

Predicting King County Property Prices

Flatiron School, Data Science, Flex Program

by

Milad Shirani

Contents

- ❖ Project Overview
- ❖ Data
 - ❖ Categorical Data
 - ❖ Numerical Data
 - ❖ Data Preparation
 - ❖ Converted Data
- ❖ Modeling and Results
- ❖ Suggestions
- ❖ Q & A

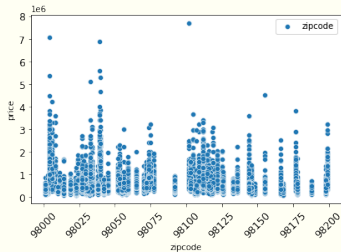
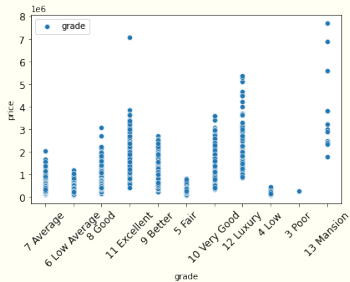
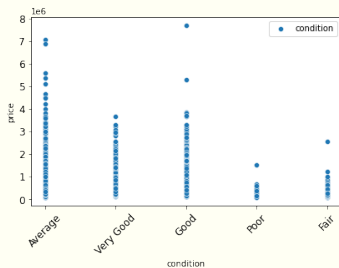
Project Overview

1. The goal of this project is predicting the property price in King County, WA, and assisting residents with methods that might help them to increase the value of their property
2. We use linear regression to find the model to predict the value of a property

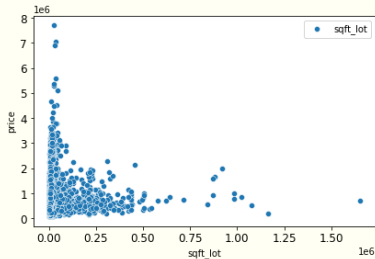
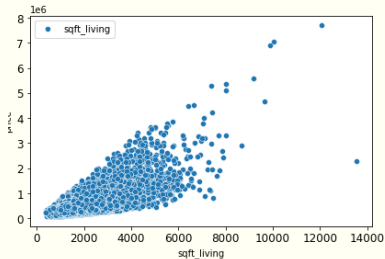
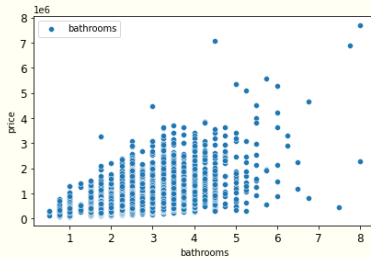
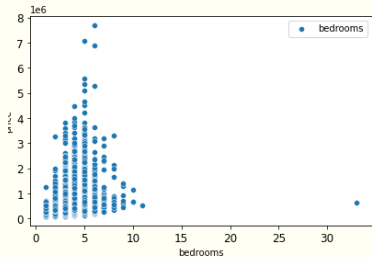
Data

1. We use 21099 data points for our analysis
2. Data is coming from [King County Open Data Source](#) and it contains information about a property such as
 - ❖ Number of bedrooms, bathrooms and floors in a house
 - ❖ Year when the property was built
 - ❖ Zipcode and condition of the property
 - ❖ Square footage of living space in the home
3. Categorical variables are:
 - ❖ waterfront, condition, grade, zipcode
4. Numerical variables are:
 - ❖ price, bedrooms, bathrooms, sqft_living, floors, yr_built, lat, long

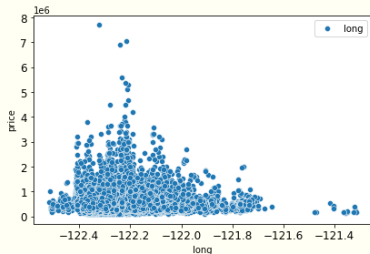
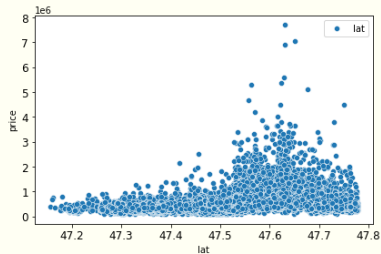
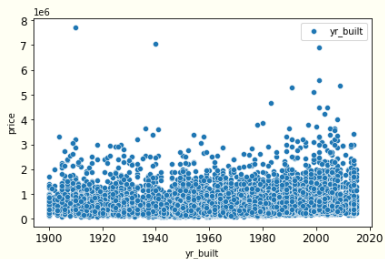
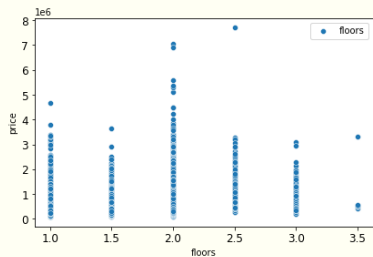
Categorical Data



Numerical Data



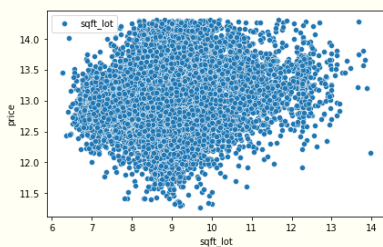
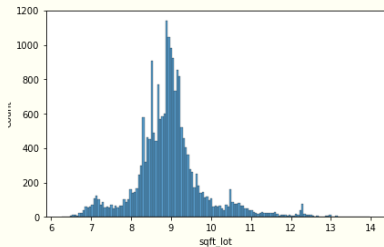
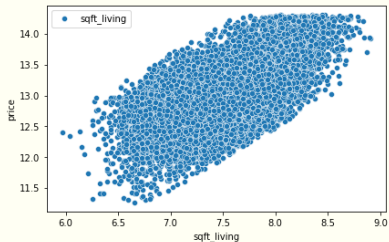
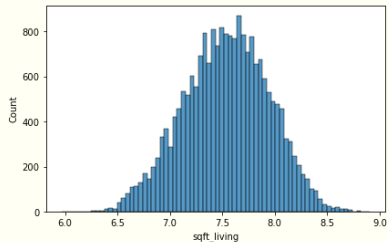
Numerical Data



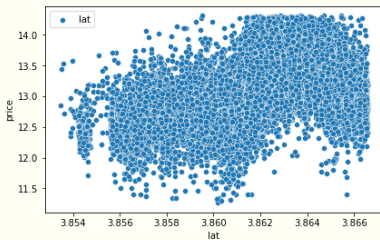
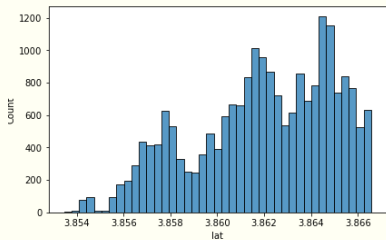
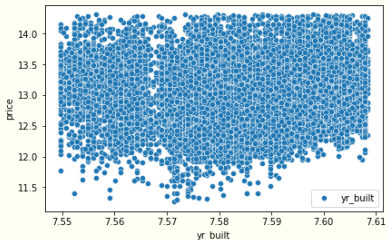
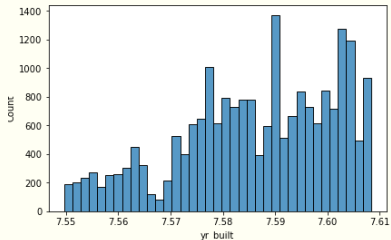
Data Preparation

1. There is one data point for "3 Poor" value in grade
2. Removing outliers from price
3. maximum number of bedroom is 8
4. maximum number of bathroom is 7
5. minimum number of bathroom is 1
6. maximum number of floors is 3.5
7. long is multiplied by -1 to become positive
8. We use the natural logarithm to convert
 - ❖ lat, long, yr_built, sqft_living, sqft_lot

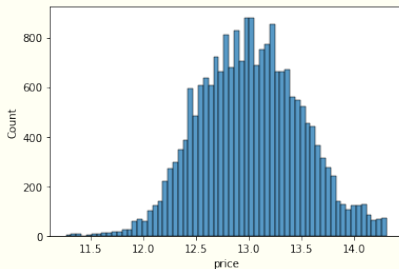
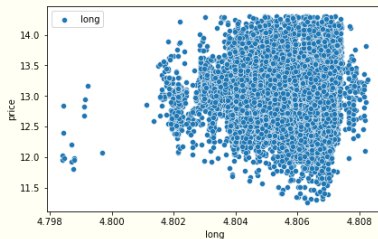
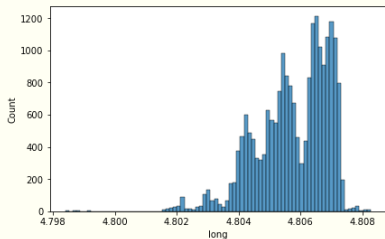
Converted Data



Converted Data



Converted Data



Modeling and Results

1. We used Linear Regression for modeling
2. $R^2 = 0.72$
3. Mean Squared Error \$143922.7
4. Interpreting the coefficient of square footage living area (c_s) is:

$$\frac{p(s + \Delta s)}{p(s)} = \exp\left(\frac{s + \Delta s}{s}\right)^{c_s}$$

	feature	coefficient
0	bedrooms	-0.0208
1	floors	0.054
2	5 Fair	-0.1196
3	7 Average	0.1789
4	4 Low	-0.2152
5	8 Good	0.3805
6	sqft_living	0.4858
7	waterfront_impute	0.5539
8	9 Better	0.5932
9	10 Very Good	0.7328
10	11 Excellent	0.861
11	12 Luxury	1.0057
12	yr_built	-7.6526
13	lat	62.5153
14	const	-174.335

Conclusion

1. Project goal is to predict the property housing in King County, WA
2. We used linear regression to model the data
3. Latitude has the highest coefficient compared to other numerical features so it has the highest impact.
4. Square Footage of Living Area has the second highest impact on the price of a property.
5. The property owners might consider the following that might increase the value of their property
 - ❖ Reduce the number of bedrooms
 - ❖ Increase the square footage of living area
 - ❖ increase the grade of the property at least to very good.

Next Steps

1. Adding other features such as renovation year or square footage of basement to the model
2. Adding some combination of features into the model
3. Considering adding polynomial features to the model

*Thank
You!*