Project 2: Predicting Housing Prices In Ames, Iowa

Unravel the merits of unrefined data and generate accurate predictions of housing prices based on our data-driven model

By: Chee Howe, Ben, Yuan, Yun Jie

Problem Statement

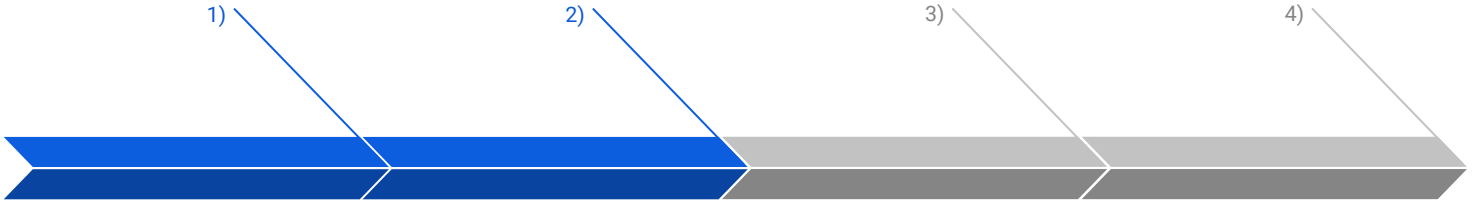
What are we searching for?

With a wide range of characteristics of houses collected, we aim to provide aid to Housing Developers by building a model which analyzes, sieves and extracts important features that would provide accurate predictions of housing prices

How do we do that?

Through the use of a regularized linear regression model, **Lasso**, and metrics such as R2 and root mean square error (RMSE) to validate it

Methodology



Business Understanding

Goal: Optimize a model that provides accurate house prices to Housing Developers in Iowa

Data Analysis

a. Data Visualization:
> Heatmap, Scatterplots
To study **correlations** between cost of houses and features
b. Data Cleaning:
> Removal of redundant features
> Imputation

Feature Selection and Model Optimization

Identifying the **top 4** features that strongly influences our output and feeding them into our model

Conclusion

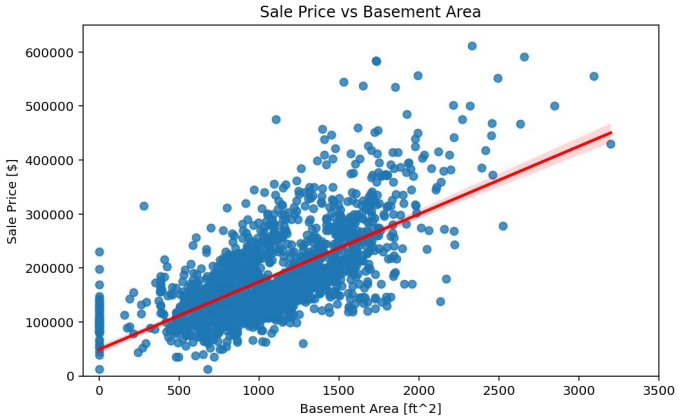
Suggestions on the use of this model and recommendations to improve our model

1) Ground Living Area

1) Ground living area has the highest impact on sale prices.
2) Using the model, price/ft^2 is ~$36.

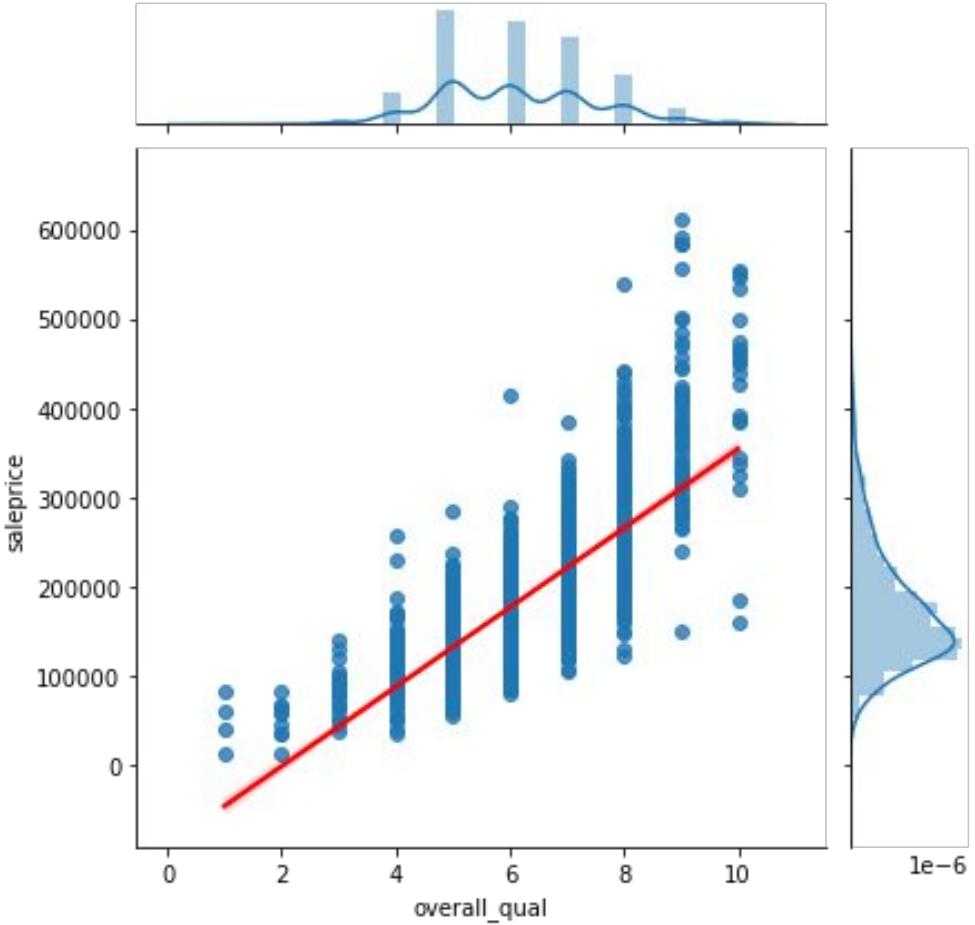
2) Basement Improvements

1) Unfinished basements have no impact on sale price!
2) Finished basement does!
3) Using the model, price/ft^2 increases from ~$0 to ~$16.



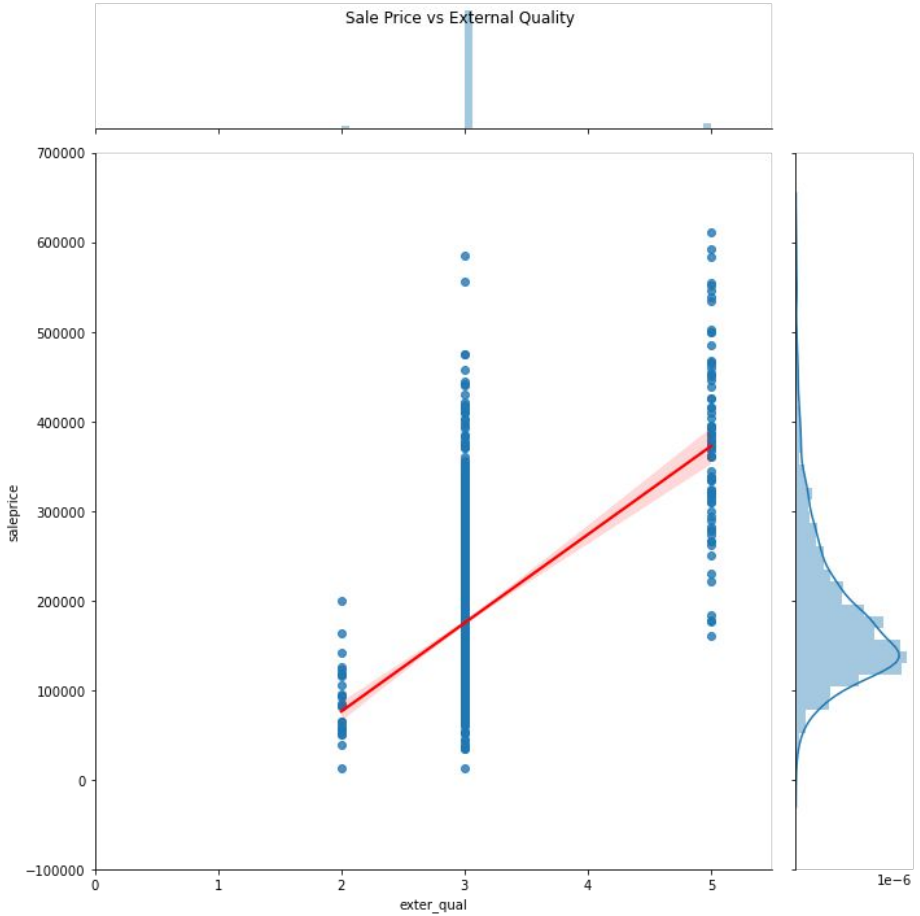
3) Overall Quality

- **Positive** correlation between Sale Price and overall quality of the houses
- Strength of correlation: 0.73
- Strong contender in predicting housing prices

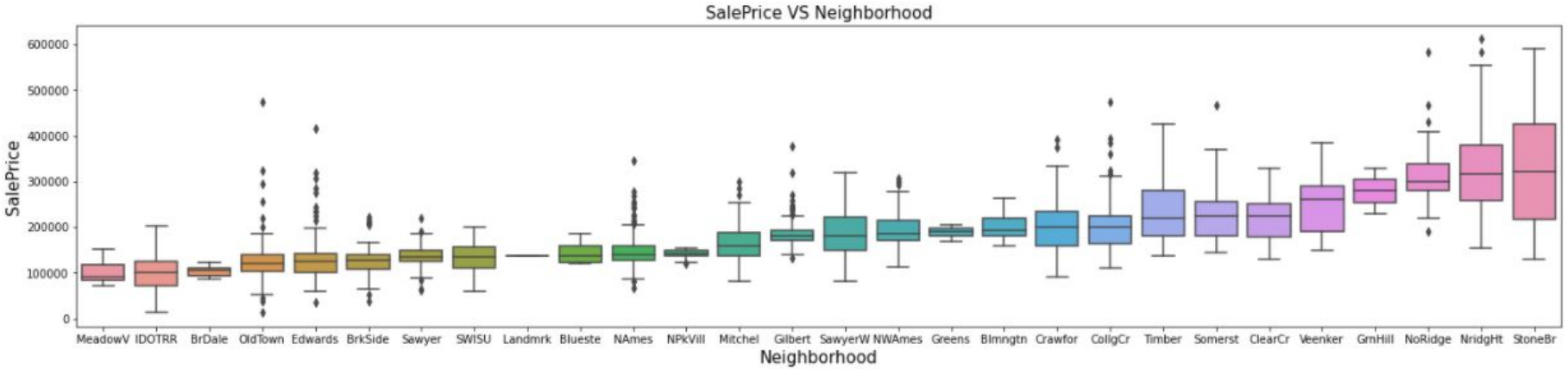


4) External Quality

- **Positive** correlation between Sale Price and overall quality of the houses
- Strength of correlation: 0.503
- No observations at 4.0, 1.0: room for improvement



4) Neighborhoods



- **Positive** correlation between Sale Price and overall quality of the houses

- Strength of correlation: 0.73

Conclusion

Versatility of this model for different consumers

1) Developers could prioritise these features to increase sale price:
    a) Increase ground living area
    b) Furnish the basement to living standards
    c) Ensure quality material and good workmanship
2) Developers can leverage on good neighborhoods in marketing effort

Suggestions to improve this model

1) To update model constantly
2) Strict guidelines during data gathering for ordinal data