KING COUNTY REAL ESTATE ANALYSIS

NAGIB GONZALEZ

overview

Nimble Home Buyers is a real estate investment company that wants to expand its business into the state of Washington. However, they are unsure of the factors that drive home prices in the area and how to optimize their investments for maximum returns. As a data scientist, your task is to analyze housing data from King County to identify the key factors that impact home prices and provide recommendations for how Nimble Home Buyers can use this information to inform their investment strategy in the Washington market

Problem

- How can Nimble Home Buyers identify the most profitable properties for investment in the state of Washington?
- What factors have the greatest impact on property values in Washington state?
- How can Nimble Home Buyers optimize the value of their investments through renovation or new construction efforts?
- What strategies can be implemented to ensure that Nimble Home Buyers generate returns for their stakeholders?

Data

King County, Washington home sales information

21380 homes

Dates starting from May 2014- April 2015

22 features

Modeling

- We used the Ordinary Least Squares model (OLS)
- It is a method used in statistics and econometrics to estimate the values of parameters in a linear regression model
- The goal of OLS is to find the line of best fit that minimizes the sum of the squared differences between the predicted values and the actual values
- OLS is used to predict a response variable based on one or more predictor variables

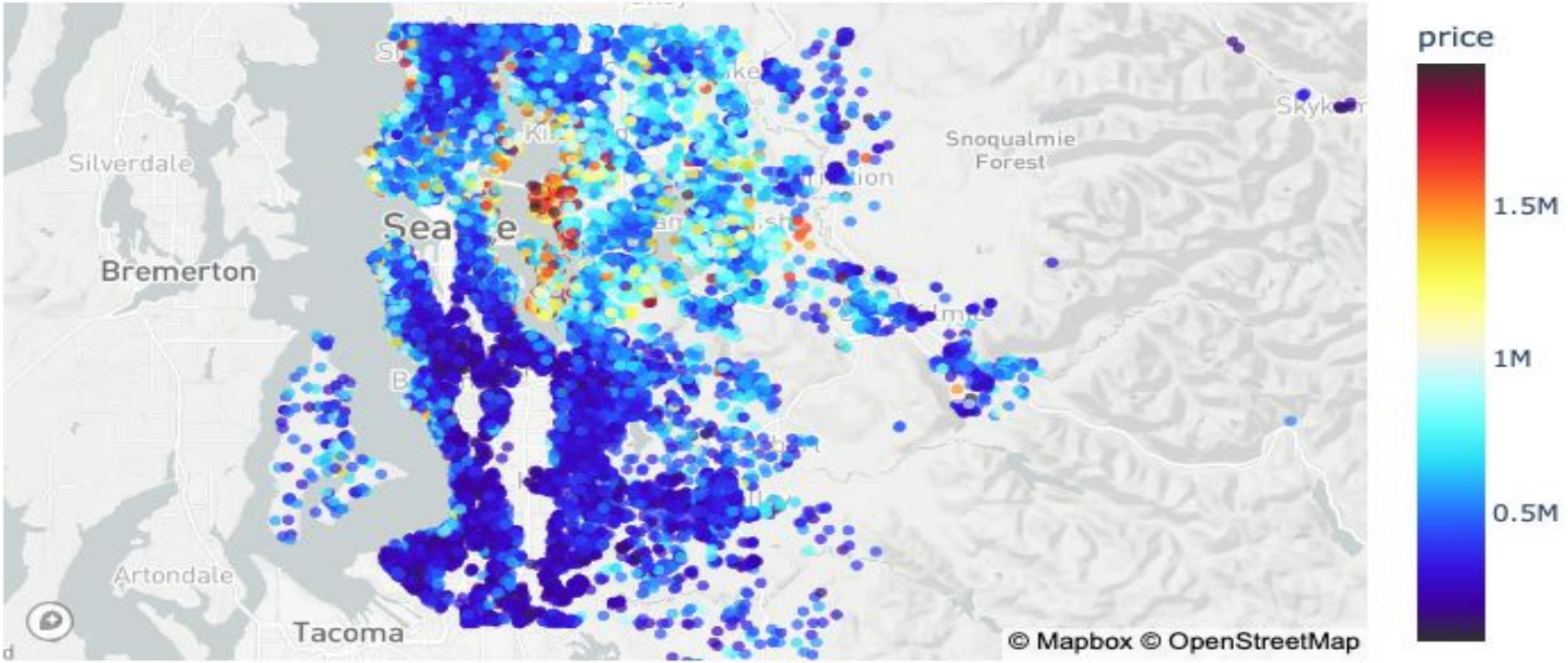
Performance

- Our Model accounted for 83.9% of the variance in price of homes in King County
- The most significant coefficients in this OLS model are for the variables of grade, waterfront, view, condition, and sqft_living.
- These coefficients have large values, indicating a strong positive relationship with the price of the property.

Some Observations

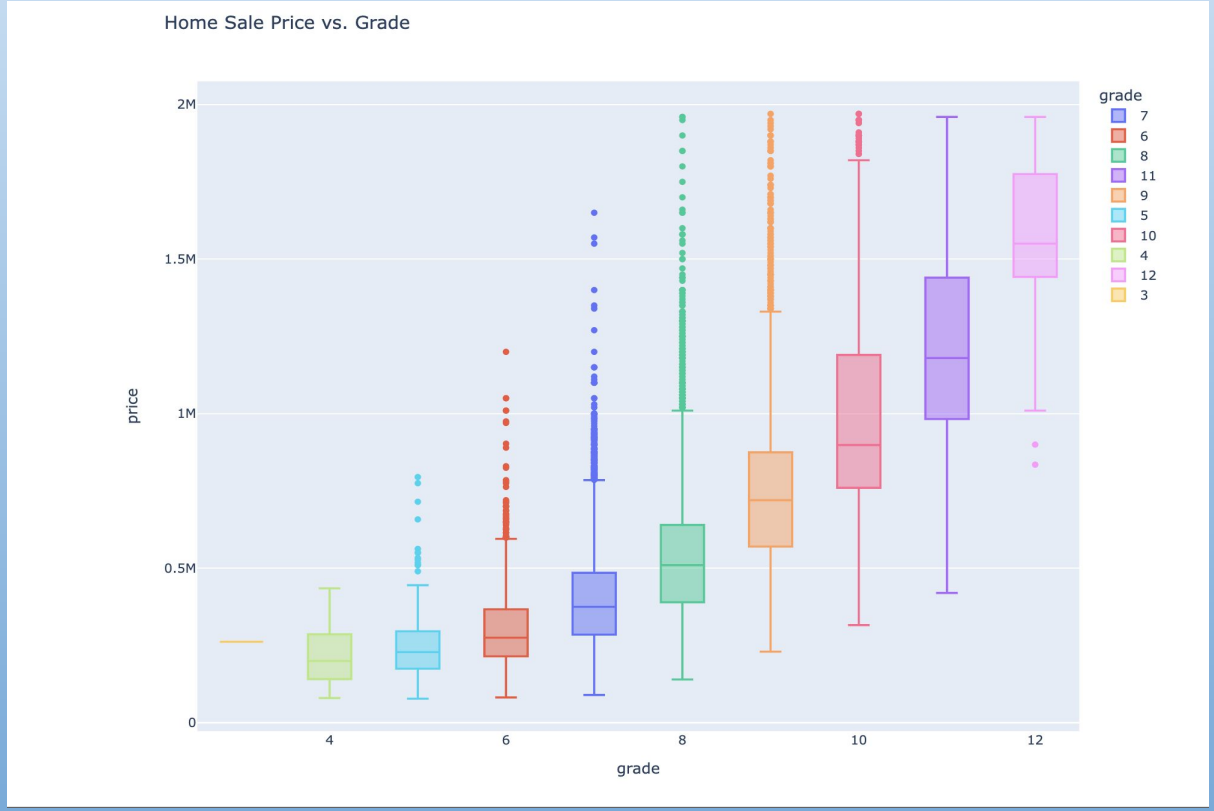
- A higher grade level increased the value of a house by 55,670 dollars per unit grade
- For every 1 sqft increase in living space, the price of the property increases by 62.33 dollars.
- Homes located on a waterfront had an increase in value of 359,800 dollars.
- The condition of the property was found to have a positive impact on its value, with an incremental increase of 26,400
- The condition of the property was found to have a positive impact on its value, with an incremental increase of 26,400

Home Prices Across King County



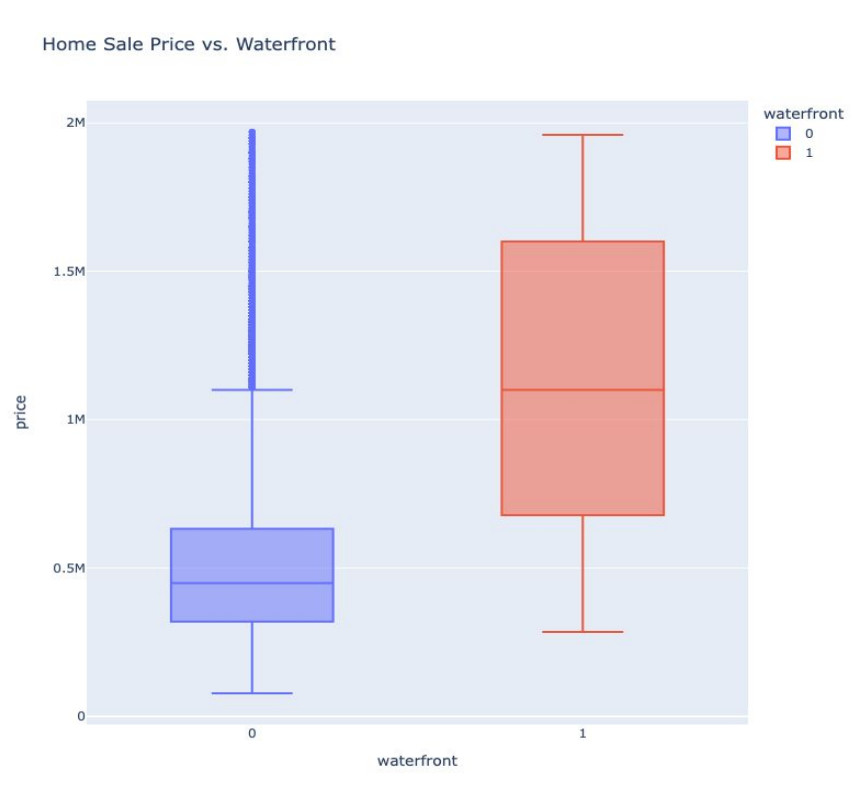
Feature Exploration

Feature Exploration



Grade:
3 =Poor
4 =Low
5 =Fair
6 =Low Average
7= Average
8= Good
9=Better'
10 =Very Good
11 =Excellent
12 =Luxury
13 =Mansion

Feature Exploration



Waterfront:
0 = Not on Waterfront
1 = Is on Waterfront

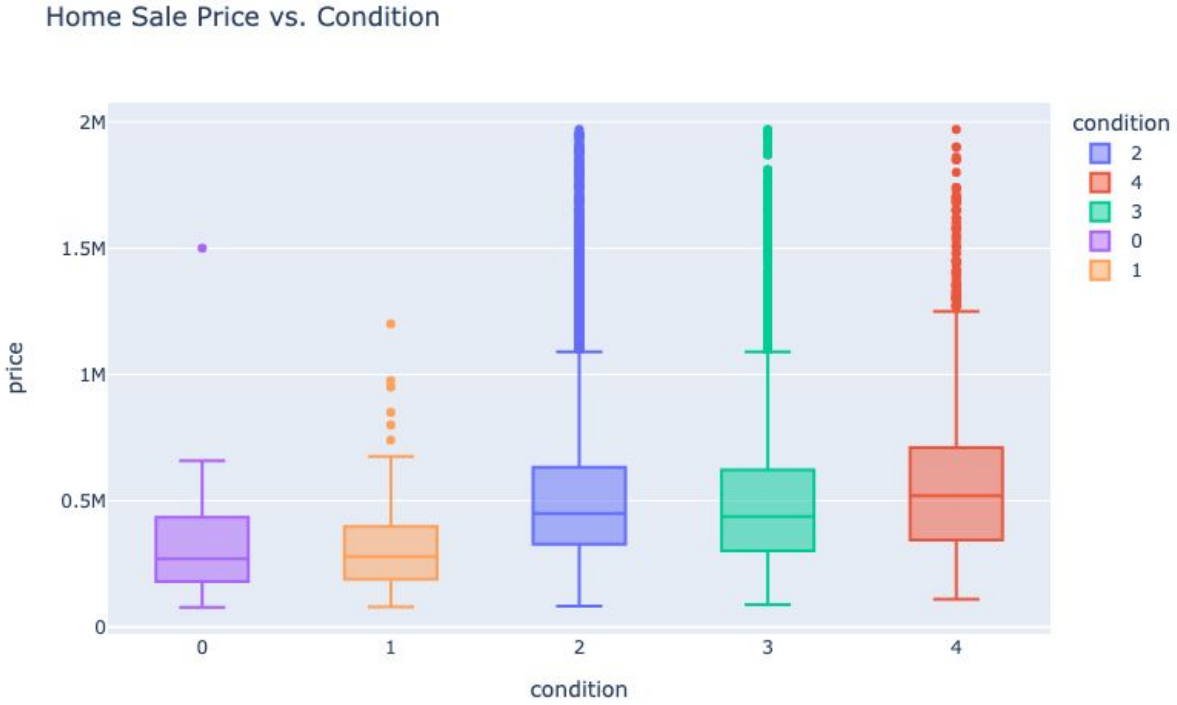
- Median home price without home front was $449,500
- Median Home price with Home front was $1,100,000

Feature Exploration

View:
0 = No view
1 = Fair View
2 = Average View
3 = Good View
4 = Excellent View

- Median home price with no view was $431,500
- Median Home price with good view $775,000
- Median price of Excellent view was $999,999

Feature Exploration



Condition:
0 = Poor
1 = Fair
2 = Average Condition
3 = Good Condition
4 = Very Good Condition

- Poor median was 270k
- Fair median was 279k
- Average median was 450k
- Good median was 438k
- Very Good median was 520k

Recommendations

1. Building grade is a crucial factor when it comes to real estate, and it was found to be the highest positively correlated feature with price in our analysis. Therefore, when renovating and developing homes, it is important to prioritize the use of the best materials. This will increase the chances of the home selling for the highest possible price.

Recommendations

2. Aim for distressed homes in poor condition for renovation and make them good or excellent condition, as this can result in an increase in value from 79,200 to 105,600

Recommendations

3. For new development, aim to acquire homes or lots located on a waterfront with a lower building grade, as they are found to sell for 359,800 dollars more than non-waterfront homes. Aim for luxury or mansion grades in development.

Limitations

- further investigation is needed as the coefficient scores for bedrooms and floors were negative.
- Our final model utilized one-hot encoding for City and zipcode columns, but not for Condition, Grade, and View.
- Implementing one-hot encoding for these variables may improve model performance, but we chose to transform them into numerical for ease of modeling.

Further Analysis

- Incorporate both recent and historical data in future analyses to gain a more comprehensive understanding of the real estate market.
- Consider the cyclical nature of the industry and examine the correlation with the date of sale.
- Analyze the impact of seasonality on real estate trends, as home sales tend to increase in the summer and decrease in the winter.

Further Analysis

- Consider using alternative data sources such as demographic data, social media data, and satellite data to gain valuable insights into the performance and potential of real estate investments.
- Consider using alternative modeling methods such as decision trees, random forests, gradient boosting, support vector regression, and neural networks to improve the accuracy and robustness of real estate data analysis.

Thank You

Questions?

Please feel free to reach out at:
nagib@nimblehomes.us

For a more in-depth and interactive analysis, as well as a closer look at the methods used in this presentation, please visit:
https://github.com/inagib21/dsc-phase-2-project-v2-3