AMES Housing Project
The price modeling process

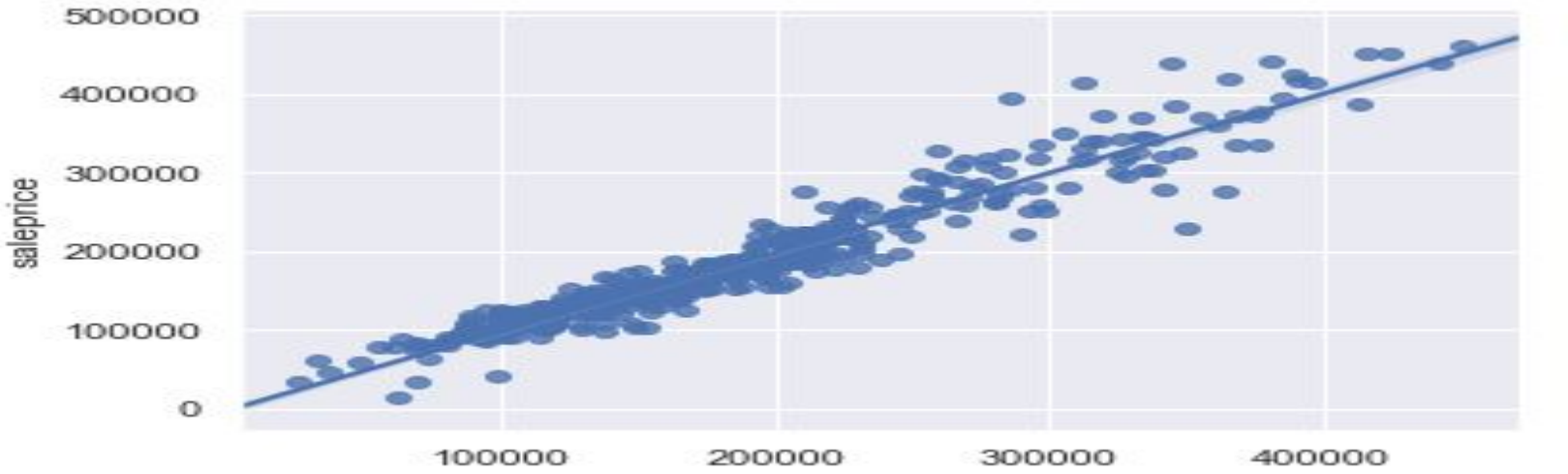
Project Scope

- High-level exploration of data
- Identification of missing data
- Exploring options to impute data
- Correlation analysis
- Selection of features
- Interaction terms
- Model evaluation and iteration

Data Cleaning

- Value_counts(): values distribution
- Data dictionary: value options

The more features, the better ?

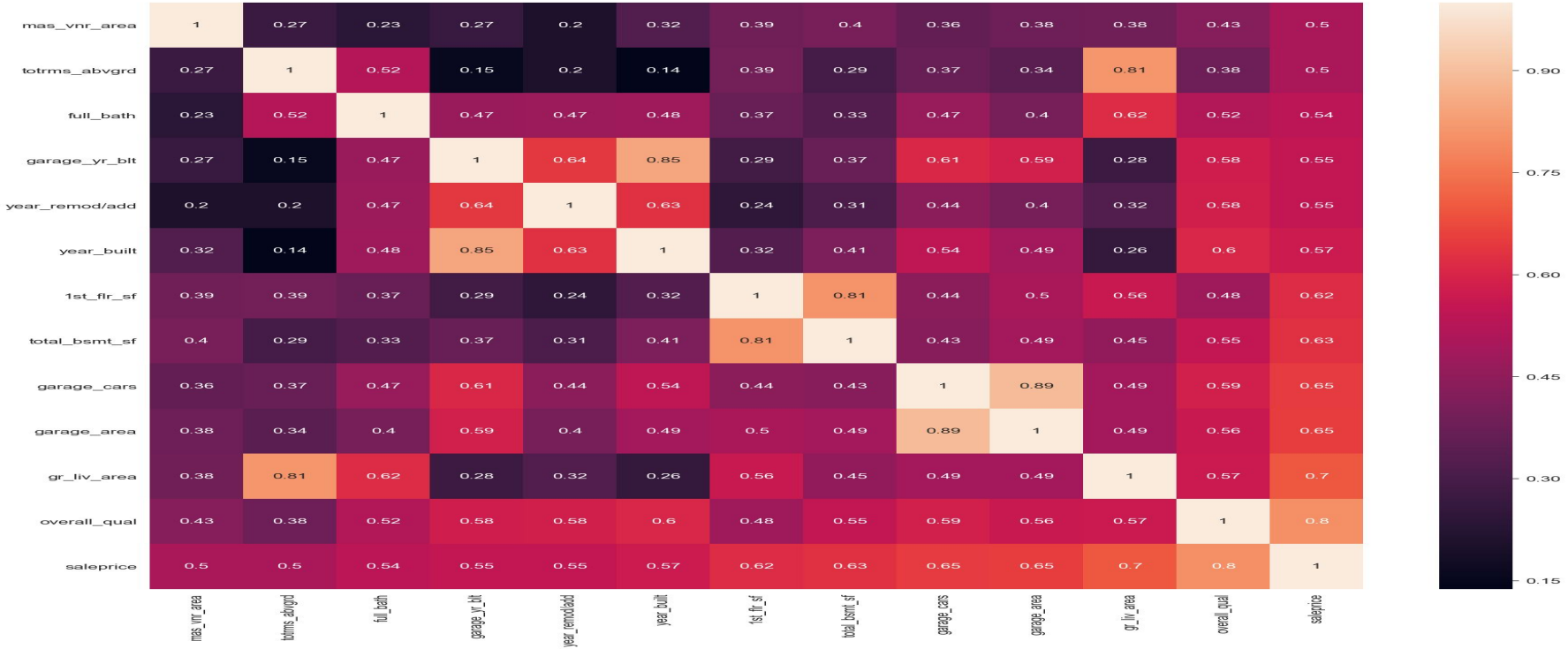


Model overfit to the training set

The basics … Correlation analysis

Only 13 features were deemed having strong correlation with the `saleprices`

Collinearity is a problem



Additional iteration … dropping four more features. The model improves.

Poly boost

PolynomialFeatures to play the interactions between features.

Recommendations and next ...

- Less is more!
- Correlation is your friend
- Beware of multicollinearity
- Let interactions play
- Use common sense

Next:

- Test logical interactions
- Fine tune Hyperparameters with Grid search