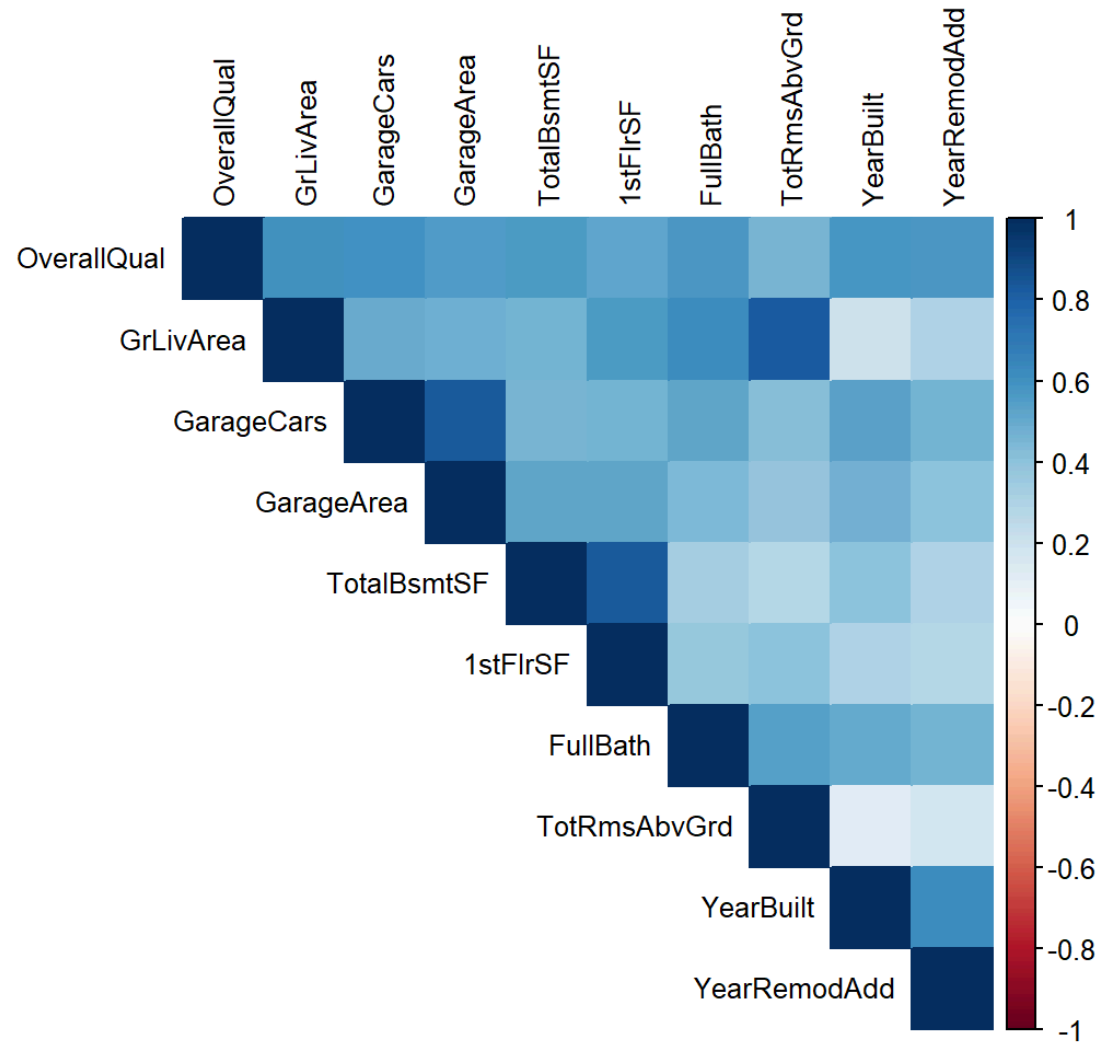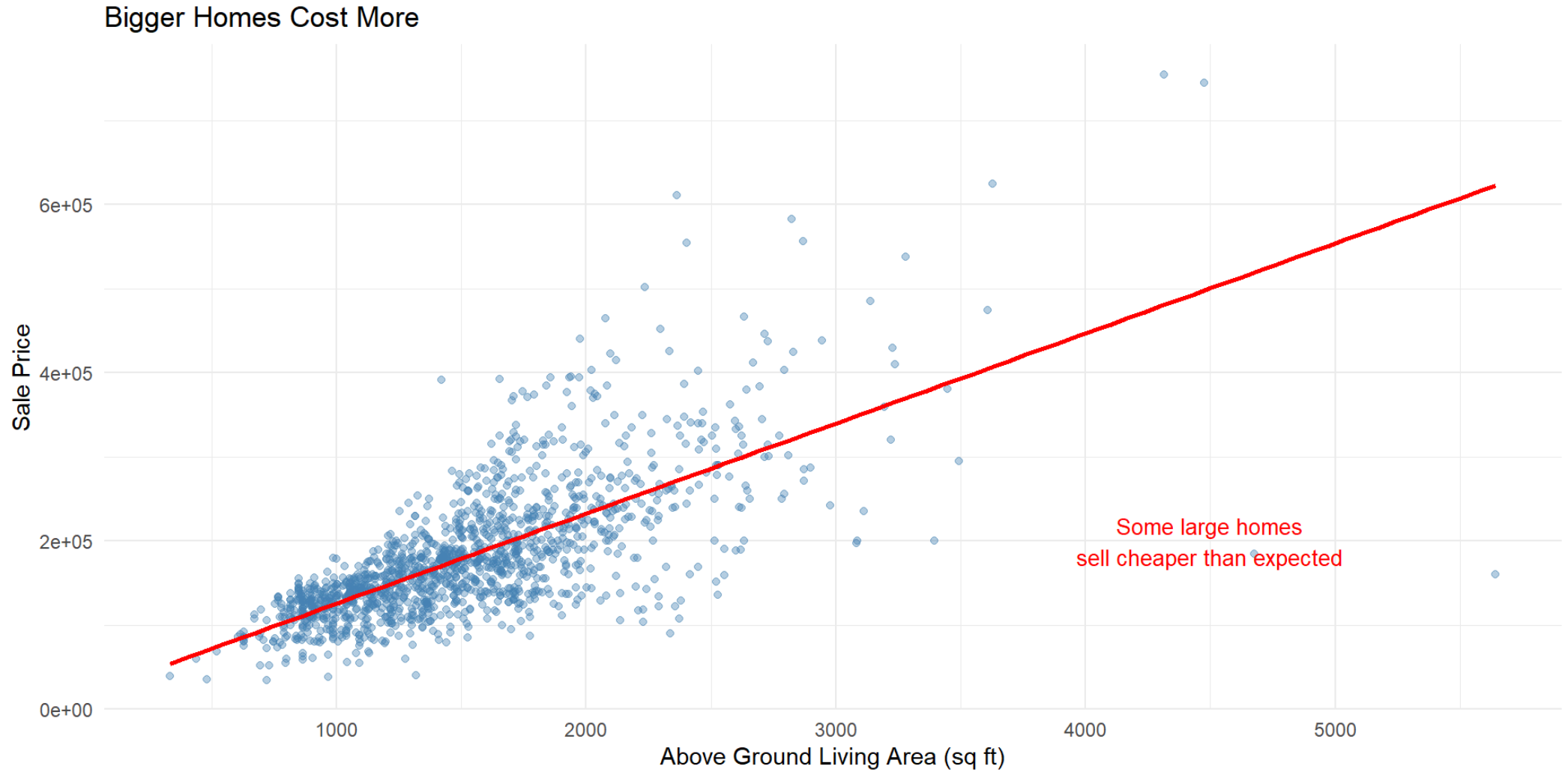# Story5_Data608

Mubashira Qari

2025-03-24

Defining Business Question

This week we will be building a model of the factors determining housing prices and using visualizations to explain their meaning and implications.

Sale Price Distribution & Top Correlations

Bigger Homes, Higher Prices

Bigger Homes Cost More

A Simple Yet Powerful Model Using Key Features

```
Call:
lm(formula = SalePrice ~ GrLivArea + OverallQual + GarageCars +
    TotalBath + Age, data = train_df)

Coefficients:
(Intercept)      GrLivArea    OverallQual    GarageCars     TotalBath           Age
  -68048.82          45.42       22967.62      16827.04       9040.93       -281.62
```
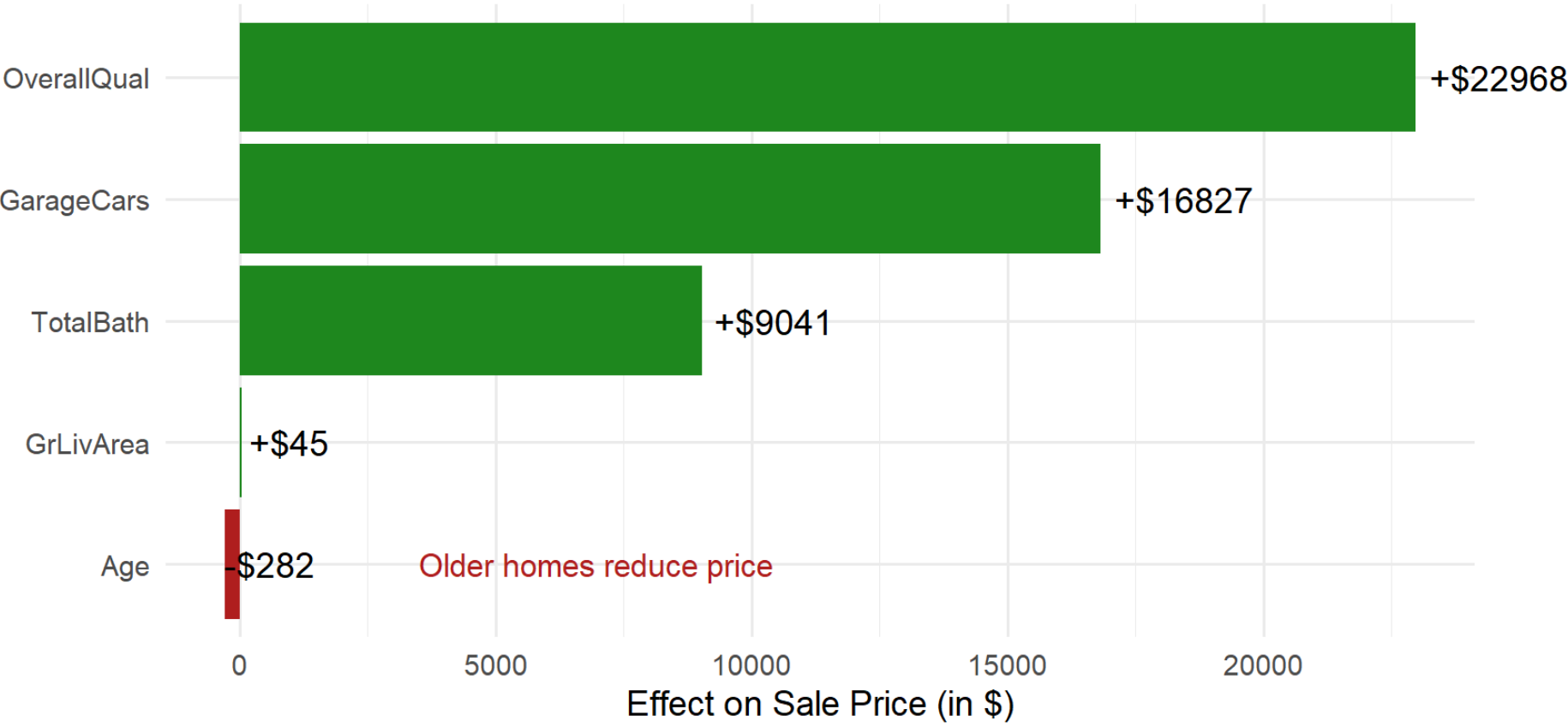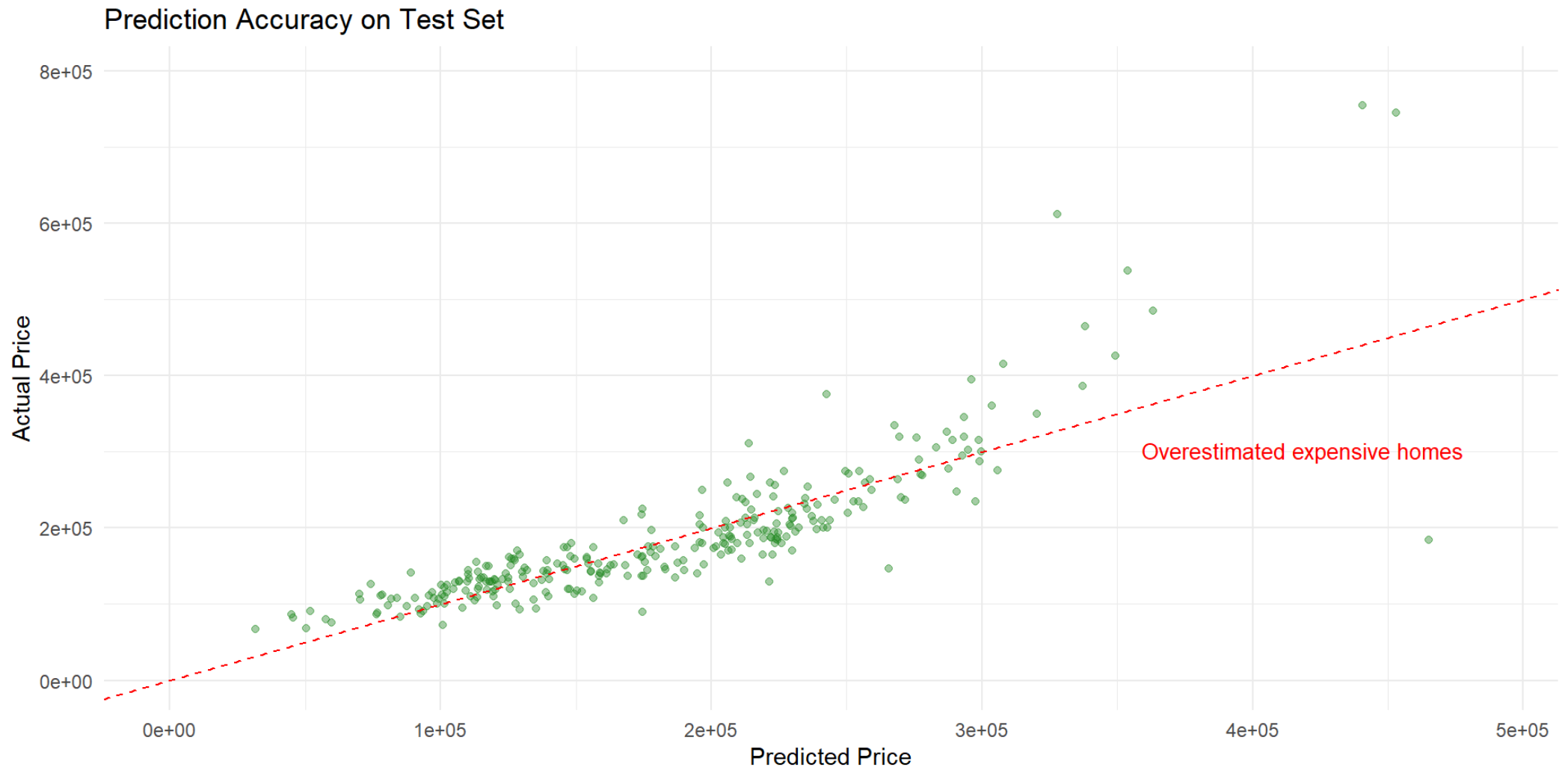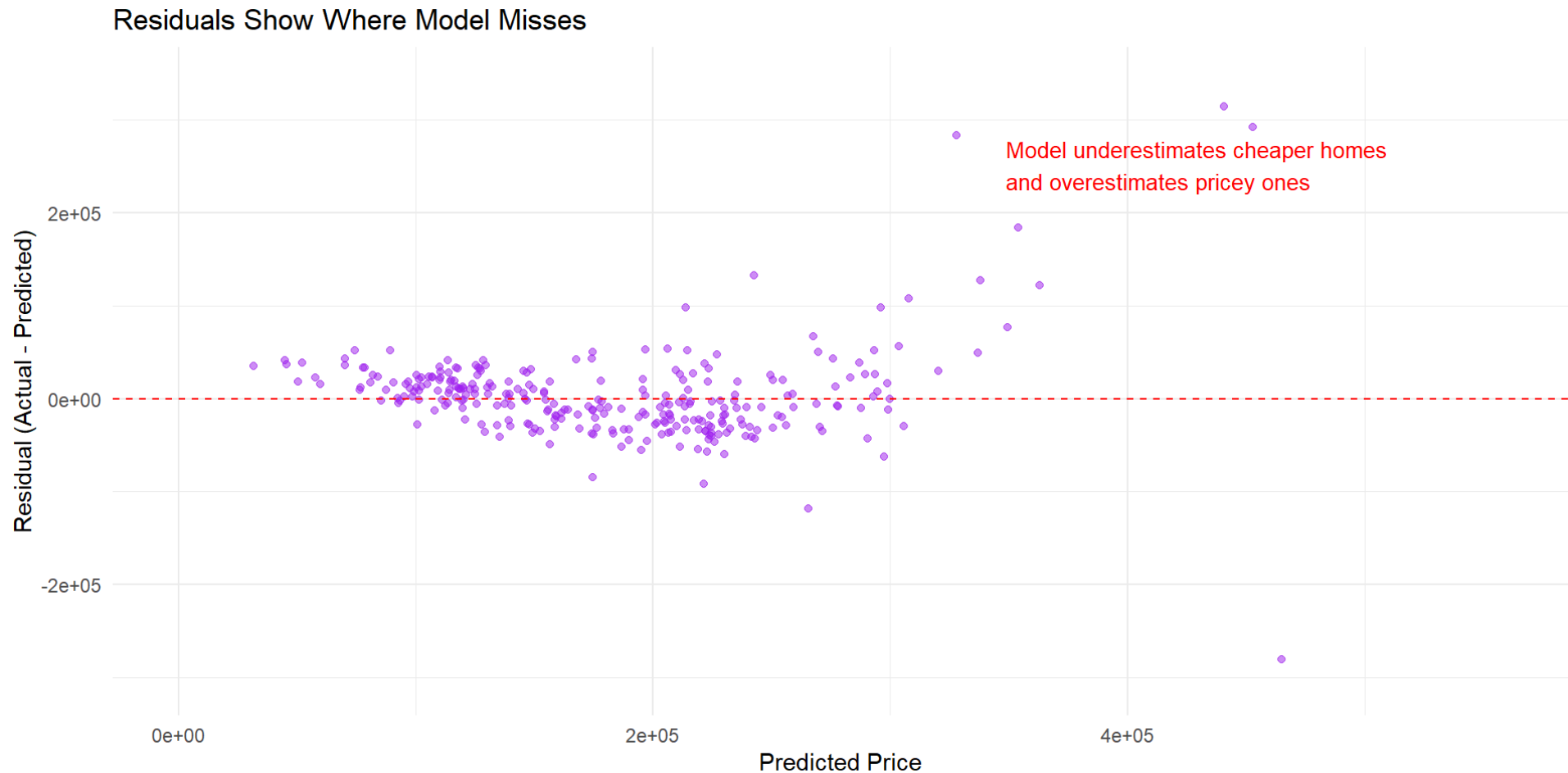
What Drives House Prices?

Quality matters most!

Impact of Each Variable on Sale Price (Based on Linear Model)

	Effect on Sale Price (in $)
OverallQual	+$22968
GarageCars	+$16827
TotalBath	+$9041
GrLivArea	+$45
Age	-$282 — Older homes reduce price

Predictions Closely Follow Actual Prices

Prediction Accuracy on Test Set

Where the Model Struggles: Prediction Errors

Residuals Show Where Model Misses

Which Factors Matter Most?

Top Features Driving Sale Price

Higher quality & garage space
boost prices most
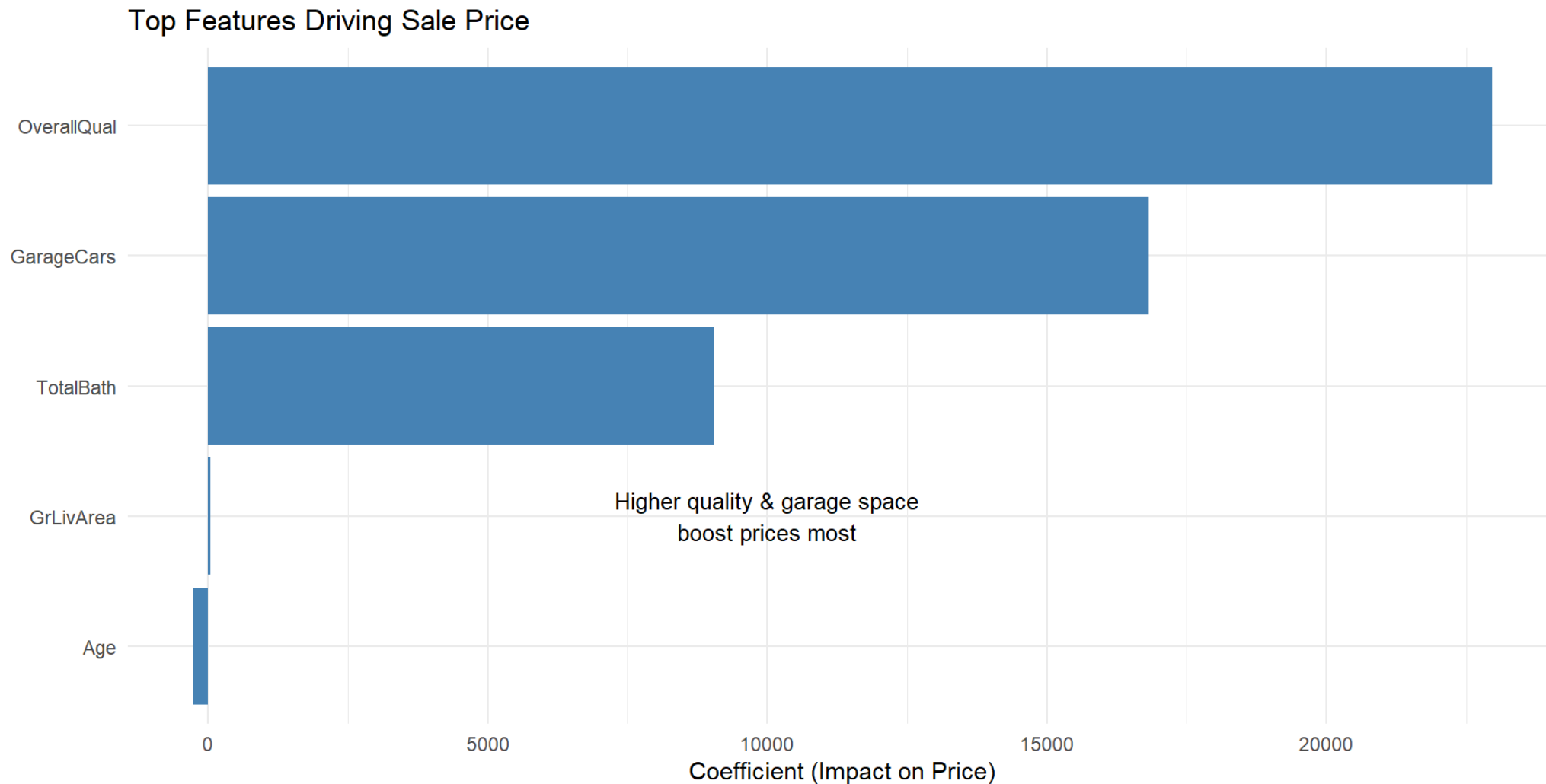
Model Explains 71% of the Variability in Housing Prices

Final Model Metrics

RMSE: $48798
R²: 0.71