



Modeling of King County house prices for new television game-show 'King's Contest'

James Pheby

November 29, 2021



Summary

Television production company King's Contest is producing a new game show in which a group of contestants and a team of experts working for the show are both given a lump sum with which to buy a house in King County, WA

The aim of the game is to sell the house on for a profit, and if the contestants make more than the experts, they get to keep the sale proceeds

In order to minimize the chance of this happening, we analyzed house sales in King County between May 2014 and May 2015 to produce a model of how various features of a house are correlated to its market value, helping us identify undervalued property



Main recommendations

- Give the contestants and experts each a \$500,000 lump sum to buy a house
- The experts should concentrate their search on the 10 locations in which the model performs the best
- They should mainly focus on living space and grade of the property and keep an eye out for bargain waterfront homes

Outline



- Business Problem
- Data Collection, Analysis and Modeling
- Results
- Conclusion



Business Problem

- To attract high ratings with an entertaining show that strikes a good balance between difficulty and achievability for the contestants
- This relies upon the expert team winning almost all of the time. We therefore need a good understanding of the features that most drive the county's house prices in order to identify undervalued property
- We also need to make sure we stack the odds in our favour in order not to keep having to pay out the big prize. Reducing uncertainty is key to this



Data and Methods

Data

- Data drawn from dataset of 21,597 house sales in King County between May 2, 2014 and May 27, 2015
- Along with the price, the dataset contains information on 19 features for each house that could potentially affect its value, ranging from location to property size and whether it is by the waterfront

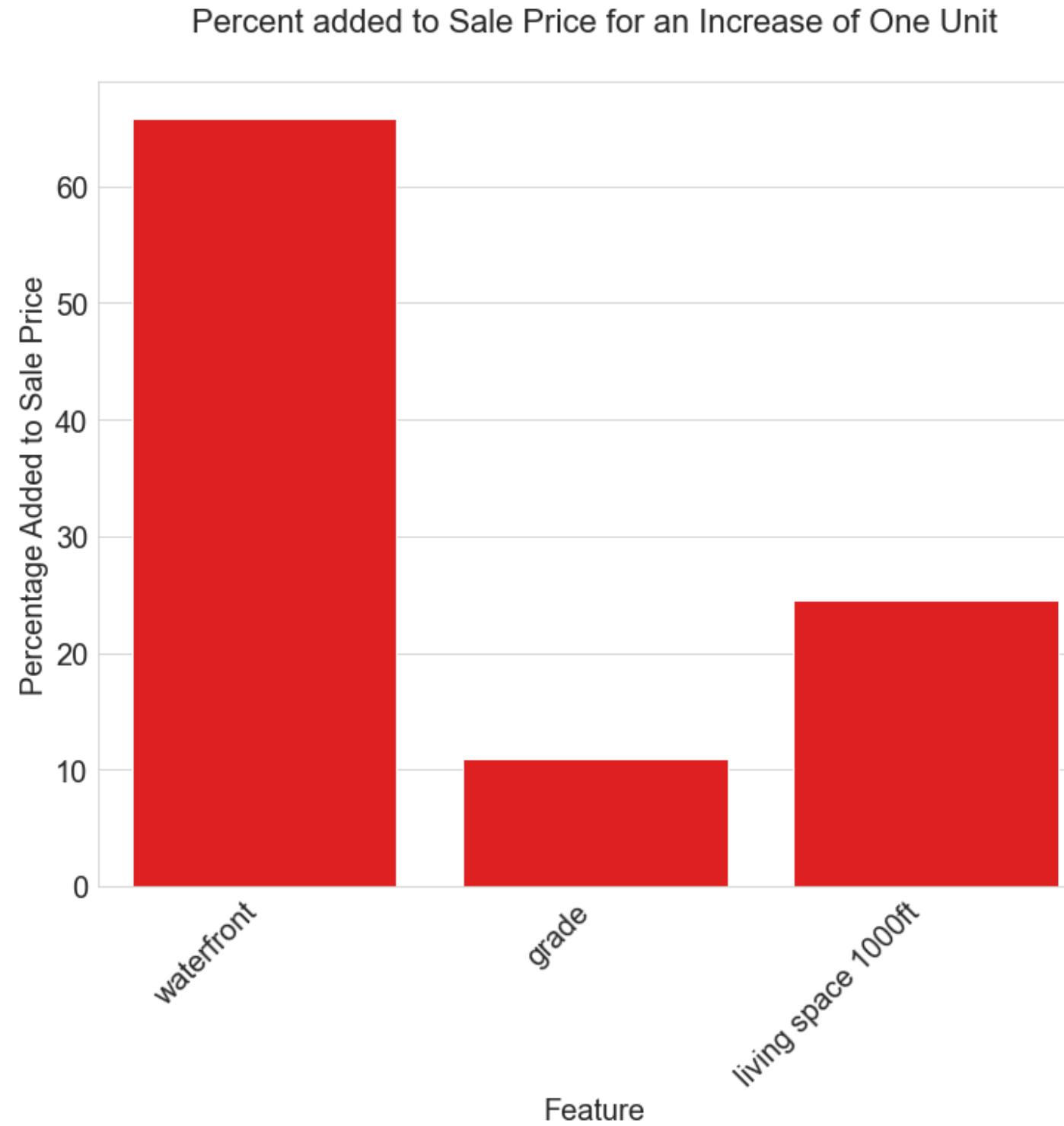


Methods

- Using the entire dataset -- except for a handful of outliers and sales where the house went on to be immediately resold -- each feature was modeled to find the strength of its relationship with the eventual sale price
- Discarding those features that had little effect on the price, a final model was created that includes only the most important features
- The model was then used to infer what percentage rise in house value we would expect to see for a one unit increase in each of the features
- The model, which accounts for around 85% of the variability in house prices, can be used to predict the price of homes newly on the market and help in finding undervalued property

Results

- Being by the water adds more than 60% to the value of a property
- Each increase in the grade of a property adds around 10% to its value
- Every 1000sqft of living space adds almost 25% to the value of a property

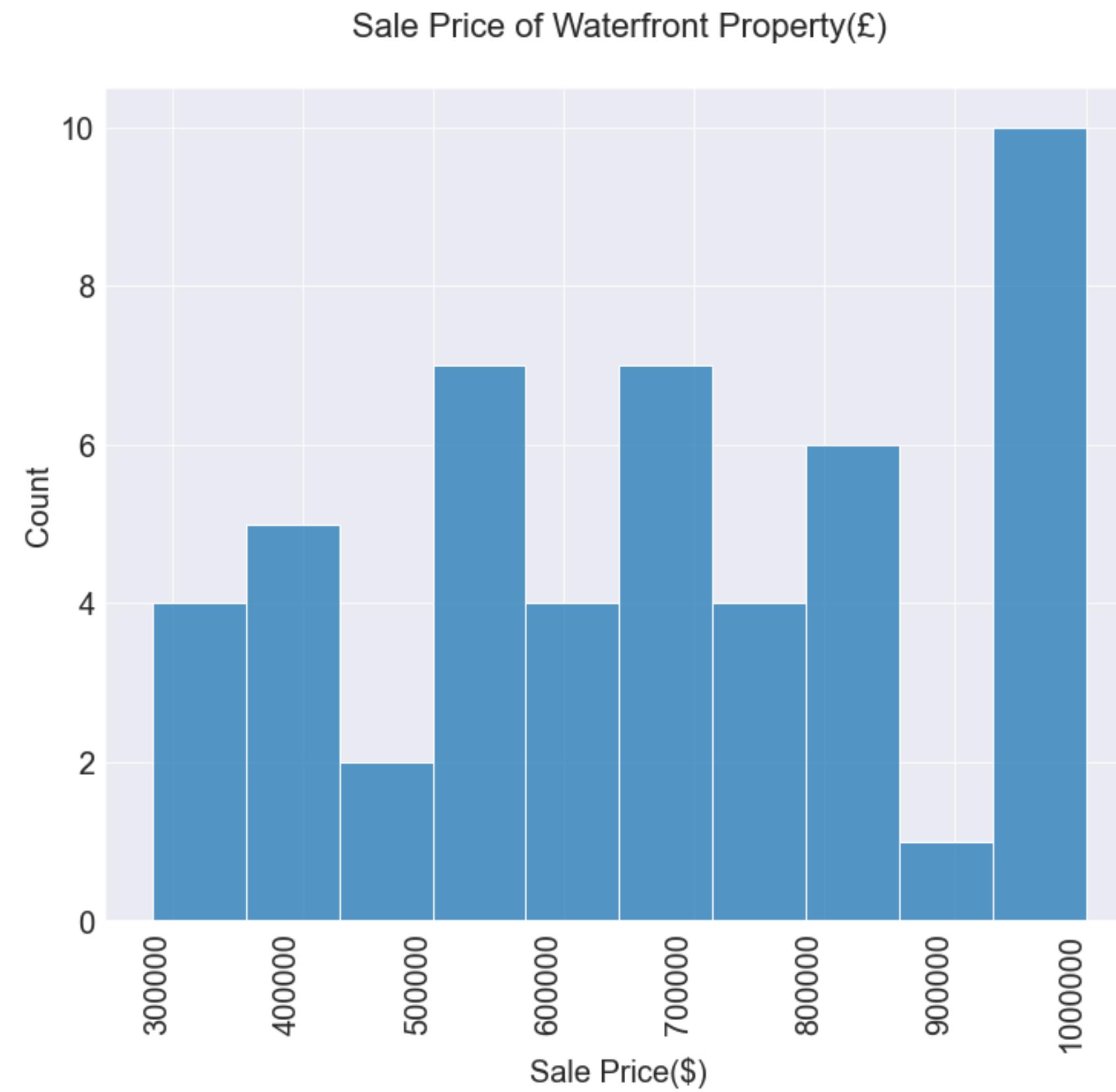




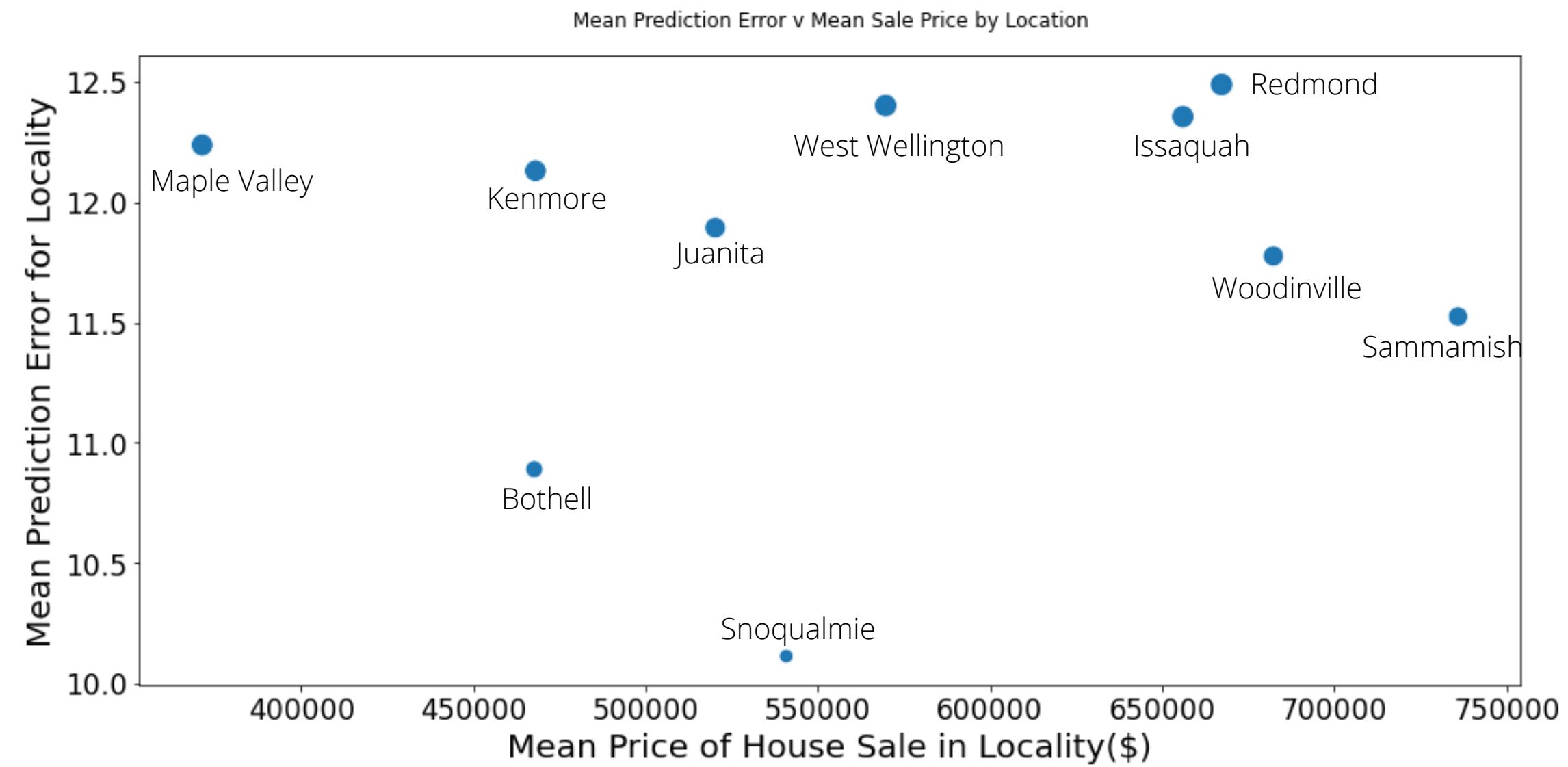
- Medina, Clyde Hill and Madison Valley are the most desirable locations, adding 55%, 45% and 37% to the value of a property
- At the bottom end, property in Federal Way, Auburn and Tukwila suffer a drag on prices of between 61% and 64%



- There is a clear correlation between living space and value, with price 'ceilings' and 'floors' for each property size. Houses near the floor are likely to offer good profit potential, whereas those near the 'ceiling' are probably best avoided
- The orange blobs are waterfront homes, They are almost all concentrated at the top of the price range



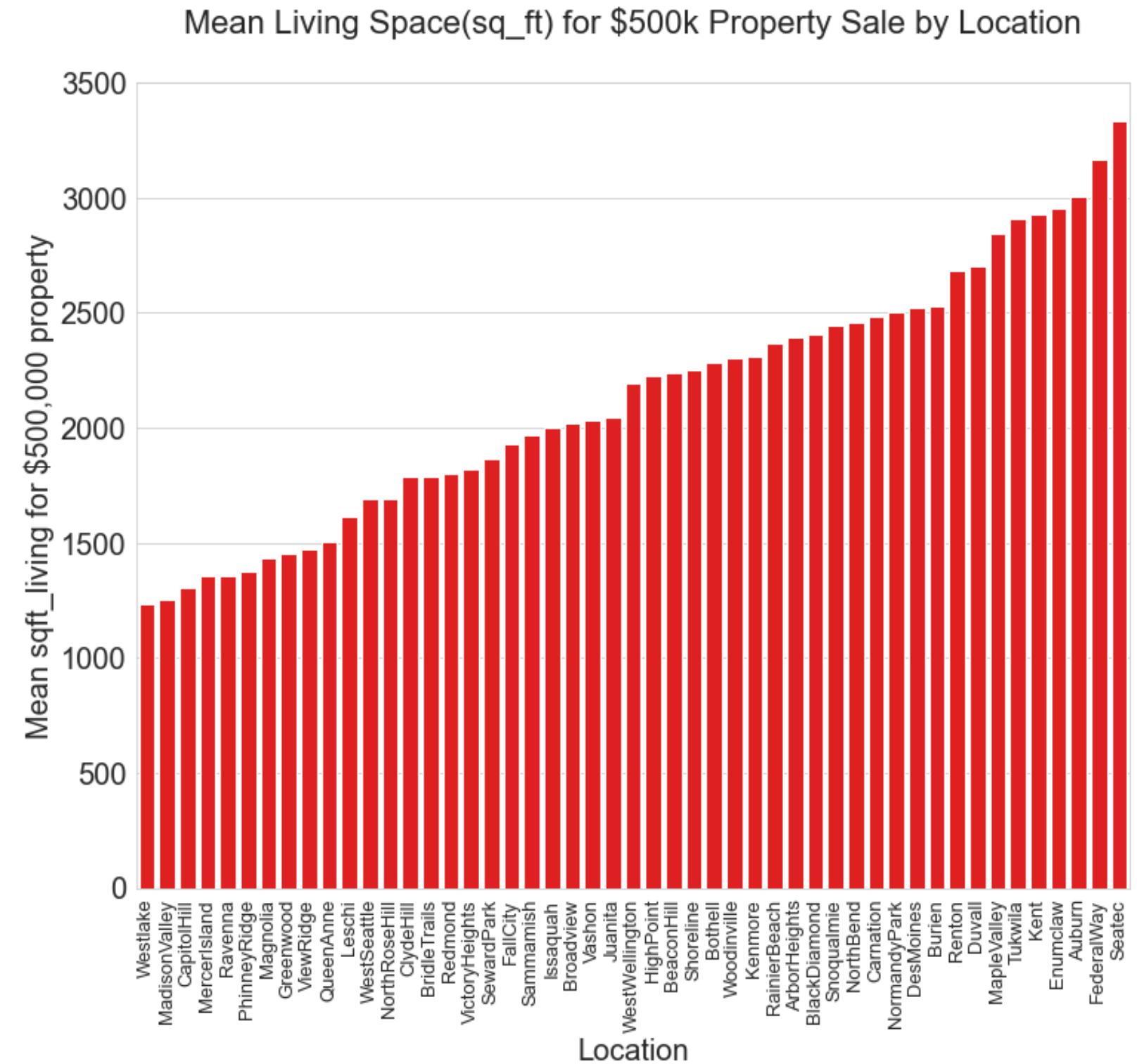
- Although providing the biggest boost to house value, there were only around 20 waterfront houses within our budget sold over the year



- The model performs best in Snoqualmie, with a mean error of around 10%, and the error is less than 13% in each of these top 10 locations
- The average property price in these locations ranges between around \$400,000 and \$750,000

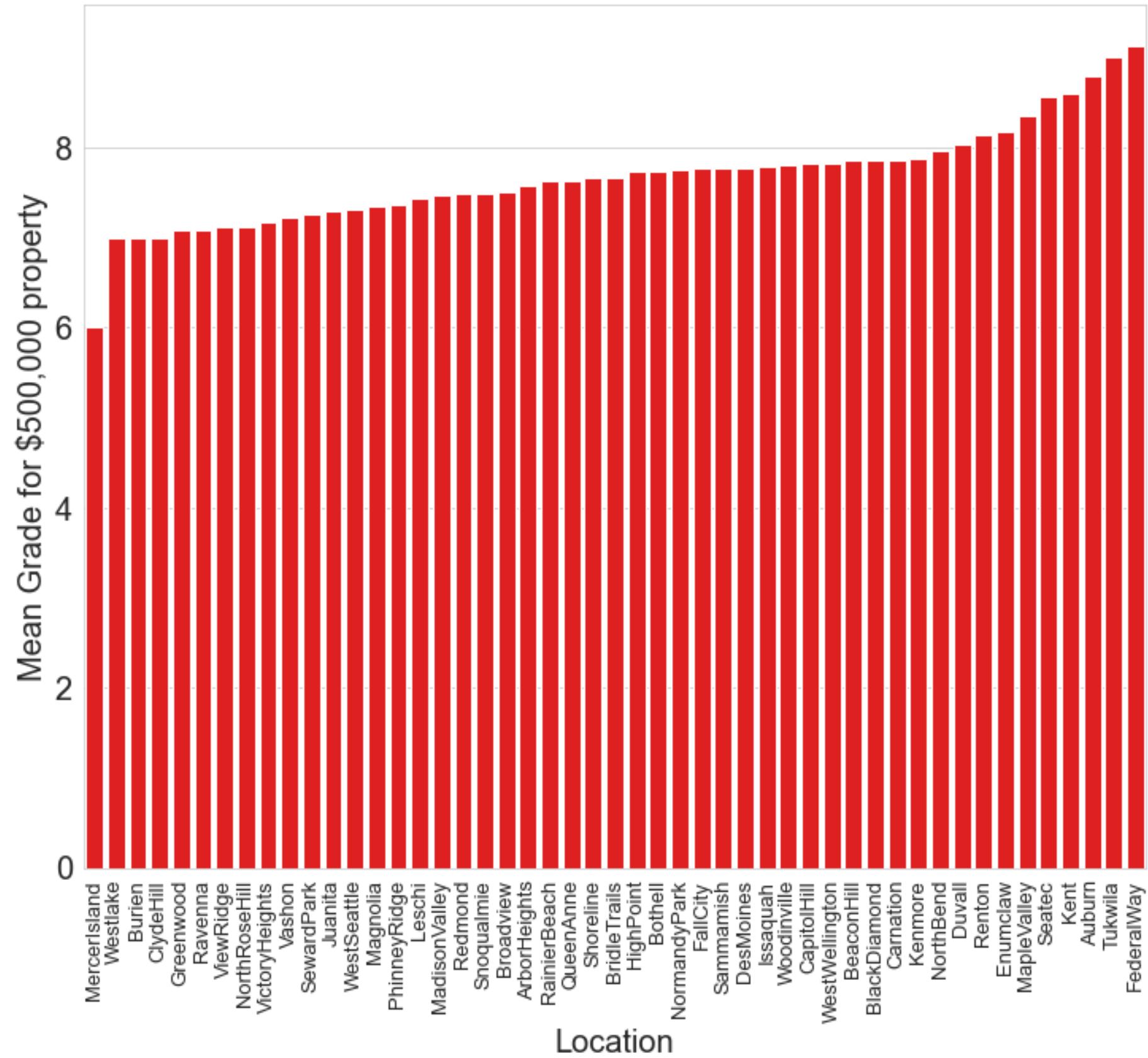


- In the top end locations, a \$500,000 budget will generally get you between 1250 and 1500 sqft of living space.
- In the lower-priced locations, the same budget will usually get you more than 2500 sqft of living space





Mean Grade for \$500k Property Sale by Location



- In the top end locations, a \$500,000 budget will generally get you a property graded around 7
- In the lower-priced locations, the same budget will usually get you a property graded over 8

The Magic formula



$$\exp(P) = \text{base} + (\beta * \text{wfront}) + (\beta * \text{size}) + (\beta * \text{place}) + (\beta * \text{grade}) + (\beta * \text{yard})$$

Don't Worry! It might look complicated, but it's a simple task for a computer. Just tell it the values of the features for any property of interest, and a moment later you have an estimate!

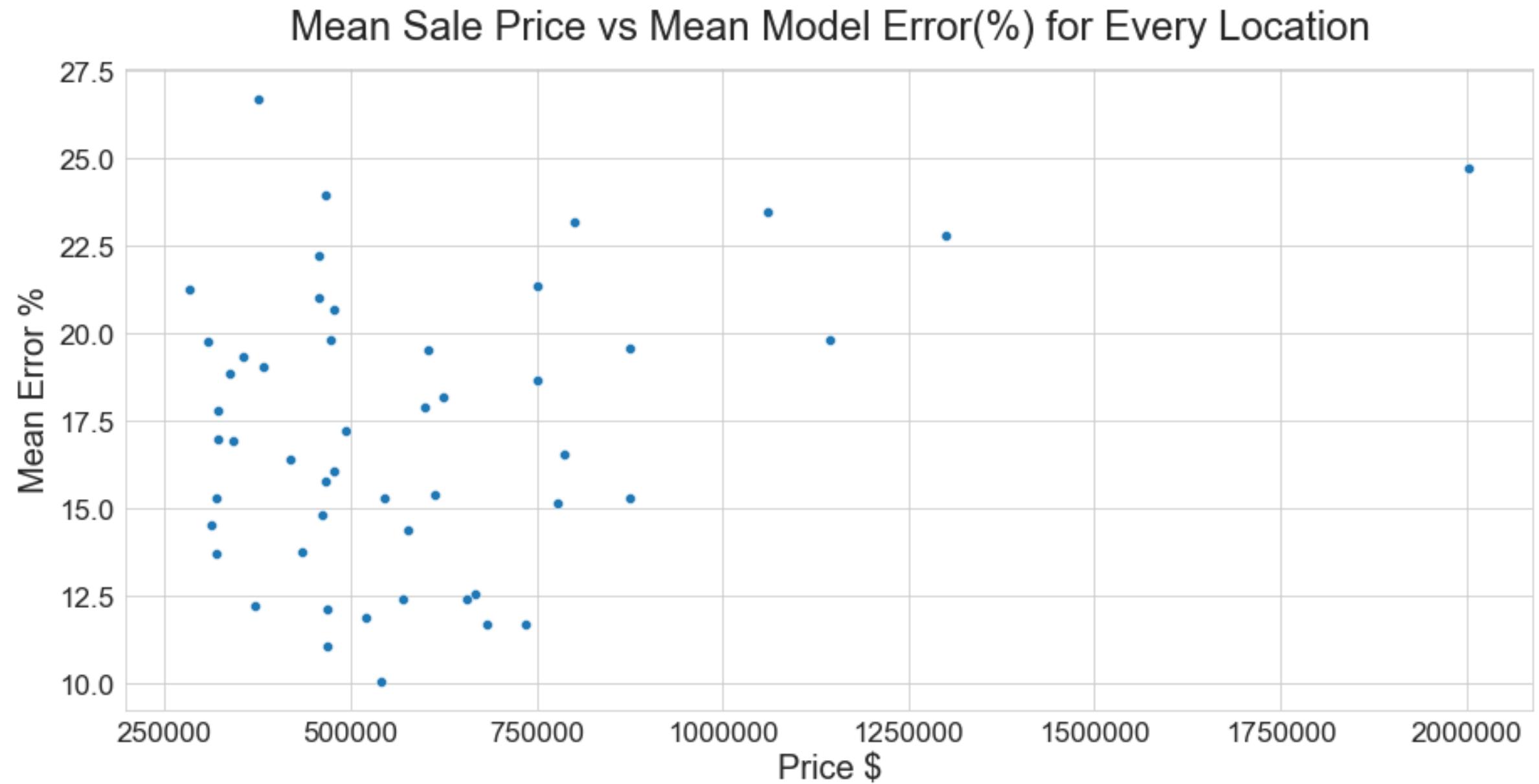


How reliable is the model?

A model is a simplified representation of reality, meaning we lose some precision in the search for clarity

The features in this model account for around 85% of a house price's variation from the mean price, meaning around 15% is down to features not included, or random error

We are over 95% percent confident for each feature that its connection to house price will be seen 'out in the wild', not just in our dataset



- The model has a 10% error rate in its best performing locations, and a 27.5% error rate in its worst performing locations
- Most locations have an error rate of between 12.5% and 20%



Conclusion

- The search should be focussed in the top 10 locations where the model performs the best. This ensures as much control as possible over the result, minimising the chances of the contestants winning the prize
- The contestants and experts should be given a lump sum of \$500,000 with which to buy a house. This is roughly the sale price of an average home in the top 10 neighbourhoods most accurately predicted by the model
- While being by the waterfront adds the most value to a property's value, such homes are very seldom on the market within our budget range. Instead, the experts should focus on location, the size of living space and the grade of a house



Conclusion

- The top 10 locations where the model performs best are: Snoqualmie, Bothell, Sammamish, Woodinville, Juanita, Kenmore, Maple Valley, Issaquah, West Wellington and Redmond
- If we go house-hunting outside these locations, \$500,000 should get you between 1250 and 1500 sqft of living space in high-end areas and more than 2500 sqft in lower-end areas
- In the top end locations, a \$500,000 budget should get you a property graded around 7 and a house graded over 8 in the cheaper areas

We also have a heatmap of house sales in the county weighted by sale price, that can be zoomed in to get a street level detail of house values





Next Steps

- The dataset is somewhat limited, with almost all the features describing immutable characteristics of the house, such as lot space, waterfront, year built
- Access to more comprehensive features, particularly changeable characteristics such as garage, fireplaces, loft conversion, porch, age of bathroom/kitchen units would open the door to a new program in which contestants are given a lump sum with which to renovate a house, and they get to keep whatever extra value is added when sold
- We should also look at characteristics found only in top-end houses for a 'high-rollers' special. Features could include swimming pool, cinema room, stables, jetty (for waterfront homes!), gym, library, terrace



Next Steps

- We need to keep collecting data on latest sale prices as we need to be ahead of the game in discerning shifts in trends, and in knowing which locations are hot, and which are not. This will keep our model updated with the latest information
- Collect data on whether there has been shift in which features add most value to a house in light of the pandemic, for instance...
- Does living space and outdoor space now come at an even higher premium given people's experiences of being locked down?
- Has proximity to urban centres become less desirable now working from home has become more popular?

Email: jpheby93@gmail.com

GitHub: <https://github.com/jimp93>

Medium: <https://medium.com/@jpheby93/>

