Capstone Project

Restaurant Location Recommendation

Introduction: Business Problem

Delhi is the capital of India and it is one of the oldest cities in India. According to the 2011 census, Delhi's city proper population was over 11 million, the second-highest in India after Mumbai. As India's national capital and centuries old Mughal capital, Delhi influenced the food habits of its residents and is where Mughlai cuisine originated. Along with Indian cuisine, a variety of international cuisines are popular among the residents. The dearth of food habits among the city's residents created a unique style of cooking which became popular throughout the world, with dishes such as Kebab, biryani, tandoori. The city's classic dishes include butter chicken, dal makhani, shahi paneer, aloo chaat, chaat, dahi bhalla, kachori, gol gappe, samosa, chole bhature, chole kulche, gulab jamun, jalebi and lassi.

In this project will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian Cuisine restaurant in Delhi, India. Finding a suitable location for restaurants in major cities like Delhi proves to be a daunting task. Various factors such as over-saturation or no demand, for the type of restaurant that the customer wants to open, effect the success or failure of the restaurant. Hence, customers can bolster their decisions using the descriptive and predictive capabilities of data science.

We need to find locations (Neighborhood) that have a **potentially unfulfilled demand** for Indian Restaurant. Also, we need locations that have **low competition and are not already crowded**. We would also prefer location as close to popular city Neighborhood, assuming the first two conditions are met.

We will use our data science powers to generate a few most promising neighbourhoods based on this criteira. Advantages of each area will then be clearly Expressed so that best possible final location can be chosen by stakeholders.

Data Acquisition and preparation

Data requirement

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighbourhood (any type of restaurant)
- number of and distance to Indian restaurants in the neighbourhood, if any
- distance of neighbourhood from popular neighbourhoods

Data Collection

In our project we will:

- acquire the names and boroughs of the neighbourhoods by scrapping a Wikipedia page.
- After we have got the names of all the neighborhoods, we will geocode them using the library geopy.geocoder (Nominatim).
- Next, we use the foursquare API to find all types of restaurants within a 1000 meter radius for every neighbourhood.

Foursquare

'The Foursquare Places API provides location-based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.'(Wikipedia)

Delhi Dataset

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.614192	77.071541
1	North West Delhi	Ashok Vihar	28.699453	77.184826
2	North West Delhi	Karala	28.735140	77.032511
3	North West Delhi	Model Town	28.641926	77.221750
4	North West Delhi	Narela	28.842610	77.091835

Restaurant Dataset

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	28.614193	77.071541	Eagle Boys Pizza	28.615595	77.070784	Pizza Place
1	Adarsh Nagar	28.614193	77.071541	Bikanerwala	28.613391	77.076084	Indian Restaurant
2	Adarsh Nagar	28.614193	77.071541	Bikano East Patel Nagar	28.616190	77.066978	Fast Food Restaurant
3	Adarsh Nagar	28.614193	77.071541	McDonald's	28.616330	77.067034	Fast Food Restaurant
4	Ashok Vihar	28.699453	77.184826	Nat Khat Caterers	28.699630	77.187832	Indian Restaurant

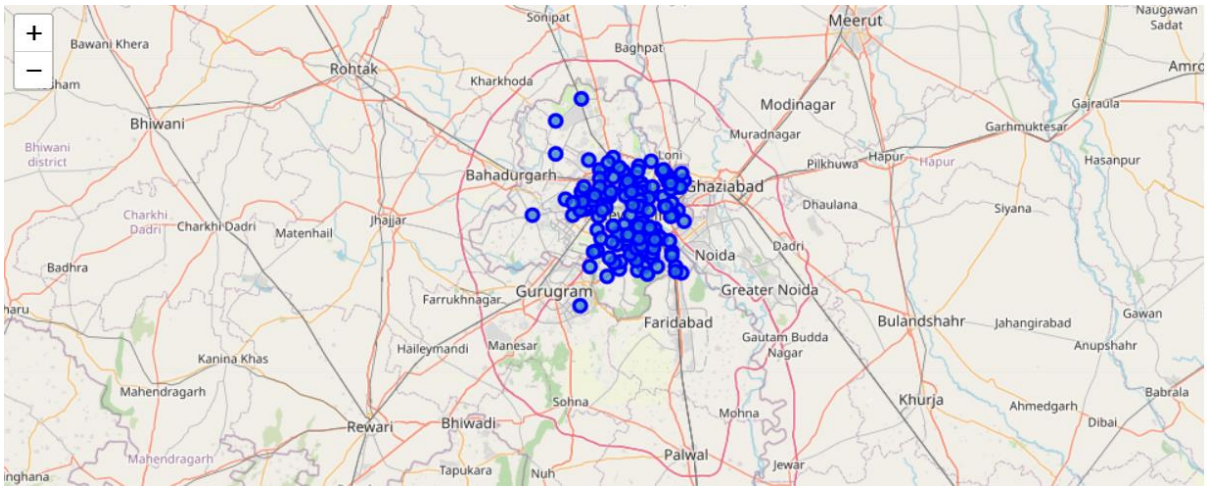
Methodology

Exploratory analysis and Clustering:

Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.

The Maps created based on the two datasets are given below.

Delhi Neighbourhood Map



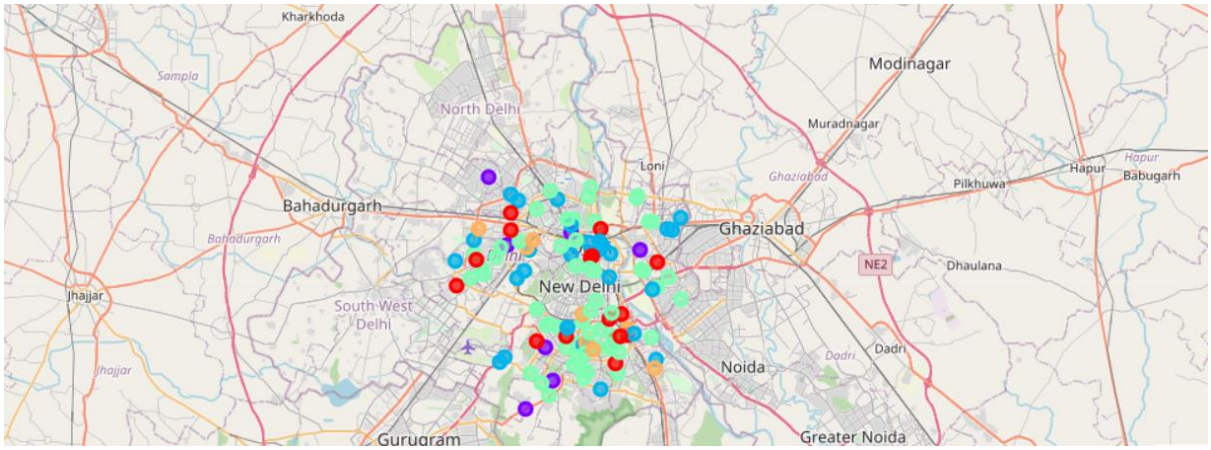
Delhi's major restaurant Map

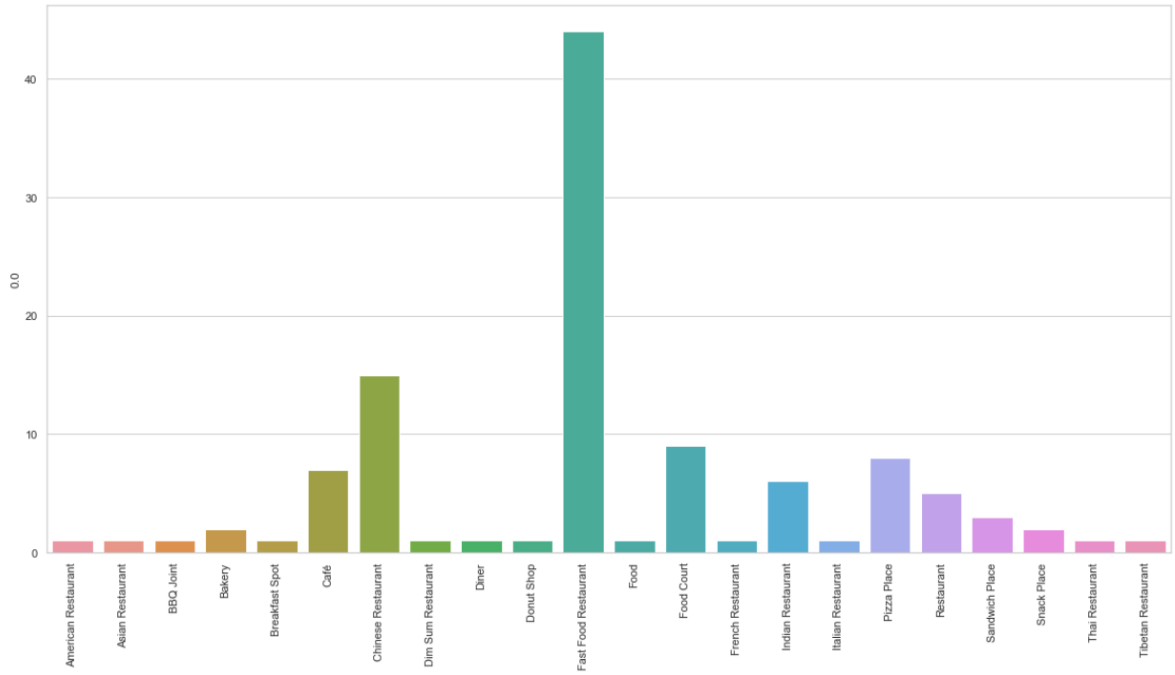
Our goal here is to find the neighborhoods with low density of Indian restaurants. But, how will we decide which neighborhoods, currently operating on minimal number of Indian restaurants, have the potential for growth and which neighborhoods do not.

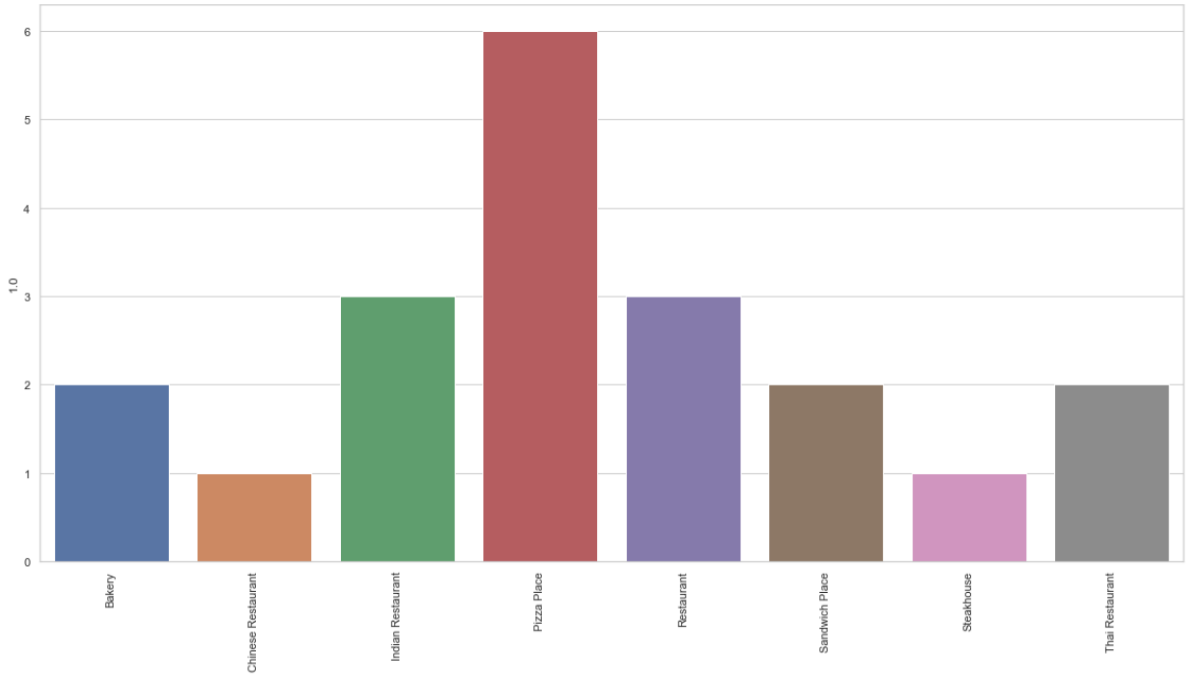
The most intuitive idea would be to find neighborhoods that have similar patterns of restaurant trends.

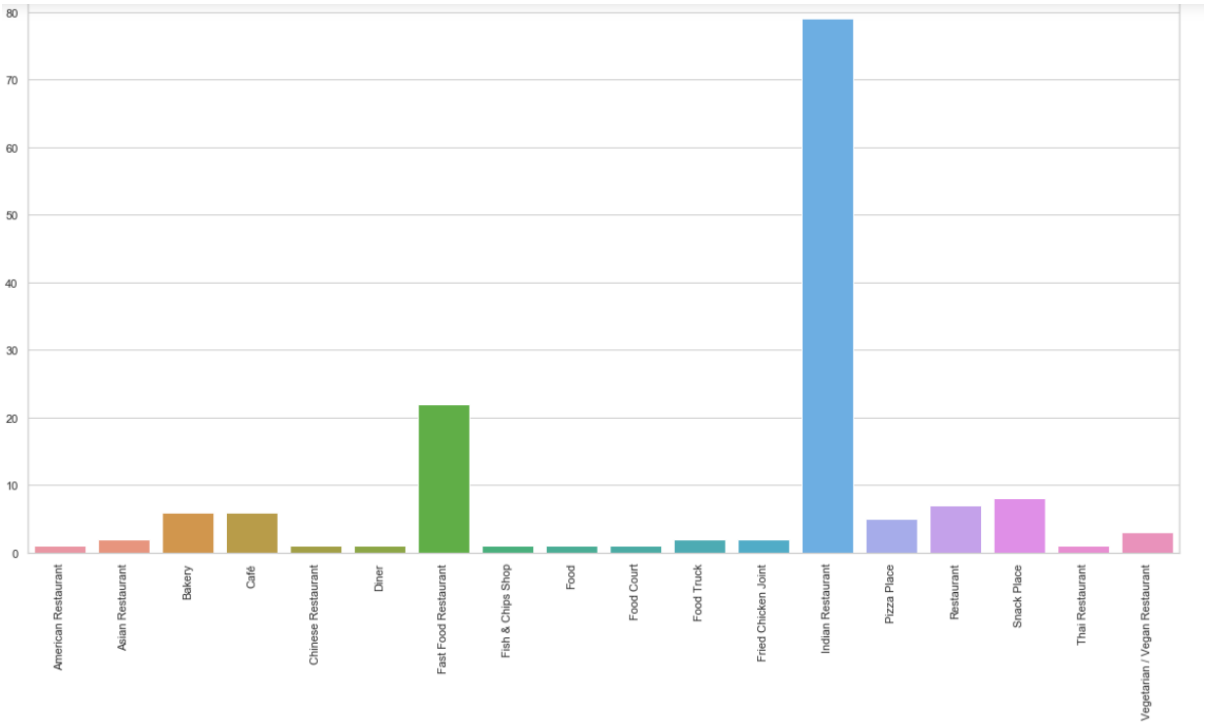
This can be achived by clustering the neighborhoods of the basis of the restaurant data we have acquired. Clustering is a predominant algorithm of unsupervised Machine Learning. It is used to segregate data entries in cluster depending of the similarity of their attributes, calculated by using the simple formula of euclidian distance.

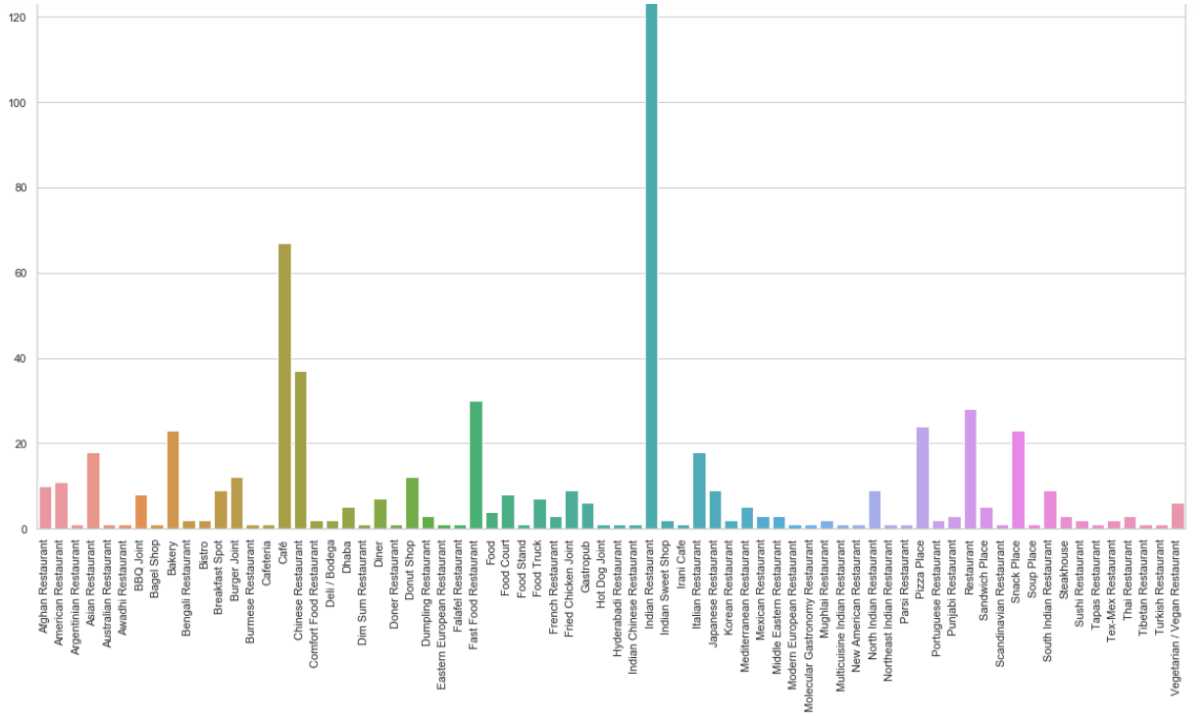
We can then analyze these clusters separately and use those clusters that show high trends of Indian Restaurants

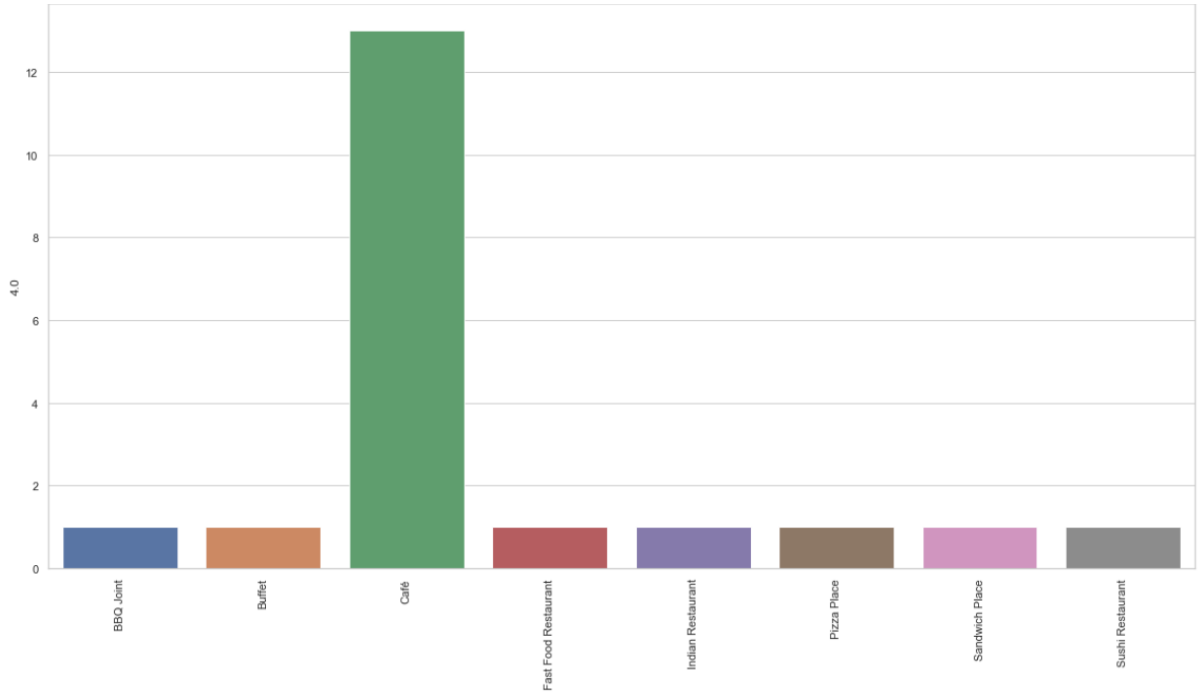





1.0
0
1
2
3
4
5
6
Bakery
Chinese Restaurant
Indian Restaurant
Pizza Place
Restaurant
Sandwich Place
Steakhouse
Thai Restaurant


0
10
20
30
40
50
60
70
80
American Restaurant
Asian Restaurant
Bakery
Café
Chinese Restaurant
Diner
Fast Food Restaurant
Fish & Chips Shop
Food
Food Court
Food Truck
Fried Chicken Joint
Indian Restaurant
Pizza Place
Restaurant
Snack Place
Thai Restaurant
Vegetarian / Vegan Restaurant





Analysing the bar graphs we can clearly see that clusters 1 and 2 have a high demand for Indian Restaurants

Recommendation

In this section:

we will, first, analyze the density of the Indian Restaurants in generally for each neighbourhood. Then we will weed out the neighbourhoods that in the highest 70 percentile of density Find out the most popular neighbourhoods Will then try to find remaining neighbourhoods that are close to them Provide the a detailed recommendation of top 10 neighbourhoods

Now, as clusters 1 and 2 have a maximum number of Indian Restaurants, we will focus our search on neighbourhoods within these two clusters.

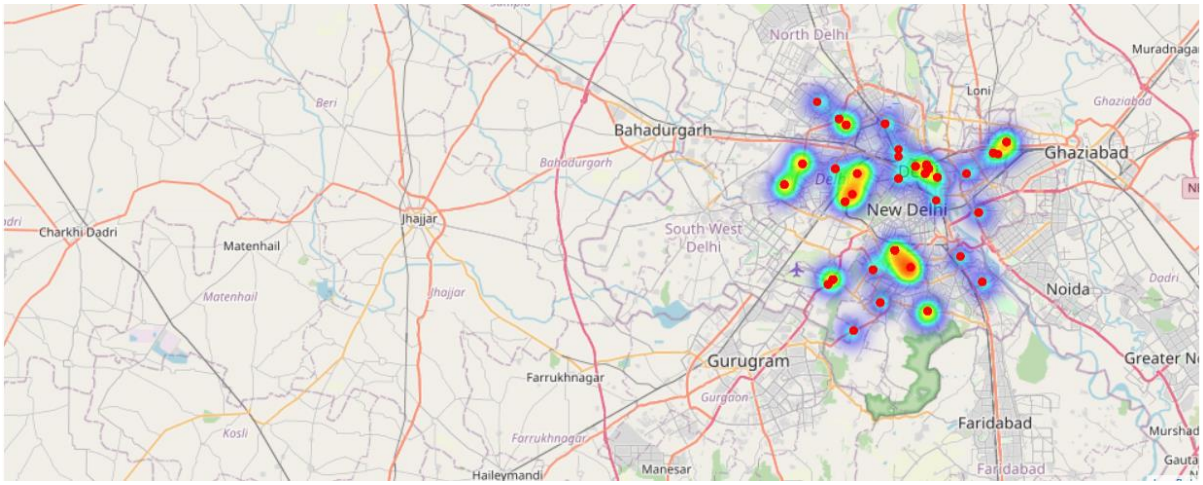
Why?

We know that when we were clustering the neighbourhoods the data used contained the mean of all types of restaurants present in the particular neighbourhood. Therefore, we can say that the neighbourhoods are clustered on their restaurant trends.

Now, clusters 2 and 3 may collectively have the highest number of Indian restaurant but there will be some neighbourhoods in these clusters which would have a demand for Indian Restaurants, as these neighbourhoods are in the same cluster, but would not have enough supply.

Results

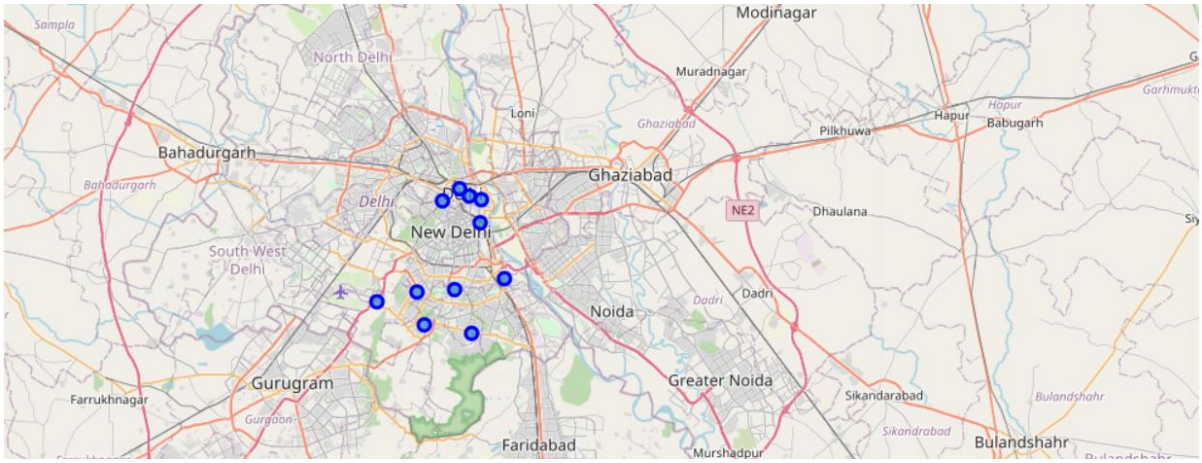
Our Analysis was done on over 113 neighbourhoods, containing over 848 restaurants within 2km radius of every neighbourhood. We segregated these neighbourhoods on the basis of types and amounts of restaurants. Five clusters were obtained, each having a unique collection of restaurants. Since, we were focused on finding optimal neighbourhoods for opening Indian restaurants, we selected cluster 2 and 3 which had the highest number of Indian restaurants. The above actions left us with the only those neighbourhoods that had a shared characteristics of and that had a high demand for Indian restaurants.



Next, we plotted a heat map for analysing the density of restaurants in the remaining neighborhoods. This allowed us to select neighborhoods that had few or no Indian restaurants and were not overcrowded by other kinds of restaurants. A total of 57 neighborhoods were left. After this, we found out the top three most popular neighborhoods(namely: Canaught Place, Hauz khas Village and Khirki Village), and the distance of every remaining neighborhoods from all three of them. Then, we extracted top 5 closest neighborhoods from each of three most popular neighborhoods mentioned above. Taking the union of the resulting three dataset we get 11 neighborhoods that satisfy all three conditons layed out in the business problem by the customer.

	index	Neighborhood	latitude	longitude
0	58	Connaught Place	28.631383	77.219792
1	95	Hauz Khas Village	28.553855	77.194713
2	103	Khirki Village	28.529885	77.218077

	Neighborhood	latitude	longitude
0	Chawri Bazaar	28.649927	77.229788
1	Pragati Maidan	28.623459	77.242512
2	Jhandewalan	28.644336	77.199927
3	Daryaganj	28.646090	77.243048
4	Lahori Gate	28.656841	77.218534
5	Gulmohar Park	28.557101	77.213006
6	Munirka	28.554886	77.171084
7	Mehrauli	28.521826	77.178323
8	Khanpur	28.512798	77.232395
9	Mahipalpur	28.544485	77.125691
10	New Friends Colony	28.567101	77.269764



Discussions

The neighbourhood recommendation obtained here are not completely accurate. This is due to the limitations in the dataset used in the project. Due to lack of cross-referencing sources, we may have missed a few neighbourhoods from our consideration. The foursquare API does not contain, or does not rely, a comprehensive dataset about the restaurants present in Delhi. Surely, in a city like Delhi with a population of over 19 million, there are much more restaurants than 848.