PROJECT 2: KING COUNTY HOUSING

By Sam Lim
Data Science Full Time
3/26/21

Overview

THE BUSINESS PROBLEM

DATA MODELING/
REGRESSION MODEL

CONCLUSION/
RECOMMENDATION

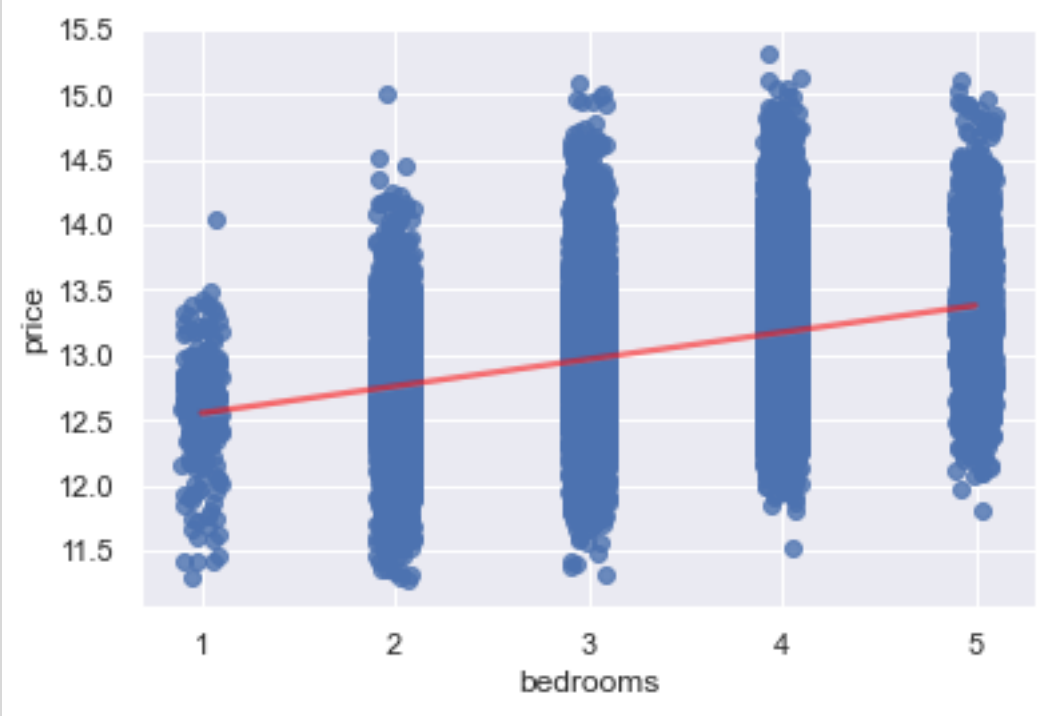
Business Problem

- A realtorship is experiencing an influx of clients as smaller families are coming into King County. As Microsoft has expanded, it has hired substantial number of workers. The task is to find a price range of a house depending on the specifications of the client.

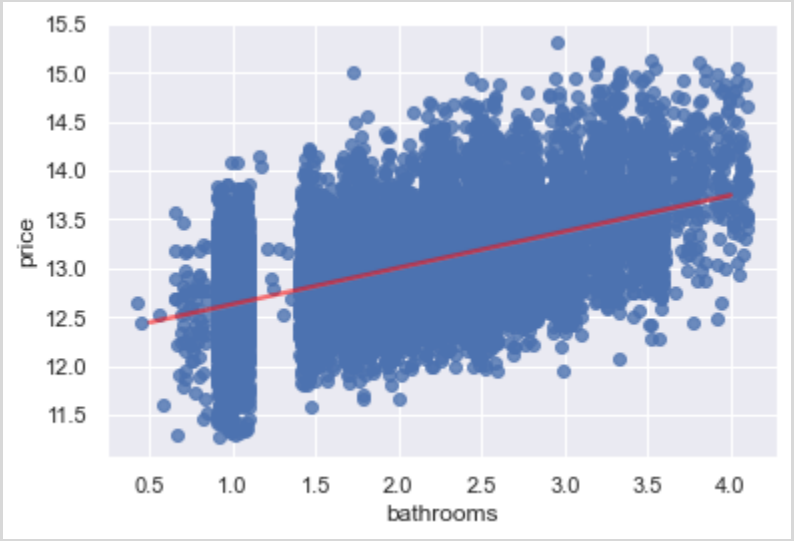
Data Modeling

- As most clients are going to be living alone or with small to mid-size families, the number of rooms should be set to no more than 5 as well as the number of bathrooms to be no more than 4.
- Setting number of bedrooms and bathrooms, the size of lot and living space, as well as the condition as the main independent variables in determining the price.

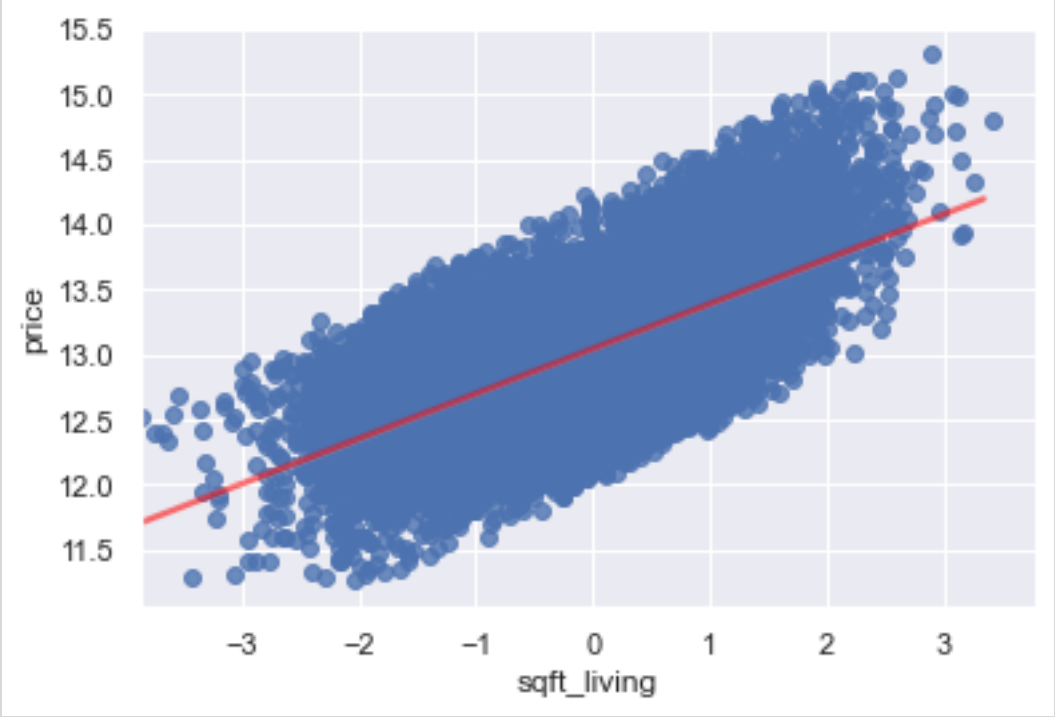
Bedrooms



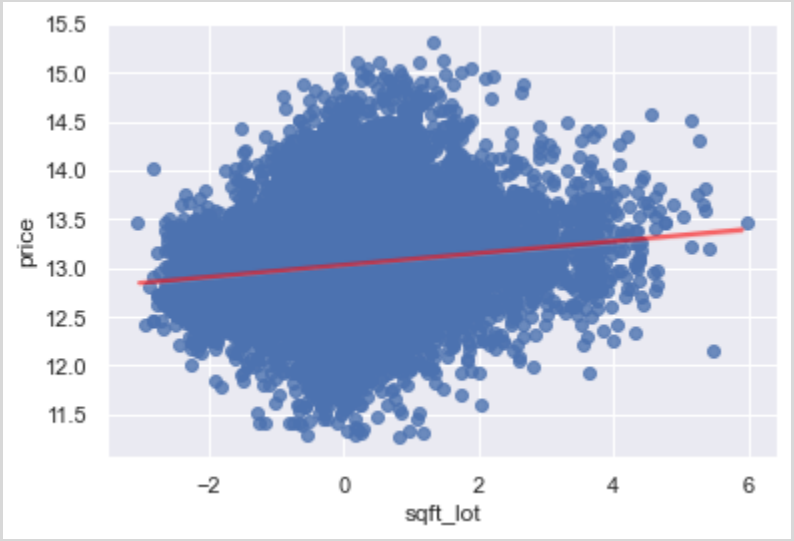
Bathrooms


15.5
15.0
14.5
14.0
13.5
13.0
12.5
12.0
11.5
price
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
bathrooms

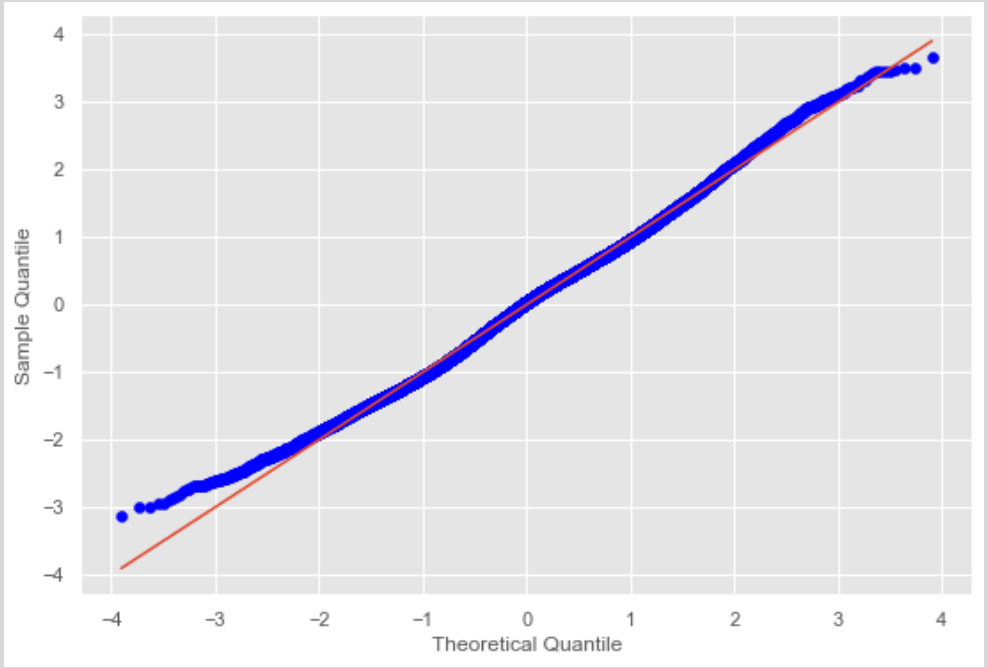
Sqft Living

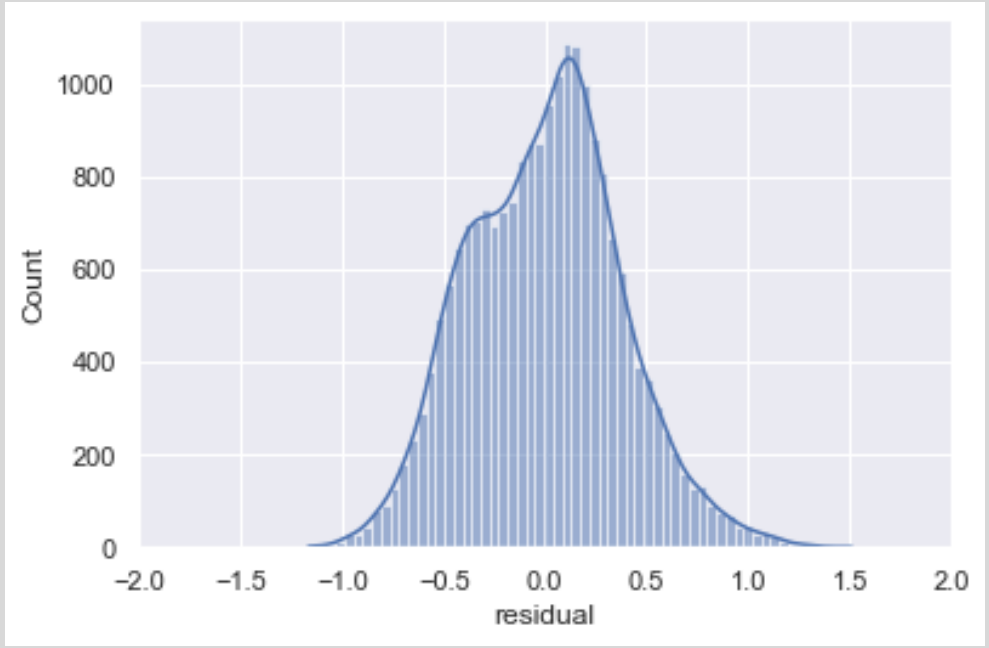


Sqft Lot



Q-Q Plot and Residual Histogram





Dep. Variable:	price	**R-squared:**	0.461
Model:	OLS	**Adj. R-squared:**	0.461
Method:	Least Squares	**F-statistic:**	2253.
Date:	Fri, 26 Mar 2021	**Prob (F-statistic):**	0.00
Time:	16:15:46	**Log-Likelihood:**	-9178.8
No. Observations:	21061	**AIC:**	1.838e+04
Df Residuals:	21052	**BIC:**	1.845e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	13.2128	0.073	181.587	0.000	13.070	13.355
cnd_2[T.1]	-0.0868	0.076	-1.136	0.256	-0.237	0.063
cnd_3[T.1]	0.0444	0.071	0.626	0.531	-0.095	0.183
cnd_4[T.1]	0.1066	0.071	1.501	0.133	-0.033	0.246
cnd_5[T.1]	0.2132	0.071	2.984	0.003	0.073	0.353
bedrooms	-0.0968	0.004	-23.666	0.000	-0.105	-0.089
bathrooms	0.0391	0.006	6.785	0.000	0.028	0.050
sqft_living	0.3929	0.005	81.695	0.000	0.384	0.402
sqft_lot	-0.0485	0.003	-16.956	0.000	-0.054	-0.043

Omnibus:	68.943	**Durbin-Watson:**	1.980
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	68.132
Skew:	0.127	**Prob(JB):**	1.60e-15
Kurtosis:	2.887	**Cond. No.**	259.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Model

Conclusion

- The biggest challenges were first, trying to figure out why the relationship between lot size and number of bedrooms to the price were negative, and much effort was put into trying to make the coefficients be positive in the model. Also, not having any information on the neighborhoods depending on the zip codes would have given better insight to my clients as it is an important factor for them.

- - Being able to differentiate between the type of housing using zip codes will be highly beneficial.
- - In the future, understanding that it is okay for coefficients to be negative even if it goes personal intuition is okay.