LendingClubLab.R

andre

2022-06-07

# LendingClub Direction  
# The data set was available to the public in the past from this url( http://www.lendingclub.com/info/download-data.action"), which is not the case anymore. Just pretend that you have download the data from YEAR 2007-2011 in a csv format.  
# save your file LoanStats3a.csv Download LoanStats3a.csvin the Introduction to Data Science folder.   
  
  
#Q1: Set the directory to the Introduction to Data Science folder.  
#setwd("C:/Users/lonihagen/Desktop/Intro.Data.Science") ### This is my directory. Change the directory to yours.  
setwd("C:/Users/andre/OneDrive/Desktop/IntroDataSci")  
  
#Q2: read the csv file using read.csv. set "stringAsFactors=FALSE" and create a dataframe called "data"   
data <- data.frame(read.csv("C:/Users/andre/Downloads/LoanStats3a.csv", stringsAsFactors = FALSE))  
  
#Q3: Print the first and the last six rows using "head" and "tail" functions  
head(data, 6)

## Notes.offered.by.Prospectus..https...www.lendingclub.com.info.prospectus.action.  
## 1 id  
## 2   
## 3   
## 4   
## 5   
## 6   
## X X.1 X.2 X.3 X.4 X.5  
## 1 member\_id loan\_amnt funded\_amnt funded\_amnt\_inv term int\_rate  
## 2 5000 5000 4975 36 months 10.65%  
## 3 2500 2500 2500 60 months 15.27%  
## 4 2400 2400 2400 36 months 15.96%  
## 5 10000 10000 10000 36 months 13.49%  
## 6 3000 3000 3000 60 months 12.69%  
## X.6 X.7 X.8 X.9 X.10  
## 1 installment grade sub\_grade emp\_title emp\_length  
## 2 162.87 B B2 10+ years  
## 3 59.83 C C4 Ryder < 1 year  
## 4 84.33 C C5 10+ years  
## 5 339.31 C C1 AIR RESOURCES BOARD 10+ years  
## 6 67.79 B B5 University Medical Group 1 year  
## X.11 X.12 X.13 X.14 X.15 X.16  
## 1 home\_ownership annual\_inc verification\_status issue\_d loan\_status pymnt\_plan  
## 2 RENT 24000 Verified 11-Dec Fully Paid n  
## 3 RENT 30000 Source Verified 11-Dec Charged Off n  
## 4 RENT 12252 Not Verified 11-Dec Fully Paid n  
## 5 RENT 49200 Source Verified 11-Dec Fully Paid n  
## 6 RENT 80000 Source Verified 11-Dec Fully Paid n  
## X.17  
## 1 url  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.18  
## 1 desc  
## 2 Borrower added on 12/22/11 > I need to upgrade my business technologies.<br>  
## 3 Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at. I plan to have it paid off as soon as possible/when i sell my old bike. I only need this money because the deal im looking at is to good to pass up.<br><br> Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at. I plan to have it paid off as soon as possible/when i sell my old bike.I only need this money because the deal im looking at is to good to pass up. I have finished college with an associates degree in business and its takingmeplaces<br>  
## 4   
## 5 Borrower added on 12/21/11 > to pay for property tax (borrow from friend, need to pay back) & central A/C need to be replace. I'm very sorry to let my loan expired last time.<br>  
## 6 Borrower added on 12/21/11 > I plan on combining three large interest bills together and freeing up some extra each month to pay toward other bills. I've always been a good payor but have found myself needing to make adjustments to my budget due to a medical scare. My job is very stable, I love it.<br>  
## X.19 X.20 X.21 X.22 X.23 X.24  
## 1 purpose title zip\_code addr\_state dti delinq\_2yrs  
## 2 credit\_card Computer 860xx AZ 27.65 0  
## 3 car bike 309xx GA 1 0  
## 4 small\_business real estate business 606xx IL 8.72 0  
## 5 other personel 917xx CA 20 0  
## 6 other Personal 972xx OR 17.94 0  
## X.25 X.26 X.27 X.28  
## 1 earliest\_cr\_line inq\_last\_6mths mths\_since\_last\_delinq mths\_since\_last\_record  
## 2 Jan-85 1   
## 3 Apr-99 5   
## 4 1-Nov 2   
## 5 Feb-96 1 35   
## 6 Jan-96 0 38   
## X.29 X.30 X.31 X.32 X.33 X.34 X.35  
## 1 open\_acc pub\_rec revol\_bal revol\_util total\_acc initial\_list\_status out\_prncp  
## 2 3 0 13648 83.70% 9 f 0  
## 3 3 0 1687 9.40% 4 f 0  
## 4 2 0 2956 98.50% 10 f 0  
## 5 10 0 5598 21% 37 f 0  
## 6 15 0 27783 53.90% 38 f 0  
## X.36 X.37 X.38 X.39 X.40  
## 1 out\_prncp\_inv total\_pymnt total\_pymnt\_inv total\_rec\_prncp total\_rec\_int  
## 2 0 5863.155187 5833.84 5000 863.16  
## 3 0 1014.53 1014.53 456.46 435.17  
## 4 0 3005.666844 3005.67 2400 605.67  
## 5 0 12231.89 12231.89 10000 2214.92  
## 6 0 4066.908161 4066.91 3000 1066.91  
## X.41 X.42 X.43 X.44  
## 1 total\_rec\_late\_fee recoveries collection\_recovery\_fee last\_pymnt\_d  
## 2 0 0 0 15-Jan  
## 3 0 122.9 1.11 13-Apr  
## 4 0 0 0 14-Jun  
## 5 16.97 0 0 15-Jan  
## 6 0 0 0 17-Jan  
## X.45 X.46 X.47 X.48  
## 1 last\_pymnt\_amnt next\_pymnt\_d last\_credit\_pull\_d collections\_12\_mths\_ex\_med  
## 2 171.62 18-May 0  
## 3 119.66 16-Oct 0  
## 4 649.91 17-Jun 0  
## 5 357.48 16-Apr 0  
## 6 67.3 18-Apr 0  
## X.49 X.50 X.51 X.52  
## 1 mths\_since\_last\_major\_derog policy\_code application\_type annual\_inc\_joint  
## 2 1 Individual   
## 3 1 Individual   
## 4 1 Individual   
## 5 1 Individual   
## 6 1 Individual   
## X.53 X.54 X.55 X.56 X.57  
## 1 dti\_joint verification\_status\_joint acc\_now\_delinq tot\_coll\_amt tot\_cur\_bal  
## 2 0   
## 3 0   
## 4 0   
## 5 0   
## 6 0   
## X.58 X.59 X.60 X.61 X.62  
## 1 open\_acc\_6m open\_act\_il open\_il\_12m open\_il\_24m mths\_since\_rcnt\_il  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.63 X.64 X.65 X.66 X.67 X.68  
## 1 total\_bal\_il il\_util open\_rv\_12m open\_rv\_24m max\_bal\_bc all\_util  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.69 X.70 X.71 X.72 X.73  
## 1 total\_rev\_hi\_lim inq\_fi total\_cu\_tl inq\_last\_12m acc\_open\_past\_24mths  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.74 X.75 X.76 X.77 X.78  
## 1 avg\_cur\_bal bc\_open\_to\_buy bc\_util chargeoff\_within\_12\_mths delinq\_amnt  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## 6 0 0  
## X.79 X.80 X.81 X.82  
## 1 mo\_sin\_old\_il\_acct mo\_sin\_old\_rev\_tl\_op mo\_sin\_rcnt\_rev\_tl\_op mo\_sin\_rcnt\_tl  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.83 X.84 X.85 X.86  
## 1 mort\_acc mths\_since\_recent\_bc mths\_since\_recent\_bc\_dlq mths\_since\_recent\_inq  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.87 X.88 X.89  
## 1 mths\_since\_recent\_revol\_delinq num\_accts\_ever\_120\_pd num\_actv\_bc\_tl  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.90 X.91 X.92 X.93 X.94 X.95  
## 1 num\_actv\_rev\_tl num\_bc\_sats num\_bc\_tl num\_il\_tl num\_op\_rev\_tl num\_rev\_accts  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.96 X.97 X.98 X.99 X.100  
## 1 num\_rev\_tl\_bal\_gt\_0 num\_sats num\_tl\_120dpd\_2m num\_tl\_30dpd num\_tl\_90g\_dpd\_24m  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.101 X.102 X.103 X.104  
## 1 num\_tl\_op\_past\_12m pct\_tl\_nvr\_dlq percent\_bc\_gt\_75 pub\_rec\_bankruptcies  
## 2 0  
## 3 0  
## 4 0  
## 5 0  
## 6 0  
## X.105 X.106 X.107 X.108  
## 1 tax\_liens tot\_hi\_cred\_lim total\_bal\_ex\_mort total\_bc\_limit  
## 2 0   
## 3 0   
## 4 0   
## 5 0   
## 6 0   
## X.109 X.110 X.111  
## 1 total\_il\_high\_credit\_limit revol\_bal\_joint sec\_app\_earliest\_cr\_line  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.112 X.113 X.114 X.115  
## 1 sec\_app\_inq\_last\_6mths sec\_app\_mort\_acc sec\_app\_open\_acc sec\_app\_revol\_util  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.116 X.117 X.118  
## 1 sec\_app\_open\_act\_il sec\_app\_num\_rev\_accts sec\_app\_chargeoff\_within\_12\_mths  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.119 X.120  
## 1 sec\_app\_collections\_12\_mths\_ex\_med sec\_app\_mths\_since\_last\_major\_derog  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.121 X.122 X.123 X.124 X.125  
## 1 hardship\_flag hardship\_type hardship\_reason hardship\_status deferral\_term  
## 2 N   
## 3 N   
## 4 N   
## 5 N   
## 6 N   
## X.126 X.127 X.128 X.129  
## 1 hardship\_amount hardship\_start\_date hardship\_end\_date payment\_plan\_start\_date  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.130 X.131 X.132  
## 1 hardship\_length hardship\_dpd hardship\_loan\_status  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.133 X.134  
## 1 orig\_projected\_additional\_accrued\_interest hardship\_payoff\_balance\_amount  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.135 X.136 X.137  
## 1 hardship\_last\_payment\_amount disbursement\_method debt\_settlement\_flag  
## 2 Cash N  
## 3 Cash N  
## 4 Cash N  
## 5 Cash N  
## 6 Cash N  
## X.138 X.139 X.140 X.141  
## 1 debt\_settlement\_flag\_date settlement\_status settlement\_date settlement\_amount  
## 2   
## 3   
## 4   
## 5   
## 6   
## X.142 X.143  
## 1 settlement\_percentage settlement\_term  
## 2   
## 3   
## 4   
## 5   
## 6

tail(data, 6)

## Notes.offered.by.Prospectus..https...www.lendingclub.com.info.prospectus.action.  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542 Total amount funded in policy code 1: 460296150  
## 42543 Total amount funded in policy code 2: 0  
## X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9 X.10  
## 42538 6500 6500 0 36 months 8.38% 204.84 A A5 < 1 year  
## 42539 5000 5000 0 36 months 7.75% 156.11 A A3 Homemaker 10+ years  
## 42540   
## 42541   
## 42542   
## 42543   
## X.11 X.12 X.13 X.14  
## 42538 NONE Not Verified 7-Jun  
## 42539 MORTGAGE 70000 Not Verified 7-Jun  
## 42540   
## 42541   
## 42542   
## 42543   
## X.15 X.16 X.17  
## 42538 Does not meet the credit policy. Status:Fully Paid n   
## 42539 Does not meet the credit policy. Status:Fully Paid n   
## 42540   
## 42541   
## 42542   
## 42543   
## X.18  
## 42538 Hi, I'm buying a used car. Anybody on facebook wants to finance me? Thanks  
## 42539 I need to make several improvements around the house - fix garage, fix back fencing, and misc other.  
## 42540   
## 42541   
## 42542   
## 42543   
## X.19 X.20 X.21 X.22 X.23 X.24 X.25 X.26 X.27 X.28 X.29 X.30  
## 42538 other Buying a car 100xx NY 4   
## 42539 other Aroundthehouse 068xx CT 8.81   
## 42540   
## 42541   
## 42542   
## 42543   
## X.31 X.32 X.33 X.34 X.35 X.36 X.37 X.38 X.39 X.40 X.41 X.42 X.43  
## 42538 0 f 0 0 7373.904962 0 6500 873.9 0 0 0  
## 42539 0 f 0 0 5619.76209 0 5000 619.76 0 0 0  
## 42540   
## 42541   
## 42542   
## 42543   
## X.44 X.45 X.46 X.47 X.48 X.49 X.50 X.51 X.52 X.53 X.54 X.55  
## 42538 10-Jun 205.32 10-Jul 7-Aug 1 Individual   
## 42539 10-Jun 156.39 10-Jul 15-Feb 1 Individual   
## 42540   
## 42541   
## 42542   
## 42543   
## X.56 X.57 X.58 X.59 X.60 X.61 X.62 X.63 X.64 X.65 X.66 X.67 X.68 X.69  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542   
## 42543   
## X.70 X.71 X.72 X.73 X.74 X.75 X.76 X.77 X.78 X.79 X.80 X.81 X.82 X.83  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542   
## 42543   
## X.84 X.85 X.86 X.87 X.88 X.89 X.90 X.91 X.92 X.93 X.94 X.95 X.96 X.97  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542   
## 42543   
## X.98 X.99 X.100 X.101 X.102 X.103 X.104 X.105 X.106 X.107 X.108 X.109  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542   
## 42543   
## X.110 X.111 X.112 X.113 X.114 X.115 X.116 X.117 X.118 X.119 X.120 X.121  
## 42538 N  
## 42539 N  
## 42540   
## 42541   
## 42542   
## 42543   
## X.122 X.123 X.124 X.125 X.126 X.127 X.128 X.129 X.130 X.131 X.132 X.133  
## 42538   
## 42539   
## 42540   
## 42541   
## 42542   
## 42543   
## X.134 X.135 X.136 X.137 X.138 X.139 X.140 X.141 X.142 X.143  
## 42538 Cash N   
## 42539 Cash N   
## 42540   
## 42541   
## 42542   
## 42543

#Q4: Investigate the first line. If you think the first row is not necessary, then delete the first row of the dataframe and assign the result to the dataframe "data" The data should now include headers in the first row.  
colnames(data) <- data[1,]  
  
  
#Q5: Related with the #Q4, R program may not know the first row is a header. You may let R know that the first row is a "header." HINT: You can do this by using "colnames" function and assign the first row (data[1, ]) as column names of the data frame. Then delete the old first row so that you don't have the identical row appearing in the first row.  
data <- data[-1,]  
  
# Next, replace "id" column with serial numbers.  
colnames(data)[colnames(data) == "id"] <- "serialNumbers"  
  
#Q6: create a vector with 1 through the number of rows. use "nrows" The length of the rows should be over 42000.  
serialNumberVector <- c(1:nrow(data))  
  
#Q7: replace "id" column with the new vector you just created. Now, "id" column is populated with serial numbers rather than empty.  
data$serialNumbers <- serialNumberVector  
  
# Next, you will drop all the columns starting from "annual\_inc\_joint" (the index of this column is 54) through the end of the columns  
# For your information:  
#a <-grep("annual\_inc\_joint", colnames(data))  
#a  
# will yield 54  
data <- data[,-54:-ncol(data)]  
  
#create a vector containing the names of all the variables that are empty.  
#empty columns are ("member\_id", "emp\_title","url", "desc","mths\_since\_last\_delinq", "mths\_since\_last\_record", "next\_pymnt\_d", "mths\_since\_last\_major\_derog")  
#list <-c("member\_id", "emp\_title","url", "desc","mths\_since\_last\_delinq", "mths\_since\_last\_record", "next\_pymnt\_d", "mths\_since\_last\_major\_derog")  
#list  
# And the column indices of the empty columns are as below:  
#match(list, names(data))  
#[1] 2 11 19 20 29 30 48 51  
list <- c("member\_id", "emp\_title","url", "desc","mths\_since\_last\_delinq", "mths\_since\_last\_record", "next\_pymnt\_d", "mths\_since\_last\_major\_derog")  
match(list, names(data))

## [1] 2 11 19 20 29 30 48 51

#Q8. Delete the columns indexed: (2, 11, 19, 20, 29, 30, 48, 51) and assign the result to the dataframe "data"  
data <- data[, -match(list, names(data))]  
  
#Q9. If you look at your Global Environment, you will find several numeric variables are "char" (character variables). Switch these columns (loan\_amt, funded\_amnt, funded\_amnt\_inv, annual\_inc), which are character vectors, to numeric vectors. Then assign this to the dataframe "data"  
data$loan\_amnt <- as.numeric(data$loan\_amnt)  
data$funded\_amnt <- as.numeric(data$funded\_amnt)  
data$funded\_amnt\_inv <- as.numeric(data$funded\_amnt\_inv)  
data$annual\_inc <- as.numeric(data$annual\_inc)  
  
#Q10. Save the data frame as a csv file. Upload your csv file.  
write.csv(data, "LendingClubLab.JF.csv")