Atefeh Moradan	**1820892**	**Kaggle accounts: atefehhh, Parastoo91**
Samin Hamidi	**1732304**	**https://www.kaggle.com/atefehhh** **https://www.kaggle.com/parastoo91**
Babak Habibi	**1799271**	**Foundation of Data Science Project Report**

Data Tidying	1) NAs removed 34 columns with NAs. All the NAs were removed. Methods used: Mean/Mode Imputation Taking other logically related columns into consideration 2) Outliers removed Data visualization to find outliers: box pilots 3) Converting categorical variables to numeric All turned into dummy
Algorithm	XGBoosting
Score	0.132

Data Tidying	1) NAs removed 34 columns with NAs. All the NAs were removed. Methods used: Mean/Mode Imputation Taking other logically related columns into consideration 2) Outliers removed var X > mean(X) + 3*sd(X) Data visualization to find outliers: box pilots 3) Converting categorical variables to numeric Ordinal instances were label encoded Nominal instances were turn into dummy 4) Data normalization Predictor was normalized The log of all numeric variables with skewness larger than 0.5 was taken
Algorithm	Lasso
Score	0.12347

Data Tidying	**1) NAs removed** 34 columns with NAs. All the NAs were removed. Methods used: Mean/Mode Imputation Taking other logically related columns into consideration **2) Outliers removed** Data visualization to find outliers: box pilots **3) Converting categorical variables to numeric** Ordinal instances were label encoded Nominal instances were turn into dummy **4) Data normalization** Predictor was normalized Those features with skewness higher than 0.5 were fed into BoxCoxTrans function **5) Feature Engineering** We added two new features: **TotalPorchSF** (combining and thus consolidating variables which are related regarding porches) **Age** (how old our house is: 2010(year houses were sold)-yearbuilt) **TotalSF**(adds up all the living spaces) **Variables removed :** • (due to multicollinearity): 'YearRemodAdd', 'GarageYrblt', 'YearBuilt' • (Unbalanced variables): 'utilities', 'street', 'functional', 'electrical', 'saletype'
Algorithm	Lasso
Score	**0.11772**