

# PREDICTING KING COUNTY, WASHINGTON HOUSING PRICES - EDA REPORT

## I. Data

This is a list of column names with their description and data type

- **id:** a notation for a house - Numeric
- **date:** Date house was sold - String
- **price:** Price is prediction target - Numeric
- **bedrooms:** Number of Bedrooms/House - Numeric
- **bathrooms:** Number of bathrooms/bedrooms - Numeric
- **sqftliving:** square footage of the home - Numeric
- **sqftlot:** square footage of the lot - Numeric
- **floors:** Total floors (levels) in house - Numeric
- **waterfront:** House which has a view to a waterfront - Numeric
- **view:** Has been viewed - Numeric
- **condition:** How good the condition is ( Overall ). 1 indicates worn out property and 5 excellent.(<http://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g>) - Numeric
- **grade:** overall grade given to the housing unit, based on King County grading system. 1 poor ,13 excellent. - Numeric
- **sqftabove:** square footage of house apart from basement - Numeric
- **sqftbasement:** square footage of the basement - Numeric
- **yrbuilt:** Built Year - Numeric
- **yrrenovated:** Year when house was renovated - Numeric
- **zipcode:** zip - Numeric
- **lat:** Latitude coordinate - Numeric
- **long:** Longitude coordinate - Numeric
- **sqftliving15:** Living room area in 2015(implies-- some renovations) This might or might not have affected the lotsize area - Numeric
- **sqftlot15:** lotSize area in 2015(implies-- some renovations) - Numeric

## II. Examining Price

The price per square foot of a house in King County, Washington

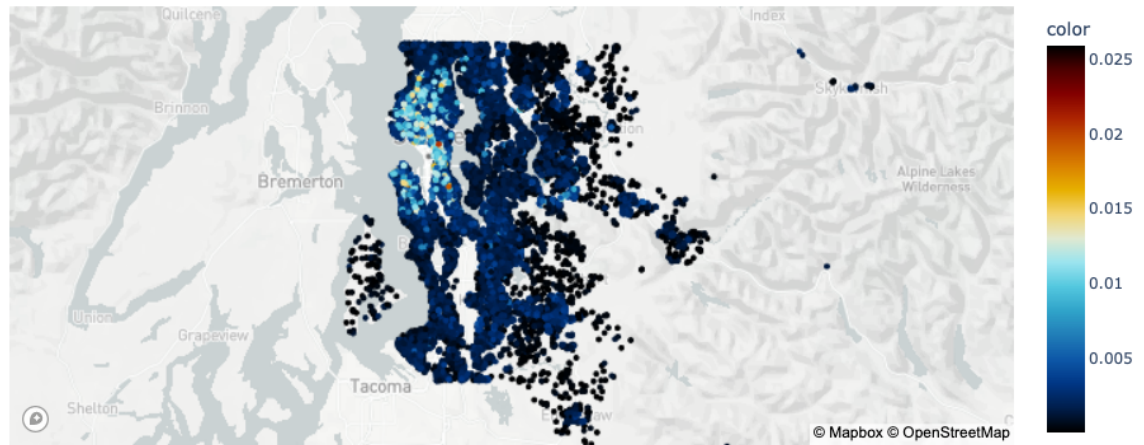


Figure 1.1 Price per sqft. of houses in King County, Washington

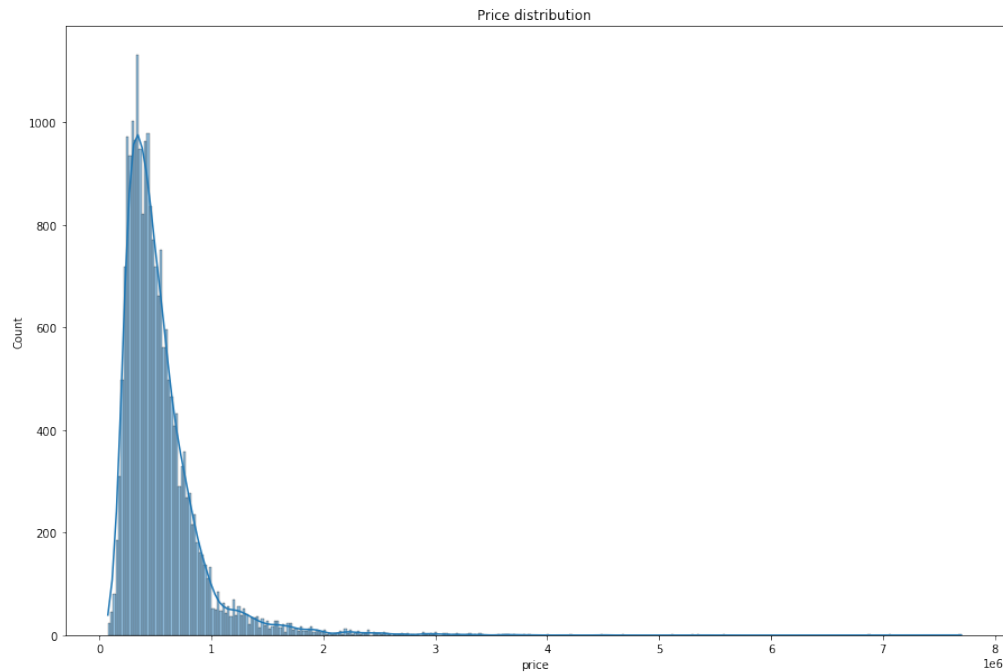


Figure 1.2 Prices of houses in King County, Washington

First we need to take a closer look at the area. Looking at the region of King County, Washington in Figure 1.1, the closer the house is to Seattle, Washington, the more expensive it is. The color key represents the house price per sqft. The map shows that the closer the house is to Seattle, Washington, the more expensive the house per square foot is. Figure 1.2 shows the distribution of the prices of houses.

### III. Compare price to other columns

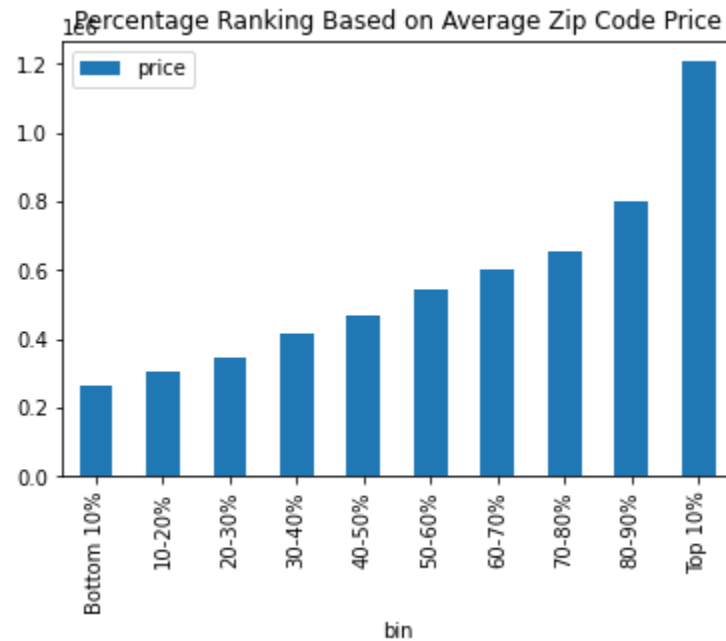


Figure 2. Percentage Ranking Based on Average Zip Code Price

I divided the data into 10 equal groups based on price (top 10%, 80-89%, 70-79%, etc) . The top 10% group had the highest bar. It seems that house prices are highly unequal between different zip codes. In Figure 2 above, the top 10% of zip codes have an average price 33.83% more than the next 10% of zip codes, and nearly 7x the average price of the bottom 10% of zip codes.

### IV. Price vs Features

The next few figures are showing the relationship between price and certain features.

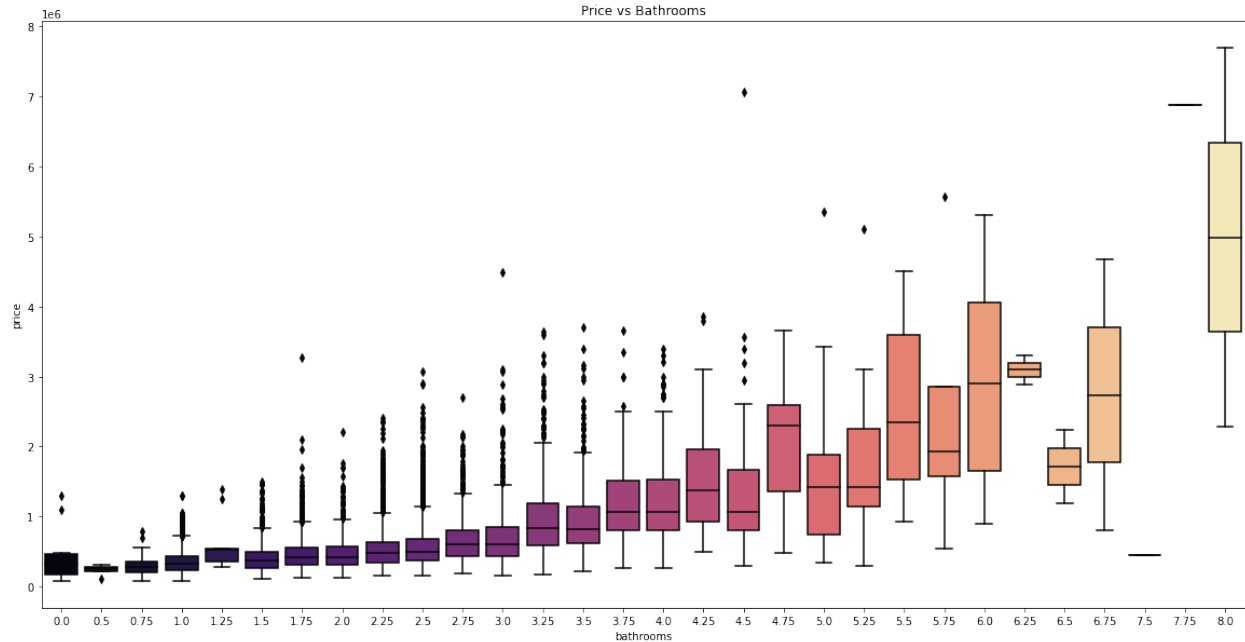


Figure 3. Price vs Bathrooms

Figure 3 above shows the relationship between price in bathrooms. Price and bathrooms have a positive correlation according to our heatmap. Typically, the houses with more bathrooms have a higher price. They also have a higher median.

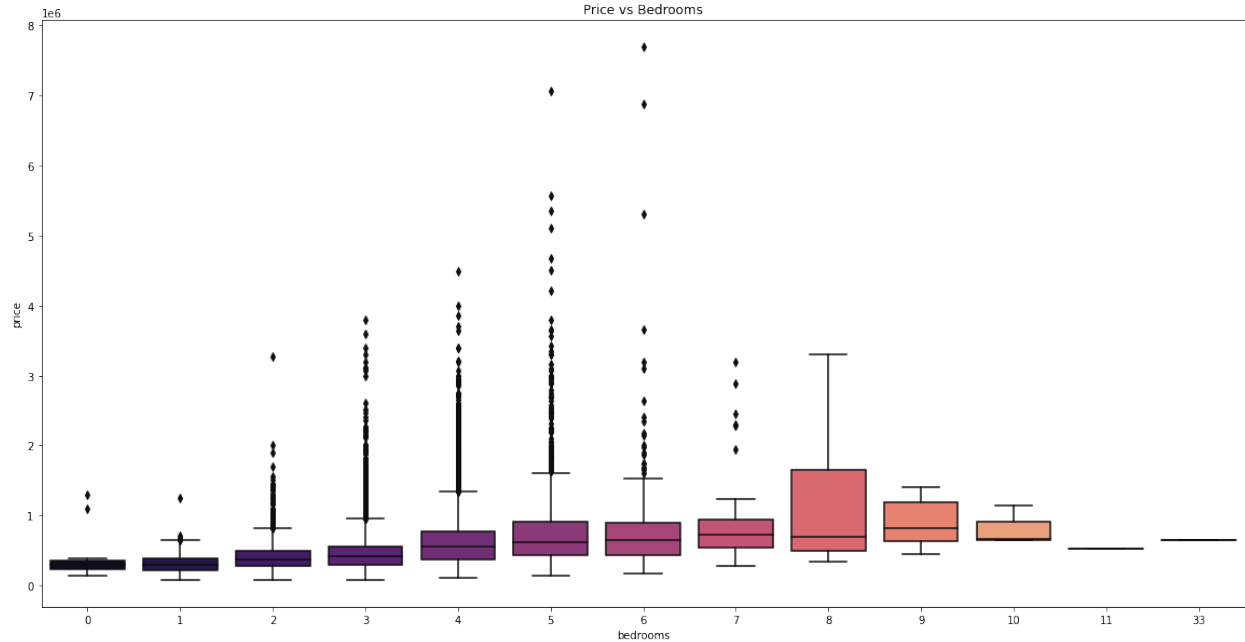


Figure 4 Price vs Bedroom

Figure 4 shows the relationship between price and the number of bedrooms. The houses with 6 bedrooms have the highest house price. The houses with 6 bedrooms have the highest range in the dataset. The houses with 5 bedrooms have the second highest range in the dataset. They

also roughly have similar medians. Based on the correlation heatmap, price and bedrooms have a correlation of 0.31.

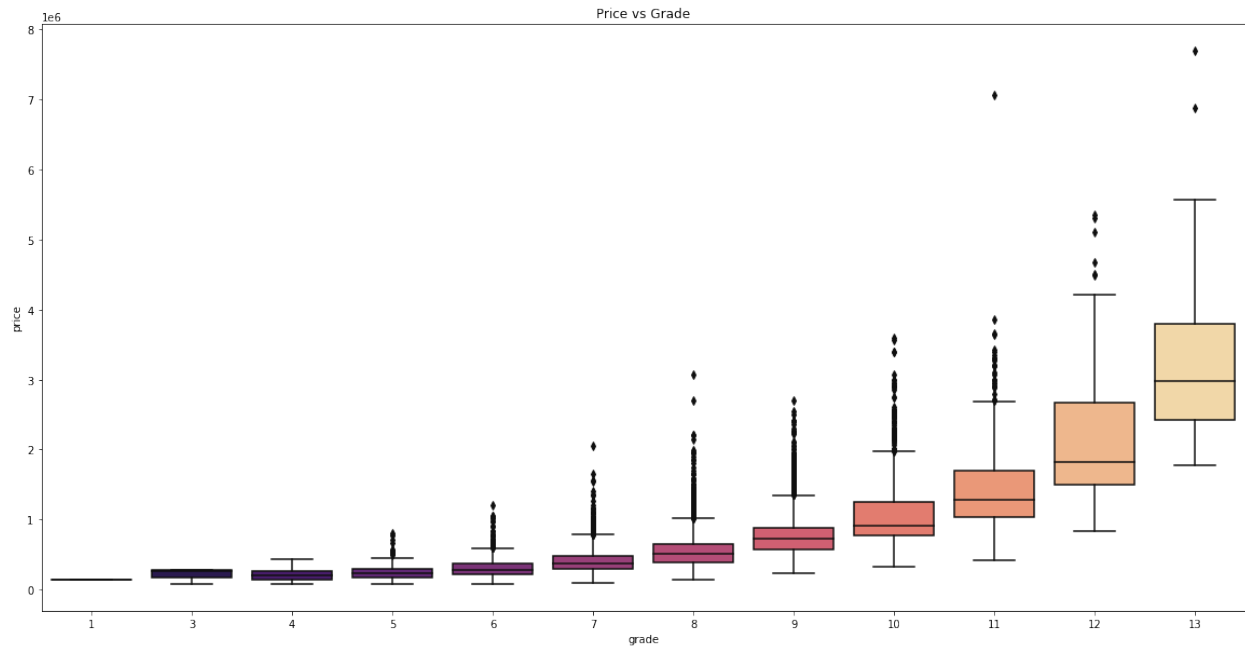


Figure 5 Price vs Grade

Figure 5 shows the relationship between price and grade. Based on the correlation heatmap, it has a positive correlation of 0.67. Typically, the houses with a higher grade have a higher house price. They also have a higher median. You can see that there is a house with a grade of 11 that has a higher outlier compared to the other outliers.

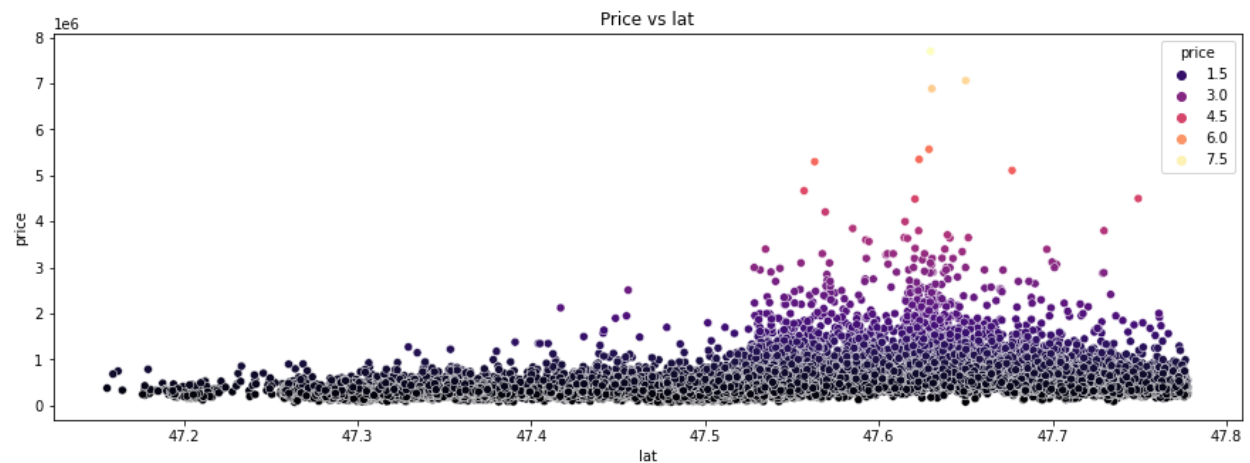


Figure 6 Price vs Latitude

Figure 6 shows the relationship between price and latitude. The houses with latitudes between 47.6 and 47.7 have some of the highest house prices. This is also where most of the outliers lay. The latitude for Seattle, Washington is 47.6062 N. This makes sense that the houses between 47.6 and 47.7 are some of the highest priced houses.

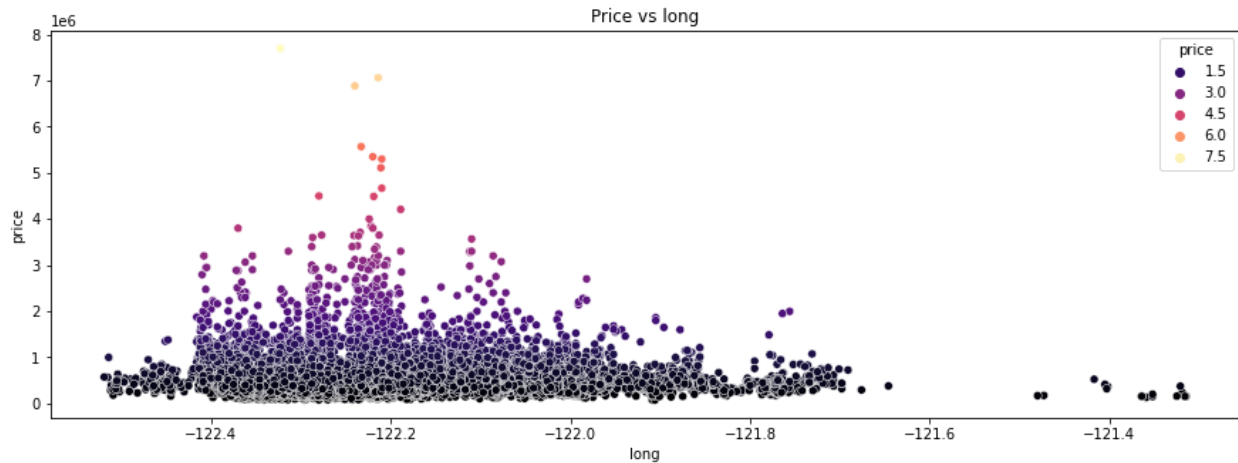


Figure 7 Price vs Longitude

Figure 7 shows the relationship between price and longitude. The more expensive homes have a longitude close to -122.2. Seattle, Washington longitude is 122.3321 W. Based on our correlation heatmap, longitude does not have a significant relationship with price. It has a correlation of 0.022.

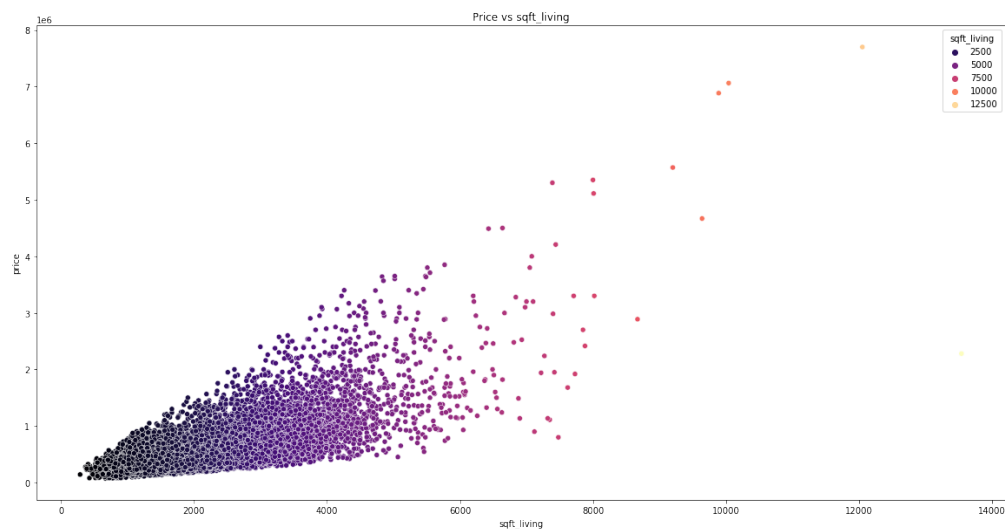


Figure 8 Price vs sqft\_living

Figure 8 shows the relationship between price and sqft\_living. The most expensive home has 12000. Most of the homes are between 2,000 and 6,000 square feet. Based on the correlation heatmap, price and sqft\_living have the highest correlated relationship with 0.70. Sqft\_living is the most important feature for determining price.

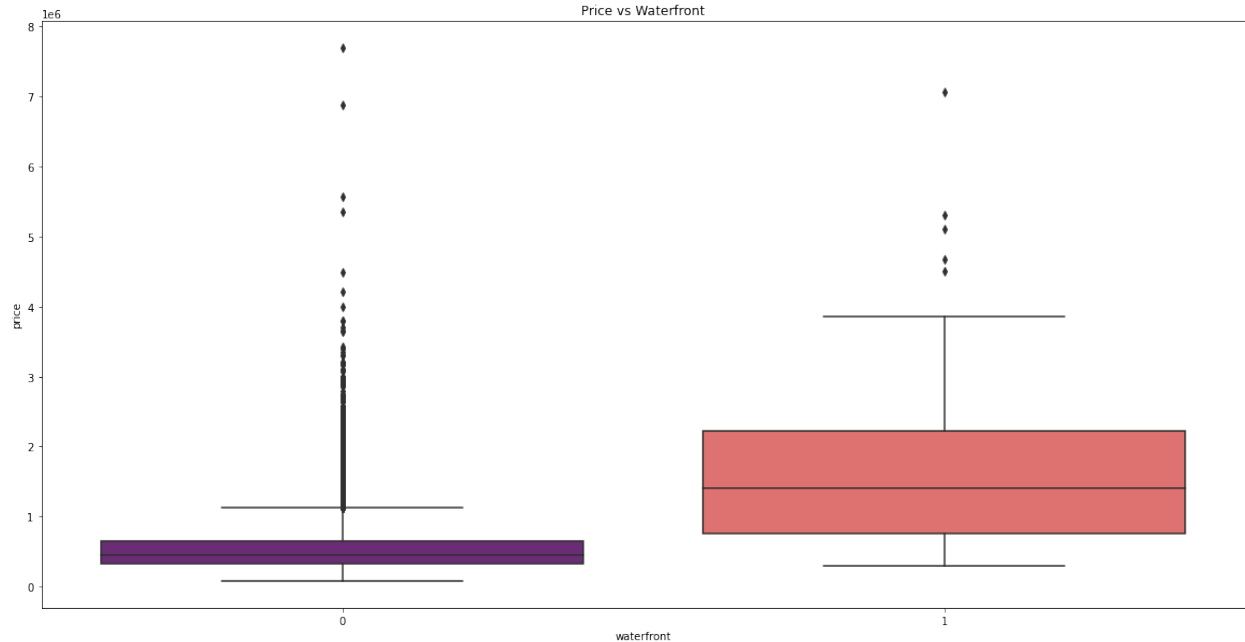


Figure 9 Price vs waterfront

Figure 9 shows the relationship between price and waterfront. Typically, houses with a waterfront view have a higher median house price, but the houses without a waterfront have higher outliers.

In this dataset waterfront is a categorical variable. If a house has a 0 for waterfront, then the house does not have a view to a waterfront. If a house has a 1 for waterfront, then the house does have a view to a waterfront.

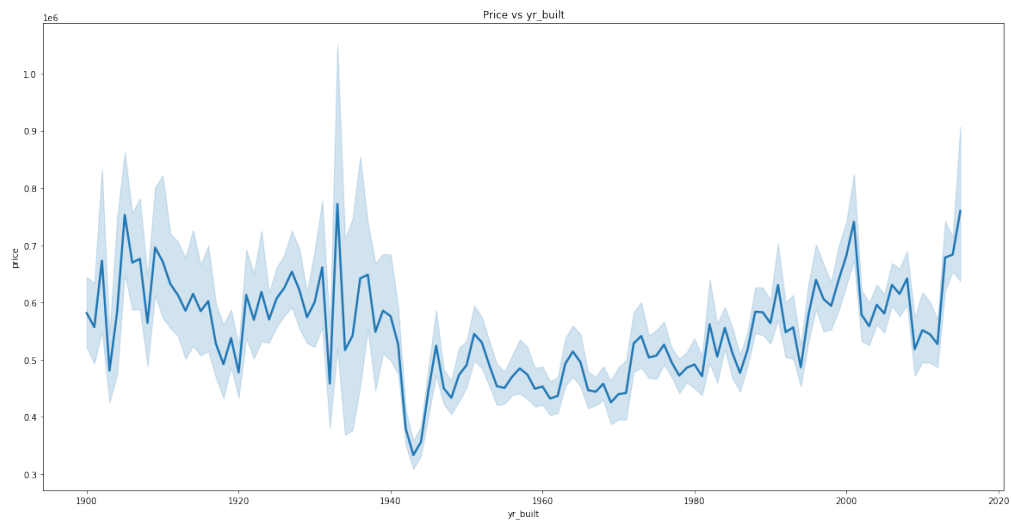


Figure 10 Price vs yr\_built

Figure 10 shows the relationship between price and the year the house was built. You can see the trend of house prices over the years. Based on the correlation heatmap, price and yr\_built does not have a high correlation. It has a correlation of 0.054. The lowest house price was between 1940 and 1960.

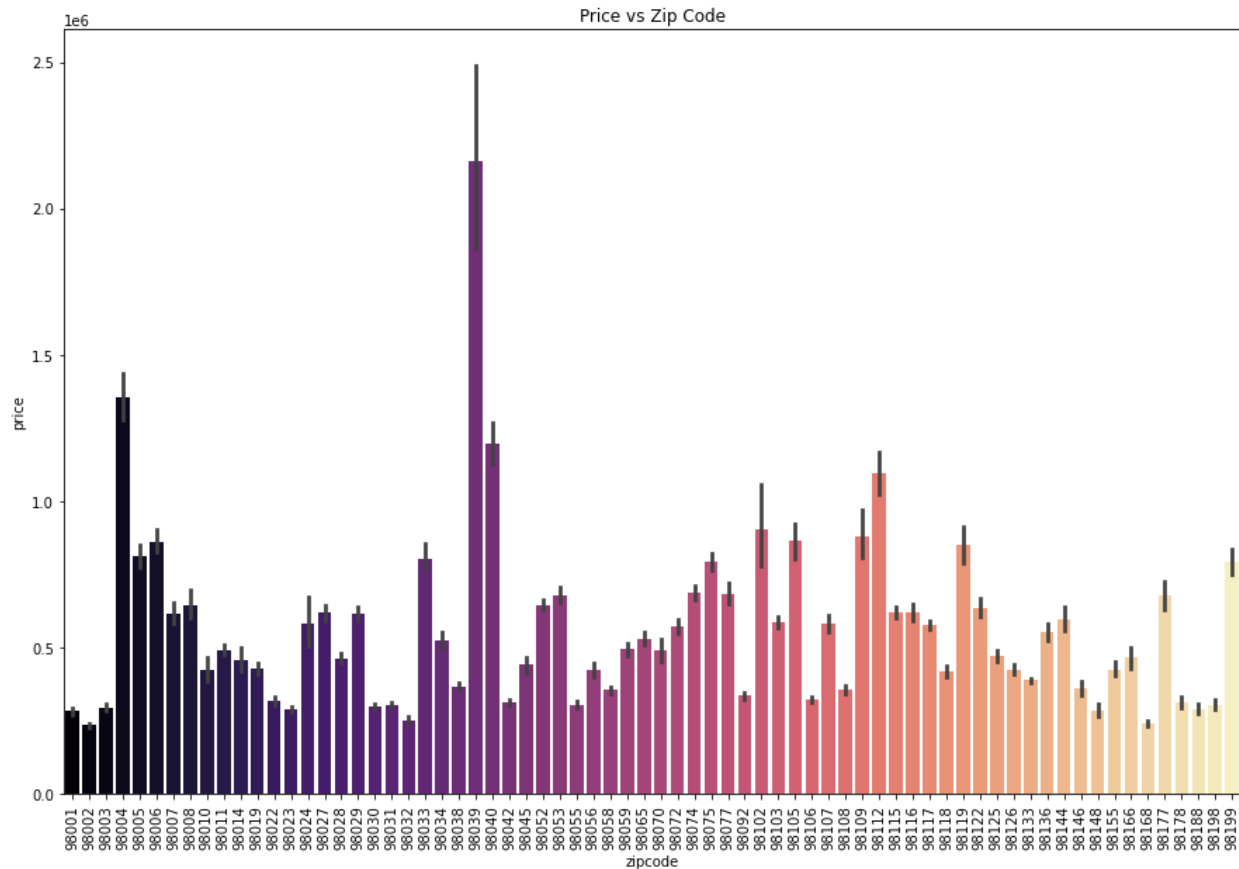


Figure 11.1 Price vs Zipcode

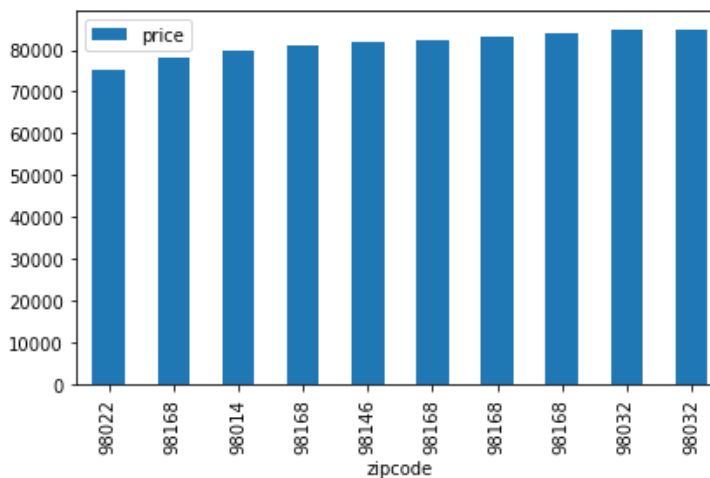


Figure 11.2 Top House Prices and their zip code

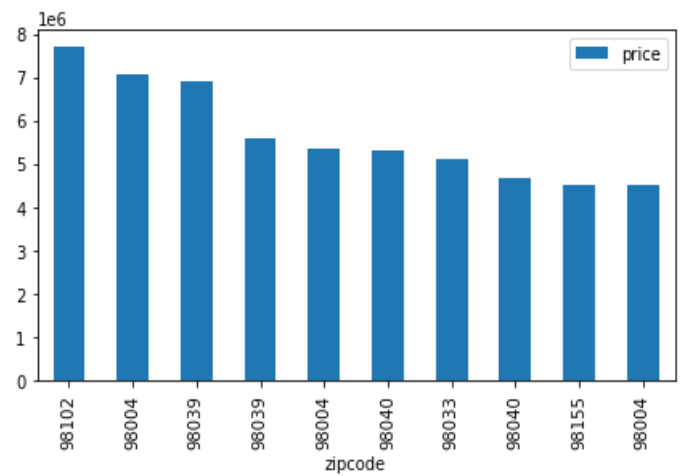


Figure 11.3 Bottom House Prices and their zip code

Figure 11.1 shows the relationship between price and the zipcode of the house location. Figure 11.2 shows the top house prices and their zip codes. Figure 11.3 shows the lowest house prices and their zip codes. Based on the correlation heatmap, price and zipcode have a correlation of -



0.053. A future home buyer/seller can keep in mind that the zip code will not be a major factor in the price.

## V. Feature Correlation

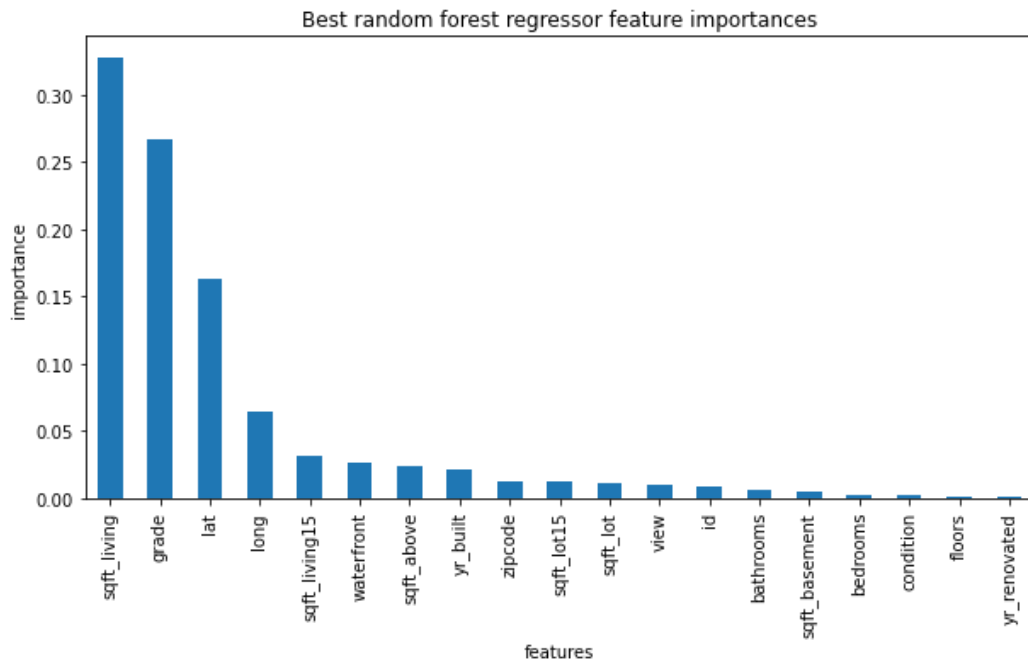


Figure 13 Random Forest Chart

Using a random forest regressor can determine the most important features. Based on Figure 13 above, the random forest chart shows that the top 4 most important features are “sqft\_living”, “grade”, “lat”, “long”. These features are what will affect the price of the house.

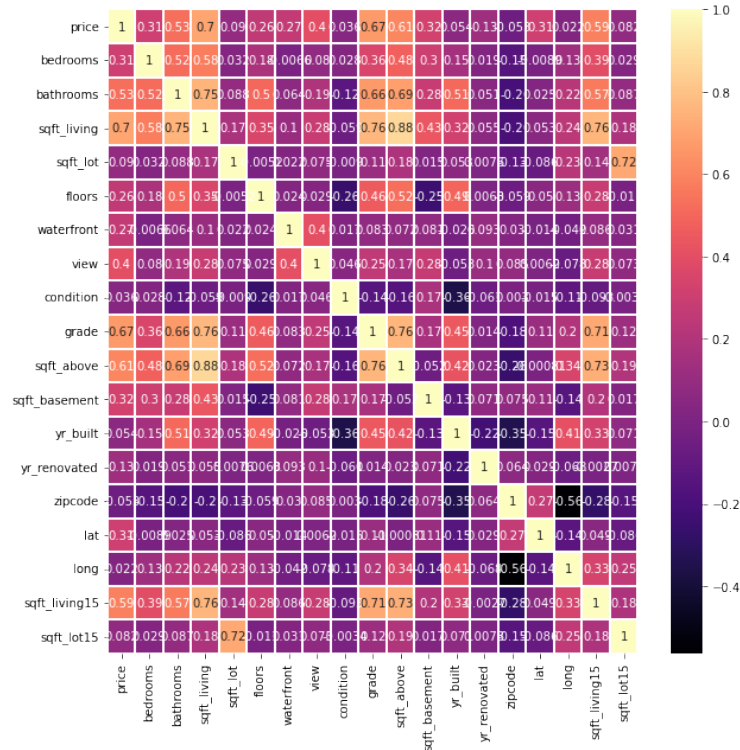


Figure 14 Heatmap Correlation map

The heatmap let us know that there's a strong positive correlation between 'price' and 'sqft\_living'(0.70), 'grade'(0.67), and 'sqft\_above'(0.61).

There is also a strong correlation between 'grade' and 'sqft\_above'(0.76) as well. These features will affect the price of the house.

## Summary

The closer the house is to Seattle, Washington, the more expensive the house price will be.

Based on the heatmap, here's a strong positive correlation between 'price' and 'sqft\_living', 'grade', and 'sqft\_above'.

These features will affect our house price. We currently can expect 'sqft\_living', 'grade', 'lat', 'long' to have the most effect on house price.