

# **Restaurant Location Analysis in Delhi-NCR**

Divyanshi Paliwal

October 1, 2019

## **1. Introduction**

### **1.1 Background**

A business in food is seldom a failure, especially in a country like India, where food is a religion. People in Delhi are fond of food and love to try different cuisines or restaurants; this is why on every street of Delhi, one can find a restaurant or a street-food vendor. There are already many restaurants in the region, but still a new restaurant with a different style, theme or cuisine opens up every month and people rush to try it out. As a restaurant owner, it is important to identify a suitable location to open a restaurant to maintain the customers and ensure profitability. An ideal location can ensure visibility of the restaurant, which is important to attract customers, and helps in maintaining the supply for the raw materials. For example, a place where people go often will be better than the place where the footfall is lesser.

### **1.2 Problem**

To identify a suitable location for a restaurant, it is important to have information on the places available, cuisines and type of restaurants. Using data related to these parameters, the project aims to find suitable locations for a restaurant, what type of restaurant will be better at these locations and cuisines that the restaurant can provide.

### **1.3 Interest**

The analysis is of interest to restaurant chains and restaurant owners who want to open a new restaurant in Delhi-NCR and aim to identify suitable location and what type of restaurant will work for the location.

## **2. Data Acquisition and cleaning**

### **2.1 Data Acquisition**

The data used in the project has been taken from one of the publically available Zomato Dataset on Kaggle. Zomato is a restaurant finder and online booking application operational in India.

However, the datasets are not exhaustive and misses many restaurants. Nevertheless, has enough data to be analysed (close to 5000 rows for Delhi, Noida and Gurgaon combined).

## **2.2 Data Cleaning**

Data downloaded has records of restaurants throughout India. Since the study focuses only on Delhi-NCR, the data needed to be filtered. For simplicity, in Delhi-NCR I have considered New Delhi, Noida and Gurgaon only. According to the Indian Government other cities like Sonapat, Greater Noida, Indirapuram etc. also belong to Delhi-NCR, but with ground knowledge it can be easily said that New Delhi, Noida and Gurgaon have more restaurants, thus it was apt to consider only these three cities as a part of Delhi-NCR.

The dataset contained comma separated locality or locations. By looking at the data, it could be seen that the value after the comma was the main locality and before the comma was not needed, hence the column 'locality' was modified accordingly. Also, the names were not consistent, for example – 'Greater Kailash 1 (GK)' and 'Greater Kailash (GK) 1', they both address to the same locality and hence needed to be corrected. The column 'highlights' contains a set of all the features provided by the restaurant, though these features were not needed as such, but their count is important to know if more number of features attract more customers. Thus, the column was modified to contain the number of features each restaurant offers. Another column, 'establishments' contained values in square brackets, therefore the brackets need to be removed as they were not adding to the data.

After all this, it was observed that some of the restaurants have not specified their 'establishment' (the type of restaurant). These rows were not removed as they were needed for other analysis, however, the empty value was replaced with a hyphen.

Also, to get latitude and longitude data Foursquare API was used and appended to the dataset.

## **2.3 Parameter Selection**

After cleaning the data, it was observed that the dataset contained columns, which were not needed as they were not adding to the data and hence, they needed to be removed. The table below shows what parameters were retained and which parameters were removed.

**Table 1: Parameters to Retain**

<i>Parameter</i>	<i>Reason to retain</i>
Name	It should be kept for identification only, though it does not add to the analysis.
Establishment	This is required to identify type of restaurant that may work in a locality.
City	For identification if a locality belongs to New Delhi, Noida or Gurgaon
Locality	This is required for identification of a suitable area to open a restaurant
Address	This is required for getting Latitude and Longitude data
Cuisines	This is required to identify what cuisines are popular
Average_cost_for_two	This is required to analyse suitable price range
Highlights	This column shows the features provided and was needed for analysis
Aggregate_rating	This is needed to identify highly rated localities and other impacts
Votes	This column is synonymous to number to people visited, which is needed for analysis

**Table 2: Parameters to Remove**

<i>Parameter</i>	<i>Reason to remove</i>
Res_id	This parameter has no value in the analysis
url	This parameter has no value in the analysis
City_id	City is enough to distinguish between places and city_id is redundant
Zipcode	This parameter has no value in the analysis
Country_id	This parameter has no value in the analysis
Locality_verbose	This is a redundant column
Timings	This parameter has no value in the analysis
Currency	This parameter has no value in the analysis
Rating_text	This is a redundant column
Photo_count	This parameter has no value in the analysis
Opentable_support	This is a redundant column
Delivery	This is a redundant column
Takeaway	This is a redundant column

### 3. Methodology

I shall now be performing analysis on the parameters selected above.

#### 3.1. Exploratory Data Analysis

##### 3.1.1 Types of restaurants in Delhi-NCR

Let us see what are the major types of restaurants in Delhi-NCR. This will give us an idea on what type of restaurants can be opened by the restaurant owner or investor. On analysis, the following plot was obtained –

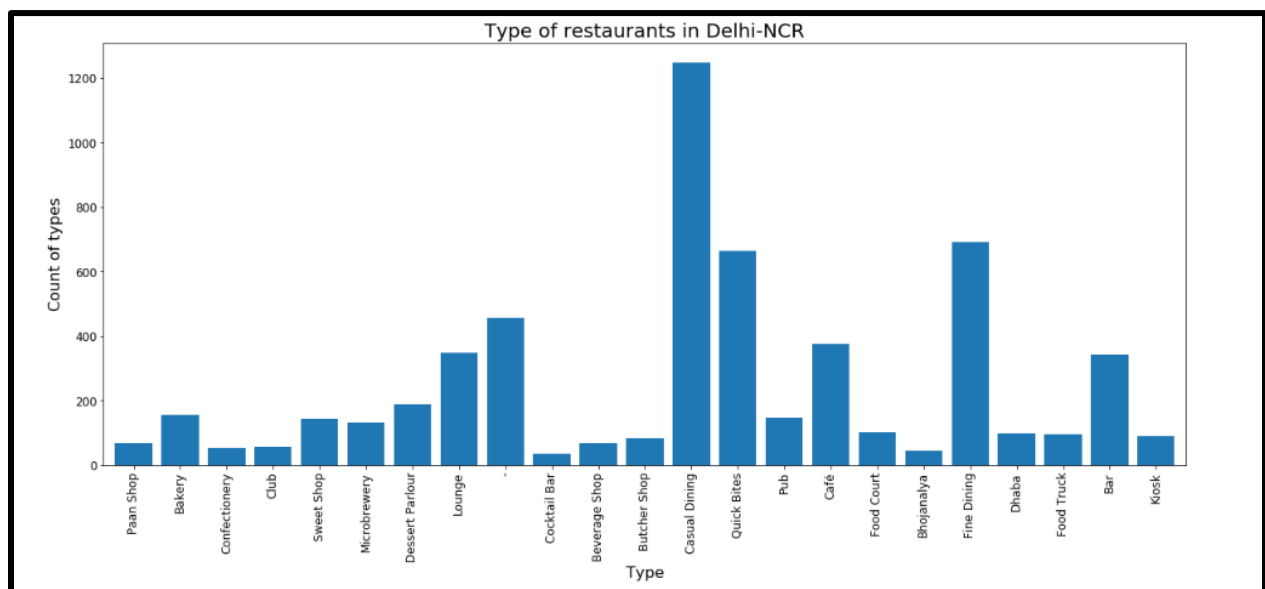


Figure 1: Bar Plot for popular restaurant types in Delhi-NCR

The bar plot above shows that in Delhi-NCR, Casual Dining is the most popular type of restaurant, followed by Fine Dining and Quick Bites.

##### 3.1.2 Popular cuisines in Delhi-NCR

Delhi-NCR has many cuisines available, let us see what cuisines are preferred the most by the restaurants in the area. The cuisine preferred the most by restaurant owner is directly linked with the preference of the people living in the region. To see popular cuisines, a word cloud has been created as follows –



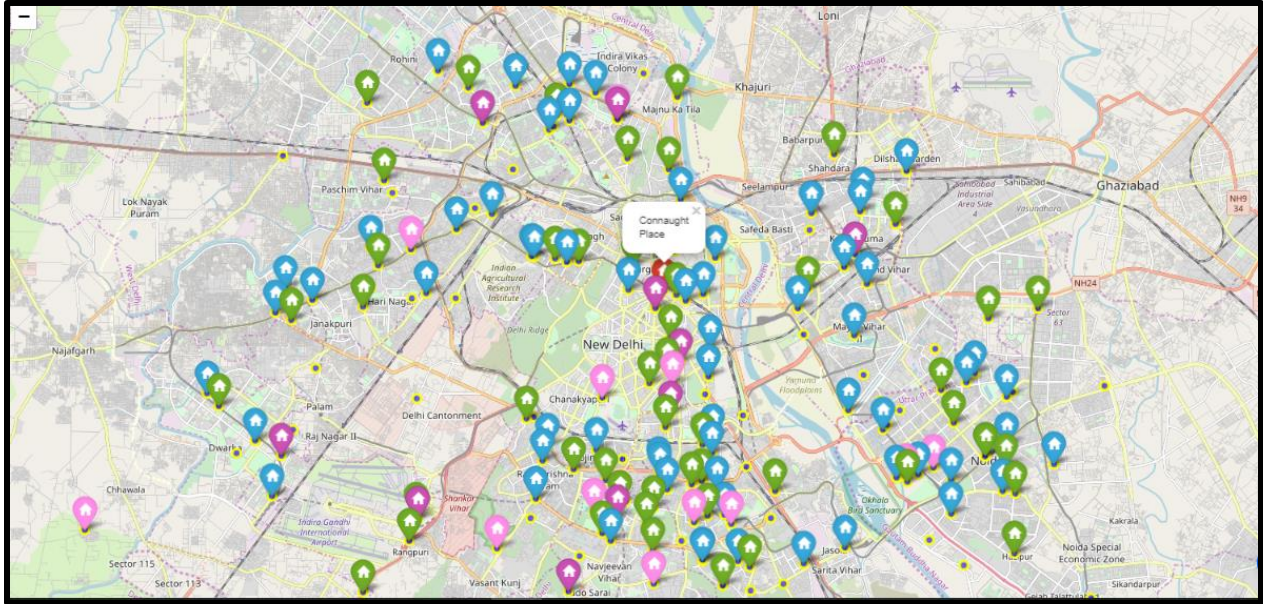
Figure 2: Word Cloud for Popular Cuisines

The word cloud shows that North Indian cuisine is preferred by the people in the region. Delhi-NCR is a region in the northern part of India, hence the liking for North Indian food seems apt. Also to note, Fast Food and Indian Chinese are other cuisines which are preferred by the people, which means that many restaurants provide these cuisines.

### 3.1.3 Locations and respective footfalls

The column 'votes' in the dataset can give an indication of footfall – more the votes, more people have been to that restaurant. In order to get this data, firstly, the dataset was grouped according to the locality and the total number of votes for each locality was counted. On performing such operations, the data obtained was plotted on map as shown below.

The markers for the localities have been colored, where 'Red' symbolizes footfalls in the range of 5,00,000 – 15,00,000. 'Pink' shows localities with footfalls in the range of 1,00,000 – 5,00,000, 'Purple' depicts localities with votes in the range of 50,000 – 1,00,000, 'Green' shows localities with footfall in the range of 10,000 – 50,000 and 'Blue' shows localities with footfalls less than 10,000.



*Figure 3: Map showing localities with footfalls*

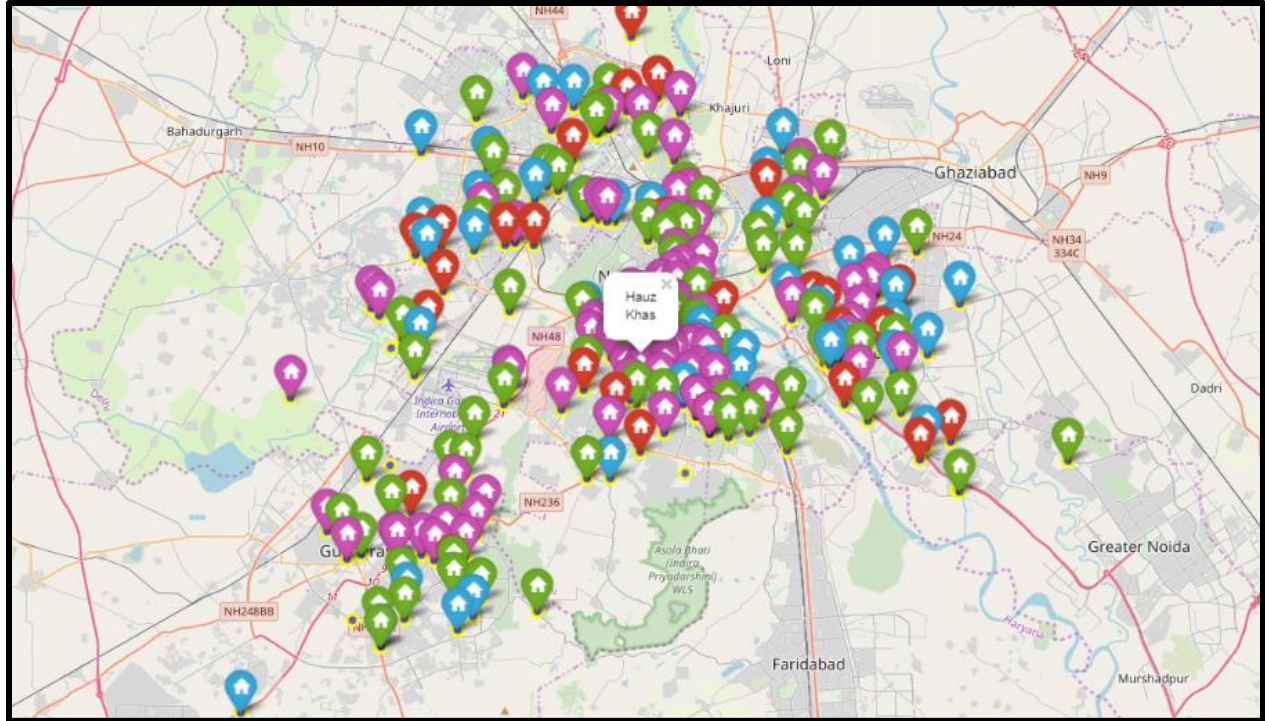
As can be seen from the map, there is only one 'Red' marker, which is for the locality – Connaught Place. On further analysis, it was found that Connaught Place has more than 13,00,000 recorded footfalls. This means that people tend to go to Connaught Place a lot as compared to other regions.

### **3.1.4 Locations with high rated restaurants**

Now, let us categorize locations on the based on average ratings of the restaurants in that location. In order to perform this, the dataset was first grouped by the location and the ratings of restaurants were averaged. The data obtained was plotted on a map as shown below.

Each location has been marked with a colored marker and these colors have a significance. 'Red' marker symbolizes locations with average rating between 1 and 2.99, 'Blue' color signifies location with average rating between 3.0 and 3.6, 'Green' color signifies locations with average rating between 3.61 and 4.0, and 'Purple' color symbolizes locations which rating between 4.01 and 5.0, which are the highest category of ratings. All these ratings are out of a scale of 5.0.





*Figure 4: Map showing localities with average restaurant ratings*

From the above map, it can be inferred that in regions in and around southern part of Delhi, there are restaurants with high average ratings and a lot of 'Purple' markers can be seen crowding the space. The southern part of Delhi has a low number of low-rated restaurant localities.

### **3.1.5 Impact of Cost on Ratings**

For a restaurant, it is also important to know what should be the ideal cost of food that they must offer in order to maintain the customer and increase the footfall. To analyze this, let us see if cost affects the rating of a restaurant. In order to perform such operation, average cost for two for each restaurant and corresponding ratings were analyzed. The box plot below shows the range of cost plotted against the ratings.

By looking at the box plot it can be inferred that the ideal range of cost is between 1000-2000 as the ratings have increased till that point and as the cost increased beyond 2000 for two, the ratings have seen a decline.

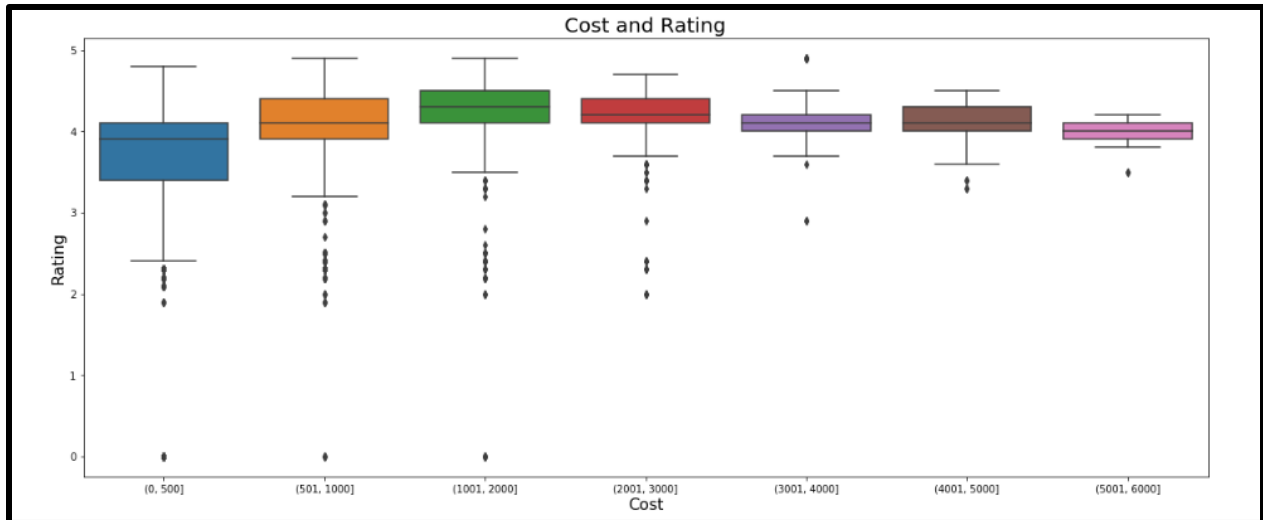


Figure 5: Box plot of Average cost vs Rating

### 3.1.6 Relation between features and footfall

A restaurant may provide various features like Credit/Debit card acceptance, Air Conditioner, Wheelchair accessibility, Alcohol and many more. Do more number of features attract more customers? The answer to this question may help the restaurant owner to select the number and type of features they want to provide to their customer. For answering this question, let find a correlation between the features and the votes or footfall for the restaurants in Delhi-NCR. The graph below depicts the relation between the two-



Figure 6: Plot to show relation between features and number of customers or footfall



The correlation statistics for the parameter are as follows –

The Pearson Correlation Coefficient is 0.3610830067884935 with a P-value of  $P = 2.364567397697596e-174$

From both Figure 6 and the value of correlation coefficient, it can be inferred that there is no strong correlation between the two parameters, but the p-value is less than 0.01, therefore, relation between the two cannot be disregarded altogether. Hence, number of features play a limited role in attracting customers.

## 4. Clustering and Further Analysis

### 4.1 Forming Clusters

Now, let us divide all the locations in 4 clusters on the basis of the type of restaurants available in the location.

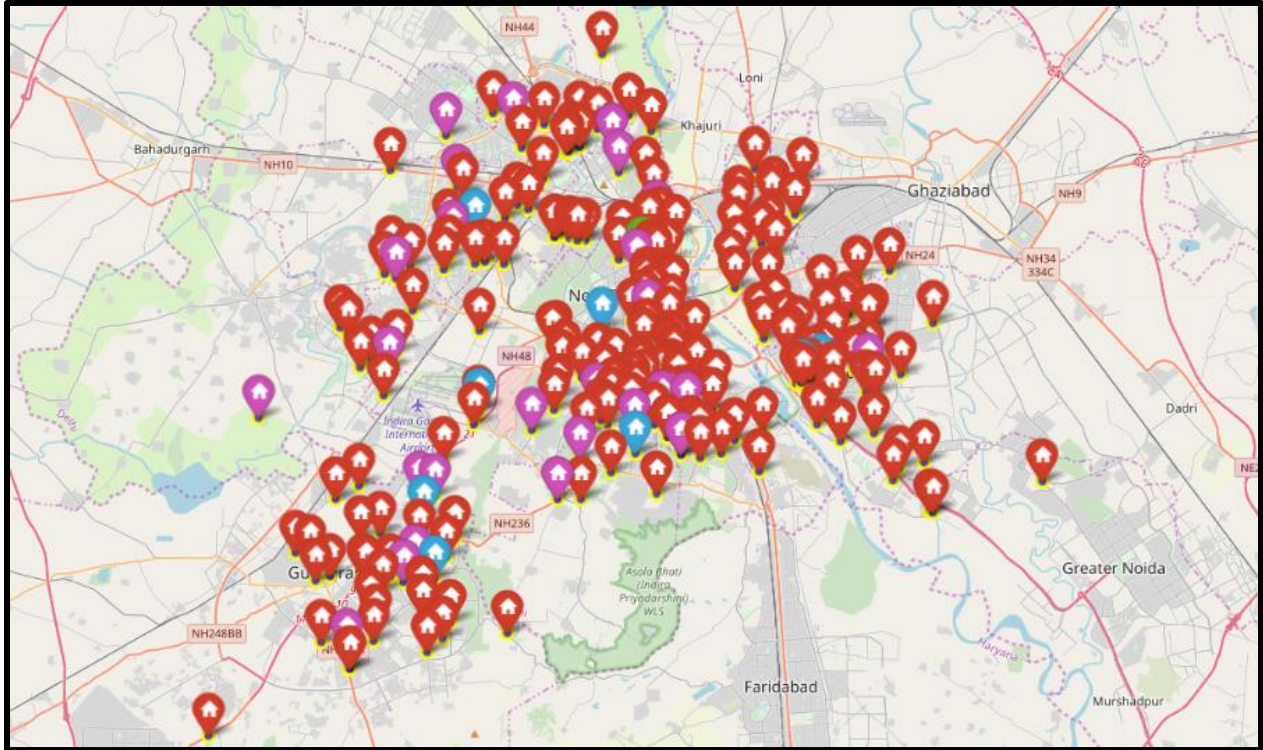
To perform this, K-means clustering was used with  $K=4$  and the dataset was group by location and a new dataframe was created as shown below –

		Microbrewery	Casual Dining	Lounge	Cocktail Bar	Pub	Café	Fine Dining	Bakery	Food Court	...	Beverage Shop	Sweet Shop	Dessert Parlour	Food Truck	Bhojanalya	Butcher Shop	Quick Bites	Bar	Kiosk
locality																				
Adchini	12	0	0	2	0	0	6	0	0	0	...	0	0	0	0	0	0	3	1	0
Alaknanda	4	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	1	1	0	0
Amar Colony	12	0	0	0	0	0	0	0	0	0	...	1	0	2	0	0	0	1	0	1
Anand Lok	10	0	10	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Anand Vihar	15	0	0	0	0	0	12	0	0	0	...	1	1	0	1	0	0	0	0	0
Ardee City	4	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0
Ashok Vihar Phase 1	8	0	0	0	0	0	0	0	2	0	...	0	0	1	0	0	2	2	0	1
Ashok Vihar Phase 2	18	0	4	0	0	0	0	0	0	0	...	0	0	0	0	0	0	14	0	0
Ashok Vihar Phase 3	3	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	2	0	0
Aurangzeb Road	27	0	0	3	0	0	1	22	1	0	...	0	0	0	0	0	0	0	0	0
Azadpur	11	0	0	0	0	0	0	0	0	0	...	0	0	0	0	2	2	6	0	0
Barakhamba Road	25	0	1	0	0	0	0	19	0	0	...	0	0	0	0	0	0	0	0	0
Bhikaji Cama Place	21	0	0	0	0	0	2	18	0	0	...	0	0	0	0	0	0	0	0	0

Figure 7: Dataframe showing locations and type of restaurant in each location.

Please note that the above dataframe snapshot is not exhaustive as this dataframe also consists of other columns like latitude and longitude data.

After getting the above dataset, using k-means clustering, the clusters obtained were plotted on map shown below. Each cluster has been color coded with 'Red' symbolizing cluster 1, 'Blue' symbolizing cluster 2, 'Green' symbolizing cluster 3, and 'Purple' symbolizing cluster 4.



*Figure 8: Map showing the four clusters*

From the above map, it can be seen that cluster 1 is the densely populated cluster with many locations clustered together. Cluster 3 has only one location, which is Connaught Place.

For further analysis to find a location for opening a restaurant, I shall focus only on cluster 1, as it is the most densely populated cluster.

#### **4.2 Dominant restaurant types in densely populated cluster**

Let's identify what type of restaurants are dominant in the cluster 1. For this the data in cluster 1 is plotted in form of a bar graph as shown below.

From the graph, it can be observed that the most dominant restaurant type is Quick Bites, followed by Casual Dining and Fine Dining. Though, these types are same as what was observed for whole Delhi-NCR region before dividing into clusters, but there is change in the order of dominance.

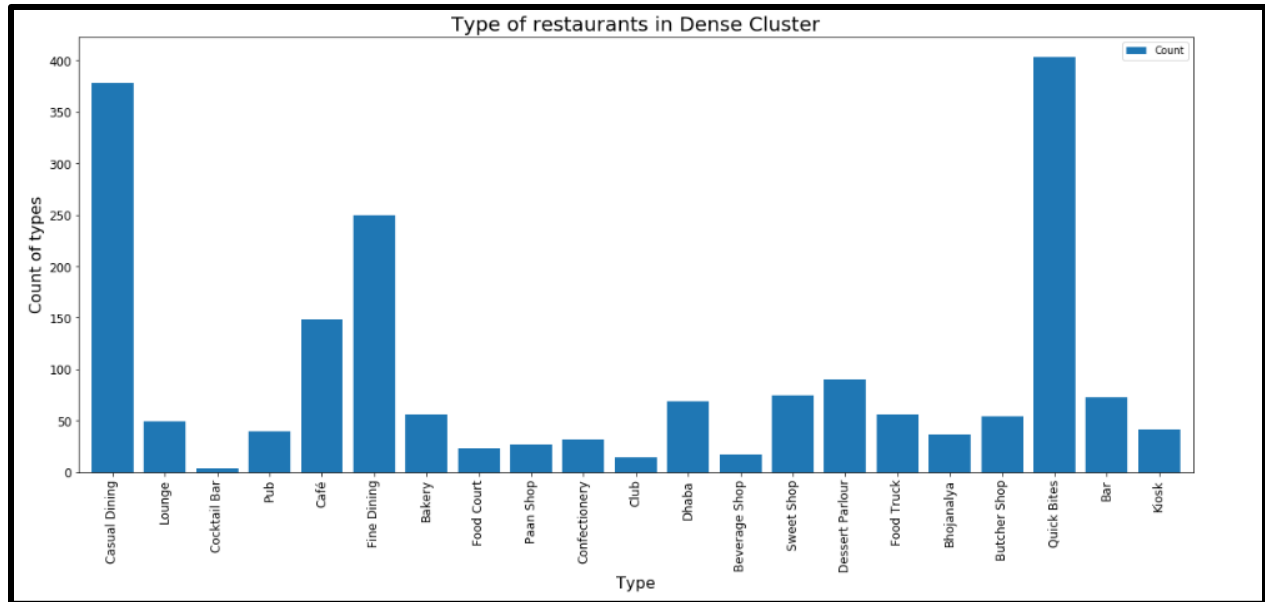


Figure 9: Bar Plot showing the different restaurant types

#### 4.3 Locations with highest number of restaurant in identified restaurant types

To further analyze the cluster and identifying an appropriate location, I would identify locations which have more number of restaurants in the Casual Dining, Quick Bites and Fine Dining categories, as compared to other locations in the cluster. The graph below shows the locations with highest number of restaurants. For simplicity only those locations which have more than 15 restaurants have been considered.

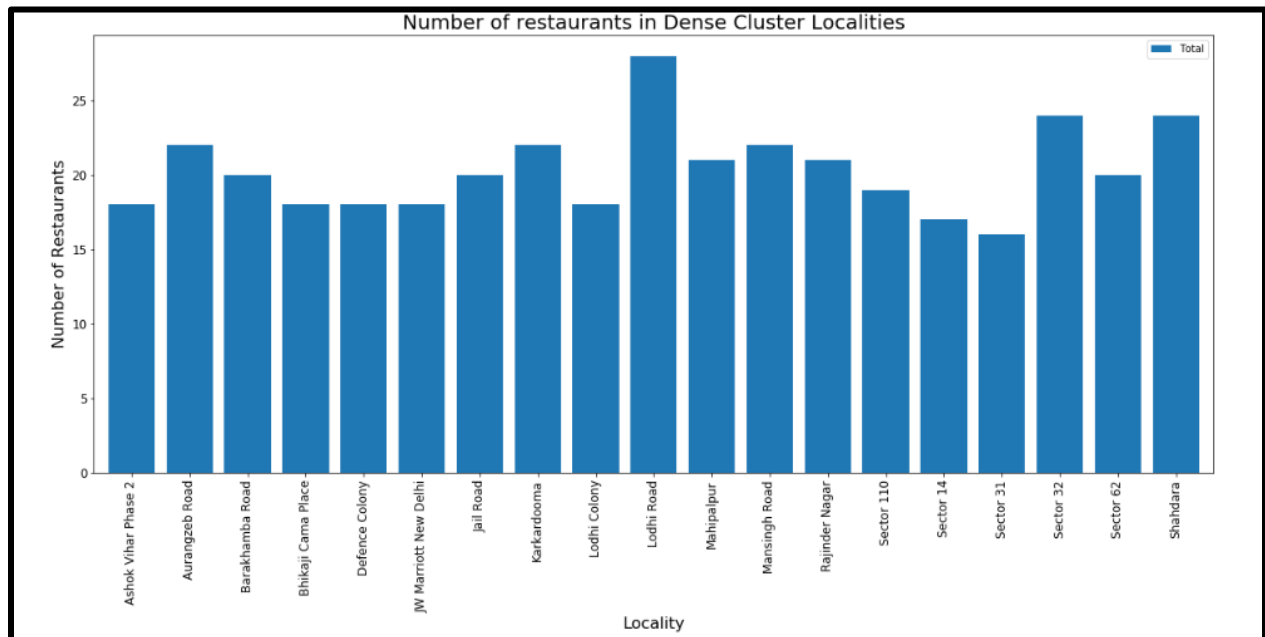
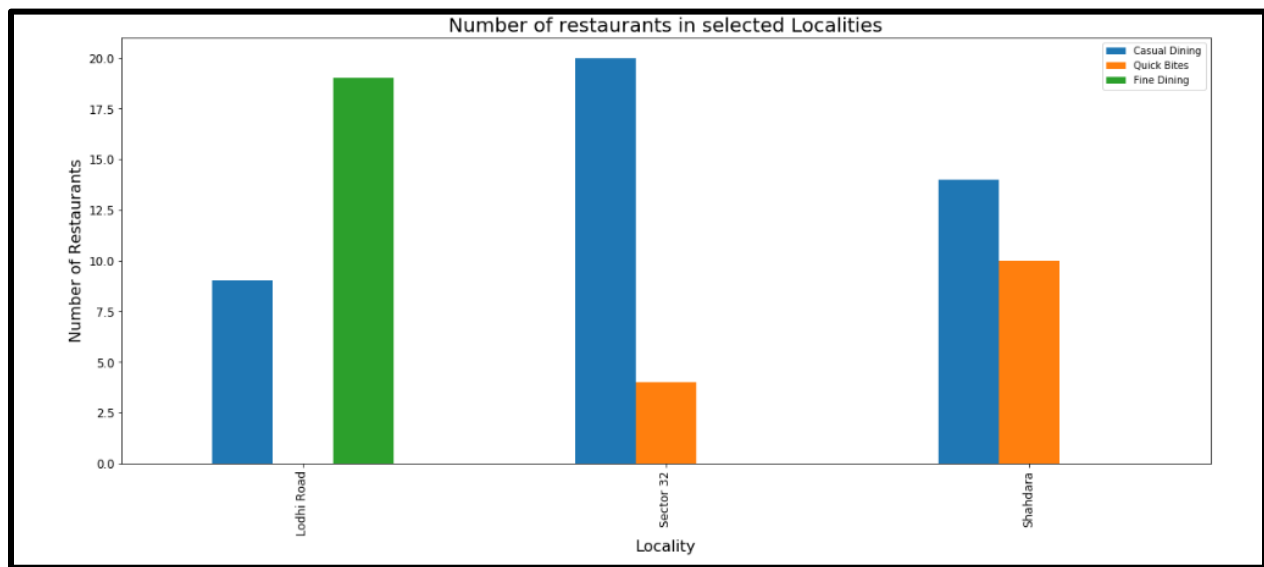


Figure 10: Bar Plot showing the locations with high number of restaurants in Casual Dining, Fine Dining and Quick Bites categories

From the graph, Lodhi Road has the highest number of restaurants in the identified restaurant types, followed by Sector 32 and Shahdara.

Now, with these three locations, let's see what type of restaurants are common in these locations. This has been shown in the graph below –



*Figure 11: Bar Plot showing the locations with popular type of restaurants.*

From the above figure, it is observed that in Lodhi Road, Fine Dining is the preferred restaurant type. This is because, Lodhi Road is a posh area and people there prefer a luxury dining experience. In Sector 32 of Noida, Casual Dining is preferred, but in Shahdara both Casual and Quick Bites are preferred, Casual Dining a little more than Quick Bites, this is because in Shahdara middle-income population thrives who would prefer an economical casual dining.

#### **4.4 A different case – Connaught Place**

In the earlier analysis, it was identified that Connaught Place has more number of footfalls than any other location and while clustering, Connaught Place was identified as a separate cluster altogether.

Let's see what made Connaught Place as a separate cluster (Cluster 3) by diving into the popular type of restaurants available in the area. The graph below shows the different restaurant types available in CP (as Connaught Place is famously called) –

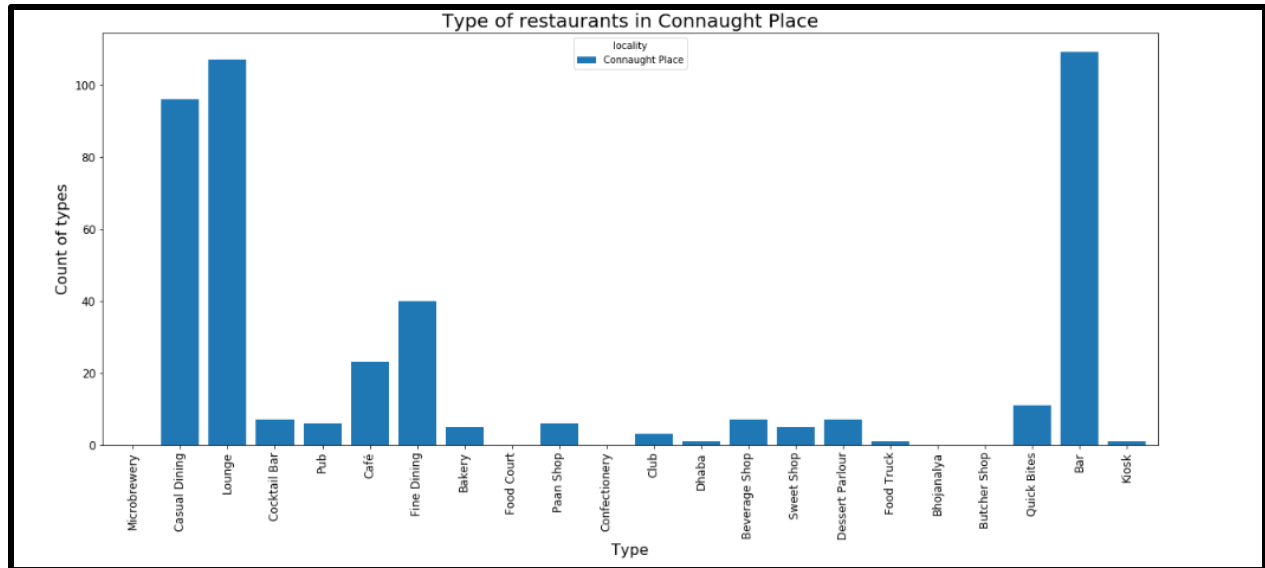


Figure 11: Bar Plot showing the popular type of restaurants in Connaught Place

The above graph shows that Connaught Place will be an ideal location if one wants to open a Bar or a Lounge. Alcohol and music is probably the reason why the footfalls in Connaught Place is so high!

## 5. Results

In Delhi-NCR the most popular cuisine is North Indian, which is not a surprise since the region lies in the north of India. Though, people also prefer to have Chinese and fast foods. Reflecting on the cuisine preferred by the population of the region, the dominant types of restaurant are Casual Dining, Fine Dining and Quick Bites. It was also observed that Connaught Place has the highest number of footfalls as compared to other locations in Delhi-NCR. It is rightly so as Connaught Place lies at the center of Delhi and has interconnections to all the major stations within the city. Also, the place has been a commercial hub since pre-independence days. It was also observed that cost has an impact on ratings. Also, the role of features in attracting customers could not be completely discarded. It was also observed that most of the highly rated restaurants are in the southern part of Delhi.

## **6. Conclusion**

After all the analysis, it can be said that to open a restaurant ideal place will be Lodhi Road if the type of restaurant to be opened is Fine Dining as the people in the region are high-incomed and prefer fine dining. If choice of restaurant is Casual Dining then either Sector 32, Noida or Connaught Place are better option than others. For Quick Bites, Shahdara can be preferred, but this place may not prove to be profitable due to limited income population in the area. If what a restaurant owner prefers to open is a lounge or a bar, then Connaught Place is the best place to open it.

Since costs have an impact on ratings, it is advisable to keep the cost between Rs. 1000- Rs. 2000 for two to make dining affordable but selective. Also, the most popular cuisine is North Indian, the restaurant can opt to serve North Indian, along with Chinese, Continental or Italian, since these are clearly identifiable in the word cloud shown before.

## **7. Discussions**

The project is based on the data available and as stated before the data was not exhaustive and missed out on many restaurants, but still there was enough data to perform the analysis. The clustering was performed on only one criteria, different criteria such as proximity to each other or based on cuisine, the locations could have been clustered. In that case, the result would have been different than that shown above. Also, if data related to reviews for each restaurant could be arranged, it would have been possible to identify the popular dish in a location. Moreover, the data was segregated into many locations, these locations could have been clubbed into a few major locations and the analysis could have been performed using that.

Furthermore, selecting a location for a restaurant also depends on the land available, price of acquiring that land and other miscellaneous charges, a data related to such parameters would have helped in a proper analysis.