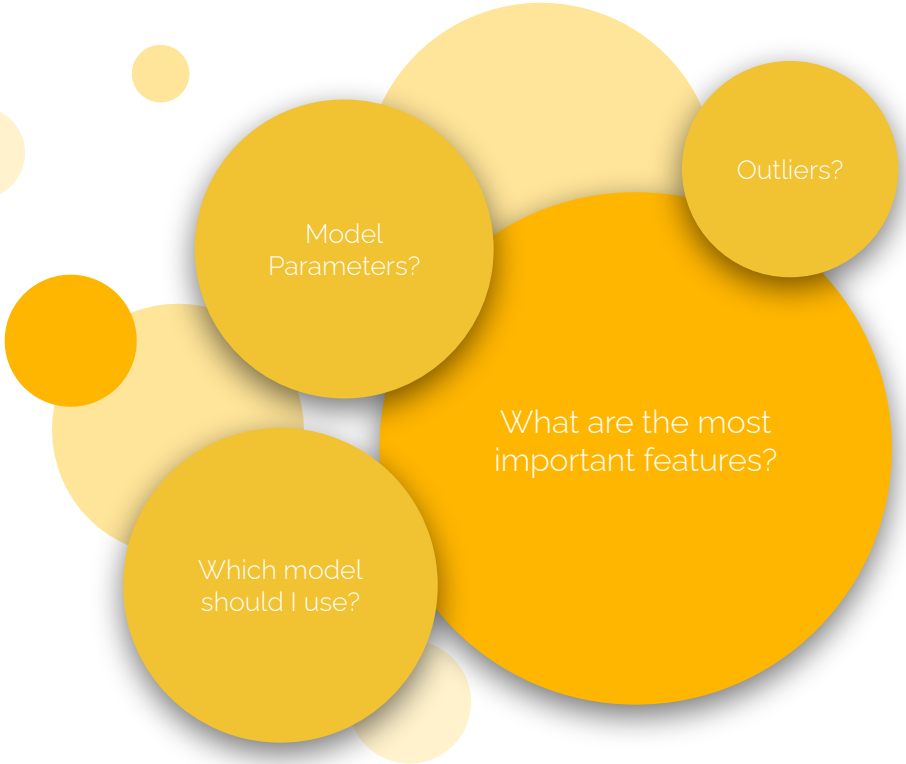
Predicting home sale prices in Ames

Hello!

I am Alex Fioto

I am here to talk about predicting home sale prices based on data from Ames, Iowa.

What are the best predictors of house sale price?



1

Check the data.

What is in there?

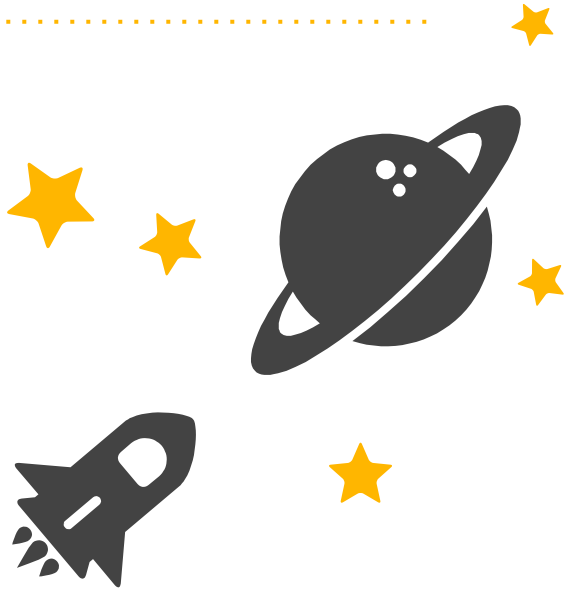
Data Cleaning

- Data types: Numerical, Ordinal, Categorical
- Imputing missing data: KNN
- Removing Outliers

2

What seem like promising features?

80 Features!



How many? Too many.

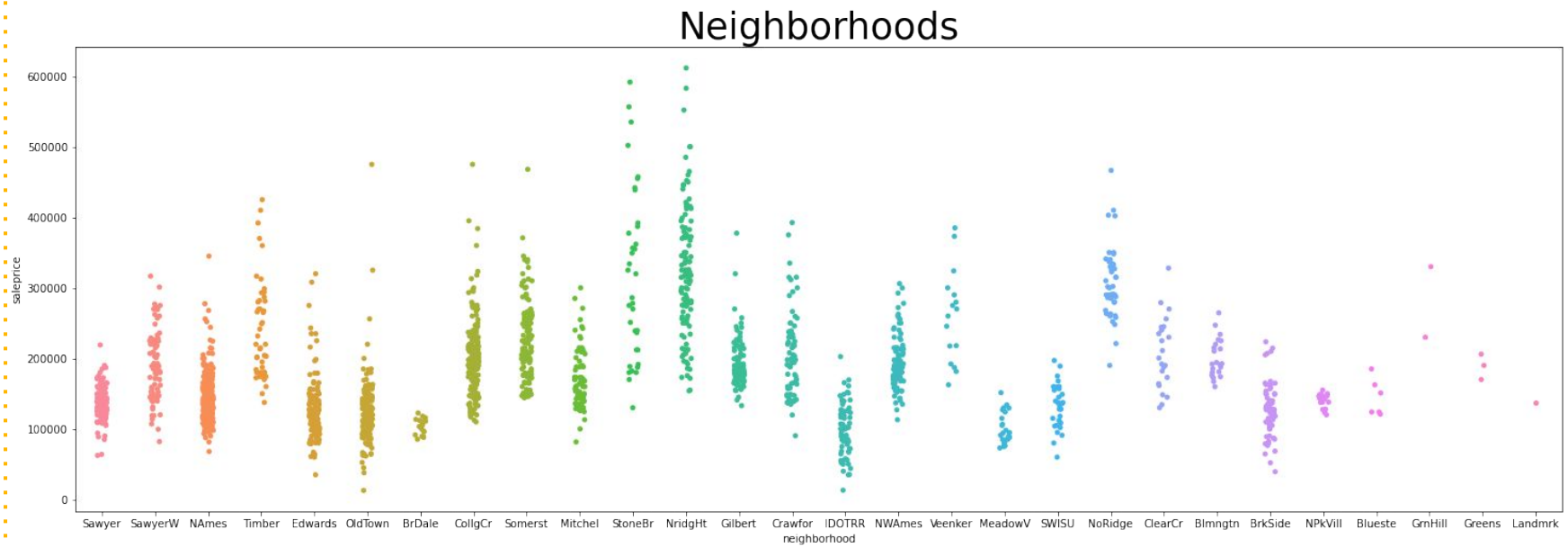
Maybe...

Selected Features

- Square Footage
- Overall Quality
- Neighborhood
- Garage
- Basement (quality and size)
- Bathrooms
- Heating and Air
- Year Built/Remodeling
- Lot Size

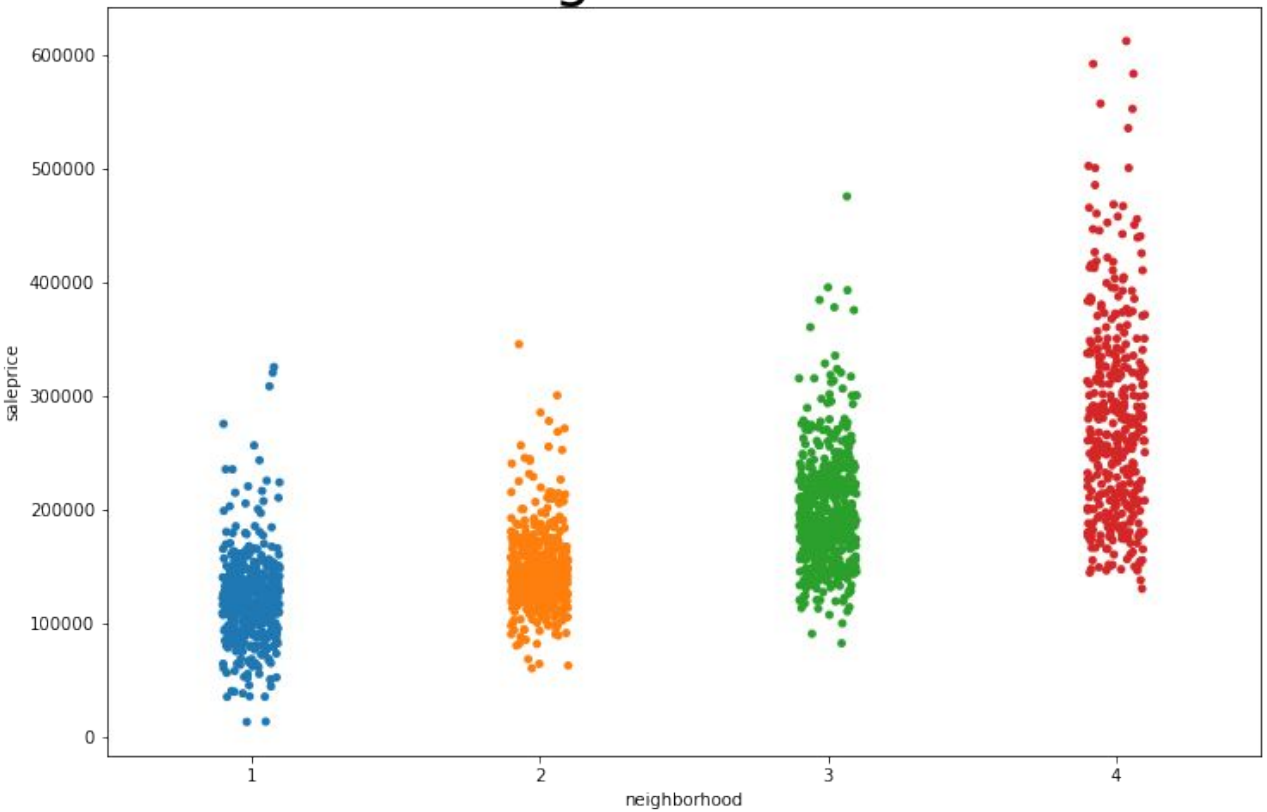
Location, location, location!

- *Harold Samuel*



Inspiration from Alvin T. Tan on medium

Neighborhoods



3

What is the best model?

Linear Models

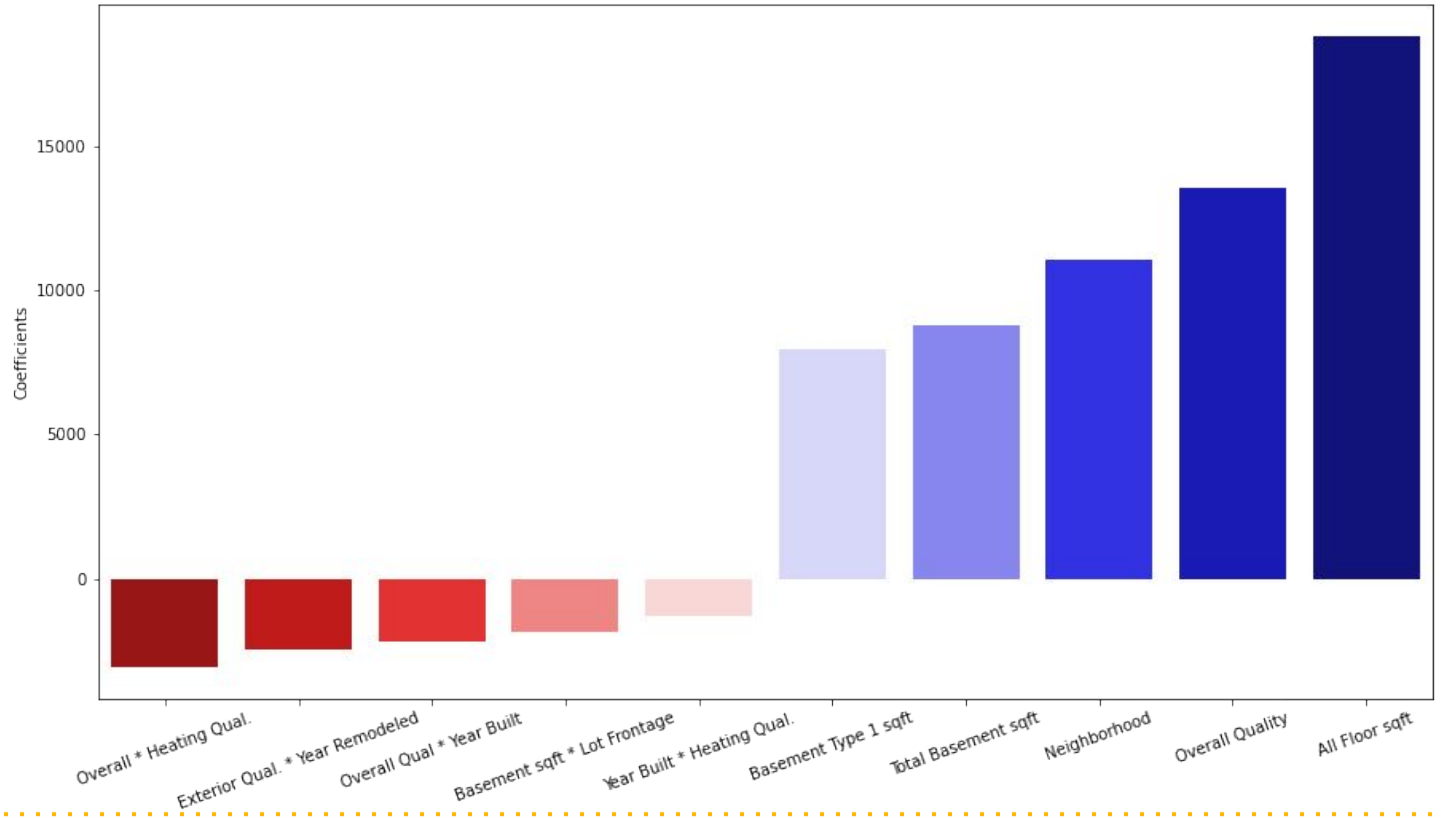
- Linear Regression
- Ridge
- Lasso
- Elastic Net

Ridge

Pipeline and Hyperparameters

- KNN Imputer (k=5)
- Standard Scaler
- Polynomial Features
- Select K Best (147 out of 150)
- Ridge (Alpha = 175)
- R2 = 0.9381
- RMSE = 18892.8

Features that Matter



Conclusion

- Quality
- Square Footage
- Neighborhood

Thanks!

Any questions?