



Module 1 Project

Seeta Rajpara
Self-Paced, Online Data Science Program
Flatiron School

King County House Sales Data Analysis Strategy

Upon examination of the column names from this dataset, I wanted to explore the variables further to understand the data structure of each column

Using basic deduction, the following column names warranted a deeper investigation into the relationship between that variable and the price of the house:

- Bedrooms, bathrooms, sqft_(living, basement, above, lot) floors, grade, yr_built, and zip code

```
In [4]: df_housedata.columns
```

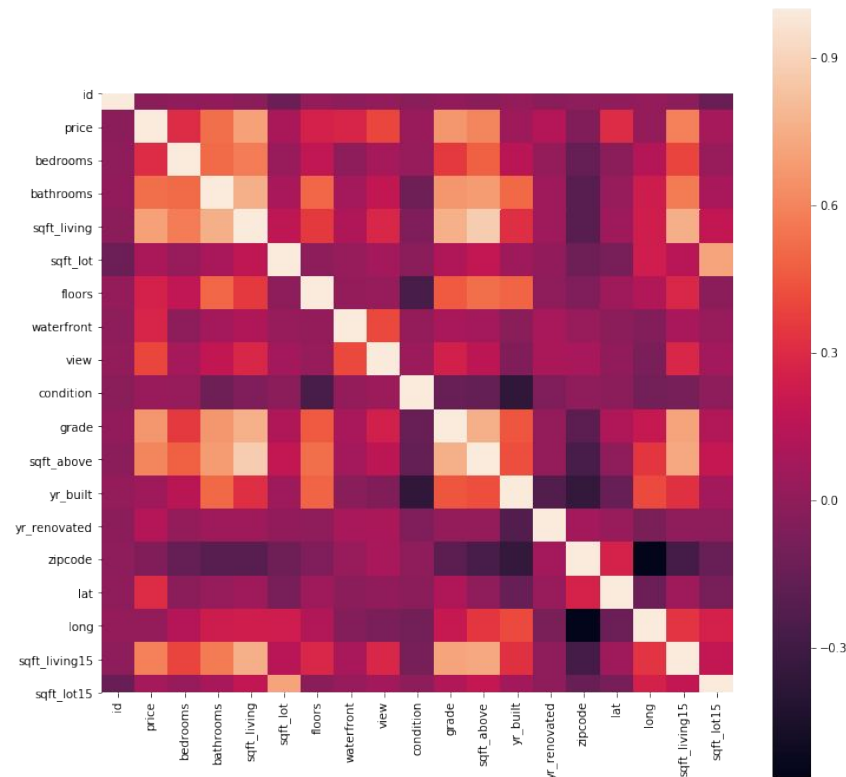
```
Out[4]: Index(['id', 'date', 'price', 'bedrooms', 'bathrooms', 'sqft_living',  
              'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'grade',  
              'sqft_above', 'sqft_basement', 'yr_built', 'yr_renovated', 'zipcode',  
              'lat', 'long', 'sqft_living15', 'sqft_lot15'],  
             dtype='object')
```

Categorical Variables (boxplot visualization):

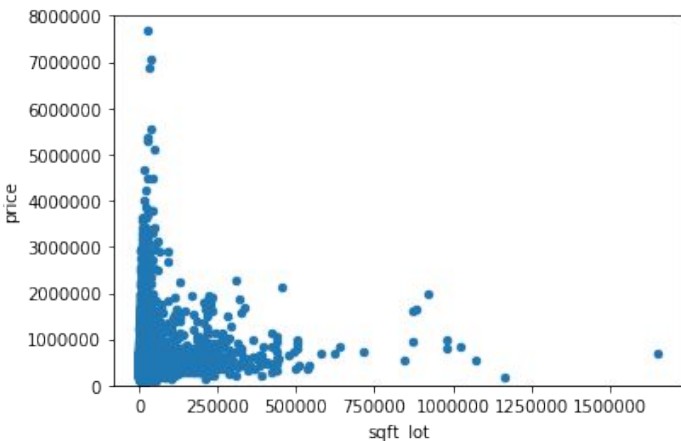
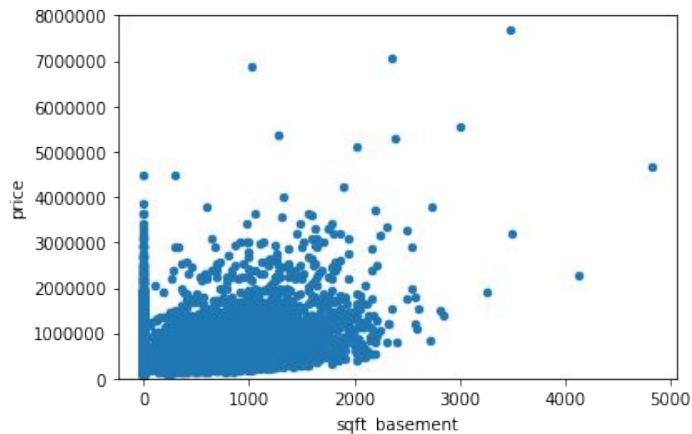
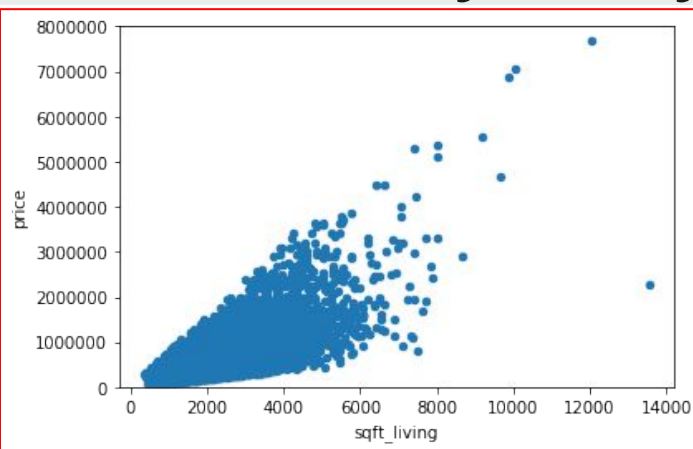
- Bedrooms, bathrooms, grade, year built, zip code, and floors

Quantitative Variables (scatterplot visualization):

- square footage (living, basement, above, lot)

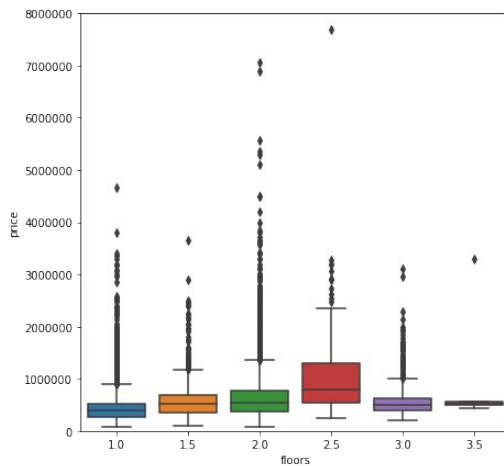
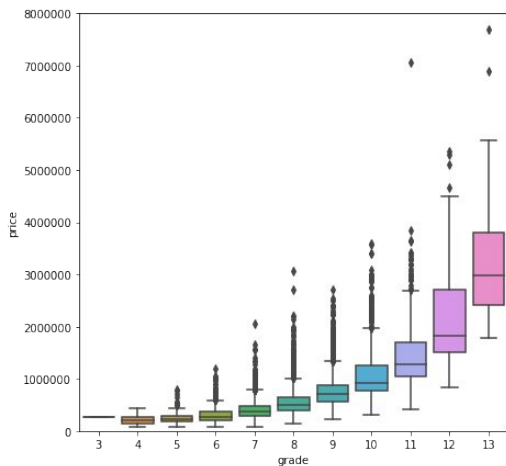


Summary: Analysis of Quantitative Variables

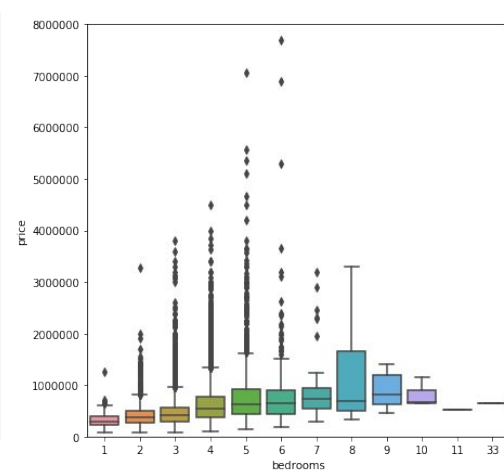
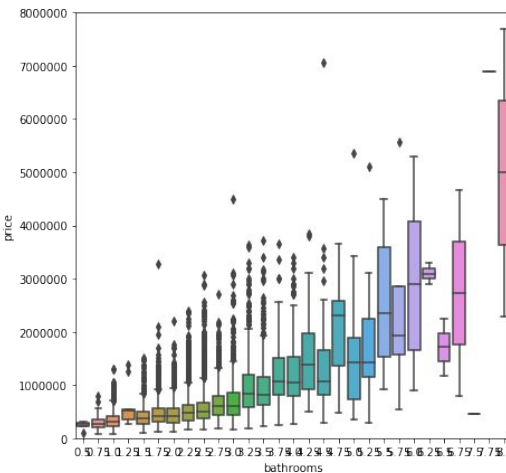


- The most meaningful plot here would be price vs. **sqft_living**: there's an obvious linear relationship with increasing square footage of living space, the price of the home increases
- The data from sqft_lot don't correlate at all, and this plot shows that this data might not be reported for each home
- Some homes do not have basements, and therefore the sqft_basement plot also can be misleading

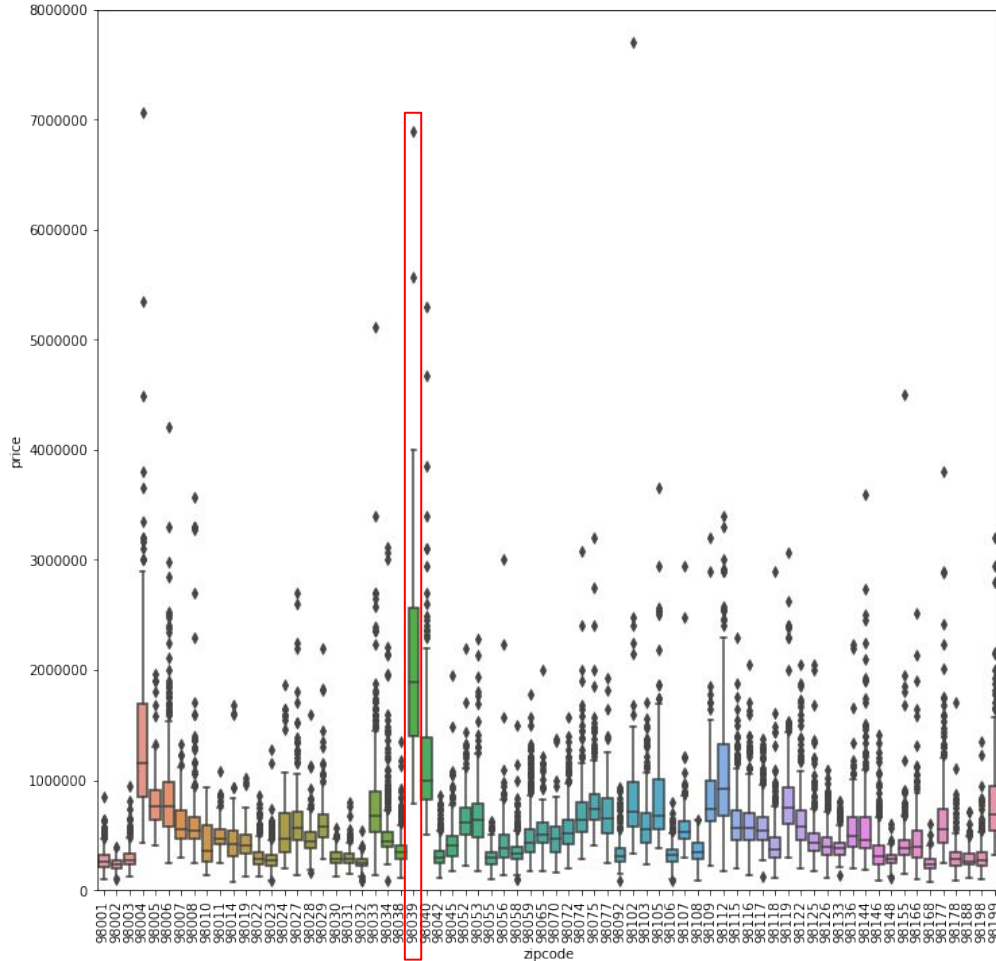
Summary: Analysis of Categorical Variables



- With increasing grade, price increases
- As the number of floors increase, up until 2.5, the price of the house also increases
- In general, the more bedrooms in the house, the higher the price; although this seems to plateau after 8 bedrooms
- With an increased number of bathrooms, the price, as expected, also increases



Zip Codes vs. Housing Prices in King County



The boxplot analysis for zip codes vs. home price in King County shows that one zipcode in particular seems to have the highest housing prices in this analysis: 98039¹

Upon further research, this zip code correlates with a Seattle suburb, Medina, which is home to billionaires like Bill Gates and Jeff Bezos²!

Other zip codes also seem to be home to expensive houses in King County.

- 1) https://en.wikipedia.org/wiki/Medina,_Washington
- 2) <https://www.businessinsider.com/where-billionaires-bill-gates-jeff-bezos-live-photos-of-medina-washington-2017-12>

Conclusions

- In this analysis of housing prices in King County, WA, we can deduce the following variables have influence in housing price:
 - Bedrooms, bathrooms, square footage of living space, grade, floors, and year built.
- In addition, various zip codes in the county are home to some very expensive homes owned by tech billionaires (previous slide)

