

# Real Estate Data Analysis Report

B V G Poojitha,Aryan Bansal

## 1 Introduction

This report presents a comprehensive analysis of a real estate dataset, focusing on data preprocessing, exploratory insights, and city-level comparisons. The goal was to uncover patterns in pricing, amenities, developer influence, and spatial efficiency across key Indian cities, with particular emphasis on Mumbai and Thane.

## 2 Data Preprocessing

The preprocessing phase involved rigorous schema standardization, where column names were normalized and redundant fields removed. Price values were converted to lakhs for interpretability, and engineered features such as price per square foot, amenity count, and luxury flags were introduced. Amenities were grouped into logical categories—like kids amenities, water facilities, and food and beverage options—to simplify analysis. Missing values were identified and addressed, and outliers were flagged using z-score thresholds on price, carpet area, and price per square foot. The cleaned dataset was exported for downstream analysis.

## 3 Price Segmentation and Market Overview

Properties were segmented into affordable (0–50L), mid-range (50–100L), and luxury (100L+) bands. Visualizations revealed that Mumbai dominates the luxury segment, while Thane offers a more balanced mix.



Figure 1: Price Segmentation Across Cities

Property type distribution showed that Mumbai features more premium formats like villas and penthouses, whereas Thane is concentrated in apartments and residential houses.



Figure 2: Property Type Distribution

Amenity count analysis indicated that luxury listings tend to offer more features, though some mid-range properties rival them in richness. Violin plots revealed that luxury does not always equate to larger carpet areas—some affordable and mid-range listings offer comparable space.

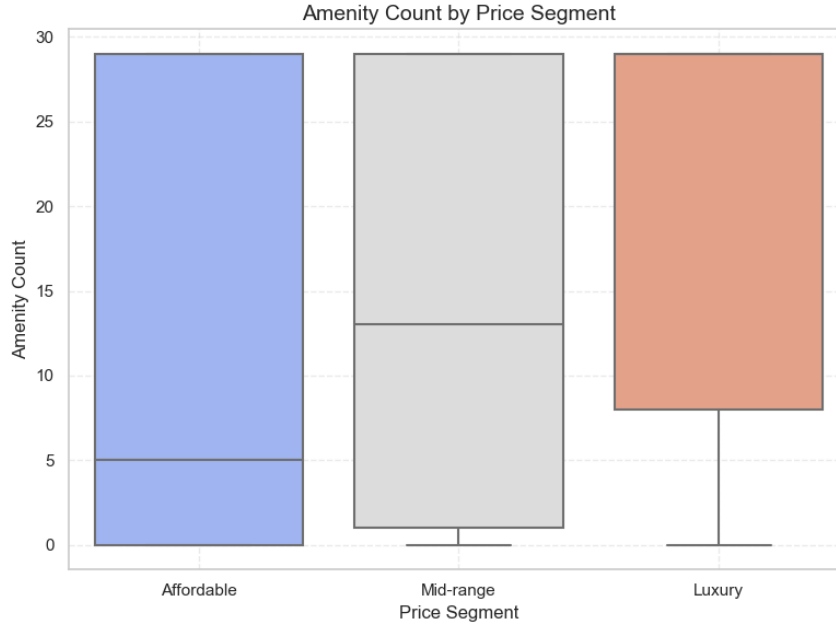


Figure 3: Amenity Count by Price Segment

## 4 Mumbai vs Thane Comparison

Thane properties generally offer larger carpet areas and better spatial efficiency, while Mumbai listings are more compact and expensive. Residential and commercial segments were both active, though Mumbai had a slightly higher share of commercial listings. Thane's commercial market is emerging, offering opportunities for expansion.

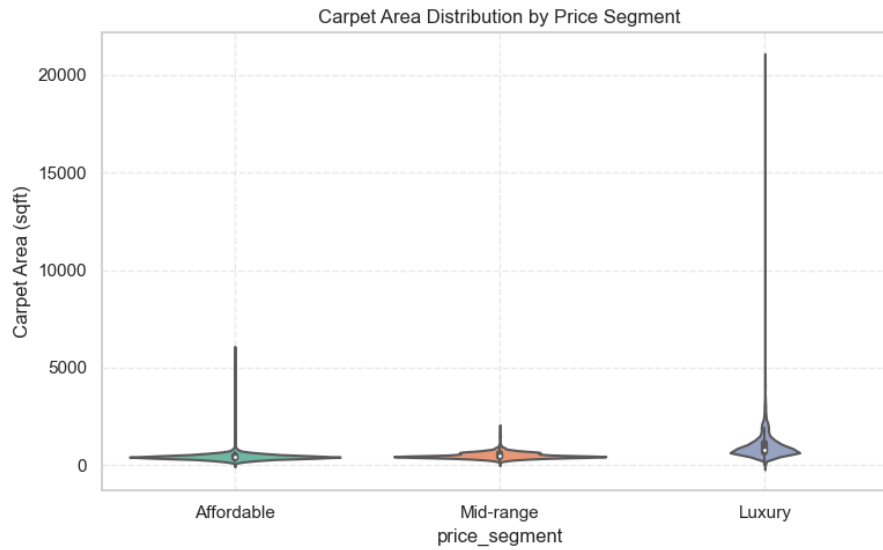


Figure 4: Carpet Area Distribution by City

Bar charts showed that Thane delivers nearly double the carpet area per lakh spent compared to Mumbai, making it attractive for budget-conscious buyers.

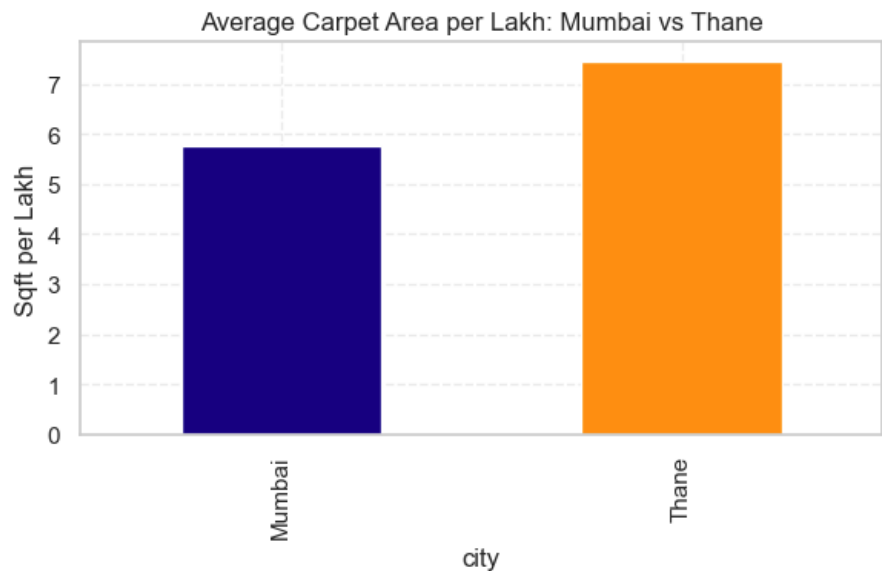


Figure 5: Carpet Area per Lakh Comparison

## 5 Location Premium and Value Efficiency

Prime-location properties command significantly higher price per square foot and offer more amenities, though they don't always provide more space. In Mumbai, the uplift is sharper, while in Thane, non-prime listings offer better value for money.

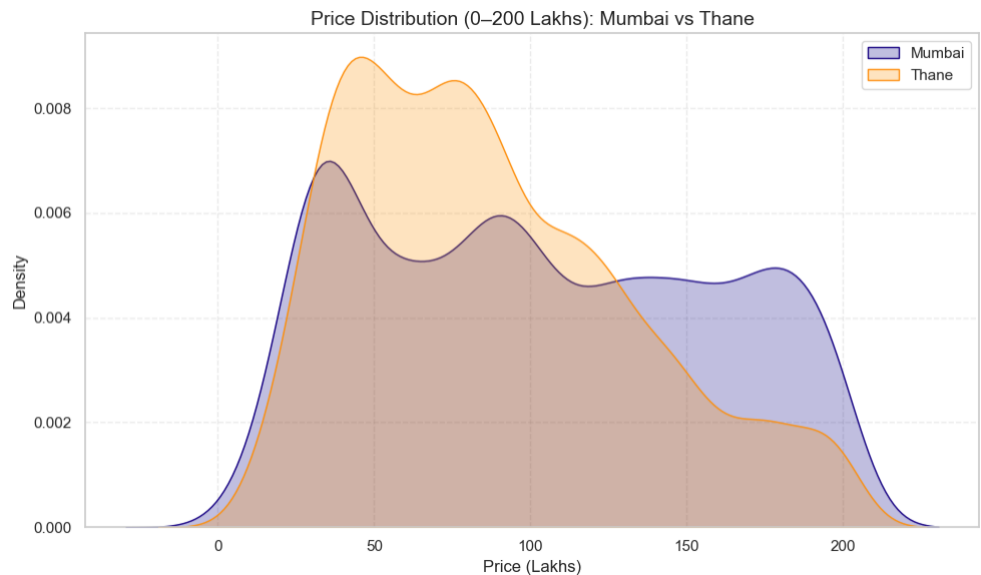


Figure 6: Price per Sqft – Prime vs Non-Prime

Cities like Gurgaon, Nagpur, and Kalyan were identified as top performers in carpet area per lakh spent. Builder floor apartments and residential houses offered the best spatial efficiency, outperforming high-rise formats.

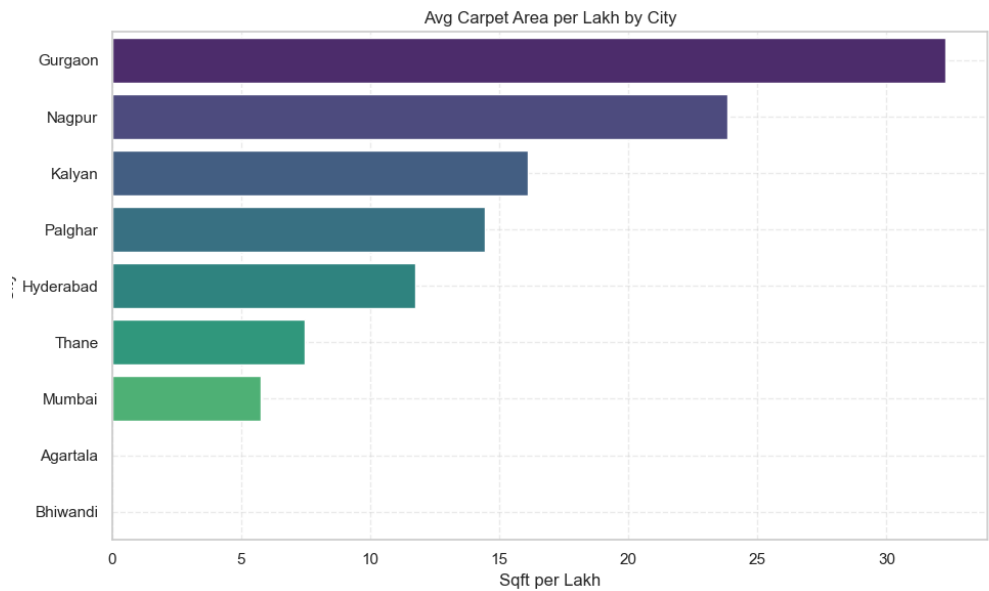


Figure 7: Carpet Area per Lakh by City

## 6 Amenity and Developer Impact

Swimming pools and gymnasiums significantly influence pricing, especially in Mumbai. Clubhouse type also added value, with premium clubhouses correlating with higher prices. Listings with all three amenities—pool, gym, and premium clubhouse—commanded the highest prices.

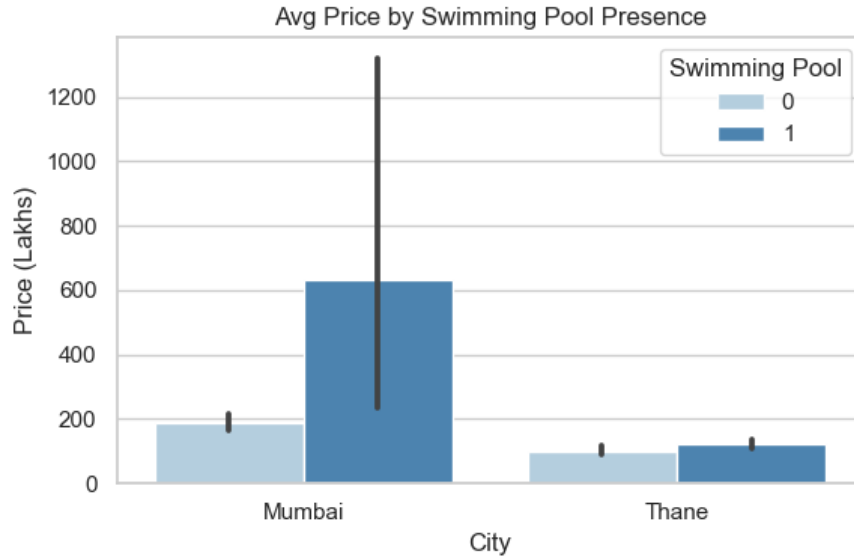


Figure 8: Swimming Pool Impact on Price

Ready-to-move properties are priced higher than under-construction listings, with the price gap more pronounced in Mumbai. Listings with near-term possession dates showed a premium, while longer timelines correlated with lower prices. Developer analysis showed that brands like Rustomjee and Lodha dominate Mumbai's luxury segment, commanding the highest average prices. In Thane, developers such as Kalpataru and Godrej offer more affordable pricing with solid amenity offerings.



Figure 9: Average Price by Developer

## 7 Conclusion

This analysis provides a data-driven guide for investors. Mumbai offers premium branding and lifestyle features, ideal for luxury-focused buyers. Thane delivers better spatial efficiency

and affordability, making it attractive for mid-range and value-conscious investors. The insights on location, amenities, possession timelines, and developer influence equip stakeholders to make informed decisions in a dynamic real estate market.