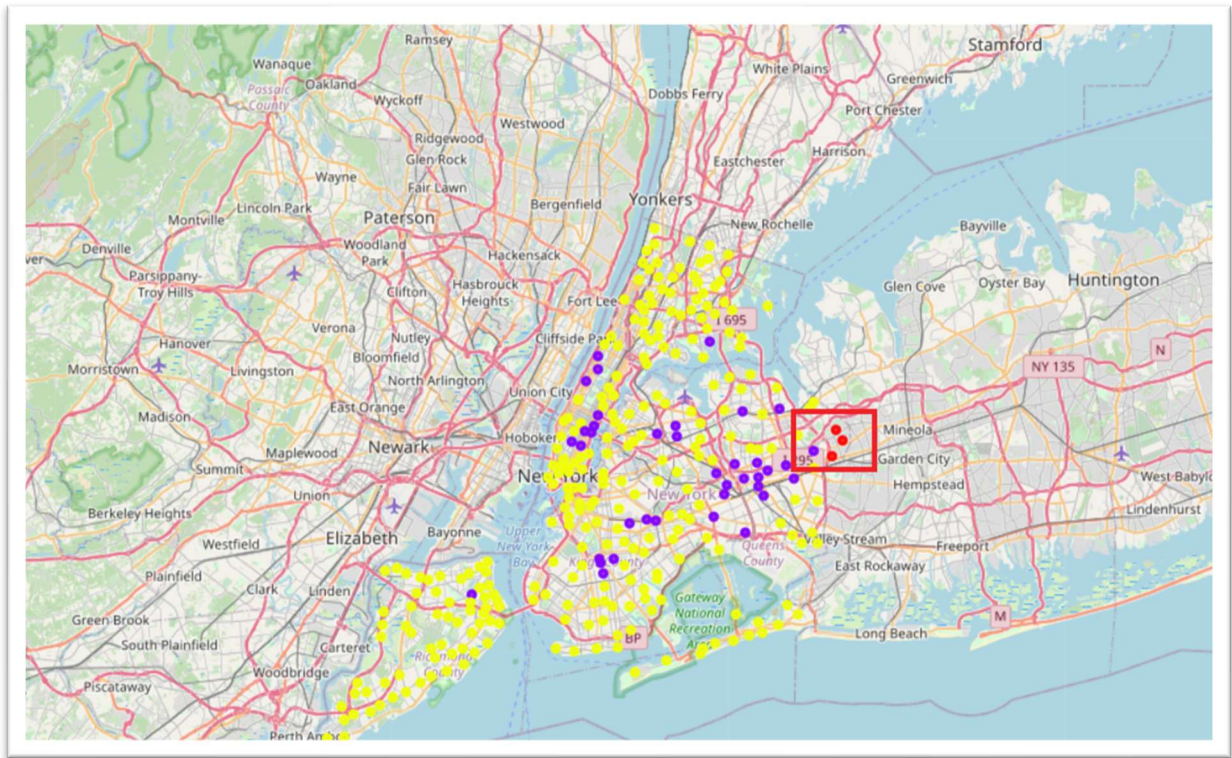
	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
155	Glen Oaks	Queens	Indian Restaurant	Pizza Place	Chinese Restaurant	Food Truck	Fast Food Restaurant	Diner	Café	American Restaurant	Sandwich Place	Bakery
156	Bellerose	Queens	Indian Restaurant	Pizza Place	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Bagel Shop	Food Truck	Donut Shop	Diner
181	Floral Park	Queens	Indian Restaurant	Pizza Place	Diner	Deli / Bodega	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	American Restaurant	Bakery	Sandwich Place

Cluster 2/5 AllTypes of Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllTypes'] == 1, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
130	Woodside	Queens	Thai Restaurant	Mexican Restaurant	Bakery	Indian Restaurant	Food Truck	Pizza Place	Latin American Restaurant	American Restaurant	Filipino Restaurant	Italian Restaurant
131	Jackson Heights	Queens	Thai Restaurant	Bakery	Indian Restaurant	Latin American Restaurant	Food Truck	Argentinian Restaurant	Pizza Place	Mexican Restaurant	South American Restaurant	Chinese Restaurant
132	Elmhurst	Queens	Thai Restaurant	Bakery	Food Truck	Indian Restaurant	Argentinian Restaurant	Mexican Restaurant	Latin American Restaurant	Filipino Restaurant	South American Restaurant	Indonesian Restaurant
136	Kew Gardens	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Chinese Restaurant	Latin American Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	New American Restaurant	Diner
137	Richmond Hill	Queens	Pizza Place	Sandwich Place	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Deli / Bodega	Latin American Restaurant	Diner	Fast Food Restaurant	Bakery
159	Briarwood	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Fried Chicken Joint	Chinese Restaurant	Donut Shop	Fast Food Restaurant	Latin American Restaurant	Sushi Restaurant	Portuguese Restaurant
194	Bellaire	Queens	Pizza Place	Deli / Bodega	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Donut Shop	Sandwich Place	Food Truck	Diner
274	Tudor City	Manhattan	American Restaurant	Indian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Sandwich Place	Korean Restaurant	Café	Steakhouse	Pizza Place	Bagel Shop

All Indian Restaurant in 3 Clusters:
Mean value of 'Indian Restaurant', No of Clusters: 3



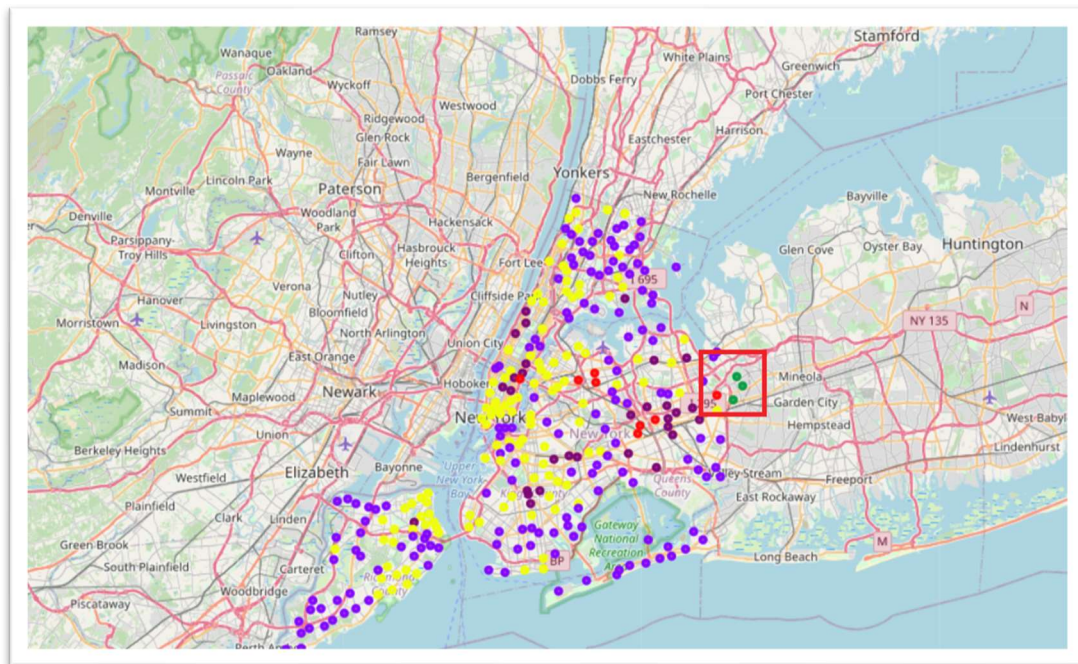
Cluster 3/3 Indian Restaurant 3 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllIndRes3'] == 2, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
155	Glen Oaks	Queens	Indian Restaurant	Pizza Place	Chinese Restaurant	Food Truck	Fast Food Restaurant	Diner	Café	American Restaurant	Sandwich Place	Bakery
156	Bellerose	Queens	Indian Restaurant	Pizza Place	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Bagel Shop	Food Truck	Donut Shop	Diner
181	Floral Park	Queens	Indian Restaurant	Pizza Place	Diner	Deli / Bodega	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	American Restaurant	Bakery	Sandwich Place

All Indian Restaurants in 5 Clusters:

Mean value of 'Indian Restaurant', No of Clusters: 5



Cluster 4/5 Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllIndRes5'] == 3, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
155	Glen Oaks	Queens	Indian Restaurant	Pizza Place	Chinese Restaurant	Food Truck	Fast Food Restaurant	Diner	Café	American Restaurant	Sandwich Place	Bakery
156	Bellerose	Queens	Indian Restaurant	Pizza Place	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Bagel Shop	Food Truck	Donut Shop	Diner
181	Floral Park	Queens	Indian Restaurant	Pizza Place	Diner	Deli / Bodega	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	American Restaurant	Bakery	Sandwich Place

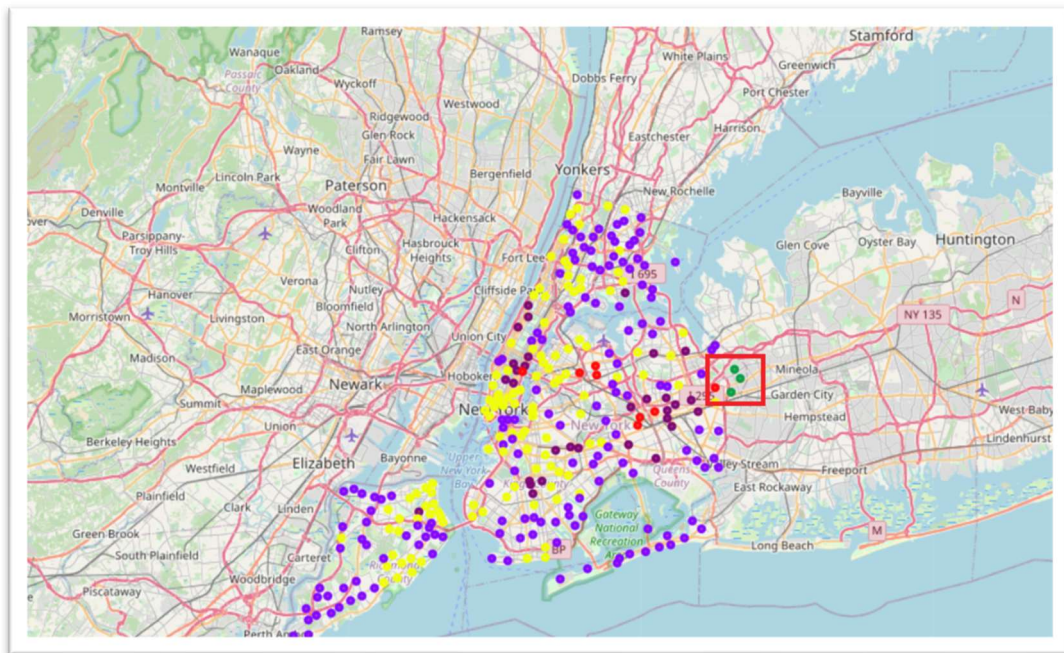
Cluster 3/5 Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllIndRes5'] == 2, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
130	Woodside	Queens	Thai Restaurant	Mexican Restaurant	Bakery	Indian Restaurant	Food Truck	Pizza Place	Latin American Restaurant	American Restaurant	Filipino Restaurant	Italian Restaurant
131	Jackson Heights	Queens	Thai Restaurant	Bakery	Indian Restaurant	Latin American Restaurant	Food Truck	Argentinian Restaurant	Pizza Place	Mexican Restaurant	South American Restaurant	Chinese Restaurant
132	Elmhurst	Queens	Thai Restaurant	Bakery	Food Truck	Indian Restaurant	Argentinian Restaurant	Mexican Restaurant	Latin American Restaurant	Filipino Restaurant	South American Restaurant	Indonesian Restaurant
136	Kew Gardens	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Chinese Restaurant	Latin American Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	New American Restaurant	Diner
137	Richmond Hill	Queens	Pizza Place	Sandwich Place	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Deli / Bodega	Latin American Restaurant	Diner	Fast Food Restaurant	Bakery
159	Briarwood	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Fried Chicken Joint	Chinese Restaurant	Donut Shop	Fast Food Restaurant	Latin American Restaurant	Sushi Restaurant	Portuguese Restaurant
194	Bellaire	Queens	Pizza Place	Deli / Bodega	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Donut Shop	Sandwich Place	Food Truck	Diner
274	Tudor City	Manhattan	American Restaurant	Indian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Sandwich Place	Korean Restaurant	Café	Steakhouse	Pizza Place	Bagel Shop

Sum of all types of Indian Restaurant in 5 Clusters:

Sum of all Mean value of All types of Indian Restaurant, No of Clusters: 5



Cluster 4/5 Sum of Means of AllTypes of Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllSum'] == 3, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
155	Glen Oaks	Queens	Indian Restaurant	Pizza Place	Chinese Restaurant	Food Truck	Fast Food Restaurant	Diner	Café	American Restaurant	Sandwich Place	Bakery
156	Bellerose	Queens	Indian Restaurant	Pizza Place	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Bagel Shop	Food Truck	Donut Shop	Diner
181	Floral Park	Queens	Indian Restaurant	Pizza Place	Diner	Deli / Bodega	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	American Restaurant	Bakery	Sandwich Place

Cluster 3/5 Sum of Means of AllTypes of Indian Restaurant 5 Clusters

```
ny_merged.loc[ny_merged['Cluster_AllSum'] == 2, ny_merged.columns[[1] + [0] + list(range(8, ny_merged.shape[1]))]]
```

	Neighborhood	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
130	Woodside	Queens	Thai Restaurant	Mexican Restaurant	Bakery	Indian Restaurant	Food Truck	Pizza Place	Latin American Restaurant	American Restaurant	Filipino Restaurant	Italian Restaurant
131	Jackson Heights	Queens	Thai Restaurant	Bakery	Indian Restaurant	Latin American Restaurant	Food Truck	Argentinian Restaurant	Pizza Place	Mexican Restaurant	South American Restaurant	Chinese Restaurant
132	Elmhurst	Queens	Thai Restaurant	Bakery	Food Truck	Indian Restaurant	Argentinian Restaurant	Mexican Restaurant	Latin American Restaurant	Filipino Restaurant	South American Restaurant	Indonesian Restaurant
136	Kew Gardens	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Chinese Restaurant	Latin American Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	New American Restaurant	Diner
137	Richmond Hill	Queens	Pizza Place	Sandwich Place	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Deli / Bodega	Latin American Restaurant	Diner	Fast Food Restaurant	Bakery
159	Briarwood	Queens	Pizza Place	Sandwich Place	Indian Restaurant	Fried Chicken Joint	Chinese Restaurant	Donut Shop	Fast Food Restaurant	Latin American Restaurant	Sushi Restaurant	Portuguese Restaurant
194	Bellaire	Queens	Pizza Place	Deli / Bodega	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Donut Shop	Sandwich Place	Food Truck	Diner
274	Tudor City	Manhattan	American Restaurant	Indian Restaurant	Japanese Restaurant	Mediterranean Restaurant	Sandwich Place	Korean Restaurant	Café	Steakhouse	Pizza Place	Bagel Shop

Part V Discussions

With all the visualizations and data provided in the result section, it is very easy to decide the neighbourhoods that we are interested on. Glen Oaks, Bellerose and Floral Park in Queens Borough are the neighbourhoods which have a good number of Indian Restaurants. As you can see, in these neighbourhoods, 'The 1st most important venue' among the restaurants itself is 'Indian Style Restaurants'. This is in-line with the population density data which I depict in the introduction section.

Interesting thing in making clustering is, we can also understand the 2nd level of demand for the Indian cuisine. They are from, Woodside, Jackson Heights, Elmhurst, Kew Gardens, Richmond Hill, Briarwood and Bellaire of Queens borough. Surprisingly, Tudor City of Manhattan borough has a nice demand for Indian cuisines.

There are many neighbourhoods which has some notable demand for Indian cuisines. They are listed below:

Manhattan:

- Upper West Side
- Murray Hill
- Manhattan Valley
- Morningside Heights
- Gramercy
- Sutton Place
- Turtle Bay
- Flatiron

Queens:

- Ozone Park
- South Ozone Park
- Bayside
- Kew Gardens Hills
- Jamaica Center
- Hollis
- South Jamaica
- Murray Hill
- Holliswood

- Jamaica Estates
- Hillcrest
- Forest Hills Gardens
- Jamaica Hills

Brooklyn:

- Flatbush
- Ocean Hill
- Prospect Park South
- Ditmas Park
- Weeksville
- Broadway Junction
- Erasmus

Staten Island:

- Silver Lake

Bronx:

- Castle Hill

Part VI Limitations and Suggestions for Future Works

This research is analysing the demand for Indian cuisines as a whole. Indian cuisines have many sub-cuisines as India is a multi-culture country. A study on this is necessary to setup the exact Indian cuisine Restaurant. Probable way, is to analyse the menu of the Indian Restaurants and the tips provided by the people. Both of these are available as premium APIs in foursquare.com.

Also, the per-capita income of the people and expenses pattern of the people in the neighbourhood is not studied. A study on this is required to understand 'Class' of the restaurant which is appropriate to setup in the neighbourhood, which in-turn give us an idea about how much investment is needed to setup the restaurant.

Part VII Conclusion

It is very surprising for me to see the results of this research, that NYC has a good demand for Indian cuisines. This research finds the best neighbourhoods in NYC to setup an Indian style restaurant. As stated in the Introduction section, the Indian population density is more in Queens borough compared to other boroughs in NYC. But that doesn't mean that, one can setup Indian restaurants at any neighbourhood in Queens borough. Even though Queens borough in NYC has the highest demand, a few neighbourhoods in other boroughs have more demand for Indian cuisines than some neighbourhoods in Queens borough. This is evident from the results.

END OF THE REPORT