



Predicting House Prices with the Kings County Housing Dataset

Avonlea Fisher
September 2020



Process

01

Data cleaning and preprocessing

02

Exploring the data

03

Building linear models

04

Interpreting model results

Feature Correlations

Correlations with Price

	Correlations	Features
0	0.336450	bedrooms
1	0.527590	bathrooms
2	0.661318	sqft_living
3	0.302772	floors
4	0.331015	view
5	0.690809	grade
6	0.581063	sqft_above
7	0.599692	sqft_living15

Notes

-Features that were highly correlated with price were considered for inclusion in the model.

-Combinations of highly multicollinear features were avoided.

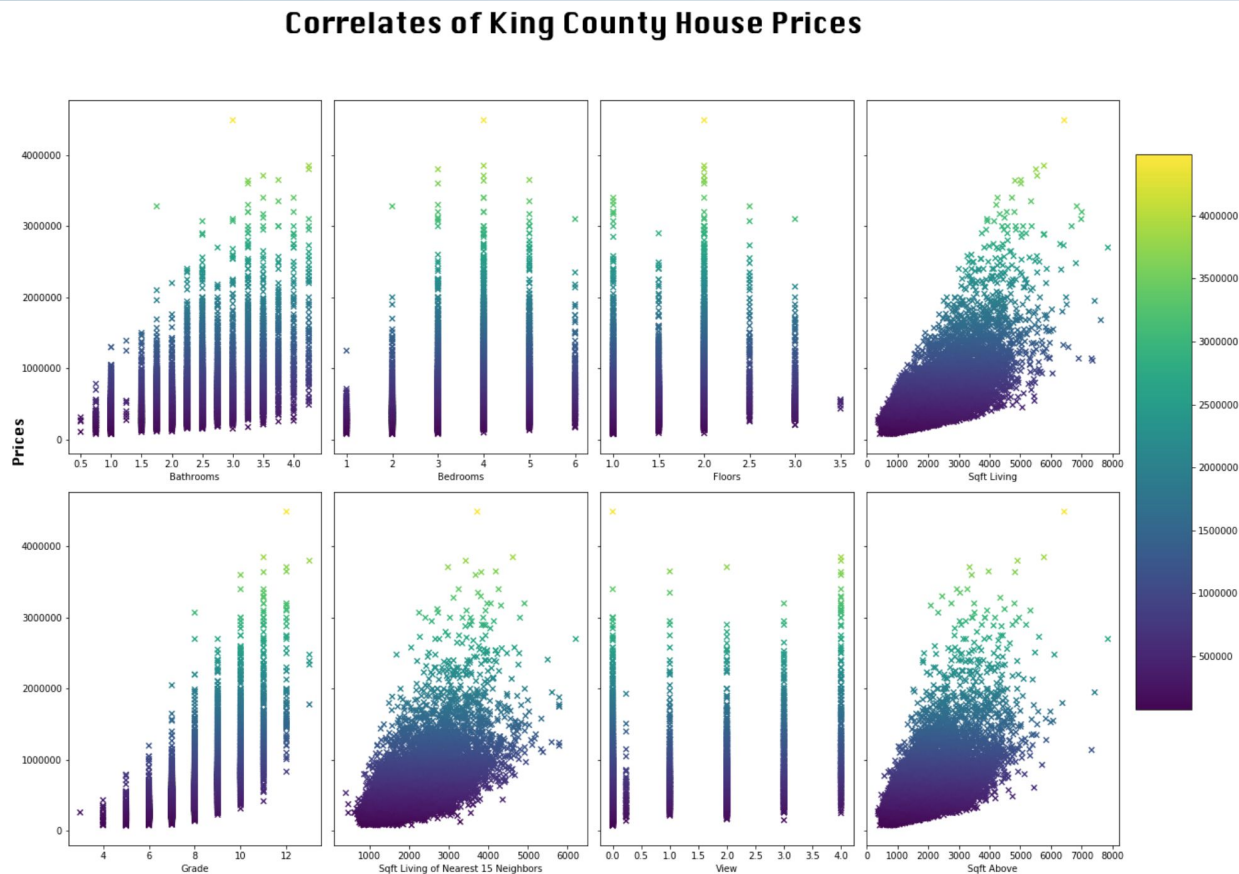
Multicollinear Features

	Correlations	Features
0	0.827893	[sqft_living, sqft_above]
1	0.919460	[sqft_lot, sqft_lot15]
2	0.827893	[sqft_above, sqft_living]
3	0.919460	[sqft_lot15, sqft_lot]
4	0.812117	[3, 4]
5	0.812117	[4, 3]



Notes

- View, floors, and bedrooms were excluded from models due to a weak linear relationship



Model and Results

price ~ sqft_living + sqft_living15 + grade + bathroom

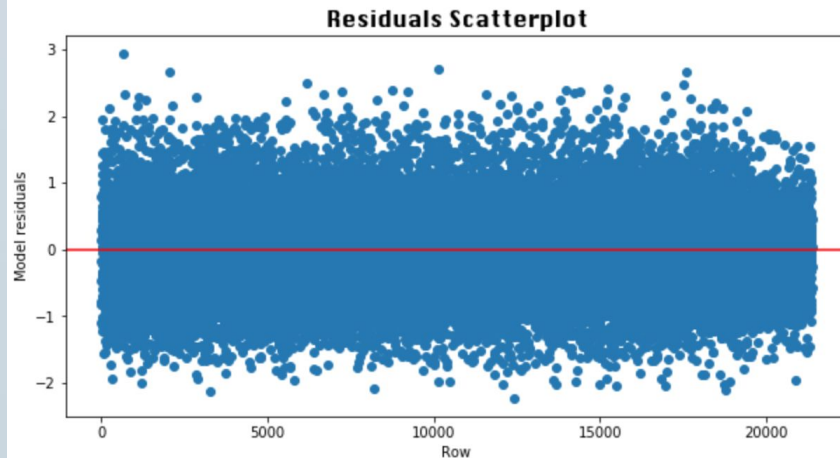


Notes: The model can account for 52% of the variability in price. We can reject the null hypothesis, which states that there is no relationship between price and the response variables chosen for the model.

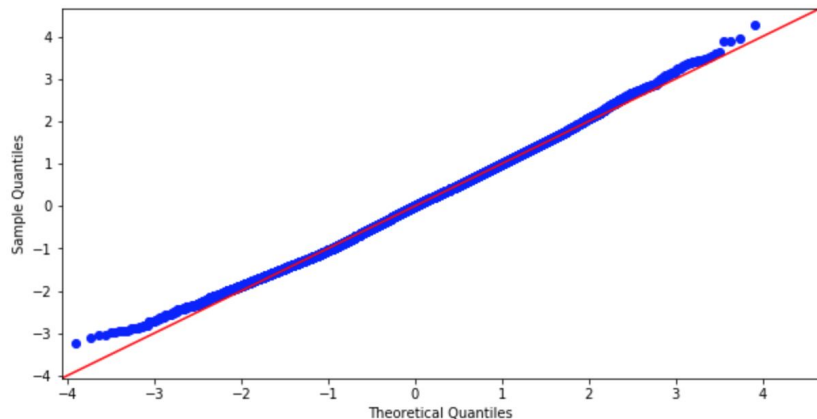


Residuals Plots

Homoscedasticity: the residuals have equal variance along the regression line.



Residuals QQ Plot

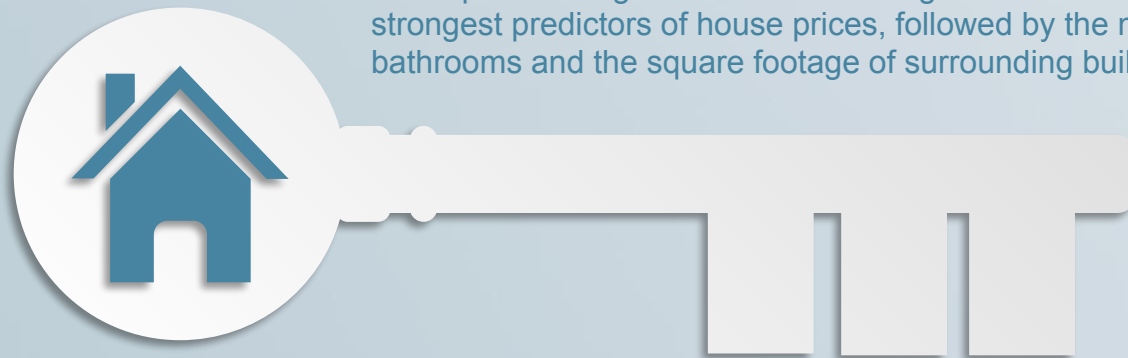


Normality: the residuals follow a normal distribution.



Conclusions

The square footage of a house and its grade are among the strongest predictors of house prices, followed by the number of bathrooms and the square footage of surrounding buildings.



Build additional bathrooms

If you have an exceptionally large room in your home, consider converting part of it to an extra bathroom. Price tends to increase with the addition of bathrooms.

Improve construction quality

A house's grade is based on the quality of its construction and design. Renovations can help boost the price of your house.

If possible, expand living area

Lot size is a weaker predictor of price than living area. If you're able to 'build out,' your house price could increase significantly.





Thank you!