King County House Prices

for Investors and Real estate agents

Christoph Blickle

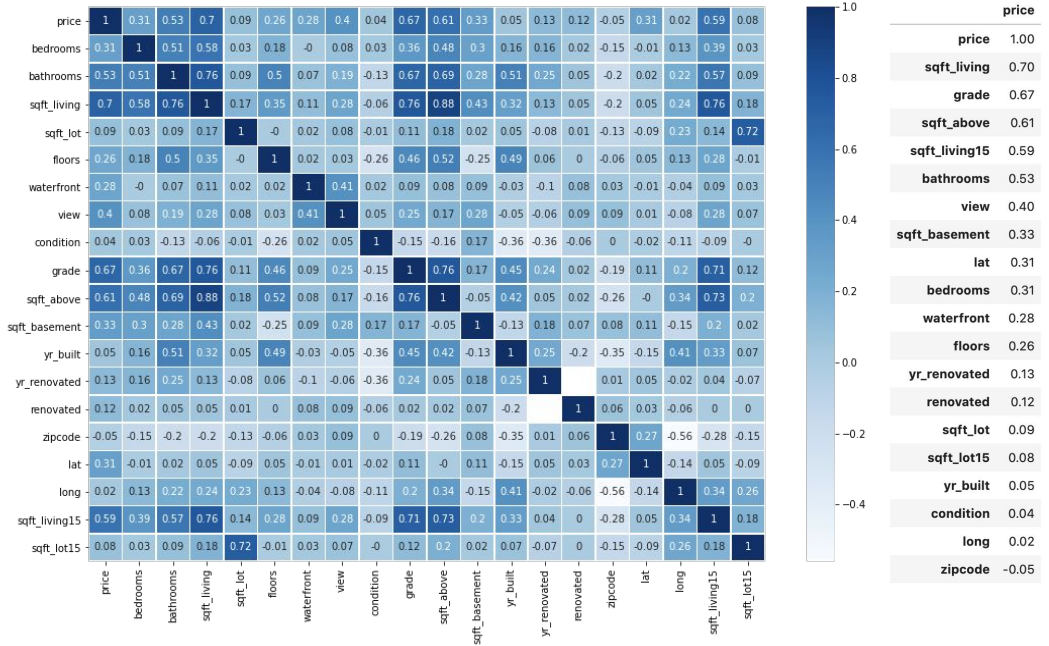
Contents

Overview of the given data

- The data for 21597 house in King County was given
- The dataset is mostly complete with missing values in only a few variables
- Strong outliers can be found in sqft_living, bedrooms, bathrooms and price
    - Probably luxury houses
- The outliers were not removed as these may also be interesting to high class investors and real estate agents

Overview of the given data

- Correlation between variables
- Correlation with price
- Highest with sqft_living and grade



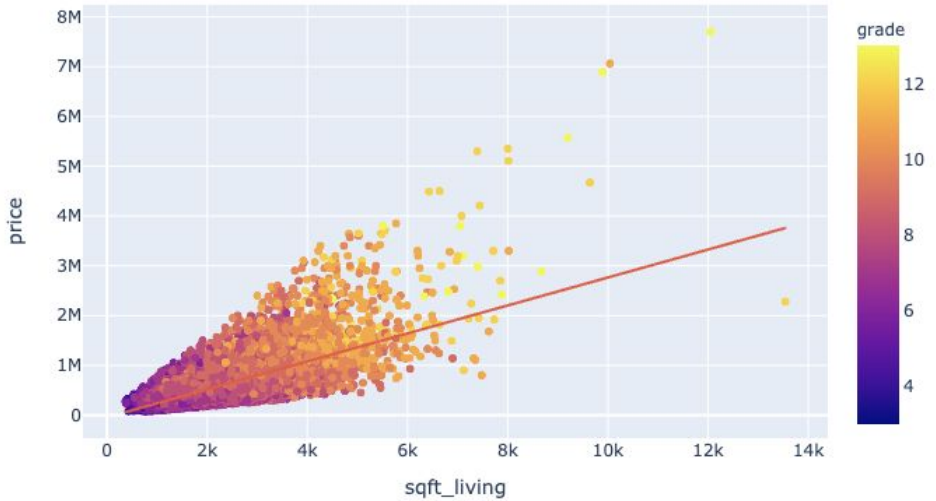
	price
price	1.00
sqft_living	0.70
grade	0.67
sqft_above	0.61
sqft_living15	0.59
bathrooms	0.53
view	0.40
sqft_basement	0.33
lat	0.31
bedrooms	0.31
waterfront	0.28
floors	0.26
yr_renovated	0.13
renovated	0.12
sqft_lot	0.09
sqft_lot15	0.08
yr_built	0.05
condition	0.04
long	0.02
zipcode	-0.05

Parameter relations

- Price is mostly influenced by sqft_living and grade
- These go hand in hand

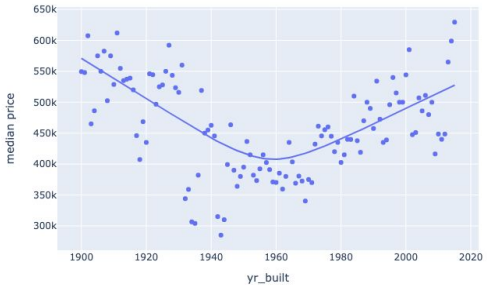
	grade
sqft_living	0.762779

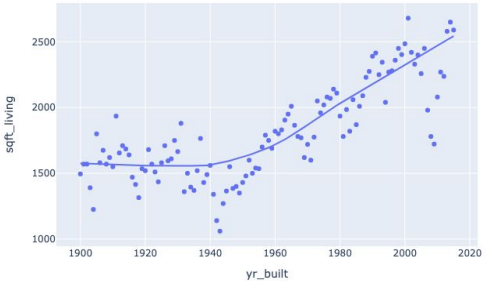
- A positive linear correlation between price and sqft_living is observable

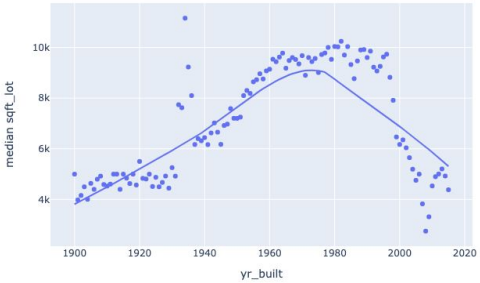


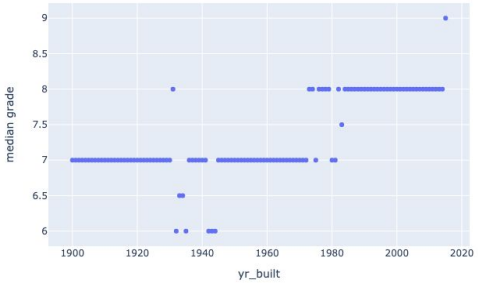
Parameter relations

- Older house are more expensive even though they offer less space
- Grades are somewhat constant with small dips in the 40s and an increase in the 90s







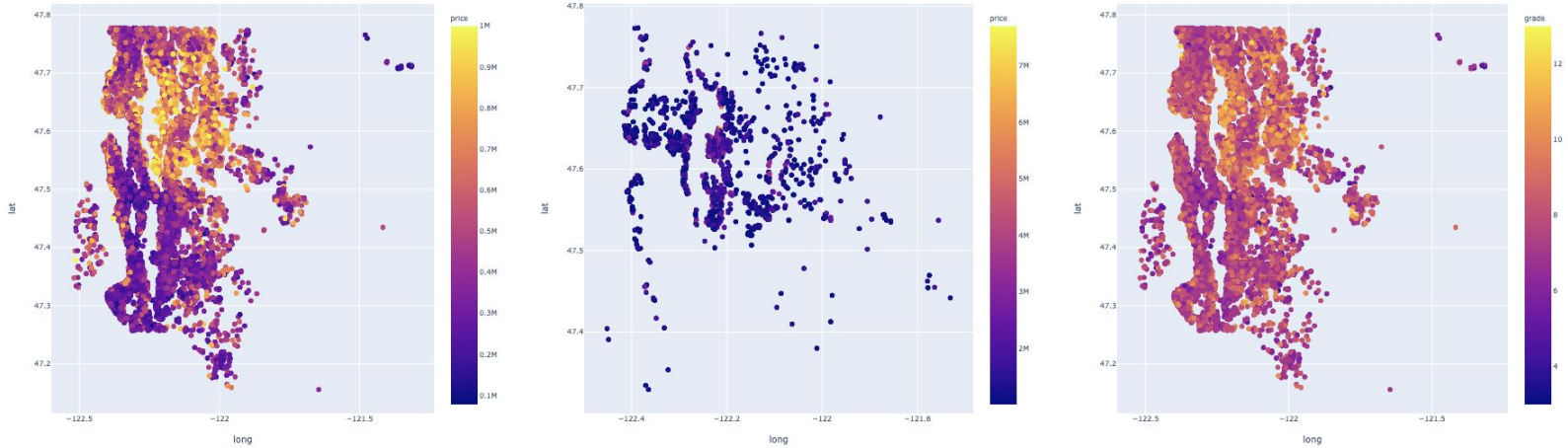


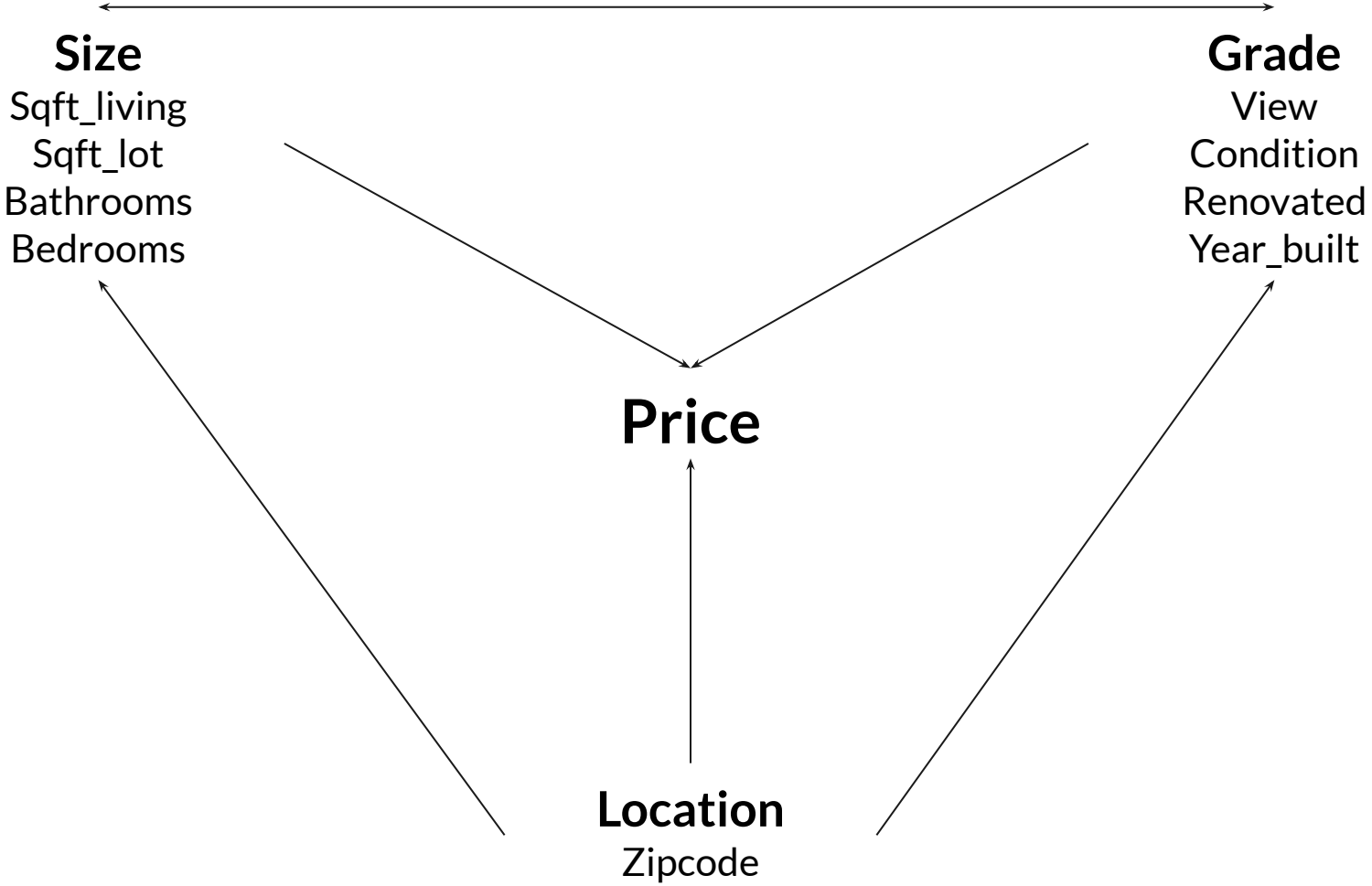
Parameter relations

- Literature researches show that seasons influence sales prices
- The data only contains one year of data
- A small fluctuation is still visible

Parameter relations

- Prices correlate with location, but this seems to be the consequence of higher grades in the area




Size
Sqft_living
Sqft_lot
Bathrooms
Bedrooms
Grade
View
Condition
Renovated
Year_built
Price
Location
Zipcode

Recommendations

- When was the house built?
  - The newer the house the more expensive it will get to buy in the future
  - Buy houses built in the 40s to the 80s
- How big is the lot in comparison to the actual living space?
  - Buy big lots to leave room for living space expansion
  - Adding new floors is more expensive making inner city houses less desirable and hard to improve
  - Add bathrooms and bedrooms if needed, but not as sole means to improve the resale price
- Why is a grade high or low?
  - Avoid houses whose grades are high due to their location, since improvement is difficult. Thus making a buy and resell less beneficial
  - Improve grade by renovation and expansion
- Keep the seasons in mind

Multivariate linear regression

- The algorithm uses only 3 variables:
  - sqft_living, grade and zipcode
- Thereby being reliable and less prone to errors when using new datasets
- The 3 variables cover most other parameters, since sqft_living also correlates with number of bathrooms, grade with view etc.
- The mean absolute percentage error at this point in time equals 18.9 %
- A better MAPE can be achieved when more variables are passed, but this could come at the cost of reliability

Summary

- The King County house price dataset from 14/15 was used to understand influences of different parameters
- The goal was to give recommendations to real estate agents and investors
- Multivariate linear regression can be used to predict house prices for new data