

# Credit EDA Case Study

by,

**Jobish Jose**



# Business Understanding

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter.
- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.



# Business Understanding

- The data contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:
- **The client with payment difficulties (Defaulters):** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- **All other cases (Non-Defaulters):** All other cases when the payment is paid on time.
- In this case study, EDA is used to understand how consumer attributes and loan attributes influence the tendency to default.



# Business Objectives

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.
- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected.
- Identification of such applicants using EDA is the aim of this case study.
- Analysis is done using Python on a Jupyter notebook.

# Analysis Approach

- Sourcing and Understanding the data
- Cleaning the data
- Handling data imbalances and outliers
- Analysing data using Univariate, Segmented  
Univariate, Bivariate and Correlation analysis
- Insights



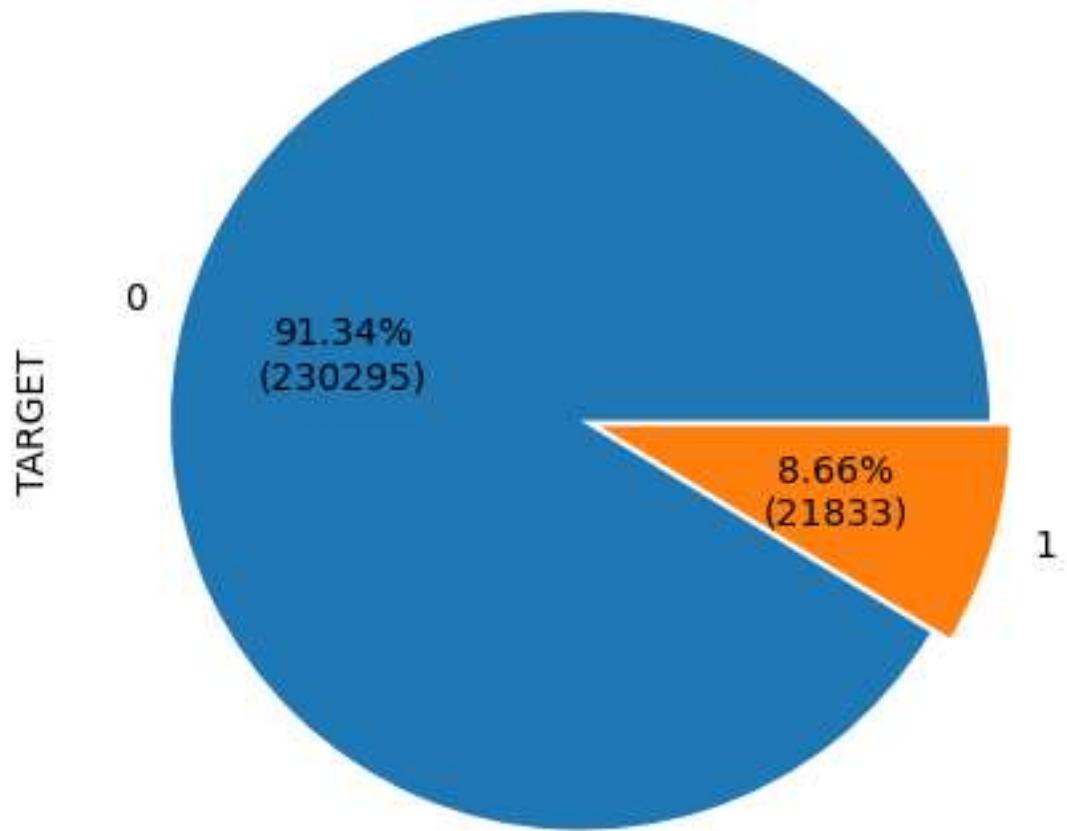
# Proportion of Defaulters and Non - Defaulters

Non – Defaulters : 91.34%

Defaulters : 8.66%

**The imbalance ratio is 10.55**

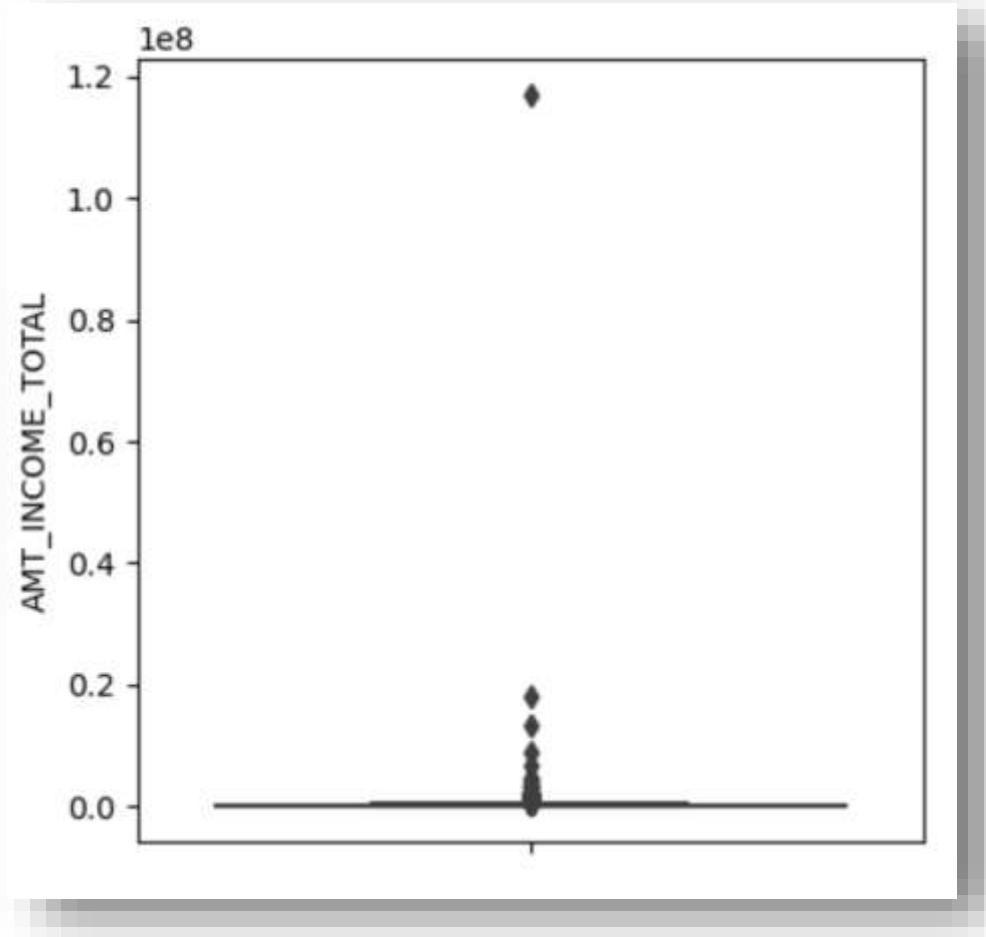
Imbalance between target0 and target1



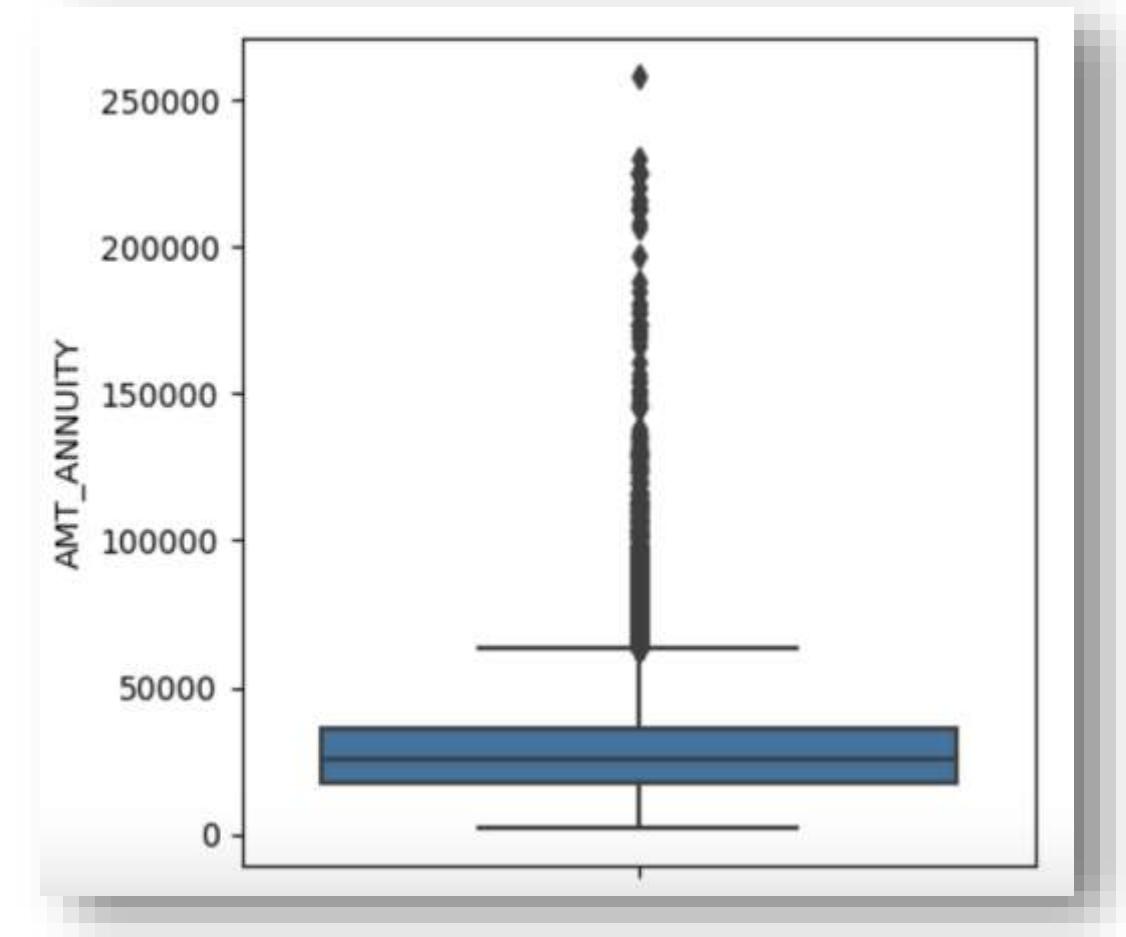
Target value 1 - Clients with payment difficulties (Defaulters)

Target value 0 – All other cases (Non– Defaulters)

# Outliers in Income Amount and Amount Annuity

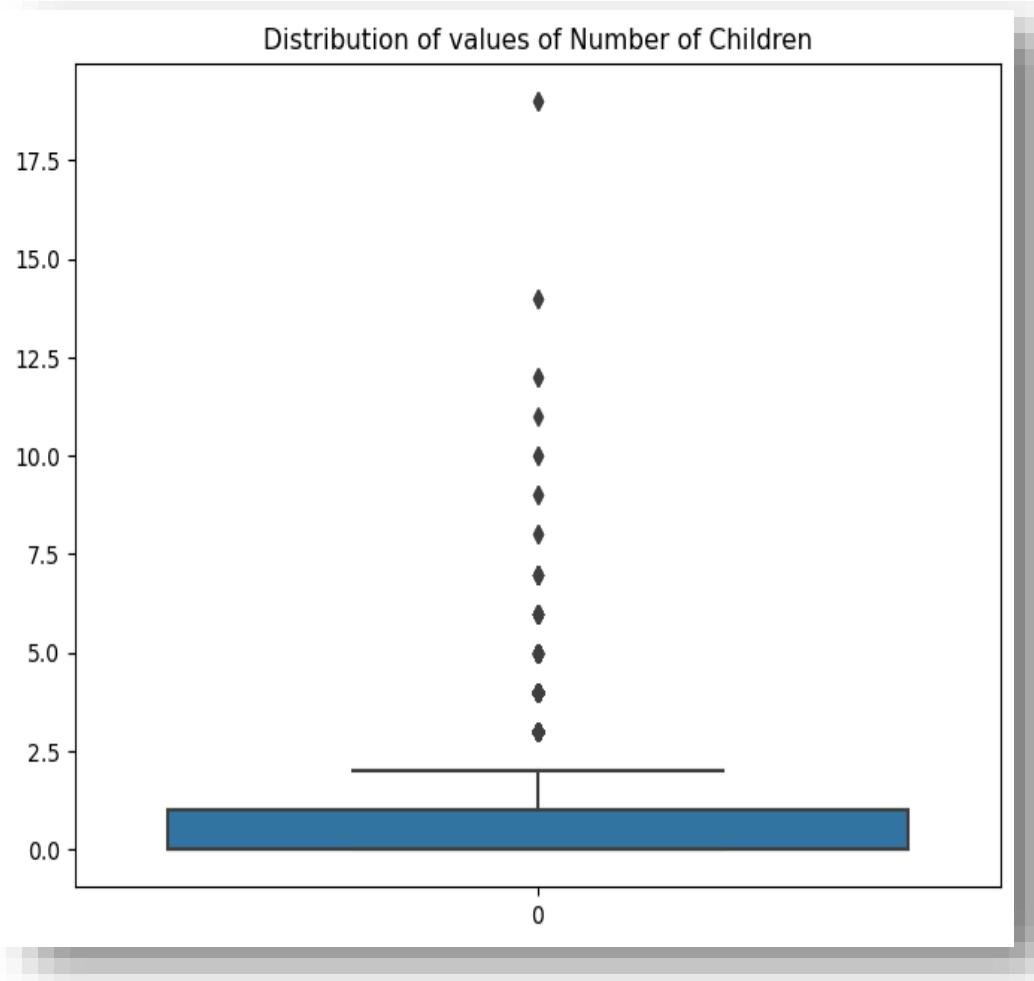


There is an outlier in the range of 1.0 to 1.2.

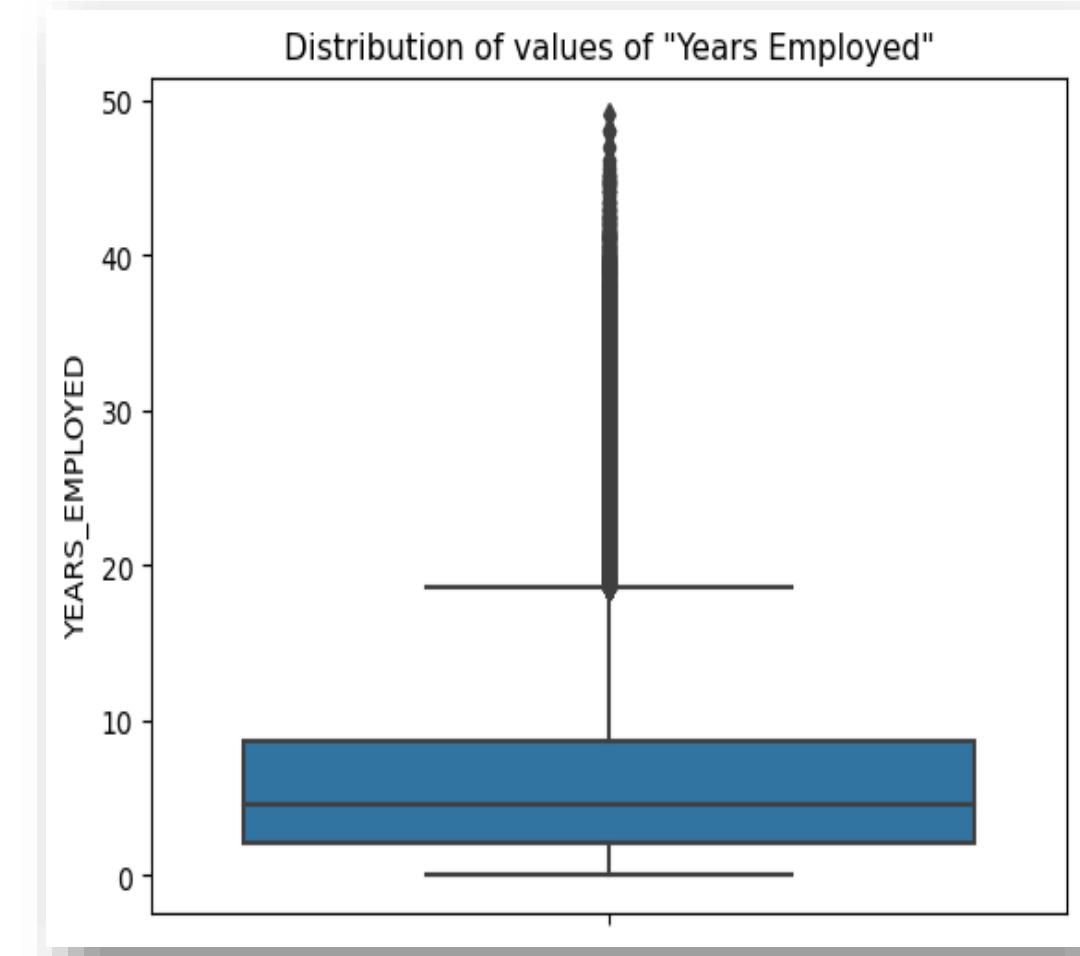


There are a lot of value points above the Max value of the box plot. The value point just above the 250K mark is clearly an outlier.

# Outliers in Number of Children and Years Employed



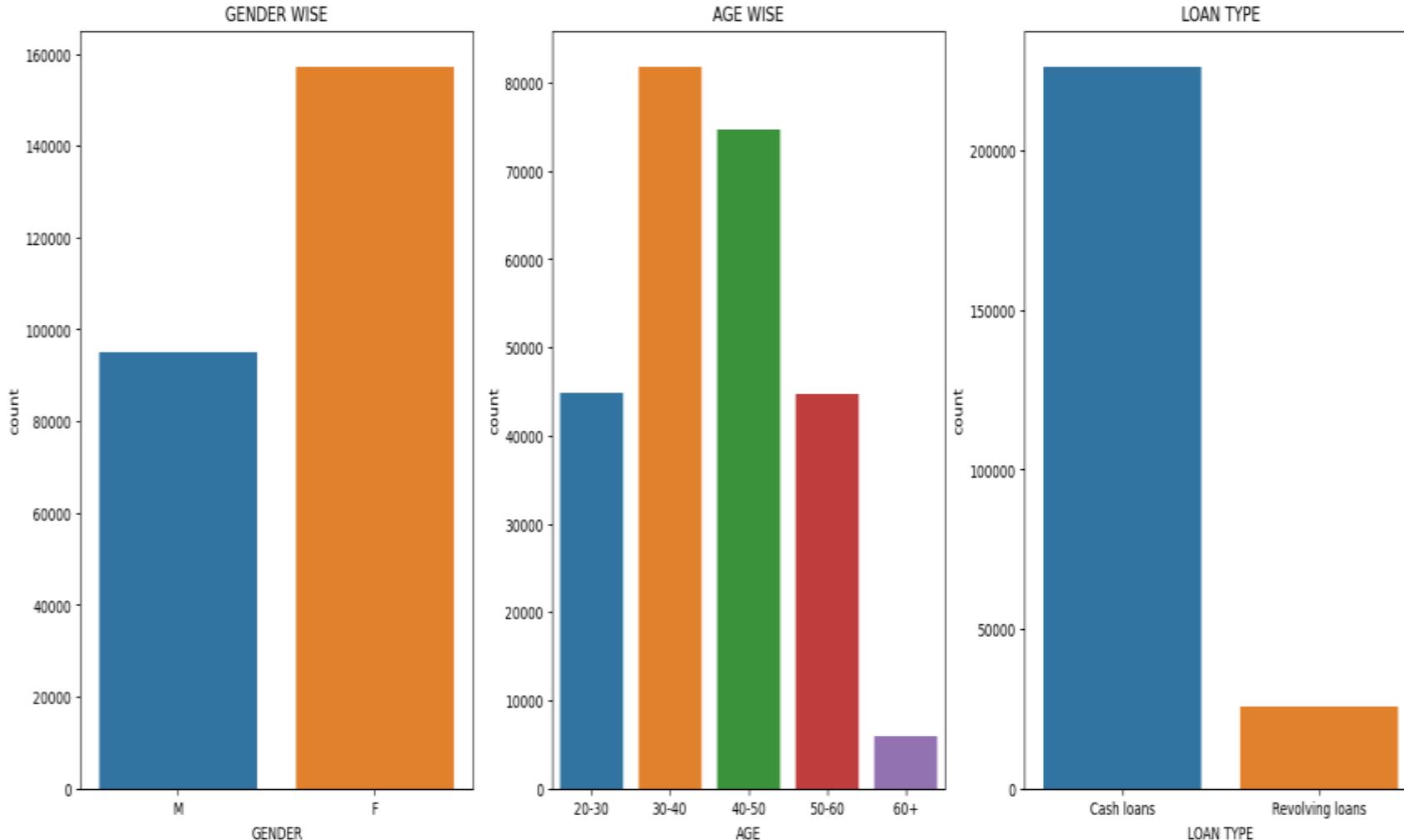
There are outliers in the "CNT\_CHILDREN" column.



Even though there are a lot of value points above the Max value of the box plot, they are genuine values and cannot be considered as outliers.

# Univariate Analysis

## Univariate Analysis of Gender, Age, Loan Type



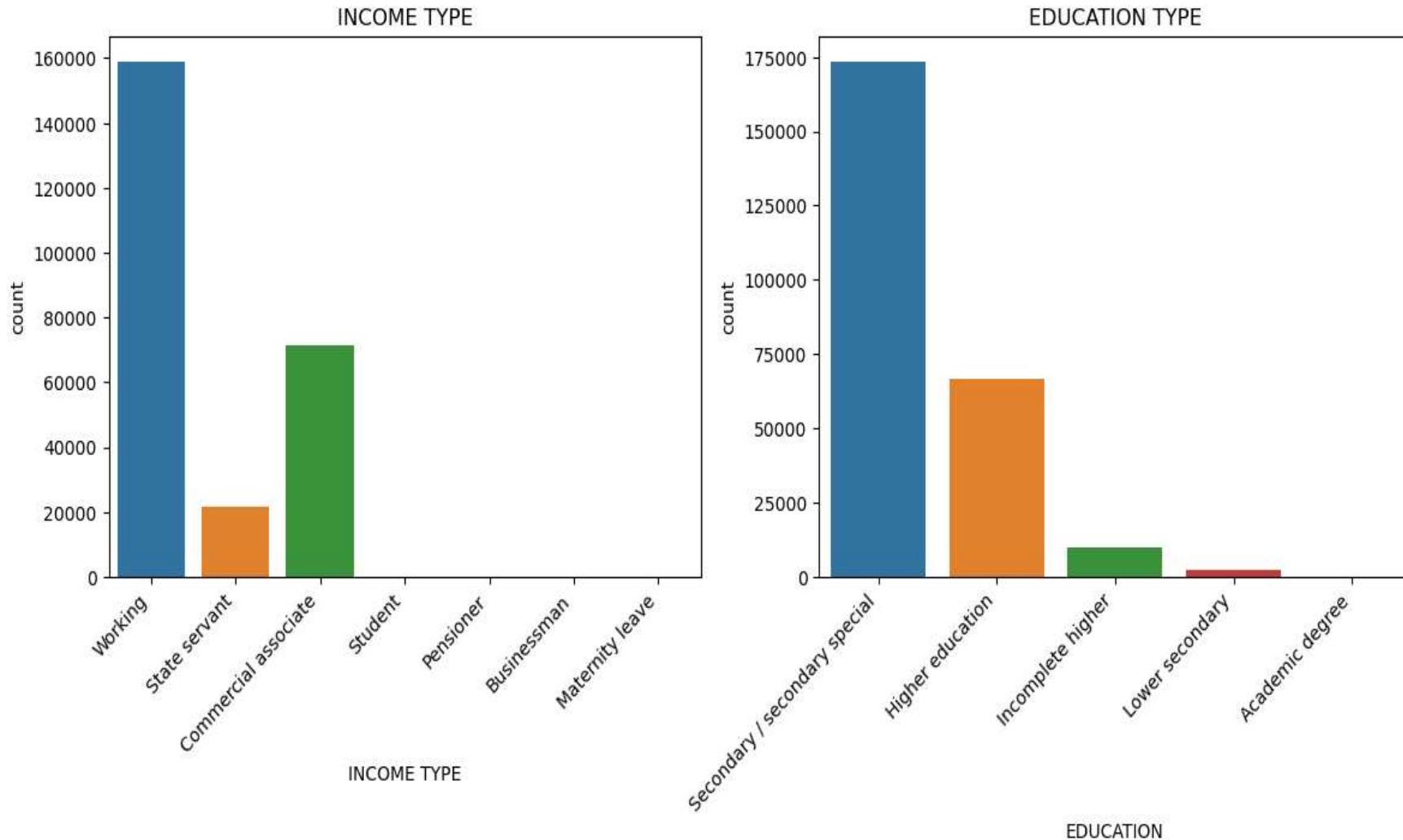
- Bank received more loan applications from female clients compared to the number of applications from male clients.
- Majority of the applications are from 30-40 age group closely followed by 40-50.
- Most of the clients applied for Cash loans; Number of people who applied for Revolving loans are comparatively too low.

# Univariate Analysis

## Univariate Analysis of Income Type and Education Type

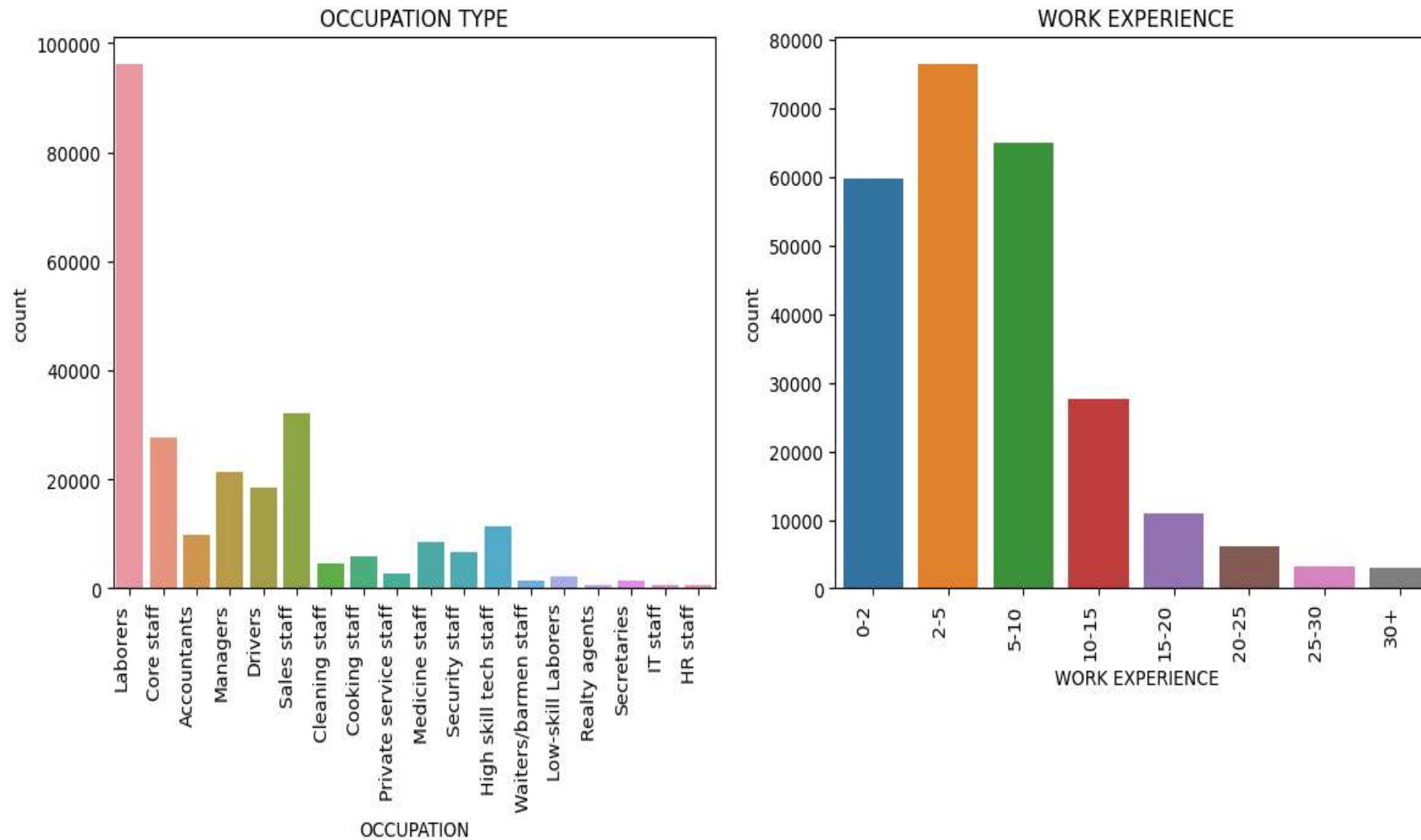
Out of the total applicants, majority of the clients are Working class followed by Commercial associate and State servant respectively.

Education type of most number of clients who applied for the loan is Secondary/Secondary Special followed by Higher Education.



# Univariate Analysis

## Univariate Analysis of Occupation Type and Work Experience



- Majority of the loan applicants are Laborers.
- Most of the applicants are with a work experience of 2-5 years closely followed by people with 5-10 years of experience.

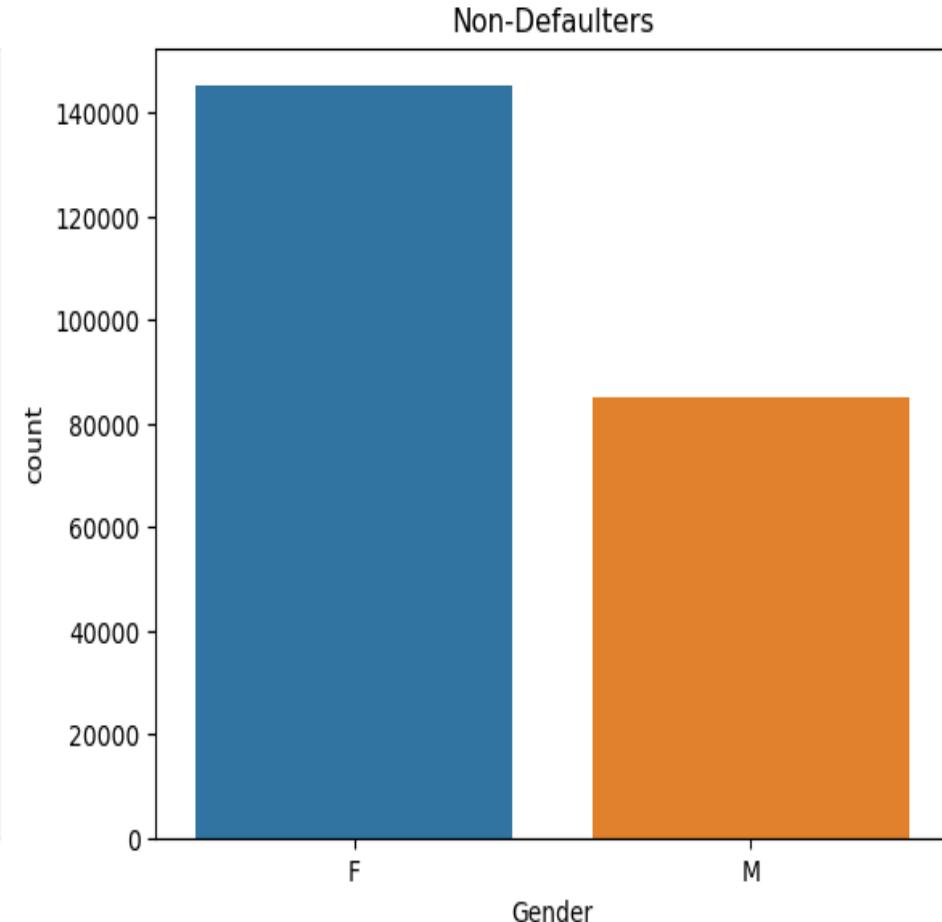
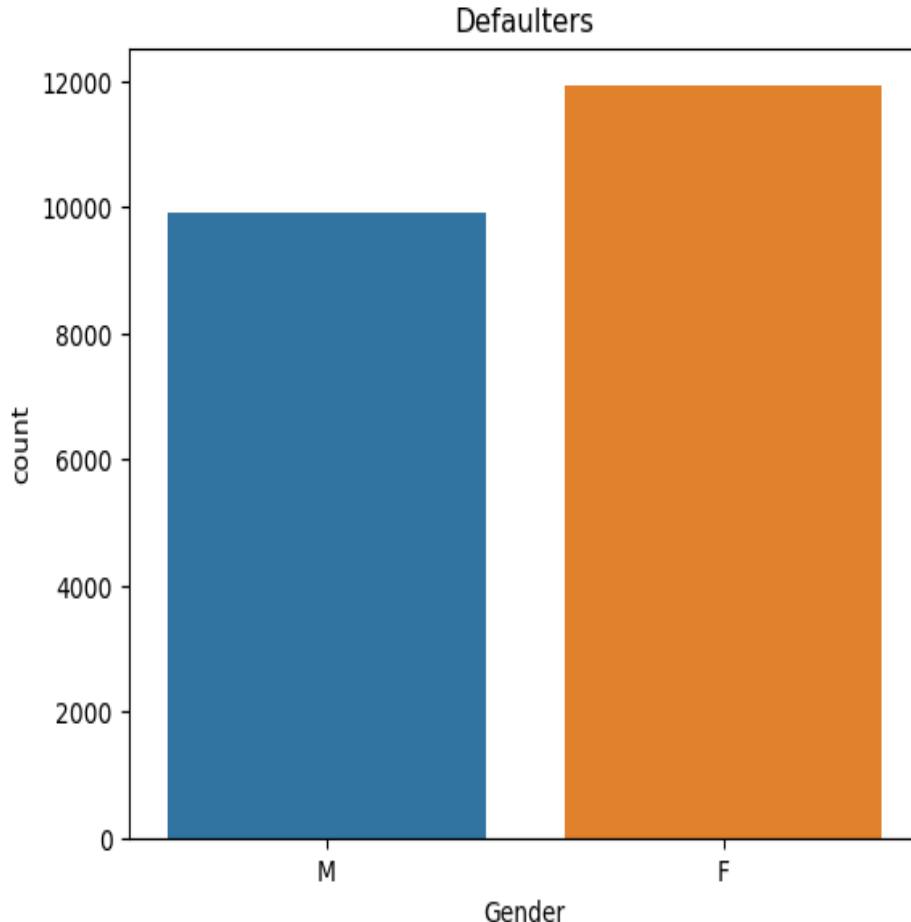
# Univariate Analysis

*for Target 1 and Target 0 (Defaulters and Non – Defaulters)*

## Univariate Analysis of Gender

Majority of  
the Defaulters  
are Female.

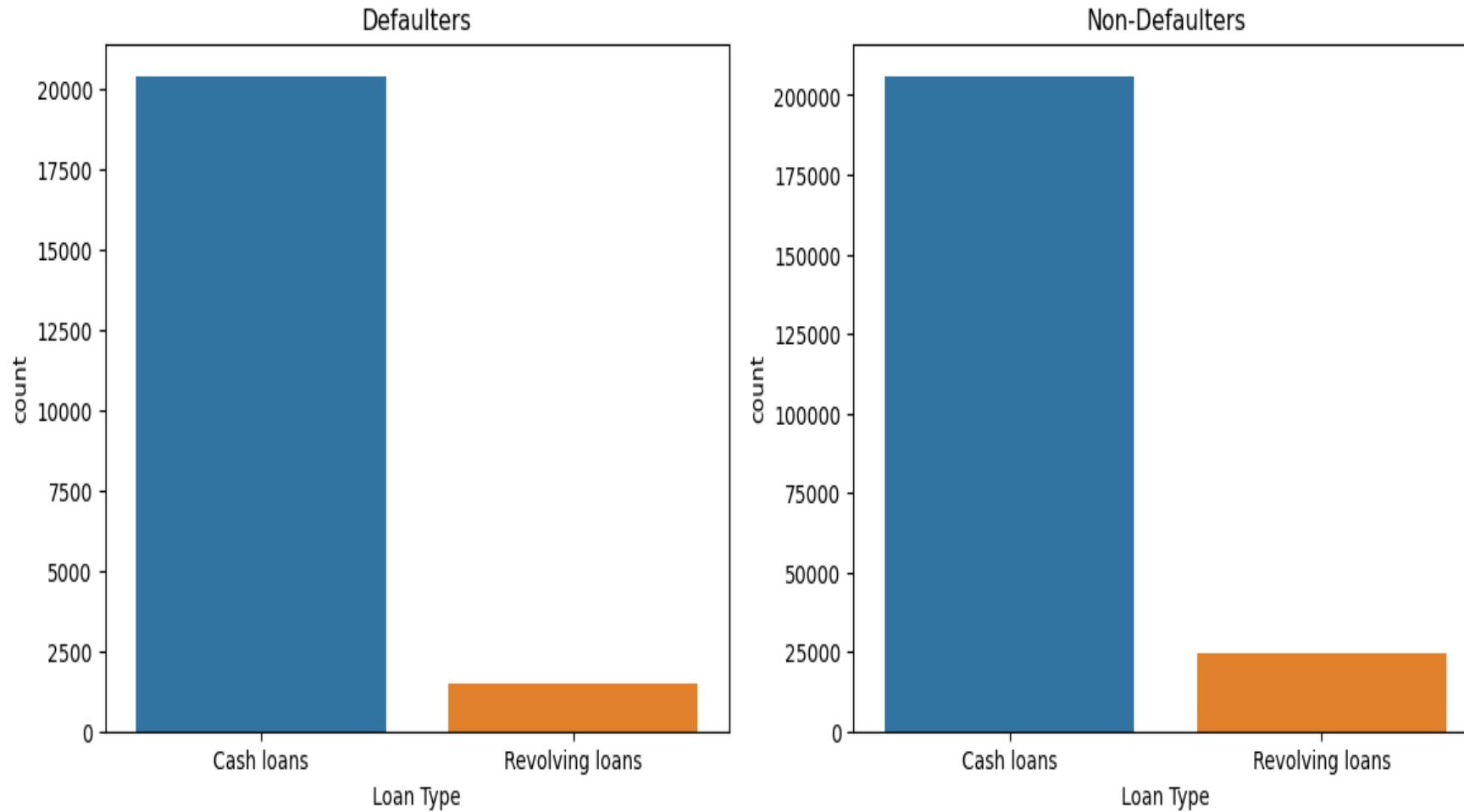
Non –  
Defaulters also  
follow the  
same trend.



# Univariate Analysis

for Target 1 and Target 0 (Defaulters and Non – Defaulters)

## Univariate Analysis of Loan Type(Contract Type)



Number of Revolving Loans are very less compared to Cash Loans.

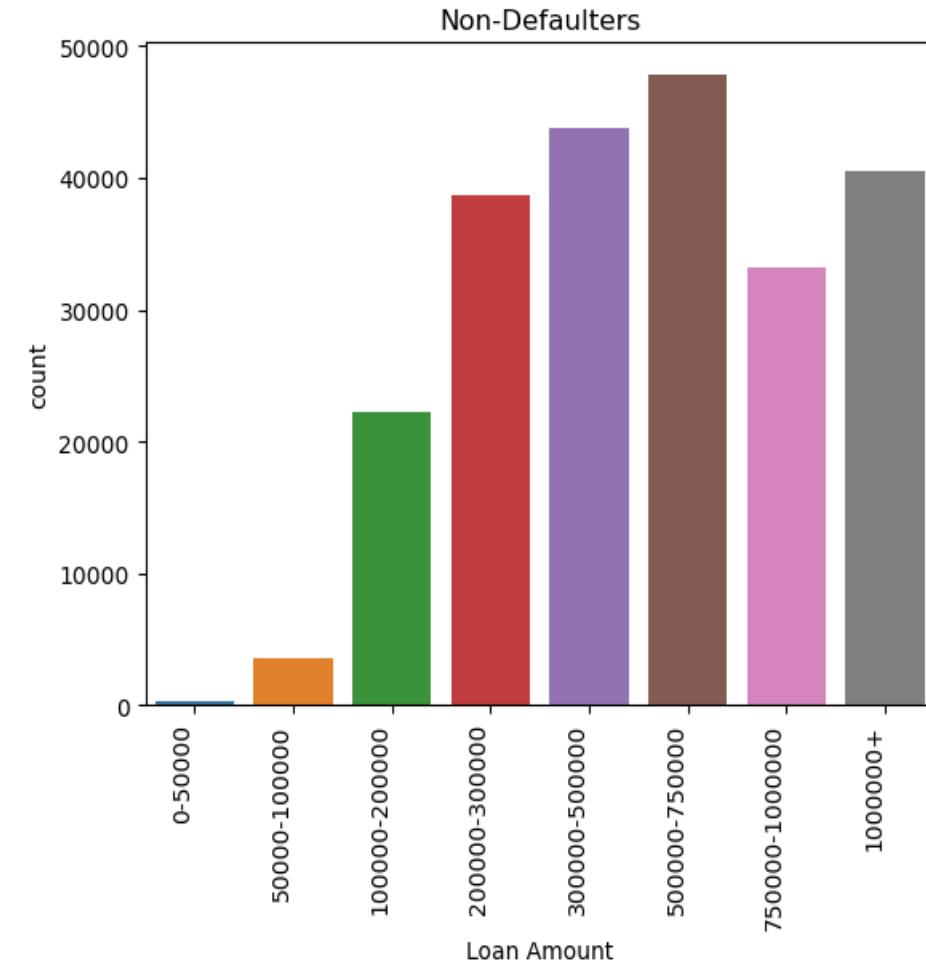
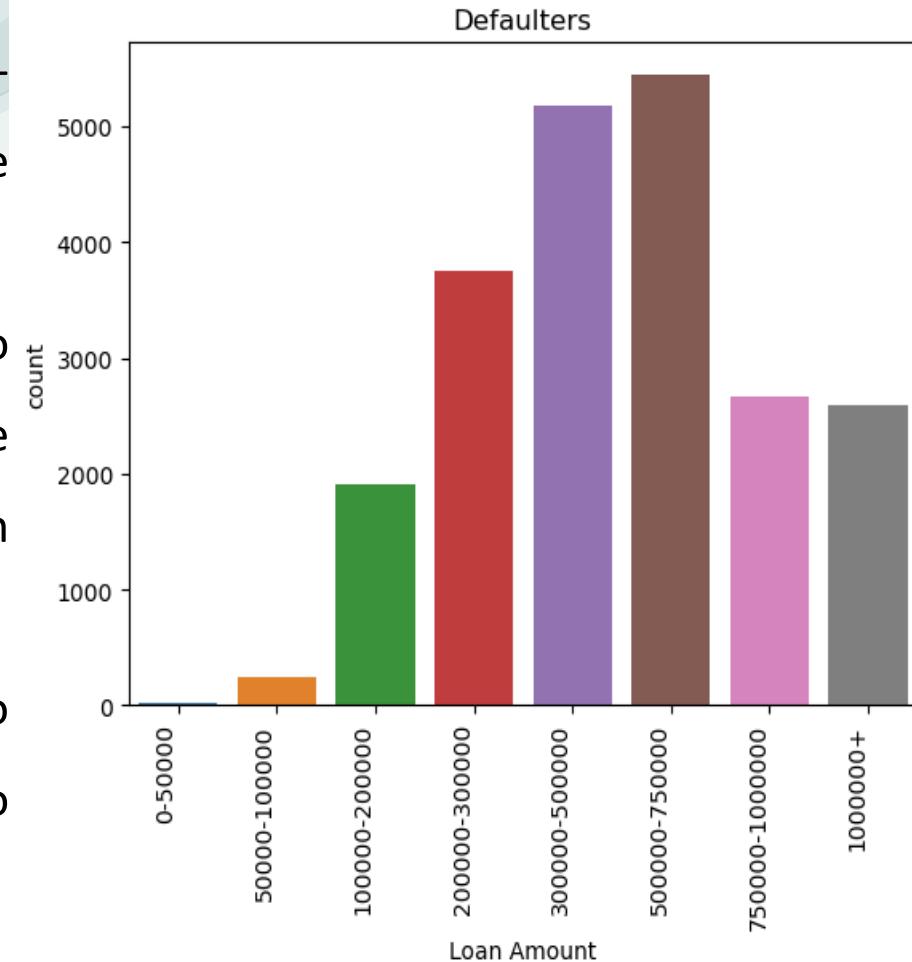
Majority of the clients applied for Cash Loans in both the Defaulters and Non- Defaulters category.

# Univariate Analysis

for Target 1 and Target 0 (Defaulters and Non – Defaulters)

## Univariate Analysis of Loan Amount(Credit Range)

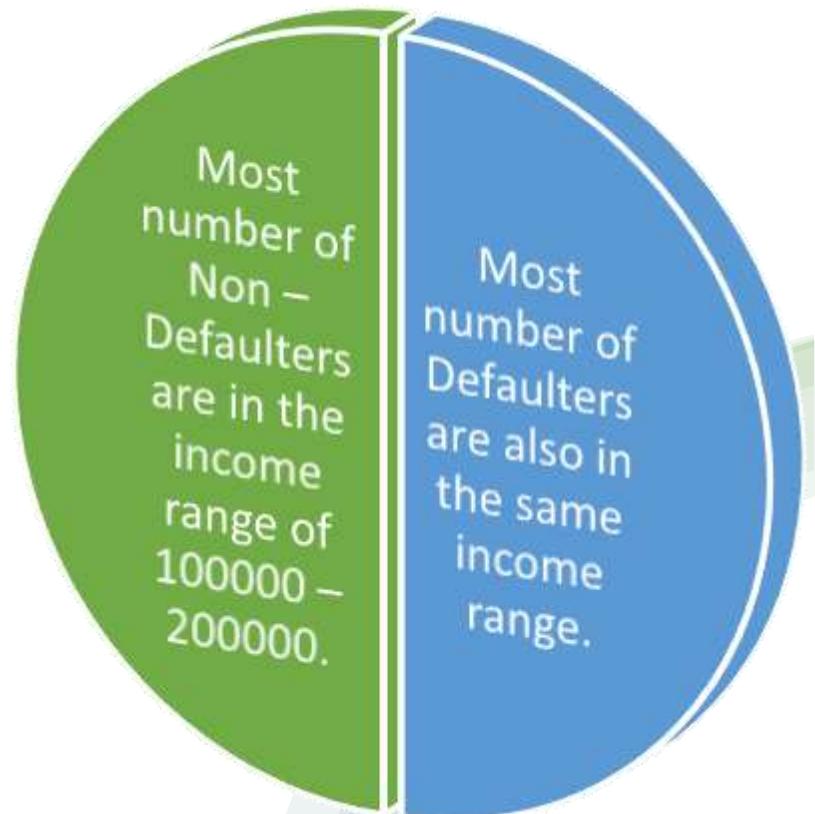
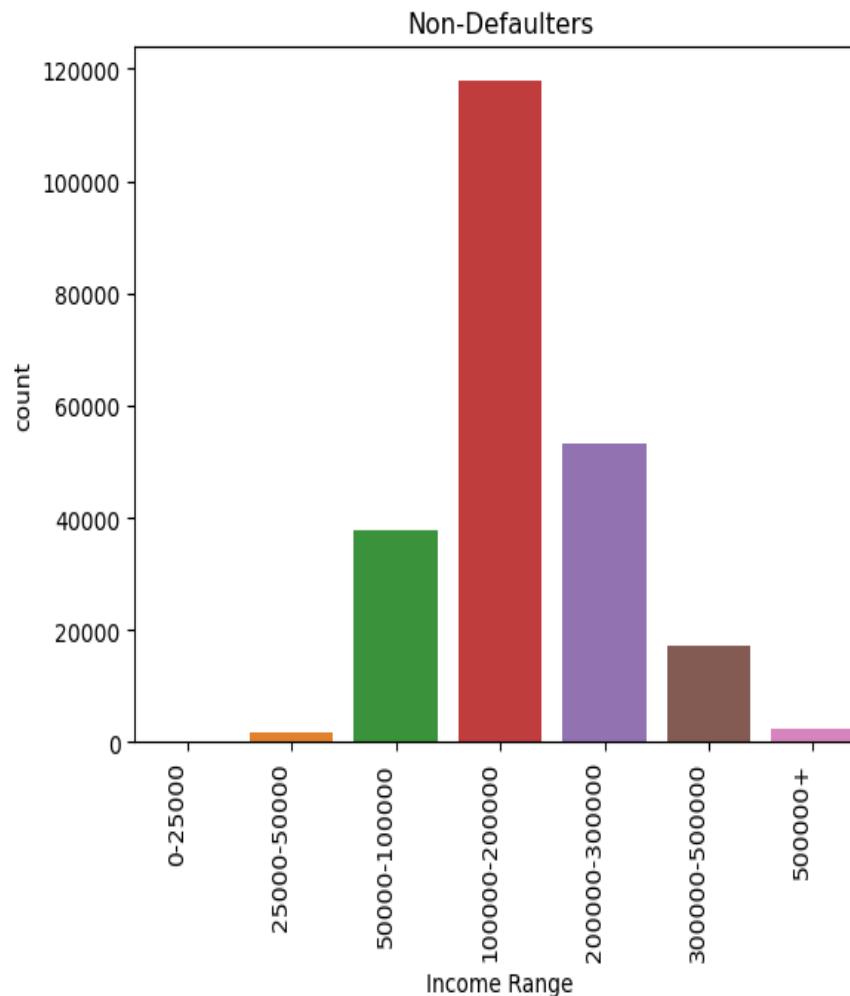
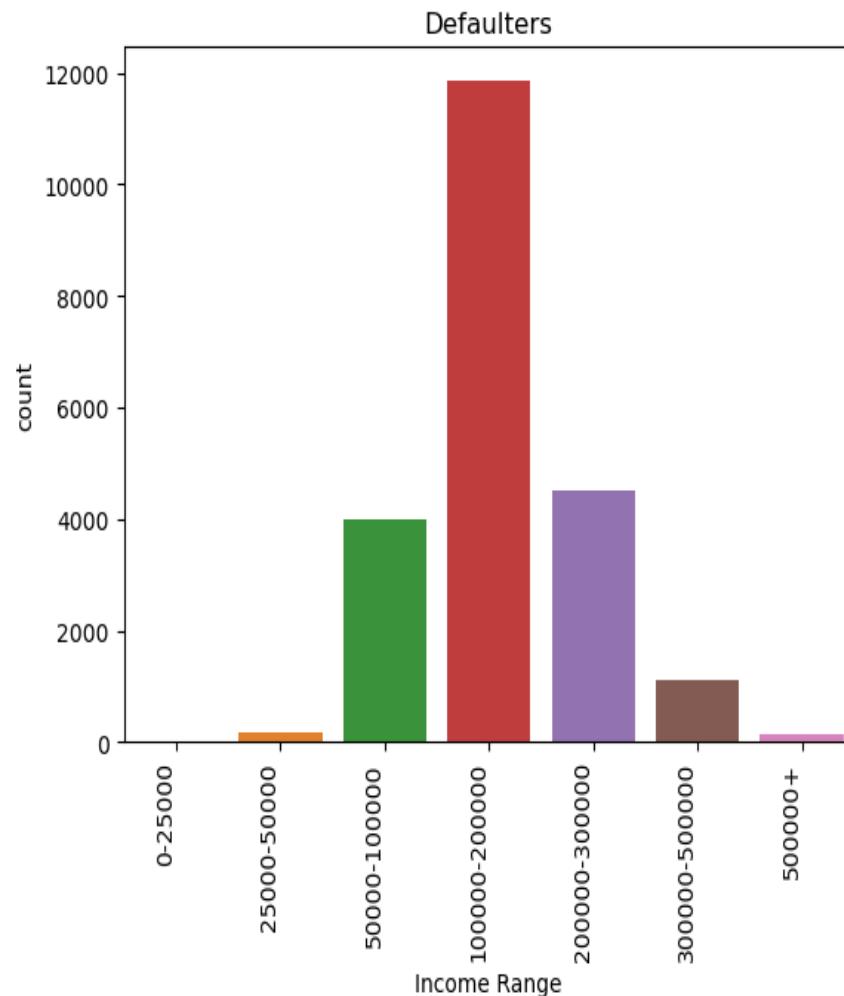
- Both the Defaulters and Non-defaulters follow the same trend.
- Most number of clients who repay the loan on time availed the loan amount in the range of 500000-750000.
- Majority of people who defaulted the repayment also fall in the same range.



# Univariate Analysis

for Target 1 and Target 0 (Defaulters and Non – Defaulters)

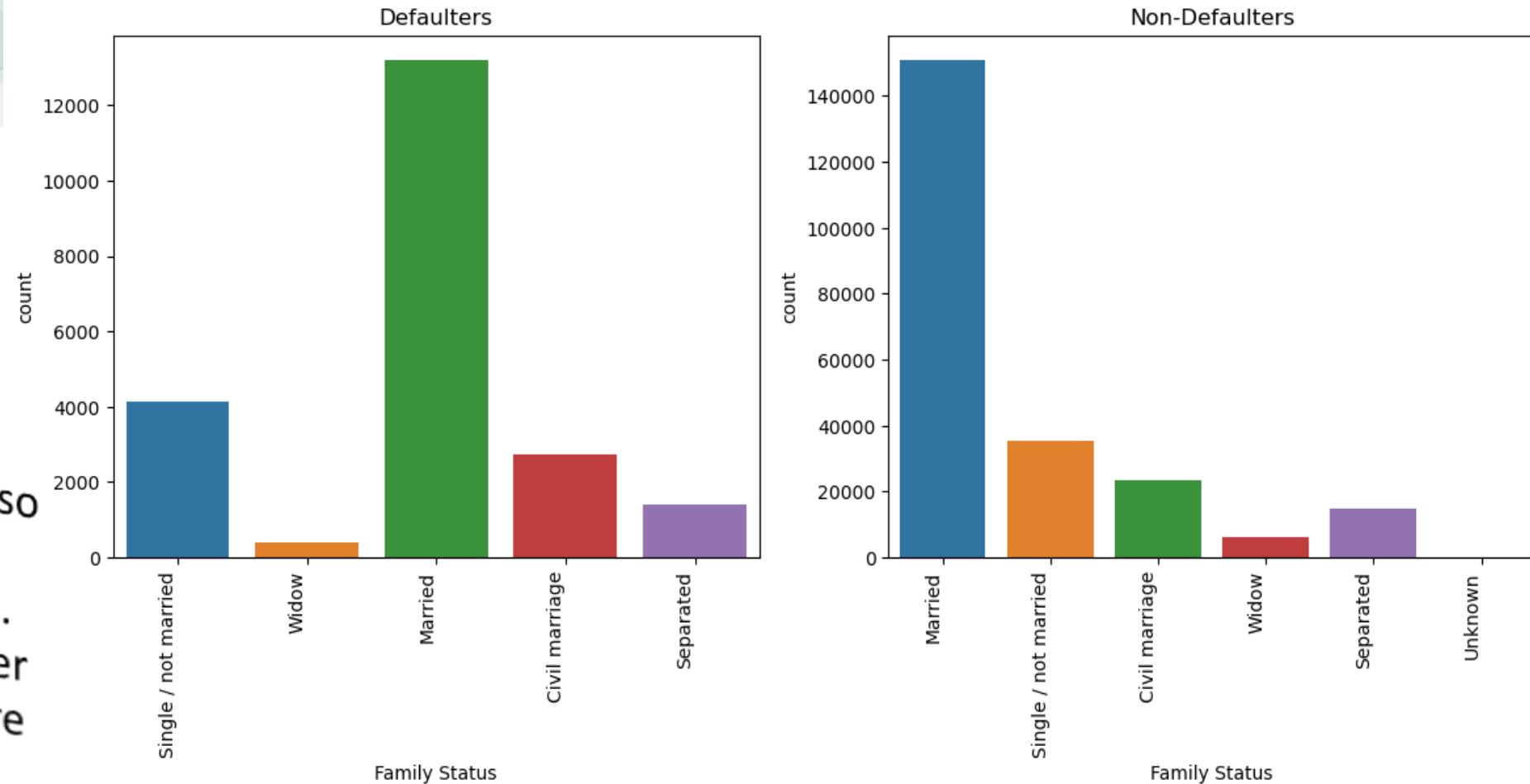
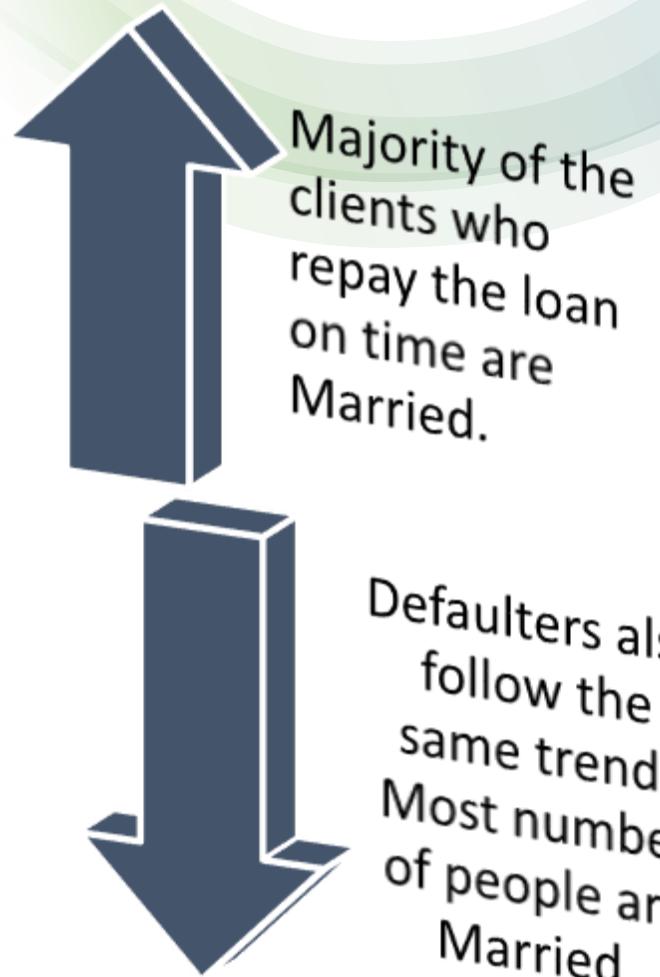
## Univariate Analysis of Income Range



# Univariate Analysis

*for Target 1 and Target 0 (Defaulters and Non – Defaulters)*

## Univariate Analysis of Family Status

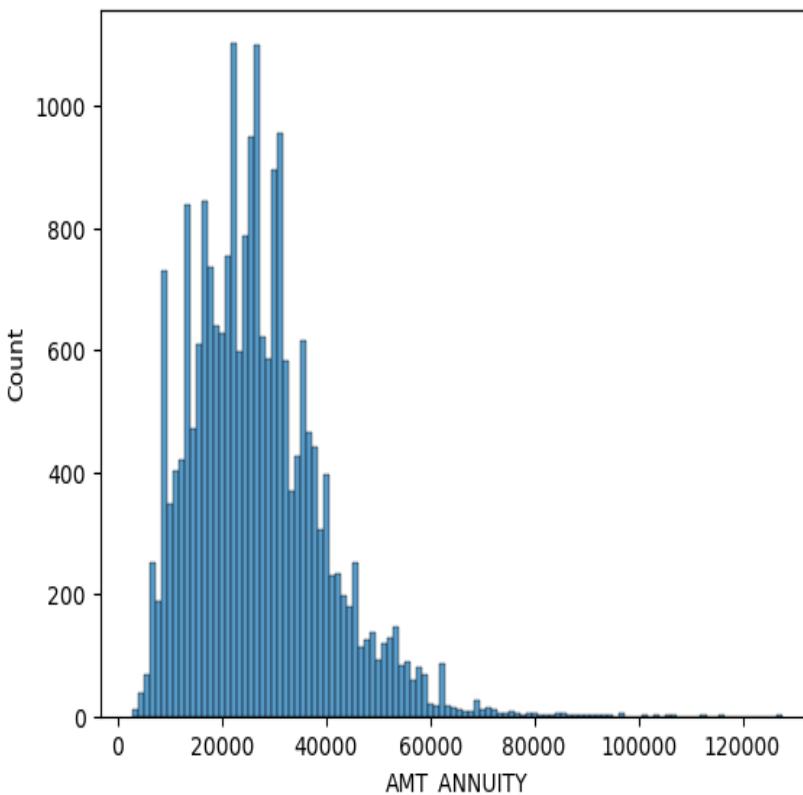


# Univariate Analysis

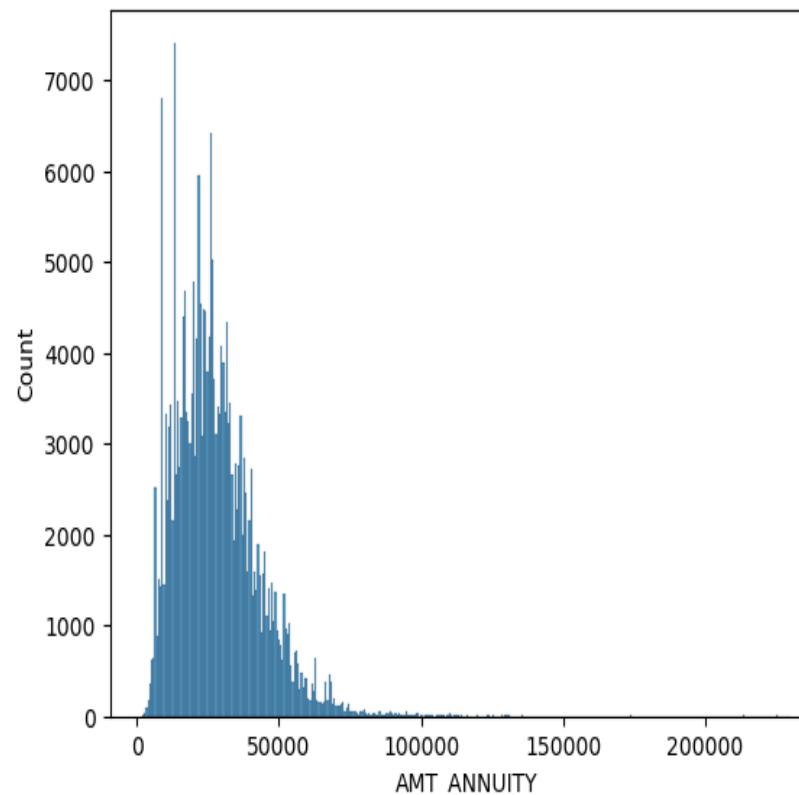
for Target 1 and Target 0 (Defaulters and Non – Defaulters)

## Univariate Analysis of Annuity Amount

Defaulters



Non Defaulters



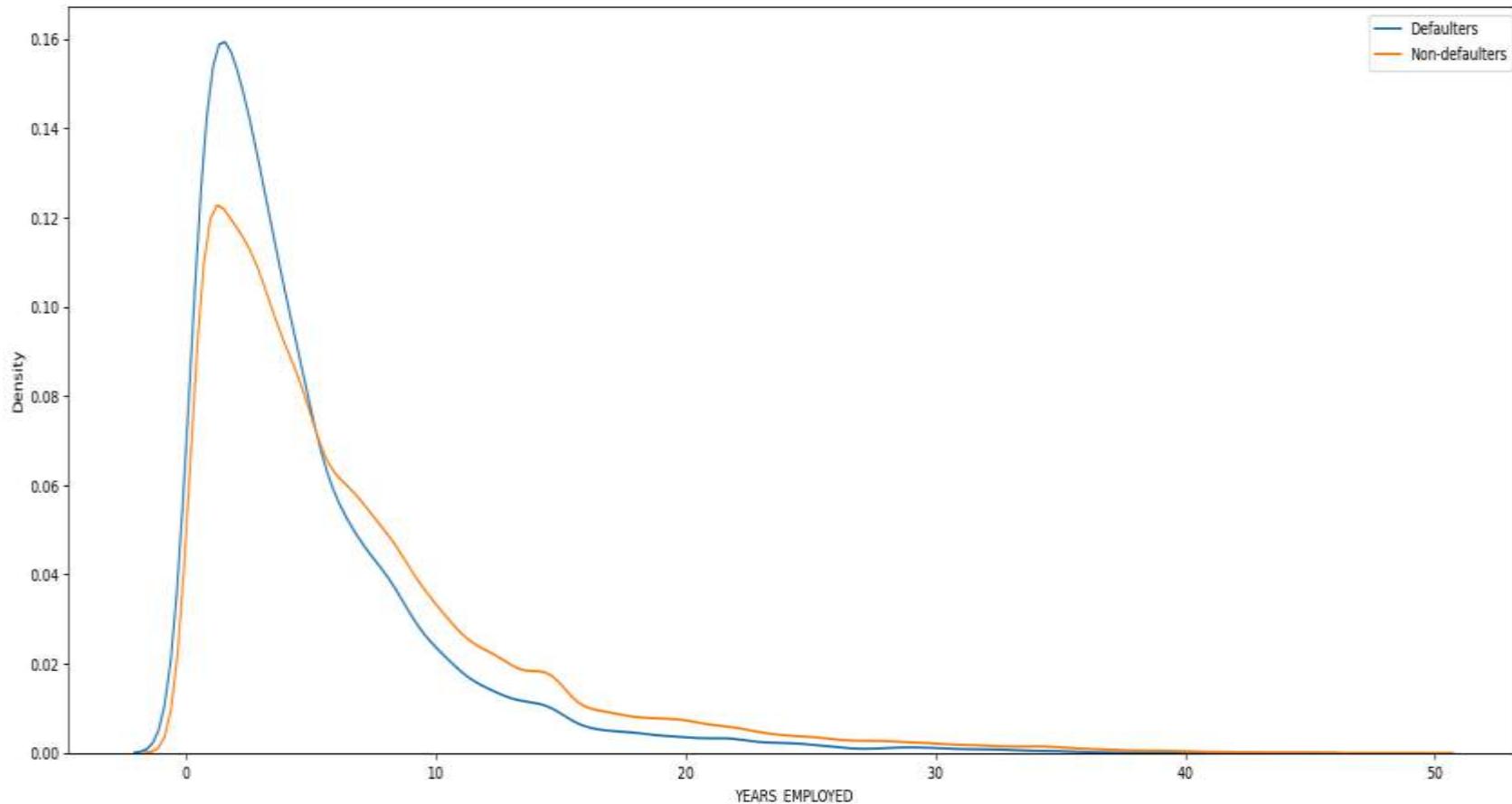
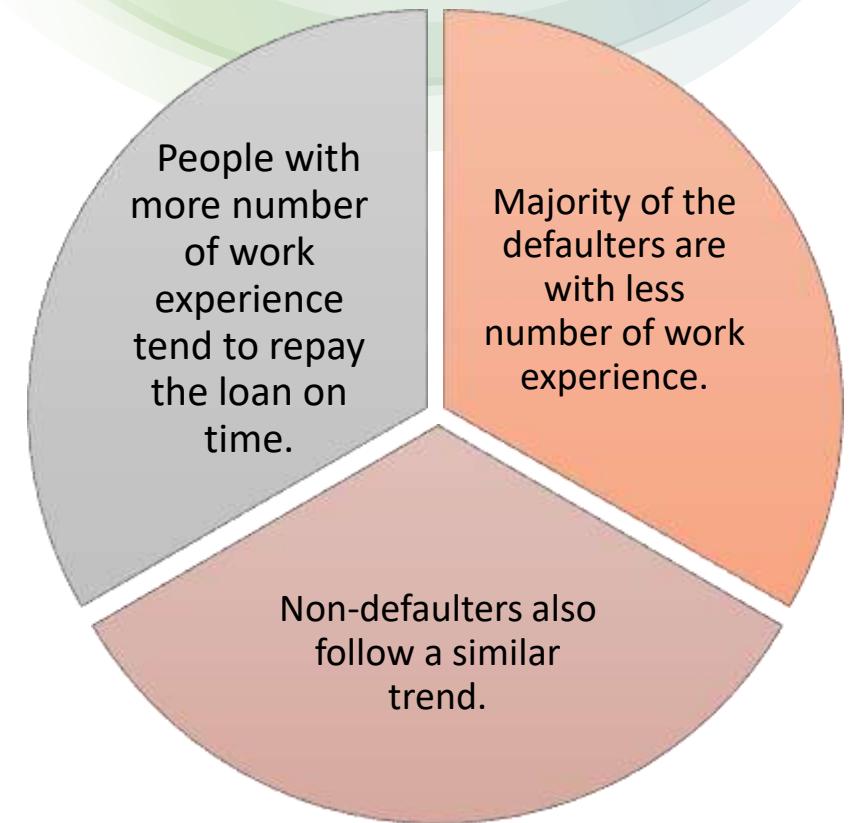
Amount Annuity is concentrated within the range of 10000 - 50000 for people who repay the loan on time.

Defaulters also follow the same trend. Most of the defaulters' Amount Annuity is within the range of 10000 – 50000.

# Univariate Analysis

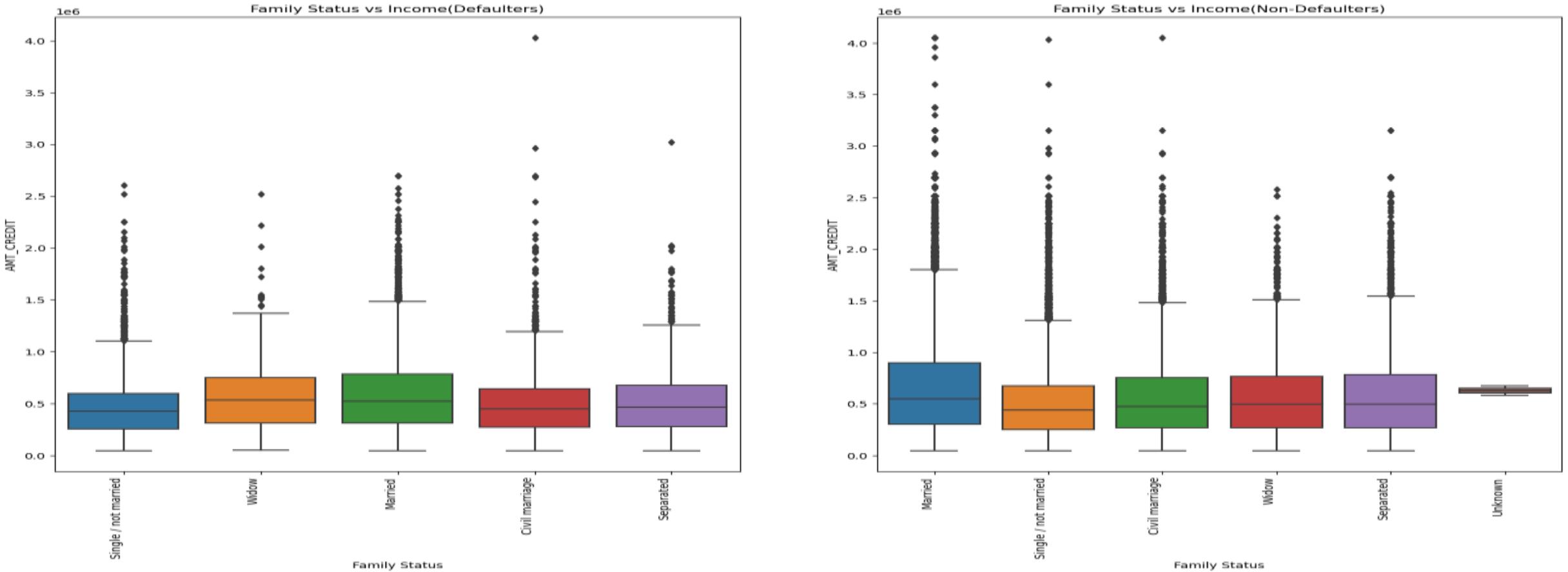
for Target 1 and Target 0 (Defaulters and Non – Defaulters)

## Univariate Analysis of Years Employed



# Bivariate Analysis

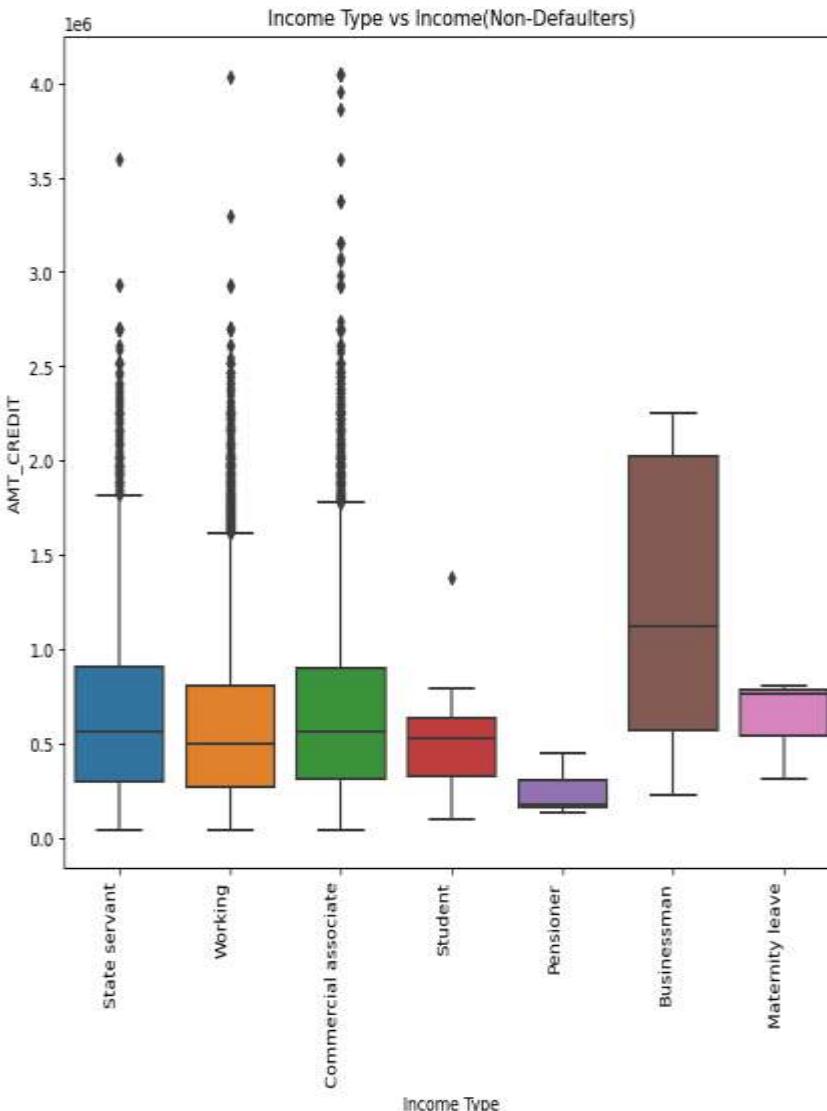
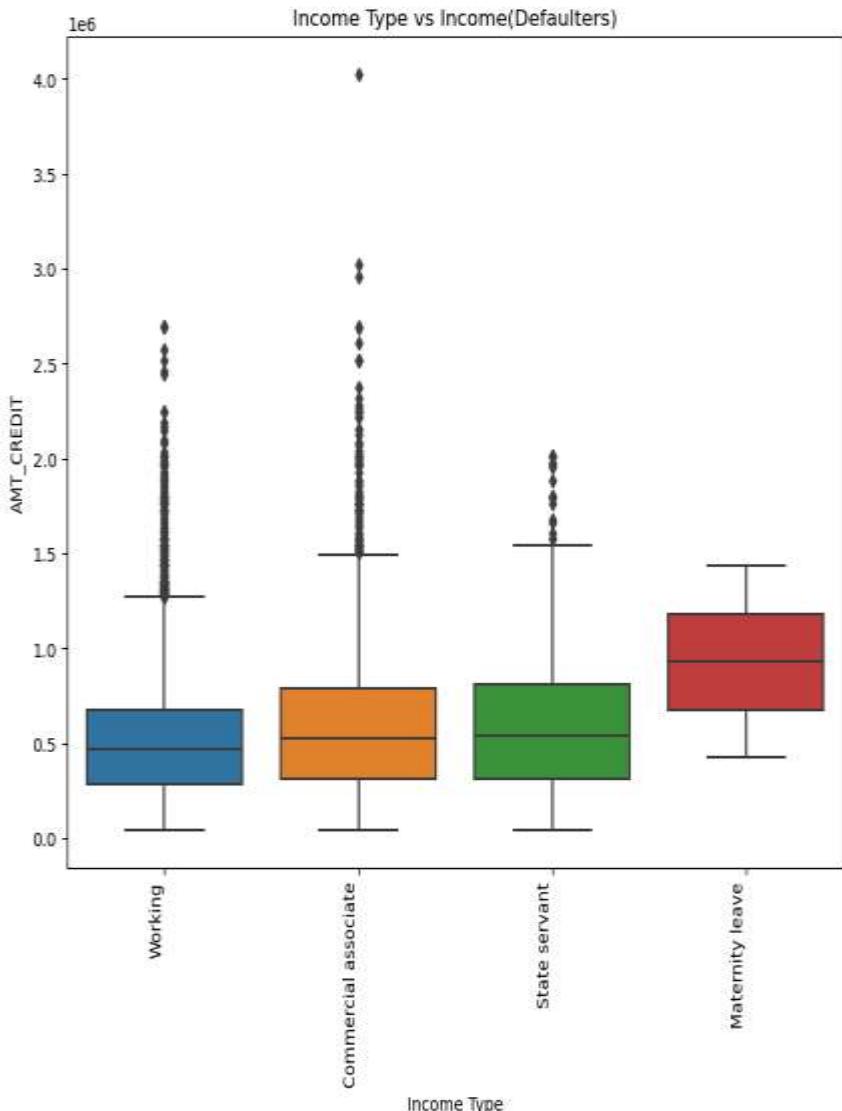
## Bivariate Analysis of Family Status and Credit Amount



- Majority of the clients who applied for comparatively higher loan amounts are married people in both the Defaulters and Non-Defaulters categories.

# Bivariate Analysis

## Bivariate Analysis of Income Type and Credit Amount



Businessman seems to be one of the best categories to be considered for loans.

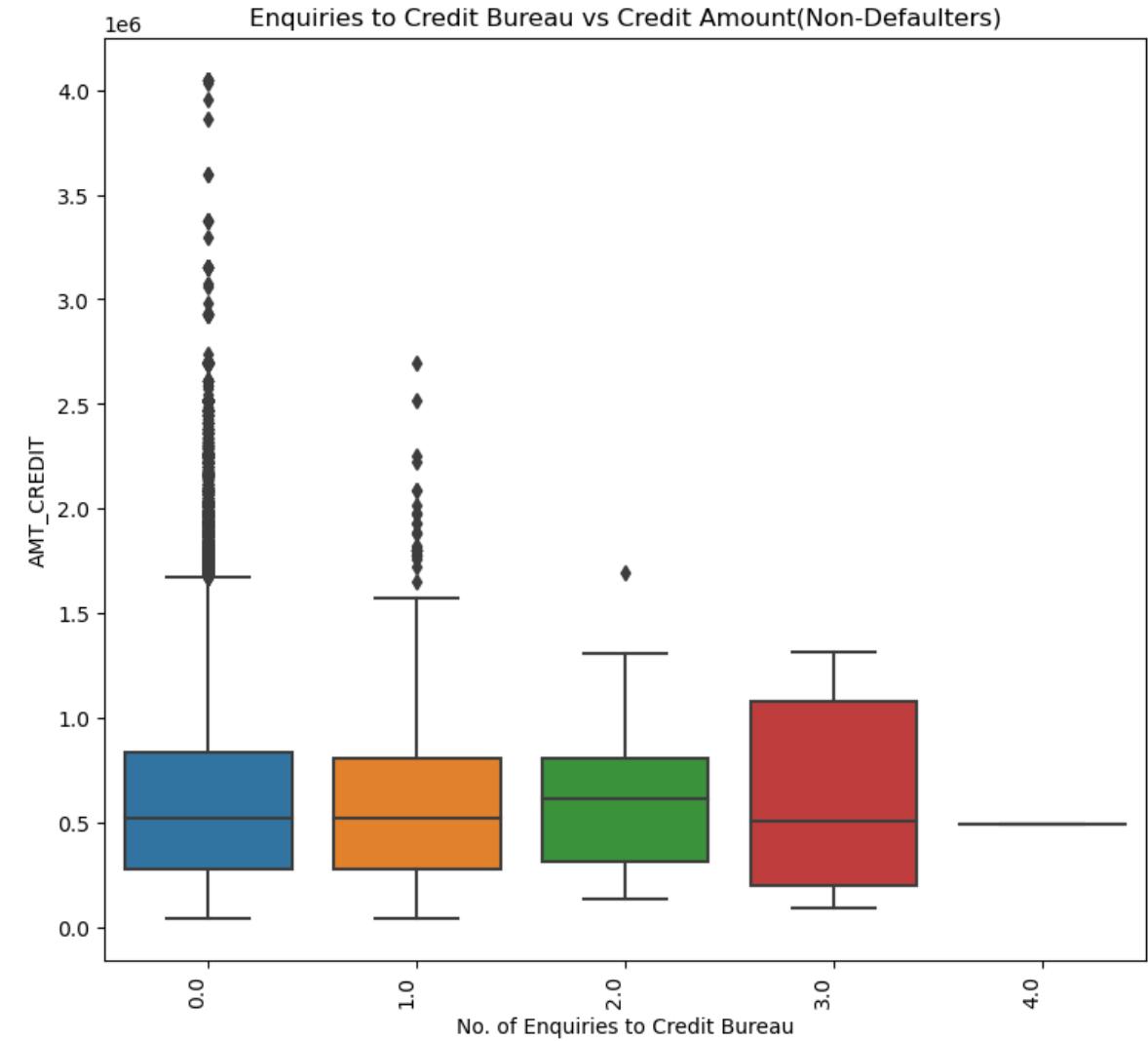
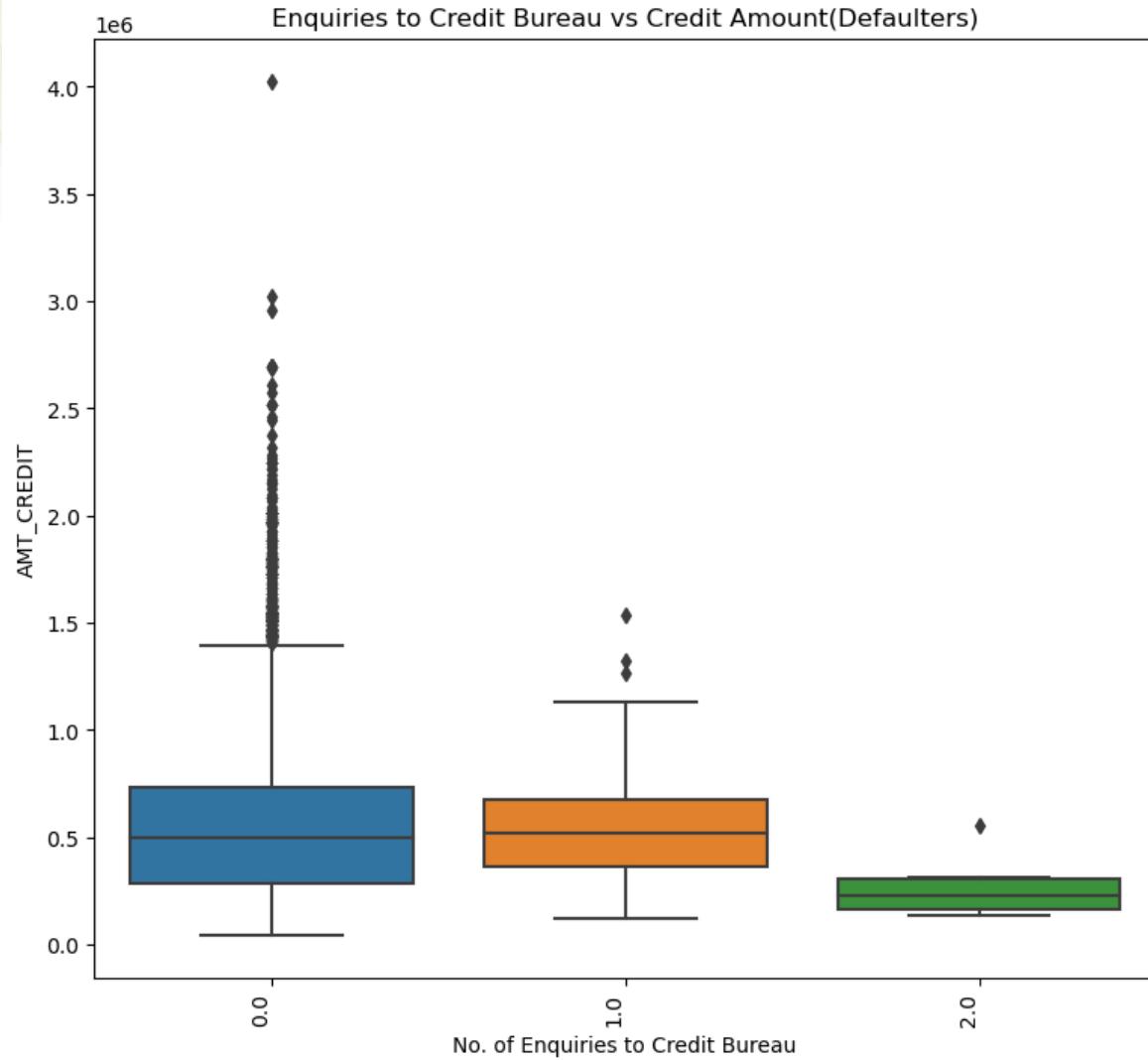
Income type – ‘Businessman’ does not even appear in the Defaulters plot. They all make sure to repay the loan on time.

Student is another good category to be considered for loans. They also make sure to repay the loan on time.

A category that would be better not to target is Female clients on Maternity leave.

# Bivariate Analysis

## Bivariate Analysis of Enquiries to Credit Bureau and Credit Amount



# Bivariate Analysis

## Bivariate Analysis of Enquiries to Credit Bureau and Credit Amount

- Clients with more number of enquiries to Credit Bureau make sure to repay the loan on time.

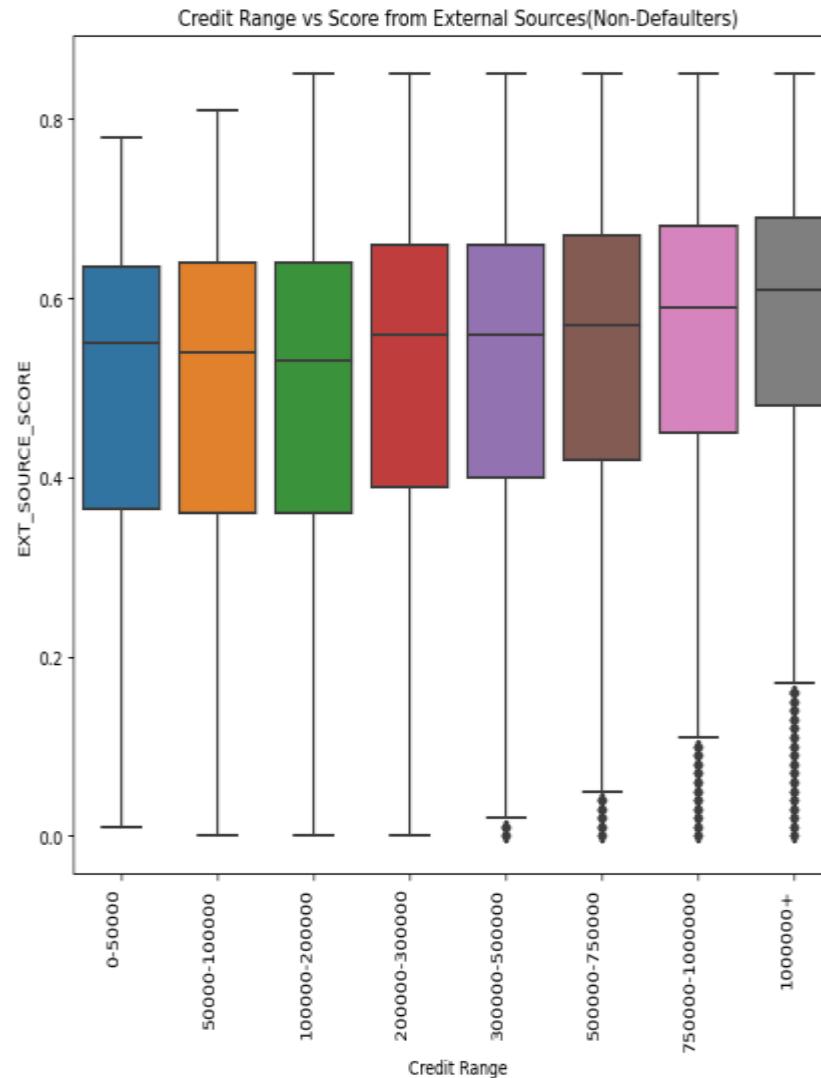
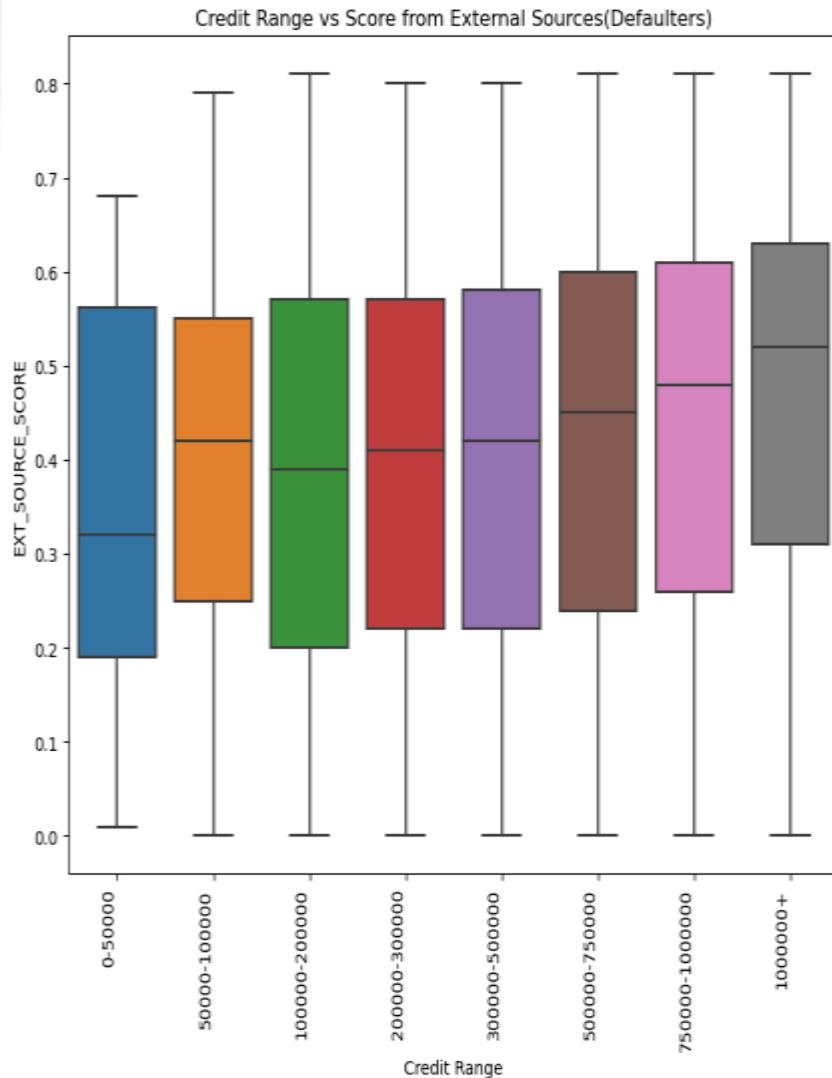
- Even though the number of value counts are too low for values 3 and 4, they don't appear in the Defaulter's plot.

- It is a possibility that those clients have approached more than one bank for the loan or maybe more than one time which might have resulted in more number of enquiries to the Credit Bureau. It can be assumed that those clients were actively looking for a loan.

- The plot shows they make sure to repay the loan on time. Hence the number of enquiries to Credit Bureau can be used as an effective measure to shortlist the target clients.

# Bivariate Analysis

## Bivariate Analysis of Credit Range and Score from External Sources

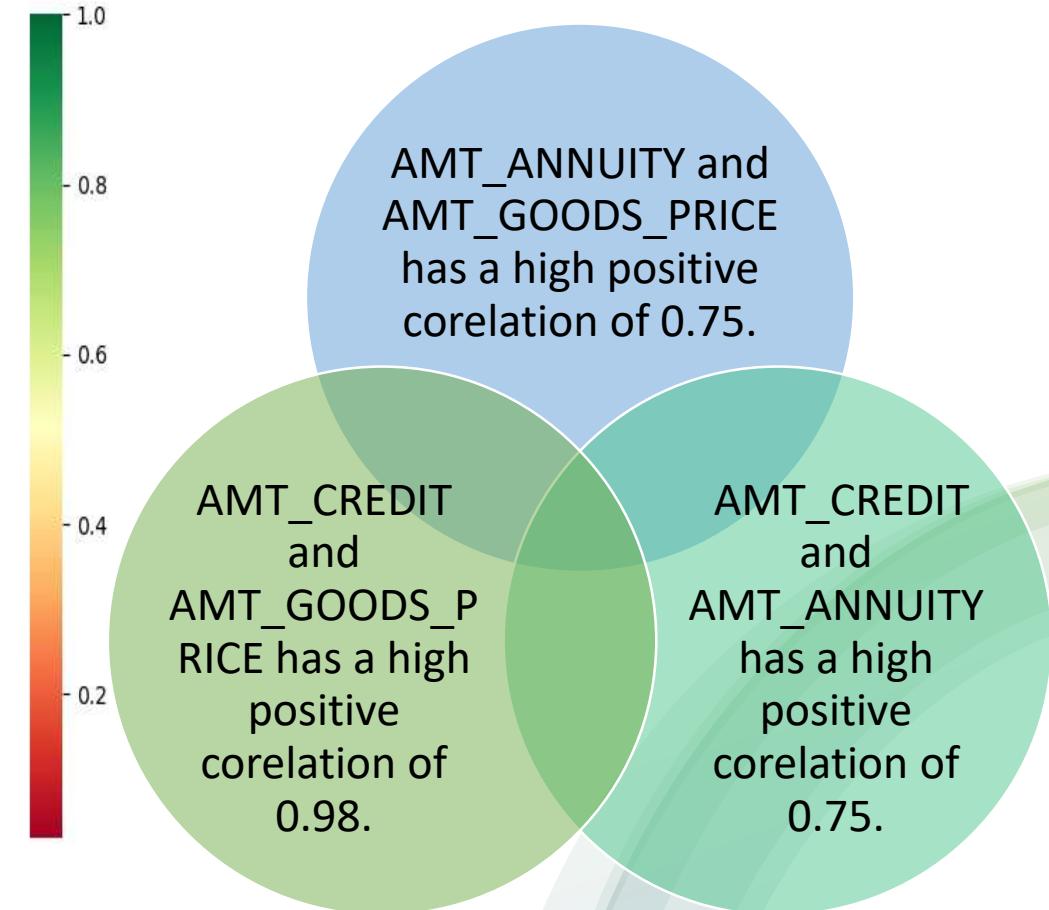
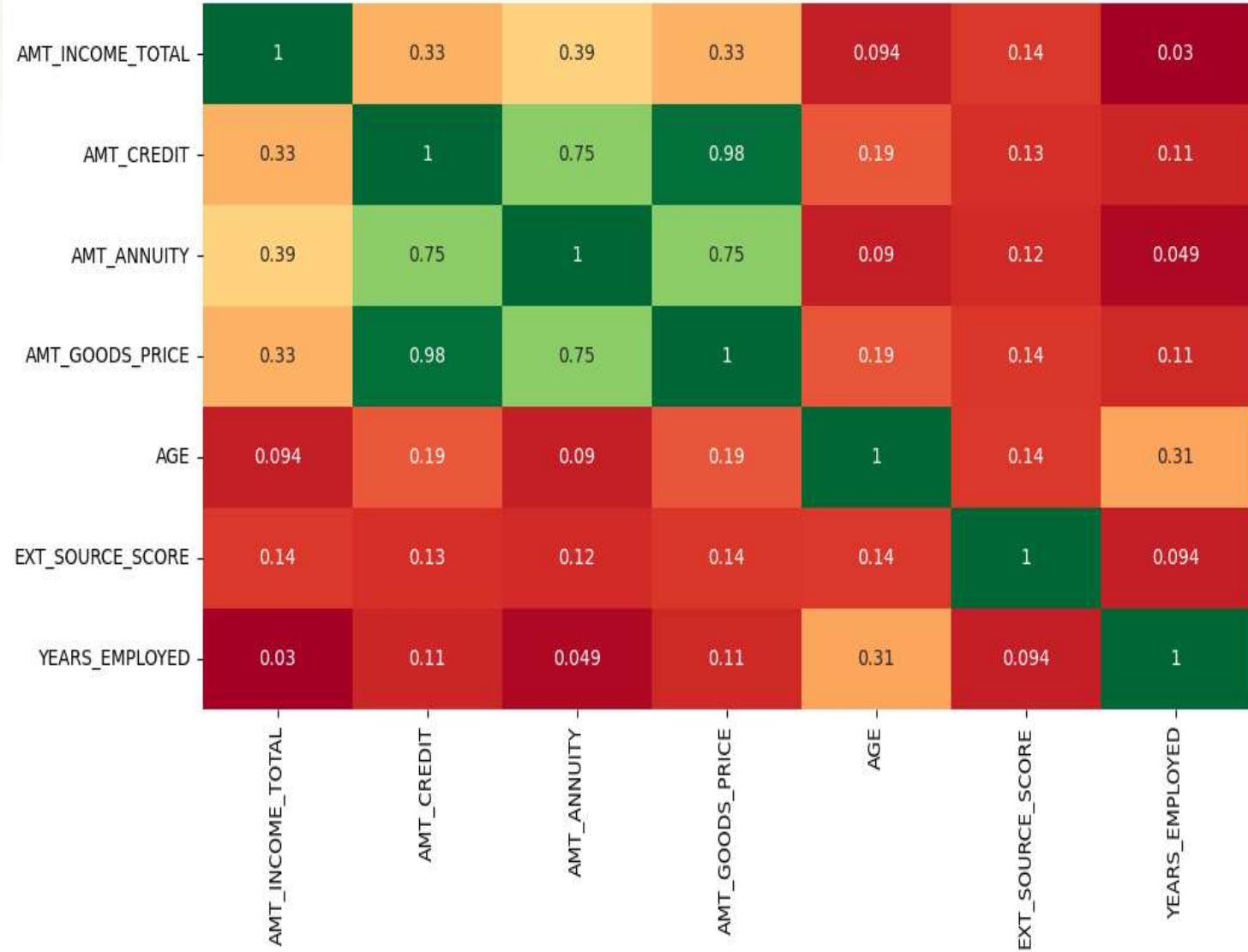


Clients with high “EXT\_SOURCE\_SCORE” (Score from External sources) make sure to repay the loan on time. Median of clients who repay the loan on time is moderately high than who are defaulters.

Hence it is a good measure to consider to arrive in a decision while targeting the clients.

# Correlation Analysis

## Clients with Payment Difficulties



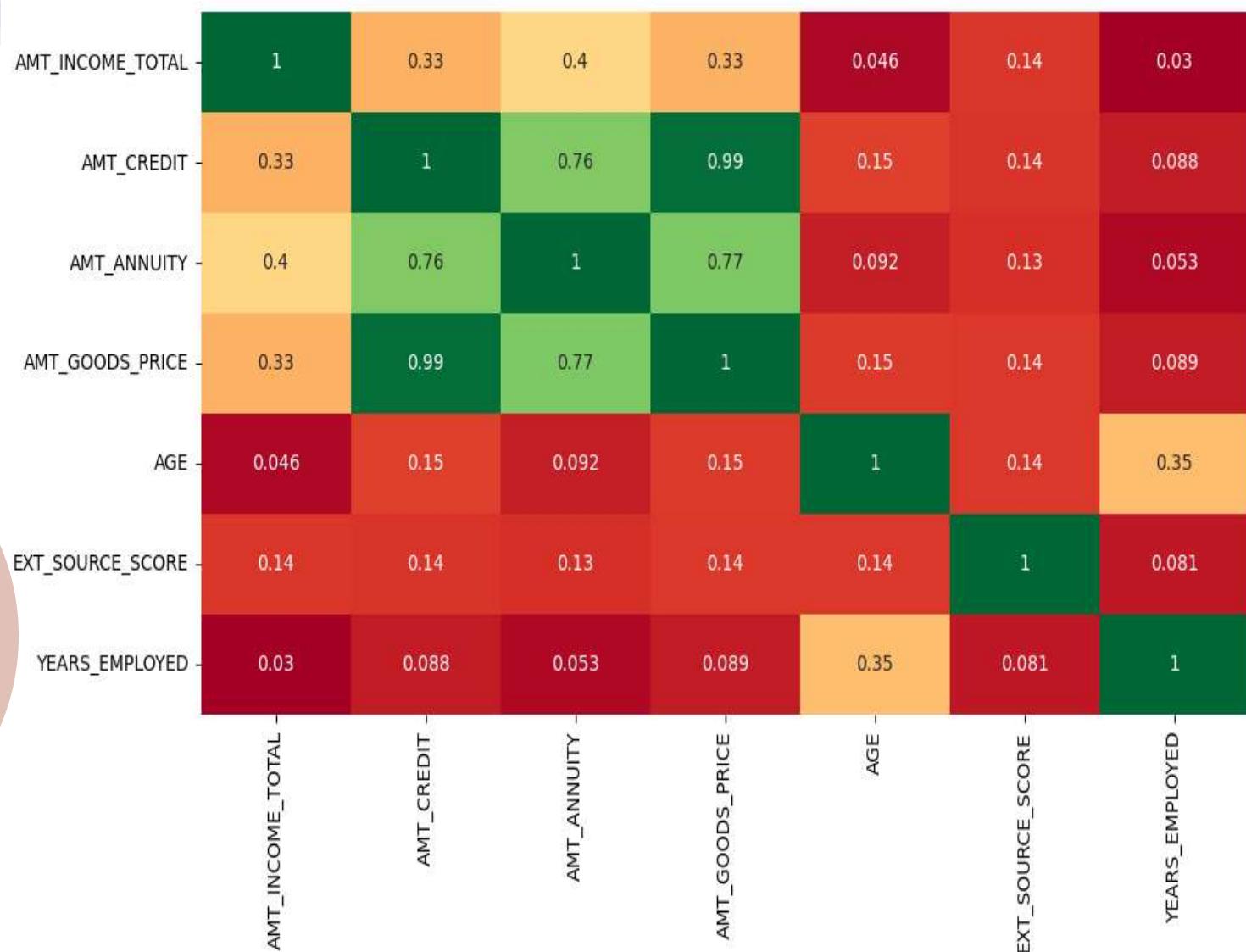
# Correlation Analysis

## Clients with No Payment Difficulties

AMT\_CREDIT and AMT\_GOODS\_PRICE has a high positive corelation of 0.99.

AMT\_CREDIT and AMT\_ANNUITY has a high positive corelation of 0.76.

AMT\_ANNUITY and AMT\_GOODS\_PRICE has a high positive corelation of 0.77.



# Correlation Analysis

***For both the Defaulters and Non-Defaulters the same pairs of columns are highly correlated.***

AMT\_GOODS\_PRICE and AMT\_CREDIT has a strong positive correlation.

This means that as Goods price increases, the Credit Amount also increases.

AMT\_ANNUITY and AMT\_CREDIT has a strong positive correlation.

This means that as Annuity Amount increases, Credit Amount also increases.

AMT\_ANNUITY and AMT\_GOODS\_PRICE has a strong positive correlation.

This means that as Annuity Amount increases, Goods Price also increases.

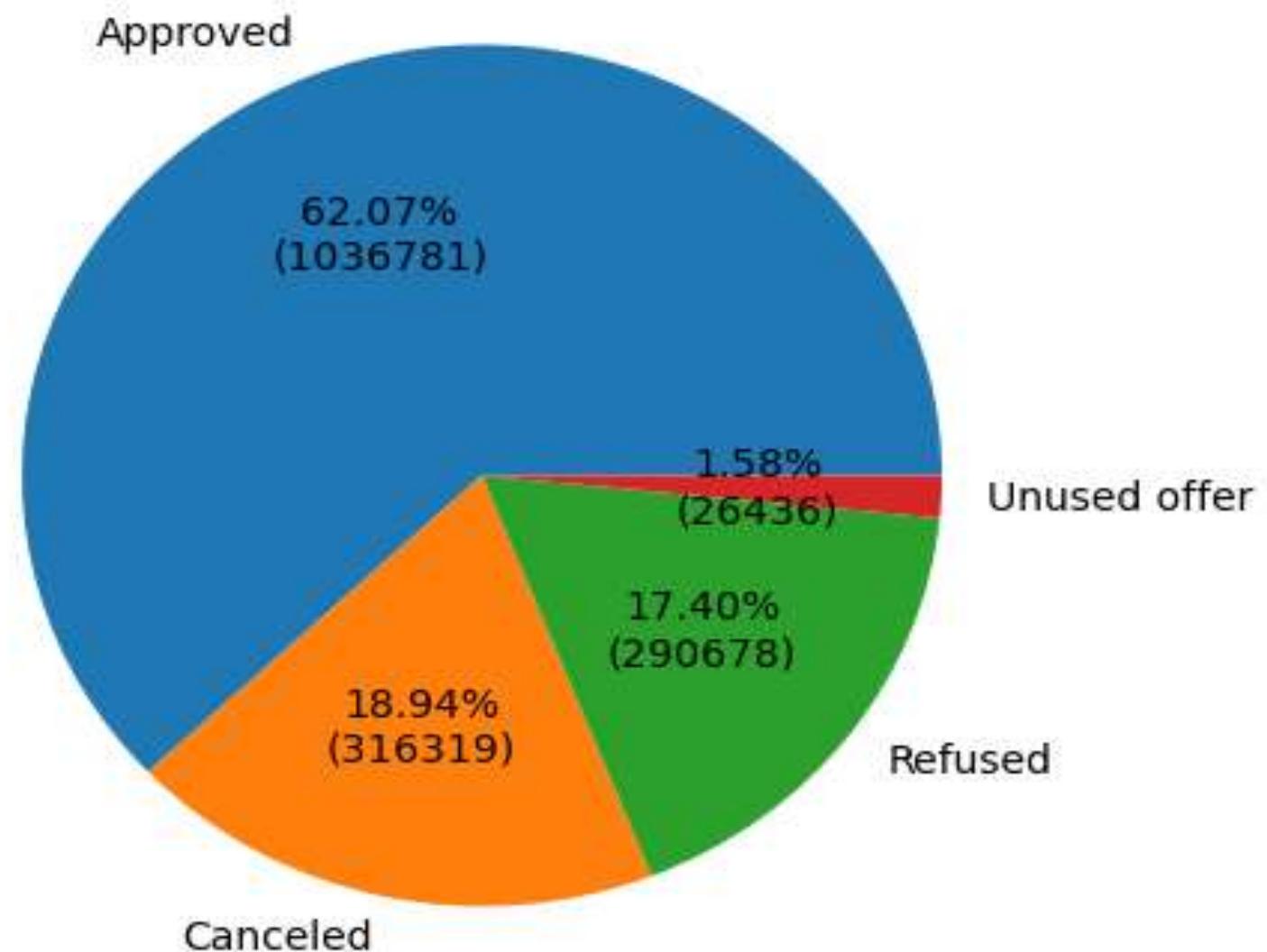
# Previous Application Analysis

## Contract Status of Previous Application

Majority of the loan applications are 'Approved' in the Previous Application Data.

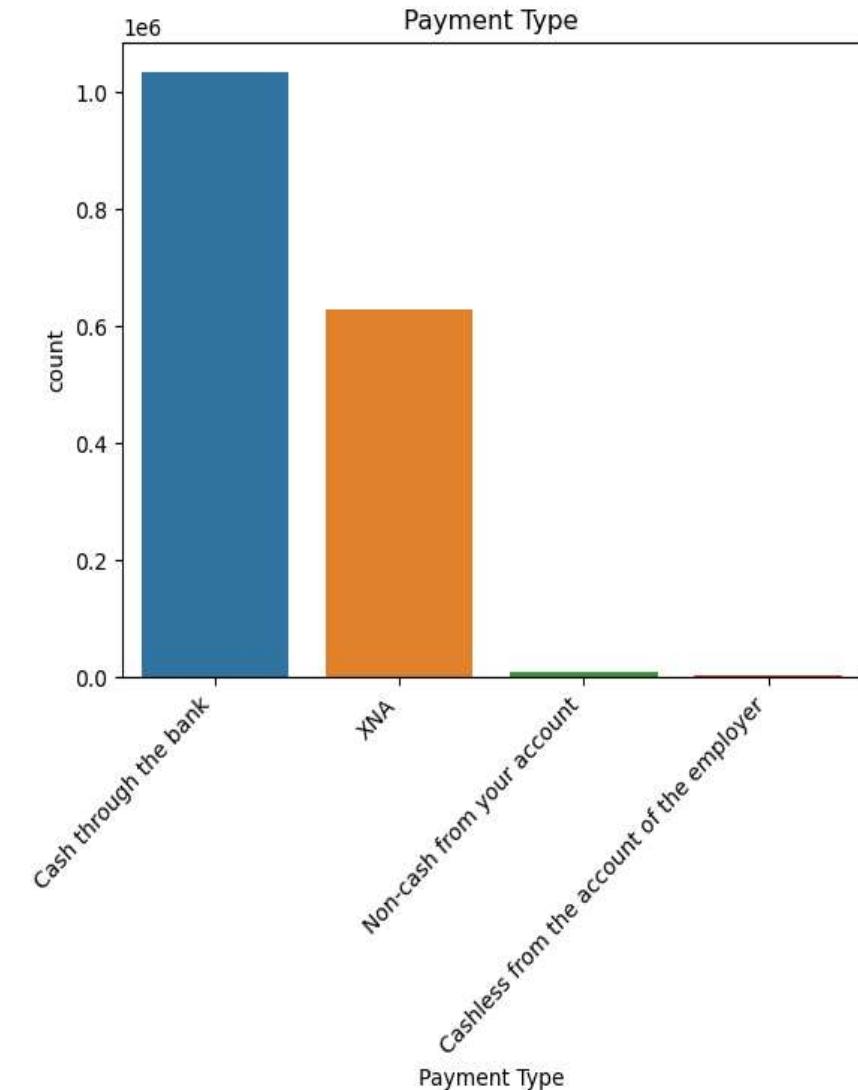
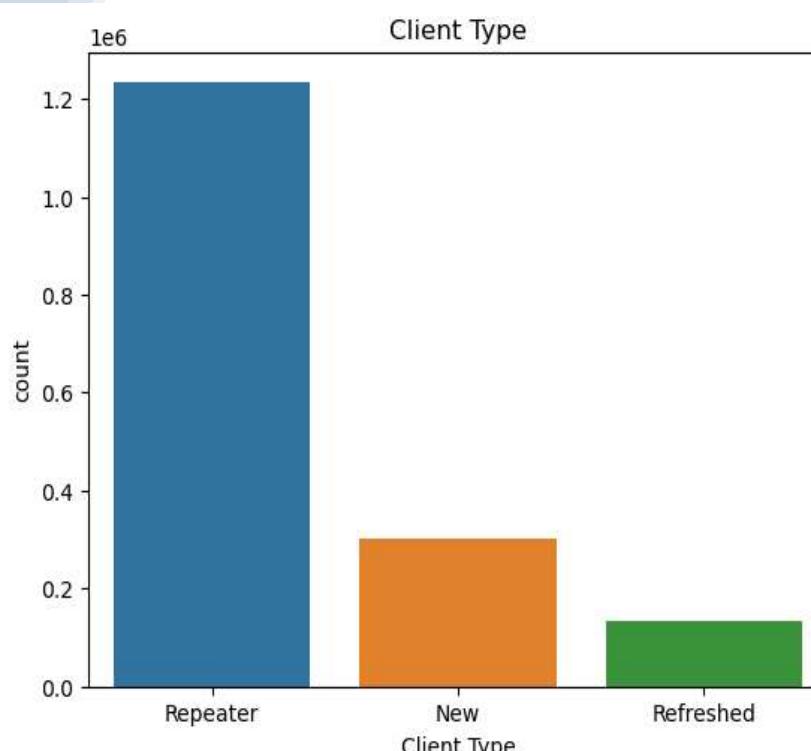
62.07% are Approved  
18.94% are Cancelled  
17.40% are Refused  
1.58% are Unused

NAME\_CONTRACT\_STATUS



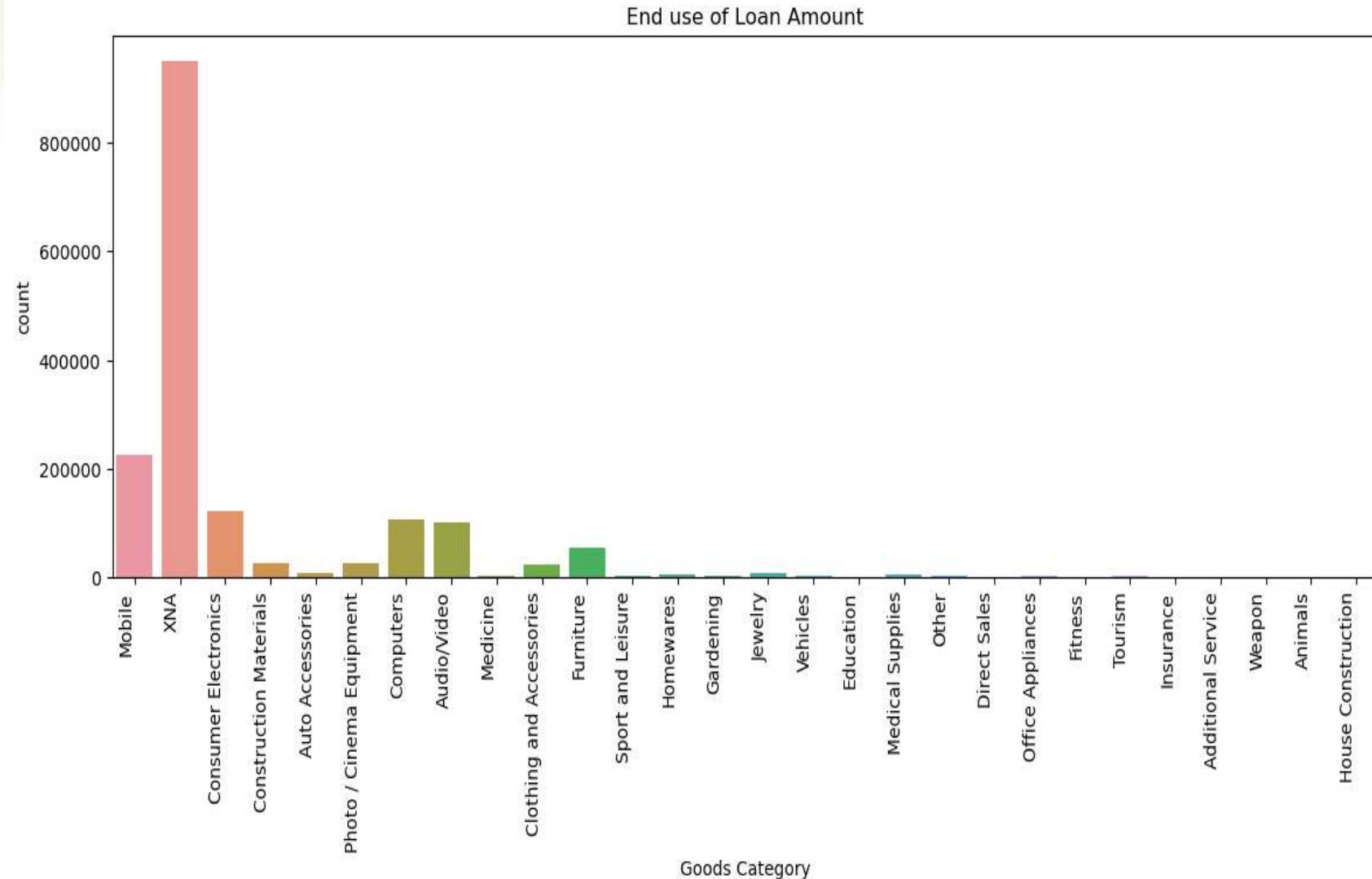
# Previous Application Analysis

## Analysis of Client Type and Payment Method



# Previous Application Analysis

## Analysis of End Use of the loan(Goods Category)



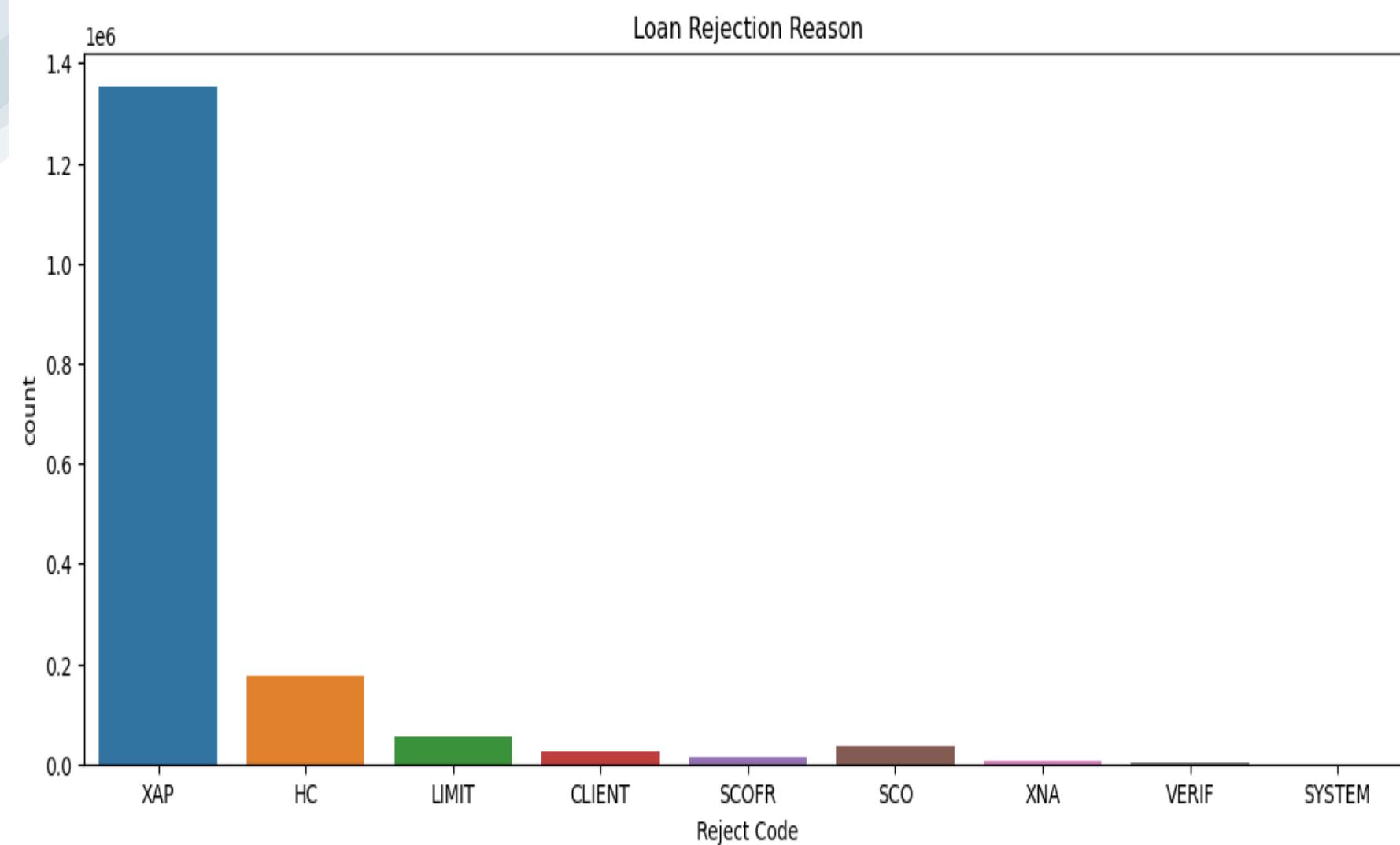
The end use of the loan amount was not known in majority of the cases(XNA).

In case of known cases, most of the clients spent their loan amounts on Mobiles, Consumer Electronics and Computers respectively.

# Previous Application Analysis

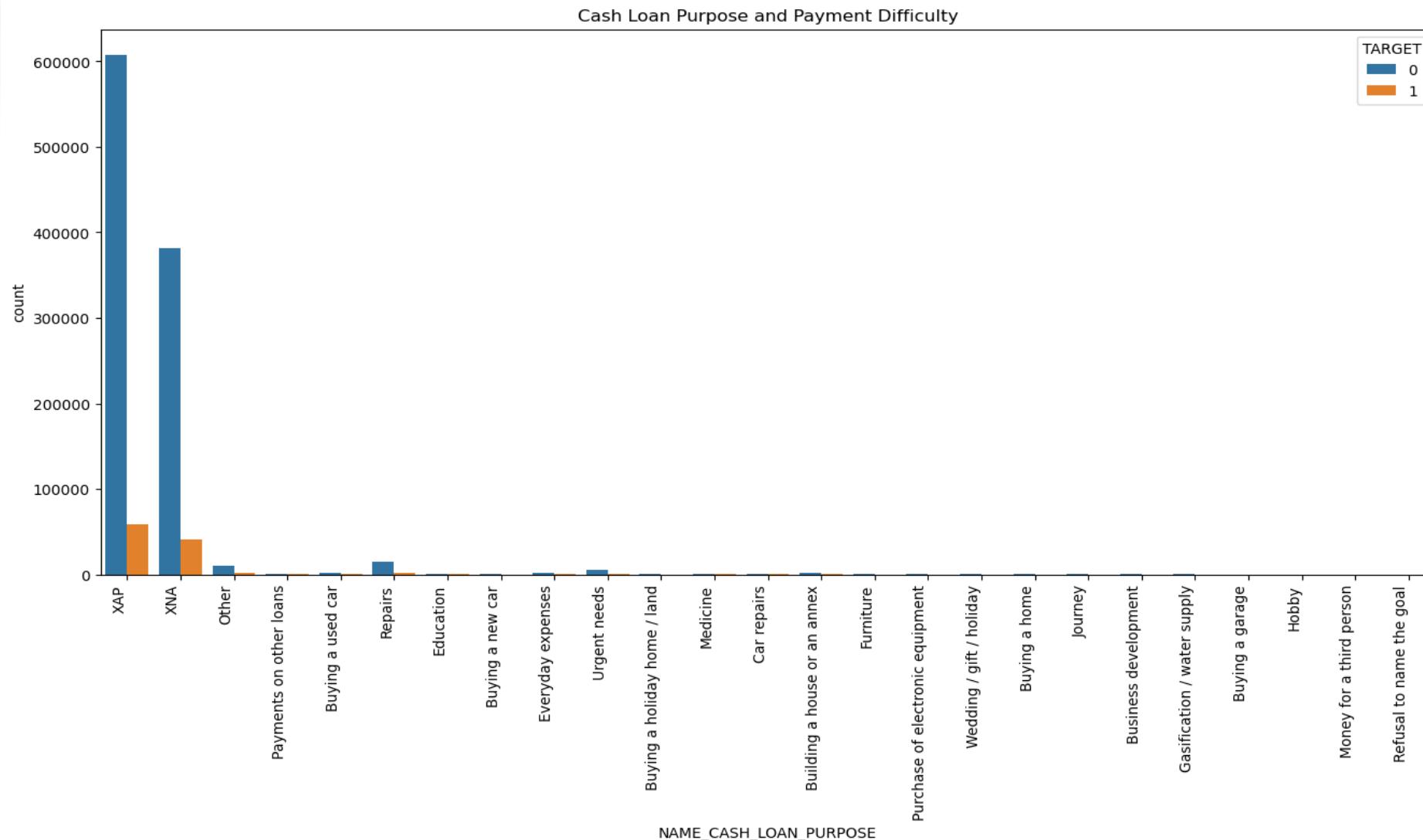
## Analysis of Reason for Rejection

Majority of the loan applications were rejected because of the Rejection Code - "HC"



# Previous Application Analysis

## Analysis of Cash Loan Purpose and Payment Difficulty



Clients who availed loans for Repairs in the Previous Application are among the top category who defaulted the repayments in the current application.(XAP and XNA are unknown values)

# Previous Application Analysis

## Correlation Analysis



# Previous Application Analysis

## Correlation Analysis

**AMT\_APPLICATION** -  
There is a high positive  
corelation with  
**AMT\_GOODS\_PRICE\_y**,  
**AMT\_ANNUITY\_y** and  
**AMT\_CREDIT\_y**.

**AMT\_CREDIT\_x** -  
There is a high positive  
corelation with  
**AMT\_GOODS\_PRICE\_x** and  
a good positive corelation  
with **AMT\_ANNUITY\_x**.

**AMT\_CREDIT\_y** -  
There is a high positive  
corelation with  
**AMT\_GOODS\_PRICE\_y**,  
**AMT\_ANNUITY\_y** and  
**AMT\_APPLICATION**.

**AMT\_ANNUITY\_x** -  
There is a good positive  
corelation with  
**AMT\_GOODS\_PRICE\_x** and  
**AMT\_CREDIT\_x**.

**AMT\_ANNUITY\_y** -  
There is a high positive  
corelation with  
**AMT\_GOODS\_PRICE\_y**,  
**AMT\_CREDIT\_y** and  
**AMT\_APPLICATION**.

**AMT\_GOODS\_PRICE\_x** -  
There is a high positive  
corelation with  
**AMT\_CREDIT\_x** and a good  
positive corelation with  
**AMT\_ANNUITY\_x**.

**AMT\_GOODS\_PRICE\_y** -  
There is a high positive  
corelation with  
**AMT\_ANNUITY\_y**,  
**AMT\_CREDIT\_y** and  
**AMT\_APPLICATION**.

# Insights



- "Businessman", "Pensioner" and "Student" are the best categories to be considered for loans.
- Clients with high "Score from External Sources" make sure to repay the loan on time. Hence it is recommended to target clients with high "Score from External Sources".
- Clients with more number of "Enquiries to Credit Bureau" make sure to repay the loan on time. Considering other factors like high "Score from External Sources" and "Income Type" with more number of "Enquiries to Credit Bureau" is recommended.
- Majority of the clients spend their loan amounts on Mobiles, Consumer Electronics and Computers. A business plan to target such clients would result in more profitability.



# Insights

- It is highly recommended not to target female clients who are on maternity leave.
- Clients who availed loan for "Repairs" are among the top category who defaulted the repayment. Hence it is recommended to be cautious while considering the loan applications for "Repairs".
- Majority of the loan applications were rejected because of the Rejection Code - "HC". Analysing on Rejection Code - "HC" in the initial stages of the loan assessment itself would be a better approach.