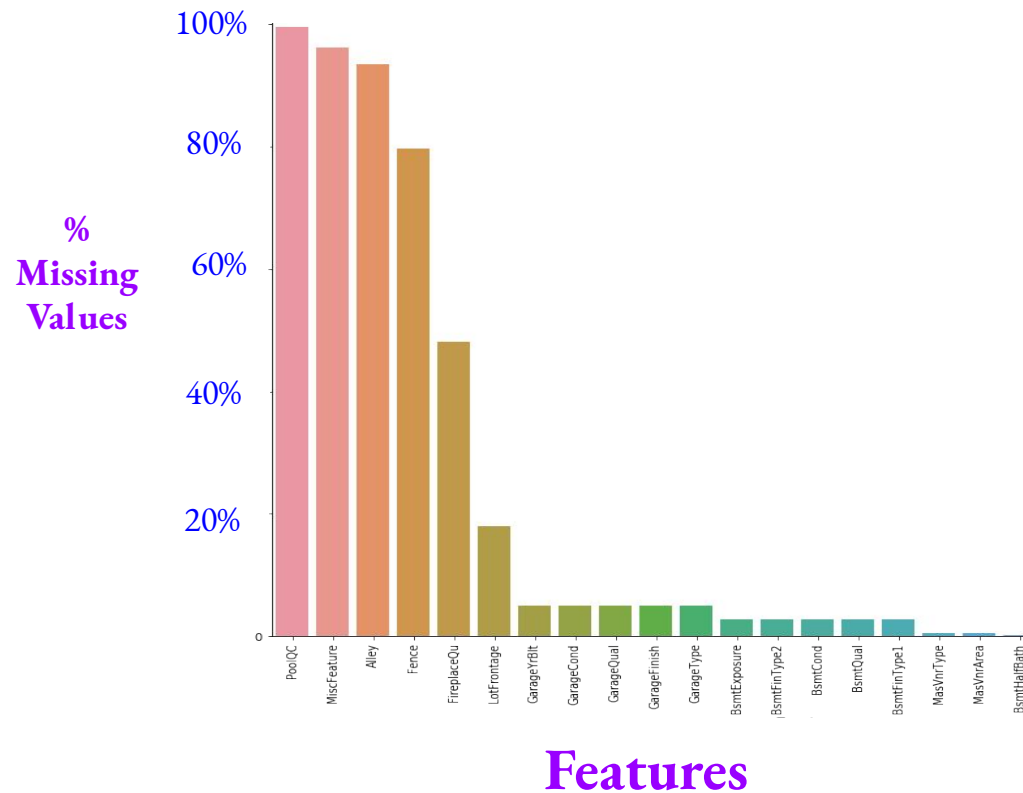


Machine Learning to Enhance Cost-Effective Decision Making by Housing Developers

Missing Data Percent by Feature



Missing Ratio

- Pool Quality Rating: 99.7%
- Miscellaneous Features: 96.2%
- Alley: 93.5%
- Fence: 79.7%
- Fireplace Quality: 48.1%
- Lot Frontage: 17.9%

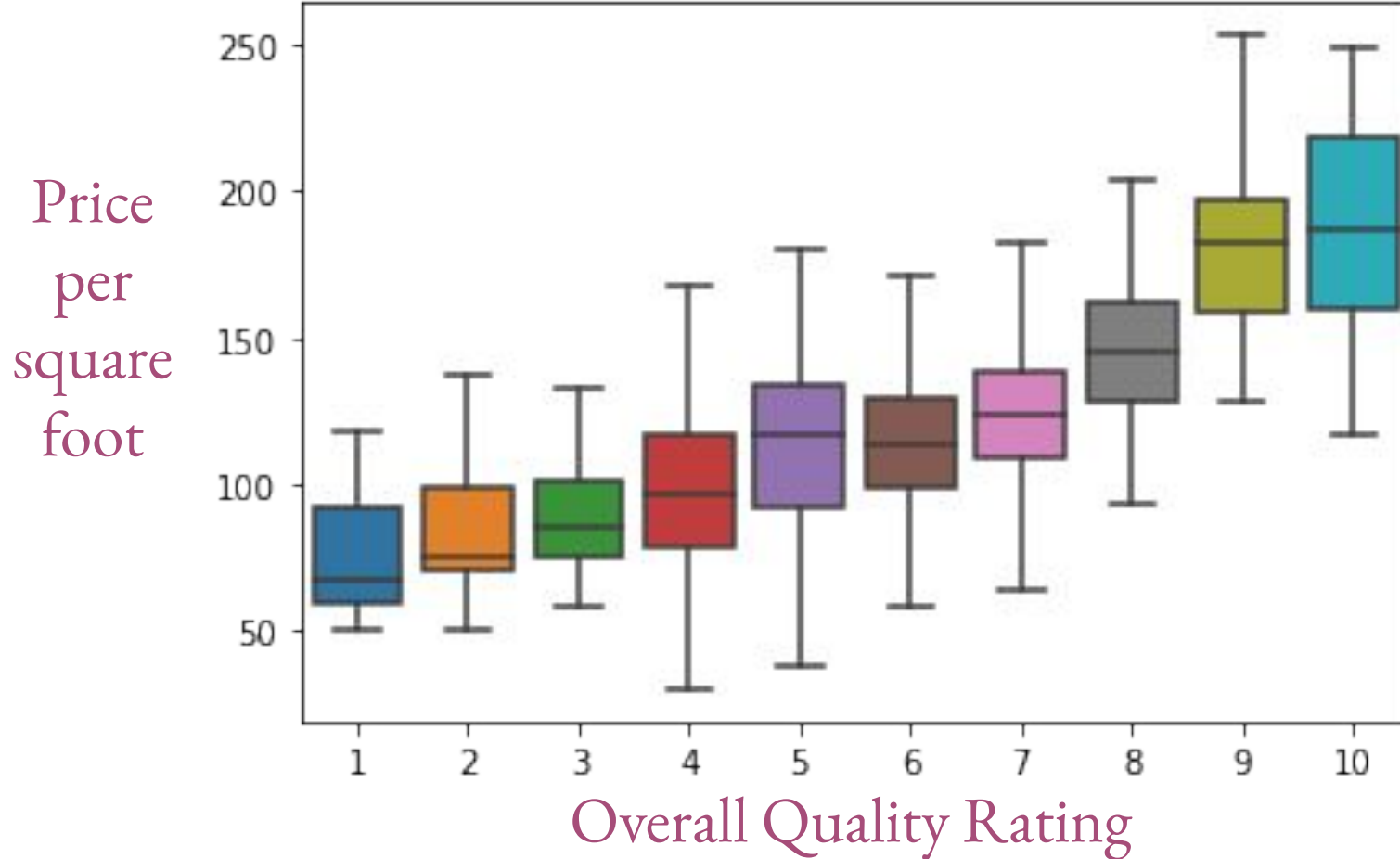
All other features less than 5% missing

Feature Cleaning (Missing Values)

- For *Pool Quality Rating*, *Miscellaneous*, *Alley*, *Fence*, and *Fireplace Quality*: fill with “**None**” value.
- For *Lot Frontage* feature: fill with **median** neighborhood Lot Frontage. Reason: maximum is 313, minimum is 21, and mean is 68.5, so the maximum values are outliers that inflate the mean.
- For *Electrical*, *Zoning Classification*, *Utilities*, *Home Functionality*, *Type of Sale*, *Kitchen Quality*, and both *Exterior Covering* features: fill with the mode, the most common value.

How does the overall
quality of a home influence
price per square foot?

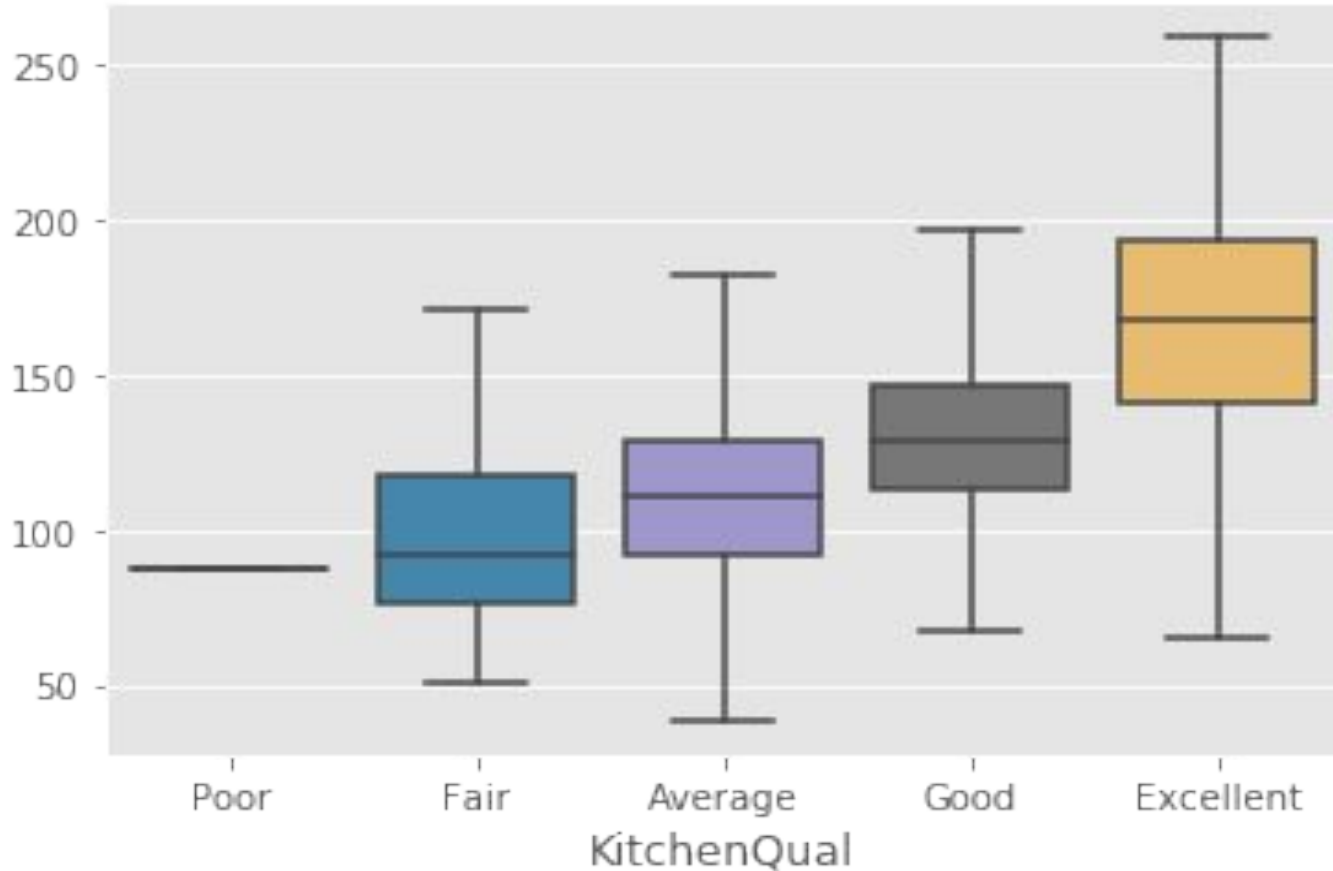
Overall Quality vs. Price per Square Foot



Does a higher kitchen
quality correlate with a
higher price per square foot?

Graphs of Kitchen and Overall Quality

Price per
square
foot



Which combinations of
bedrooms and bathrooms
have highest price per
square foot?

Bed and Bath Combinations vs Price per Square Foot

(Bed,Bath)

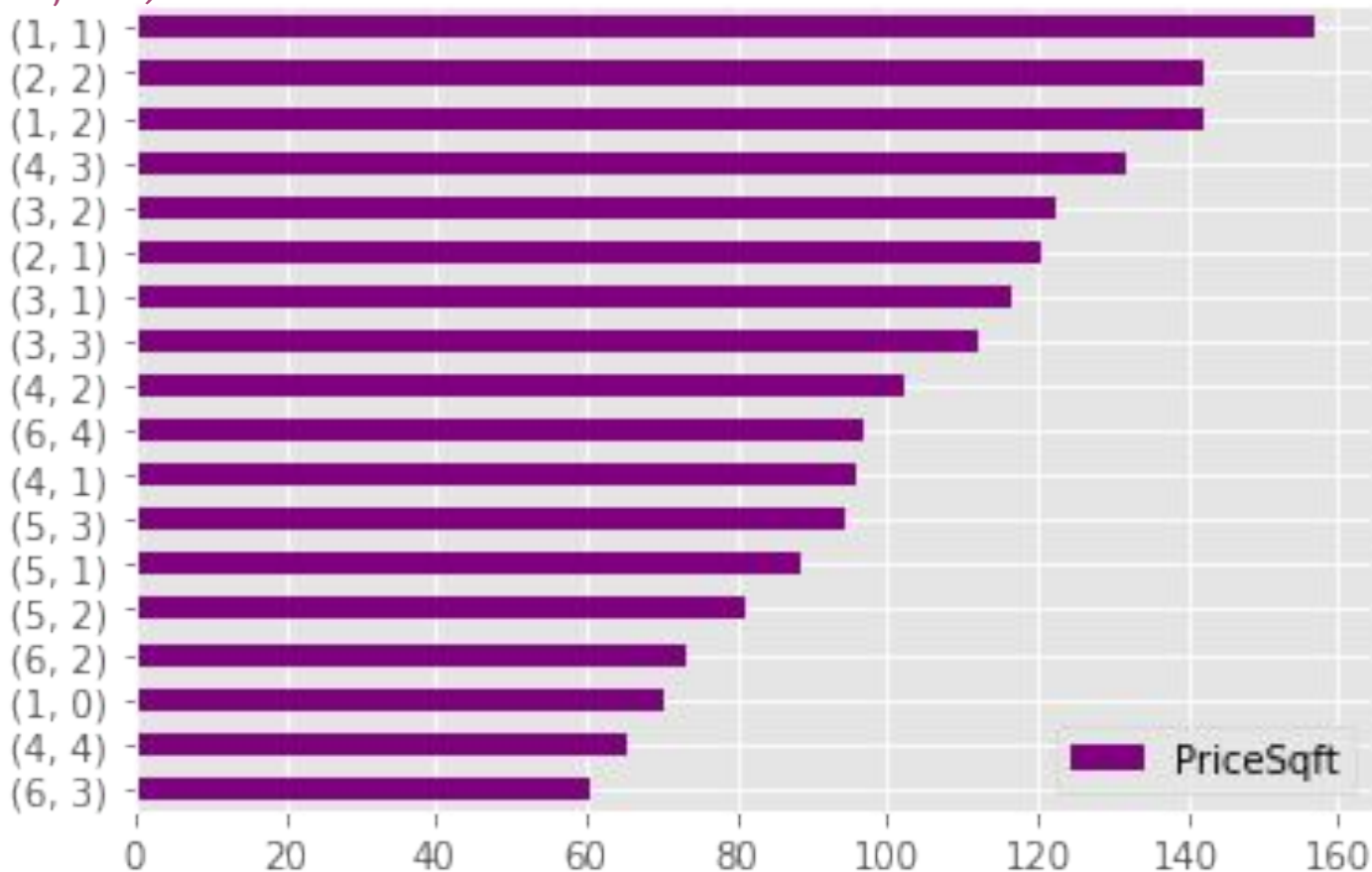
(1, 1)
(2, 2)
(1, 2)
(4, 3)
(3, 2)
(2, 1)
(3, 1)
(3, 3)
(4, 2)
(6, 4)
(4, 1)
(5, 3)
(5, 1)
(5, 2)
(6, 2)
(1, 0)
(4, 4)
(6, 3)

Number of
Bedrooms,
Number of
Bathrooms

0 20 40 60 80 100 120 140 160

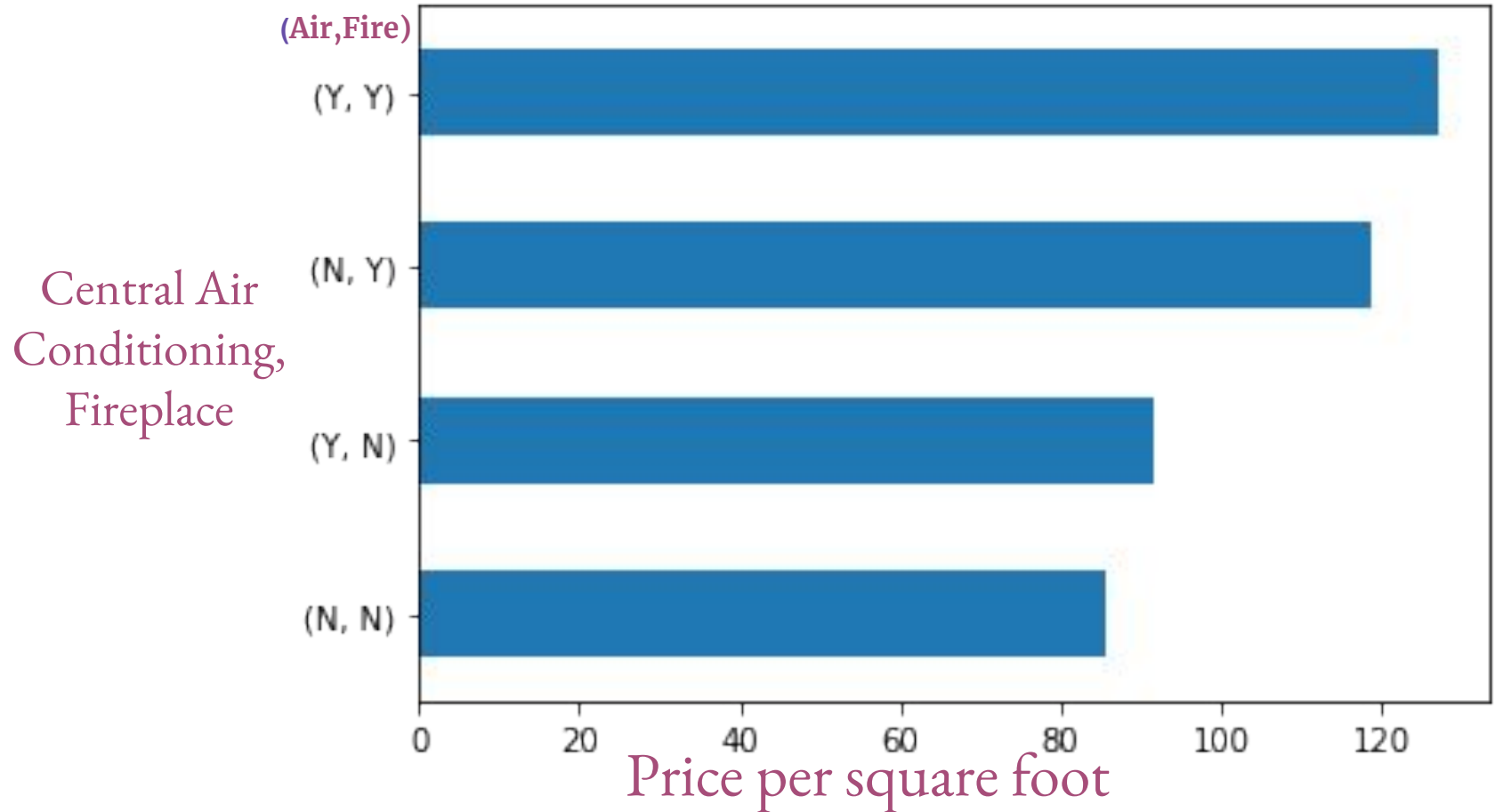
Price per square foot

PriceSqft



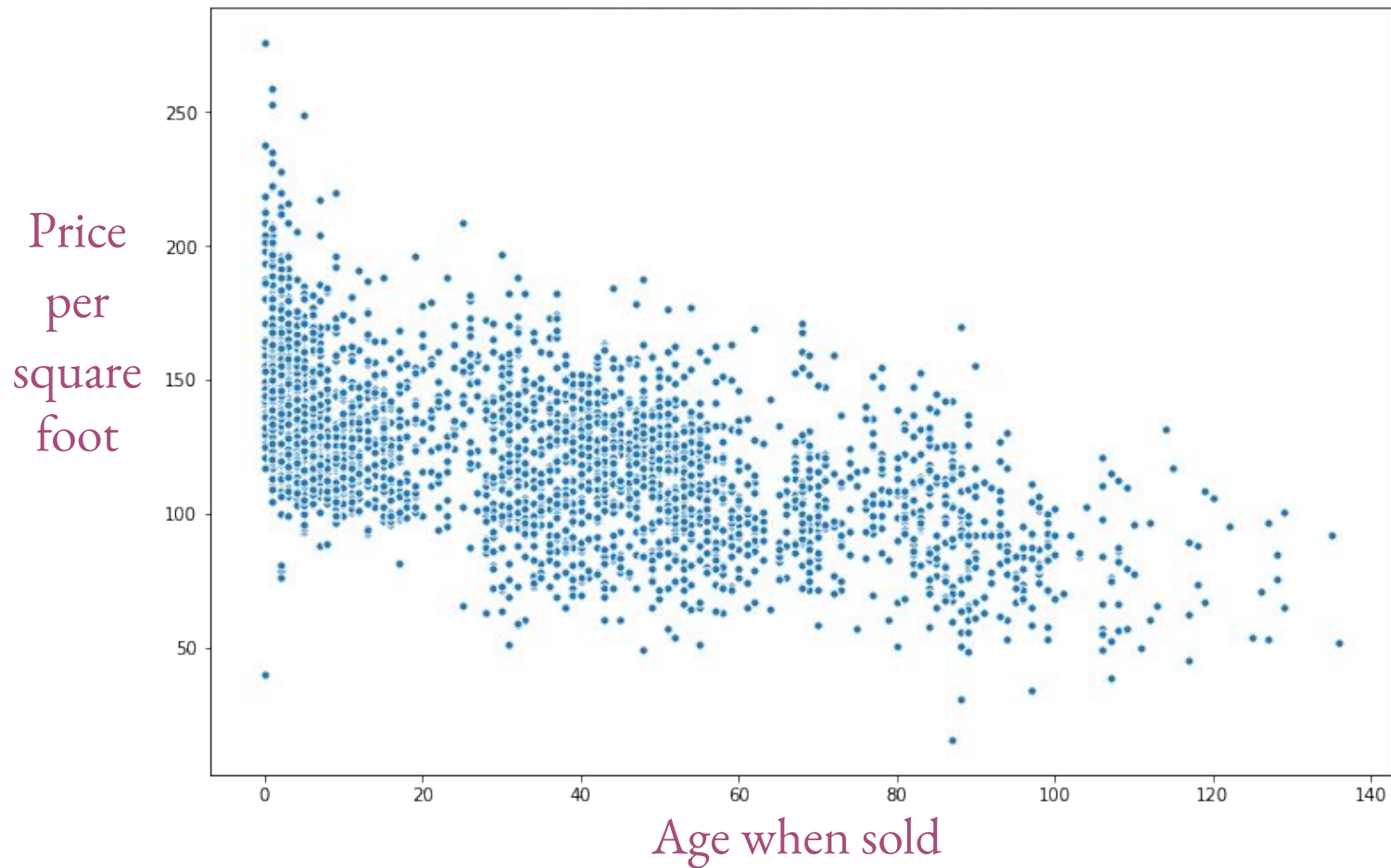
Does Central Air
Conditioning or a Fireplace
affect price per square foot?

Central AC and Fireplace vs. Price per Square Foot



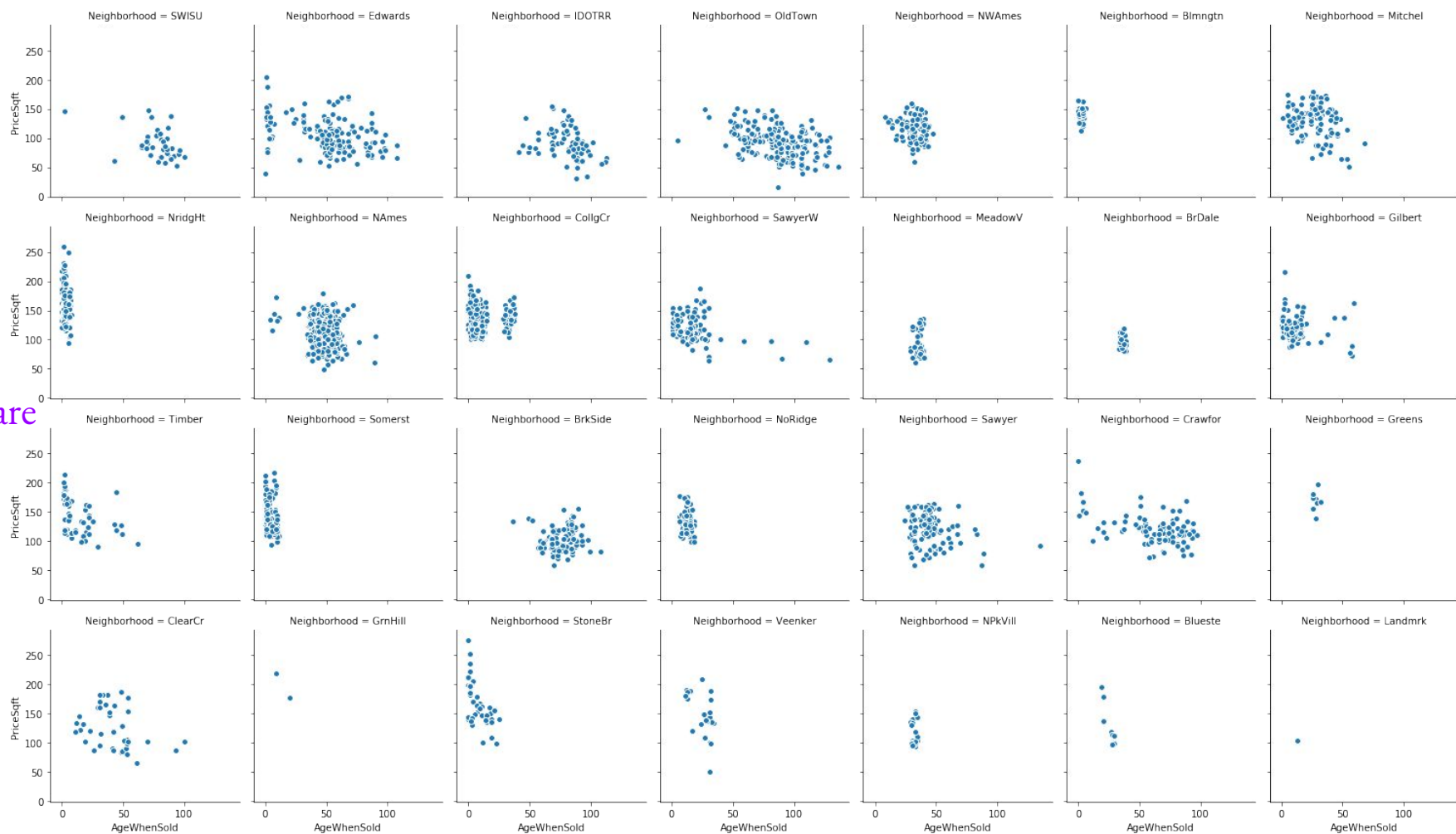
How does the age of a house
influence price per square foot?

Age of House vs. Price per Square Foot



Does this vary by
neighborhood?

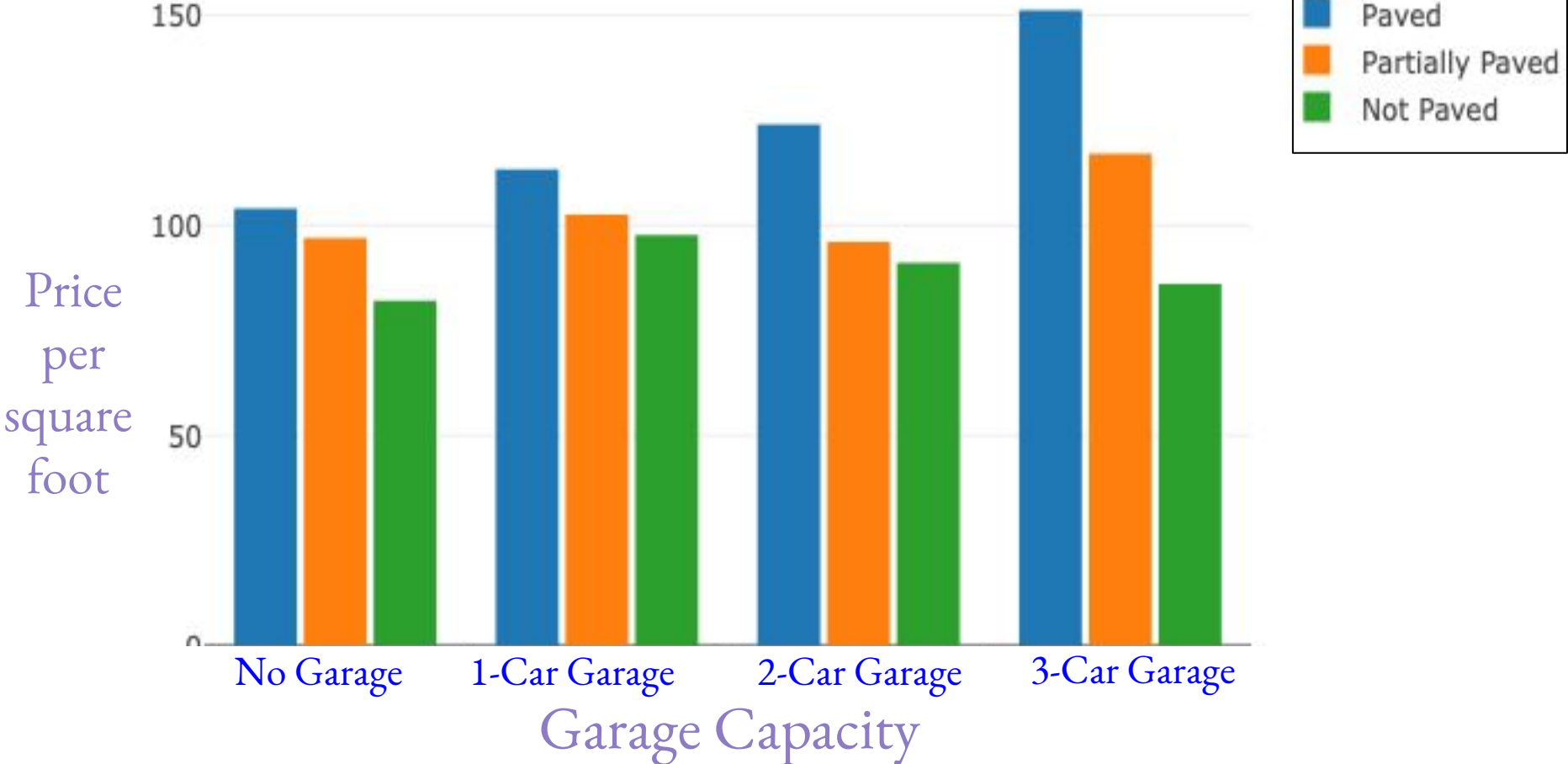
Price
per square
foot



Age when sold

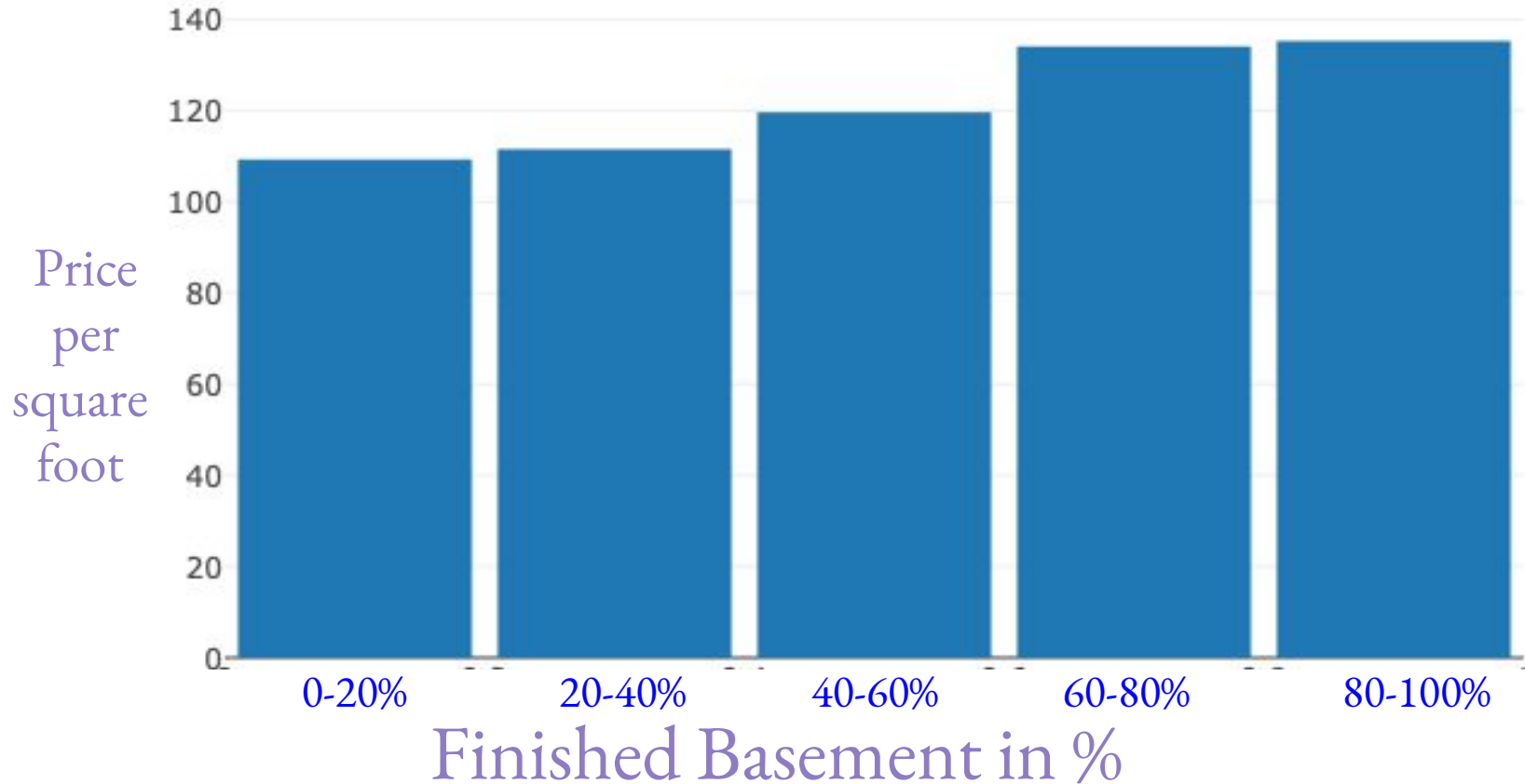
Does capacity of the garage
and state of the driveway
result in a higher price per
square foot?

Garage Capacity, Driveway, and Price per Square Foot



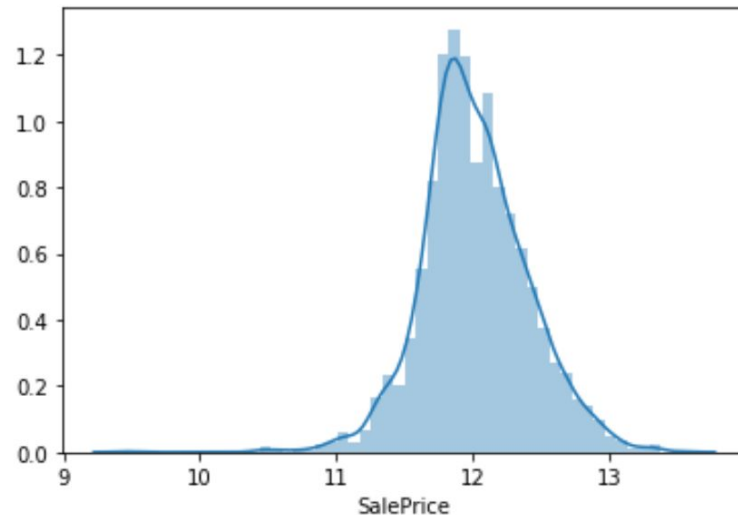
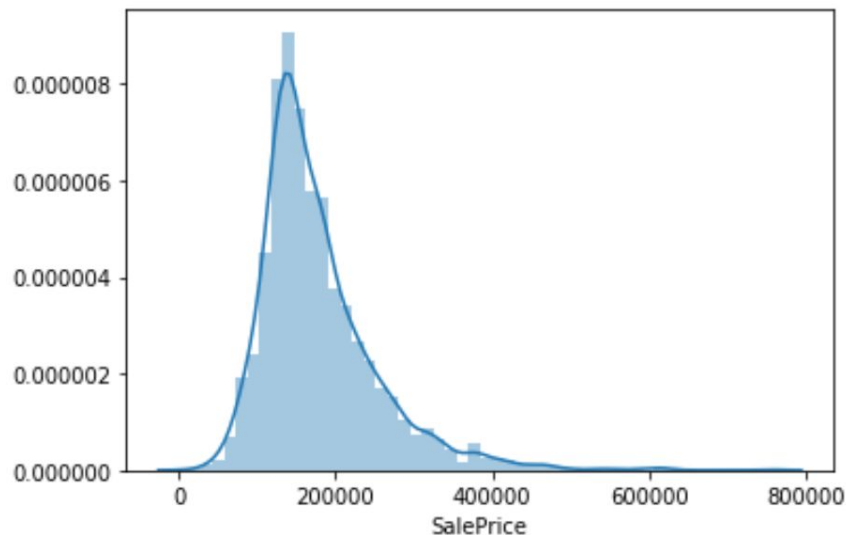
How does a finished basement
affect the average price per
square foot?

Finished Basement in % vs. Price per Square Foot

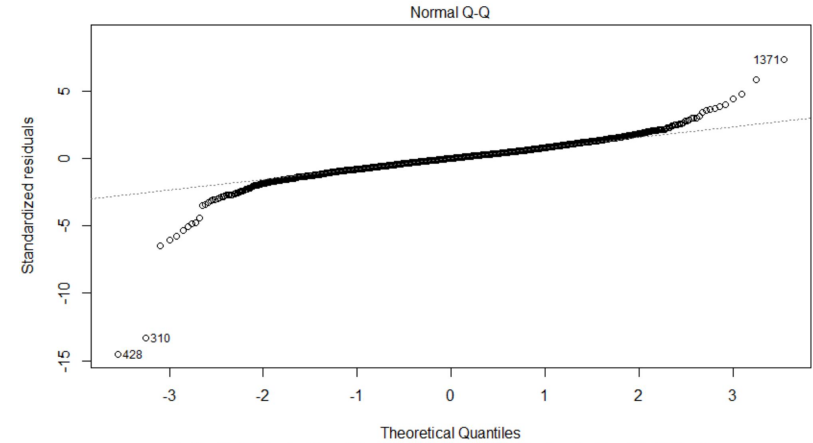
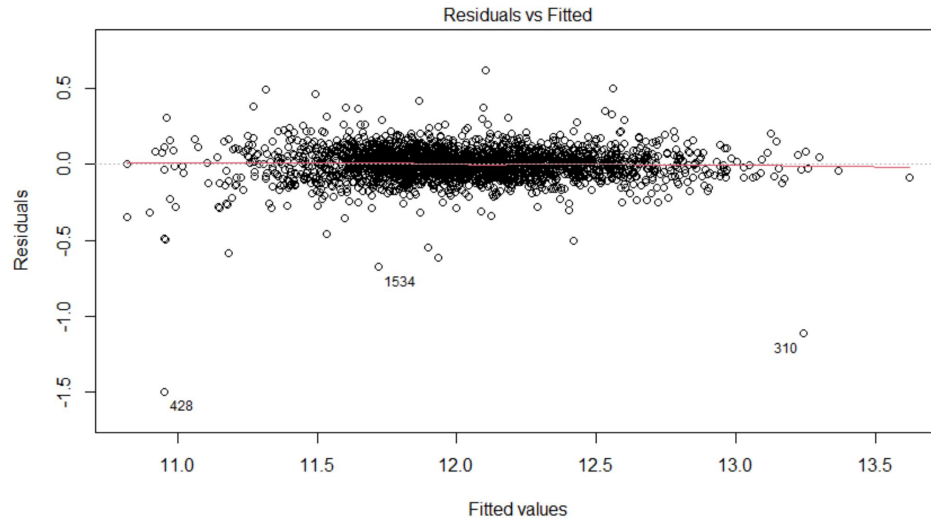


What does the data tell us about
features significant to
predicting sale price?

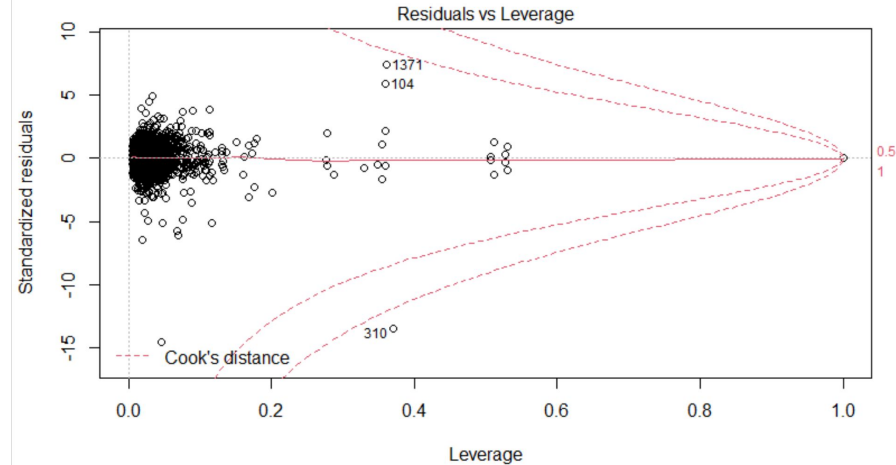
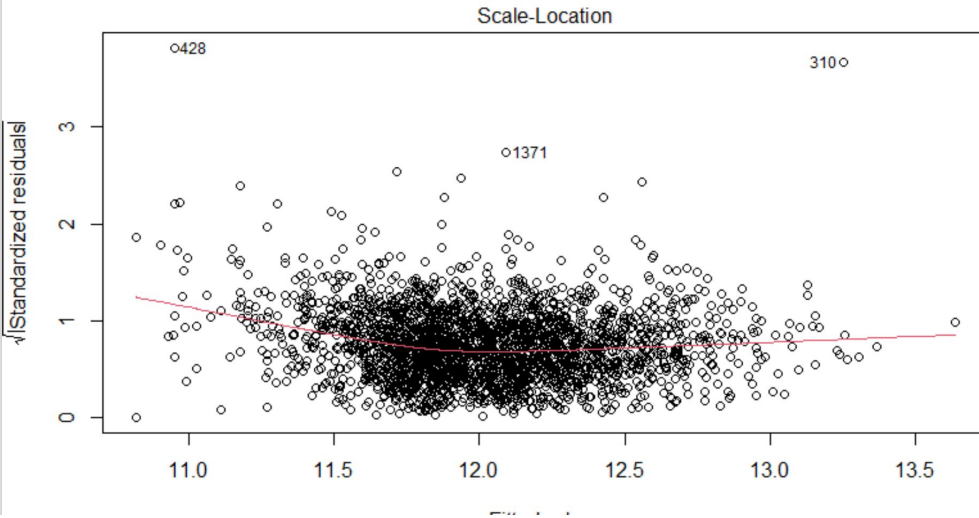
Log Transformation for the Sale Price



Conditions for Multiple Linear Regression



Conditions for Multiple Linear Regression



Stepwise regression using Bayesian Information Criterion (BIC)

$\log(\text{SalePrice}) \sim \text{OverallQual} + \text{GrLivArea} + \text{Neighborhood} + \text{BsmtFinSF1} +$
 $\text{OverallCond} + \text{ageWhenSold} + \text{TotalBsmtSF} + \text{BldgType} + \text{GarageCars} +$
 $\text{Fireplaces} + \text{SaleCondition} + \text{CentralAir} + \text{LotArea} + \text{Condition2} +$
 $\text{KitchenQual} + \text{BsmtExposure} + \text{YearRemodAdd} + \text{ScreenPorch} +$
 $\text{MSZoning} + \text{Functional} + \text{BsmtFullBath} + \text{EnclosedPorch} + \text{HeatingQC} +$
 $\text{PavedDrive} + \text{bb_diff} + \text{BsmtFinSF2}$

Adjusted r-squared : 0.9263

Lasso + GridSearchCV



best parameter
alpha = $5.995e-5$



mean cross-validated score:
0.905



performance score:
0.926

Ridge + GridSearchCV



best parameter
alpha = 0.13



mean cross-validated score:
0.904



performance score:
0.924

Random Forest + GridSearchCV

best parameters:

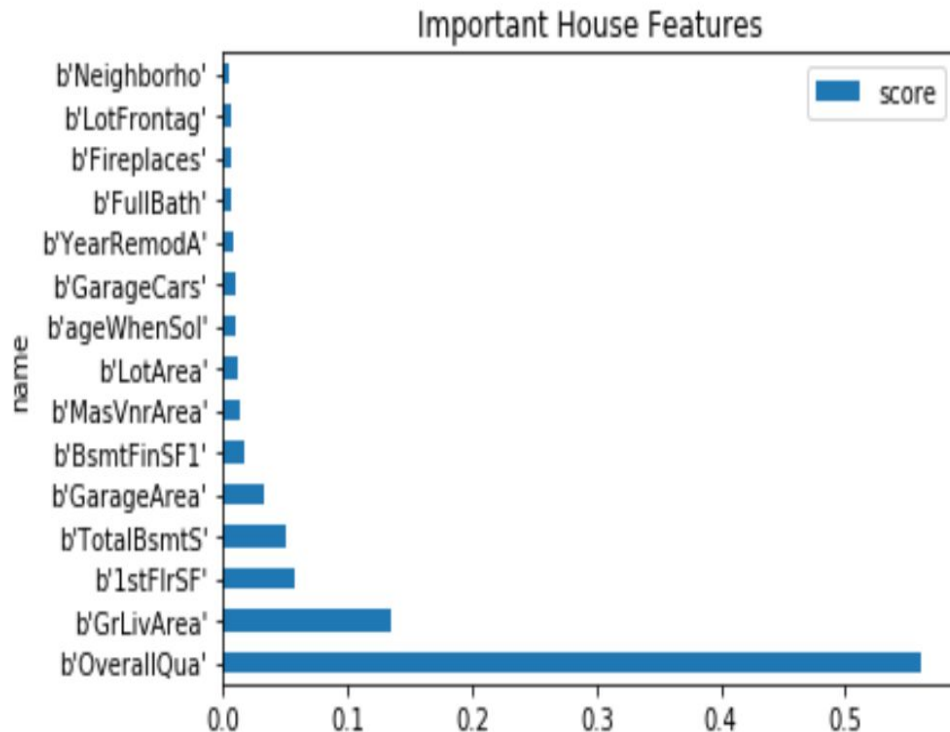
min_samples_leaf = 2

Min_samples_split = 7

n_estimators = 20

Mean cross_validation score:
0.882

performance score
0.8952



RECOMMENDATIONS

- Overall house quality

A one point increase in overall quality rating **increases** sale price by **0.06%**.

- Kitchen Quality

Excellent kitchen quality sells **0.063% higher** compared to other houses.

- Central Air vs Fireplace

Central air gives a **0.056% higher** sale price compared to other houses.

- **Driveway**

A home with a paved driveway sells **0.048% higher** than a home without a paved driveway

- **Bedroom and Bathroom**

For every *one room increase* between *number of bedrooms and bathrooms*, sale price **drops 0.01%**.