

# Coursera Capstone

## IBM Applied Data Science Capstone

*Are you thinking about opening a new Indian Restaurant near New York City (NYC)?*

*By: Pronab Das*

*July 2020*



## Introduction

New York City, world's major commercial, financial and cultural center. It has been a center for trade and economic growth. New York is known worldwide as a cultural melting pot. New York City Metropolitan Area contains the largest metropolitan Asian Indian population in the Western Hemisphere. According to the 2018 Asian American Federation, New York City is home to approximately 650,000 people from the Indian subcontinent and with its diverse culture, comes diverse food items.

Restaurant is a place where people come to have food and drinks for a cost. Restaurants are a notoriously difficult business to own or operate. Not only is it a relatively capital and employee-intensive business, restaurants are also highly regulated, low margin and in most cases have a plethora of competition to deal with. There are more dining establishments and more diners now than during any other time in history; the National Restaurant Association reports over 1 million restaurant locations in the U.S. alone, a boon considering more than half the American population visits at least one of them each week. To fill a niche and remain relevant, prospective, as well as established, restaurants must hedge their bets with well-rounded and well-directed dataset. Data

science provides valuable insights regarding market trends and evolving consumer lifestyles so that restaurateurs can better address and meet public demand.

## **Business Problem**

My client wants to open an Indian restaurant in this one of the world's most diverse region. No single restaurant is ever going to appeal to everyone. Some people like quiet, intimate settings; others prefer boisterous ones. Some people want to bring their kids along; others want to dine alone. The preferences are as varied as the possible offerings, with specific generational cohorts preferring one thing, "people who like Indian food" preferring another and everyone else liking a million other things in between.

So, what are all the factors we must consider determining which neighborhood will have best potential to open an Indian restaurant?

Population, Competition in the neighborhood, Surrounding amenities like parks, visiting sites, entertainment, Market Places, Menu items and so on...Our solution needs to be data driven for avoiding or considering low risk criteria and high success rate and thus applying data science techniques and methodologies would lead us to take correct decision.

## **Target Audience**

- Business personnel who wants to invest or open a restaurant
- Freelancer who loves to have their own restaurant as a side business
- Small business owner finding the best location for opening a restaurant
- Budding Data Scientists who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it

## Data

We will be analyzing data and provide insight as per below to the client to determine which neighbourhood in New York City will be best choice to start a restaurant.

- *NYC Population & Demographic characteristics*

*Data sources:*

Population – [https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)

Demographic Characteristics -  
[https://en.wikipedia.org/wiki/Demographics\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Demographics_of_New_York_City)

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

- *Who are the competitors in that location?*

*Data source:*

[https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

New York City data that contains Borough, Neighborhoods along with their latitudes and longitudes. These datasets will be used to explore various neighborhoods and each Indian restaurants venue in the neighborhood

*Data Source:*

[Foursquare API](#)

By using this API we will get all the venues in the neighborhoods. The data going to be collected/acquired from the Foursquare API about the various restaurants in each neighborhood of New York city especially Indian cuisine, which will be used for acquiring the information regarding same for all the venues of each neighborhood.

## Approach

- Acquire the data from the various sources as mentioned above
- Filter the data to extract only Indian Cuisine related data acquired from Foursquare
- Acquire the tips and get the rating of every restaurant in the neighborhood of New York
- Analyze the data using Clustering
- Visualize the findings using Matplot library and Folium for better understanding and data insight
- Compare the Neighborhoods to find the best place for starting up a restaurant
- Inference From these Results and related Conclusions