FALCON
INFORMATICS

Demystifying Real Estate Prices

CASE STUDY

Author
Insiya Maryam

Table of **Contents**

Overview

Predicting property prices using machine learning and real estate data analytics. By analyzing property listings from different city tiers across India and identifying the key determinants influencing property valuation.

The dataset used an open dataset containing more than 300k rows and 30 columns, covering real estate listings from Tier-1, Tier-2, and Tier-3 cities.

This case study enables stakeholders to estimate property prices, explore spatial trends, and make informed investment decisions.

Introduction

Real estate is a dynamic sector influenced by numerous variables - location, property type, size, city development, and accessibility to amenities. Traditional valuation models often fail to capture non-linear relationships between these factors.

This project uses data-driven techniques to uncover patterns in property pricing and deploys a predictive model that assists buyers, sellers, and real estate analysts. Using open-source data, the model predicts prices for properties across India and integrates geographic visualization to highlight the spatial impact on pricing.

Problem Statement

Real estate often face challenges in accurately estimating property prices due to:

- High variability across city tiers and neighborhoods
- Lack of structured, analyzed spatial data
- Unpredictable correlations between property attributes and pricing

The case study aims to:

1. Build a machine learning model capable of predicting property prices with high accuracy.
2. Analyze spatial and categorical factors affecting price variations.
3. Develop an interactive web application that visualizes nearby amenities and their influence on property value.

By combining predictive analytics with geospatial visualization, the project seeks to “demystify” how real estate values are determined across different regions.

Methodology

1. **Data Description**

The dataset attributes includes:

- Price, Area, Bedrooms, Bathrooms
- City, Furnishing Status, Property Type
- Latitude and Longitude

City Name
Ahmedabad
Bangalore
Chennai
Delhi
Hyderabad
Kolkata
Lucknow
Mumbai

	Property_id	Property_type	Property_status	Price_per_unit_area	Posted_On	builder_id	Builder_name	Property_building_status	City_id	City_name
0	15446514	Apartment	Under Construction	4,285	1 day ago	100563465.0	Arkiton life Space	ACTIVE	1	Ahmedabad
1	15367414	Apartment	Under Construction	7,000	2 days ago	100009433.0	Keshav Narayan Group	ACTIVE	1	Ahmedabad
2	14683118	Apartment	Ready to move	5,752	2 days ago	100207731.0	Vishwa Developers Ahmedabad	ACTIVE	1	Ahmedabad
3	5476295	Apartment	Ready to move	2,486	5 days ago	101303.0	Satyam Developers	ACTIVE	1	Ahmedabad
4	15477040	Apartment	Under Construction	5,324	8 days ago	1484209.0	Navkar Buildcon Ahmedabad	ACTIVE	1	Ahmedabad

_furnished	listing_domain_score	is_plot	is_RERA_registered	is_Apartment	is_ready_to_move	is_commercial_Listing	is_PentaHouse	is_studio	Listing_Category
nfurnished	4.0	False	True	True	False	False	False	False	sell
nfurnished	4.0	False	True	True	False	False	False	False	sell
nfurnished	4.0	False	False	True	True	False	False	False	sell
nfurnished	4.0	False	False	True	True	False	False	False	sell
nfurnished	4.0	False	True	True	False	False	False	False	sell

Methodology

2. Data Preprocessing

The raw data contained missing values, inconsistencies, and duplicates.

- Imputed missing numerical values using median and categorical ones with mode.
- Removed duplicate listings and extreme outliers using IQR-based filtering.
- Encoded categorical variables such as City, Furnishing, and Property Type using label encoding.
- Standardized numerical columns for improved model convergence.
- Validated and corrected invalid latitude-longitude coordinates.

3. Exploratory Data Analysis (EDA)

EDA was conducted to understand key trends, correlations, and patterns:

- Price Distribution: Skewed toward higher values, showing a long tail of luxury properties.
- City-Level Trends: Mumbai, Bangalore, and Hyderabad emerged as top-valued metros, while Tier-2 cities like Indore and Pune had balanced mid-range markets.
- Feature Correlation: Strong positive correlation between Area and Price; moderate influence from Bedrooms and Furnishing.
- Furnishing Impact: Furnished homes were priced higher than semi/unfurnished ones.
- Spatial Analysis: City-wise scatterplots revealed clear clusters of high-value zones in metro areas.

These analyses provided the foundation for selecting relevant predictive features.

Methodology

4. Feature Engineering

- Derived Price_per_sqft feature for better price scaling.
- Encoded categorical attributes numerically.
- Removed redundant features based on high correlation.
- Integrated geographic proximity metrics for later use in visualization.

5. Model Development

Multiple models were tested — Linear Regression, Ridge Regression, and Random Forest.

Among them, Random Forest Regressor achieved the highest performance due to its ability to model non-linear patterns and manage categorical-numeric mixes.

Training Details:

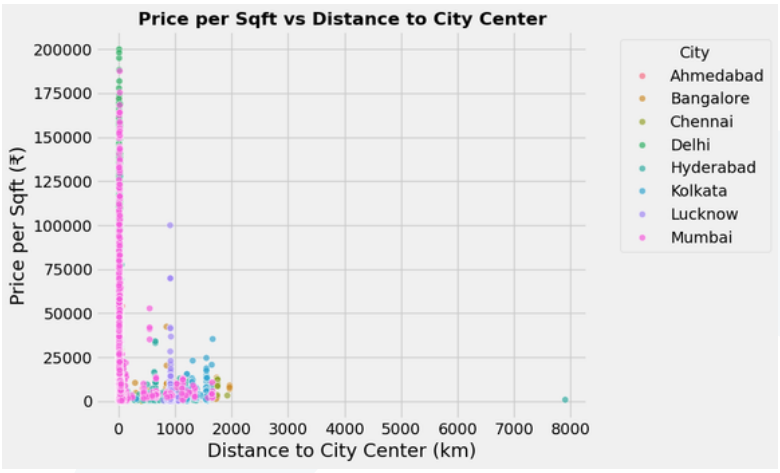
- Data split: 80% training, 20% testing
- Final model saved as: Real_Estate_RF_Model.joblib

Performance Metrics:

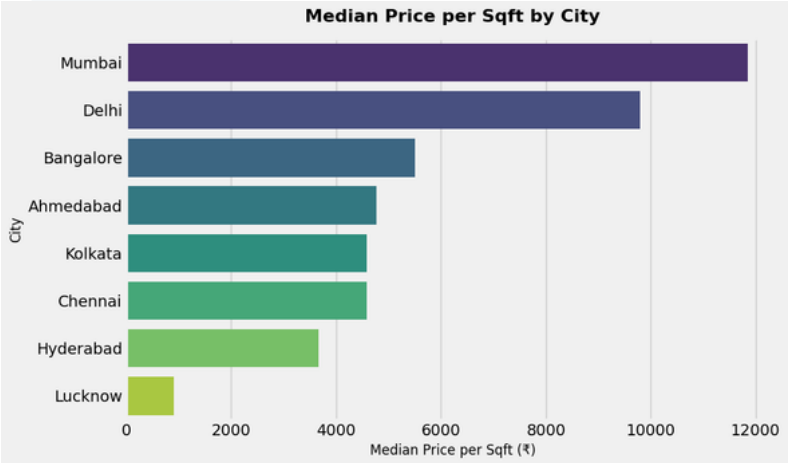
- R^2 Score: ~0.87
- RMSE: ~5.8 lakh INR

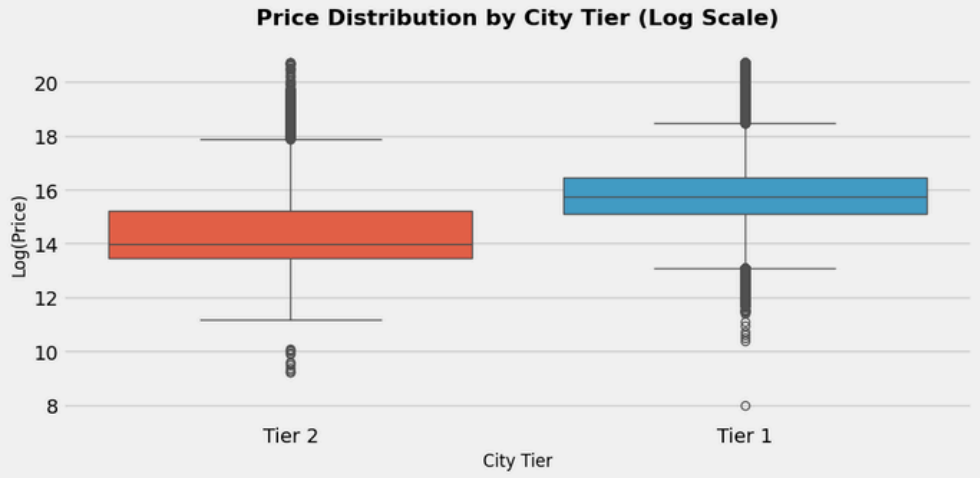
Feature importance indicated Area, City, Bedrooms, and Furnishing as top predictors. The model effectively captured city-level and regional price variations.

Methodology




```
City_name
Mumbai        11855.670103
Delhi          9791.666667
Bangalore      5506.072874
Ahmedabad      4766.612357
Kolkata        4590.891089
Chennai        4583.333333
Hyderabad      3666.666667
Lucknow         900.000000
Name: price_per_sqft_derived
```





Methodology

6. Output

A web application was created to operationalize the model.

Named "Real Estate Price Prediction & Nearby Amenities Visualizer", it enables users to:

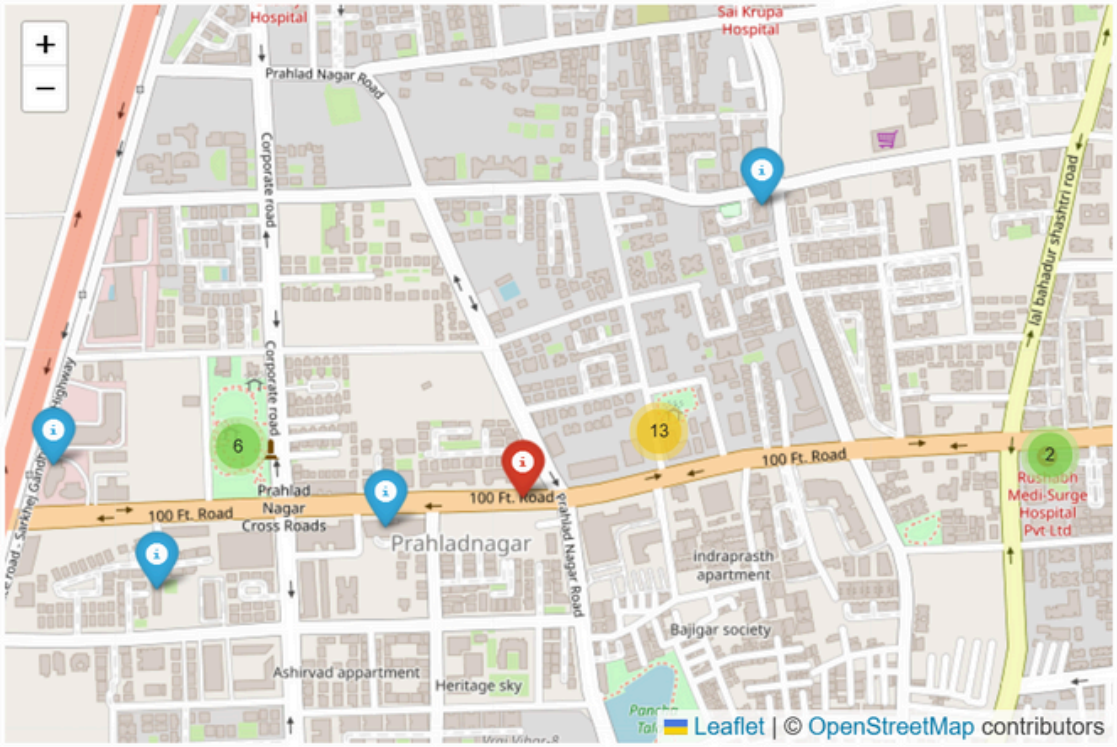
- Select property parameters (city, area, furnishing, etc.)
- Predict the estimated price using the trained Random Forest model
- View nearby amenities (schools, hospitals, restaurants) using Folium maps integrated with OpenStreetMap (Overpass API)

This tool transforms static model predictions into an interactive, spatially aware experience for real-world usability.

Methodology

FALCON INFORMATICS

Amenities Around Selected Property



Found 25 amenities nearby.

FALCON INFORMATICS

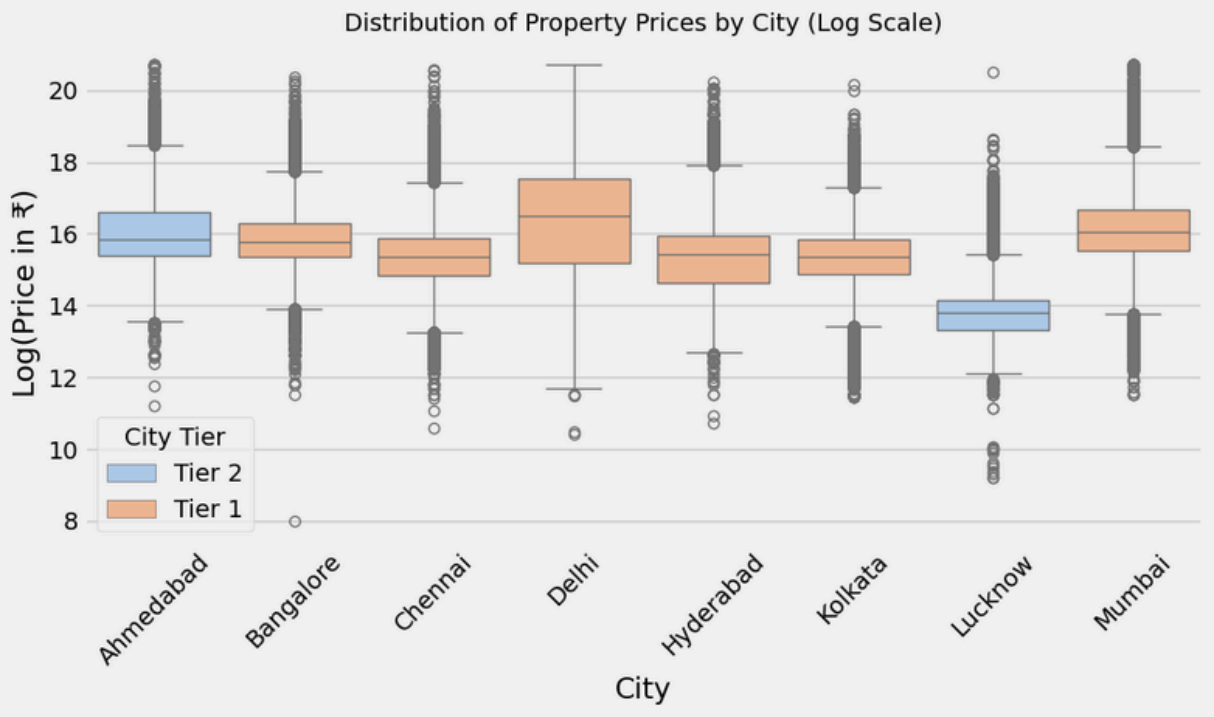
Found 25 amenities nearby.

Nearby Amenities List

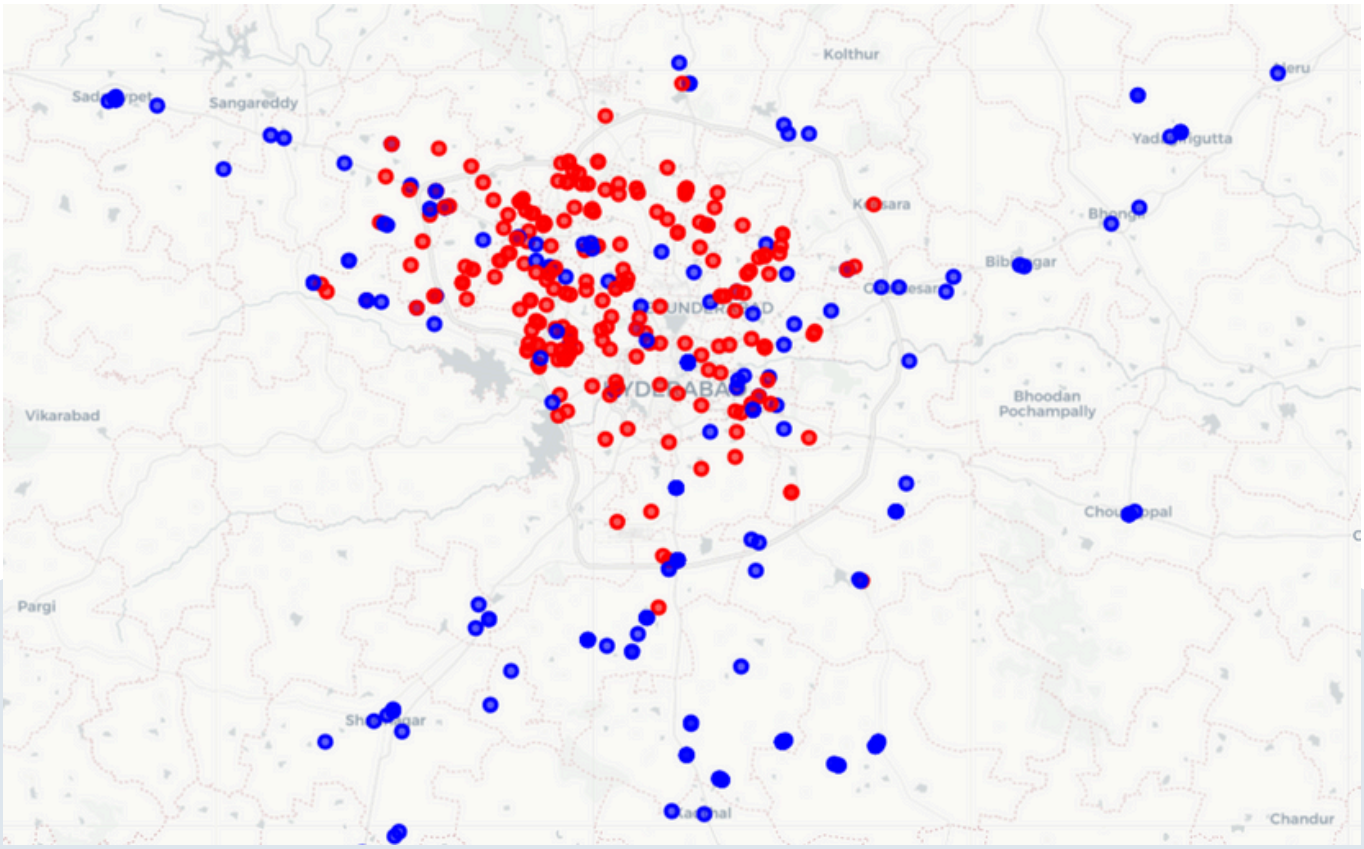
	Amenity Type	Name
0	Fuel	Shell
1	Bank	Yes Bank
2	Bar	Sphere Louge
3	Shelter	Unnamed
4	Shelter	Unnamed
5	Fountain	Unnamed
6	Ice_Cream	Amul
7	Cafe	In-Garden Cafe
8	Hospital	Rushabh Medi-Surge Hospital Pvt Ltd
9	Hospital	Sahjanand Orthopaedic Hospital

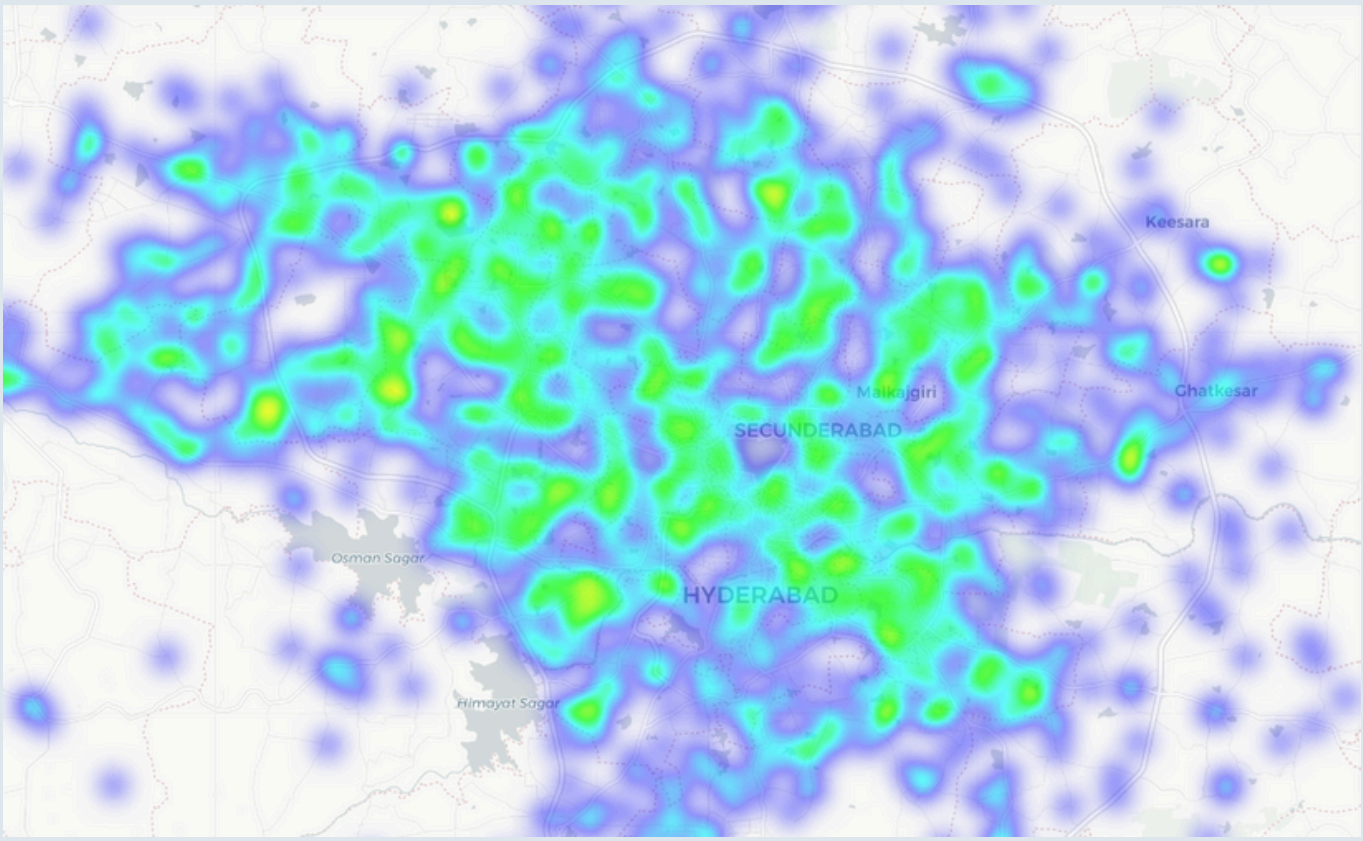
Key Insights

1. Area and Location are the strongest determinants of price — metro cities show exponential price jumps with increasing area.
2. Furnishing status significantly influences value — furnished units are 15–25% more expensive on average.
3. Spatial accessibility plays a vital role — proximity to key amenities leads to higher property appreciation.
4. Tier-2 cities are showing growing consistency, indicating potential for investment stability.
5. Random Forest outperforms linear models, offering both interpretability and high prediction accuracy.
6. The Web app bridges analysis and visualization, making it easier for businesses and end-users to make data-driven decisions.



Key Insights





Conclusion

The Demystifying Real Estate Prices project demonstrates how machine learning and geospatial analytics can be effectively combined to understand and predict property values.

By using a diverse dataset of over 3 lakh records, the Random Forest model achieved strong predictive accuracy and meaningful interpretability.

The integrated Streamlit application provides a practical interface for users to visualize predicted prices and surrounding amenities — bridging the gap between data analytics and real-world decision-making.

Future enhancements may include:

- Incorporating temporal data for trend forecasting
- Building rental and ROI prediction modules
- Extending API deployment for commercial integration

This solution marks a significant step toward data-driven real estate intelligence in the Indian market.

FALCON INFORMATICS

TRANSFORMING BUSINESSES WITH ERP SYSTEMS, DATA SCIENCE, AND INTELLIGENCE

Delivering Data-Driven Insights for Smarter Decisions and Sustainable Growth

Contact Us

www.falconinformatics.com