

CREDIT EDA ASSIGNMENT

DR SAMIR YASHWANT KSHIRSAGAR

CREDIT EDA ASSIGNMENT

- PROBLEM STATEMENT – I

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history.
- Some consumers use it to their advantage by becoming a defaulter
- **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 the sample
- When the company receives a **loan application**, the company has to decide for loan approval based on the applicant's profile.
- Two types of risks are associated with the bank's decision:
 - I. If the applicant is likely to repay the loan – **Non-defaulters** (then not approving the loan results in a loss of business to the company)
 - II. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, - **Defaulters** (then approving the loan may lead to a financial loss for the company)

CREDIT EDA ASSIGNMENT

- Data given below contains the information about the loan application
- contains two types of scenarios:
 - A. **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
 - B. **All other cases:** All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

- Approved: The Company has approved loan Application
- Cancelled: The client cancelled the application sometime during approval
- Refused: The company had rejected the loan
- Unused offer: Loan has been cancelled by the client but at different stages of the process.

CREDIT EDA ASSIGNMENT

- Business Objectives
 - The 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
 - To ensure that the consumers capable of repaying the loan are not rejected.
 - Aim of the study is to use [EDA to analyse the patterns present in the data](#) and to ensure that the applicants capable of repaying the loan are not rejected.
 - Use EDA [to understand how consumer attributes](#) and [loan attributes](#) influence the tendency to default.

CREDIT EDA ASSIGNMENT

- PROBLEM STATEMENT - II

- Approach of the analysis in a presentation.
- We will analyze three data set as below:
 - I. *'application_data.csv'* that contains all the information of the client at the time of application
 - II. *'previous_application.csv'* contains information about the client's previous loan data. It contains the data on whether the previous application had been **Approved, Cancelled, Refused or Unused offer**
 - III. *'columns_description.csv'* is data dictionary which describes the meaning of the variables.

CREDIT EDA ASSIGNMENT

Assumptions:

- Let INP0 = application dataset
- Let INP2 = previous application dataset
- Let INP_All = Merging of the two database Application and previous application database.

CREDIT EDA ASSIGNMENT

- APPROACH AND METHODOLOGY

A. Steps for analysis

- Load and clean the data using pandas libraries
- Use of visualization techniques to plot the graph and understand the trends
- Identify the variables and identify the null variable(>40 %) and dropping of null variables
- Identify variables that are defaulters as 'Target' and non-defaulters as 'Non-Target' variable
- Clean the data variables and analyze univariate and bivariate variables
- Merge the application dataset and previous application dataset
- Find the correlations and analyze top variables for conclusions

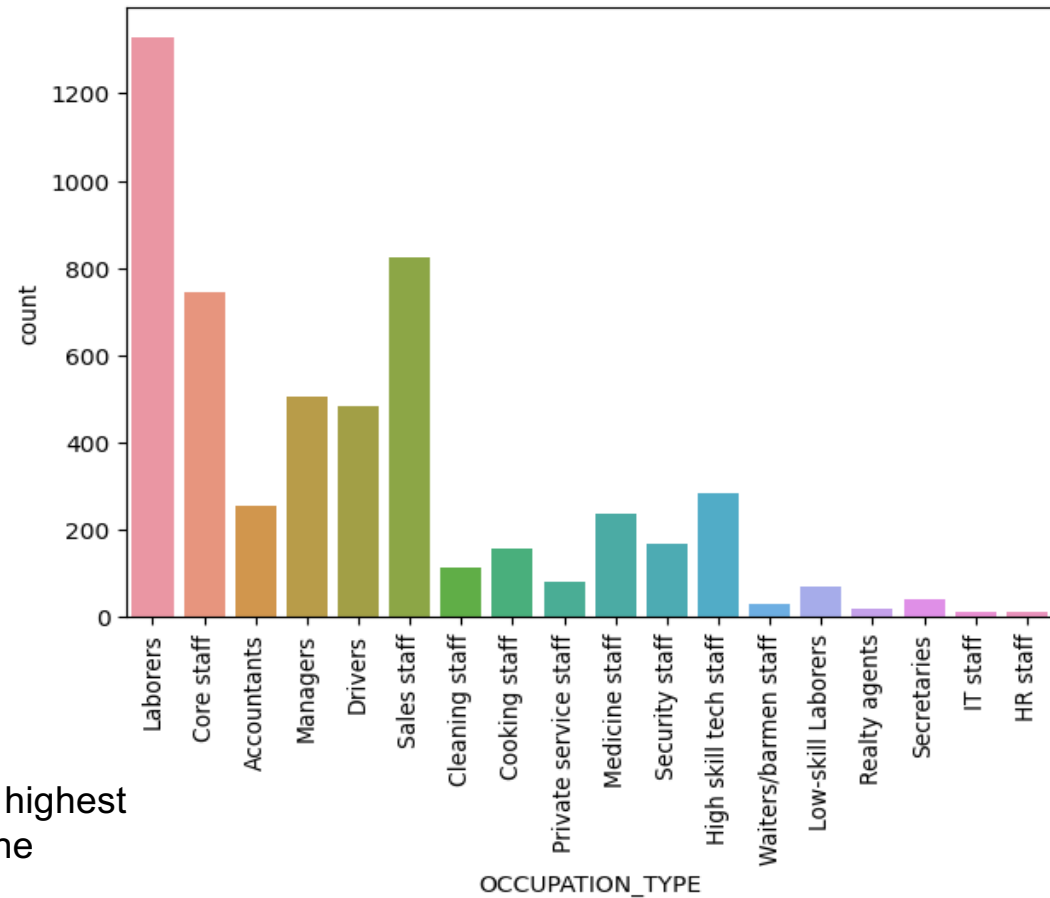
CREDIT EDA ASSIGNMENT

GRAPHS AND INSIGHTS :

- *Analysis for 'application_data.csv'*
- *Data Cleaning*
 - Identifying and Treating null values
 - checking for columns having greater than 40% null value
 - Dropping all columns having more than 40% null values
- *Analysis of variables*
- Distribution of OCCUPATION_TYPE

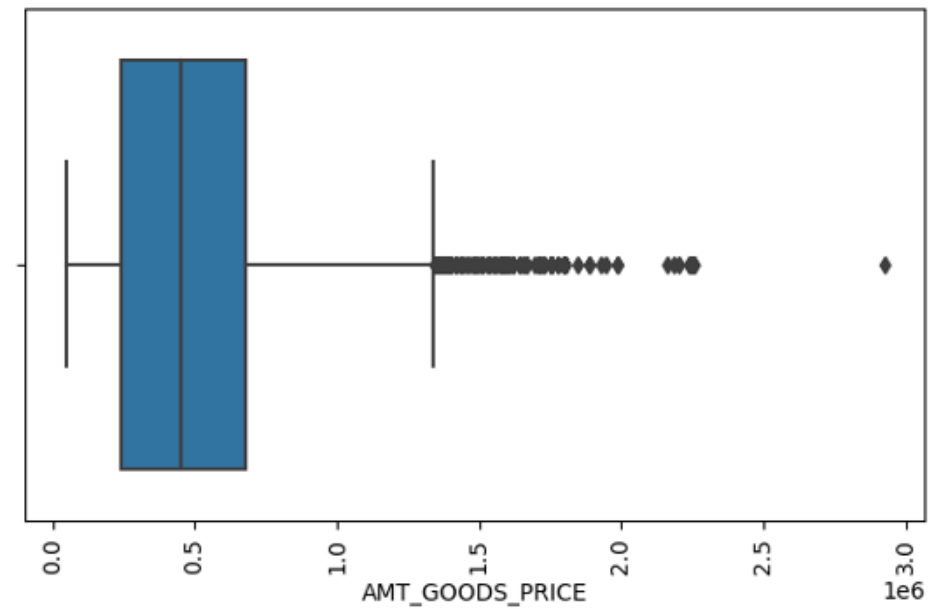
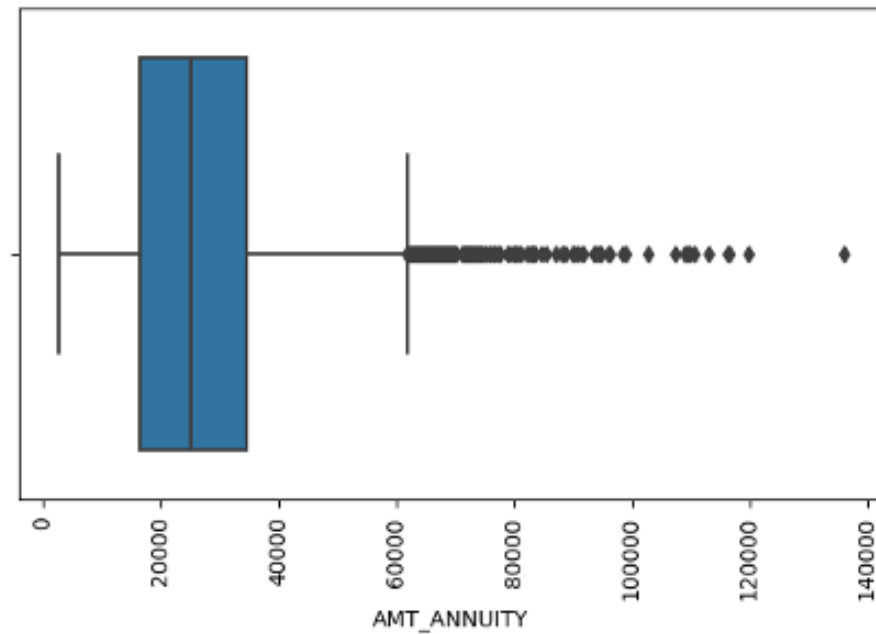
Analysis:

Looking at the plot, Laborers, Core staff has the highest number of loan applicants, where HR staff has the lowest

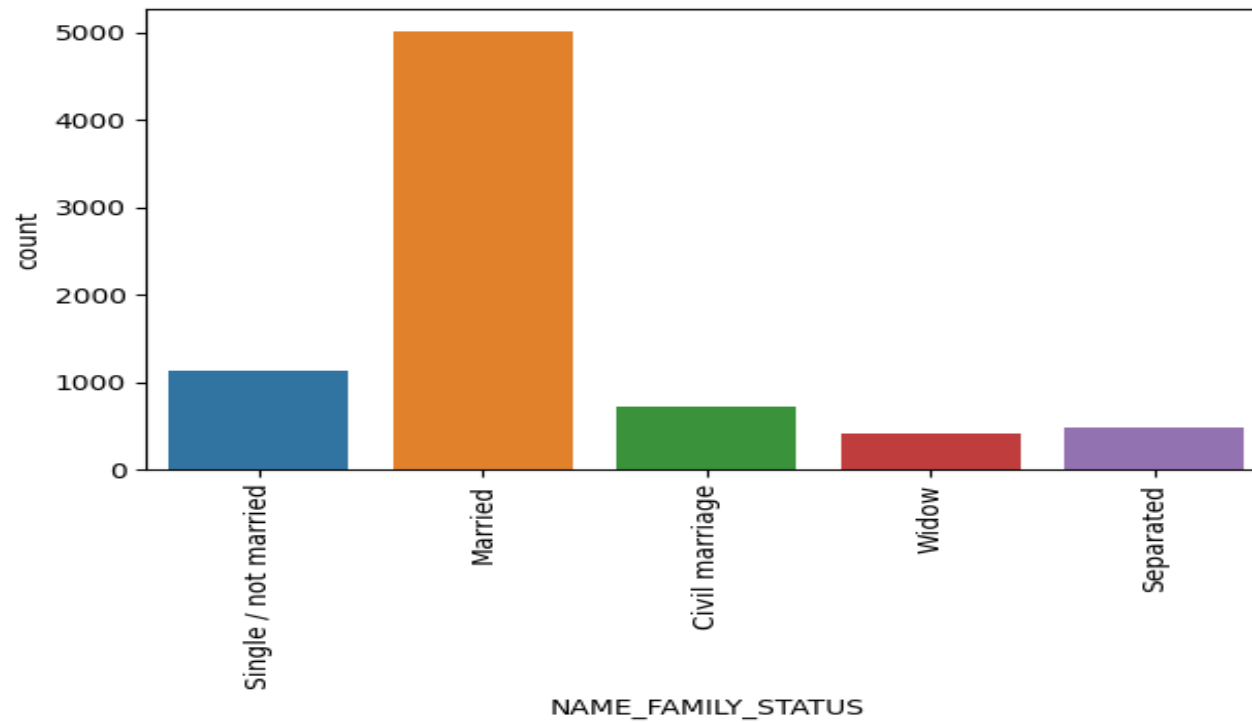


CREDIT EDA ASSIGNMENT

- Distribution of Variables:
- Distribution of AMT_GOODS_PRICE and AMT_ANNUIITY



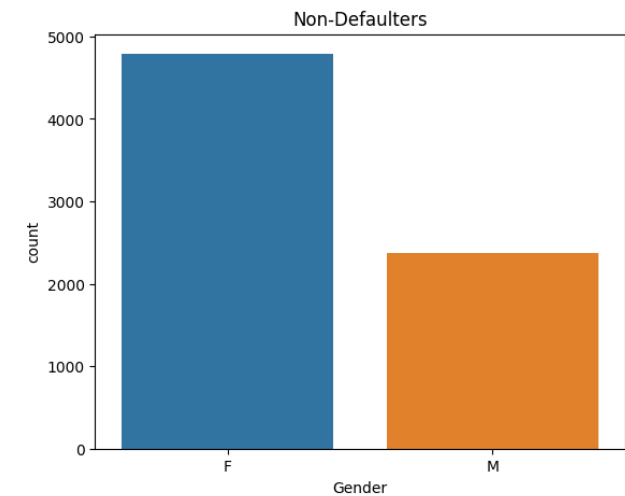
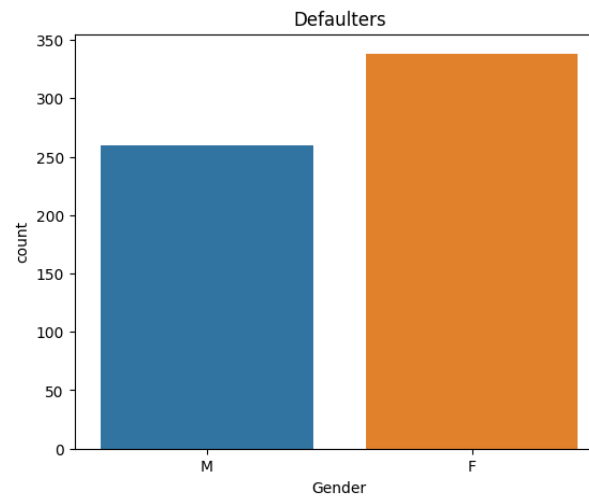
CREDIT EDA ASSIGNMENT



- Distribution of Family status

CREDIT EDA ASSIGNMENT

- Univariate Analysis (Countplot) - categorical columns
- Dividing dataset in to categories:
 - Client with payment difficulties (Target=1)= Defaulters
 - All other cases (Target=0)=Non-Defaulters
- Plotting two plots for defaulters and non defaulters on basis of gender



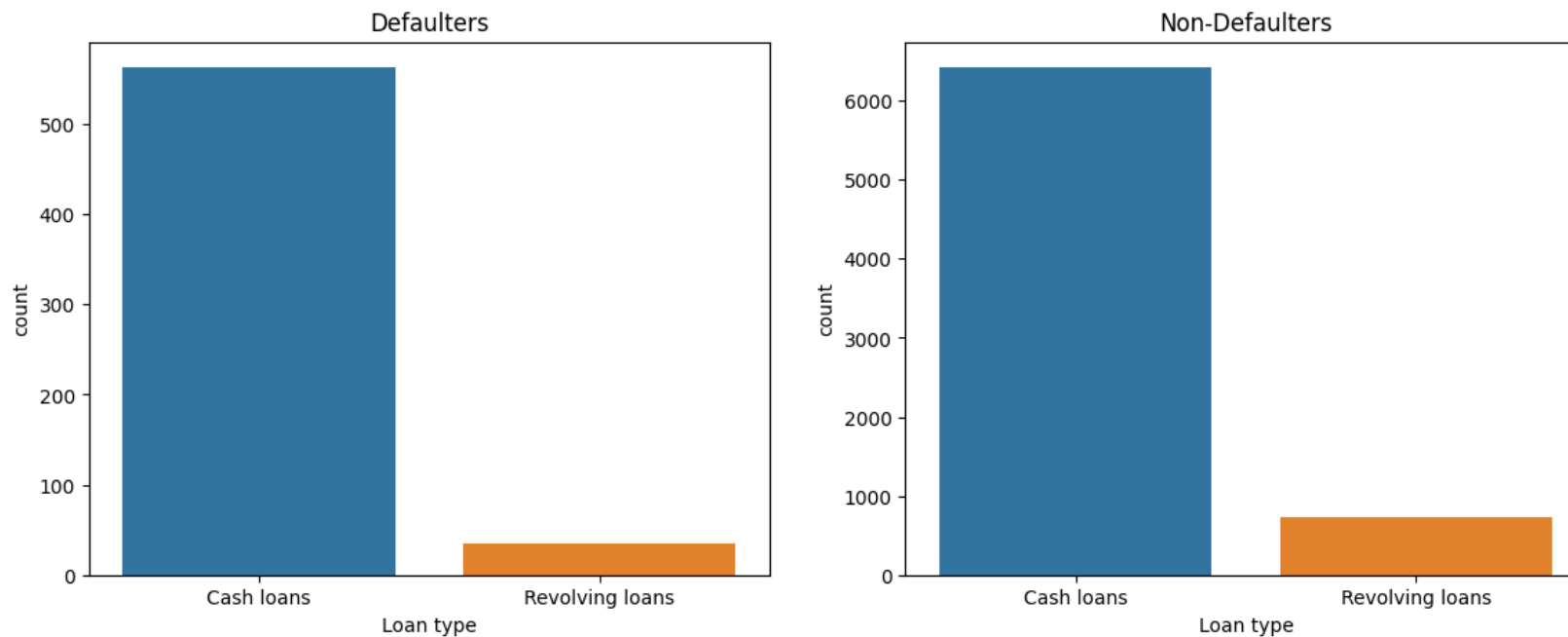
Analysis:

Defaulters - We can see that females are more in number of defaulters than male.

Non-defaulters - The same pattern continues for non-defaulters as well. The females are more in number here than male.

CREDIT EDA ASSIGNMENT

- #Defaulters and non-defaulters on the basis of Loan type



Analysis - We see in both the cases that Revolving loans are very less in number compared to Cash loans.

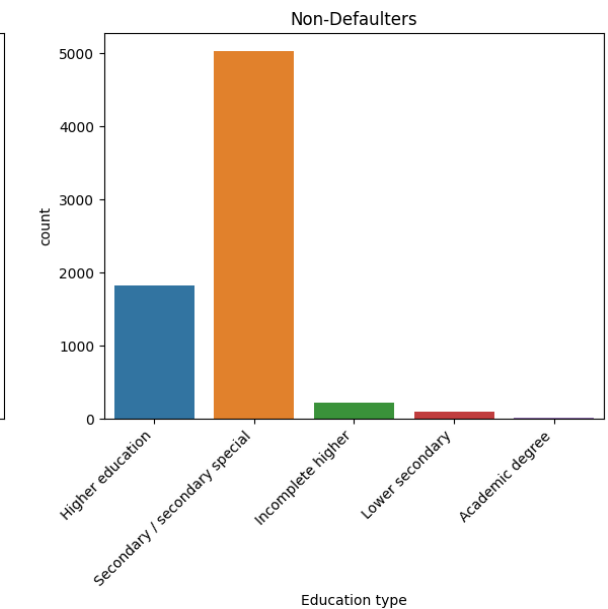
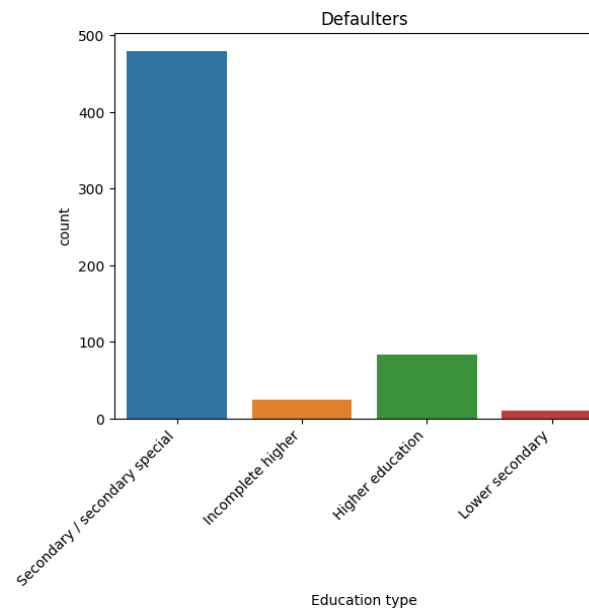
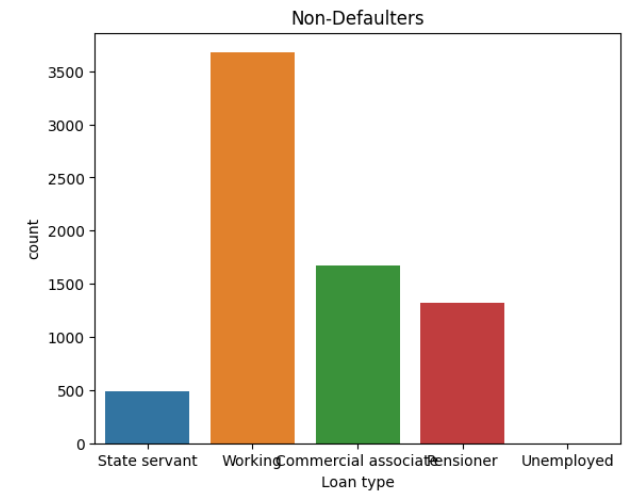
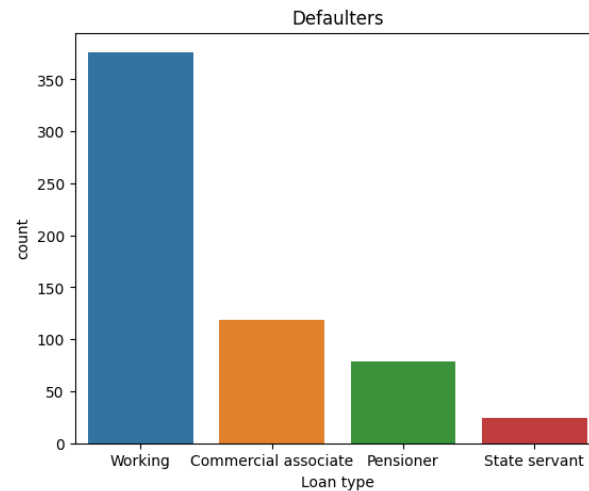
CREDIT EDA ASSIGNMENT

- #Defaulters and non-defaulters on the basis of income and Education type

Analysis:

There are more defaulters for secondary and secondary education type

However, more number of working and commercial associate and pensioner are non-defaulters



CREDIT EDA ASSIGNMENT

- Checking Data Imbalance

- Observation:

- We can see that there is data imbalance in below columns:-

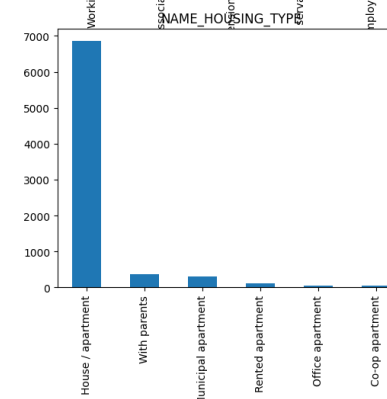
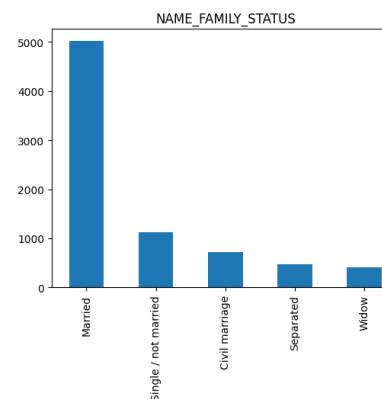
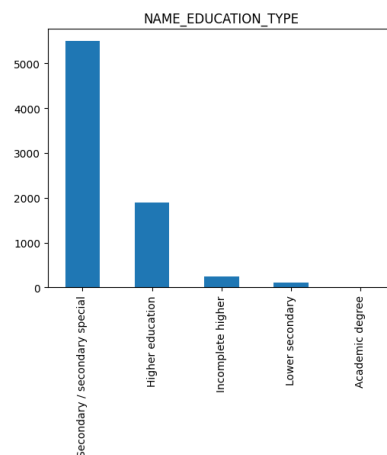
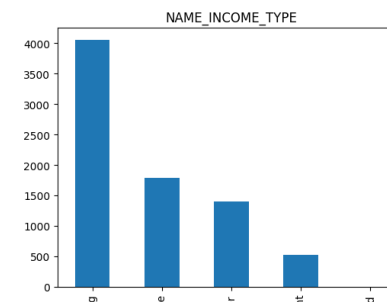
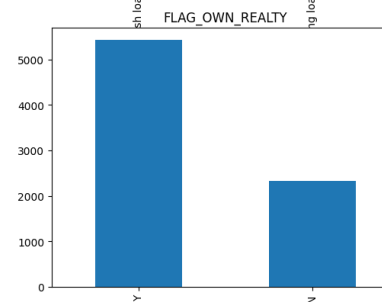
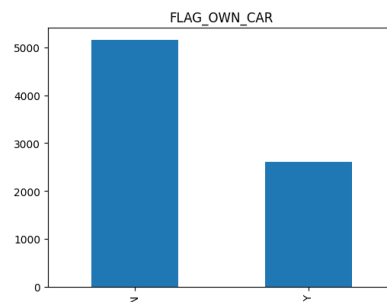
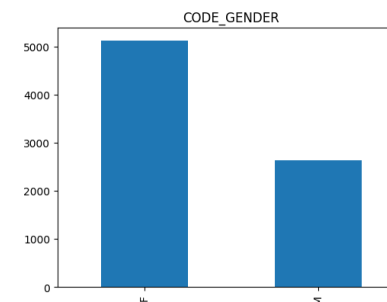
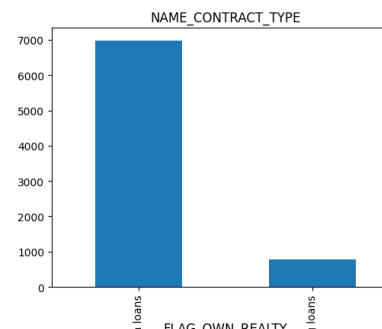
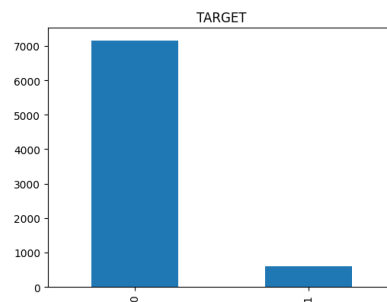
- TARGET - There are very few defaulters(1) compare to non defaulters(0)

- NAME_CONTRACT_TYPE - There are very few Revolving loans than Cash loans

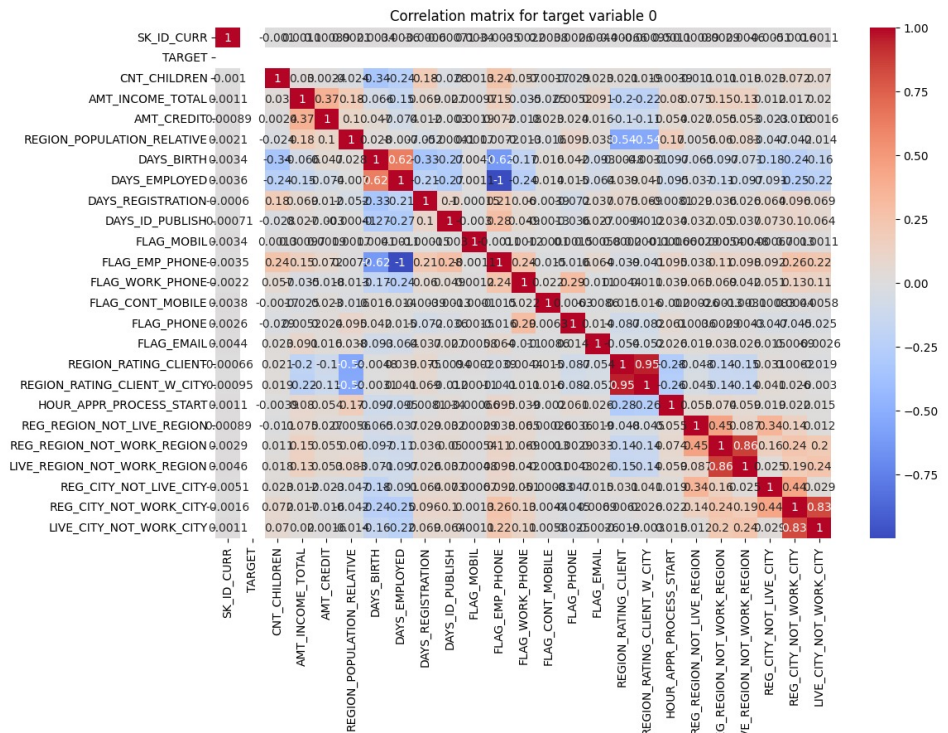
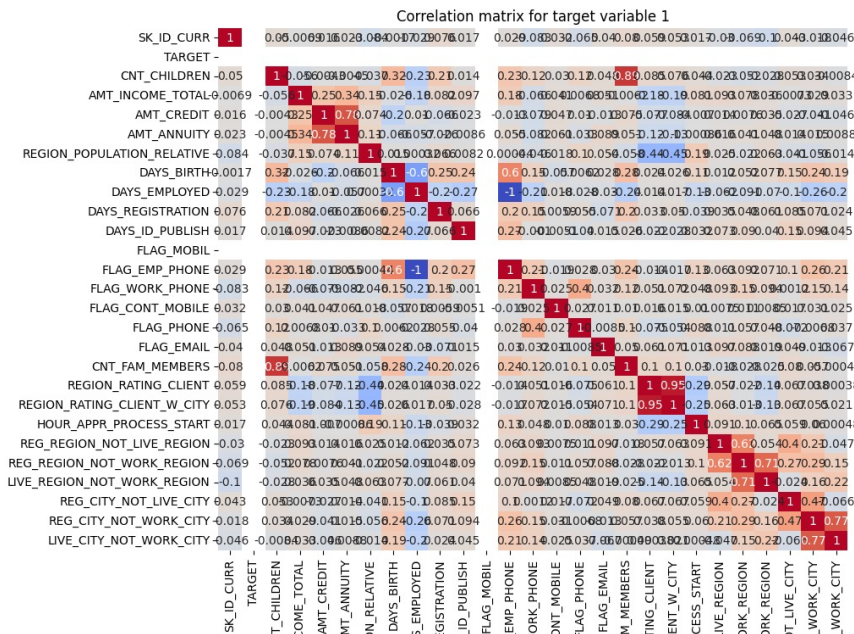
- NAME_EDUCATION_TYPE - Most of the loans applied by Secondary/Secondary special educated people

- NAME_FAMILY_STATUS - Most of the loans applied by Married people.

- NAME_HOUSING_TYPE - Most of the application came from Home/apartment owner



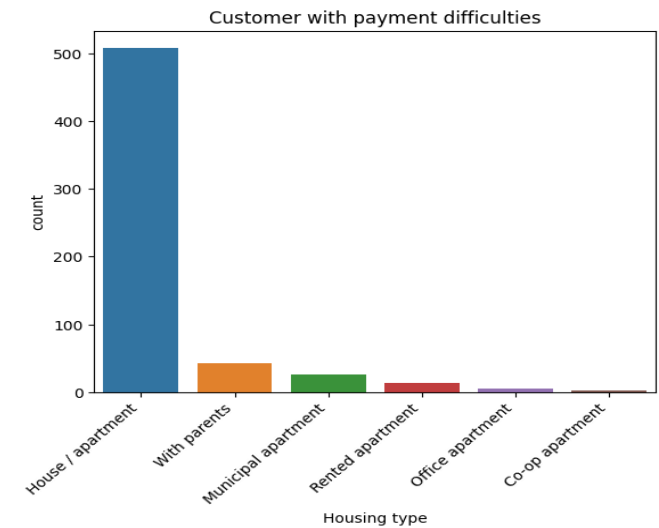
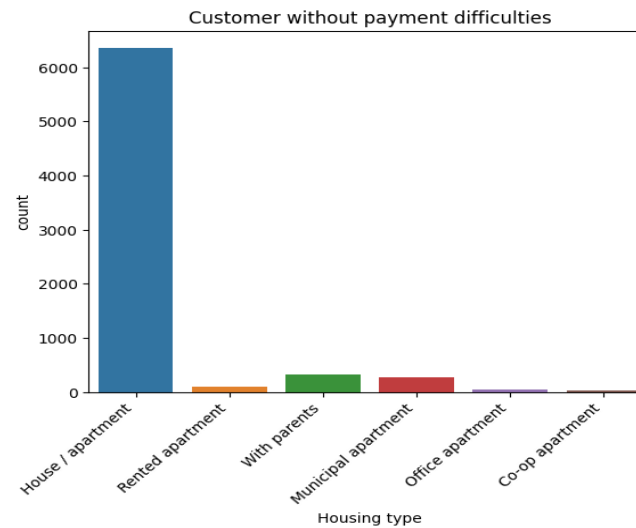
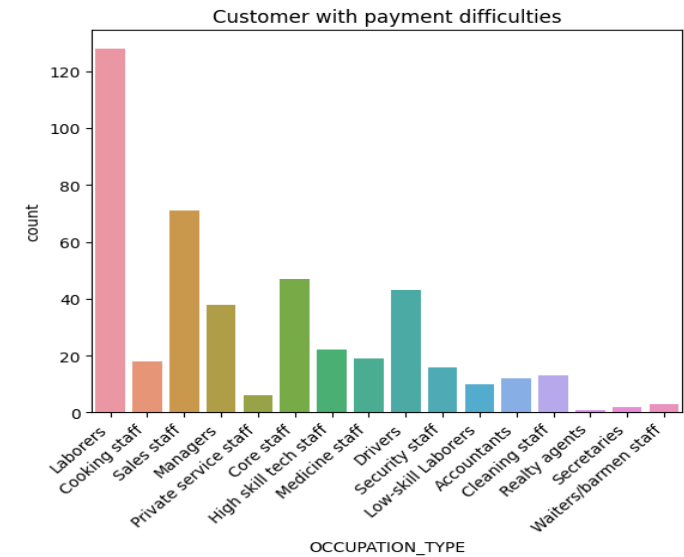
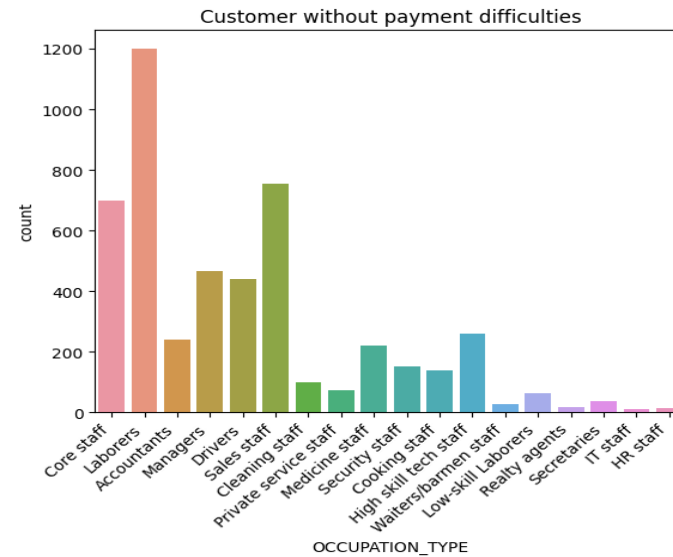
CREDIT EDA ASSIGNMENT



Correlation matrix between Target Variable 1 and Target variable 0 in application data

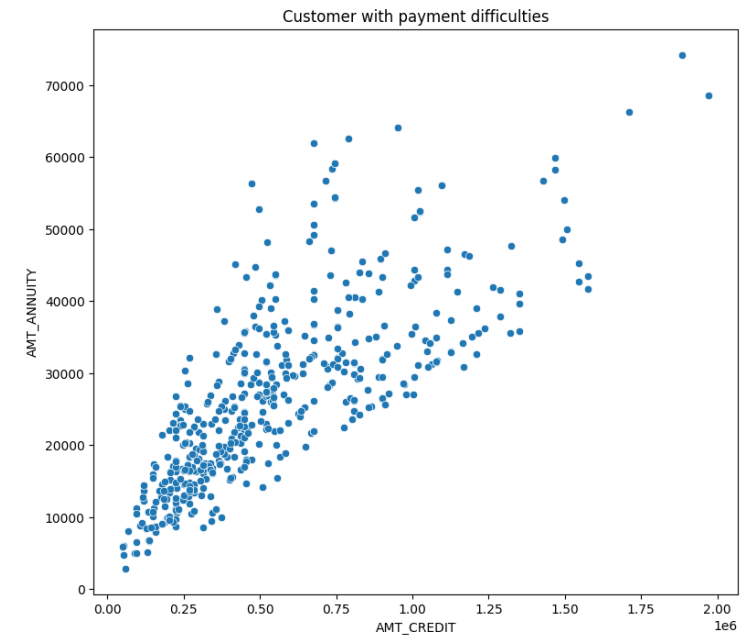
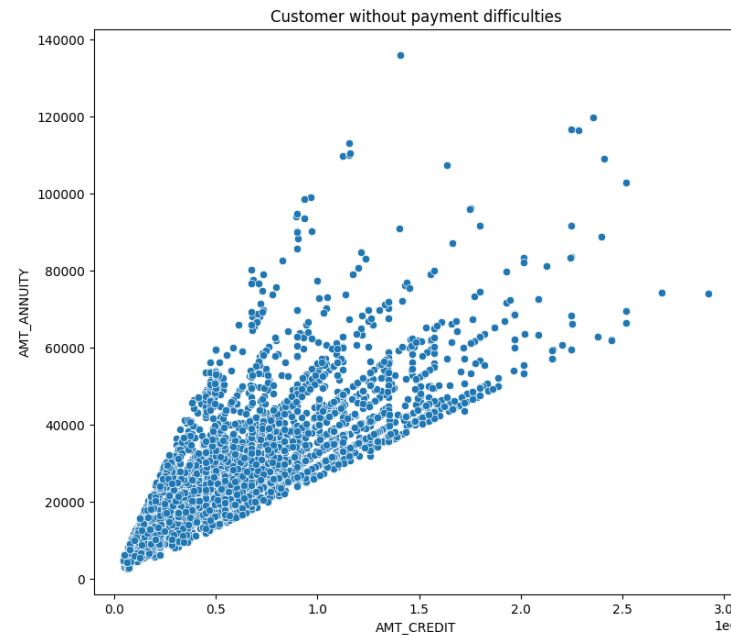
CREDIT EDA ASSIGNMENT

- Analysis for Customer without payment difficulties
- Customer without payment difficulties for Occupation type and housing type
- Analysis – Laborers and Sales Staff have more payment difficulties than other
- Most of the customers take loan for House /apartments than any other type



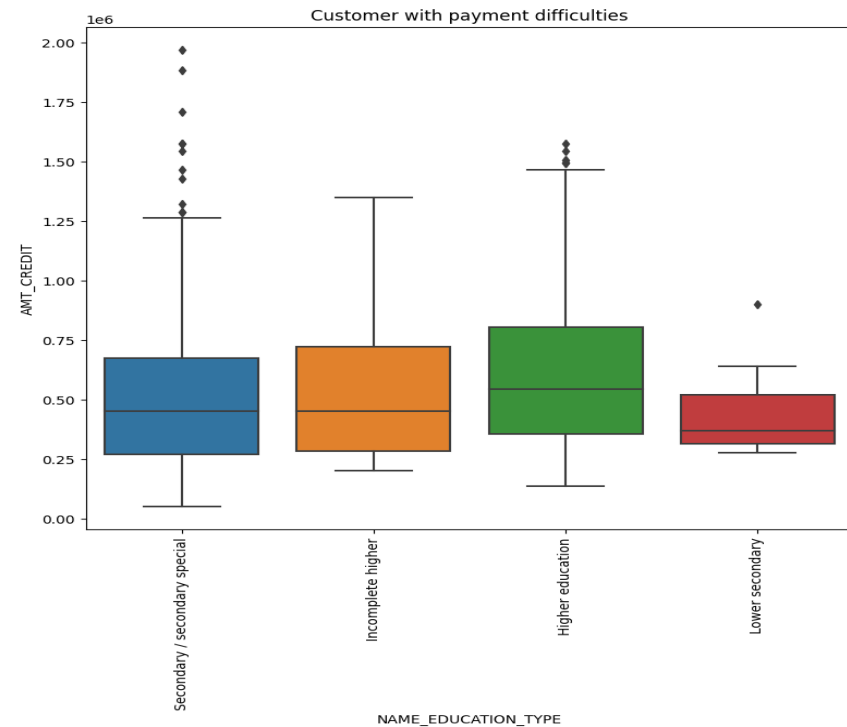
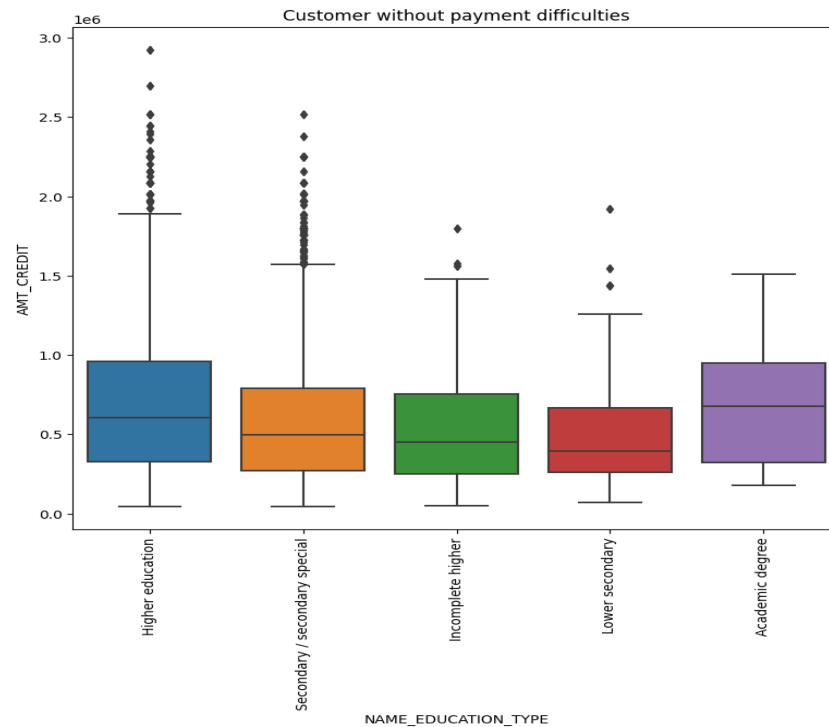
CREDIT EDA ASSIGNMENT

- Bivariate analysis:
- Between 'AMT_CREDIT' and 'AMT_ANNUITY'
- Analysis for Customer without payment difficulties
- Customer without payment difficulties



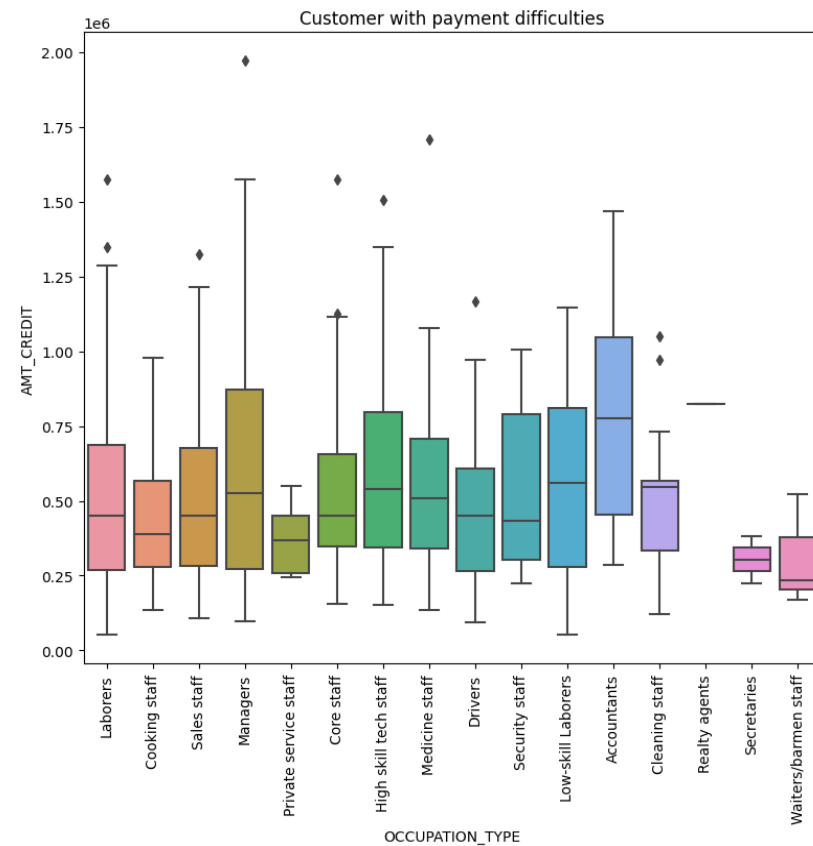
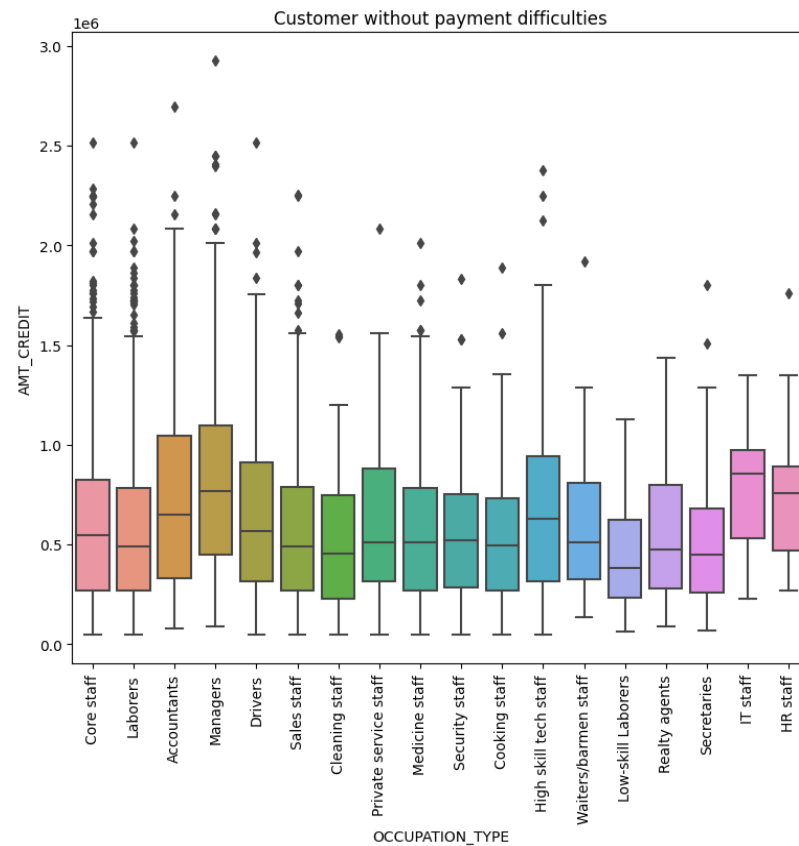
Analysis: More credit for the annuity is been taken by the people without payment difficulties

CREDIT EDA ASSIGNMENT



- Numerical- categorical bivariate analysis – AMT_CREDIT and NAME_EDUCATION_TYPE for customers with or without payment difficulties
- #Analysis - the number of outliers are more in consumers without payment difficulties for secondary and higher education, while more outliers are seen for secondary, secondary special education for payment with difficulties

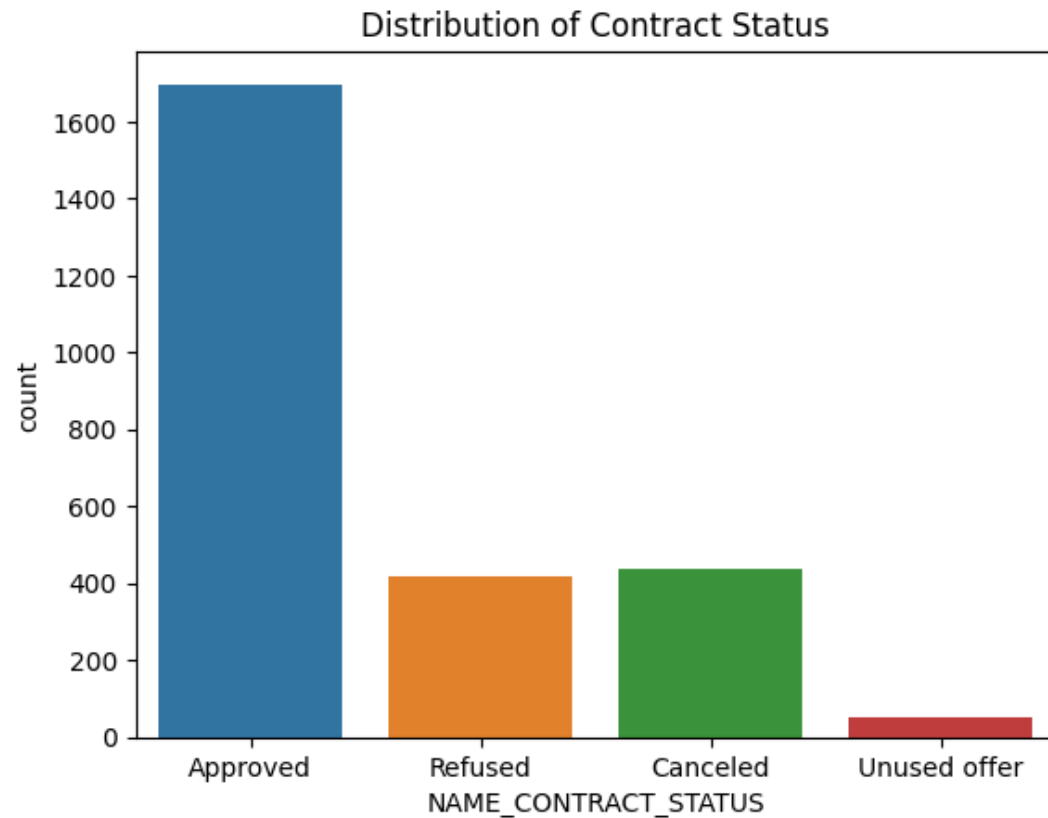
CREDIT EDA ASSIGNMENT



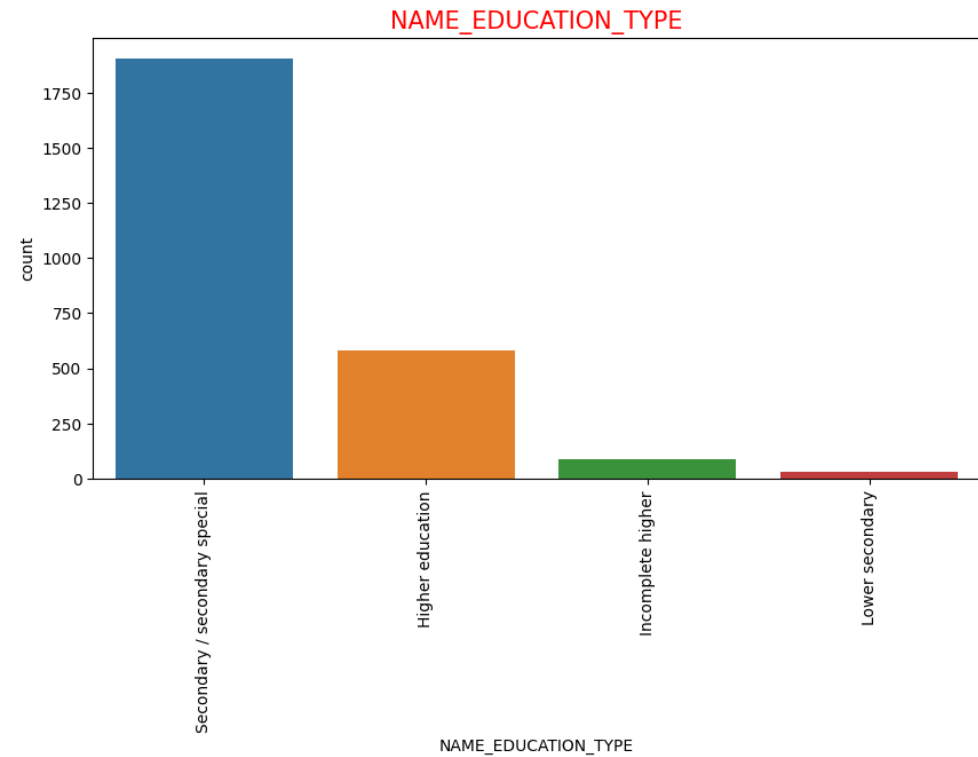
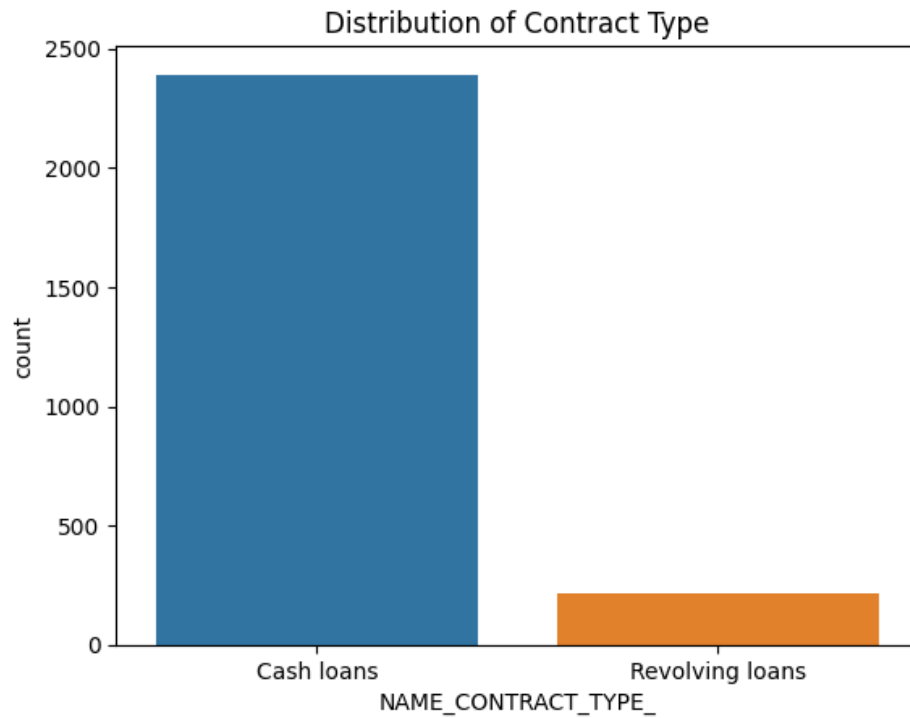
- Analysis for Customer without payment difficulties
- Customer without payment difficulties for Occupation type and AMT_CREDIT

CREDIT EDA ASSIGNMENT

- Analysis for "previous_application.csv"
- Distribution of CONTRACT_STATUS

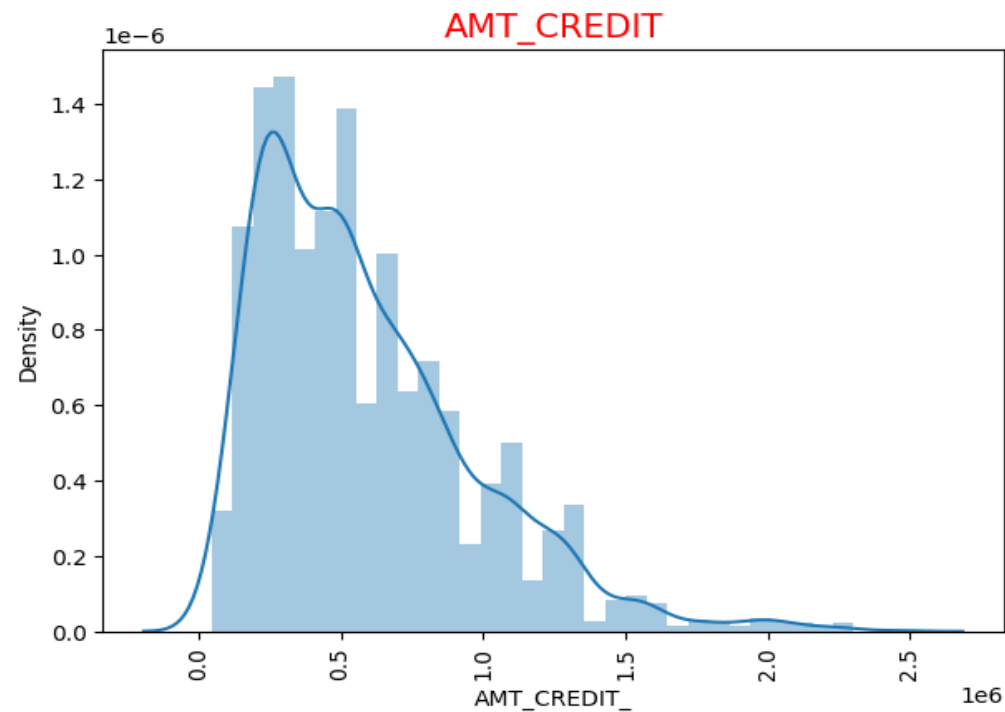


CREDIT EDA ASSIGNMENT



- Distribution of Contract Type and Education type

CREDIT EDA ASSIGNMENT



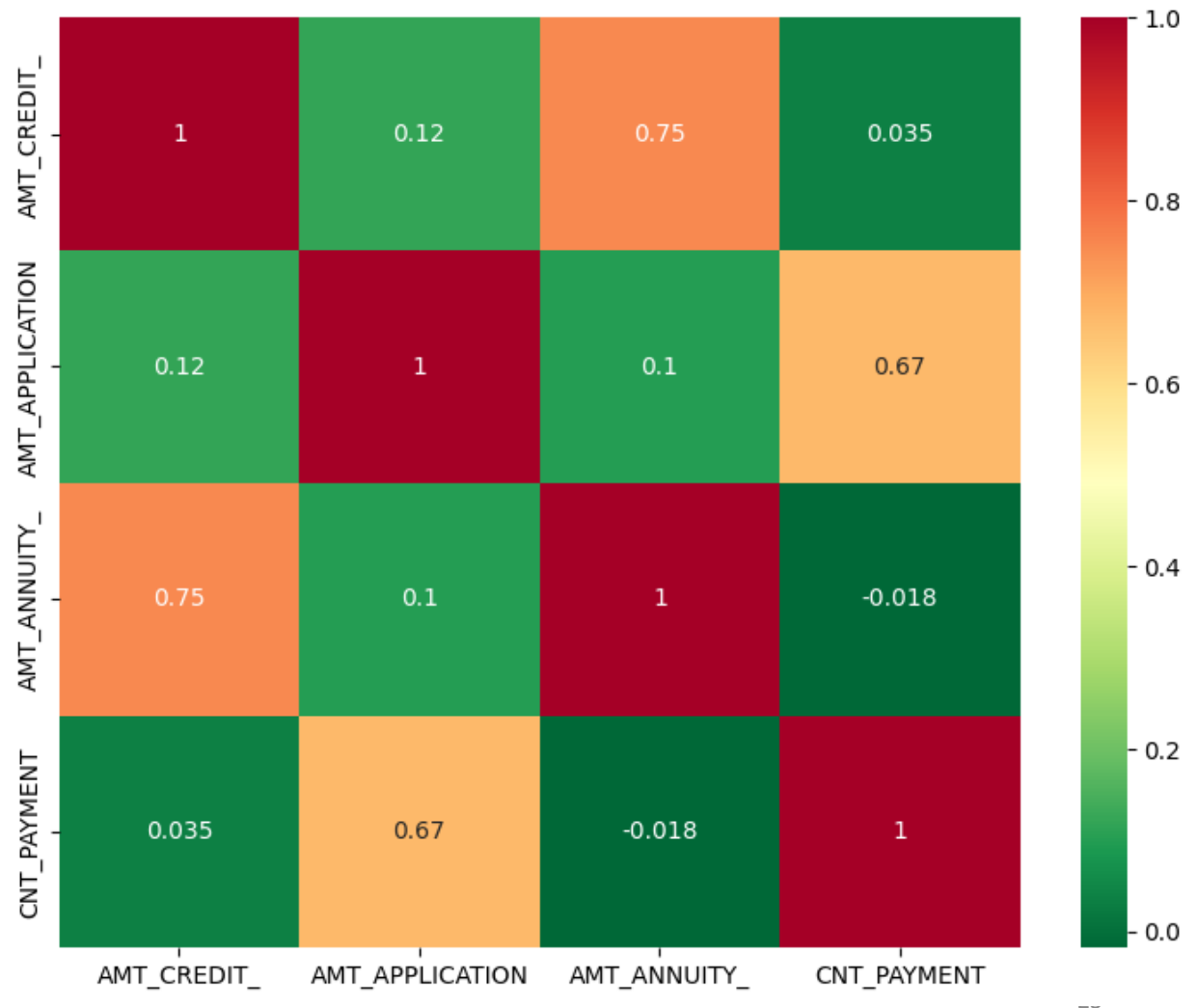
Analysis of AMT_CREDIT by distplot

Most of the Credited amount has been distributed between 20k to 1.5lakh

CREDIT EDA ASSIGNMENT

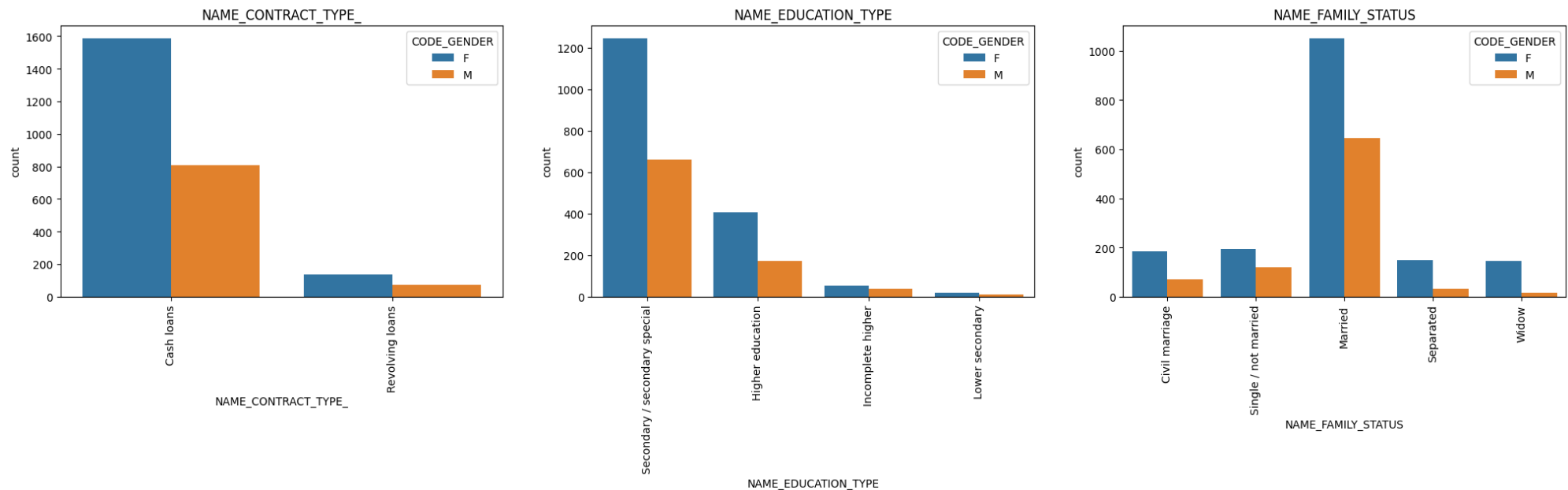
- Bivariate Analysis
- Correlation between
'AMT_CREDIT_',
'AMT_APPLICATION',
'AMT_ANNUITY_',
'CNT_PAYMENT'

Observation:
#AMT_CREDIT &
AMT_ANNUITY : Highly
Correlated, nearly 75%
'CNT_PAYMENT'
& 'AMT_APPLICATION' : Highly
Correlated, nearly 67%.



CREDIT EDA ASSIGNMENT

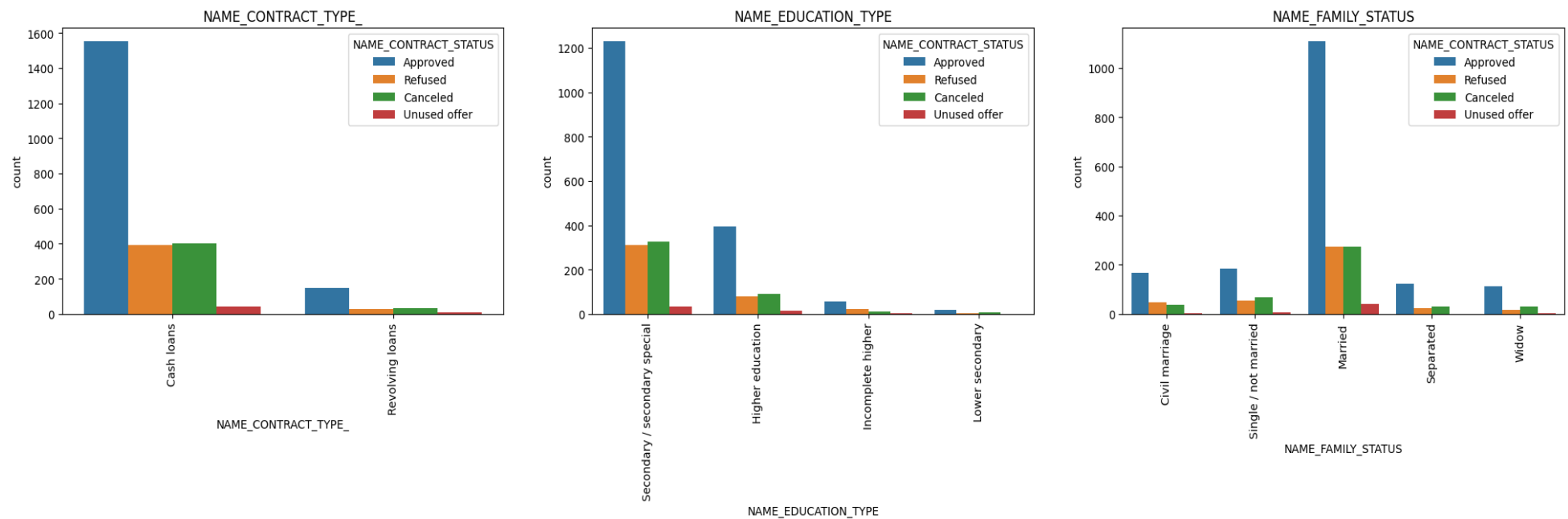
- Analysis of Contract type, Education type and Family status based on Gender:



Analysis: Cash loans are more taken by Female than Male.
Secondary/Secondary Special Education type of female are more than male.
More Married female member are taken loans compare to male

CREDIT EDA ASSIGNMENT

Analysis of Contract type, Education type and Family status based on Gender:



Analysis:

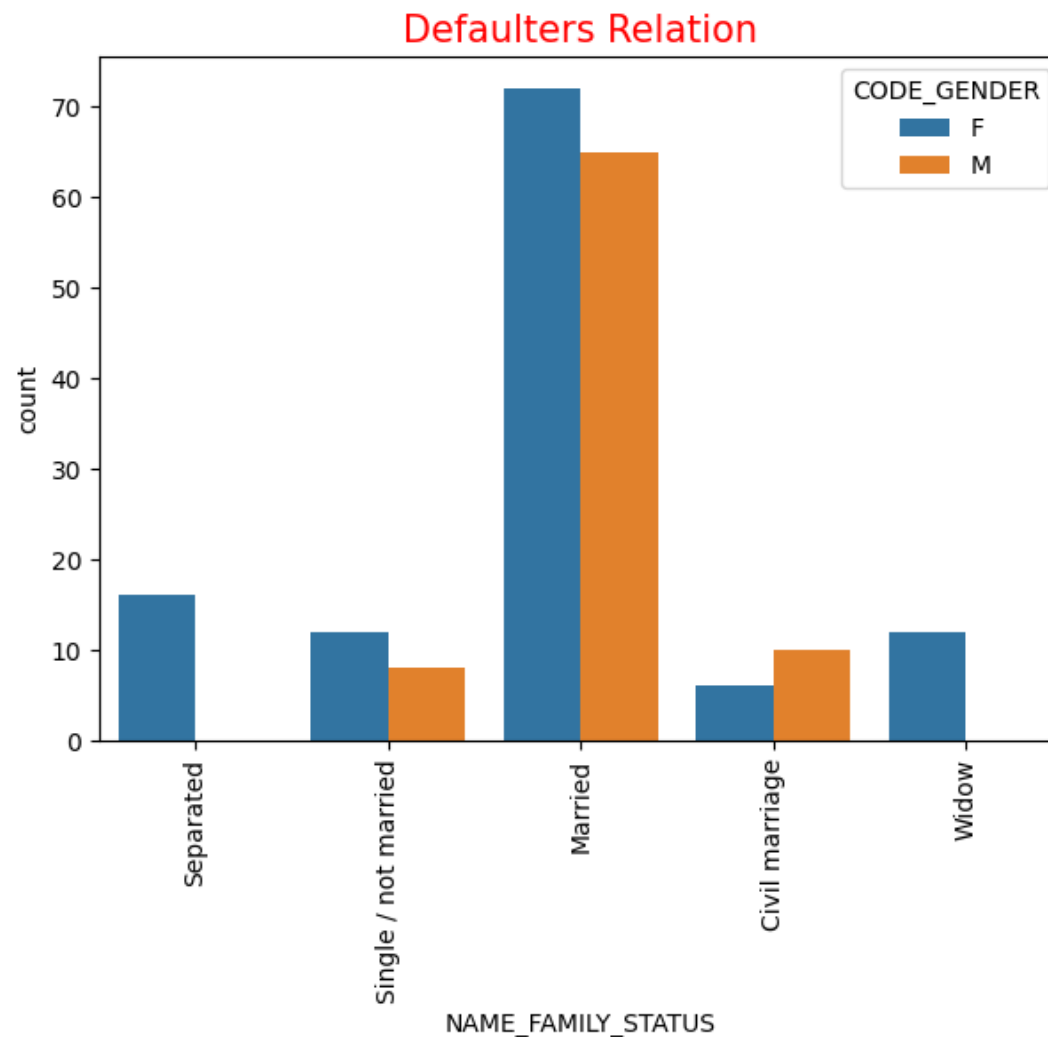
Highest no of contract got approval in Cash Loans.

Highest no of contract got approval in secondary /secondary special education type.

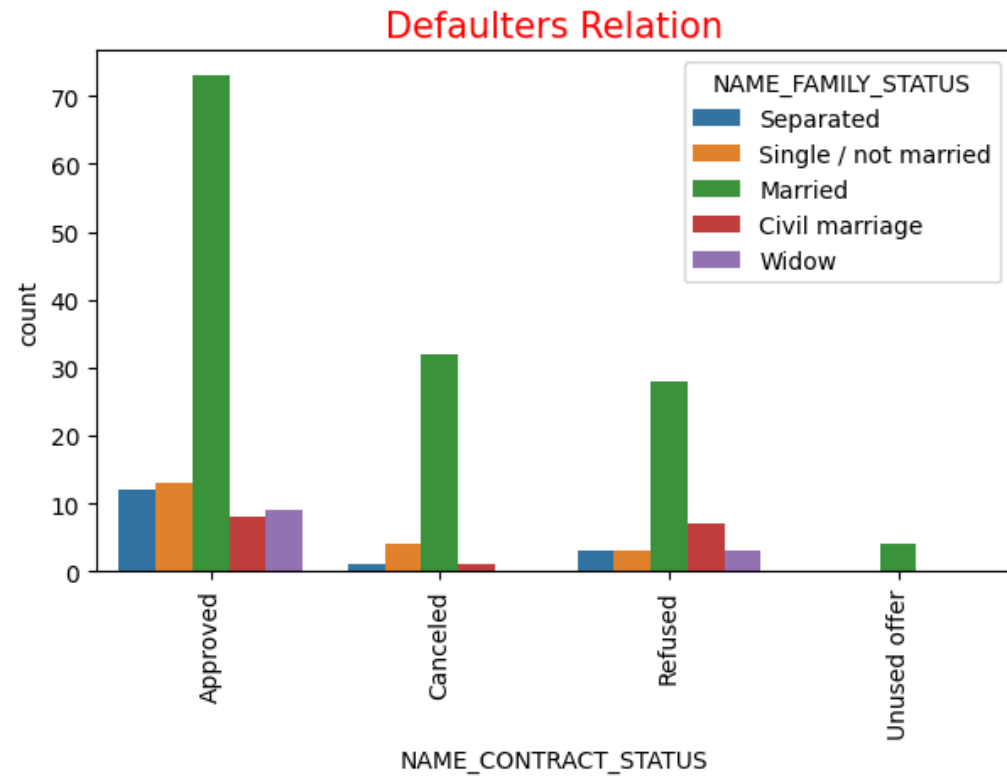
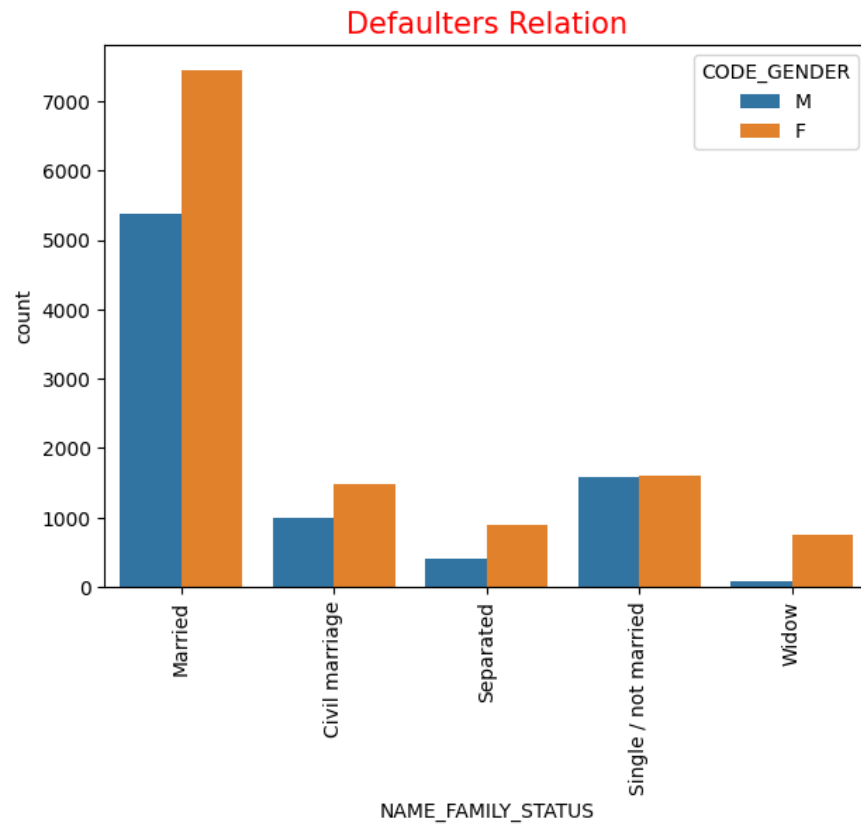
Highest no of contract got approved to those who are married

CREDIT EDA ASSIGNMENT

- Dividing previous application dataset in to categories after merging(INP_All):
 - I. Client with payment difficulties (Target=1)= INP_All_defaults
 - II. All other cases (Target=0)=INP_All_target_value_0
- Plotting two plots for defaulters and non defaulters on basis of gender and family status



CREDIT EDA ASSIGNMENT



CREDIT EDA ASSIGNMENT

- RECOMMENDATIONS AND CONCLUSIONS –
- In Contract type, female with consumer loans type are in more no of defaulters.
- More no of approved people are married but married people are high in defaulters than separated or widow
- Secondary/ secondary special people with education are highest in defaulters
- Low as well as high credit amount people carry a risk of becoming Defaulters
- Female are High in number than who face difficulties than males in Family status.
- Customers who are Married should be preferred for the loan while loans to pensioner, separated and widow carry less risk
- High number of working class, commercial associate and pensioner are non-defaulters in income type.
- Laborers, Sales Staff, Core staff and Drivers have more payment difficulties than other.
- Male customers with Academic degree should be preferred for loan compared to other groups.

Thank You.