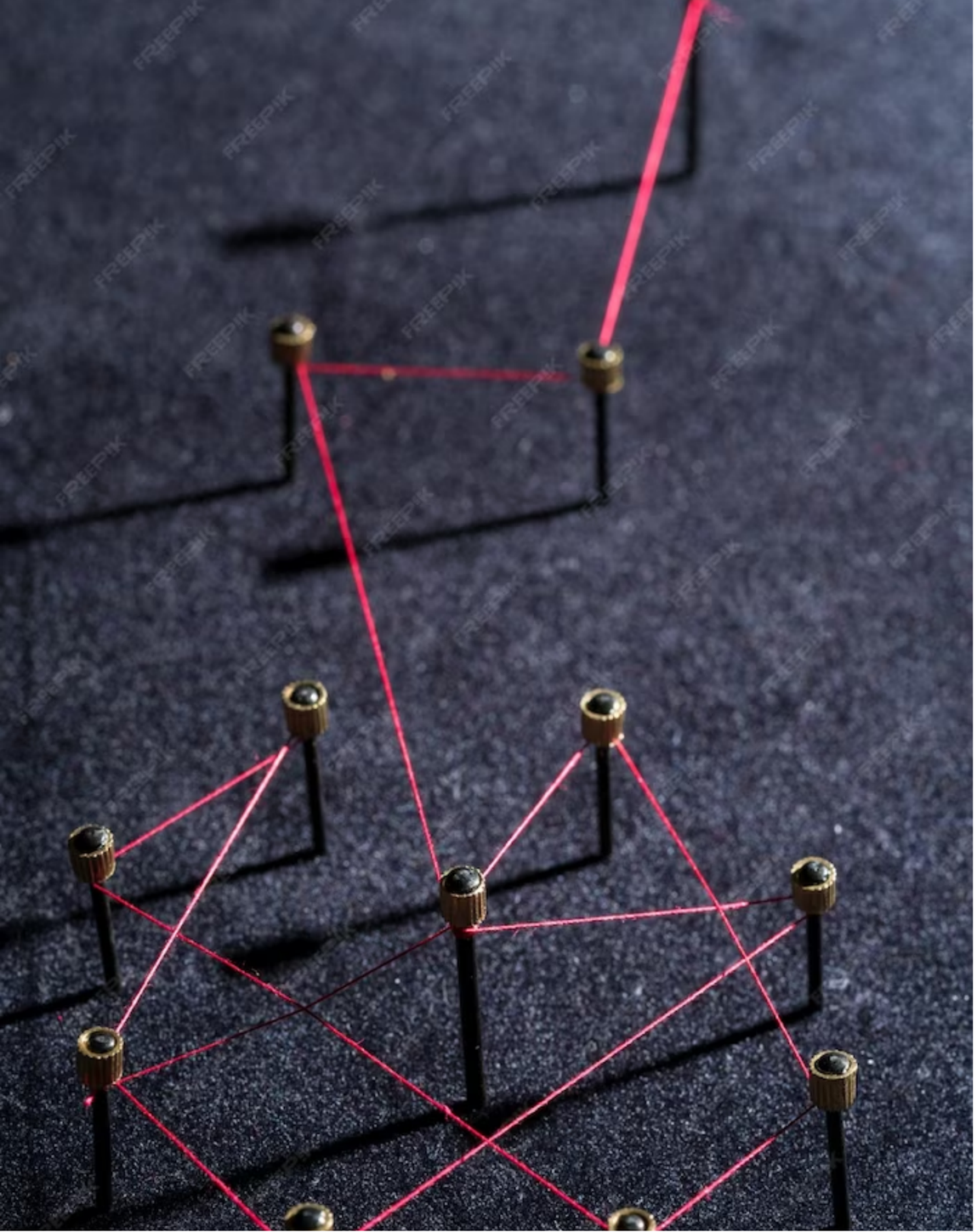
Unveiling Insights: Leveraging Linear Regression and K-Nearest Neighbors for Exploratory Data Analysis and Predictive Modeling of Real Estate Prices



Linear Regression

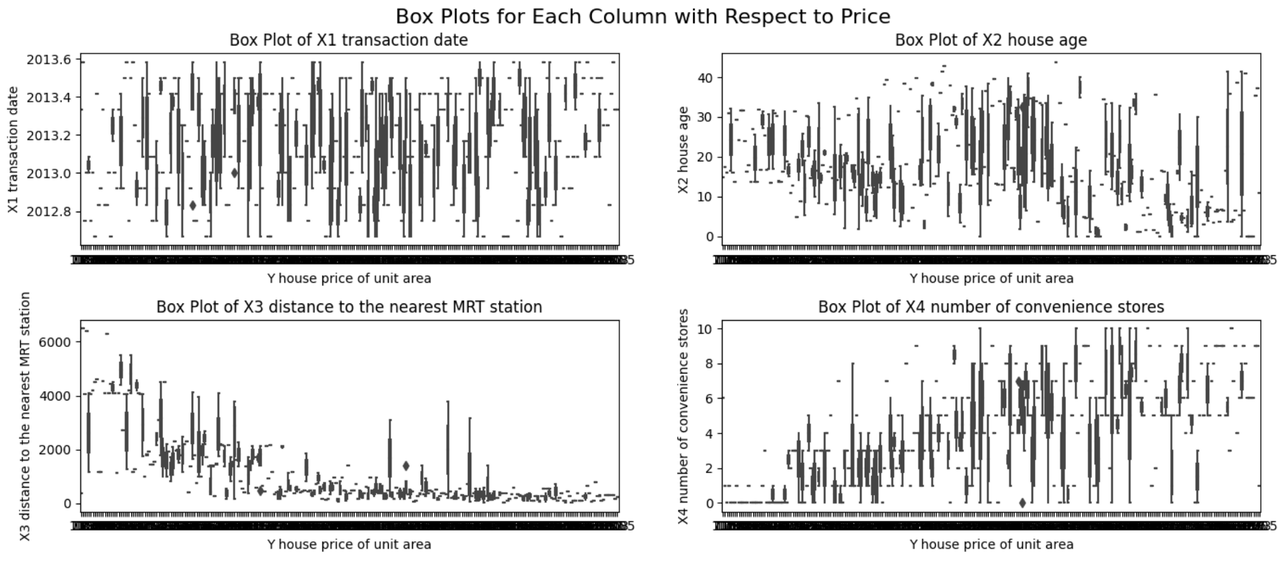
Linear Regression is a **supervised learning** algorithm used for predicting continuous values. By fitting a line to the data points, we can estimate the relationship between the independent variables and the dependent variable, i.e., real estate prices. We will explore the assumptions, model evaluation, and interpretation of results in the context of real estate price prediction.

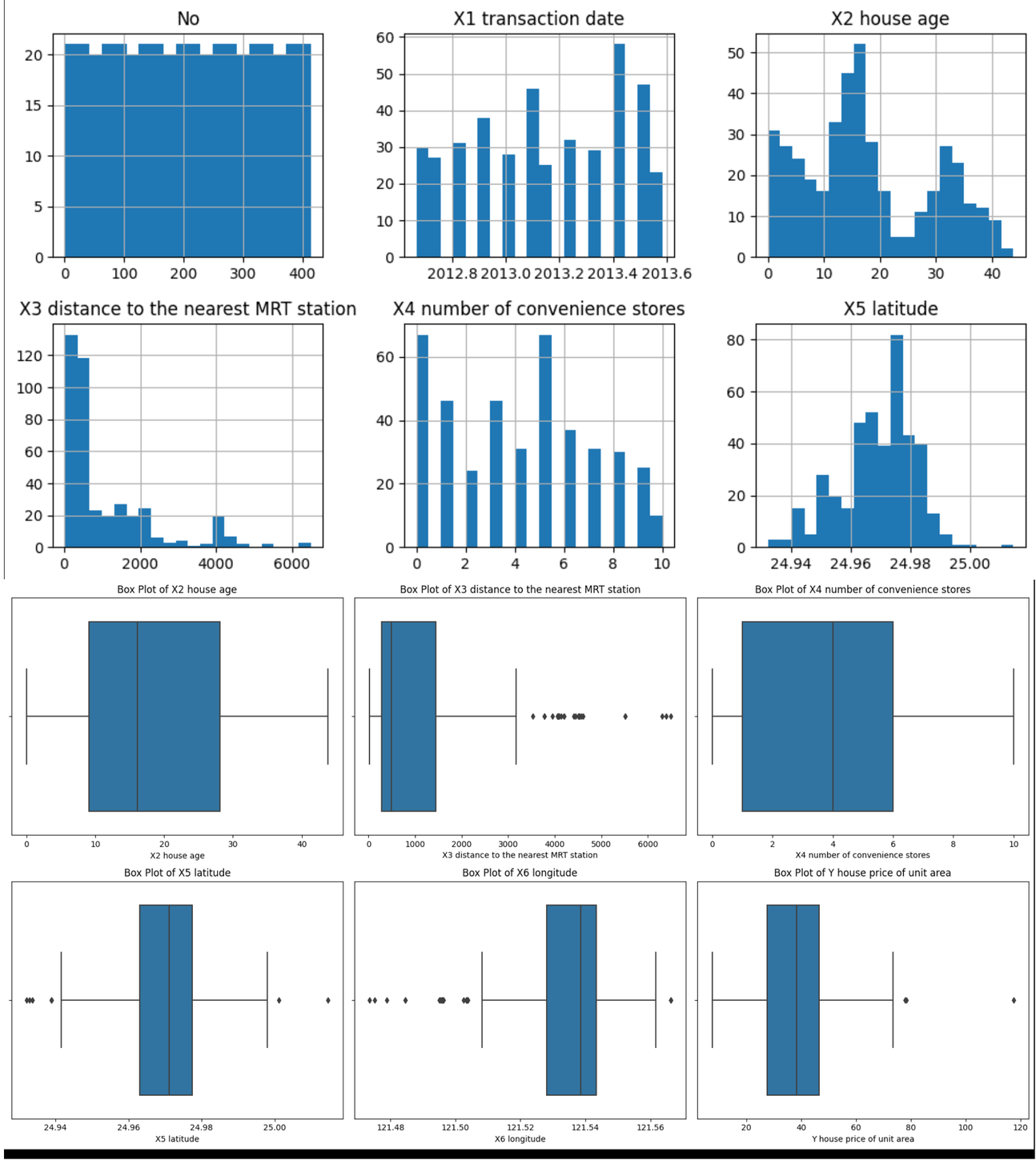
K-Nearest Neighbors

K-Nearest Neighbors is a versatile algorithm for both classification and regression tasks. It predicts the value of a data point based on the values of its k nearest neighbors. We will discuss the intuition behind KNN, the selection of k, and the impact of feature scaling. Additionally, we will explore how KNN can be applied to predict real estate prices.

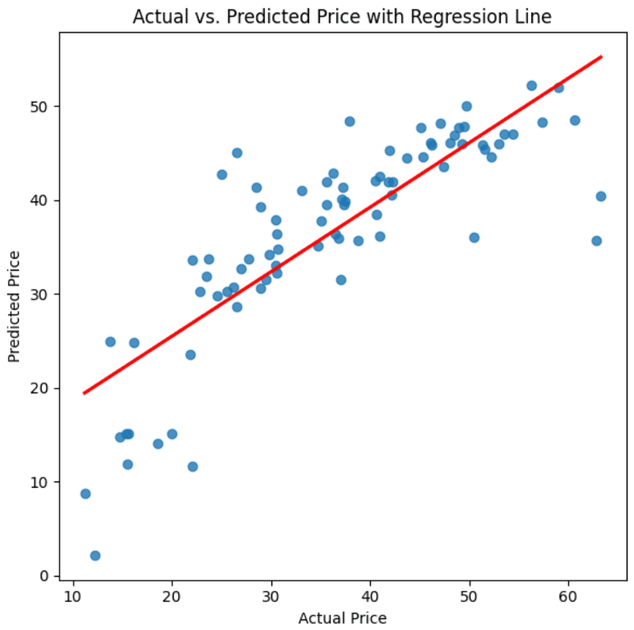
Exploratory Data Analysis

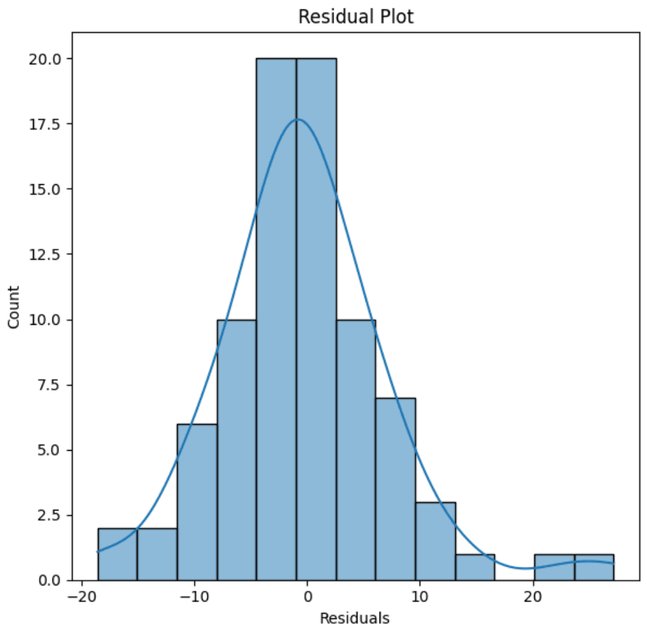
Before diving into predictive modeling, it is crucial to perform **exploratory data analysis**. This involves understanding the dataset, identifying patterns, and gaining insights into the factors influencing real estate prices. Through visualizations and statistical techniques, we can uncover valuable information that will guide our modeling efforts.

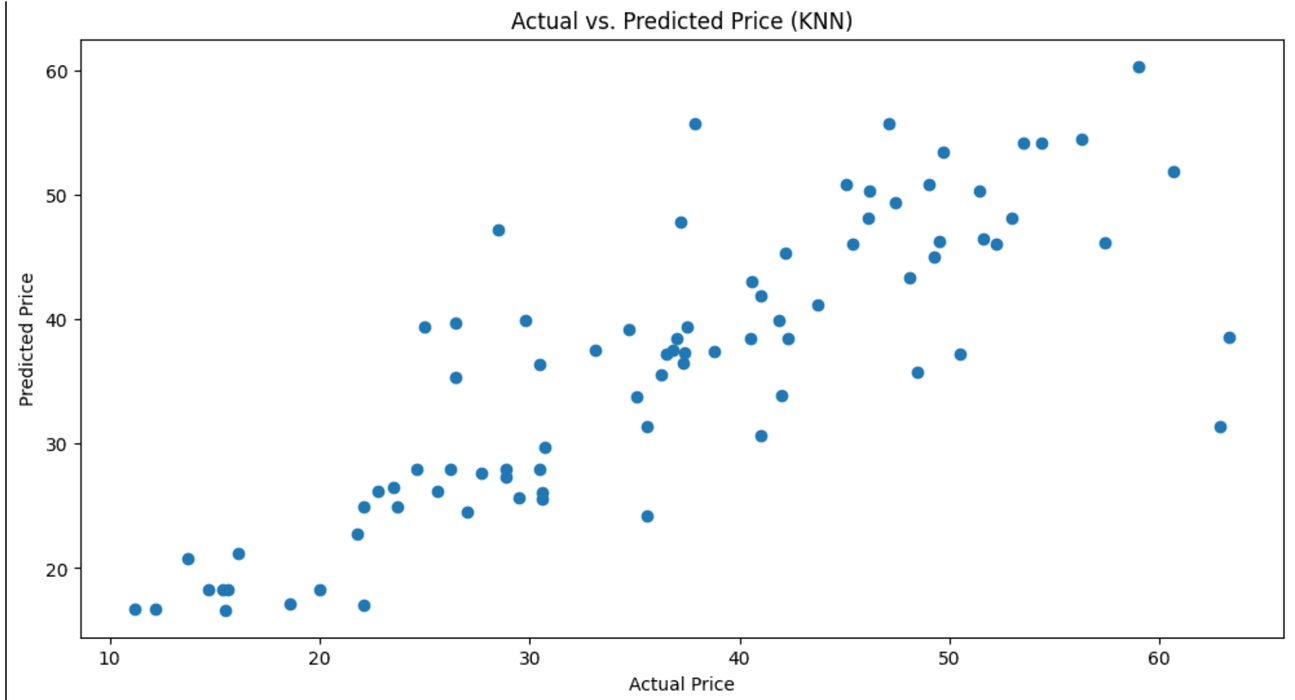


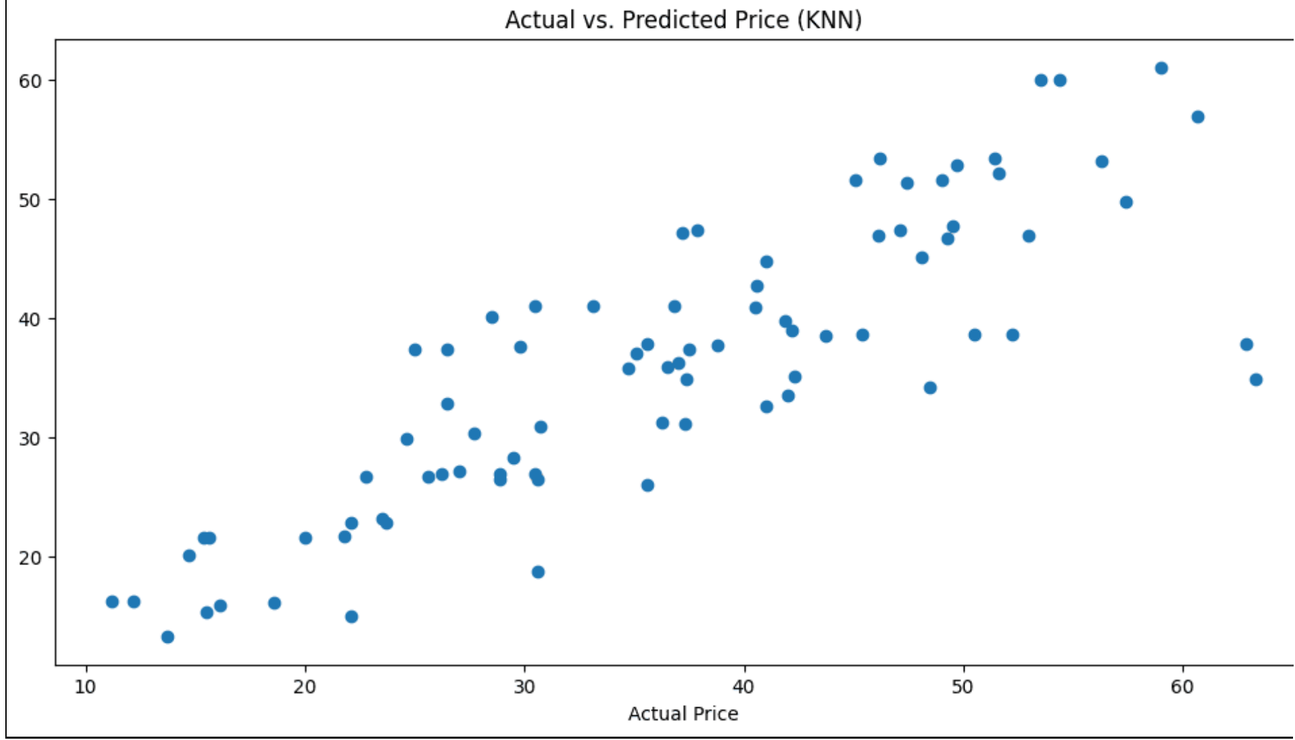


Linear Regression: The model coefficients indicate that house age, distance to the nearest MRT station, number of convenience stores, latitude, and longitude are significant factors in predicting house prices. The model achieved an R-squared value of 0.675, indicating moderate predictive performance.KNN Model: The KNN model with the best-performing hyperparameter (K=2) outperformed linear regression. It achieved an R-squared value of 0.704, explaining approximately 70.4% of the variance in house prices.Comparison: The KNN model with hyperparameter tuning demonstrated superior performance, highlighting the importance of hyperparameter optimization in machine learning.Conclusion: In this research, the tuned KNN model emerged as the best choice for predicting house prices, achieving an R-squared value of 0.704 and outperforming linear regression. Hyperparameter tuning played a pivotal role in improving model performance.

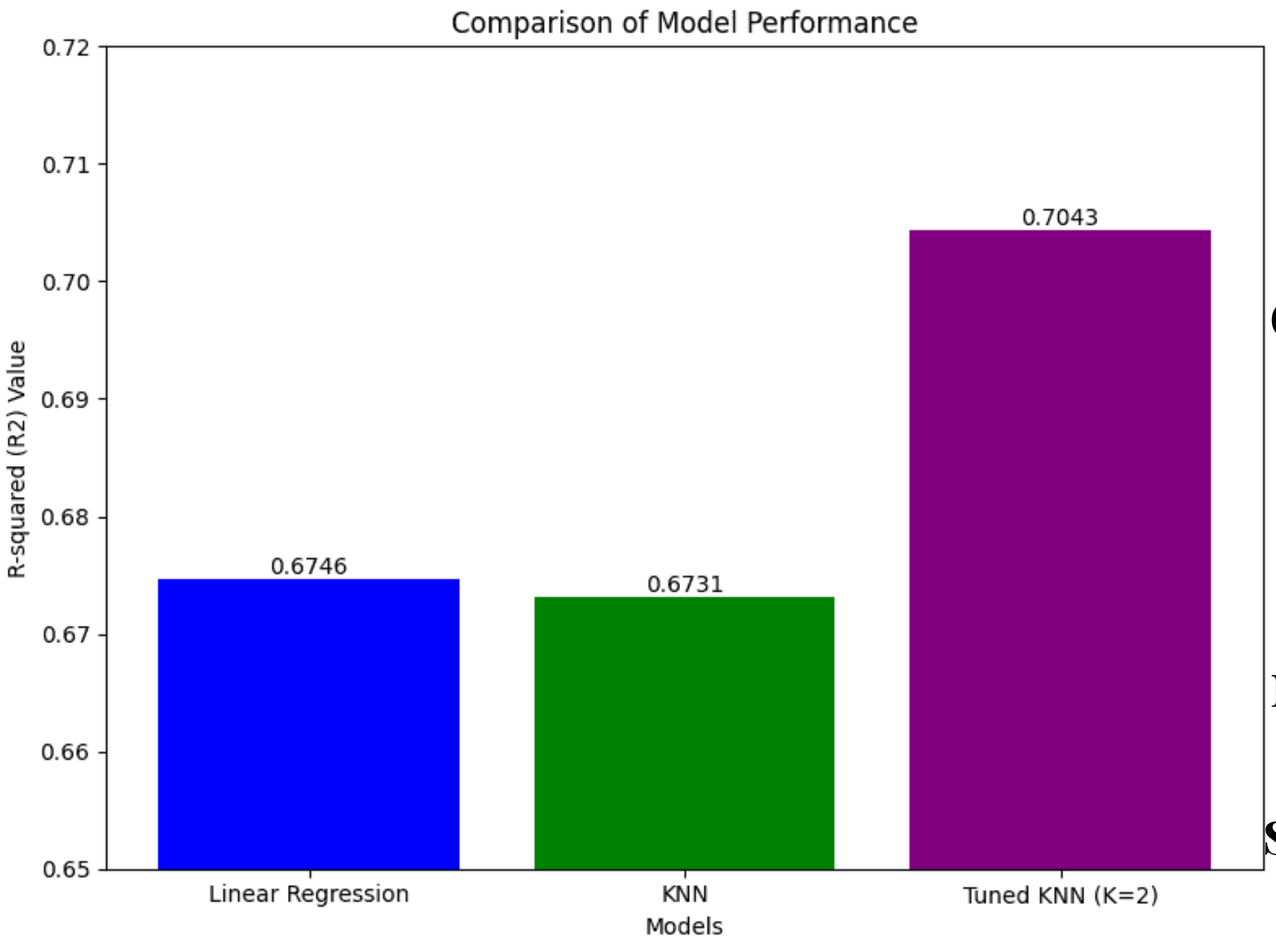








Model Comparison and Evaluation



Linear Regression and KNN: The linear regression model exhibited moderate predictive performance with an R-squared value of 0.675. However, the KNN model, especially with hyperparameter tuning (K=2), outperformed linear regression, achieving a higher R-squared value of 0.704. This highlights the significance of hyperparameter optimization in improving model accuracy. In summary, the KNN model with optimized parameters proved to be the superior choice for predicting house prices in this research.

Conclusion

In conclusion, our research successfully explored the factors influencing house prices in the real estate dataset. Through comprehensive analysis and model comparison, we found that the K-nearest neighbors (KNN) model, particularly with optimized parameters (K=2), outperformed linear regression. This highlights the importance of hyperparameter tuning in improving model accuracy. The absence of missing values in our dataset ensured the reliability of our results. Our findings can be applied to make informed decisions in the real estate market, benefiting buyers, sellers, and investors.