

House Price Prediction (Summary)

Project Overview:

This project focuses on predicting house prices using key property features such as:

- Square footage
- Number of bedrooms
- Location

We simulated a realistic dataset to remove the need for external file uploads or downloads. The process includes data preprocessing, model training, evaluation, and visualization.

What I Did:

1. Created a synthetic dataset with 500 entries representing homes with:
 - Random square footage between 500 and 3500
 - Bedroom count between 1 and 5
 - Locations categorized into 'Downtown', 'Suburb', or 'Countryside'
 - Prices calculated based on realistic market assumptions with added noise
2. Preprocessed the data:
 - Standardized numerical features
 - One-hot encoded categorical location data
3. Trained a Gradient Boosting Regression model to predict housing prices.
4. Evaluated the model using:
 - Mean Absolute Error (MAE)
 - Root Mean Squared Error (RMSE)
5. Visualized results with a scatter plot comparing actual vs predicted prices.

What I Learned:

- How to build and structure a machine learning pipeline using scikit-learn
- How to handle different types of features (numerical and categorical)
- The importance of feature scaling and encoding in preprocessing
- How to evaluate regression models using MAE and RMSE

- How to create and work with synthetic data when real data is not available

Next Steps:

- Experiment with other models like Linear Regression, Random Forest, or XGBoost
- Add more features like bathrooms, year built, or neighborhood score
- Perform hyperparameter tuning to optimize model performance
- Save and load trained models using joblib or pickle