Housing rent price and venues Data Analysis of Navi Mumbai

A.Introduction

Navi Mumbai is a planned city off the west coast of the Indian state of Maharashtra .Navi Mumbai has a population of 1,119,477[1] as per the 2011 provisional census. The population of Bombay(now Mumbai) rose from 2.966 million in 1951 to 4.152 million in 1961 and to 5.970 million in 1971, registering 40.0 and 43.80 per cent growths during the first and second decades respectively. The rapid rate of growth of population, made possible by the increasing industrial and commercial importance of the city, resulted in a fast deterioration in the quality of life for the majority of people living in the city.Hence it was recommended to the then govt,that a new metro center be developed in the outskirt of Mumbai to accommodate the new growth and population

Navi Mumbai is a very fast growing metro.It is very well connected to other parts of the state and country and is comparatively less polluted than Mumbai .It has a very robust infrastructure and a very good public transport system.Apart from major railway junctions,a new airport is being planned in Kharghar which will shoot up the property prices.

Navi Mumbai is now becoming a major industrial/Start up hub now which will lead to even more influx of young people,students,and many other migrants looking for work.Being situated right next to the most expensive place in the world,it will soon become crowded city leading people looking for cheaper houses to rent and also places to hangout in and also well connected to Mumbai/Pune. If we think of the 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 and very well connected to other parts. it is difficult to obtain information that will guide residents in that direction.This is our Business problem.We need a platform to show all the venues and the rent info of the localities in Navi Mumbai

Hence,we can create a map and information chart where the real estate index is placed on Navi Mumbai and each district is clustered according to the venue density.

B.Data Description

To consider the problem we can list the datas as below:

- I used an article by makaan.com to find the rent trend of different areas of Navi Mumbai.I scarped the data and analyzed it
- I used Geocoder library to get the latitude and longitude of all the areas given in makaan.com's article
- I used Foursquare API to get the most common venues of given areas of Navi Mumbai [3].

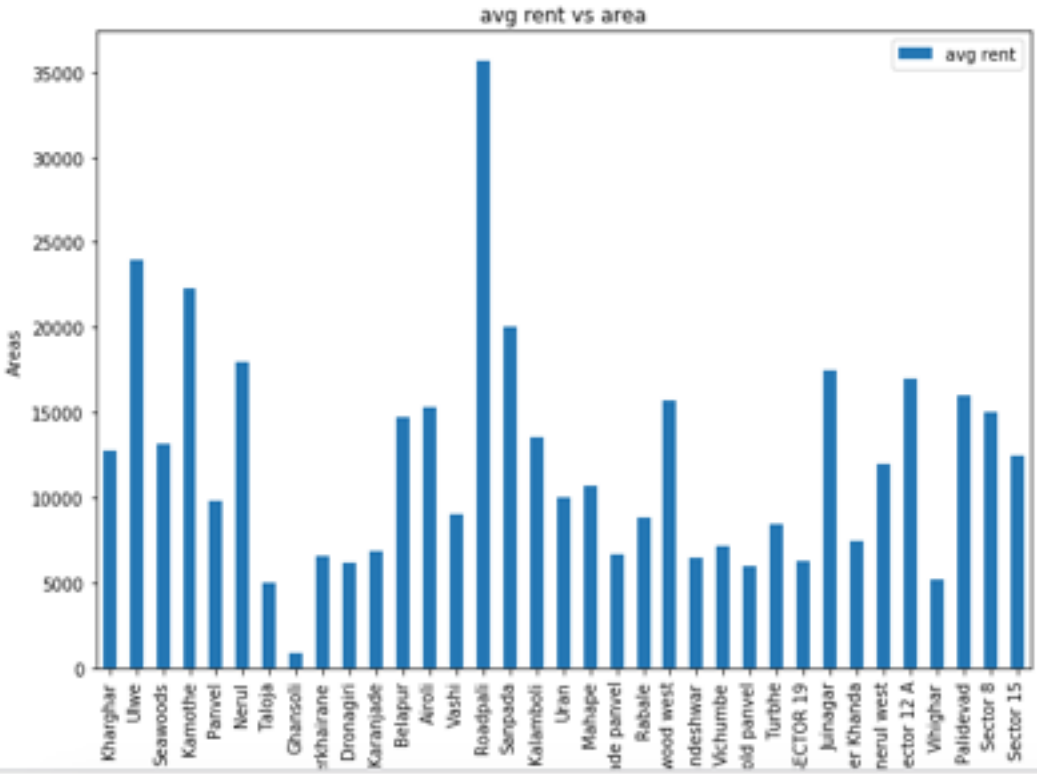
C. Methodology

My master data,which has the avg rent,lat ,long and the locality name is as follows:

	locality	1bhk rent range	avg rent	lat	lon
0	Kharghar	12800	12800.0	19.0258	73.0592
1	Ulwe	24000	24000.0	18.9871	73.0404
2	Seawoods	9,300 - 17,000	13150.0	19.0222	73.0187
3	Kamothe	7,500 - 90,000	22250.0	19.0164	73.0807
4	Panvel	4,500 - 20,000	9850.0	18.9656	73.1029
5	Nerul	18000	18000.0	19.0336	73.0181
6	Taloja	5000	5000.0	19.0615	73.1161
7	Ghansoli	5,000 - 14,000	832.0	19.1193	72.9995
8	Koperkhairane	5,500 - 7,500	6571.0	19.1001	72.9984
9	Dronagiri	4,500 - 12,500	6167.0	18.903	72.9895

Total area with lat and long were 27.I used python folium library to visualize geographic details of Navi Mumbai and its areas and I created a map of Navi Mumbai with its areas superimposed on top. I used latitude and longitude values to get the visual as below-

I analyzed the average rent data of all the places in Navi Mumbai to gauge which area is the most expensive ones and how are the rent prices distributed.



Observations from the graph-

- .Roadpali is the most expensive area to stay and Ghansoli is the cheapest area
- Max number of areas have rent between 7500-15000,only 1 area has above 30k
- Ranges can be defined as follows

  < 7500:L RENT

  > 7500 and < 15000:L MIDDLE RENT

  > 15000 and < 21000:MIDDLE RENT

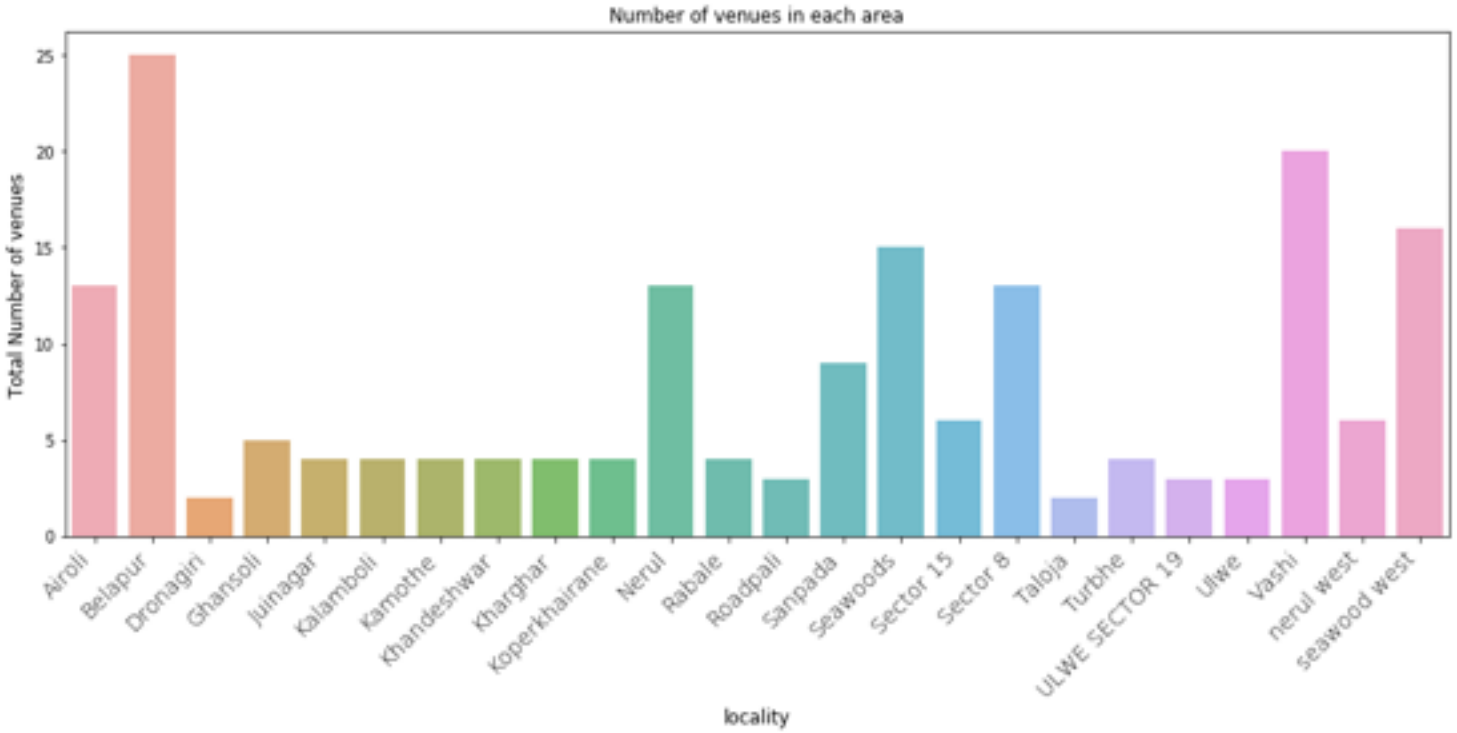
  > 21000 and <28000:H MIDDLE RENT

  <28K:H RENT

I utilized the Foursquare API to explore the locality and segment them. I designed the limit as 100 venue and the radius 700 meter for each locality from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Foursquare API.Total venues returned by Foursquare were 186

	Venue	Venue Latitude	Venue Longitude	Venue Category
	Take Away - 24hrs Food Service	19.029934	73.059587	Fast Food Restaurant
	Kharghar Railway Station	19.026095	73.059125	Train Station
	Cafe Coffee Day	19.029363	73.062915	Café
	Virgin Street Cafe	19.029540	73.060551	Burger Joint
	Pizza Hut	18.986552	73.037475	Pizza Place

Total venues returned by Foursquare were 186.I counted the total number of venues generated by foursquare for each locality and plotted them



Observation from the plot are as follows:

- Belapur has maximum number of venues around 700mts,then Vashi and then Seawood west
- Dronagiri and Taloja have minimum venues
- There are some areas which have no venues around them-Panvel,old Panvel,Turbhe.I removed them from the venue data to form as better clusters as possible
- There are 60 uniques categories.

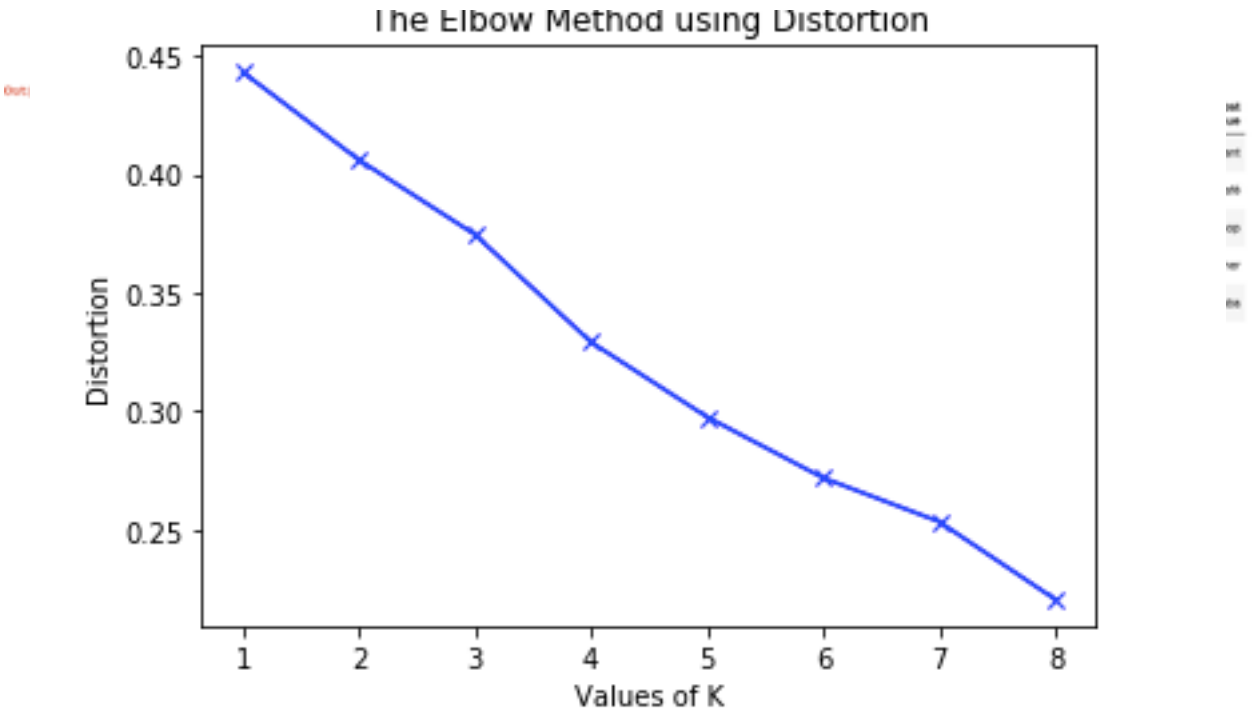
	locality	lat	lon	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kharghar	19.025773	73.059185	Take Away - 24hrs Food Service	19.029934	73.059587	Fast Food Restaurant
1	Kharghar	19.025773	73.059185	Kharghar Railway Station	19.026095	73.059125	Train Station
2	Kharghar	19.025773	73.059185	Cafe Coffee Day	19.029363	73.062915	Café
3	Kharghar	19.025773	73.059185	Virgin Street Cafe	19.029540	73.060551	Burger Joint
4	Ulwe	18.987113	73.040393	Pizza Hut	18.986552	73.037475	Pizza Place
5	Ulwe	18.987113	73.040393	Sagar Khara Dhaba	18.987031	73.038108	Dhaba
6	Ulwe	18.987113	73.040393	The BAKEHOUSE	18.990475	73.038651	Bakery
7	Seawoods	19.022192	73.018738	Seawoods Grand Central	19.020247	73.017808	Shopping Mall
8	Seawoods	19.022192	73.018738	Gupta Sandwich And Snacks	19.024901	73.020976	Sandwich Place
9	Seawoods	19.022192	73.018738	Starbucks	19.021478	73.018222	Coffee Shop

I created a table which shows list of top 10 venue category for each locality in below table.

	locality	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	Airoli	Chinese Restaurant	Bus Station	Food Court	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Indian Restaurant	Italian Restaurant	Coffee Shop	Platform	Restaurant
1	Belapur	Indian Restaurant	Fast Food Restaurant	BBQ Joint	Bakery	Restaurant	ATM	Bus Station	Diner	Chinese Restaurant	Café
2	Dronagiri	ATM	Food & Drink Shop	Vegetarian / Vegan Restaurant	Gym	Garden	Food Court	Fast Food Restaurant	Diner	Dhaba	Dessert Shop
3	Ghansoli	Pharmacy	Surf Spot	Department Store	Park	Gym	Volleyball Court	Food Court	Food & Drink Shop	Fast Food Restaurant	Diner
								Comfort			

We have some common venue categories in localities. This is the reason I used unsupervised learning K-means algorithm to cluster the locality. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the localities into 5 clusters because when I analyze the K-Means with elbow method it ensured me the 5 degree for optimum k of the K-Means.



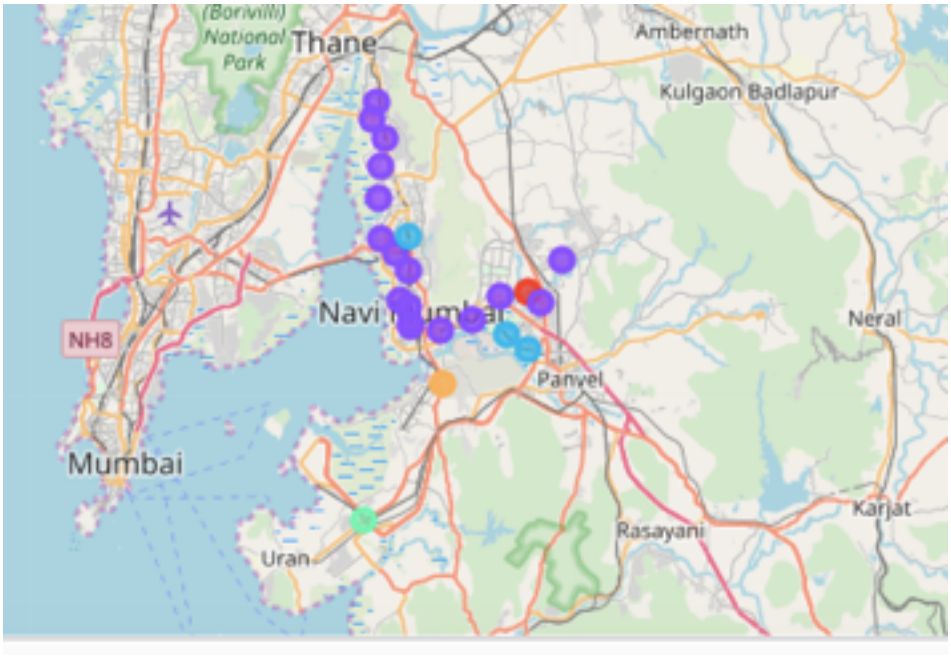
Below you can see a new Cluster table with locality name,its lat/lon,cluster label and 10 most common venue.

	locality	1bhk rent range	avg rent	lat	lon	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	Kharghar	12800	12800.0	19.0258	73.0592	1.0	Train Station	Fast Food Restaurant	Burger Joint	Café	Volleyball Court	Cocktail Bar	Garden	Food Court	Food & Drink Shop	Diner
1	Ulwe	24000	24000.0	18.9871	73.0404	4.0	Pizza Place	Bakery	Dhaba	Volleyball Court	Garden	Food Court	Food & Drink Shop	Fast Food Restaurant	Diner	Dessert Shop
2	Seawoods	9,300 - 17,000	13150.0	19.0222	73.0187	1.0	Chinese Restaurant	Coffee Shop	Gym	Café	Sandwich Place	Department Store	Shopping Mall	Dessert Shop	Fast Food Restaurant	Indian Restaurant
3	Kamothe	7,500 - 90,000	22250.0	19.0164	73.0807	2.0	Train Station	Volleyball Court	Multiplex	Clothing Store	Garden	Food Court	Food & Drink Shop	Fast Food Restaurant	Diner	Dhaba
4	Nerul	18000	18000.0	19.0336	73.0181	1.0	Indian Restaurant	Diner	Bakery	Garden	Café	Fast Food Restaurant	Ice Cream Shop	Food Court	Train Station	Food & Drink Shop

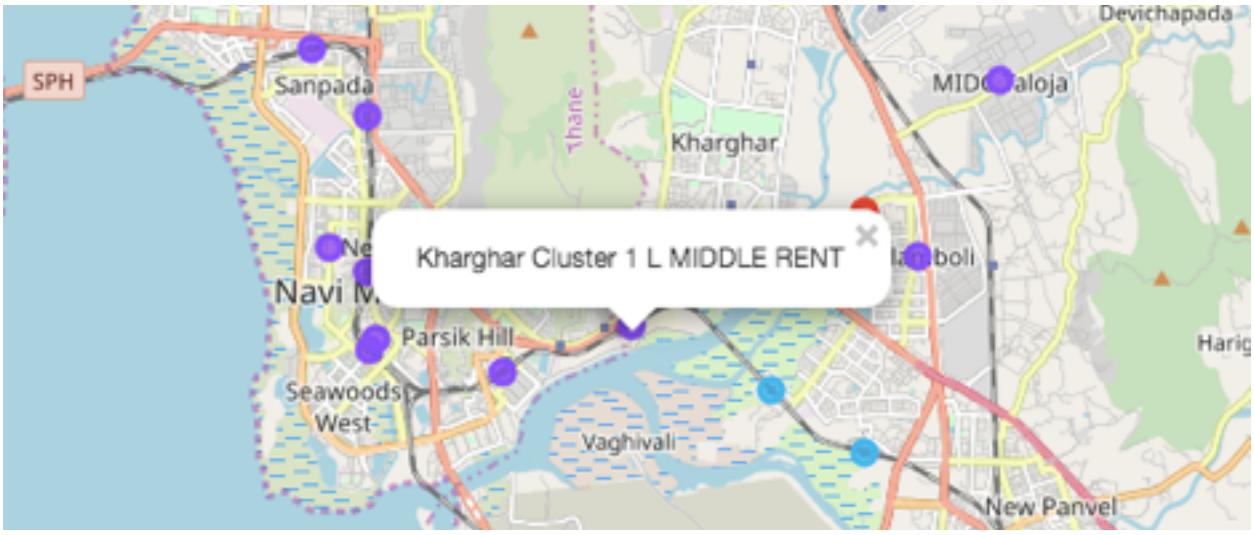
I Merged the cluster table with rent labels table.Hence each locality now has the kind of rental category it belongs to .For example,Row 1 shows,Kharghar locality,it's cluster label,10 most common venues and the Lower Middle rental category it belongs too.Hence,it gets easier for a person to choose the kind of area they want to live in.

	locality	1bhk rent range	avg rent	lat	lon	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	Rent label
0	Kharghar	12800	12800	19.0258	73.0592	1	Train Station	Fast Food Restaurant	Burger Joint	Café	Volleyball Court	Cocktail Bar	Garden	Food Court	Food & Drink Shop	Diner	L MIDDLE RENT
1	Ulwe	24000	24000	18.9871	73.0404	4	Pizza Place	Bakery	Dhaba	Volleyball Court	Garden	Food Court	Food & Drink Shop	Fast Food Restaurant	Diner	Dessert Shop	L MIDDLE RENT
2	Seawoods	9,300 - 17,000	13150	19.0222	73.0187	1	Chinese Restaurant	Coffee Shop	Gym	Café	Sandwich Place	Department Store	Shopping Mall	Dessert Shop	Fast Food Restaurant	Indian Restaurant	L MIDDLE RENT
3	Kamothe	7,500 - 90,000	22250	19.0164	73.0807	2	Train Station	Volleyball Court	Multiplex	Clothing Store	Garden	Food Court	Food & Drink Shop	Fast Food Restaurant	Diner	Dhaba	L MIDDLE RENT
4	Nerul	18000	18000	19.0336	73.0181	1	Indian Restaurant	Diner	Bakery	Garden	Café	Fast Food Restaurant	Ice Cream Shop	Food Court	Train Station	Food & Drink Shop	L MIDDLE RENT

Below is the cluster maps for all the localities. As you can see, 5 clusters were created represented by different colors.



In the final section,I added labels to the cluster markers,if you click on any cluster marker,it ll show you the locality,the cluster it belongs and the rental category it belongs too.Just what someone renting a place wants.



D.Discussion

Navi Mumbai is a big city witch is still developing. As the development is still going on, it is still very difficult to find out which will be the best area for a particular person.You will need more info at that stage.I could not add all the railway stations label,which again would have been really helpful.I used the Kmeans algorithm as part of this clustering study and set the optimum k value to 5.More clusters could have been considered instead of 5.The study ended by visualizing the data and clustering information on the Navi Mumbai map.In future,one can add the railway station info,mall info,or number of schools info around the area.

E.Conclusion

People are turning to Navi Mumbai to get out of Mumbai and still be around the city.Navi Mumbai has head offices of many Start Ups and industries,hence many youngsters are looking to rent places in these areas.For this reason,people can use this platform to get relevant information on where to stay.Not only this,home buyers can also look here for investment opportunities by using the similar data.

F. References

- Navi Mumbai- Wikipedia
- https://www.makaan.com/price-trends/property-rates-for-rent-in-navi-mumbai-mumbai-10045
- Foursquare API
- Google Maps
- Housing Sales Prices & Venues Data Analysis of Istanbul by Sercan Yıldız