INVESTOR UPDATE

King County Property Investment

Overview

Evaluate house price

King County Evaluation

Geographical House price distribution

Methodology

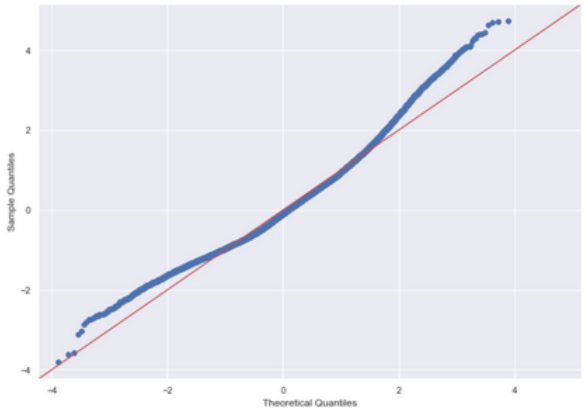
- Train/test split
- Perform feature selection
- Regression modelling
- Test for regression assumptions
  - normality
  - heteroscedasticity
  - independence
- Execute final model

When it comes to valuing a property, what features of real estate can best predict the price?

Dep. Variable:	price	**R-squared:**	0.412
Model:	OLS	**Adj. R-squared:**	0.412
Method:	Least Squares	**F-statistic:**	7102.
Date:	Fri, 27 Mar 2020	**Prob (F-statistic):**	0.00
Time:	11:44:26	**Log-Likelihood:**	-2.7330e+05
No. Observations:	20292	**AIC:**	5.466e+05
Df Residuals:	20289	**BIC:**	5.466e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	5.022e+05	1207.570	415.882	0.000	5e+05	5.05e+05
bedbath	-2.182e+04	1878.224	-11.620	0.000	-2.55e+04	-1.81e+04
sqft_living	1.82e+05	2071.778	87.862	0.000	1.78e+05	1.86e+05

Omnibus:	1539.437	**Durbin-Watson:**	1.972
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	2030.286
Skew:	0.670	**Prob(JB):**	0.00
Kurtosis:	3.778	**Cond. No.**	2.66



Size!

- Living Space
- Lot Size

Bedrooms

Bathrooms

When it comes to valuing a property, what features of real estate can best predict the price?

Dep. Variable:	price	**R-squared:**	0.634
Model:	OLS	**Adj. R-squared:**	0.632
Method:	Least Squares	**F-statistic:**	332.3
Date:	Fri, 27 Mar 2020	**Prob (F-statistic):**	0.00
Time:	11:42:46	**Log-Likelihood:**	-42468.
No. Observations:	3086	**AIC:**	8.497e+04
Df Residuals:	3069	**BIC:**	8.507e+04
Df Model:	16		
Covariance Type:	nonrobust		

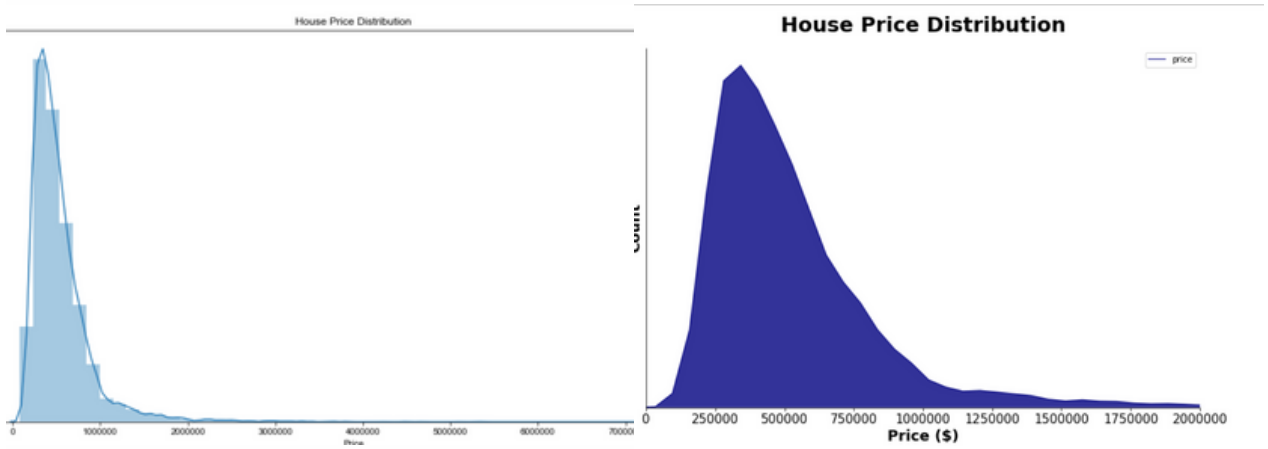
	coef	std err	t	P>|t|	[0.025	0.975]
const	5.392e+05	4138.091	130.302	0.000	5.31e+05	5.47e+05
sqft_above	1.314e+05	7668.352	17.133	0.000	1.16e+05	1.46e+05
sqft_basement	8.855e+04	4756.422	18.617	0.000	7.92e+04	9.79e+04
yr_renovated	3.21e+04	4191.354	7.659	0.000	2.39e+04	4.03e+04
sqft_living15	3.306e+04	6693.876	4.938	0.000	1.99e+04	4.62e+04
sqft_lot15	-2.307e+04	4272.581	-5.400	0.000	-3.14e+04	-1.47e+04
floor_2	2832.0116	5361.784	0.528	0.597	-7681.038	1.33e+04
floor_3	1.369e+04	4313.084	3.173	0.002	5228.588	2.21e+04
wtrfrnt_1	3.027e+04	4834.286	6.261	0.000	2.08e+04	3.97e+04
view_1	2.565e+04	4194.109	6.116	0.000	1.74e+04	3.39e+04
view_2	2.122e+04	4240.295	5.004	0.000	1.29e+04	2.95e+04
view_3	2.873e+04	4289.742	6.697	0.000	2.03e+04	3.71e+04
view_4	3.874e+04	5053.529	7.666	0.000	2.88e+04	4.86e+04
grade_10	5.867e+04	4712.776	12.448	0.000	4.94e+04	6.79e+04
grade_11	7.115e+04	4628.920	15.370	0.000	6.21e+04	8.02e+04
grade_12	3.504e+04	4397.958	7.967	0.000	2.64e+04	4.37e+04
grade_13	7.461e+04	4314.845	17.290	0.000	6.61e+04	8.31e+04

Size!

- Square foot basement
- Square food living

Grade

In King County as a whole, how are housing sales prices distributed?

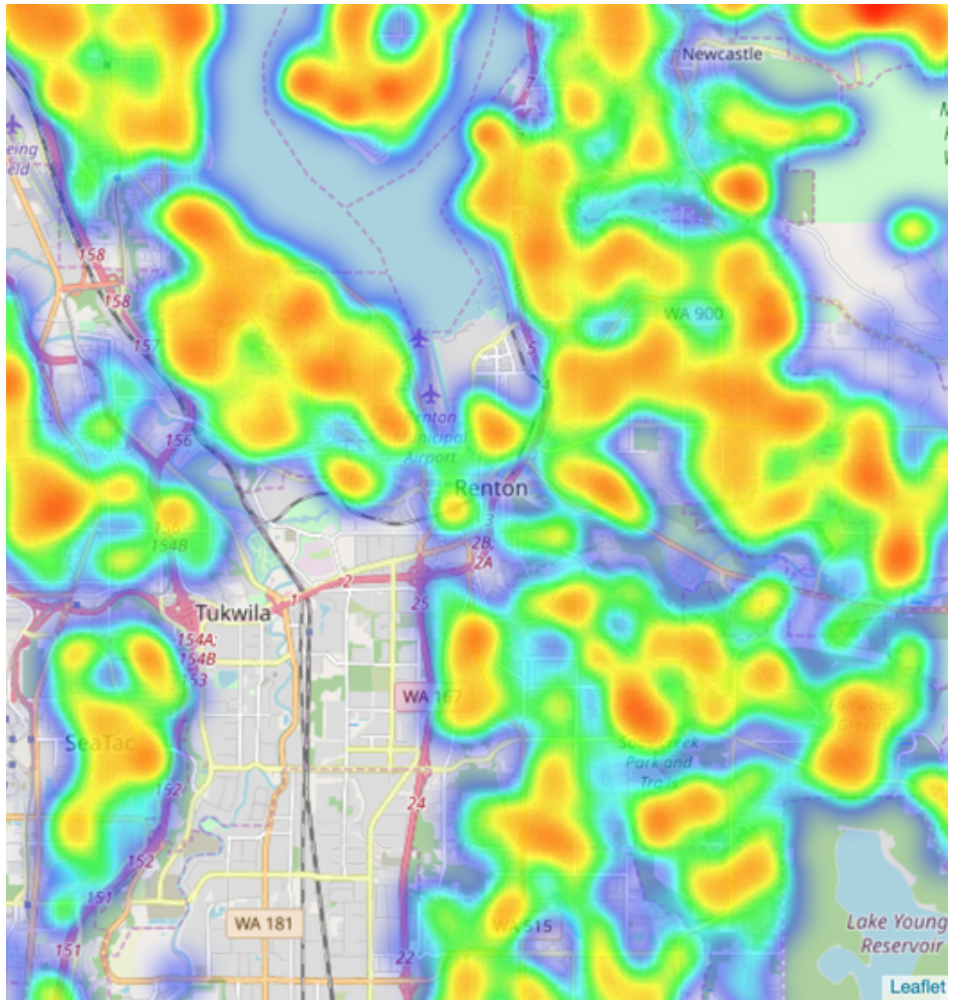


Normal Distribution

Let's talk about Outliers...

What does this tell us?

How are they distributed geographically?



Area

Lattitude, Longitude

Price

Visual Distribution

Closer Look Into Feature Variables

```
In [1099]: # What were the top five zipcodes with the largest volume of sales within our data?
           df.zipcode.value_counts().head()

Out[1099]: 98038    425
           98103    415
           98052    410
           98042    408
           98117    394
           Name: zipcode, dtype: int64
```

Top 5 zip codes with the largest volume of sales (turnover of real estate)

Key Takeaways:

Overall Assumption with key variables: Don't assume!

- Really cool place to live!
- Tourism
- Economy *was* good

Meet the Team

Anila

Data Scientist

Former Waterslide Tester

Marwan Kalo

Data Scientist

Former Hypnotherapist

References

- https://bit.ly/39mjB5f
- https://bit.ly/33PtF5p
- https://bit.ly/2UIUy79