

Ames Housing Data and Kaggle Challenge

By Binita Patel
Data Scientist





Table of contents:

- 01 Data Science Problem
- 02 Procedure/Methodology
- 03 Primary Findings
- 04 Next Steps and Recommendations

Data Science Problem





- Identify best predictors of house sale prices
- Give potential customers best price
- Increase Revenue

Data Science Problem

Procedure/Methodology



Procedure/Methodology



EDA



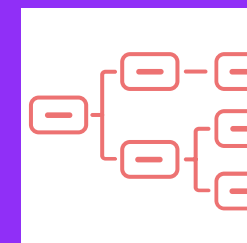
**Data
Cleaning**



Modelling



Evaluation



**Prediction
and Kaggle
Submission**

Primary Findings



	Models						
	Baseline Model	Model 1: Top 10	Model 2: Top 5	Model 3: Top 15	Model 4: Top 15 + RidgeCV	Model 5: Top 15 + LassoCV	Model 6: Top 15 + 3 cat vars + LassoCV
R-Squared Score (Train)	-0.03	0.80	0.79	0.82	0.91	0.88	0.89
R-Squared Score (Test)	-0.03	0.72	0.71	0.72	0.77	0.77	0.8
RMSE	70,375	36,300	37,291	34,757	25,085	27,809	26,292

Next Steps and Recommendations





Next Steps and Recommendations



- Variables correlated to House Prices
- Look at more categorical data
- Look at different alphas
- More time and resources can lead to better results

Thank you!



Binita Patel
Email : binita.patel.bp@gmail.com