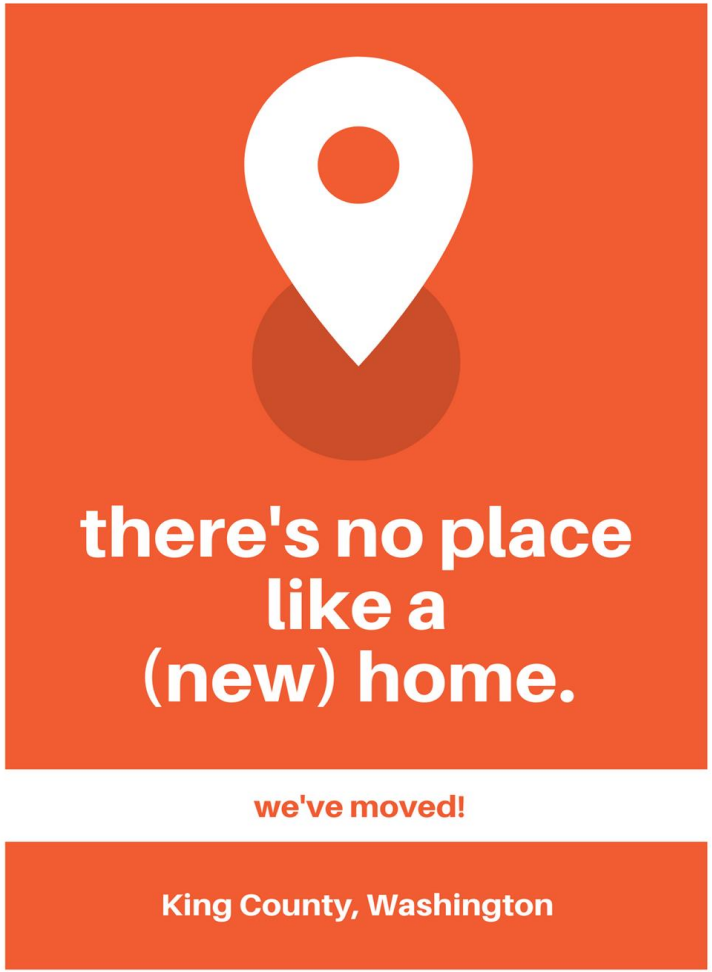


House Price Prediction in — King County, Washington.

Pengju Sun
03/22/2021

D·R·HORTON®

America's Builder

Business Problem

King County House Data

Methods

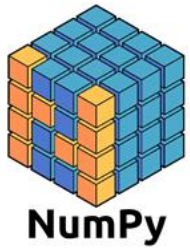

Pandas


matplotlib

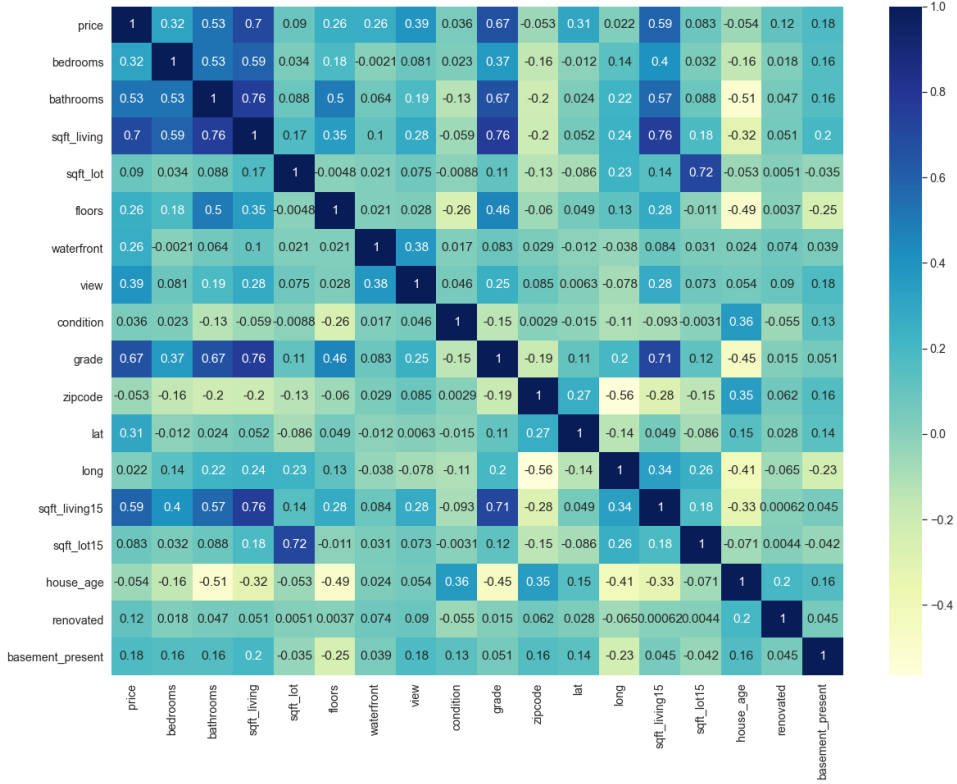

statsmodels


seaborn


scikit
learn

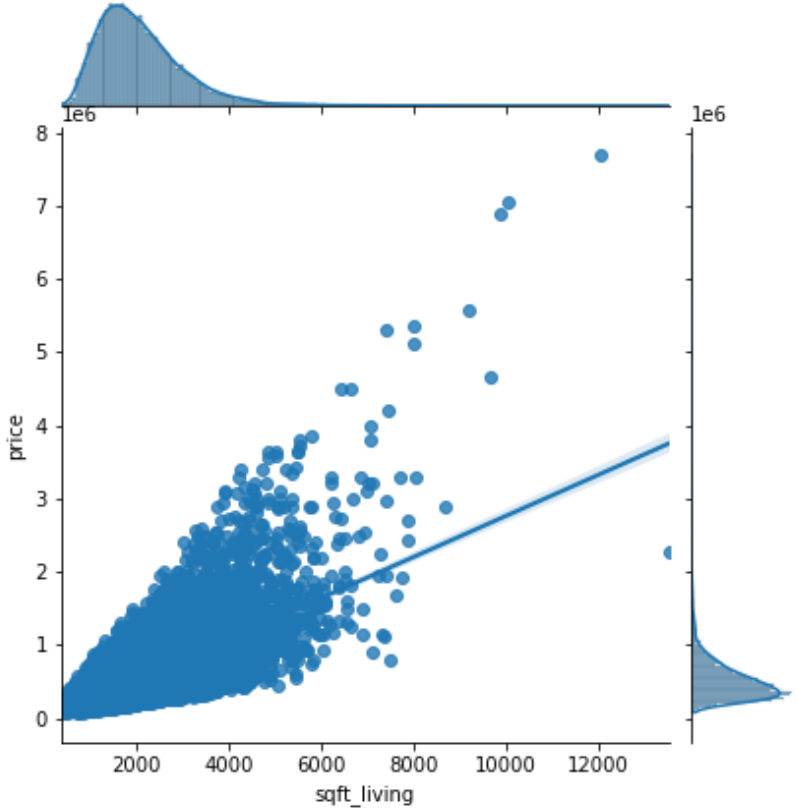

NumPy

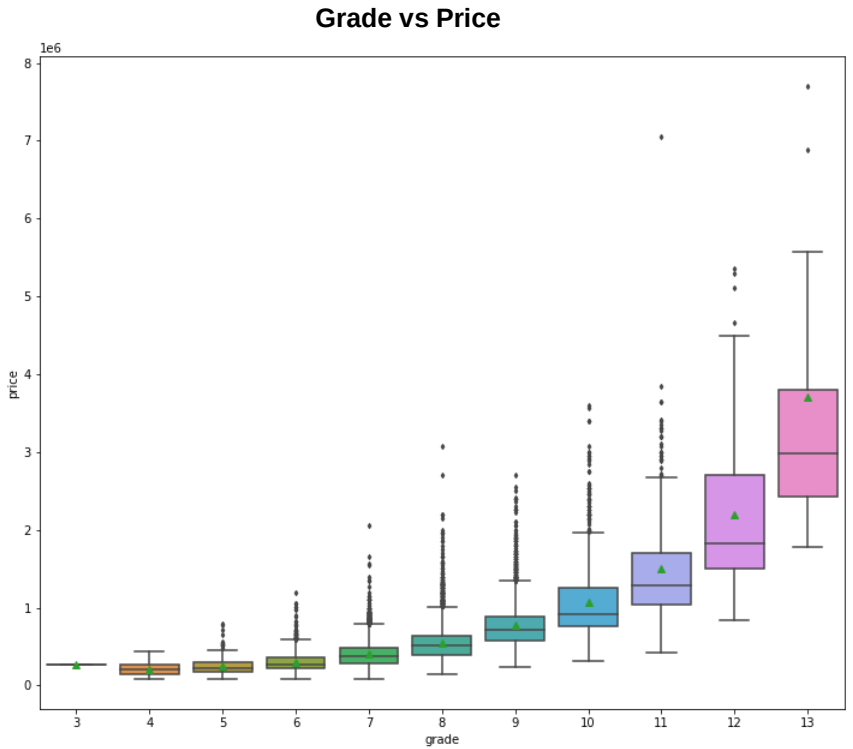
Results



Features	Correlation
price	1.000000
sqft_living	0.701917
grade	0.667951
sqft_above	0.605368
sqft_living15	0.585241
bathrooms	0.525906
view	0.393497
sqft_basement	0.321108
bedrooms	0.308787
lat	0.306692
waterfront	0.264306
floors	0.256804
yr_renovated	0.117855
sqft_lot	0.089876
sqft_lot15	0.082845
condition	0.036056
long	0.022036
zipcode	-0.053402
house_age	-0.053890

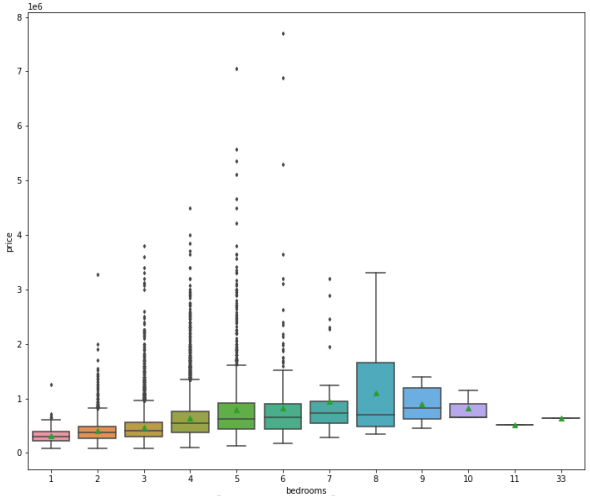
Results


1e6
1e6
price
sqft_living
0
1
2
3
4
5
6
7
8
2000
4000
6000
8000
10000
12000

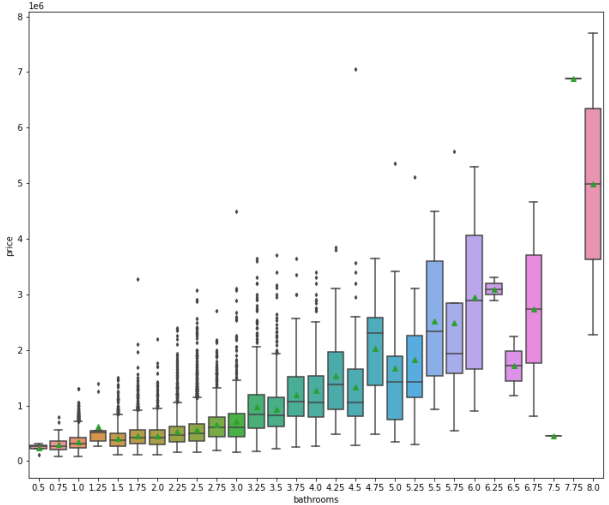

Grade vs Price
1e6
price
grade
0
1
2
3
4
5
6
7
8
3
4
5
6
7
8
9
10
11
12
13

Results

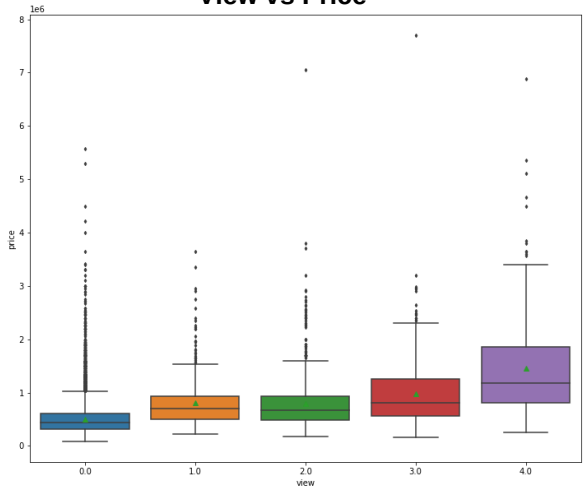
Bedrooms vs Price



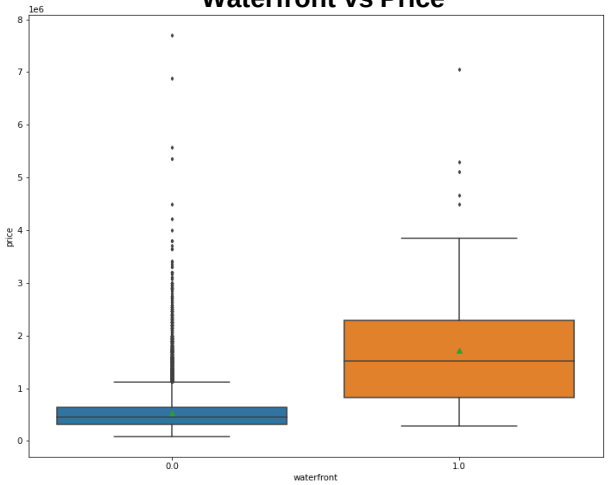
Bathrooms vs Price



View vs Price



Waterfront vs Price



Results

Multiple Linear Regression Model:

R2 = 0.763
MAE = 110K

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.763
Model:	OLS	**Adj. R-squared:**	0.763
Method:	Least Squares	**F-statistic:**	2225.
Date:	Sat, 20 Mar 2021	**Prob (F-statistic):**	0.00
Time:	21:03:28	**Log-Likelihood:**	-996.24
No. Observations:	17277	**AIC:**	2044.
Df Residuals:	17251	**BIC:**	2246.
Df Model:	25		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.0468	0.002	6685.045	0.000	13.043	13.051
bedrooms	-0.0121	0.003	-4.637	0.000	-0.017	-0.007
bathrooms	0.0571	0.004	15.713	0.000	0.050	0.064
sqft_living	0.1429	0.005	30.710	0.000	0.134	0.152
sqft_lot	0.0199	0.003	6.947	0.000	0.014	0.026
floors	0.0443	0.003	16.397	0.000	0.039	0.050
condition	0.0414	0.002	19.076	0.000	0.037	0.046
zipcode	-0.0417	0.003	-16.559	0.000	-0.047	-0.037
lat	0.1983	0.002	93.970	0.000	0.194	0.202
long	-0.0285	0.003	-10.841	0.000	-0.034	-0.023
sqft_living15	0.0748	0.003	22.177	0.000	0.068	0.081
sqft_lot15	-0.0083	0.003	-2.858	0.004	-0.014	-0.003
house_age	0.0874	0.003	29.273	0.000	0.082	0.093
wf_1.0	0.0312	0.002	13.141	0.000	0.027	0.036
vw_1.0	0.0200	0.002	10.113	0.000	0.016	0.024
vw_2.0	0.0241	0.002	11.945	0.000	0.020	0.028
vw_3.0	0.0245	0.002	12.094	0.000	0.021	0.028
vw_4.0	0.0308	0.002	12.712	0.000	0.026	0.036
gd_8	0.0792	0.002	32.002	0.000	0.074	0.084
gd_9	0.1105	0.003	40.030	0.000	0.105	0.116
gd_10	0.0983	0.003	36.368	0.000	0.093	0.104
gd_11	0.0665	0.002	26.720	0.000	0.062	0.071
gd_12	0.0339	0.002	15.486	0.000	0.030	0.038
gd_13	0.0166	0.002	8.120	0.000	0.013	0.021
rn_1	0.0156	0.002	7.567	0.000	0.012	0.020
bs_1	0.0235	0.002	9.936	0.000	0.019	0.028

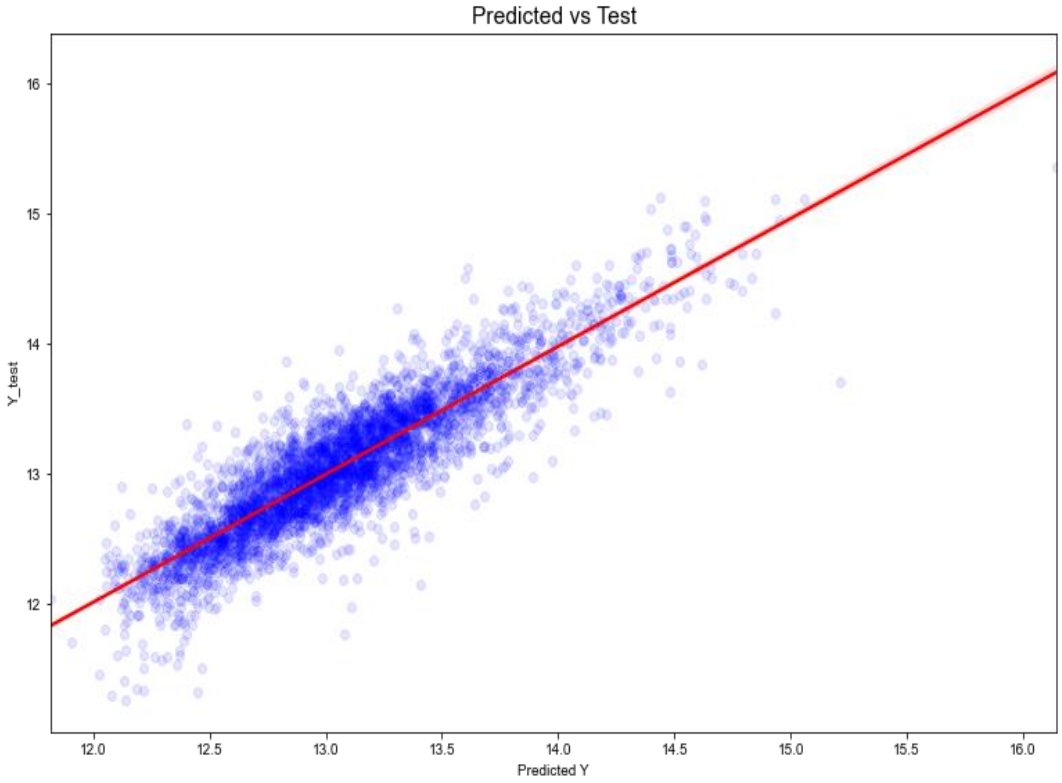
Omnibus:	357.067	**Durbin-Watson:**	1.990
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	725.406
Skew:	-0.102	**Prob(JB):**	3.02e-158
Kurtosis:	3.983	**Cond. No.**	5.83

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

Results

Polynomial Regression Model:

R2 = 0.79
MAE = 100K



Recommendation

The recommendations are as follows:

- Increase square-footage of living space.
- Increase the number of bedrooms and bathrooms as the square feet of houses increase.
- Attain the highest possible building grade
- Build and develop homes with waterfront
- Build and develop homes with good view

Future Work

- Reduce noise in the data to improve the accuracy of the models
- Investigate on the location importance of the houses.
- Investigate certain features, such as constructional/architectural values of house, to see what trends we could discern from that.

Thank you!

For any questions or comments, please feel free to reach out!
Email: pjsun2012@gmail.com
GitHub: @pjsun2012