EDA CASE STUDY ON CREDIT RISK ANALYSIS

CREATED BY : MS NEHA MAHESH PATIL

PROBLEM STATEMENT

- There are two types of risk associated with any loan request-

- Target-0 if the client/customer is likely to repay the loan, then not approving the loan results in a loss of business to a company.

- Target-1 if the if the client/customer is not likely to repay the loan, that is he/she is likely to default, then approving the loan may lead to a financial loss for a company.

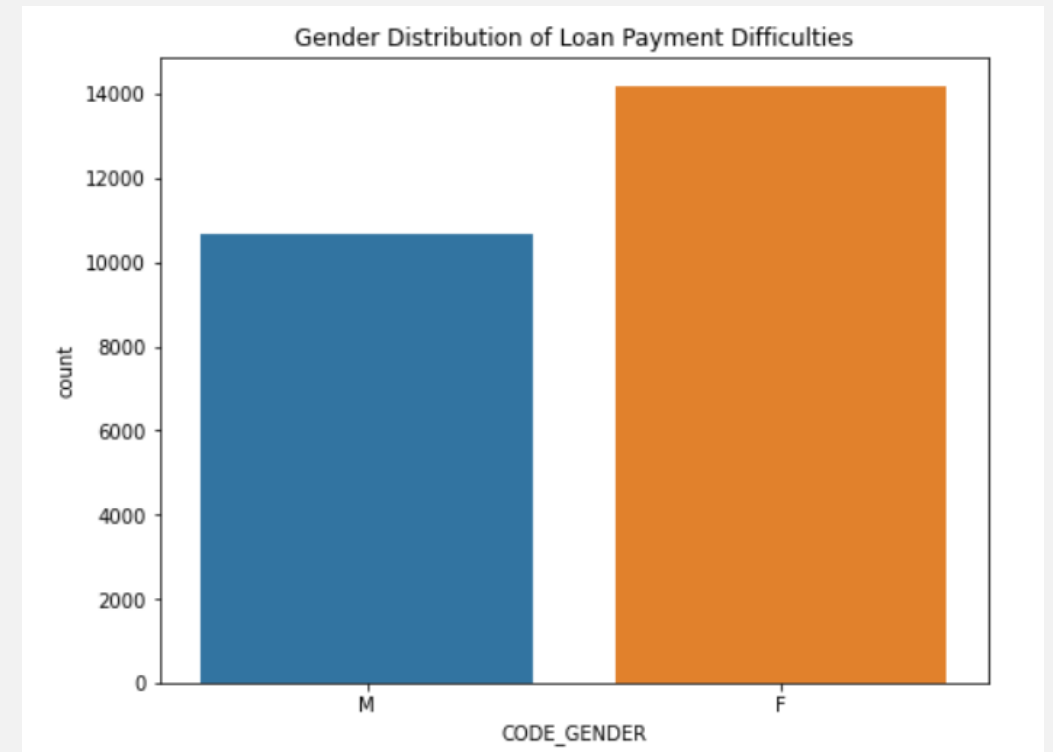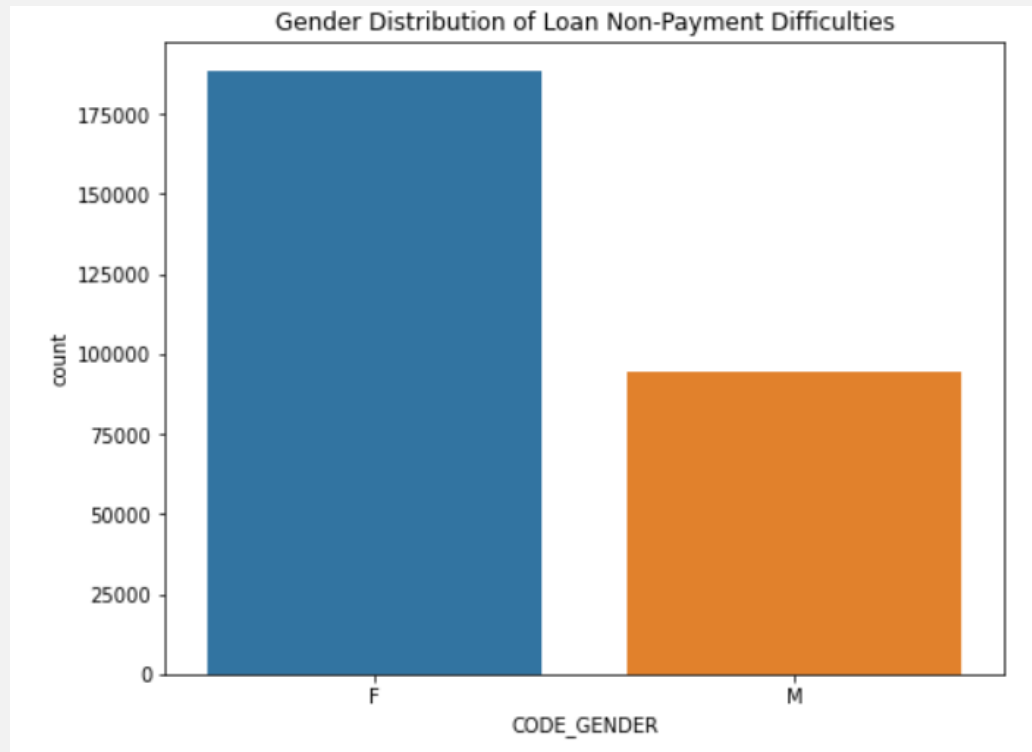# STEPS FOLLOWED FOR THE ANALYSIS

- Understanding the data and its purpose.

- Checked for Data quality issues by inspecting and removing , imputing the missing values.

- Checked for Data imbalance and analyzed the current and previous application dataset using univariate analysis, bivariate analysis and checked correlations.

- Divided the dataset into two parts based on the target variables(0,1) for further analysis.

- Merged the current application with the previous application and performed the analysis on the same.

ANALYSIS BASED ON TARGET VARIABLE

 - It is clear that there is imbalance between the defaulters and non-defaulters.

- The percentage of non-defaulters is more than that of defaulters.

- Defaulters are 8% of the total distribution.

Target Variable- Defaulter V/S Non-Defaulter
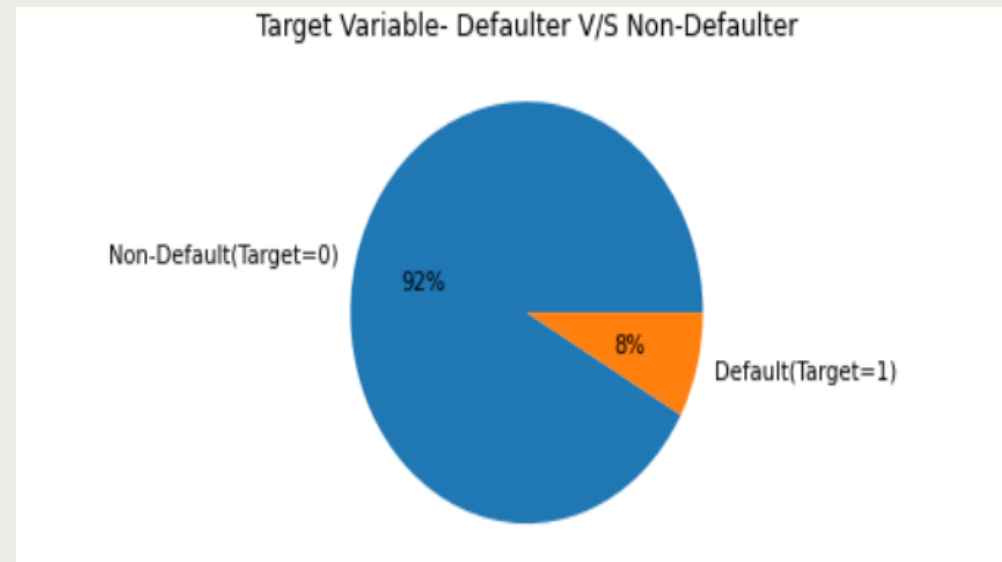
UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON GENDER

Gender Distribution of Loan Non-Payment Difficulties

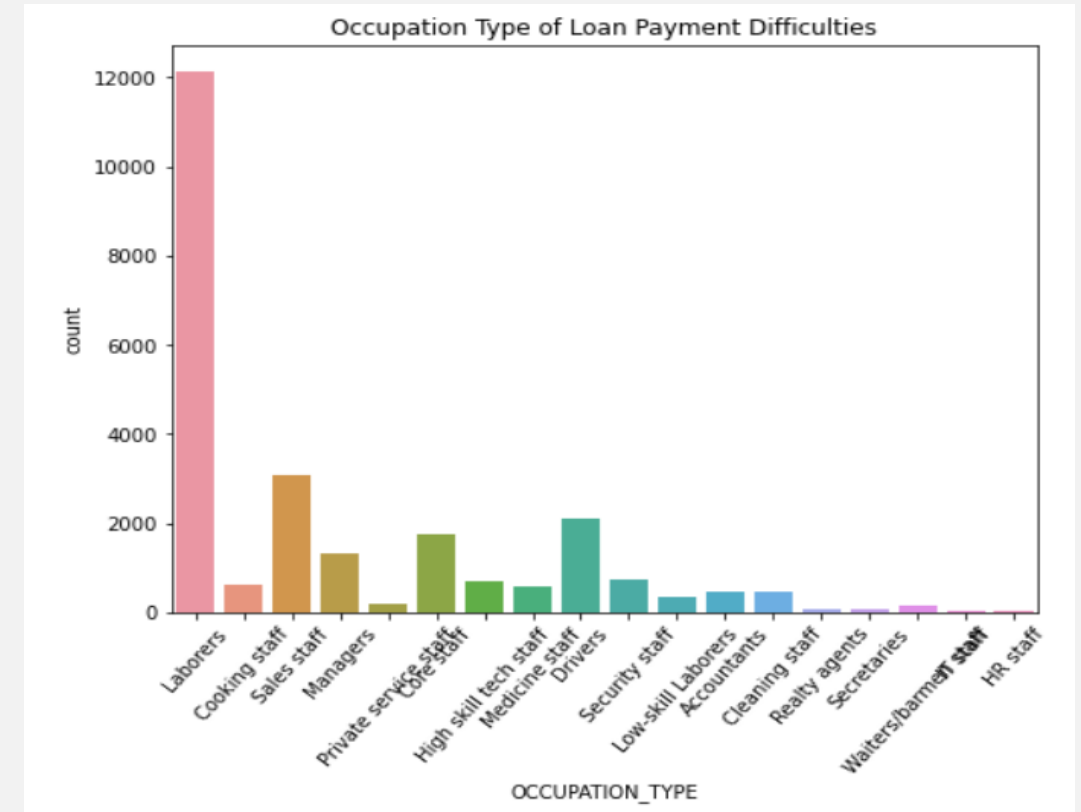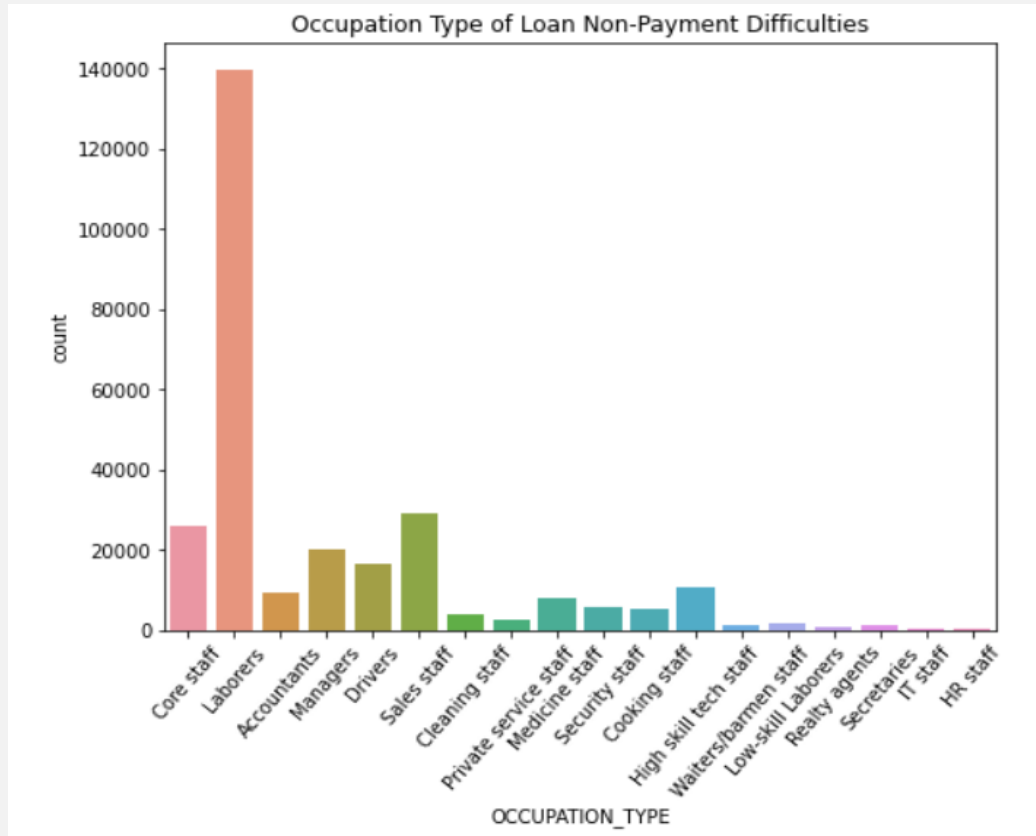Gender Distribution of Loan Payment Difficulties

- Comparing the Payment Difficulties and Non Payment Difficulties based on the gender, we observe that Females are the majority in both the cases.

- Also, there is an increase in Male Payment Difficulties from Non-Payment Difficulties.

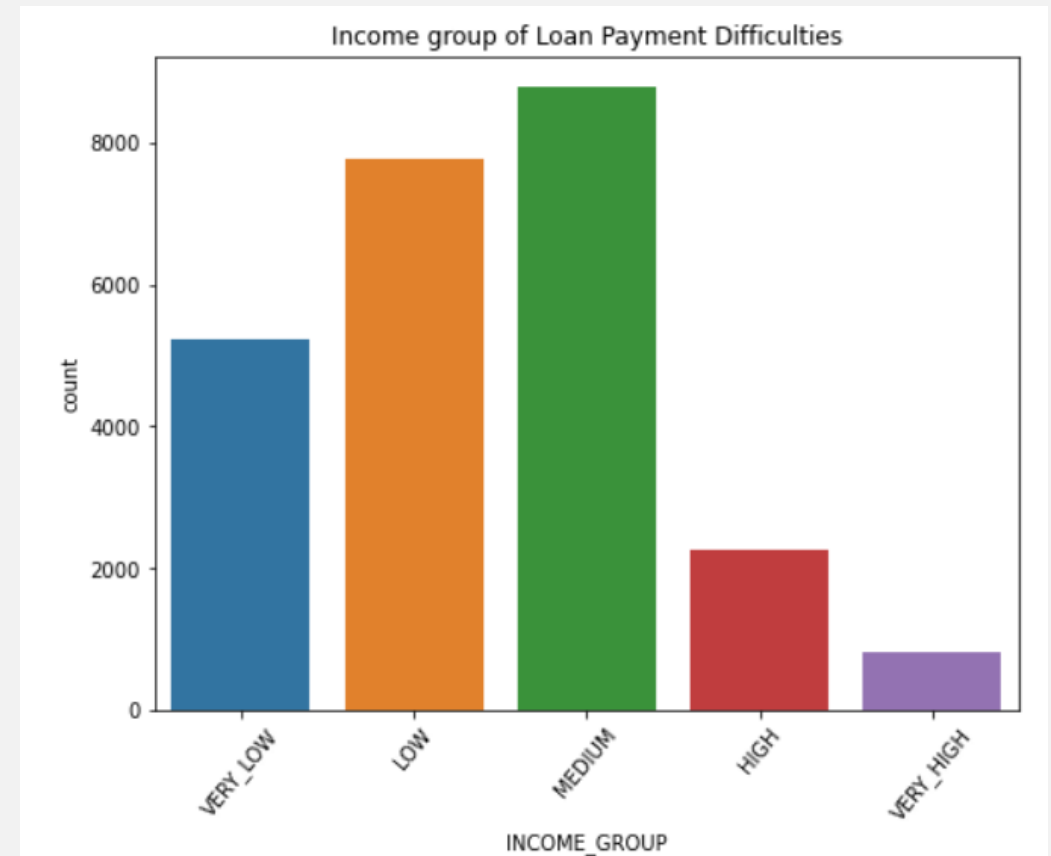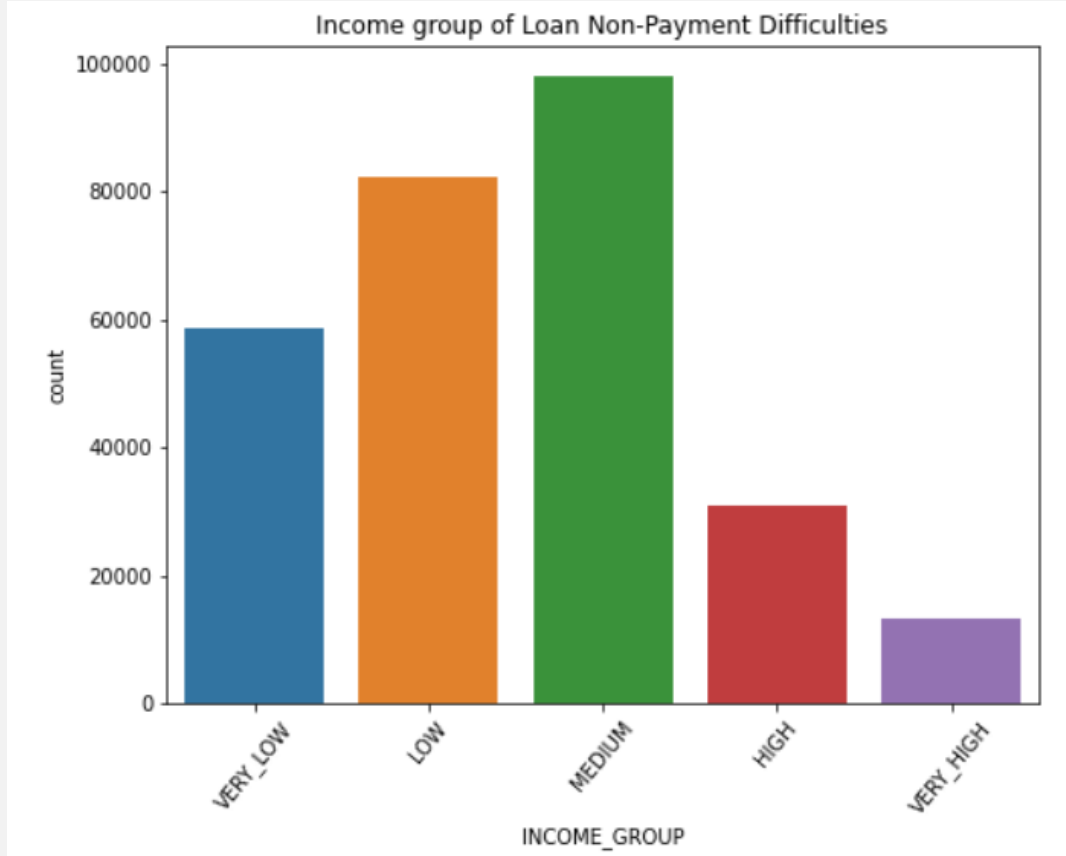# UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON INCOME TYPE
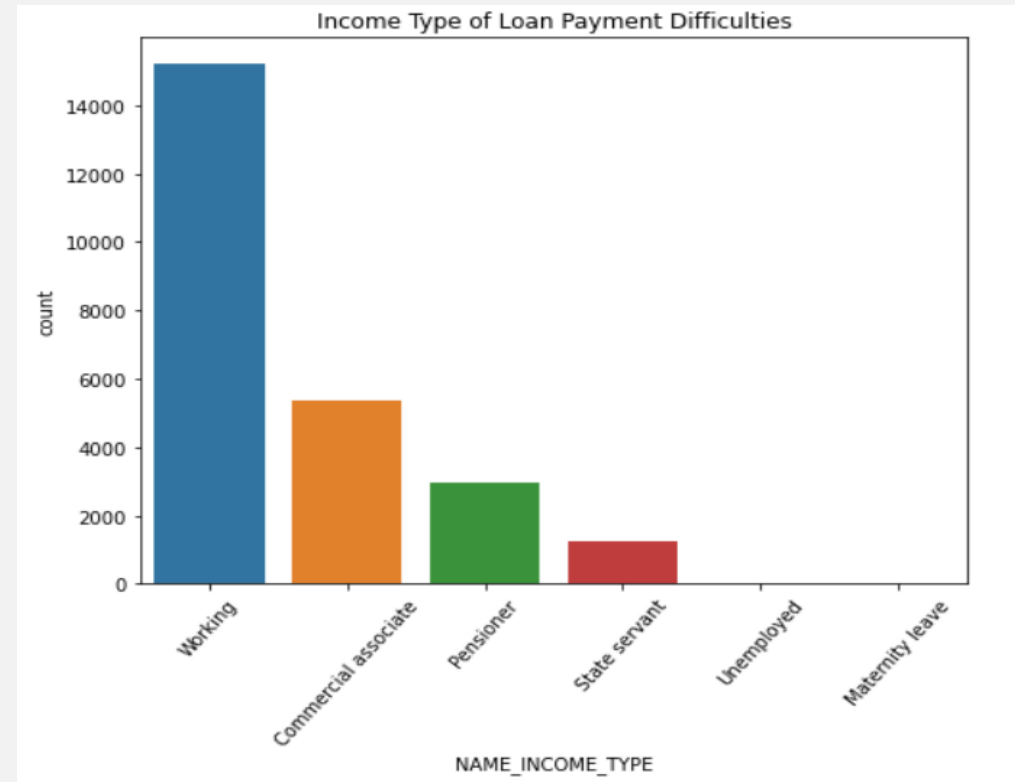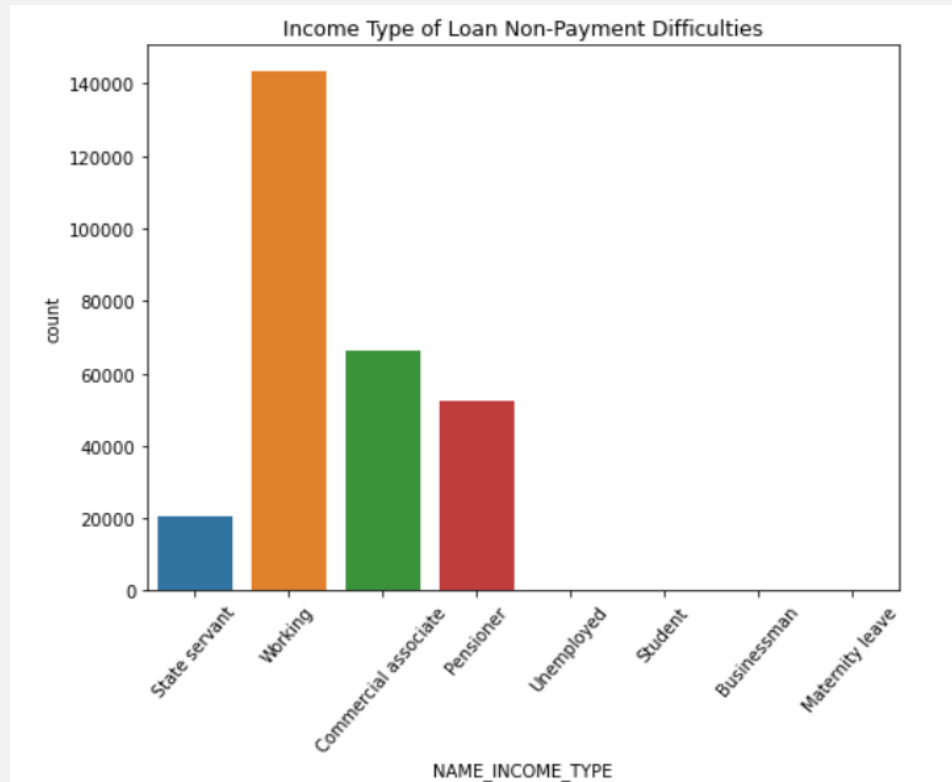
- We can conclude that students don't default. They don't need to pay during the time they are students.

- We can also see that businessmen never default.

- Working class people has more loan payment difficulties as compared to non-payment difficulties/non-defaulters.

UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON OCCUPATION TYPE

Occupation Type of Loan Non-Payment Difficulties

Occupation Type of Loan Payment Difficulties

- We can see that labourers are the majority in both the cases. Also, labourers are more likely to pay the loan on time as compared to the loan payment difficulties.

- HR Staff, IT staff, waiters are least likely to be part of loan payment process.
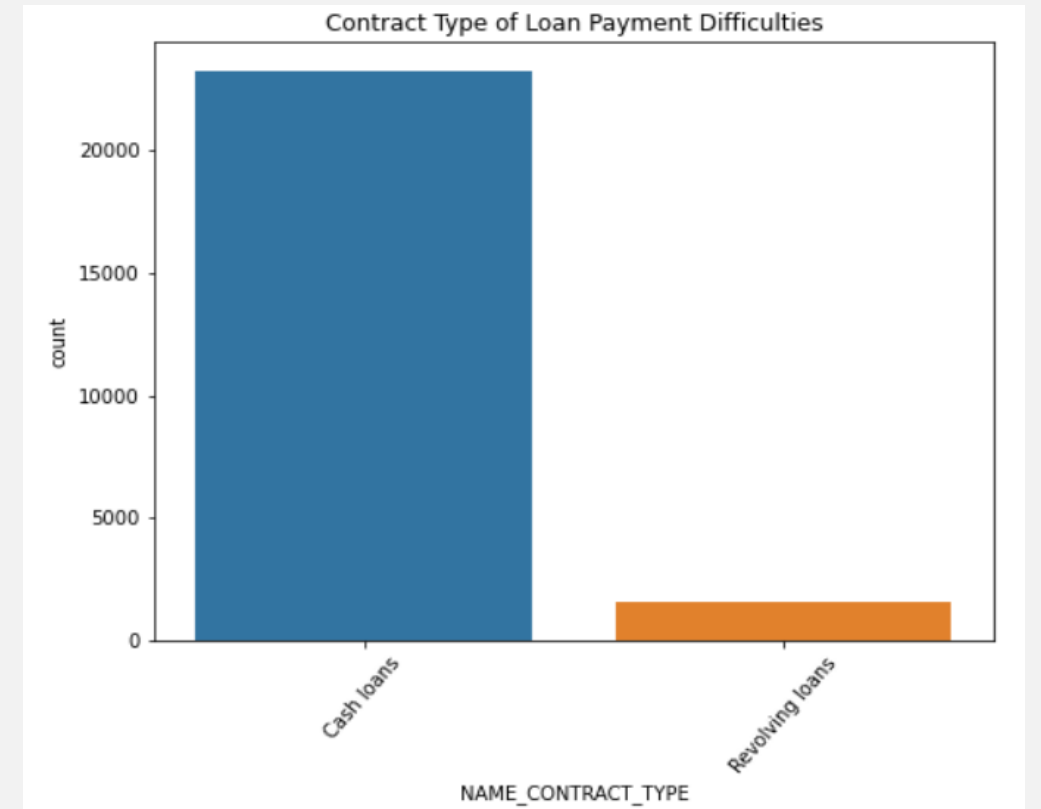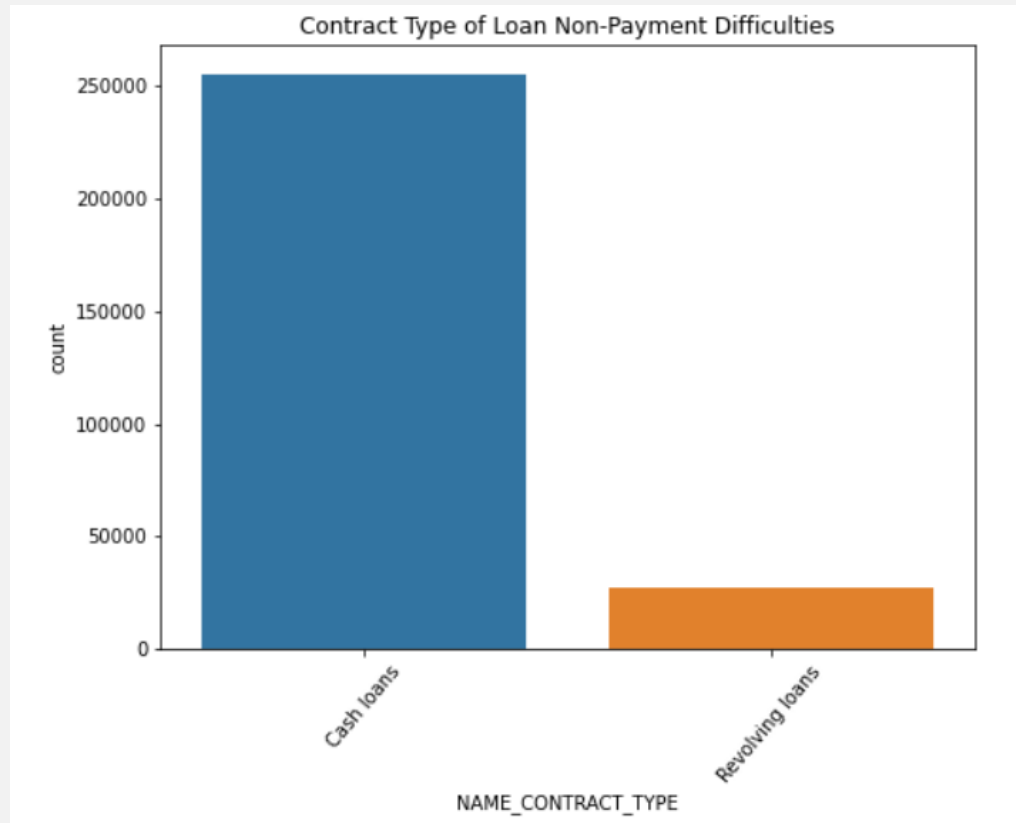
UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON INCOME GROUP

Income group of Loan Non-Payment Difficulties

Income group of Loan Payment Difficulties

- We observe an increase in the loan payment difficulties whose income is low when compared with the other income groups of payment difficulties and non-payment difficulties.
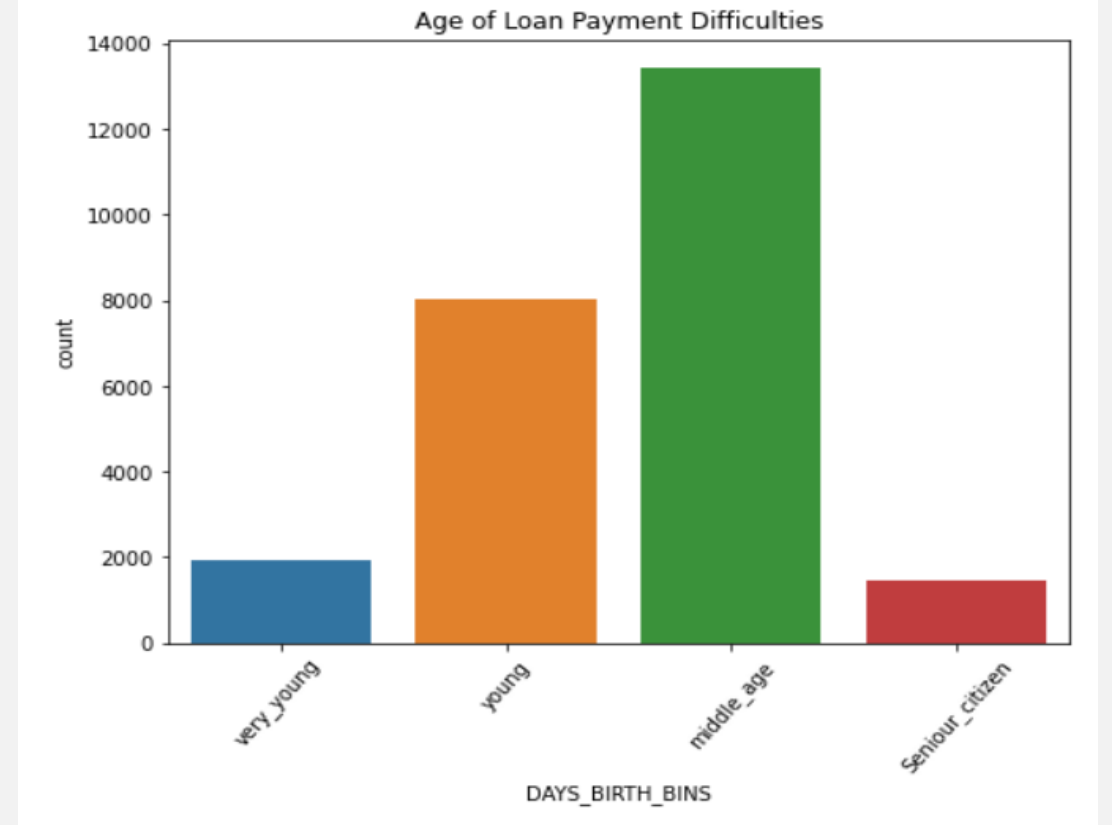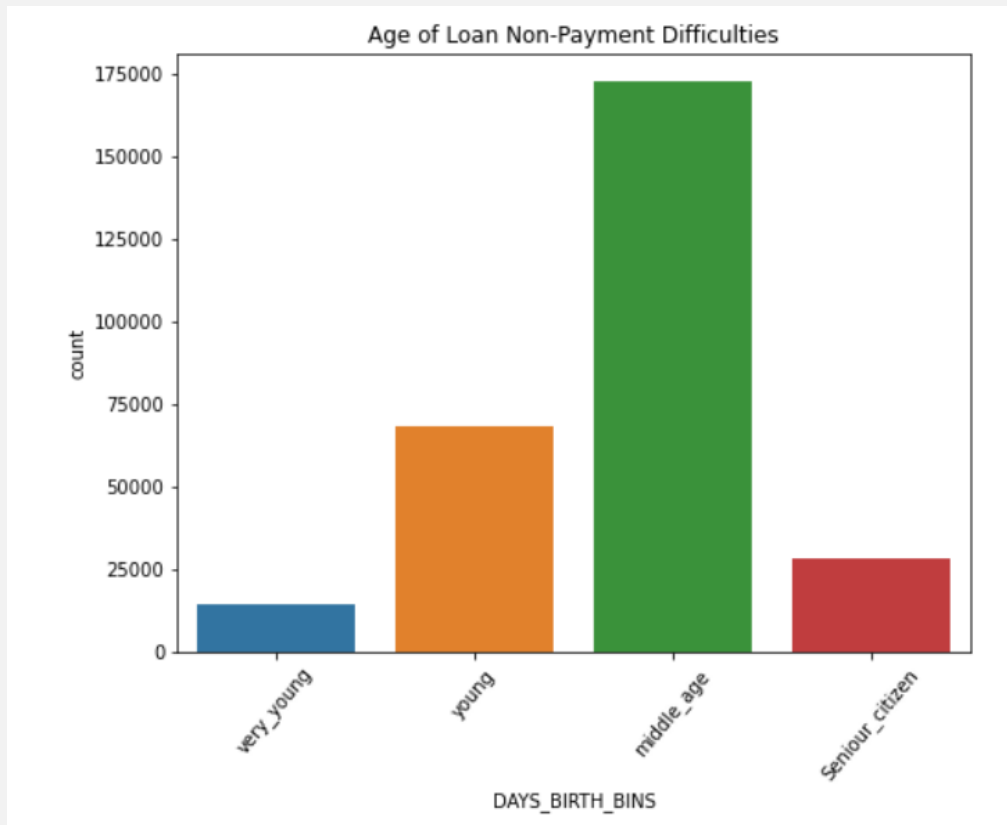
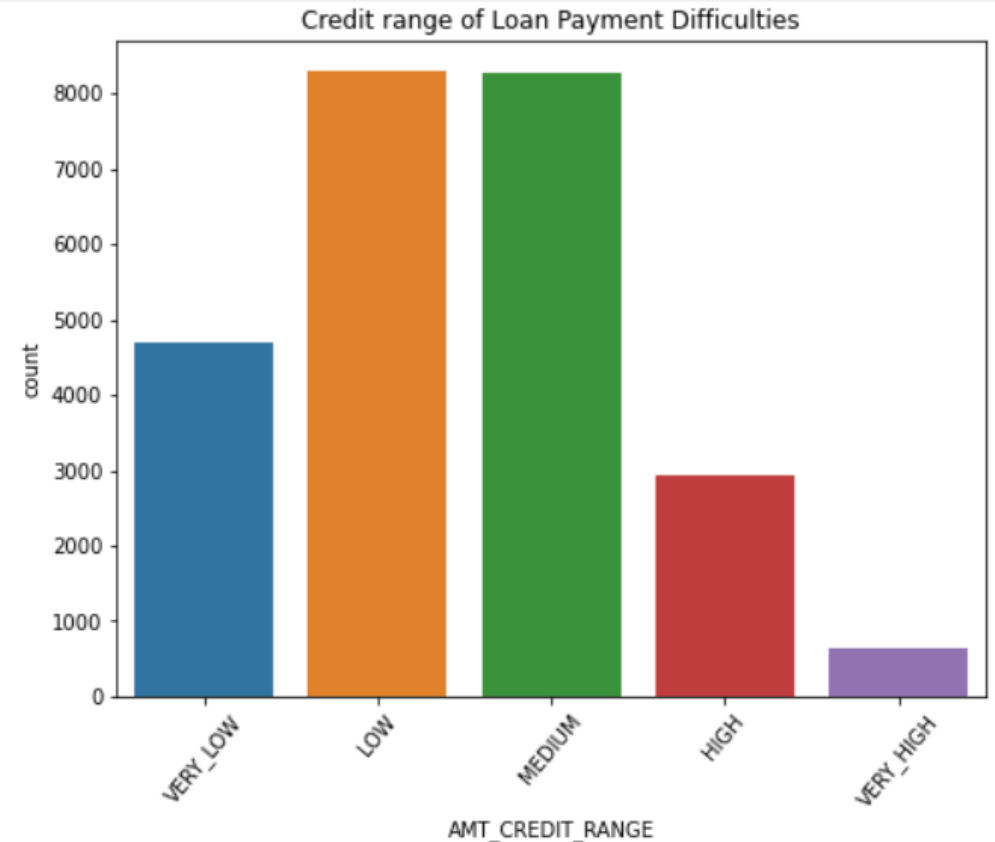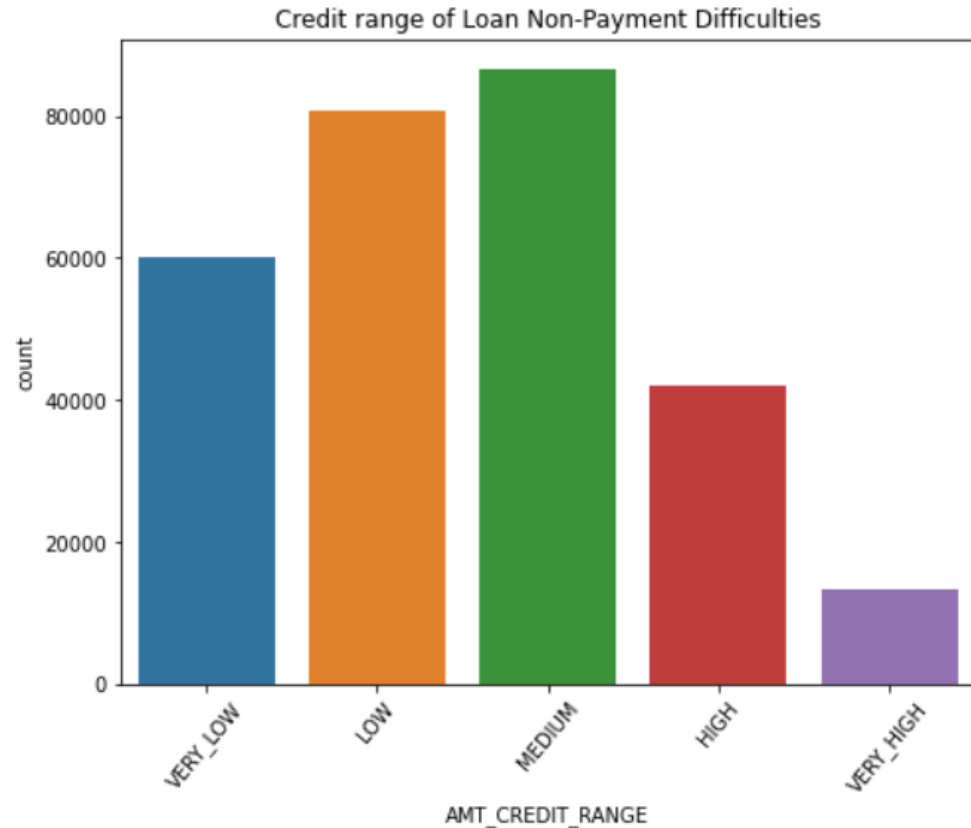# UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON CONTRACT TYPE

- We can observe that cash loans are preferred by both Loan Payment difficulties and Loan Non-Payment difficulties.

- Also, there is decrease in the loan payment difficulties in case of revolving loans.

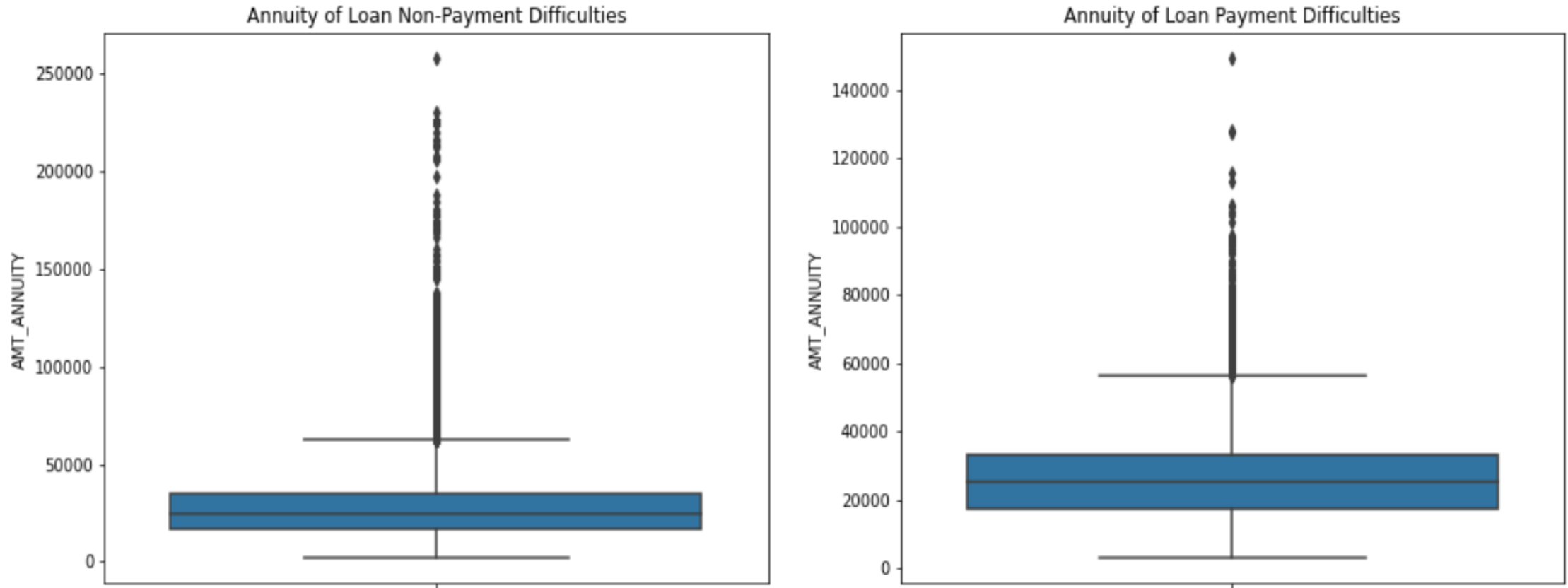# UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON AGE

- We observe that there is increase in loan payment difficulties who are young in age when compared to other age groups in loan payment and non-payment difficulties.

UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON CREDIT RANGE

Credit range of Loan Non-Payment Difficulties
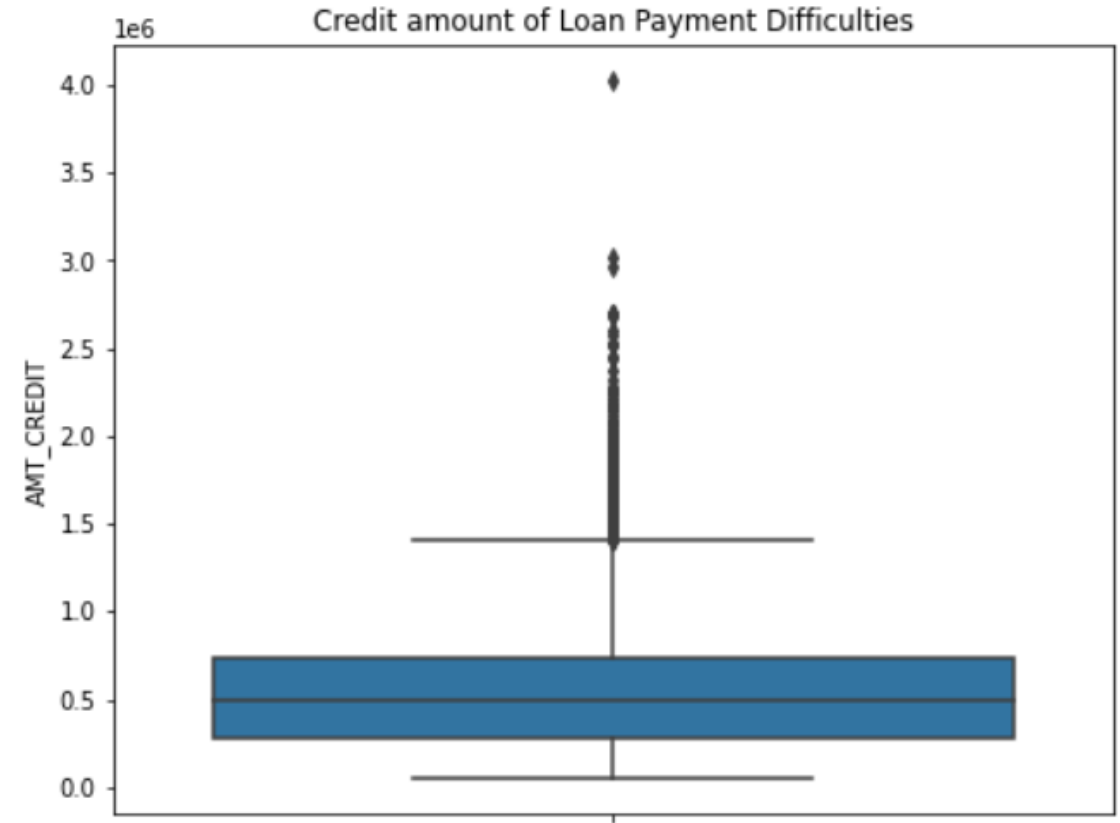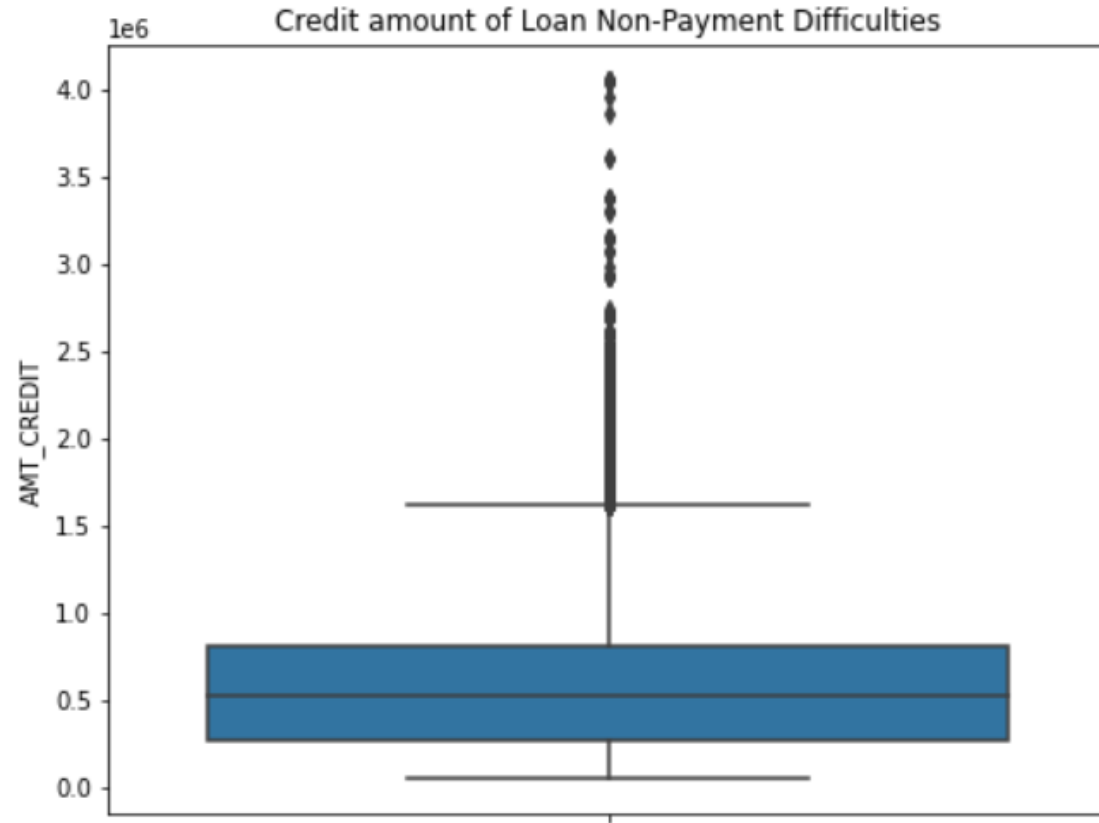

Credit range of Loan Payment Difficulties

- We can observe that the low credit range has more loan payment difficulties when compared to the remaining credit ranges in loan payment and non-payment difficulties.

UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON ANNUITY


- We can observe some outliers and first quartile is bigger than third quartile in case of loan payment difficulties. This means that most of the annuity clients are from first quartile.

UNIVARIATE ANALYSIS OF CURRENT APPLICATION BASED ON CREDIT AMOUNT

- We can observe some outliers and first quartile is bigger than third quartile in case of loan payment difficulties. This means most of the credits of clients are present in first quartile.

BIVARIATE ANALYSIS OF CURRENT APPLICATION FOR TARGET-0
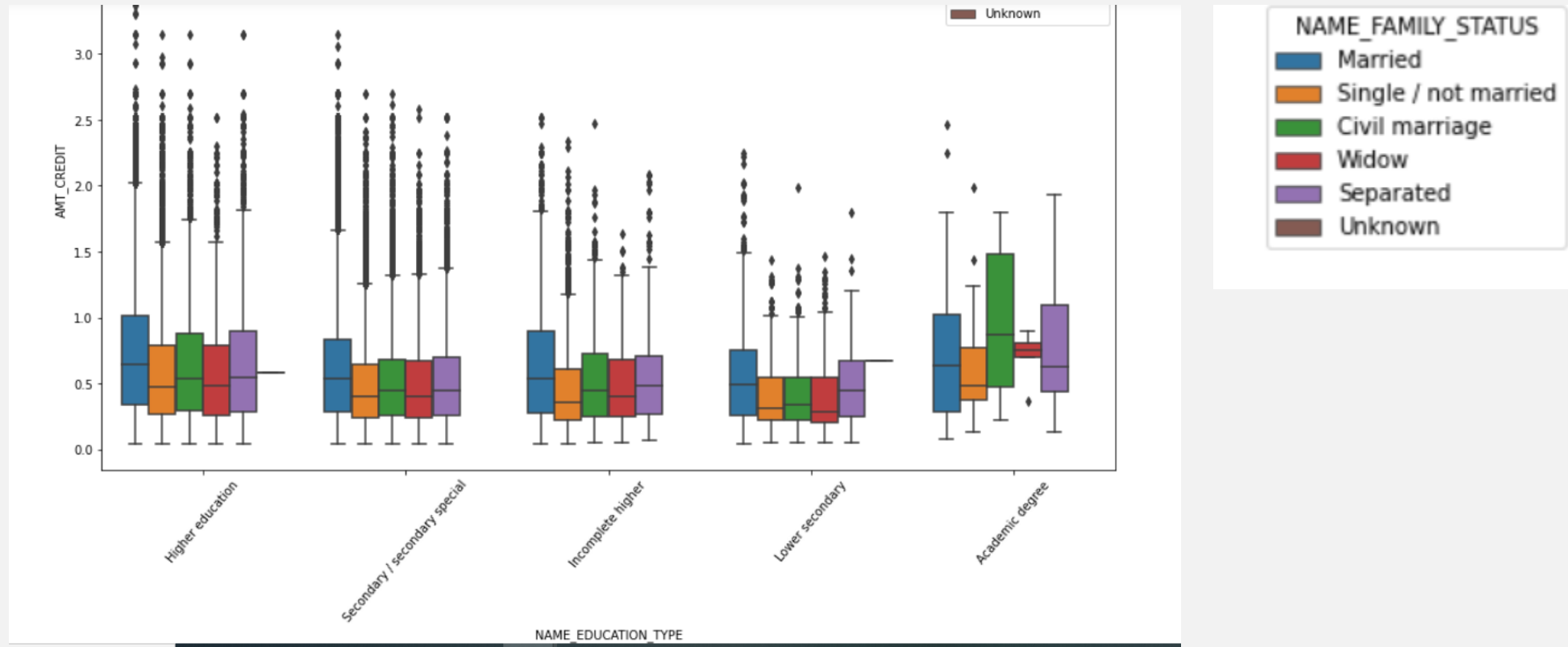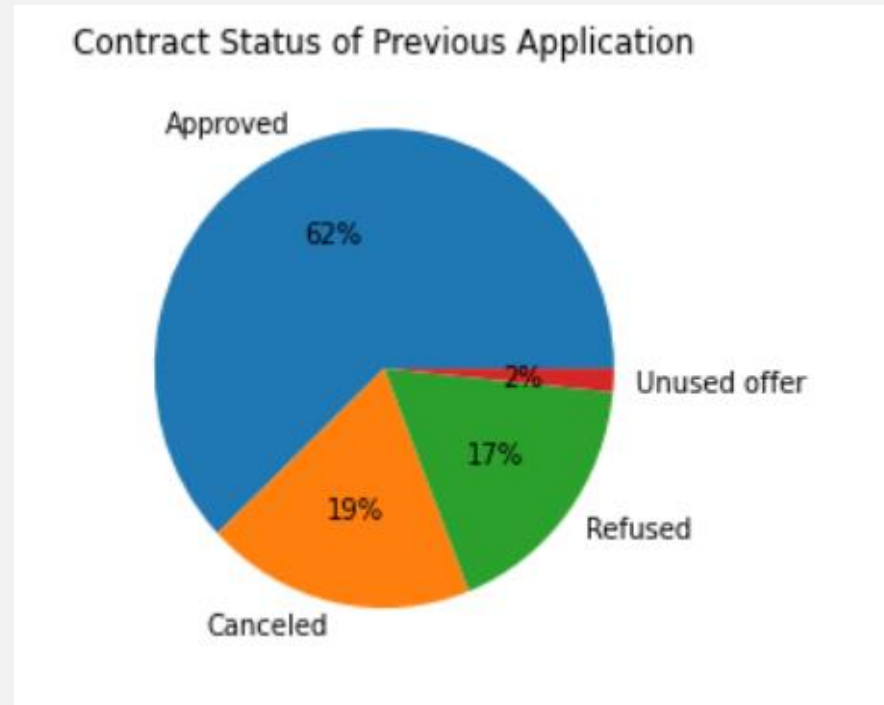

- - From the above box plot we can conclude that Family Status of 'Civil Marriage', 'Marriage' and 'Separated' of Academic degree education are having higher number of credits than others.

- - Also, Higher education of family status 'marriage', 'single' and 'civil marriage' have more outliers.

- - 'Civil Marriage' in Academic Degree have most of the credits in the third quartile

BIVARIATE ANALYSIS OF CURRENT APPLICATION FOR TARGET-1
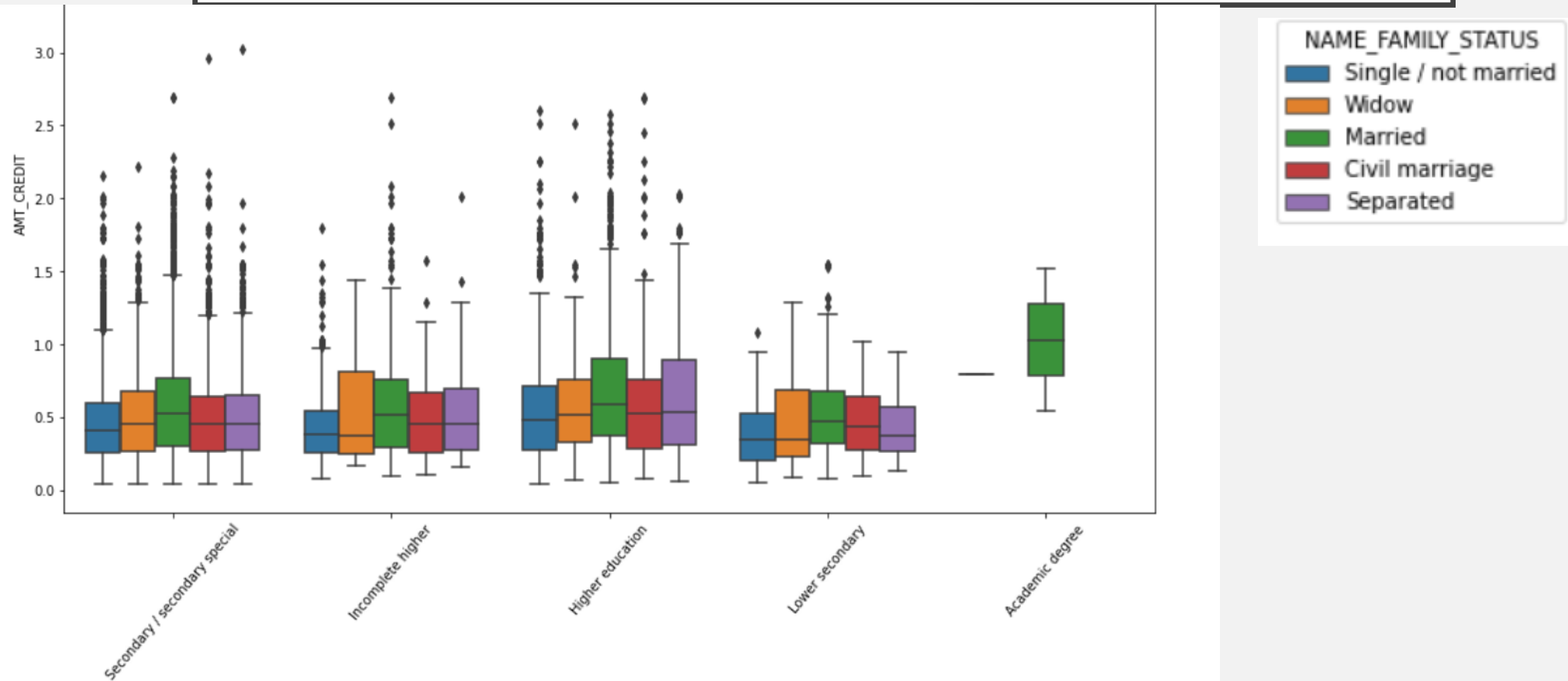
- • - From the above boxplot we can conclude that the family status of 'Marriage' of Academic Degree ducation have higher number of credits than others in case of loan payment difficulties.

- • - Also, Academic Degree of family status have no outliers.

- • - In most of the cases, the graphs of Loan non-payment difficulties and payment difficulties appear similar.
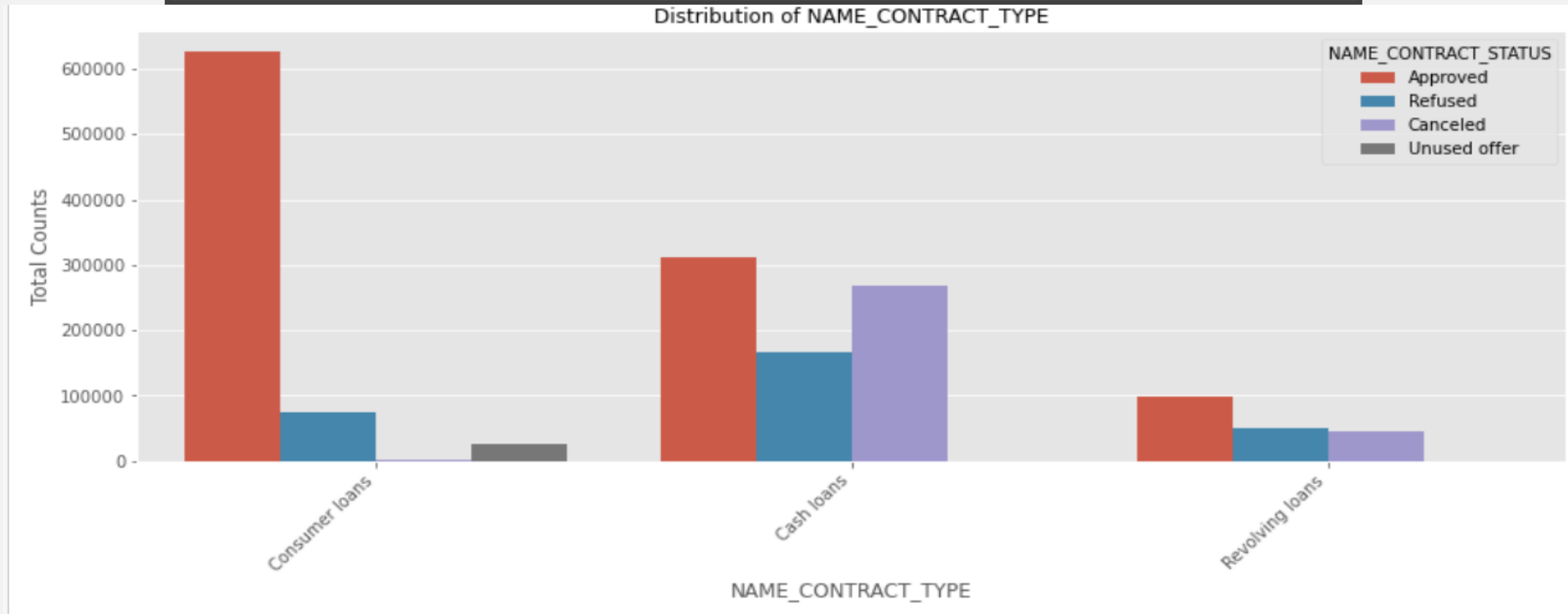
UNIVARIATE ANALYSIS OF PREVIOUS APPLICATION BASED ON CONTRACT STATUS

Contract Status of Previous Application

- We can see that majority of the loans are approved and very less percentage of loans are unused offer.

# UNIVARIATE ANALYSIS OF PREVIOUS APPLICATION BASED ON CONTRACT TYPE

Distribution of NAME_CONTRACT_TYPE

- From the above plot, we can conclude that most of the applications are for consumer loans and cash loans.

- Also, the cash loans are refused more often than others.

BIVARIATE ANALYSIS BY COMBINING CURRENT AND PREVIOUS APPLICATION

% of Loan payment difficulties for NAME_CONTRACT_STATUS and NAME_CLIENT_TYPE
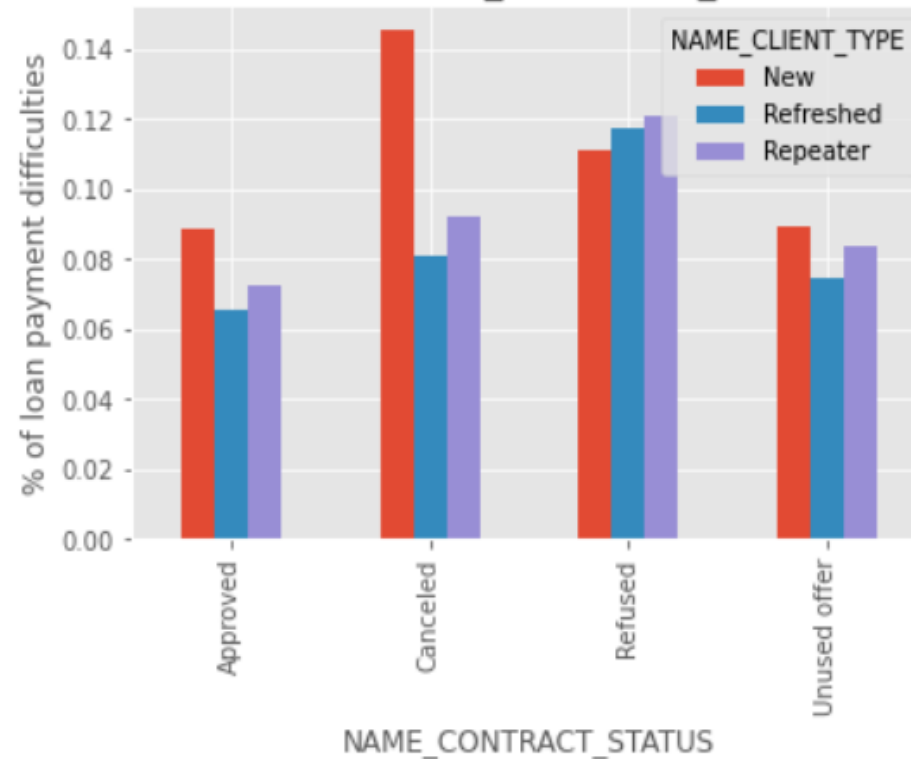
- It can be observed that the client who were 'New' and had 'Canceled' previous application tend to have more percentage of loan payment difficulties in current application.

INSIGHTS

- The proportion of defaulters is around 8%

- The bank lends more to females.

- Higher amount loans, less income defaults less.

- Loans previously refused or canceled has higher default rates.

- More cash loans go into default, bank should give more revolving loans.

- Proportion of working defaults more/ State servants less.