

STAT3622 Homework 1: Lending Club

Date: October 4, 2018

Submit in HTML format through Moodle on or before October 21, 2018. You also need to supplement the codes for reproducing your results for grading purpose.

The peer-to-peer (P2P) lending is a method of lending money through online services by matching individual lenders and borrowers. It has been one of the hottest FinTech applications, including Lending Club in US, WeLend in Hong Kong, and Yirendai in mainland China. In this homework, we use the publicly available real data of Lending Club that can be downloaded from [www.lendingclub.com]. The original data is pretty big. We have manipulated the data by aggregating, subsetting and subsampling, then ended up a relatively small data with the following summary statistics:

```
## Loan applications dating from 201501 to 201806
```

```
## Loan purposes:
```

```
##
```

```
##      Credit_Card Debt_Consolidation      Housing
##      54139          177386          31822
##      Other          Purchase
##      65926          29591
```

```
## Loan application attributes:
```

```
## Amount_Requested Risk_Score Debt_Income_Ratio Employment_Length
## Min. : 1000      Min. :300.0      Min. : -0.0100      Min. : 0.000
## 1st Qu.: 5000    1st Qu.:605.0    1st Qu.: 0.1083     1st Qu.: 0.000
## Median :10000    Median :652.0    Median : 0.2064     Median : 0.000
## Mean :13279     Mean :646.1     Mean : 0.9073      Mean : 1.564
## 3rd Qu.:20000    3rd Qu.:685.0    3rd Qu.: 0.3413     3rd Qu.: 0.000
## Max. :82700     Max. :990.0     Max. :3082.0000     Max. :11.000
```

```
## Sample size of accepted applications: 71695
```

```
## Sample size of declined applications: 287169
```

Such data can be downloaded from the [STAT3622 course website].

1. (20%) Plot the acceptance rates of loan applications over the months. (Hint: $\text{rate} = \text{accept}/(\text{accept} + \text{decline})$.) Interpret the resulting chart.
2. (20%) Draw a side-by-side barchart of loan purposes grouped by the status of accept and decline. Interpret the resulting chart.
3. (60%) Use R:ggplot2 package to visualize the distributional difference for each continuous attribute between accept and reject. Perform suitable data processing whenever necessary.
 - a) (15%) Attribute **Amount_Requested**. Interpret the chart.
 - b) (15%) Attribute **Risk_Score**. Interpret the chart.
 - c) (15%) Attribute **Debt_Income_Ratio**. Interpret the chart.
 - d) (15%) Attribute **Employment_Length**. Interpret the chart.