King’s county Housing Analysis

By: Jordan Johnson

Outline of the presentation.

- Business Problem
- Data Used
- Methods
- Results
- Conclusion
- Best month to buy a house.
- Next Step

Business Problem

- A Seattle real estate firm would like to create a tool that allows their clients to enter information about their home and receive a prediction for their home's sale price. They have hired you to produce a predictive model that can predict the sale price of a home as accurately as possible.

Methods

- Cleaned the data
- Logged the date
- Removed outliers
- One Hot Encoded the categorical features.
- Made several models to find a good prediction.
- Looked for features that were most correlated with price
- Used the date to find out when are the best months to buy a house.

Results

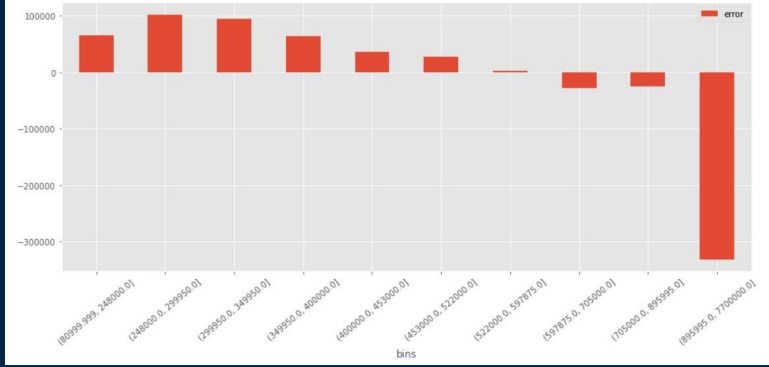
I created nine models with different features and methods in order to get a lower test and train error. Adding more larger and more specific data like the location of the house(zipcode), mad the model fit better.

```
model_1_train = X_train[['sqft_living','grade','sqft_above']]

model_1_test = X_test[['sqft_living','grade','sqft_above']]

model(model_1_train, model_1_test, y_train, y_test, history)
```

```
Train error: 248832.0080114243
Test error: 243814.33811150488
```

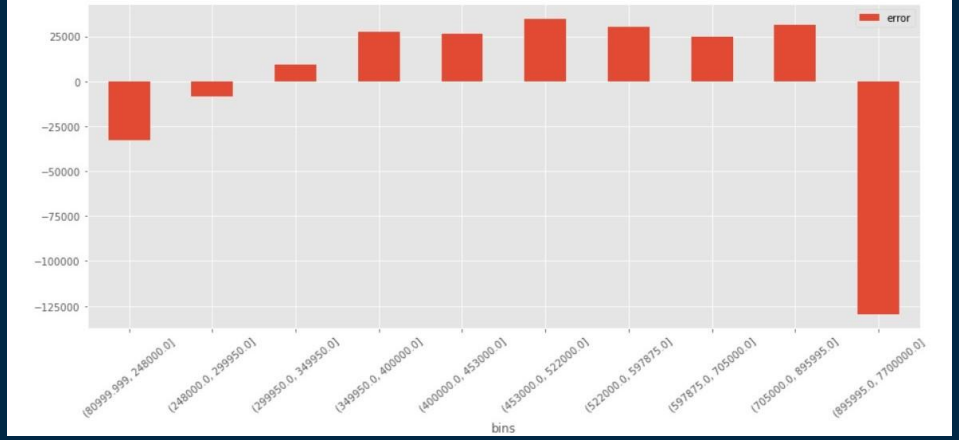



```
X = df[['sqft_living', 'bedrooms', 'grade', 'zipcode','floor_bath','view','waterfront']]
y = df['price']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.3, random_state=2021)
```

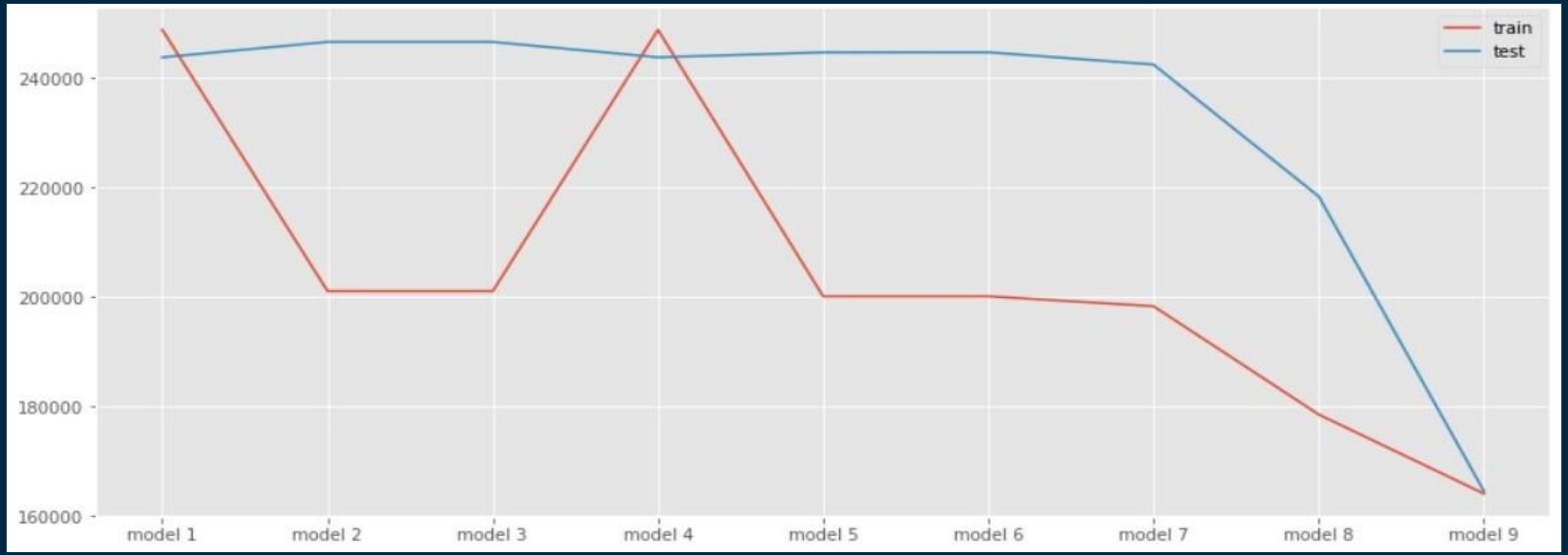
```
numeric = ['sqft_living','bedrooms','grade','floor_bath']
cat= ['zipcode','view','waterfront']
```

```
Train error: 164055.86123491873
Test error: 164525.45580753023
```



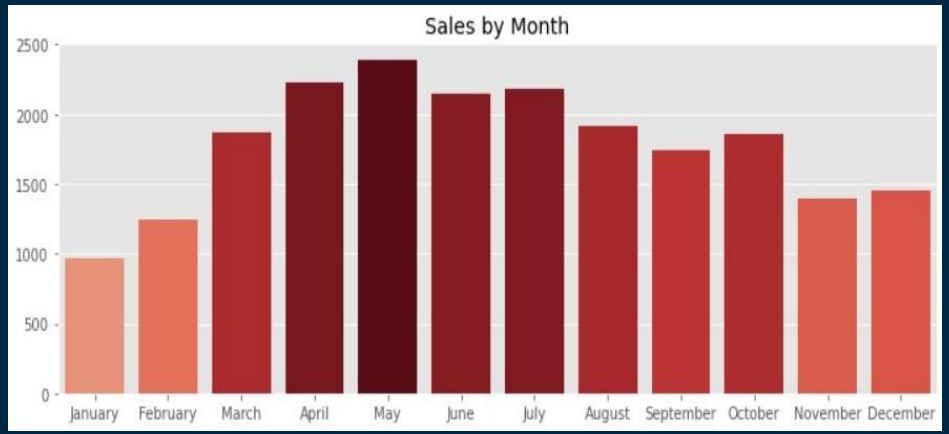
Results Cont.

When making models 1-8 i tried to add more features, drop outliers, and log the data. The ninth model worked the best when I One hot encoded the data and added the categorical values. The final train error was :164055.86123491873 and
the final test error was 164525.45580753023



Best month to buy a house.

The best times to buy a house would be in April, May, and July. So the best time to buy a house is in the summer month.

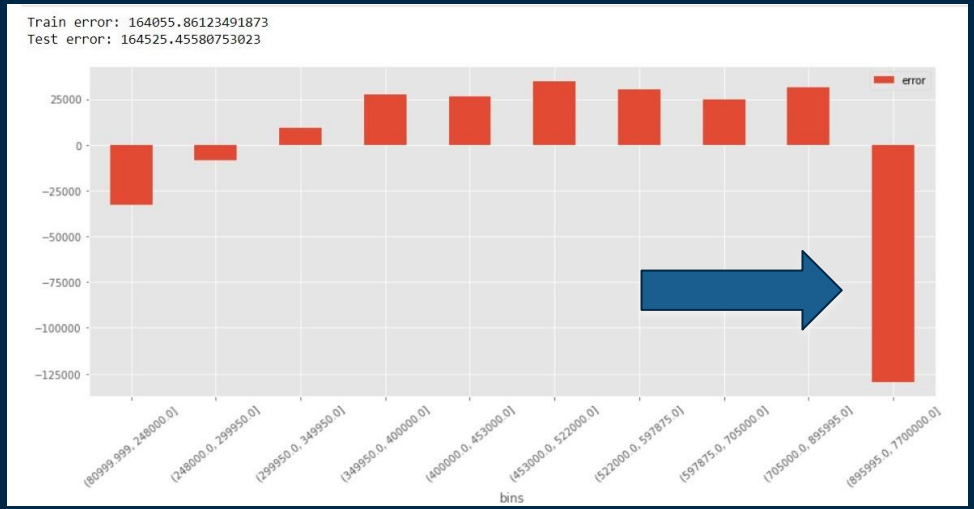


Conclusions

Based on the data, the categorical features were important when fitting this model. Zip Code, waterfront, and the view were features that changed the model and caused it to fit well. One hot encoding, logging, and removing outliers helped the most when trying to get the closest prediction. It would be good to buy a house in the summer months, with a good view, waterfront, and a good sized house.

Next Step

- Try to use all the available data to make models with. Like house renovations.
- Try to lower this error.



Thank You