### **Introduction**

In the rapidly evolving real estate market, accurately predicting property prices is crucial for both buyers and sellers. This project focuses on developing a robust machine learning model that leverages key features such as the number of bedrooms, location, and price per square foot to predict real estate prices. By utilizing advanced data preprocessing techniques, including feature engineering and outlier detection, the model aims to provide reliable and actionable insights for users. The model is deployed through a user-friendly Flask web application, allowing real-time predictions and seamless user interaction.

**Analysis**

**1. Data Preprocessing and Feature Engineering:**

*Dropping Irrelevant Features:*

It started by dropping features such as area\_type, society, balcony, and availability which were not necessary for building the model.

*Handling Missing Data:*

Removing rows with missing values to ensure the dataset was clean.

*Feature Creation:*

Added a new feature bhk by extracting it from the size column.

Created price\_per\_sqft by dividing the price by the total square feet area.

*Handling NonNumeric Data:*

Converted the total\_sqft feature to a numeric value, handling ranges by averaging and dropping complex cases.

**2. Dimensionality Reduction:**

*Location Encoding:*

Locations with less than 10 data points were categorized as other to reduce the number of categories for better model performance.

*OneHot Encoding*:

Applied one hot encoding to the location feature to convert categorical data into numerical.

**3. Outlier Detection and Removal:**

*Business Logic*:

Removed outliers based on domain knowledge, such as apartments with less than 300 sq ft per bedroom.

*Statistical Methods:*

Used standard deviation and mean to remove price outliers within each location.

*BHKBased Outlier Removal:*

Removed properties where the price per square foot of a higher BHK was less than that of a lower BHK in the same location.

**4. Model Building:**

*TrainTest Split:*

Split the dataset into training and testing sets with an 80/20 split.

*Linear Regression:*

Built an initial model using Linear Regression.

*CrossValidation:*

Used K Fold cross validation to evaluate the model performance, ensuring consistency.

*Model Selection with GridSearchCV:*

Comparing different models (Linear Regression, Lasso, DecisionTree) using GridSearchCV to find the best hyperparameters.

**5. Model Deployment:**

*Pickle for Model Persistence:*

The trained model was serialized using pickle for future use.

*JSON for Feature Names:*

The feature names were saved in a JSON file to be used in the application.

**6. Web Application:**

*Frontend Development:*

Created a frontend interface using HTML, CSS, and JavaScript. The application allowed users to input property features and get price predictions.

*AJAX for Asynchronous Requests:*

Used jQuery's AJAX method to send user inputs to a Flask backend and receive predictions.

**7. Flask Backend:**

*Handling Requests:*

The Flask app would handle incoming requests, load the model, and return predictions.

*API Endpoint:*

Setting up an endpoint /predict\_home\_price to handle POST requests with user inputs and return estimated prices.

**Future Enhancements:**

* Additional Features: Incorporate more features such as the age of the property, amenities, and nearby facilities for better predictions.
* Advanced Models: Explore other advanced machine learning models or ensembles to improve prediction accuracy.
* RealTime Data: Integrate real time data feeds for dynamic predictions based on the latest market trends.

### **Conclusion**

This real estate price prediction project successfully demonstrates the application of machine learning in forecasting property values with high accuracy. The integration of advanced feature engineering and model tuning techniques significantly improved prediction performance. The deployment of the model via a Flask web application enhances accessibility, making it a valuable tool for end-users seeking real-time price predictions. Overall, this project showcases the potential of data-driven approaches in addressing complex problems in the real estate industry, offering a scalable solution that can adapt to evolving market trends.