

# Capstone Two – Predicting Housing Prices in Washington, D.C.

Gbatchin Kochoni

# Project Overview

- Goal: Predict residential property prices in Washington, D.C.
- Dataset: DC\_Properties.csv
- Audience: Homebuyers, real estate analysts, city planners
- Key Steps: Wrangling, EDA, Feature Engineering, Preprocessing, Modeling, Evaluation

# Dataset Overview

- Target Variable: PRICE
- Features: ROOMS, BEDRM, BATHRM, LANDAREA, WARD, HEAT, STRUCT
- ~150,000 records
- Cleaned: Removed duplicates, invalid prices, null rows

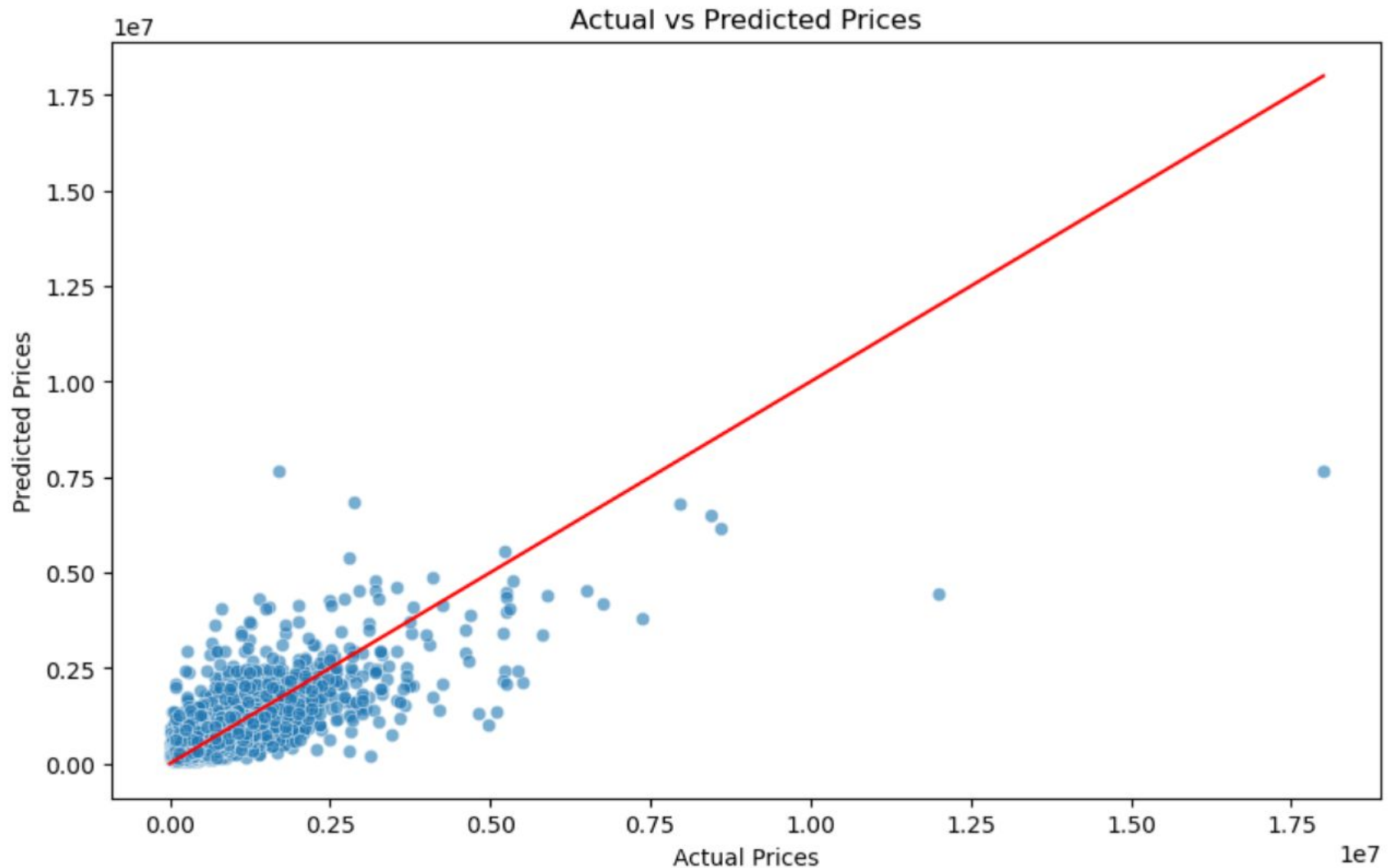
# Modeling Approach

- Baseline: Linear Regression
- Tree-based: Decision Tree (tuned), Random Forest (tuned)
- Evaluation Metrics: MAE, RMSE,  $R^2$
- GridSearchCV used for hyperparameter tuning

# Model Comparison (Test Set)

Model	MAE (\$)	RMSE (\$)	R <sup>2</sup>	Selected
Linear Regression	233,660.42	400,869.24	0.4954	No
Decision Tree (Tuned)	276,701.55	468,127.76	0.3119	No
Random Forest (Tuned)	225,166.54	376,057.71	0.5559	Yes

# Visualizing the model's performance



# Conclusion and Next Steps

- Random Forest Regressor selected for deployment
- Strong performance: Best MAE, RMSE,  $R^2$
- Next steps:
  - Add external features (crime, school data)
  - Test XGBoost, Gradient Boosting
  - Deploy as web or API