Kings County Housing Analysis

●●●

Bedrooms

As bedrooms increased, the average price of a home also increased. Thought this was important, so I wanted to include this in the model.
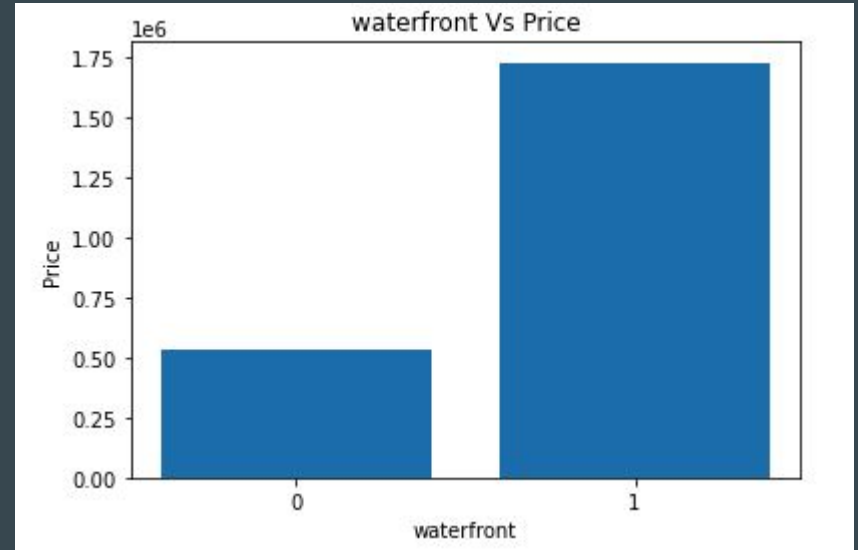
p-value=0.0

Waterfront

A waterfront view makes a difference. If there is a waterfront view, you should expect a higher average price than if there wasn't a waterfront view.
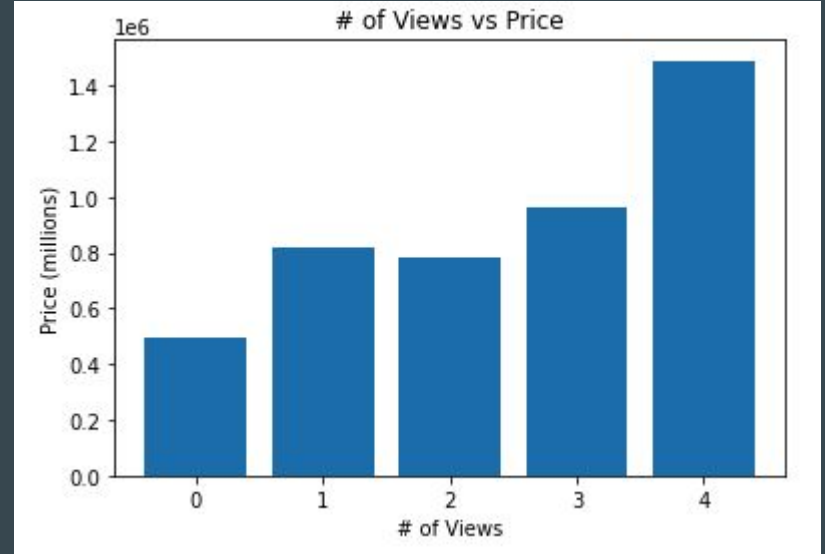
P-value = 1.677044845201236e-309

Views

Both if they house has been viewed and the amount of times the house has been viewed both show a strong positive relationship with the average price.

pvalue=0.0

Grade Quality

If the grade given to the housing unit is high, the average price tends to be high as well.
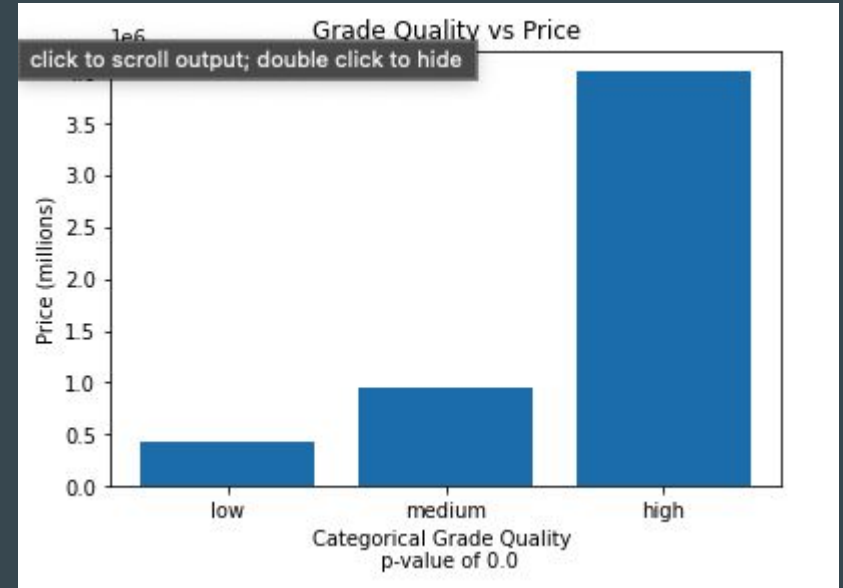
Grade Quality Ranges:

Low   ->   0-8

Med   ->   9-12

High   ->   13

pvalue=0.0

Seasons

Hard to see but there is a difference in average price and seasons where Spring is the highest, Summer is next, Fall, and then Winter.
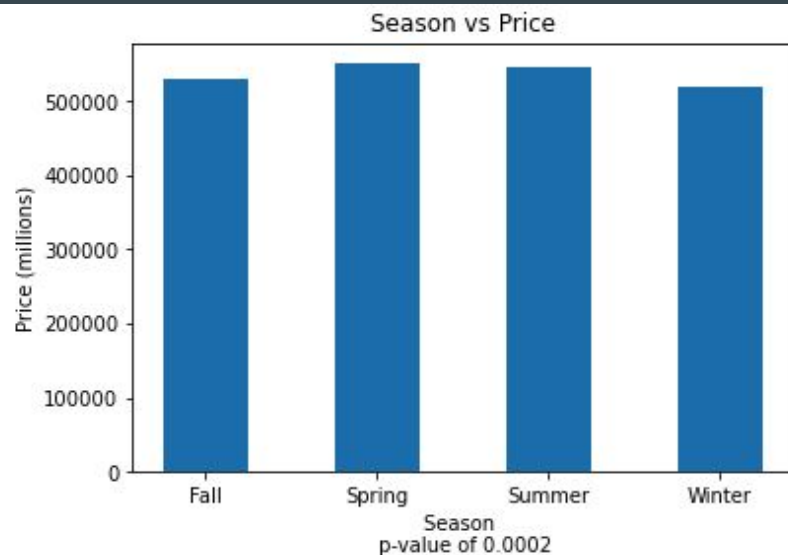
Average Prices:

Fall $528,954.88

Spring $550,809.70

Summer $545,380.28

Winter $519,066.99

Zip Codes

Lastly, the average price across Zip Codes is different as well so this should be a factor in our model as well.

Zip Codes ['98039', '98004', '98040', '98112'] have the highest average price for a home.

Results

Using this model, about 71% of the target variable price can be explained by these features.

OLS Regression Results			
Dep. Variable:	price	**R-squared:**	0.714
Model:	OLS	**Adj. R-squared:**	0.713
Method:	Least Squares	**F-statistic:**	522.1
Date:	Wed, 02 Jun 2021	**Prob (F-statistic):**	0.00
Time:	04:23:22	**Log-Likelihood:**	-2.3345e+05
No. Observations:	17212	**AIC:**	4.671e+05
Df Residuals:	17129	**BIC:**	4.677e+05
Df Model:	82		
Covariance Type:	nonrobust		

zipcode_98178	-3278.0511
zipcode_98188	2.186e+04
zipcode_98198	-1.904e+04
zipcode_98199	4.118e+05
bedrooms_2	5.453e+04
bedrooms_3	1.259e+05
bedrooms_4	1.975e+05
bedrooms_5	2.56e+05
bedrooms_6	2.766e+05
bedrooms_7	3.862e+05
season_sold_Spring	2.418e+04
season_sold_Summer	1.073e+04
season_sold_Winter	8662.5421
waterfront	6.505e+05
view	9.256e+04
grade_cat_low	-2.075e+06
grade_cat_medium	-1.765e+06

Results 2:

If we also include sqft_living to the previous model, we see that 71% getting bumped up to 80%.

Dep. Variable:	price	R-squared:	0.806
Model:	OLS	Adj. R-squared:	0.805
Method:	Least Squares	F-statistic:	856.3
Date:	Wed, 02 Jun 2021	Prob (F-statistic):	0.00
Time:	11:51:49	Log-Likelihood:	-2.3012e+05
No. Observations:	17212	AIC:	4.604e+05
Df Residuals:	17128	BIC:	4.611e+05
Df Model:	83		
Covariance Type:	nonrobust		

zipcode_98178	2.542e+04
zipcode_98188	3.416e+04
zipcode_98198	4880.5078
zipcode_98199	3.948e+05
bedrooms_2	-3082.1417
bedrooms_3	-1.556e+04
bedrooms_4	-4.073e+04
bedrooms_5	-6.081e+04
bedrooms_6	-9.841e+04
bedrooms_7	-1.293e+05
season_sold_Spring	2.758e+04
season_sold_Summer	8592.8148
season_sold_Winter	8261.8756
waterfront	6.225e+05
view	5.841e+04
grade_cat_low	-1.304e+06
grade_cat_medium	-1.185e+06
sqft_living	198.2551

Comparing coefficients

zipcode_98178	-3278.0511
zipcode_98188	2.186e+04
zipcode_98198	-1.904e+04
zipcode_98199	4.118e+05
bedrooms_2	5.453e+04
bedrooms_3	1.259e+05
bedrooms_4	1.975e+05
bedrooms_5	2.56e+05
bedrooms_6	2.766e+05
bedrooms_7	3.862e+05
season_sold_Spring	2.418e+04
season_sold_Summer	1.073e+04
season_sold_Winter	8662.5421
waterfront	6.505e+05
view	9.256e+04
grade_cat_low	-2.075e+06
grade_cat_medium	-1.765e+06

71%

(Without Sqft_living)

80%

(With Sqft_living)

zipcode_98178	2.542e+04
zipcode_98188	3.416e+04
zipcode_98198	4880.5078
zipcode_98199	3.948e+05
bedrooms_2	-3082.1417
bedrooms_3	-1.556e+04
bedrooms_4	-4.073e+04
bedrooms_5	-6.081e+04
bedrooms_6	-9.841e+04
bedrooms_7	-1.293e+05
season_sold_Spring	2.758e+04
season_sold_Summer	8592.8148
season_sold_Winter	8261.8756
waterfront	6.225e+05
view	5.841e+04
grade_cat_low	-1.304e+06
grade_cat_medium	-1.185e+06
sqft_living	198.2551