

Project 2:

Ames Housing Price

Dan Chen
July 23, 2021

Overview

1. Problem Statement
2. Data Clean and EDA
3. Feature Engineering
4. Model and Evaluation
5. Prediction and Result
6. Recommendations



Problem Statement

Based on the analysis on Ames Housing Train Dataset, we will set up a model to predict the housing price with Test Dataset and make recommendation to Ames homeowners how to invest and maintain their properties values

Datasets

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

train.csv

test.csv

Missing Values in Features

Pool QC	2042
Misc Feature	1986
Alley	1911
Fence	1651
Fireplace Qu	1000
Lot Frontage	330
Garage Finish	114
Garage Qual	114
Garage Yr Blt	114
Garage Cond	114
Garage Type	113
Bsmt Exposure	58
BsmtFin Type 2	56
BsmtFin Type 1	55
Bsmt Cond	55
Bsmt Qual	55
Mas Vnr Area	22
Mas Vnr Type	22
Bsmt Half Bath	2
Bsmt Full Bath	2

Categorical Features

1. Fill in missing values with object values
2. Plot categorical features vs housing price
3. Conduct ANOVA analysis
4. Remove irrational categorical features

```
['MS Zoning', 'Street', 'Land Contour', 'Neighborhood', 'Condition 1',  
'Bldg Type', 'House Style', 'Roof Matl', 'Exterior 1st', 'Mas Vnr Type',  
'Exter Qual', 'Exter Cond', 'Foundation', 'Bsmt Qual', 'Bsmt Cond',  
'Bsmt Exposure', 'BsmtFin Type 1', 'Heating QC', 'Central Air',  
'Electrical', 'Kitchen Qual', 'Functional', 'Garage Type',  
'Garage Finish', 'Garage Qual', 'Garage Cond', 'Paved Drive',  
'Sale Type'],
```

Numerical Features

1. Fill in missing values with 0s
2. Aggregate related features
3. Remove weak housing price correlated numerical features
4. Remove irrational numerical features

Corr vs Sale Price

Overall Qual	0.807756
Year Remod/Add	0.562313
Mas Vnr Area	0.473979
BsmtFin SF 1	0.438484
Total Bsmt SF	0.652809
Gr Liv Area	0.711655
TotRms AbvGrd	0.486419
Fireplaces	0.478641
Garage Cars	0.652872
Garage Area	0.652043
SalePrice	1.000000
HouseAge	-0.583435
Flr SF	0.720825
Bath	0.620621
Attach SF	0.445135
Lot SF	0.278537

Prediction Models

1. Linear Regression
2. Ridge Regression - CV optimized
3. Lasso Regression - CV optimized
4. ElasticNet Regression - CV optimized



```
R2 Score in Each Model  
train  
test
```

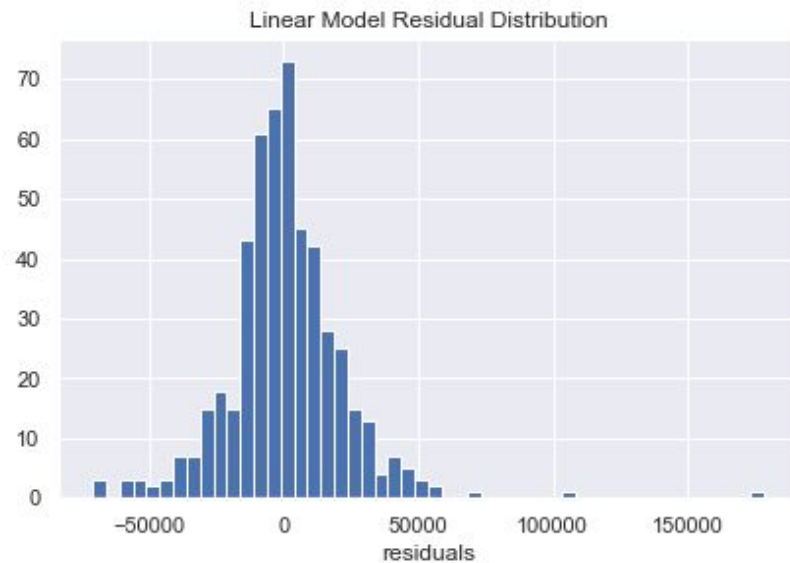
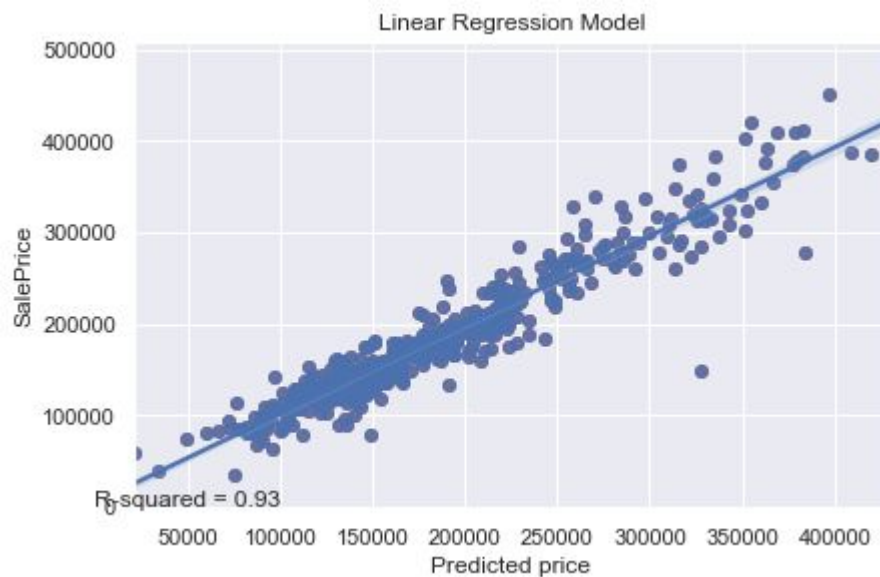
```
===== LR =====  
0.9363390720274143  
0.9150219973069138
```

```
===== Ridge =====  
0.9305219240417956  
0.9192954187168814
```

```
===== Lasso =====  
0.9362716330781822  
0.9158670964292007
```

```
===== ElasticNet =====  
0.9215335345996165  
0.9153459819875953
```

Linear Regression



Prediction

1. Import test data
2. Do same data clean process as train set
3. Predict housing price with Linear Regression
4. Export submission.csv
5. Explore TOP 20 strong price correlated features

Recommendations

1. There are positive and negative correlated features
2. Maintain/Increase positive features and Improve negative ones

GOOD / BAD	Recommendation
Wood Shingles	Pay attention to roof materials, using wood shingles could add more values
Typical home functionality with minor deductions	Maintain typical home functionality
No Garage	House without garage ??? value increase fast
Sale Type - 15%DownPay	15%DownPay could increase the sale price
Neighborhood	Move or invest in good neighborhood
Bldg Type - Townhouse End/Inside unit	Townhouse End/Inside unit probably not good investment
Poor garage quality	Remodeling your garage

THANKS!!!