**Housing Prices and Venues Data Analysis of Mumbai and Navi Mumbai**

# 1. **Introduction**:

## **Description and Discussion of the Background** : *Mumbai* is the financial and the commercial capital of India. The city hosts a bustling commercial life as it is home to the Hindi film industry, popularly known as Bollywood. The city is dotted with numerous pubs, bars and restaurants. It is also considered to be the central hub for industries, trade and business. Mumbai is also known for its high standard of living and pricy real estates. As the city presents good opportunities for business , it attracts people from all over the country. The contribution of Mumbai is 6.16% to the GDP of India. The Metro GDP ranking recognizes the city as one of the most productive metro areas of India. *Navi Mumbai* , located off the west coast of Maharashtra, is gradually emerging as a major commercial hotspot. It is a well-planned city with developed infrastructure and good connectivity to Mumbai and Pune. As the real estate prices are on a lower side as compared to Mumbai, many corporates and start ups have set up their offices here. There are some regions that are quite famous for affordable residences. The city boasts good social infrastructure with many flyovers, broad roads and parking lots. Many utility services, restaurants, banks,multiplexes,schools, and hospitals can be located here.

## **Business Problem** : The goal of this project is to cluster the different neighborhoods of Mumbai and Navi Mumbai by analyzing data related to category of venues . The project will also cover an analysis of the average real estate price of different neighborhoods. This analysis will benefit potential investors, businesspeople and the residents of the two cities. Investors and businesspeople can identity their preferred districts in terms of lower real estate cost and low presence of the business they want to install. City residents can identify locations for residence as well as plan to visit different neighborhoods based on the venue categories.

# 2**. Data Description**

## 2.1 **Description of the data** : The data for the latitude, longitude of different neighborhoods of Mumbai, Navi Mumbai are collected and their corresponding real estate price per square feet are collected. All these data are collected independently of each other as there are no websites that can serve as a good repository of the entire dataset. Foursquare API is used to collect data related to the different venues and venue categories for a neighborhood.

## 2.2 **Source of the data**: The data collection sources are as follows:

## 2.2.1 Latitude and Longitude: <https://www.latlong.net/>

### 2.2.2 Neighborhood Lists: Google Maps and Wikipedia

### 2.2.3 Average Real Estate Price per square ft : <https://www.makaan.com/>

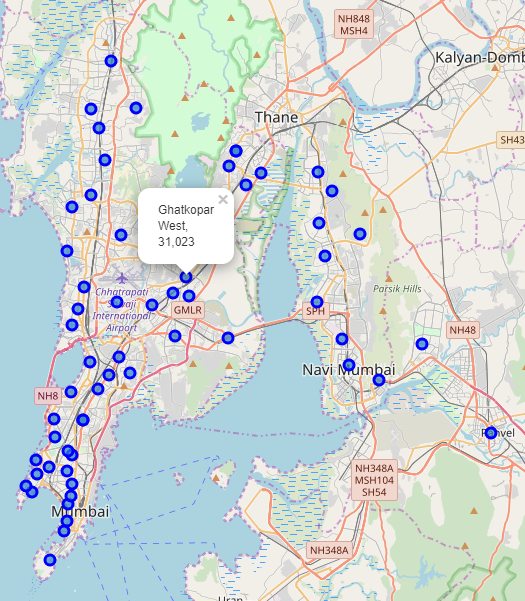
### 2.2.4 Venue Data : Foursquare API

## **2.3 The Data :** The following table depicts the first 10 rows of the data that will be used for analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | | Latitude | | Longitude | | Average price Rs/sqft |
| Colaba | 18.9067 | | 72.8147 | | 58,643 | |
| Nariman Point | 18.9256 | | 72.8242 | | 67,619 | |
| Churchgate | 18.9322 | | 72.8264 | | 59,969 | |
| Marine Lines | 18.9431 | | 72.8272 | | 32,463 | |
| Kalbadevi | 18.9487 | | 72.8289 | | 31,269 | |
| Walkeshwar | 18.9508 | | 72.8021 | | 1,18,331 | |
| Malabar Hill | 18.9548 | | 72.7985 | | 1,32,879 | |
| Bhuleshwar | 18.9562 | | 72.8296 | | 22,720 | |
| Kemps Corner | 18.9629 | | 72.8054 | | 75,444 | |
| Kamathipura | 18.9649 | | 72.8261 | | 34,507 | |

# 3. **Methodology**

After importing the data in a dataframe, the neighborhoods of Mumbai and Navi Mumbai have been visualized using the folium library and have used the pop up to display the neighborhood name and Average price in Rs/square ft . PFB the output of the same.



The Foursquare API is used to fetch the data regarding the venues and the category of venues in the respective neighborhoods in the two cities. The radius and limit are set to 500 and 100 respectively.

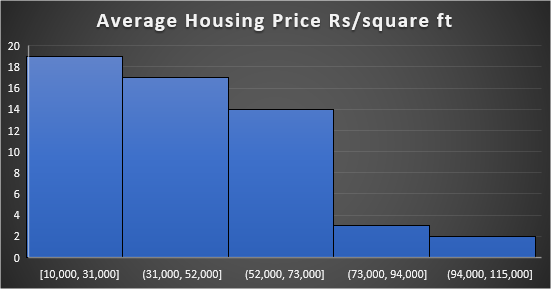
The venue and the venue categories when the API were called using the coordinates of Mumbai are as follows,



The venue and venue categories are queried for the entire data set of neighborhoods. A sample output with the venue categories arranged as 1st Most Common Venue to 10th Most Common Venue are as follows:



The real estate prices have been categorized into a histogram and they have been segmented five bins. Following is the histogram chart and the nomenclature for each of the bins.



Bin 1 [10,000-31,000] Low Level Housing Price

Bin 2 [31,000-52,000] Mid-Level 1 Housing Price

Bin 3 [52,000-73000] Mid-Level 2 Housing Price

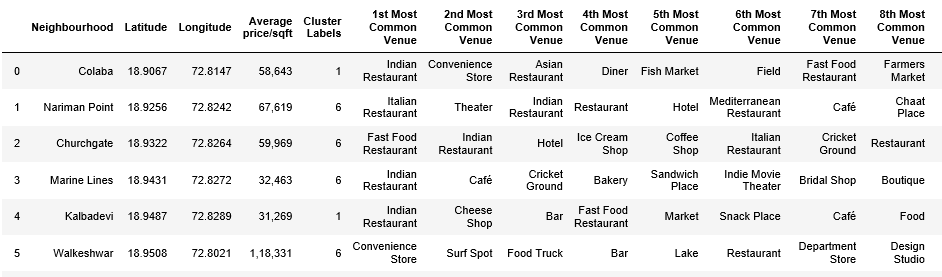
Bin 4 [73,000-94000] High-Level 1 Housing Price

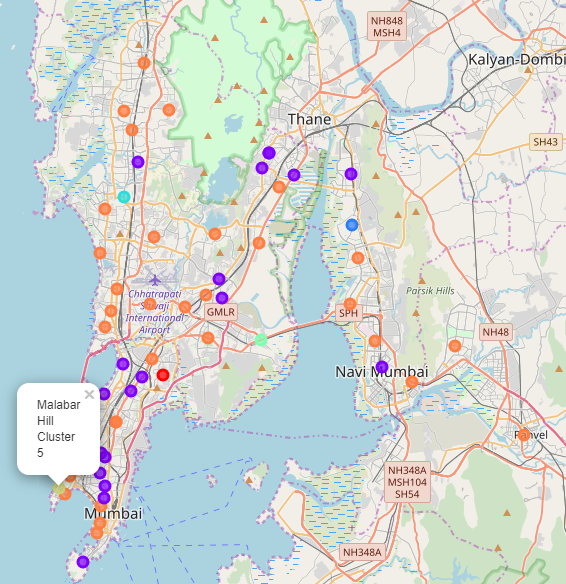
Bin 5 [94,000-115,000] High-Level 2 Housing Price

The major neighborhoods included in this project falls below High-Level 1 Housing Price.

The following bubble chart also shows the spread of the housing prices as per the geographical coordinates.

K means clustering is used to cluster the neighborhoods . The number of clusters used are 7 as this produces the optimum results. The results of the K-means clustering are also visualized using folium library. The different clusters are visualized with different colors.





# **4. Results**

The different clusters can be analyzed as follows:

**Cluster 1** can be named as ‘**Foodies Paradise’** as we can see different categories of restaurants in these neighborhoods. The most common being Indian Restaurant, Fast Food Restaurant, Sea Food restaurant and coffee shops. The place has good connectivity owing to the number of bus stations. The real estate prices are also in the affordable range as most are in Low Level, Mid-Level 1 and Mid-Level 2. These neighborhoods can also be quite good for residential purpose.

A sneap peak into **Foodies Paradise** is as follows

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | Housing Price Bin |
| Colaba | Indian Restaurant | Convenience Store | Asian Restaurant | Diner | Mid Level 2 |
| Kalbadevi | Indian Restaurant | Cheese Shop | Bar | Fast Food Restaurant | Mid Level 1 |
| Bhuleshwar | Indian Restaurant | Market | BBQ Joint | Ice Cream Shop | Low Level |
| Kamathipura | Indian Restaurant | Fast Food Restaurant | Department Store | Dessert Shop | Mid Level 1 |
| Byculla | Indian Restaurant | Hotel | Asian Restaurant | Restaurant | Mid Level 2 |
| Agripada | Restaurant | Indian Restaurant | Coffee Shop | Gym | Mid Level 1 |
| Worli | Seafood Restaurant | Ice Cream Shop | Indian Restaurant | Clothing Store | Mid Level 2 |
| Prabhadevi | Indian Restaurant | Electronics Store | CafÃ© | Chinese Restaurant | Mid Level 2 |
| Dadar | Indian Restaurant | Coffee Shop | CafÃ© | Hotel | Mid Level 1 |
| Matunga | Indian Restaurant | Ice Cream Shop | CafÃ© | Snack Place | Mid Level 1 |
| Nerul | Indian Restaurant | Diner | Fast Food Restaurant | Bakery | Low Level |
| Mahim | Indian Restaurant | Fast Food Restaurant | Chinese Restaurant | CafÃ© | Mid Level 2 |
| Ghatkopar East | Indian Restaurant | Park | Dessert Shop | Gym / Fitness Center | Mid Level 1 |
| Ghatkopar West | Indian Restaurant | Bakery | Juice Bar | Coffee Shop | Mid Level 1 |
| Mulund East | Restaurant | Indian Restaurant | Bus Station | Yoga Studio | Low Level |
| Airoli | Food Court | Platform | Indian Restaurant | Gym | Low Level |
| Bhandup West | Indian Restaurant | Fast Food Restaurant | Coffee Shop | Cosmetics Shop | Mid Level 2 |
| Goregaon | Indian Restaurant | Seafood Restaurant | Design Studio | Fast Food Restaurant | Low Level |
| Mulund West | Indian Restaurant | Coffee Shop | Bus Station | Pub | Mid Level 1 |

**Cluster 6** can be named as **Teens Hangout Place** owing to the variety of different venues present. These neighborhoods are dominant with Fast Food Restaurants, Pizza Palace, Caffe, Pubs , Lounge , Shopping malls etc. Most of the neighborhoods are a bit costly due to the high real estate prices. These are mostly in the range of Mid Level 2 and High Level. These neighborhoods are a bit luxurious and would be ideal for high end apartments .

The following table gives a brief overview of the Cluster

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | Housing Price Bin |
| Nariman Point | Italian Restaurant | Theater | Indian Restaurant | Restaurant | Mid Level 2 |
| Churchgate | Fast Food Restaurant | Indian Restaurant | Hotel | Ice Cream Shop | Mid Level 2 |
| Marine Lines | Indian Restaurant | CafÃ© | Cricket Ground | Bakery | Mid Level 1 |
| Walkeshwar | Convenience Store | Surf Spot | Food Truck | Bar | High Level 2 |
| Kemps Corner | CafÃ© | Dessert Shop | Bakery | Coffee Shop | High Level 1 |
| Tardeo | Pizza Place | Indian Restaurant | Snack Place | Coffee Shop | High Level 2 |
| Cumbala Hill | Sandwich Place | Bar | Other Great Outdoors | Park | High Level 1 |
| Mahalaxmi | Art Gallery | Lounge | Hotel | Theater | Mid Level 2 |
| Panvel | Shopping Mall | Theater | Multiplex | Yoga Studio | Mid Level 2 |
| Lower Parel | Indian Restaurant | Playground | Coffee Shop | Roof Deck | Mid Level 2 |
| Parel | Indian Restaurant | Playground | Coffee Shop | Roof Deck | Mid Level 1 |
| CBD Belapur | Plaza | Vegetarian / Vegan Restaurant | Fast Food Restaurant | Indian Restaurant | Low Level |
| Sion | Indian Restaurant | Gym / Fitness Center | Vegetarian / Vegan Restaurant | CafÃ© | Mid Level 1 |
| Kharghar | Pizza Place | Ice Cream Shop | Fast Food Restaurant | Outdoor Sculpture | Mid Level 2 |
| Juinagar | Concert Hall | Park | Market | Donut Shop | Low Level |
| Chembur | Bakery | Pizza Place | Fast Food Restaurant | CafÃ© | Mid Level 2 |
| Bandra | Indian Restaurant | CafÃ© | Chinese Restaurant | Gym / Fitness Center | High Level 1 |
| Khar | Indian Restaurant | Bar | Lounge | Italian Restaurant | Mid Level 2 |
| Kurla | Jewelry Store | Health & Beauty Service | Music Venue | Diner | Mid Level 1 |
| Santacruz East | Bakery | Pet Store | Ice Cream Shop | Burger Joint | Mid Level 2 |

**Cluster 5** can be also called as a ‘Tourist spot’ as the venues are mostly Hotel, Boarding House. The real estate price is also very high in these neighborhoods. Following is an example from Cluster 5.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Housing Price Bin |
| Malabar Hill | 5 | Hotel | Boarding House | Park | Dessert Shop | Donut Shop | High Level 2 |

**Cluster 0** can be also called as a ‘Historic spot’ as there are some places for sigh seeing and very less number of venues. It is also a ‘Low Level’ in terms of real estate price.

Following is an example of Cluster 0.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Housing Price Bin |
| Antop Hill | 0 | Pizza Place | Historic Site | Diner | Fish Market | Field | Low Level |

**Cluster 4** can be called as ‘Silent Neighborhood’ as there very less number of restaurants, hang out spots. The most common venue for these neighborhoods are Train Station. It is also a ‘Low Level’ in terms of real estate price.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Housing Price Bin |
| Mankhurd | 4 | Train Station | Sports Bar | Yoga Studio | Diner | Fish Market | Low Level |

**Cluster 2** can be called as ‘Residential Neighborhood’ as the venues prominent here are Department Store, Yoga Studio, Donut Shop, Fish market etc. It is also a ‘Low Level’ in terms of real estate price.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Housing Price Bin |
| Ghansoli | 2 | Department Store | Yoga Studio | Donut Shop | Flower Shop | Fish Market | Low Level |

**Cluster 3** can be called as ‘ Dating Spots’ as the venues present here are Icecream shop, Café, Bakery, Flower Shop etc. It is also a ‘Low Level’ in terms of real estate price

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Neighborhood | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Housing Price Bin |
| Jogeshwari | 3 | Ice Cream Shop | CafÃ© | Bakery | Food Court | Flower Shop | Low Level |

# 5. **Discussion**

The analysis can be beneficial for residents as well as budding entrepreneurs . Residents can choose areas where there are affordable housing prices combined with the easy access to their preferred venue categories. The businesspeople can set up their niche in neighborhoods where the presence of their business category is very sparse. They can explore each cluster and each neighborhood to draw out inferences suitable to their business requirements. The analysis is restricted to only 62 neighborhoods, however Mumbai and Navi Mumbai are much bigger landscapes, so this study can be expanded covering more neighborhoods which would give better and accurate results.

# 6. **Conclusion**

Mumbai and Navi Mumbai are cities where people from different parts of the country to start their careers or set up new business ventures. However, the lack of data and appropriate analysis often serves as a hindrance to them . I hope this project would be beneficial to all these people.