**Predicting Property Prices using Machine Learning**

**1. Introduction to the Project**

The real estate market is influenced by multiple factors such as property size, location, amenities, and economic conditions. Accurately predicting property prices is crucial for buyers, sellers, and real estate professionals. This project aims to develop a machine learning model to predict property prices using historical data, employing various preprocessing, exploratory data analysis (EDA), and modelling technique like Random forest and Linear regression.

**2. Objectives of the Project**

1. Collect and clean real estate data from a Given Dataset.
2. Handle ordinal and nominal features separately using appropriate encoding techniques.
3. Use data preprocessing techniques like scaling, Principal Component Analysis (PCA), and imputation for handling missing data.
4. Perform exploratory data analysis (EDA) to identify key factors affecting property prices. Along with Correlation Analysis .
5. Develop a predictive model using machine learning techniques such as Linear Regression, Random Forest, and Gradient Boosting.
6. Evaluate model performance using metrics such as R^2 score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).
7. Deploy the best-performing model for price prediction.

**3. Flow Chart of Operations**

1. **Data Collection**: Load and inspect the Dataset into the terminal
2. **Data Cleaning**: Handle missing values and outliers and treating them .
3. **EDA**: Analyze and visualize relationships between features and property price.
4. **Feature Engineering**: Create new relevant features.
5. **Data Preprocessing**: Encode categorical features and scale numerical ones.
6. **Model Training & Selection**: Train multiple machine learning models and evaluate them.
7. **Hyperparameter Tuning**: Optimize model performance using GridSearchCV.
8. **Model Evaluation**: Compare performance across different models.
9. **Deployment & Insights**: Deploy the final model and interpret key insights.

**4. Python Codes**

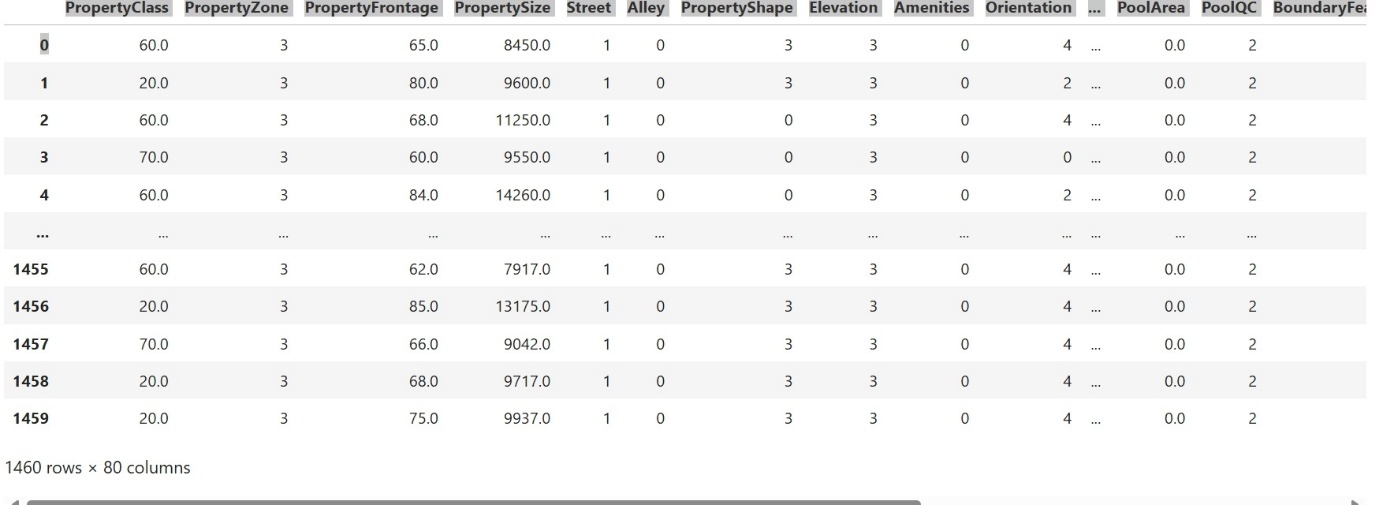
Python code is implemented to perform all the Below steps, including:

* Data Loading & Cleaning
* Exploratory Data Analysis
* Feature Engineering
* Model Training & Evaluation
* Hyperparameter Tuning

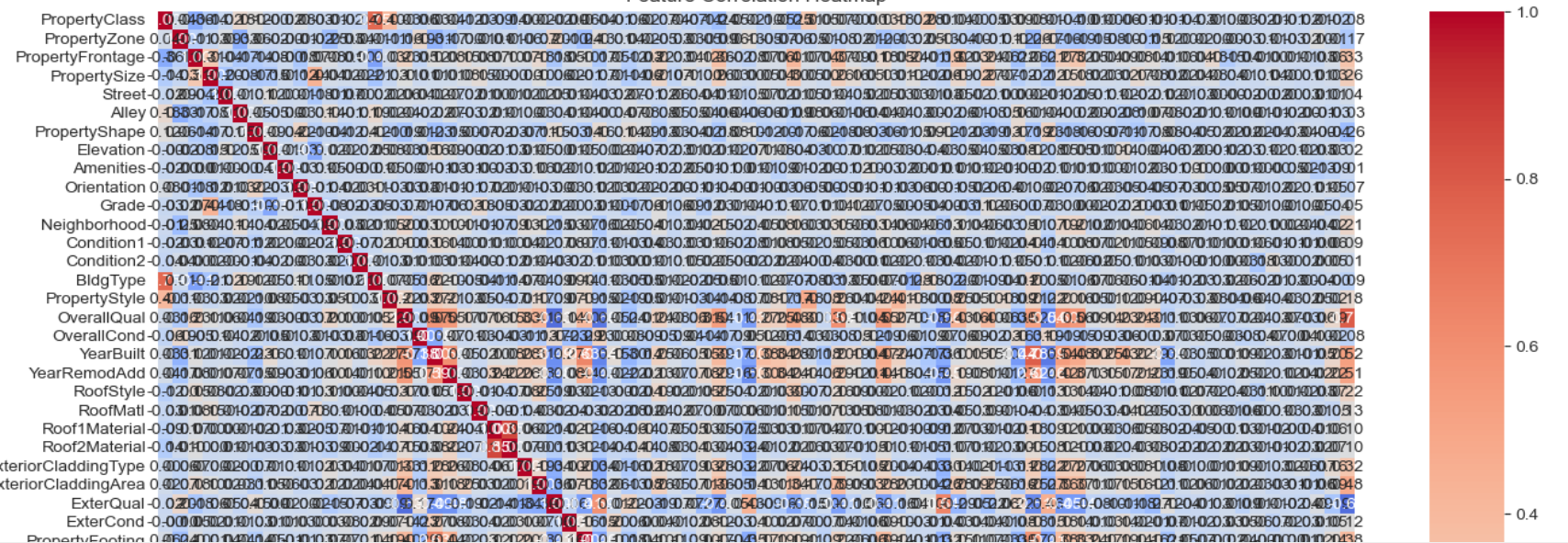
(Refer to the attached Jupyter Notebook for complete code.)

**5. Screenshots of the Outputs**

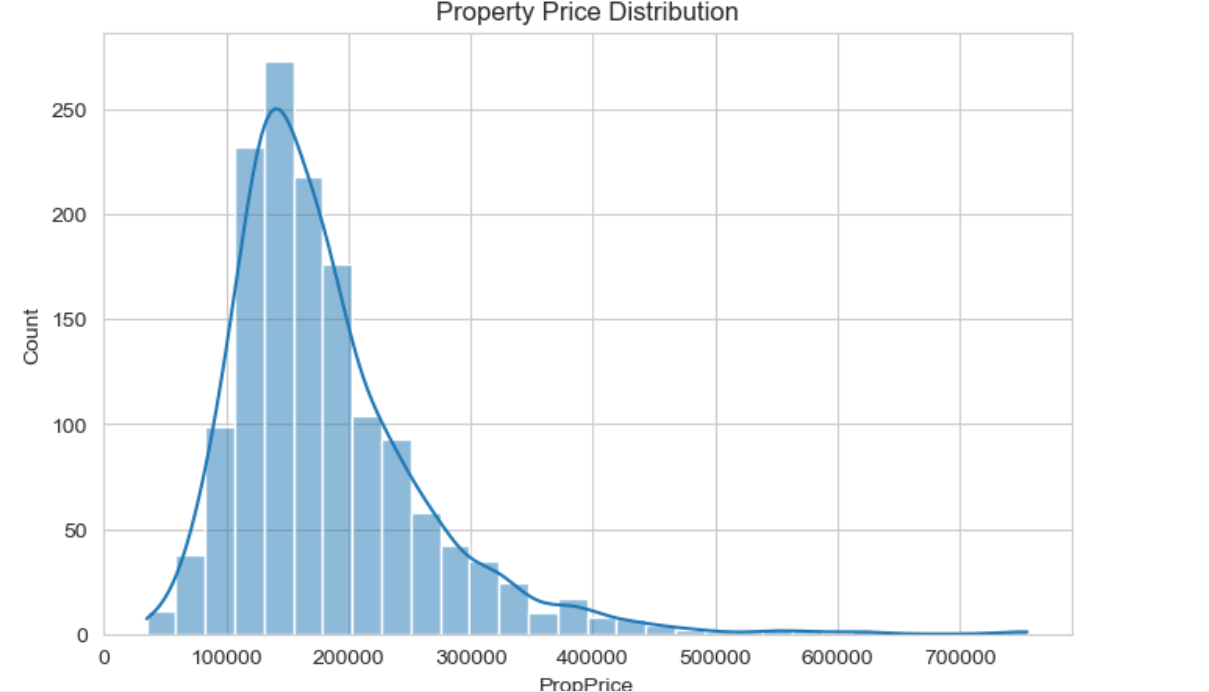
* Dataset overview after cleaning.



* Correlation heatmap.



* Distribution of property prices.



* Model performance comparison table.

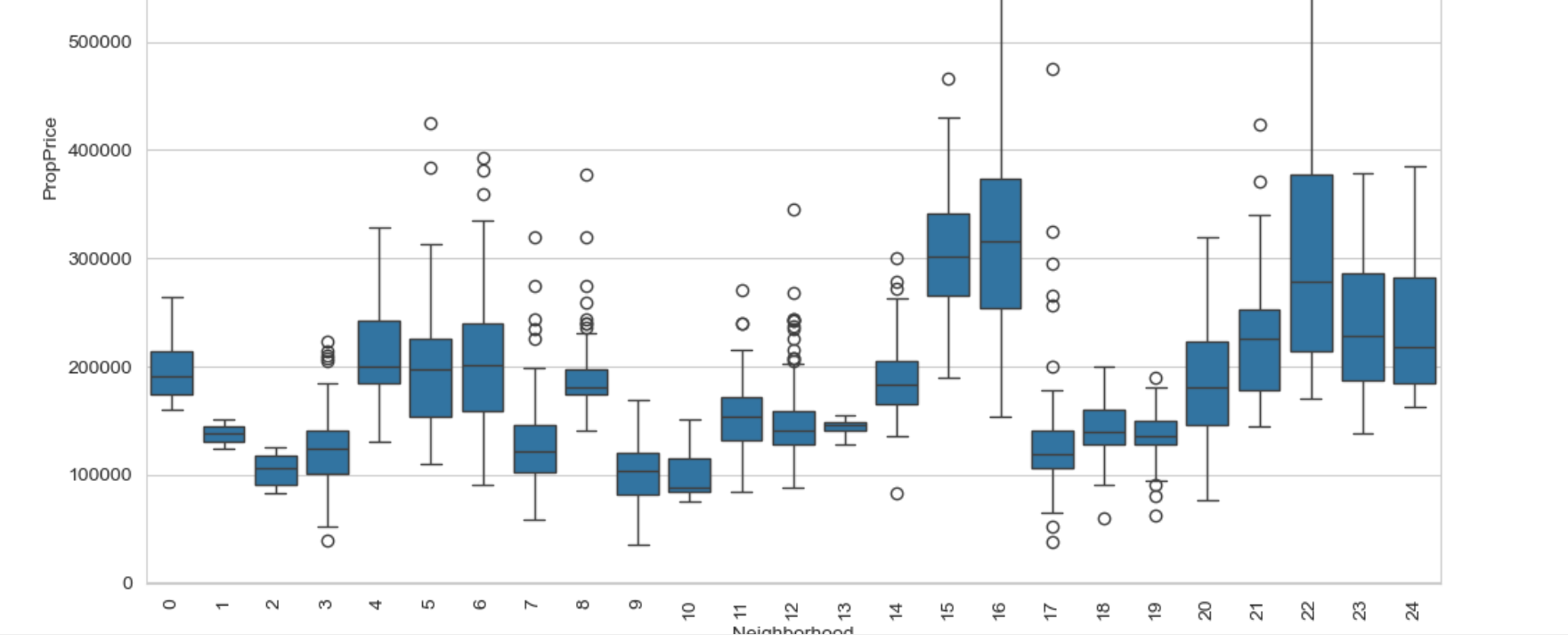
|  |  |  |
| --- | --- | --- |
| METRIC | LINEAR REGRESSION | RANDOM FOREST |
| R2 SCORE | 0.84 | 0.89 |
| MAE | 21566 | 17591 |
| MSE | 32295 | 28730 |

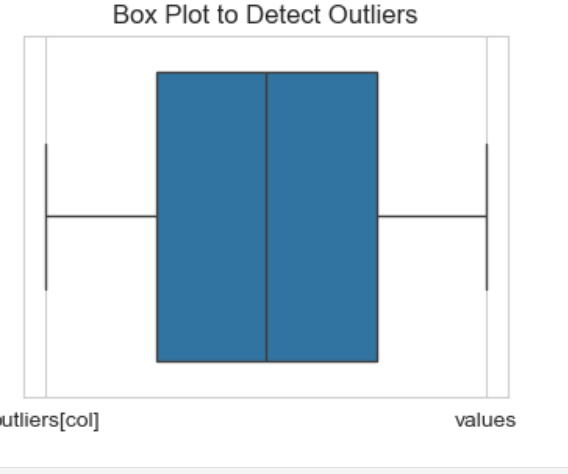
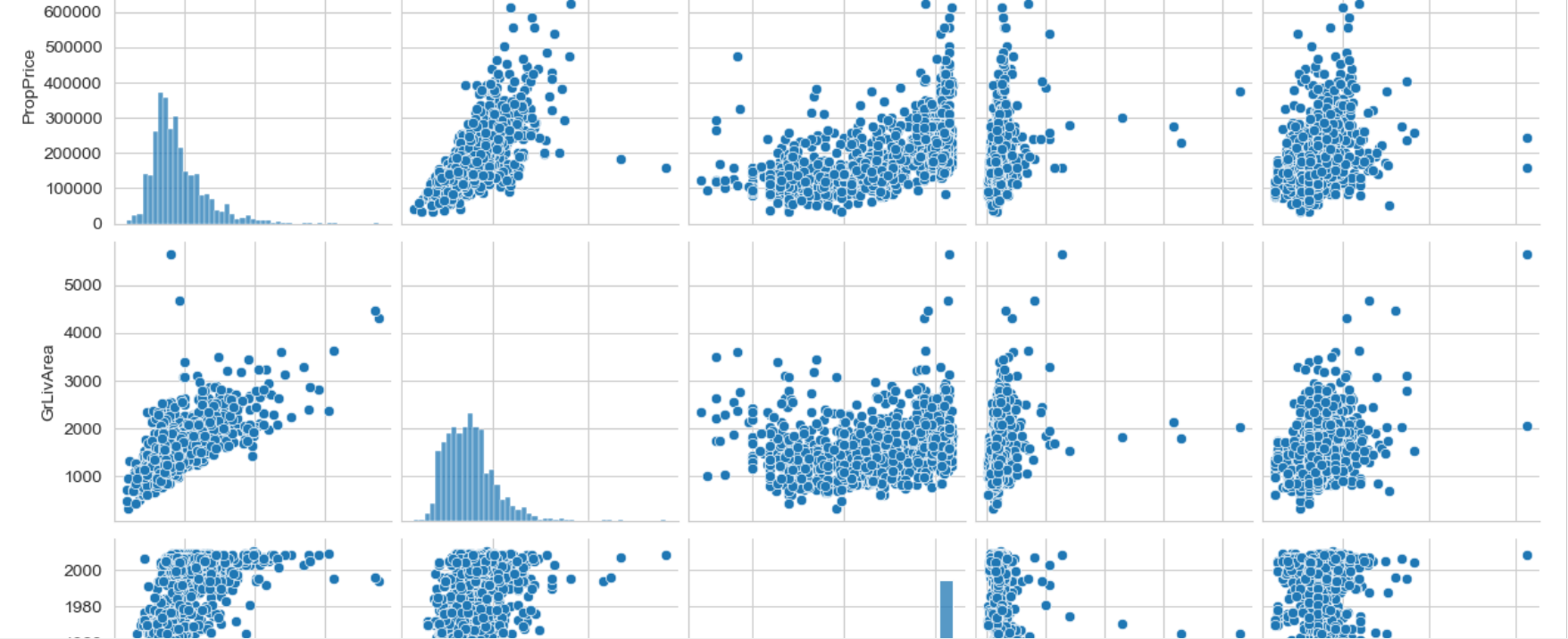
* Hyperparameter tuning results
* Best Parameters: {'max\_depth': 20, 'min\_samples\_leaf': 2, 'min\_samples\_split': 2, 'n\_estimators': 300}

**6. Report on EDA (with Graphs)**

Key insights derived from EDA:

* **Correlation Analysis**: Identified strong correlations between property price and features like Neighborhood



* **Outlier Detection**: Used boxplots to identify anomalies in price distribution.
* 
* **Feature Relationships**: Pair plots and scatter plots visualized key interactions.
* 

**7. Learning Outcomes**

* Importance of handling categorical and numerical data separately.
* Effective ways to clean and preprocess real estate data.
* How to engineer new features for better predictive accuracy.
* How to compare machine learning models and select the best one.
* Techniques for improving model performance using hyperparameter tuning.
* Use scaling and Xgboost

**8. Conclusion**

This project successfully developed a machine learning model like Random forest and Linear Regression capable of predicting property prices with an expected R^2 score of 84 % in case of Linear Regression And 89 % in case of Random Forest Regressor , Further Hyperparameter Tuning done and By using XGboost Model Performance Increased Further to 90 % . Insights from the study highlight the significance of key property features in price determination, assisting stakeholders in making informed decisions.

**9. Citations – Books and Websites Used for Research**

* Python Machine Learning – Sebastian Raschka
* Hands-On Machine Learning with Scikit-Learn
* Kaggle: Housing Price Prediction Datasets
* Scikit-learn Documentation ([https://scikit-learn.org](https://scikit-learn.org/))
* Seaborn & Matplotlib Documentation for data visualization

**Attachments:**

1. **Jupyter Notebook** – Capstoneproject1 Contains complete Python code and outputs
2. **Dataset File** – Property\_Data
3. **Project Summary Report (Word File)**