

Assignment 1: Report

Sajja Patel - 2021101107, Bhumika Joshi - 2022121006

August 2024

1 Data Preprocessing

In order to do any analysis, we first need to understand the data that we are working with. We also need to make sure that the data is available in a usable format, doesn't have inconsistencies, null values, or outliers that affect our analysis. We also need to identify useful features from the redundant and non-essential features in our dataset. All this can be done during the data preprocessing step.

In our case, we have dropped the following columns (feature selection) due to the given reasons:

- 'Covered Area.1', 'Carpet Area Unit': Repeated Columns
- 'Pantry Type', 'Land Area / Covered Area', 'Land Area Unit', 'Possession By': Almost Empty Columns
- 'Price (English)': Redundant Information (Numerical Already Present)
- 'Location': Information present as combination of 'Area Name' and 'City'
- After selecting only 'sq-ft' as unit of measurement, we have also dropped 'covArea Unit' and 'Unit of Carpet Area' since they don't give any new information.

Other changes made are:

- (Outlier detection) entries which are identified as outliers using inter-quartile range method (IQR) in terms of 'sqft Price' are removed. This has also been done for 'Price' and 'Carpet Area' wherever necessary.
- (Missing Value Imputation) All numerical missing values have been imputed by the mean while all the categorical missing values have been imputed by the mode of the respective column.
- (Redundant Data Detection) All the duplicate rows have been dropped from the dataset.

2 Question 1 & 2

For the first two questions, the goal is to divide the real estate opportunities into three distinct price ranges (Low, Medium, High Budget) and analyze how these budget categories are distributed across different cities. Price Categories:

Low Budget: Properties priced below Rs.15,058,859.70

- 4,812 properties

Medium Budget: Properties priced between Rs.15,058,859.70 and Rs.25,098,099.50

- 1,764 properties

High Budget: Properties priced above Rs.25,098,099.50

- 2,023 properties

Mumbai and Thane emerge as significant markets for investors across all budget ranges, particularly for Low and high Budget properties, indicating diverse opportunities for investors with varying budget constraints. Smaller cities might provide niche opportunities, but the number of available properties is relatively low.

This analysis highlights that investors seeking high end properties should focus on Mumbai and Thane, whereas low budget seekers may also find valuable options in other cities.

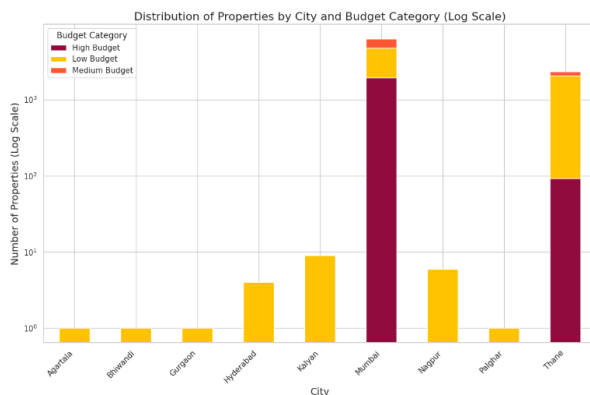


Figure 1: Distribution of Properties by City and Budget Category (Log Scale)

3 Question-3

The aim here is to compare properties in Mumbai and Thane using features like property types, sizes, and price.

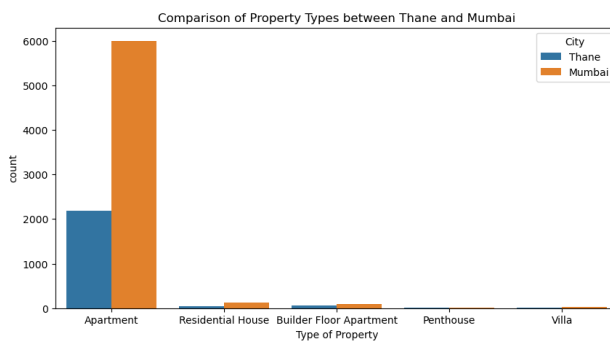


Figure 2: Comparison of Property Types between Thane and Mumbai

Here we can see that Most of the properties seem to be Apartments. Mumbai has more properties as compared to Thane.

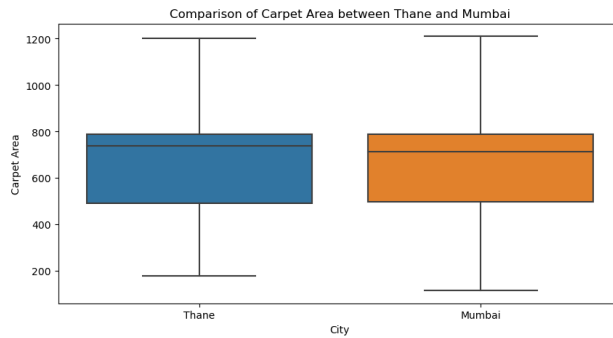


Figure 3: Comparison of Carpet Area between Thane and Mumbai

From the above, we can conclude that there isn't much difference in carpet area distribution between Mumbai and Thane. The distribution is slightly more left skewed in the case of Thane. That means that slightly higher proportion of properties have higher carpet area in Thane as compared to Mumbai.

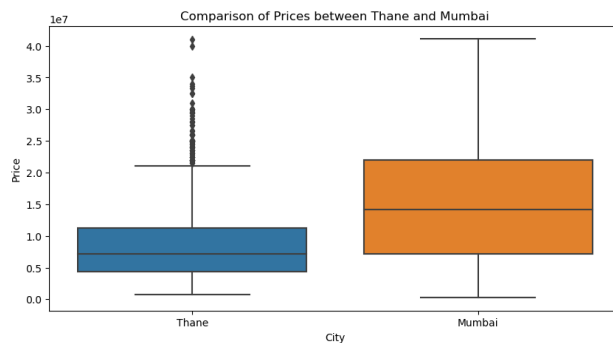


Figure 4: Comparison of Prices between Thane and Mumbai

From the above box-plot, we can see that Mumbai is costlier than Thane. Thane has a narrower range and a bunch of outliers. Thus, if an investor is looking for cheaper properties to invest in, they may consider Thane as their preferred city and if an investor is more interested in expensive real estate, they may narrow down and consider properties in Mumbai.

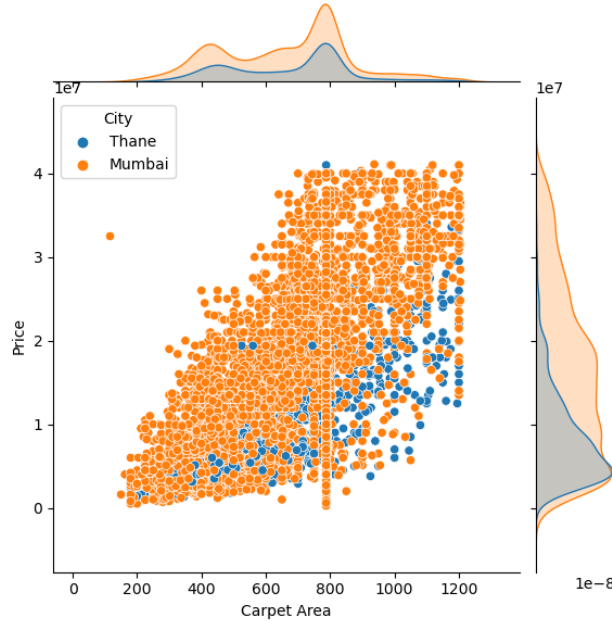


Figure 5: Carpet Area vs Price: Thane vs Mumbai

The Price to Carpet Area Ratio seems to be lesser in Thane, meaning we can possibly get larger properties for cheaper in Thane as compared to Mumbai. Carpet area frequency distribution seems to be similar in both cities (though smaller in Thane because of less data points). Price frequency distribution peaks at the same spot for both cities but seems to decline sharply in the case of Thane. Mumbai has more properties in the higher price range.

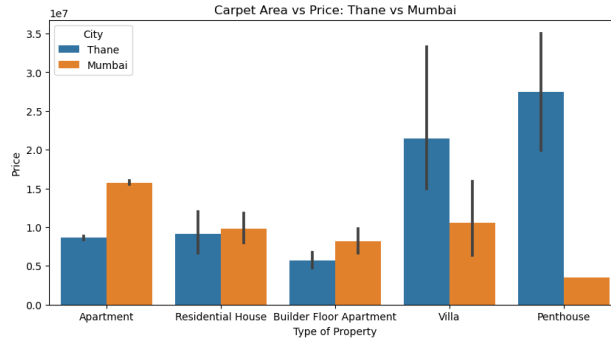


Figure 6: Carpet Area vs Price: Thane vs Mumbai

Avg. Price is the least in the case of Builder Floor Apartments in both cities. Residential Houses and Apartments are also on the cheaper side in both cities, though apartments seem to be significantly costlier in Mumbai than in Thane. Villas and Penthouses are expensive, but they seem to be significantly higher priced in Thane as compared to Mumbai, contrary to the other property types which are cheaper in Thane.

4 Question-4

Here we seek to find more and less expensive properties in each city. We also look at aspects like prime location, etc. First, we look at expensive properties. We classify a property as expensive if it is above the

75th percentile in the price category.

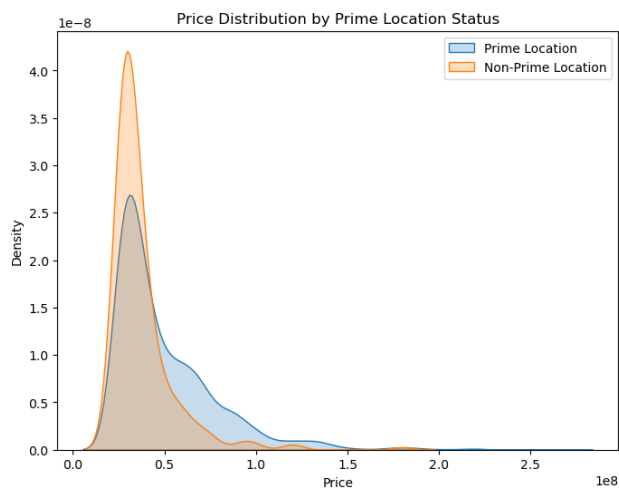


Figure 7: Price Distribution by Prime Location Status (High Budget)

In the above, we can see that most High-Budget properties have a similar price range regardless of whether they are in prime location or not. However at the higher price range among these, most are located in a prime location.

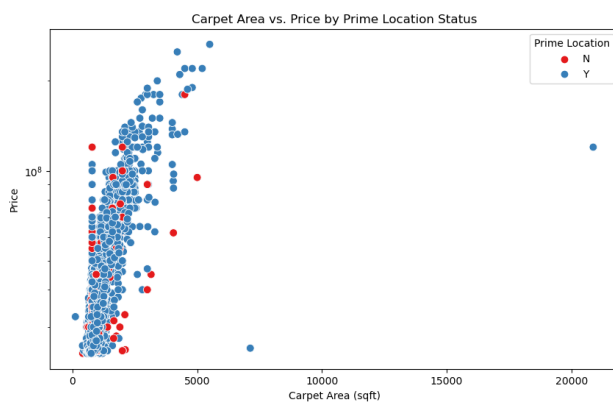


Figure 8: Carpet Area vs. Price by Prime Location Status (High Budget)

Higher price seems here to correlate with higher carpet area in the case of prime location properties. This does not seem to happen as distinctly in the case of non-prime locations.

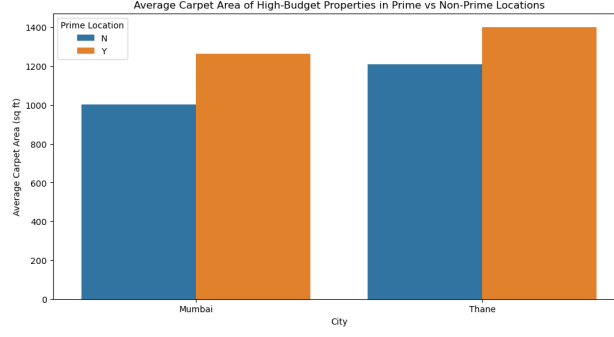


Figure 9: Average Carpet Area of High-Budget Properties in Prime vs Non-Prime Locations

In the above bar plot, we see that properties at prime locations seem to have higher carpet area on average as compared to ones located at non-prime locations. Next, we look at less expensive properties. A property has been classified as non-expensive if its price falls under the 25th percentile.

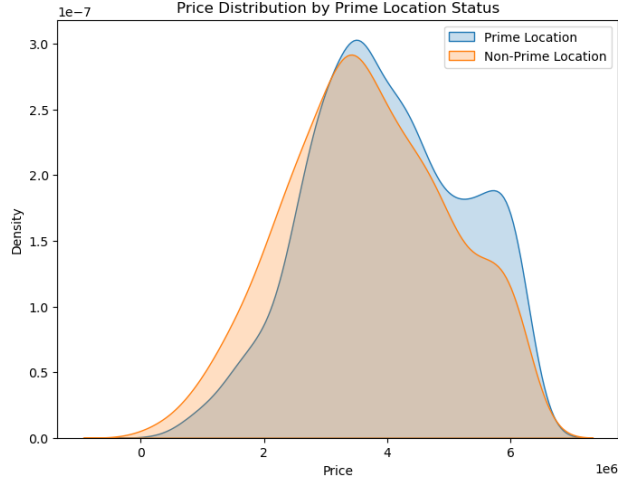


Figure 10: Price Distribution by Prime Location Status (Low Budget)

In the above, we see that price distribution is similar in the case of both prime location and non-prime location in the case of low budget properties. Again there is a small spike at the higher end of the price range in the case of prime location properties.

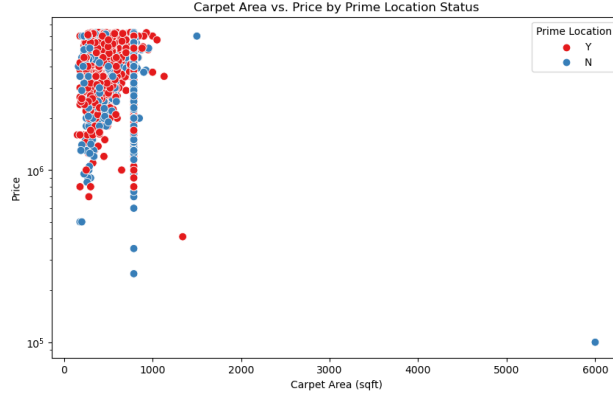


Figure 11: Carpet Area vs. Price by Prime Location Status (Low Budget)

Here, we observe that while there is some amount of linearity in the correlation between price and carpet area, most properties at the higher end of the price range span across the range of carpet area. Smaller carpet area properties still tend to be cheaper.

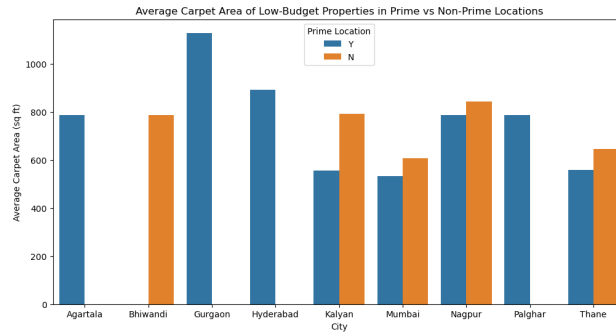


Figure 12: Average Carpet Area of Low-Budget Properties in Prime vs Non-Prime Locations

In the above plot, we can see that unlike High-Budget Properties that were available only in Thane and Mumbai, Low-Budget Properties are also available in some other cities. Bhiwandi only has non-prime location properties, while Agartala, Gurgaon, Hyderabad and Palghar have only prime location properties available.

5 Question-5

In this analysis, we delve into price per square foot as a vital indicator for investors looking to maximize their real estate investment by securing larger properties at lower costs. By analyzing the average price per unit area across different cities, we provide valuable insights for investors aiming to identify cities where real estate prices are more affordable, making it easier to acquire larger spaces at lower rates.

For example, cities with the lowest price per square foot are often appealing to investors with long-term strategies focused on capitalizing on future growth potential. These cities might offer significant opportunities for cost-effective investments in larger properties, which could appreciate over time.

Additionally, the influence of developers plays a significant role in determining price per square foot. Well-known developers, such as Brigade Constructions, might offer properties at competitive rates due to their reputation and economies of scale. This is critical for investors who prioritize working with established developers but still seek lower prices.

Thus, the visualizations presented not only highlight cities with lower price per square foot but also showcase

the top 10 developers with the most affordable rates, offering a comprehensive view of where investors might find the best deals based on location and developer reputation. This dual perspective helps investors align their investment strategies with their budgetary constraints and developer preferences.

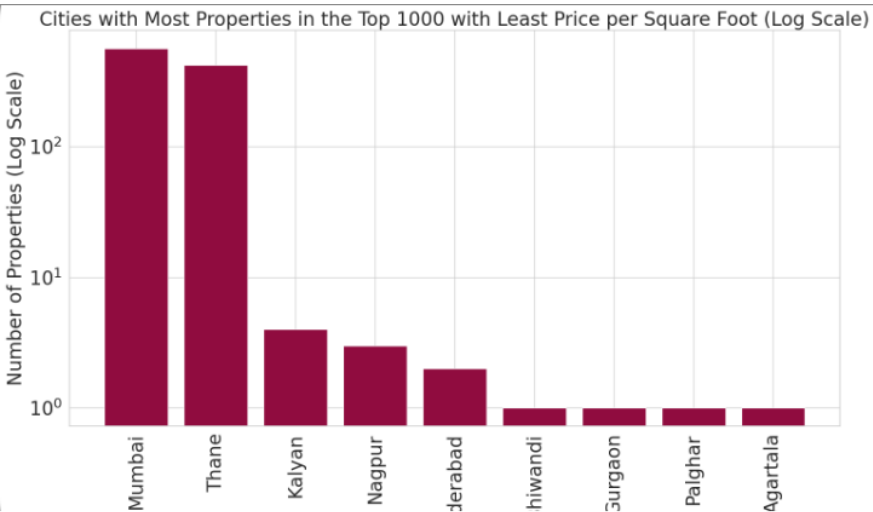


Figure 13: Cities with Most Properties in the Top 1000 with Least Price per Square Foot

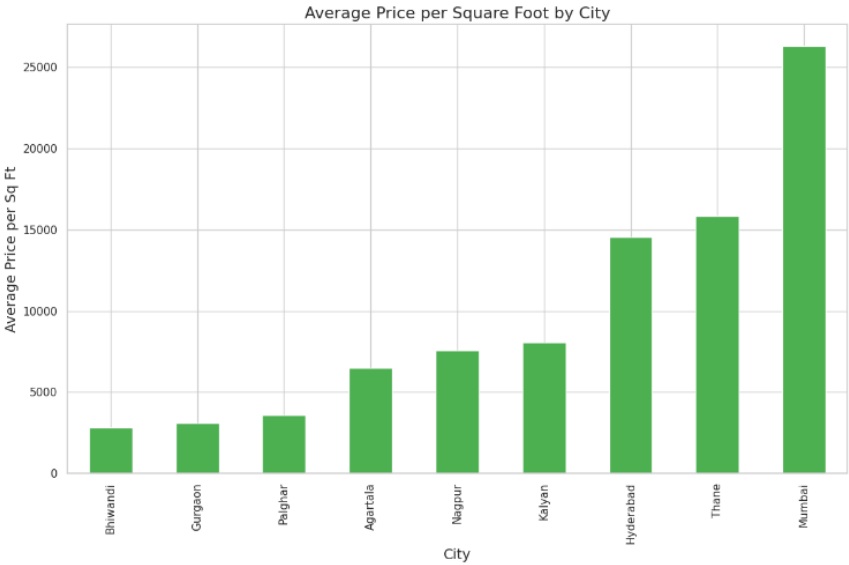


Figure 14: Average Price per Square Foot by City

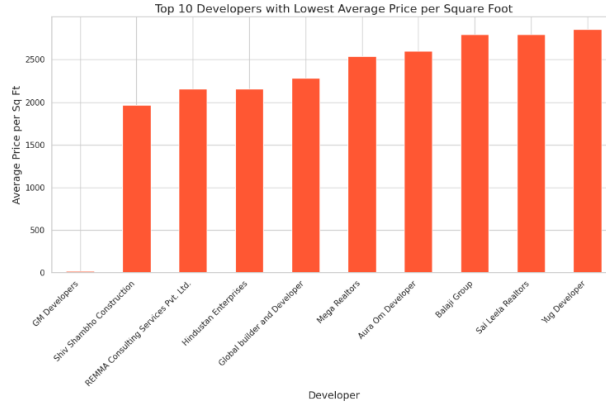


Figure 15: Top 10 Developers with Lowest Average Price per Square Foot'

6 Question-6

Here we seek to find commercial hotspots and compare their prices, area, amenities offered, etc. to find the ideal locations for commercial properties. First, we look at the aspect of price and carpet area in the case of commercial properties.

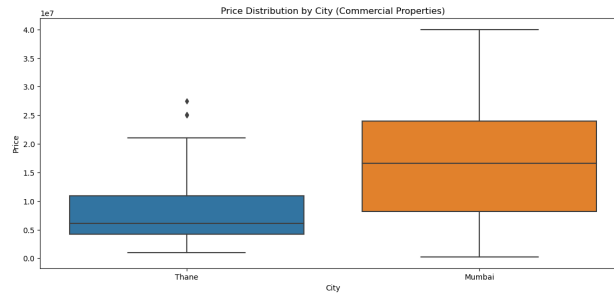


Figure 16: Price Distribution by City (Commercial Properties)

Commercial properties are significantly less expensive in Thane as compared to Mumbai.

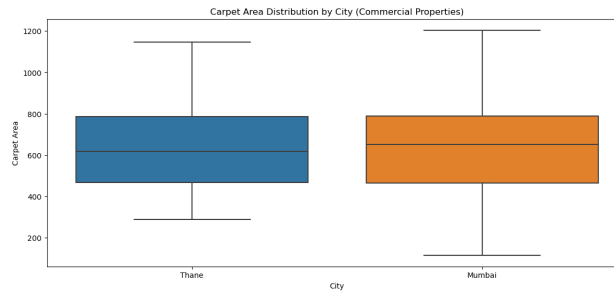


Figure 17: Carpet Area Distribution by City (Commercial Properties)

Carpet area remains similar across commercial properties in both cities.

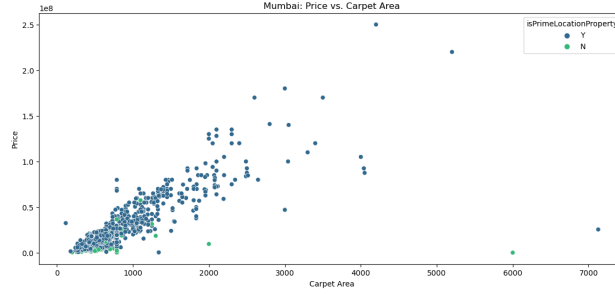


Figure 18: Mumbai: Price vs. Carpet Area (Commercial Properties)

Properties in non-prime location seem to be cheaper and have less carpet area as compared to prime location properties in Mumbai. Prices seem to increase linearly with carpet area.

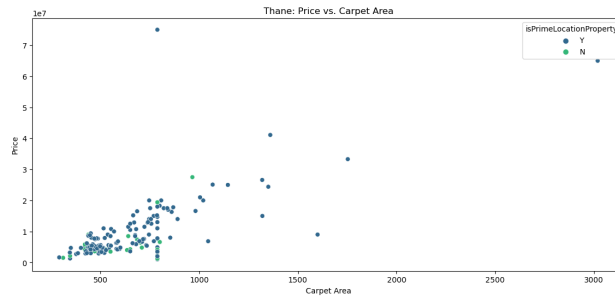


Figure 19: Thane: Price vs. Carpet Area (Commercial Properties)

Prices generally increase with carpet area. There are more non-prime location properties in the smaller and cheaper side, however they are scattered throughout. They tend to be less expensive than prime location properties for the same carpet area.

Next, we analyze commercial properties based on the amount of amenities they provide.

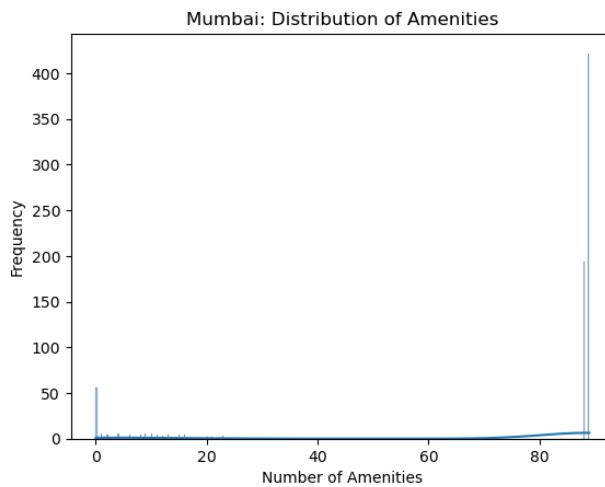


Figure 20: Mumbai: Distribution of Amenities (Commercial Properties)

Most properties seem to have most of the available amenities. However a small amount of properties

seem to have significantly lesser amount of listed amenities in Mumbai.

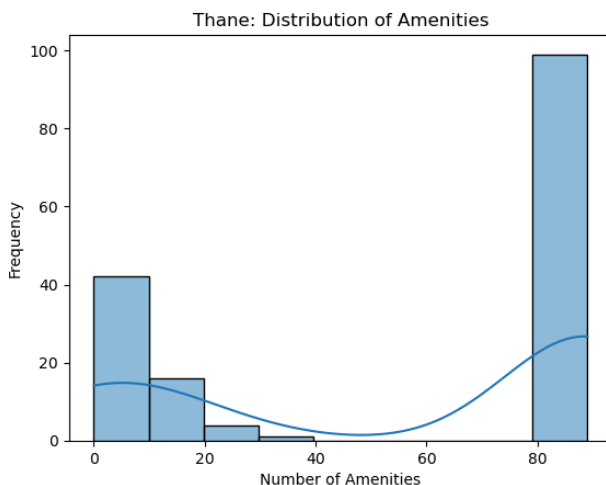


Figure 21: Thane: Distribution of Amenities (Commercial Properties)

There are some properties with little to none of the listed amenities. The amount of properties declines as we increase no. of amenities until we reach the highest bracket where majority of the properties are located in Thane.

Moving on, we analyze the locations of these commercial properties. Based on plotting the data using bar plots, we have found that in Mumbai, Andheri West seems to have the most commercial properties, followed by Kandivali West and Andheri East. Most of the properties in Andheri West are in prime location. It is also the location with the highest no. of properties in prime locations. Areas like 4 Bungalows, Dattaguru Nagar, Aarey Milk Colony have all the listed amenities available. Andheri west, on an average, has around 60 of the listed amenities available.

On the other hand, in Thane, Dombivli East has the maximum commercial properties followed by Dombivli West and Thane West. Dombivli East again has the highest number of properties in the prime location and most of its properties are in the prime location. Dombivli East, Dombivli West and Thane West have around 50-60 of the listed amenities available on an average. Areas like Bhoiwada, Kolshet, Owale have all the listed amenities available on an average.

7 Question-7

In this analysis, we focus on how the inclusion of specific amenities, such as swimming pools, gymnasiums, and clubhouses, impacts property prices in Mumbai and Thane. Investors are particularly interested in understanding whether these amenities add value to properties and whether there are any notable differences in amenity preferences between the two cities.

Key Findings:

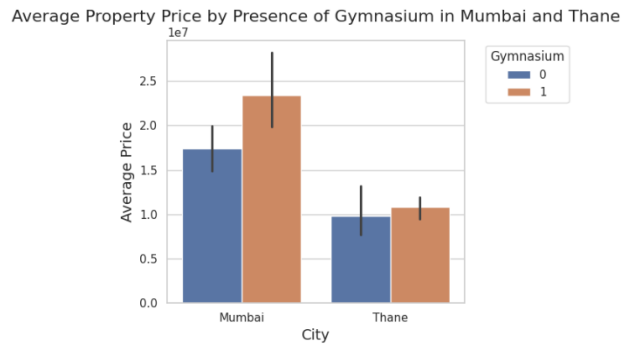
- **Swimming Pool:**

- *Mumbai*: Properties with swimming pools command higher prices compared to those without. This suggests that having a swimming pool is a valuable amenity in Mumbai's real estate market.
- *Thane*: Similarly, properties with swimming pools in Thane are priced higher than those without, indicating that this amenity adds significant value in Thane as well.



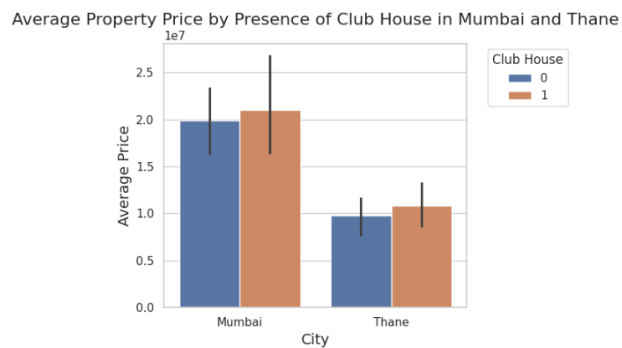
• Gymnasium:

- *Mumbai*: Properties that feature a gymnasium have higher prices than those without, indicating that access to a gym is a desirable feature for buyers in Mumbai.
- *Thane*: The trend is consistent in Thane, where properties with a gymnasium are valued higher than those without. This amenity appears to be a significant factor in both cities.



• Club House:

- *Mumbai*: The inclusion of a clubhouse also correlates with higher property prices, further solidifying the importance of this amenity in Mumbai.
- *Thane*: In Thane, properties with a clubhouse feature also tend to be more expensive, reflecting its value in the local market.



Across both cities, the presence of amenities like swimming pools, gymnasiums, and clubhouses consistently adds value to properties. Although there may be slight variations in the degree of impact, the overall trend

shows that buyers in both Mumbai and Thane are willing to pay a premium for these features. This indicates that these amenities are universally appealing and contribute to higher property valuations, making them critical considerations for investors targeting these markets.



Figure 22: Price Impact of Amenities in Mumbai and Thane

8 Question-8

Here, we wish to find out how possession status and availability affect the prices of properties.



Figure 23: Mumbai: Average Price by Possession Status

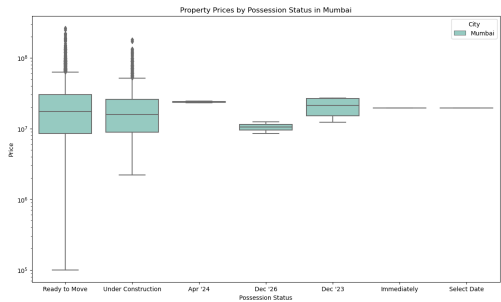


Figure 24: Property Prices by Possession Status in Mumbai

Ready-to-Move properties have the highest average price. Properties available by Apr'24 are also expensive.

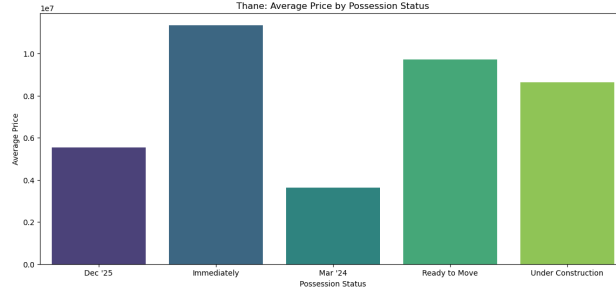


Figure 25: Thane: Average Price by Possession Status

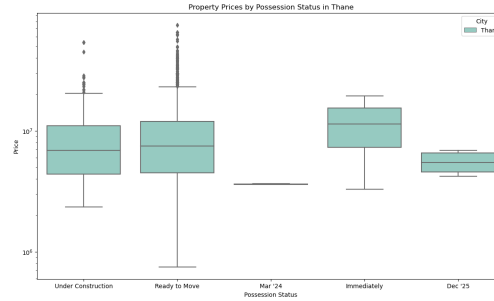


Figure 26: Property Prices by Possession Status in Thane

Immediately available and ready to move properties seem to be expensive in Thane.

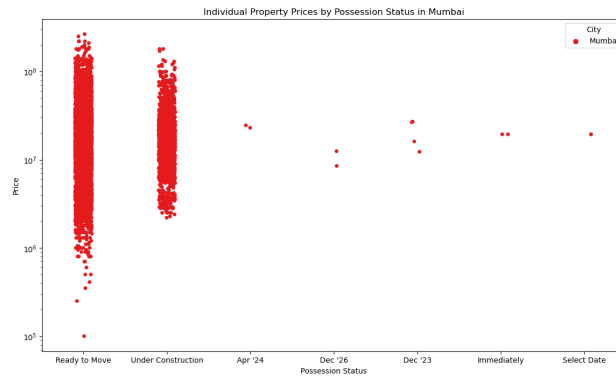


Figure 27: Individual Property Prices by Possession Status in Mumbai

From the above plot, we can see that in Mumbai most of the properties are ready to move in or under construction. These also have the highest prices. Also, properties available from Oct '18 are the cheapest while those available from Oct '27 are the costliest.

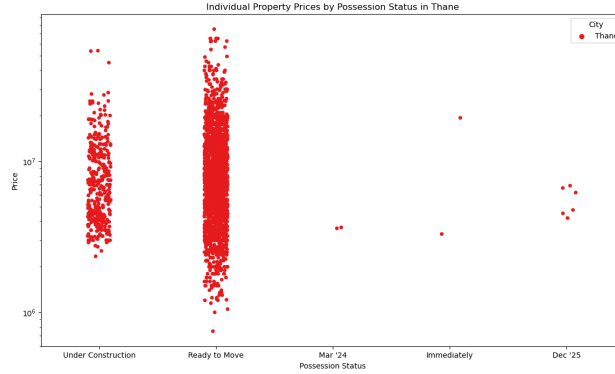


Figure 28: Individual Property Prices by Possession Status in Thane

In Thane, similar to Mumbai, most of the properties are ready to move in or under construction. These also have the highest prices. Over here, properties available from Oct '17 are the cheapest while those available from Jul '23 are the costliest.

9 Question-9

In this analysis, we aim to evaluate the impact of developers on property prices and amenities. Specifically, we explore whether certain developers are associated with higher-end properties and better amenities such as swimming pools, clubhouses, and gymnasiums. By analyzing the top 10 developers in terms of cost per square foot, we can determine which developers cater to high-end markets and which provide more affordable options.

Key Findings:

- **Top 10 Cheapest Developers (by Price per Square Foot):**

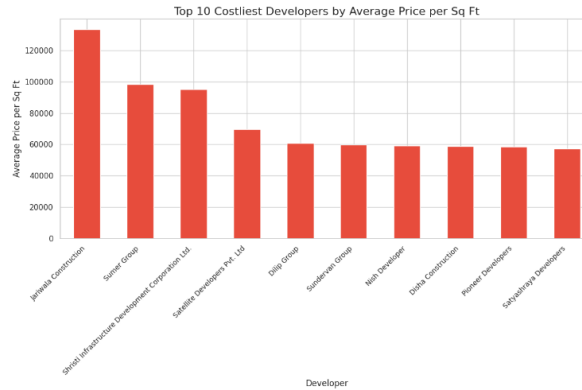
- These developers generally offer properties at more affordable rates per square foot, making them attractive to budget-conscious buyers.
- The presence of amenities such as swimming pools, gymnasiums, and clubhouses is less common among these developers, indicating that these properties may prioritize cost savings over luxury features.



- **Top 10 Costliest Developers (by Price per Square Foot):**

- Developers in this category are associated with high-end properties that command premium prices per square foot.

- These properties are more likely to feature luxurious amenities such as swimming pools, gymnasiums, and clubhouses, which add significant value to the properties.
- These developers often cater to buyers seeking luxury living and are willing to pay a premium for enhanced features and amenities.



Amenity Analysis:

- **Swimming Pool:**

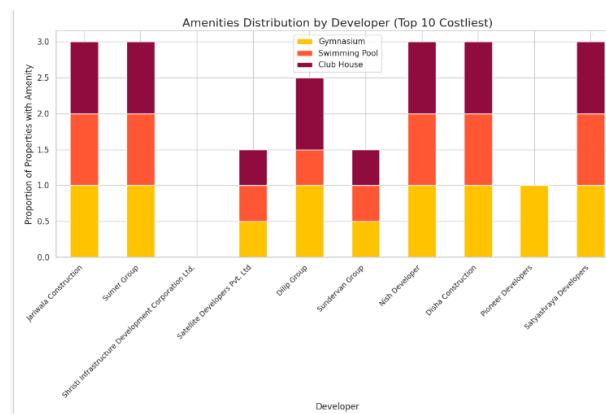
- Properties developed by costlier developers are more likely to include swimming pools, emphasizing their focus on luxury and lifestyle features.

- **Gymnasium:**

- Gymnasium facilities are commonly found in properties developed by higher-end developers, indicating a focus on providing comprehensive fitness amenities for residents.

- **Club House:**

- Clubhouses are another feature predominantly found in properties developed by the costliest developers, adding to the exclusivity and social aspects of these residential complexes.



The analysis reveals a clear distinction between affordable and premium developers. Higher-end developers consistently offer properties with more luxurious amenities, reflecting their target market's preference for upscale living environments. On the other hand, cheaper developers tend to focus on providing cost-effective properties, often without the inclusion of high-end amenities. This differentiation offers investors a clear choice based on their budget and lifestyle preferences, whether they seek affordable housing or premium properties with extensive amenities.