Predicting the prices of real estate

OMAR STEITIEH

Introduction

It is hard sometimes to look for real estate within your budget and its is very time consuming to go around visiting different real estates. This project aims to make this easier and being able to look for desired houses easily

Data acquisition and cleaning

1. Houses around Washington data was provided by another course however the data was so unpredictable due to having many factors that affect the price of a certain real estate.

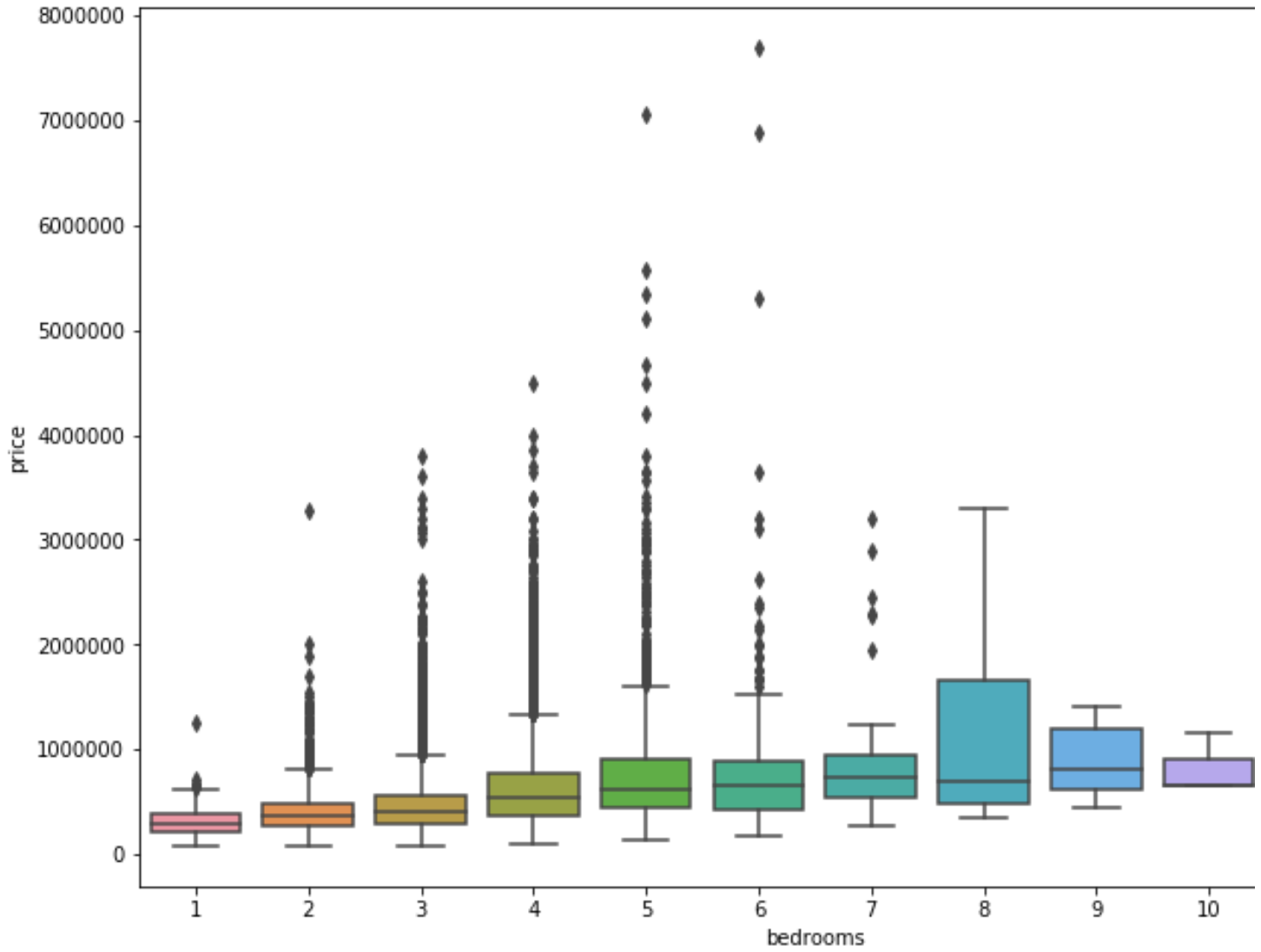
2. A couple problems came up. First, few real estates had large number of bedrooms which causes the data to be uncertain therefore number of bedrooms greater than 3 was not included in the study. Secondly, Data cannot be sorted due to the different categories each rea estate holds.

```
long           -0.055727
zipcode        -0.031147
month          -0.009857
year            0.025255
sqft_lot15      0.056414
id              0.060359
yr_renovated    0.064344
yr_built        0.064555
condition       0.066904
sqft_lot        0.087743
bedrooms        0.105407
floors          0.172002
waterfront      0.272824
sqft_basement   0.340887
lat             0.348405
bathrooms       0.365869
view            0.383350
sqft_living15   0.505287
sqft_above      0.537542
grade           0.556985
sqft_living     0.623682
price           1.000000
Name: price, dtype: float64
```

Exploratory Data

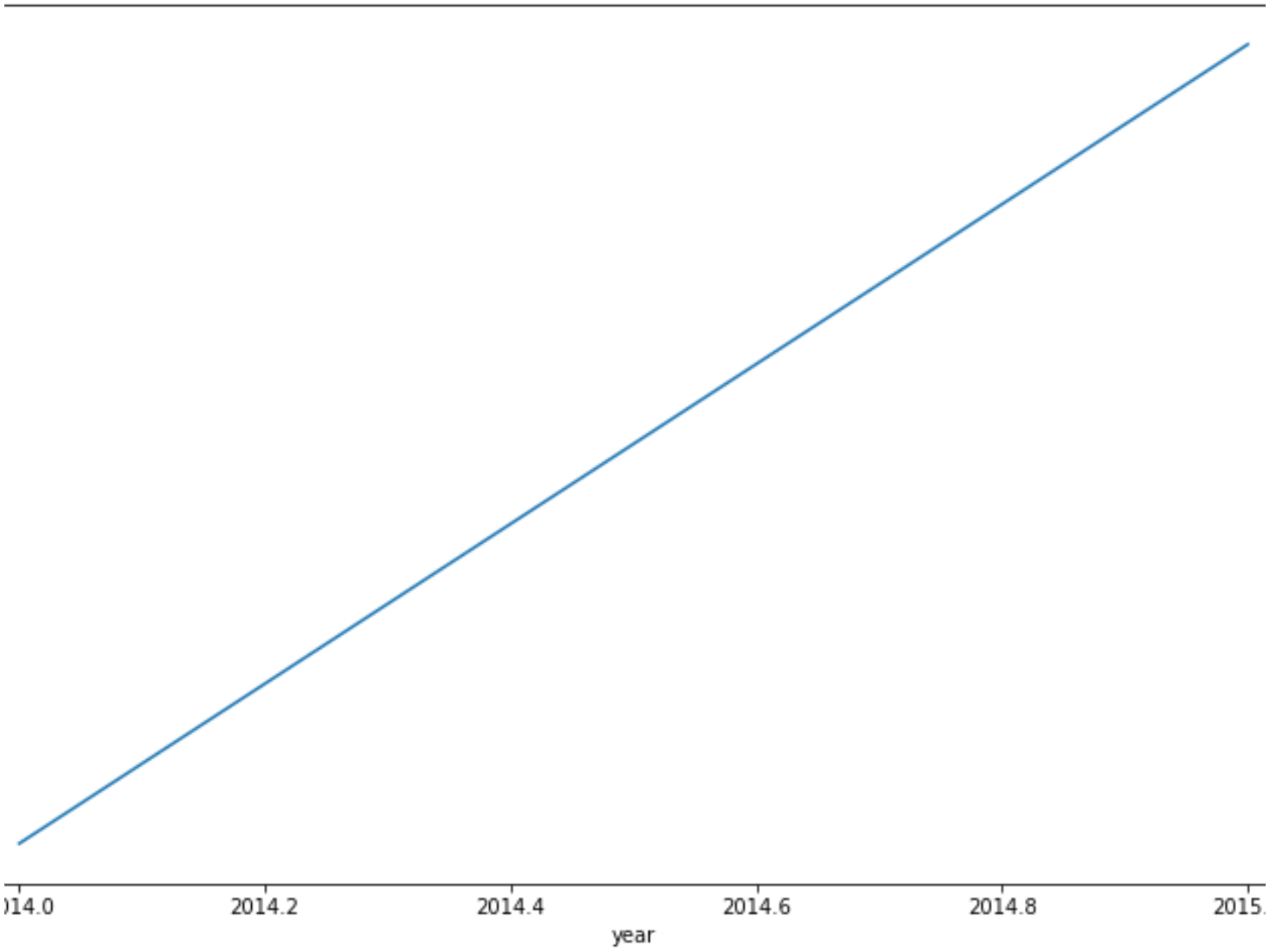
The main feature people tend to focus on while looking for real estate is the Price (our target variable). What affects the price the most? It is how bug is it and its grade.

The following figure values will say sort underscore values and here I can see things that are either highly positively correlated or highly negatively correlated.



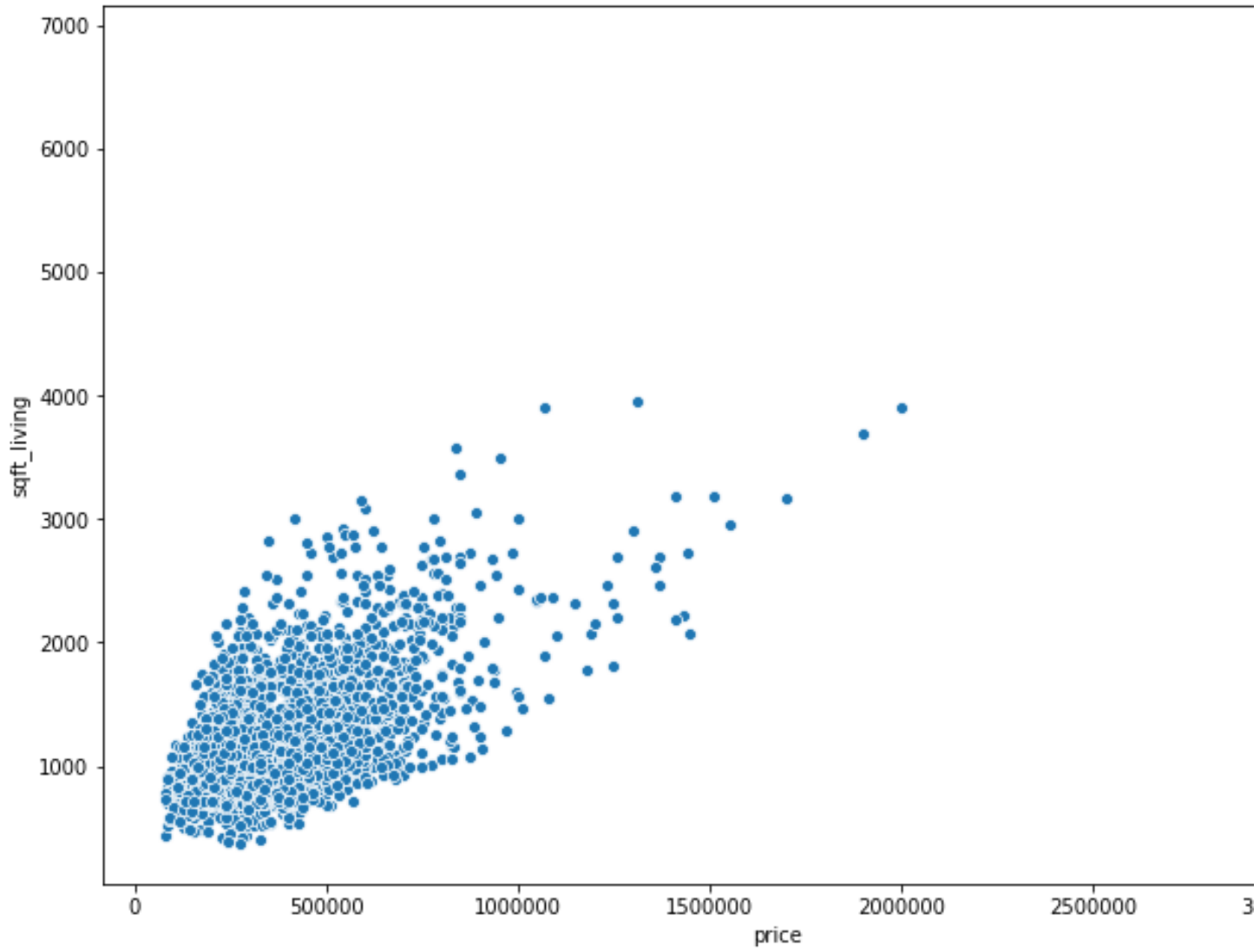
Relationship between Price and number of bedrooms

As you can see there are different varieties of bedrooms but at least we can have an idea of the correlation of the number of bedrooms against Price.



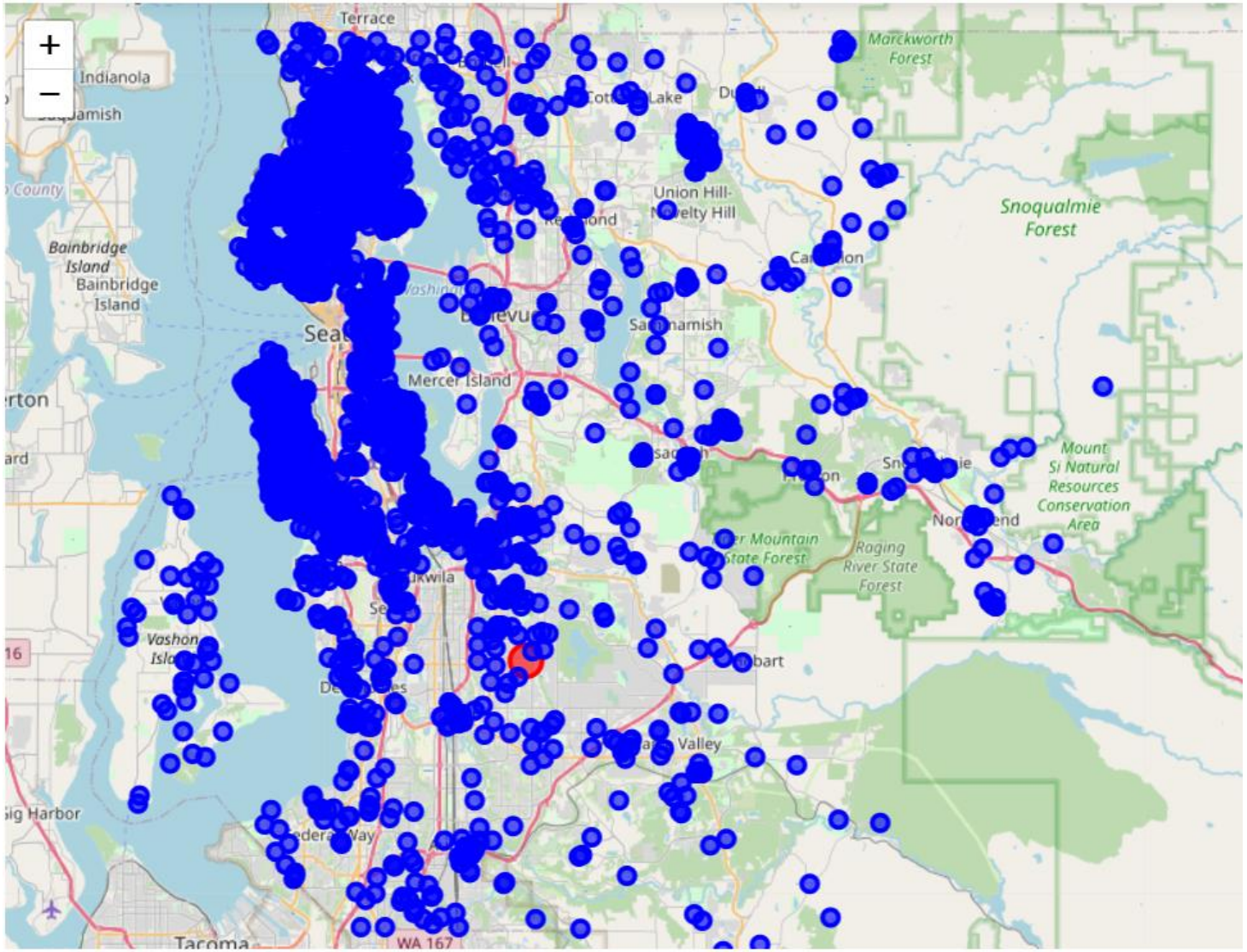
Relationship between Price and Year built

As we can see it is a linear relationship as newer as it gets the higher the price disregarding other categories.



Relationship between Price and Square feet of the real estate

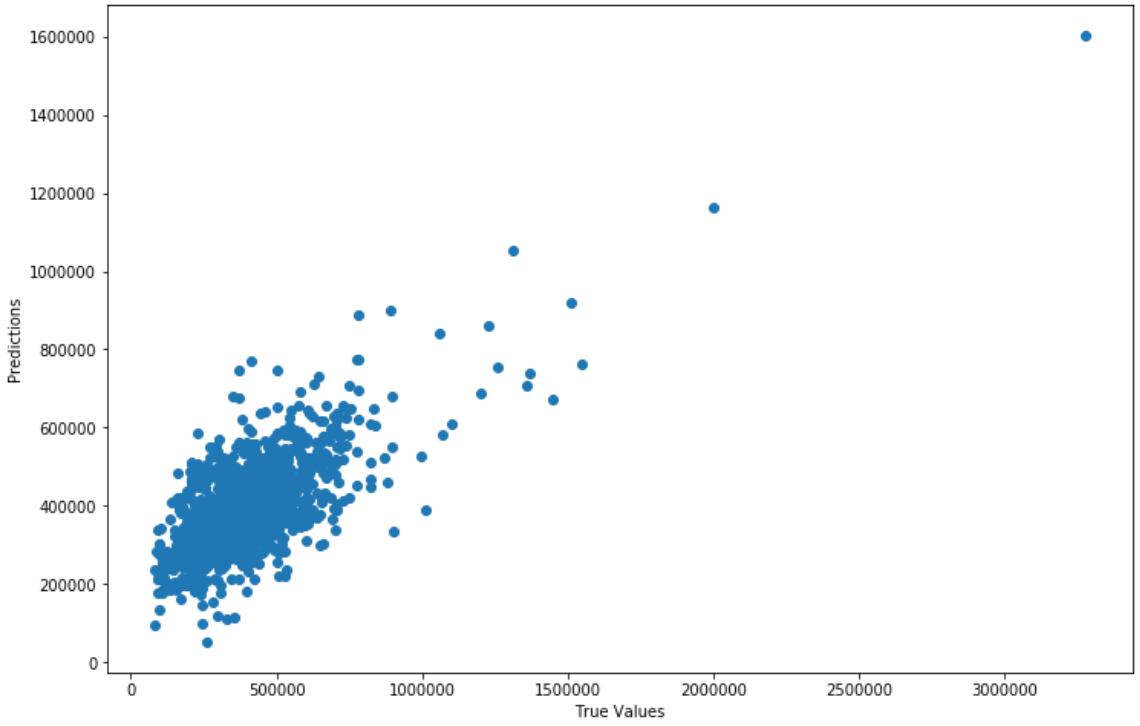
This variable has the maximum correlation with the price according to figure 1.

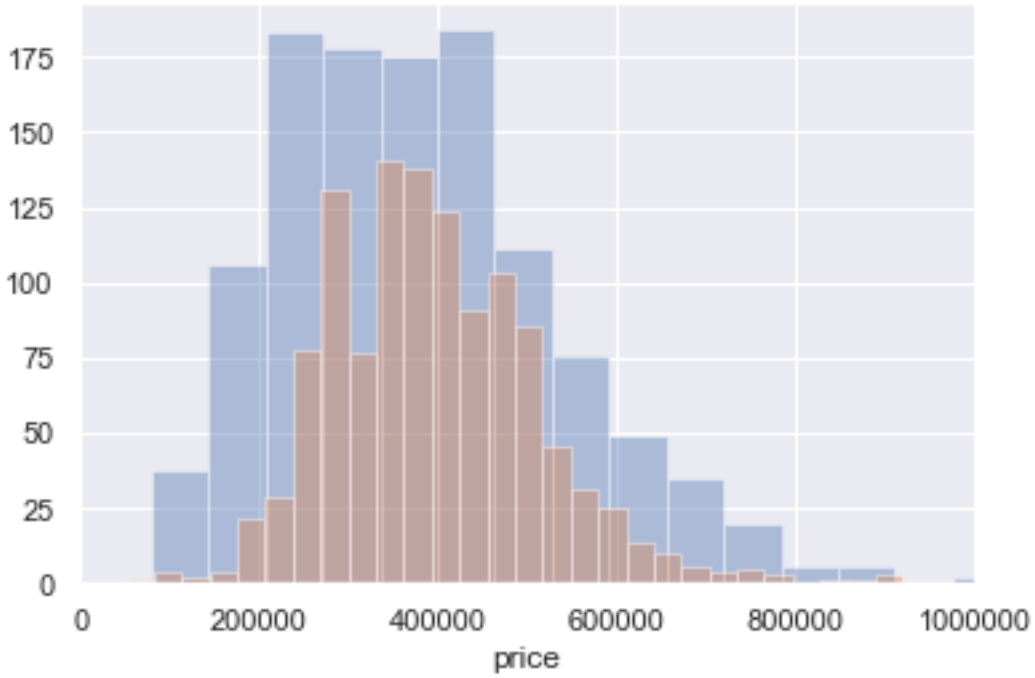


Foursquare API

Foursquare API was used, to be able to determine all listed real estates surrounding my location (in red).

Predictive Modeling





Results

After applying the linear regression. We have achieved an accuracy of 47.8%. due to the various categories available.

Discussion and conclusion

In this study we have achieved a low accuracy due to the various information that can be fed to the model. Also, the model in this study mainly focused on individual features. However for future references we might have to categorize the such huge data to be able to study each category alone and that can be done by classification.