Lending Club Data

Siran Zhao

December 3, 2017

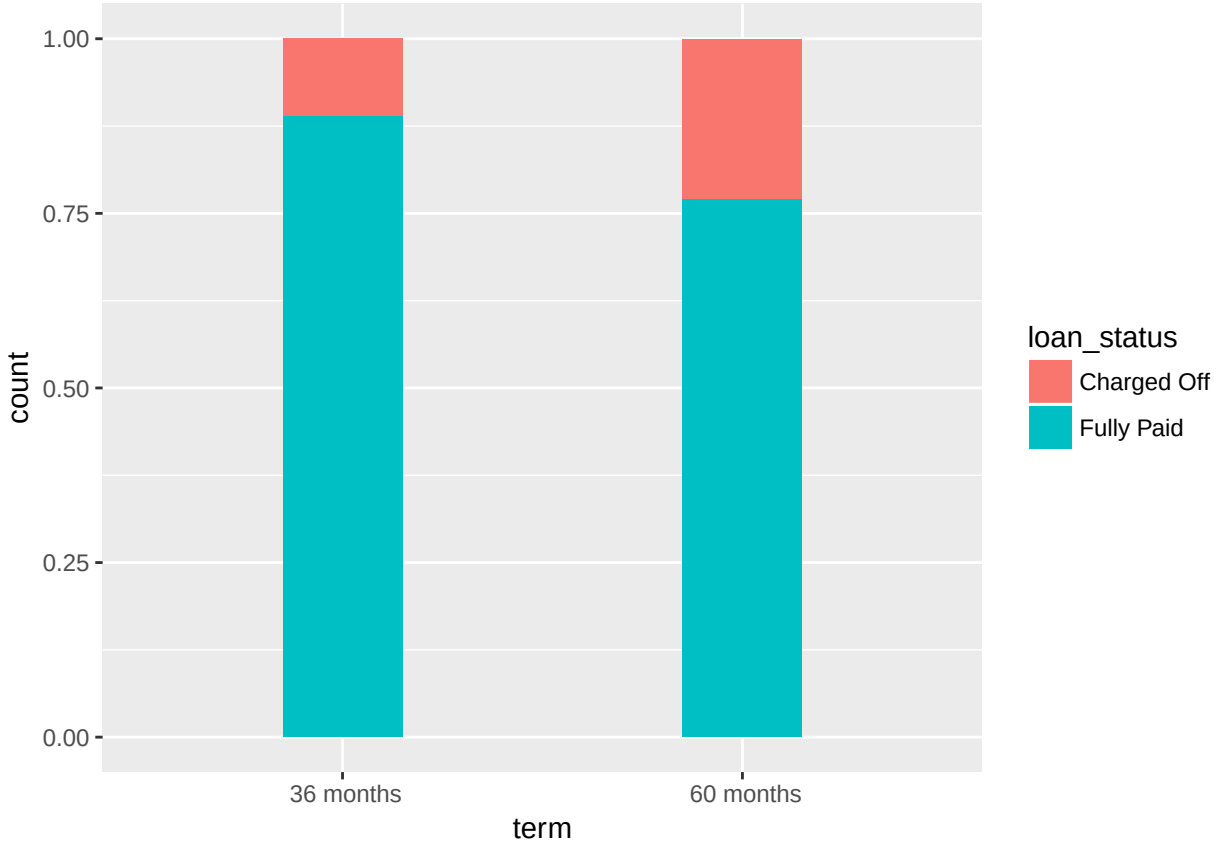
OVERVIEW

As more and more peoples like to lend money from lending club. Under this circumstance, I think if we can predict the probability that whether a person are going to charge off his(her) loans, then we can control the risk we are going to face.

Data Cleaning

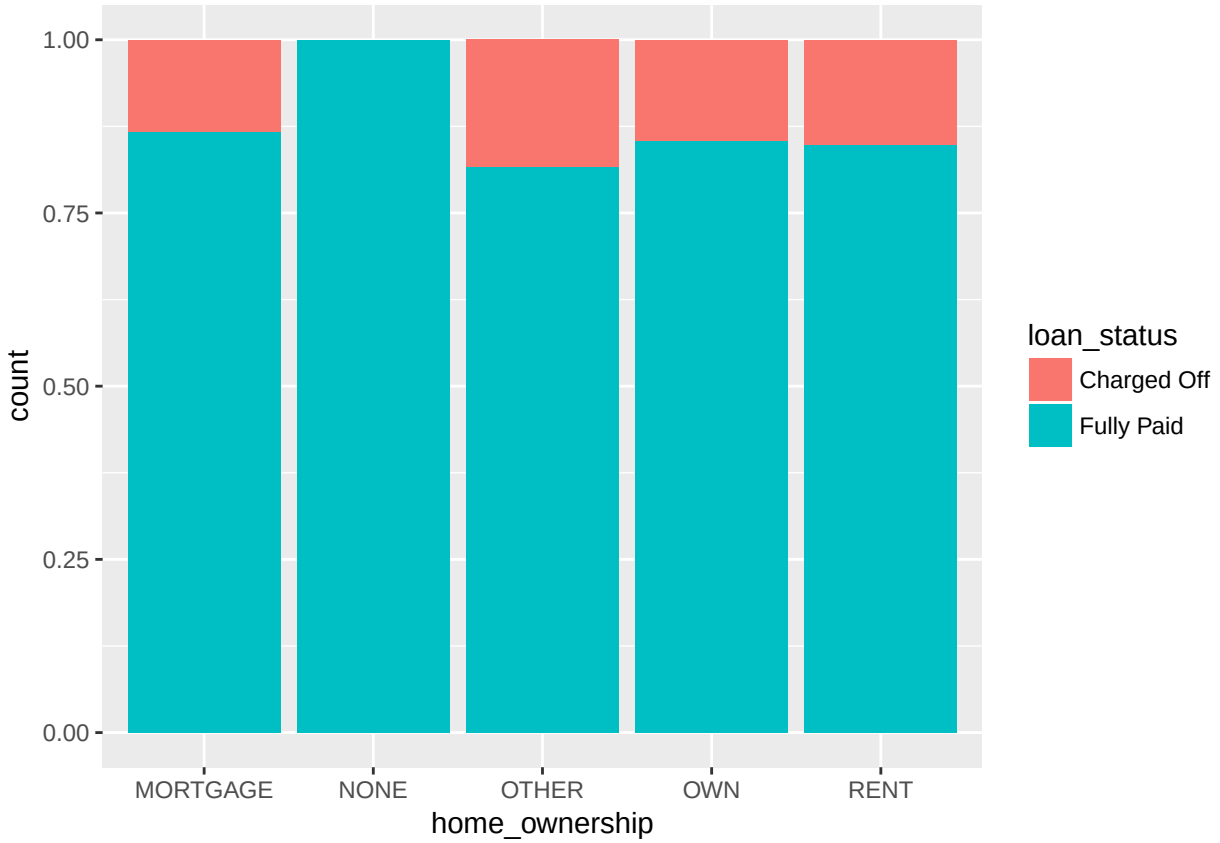
I select 10 columns from the original data set, then assign their format in order to fit the models. Because the scale of intterest rate is not appropriate, so I centralized them.

potion Graph between Loan_term and loan_status



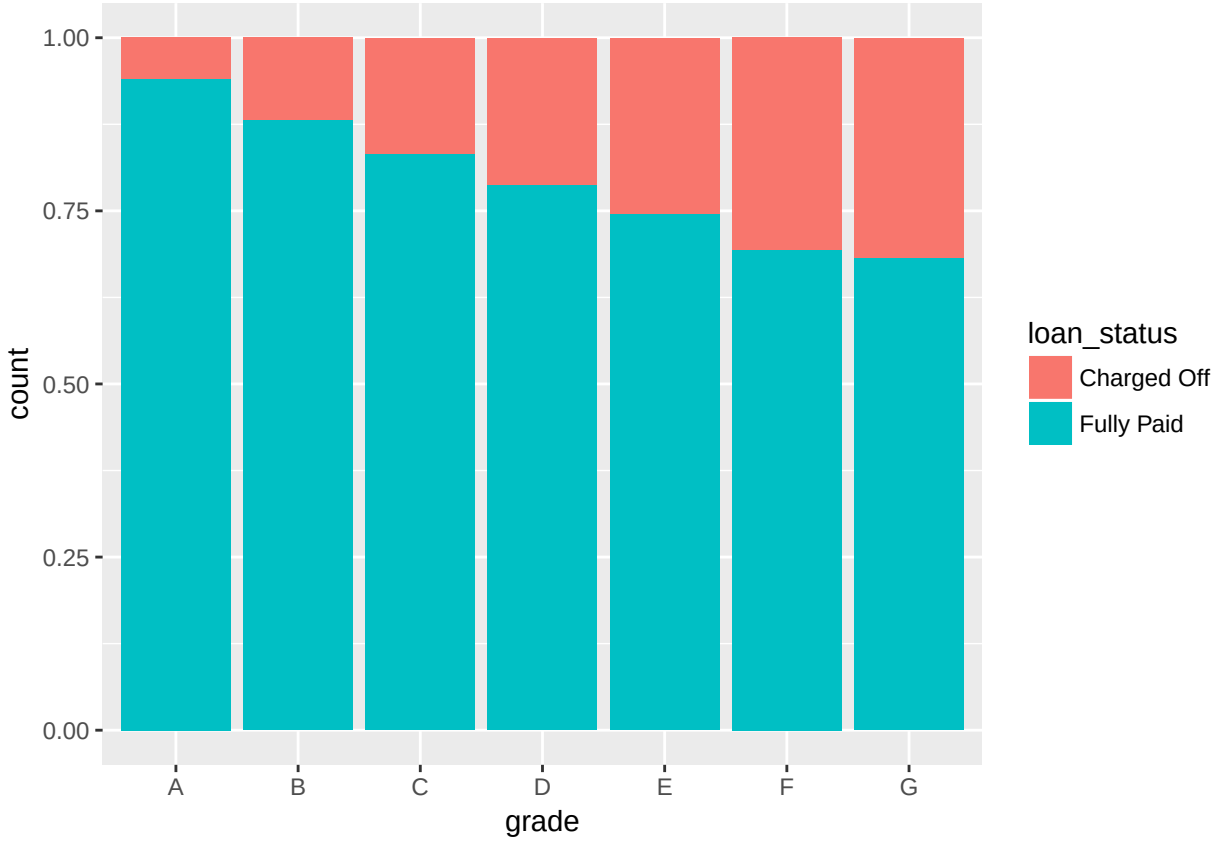
36 months ratio of charged off is lower than 72 months. So there must be some connection between terms and results.

potion graph between home_ownership and loan_status



From the graph above, we can find slightly difference in charge off rate among different kinds of ownership.

Ratio of Charged off and fully paid loan among Loan Grade.



From this graph, we can find that as the Grade of the loans went down, the ratio of charged off became higher. So I think we can fit the models with the elements above.

Fitting Models

At first we tried to fit a model only with term of loans in different grade levels. Then we added two more into the model to see whether it can make my models more accurate.

AIC Compairision

```
##       AIC       BIC    logLik  deviance  df.resid
##  7881.668  7903.299 -3937.834  7875.668  9997.000
```

```
##       AIC       BIC    logLik  deviance  df.resid
##  7873.619  7916.881 -3930.810  7861.619  9994.000
```

```
##       AIC       BIC    logLik  deviance  df.resid
##  7853.875  7904.347 -3919.937  7839.875  9993.000
```

From the AIC Score above, we can find that as we added home_ownership and annual interest of the loan into the fitting models, the AIC Score went down. This meant the model fit better with this two elements.

Summaries for 3 models.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pay.or.not ~ term + (1 | grade)
##    Data: sample.loan.data
##
##      AIC      BIC   logLik deviance df.resid
##   7881.7   7903.3  -3937.8   7875.7     9997
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8371  0.2606  0.3483  0.4506  0.7150
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  grade  (Intercept) 0.2567   0.5067
## Number of obs: 10000, groups:  grade, 7
##
## Fixed effects:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     1.72487    0.19936   8.652  < 2e-16 ***
## term 60 months -0.51506    0.06555  -7.858 3.91e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## term60mnths -0.169
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pay.or.not ~ term + home_ownership + (1 | grade)
##    Data: sample.loan.data
##
##      AIC      BIC   logLik deviance df.resid
##   7873.6   7916.9  -3930.8   7861.6     9994
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0051  0.2497  0.3600  0.4359  0.8734
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  grade  (Intercept) 0.2467   0.4967
## Number of obs: 10000, groups:  grade, 7
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.83570    0.19972   9.192  < 2e-16 ***
## term 60 months      -0.55367    0.06679  -8.290  < 2e-16 ***
## home_ownershipOTHER -1.32089    0.44586  -2.963  0.00305 **
```

```
## home_ownershipOWN   -0.06899    0.11787  -0.585  0.55831
## home_ownershipRENT  -0.17098    0.06205  -2.755  0.00586 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) trm60m h_OTHE hm_OWN
## term60mnths -0.202
## hm_wnrOTHER -0.031  0.057
## hm_wnrshOWN -0.095  0.058  0.043
## hm_wnrsRENT -0.198  0.180  0.085  0.292
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: pay.or.not ~ int_rate + term + home_ownership + (1 | grade) -
##     1
##    Data: sample.loan.data
##
##      AIC      BIC   logLik deviance df.resid
##   7853.9   7904.3  -3919.9   7839.9     9993
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3176  0.2650  0.3520  0.4353  0.9186
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  grade  (Intercept) 0.005795 0.07612
## Number of obs: 10000, groups:  grade, 7
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## int_rate            -0.40591    0.04518   -8.98  < 2e-16 ***
## term 36 months       2.10764    0.06449   32.68  < 2e-16 ***
## term 60 months       1.61755    0.07090   22.81  < 2e-16 ***
## home_ownershipOTHER -1.32691    0.44394   -2.99  0.00280 **
## home_ownershipOWN   -0.06031    0.11800   -0.51  0.60925
## home_ownershipRENT  -0.16440    0.06213   -2.65  0.00815 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             int_rt trm36m trm60m h_OTHE hm_OWN
## term36mnths -0.071
## term60mnths -0.442  0.489
## hm_wnrOTHER  0.005 -0.099 -0.036
## hm_wnrshOWN -0.009 -0.297 -0.212  0.043
## hm_wnrsRENT -0.056 -0.588 -0.361  0.087  0.293
```

From above summaries, 1,Interest rate have negative influence on Fully paid probability. 2,As last-time of loans grew longer, it have weaker positive influence on Fully paid probability. 3.Basis of home_ownership is MORTGAGE, and we can find who have a Other or Rent contracts of the houses are more likely to Charge off.

Predict the probability of Fully Paid and Charged off among 9227 customers' loan whose status under Current.

After finishing the models, I tried to predict loan status within 10000 data from 2017 Quarter 2 in which customers' status are still pending.

```
##    loan_amnt      funded_amnt            term        int_rate
##  Min.   : 1200   Min.   : 1200            :  0   Min.   :3.419
##  1st Qu.:12000   1st Qu.:12000    36 months:124   1st Qu.:3.807
##  Median :17913   Median :17913    60 months:488   Median :4.971
##  Mean   :19048   Mean   :19048                    Mean   :4.789
##  3rd Qu.:25000   3rd Qu.:25000                    3rd Qu.:5.261
##  Max.   :40000   Max.   :40000                    Max.   :6.200
##
##      grade       sub_grade    home_ownership   annual_inc
##  D      :280   D3     :115   RENT    :281    Min.   :   2
##  E      :195   E5     :100   MORTGAGE:265    1st Qu.:2300
##  F      : 94   D5     : 60   OWN     : 66    Median :4832
##  G      : 43   D4     : 56           :  0    Mean   :4552
##         :  0   E4     : 51   35000   :  0    3rd Qu.:6951
##  A      :  0   E3     : 35   40000   :  0    Max.   :9066
##  (Other):  0   (Other):195   (Other) :  0
##           loan_status    addr_state       prob         pre.status
##  Current        :612   CA     : 76   Min.   :0.2571   Length:612
##                 :  0   NY     : 63   1st Qu.:0.3849   Class :character
##  Charged Off    :  0   FL     : 51   Median :0.4350   Mode  :character
##  Default        :  0   TX     : 50   Mean   :0.4190
##  Fully Paid     :  0   IL     : 28   3rd Qu.:0.4758
##  In Grace Period:  0   NJ     : 28   Max.   :0.5000
##  (Other)        :  0   (Other):316
```

I filtered the people who are likely to charge off in the future with Probability larger than 50%. And count the number are 612. We should be aware of these peoples.