

# CREDIT EDA ASSIGNMENT



❖ Dipansu Prusty

# BUSINESS OBJECTIVE

---

- 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.
- In other words, 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 utilize this knowledge for its portfolio and risk assessment. To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough).

# PROBLEM STATEMENT

---

A Company is lending various types of loans to Customers.

When the company receives a loan application, the company has to decide on 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.

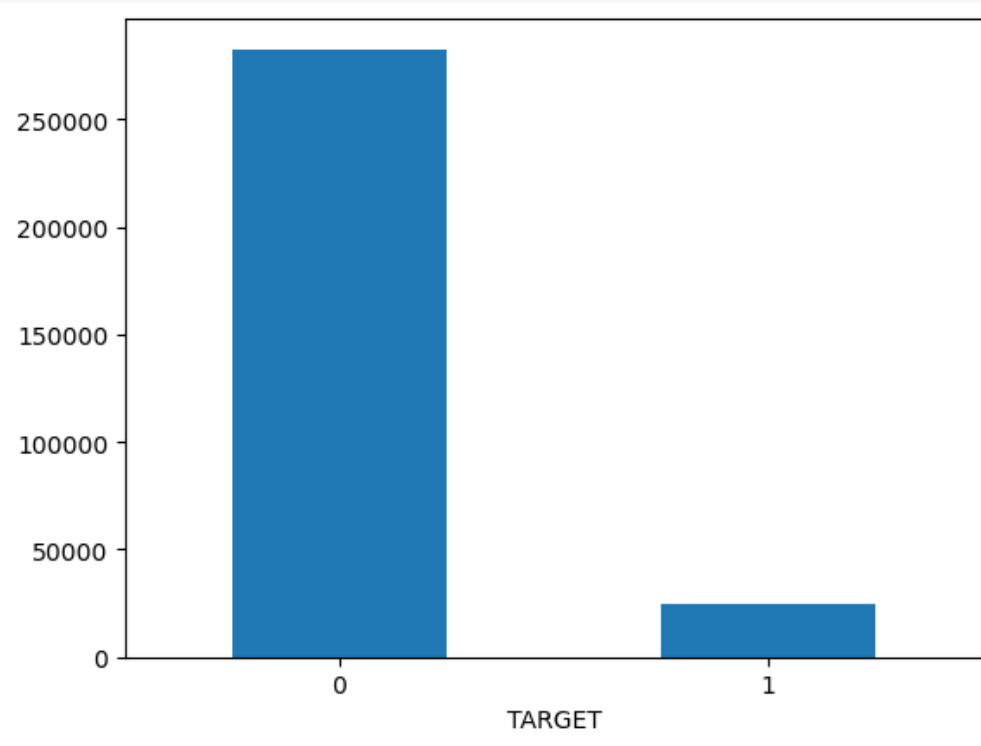
# STEPS TAKEN

---

- Data Sourcing (already provided in assignment)
- Data loading and Data Cleaning :
  - Fixing the rows and columns
  - Imputing and Removing Missing columns
  - Handling Outliers
- Univariate Analysis
- Bivariate and Multivariate Analysis
- Conclusions

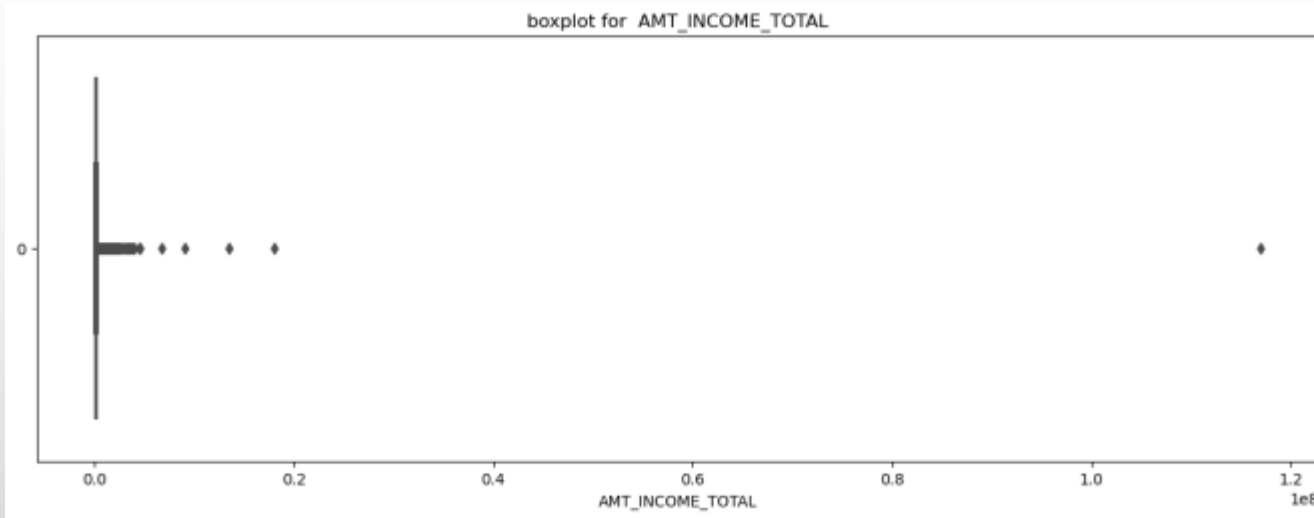
# DATA IMBALANCE

Non-Defaulter & Defaulter

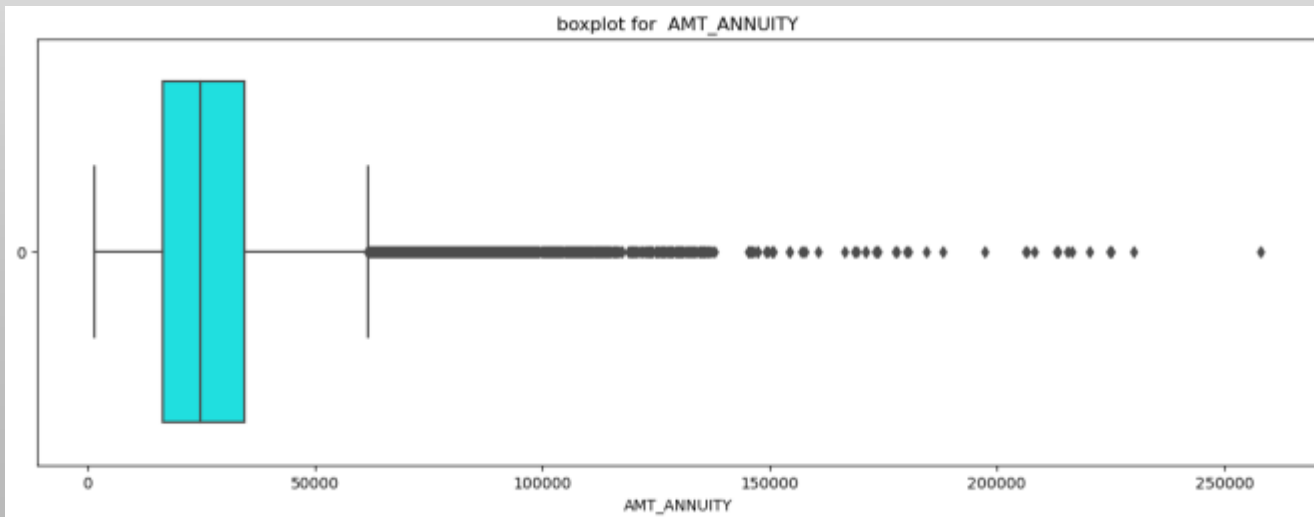


- Here 0 represents “Non-Defaulter” & 1 represents “Defaulter”
- The count for “Non-Defaulter” is 282686 & “Defaulter” is 24825 .So, This is not a balanced Data
- The Imbalance percentage for “Non-Defaulter” customers is 91.92% and for “Defaulter” customers is 8.07%
- The Ratio of imbalance is 11.39

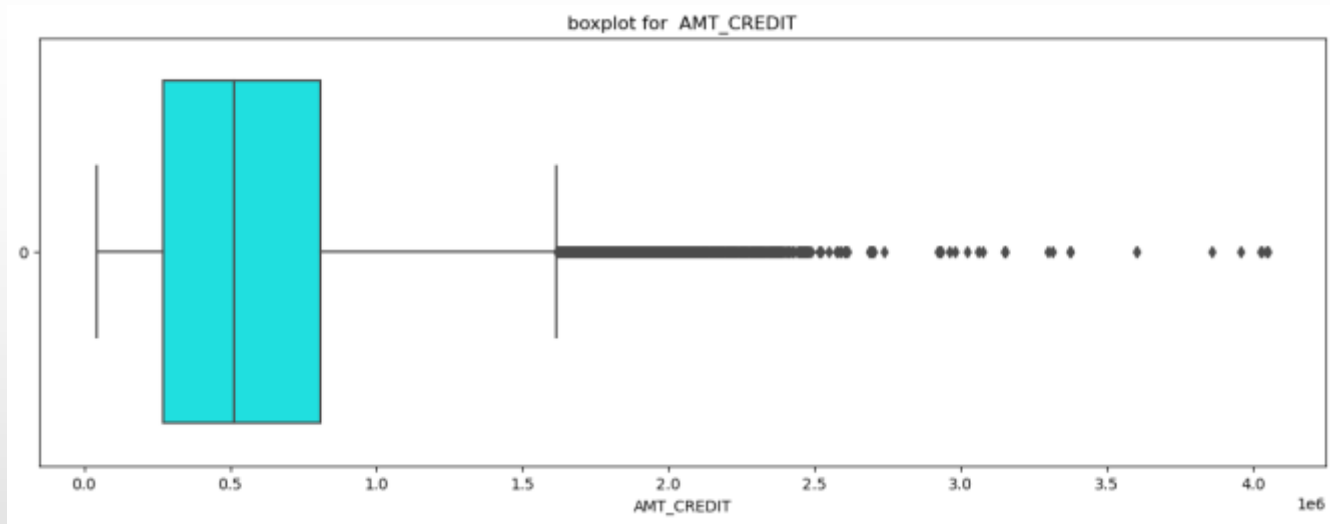
# ANALYSIS OF OUTLIERS



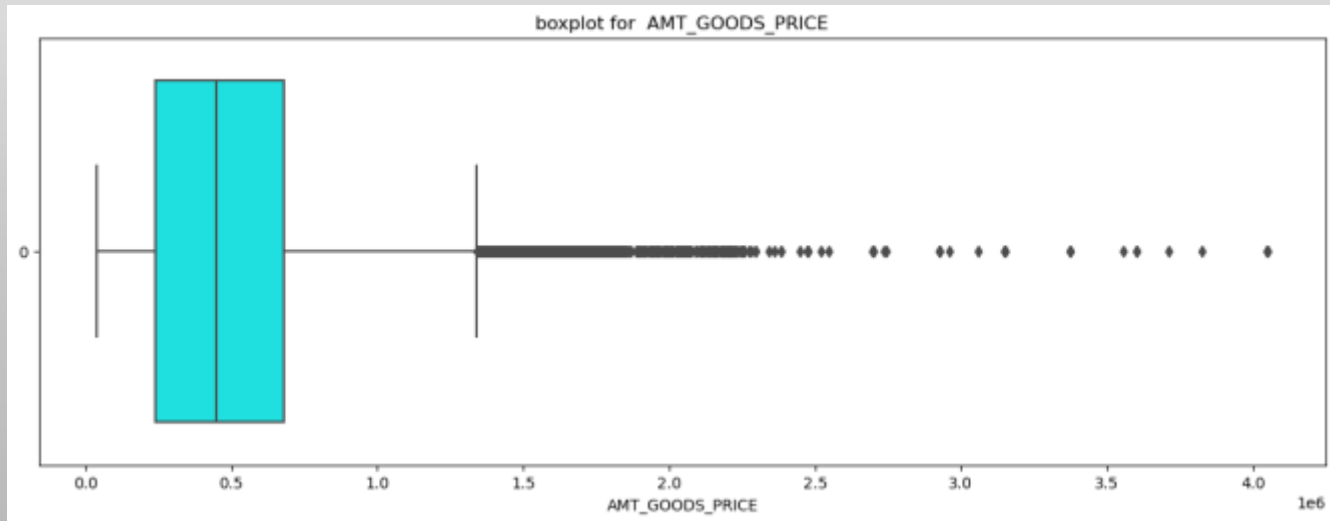
- AMT\_INCOME\_TOTAL has an outlier which is far from main data set that indicates some applicant has high income



- AMT\_ANNUITY has many outliers which is very large



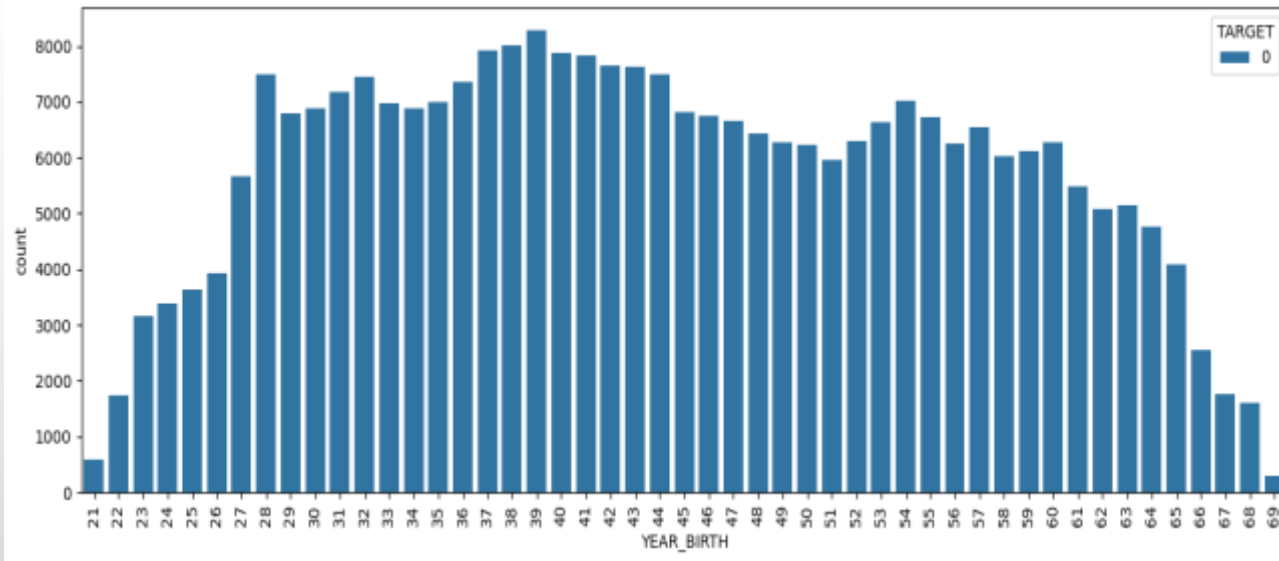
- AMT\_CREDIT has many outliers which is very large



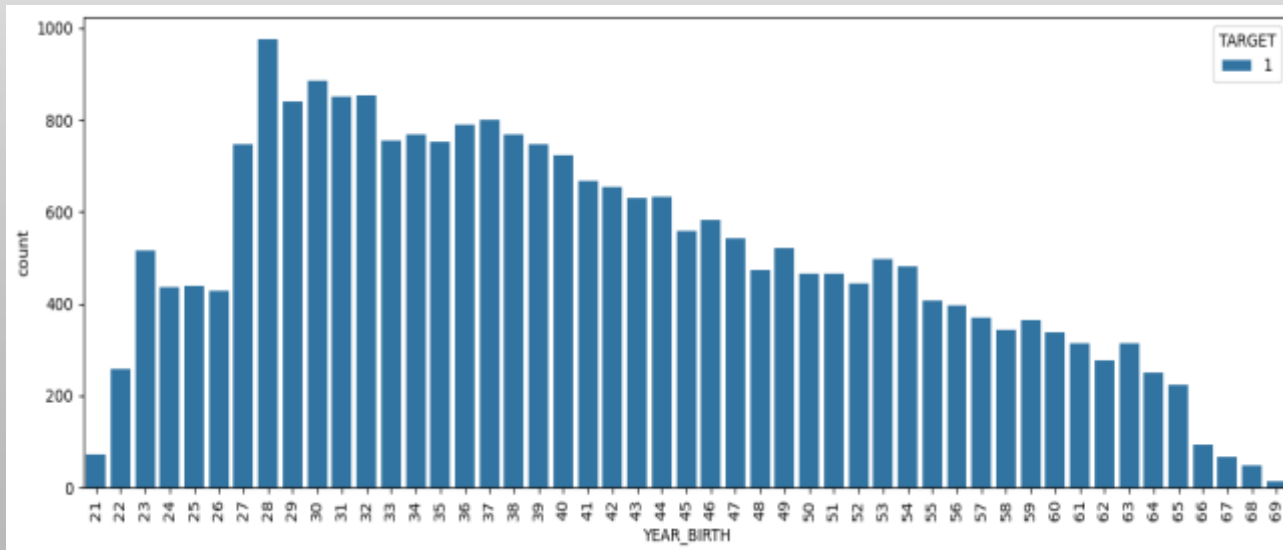
- AMT\_ANNUIITY has many outliers which is very large

# CATEGORICAL ANALYSIS

## AGE ANALYSIS



- Most Non-Defaulters are from 37yr - 43yr age group

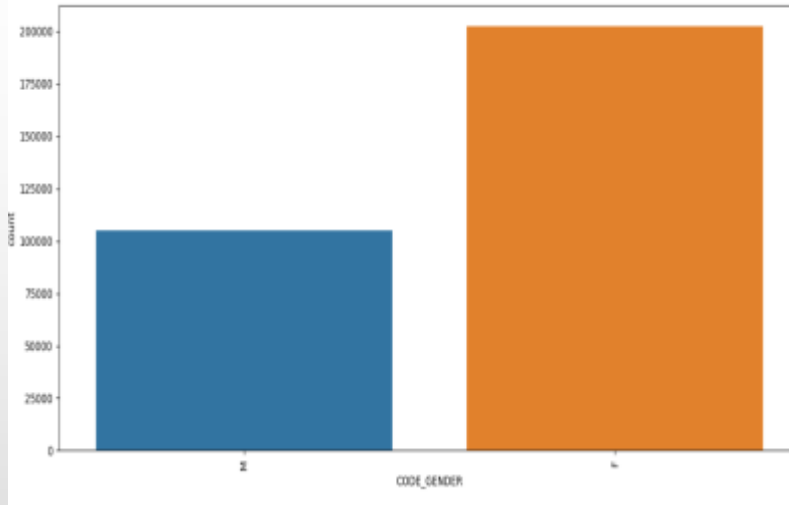


- Most Defaulters are from 27yr - 32yr age group
- 27 year age has huge spike for Defaulter



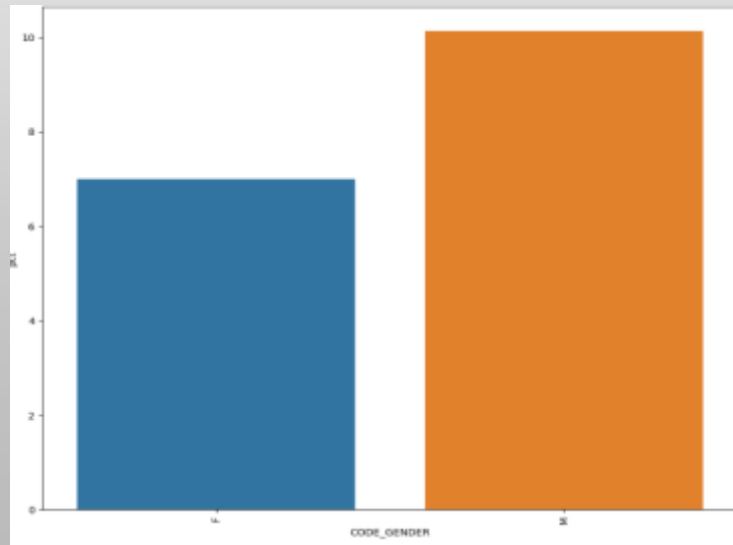
# GENDER TYPE ANALYSIS

Counts of male & female taken loan



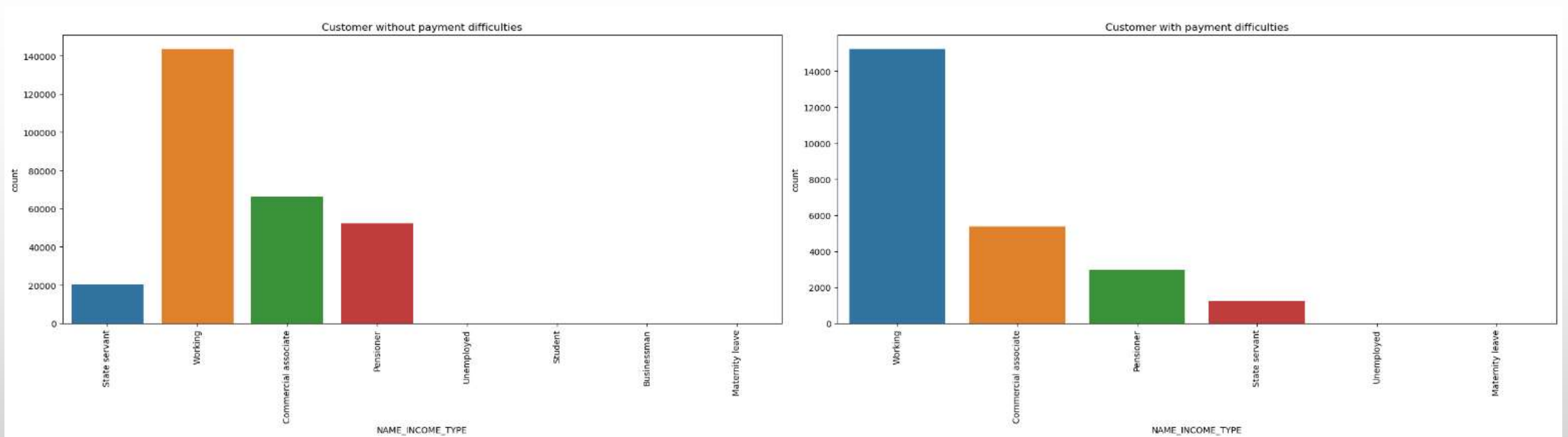
- most of the loans have been taken by female

Male to female Defaulter ratio



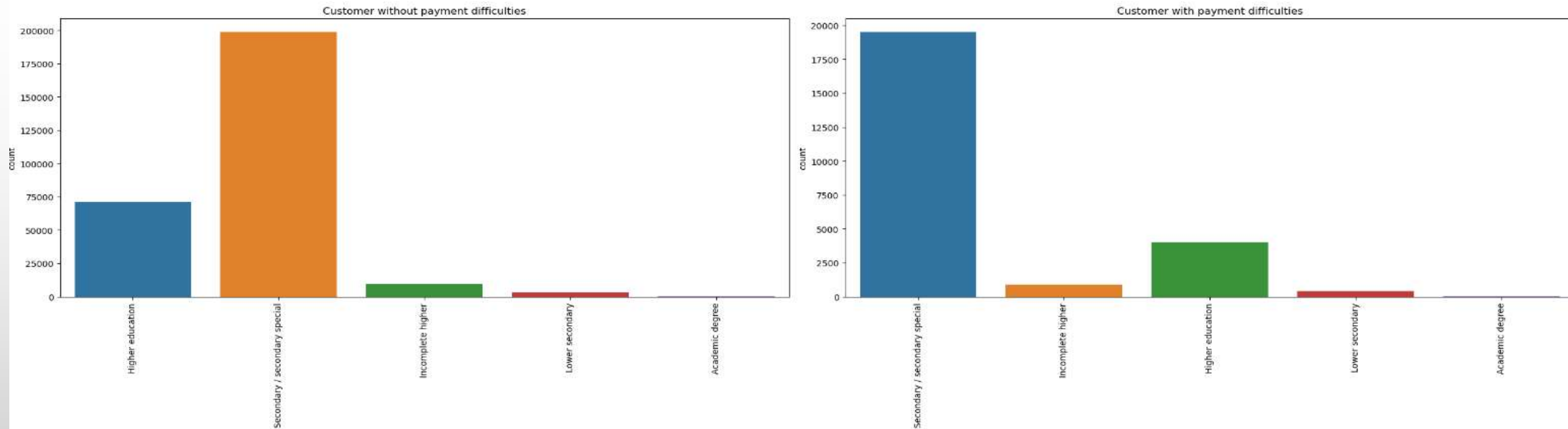
- default rate for females are just 7% lower which is safer and lesser than male

# INCOME TYPE ANALYSIS



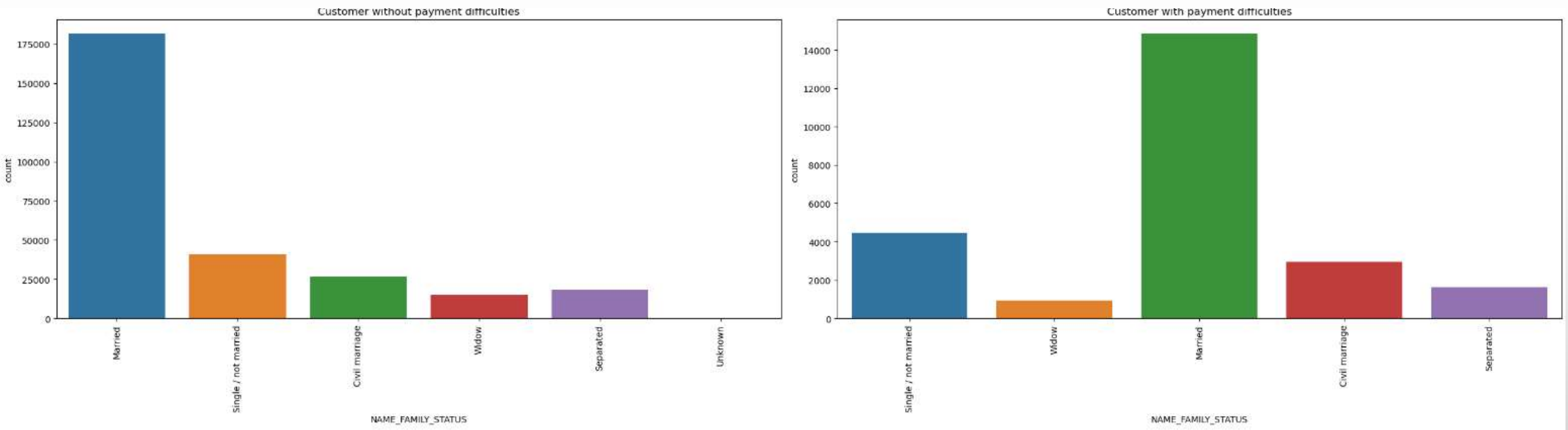
- Bank should focus less on People whose income type is working are more likely to be defaulters than non-defaulters.
- Number for Pensioner is low in the non default case when compared to default case
- the safest segments are working, commercial associates and pensioners

# EDUCATION TYPE ANALYSIS



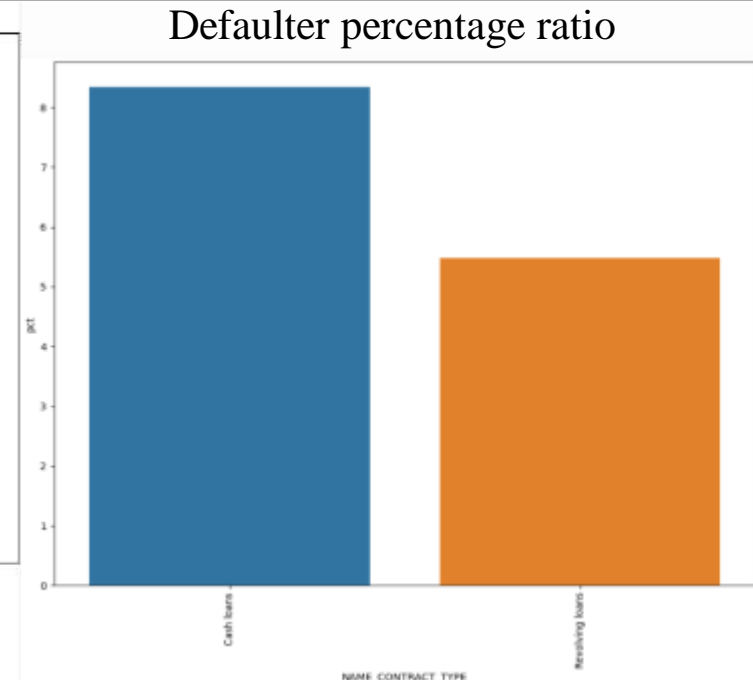
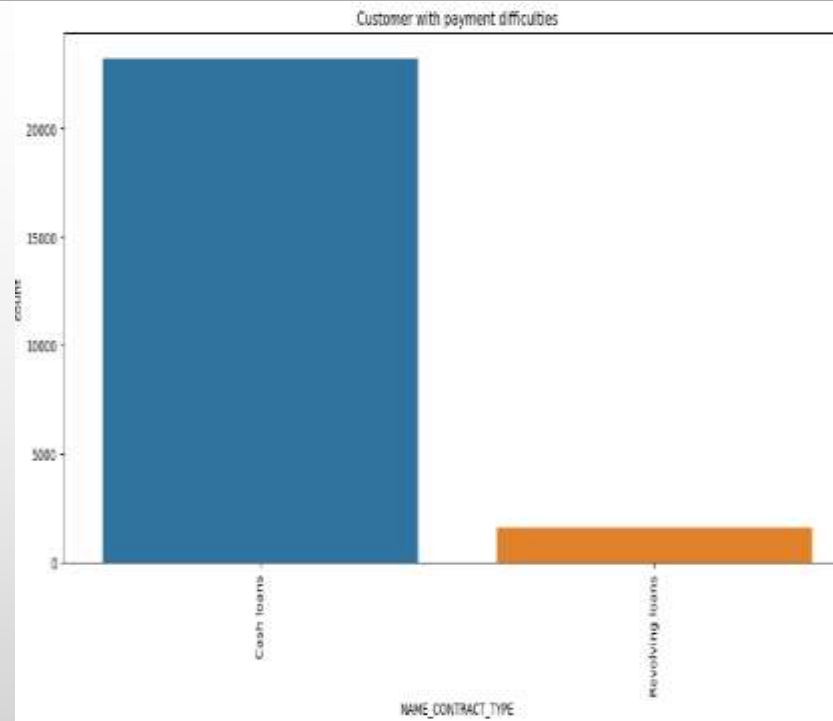
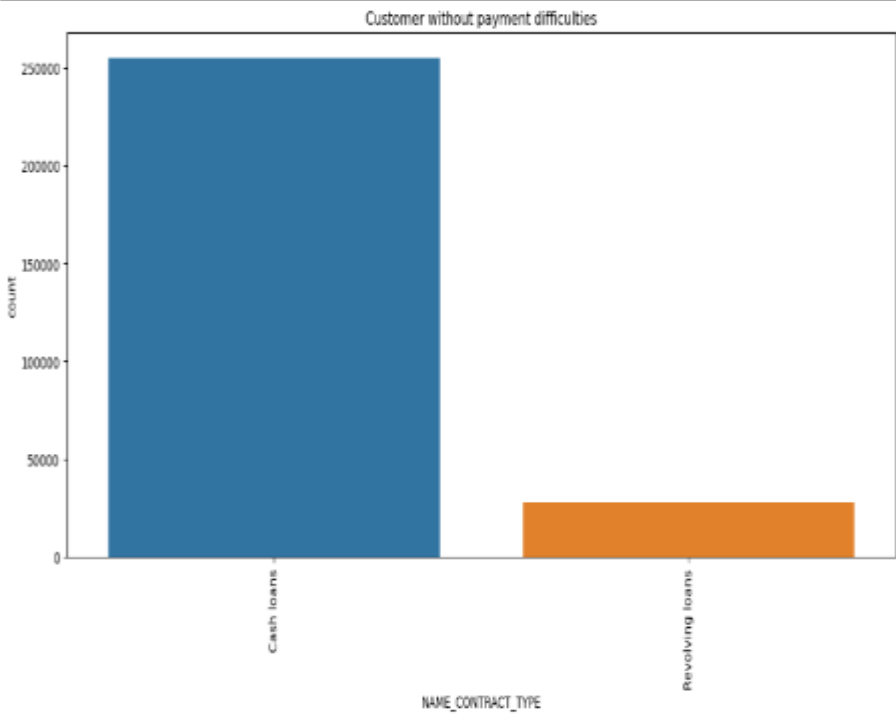
- Majority of the clients have Secondary/secondary special education , Secondary/special educated people are applying for loans are high, Academic degree educated people are applying for loans are least
- clients who has Academic degree are with lower default rate.
- Higher education is the safest segment to give the loan .

# FAMILY STATUS ANALYSIS



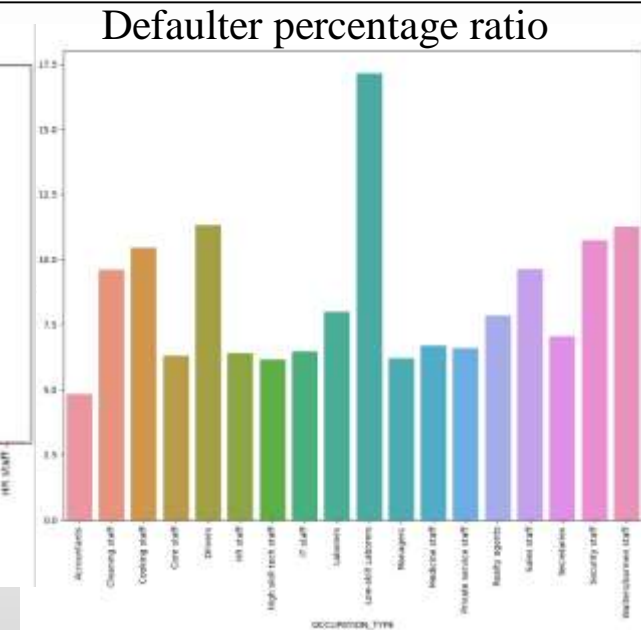
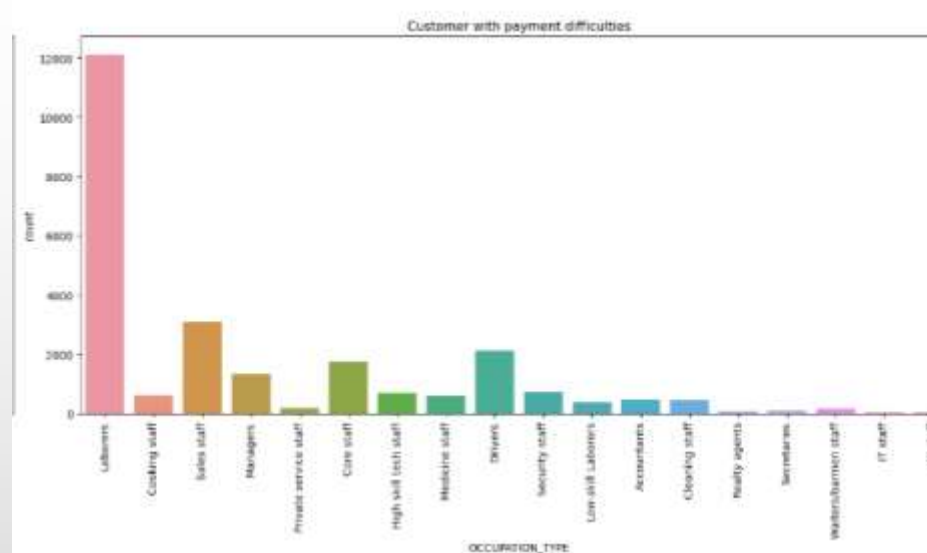
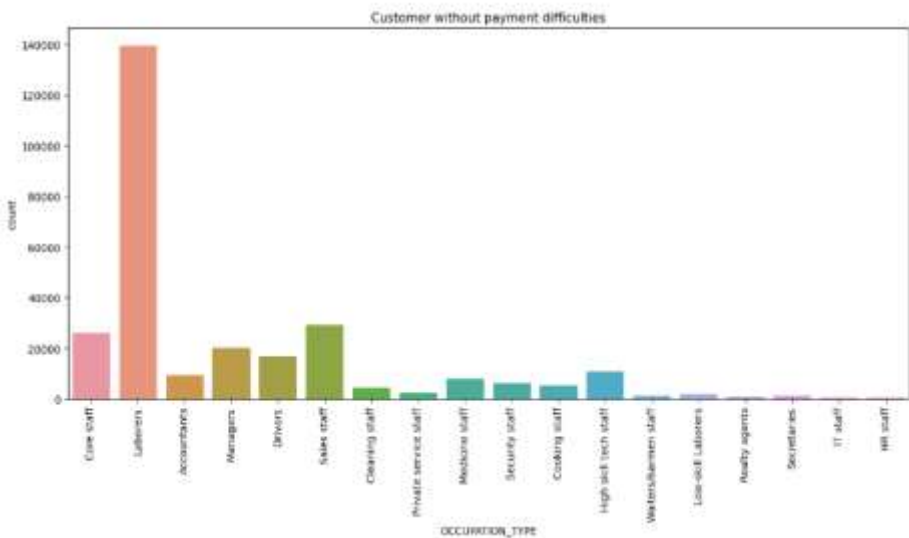
- Most of the people who have taken loans are Married, followed by Single/not married and civil
- clients in a Civil marriage have higher default rate
- Married people are safe to target

# CONTRACT TYPE ANALYSIS



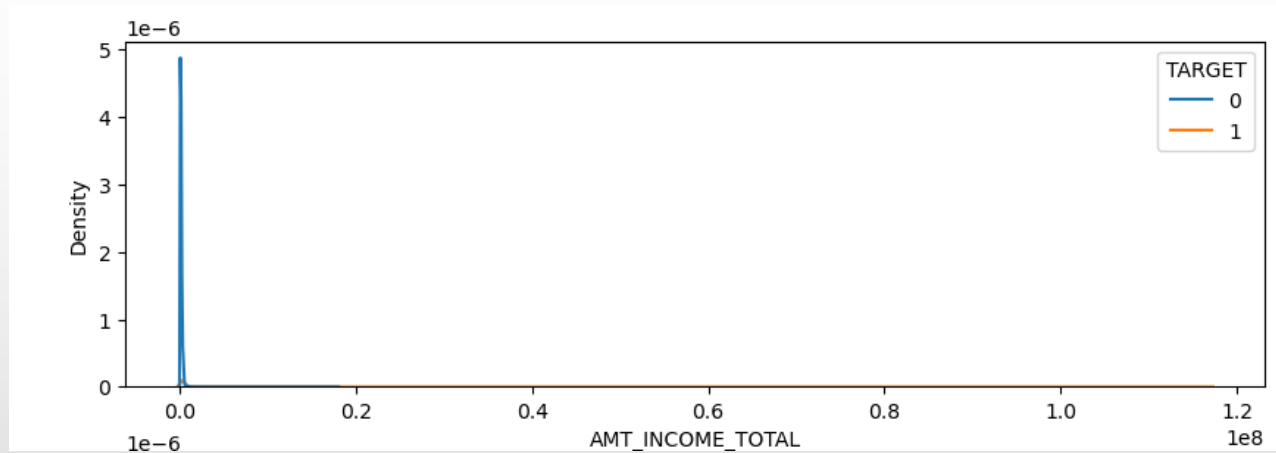
- most of the customers have taken cash loan
- The percentage of defaulter is around 8% in cash loan but as comparing to revolving loan 5.5% but the values of defaulter in revolving loan is higher
- customers who have taken cash loans are less likely to default

# OCCUPATION TYPE ANALYSIS

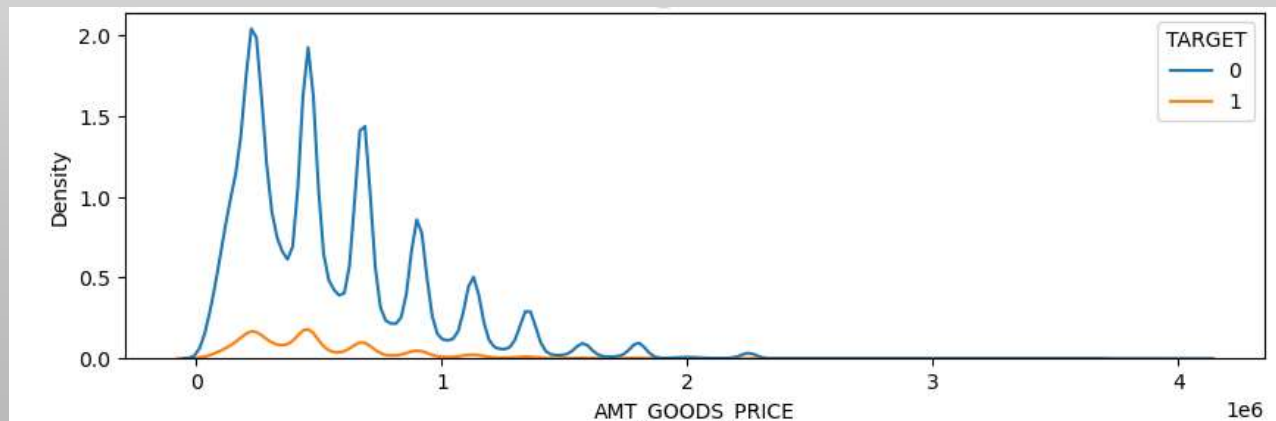


- Low-Skill Laborers and drivers are highest defaulters
- Accountants are less defaulters
- Core staff, Managers and Laborers are safer to target with a default rate of  $\leq 7.5$  to 10%

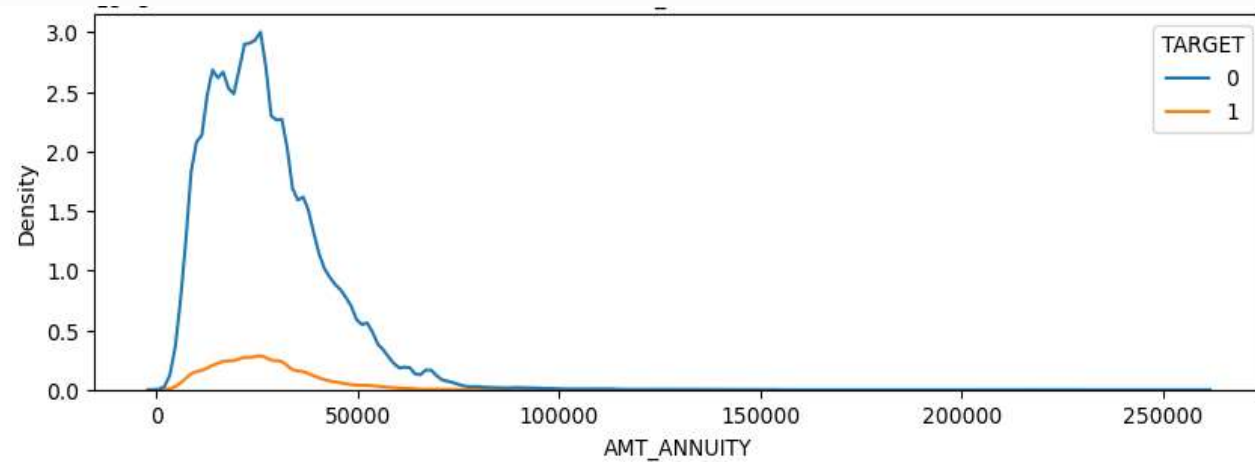
# UNIVARIATE NUMERICAL ANALYSIS



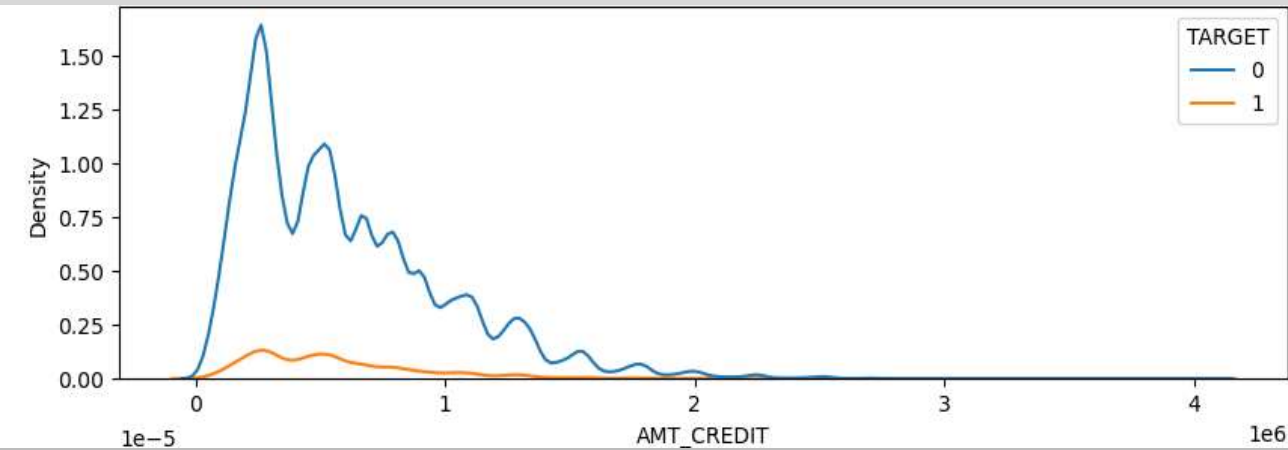
- mostly the customers have income between 0 to 1 ml



- most of the loans were given for the goods price ranging between 0 to 1 ml



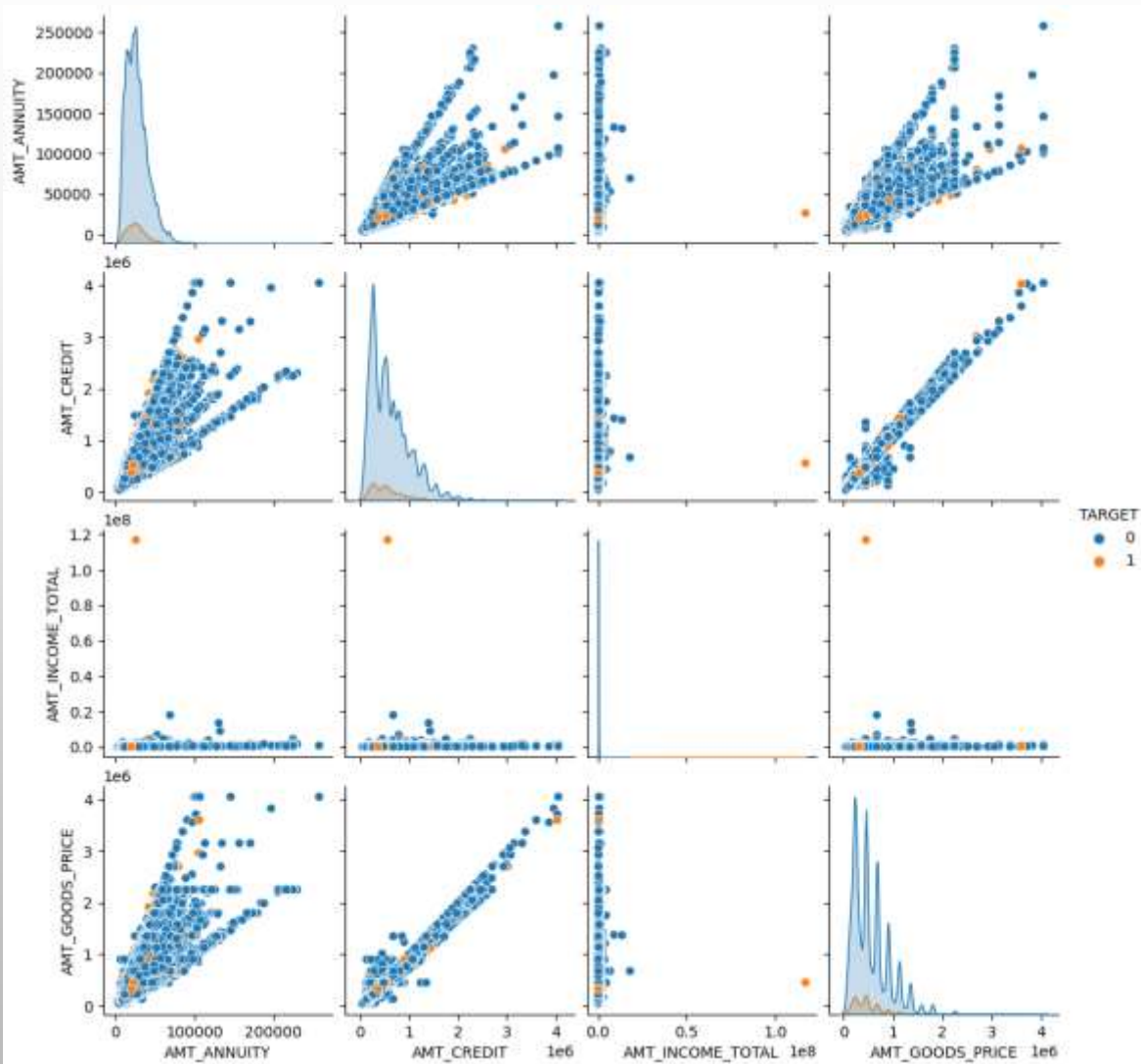
- most of the customers are paying annuity of 0 to 50 K



- most of the loans were given for the credit amount of 0 to 1 ml
- for  $\text{AMT\_CREDIT} > 750000$ , there are more clients with On-Time Payments

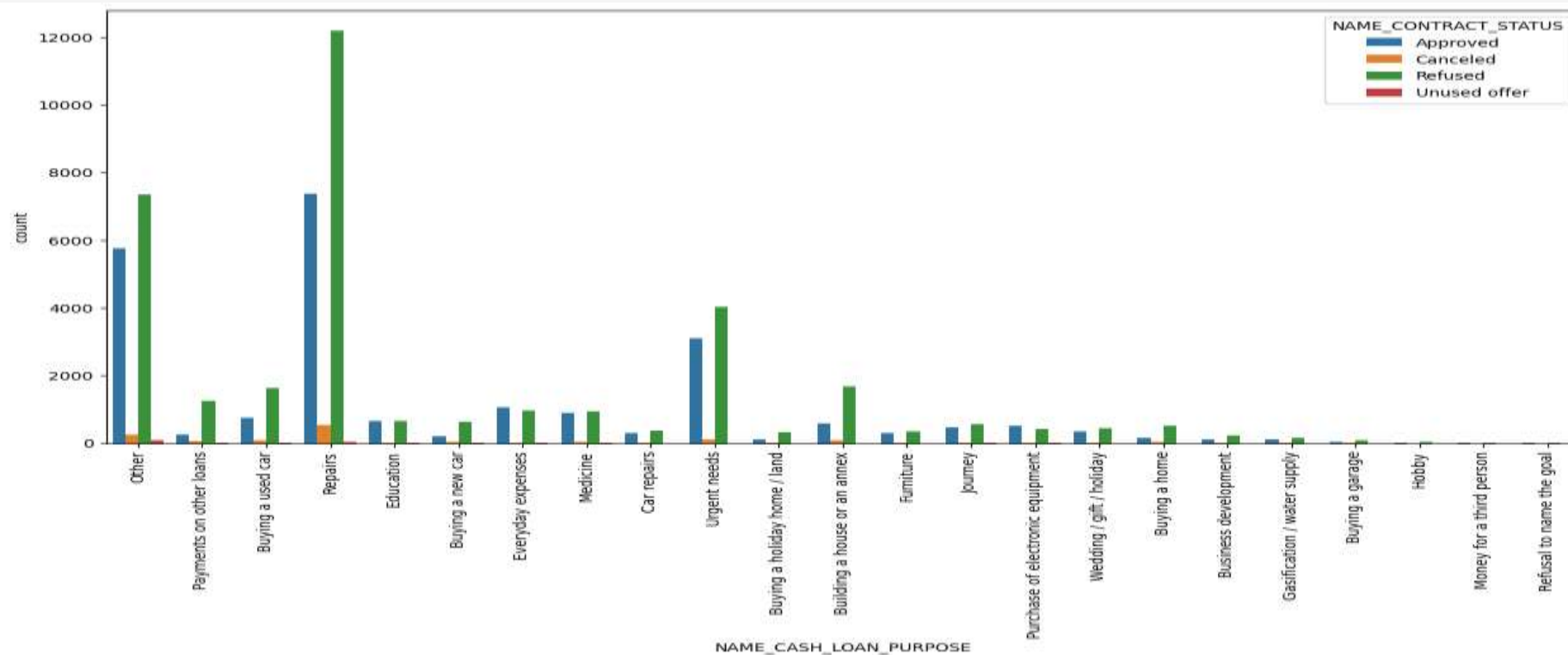


# BIVARIATE ANALYSIS

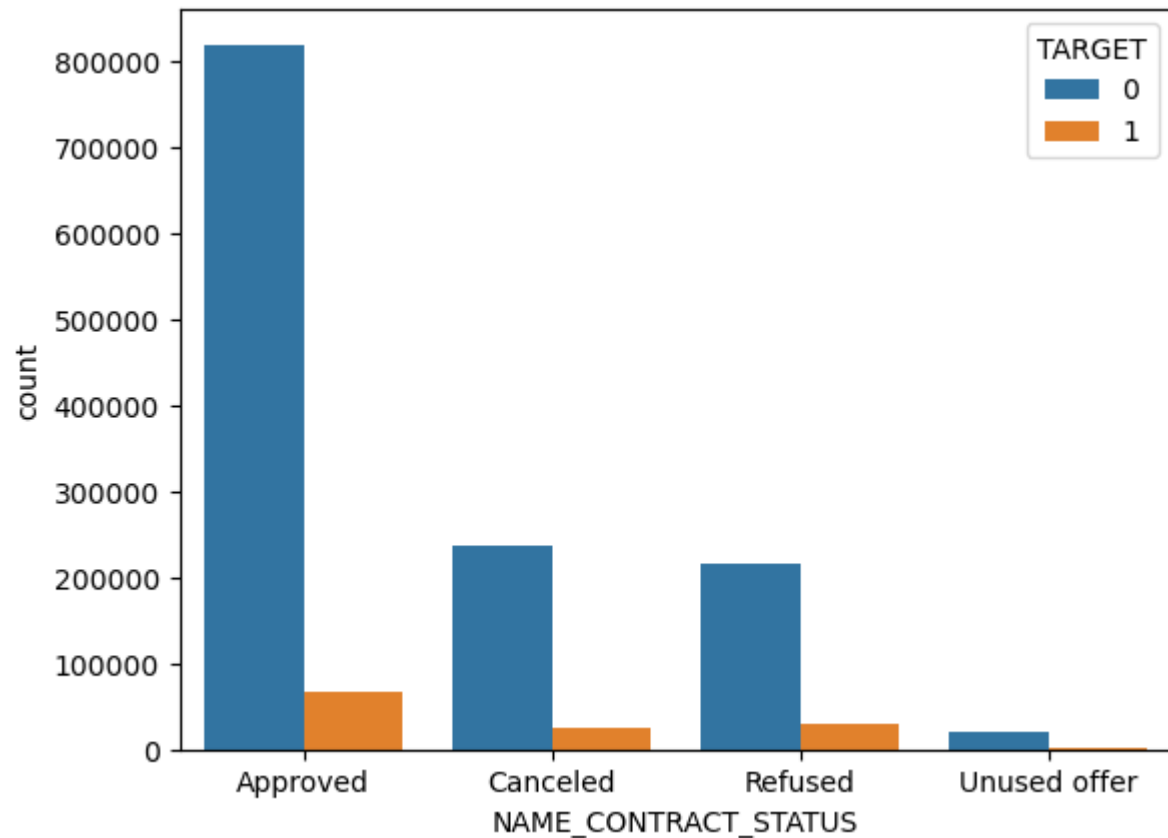


- AMT\_CREDIT and AMT\_GOODS\_PRICE are linearly correlated ,This means that as Goods price increases, so does Credit Amount\
- AMT\_ANNUIITY and AMT\_CREDIT have strong positive correlation ,This means that as annuity price increases, so does Credit Amount.

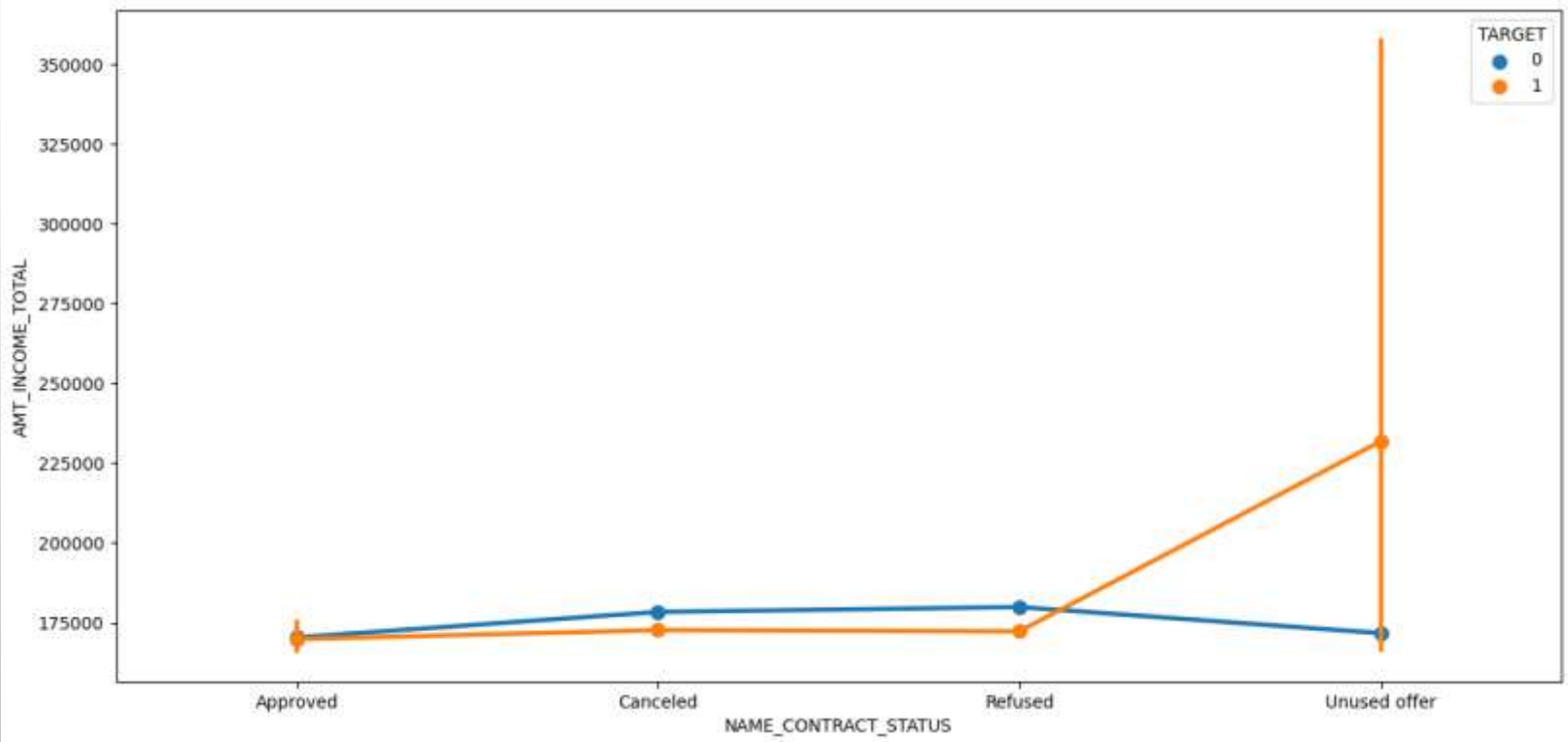
# MERGED DATA ANALYSIS



- repairing purpose clients had applied mostly previously and the same purpose has most number of cancelations and refused



- most of the client which were previous either canceled or refused 80-90% of them are not defaulter in the current data



- In previous data unused offer have maximum number of defaulters despite of having high income clients

# CONCLUSIONS

---

- Bank should target the customers having low income i.e. below 1 ml
- The credit amount should not be more than 1 ml
- Working as Accountants, Core staff, Managers and Laborers are targetable occupation for loans
- Female clients can be prefer than males for the loan approval
- Highly educated peoples have low defaults so they are good for the profit of bank
- Clients having house/apartment and are married and having children not more than 5 have also low Default rate so they are also preferable
- Low-Skill Laborers and drivers should be avoided
- Previous unused and high income customer should be avoided

# CONCLUSIONS

---

- org. Transport type 3 should be avoided
- working in Others, Business Entity Type 3, Self Employed org. type
- Working as Accountants, Core staff, Managers and Laborers are targetable occupation for loans
- income bracket could be below 1 ml
- Student and Businessmen are preferable as low default rates
- Bank should Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customer in future do better analysis because 80-90% of the customer who were previous canceled/refused are re-paying customer.