Predicting King County Property Prices

by

Milad Shirani

Contents

Project Overview

1. King County in Washington State has decided to help newly married couples to find a property in this county and the county wants to estimate the value of a property

2. King County in Washington State has asked us to suggest them two ways that they can apply to increase the value of a property.

3. We use linear regression to find the model to predict the value of a property
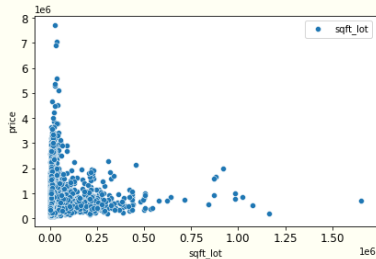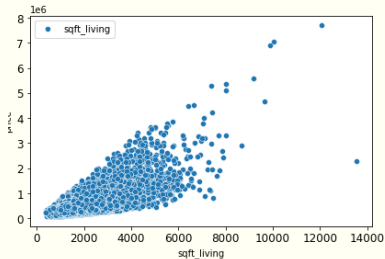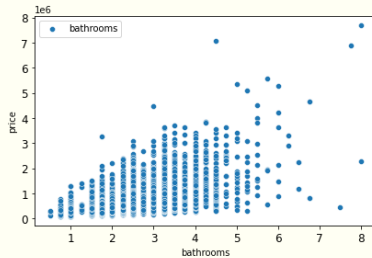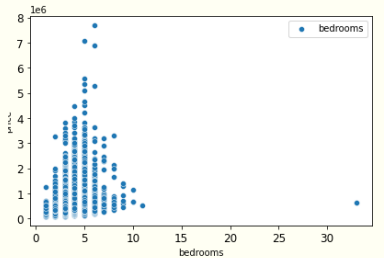
Data

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, zipecode
4. Numerical variables are:
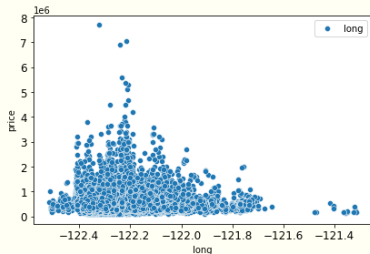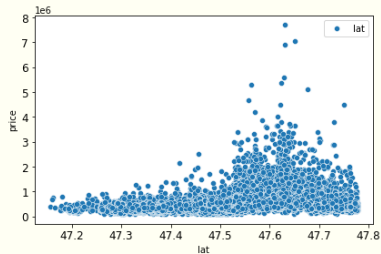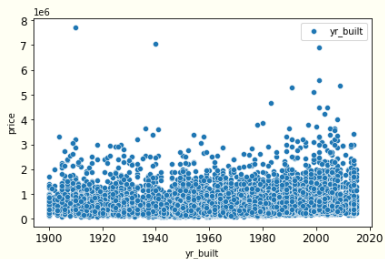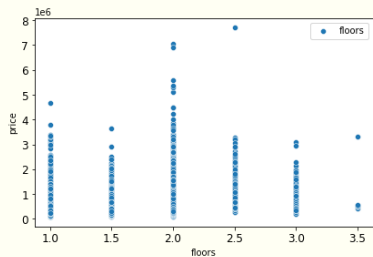   - price, bedrooms, bathrooms, sqft_living, floors, yr_built, lat, long

Categorical Data
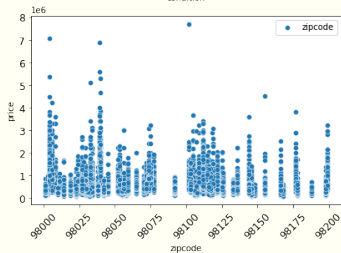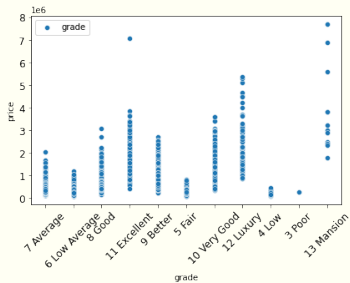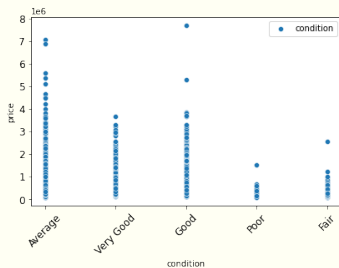
Numerical Data

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
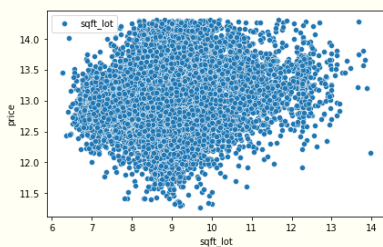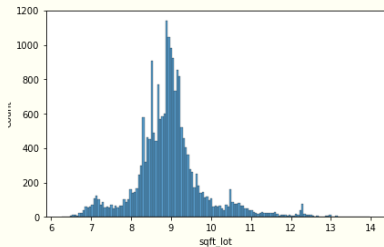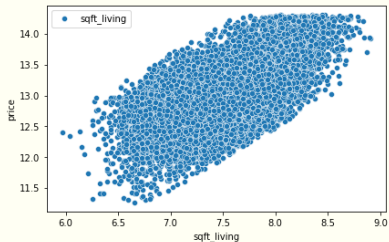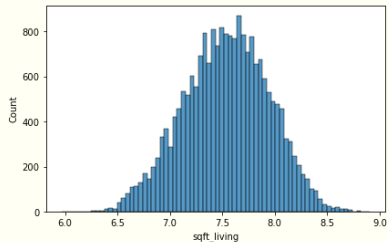
Converted Data
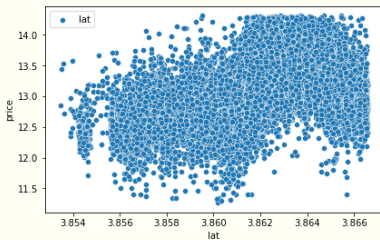
Converted Data
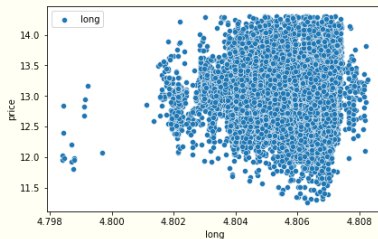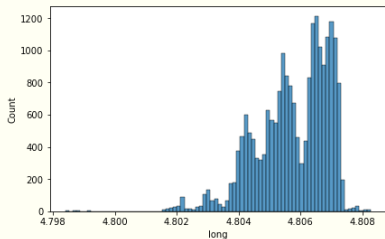
Modeling and Results

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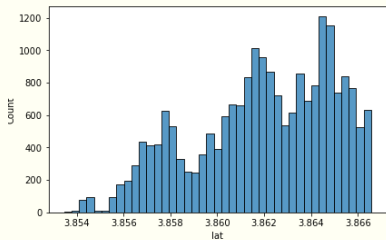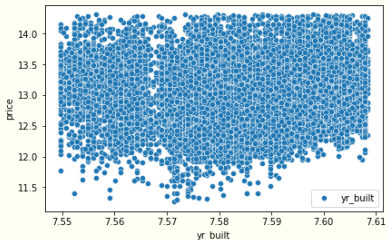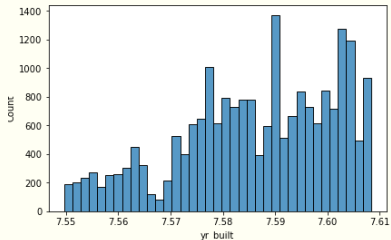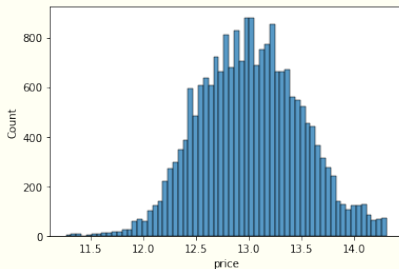4. Interpreting the coefficient of square footage living area (c_s) is:

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

	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. King County asked us to estimate the value of properties and suggest them two ways that can increase properties' values.
2. We used linear regression to model the data
3. We concluded that
   - Latitude and Square Footage of Living Area are the first and second numerical features that have the highet impact on the price of a property
4. 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 lease 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!