# Reddic Housing LLC

House Price Estimate Report


presented by

Andrew Johnson

Michael Porter Oleson

Dallin Neeley

Daniel Dominguez

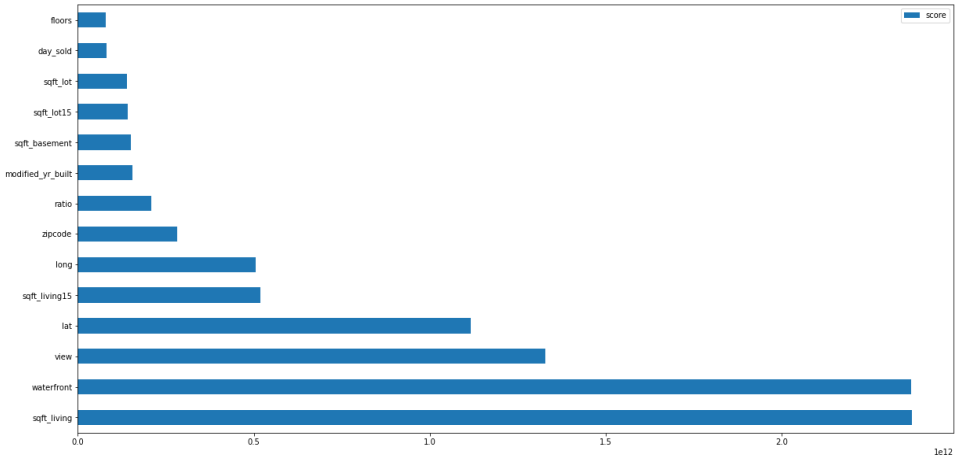
Owen Rowe

# I. Overview

Data Science Team 3 is pleased to present our findings for Reddic Hosuing LLC. We were tasked with creating a machine learning model that will help the company predict housing prices using housing data from the Seattle, Washington area. We found that a regression model would fit our purposes most appropriatly. We are confident that our model will allow the company to predict housing prices within $**110,312** dollars.

# II. Property Features Affecting Price

How are the different features of a property, (the square footage, view, bedrooms, etc.), affecting the price of the house? You are currently able to find that by looking at spreadsheets and by doing manual comparisons between the different features of the properties. Fortunately, we are able to find out this kind of information using our machine learning model. Below is a graph that shows each feature that our machine learning model is looking at and how much of a difference that it makes to our model.




*How much each housing feature gets us closer to the actualprice of the house.*

From this graph you can see some surprising results! According to our model, sqft_living and waterfront are very important features in helping us to be able to figure out the price. It turns out that floors and the day that the house was sold are minimally important to the price, but still make a difference.

## The Difference in Ranges

A lot of the features that we were given with the data have very different ranges. For example, square footage is in the thousands, while there are only four different numbers for view. The way we solved this problem was by normalizing the data. This means that we put each value

between 0 and 1 by dividing a feature by the biggest value that feature has. For example, if we had a house with a square footage of 3,000 sqft and the biggest value we have is 9,000 sqft, then we will divide the 3,000 by 8,000 to get 0.375. This way, our model will treat all features at the same level.

## III. The Accuracy of the Model

We have shown that our model can tell us the most important features are, but is the model acurate? There are multiple ways to look and see just how accurate our model is. The best way that we found is by using something called the “R squared error”. What this “R squared error” does is look at how well our model can explain each housing price. If our model is very far off from guessing the price, the error value will be smaller, but if our model does well at explaining the prices of houses, the number will be bigger. We managed to get a **85%**.

### Insurance and Unsavory Neighborhoods

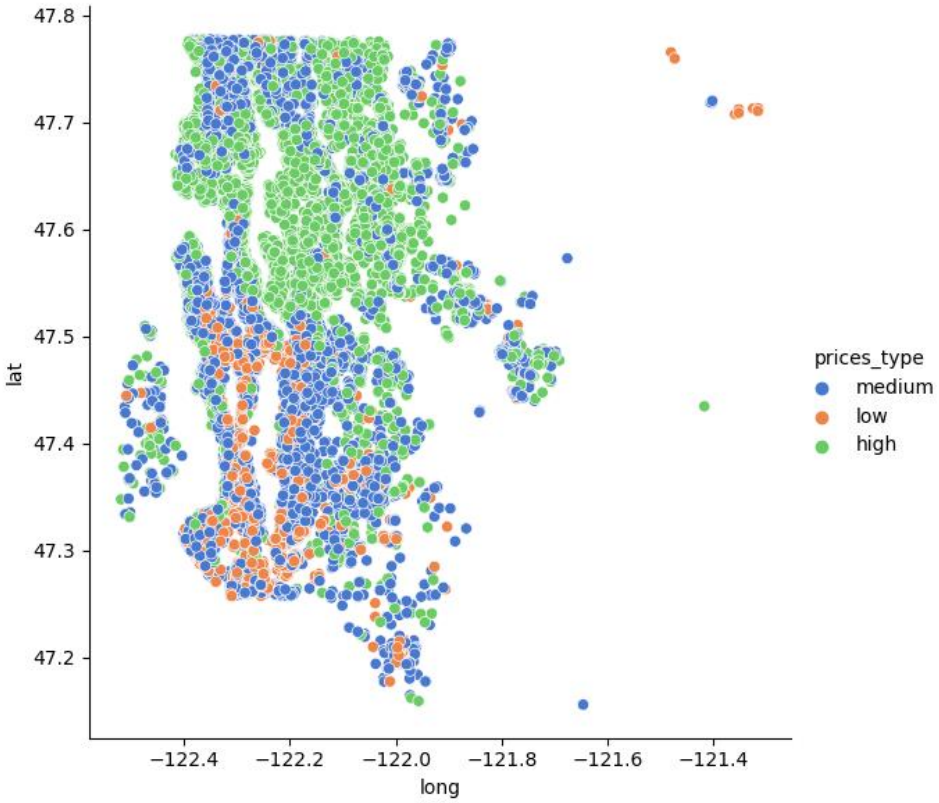
A question came up, while we were trying to find out what all of you wanted from our project, about making sure that homes in unsavory neighborhoods are estimated at a lower price. We believe that doing this would be unethical. If there truly is some correlation between lower prices and unsavory neighborhoods, the model will find it in the data we have been given in regards to location. With both of these ideas in mind, there is no good reason to artificially reduce the prices of certain properties.

During our project, while gathering feedback from everyone, we encountered a query regarding the pricing of homes in less desirable neighborhoods. While it is true that such areas may have a correlation with lower property values, we firmly believe that adjusting prices based on this factor would be unethical. Our data already accounts for location, and any such correlations will naturally emerge from the model. Therefore, we see no justification for artificially decreasing the value of certain properties.

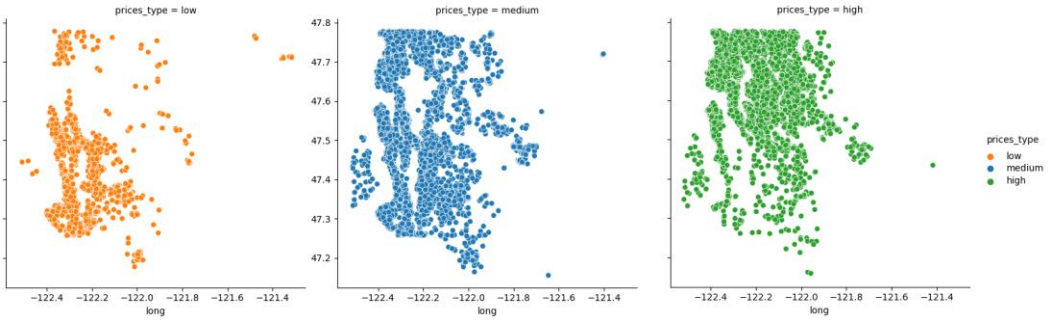
### Location, Location, Location...

While trying to understand the importance of the location we tried to visualize it. First we tried to throw the zipcodes but without seeing something that would let us understand if there was any patterns to understand. Then we use the latitude and longitude and put a gradient in the price. Since that graph was really hard to read we decided to bin the prices into “low”, “medium” and “high”. “Low price” being houses in the range from $1-$250K; “Medium price” houses being in the range of $250K to $500K; and “High price” houses above $500K.

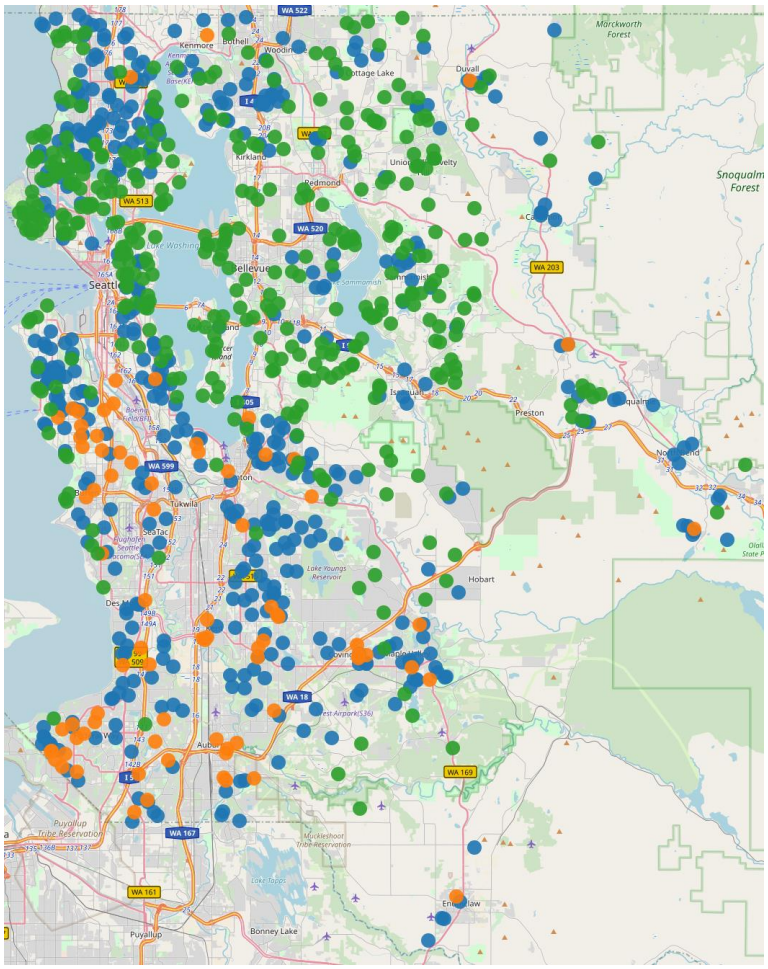
Once we did this we were able to see there was a pattern to understand about the houses and the different prices, but still hard to understand them and see them. Once this is done we can see this a better way if we split it by prices in the visual.




This is how we were able to understand that certain types of houses by their price tend to be located in certain areas and be souranded by similar price houses in the binnin that we did.




Now observing the plot against a map looks like this:




*Now when we plot it against a map the dispersion helps a lot to understand where the different price-based neighborhoods are stablished. So Location is very important to help us dictate the price that a certain house will receive according to de location.*

### Conclusion.

In conclusion, our team has successfully developed a machine learning model for Reddic Housing LLC that predicts housing prices with an accuracy of 85%, using various property features such as square footage, waterfront, view and latitude being the top ones. Through the use of normalization techniques, the model was able to treat all features equally, regardless of their range within the model. We believe that our model will greatly benefit Reddic Housing LLC by providing them with a tool that can accurately predict housing prices and enable them to make informed decisions about their business operations.

## IV. Python Notebooks

Below are Github Gist links to the notebooks we used during this case study:

https://colab.research.google.com/drive/1PdKjLCmICBylJ1mIFGObpXsX_GZxMys1?usp=sharing