Name: MD Abdul Kader
Roll: 180030

Project Activity -2

Section-A

GitHub Link:
https://github.com/AbdulKader05/Big-Data-Project-Activity-2.git
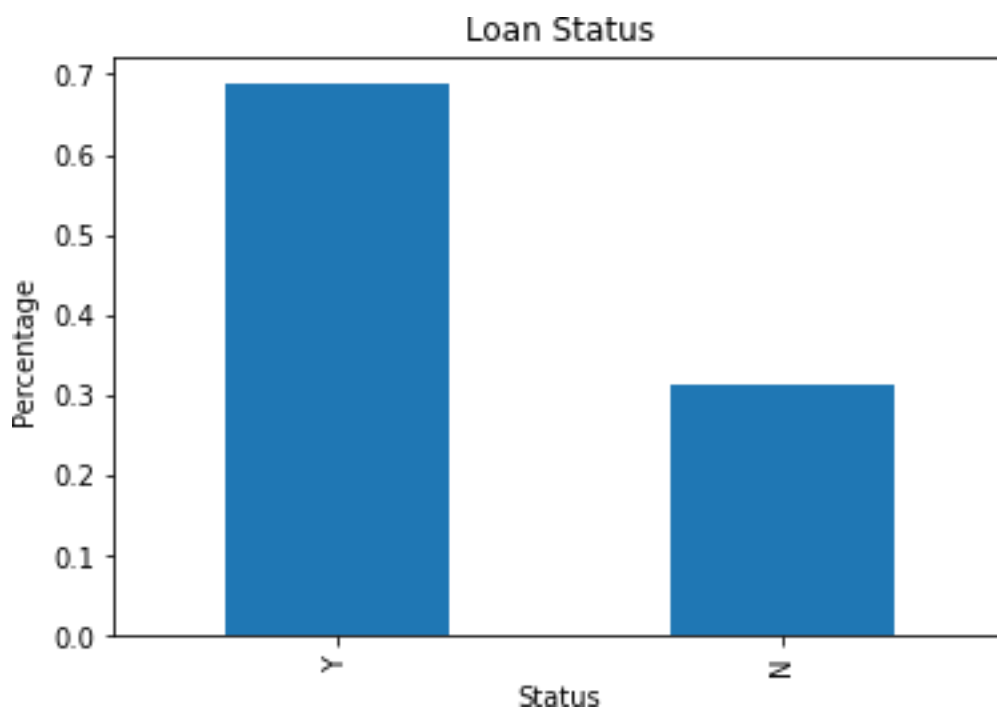
Shape train dataset (614, 13)
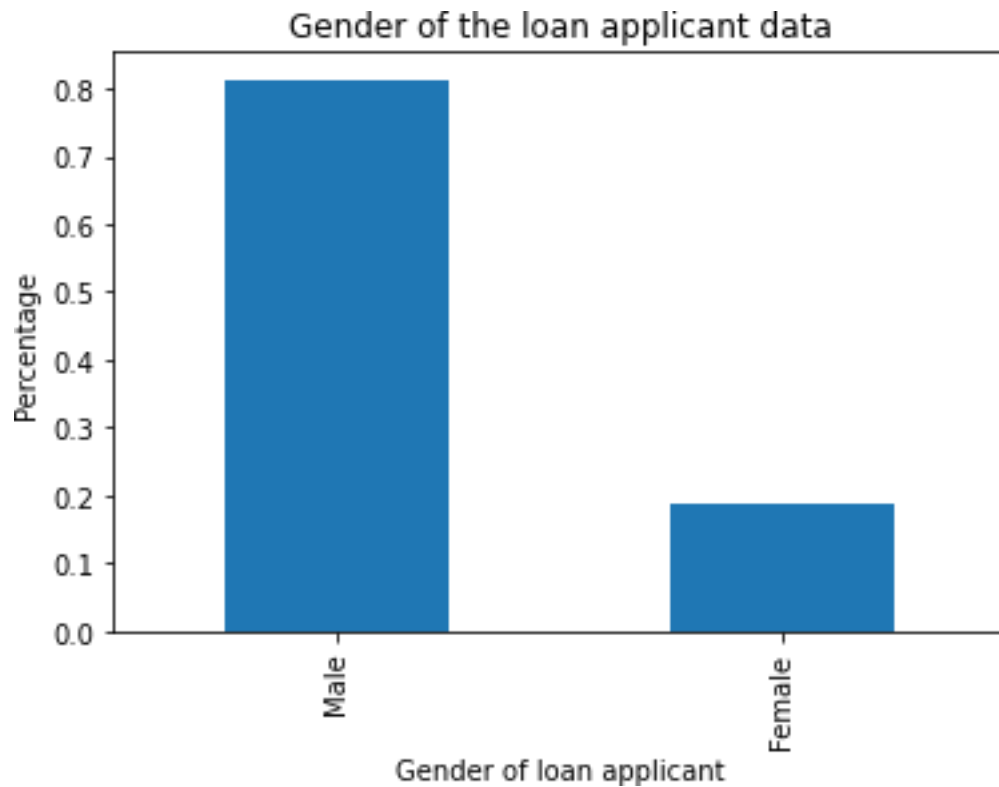
Shape test dataset (367, 12)

loan of 422(approx 70%) people out of 614 has been approved.

Categorical features: These features have categories (Gender, Married, Self

_Employed, Credit_History, Loan_Status)

QUESTION 1-a: Find out the number of male and female in loan applicants dat a.

Name: MD Abdul Kader
Roll: 180030

Gender of the loan applicant data

Answer of   QUESTION 1-a:

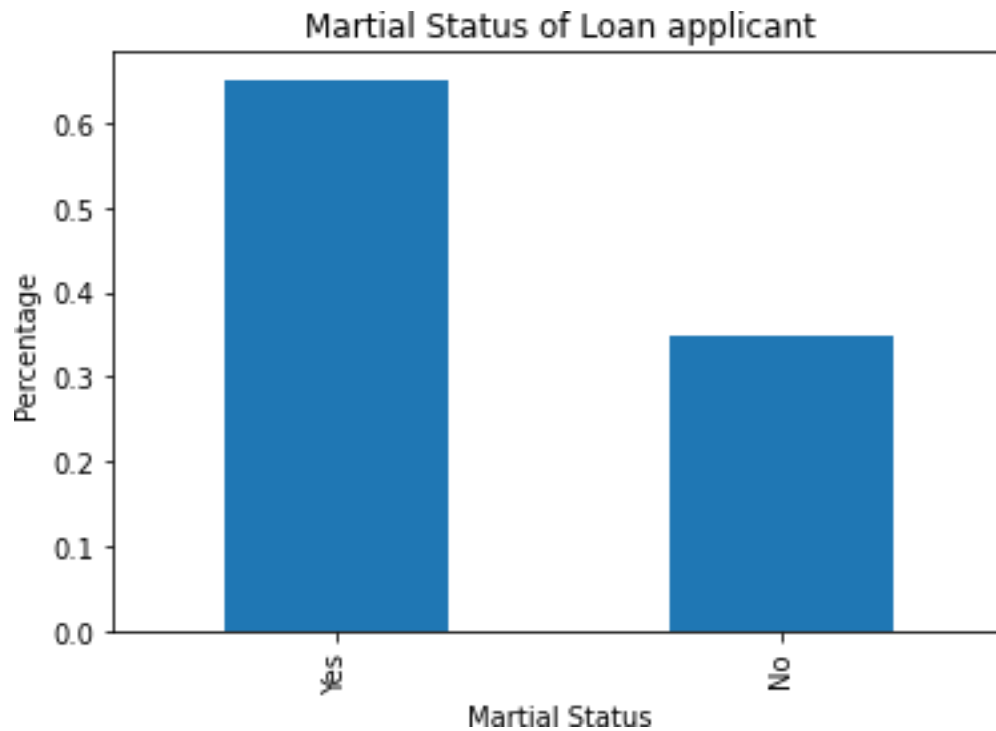There are 81% of Male and 19% of Female in loan applications Total number of people:

611

Married: 398

Unmarried: 213

QUESTION 1-b: Find out the number of married and unmarried loan applicants.

Martial Status of Loan applicant
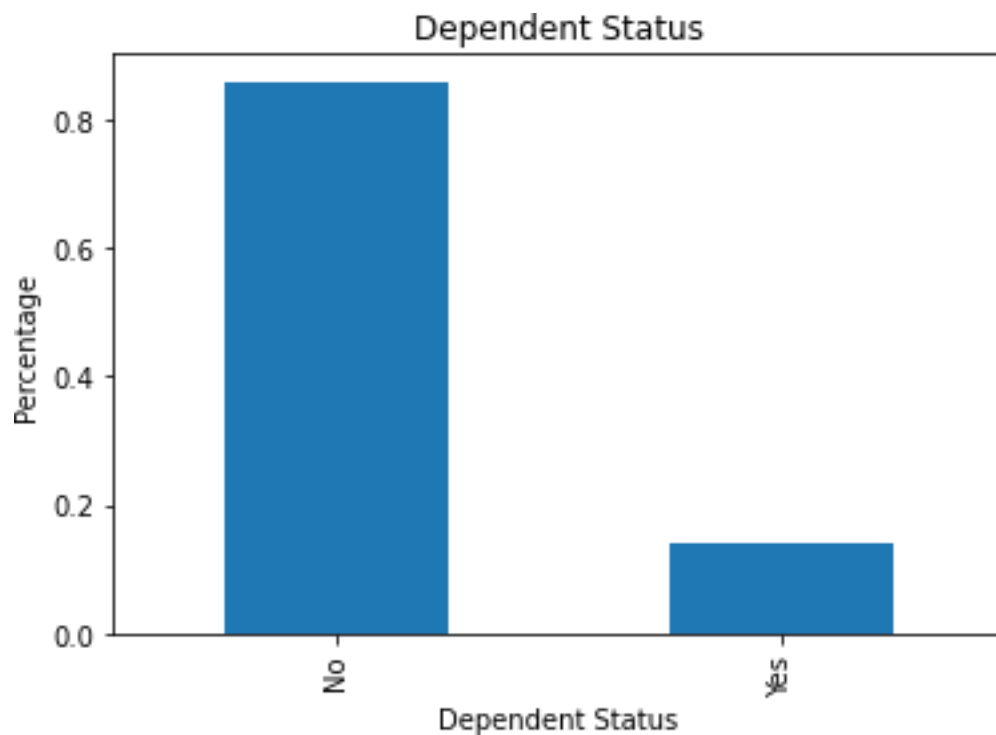

Answer   of QUESTION 1-b:

Number of married people : 65% Number of

unmarried people : 35%

QUESTION 1-c: Find out the overall dependent status in the dataset.
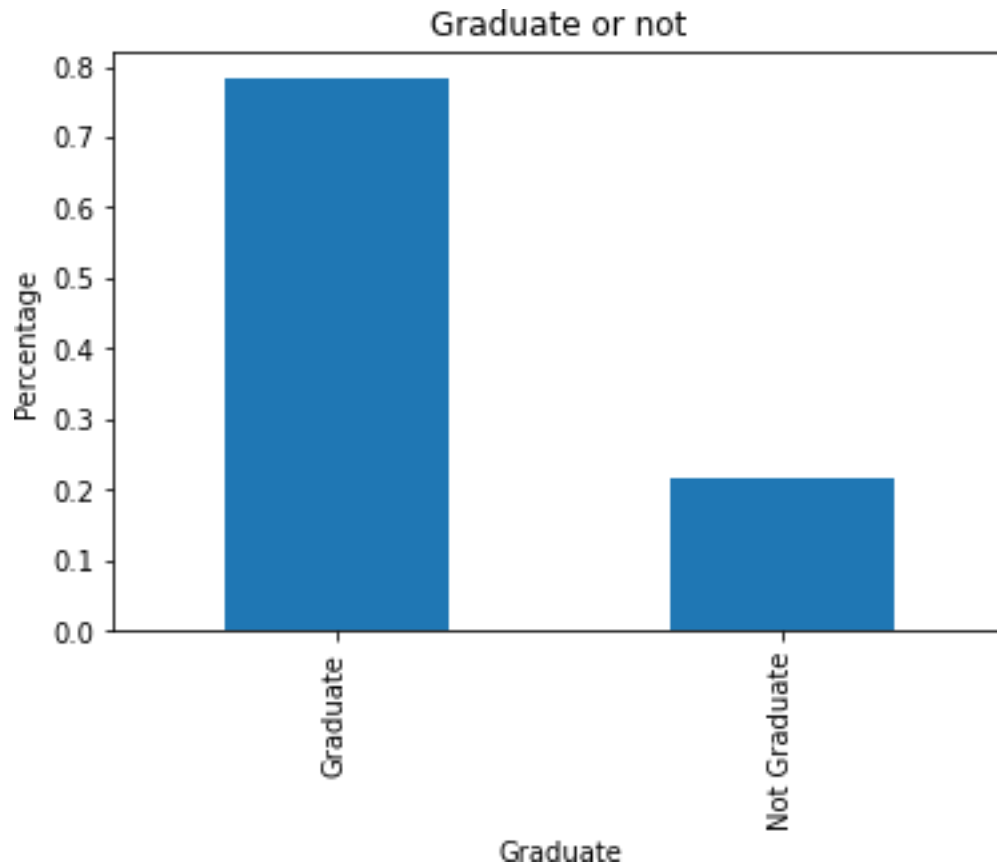
Dependent Status

Answer for QUESTION 1-c:

In a total of 582 people - 14% are SelfEmployed and - 86% are Not SelfEmplo yed

QUESTION 1-d: Find the count how many loan applicants are graduate and non graduate.
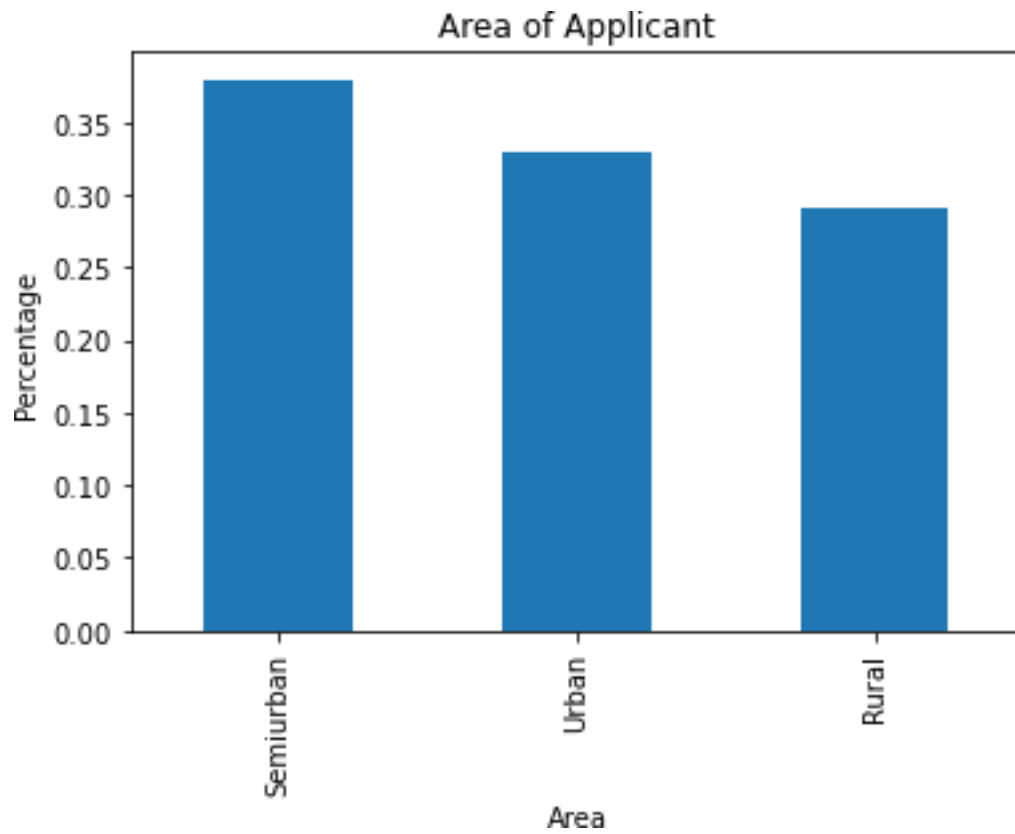

Answer for QUESTION 1-d:

Total number of People : 614 78% are

Graduated

22% are not Graduated

QUESTION 1-e: Find out the count how many loan applicants property lies in urban, rural and semi-urban areas.

Area of Applicant
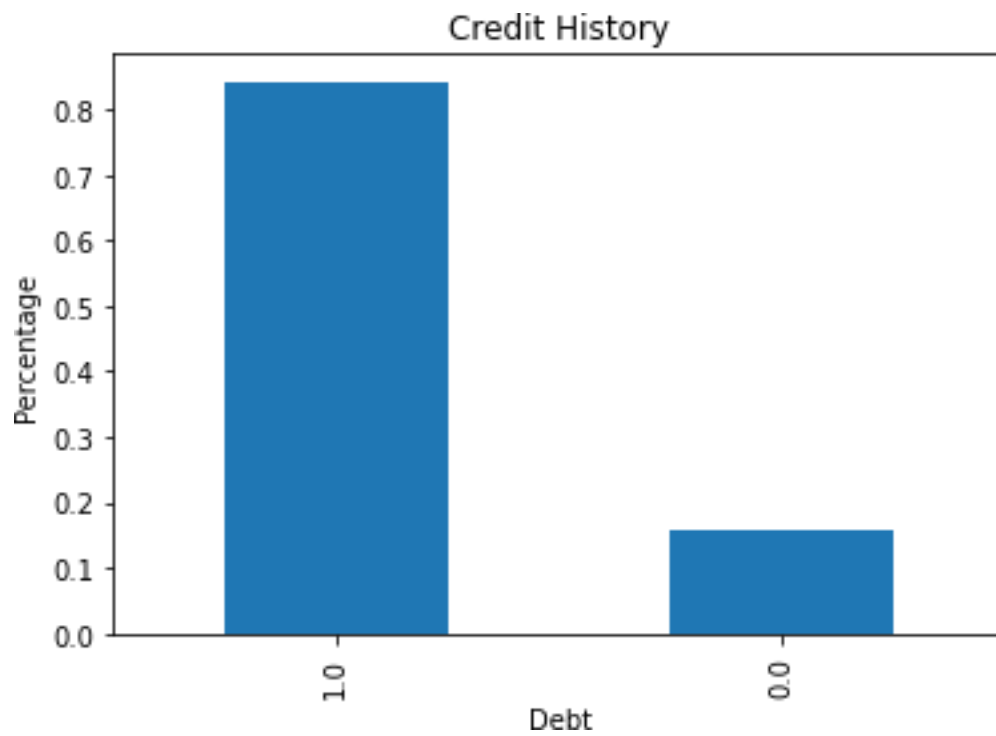
Answer of Question 1-E:

Applicants from Semiurban area = 38% Applicants
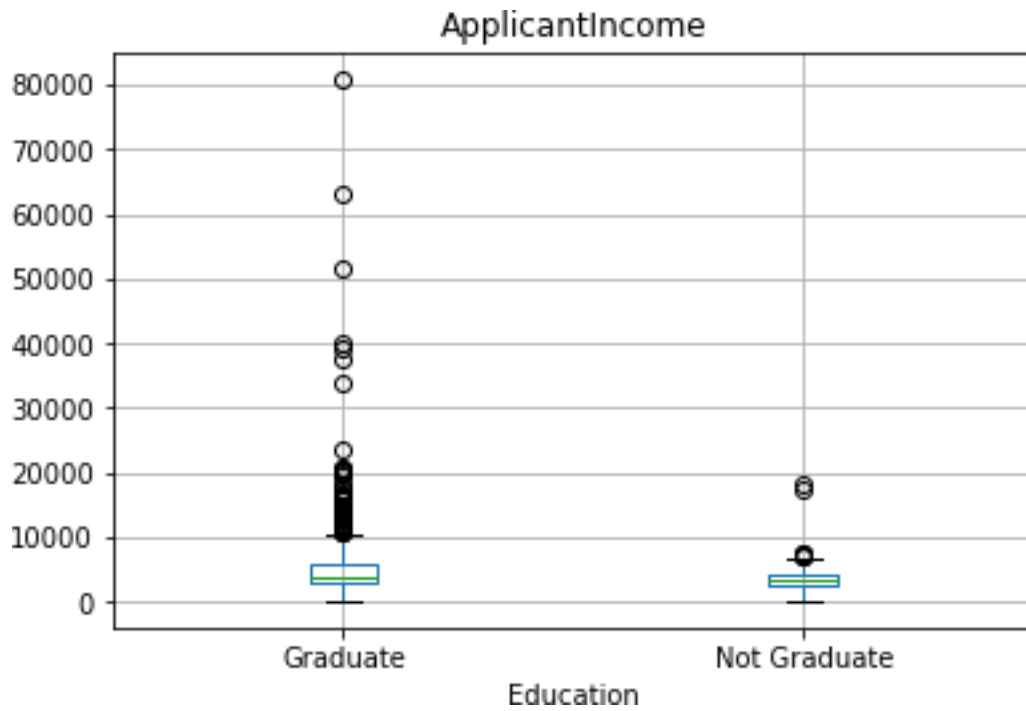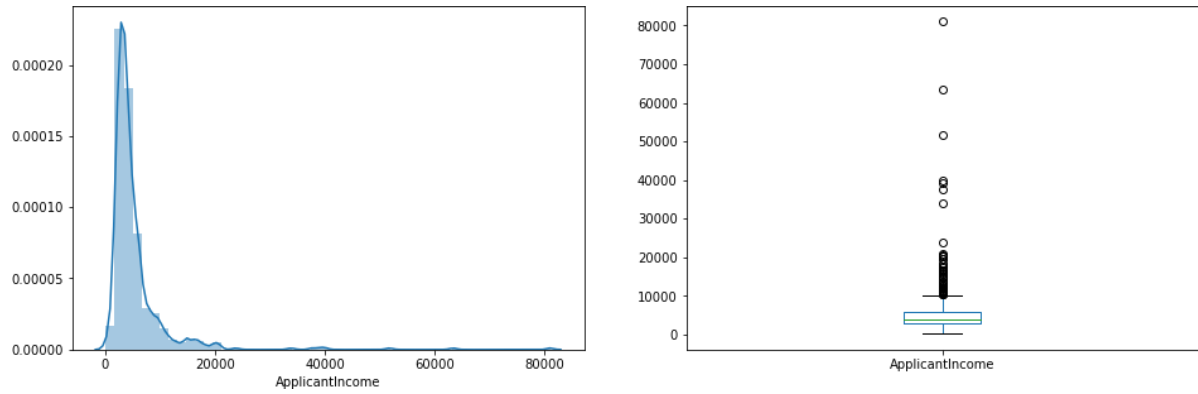from Urban area = 33% Applicants from Rural area =
29%

Credit History

QUESTION 3:

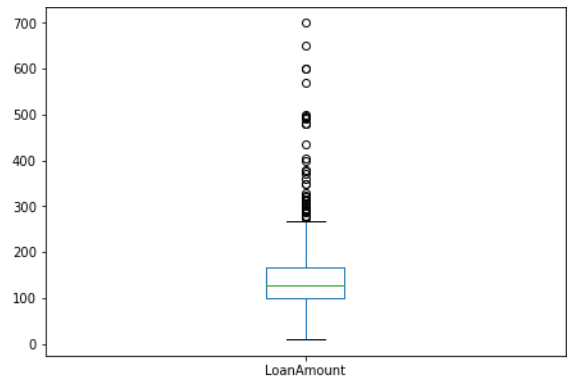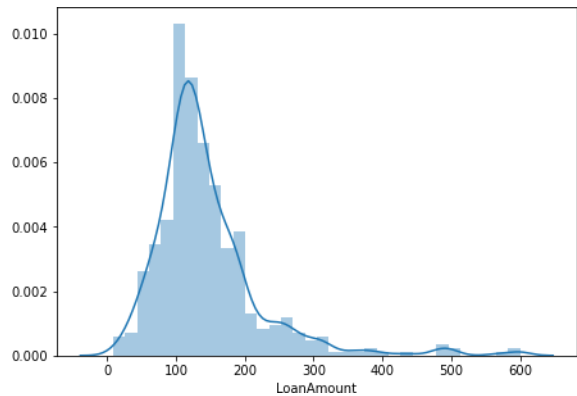To visualize and plot the distribution plot of all numerical attributes of the given train dataset i.e.

ApplicantIncome, CoApplicantIncome and LoanAmo unt.

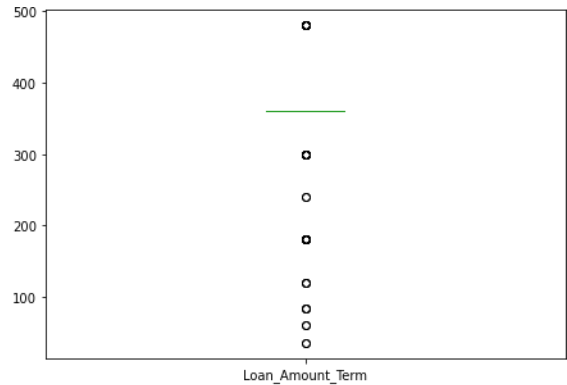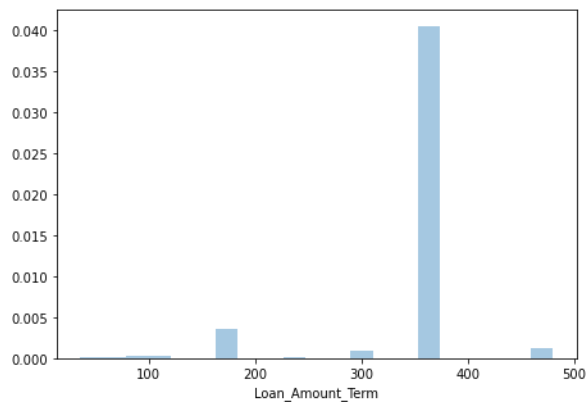Applicant Income distribution

Loan Amount distribution

Loan Amount   Distribution

Question 4

Relation between Loan_Status and Gender Loan_Status

	N	Y
Gender		
Female	37	75
Male	150	339
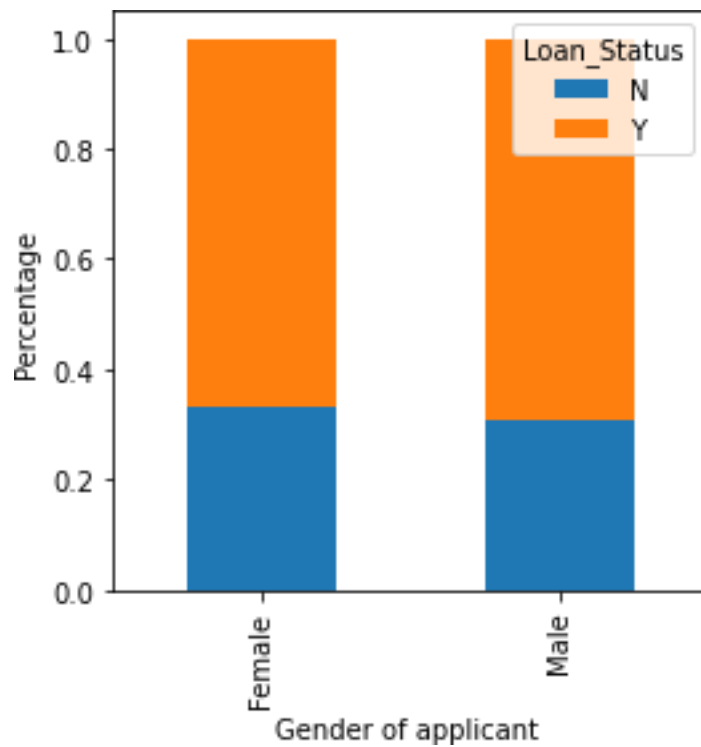
Conclusion from Relation between Loan_Status and Gender Female whose Loan

was approved = 75

Male whose Loan was approved = 339 Female whose Loan

was not0 approved = 37 Female whose Loan was approved

= 150

Ee can observe that the proportion of Male applicants is higher for the app roved loans.

Relation between Loan Status and martial status Loan_Status

	N	Y
Married		
No	79	134
Yes	113	285
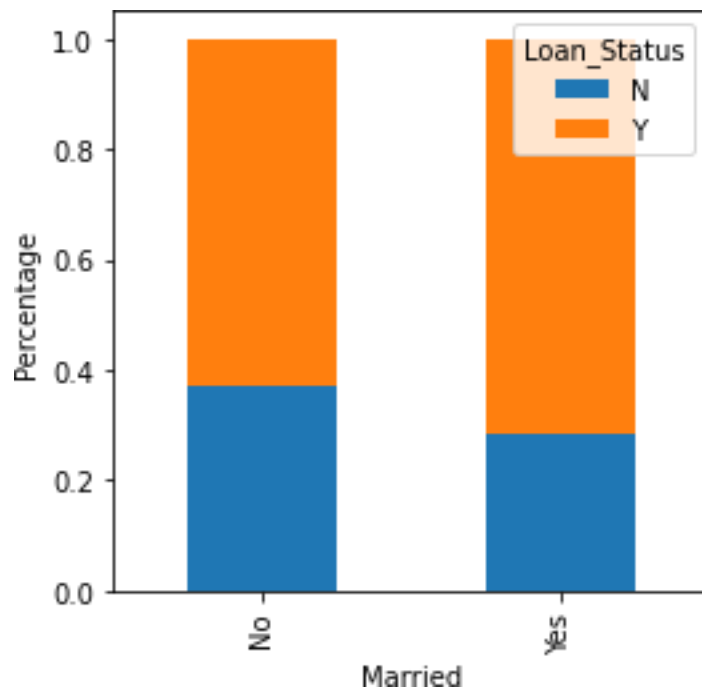
Conclusion of relation between Loan_Status and Married status Married people whose

Loan was approved = 285

Married people whose Loan was not approved = 113 Unmarried people

whose Loan was approed = 134 Unmarried people whose Loan was not

approed = 79

We can observe that the proportion of Married applicants is higher for the approved loans.

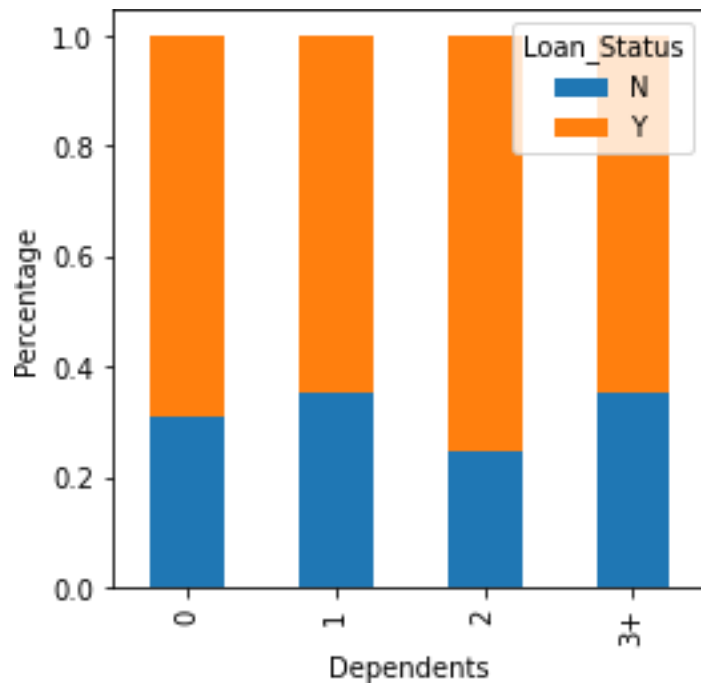Relation between LoanStatus and Dependent Loan_Status

	N	Y
Dependents		
0	107	238
1	36	66
2	25	76
3+	18	33

Name: MD Abdul Kader
Roll: 180030

Conclusion of relation between Loan_Status and Dependents Number of dependents

on the loan applicant

0 and Loan was approed : 238

0   and Loan was not approed : 107

1   and Loan was approed : 66

1   and Loan was not approed : 36

2   and Loan was approed : 76

2 and Loan was not approed : 25 3+ and Loan
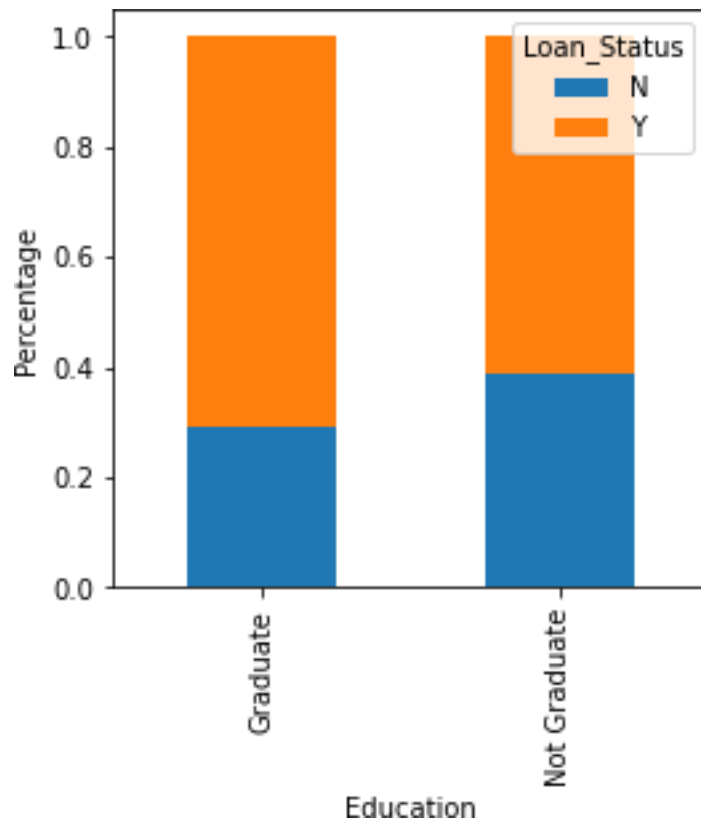
was approed : 33

3+ and Loan was not approed : 18

We can observe that the distribution of applicants with 1 or 3+ dependents is similar across both the categories

of Loan_Status.

Relation between Loan Status and Education

Loan_Status	N	Y
Education		
Graduate	140	340
Not Graduate	52	82

Name: MD Abdul Kader
Roll: 180030

Conclusion of relation between Loan_Status and Education. People who are

Graduate and Loan was approed : 340

People who are Graduate and Loan was no approed : 140 people who are Not

Graduate and Loan was approed : 82 People who are Not Graduate and Loan was not
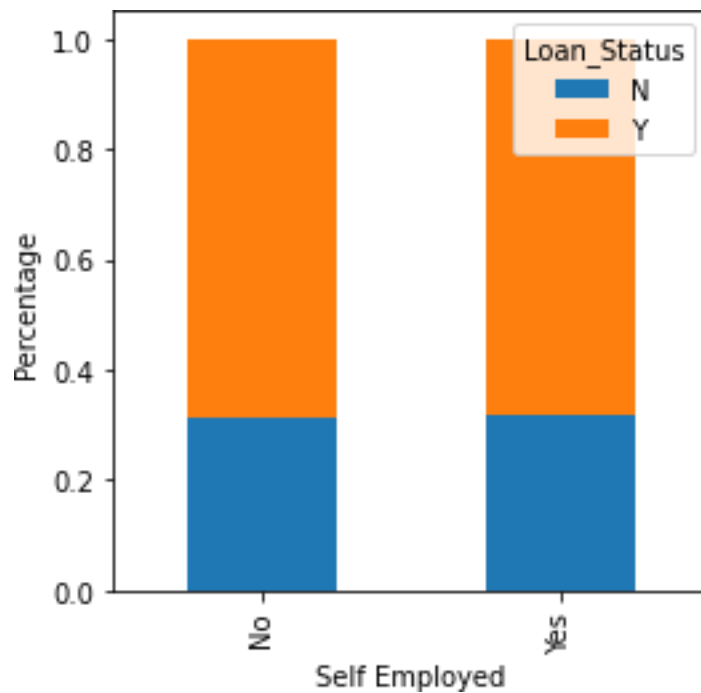
approed : 52

We can observe that the proportion of Graduate applicants is higher for the approved loans.

Relation between Loan Status and Self Employed

Loan_Status	N	Y
Self_Employed		
No	157	343
Yes	26	56

Conclusion from Relation between Loan_Status and Self_Employed People who are

Self_Employed and Loan was approed : 56

People who are Self_Employed and Loan was not approed : 26 People who are not

Self_Employed and Loan was approed : 343 People who are not Self_Employed and Loan

was not approed : 157

There is nothing thatnwe can signify and   infer from Self_Employed vs Loan_ Status plot.

Relation between Loan_Status and Credit_History Loan_Status

	N	Y
Credit_History		
0.0	82	7
1.0	97	378

Conclusion from relation between Loan Status and Credit History

People   with   credit   history   as 1 and loan was approved : 378

People   with   credit   history   as 1 and loan was not approved :   97

People   with   credit   history   as 0 and loan was approved : 7

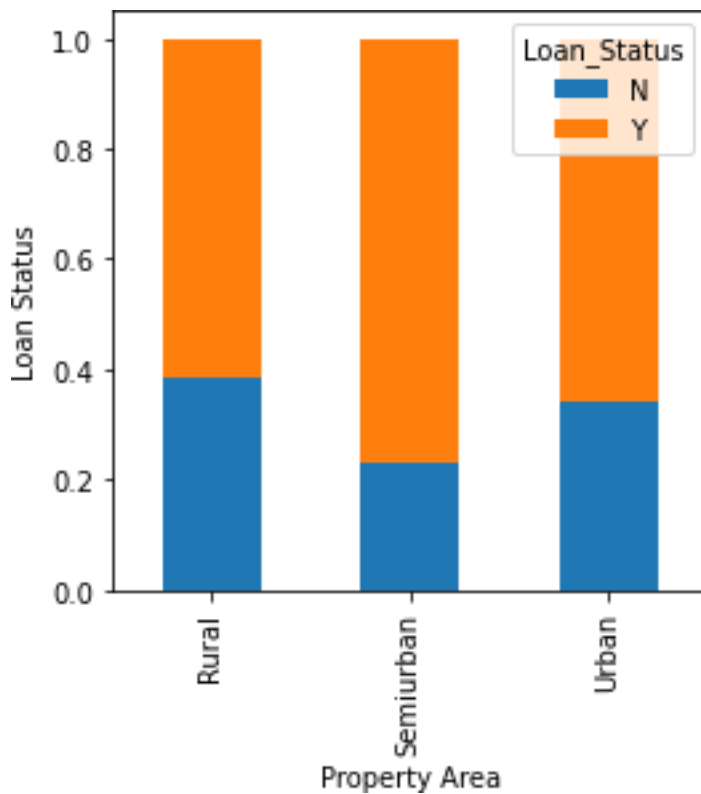People   with   credit   history   as 0 and loan was not approved :   82

We can observe that, it seems people with credit history as 1 are more like ly to get their loans approved.

Relation between Loan Status and Property Area

Loan_Status	N	Y
Property_Area		
Rural	69	110
Semiurban	54	179
Urban	69	133

conclusion from Relation between Loan Status and Property Area People who are from Rural area and loan was approved : 110 People who are from Rural area and loan was not approved : 69 People who are from Semiurban area and loan was approved : 179 People who are from Semiurban area and loan was not approved : 54 People who are from Urban area and loan was approved : 133

People who are from Semiurban area and loan was not approved : 69
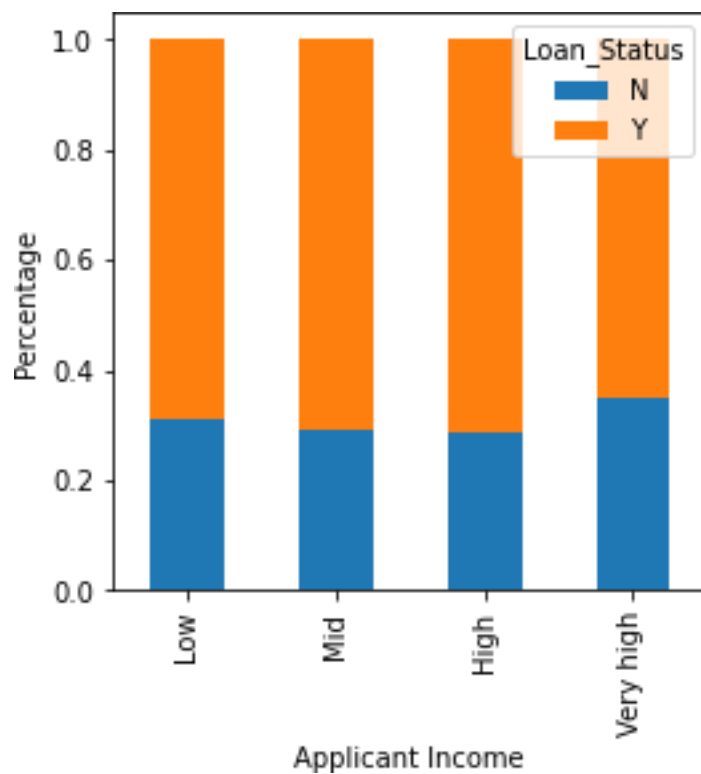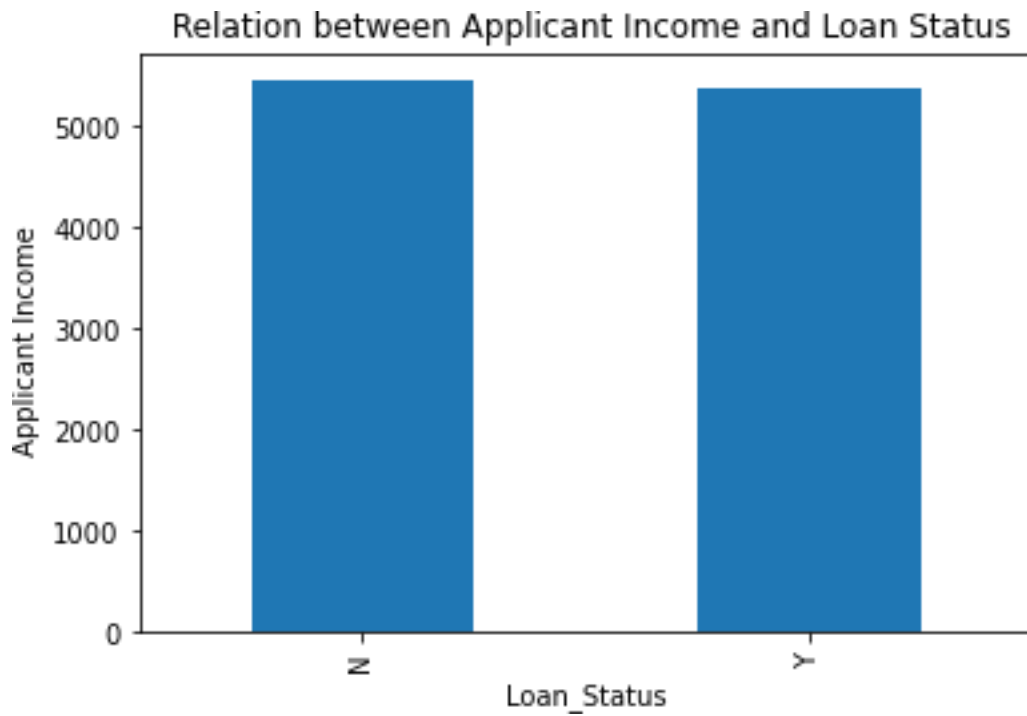
We can observe that theProportion of loans getting approved in semiurban ar ea is higher as compared to that in rural or urban areas

Relation between Loan_Status and Income Loan_Status

	N	Y
Income_bin		
Low	26	57
Mid	51	123
High	32	79
Very high	39	73

Name: MD Abdul Kader
Roll: 180030
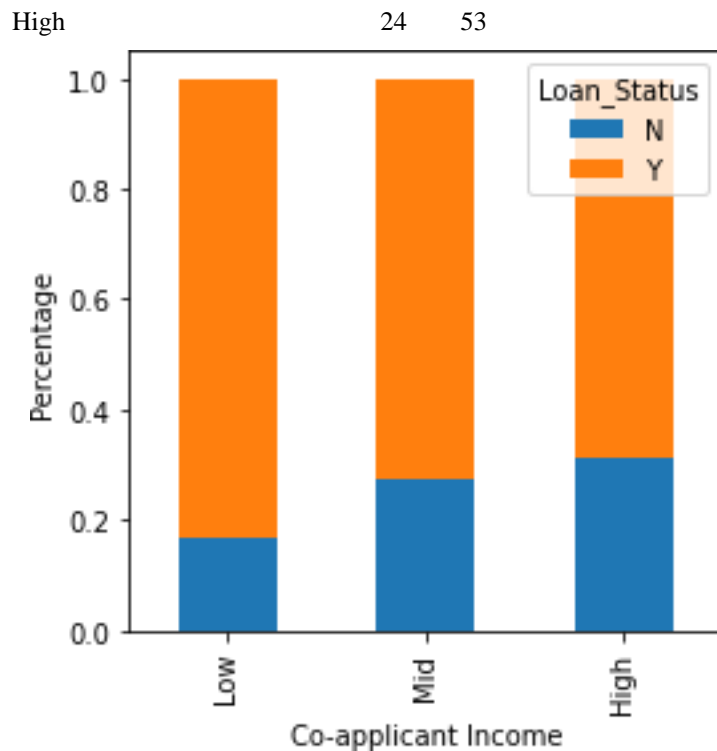
Relation between Applicant Income and Loan Status

We can observe that applicant income does not affect the chances of loan ap proval

Loan_Status	N	Y
CoapplicantIncome_bin		
Low	3	15
Mid	46	123

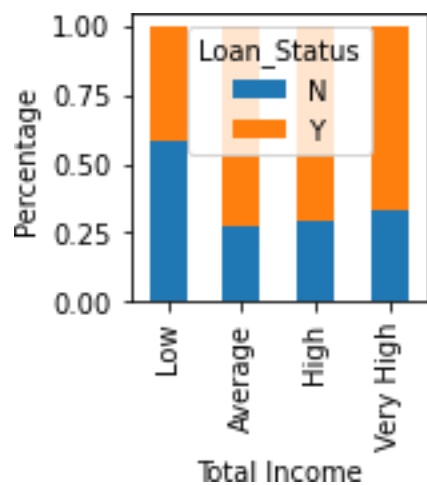High   24   53

Co-applicant Income

We can observe that if coapplicant's income is less the chances of loan app roval are high. This might be because of if there is no coapplicant then hi s income is marked as ZERO. So I think so we cannot infer any conclusion fr om here.

Loan_Status	N	Y
TotalIncome_bin		
Low	14	10
Average	32	87
High	65	159
Very High	81	166

Total Income

We can observe that low income people have less change of getting loan appr oved compared to Average, High and Very High Income

Name: MD Abdul Kader
Roll: 180030

Applicants whose TotalIncome was Low and loan was approved : 10 Applicants whose TotalIncome was Low and loan was not approved : 14 Applicants whose TotalIncome was Aerage and loan was apprvoed : 87 Applicants whose TotalIncome was Average and loan was not approved : 32 Applicants whose TotalIncome was High and loan was approved : 159 Applicants whose TotalIncome was High and loan was not approved : 65 Applicants whose TotalIncome was Very High and loan was approved : 166 Applicants whose TotalIncome was Very High and loan was not approed : 81 Relation between Loan Status and Loan Amount

Loan_Status	N	Y
LoanAmount_bin		
Low	38	86
Average	83	207
High	27	39
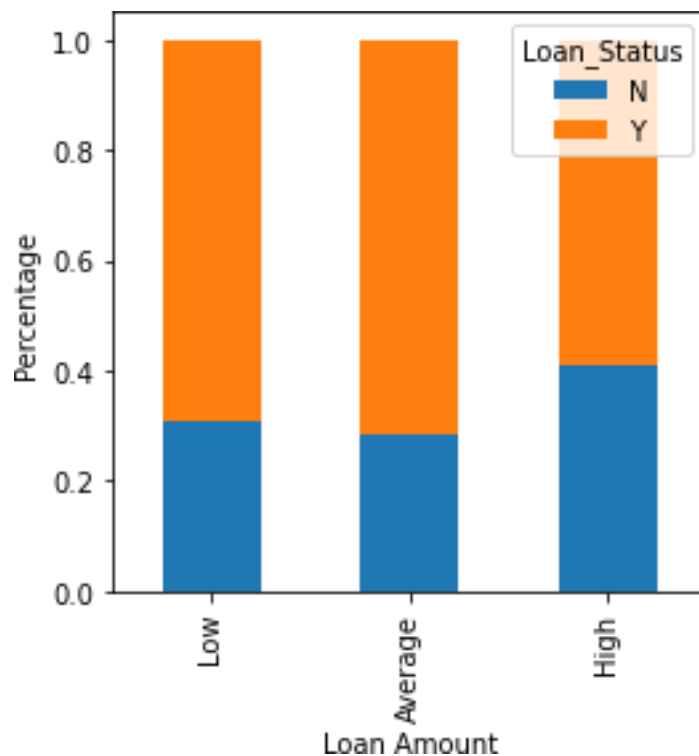
Conclusions from Relation between Loan Status and Loan Amount Applicants whose Loan Amount was low and Loan was approved : 86 Applicants whose Loan Amount was low and Loan was not approved : 38 Applicants whose Loan Amount was Average and Loan was approved : 207 Applicants whose Loan Amount was Average and Loan was not approved : 83 Applicants whose Loan Amount was High and Loan was approved : 39 Applicants whose Loan Amount was High and Loan was not approved : 27
We can observe that the proportion of approved loans is higher for Low and Average Loan Amount as compared to that of High Loan Amount
From the above HEAT MAP the variables with darker color means their correla tion is more.

Name: MD Abdul Kader
Roll: 180030

More correlated variables:
ApplicantIncome - LoanAmount
Credit_History - Loan_Status