



# Credit EDA Assignment (Bank Case Study)

THIS ASSIGNMENT AIMS TO GIVE YOU AN IDEA OF APPLYING EDA IN A REAL BUSINESS SCENARIO.

BY: MAHESH KUMAR PATRA

## Objectives:

- This case study aims to identify patterns which indicate if a client has difficulty paying their installments 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.
- 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.

## Assumptions:

- Drop the rows having missing values with more than 40% in application\_data.
- Drop the rows having missing values with more than 40% in previous\_application.

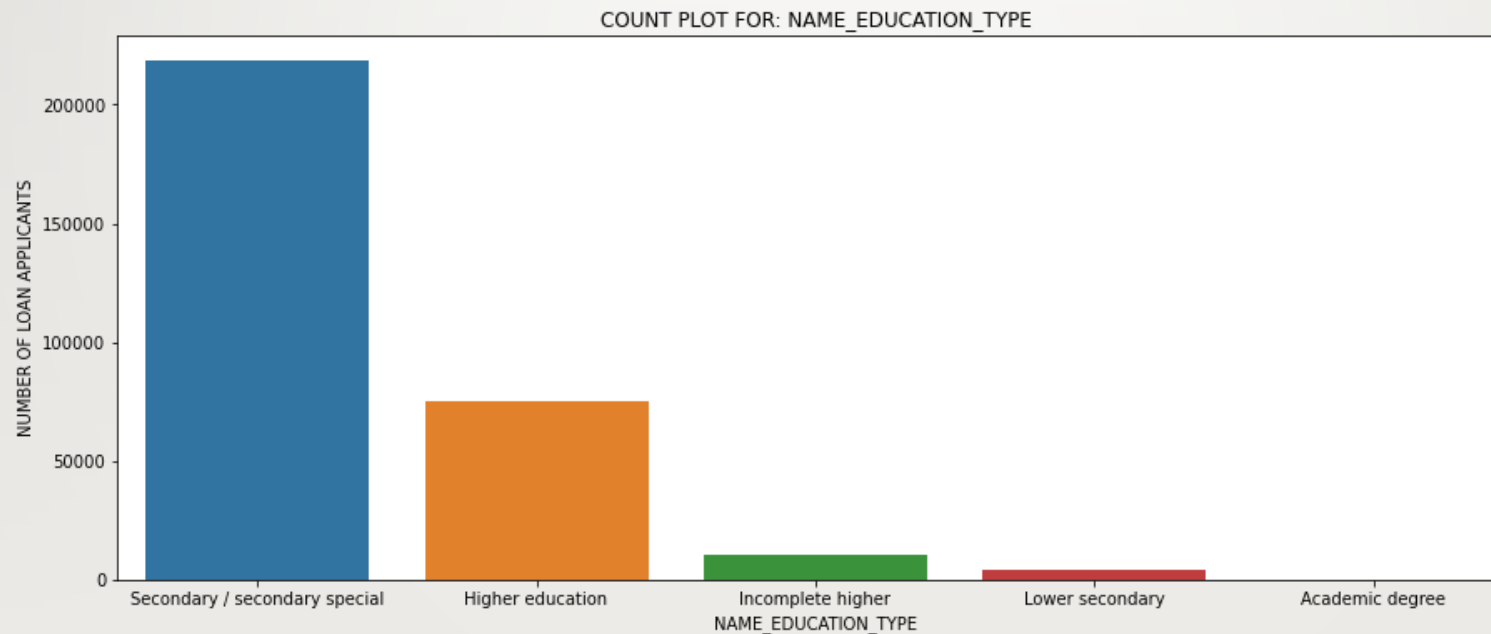
## Approach:

- Create a new feature called AGE\_GROUP in application\_data.
- Create a new feature called YEARS\_WORK\_EXP in application\_data.
- Convert the negative values to absolute values
- Convert the dtype of CNT\_FAM\_MEMBERS feature from float to int.
- In 'CODE\_GENDER' feature we have 4 rows of 'XNA' which is not belongs to any gender, so we need to drop that rows from application\_data.

# Univariate, Bivariate, Segmented univariate Analysis Graphs:

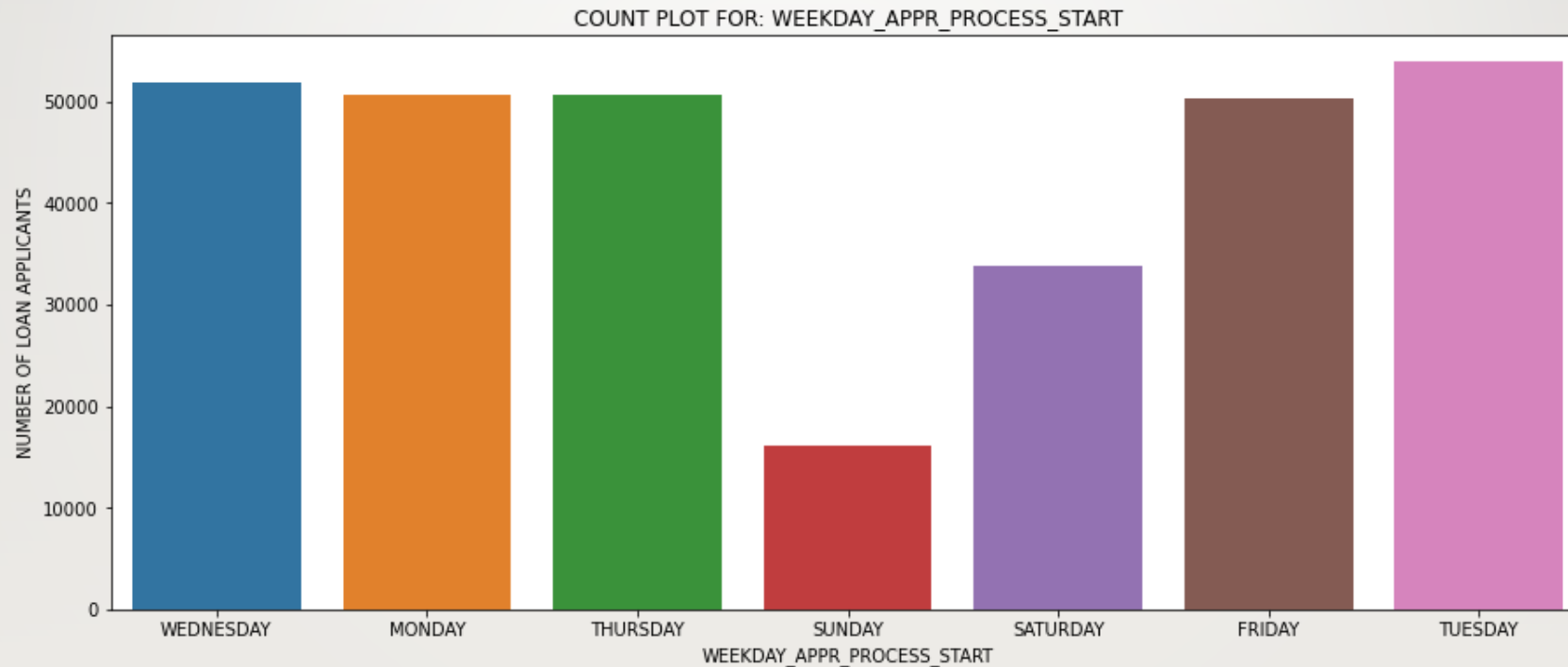
## Univariate Analysis on application data

### \* Education of the client vs Number of loan applicants



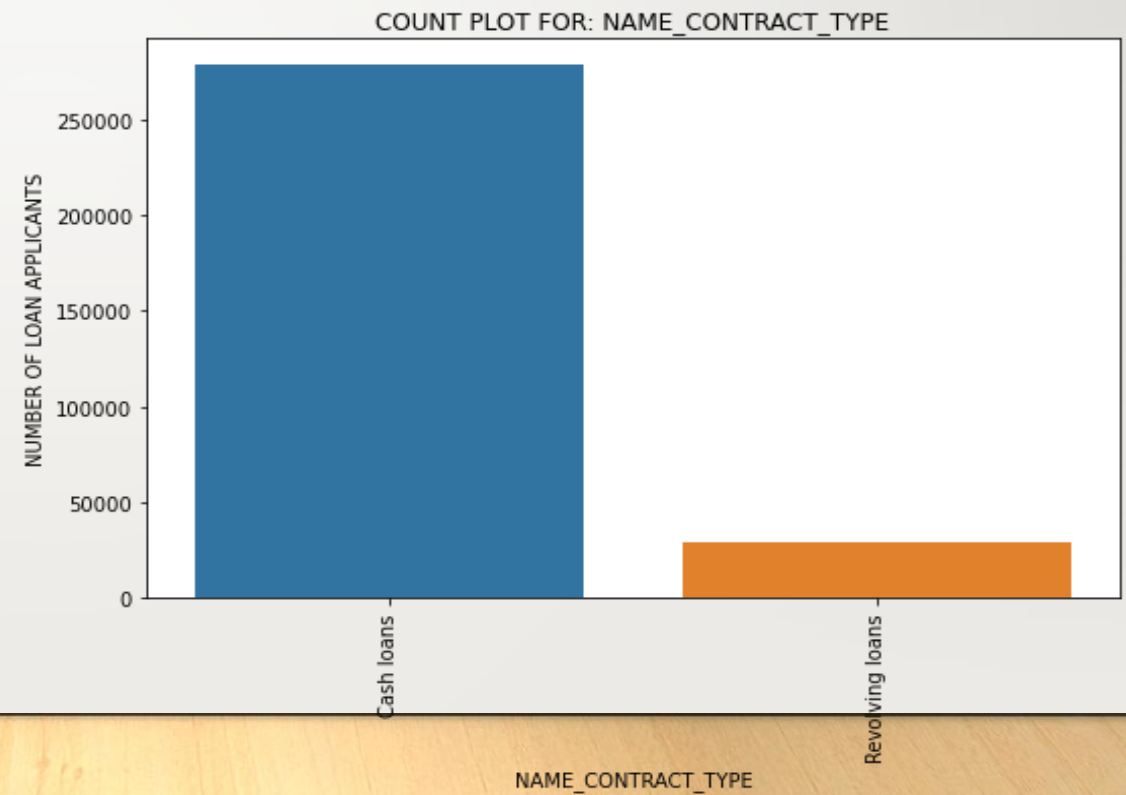
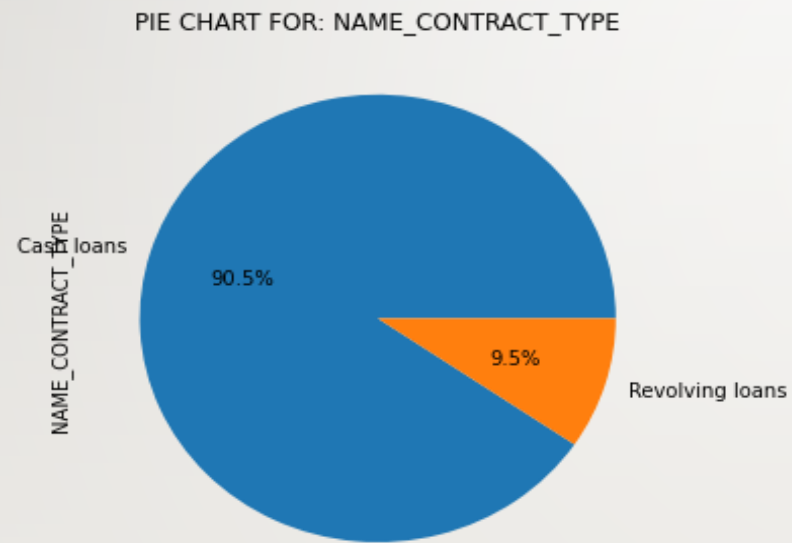
# Univariate Analysis on application data

WEEKDAY\_APPR\_PROCESS with NUMBER OF LOAN APPLICANTS

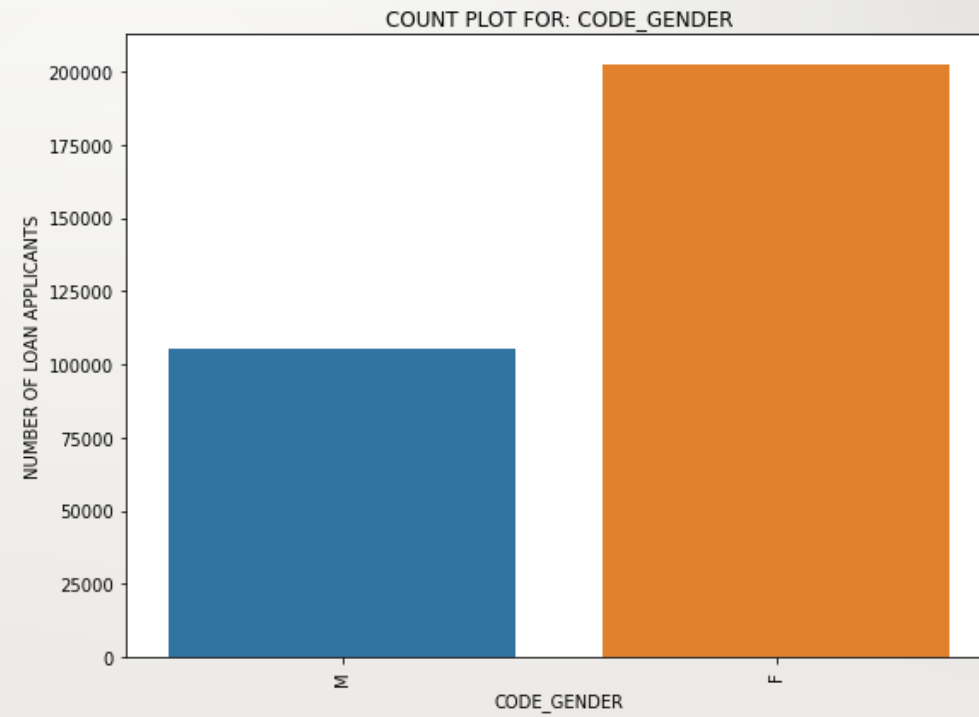
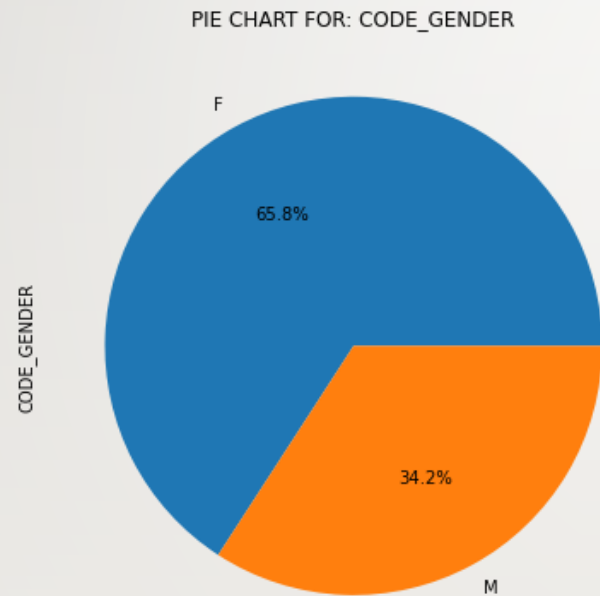




## NAME\_CONTRACT\_TYPE with NUMBER OF LOAN APPLICANTS

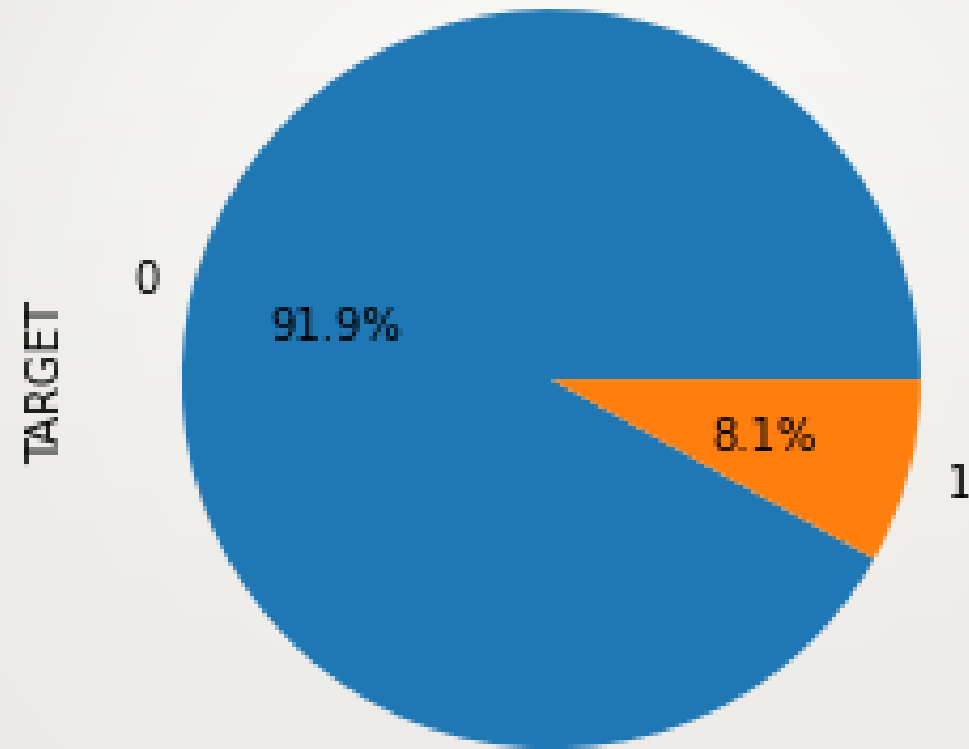


## CODE\_GENDER with NUMBER OF LOAN APPLICANTS





## Checking the imbalance of data based on 'TARGET' feature

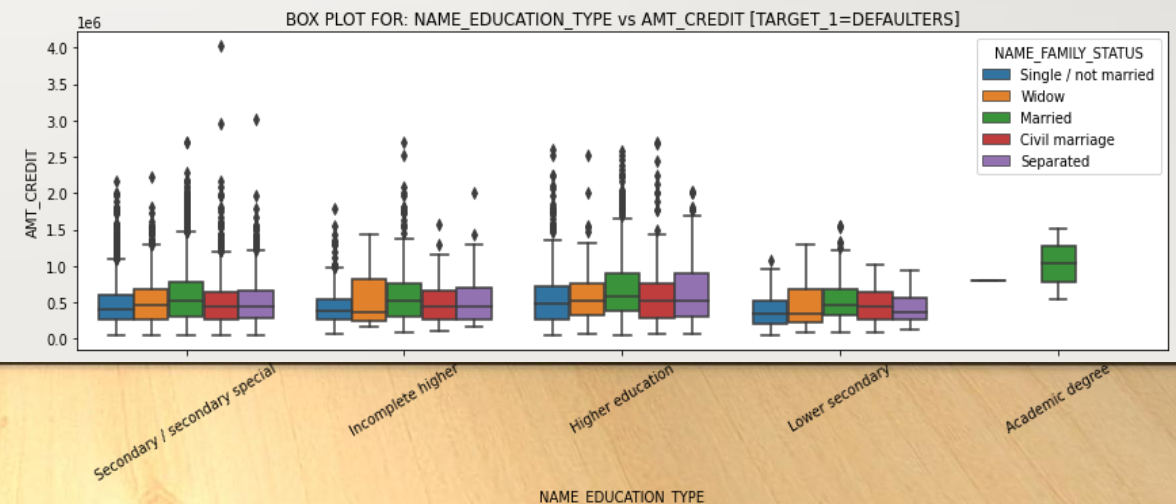
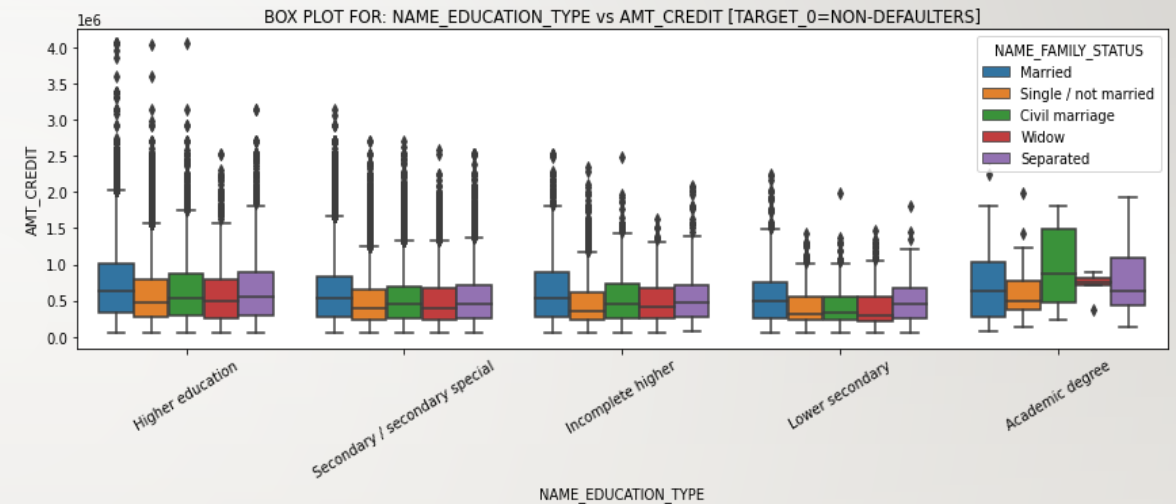


# Bivariate Analysis on application data

NAME\_EDUCATION\_TYPE vs AMT\_CREDIT

The Academic degree background applicants facing less payment difficulties and Defaulted less no of loans compare to other educational background.

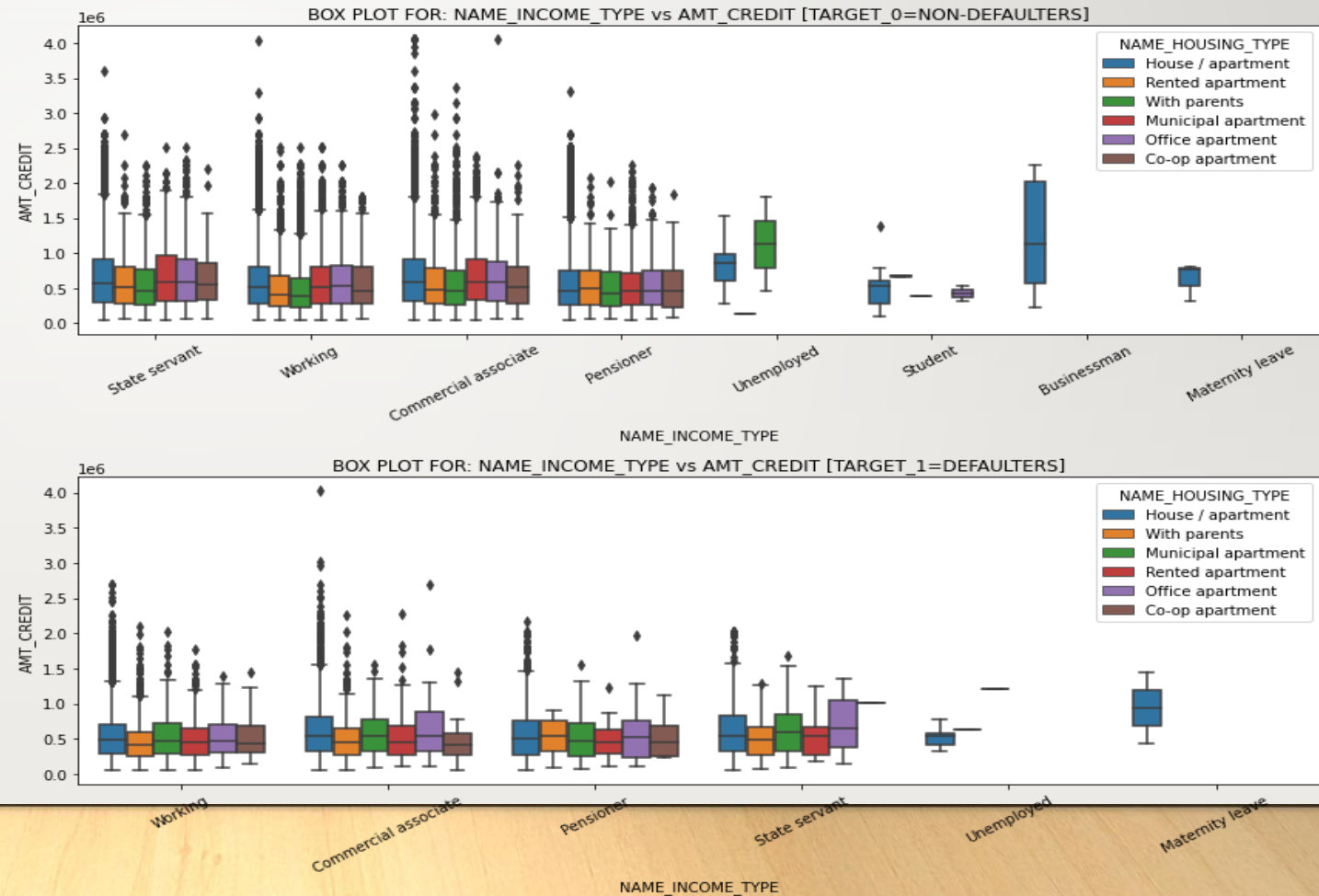
The Higher and secondary educational background applicants with married status are facing more payment difficulties with higher loan amounts.



# NAME INCOME TYPE vs AMT CREDIT based on NAME HOUSING TYPE

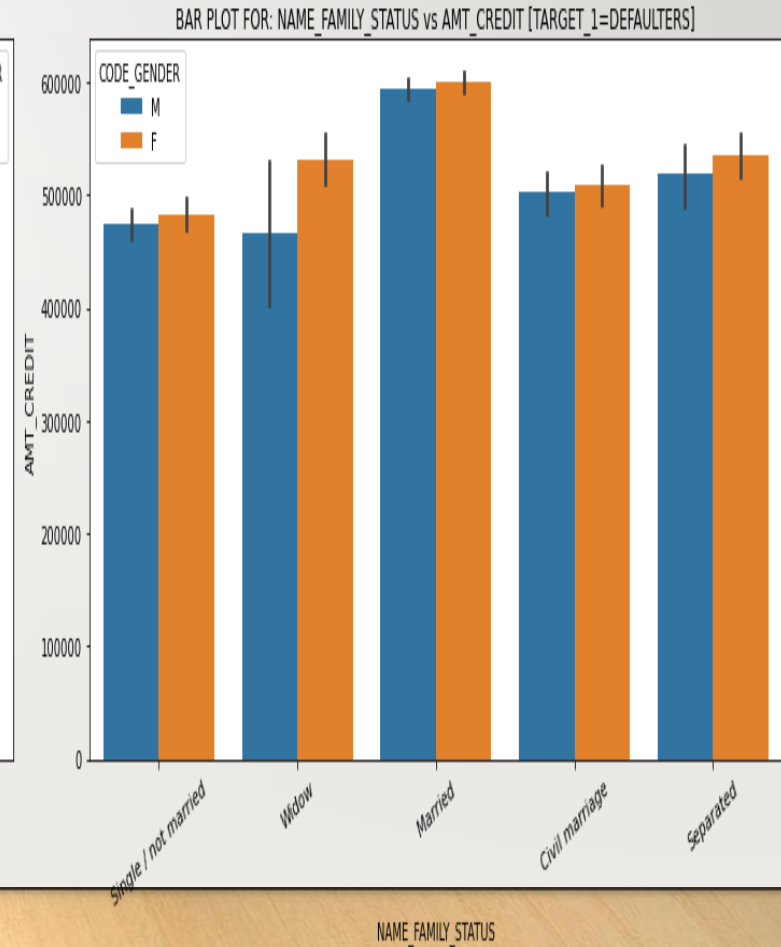
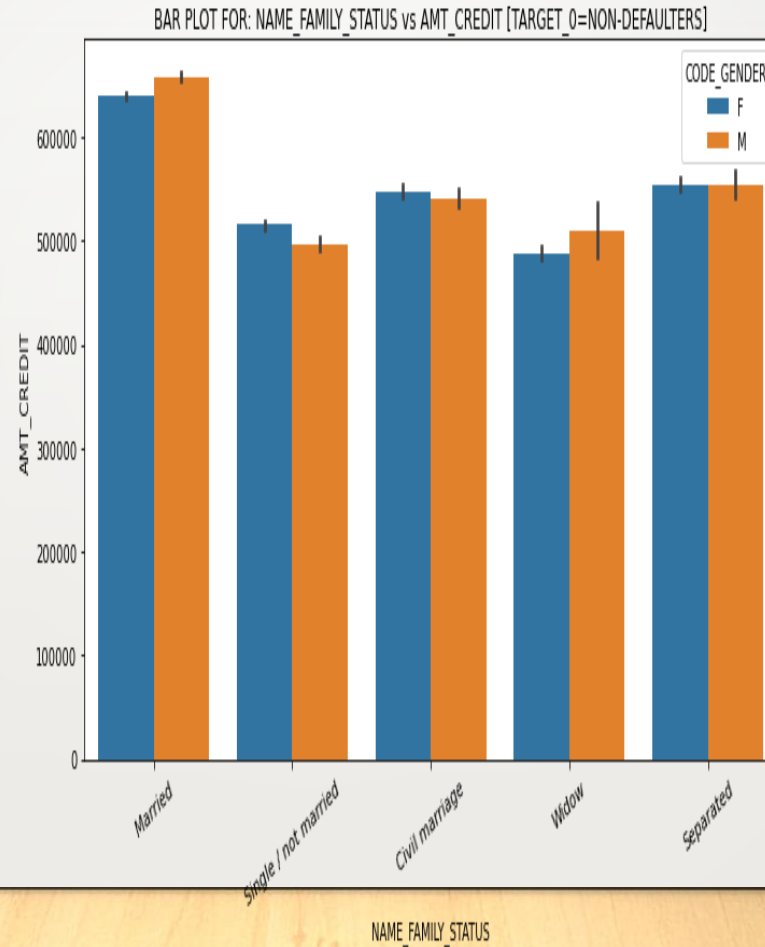
Businessman people are defaulted less number of loan compared to others.

Commercial associate and working people who lives in apartment have defaulted loans with higher loan amounts.



# NAME\_FAMILY\_STATUS vs AMT\_CREDIT based on CODE\_GENDER

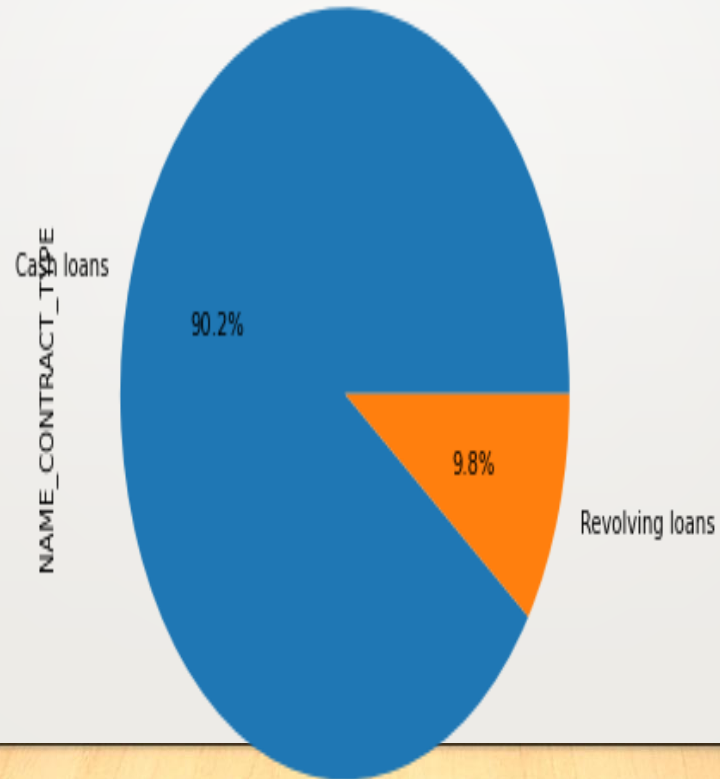
Number of Defaulters are higher in females compare to male .



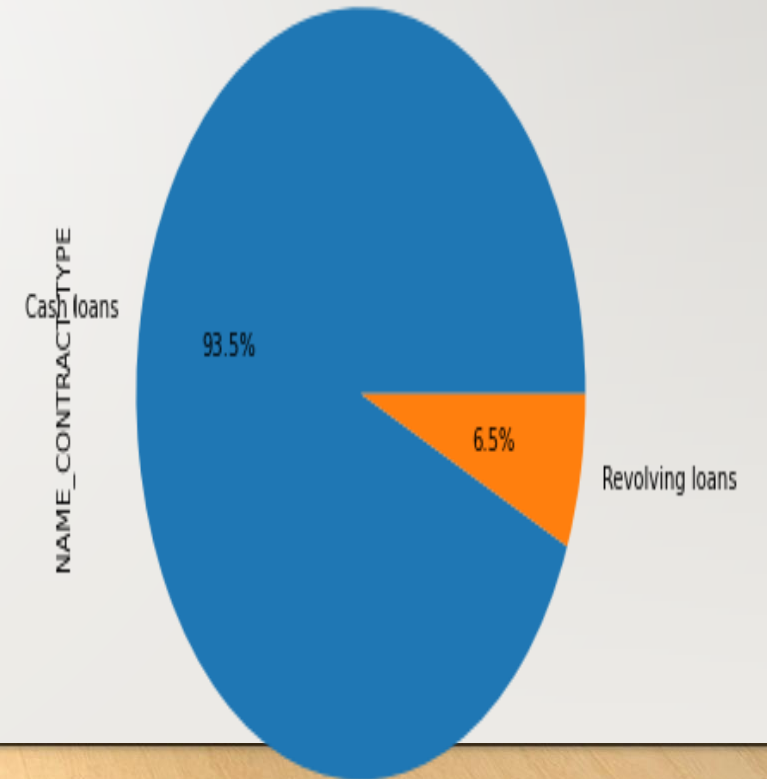
# Segmented univariate Analysis on application data

Cash loans  
contract type  
defaulted loans  
higher than  
revolving contract  
type.

PIE CHART FOR: NAME\_CONTRACT\_TYPE [TARGET\_0=NON-DEFAULTERS]



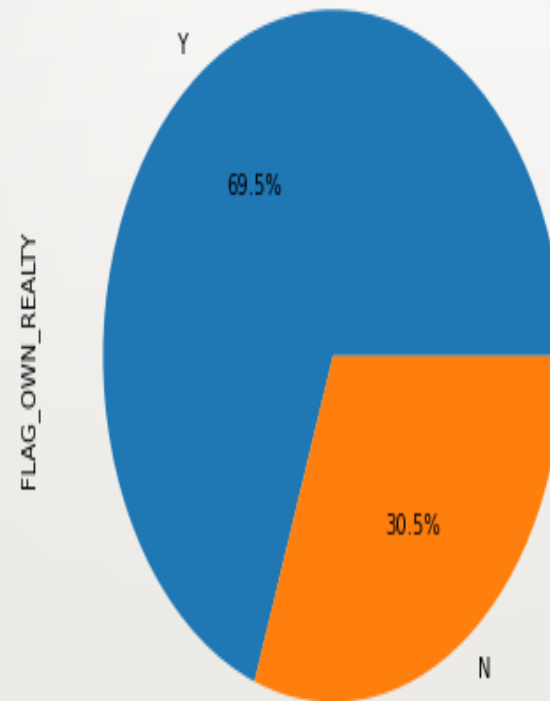
PIE CHART FOR: NAME\_CONTRACT\_TYPE [TARGET\_1=DEFAULTERS]



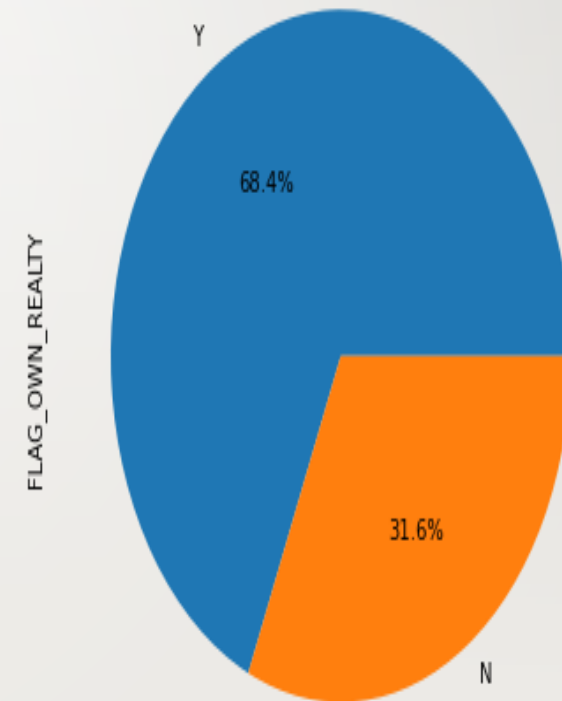


Clients have own house or flat defaulted loans higher than clients don't have house or flat.

PIE CHART FOR: FLAG\_OWN\_REALTY [TARGET\_0=NON-DEFAULTERS]

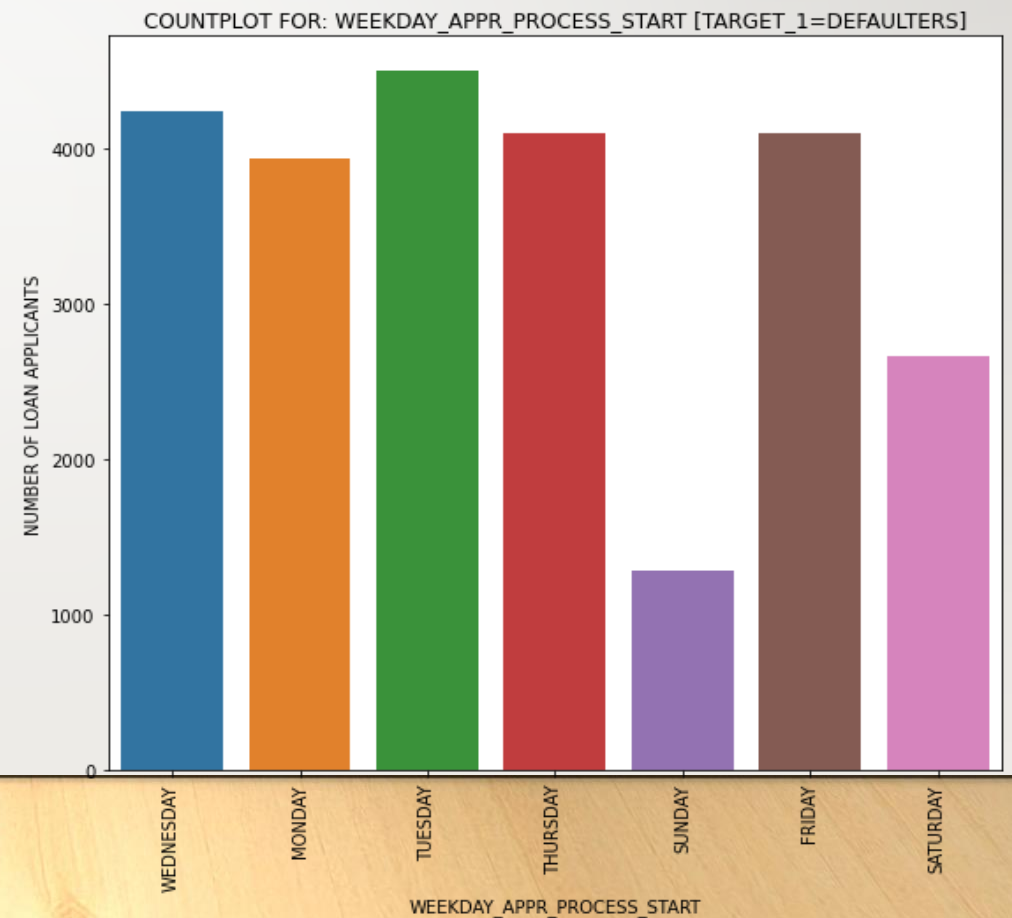
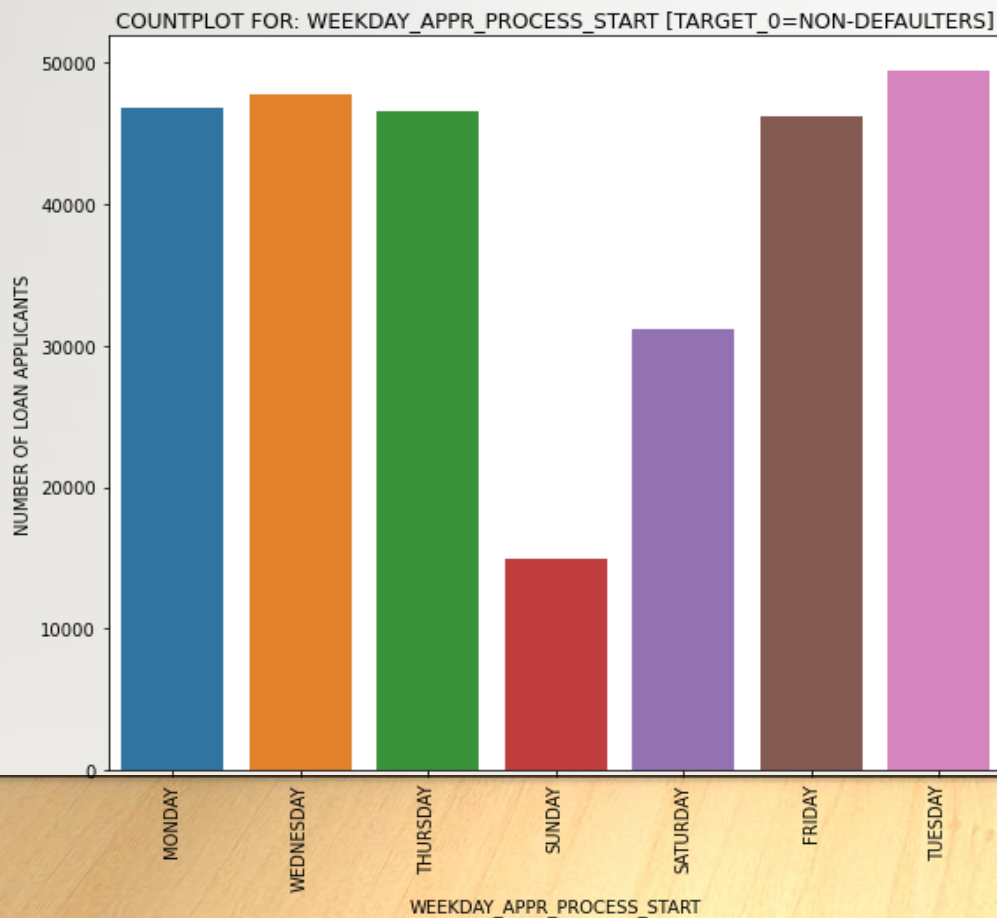


PIE CHART FOR: FLAG\_OWN\_REALTY [TARGET\_1=DEFAULTERS]

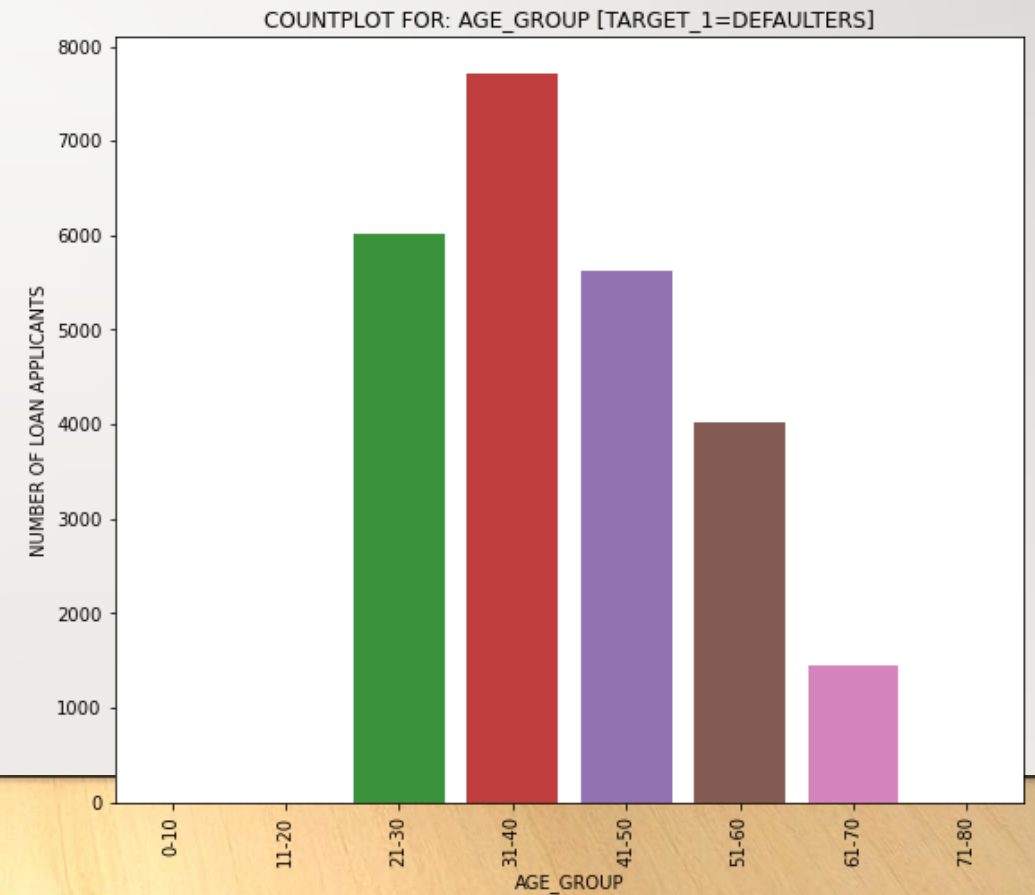
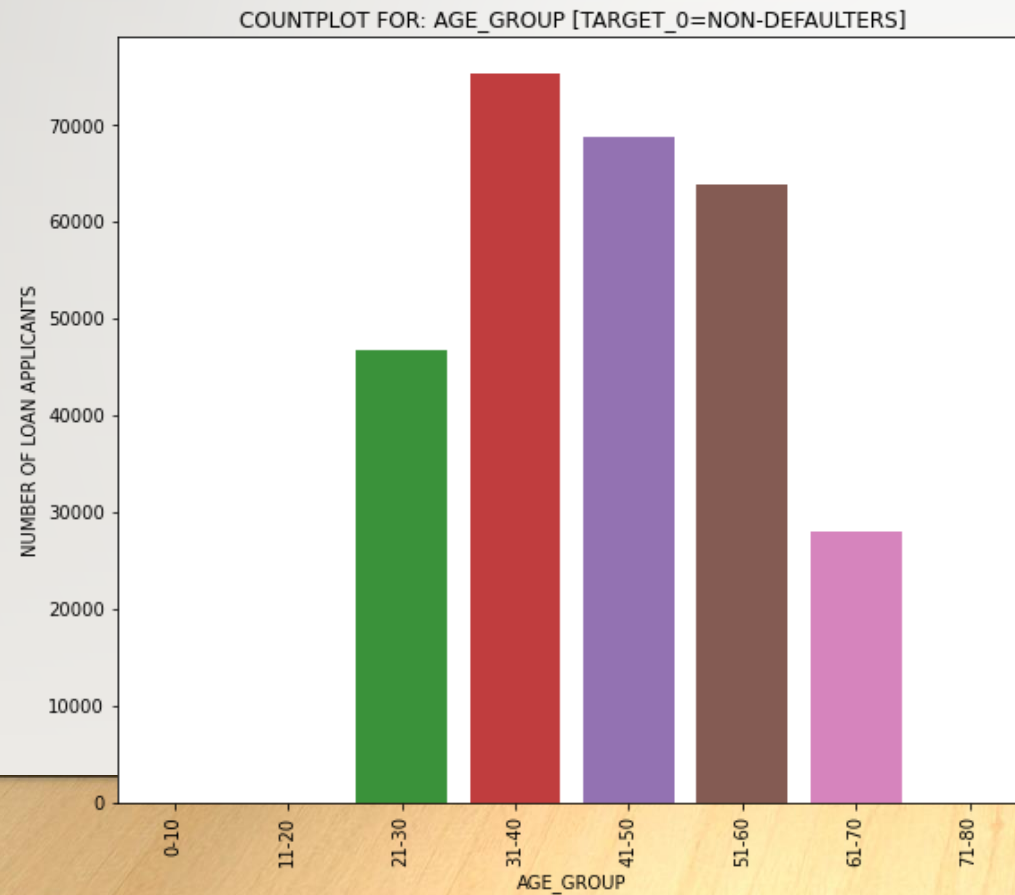




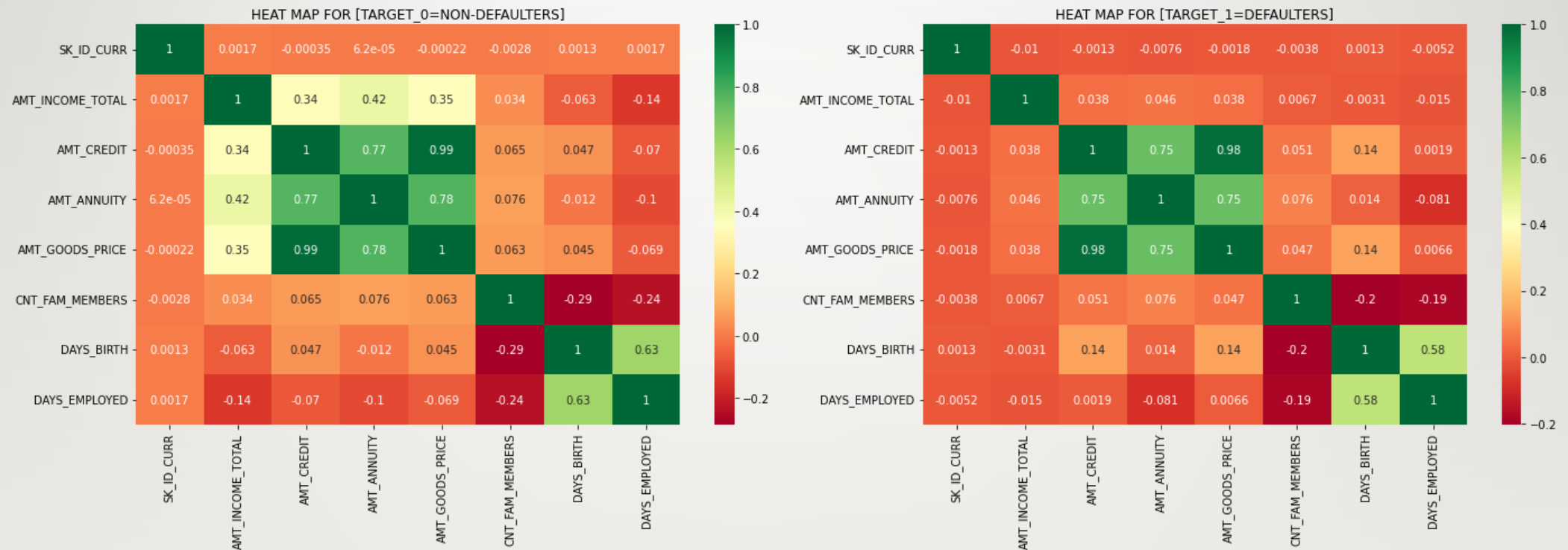
Maximum loans are applied on Tuesday.



Age group between 31-40 are facing payment difficulties.

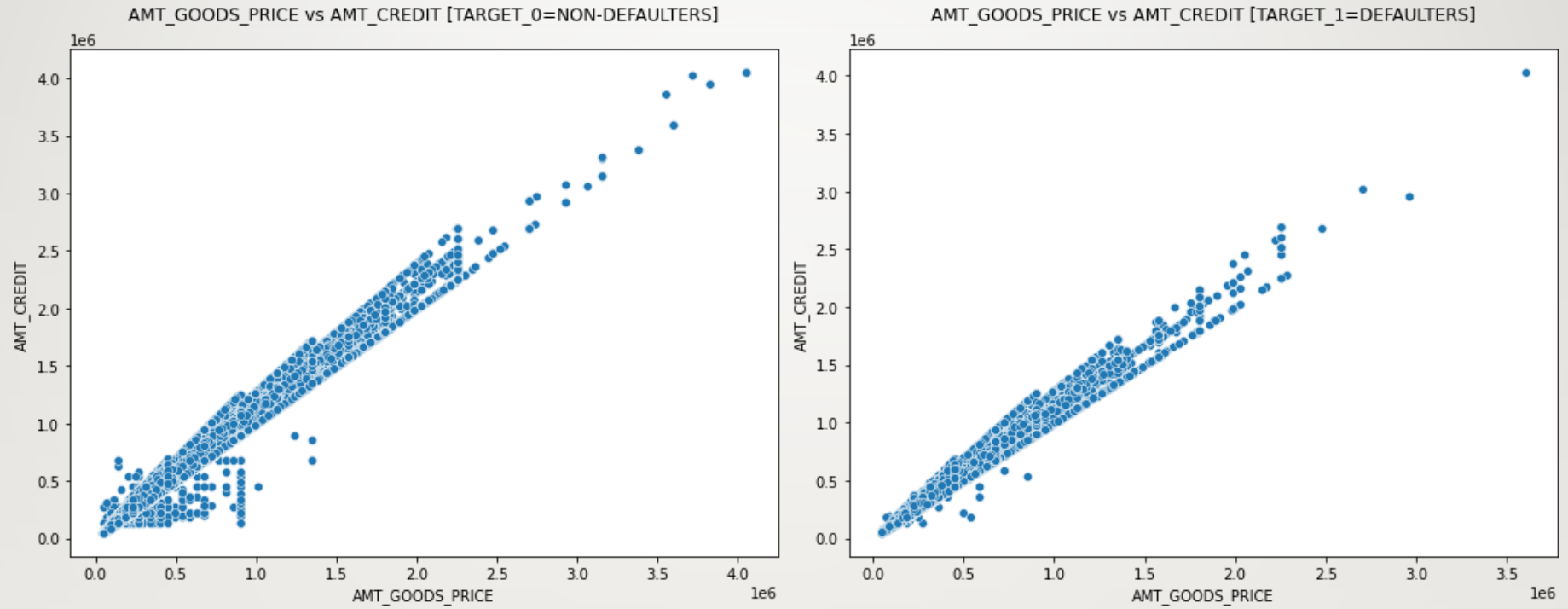


# Multivariate Analysis on application\_data



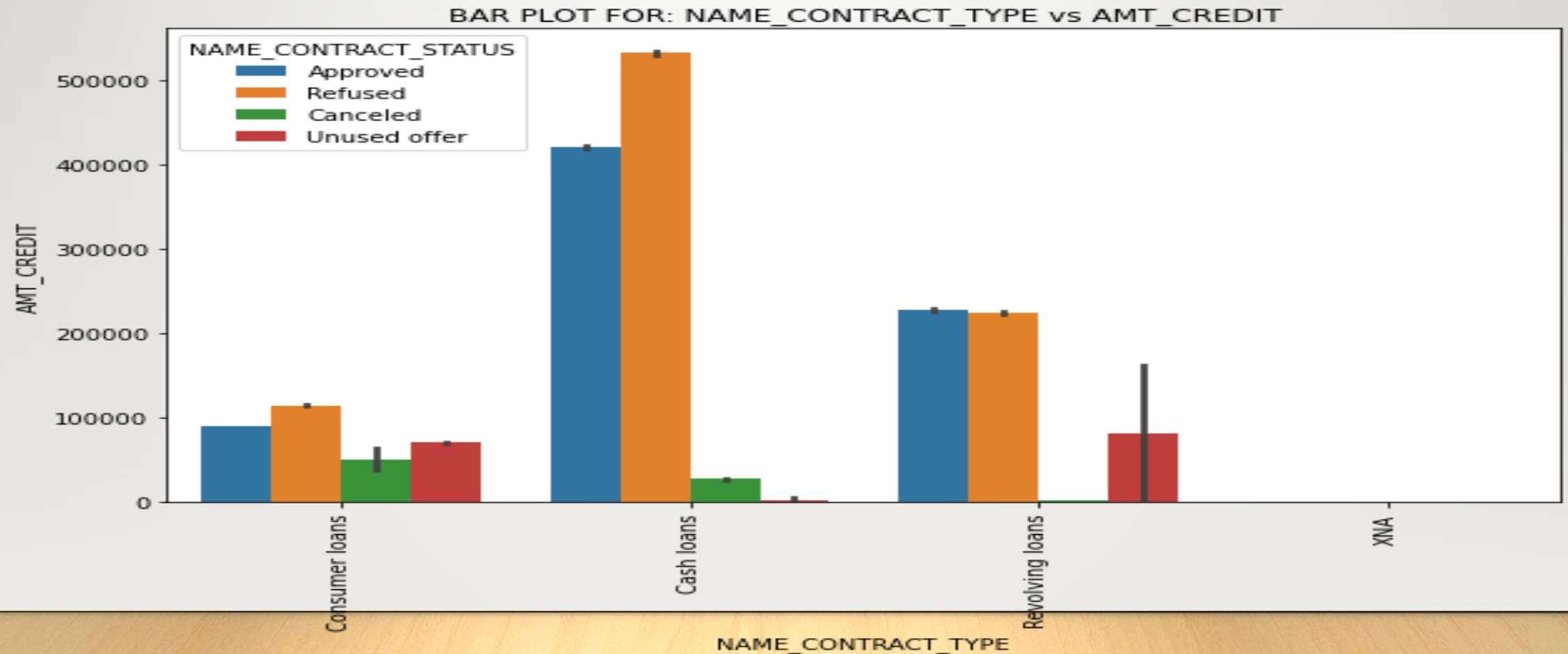
From the above heat map , There is a strong correlation between credit amount of the loan and the price of the goods for which the loan is given.

From the above heat map , There is a strong correlation between Client's age and clients employment.

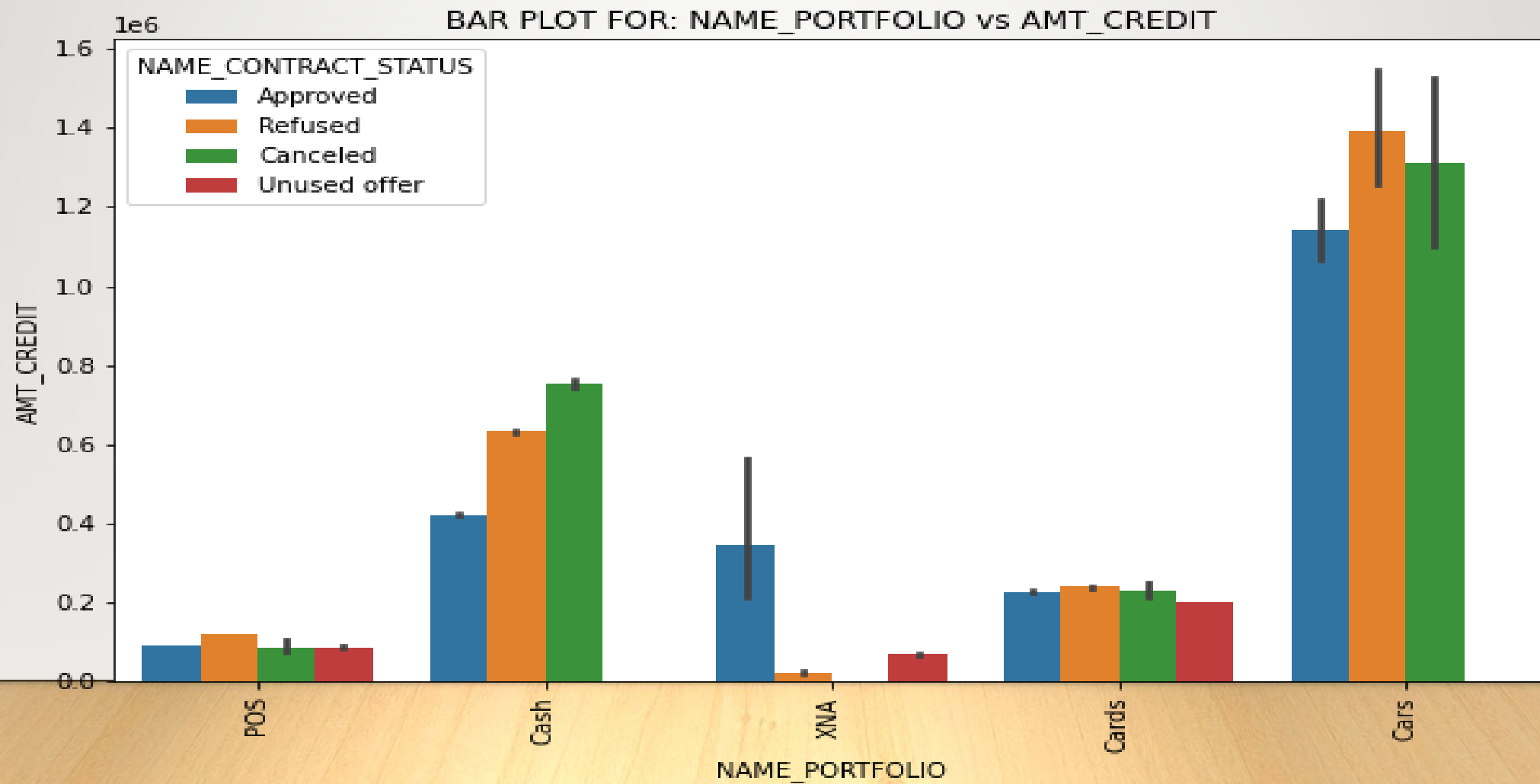


The above scatterplot represents , There is a strong correlation between Client's age and clients employment.

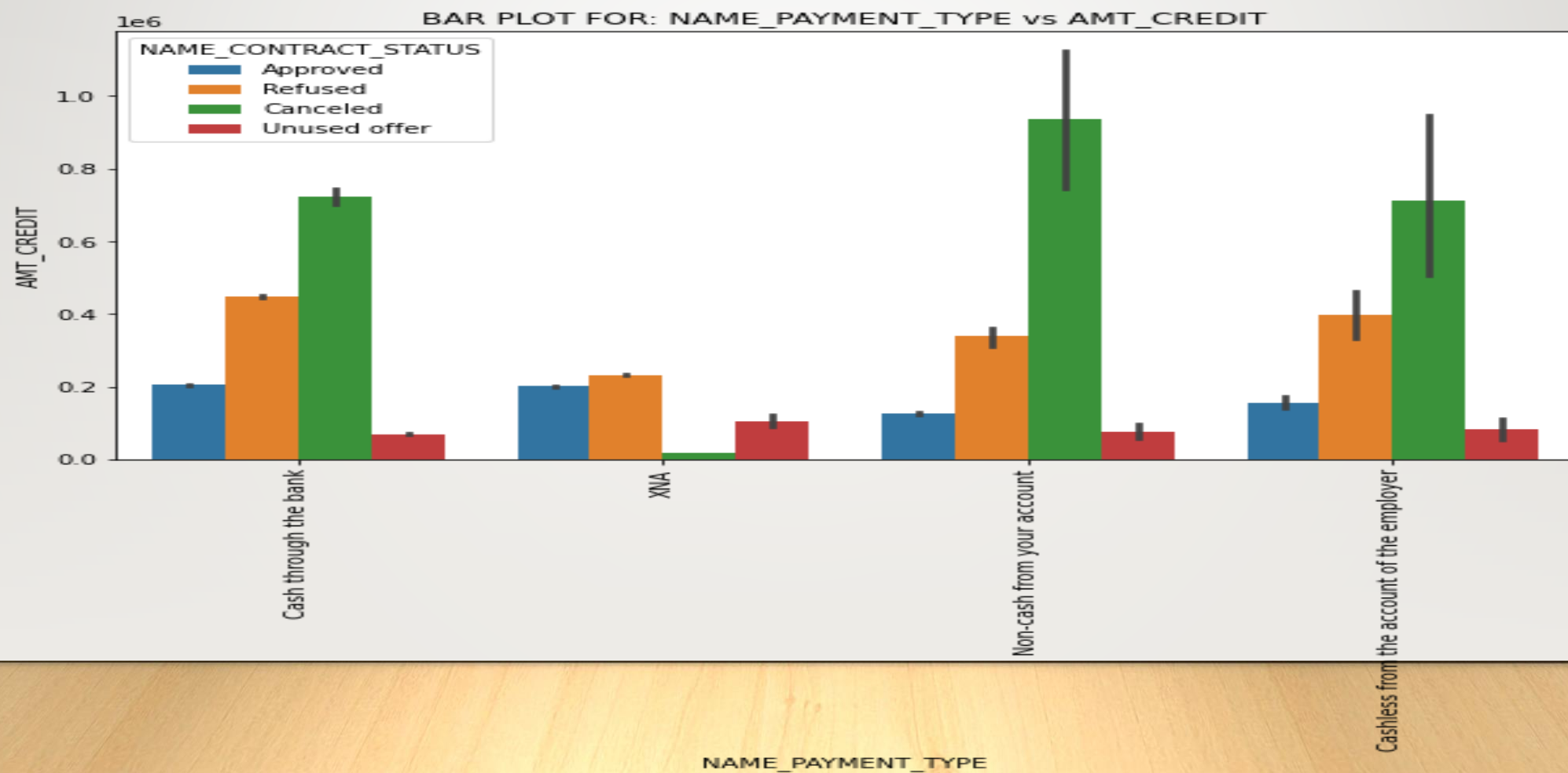
# Analysis on Previous application



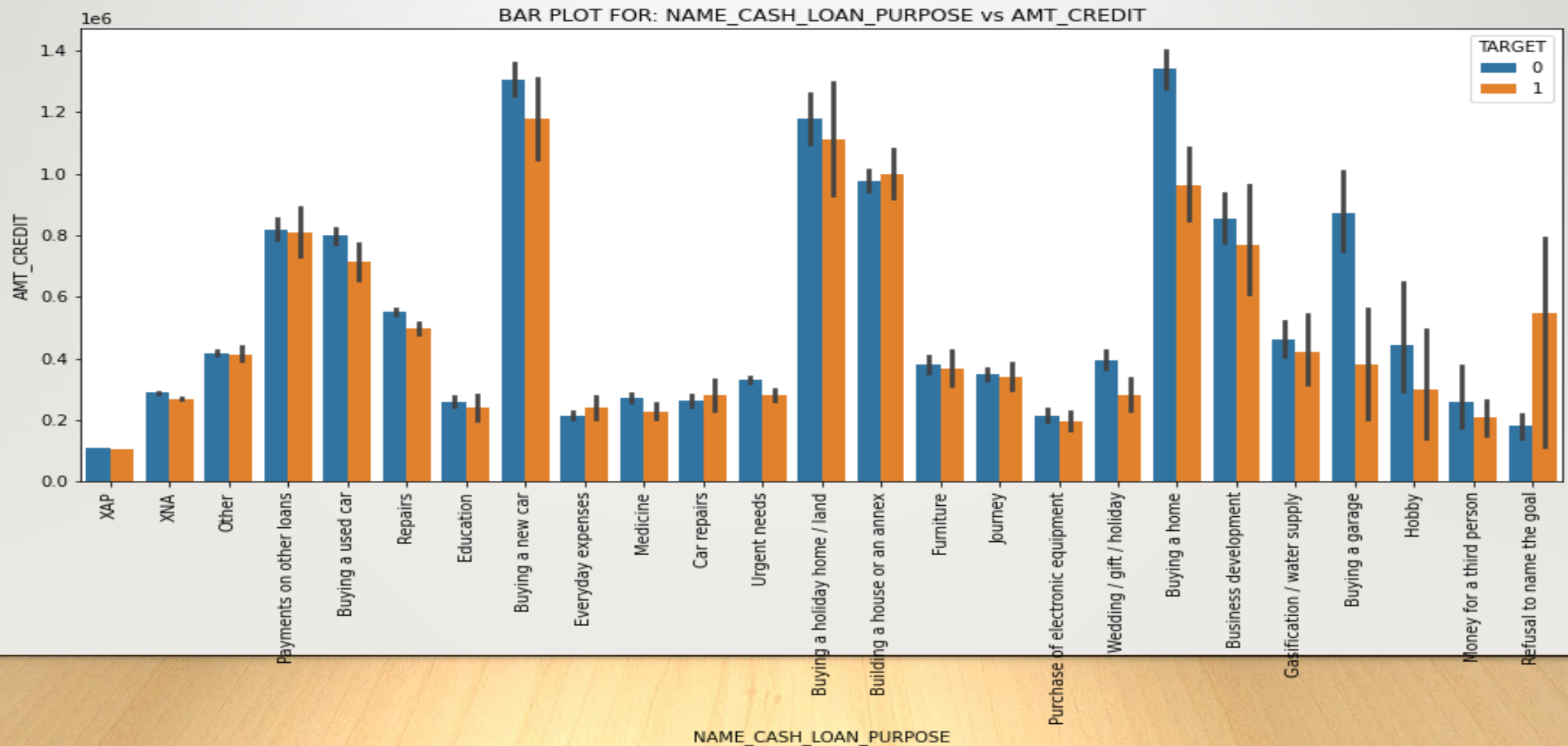
1000000

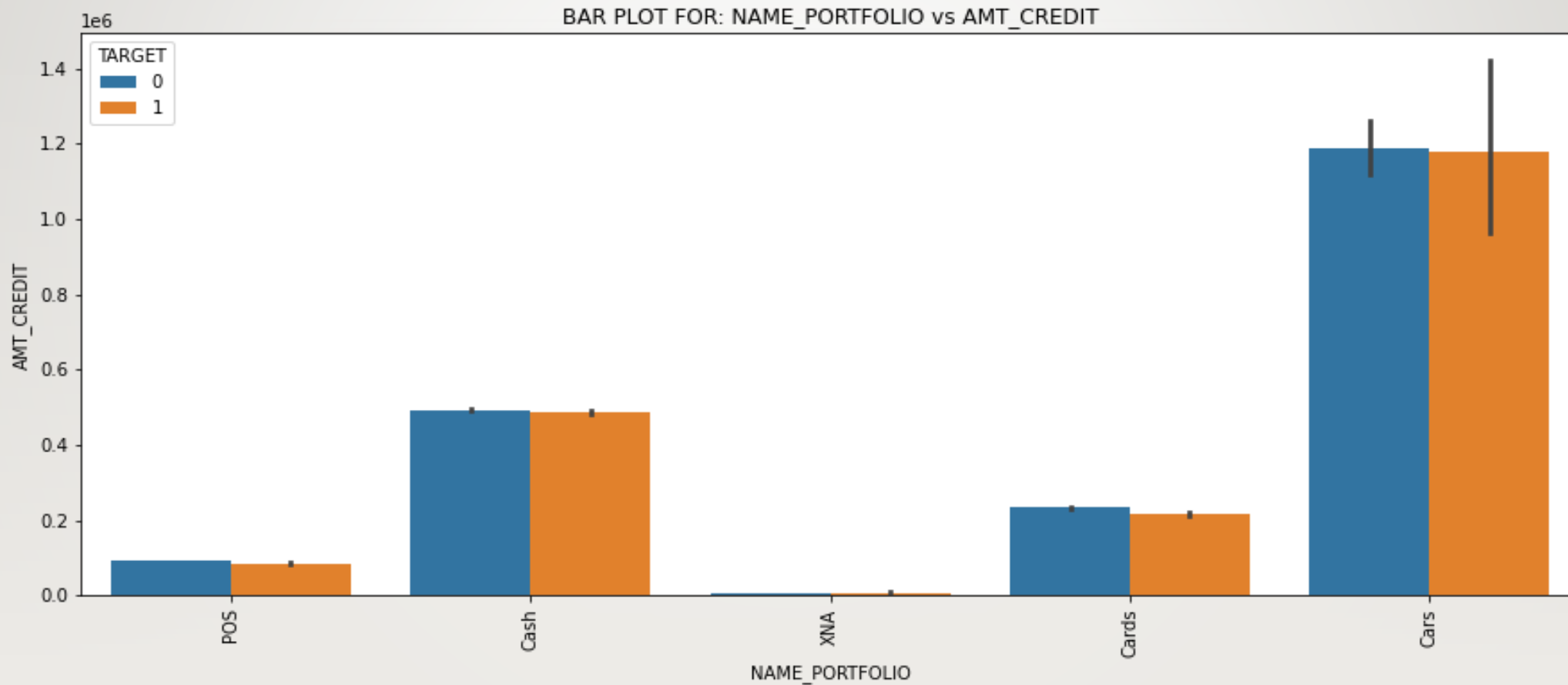


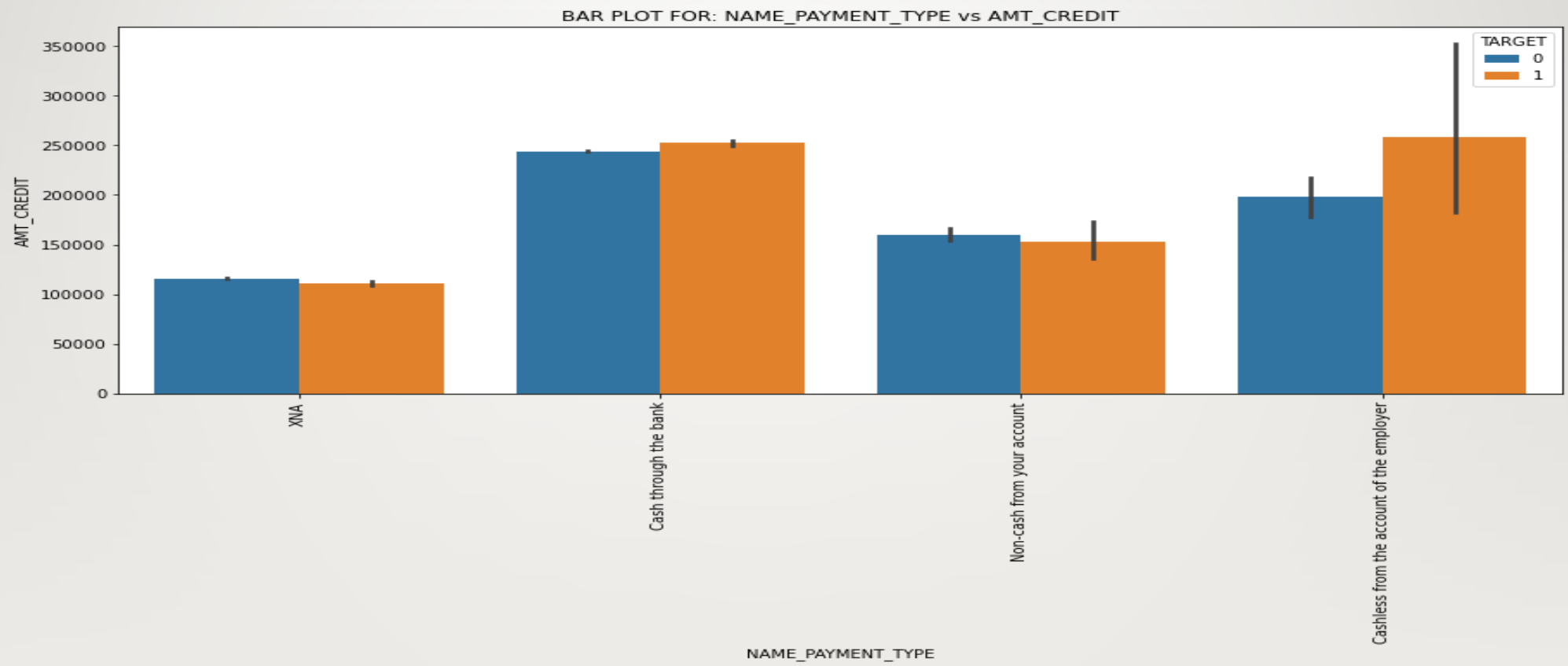


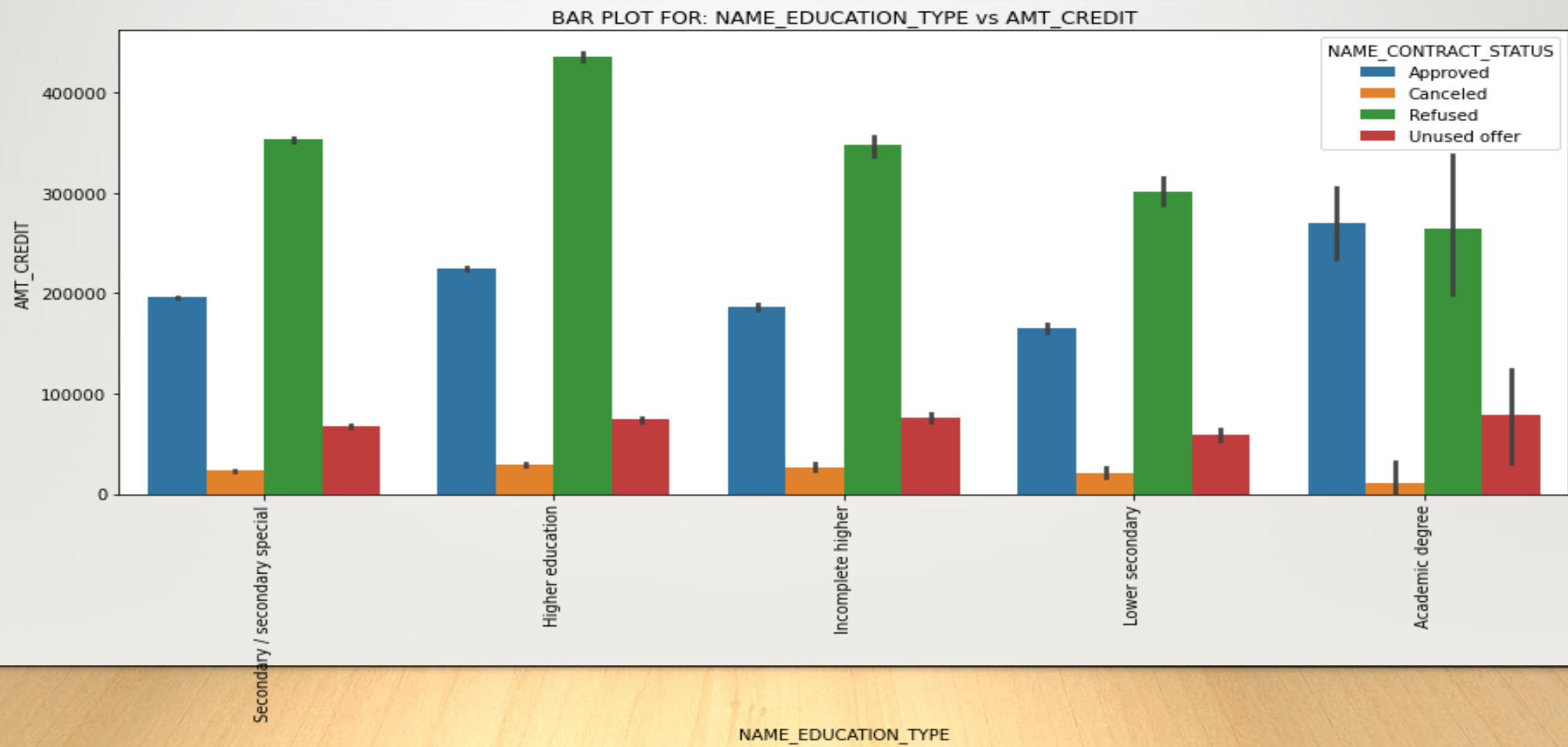


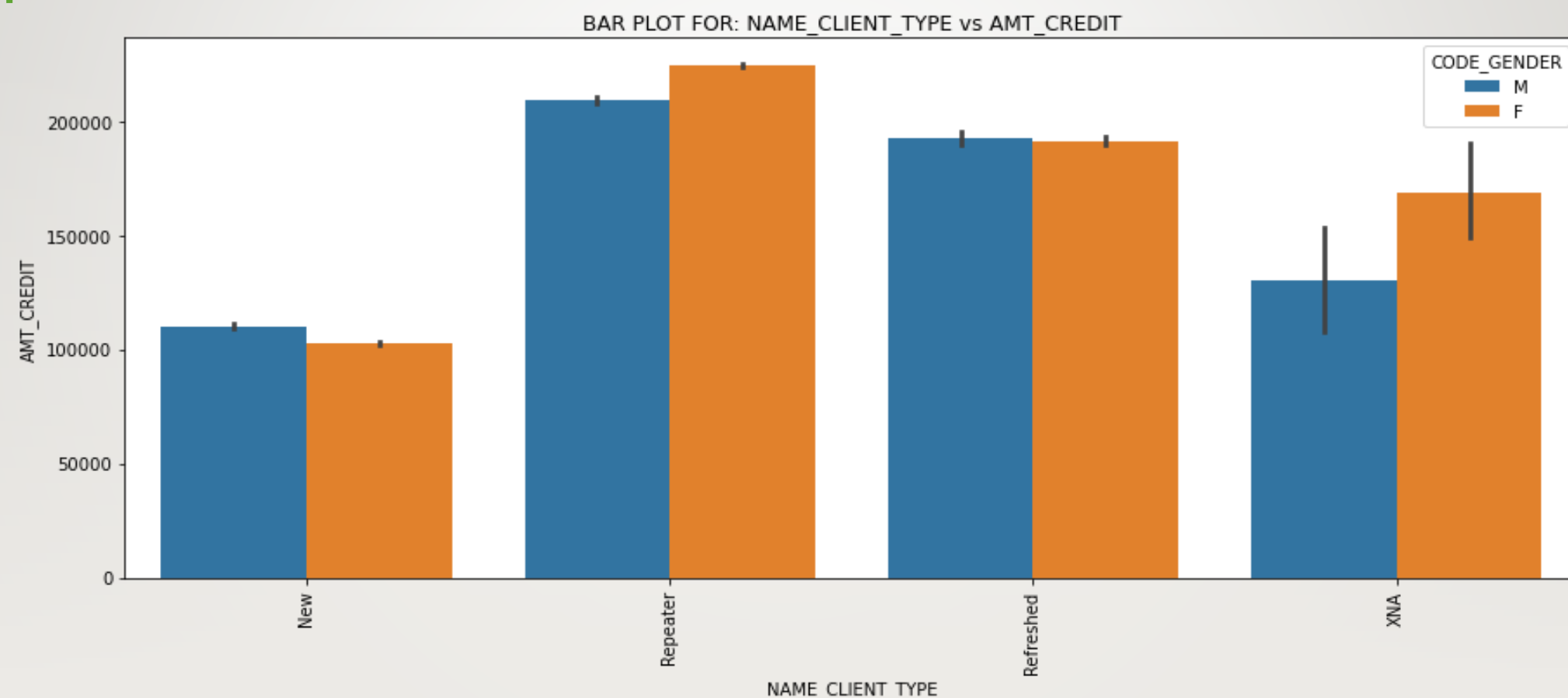
# Analysis on merged dataframe















**THANK YOU**