

prosper_clean_up

Chris Richardson

February 18, 2019

```
setwd('C:/Users/crich/Desktop/udacity')

# df <- read.csv('C:/Users/crich/Desktop/udacity/prosperLoanData.csv', sep = ',')
# testDF <- df
testDF <- read.csv('C:/Users/crich/Desktop/udacity/prosperData.csv', sep = ',')

## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =
## dec, : EOF within quoted string

# colnames(df)
dim(testDF)

## [1] 1063    81

81 Features is too much. Which is why I am going to filter any unnecessary columns

testDF <- testDF[,c("ListingNumber", "CreditGrade", "Term", "LoanStatus", "BorrowerAPR", "BorrowerRate",
"EstimatedReturn", "ProsperScore", "ListingCategory.numeric.",
"BorrowerState", "Occupation", "EmploymentStatus", "EmploymentStatusDuration",
"IsBorrowerHomeowner", "CurrentlyInGroup", "GroupKey", "DateCreditPulled",
"CreditScoreRangeLower", "CreditScoreRangeUpper", "FirstRecordedCreditLine",
"CurrentCreditLines", "OpenCreditLines", "TotalCreditLinespast7years",
"CurrentDelinquencies", "AmountDelinquent", "DelinquenciesLast7Years",
"PublicRecordsLast10Years", "PublicRecordsLast12Months", "RevolvingCreditBalance",
"BankcardUtilization", "AvailableBankcardCredit", "TotalTrades", "TradesNeverDelinquent..percentage.",
"TradesOpenedLast6Months", "DebtToIncomeRatio", "IncomeRange", "IncomeVerifiable",
"StatedMonthlyIncome", "TotalProsperPaymentsBilled", "OnTimeProsperPayments",
"ProsperPaymentsLessThanOneMonthLate", "ProsperPaymentsOneMonthPlusLate", "ProsperPrincipalBorrowed",
"ProsperPrincipalOutstanding", "ScorexChangeAtTimeOfListing", "LoanCurrentDaysDelinquent",
"LoanFirstDefaultedCycleNumber", "LoanMonthsSinceOrigination", "LoanOriginalAmount",
"MonthlyLoanPayment", "PercentFunded", "Recommendations", "InvestmentFromFriendsCount",
"InvestmentFromFriendsAmount"))]

# Copy original DF before experimenting with DF
commitDF <- testDF

# remove unnecessary chars
commitDF$lowerIncome <- sub("^\\$?([0-9,]+).*", "\\1", commitDF$IncomeRange)

# change text to 0s for Tableau Readability, will omit in Tableau*
commitDF$lowerIncome[commitDF$lowerIncome == "Not displayed"] <- NA
commitDF$lowerIncome[commitDF$lowerIncome == 'Not employed'] <- 0

# remove unnecessary chars
commitDF$upperIncome <- sub("^[-]+-", "\\1", commitDF$IncomeRange)

# change text to 0s for Tableau Readability, will omit in Tableau*
commitDF$upperIncome[commitDF$upperIncome == "Not displayed"] <- NA
commitDF$upperIncome[commitDF$upperIncome == 'Not employed'] <- 0
```

```
# Format number for Tableau
commitDF$upperIncome[commitDF$upperIncome == "$100,000+"] <- "100,000"

write.csv(commitDF, 'C:/users/crich/desktop/udacity/prosperLoanCleanedData.csv', row.names = F)
```