**Coursera Capstone Project: Applied Data Science**

1. **Introduction**

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, accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States. New York City has also been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, and French etc. So as part of this project, we will list and visualize all major parts of New York City that has great Indian/Pakistani restaurants.

1. **Business Problem**

People from Pakistan working in New York need to eat breakfast, lunch, dinner and snacks and that food needs to be halal for People from Pakistan, in this project we will:

* List and visualize the best foods spots offering Pakistani food
* Find best location for Pakistani Food
* What areas have most potential in offering Pakistani food?
* Which areas don’t offer Pakistani food?
* What should be the best place to live for a Pakistani considering halal food.

1. **Data**

The data for this project has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used. The Neighborhoods data of the neighborhoods in New York can be extracted out by web scraping using BeautifulSoup library for Python. The neighborhood data is scraped from a Wikipedia webpage.

For this project we need the following data:

* New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.
* Data source : <https://cocl.us/new_york_dataset>
* Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City
* Indian restaurants in each neighbourhood of New York City.
* Data source : Foursquare API
* Description: By using this API we will get all the venues in each neighbourhood. We can filter these venues to get only Indian restaurants.
* GeoSpace data
* Data source : <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>
* Description: By using this geo space data we will get the New York Borough boundaries that will help us visualize choropleth map.

**Methodology:**

1. We begin by collecting the New York city data from the following link "[https://cocl.us/new\_york\_dataset"](https://cocl.us/new_york_dataset%22)
2. We will find all venues for each neighbourhood using Foursquare API
3. We will then filter out all venues with Indian restaurant for further analysis.

1. Next using Foursquare API, we will find the Ratings, Tips, and Number of Likes for all the Indian Restaurants.



1. We will then sort Neighbourhoods and Borough the data keeping Ratings as the constraint.





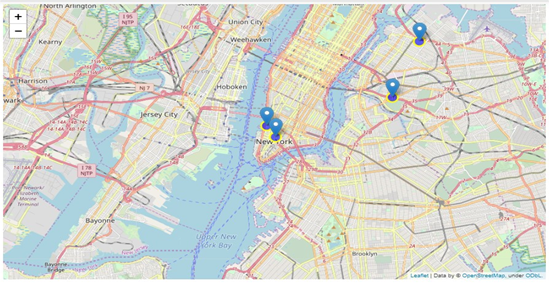
1. Next we will consider all the neighbourhoods with average rating greater or equal 9.0 to visualize on map.



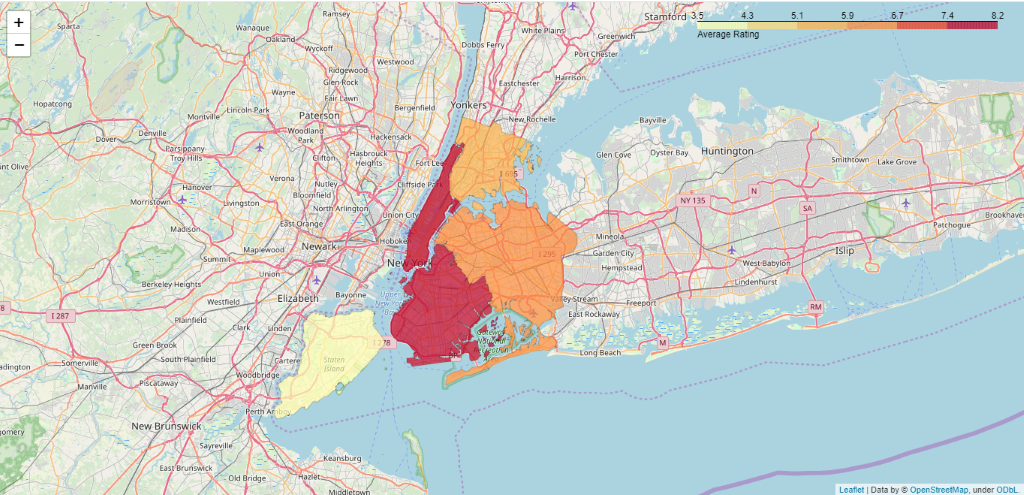
1. We will join this dataset to original New York data to get longitude and latitude.



1. Finally, we will visualize the Neighbourhoods and Borough based on average Rating using python’s Folium library.

Neighbourhoods based on average rating:

Borough based on average rating:



**Result:**

So now we can answer the questions asked above in the Questions section:

Answers:

1. The following location in New York City has great Indian restaurants.



#### Astoria (Queens), Blissville (Queens), Civic Center (Manhattan) are some of the best neighbourhoods for Indian cuisine.

#### Manhattan have potential Indian Restaurant Market.

#### Staten Island ranks last in average rating of Indian Restaurants.

#### Manhattan is the best place to stay if you prefer Indian Cuisine.

**Conclusion:**

There is always room for improvement and hence the above solution I have provided can also be improved for best results depending upon the data we have.