

# Final Project

*MuthukumarSrinivasan*

*Friday, May 27, 2016*

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
require(ggplot2)
```

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

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
library(ggplot2)
```

```
finalProject<-read.csv('loan.csv',header=TRUE)
```

```
finalProject1000Rows<-read.csv('loan-1000Rows.csv',header=TRUE)
```

```
summary(finalProject)
```

```
##           id           member_id      loan_amnt      funded_amnt
## Min.      : 54734   Min.      : 70473   Min.      : 500   Min.      : 500
## 1st Qu.: 9206643   1st Qu.:10877134   1st Qu.: 8000   1st Qu.: 8000
## Median :34433267   Median :37095283   Median :13000   Median :13000
## Mean    :32465133   Mean    :35001825   Mean    :14755   Mean    :14742
## 3rd Qu.:54908135   3rd Qu.:58471347   3rd Qu.:20000   3rd Qu.:20000
## Max.    :68617057   Max.    :73544841   Max.    :35000   Max.    :35000
##
## funded_amnt_inv      term      int_rate      installment
## Min.      : 0      36 months:621125   Min.      : 5.32   Min.      : 15.67
## 1st Qu.: 8000      60 months:266254   1st Qu.: 9.99     1st Qu.: 260.70
## Median :13000                                     Median :12.99     Median : 382.55
## Mean    :14702                                     Mean    :13.25     Mean    : 436.72
## 3rd Qu.:20000                                     3rd Qu.:16.20     3rd Qu.: 572.60
## Max.    :35000                                     Max.    :28.99     Max.    :1445.46
##
## grade      sub_grade      emp_title      emp_length
## A:148202   B3      : 56323      : 51457   10+ years:291569
## B:254535   B4      : 55626   Teacher   : 13469   2 years  : 78870
## C:245860   C1      : 53387   Manager   : 11240   < 1 year : 70605
## D:139542   C2      : 52236   Registered Nurse: 5525   3 years  : 70026
## E: 70705   C3      : 50161   Owner     : 5376    1 year   : 57095
## F: 23046   C4      : 48857   RN        : 5355    5 years  : 55704
## G: 5489    (Other):570789   (Other)    :794957   (Other)  :263510
```

```

##   home_ownership      annual_inc      verification_status
## ANY      :      3   Min.      :      0   Not Verified   :266750
## MORTGAGE:443557   1st Qu.:  45000   Source Verified:329558
## NONE     :      50   Median :  65000   Verified       :291071
## OTHER    :     182   Mean    :  75028
## OWN      :  87470   3rd Qu.:  90000
## RENT     :356117   Max.    :9500000
##          NA's      :4
##      issue_d      loan_status      pymnt_plan
## Oct-2015: 48631   Current      :601779   n:887369
## Jul-2015: 45962   Fully Paid   :207723   y:    10
## Dec-2015: 44342   Charged Off  : 45248
## Oct-2014: 38782   Late (31-120 days): 11591
## Nov-2015: 37530   Issued       :  8460
## Aug-2015: 35886   In Grace Period :  6253
## (Other) :636246   (Other)      :  6325
##
##                                     url
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000007:    1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000030:    1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000033:    1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000045:    1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000067:    1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1000095:    1
## (Other)                                     :887373
##
##                                     desc
##                                     :761350
##                                     : 246
## Debt Consolidation                      : 13
## Borrower added on 03/17/14 > Debt consolidation<br>: 11
## Borrower added on 03/10/14 > Debt consolidation<br>: 10
## Borrower added on 02/19/14 > Debt consolidation<br>: 9
## (Other)                                :125740
##
##      purpose      title
## debt_consolidation:524215   Debt consolidation      :414001
## credit_card      :206182   Credit card refinancing:164331
## home_improvement : 51829   Home improvement     : 40112
## other            : 42894   Other                 : 31892
## major_purchase   : 17277   Debt Consolidation    : 15760
## small_business    : 10377   Major purchase        : 12051
## (Other)          : 34605   (Other)               :209232
##
##      zip_code      addr_state      dti      delinq_2yrs
## 945xx : 9770   CA      :129517   Min.      : 0.00   Min.      : 0.0000
## 750xx : 9417   NY      : 74086   1st Qu.: 11.91   1st Qu.: 0.0000
## 112xx : 9272   TX      : 71138   Median : 17.65   Median : 0.0000
## 606xx : 8641   FL      : 60935   Mean    : 18.16   Mean    : 0.3144
## 300xx : 8126   IL      : 35476   3rd Qu.: 23.95   3rd Qu.: 0.0000
## 100xx : 7609   NJ      : 33256   Max.    :9999.00   Max.    :39.0000
## (Other):834544   (Other):482971   NA's      :29
##
## earliest_cr_line  inq_last_6mths  mths_since_last_delinq
## Aug-2001: 6659   Min.      : 0.0000   Min.      : 0.0
## Aug-2000: 6529   1st Qu.: 0.0000   1st Qu.: 15.0
## Oct-2000: 6322   Median : 0.0000   Median : 31.0
## Oct-2001: 6154   Mean    : 0.6946   Mean    : 34.1
## Aug-2002: 6086   3rd Qu.: 1.0000   3rd Qu.: 50.0

```

```

## Sep-2000: 5918 Max. :33.0000 Max. :188.0
## (Other) :849711 NA's :29 NA's :454312
## mths_since_last_record open_acc pub_rec
## Min. : 0.0 Min. : 0.00 Min. : 0.0000
## 1st Qu.: 51.0 1st Qu.: 8.00 1st Qu.: 0.0000
## Median : 70.0 Median :11.00 Median : 0.0000
## Mean : 70.1 Mean :11.55 Mean : 0.1953
## 3rd Qu.: 92.0 3rd Qu.:14.00 3rd Qu.: 0.0000
## Max. :129.0 Max. :90.00 Max. :86.0000
## NA's :750326 NA's :29 NA's :29
## revol_bal revol_util total_acc initial_list_status
## Min. : 0 Min. : 0.00 Min. : 1.00 f:456848
## 1st Qu.: 6443 1st Qu.: 37.70 1st Qu.: 17.00 w:430531
## Median : 11875 Median : 56.00 Median : 24.00
## Mean : 16921 Mean : 55.07 Mean : 25.27
## 3rd Qu.: 20829 3rd Qu.: 73.60 3rd Qu.: 32.00
## Max. :2904836 Max. :892.30 Max. :169.00
## NA's :502 NA's :29
## out_prncp out_prncp_inv total_pymnt total_pymnt_inv
## Min. : 0 Min. : 0 Min. : 0 Min. : 0
## 1st Qu.: 0 1st Qu.: 0 1st Qu.: 1915 1st Qu.: 1900
## Median : 6458 Median : 6456 Median : 4895 Median : 4862
## Mean : 8403 Mean : 8400 Mean : 7559 Mean : 7521
## 3rd Qu.:13659 3rd Qu.:13654 3rd Qu.:10617 3rd Qu.:10566
## Max. :49373 Max. :49373 Max. :57778 Max. :57778
##
## total_rec_prncp total_rec_int total_rec_late_fee recoveries
## Min. : 0 Min. : 0.0 Min. : 0.0000 Min. : 0.00
## 1st Qu.: 1201 1st Qu.: 441.5 1st Qu.: 0.0000 1st Qu.: 0.00
## Median : 3215 Median : 1073.3 Median : 0.0000 Median : 0.00
## Mean : 5758 Mean : 1754.8 Mean : 0.3967 Mean : 45.92
## 3rd Qu.: 8000 3rd Qu.: 2238.3 3rd Qu.: 0.0000 3rd Qu.: 0.00
## Max. :35000 Max. :24205.6 Max. :358.6800 Max. :33520.27
##
## collection_recovery_fee last_pymnt_d last_pymnt_amnt
## Min. : 0.000 Jan-2016:470150 Min. : 0.0
## 1st Qu.: 0.000 Dec-2015:150861 1st Qu.: 280.2
## Median : 0.000 : 17659 Median : 462.8
## Mean : 4.881 Oct-2015: 16000 Mean : 2164.2
## 3rd Qu.: 0.000 Jul-2015: 14483 3rd Qu.: 831.2
## Max. :7002.190 Nov-2015: 13981 Max. :36475.6
## (Other) :204245
## next_pymnt_d last_credit_pull_d collections_12_mths_ex_med
## Feb-2016:553406 Jan-2016:730574 Min. : 0.00000
## :252971 Dec-2015: 19308 1st Qu.: 0.00000
## Jan-2016: 78195 Nov-2015: 11490 Median : 0.00000
## Mar-2011: 107 Oct-2015: 10419 Mean : 0.01438
## Apr-2011: 101 Sep-2015: 10087 3rd Qu.: 0.00000
## Feb-2011: 91 Jul-2015: 8642 Max. :20.00000
## (Other) : 2508 (Other) : 96859 NA's :145
## mths_since_last_major_derog policy_code application_type
## Min. : 0.0 Min. :1 INDIVIDUAL:886868
## 1st Qu.: 27.0 1st Qu.:1 JOINT : 511
## Median : 44.0 Median :1

```

```

## Mean      : 44.1              Mean      :1
## 3rd Qu.: 61.0              3rd Qu.:1
## Max.      :188.0            Max.      :1
## NA's      :665676
## annual_inc_joint   dti_joint       verification_status_joint
## Min.      : 17950   Min.      : 3.0           :886868
## 1st Qu.: 76033   1st Qu.:13.2       Not Verified   : 283
## Median :101771   Median :17.6       Source Verified: 61
## Mean      :109981   Mean      :18.3       Verified      : 167
## 3rd Qu.:132800   3rd Qu.:22.6
## Max.      :500000   Max.      :43.9
## NA's      :886868   NA's      :886870
## acc_now_delinq     tot_coll_amt     tot_cur_bal     open_acc_6m
## Min.      : 0.000000   Min.      : 0   Min.      : 0   Min.      : 0.0
## 1st Qu.: 0.000000   1st Qu.: 0   1st Qu.: 29853   1st Qu.: 0.0
## Median : 0.000000   Median : 0   Median : 80559   Median : 1.0
## Mean      : 0.004991   Mean      : 226   Mean      : 139458   Mean      : 1.1
## 3rd Qu.: 0.000000   3rd Qu.: 0   3rd Qu.: 208205   3rd Qu.: 2.0
## Max.      :14.000000   Max.      :9152545   Max.      :8000078   Max.      :14.0
## NA's      :29         NA's      :70276   NA's      :70276   NA's      :866007
## open_il_6m         open_il_12m         open_il_24m         mths_since_rcnt_il
## Min.      : 0.0     Min.      : 0.0     Min.      : 0.0     Min.      : 0.0
## 1st Qu.: 1.0     1st Qu.: 0.0     1st Qu.: 0.0     1st Qu.: 6.0
## Median : 2.0     Median : 0.0     Median : 1.0     Median : 12.0
## Mean      : 2.9     Mean      : 0.8     Mean      : 1.7     Mean      : 20.9
## 3rd Qu.: 4.0     3rd Qu.: 1.0     3rd Qu.: 2.0     3rd Qu.: 23.0
## Max.      :40.0     Max.      :12.0     Max.      :19.0     Max.      :363.0
## NA's      :866007   NA's      :866007   NA's      :866007   NA's      :866569
## total_bal_il       il_util         open_rv_12m         open_rv_24m
## Min.      : 0     Min.      : 0.0     Min.      : 0.0     Min.      : 0
## 1st Qu.: 10252   1st Qu.: 58.6     1st Qu.: 0.0     1st Qu.: 1
## Median : 24685   Median : 74.9     Median : 1.0     Median : 2
## Mean      : 36553   Mean      : 71.5     Mean      : 1.4     Mean      : 3
## 3rd Qu.: 47858   3rd Qu.: 87.6     3rd Qu.: 2.0     3rd Qu.: 4
## Max.      :878459   Max.      :223.3     Max.      :22.0     Max.      :43
## NA's      :866007   NA's      :868762   NA's      :866007   NA's      :866007
## max_bal_bc         all_util         total_rev_hi_lim     inq_fi
## Min.      : 0     Min.      : 0.0     Min.      : 0     Min.      : 0.0
## 1st Qu.: 2411   1st Qu.: 47.7     1st Qu.: 13900   1st Qu.: 0.0
## Median : 4483   Median : 61.9     Median : 23700   Median : 0.0
## Mean      : 5888   Mean      : 60.8     Mean      : 32069   Mean      : 0.9
## 3rd Qu.: 7772   3rd Qu.: 75.2     3rd Qu.: 39800   3rd Qu.: 1.0
## Max.      :127305   Max.      :151.4     Max.      :9999999   Max.      :17.0
## NA's      :866007   NA's      :866007   NA's      :70276   NA's      :866007
## total_cu_tl       inq_last_12m
## Min.      : 0.0     Min.      : -4
## 1st Qu.: 0.0     1st Qu.: 0
## Median : 0.0     Median : 2
## Mean      : 1.5     Mean      : 2
## 3rd Qu.: 2.0     3rd Qu.: 3
## Max.      :35.0     Max.      :32
## NA's      :866007   NA's      :866007

```

```
str(finalProject)
```

```
## 'data.frame': 887379 obs. of 74 variables:
## $ id : int 1077501 1077430 1077175 1076863 1075358 1075269 1069639 1072053
## $ member_id : int 1296599 1314167 1313524 1277178 1311748 1311441 1304742 1288686
## $ loan_amnt : num 5000 2500 2400 10000 3000 ...
## $ funded_amnt : num 5000 2500 2400 10000 3000 ...
## $ funded_amnt_inv : num 4975 2500 2400 10000 3000 ...
## $ term : Factor w/ 2 levels " 36 months"," 60 months": 1 2 1 1 2 1 2 1 2 2 ...
## $ int_rate : num 10.6 15.3 16 13.5 12.7 ...
## $ installment : num 162.9 59.8 84.3 339.3 67.8 ...
## $ grade : Factor w/ 7 levels "A","B","C","D",...: 2 3 3 3 2 1 3 5 6 2 ...
## $ sub_grade : Factor w/ 35 levels "A1","A2","A3",...: 7 14 15 11 10 4 15 21 27 10 ...
## $ emp_title : Factor w/ 299273 levels "","Property Manager",...: 1 224800 1 9376 2 ...
## $ emp_length : Factor w/ 12 levels "< 1 year","1 year",...: 3 1 3 3 2 5 10 11 6 1 ...
## $ home_ownership : Factor w/ 6 levels "ANY","MORTGAGE",...: 6 6 6 6 6 6 6 6 5 6 ...
## $ annual_inc : num 24000 30000 12252 49200 80000 ...
## $ verification_status : Factor w/ 3 levels "Not Verified",...: 3 2 1 2 2 2 1 2 2 3 ...
## $ issue_d : Factor w/ 103 levels "Apr-2008","Apr-2009",...: 22 22 22 22 22 22 22 22 ...
## $ loan_status : Factor w/ 10 levels "Charged Off",...: 6 1 6 6 2 6 2 6 1 1 ...
## $ pymnt_plan : Factor w/ 2 levels "n","y": 1 1 1 1 1 1 1 1 1 1 ...
## $ url : Factor w/ 887379 levels "https://www.lendingclub.com/browse/loanDetail.action?loanid=1077501",...: 1 1 1 1 1 1 1 1 ...
## $ desc : Factor w/ 124471 levels "","- Pay off Dell Financial: $ 1300.00 - Pay off Dell Financial: $ 1300.00",...: 1 1 1 1 1 1 1 1 ...
## $ purpose : Factor w/ 14 levels "car","credit_card",...: 2 1 12 10 10 14 3 1 12 1 ...
## $ title : Factor w/ 63146 levels "","'08 & '09 Roth IRA Investments",...: 10497 10497 10497 10497 10497 10497 10497 10497 ...
## $ zip_code : Factor w/ 935 levels "007xx","008xx",...: 810 296 572 856 909 803 267 267 267 267 ...
## $ addr_state : Factor w/ 51 levels "AK","AL","AR",...: 4 11 15 5 38 4 28 5 5 44 ...
## $ dti : num 27.65 1 8.72 20 17.94 ...
## $ delinq_2yrs : num 0 0 0 0 0 0 0 0 0 0 ...
## $ earliest_cr_line : Factor w/ 698 levels "","Apr-1955",...: 265 43 572 210 276 575 342 287 287 287 ...
## $ inq_last_6mths : num 1 5 2 1 0 3 1 2 2 0 ...
## $ mths_since_last_delinq : num NA NA NA 35 38 NA NA NA NA NA ...
## $ mths_since_last_record : num NA NA NA NA NA NA NA NA NA NA ...
## $ open_acc : num 3 3 2 10 15 9 7 4 11 2 ...
## $ pub_rec : num 0 0 0 0 0 0 0 0 0 0 ...
## $ revol_bal : num 13648 1687 2956 5598 27783 ...
## $ revol_util : num 83.7 9.4 98.5 21 53.9 28.3 85.6 87.5 32.6 36.5 ...
## $ total_acc : num 9 4 10 37 38 12 11 4 13 3 ...
## $ initial_list_status : Factor w/ 2 levels "f","w": 1 1 1 1 1 1 1 1 1 1 ...
## $ out_prncp : num 0 0 0 0 767 ...
## $ out_prncp_inv : num 0 0 0 0 767 ...
## $ total_pymnt : num 5861 1009 3004 12226 3242 ...
## $ total_pymnt_inv : num 5832 1009 3004 12226 3242 ...
## $ total_rec_prncp : num 5000 456 2400 10000 2233 ...
## $ total_rec_int : num 861 435 604 2209 1009 ...
## $ total_rec_late_fee : num 0 0 0 17 0 ...
## $ recoveries : num 0 117 0 0 0 ...
## $ collection_recovery_fee : num 0 1.11 0 0 0 0 0 0 2.09 2.52 ...
## $ last_pymnt_d : Factor w/ 99 levels "","Apr-2008",...: 42 7 58 42 43 42 43 42 6 80 ...
## $ last_pymnt_amnt : num 171.6 119.7 649.9 357.5 67.8 ...
## $ next_pymnt_d : Factor w/ 101 levels "","Apr-2008",...: 1 1 1 1 35 1 35 1 1 1 ...
## $ last_credit_pull_d : Factor w/ 104 levels "","Apr-2009",...: 43 102 43 42 43 104 43 25 14 6 ...
## $ collections_12_mths_ex_med : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ mths_since_last_major_derog: num NA NA NA NA NA NA NA NA NA NA NA ...
## $ policy_code : num 1 1 1 1 1 1 1 1 1 1 1 ...
## $ application_type : Factor w/ 2 levels "INDIVIDUAL","JOINT": 1 1 1 1 1 1 1 1 1 1 1 ...
## $ annual_inc_joint : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ dti_joint : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ verification_status_joint : Factor w/ 4 levels "", "Not Verified", ...: 1 1 1 1 1 1 1 1 1 1 1 ...
## $ acc_now_delinq : num 0 0 0 0 0 0 0 0 0 0 0 ...
## $ tot_coll_amt : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ tot_cur_bal : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_acc_6m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_il_6m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_il_12m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_il_24m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ mths_since_rcnt_il : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_bal_il : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ il_util : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_rv_12m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_rv_24m : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ max_bal_bc : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ all_util : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_rev_hi_lim : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ inq_fi : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ total_cu_tl : num NA NA NA NA NA NA NA NA NA NA NA ...
## $ inq_last_12m : num NA NA NA NA NA NA NA NA NA NA NA ...
```

```
summary(finalProject1000Rows)
```

```
##          id          member_id      loan_amnt      funded_amnt
## Min.   : 765927   Min.   : 943135   Min.   : 1000   Min.   : 1000
## 1st Qu.:1058868   1st Qu.:1290204   1st Qu.: 7050   1st Qu.: 7000
## Median :1061891   Median :1293501   Median :11500   Median :11000
## Mean   :1060005   Mean   :1291575   Mean   :12704   Mean   :12198
## 3rd Qu.:1065345   3rd Qu.:1299062   3rd Qu.:16050   3rd Qu.:16000
## Max.   :1077501   Max.   :1314167   Max.   :35000   Max.   :35000
##
## funded_amnt_inv      term      int_rate      installment
## Min.   : 1000      36 months:725   Min.   : 6.03   Min.   : 33.94
## 1st Qu.: 7000      60 months:274   1st Qu.: 9.91   1st Qu.: 223.44
## Median :10800                                     Median :12.69   Median : 337.75
## Mean   :12137                                     Mean   :12.96   Mean   : 366.99
## 3rd Qu.:15947                                     3rd Qu.:15.96   3rd Qu.: 469.24
## Max.   :35000                                     Max.   :23.91   Max.   :1243.85
##
## grade      sub_grade      emp_title      emp_length
## A:202      B3          : 84          : 51      10+ years:234
## B:354      B4          : 74      Bank of America      : 3      2 years :110
## C:207      B1          : 71      Department of Defense : 3      5 years : 96
## D:138      B5          : 67      IBM                  : 3      3 years : 91
## E: 72      A4          : 60      Kaiser Permanente    : 3      4 years : 90
## F: 21      C1          : 59      United States Air Force: 3      1 year  : 82
## G: 5       (Other):584      (Other)              :933      (Other) :296
## home_ownership      annual_inc      verification_status      issue_d
## MORTGAGE:328      Min.   : 12000      Not Verified :360      Dec-11:999
## OWN          : 75      1st Qu.: 40000      Source Verified:273
```

```

## RENT      :596      Median : 55000      Verified      :366
##           Mean      : 62119
##           3rd Qu.: 75500
##           Max.     :276000
##
##           loan_status pymnt_plan
## Charged Off      :176      n:999
## Current          : 91
## Default          : 1
## Fully Paid       :729
## Late (31-120 days): 2
##
##
##
##                                     url
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1010310: 1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1018385: 1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1020855: 1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1023158: 1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1023566: 1
## https://www.lendingclub.com/browse/loanDetail.action?loan_id=1024951: 1
## (Other)                                     :993
##
##
## Borrower added on 12/06/11 > Debt consolidation<br>
##
## Borrower added on 11/28/11 > Thanks.<br> Borrower added on 12/05/11 > Hello,<br>I am in need of t
## Borrower added on 11/30/11 > Payoff credit cards and close<br>Borrower added on 11/30/11 > Payoff
## Borrower added on 11/30/11 > this funding will be used to purchase tools and a van for a hardwood
## (Other)
##           purpose                                title                zip_code
## debt_consolidation:527      Debt Consolidation Loan: 91      606xx : 17
## credit_card      :252      Debt Consolidation      : 84      300xx : 16
## other            : 63      Credit Card Loan        : 35      900xx : 16
## home_improvement : 37      Consolidation          : 20      945xx : 16
## small_business   : 35      Credit Card          : 18      113xx : 15
## car              : 22      debt consolidation    : 18      112xx : 13
## (Other)          : 63      (Other)                :733      (Other):906
## addr_state      dti          delinq_2yrs          earliest_cr_line
## CA              :203      Min.      : 0.72      Min.      :0.00000      May-00 : 16
## NY              : 93      1st Qu.:10.02      1st Qu.:0.00000      Oct-00 : 15
## FL              : 82      Median :14.80      Median :0.00000      Apr-00 : 14
## TX              : 68      Mean    :14.64      Mean    :0.07908      Nov-98 : 13
## NJ              : 49      3rd Qu.:19.63      3rd Qu.:0.00000      Jan-00 : 12
## IL              : 40      Max.    :29.85      Max.    :3.00000      Oct-02 : 12
## (Other):464                                     (Other):917
## inq_last_6mths      mths_since_last_delinq mths_since_last_record
## Min.      :0.0000      Min.      : 2.00      Min.      : 33
## 1st Qu.:0.0000      1st Qu.:24.00      1st Qu.: 83
## Median :1.0000      Median :40.00      Median : 91
## Mean    :0.8579      Mean    :41.12      Mean    : 87
## 3rd Qu.:1.0000      3rd Qu.:58.00      3rd Qu.:107
## Max.    :5.0000      Max.    :82.00      Max.    :118
##           NA's      :774      NA's      :979
## open_acc      pub_rec      revol_bal      revol_util

```

```

## Min. : 2.000 Min. :0.00000 Min. : 0 Min. : 0.00
## 1st Qu.: 7.000 1st Qu.:0.00000 1st Qu.: 7636 1st Qu.:51.75
## Median : 9.000 Median :0.00000 Median :12183 Median :68.70
## Mean : 9.386 Mean :0.02002 Mean :14106 Mean :64.80
## 3rd Qu.:11.000 3rd Qu.:0.00000 3rd Qu.:18358 3rd Qu.:82.65
## Max. :34.000 Max. :1.00000 Max. :93718 Max. :99.80
##
## total_acc initial_list_status out_prncp out_prncp_inv
## Min. : 3.00 f:999 Min. : 0.0 Min. : 0
## 1st Qu.:13.00 1st Qu.: 0.0 1st Qu.: 0
## Median :19.00 Median : 0.0 Median : 0
## Mean :20.49 Mean : 393.5 Mean : 393
## 3rd Qu.:26.00 3rd Qu.: 0.0 3rd Qu.: 0
## Max. :79.00 Max. :10069.4 Max. :10062
##
## total_pymnt total_pymnt_inv total_rec_prncp total_rec_int
## Min. : 0 Min. : 0 Min. : 0 Min. : 0.0
## 1st Qu.: 6904 1st Qu.: 6894 1st Qu.: 5110 1st Qu.: 960.5
## Median :11599 Median :11593 Median : 9600 Median : 1807.9
## Mean :13219 Mean :13124 Mean :10396 Mean : 2703.6
## 3rd Qu.:17560 3rd Qu.:17405 3rd Qu.:14000 3rd Qu.: 3279.7
## Max. :45755 Max. :44520 Max. :35000 Max. :19834.1
##
## total_rec_late_fee recoveries collection_recovery_fee
## Min. : 0.000 Min. : 0.0 Min. : 0.000
## 1st Qu.: 0.000 1st Qu.: 0.0 1st Qu.: 0.000
## Median : 0.000 Median : 0.0 Median : 0.000
## Mean : 1.081 Mean : 118.7 Mean : 9.818
## 3rd Qu.: 0.000 3rd Qu.: 0.0 3rd Qu.: 0.000
## Max. :94.480 Max. :11349.5 Max. :2043.180
##
## last_pymnt_d last_pymnt_amnt next_pymnt_d last_credit_pull_d
## Dec-14 :163 Min. : 0.0 :905 Jan-16 :420
## Jan-15 :146 1st Qu.: 281.2 Feb-16: 76 Dec-14 :130
## Dec-15 : 76 Median : 530.8 Jan-16: 18 Jan-15 : 33
## Dec-13 : 31 Mean : 2762.7 Dec-15 : 25
## Oct-13 : 31 3rd Qu.: 3516.3 Sep-15 : 17
## Feb-14 : 25 Max. :28412.4 Feb-14 : 16
## (Other):527 (Other):358
## collections_12_mths_ex_med mths_since_last_major_derog policy_code
## Min. :0 Mode:logical Min. :1
## 1st Qu.:0 NA's:999 1st Qu.:1
## Median :0 Median :1
## Mean :0 Mean :1
## 3rd Qu.:0 3rd Qu.:1
## Max. :0 Max. :1
##
## application_type annual_inc_joint dti_joint
## INDIVIDUAL:999 Mode:logical Mode:logical
## NA's:999 NA's:999
##
##
##
##

```



```
##
## verification_status_joint acc_now_delinq tot_coll_amt tot_cur_bal
## Mode:logical Min. :0 Mode:logical Mode:logical
## NA's:999 1st Qu.:0 NA's:999 NA's:999
## Median :0
## Mean :0
## 3rd Qu.:0
## Max. :0
##
## open_acc_6m open_il_6m open_il_12m open_il_24m
## Mode:logical Mode:logical Mode:logical Mode:logical
## NA's:999 NA's:999 NA's:999 NA's:999
##
##
##
##
## mths_since_rcnt_il total_bal_il il_util open_rv_12m
## Mode:logical Mode:logical Mode:logical Mode:logical
## NA's:999 NA's:999 NA's:999 NA's:999
##
##
##
##
## open_rv_24m max_bal_bc all_util total_rev_hi_lim
## Mode:logical Mode:logical Mode:logical Mode:logical
## NA's:999 NA's:999 NA's:999 NA's:999
##
##
##
##
## inq_fi total_cu_tl inq_last_12m
## Mode:logical Mode:logical Mode:logical
## NA's:999 NA's:999 NA's:999
##
##
##
##
##
```

```
str(finalProject1000Rows)
```

```
## 'data.frame': 999 obs. of 74 variables:
## $ id : int 1077501 1077430 1077175 1076863 1075358 1075269 1069639 1072053
## $ member_id : int 1296599 1314167 1313524 1277178 1311748 1311441 1304742 1288686
## $ loan_amnt : int 5000 2500 2400 10000 3000 5000 7000 3000 5600 5375 ...
## $ funded_amnt : int 5000 2500 2400 10000 3000 5000 7000 3000 5600 5375 ...
## $ funded_amnt_inv : num 4975 2500 2400 10000 3000 ...
## $ term : Factor w/ 2 levels " 36 months"," 60 months": 1 2 1 1 2 1 2 1 2 2 ..
## $ int_rate : num 10.6 15.3 16 13.5 12.7 ...
## $ installment : num 162.9 59.8 84.3 339.3 67.8 ...
## $ grade : Factor w/ 7 levels "A","B","C","D",...: 2 3 3 3 2 1 3 5 6 2 ...
```

```

## $ sub_grade : Factor w/ 34 levels "A1","A2","A3",...: 7 14 15 11 10 4 15 21 27 10 .
## $ emp_title : Factor w/ 919 levels "","16th MP BDE, U.S. Army",...: 1 673 1 27 845 8
## $ emp_length : Factor w/ 12 levels "< 1 year","1 year",...: 3 1 3 3 2 5 10 11 6 1 ..
## $ home_ownership : Factor w/ 3 levels "MORTGAGE","OWN",...: 3 3 3 3 3 3 3 3 2 3 ...
## $ annual_inc : num 24000 30000 12252 49200 80000 ...
## $ verification_status : Factor w/ 3 levels "Not Verified",...: 3 2 1 2 2 2 1 2 2 3 ...
## $ issue_d : Factor w/ 1 level "Dec-11": 1 1 1 1 1 1 1 1 1 1 ...
## $ loan_status : Factor w/ 5 levels "Charged Off",...: 4 1 4 4 2 4 2 4 1 1 ...
## $ pymnt_plan : Factor w/ 1 level "n": 1 1 1 1 1 1 1 1 1 1 ...
## $ url : Factor w/ 999 levels "https://www.lendingclub.com/browse/loanDetail.",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ desc : Factor w/ 567 levels ""," ", " Borrower added on 11/28/11 > Thanks.<...",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ purpose : Factor w/ 13 levels "car","credit_card",...: 2 1 11 9 9 13 3 1 11 9 .
## $ title : Factor w/ 543 levels " Bill pay of",...: 91 32 467 458 448 380 332 56
## $ zip_code : Factor w/ 378 levels "010xx","011xx",...: 303 150 227 321 362 299 129
## $ addr_state : Factor w/ 42 levels "AK","AL","AR",...: 4 10 12 5 30 4 22 5 5 35 ...
## $ dti : num 27.65 1 8.72 20 17.94 ...
## $ delinq_2yrs : int 0 0 0 0 0 0 0 0 0 0 ...
## $ earliest_cr_line : Factor w/ 268 levels "Apr-00","Apr-01",...: 100 21 209 87 110 212 119
## $ inq_last_6mths : int 1 5 2 1 0 3 1 2 2 0 ...
## $ mths_since_last_delinq : int NA NA NA 35 38 NA NA NA NA NA ...
## $ mths_since_last_record : int NA NA NA NA NA NA NA NA NA NA NA ...
## $ open_acc : int 3 3 2 10 15 9 7 4 11 2 ...
## $ pub_rec : int 0 0 0 0 0 0 0 0 0 0 ...
## $ revol_bal : int 13648 1687 2956 5598 27783 7963 17726 8221 5210 9279 ...
## $ revol_util : num 83.7 9.4 98.5 21 53.9 28.3 85.6 87.5 32.6 36.5 ...
## $ total_acc : int 9 4 10 37 38 12 11 4 13 3 ...
## $ initial_list_status : Factor w/ 1 level "f": 1 1 1 1 1 1 1 1 1 1 ...
## $ out_prncp : num 0 0 0 0 767 ...
## $ out_prncp_inv : num 0 0 0 0 767 ...
## $ total_pymnt : num 5861 1009 3004 12226 3242 ...
## $ total_pymnt_inv : num 5832 1009 3004 12226 3242 ...
## $ total_rec_prncp : num 5000 456 2400 10000 2233 ...
## $ total_rec_int : num 861 435 604 2209 1009 ...
## $ total_rec_late_fee : num 0 0 0 17 0 ...
## $ recoveries : num 0 117 0 0 0 ...
## $ collection_recovery_fee : num 0 1.11 0 0 0 0 0 0 2.09 2.52 ...
## $ last_pymnt_d : Factor w/ 50 levels "","Apr-12","Apr-13",...: 21 3 29 21 22 21 22 21 1
## $ last_pymnt_amnt : num 171.6 119.7 649.9 357.5 67.8 ...
## $ next_pymnt_d : Factor w/ 3 levels "","Feb-16","Jan-16": 1 1 1 1 2 1 2 1 1 1 ...
## $ last_credit_pull_d : Factor w/ 47 levels "Apr-12","Apr-13",...: 20 45 20 19 20 47 20 11 5 1
## $ collections_12_mths_ex_med : int 0 0 0 0 0 0 0 0 0 0 ...
## $ mths_since_last_major_derog : logi NA NA NA NA NA NA NA ...
## $ policy_code : int 1 1 1 1 1 1 1 1 1 1 ...
## $ application_type : Factor w/ 1 level "INDIVIDUAL": 1 1 1 1 1 1 1 1 1 1 ...
## $ annual_inc_joint : logi NA NA NA NA NA NA NA ...
## $ dti_joint : logi NA NA NA NA NA NA NA ...
## $ verification_status_joint : logi NA NA NA NA NA NA NA ...
## $ acc_now_delinq : int 0 0 0 0 0 0 0 0 0 0 ...
## $ tot_coll_amt : logi NA NA NA NA NA NA NA ...
## $ tot_cur_bal : logi NA NA NA NA NA NA NA ...
## $ open_acc_6m : logi NA NA NA NA NA NA NA ...
## $ open_il_6m : logi NA NA NA NA NA NA NA ...
## $ open_il_12m : logi NA NA NA NA NA NA NA ...
## $ open_il_24m : logi NA NA NA NA NA NA NA ...

```

```
## $ mths_since_rcnt_il      : logi NA NA NA NA NA NA ...
## $ total_bal_il           : logi NA NA NA NA NA NA ...
## $ il_util                : logi NA NA NA NA NA NA ...
## $ open_rv_12m           : logi NA NA NA NA NA NA ...
## $ open_rv_24m           : logi NA NA NA NA NA NA ...
## $ max_bal_bc             : logi NA NA NA NA NA NA ...
## $ all_util               : logi NA NA NA NA NA NA ...
## $ total_rev_hi_lim       : logi NA NA NA NA NA NA ...
## $ inq_fi                 : logi NA NA NA NA NA NA ...
## $ total_cu_tl            : logi NA NA NA NA NA NA ...
## $ inq_last_12m           : logi NA NA NA NA NA NA ...
```

```
columnNames<-colnames(finalProject)
columnNames
```

```
## [1] "id" "member_id"
## [3] "loan_amnt" "funded_amnt"
## [5] "funded_amnt_inv" "term"
## [7] "int_rate" "installment"
## [9] "grade" "sub_grade"
## [11] "emp_title" "emp_length"
## [13] "home_ownership" "annual_inc"
## [15] "verification_status" "issue_d"
## [17] "loan_status" "pymnt_plan"
## [19] "url" "desc"
## [21] "purpose" "title"
## [23] "zip_code" "addr_state"
## [25] "dti" "delinq_2yrs"
## [27] "earliest_cr_line" "inq_last_6mths"
## [29] "mths_since_last_delinq" "mths_since_last_record"
## [31] "open_acc" "pub_rec"
## [33] "revol_bal" "revol_util"
## [35] "total_acc" "initial_list_status"
## [37] "out_prncp" "out_prncp_inv"
## [39] "total_pymnt" "total_pymnt_inv"
## [41] "total_rec_prncp" "total_rec_int"
## [43] "total_rec_late_fee" "recoveries"
## [45] "collection_recovery_fee" "last_pymnt_d"
## [47] "last_pymnt_amnt" "next_pymnt_d"
## [49] "last_credit_pull_d" "collections_12_mths_ex_med"
## [51] "mths_since_last_major_derog" "policy_code"
## [53] "application_type" "annual_inc_joint"
## [55] "dti_joint" "verification_status_joint"
## [57] "acc_now_delinq" "tot_coll_amt"
## [59] "tot_cur_bal" "open_acc_6m"
## [61] "open_il_6m" "open_il_12m"
## [63] "open_il_24m" "mths_since_rcnt_il"
## [65] "total_bal_il" "il_util"
## [67] "open_rv_12m" "open_rv_24m"
## [69] "max_bal_bc" "all_util"
## [71] "total_rev_hi_lim" "inq_fi"
## [73] "total_cu_tl" "inq_last_12m"
```