Predicting King County Property Prices

Flatiron School, Data Science, Flex Program

by

Milad Shirani

February 2022

Contents

Data

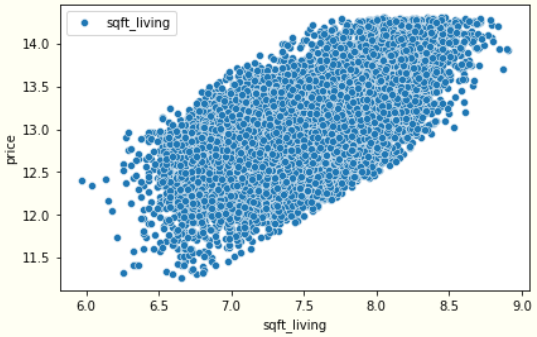
1. We use 21099 data points for our analysis
2. Data is coming from King County Open Data Source and it contains information about a property such as
   - Number of bedrooms,bathrooms and floors in a house
   - Year when the property was built
3. Categorical variables are:
   - waterfront, condition, grade, zipecode
4. Numerical variables are:
   - price, bedrooms, bathrooms, sqft_living, floors, yr_built, lat, long
5. Data preparation contains removing outliers and setting maximum and minimum values for
   - bedrooms, bathrooms, sqft_living, floors, grade

Modeling and Results

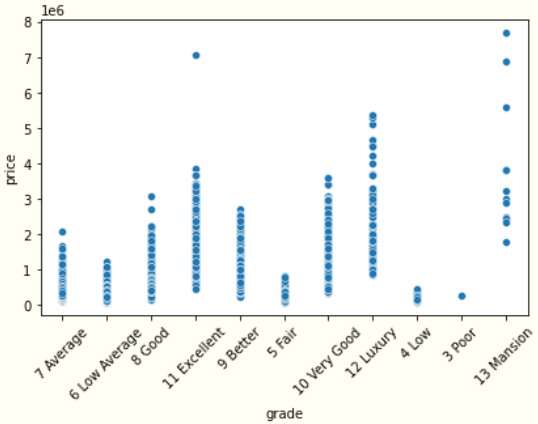
1. We used Linear Regression for modeling
2. $R^2 = 0.72$
3. Root Mean Squared Error $143922.7

Interpretation of Coefficients

1. 1% change in the scaled ‘sqft_living‘ gives 0.4858% change in scaled ‘price‘



2. Changing ‘grade‘ from ‘7 Average‘ to ‘12 Luxury‘ gives $0.8268 changes in scaled ‘price‘.



Conclusion

1. We used linear regression to model the data
2. We concluded that
   - Latitude and Square Footage of Living Area are the first and second numerical features that have the highest impact on the price of a property
3. 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

Q and A