KINGS COUNTY HOUSING PROJECT

PRESENTED BY:
FRED MUTUMA

SUMMARY

I used regression modeling to analyze house sales in a northwestern county.The seller can use the model to predict the selling price of their house and if they need to do any renovation before selling their home.
The buyer can have some suggestions about which kind of house they can afford based on their budget.

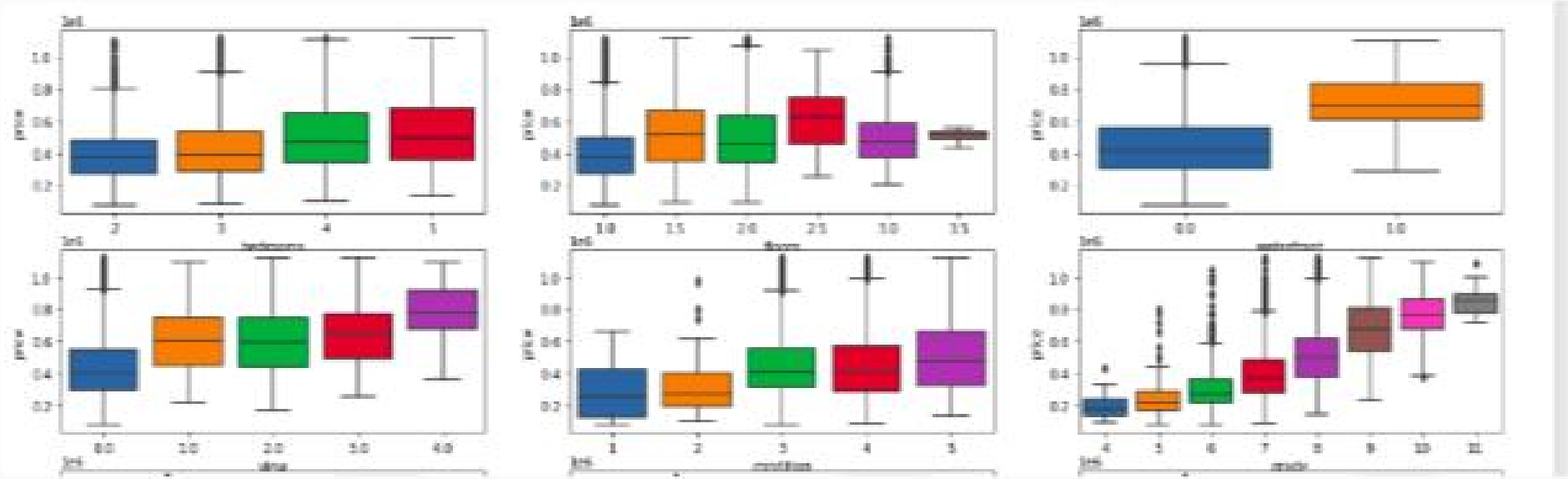
OBJECTIVES

1. Analyse and polish the data which have no meaning or is null to the price.
2. Check if there are some high correlated features in which some of them can be removed.
3. Build the linear regression model.
4. Check how the features can contribute to the house change.

THE METHOD

For this project, I used linear regression combined with few plots to validate my regression model

Distribution of Prices Across Different Features



CONCLUSION #1

The number of bedrooms had negative correlation to the house price. More bathrooms, floors, views and conditions will increase the house price in general. Grades increase the house price a lot each level though still negative. To the zip-code, the house in some area is much higher than others.

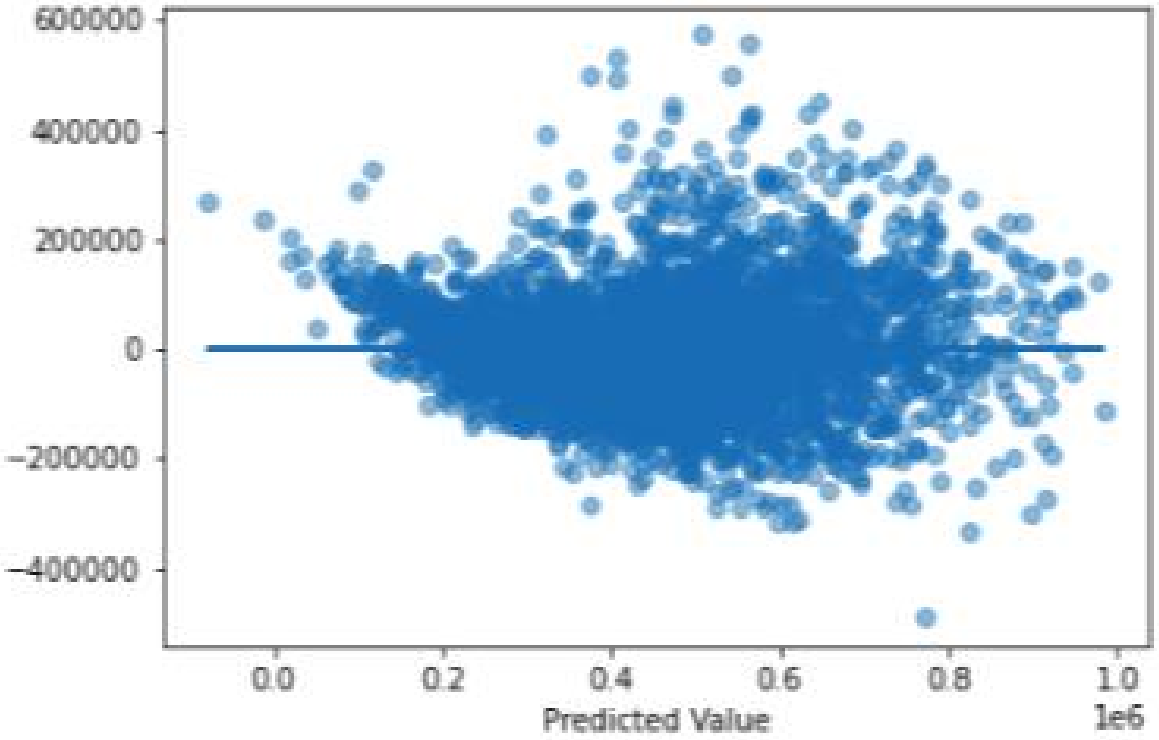
The Coefficient of all the features show how each of the feature affect the house price. Briefly, for the house size, the sqft_living had value 116.5640 which suggests that increasing 1 sqrt of living area, the house pirce will increase 116 dollars. However, the sqft_lot and sqft_basement had negtive correlation to the house price even though the correlation value is very low compare to sqft_living.

CONCLUSION #3

CONCLUSION #2

The house price in month February to July is obviously higher than other months. If there is a waterfront, the house price will increase by 161000. If the house is renovated, the house price can increasing around 72000. If the renovated is within 10 years, the house price can increase around 169000.

qqplot for Validation.



The validation of prediction and real data shows that the prediction price for most house whose price is low (20% - 40% of the max price) is close to the real price. The qqplot shows that the house price is well predicted when the house price is not very high. However, for the high value price house, the prediction is not very accurate. There is a lot of shift of prediction price when the house value increase especially when house price is more than 200000.

Recommendation.

To the buyer, We had our prediction model which can predict the house price and give buyer some suggestion about the price they want. However, the predicted house price is higher than the selling price when the price is over 750000. To the seller, our model give them some suggestion how to increase the potential selling value. For example, they can try to renovate the house and make water front if possible and increas the grade of the house.


THANK YOU