## Background

Opening and running a business in Food industry is difficult, as there are a lot of challenges one will face, when opening such establishment. For example, there is an issue of where to operate from as the location is really an important factor. There have been a lot of cases where establishments have failed just because their location was not appropriate. But this problem could be solved using data.

## Problem Statement and Description

In the Union Territory of New Delhi, there are a lot of places for the residents to go out and eat, and there are people who would like to take the opportunity of serving such people by opening his/her own restaurant. Food business always thrives on the quality of the food served, variety of the dishes and area in which it is operating. But, in large cities like New Delhi, it could be difficult to decide where should one open his new restaurant, after which he would have to decide which cuisines to offer and many more things. No worries, these problems could be easily solved by using the data and applying certain tools on it.

**So, to find appropriate answers to the problem we must-**

1. Classify different areas in different Clusters
2. Decide appropriate location regarding where to open a restaurant

## Data Requirements

Since we have established a problem, we must go ahead and decide what kinds of data would be needed. To solve the problem for deciding places where people usually go to eat, we must determine that by clustering the neighbourhoods into different categories, and then determine which could be a more appropriate place to operate at. Then we would need the

**Important links:**

https://simple.wikipedia.org/wiki/List_of_districts_in_Delhi

https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi

These are the sources used in identifying areas in Delhi for further segmentation, analysis and finding optimum place for a person to open a restaurant.

Apart from this data, Foursquare API will be used to gather data regarding the places in each neighbourhood, so that a more detailed and thorough analysis could be conducted by clustering different areas according to the types of places present in the locality and proper findings can be shown.

## Data Gathering and Transformation

Initially, names of the Neighbourhoods of Delhi are collected by scrapping data from a Wikipedia web page. After collecting names of Neighbourhoods Geolocator library is used to locate the coordinates of each neighbourhood so that these markers could be used for-

1. Plotting markers of the locations on Map to visualize the areas.
2. Using coordinates in Foursquare API to make several calls and gather data of different places in each neighbourhood for classification of them in clusters.
3. After classifying, they will be marked on map in different colours for visualizing different clusters in the city.

**Neighbourhoods in Delhi marked using Folium**
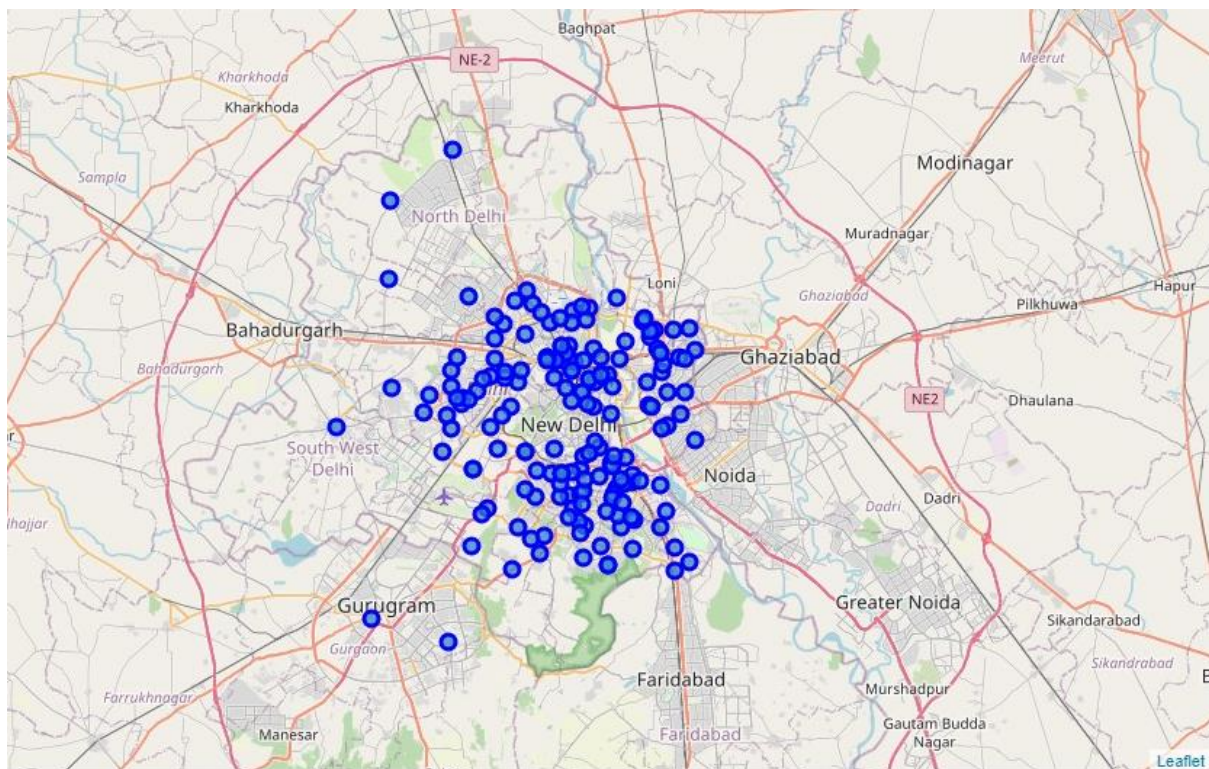


So, in total location data of 173 neighbourhoods was collected using Geolocator Library in python. This data was then plotted on the Map using Folium Maps library, which will be used again when plotting for classified clusters.

**Top Venue Categories in Delhi**
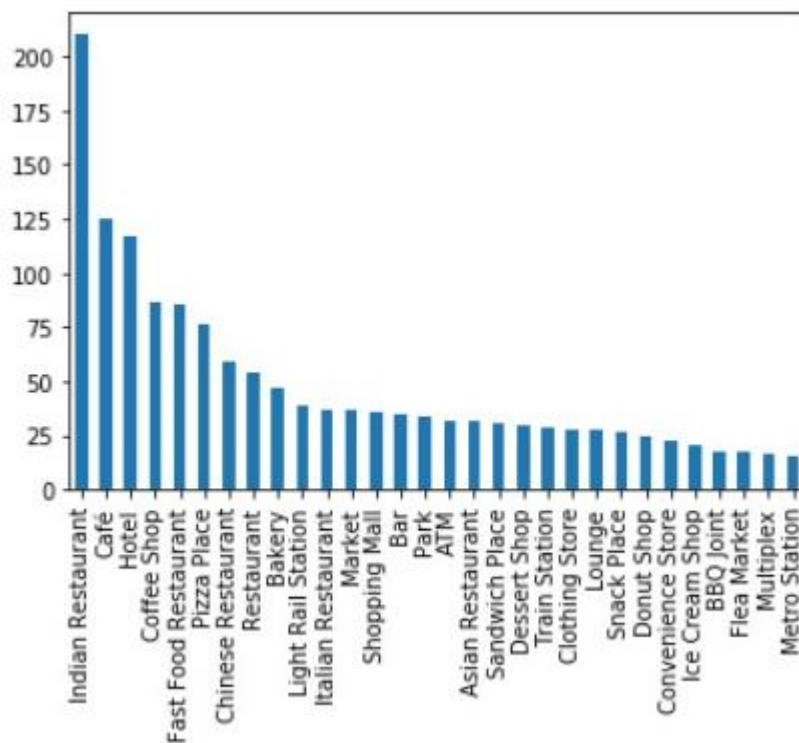


After collecting data from Foursquare, a total of 2080 locations were found in their database located in Delhi. Those 2080 locations were divided into 218 different categories, of those top 30 categories with respect to the number of venues in each category, is represented as a bar graph plotted above. So, we can easily see that the first 9 venue categories are related to food, of which the most venues were Indian Restaurants, followed by Café and then Hotel. Most common category is related to food and is quite popular.

In the Jupyter notebook, for each neighbourhood, top 5 venues are shown for a better understanding of each area.

## Methodology

The data collection, transformation, and analysis, all was done in the Jupyter Notebook using python. Sources of data were different for different data. Mainly there were 3 sources:

1. Wikipedia
2. Geolocator
3. Foursquare

After collecting and transforming data into meaningful dataset, an algorithm called as K-Means-Clustering was applied to the collected data. This technique gathers the specifications/features of the data and divides them into clusters of different properties. Each cluster has similar kinds of elements and different clusters have different properties.

In this case we divided these Neighbourhoods into 10 different clusters, and then plotted them on a map, using Folium.

**The map of Delhi showing neighbourhoods divided into 10 clusters**
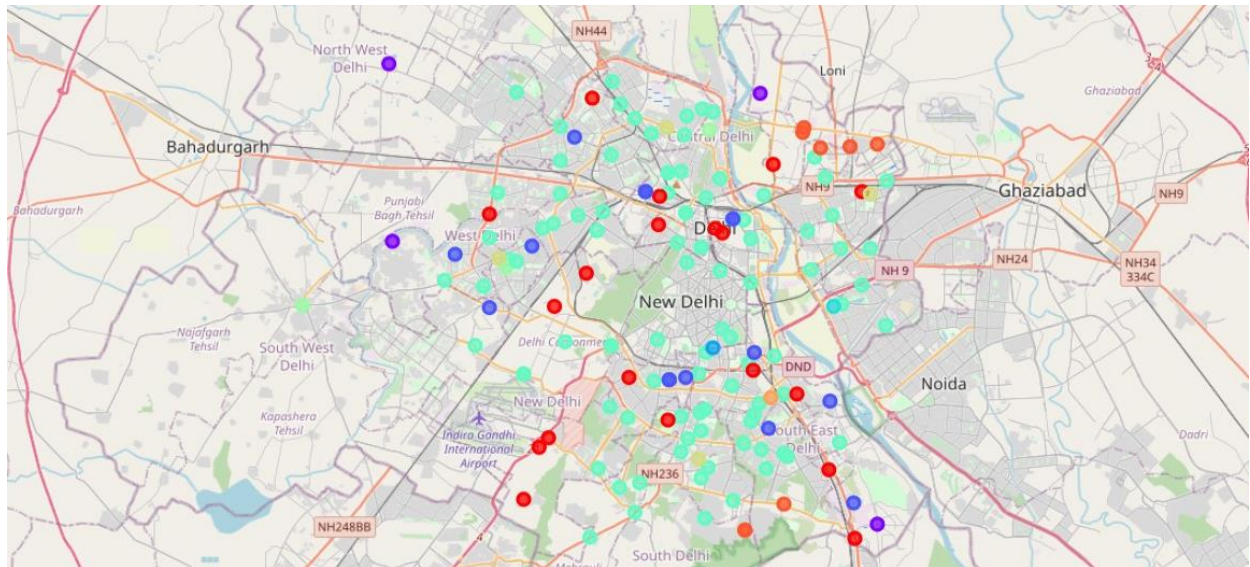


## Results

After dividing the neighbourhoods into 10 clusters, we can now view the properties of types of neighbourhoods in each cluster and then choose appropriate neighbourhoods where, one can open a Restaurant.

**These are the top 5 most common venues in each cluster:**

```
---Cluster1---

Hotel
Frozen Yogurt Shop
Fried Chicken Joint
French Restaurant
Food Truck

---Cluster2---

ATM
Furniture / Home Store
Frozen Yogurt Shop
Fried Chicken Joint
French Restaurant
```

---Cluster3---

Indian Restaurant
Fried Chicken Joint
Food Truck
French Restaurant
Women's Store

---Cluster4---

Chinese Restaurant
Women's Store
Farmers Market
Furniture / Home Store
Frozen Yogurt Shop

---Cluster5---

Hindu Temple
Women's Store
Fast Food Restaurant
Furniture / Home Store
Frozen Yogurt Shop

---Cluster6---

Indian Restaurant
Fast Food Restaurant
Café
Pizza Place
Coffee Shop

---Cluster7---

Furniture / Home Store
Frozen Yogurt Shop
Fried Chicken Joint
French Restaurant
Food Truck

---Cluster8---

French Restaurant
Frozen Yogurt Shop
Food Truck
Fried Chicken Joint
Furniture / Home Store

```
---Cluster9---

Department Store
Fast Food Restaurant
Gaming Cafe
Furniture / Home Store
Frozen Yogurt Shop

---Cluster10---

ATM
Frozen Yogurt Shop
Fried Chicken Joint
French Restaurant
Food Truck
```

**Number of Neighbourhoods in each cluster:**

```
5.0   96
0.0   19
2.0   15
9.0   10
7.0   4
1.0   4
6.0   3
8.0   1
3.0   1
4.0   1
Name: Cluster Labels, dtype: int64
```

With these findings, we can surely say that

- Most common type cluster is 5, which has Fast Food Restaurants
- Clusters- 1,2,3,8,7 commonly have French Restaurants
- Clusters-3,6 commonly have Indian Restaurants

## Discussion

Now that we have findings backed by data, one can easily plan with these findings, about whether he will go into areas where that venue common and will have to compete with them in order to survive in the market. Also, he can go into areas where that category of venue is not there at present and could possibly attract people by introducing something new in their neighbourhood. It is completely dependent on the person on how he would use these findings

## Conclusion

These findings from the analysis could be very useful for someone looking to get into food business and would help him a lot to understand the characteristics of the neighbourhoods in Delhi, and Delhi as whole. He would be able to get a fair idea of what kind of challenges and opportunities each neighbourhood offers. This would help him in planning about those in advance and would be very beneficial in the initial stages of business.