Ames Housing Dataset

Elliot Lee

01 **Problem Statement**
Value of our analysis
How to evaluate success?

02 **Methodology**
Steps taken to create our model

03 **Key Takeaway and Findings**
Insights from our model

04 **Conclusion**
Recommendations and future steps

01

Problem Statement

Value of our analysis
How to evaluate success?

Problem Statement

Develop Price Prediction Model

Allow homeowners, homeseekers and real estate agents to estimate sale value of properties

Lasso Regression Model

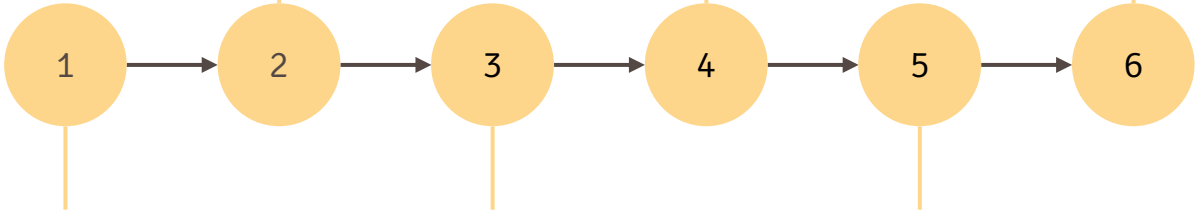
Accuracy of model evaluated based on **Root Mean Squared Error** metric

02

Methodology

Steps taken to create our model

Methodology



Exploratory Data Analysis
of our training dataset

81 columns,
2051 rows

Target variable:
SalePrice

Data cleaning

Identification of outliers;
imputation of null values

Feature Engineering

Create new features and interaction terms

Encoding of categorical variables

Feature selection
through data visualization and filtering

Modelling

Linear Regression
Ridge
Lasso

Model iteration and RFE

Model interpretation and evaluation

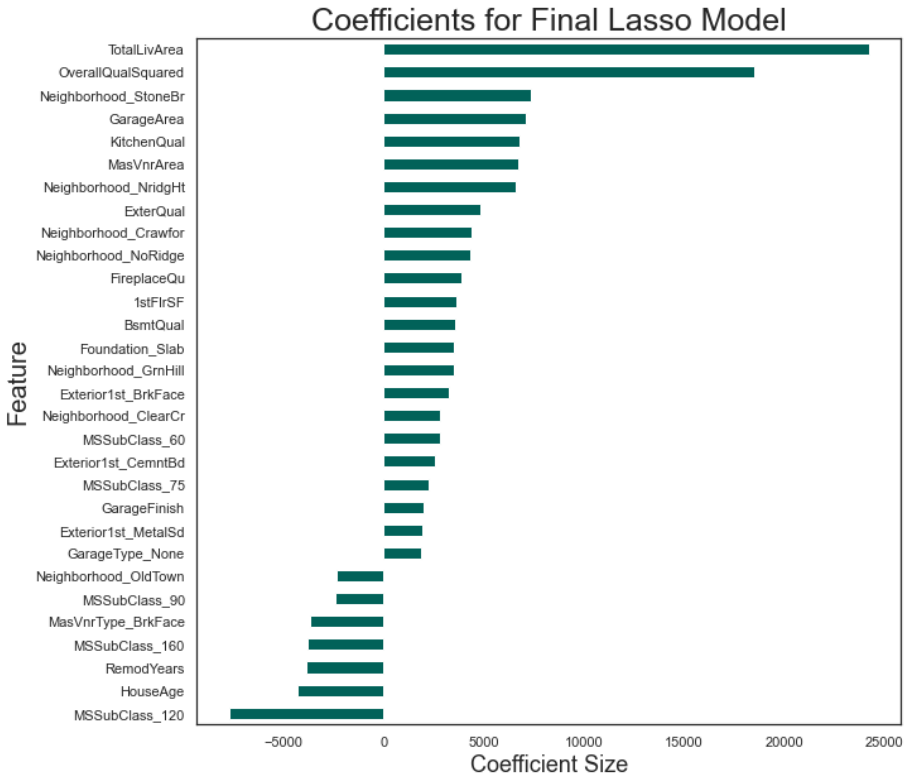
RMSE, interpretation of predictor features

03

Key Takeaways and Findings

Insights from our model

Key Takeaways and Findings



Lasso Model, 30 features

R2 Score

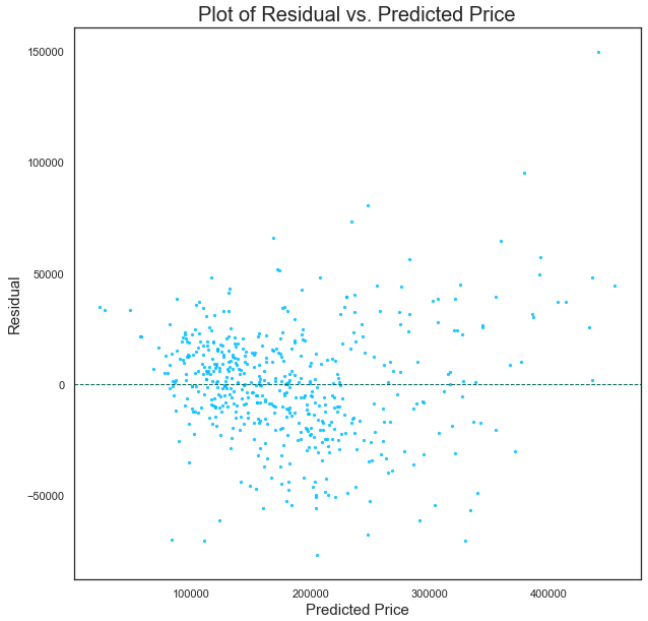
Model Train Score: 0.8876
Model Test Score: 0.8986
Model 5-fold CV Score: 0.8758

RMSE

Model Train Score: 26548.4062
Model Test Score: 25317.1156
Model 5-fold CV Score: 27776.0085

TotalLivArea and **OverallQualSquared** show the greatest predictive strength

Key Takeaways and Findings



Model performs reasonably well – majority of predicted prices **roughly coincide with actual prices**

However, tends to **undervalue houses > 350k**

Residual plot shows that **errors are not homoscedastic**

Residuals are more **sparsely distributed** as price increases

Key Takeaways and Findings

Prediction

Used to predict the SalePrice of a new house

Inference

Infer the relative effect of each predictor variable on SalePrice in terms of magnitude and direction

04

Conclusion

Recommendations and future steps

Conclusions

Features adding value to a home include the **total living area**, **overall quality and finish** and **neighbourhood** that the house is located in

Features lowering value include **brick masonry veneers,** as well as the **age of the house** and **time since last remodelled**

Lasso Regression Model with **30 features** obtained a RMSE score of **25371.1156**

Linear Regression, while a simplistic ML algorithm allows us greater interpretability of predictive features including the relative magnitude and direction

Further improvements to our model can involve gathering more information on Ames housing; using other machine learning algorithms to perform Recursive Feature Elimination; and performing a logarithmic transformation of our target variable to mitigate the effects of skew