

# GROUP ICHIGO

IVY



PATIENCE



JUDY



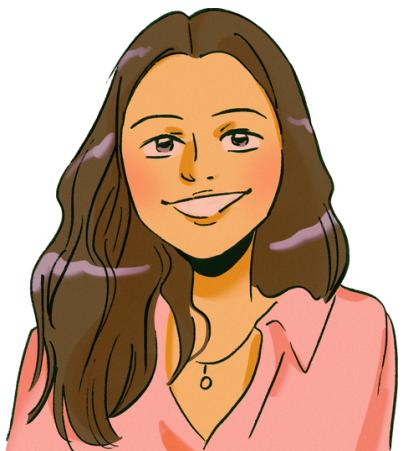
MARYANN



REBECCA



WEMA



HANNAH



BRADLEY



BELLA



# HOUSING PRICE PREDICTION



# *INTRODUCTION;*



- Buying a new house is, without a doubt, a daunting process, both for the agency and the individual.
- The journey is often filled with lots of fraud, and the research needed to strike a good deal is exhausting.

# PROBLEM STATEMENT

Our client is a real estate agency who plans on expanding their real estate business within the northwestern county. They want us to help them investigate the relationship between the price of their properties and several predictor variables including square footage, number of bathrooms, and number of bedrooms so as to identify the key factors that influence the sale price of a house.

# MAIN OBJECTIVE

The main goal of the project is to build an optimal linear regression model to predict the final prices of each house using the features in the test data.

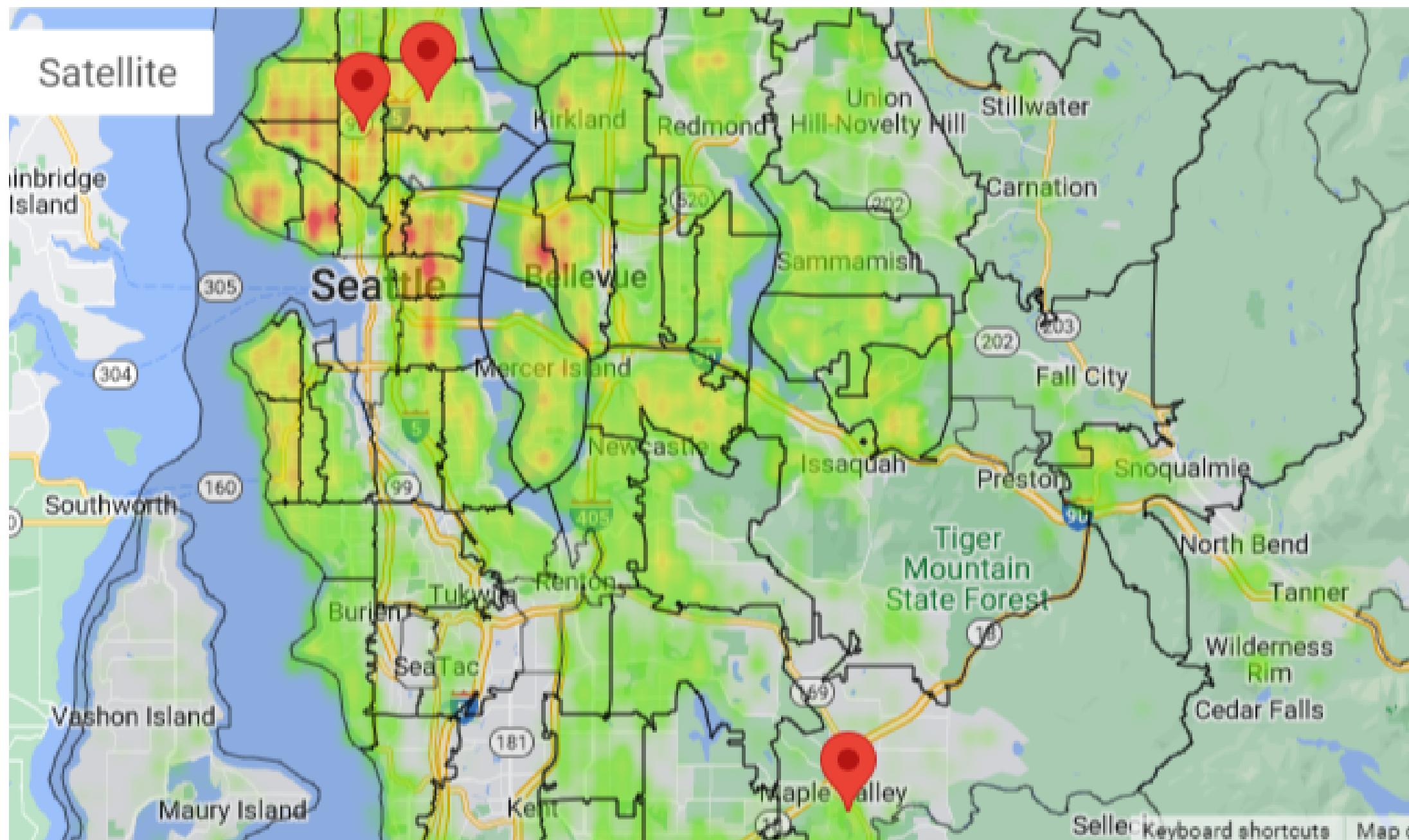
# Specific Objectives

- Exploratory Data Analysis of the dataset to understand each data feature and their relationship among each other and to the target, the housing sale price.
- To identify the variables affecting house prices.
- To create a linear model that quantitatively relates house prices with variables.
- Evaluating our model on how well these variables can predict house prices.
- Come up with recommendations upon interpreting our regression model.

# EXPLORATORY DATA ANALYSIS

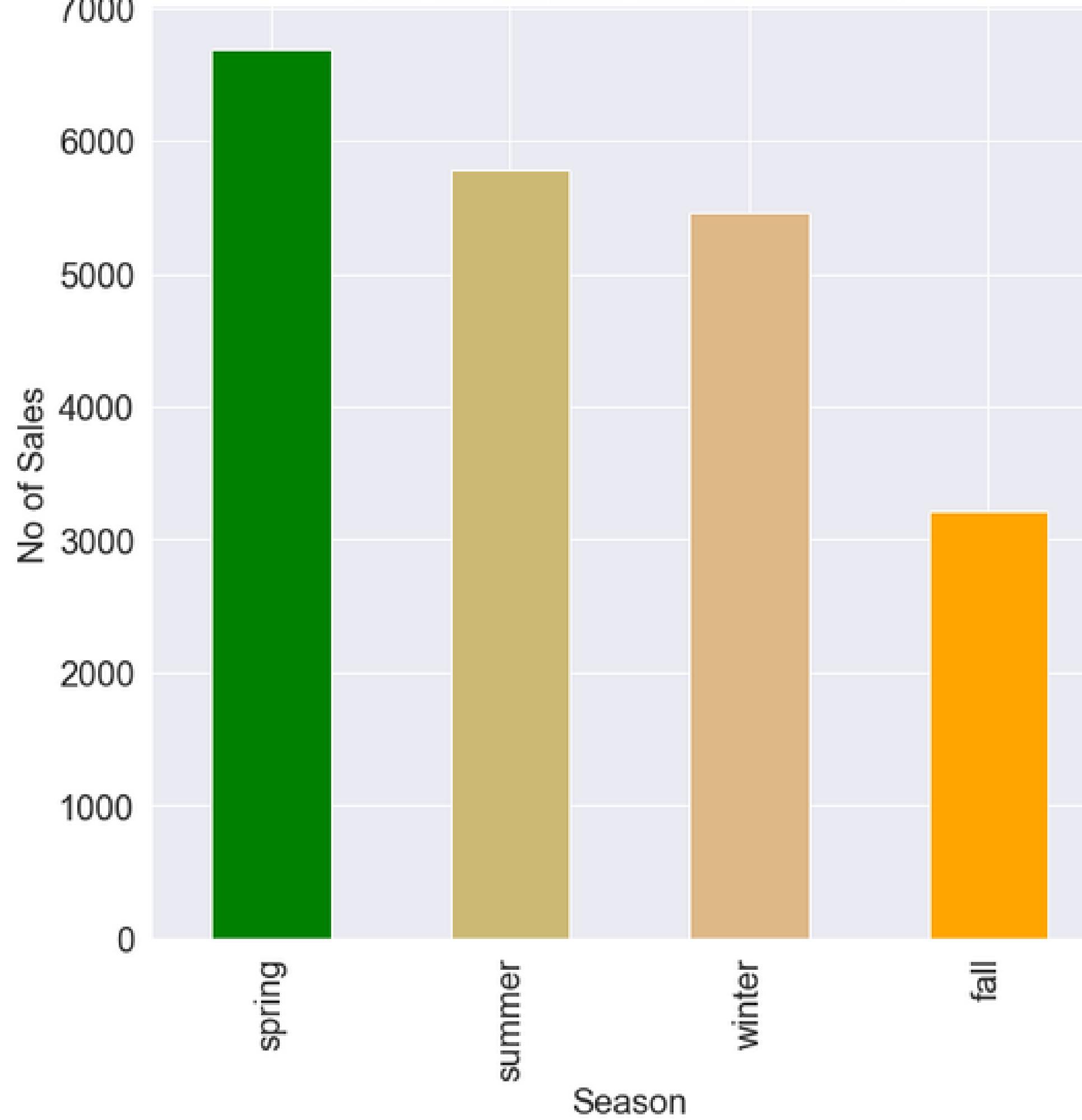


# ZIPCODE HEATMAP



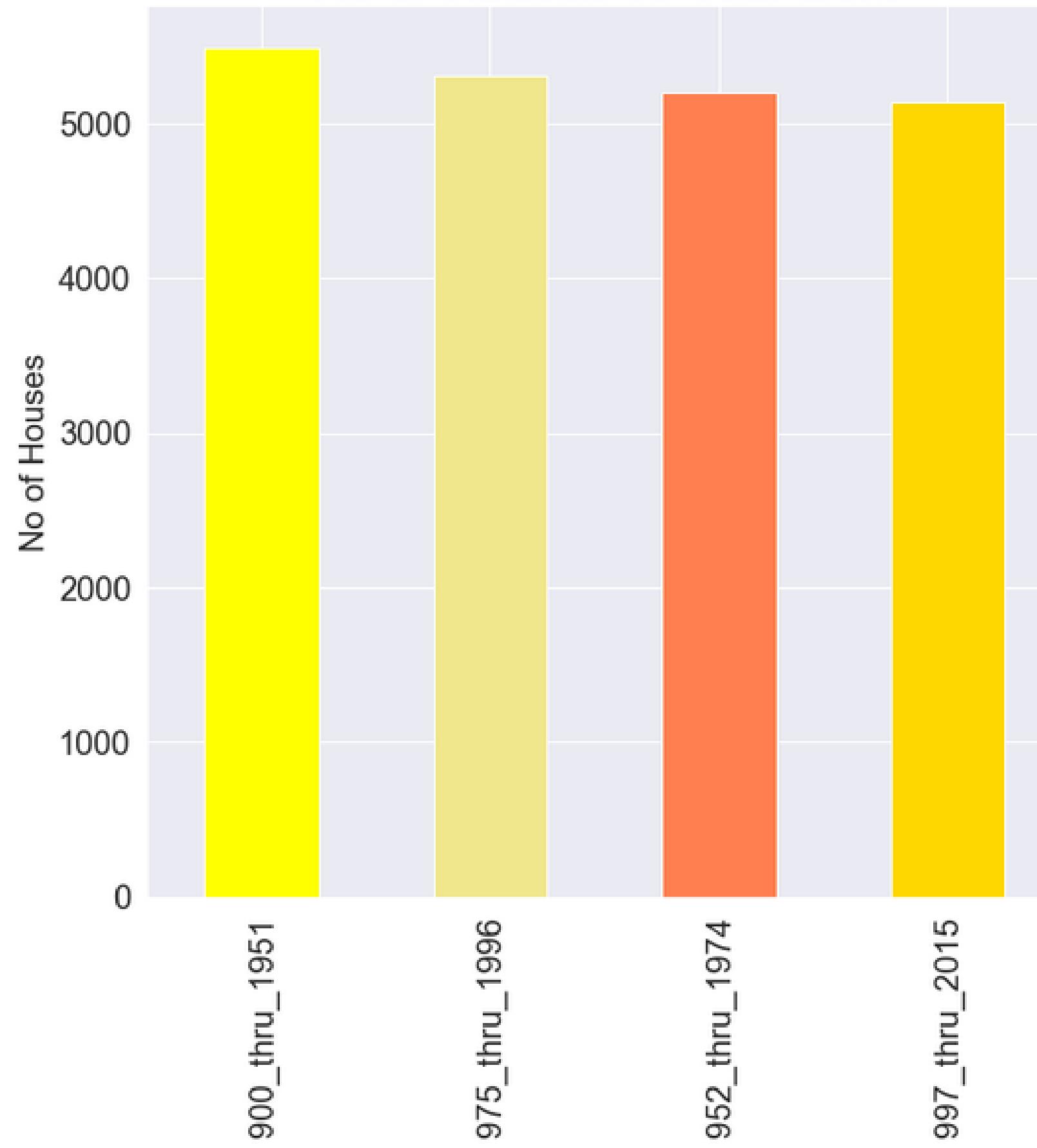
The zipcodes for the houses shows most of the data came from Seattle

## No of Sales vs Seasons



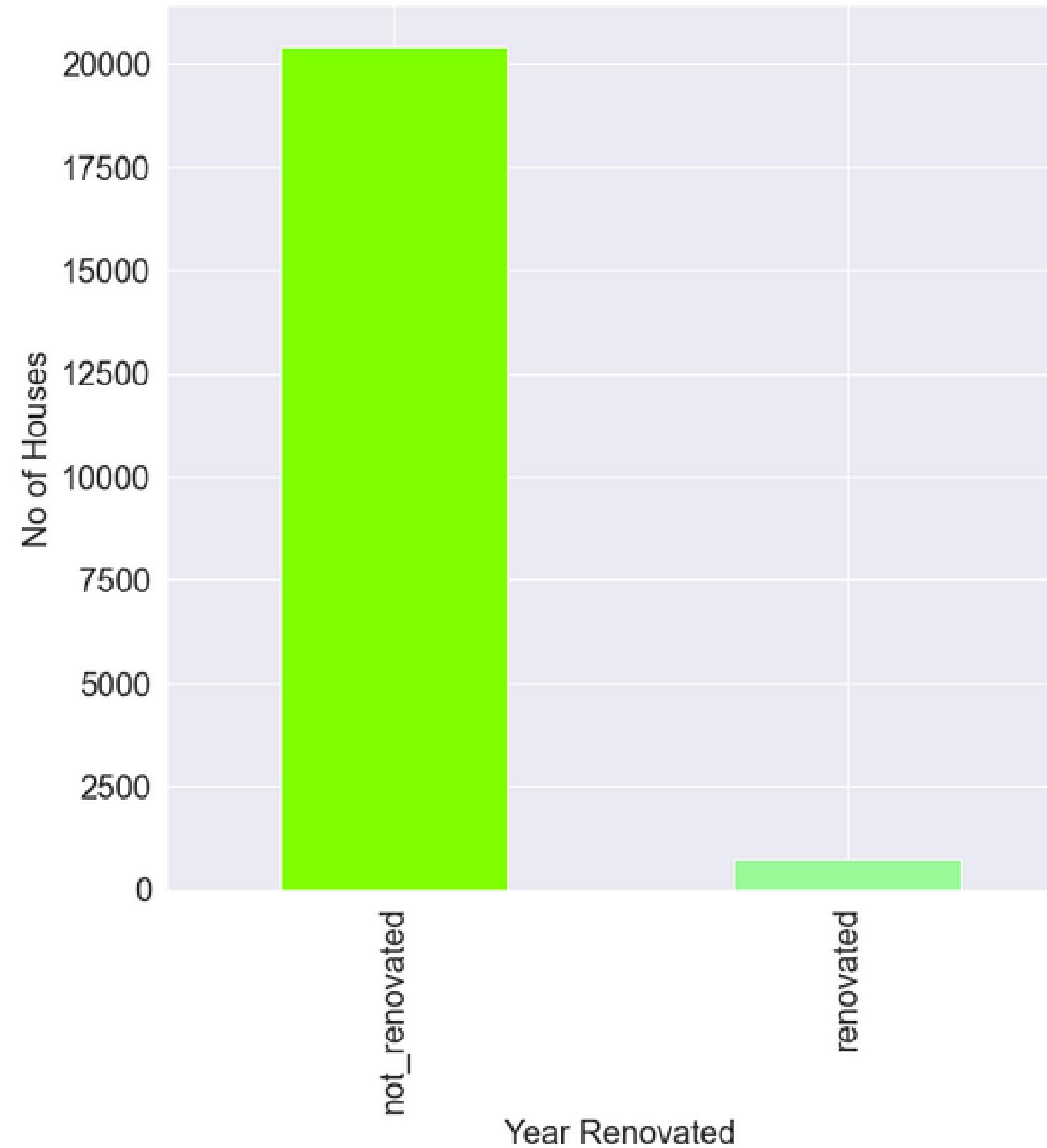
Most houses were sold in Spring with the least number of houses being sold in Fall.

## No of Houses vs Year Built

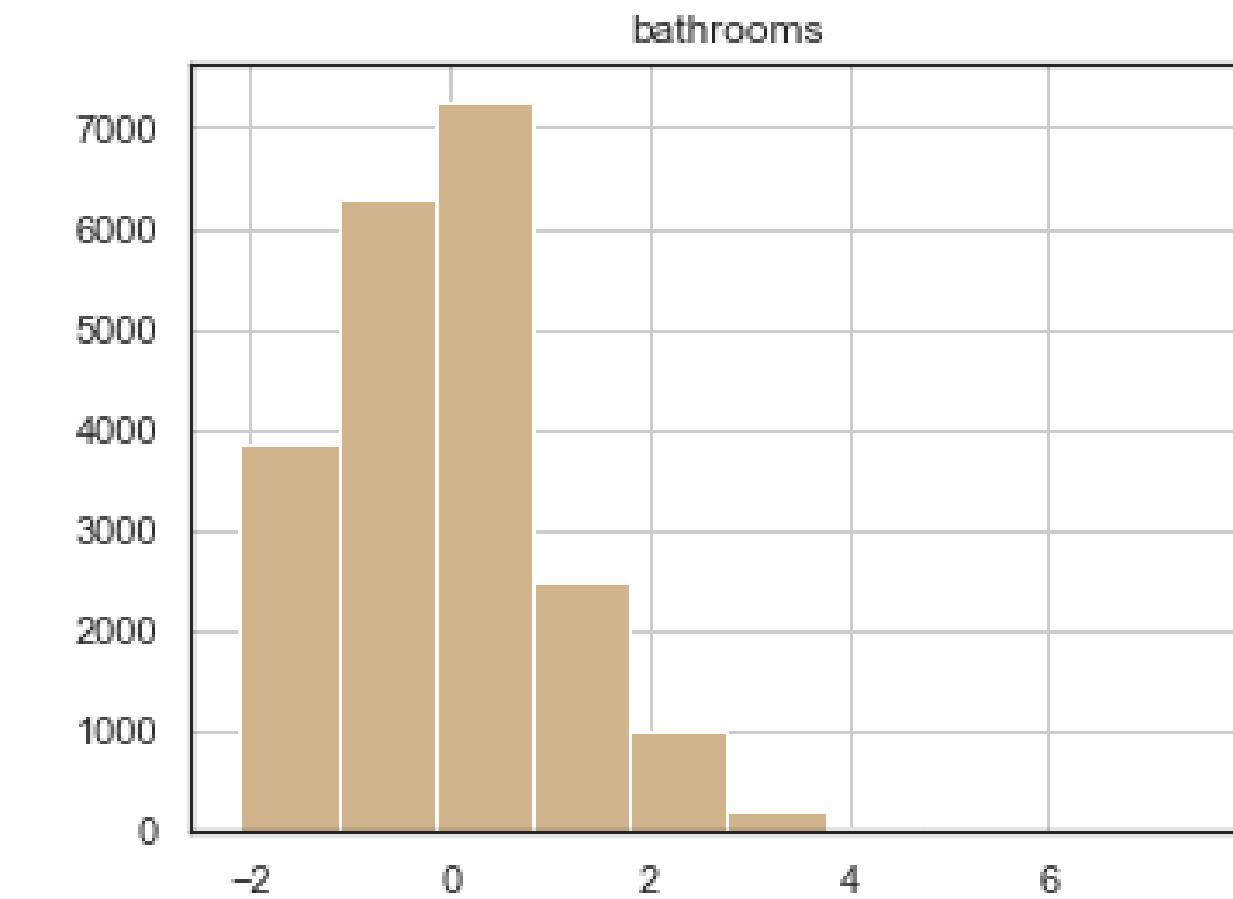
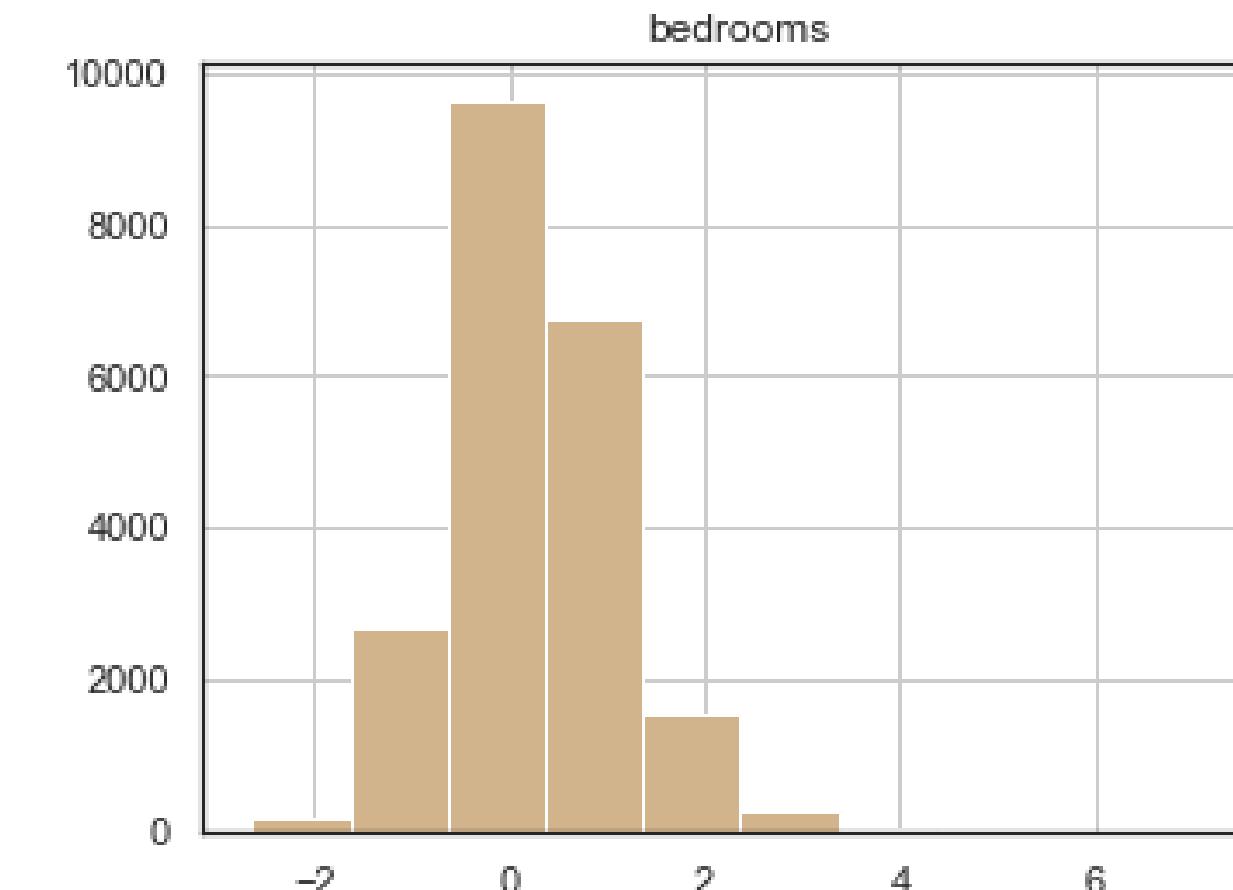
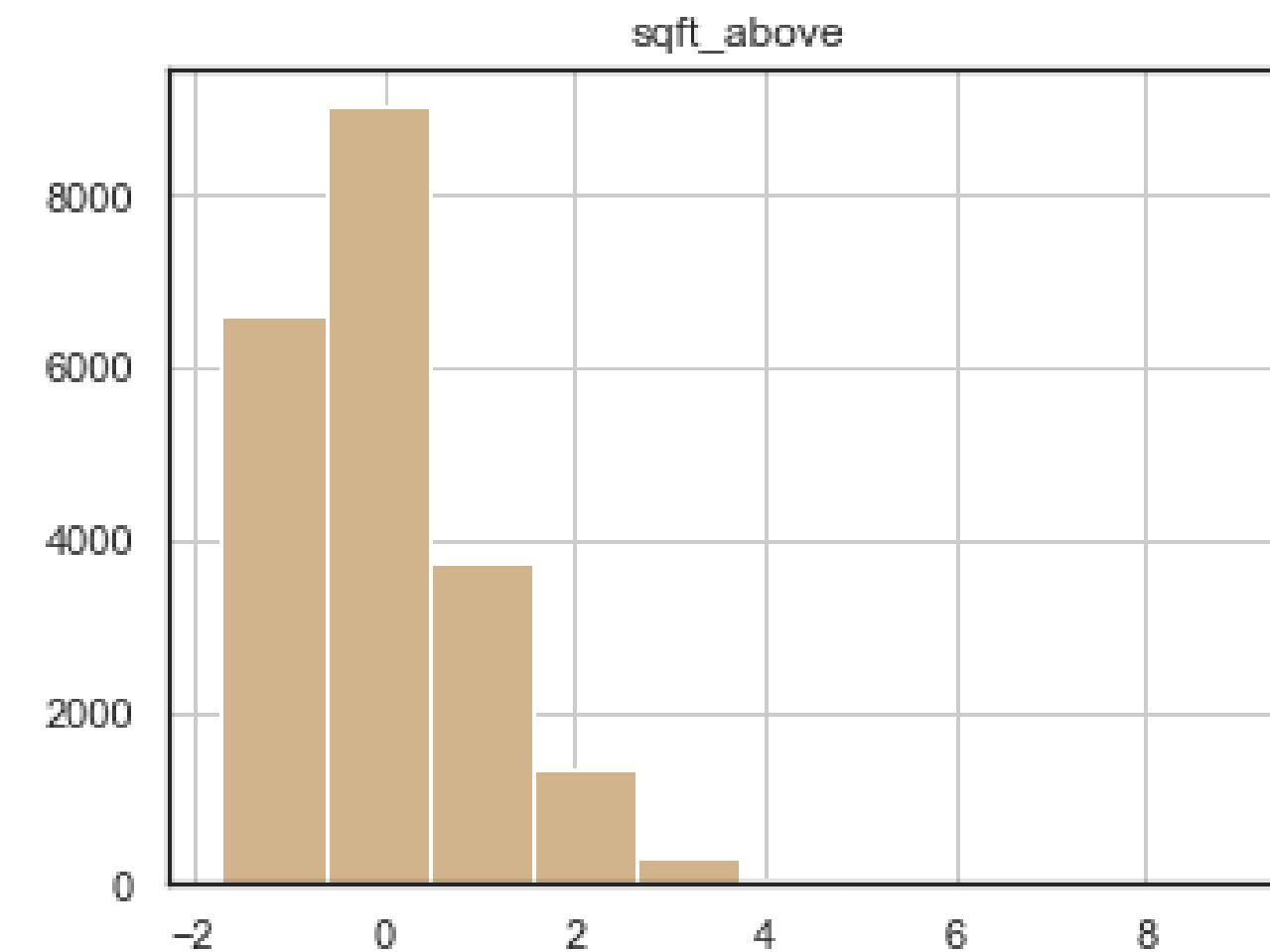
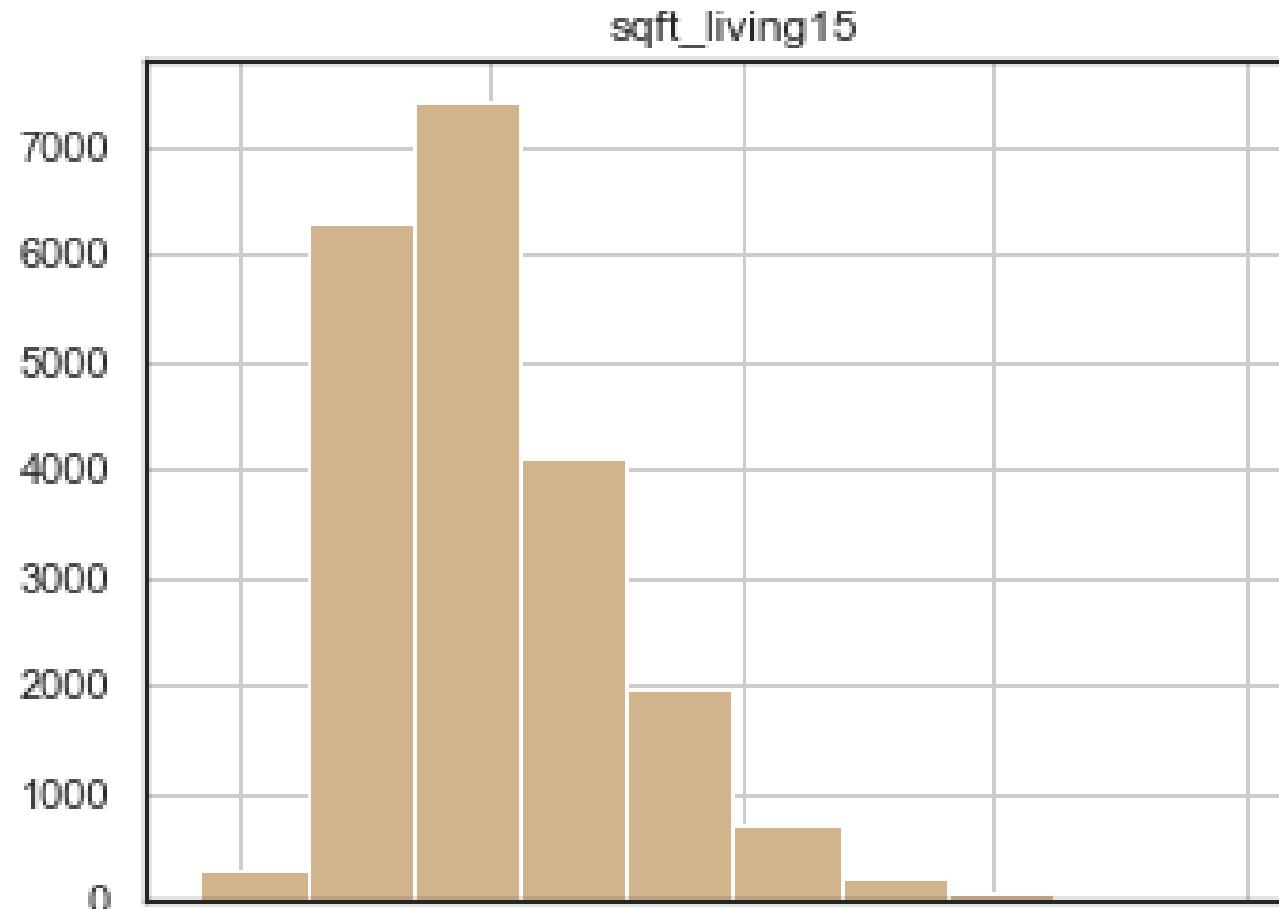


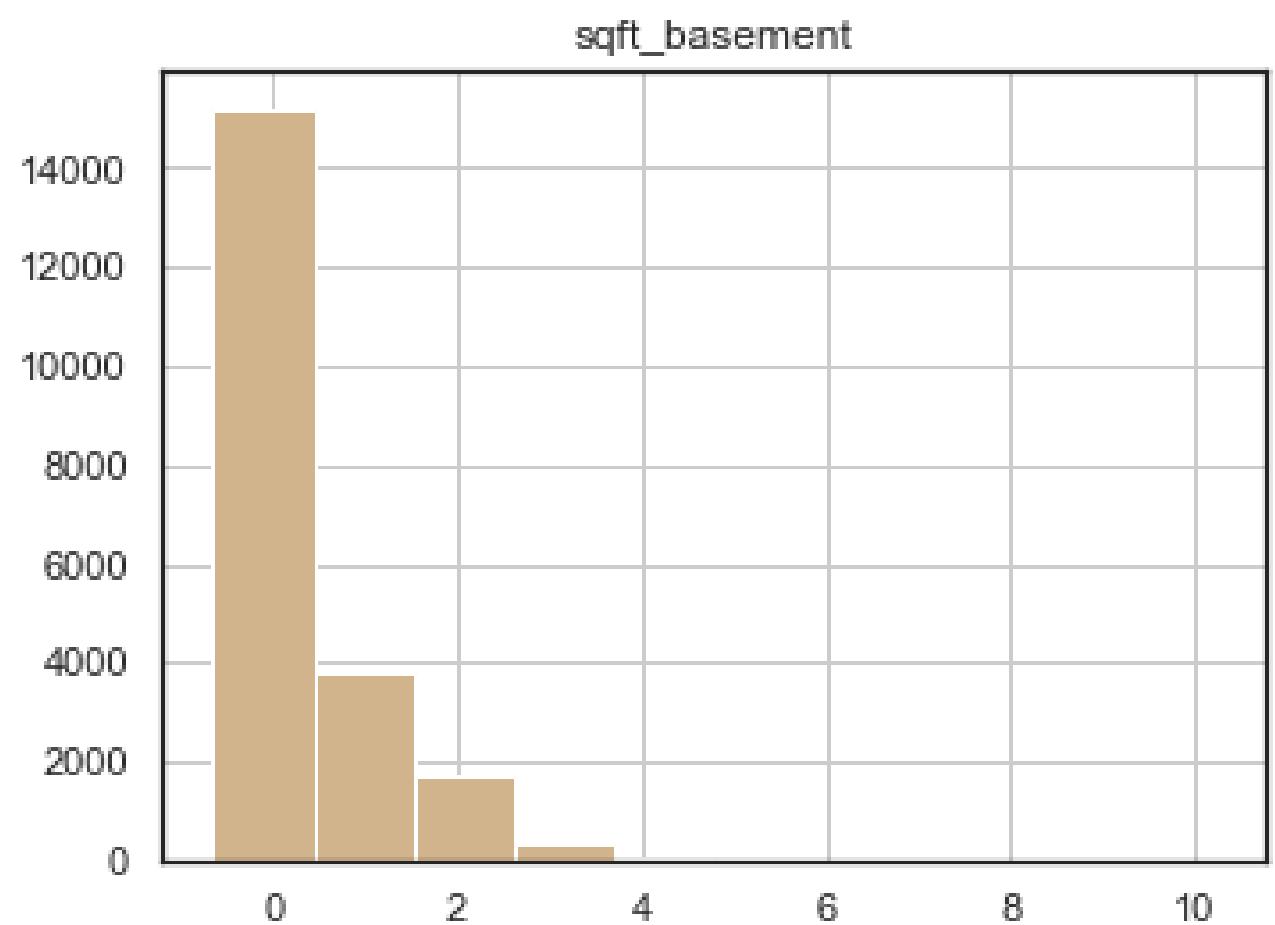
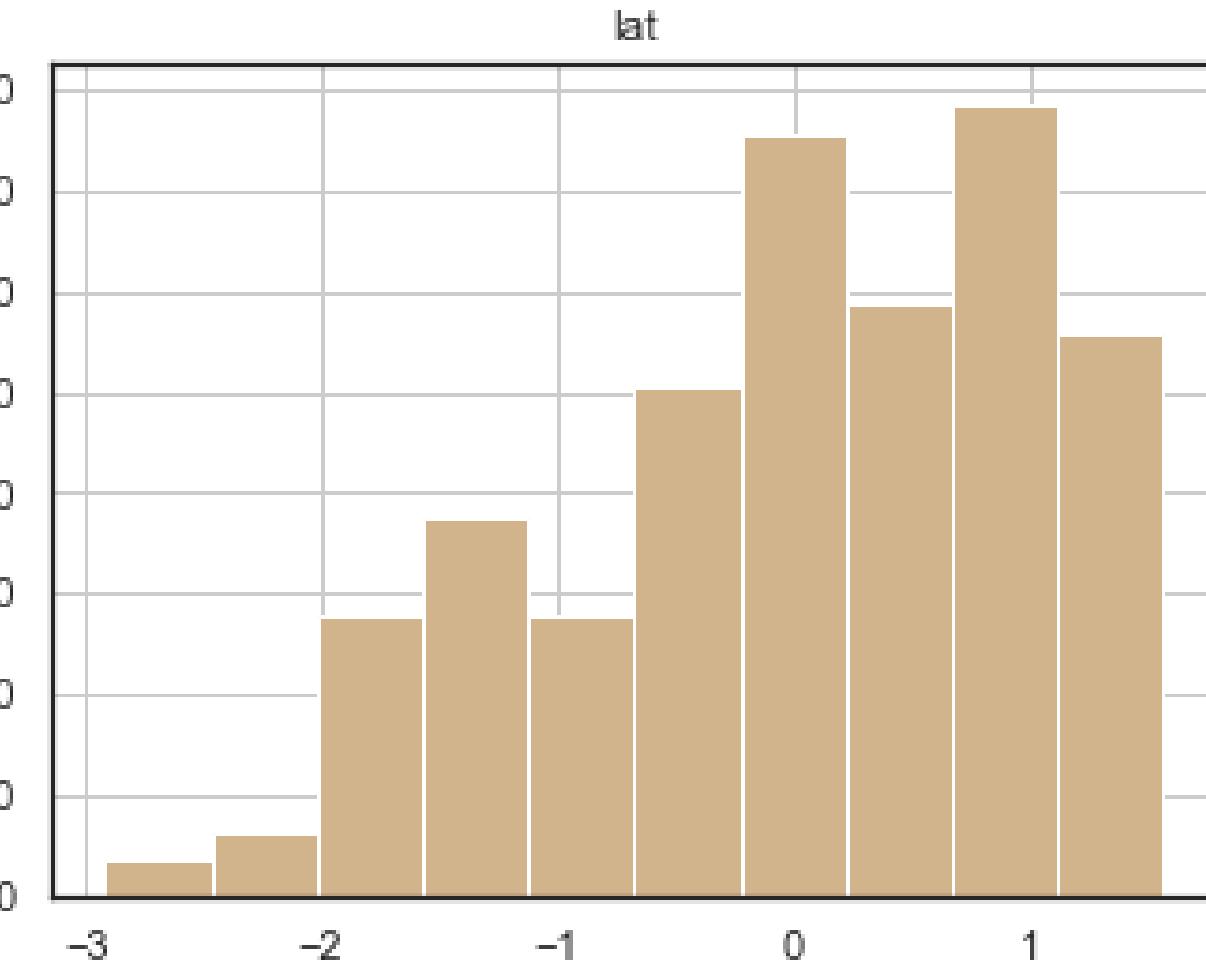
The year a house was built didn't have much of an influence on the purchase decision made by clients. However, those built from 1900 through 1951 were sold most.

### No of Houses vs Renovated

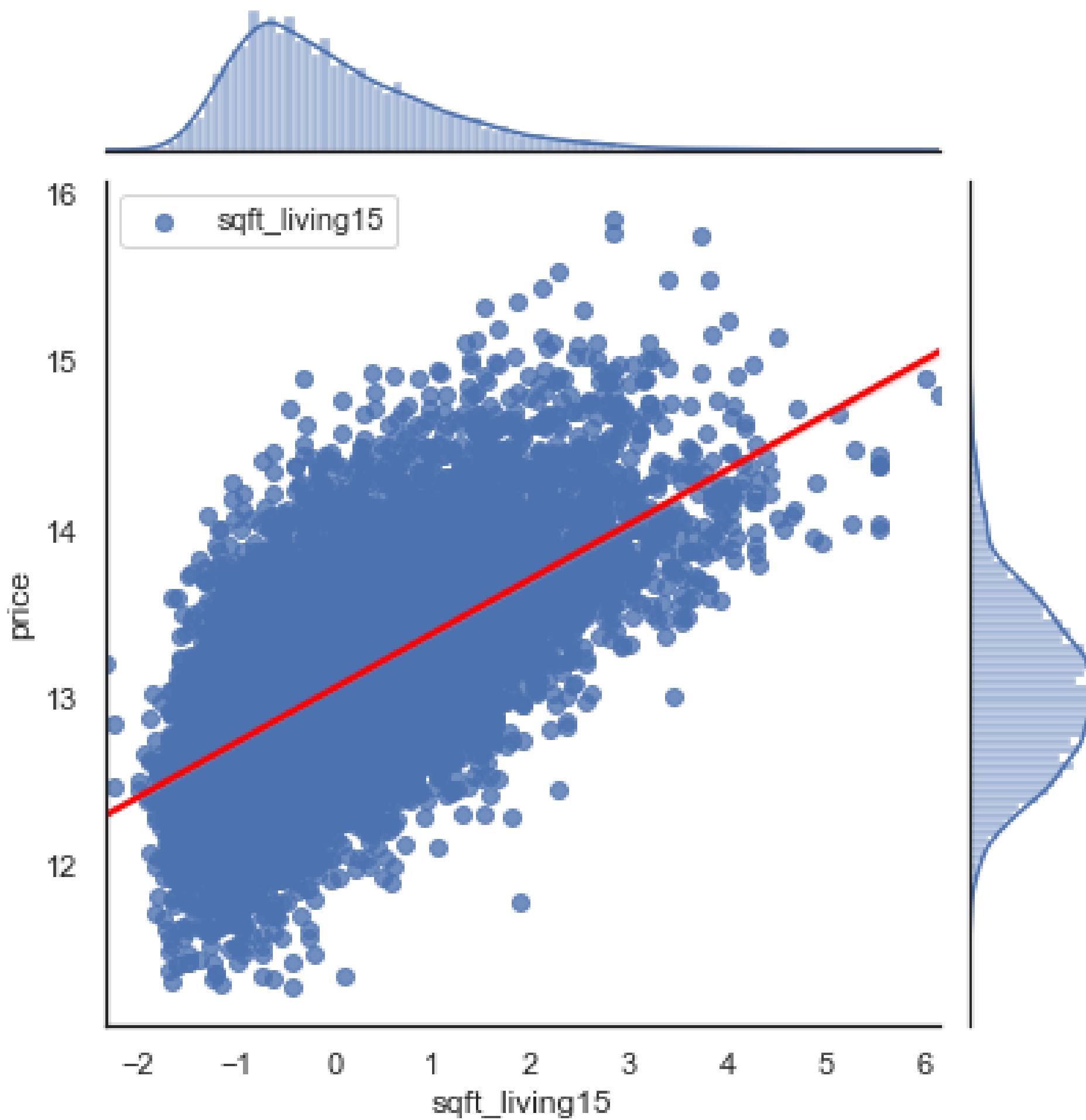


From the data in the bar-graph, most houses have not been renovated

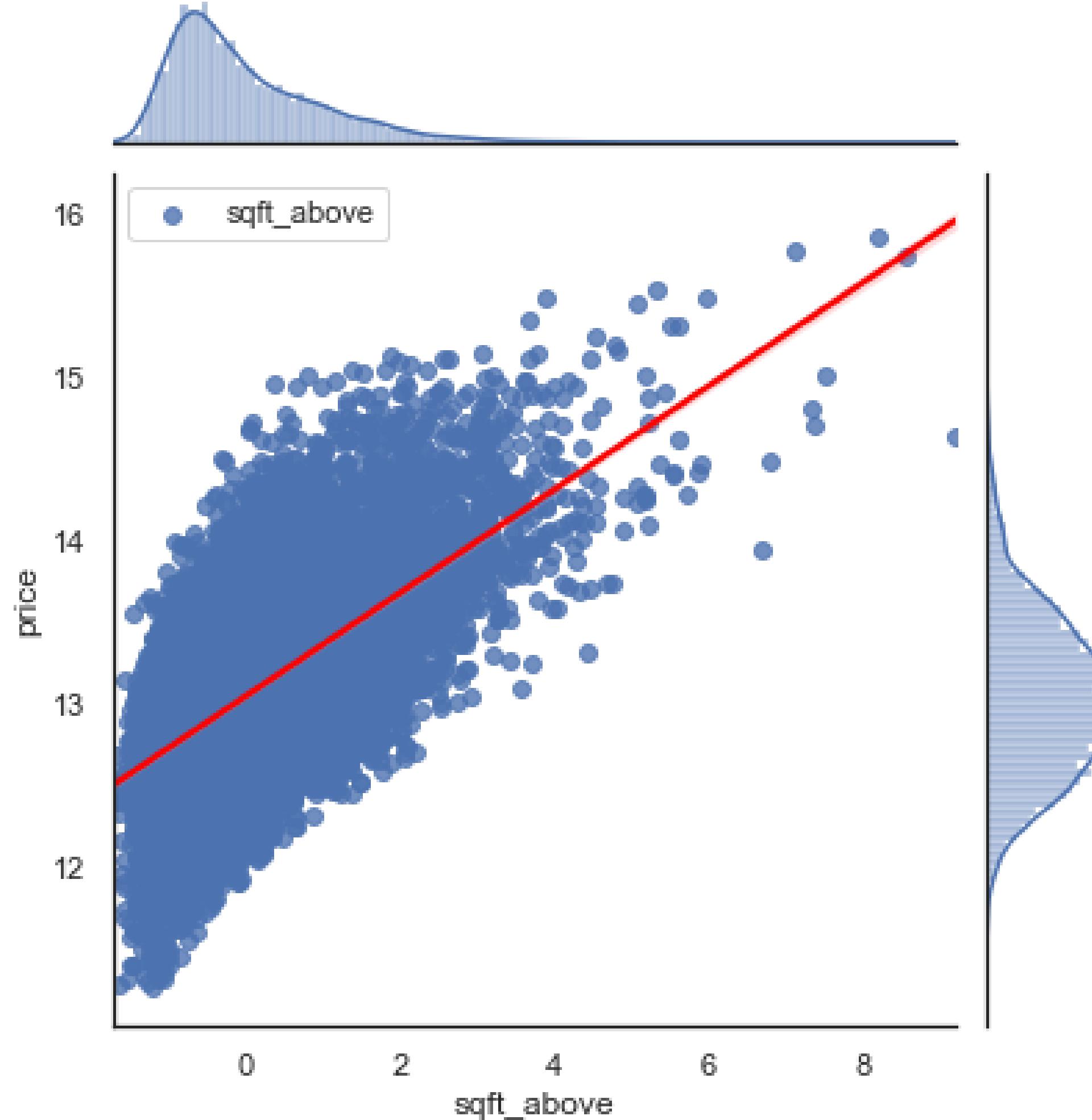




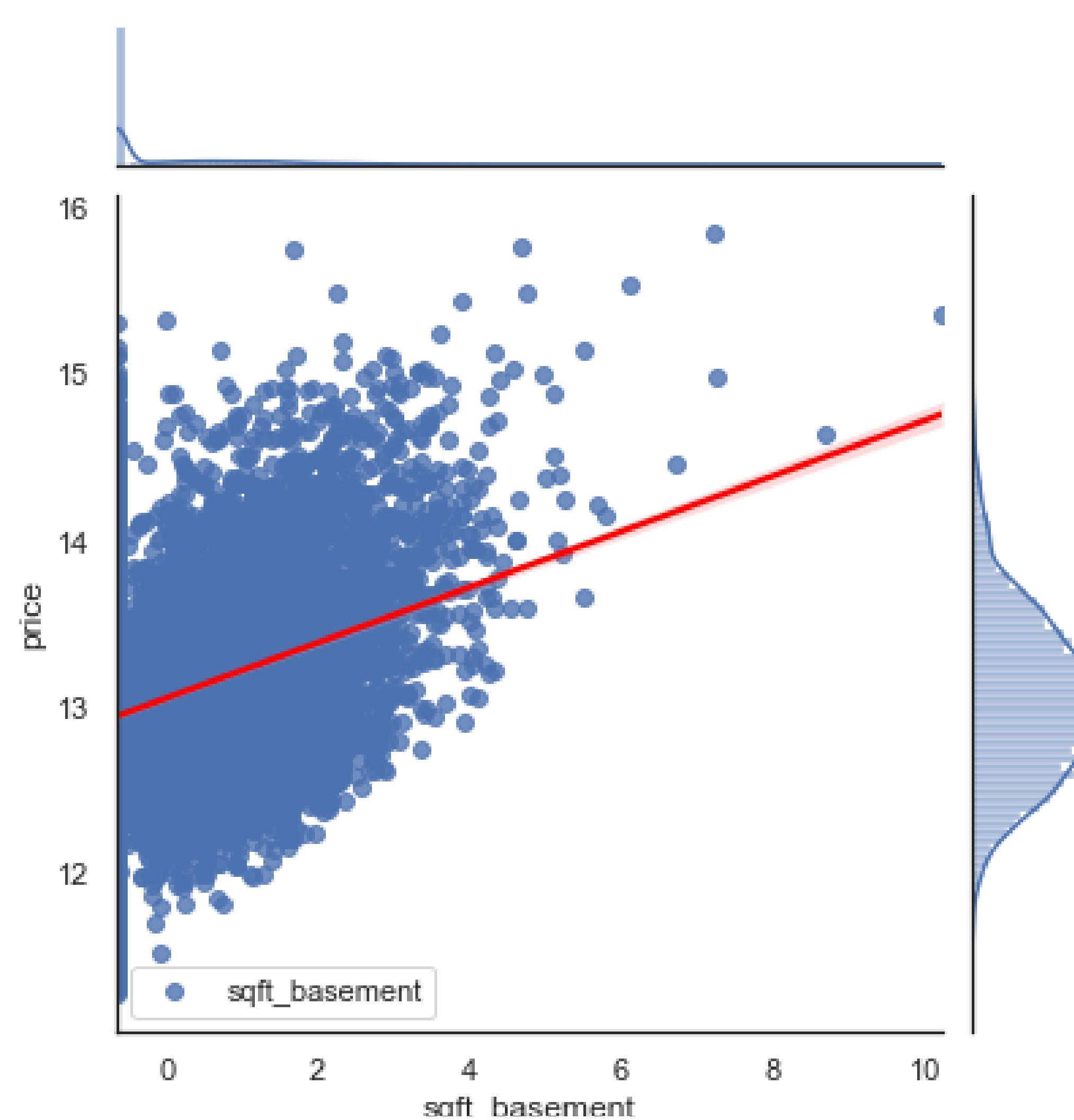
All the histograms, with  
an exception of the lat  
histogram,  
are highly positively  
skewed



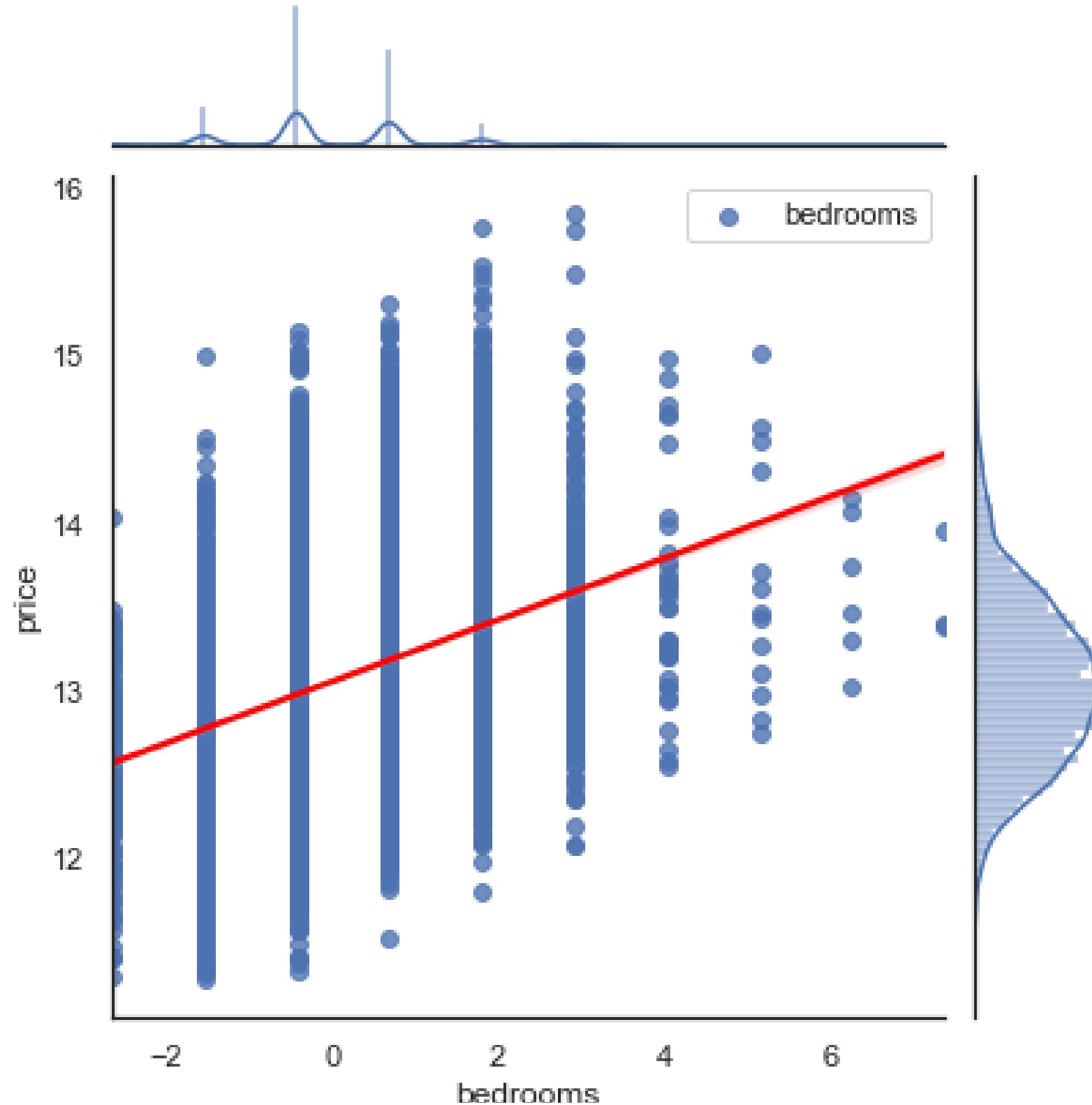
`sqft_living15` has a very  
strong positive  
relationship with the  
price



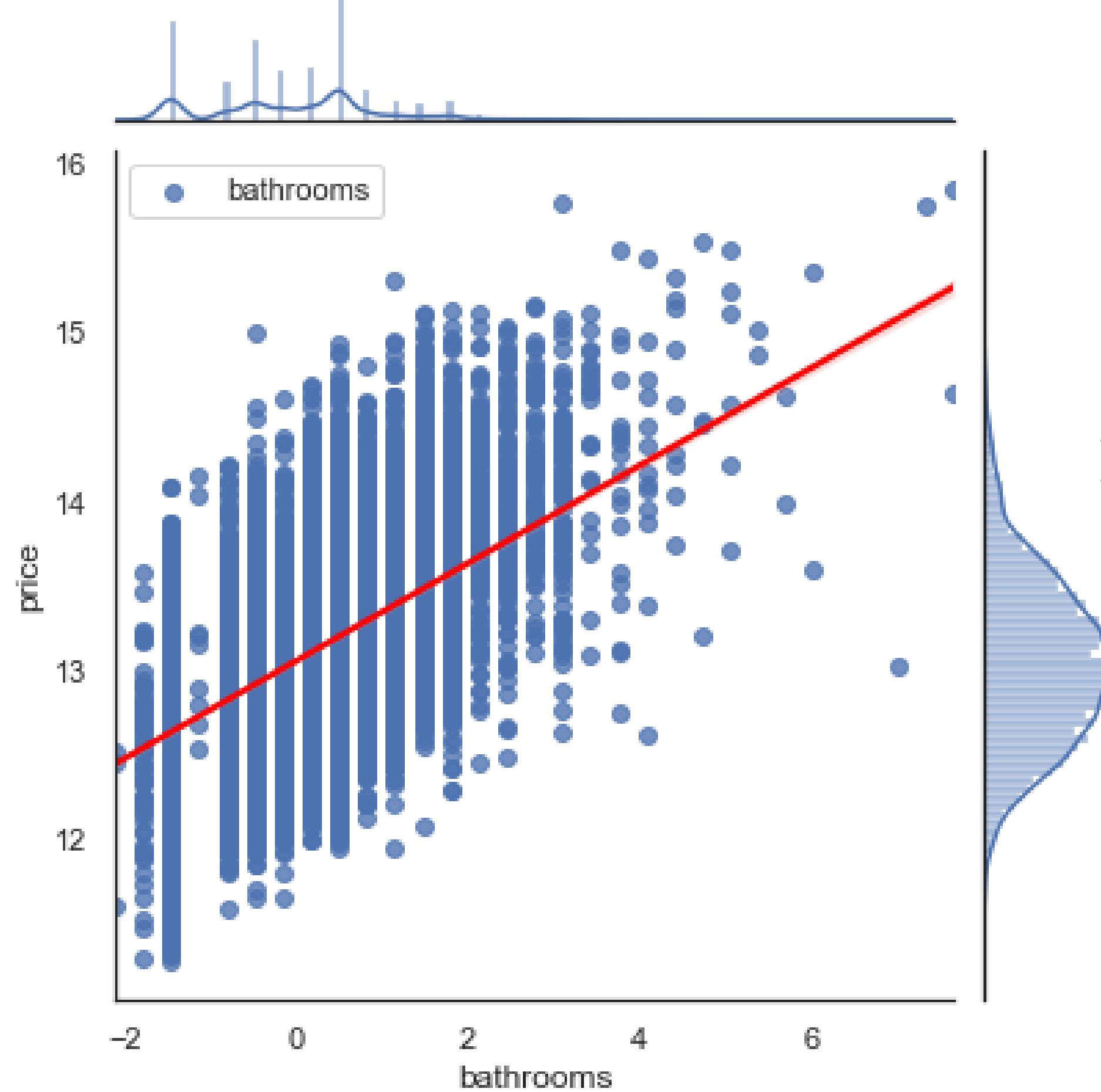
From the data `sqft_above` has a very strong positive relationship with price



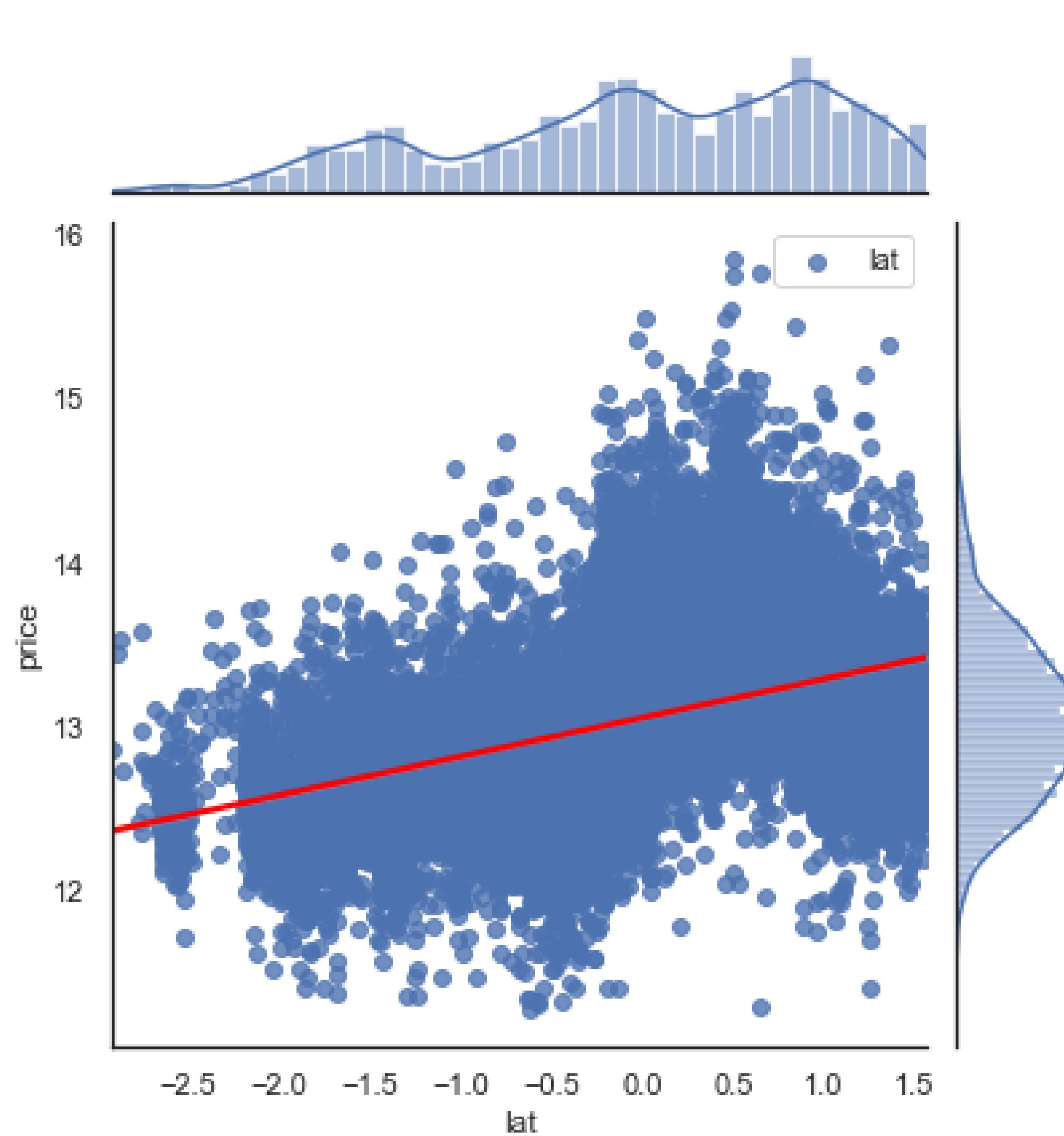
`sqft_basement` has a  
very strong positive  
relationship with price.



Bedrooms has a very strong positive relationship with price.



Bathrooms from the data has a very strong positive relationship with price



lat has a very strong positive  
relationship with price.

# MODELING



# MODELING

- A multilinear regression model was used to predict the sale of property based on various variables like the square foot, number of bedrooms and zipcode.
- The model was trained on the training set and evaluated on the set using Mean Squared Error and R-squared as the evaluation metrics for success.

# METRIC OF SUCCESS

- The model gave us a mean squared error of 0.04 on the test set, and was able to achieve an R-squared score of 0.85, which shows that 85% of the variance in the data is explained by the predictors, which shows a good fit.
- This indicated that the model is able to accurately predict the sale price of a property based on the input variables.

## CONCLUSION:

Our final model only includes the predictor variables that had r<sup>2</sup> values of .1 or higher.

The final predictors were sqft\_living15 with an r<sup>2</sup> value of 0.385, sqft\_above with 0.362, bathrooms with 0.304, lat with 0.202, bedrooms with 0.123, sqft\_basement with .101, and zipcode with .531.

Floors and view both had r<sup>2</sup> values of .097 which we rounded up to .1. and all of our predictors are statistically significant with p-values smaller than .05.

Our r-squared value overall was .85

# RECOMMENDATIONS

- Square footage is the best predictor of a house's price in King County and homeowners who are interested in selling their homes at a higher price should focus on expanding it. When expanding square footage, homeowners should consider building additional bathrooms and bedrooms, as this analysis suggests that number of bathrooms and bedrooms is highly positively related to price.
- We should use the current model to predict price ranges by zip code so that home buyers can see what their macro options are for housing by location and price. The model can also provide general price modifications to a house such as knowing how much an additional bedroom or bathroom would cost, and how the price changes depending on the addition or subtraction of square footage.

# LIMITATIONS AND STEPS TO TAKE;

- The accuracy of our linear regression models might not be sufficient for launching the business, but they are a step in the right direction. An expert could do a better job making predictions, but of course not at the scale and speed of our models.
- To make further improvements to our model, we need additional data. To start, better location information, specifically neighborhood level data, historical price information for each house and real time localized information about the housing market.

ANY QUESTIONS?



THANK YOU

Stay  
Fearless

