Predicting Home Prices using Bengaluru House Prices Dataset

Author: Jubair Abdulla

Date: 2025.05.19

CONTENTS

Project Objective & Use Case

1 Objective

Predict house prices in Bengaluru using features like size, location, BHK, etc.

2 Problem Type

Regression — target variable is continuous (price in Lakhs INR).

3 Use Case

Helps users estimate home value and assists in decision-making for buying/selling real estate.

Data Preprocessing & Feature Engineering

Dataset Cleaning

Handled missing values

Feature Selection

Dropped irrelevant features: area_type, availability, society, balcony

Feature Engineering

Transformed size (e.g., "2 BHK") to numeric BHK

Converted range values like '2100-2850' sqft into their average

Categorical Encoding

Applied One-Hot Encoding to location column

EDA & Outlier Handling

Visual Exploration

Scatter plot of 2 BHK vs 3 BHK to compare sqft vs price

Identified irregular patterns in square footage and pricing

Distribution Analysis

Most homes have a price per square foot in the range of 3000 to 7000.

Outlier Removal

Removed entries with extreme bathroom counts (e.g., > 10 bathrooms)

Filtered out inconsistent data points (e.g., too much price difference for similar sqft)

Modeling with Linear Regression

Algorithm Used:

Linear Regression

Features Used:

total_sqft, bath, bhk, and encoded location columns

Data Preparation:

Created input array using numpy with correct feature positions
Trained using scikit-learn’ s LinearRegression model

Model Evaluation & Hyperparameter Tuning

Evaluation Metrics:

R^2 Score

Mean Squared Error (MSE)

Train-Test Split:

Used train_test_split() to evaluate performance

Cross Validation:

Applied ShuffleSplit cross-validator

Hyperparameter Tuning:

Compared Linear Regression, Lasso, and Decision Tree Regressor using GridSearchCV

Selected best performing model based on R^2

UI Interface (Price Estimation Tool)

1

Frontend

Built with HTML/CSS and JavaScript

Drop-down menu to select location

Inputs for sqft, BHK, bath

2

Backend

Flask-based API

Model and columns loaded from pickle and JSON files

3

Output

Returns predicted price in Lakhs INR when inputs are submitted

Business Insights & Conclusion

Insights

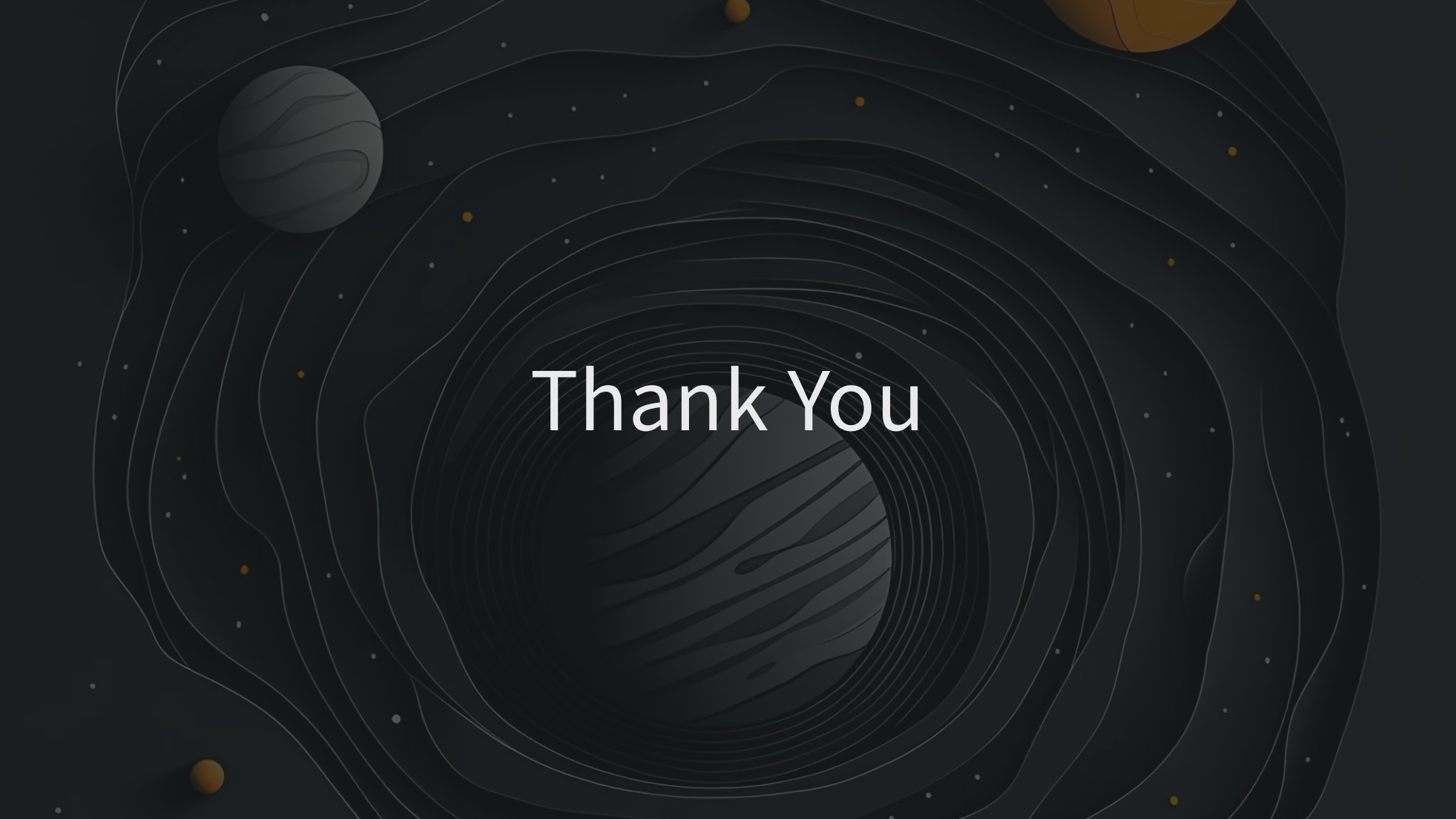
Location significantly affects home price — more than sqft or BHK alone

Price trends can guide investment opportunities in real estate

Conclusion

Model performs well for basic estimation

This project demonstrates how machine learning can simplify complex real estate decisions for buyers and sellers.


Thank You