I began by reading the data in and studying the features. I compared a few pair plots and discovered a discrepancy in one of the factors wherein almost all observations, except one, were the same. The variable in question was the utilities variable. Only one observation, ID 945, was different which seems to imply that the home doesn’t receive sewage.

for i in range(len(df['Utilities'])):

if df['Utilities'][i]!='AllPub':

print(df.iloc[i])

Multicollinearity is not immense at first glance aside from a few variables. One variable of concern is the year the garage is built inflating variance due to collinearity with YearBuilt.

