

# CREDIT EDA ASSIGNMENT

BY-

APRAJITA SRIVASTAVA

# PROBLEM STATEMENT

- ❖ The company as lending institution faces two types of challenges-
  - 1)Giving loan to defaulter results in loss of business to company
  - 2)Not approving loan to clients having repayment capacity results in financial loss to company.
- ❖ To solve this problem ,two csv files have been taken(application\_data and previous application).Analysis is done on them in Jupyter notebook.

# ANALYSIS

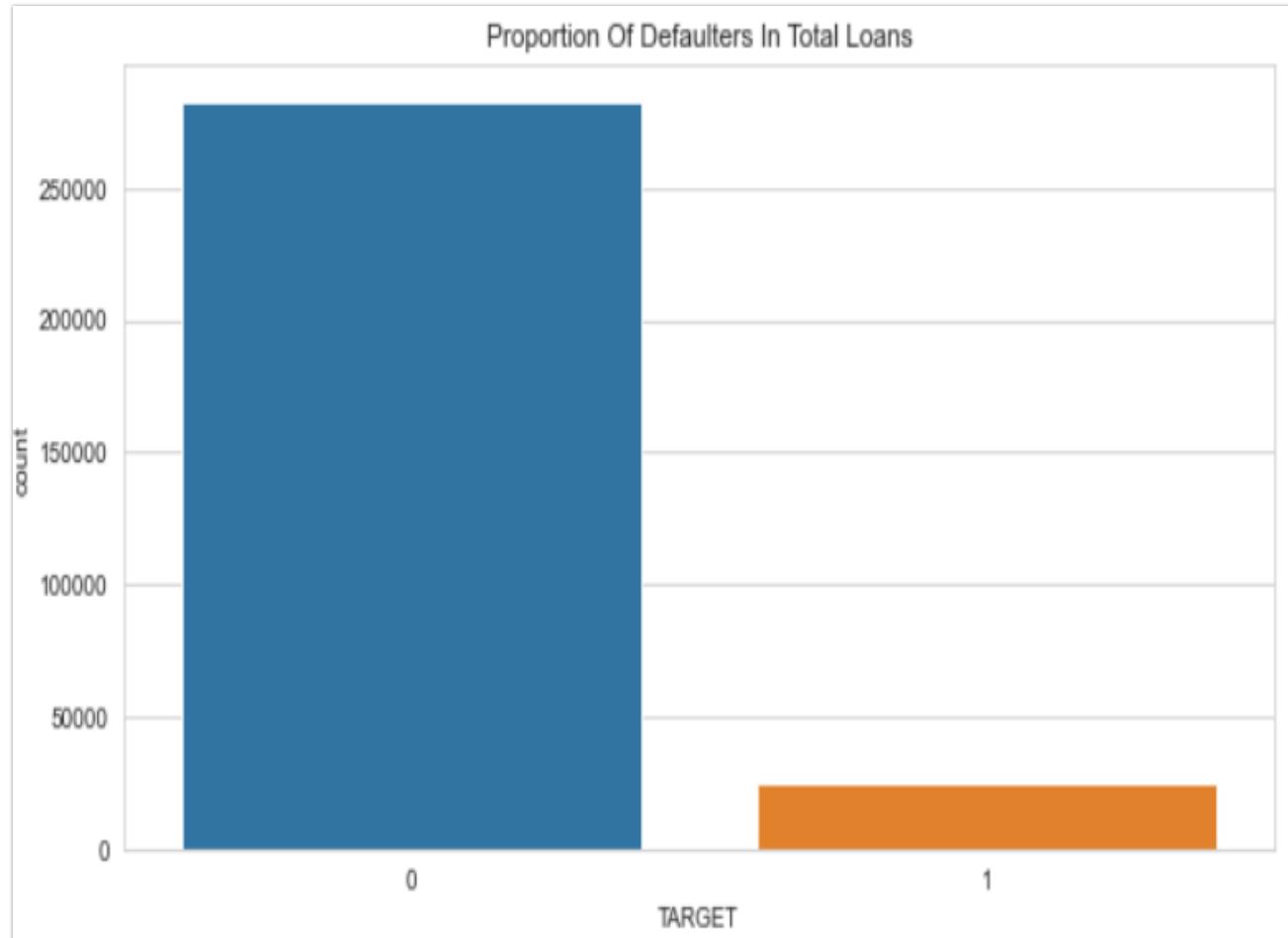
The steps involved are -

- ❖ Checking the structure of data
- ❖ Data quality check and missing value-1:finding percentage of missing values and removing high missing value.
- ❖ Data quality check and missing value-2:checking data types,outliers for numerical columns and binning.
- ❖ Imbalance percentage,correlation,univariate analysis and bivariate analysis.

## DATA IMBALANCE

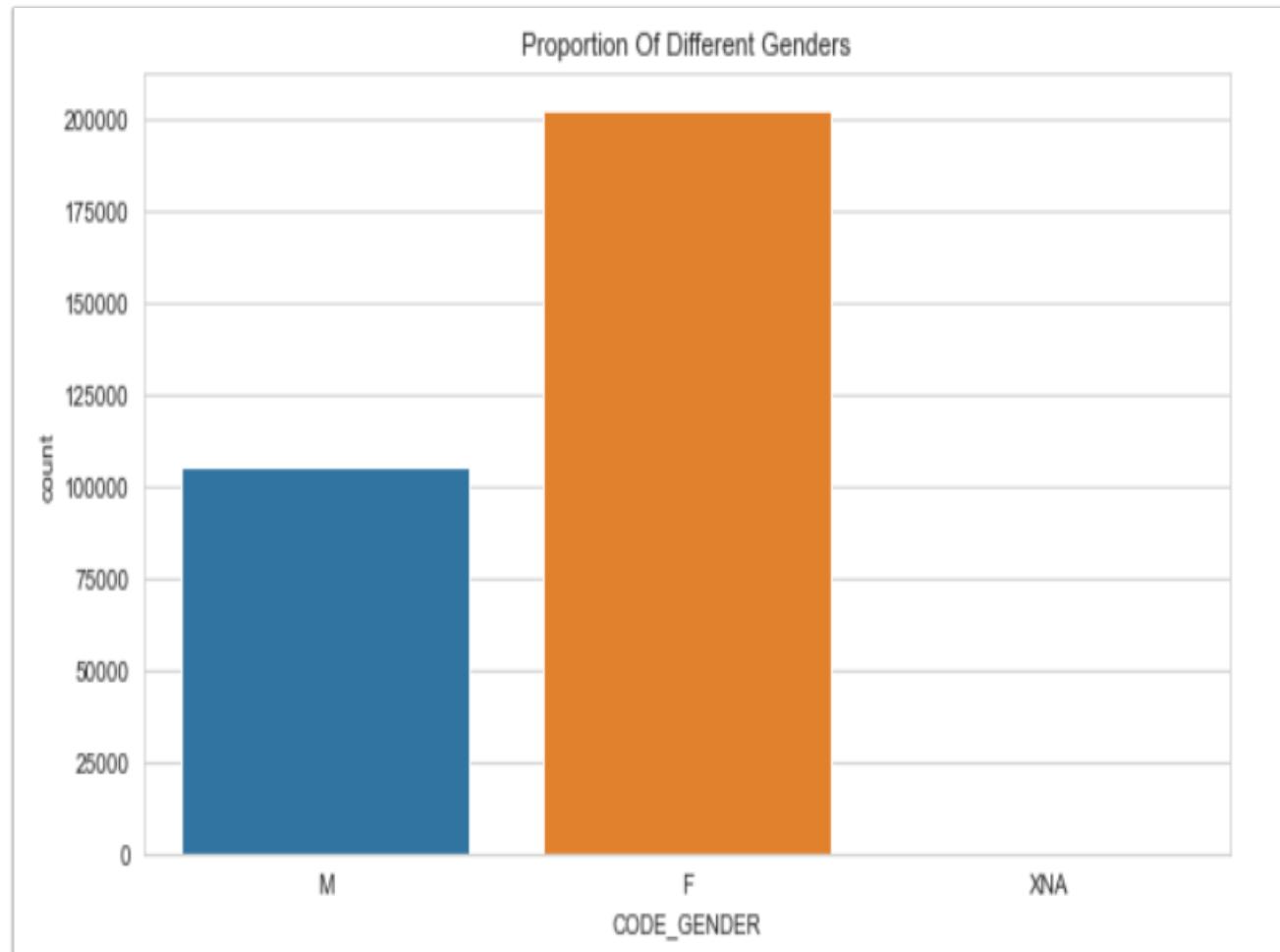
1) PROPORTION OF  
DEFAULTERS IN  
APPLICATION  
DATASET-

91.9% :Non-defaulter 8.07%  
:Defaulter



## 2)PROPORTION OF GENDER IN APPLICATION DATASET-

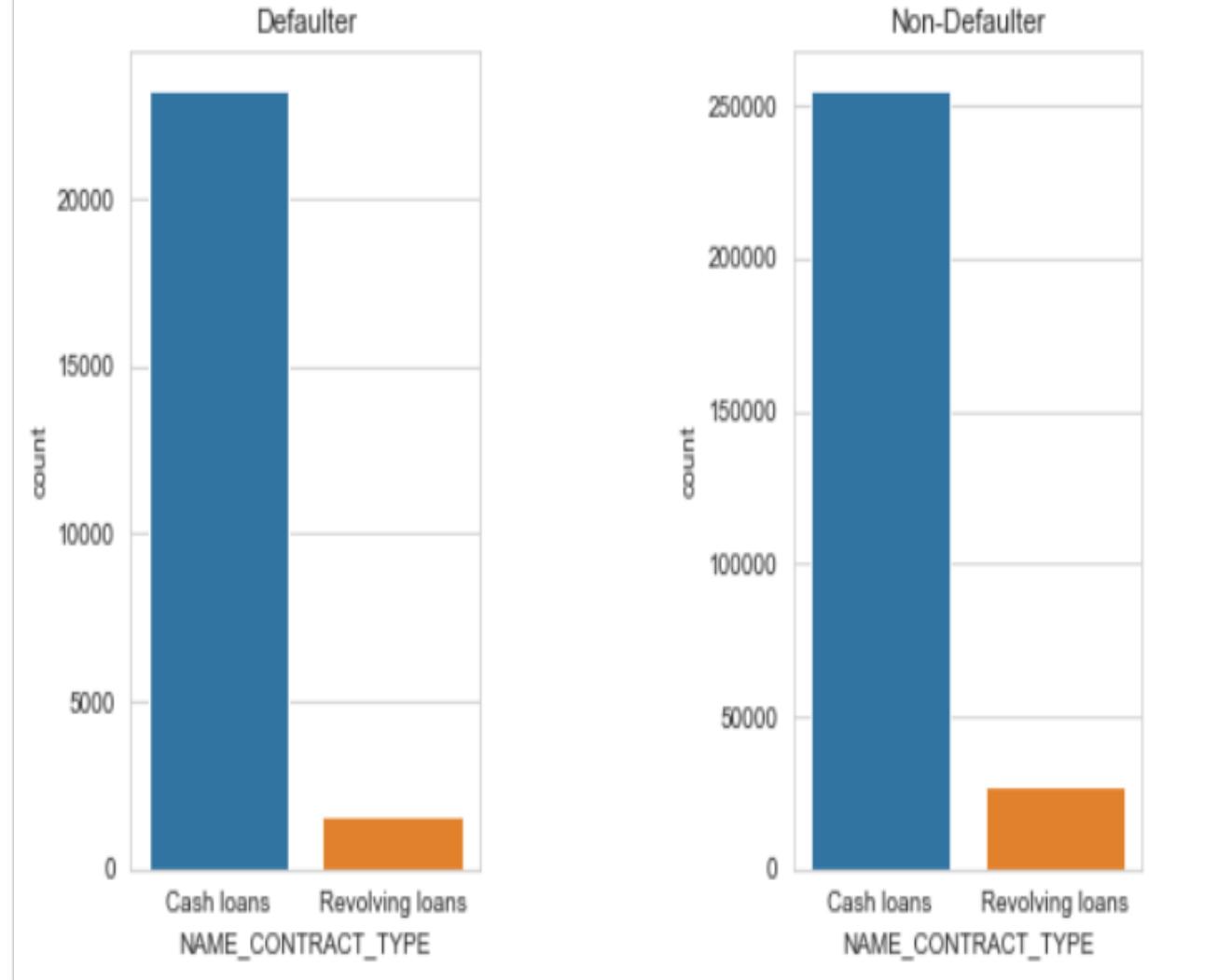
Female-65.8% Male-34.1%



## UNIVARIATE ANALYSIS OF APPLICATION DATA

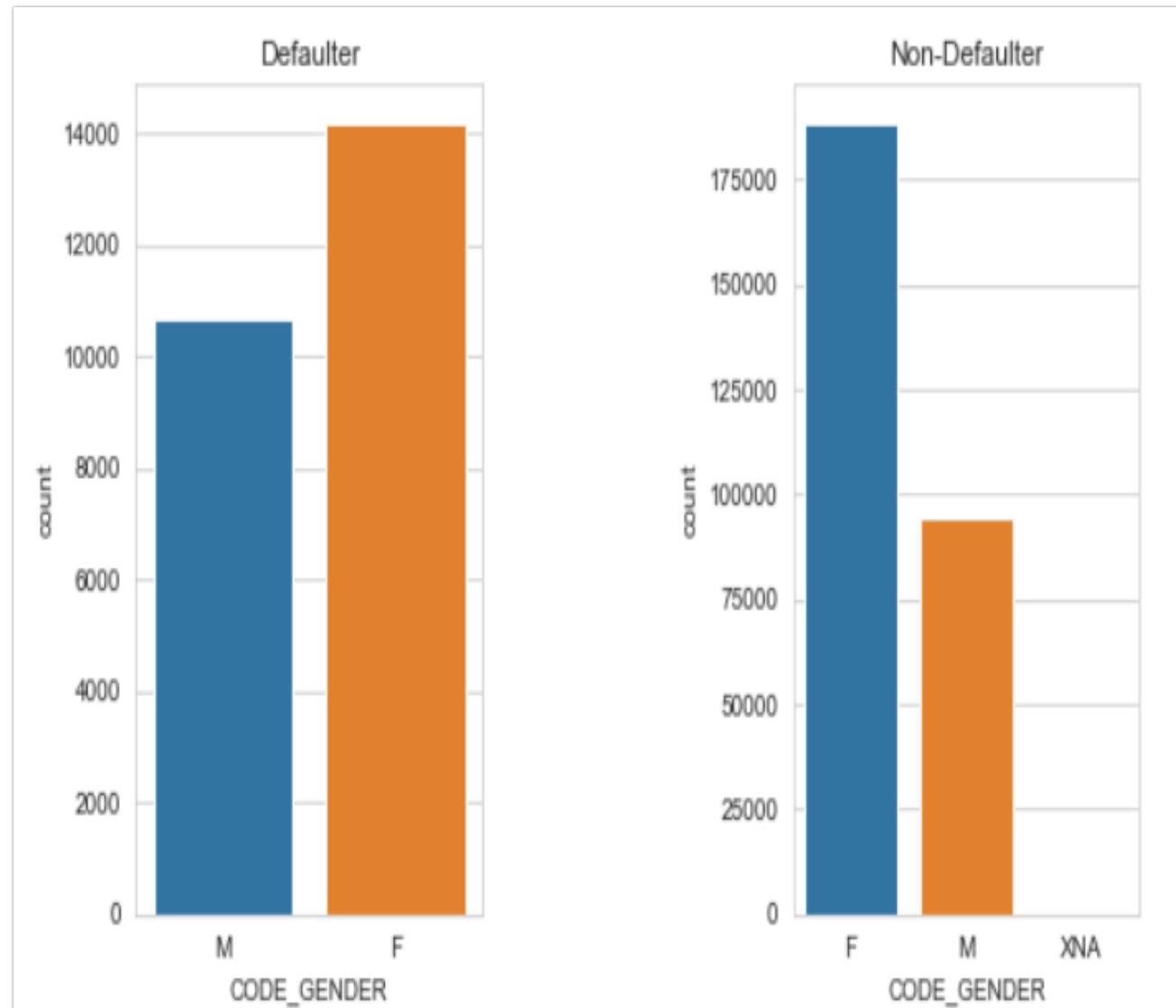
If loan is cash or  
revolving(NAME\_CONT  
RACT\_TYPE) :

- ❖ Revolving loans are just a small fraction of the total number of loans.
- ❖ Less defaulter for revolving loans.



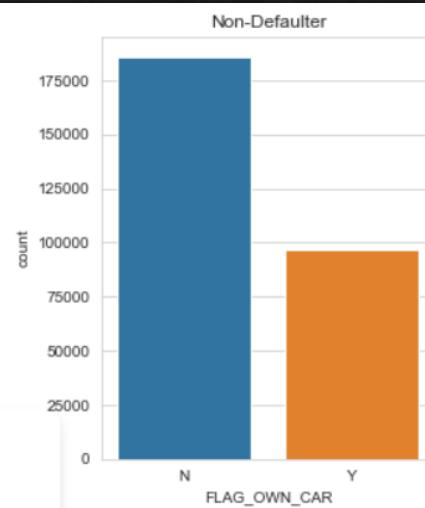
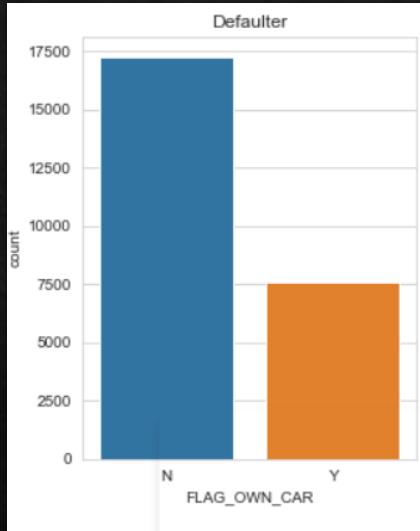
Gender of the client  
shows  
(CODE\_GENDER):

- ❖ More females have been given loan.
- ❖ They are more likely to default.



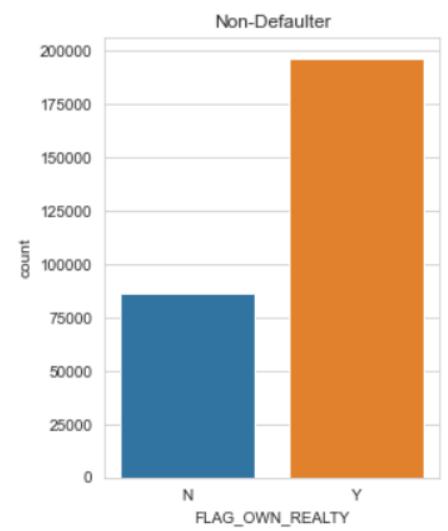
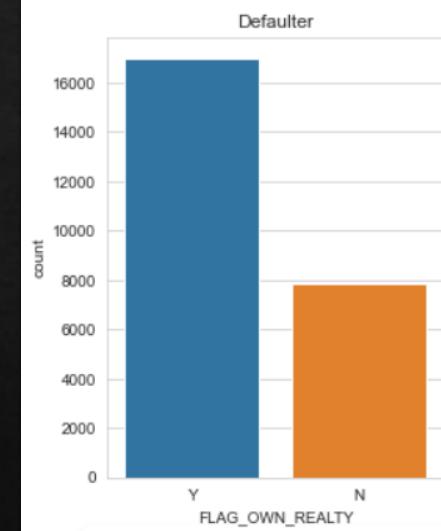
If the client owns a car:

Not owning a car default less.



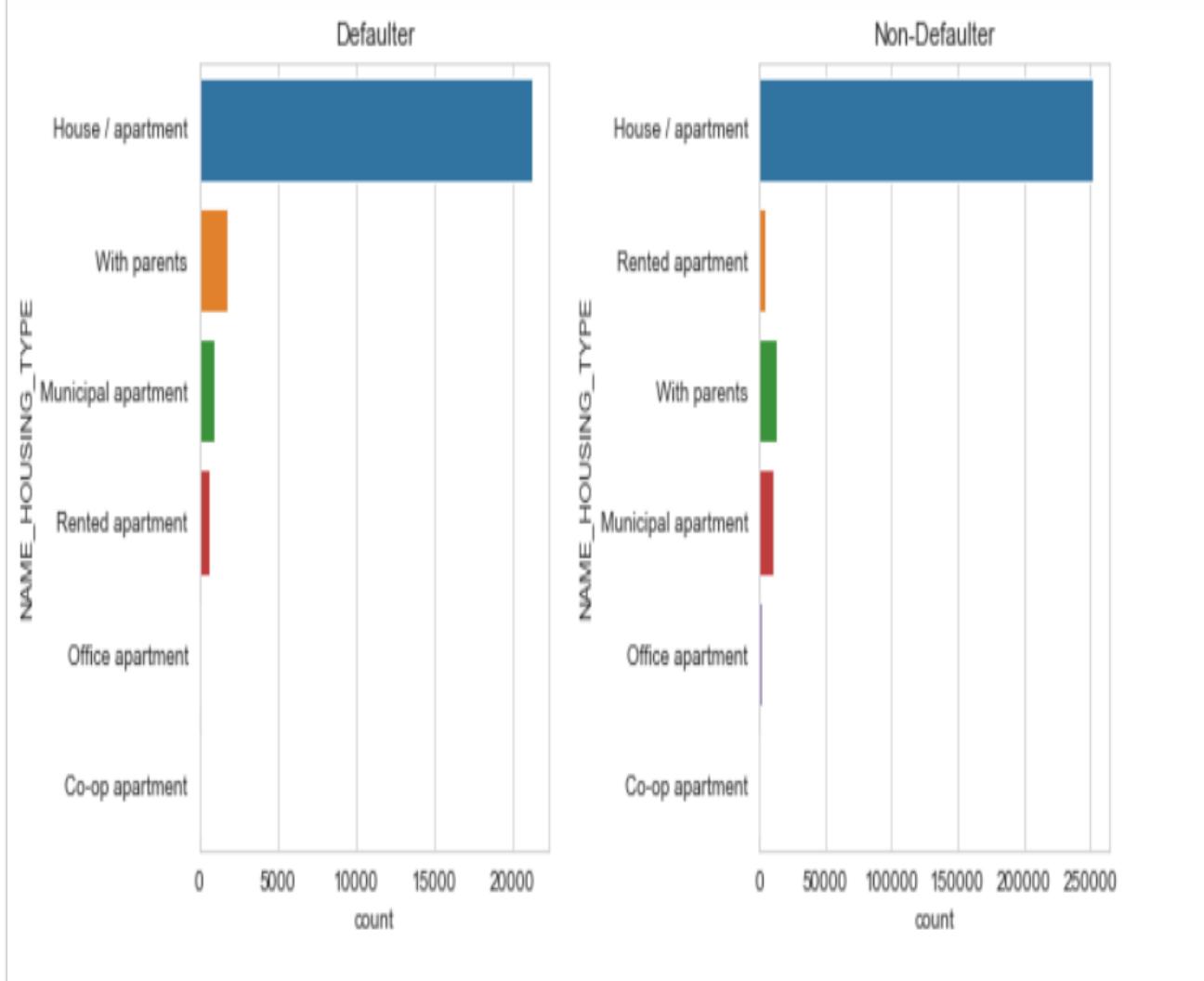
If the client owns a house or flat:

Not owning a car default less.



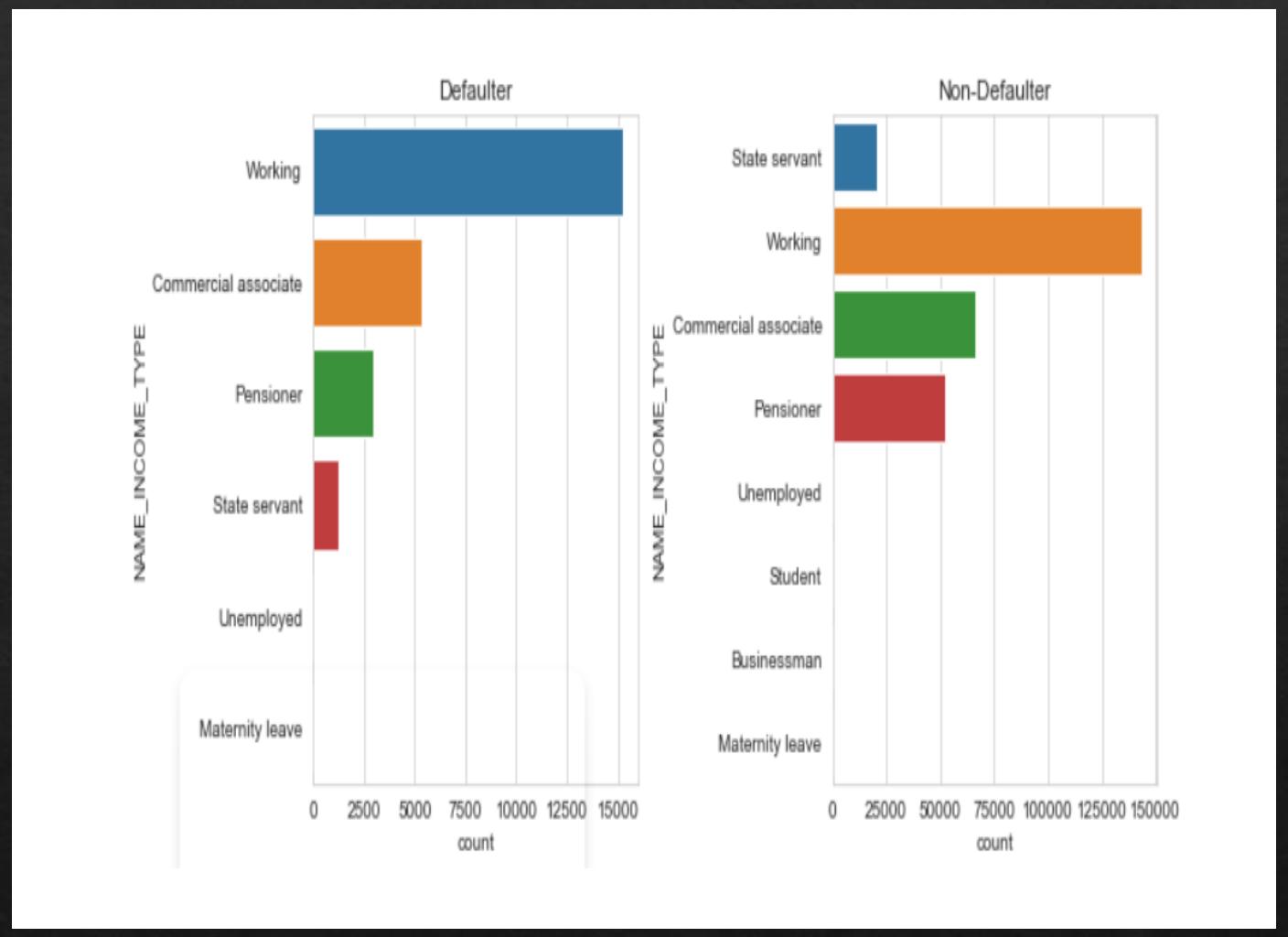
## Housing situation of client:

- ❖ People living in own house default less.



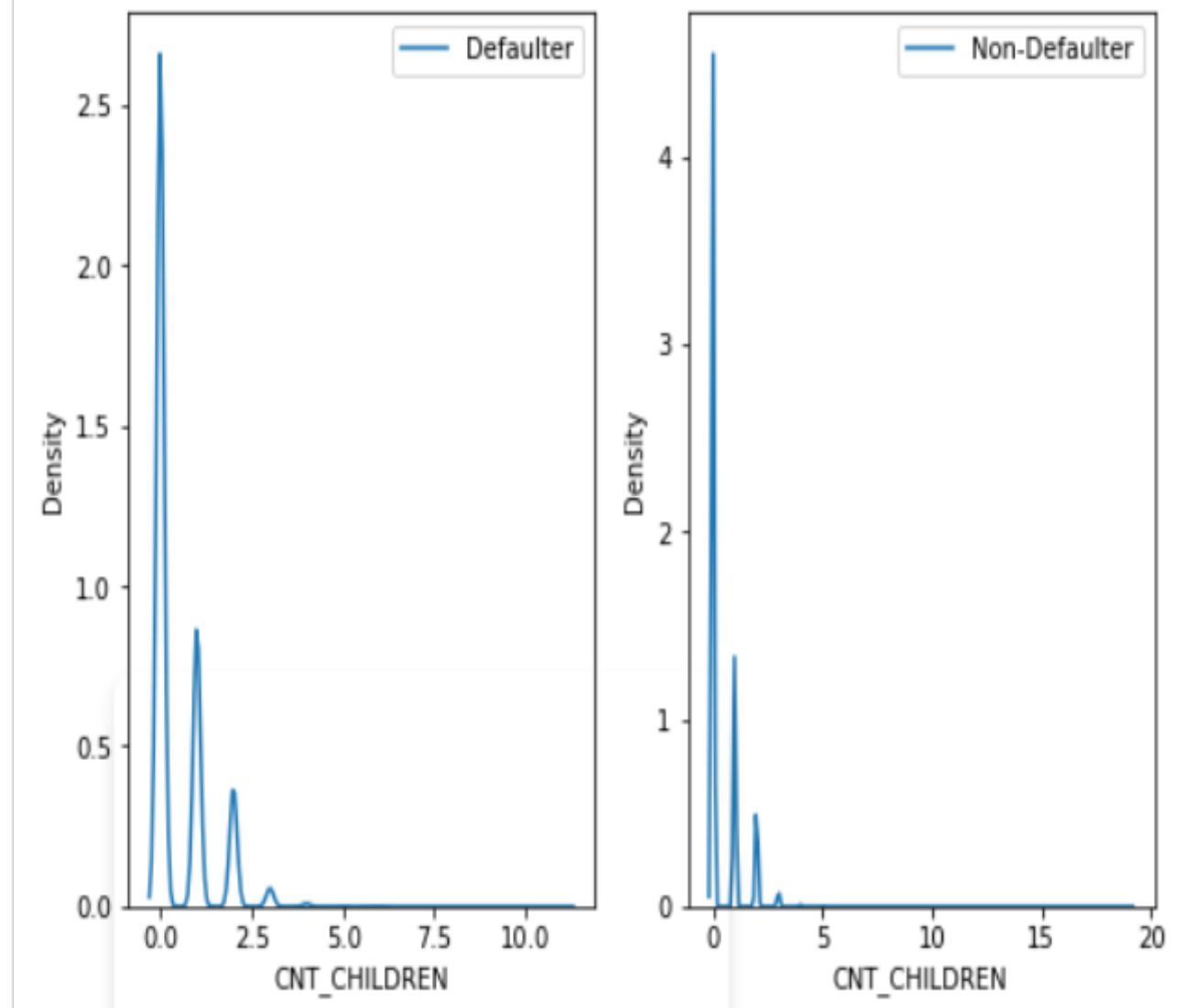
### Client income types:

- ❖ Most of applicants for loans income type is Working, followed by Commercial associate, Pensioner and State servant.
- ❖ Student and Businessmen though less in numbers, do not have default record. Safest two categories for providing loan.



Most of the applicants do not have children:

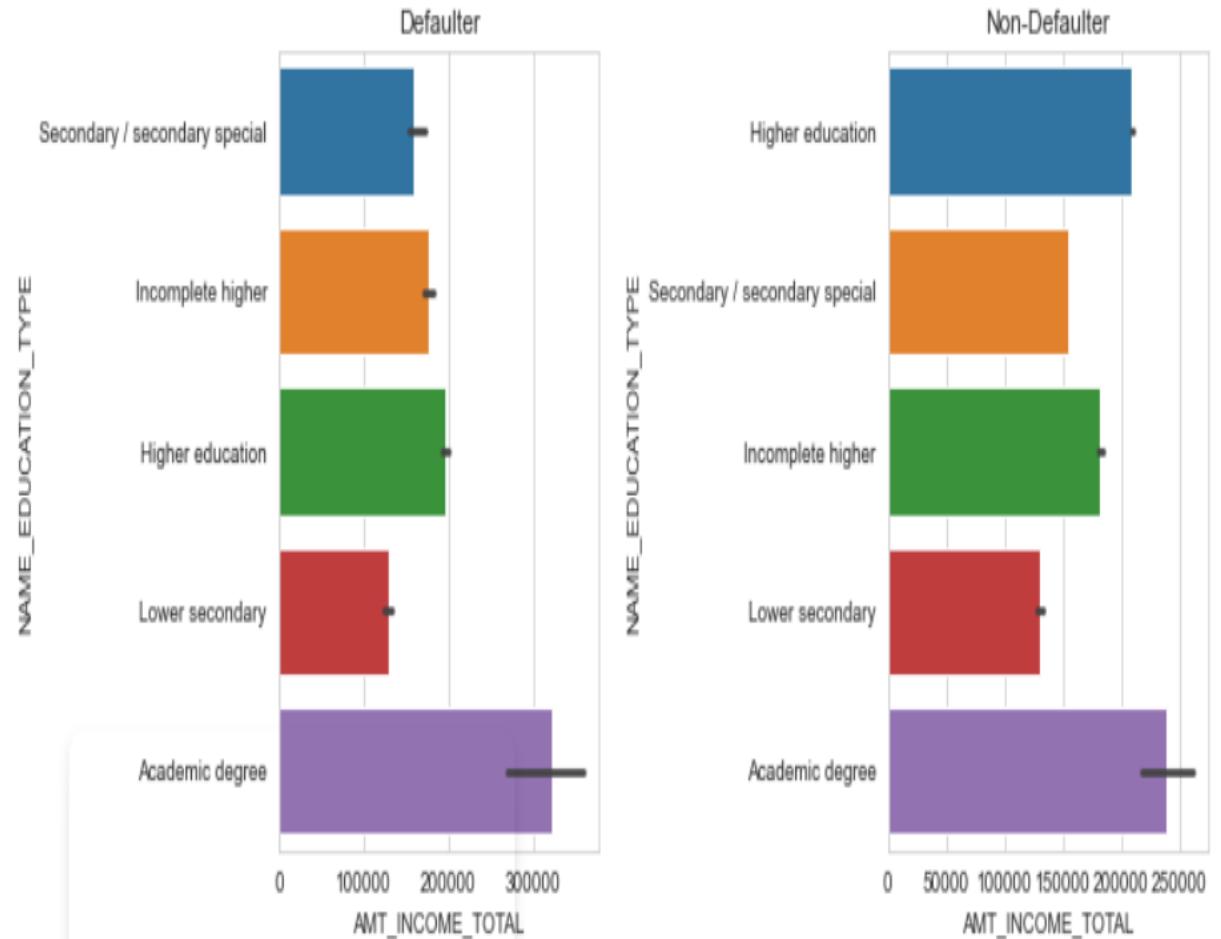
- ❖ Very few clients have more than 3 children.
- ❖ Client who have more than 4 children has a very high default rate with child count 9 and 11 showing 100% default rate



# BIVARIATE ANALYSIS

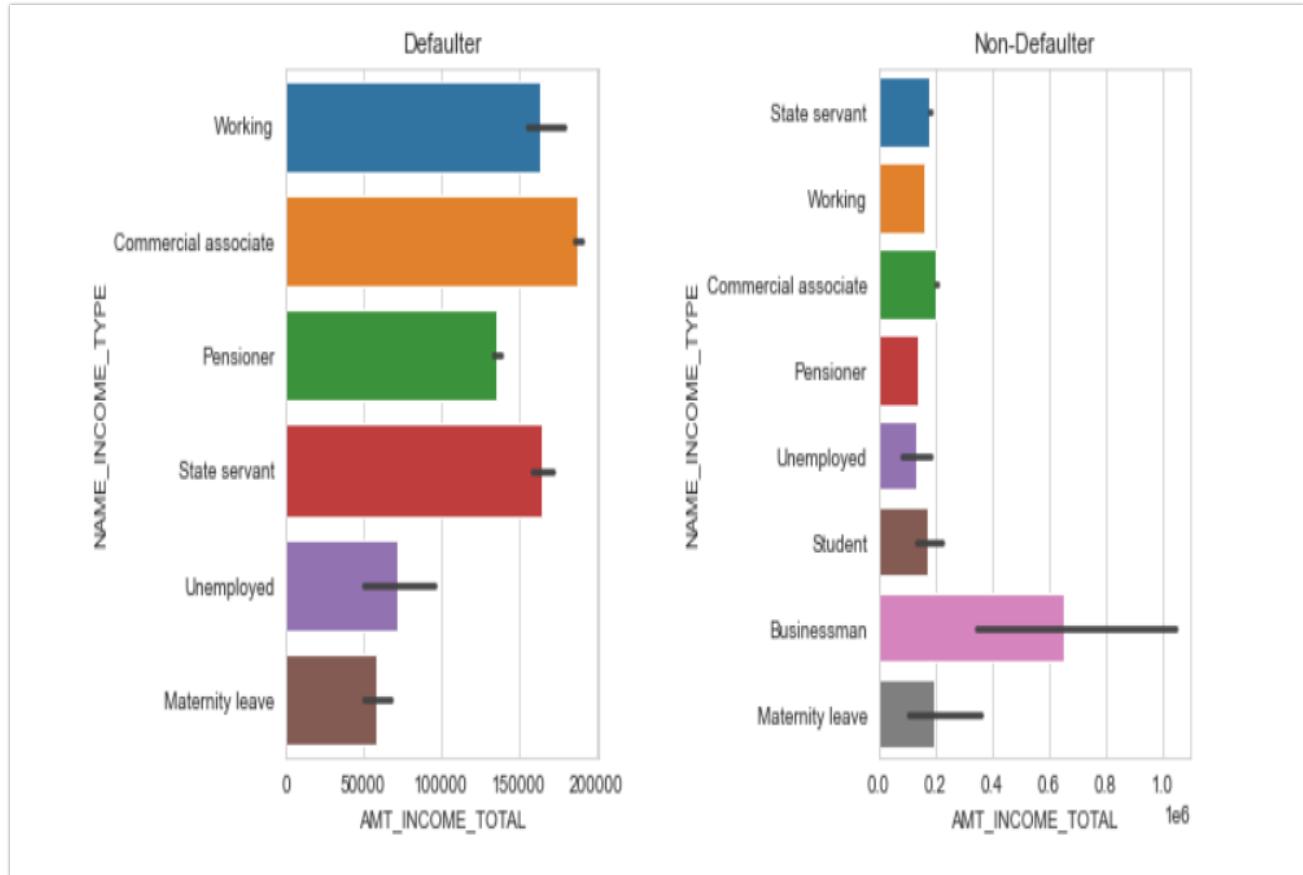
Education type vs Total income:

- ❖ Applicants with Academic Degree and High Income are least likely to default
- ❖ Applicants with Lower Secondary Education and Mid Income Range are very likely to default followed by Lower Secondary - Low Income Range



### Income type vs Total income:

- ❖ Students and Businessmen are highly unlikely to Default on loans
- ❖ Businessmen have a higher credit availability because of high income.
- ❖ Women on Maternity Leave are likely to Default may be because of less payment from the organization they are working



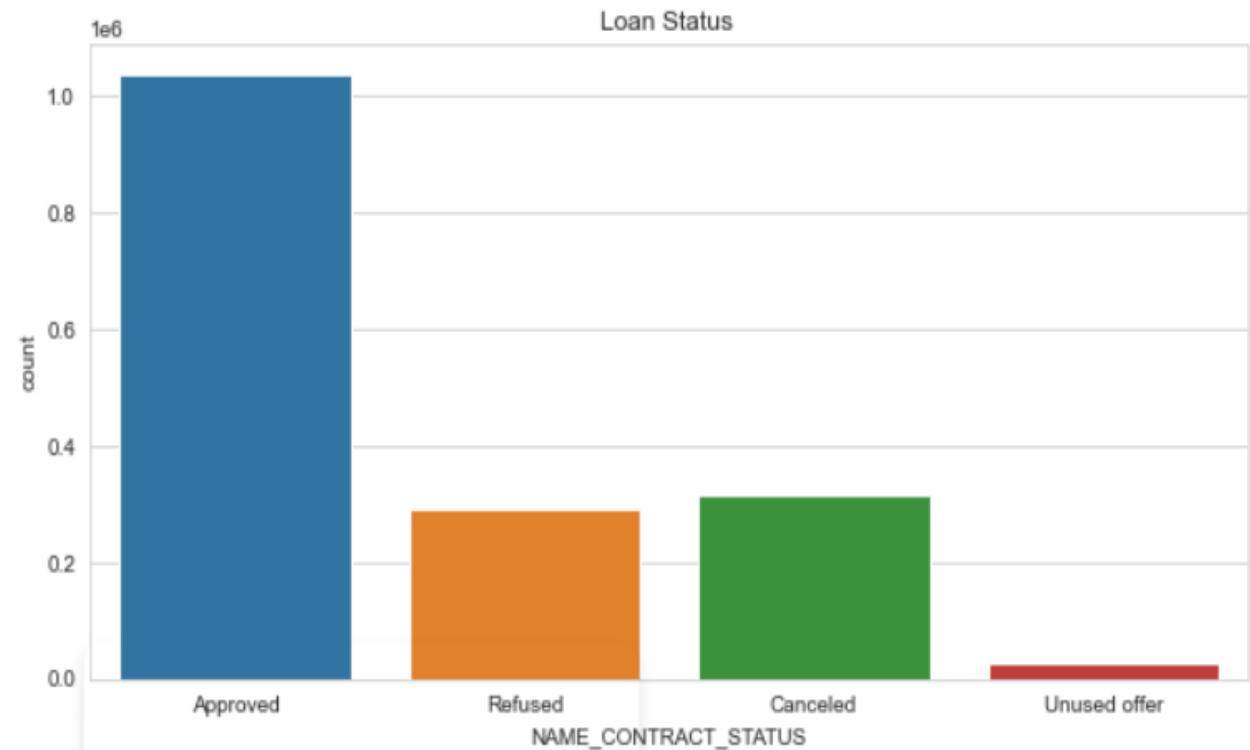
## PREVIOUS APPLICATION DATASET

- ❖ The analysis of this data set is done to know the loan history of client
- ❖ The analysis are-

1)univariate-

Loan status of the client:

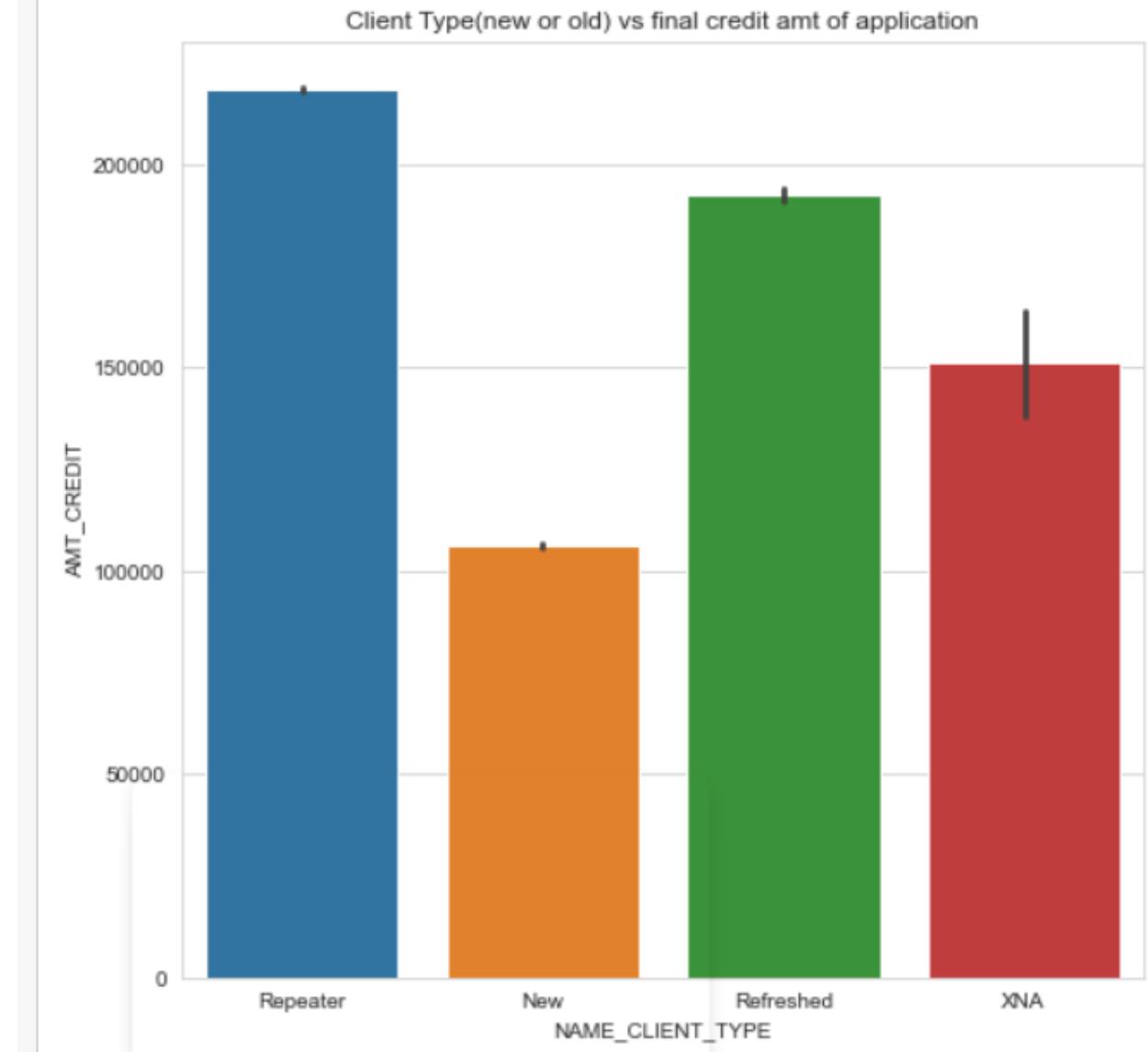
- a)Maximum loans approved
- b)unused offers are very less



## 2)Bivariate-

Client Type(new or old) vs final credit amt of application:

- ❖ Repeater and refreshed has high final credit amount.



# CONCLUSION

EDA for the dataset is done are following are inferred:

- ❖ Defaulters are 8.07. They need to be taken into consideration while giving loan.
- ❖ Bank lends more to female. This means they support gender equality.
- ❖ More cash loan goes into default. So need to minimise it.
- ❖ The previous application dataset should be used for knowing more of the client loan history

Thank you