

ADVANCED HOUSE PRICE PREDICTION ANALYSIS

=====

Creating synthetic housing dataset...

=== ADVANCED DATA PREPROCESSING ===

Dataset shape: (1000, 11)

Missing values per column:

sqft_living 0

bedrooms 0

bathrooms 100

floors 0

waterfront 0

view 100

condition 0

grade 0

yr_built 100

zipcode 0

price 0

dtype: int64

=== FEATURE ENGINEERING ===

=== OUTLIER DETECTION===

Total unique outliers: 40

===MACHINE LEARNING PIPELINE===

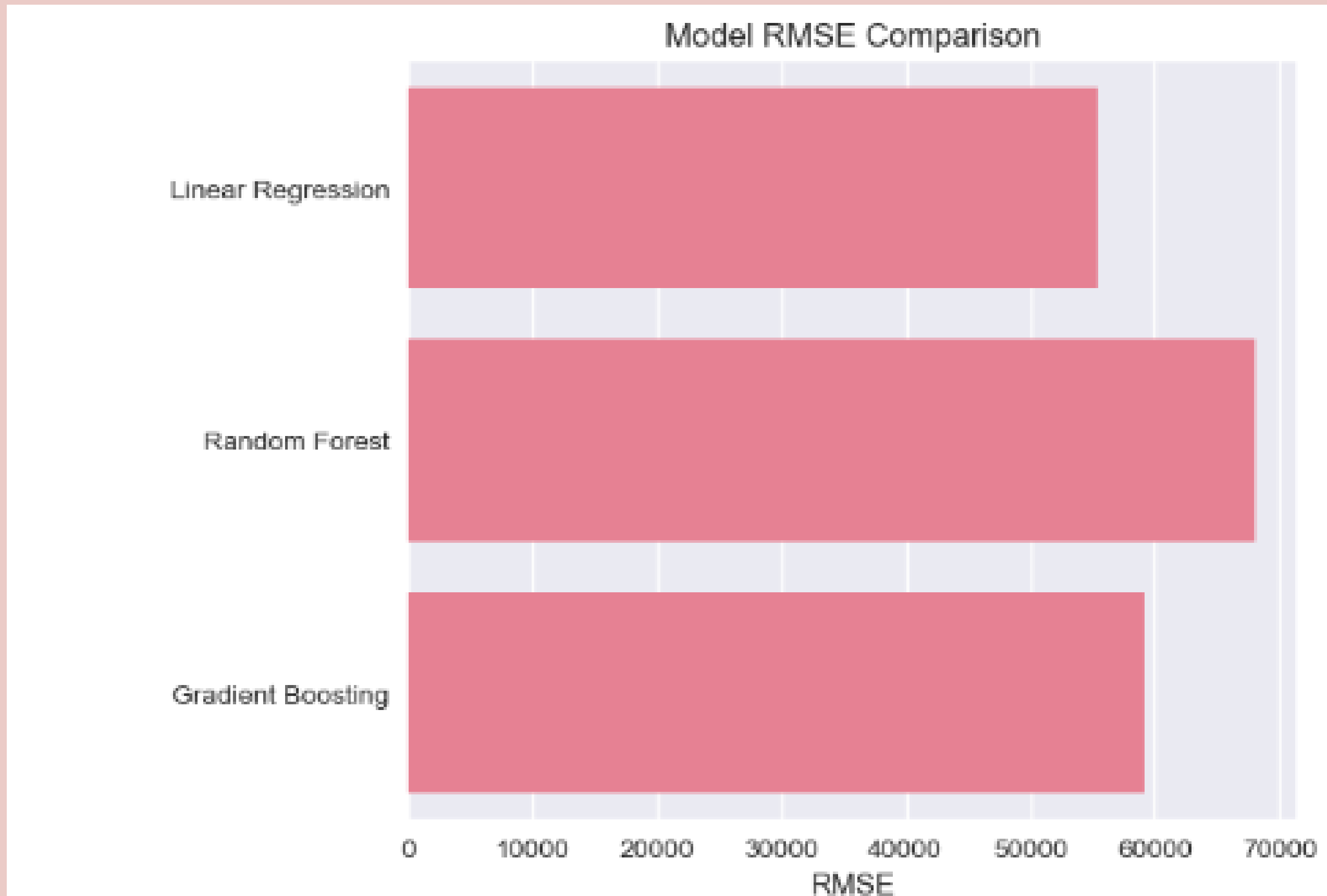
Training Linear Regression...

Training Random Forest...

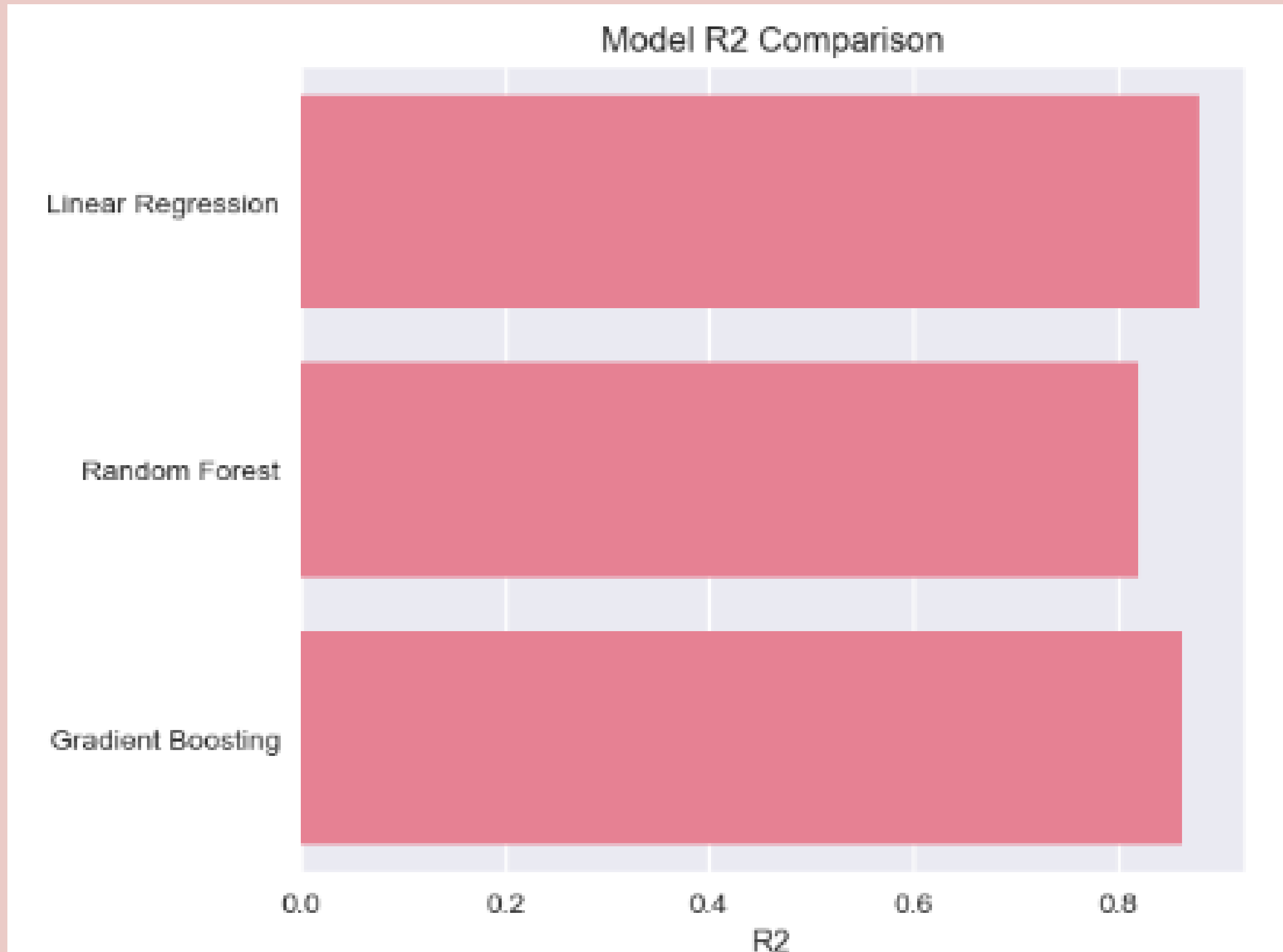
Training Gradient Boosting...

=== CREATING ADVANCED VISUALIZATIONS ===

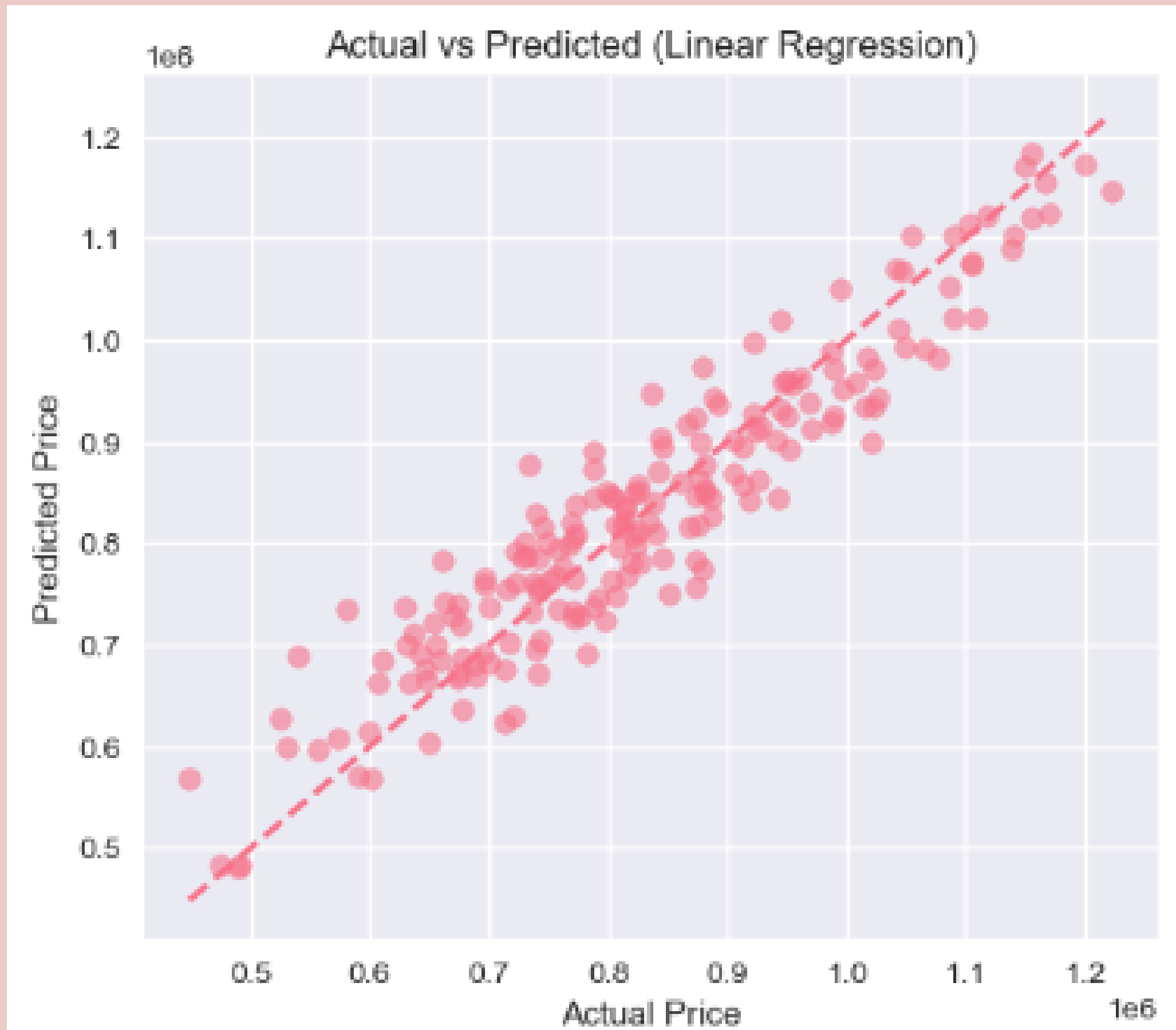
Model RMSE Comparison



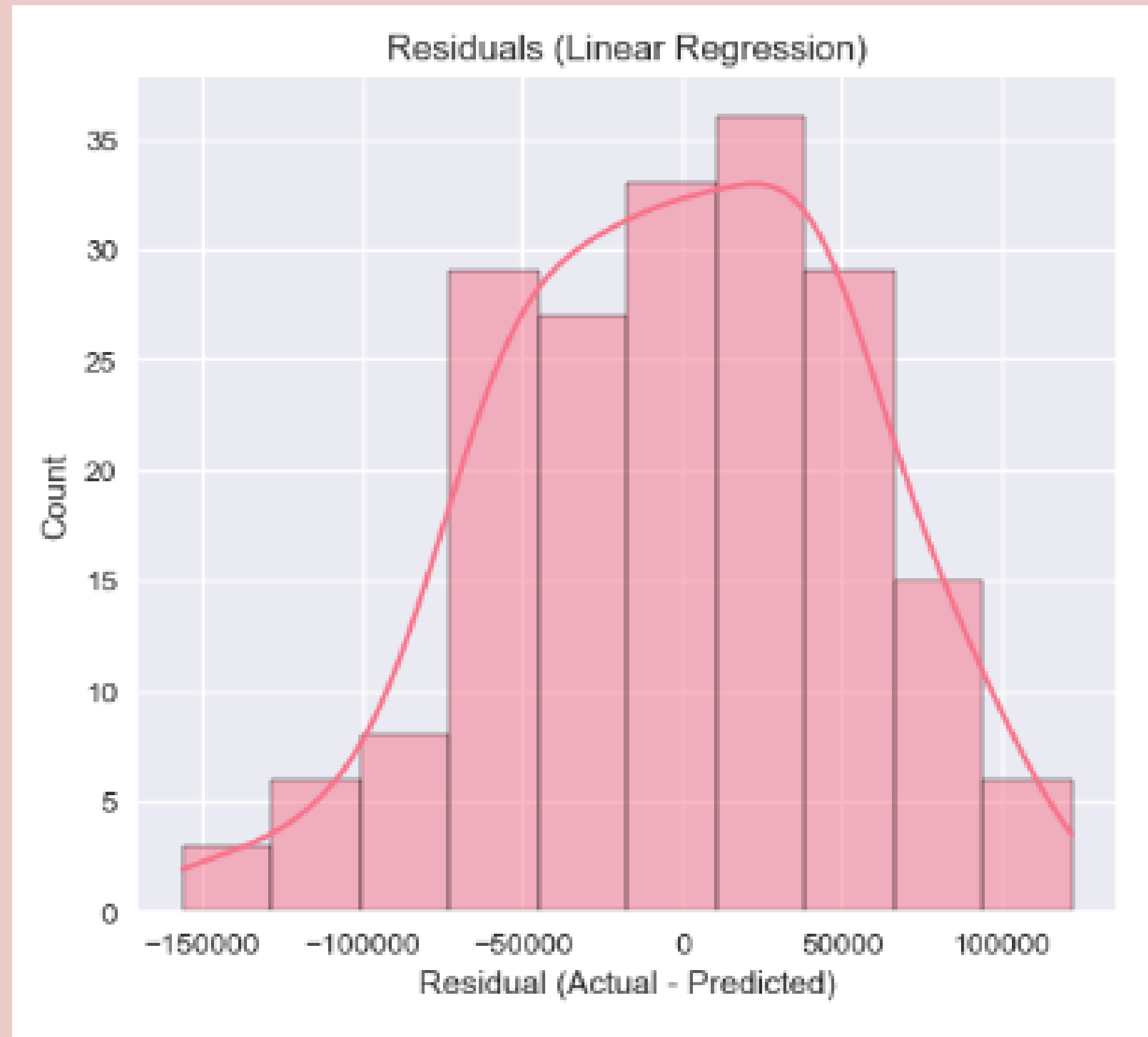
Model R² Comparison



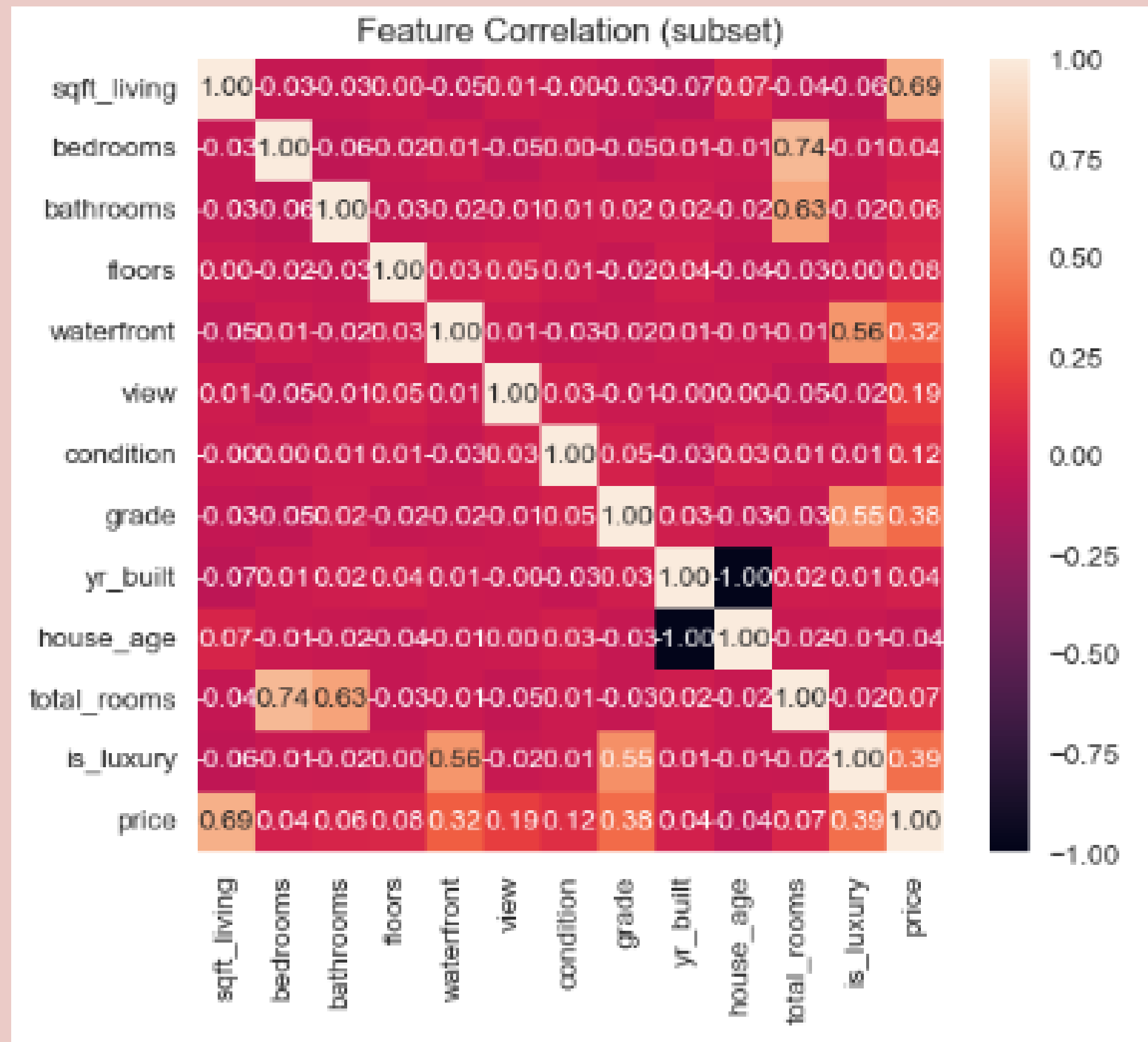
Actual vs Predicted Prices



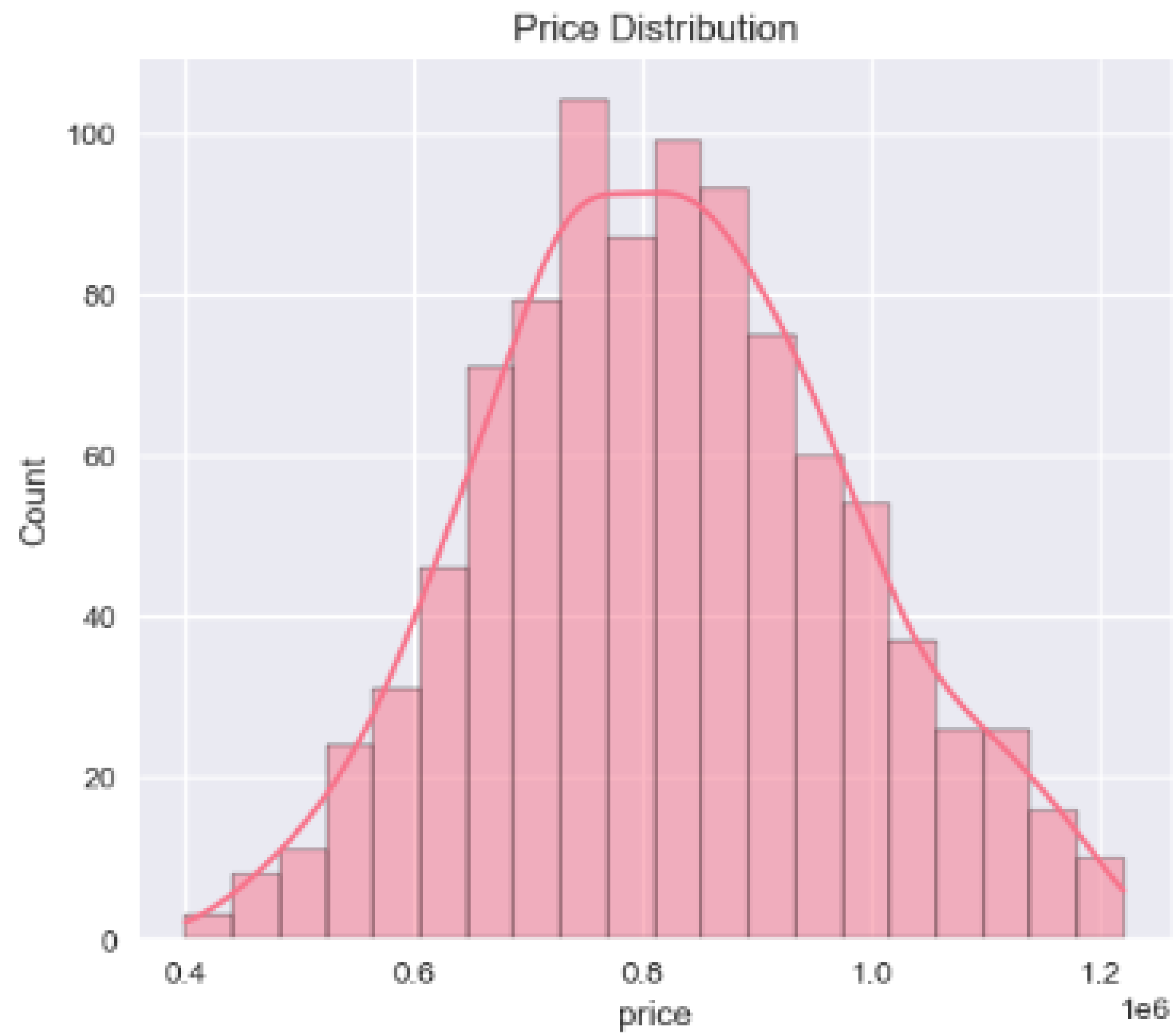
Residual Distribution



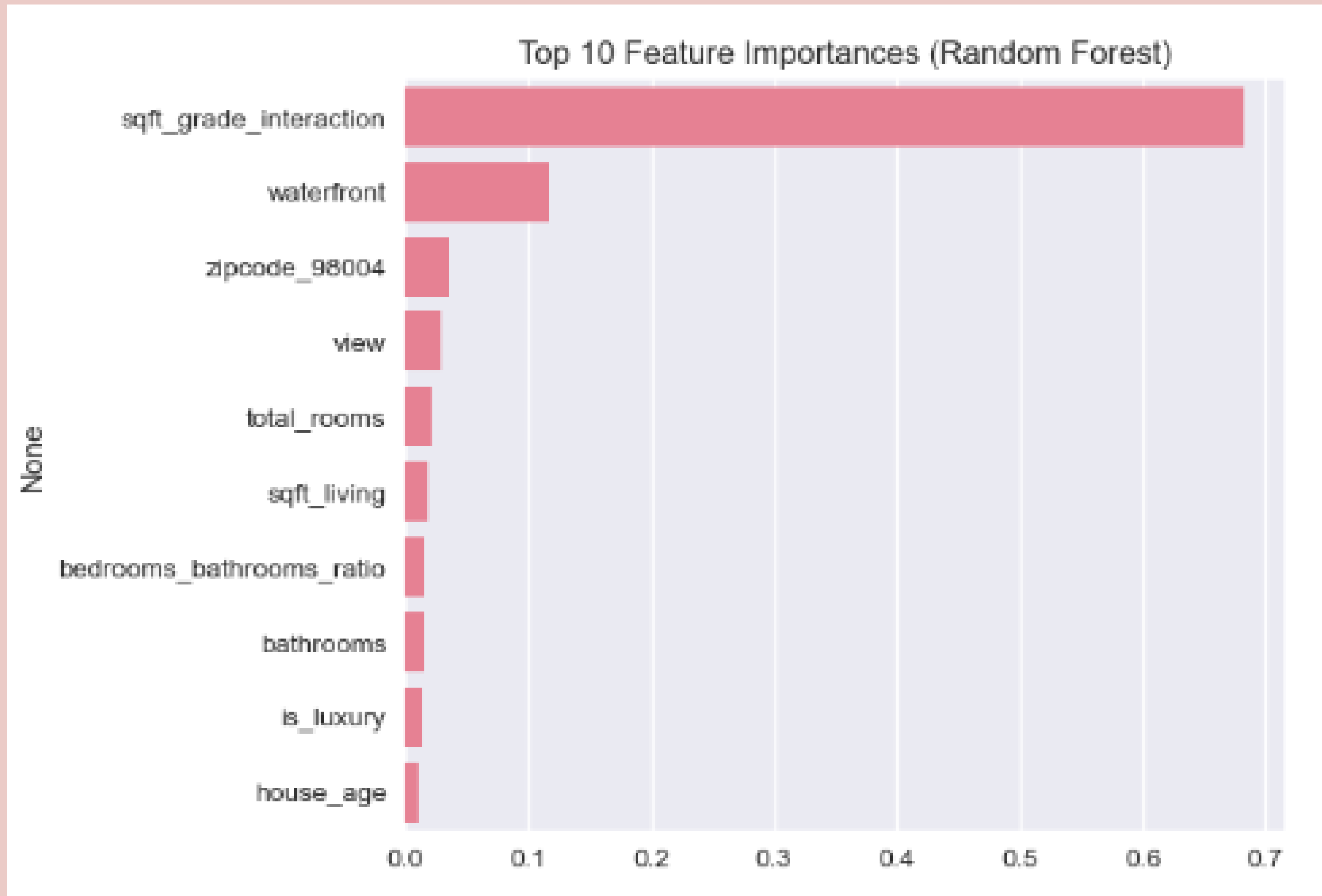
Feature Correlation Heatmap



Price Distribution



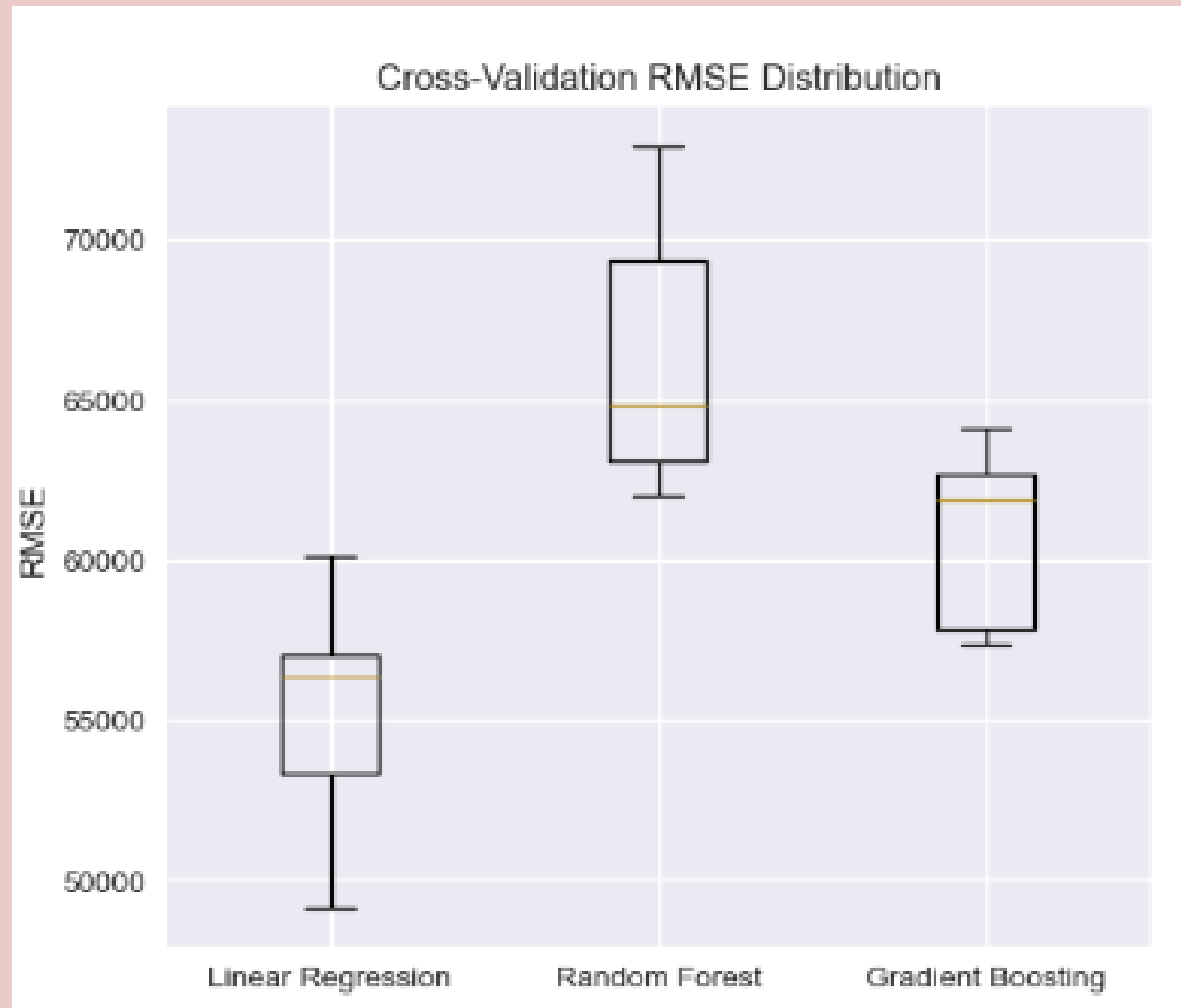
Top 10 Feature Importances (Random Forest)



Price vs. Square Footage



Cross-Validation RMSE Distribution



```
=====
FINAL RESULTS SUMMARY
=====
```

```
Linear Regression:
```

- RMSE: \$55,452.22
- MAE: \$45,526.63
- R^2 : 0.8804
- Best Parameters: {'poly__degree': 1, 'selector__k': 10}

```
Random Forest:
```

- RMSE: \$68,008.51
- MAE: \$55,045.75
- R^2 : 0.8200
- Best Parameters: {'max_depth': 10, 'min_samples_split': 5, 'n_estimators': 100}

```
Gradient Boosting:
```

- RMSE: \$59,102.51
- MAE: \$48,118.66
- R^2 : 0.8641
- Best Parameters: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}

```
BEST MODEL: Linear Regression
```

```
└─ RMSE: $55,452.22
```