# Attention Entrepreneurs, Predictors of Locations of Indian Restaurants in New York City

### Introduction

#### Background

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014. While English remains the most widely spoken language, there are areas in the outer boroughs in which up to 25% of people speak English only as an alternate language.

Indians are the largest South Asian group, comprising 2.4% of the city's population, among the largest in the US cities. New York city also sees a lot of tourists consuming at local business. At 65.2 million the number of tourists is many times the population of the city at 8.1 million. New York city also sees a lot of commute from outside the city. Tourists and commuters indicate that residential data alone might not be a good indicator of the consumption patterns in the city. All of these factors make New York city a unique problem for data analysis.

Indian are highly popular in the united states. These restaurants attract a diverse set of people, and not just Indian origin people. This suggests factors other than just ethnicity of the people wouldn't dictate the presence of an Indian restaurant. Determining the reason for the presence of an Indian restaurant would open up data savvy entrepreneur to find neighborhoods to open an Indian restaurant.

Business Problem: Determine the variable that can predict the presence of Indian restaurants in NYC

### Data

#### Data Sources

The data used would two sets of categories the Target and the parameters.

**Target data**

* NYC neighborhood longitudes and latitudes: https://cocl.us/new\_york\_dataset
* No. of Indian restaurants in New York City boroughs and neighborhoods: Foursquare API

**Parameter data**

* People of Indian Origin in New York City boroughs: <https://www.baruch.cuny.edu/nycdata/population-geography/foreign-birthcountry.htm>
* General population statistic data: <https://guides.newman.baruch.cuny.edu/nyc_data/nbhoods>
* No. of Chinese restaurants in New York City boroughs and neighborhoods: Foursquare API

#### Data Extraction and Exploratory Data analysis

A loop is used to cycle through every neighborhood longitude and latitude. In this loop, foursquare API search calls are given, with venue category for Indian restaurants in a radius of 2 km. Restaurant details are discarded but only the values o number of Indian restaurants are retained. This data is put back into the data frame with neighborhoods as a column. This data is used to construct a map in folium. Figure 1 shows the number of Indian restaurants in each NYC neighborhood. The size of the marker indicates the number.

The no. of Indian restaurants is aggregated for every borough from their neighborhoods. The number of Indian restaurants in every borough are shown in figure 2. Most of the Indian restaurants are in Manhattan. Brooklyn and Queens have the next most number of Indian restaurants.



Fig 2. The number of Indian restaurants in every New York city borough

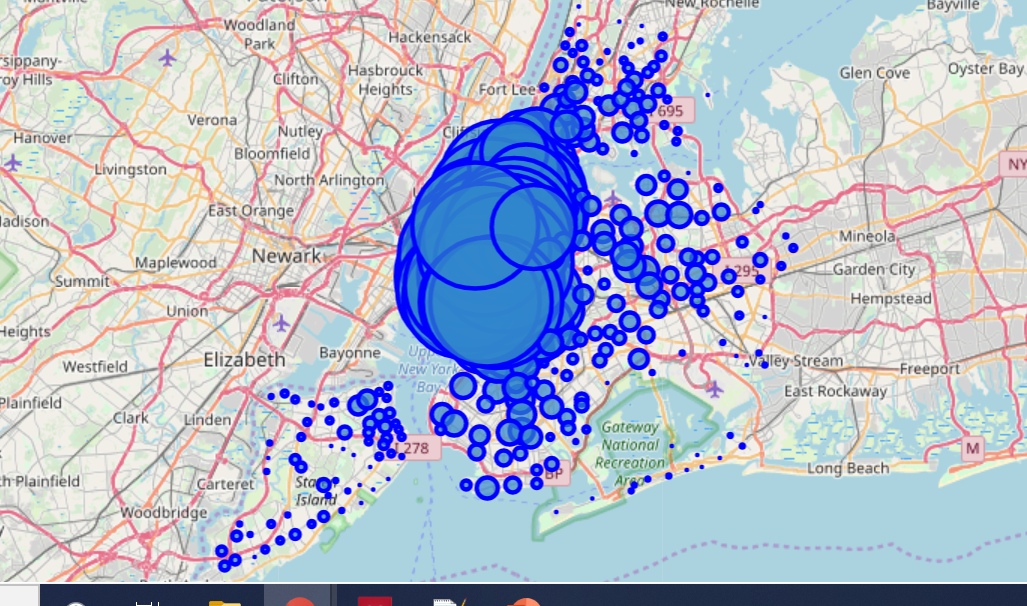


Fig 1. No. of Indian restaurants in every NYC neighborhood

The Indian origin population is plotted in figure 3. The population is in figure 3(a) the population as a percentage of the whole population is in figure 3(b). From the plots, it can be deduced that most Indian origin people in NYC live in Queens. Much less no. of Indians live in other boroughs. Queens also contains the most no. of Indian origin people by fraction of the total borough population; ~ 4% of the total population. Indian as a fraction of the population in Manhattan and Staten Island is also large: >3% of the total population

For initial exploration the figures 2 and 3 are compared. While the most no. of Indian restaurants are in Manhattan, most no. of Indians live in Queens. In Queens, the no. of Indian restaurants are only 3rd highest among all boroughs.

|  |  |
| --- | --- |
| (a) | (b) |

Fig 3. The Indian origin population in New York City by borough. (a) Population and (b) Population percentage

The same procedure as before is followed to extract the no. of Chinese restaurants in every New York City neighborhood. The results are plotted in folium map as before and shown in in figure 4.



Fig 4. No. of Chinese restaurants in every NYC neighborhood

### Methodology

The objective of the methodology is to find correlation between number of Indian restaurants and any of the parameters, in NYC neighborhoods. Regression techniques are used. Linear and polynomial fit is used for each of the parameters. The fitting tools used are form Scikit learn libraries. The value is used to determine the quality of the fit.

Since the Indian population data is available only for boroughs, the fit is performed with number of Indian restaurants per borough. The regression fit for number of Indian and Chinese restaurants is performed for each neighborhood.

### Results and Discussion

|  |  |
| --- | --- |
|  |  |

Fig 5. The linear regression to obtain the No. of Indian restaurants from (a) The Indian origin population; (b) The percentage of Indian origin population

Figure 5 shows the linear fit between number of Indian restaurants and Indian population and population percentage per NYC borough. The value for both the parameters indicate a poor fit, as expected from the initial data exploration. This indicates that there is no correlation between the Indian population and Indian restaurants.

Figure 6 shows the linear fit between the number of Indian restaurants and the number of Chinese restaurants in every neighborhood. The neighborhoods with 50 restaurants is filtered out from the model, because it might indicate that the limit of 50 search results in folium is reached. The value for this fit is 0.33520. While this value doesn’t indicate a good fit, it can imply that some correlation between the number of Indian and Chinese restaurants exist.



Fig 6. linear regression plot to obtain the No. of Indian restaurants from the number of Chinese restaurants

This plot shows some useful insights for entrepreneurs. The data points below the line in figure 6 indicate that there are disproportionately more Chinese restaurants and Indian restaurants in those neighborhoods. This presents a business opportunity to open more Indian restaurants. The neighborhoods with data points above the line indicate that the number of Indian restaurants are disproportionately high in those neighborhoods. The recommendation for existing businesses in these neighborhoods is to close down a few Indian restaurants.

### Conclusion

Indian origin population is not related to no. of Indian restaurants. No. of Indian restaurants is better correlated to no. of Chinese restaurants. Neighborhoods to open Indian restaurants in NYC are identified.

### Future Directions

Other demographic data, commuter and tourist information for city like NYC is very important for analysis in the future. Tourists and commuters are important consumers in New York. Considering these would improve our analysis and modeling accuracy.