How much?

...

Predicting the sale price for a house in King County

Yu Fen Lin and Karen Warmbein

Project on github

December 6, 2019

Problem

- Validate 3 claims about the sale prices of homes
  - Limit to King County sales in 2018
  - Data from King County Department of Assessments[1]

1. Claim: Higher square footage increases home sale price[2,3]
2. Claim: Having a porch increases home sale price[4,5]
3. Claim: Having a beachfront or lakefront increases home sale price[6]

The model

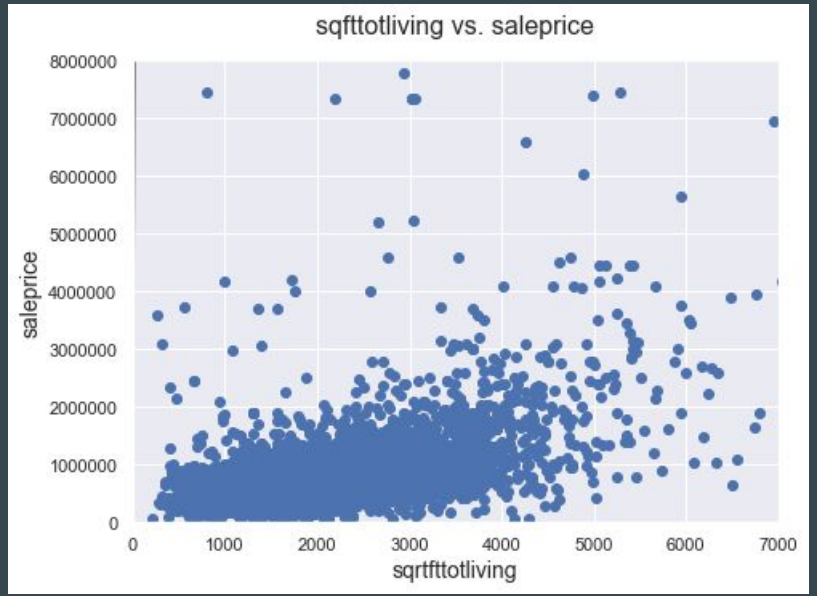
13 predictive features on ~5200 sale price records

- Square foot of total living
- Waterfront footage
- Traffic noise
- Full baths
- Additional cost
- Sewer system
- Inadequate parking
- Age of house
- Porch
- Waterfront location (4)

Does a larger house impact sale price?

Claim: Higher square footage increases home sale price

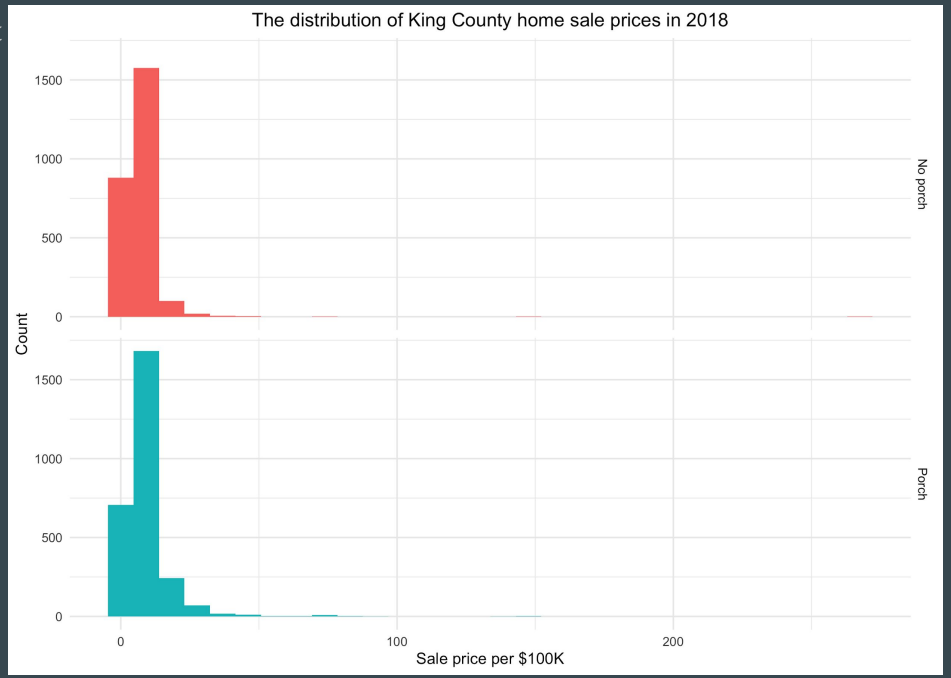
- Model: Each square foot in the house increases the sale price by $375



Does a porch increase sale price?

Claim: Having a porch increases home sale price

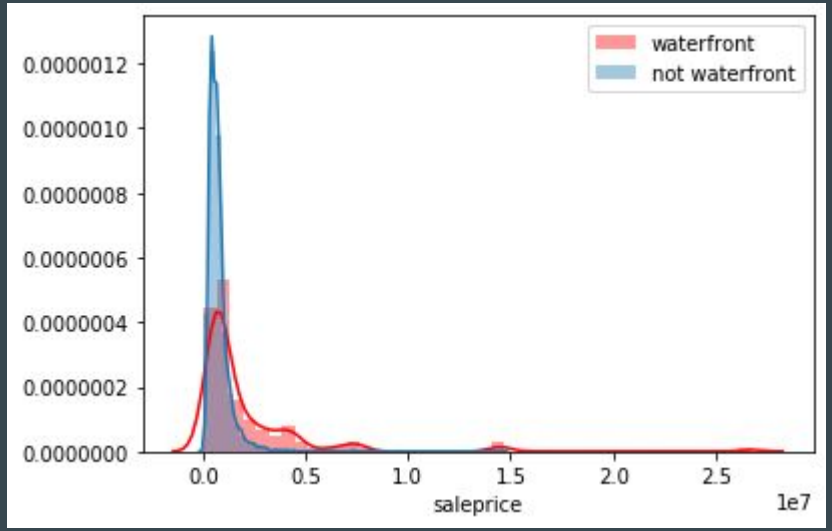
- Model: Having a porch is not statistically significant



Does waterfront property increase sale price?

Claim: Having a beachfront or lakefront increases home sale price

- Model: used 4 features
- Depends where:
  - Lake Sammamish: $3.5 million
  - Lake Washington $1.5 million
  - Puget Sound: not statistically significant
  - Other: decreased by $390,000



Recommendations

- Porch vs additional room
  - Think: bigger house
- Location by Lake Sammamish
  - Increase sale price

Next Steps

- Investigate adding more predictive features
  - Education/Schools
  - Zip codes
- Does our model violate assumptions of linearity?
  - Or, do we have unreasonable outliers in our data
- Sale price appears to be 2 groups, does our model work on both (separately)?
- House age vs other features - dependent?
- Consider economy in 2018 with results from another year

Questions?

Appendix

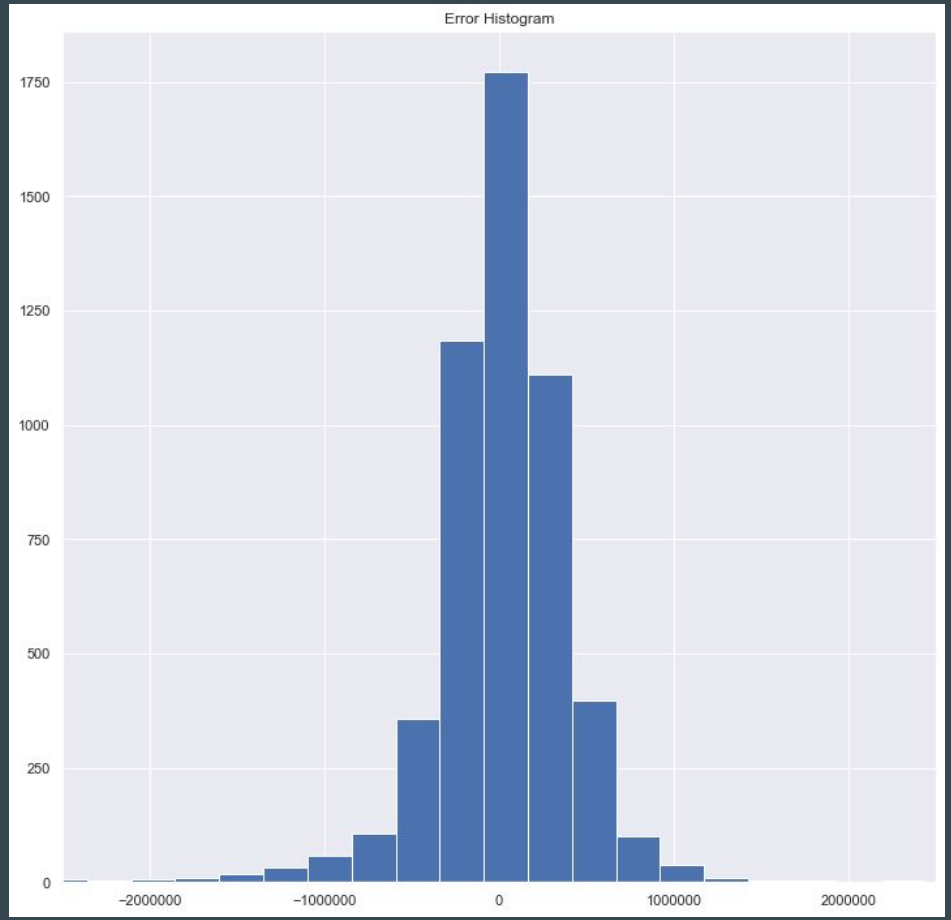
- Data reduction
- Assumptions of a linear model
- Model - Linear Regression analysis results
- Sources

Data Reduction

- Data from 3 csv files to db
  - Parcel
  - Residential building
  - Real property sales
- Criteria used:
  - Exclude empty zip code fields
  - Principal use defined as residential
  - Sale document listed as 2018
  - Exclude data where home sale price is 0
- Reduced from ~205,000 records to ~5,200

Assumptions?

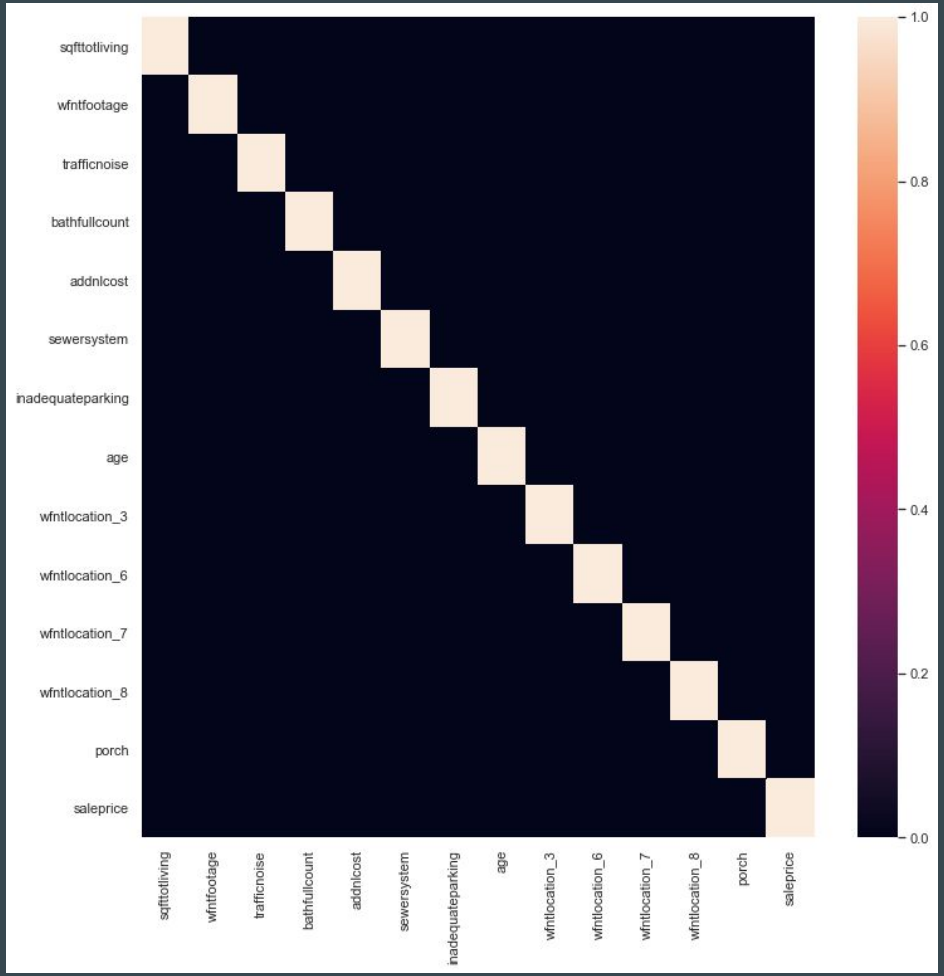
1. Normality of the residual distribution?



Assumptions?

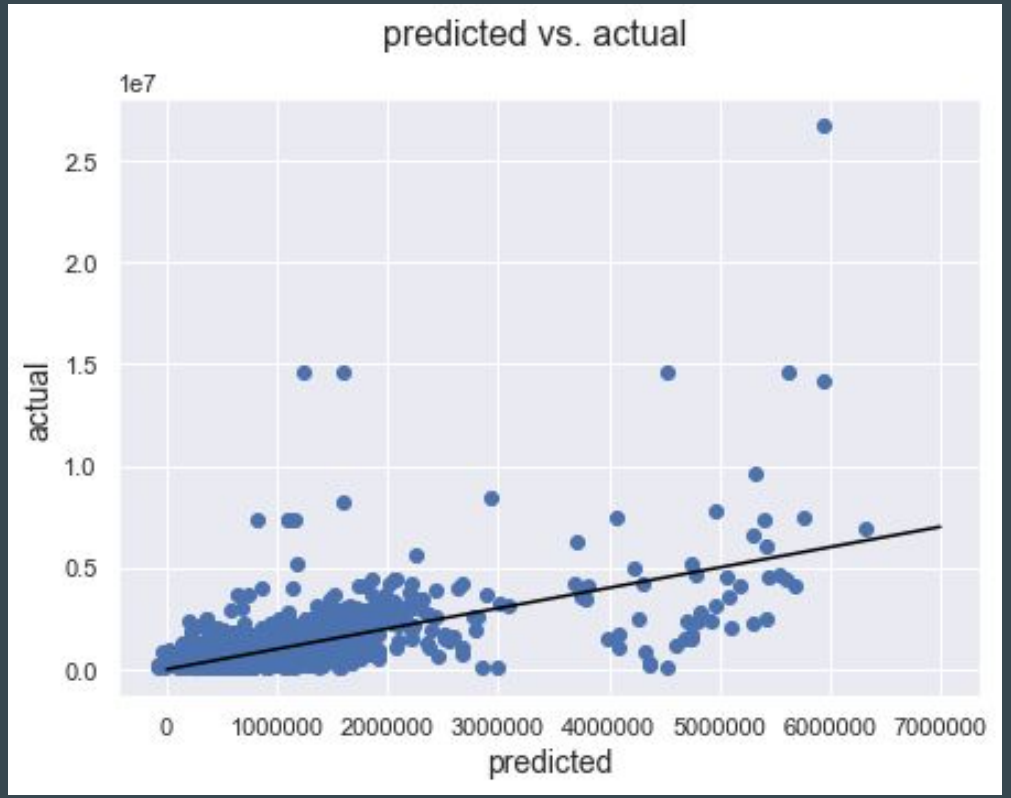
2. Multicollinearity violated?

- All our features are independent, none have a high (> 0.7) correlation
- Found that building grade and square feet of living were highly correlated, so we removed building grade from our model



Assumptions?

3. Linear model? Homoscedastic?



Model - Linear Regression Results

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.397
Model:	OLS	**Adj. R-squared:**	0.395
Method:	Least Squares	**F-statistic:**	265.5
Date:	Thu, 05 Dec 2019	**Prob (F-statistic):**	0.00
Time:	17:01:45	**Log-Likelihood:**	-78018.
No. Observations:	5261	**AIC:**	1.561e+05
Df Residuals:	5247	**BIC:**	1.562e+05
Df Model:	13		
Covariance Type:	nonrobust		

Omnibus:	8882.401	**Durbin-Watson:**	1.884
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	16032575.867
Skew:	11.133	**Prob(JB):**	0.00
Kurtosis:	272.523	**Cond. No.**	6.68e+04

	coef	std err	t	P>|t|	[0.025	0.975]
const	-4.708e+05	4.98e+04	-9.462	0.000	-5.68e+05	-3.73e+05
sqfttotliving	374.9207	11.270	33.268	0.000	352.827	397.014
wfntfootage	2986.0062	542.663	5.503	0.000	1922.161	4049.852
trafficnoise	1.756e+04	1.47e+04	1.194	0.233	-1.13e+04	4.64e+04
bathfullcount	-2.149e+04	1.62e+04	-1.324	0.186	-5.33e+04	1.03e+04
addnlcost	-3.0561	3.665	-0.834	0.404	-10.240	4.128
sewersystem	2.31e+05	2.14e+04	10.814	0.000	1.89e+05	2.73e+05
inadequateparking	2.375e+04	9478.805	2.506	0.012	5171.179	4.23e+04
age	1122.5276	329.606	3.406	0.001	476.363	1768.692
wfntlocation_3	-4.719e+04	1.1e+05	-0.428	0.669	-2.63e+05	1.69e+05
wfntlocation_6	3.454e+06	1.12e+05	30.914	0.000	3.24e+06	3.67e+06
wfntlocation_7	1.495e+06	2.06e+05	7.242	0.000	1.09e+06	1.9e+06
wfntlocation_8	-3.855e+05	1.35e+05	-2.851	0.004	-6.51e+05	-1.2e+05
porch	3.618e+04	1.91e+04	1.891	0.059	-1328.356	7.37e+04

Sources

1. Assessment Data: https://info.kingcounty.gov/assessor/DataDownload/default.aspx
2. Gomez, J. 2019. "8 critical factors that influence a home's value". OpenDoor. Available at: https://www.opendoor.com/w/blog/factors-that-influence-home-value
3. Buczynski, B. 2019. "5 Proven Ways to Increase Home Value". NerdWallet. Available at: https://www.nerdwallet.com/blog/mortgages/how-to-increase-home-value/
4. Taylor, A.B. 2019. "11 Features That Will Sell Your Home Faster". Kiplinger. Available at: https://www.kiplinger.com/slideshow/real-estate/T010-S001-home-features-today-s-buyers-want-most/index.html
5. Crow, S. 2019. "50 Clever Ways to Instantly Add Value to Your Home". BestLife. Available at: https://bestlifeonline.com/home-value-upgrades/
6. Unknown author. 2018. "5 Features That Make A Property Valuable". House Flipping School. Available at: https://houseflippingschool.com/5-features-valuable/