Analysis of Potential Small Business and Real Estate Investment Opportunities in Smart Cities.

- Author: Leron Menezes

1. Introduction

India is a developing country, from a Gross Domestic Product of $1.676 Trillion in 2010 it has grown to a $2.875 Trillion economy in 2019 (The World Bank, 2020). This means that it has a real potential for growth and investment opportunities. Through this study, the aim is to identify the ideal business type and location to make an investment in the below selected cities:

a. Ahmedabad
b. Bengaluru
c. Kalyan - Dombivili
d. Nagpur
e. Pimpri - Chinchwad
f. Pune
g. Thane
h. Visakhapatnam

The target audience of this study would be real estate investors, entrepreneurs, members of the business community who are looking to expand their real estate portfolios and businesses into new locations with significant promise of growth and low input costs. This study is also aimed at local government administrative officials who are looking to explore the types of businesses in their localities and how they can make decisions to influence growth in investments in their administrative areas.

These cities have been selected due to their status as a smart city. With the launch of the Smart Cities Mission in 2015, India started paving a new pathway towards transforming urban management with the power of digital technologies. The concept recognizes the value of enhancing engagement among all four stakeholders of the quadruple-helix model—Government, citizens, academia, and industry, along with improvements in the internal workflow and decision-making processes of city Governments. In this context, the need for city Governments to take ‘digital leadership’ has become more pronounced (Ministry of Home Affairs, Government of India, 2020).

This project aims to leverage the data available through the smart cities programme to identify the best investment opportunities that are available for small businesses or real estate in these cities.

1.1 What are Smart Cities?

The first question is what is meant by a 'smart city'. The answer is, there is no universally accepted definition of a smart city. It means different things to different people. The conceptualisation of Smart City, therefore, varies from city to city and country to country, depending on the level of development, willingness to change and reform, resources and aspirations of the city residents. A smart city would have a different connotation in India than, say, Europe. Even in India, there is no one way of defining a smart city.

Some definitional boundaries are required to guide cities in the Mission. In the imagination of any city dweller in India, the picture of a smart city contains a wish list of infrastructure and services that describes his or her level of aspiration. To provide for the aspirations and needs of the citizens, urban planners ideally aim at developing the entire urban eco-system, which is represented by the four pillars of comprehensive development-institutional, physical, social and economic infrastructure. This can be a long term goal and cities can work towards developing such comprehensive infrastructure incrementally, adding on layers of 'smartness' (Government of India, n.d.).

1.2 The Idea

Through the use of the Foursquare API the small businesses and venues in the cities will be identified. Assuming that the theory of the network effect holds in any urban agglomeration, any investment in these sorts of businesses in these cities should see a positive return on investment due to an existing infrastructure, business friendly policies and a customer and supplier pool. A network effect (also called network externality or demand-side economies of scale) is the effect described in economics and business that an additional user of goods or services has on the value of that product to others. When a network effect is present, the value of a product or service increases according to the number of others using it. The classic example is the telephone, where a greater number of users increases the value to each. A positive externality is created when a telephone is purchased without its owner intending to create value for other users, but does so regardless. Online social networks work similarly, with platforms like Twitter, Facebook, and WhatsApp increasing in value to each member as more users join (Wikipedia, n.d.).

After the popular types of businesses in the area are obtained, the locations are clustered based on those businesses to form Business Clusters.

a. Transport
b. Utilities
c. Education
d. Taxes
e. Healthcare
f. Crime

Once the business clusters are identified using the above measures in-cooperation with housing price index, a decision on the type of business and location to make an investment can be made.

2. The Data

2.1 Location data:

- A list of postal codes from all over India, was obtained with the available latitude and longitude data from the www.geonames.org. This data is licensed under the creative commons and is free to use.
- Using the Foursquare API we will obtain the list of venues in these postal codes.

2.2 City Indicators:

- The data related to Transport, Utilities, Education, Taxes, Healthcare, Crime and Natural Calamities were obtained from the Smart Cities portal.

2.3 Housing Price Index:

- The housing price index data was obtained from the National Housing Bank City-wise Housing Price Indices data

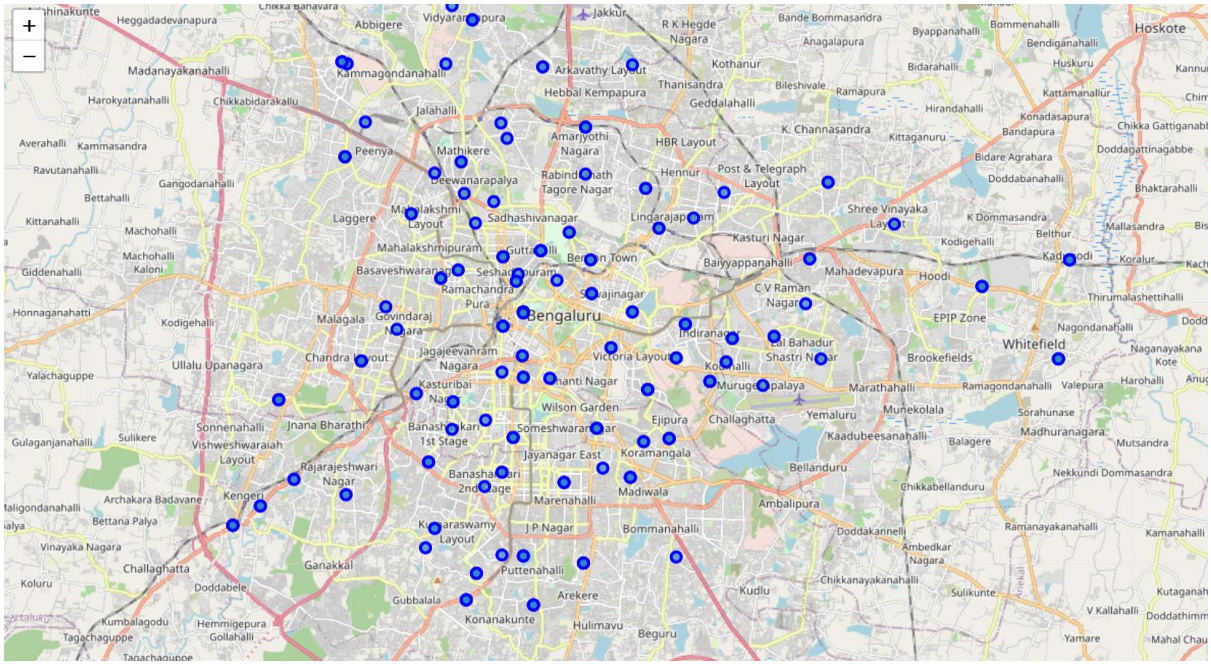
3. Methodology

The concerned postal codes in the city was obtained using google. Then these postal codes were picked up from the below data set which was created by **geonames.org**

	country_code	postal_code	place_name	admin_name_1	admin_code_1	admin_name_2	admin_code_2	admin_name_3	admin_code_3	latitude	longitude	accuracy
0	IN	744101	Marine Jetty	Andaman & Nicobar Islands	1	South Andaman	NaN	Portblair	NaN	11.6667	92.7500	3
1	IN	744101	Port Blair	Andaman & Nicobar Islands	1	South Andaman	NaN	Port Blair	NaN	11.6667	92.7500	4
2	IN	744101	N.S.Building	Andaman & Nicobar Islands	1	South Andaman	NaN	Portblair	NaN	11.6667	92.7500	3
3	IN	744102	Haddo	Andaman & Nicobar Islands	1	South Andaman	NaN	Port Blair	NaN	11.6833	92.7167	4
4	IN	744102	Chatham	Andaman & Nicobar Islands	1	South Andaman	NaN	Portblair	NaN	11.7000	92.6667	3

However, as the data set by geonames was not exhaustive for more location data the **Geocoder Open Street Maps API** was used.

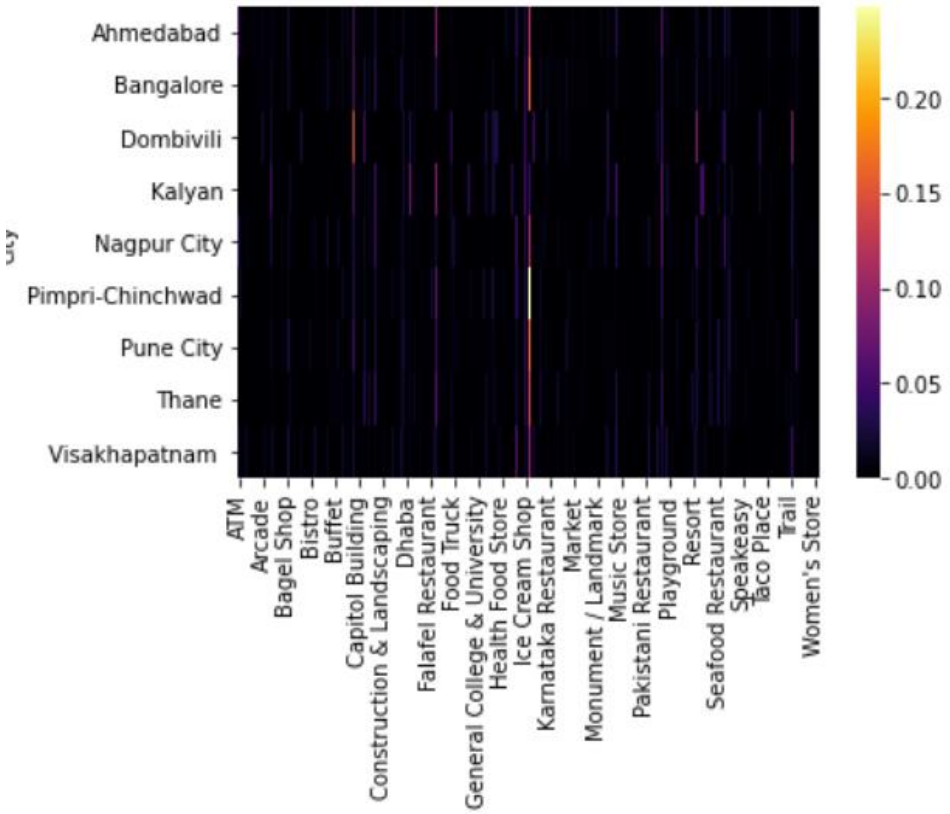
After the locations were obtained, a quick visualisation on the sanity of the locations were done using the **Folium** library. Locations which were too far out or unnecessary were removed from the data set.



The **FourSquare API** was used then to obtain the venues and businesses at each postal code. A limit of 200 venues were set on the result of each location call and a radius of 2.5 kilometres.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	380001	23.019812	72.59372	Manek Chowk Khau Gali	23.023505	72.588539	Snack Place
1	380001	23.019812	72.59372	Manek Chowk	23.023626	72.588553	Fast Food Restaurant
2	380001	23.019812	72.59372	Lucky Tea	23.027829	72.581394	Tea Room
3	380001	23.019812	72.59372	Agashiye	23.027104	72.581614	Indian Restaurant
4	380001	23.019812	72.59372	Jama Masjid	23.024323	72.587042	Historic Site

From the resulting data that was created using the venue data, to identify the popular venues by city a heat map was created. What is observed is that, ice cream joints and Indian restaurants are the more common business types across all the selected cities.

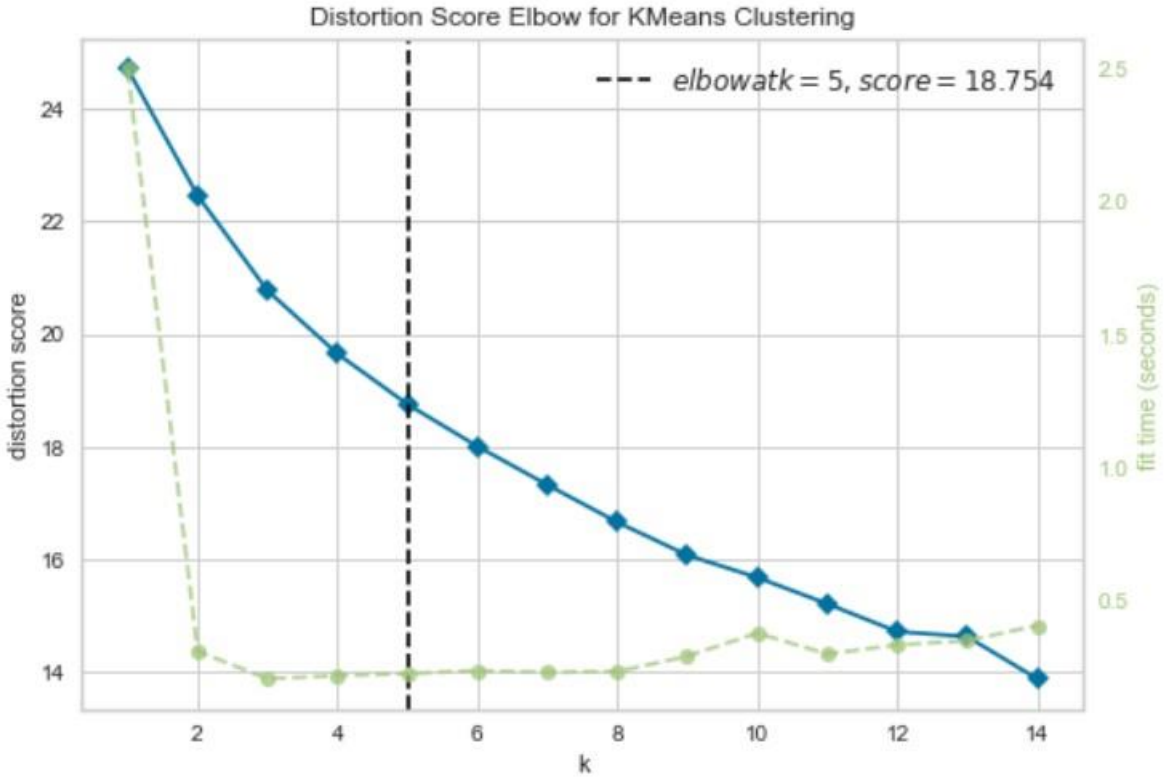


The below image of the data consolidated by city proves the assumption from the heat map true.

	city	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	Fast Food Restaurant	Pizza Place	Café	ATM	Hotel	Multiplex	Shopping Mall	Bakery	Sandwich Place
1	Bangalore	Indian Restaurant	Café	Ice Cream Shop	Fast Food Restaurant	Pizza Place	Coffee Shop	Restaurant	Department Store	Hotel	Chinese Restaurant
2	Dombivili	Café	Restaurant	Train Station	Chinese Restaurant	Indie Movie Theater	Pizza Place	Ice Cream Shop	Food Court	Supermarket	Shopping Mall
3	Kalyan	Fast Food Restaurant	Pizza Place	Diner	Café	Ice Cream Shop	Road	Coffee Shop	Asian Restaurant	Multiplex	Indian Restaurant
4	Nagpur City	Indian Restaurant	Pizza Place	Hotel	Coffee Shop	Restaurant	Shopping Mall	ATM	Train Station	Café	Multiplex
5	Pimpri-Chinchwad	Indian Restaurant	Fast Food Restaurant	Hotel	Pizza Place	Café	Snack Place	Gym	Ice Cream Shop	Coffee Shop	Gift Shop
6	Pune City	Indian Restaurant	Café	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Ice Cream Shop	Bakery	Snack Place	Seafood Restaurant	Restaurant	Coffee Shop
7	Thane	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Pizza Place	Chinese Restaurant	Ice Cream Shop	Restaurant	Multiplex	Seafood Restaurant	Hotel
8	Visakhapatnam	Indian Restaurant	Hotel	Train Station	Bakery	Multiplex	Pharmacy	Platform	Pizza Place	Park	Café

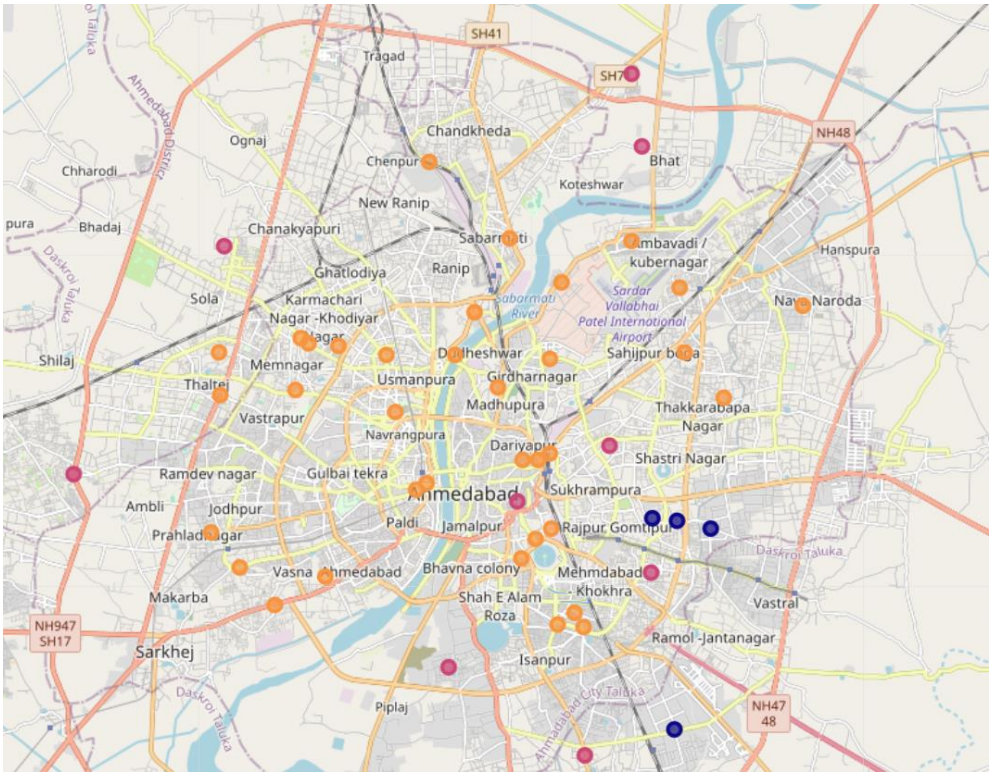
Now to identify the clusters within the cities, to identify common business clusters by the types of venues that exist therein, we cluster the data using the **KMeans unsupervised machine learning algorithm**. This is one of the most common unsupervised learning techniques.

In order to identify the ideal number of clusters the **elbow method** is used. The elbow method gives the number of clusters vs the level of distortion that occurs by increasing the number of clusters by one. After doing this analysis, the ideal number of clusters identified for clustering was 5.

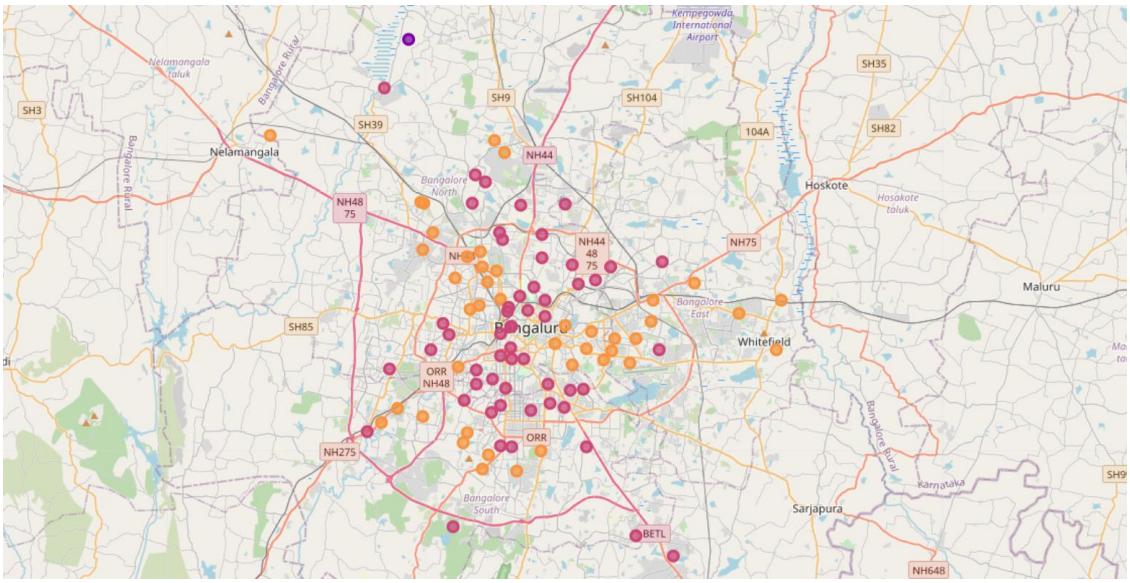


The clustering of the data by 5 clusters was performed and the result of the clustering was plotted to the map. Each cluster was given a colour.

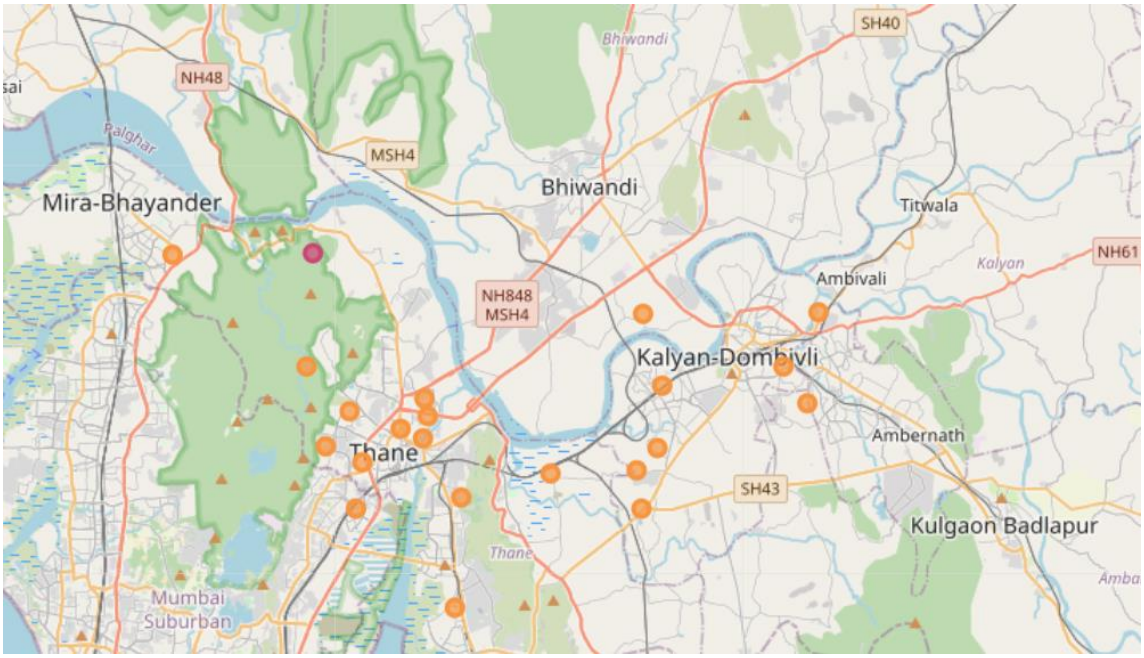
Ahmedabad



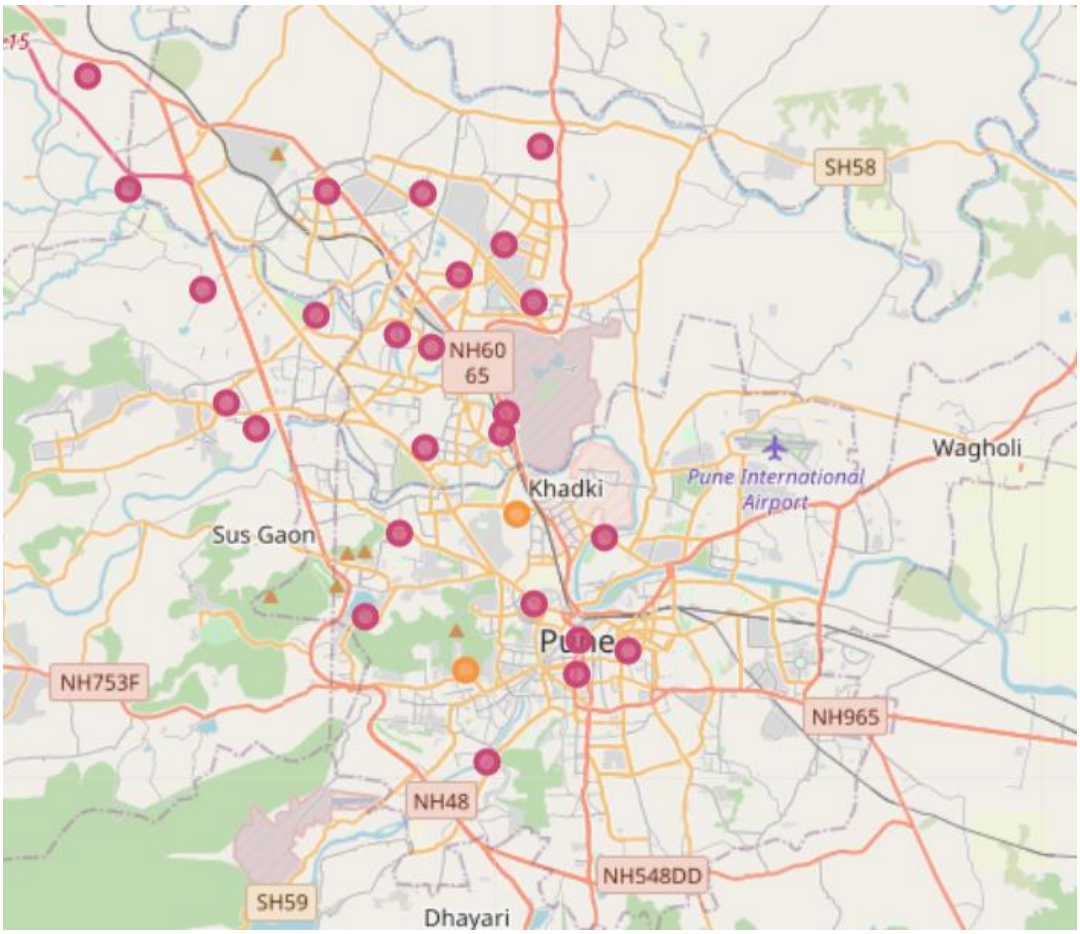
Bangalore



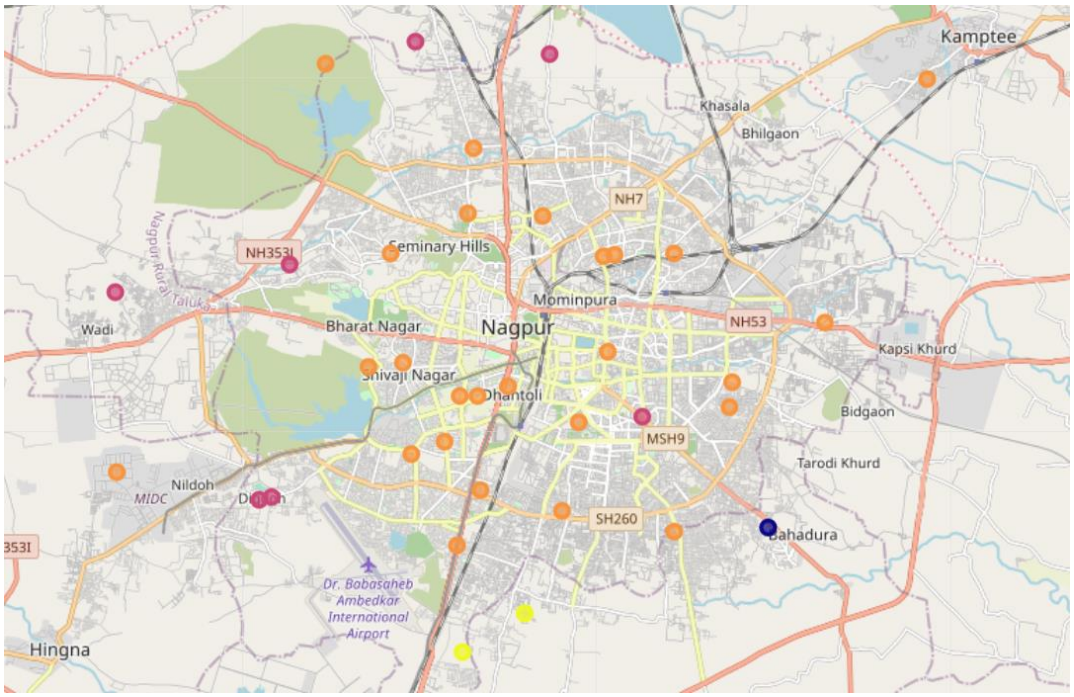
Thane and Kalyan-Dombivli


Bhiwandi
SH40
NH48
Ralghar
MSH4
Mira-Bhayander
Bhiwandi
Titwala
Kalyan
NH61
Ambivali
NH848
MSH4
Kalyan-Dombivli
Thane
Ambernath
SH43
Kulgaon Badlapur
Thane
Mumbai
Suburban

Pune and Pimpri- Chinchwad



Nagpur

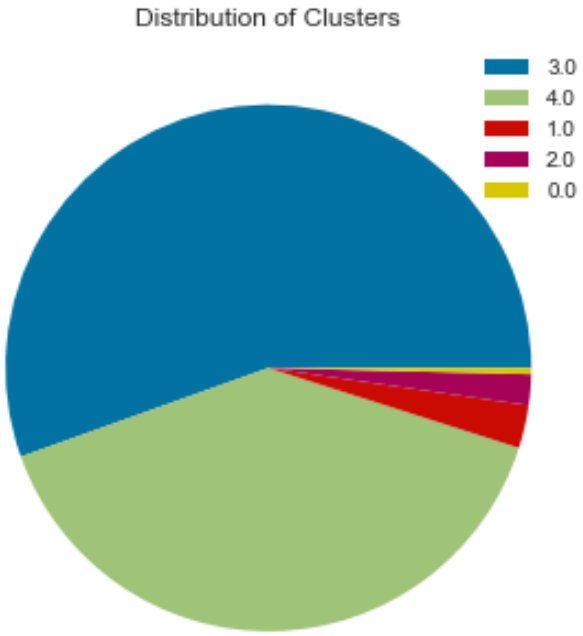


Visakhapatnam

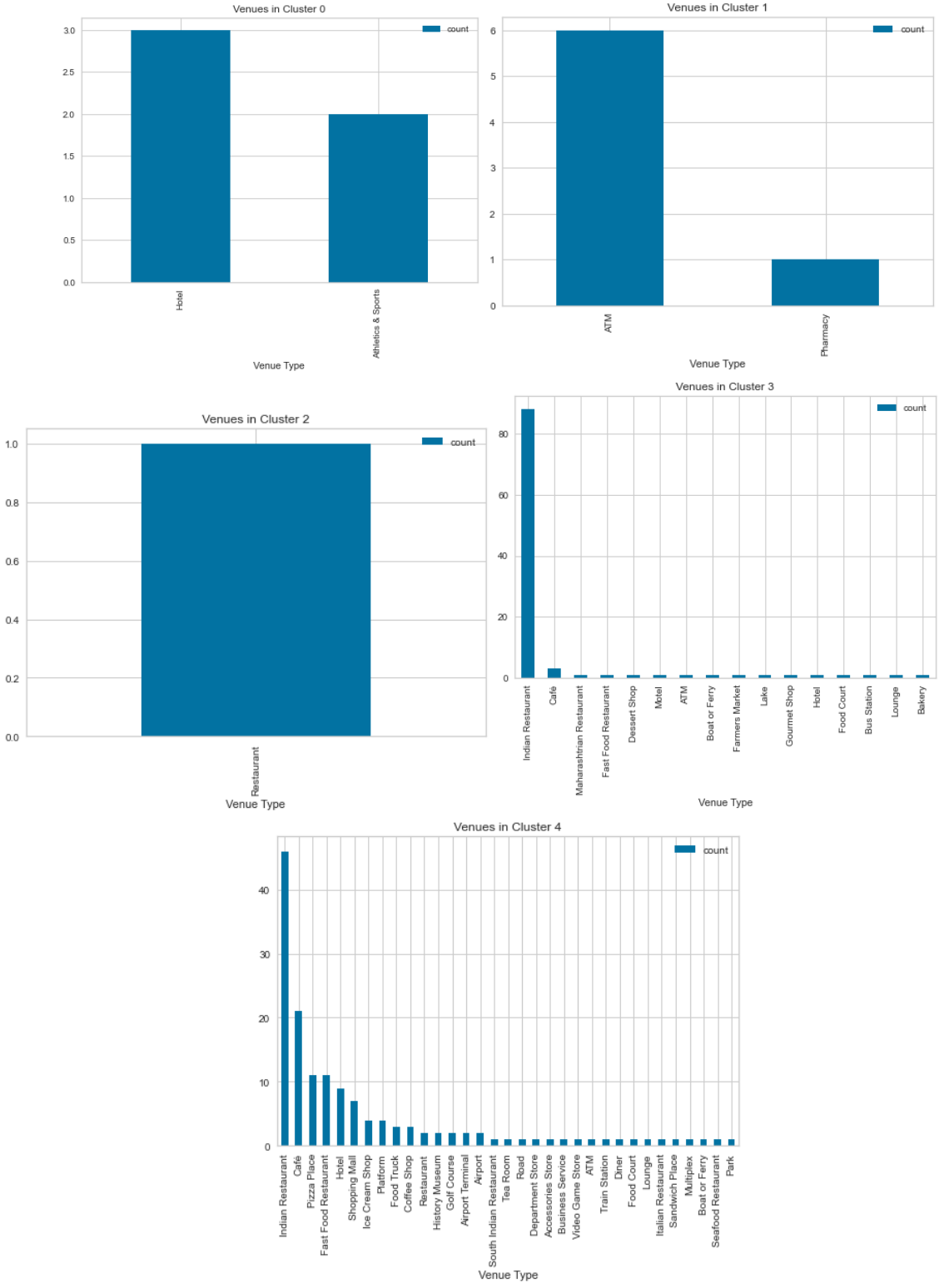
The distribution of the clusters is as below:

Cluster Number	Count
0	5
1	7
2	1
3	105
4	147

We observe that clusters 3 and 4 are the most.



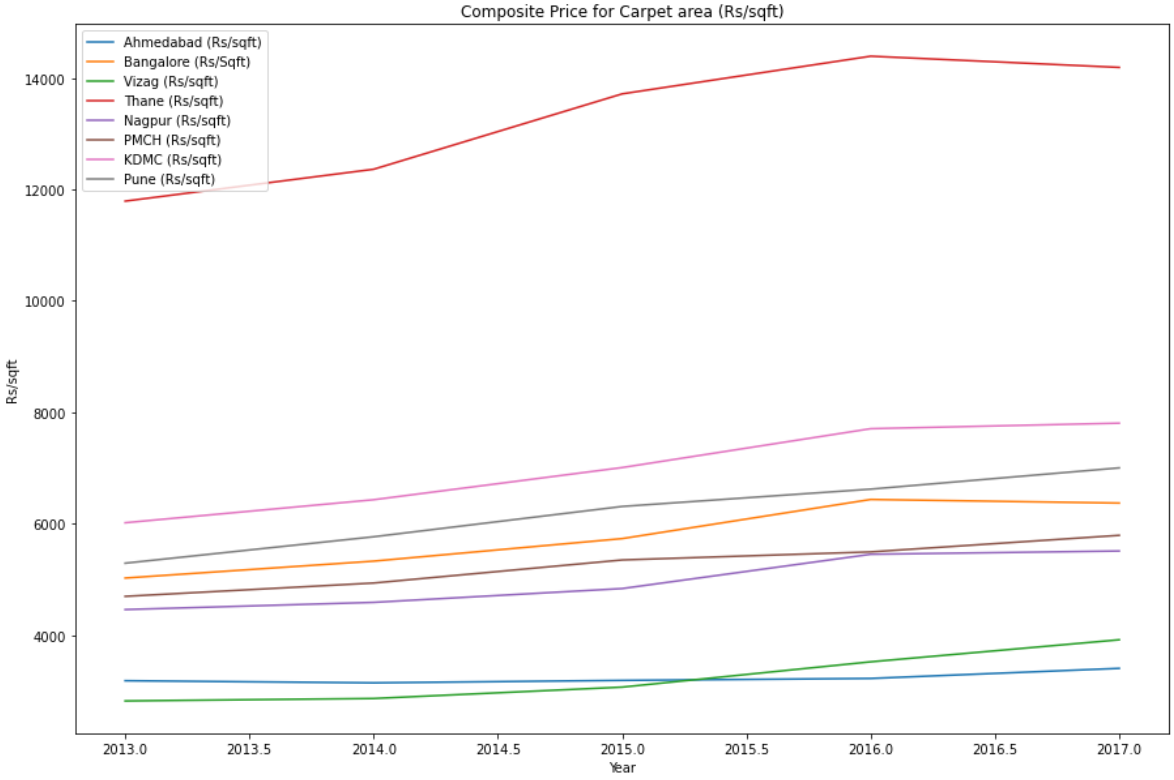
The popular types of businesses by clusters are explored below.



From the above images it can be observed that the clusters 3 and 4 have the most number of businesses and venues. Also from the map shared, it can be observed that cluster number 3 is always found at city centres and cluster number 3 in the suburbs. This is crucial to our decision in selecting a business and a location. Also we see that cafes and Indian restaurants are the most common types of businesses in these clusters.

Now the data related to the cities are explored to find the suitable city to set up a business.

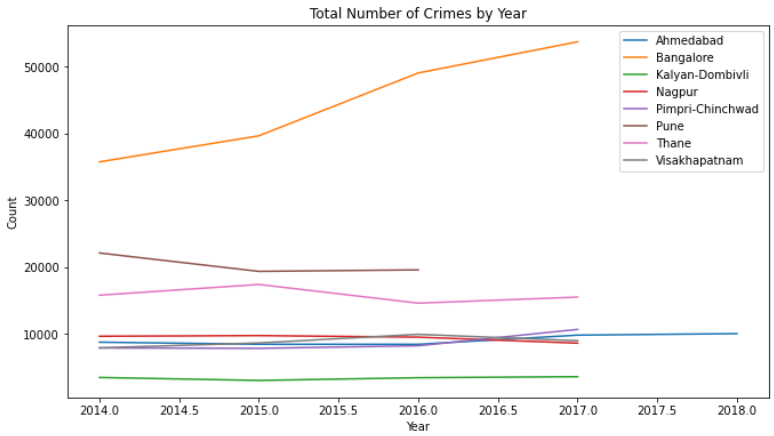
From the **National Housing Bank** (National Housing Bank, n.d.) we obtain the composite prices of the selected cities by Carpet area (Rs/sqft).



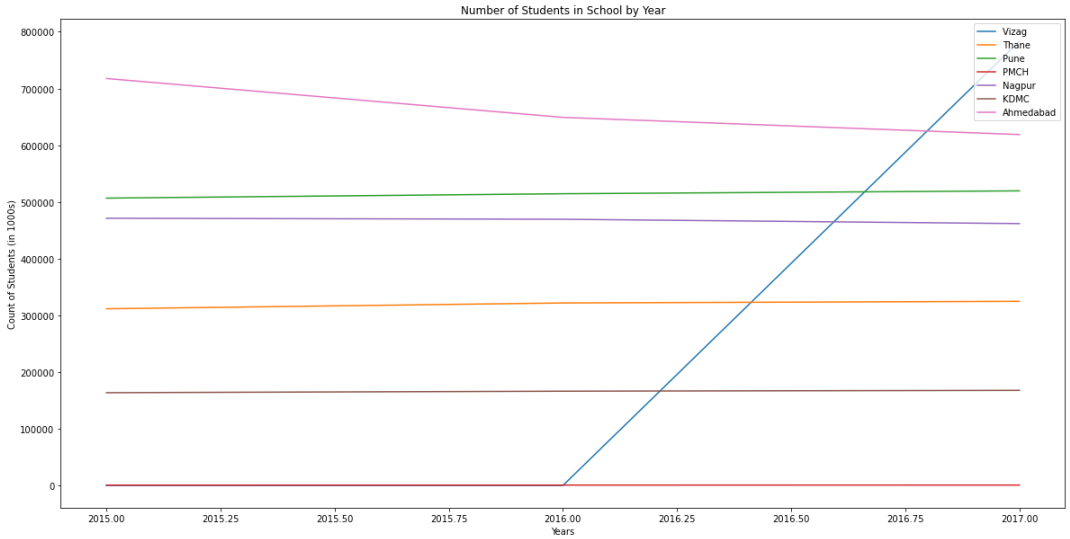
What is observed is the cost of real estate in Thane is extremely high, this is due to it's proximity to Mumbai. However, other smart cities like Kalyan-Dombivli, Pune, Pimpri-Chinchwad, Nagpur and Bangalore are relatively close to each other. But Ahmedabad and Visakhapatnam are the least expensive cities.

So as per this analysis, Bangalore and Pimpri Chinchwad are the cities which are midrange and Ahmedabad and Visakhapatnam are in the lower range.

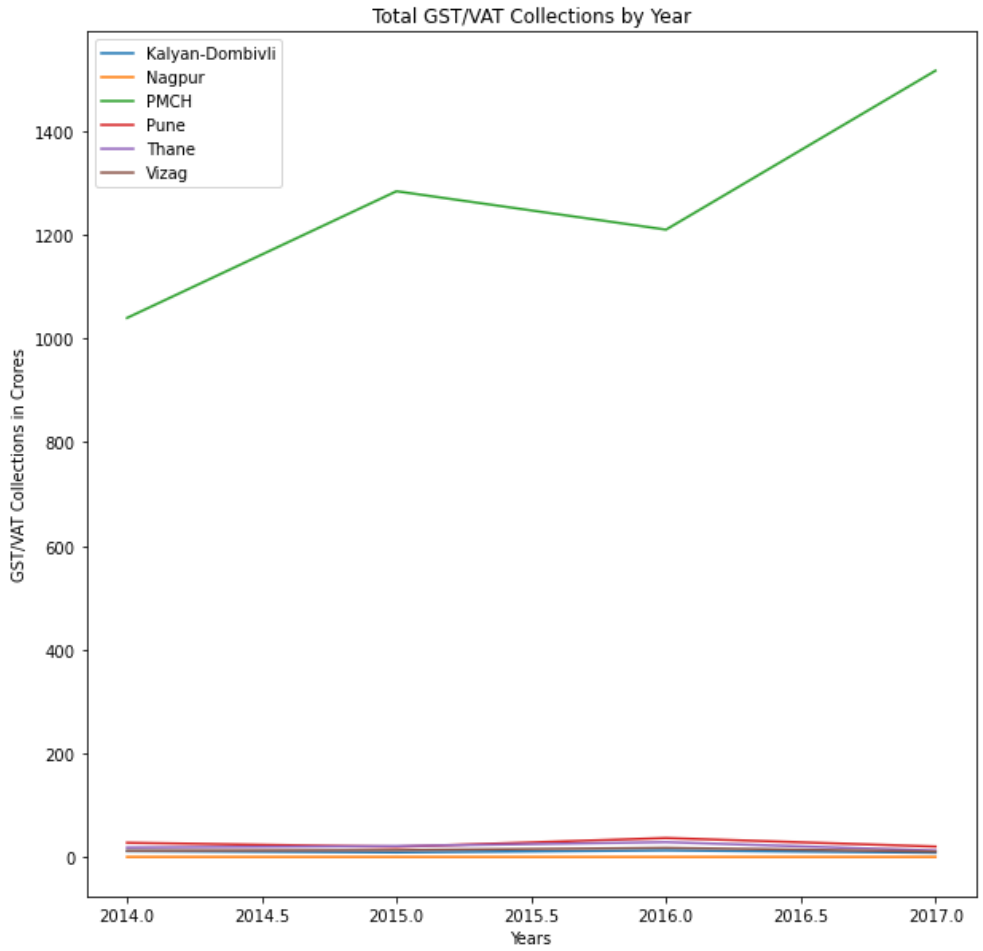
Next we look at Total Crimes per year in these cities. It can be observed that Bangalore has the highest number of reported crimes, whereas Kalyan-Dombivli has the least.



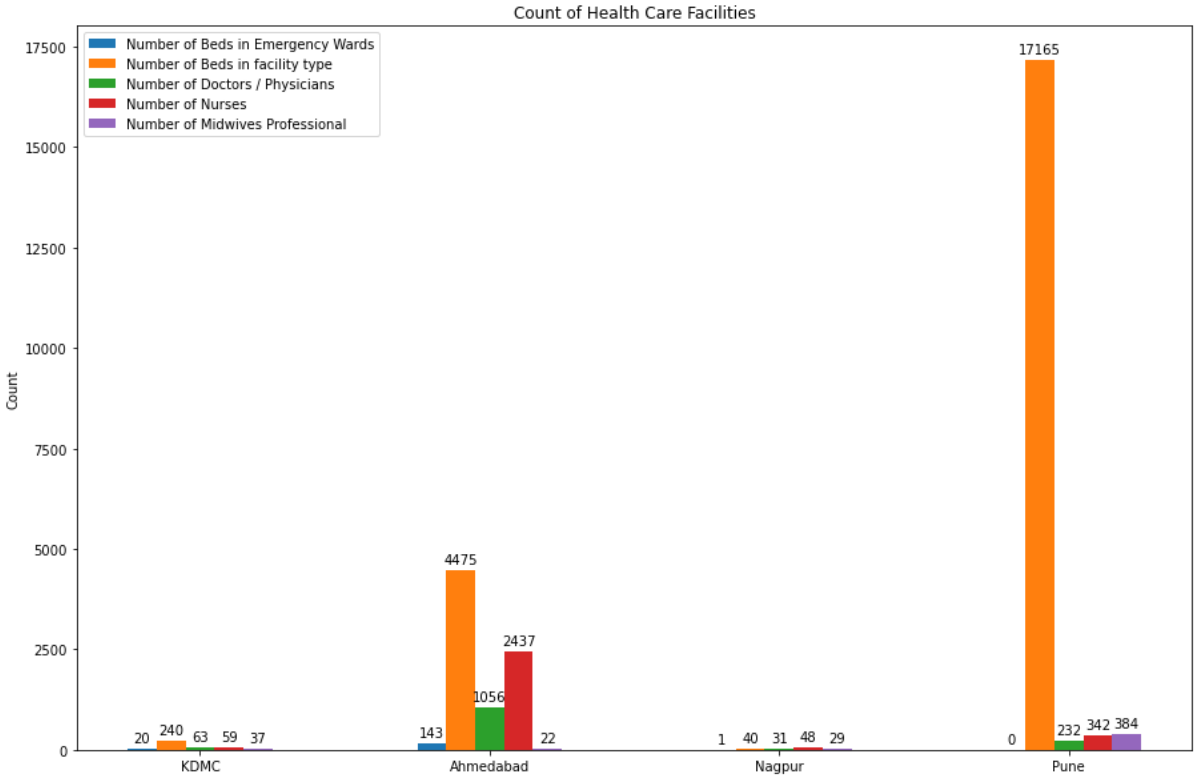
Now the number of students in school are analysed. From this it can be seen that data for Visakhapatnam is missing till 2016. But it is relatively constant for most cities. However, Ahmedabad's school population is decreasing.



The collections for Sales Tax from these cities are plotted below, it can be observed that due to Pimpri Chinchwads huge automobile sector, the sales tax collected in the city is the highest. Whereas the remaining cities are relatively close, with the exception of Nagpur which pays the least.



Lastly we look at the available healthcare facilities in these cities, as this data is not available for most cities we are only able to do this analysis for Kalyan-Dombivli, Ahmedabad, Nagpur and Pune. Pune leads the rest with number of bed facilities, however, Ahmedabad has a greater number of staff available for Healthcare.



4. Results

From the above data exploration and clustering, the cities for consideration for investment based on the price of real estate and the other measures we observe that the cities with the best opportunity are Ahmedabad and Pune.

The cost of real estate is low and the venues that are present are from Cluster 3 and 4 and the indicators for these are positive.

The most common types of businesses are restaurants and café's, the clusters that are in the city centre have a wider variety of venues than the ones in the suburbs. That coupled with the fact that, these cities each have their own unique strong point (Ahmedabad is a buddying financial services centre and Pune is an Automobile hub) means growth for the investor.

The remaining cities have high real estate costs and negative indicators in the above KPI's that were explored.

5. Discussion

The cities with high real estate costs are near existing Tier 1 cities, hence there will be existing demand. However, it will mean lower ROI for a small business investor. Then amongst these cities is Bangalore, which is a Tier 1 city with low real estate costs. But the population in Bangalore is spread out across a large area with multiple sub industrial areas, hence the low demand for cluster 4 type businesses.

Cities like Ahmedabad and Pune are however have city centre like clusters and a growing business and citizen population.

But cities like Nagpur and Visakhapatnam are mostly industrial areas, with low civic population hence the low real estate cost and low demand for specialised venues.

Cities like Thane and Kalyan-Dombivli are part of the MMRDA, which already has existing infra concentrated around Mumbai. Which makes Mumbai the centre of business. Hence they are not considered.

6. Conclusions

The smart cities initiative is still starting out, the maturity of data available and the availability of data across the different cities still have opportunities for growth. The real estate costs due to COVID 19 have stagnated and may even reduce in the coming years. So this analysis is very important for the investor to consider. In this manner they might get a bargain for their investment.

References

1. Government of India. (n.d.). *Smart Cities*. Retrieved from About Smart Cities: http://smartcities.gov.in/content/innerpage/what-is-smart-city.php
2. Ministry of Home Affairs, Government of India. (2020, August). *DataSmart Cities Strategy.* Retrieved from DataSmart Cities: https://smartnet.niua.org/dsc/datasmartcity-stratergy.php
3. National Housing Bank. (n.d.). *NBH Residex*. Retrieved from NBH ONLINE: https://residex.nhbonline.org.in/
4. The World Bank. (2020, August). *worldbank.org*. Retrieved from Data: https://data.worldbank.org/country/india
5. Wikipedia. (n.d.). *Network Effect*. Retrieved from Wikipedia.Org: https://en.wikipedia.org/wiki/Network_effect