

Unsung Agency House Modelling Project



Agenda:

- Go through the business problem.
- Give the benefits that could be gotten by solving the business problem.
- Highlight the findings from the analysis and modelling done.
- Give some recommendations and a conclusion from the findings.

Business problem:

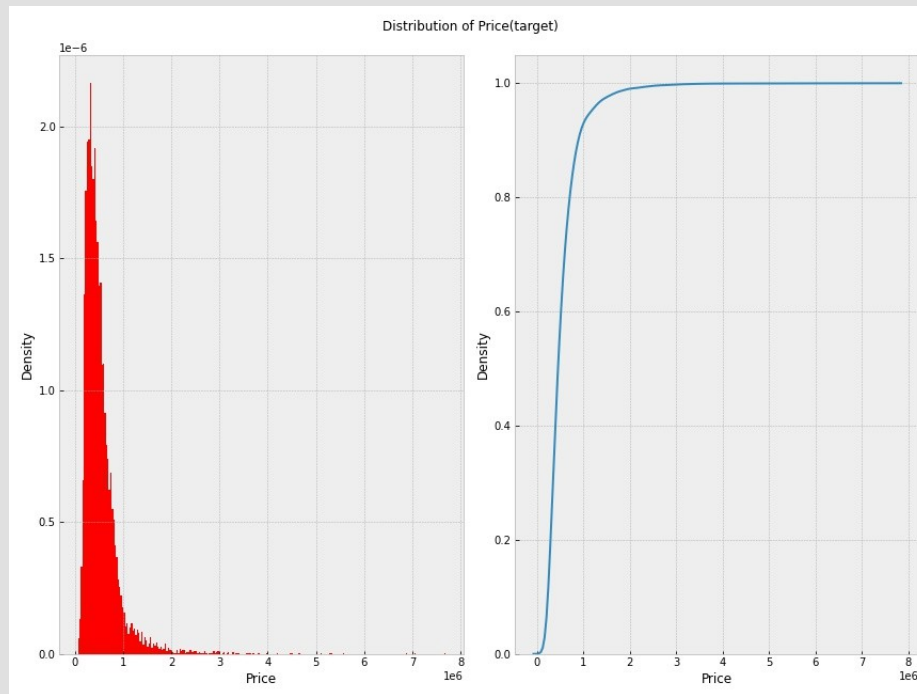
- Unsung agency is in need of a model that would help predict the price of a home based on the sort of attributes or features a home has.
- A more transparent valuation of specific features in a home is also required.

What sort of benefits would the model bring?

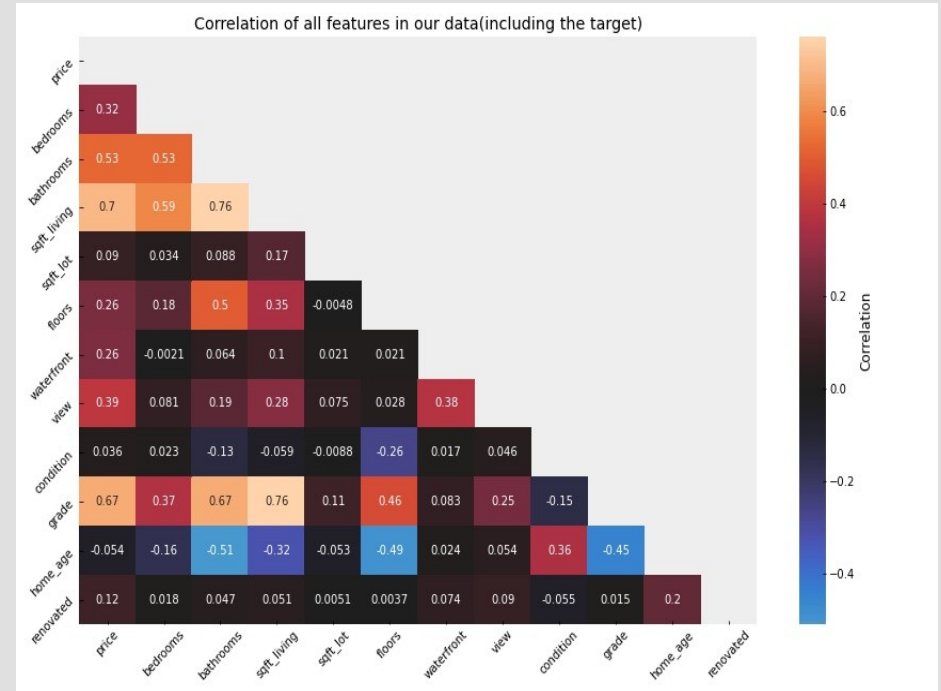
- I The agency can save on the time it takes to determine the price of a house because with the model, this process will be automated.
- II Factual advice about what attributes increase the valuation of a house can be given to clients seeking to either buy or sell a home. The best part is an approximate figure is attached to each value.

What are some of the things we found out?

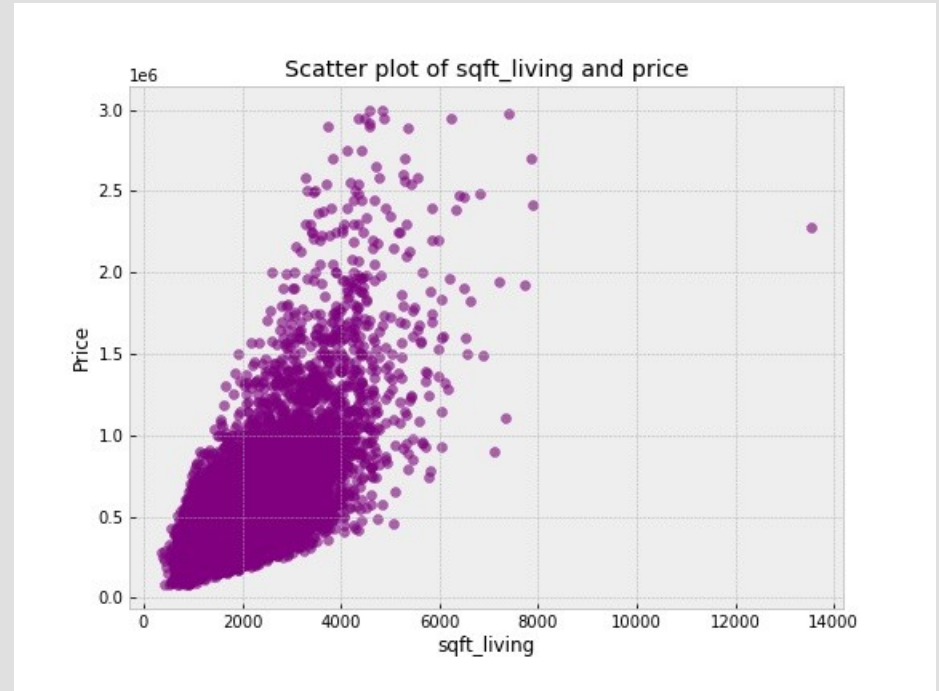
- Most homes in King County are priced between 78,000 and 1 million dollars.
- Houses that are priced past 3 million are very little in number.



- The top correlated features to price are:
 - Sqft living – area of a house
 - Grade – King County grading system
 - Bathrooms
 - Bedrooms



- Sqft living is the most correlated feature of a home to price.
- This ideally means that the larger a home is, the higher the price is going to be.



- This is the baseline model.
- It uses sqft living as the only feature that determines the price of a house.
- The model score is 37.8% which is very low so we have to try and create a model with a better score.

Dep. Variable:	price	R-squared:	0.378
Model:	OLS	Adj. R-squared:	0.378
Method:	Least Squares	F-statistic:	9178.
Date:	Mon, 04 Jul 2022	Prob (F-statistic):	0.00
Time:	20:42:39	Log-Likelihood:	-2.1163e+05
No. Observations:	15117	AIC:	4.233e+05
Df Residuals:	15115	BIC:	4.233e+05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-3.475e+06	4.2e+04	-82.741	0.000	-3.56e+06	-3.39e+06
transformed_sqft_living	5.319e+05	5552.456	95.803	0.000	5.21e+05	5.43e+05

Omnibus:	12767.815	Durbin-Watson:	1.982
Prob(Omnibus):	0.000	Jarque-Bera (JB):	812994.778
Skew:	3.690	Prob(JB):	0.00
Kurtosis:	38.160	Cond. No.	137.

Dep. Variable:	price	R-squared:	0.657
Model:	OLS	Adj. R-squared:	0.657
Method:	Least Squares	F-statistic:	1379.
Date:	Mon, 04 Jul 2022	Prob (F-statistic):	0.00
Time:	20:42:59	Log-Likelihood:	-2.0712e+05
No. Observations:	15117	AIC:	4.143e+05
Df Residuals:	15095	BIC:	4.145e+05
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.588e+07	5.2e+05	30.535	0.000	1.49e+07	1.69e+07
waterfront	5.962e+05	2.3e+04	25.870	0.000	5.51e+05	6.41e+05
renovated	5.695e+04	1.02e+04	5.570	0.000	3.69e+04	7.7e+04
bedrooms_high	-1.553e+05	5.27e+04	-2.945	0.003	-2.59e+05	-5.19e+04
bedrooms_low	3.643e+04	4415.916	8.249	0.000	2.78e+04	4.51e+04
bathrooms_high	4.366e+05	4.57e+04	9.548	0.000	3.47e+05	5.26e+05
bathrooms_low	-9.081e+04	7795.700	-11.649	0.000	-1.06e+05	-7.55e+04
floors_high	8.995e+04	1.14e+04	7.883	0.000	6.76e+04	1.12e+05
floors_low	-3.287e+04	4956.060	-6.633	0.000	-4.26e+04	-2.32e+04
view_low	-5.365e+04	1.73e+04	-3.100	0.002	-8.76e+04	-1.97e+04
view_medium	-1.176e+05	1.29e+04	-9.137	0.000	-1.43e+05	-9.24e+04
view_none	-1.858e+05	1.02e+04	-18.153	0.000	-2.06e+05	-1.66e+05
condition_high	3.125e+04	4167.747	7.498	0.000	2.31e+04	3.94e+04
condition_low	-3.907e+04	1.84e+04	-2.121	0.034	-7.52e+04	-2959.176
grade_excellent	8.619e+05	2.93e+04	29.387	0.000	8.04e+05	9.19e+05
grade_fair	-1.267e+05	7012.973	-18.063	0.000	-1.4e+05	-1.13e+05
grade_good	2.809e+05	8611.499	32.622	0.000	2.64e+05	2.98e+05
grade_poor	-3.029e+05	5.36e+04	-5.649	0.000	-4.08e+05	-1.98e+05
scaled_home_age	2.955e+05	9618.919	30.724	0.000	2.77e+05	3.14e+05
transformed_sqft_living	-4.387e+06	1.39e+05	-31.457	0.000	-4.66e+06	-4.11e+06
transformed_sqft_lot	-3.82e+04	2316.571	-16.491	0.000	-4.27e+04	-3.37e+04
transformed_sqft_living_sq	3.19e+05	9338.724	34.156	0.000	3.01e+05	3.37e+05

Omnibus:	7289.172	Durbin-Watson:	2.012
Prob(Omnibus):	0.000	Jarque-Bera (JB):	189210.711
Skew:	1.765	Prob(JB):	0.00
Kurtosis:	19.969	Cond. No.	1.80e+04

- This a summary of the final model.
- The model's score is 65.7% meaning that it covers 65.7% of the variability in price of homes.
- A basic interpretation of it is:
 - The base price of a house is 15.8 million dollars.
 - From the base price adjustments to the price are made based on the present attributes in a house.

For example: Having a high number of bathrooms in a home increases the price by 436,600 dollars while having a low number of bathrooms in a home, reduces the price by 90,810 dollars.

Final recommendations:

- More investigation into the relation of certain features to price should be done to have a better understanding of how said features affect the price of a house.
- A more complete dataset where we don't have to fill in values would be much better because with such data a more accurate model could be made.