**Selecting a city in US where and Indian restaurant to be opened.**

**Mohan Kakarla**

**22nd August 2020**

1. **Introduction**
   1. **Background**

One of the Indian Restaurant Chain owner having many restaurants in many Indian cities wants to open a restaurant in a city of US. How a Restaurant business runs well in a foreign country is based on various parameters like what is the ethnic mix in the city and percentage of Indian population who are not getting Indian food when they are away from the country they are born. Additionally one should open restaurant at a place where many other restaurants are already available and is an attraction for those who crave for the delicacy. Additionally if more Indians are already there in the city then relatives will also come for tours, when the host will take them out for a meal at least once during their staying.

* 1. **Problem**

Available data of Cities and restaurants will provide valuable information to decide on the city and neighbourhood where and Indian restaurant will work and flourish. This project will predict the City and neighbourhood where the stakeholder should open his Indian restaurant.

* 1. **Stakeholders**

Restaurant owner is the first stakeholder who will be interested to know where he should open the restaurant. Indians who are already staying for years are also interested as they will be having the delicacies of their motherland in US. Additionally Tourists are also very much interested as they want to taste their delicacies and enjoy. At time locals also want to taste the delicacies available in their neighbourhood.

1. **Data acquisition and cleaning**
   1. **Data sources**

US has 52 states and 300+ plus cities having more that 100K population, project is to evaluate various cities and its Metropolitan Statistical Area with Indian American Population. We will select the city having highest Indian American population where the client can open the restaurant.

First data set is of the list of US cities where most Indians stay having the count of Indians and Total population and Percentage of Indians out of the total population.

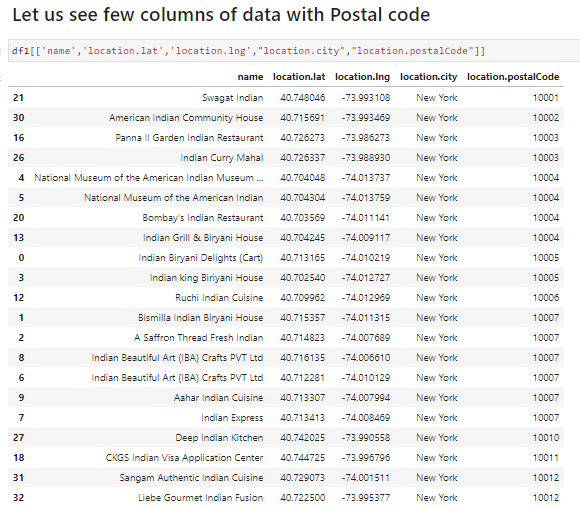
We need to select the city where the Indian American population is most across all cities of US by fetching the city with most Indian American population.



Second dataset that I will be using is the Foursquare data, was used to shortlist the neighbourhood where the restaurant is to be opened by knowing how many restaurants are there in each neighbourhood.

But we can see below that we do have neighbourhood data for all neighbourhoods.



Hence alternatively we can shortlist the area where stakeholder can open his indian restaurant with the help of Postal code as given below.

After fetching counts by Postal code, we conclude that Postal code 10007 is the area where maximum Indian restaurants are there in New York.



1. **Methodology**

In this project we will direct our efforts on detecting areas of New York that have high restaurant density, particularly those with high number of Indian restaurants. We will limit our analysis to area ~10km around city center.In first step we have collected the required data: location and type (category) of every restaurant within 10km from New York center. We have also identified Indian restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of ‘restaurant density’ across different areas of New york - we will use heatmaps to identify a few promising areas close to center with high number of restaurants in general and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with more than ten restaurants in radius of 500 meters, and we want locations with Indian restaurants in radius of 800 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final ‘street level’ exploration and search for optimal venue location for the stakeholder.

1. **Results and discussion**

Data analysis from Foursquare present that even though there are great number of restaurants in NewYork, there are areas of high density close to city center. Highest concentration of restaurants was detected in and around Postal code 10007 which is of MSA name as New York-Northern New Jersey-Long Island, NY-NJ-CT-PA.

Hence recommending stakeholder to open his Indian restaurant around this area which will provide flourishing business.

1. **Conclusion**

Purpose of this project was to identify which city in USA is having highest Indian population and we found that New York is having the highest Indian population. Then we identified New York areas in the city center having high volume of foot fall which will enable flourishing business for the restaurant to be opened. This project identifies the stakeholder where to open Indian Restaurant. By calculating restaurant density distribution from Foursquare data we have shortlist existing density of restaurants where many locals, Tourists and office goers visit to have a meal.

**Concluding that the Stake holder to open the Indian restaurant in Postalcode 10007 as it has maximum number of Indian restaurants with MSA name as New York-Northern New Jersey-Long Island, NY-NJ-CT-PA**

Final decision of where to open Indian restaurant is with the stakeholder depending on his Capital budget which he is willing to invest.