



HOUSING PREDICTION

---

IOWA

---

ZHUOQUAN CHEN



# CONTENT

Problem Statement

Data Cleaning

Data Pre-processing

Features Engineering

Modeling & Regularization

Prediction

Recommendation

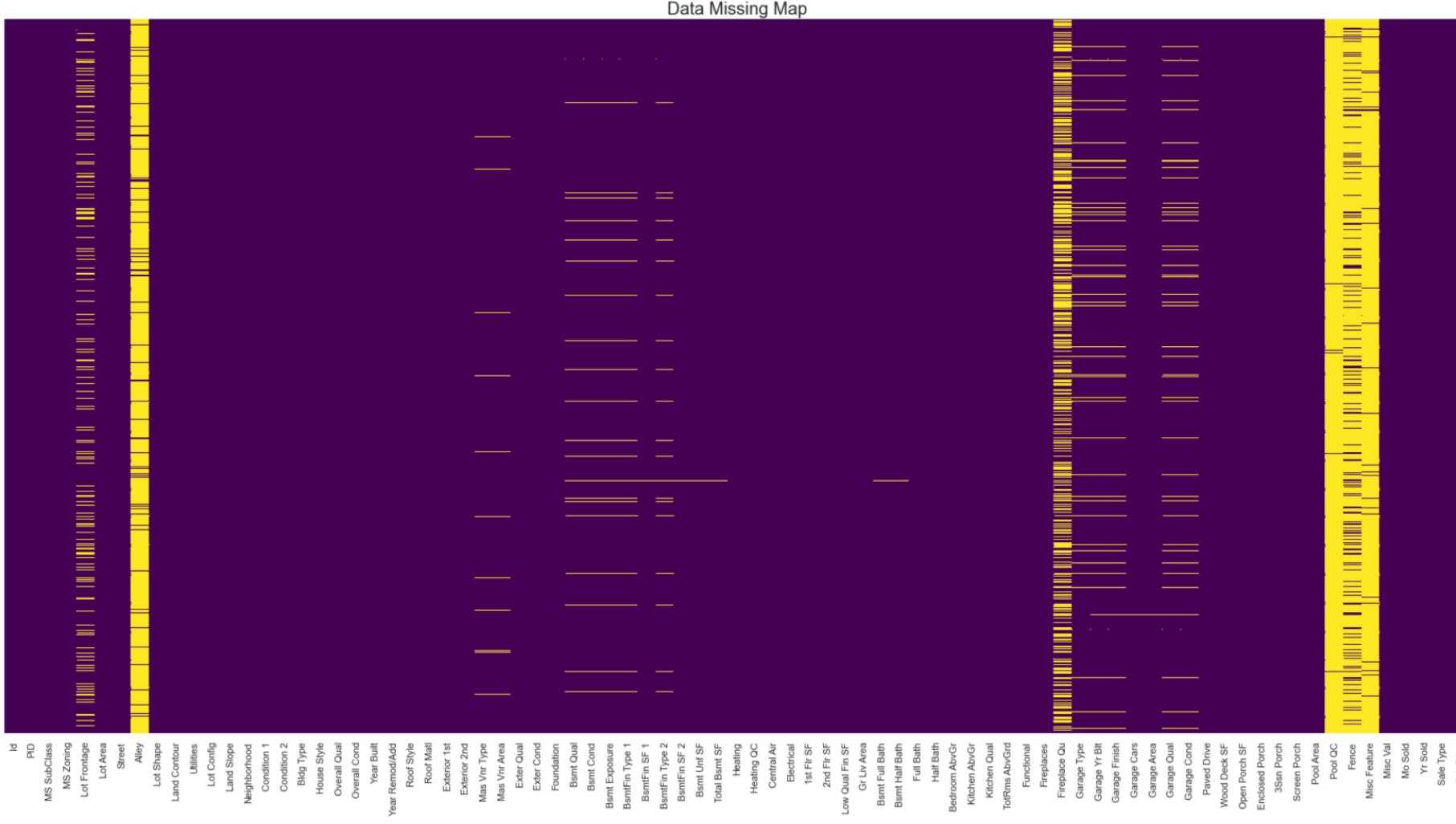


## PROBLEM STATEMENT

My team was assigned a new task recently with conducting analysis and creating a regression model based on the Ames Housing Dataset, which will predict the house prices at the sale in Ames, Iowa. Through this analysis and modeling, we can gain more insights into which factors can impact the market prices in Ames.



# DATA CLEANING



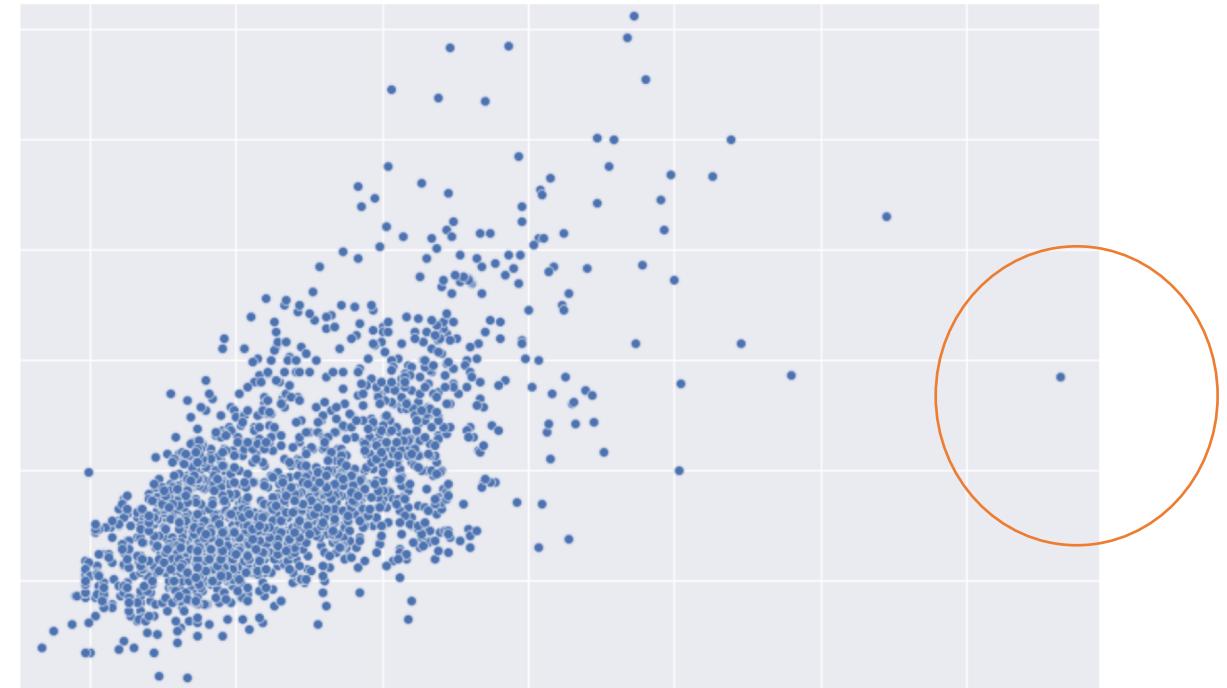


# DATA CLEANING

1> Replacing null null value

- Most null values correspond to 'NA'

2> Removing outliers





# DATA PRE-PROCESSING

Features with higher correlations to sale price.

```
1 df_train.corr()[['saleprice']]
```

	saleprice
id	-0.051008
pid	-0.253495
ms_subclass	-0.087205
lot_frontage	0.338076
lot_area	0.304151
lot_shape	-0.297281
utilities	0.026586
land_slope	-0.051304
overall_qual	0.803981
overall_cond	-0.096381
year_built	0.574203
year_remod/add	0.552135
exter_qual	0.715556
exter_cond	0.039348
bsmt_qual	0.613768
bsmt_cond	0.227207
bsmt_exposure	0.424001
bsmtfin_type_1	0.349591
bsmtfin_sf_1	0.444887
bsmtfin_type_2	0.016219
bsmtfin_sf_2	0.019421
bsmt_unf_sf	0.188498



## FEATURE ENGINEERING

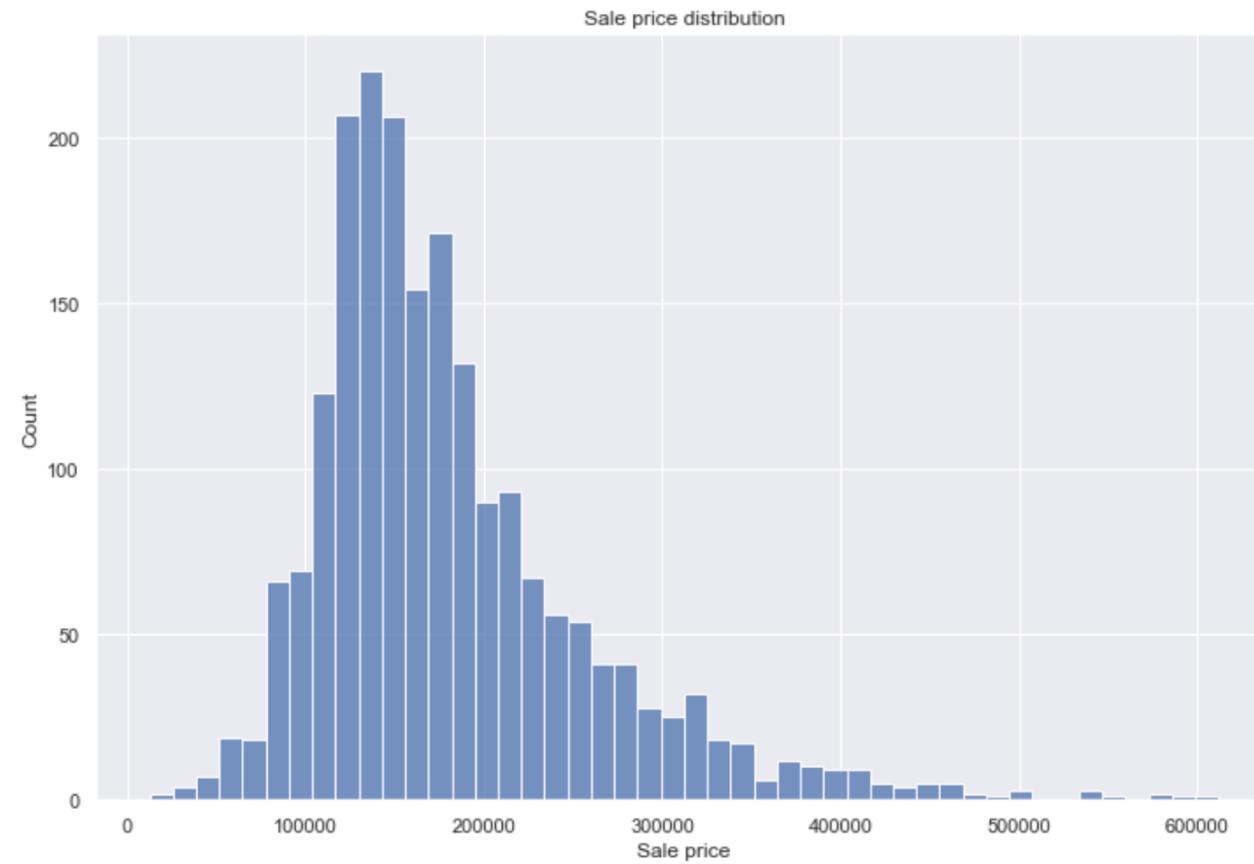
In Ames dataset, the key problem is how to deal with:

**Numerical Feature**  
**Categorical Feature**



# FEATURE ENGINEERING

For example: saleprice is a numeric attribute, and its distribution looks right skew.

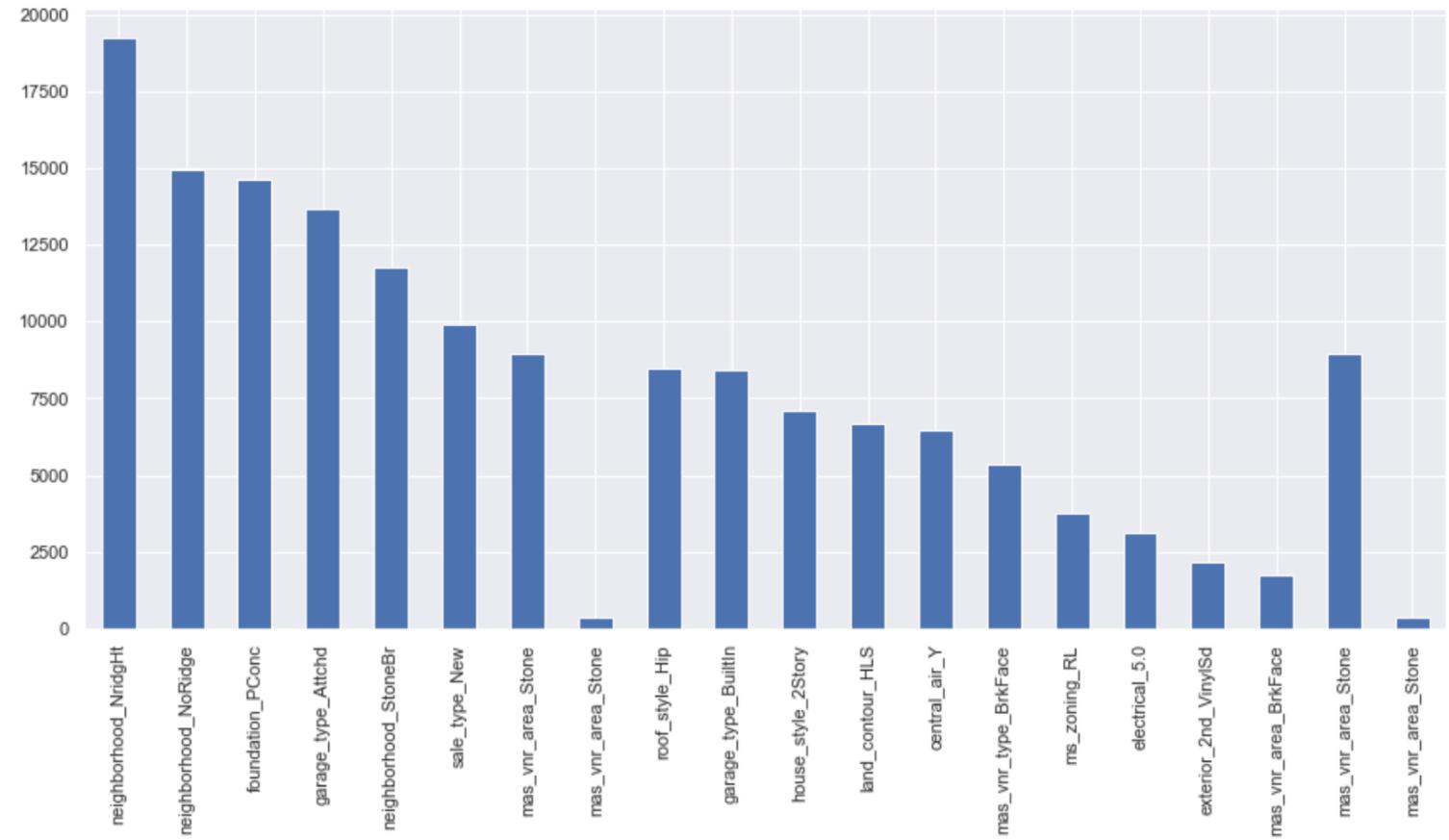




# MOEDLING & REGULARIZATION

## Linear Regression

- Ridge Regularization
- Lasso Regularization





## PREDICTION

North Ridge neighborhood has the highest positive effect on property sale price, the following is Stone Brook neighborhood.

neighborhood_NoRidge	100365.516848
neighborhood_StoneBr	88252.868801
neighborhood_NridgHt	82088.575601
sale_type_New	37353.865515
garage_type_BuiltIn	35476.661350
mas_vnr_area_Stone	35092.413338
land_contour_HLS	33740.093408
foundation_PConc	29586.152200
garage_type_Attchd	28008.072108
central_air_Y	26308.328622
roof_style_Hip	21829.047967
house_style_2Story	16021.356040
mas_vnr_type_BrkFace	15533.361464
electrical_5.0	11244.161894
ms_zoning_RL	9181.071369
exterior_2nd_VinylSd	4542.205137
mas_vnr_area_BrkFace	12.697129
mas_vnr_area_Stone	0.000000
dtype:	float64



## RECOMMANDATION & CONCLUSION

The top feature that highly influences a house price is at a coefficient of 19206.483386 which located at Northridge neighborhood. It is a reasonable predictor as this neighborhood is very convenient and Gilbert Middle School and Gilbert High School located here, which are the best school high quality schools in Iowa. It is also very close to downtown and is just about a 10 minute drive away.

In addition, remodeling your house help! And also the concrete foundation can keep your house has a high price.

	saleprice
ms_zoning_RL	0.231297
land_contour_HLS	0.205066
neighborhood_NoRidge	0.269008
neighborhood_NridgHt	0.440740
neighborhood_StoneBr	0.251010
house_style_2Story	0.203573
roof_style_Hip	0.261220
exterior_1st_VinylSd	0.347296
exterior_2nd_VinylSd	0.343395
mas_vnr_type_BrkFace	0.265216
mas_vnr_type_Stone	0.299993
mas_vnr_area_BrkFace	0.265216
mas_vnr_area_Stone	0.299993
foundation_PConc	0.530752
central_air_Y	0.279440
electrical_5.0	0.256996
garage_type_Attchd	0.366813
garage_type_BuiltIn	0.209413
sale_type_New	0.357099