# Opening a New Restaurant in the Capital City of India-Delhi

Ayush Rawat

July 15, 2020

## 1. Introduction

### 1.1 Background

New Delhi is the capital city of India. A population of more than 11 million is complemented by its market for venues ranging from basic necessity providers to restaurants and hotels. Delhi is a city known for the quality and taste of food available throughout the country. People often go out for food walks through the city hitting multiple food spots throughout the day. The variety of quality food cuisines available in New Delhi is appealing to the population of the country and many choose to visit the capital city for its rich historic sites and for the food. It is also an attractive city for people looking to invest in the food market due to the demand of quality food. Therefore, it will be advantageous to have a system in place that can help the stakeholders narrow down the possible locations to invest in.

### 1.2 Problem

Data of the existing venues in and around the various neighbourhoods of the city can help in narrowing down the possible locations. This project aims to recommend a location on basis of the demand of a particular type of venue and its distance from popular places in the city to maximize the chances of a successful business.

### 1.3 Interest

Stakeholders who are looking to invest and open a new restaurant in the city are facing a new challenge: finding a location where the restaurant market is not saturated in order to boost the chances of being a go-to restaurant for people passing by or those who live in the surrounding areas. Being a hit rests on the experience the restaurant provides.

## 2. Data Acquisition and Cleaning

### 2.1 Data Sources

The data required to perform analysis and make accurate recommendations can be found online. Delhi's Neighbourhood and Borough's data can be found in Wikipedia on this link. The details of locations of these Neighbourhoods can be found using Geopy API. Data containing information on the venues in and around those Neighbourhoods can be found using Foursquare API.

### 2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. The data of strings needs to be cleaned off by removing spaces and special characters to avoid confusion during the advanced steps of the project. Since the information is compiled from multiple sources it needs to be joined in a way that doesn't corrupt the data. Joins can be performed over the different sets of data obtained from different sources on their common features like Neighbourhoods found in the Neighbourhood data need to be joined with the dataframe containing location details of those Neighbourhoods. Additionally, data obtained from Foursquare API needs to be joined to the resulting dataframe on basis of the Neighbourhood details. It is clear that the Neighbourhood details are a pivot for us across all dataframes as we have to make a Neighbourhood recommendation of a potential location to open a new venue.

### 2.3 Feature Selection

After Data Cleaning, the data compiled will still contain information that is not relevant to the purpose of this project like additional details of the venue received by Foursquare API. The location data and the type of venue details are the only features which are relevant to serve the purpose of this project. The data being scrapped from Wikipedia is already selection and doesn't need to be trimmed. Geopy API will be used to just extract the latitude and longitude details of each of the Neighbourhoods.

Table 1. Final form of data compiled from different sources before analysis

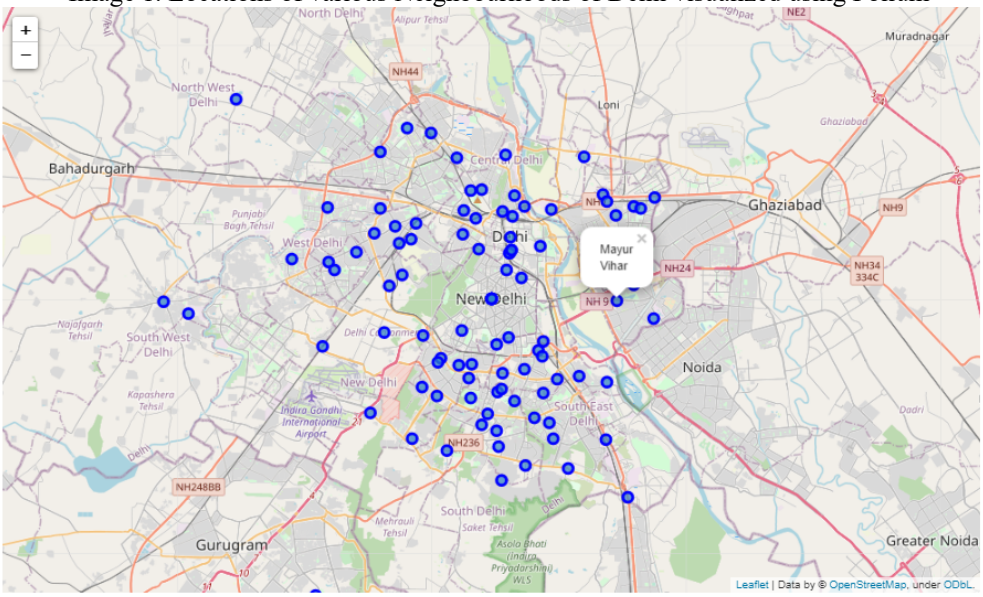
| | Neighbourhood | Neighbourhood Latitude | Neighbourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Adarsh Nagar | 28.714401 | 77.167288 | Pahalwan Dhaba | 28.714594 | 77.172155 | Indian Restaurant |
| 1 | Ashok Vihar | 28.699453 | 77.184826 | Nat Khat Caterers | 28.699630 | 77.187832 | Indian Restaurant |
| 2 | Ashok Vihar | 28.699453 | 77.184826 | Bakers Stop | 28.700495 | 77.188716 | Bakery |
| 3 | Ashok Vihar | 28.699453 | 77.184826 | Invitation Banquet | 28.696018 | 77.185953 | Diner |
| 4 | Ashok Vihar | 28.699453 | 77.184826 | Gola Northend | 28.701242 | 77.189288 | Indian Restaurant |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 526 | Tilak Nagar | 28.636548 | 77.096496 | Bikanerwala | 28.633524 | 77.097492 | Fast Food Restaurant |
| 527 | Vikaspuri | 28.638419 | 77.070836 | Domino's Pizza | 28.638000 | 77.075000 | Pizza Place |
| 528 | Vikaspuri | 28.638419 | 77.070836 | McDonald's | 28.639752 | 77.075190 | Fast Food Restaurant |
| 529 | Vikaspuri | 28.638419 | 77.070836 | Asian Garden | 28.639661 | 77.074482 | Indian Restaurant |
| 530 | Vikaspuri | 28.638419 | 77.070836 | Mela | 28.635672 | 77.067237 | Indian Restaurant |

## 3. Exploratory Data Analysis

### 3.1 Analysing the Location of Neighbourhoods on Map

Folium is used to analyse the location of each Neighbourhood spread over Delhi. Each Neighbourhood is marked with a pin and reveals the name of the neighbourhood when clicked.

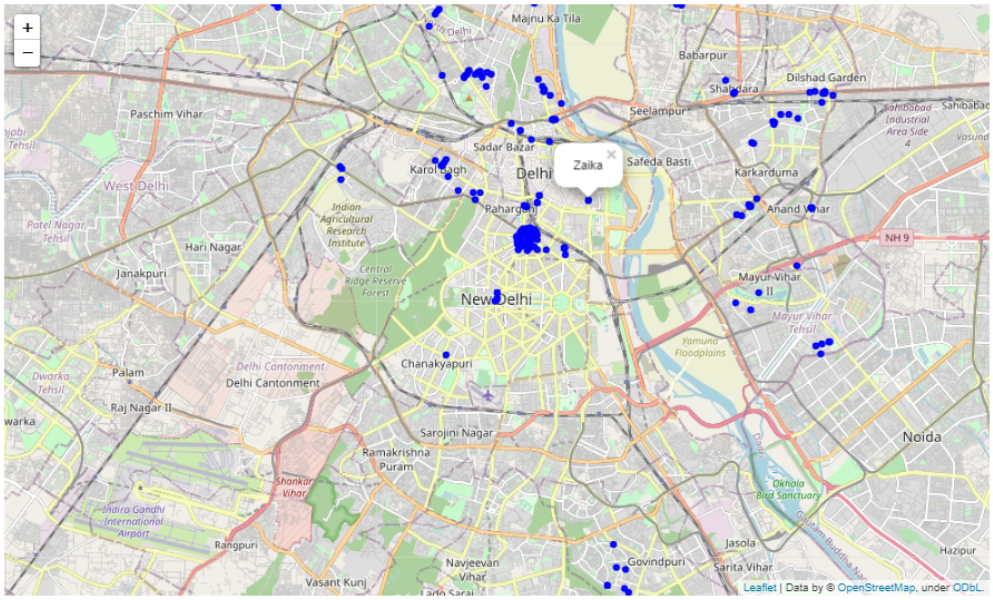
Image 1. Locations of various Neighbourhoods of Delhi visualized using Folium




### 3.1 Analysing the Location of Venues across Neighbourhoods on a Map

Another map leaflet is generated to visualize the different types of venues and their locations using pins as markers. The number of venues that can be visualized is limited by the number of markers a leaflet can render successfully which is somewhere around 200. So 200 venue locations are visualized on the map.

Image 2. Locations of various venues of Delhi spread across Neighbourhoods visualized using Folium



## 4. Classification Modelling

To solve the problem introduced in the project, a Classification model that can cluster the Neighbourhoods based on the types of most common venues will be needed. K Means algorithm will be used in this project to form the clusters of Neighbourhoods with similar types of venues.

### 4.1 Transforming the Data to a Suitable format for KMeans Clustering Algorithm

To apply KMeans Clustering algorithm, there is a fixed number of clusters that needs to be determined in order to get the results which best suit the problem at hand. Generally this can be done using some performance metrics which return a numerical error value in order to better evaluate the result and the solution can be selected as the one with least error value. In this case however, there are no numerical values involved, this problem revolved around recommending a Neighbourhood to a stakeholder to maximize the chances of success in terms of location. A fixed value of 5 clusters is chosen as a starting point to apply the KMeans Clustering Algorithm over the dataset.

The Data is used in a way that a table of the top 10(abundant) types of venues is displayed for each of the Neighbourhoods in Delhi.

Table 2. Top 10(abundant) types of Venues for each Neighbourhoods

| | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Adarsh Nagar | Indian Restaurant | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | Fish & Chips Shop | Fast Food Restaurant |
| 1 | Alaknanda | BBQ Joint | Indian Restaurant | New American Restaurant | Pizza Place | Steakhouse | Middle Eastern Restaurant | Restaurant | Dumpling Restaurant | Diner | Doner Restaurant |
| 2 | Anand Vihar | Indian Restaurant | Fast Food Restaurant | Pizza Place | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food |
| 3 | Ashok Nagar | Indian Restaurant | Food Truck | Food Court | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food | Fish & Chips Shop | Fast Food Restaurant |
| 4 | Ashok Vihar | Indian Restaurant | Bakery | Diner | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 86 | Vikaspuri | Indian Restaurant | Fast Food Restaurant | Pizza Place | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food |
| 87 | Vishwas Nagar | Indian Restaurant | Café | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | Fish & Chips Shop |
| 88 | Vivek Vihar | Indian Restaurant | Japanese Restaurant | Wings Joint | Hot Dog Joint | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | Fish & Chips Shop |
| 89 | Wazirabad | Snack Place | Wings Joint | Dhaba | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | Fish & Chips Shop | Fast Food Restaurant |
| 90 | Yamuna Vihar | Dumpling Restaurant | North Indian Restaurant | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | Fish & Chips Shop |

### 4.2 Applying the KMeans Clustering Algorithm

Once the data is transformed to a suitable format, KMeans Clustering Algorithm can be applied to the data and clusters can be formed. 5 clusters were created and each Neighbourhood was allotted a Cluster Label according to the kind of most abundant types of venues in each Neighbourhood.

The column of Cluster Label was added to the Table 2 to match each of the Neighbourhoods with the allotted Cluster as per results from KMeans Algorithm.

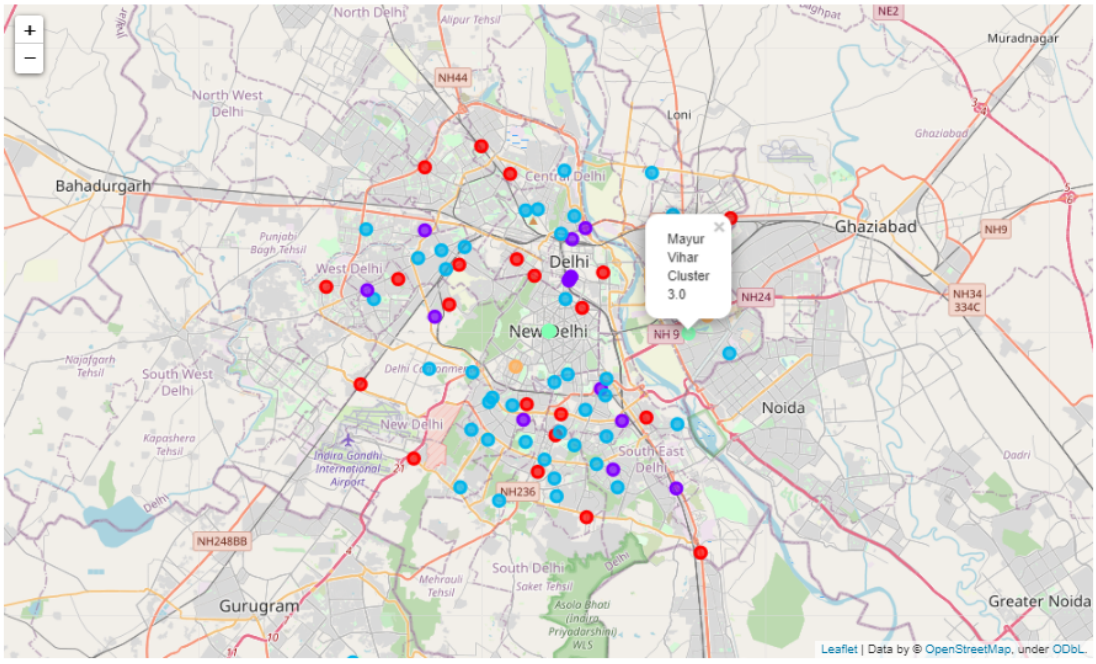
Table 3. Data Table with Cluster Labels allotted by KMeans Algorithm

| | Borough | Neighbourhood | latitude | longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Cor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | North West Delhi | Adarsh Nagar | 28.714401 | 77.167288 | 0.0 | Indian Restaurant | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | Food | |
| 1 | North West Delhi | Ashok Vihar | 28.699453 | 77.184826 | 0.0 | Indian Restaurant | Bakery | Diner | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | |
| 3 | North West Delhi | Model Town | 28.613895 | 77.209006 | 3.0 | Indian Restaurant | Food Truck | Food Court | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food | |
| 5 | North West Delhi | Pitam Pura | 28.703268 | 77.132250 | 0.0 | Indian Restaurant | Fried Chicken Joint | Café | Wings Joint | Dim Sum Restaurant | French Restaurant | Food Truck | Food Court | |
| 7 | North Delhi | Civil Lines | 28.676851 | 77.225030 | 2.0 | Burger Joint | Café | Asian Restaurant | Indian Restaurant | Italian Restaurant | Chinese Restaurant | Dumpling Restaurant | Falafel Restaurant | Fast Rest |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 97 | West Delhi | Patel Nagar | 28.659809 | 77.156957 | 2.0 | Café | Wings Joint | Dim Sum Restaurant | Gastropub | Fried Chicken Joint | French Restaurant | Food Truck | Food Court | |
| 98 | West Delhi | Punjabi Bagh | 28.668945 | 77.132461 | 1.0 | Fast Food Restaurant | Bakery | Pizza Place | Breakfast Spot | Sandwich Place | Wings Joint | Doner Restaurant | Donut Shop | Dur Rest |
| 99 | West Delhi | Rajouri Garden | 28.642152 | 77.116060 | 0.0 | Indian Restaurant | BBQ Joint | Bakery | Food Truck | Wings Joint | Diner | Fried Chicken Joint | French Restaurant | |
| 100 | West Delhi | Tilak Nagar | 28.636548 | 77.096496 | 1.0 | Fast Food Restaurant | Café | Wings Joint | Dim Sum Restaurant | Gastropub | Fried Chicken Joint | French Restaurant | Food Truck | |
| 101 | West Delhi | Vikaspuri | 28.638419 | 77.070836 | 0.0 | Indian Restaurant | Fast Food Restaurant | Pizza Place | Wings Joint | Dim Sum Restaurant | Fried Chicken Joint | French Restaurant | Food Truck | |

## 4.3 Visualizing the Clusters

The clusters formed by using the KMeans Algorithm are then displayed on a map using Folium. Each of the Neighbourhood Clusters will be shown in a different colour to differentiate it from other clusters.

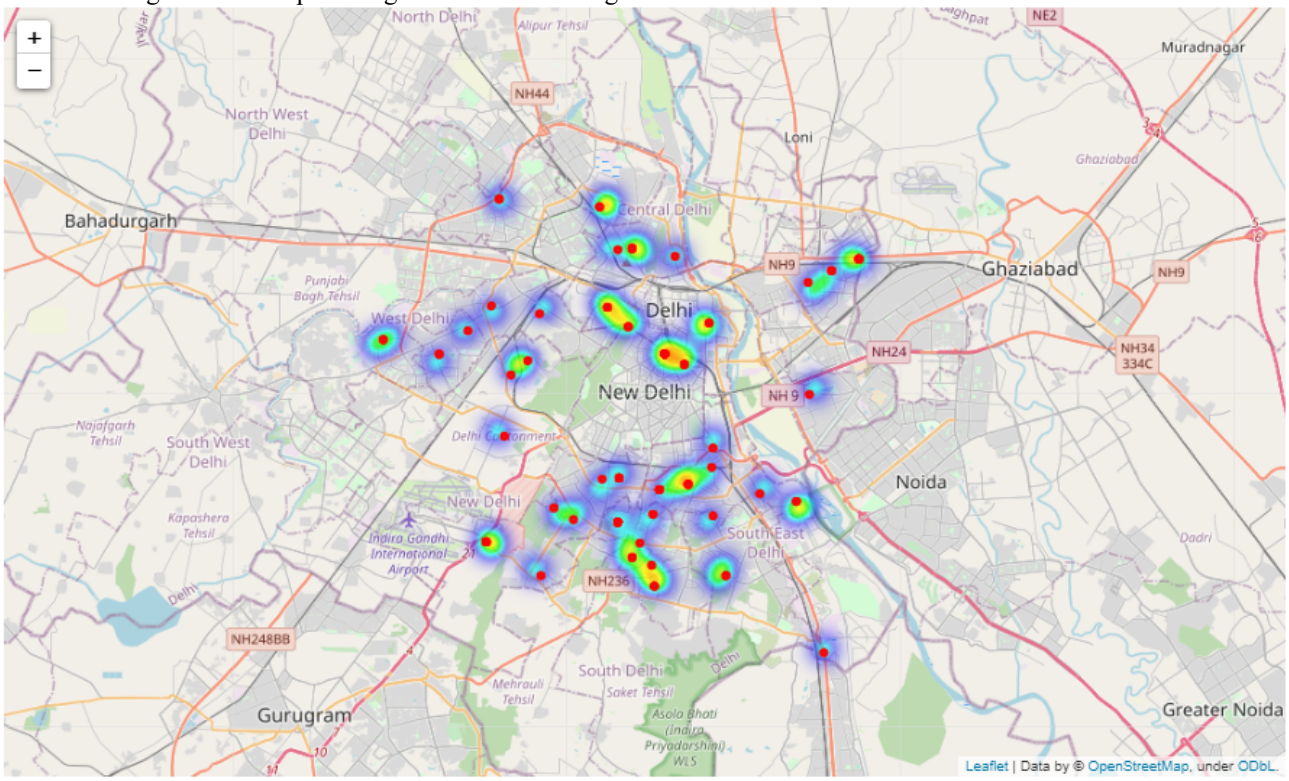
Image 3. Locations of all Neighbourhoods across different Clusters visualized using Folium



### 4.4 Analysing the Clusters

The clusters formed by the KMeans algorithm are then analysed. On plotting the number of venues against their types found in each cluster on a bar graph, it is clear that the demand for Indian Restaurants is high among the Clusters 0 and 2. Heat Maps are plotted to show the Indian Restaurants in the clusters 0 and 2. The last constraint is added to the data where neighbourhoods of higher than 30 percentile in terms of number of Indian Restaurants and 60 percentile in terms of number of venues are removed. This removed the neighbourhoods with high density of Indian Restaurants. Since the restaurant stakeholders intend to open their restaurant near a popular neighbourhood, we recommend locations based on their distance from the popular neighbourhoods.

Image 4. Heat-map of Neighbourhoods with highest demand but least densities of Indian Restaurants



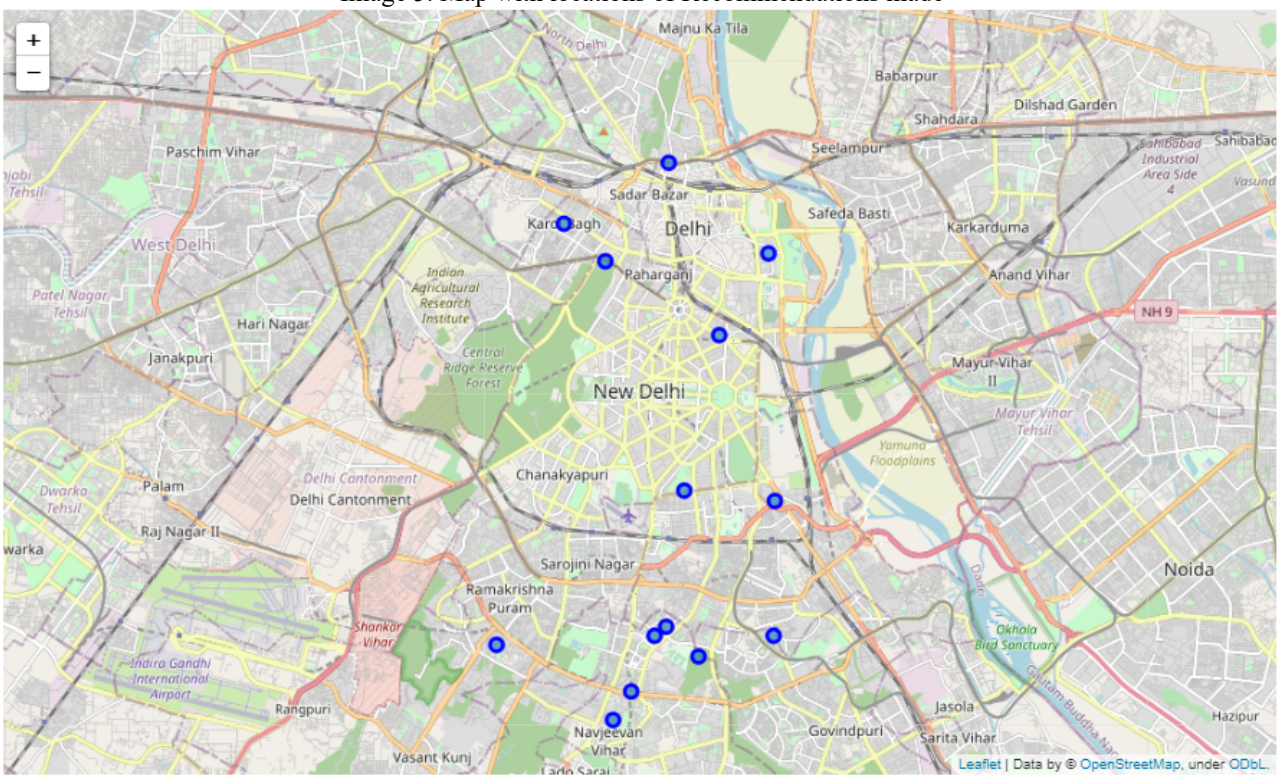

### 4.5 Making the Recommendation

Once the distances of the Neighbourhoods with higher demand but lower densities from popular neighbourhoods are calculated, the neighbourhoods can be ranked based on the distance. The Neighbourhoods are ranked with minimum distance from popular neighbourhood being ranked the highest and the most distant neighbourhoods will be lower down the ranks. A list can be returned in the ranked order for the Stakeholder to make the decision as per their personal preferences. The recommended locations are then displayed on a map with markers.

Table 4. Recommended Neighbourhoods based on lower density and higher demand for Indian Restaurants

| | Neighbourhood | latitude | longitude |
|---|---|---|---|
| 0 | Barakhamba Road | 28.626952 | 77.229950 |
| 1 | Jhandewalan | 28.644319 | 77.199917 |
| 2 | Daryaganj | 28.646090 | 77.243048 |
| 3 | Karol Bagh | 28.652998 | 77.189023 |
| 4 | Tis Hazari | 28.667163 | 77.216631 |
| 5 | Hauz Khas | 28.544256 | 77.206707 |
| 6 | Gulmohar Park | 28.557101 | 77.213005 |
| 7 | Sarvodaya Enclave | 28.537478 | 77.202089 |
| 8 | Neeti Bagh | 28.559251 | 77.216166 |
| 9 | Munirka | 28.554886 | 77.171084 |
| 10 | East of Kailash | 28.557032 | 77.244614 |
| 11 | Siri Fort | 28.552146 | 77.224698 |
| 12 | Nizamuddin West | 28.588365 | 77.244955 |
| 13 | Lodi Colony | 28.590702 | 77.220921 |

Image 5. Map with locations of Recommendations made




## 5. Conclusion

In this project, I analysed the data of venues in and around the Neighbourhoods of Delhi to create a recommendation system. The venue location data used was taken from Foursquare API. I used KMeans Algorithm to form clusters of neighbourhoods and then used them to filter places with higher demand for Indian Restaurant and lower existing densities of the same. These recommendations were made by calculating the distance of potential recommendations from the popular neighbourhoods.

## 6. Future Directions

In the future we can create a system which uses the ratings of the restaurants in Neighbourhoods and the income statistics of Neighbourhoods to better target a specific audience. Similarly, we can potentially analyse the communities living in these Neighbourhoods and their preferences of cuisines and types of restaurants before making a decision.