Group\_EDA\_histgram:

* Plot the distribution of target variable
* Clean and transform 5 features
* Do the plot for categorical variables and numeric variable separately
* Save new data as train\_v1.csv

Group\_EDA\_missing:

* Calculate the percentage of missing value
* Filled the missing value with median for every feature
* Training data extreme values/outliers cleaning
* Save new data as train\_v2\_clean.csv

Group\_FE\_encoding:

* Encoding the categorical variables

Group\_FE\_polynomial\_features:

* Generate polynomial features based on several original features

Group\_FE\_domain\_knowledge:

* Generate domain knowledge features (there are many features, please look at my script)
* The final version of data is train\_v4\_clean\_polynomial\_domain.csv so far