House Prices Prediction

By Zhi Jiang

How to predict a house price?

Need to know some features of the house:

—number of bedrooms,

—total living area, etc

Some Features are more important than others:

—number of bedrooms $>$ height of basement ceiling.

Machine learning model

Data Cleaning:

-Feature Selection

-Missing data

-Categorical features

Build Models:

-Select suitable models

-Model training

-Hyperparameter tuning

-Model evaluation

Prediction:

- Select best performed model

-Make house price predictions for new data

Data Cleaning

Data Set

- train.csv

  1460 data points,
  79 features,
  1 target, i.e. house price.

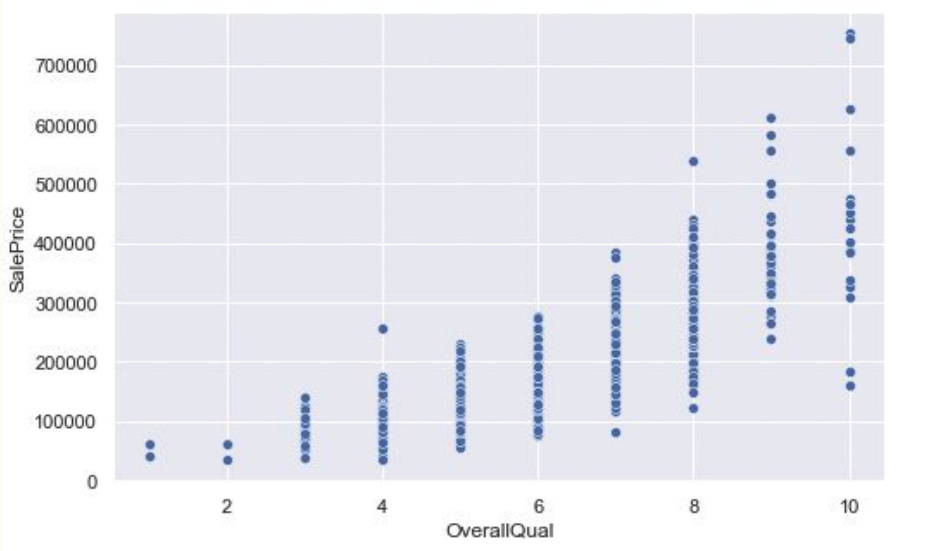
- test.csv

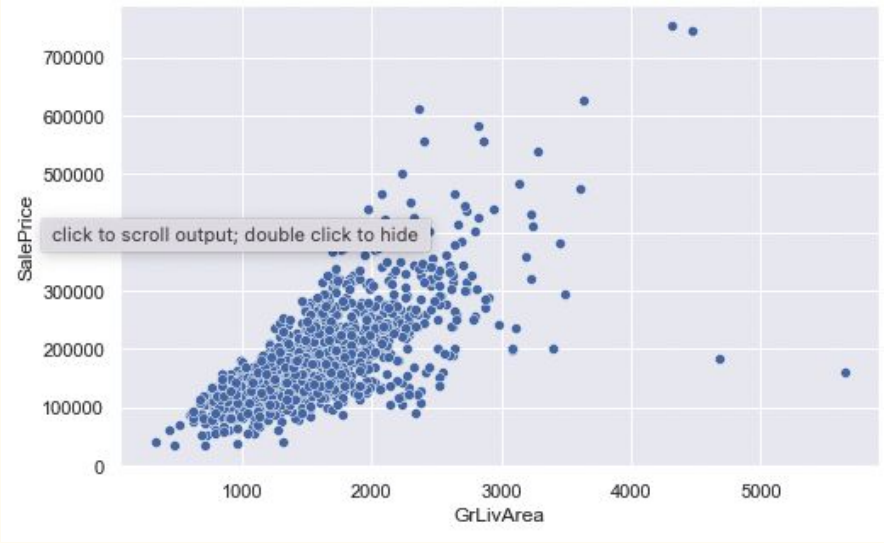
  1459 data points,
  79 features.

Some of the Features

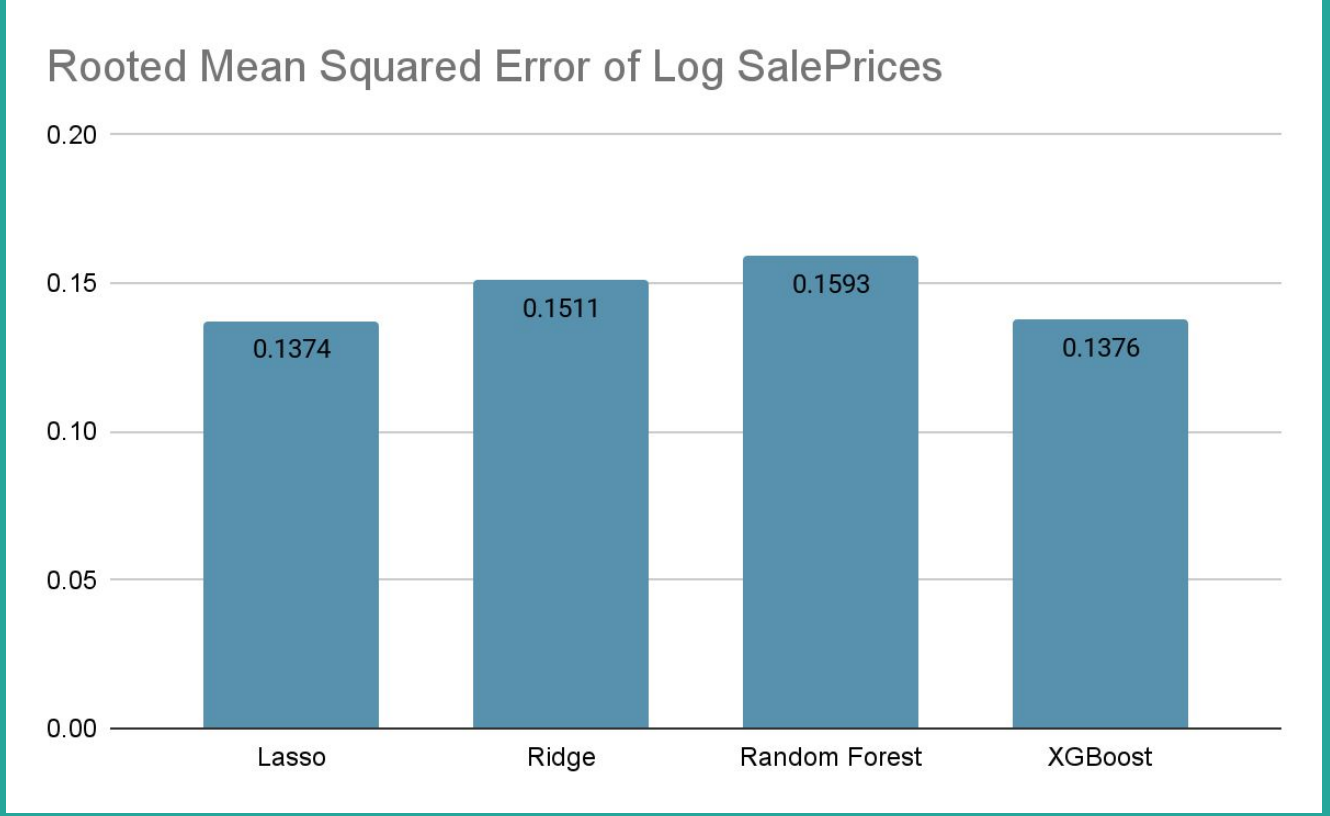
- GrLivArea: Above ground living area
- OverallQual: Overall material and finish quality
- Bedroom: Number of bedrooms
- YearBuilt: Original construction date

Data Exploration





Build Models



We select XGBoost as our final model to make predictions. It has low RMSE and runs pretty fast.

We also take Lasso Regression as comparison, as it also has pretty low RMSE.

Predictions

XGBoost

score:0.14546

	Id	SalePrice
0	1461	126442.718750
1	1462	160120.390625
2	1463	179746.546875
3	1464	188976.390625
4	1465	189034.265625

Lasso Regression

score:0.15212

	Id	SalePrice
0	1461	119607.628818
1	1462	161446.293476
2	1463	173668.848506
3	1464	189560.270305
4	1465	214366.284691

Conclusion

In this project, we analyzed the house price data set carefully, and build high performance machine learning model to predict the house prices.

We finally select XGBoost Regressor as our final model, as it has low RMSE and runs fast. We made predictions based on this model!

Thank you for attention