



The Next Hot Real Estate Market

How to make sense of Ames housing data



Overview

- Cleaning and EDA
- Feature Engineering
- Modeling



Data Cleaning

```
def series_of_null_features(df):  
    return df.loc[:,df.isnull().sum() > 0].isnull().sum()
```

```
>> lot_frontage      330  
>> alley             1911  
>> mas_vnr_type      22  
>> mas_vnr_area      22  
>> bsmt_qual         55  
>> bsmt_cond         55  
>> bsmt_exposure     58  
>> bsmtfin_type_1    55  
>> bsmtfin_sf_1      1  
>> bsmtfin_type_2    56  
...
```



Feature Engineering

LassoCV as feature selection

Variance Threshold (.2)

SelectKBest (20)



Modeling

	Feature Select	Model type	Public score	Private score
1	Drop Categoricals	OLM	\$41571	\$29853
2	Lasso	Ridge	\$54315	\$19472
3	VarThresh + SelectKBest	Elastic Net	\$66166	\$22200



In Conclusion

Reducing features to reduce variance

Try other models

Interaction terms