IBM Data Science Professional Course

Coursera Capstone Project

The Battle of Neighborhoods

Starting a new Indian Restaurant in Delhi, India

By – Mukta Prafulla Sapre

July 2020

Introduction:

Delhi is a union territory in India containing New Delhi, a capital of India. According to Census 2011, Delhi's city population is about 11 Million which second-highest in India. Delhi's urban area has a 2016 estimated population of over 26 Million which is world's second-largest urban area. According to Euromonitor International, Delhi ranked as 28th-most visited city in the world and first in India by foreign visitors in 2015. Therefore, there are already many Restaurants, Hotels, Coffee Shops etc. They all have different types of food from different parts of the country and also world. Increasing population and tourism are increasing the demand for good quality Restaurants.

According to Indian Medical Science (Ayurveda), we should eat only local food for good health. Understanding Indian culture is actually incomplete without even tasting the Indian food. The tourists mostly from foreign countries always want to taste Indian Regional food. Many people have to avoid foreign foods to maintain good health. Regularly eating continental food is not hygienic. It also creates so many economic problems. The people should always prefer Local food. Now days, due to Covid-19 pandemic, people are realising the importance of

eating home-made or local food. The situation between India and China is rapidly lowering the Chinese market in India. The Chinese food has occupied the large part of food industry in India. Hence, preferring to eat Indian food is the best solution to make the country self-dependant.

Actually it is very complicated process to open a Restaurant serving the Indian food from all parts of India. Because, there is already an oversupply of such Restaurants in Delhi as it is a capital of India. But finding a place for such Restaurant is very important step. It will decide whether the Restaurant will run successfully or not.

Business Problem:

The aim of the project is to analyze and find area in Delhi, India where a new Indian Regional Food Restaurant can be started and run successfully. Actually there is oversupply of Restaurants, Hotels and food options in Delhi. There are many restaurants serving Indian food. But as mentioned earlier, increasing the number of such local food restaurants is very important to raise Indian economy and make India self-dependent.

Using various types of Data science and Machine Learning techniques and mainly Foursquare API, the project provides a solution to the Business Problem, '**Where to start new Indian Regional Food Restaurant in Delhi, India?'**

Interest:

This project is useful to the businessmen, investors and developers who want to start a new Indian Restaurant in Delhi, India. As Delhi is capital of India, there are varieties of food Restaurants. It is required to find area in Delhi with very less number of Indian food Restaurants. My project will help Businessmen to start and develop their Business in the Hoteling sector.

Starting an Indian Restaurant will increase the demand of raw food materials that produce in remote areas of country. Indian farmers will get their productions for sell in market. It will definitely help to solve a problem of unemployment. And therefore, it will automatically raise the Indian economy.

Data:

For this project, following data is necessary:

- The list of suburbs (neighborhoods) in Delhi, India. This data is the main dataset of the project which decides the scope of the project. I used the data on website **https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi**
- The latitude and longitude coordinates of each neighborhood should be found using **Geocoder**. This helps to draw maps with neighborhoods superimposed on top.
- The venues data particularly related to Indian Restaurants from **Foursquare**. It will be used for clustering of neighborhoods.

How the data can be used to solve the problem?

The Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi) contains the list of 138 suburbs (neighborhoods) in Delhi, India. The data should be extracted and transformed into dataframe. This can be done using **BeautifulSoup** in Jupyter Notebook. The latitude and longitude coordinates of all the neighborhoods can be found with the help of **Geocoder**.

The Foursquare has access firmographic data and rich community-sourced content for more than 60 million commercial places around the world. It has data of 62M+ global venues, 190+ countries, 50 territories, 900+ venue categories, 30+ attribute fields and more than 150K developers on platform. Using **Foursquare,** all the venues in all the neighborhoods and their categories can be listed. Then the data with venue category 'Indian Restaurant' should be extracted. It will help to create clusters.

The data should be visualized with the help of maps using **Folium**. Finally, using K-means clustering, the data will be transformed into a specified number of clusters that will help to understand results and conclusion of the analysis.

Methodology:

I have already discussed the data I used in this project, sources where I found it and how it can be used to solve the problem. I used the Python Jupyter Notebook in IBM Watson Studio to extract all the data and analyse.

The data required to work on this project is available on the Wikipedia page: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi .In Jupyter Notebook, the data was extracted and transformed into dataframe using BeautifulSoup. The dataframe contains the column 'Neighborhood'. There are 138 neighborhoods in the data. Then the latitude and longitude coordinates for each neighbourhood was found using Geocoder. Then using these values, the map of Delhi was drawn with all the neighborhoods superimposed on top with coloured circles with the help of Folium. Using Foursquare API, venues respective to each neighbourhood was found. I had to register on Developer portal on the website: Then I had to create an app and copy the Client ID, Client Secret and Version in Notebook to get access. I made calls to get 100 venues in 2000 meters radius. Then I extract the data into dataframe with columns neighbourhood, venue-name, venue-category, latitude and longitude. Checking how many venues are there for each neighbourhood, I also found that there are 70 unique

categories of venues. Then the dataframe was created to measure number of venues in each neighbourhood but this time category-wise. The rows were grouped by neighborhood and by taking the mean of frequency of occurrence of each category. The dataframe containing venue category ‘Indian Restaurant’ was separated.

K-Means clustering and some further analysis has discussed in Exploratory Analysis.

Exploratory Analysis:

K-means clustering algorithm is one of the simplest unsupervised learning algorithms that solve all the well-known clustering problems like the problem of this project. The algorithm classifies the data into K centres which are as small as possible. I transformed the data into 3 clusters based on the frequency derived. It helps to identify the neighborhood with higher concentration of ‘Indian Restaurant’ and lower concentration of ‘Indian Restaurant’. The result allows determining the area where the number of ‘Indian Restaurant’ is to be increased i.e. which area is convenient and suitable to start a new Indian Restaurant.

Then I had to determine which cluster has less number of Indian restaurants, which cluster has moderate number of Indian restaurants and which cluster has very high number of

Indian restaurants. Visualizing these clusters with the help of map is really very simple way to understand the results. The area where the new Indian Restaurant can be started will be the area with lower number of Indian restaurants.

After classifying the data into clusters and completing the process of K-Means clustering, I separated the data with venue categories that are related to food restaurants i.e. Indian Restaurant, Pizza Place, Fast Food Restaurant, American Restaurant, Middle Eastern Restaurant, Chinese Restaurant, Sandwich Place, Hotel, Diner, Food Truck, Thai Restaurant, Italian Restaurant, Salad Place, Vegetarian / Vegan Restaurant and Food Court. Calculating total number of Restaurants in each category, following dataframe was found. From this dataset, Bar Graph and Pie Chart were drawn.

Category	Total
Indian Restaurant	145
Pizza Place	182
Fast Food Restaurant	94
American Restaurant	1
Middle Eastern Restaurant	27
Chinese Restaurant	41
Sandwich Place	42
Hotel	121
Diner	27

Food Truck	26
Thai Restaurant	27
Italian Restaurant	27
Salad Place	17
Vegetarian / Vegan Restaurant	18
Food Court	13

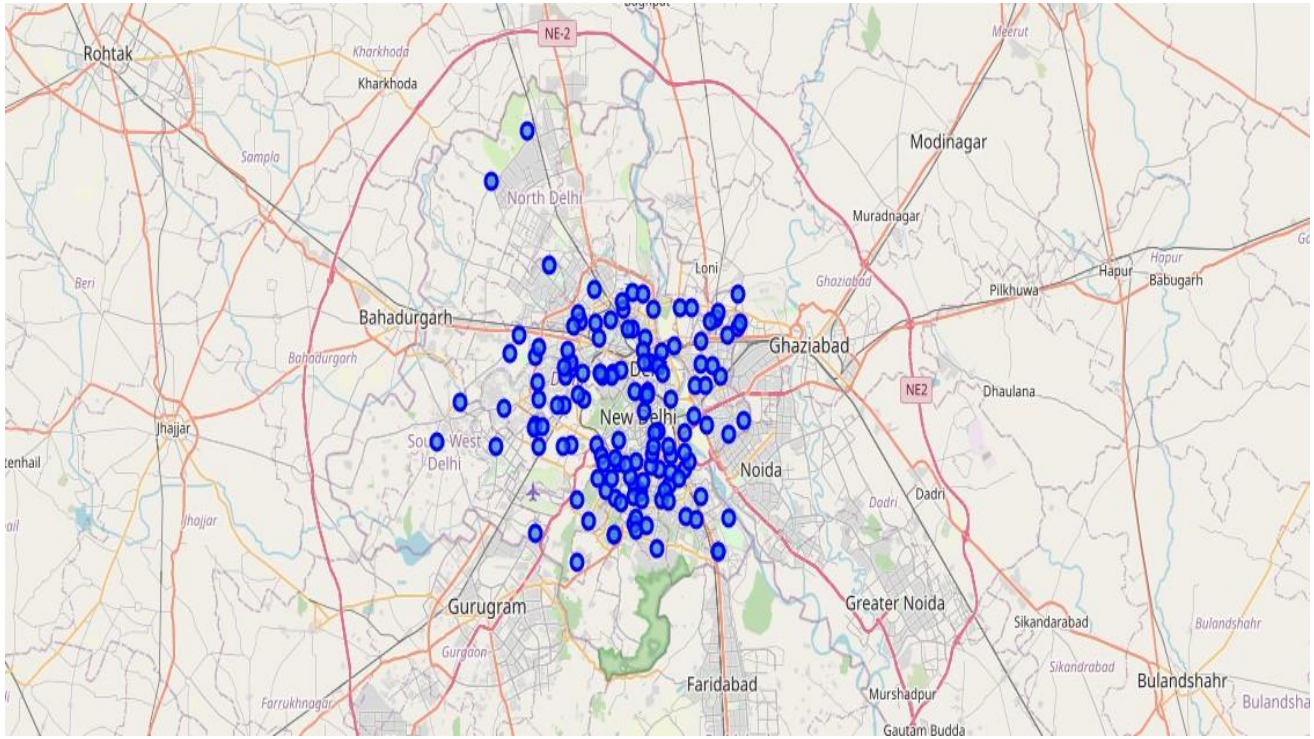
After this, the above categories were classified into three types that are Indian, Foreign and Both. ‘Indian’ type includes the Restaurants that serve only Indian Food from different parts of India. ‘Foreign’ type includes the Restaurants that serve only foreign food from different countries in the world. ‘Both’ includes the Restaurants that serve Indian food as well as foreign food. I calculated the number of Restaurants in each type and also its percentage. This constructs following dataframe. The Bar Graph and Pie Charts were drawn using following tables.

Types of Food	Number of Restaurants
Indian	145
Foreign	458
Both	205

Types of Food	Percentage
Indian	17.94554455
Foreign	56.68316832
Both	25.37128713

Results:

Following picture shows the map of Delhi with all the neighborhoods superimposed on top.



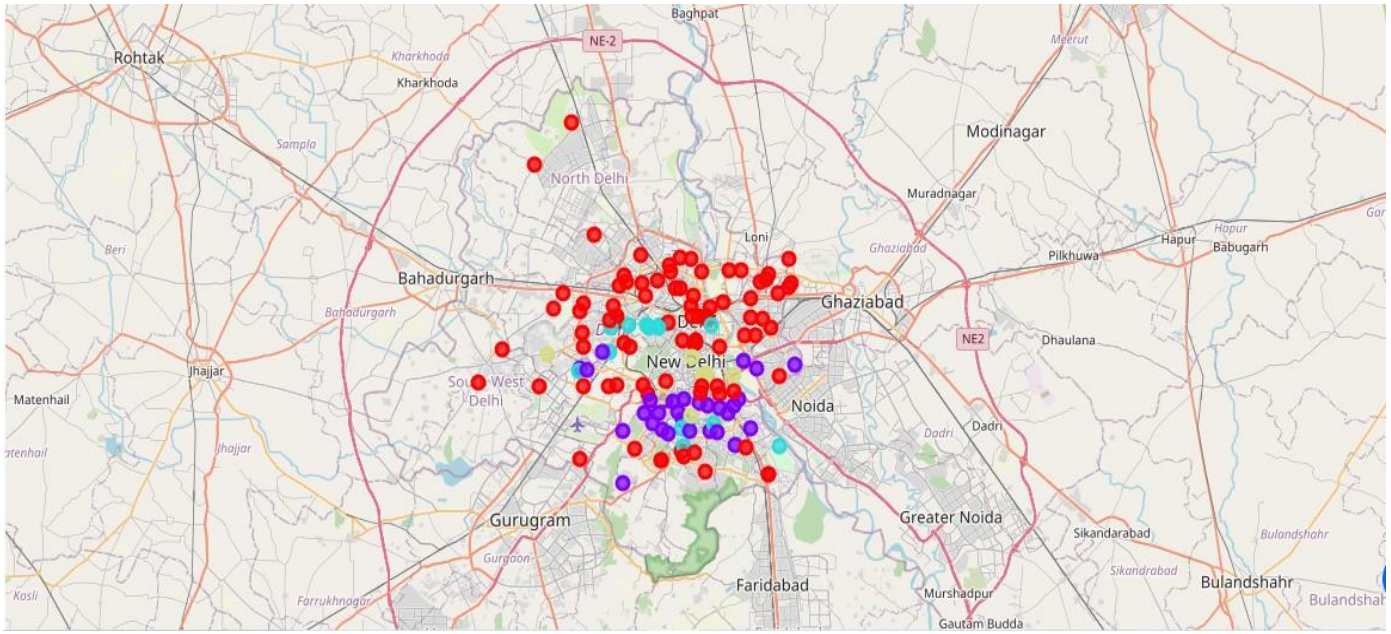
Classifying the data into clusters, I found 3 clusters as follows:

Cluster 0: Neighborhoods with less number of Indian Restaurants.

Cluster 1: Neighborhoods with large number of Indian Restaurants.

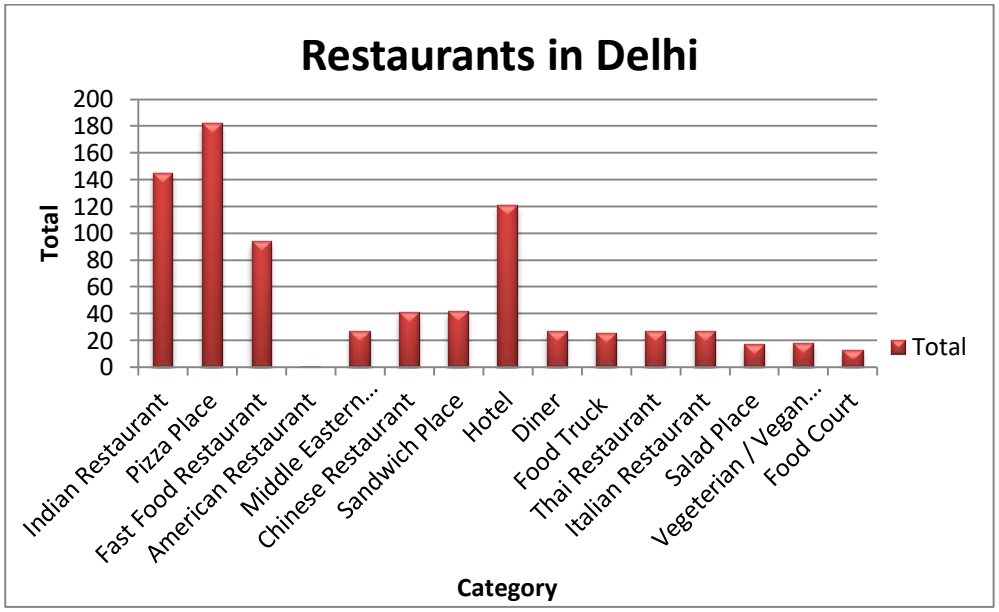
Cluster 2: Neighborhoods with moderate number of Indian Restaurants.

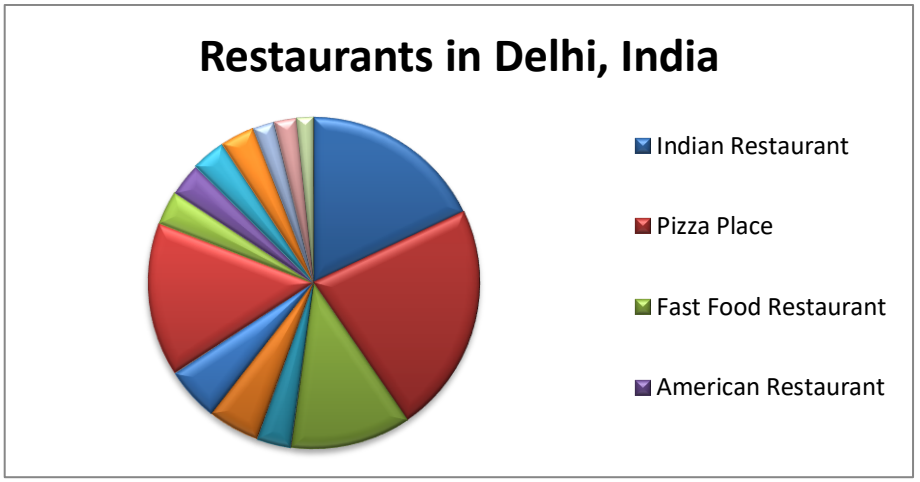
These clusters are visualized with the help of map.

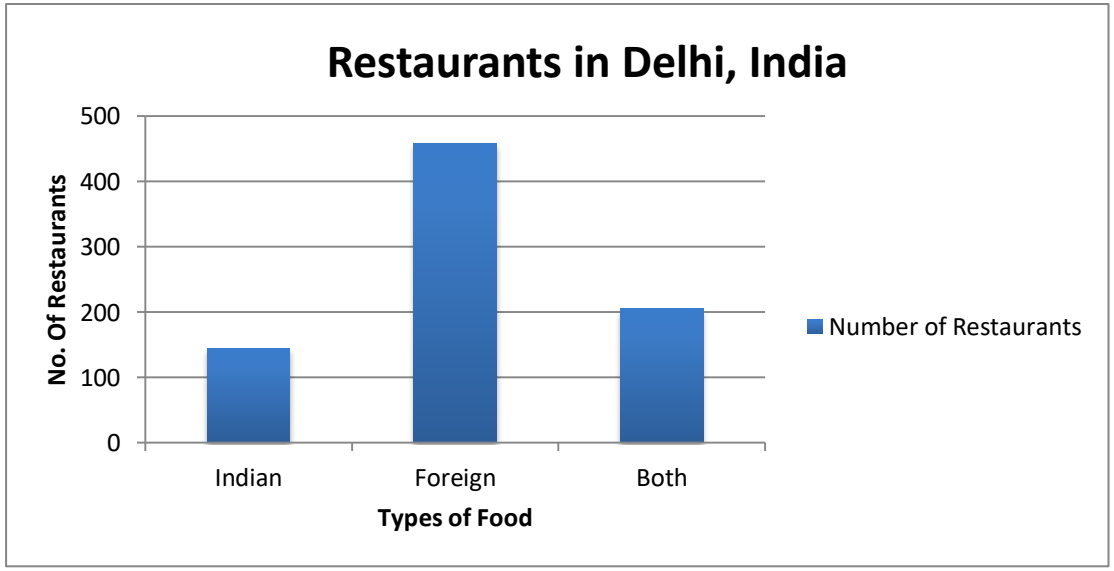


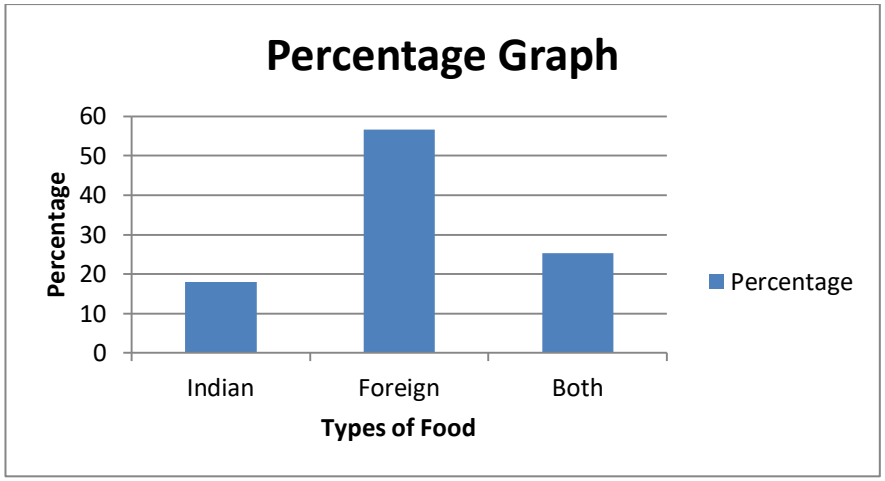
In the above map, cluster 0 is represented by Blue colour, Cluster 1 by Red colour and Cluster 2 by Purple colour.

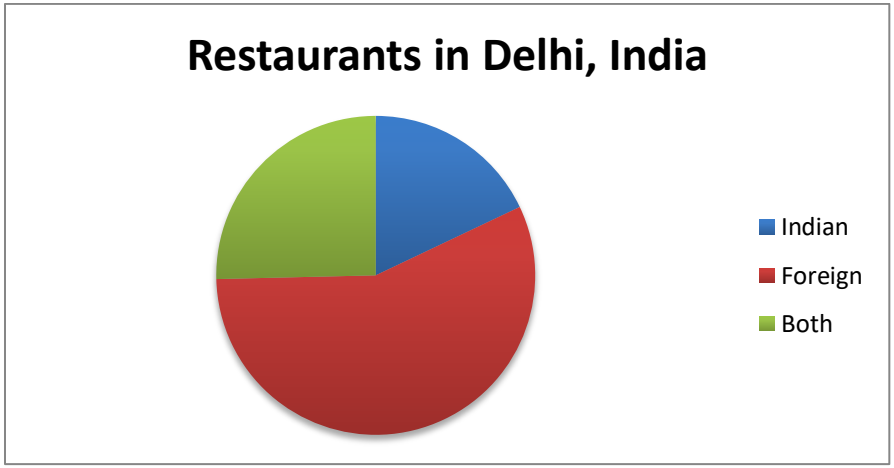
Then the bar graphs and pie charts are as follows:




Restaurants in Delhi, India
Indian Restaurant
Pizza Place
Fast Food Restaurant
American Restaurant


Restaurants in Delhi, India
500
400
300
200
100
0
No. Of Restaurants
Indian
Foreign
Both
Types of Food
Number of Restaurants


Percentage Graph
60
50
40
30
20
10
0
Percentage
Indian
Foreign
Both
Types of Food
Percentage


Restaurants in Delhi, India
Indian
Foreign
Both

Discussion:

The observations can be noted from the above mentioned results. The results of K-Means clustering show that there is highest concentration of Indian restaurants in northern part of Delhi. The number of Indian restaurants is moderate in southern part. The central part of Delhi has very few numbers of Indian Restaurants. This clearly shows that there is a greater opportunity to start a new Indian Restaurant in central part of Delhi. Since, there are large numbers of Indian restaurants in northern part; there will be large competition for a new restaurant. Therefore, it will not be convenient to open a new restaurant there. So, central part of Delhi will be more suitable and potential to start a new Indian Restaurant.

From another analysis of types of restaurants in Delhi, it is shown that the overall availability of Indian food is less in comparison with the foreign food. There are less numbers of restaurants that serve only Indian food.

Therefore, my project suggests the businessmen, investors, developers and people interested in hoteling sector to start new Indian restaurants in the Central part of Delhi. The number of Indian restaurants has to be increased in Delhi. I also recommend the new people in

this business to study my project for starting the new Restaurant. I recommend all these people to increase their business in Indian Food to help the country to be self-dependent.

Conclusion:

As mentioned above, the new Indian Restaurant can be started in Central areas of Delhi. The availability of Indian food should be increased in comparison with foreign food. The number of only Indian Food Restaurants is very less in Delhi in comparison with Foreign Restaurants. Therefore, the number of restaurants only serving Indian food should be increased in entire Delhi. Especially, it has to be increased in areas in Central part of Delhi.

Limitations and future directions:

My project is based on the data of only number of restaurants in Delhi. But, there are so many factors that affect the hoteling sector in that area such as population, location, class of living population in respective areas, choice of majority people etc. etc… It is practically complicated and time consuming to collect all these data and analyse.

Now days, the percentage of online food deliveries is rapidly increasing. Someone can consider this factor in this project while doing further analysis.

I hope my project will help maximum people to make good career in hoteling sector. Anyone can give me more suggestions on this project so that I can improve it. Anyone can point out my mistakes and suggest more ideas. Thank you!!