R/Python ML Challenge

Use the following Lending Club Loan Dataset from Kaggle and download the data set named **Lending_Club_v2.csv.** This competition dataset was used to predict Loan Default (learn more). Using this data create an R/Python script or notebook that does the following:

1. EDA
2. Do feature reduction
3. Transform any characteristics or categorical variables into numeric.
4. Do Feature selection
5. Build a model - `xgBoost` with the target being defined below.
6. Show train and test results.
7. It will be helpful if you add comments (headers) for each section of the code.

Assuming the dataset is imported in **R** as a dataframe define the target as follow

```
bad_indicators <- c("Charged Off ",
                    "Charged Off",
                    "Default",
                    "Does not meet the credit policy. Status:Charged Off",
                    "In Grace Period",
                    "Default Receiver",
                    "Late (16-30 days)",
                    "Late (31-120 days)")
ds$target <- ifelse(ds$loan_status %in% bad_indicators,1,0)
```

If you are working in **Python**, you can define the target as follows:

```
bad_indicators  =  ["Charged Off ",
                    "Charged Off",
                    "Default",
                    "Does not meet the credit policy. Status:Charged Off",
                    "In Grace Period",
                    "Default Receiver",
                    "Late (16-30 days)",
                    "Late (31-120 days)"]
df['target'] = df['loan_status'].isin(bad_indicators).astype(int)
```

Note: There is no need for extensive feature manipulation or model iteration. We are looking for clear code that performs each of the steps mentioned before. Once you have the code in github, please send us the link.