



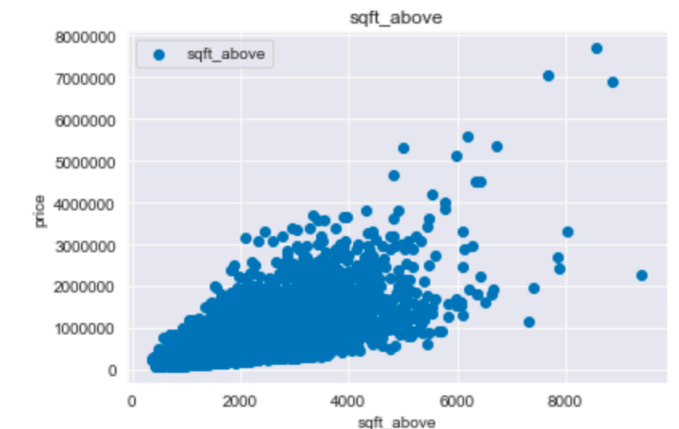
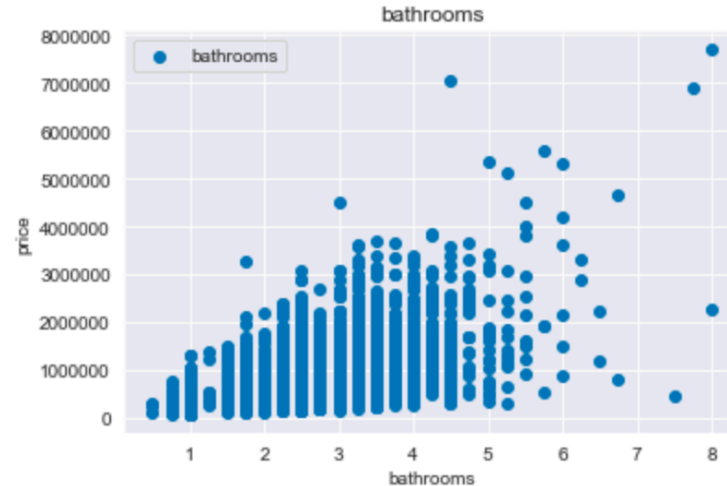
Predicting house prices

kc_house_data.csv

Methodology: Preparing and exploring data

- Identifying missing values in the dataset
- Replacing missing values for 'view' and 'yr_renovated' with the median for the respective column
- Replacing missing values for 'waterfront' with 0

Columns 'waterfront', 'view', and 'yr_renovated' had missing values



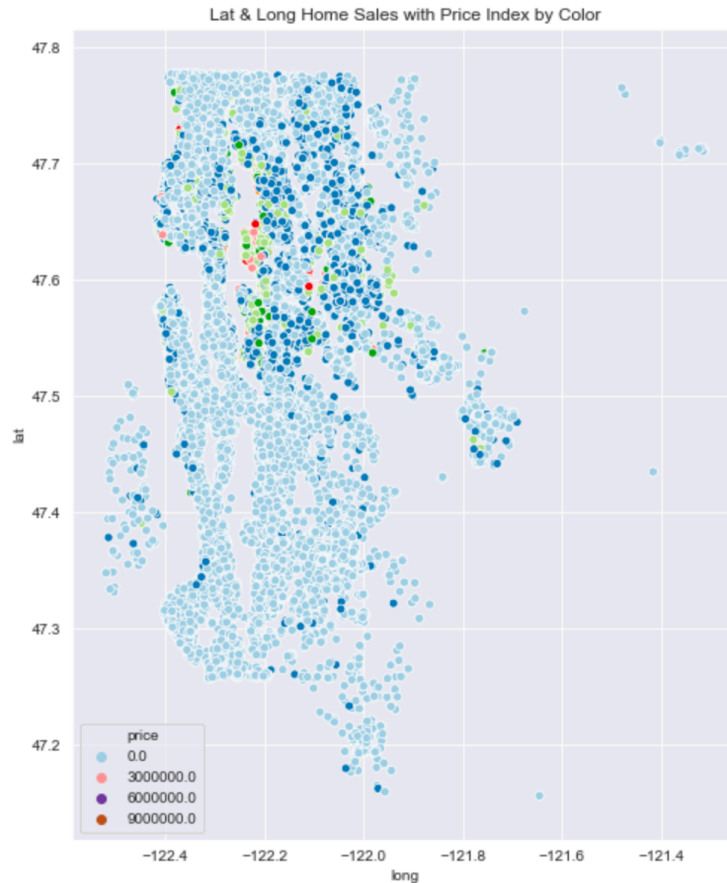
Initial model

- The initial model was based on cleaned data using 'sqft_living', 'view', and 'yr_renovated' columns as independent variables since they were shown to have the most direct relationship with price
- The R-squared of 0.705 showed that the data fitted the regression line relatively closely
- There was room for improvement in the model given that R-squared could be higher

Improved model

- Using a modified set of independent variables
- **'Sqft_living'** and **'bathrooms'** is used instead of 'sqft_living' and 'sqft_above'
- **'Has_renovation'** and **'been_viewed'** are used as independent variables and are a modification of the columns 'view' and 'yr_renovated' since the exact year of renovation and number of views are less relevant than the fact whether a property has been viewed or renovated
- R-squared has improved from 0.705 to **0.725** indicated that the data fits regression line more closely, however there is still room for improvement based on both the **R-squared value** and the somewhat elevated **test root mean squared error**

Location matters



As can be seen from the picture, properties with the highest prices are concentrated in the northern part of the county. Further research and analysis could be helpful to explain the underlying reasons.

Summary & Recommendations

- The size of the house indicated by 'sqft_living' and 'bathrooms' is almost directly related to its price, meaning as those two variables increase, the price of the house increases
- The exact number of views and the year of renovation do not matter as much as whether the property has been viewed or renovated
- Focus on houses that **have not** been viewed or renovated, regardless of size, to find properties with the most potential for price increase through renovation and marketing
- Be cognizant of the underlying reasons why the property has not been viewed (poor location, lack of certain desired characteristics/amenities)