King County Housing Price Prediction

Analysis of King County housing data to predict house prices

overview

- Data Overview
- Business Understanding

Findings

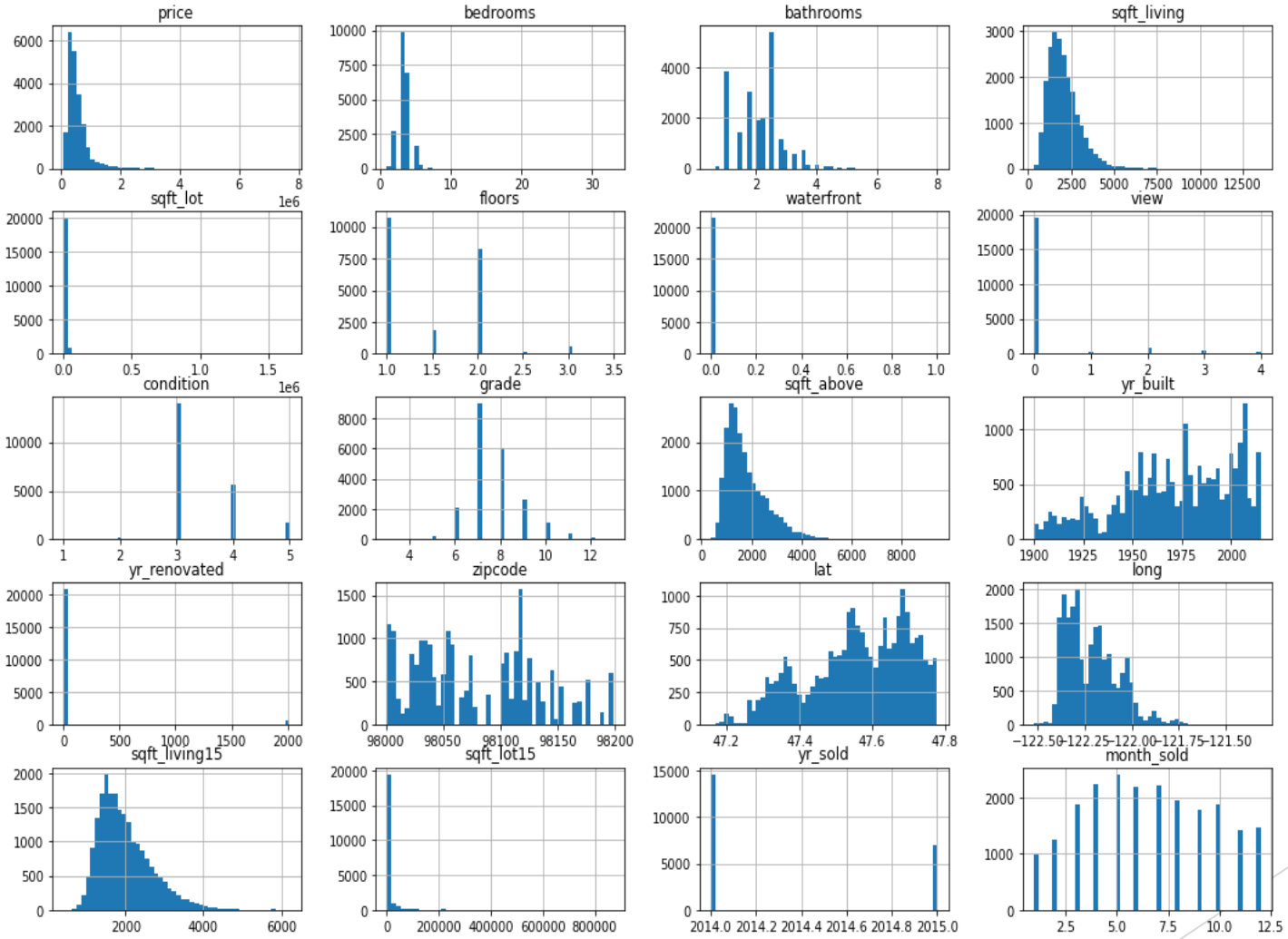
- Discussion
- Recommendation
- Conclusion

Data Overview

- Data contains 21 features of houses sold between May 2014 and May 2015. There are a total of 21597 records of houses were sold in the dataset

Variable Distribution

- Skewed data



Business Understanding

- The predictive model is for real estate investors.
- It will help determine house prices
- It will determine which features favour prices to drive profit margins high

Findings

- Top two most significant feature are:
  - Grades
  - square footage of houses.

- Least Significant Features:
  - Sqft_basement
  - Sqft_lot
  - Sqft_lot15

Correlation of Price with Square Footage of Homes

Price increases with square footage of the home

Best Model

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.537
Model:	OLS	**Adj. R-squared:**	0.537
Method:	Least Squares	**F-statistic:**	6671.
Date:	Mon, 04 Jul 2022	**Prob (F-statistic):**	0.00
Time:	04:08:49	**Log-Likelihood:**	-2.3933e+05
No. Observations:	17277	**AIC:**	4.787e+05
Df Residuals:	17273	**BIC:**	4.787e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-6.17e+05	1.5e+04	-41.058	0.000	-6.46e+05	-5.88e+05
sqft_living	176.9738	3.651	48.478	0.000	169.818	184.129
grade	9.882e+04	2663.123	37.106	0.000	9.36e+04	1.04e+05
sqft_living15	16.6291	4.541	3.662	0.000	7.729	25.530

Omnibus:	13140.968	**Durbin-Watson:**	1.996
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	675087.896
Skew:	3.185	**Prob(JB):**	0.00
Kurtosis:	32.953	**Cond. No.**	2.45e+04

R squared 54%

Recommendations

- Homes with better grades
- Homes with appr 2000sqft
- Size of homes similar to neighbouring Homes

Limitations of Model

- Multicollinearity unless Transformed

through natural log and feature scaling