**Coursera Capstone Project**

**IBM Data Science Specialization**

**Opening an Indian Cuisine Restaurant in New York City, USA**

**By: Tushar Singhal**

**August 2020**

**1. Introduction**

**1.1 Background**

While opening a restaurant can be a very lucrative business, a lack of demand causes many restaurants to close within the first year of opening. There are many different factors that can account for a restaurant’s success such as location, competition and quality of the food. This is an important question that every business owner must face when choosing whether to open a restaurant or not, as well as location of the business. To demonstrate the process of picking a location for a client opening a business, the project will focus on answering were to open the restaurant. If there are too many Indian Restaurants in the local vicinity, the profitability of the restaurant will be severely decreased. Additionally, starting a restaurant in a location with higher income would increase the profitability of the business over starting in a poorer area.

**1.2 Business Problem**

The following question: “If the client wanted to open an Indian Restaurant in New York City, what areas are the best options to open the restaurant?” For an Indian Restaurant, the location and competition are both determined by where the restaurant is opened.

**1.3 Target audience of this project**

This project is mainly useful to an anyone who wishes to open an Indian Cuisine Restaurant in New York City. The insight from the project will be helpful for determining the best possible location of the restaurant. It will help it understanding whether there is a lot or little competition in a given neighbourhood of the city. Accordingly, the person who is interested in opening it can take the necessary decision with the help of the gathered intel.

**2. Data**

To solve the problem, the following data is needed:

* List of neighbourhoods in New York City. This defines the scope of this project which is confined to the city of New York.
* Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
* Venue data.

Sources of data and methods to extract them,

The data for the neighbourhoods of New York is obtained from previous week of the course. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data in order to help us to solve the business problem put forward. This is a project will make use of many data science skills, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).