



# Bangalore House Price Prediction

## Model Development & Deployment Documentation

---

## 1. Project Overview

This project focuses on analyzing Bangalore housing data, building a Machine Learning model to predict house prices, and deploying the trained model using **Streamlit** as a web application.

The work is divided into two main parts:

1. **Data Analysis & Model Development** (`House_Pricing_Predictions.ipynb`)
  2. **Model Deployment using Streamlit** (`app.py`)
- 

## 2. Dataset Description

The dataset used is the **Bangalore Housing Prices Dataset** sourced from Kaggle.

### Dataset Link:

<https://www.kaggle.com/datasets/aryanfelix/bangalore-housing-prices>

### Key Features Used:

- `total_sqft` – Total area of the house
  - `bath` – Number of bathrooms
  - `balcony` – Number of balconies
  - `bhk` – Number of bedrooms
  - `location` – Area/location of the property
- 

## 3. Data Exploration (EDA)

Initial data exploration was performed in `House_Pricing_Predictions.ipynb` to understand the structure and quality of the dataset.

### Steps Performed:

- Loaded the dataset using Pandas
- Checked dataset shape, column names, and data types
- Identified missing values and inconsistent entries

- Analyzed location-wise distribution of properties
- Explored relationships between:
  - Square footage and price
  - BHK and price
  - Location and price

This step helped in identifying important features influencing house prices.

---

## 4. Data Cleaning & Preprocessing

To prepare the data for modeling, the following preprocessing steps were applied:

- Removed rows with missing or invalid values
- Converted `total_sqft` into numerical format
- Removed extreme outliers based on business logic
- Reduced location categories by grouping rare locations into an “other” category
- Converted categorical `location` feature using **One-Hot Encoding**

After preprocessing, the dataset was ready for model training.

---

## 5. Feature Engineering

The final features used for training were:

- Numerical features:
  - `total_sqft`
  - `bath`
  - `balcony`
  - `bhk`
- Encoded categorical features:
  - Location-based one-hot encoded columns

The processed feature list was saved in `columns.json` for use during deployment.

---

## 6. Model Training

A regression-based Machine Learning model was trained using Scikit-learn.

### Training Process:

- Split data into training and testing sets
- Applied regression algorithm to learn patterns in housing prices
- Evaluated model performance using accuracy metrics

- Tuned parameters to improve prediction reliability

The trained model was serialized using **Pickle** and saved as:

`banglore_home_prices_model.pkl`

---

## 7. Model Evaluation

The model was evaluated based on:

- Prediction accuracy on unseen test data
- Logical price estimates across different locations
- Consistency of predictions for similar property inputs

This ensured the model generalizes well and provides realistic predictions.

---

## 8. Streamlit Application Integration

After training the model, it was integrated into a **Streamlit web application**.

### Streamlit Workflow:

1. Load trained model (`.pkl` file)
  2. Load feature columns from `columns.json`
  3. Take user inputs:
    - Location
    - Square feet
    - Bathrooms
    - Balconies
    - BHK
  4. Convert inputs into model-compatible format
  5. Predict house price using the trained model
  6. Display estimated price in **Lakhs (₹)**
- 

## 9. Deployment using Streamlit Cloud

The Streamlit application was deployed using **Streamlit Cloud**.

### Live Application Link:

<https://bangalorehousepriceprediction-prasadshetty.streamlit.app/>

The deployment enables users to interact with the trained ML model through a web interface without requiring local setup.

---

## 10. Project Structure

```
Bangalore-House-Price-Prediction/
├── app.py                                # Streamlit application
├── requirements.txt                        # Dependencies
├── banglore_home_prices_model.pickle
├── columns.json
└── README.md

├── data/
│   └── BHP.csv                            # Dataset

└── notebook/
    └── House_Pricing_Predictions.ipynb      # EDA & Model Training
```

---

## 11. Key Learnings

- Exploratory Data Analysis (EDA) on real-world datasets
  - Data cleaning and feature engineering techniques
  - Training and evaluating regression models
  - Model serialization using Pickle
  - Deploying ML models using Streamlit Cloud
  - Building end-to-end Machine Learning projects
- 

## 12. Conclusion

This project demonstrates an end-to-end Machine Learning workflow starting from raw data analysis to model deployment.

The integration of EDA, model training, and Streamlit deployment showcases practical skills required for real-world data science applications.