Credit EDA Case Study Assignment

Submitted By: Priyanka T

Purpose:

- To develop a basic understanding of risk analytics in banking and financial services and how to use data to reduce the losses. EDA is used to analyse the patterns present in the data and ensures that customers capable of loan repayment are not rejected.

Steps:

1. Data Sourcing and Understanding(application data)
2. Data Handling and Cleaning
   1. Checking Duplicates
   2. Checking Data Types
   3. Missing Value Handling
   4. Outliers Handling
   5. Errors in Data(Negative Values,XNA,etc)
   6. Binning
3. Data Imbalance Handling and Data Segmentation
4. Univariate, Bivariate and Multivariate Analysis
5. Previous data Analysis
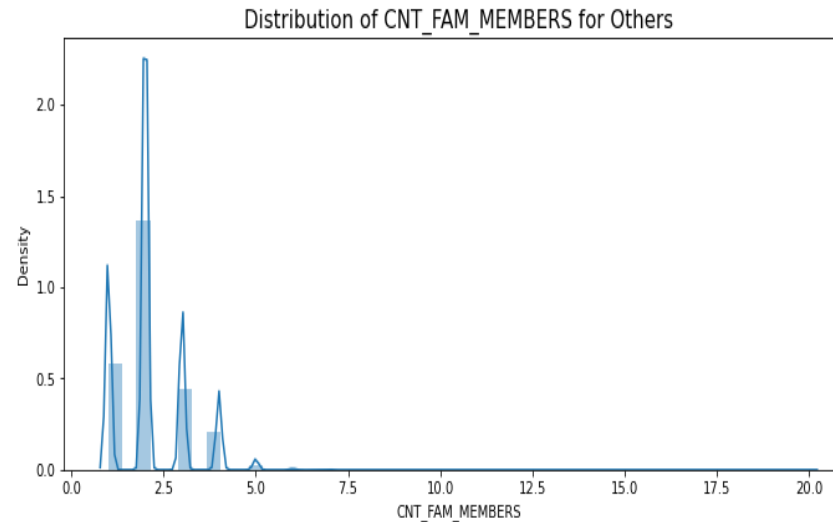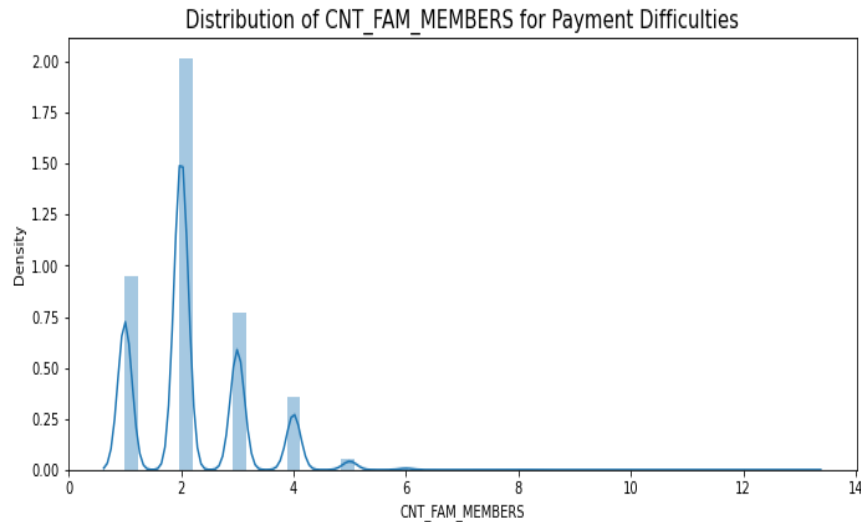6. Data Merging and Analysis
7. Recommendation

Data Imbalance
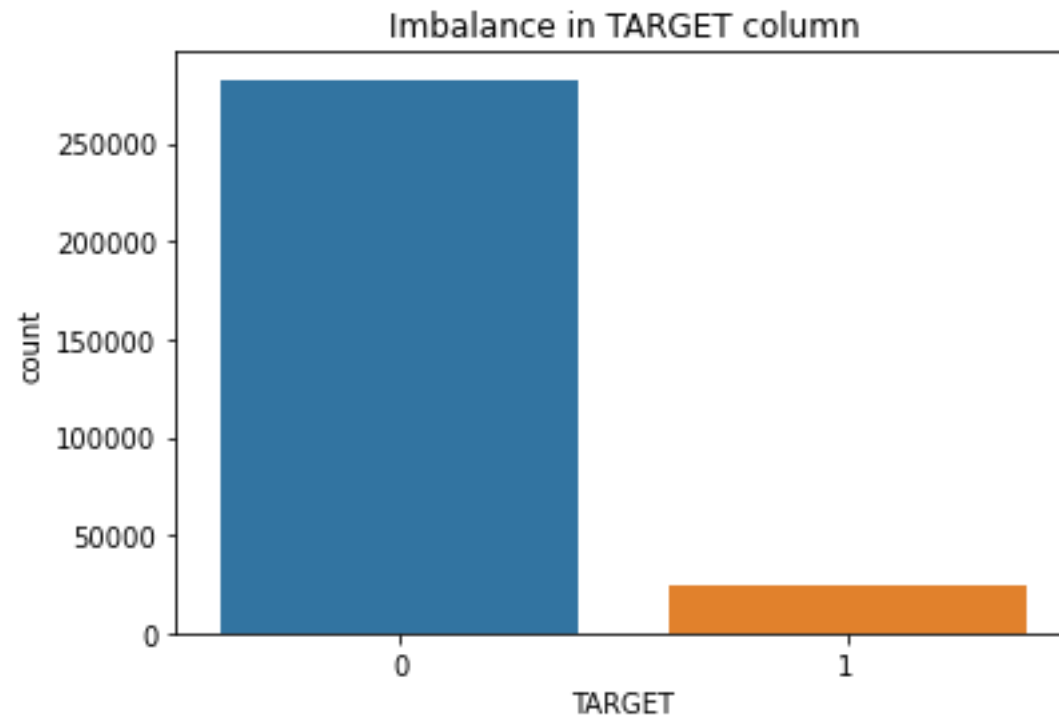
1: Having Payment Difficulties
0: Others

Univariate Analysis of Segmented Data:

Univariate analysis for numerical (or continuous) columns/categories

For CNT_FAM_MEMBERS there is not much variation in Distribution of CNT_FAM_MEMBERS for Payment Difficulties and Others, but family of 3 applies more for loan.

Distribution of AMT_ANNUITY for Payment Difficulties

Distribution of AMT_ANNUITY for Others
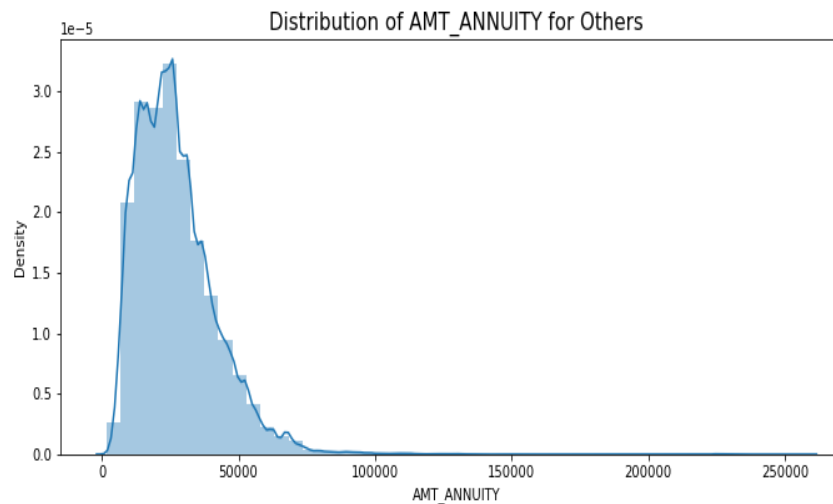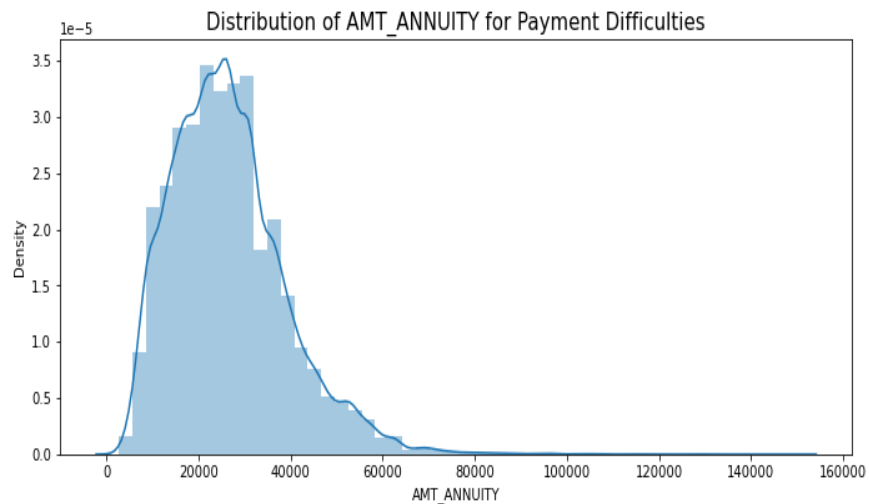
AMT_ANNUITY facing Payment Difficulties is more in the range of 135000 to 325000
AMT_ANNUITY for Others is more in range 5000 to 40000
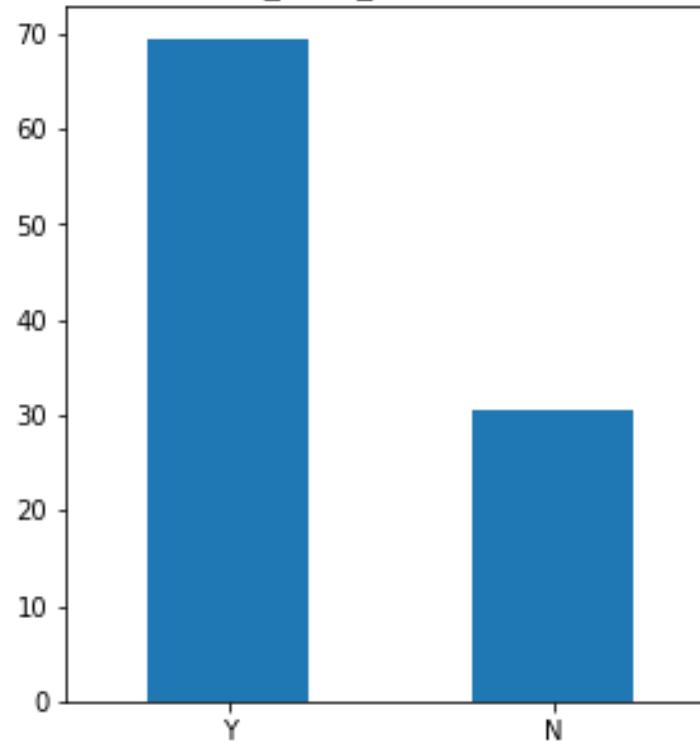
Univariate analysis for Categorical columns
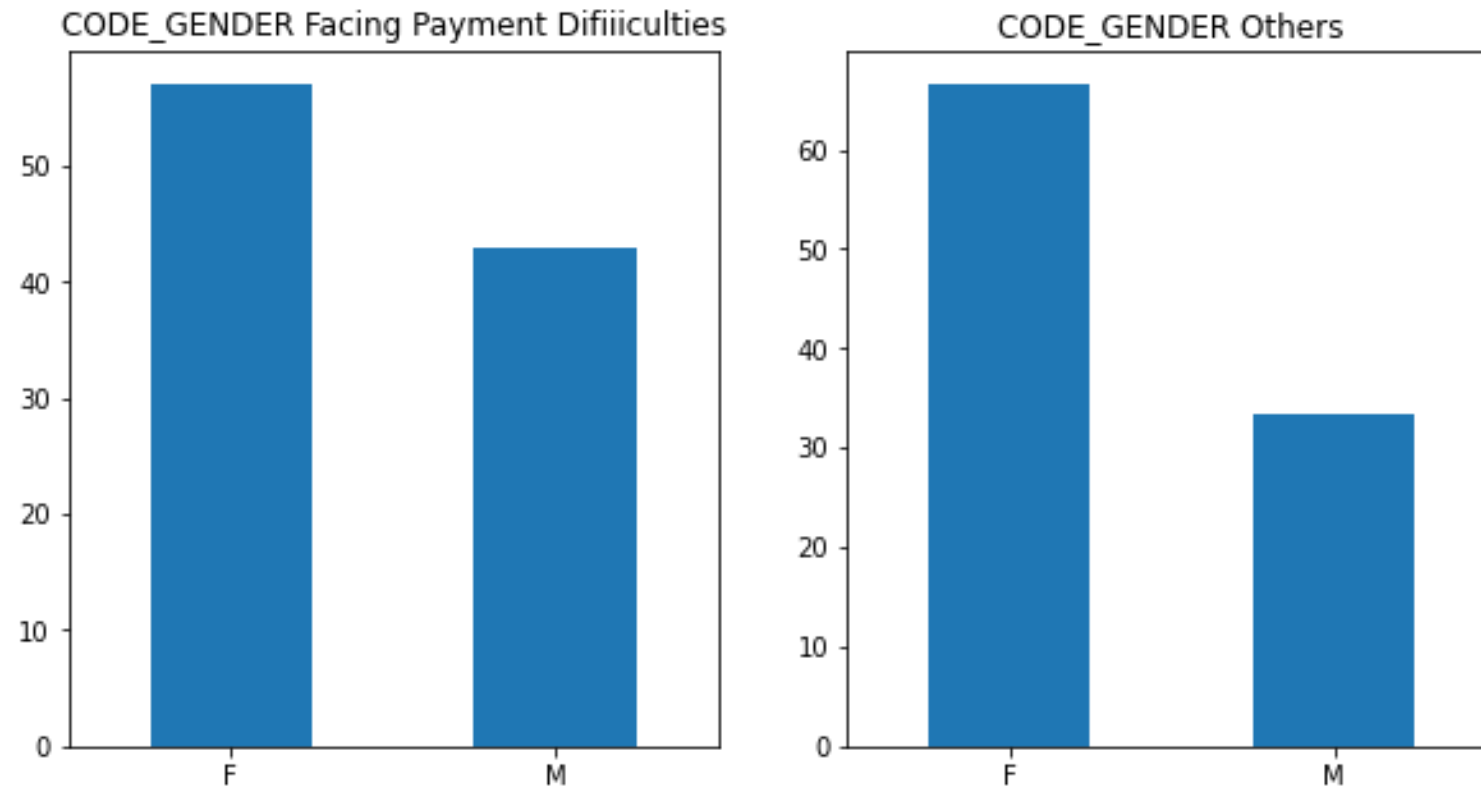
In CODE_GENDER Female are Facing more Payment Difficulties
in TARGET_1
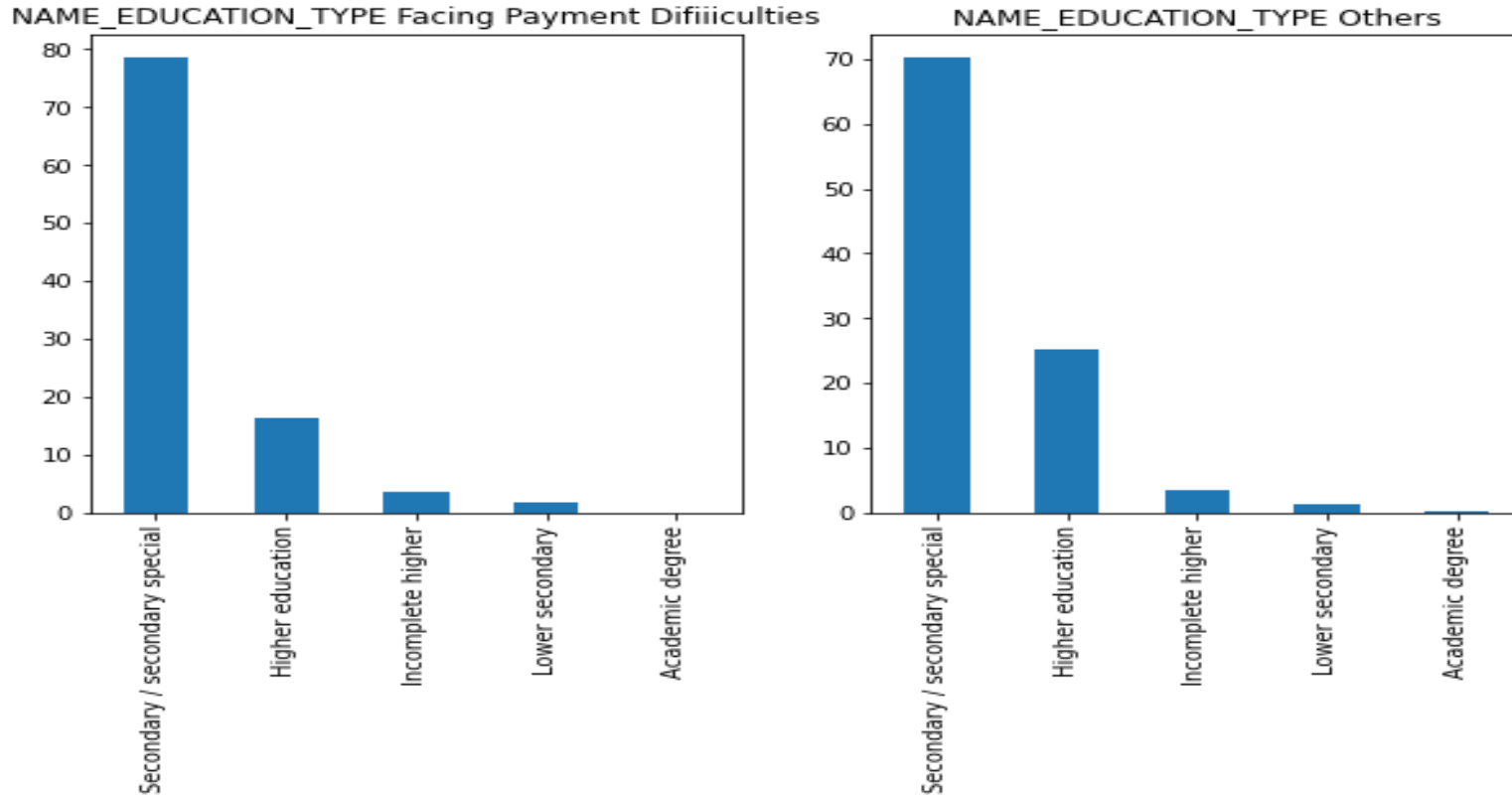But Overall Female are applying more for loan
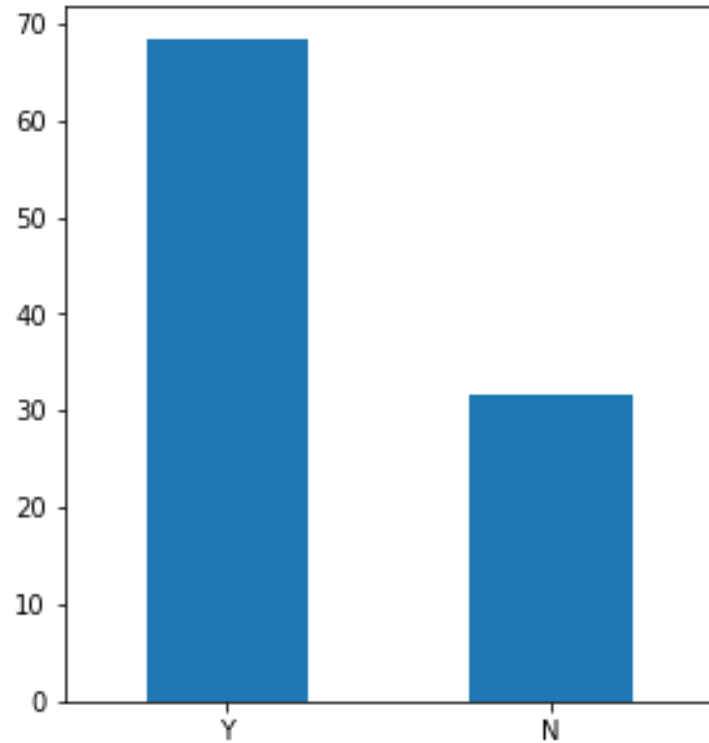
In FLAG_OWN_REALTY customer who own house are applying more for loan and around 68% are defaulters who own a house
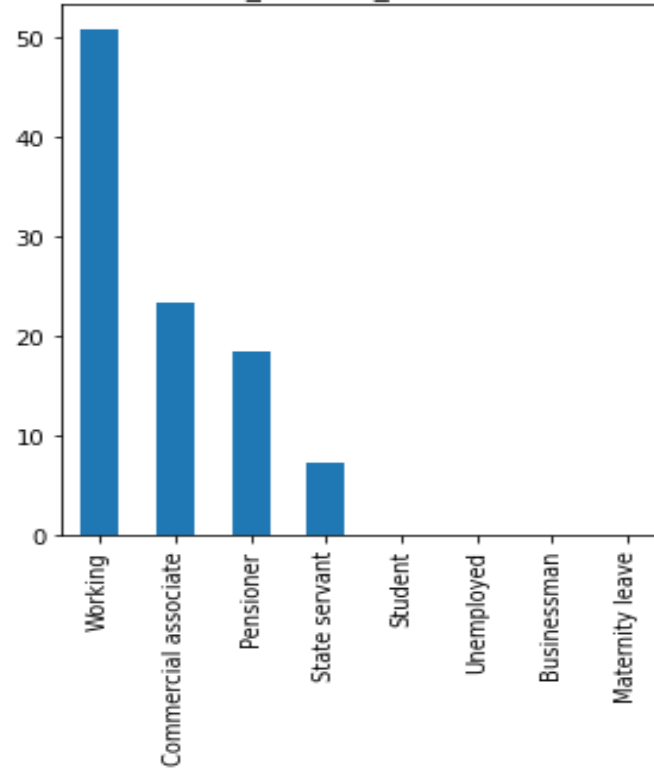
In NAME_EDUCATION_TYPE Secondary/Secondary special are more likely to have payment difficulties followed by Higher education. But Secondary/Secondary special apply more for loan followed by Higher education.
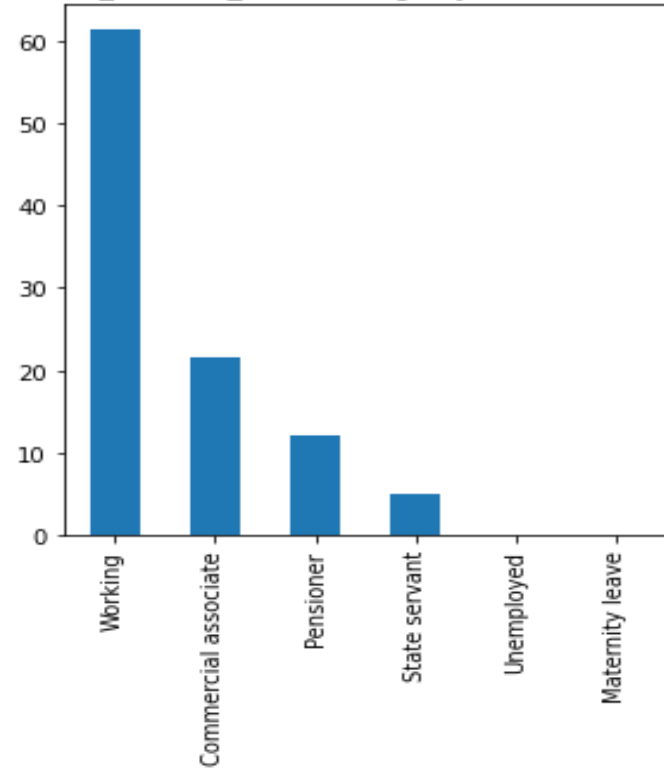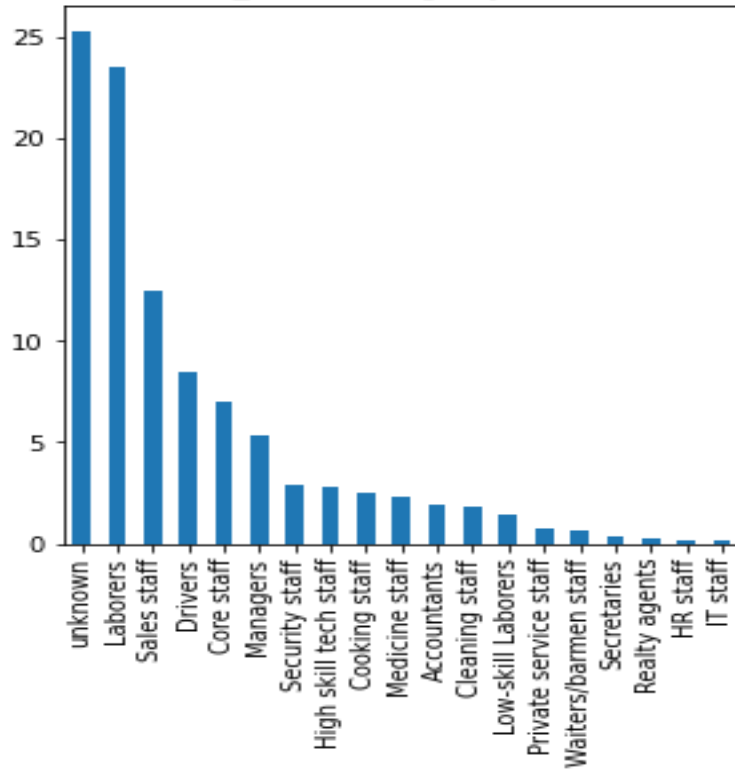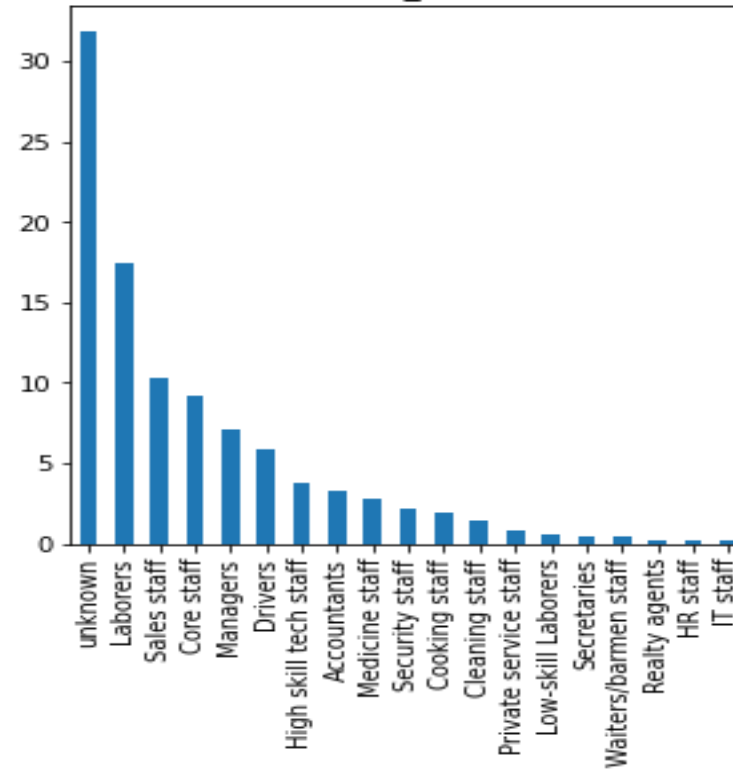
In NAME_INCOME_TYPE Working class are more likely to have payment difficulties followed by Commercial associate. But Working class apply more for loan followed by Commercial associate then Pensioner.

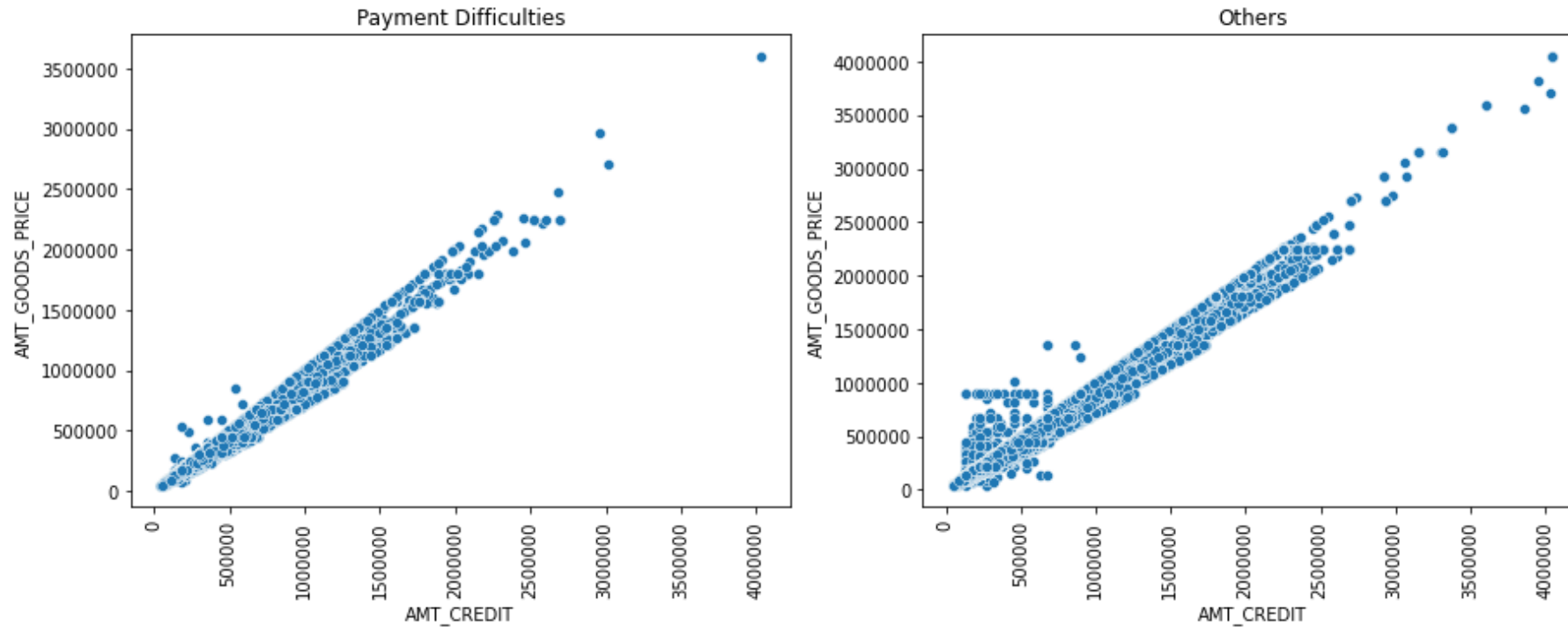OCCUPATION_TYPE Facing Payment Difiiiculties

OCCUPATION_TYPE Others

In OCCUPATION_TYPE Laborers are more likely to have payment difficulties. But Laborers apply more for loan followed by Sales staff, Core staff etc.

Bivariate Analysis
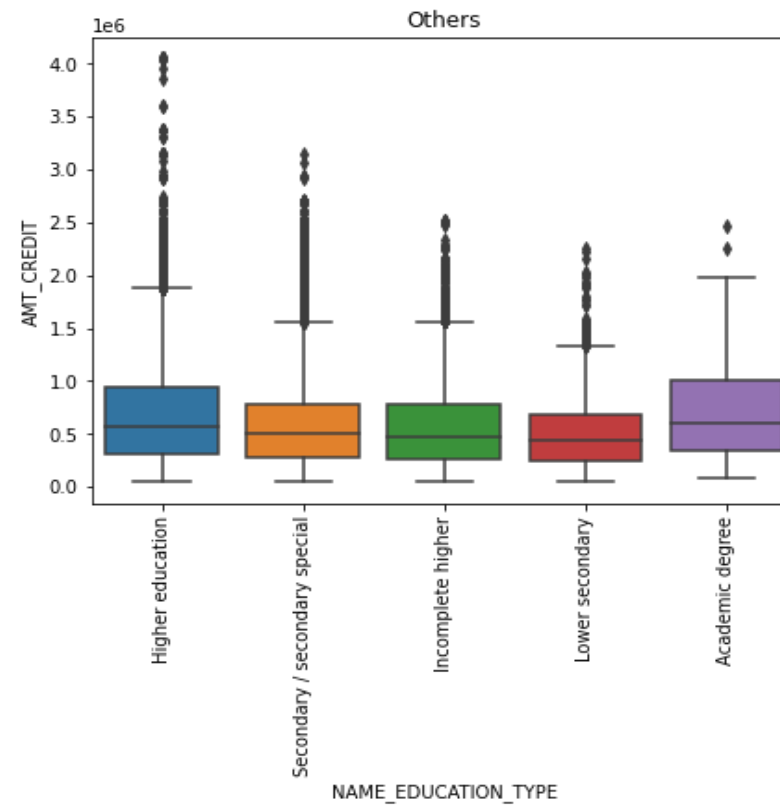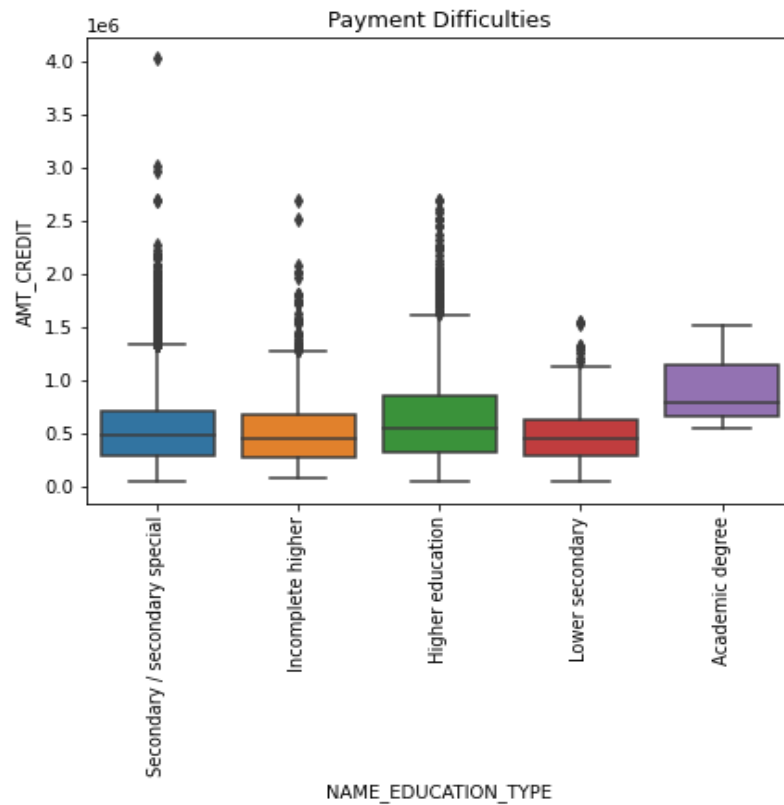
Numerical Vs Numerical Category

AMT_CREDIT and AMT_GOODS_PRICE show linear relation i.e if
the AMT_GOODS_PRICE increases the AMT_CREDIT also increases
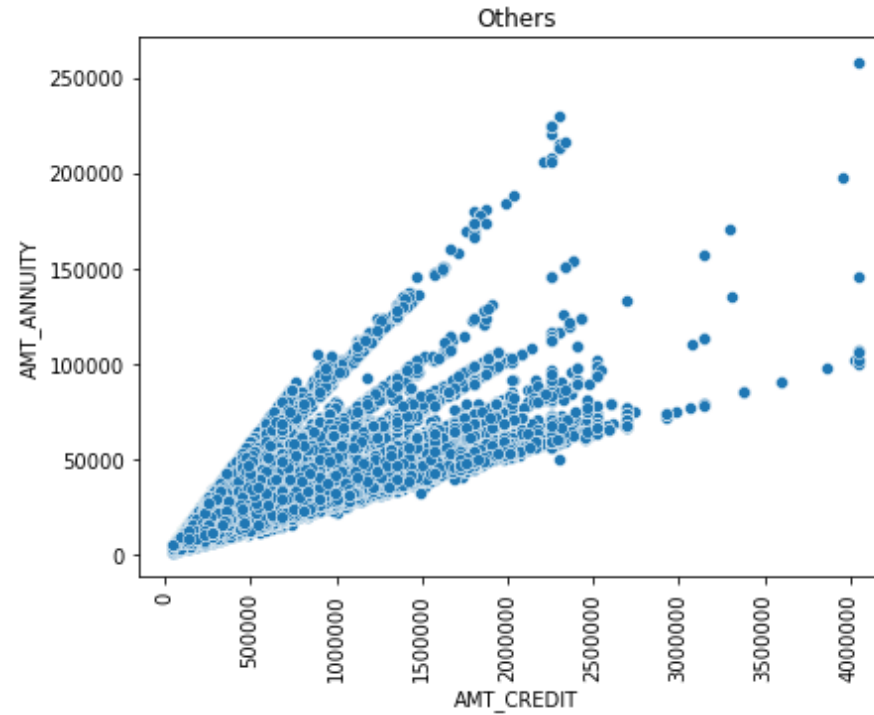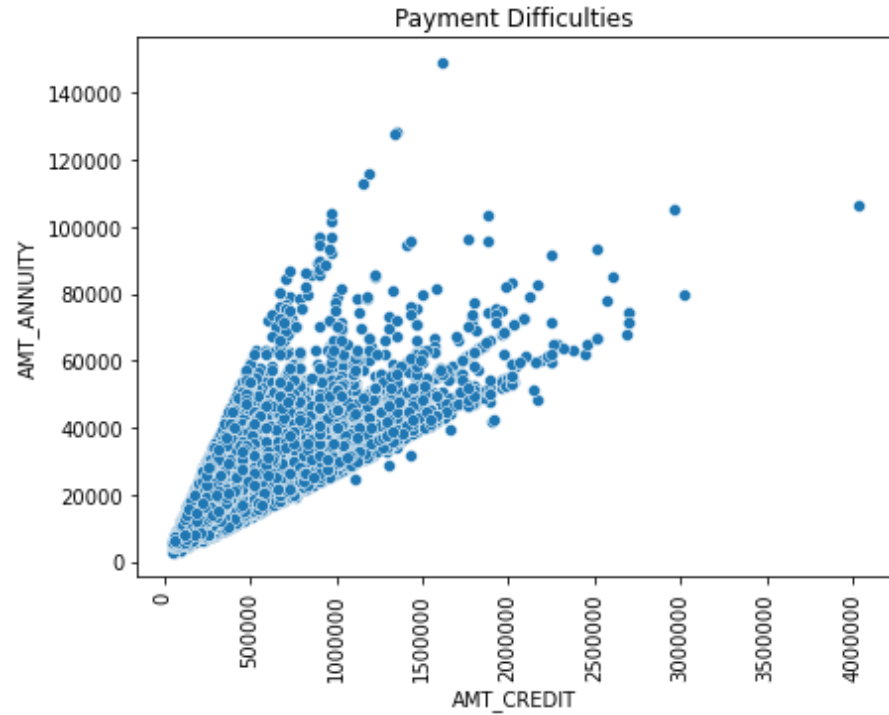
AMT_ANNUITY and AMT_GOODS_PRICE show almost linear relation i.e if the AMT_ANNUITY increases the AMT_GOODS_PRICE also increases
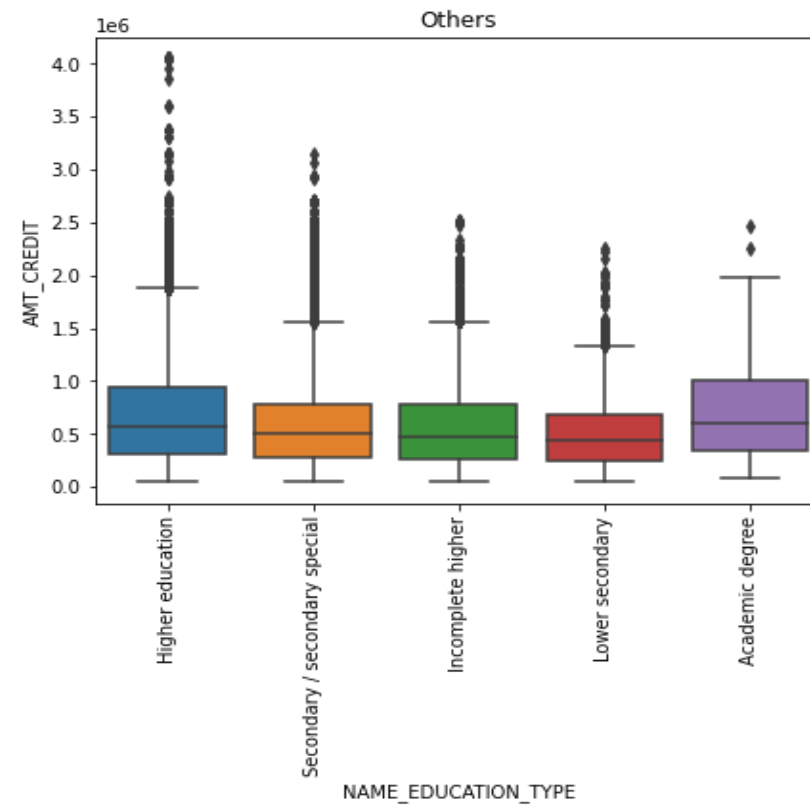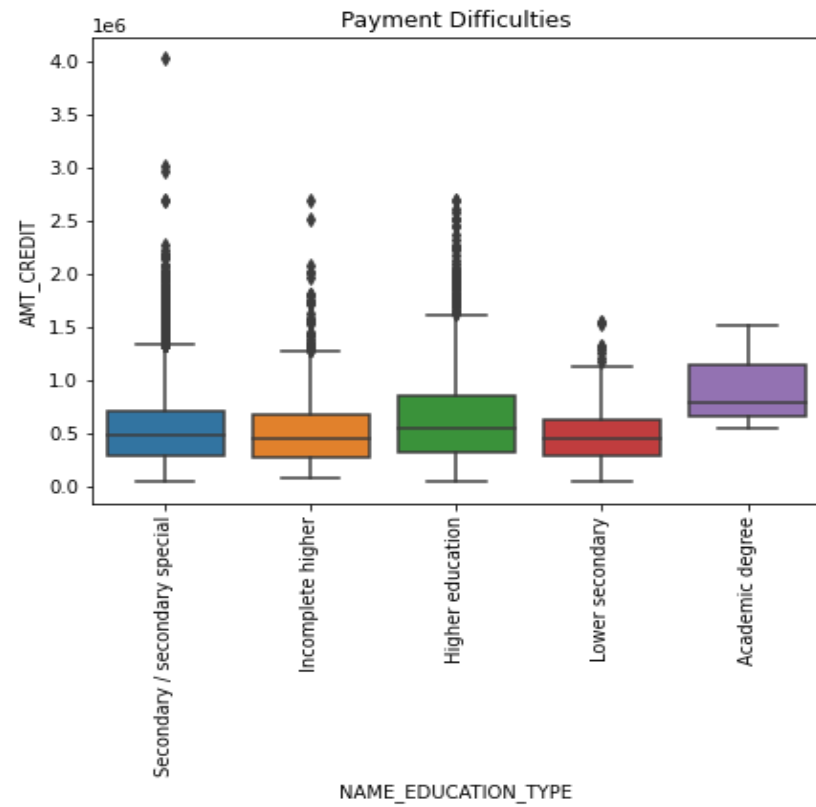
Categorical Vs Numerical Category

In NAME_EDUCATION_TYPE Academic degree category faces difficulties in getting loan and Higher education and Academic degree gets more AMT_CREDIT.
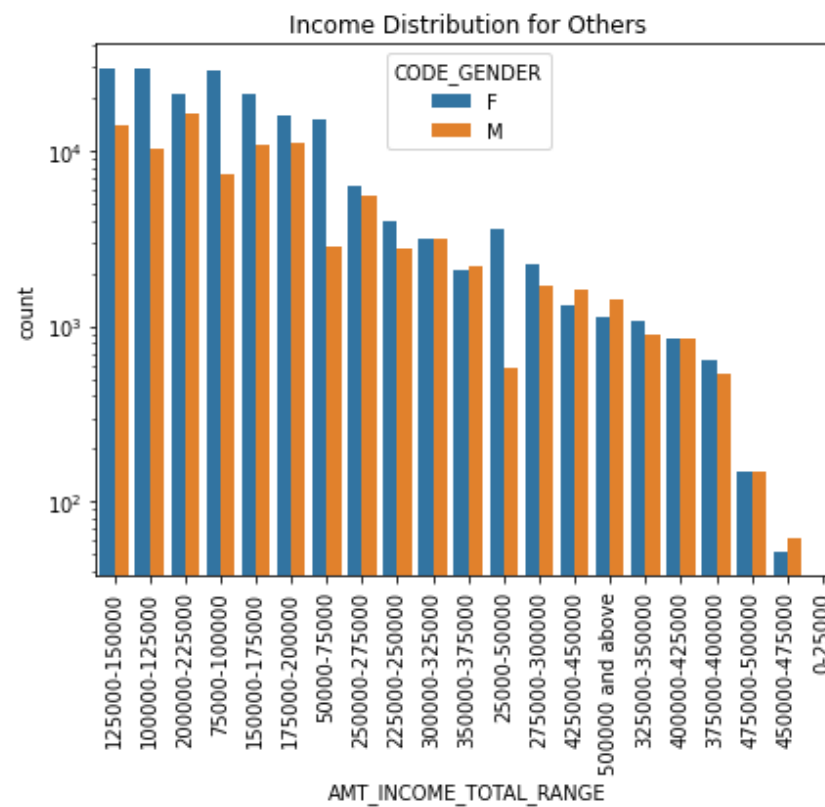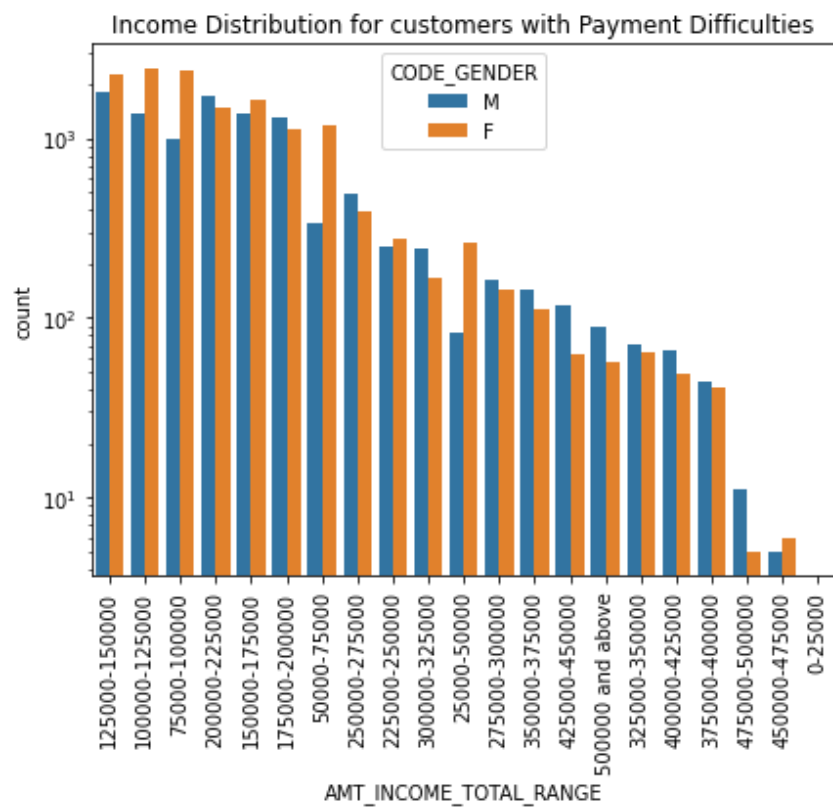
AMT_INCOME_TOTAL_RANGE in the range of 50000 to 150000 have more face difficulties in loan repayment and Female apply more for loan as compared to Male
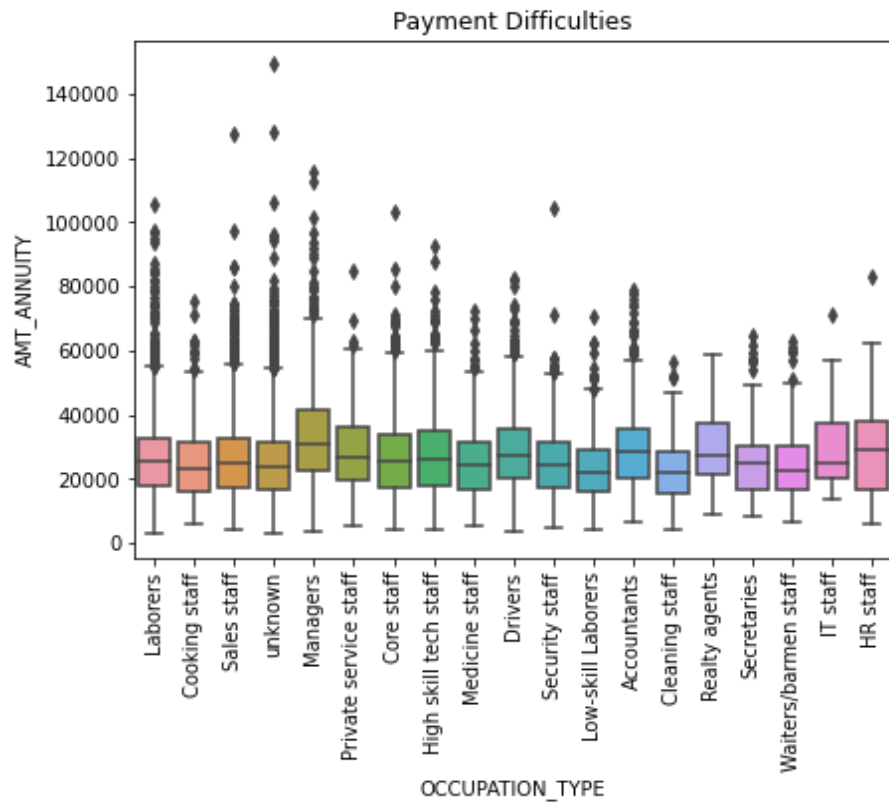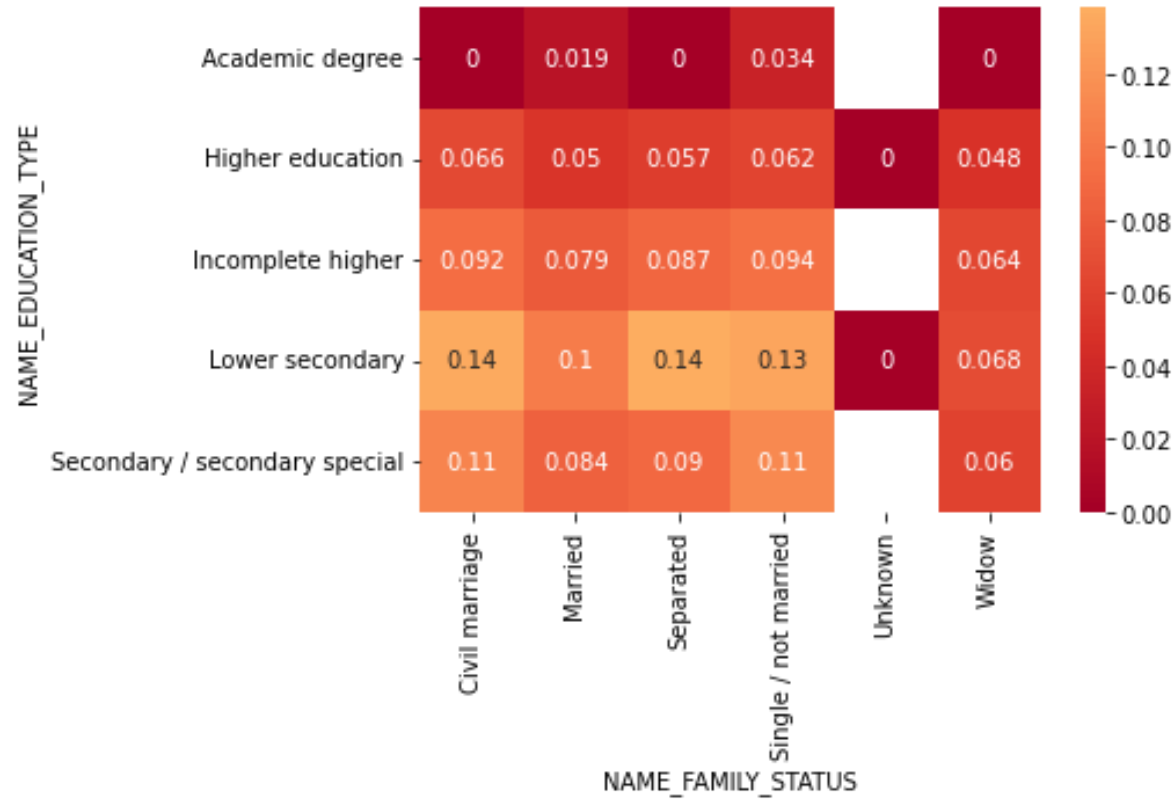
In NAME_EDUCATION_TYPE Academic degree category faces difficulties in getting loan and Higher education and Academic degree gets more AMT_CREDIT.
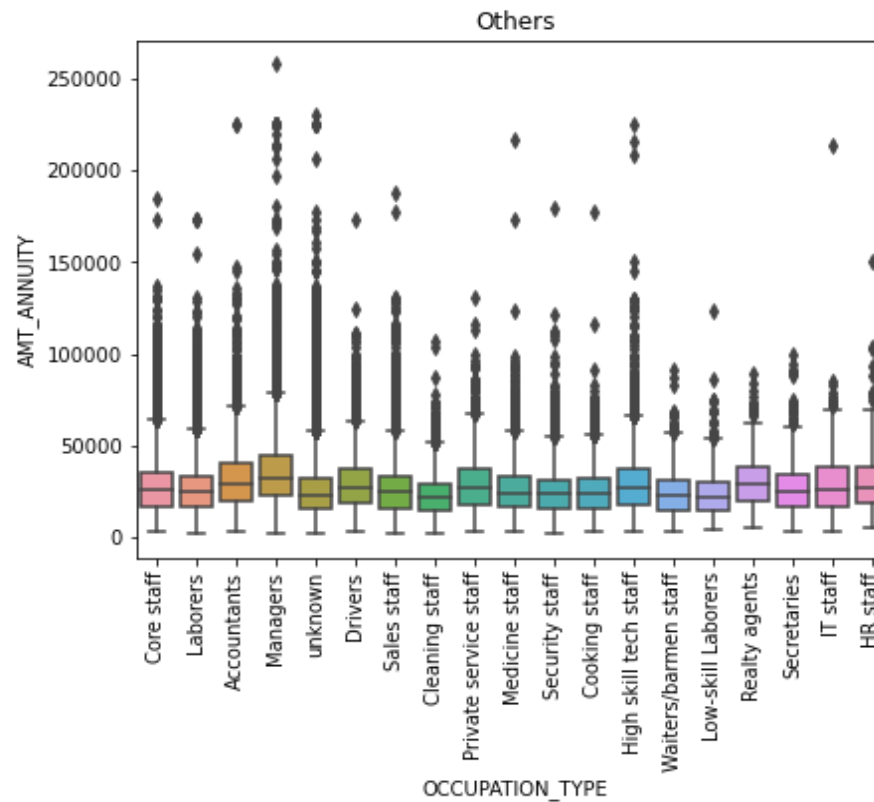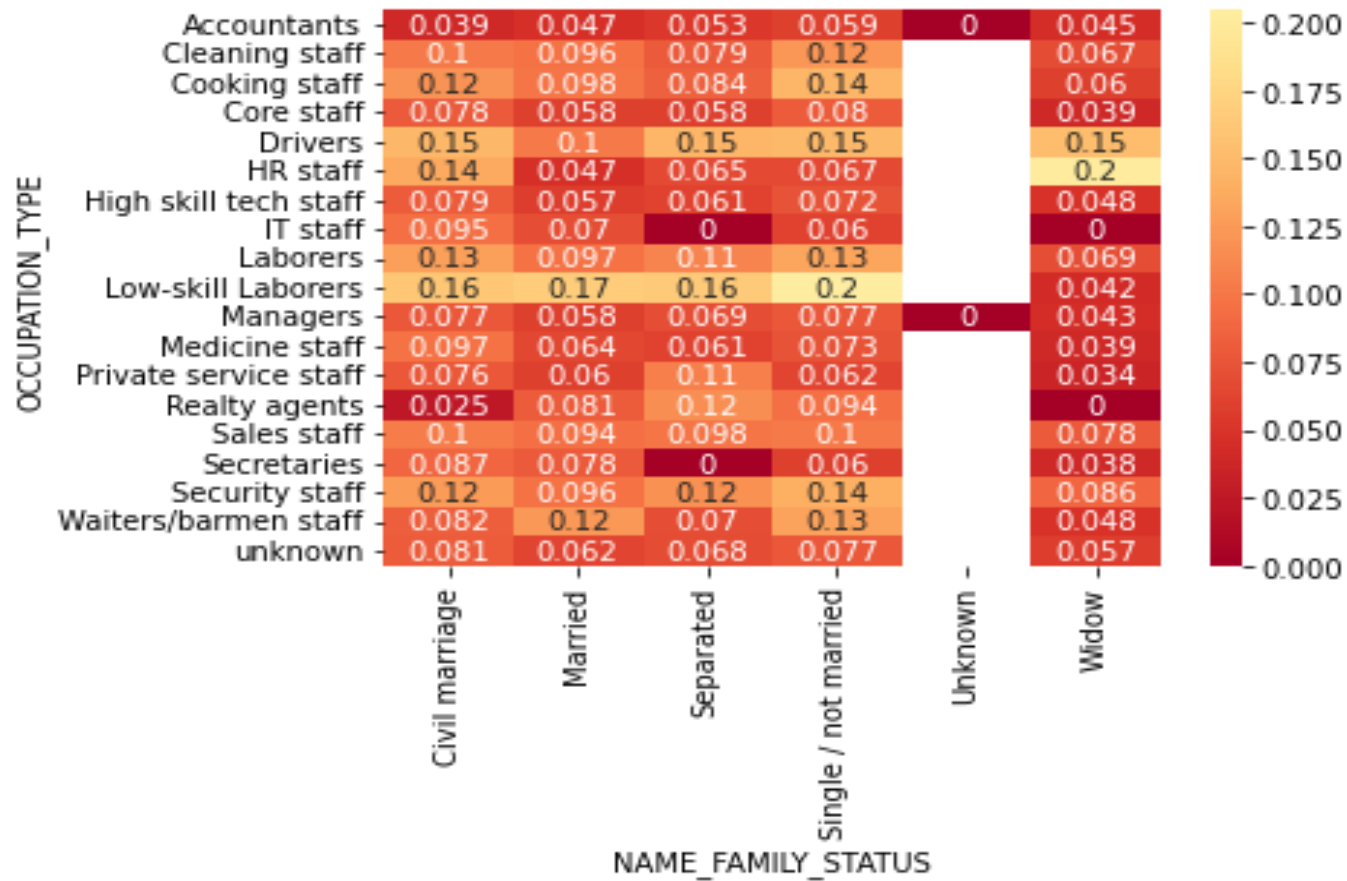
In OCCUPATION_TYPE Managers, High skill tech staff and Realty agents pays loan on time, Managers also face difficulties in loan repayments.
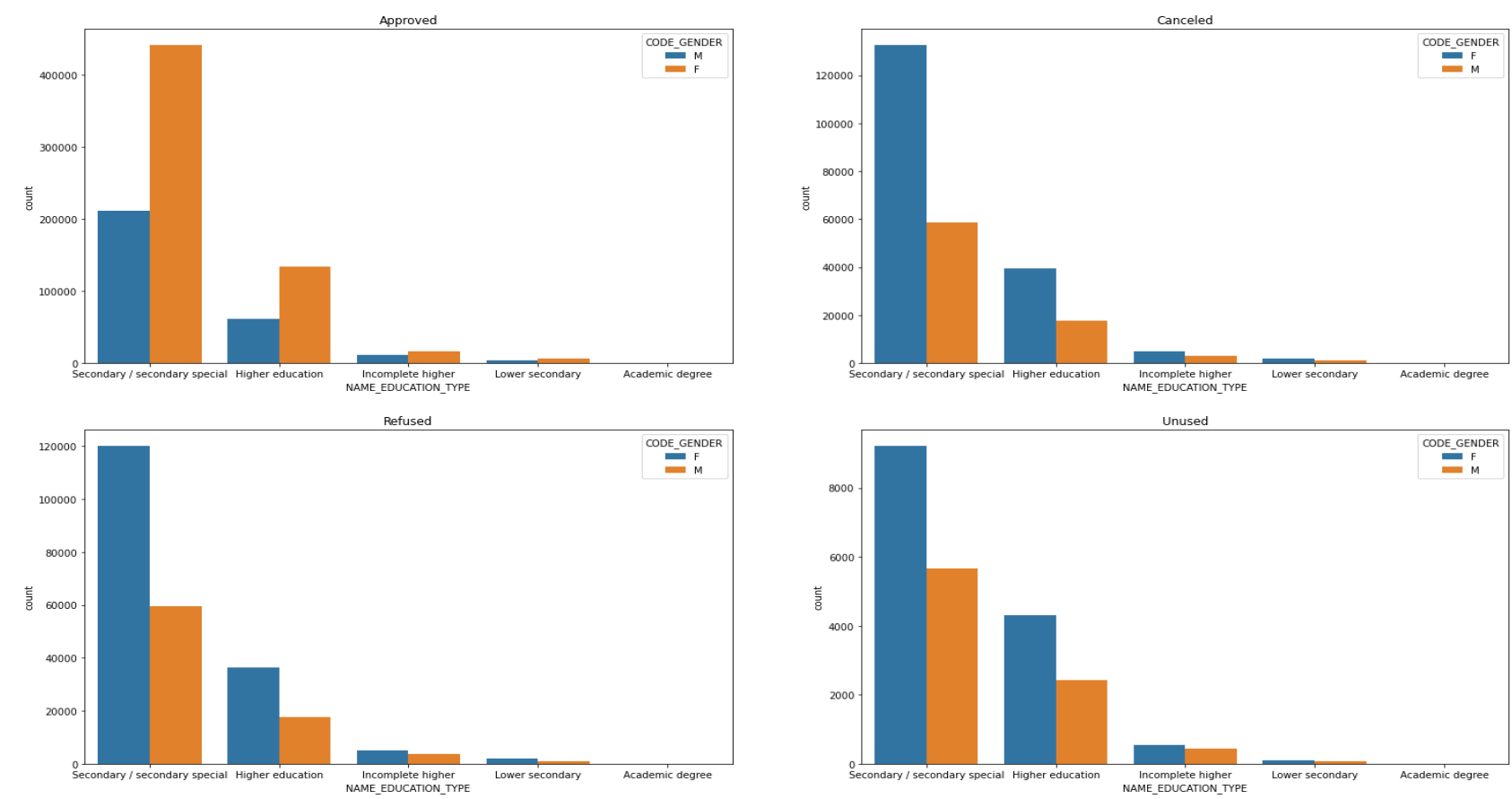
Multivariate Analysis

NAME_EDUCATION_TYPE,NAME_FAMILY_STATUS and TARGET
having lower value i.e 0 or dark color face less difficulties in
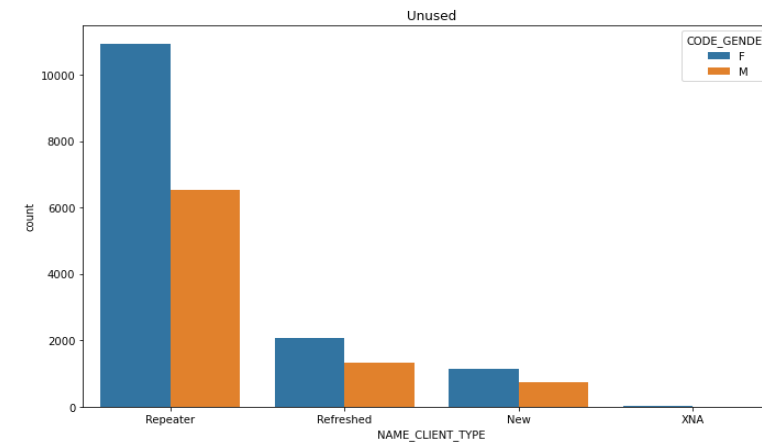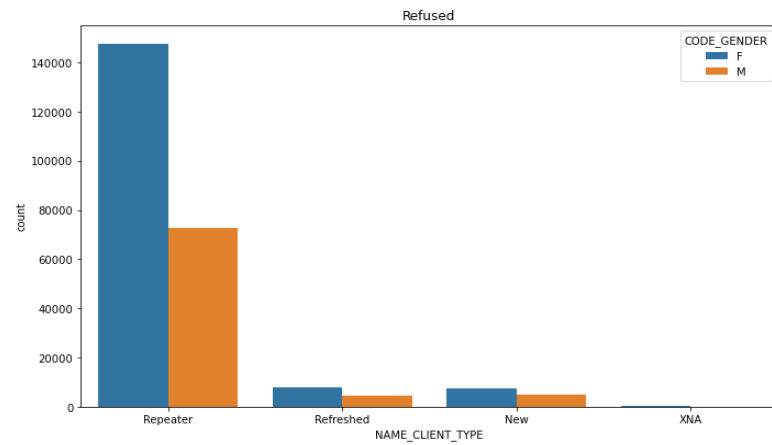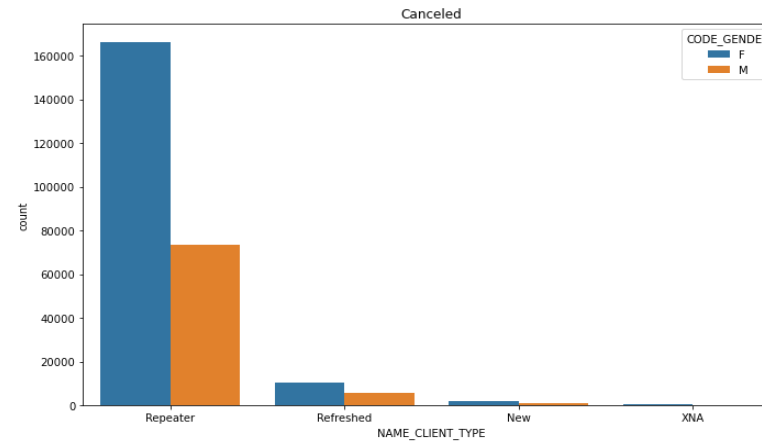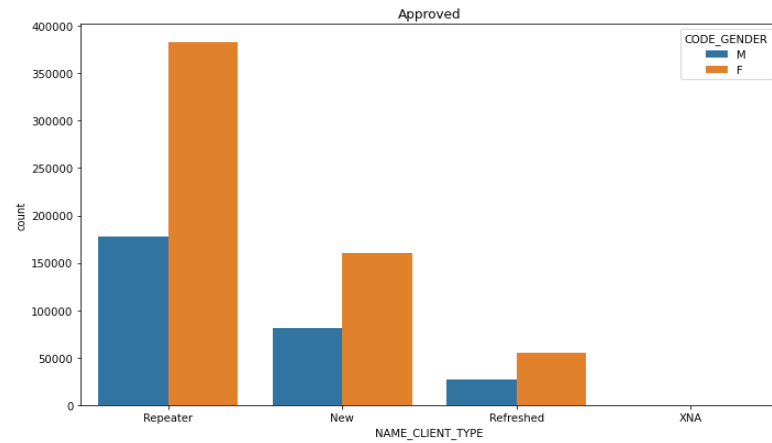loan repayment and getting loan.

OCCUPATION_TYPE,NAME_FAMILY_STATUS and TARGET having lower value i.e 0 or dark color face less difficulties in loan repaymant and getting loan.

Merged Data Analysis

Approved loans for Female having secondary/secondary special education type

In NAME_CLIENT_TYPE Female Repeater, New, Refreshed Client loans are approved more.

Recommended Category To Which Loan Can Be Credited

1. Female Client
2. Secondary/Secondary special and Higher Education
3. Manager, Accountants, High skill tech staff

Risky Groups

1. Male, Married
2. Lower secondary educated