Exploring Neighborhoods of a City

- Deepesh Wadhwani

1. Introduction

1.1 Background:

- This is a capstone project – The Battle of Neighborhood.
- The Battle of Neighborhood is all about exploring restaurants or eateries of neighborhoods within the city of choice (here I have chosen Ahmedabad as city of my choice).
- **Foursquare API** is used to get neighborhood data.
- **Python** is used as a tool for analysis of neighborhood based on different parameters like average cost for two persons, cuisines offered and rating.

1.2 City of choice:

- Here I have chosen Ahmedabad as city of my choice.
- In Indian state of Gujarat there is a city called Ahmedabad, it is largest city of Gujarat and former capital of the state as well.
- It is economic and industrial hub of the state.
- It is administrative headquarters of Ahmedabad district and seat of Gujarat high court, which makes it one of the frequently visited city of the state.
- Being a designated as world heritage city by UNESCO it attracts tourists from all over the world.

1.3 Business Problem:

- For travellers looking for a restaurant at nearby place with ratings and different cuisines and cost, so that they can make decision as per their choice.
- Owner can use this data for improvisation of their business strategies.

1.4 Intended Audience:

- Travelers who are new in the city.
- Owners of restaurants can know about the rising competition and can improve their business strategies.

2. Data Description

2.1 Data Acquisition:

- The data acquired for this project are from two resources:
  1. https://ahmedabadcity.gov.in/portal/jsp/Static_pages/amc_zone_list.jsp
  2. https://www.kaggle.com/meemr5/zomato-ahmedabad-gandhinagar-restaurants-eda/data
- The Neighborhoods of Ahmedabad City is acquired from first resource i.e. ahmedabadcity.gov
- Restaurants data is acquired from second resource as csv file obtained from a Kaggle website. Or this data can also be obtained using Zomato API.

2.2 Neighborhoods:

- Data is scraped using Beautiful Soup Library of python.
- Dataset contains following columns:
  1. Zone Name
  2. Ward(neighborhood) No
  3. Ward Name
  4. Ward Office Address
- The Neighborhoods of Ahmedabad City is divided into six zones:
  1. East
  2. West
  3. New West
  4. North
  5. South
  6. Center
- Neighborhoods are defined based on above mentioned zones.

2.3 Restaurants:

- The data obtained from CSV file majorly contains:
  1. Name of Restaurant
  2. Locality of Restaurant
  3. Average cost for two people
  4. Aggregate rating
  5. Rating in text

2.4 Use of Foursquare API:

- To obtain Venues of neighborhood of the city.
- These obtained venues will be filtered by category.

2.5 Data Cleaning:

- Data Scraped from two resource were merged into one table and based on cleaned data, analysis has been done. There were some columns which were not required for analysis, so those columns were dropped.
- Several problems were faced while wrangling dataset. First, dataset containing neighborhoods column had several neighborhoods with different name as compared to Geopy library. So, these neighborhoods were identified and their names were changed according to the Geopy library. Geopy library is used to fetch latitude and longitude of locality.
- There were 196 Venue data fetched using foursquare API. To overcome problem of fewer number of restaurants fetched for big city like Ahmedabad another dataset which is a CSV file of Zomato (online food ordering platform) Restaurants was merged.
- Second problem was faced while merging two data set. The problem was both datasets having different number of columns. The datasets merged based on similarity of venues obtained, so columns were reduced to 10 as required for analysis.
- Third problem was when datasets merged there were many missing values. These missing values were removed using pandas library's function 'dropna'.
- Fourth problem was Venue categories values were not in a standard format. To over come this problem pandas dataframe 'Replace' function is used.
- Fifth problem was when two datasets were merged there were some venues that were same. First, they were identified using pandas 'nunique' function. After recognizing similar restaurants, they were dropped using 'drop_duplicate' function.

	Zone_Name	Neighborhood	Latitude	Longitude
0	North Zone	Sardarnagar	17.093462	78.165887
1	North Zone	Noble nagar	12.992753	80.187590
2	North Zone	Kubernagar	23.080409	72.634050
3	North Zone	Saijpur	22.364376	72.813849
4	North Zone	Meghaninagar	23.055642	72.615645

Figure 1 Ahmedabad Neighbourhoods

- Data from first source was cleaned and two columns namely Zone Name, Neighborhood were obtained, further neighborhoods latitude and longitude were obtained using Geopy Library as in a table below.
- **Foursquare** API is used to get near by venues based on neighborhoods of Ahmedabad.
- Total Venues obtained through API are 196 containing 7 columns as shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Sardarnagar	23.083000	72.620400	Shambhu's	23.083308	72.620098	Coffee Shop
1	Sardarnagar	23.083000	72.620400	The Gateway Hotel Ummed Ahmedabad	23.082470	72.616503	Speakeasy
2	Sardarnagar	23.083000	72.620400	Narmada	23.082377	72.616582	Indian Restaurant
3	Sardarnagar	23.083000	72.620400	GAD - Gateway All Day	23.082461	72.616506	Coffee Shop
4	Sardarnagar	23.083000	72.620400	Taj Residency Ummed	23.081322	72.616794	Hotel

Figure 2 Ahmedabad Venues

- Based on Latitude and Longitude and venue name obtained Neighborhoods and Restaurants data will be merged.
- Cluster will be formed based on category of venue
- Ten most preferred restaurants will be obtained for each neighborhood.

3. Exploratory Data Analysis (Methodology)

- Major components of a Dataset are Neighborhood, latitude, longitude based on which analysis is done.

	Zone_Name	Neighborhood	Latitude	Longitude
0	North Zone	Sardarnagar	17.093462	78.165887
1	North Zone	Noble nagar	12.992753	80.187590
2	North Zone	Kubernagar	23.080409	72.634050
3	North Zone	Saijpur	22.364376	72.813849
4	North Zone	Meghaninagar	23.055642	72.615645

Figure 3 Neighbourhoods of Ahmedabad

- To get Geographical details of Ahmedabad and its neighborhood a map is created using python Folium library as shown below.
- Using Foursquare API to explore nearby areas within radius of half kms and kept the limit of 100 venues to be fetched.
- Venues have been merged with Zomato (online food offering platform) Restaurants data which was in CSV format.
- When both data set were merged there were duplicate venues, so data cleaning has been done. Finally, there were 1208 unique venues were obtained.
- One-hot encoding was used to get unique values of categorical data (venues category).

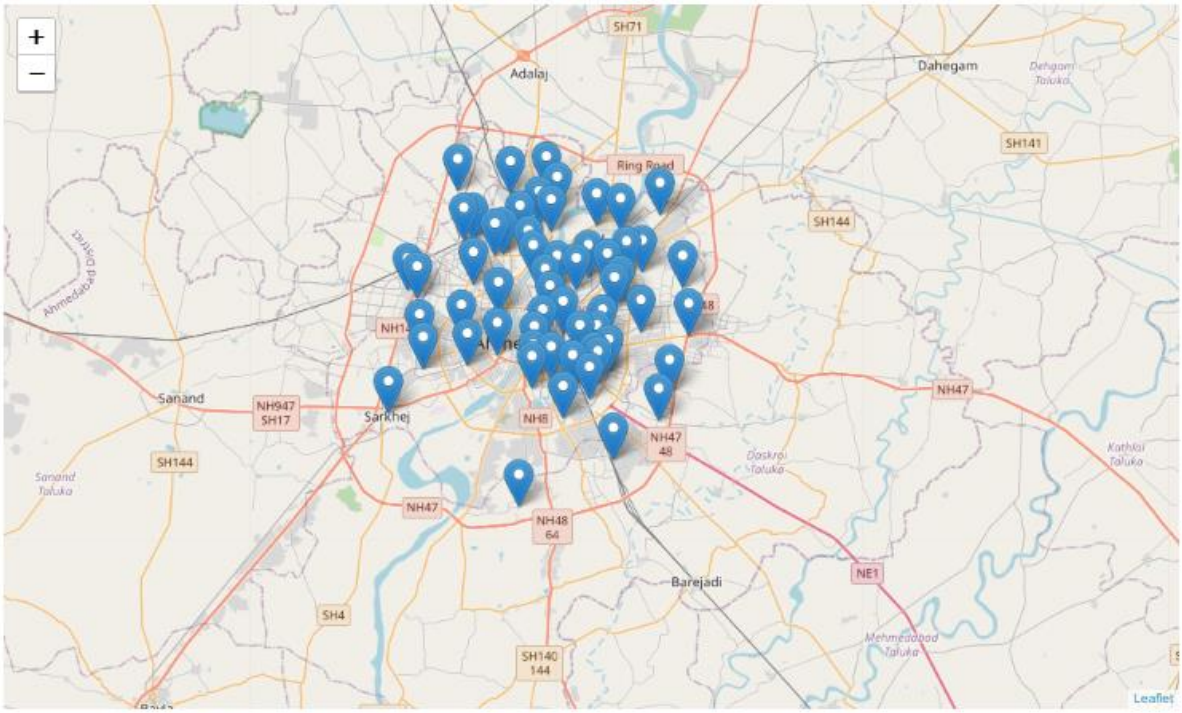

Figure 4 Geographical Details of Ahmedabad

	Neighborhood	Bakery	Beverage Shop	Bhojanalya	Café	Casual Dining	Confectionery	Dessert Parlour	Dhaba	Fine Dining	Food Court	Food Truck	Kiosk	Paan Shop	Quick Bites	Sweet Shop
0	Bodakdev	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	Ahmedabad One Mall, Vastrapur	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
2	Bodakdev	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	Prahlad Nagar	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	Vastrapur	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Figure 5 one-hot encoding unique values of venue category

- Top 10 most preferred venues were fetched for each neighborhood location.

	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
0	Sigma Legacy Building, Vastrapur	Dessert Parlour	Casual Dining	Sweet Shop	Quick Bites	Paan Shop	Kiosk	Food Truck	Food Court	Fine Dining	Dhaba
1	10 Acres Mall, Kankaria	Quick Bites	Café	Sweet Shop	Paan Shop	Kiosk	Food Truck	Food Court	Fine Dining	Dhaba	Dessert Parlour
2	4D Square Mall, Chandkheda	Quick Bites	Dessert Parlour	Casual Dining	Sweet Shop	Café	Beverage Shop	Paan Shop	Kiosk	Food Truck	Food Court
3	Aarya Grand Hotels & Resorts, Sola	Casual Dining	Café	Sweet Shop	Quick Bites	Paan Shop	Kiosk	Food Truck	Food Court	Fine Dining	Dhaba
4	Acropolis Mall, Thaltej	Dessert Parlour	Casual Dining	Café	Sweet Shop	Quick Bites	Paan Shop	Kiosk	Food Truck	Food Court	Fine Dining

Figure 6 Top 10 preferred venues for each neighbourhood

- As observed in above fig5. I have many common venues in neighborhoods.
- To calculate similarity between venues I have used unsupervised K-Means Clustering algorithm.
- The Elbow methodology is the one which determines optimal value of k clusters.

- For Elbow methodology I have used python yellowbrick library.

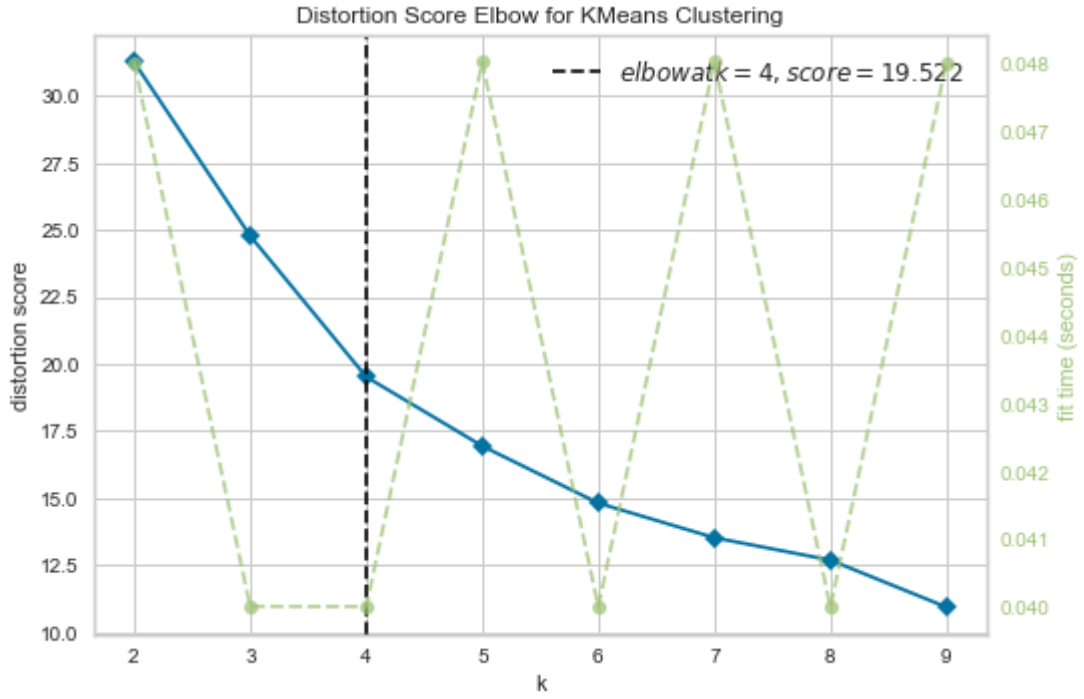

Figure 5 Elbow method for optimal k clusters

- From above fig7. It can be noted that the optimal value for k is 4.
- Cluster labels were obtained for each venue neighborhood using K-Means algorithm.

	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category	res_id	cuisines	average_cost_for_two	aggregate_rating	rating_text	Cluster Labels
0	Bodakdev	Ph Se Food	23.038755	72.510887	Casual Dining	18757684.0	North Indian, Chinese, Fast Food	900.0	4.5	Excellent	0
1	Ahmedabad One Mall, Vastrapur	Millhouse	23.039770	72.531508	Casual Dining	18658177.0	Continental, Italian	1200.0	4.5	Excellent	0
2	Bodakdev	The Red Bistro	23.039852	72.509181	Casual Dining	18663856.0	Mexican, Italian, North Indian	1400.0	4.6	Excellent	0
3	Prahlad Nagar	Kadak Bhagat	23.012411	72.514392	Casual Dining	18819827.0	North Indian, Chinese, Fast Food	1000.0	4.4	Very Good	0
4	Vastrapur	Urban Khichdi	23.028315	72.542458	Casual Dining	18718880.0	North Indian, Continental	600.0	4.7	Excellent	0

Figure 6 Cluster labels for each venue neighbourhood

4. Results

- Four clusters obtained and were plotted using venue's latitude and longitude on map of Ahmedabad using python folium library.
- As shown below fig9. Clusters were spread all over Ahmedabad.

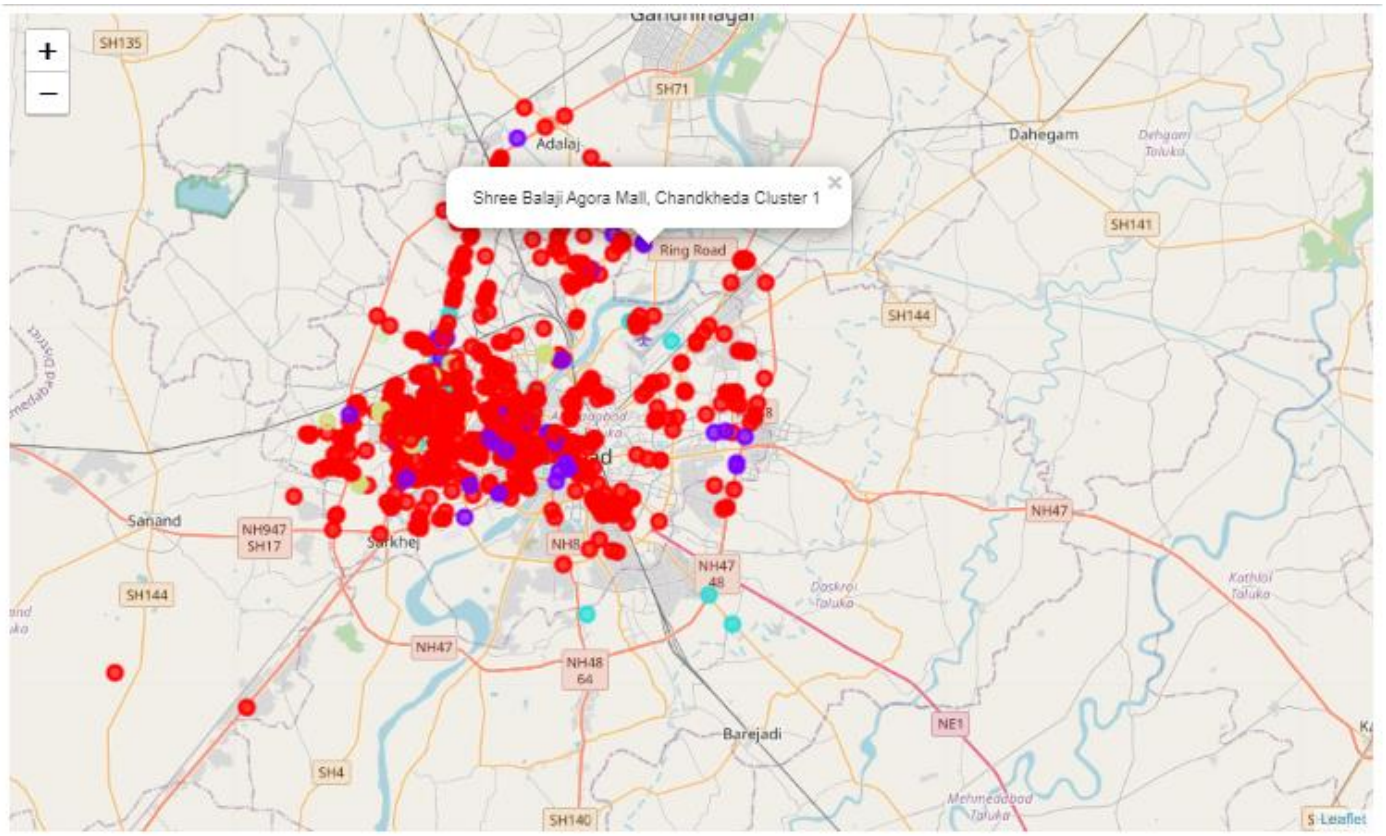

Figure 7 geographic details of venues clusters obtained in each neighbourhood of Ahmedabad.

- Details obtained for each cluster can be found in below table.

Table 1 Features of clusters

Sr. No	Cluster Name	Total Venues in a Cluster	Aggregate of Avg cost (in INR) of two people	Average Rating of Venues in a Cluster	Average Rating Text
1	Clust1	1049	445.38	3.19	Average
2	Clust2	66	957.57	3.82	Good
3	Clust3	25	1578.0	3.78	Good
4	Clust4	68	414.70	2.42	Poor

- Features like Aggregate of Avg cost (in INR) of two people, Average Rating of Venues in a Cluster, Average Rating Text, Top 5 common cuisines were plotted on graphs for each cluster. Graphs of each cluster features mentioned above can be found here.

- **Cluster 1 graphs:**

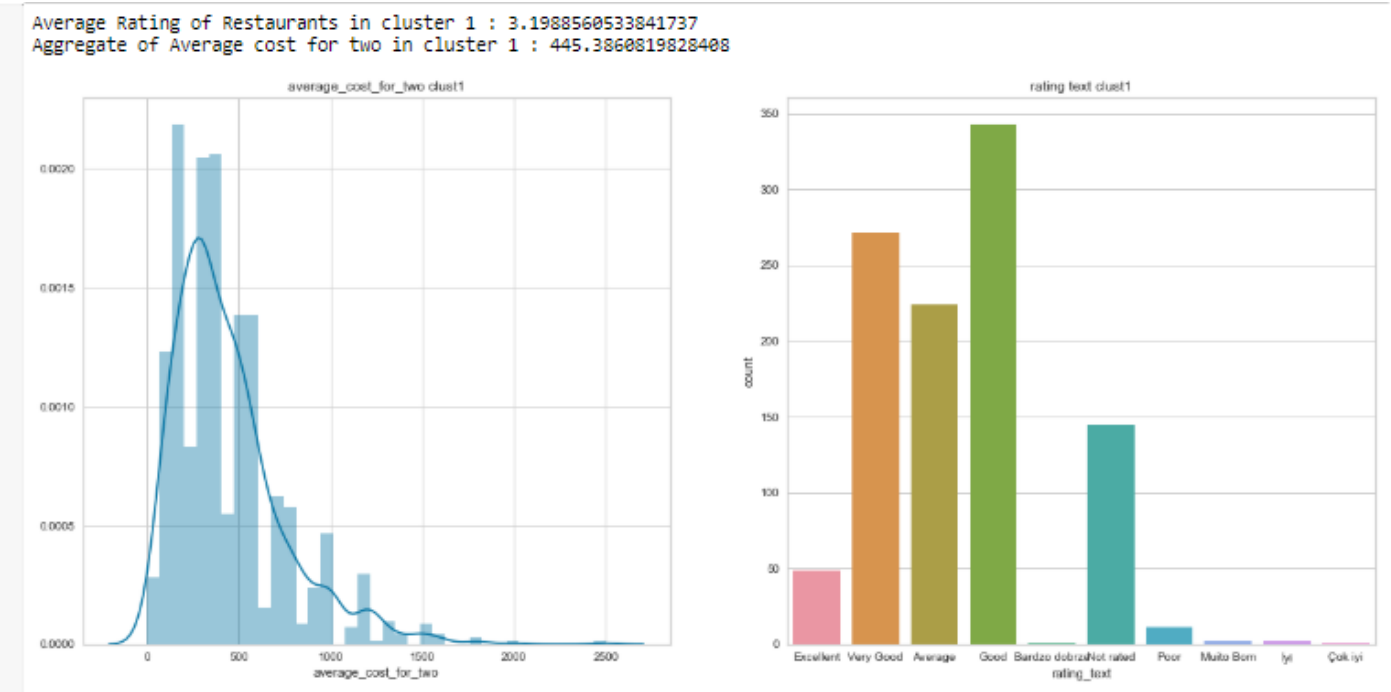

Figure 10 Clust1 left fig: Average cost of two, right fig: Rating text

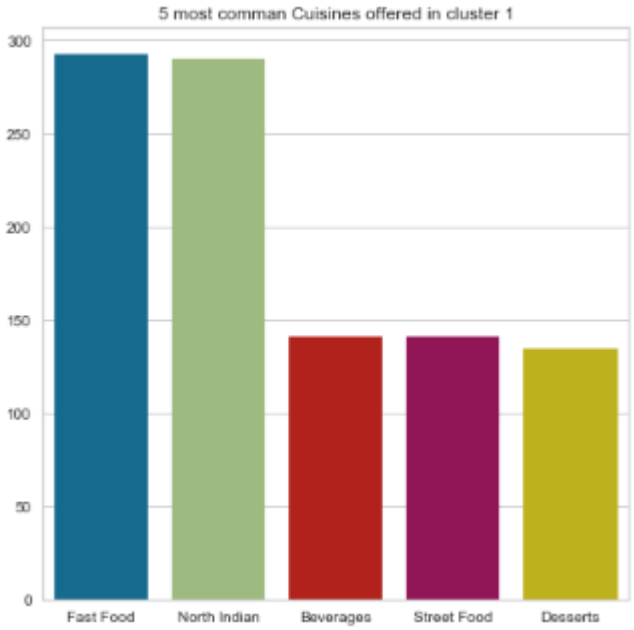

Figure 11 clust1 top 5 common cuisines

- **Cluster 2 graphs:**

```
Average Rating of Restaurants in cluster 2 : 3.8212121212121217
Aggregate of Average cost for two in cluster 2 : 957.5757575757576
```

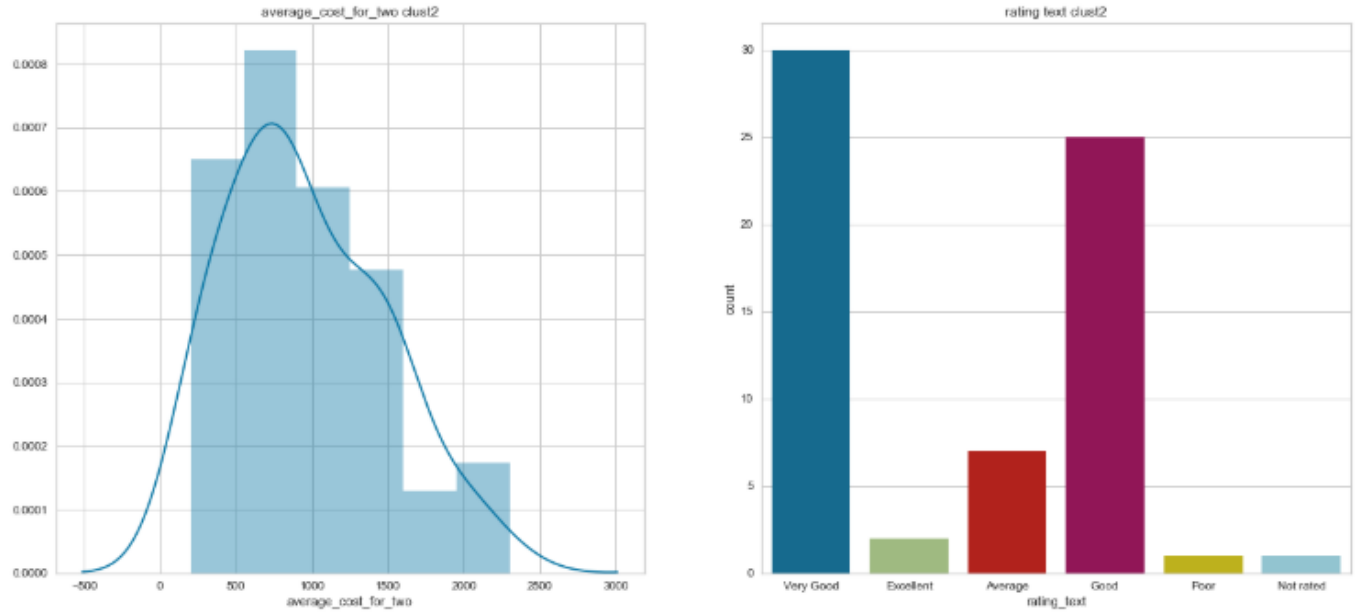

Figure 9 Clust2 left fig: Average cost of two, right fig: Rating text

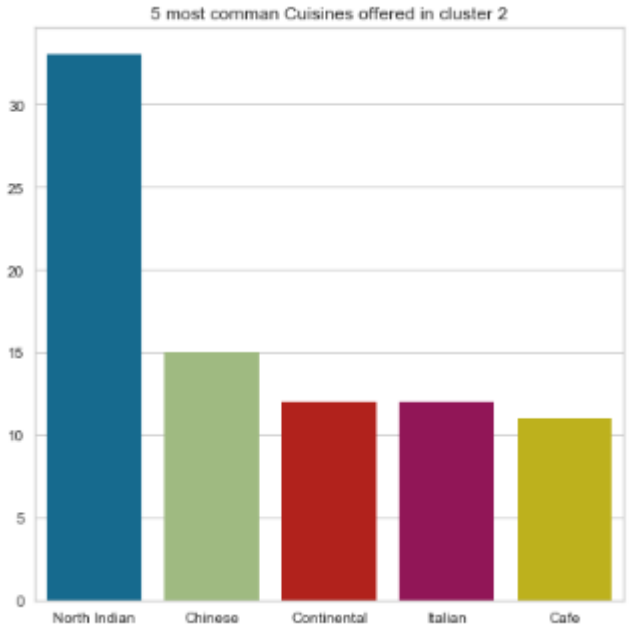

Figure 9 clust2 top 5 common cuisines

- **Cluster 3 graphs:**

```
Average Rating of Restaurants in cluster 3 : 3.7800000000000007
Aggregate of Average cost for two in cluster 3 : 1578.0
```

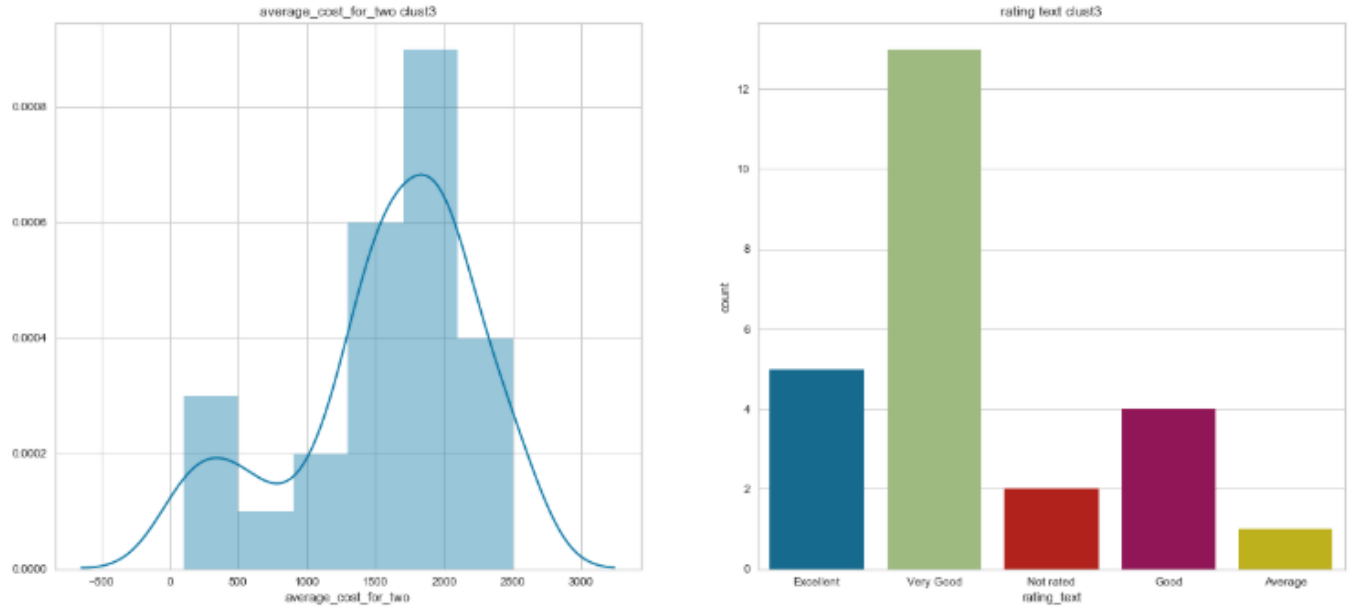

Figure 14 Clust3 left fig: Average cost of two, right fig: Rating text

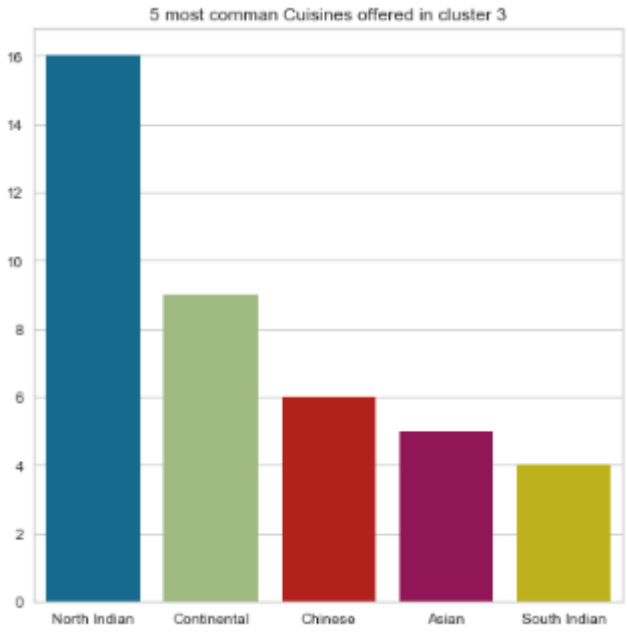

Figure 10 clust3 top 5 common cuisines

- **Cluster 4 graphs:**

```
Average Rating of Restaurants in cluster 4 : 2.427941176470587
Aggregate of Average cost for two in cluster 4 : 414.70588235294116
```

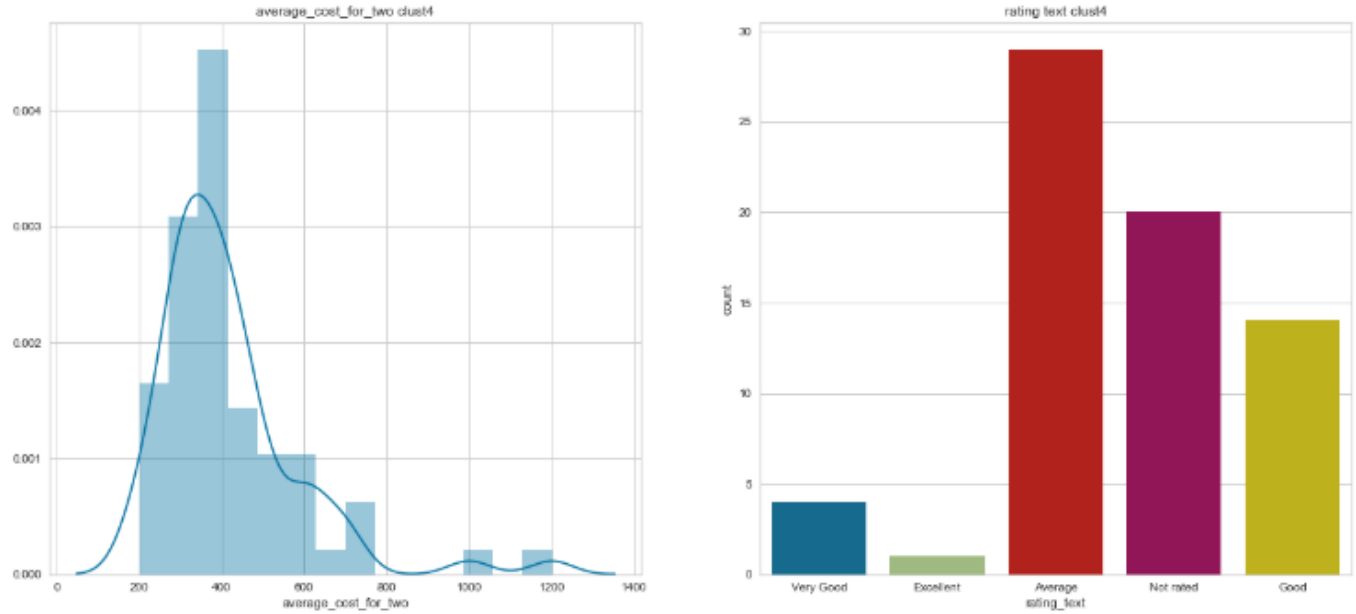

Figure 12 Clust4 left fig: Average cost of two, right fig: Rating text

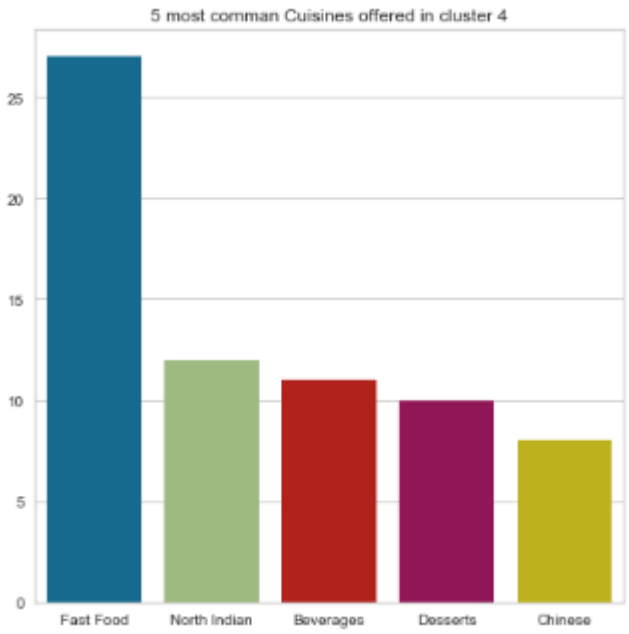

Figure 12 clust4 top 5 common cuisines

5. Discussion

- From table 1 above one can observe that each cluster has different number of venues. Cluster 1 has maximum venues being spread all over Ahmedabad. Cluster 2 is second most populous followed by cluster 4. Cluster 3 has least number of venues.
- From graphs top 5 most common cuisines offered one can infer that North Indian and Fast Food cuisines are most common foods available in each cluster. In fig11 of cluster 1 one can observe that these cuisines are offered by 250+ venues.
- Cluster 1 has average rating of 3.19 in text rating it is ‘Average’ rating and cost for two people is INR 445 which is quite low compared to other clusters.
- Cluster 2 has average rating of 3.8 in text rating it is ‘Good’ rating and cost for two people is INR 957 which is medium range pricing compared to other clusters.
- Cluster 3 has average rating of 3.78 in text rating it is ‘Good’ rating and cost for two people is INR 1578 which is high range pricing compared to other clusters.
- Cluster 4 has average rating of 2.4 in text rating it is ‘Poor’ rating and cost for two people is INR 414 which is low range pricing compared to other clusters.
- Cluster 1 and 4 have average and poor rating respectively can be chosen for business purpose as food quality offered in these clusters have chance of being improved, business offering good quality food in both clusters will face medium to low competition.

6. Conclusion

- As a result, all the clusters are offering wide variety of cuisines but for big city like Ahmedabad very few venues (cluster 2 & 3) are offering good quality food.
- The cuisines offered in clusters 2 & 3 cost medium to high, which may not be affordable for every person. So, there are fair chances for improvement of food quality in a low-medium price range venue.
- Thus, travelers looking for variety of foods offered across Ahmedabad can use this information also business owners can use this information to improvise their business strategies or person looking for opening a new venue can use this information and can know what competition he/she will face with its nearby venues.