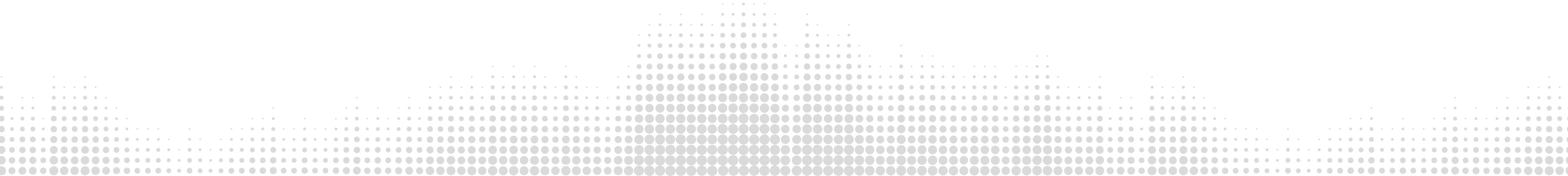


House price prediction

Analysis by: Roman Kaltschew & Selena Eberwein

Agenda

- 01 Data introduction
 - 02 Model presentation
 - 03 House grade
 - 04 Scenic features
 - 05 Geographics
 - 06 Conclusion
- 



Data Introduction

SEATTLE HOUSE
PRICES

Data:

House prices of 21.500 properties sold in Seattle from May 2014 - May 2015

Features:

date sold, number of bedrooms, number of bathrooms, sqft living, sqft lot, number of floors, waterfront, view, condition of the house, grade, sqft above, sqft basement, year built, year of renovation, zipcode, latitude, longitude, sqft living area, sqft lot area, price

Objective:

Building a model that will predict the price of a house based on features provided in the dataset and understanding which factors are responsible for higher property value - \$650K and above

Predictive Model

Cleaning

Kept 99% of the rows
Dropped 5 columns

Dropped features: id, date, year of renovation, sqft above, year sold

Outliers

Removed houses > \$ 4Mio

No further outliers were removed.
Some values were adjusted because of probable typos.

Feature engineering

Aggregated year built and year renovated to age of the house

Since few houses were renovated we considered those as "new built".

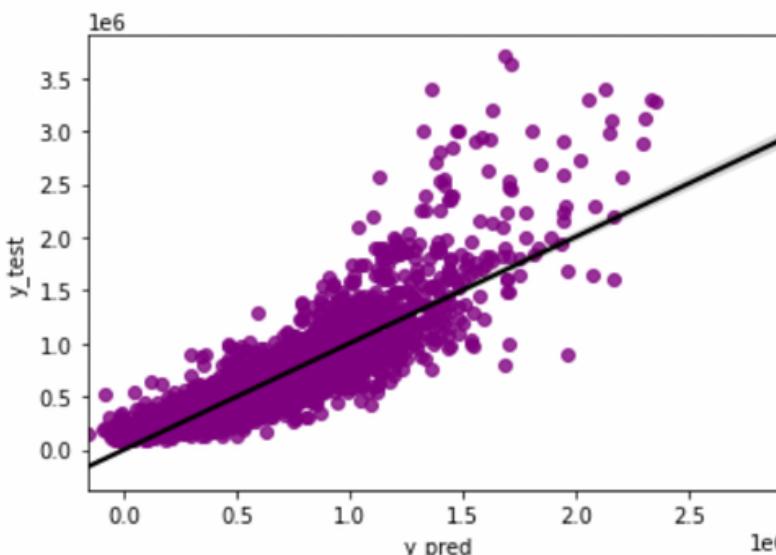


Feature engineering

Grouped zipcodes in 5 categorys by avg price

Improved the model by 4% in adjusted R².

Residuals plot

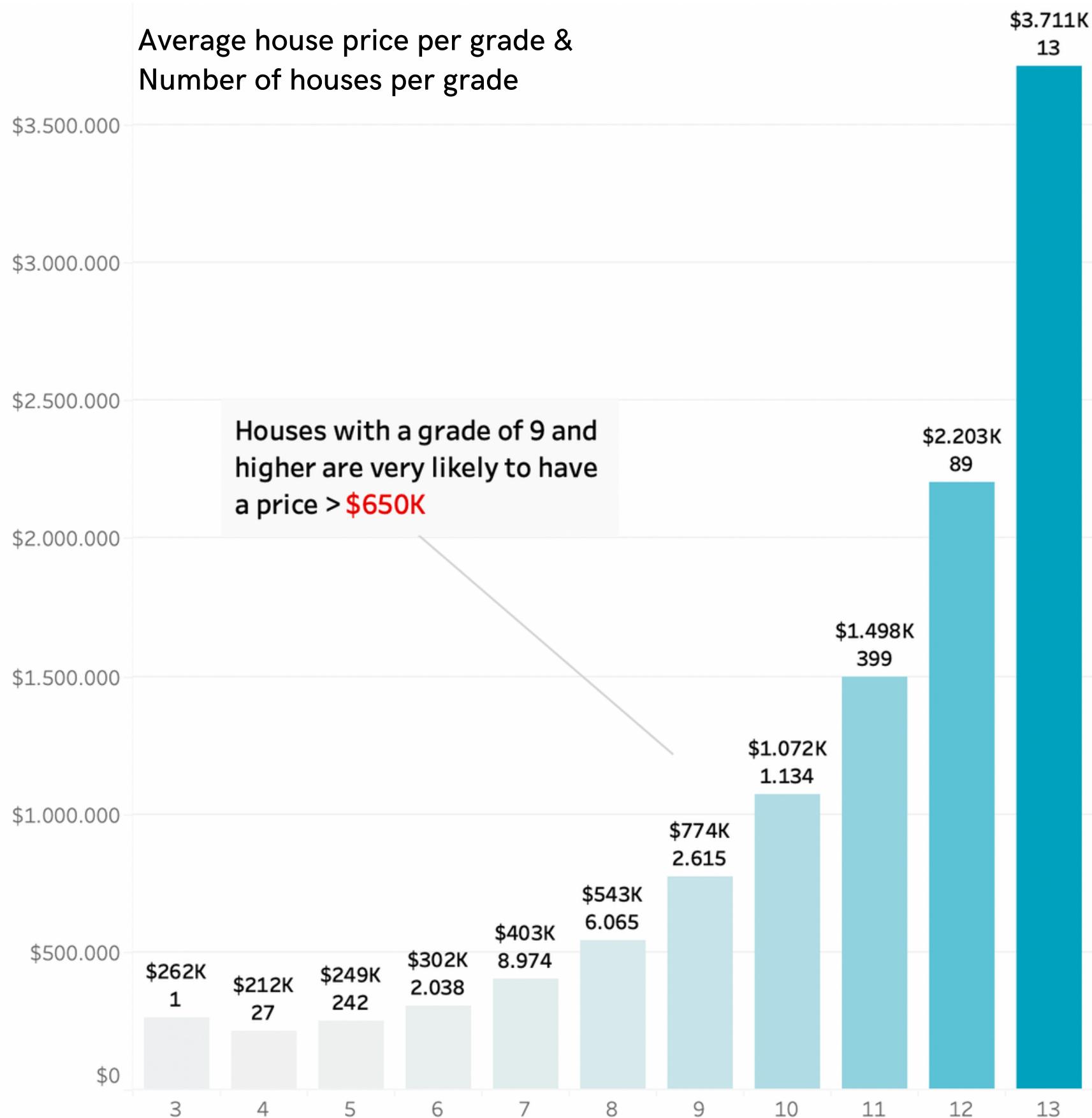


Accuracy

MAE: \$115.418
Adjusted R2: 0.7477

Our model is able to account for approx. 75% of the variation in sales price of a property and it is able to predict the sales price within \$115.418.

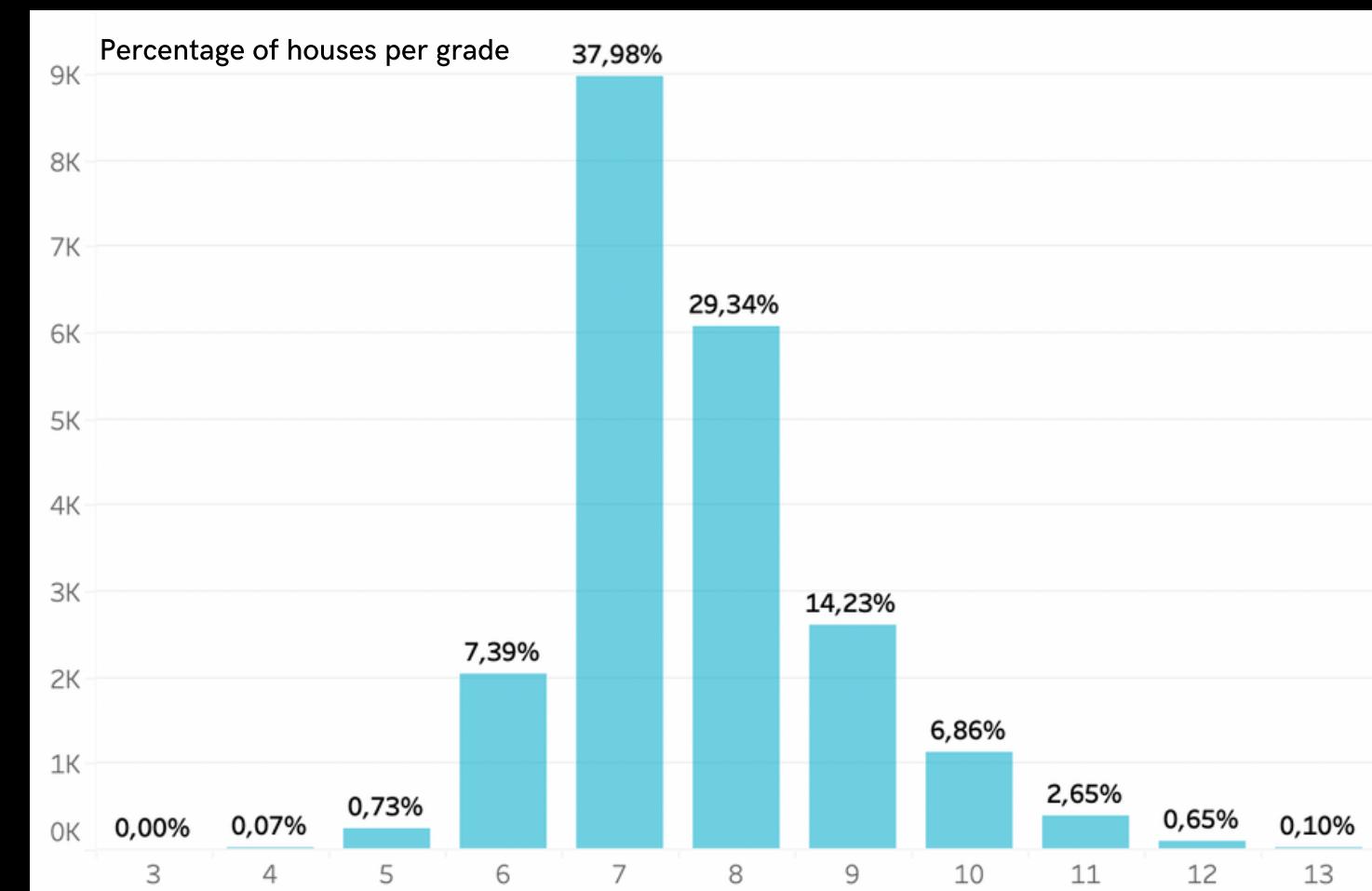
Average house price per grade &
Number of houses per grade



Grade is a strong predictor for price

With higher grade, prices increase exponentially

However, 70% of the houses are graded with 7 or 8 which means average grading.



\$1.662.524

Is the average price of houses having a waterfront

80% of the houses having a waterfront cost more than \$650K

27.2%

Is the minimum increase of the price per sqft for houses having a view

Average price/ Sqft living per view grade



Waterfront and view ramp up prices

Only few houses come with these features

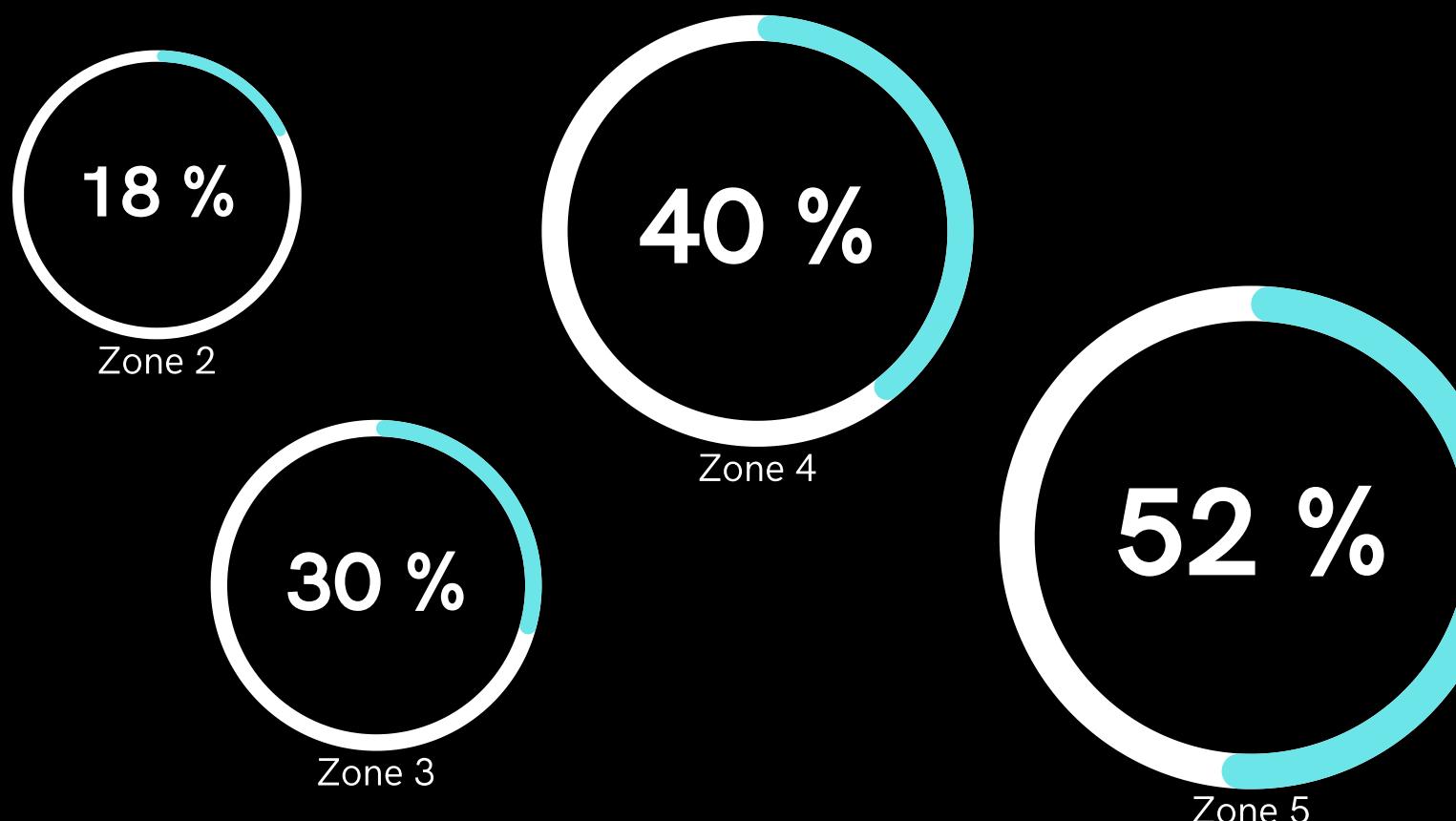
Less then 1% of houses have a waterfront and only 10% have a view graded 1 or higher



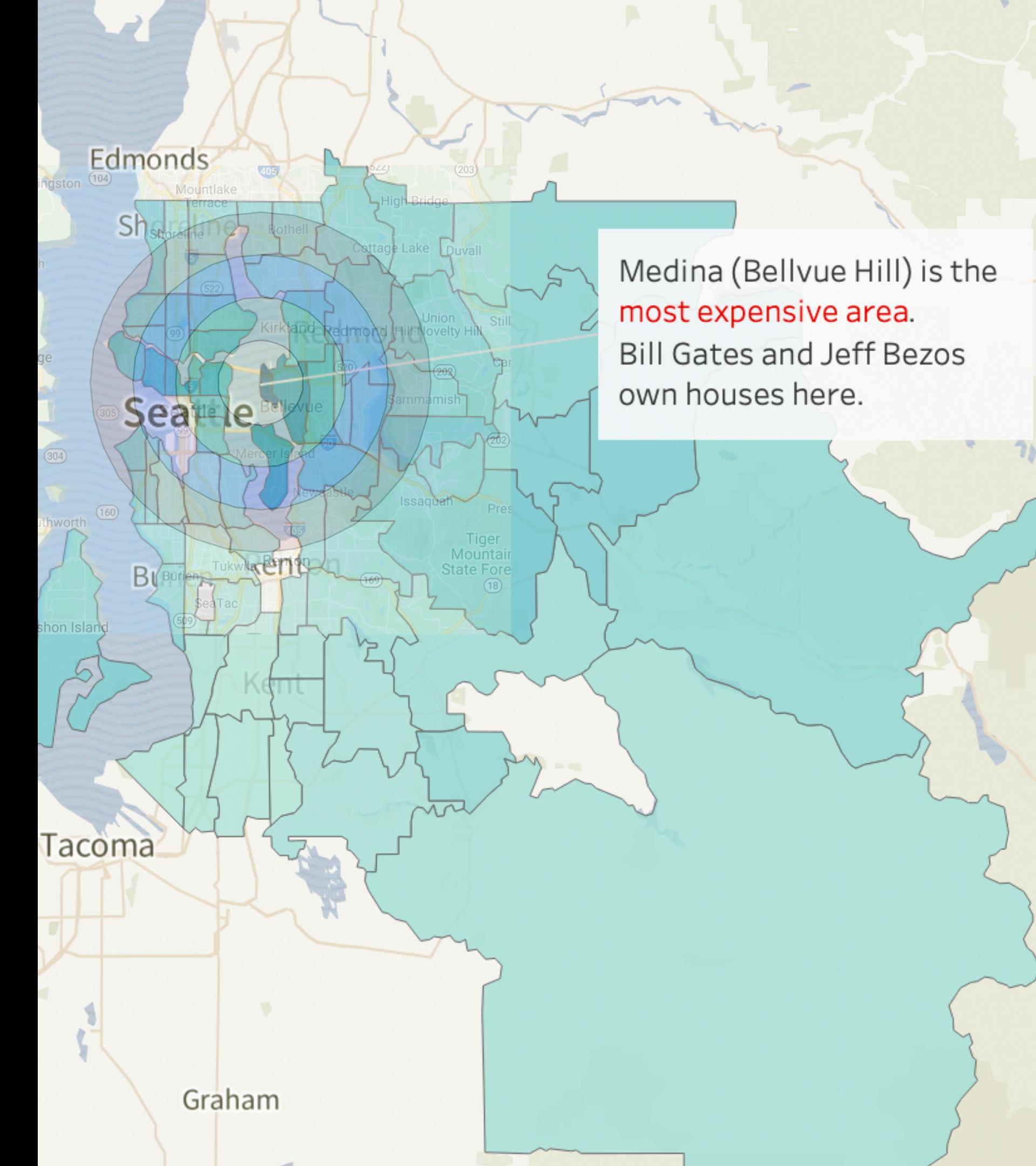
Price zone

After the house size (sqft), living close to the city center is the no1 driver for price

Decrease per zone of the price/sqft living



Average for a three bedroom house



Conclusion

Prediction for house prices is a complex operation, for that linear Regression might not be the best model, since most of the features are categorical.

Another suitable model would be the Random Forest Classification Model, which was not applied in our analysis.

For linear Regression consider adding features, such as:

- Schools distance
- Traffic
- Overall popularity of the district
- Supply and demand

SEATTLE HOUSE
PRICES

Copyright

Roman Kaltschew
Selena Eberwein

GitHub

https://github.com/Romes-Kalt/Housing_linear_regression
https://github.com/SEberwein/Housing_linear_regression

Data

https://github.com/raafat-hantoush/IH_RH_DA_FT_OCT_2021/blob/main/Class_Materials/MidTerm_Project/Data/Data_MidTerm_Project_Real_State_Regression.xls