

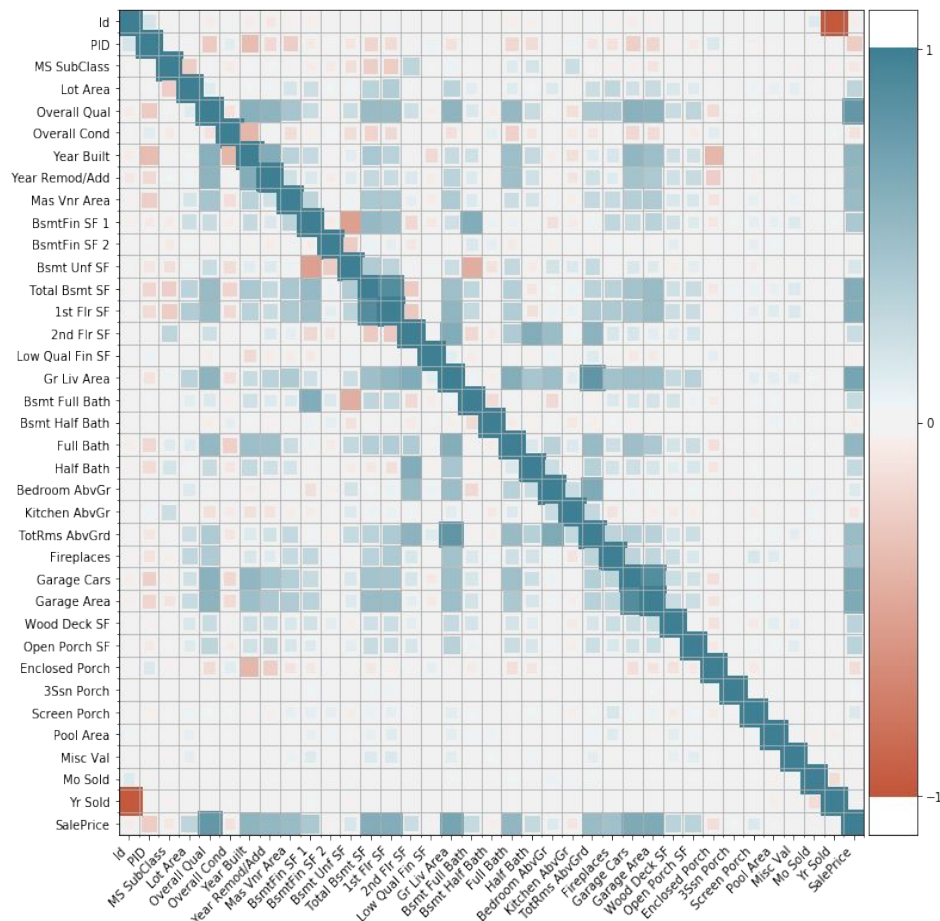
GA Project 2

Predicting housing price in Ames, IA

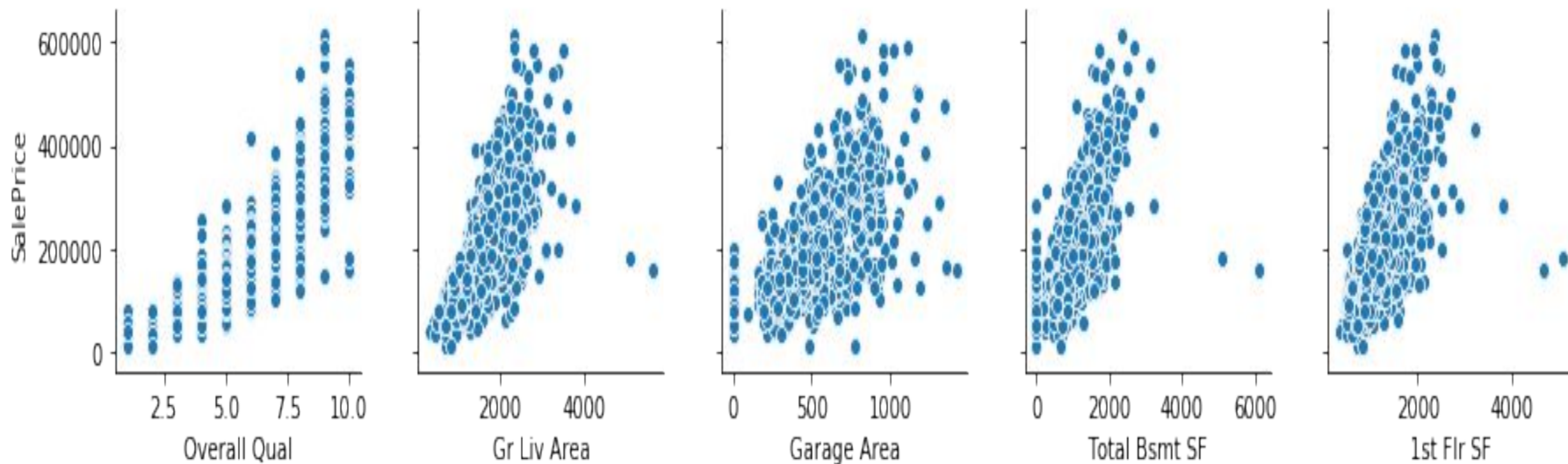


Exploratory Analysis

- Plot a heat map to see which columns are most highly correlated with sale price.
- Custom function adjusts size of squares according to magnitude of corr.
- Much easier to spot which variables are most correlated.
- credit to Drazen Zaric



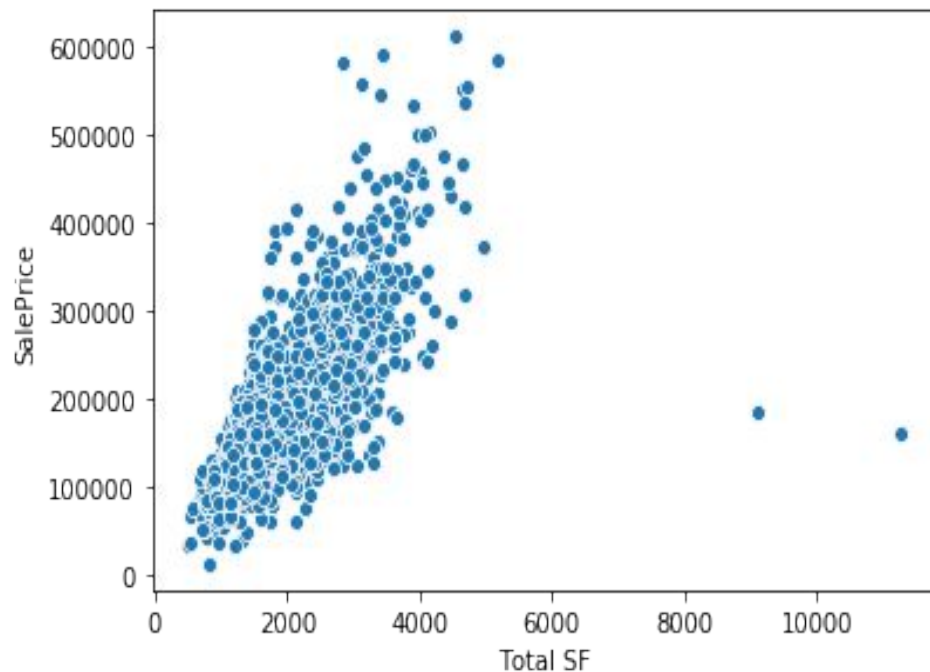
Most important numerical features.



These are the numerical features we will use for regression model. For simplicity we have decided to focus on numerical qualities as a baseline. We will use dummy variables to incorporate categorical features in the future.

Feature Engineering

- We create a column called 'Total SF' = '1st Floor SF' + '2nd Floor SF' + 'BsmtFin SF 1' + 'BsmtFin SF 2'
- We remove outliers where Total SF > 8000



Simple Linear Regression

- MSE = 129033484
- Ranked 92 out of 94...
- Try incorporating more features
- Use more comprehensive feature engineering
- Try averaging results from different models
- Use Grid search to tune parameters

