AMES HOUSING PRICE PREDICTION

"Wait how'd you get that price?"

By: Damar Shipp

PROBLEM STATEMENT

Housing prices are influenced by numerous factors such as the year it was built, the number of bedrooms, neighborhoods, and more. The goal is to create a model that can predict the price of the houses based on these variables with high accuracy. Given the string of data and complexity to find the potential correlations between variables, this problem involves both variable selection and the application of machine learning techniques to return accurate predictions.

TABLE OF CONTENTS

OBJECTIVES

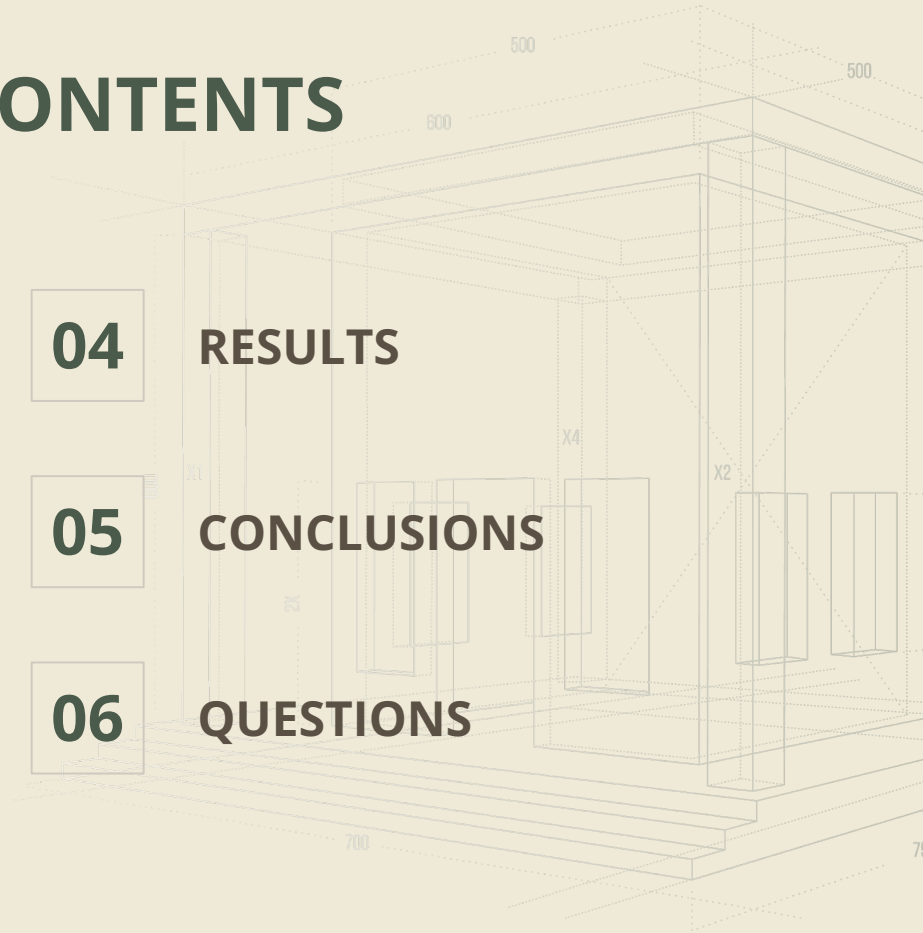## OUR AIM

Given the string of data and complexity to find the potential correlations between variables, this problem involves both variable selection and the application of machine learning techniques to return accurate predictions.

THE GOAL

Develop a regression model that predicts the final price of a house using the features of the house with a high accuracy.

DATA

AMES IOWA HOUSING DATA

- Kaggle data about Ames housing final sale prices.

Assumptions

- Dataset is accurate
- Selected features are the most ideal
- Economic conditions are the same

Challenges

- It includes 2008 which is were the housing crisis occurred and prices were desperately sold at lower prices.
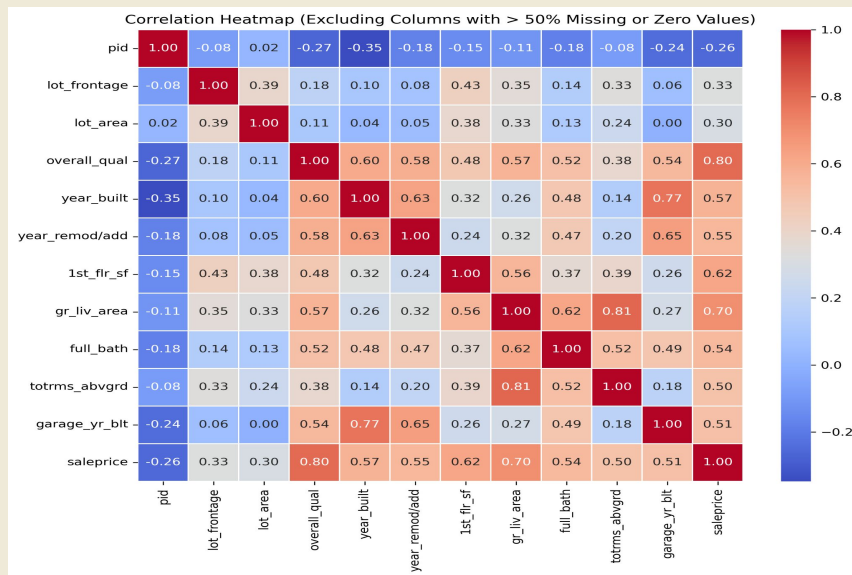
METHODOLOGY

Process of the creation

Trained a model with features engineered to help get to the best accuracy. Each attempt more features were engineered.

EQUIPMENT & MATERIALS

A Linear Regression model was the machine learning model of choice. I know there are better models but at this point in my career this is what I am most familiar with.

FEATURE IMPORTANCE

Feature importance was based on the the correlation of the features to my target (the final sale price)

Correlation Heatmap (Excluding Columns with > 50% Missing or Zero Values)

|	pid	lot_frontage	lot_area	overall_qual	year_built	year_remod/add	1st_flr_sf	gr_liv_area	full_bath	totrms_abvgrd	garage_yr_blt	saleprice
pid	1.00	-0.08	0.02	-0.27	-0.35	-0.18	-0.15	-0.11	-0.18	-0.08	-0.24	-0.26
lot_frontage	-0.08	1.00	0.39	0.18	0.10	0.08	0.43	0.35	0.14	0.33	0.06	0.33
lot_area	0.02	0.39	1.00	0.11	0.04	0.05	0.38	0.33	0.13	0.24	0.00	0.30
overall_qual	-0.27	0.18	0.11	1.00	0.60	0.58	0.48	0.57	0.52	0.38	0.54	0.80
year_built	-0.35	0.10	0.04	0.60	1.00	0.63	0.32	0.26	0.48	0.14	0.77	0.57
year_remod/add	-0.18	0.08	0.05	0.58	0.63	1.00	0.24	0.32	0.47	0.20	0.65	0.55
1st_flr_sf	-0.15	0.43	0.38	0.48	0.32	0.24	1.00	0.56	0.37	0.39	0.26	0.62
gr_liv_area	-0.11	0.35	0.33	0.57	0.26	0.32	0.56	1.00	0.62	0.81	0.27	0.70
full_bath	-0.18	0.14	0.13	0.52	0.48	0.47	0.37	0.62	1.00	0.52	0.49	0.54
totrms_abvgrd	-0.08	0.33	0.24	0.38	0.14	0.20	0.39	0.81	0.52	1.00	0.18	0.50
garage_yr_blt	-0.24	0.06	0.00	0.54	0.77	0.65	0.26	0.27	0.49	0.18	1.00	0.51
saleprice	-0.26	0.33	0.30	0.80	0.57	0.55	0.62	0.70	0.54	0.50	0.51	1.00

RESULTS & INSIGHTS

MODEL 1

The model does not capture the skewness in the data therefore the kaggle submission comes out lower than expected.

MODEL 2

This model was log–transformed and it allowed for the capture of the skewness and factored into the price estimation.

Actual vs Predicted Sale Prices
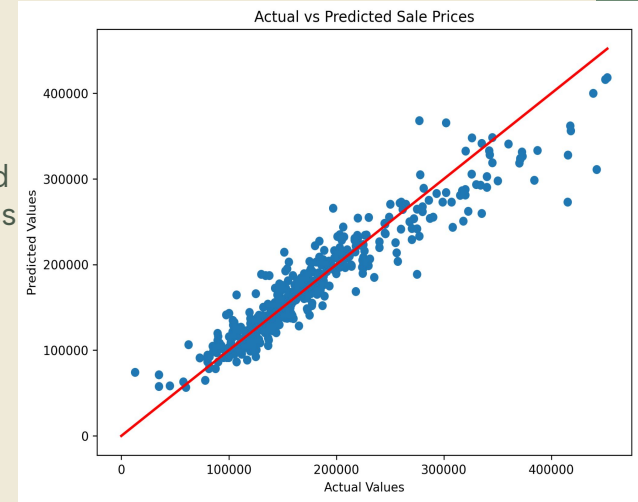
MODEL	R-SQUARED SCORE	R-SQUARED SCORE (1)
Linear Regression	86.04%	89.08% (Chosen Model)

CONCLUSIONS

- This model provides a more accurate way to assess the determination of the final sale price, helping investors or sellers reduce financial losses and make smarter decisions.

NEXT STEPS

- Add more hyperparameters
- Develop customer-facing tools for users to interact with
- Test the model on new date were the sale price is not included and check the predicted to the actual price
- Explore more models

RECOMMENDATIONS

Incorporate economic factors into home final sale price.

Utilize log-transformation to scale the sale price based on the features and then convert it back to the predicted sales price.

Deploy the model to companies or create a interactive website so investors or real estate companies can get a final sale price.

THANKS

Do you have any questions?

LinkedIn
www.linkedin.com/in/damar-shipp-jr-614b71186

Portfolio
https://github.com/DamarTheMunginizer