**House Price Analysis and Prediction**

**Project Overview**

This project aims to analyze and predict house prices using a dataset that includes real estate property listings with details on attributes, location, and pricing.

**Methodology**

**Data Cleaning and Exploration**

* Filled missing values for 'agency' and 'agent' columns with 'Unknown'.
* Dropped rows with missing values where necessary.

**Identify and Handle Inconsistencies**

* Stripped leading/trailing spaces and converted categorical data to lowercase.
* Validated 'price', 'latitude', 'longitude', 'baths', and 'bedrooms' to ensure data integrity.

**Checking Duplicates**

* Identified and removed duplicate records to ensure unique entries.

**Outlier Detection and Handling**

* Used scatter plots and box plots to visualize outliers in numerical features.
* Applied the IQR method and Winsorization to handle outliers.

**Analyze the Distribution of House Prices**

* Utilized histograms and KDE plots to visualize price distribution.
* Created scatter plots and bar plots to identify relationships between features and price.

**Correlation Analysis**

* Generated a correlation heatmap to analyze the relationships between numerical features.

**Visualization of Geographical Data**

* Used Folium to create a map visualizing property locations and prices.

**Create New Features**

* Calculated the age of the house.
* Converted area measurements to a numeric format (square feet).
* Computed price per square foot.
* Derived additional features like the number of bedrooms per floor and bathrooms per bedroom.

**Encode Categorical Features**

* Applied one-hot encoding to categorical features.

**Predictive Modeling**

**Model Training and Evaluation**

* Split the dataset into training and testing sets.
* Trained Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor models.
* Evaluated models using mean absolute error, mean squared error, and R-squared metrics.

**Challenges**

* Handling and cleaning a large dataset with numerous attributes.
* Managing and visualizing outliers to maintain data integrity.
* Engineering has meaningful features to improve model performance.

**Conclusion**

Through thorough data cleaning, exploration, and feature engineering, followed by model training and evaluation, this project successfully analyzed and predicted house prices. The use of various visualizations and statistical methods ensured a robust approach to understanding the dataset and deriving insights. The predictive models developed can assist in estimating house prices based on the attributes provided.