

The background is a dark blue gradient. It features several vertical white lines of varying lengths. Scattered throughout are small squares in teal, orange, and pink. Some squares are solid, while others are outlined. The overall aesthetic is modern and tech-oriented.

Real Estate through DATA SCIENCE

Jahnvi Kalpathy
DSIR 1031
2022-12-09

Problem Statement

Creating a monetized calculator for real-estate price predictions:

- Premium short-term subscription service
 - For potential buyers/sellers to better anticipate the market
- Buyers: tweak parameters to find dream house in budget!
- Sellers: input data to calculate best evaluation of your home!

Questions of Interest:

- Minimize RMSE
- Greatest impact features
- How can a seller improve their value?

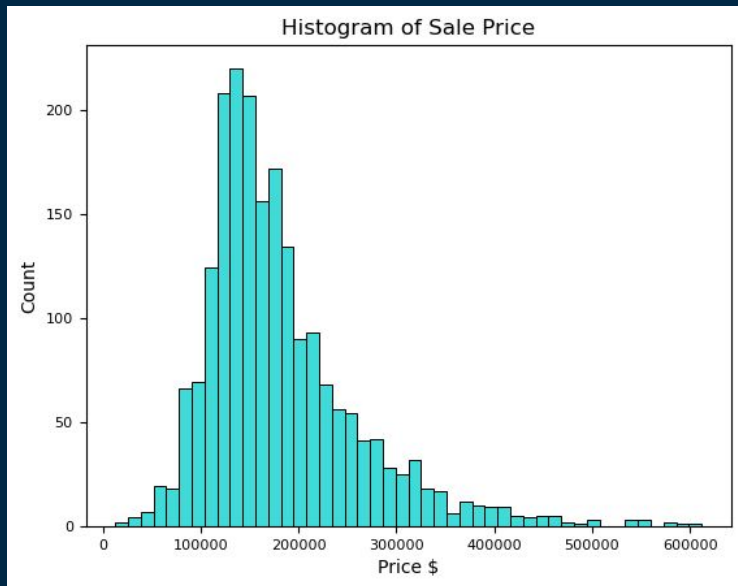


Ames, Iowa



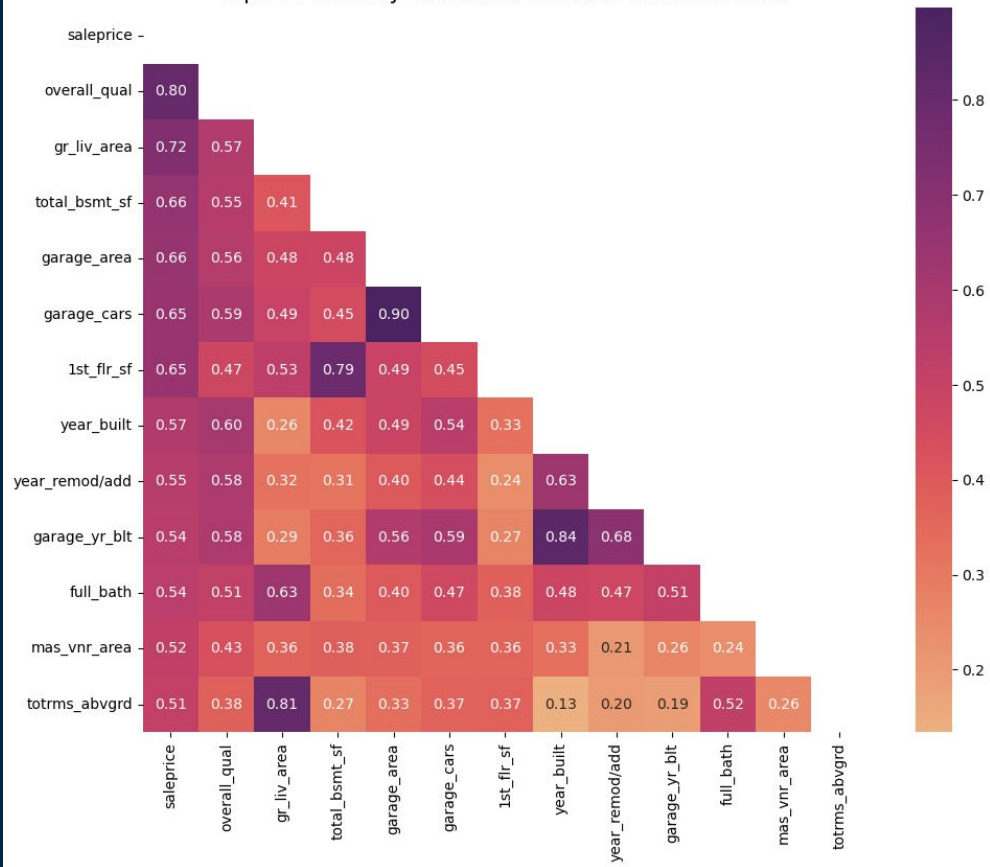
- Agricultural and research center of Iowa
- Population: 66,000
- Real Estate Market: over 3,000 tracked sales from 2006-2010
- Ames Assessor's Office: tracks **81** features, including **SalePrice**

EDA



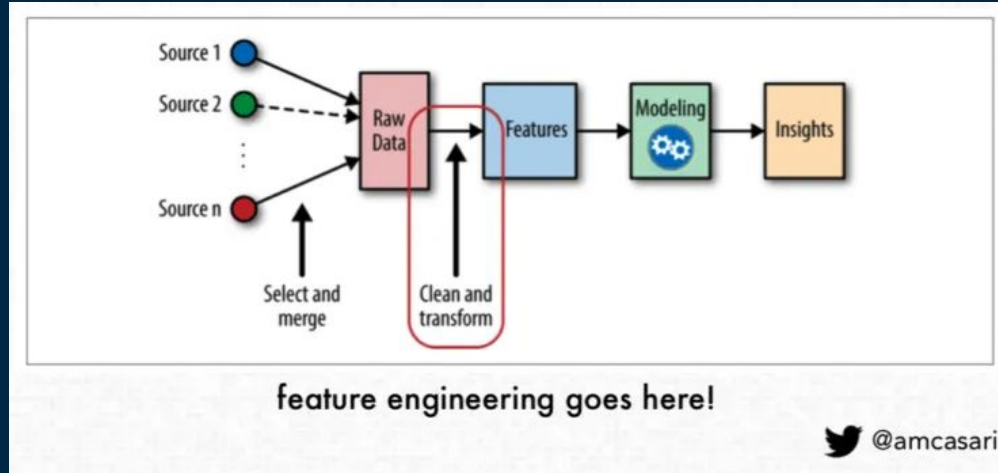
- Normality in Histograms
- Feature vs Price
- Null values?

Top 13 Positively Correlated Features with Sale Price



Correlation Matrix

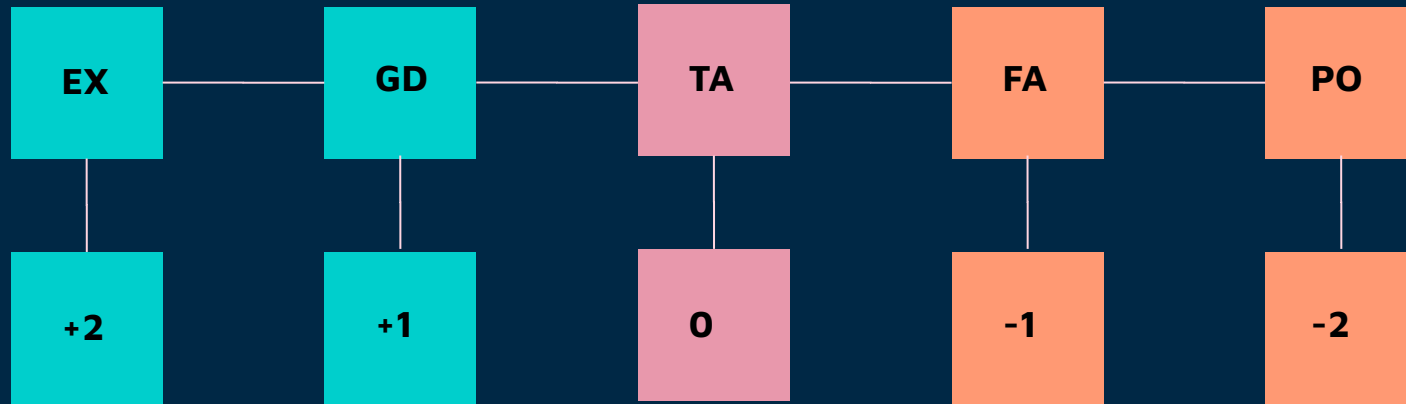
Feature Engineering



Feature Engineering from [Domino Labs](#)

- Nulls
 - Threshold: 25%
 - Imputation: 0/'None'
- Combining Features
 - 'Outdoor Finished Space'
- Dropping Redundant Features
 - High pairwise correlation
 - Contributed to new feature

Encoding Categorical Data



Feature (Ordinal): quality or condition

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

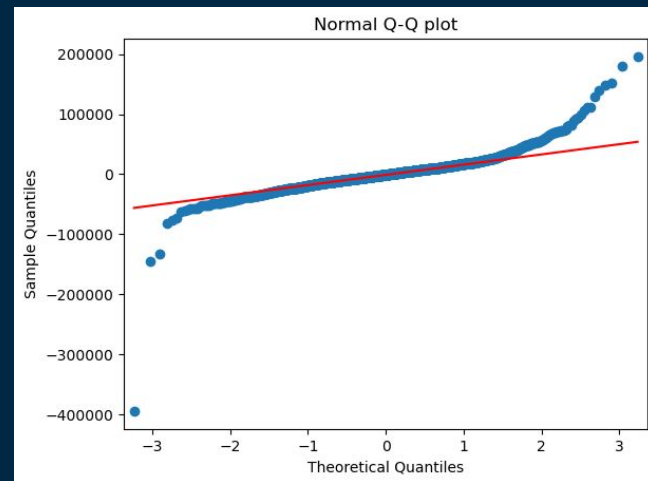
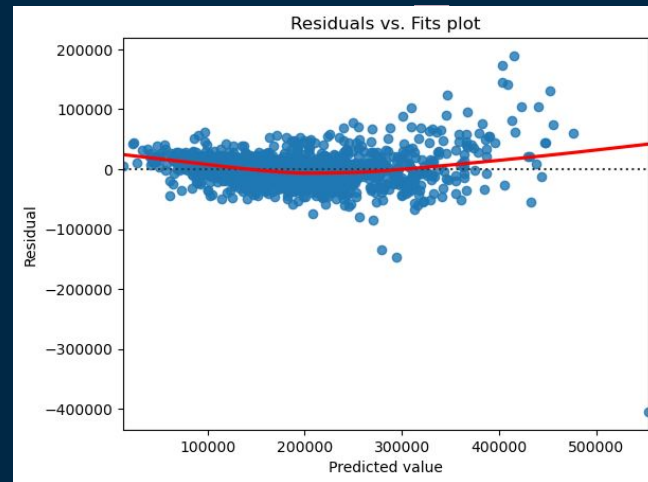
- If 'None' → becomes 0
- If Nominal → dummies

Modeling & Assumptions

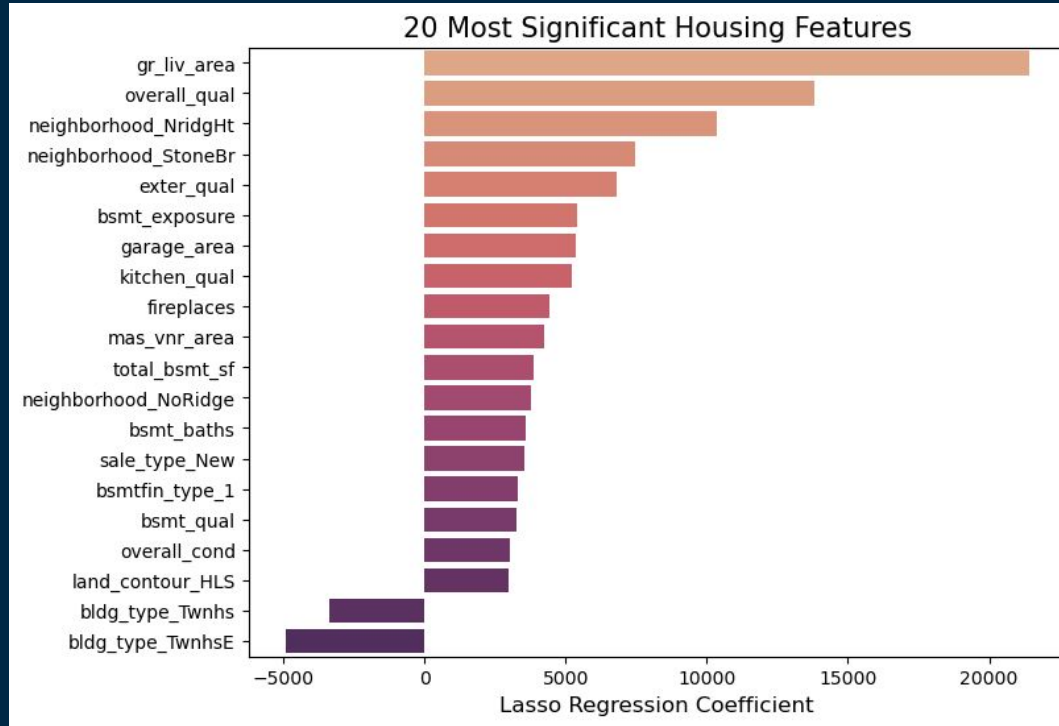
model	cross_val R^2	train R^2	val R^2	train RMSE	val RMSE
baseline	NaN	NaN	-0.006	80726.0	73049.0
LinReg	-6.95e+22	0.927	0.756	21831.0	35992.0
Ridge	0.864	0.900	0.811	25533.0	31643.0
Lasso	0.854	0.895	0.822	26212.0	30691.0
ElasticNet	0.865	0.901	0.811	25361.0	31703.0

Comparing Models

- All are overfitting
- Lasso: balance good score/high variance
- Kaggle: 29468.95
- Assumptions of SLR are met
- Multicollinearity: high VIF's still



Conclusions & Recommendations



- Sellers
 - Remodel spaces that do not contribute to 'living area'
 - Improve quality of finish (stucco)
- Buyers
 - Townhomes are more bang for your buck
- Zillow
 - Improve model for \$\$



Do you have any questions?

jahnavi.kal@gmail.com
General Assembly DSIR 10/31

THANKS



CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)