

there's no place like a (new) home.

we've moved!

**King County, Washington** 

# **House Price Prediction** — King County, Washington.

Pengju Sun 03/22/2021

# D·R·HORTON Business Problem

America's Builder



## **King County House Data**



## **Methods**

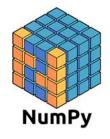










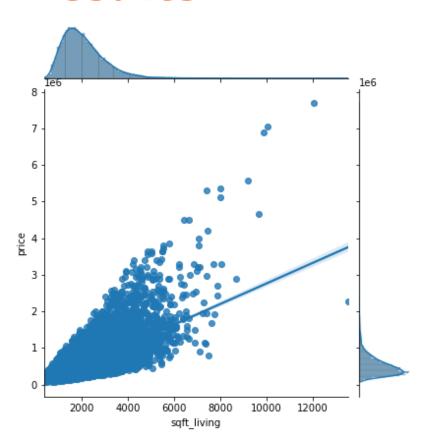


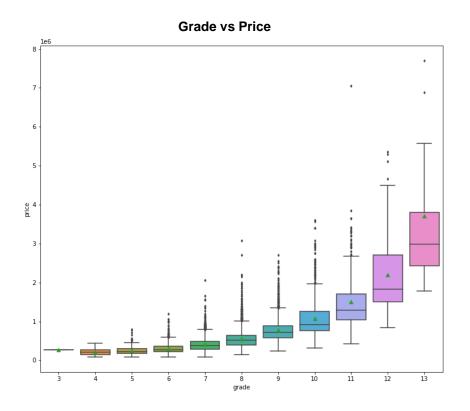
price	1	0.32	0.53	0.7	0.09	0.26	0.26	0.39	0.036	0.67	-0.053	0.31	0.022	0.59	0.083	-0.054	0.12	0.18
bedrooms	0.32	1		0.59	0.034		-0.0021	0.081	0.023		-0.16	-0.012	0.14		0.032	-0.16	0.018	0.16
bathrooms	0.53			0.76	0.088		0.064		-0.13	0.67	-0.2	0.024		0.57	0.088	-0.51	0.047	0.16
sqft_living	0.7	0.59	0.76		0.17		0.1		-0.059	0.76	-0.2	0.052		0.76		-0.32	0.051	0.2
sqft_lot	0.09	0.034	0.088	0.17	1	-0.0048	3 0.021	0.075	-0.0088	0.11	-0.13	-0.086		0.14	0.72	-0.053	0.0051	-0.035
floors	0.26				-0.0048		0.021	0.028	-0.26	0.46	-0.06	0.049	0.13		-0.011	-0.49	0.0037	-0.25
waterfront	0.26	-0.0021	0.064	0.1	0.021	0.021	1		0.017	0.083	0.029	-0.012	-0.038	0.084	0.031	0.024	0.074	0.039
view	0.39	0.081			0.075	0.028			0.046		0.085	0.0063	-0.078		0.073	0.054	0.09	0.18
condition	0.036	0.023	-0.13	-0.059	-0.0088	-0.26	0.017	0.046	1	-0.15	0.0029	-0.015	-0.11	-0.093	-0.0031		-0.055	0.13
grade	0.67	0.37	0.67	0.76	0.11	0.46	0.083		-0.15	1	-0.19	0.11	0.2	0.71	0.12	-0.45	0.015	0.051
zipcode	-0.053	-0.16	-0.2	-0.2	-0.13	-0.06	0.029	0.085	0.0029	-0.19	1	0.27	-0.56	-0.28	-0.15		0.062	0.16
lat	0.31	-0.012	0.024	0.052	-0.086	0.049	-0.012	0.0063	-0.015	0.11		1	-0.14	0.049	-0.086	0.15	0.028	0.14
long	0.022	0.14	0.22	0.24		0.13	-0.038	-0.078	-0.11	0.2	-0.56	-0.14	1	0.34		-0.41	-0.065	-0.23
sqft_living15	0.59	0.4	0.57	0.76	0.14		0.084		-0.093	0.71	-0.28	0.049		1	0.18	-0.33	0.00062	0.045
sqft_lot15	0.083	0.032	0.088		0.72	-0.011	0.031	0.073	-0.0031	0.12	-0.15	-0.086				-0.071	0.0044	-0.042
house_age	-0.054	-0.16	-0.51	-0.32	-0.053	-0.49	0.024	0.054	0.36	-0.45	0.35	0.15	-0.41	-0.33	-0.071	1	0.2	0.16
renovated	0.12	0.018	0.047	0.051	0.0051	0.0037	0.074	0.09	-0.055	0.015	0.062	0.028	-0.065	0.00062	20.0044		1	0.045
basement_present	0.18	0.16	0.16	0.2	-0.035	-0.25	0.039	0.18	0.13	0.051	0.16	0.14	-0.23	0.045	-0.042	0.16	0.045	1
	price	pedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	zipcode	lat	long	sqft_living15	sqft_lot15	house_age	renovated	asement_present

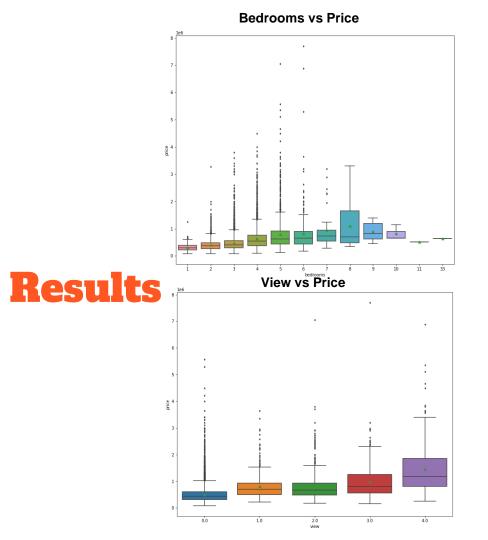
Features	Correlation
price	1.000000
sqft_living	0.701917
grade	0.667951
sqft_above	0.605368
sqft_living15	0.585241
bathrooms	0.525906
view	0.393497
sqft_basement	0.321108
bedrooms	0.308787
lat	0.306692
waterfront	0.264306
floors	0.256804
yr_renovated	0.117855
sqft_lot	0.089876
sqft_lot15	0.082845
condition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

- -0.2

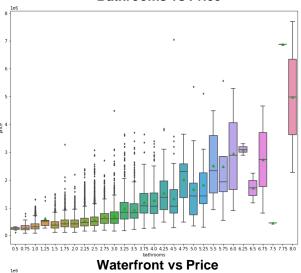
- -0.4

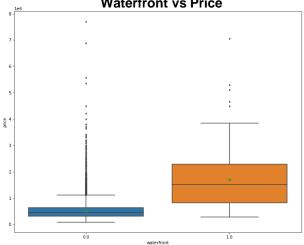












## Multiple Linear Regression Model:

**R2 = 0.763 MAE = 110K** 

OLS Regression	Results							
Dep. Varia	ble:		у	R-squared:				
Мо	del:		OLS Ad	Adj. R-squared:				
Meth	nod: L	east Squ	ares	F-statistic:				
D	ate: Sat	20 Mar 2	2021 <b>Prob</b>	(F-stat	istic):	0.00		
Ti	me:	21:0	3:28 <b>Lo</b> s	g-Likelil	nood:	-996.24		
No. Observation	ons:	17	7277		AIC:	2044		
Df Residu	als:	17	7251		BIC:	2246.		
Df Mo	del:		25					
Covariance Ty	/pe:	nonro	bust					
	coef	std err	t	P> t	[0.025	0.97		
const	13.0468	0.002	6685.045	0.000	13.043	13.05		
bedrooms	-0.0121	0.003	-4.637	0.000	-0.017	-0.00		
bathrooms	0.0571	0.004	15.713	0.000	0.050	0.06		
sqft_living	0.1429	0.005	30.710	0.000	0.134	0.15		
sqft_lot	0.0199	0.003	6.947	0.000	0.014	0.02		
floors	0.0443	0.003	16.397	0.000	0.039	0.05		
condition	0.0414	0.002	19.076	0.000	0.037	0.04		
zipcode	-0.0417	0.003	-16.559	0.000	-0.047	-0.03		
lat	0.1983	0.002	93.970	0.000	0.194	0.20		
long	-0.0285	0.003	-10.841	0.000	-0.034	-0.02		
sqft_living15	0.0748	0.003	22.177	0.000	0.068	0.08		
sqft_lot15	-0.0083	0.003	-2.858	0.004	-0.014	-0.00		
house_age	0.0874	0.003	29.273	0.000	0.082	0.09		
wf_1.0	0.0312	0.002	13.141	0.000	0.027	0.03		
vw_1.0	0.0200	0.002	10.113	0.000	0.016	0.02		
vw_2.0	0.0241	0.002	11.945	0.000	0.020	0.02		
vw_3.0	0.0245	0.002	12.094	0.000	0.021	0.02		
vw_4.0	0.0308	0.002	12.712	0.000	0.026	0.03		
gd_8	0.0792	0.002	32.002	0.000	0.074	0.08		
gd_9	0.1105	0.003	40.030	0.000	0.105	0.11		
gd_10	0.0983	0.003	36.368	0.000	0.093	0.10		
gd_11	0.0665	0.002	26.720	0.000	0.062	0.07		
gd_12	0.0339	0.002	15.486	0.000	0.030	0.03		
gd_13	0.0166	0.002	8.120	0.000	0.013	0.02		
rn_1	0.0156	0.002	7.567	0.000	0.012	0.02		
bs_1	0.0235	0.002	9.936	0.000	0.019	0.02		
Omnibus: 357.067 Durbin-Watson: 1.990								
Prob(Omnibus	i): 0.00	00 Jarq	ue-Bera (JI	3): 7	25.406			
Skev	w: -0.10	02	Prob(JI	3.0	2e-158			
Kurtosi	s: 3.98	33	Cond. N	lo.	5.83			

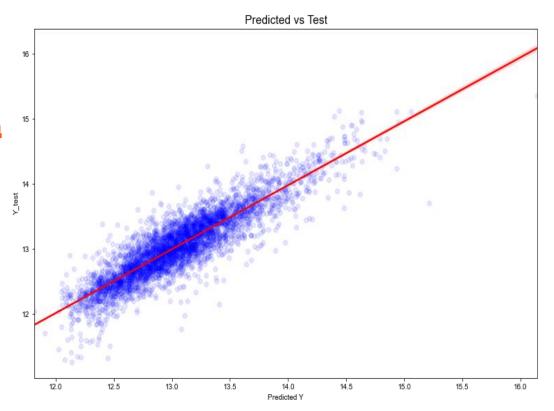
OLS Regression Results

#### Notes

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

## 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: <a href="mailto:pjsun2012@gmail.com">pjsun2012@gmail.com</a>

GitHub: @pjsun2012