

King County House Sales

Module 1 Project

OLIVIA LO

PT 041519



Background

- King County House Sales dataset
- 21,000 rows of house sales data
- Sep 9, 2014 to Oct 1, 2015
- Use multivariate linear regression to predict the sales price of houses



What factors are affecting prices?

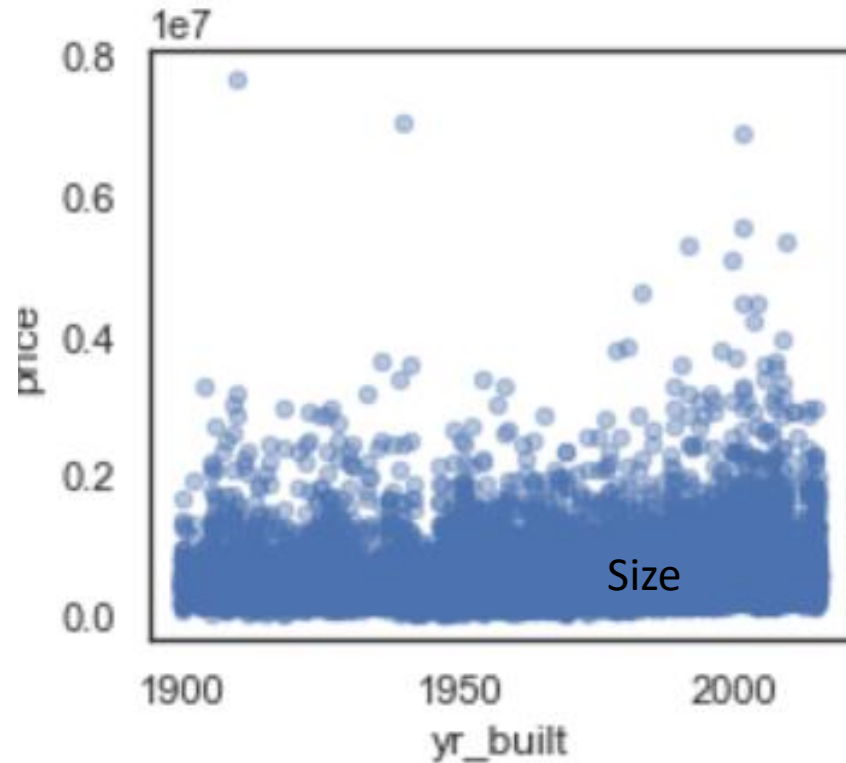
Bedrooms

Bathrooms

Size

Locations

What factors are NOT affecting prices?



Yr_built

Multivariable Regression Model

Dep. Variable:	price	R-squared:	0.774
Model:	OLS	Adj. R-squared:	0.773
Method:	Least Squares	F-statistic:	3579.
Date:	Tue, 25 Jun 2019	Prob (F-statistic):	0.00
Time:	21:40:16	Log-Likelihood:	-739.25
No. Observations:	20967	AIC:	1520.
Df Residuals:	20946	BIC:	1687.
Df Model:	20		
Covariance Type:	nonrobust		

The model can predict house prices with 77% of accuracy.

Interpretation



Increase in the interior living space in the nearest neighborhood = +26% in price



Increase in the living space of the house (apart from basement) = +38% in price



Increase in the housing grade = +16% in price

Further Analysis

- Adjust variables by include or exclude the ones that are highly relevant
- Better analyze the categorical data such as “Zipcode” and “Grade”
- Explore “Waterfront” and its effect on prices or how this may relate to “Zipcode”

Thank you!

OLIVIA LO

PT 041519

A solid orange horizontal bar spanning the width of the slide at the bottom.