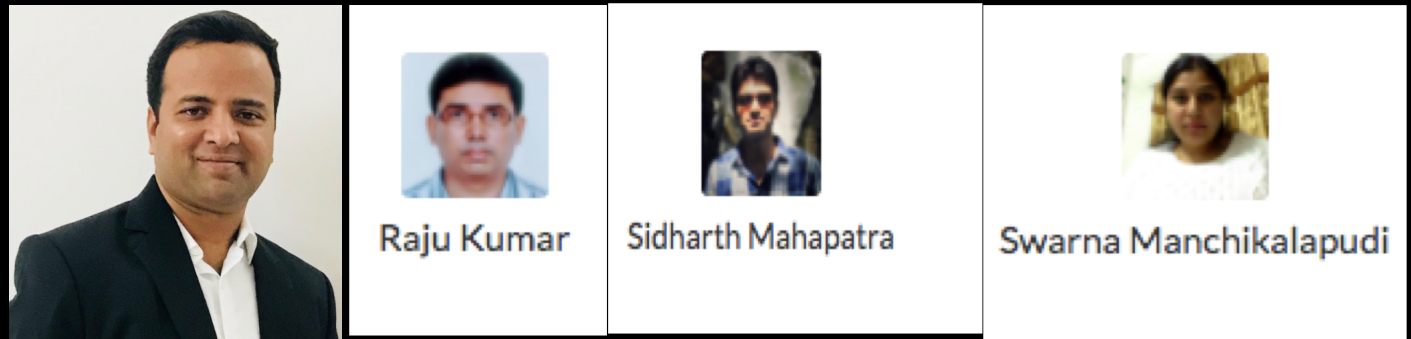
Gramener Case Study

Group Name:

1. **Sathish Devunuri (group facilitator)**
2. **Raju Kumar**
3. **Sidharth Mahapatra**
4. **Swarna Manchikalapudi**



Abstract


BANK


LOAN
APPLICATION
APPROVED


LOAN
APPLICATION
REJECTED


DEFAULT
ON
LOANS


LOANS
BORROW
Repay Loans
REPAY


Competitive Returns


Meeting Your
Financial Needs


ANNUAL
INCOME
> 100.000 $
> 50.000 $
> 15.000 $
< 15.000 $


CLOSED


PAST DUE


NOTICE OF
LOAN DEFAULT

Business Understanding

1. Business Understanding?

- **Business Objective:**
  - **The finance company is looking for the attributes in a applicant profile which can help them is deciding whether to approve or decline the loan application.**

- **Goal of Analysis:**
  - **To find out the relation between the different attributes and their impact on loan default. And suggest which attributes contributes a significant difference in Loan Default.**

Data Understanding & Driving Variables

Loan Amount

Funded Amount

Interest Rate

Purpose

Tenure

Employee Length

Home Ownership

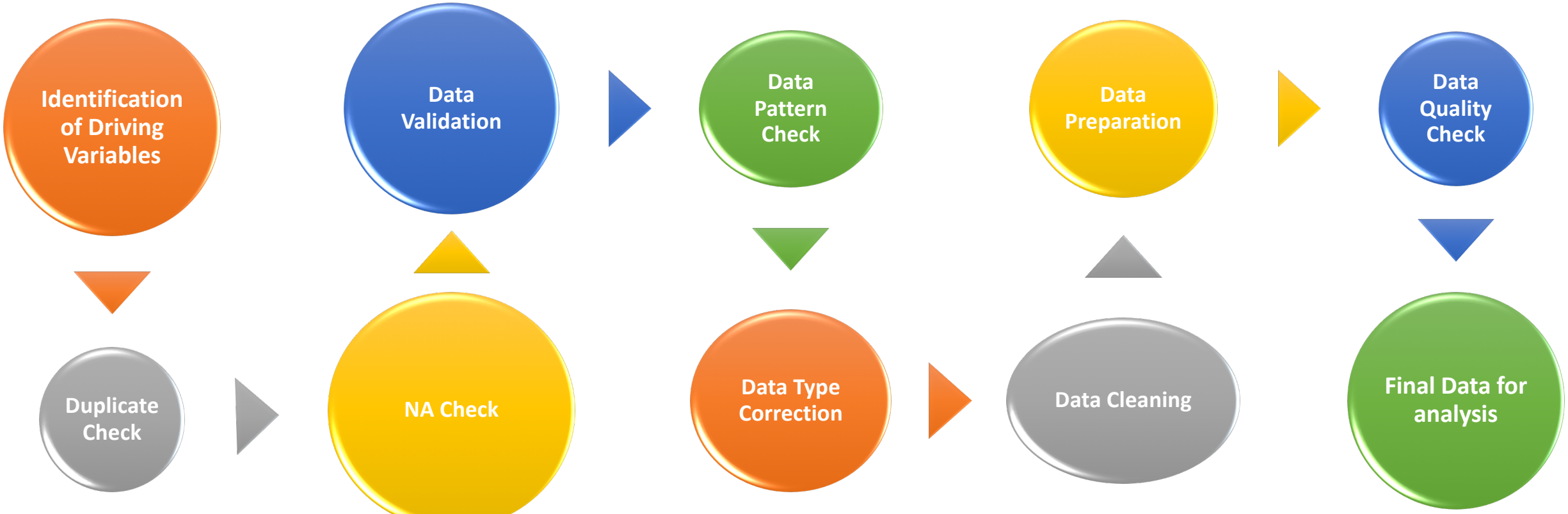
© Can Stock Photo - csp6879400

Annual Income

Grade

&

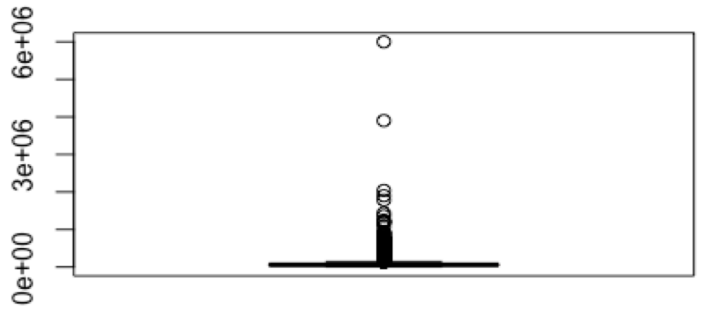
Data Cleaning and preparation



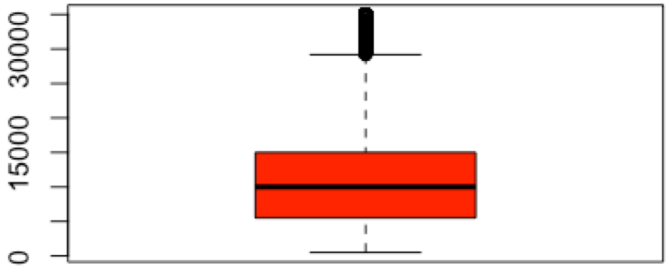
Exploratory Data Analysis - Univariate Analysis

Outlier detection of Annual Income, loan amount and funded amount have outliers present

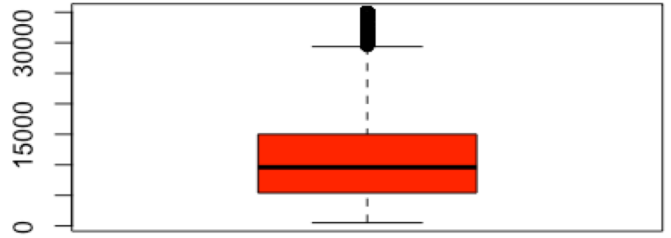
annual_inc



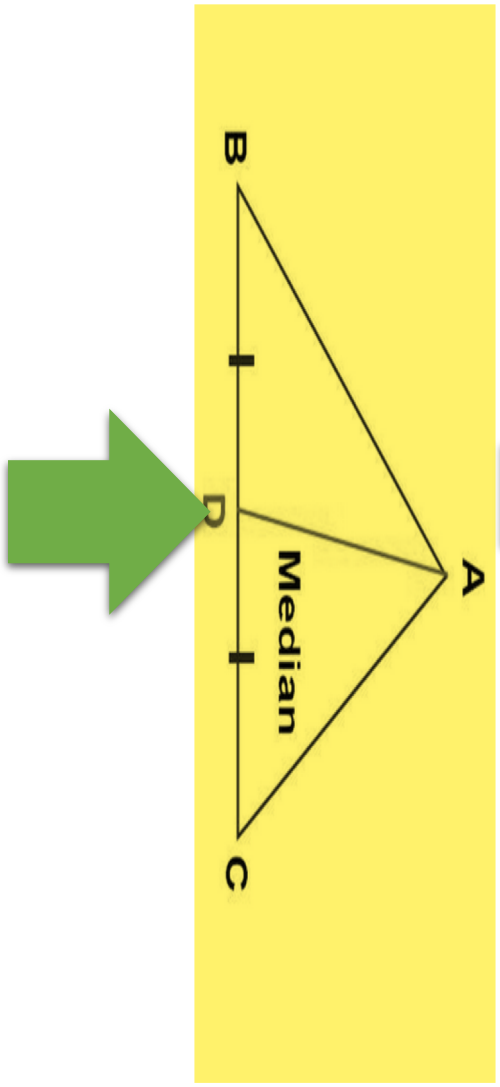
loan_amnt



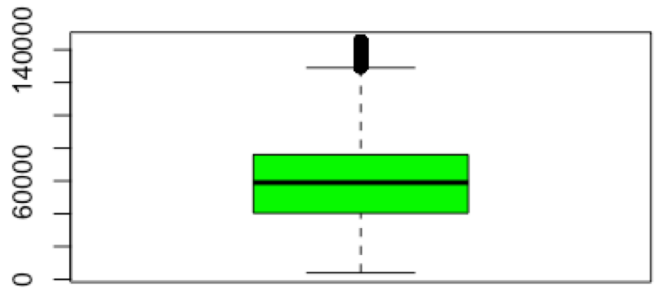
funded_amnt



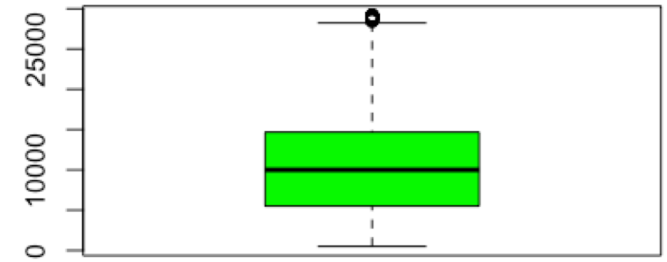
Outliner Present



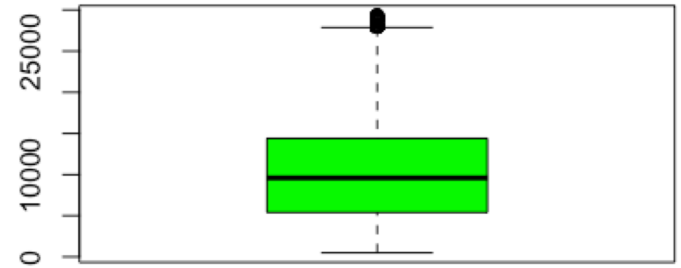
annual_inc



loan_amnt

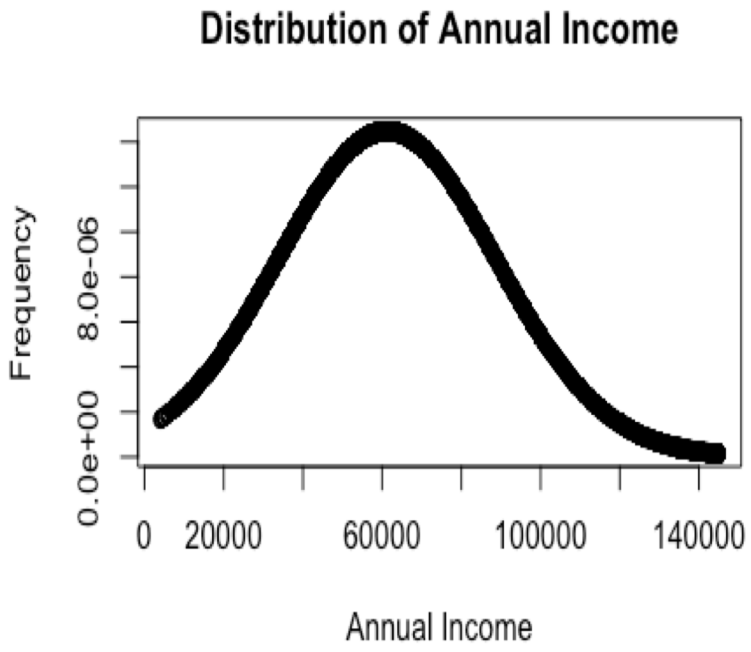


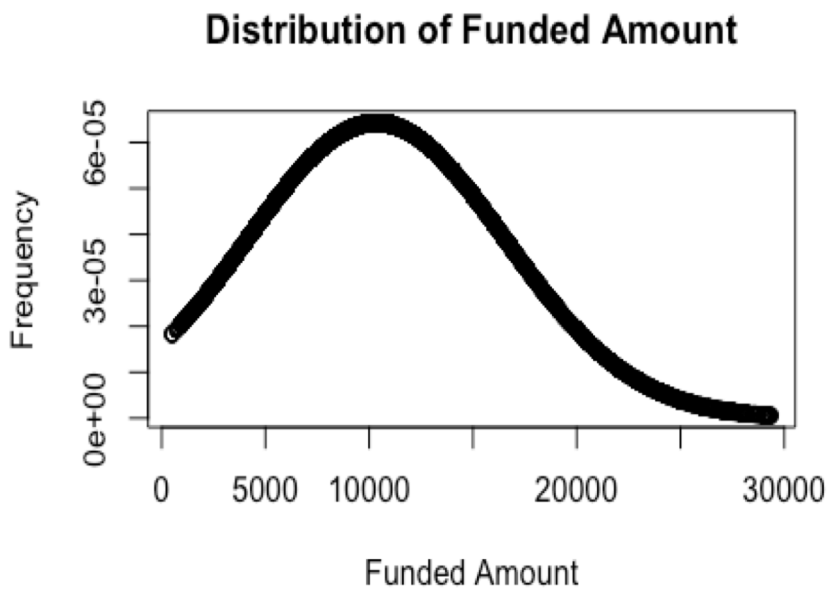
funded_amnt

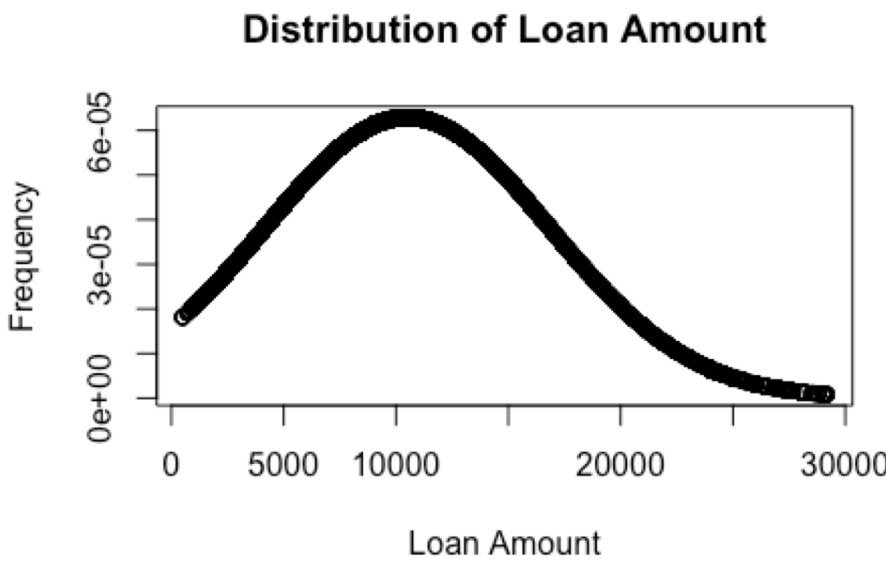


Outliner corrected with Median

Distribution of the Outlier Variable

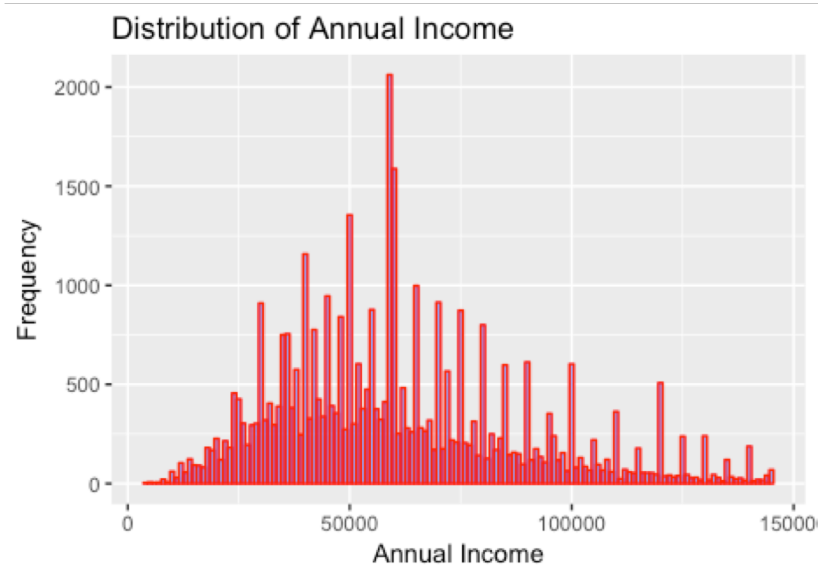


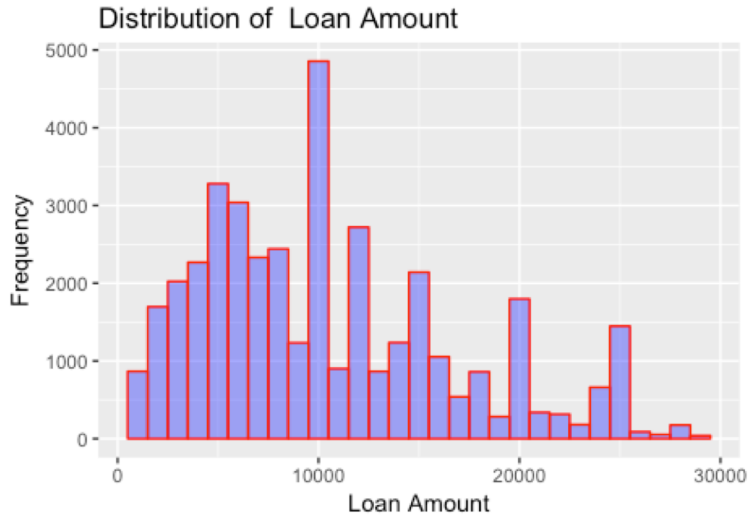


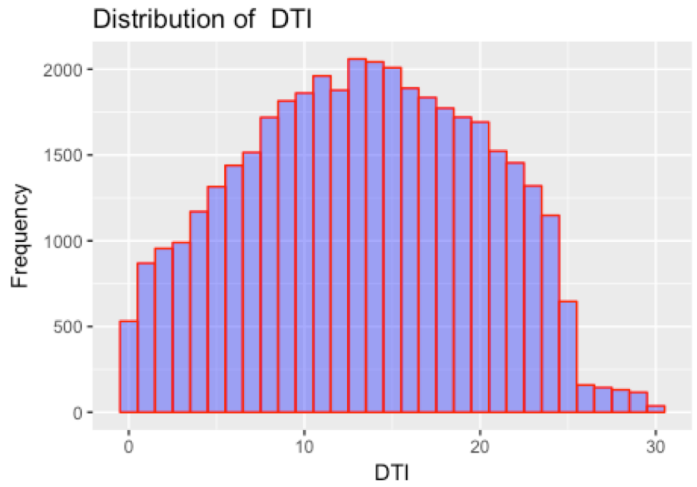


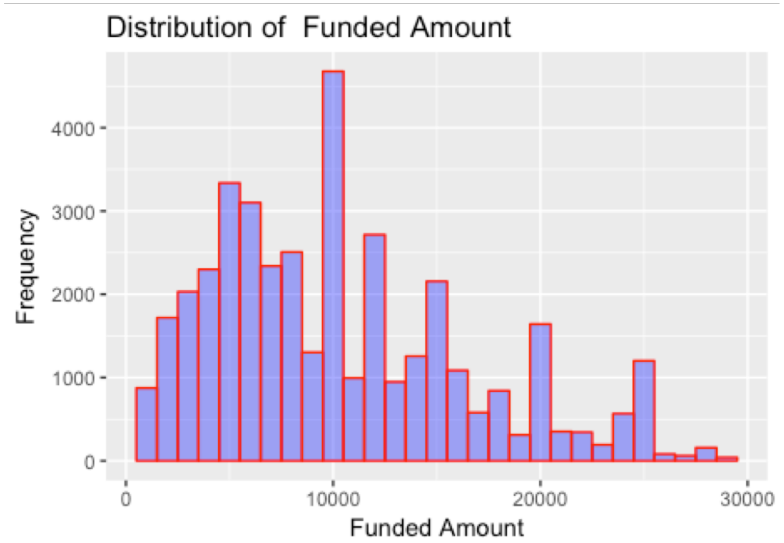
After treating all the outlier values of continuous variable by using median , The Distribution shows in normal

Histograms to Visually verify the results of Distribution



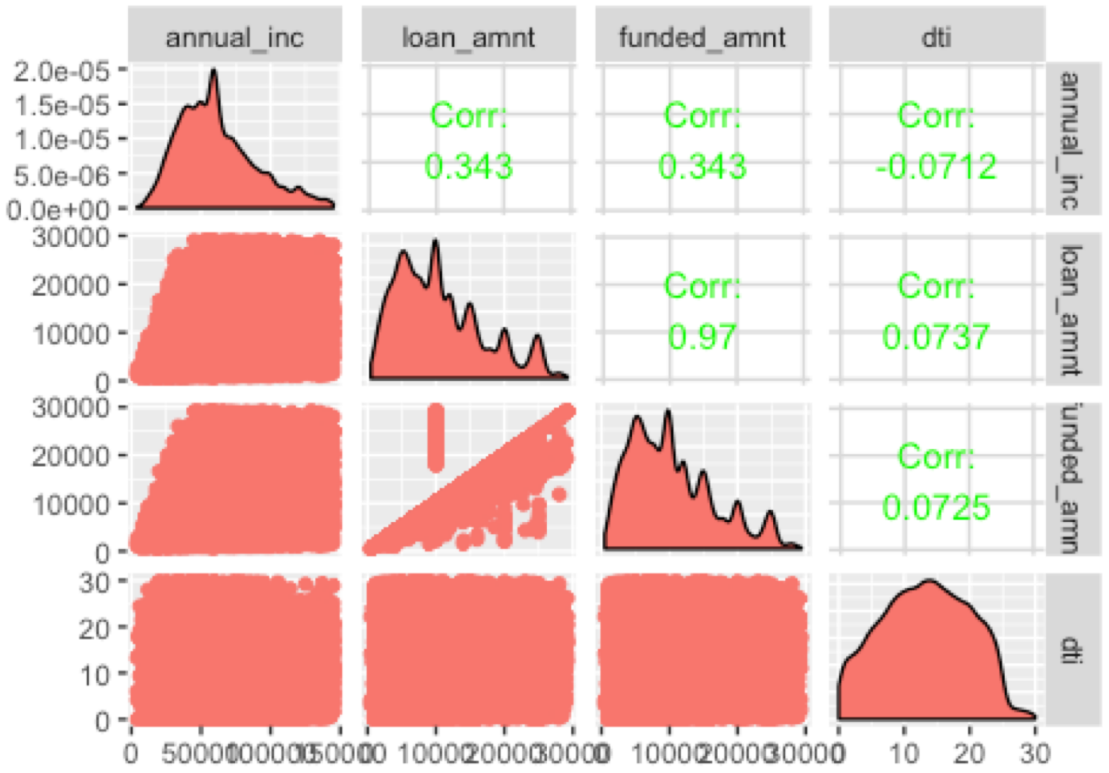


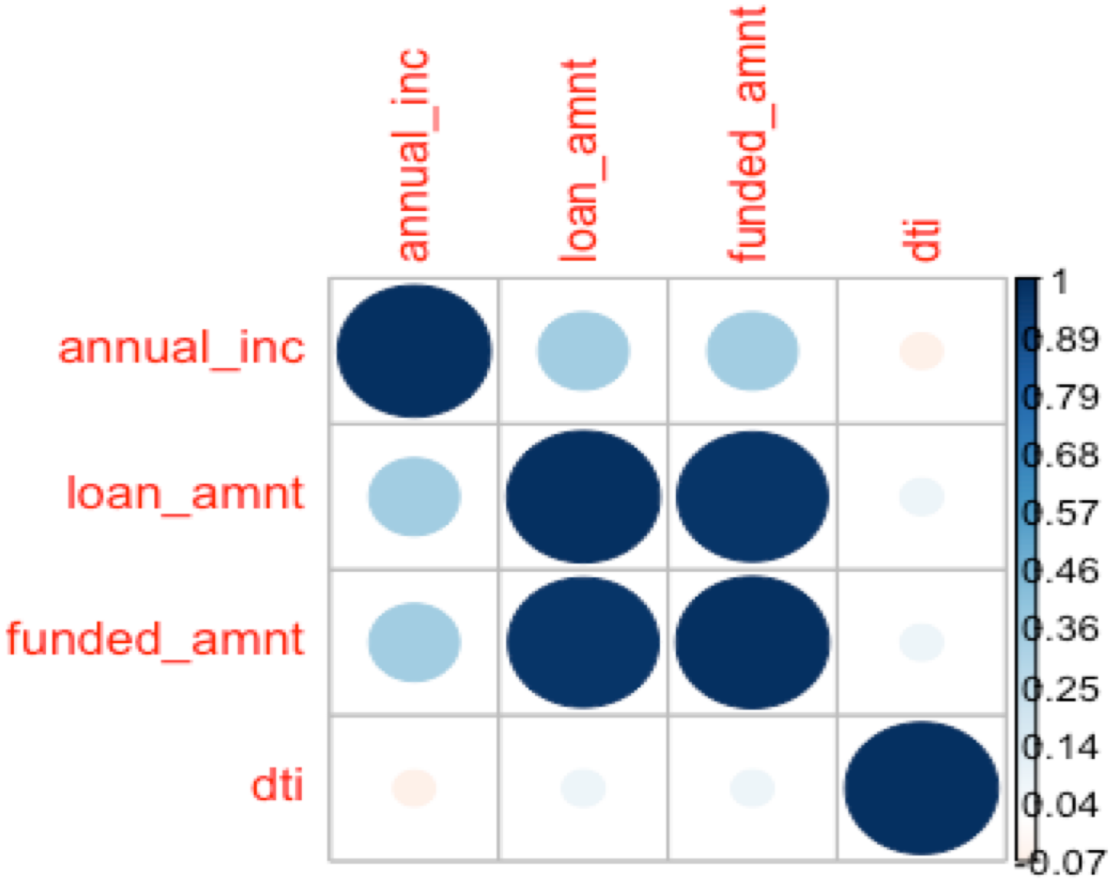




Exploratory Data Analysis - Multivariate Analysis

correlation between any of the continuous variables

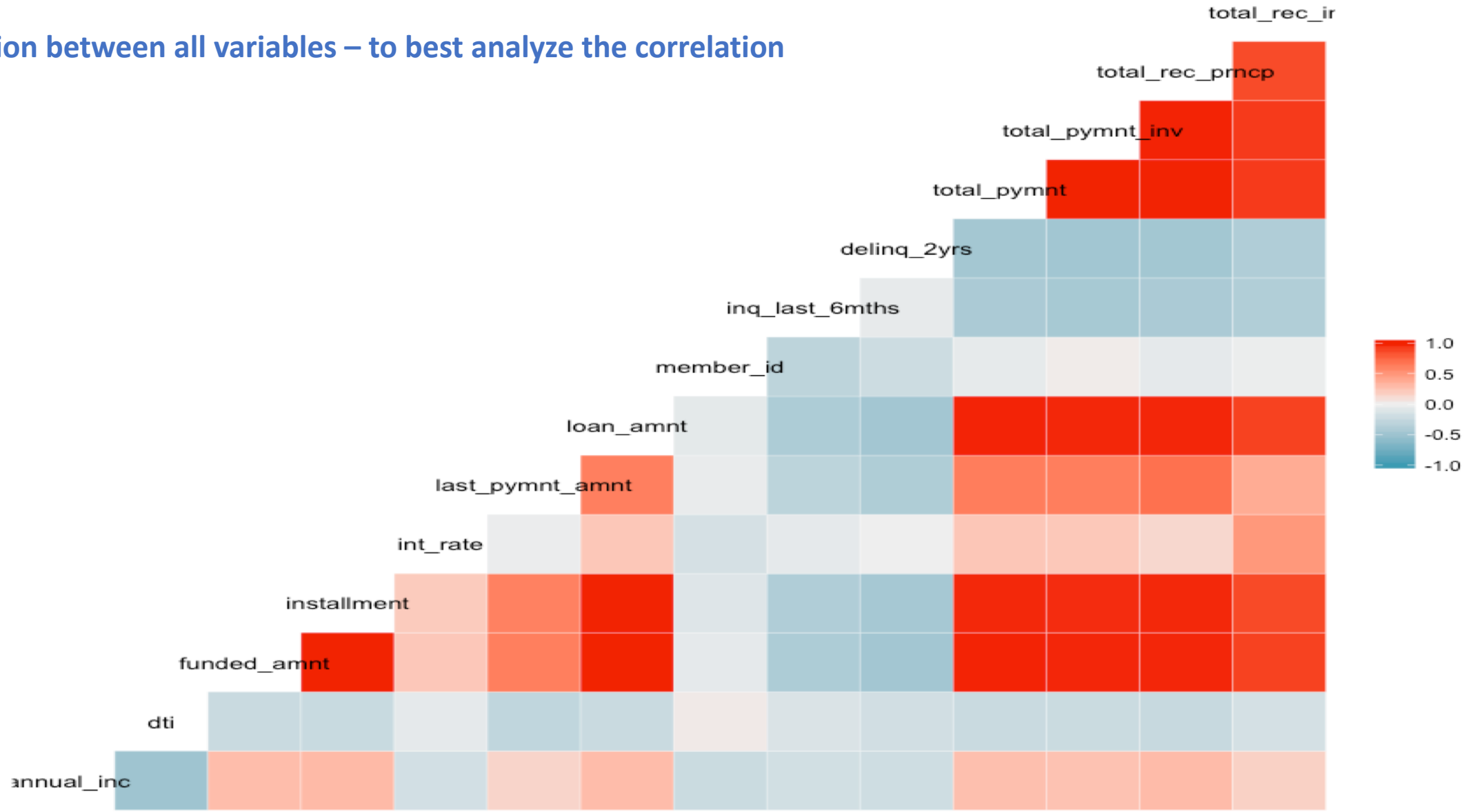




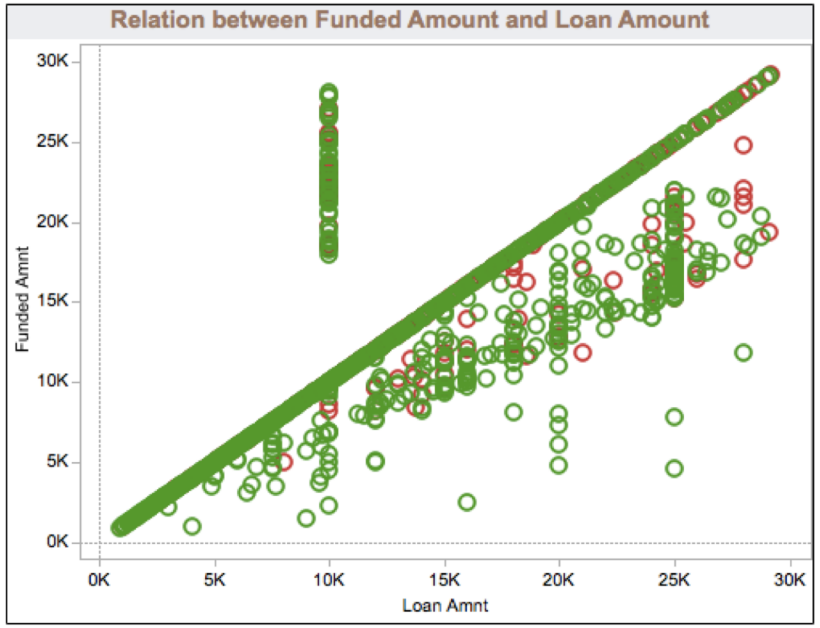
- Based on the above correlation analysis between the key variables we can see below output
- Funded Amount and Loan Amount, and the correlation value is 0.97004145 which is equal to 1 positive correlated
- This means if the loan amount increases funded amount also increases which is what we expect

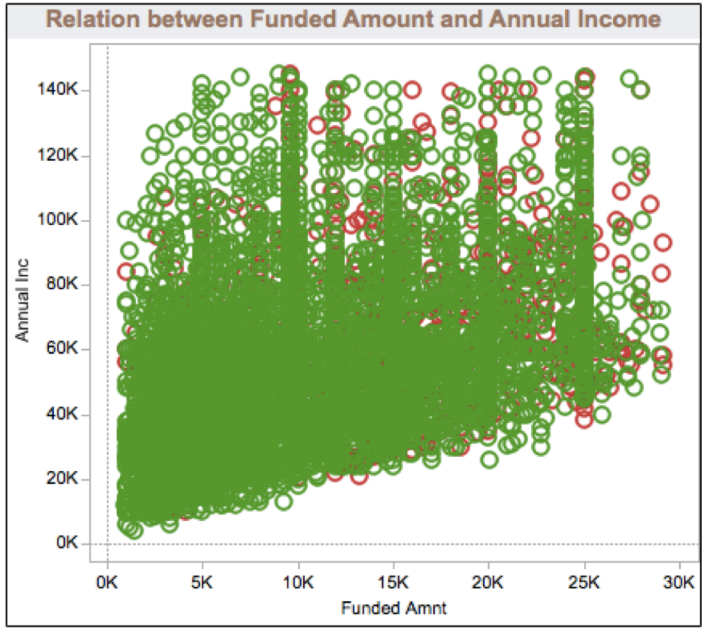
Exploratory Data Analysis - Multivariate Analysis

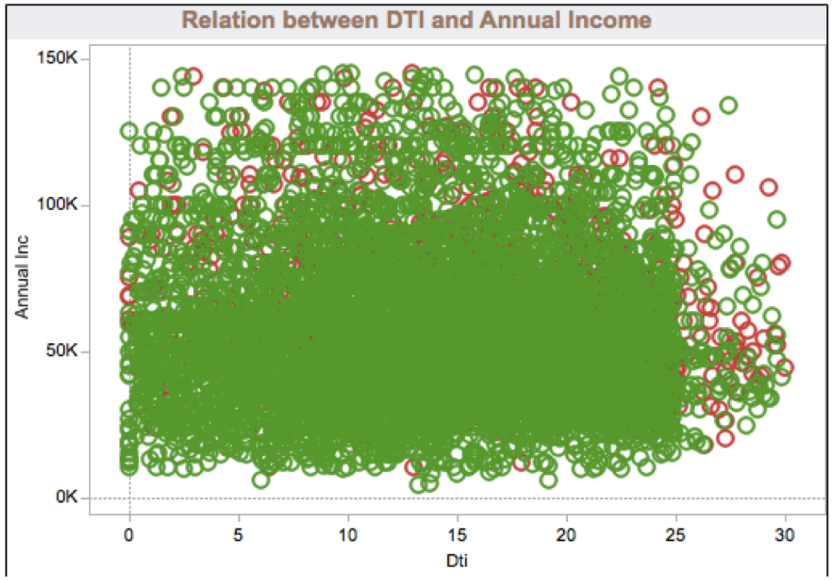
correlation between all variables – to best analyze the correlation

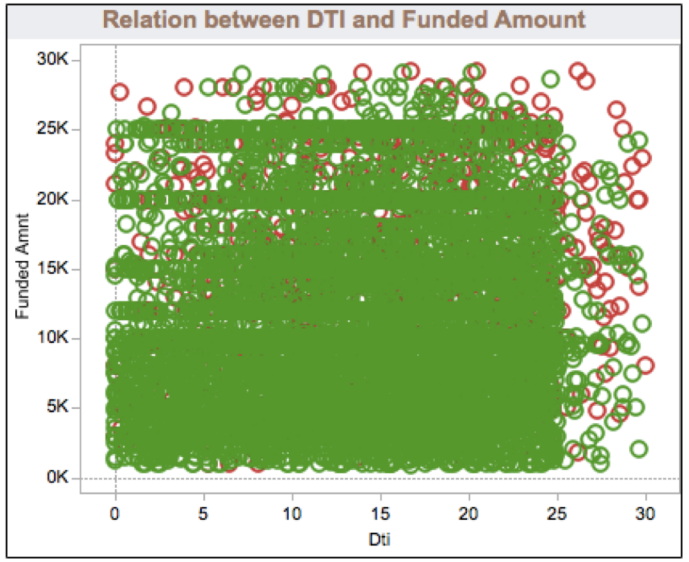


Exploratory Data Analysis - Multivariate Analysis









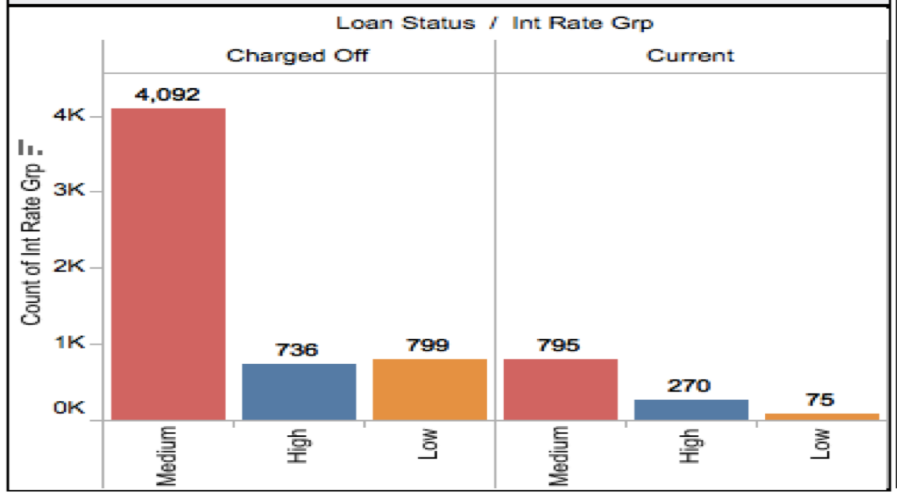
correlation analysis between the key variables

- ❑ **Loan Amount - Funded Amount**
- ❑ **Annual Income – Funded Amount**
- ❑ **Annual Income – DTI**
- ❑ **Funded Amount – DTI**

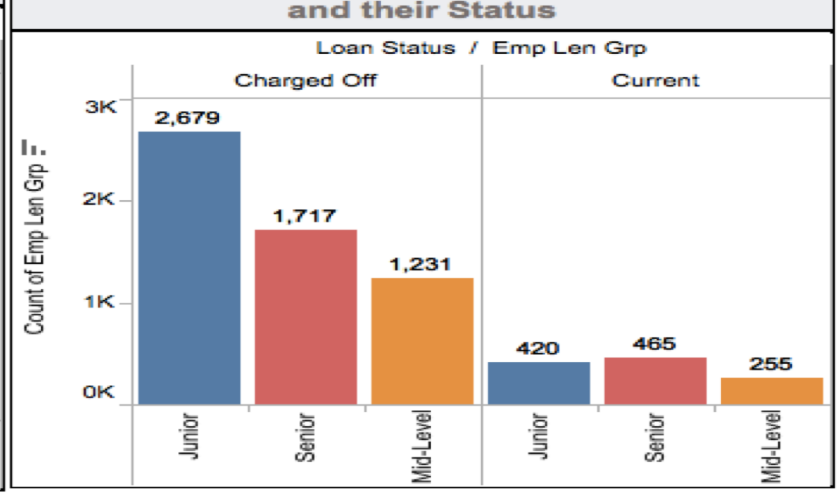
Exploratory Data Analysis - Multivariate Analysis

Multivariate Analysis

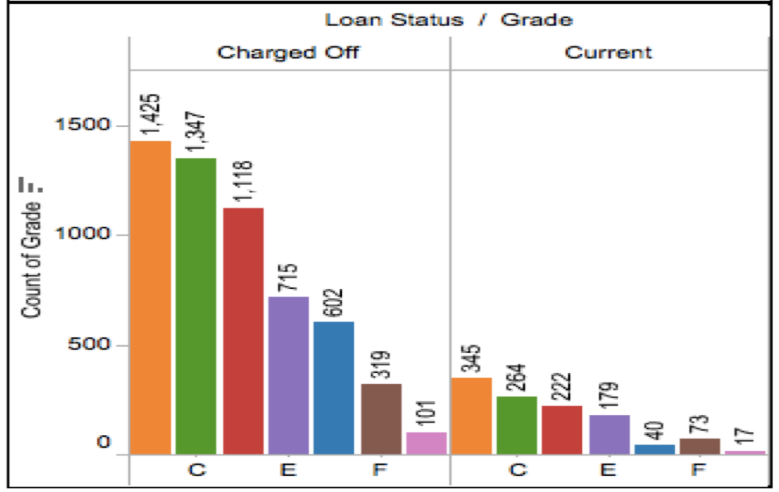
Interest Rate of the Loan and their Status



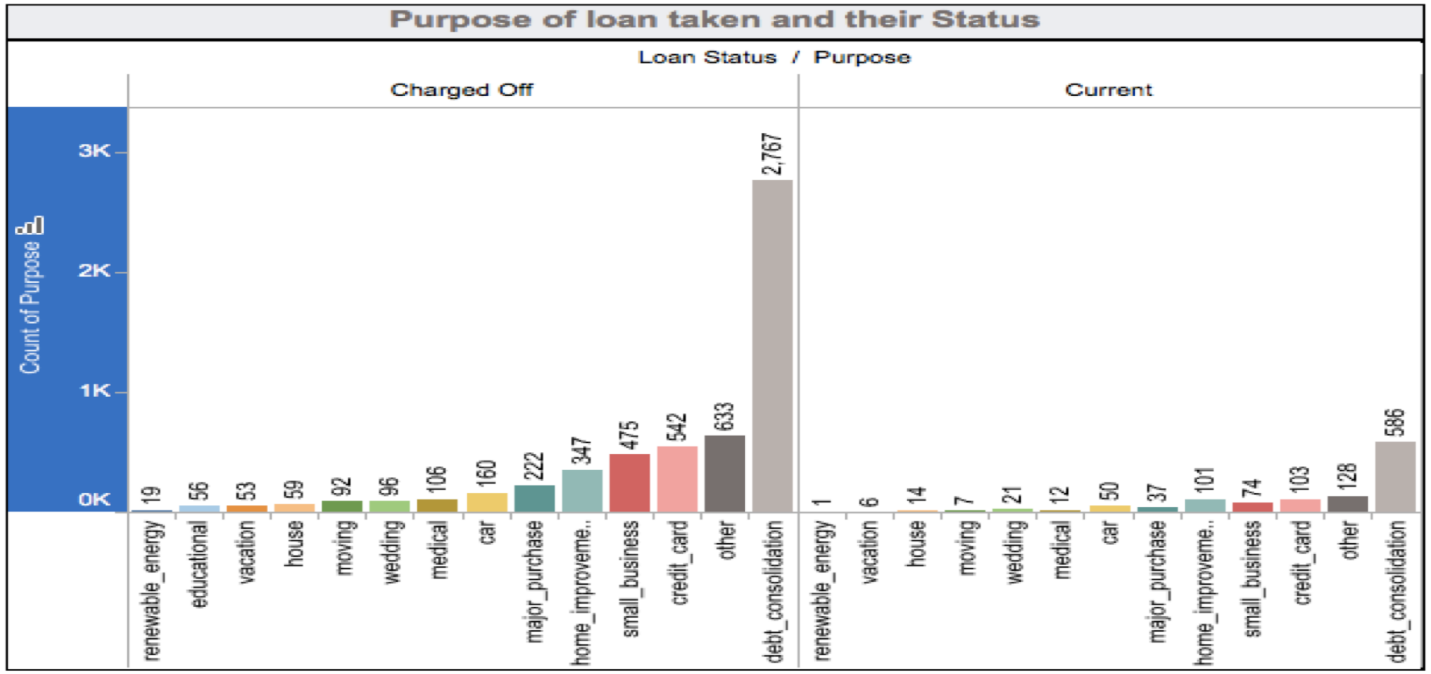
No of loan taken by each employment group and their Status



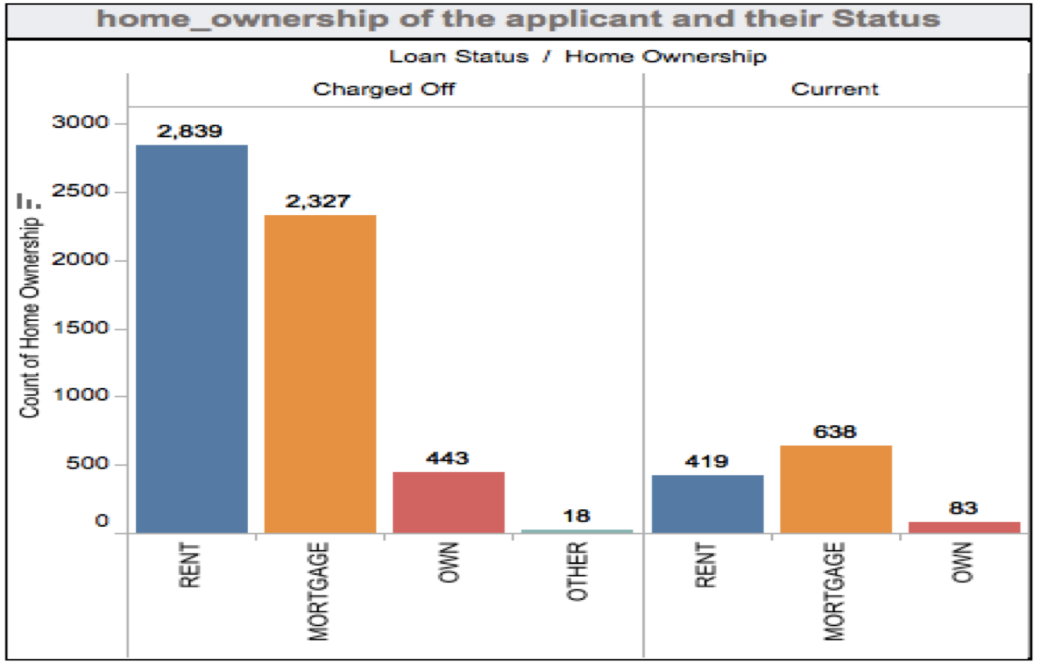
Interest Rate of the Loan and their Status



Purpose of loan taken and their Status



home_ownership of the applicant and their Status

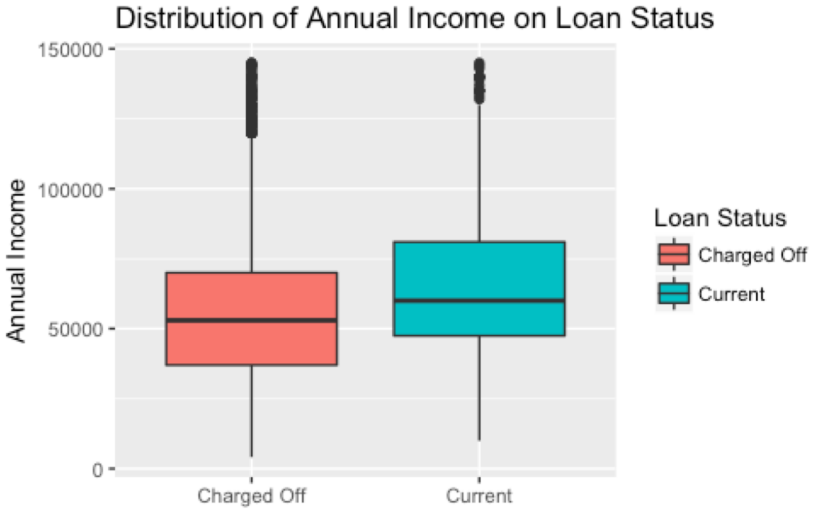


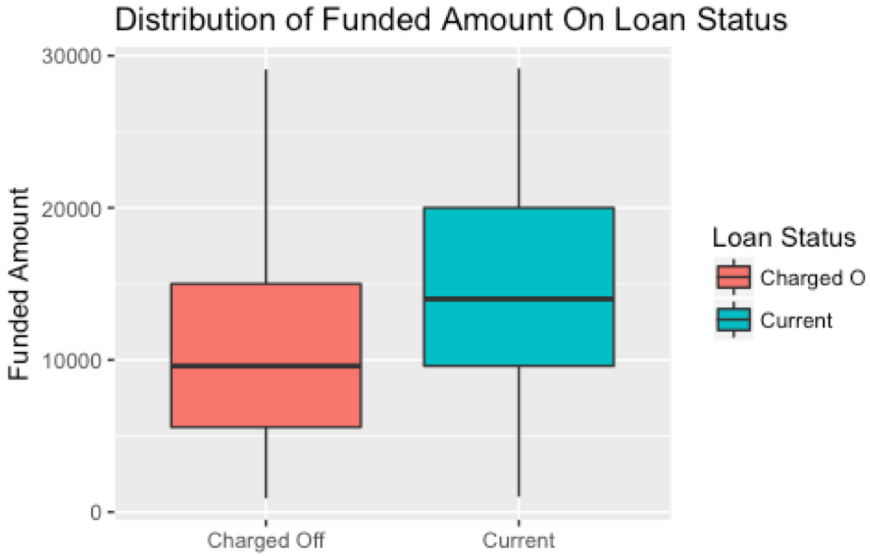
Categorical variable analysis

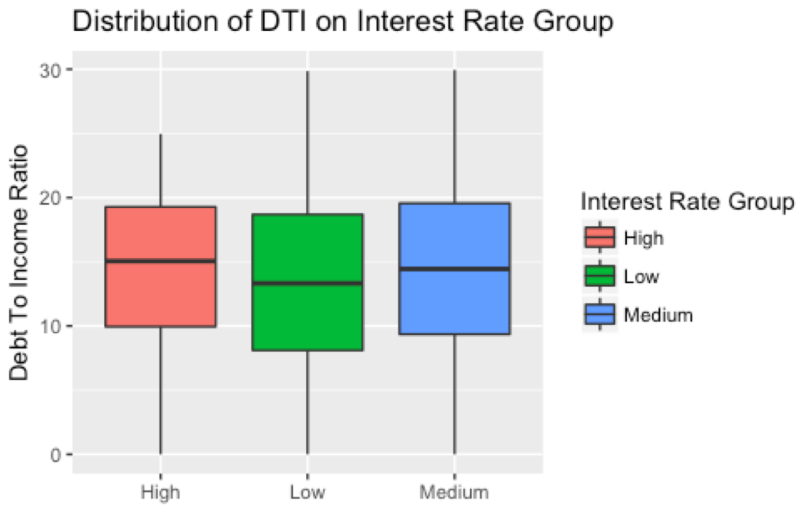
- Junior Employment Group 1 to 4 years tend to default more than rest
- High no of default are for debt_consolidation
- Applicant with home ownership of Mortgage and Rent tends to default more
- If the interest rate is medium people tends to default more
- Grades like B,C,D have trends to default more

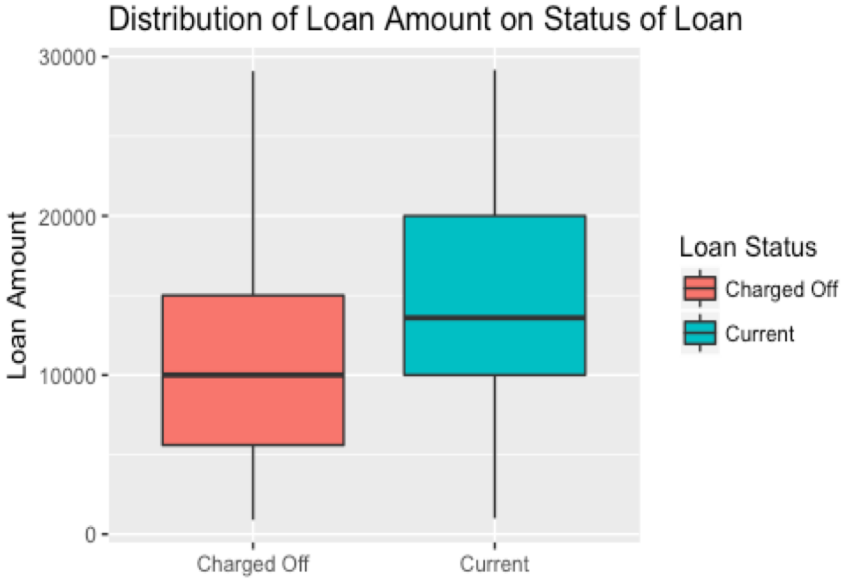
Exploratory Data Analysis - Multivariate Analysis

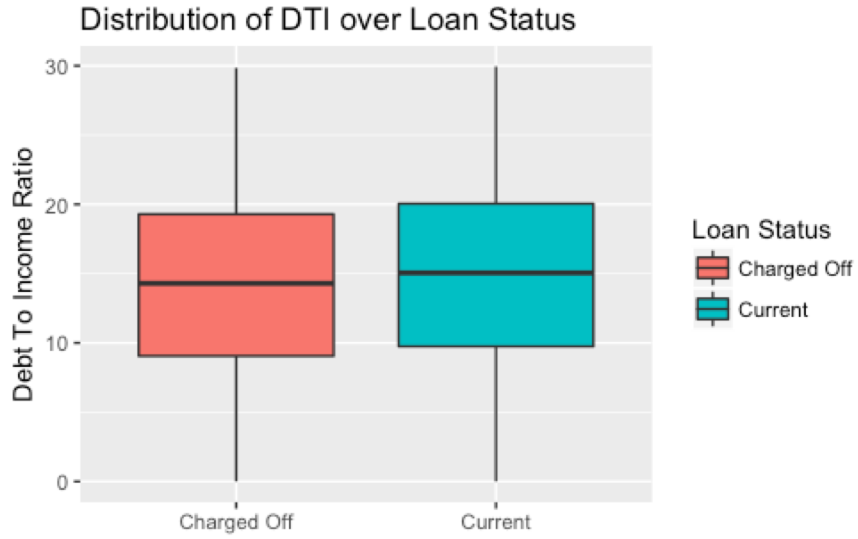
Variable X Differ on Variable Y : Analysis between key variables











Hypothesis Test

A1 : annual_inc : Default and A2 : annual_inc : Current

#Null Hypothesis : H0 : A1 = A2
#Alternative Hypothesis : H1 : A1 <> A2

```
data:  A1 and A2
t = 10.454, df = 1645.3, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7222.153 10557.989
sample estimates:
mean of x mean of y
 65622.20  56732.13
```

#Since the p-value is less than alpha 2.2e-16 or is less than 0.0001 we can reject the null Hypothesis and say that
#there is significant difference between means of A1 and A2

L1 : Loan Amount Default and L2 : Loan Amount Current

#Null Hypothesis : H0 : L1 = L2
#Alternative Hypothesis : H1 : L1 <> L2

```
data:  L1 and L2
t = 15.453, df = 1660.2, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2854.785 3684.849
sample estimates:
mean of x mean of y
 14330.33  11060.51
```

#Since the p-value is less than alpha 2.2e-16 or is less than 0.0001 we can reject the null Hypothesis and say that
#there is significant difference between means of L1 and L2

F1 : Funded Amount Default and F2Funded Amount Current

#Null Hypothesis : H0 : F1 = F2
#Alternative Hypothesis : H1 : F1 <> F2

```
data:  F1 and F2
t = 16.61, df = 1643.3, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3069.266 3891.218
sample estimates:
mean of x mean of y
 14354.82  10874.58
```

#Since the p-value is less than alpha 2.2e-16 or is less than 0.0001 we can reject the null Hypothesis and say that
#there is significant difference between means of F1 and F2

D1 : DTI Default and D2 : DTI Current

#Null Hypothesis : H0 : D1 = D2
#Alternative Hypothesis : H1 : D1 <> D2

```
data:  D1 and D2
t = 3.449, df = 1614.4, p-value = 0.0005771
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3232165 1.1755535
sample estimates:
mean of x mean of y
 14.75001  14.00062
```

#Since the p-value is less than alpha 0.025 or P value = 0.0005771 we can reject the null Hypothesis and say that
#there is significant difference between means of D1 and D2 with 95% Confidence

Hypothesis Test

A3 : Annual_Income_High_Interest_Rate and A4 :Annual_Income_Low_Interest

#Null Hypothesis : H0 : A3 = A4
#Alternative Hypothesis : H1 : A3 <> A4

data: A3 and A4
t = 11.833, df = 1866.3, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
11668.85 16305.57
sample estimates:
mean of x mean of y
66158.78 52171.57

#Since the p-value is less than alpha 2.2e-16 we can reject the null Hypothesis and say that
#there is significant difference between means of A3 and A4 with 95% Confidence

L3 : Loan_Amount_High_Interest_Rate and L4 : Loan_Amount_Low_Interest

#Null Hypothesis : H0 : L3 = L4
#Alternative Hypothesis : H1 : L3 <> L4

data: L3 and L4
t = 23.011, df = 1840.6, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
5786.945 6865.324
sample estimates:
mean of x mean of y
14685.860 8359.725

#Since the p-value is less than alpha 2.2e-16 we can reject the null Hypothesis and say that
#there is significant difference between means of L3 and L4 with 95% Confidence

F3 : Funded_Amount_High_Interest_Rate and F4 : Funded_Amount_Low_Interest

#Null Hypothesis : H0 : F3 = F4
#Alternative Hypothesis : H1 : F3 <> F4

data: F3 and F4
t = 24.679, df = 1822, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
6043.158 7086.602
sample estimates:
mean of x mean of y
14693.713 8128.833

#Since the p-value is less than alpha 2.2e-16 we can reject the null Hypothesis and say that
#there is significant difference between means of F3 and F4 with 95% Confidence

D3 : DTI_High_Interest_Rate and D4 :DTI_Low_Interest

#Null Hypothesis : H0 : D3 = D4
#Alternative Hypothesis : H1 : D3 <> D4

data: D3 and D4
t = 3.2121, df = 1741.4, p-value = 0.001342
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.3844607 1.5902425
sample estimates:
mean of x mean of y
14.36935 13.38200

#Since the p-value 0.001342 is greater than alpha 0.025 we cannot reject the null Hypothesis and say that
#there is no significant difference between means of D3 and D4

Conclusion

1. Since we have seen significant difference between the Annual Income, Funded Amount etc., we can conclude that it we concentrate our decision to give loan to only those income group which do not fall under default we can escape from crating type 1 error.

2. Since we found that Junior account for 48% of the default, we can decide which employment group should be given loan.

3. We have also found that 100% of the loan given for education, followed by renewable energy 95%, moving 93 are default, So we can refrain ourselves from giving loan for these purposes.

Thank You!

Raju Kumar

Sidharth Mahapatra

Swarna Manchikalapudi