MODULE 1 PROJECT

King County Housing

By Desmond Webb

King County Housing

- 21597 entries

- 21 variables

- May 2014 to May 2015

- Living space

- Lot space

- Neighborhood scale

Pricing Model

- Purpose:
  - *Predicting home price within King county*
  - *Exploratory data analysis*
- Predictors
  - *14 chosen*
- Hypothesis: Home price is predicted by selected predictors
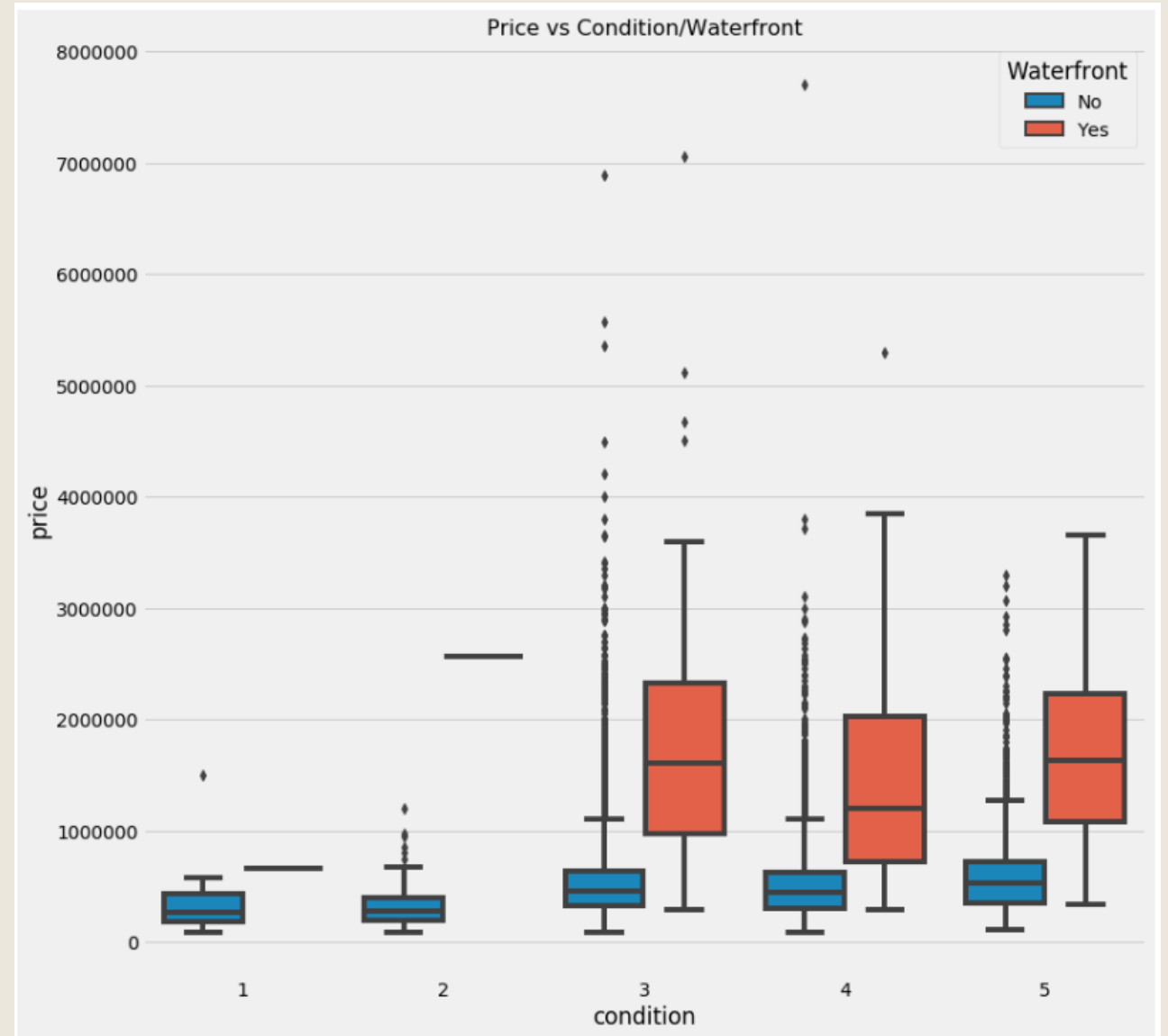
Findings

- Hypothesis: supported
  - *p-value can be trusted with caution*
- Large conditional number
  - *too much noise*
- Positive trend (price)

OLS Regression Results

Dep. Variable:	price	R-squared:	0.665
Model:	OLS	Adj. R-squared:	0.665
Method:	Least Squares	F-statistic:	3058.
Date:	Mon, 08 Jul 2019	Prob (F-statistic):	0.00
Time:	09:01:15	Log-Likelihood:	-2.9559e+05
No. Observations:	21597	AIC:	5.912e+05
Df Residuals:	21582	BIC:	5.913e+05
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.27e+07	3.54e+06	9.231	0.000	2.58e+07	3.96e+07
bedrooms	-1.457e+05	6390.217	-22.808	0.000	-1.58e+05	-1.33e+05
sqft_living	318.8158	2.830	112.640	0.000	313.268	324.364
sqft_lot	-2.022e+04	4225.777	-4.784	0.000	-2.85e+04	-1.19e+04
floors	1.139e+04	6108.135	1.865	0.062	-583.409	2.34e+04
waterfront	5.96e+05	1.92e+04	31.075	0.000	5.58e+05	6.34e+05
view	7.098e+04	2193.487	32.362	0.000	6.67e+04	7.53e+04
condition	3.114e+04	2477.779	12.569	0.000	2.63e+04	3.6e+04
sqft_basement	-76.6356	4.570	-16.769	0.000	-85.593	-67.678
yr_built	-2.883e+06	1.35e+05	-21.322	0.000	-3.15e+06	-2.62e+06
yr_renovated	35.2610	4.167	8.462	0.000	27.094	43.428
zipcode	-749.1309	34.683	-21.599	0.000	-817.113	-681.149
lat	6.913e+05	1.12e+04	61.864	0.000	6.69e+05	7.13e+05
long	-2.452e+05	1.39e+04	-17.593	0.000	-2.73e+05	-2.18e+05
sqft_lot15	-9774.6761	4555.973	-2.145	0.032	-1.87e+04	-844.633

Omnibus:	15149.875	Durbin-Watson:	1.995
Prob(Omnibus):	0.000	Jarque-Bera (JB):	809022.411
Skew:	2.790	Prob(JB):	0.00
Kurtosis:	32.460	Cond. No.	2.40e+08

Findings Cont...

- Price of Condition 3/5 homes
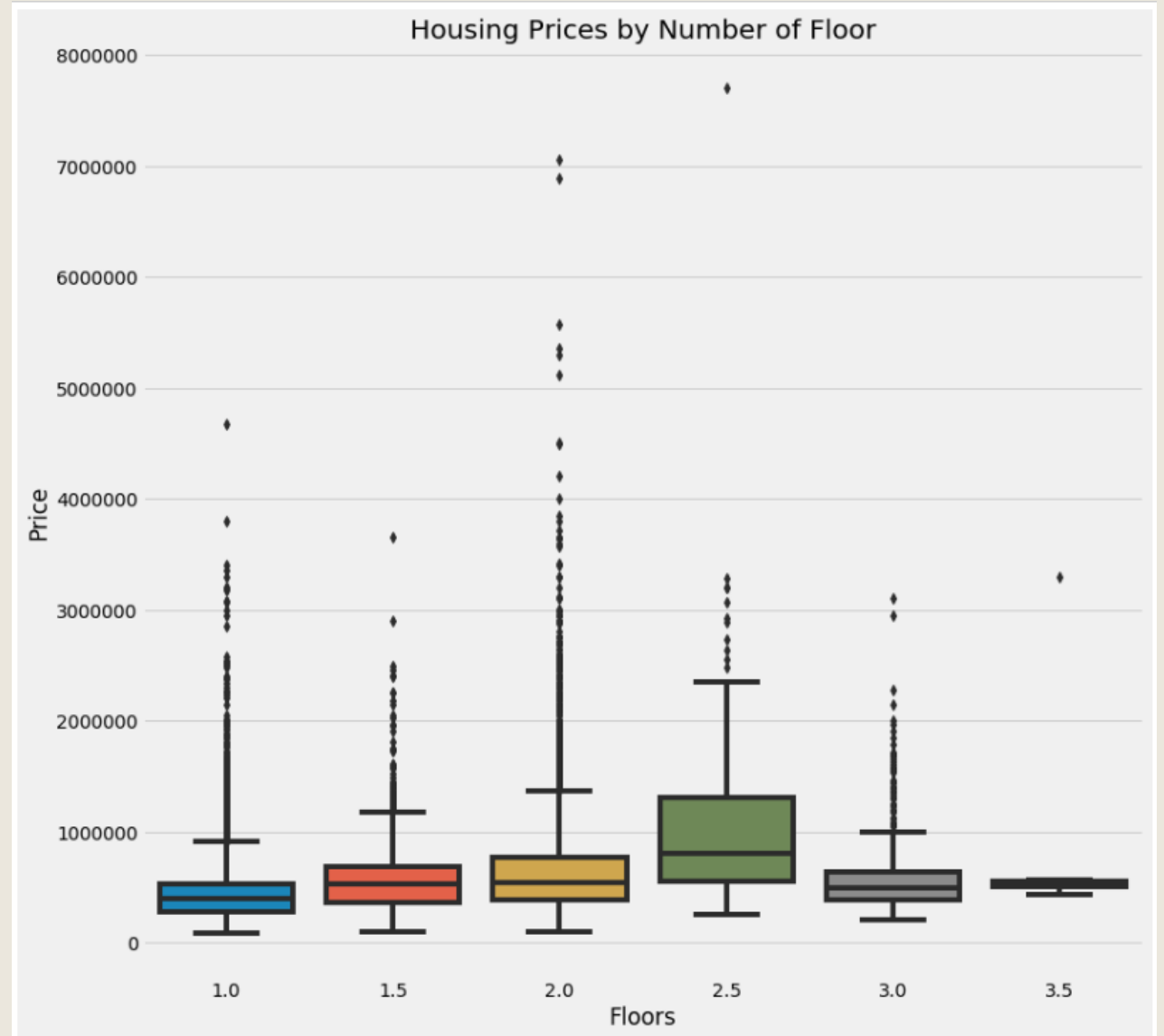 - *median*

- Condition 3 homes variance
  - *outliers*

Price vs Condition/Waterfront

Findings Cont...

■ Cumulative lot size does not predict well

Housing Prices by Sqft of Lots
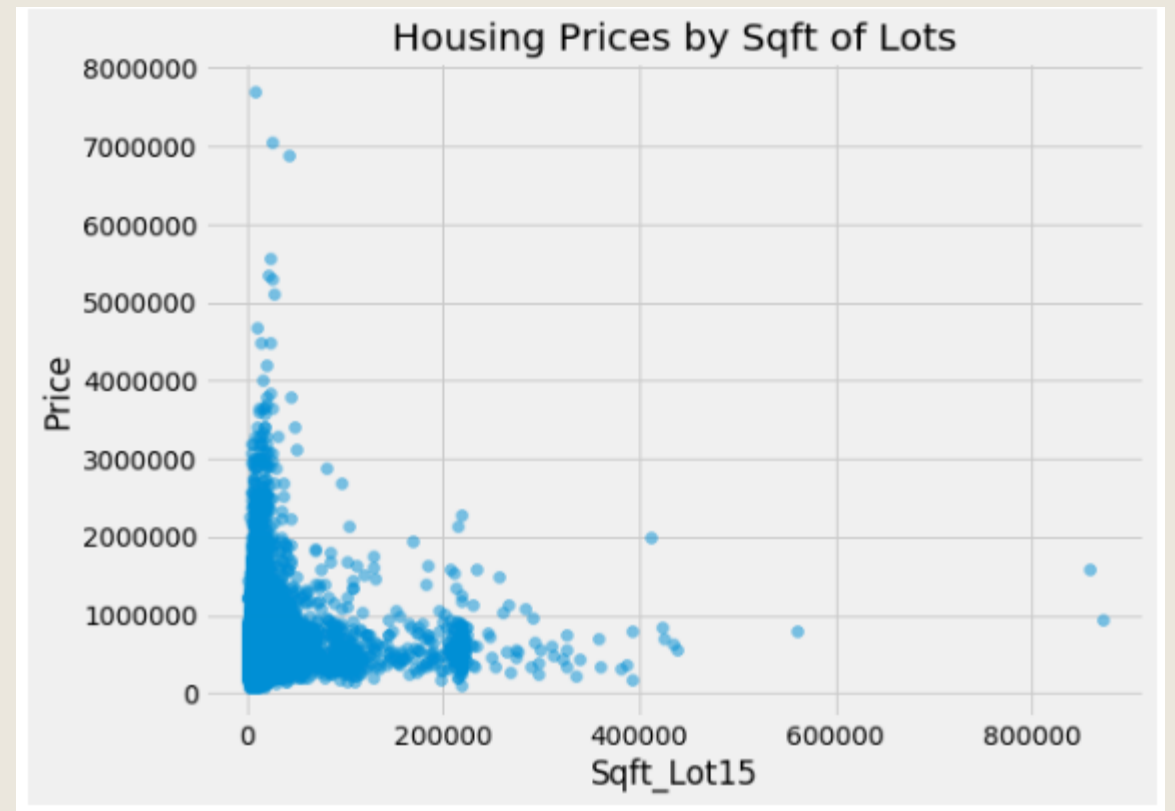
Findings Cont...

■ Number of floors predicts price...to a degree

Housing Prices by Number of Floor

How We Improve

- Data acquisition
  - *Data custodian*
  - *Cross referencing*
    - view
    - Waterfront
- Model has potential iterability

What Now?

- ■ Develop a better model
  - – *Additional multivariate techniques*
    - ■ MANOVA

Thank You