

Analysis On Best Neighborhood In Ahmedabad City (India) For Opening A Specialty Pizza Outlet

By Rushikesh R. Mokashi
November 2020

Table of Contents

1. Introduction
2. Data Requirement, Acquisition and Cleaning
3. Methodology
4. Results
5. Discussion
6. Conclusion
7. References

1. Introduction

1.1 About The City^[1]

Ahmedabad (also spelled **Ahmadabad**) was founded by King Karandev 1 in the 11th century and originally called "Karnavati". The Hindu kingdom of Karnavati ruled for more than 400 years until 1411 when Sultan Ahmed Shah conquered Gujarat and the city was renamed Ahmedabad after him.

Ahmedabad is the sixth largest city (pop. 6.5 million) in India, and it's India's first UNESCO World Heritage City. It is the commercial hub of the state of Gujarat. The city was built in open and spacious plains to the East of Sabarmati. Ahmedabad was a cradle of the non-violent movement for India's independence, being host to the Sabarmati ashram of Mahatma Gandhi.

Ahmedabad is a gastronomic paradise. You will find at least one restaurant in every nook and corner, anywhere in Ahmedabad. Gujarati people are fond of eating and the food generally tends to be on the sweeter side. Ahmedabad is famous for its ice-creams because of abundance of dairy products in the state and Gujaratis' penchant for sweets as well.

If you have not stayed in India for long, select a food place which is famous and well organized (hotels, big restaurants). Otherwise, it's fun to eat at one of the road side stalls, especially at the Khau Galli in Law Garden area and Khan Pan Bazaar in Manek Chawk. Various stalls start operating in both these places in the evening and run late in the night. You will enjoy the food like Paani - Puri (Gol Gappa), Indian Chat, Indian Sandwiches, North Indian, Indian-Chinese

and some continental cuisines, amazingly at road-side stalls and for a very reasonable price of less than a US\$1 per item.

1.2 The Problem At Hand

As an investor who wants to invest in a restaurant business and open a specialty pizza restaurant, that serves a wide range of pizza choices in the city. The investor would like to explore the neighborhoods within the city and find out places that would be most profitable with less competition.

1.3 Interested Stakeholders

Anyone who is interested in exploring the neighborhoods of Ahmedabad city with a view of starting a pizza outlet in the city.

2. Data Requirement, Acquisition and Cleaning

2.1 Data Required

The problem at hand requires us to explore different neighborhoods around the city and understand various pizza places around each of these neighborhoods. In order to do this we would need the following data.

- i. A list of all the neighborhoods
- ii. Geo-spatial data of these neighborhoods
- iii. pizza places around each of the neighborhoods

2.2 Data Sources

No ready-to-use dataset was available for the purpose of this analysis. Data has been acquired from variety of sources. The postal codes for listing different neighborhoods in the city and the geo-coordinate data was available from www.geonames.org.^[2] The pizza places data was gathered using Foursquare APIs. Some of the missing geo-spatial data was collected using Nominatim Library and rest had to be manually entered using google maps.

2.3 Data Cleaning

After downloading the data file from www.geonames.org, only data pertaining to Ahmedabad city was selected and copied to another file. Since the geo-spatial data available in the file was for the postal code and not for the individual neighborhood, only the neighborhoods were selected from the file.

	Country Code	Postal Code	Vicinity	State	City	County	Latitude	Longitude
0	IN	380001	Lal Darwaja	Gujarat	Ahmedabad	Ahmadabad City	23.0256	72.5769
1	IN	380001	Gheekanta Road	Gujarat	Ahmedabad	Ahmadabad City	23.0256	72.5769
2	IN	380001	Ahmedabad G.P.O.	Gujarat	Ahmedabad	Ahmadabad City	23.0256	72.5769
3	IN	380001	Raipur (Ahmedabad)	Gujarat	Ahmedabad	Ahmadabad City	23.0256	72.5769
4	IN	380001	Dariapur (Ahmedabad)	Gujarat	Ahmedabad	Ahmadabad City	23.0256	72.5769

Once all the neighborhood data was available, the Geo-spatial data was collected by iterating through the neighborhoods and getting location data from geocoder library Nominatim. This data along with the neighborhoods was saved in a .csv file.

	Vicinity	V Lat	V Long
0	Lal Darwaja	23.0242	72.5786
1	Gheekanta Road	23.0325	72.5977
2	Ahmedabad G.P.O.		
3	Raipur (Ahmedabad)	23.0205	72.59
4	Dariapur (Ahmedabad)	23.0374	72.593

Once this .csv file was obtained, the missing location data for a select few neighborhoods in the Ahmedabad city county were manually input in the file using google maps.

For the purpose of this study only 50 neighborhoods were selected. These are as follows:

1. Lal Darwaja	26. Jodhpur Char Rasta
2. Manek Chowk	27. I I M
3. Shahibag	28. Manekbag
4. Sabarmati	29. Polytechnic (Ahmedabad)
5. Kabir Chowk	30. Azad Society
6. Motera	31. Meghaningar
7. Ambawadi (Ahmedabad)	32. Gita Mandir Road
8. Ellisbridge	33. Civil Hospital
9. Paldi (Ahmedabad)	34. Gandhi Ashram (Ahmedabad)
10. Sarkhej Road	35. Jivraj Park
11. Anandnagar (Ahmedabad)	36. Memnagar
12. Shardanagar	37. Thaltej Road
13. Narayannagar	38. Bodakdev
14. Jawahar Chowk	39. Juhapura
15. Maninagar	40. Bopal
16. Ashram Road	41. Ghuma
17. Darpan Society	42. Ambli
18. Gujarat University	43. Shela
19. Navrangpura	44. Thaltej
20. Stadium Marg	45. Shilaj
21. Vadaj	46. Sola
22. Naranpura Vistar	47. Bhadaj
23. Shastrinagar (Ahmedabad)	48. Ognaj
24. Nava Vadaj	49. Ghatlodia
25. Navjivan	50. Sola H B C

Based on the location data collected for the above neighborhoods, pizza places in each of the above neighborhoods were collected using the Foursquare™ APIs (www.foursquare.com). Using the available API, I explored different venues within a radius of 1100 meters / 1.1 KMs.

Following is a snapshot of the data gathered via the API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
0	Lal Darwaja	23.02419	72.578556	56fcee27498eef4f589db68f	Pizza Hut	23.032067	72.570223	Pizza Place	1223
1	Lal Darwaja	23.02419	72.578556	510947c5e4b0c06b6e9fe820	Chef's Pizza	23.026209	72.580739	Pizza Place	317
2	Lal Darwaja	23.02419	72.578556	5300f3f9498efe1cddbdc2e0	Pizza Hut Delivery Express	23.031657	72.570516	Pizza Place	1170
3	Lal Darwaja	23.02419	72.578556	51d31558498eab154179e566	Manek Pizza & Sandwich	23.023835	72.588502	Fast Food Restaurant	1019
4	Lal Darwaja	23.02419	72.578556	4fd73e73e4b0ed2ec945faa6	Pinky Pizza	23.027697	72.589634	Fast Food Restaurant	1200

After the above data was collected, I checked to see if same venues are repeated because of the intersection of radii of the neighborhoods. To eliminate the duplicate venue entries, I sorted the above data in ascending order based on 'venue id' and 'distance' from the neighborhood coordinates.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
153	Jodhpur Char Rasta	23.025341	72.527761	4bace359f964a52011163be3	Pizza hut	23.026240	72.529440	Italian Restaurant	199
183	Azad Society	23.029157	72.532752	4bace359f964a52011163be3	Pizza hut	23.026240	72.529440	Italian Restaurant	469
158	I I M	23.031270	72.537133	4bace359f964a52011163be3	Pizza hut	23.026240	72.529440	Italian Restaurant	966
52	Anandnagar (Ahmedabad)	23.016075	72.509295	4bc829dd14d79521b49c68e9	Domino's Pizza	23.024285	72.507111	Pizza Place	940
41	Anandnagar (Ahmedabad)	23.016075	72.509295	4bd6e2decfa7b7132e3428da	Jasuben pizza	23.015231	72.507146	Pizza Place	239

Neighborhood		
Venue ID	Venue	
4f5ccb33e4b05cfda5d7cbad	Domino's Pizza	6
561d52eb498ed81f498c4fbe	Domino's Pizza	6
541d99c5498e3fbc999faec7	bon pizza	6
5162ba38e4b03d1defd89f4f	Jasuben old pizza	5
56fcee25498ee94f64c5b967	Pizza Hut	5

I removed the duplicates and only kept the first record of the duplicate venues. The one closest to the respective neighborhood.

Venue ID	Venue	Neighborhood
4bace359f964a52011163be3	Pizza hut	1
4bc829dd14d79521b49c68e9	Domino's Pizza	1
5319778a498e2932f30def4b	Neopolitan Pizza	1
530e081011d20a35828c5200	Pizza Hut	1
5300f3f9498efe1cdbc2e0	Pizza Hut Delivery Express	1

3. Methodology

3.1 Exploratory Data Analysis

In this section I dived into more details of the data which I had collected and cleaned thus far. I tried to figure out if there were multiple branches of any of the pizza outlets.

	Venue	Count
0	Domino's Pizza	18
1	Pizza Hut	9
2	U S Pizza	6
3	Pizza Zone	3
4	La Pino's Pizza	2
5	Pizza World	2
6	Enjoy Pizza	2
7	Sam's Pizza	2
8	Pizza Americano	2
9	Pomos Pizza	1

From the above we can see that there are quite a few outlets having multiple branches in the city. After finding the above information, I explored further to see if all the branches are classified into a similar venue category.

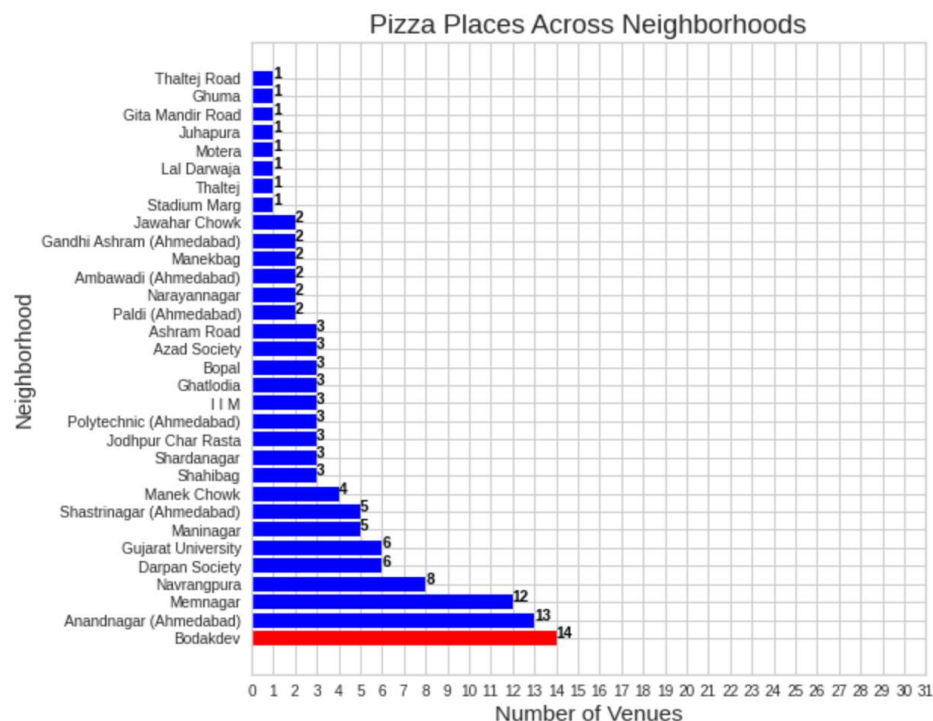
	Venue	Category List
0	Bon pizza	[Food Truck, Pizza Place]
1	Enjoy Pizza	[Italian Restaurant, Pizza Place]
2	Pizza Hut	[Pizza Place, Italian Restaurant]
3	Pizza hut	[Pizza Place, Italian Restaurant]
4	U S Pizza	[Italian Restaurant, Pizza Place]
5	U.S Pizza	[Italian Restaurant, Pizza Place]
6	U.S.Pizza	[Italian Restaurant, Pizza Place]
7	bon pizza	[Food Truck, Pizza Place]

As you can see, different branches of the same outlet are classified into different venue categories. I then, re-classified the above venues into a uniform Category viz. 'Pizza Place'. I then tried to check the unique categories into which these venues were classified which served pizzas. I found that these venues were classified into the following categories.

'Pizza Place', 'Arcade', 'Asian Restaurant', 'Sandwich Place', 'Fast Food Restaurant', 'Italian Restaurant', 'Food Truck'

Since all of the above venues served pizzas, I re-classified each of these venues as 'Pizza Place'.

Pizza Places Around The Neighborhoods



I got the above information regarding the neighborhood and its respective counts in a tabular format:

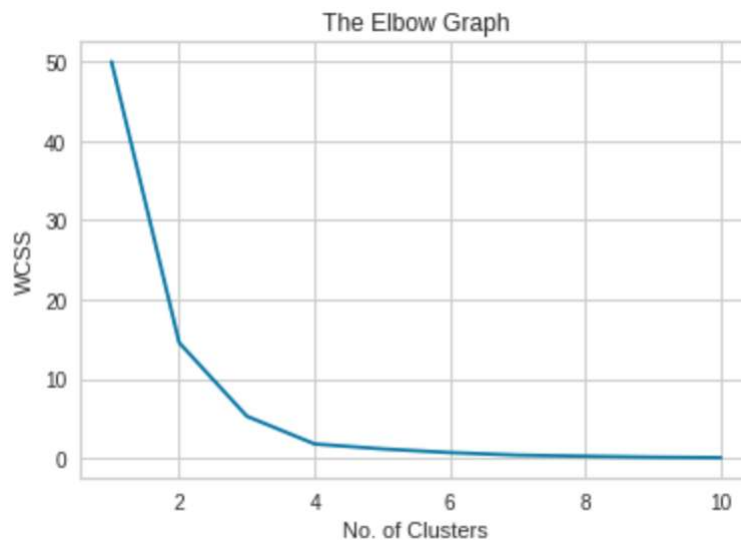
	Neighborhood	Venue
0	Bodakdev	14
1	Anandnagar (Ahmedabad)	13
2	Memnagar	12
3	Navrangpura	8
4	Darpan Society	6
5	Gujarat University	6
6	Maninagar	5

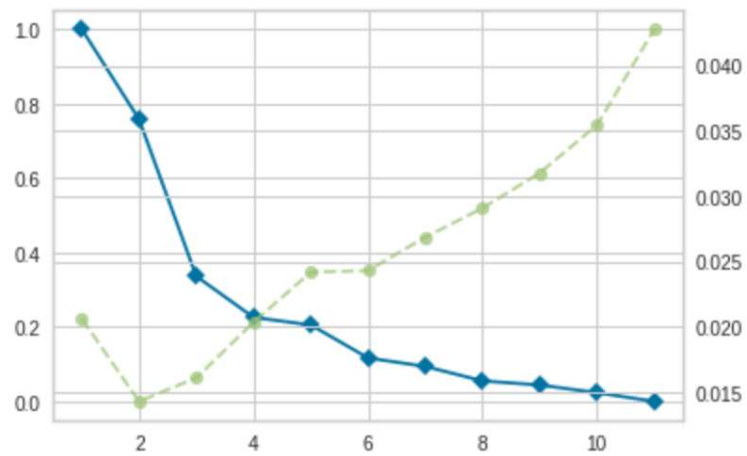
3.2 Unsupervised Machine Learning (K-Means Clustering)

As we are interested in exploring the neighborhoods that would be most conducive for opening the pizza outlet, we would try to find out clusters among these neighborhoods that might give us some insights.

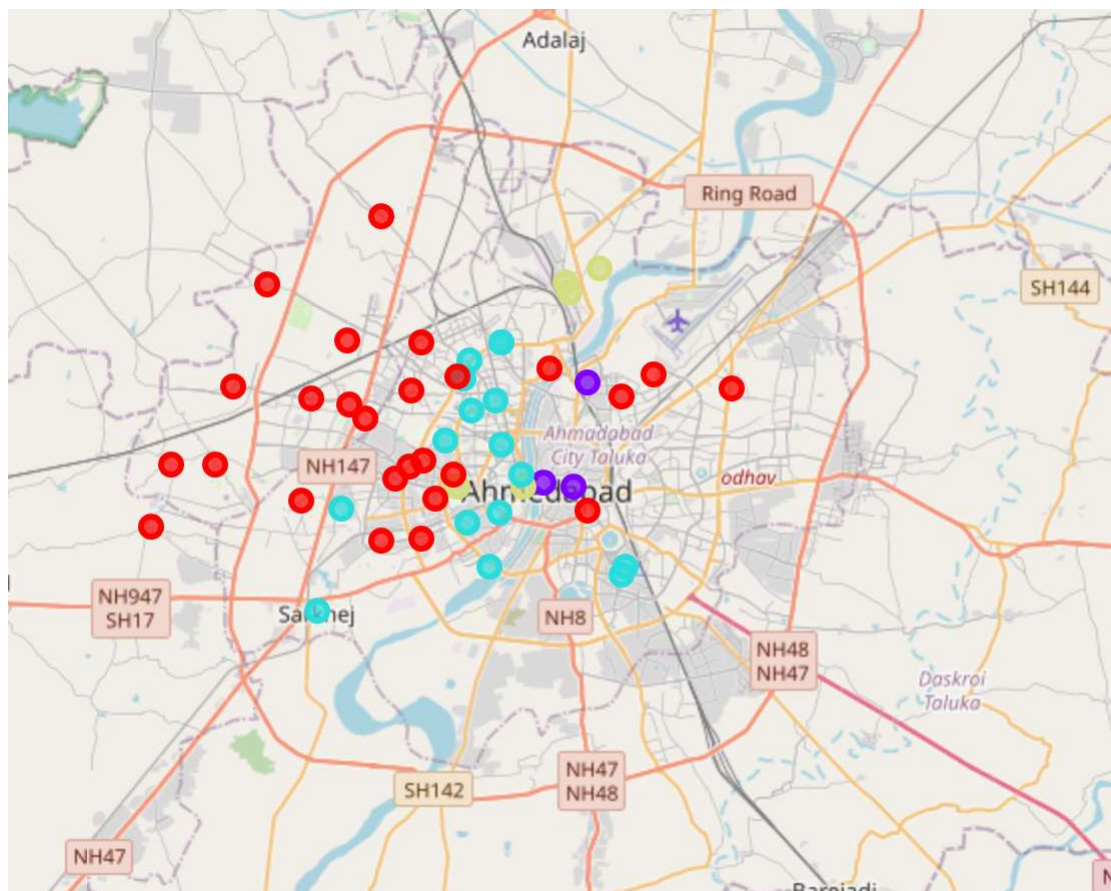
To uncover the clusters, I used the K-Means method. To perform that, we need data of the neighborhoods and their respective counts of the pizza outlets. We also need to combine the neighborhoods that do not have any outlets in its surrounding.

After doing the above, the next step was to find the optimal value of K. To do this I iterated the K-Means algorithm starting from K=1 to K=10. The following is the elbow graph created based on the selected K values.





Referring to both the graphs above, I selected the value of $K=4$ and performed the K-Means clustering. After the execution of the algorithm, each neighborhood was then classified into different clusters. Below is the map of the city which shows the various neighborhood classified into their respective clusters.



Following are the various clusters:

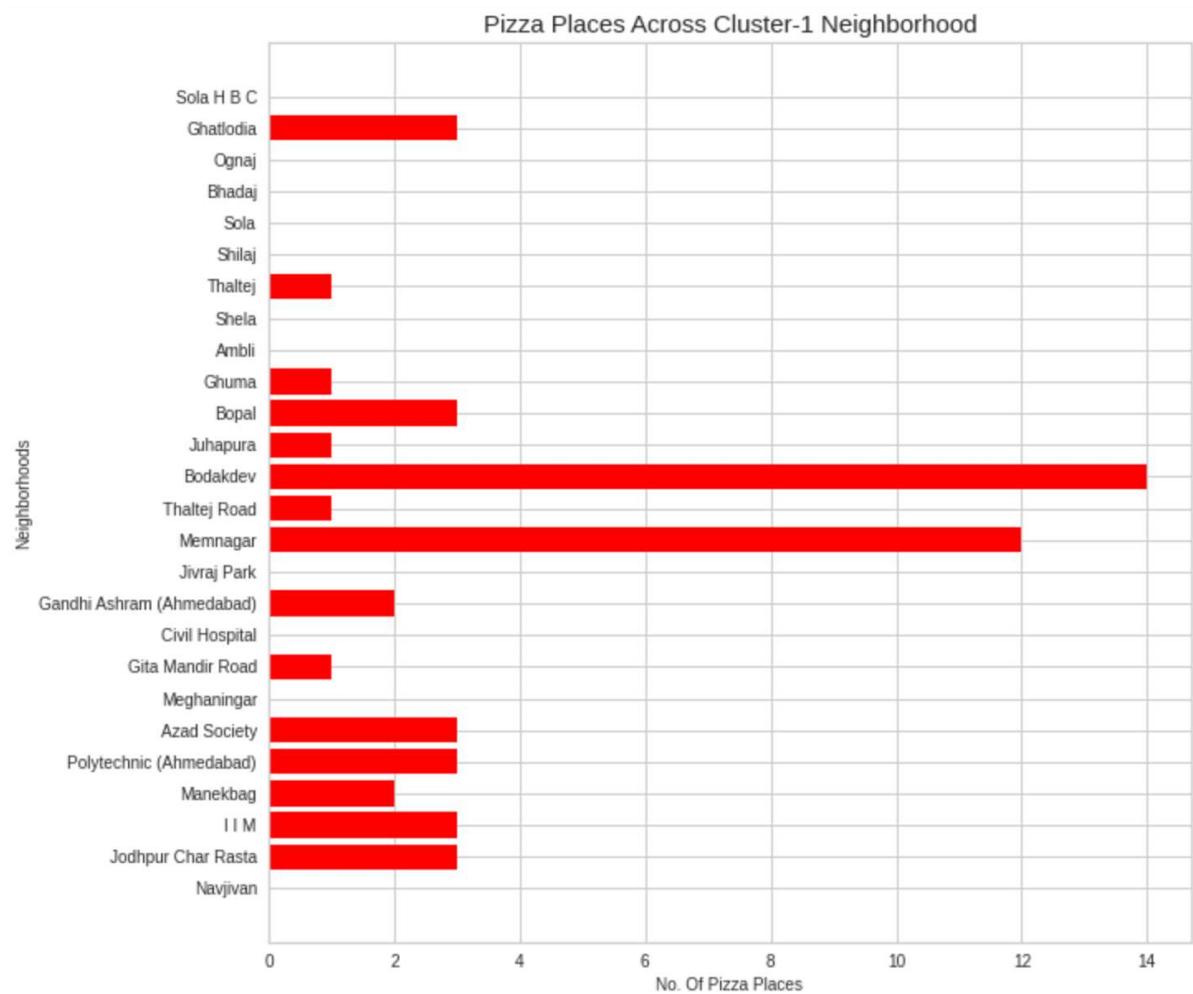


4. Results

In this section let's explore each of the clusters that were identified in the previous section.

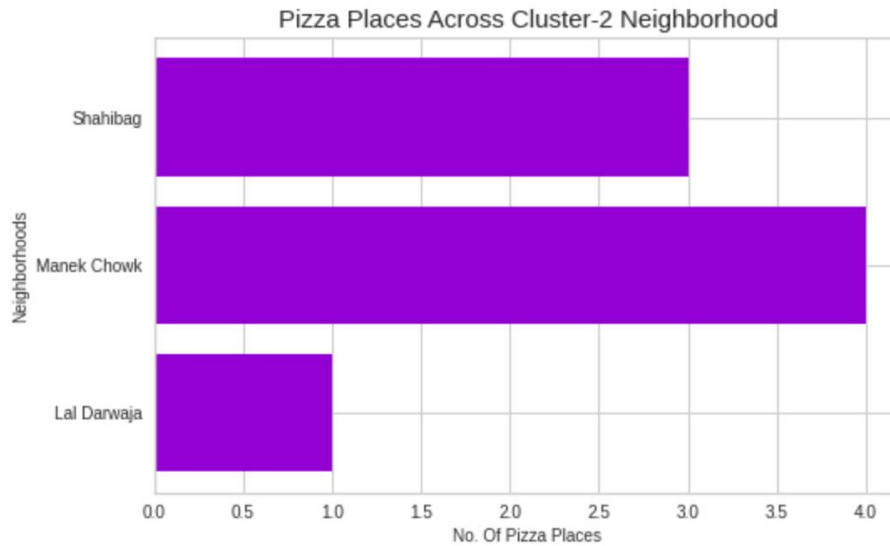
Cluster-1

Total Neighborhoods in Cluster-1: 26



Cluster-2

Total Neighborhoods in Cluster-2: 3



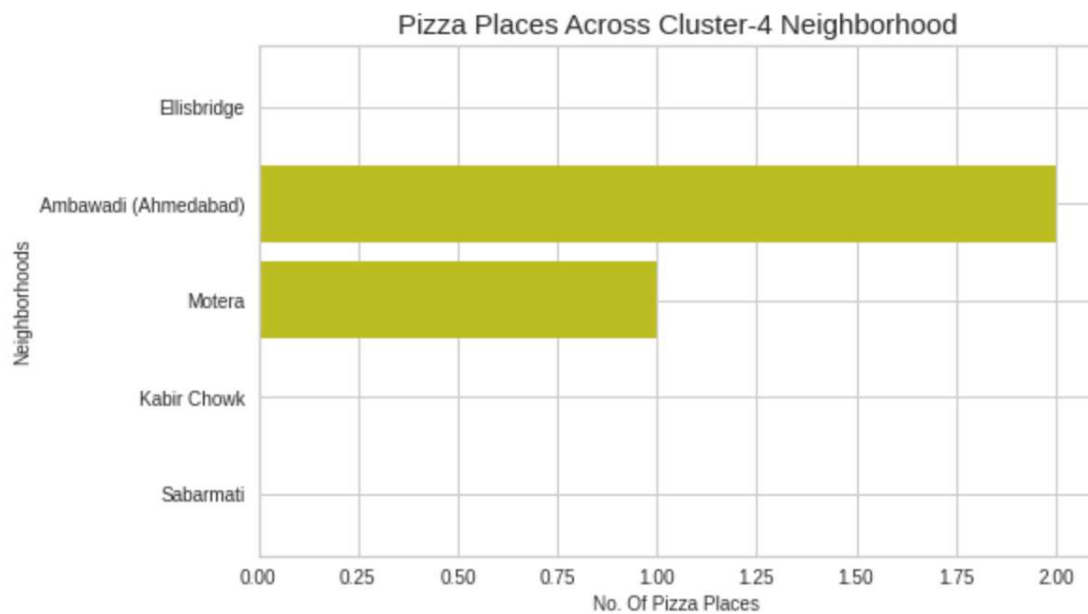
Cluster-3

Total Neighborhoods in Cluster-3: 16

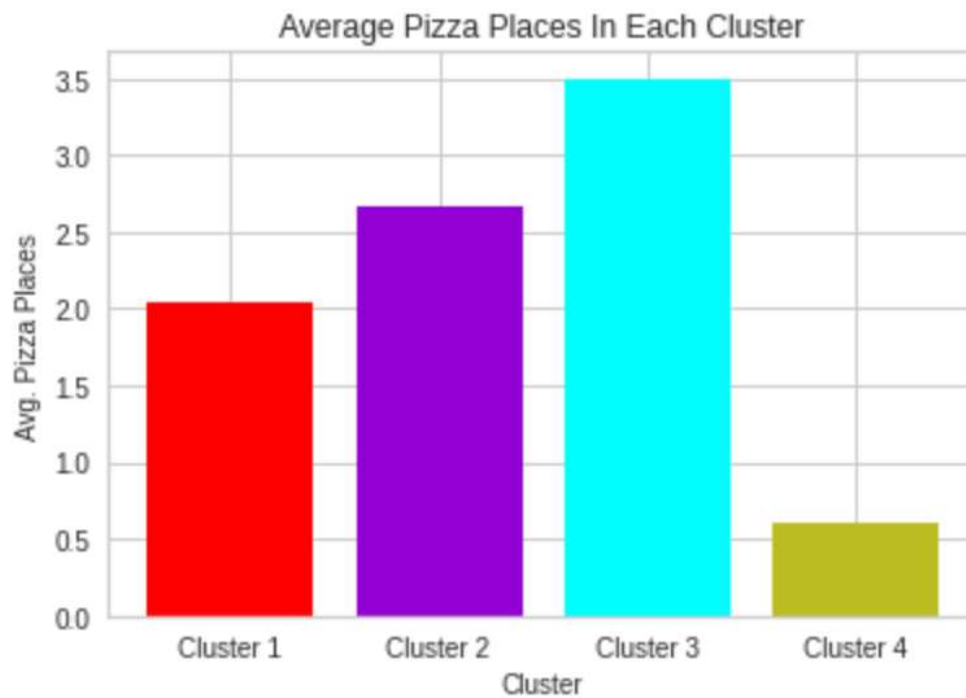


Cluster-4

Total Neighborhoods in Cluster-4: 5

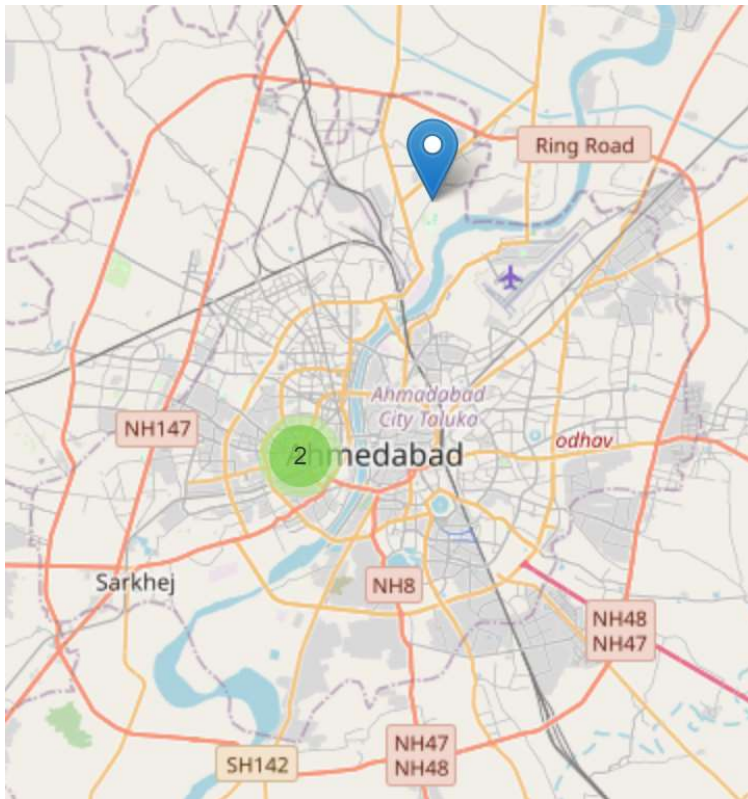


Based on the above charts, I plotted the average pizza places in each clusters.



As you can see from the chart above, the average number of pizza places is lower in cluster 4.

Location of Pizza Places In Cluster 4



In the previous graphs we saw how each clusters fared against each other. With cluster 4 having the lowest average number of pizza places (≈ 0.6) and cluster 3 having the highest average number of pizza places (≈ 3.5). Based on these information, we can propose cluster 4 to be the most opportune cluster. However, just the average number of pizza places cannot be the only criteria to narrow down on a neighborhood that is more profitable.

In order to look at other parameters, I explored the neighborhoods with high number of pizza places to see the top venue categories in these neighborhoods. This gives us an idea of the preferences and popularity of food choices.

Top 10 neighborhoods having highest number of pizza places:

	V Lat	V Long	Neighborhood	Cluster Labels	Venue
37	23.044227	72.517458	Bodakdev	0	14
10	23.016075	72.509295	Anandnagar (Ahmedabad)	2	13
35	23.053084	72.532931	Memnagar	0	12
18	23.036000	72.564343	Navrangpura	2	8
17	23.037307	72.544748	Gujarat University	2	6
16	23.046712	72.553567	Darpan Society	2	6
14	22.997713	72.606717	Maninagar	2	5
22	23.062423	72.553180	Shastrinagar (Ahmedabad)	2	5
1	23.023180	72.588952	Manek Chowk	1	4
29	23.029157	72.532752	Azad Society	0	3

Top 10 recommended venue categories in these neighborhoods:

Got the below information using Foursquare API that provides recommended places in the given neighborhood.

	Venue Category	Venue count
0	Indian Restaurant	34
1	Café	33
2	Fast Food Restaurant	20
3	Pizza Place	13
4	Coffee Shop	12
5	Hotel	11
6	Sandwich Place	11
7	Restaurant	11
8	Vegetarian / Vegan Restaurant	8
9	Ice Cream Shop	8

I then compared the above information with the top 10 venues in the neighborhoods of each cluster. The objective behind this exercise was to see if we can derive any similarity between these neighborhoods viz. The ones having highest pizza places in comparison to the neighborhoods in other clusters. To do this I used the Jaccard Similarity Index metrics as shown below.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

And the following were the observations for each clusters.

Cluster	Jaccard Similarity Index
Cluster-1	0.82 \approx 82%
Cluster-2	0.18 \approx 18%
Cluster-3	0.82 \approx 82%
Cluster-4	0.54 \approx 54%

5. Discussion

All of the above analysis has been undertaken using the venue information available with Foursquare. There are high chances that some of the venues were not covered in our analysis.

However, based on the results explained in the previous section we can see that cluster-2 has an average of \approx 2.6 pizza places in its neighborhood. And based on the Jaccard score of 18% the neighborhoods in cluster 2 are not very similar to the neighborhoods having highest number of pizza places.

Cluster 1 and Cluster 3 have a similar Jaccard score of \approx 82% which indicates a good similarity with the neighborhoods having highest number of pizza outlets. However, we should also see that the neighborhoods that have the highest pizza outlets also belongs to these clusters viz. Bodakdev, Anandnagar, Memnagar and Navrangpura.

Cluster 4 on the other hand has a Jaccard score of \approx 54%. This combined with the average pizza outlets of 0.6 seems to be a good indicator about a potential market.

6. Conclusion

The analysis above shows the various neighborhoods across the city of Ahmedabad and how each neighborhood is similar or diverse when it comes to food choices. The analysis shows how some of the neighborhoods have too many pizza outlets while some of the neighborhoods remain untapped. This analysis could be improved by combining the information like household incomes in each neighborhood to understand the spending capacity of the population in these neighborhoods.ⁱ

This analysis should help anyone who intends to invest in opening a pizza outlet in the city.

7. References

1. <https://www.coursera.org/learn/applied-data-science-capstone/home/welcome>
2. www.foursquare.com
3. www.geonames.org

i^[1] <https://wikitravel.org/en/Ahmedabad>

^[2] This work is licensed under a Creative Commons Attribution 3.0 License. see <http://creativecommons.org/licenses/by/3.0/>