Analyzing the Neighborhoods in Delhi for Opening New Hotel and Paying Guest

Applied Data Science Capstone Project

By: AKSHAY KUMAR

Date: February-28-2021

Table of Contents

1.INTRODUCTION	3
2.DATA COLLECTION	3
2.1 NEIGHBORHOOD DATA OF DELHI 2.2 COORDINATES OF THESE NEIGHBORHOOD 2.3 VENUES IN THESE NEIGHBORHOOD	4
3.METHODOLOGY	5
3.1 DATA VISUALIZATION 3.2 FEATURE EXTRACTION 3.3 CLUSTERING ANALYSIS	6
4.RESULT	9
5. DISCUSSION	12
6.CONCLUSION	13

1.Introduction

Delhi, India's capital territory, a massive metropolitan area in Northern India spread in a Total area of 1484KM² and It is one the Famous and favorite area for tourists and Students due to its Top Universities like IIT and DU and tourist's places like Red Fort, Lotus Temple and others. This massive city is divided in 9 parts which are North West Delhi, North Delhi, North East Delhi, Central Delhi, New Delhi, East Delhi, South Delhi, South West Delhi, West Delhi and these are contains around 118 Neighborhood (sub-Areas).

Every year tourists and Students comes in heavy number in Delhi. In 2019 alone around 15 million tourists visited Delhi and around 200 thousand students came in Delhi University alone.

Now for staying and living these Tourists and Students needs Places to stay on one hand Tourists prefer Hotels because they stay for shorter period and on the other hand Students Prefer Paying Guests (PG) because PGs are cheap and good for long stay like for months or years.

Opening a new Hotel or Paying Guest in Delhi can be very Profitable But as earlier we have seen that area is massive divided in to 9 areas and 118 sub areas so choosing area for opening new Hotel or PG can be problematic without analyzing those area. Our Aim in this Projects is to Analyze the Neighborhood of Delhi based on Popularity of venues they have and find out which places are highly recommended for opening a new Hotel and Paying Guest (PG).

2.Data Collection

We need Multiple Data for this project Which are as follows: -

- 1. The Neighborhood data of Delhi.
- 2. Coordinates of these Neighborhood
- 3. Venues in These Neighborhood

2.1 Neighborhood Data of Delhi

Neighborhood data means the Sub area in Major areas of Delhi, for ex- Central Delhi is one of the 9 Major areas in Deli under which Chandni Chowk and Daryaganj are two sub areas. The data is scrapped from https://en.wikipedia.org/wiki/List of neighbourhoods of Delhi. The data is not available in any table format so we have used external library **BeautifulSoup** we have used this library for getting the Data in HTML format and after parsing the data and doing some processing we have stored the whole Data in pandas dataframe.

These are Top 10 rows from the Table.



2.2 Coordinates of these Neighborhood

Geographical Coordinate of Neighborhoods which are basically Longitude and Latitude are gained using the GeoPy Library. This Coordinates Come in use when we wants plot these Neighborhood on map or when we to extract the Venues Using Foursquare API. We have extracted coordinates for all the Neighborhoods in Data we have.

After getting coordinates of these Neighborhood, we have two new columns to our data table which are Longitude and Latitude.

The Top 10 rows of the new table is shown below.

	Area	Neighbourhood	Latitude	Longitude
0	North West Delhi	Adarsh Nagar	28.7204	77.1726
1	North West Delhi	Ashok Vihar	28.6904	77.1761
2	North West Delhi	Begum Pur	28.7326	77.0521
3	North West Delhi	Karala	28.7373	77.0396
4	North West Delhi	Model Town	28.705	77.1895
5	North West Delhi	Narela	28.8398	77.077
6	North West Delhi	Pitam Pura	28.6959	77.1373
7	North West Delhi	Rohini Sub City	28.7336	77.104
8	North West Delhi	Shalimar Bagh	28.7142	77.1574
9	North Delhi	Civil Lines	28.6767	77.2177

2.3 Venues in These Neighborhood

Venues in the radius of around 1000m or 1km in These Neighborhood have been extracted using Foursquare API. These Venues is used to study the Neighborhood like what are most common venues we have near these neighborhoods like food centers, gym, pub or hotels. Number of venues we received in under the limit of 200.

Top 10 rows of our new Data set is shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	28.72035	77.17264	Giani's	28.717900	77.173907	Ice Cream Shop
1	Adarsh Nagar	28.72035	77.17264	Adarsh Nagar Metro Station	28.716598	77.170436	Light Rail Station
2	Adarsh Nagar	28.72035	77.17264	Vishyavidyalaya Metro Station@Entry gate #1 n	28.715596	77.170981	Train Station
3	Adarsh Nagar	28.72035	77.17264	Pahalwan Dhaba	28.714594	77.172155	Indian Restaurant
4	Adarsh Nagar	28.72035	77.17264	Love Back Spells	28.715262	77.166102	Astrologer
5	Adarsh Nagar	28.72035	77.17264	Sargam Electronics	28.728114	77.168269	Electronics Store
6	Adarsh Nagar	28.72035	77.17264	Golden Girl Imitation Jewellery	28.728386	77.168960	Jewelry Store
7	Ashok Vihar	28.69037	77.17609	Bellagio	28.696361	77.180021	Asian Restaurant
8	Ashok Vihar	28.69037	77.17609	Subway	28.696321	77.179983	Sandwich Place
9	Ashok Vihar	28.69037	77.17609	Domino's Pizza	28.693000	77.177000	Pizza Place

3.Methodology

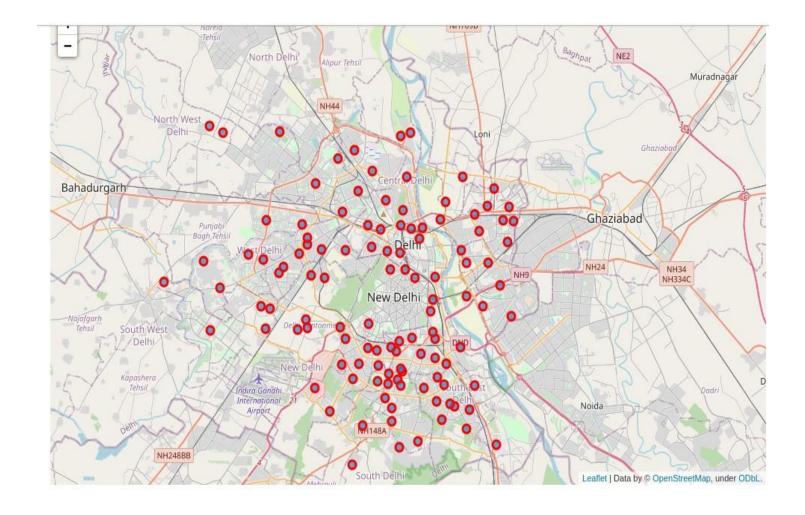
This Section will provide the details used in this project.

3.1 Data Visualization

In order to understand the data better, we used some Basic Visualization technique.

Folium map is to visualize the spread of the Neighborhoods in Delhi. With the help of Longitude and Latitude we have extracted earlier we have plotted the Neighborhoods on maps.

Below map shows the Visualization of Neighborhoods in Delhi.



3.2 Feature Extraction

Feature extraction was carried out to obtain features from the Foursquare API data which was used for Clustering and analysing the Neighbourhoods based on their Venues. In order to this, we have created the "Venue Category" and converted those categories to numerical values. Using the One-hot Encoding method we have achieved the whole process easily. Then, if a neighbourhood venue belongs to that category, it would get a value of 1 for that row in that specific category column and if a neighbourhood venue does not belong to the particular category, the value would be 0. This process was repeated for all venues in all neighbourhoods and the result was a sparse matrix containing the neighbourhood name and all unique category columns with either 1 or 0 based on. whether the neighbourhood venue belonged to that category or not. This dataframe was then grouped by the neighbourhood name and the average value was taken for all categories.

Top 10 rows for the Data is Shown below.

[164]:

	Neighborhood	Zoo	ATM	Accessories Store	Airport	Airport Lounge	American Restaurant	Antique Shop	Arcade	Art Gallery		Trail	Train Station	Turkish Restaurant	Udupi Restaurant	University
0	Adarsh Nagar	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.142857	0.0	0.0	0.0
1	Alaknanda	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	***	0.0	0.000000	0.0	0.0	0.0
2	Anand Vihar	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.000000	0.0	0.0	0.0
3	Ashok Nagar	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.000000	0.0	0.0	0.0
4	Ashok Vihar	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.000000	0.0	0.0	0.0
5	Badarpur	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.333333	0.0	0.0	0.0
6	Bali Nagar	0.0	0.0	0.0	0.0	0.0	0.045455	0.0	0.000000	0.000000		0.0	0.000000	0.0	0.0	0.0
7	Barakhamba Road	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.033333	0.066667		0.0	0.000000	0.0	0.0	0.0
8	Begum Pur	0.0	0.5	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000		0.0	0.000000	0.0	0.0	0.0
9	Chanakyapuri	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	***	0.0	0.000000	0.0	0.0	0.0

10 rows × 181 columns

We cannot see all the Category in this image but it gives the idea what we did. We can see that most of the values are 0 since there 181 unique categories and not all neighbourhoods had venues belonging to each category. After doing some Processing we have decided to work only with only top 10 common venues because it is not efficient and even for analyse purpose working with all the venues we can easily analyse the neighbourhood by their top 10 venues and that is why decided to work with that. After some Processing we have the new Data set in which we have only top 10 common venues for all the neighbourhood.

The top 10 rows of new data set are shown below.

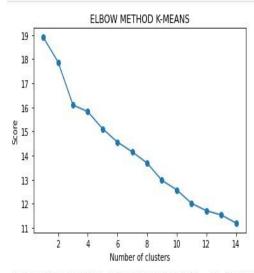
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Adarsh Nagar	Jewelry Store	Ice Cream Shop	Astrologer	Light Rail Station	Train Station	Electronics Store	Indian Restaurant	Indian Chinese Restaurant	Garden Center	Garder
Alaknanda	Restaurant	BBQ Joint	Gym	Market	Thai Restaurant	Food & Drink Shop	Women's Store	Food Court	Garden	Furniture Home Store
Anand Vihar	Italian Restaurant	Indian Restaurant	Light Rail Station	Food Court	Women's Store	Food & Drink Shop	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop
Ashok Nagar	Smoke Shop	Indian Restaurant	Mobile Phone Shop	Women's Store	Food Court	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
Ashok Vihar	Snack Place	Indian Restaurant	Sandwich Place	Pizza Place	Market	South Indian Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Departmen Store
Badarpur	Indian Restaurant	Train Station	Snack Place	Light Rail Station	Food & Drink Shop	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Join
Bali Nagar	Fast Food Restaurant	Pizza Place	Indian Restaurant	Café	Shoe Store	American Restaurant	Multiplex	Hookah Bar	Light Rail Station	Convenience Store
Barakhamba Road	Indian Restaurant	Café	Hotel	Art Gallery	Theater	Lounge	Bistro	Molecular Gastronomy Restaurant	Concert Hall	South Indian Restauran
Begum Pur	Tourist Information Center	ATM	Food & Drink Shop	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restauran
Chanakyapuri	Indian Restaurant	Restaurant	Pub	Café	Chinese Restaurant	Nightclub	Asian Restaurant	Moroccan Restaurant	Karnataka Restaurant	Park

3.3 Clustering Analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. In our case we want to group the neighborhood based on Venues. For achieving this we have Used K-means algorithm which is an Unsupervised Machine Learning Clustering algorithm.

Now for K-means we need to define one parameter manually which is number of clusters we want. To get the number of clustering we have used a Technique called Elbow method in which we used a set of different cluster number say from 1 to 15 and then we visualize the values and checked at which position error/ WSS score is decreasing slowly or Linearly. After analyzing the graph and WSS values we have found that after cluster number 4 and 5 the error values start decreasing slowly/ Linearly So we have decided to go with number of clusters 5.

The graph and values is shown below.



[18.90333761811746, 17.875527366551708, 16.098899562655987, 15.821111107287777, 15.111060750744551, 14.561845749234964, 14.153340 526659019, 13.690486129685599, 12.973819463018932, 12.579860951691309, 12.020602319889422, 11.713671637031176, 11.54026956932773 7, 11.202851935453094]

4.Result

After running the K-means algorithm we were able to divide our Neighborhood in 5 group/Clusters. Then we have separated the data based on clusters for better Analysis. Below are details of clusters.

Cluster 0

1:	Area	Neighbourhood	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 Most Common Venue	10th Most Common Venue
4	North West Delhi	Model Town	0	Pizza Place	Men's Store	Fast Food Restaurant	Convenience Store	Coffee Shop	Food Truck	Shopping Mall	Snack Place	Bank	Bakery
5	North West Delhi	Narela	0	Furniture / Home Store	Multiplex	Women's Store	Golf Course	Garden Center	Garden	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck
6	North West Delhi	Pitam Pura	0	Chinese Restaurant	Snack Place	Department Store	Mobile Phone Shop	Multiplex	Women's Store	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
7	North West Delhi	Rohini Sub City	0	Pizza Place	Juice Bar	Multiplex	Food & Drink Shop	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
11	North Delhi	Kamla Nagar	0	Breakfast Spot	Fast Food Restaurant	Pizza Place	Dessert Shop	Café	Chinese Restaurant	Coffee Shop	Donut Shop	Indian Restaurant	Dumpling Restaurant
13	North Delhi	Kotwali	0	Shopping Mall	Convenience Store	Cafeteria	Coffee Shop	Café	Women's Store	Food Truck	Garden	Furniture / Home Store	Frozen Yogurt Shop
16	North Delhi	Sarai Rohilla	0	Fast Food Restaurant	Dessert Shop	Kitchen Supply Store	Light Rail Station	Mobile Phone Shop	Women's Store	Food Court	Garden Center	Garden	Furniture / Home Store
18	North Delhi	Tis Hazari	0	Food Truck	Miscellaneous Shop	Light Rail Station	Metro Station	Women's Store	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop
19	North Delhi	Timarpur	0	Asian Restaurant	Tibetan Restaurant	Bus Station	Café	Hotel	Restaurant	Metro Station	Coffee Shop	Food Court	Garden
22	North East Delhi	Naveen Shahdara	0	Light Rail Station	Photography Studio	Vegetarian / Vegan Restaurant	Food	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant

Cluster 1

143]:		Area	Neighbourhood	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 Most Common Venue	10th Most Common Venue
	53	South Delhi	Chattarpur	1	Coffee Shop	Dessert Shop	Restaurant	Farm	Boutique	Falafel Restaurant	Event Space	Garden Center	Garden	Furniture / Home Store

Cluster 2

	Area	Neighbourhood	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 Most Common Venue	10th Most Common Venue
0	North West Delhi	Adarsh Nagar	2	Jewelry Store	Ice Cream Shop	Astrologer	Light Rail Station	Train Station	Electronics Store	Indian Restaurant	Indian Chinese Restaurant	Garden Center	Garden
1	North West Delhi	Ashok Vihar	2	Snack Place	Indian Restaurant	Sandwich Place	Pizza Place	Market	South Indian Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Department Store
8	North West Delhi	Shalimar Bagh	2	Coffee Shop	Indian Restaurant	Convenience Store	Restaurant	Hotel	Food	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
9	North Delhi	Civil Lines	2	Pizza Place	Donut Shop	Grocery Store	Light Rail Station	Trail	Chinese Restaurant	Hotel	Dessert Shop	Italian Restaurant	Indian Restaurant
10	North Delhi	Gulabi Bagh	2	ATM	Electronics Store	Pizza Place	Indian Restaurant	Train Station	Business Service	Café	Food Court	Garden	Furniture / Home Store
12	North Delhi	Kashmiri Gate	2	Indian Restaurant	Fast Food Restaurant	Market	Snack Place	Bus Station	Pizza Place	Metro Station	Boutique	Dessert Shop	Historic Site
14	North Delhi	Mori Gate; Delhi	2	Indian Restaurant	Snack Place	Fast Food Restaurant	Market	Light Rail Station	Dessert Shop	Boutique	Miscellaneous Shop	Metro Station	Historic Site
15	North Delhi	Sadar Bazaar	2	Indian Restaurant	Indian Sweet Shop	Multiplex	Convenience Store	Shoe Store	Women's Store	Food Court	Garden	Furniture / Home Store	Frozen Yogurt Shop
17	North Delhi	Shakti Nagar	2	Train Station	Indian Restaurant	Light Rail Station	Fast Food Restaurant	Women's Store	Food & Drink Shop	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop
21	North East Delhi	Dilshad Garden	2	Pizza Place	Indian Restaurant	Shop & Service	Light Rail Station	Metro Station	Food & Drink Shop	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint

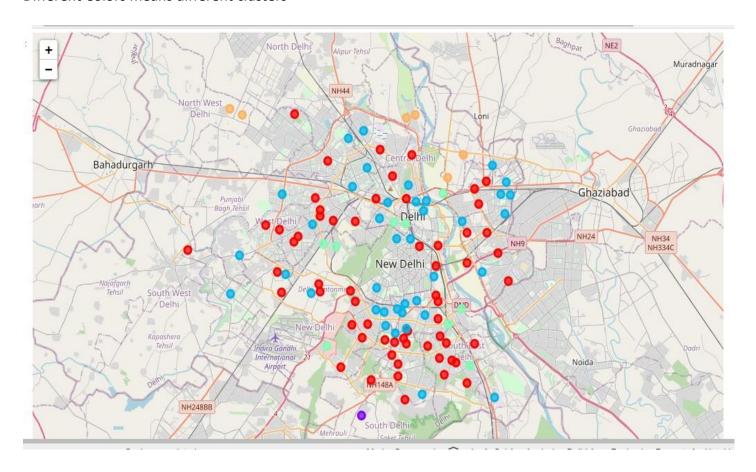
Cluster 3

	Area	Neighbourhood	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 Most Common Venue	10th Most Common Venue
25	North East Delhi	Shastri Park	3	Metro Station	IT Services	Hotel	Women's Store	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restauran
29	Central Delhi	Jhandewalan	3	Indian Restaurant	Hotel	Pizza Place	Light Rail Station	High School	Food	Women's Store	Food Court	Garden	Furniture Home Store
33	Central Delhi	Paharganj	3	Hotel	Indian Restaurant	Pizza Place	Fast Food Restaurant	Tibetan Restaurant	Breakfast Spot	Snack Place	Café	Bar	Coffee Shop
74	South Delhi	New Friends Colony	3	Hotel	Gym	Café	Food Court	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
76	South Delhi	Okhla	3	Pizza Place	Convenience Store	Hotel	Food & Drink Shop	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restauran
86	South Delhi	Sriniwaspuri	3	Hotel	Flea Market	Café	Shopping Mall	Plaza	Gym	Gym / Fitness Center	Fast Food Restaurant	Coffee Shop	Sculpture Garder
94	South West Delhi	Inderpuri	3	Hotel	Snack Place	Sandwich Place	Indian Restaurant	Multiplex	Arcade	Fast Food Restaurant	Café	Garden	Furniture Home Store
95	South West Delhi	Mahipalpur	3	Hotel	Indian Restaurant	Road	Women's Store	Food	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
99	South West Delhi	Naraina Vihar	3	Hotel	Café	Snack Place	Pizza Place	Furniture / Home Store	Indian Restaurant	Multiplex	Arcade	Fast Food Restaurant	Sandwich Place

Cluster 4

	Area	Neighbourhood	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 Most Common Venue	10th Most Common Venue
2	North West Delhi	Begum Pur	4	Tourist Information Center	ATM	Food & Drink Shop	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	Frenci Restauran
3	North West Delhi	Karala	4	Tourist Information Center	Women's Store	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truc
20	North Delhi	Wazirabad	4	ATM	Accessories Store	Women's Store	Food Court	Golf Course	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozei Yogurt Sho
23	North East Delhi	New Usmanpur	4	Tourist Information Center	ATM	Food & Drink Shop	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	Frenci Restauran
26	North East Delhi	Yamuna Vihar	4	Pizza Place	ATM	Food & Drink Shop	Gastropub	Garden Center	Garden	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	Frenci Restauran
80	South Delhi	Sangam Vihar	4	ATM	Accessories Store	Shipping Store	Clothing Store	Women's Store	Food Court	Garden Center	Garden	Furniture / Home Store	Frozei Yogurt Sho

Based on the clusters shown above, the grouped neighborhoods can be plotted on a map of Delhi. Please Note Different Colors means different clusters



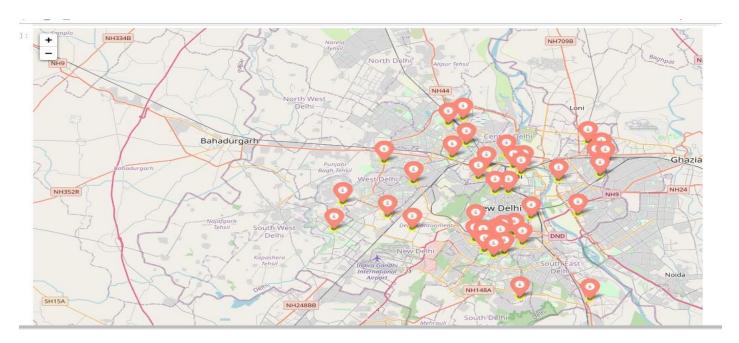
5. Discussion

After analyzing All clusters Data we can see that some clusters are Highly recommended Like Clusters 2 and 3 On the other hand Some Clusters are not good for investing like Clusters 1 and 4. Neighborhoods in clusters 1 and 4 have very less numbers of restaurants, hotels, cafe and Bars in their top 10 common venues. These clusters contain a higher numbers of other venues like ATM, Tourist Information Centre, Clothing store etc. because of this, neighborhood in these Clusters is not well for opening a new Hotel or Paying Guests. Now if we look at clusters 2 and 3 they contain a much higher number of restaurants, hotels, Gym, cafes, pubs etc. and that is why neighborhood in these city are highly recommended for opening a hotel or Paying Guests.

Now If we compare 2 and 3 we can clearly see that neighborhood in cluster 2 have more number of food centers, pubs and cafe etc and less number of hotels so for opening a new hotel or paying guests can be beneficial here since there are very less competition in these neighborhood and if we look at clusters 3 there are already more number of hotels so opening a new Hotel or Paying Guests where competition is very high is not recommended.

Now it is up to the owner what type of customers they are targeting if their target is Students then neighborhood in cluster 2 is very well suited for that since it contains all the things student wants like food centers, cafe, pubs and Gym, snack places etc. Even if target customers are Tourists then cluster 2 is highly recommended.

Recommended Neighborhood is plotted on the map.



6.Conclusion

We have successfully analyzed the neighborhoods in Delhi, India for determining which would be the best neighborhoods for opening a new Hotel or Paying Guests. Based on our Analysis of Clusters we can clearly see that opening a new Hotel and Paying Guests is Highly recommended in Cluster 2 Since there are very less numbers of hotels or paying guests and High number of food centers, gyms, pubs, snack places etc. that means opening a new Hotel or Paying Guest in these are can be very beneficial since there are less competition and high availability of food and other stuff which every human needs.