



Loan Data Analysis

Here in this Data analysis we will see that how the different factors effects on the loan and which factor is most affected on take a loan.

Here we analyse that applicant is capable for take loan.

Machine Learning Model Building Stages

1

**Data Exploration
and
Visualization**

2

**Data Pre-
procesasing**

3

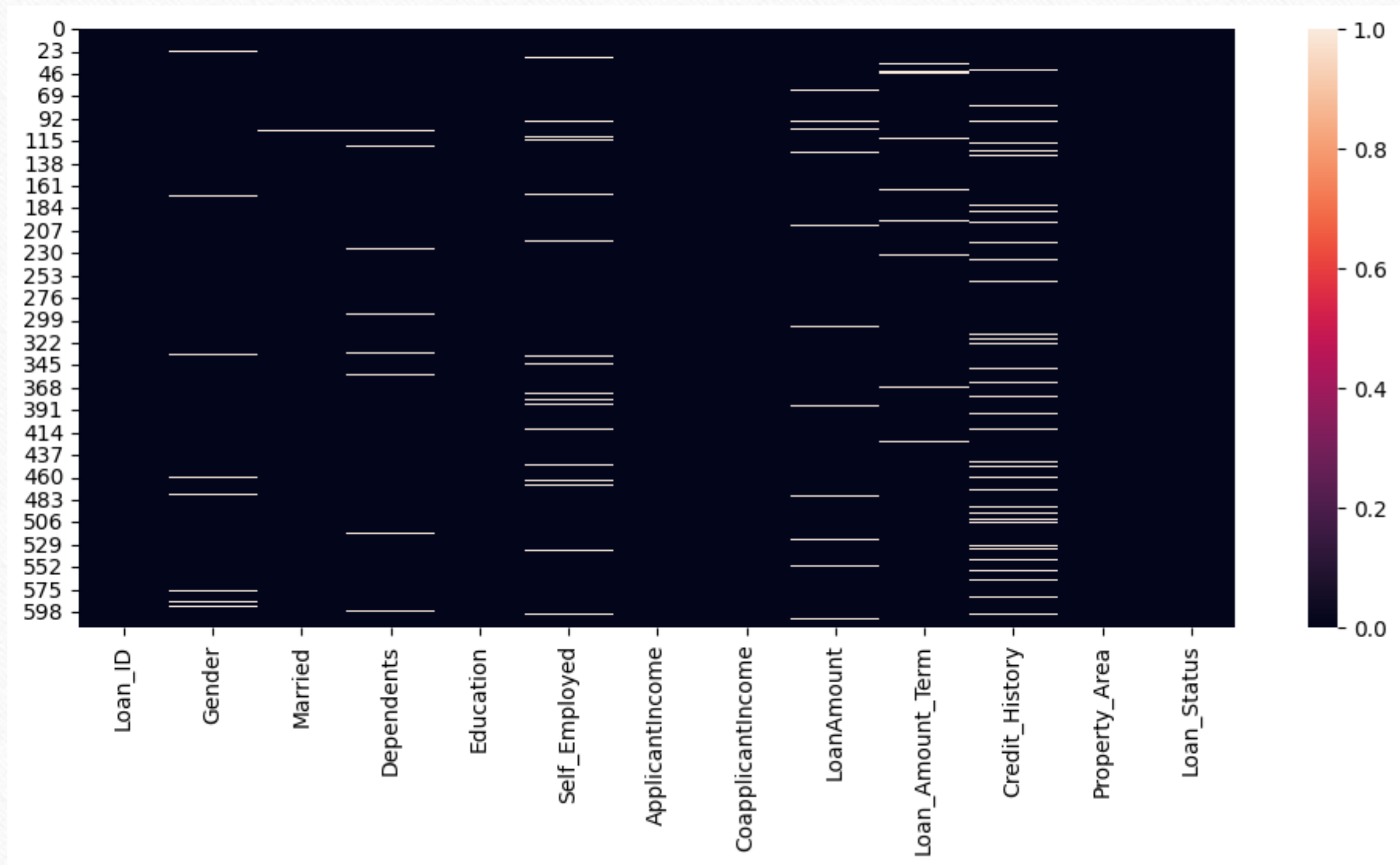
**Model building
and Evaluation**

Overview of Data

| column name | column description | column name | column description |
|---------------|---|-------------------|--|
| Loan ID | Unique Loan ID | ApplicantIncome | Income of applicant |
| Gender | Male / Female | CoapplicantIncome | Co-applicant income |
| Married | Married Status (Yes / No) | LoanAmount | Amount of the Loan |
| Dependents | Number of Dependents | Loan_Amount_Term | Time period of Loan |
| Education | Applicant Education (Graduate / Not Graduate) | Credit_History | Credit History |
| Self_Employed | Self Employed (Yes / No) | Property Area | Property (Urban / Rural / Semiurban) |

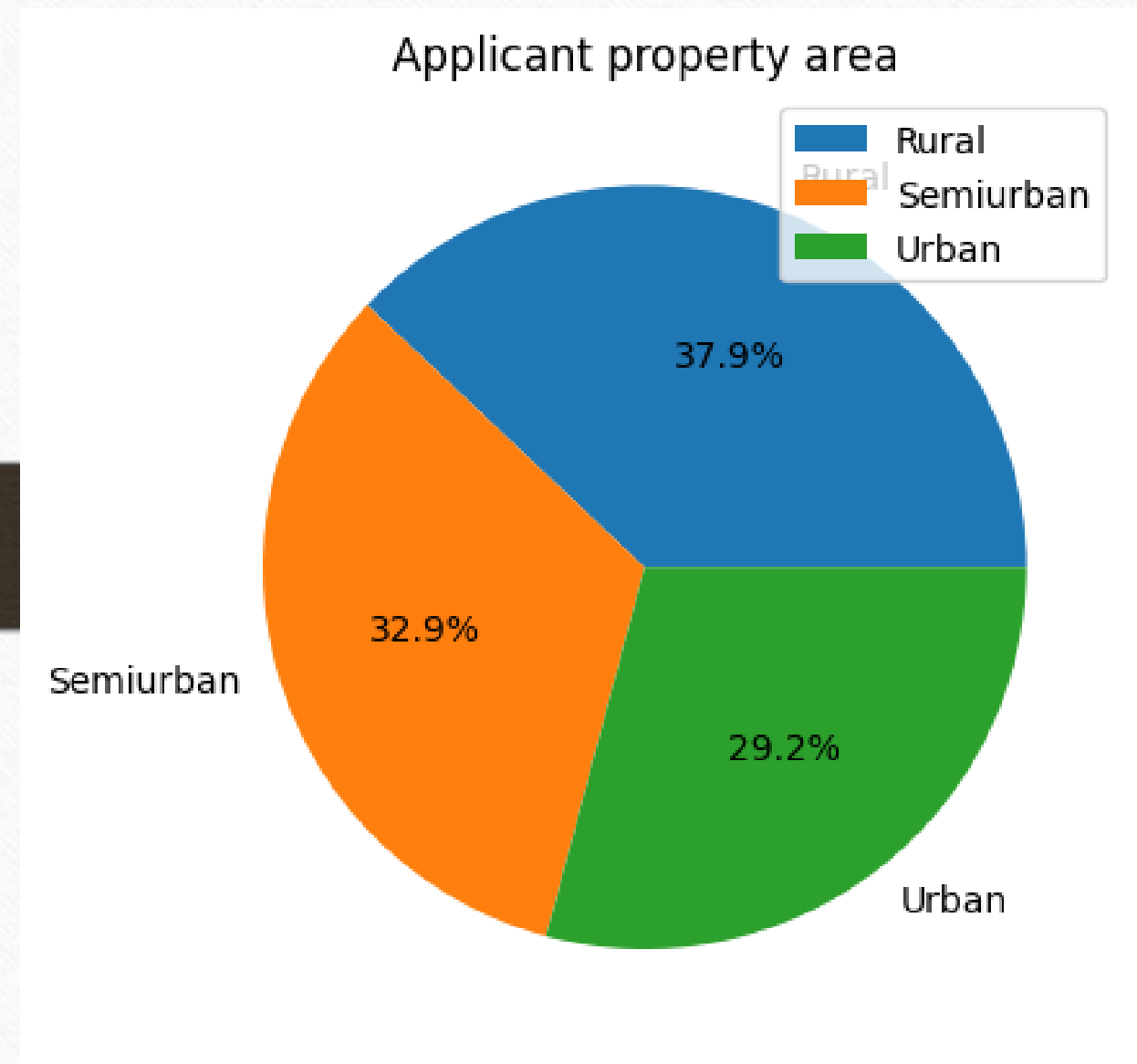
Loan Status : The loan is approved or not - (Yes / No)

First we check in our data that any value is missing or not



Here we can see that there are some columns have the missing values so we can fill the data using the mean , median and mode or remove this data.

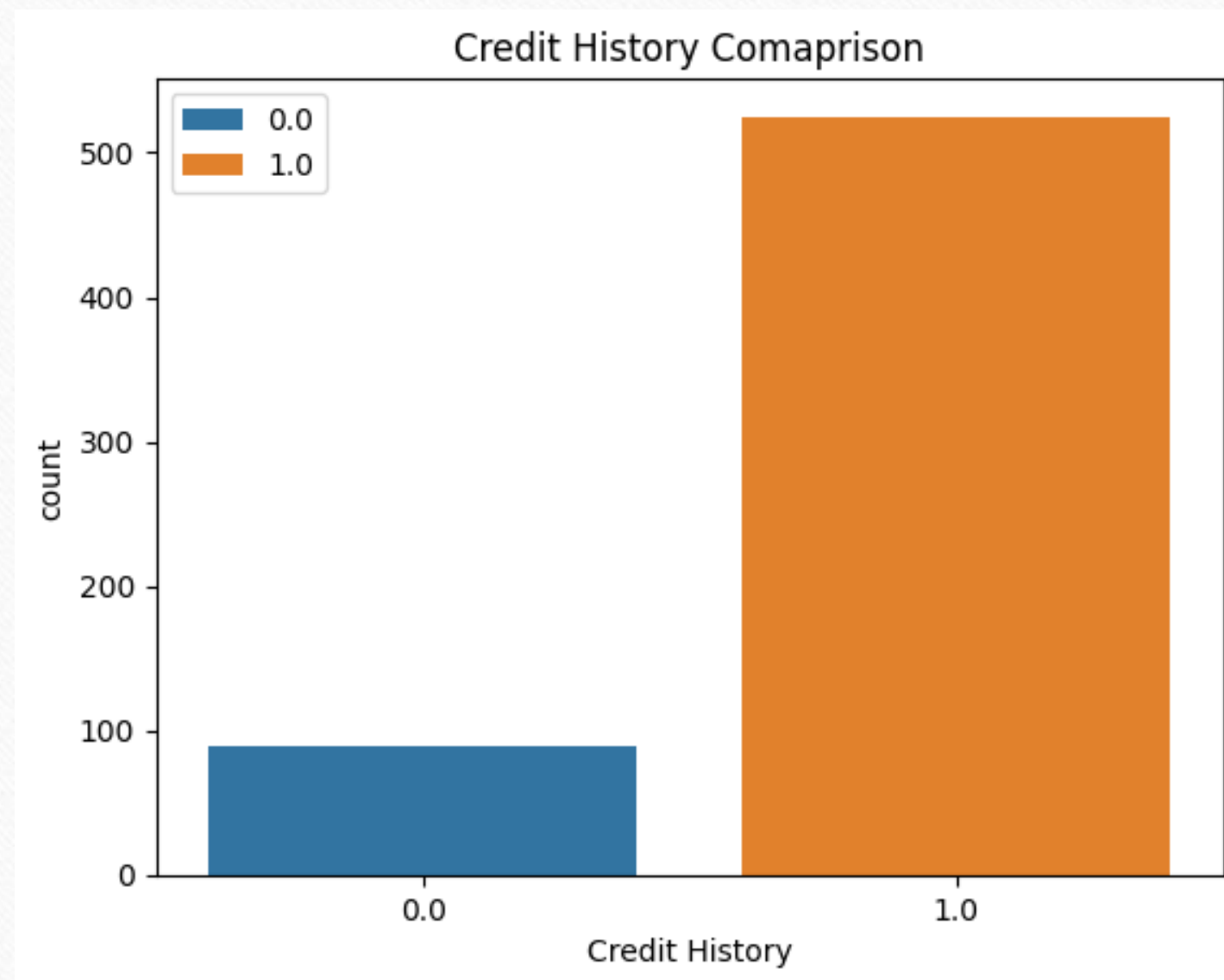
Now, let's visualize each variable separately.



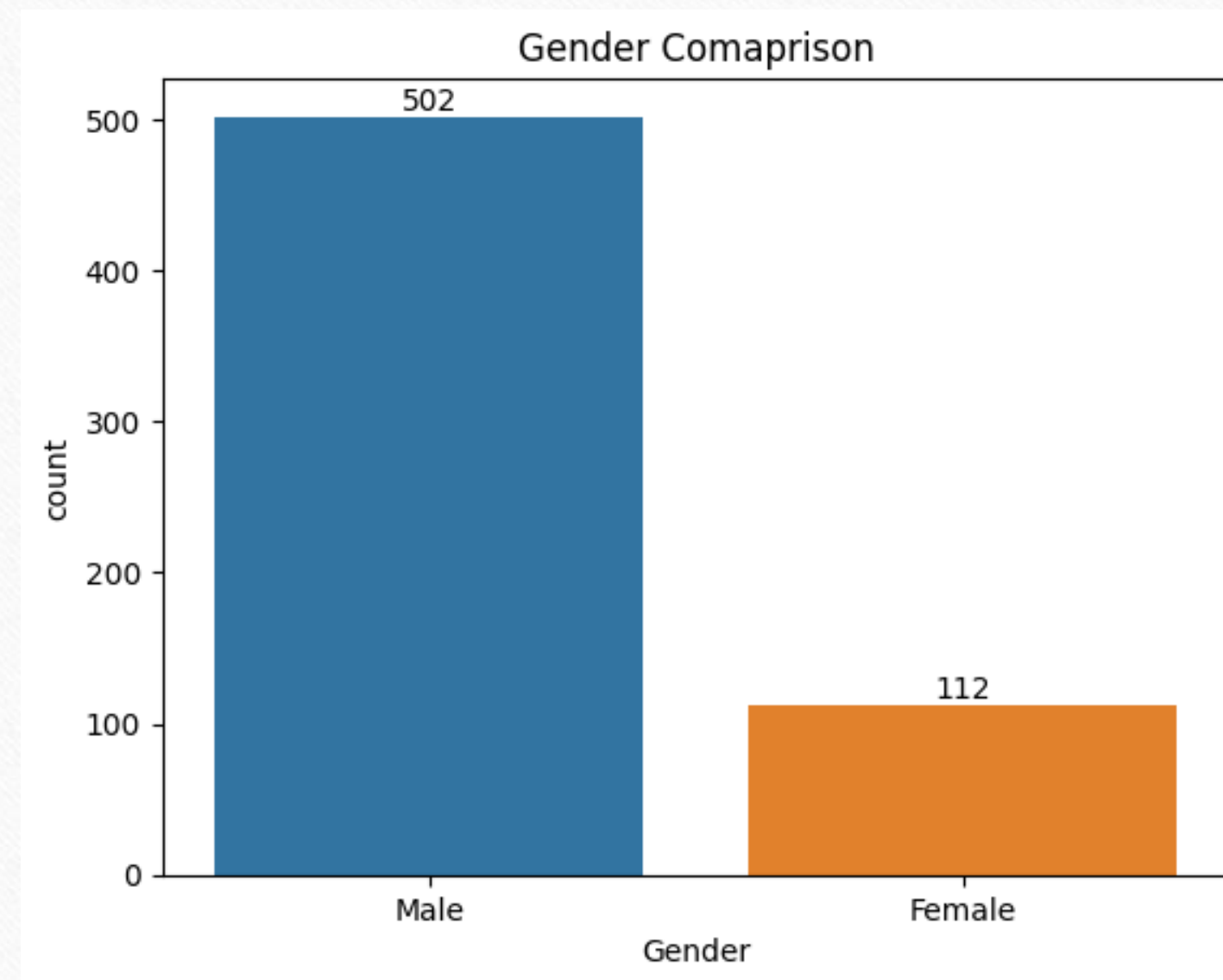
Different types of variables are Categorical, ordinal, and numerical

- **Categorical features:** These features have categories (Gender, Married, Self_Employed, Credit_History, Loan_Status)
- **Ordinal features:** Variables in categorical features having some order involved (Dependents, Education, Property_Area)
- **Numerical features:** These features have numerical values (ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term)

Now, let's visualize each variable separately.

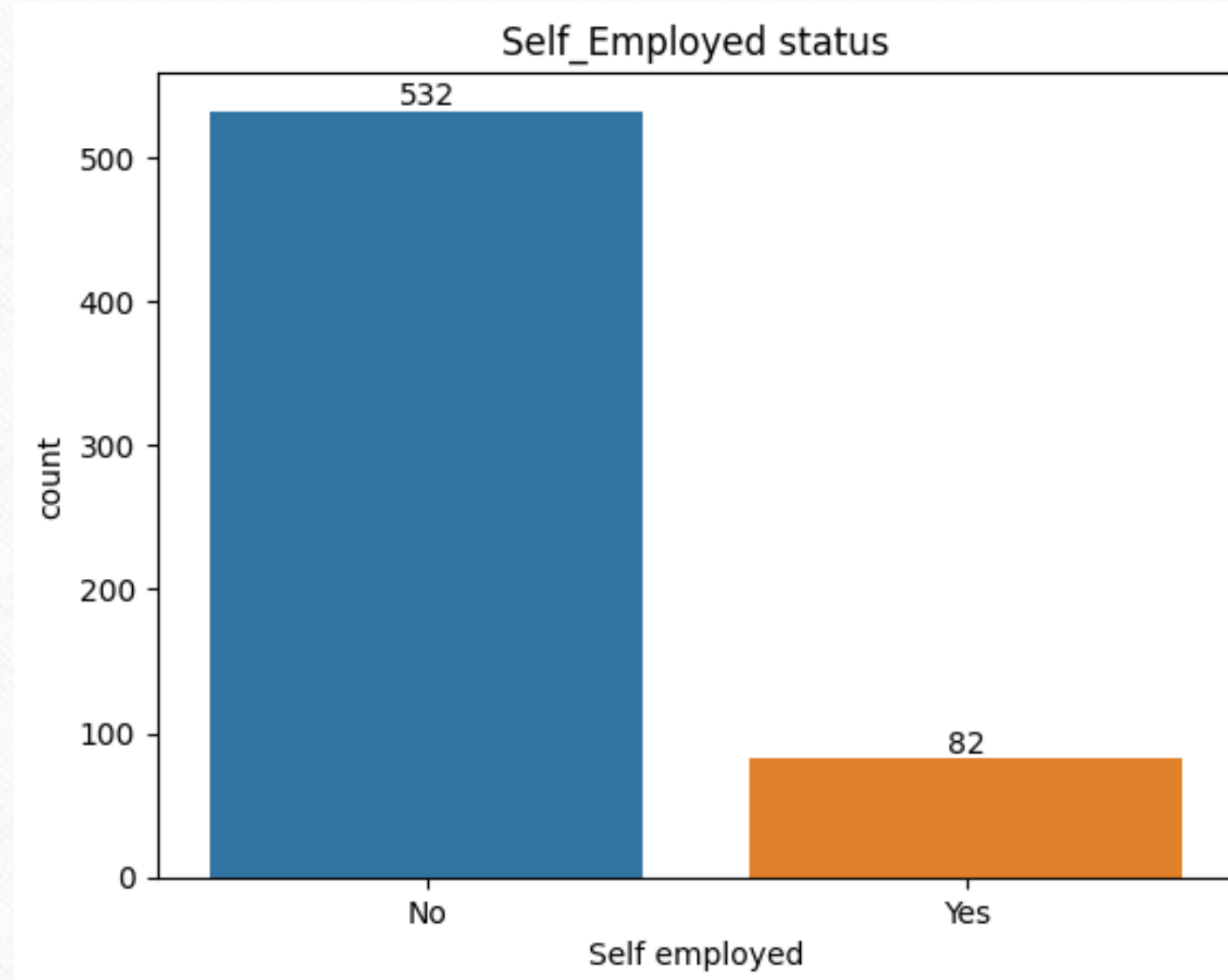


| Credit_History | |
|----------------|-----|
| 0.0 | 89 |
| 1.0 | 525 |

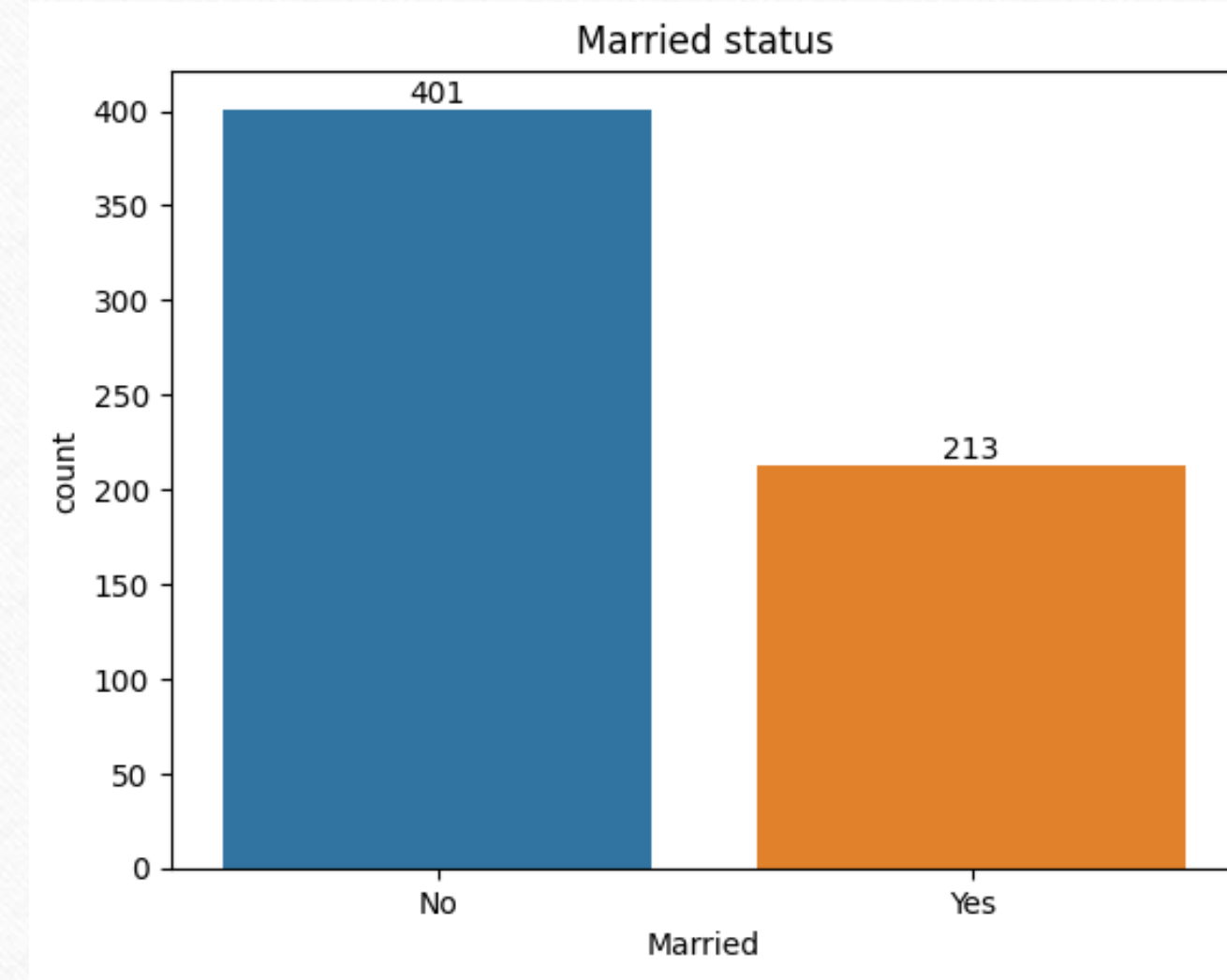


| Gender | |
|--------|-----|
| Male | 502 |
| Female | 112 |

Now, let's visualize each variable separately.

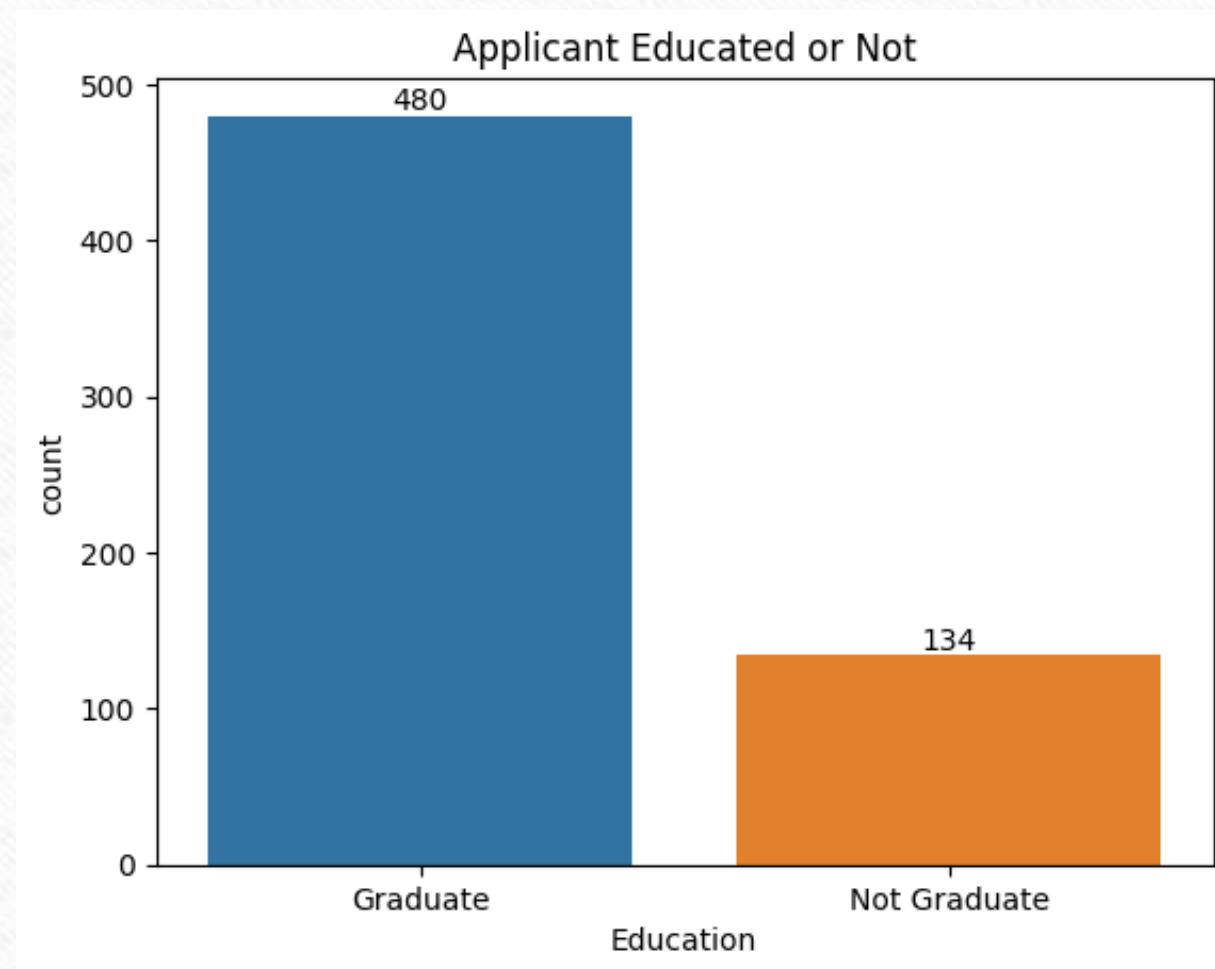


| self employed | |
|---------------|-----|
| Yes | 82 |
| No | 532 |

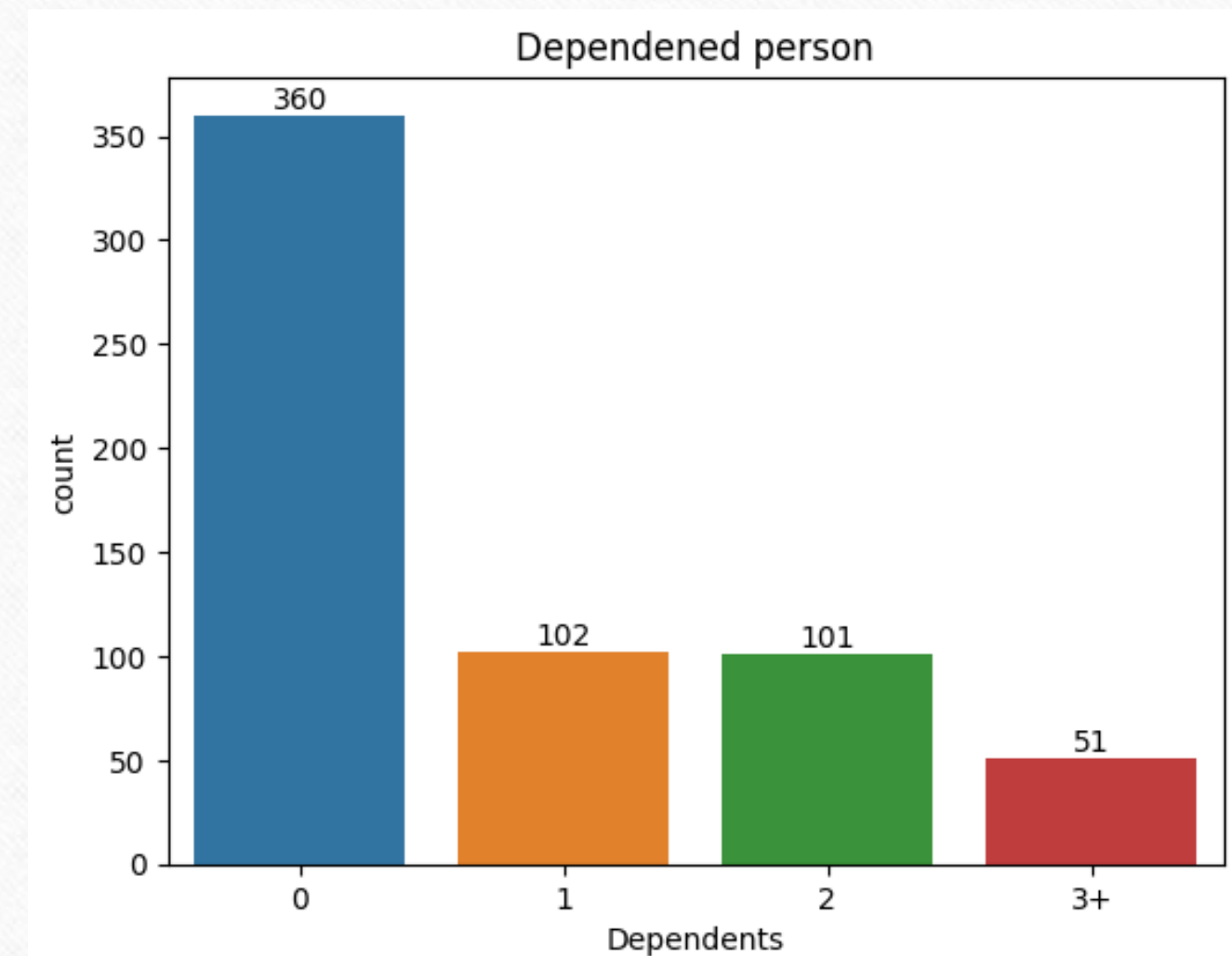


| Married status | |
|----------------|-----|
| Married | 401 |
| Unmarried | 213 |

Now, let's visualize each variable separately.



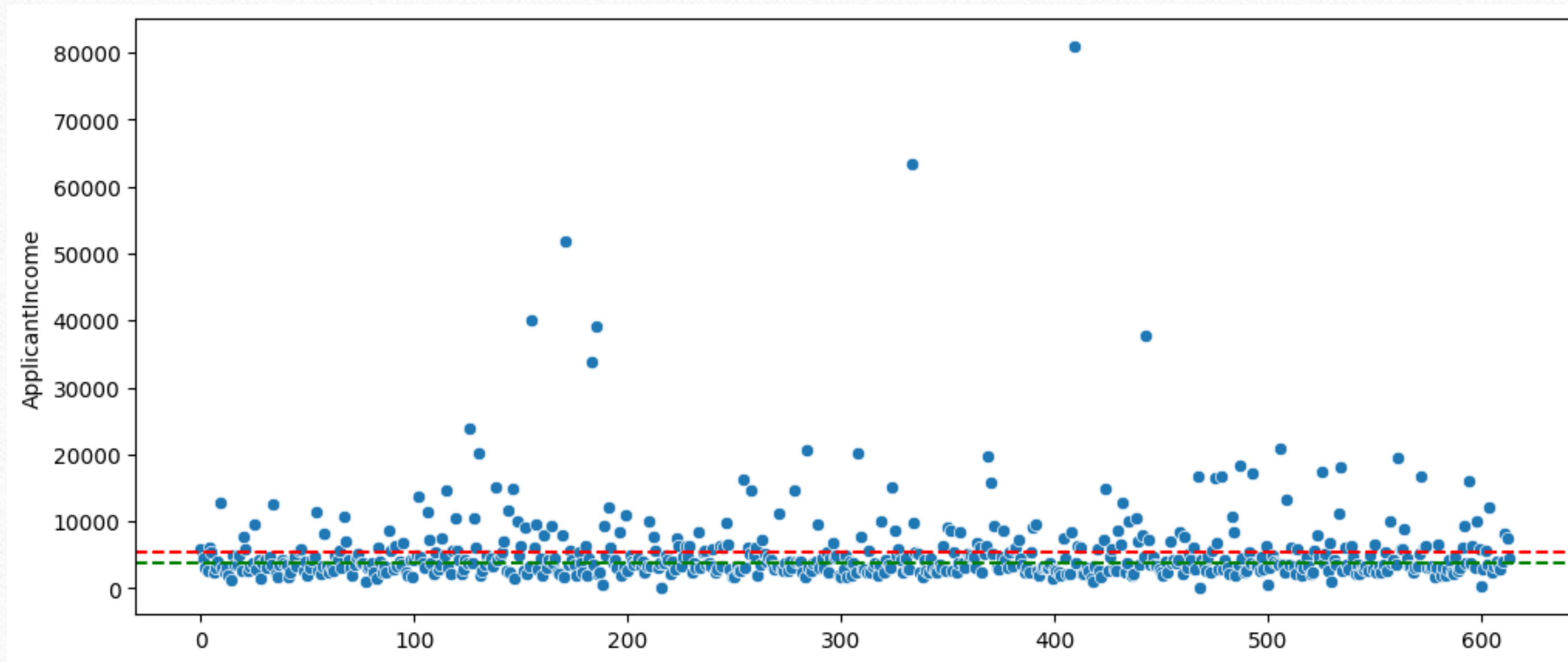
| Education | |
|--------------|-----|
| Graduate | 480 |
| Not Graduate | 134 |



| Dependents | |
|------------|-----|
| 0 | 360 |
| 1 | 102 |

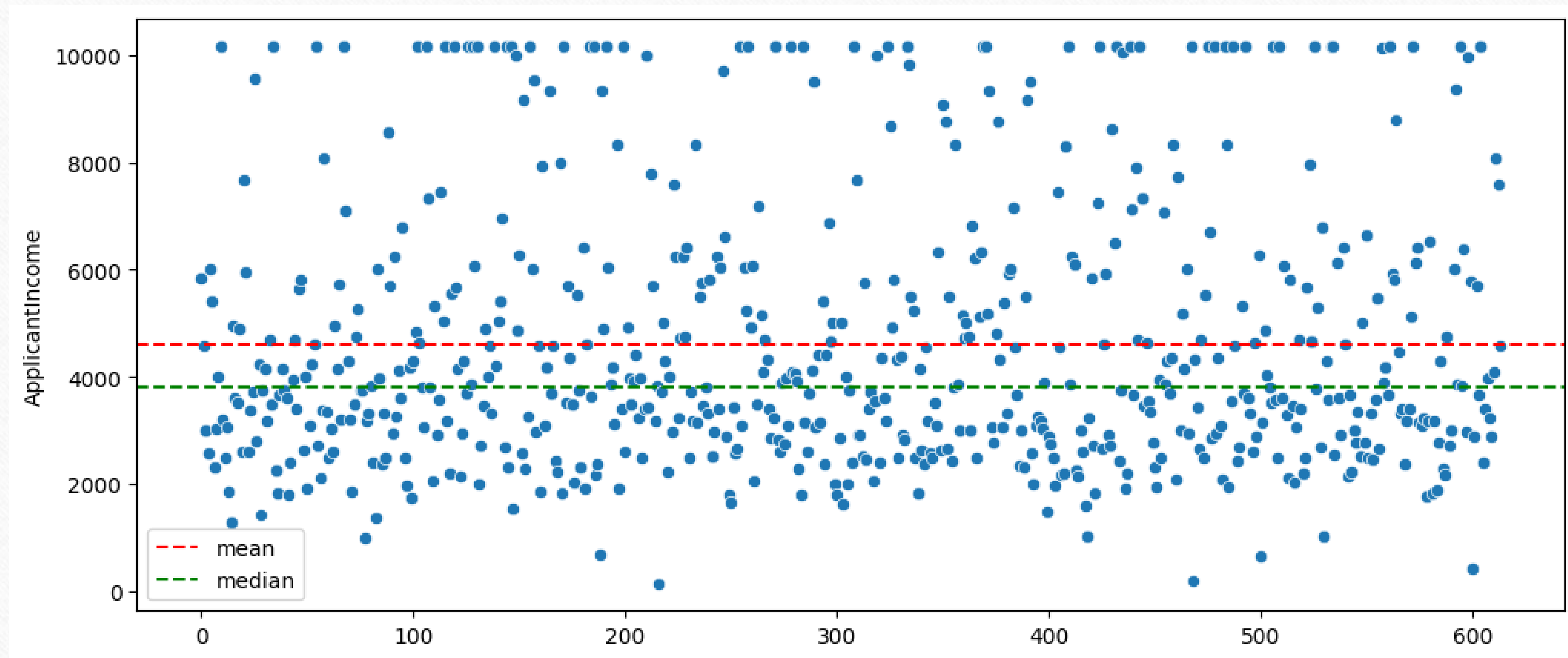
| Dependents | |
|------------|-----|
| 2 | 101 |
| 3+ | 51 |

Let's Find outliers of our Data.

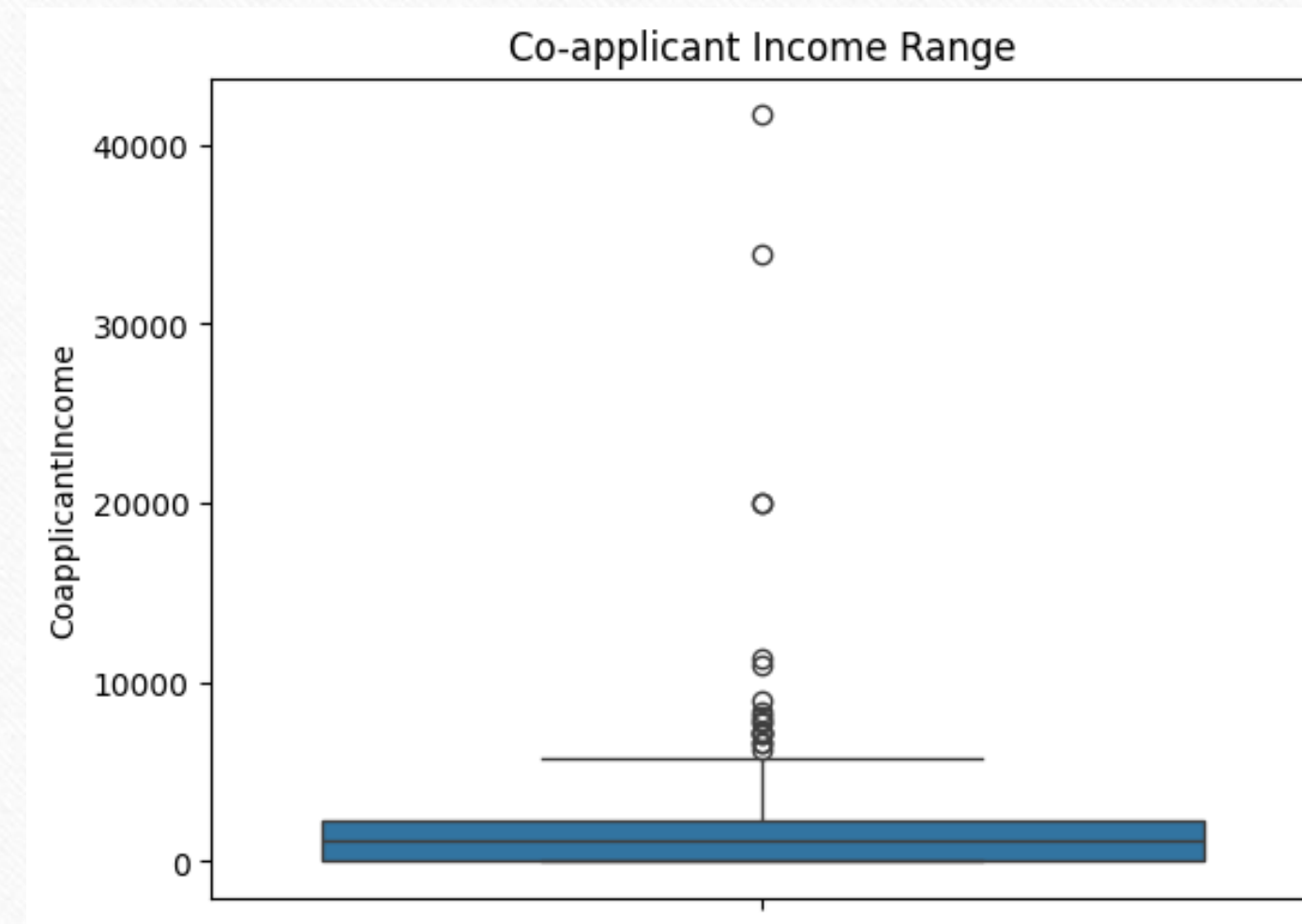
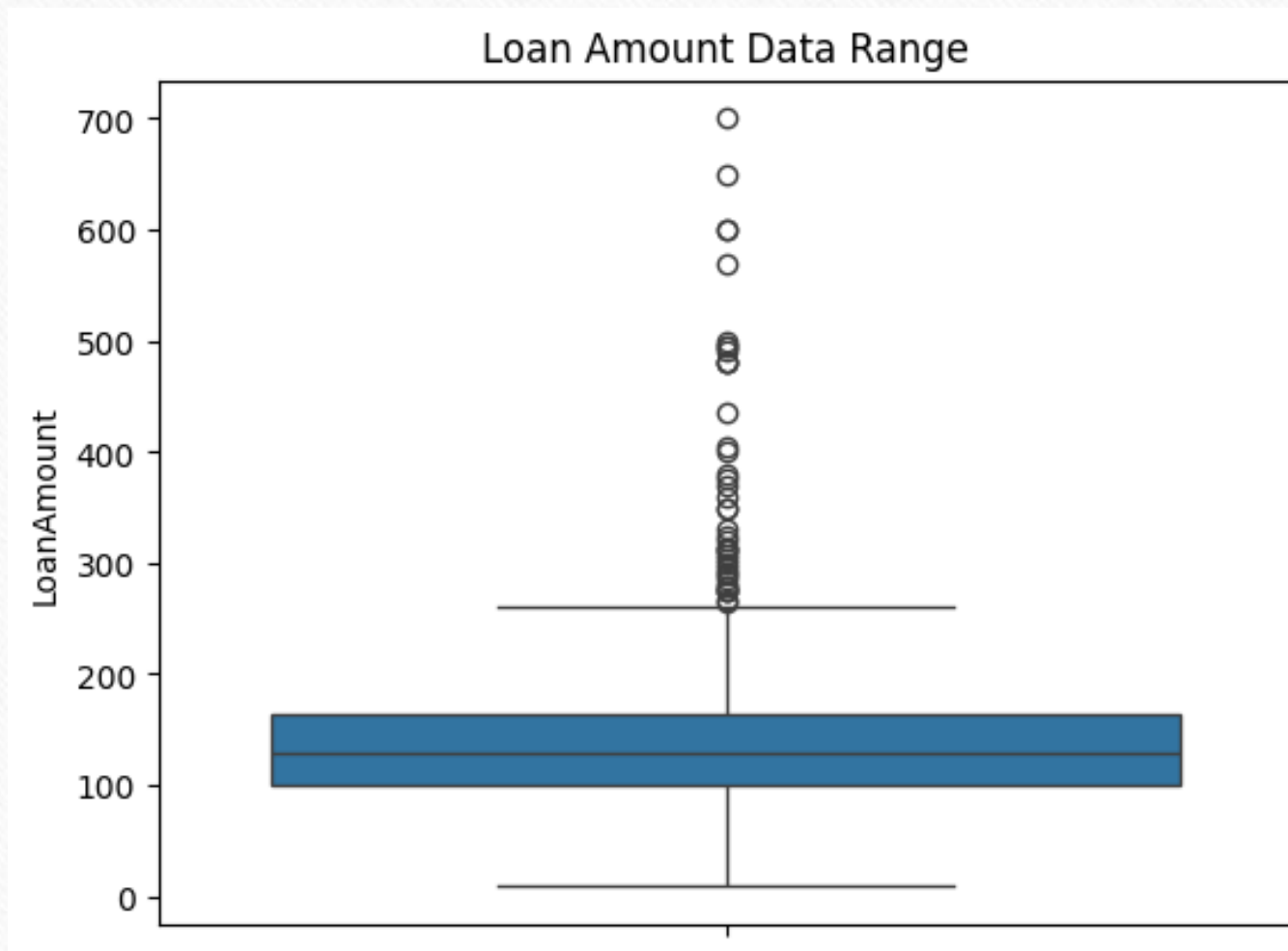


- Here in this graph we can show the applicant Income range. Here mostly data is lie between the range of 15000 around.

- It's most important thing to remove the anomaly data. If we can not remove it so our prediction may be wrong

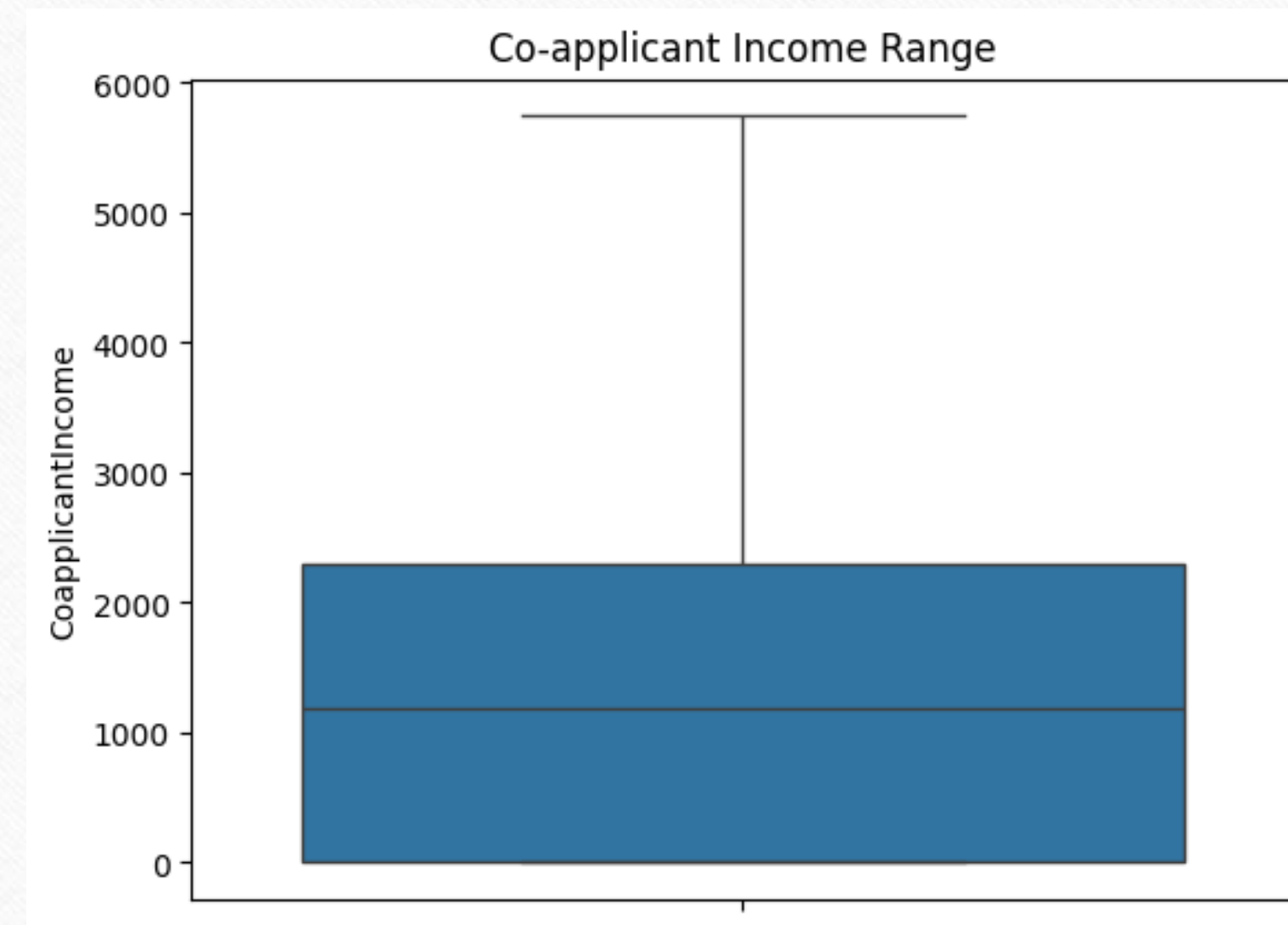
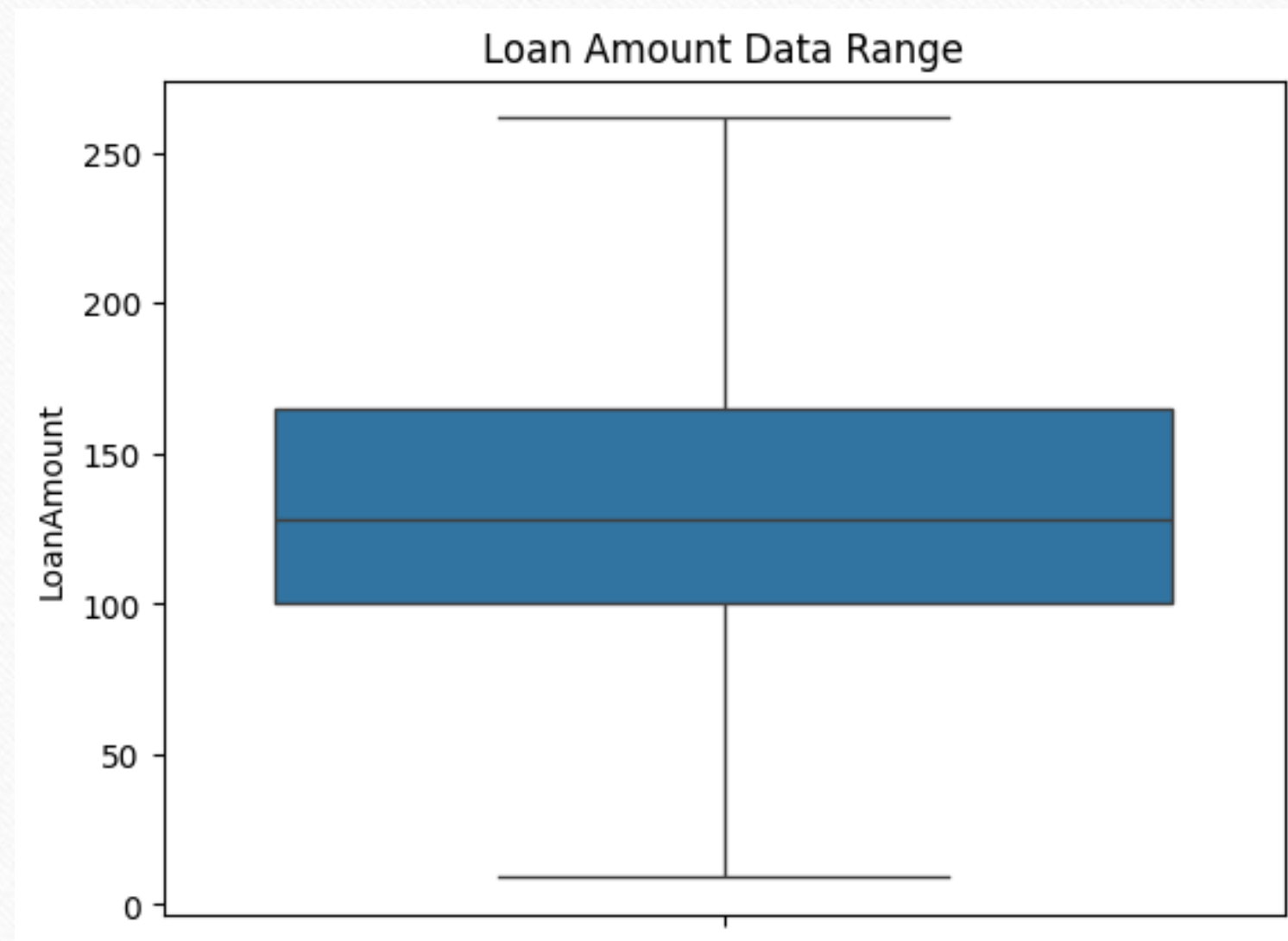


Here we show the different columns has outliers or not.

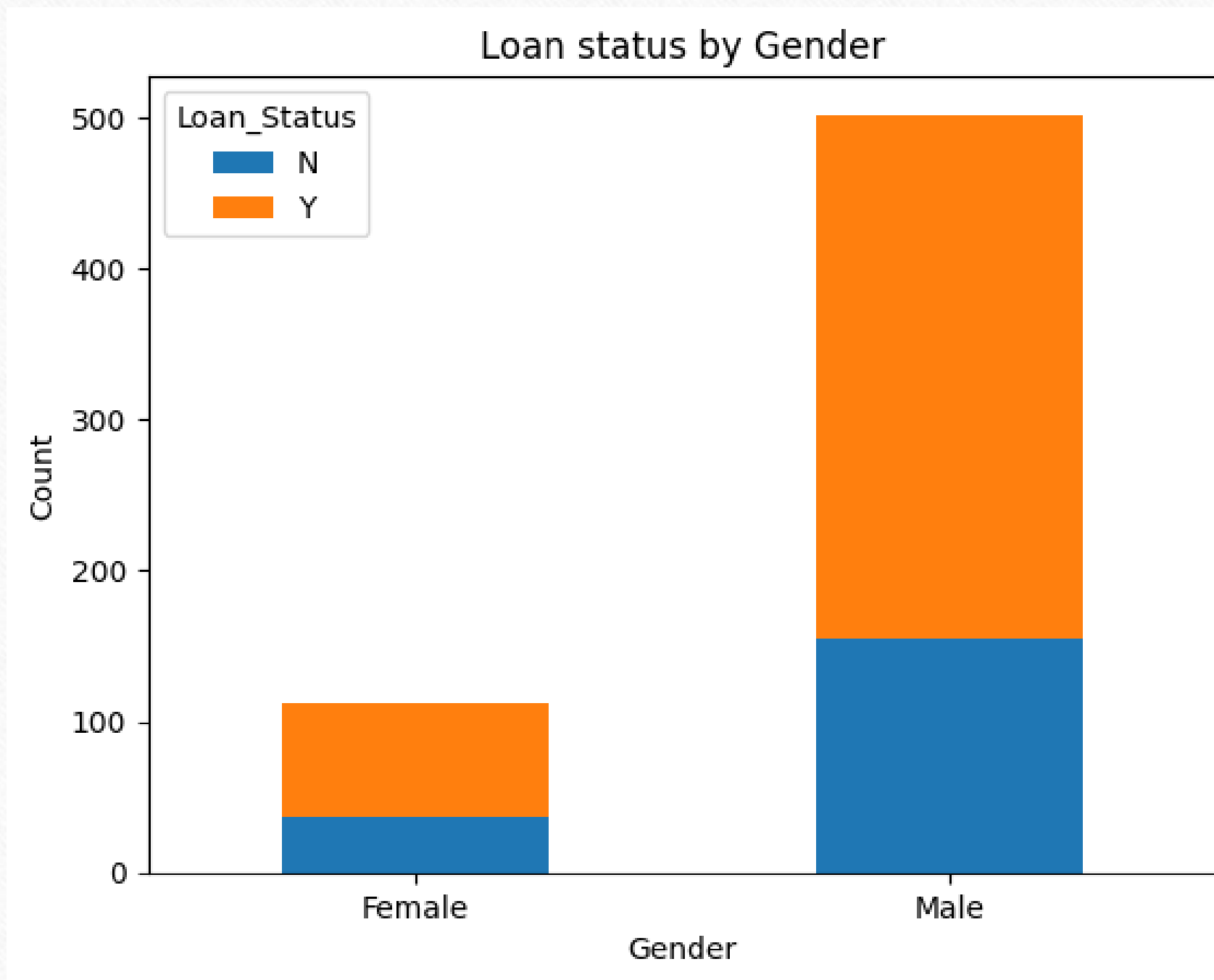


- Here the show two columns on is loan amount data Range and another is co-applicant income range. In the loan amount data range data is mostly lie between the 0 to 300 and another data is outliers as well as in the coapplicant income data is mostly lier between the 0 to 9000 another data is called the outliers.

- After removing the outliers.



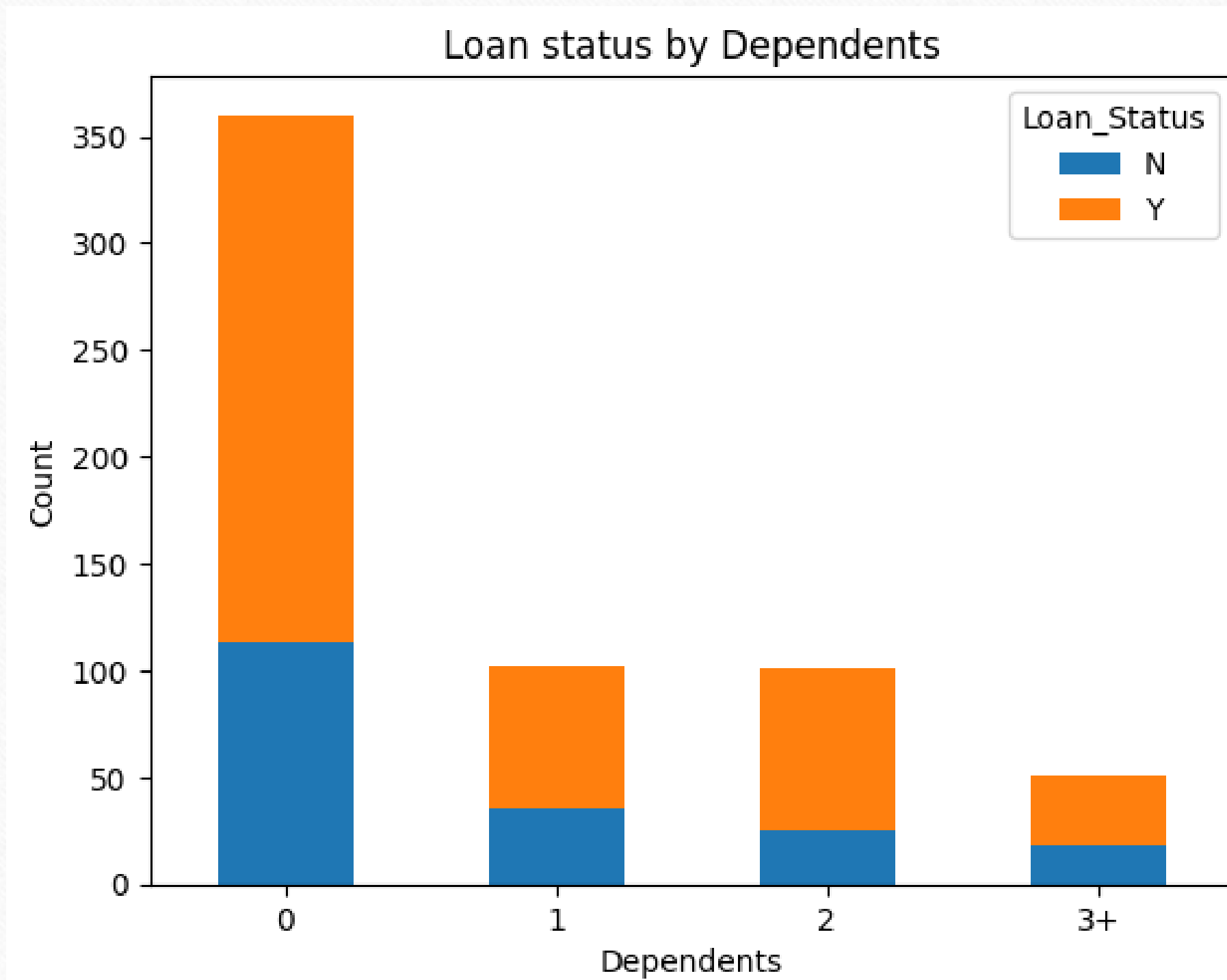
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Gender



- we will find the relation between the target variable (Loan status) and categorical independent variables (Gender).

| Gender | Total | Loan(%) (Y) |
|--------|-----------|----------------|
| Male | 502 (82%) | 347 (69%) |
| Female | 112 (18%) | 75 (66%) |

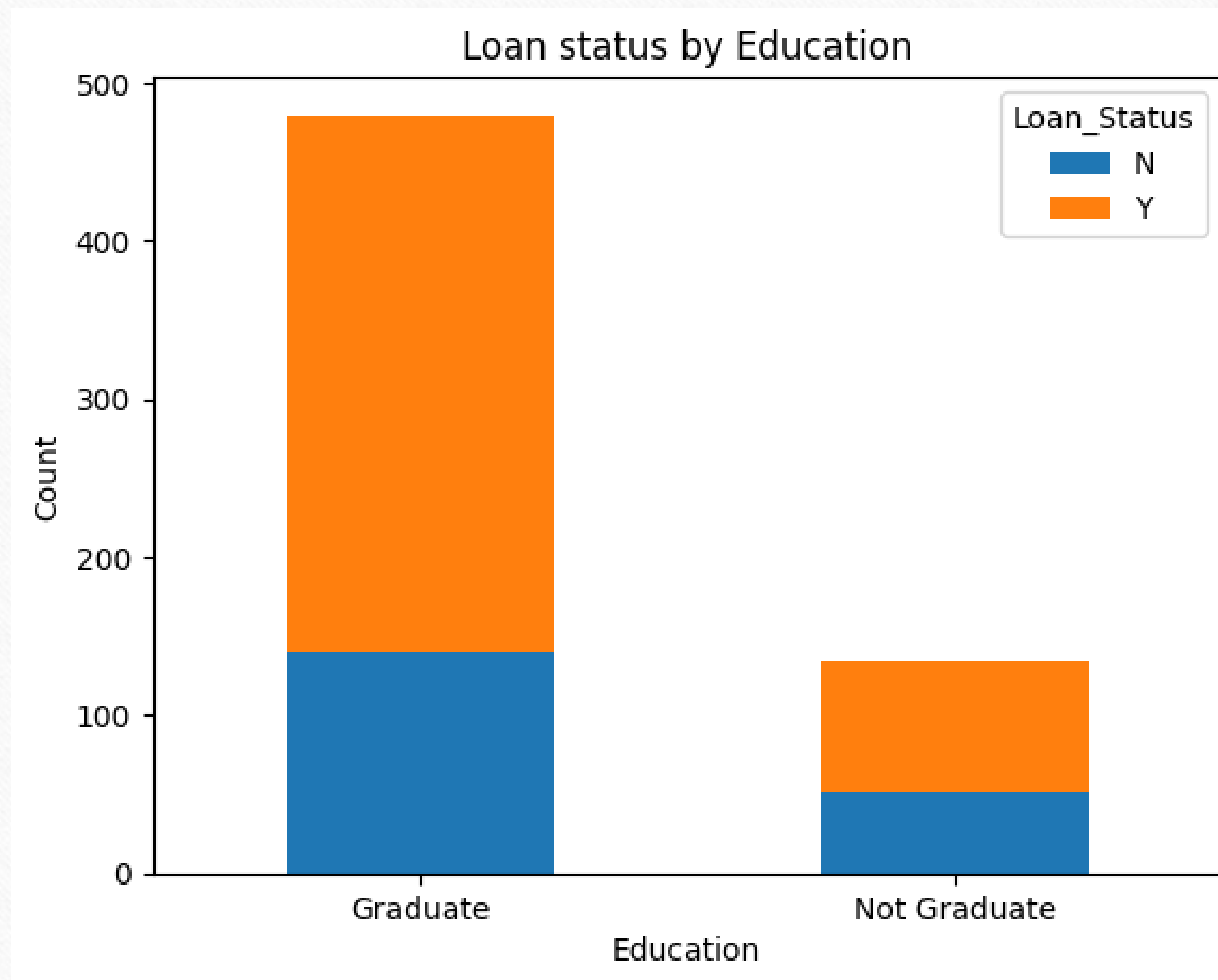
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Dependents



- we will find the relation between the target variable (Loan status) and categorical independent variables (Dependents).

| Dependents | Total | Loan Status (Y) |
|------------|-----------|-----------------|
| 0 | 360 (59%) | 247 (68%) |
| 1 | 102 (17%) | 66 (64%) |
| 2 | 101 (16%) | 76 (75%) |
| 3+ | 51 (8%) | 33 (64%) |

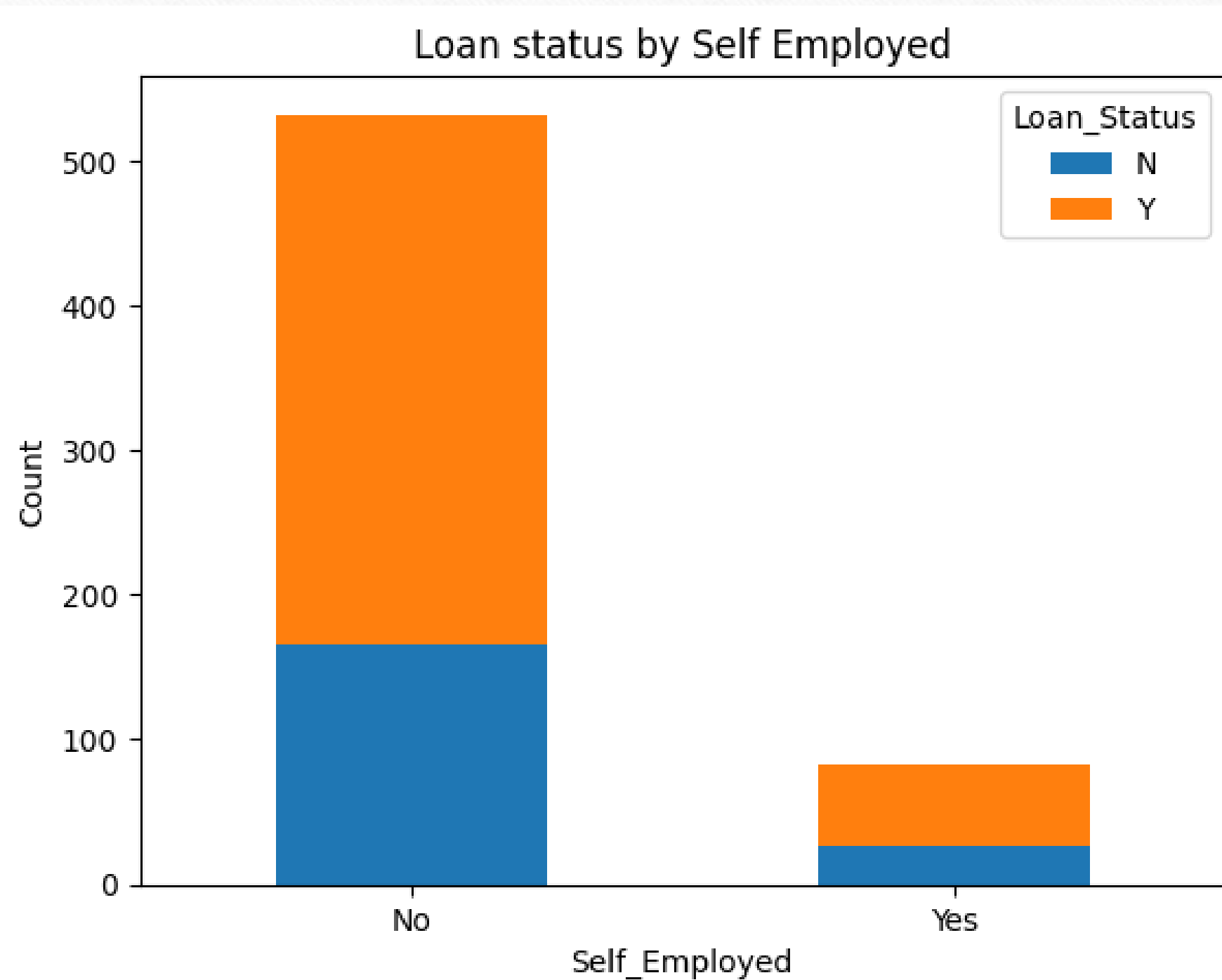
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Education



- we will find the relation between the target variable (Loan status) and categorical independent variables (Education).

| Gender | Total | Loan(%) (Y) |
|--------------|-----------|----------------|
| Graduate | 480 (78%) | 340 (71%) |
| Non Graduate | 134 (22%) | 82 (61%) |

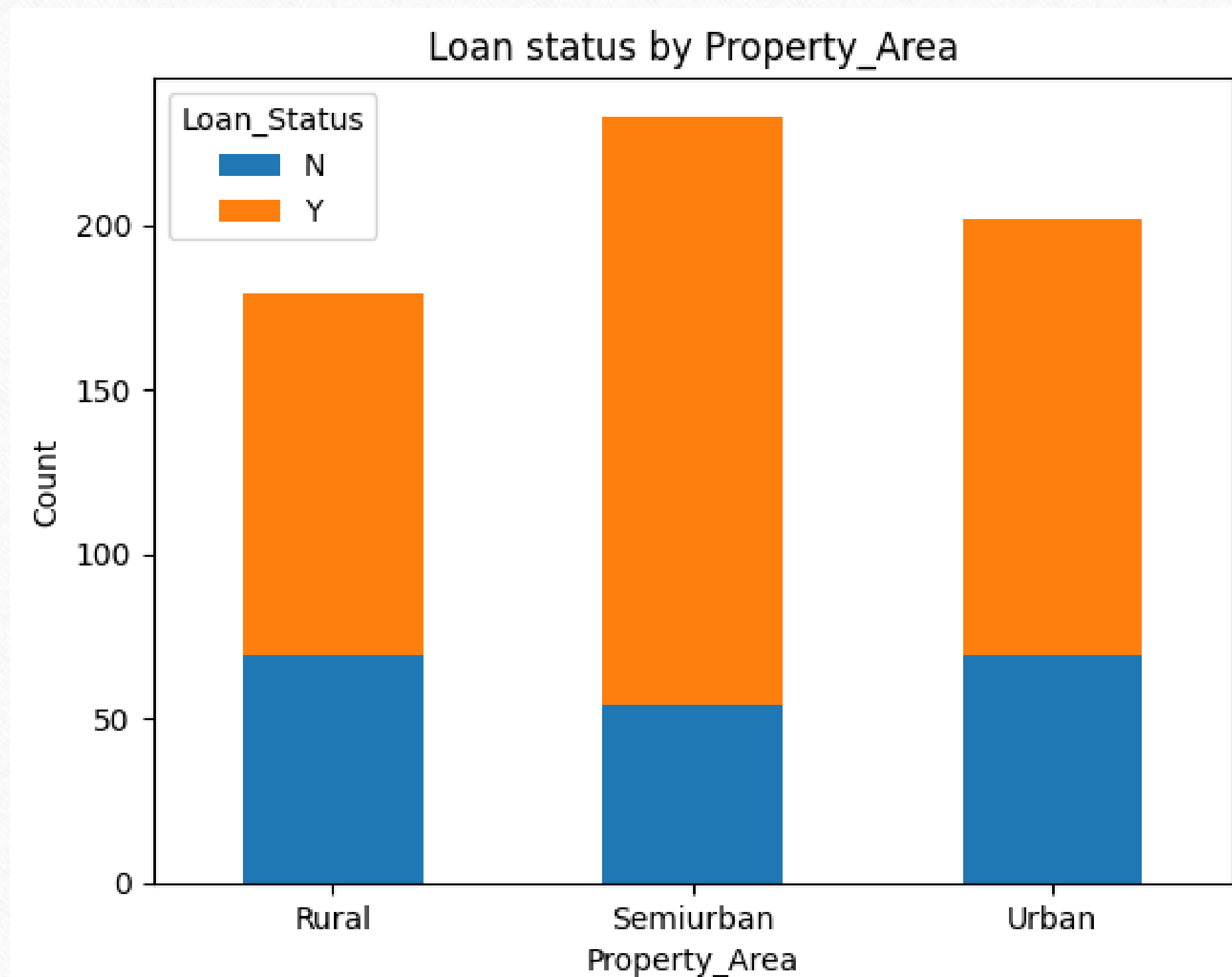
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Education



- we will find the relation between the target variable (Loan status) and categorical independent variables (Self-Employed).

| Gender | Total | Loan(%) (Y) |
|--------|-----------|----------------|
| No | 532 (87%) | 366 (69%) |
| Yes | 82 (13%) | 56 (68%) |

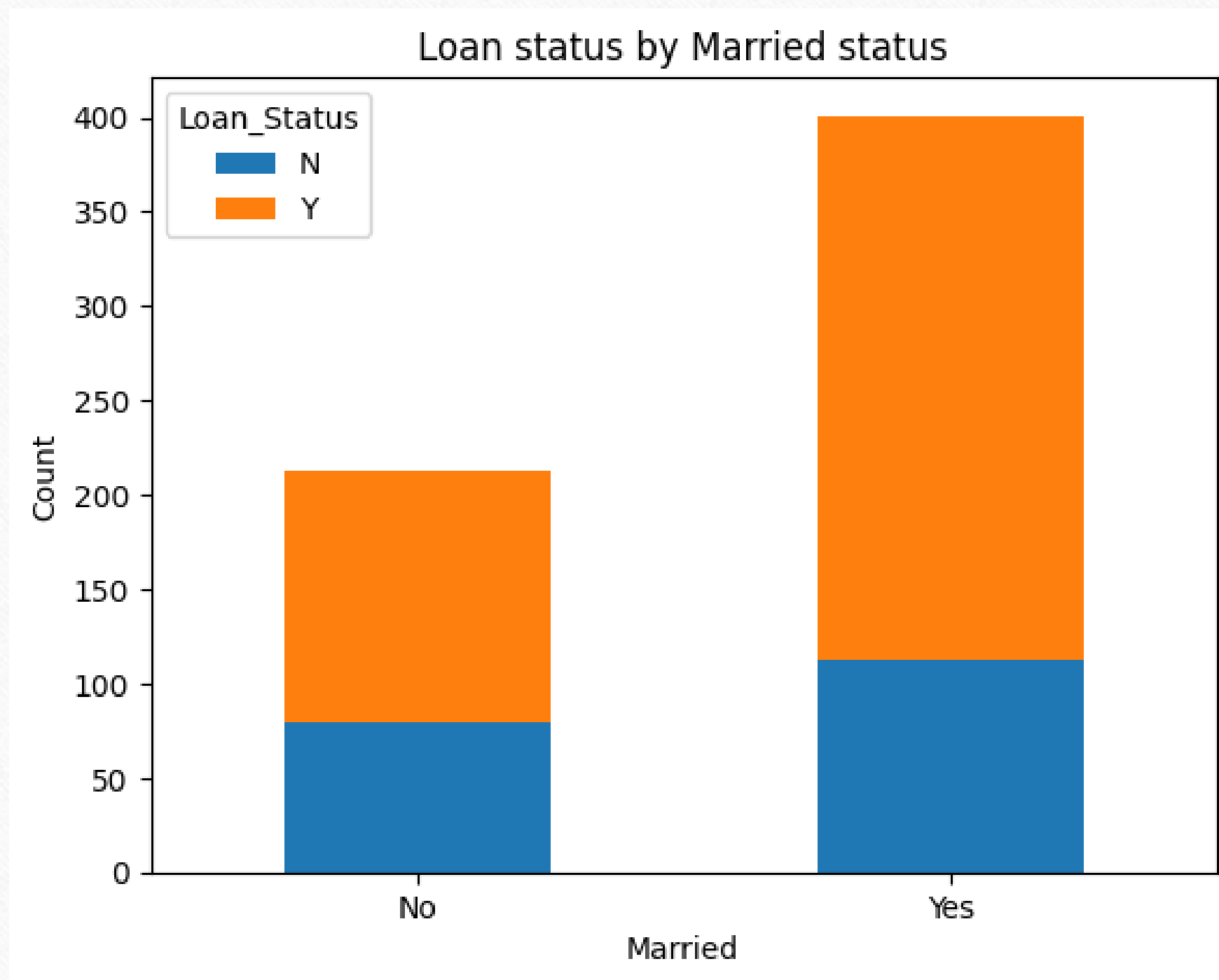
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Education



- we will find the relation between the target variable (Loan status) and categorical independent variables (Property Area).

| Gender | Total | Loan(%) (Y) |
|-----------|-----------|----------------|
| Rural | 179 (29%) | 110 (61%) |
| Semiurban | 233 (38%) | 179 (77%) |
| Urban | 202 (33%) | 133 (66%) |

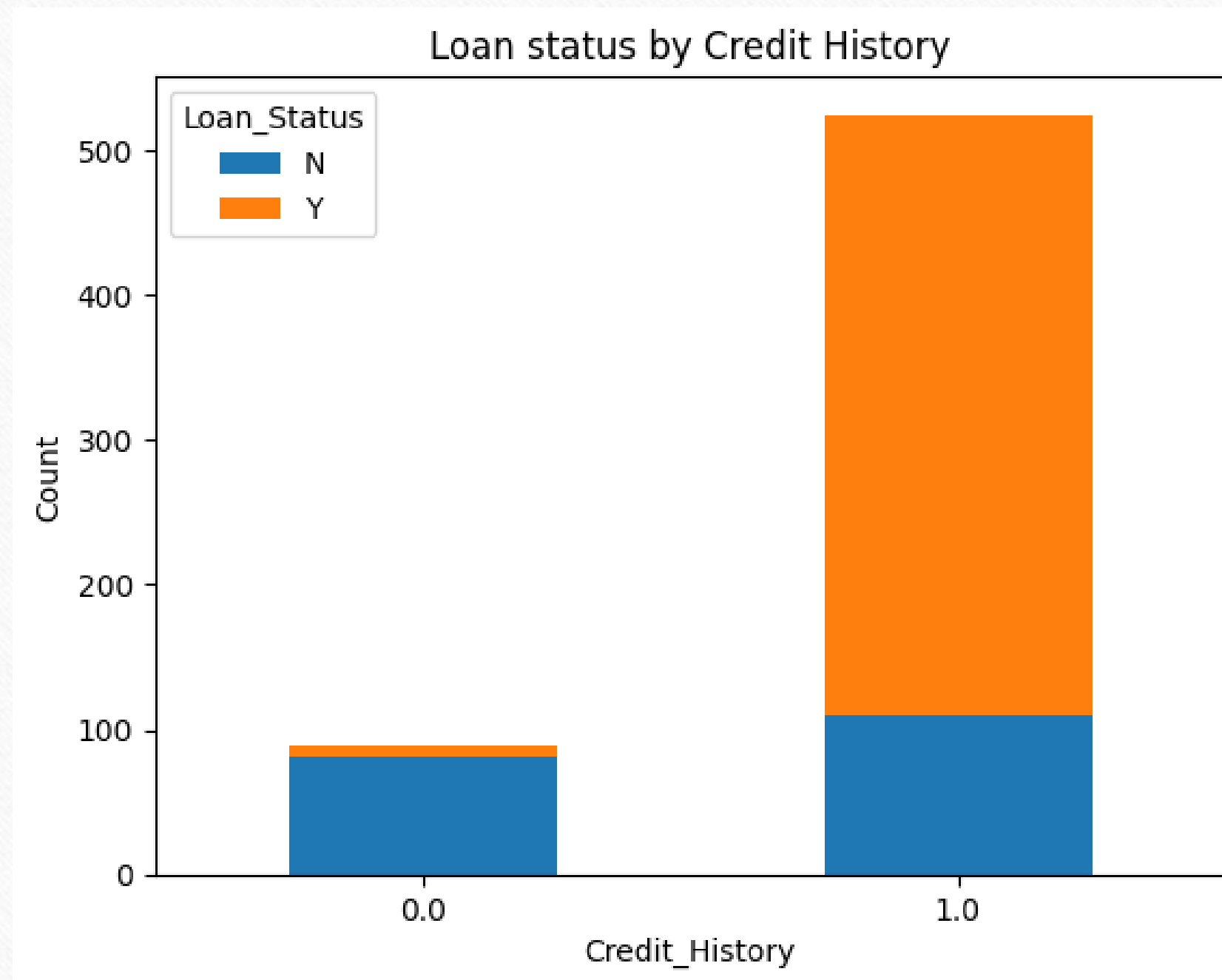
- Here we analyze that every factor is impact the loan or not. Here we will analyse the Education



- we will find the relation between the target variable (Loan status) and categorical independent variables (Married).

| Gender | Total | Loan(%) (Y) |
|--------|-----------|----------------|
| No | 213 (35%) | 134 (63%) |
| Yes | 401 (65%) | 288 (72%) |

- Here we analyze that every factor is impact the loan or not. Here we will analyse the Education



- we will find the relation between the target variable (Loan status) and categorical independent variables (Credit History).

| Gender | Total | Loan(%) (Y) |
|--------|-----------|----------------|
| 0 | 89 (14%) | 7 (8%) |
| 1 | 525 (86%) | 415 (80%) |



Thank You