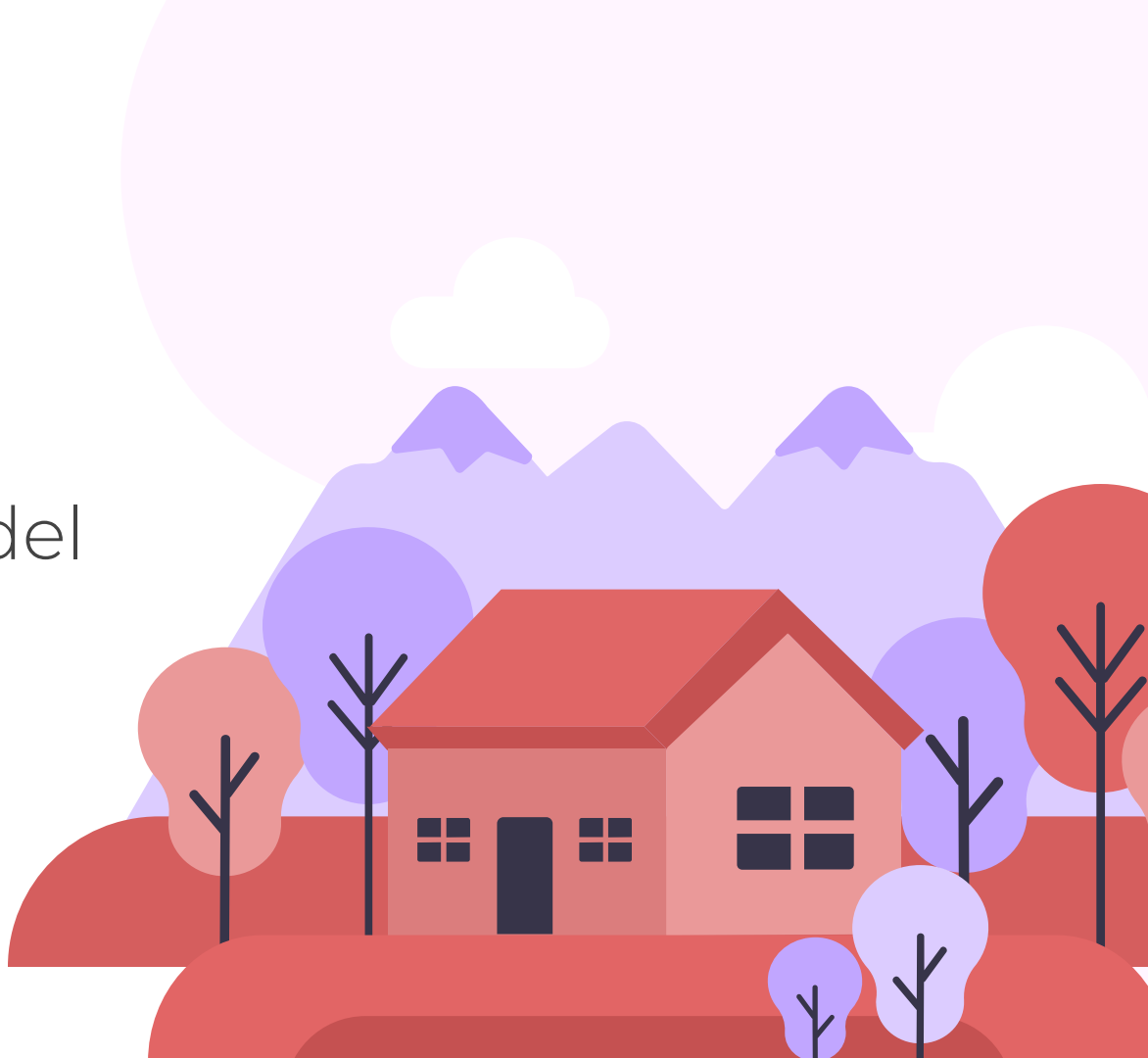
Project 2

Ames, Iowa

Regression Model



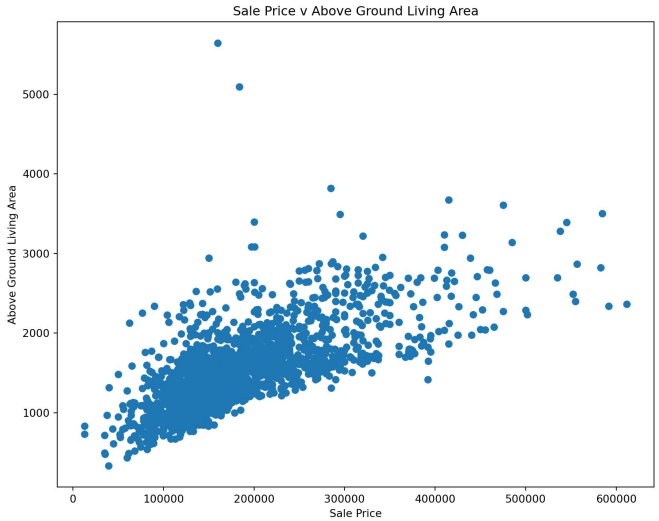
The Goal

The goal of this project is the creation of a regression model that predicts housing prices in Ames Iowa.
Predictions generated by the model were submitted to Kaggle for evaluation.

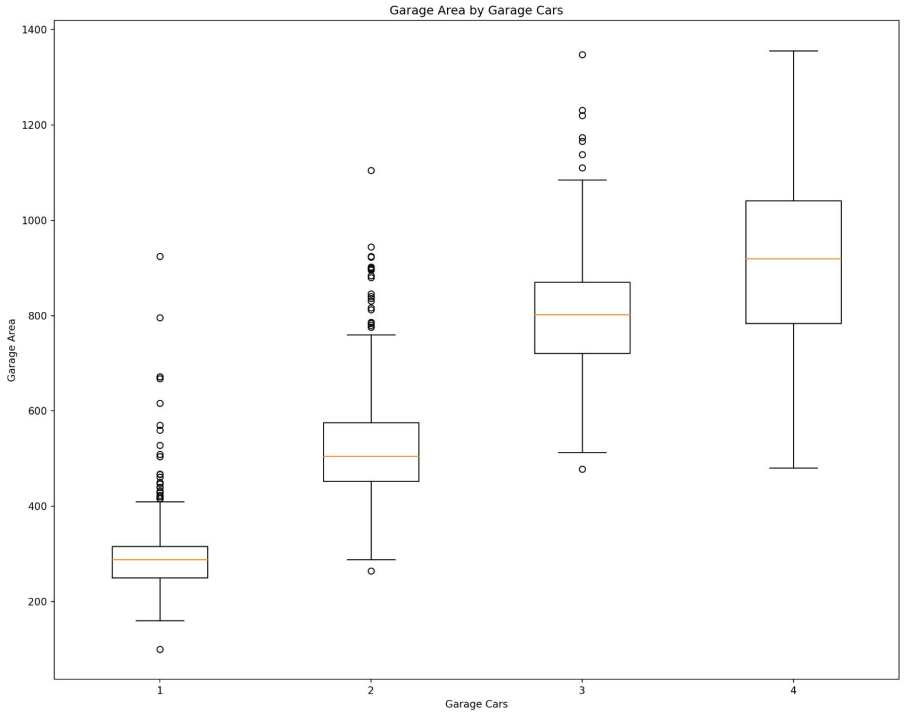
Image from The Daily Iowan

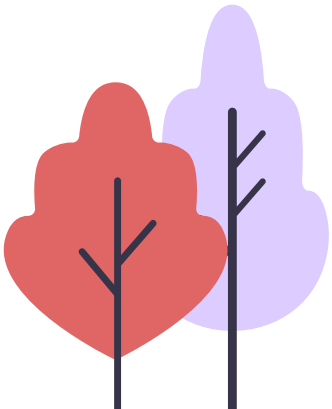
Cleaning

- In order to build the model several columns in the dataset had to be adjusted
- Garage Quality, Garage Condition, Garage Cars, Pool QC, and Fence had missing values
- In addition, two outliers were spotted and removed in the training data



Sidenote: Collinearity



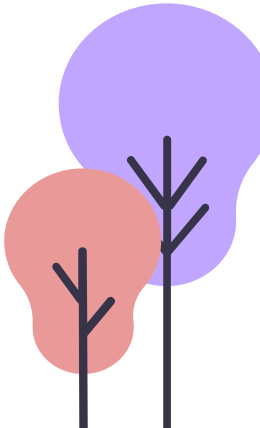


Baseline

- The null model predicts the average and does not take any input
  - It does not explain any of the variance in prices, so its R-squared score is 0
- The second model I created predicts price based on Overall Condition, Overall Quality, Year Sold, and Total Rooms Above Ground
  - It achieved an R-squared score of 0.697

Adding Complexity

- The third model adds Lot Area, Neighborhood, Building Type, Heating, Heating QC, Central Air, Electrical System, Kitchen Quality, Pool Area, Pool QC, Fence, and Miscellaneous Value
  - It uses a One Hot Encoder to process the categorical variables
- With these additions, its R-squared score reaches 0.83



The Best Model

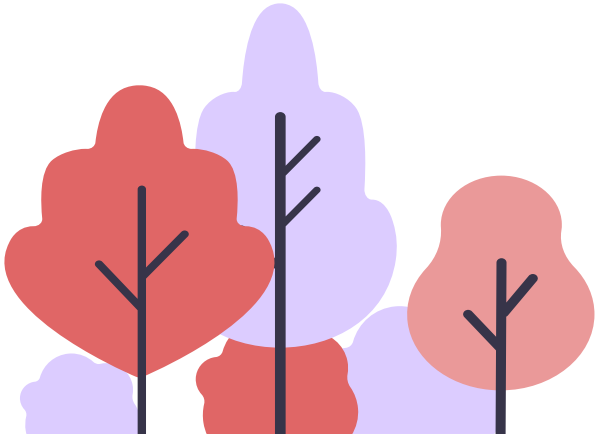
- The fourth model has the best performance overall
  - It uses 27 inputs
  - It uses a One Hot Encoder to transform categorical variables and Polynomial Features on selected numeric variables
- It had an R-squared of 0.893 on testing data and the best Kaggle score

Too Much Variance

- The penultimate model added even more variables and had the lowest bias of any model. However, it was overfit and performed worse on the testing data and on Kaggle.
- The final model used the inputs as the previous, but used Ridge Regularization to try to reduce variance
    - It helped, but not enough to overtake the fourth model

Conclusion

- Overall the fourth model is the best due to it having low bias without variance getting too high.
- It can be used by prospective buyers and sellers in the housing market in Ames in order to predict the value of a house.



THANKS

Does anyone have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.