Where to eat in Gurgaon!!

Vivek Chauhan

1 Introduction

1.1 Background

Gurgaon ,officially known as Gurugram is a city located in the northern Indian state of Haryana. It is situated (19mi) to the southwest of national capital of New Delhi and is part of NCT(National Capital Territory of Delhi).As of 2011,Gurgaon had a population of 876900.

Today, Gurgaon has local offices for more than 250 Fortune 500 companies. Gurgaon is categorised as very high on the Human Development Index, with an HDI of 0.889 (2017), which is also the highest in India. Gurgaon has budding start-up culture and also has well established firms. It attracts bright minds from across the country and thus serves to diverse food demands of city denizens. Millennium City Gurgaon is famous for many things like popular Shopping Malls, Multiplexes, Pubs and Hotels etc. In Gurgaon, there are many small food joints also which are very popular with the local audience. People looking to rent/buy places in Gurgaon thus need to look carefully at the places surrounding it. Thus it will be advantageous to explore the locations based on food preferences, venues of interest around the locality.

1.2 Business Problem

To analyze and enable people to choose the best locality to stay based on their preferences in food. We will cluster the localities based on their similarities of food places and venues around it. Our task is to enable the natives to find common places around them, best and worst places for particular cuisines ,group localities on food places, average ratings etc and to educate new dwellers in the city to choose place of residence based on food preferences and venues of interest. Also if a new “North Indian” cuisine serving food outlet has to be opened, which locality is best suited for it can be answered after this analysis.

1.3 Target Audience

New Dwellers coming into the city looking for place to put up. Natives looking to find out best and worst places around them for particular cuisines. People looking to explore venues around them in city. Also the people looking to open places in Gurgaon would be interested to know the competitors and popular spots around.

2. Data Acquisition and Cleaning

2.1 Data sources

Most food places in Gurgaon are associated with Zomato and the dataset presents a comprehensive overview of these places. The columns of relevance in the dataset were namely Restaurant Name, Country Code, City, Address, Locality, Longitude, Latitude, Cuisines, Average Cost for two, Price range, Aggregate rating, Rating text. This dataset was used to explore answer questions related to the best and worst places for particular cuisines,the localities having maximum and minimum number of food places, etc. It was taken from *https://www.kaggle.com/shrutimehta/zomato-restaurants-data* . Then we used the Foursquare API (*https://developer.foursquare.com/*) developer account to explore the venues near all these places to cluster them. It enabled us to look into details the localities which might serve a particular cuisine.

2.2 Data Cleaning and Feature Selection

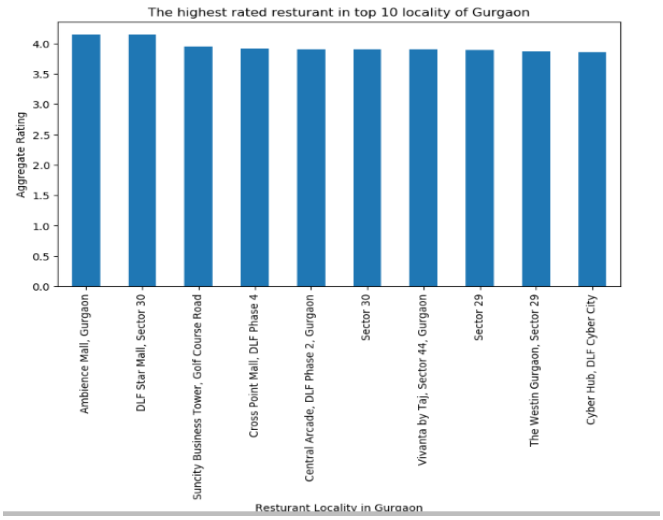
The Zomato dataset used contained the restaurants from multiple countries and multiple cities in India. Another .xlsx file containing the country code was used to bisect dataframe from master dataset. The Country Code for India used was 1 and city field contained the cities it served and Gurgaon was used to find the restaurants of interest. There were 9551 restaurants with 8652 being from India having 21 features to choose from in master dataset. Gurgaon is served by 1118 restaurants and is the dataframe of interest to us.

The redundant fields such as Locality verbose and Currency type did not serve any major purpose and hence were not used. Our columns of interest were Restaurant Name, Locality, Longitude, Latitude, Cuisines, Average Cost for two, Aggregate rating. Removed all places having Longitude value of 0 and Aggregate rating of 0 to clean our data.

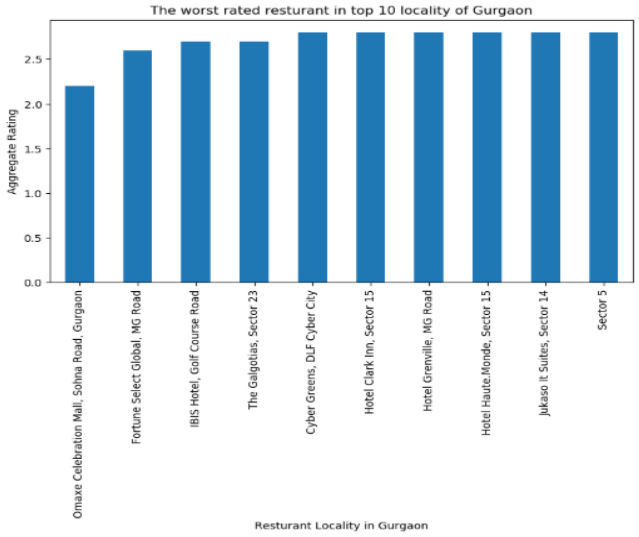
3 Methodology

3.1 Exploratory Data Analytics (EDA)

3.1.1 Finding the highest rated restaurant in top 10 locality of Gurgaon:



3.1.2 Finding the lowest rated restaurant in top 10 locality of Gurgaon:



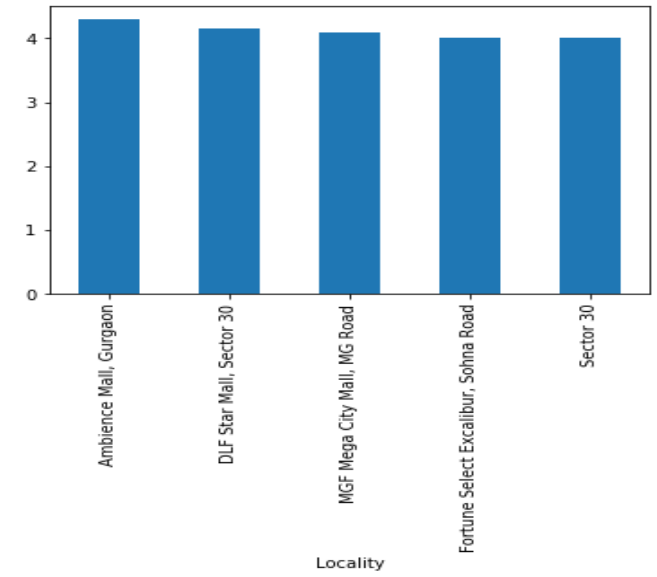
3.1.3 Finding which locality have most number of food places

Locality	
DLF Phase 3	68
Cyber Hub, DLF Cyber City	41
Sector 14	41
Sohna Road	41
Sector 29	38
Old Railway Road	36
Sector 31	30
Golf Course Road	27
Sector 56	26
MG Road	24

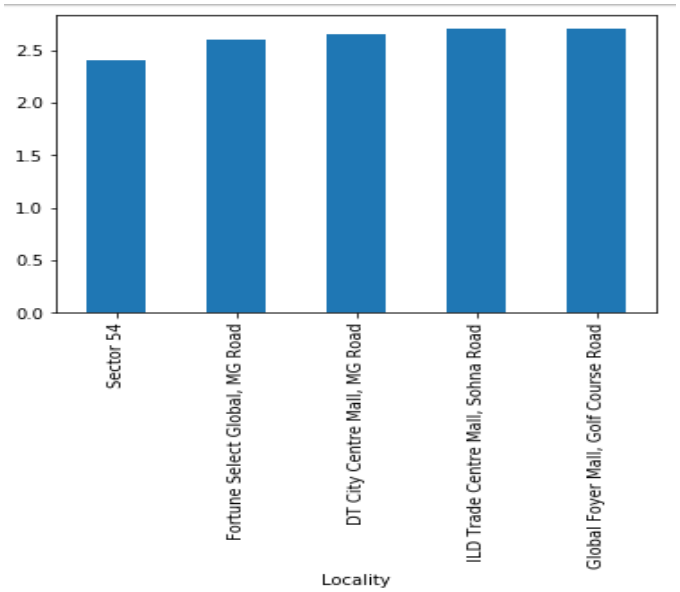
3.1.4 Finding which locality has minimum number of food places

Locality	
Central Arcade, DLF Phase 2, Gurgaon	1
Country Inn & Suites By Carlson, Udyog Vihar	1
Country Inn & Suites by Carlson, Gurgaon	1
Country Inn & Suites, Sohna Road	1
Courtyard by Marriott, Sushant Lok	1
Cyber Greens, DLF Cyber City	1
DLF Phase 2	1
Fortune Select Global, MG Road	1
Hotel Clark Inn, Sector 15	1
Hotel Grenville, MG Road	1

3.1.5 Finding localities having highest rated places which serve North Indian Cuisine



3.1.6 Finding localities having lowest rated places which serve North Indian Cuisines



3.2 k-means Clustering

Clustering can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as Euclidean-based distance or correlation-based distance.

K-means algorithm is an iterative algorithm that tries to partition the dataset into *K* pre - defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster. The objective function is:

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} ||x^i - \mu_k||^2 \qquad (1)$$

where wik=1 for data point xi if it belongs to cluster *k*; otherwise, wik=0. Also, μk is the centroid of xi's cluster.

It's a minimization problem of two parts. We first minimize J w.r.t. wik and treat μk fixed. Then we minimize J w.r.t. μk and treat wik fixed. Technically speaking, we differentiate J w.r.t. wik first and update cluster assignments (*E-step*). Then we differentiate J w.r.t. μk and recompute the centroids after the cluster assignments from previous step (*M-step*). Therefore, E-step is:

$$\frac{\partial J}{\partial w_{ik}} = \sum_{i=1}^{m} \sum_{k=1}^{K} ||x^i - \mu_k||^2$$

$$\Rightarrow w_{ik} = \begin{cases} 1 & \text{if } k = argmin_j ||x^i - \mu_j||^2 \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

In other words, assign the data point xi to the closest cluster judged by its sum of squared distance from cluster's centroid.And M-step is:

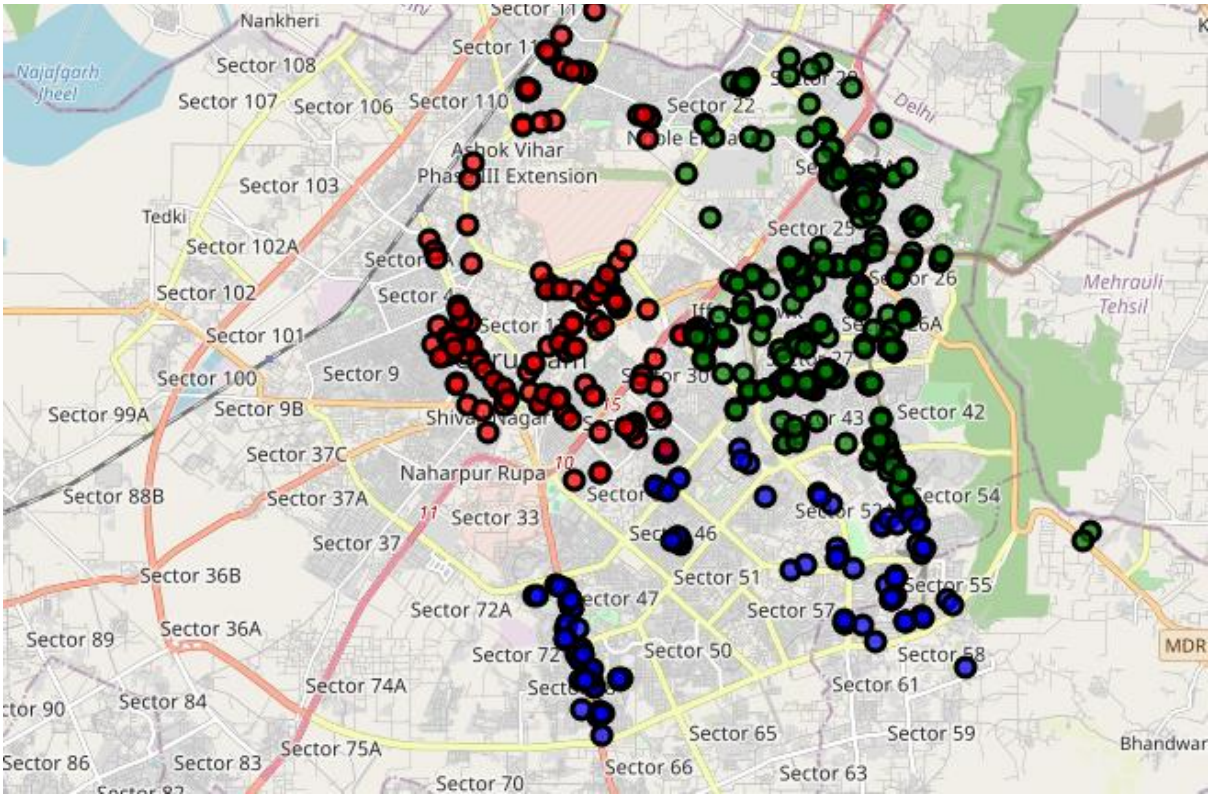
$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^{m} w_{ik}(x^i - \mu_k) = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{m} w_{ik} x^i}{\sum_{i=1}^{m} w_{ik}} \qquad (3)$$

Which translates to recomputing the centroid of each cluster to reflect the new assignments. We implement it using sklearn module kMeans class to fit our model.

4 Results

In first part of our analysis we clustered based on location, and in EDA we find out the best and worst places for Northern cuisines based on average rating number of places at a particular location.



Exploring the most common venues around a place educates the person looking to buy/rent a place or open a food place.

	Locality	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	ILD Trade Centre Mall, Sohna Road	Brewery	Movie Theater	Indian Restaurant	BBQ Joint	Shopping Mall	Fast Food Restaurant	Sporting Goods Shop	Multiplex	Department Store	Dim Sum Restaurant
1	Ambience Mall, Gurgaon	Indian Restaurant	Fast Food Restaurant	Italian Restaurant	American Restaurant	Café	Asian Restaurant	Multiplex	Shopping Mall	Bar	Clothing Store
2	Ansal Plaza Mall, Palam Vihar	Pizza Place	Gym	Sandwich Place	Café	Cosmetics Shop	Dairy Store	Deli / Bodega	Department Store	Dessert Shop	Flea Market
3	Ardee City	Convenience Store	Indian Restaurant	Thai Restaurant	Beer Garden	Donut Shop	Flea Market	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store
4	Best Western Skycity Hotel, Sector 15, Gurgaon	Hotel	Farmers Market	Asian Restaurant	Diner	Flea Market	Fast Food Restaurant	Falafel Restaurant	Electronics Store	Donut Shop	Dim Sum Restaurant
5	Central Arcade, DLF Phase 2, Gurgaon	Department Store	Indian Restaurant	Shopping Mall	Coffee Shop	Bar	Scenic Lookout	Brewery	Café	Donut Shop	Mexican Restaurant
6	Central Plaza Mall, Golf Course Road	Chinese Restaurant	Italian Restaurant	Falafel Restaurant	Hotel	Business Service	Sandwich Place	Café	Diner	Ice Cream Shop	Dim Sum Restaurant
7	Country Inn & Suites By Carlson, Udyog Vihar	Clothing Store	Hotel	Fast Food Restaurant	Campground	Donut Shop	Flea Market	Farmers Market	Falafel Restaurant	Electronics Store	Dim Sum Restaurant
8	Country Inn & Suites by Carlson, Gurgaon	Hotel	Japanese Restaurant	Resort	Café	Indian Restaurant	Go Kart Track	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store
9	Country Inn & Suites by Carlson, Sector 12	Hotel	Indian Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Donut Shop	Diner	Dessert Shop

In the second part we explored the venues around all locations and clustered them using k-means algorithm. We took k=3(number of clusters)

Cluster 1: Breweries, Indian Restaurant

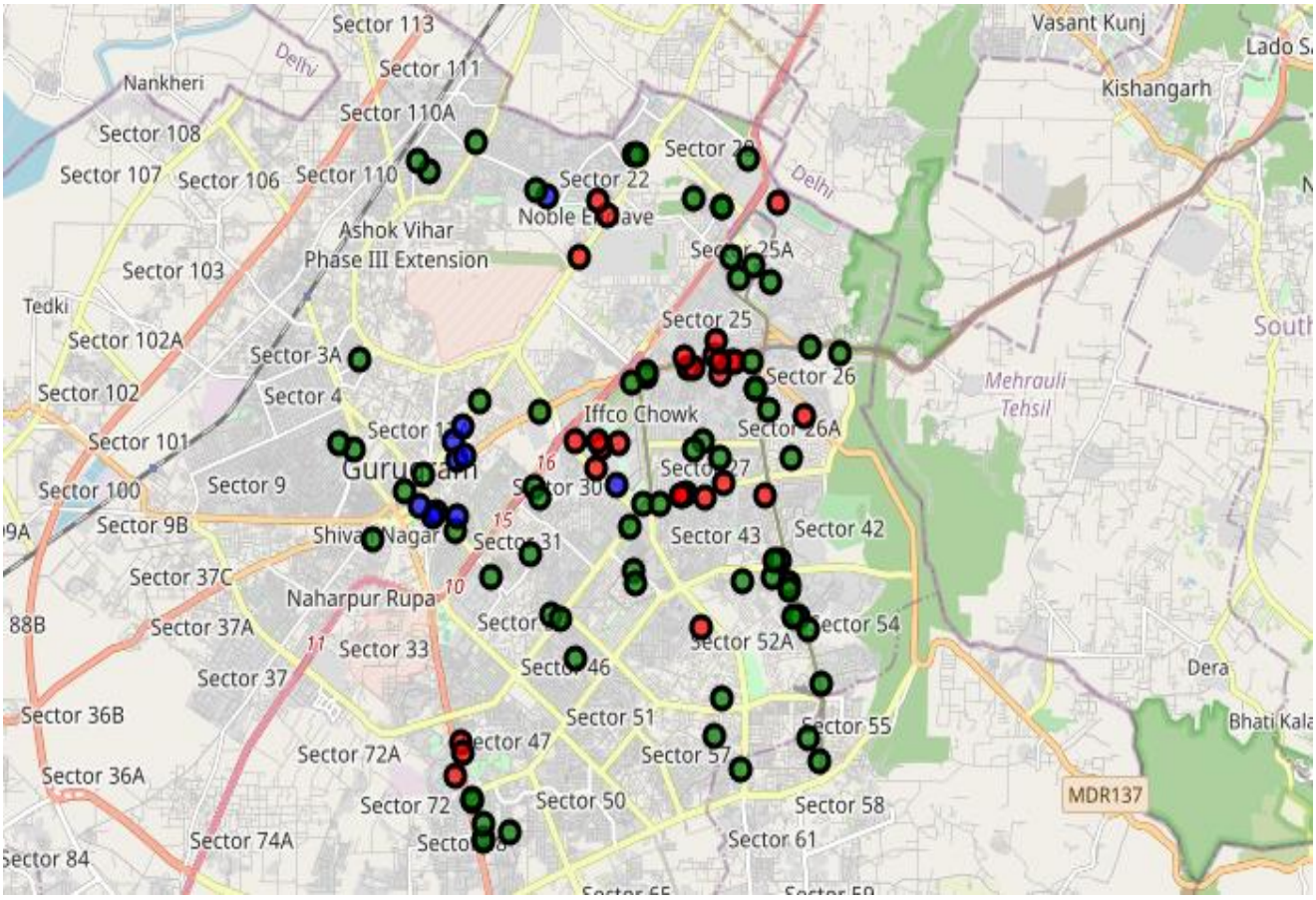
Cluster 2: Café, coffee ,Pizza Fast food, Chinese and Bars

Cluster 3: Hotels

Cluster 1: Green Color

Cluster 2: Blue Color

Cluster 3: Red Color



The clusters are formed on basis of similarity in cuisines available, location and common places near it. Hence if a person likes North Indian cuisines, he should rent a place in Cluster 1 around Sector 30 and ambience mall location and a person wanting to open a café place should target finding a place in Cluster 2 and cluster 3 is suitable for opening of hotels.

6 Conclusions

In this project, we have gone through the process of identifying the localities based on food places and later went on to cluster places in 3 clusters based on similarity on multiple factors discussed earlier sections. We were able to answer the questions pertaining to best and worst food places and also analysed the localities having the maximum restaurants. We finally made a dataframe with all common venues around locations to explore. A person looking to buy/rent place could now make an educated decision considering food preferences and places around him. The findings of this project will be better able to guide the stakeholders for their decisions. We performed EDA and clustering on these datasets in our pursuit of solutions. We were able to find satisfactory answers to the questions we posed before the study. The study is based on limited data, but it is nevertheless a significant step in shedding light on the hostel scene in Gurgaon. This study can be repeated easily for other cities of the world.

7 Future Directions

For this project we have taken the location of the place as the criterion for assessment. We could also consider the corporate offices, work places and places available for buying/renting to give a potential buyer/renter a comprehensive overlook because these are the two most essential factors one considers. Other data pertaining to cost of places, crime rate will also be relevant to suggesting place. Home delivery is also overlooked in this case since we did not have the ranges they would cover and as an approximation considered that they would be serving to nearby places, but the availability of such data would be highly beneficial. The comparative study of such data with other cities could also serve to people confused between two cities to move to. Effect of corporate and work force on food places could be an interesting study.