

Clustering and Segmenting Neighbourhoods of Ahmedabad and it's Price Index

Ishaan Buch

July 24, 2020

1. Introduction

Ahmedabad is the largest city and former capital of the Indian State of Gujarat. It is the administrative headquarters of the Ahmedabad district and the seat of the Gujarat High Court. Ahmedabad's population of 5,633,927 (as per 2011 population census) makes it the fifth-most populous city in India.

As you can see from the figures, Ahmedabad is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense. When we think of it by the investor, we expect from them to prefer the districts where there is a lower real estate cost and the type of business they want to install is less intense. If we think of the city residents, they may want to choose the regions where real estate values are lower, too. At the same time, they may want to choose the district according to the social places density. However, it is difficult to obtain information that will guide investors in this direction, nowadays.

When we consider all these problems, we can create a map and information chart where the real estate index is placed on Ahmedabad and each district is clustered according to the venue density. This model helps the investors to view all the similar areas in the region and help them choose an area which is suitable for their cause with an affordable price.

2. Data Acquisition and Cleaning

The Area wise rates of the city of Ahmedabad form [makaan.com](https://www.makaan.com). The data is available on the internet, so for conducting the analysis it had to be extracted. Certain tools were used to extract the data into an .csv file. The .csv file has the rates of all the areas of the city per square foot. I cleaned the data and manipulated it for ease of use.

I used [Foursquare API](#) to get the most common venues of given areas of Ahmedabad. Foursquare is a social location service that allows users to explore the world around

them. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server.

There are not too many public datas related to demographic and social parameters for the city of Ahmedabad. Due to the limited availability of data, certain factors could not be included for the segmentation of the neighbourhoods. Only proper clustering was able to be performed.

3.Exploratory Data Analysis

The dataset did not include the exact locations of the areas which could help in locating them. The first step was to find the corresponding latitude and longitude of all the areas of the city. As it is not one of the most popular cities, neither it is a city which has a lot of data regarding its areas. The first step was crucial in removing the redundant data. As the dataset included many streets or junctions which were located in the area. This would lead to the data of the same area being repeated in the dataset.

Many areas did not have digital identity, as they were not so common. So it was important to remove those areas. The table below shows the final dataset after finding the pinpoint locations of the areas. And place all the locations on the map.

	Location	Name	Pincode	State	District	Lat	Long	Buy Rates 1	Buy Rates 3	Buy Rates Average	Trends
0	Abasana	Abasana	382120	Gujarat	Ahmedabad	23.3662	72.054	NaN	NaN	NaN	NaN
1	Adroda	Adroda	382220	Gujarat	Ahmedabad	22.8579	72.4062	NaN	NaN	NaN	NaN
2	Adval	Adval	382460	Gujarat	Ahmedabad	22.3755	71.9772	NaN	NaN	NaN	NaN
3	Ahmedabad.	Ahmedabad.	380001	Gujarat	Ahmedabad	23.0198	72.5935	NaN	NaN	NaN	NaN
4	Alampur	Alampur	363610	Gujarat	Ahmedabad	22.3515	71.7118	NaN	NaN	NaN	NaN

Table 3.1. Displays all the areas with their coordinates

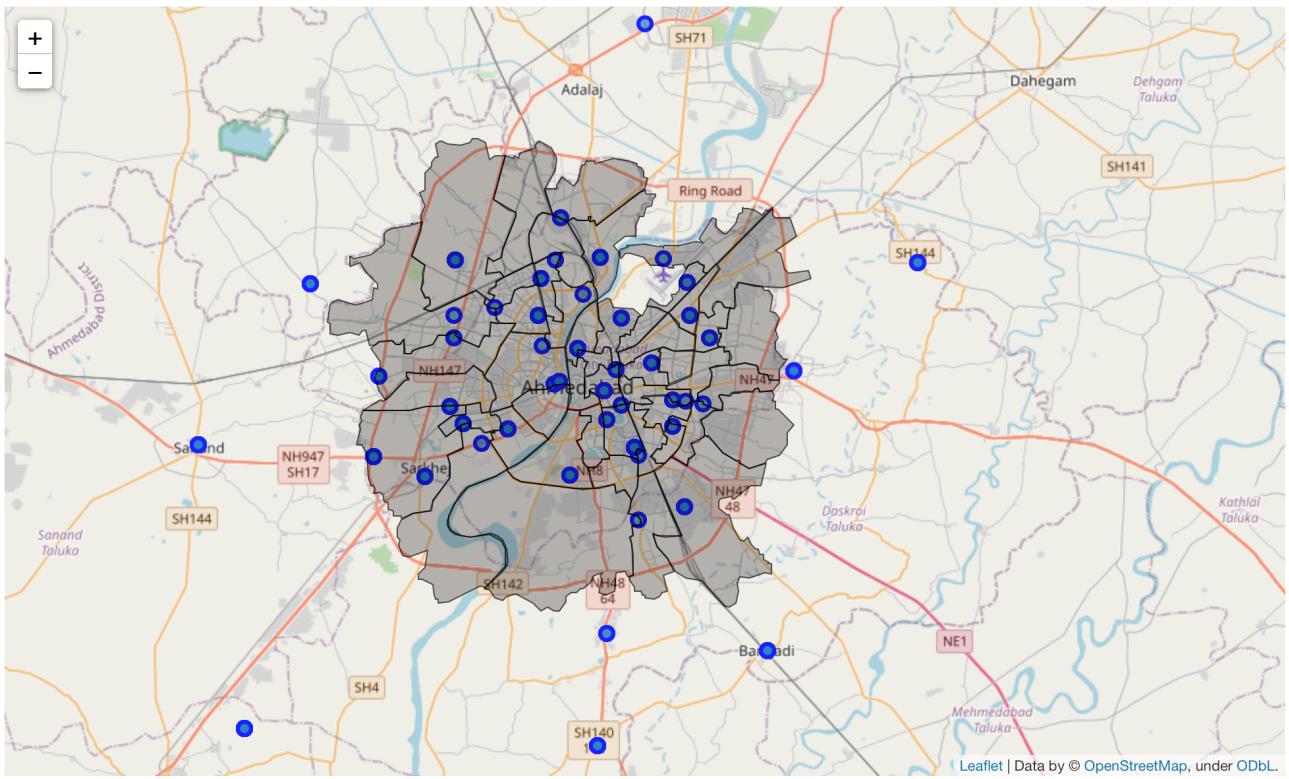


Fig 3.1 Display the map with the plotted coordinates

The next step was to locate all the popular stores or the venues will are the most common in those areas. The main aim to find these venues was to help in clustering the areas. The clustering of the areas was done on the basis of all the common venues of the area. The first step in this was to list all the venues in a particular area, and listing them into a datagram with their coordinates. And then counting the number of venues for each area.

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ahmedabad.	23.019845	72.593461	MRPSE SASTA	23.019333	72.594175	Vegetarian / Vegan Restaurant
1	Ahmedabad.	23.019845	72.593461	Astodia Darwaza	23.017156	72.591002	Historic Site
2	Ahmedabad.	23.019845	72.593461	Mangaldas ni Haveli II	23.023121	72.591255	Bed & Breakfast
3	Ahmedabad.	23.019845	72.593461	Purohit Dinning hall	23.023365	72.595322	Restaurant
4	Ahmedabad.	23.019845	72.593461	Lucky Tea	23.016723	72.590728	Tea Room
5	Ahmedabad.	23.019845	72.593461	Bhagyoday Restaurant	23.015734	72.592683	Indian Restaurant
6	Ambaliara	22.722057	72.436105	YES Bank ATM	22.719334	72.438113	Bank
7	Ambareli	22.722057	72.436105	YES Bank ATM	22.719334	72.438113	Bank
8	Ambawadi	23.022584	72.567249	The Westend	23.022107	72.567869	Hotel
9	Ambawadi	23.022584	72.567249	Wow Mughlai Handi and BBQ Grill	23.020332	72.569546	Indian Restaurant

Table 3.2 Displays all the venues in a certain area

After making a list of all the venues located, the next step was to detail all the areas and which venues did they have. On that basis, I was able to calculate the frequency of a particular type of venues in the area.

----Ahmedabad.----

		venue	freq
0	Restaurant	0.17	
1	Indian Restaurant	0.17	
2	Tea Room	0.17	
3	Bed & Breakfast	0.17	
4	Historic Site	0.17	

----Ambaliara----

	venue	freq
0	Bank	1.0
1	ATM	0.0
2	Moving Target	0.0
3	Planetarium	0.0
4	Pizza Place	0.0

Fig 3.2 Display the frequency of each venue in the area

With the frequency we are able to calculate the most common venues in the area, and then we add it to a data frame which consists of 1st, 2nd, till the 10th most common venue of the area.

	Location	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	Ahmedabad.	Indian Restaurant	Vegetarian / Vegan Restaurant	Restaurant	Tea Room	Historic Site	Bed & Breakfast	Football Stadium	Fast Food Restaurant	Fireworks Store	Fish Taverna
1	Ambaliara	Bank	Zoo	Gift Shop	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market	Food Court	Food Truck	Football Stadium
2	Ambareli	Bank	Zoo	Gift Shop	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market	Food Court	Food Truck	Football Stadium
3	Ambawadi	Indian Restaurant	Hotel	Snack Place	Vegetarian / Vegan Restaurant	Hotel Bar	IT Services	Electronics Store	Fast Food Restaurant	Fireworks Store	Fish Taverna
4	Ambawadi Vistar	Fast Food Restaurant	Indian Restaurant	BBQ Joint	Café	Restaurant	Coffee Shop	Mexican Restaurant	Pizza Place	Tea Room	Vegetarian / Vegan Restaurant
5	Amraiwadi	ATM	Indian Restaurant	American Restaurant	Gift Shop	Fireworks Store	Fish Taverna	Flea Market	Food Court	Food Truck	Football Stadium
6	Anandnagar	Indian Restaurant	Donut Shop	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market	Food Court	Food Truck	Football Stadium	Furniture / Home Store
7	Andej	Lake	Zoo	Furniture / Home Store	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market	Food Court	Food Truck	Football Stadium
8	Asarwa Chakla	Indian Restaurant	Restaurant	Tennis Stadium	Food Truck	Football Stadium	Electronics Store	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market
9	Asarwa Ext south	Bookstore	IT Services	Toy / Game Store	Snack Place	Zoo	Fast Food Restaurant	Fireworks Store	Fish Taverna	Flea Market	Food Court

Table 3.3 Display the 10 most common venues in the area

Now with all this data we are able to cluster the areas into five different groups on the basis of the most common venues present in them and plot them on the map of Ahmedabad.

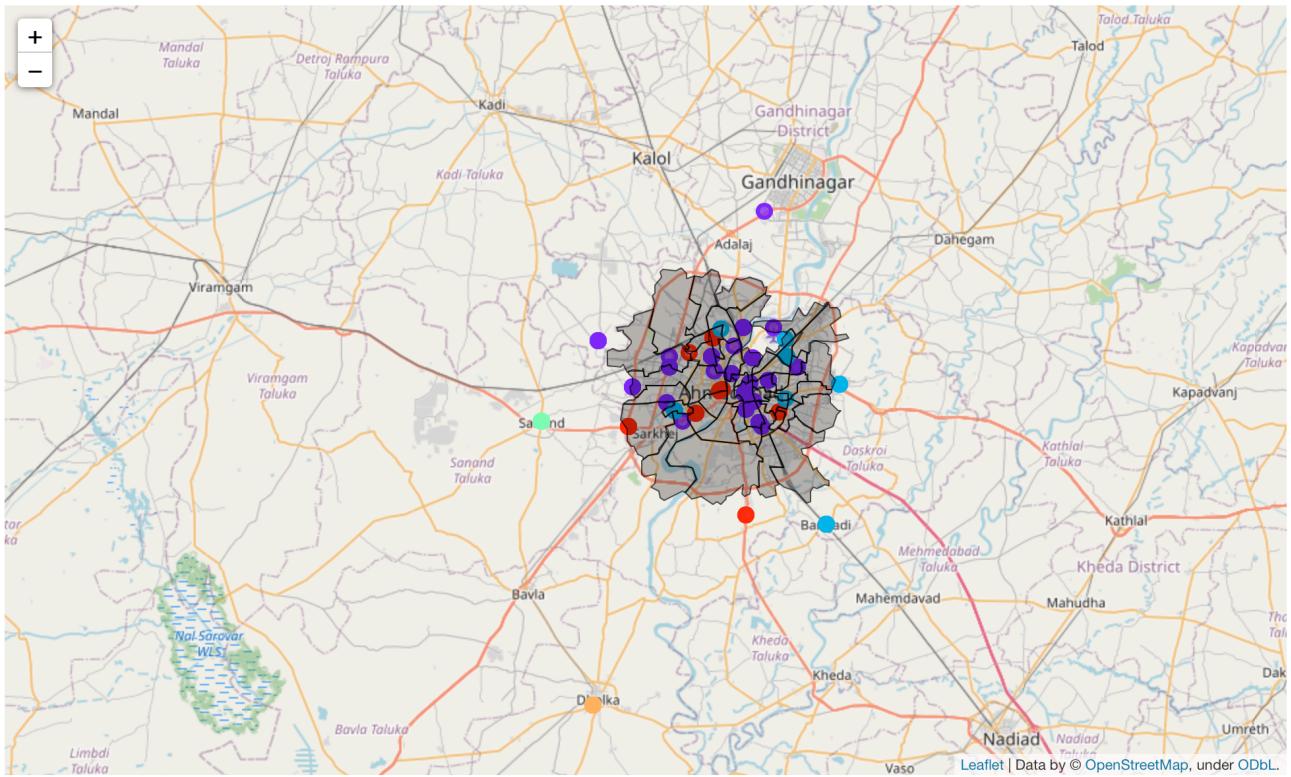


Fig 3.3 Displays the clusters plotted on the map of Ahmedabad

4. Conclusion

In this study, I classified and clustered all the areas in the Ahmedabad district of Gujarat so that buyers are able to identify the a suitable area for their purpose keeping an eye on the price of the area. This study helps the buyer find an area which is financially affordable and satisfies all their needs, and it also provides alternatives to them. A buyer looking for an house can find a suitable area, a businessman can find an area to for his/her business.

5. Future Directions

With more data available for the city of Ahmedabad, we will be able to carry out more analysis for the same. As different aspects like schooling data, crime etc are not available to us, with the help of those datasets we can predict the prices of the areas. In this study, we use a dataset already provided to us online. We need more data in order to have a better prediction model for the land prices in the city.