

# EDA Case Study on Bank Loan

PRESENTATION BY:

**SHEKHAR ANAND**

# BUSINESS OBJECTIVES

1. The company wants to understand the driving factors 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.
2. Develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

# PROBLEM STATEMENT

1. Identify patterns and relationship 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.
2. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using different EDA approach.

# METHODOLOGY

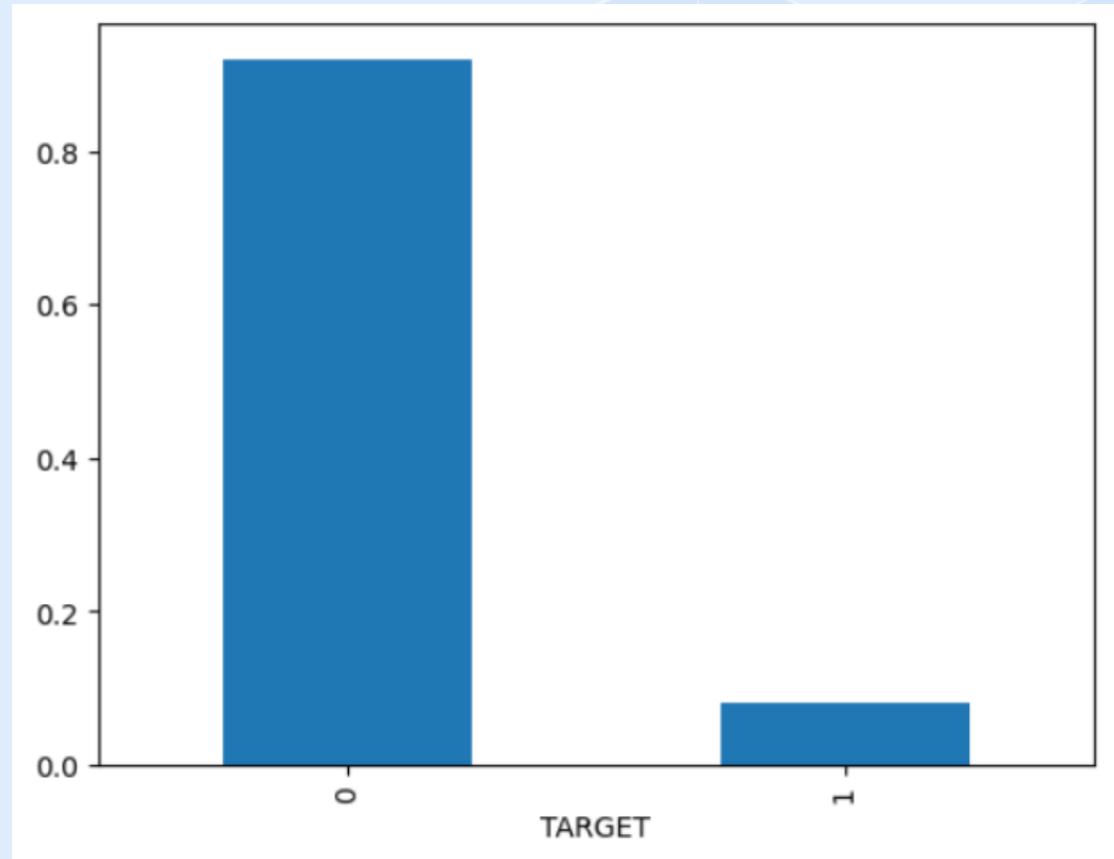
- Load the dataset
- Identify the percentage of missing Data in the dataset and remove the values which is greater than 35%.
- Impute the Null values
- Dealing with Target Variable
- Identify Outliers in the Dataset: Detect and identify outliers in the dataset
- Data Cleaning
- Perform Univariate, Bivariate and Multi-Variate Analysis on the both Dataset Application and Previous Application.
- Do merging of current and previous application data set and do the analysis.

## DETAIL OF OUTLIERS IN APPLICATION DATA

- There are few columns where the outlier values are standing out from the rest of the values.
- Few columns were having continuous range of values above the upper fence by plotting the graph using boxplot.
- List of Outliers – 'CNT\_CHILDREN', 'AMT\_INCOME\_TOTAL', 'AMT\_CREDIT', 'AMT\_ANNUITY', 'AMT\_GOODS\_PRICE', 'DAYS\_BIRTH', 'DAYS\_EMPLOYED', 'DAYS\_REGISTRATION'
- We are not deleting the outlier data points due to the importance of data while doing analysis.

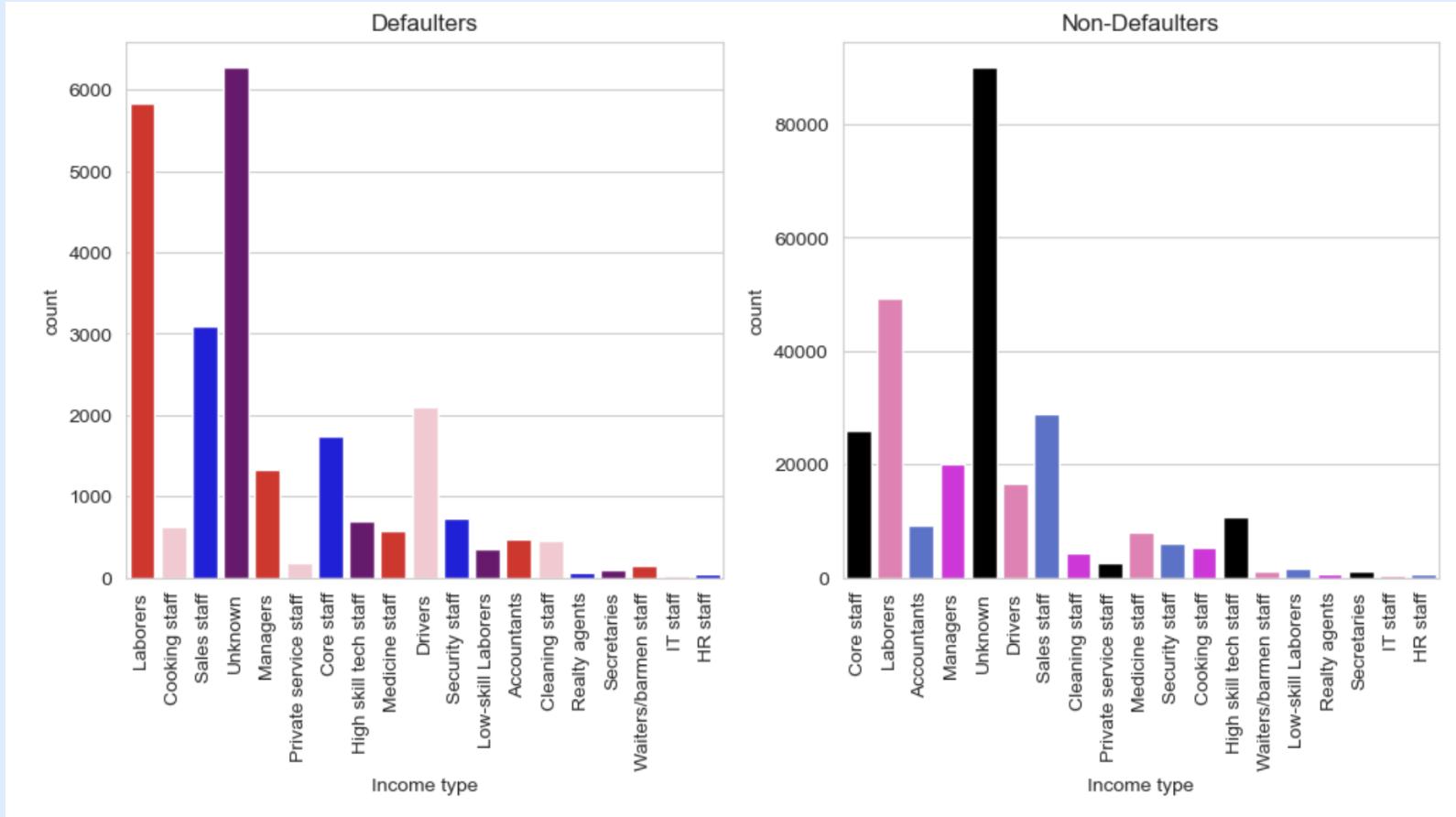
## TARGET COLUMN

- Ratio of Target column 0 and 1 is 11.38
  - 0 – NO - 2,82,686
  - 1 – Yes - 24,825
- 1 in every 11 applicant has payment difficulty



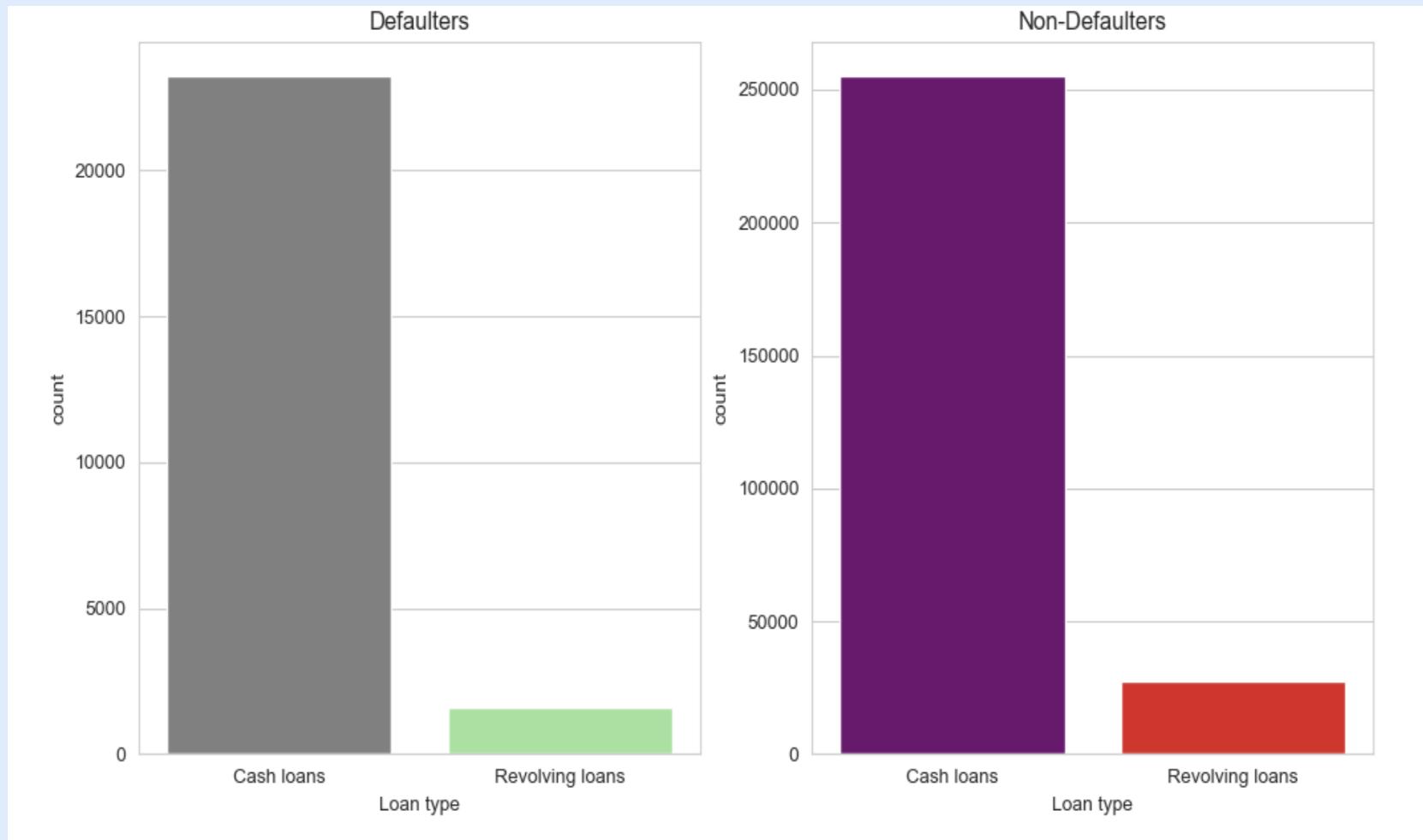
# UNIVARITAE ANALYSIS ON APPLICATION DATASET

## ANALYSIS OF OCCUPATION TYPE



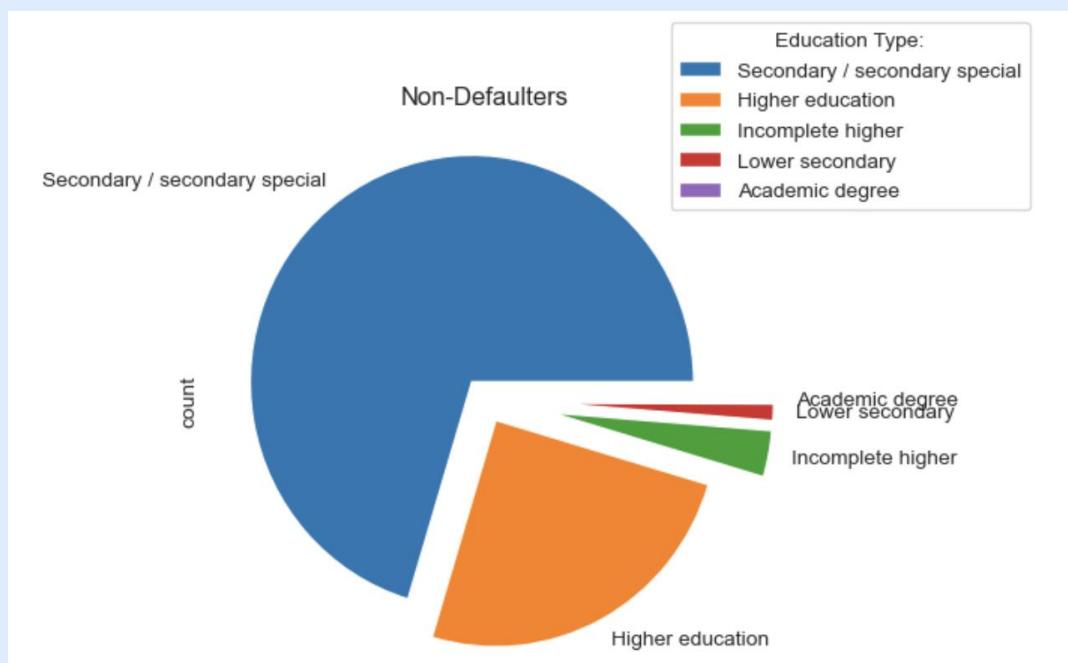
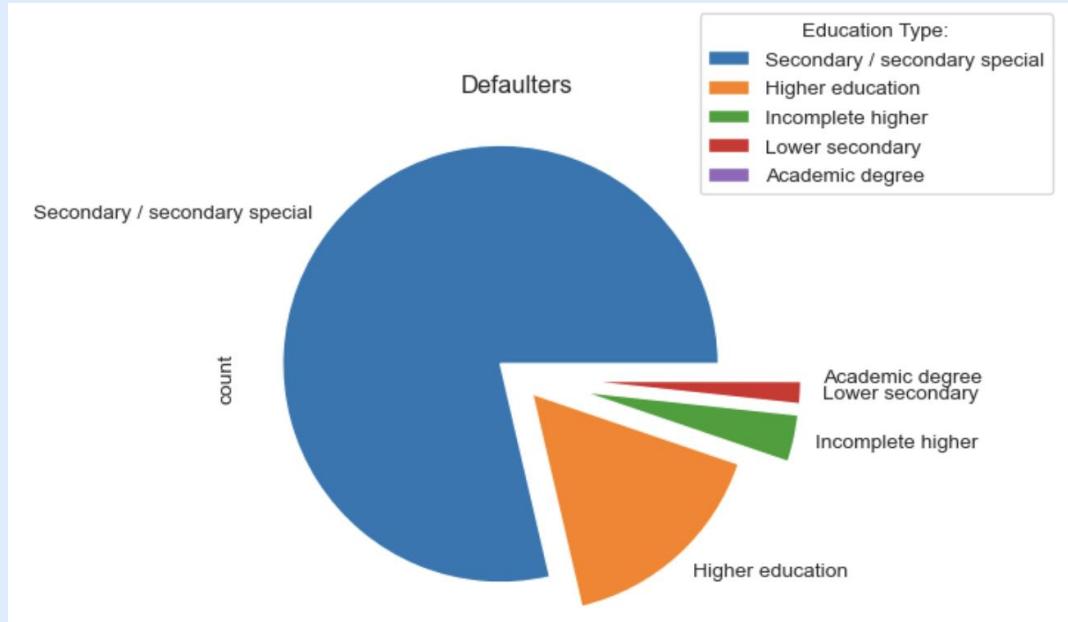
- Unknown is more in both Defaulter(1) and non-defaulter(0)
- Laborers is 2nd more Defaulter(1) and non-defaulter(0)

## ANALYSIS OF LOAN TYPE



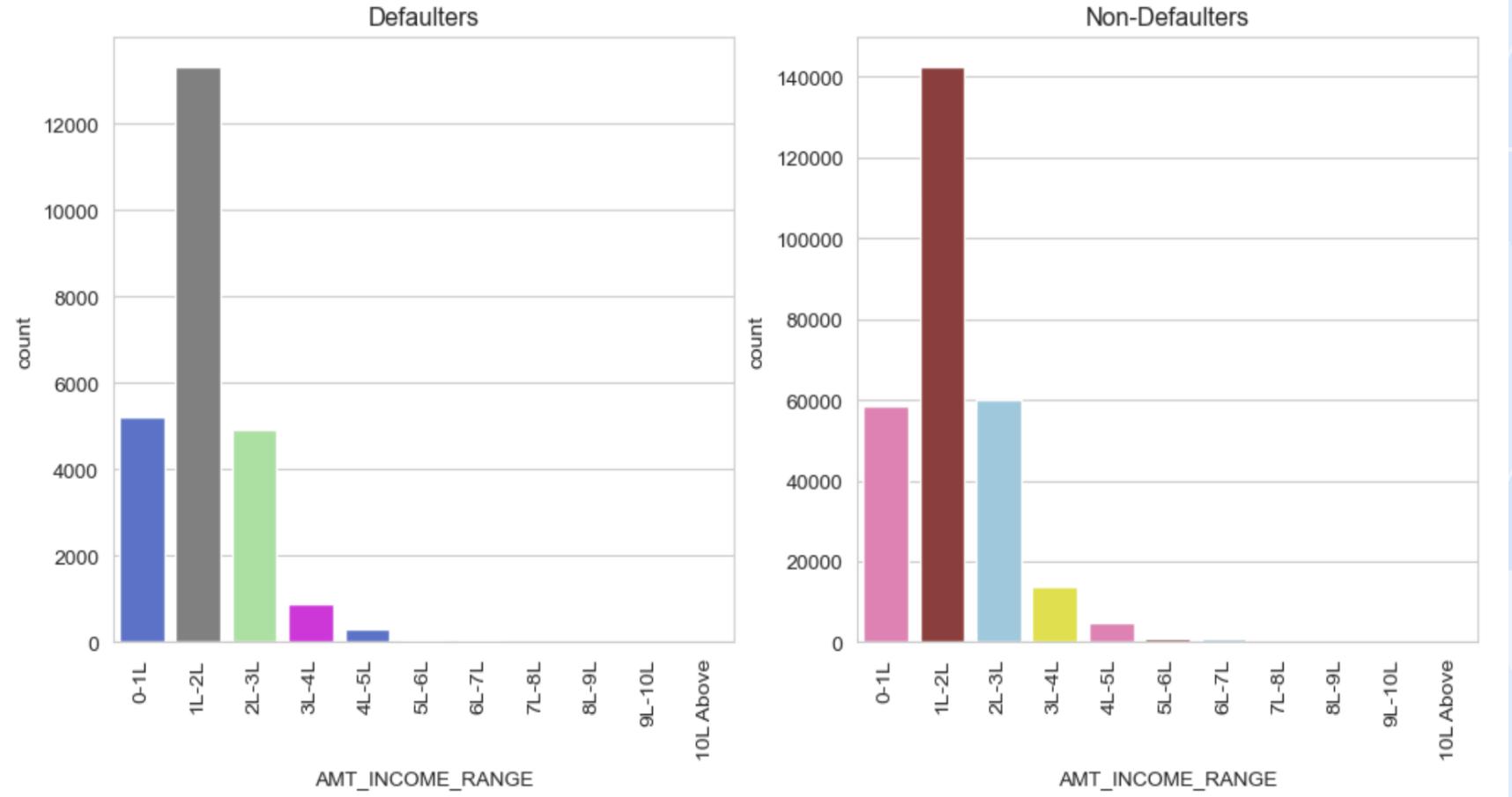
- Cash Loans is more in both Defaulters(1) and non-defaulters(0)

## ANALYSIS OF EDUCATION TYPE



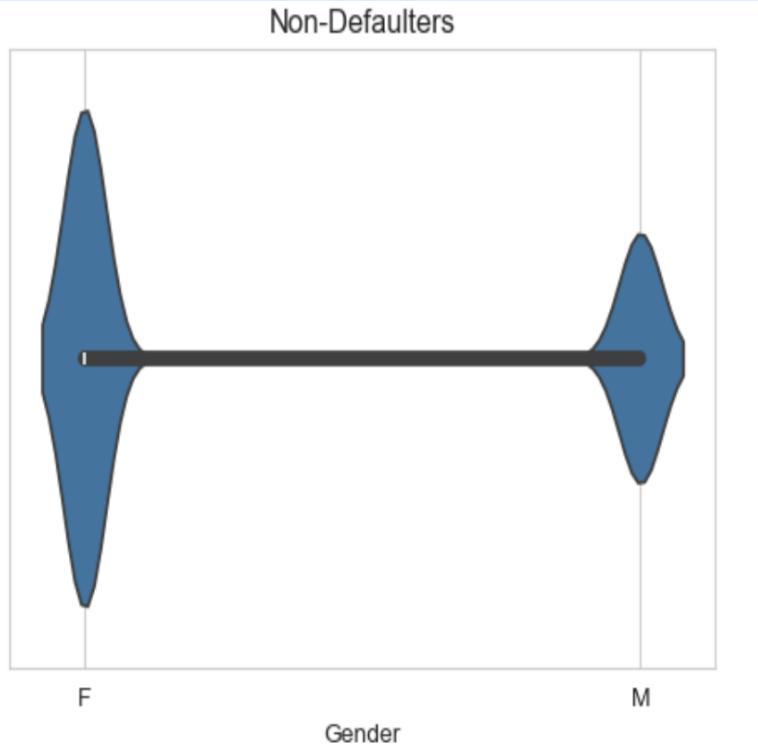
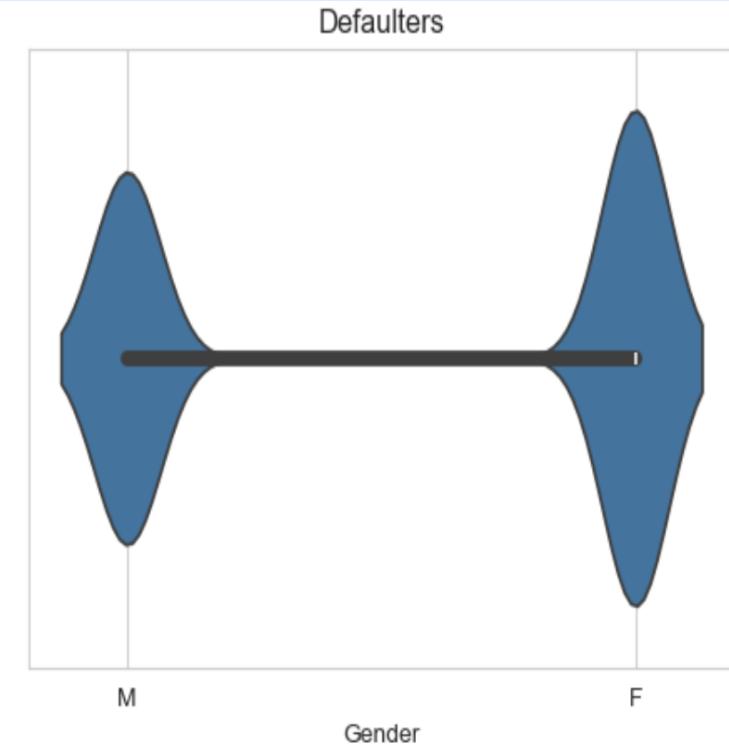
- Secondary/Secondary special is more in both Defaulter(1) and non-defaulter(0)
- Academic degree is less in both Defaulter(1) and non-defaulter(0) so, we can be consider for loans.

## ANALYSIS OF AMOUNT INCOME RANGE



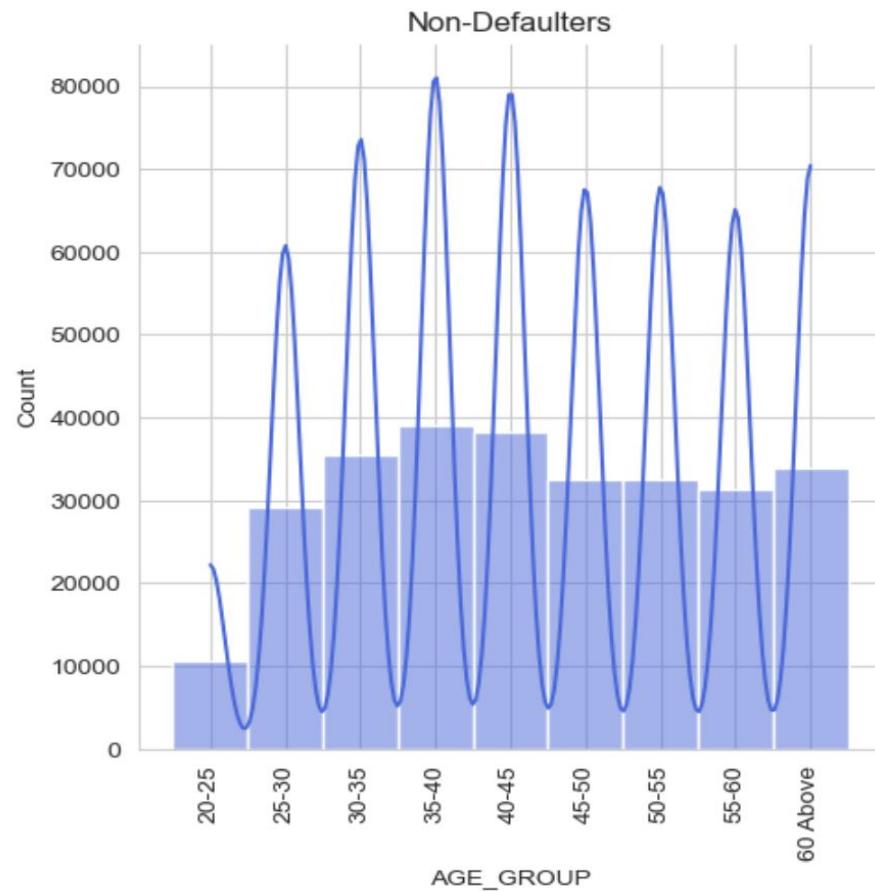
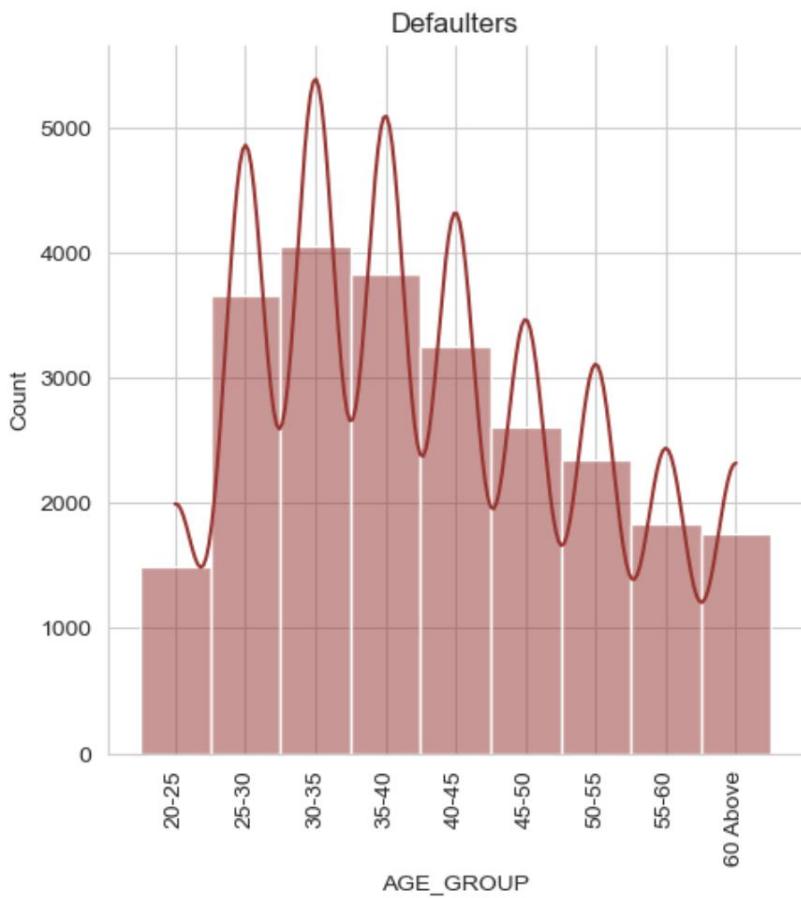
- 1-2L Income range is more in both Defaulter(1) and non-defaulter(0)

## ANALYSIS OF GENDER



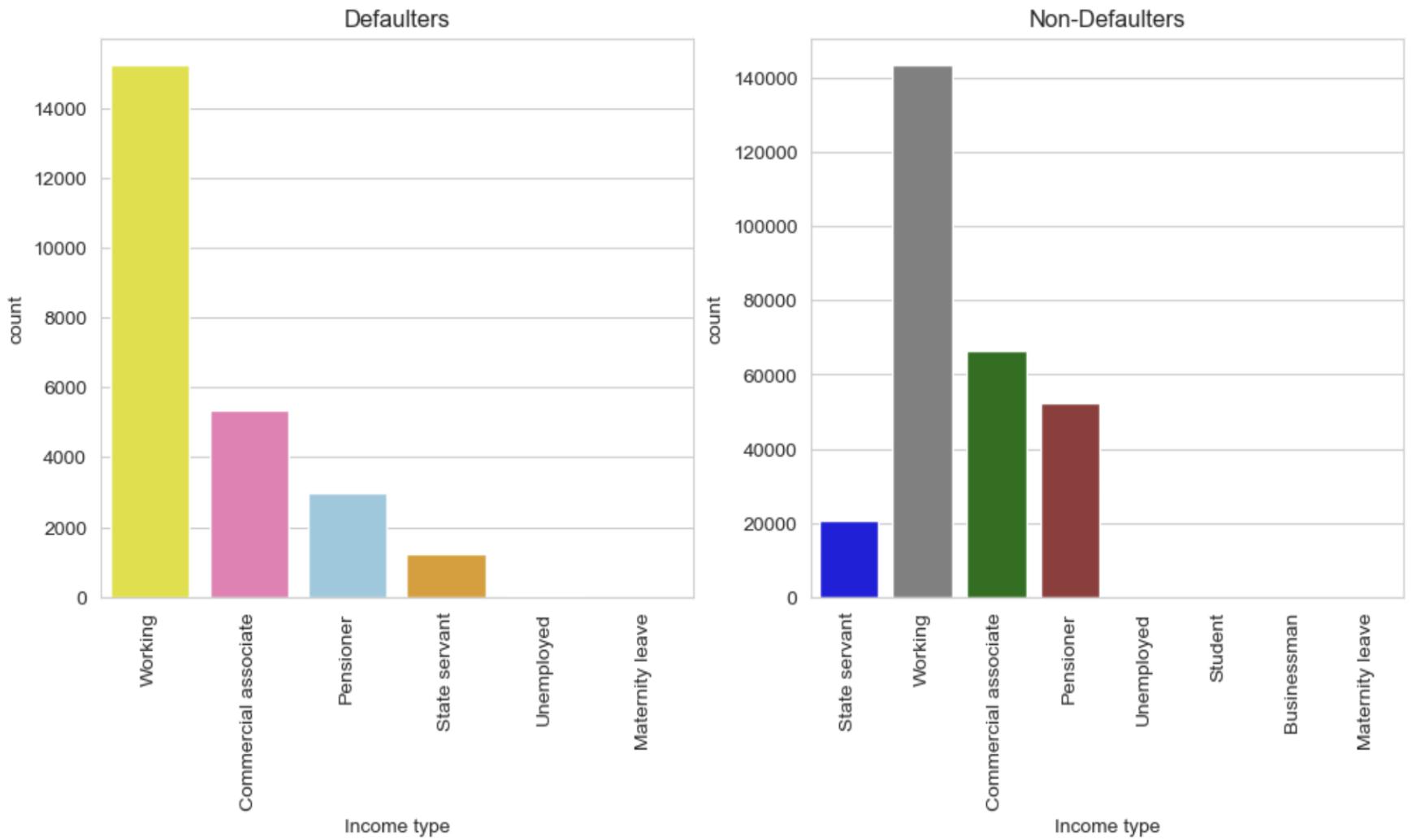
- Female applicants are more compared to Male applicants.
- Number of female is more in both Defaulter(1) and non-defaulter(0).

## ANALYSIS OF AGE GROUP



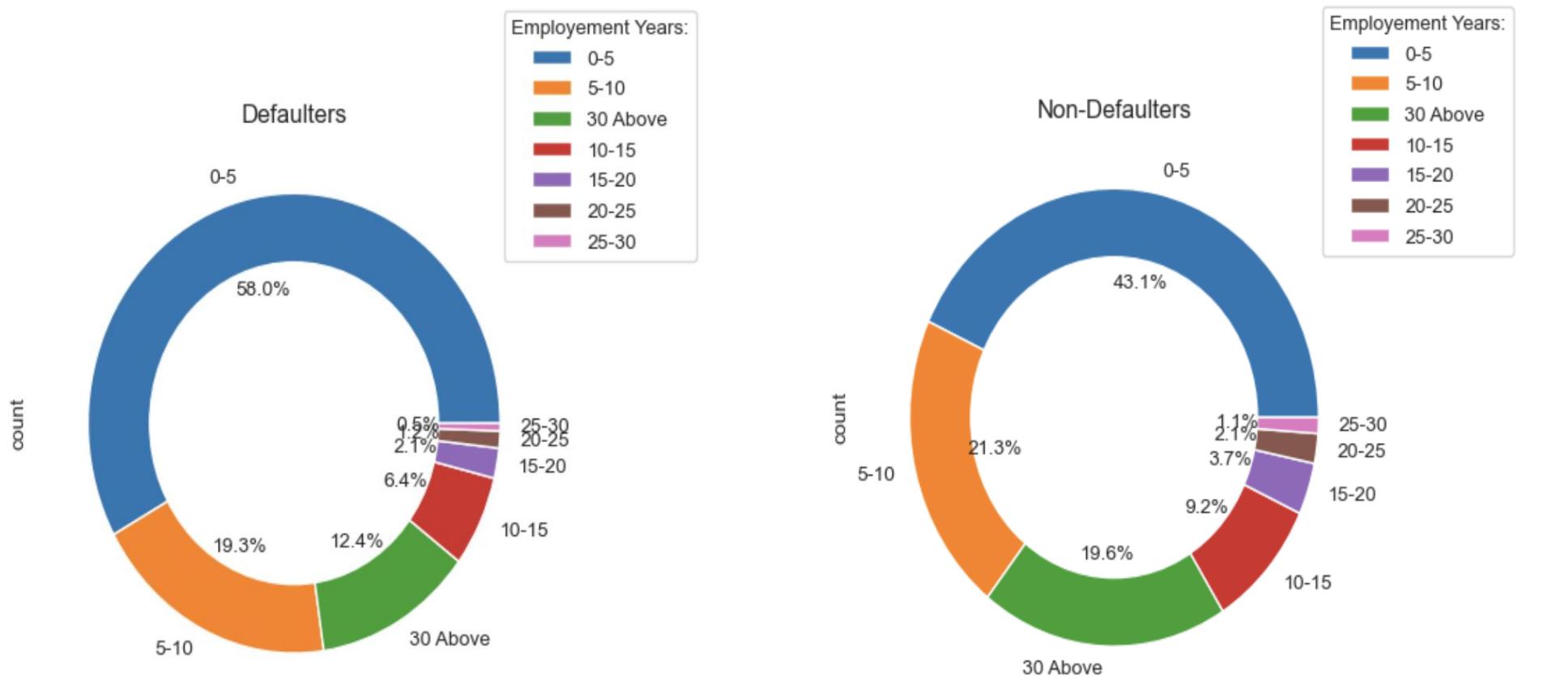
- Group age 30-35 is more defaulter(1)
- Group age 35-40 is more non-defaulter(0).

# ANALYSIS OF INCOME TYPE



- Working Professional people count has more in both Defaulter(1) and non-defaulter(0) in comparison to others.

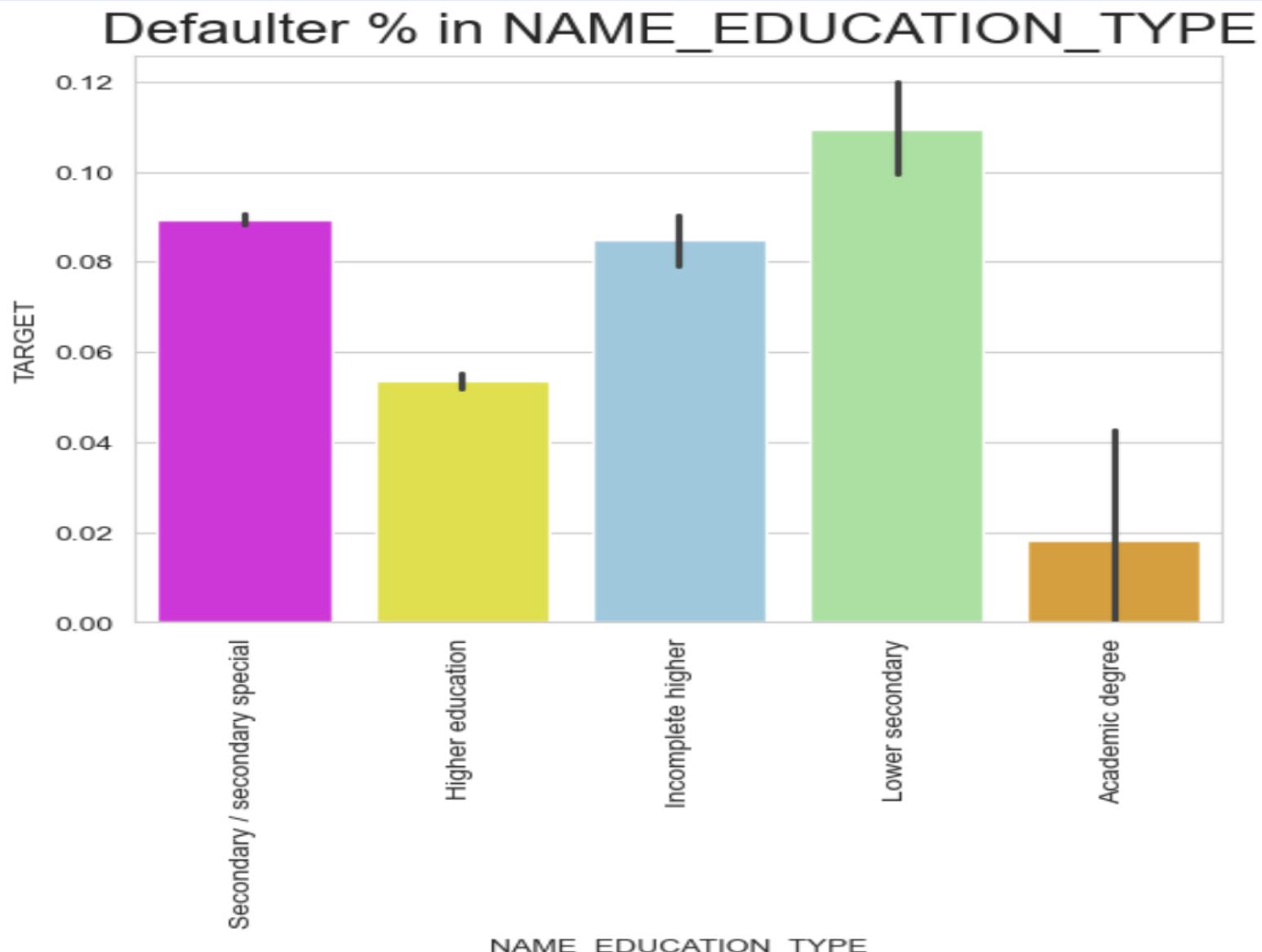
# ANALYSIS OF EMPLOYMENT YEARS



- Majority of the Employee is from 0-5 range for both defaulters(1) and non-defaulters(0).
- Work experience and loan application are inversely proportional.
- As work experience increase loan applicants are decreasing.

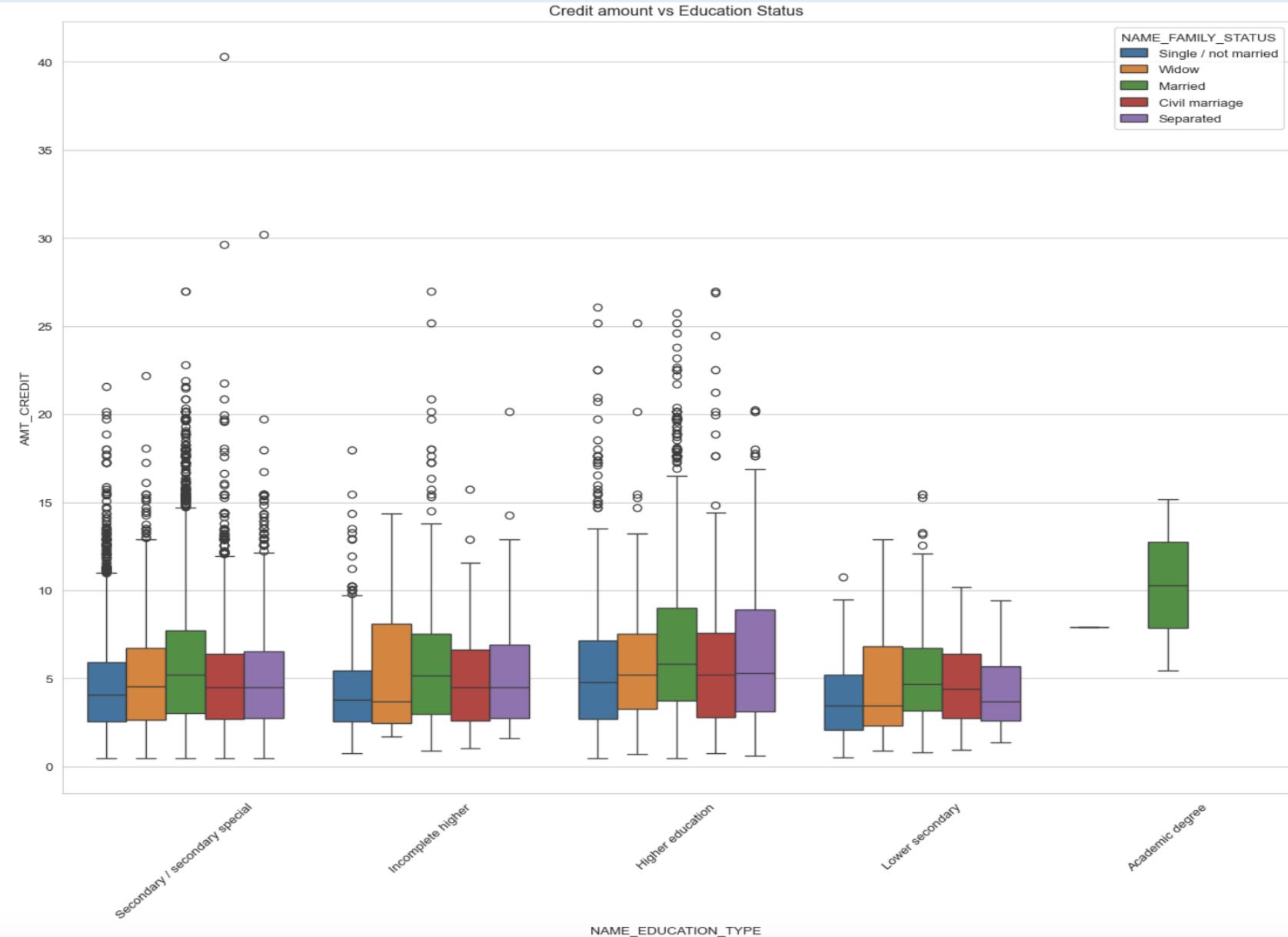
# BIVARIATE ANALYSIS ON APPLICATION DATASET

## ANALYSIS OF EDUCATION TYPE



- Lower secondary is highest percentage of defaulters.
- Academic degree is lowest percentage of defaulters. So Academic degree education can consider for loan.

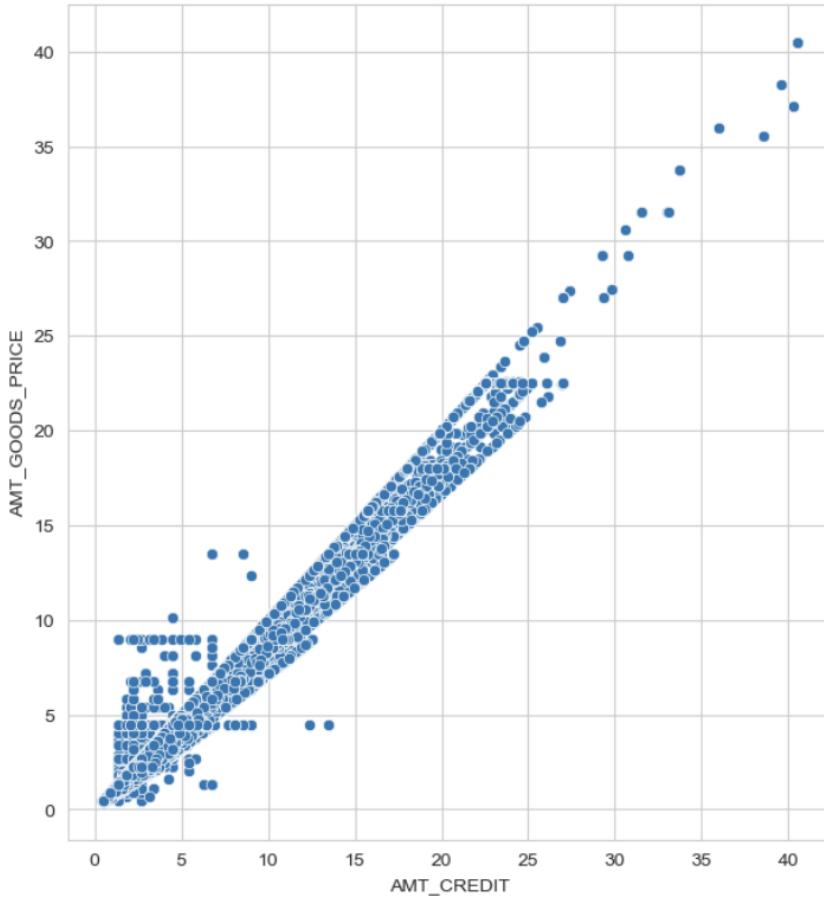
# ANALYSIS OF EDUCATION STATUS AND AMOUNT CREDIT



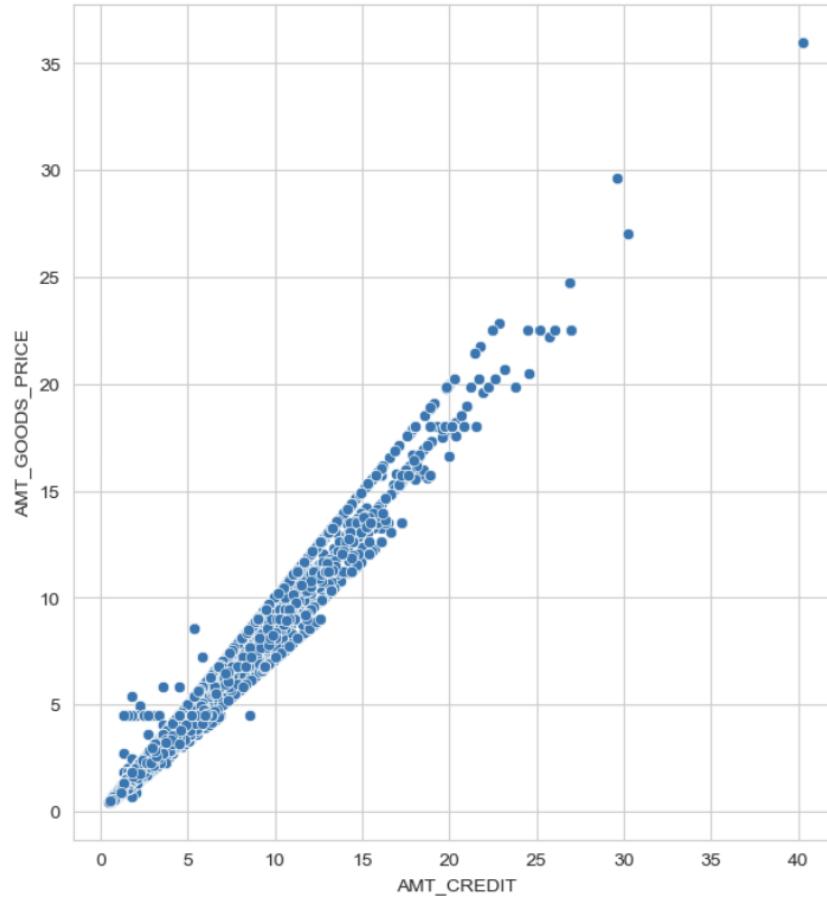
- Married people with Academic degree is having highest amount credited

## ANALYSIS OF AMOUNT GOODS PRICE AND AMOUNT CREDIT

CREDIT vs GOODS PRICE for Target-0

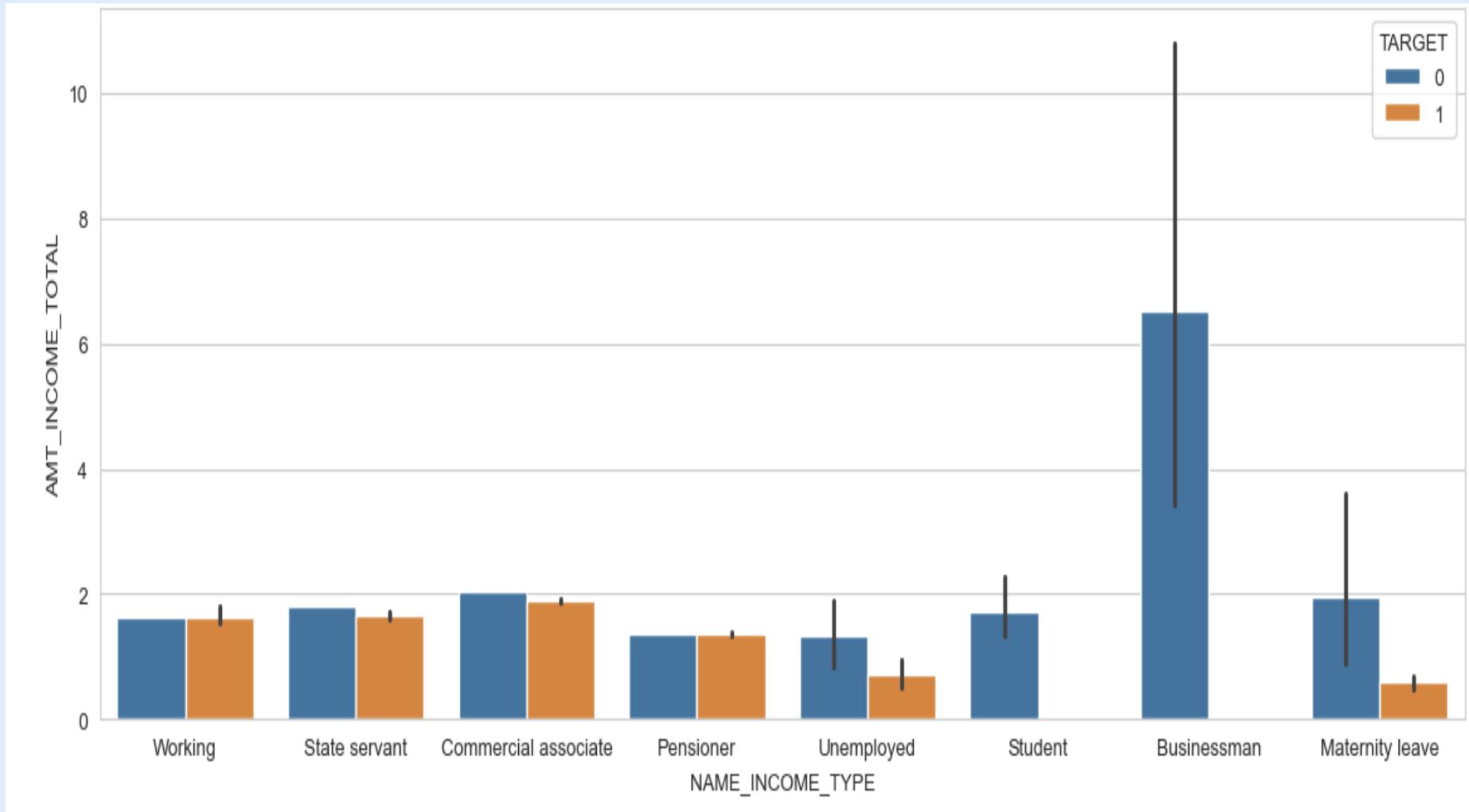


CREDIT vs GOODS PRICE for Target-1



- AMT CREDIT and AMT GOODS PRICE are highly correlated, which means if increase in goods price, the credit increased directly and vice versa.

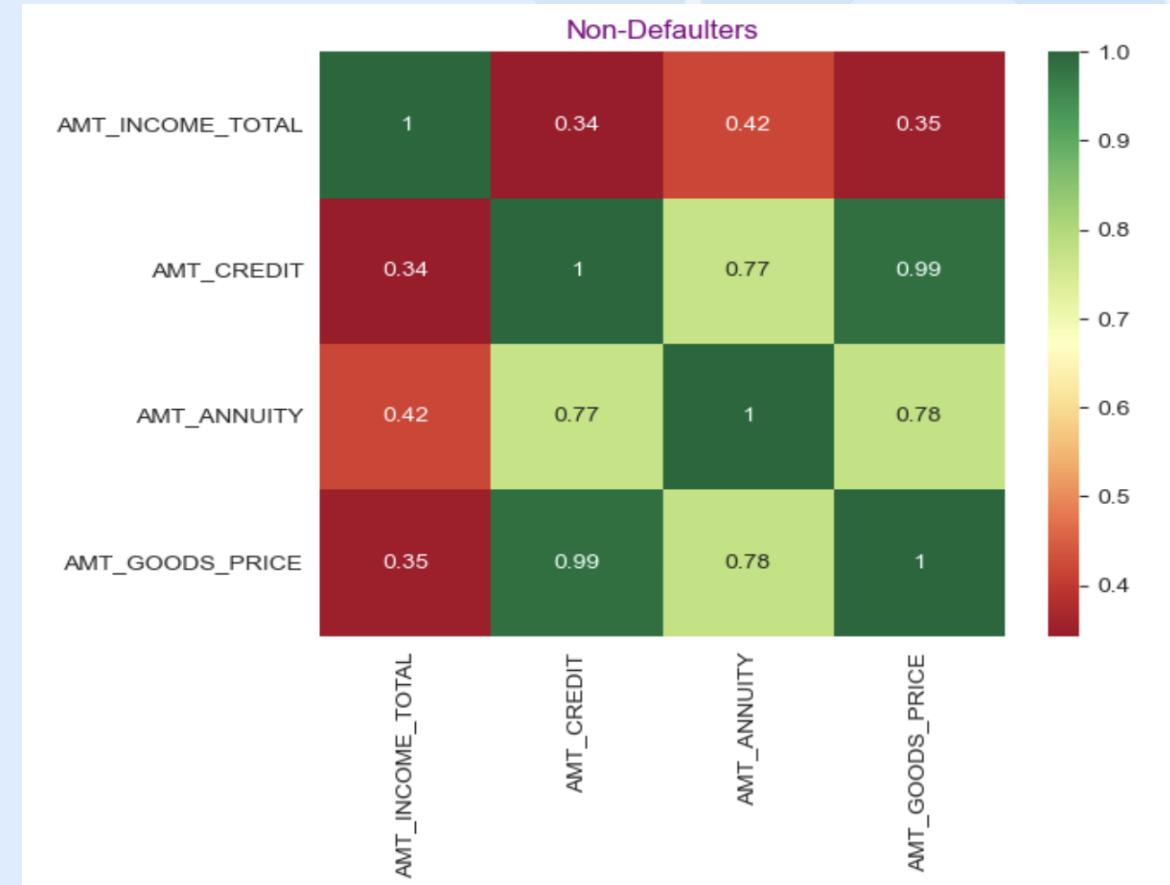
## ANALYSIS OF NAME INCOME TYPE AND AMOUNT INCOME TOTAL



- No defaulters for Businessman and student having income. Bank can consider them for loan.

# MULTI-VARIATE ANALYSIS ON APPLICATION DATASET

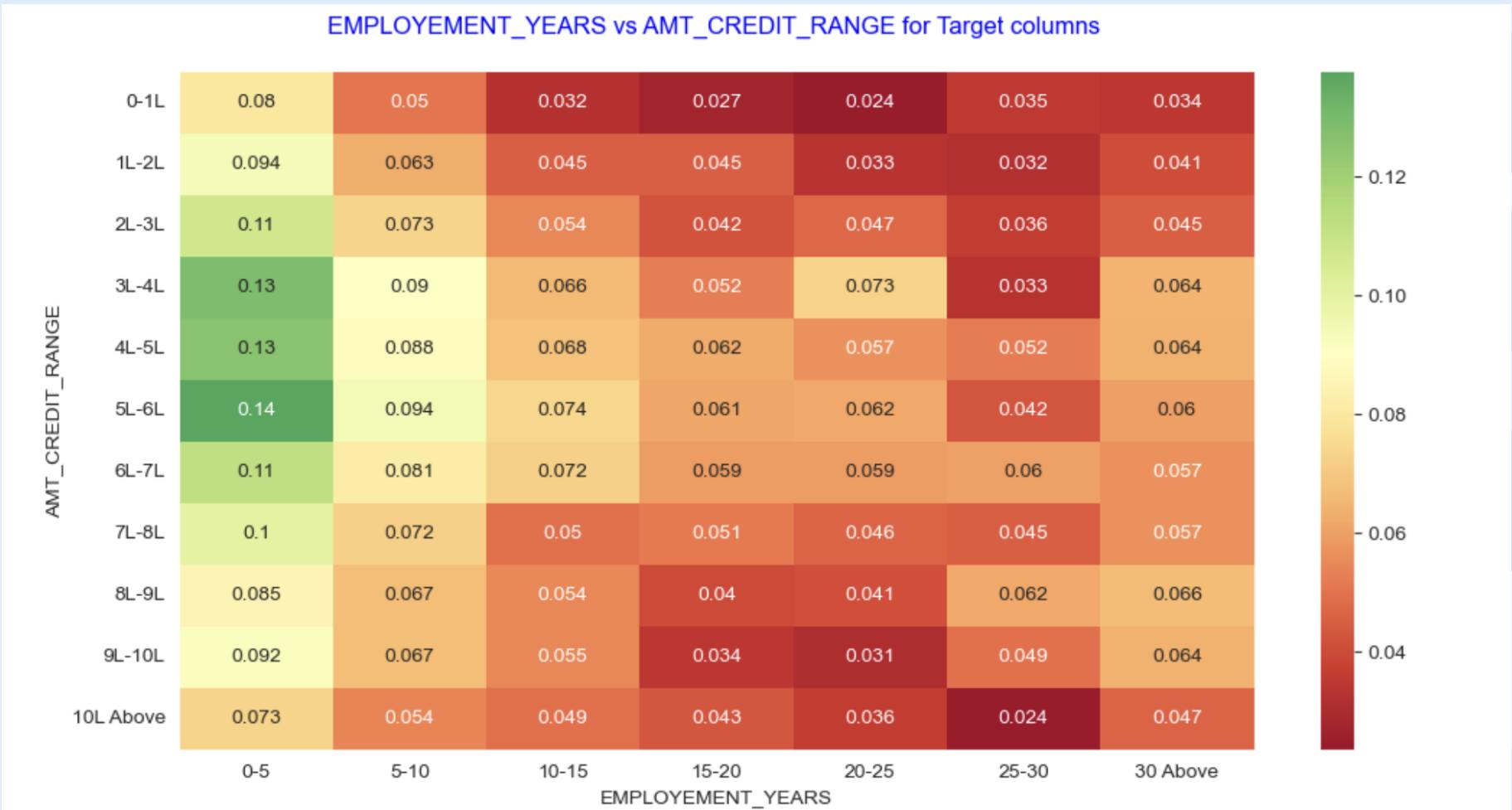
## ANALYSIS OF AMT\_INCOME\_TOTAL, AMT\_CREDIT, AMT\_ANNUITY AND AMT\_GOODS\_PRICE



- Highly corelate columns(Defaulter)
  - AMT\_CREDIT and AMT\_ANNUITY (0.75)
  - AMT\_CREDIT and AMT\_GOODS\_PRICE (0.98)
  - AMT\_ANNUITY and AMT\_GOODS\_PRICE (0.75)

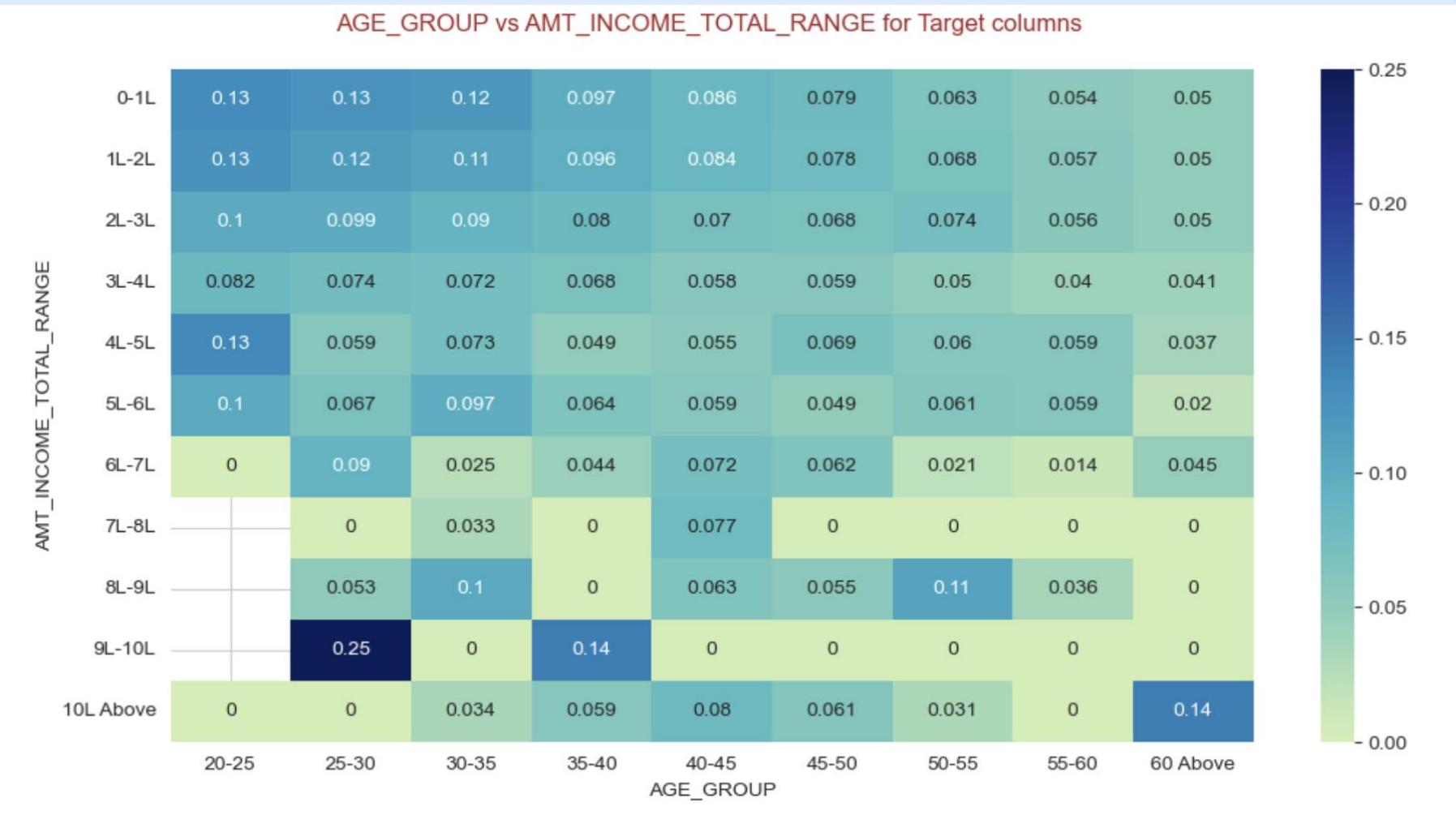
- Highly corelate columns(Non-Defaulter)
  - AMT\_CREDIT and AMT\_ANNUITY (0.77)
  - AMT\_CREDIT and AMT\_GOODS\_PRICE (0.99)
  - AMT\_ANNUITY and AMT\_GOODS\_PRICE (0.78)

## ANALYSIS OF EMPLOYEMENT\_YEARS VS AMT\_CREDIT\_RANGE FOR TARGET COLUMNS



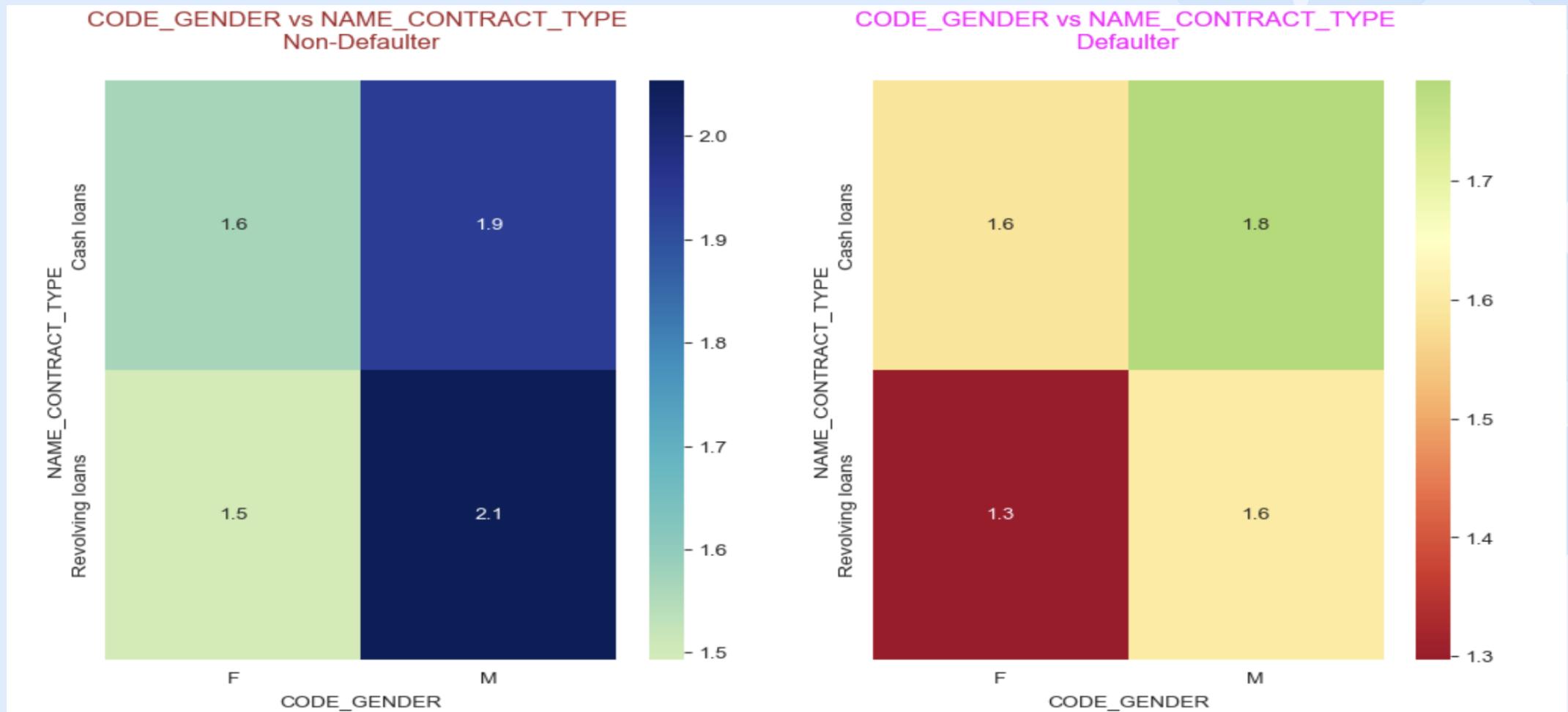
- Employment years 0-5 have highest Amount credit for 5-6 Lakh range.
- Employment years 25-30 has lowest Amount credit for 10 lakh above.

## ANALYSIS OF EMPLOYEMENT\_YEARS VS AMT\_CREDIT\_RANGE FOR TARGET COLUMNS



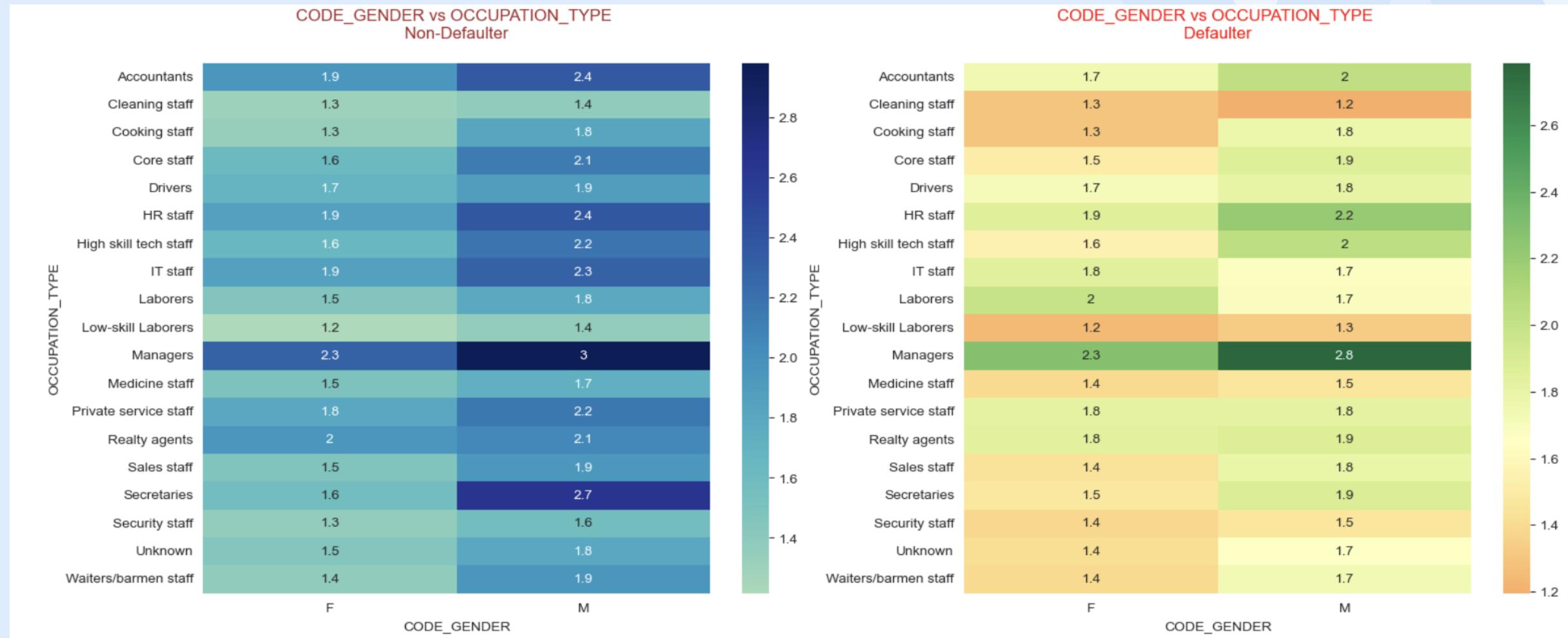
- Age group 25-30 has highest total income.

## ANALYSIS OF CODE\_GENDER VS NAME\_CONTRACT\_TYPE



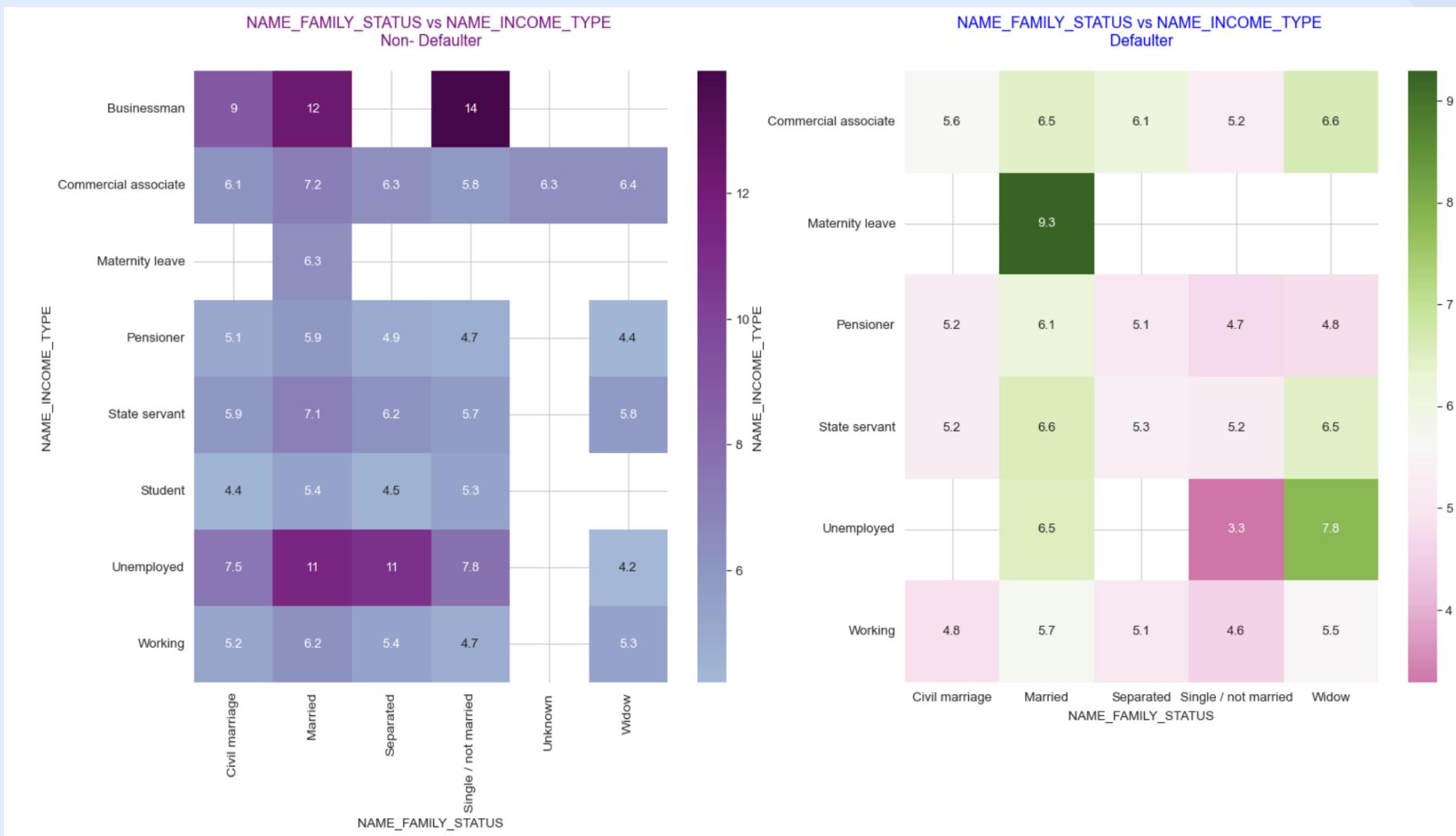
- Male with Revolving loans contract type has highest Non-defaulter and Female with Revolving loans contract type has lowest Non-defaulter.
- Male with Cash loans contract type has highest Defaulter and Female with Revolving loans contract type has lowest defaulter.

# ANALYSIS OF CODE\_GENDER VS OCCUPATION\_TYPE



- Male with Managers occupation type has highest Non-defaulter and Defaulter and Female with Low-skill laborers occupation type has lowest Non-defaulter and Defaulter.

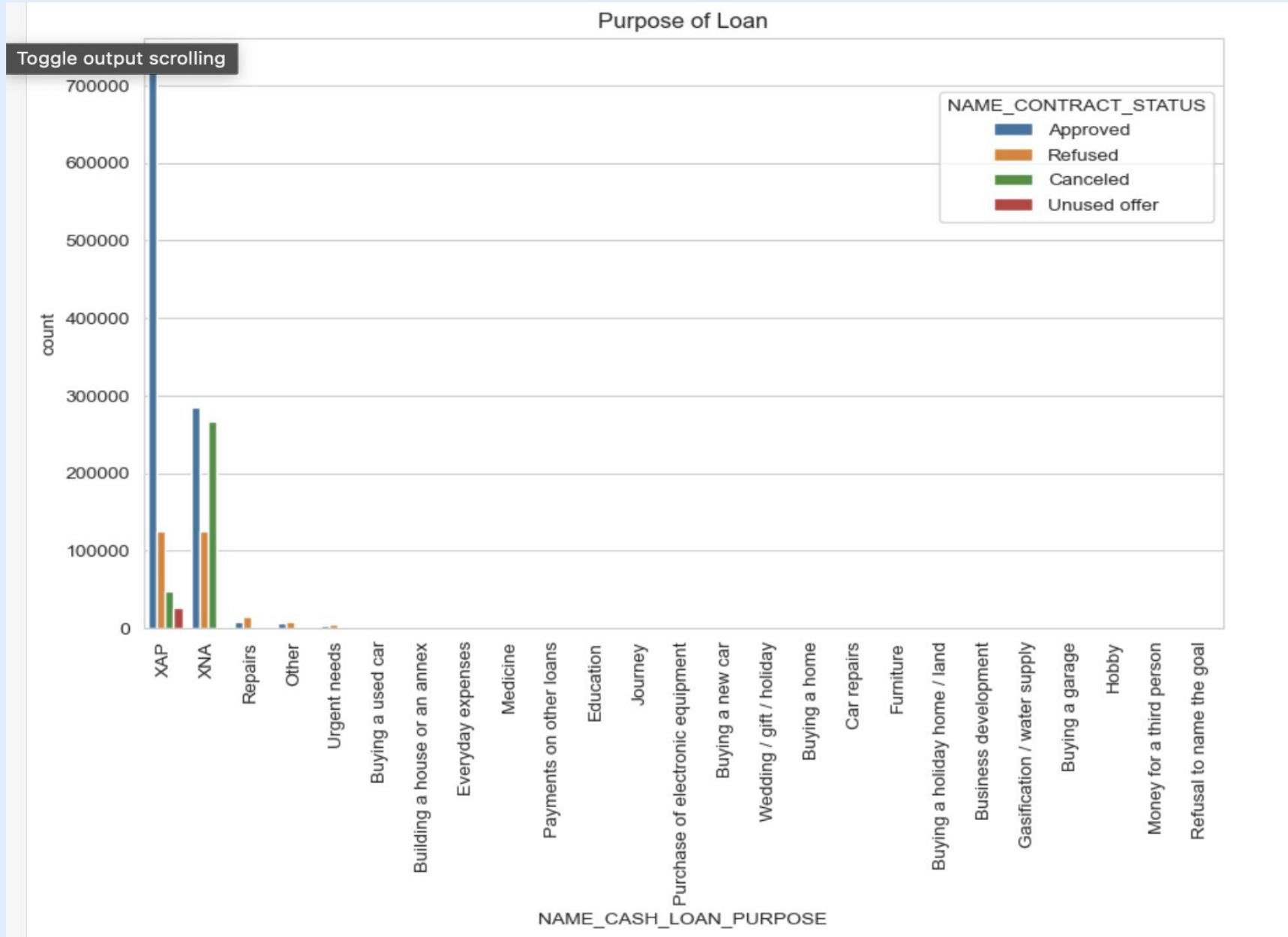
# ANALYSIS OF NAME\_FAMILY\_STATUS VS NAME\_INCOME\_TYPE



- Single family status with Businessman Income type has highest Non-defaulters.
- Married family status with Maternity leave income type has highest Defaulters.

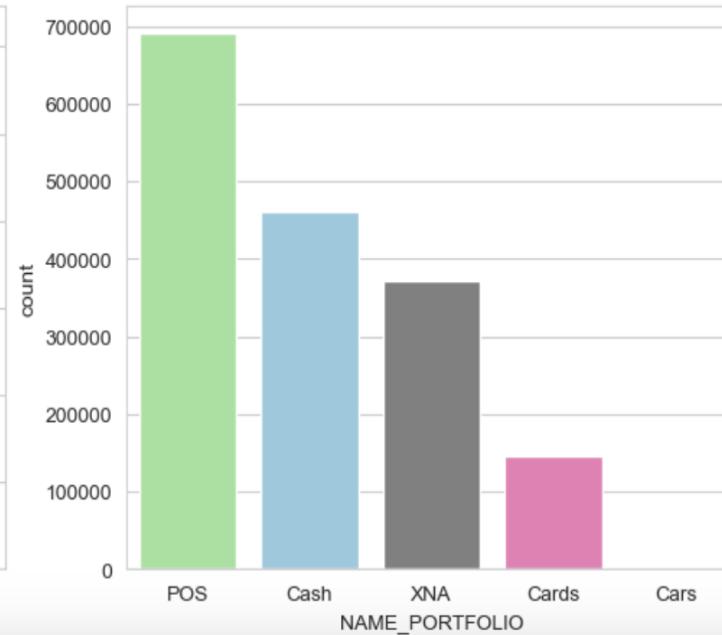
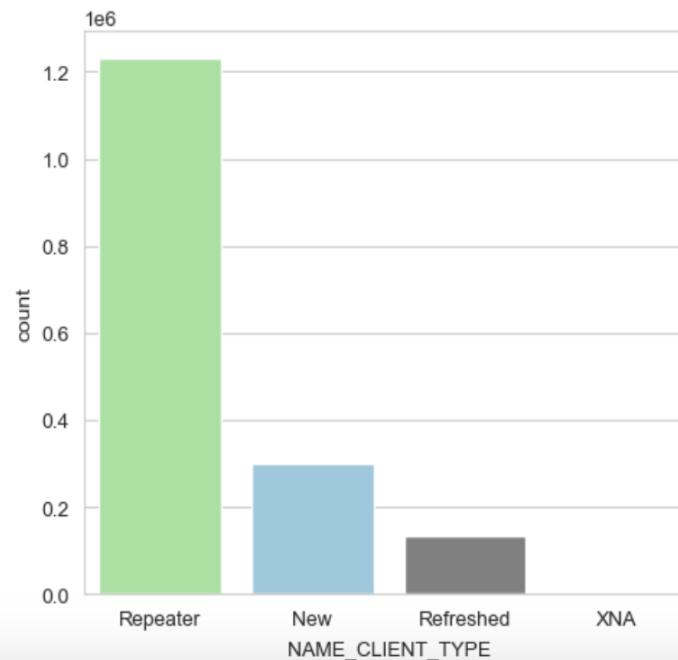
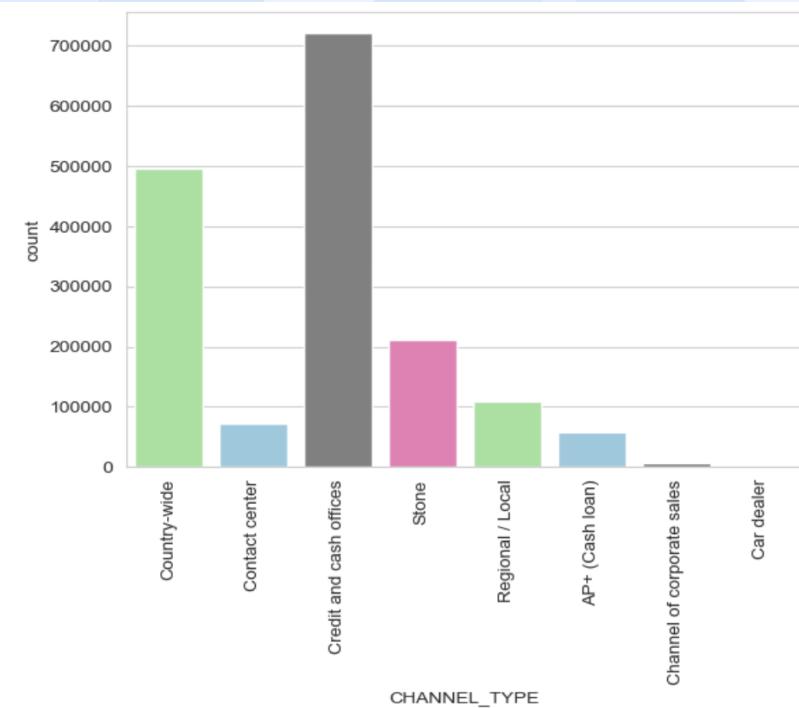
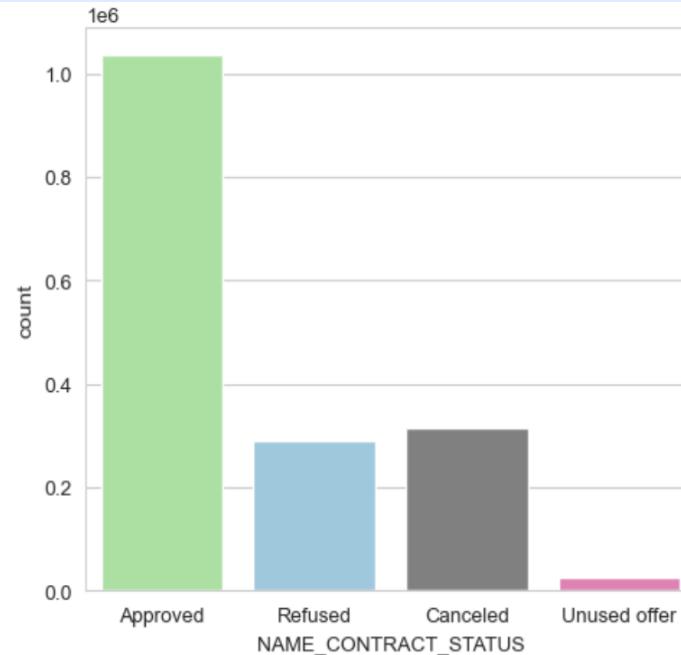
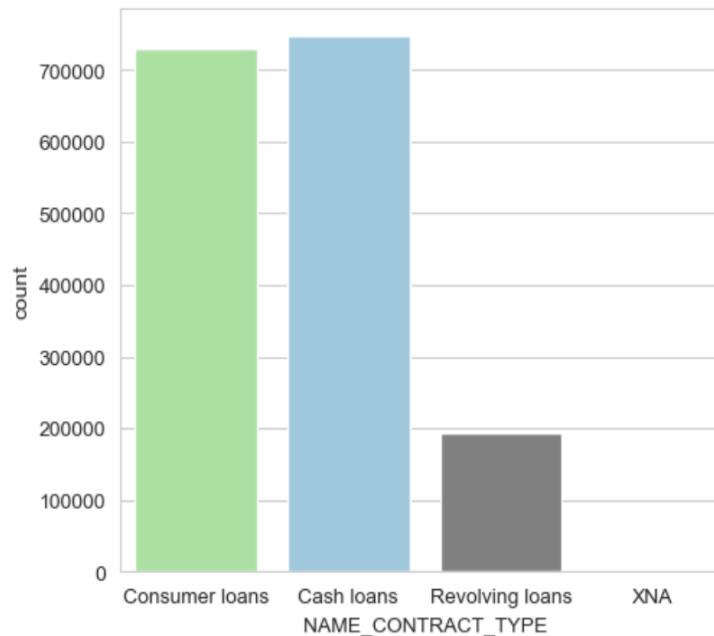
# UNIVARIATE ANALYSIS ON PREVIOUS APPLICATION DATASET

# ANALYSIS OF LOAN PURPOSE



- Repairs cash loans has highest refused contract status.
- For Repairs loan rejection is more than approval.
- For XNA and XAP loan approval is more compared to other categories

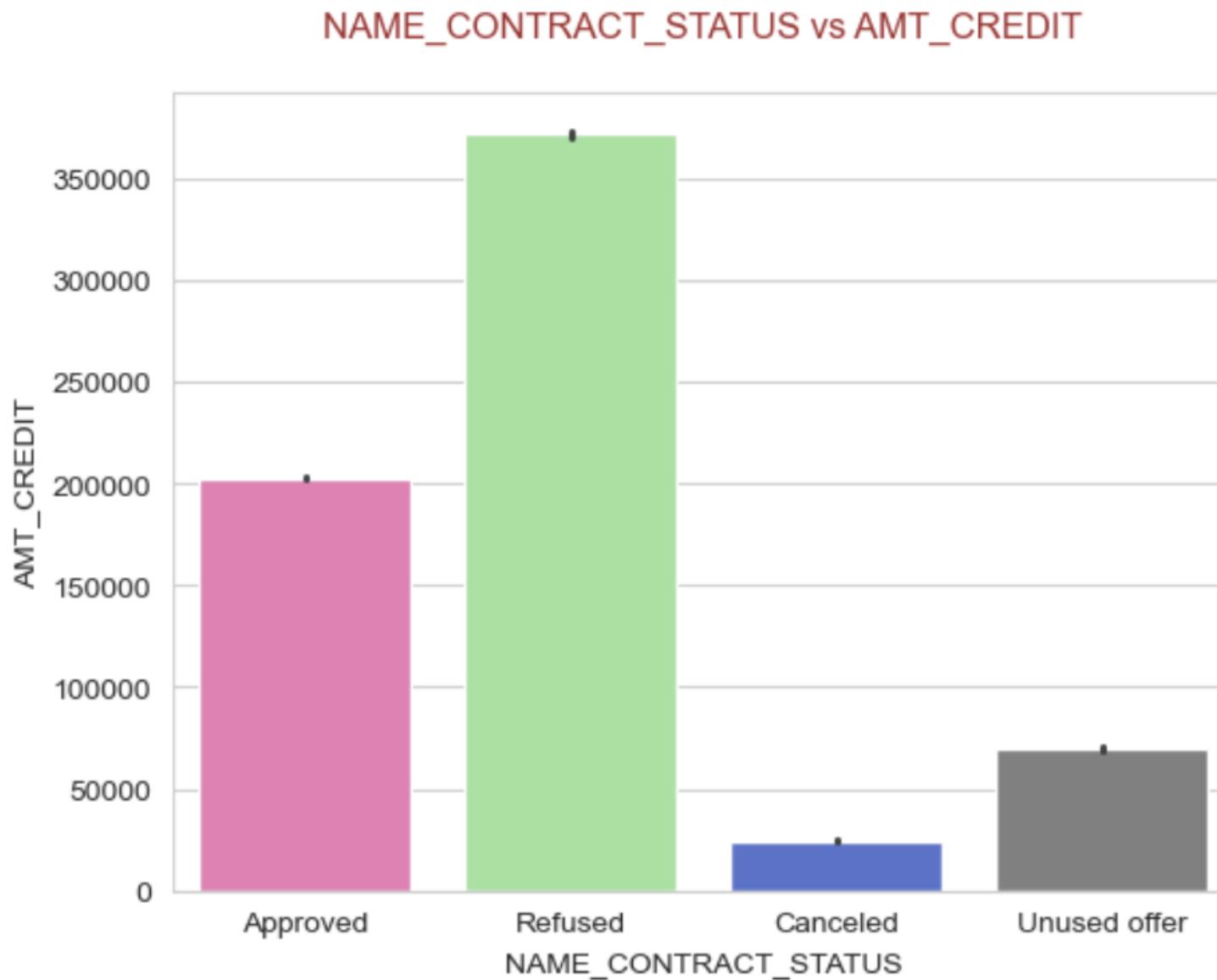
# ANALYSIS OF CONTRACT TYPE AND STATUS, CLIENT AND CHANNEL TYPE, CONTRACT STATUS



- Count of Revolving loan is less compared to consumer loans and cash loans.
- There are very few Refused loans. Almost negligible Canceled loans.
- Number of Repeater applicants is more compared to new and refreshed applicants.
- Number of application for Cards and Cars in previous application is very less and the highest number is for POS.
- Credit and Cash offices channel count is more compared to other channel types.

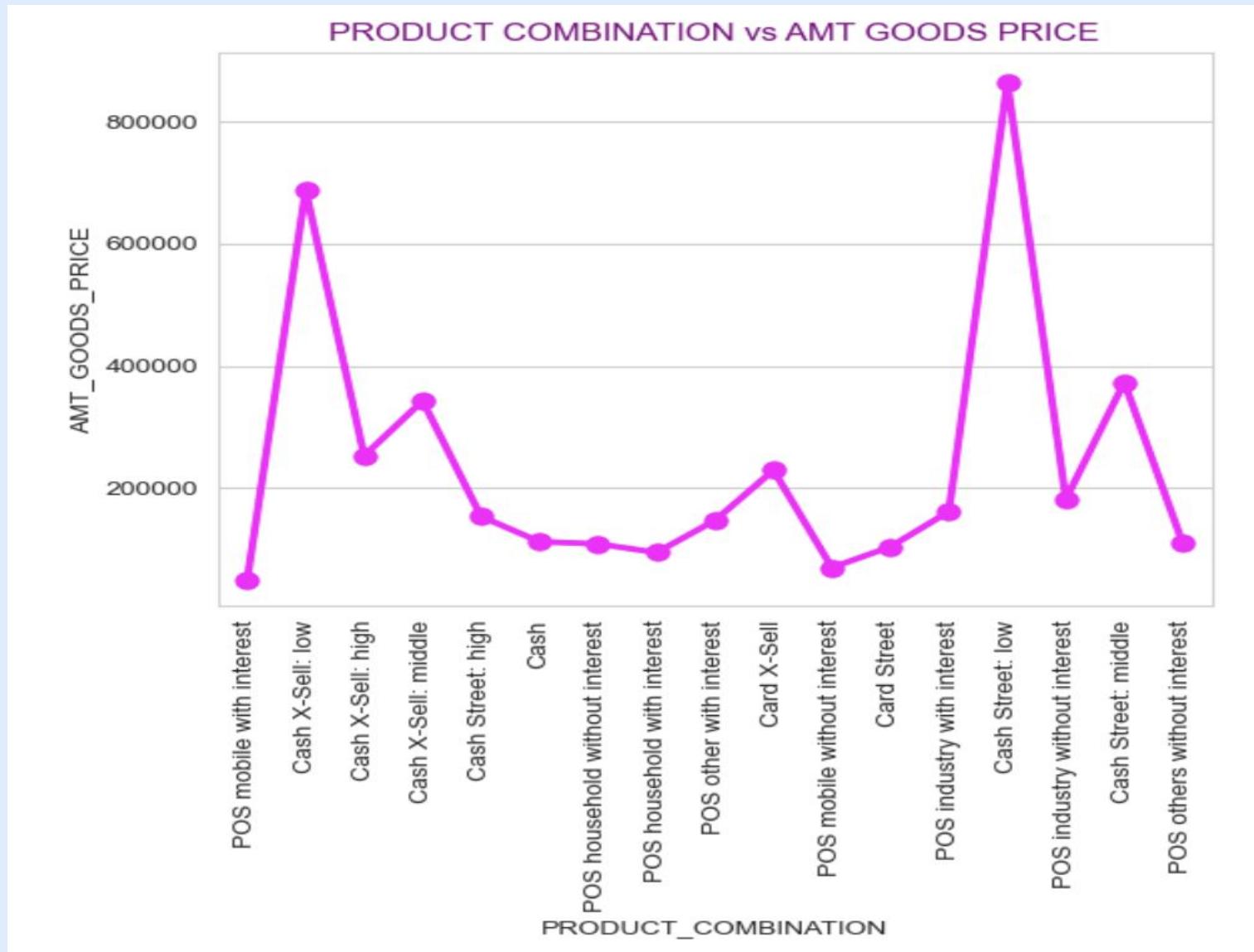
# BIVARIATE ANALYSIS ON PREVIOUS APPLICATION DATASET

## ANALYSIS OF CONTRACT STATUS AND AMOUNT CREDIT



- Refused Contract status has highest Amount credit.
- Cancel contract status has lowest amount credit.

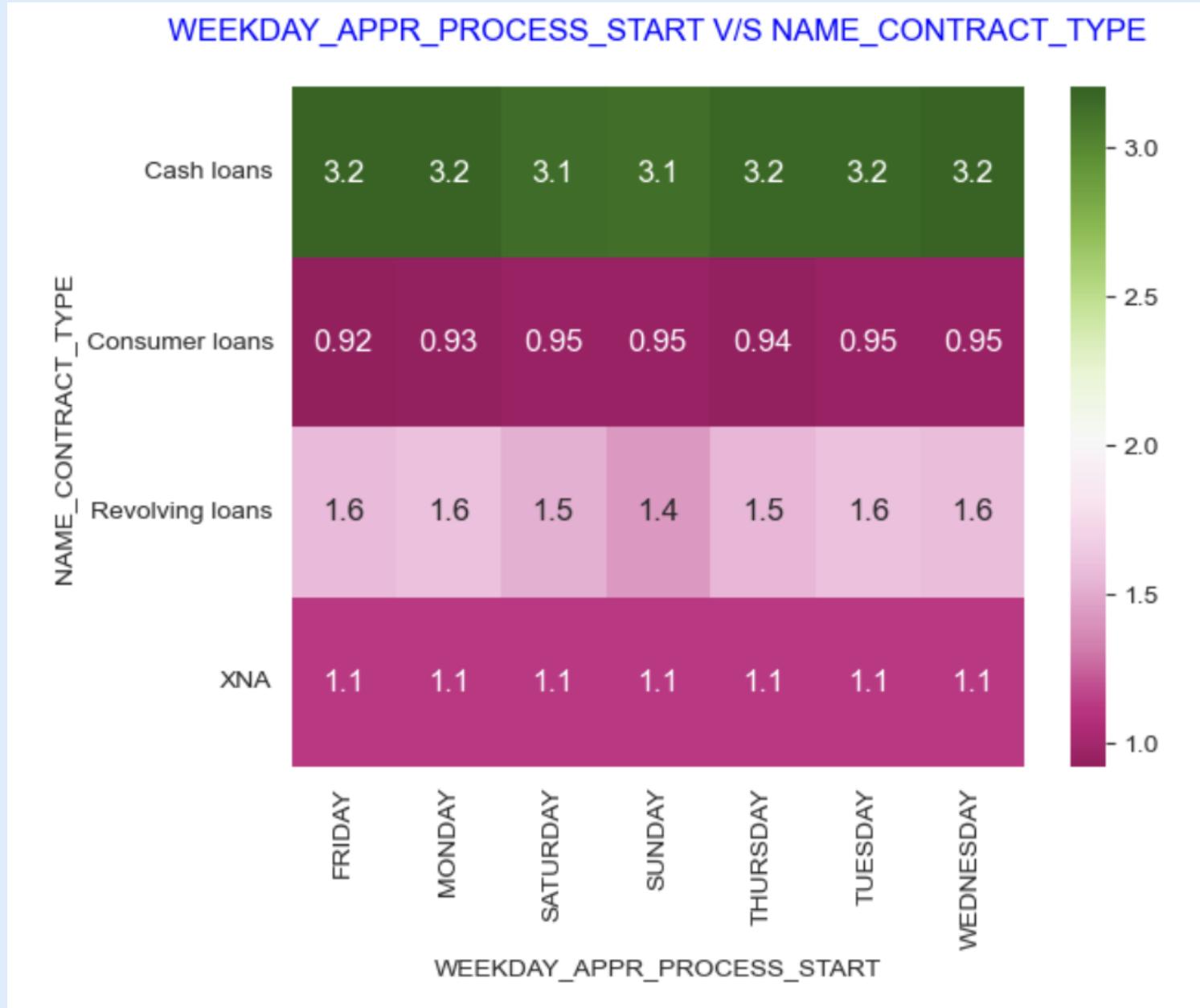
# ANALYSIS OF PRODUCT COMBINATION AND AMOUNT GOODS PRICE



- Cash Street low product has highest amount goods price.
- POS mobile with interest has lowest amount goods price.

# MULTI-VARIATE ANALYSIS ON PREVIOUS APPLICATION DATASET

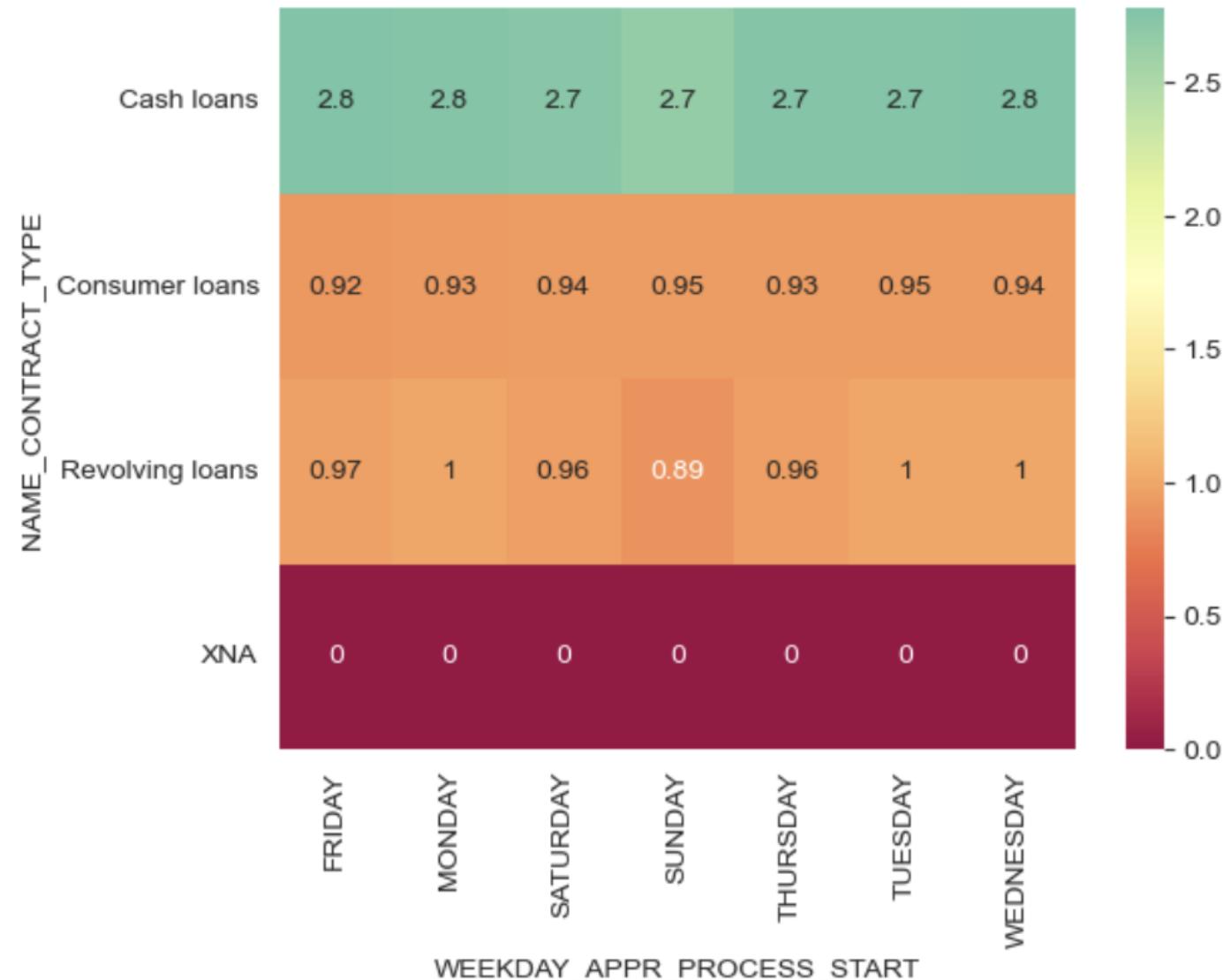
# ANALYSIS OF WEEKDAY APPR PROCESS START AND CONTRACT TYPE ON AMOUNT GOODS PRICE



- Cash loans contract type has highest defaulter on Monday, Tuesday, Wednesday, Thursday, Friday.
- Consumer loans contract type has lowest defaulter on Friday

# ANALYSIS OF WEEKDAY APPR PROCESS START AND CONTRACT TYPE ON AMOUNT APPLICATION

WEEKDAY\_APPR\_PROCESS\_START V/S NAME\_CONTRACT\_TYPE

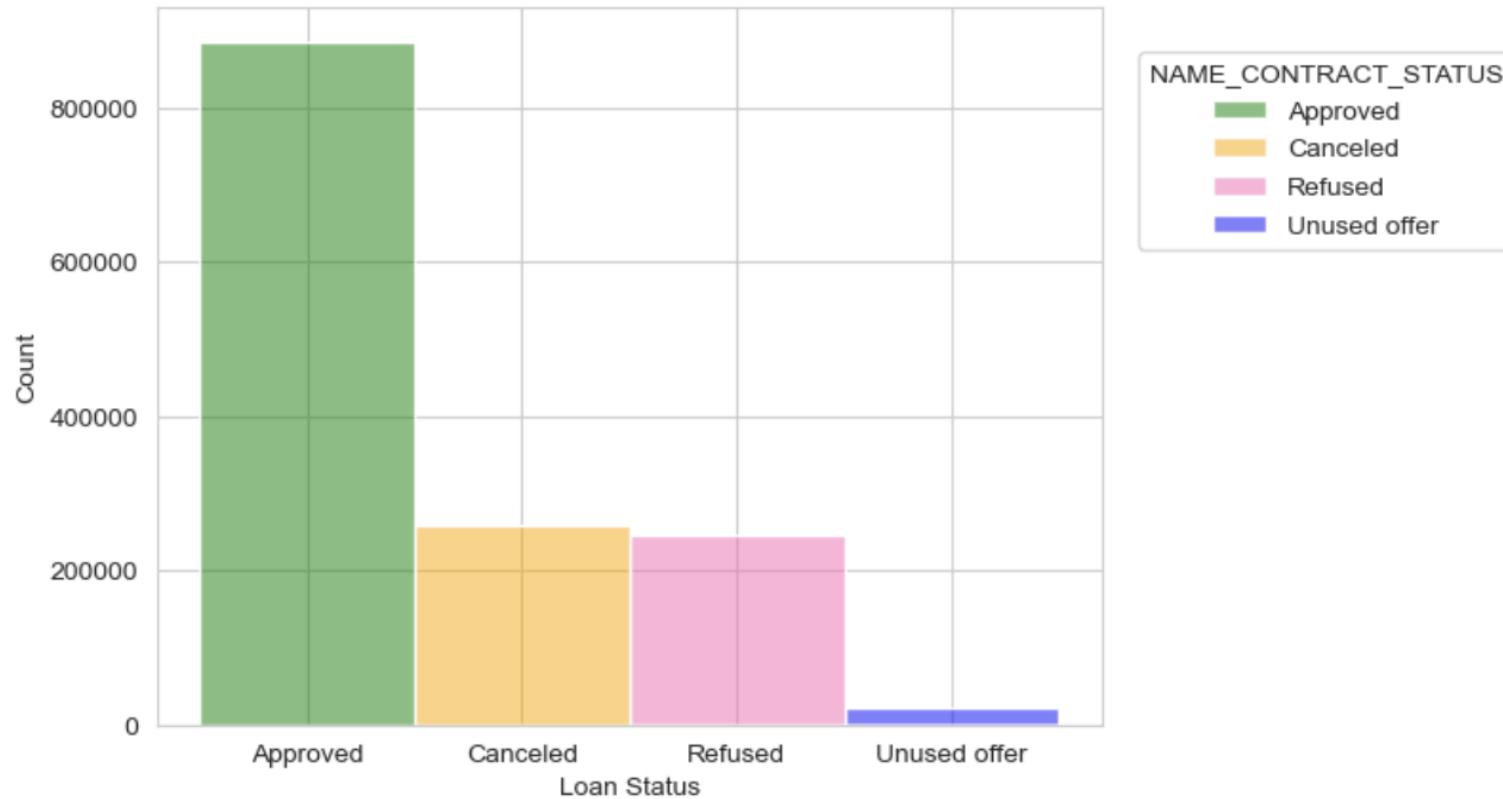


- Cash loans contract type has highest non-defaulter on Monday, Wednesday, Friday
- Revolving loans contract type has lowest non-defaulter on Sunday.

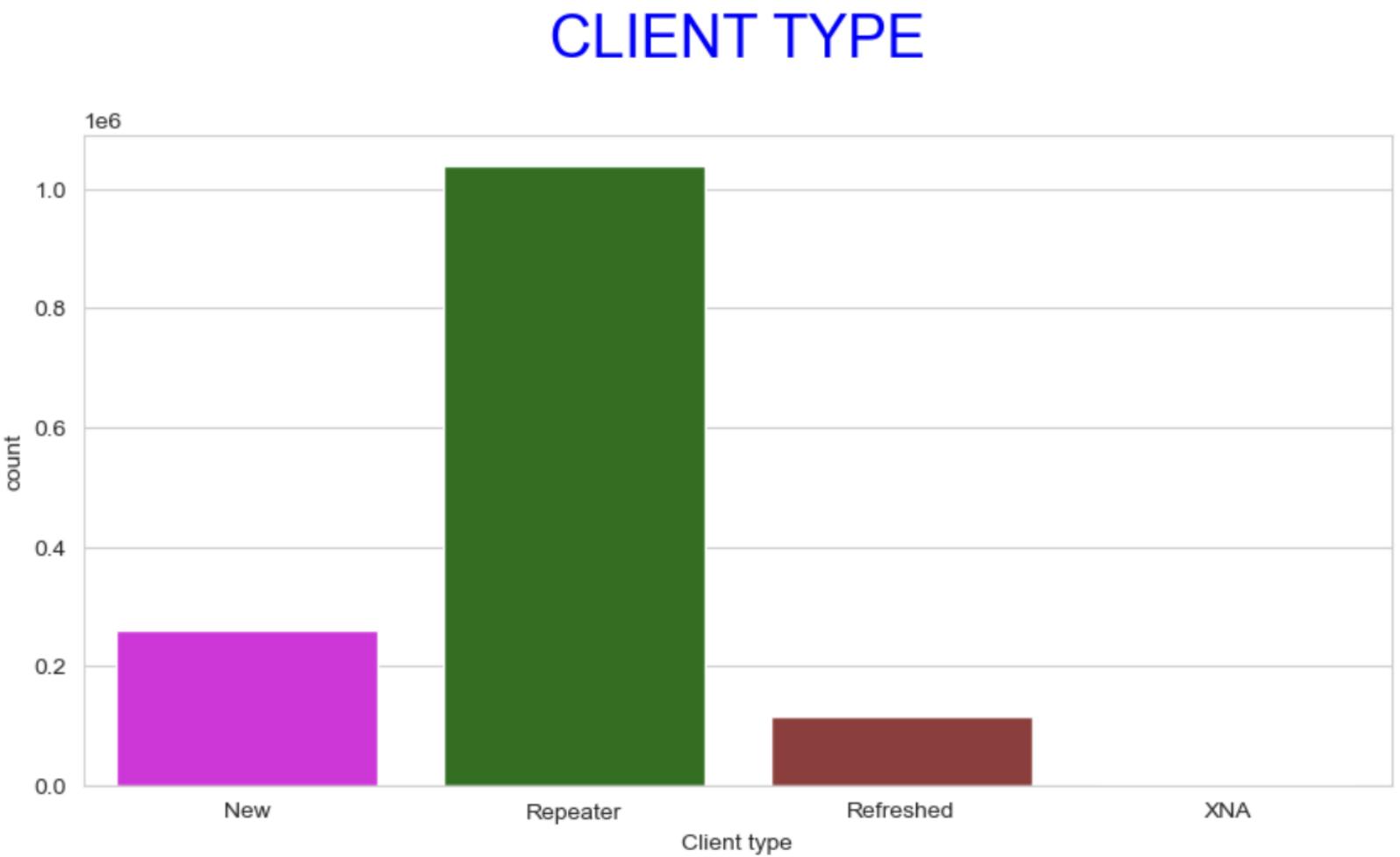
# ANALYSIS OF MERGED DATA SET ON: CURRENT APPLICATION AND PREVIOUS APPLICATION

# ANALYSIS OF LOAN STATUS

## Loan Status



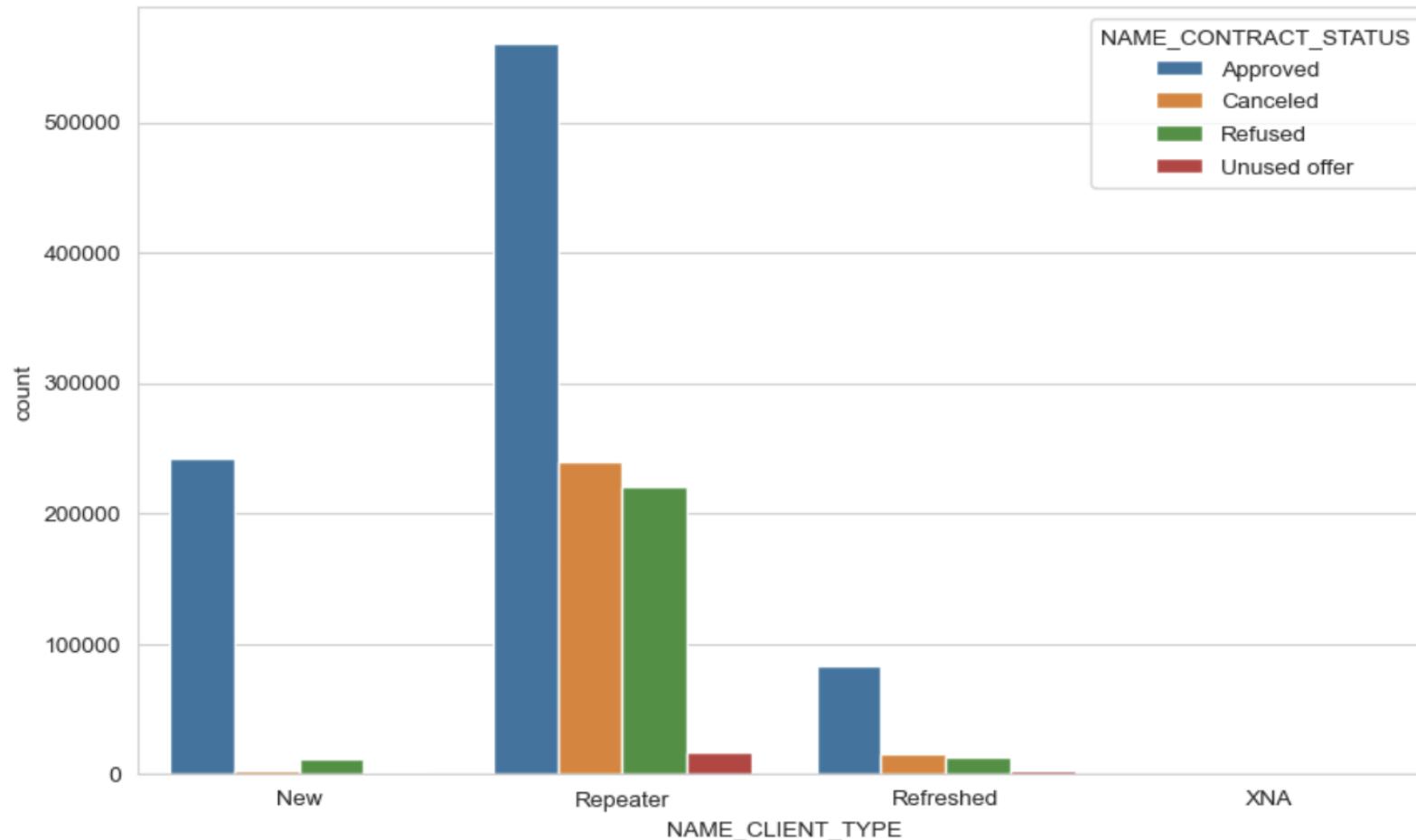
- There are huge number of Approved loan than Refused and Canceled



- Maximum loan applications are repeated

# ANALYSIS OF CLIENT TYPE ON CONTRACT STATUS

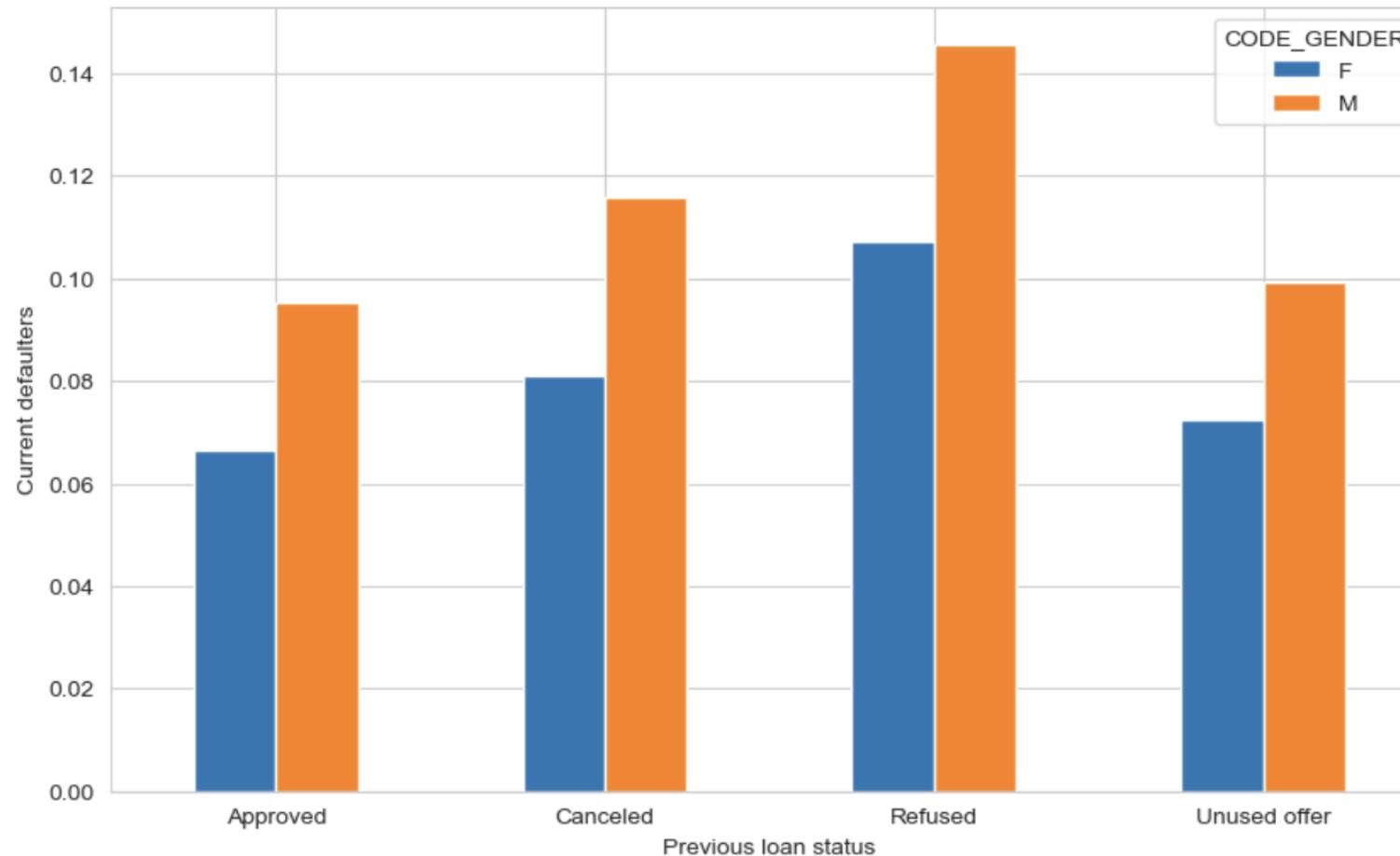
## Client Type on Contract Status



- Repeater has approved more loan in comparison to New and Refreshed

# ANALYSIS OF PREVIOUS LOAN STATUS AND CURRENT DEFAULTERS ON GENDER

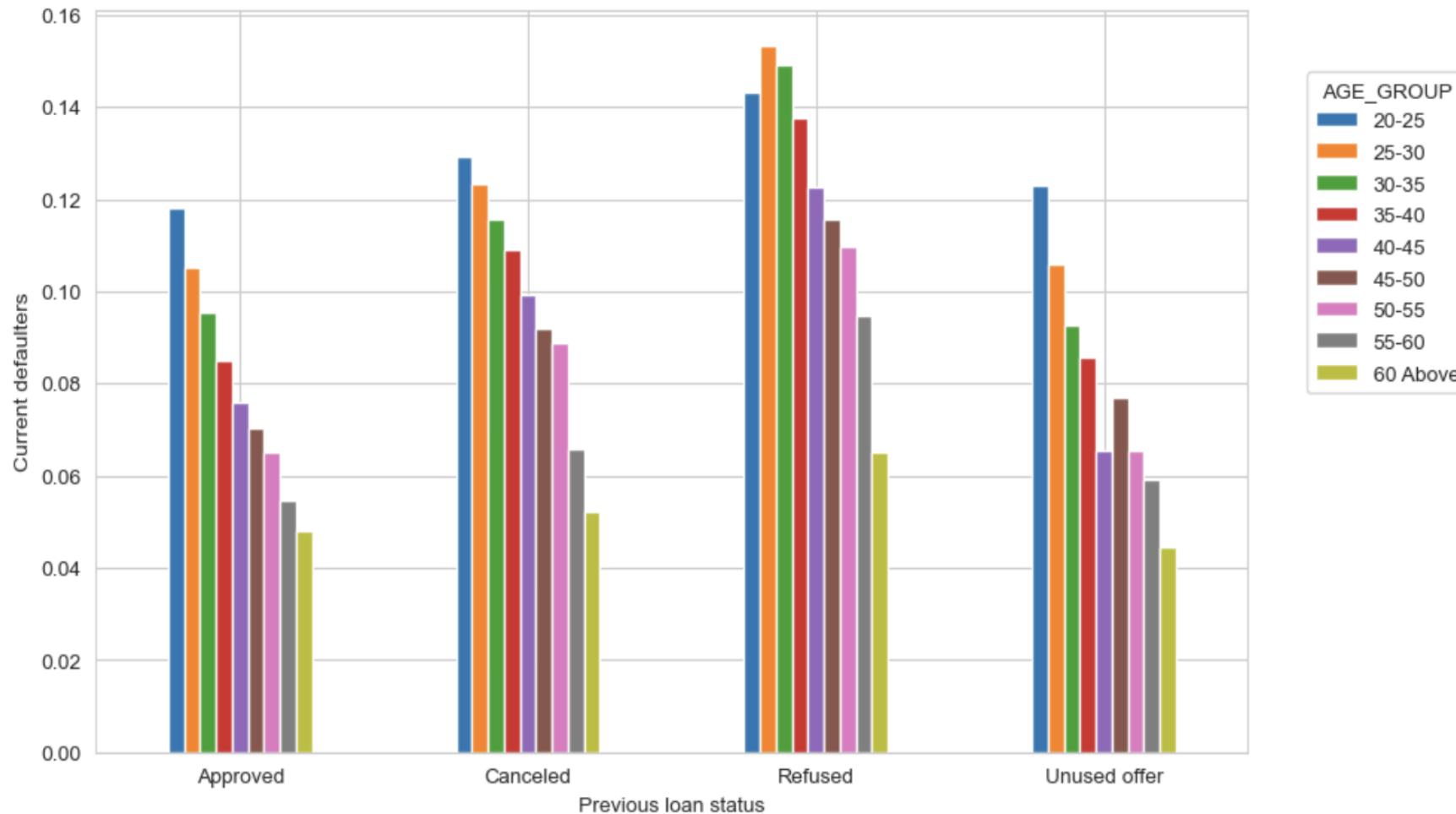
Previous Loan Status V/S Current Defaulters for Gender



- Previously Refused client is more defaulted than previously Approved clients.
- In all the cases Males are more defaulted than Females.

# ANALYSIS OF PREVIOUS LOAN STATUS AND CURRENT DEFAULTERS ON AGE GROUP

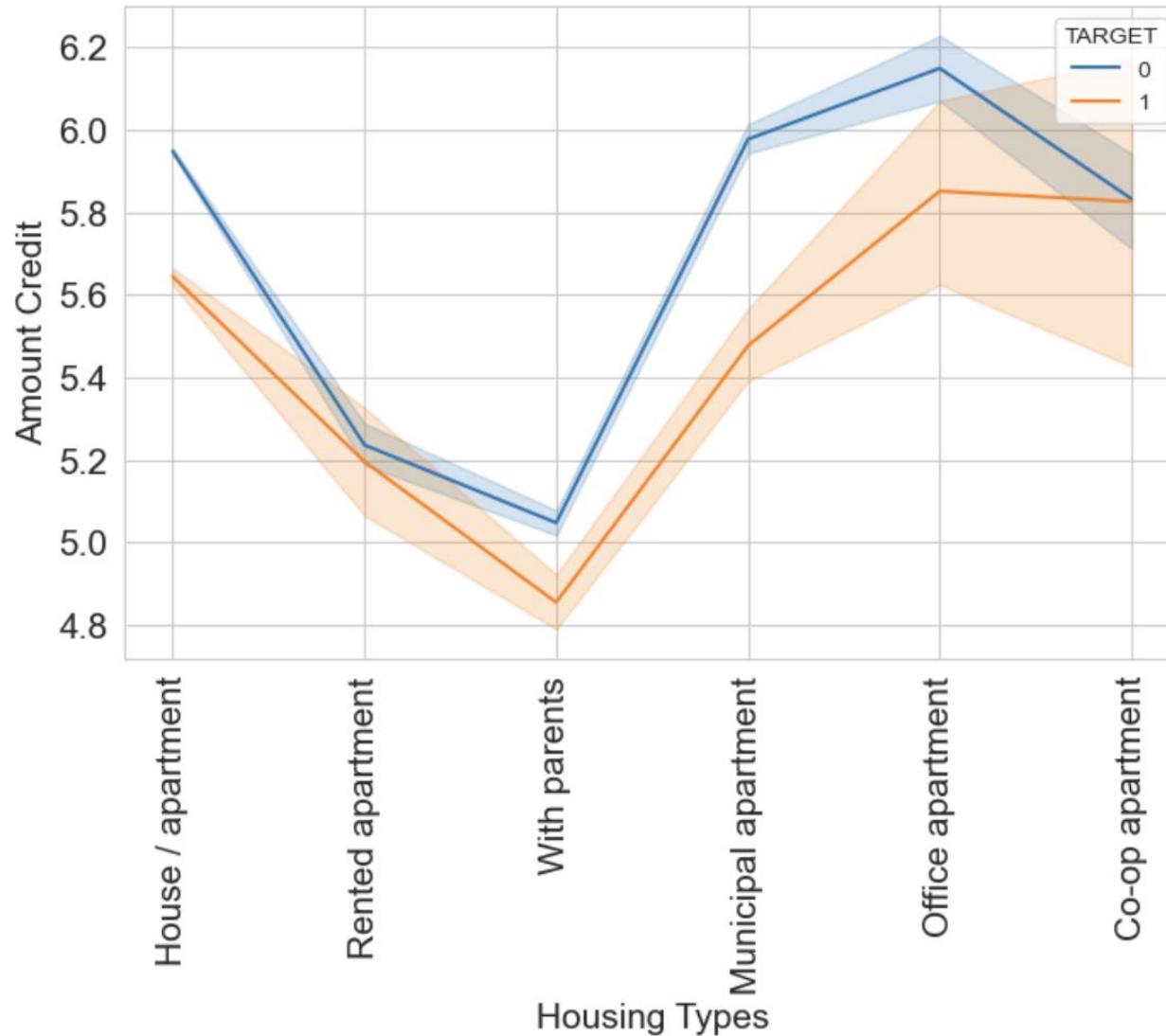
## Previous Loan Status V/S Current Defaulters for Age Group



- Age group 20-25 has more defaulter in all loan application status except refused
- Age group 60 above has less defaulter in all loan application status.
- Age group 25-30 has more refused defaulter

# ANALYSIS OF CREDIT AMOUNT AND HOUSING TYPE ON TARGET

Credit amount vs Housing type



- Office apartment is having higher credit of Non-Defaulter
- co-op apartment is having higher credit of Defaulter
- Bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.

# CONCLUSIONS

- Revolving Loans are having less defaulters than cash loan type.
- Applicants living in own House/apartment are most likely to pay on time.
- Default rate of female customers is lower than Males.
- Bank can consider more experienced workers for loan as the work experience increases the default rate decreases.
- Applicants with Academic degree can be considered for loan as they are less in defaulters.
- Unemployed and Maternity leave segments faced difficulty in repaying the loan.
- Bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.

# THANK YOU

SHEKHAR ANAND

SHEKHARANAND650@GMAIL.COM