DSC680 Project1 - College Recommendation Engine - Part1

Amie Davis

9 April, 2021

Data Sources:

Scorecard data from the US Department of Education https://collegescorecard.ed.gov/data/

References:

https://www.listendata.com/2015/06/simplest-dimensionality-reduction-with-r.html

Load Libraries

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
library(dplyr)
library(plyr)
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.0.2
```

1. Import the Data

```
# Load college data
college_df <- read_csv("Data/Scorecard/MERGED2018_19_PP.csv",
  na =c("NULL", "PrivacySuppressed")
)
```

2. Retain specific features

```
retain_df <- college_df %>% select(
UNITID,    #Unique Identifier
SAT_AVG,
ACTCMMID,    #ACT Mid-point
TUITIONFEE_OUT,    #Out of state tuition
LOCALE,
LATITUDE,
LONGITUDE,
ADM_RATE_ALL,
UGDS,     #Size
```

```
DISTANCEONLY,      #Online Flag
C200_4_POOLED,     # 4 Yr Completion Rate
RET_PT4_POOLED_SUPP,     # 3 Yr Retention Rate
DBRR1_FED_UGNOCOMP_RT,
DBRR4_FED_UGCOMP_RT,
DBRR1_PP_UG_RT,
C100_4_POOLED,
FTFTPCTFLOAN_POOLED_SUPP,
DBRR5_FED_UG_RT,
DBRR5_PP_UG_RT,
DBRR10_FED_UG_RT,
DBRR10_PP_UG_RT
)
```

```
# Handle NPT41_PRIV
# Private school students with reported income < $30
# Public schools show as NA
# Set to same value as public field
college_df$NPT41_PRIV[is.na(college_df$NPT41_PRIV)] <- college_df$NPT41_PUB[is.na(college_df$NPT41_PRIV

# Set negative outlier to zero
college_df$NPT41_PRIV[college_df$NPT41_PRIV < 0] <- 0
```

3. Feature Reduction

```
# Original matrix dimensions
dim(college_df)
```

```
## [1] 6806 2383
```

```
# Remove columns with only NULL values
college_df <- Filter(function(x)!all(is.na(x)), college_df)

# Remove columns with only 0 values
college_df[, -which(numcolwise(sum)(college_df) == 0)]
```

```
## # A tibble: 6,806 x 0
```

```
dim(college_df)
```

```
## [1] 6806  951
```

```
# Remove unnecessary features
  college_df[ ,c(
    "T4APPROVALDATE", #Approval Date
"INSTURL", # Institution URL
"NPCURL", #Pricing URL
"ACCREDAGENCY",
"ACCREDCODE",
"POOLYRS200",       # Part of Formula
"ICLEVEL",
"CDR3_DENOM",       # Part of Formula
"POOLYRSRET_PT",    # Part of Formula
"SCHTYPE",          # Same as CONTROL field
"PLUS_DEBT_ALL_N",  # Part of Formula
```

```
"DBRR1_FED_UGNOCOMP_N",     # Part of Formula
"DBRR4_FED_UGCOMP_NUM",     # Part of Formula
"DBRR4_FED_UGCOMP_DEN",
"DBRR1_PP_UG_DEN",      # Part of Formula
"DBRR1_PP_UG_NUM",
"DBRR4_PP_UG_DEN",
"DBRR4_PP_UG_NUM",
"DBRR4_FED_UG_NUM",
"DBRR4_FED_UG_DEN",
"POOLYRS100",       # Part of Formula
"FTFTPCTFLOAN_POOLED_SUPP",
"DBRR5_FED_UG_NUM",
"DBRR5_FED_UG_DEN",
"DBRR5_PP_UG_NUM",
"DBRR5_PP_UG_DEN",
"DBRR10_FED_UG_DEN",
"DBRR10_FED_UG_NUM",
"DBRR10_PP_UG_NUM",
"DBRR10_PP_UG_NUM"
)] <- list(NULL)
```

```
warnings()
```

```
# Remove variables having high missing percentage (50%)
college_df <- college_df[, colMeans(is.na(college_df)) <= .5]
dim(college_df)
```

```
## [1] 6806  527
```

```
# Remove Zero and Near Zero-Variance Predictors
#    install.packages("caret")
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.4
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
nzv <- nearZeroVar(college_df)
college_df <- college_df[, -nzv]
dim(college_df)
```

```
## [1] 6806  358
```

```
# Look for correlation

# Separate numeric columns
num_df <- select_if(college_df, is.numeric)

# Replace null values with 99999
num_df[is.na(num_df)] <- 99999

# Calculate correlation matrix
cor_matrix <- cor(num_df)

# Find attributes that are highly corrected
```

```r
highlyCorrelated <- findCorrelation(cor_matrix, cutoff=0.7)

# Identifying Variable Names of Highly Correlated Variables
highlyCorCol <- colnames(num_df)[highlyCorrelated]

# Print highly correlated attributes
#highlyCorCol

# Remove highly correlated variables and create a new dataset
college_df <- college_df[, -which(colnames(college_df) %in% highlyCorCol)]
dim(college_df)
```

```
## [1] 6806   28
```

```r
#Add back in retained columns
college_df <- cbind(retain_df, college_df)
dim(college_df)
```

```
## [1] 6806   49
```

4. Apply filters

```r
# Create subset to include only institutions that primarily offer bachelor's degrees
college_df <- filter(college_df, PREDDEG == 3)
college_df[ ,c("PREDDEG")] <- list(NULL)

# Create subset to remove colleges
#college_df <- filter(college_df, PREDDEG == 3)

dim(college_df)
```

```
## [1] 2058   48
```

5. Re-Label data fields

```r
names(college_df)[names(college_df) == "ACTCMMID"] <- "ACT_MEDIAN"
names(college_df)[names(college_df) == "TUITIONFEE_OUT"] <- "TUITION_OUT_ST"
names(college_df)[names(college_df) == "UGDS"] <- "SIZE"
names(college_df)[names(college_df) == "DISTANCEONLY"] <- "ONLINE_ONLY"
names(college_df)[names(college_df) == "INSTNM"] <- "SCHOOL_NAME"
names(college_df)[names(college_df) == "MAIN"] <- "MAIN_CAMPUS"
names(college_df)[names(college_df) == "NUMBRANCH"] <- "NUM_BRANCH"
names(college_df)[names(college_df) == "CONTROL"] <- "OWNERSHIP"
names(college_df)[names(college_df) == "NUM41_PRIV"] <- "PVT_INCOME_0_30K"
names(college_df)[names(college_df) == "NUM42_PRIV"] <- "PVT_INCOME_30_40K"
names(college_df)[names(college_df) == "TUITIONFEE_IN"] <- "TUITION_IN_ST"
names(college_df)[names(college_df) == "INEXPFTE"] <- "INSTR_EXP_PER_FTE"
names(college_df)[names(college_df) == "DEBT_N"] <- "MEDIAN_DEBT"
names(college_df)[names(college_df) == "C200_4_POOLED"] <- "Y4_COMPLETION_RT"
names(college_df)[names(college_df) == "CDR3"] <- "Y3_LN_DEFAULT_RT"
names(college_df)[names(college_df) == "DEBT_N"] <- "MEDIAN_DEBT"
names(college_df)[names(college_df) == "C100_4_POOLED"] <- "Y4_COMPLETION_RT_POOLED"
names(college_df)[names(college_df) == "RET_PT4_POOLED_SUPP"] <- "PT_RETENTION_RT"
```

```
names(college_df)[names(college_df) == "PLUS_DEBT_INST_MD"] <- "MEDIAN_PLUS_LN_DEBT"
names(college_df)[names(college_df) == "PLUS_DEBT_ALL_COMP_MD_SUPP"] <- "MEDIAN_PLUS_LN_DEBT_GRADS"
names(college_df)[names(college_df) == "DBRR1_FED_UGNOCOMP_RT"] <- "UG_INCOMP_1Y_REPAY_RT"
names(college_df)[names(college_df) == "DBRR4_FED_UGCOMP_RT"] <- "UG_GRAD_4Y_REPAY_RT"
names(college_df)[names(college_df) == "DBRR1_PP_UG_RT"] <- "UG_PLUS_1Y_REPAY_RT"
names(college_df)[names(college_df) == "DBRR4_PP_UG_RT"] <- "UG_PLUS_4Y_REPAY_RT"
names(college_df)[names(college_df) == "BBRR2_FED_UG_DFLT"] <- "Y2_LN_DEFAULT_RT"
names(college_df)[names(college_df) == "BBRR2_FED_UG_DLNQ_SUPP"] <- "Y2_LN_DELINQ_RT"
names(college_df)[names(college_df) == "BBRR2_FED_UGCOMP_FBR_SUPP"] <- "UG_GRAD_2Y_FORBEAR_RT"
names(college_df)[names(college_df) == "LPPPLUS_CNT"] <- "NUM_STU_PLUS_LN_BAL"
names(college_df)[names(college_df) == "FEDSCHCD"] <- "FED_SCHOOL_CD"
names(college_df)[names(college_df) == "FTFTPCTPELL_POOLED_SUPP"] <- "PELL_GRANT_AWD_RT"
names(college_df)[names(college_df) == "FTFTPCTFLOAN_POOLED_SUPP"] <- "FED_LN_AWD_RT"
names(college_df)[names(college_df) == "DBRR5_FED_UG_RT"] <- "Y5_UG_REPAY_RT"
names(college_df)[names(college_df) == "DBRR5_PP_UG_RT"] <- "Y5_PLUS_LN_REPAY_RT"
names(college_df)[names(college_df) == "BBRR2_FED_UG_PAIDINFULL_SUPP"] <- "FED_UG_PAIDINFULL"
```

5. Convert data types

```
college_df$ONLINE_ONLY <- factor(college_df$ONLINE_ONLY)
college_df$LOCALE <- factor(college_df$LOCALE)
college_df$MAIN_CAMPUS <- factor(college_df$MAIN_CAMPUS)
college_df$OWNERSHIP <- factor(college_df$OWNERSHIP)
college_df$REGION <- factor(college_df$REGION)
college_df$ST_FIPS <- factor(college_df$ST_FIPS)
```

5. Handle null values

```
# If admission rate is NA, set to 100%
college_df$ADM_RATE_ALL[is.na(college_df$ADM_RATE_ALL)]= 1

# If school doesn't collect test scores, set to 0
college_df$SAT_AVG[is.na(college_df$SAT_AVG)]= 0
college_df$ACT_MEDIAN[is.na(college_df$ACT_MEDIAN)]= 0

# If no rate, set to 0
college_df$Y4_COMPLETION_RT[is.na(college_df$Y4_COMPLETION_RT)]= 0
college_df$PT_RETENTION_RT[is.na(college_df$PT_RETENTION_RT)]= 0
college_df$UG_INCOMP_1Y_REPAY_RT[is.na(college_df$UG_INCOMP_1Y_REPAY_RT)]= 0
college_df$UG_GRAD_4Y_REPAY_RT[is.na(college_df$UG_GRAD_4Y_REPAY_RT)]= 0
college_df$UG_PLUS_1Y_REPAY_RT[is.na(college_df$UG_PLUS_1Y_REPAY_RT)]= 0
college_df$Y4_COMPLETION_RT_POOLED[is.na(college_df$Y4_COMPLETION_RT_POOLED)]= 0
college_df$Y3_LN_DEFAULT_RT[is.na(college_df$Y3_LN_DEFAULT_RT)]= 0
college_df$Y2_LN_DEFAULT_RT[is.na(college_df$Y2_LN_DEFAULT_RT)]= 0
college_df$Y2_LN_DELINQ_RT[is.na(college_df$Y2_LN_DELINQ_RT)]= 0
college_df$UG_GRAD_2Y_FORBEAR_RT[is.na(college_df$UG_GRAD_2Y_FORBEAR_RT)]= 0
college_df$FED_LN_AWD_RT[is.na(college_df$FED_LN_AWD_RT)]= 0
college_df$Y5_UG_REPAY_RT[is.na(college_df$Y5_UG_REPAY_RT)]= 0
college_df$Y5_PLUS_LN_REPAY_RT[is.na(college_df$Y5_PLUS_LN_REPAY_RT)]= 0

college_df$SIZE[is.na(college_df$SIZE)]= 0
```

```
college_df$TUITION_OUT_ST[is.na(college_df$TUITION_OUT_ST)]= 0
college_df$TUITION_IN_ST[is.na(college_df$TUITION_IN_ST)]= 0
college_df$SIZE[is.na(college_df$SIZE)]= 0
college_df$MEDIAN_PLUS_LN_DEBT[is.na(college_df$MEDIAN_PLUS_LN_DEBT)]= 0
college_df$INSTR_EXP_PER_FTE[is.na(college_df$INSTR_EXP_PER_FTE)]= 0
college_df$NUM_STU_PLUS_LN_BAL[is.na(college_df$NUM_STU_PLUS_LN_BAL)]= 0
```

6. Review Features

```
# Use list to review all fields in df
str(college_df, list.len=ncol(college_df))
```

```
## 'data.frame':    2058 obs. of  48 variables:
##  $ UNITID                   : num  100654 100663 100690 100706 100724 ...
##  $ SAT_AVG                  : num  957 1220 0 1314 972 ...
##  $ ACT_MEDIAN               : num  18 25 0 28 18 27 0 21 28 26 ...
##  $ TUITION_OUT_ST           : num  18354 19704 6900 22362 19396 ...
##  $ LOCALE                   : Factor w/ 12 levels "11","12","13",..: 2 2 2 2 2 2 7 2 3 2 ...
##  $ LATITUDE                 : num  34.8 33.5 32.4 34.7 32.4 ...
##  $ LONGITUDE                : num  -86.6 -86.8 -86.2 -86.6 -86.3 ...
##  $ ADM_RATE_ALL             : num  0.899 0.921 1 0.809 0.977 ...
##  $ SIZE                     : num  4990 13186 351 7458 3903 ...
##  $ ONLINE_ONLY              : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...
##  $ Y4_COMPLETION_RT         : num  0.35 0.569 0.31 0.523 0.282 ...
##  $ PT_RETENTION_RT          : num  0.25 0.451 0 0.35 0.25 ...
##  $ UG_INCOMP_1Y_REPAY_RT    : num  1.09 1.03 1.06 1 1.09 ...
##  $ UG_GRAD_4Y_REPAY_RT      : num  1.245 1.036 1.212 0.924 1.252 ...
##  $ UG_PLUS_1Y_REPAY_RT      : num  1.16 1.06 0 1.01 1.17 ...
##  $ Y4_COMPLETION_RT_POOLED  : num  0.0464 0.3324 0.1786 0.2234 0.1042 ...
##  $ FED_LN_AWD_RT            : num  0.729 0.541 0 0.44 0.746 ...
##  $ Y5_UG_REPAY_RT           : num  1.296 1.066 1.249 0.977 1.307 ...
##  $ Y5_PLUS_LN_REPAY_RT      : num  1.171 0.896 0 0.834 1.261 ...
##  $ DBRR10_FED_UG_RT         : num  1.228 0.849 1.027 0.751 1.277 ...
##  $ DBRR10_PP_UG_RT          : num  0.889 0.648 NA 0.518 0.834 ...
##  $ OPEID                    : chr  "00100200" "00105200" "02503400" "00105500" ...
##  $ OPEID6                   : chr  "001002" "001052" "025034" "001055" ...
##  $ SCHOOL_NAME              : chr  "Alabama A & M University" "University of Alabama at Birmingham" "
##  $ CITY                     : chr  "Normal" "Birmingham" "Montgomery" "Huntsville" ...
##  $ STABBR                   : chr  "AL" "AL" "AL" "AL" ...
##  $ ZIP                      : chr  "35762" "35294-0110" "36117-3553" "35899" ...
##  $ MAIN_CAMPUS              : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ NUM_BRANCH               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ OWNERSHIP                : Factor w/ 3 levels "1","2","3": 1 1 2 1 1 1 1 1 1 2 ...
##  $ ST_FIPS                  : Factor w/ 54 levels "1","2","4","5",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REGION                   : Factor w/ 10 levels "0","1","2","3",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ NPT41_PRIV               : num  13893 14550 15322 17561 11344 ...
##  $ PVT_INCOME_30_40K        : num  NA NA 0 NA NA NA NA NA NA 27 ...
##  $ TUITION_IN_ST            : num  9744 8568 6900 10714 11068 ...
##  $ INSTR_EXP_PER_FTE        : num  5384 16454 4206 9242 8527 ...
##  $ Y3_LN_DEFAULT_RT         : num  0.182 0.057 0.11 0.059 0.203 0.051 0.075 0.104 0.036 0.048 ...
##  $ MEDIAN_DEBT              : num  3606 7504 514 3021 3609 ...
##  $ MEDIAN_PLUS_LN_DEBT      : num  14838 16145 0 13524 15351 ...
```

```
##  $ MEDIAN_PLUS_LN_DEBT_GRADS: num  16106 16954 NA 16550 18952 ...
##  $ Y2_LN_DEFAULT_RT         : num  0.1726 0.0604 0.1572 0.055 0.1738 ...
##  $ Y2_LN_DELINQ_RT          : num  0.0748 0.0345 0.0502 0.0314 0.0617 ...
##  $ FED_UG_PAIDINFULL        : num  0.00793 0.05967 NA 0.09295 0.00787 ...
##  $ UG_GRAD_2Y_FORBEAR_RT    : num  0.266 0.156 0 0.1 0.287 ...
##  $ LPSTAFFORD_CNT           : num  31374 56997 4463 19702 34246 ...
##  $ NUM_STU_PLUS_LN_BAL      : num  5201 3727 16 1397 4602 ...
##  $ LPPPLUS_AMT              : num  1.14e+08 8.74e+07 3.10e+05 2.98e+07 9.54e+07 ...
##  $ FED_SCHOOL_CD            : chr  "001002" "001052" "016885" "001055" ...
```

```r
# Summary Statistics
summary(college_df)
```

```
##      UNITID          SAT_AVG         ACT_MEDIAN    TUITION_OUT_ST
##  Min.   :100654   Min.   :   0.0   Min.   : 0.00   Min.   :    0
##  1st Qu.:157089   1st Qu.:   0.0   1st Qu.: 0.00   1st Qu.:14604
##  Median :196242   Median :1042.0   Median :20.00   Median :22425
##  Mean   :223714   Mean   : 700.1   Mean   :14.21   Mean   :24319
##  3rd Qu.:230464   3rd Qu.:1145.8   3rd Qu.:24.00   3rd Qu.:33284
##  Max.   :492962   Max.   :1566.0   Max.   :36.00   Max.   :59430
##
##      LOCALE       LATITUDE       LONGITUDE        ADM_RATE_ALL
##  11     :536   Min.   :13.43   Min.   :-158.06   Min.   :0.0000
##  21     :416   1st Qu.:34.75   1st Qu.: -95.33   1st Qu.:0.5835
##  13     :269   Median :39.65   Median : -84.51   Median :0.7430
##  12     :248   Mean   :38.02   Mean   : -88.23   Mean   :0.7253
##  32     :192   3rd Qu.:41.70   3rd Qu.: -76.61   3rd Qu.:0.9332
##  33     :110   Max.   :64.86   Max.   : 144.84   Max.   :1.0000
##  (Other):287
##       SIZE         ONLINE_ONLY Y4_COMPLETION_RT PT_RETENTION_RT
##  Min.   :    0.0   0:2028      Min.   :0.0000   Min.   :0.0000
##  1st Qu.:  599.2   1:  30      1st Qu.:0.3206   1st Qu.:0.0000
##  Median : 1650.5               Median :0.4968   Median :0.0000
##  Mean   : 4397.3               Mean   :0.4740   Mean   :0.1422
##  3rd Qu.: 4524.8               3rd Qu.:0.6500   3rd Qu.:0.3135
##  Max.   :88921.0               Max.   :1.0000   Max.   :0.9565
##
##  UG_INCOMP_1Y_REPAY_RT UG_GRAD_4Y_REPAY_RT UG_PLUS_1Y_REPAY_RT
##  Min.   :0.0000        Min.   :0.0000      Min.   :0.0000
##  1st Qu.:0.9208        1st Qu.:0.7442      1st Qu.:0.8868
##  Median :0.9926        Median :0.8906      Median :0.9782
##  Mean   :0.8500        Mean   :0.8205      Mean   :0.8521
##  3rd Qu.:1.0281        3rd Qu.:1.0095      3rd Qu.:1.0402
##  Max.   :1.0989        Max.   :1.3177      Max.   :1.2336
##
##  Y4_COMPLETION_RT_POOLED FED_LN_AWD_RT    Y5_UG_REPAY_RT   Y5_PLUS_LN_REPAY_RT
##  Min.   :0.0000          Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.1512          1st Qu.:0.3305   1st Qu.:0.7402   1st Qu.:0.5677
##  Median :0.3140          Median :0.5895   Median :0.9046   Median :0.6975
##  Mean   :0.3384          Mean   :0.5131   Mean   :0.8473   Mean   :0.6431
##  3rd Qu.:0.5087          3rd Qu.:0.7372   3rd Qu.:1.0375   3rd Qu.:0.8252
##  Max.   :1.0000          Max.   :1.0000   Max.   :1.3762   Max.   :1.3787
##
##  DBRR10_FED_UG_RT DBRR10_PP_UG_RT     OPEID              OPEID6
##  Min.   :0.0000   Min.   :0.0000   Length:2058        Length:2058
```

```
##  1st Qu.:0.4883   1st Qu.:0.5029   Class :character   Class :character
##  Median :0.6515   Median :0.5959   Mode  :character   Mode  :character
##  Mean   :0.6675   Mean   :0.5902
##  3rd Qu.:0.8425   3rd Qu.:0.6752
##  Max.   :1.3692   Max.   :1.0940
##  NA's   :207      NA's   :362
##  SCHOOL_NAME            CITY              STABBR              ZIP
##  Length:2058        Length:2058        Length:2058        Length:2058
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  MAIN_CAMPUS   NUM_BRANCH     OWNERSHIP    ST_FIPS         REGION
##  0: 214      Min.   : 1.000   1: 608    36     : 172   5      :482
##  1:1844      1st Qu.: 1.000   2:1275    6      : 147   2      :399
##              Median : 1.000   3: 175    42     : 130   3      :295
##              Mean   : 2.709             48     : 102   8      :230
##              3rd Qu.: 1.000             25     :  77   4      :210
##              Max.   :74.000             39     :  77   1      :164
##                                         (Other):1353   (Other):278
##    NPT41_PRIV    PVT_INCOME_30_40K TUITION_IN_ST   INSTR_EXP_PER_FTE
##  Min.   :    0   Min.   :  0.00    Min.   :    0   Min.   :     0
##  1st Qu.:10576   1st Qu.:  8.00    1st Qu.: 8980   1st Qu.:  5907
##  Median :15258   Median : 27.00    Median :16541   Median :  8512
##  Mean   :15767   Mean   : 36.03    Mean   :21140   Mean   : 10927
##  3rd Qu.:19947   3rd Qu.: 48.00    3rd Qu.:32008   3rd Qu.: 12144
##  Max.   :54584   Max.   :331.00    Max.   :59430   Max.   :161644
##  NA's   :226     NA's   :758
##  Y3_LN_DEFAULT_RT   MEDIAN_DEBT    MEDIAN_PLUS_LN_DEBT
##  Min.   :0.00000   Min.   :   10   Min.   :    0
##  1st Qu.:0.02800   1st Qu.:  594   1st Qu.: 9216
##  Median :0.05500   Median : 1322   Median :15835
##  Mean   :0.06583   Mean   : 5020   Mean   :17848
##  3rd Qu.:0.09000   3rd Qu.: 4206   3rd Qu.:24583
##  Max.   :0.50000   Max.   :93095   Max.   :84671
##                    NA's   :129
##  MEDIAN_PLUS_LN_DEBT_GRADS Y2_LN_DEFAULT_RT  Y2_LN_DELINQ_RT
##  Min.   :  4000            Min.   :0.00000   Min.   :0.00000
##  1st Qu.: 13414            1st Qu.:0.00000   1st Qu.:0.00000
##  Median : 20510            Median :0.04529   Median :0.01758
##  Mean   : 23469            Mean   :0.05110   Mean   :0.02309
##  3rd Qu.: 30310            3rd Qu.:0.07753   3rd Qu.:0.04062
##  Max.   :113945            Max.   :0.30105   Max.   :0.13710
##  NA's   :407
##  FED_UG_PAIDINFULL UG_GRAD_2Y_FORBEAR_RT LPSTAFFORD_CNT    NUM_STU_PLUS_LN_BAL
##  Min.   :0.0060    Min.   :0.00000       Min.   :     10   Min.   :    0.0
##  1st Qu.:0.0624    1st Qu.:0.00000       1st Qu.:   4589   1st Qu.:  257.2
##  Median :0.0951    Median :0.08521       Median :  10954   Median :  876.0
##  Mean   :0.1084    Mean   :0.09815       Mean   :  45938   Mean   : 2217.0
##  3rd Qu.:0.1400    3rd Qu.:0.14811       3rd Qu.:  31458   3rd Qu.: 2159.2
##  Max.   :0.5279    Max.   :0.47761       Max.   :1429109   Max.   :32775.0
##  NA's   :427                             NA's   :89
```

```
##   LPPPLUS_AMT        FED_SCHOOL_CD
##  Min.   :4.942e+04   Length:2058
##  1st Qu.:9.374e+06   Class :character
##  Median :2.613e+07   Mode  :character
##  Mean   :7.359e+07
##  3rd Qu.:6.343e+07
##  Max.   :1.318e+09
##  NA's   :211
```

4. Output Data Frame

```
# Output file for further analysis
write.csv(college_df,"Data/Scorecard/Cleaned_Scorecard.csv")
```