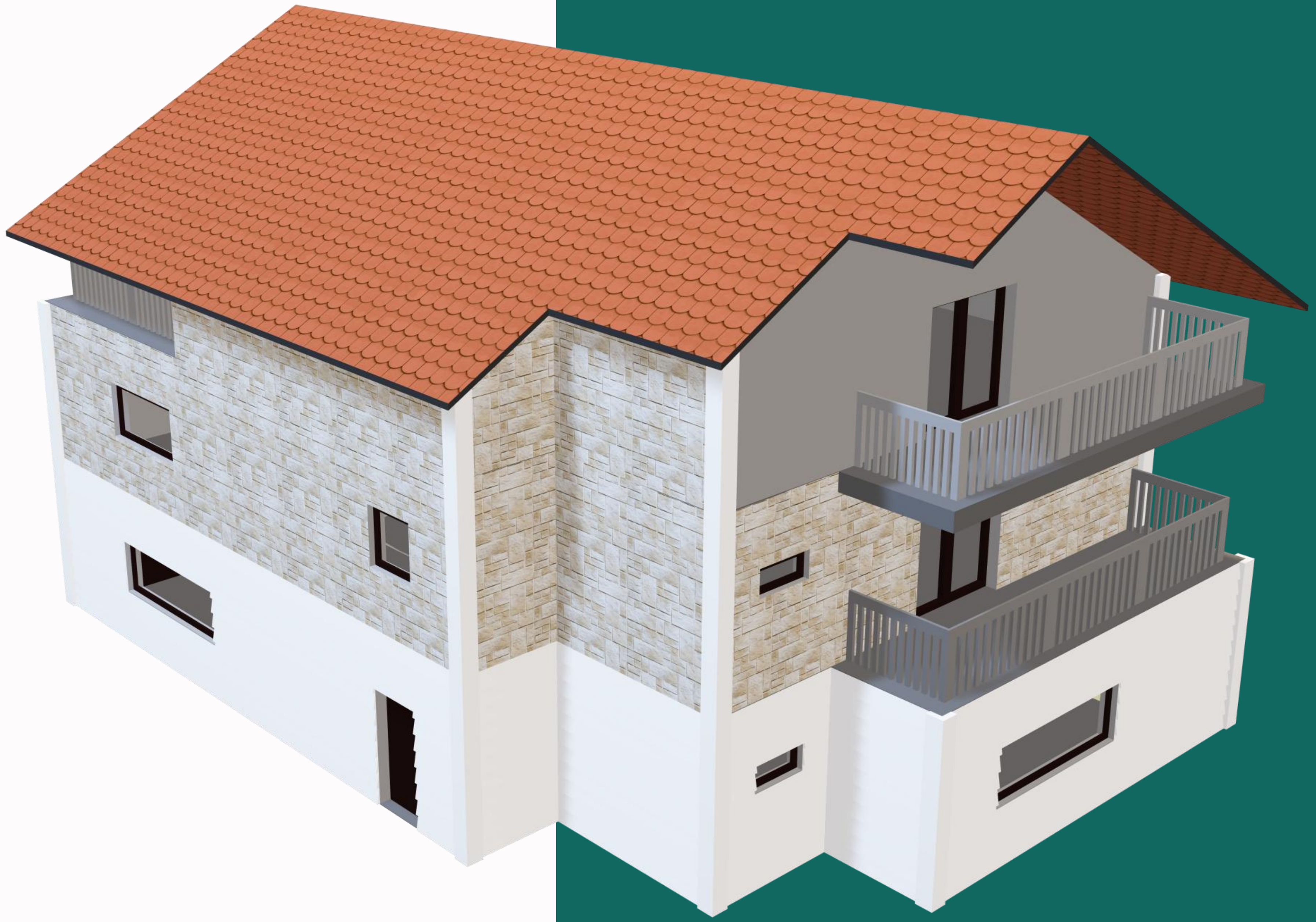
ITC AR-HP

Project: Augmented Reality House Price



Overview

Introduction

This project introduces an AR-based house price prediction model, combining K-Nearest Neighbors for spatial clustering and Linear Regression for accurate forecasting.

- By integrating location data (amenities, sqft) with property features (rooms, age, lot size), the model delivers real-time price predictions in an interactive AR environment.
- Built in Python, it empowers real estate professionals with intuitive, data-driven insights to support better investment decisions.

Problem Statement

- Real estate pricing is often complex and influenced by various factors
- Property Values fluctuate based on factors like location, size, features and market trends
- Booking Appointments

Why AR for Real Estate ?

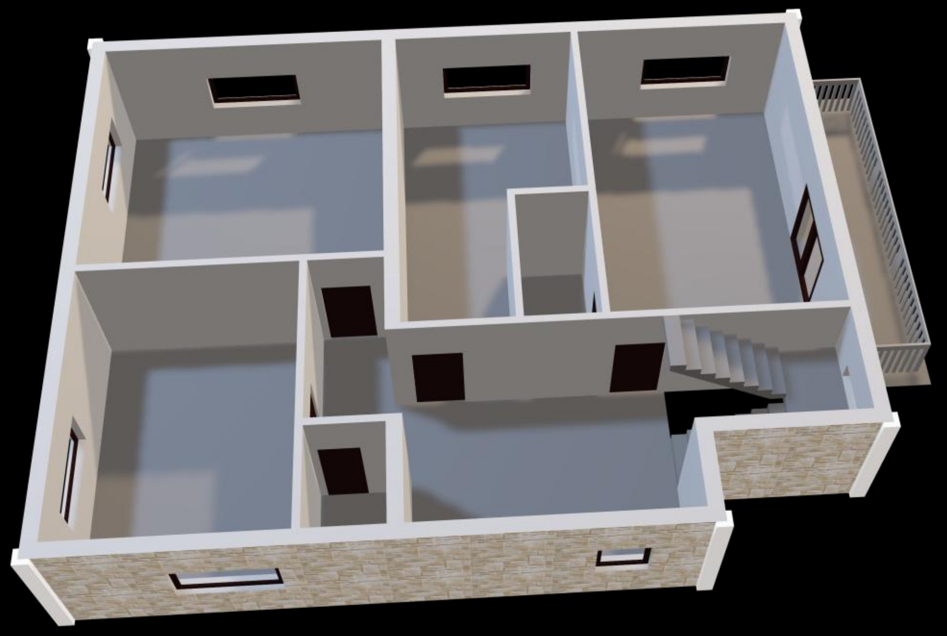
- AR allows users to view property features, trends and price predictions in real time.
- By combining AR with Prediction modeling, users can access instant insights about property values based on historical data.
- AR provides and interactive and engaging experience, making data more accessible and intuitive
- Blueprint without any additional tools or cost

$565,000 , 3bd + 4ba
85637, Alpha Road

This property has been sold, and the sale price is confidential.

$795,000 , 8bd + 5ba
85637, Beta Road

1st Floor



2nd Floor

3rd Floor

Project Objectives

Accurately Predict House Prices

Develop a hybrid model that combines K-Nearest Neighbors (**KNN**) for spatial analysis and **Linear Regression** for forecasting, enabling precise house price predictions based on both location and property-specific features.

Analyze Spatial and Property Based Factors

Identify and quantify the impact of various spatial factors (e.g., neighborhood amenities, proximity to schools, hospitals) and property attributes (e.g., property size, age, number of bedrooms) on housing prices across US.

Generate Actionable Insights for Stakeholders

Provide stakeholders with actionable insights by analyzing the contribution of each factor to house pricing, enabling informed decision-making in real estate investments, pricing strategies, and property valuation.

Scope of Work

This project involves building a comprehensive, data-driven solution for accurately forecasting house prices across the U.S. using custom machine learning models.

Key activities

- Include data preprocessing
- Feature engineering
- Model development and
- Evaluation

The final model aims to support stakeholders in making informed investment and pricing decisions

Data Collection and Processing

Collect and preprocess data on site location, property details, and demographics. Engineer features like price per square foot and room count to improve model accuracy.

Model Development and Custom Implementation

Build custom KNN for spatial analysis and Linear Regression for Price forecasting, implemented with statical model for a tailored solution.

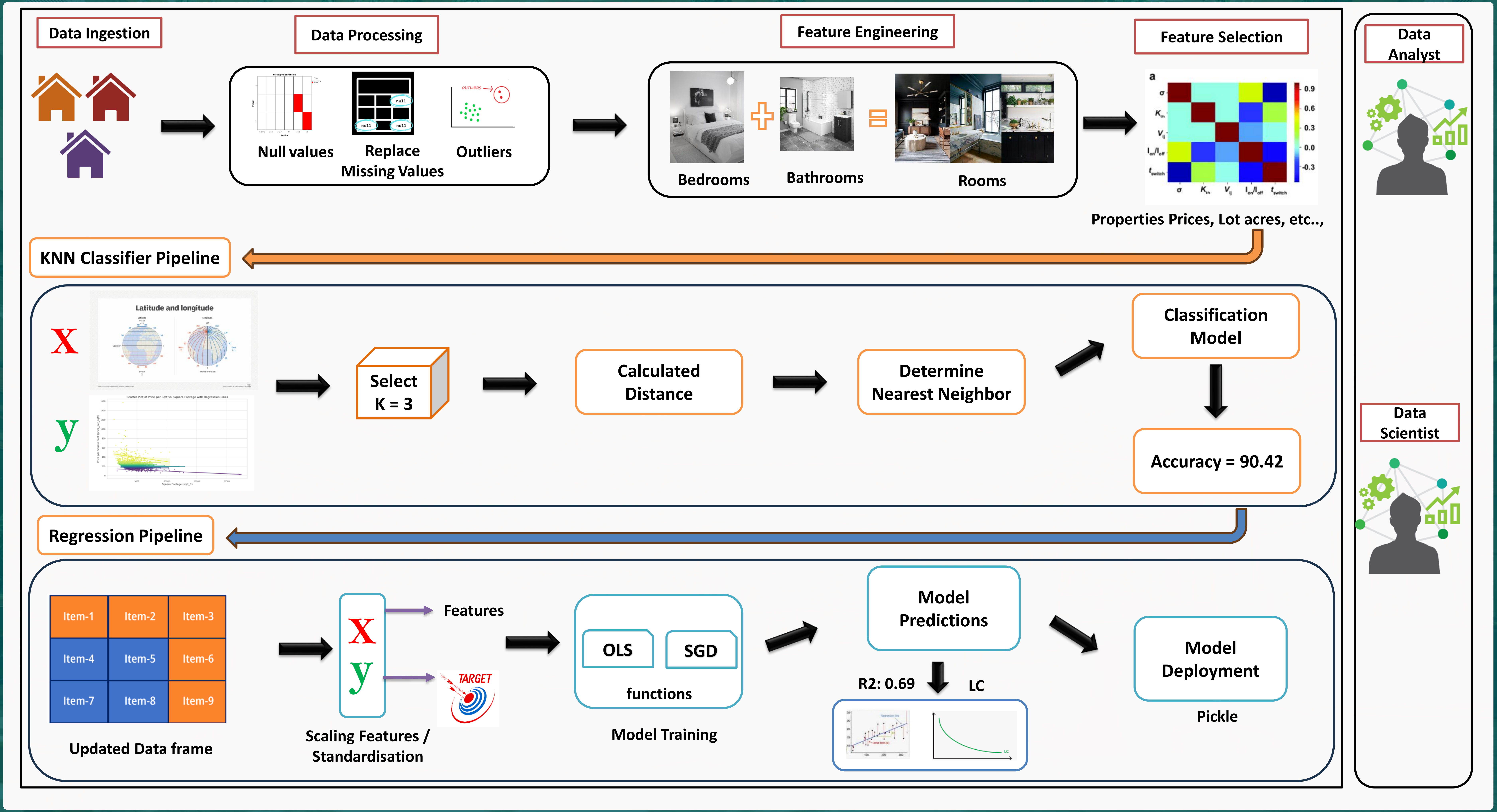
Model Evaluation and Visualization

Assess model accuracy with R^2 scores, analyze feature importance, and visualize predictions against actual values for insights on site performance.

Result Analysis and Reporting

Summarizing model accuracy, key insights, and recommendations, with visualizations to support stakeholder understanding and decision-making.

Architecture



Data Ingestion
Data Processing
Null values
Replace Missing Values
Outliers
Feature Engineering
Bedrooms
Bathrooms
Rooms
Feature Selection
Properties Prices, Lot acres, etc..,
KNN Classifier Pipeline
Latitude and longitude
x
y
Select K = 3
Calculated Distance
Determine Nearest Neighbor
Classification Model
Accuracy = 90.42
Regression Pipeline
Item-1
Item-2
Item-3
Item-4
Item-5
Item-6
Item-7
Item-8
Item-9
Updated Data frame
Features
TARGET
Scaling Features / Standardisation
OLS
SGD
functions
Model Training
Model Predictions
R2: 0.69
LC
Model Deployment
Pickle
Data Analyst
Data Scientist

Technical Part

Dataset Cleaning

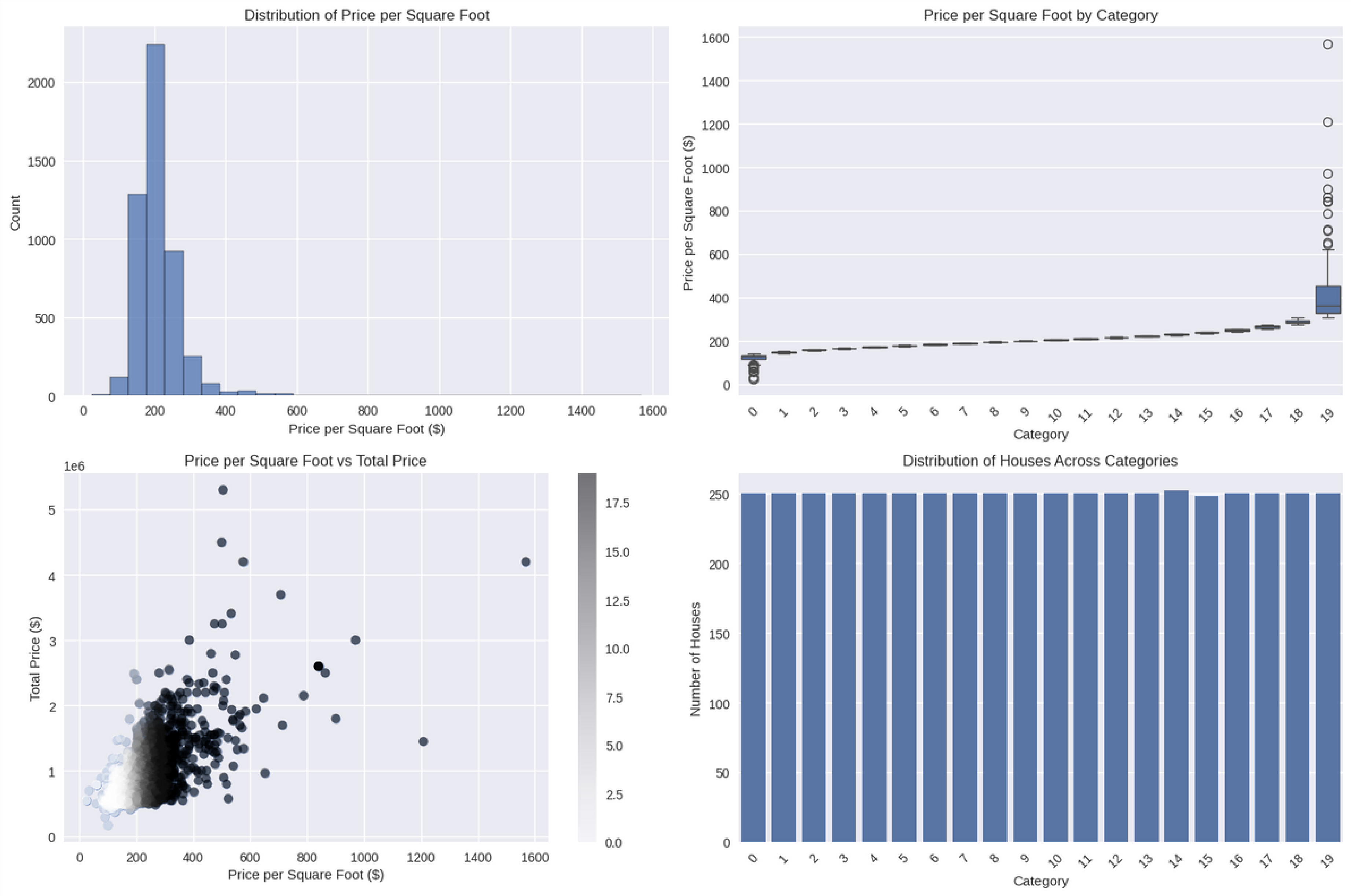
The dataset used for the base model training comprises of data shape (5000, 16) in which the columns are further explained.

Specific Steps considered were:

- Calculating price per square foot as a derived feature
- Creating 20 distinct price categories using quantile binning
- Engineering a combined 'rooms' feature from bedrooms and bathrooms

Distinct Price Categories

20 distinct price categories were creted using quantile binning:



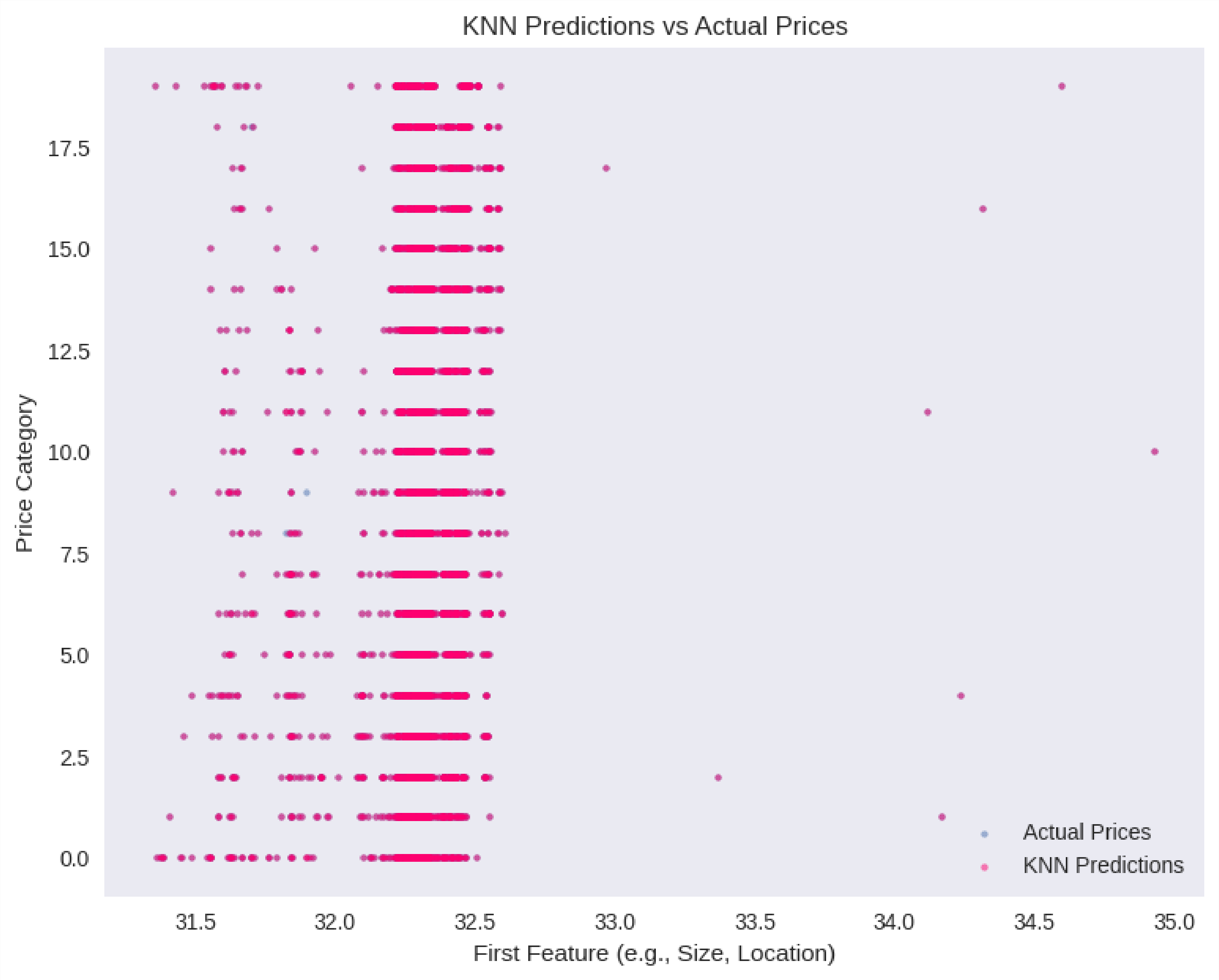
KNN Algorithm Implementation

Implemented a custom KNN algorithm for spatial price category prediction:

- Used latitude and longitude as features as **X0** and **X1.**
- Visualized predictions against actual values

The custom KNN algorithm predicts spatial price categories based on latitude and longitude by calculating Euclidean distances to training points, identifying the k nearest neighbors, determining the majority class among those neighbors, and storing the predicted categories, which are then added to the original DataFrame for comparison against actual values.

Visuals Interpretation



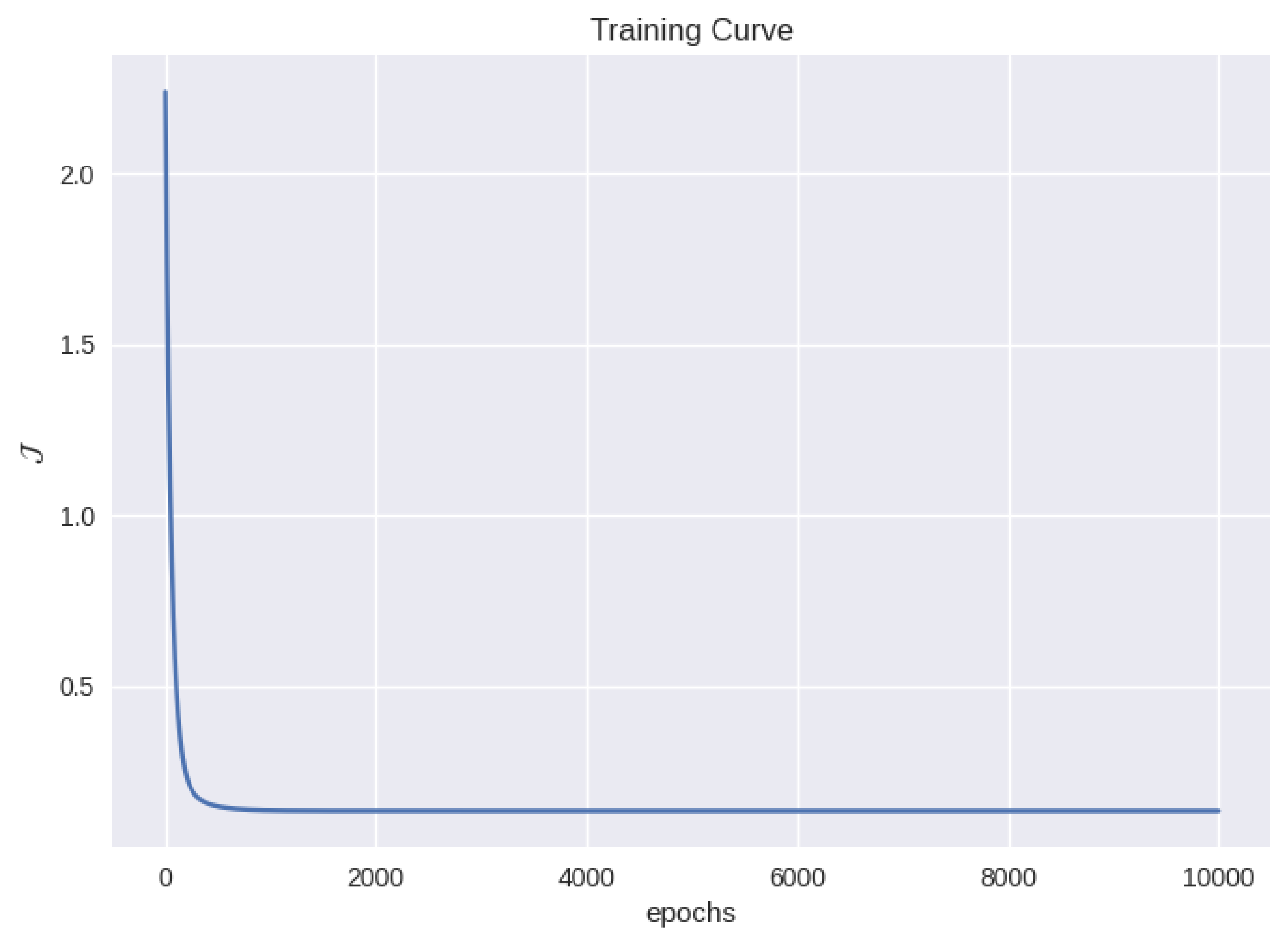
The plot shows the distribution of actual price categories against the predicted values for the first feature in X_spatial. With a 90% accuracy, the KNN model is performing very well, as indicated by how closely the y_pred_knn points align with y_knn. This visualization helps you see how well the KNN model captures the actual pricing trend in relation to the spatial feature.

Linear Regressor Implementation

The OLS algorithm implements a multivariable linear regression model to predict house prices using features such as KNN-derived price categories, square footage, lot size, number of fireplaces, garage presence, and number of rooms, by standardizing both the features and the target variable to improve gradient descent performance; it includes an Ordinary Least Squares function to compute the cost, utilizes stochastic gradient descent for weight optimization, evaluates the model's performance with an **R-squared score of 0.7287**, calculates feature importance to identify significant predictors, and determines the Mean Absolute Error to quantify the average prediction error in actual prices.

Linear Regressor Training Curve

```
R-squared score: 0.7287
knn_price_category importance: 0.6094
sqrt_ft importance: 0.7330
lot_acres importance: 0.2384
fireplaces importance: 0.0760
garage importance: 0.0419
rooms importance: 0.2001
```



Interpretation

R-squared Score: 0.7287

This indicates that 72.87% of the variance in property prices is explained by the model, suggesting a strong fit.

Feature Importances:

- sqrt_ft: 73.30% ----- Critical determinant of property price.
- knn_price_category: 60.94% ---- Influential in capturing neighborhood pricing trends.
- rooms: 20.01% ------ Notable impact on property value.
- lot_acres: 23.84% ------ Moderate influence on pricing.
- fireplaces: 7.60% ------- Minimal effect on value.
- garage: 4.19% ----- Low influence on predicted price.

Statistics

Based on Accuracy, Results after Implementation

New Customers

Recurring Customers

Conclusion

The predictive model enhances your United States House Price Prediction case study by delivering data-driven insights into key factors influencing property prices, empowering real estate professionals and investors to make informed decisions on property valuation and investment opportunities.

By accurately analyzing location and property-specific data, the model improves the accuracy of price forecasts and provides a deeper understanding of the factors that drive housing market dynamics.

Summary

- Utilized Python with Pandas, NumPy, and Matplotlib libraries for data analysis and visualization Performed initial data processing including:
  - Calculating price per square foot as a derived feature
  - Creating 20 distinct price categories using quantile binning
  - Engineering a combined 'rooms' feature from bedrooms and bathrooms
- **Implemented a custom KNN algorithm for spatial price category prediction:**
  - Used latitude and longitude as features and Built from scratch without scikit-learn dependencies
  - Visualized predictions against actual values
- **Developed a custom Linear Regression model:**
  - Features: KNN price category, square footage, lot acres, fireplaces, garage, rooms
  - Implemented using the Normal Equation method and added bias terms and performed feature scaling.
- **Model Evaluation and Analysis:**
  - Calculated and reported R^2 scores for both training and test sets
  - Generated detailed regression equation with coefficients
  - Analysed feature importance based on standardized coefficients

THANK YOU