Math 4330-Project

Cao The Cong Toan #215230170

December 2020

Contents

Loan Status Prediction

1) Overview of the problem

This project is about Loan Status Prediction.Data is accessed from https://www.kaggle.com/ . There are 12 variable in the data. Categorical regression analysis is used to build a fit mode, and determinant significant variables.The studying responce in this project is loan status (Yes or No).The purpose of this study is finding the considered variable to make prediction on the approval loan application.

2)Data and modeling approach,

2.1)Exporing and Dealing with missing data

```
data<-read.csv("D:\\math\\math 4330\\project\\loan_data_set13.csv")
str(data)
```

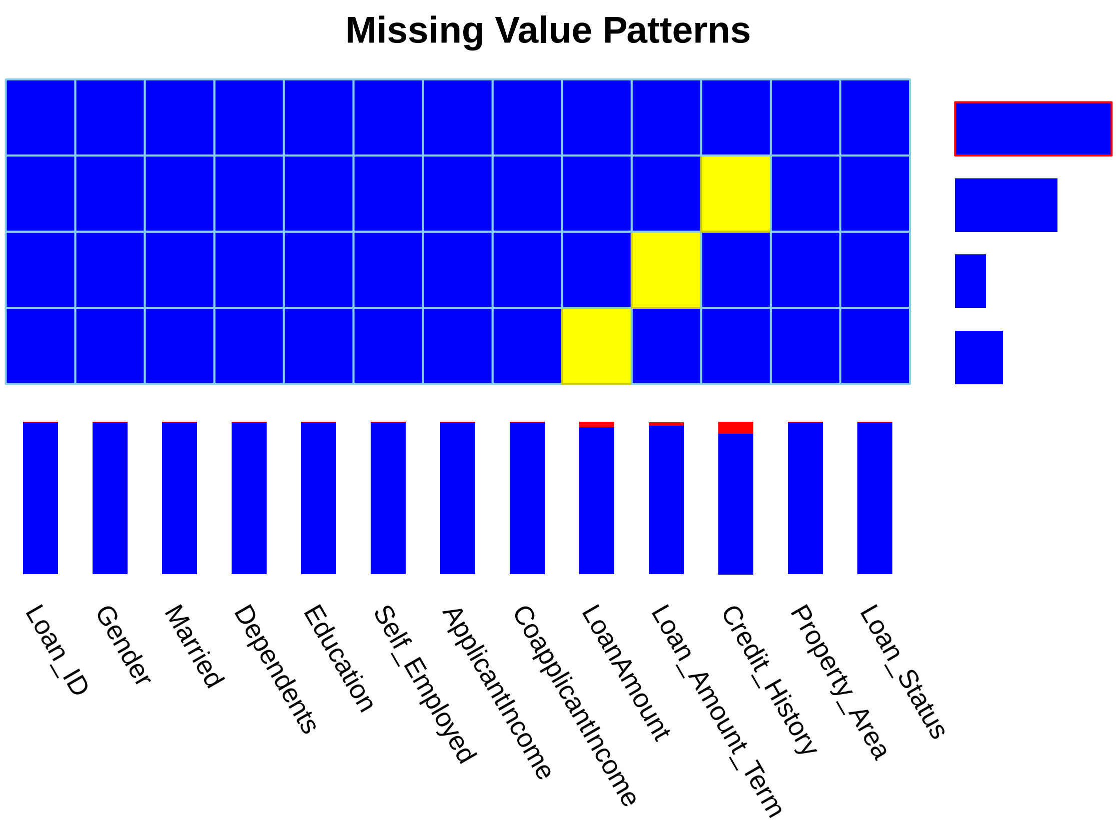
```
## 'data.frame':    566 obs. of  13 variables:
##  $ Loan_ID          : chr  "LP001002" "LP001003" "LP001005" "LP001006" ...
##  $ Gender           : chr  "Male" "Male" "Male" "Male" ...
##  $ Married          : chr  "No" "Yes" "Yes" "Yes" ...
##  $ Dependents       : chr  "0" "1" "0" "0" ...
##  $ Education        : chr  "Graduate" "Graduate" "Graduate" "Not Graduate" ...
##  $ Self_Employed    : chr  "No" "No" "Yes" "No" ...
##  $ ApplicantIncome  : int  5849 4583 3000 2583 6000 5417 2333 3036 4006 12841 ...
##  $ CoapplicantIncome: num  0 1508 0 2358 0 ...
##  $ LoanAmount       : int  NA 128 66 120 141 267 95 158 168 349 ...
##  $ Loan_Amount_Term : int  360 360 360 360 360 360 360 360 360 360 ...
##  $ Credit_History   : int  1 1 1 1 1 1 1 0 1 1 ...
##  $ Property_Area    : chr  "Urban" "Rural" "Urban" "Urban" ...
##  $ Loan_Status      : chr  "Y" "N" "Y" "Y" ...
```

```
devtools::install_github('gmonette/spida2')
```

```
## WARNING: Rtools is required to build R packages, but is not currently installed.
##
## Please download and install Rtools 4.0 from https://cran.r-project.org/bin/windows/Rtools/.
```

```
## Skipping install of 'spida2' from a github remote, the SHA1 (f79100cc) has not changed since last ins
##   Use `force = TRUE` to force installation
```

```
library(spida2)
tablemissing(data)
```




```
##       Loan_ID Gender Married Dependents Education Self_Employed ApplicantIncome
## 1           1      1       1          1         1             1               1
## 2           1      1       1          1         1             1               1
## 3           1      1       1          1         1             1               1
## 4           1      1       1          1         1             1               1
## Total       0      0       0          0         0             0               0
##       CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History
## 1                     1          1                1              1
## 2                     1          1                1              0
## 3                     1          1                0              1
## 4                     1          0                1              1
## Total                 0         20               13             43
##       Property_Area Loan_Status Total
## 1                 1           1   490
## 2                 1           1    43
## 3                 1           1    13
## 4                 1           1    20
## Total             0           0   566
```

Handling missing data steps:

To reduce the data loss, I replace the missing data of Loan amount to the mean value,then delete the categorical missing data.

#Method:

Handling Missing values

Replace numerical values with mean

Ignore/remove categorical values

```
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 4.0.3
```

```
data$LoanAmount[is.na(data$LoanAmount)] <- mean(data$LoanAmount, na.rm = TRUE)

data1<-na.omit(data)   #Omit the empty categroical rows

dt<-subset(data1,select=-c(Loan_ID))# Deleting the ID column
 dt$Loan_Status <- ifelse(dt$Loan_Status == "Y",1,0)  #encoding the responsd Loan Status
dt$Loan_Status<-factor(as.character(dt$Loan_Status))
```

Amount of missing data is deleted:

Gender :12 rows

Self employ:32 rows

Married 3 rows

Loan_Amount_Term:13 rows

Credit History: 43 rows

Total delete rows: 103

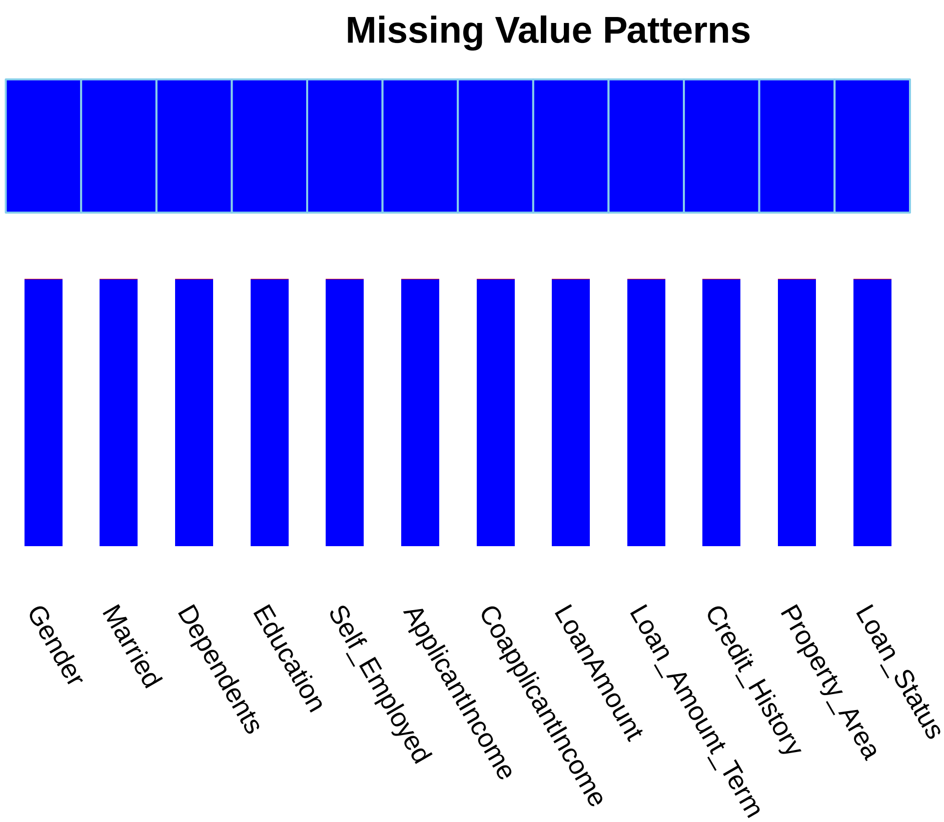
Intial data is 614rows.

Remaining data is 510 rows

17% data was removed.

After processed data

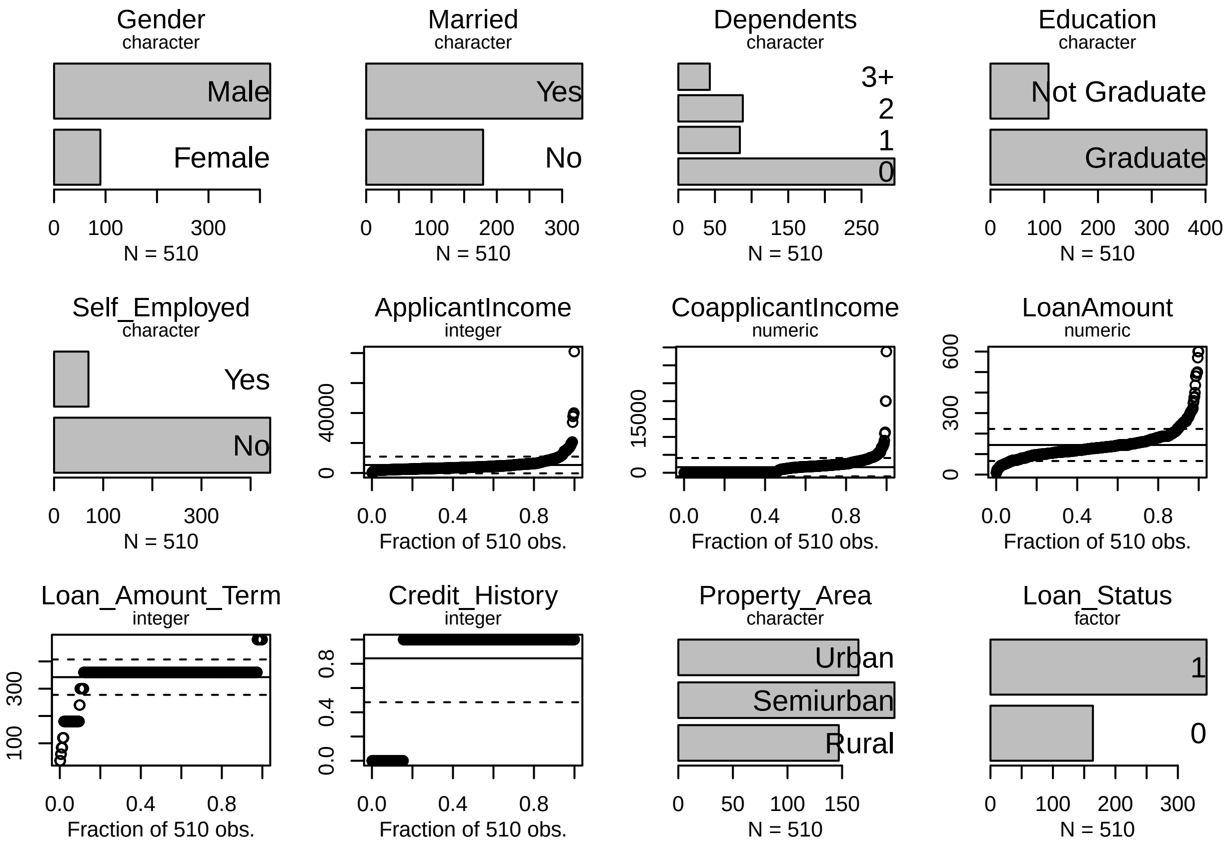
```
tablemissing(dt)
```




```
##       Gender Married Dependents Education Self_Employed ApplicantIncome
## 1          1       1          1         1             1               1
## Total      0       0          0         0             0               0
##       CoapplicantIncome LoanAmount Loan_Amount_Term Credit_History
## 1                     1          1                1              1
## Total                 0          0                0              0
##       Property_Area Loan_Status Total
## 1                 1           1   510
## Total             0           0   510
```

Xqplot

```r
xqplot(dt)
```

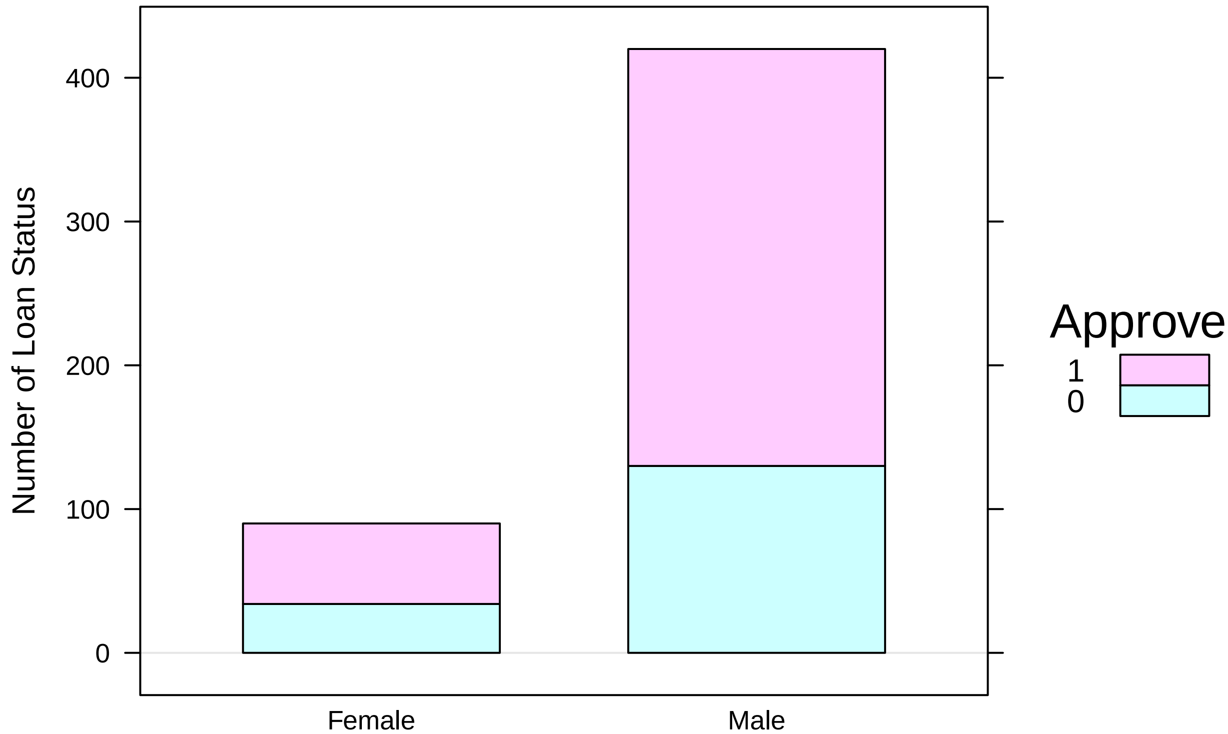


2.2)Relationships between pairs of variables

```
tab(dt,~Gender+Loan_Status)
```

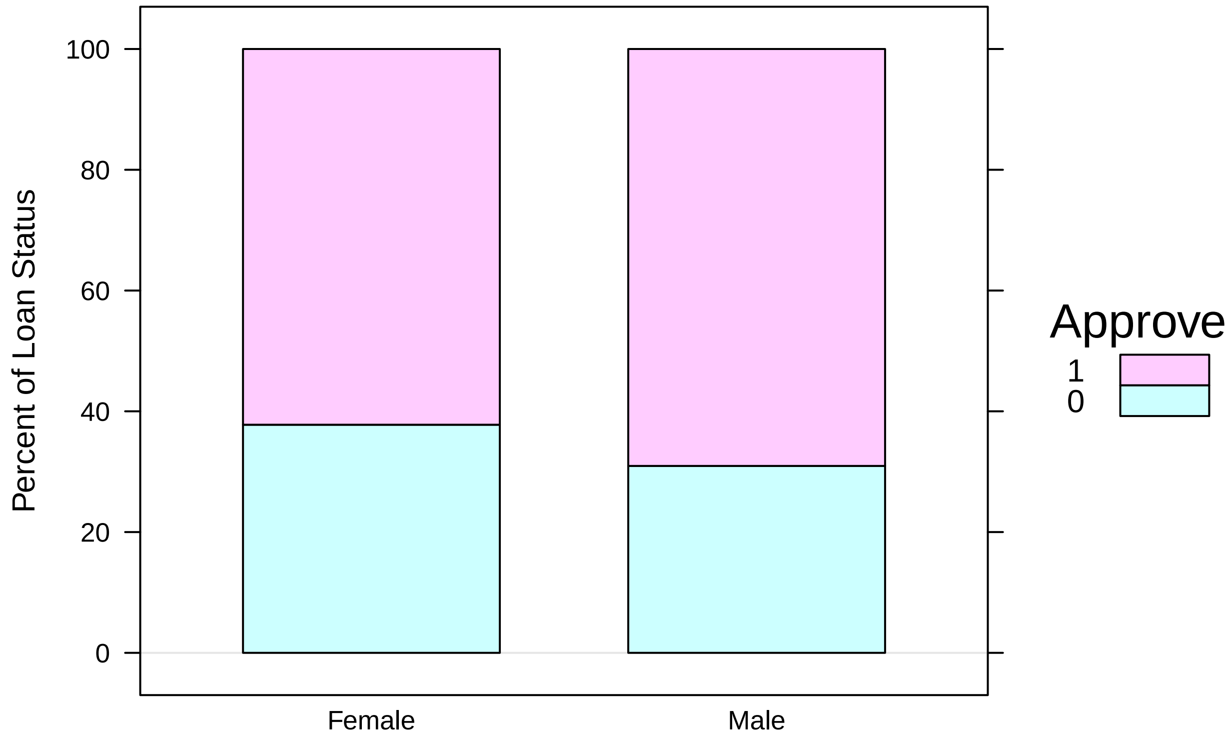
```
##          Loan_Status
## Gender      0   1 Total
##   Female   34  56    90
##   Male    130 290   420
##   Total   164 346   510
```

```
tab__(dt,~Gender+Loan_Status) %>%
barchart(
horizontal = FALSE,
ylab = 'Number of Loan Status',
auto.key = list(
  space = 'right',
  reverse.rows = T,
  title = 'Approve'
     )
)
```



This chart shows the majority of loan applicant is male.

```
  # percent comparision
   tab__(dt,~Gender+Loan_Status,pct=1) %>%
   barchart(
   horizontal = FALSE,
   ylab = 'Percent of Loan Status',
   auto.key = list(
     space = 'right',
     reverse.rows = T,
     title = 'Approve'
   )
 )
```



The chart show the relatively no difference between male and female in getting the loan approve.

```
#fisher and chisq test
tab__(dt,~Gender+Loan_Status) %>% fisher.test
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  .
## p-value = 0.2155
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8148841 2.2255519
## sample estimates:
## odds ratio
##   1.353614
```

```
tab__(dt,~Gender+Loan_Status) %>% chisq.test
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  .
## X-squared = 1.2853, df = 1, p-value = 0.2569
```

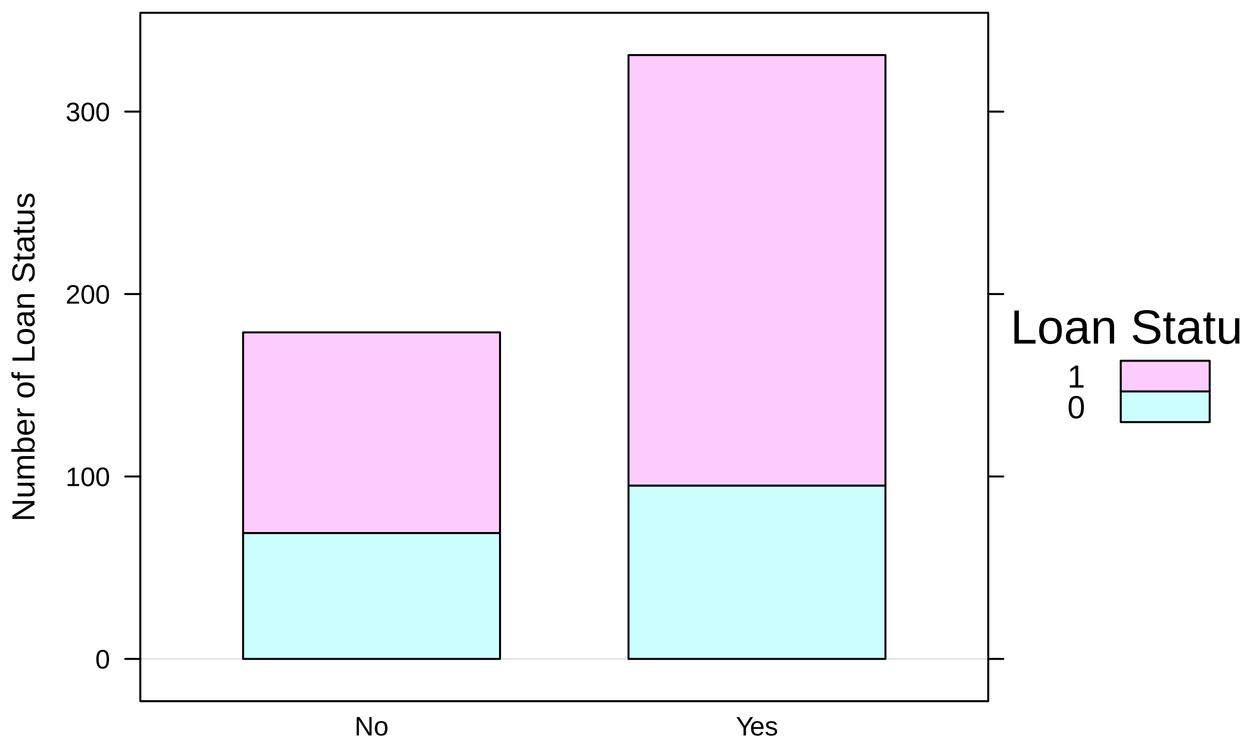
P value lager than 0.05. There is no significant differnece between male and female in getting the loan approve

Married vs Loan Status

```r
tab(dt,~Married+Loan_Status)
```

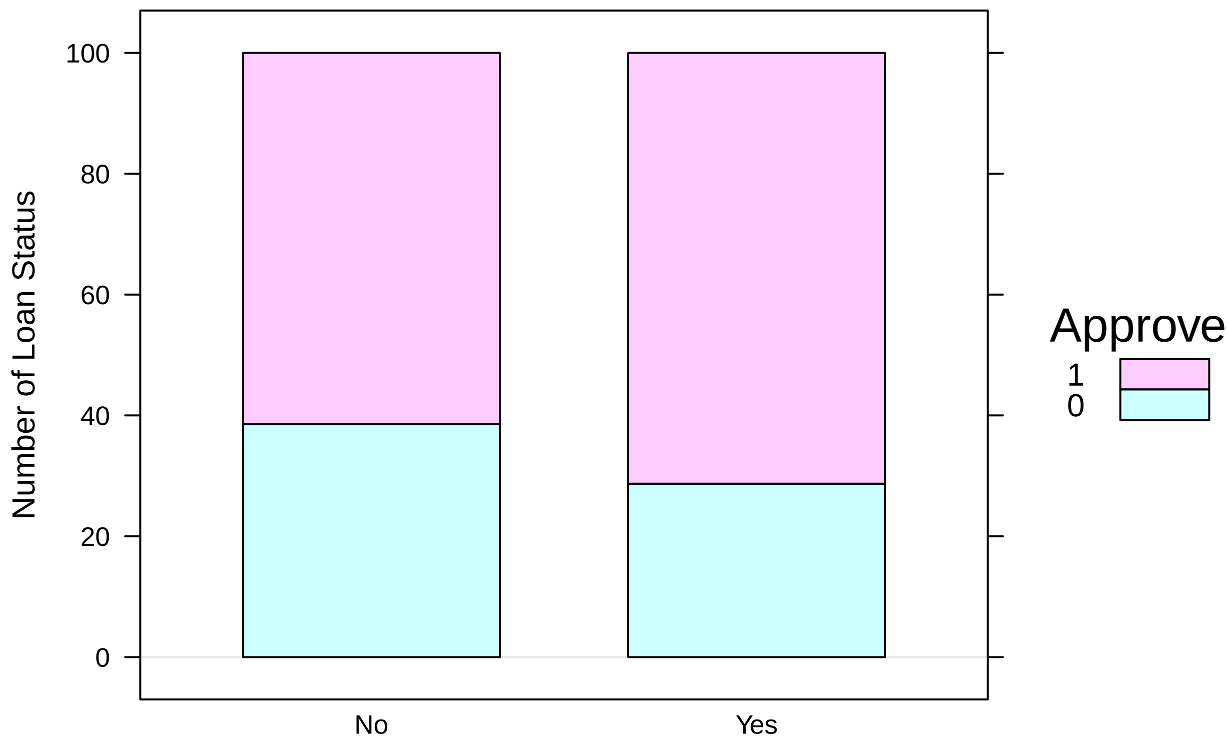
```
##        Loan_Status
## Married   0   1 Total
##   No     69 110   179
##   Yes    95 236   331
##   Total 164 346   510
```

```r
tab__(dt,~Married+Loan_Status) %>%
  barchart(
    horizontal = FALSE,
    ylab = 'Number of Loan Status',
    auto.key = list(
      space = 'right',
      reverse.rows = T,
      title = 'Loan Status'
    )
  )
```



The chart shows the number of married applicants is higher. In my opinion,the married applicant has hig

```r
# percent comparision
tab__(dt,~Married+Loan_Status,pct=1) %>%
  barchart(
    horizontal = FALSE,
    ylab = 'Number of Loan Status',
    auto.key = list(
      space = 'right',
      reverse.rows = T,
      title = 'Approve'
    )
  )
```




```r
#fisher and chisq test
tab__(dt,~Married+Loan_Status) %>% fisher.test
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  .
## p-value = 0.02871
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.040624 2.326569
## sample estimates:
## odds ratio
```

```
##   1.556812
```

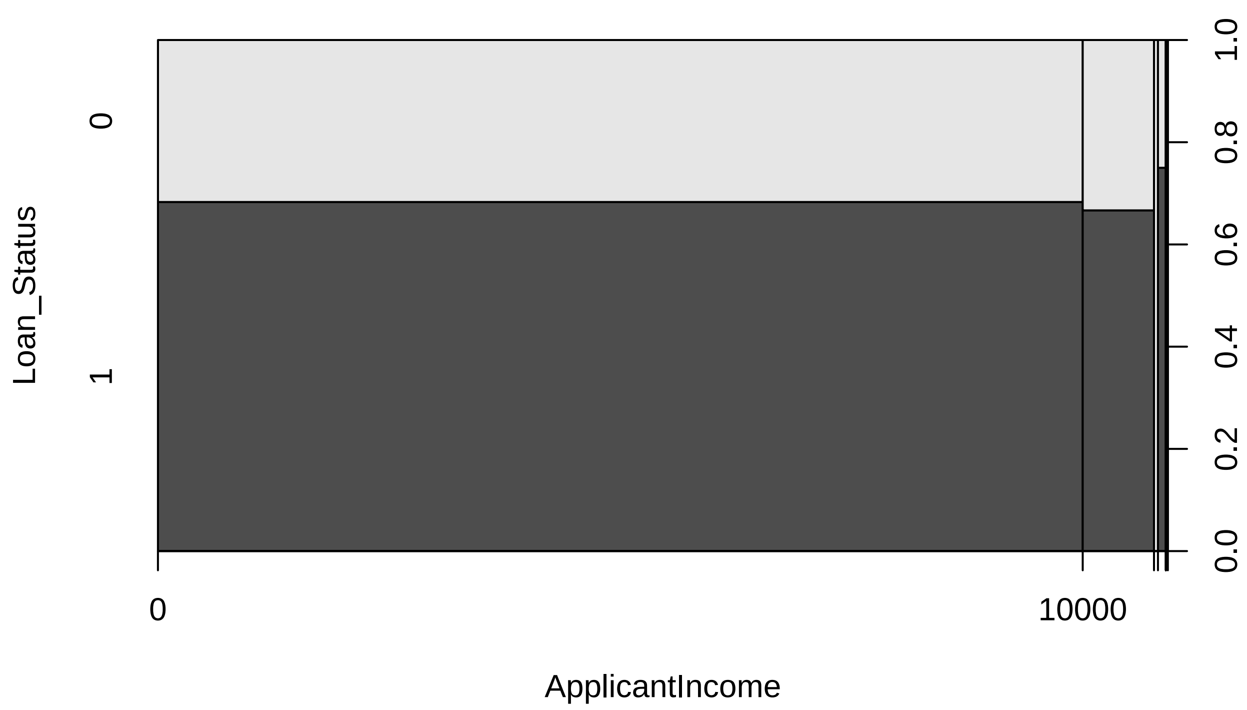
```
tab__(dt,~Married+Loan_Status) %>% chisq.test
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  .
## X-squared = 4.7215, df = 1, p-value = 0.02979
```

P-value is smaller than 0.05, so It shows that the married applicant tend to have higher chance to get a

Loan status vs ApplicantIncome

```
spineplot(Loan_Status~ApplicantIncome,dt)
```



II) Modeling.

3.1 General model

```r
fullmod<-glm(Loan_Status~.,family=binomial,data=dt)
summary(fullmod)
```

```
##
## Call:
## glm(formula = Loan_Status ~ ., family = binomial, data = dt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2715  -0.4063   0.5065   0.7231   2.3840
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -2.518e+00  9.005e-01  -2.797  0.00516 **
## GenderMale              2.719e-01  3.261e-01   0.834  0.40438
## MarriedYes              5.325e-01  2.822e-01   1.887  0.05914 .
## Dependents1            -2.199e-01  3.366e-01  -0.653  0.51360
## Dependents2             3.119e-01  3.638e-01   0.857  0.39127
## Dependents3+            1.431e-01  4.672e-01   0.306  0.75945
## EducationNot Graduate  -5.630e-01  2.865e-01  -1.965  0.04941 *
## Self_EmployedYes       -1.577e-01  3.406e-01  -0.463  0.64332
## ApplicantIncome         1.903e-06  2.956e-05   0.064  0.94867
## CoapplicantIncome      -4.848e-05  4.278e-05  -1.133  0.25718
## LoanAmount             -2.798e-03  1.768e-03  -1.583  0.11340
## Loan_Amount_Term       -5.022e-04  1.924e-03  -0.261  0.79409
## Credit_History          3.723e+00  4.269e-01   8.721  < 2e-16 ***
## Property_AreaSemiurban  9.373e-01  2.967e-01   3.159  0.00158 **
## Property_AreaUrban      1.416e-02  2.870e-01   0.049  0.96064
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 640.61  on 509  degrees of freedom
## Residual deviance: 465.47  on 495  degrees of freedom
## AIC: 495.47
##
## Number of Fisher Scoring iterations: 5
```

Null model

```r
nothing<-glm(Loan_Status~1,family = binomial,data=dt)
summary(nothing)
```

```
##
## Call:
## glm(formula = Loan_Status ~ 1, family = binomial, data = dt)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -1.5064  -1.5064   0.8809   0.8809   0.8809
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.7466     0.0948   7.875 3.41e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 640.61  on 509  degrees of freedom
## Residual deviance: 640.61  on 509  degrees of freedom
## AIC: 642.61
##
## Number of Fisher Scoring iterations: 4
```

From coefficent analysis, it show that Credit History, Property area, education are the significant variables.The model is built below.

```
remod<-glm(Loan_Status~Credit_History+Married+Property_Area+Education,family=binomial,data=dt)
summary(remod)
```

```
##
## Call:
## glm(formula = Loan_Status ~ Credit_History + Married + Property_Area +
##     Education, family = binomial, data = dt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1392  -0.4074   0.5575   0.7072   2.4664
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -2.99259    0.47482  -6.303 2.93e-10 ***
## Credit_History          3.70578    0.42201   8.781  < 2e-16 ***
## MarriedYes              0.54529    0.23519   2.319  0.02042 *
## Property_AreaSemiurban  0.92263    0.29087   3.172  0.00151 **
## Property_AreaUrban      0.08406    0.27611   0.304  0.76078
## EducationNot Graduate  -0.39826    0.27254  -1.461  0.14393
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 640.61  on 509  degrees of freedom
## Residual deviance: 473.54  on 504  degrees of freedom
## AIC: 485.54
##
## Number of Fisher Scoring iterations: 5
```

Modeling stepwise

```r
backwards<-step(fullmod)
```

```
## Start:  AIC=495.47
## Loan_Status ~ Gender + Married + Dependents + Education + Self_Employed +
##     ApplicantIncome + CoapplicantIncome + LoanAmount + Loan_Amount_Term +
##     Credit_History + Property_Area
##
##                     Df Deviance    AIC
## - Dependents         3   467.20 491.20
## - ApplicantIncome    1   465.47 493.47
## - Loan_Amount_Term   1   465.54 493.54
## - Self_Employed      1   465.68 493.68
## - Gender             1   466.16 494.16
## - CoapplicantIncome  1   466.75 494.75
## <none>                   465.47 495.47
## - LoanAmount         1   467.95 495.95
## - Married            1   469.04 497.04
## - Education          1   469.24 497.24
## - Property_Area      2   479.41 505.41
## - Credit_History     1   606.95 634.95
##
## Step:  AIC=491.2
## Loan_Status ~ Gender + Married + Education + Self_Employed +
##     ApplicantIncome + CoapplicantIncome + LoanAmount + Loan_Amount_Term +
##     Credit_History + Property_Area
##
##                     Df Deviance    AIC
## - ApplicantIncome    1   467.20 489.20
## - Loan_Amount_Term   1   467.24 489.24
## - Self_Employed      1   467.37 489.37
## - Gender             1   468.13 490.13
## - CoapplicantIncome  1   468.39 490.39
## <none>                   467.20 491.20
## - LoanAmount         1   469.66 491.66
## - Education          1   470.94 492.94
## - Married            1   471.96 493.96
## - Property_Area      2   480.67 500.67
## - Credit_History     1   608.86 630.86
##
## Step:  AIC=489.2
## Loan_Status ~ Gender + Married + Education + Self_Employed +
##     CoapplicantIncome + LoanAmount + Loan_Amount_Term + Credit_History +
##     Property_Area
##
##                     Df Deviance    AIC
## - Loan_Amount_Term   1   467.24 487.24
## - Self_Employed      1   467.38 487.38
## - Gender             1   468.13 488.13
## - CoapplicantIncome  1   468.49 488.49
## <none>                   467.20 489.20
```

```
## - LoanAmount         1   470.58 490.58
## - Education          1   470.97 490.97
## - Married            1   471.97 491.97
## - Property_Area      2   480.67 498.67
## - Credit_History     1   609.00 629.00
##
## Step:  AIC=487.24
## Loan_Status ~ Gender + Married + Education + Self_Employed +
##     CoapplicantIncome + LoanAmount + Credit_History + Property_Area
##
##                     Df Deviance    AIC
## - Self_Employed      1   467.41 485.41
## - Gender             1   468.19 486.19
## - CoapplicantIncome  1   468.53 486.53
## <none>                   467.24 487.24
## - LoanAmount         1   470.64 488.64
## - Education          1   470.97 488.97
## - Married            1   472.13 490.13
## - Property_Area      2   480.70 496.70
## - Credit_History     1   609.01 627.01
##
## Step:  AIC=485.41
## Loan_Status ~ Gender + Married + Education + CoapplicantIncome +
##     LoanAmount + Credit_History + Property_Area
##
##                     Df Deviance    AIC
## - Gender             1   468.39 484.39
## - CoapplicantIncome  1   468.70 484.70
## <none>                   467.41 485.41
## - LoanAmount         1   471.05 487.05
## - Education          1   471.18 487.18
## - Married            1   472.26 488.26
## - Property_Area      2   480.83 494.83
## - Credit_History     1   609.22 625.22
##
## Step:  AIC=484.39
## Loan_Status ~ Married + Education + CoapplicantIncome + LoanAmount +
##     Credit_History + Property_Area
##
##                     Df Deviance    AIC
## - CoapplicantIncome  1   469.38 483.38
## <none>                   468.39 484.39
## - LoanAmount         1   471.91 485.91
## - Education          1   471.94 485.94
## - Married            1   475.79 489.79
## - Property_Area      2   481.17 493.17
## - Credit_History     1   610.90 624.90
##
## Step:  AIC=483.38
## Loan_Status ~ Married + Education + LoanAmount + Credit_History +
##     Property_Area
##
##                  Df Deviance    AIC
## <none>                469.38 483.38
```

```
## - Education       1   472.72 484.72
## - LoanAmount      1   473.54 485.54
## - Married         1   476.50 488.50
## - Property_Area   2   482.26 492.26
## - Credit_History  1   611.73 623.73
```

```r
formula(backwards)
```

```
## Loan_Status ~ Married + Education + LoanAmount + Credit_History +
##     Property_Area
```

Result:

```r
formula(backwards)
```

```
## Loan_Status ~ Married + Education + LoanAmount + Credit_History +
##     Property_Area
```

```r
backwards[["aic"]]
```

```
## [1] 483.3796
```

```r
backwards[["deviance"]]
```

```
## [1] 469.3796
```

Forward direction

```r
   # Forward direction.
forwards <- step(nothing,scope=list(lower=formula(nothing),upper=formula(fullmod)), direction="forward")
```

```
## Start:  AIC=642.61
## Loan_Status ~ 1
##
##                     Df Deviance    AIC
## + Credit_History     1   494.24 498.24
## + Property_Area      2   626.84 632.84
## + Married            1   635.51 639.51
## + Education          1   636.11 640.11
## + LoanAmount         1   638.30 642.30
## <none>                   640.61 642.61
## + Gender             1   639.06 643.06
## + CoapplicantIncome  1   639.81 643.81
## + ApplicantIncome    1   639.85 643.85
## + Self_Employed      1   640.14 644.14
## + Loan_Amount_Term   1   640.58 644.58
## + Dependents         3   637.94 645.94
##
```

```
## Step:  AIC=498.24
## Loan_Status ~ Credit_History
##
##                     Df Deviance    AIC
## + Property_Area      2   480.79 488.79
## + Married            1   488.91 494.91
## + Education          1   491.89 497.89
## <none>                   494.24 498.24
## + LoanAmount         1   492.57 498.57
## + Gender             1   493.11 499.11
## + CoapplicantIncome  1   493.28 499.28
## + Self_Employed      1   493.94 499.94
## + ApplicantIncome    1   494.01 500.01
## + Loan_Amount_Term   1   494.12 500.12
## + Dependents         3   491.91 501.91
##
## Step:  AIC=488.79
## Loan_Status ~ Credit_History + Property_Area
##
##                     Df Deviance    AIC
## + Married            1   475.62 485.62
## + Gender             1   478.68 488.68
## <none>                   480.79 488.79
## + Education          1   478.88 488.88
## + LoanAmount         1   479.22 489.22
## + CoapplicantIncome  1   479.87 489.87
## + Self_Employed      1   480.47 490.47
## + ApplicantIncome    1   480.54 490.54
## + Loan_Amount_Term   1   480.63 490.63
## + Dependents         3   477.74 491.74
##
## Step:  AIC=485.62
## Loan_Status ~ Credit_History + Property_Area + Married
##
##                     Df Deviance    AIC
## + LoanAmount         1   472.72 484.72
## + Education          1   473.54 485.54
## <none>                   475.62 485.62
## + CoapplicantIncome  1   474.32 486.32
## + Gender             1   475.20 487.20
## + Self_Employed      1   475.21 487.21
## + ApplicantIncome    1   475.26 487.26
## + Loan_Amount_Term   1   475.60 487.60
## + Dependents         3   473.87 489.87
##
## Step:  AIC=484.72
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount
##
##                     Df Deviance    AIC
## + Education          1   469.38 483.38
## <none>                   472.72 484.72
## + CoapplicantIncome  1   471.94 485.94
## + Gender             1   472.19 486.19
## + Self_Employed      1   472.51 486.51
```

```
## + ApplicantIncome    1   472.60 486.60
## + Loan_Amount_Term   1   472.72 486.72
## + Dependents         3   470.98 488.98
##
## Step:  AIC=483.38
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount +
##     Education
##
##                      Df Deviance    AIC
## <none>                   469.38 483.38
## + CoapplicantIncome  1   468.39 484.39
## + Gender             1   468.70 484.70
## + Self_Employed      1   469.20 485.20
## + ApplicantIncome    1   469.32 485.32
## + Loan_Amount_Term   1   469.32 485.32
## + Dependents         3   467.59 487.59
```

Result:

```r
formula(forwards)
```

```
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount +
##     Education
```

```r
forwards[["aic"]]
```

```
## [1] 483.3796
```

```r
forwards[["deviance"]]
```

```
## [1] 469.3796
```

```r
bothway <- step(nothing,scope=list(lower=formula(nothing),upper=formula(fullmod)), direction="both")
```

```
## Start:  AIC=642.61
## Loan_Status ~ 1
##
##                      Df Deviance    AIC
## + Credit_History     1   494.24 498.24
## + Property_Area      2   626.84 632.84
## + Married            1   635.51 639.51
## + Education          1   636.11 640.11
## + LoanAmount         1   638.30 642.30
## <none>                   640.61 642.61
## + Gender             1   639.06 643.06
## + CoapplicantIncome  1   639.81 643.81
## + ApplicantIncome    1   639.85 643.85
## + Self_Employed      1   640.14 644.14
## + Loan_Amount_Term   1   640.58 644.58
## + Dependents         3   637.94 645.94
##
```

```
## Step:  AIC=498.24
## Loan_Status ~ Credit_History
##
##                     Df Deviance    AIC
## + Property_Area      2   480.79 488.79
## + Married            1   488.91 494.91
## + Education          1   491.89 497.89
## <none>                   494.24 498.24
## + LoanAmount         1   492.57 498.57
## + Gender             1   493.11 499.11
## + CoapplicantIncome  1   493.28 499.28
## + Self_Employed      1   493.94 499.94
## + ApplicantIncome    1   494.01 500.01
## + Loan_Amount_Term   1   494.12 500.12
## + Dependents         3   491.91 501.91
## - Credit_History     1   640.61 642.61
##
## Step:  AIC=488.79
## Loan_Status ~ Credit_History + Property_Area
##
##                     Df Deviance    AIC
## + Married            1   475.62 485.62
## + Gender             1   478.68 488.68
## <none>                   480.79 488.79
## + Education          1   478.88 488.88
## + LoanAmount         1   479.22 489.22
## + CoapplicantIncome  1   479.87 489.87
## + Self_Employed      1   480.47 490.47
## + ApplicantIncome    1   480.54 490.54
## + Loan_Amount_Term   1   480.63 490.63
## + Dependents         3   477.74 491.74
## - Property_Area      2   494.24 498.24
## - Credit_History     1   626.84 632.84
##
## Step:  AIC=485.62
## Loan_Status ~ Credit_History + Property_Area + Married
##
##                     Df Deviance    AIC
## + LoanAmount         1   472.72 484.72
## + Education          1   473.54 485.54
## <none>                   475.62 485.62
## + CoapplicantIncome  1   474.32 486.32
## + Gender             1   475.20 487.20
## + Self_Employed      1   475.21 487.21
## + ApplicantIncome    1   475.26 487.26
## + Loan_Amount_Term   1   475.60 487.60
## - Married            1   480.79 488.79
## + Dependents         3   473.87 489.87
## - Property_Area      2   488.91 494.91
## - Credit_History     1   621.72 629.72
##
## Step:  AIC=484.72
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount
##
```

```
##                      Df Deviance    AIC
## + Education          1   469.38 483.38
## <none>                   472.72 484.72
## - LoanAmount         1   475.62 485.62
## + CoapplicantIncome  1   471.94 485.94
## + Gender             1   472.19 486.19
## + Self_Employed      1   472.51 486.51
## + ApplicantIncome    1   472.60 486.60
## + Loan_Amount_Term   1   472.72 486.72
## + Dependents         3   470.98 488.98
## - Married            1   479.22 489.22
## - Property_Area      2   485.90 493.90
## - Credit_History     1   617.69 627.69
##
## Step:  AIC=483.38
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount +
##     Education
##
##                      Df Deviance    AIC
## <none>                   469.38 483.38
## + CoapplicantIncome  1   468.39 484.39
## + Gender             1   468.70 484.70
## - Education          1   472.72 484.72
## + Self_Employed      1   469.20 485.20
## + ApplicantIncome    1   469.32 485.32
## + Loan_Amount_Term   1   469.32 485.32
## - LoanAmount         1   473.54 485.54
## + Dependents         3   467.59 487.59
## - Married            1   476.50 488.50
## - Property_Area      2   482.26 492.26
## - Credit_History     1   611.73 623.73
```

Both way direction

Result:

```
formula(forwards)
```

```
## Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount +
##     Education
```

```
forwards[["aic"]]
```

```
## [1] 483.3796
```

```
forwards[["deviance"]]
```

```
## [1] 469.3796
```

The final result of modeling:

Final additive model:

Loan_Status ~ Credit_History + Property_Area + Married + LoanAmount + Education

In using step-wise strategy include backward,forward, and both directions, the same model is given and 5 significant variables are choosen(Credit History, Propery Area, Married,Loan Amount).The deviance of the model is 469.38 and AIC is 483.3796 with 503 degree of free doom.

3.2 Interaction between variables.

```
There will be 10 interactions for 5 variables (because 5 choice 2 = 10),
but the interaction between Credit History and Loan amount is significant.
```

Interaction between Credit History and Loan amount

```r
inta<-glm(Loan_Status~ Credit_History*LoanAmount+Married+Education+Property_Area,family=binomial,data=dt
summary(inta)
```

```
##
## Call:
## glm(formula = Loan_Status ~ Credit_History * LoanAmount + Married +
##     Education + Property_Area, family = binomial, data = dt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2936  -0.3970   0.5283   0.7231   2.5187
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -3.572883   0.762412  -4.686 2.78e-06 ***
## Credit_History            4.901553   0.769979   6.366 1.94e-10 ***
## LoanAmount                0.003413   0.003243   1.052  0.29262
## MarriedYes                0.652805   0.244754   2.667  0.00765 **
## EducationNot Graduate    -0.543352   0.283350  -1.918  0.05516 .
## Property_AreaSemiurban    0.854511   0.293366   2.913  0.00358 **
## Property_AreaUrban       -0.024158   0.282443  -0.086  0.93184
## Credit_History:LoanAmount -0.007420   0.003538  -2.097  0.03595 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 640.61  on 509  degrees of freedom
## Residual deviance: 465.60  on 502  degrees of freedom
## AIC: 481.6
##
## Number of Fisher Scoring iterations: 5
```

#Checking the adding interaction to model

```r
anova(backwards,inta,test="LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: Loan_Status ~ Married + Education + LoanAmount + Credit_History +
##     Property_Area
## Model 2: Loan_Status ~ Credit_History * LoanAmount + Married + Education +
##     Property_Area
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       503     469.38
## 2       502     465.60  1   3.7814  0.05183 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significant difference interaction between Credit_History and LoanAmount, because the p-value of Credit_History and Loan Amount is 0.05183 (smaller than 0.05).This is called 'a test of parallelism'. There is a improvement in deviance and AIC in adding interaction to the model. The coeficient of Credit_History:LoanAmount is -0.007420.It mean when the large amount of the loan with a good credit history tend to decrease the probabilty.

```r
library(interactions)
```

```
## Warning: package 'interactions' was built under R version 4.0.3
```

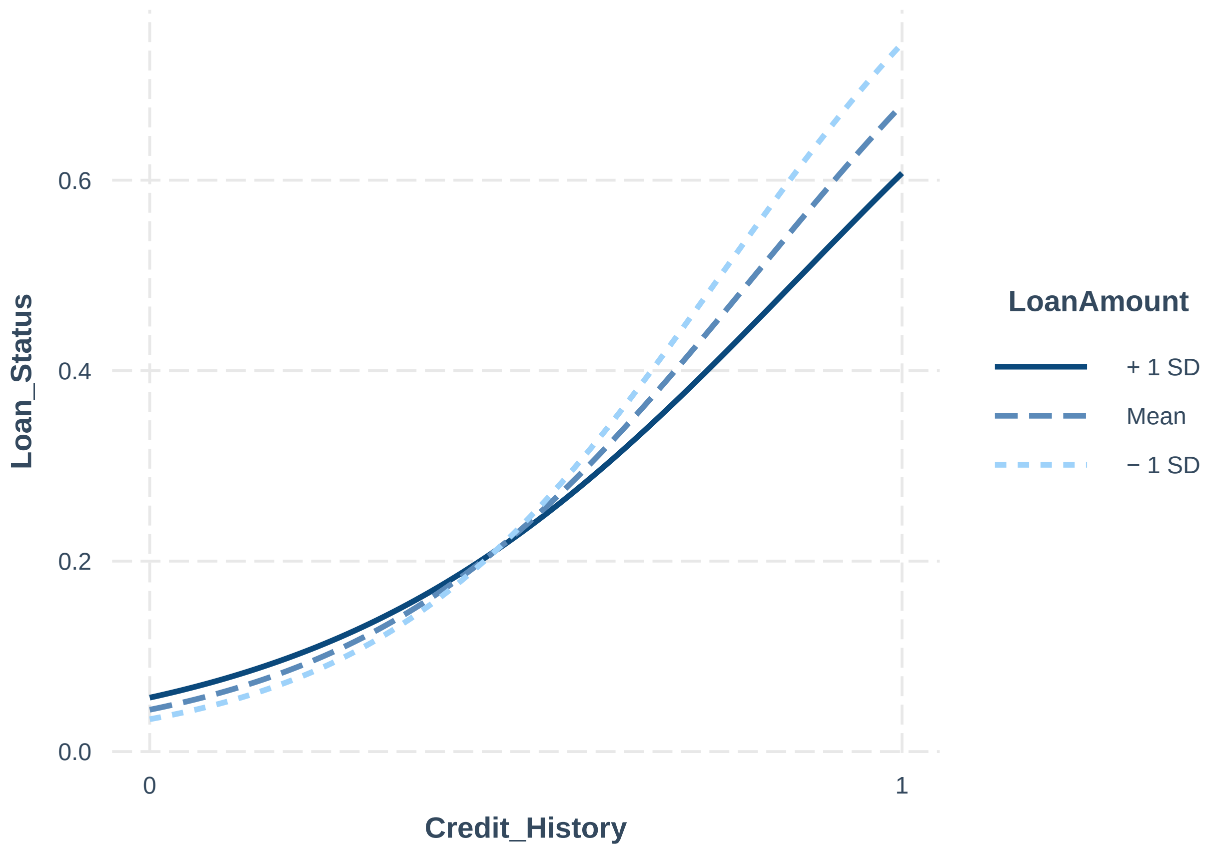
```r
library(jtools)
```

```
## Warning: package 'jtools' was built under R version 4.0.3
```

```
##
## Attaching package: 'jtools'
```

```
## The following object is masked from 'package:spida2':
##
##     center
```

```r
interact_plot(inta, pred = Credit_History, modx = LoanAmount)
```



The graph shows the intersection between 1+Sd,mean,-1 SD amount as an evidence of the interaction between Credit History and Loan amount.

Iv) Conclusion:

```
The final model and good fit is:

Loan_Status ~ Credit_History + LoanAmount + Married + Education +Property_Area+Credit_History:LoanAmou
```

5 significant variables to explain the response (Loan Status):

Credit_History

Loan Amount

Married

Education

Property_Area

```
Those variables are very meaningful in decision whether the bank should approve the application or not.
```