Nikhil Sharma

Real Estate Data Analysis Report

17 July 2023

1. **Introduction**

   This report presents findings from the analysis of the "Real Estate Data" dataset. The data is about housing real estate from Indian cities and has been sourced from Rakkesh Aravind G on Kaggle. This dataset is web scraped from a real estate website, collecting all the necessary information on resale and new properties. It contains around 14,000+ rows of data having properties from various Indian cities like Chennai, Mumbai, Bangalore, Delhi, Pune, Kolkata, and Hyderabad.

   The dataset includes the following columns:

- Name: Property Name
- Property Title: Property Ad Title
- Price: Property Price
- Location: Property Located Locality and Region
- Total Area: Total SQFT of the property
- Price Per SQFT: Price of Per SQFT of the property
- Description: Small paragraph about the property
- Baths: Number of baths in the property
- Balcony: Whether the Property has a balcony or not

The analysis aims to understand real estate trends in these cities, providing insights for researchers, data analysts, and real estate enthusiasts.

2. **Data Analysis**
   The analysis used Python, including the NumPy, Pandas, Matplotlib, and Seaborn libraries. The Jupyter Notebook code was used for the analysis, which explored and understood the dataset. The key findings are as follows:

2.1 **Data Shape and Columns** The dataset consists of 14,528 rows and 9 columns. The columns are: 'Name', 'Property Title', 'Price', 'Location', 'Total_Area', 'Price_per_SQFT', 'Description', 'Baths', and 'Balcony.'

2.2 **Data Types and Cleaning** The data types of the columns were checked and cleaned to align with their respective attributes. The "Price" column was cleaned from a string representation to a numeric one for further analysis.

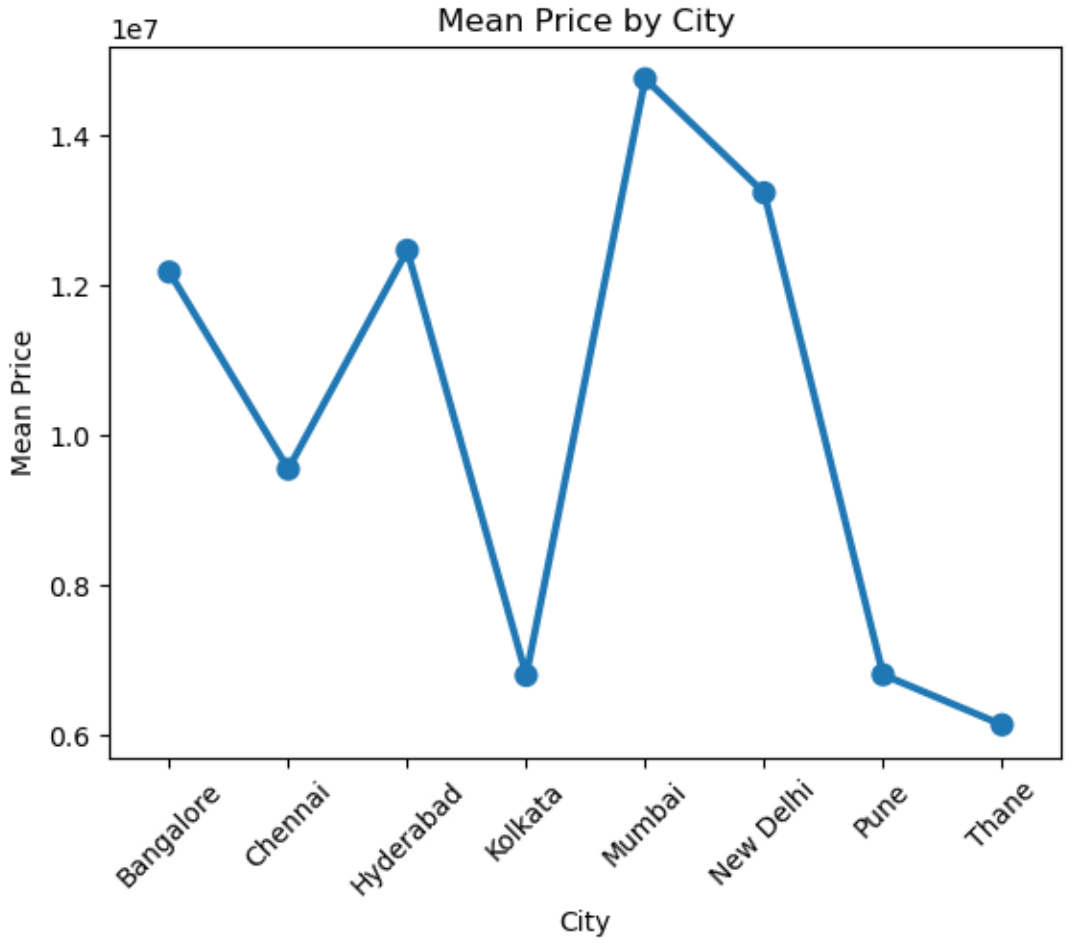
3. **Findings**

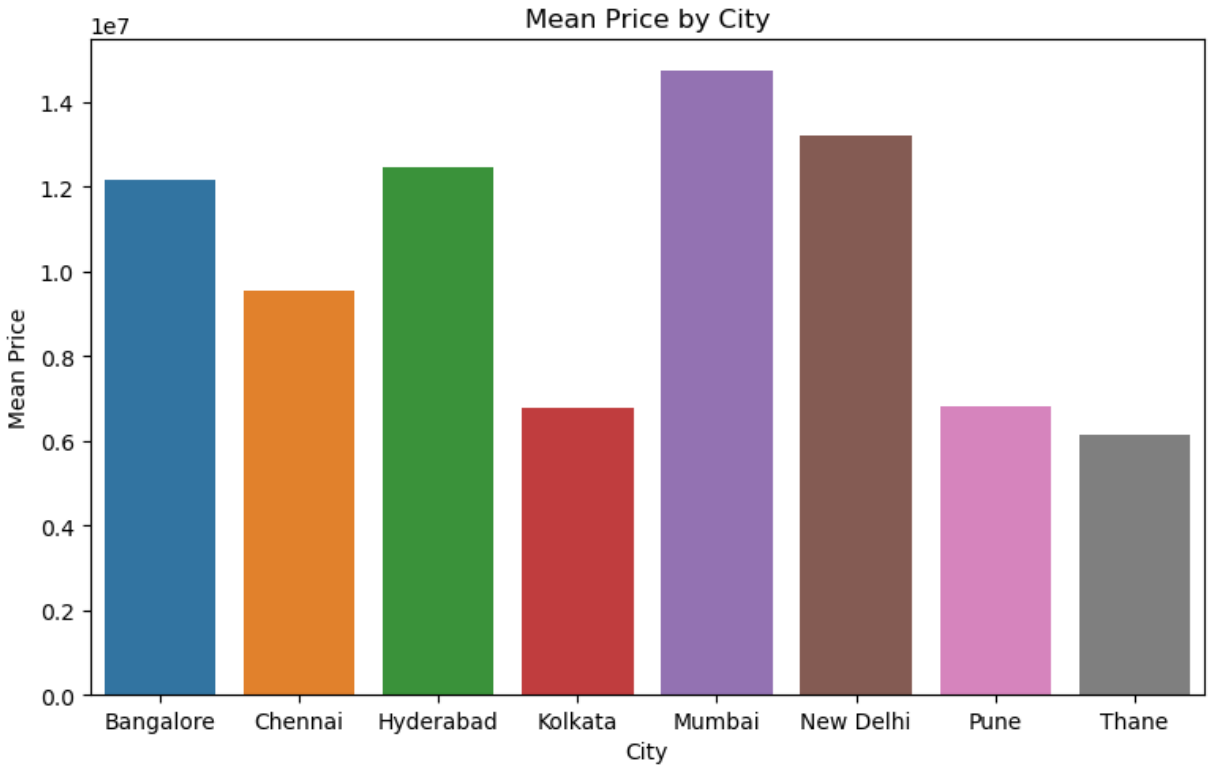
3.1 **Property Details and Habits**

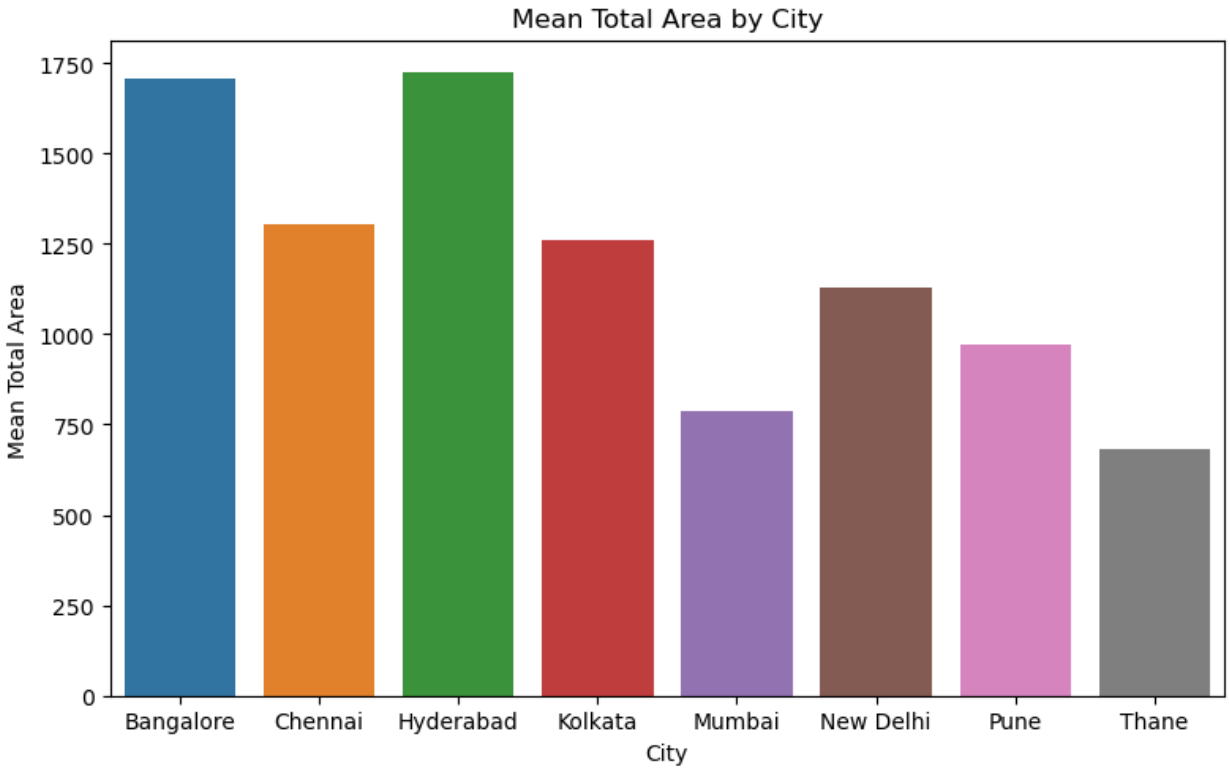
Location Analysis: 2 BHK Flat for sale in Electronic City, Bangalore, 2 BHK Flat for sale in Hadapsar, Pune, and 1 BHK Flat for sale in Mira Road East, Mumbai have the highest number of titles at 53, 39, and 38 respectively.
Sector 12 Dwarka, New Delhi has the most locations at 44, with Wagholi, Pune in second at 39, and Bannerughatta, Bangalore in third with 35 locations.

Price Distribution:

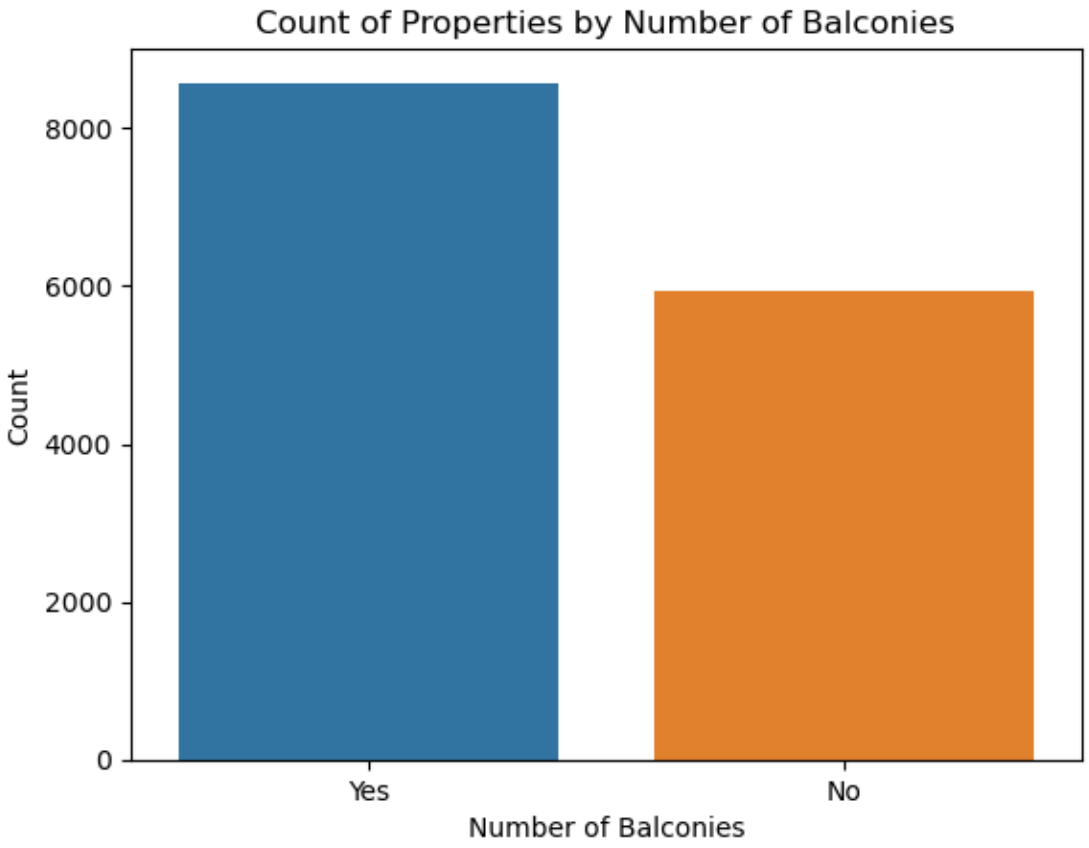
- Mumbai has the highest average price followed by New Delhi and Hyderabad.

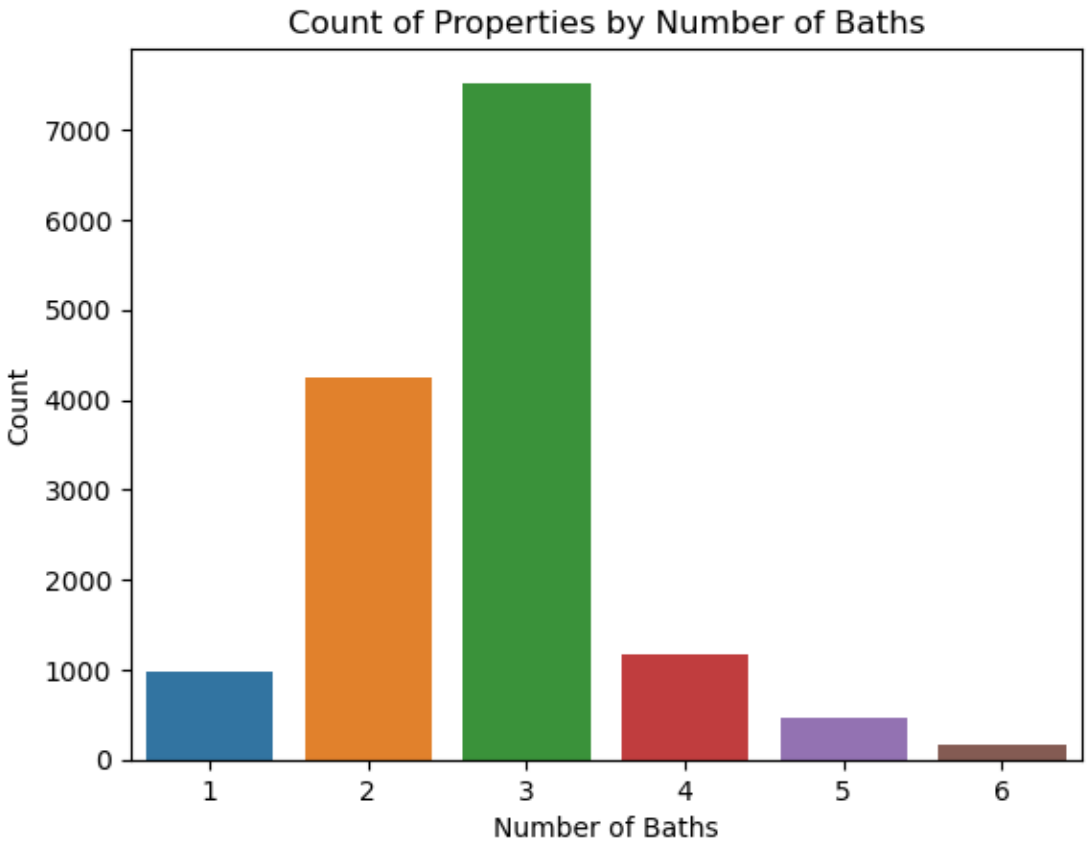




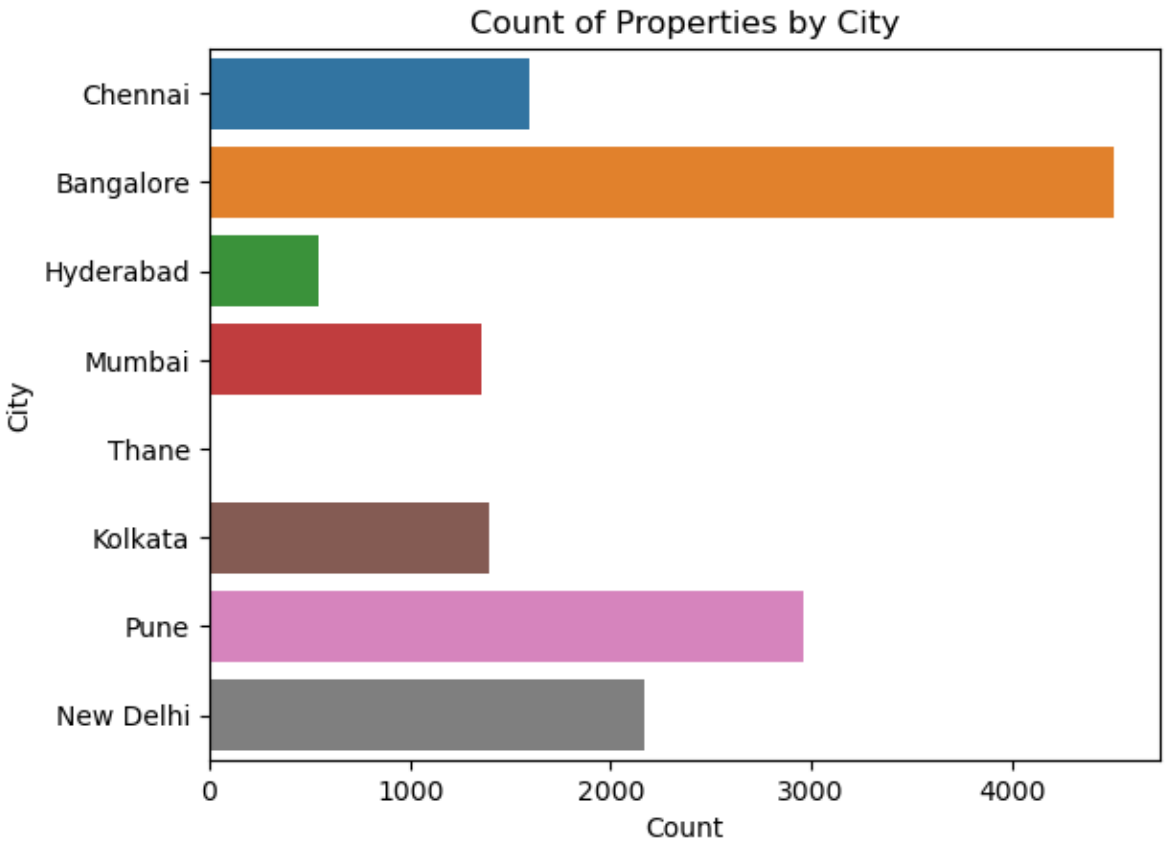


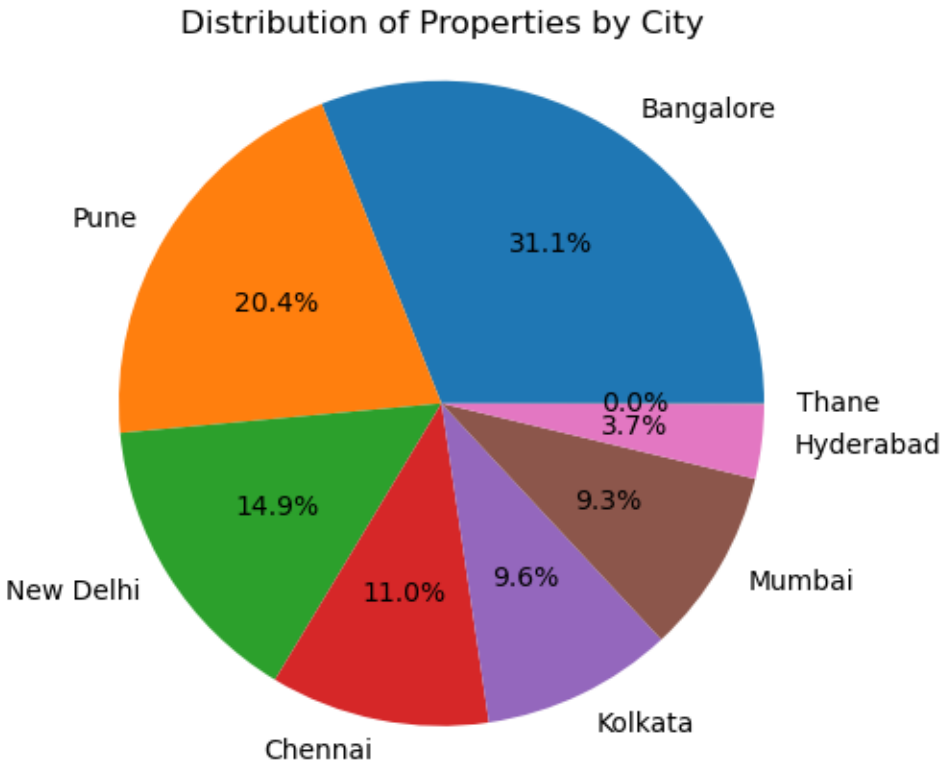
- Most properties have a balcony and 3 bathrooms.


Count of Properties by Number of Balconies
8000
6000
4000
2000
0
Count
Yes
No
Number of Balconies



- Bangalore has the most entries at 4,511. Pune comes second at 2,964, and New Delhi in third at 2,165.





- Bangalore and Hyderabad have the highest mean total area.

4. **Conclusion**

   The analysis provided valuable insights into real estate property details and habits. Properties varied greatly in terms of price, area, number of bathrooms, and balconies, suggesting a diverse real estate market.

5. **Recommendations**

   Based on the analysis, the following is recommended:

   - Investigate the relationship between property attributes (like location, total area, number of bathrooms and balconies) and price.
   - Conduct market research to understand property preferences and trends in different locations.
   - Implement targeted marketing campaigns for specific property types and locations.
   - Explore the impact of property characteristics on their price.

6. **Acknowledgments**

   I would like to express our gratitude to Rakkesh Aravind G for providing the "Real Estate Data" dataset.

7. **References**

   The analysis used the following libraries and tools:

- NumPy: A library for numerical computing in Python.
- Pandas: A data manipulation and analysis library.
- Matplotlib: A plotting library.
- Seaborn: A data visualization library.
- Jupyter Notebook: An interactive coding environment.

For a detailed understanding of the analysis process, including the Jupyter Notebook code used, please refer to the original analysis documentation.

Best regards,

Nikhil Sharma