
KING COUNTY HOUSING DATA ANALYSIS

FARNAZ GOLNAM





Project Summary

In this project, we are working with the King County House Sales dataset and model this dataset with a multivariate linear regression to predict the sale price of houses as accurately as possible.

Exploratory data analysis

Question1. How does the location affect the housing prices?

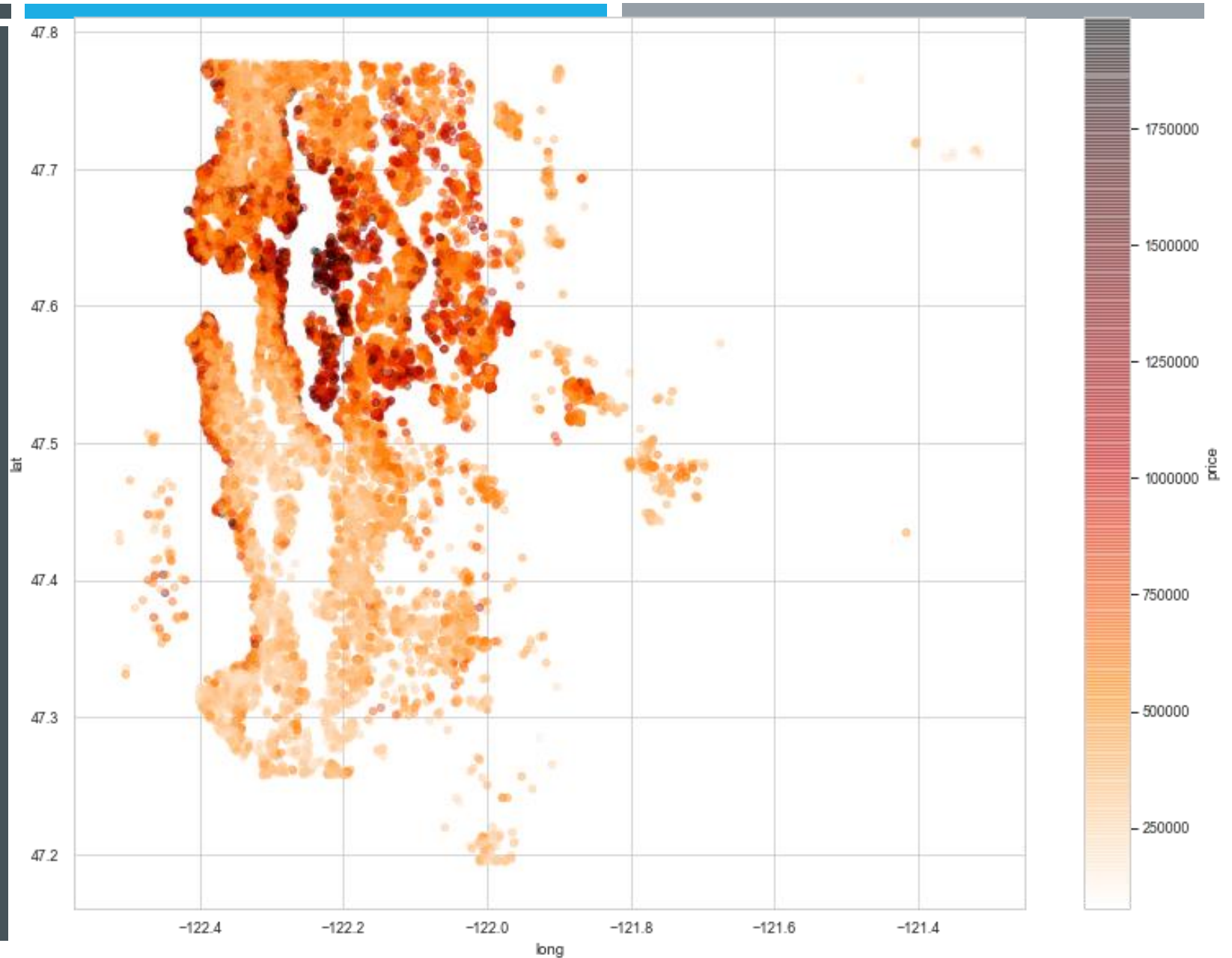
Question2. How does renovation affect the housing price?

Question3. How much difference is between the price of houses with the waterfront view and others?

Question4. what are the most effective factors on the housing price?

1. LOCATION

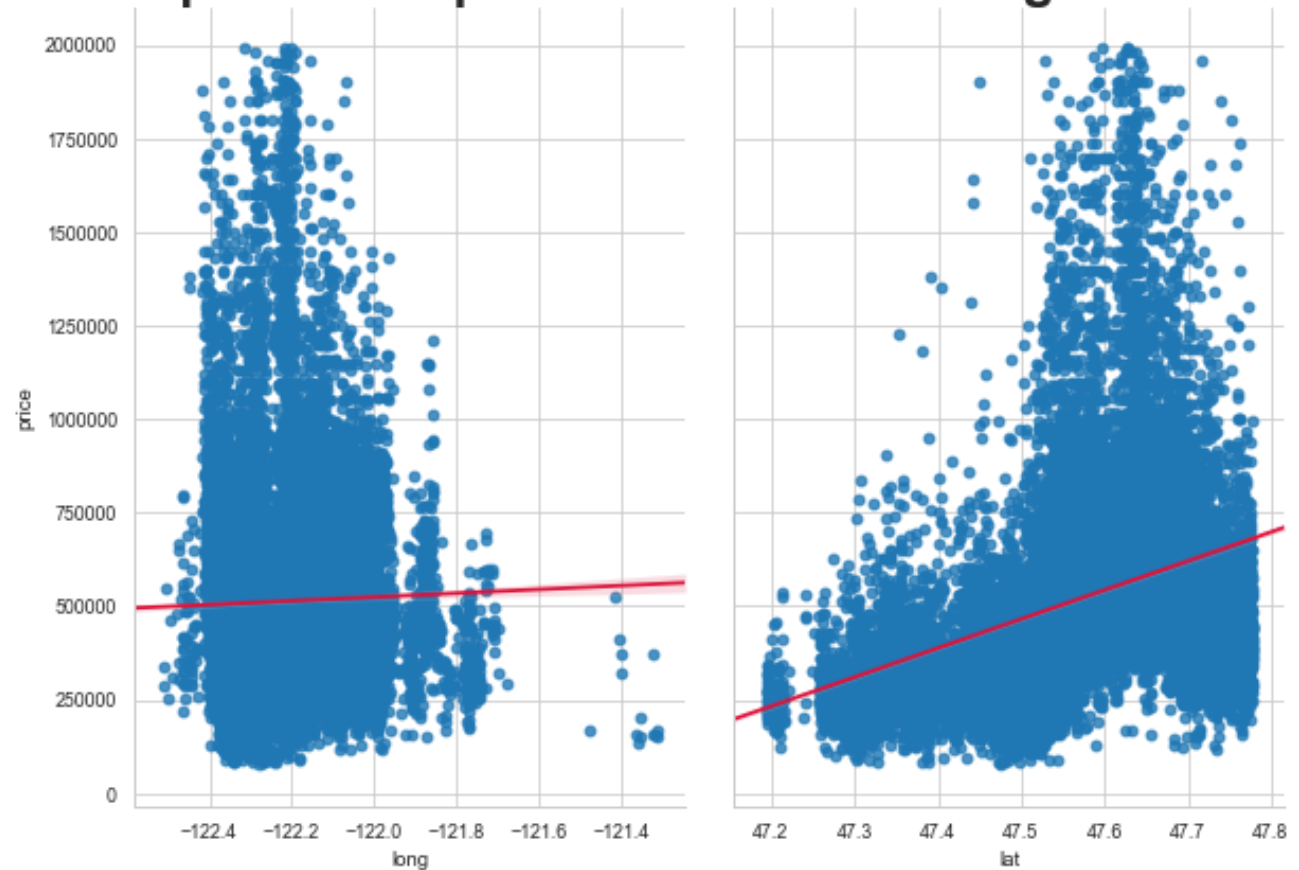
as we can see the latitude(northern-southern) direction has a linear and strong effect on the price but longitude coordinate(western eastern) direction, doesn't have much of effect on price.



LONGITUDE VS LATITUDE

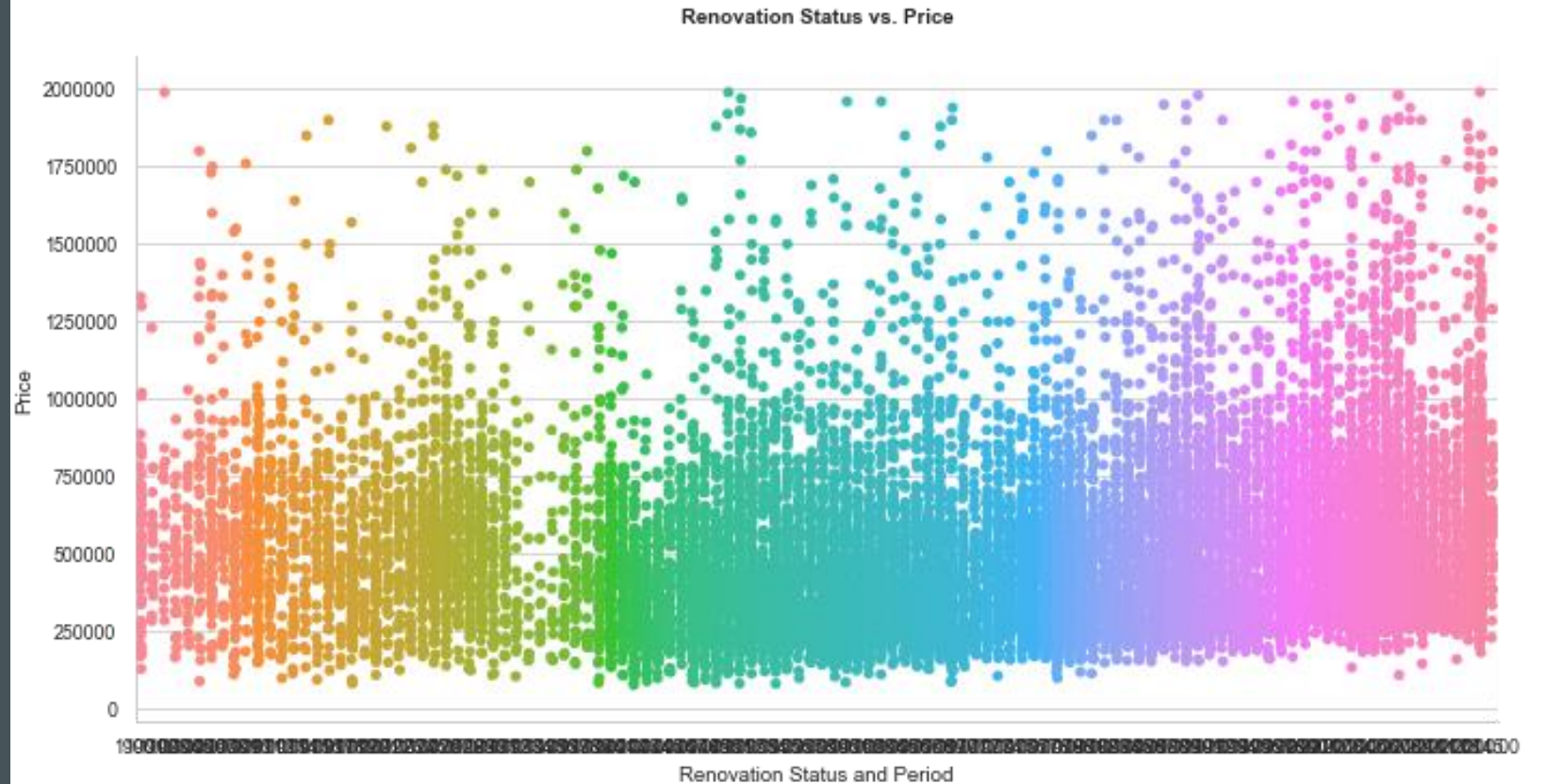
The pair plot shows the relationship between latitude and longitude coordinate with price and strangely indicates that Longitude has way less effect on price than latitude.

Relationship between price and Latitude-Longitude coordinate



2.RENOVATION

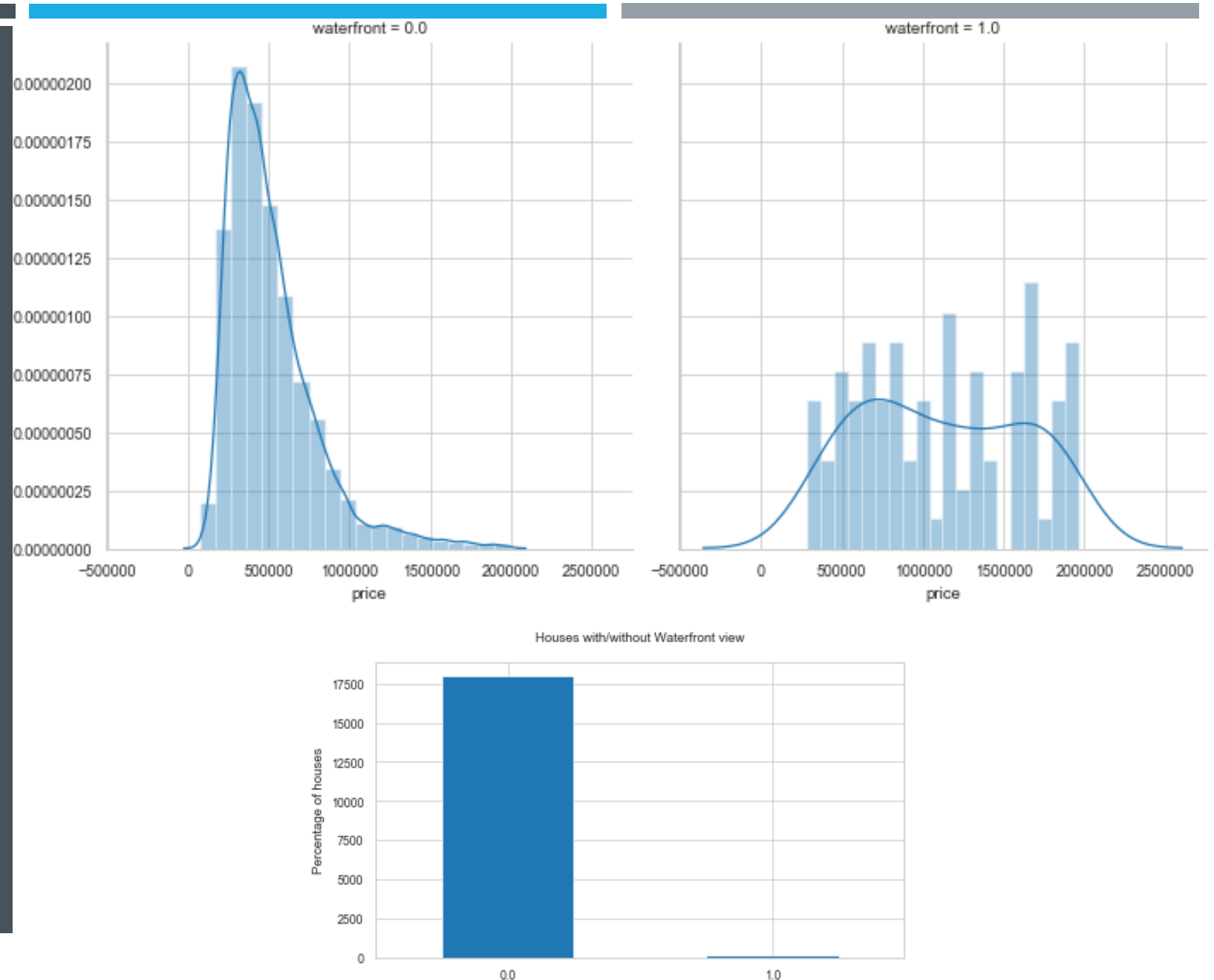
As shown, the more recent the renovation happened the more the prices spread upwards



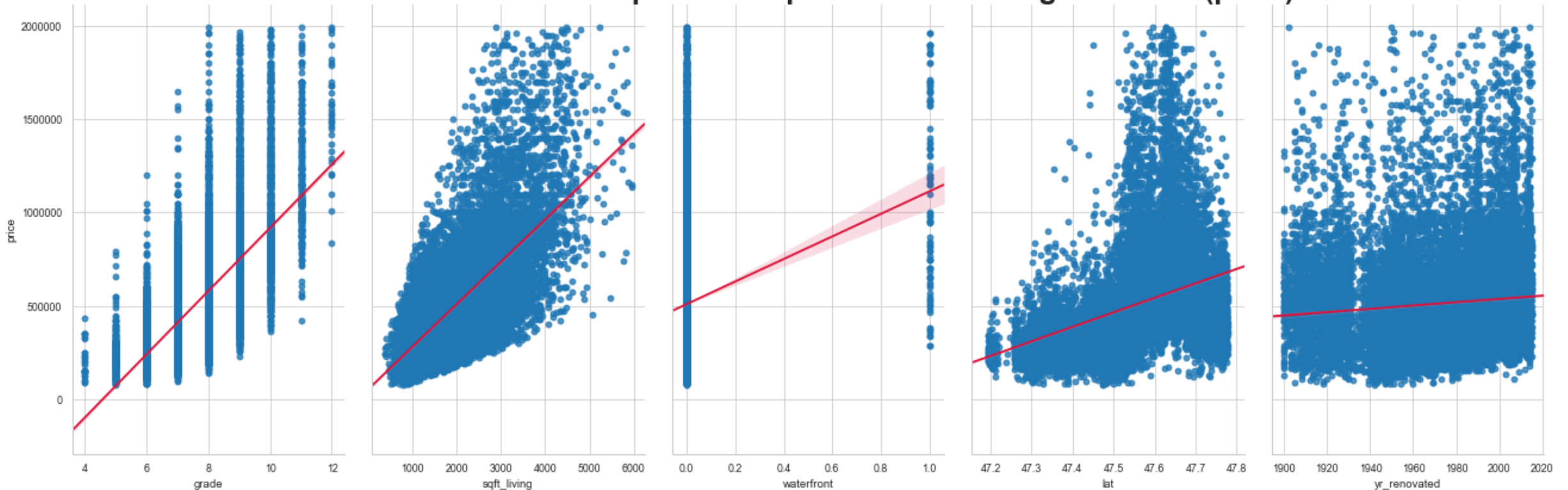
3.WATERFRONT

As shown in the KDE plots, the distribution of price is completely different in waterfront and non-waterfront view. the mean of the price is higher for the waterfront properties.

only small percentage (less than 1%) of housing options have waterfront view.



Pair Plots of Relationship between predictors and target variable(price)

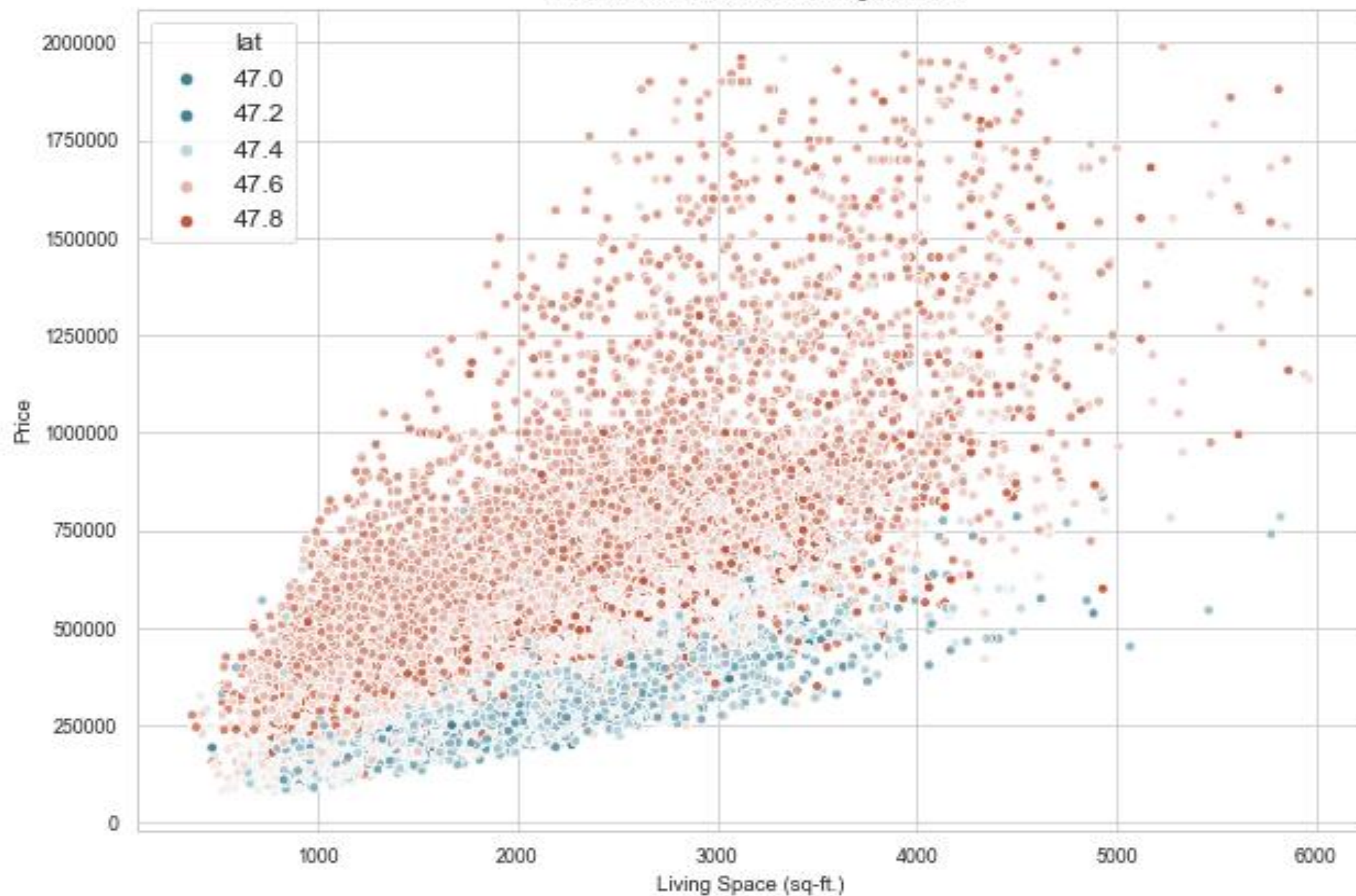


MOST IMPORTANT FACTORS AFFECT THE HOUSING PRICE:

1. **GRADE:** overall grade given to the housing unit, based on King County grading system has the most effect on the price
2. **SQFT_LIVING:** square footage of the home is the second most effective factor
3. **LOCATION:** the latitude coordinate of the house is the third most effective factor
4. **WATERFRONT:** having a waterfront view is the forth important factor.
5. **RENOVATION:** Year when house was renovated is our last factor in the model.t the housing price



Price vs Home Size, color showing Location



RECOMMENDATIONS

IN THIS PROJECT A MODEL HAS BEEN CREATED TO EXPLORE THE MOST IMPORTANT FACTORS IN HOUSING MARKET AND PREDICT THE PRICES:

PRICE ESTIMATION EQUATION

COEFFICIENTS GIVE US THE STATISTICAL RELATIONSHIP BETWEEN VARIABLES

$$\text{Estimated_price} = -0.067 + 0.5 \text{ Grade} + 0.39 \text{ Sqft_living} + 0.25 \text{ Lat} + 0.2 \text{ Waterfront} - 0.1 \text{ Yr_renovated}$$



CONCLUSION

The dataset deals with the real world data, there are always changes, updates, additions and errors in data and their collections, so there will always be some degree of error in any model, for instance, data and likely to have a multiplicity of additional factors that are effecting housing prices which are missing from the data collected.

It is the matter of getting to the best possible approximate with the available data.



FURTHER WORK

If there were more data (neighborhood crime rates, proximity/quality of schools, accessibility/quality of public transportation, etc), we could likely improve our model considerably.

Also, Perhaps a more complicated non-linear regression (polynomials) might make a better model, so we will continue to update and gather more data and explore different versions of models to keep improve our results.



THANK YOU

FARNAZ GOLNAM

