Coursera Capstone
IBM Applied Data Science Capstone

Exploring a place where opening a Shopping Complex would be more beneficial in Delhi NCR, India

By- Ajitesh Kohli

May 2 , 2020

1. Introduction

1.1 Background

Delhi NCR (National Capital Region) in India comprises of holds true as the cultural hub for the youth central location – it has a lot of companies and businesses so a lot of youth from all over the country are here for jobs and much more. Prominent cities of NCR include Delhi, Ghaziabad, Faridabad and Gurugram(also known as Gurgaon)
If you love shopping, there is an abundant selection of stores in Delhi NCR. From the stand-alone stores to the malls, the variety of clothing and products seems endless. No matter what your budget, you will find a vendor which caters to your needs. Also there are a lot of eating joints, cafés, pubs, bars which attract people for their beautiful ambience and food.Many of them are located in shopping malls also.

1.2 Business Problem and Targeted Audience

A businessman might want to know where it would be best suitable for opening a shopping mall in Delhi NCR on the basis of the places where the audience footfall is the maximum. Large audience footfall means many brands would like to invest in that area to attract more customers.

2. Data Collection

The data for the analysis was being accumulated by using **Foursquare API.**
From Foursquare, the details of the venues: name, type of venue, geographical location , city were fetched.
The complete data was not available for the free Foursquare account.

The total data comprised of only 178 venues.

3. Methodology

3.1 Data Preprocessing

The available data was checked for missing values. There were around 5-6 venues whose cities were not mentioned so instead of imputing the value, the observations were dropped.

3.2 Data Analysis

According to the analysis of the available data, it was found that maximum venues are Restaurants, followed by Hotel, then Cafés and then shopping malls, followed by others as shown in the bar graph below.

Out of the total 8 big shopping malls available in the data, 4 are in Delhi and 4 are in Gurgaon.

Count of venues of each category
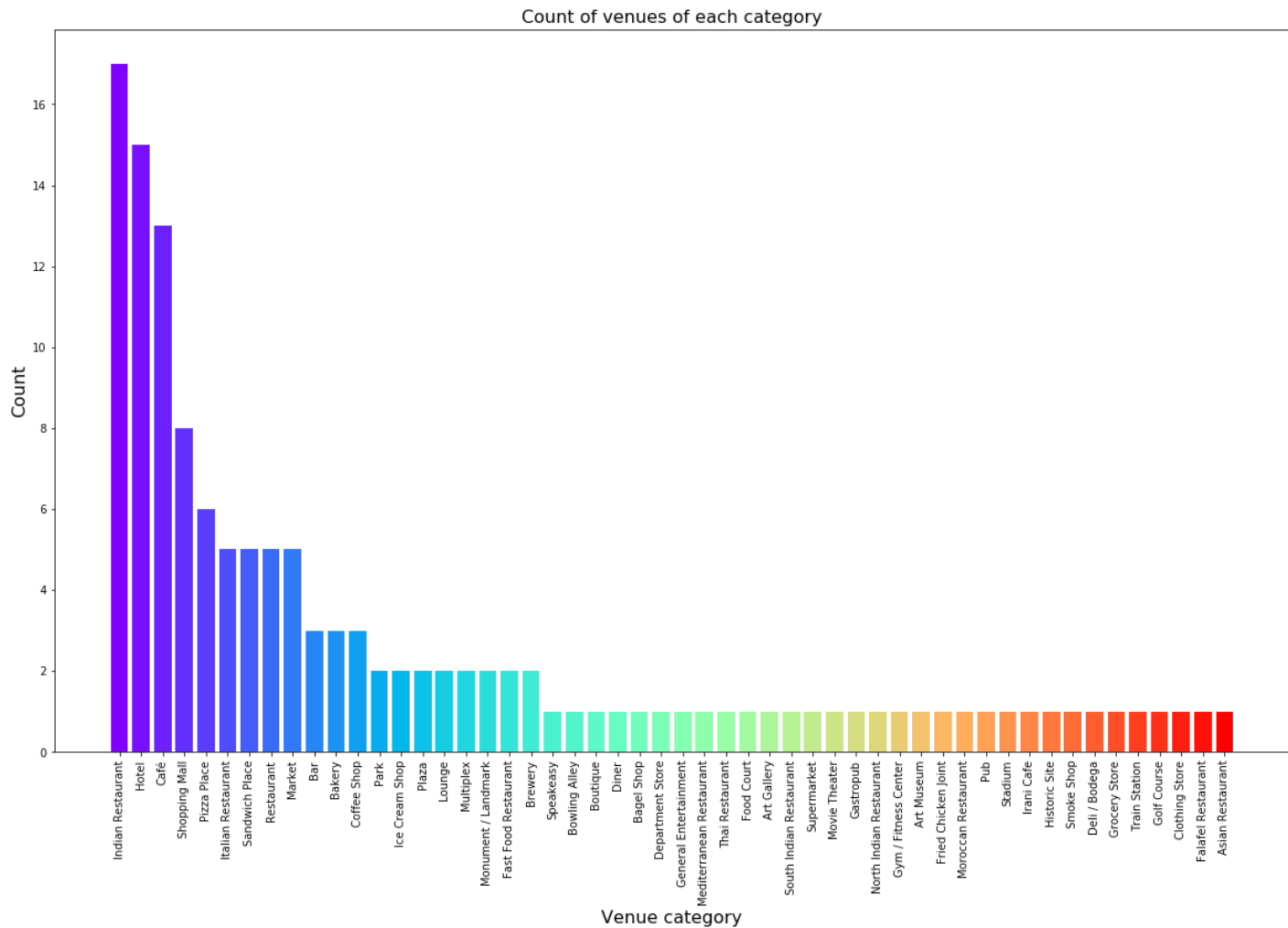
Now upon further analysis, the data was filtered by grouping on the basis of City and Type of Category.

The "categories" column was one hot encoded and then the total observations(rows) were grouped by city and the mean of the frequency of occurrence of each category.

The data was grouped into 7 regions(namely Delhi, Gurgaon, New Delhi, Faridabad, Noida, Sonipat and Ghaziabad).

It was further sorted in descending order on the basis of the occurrence in that region as shown below in the diagram.

|	city	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	Delhi	Indian Restaurant	Hotel	Speakeasy	Italian Restaurant	Lounge	Restaurant	Café	Department Store	General Entertainment	Gastropub
1	Farīdābād	Pizza Place	Café	Hotel	Deli / Bodega	Indian Restaurant	Fried Chicken Joint	Sandwich Place	Restaurant	Multiplex	General Entertainment
2	Ghāziābād	Food Court	Train Station	Thai Restaurant	Gym / Fitness Center	Grocery Store	Golf Course	General Entertainment	Gastropub	Fried Chicken Joint	Fast Food Restaurant
3	Gurgaon	Indian Restaurant	Café	Hotel	Shopping Mall	Sandwich Place	Pizza Place	Fast Food Restaurant	Coffee Shop	Market	Brewery
4	NCR	Hotel	Thai Restaurant	Gym / Fitness Center	Grocery Store	Golf Course	General Entertainment	Gastropub	Fried Chicken Joint	Food Court	Fast Food Restaurant

Lastly, we'll **cluster the venues** using K means clustering based on the available information of each venue. This will allow us to clearly identify which places have the largest number of venues and will be suitable for opening a shopping mall.

The data was organized in 5 clusters

Results and Discussion

Upon clustering , we got the different clusters.

Based on the data available, Delhi region has the most similar cluster with maximum number of venues, followed by Gurgaon, Faridabad and Greater Noida.
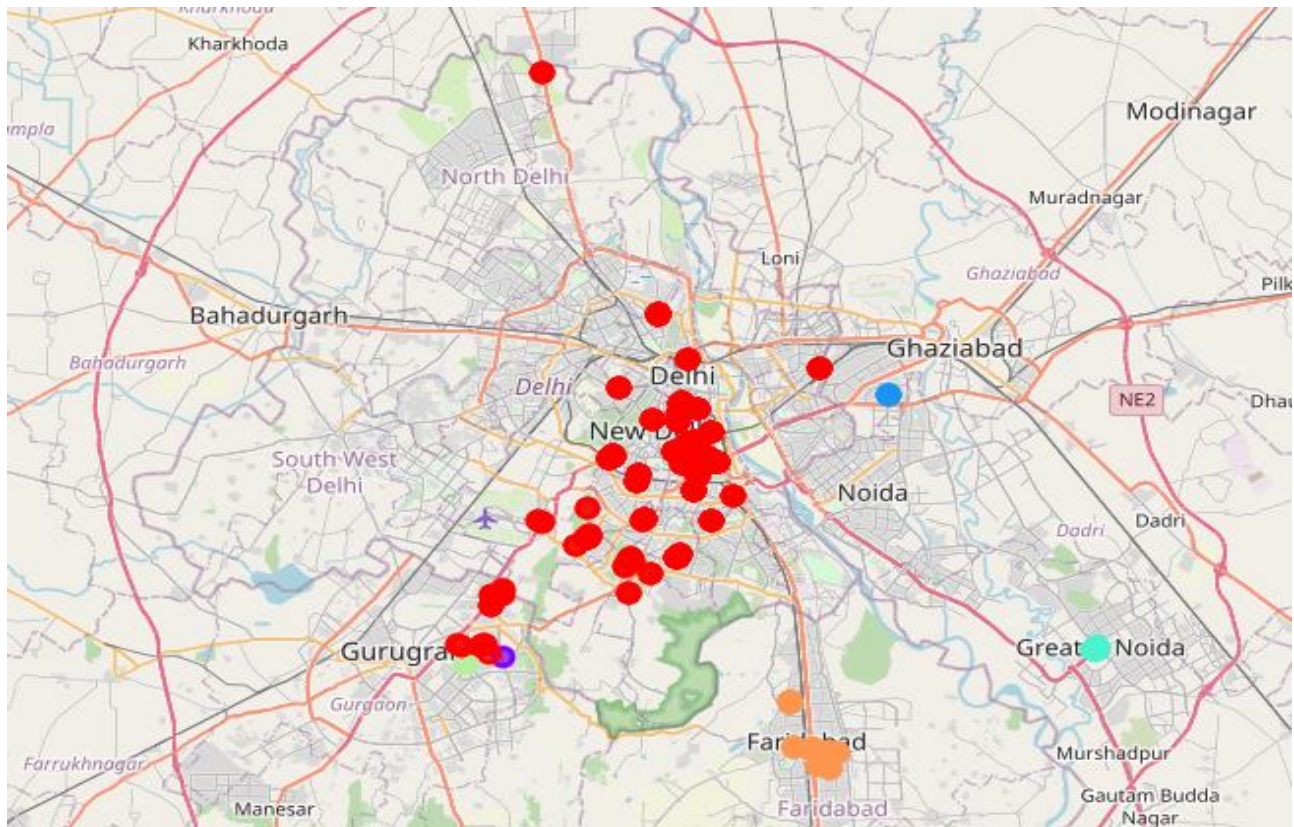
Conclusion and Future Direction

So for a business establishment according to the data in hand, it would be advisable to open a new mall in either Delhi region or Gurgaon region.

If a more elaborative data is available, results can be further improved.