Coursera Capstone

IBM Applied Data Science Capstone

Are you thinking about opening a new Indian Restaurant near New York City (NYC)?

By: Pronab Das

July 2020

1. Introduction

New York City, world's major commercial, financial and cultural center. It has been a center for trade and economic growth. New York is known worldwide as a cultural melting pot. New York City Metropolitan Area contains the largest metropolitan Asian Indian population in the Western Hemisphere. According to the 2018 Asian American Federation, New York City is home to approximately 650,000 people from the Indian subcontinent and with its diverse culture, comes diverse food items.

Restaurant is a place where people come to have food and drinks for a cost. Restaurants are a notoriously difficult business to own or operate. Not only is it a relatively capital and employee-intensive business, restaurants are also highly regulated, low margin and in most cases have a plethora of competition to deal with. There are more dining establishments and more diners now than during any other time in history; the National Restaurant Association reports over 1 million restaurant locations in the U.S. alone, a boon considering more than half the American population visits at least one of them each week. To fill a niche and remain relevant, prospective, as well as established, restaurants must hedge their bets with well-rounded and well-directed dataset. Data

science provides valuable insights regarding market trends and evolving consumer lifestyles so that restaurateurs can better address and meet public demand.

2. Business Problem

My client wants to open an Indian restaurant in this one of the world's most diverse region. No single restaurant is ever going to appeal to everyone. Some people like quiet, intimate settings; others prefer boisterous ones. Some people want to bring their kids along; others want to dine alone. The preferences are as varied as the possible offerings, with specific generational cohorts preferring one thing, "people who like Indian food" preferring another and everyone else liking a million other things in between.

So, what are all the factors we must consider determining which neighborhood will have best potential to open an Indian restaurant?

Population, Competition in the neighborhood, Surrounding amenities like parks, visiting sites, entertainment, Market Places, Menu items and so on…Our solution needs to be data driven for avoiding or considering low risk criteria and high success rate and thus applying data science techniques and methodologies would lead us to take correct decision.

3. Target Audience

- Business personnel who wants to invest or open a restaurant
- Freelancer who loves to have their own restaurant as a side business
- Small business owner finding the best location for opening a restaurant
- Budding Data Scientists who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it

4. Data

4.1. Data Sources

We will be analyzing data and provide insight as per below to the client to determine which neighborhood in New York City will be best choice to start a restaurant.

- *NYC Population & Demographic characteristics*

Data sources:

Population – https://en.wikipedia.org/wiki/New_York_City

Demographic Characteristics - https://en.wikipedia.org/wiki/Demographics_of_New_York_City

Web scraping techniques was used to get NYC's population density and demographics data from Wikipedia. Preliminary finding indicates that Queens being the second most populous urban area in New York City (NYC), behind Brooklyn; and the most ethnically diverse urban area in NYC with the highest Asian ethnic minority population.

- *Who are the competitors in that location?*

Data source:

https://cocl.us/new_york_dataset

New York City data that contains Borough, Neighborhoods along with their latitudes and longitudes. These datasets will be used to explore various neighborhoods and each Indian restaurants venue in the neighborhood

Data Source:

Foursquare API

By using this API we will get all the venues in the neighborhoods. The data going to be collected/acquired from the Foursquare API about the various restaurants in each neighborhood of New York city especially Indian cuisine, which will be used for acquiring the information regarding same for all the venues of each neighborhood.

4.2. Data Preparation

- We prepared New York City (NYC) population and demographic data from one of the data sources which would be helpful analyzing and identifying best neighborhood as one of the important factors.
- Also identified list of neighborhoods using Foursquare API & New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
- Based on the above list we found top 100 venues within a radius of 1000 meters for a given latitude and longitude.
- We prepared list of Indian restaurants in the neighborhoods based on Tips, Likes and Ratings using Foursquare API.
- Based on the above list we prepared list of top rated Indian restaurants with maximum tips, likes and highest average ratings.

5. Methodology

5.1. Approach

In this project we will be detecting areas of New York City (NYC) that have low restaurant density, particularly those with low number of Indian restaurants.

We have collected the required location data to get details of boroughs and neighborhoods along with latitude and longitude. We have also identified Indian restaurants (according to Foursquare categorization) by those boroughs and neighborhoods.

Next step in our analysis will be calculation and exploration of restaurant density across different areas of New York City (NYC) along with total population across different areas.

In the final step we will focus on most promising areas. We will take into consideration locations with fewer Indian restaurants, high population density and explore data based on Tips, Likes and Ratings. We will present map of all such locations and search for optimal venue location.

5.2. Data Analysis

5.2.1. Neighborhoods based on New York City (NYC) population and demographic data

Preliminary finding indicates that Queens is the most populous urban area in New York City (NYC). It is also the most ethnically diverse urban area in NYC with the highest Asian ethnic minority population.

Even though Manhattan is the second most populous urban area in New York City (NYC), it has a population density of 27,826 people per square km, making it highest of any borough in the United States. It has the second highest Asian ethnic minority population in NYC.

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

5.2.2. List of Boroughs and Neighborhoods in New York City (NYC)

We could see that based on the dataset, there are a total of 306 different Neighborhoods in New York City (NYC) which can be explored to open the restaurant.

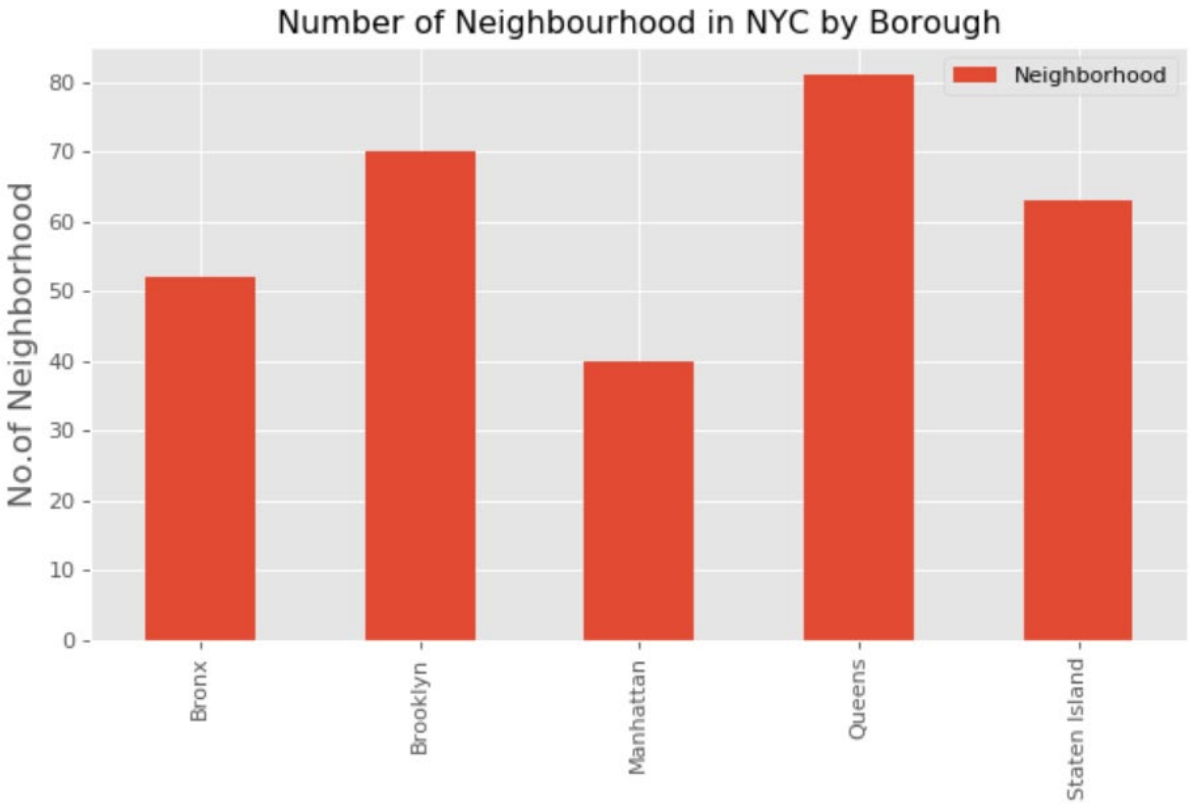
Out[10]:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

```
]: new_york_data.shape
```

Out[11]: (306, 4)

We also observed that Queens is the most densely populated borough in New York City (NYC) with more than 80 neighborhoods followed by Brooklyn, Staten Island, Bronx and Manhattan.



5.2.3. List of Boroughs and Neighborhoods in New York City (NYC) with Indian restaurants

we could see that there are total 138 different neighborhoods out of 306 total neighborhoods in New York City (NYC) which currently contains Indian restaurant.

```
indian_rest_ny.head()
```

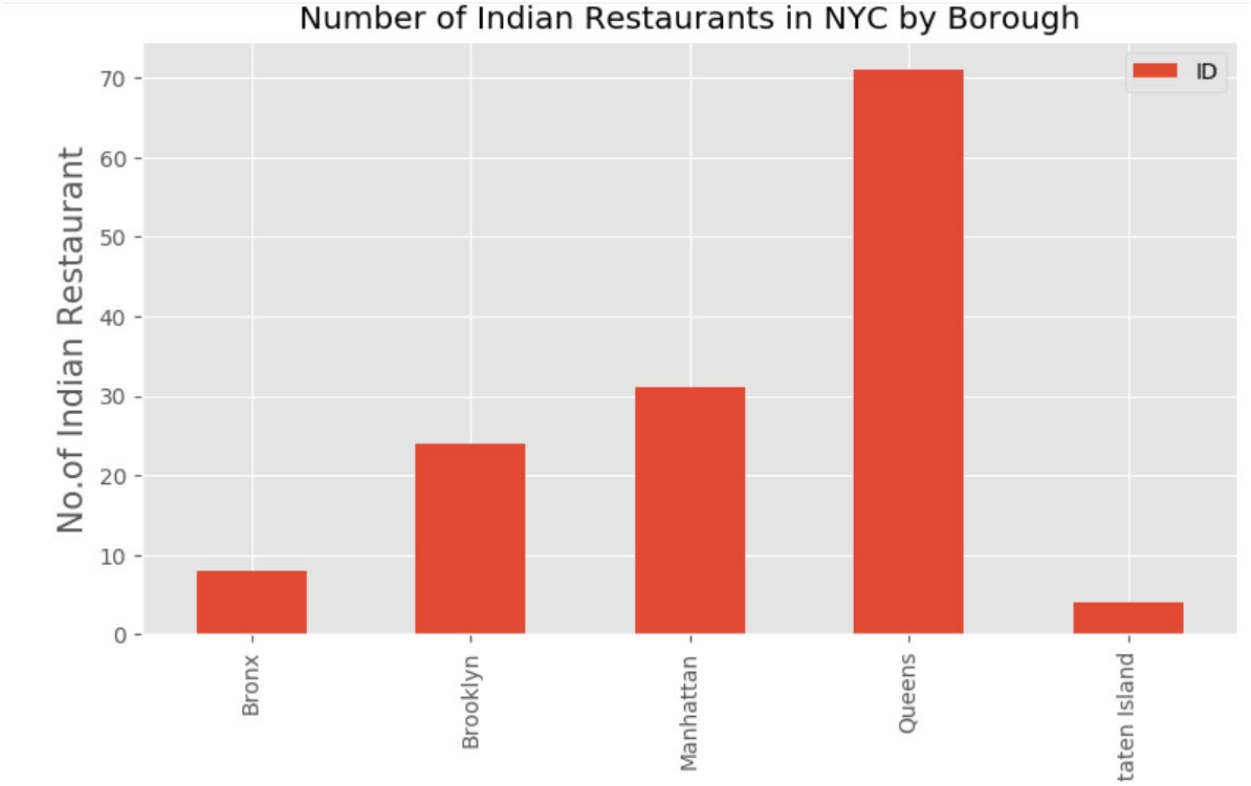
Out[14]:

	Borough	Neighborhood	ID	Name
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
1	Bronx	Williamsbridge	570c3a3fcd10eecd0d0434cd	Agra Grill
2	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill
3	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine
4	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird

```
indian_rest_ny.shape
```

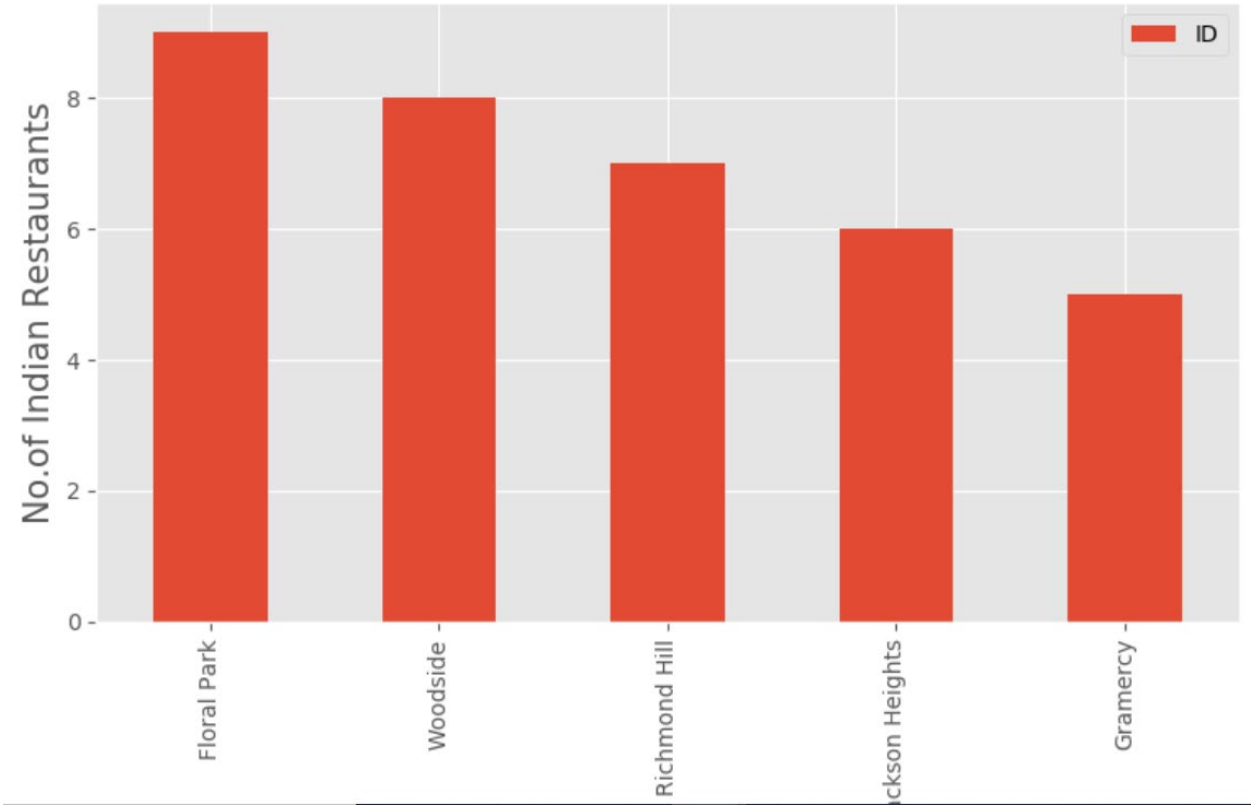
Out[15]: (138, 4)

We have also observed that Queens has the highest number of Indian Restaurants with more than 70 restaurants followed by Manhattan, Brooklyn, Bronx and Staten Island.



5.2.4. List of Neighborhoods in Queens with most number of Indian restaurants

We could see that Floral Park in Queens has the most Indian Restaurants with a total count of 11.



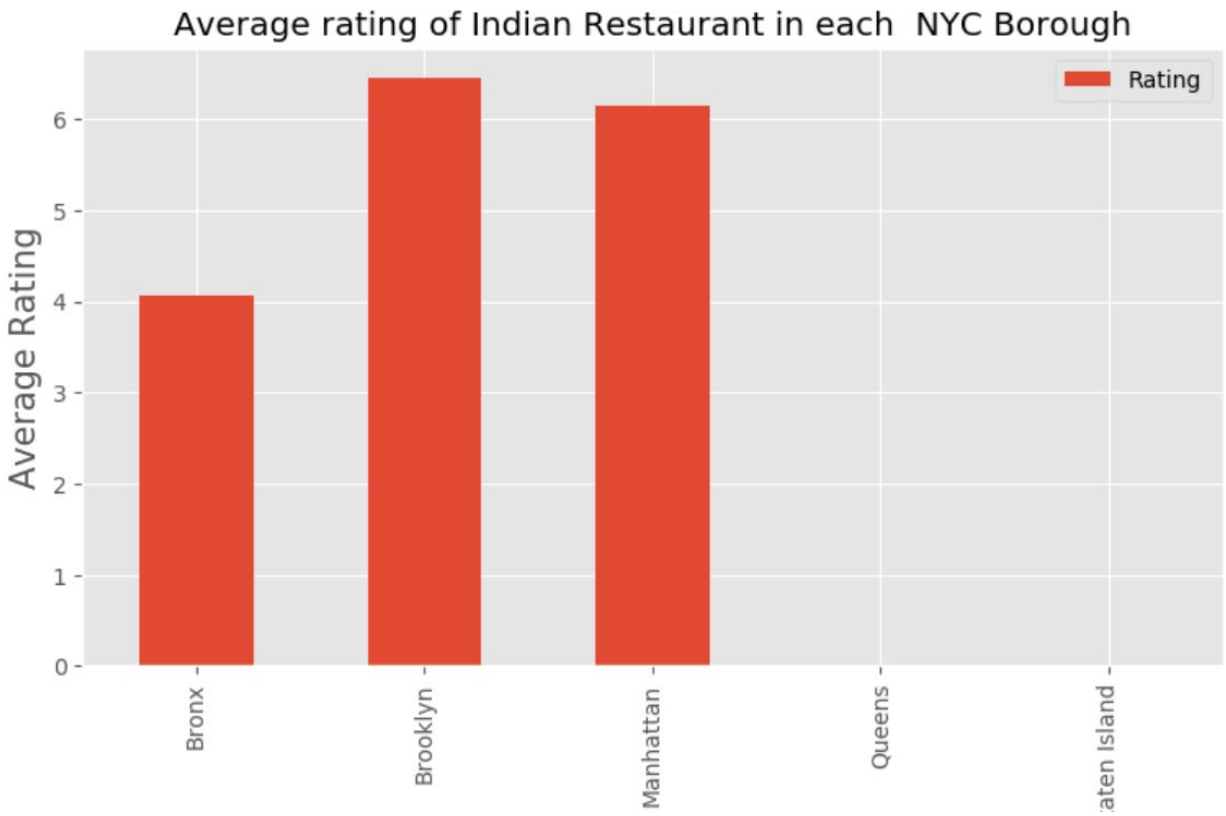
5.2.5. List of Neighborhoods & Boroughs in New York City (NYC) area with Indian restaurants based on top Likes, Tips & Ratings

Tribeca, Fort Greene and West Village are some of the top rated neighborhoods on Indian restaurants.

	Neighborhood	Average Rating
62	Tribeca	9.00
25	Fort Greene	8.80
12	Chelsea	8.70
67	West Village	8.70
65	Upper West Side	8.50
71	Yorkville	8.40
14	Clinton Hill	8.35
54	Roosevelt Island	8.30
40	Manhattan Valley	8.20
42	Morningside Heights	8.20

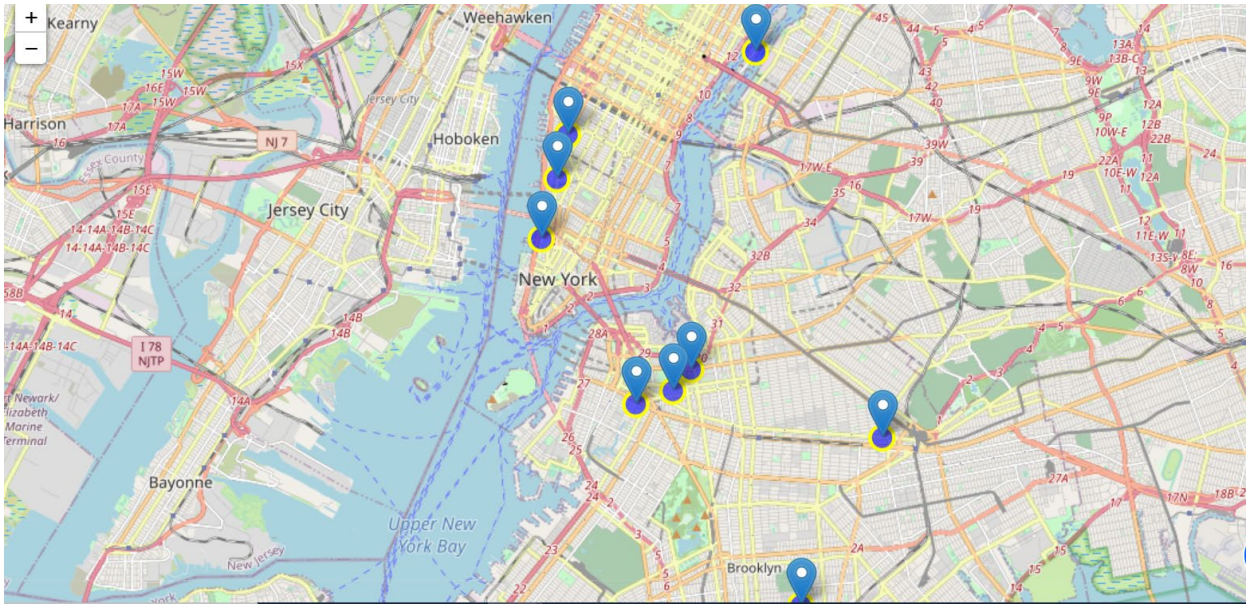
Manhattan & Brooklyn are top rated boroughs based on average rating on Indian restaurants.

	Borough	Average Rating
1	Brooklyn	6.450000
2	Manhattan	6.135484
0	Bronx	4.062500
3	Queens	0.000000
4	Staten Island	0.000000



	Borough	Neighborhood	Latitude	Longitude	Average Rating
11	Manhattan	Tribeca	40.721522	-74.010683	9.00
6	Brooklyn	Fort Greene	40.688527	-73.972906	8.80
2	Manhattan	Chelsea	40.744035	-74.003116	8.70
3	Staten Island	Chelsea	40.594726	-74.189560	8.70
14	Manhattan	West Village	40.734434	-74.006180	8.70
12	Manhattan	Upper West Side	40.787658	-73.977059	8.50
16	Manhattan	Yorkville	40.775930	-73.947118	8.40
4	Brooklyn	Clinton Hill	40.693229	-73.967843	8.35
10	Manhattan	Roosevelt Island	40.762160	-73.949168	8.30
8	Manhattan	Morningside Heights	40.808000	-73.963896	8.20

5.2.6. Visualizing the data in a map



6. Results & Discussions

Based on the population and demographic data we identified that Queens is the most populous area in New York City (NYC) along with highest Asian ethnic population followed by Manhattan.

After analyzing restaurants data, there are a total of 306 different Neighborhoods in New York City (NYC) which can be explored to open the restaurant. Among those Queens is the most densely populated borough in New York City (NYC) with more than 80 neighborhoods followed by Brooklyn, Staten Island, Bronx and Manhattan. After focusing on Indian cuisine and restaurant we identified that there are total 138 different neighborhoods out of 306 total neighborhoods in New York City (NYC) which currently contains Indian restaurant. Looking into further details we could find that Queens has the highest number of Indian Restaurants with more than 70 restaurants followed by Manhattan, Brooklyn, Bronx and Staten Island. By drilling down to neighborhood level we identified that Floral Park in Queens has the most Indian Restaurants with a total count of 11 followed by Woodside, Gramercy.

Based on the tips, ratings and likes data for all 138 Indian Restaurants we identified Tribeca neighborhood under Manhattan borough has maximum likes as well as highest average ratings of the restaurants serving Indian cuisine whereas Gramercy neighborhood under Queens borough has highest number of tips. We have also seen that Tribeca, Fort Greene, Chelsea and West Village are

the top neighborhoods with highest average ratings whereas Brooklyn and Manhattan are top boroughs in that list. We could also see that Bronx is the lowest rated among New York City (NYC) borough for Indian restaurants.

It is important to note that it's entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. On the other hand considering neighborhood with highest average rating may impose bigger competition to sustain and maintain among the competitors in the same neighborhood. Recommended neighborhoods should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors taken into account and all other relevant conditions met.

7. Conclusion

Based on the Data Analysis and Visualization we can come to a better understanding of the interests we had earlier to be discovered which are specified as follows:

- Queens and Manhattan boroughs are most populous along with Asian ethnic population
- Tribeca (Manhattan), West Village (Manhattan) and Gramercy (Queens) are some of the best neighborhoods for Indian cuisine
- Bronx and Staten Island are ranks last in the average ratings of Indian restaurant, therefore not suitable to start with
- Manhattan and Queens borough has the best potential in the Indian restaurant market based on the population, Asian community and average Indian restaurant rating
- Manhattan also has a population density of 27,826 people per square km, highest of any borough in the United States which makes her to be a potential and long term sustainable market

Final decision on optimal restaurant neighborhood will be made by stakeholders based on specific characteristics of neighborhoods and nearby locations in every recommended borough taking into consideration additional factors like attractiveness of each location, amenities, social and economic dynamics of every neighborhood etc.