AMES, IOWA HOUSING MARKET

JEN ZIEGER

AMES, IOWA HIGH-LEVEL STATS

- Ames is located about 30 miles north of Des Moines in central Iowa
- Home to Iowa State University
- According to the US Census Bureau:
  - Population in 2020: 66,427
  - Households, 2015-2019: 25K+
  - Owner-occupied housing unit rate, 2015-2019: 40.7%

As a new member of a residential real estate firm, I've been tasked with building a model that can be used in-house to determine what features and model best predict housing prices and potentially build out the model to sell to other companies across the country.

For the beta version of the model, I will use a dataset containing assessed values for individual residential properties sold in Ames, IA, from 2006 to 2010.

PROBLEM STATEMENT

OVERVIEW

Data Cleaning & EDA

- Replacement or removal of null values
- Correlation evaluation for features

Feature Engineering

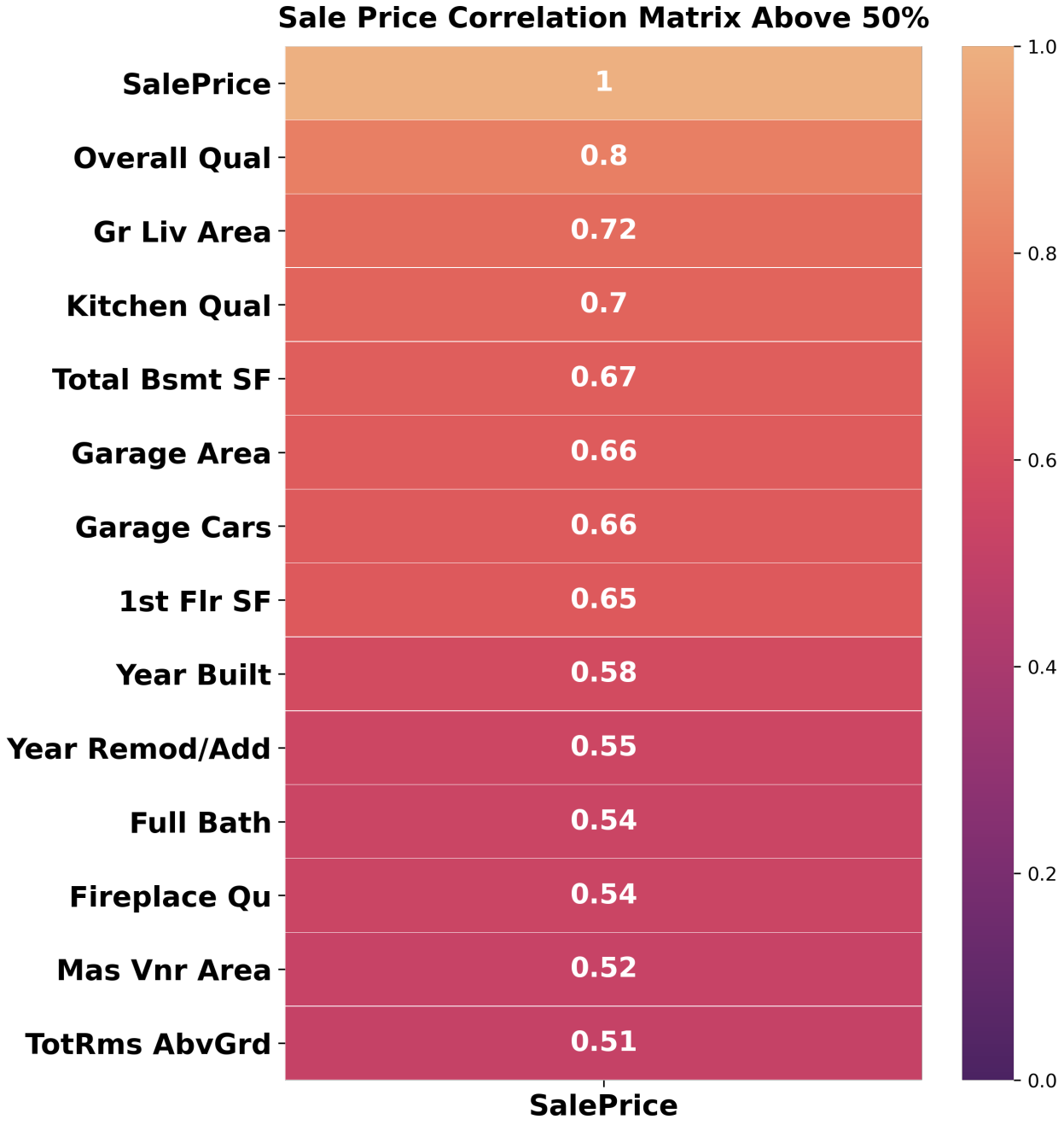
- One-Hot Encoding | Ordinal Encoding | Polynomial Features

Regression Techniques

- Linear | Ridge | Lasso

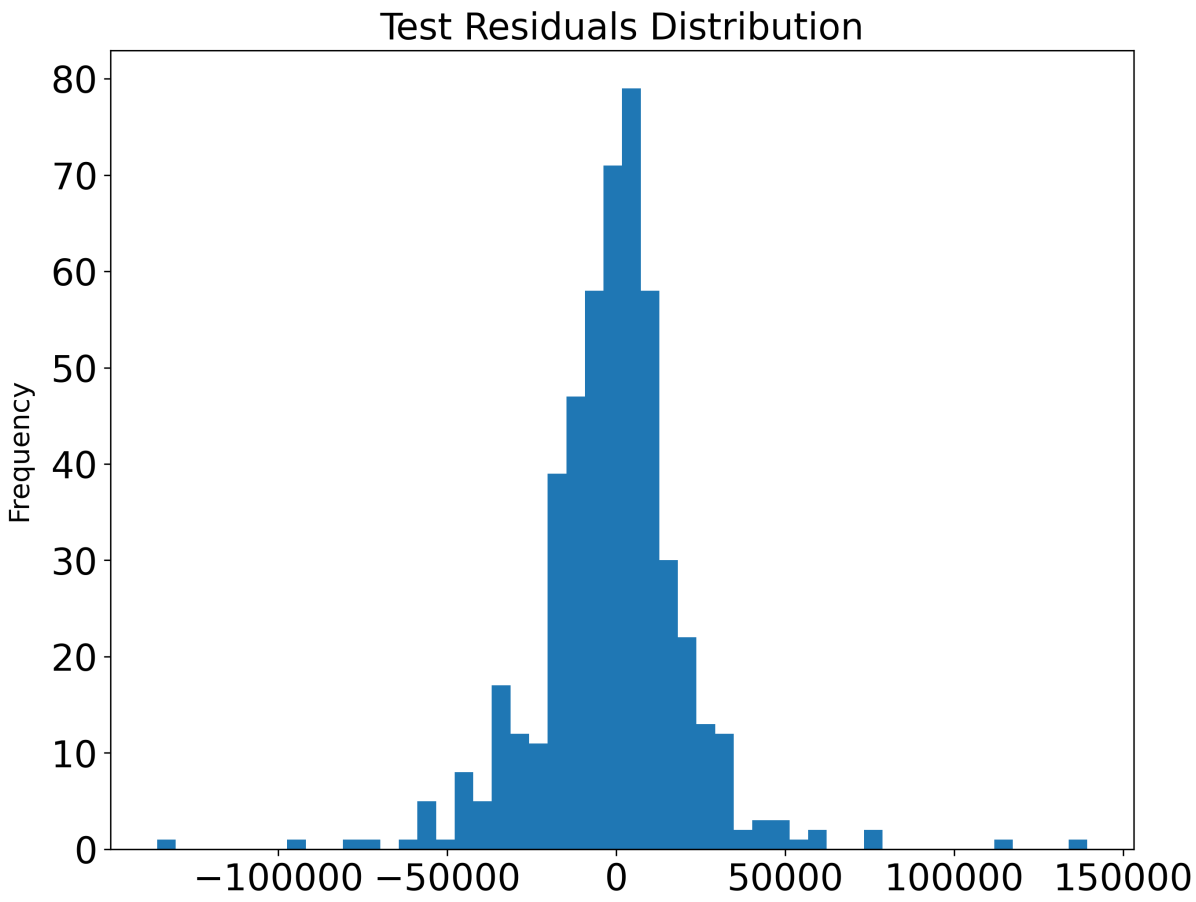
Evaluation

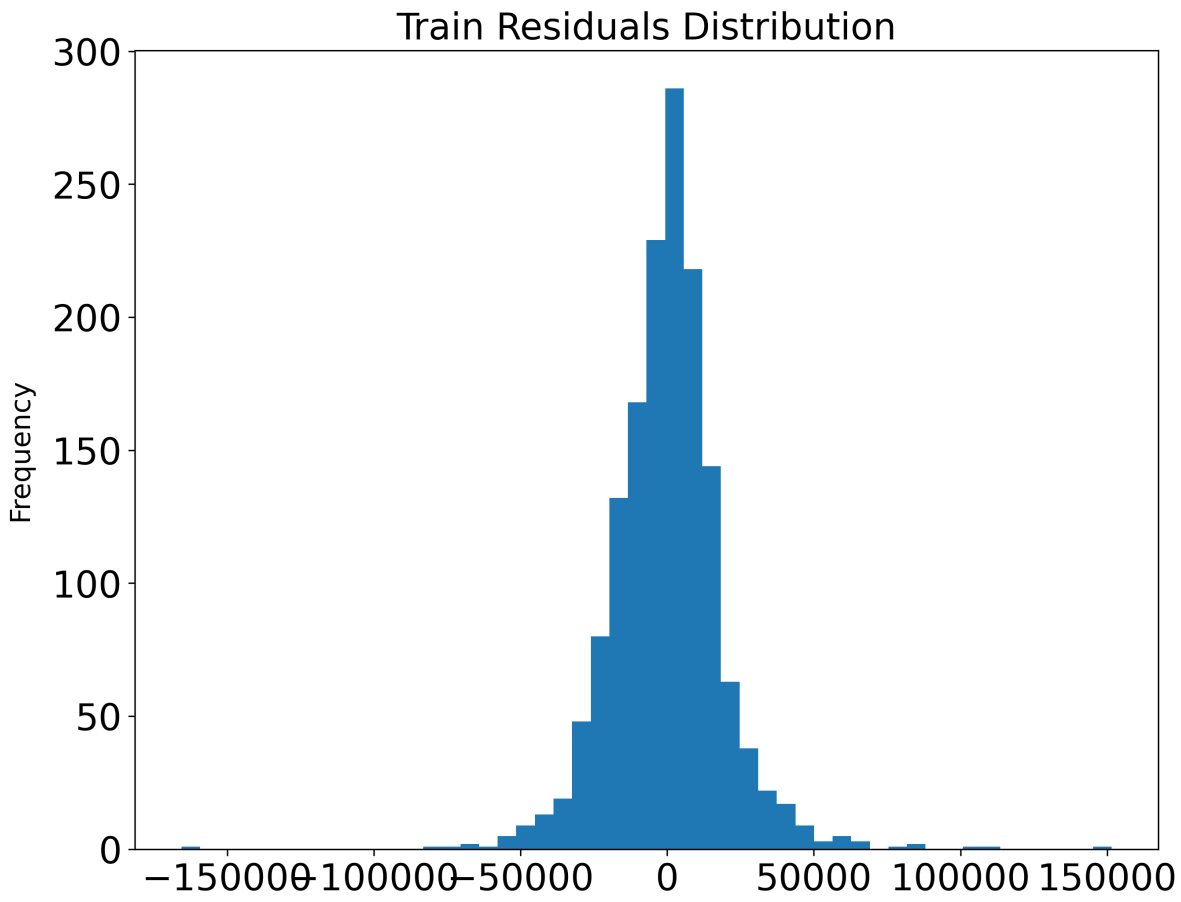
- R Squared Score | RMSE

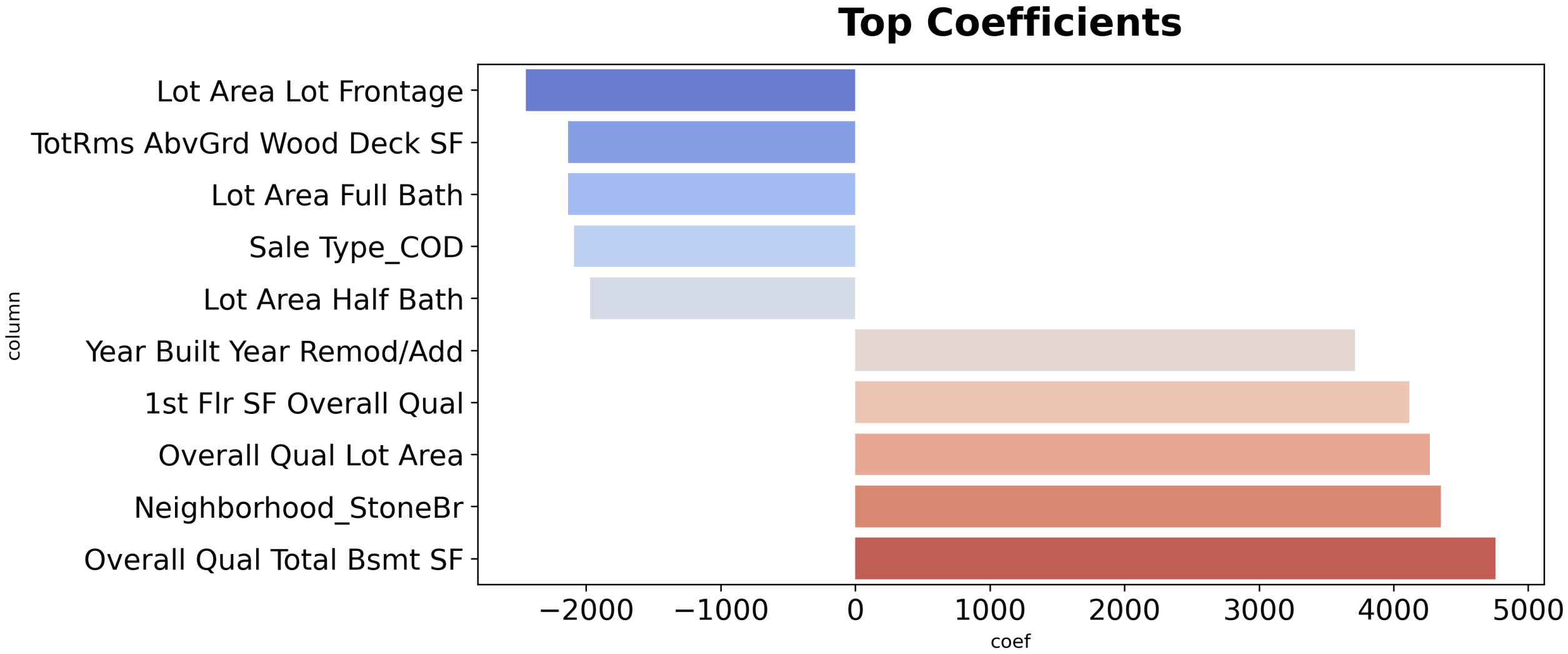

Sale Price Correlation Matrix Above 50%
SalePrice
Overall Qual
Gr Liv Area
Kitchen Qual
Total Bsmt SF
Garage Area
Garage Cars
1st Flr SF
Year Built
Year Remod/Add
Full Bath
Fireplace Qu
Mas Vnr Area
TotRms AbvGrd
1
0.8
0.72
0.7
0.67
0.66
0.66
0.65
0.58
0.55
0.54
0.54
0.52
0.51
SalePrice
1.0
0.8
0.6
0.4
0.2
0.0

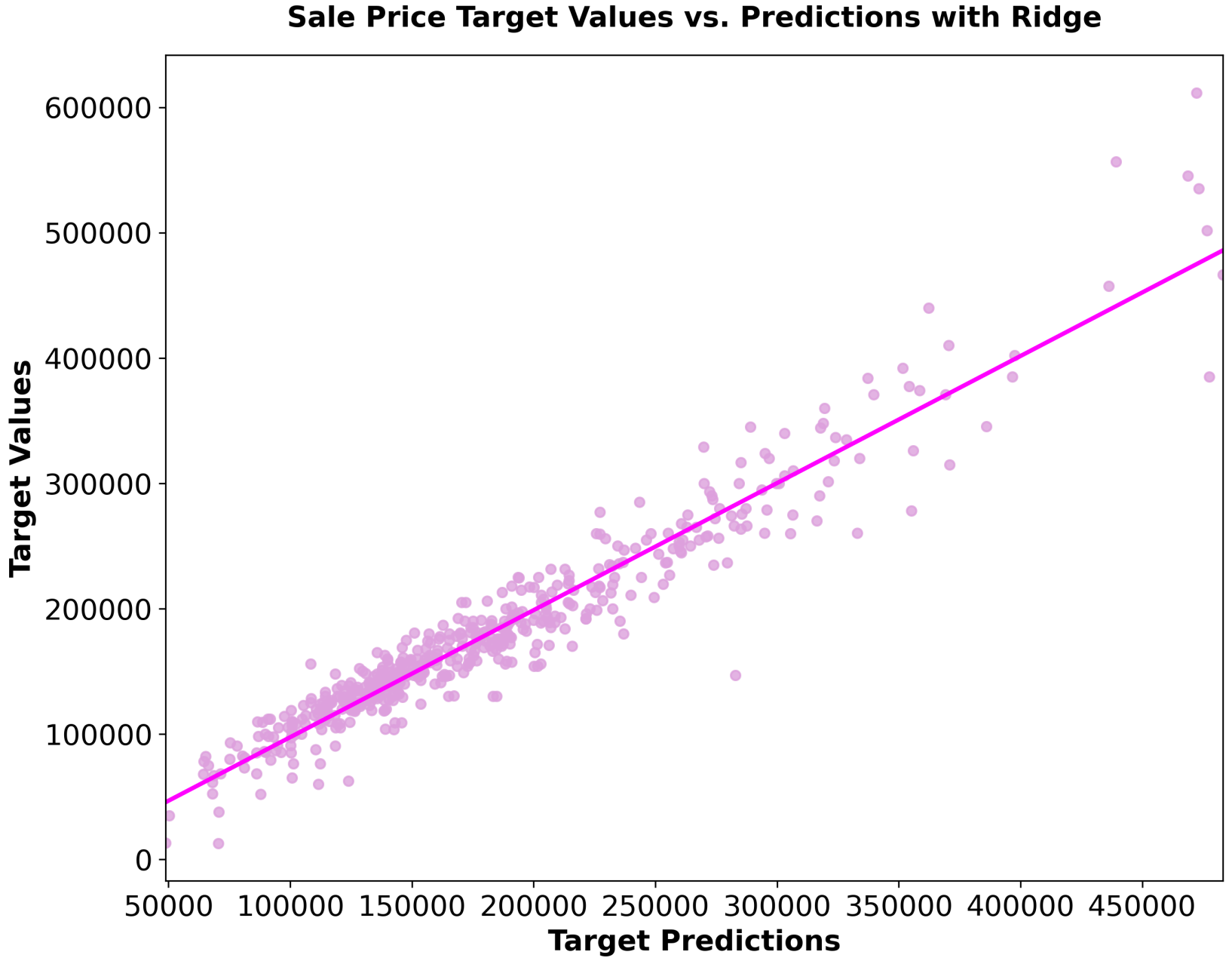
EVALUATION

Model	Train R^2 Score	Test R^2 Score	Train RMSE	Test RMSE
Baseline	0	0	79,480.44	78,777.56
Linear Regression	96.28%	89.11%	15,229.96	26,382.46
Ridge Regression - CV	94.26%	92.04%	18,923.95	22,551.22
Ridge Regression - Grid	94.11%	92.05%	19,163,64	22,536.36
Lasso	94.98%	91.06%	17,687.08	23,898.93


Test Residuals Distribution
Frequency
0
10
20
30
40
50
60
70
80
−100000
−50000
0
50000
100000
150000


Train Residuals Distribution
Frequency
0
50
100
150
200
250
300
−150000
−100000
−50000
0
50000
100000
150000


Top Coefficients
Lot Area Lot Frontage
TotRms AbvGrd Wood Deck SF
Lot Area Full Bath
Sale Type_COD
Lot Area Half Bath
Year Built Year Remod/Add
1st Flr SF Overall Qual
Overall Qual Lot Area
Neighborhood_StoneBr
Overall Qual Total Bsmt SF
column
−2000
−1000
0
1000
2000
3000
4000
5000
coef


Sale Price Target Values vs. Predictions with Ridge
Target Values
Target Predictions
0
100000
200000
300000
400000
500000
600000
50000
100000
150000
200000
250000
300000
350000
400000
450000

HOW THIS MODEL COULD BE USED

Prediction:

- This model can be used to predict housing prices in areas such as Ames

- Online real estate companies - to predict housing prices

Inference:

-Sell to housing developers who can use this model to decide where to invest upgrades and features when they build new houses

- Sell to rental owners can use this model to charge for damage and maintain the value of their homes

Best Model

- Ridge Regression
- Predicted 92% of the testing housing Sale Prices with a $22,536 RMSE

Recommendation

- Try other models and regularization methods
- Reevaluate features
- Gather more current data
- Try using the model in other housing markets

CONCLUSION AND RECOMMENDATIONS

THANK YOU

SOURCES

- Population and Housing Stats: https://www.census.gov/quickfacts/fact/table/amescityiowa/POP010220#POP010220
- Ames Data Documentation: http://jse.amstat.org/v19n3/decock/DataDocumentation.txt