

# **IBM Data Science Professional Certificate**

## **Applied Data Science Capstone Project The Battle of Neighborhoods**

**Submitted by**

**Annadurai Palanisamy**

### **Declaration: -**

I hereby declare that I have completed my training (IBM Data Science) from Coursera. I have declared that I have worked with full dedication during this training and my learning outcomes to become a Data Scientist.

## **Acknowledgement: -**

Firstly, thanks to Coursera for providing the opportunity to learn about Data Science and work on a Capstone Project. I am deeply grateful for their help & valuable guidance to this project.

I would extend my gratitude to the Instructor Alex Aklson who explained in the videos very detailed way.

Apart from my efforts, the success of my project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this training.

Thank you...

# Introduction

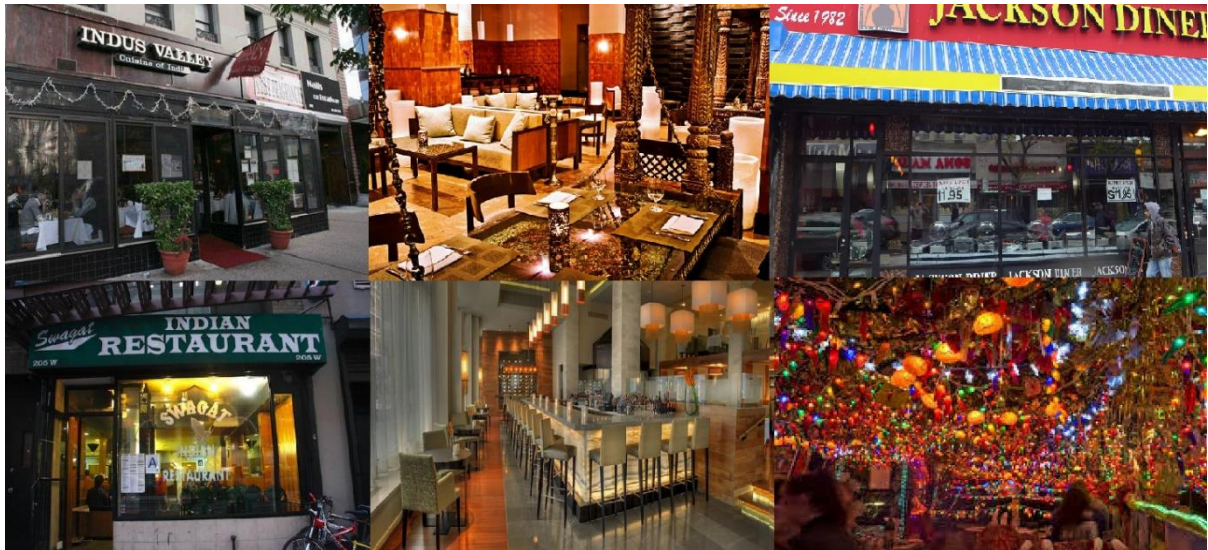
New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014, accounting for over 40% of the population of New York State 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.



Throughout its history, New York City has 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, French etc.

So as part of this project, we will list and visualize all major parts of New York City that has great Indian restaurants.



## Indian Restaurants in New York City

### Data

Data For this project we need the following data :

New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude. Data source : [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of New York city. Indian restaurants in each neighborhood of New York city. Data source : Fousquare API Description : By using this api we will get all the venues in each neighborhood. 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.

### Approach

- Collect the New York city data from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- Using FourSquare API we will find all venues for each neighborhood.
- Filter out all venues that are Indian Restaurants.
- Find rating , tips and like count for each Indian Restaurants using FourSquare API.
- Using rating for each restaurant , we will sort that data.
- Visualize the Ranking of neighborhoods using folium library(python)

## **Open Source tools for Data Science**

In this course, I have learned about various open source tools for Data Science.

- Skill Network Labs
- Jupyter Notebooks
- IBM Watson studio

## **Applied Data Science Capstone**

In this course I have learned about FourSquare API ( It is a restful API to retrieve the data about venues in different neighborhoods around the world and I have applied this learnings to complete my Capstone Project.

## **Conclusion**

- Astoria(Queens), Blissville(Queens), Civic Centre(Manhattan) are some of the best neighborhoods 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.
- Limitations
- The ranking is purely on basis of rating of restaurants
- The accuracy of data depends purely depends on the data provided by FourSquare

## **Limitations**

- The ranking is purely on basis of rating of restaurants
- The accuracy of data depends purely depends on the data provided by FourSquare