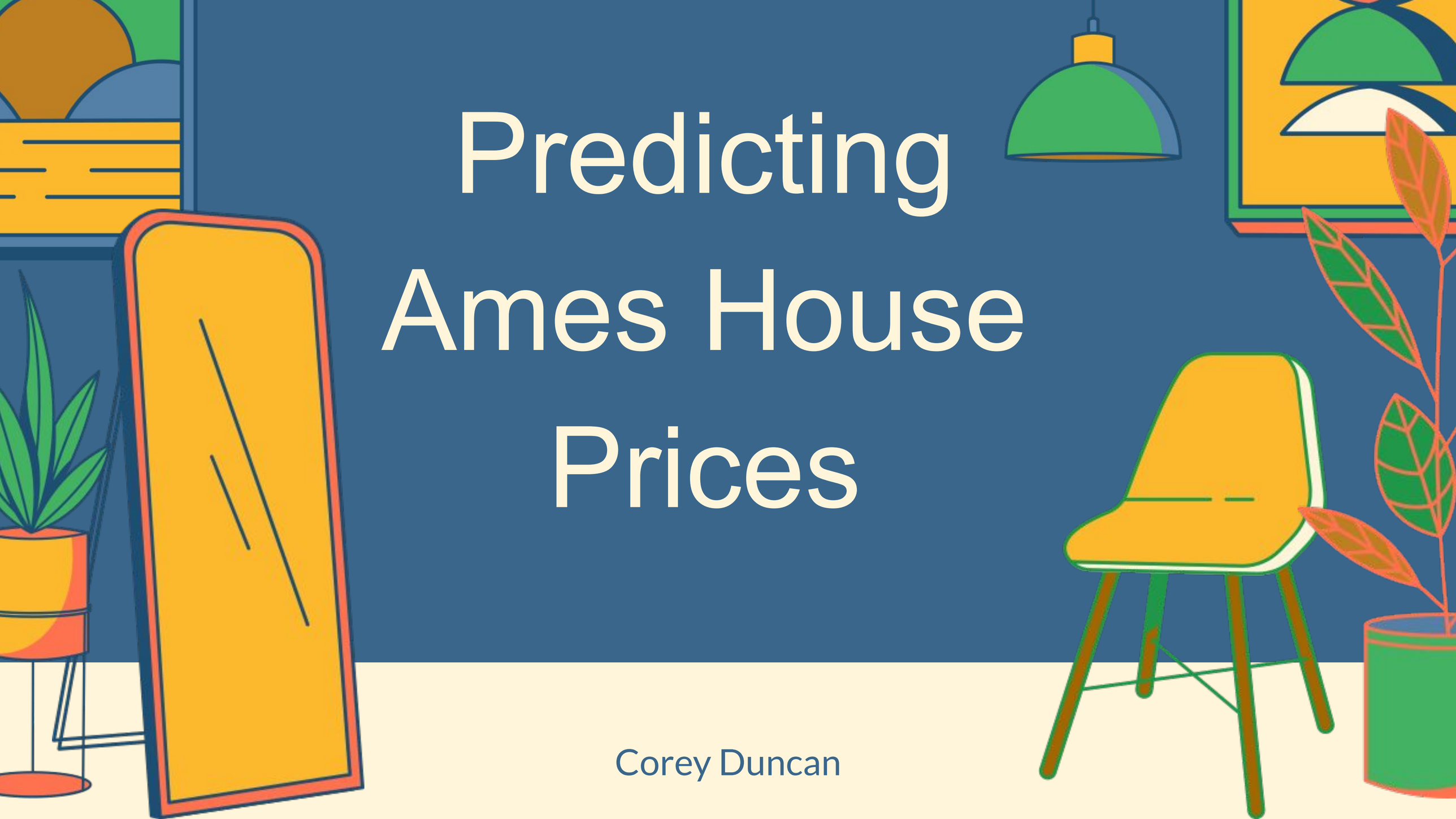
Predicting Ames House Prices



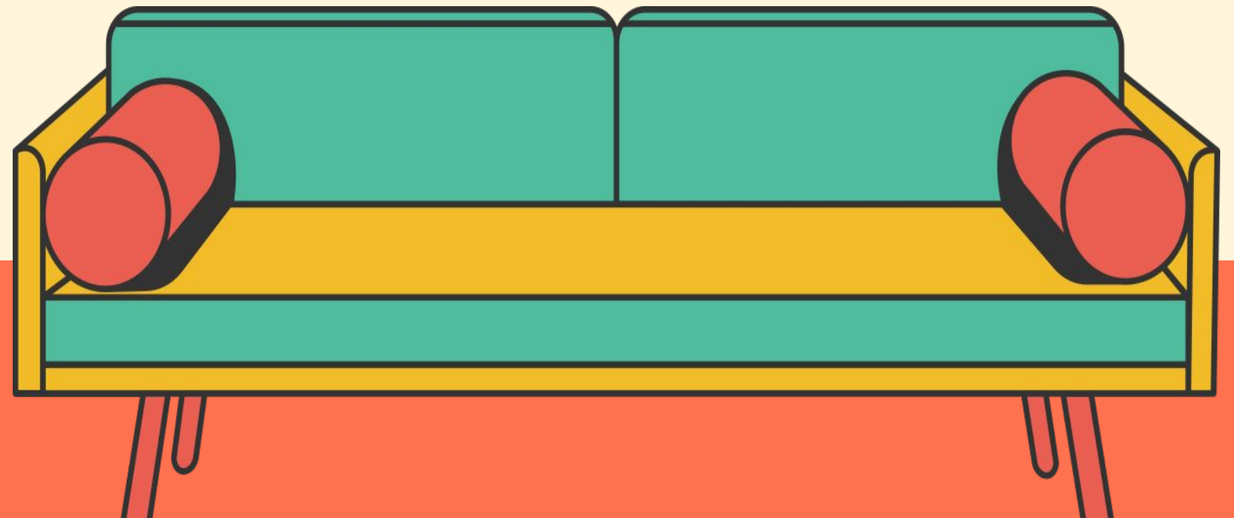
Corey Duncan

Why?

To assist in making informed decisions, thus enhancing market efficiency and transparency.

Presentation Outline

1. Problem Statement
2. Models
3. Other findings
4. Conclusions



Problem Statement

- Given a set of data on Ames housing features and sale prices, how accurate of a model can we build to predict future sale prices?
- Success will be measured by R2 score achieved

Models

Model #1

- For my first model, I picked 10 variables at my discretion.
- The resulting model was better than baseline, but not by much.

Model 1

Train R2	0.766
Test R2	0.752
Baseline Score	49334.66
Kaggle Score	45739.82

Model #2

- For my second model, I wanted to be more accurate by looking at the correlation of my numeric variables.
- I also ran ANOVA tests on my categorical variables.
- Removed other non matching variables from first model.

Model 2

Train R2	0.842
Test R2	0.844
Previous Model	45739.82
Kaggle Score	31138.39

- Included the top 10 correlated numeric variables
- The 5 highly correlated categorical variables

Model #3

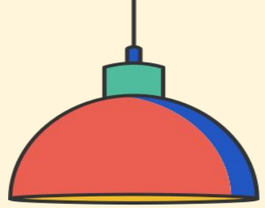
- Tried using different kinds of regression
- Tried removing columns with multicollinearity
- Ended up getting best score by combining all features of first two models

Model 3

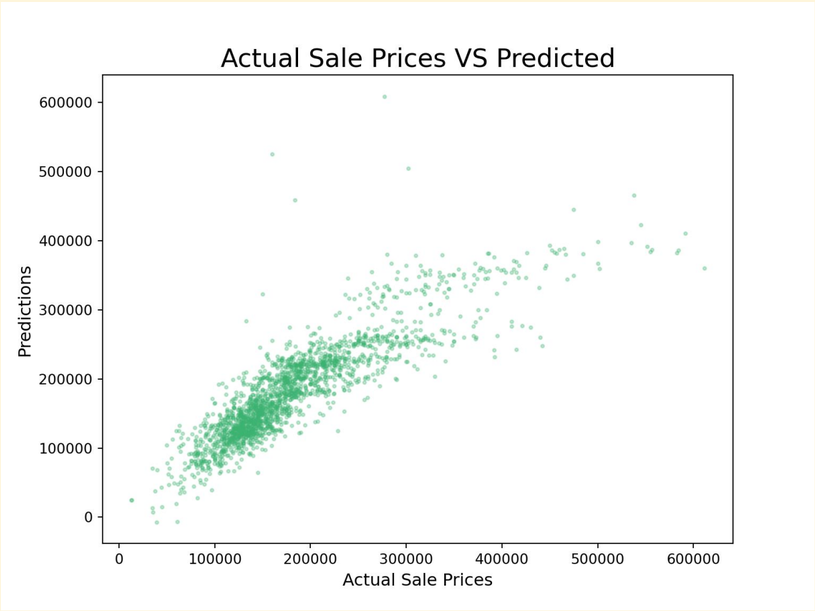
Train R2	0.852
Test R2	0.860
Previous Model	31138.39
Kaggle Score	29814.28

Only a slight improvement over the previous model, but still an improvement!

Side by Side

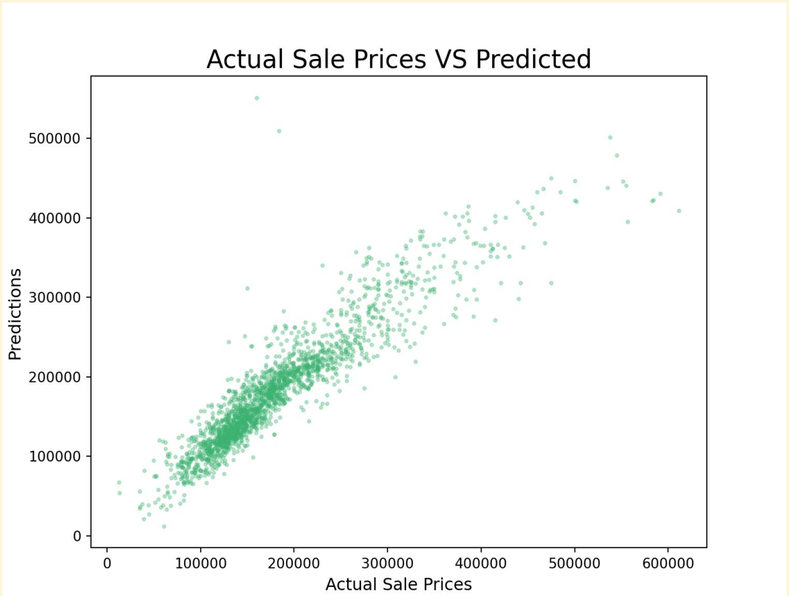


Model 1



Model 2

Model 3



Other Findings

- Lasso model
- Ridge Cross Validation model
- Tried reducing the number of variables used to only the most correlated
- Tried removing features with multicollinearity

Conclusions

I was able to achieve my most accurate model by:

- Using Anova tests to determine which categorical variables to use
- Testing my numeric features for correlation and using the best ones
- Using more features

Thank you for listening!

Works Cited

Kaggle Competition
https://www.kaggle.com/competitions/dsb-521-ames-housing-challenge

ANOVA Tests
https://www.scribbr.com/statistics/one-way-anova/

GA Lessons 301-306, 401, 403

Questions?