**Applied Data Science Capstone by IBM/Coursera**

**Avani Bhatnagar**

**August 11, 2021**

**Introduction**

Location is very important when it comes to opening a restaurant. If a family wants to open up a restaurant in a city, they need to know different things about the location to be able to strategically select the correct neighborhood which will attract the most crowd, is affordable to open up a restaurant at, has lesser competition, population etc. In this situation, a family living in New York decides to open an Indian restaurant in the city. They are in the very initial stages of planning this business and one of their first points of focus is deciding the location of the restaurant. They know a lot of people enjoy Indian food, but for the most profit they need to able to find something that will also consider factors such as population, relative competitiveness in that area, etc.

Their goal is find maximum success while putting in least risks since they are using a lot of their family savings into this!

**Data Acquisition and Cleaning**

The Data is the most important part of any data science/analysis project. We need to make sure we have all the relevant data before we can dig into the "science" part of the process.

* New York Data: We use this dataset: <https://en.wikipedia.org/wiki/New_York_City> to get information about the various New York neighborhoods which will help us explore and decide the best location for an Indian restaurant.
* Foursquare API: We use Foursquare API to get various restaurants in New York such as Indian Cuisine. This will help the business get an idea of what's the best location to open their restaurant in terms on relative competition in the same business.
* Geospatial data - We use <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm> to get the location details of New York - it gives us the Latitude and Longitude values for each Borough in New York. This will be greatly helpful in generating maps and visualizing the overall results.
* Demographic data - We use <https://en.wikipedia.org/wiki/Demographics_of_New_York_City> to get the demographic data for NYC's jurisdiction.

**Exploratory Data Analysis**

We start by looking at all boroughs of NYC with their population, gdp per capita and persons per square mile.

Table

Description automatically generated

Then we visualize it to get an idea about the population density of each borough

Chart, bar chart

Description automatically generated

Then we look at the demographics of all the boroughs to get an idea of what % of people live there are Asian since we are concerned with Indian Cuisine.

Graphical user interface

Description automatically generated with medium confidence

**Solution/Finding the best Neighborhood**

In the process of finding the best neighborhood we visualize all boroughs by number of neighborhoods to get a population density of each.

Chart, bar chart

Description automatically generated

Then we use the Foursquare API to get the Indian Restaurants in each neighborhood.

Chart, bar chart

Description automatically generated

Then we combine data sets to get a list of Neighborhoods with their boroughs and Indian Restaurant name, likes, and ratings.

Graphical user interface, text, application

Description automatically generated

We then find the restaurants with the highest ratings.Table

Description automatically generated

We group them by boroughs to get average rating for each borough.

Chart, bar chart

Description automatically generated

We visualize the neighborhoods with highest Indian restaurant ratings to help our user visualize the likely spots to open an Indian restaurant.

Map

Description automatically generated

**Conclusion**

From all our analysis and results & findings, we can come to a conclusion that Manhattan would be a great place to open an indian restaurant since it has the highest population density and has the 2nd highest Asian population which will attract a large crowd. It also has some of the highest rated Indian restaurants so people probably stay in that area when looking for Indian food.

Another conclusion is that Queens can also be a good option if they want to be close to other Indian restaurants since it has the highest Asian population and it also has highest number of Indian restaurants.

This is a basic data science analysis. We can do further analysis using other algorithms in the future