

IBM - Applied Data Science Capstone - The battle of neighbourhoods

Introduction

New York City represents one of the most attractive cities of the United States of America and of the world. Indeed, it offers an extremely wide range of business opportunities: from finance and real estate to digital marketing, tourism, art and fashion, this city attracts talents from all over the world. However, the competition among all these ambitious talents is very high and developing a successful business can be very hard and risky. As a consequence, any plan aimed at starting a new business in the city should be analysed by taking into consideration not only the idea behind it, but also the competitive landscape. This analysis aims at clarifying the business environment that New York City represents, with the final goal of providing applicable insights for new entrepreneurs, to successfully confront themselves with this market and start a profitable business.

The new business in analysis is represented by a successful Indian restaurant, named "The Elephant", trying to expand in North America and open a new restaurant in New York City and aiming at becoming a point of reference for its unique offer in the heterogeneous mix of cuisines located in New York City.

The culture of food and the industry behind it in New York City mainly involves: North American cuisine Asian cuisine (Japanese, Indian, Chinese and Tibetan); Middle Eastern cuisine; European cuisine (Italian, French and Spanish); Ecommerce-based American food providers.

This wide mix of choices makes it very hard for a new restaurant to emerge and stay profitable. One of the first choices to be made, that will also have a big impact on the success of the business, is its location.

The choice of the location depends on several variables: population density of the areas; Asian ethnic minority population size of the areas; distance from top rated Indian restaurants located in the area.

The objective of this analysis will be to help "The Elephant" choosing the best New York City's spot for the business to start and remain profitable in the long run. Our team of data scientists will analyze data relative to the previously cited characteristics of the neighbourhoods, to identify the optimal geographical location for this new business. Therefore, the resulting insights will be also useful for any entrepreneur interested in starting a new Indian restaurant in the city.

Data description

The focus of this analysis is New York City. The final goal is to recommend our client the best choice in terms of NYC's neighbourhoods to start a restaurant. The data embedded in the analysis are listed below. More specifically, our data science team will provide data and insight to:

- Part 1: NYC Population & Demographic characteristics

Data source:

https://en.wikipedia.org/wiki/New_York_City; https://en.wikipedia.org/wiki/Demographics_of_New_York_City.

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

- Part 2: NYC Competitors landscape

Data source: https://cocl.us/new_york_dataset and Foursquare API.

These data have been used to explore various neighbourhoods and each Indian restaurants venues in the neighbourhood.

Methodology

➤ PART 1 NYC population and demographics characteristics

Data source:

https://en.wikipedia.org/wiki/New_York_City; https://en.wikipedia.org/wiki/Demographics_of_New_York_City.

We ran an exploratory analysis with web scraping techniques, to get New York City's population density and demographics data from Wikipedia. The final goal was to recommend our client which neighbourhood in New York City is the best choice to start a restaurant.

	jurisdiction	%_white	%_black_or_african_american	%_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

Table 1 – NYC exploratory summary

According to these preliminary finding summarized in “Table 1”, Queens is the second most populous urban area in New York City (NYC), behind Brooklyn. Moreover, it's the most ethnically diverse urban area in New York City, with the highest Asian ethnic minority population.

➤ PART 2: Competitors landscape

Data source: https://cocl.us/new_york_dataset and Foursquare API.

These data have been used to explore various neighbourhoods and each Indian restaurants venues in the neighbourhood.

We gathered the neighbourhoods using FourSquare API & New York City data, which contains a list of boroughs and neighbourhoods, along with their latitude and longitude. We then defined the following functions:

- a function to interact with FourSquare API and get top 100 venues within a radius of 1000 metres for a given latitude and longitude;
- a function to get venue details (e.g. count, rating, tip counts for a given venue id), to be used for ranking;
- a function to get NYC data such as boroughs, neighbourhoods and their latitude and longitude.

As a result, we realized that there are 306 different neighbourhoods in New York City.

As displayed in “Figure 1”, we concluded that Queens is the most densely populated borough in New York City with a total of 80 neighbourhoods. Moreover, we also concluded that Queens has the highest number of Indian Restaurants (see “Figure 2”). We also noticed that Floral Park in Queens is characterized by the majority of Indian restaurants, with a total count of 11 (see “Figure 3”). Finally, as

shown in “Figure 4”, we concluded that Brooklyn has the highest average rating for Indian Restaurants.

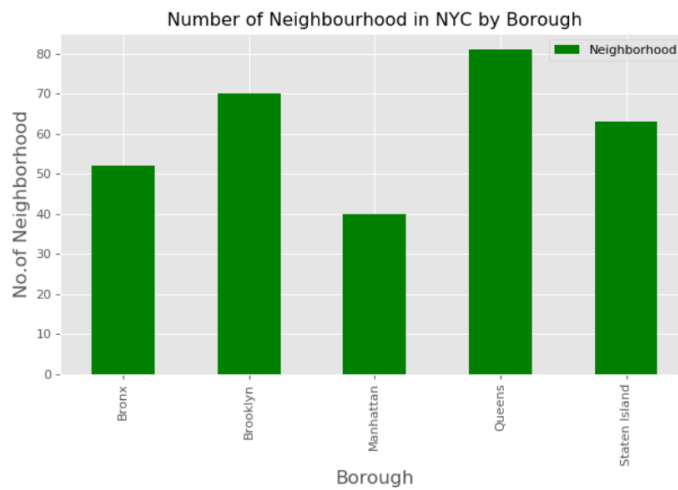


Figure 1 – Number of neighbourhoods in NYC by borough

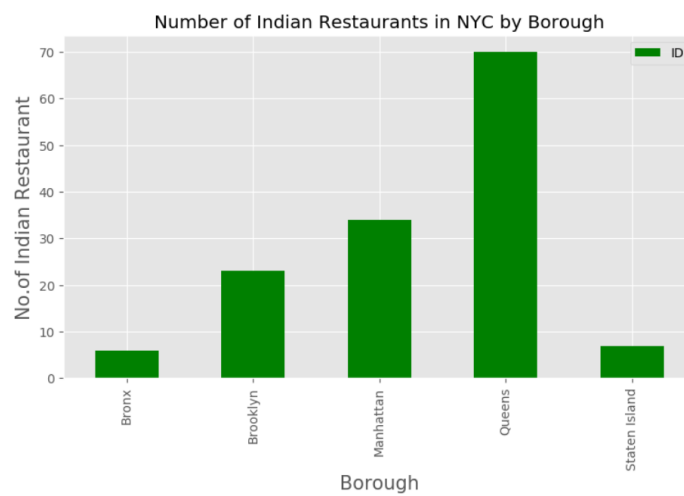


Figure 2 – Number of Indian restaurants in NYC by borough

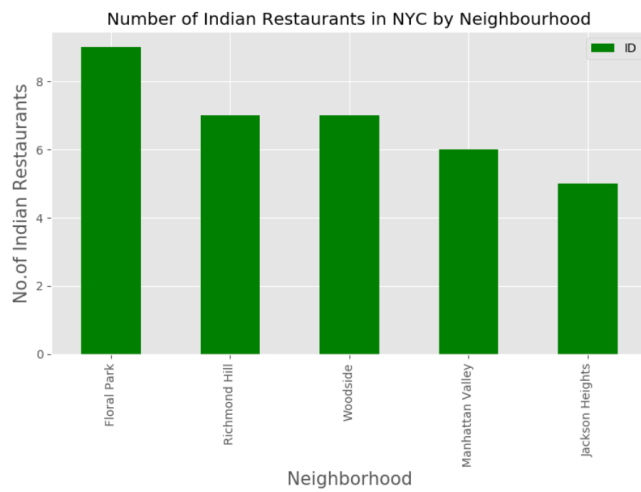


Figure 3 – Number of Indian restaurants in NYC by neighbourhood

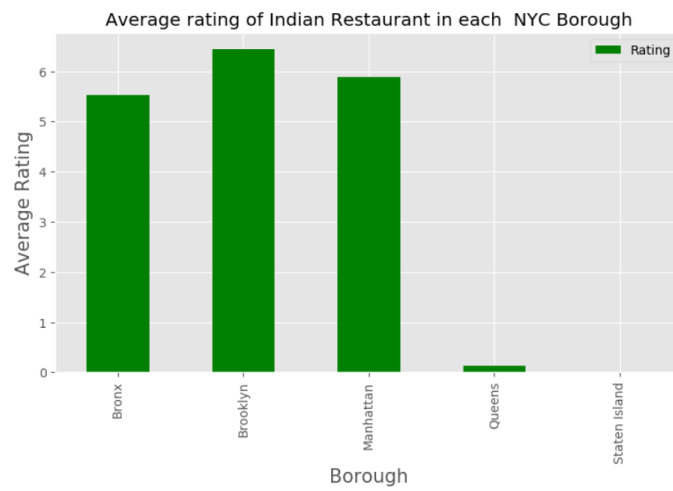


Figure 4 – Number of Indian restaurants in each NYC borough

Finally, we plotted all these neighbourhoods on a map (Figure 5”), by filtering the ones with average rating greater or equal to 8.0 (see “Table 2”).

	Neighborhood	Average Rating
0	Astoria	9.10
1	Bay Ridge	8.30
11	Central Harlem	8.10
13	Clinton Hill	8.70
17	East Flatbush	8.10
19	East Village	8.60
24	Fort Greene	8.70
28	Gramercy	8.60
42	Morningside Heights	8.00
43	Murray Hill	8.80
46	North Side	8.50
47	Ocean Hill	8.05
57	Roosevelt Island	8.40
59	South Side	8.50
66	Tribeca	9.10
72	Woodlawn	8.00
74	Yorkville	8.20

Table2 – Filtered neighbourhoods (i.e., with average rating greater or equal to 8.0)

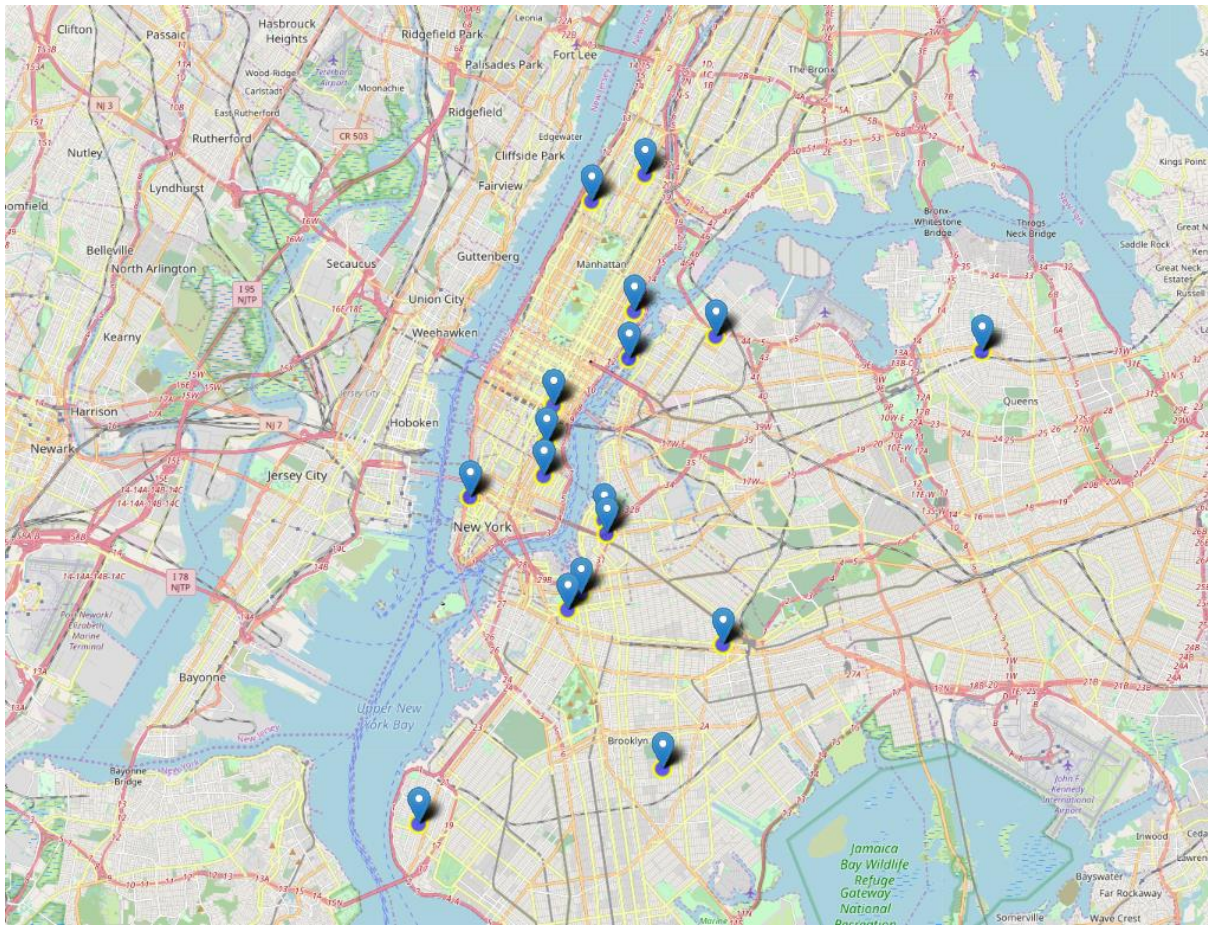


Figure 5 – NYC map

As we can see in “Figure 5”:

- Murray Hill, Tribeca, Midtown in Manhattan are some of the best neighbourhoods for Indian cuisine.
- Bronx has the lowest rated Indian restaurants in NYC.
- Manhattan is the best place to stay if you prefer Indian Cuisine.

Results (summary)

We ran a preliminary analysis, according to which Queens is the second most populous urban area in New York City (NYC), behind Brooklyn; moreover, it's the most ethnically diverse urban area in New York City, with the highest Asian ethnic minority population.

We then analyzed the data more in depth and we realized that there are 306 different neighbourhoods in New York City, with the following highlighted characteristics:

- Queens is the most densely populated borough in New York City with 80 neighbourhoods
- Moreover, Queens has the highest number of Indian restaurants.
- Floral Park in Queens has the most Indian restaurants with (11 in total).
- Brooklyn has the highest average rating of Indian restaurants.

Finally, we visualized on a map all the neighbourhoods with average rating greater or equal 8.0, to visually check what are the best neighbourhoods in New York that has highest average rating for Indian Restaurants, with the following highlights:

- Murray Hill, Tribeca, Midtown in Manhattan are some of the best neighbourhoods for Indian cuisine.
- Bronx has the lowest rated Indian restaurants in NYC.
- Manhattan is the best place to locate for people who prefer Indian Cuisine.

Discussion (highlights by area)

- Queens: area with highest density of population borough in New York City (i.e., 80 neighbourhoods), with the highest number of Indian restaurants (11 in total only in Floral Park).
- Brooklyn: area with the highest average rating of Indian Restaurants.
- Bronx: area with the lowest rated Indian restaurants in NYC.
- Manhattan: Murray Hill, Tribeca, Midtown are the best Manhattan's neighbourhoods for Indian cuisine.
- Manhattan is the best place for people who prefer Indian Cuisine.

Conclusion

Based on all these results, we recommend Midtown in Manhattan as the best location in NYC for "The Elephant", considering that it represents:

- the third most populous urban area in New York City.
- an area with population density of 27,826 people per square km, highest of any borough in the United States.
- an area with some of the top-rated Indian restaurants and thus an existing customer base and interest toward this cuisine.
- an area with the second highest Asian ethnic minority population in NYC.