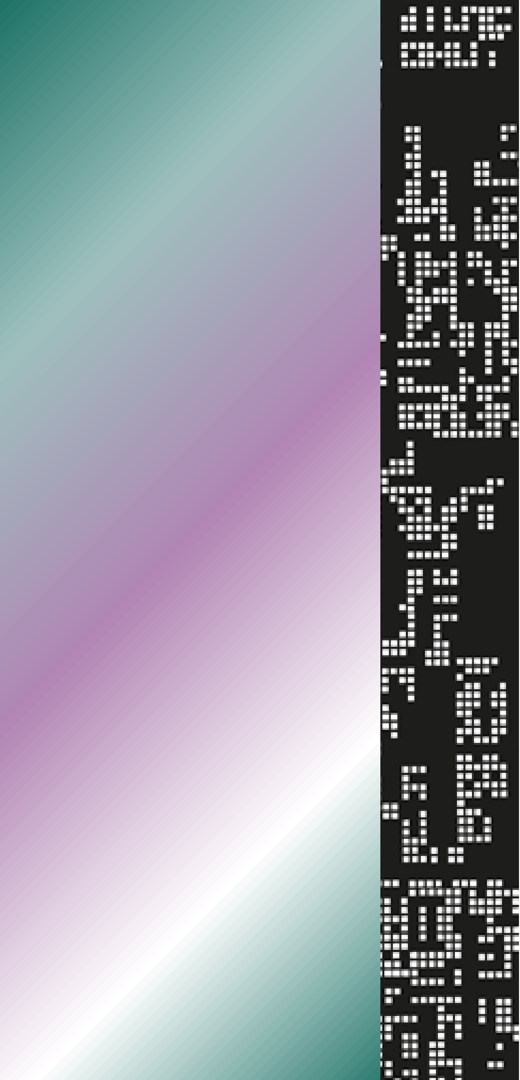


Credit EDA Case Study

Hrishikesh Pradhan

Table of contents

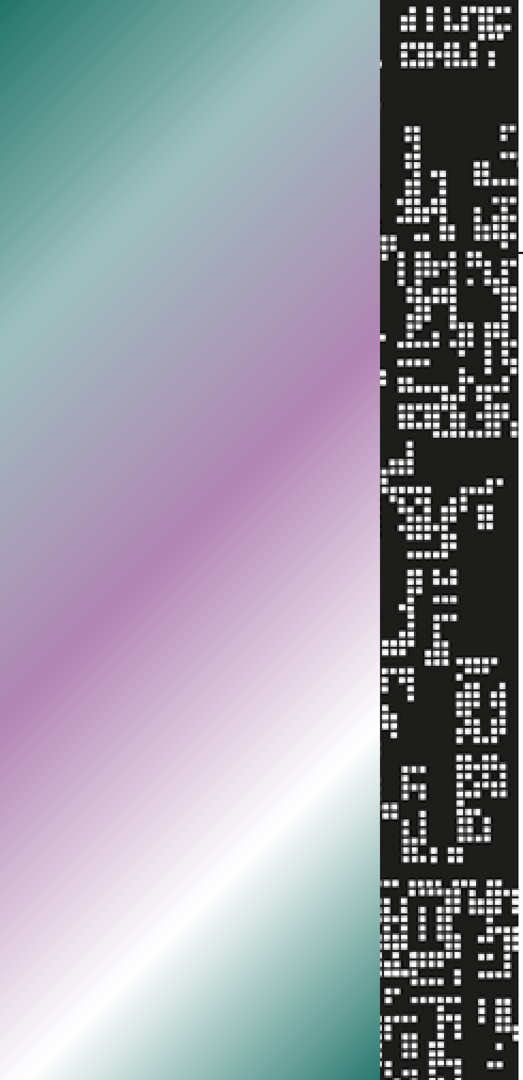
01 Problem Statement

02 Importing & Cleaning

03 Approach

04 Data visualisation & Insights

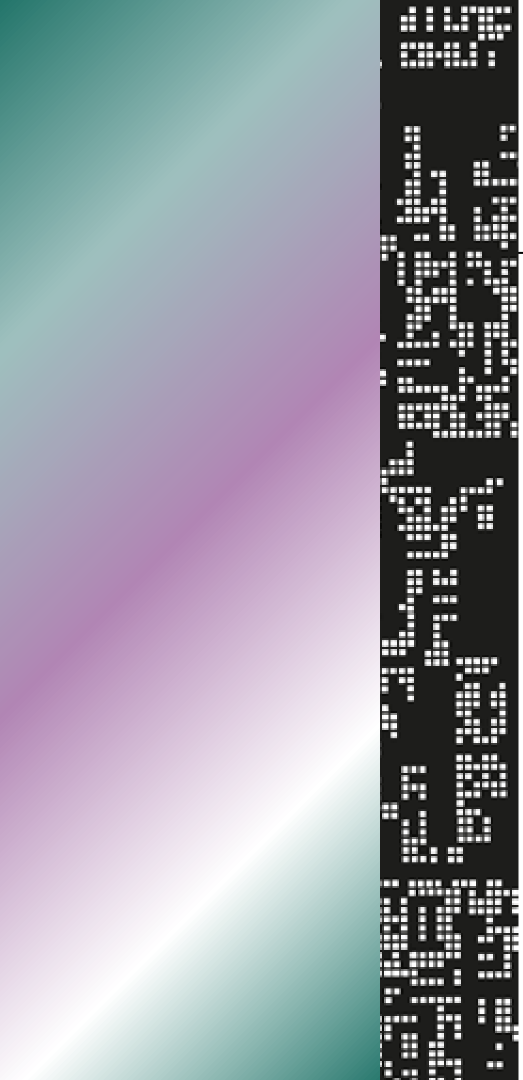
01 Problem Statement



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.

- 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.



Problem Statement

❑ The data which we have analysed 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: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample. (in our analysis, it is mentioned as target =1)
- All other cases: All other cases when the payment is paid on time. (in ouranalysis, it is mentioned as target =0)

02

Importing & Cleaning

Importing & Cleaning

❑ Importing

1. Importing all the requited libraries
2. Importing the warnings
3. Setting rows and columns to max for full view
4. Reading the dataset application_data as app_df
5. Checking the datatypes and number of values present in each column

Importing & Cleaning

❑ Data Cleaning

1. Checking the amount of missing values in each column
2. Converting the missing values into percentages
3. Removing the columns where more than 50% of the data is missing and sorting them in descending order for better visualisation

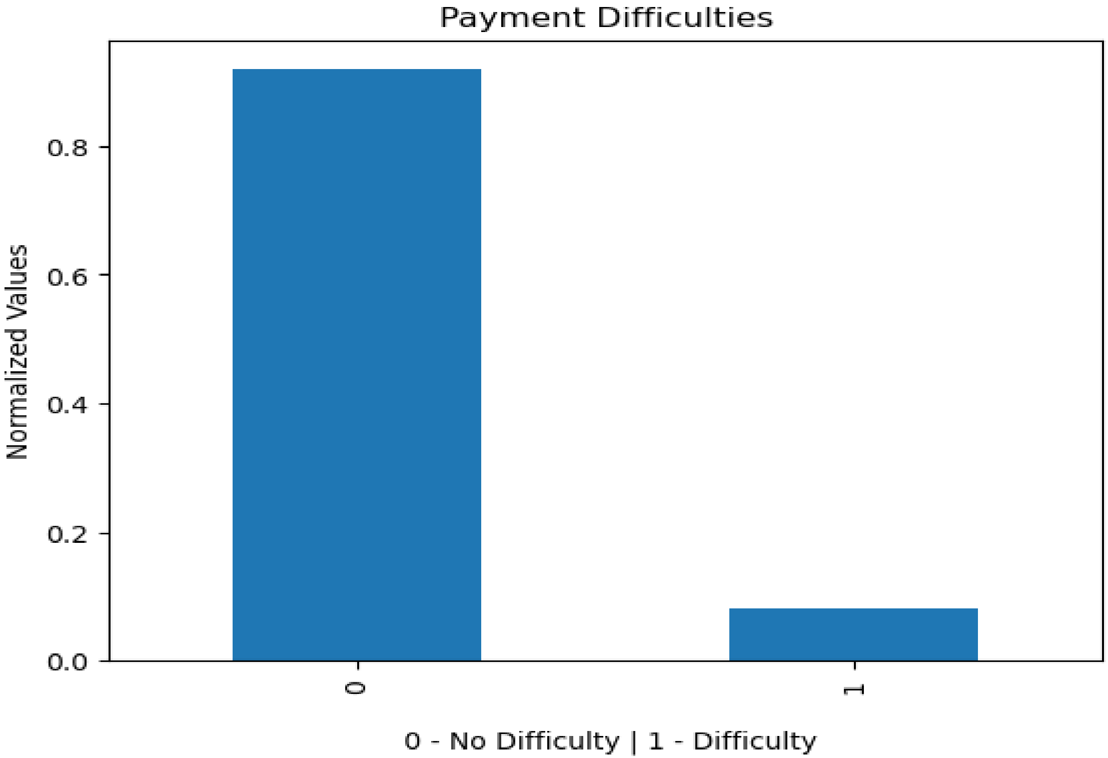
03 Approach

Approach to Data Cleaning

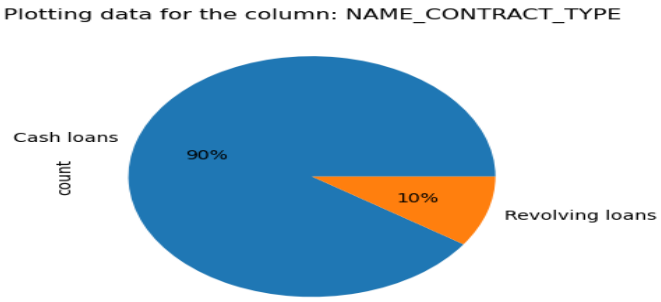
1. Looking at the head of all the columns and understanding them
2. Checking for null values in columns with the help of histogram to visualise it's distribution and replacing them with either mean, median or mode.
3. After visualisation of application_data, we import previous_data and merge them together with common column "SK_ID_CURR" as it had duplicated values

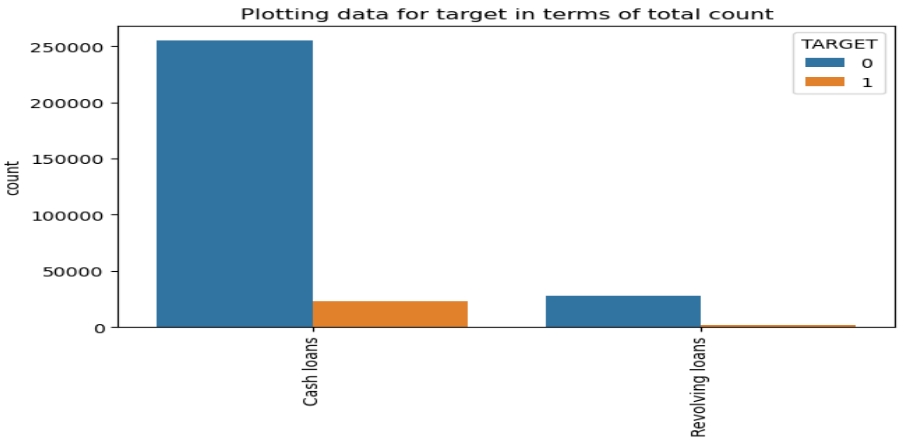
04

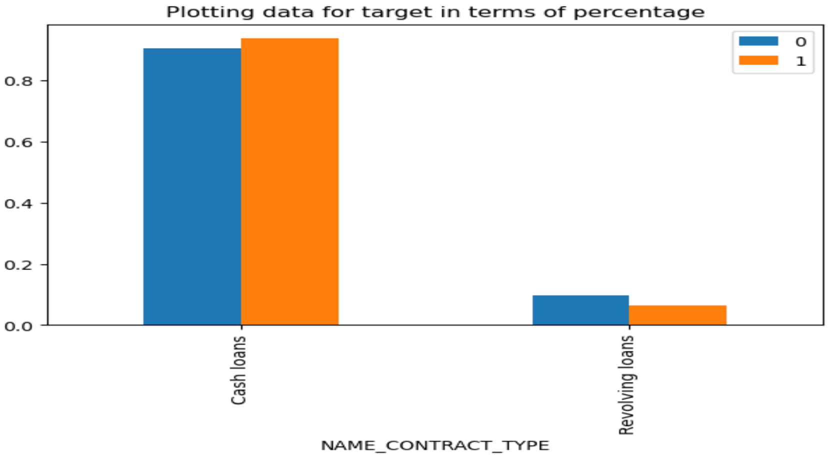
Data visualisation & Insights



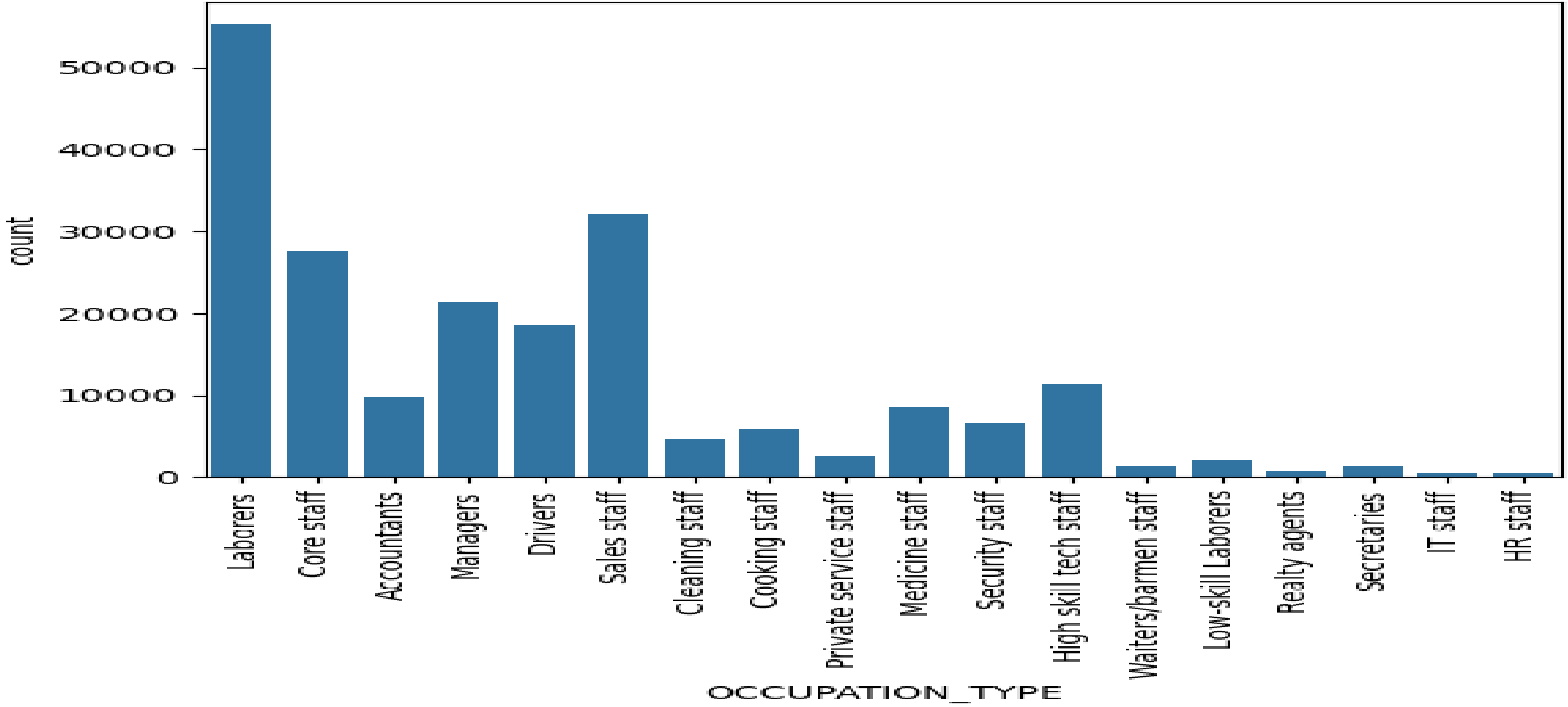
People facing difficulty in payment is around 90%



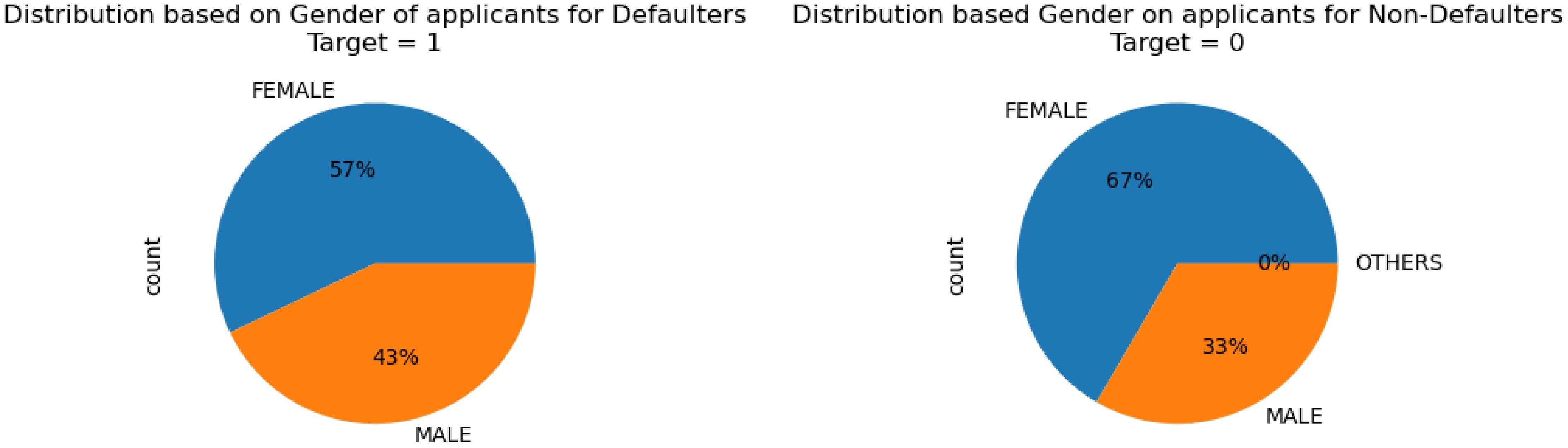




Total 90% of loans are in the form of cash loans and rest are revolving. People face more difficulty in case of cash loans compared to revolving loans.

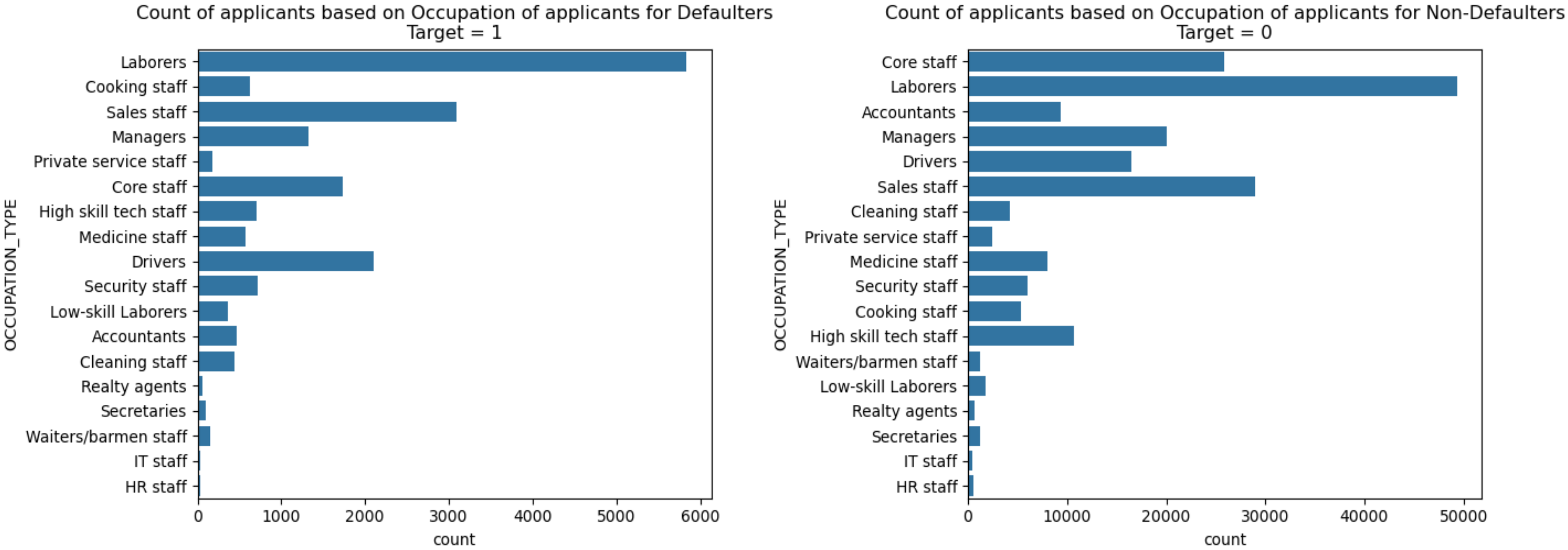


Column: OCCUPATION_TYPE
OCCUPATION_TYPE column has 31% missing data, which is also a large number. So, it would be appropriate to remove this column, but the datain this column seems to look important. So i would not remove the missing data.



Observation:

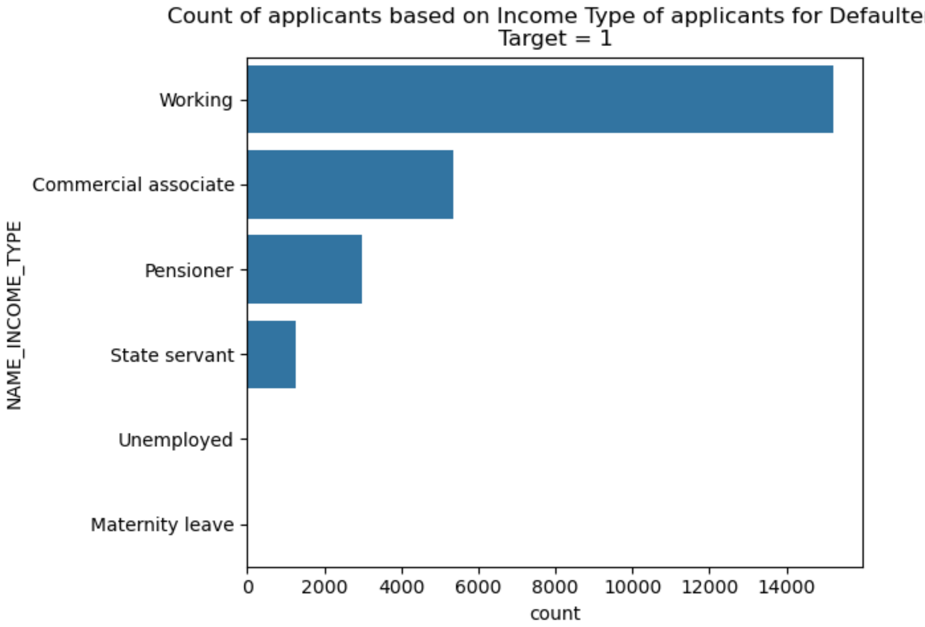
1. Close to 60% of the applicants are Females in Defaulters
2. Close to 70% of the applicants are Females in Non-Defaulters

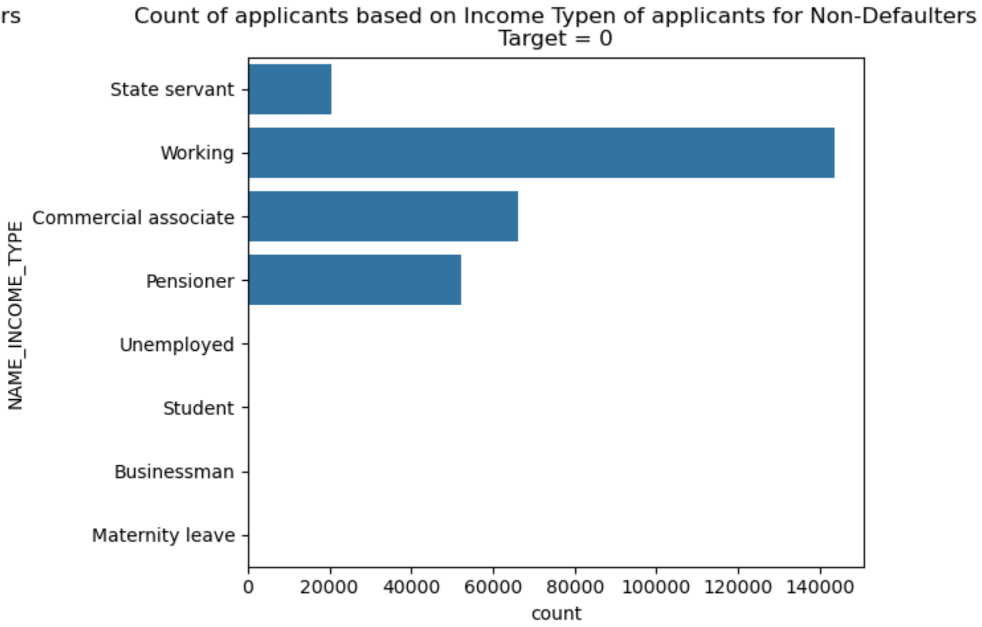


Count of applicants based on Occupation of applicants for Defaulters
Target = 1

Count of applicants based on Occupation of applicants for Non-Defaulters
Target = 0

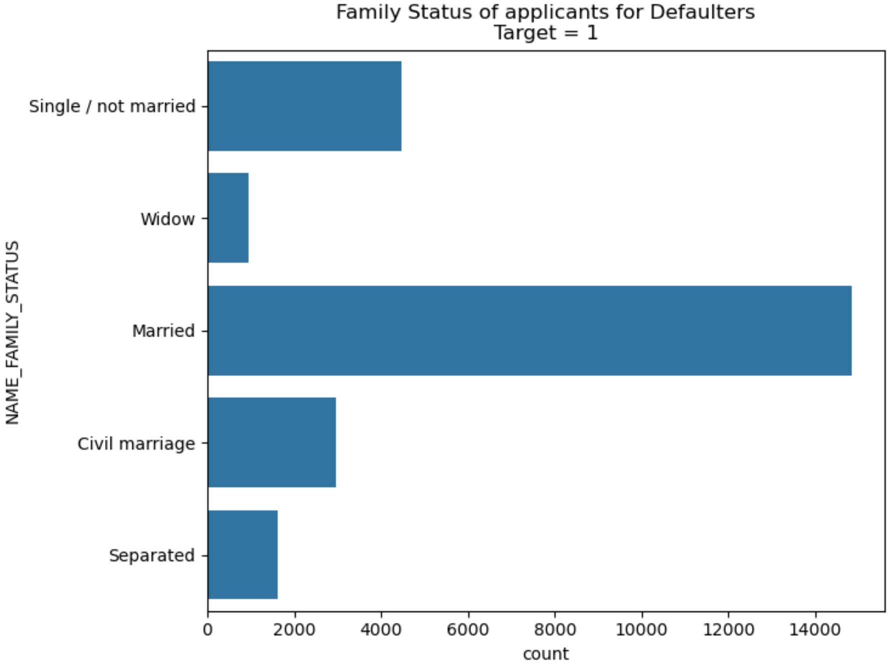
Occupation of most of the applicants is Labourer and next to them are Sales staff

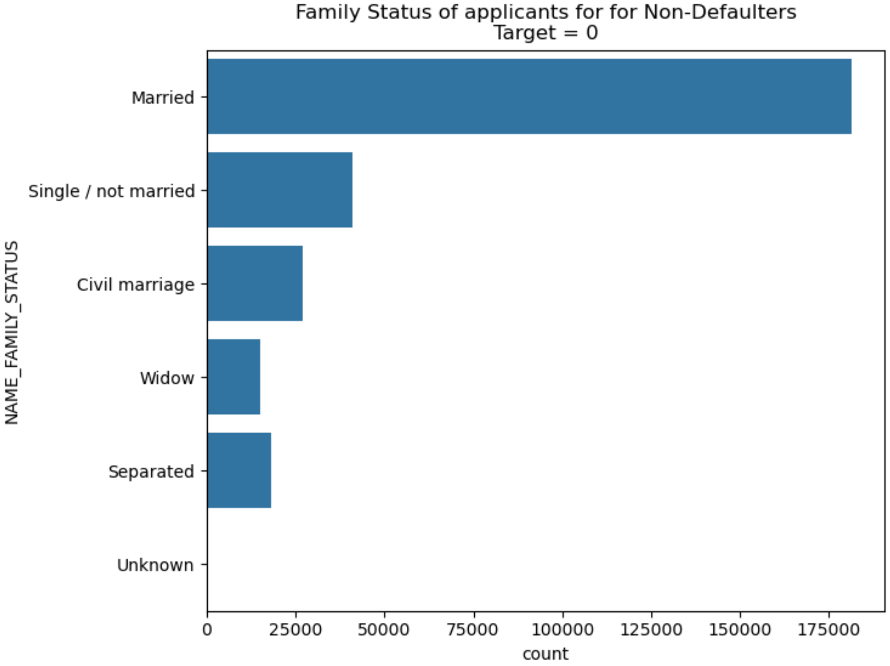

Count of applicants based on Income Type of applicants for Defaulters
Target = 1


Count of applicants based on Income Typen of applicants for Non-Defaulters
Target = 0

Observations:

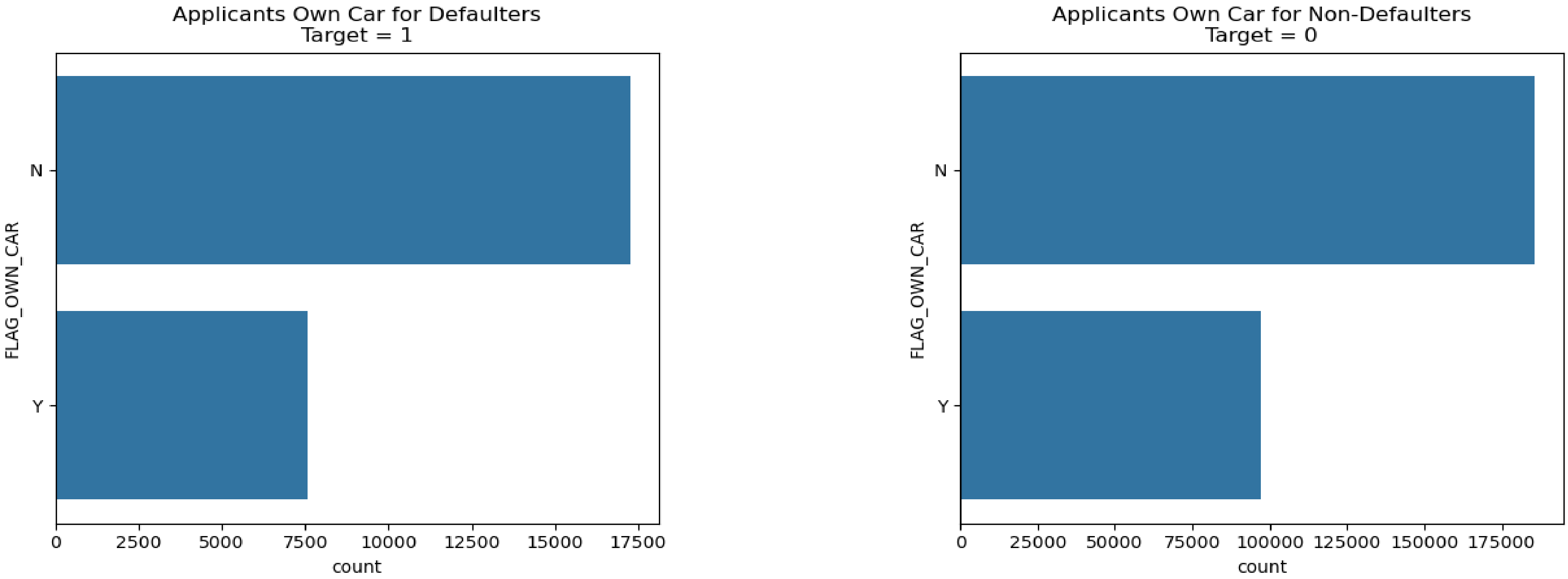
1. From the above graph, we can notice that the students are falling in non-defaulters. The reason could be they are not required to pay during their college tenure.
2. Most of the loans are distributed to working class people
3. Pensioners are also good in number for applying loans and mostly they are non-defaulters as we can see in plots





Observations:

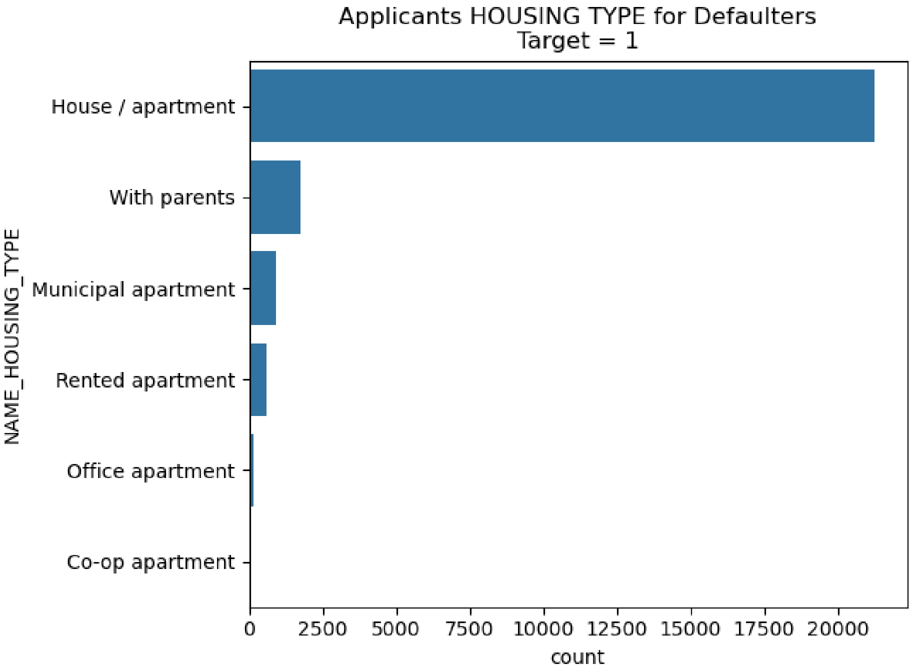
1. Most of the applicants are married
2. Next to married, 2nd highest applicants are Single/Non-married
3. Most of the applicants are married in both defaulter and non-defaulter categories

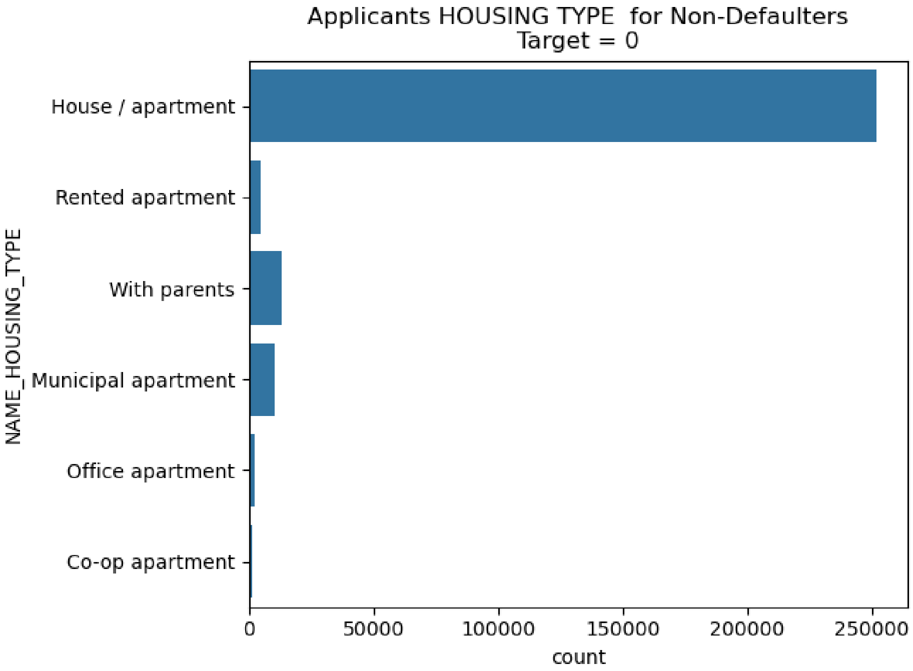


Observations:

1. Most of the applicants don't own a car
2. It can be seen that people with cars contribute almost same to the non-defaulters and defaulters.

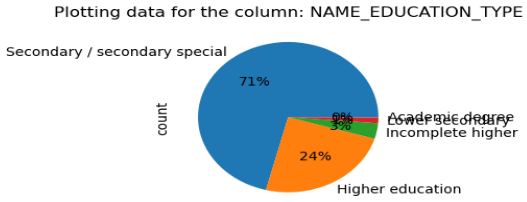
We can conclude that the number of default of people having car is low compared to people who don't.

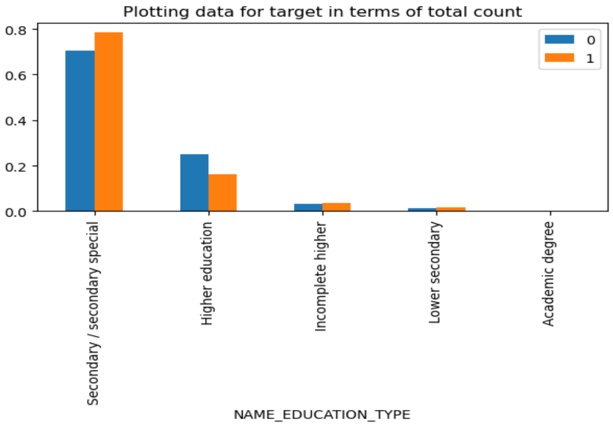


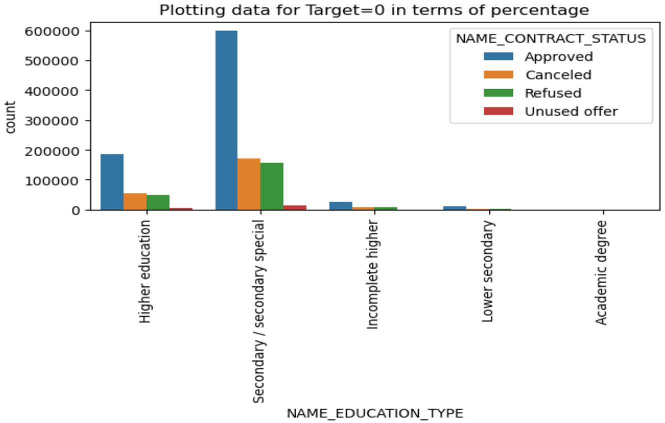


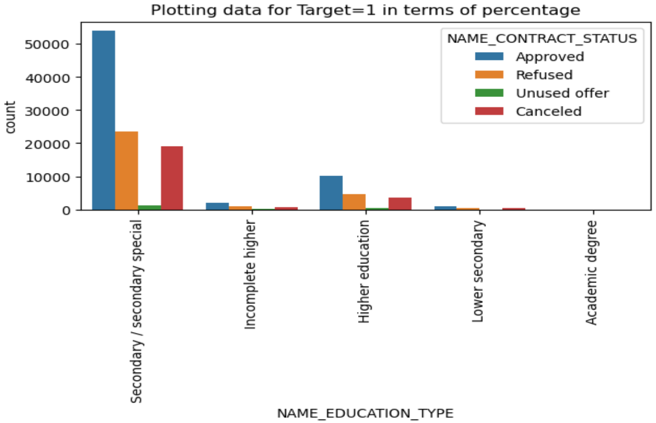
Observations:

1. Most of the applicants who own a house are non-defaulters and who don't own a house are defaulters. Its a very interesting trend here. We can say that applicants who own a house are tend to be non-defaulters


Plotting data for the column: NAME_EDUCATION_TYPE


Plotting data for target in terms of total count


Plotting data for Target=0 in terms of percentage

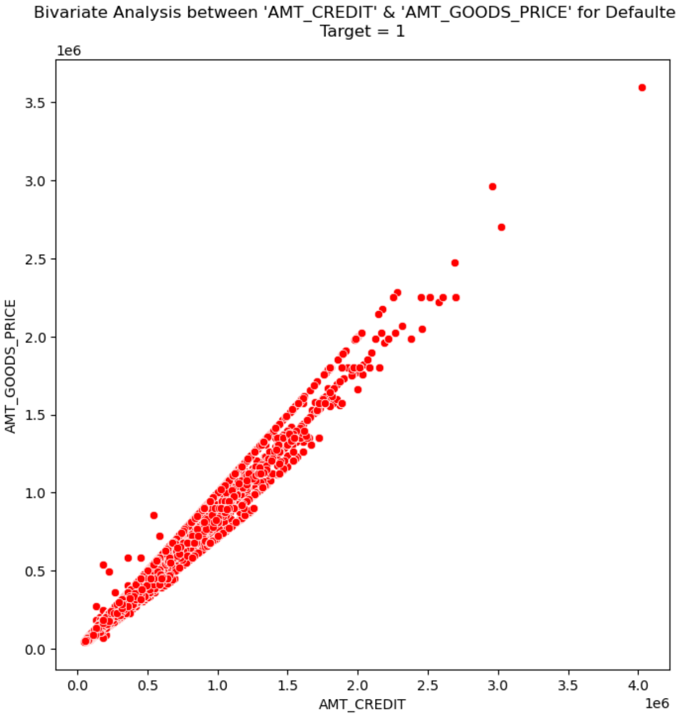

Plotting data for Target=1 in terms of percentage

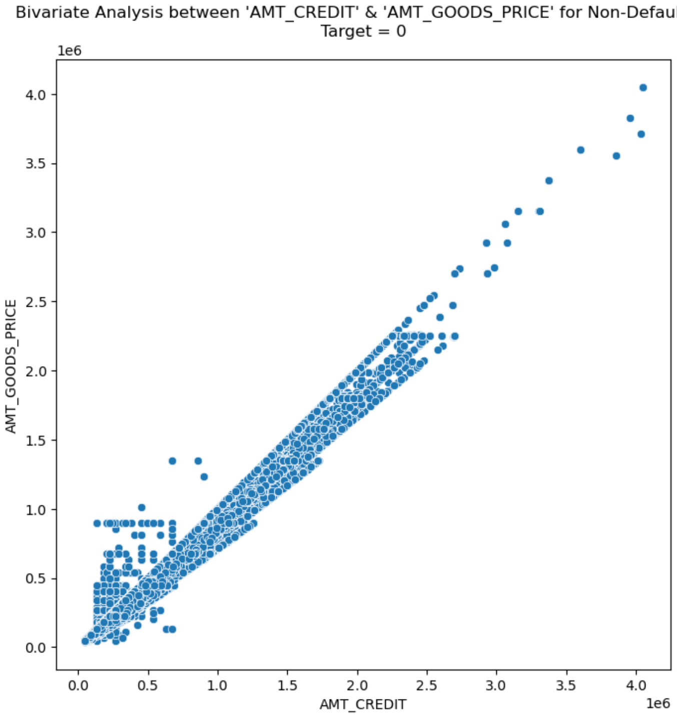
Observations:

1. Most of the applicants have completed Secondary Education in both defaulter and non-defaulter categories

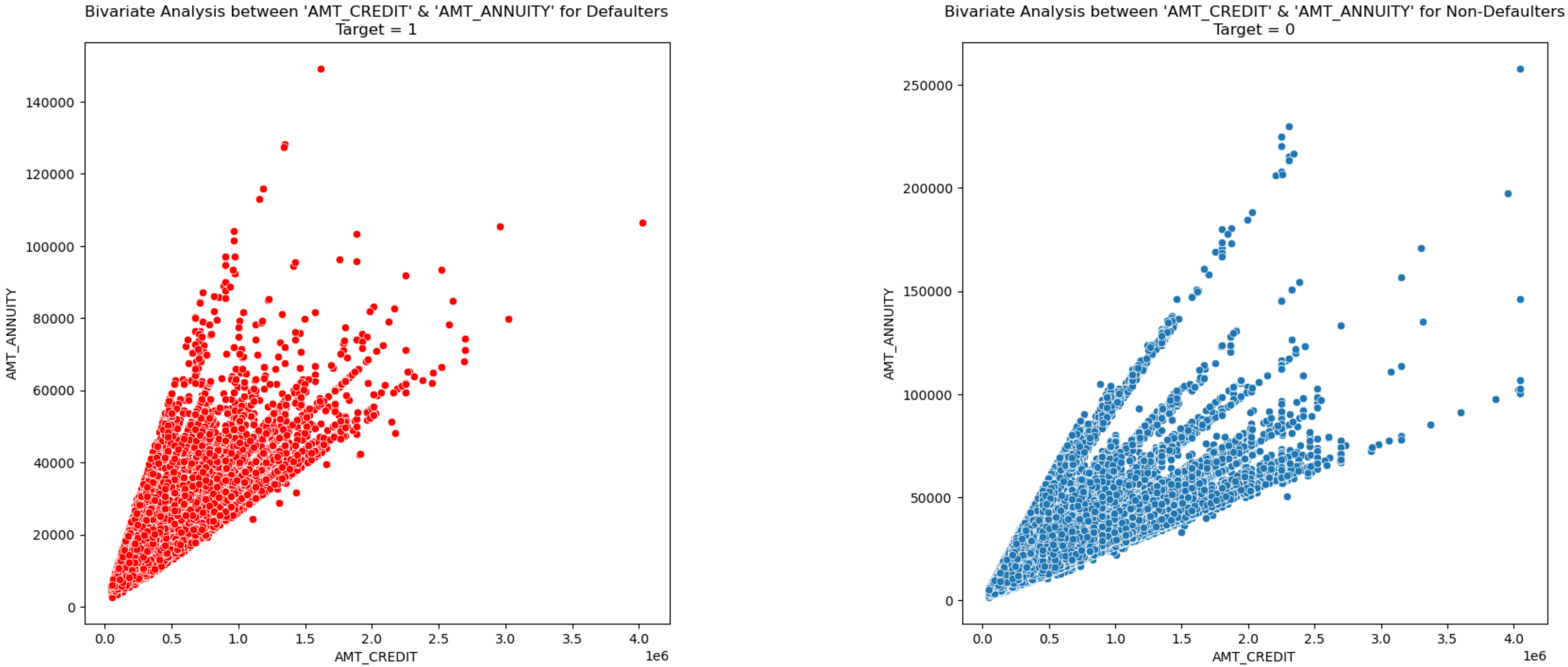
2. Many of the applicants have completed Higher Education in both defaulter and non-defaulter categories

3. Academic degree holders are almost neglible in number in both defaulter and non-defaulter categories





Amount Credit and Amount of Good price are showing same trend and which is mostly true as credit amount may be same or less than goods price



Amount_Credit and Amount_Annuity both are showing similar trend.

Key Insights

1. The ratio of non-defaulters to defaulters seems to be 1:10
2. Women tend to take loans more than the men, especially women from education type Secondary/secondary special.
3. The Business entity type 1 is tending to take more loans that other business categories. The second most common category to take loan is self employed.
4. Business type tend to pay more annuity amount and prefer revolving loans.
5. People who has rented apartment tend to take more loans among the other housing type category.
6. Education type Secondary/secondary special tend to have more income and also take more loans.
7. Married and single people tend to take more loans.
8. People who already own a property tend to take more loans. Maybe the banks are inclined to provide them loans without much hassles since they have a property to count on if they become defaulters.
9. Business people tend to take only revolving loans and they are good enough to pay them back too in general.
10. People generally tend to take loans for electronics and banks also approve them.

Thank You