EDA CASE STUDY:

CREDIT RISK ANALYSIS

HIMANSHI SAINI

INTRODUCTION

The finance company has to decide on loan approval based on the applicant's profile. Which controls the loss of business to the company and avoids financial loss for the company. The loan-providing company finds it hard to give loans to the people due to insufficient or non-existent credit history

Two types of risks are associated with the bank's decision:

1. If the applicant's is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. If the applicant is not likely to repay the loan, and is likely to default, then approving the loan may lead to financial loss for the company.

PURPOSE

This case study aims to analyze the patterns that 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 at a higher interest rate, etc.

Main purpose:-

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.

STEPS TO ANALYSE DATA

1. Data sourcing
2. Data – understanding
3. Data- cleaning
4. Check for data quality
5. Binning
6. Check for data imbalance
7. All types of Univariate analysis
8. All types of bivariate analysis
9. Correlation
10. Merge application data with the previous application data
11. Data analysis by univariate, bivariate, and correlation
12. Recommendations and risks

IDEAL ANALYSIS

WHICH TYPE OF DATA SHOULD CONTAIN FOR IDEAL ANALYSIS:

- ❑ Correct datatype
- ❑ No irregularities
- ❑ No missing values
- ❑ No outliers
- ❑ No typing mistakes
- ❑ Errors
- ❑ Duplicates
- ❑ Standardize the format of data
- ❑ Filter data

DATA QUALITY CHECK

The most important thing to do is to deal with missing or null values(NaN):

- To deal with missing values there are only two methods i.e to remove if columns have more than 50% of data missing or we can impute them by replacing them with the mean, median, and mode of a given column

UNDERSTANDING DATA: COMPARE THE DATA NON-DEFAULTERS AND DEFAULTERS ON THE BASIS OF TYPES OF LOANS

TO UNDERSTAND WHICH KINDS OF LOANS ARE A PRIORITY FOR NON-DEFAULTERS AND DEFAULTERS :

- CHART-1:

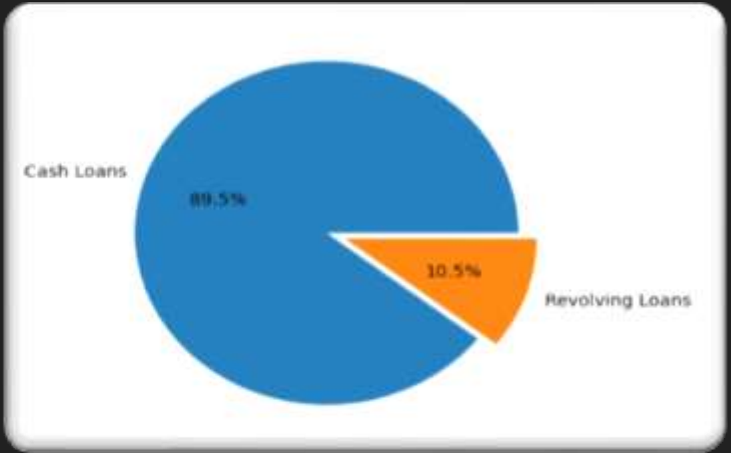
For NON-DEFAULTERS

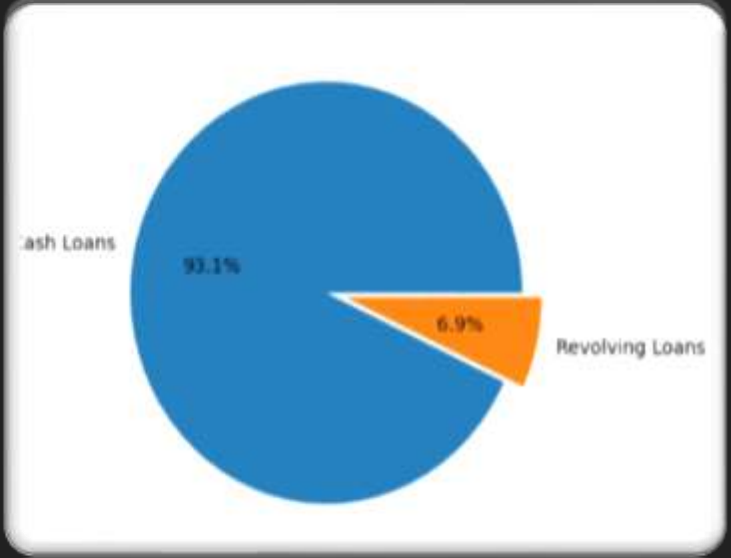
- CHART-2:

For DEFAULTERS

In both given pie charts we can see that for the current application process, cash loans are preferred over revolving loans.

The revolving loans are very few in comparison to cash loans





DATA BASED ON GENDER FOR NON-DEFAULTERS AND DEFAULTERS

COMPARISON OF DATA FOR NON-DEFAULTERS AND DEFAULTERS: -

- ❑ Chart-1:

For non- defaulters

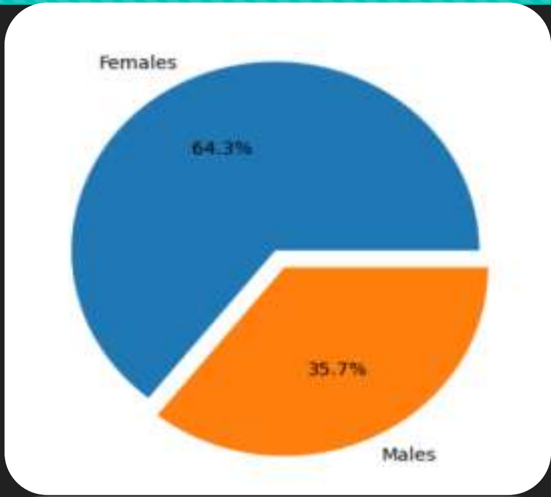
In this case, FEMALES are more non-defaulters for issuing the loan

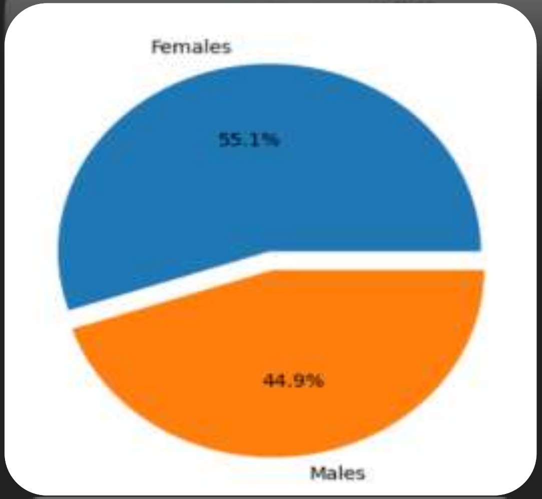
- ❑ Chart -2:

For defaulters

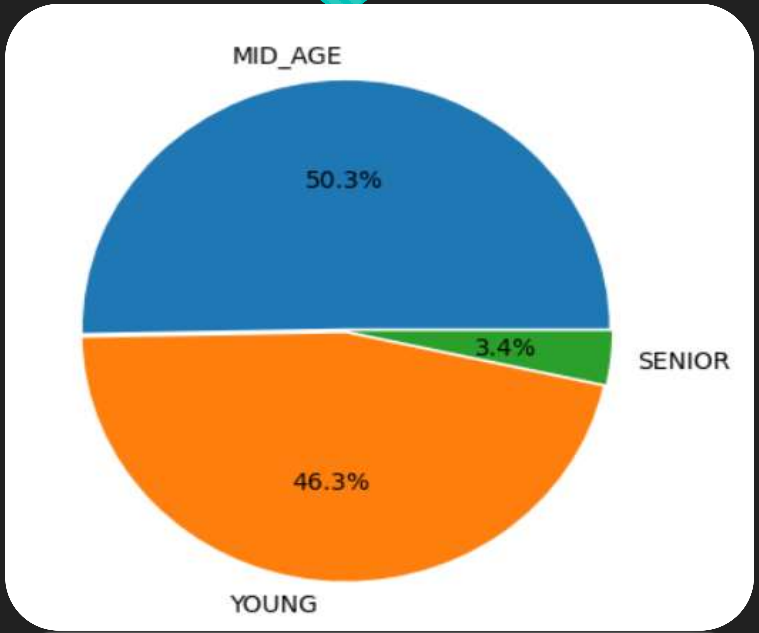
Here, females are more defaulted than males

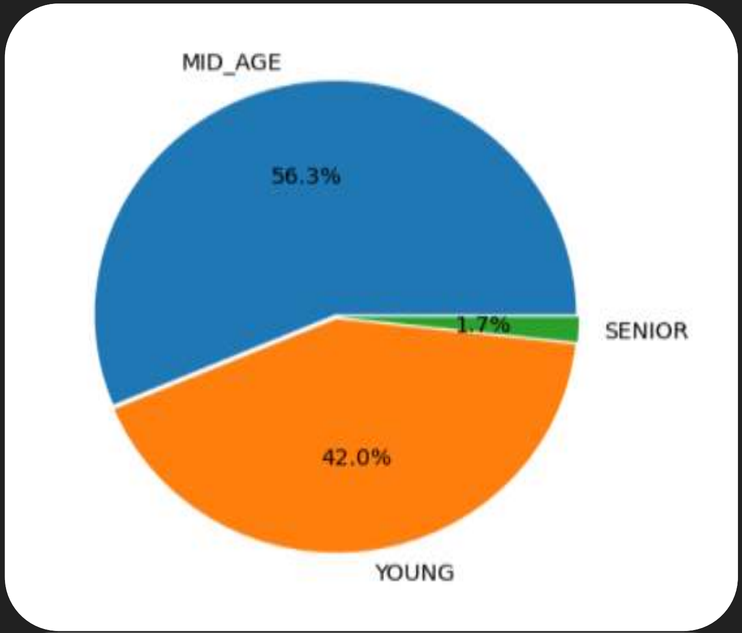
Hence, in both cases, males are less than females.





DATA VISUALISATION BASED ON AGE GROUP



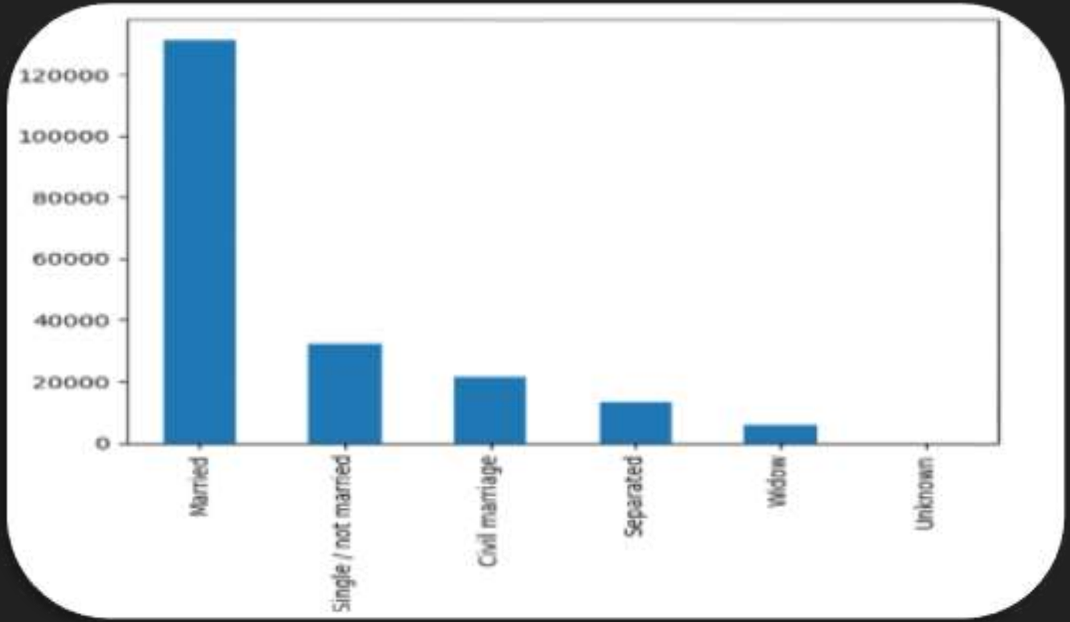


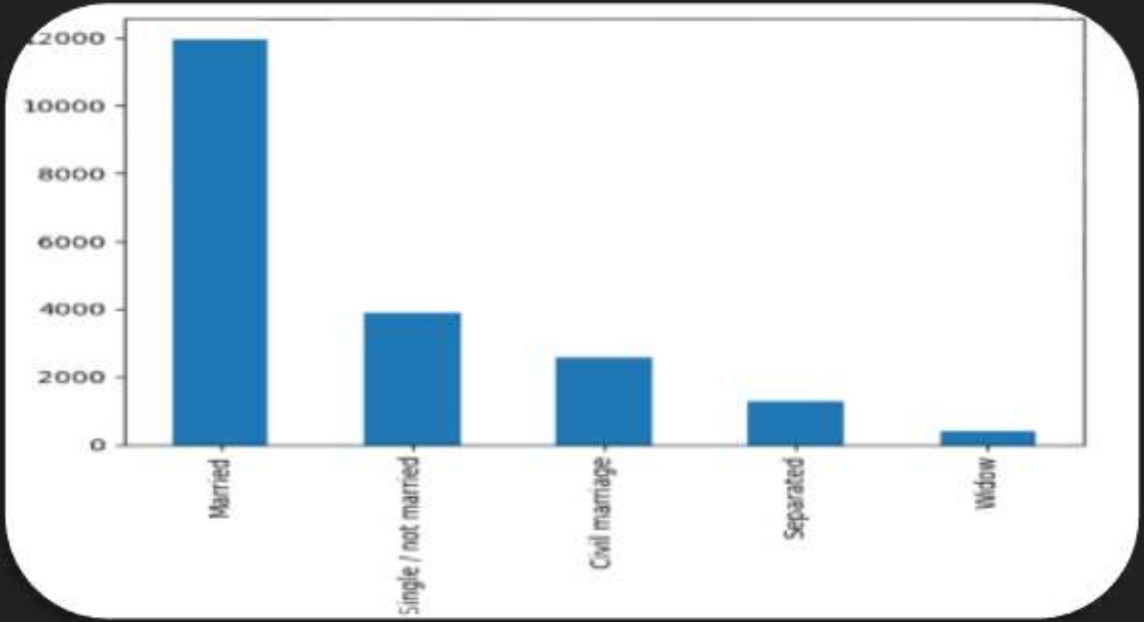
COMPARISON B/W NON-DEFAULTERS AND DEFAULTERS:

Here, “middle-aged” and “young” people are more for non-defaulters and defaulters also

Very few senior issued loans

VISUALISE THE DATA BASED ON TYPES OF MARRIAGES

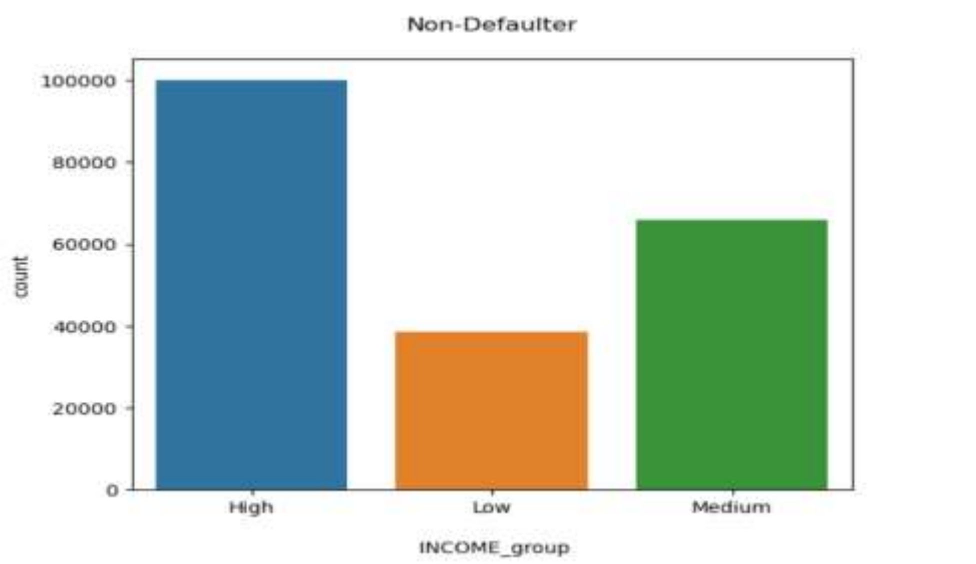


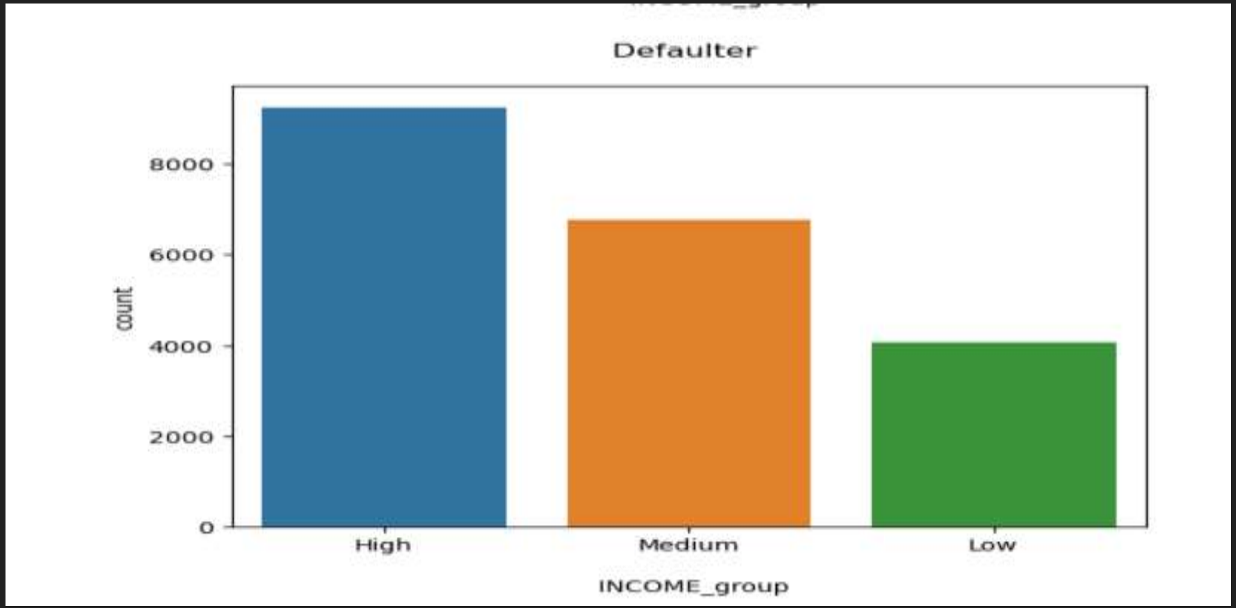


COMPARISON B/W NON- DEFAULTERS AND DEFAULTERS:

In both cases, married people are more defaulters and non-defaulters than other categories of marital status for issuing loan

UNIVARIATE ANALYSIS: based on income group

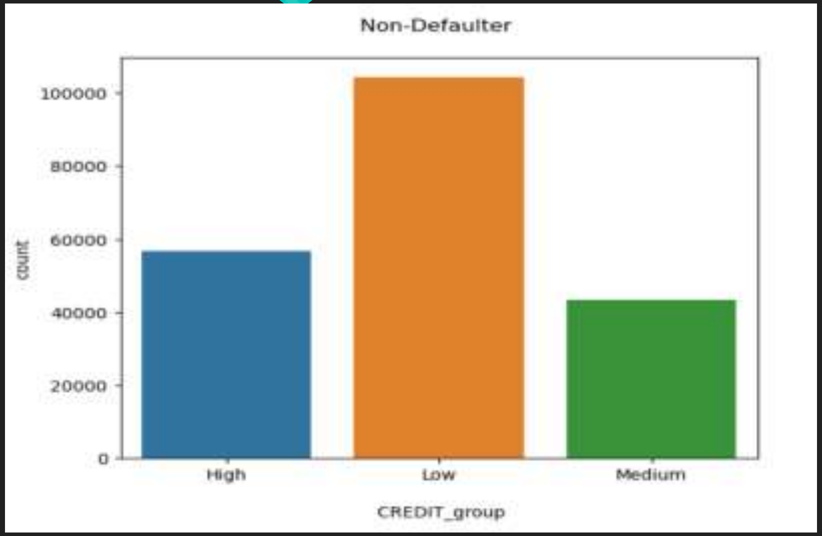


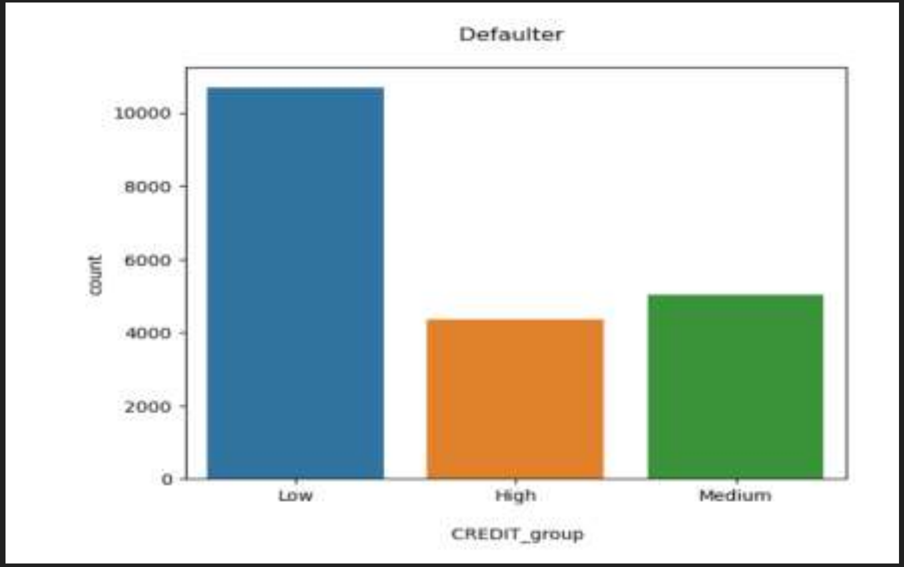


- **UNIVARIATE CATEGORICAL ANALYSIS BASED ON INCOME GROUP**:

In both cases high-income people have more non-defaulters and defaulters

CATEGORICAL UNIVARIATE ANALYSIS: based on credit group



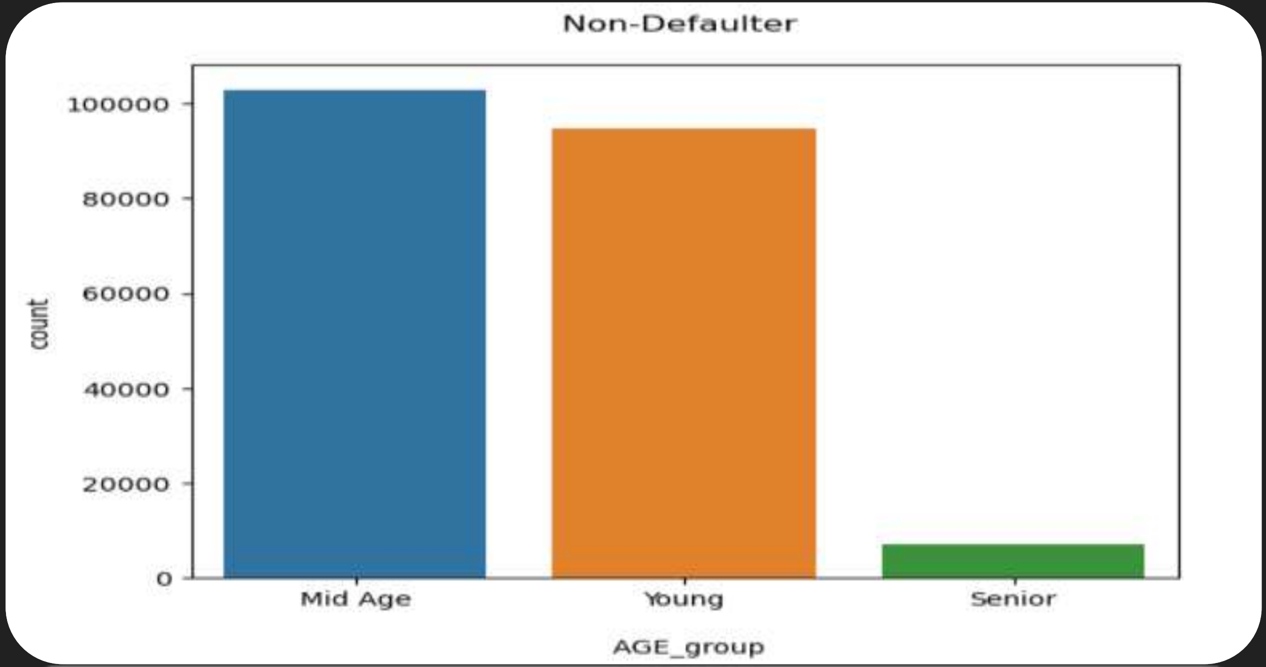


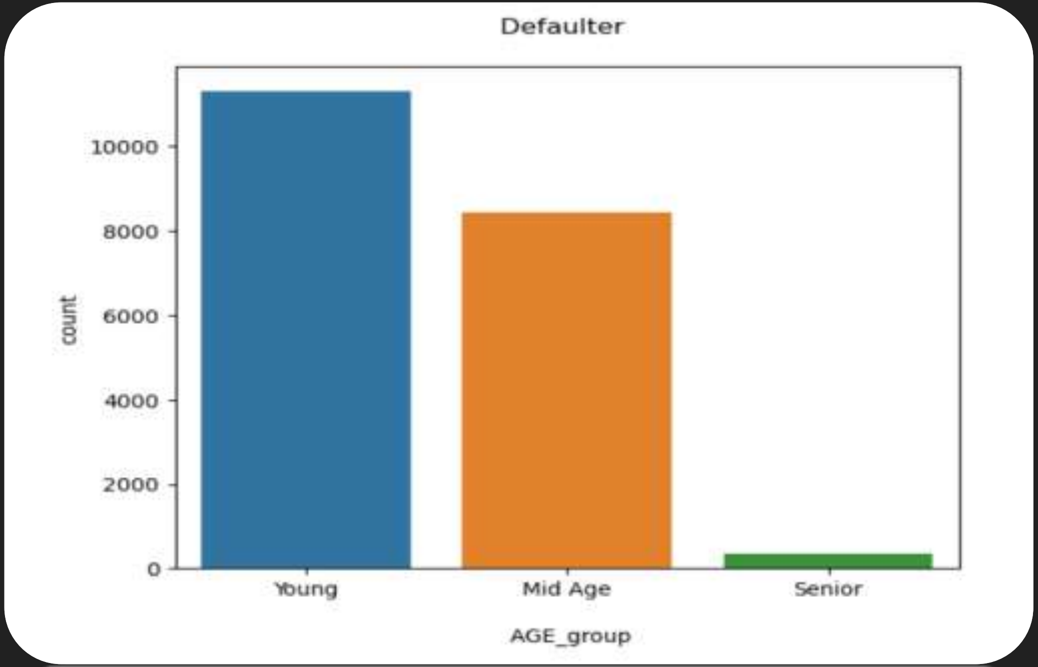
COMPARISON BASED ON CREDIT GROUP:

In both cases, people of low credit groups are defaulters and non-defaulters.

We can see that high credit group people are very few

UNIVARIATE ANALYSIS: based on age group

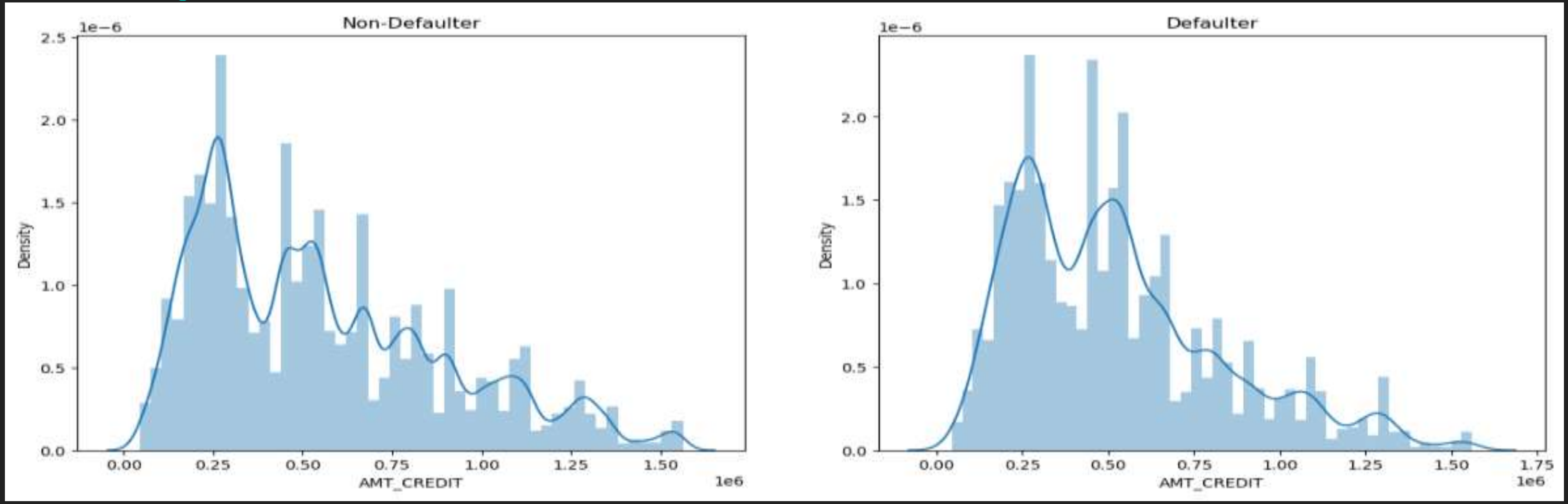




COMPARISON BASED ON AGE GROUP:

- For non-defaulters, MID-AGE people are more non-defaulters than senior and young
- For defaulters, YOUNG are more defaulters than mid-age or senior

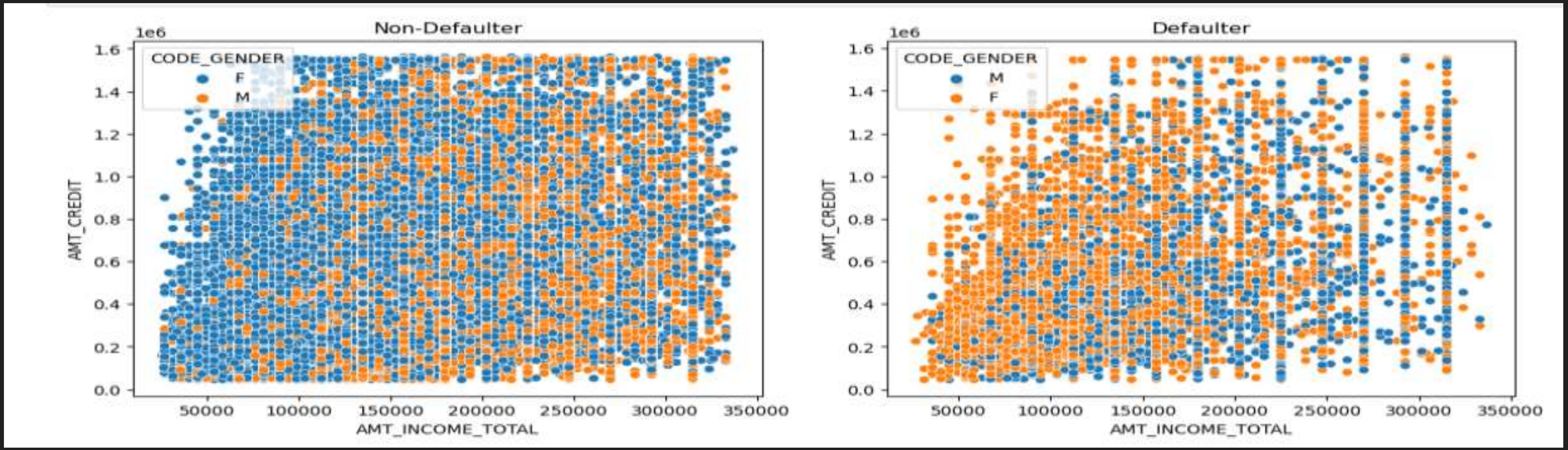
CONTINUOUS UNIVARIATE ANALYSIS: based on credit amount



COMPARISON BASED ON AMOUNT CREDIT:

- Here, In the first graph we can see the trend that, the possibility of being a NON-DEFAULTER is also lesser as the amount of credit loan is lesser.
- Now in the Second one, we can see that the possibility of being a DEFAULTER is more as the amount of credit loan is lesser.

CONTINUOUS BIVARIATE ANALYSIS: credit amount of loan on the basis of client income for both male and female



COMPARISON:

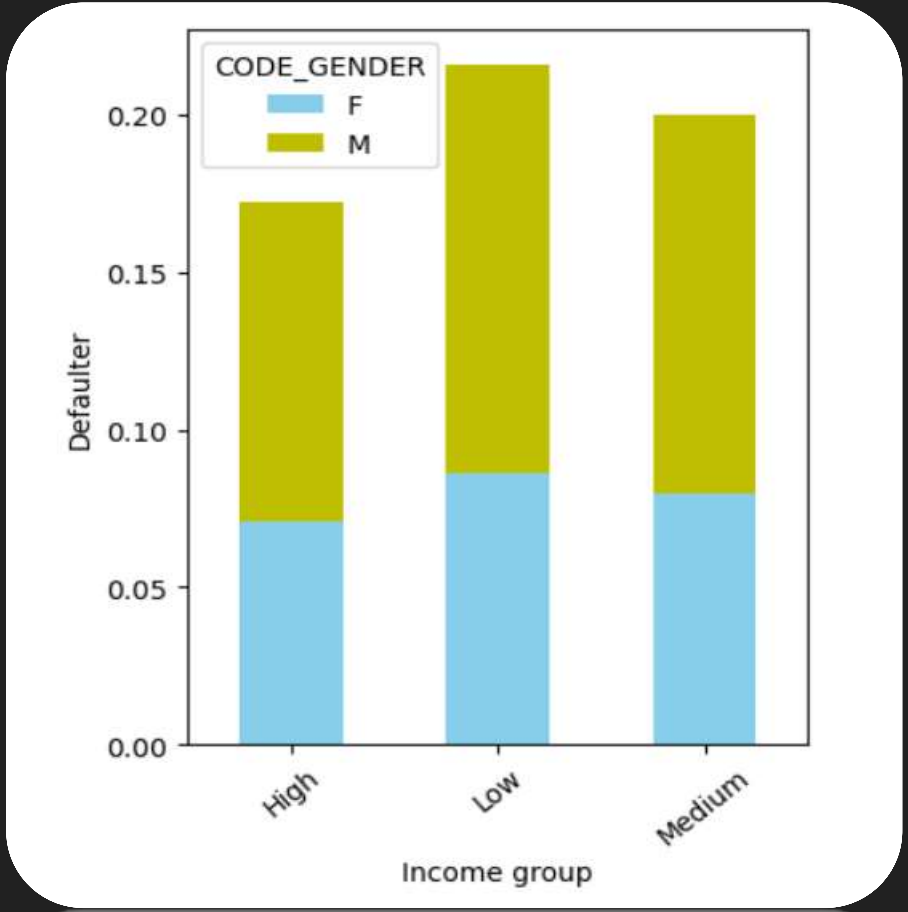
- For non-defaulters, we are unable to get any insights from this graph
- For defaulters, more values are concentrated more on lower income and lower credit. So, we can say that the credit amount is directly proportional to the total income amount for both genders.

TWO-SEGMENTED BIVARIATE ANALYSIS: based on gender for income group

COMPARISON:

We can see that MALES are more defaulted than FEMALES

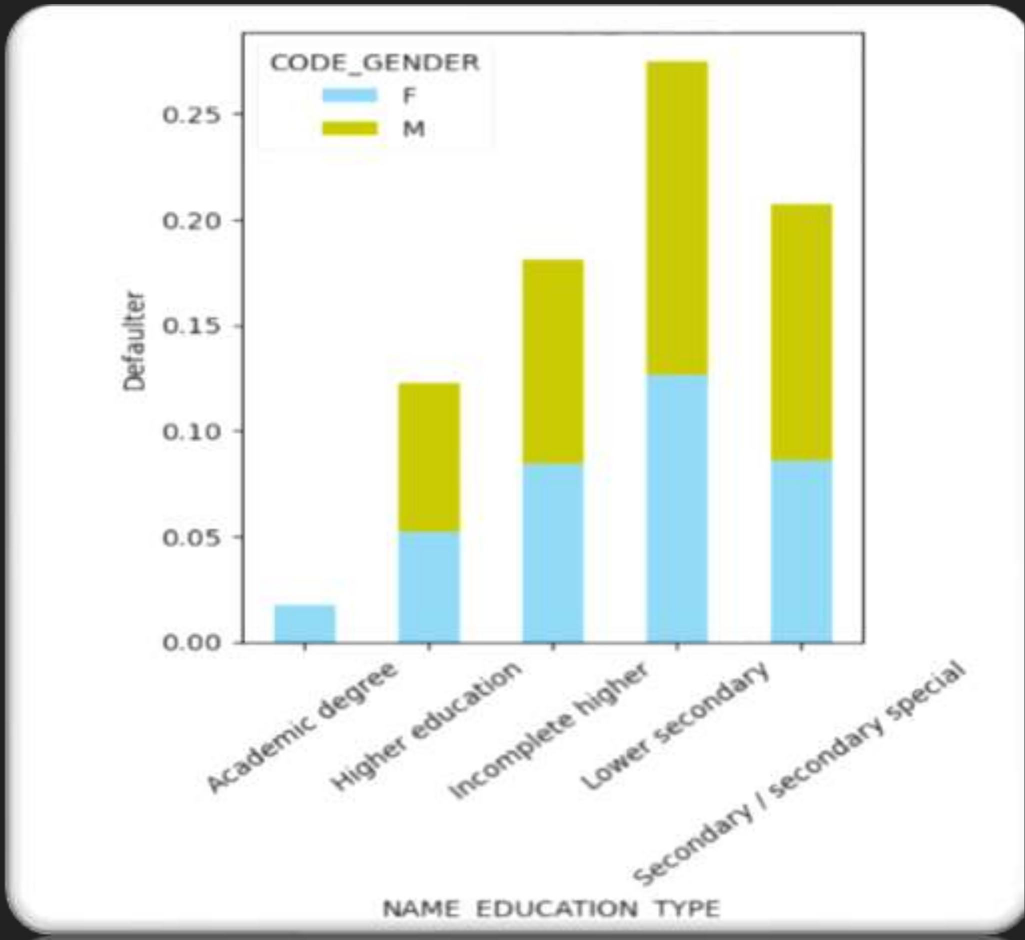
For all groups



TWO-SEGMENTED BIVARIATE ANALYSIS: based on gender for education type

COMPARISON:

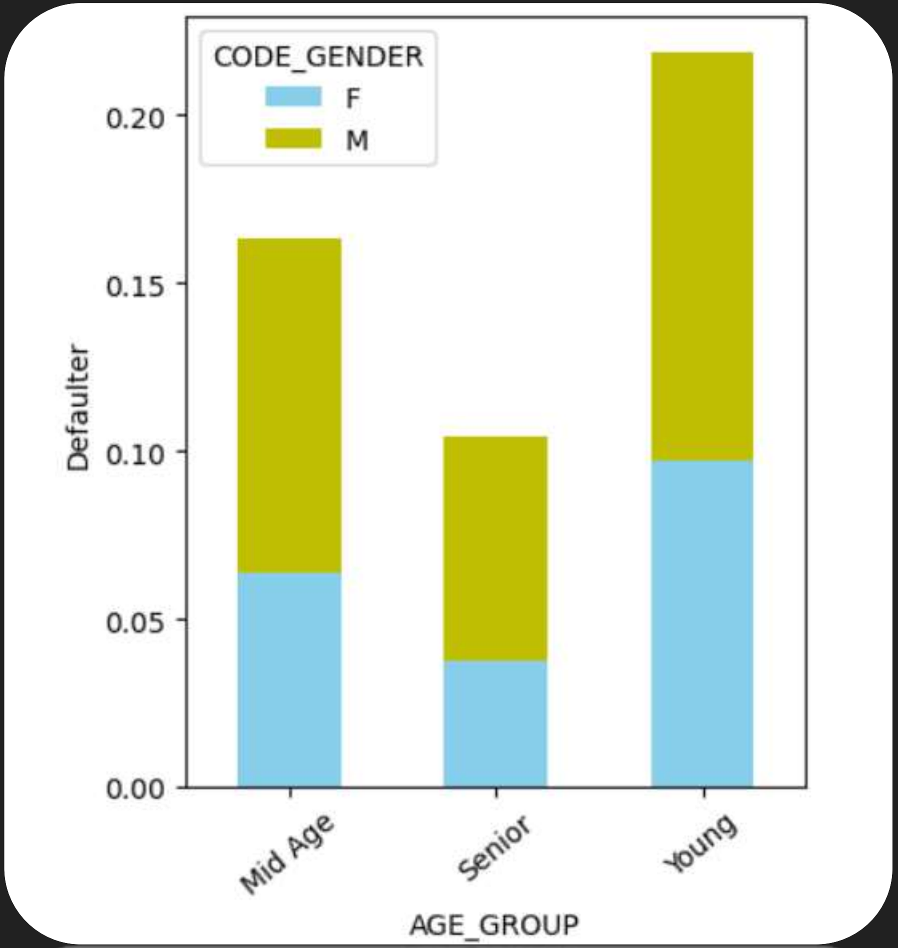
- Lower secondary educated clients are more defaulted
- The higher educated people are less defaulted
- Males are more defaulted than females



TWO-SEGMENTED BIVARIATE ANALYSIS: based on gender for age group

COMPARISON:

- ❑ Young clients are more defaulted with mid age clients.
- ❑ Females are less defaulted than males in each group



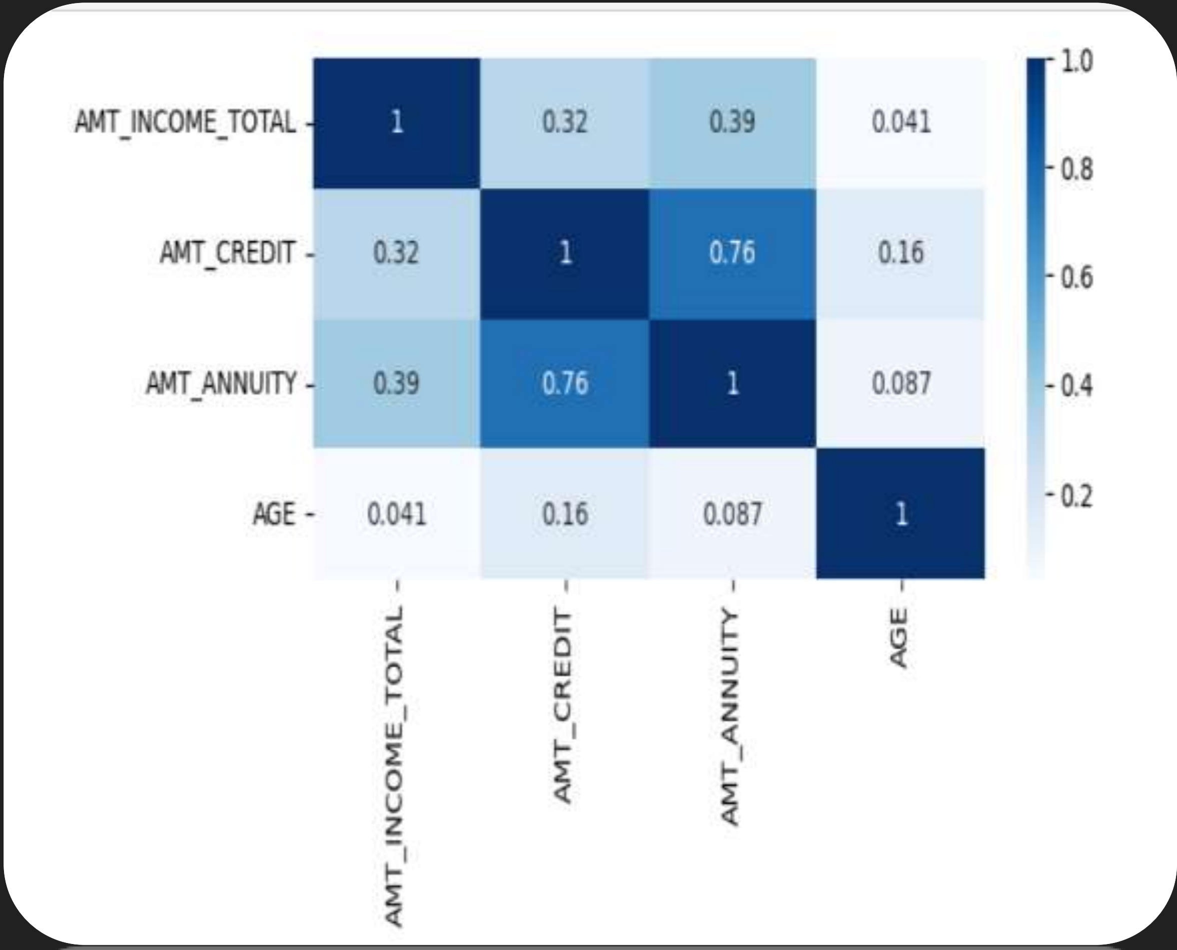
CORRELATION: for non-defaulters

HIGHLY CORRELATED VALUES:

- AMT_INCOME_TOTAL and AMT_CREDIT
- AMT_INCOME_TOTAL and AMT_ANNUITY
- AMT_CREDIT and AMT_ANNUITY

MODERATELY CORRELATED:

- AMT_INCOME_TOTAL and AGE
- AMT_CREDIT and AGE
- AMT_ANNUITY and AGE



	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AGE
AMT_INCOME_TOTAL	1	0.32	0.39	0.041
AMT_CREDIT	0.32	1	0.76	0.16
AMT_ANNUITY	0.39	0.76	1	0.087
AGE	0.041	0.16	0.087	1

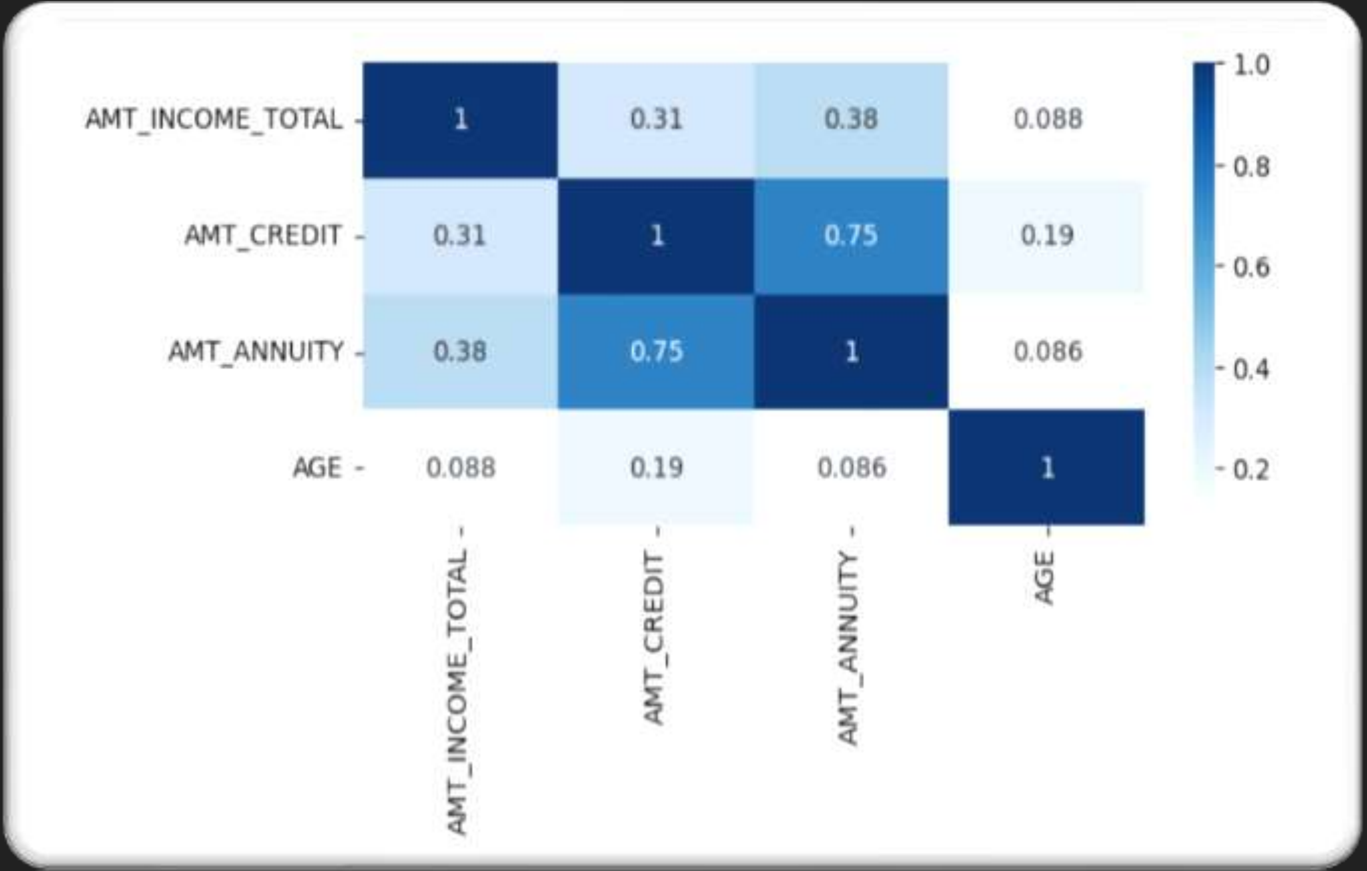
CORRELATION: for defaulters

HIGHLY CORRELATED:

- AMT_CREDIT and AMT_ANNUITY
- AMT_INCOME_TOTAL and AMT_ANNUITY
- AMT_INCOME_TOTAL and AMT_CREDIT

MODERATELY CORRELATED:

- AGE and AMT_INCOME_TOTAL
- AGE and AMT_CREDIT
- AGE and AMT_ANNUITY

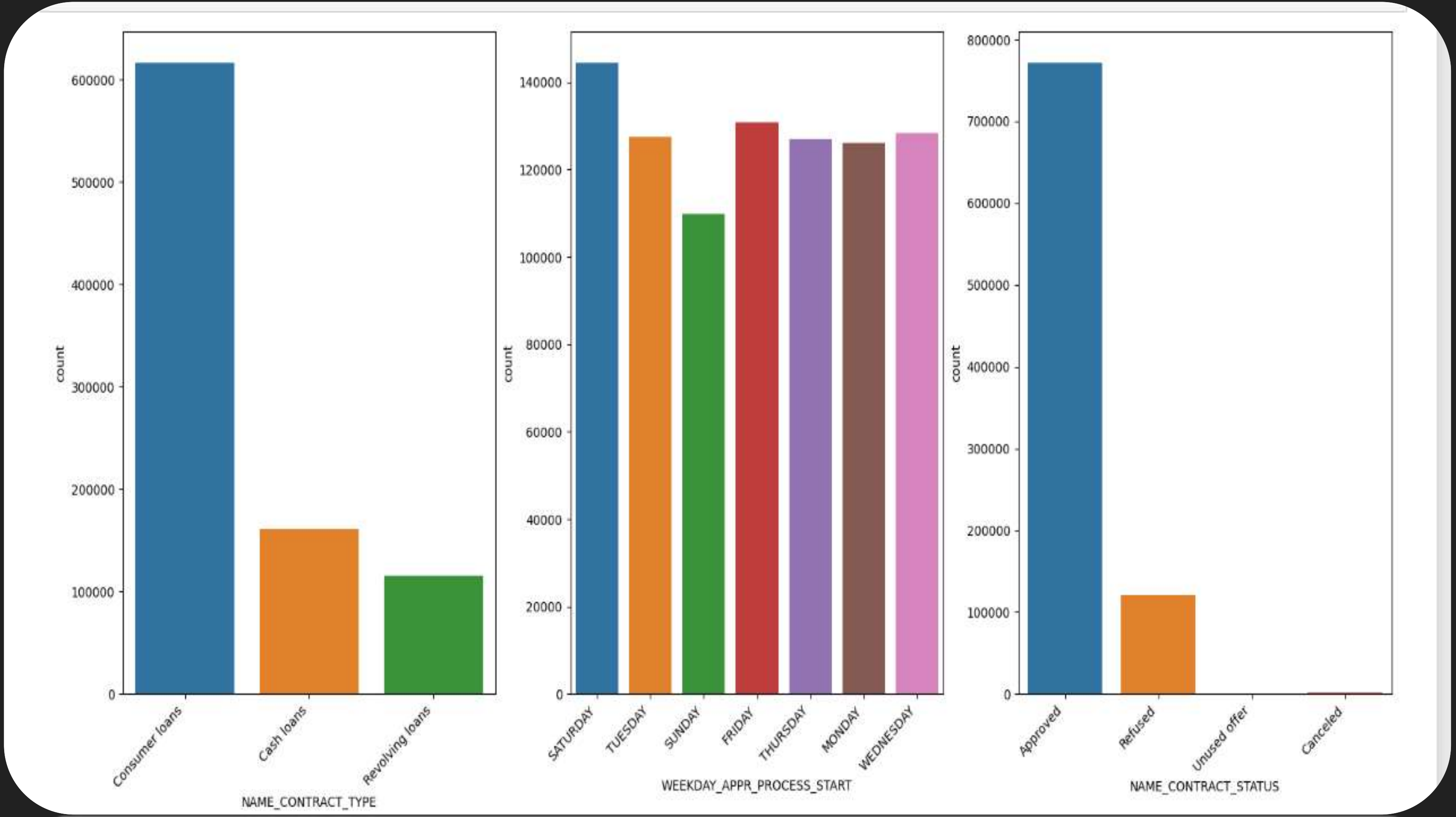


AFTER MERGING DATA: current application process with previous application process

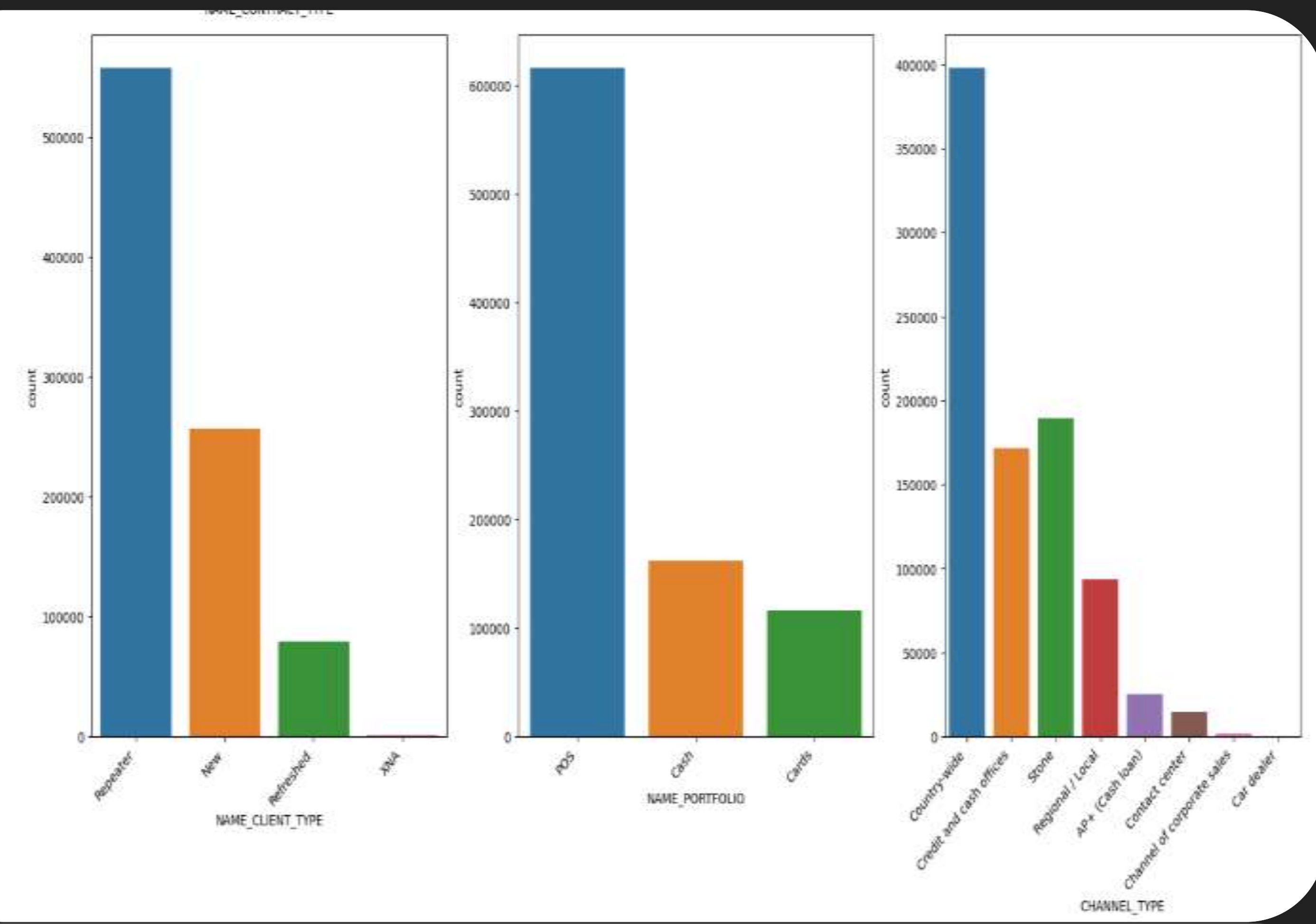
Do all the same processes after merging the data

UNDERSTANDING DATA

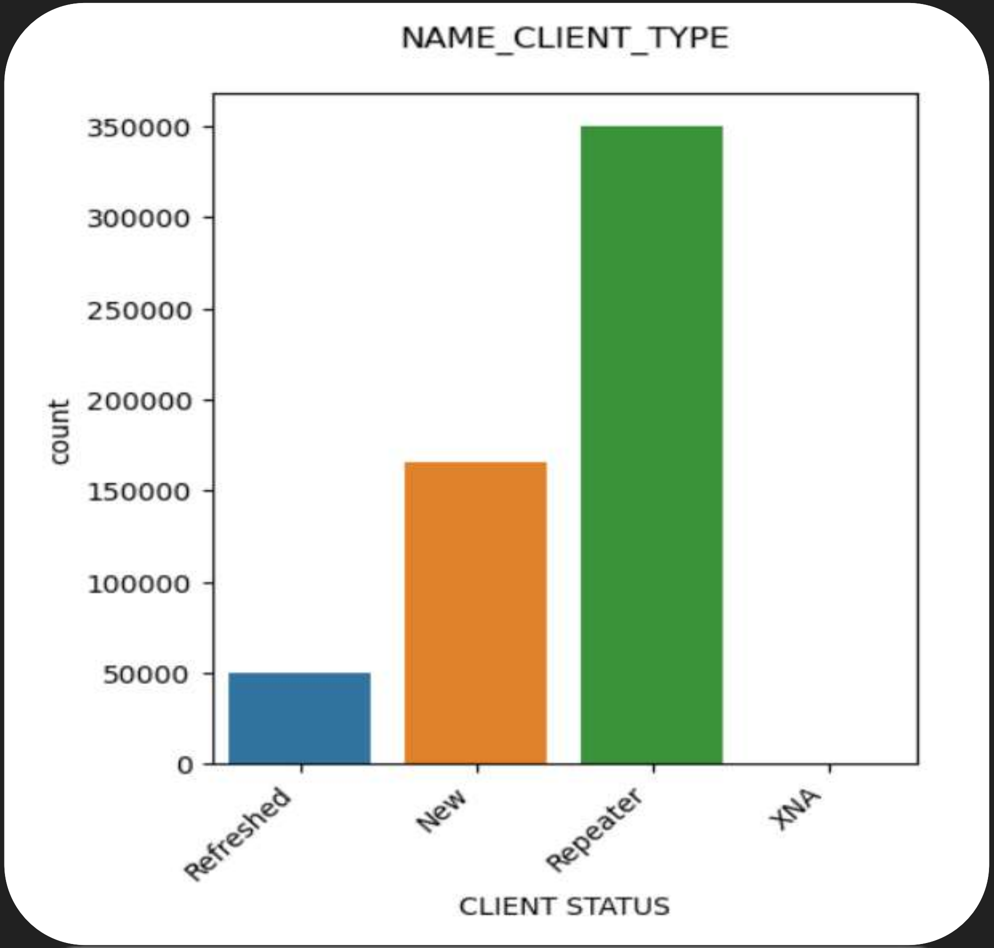
- **Consumer** and **Cash loans** are more preferred than others.
- based on the day for starting the application process **Sunday** has the least applicants
- Most of the loans are of **approved** category



- ❑ Most of the people are a **repeater** and **approved** for loans in the previous application process
- ❑ The most used channel type is country-wide followed by credit and stone type while others are rarely used.



UNIVARIATE CATEGORICAL ANALYSIS : based on client type



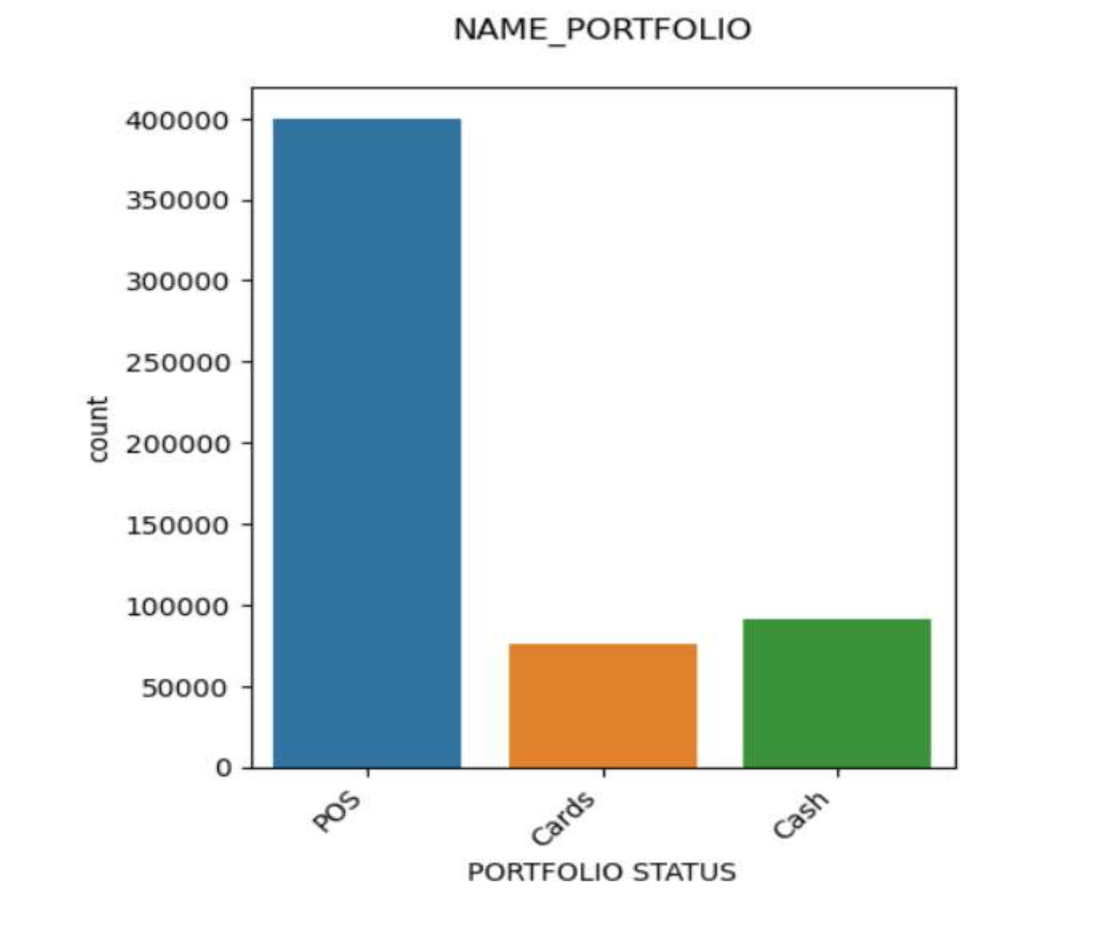
ANALYSIS:

Most of the clients are repeater

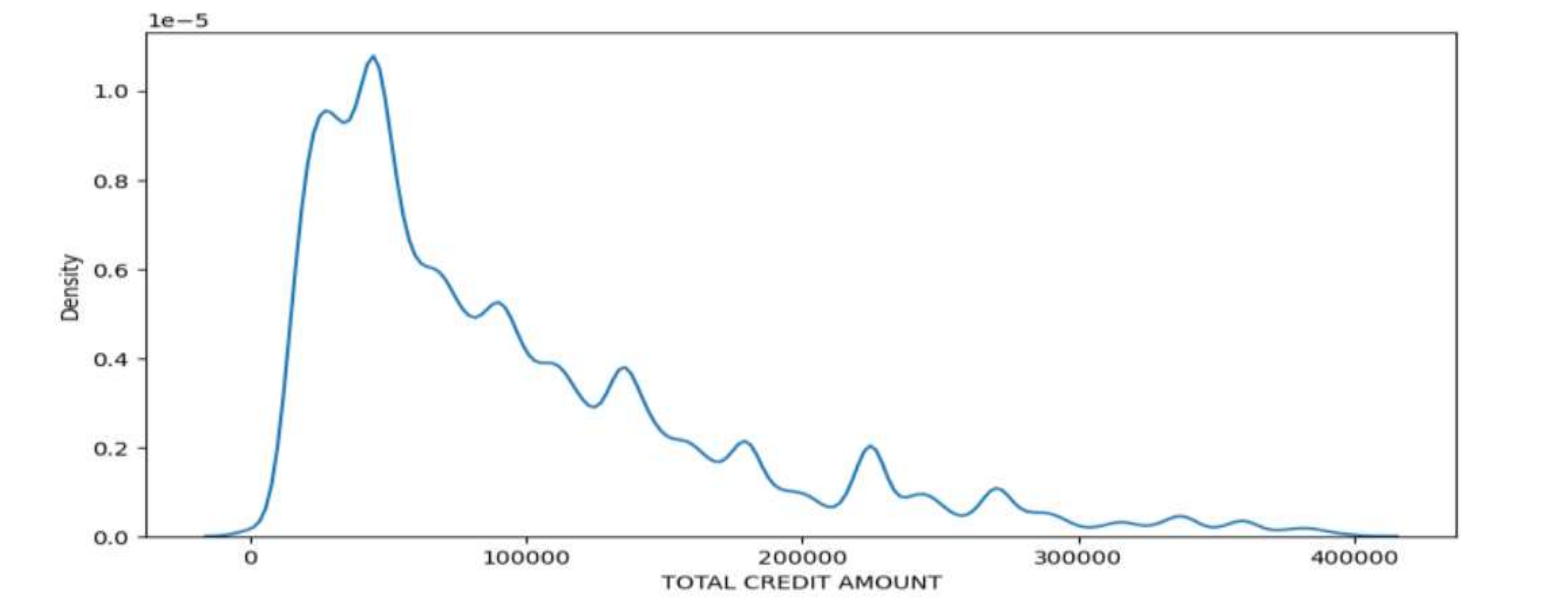
UNIVARIATE ANALYSIS: based on portfolio status

ANALYSIS:

Application for cards are very few. But highest for POS



CONTINUOUS UNIVARIATE ANALYSIS: FOR total credit



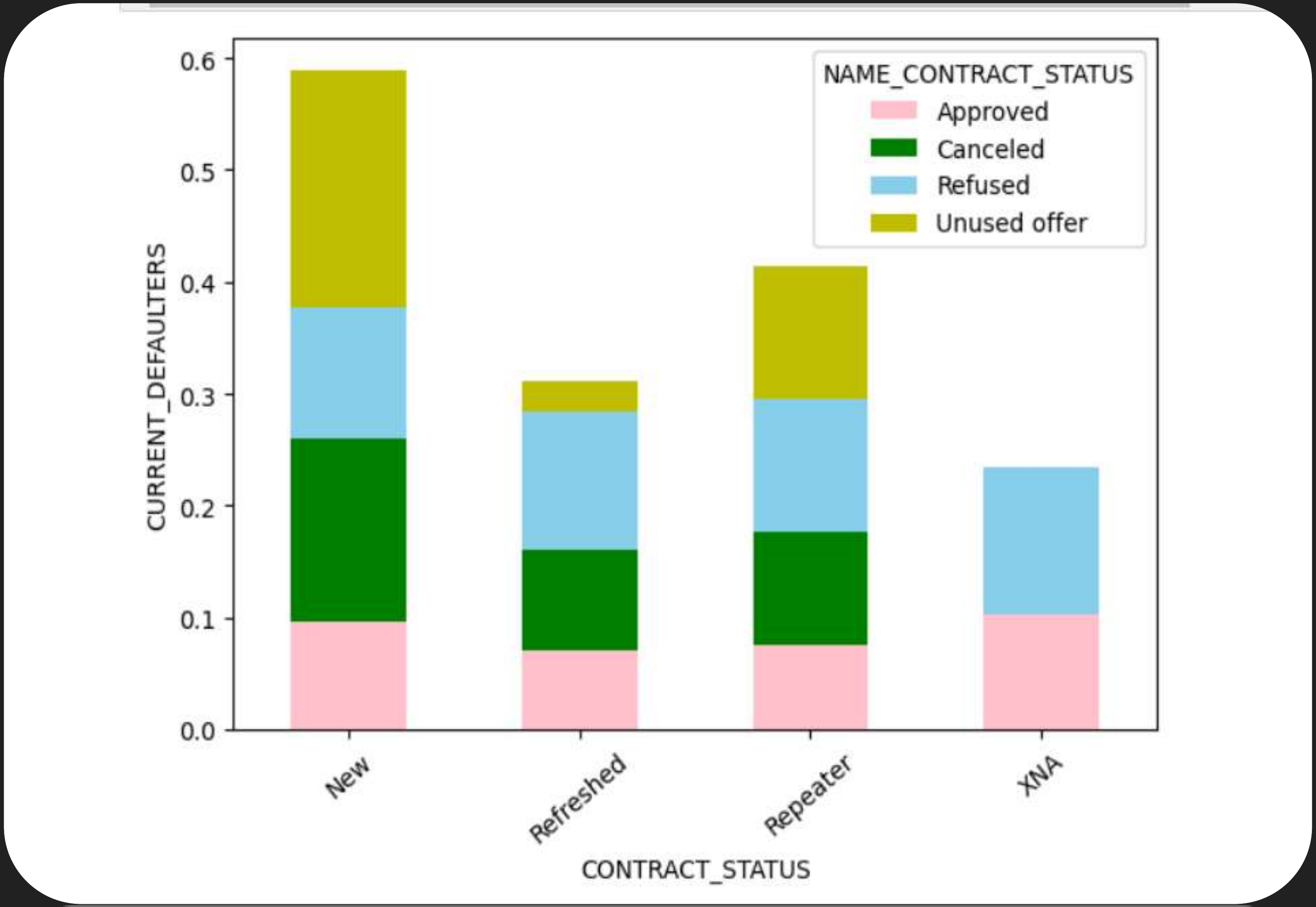
ANALYSIS:

The total credit amount is almost equal to 2,50,000-3,00,0000 and most of have 50,000-60,000

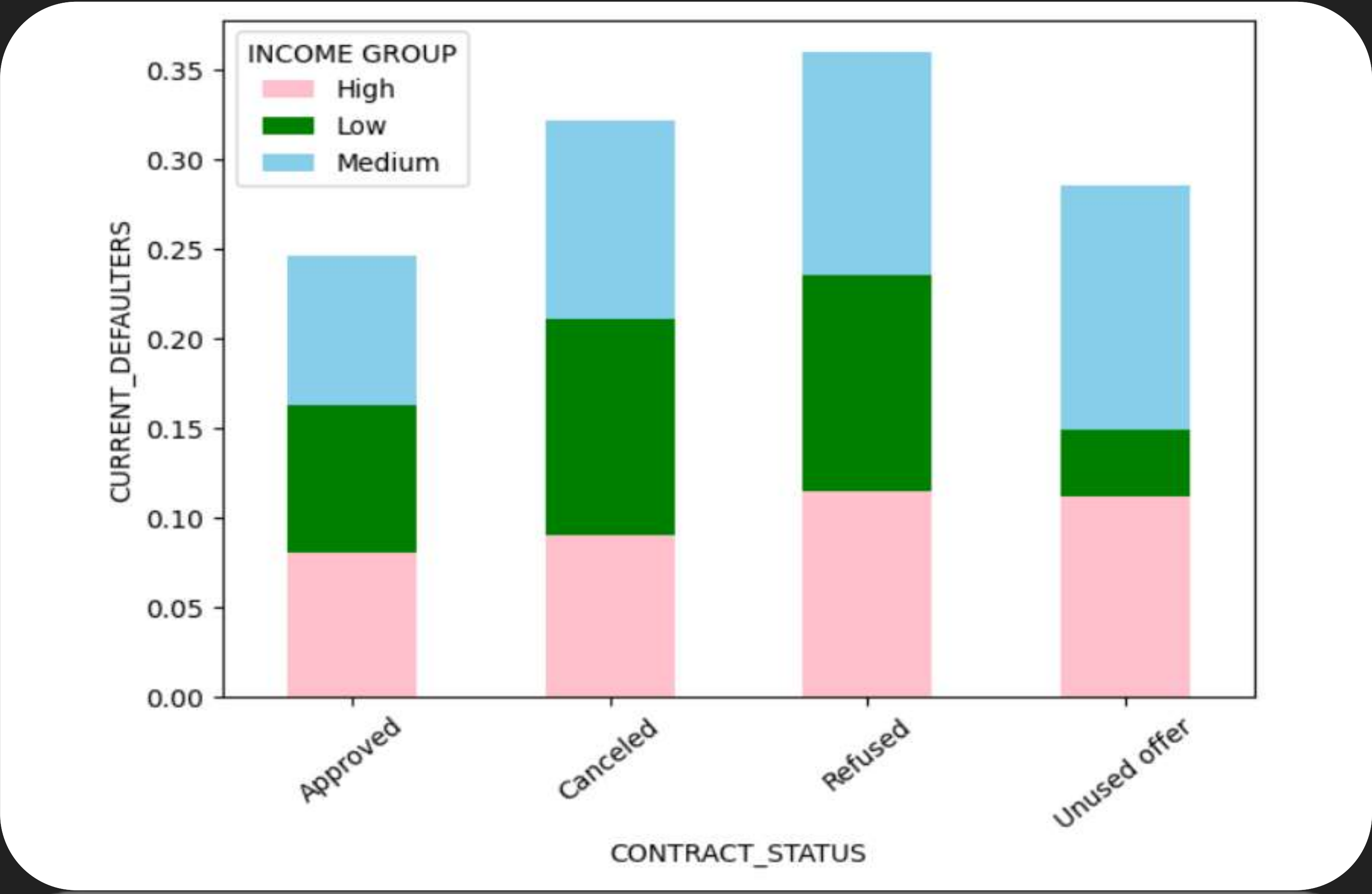
BIVARIATE CATEGORICAL ANALYSIS: current loan defaulter with respect to the previous loan application status and client types

ANALYSIS:

- Most people are NEW with mostly unused and cancelled offers.
- For the REPEATER CASE people have refused an unused offer
- Very few are present for the APPROVED category.



BIVARIATE CATEGORICAL ANALYSIS: current loan defaulter with respect to the previous loan application status and income group



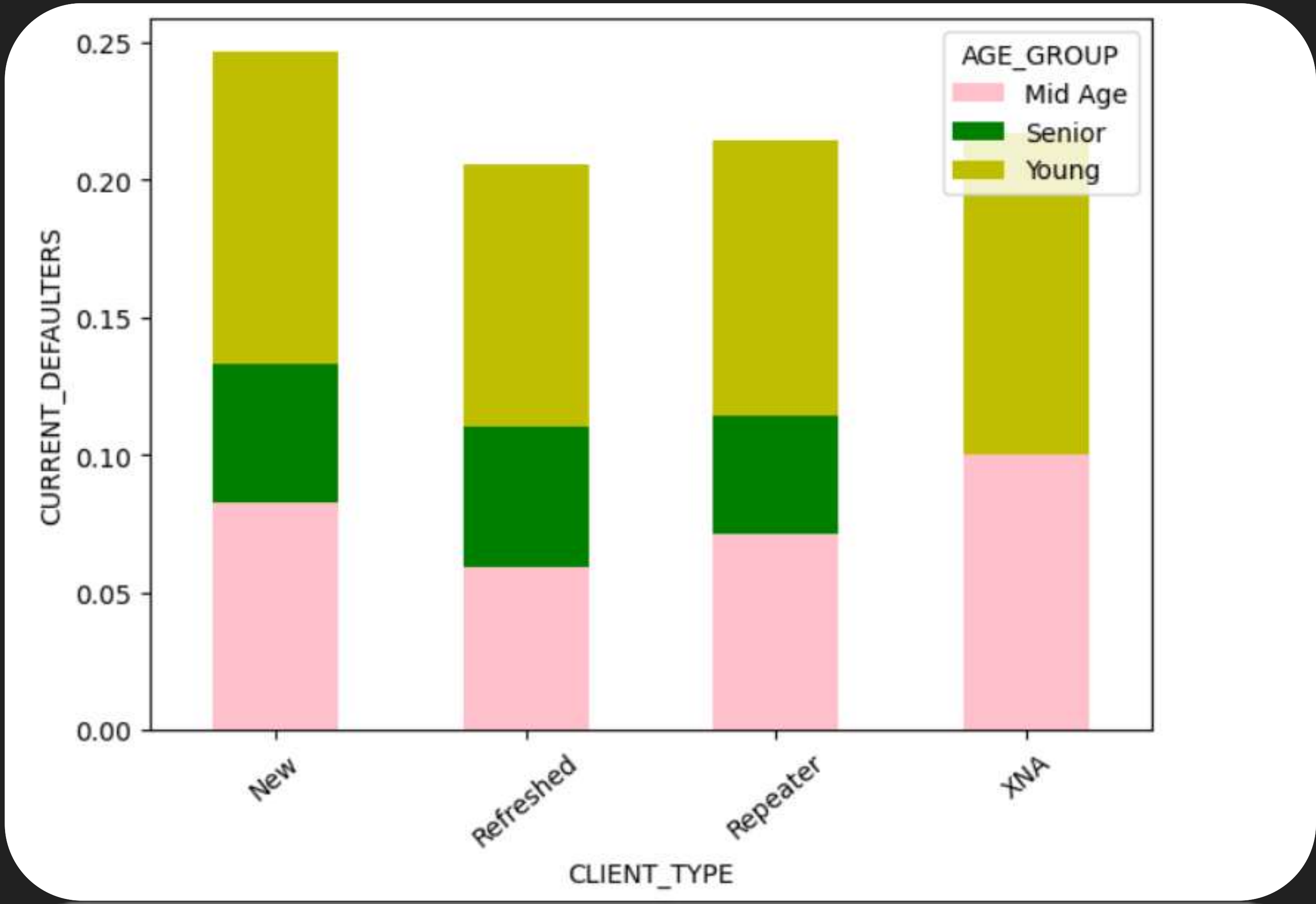
ANALYSIS:

- For the Previous application process REFUSED- category and approved category offer all income groups almost same
- But for UNUSED OFFER the least has a low-income group

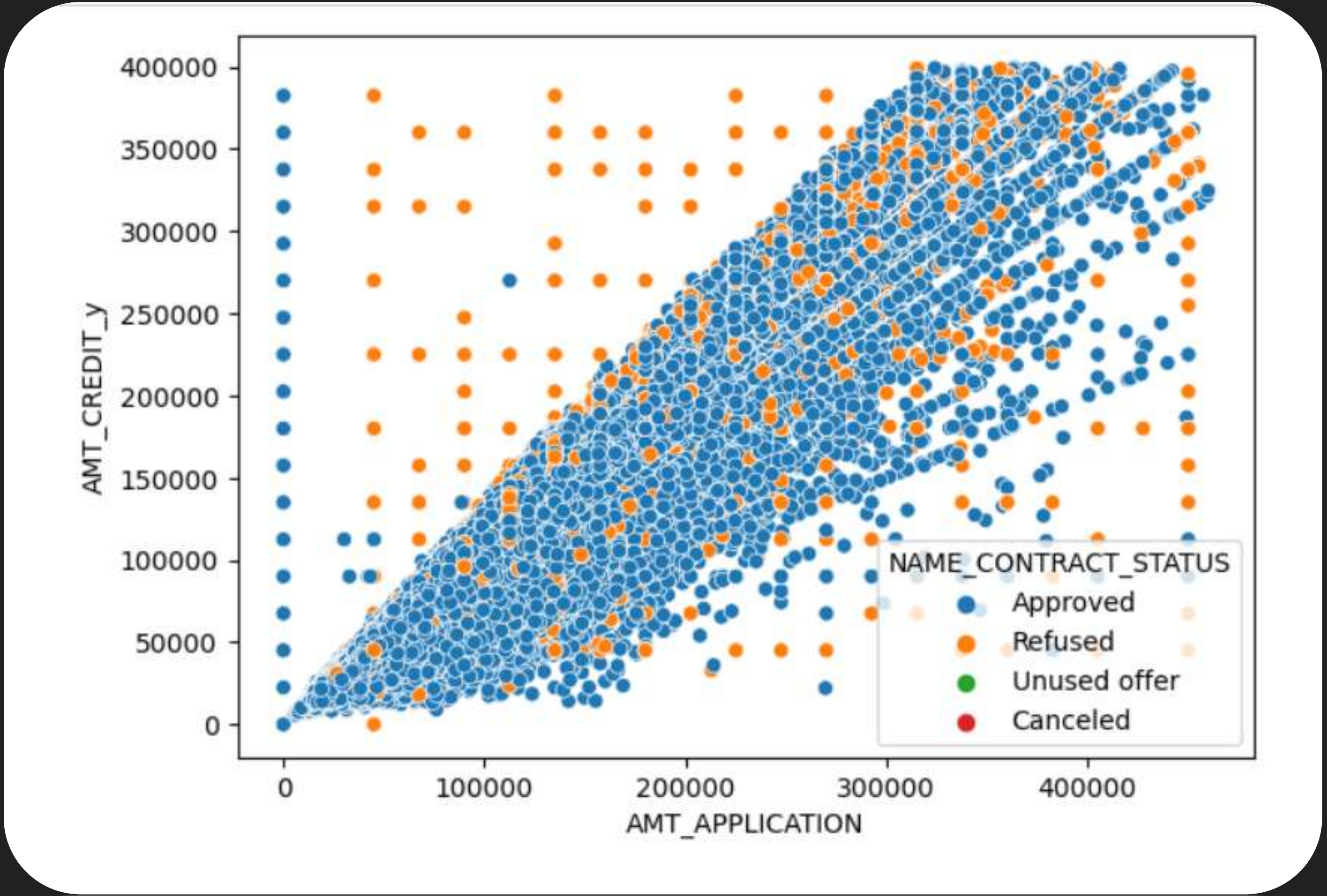
BIVARIATE CATEGORICAL ANALYSIS: current loan defaulter with respect to the previous loan application status and age group

ANALYSIS RESULT: -

- For all category NEW clients have more young people and least senior
- In all the categories young people are more defaulter



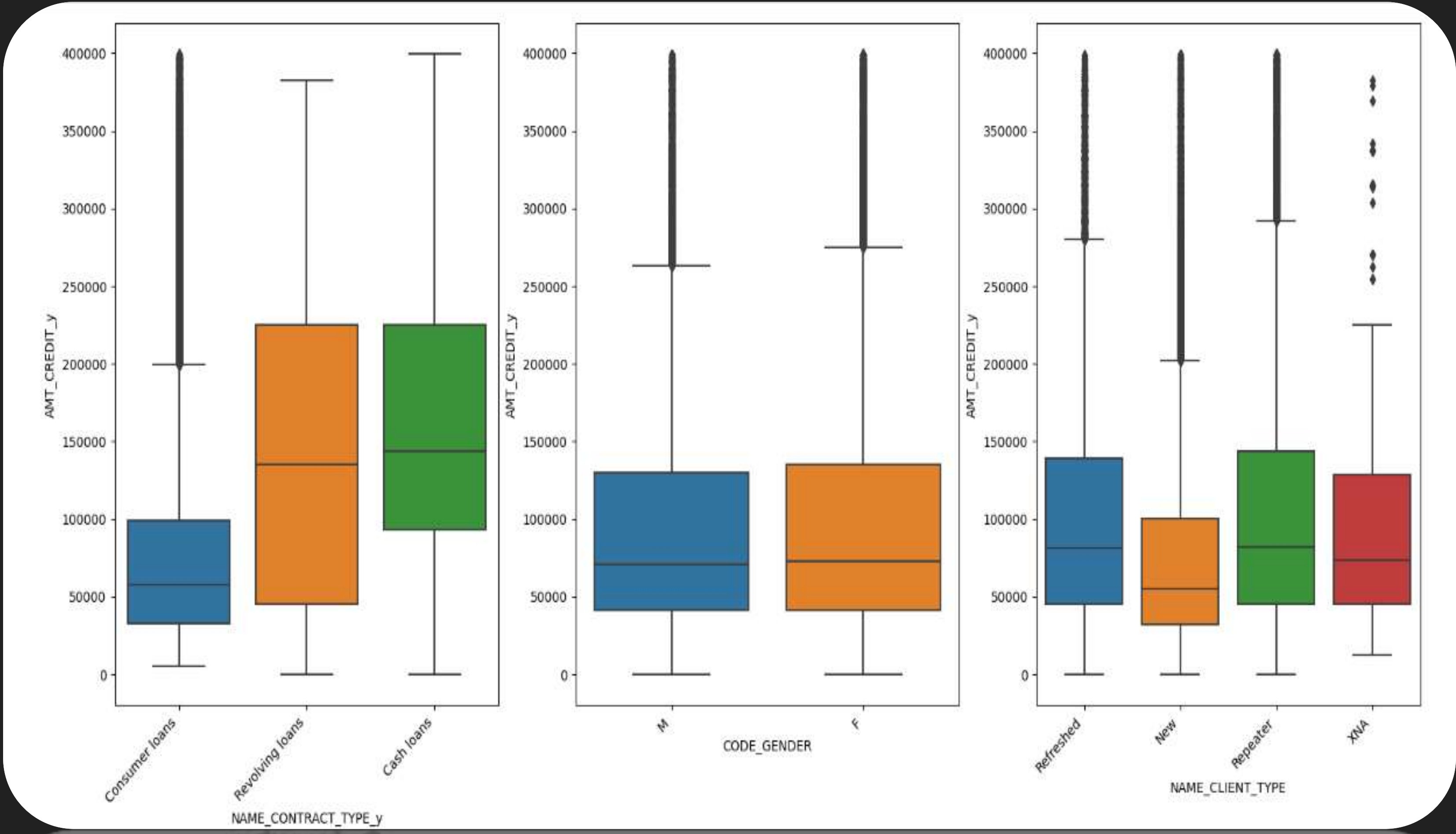
CONTINUOUS VARIABLE BIVARIATE ANALYSIS: application amount and credited amount



RESULT:-

- we can see most of the people are of an approved category
- We can see the credited amount is increased with the application amount

NUMERICAL-CATEGORICAL BIVARIATE ANALYSIS: with respect to the credit amount



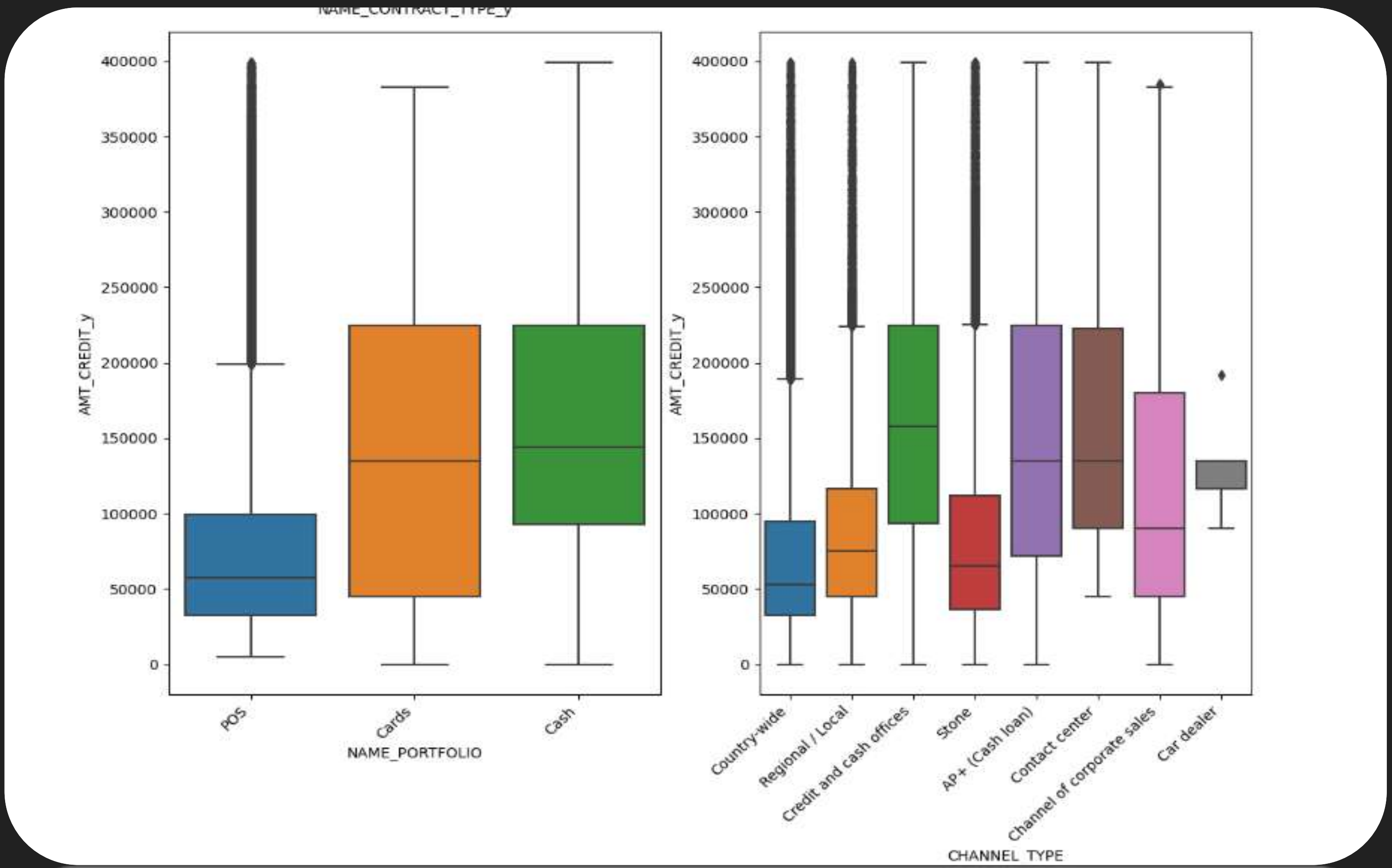
○ ANALYSIS RESULT: -

- The CASH LOANS are more credited than revolving and consumer loans
- Females have more credited amount than males.
- Repeater clients have more credits

NUMERICAL-CATEGORICAL BIVARIATE ANALYSIS: with respect to the credit amount

ANALYSIS:

- The amount credited in the previous application is max for the cards
- The credit amount of the loan is more from the application channel type as car dealer followed by Channel of corporate sales, Credit and cash offices, and Contact center. The amount is very low for Regional, Stone, and Country-wide channels.



CORRELATION

ANALYSIS RESULT: -

HIGHLY CORRELATED VALUES:

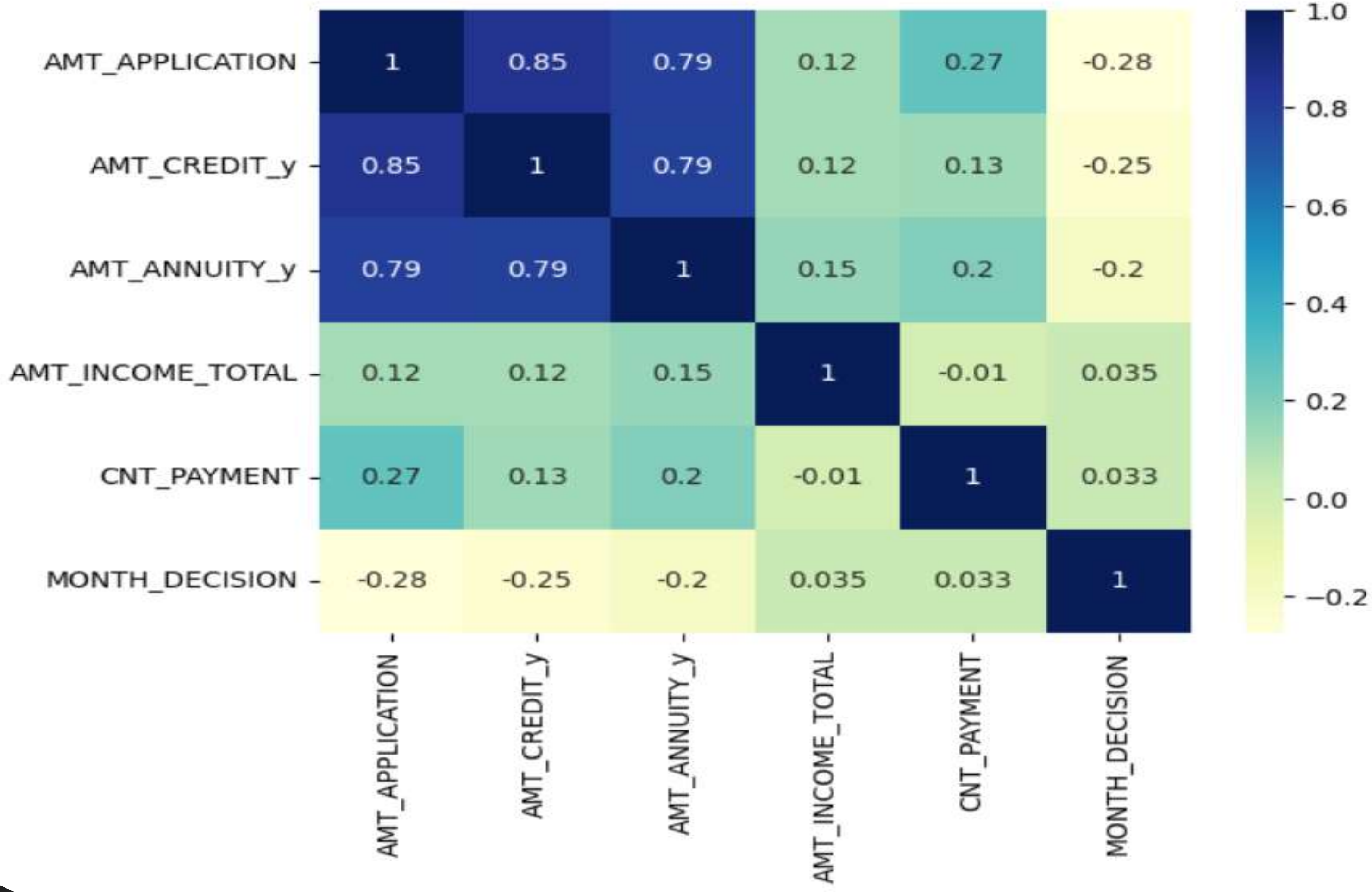
- AMT_APPLICATION and AMT_ANNUITY
- AMT_APPLICATION and AMT_CREDIT
- AMT_CREDIT and AMT_ANNUITY

MODERATELY CORRELATED:

- CNT_PAYMENT and AMT_APPLICATION
- CNT_PAYMENT and AMT_CREDIT

LEAST CORRELATED:

- AMT_APPLICATION and MONTH_DECISION
- MONTH_DECISION and AMT_CREDIT
- MONTH_DECISION and AMT_ANNUITY



	AMT_APPLICATION	AMT_CREDIT_y	AMT_ANNUITY_y	AMT_INCOME_TOTAL	CNT_PAYMENT	MONTH_DECISION
AMT_APPLICATION	1	0.85	0.79	0.12	0.27	-0.28
AMT_CREDIT_y	0.85	1	0.79	0.12	0.13	-0.25
AMT_ANNUITY_y	0.79	0.79	1	0.15	0.2	-0.2
AMT_INCOME_TOTAL	0.12	0.12	0.15	1	-0.01	0.035
CNT_PAYMENT	0.27	0.13	0.2	-0.01	1	0.033
MONTH_DECISION	-0.28	-0.25	-0.2	0.035	0.033	1

RECOMMENDATIONS

EDA for bank dataset recommends where the loan can be credited:

- Bank should lend more loans to females
- People with higher education
- According to age, older people are more default less
- Giving loans to married people is safer than singles
- People with higher income
- Client whose previous loan was approved
- Clients working as state servants

RISKS

RISKY GROUP:

- Lower secondary educated clients are the most in number to default when their previous loans were canceled or refused
- Male clients with civil marriage
- Previously refused loan status group.