# iPredict

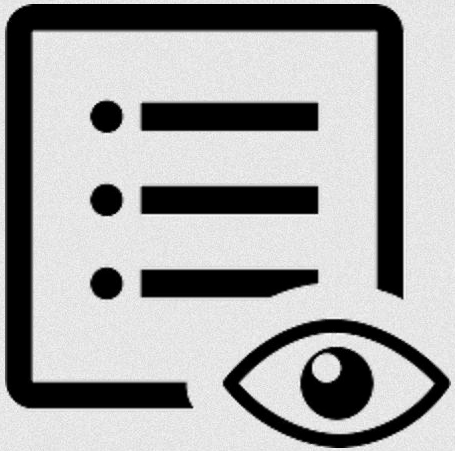
*by MyProperty*

**Your one-stop portal to predict future house prices**

Desmond, Ian, Richelle, Ridzuan

28 Sep 2022

# Table of contents



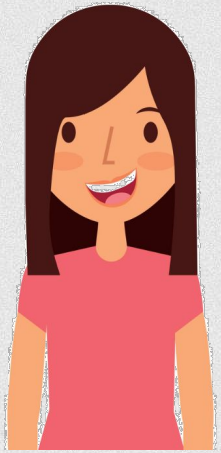
# 01.

# Background

Where and what we are trying to solve

# Employee Personas

I literally spend 4 hours trying to come up with a price but my buyer is still not satisfied..

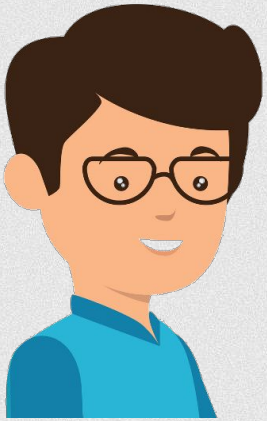


Aileen
2 years experience

All I do is look through the properties listed online, but I'm still lost...

Susan
8 years experience

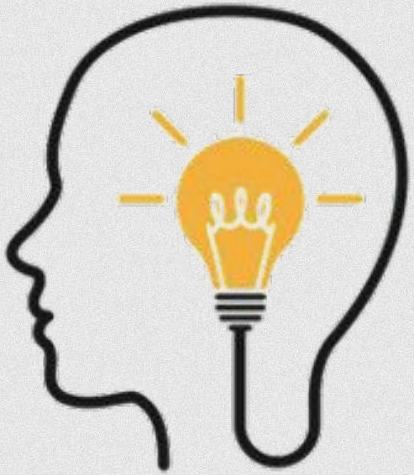
There are just way too many features to consider all the time...



Matthew
10 years experience

# Problem Statement

How to help realtors effectively and efficiently predict the market value of houses in Ames, Iowa?



# Ames Housing Dataset (2006-2010)

## 2930 observations, 82 variables

| Variables | Description | Responses |
|---|---|---|
| Exterior quality | Quality of the material on the exterior | • Excellent<br>• Good<br>• Average/typical<br>• Fair<br>• Poor |
| Exterior condition | Present condition of the material on the exterior | |
| Kitchen quality | Kitchen quality | |
| Basement quality | Height of the basement | • Excellent (100+ inches)<br>• Good (90-99 inches)<br>• Typical (80-89 inches)<br>• Fair (70-70 inches)<br>• Poor (<70 inches)<br>• NA (no basement) |

Source: Ames, Iowa Assessor's Office

# Selected Features

**Ridge model - 16 features selected**
R2: 0.93 RMSE: 21072

| | |
|---|---|
| Overall material and finish quality | Neighborhood |
| Exterior material quality | Overall condition rating |
| Above grade (ground) living area square feet | Lot size in square feet |
| Kitchen quality | Size of garage in car capacity |
| Screen porch area in square feet | Fireplace quality |
| Original construction date | Basement finished area |
| Proximity to main road or railroad | Home functionality rating |
| Total square feet of basement area | Height of basement |

# 02.

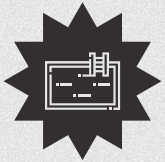
# Data Cleaning

How we cleaned and prepared the data

# Our Cleaning Process

**Missing values**

**Multi-collinearity**
(numeric features)



**80% of the same responses**
(categoric features)

**Small variance**
(categorical variables)

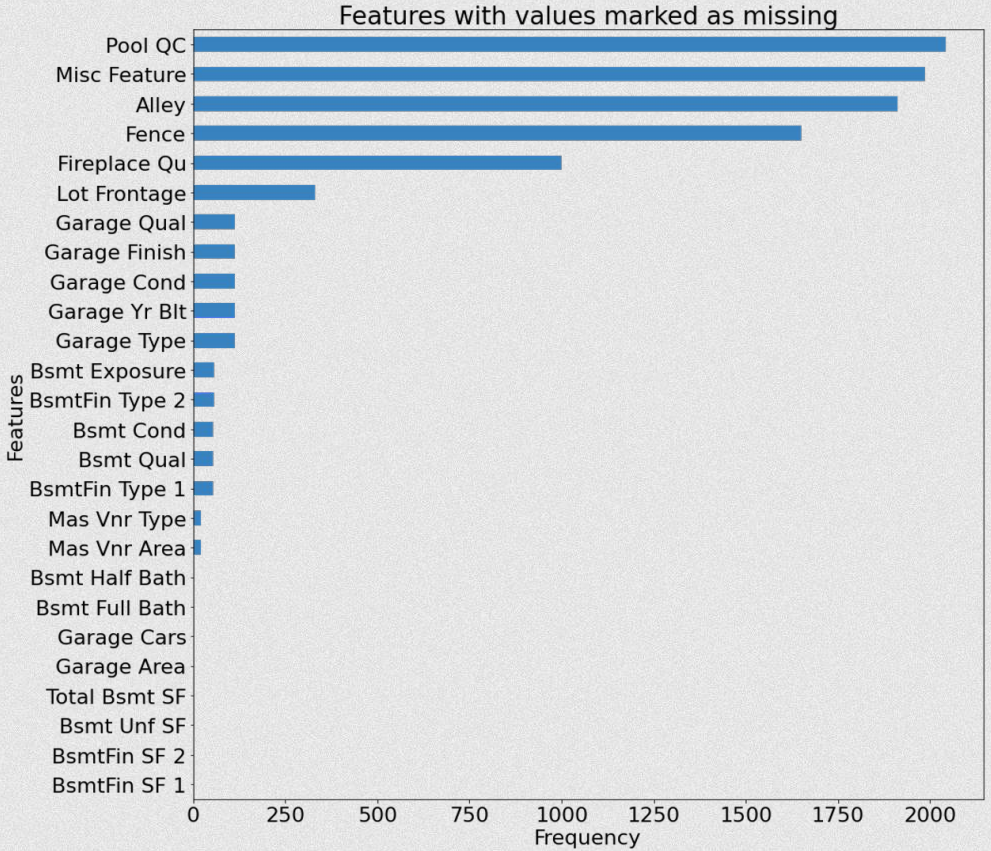
# Columns with Missing Values

**Columns with missing values**

Missing values due to non-existent house features

Solution:

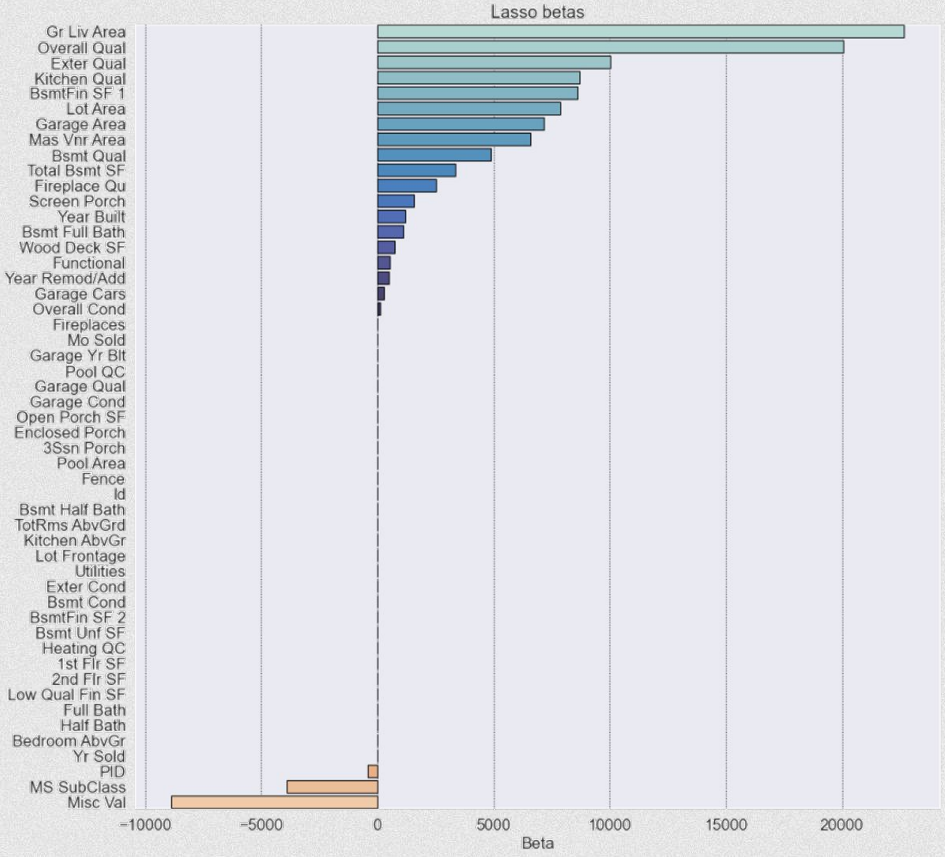
- “0” to replace missing numeric variables
- “None” to replace missing categorical variables
- Mean/mode to replace remaining variables



# Numeric Feature Selection

**Preprocessing features by Lasso**

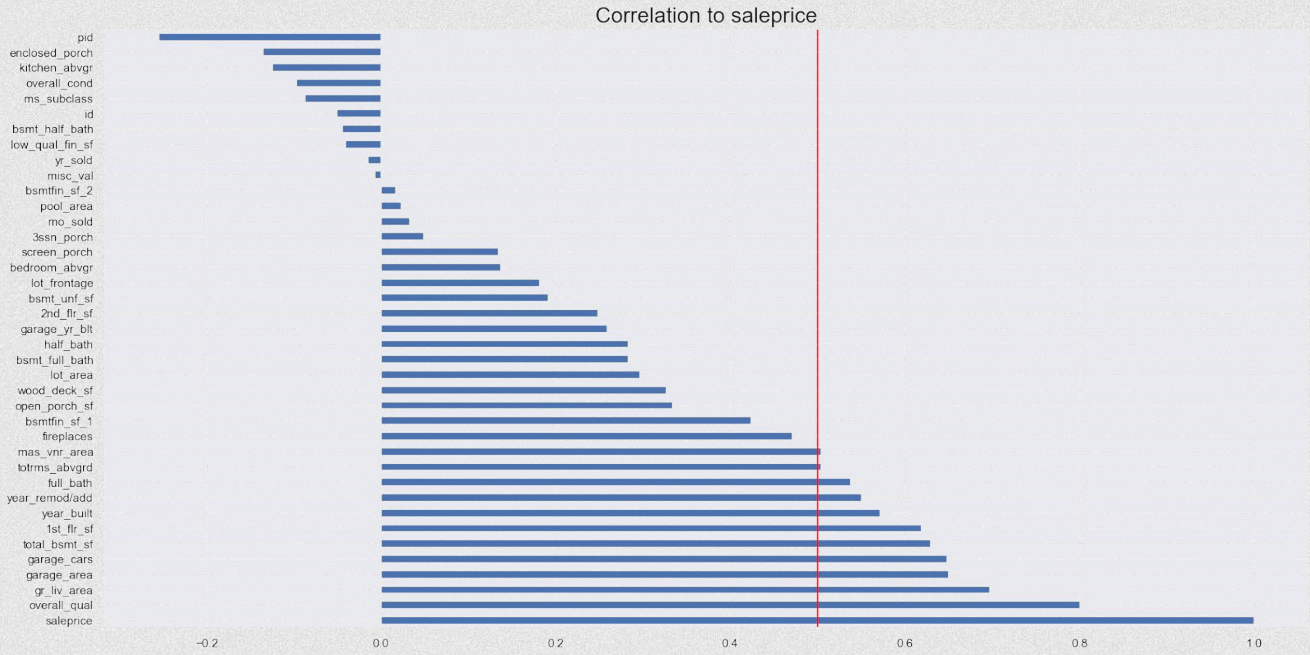
- 29 features were dropped with lasso beta of 0
- These features had no impact to sale price prediction → interfering noise with other model estimators (e.g., Linear Regression or Ridge)



# Numeric Feature Selection

**Comparison with correlation between base numeric features and sale price**

- 11 features with more than 0.5 positive correlation with sale price
- No feature that had above -0.5 negative correlation with sale price



# Numeric Feature Selection

## Final selection

Utilised multicollinearity reduction

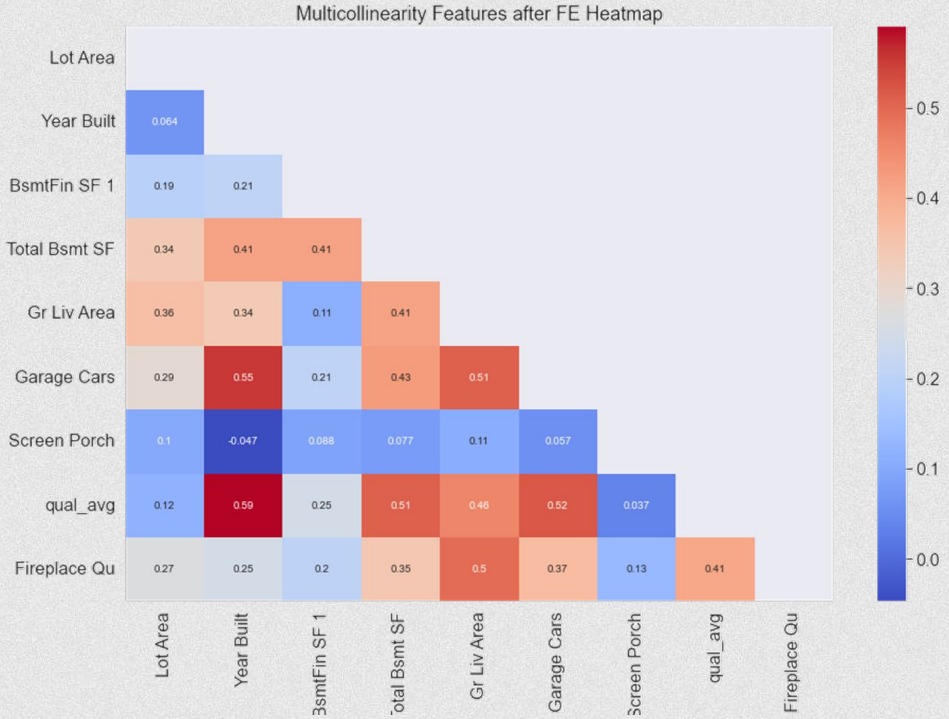
- Lasso coefficient (betas)
- Correlation between features only
- Correlation between features and sale price
- Feature engineering

Above grade (ground) living area square feet (gr_liv_area) was chosen over total rooms above ground as one of the features

- Highest lasso beta
- Above 0.5 correlation with sale price
- High correlation with total rooms above ground

Feature engineering

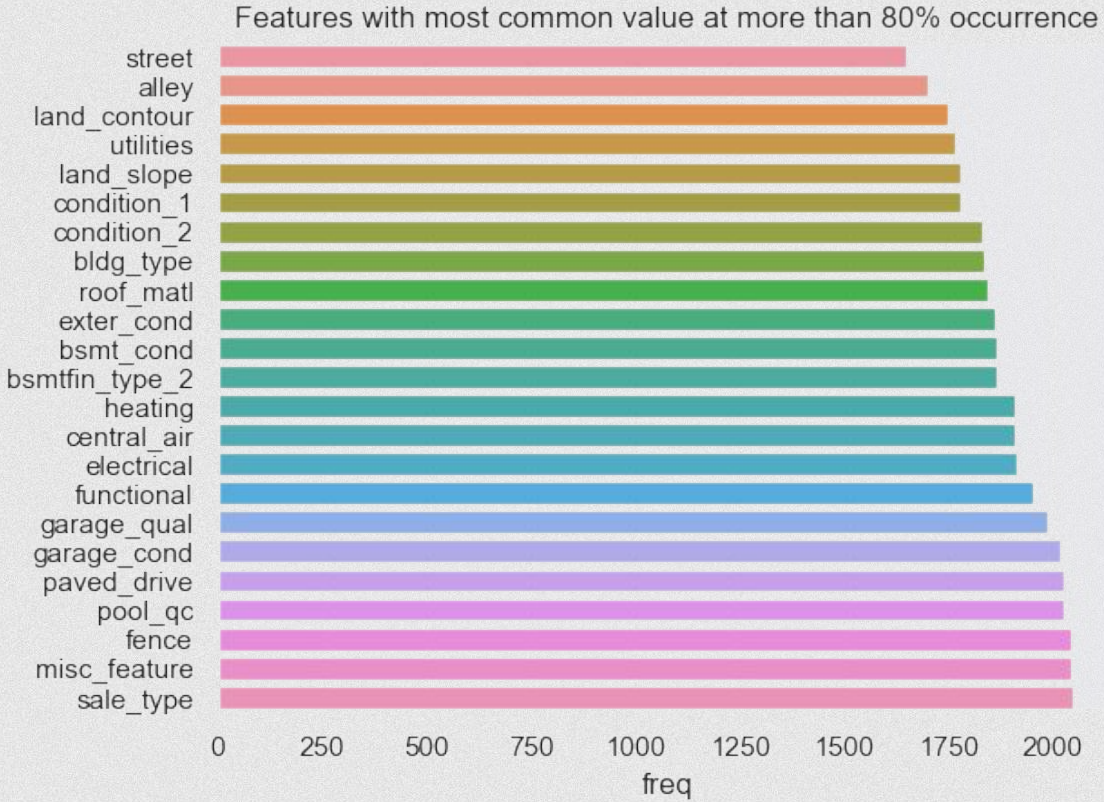
- 6 features to make qual_average



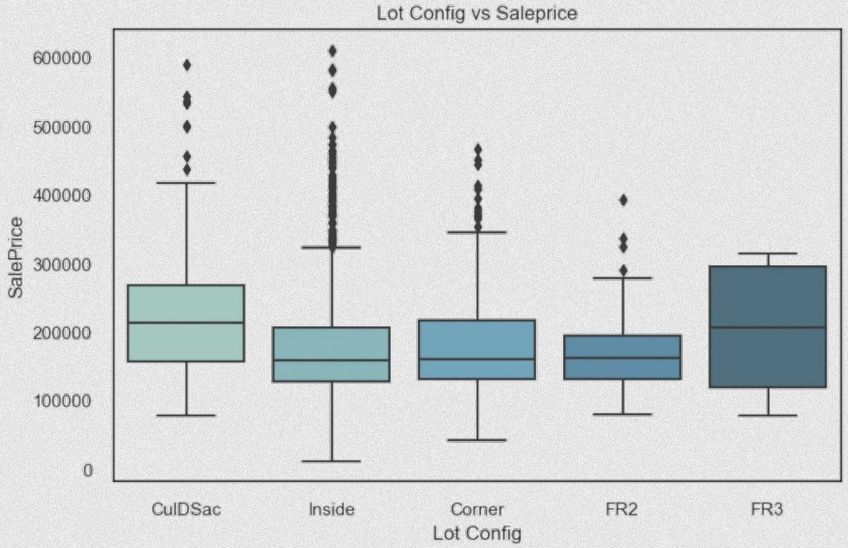
# Categoric Feature Selection

**More than 80% common value occurrence**

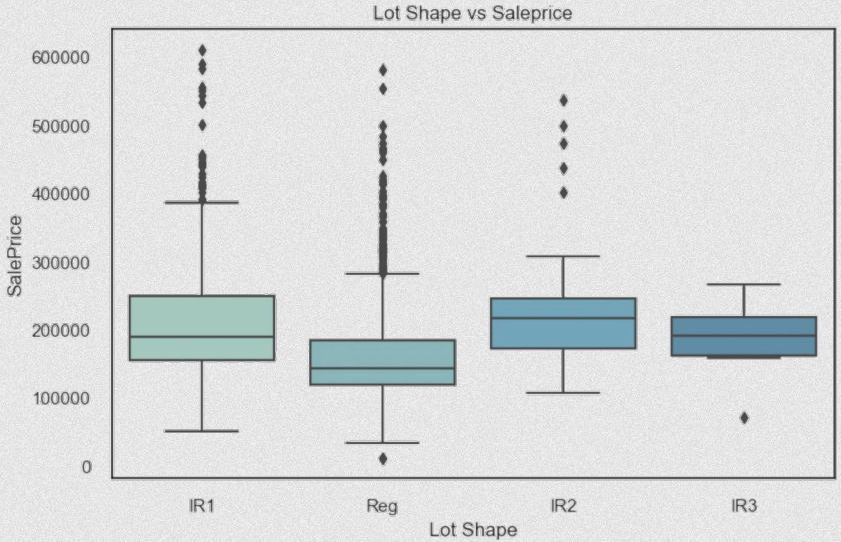
- These features were dropped and not considered for the the subsequent evaluations in the boxplots with sales price



# Small Variance among Variables

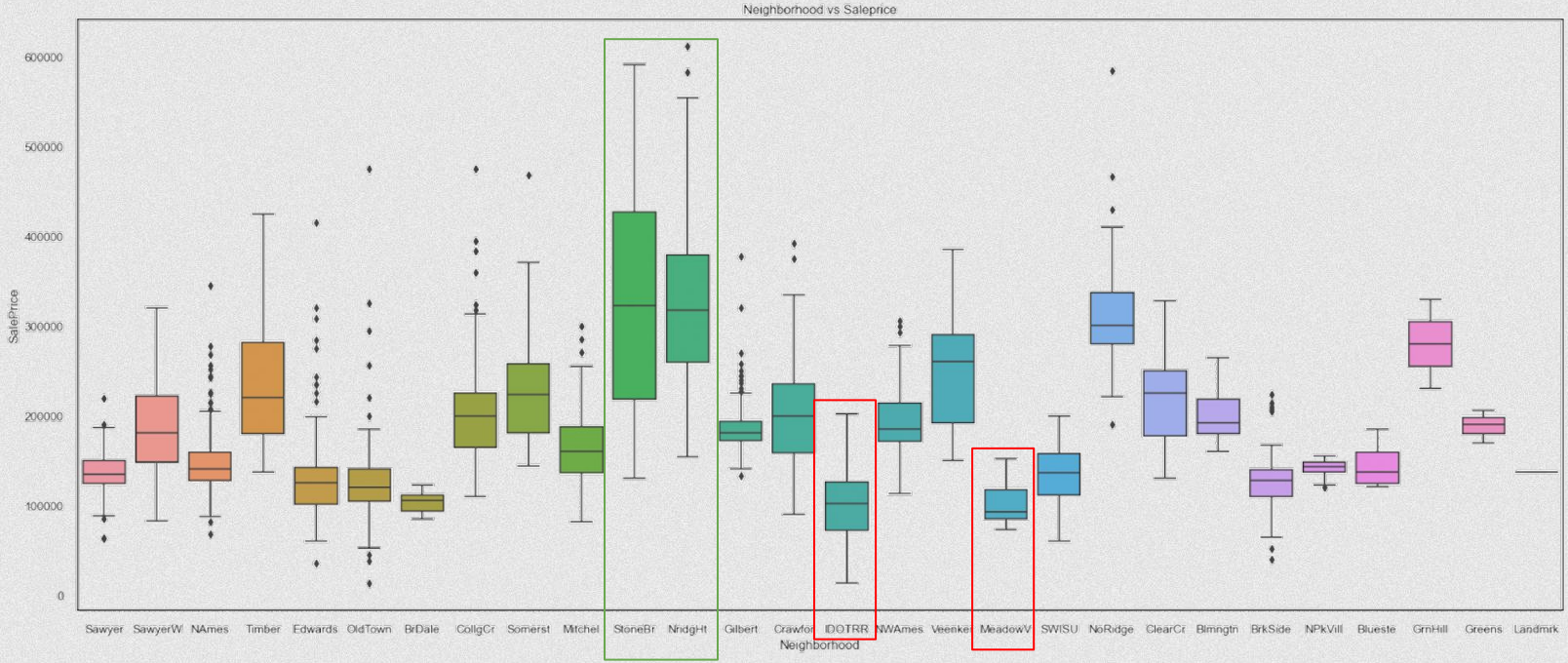



Most outliers: Inside and Corner Lot configurations



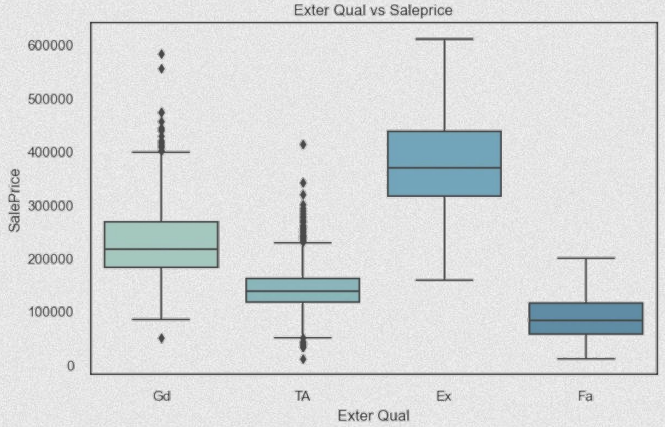

Most outliers: IR1 (slightly irregular) and Reg (regular)

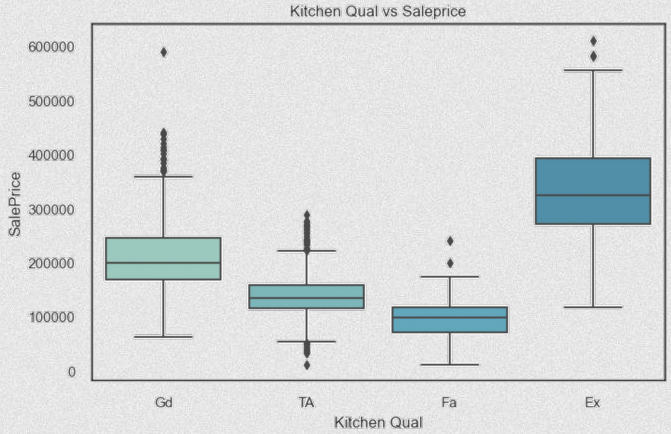
# Boxplot of Neighborhoods with Sale Price

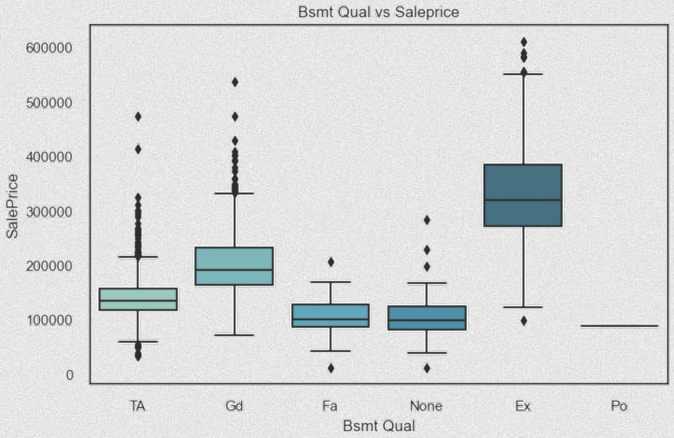


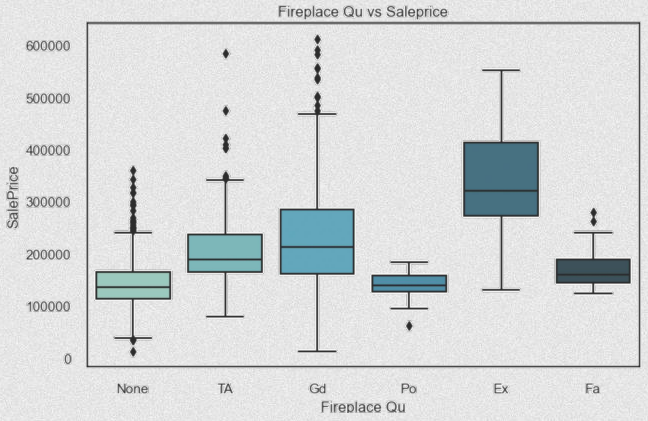

**Neighborhoods with the highest median sale prices: Stone Brook and Northridge Heights**
**Neighborhoods with the lowest median sale prices: Meadow Village and Iowa DOT**

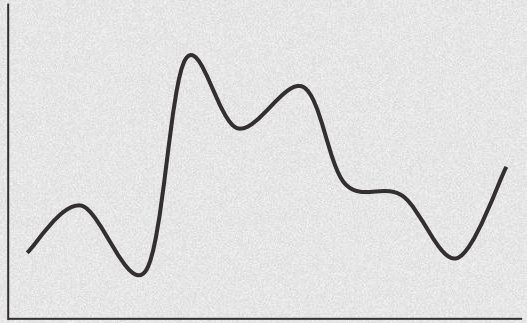
# Boxplots of Various Features with Sale Price



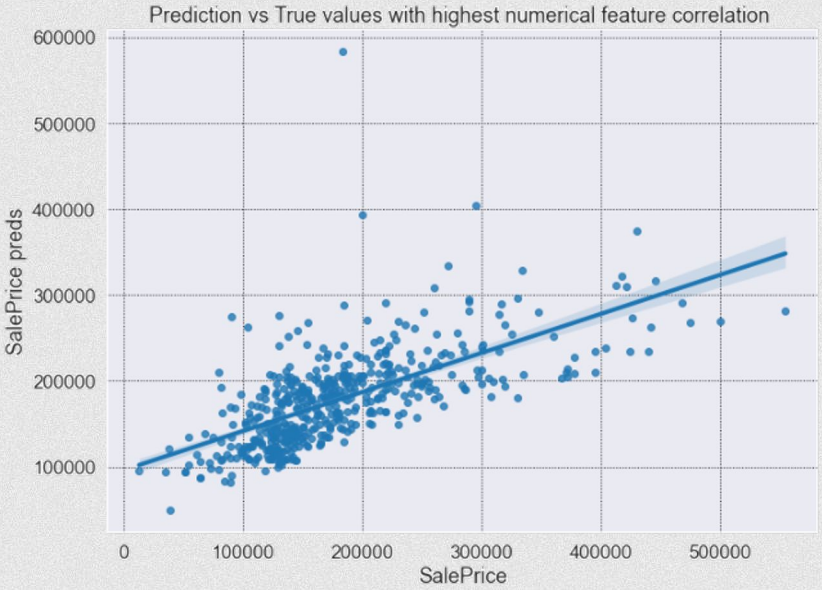


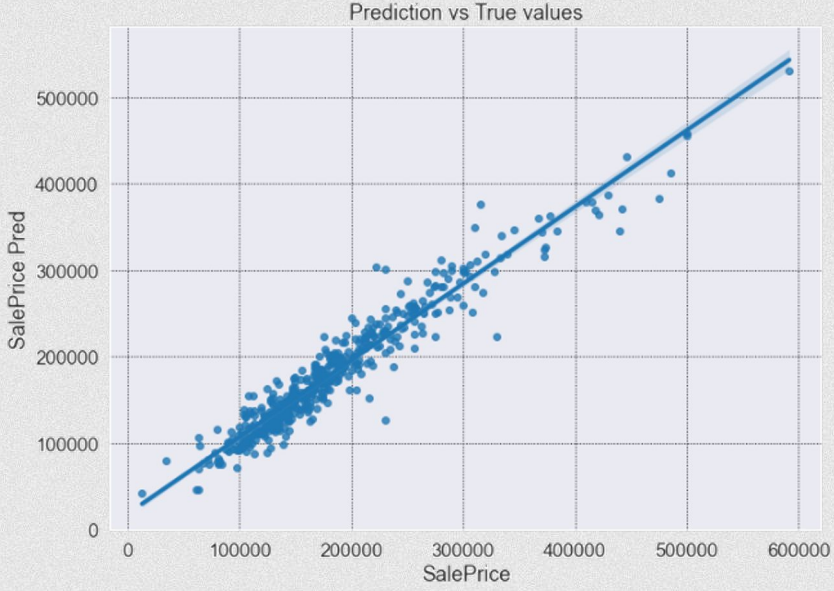
# 03.

# Model

Regression model to predict sale prices

# Baseline Model vs Final Best Model

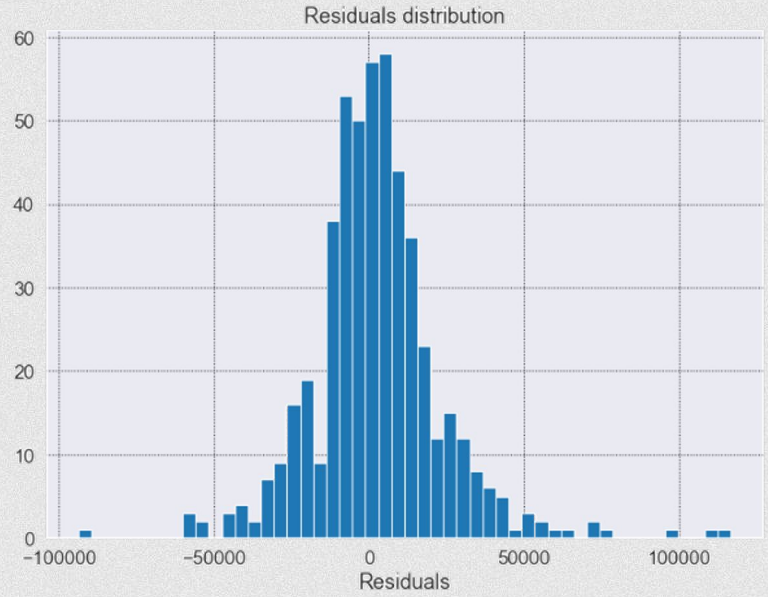


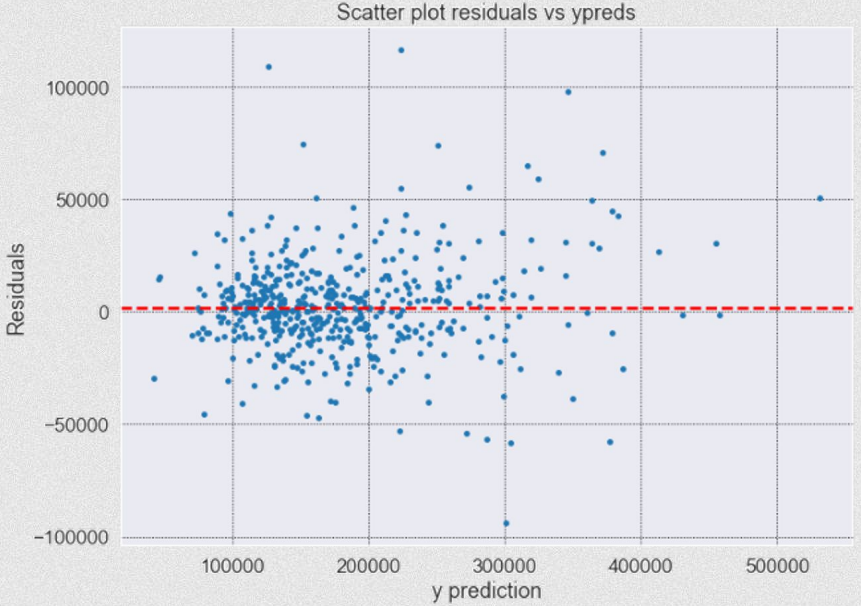




**R2: 0.44**
**RMSE: 59355**

**R2: 0.93**
**RMSE: 21072**

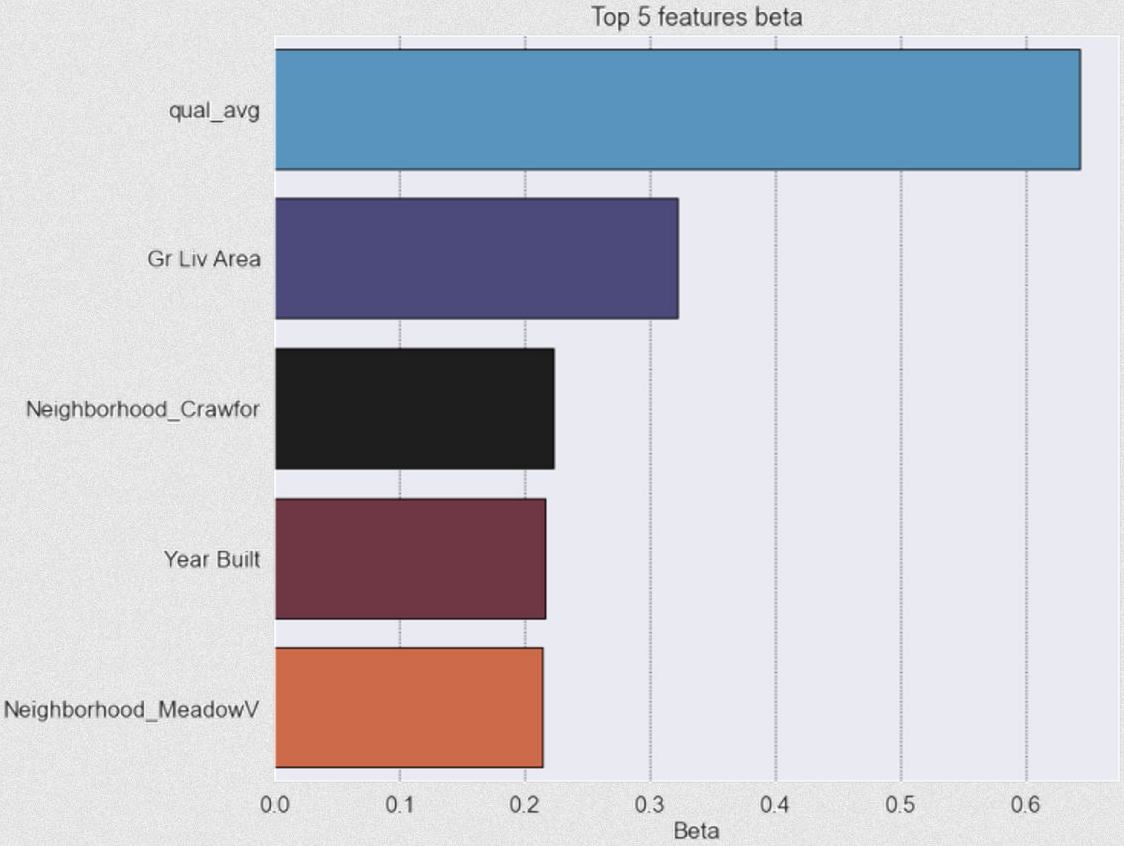
# Residual Distribution and Average

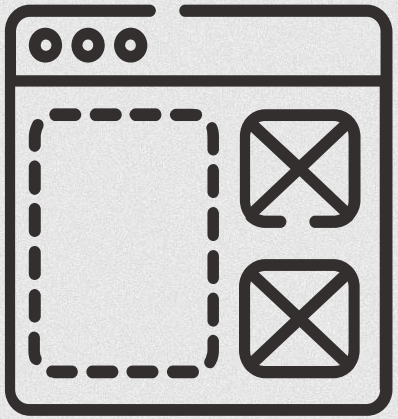



**Residual distribution is close to nominal and the average is near 0**

# Top 5  Features absolute Beta





# 04.

# Application

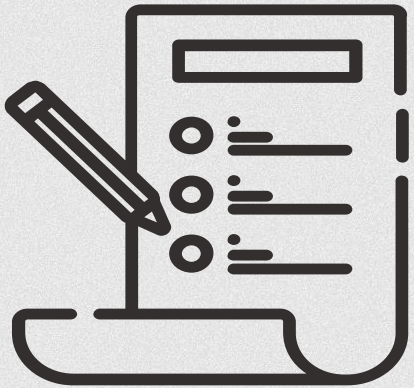
Predictor application

# Sale price $226631.25 (Predicted)

| | | | |
|---|---|---|---|
| Overall material and finish quality | 7 | Neighborhood | Collg Cr |
| Exterior material quality | 4 | Overall condition rating | 5 |
| Above Ground living area (Sq ft) | 1823 | Lot size (Sq ft) | 12192 |
| Kitchen quality | 4 | Size of garage in car capacity | 2 |
| Screen porch area (Sq ft) | 0 | Fireplace quality | 0 |
| Original construction date | 2000 | Basement finished area (Sq ft) | 663 |
| Proximity to main road or railroad | 3 | Home functionality rating | 7 |
| Total basement area (Sq ft) | 928 | Height of basement | 4 |

# Limitations of the Predictor

- Predictor is applicable only to the Ames, Iowa housing sale prices.
- Higher sale prices predictions show more variance due to insufficient training data for higher sale prices.
- Data collected more than 10 years old.
- Predictions are limited to the features given at the point of time.



# 05.

# Conclusion

Summary and recommendations

# Conclusion

1. Increased accuracy but would need to consider other factors to generalise to other markets
   (eg. Government policy - Singapore government's revision of additional buyer stamp duty in December 2021 as part of cooling measures)
2. Narrowed down the features from 82 to 16

# Recommendations

1. Collect more recent data

2. Include features from external sources
(e.g., timestamp related to sale, mortgage interest rates at time of sale, Iowa population growth, employment figures)

Create an app for a seamless experience!

# Questions?

# Thank you!