Housing Sales Prices & Venues Data Analysis of Mumbai

Sumit Singh

April 28, 2020

1 Introduction

1.1 Background

Mumbai is one of the largest metropolises in the world where over 19.98 million people live and it has a population density of 21000 people per square kilometer. As a resident of this city, I decided to use Mumbai in my project. The city is divided into 4 suburban zone in total. However, the fact that the four suburban zones are squeezed into an area of approximately 603 square kilometers causes the city to have a very intertwined and mixed structure [1].

1.2 Problem

As you can see from the figures, Mumbai 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 suburban zone 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 region 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 Mumbai and each neighbourhood is clustered according to the venue density.

1.3 Interest

Obviously, The interest of investor and person want to live in Mumbai with least real state value and high venue density. The problem solve here is to answer to those questions.

2 Data Description

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

I found the neighbourhoods of Mumbai from Wikipedia [1], [2]. The coordinate of neighbourhoods was obtained from google map [5].
I cleaned the data and reduced it to city of Istanbul where I used it to create choropleth map of Housing Sales Price for each neighbourhoods was obtained from 99acres property dealing site [4]. I used Forsquare API to get the most common venues of given Neighbourhoods of Mumbai [3]. I used Google Map, 'Search Nearby' option to get the center coordinates of the each Neighbourhood. [5]. All compiled data was also stored in Github repository

3   Methodology

As a database, I used GitHub repository in my study. My master data which has the main components Neighborhood, Suburbs_Zone, Latitude, Longitude, Avarage House Price and Grid information of the city.

	Neighborhood	Suburbs_Zone	Latitude	Longitude	Avarage House Price	grid
0	Andheri	Western Suburbs	19.119000	72.847000	18000	19
1	Marol	Western Suburbs	19.117300	72.884000	16000	19
2	Bhayandar	Western Suburbs	19.290000	72.850000	7500	ot
3	Bandra	Western Suburbs	19.054444	72.840556	32000	23
4	Borivali (West)	Western Suburbs	19.238100	72.852300	16000	11

I used python folium library to visualize geographic details of Mumbai and its Neighbourhoods and I created a map of Mumbai with Neighbourhoods superimposed on top coloured with Suburbs_Zone. I used latitude and longitude values to get the visual as below:
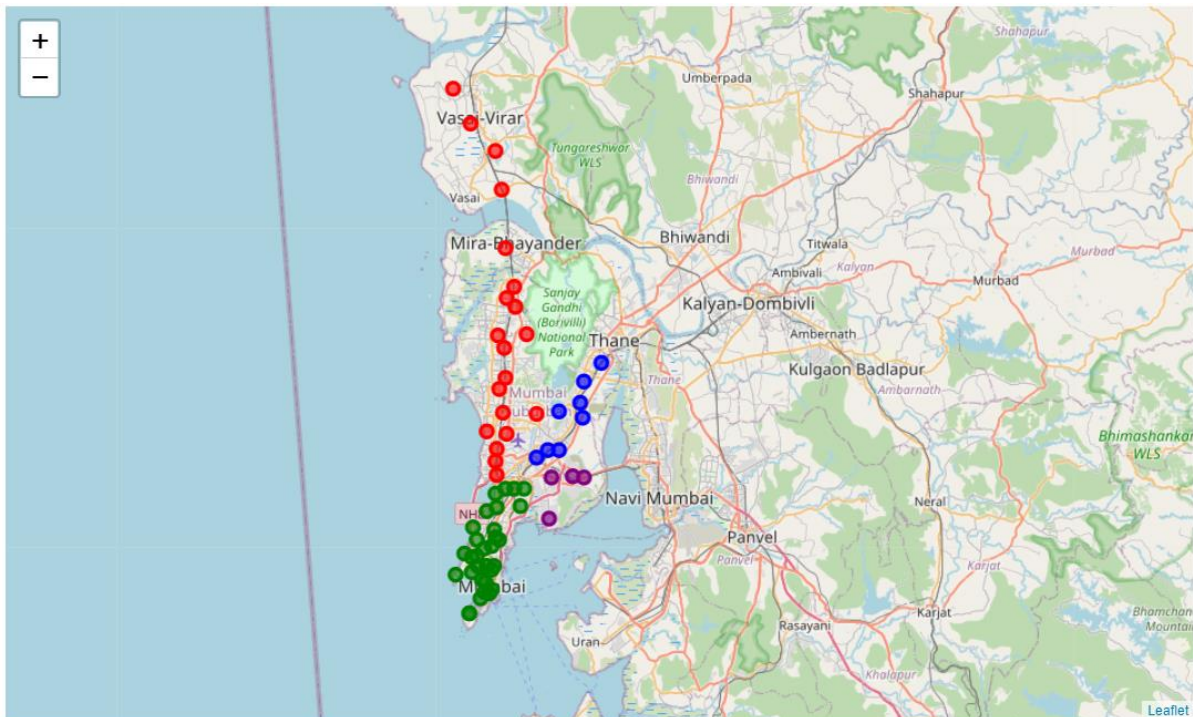

I utilized the Foursquare API to explore the Neighbourhoods and segment them. I designed the limit as 100 venue and the radius 500 meter for each Neighbourhoods from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

In summary of this data 937 venues were returned by Foursquare. Here is a merged table of boroughs and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Andheri	19.119	72.847	Merwans Cake shop	19.119300	72.845418	Bakery
1	Andheri	19.119	72.847	Narayan Sandwich	19.121398	72.850270	Sandwich Place
2	Andheri	19.119	72.847	McDonald's	19.119691	72.846102	Fast Food Restaurant
3	Andheri	19.119	72.847	Cafe Alfa	19.119667	72.843560	Indian Restaurant
4	Andheri	19.119	72.847	Vaibhav Restaurant	19.118235	72.847991	Indian Restaurant

The result doesn't mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude informations and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighborhood informations with more Latitude and Longitude informations. Below graph shows top 20 Neighborhood as per Number of Venue.

Top 20 Neighborhood as per Number of Venue
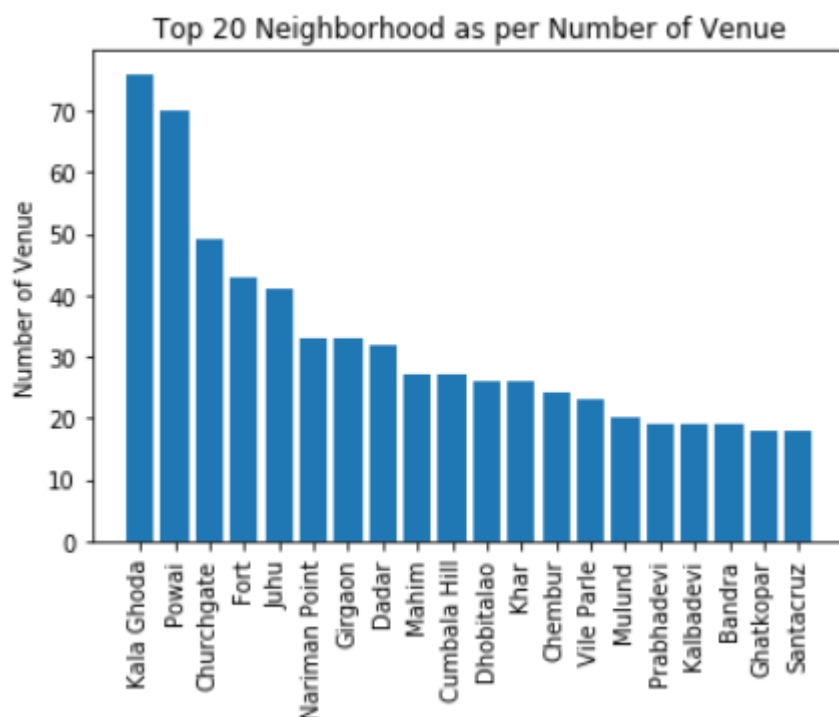
In summary of this graph 167 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each Neighborhoods in below table.

|	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	Agripada	Bakery	Indian Restaurant	Coffee Shop	Gym	Diner	Flower Shop	Flea Market	Field	Fast Food Restaurant	Electronics Store
1	Andheri	Indian Restaurant	Fast Food Restaurant	Bakery	Platform	Restaurant	Sandwich Place	Burger Joint	Food Truck	Food Court	Gym
2	Antop Hill	Indian Restaurant	Grocery Store	Train Station	Diner	Flower Shop	Flea Market	Field	Fast Food Restaurant	Electronics Store	Dumpling Restaurant
3	Bandra	Indian Restaurant	Café	Paper / Office Supplies Store	Italian Restaurant	Brewery	Breakfast Spot	Restaurant	Furniture / Home Store	Lake	Platform
4	Bhandup	Indian Restaurant	Bar	Donut Shop	Food	Flower Shop	Flea Market	Field	Fast Food Restaurant	Electronics Store	Dumpling Restaurant

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

First, I will run K-Means to cluster the boroughs into 5 clusters.

Here is my merged table with cluster labels for each Neighbourhoods.

	Neighborhood	Suburbs_Zone	Latitude	Longitude	Avarage House Price	grid	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 Com V
0	Andheri	Western Suburbs	19.119000	72.847000	18000	19	0.0	Indian Restaurant	Fast Food Restaurant	Bakery	Platform	Restaurant	Sandwich Place	Bi
1	Marol	Western Suburbs	19.117300	72.884000	16000	19	2.0	Snack Place	Indian Restaurant	Hotel	Gym	Diner	Hotel Pool	D
2	Bhayandar	Western Suburbs	19.290000	72.850000	7500	ot	3.0	Shipping Store	Zoo	Donut Shop	Food	Flower Shop	Flea Market	
3	Bandra	Western Suburbs	19.054444	72.840556	32000	23	2.0	Indian Restaurant	Café	Paper / Office Supplies Store	Italian Restaurant	Brewery	Breakfast Spot	Resta
4	Borivali (West)	Western Suburbs	19.238100	72.852300	16000	11	0.0	Park	Intersection	Pizza Place	Shop & Service	Soccer Field	Italian Restaurant	

We can also estimate the number of 1st Most Common Venue in each cluster. Thus, we can create a bar chart which may help us to find proper labels for each cluster.

Number of Venues in Each Cluster
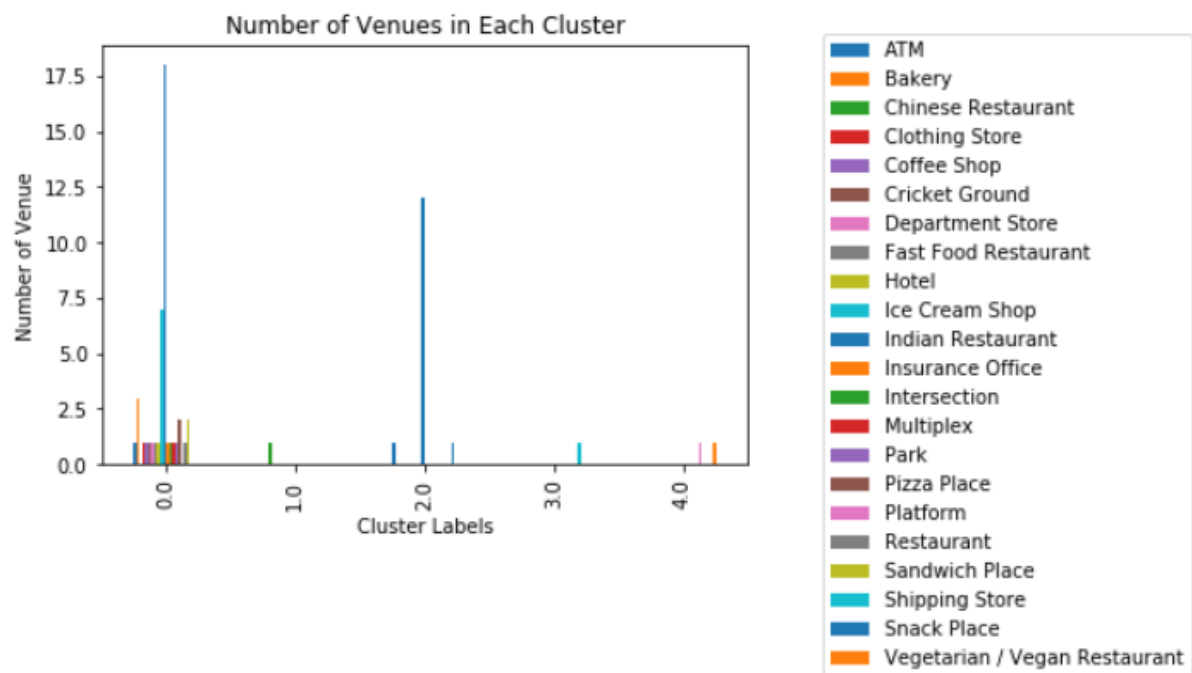
		Count
Cluster Labels	**1st Most Common Venue**	
0.0	ATM	1
	Bakery	3
	Clothing Store	1
	Coffee Shop	1
	Cricket Ground	1
	Department Store	1
	Fast Food Restaurant	1
	Hotel	1
	Ice Cream Shop	7
	Indian Restaurant	18
	Insurance Office	1
	Intersection	1
	Multiplex	1
	Park	1
	Pizza Place	2
	Restaurant	1
	Sandwich Place	2
1.0	Chinese Restaurant	1
2.0	ATM	1
	Indian Restaurant	12
	Snack Place	1
3.0	Shipping Store	1
4.0	Platform	1
	Vegetarian / Vegan Restaurant	1

When we examine above graph we can label each cluster as follows:

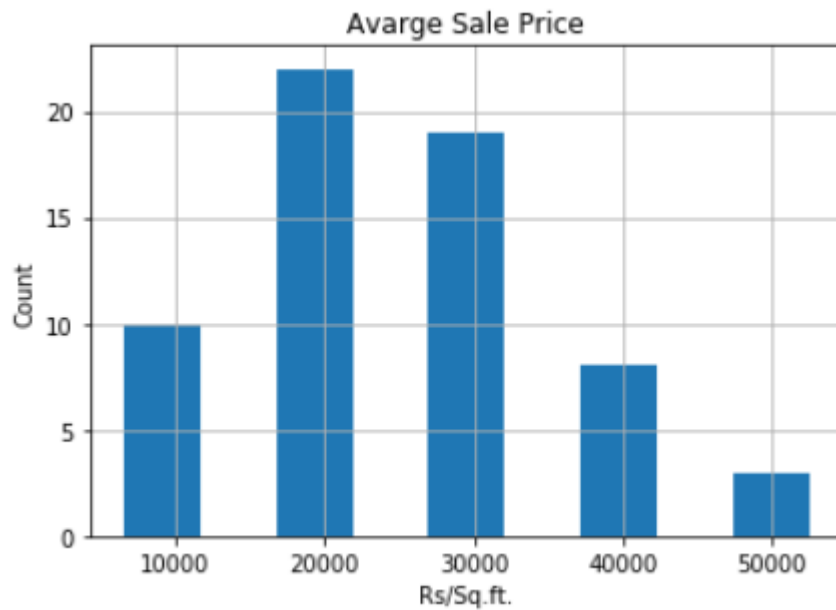Cluster 0 : "Multiple Social Venues"

Cluster 1 : "Chinese Restaurant Venues"

Cluster 2 : "Indian Restaurents Venues"

Cluster 3 : "Shipping Store Venues"

Cluster 4 : "Vegetarian Restaurant and Platform Venues"

We can also examine that what is the frequency of average housing sales prices in different ranges. Thus, histogram can help to visualization:

As it seems in above histogram, we can define the ranges as below:

15000 ASP : "Low Level HSP"

15000–25000 ASP : "Mid-1 Level HSP"
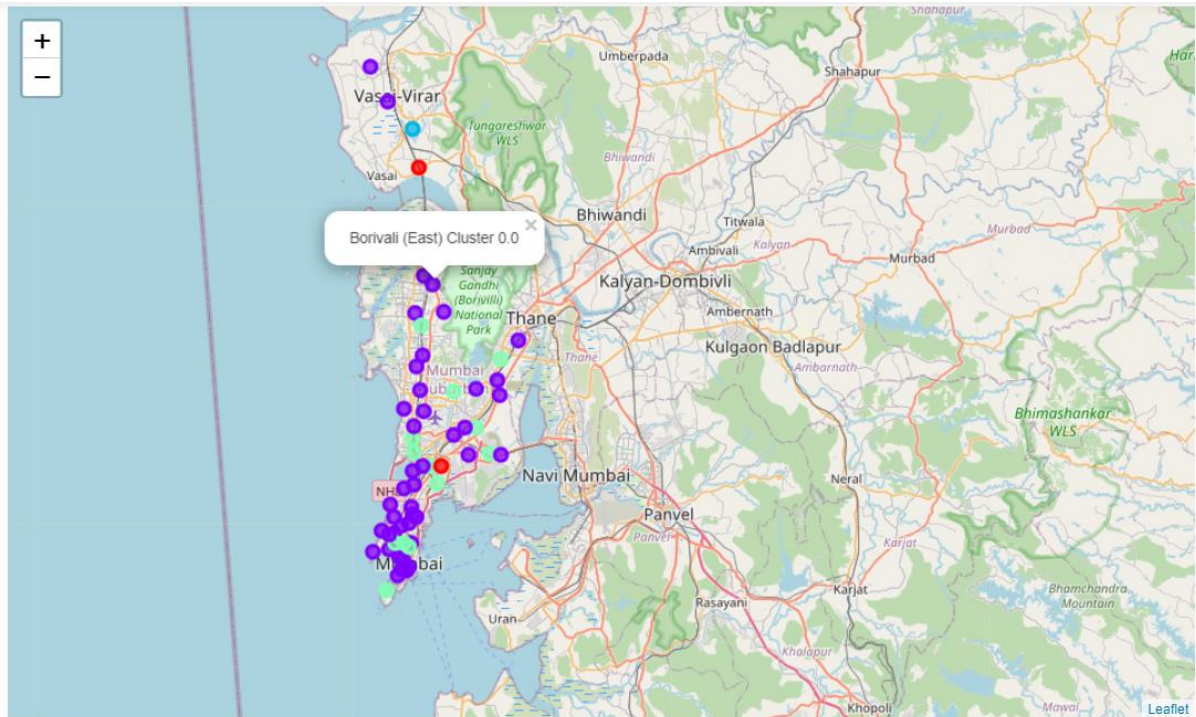
25000–35000 ASP : "Mid-2 Level HSP"

35000–45000 ASP : "High-1 Level HSP"

> 45000 AHP : "High-2 Level HSP"

4   Results

You can also see a clustered map Neighbourhoods of Mumbai in the below.

In summary section, one of my aim was also visualize the Average Housing Sale Prices for per square meter with choropleth style map. Thus, first I downloaded a json file of Mumbai Grid wise.

In final section, I created choropleth map which also has the below informations for each Neighbourhoods:
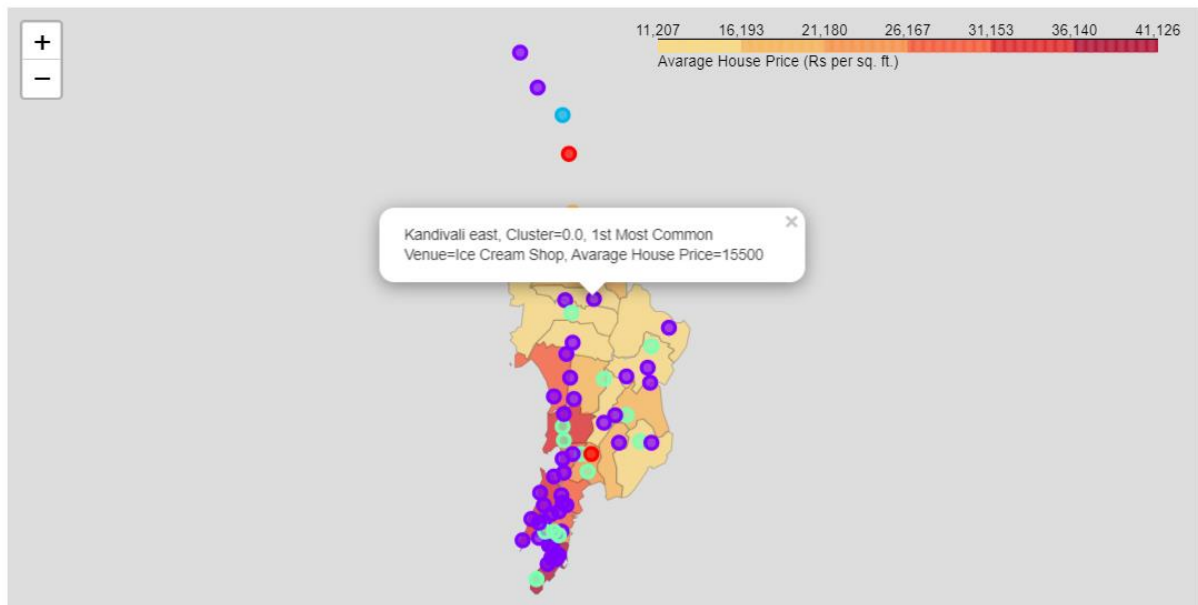
Neighbourhood name,

Cluster,

1st Most Common Venue,

Average Housing Price.

Kandivali east, Cluster=0.0, 1st Most Common Venue=Ice Cream Shop, Avarage House Price=15500

Avarage House Price (Rs per sq. ft.)

5 Discussion

As I mentioned before, Mumbai is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 24 Grid in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropol.

I used the Kmeans algorithm as part of this clustering study. I set the k value to 5. I also performed data analysis through this information by adding the coordinates of Neighbourhood with Grid and home sales price averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the Mumbai map. In future studies, web or telephone applications can be carried out to direct investors.

6 Conclusion

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

7 References

[1] https://en.wikipedia.org/wiki/Mumbai
[2] https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai
[3] https://developer.foursquare.com/

[4] https://www.99acres.com/property-rates-and-price-trends-in-mumbai

[5] https://www.google.co.in/maps/