Predicting Value based on Basements and Proximity to ISU

Dominic Martorano

DSB-EC-1211

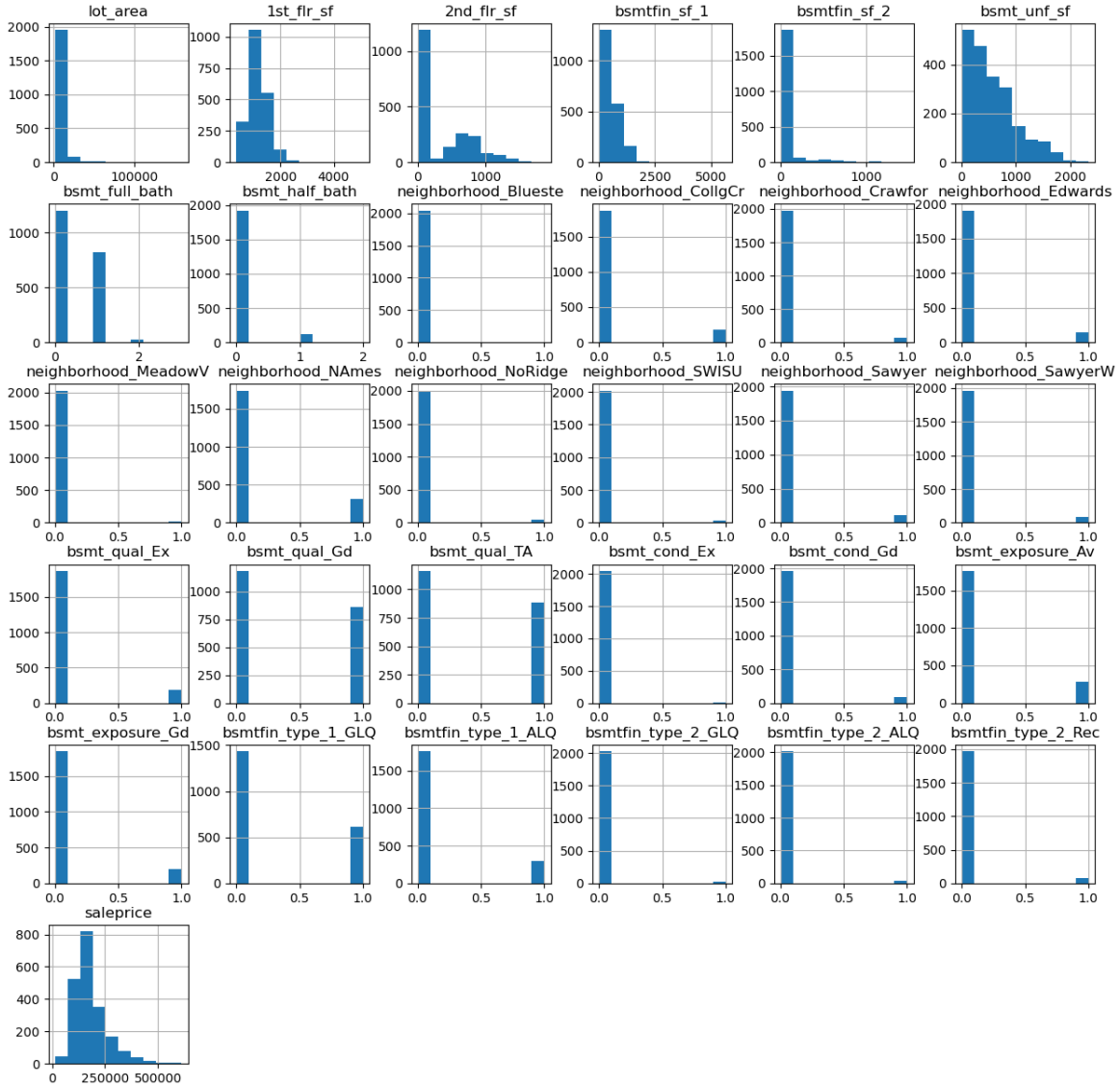
Problem Statement

- Real estate agent in Ames needs model predicting sale price based on location in Ames and basement usability

Breaking Down Data

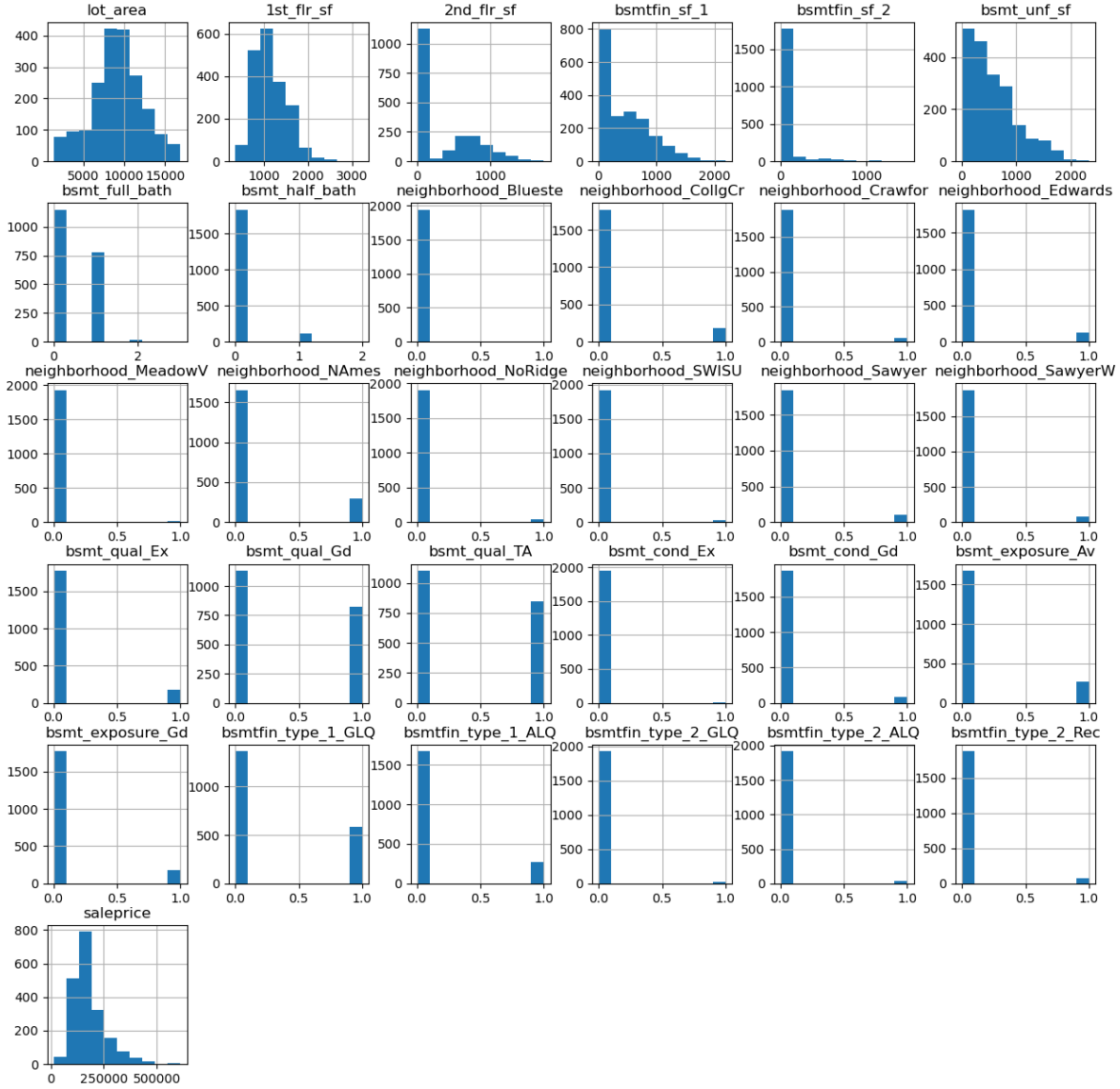
- Model to be created with data based around neighborhood and basement-related data
- All categorical data to be transformed into dummy variables to promote modelability

Identifying Outliers



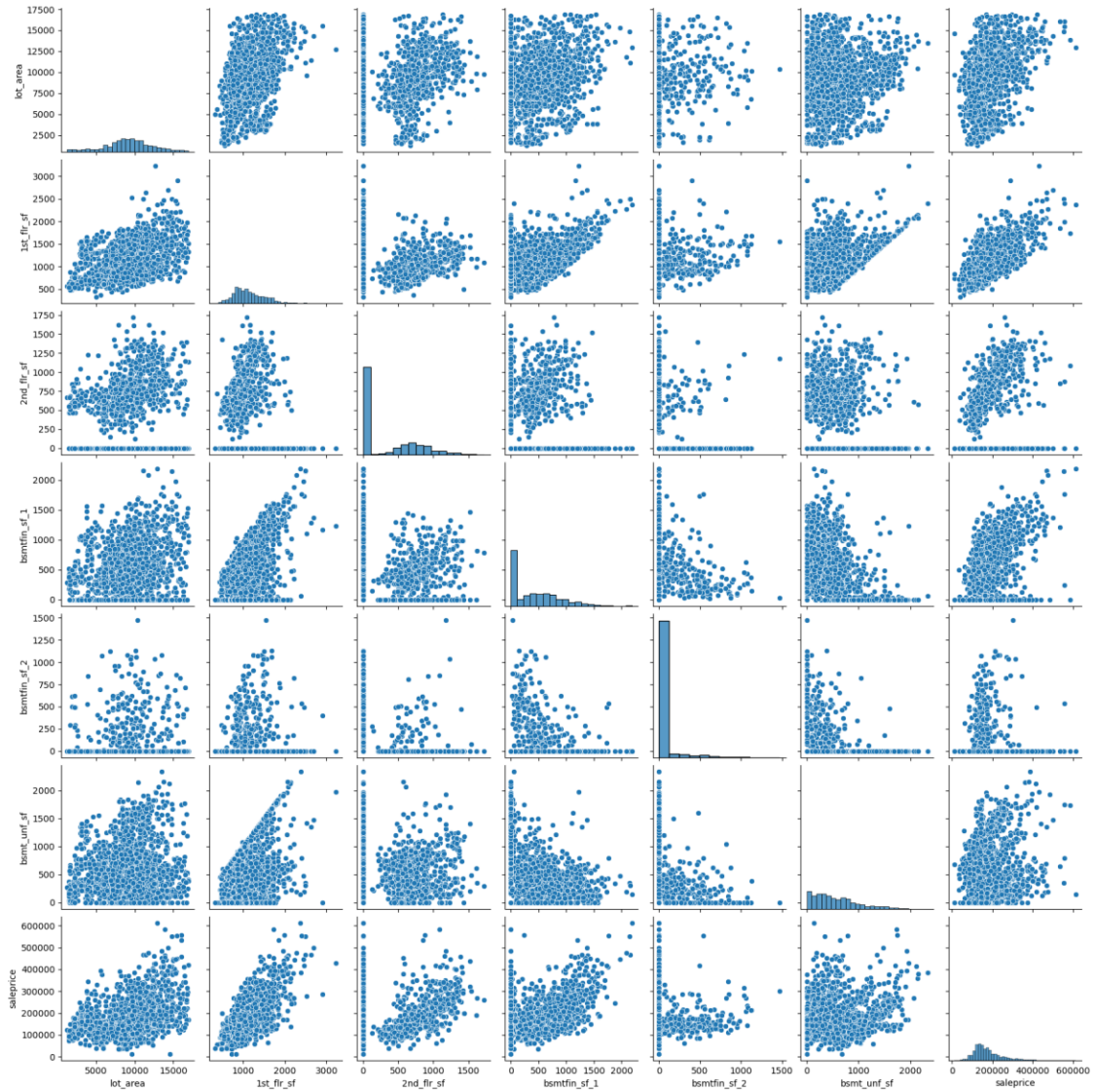
- Features related to area have heavy left-side skew, likely due to large lots/homes

Addressing Outliers



- Some homes still on fairly large lots, other variables still with large left-side skew

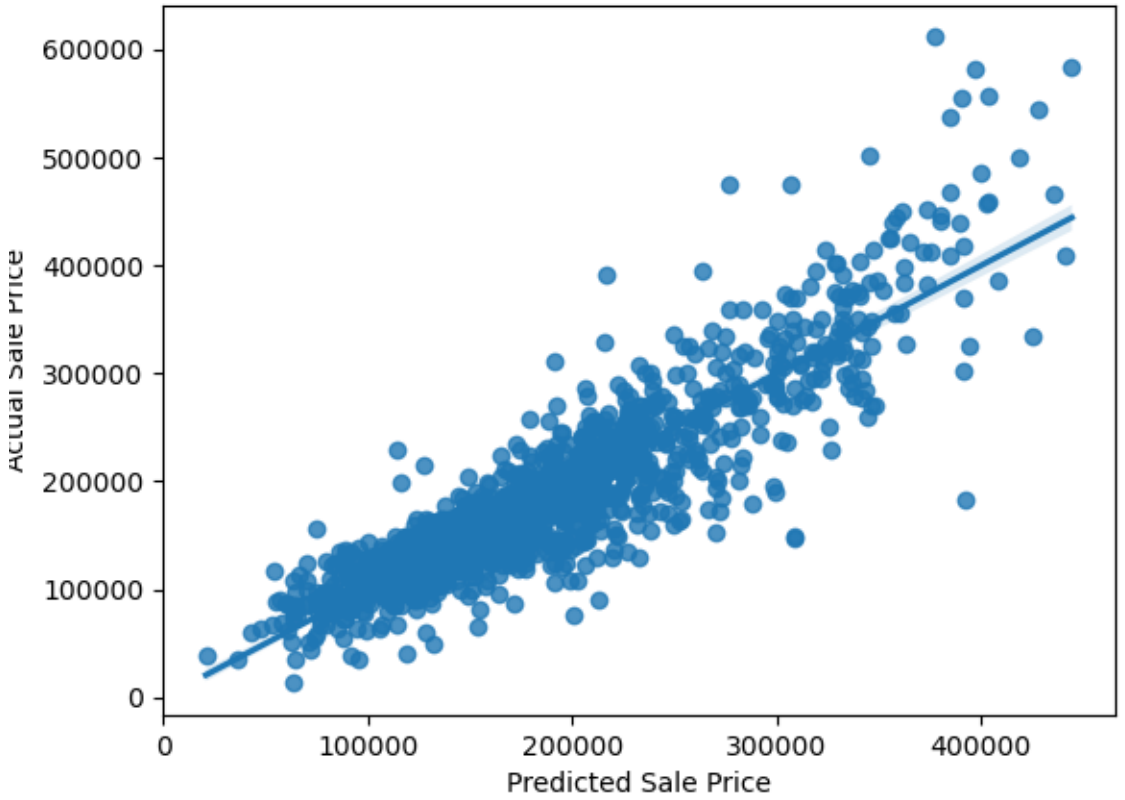
Possible Interactions



- Most prevalent interaction observed between 1st floor and basement square footage

Model Selection and Evaluation

- Tested cross-validation scores for LR, LASSO, and Ridge
- All scores around 71, but Ridge was highest
- Fit and evaluated Ridge to find that model was only very slightly overfit



Conclusions

- Area-related features found to be most influential
- Basement-related features all had positive impacts on sale price
- Neighborhood features impacts more variable and less impactful