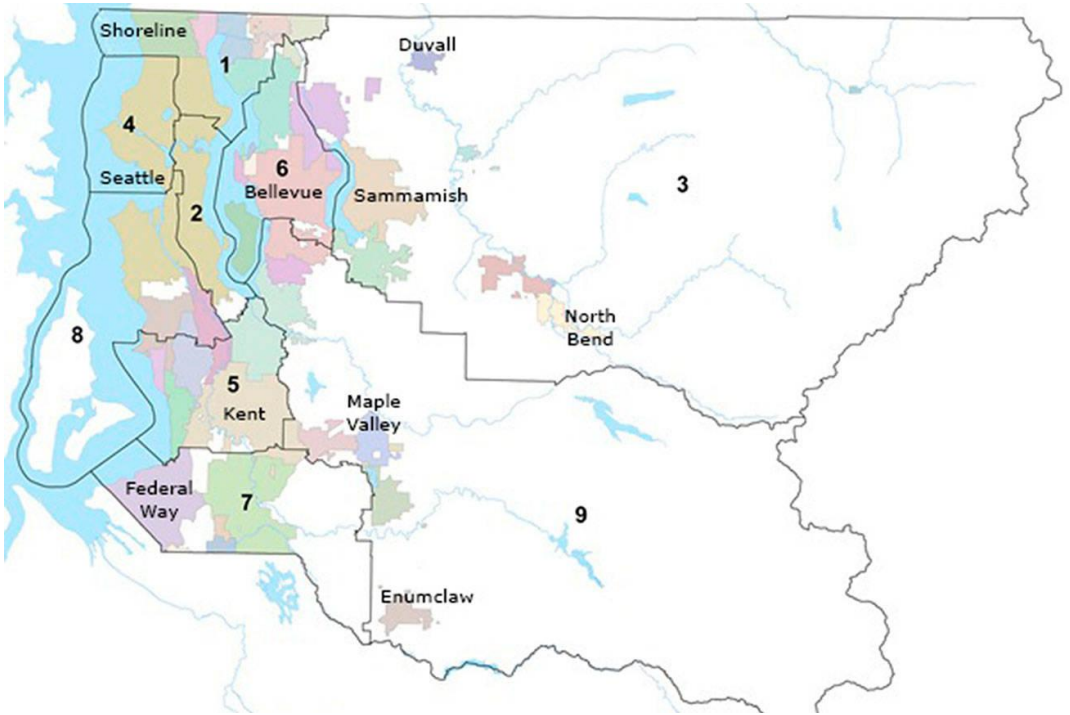
King County, Washington: Data Analysis & Strategy Recommendations

Pete VanZandt, Emma Choate, Alex Valencia
July 2nd, 2021

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

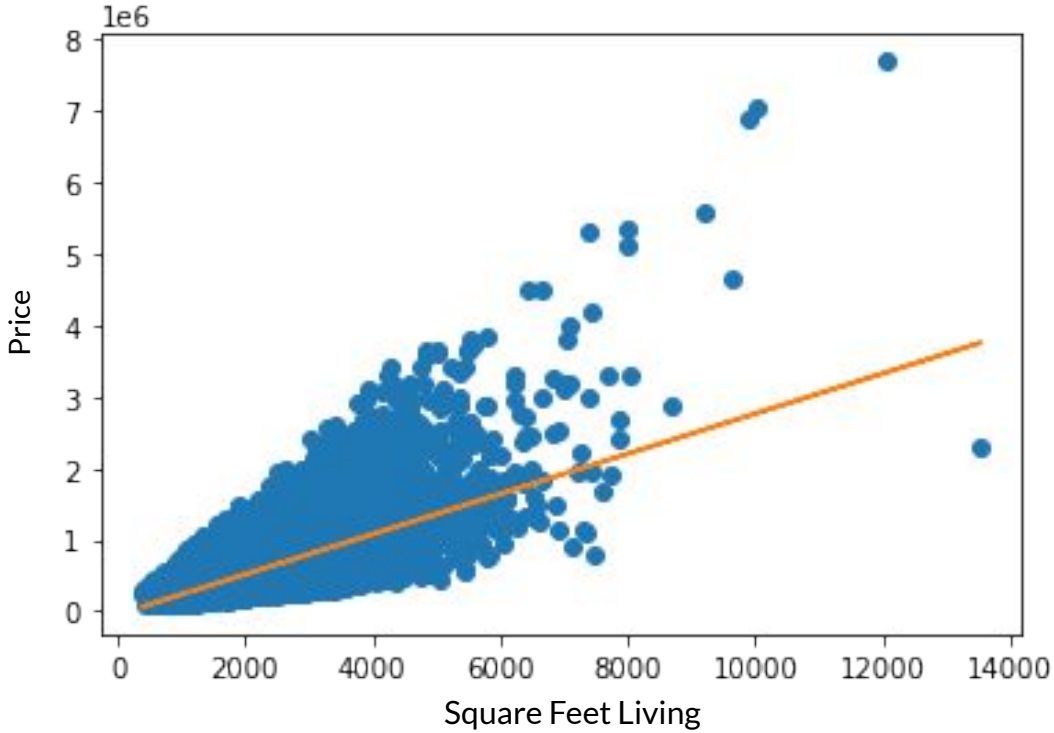


Business Problem

A Seattle real estate firm wants their clients to have a prediction for the sale price of their home in King County. In order to achieve this, they would like a tool that can be as accurate as possible in their predictions of the sale price. We have come in to create this tool and use it for the real estate firm's clients.

The challenge

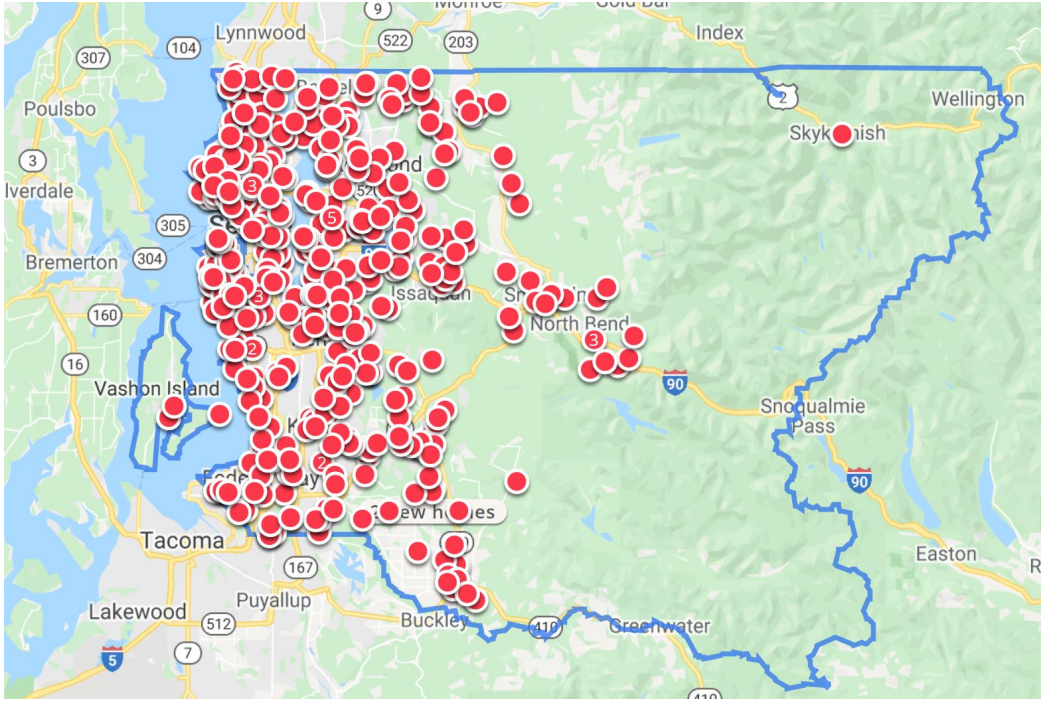
- King county is large
- Housing dominated by Seattle
- House prices vary considerably



Data

We used data on houses from King County Washington. This data gave information about the homes such as:

- 2014-2015
- 21,597 homes
- Homes within the price range of $78,000 and $7.7 million
- Location of the map (clustered around seattle)
- Number of bedrooms and bathrooms
- Square feet of the home and the surrounding homes
- Grade
- Etc.



https://www.zillow.com/king-county-wa/

Methods

- Imported necessary packages
- Inspected & cleaned the data
- Data was sourced to identify neighborhoods
- Started with a simple model of the data to see correlations
- Created a complex model using interactions found in order to make predictions

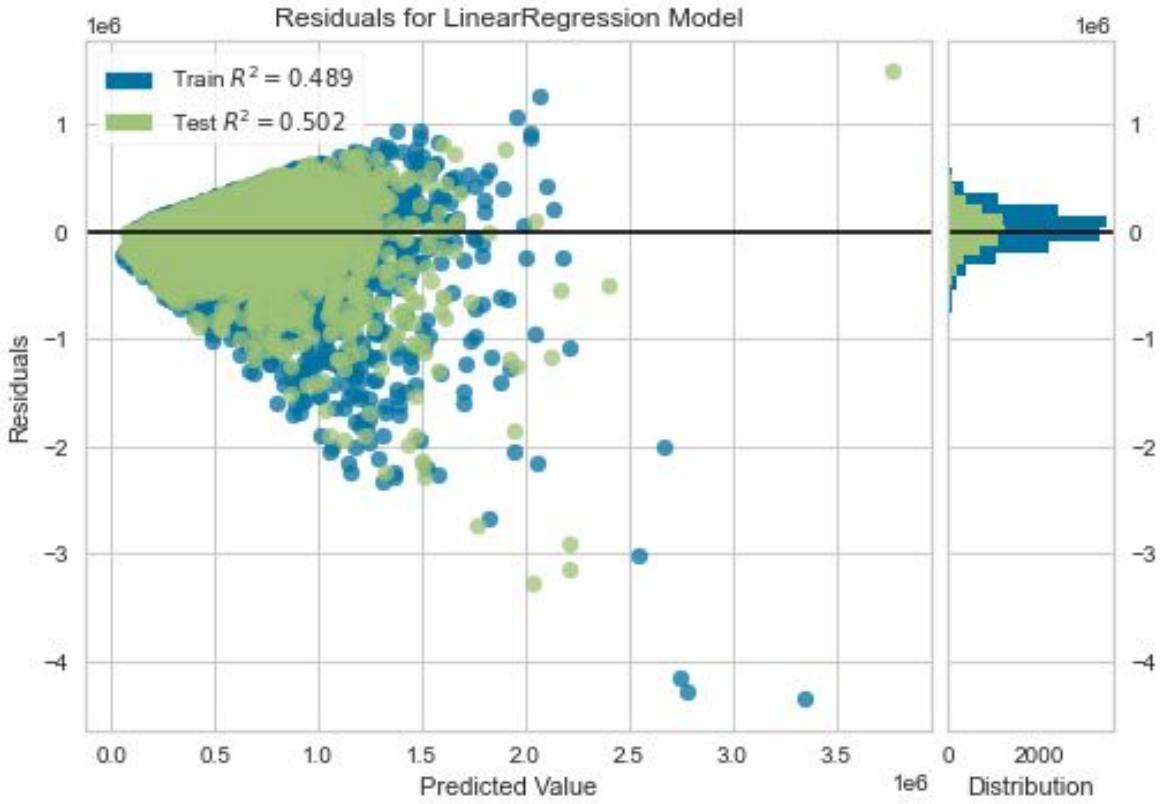
Baseline Model

-Only feature is square feet living

- Test Error (RMSE): 259,878

- Train Error: 262,419

-Most interpretable, least predictive



Complex Model

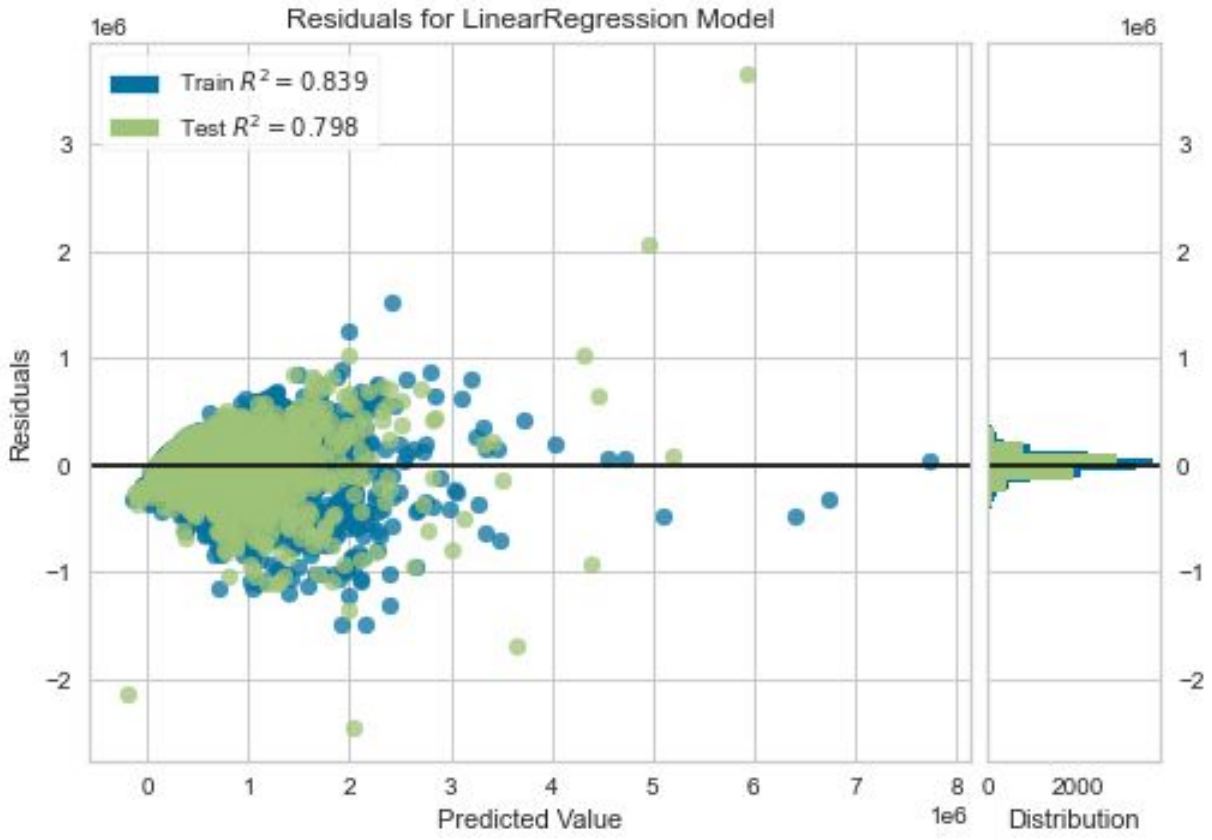
-Polynomial linear regression model

-Comparing 253 features

-Train Error (RMSE): 147,269

-Test Error: 165,436

-Most predictive, least interpretable



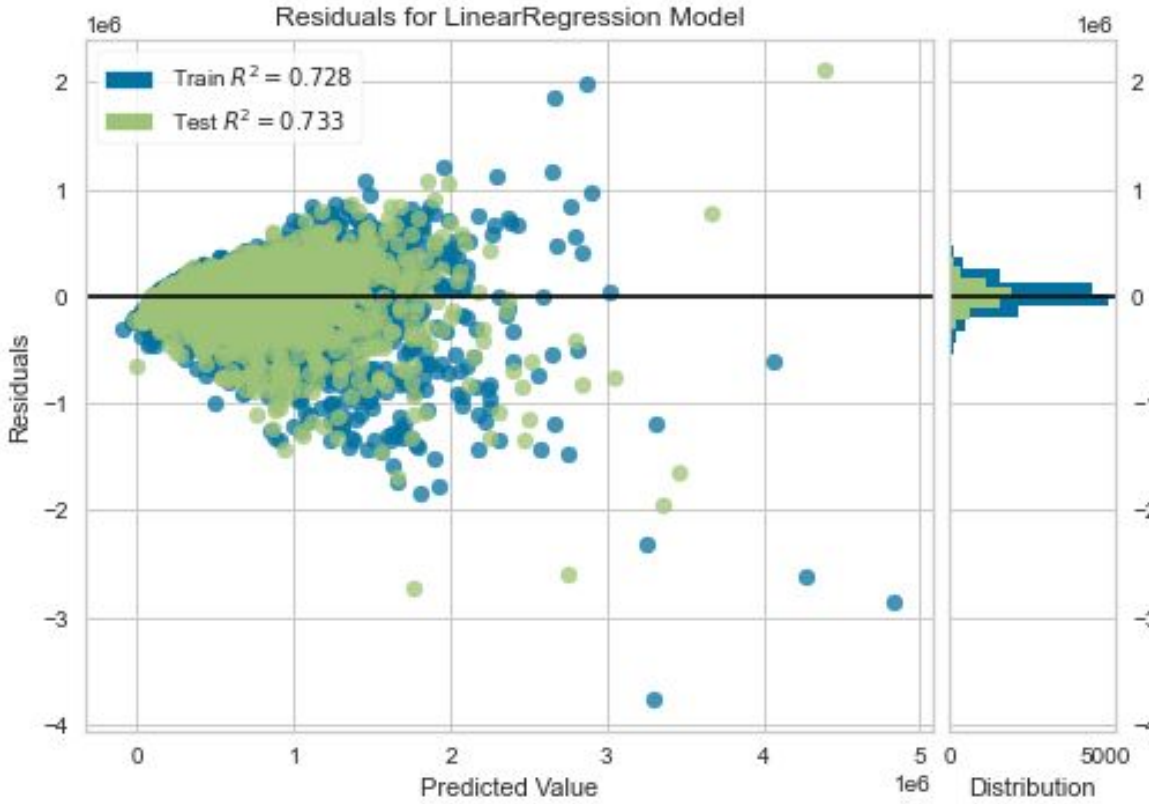
Suggested Model

-Iterative modeling with poor predictors removed

- 13 features

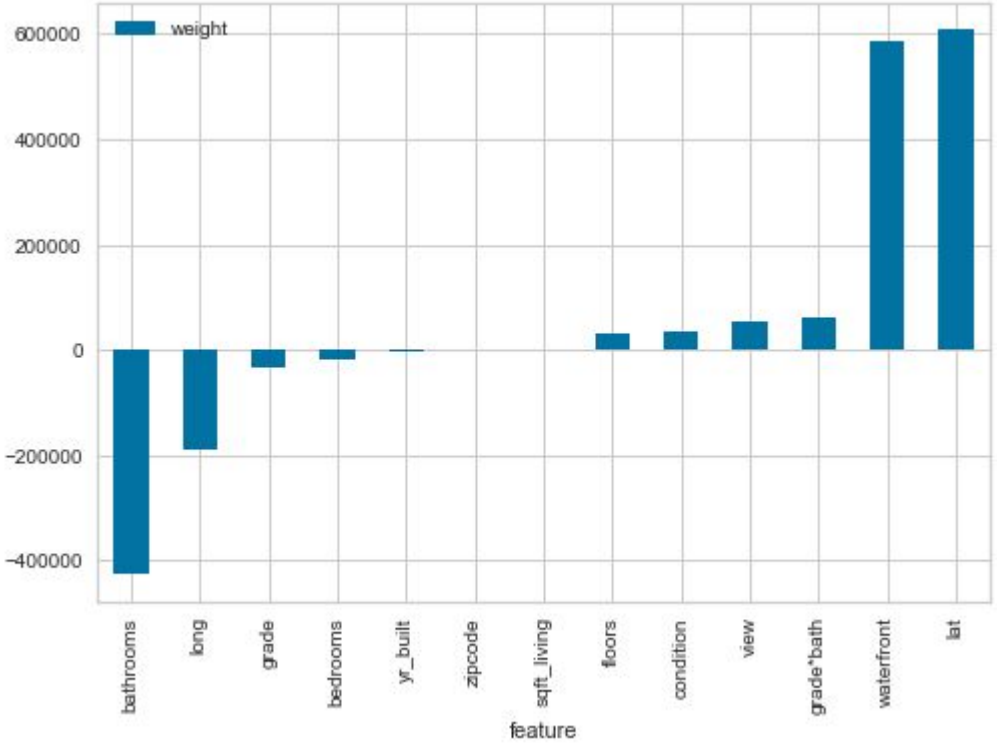
- Test Error (RMSE): 190,102

-Train Error: 191,327



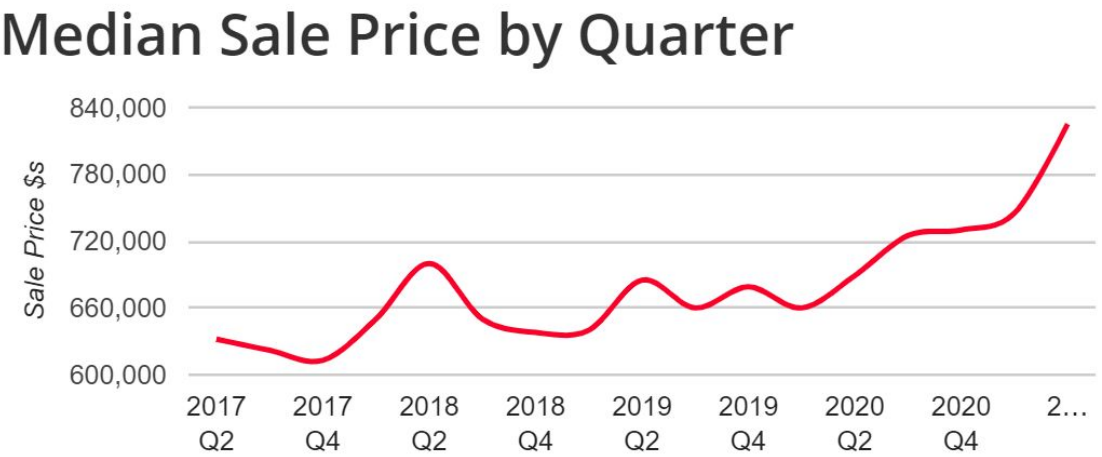
Most important features of suggested model

- 13 features and their weights on the overall price estimate
- Latitude = 608,000
- Bathrooms= -425,000



Future Directions

- Source more current data
- Possibly use nonlinear/more advanced machine learning models
- Look deeper into the data that the model consistently couldn't handle



https://localscape.property/#kingcountyassessor/

Conclusions

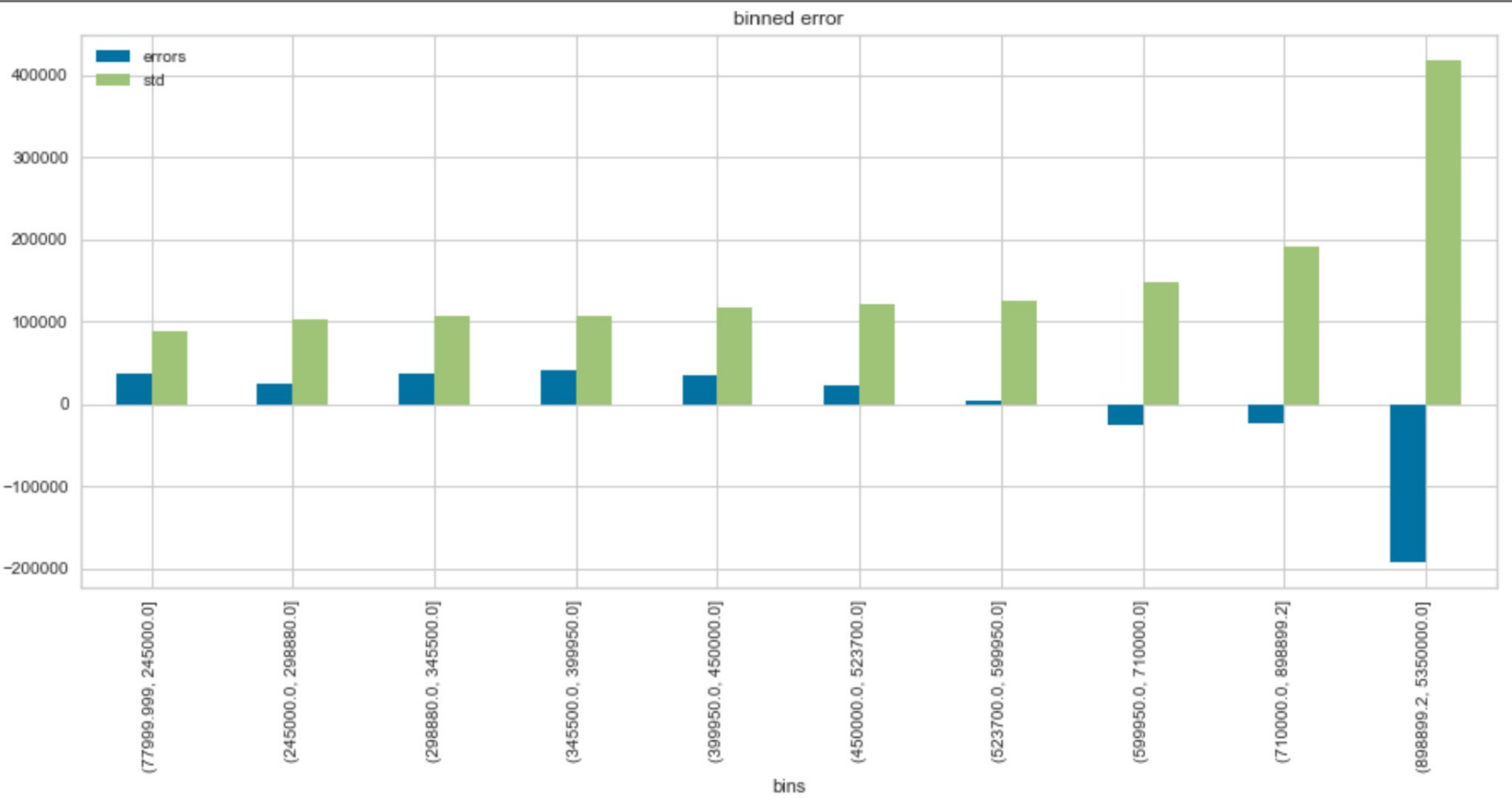
- We were able to create a model to predict the selling price of a home across a wide range of prices and conditions

- Our model could be used on recent data to help the real estate company provide their customers with helpful insights to predict the selling price of their home based on its features

- Breaking down homes into price ranges, using nonlinear modeling, and tracking home data trends over time may produce more reliable predictions

Thank You!

Email: pevanzandt@gmail.com, emmachoate11@gmail.com, asvalencia1688@gmail.com
GitHub: @pvanzand, @EmmaChoate, @asval211
LinkedIn: linkedin.com/in/emma-choate-8287671ba/, linkedin.com/in/asvalencia/


binned error
errors
std
400000
300000
200000
100000
0
−100000
−200000
(77999.999, 245000.0]
(245000.0, 298880.0]
(298880.0, 345500.0]
(345500.0, 399950.0]
(399950.0, 450000.0]
(450000.0, 523700.0]
(523700.0, 599950.0]
(599950.0, 710000.0]
(710000.0, 898899.2]
(898899.2, 5350000.0]
bins