Ames Housing Prices

Bobble Bee's investment fund:

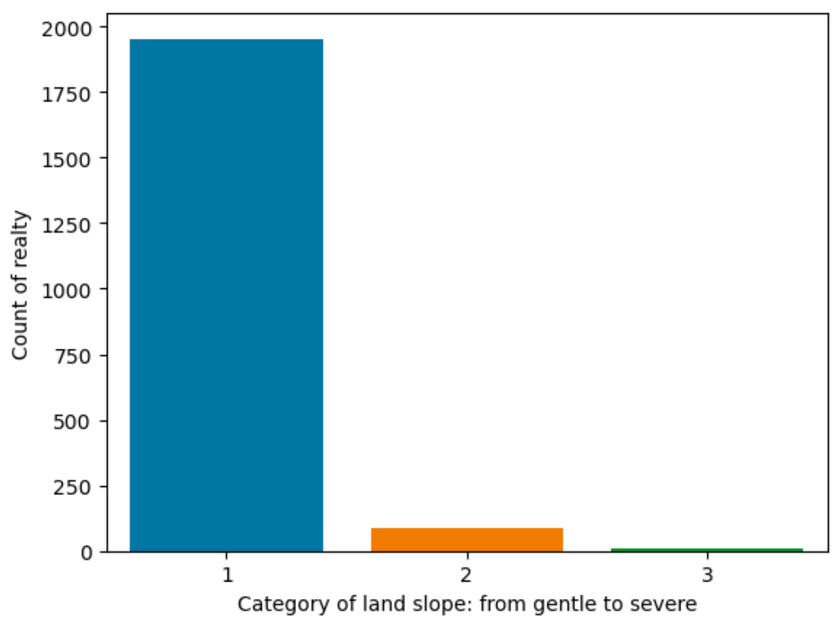
Maximize their profit investing in Ames housing

How?

Selecting realty that are predicted to be worth more than the actual price

Filtering variables

1. Variables that don't show variability
Ex: 95% of data fit into one category



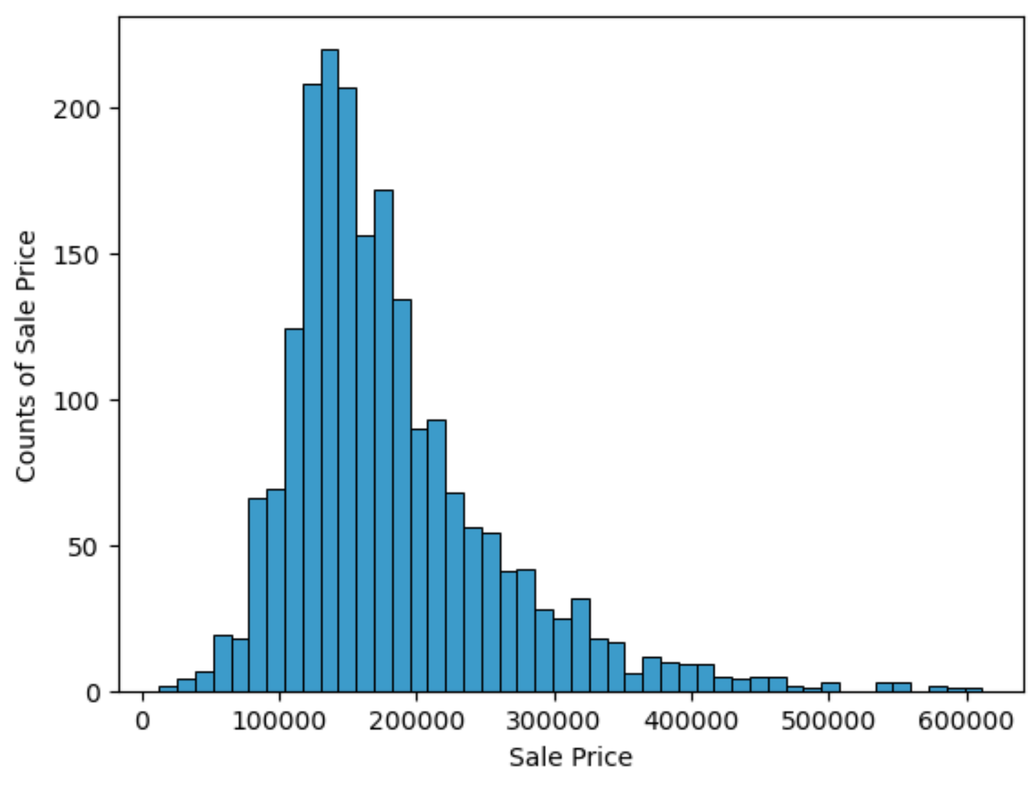
2. Underrepresentative variables
Variables that are present in only 20% or less of data

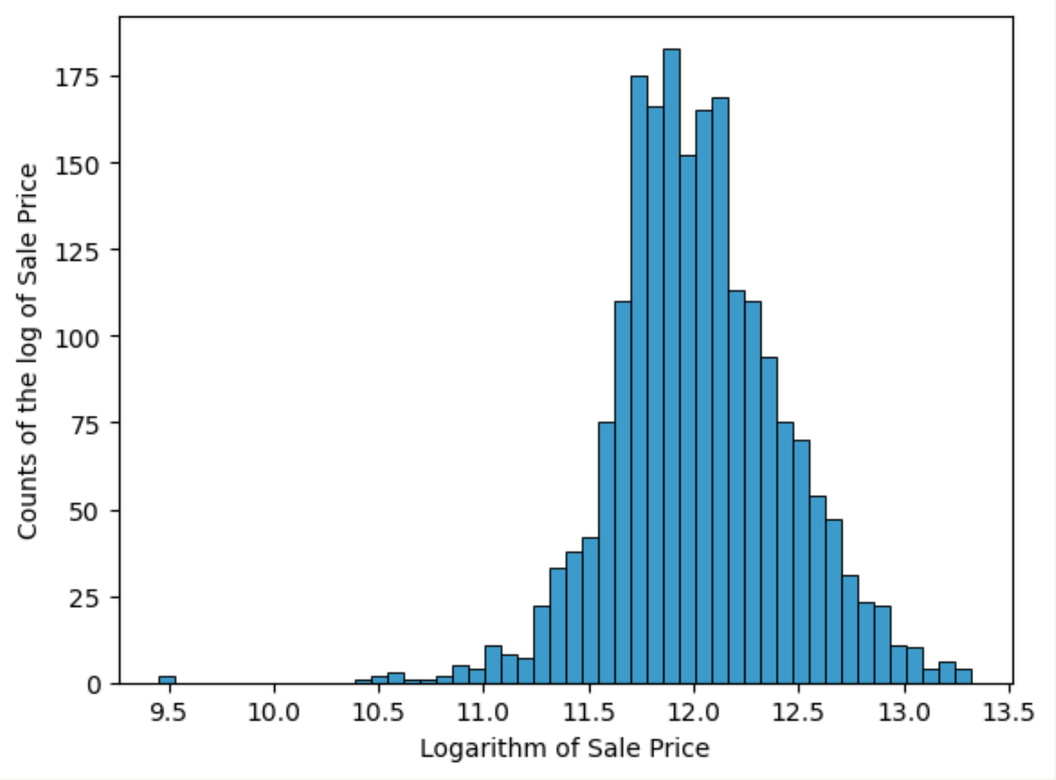
Variables and non-nulls	
Alley	140
Pool QC	9
Misc Feature	65
Fence	400

R2 Score of 88.5%

Transforming the target variable

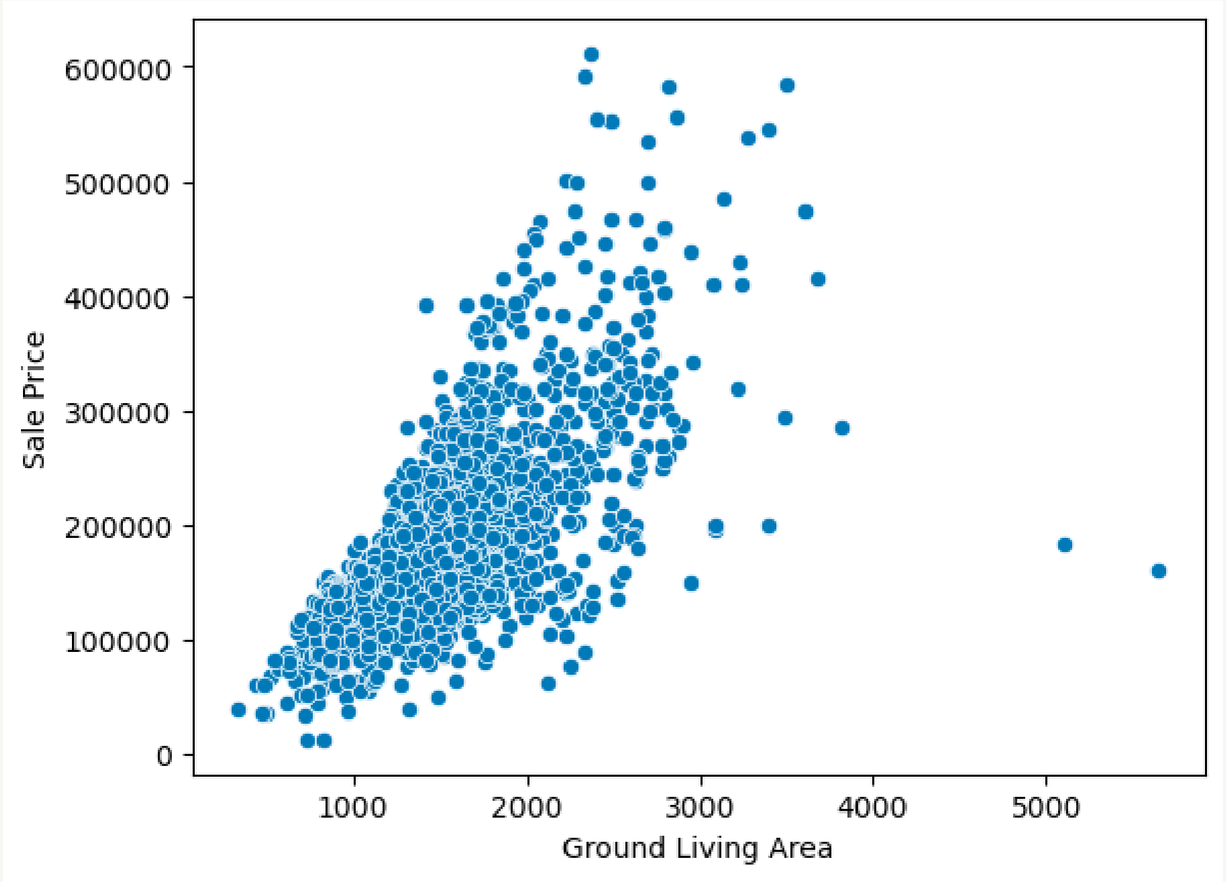
- Logarithm of sale price
- More Normal distribution
- Improved linear regression metrics: R2 score went from 88.5% to 90.3%





Removing outliers

- Data points with ground living area greater than 40000 square feet
- R2 score went from 90.3% to 92.3%



Regularization - Ridge Regression

Although filtering data to avoid overfitting as much as possible,
Regularization was also used to reduce possible overfitting

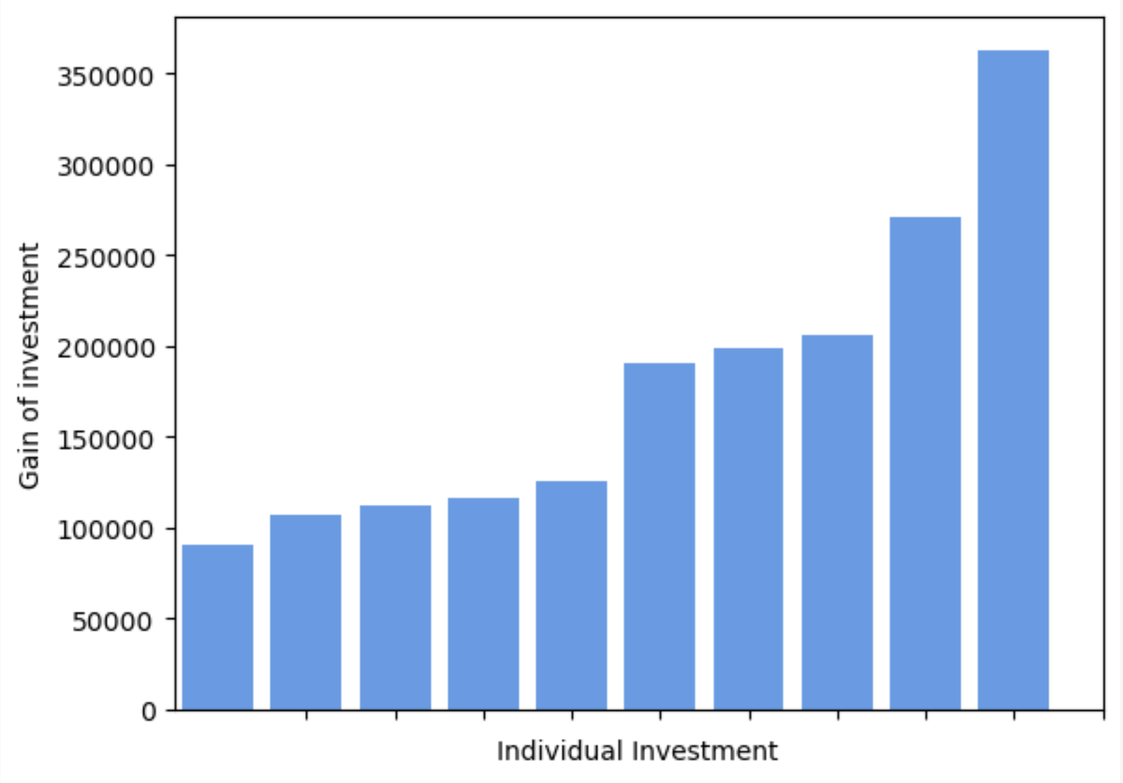
Ridge regression generated an R2 of 91.2% in the test set

Conclusion and Recommendations

(Value Predicted) - (Actual Value)

- There are 2 opportunities of gains of approximately $250,000 or more each

- There are 9 opportunities of gains of approximately $100,000 or more each



Any questions?