iPredict

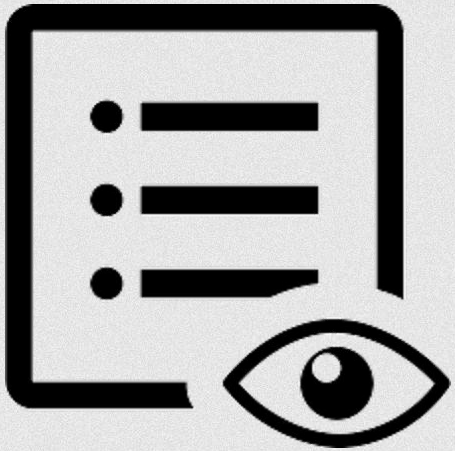
by MyProperty

Your one-stop portal to predict future house prices

Desmond, Ian, Richelle, Ridzuan

28 Sep 2022

Table of contents



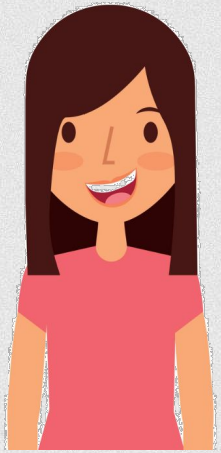
01.

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..

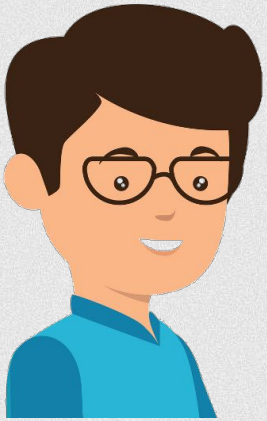


Aileen
2 years experience

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

Susan
8 years experience

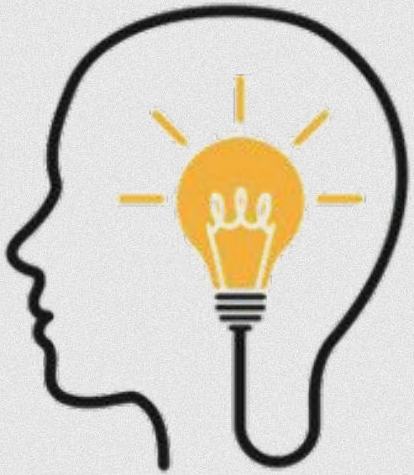
There are just way too many features to consider all the time...



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 • Good • Average/typical • Fair • Poor
Exterior condition	Present condition of the material on the exterior	
Kitchen quality	Kitchen quality	
Basement quality	Height of the basement	• Excellent (100+ inches) • Good (90-99 inches) • Typical (80-89 inches) • Fair (70-70 inches) • Poor (<70 inches) • 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.

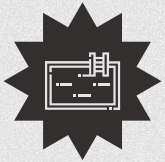
Data Cleaning

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)

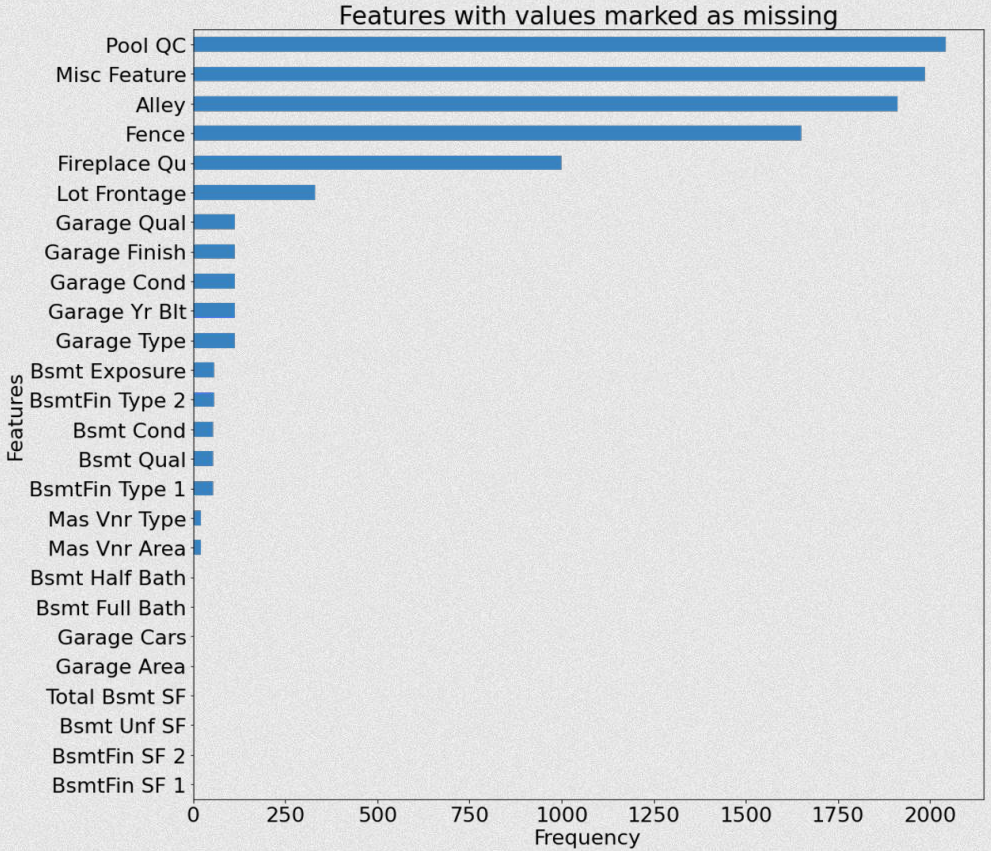
Columns with Missing Values

Columns with missing values

Missing values due to non-existent house features

Solution:

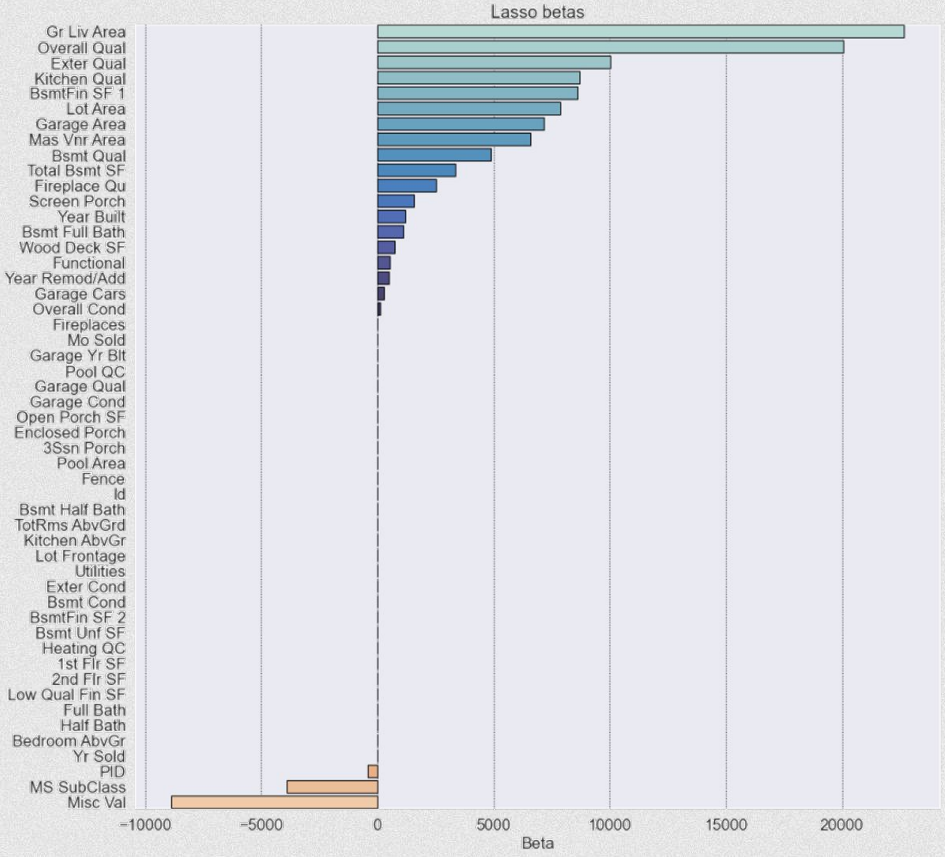
- “0” to replace missing numeric variables
- “None” to replace missing categorical variables
- Mean/mode to replace remaining variables



Numeric Feature Selection

Preprocessing features by Lasso

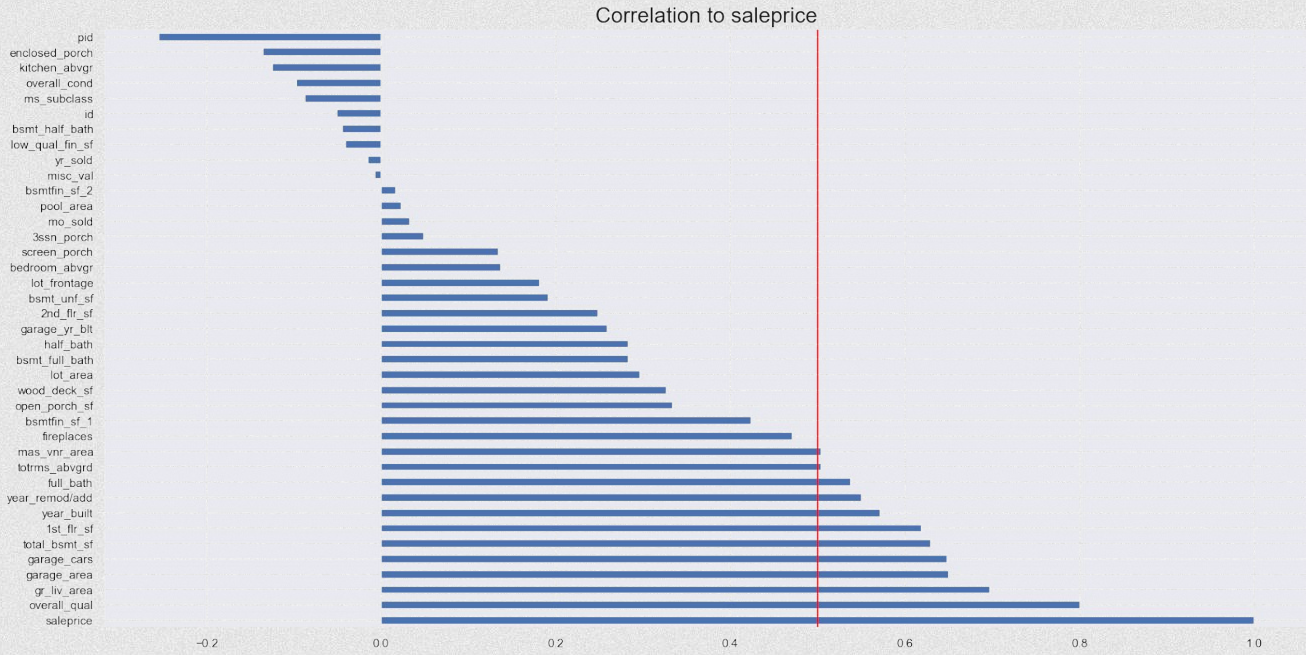
- 29 features were dropped with lasso beta of 0
- These features had no impact to sale price prediction → interfering noise with other model estimators (e.g., Linear Regression or Ridge)



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

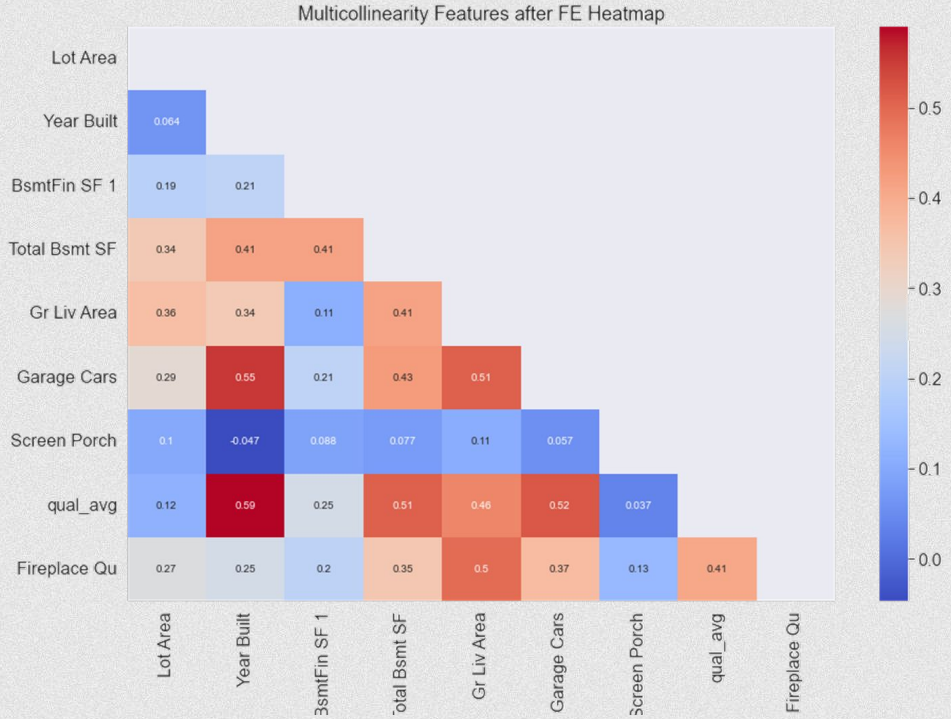
- Lasso coefficient (betas)
- Correlation between features only
- Correlation between features and sale price
- Feature engineering

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

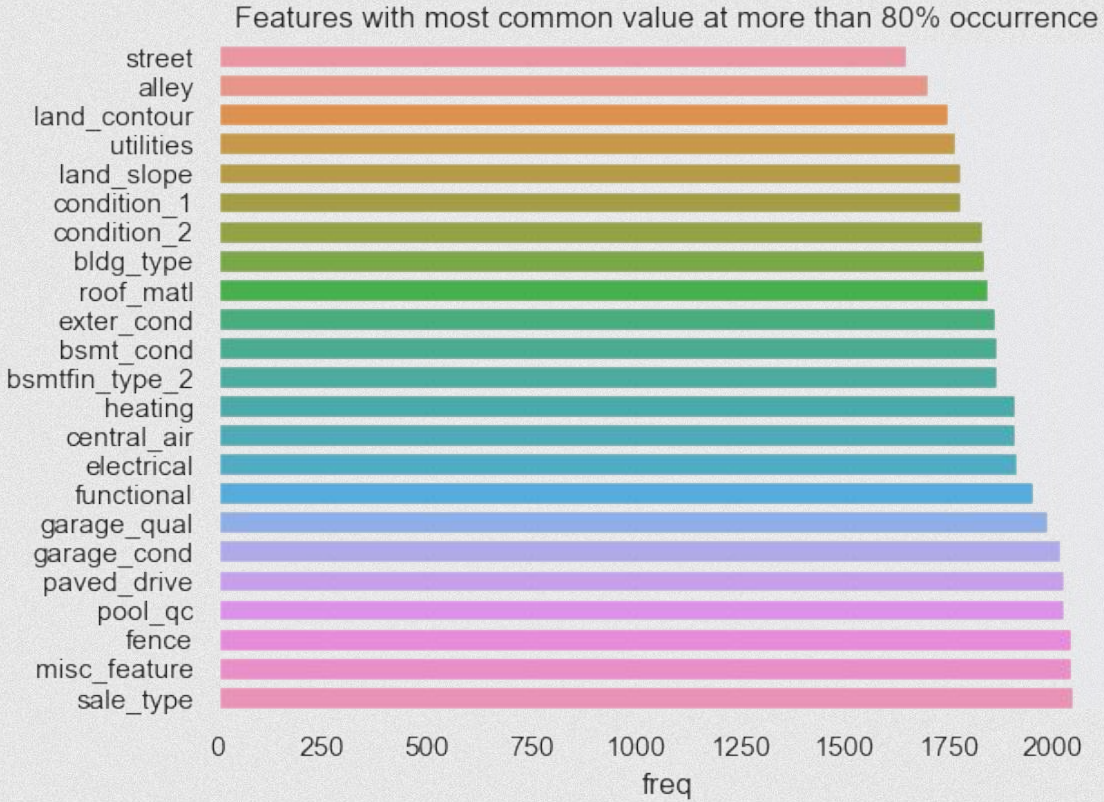
- 6 features to make qual_average



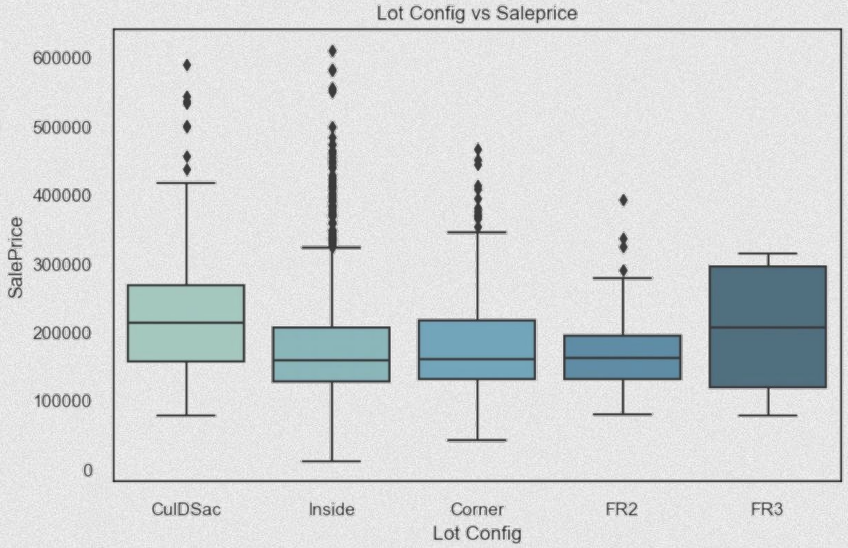
Categoric Feature Selection

More than 80% common value occurrence

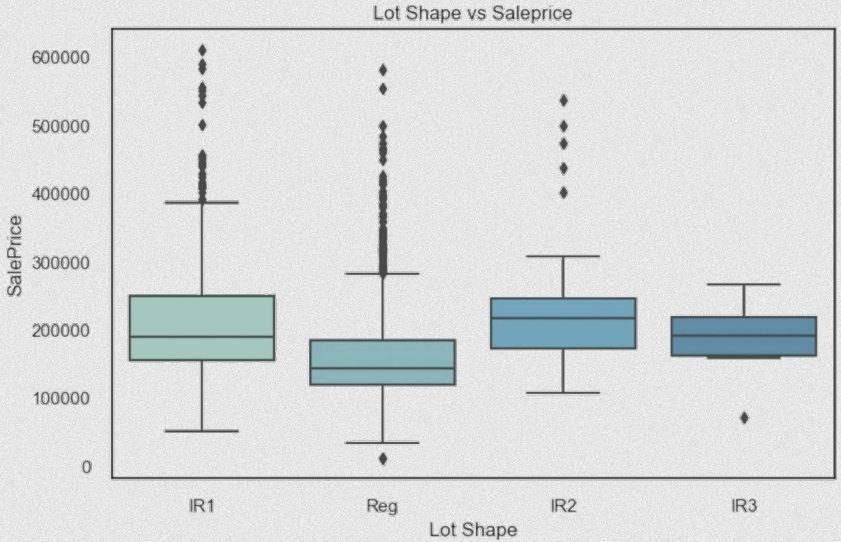
- These features were dropped and not considered for the the subsequent evaluations in the boxplots with sales price



Small Variance among Variables

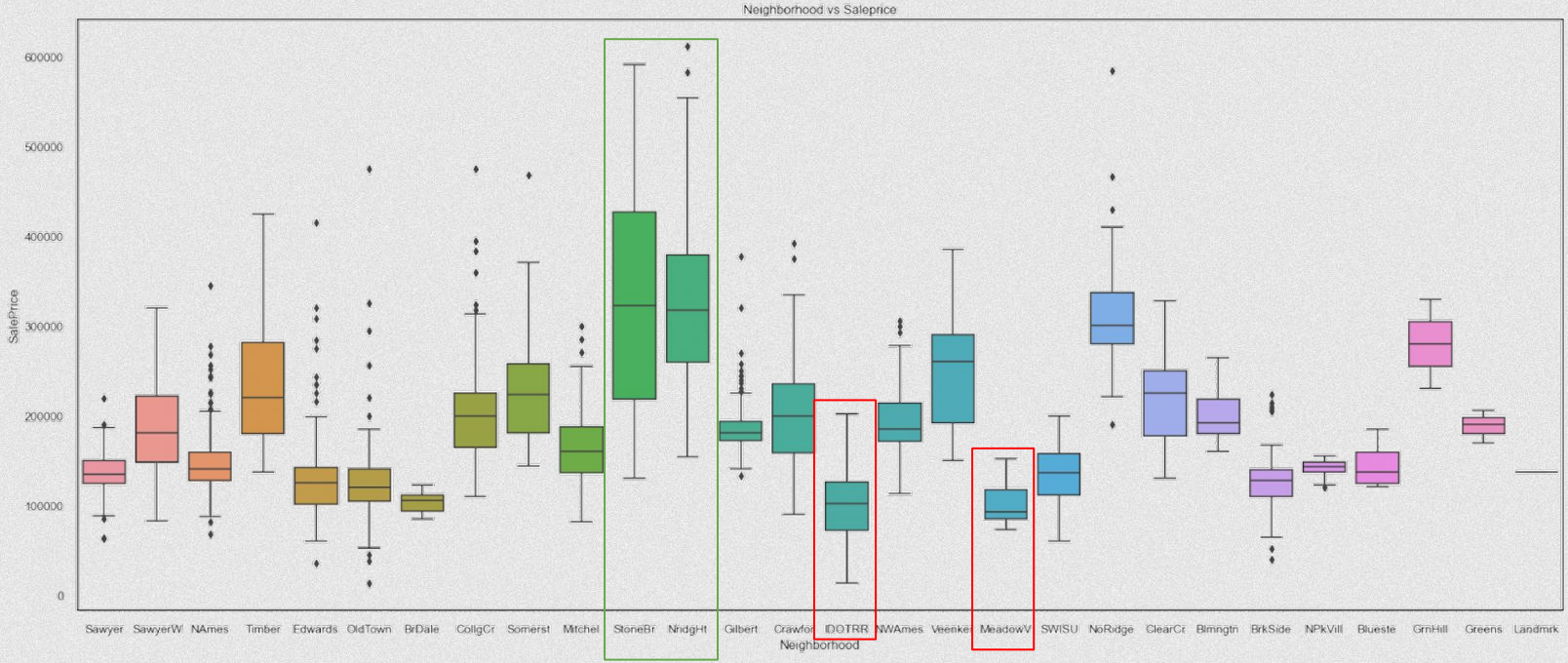


Most outliers: Inside and Corner Lot configurations



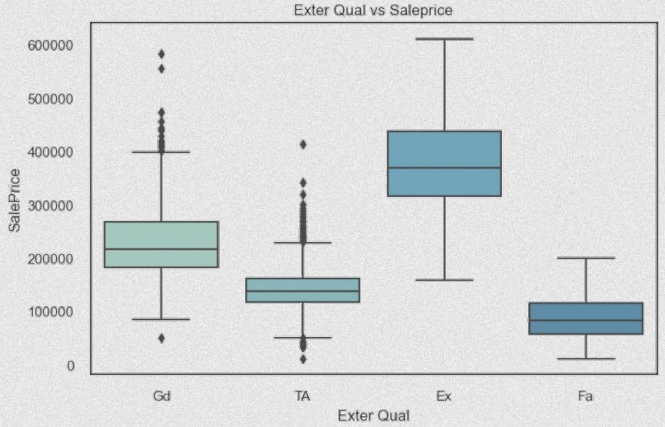
Most outliers: IR1 (slightly irregular) and Reg (regular)

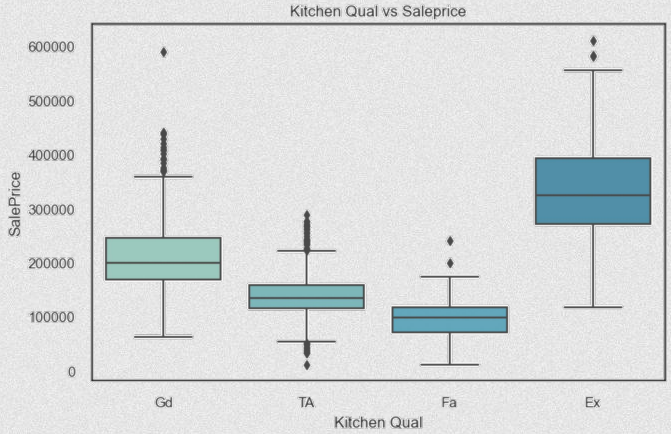
Boxplot of Neighborhoods with Sale Price

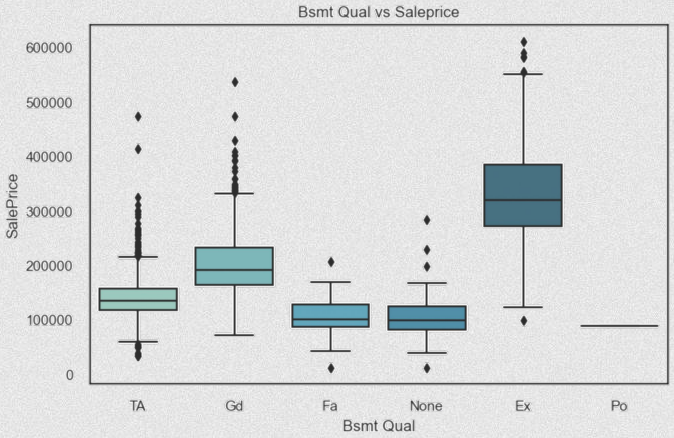


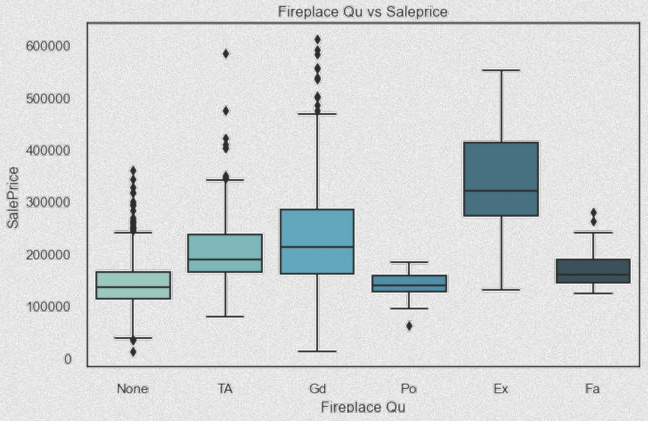
Neighborhoods with the highest median sale prices: Stone Brook and Northridge Heights
Neighborhoods with the lowest median sale prices: Meadow Village and Iowa DOT

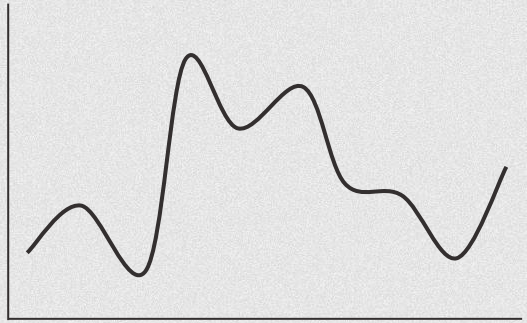
Boxplots of Various Features with Sale Price









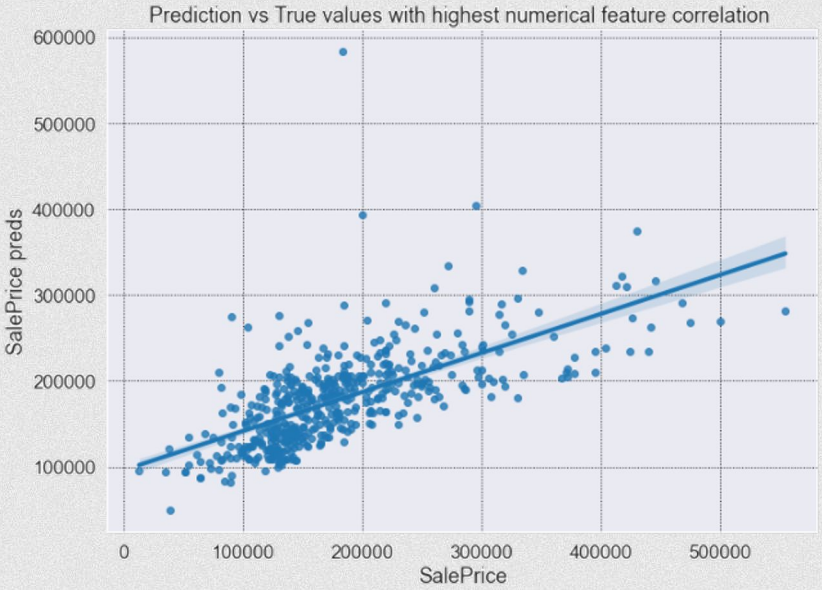


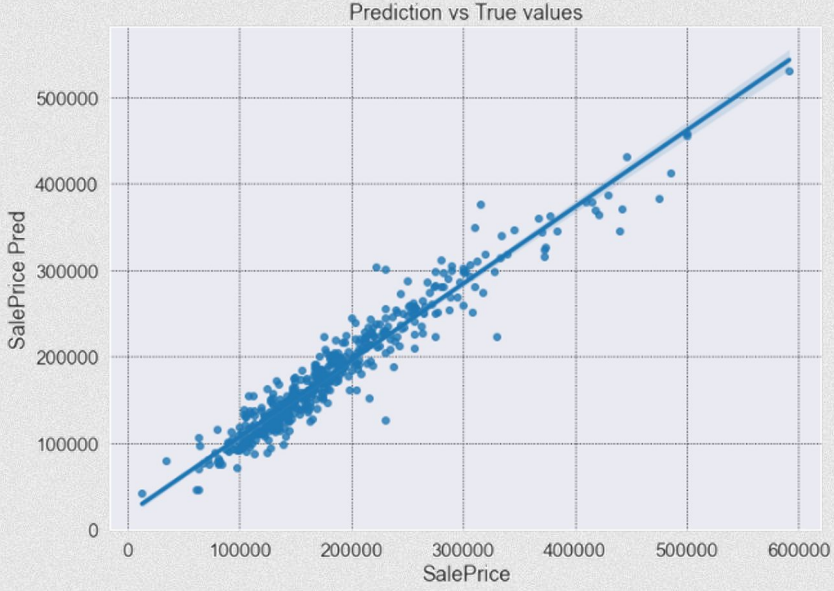
03.

Model

Regression model to predict sale prices

Baseline Model vs Final Best Model

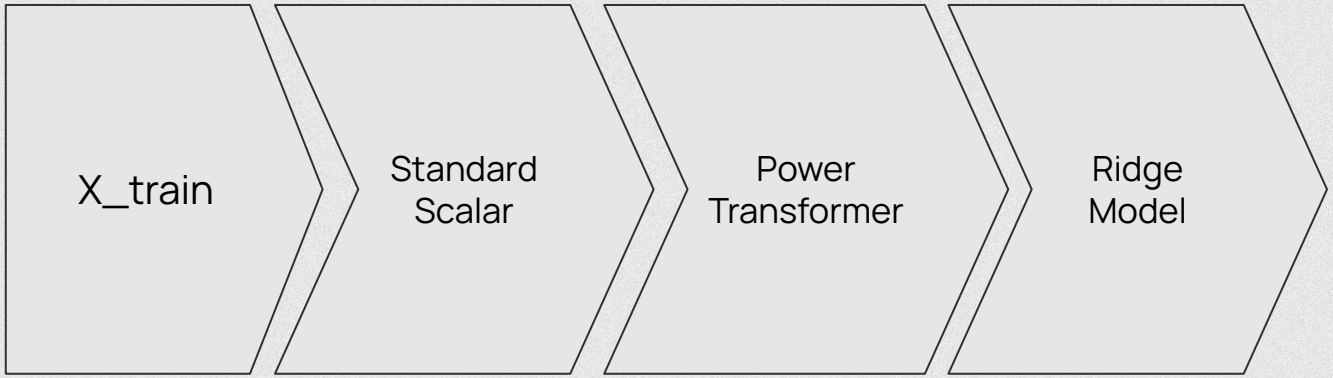




R2: 0.44
RMSE: 59355

R2: 0.93
RMSE: 21072

Model Fitting Process

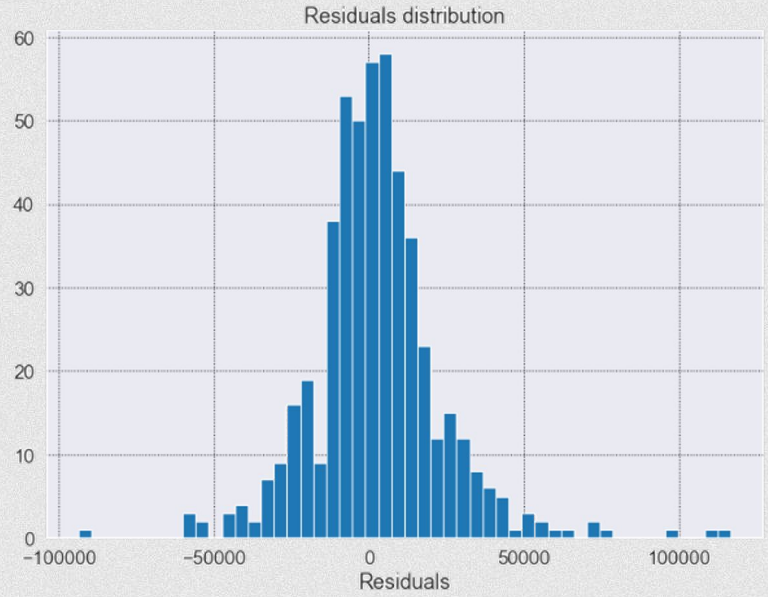


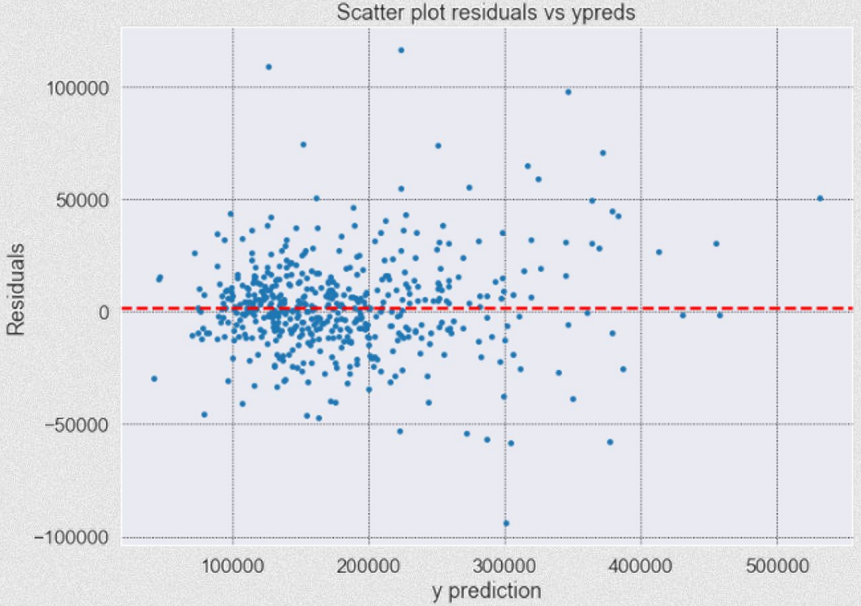
X_train features scaled by Standard Scalar

Scaled X Features transformed by Power Transformer

Transformed X features fitted with Ridge Model (hyperparameter tuned)

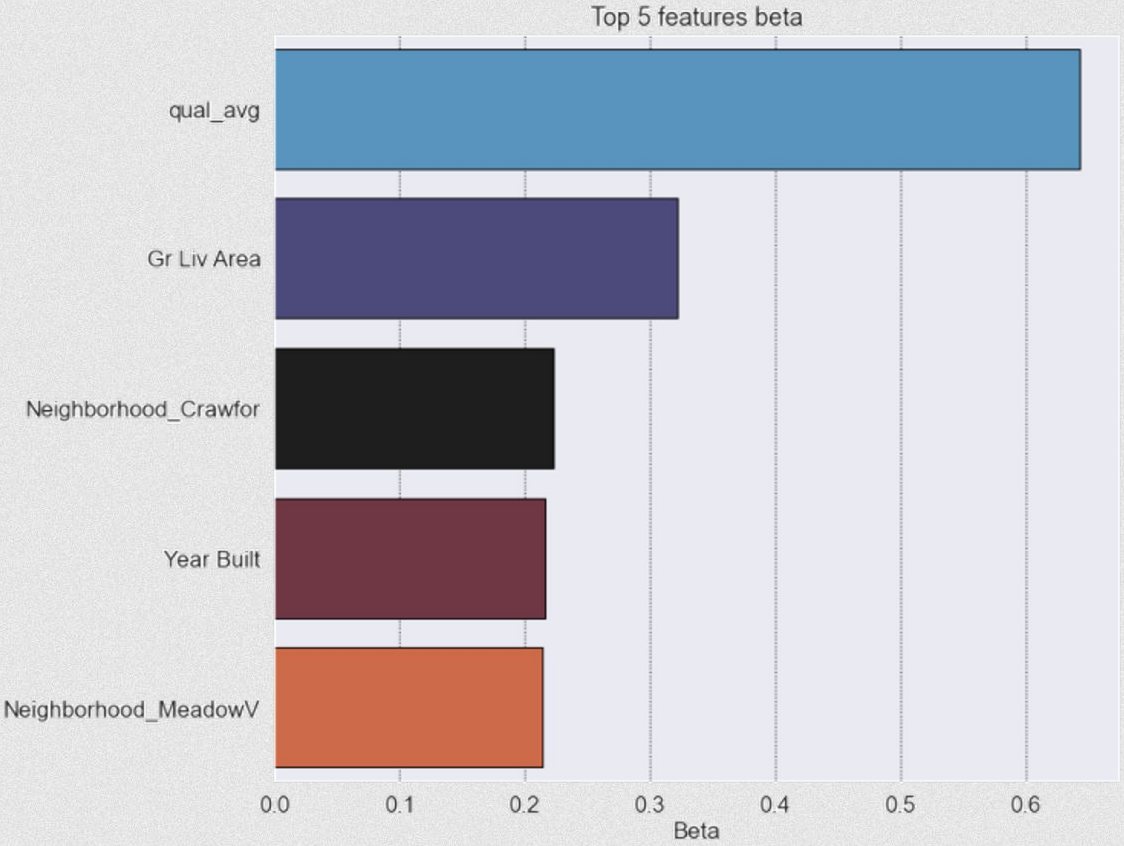
Residual Distribution and Average



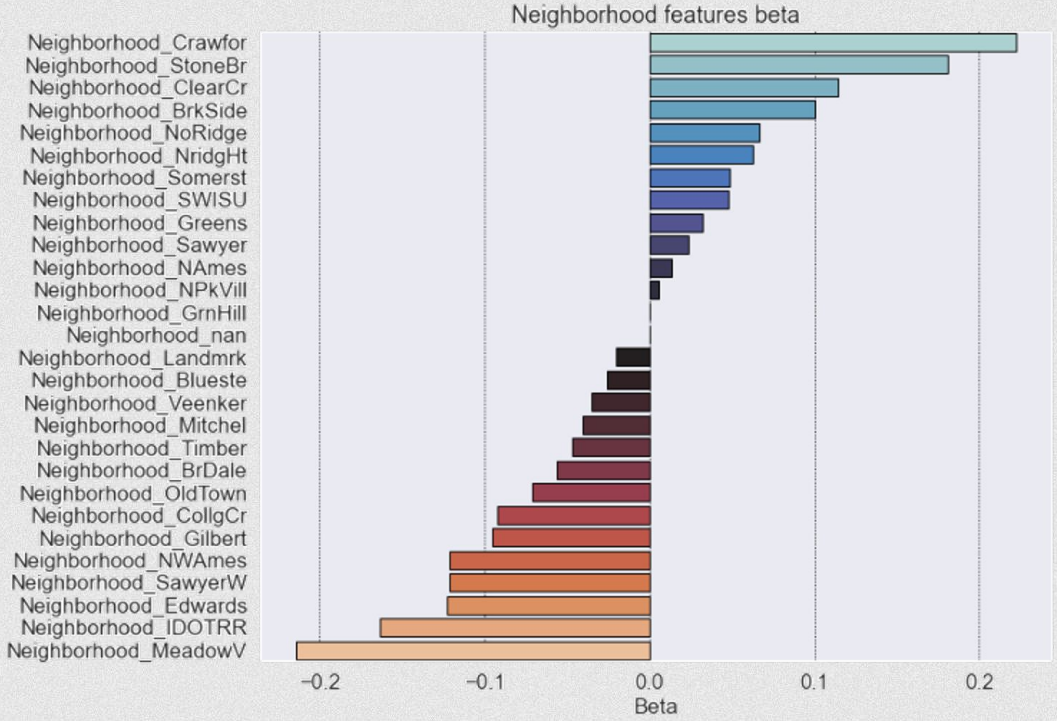


Residual distribution is close to nominal and the average is near 0

Top 5 Features absolute Beta

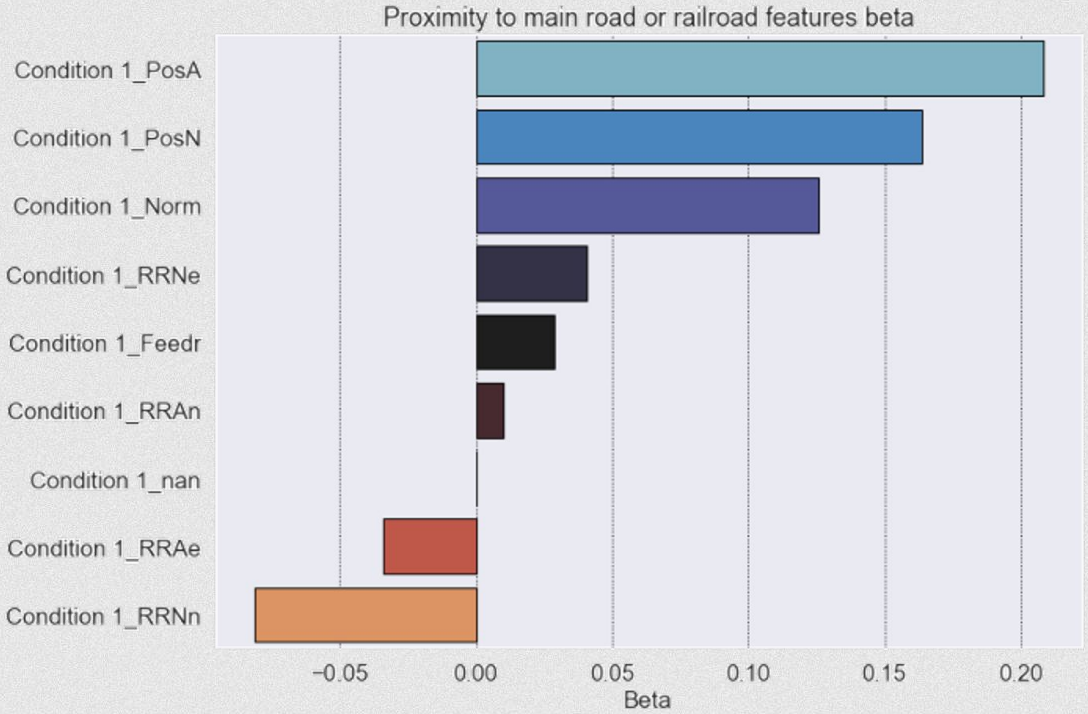


Neighborhood Features Betas



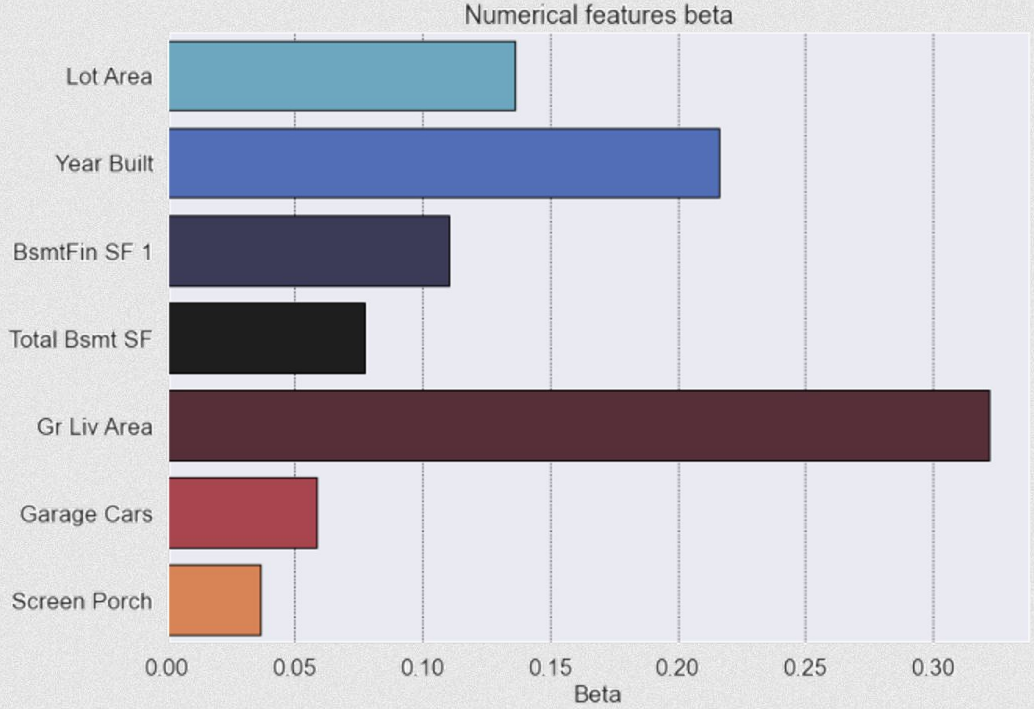
Meadow Village has the highest absolute beta for Neighborhood features

Proximity Features Betas



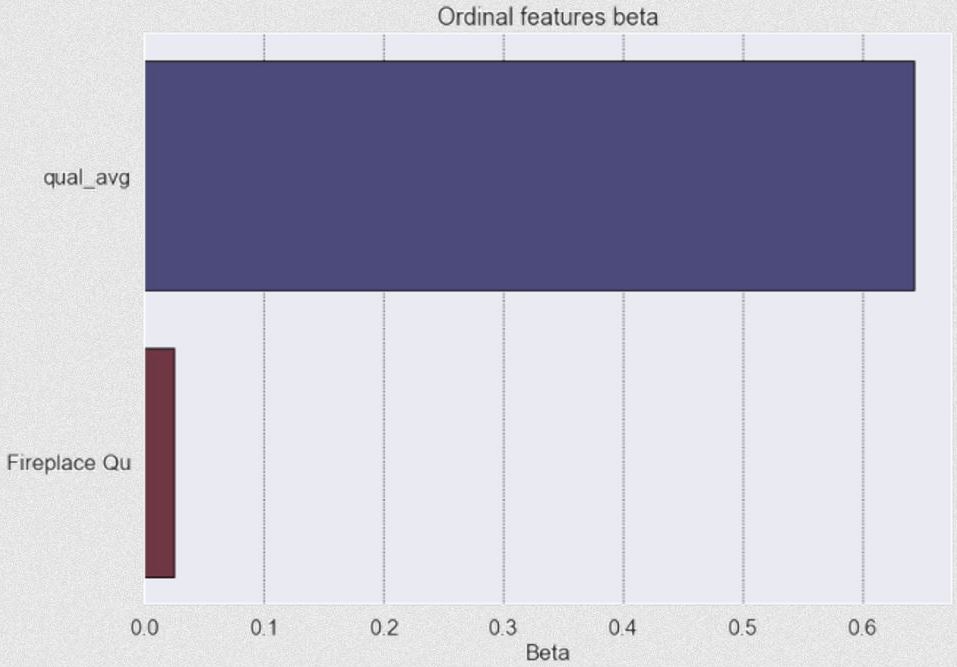
Adjacent to positive off-site feature (PosA) has the highest absolute beta for Proximity features

Numerical Features Betas



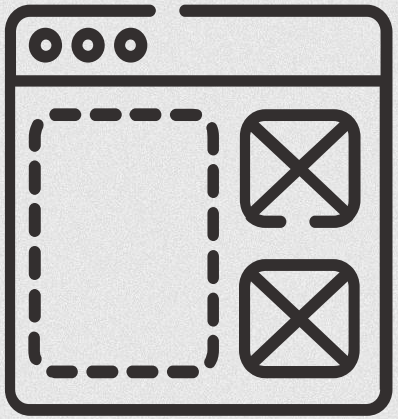
Above grade (ground) living area square feet has the highest absolute beta for numerical features

Ordinal Features Betas



Quality average as the highest absolute beta for ordinal features

Qual_avg = (Exter Qual + Functional + Overall Qual + Kitchen Qual + Bsmt Qual +Overall Cond)/6



04.

Application

Predictor application

What Data to Feed the Predictor?

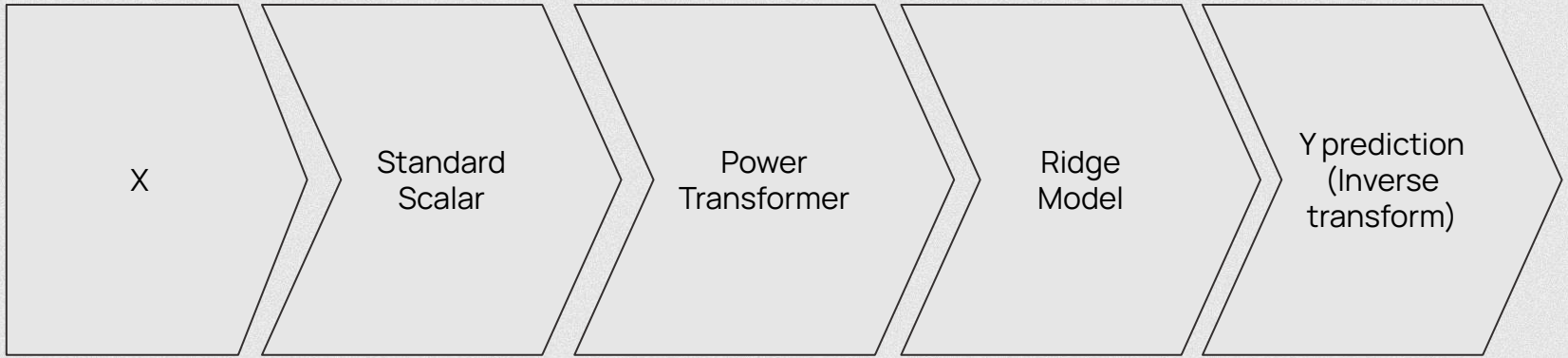
```
predict_sale_price()

------------------------------------------------------------------
1) Input Original construction date--> 2000
------------------------------------------------------------------
Overall Quality : 10-Very Excellent, 9, 8, 7, 6, 5, 4, 3, 2, 1-Very Poor

2) Input Overall material and finish quality, just the number only--> 7
------------------------------------------------------------------

Overall Condition : 10-Very Excellent, 9, 8, 7, 6, 5, 4, 3, 2, 1-Very Poor
3) Input Overall condition rating, just the number only--> 5
------------------------------------------------------------------
4) Input Lot size in square feet--> 12192
------------------------------------------------------------------
Exterior material quality : 4-Excellent, 3, 2, 1-Very Poor

5) Input Exterior material quality, just the number only--> 4
------------------------------------------------------------------
6) Input Size of garage in car capacity--> 2
------------------------------------------------------------------
7) Input Screen porch area in square feet--> 0
------------------------------------------------------------------
8) Input Above grade (ground) living area in square feet--> 1823
------------------------------------------------------------------
Kitchen quality : 5-Excellent, 3, 2, 1-Poor

9) Input Kitchen quality, just number only--> 4
------------------------------------------------------------------
Fireplace quality : 5-Excellent, 4, 3, 2, 1, 0-No Fireplace

10) Input Fireplace quality--> 0
------------------------------------------------------------------
Basement height :'
    5 : Excellent (100+ inches)
    4 : Good (90-99 inches)
    3 : Typical (80-89 inches)
    2 : Fair (70-79 inches)
    1 : Poor (<70 inches)
    0 : No Basement


11) Input the height of basement, just key in the number --> 4
------------------------------------------------------------------
12) Input Basement finished area in square feet--> 663
------------------------------------------------------------------
13) Input total Basement area in square feet--> 928
------------------------------------------------------------------
Home functionality Rating : 7-Typical Functional, 6, 5, 4, 3, 2, 1-Salvage only

14) Input Home functionality rating--> 7
------------------------------------------------------------------
```

```
14) Input Home functionality rating--> 7
------------------------------------------------------------------

   Abbreviation                        Neighborhood
0       Blmngtn                 Bloomington Heights
1       Blueste                           Bluestem
2        BrDale                           Briardale
3       BrkSide                           Brookside
4       ClearCr                          Clear Creek
5       CollgCr                        College Creek
6       Crawfor                           Crawford
7       Edwards                           Edwards
8       Gilbert                           Gilbert
9        IDOTRR                Iowa DOT and Rail Road
10      MeadowV                      Meadow Village
11      Mitchel                           Mitchell
12        Names                         North Ames
13      NoRidge                       Northridge
14      NPkVill                     Northpark Villa
15      NridgHt                   Northridge Heights
16       NWAmes                       Northwest Ames
17      OldTown                           Old Town
18        SWISU  South & West of Iowa State University
19       Sawyer                             Sawyer
20      SawyerW                         Sawyer West
21      Somerst                            Somerset
22      StoneBr                         Stone Brook
23       Timber                          Timberland
24      Veenker                            Veenker
25         None                        unspecified

15) Input Neighborhood area, key in the abbreviation(case sensetive)--> CollgCr
------------------------------------------------------------------

Proximity to main road or railroad
    1 : Adjacent to arterial street(Artery)
    2 : Adjacent to feeder street(Feedr)
    3 : Normal(Norm)
    4 : Within 200' of North-South Railroad(RRNn)
    5 : Adjacent to North-South Railroad(RRAn)
    6 : Near positive off-site feature--park, greenbelt, etc.(PosN)
    7 : Adjacent to postive off-site feature(PosA)
    8 : Within 200' of East-West Railroad(RRNe)
    9 : Adjacent to East-West Railroad(RRAe)

16) Input the number representation only--> 3
------------------------------------------------------------------

Thank You
The saleprice of this house is near $226631.35.
```

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

From X to Y Predictions



X Features scaled by Standard Scalar

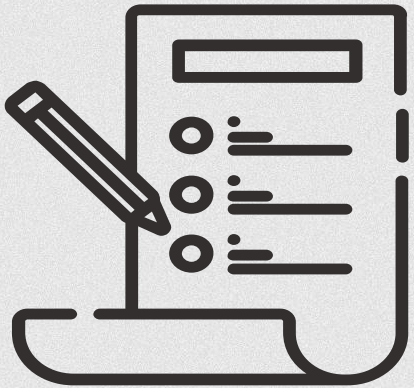
Scaled X Features transformed by Power Transformer

Transformed X Features predicted with Ridge Model

Transformed predicted Y Inverse transform to normal Y prediction

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!