



Across the Pond Tech

PREDICTING SALE PRICES FOR LAKE AND COMPANY REALTY

Abass Ibrahim, Nathaniel Martin, Matthew Samson



Summary

Predictive Model

Our predictive model is able to predict housing prices with a typical error of \$160k.

Inferential Model

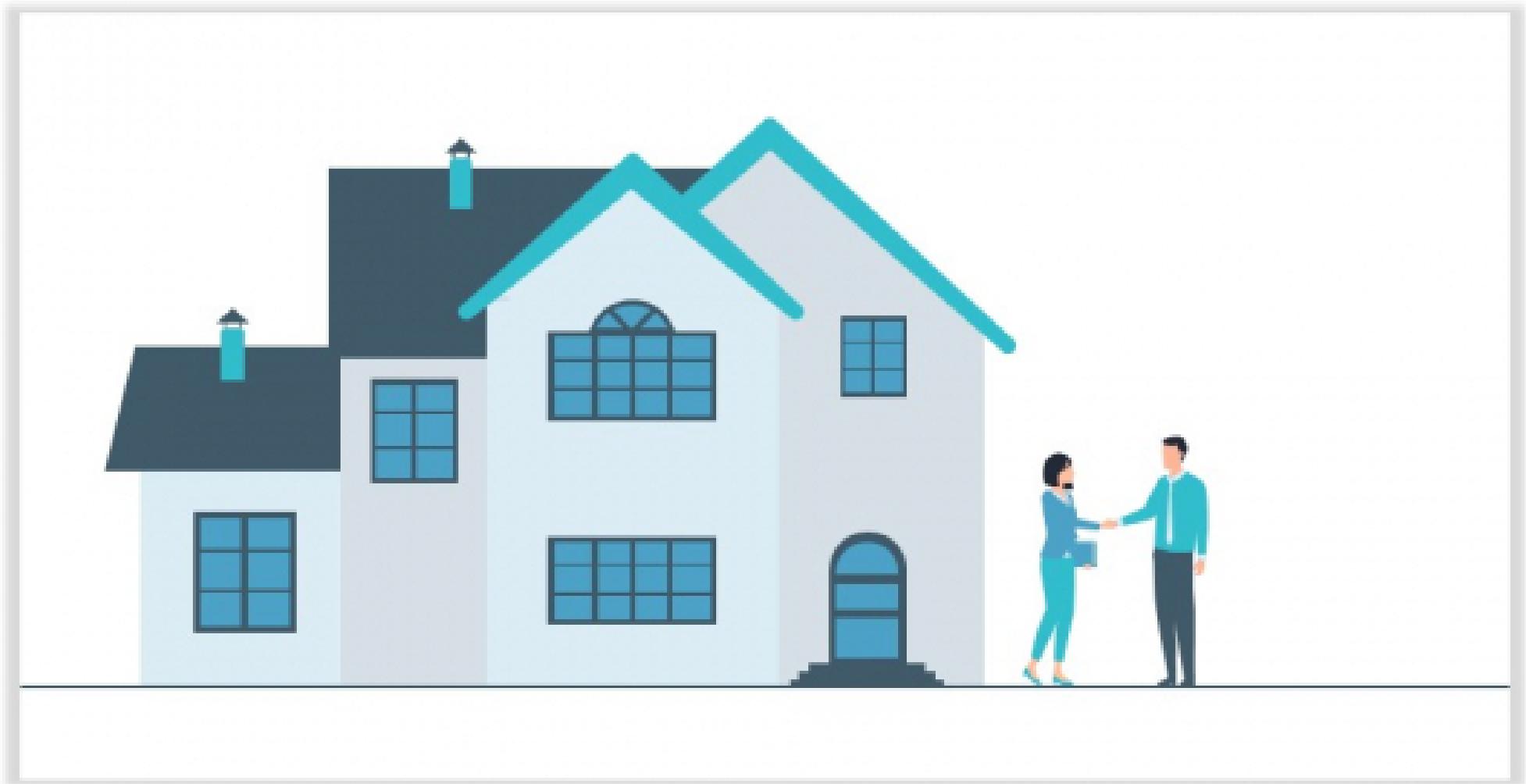
Our inferential model measures the most important features in understanding sale price, particularly distance from city centers, living area, number of bedrooms, and room sizes.

Conclusion

We recommend more data collection and more advanced modelling.

The Business Problem:

- Many different (and some related) features makes knowing what price a house for difficult.
- This makes it difficult to tell which features are the most relevant to the sale price.
- Multiple linear regression can help sort this out.



Action Plan House

Data and Methodology

Data

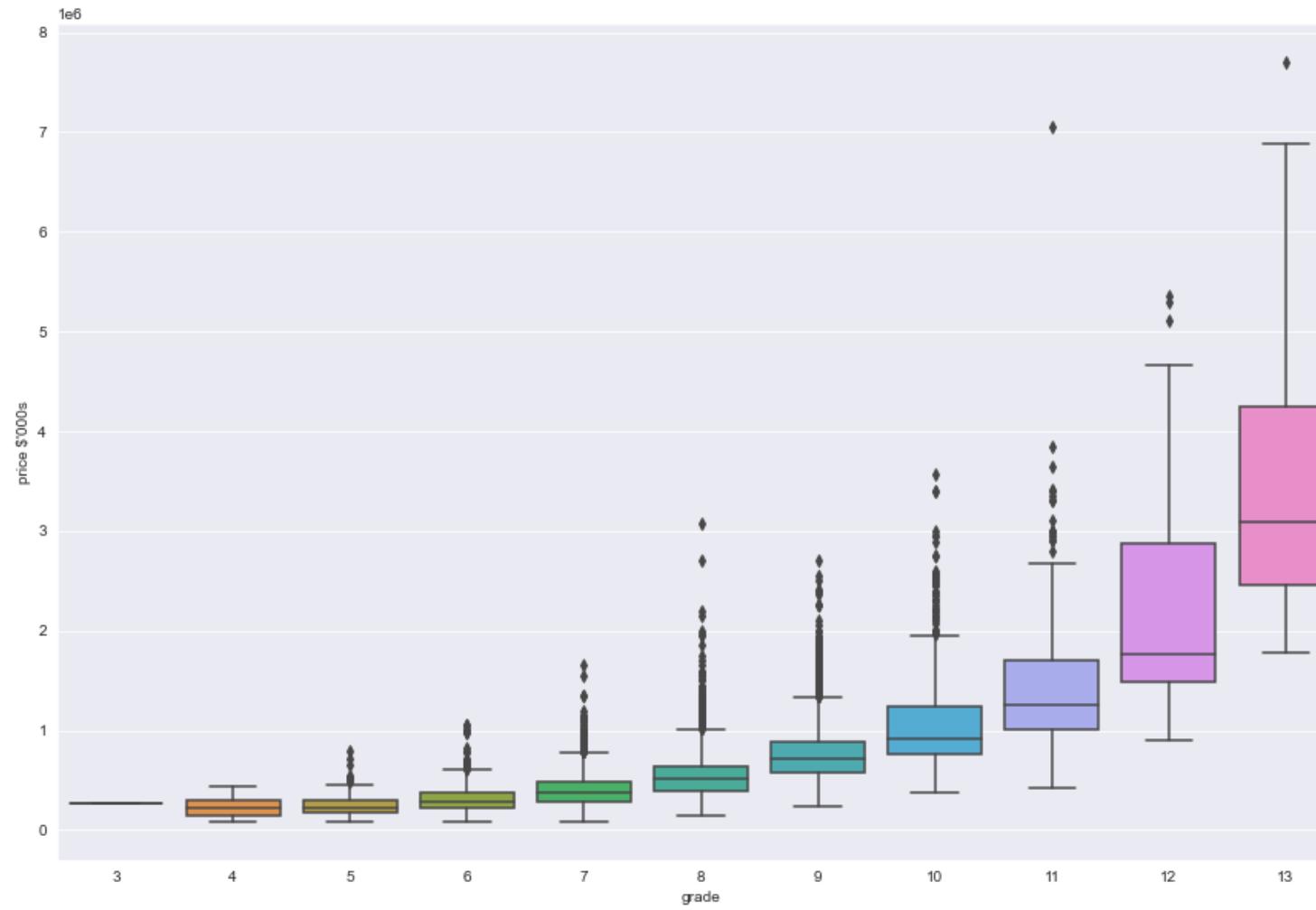
- Dataset from the King County Assessor's public website, which lists house sale prices and other data for over 21,000 houses in the period from May 2014 - May 2015

- Manually scraped additional data from UnitedStatesZipCodes.org

Methodology

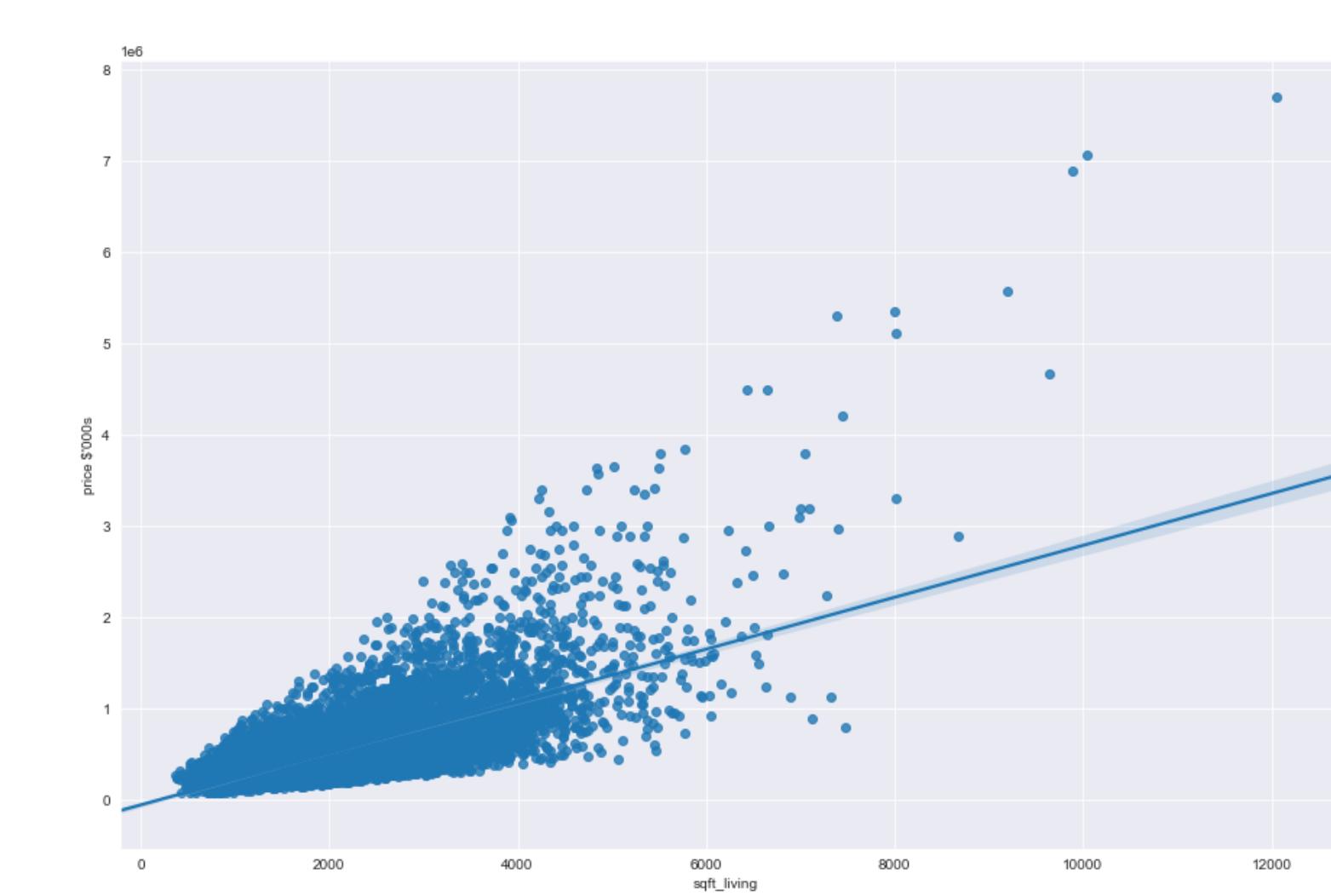
- We built both predictive and inferential models using multiple linear regression. We engineered several additional features using both datasets.

Initial Findings



Price vs Grade

Exponential increase in price



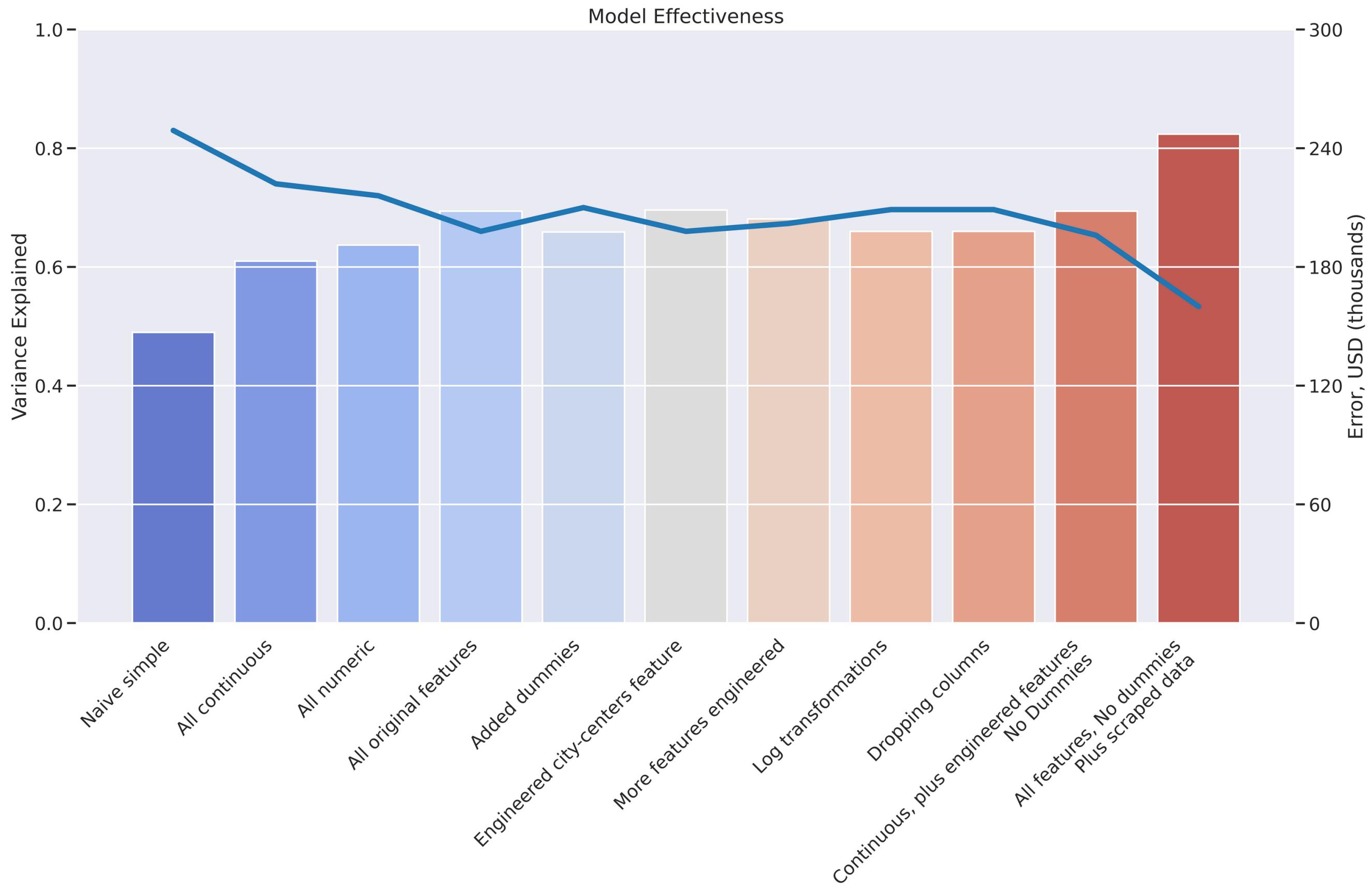
Price vs Square foot living space

Linear increase in price

The progressive improvements of the linear models

Bar graph is explainable variance. Higher is better.

Line graph is Error. Lower is better.



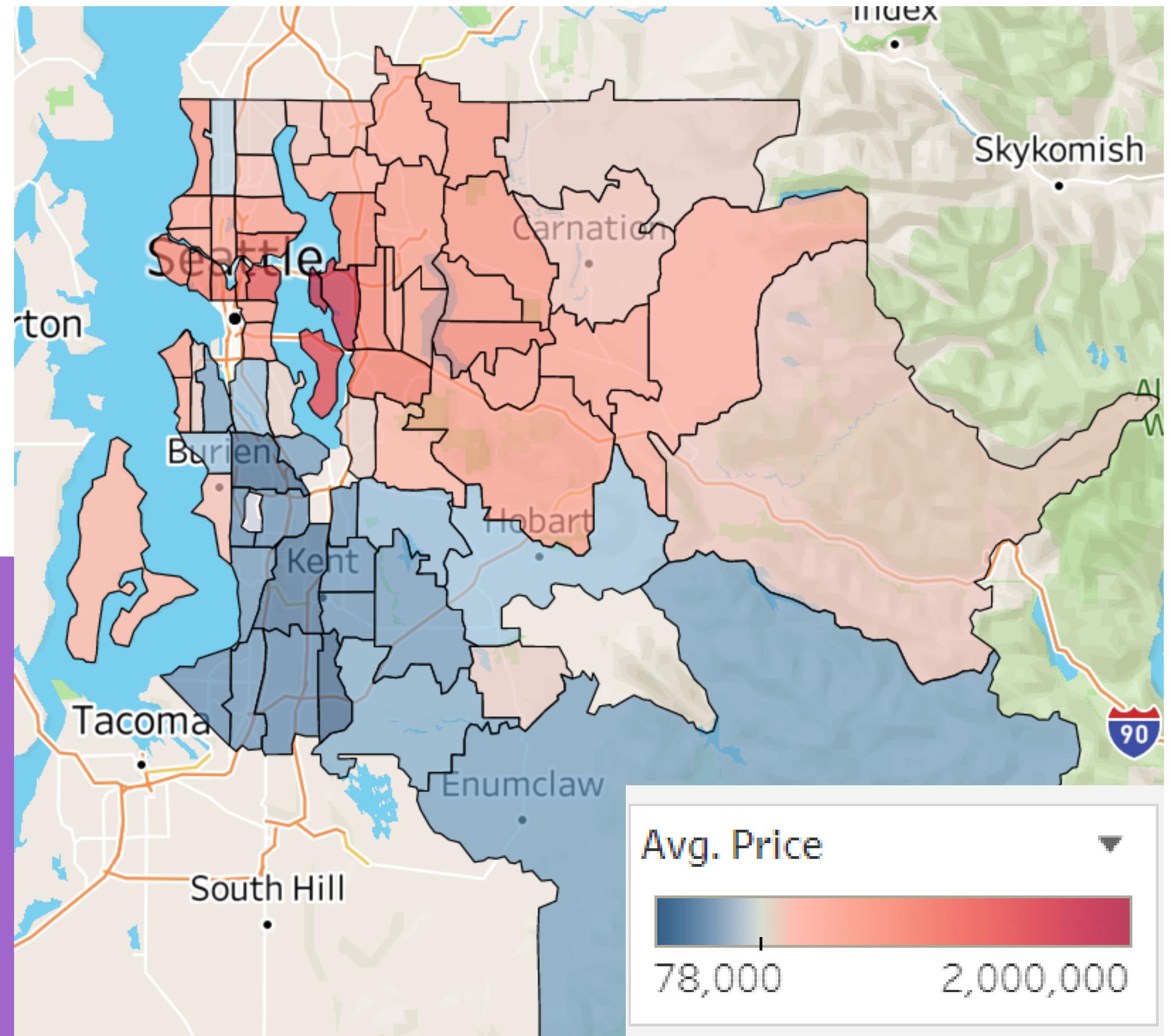
Price Map

Regional House Price

We see a cluster of high prices around Seattle Centre

.....

We also see higher prices north and lower prices south, indicating that latitude is correlated with price.





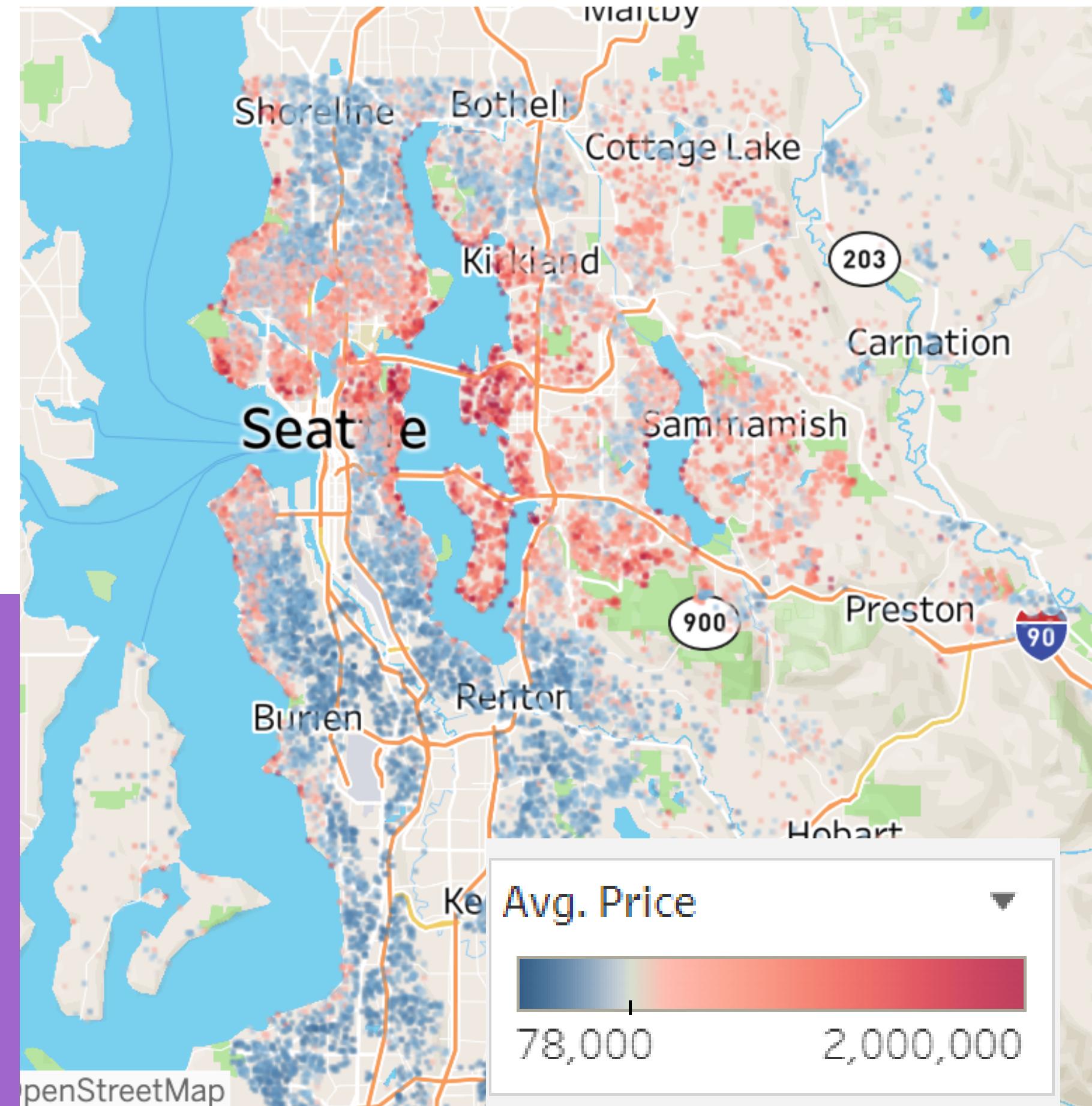
Most Expensive regions

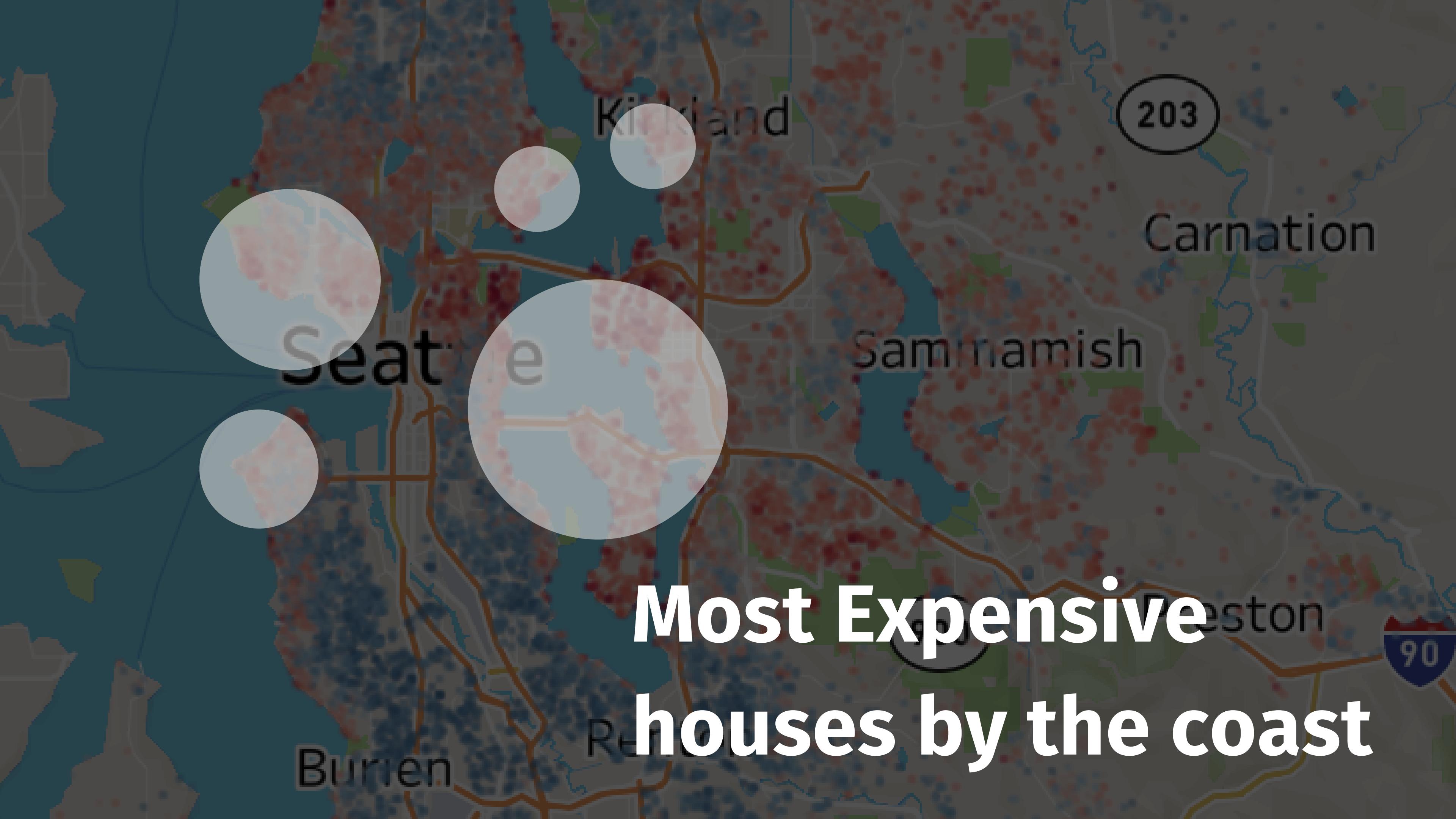
Price Map

Individual House Price

We see a cluster of high prices around Water

Specifically, we see along the coast prices are generally higher. However they still decrease as we go South.





**Most Expensive
houses by the coast**

Predictive Model

82%

82% of the variance of our data from the mean was explained by our model.

The model is a multilinear model which takes into account the majority of our data, both categorical and continuous.

\$160k

The prediction error by the model amounts to \$160k.

We gathered data that accounted for factors such as, area of water in the zipcode, number of schools in the region, average income in the region and population density.

Inferential Model

59%

59% of the variance of our data from the mean was explained by this model.

The model was limited by factors such as correlation between predictors and led us to only pick the 4 most influential on price.

These were the total distance from Seattle and Redmond, living area, number of bedrooms and room sizes.

\$249k

The prediction error of the model was around \$249k.

This rise in error in comparison to the prediction is most likely caused by the lack of predictors.



How much does
a predictor
effect price?
If we raise the
value of the
predictor, how
will price
change?

Inferential model key insights

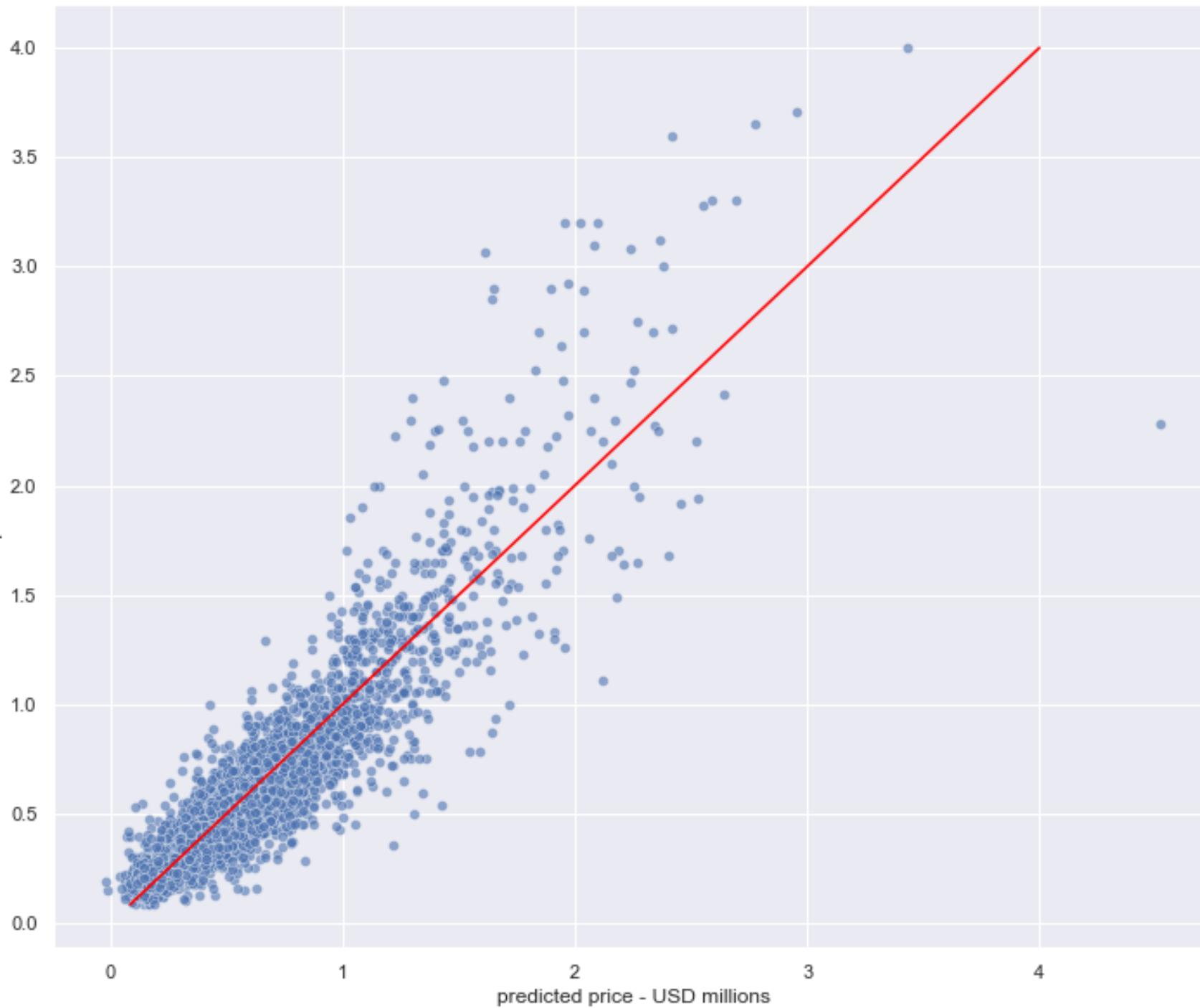
For every 1% increase in square foot living space, the price increases by 0.6%.

For every 1% increase in distance moved away from Seattle and Redmond, price *decreases* by 0.5%.

The model explains 56% of variation of our data from the mean.

Conclusion:

- Using our predictive model, we can reasonably expect to be within \$160k when predicting the sell price of a house.
- Distance from city centers, living area, number of bedrooms, and room sizes are the most important features when interpreting the likely sell price of a house
- This can be further improved with better data and more advanced modeling.



Contact us

Reach out if you have any
questions or clarifications

■ **Abass Ibrahim**

linkedin.com/in/abass-ibrahim/

■ **Nathaniel Martin**

linkedin.com/in/nathaniel-martin-73b037227

■ **Matthew Samson**

linkedin.com/in/matthew-samson-0a1b86165/



Thank you!

Feel free to approach us if
you have any questions.