



CREDIT EDA **CASE STUDY**

Problem Statement

01

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:

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

02

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.



APPROACH

The approach followed in this CASE STUDY for this kind of huge dataset was as follows

Cleaning The Dataset

01

Getting an idea of the shape of the Dataset

02

Checking for columns with Null Values more than 40% and dropping those columns

03

Checking for columns with null values lower than 15% and imputing them with Mean, Median, Mode accordingly

04

Checking for a few more unwanted columns and dropping those

05

Checking the Datatypes of the columns and making adjustment to it accordingly

03

Adjusting the values of few columns to +ve and also naming the columns accordingly



Analysis

01

Checking for outliers with the help of Boxplot

02

Binning for continuous columns to make them categorical columns and easier to visualize

03

Dividing the dataset on basis of Target Variable and plotting graphs to get better insights

04

Finding correlations with the help of Heatmap and recording the top 10 correlations for both target variable datasets.

05

Doing Univariate Analysis for numerical and categorical variable in the dataset for both target variables using different plots.

05

Doing Bivariate Analysis for all Numerical to Numerical, Categorical to Categorical and Numerical to Categorical using different plots.



Merging the Datasets

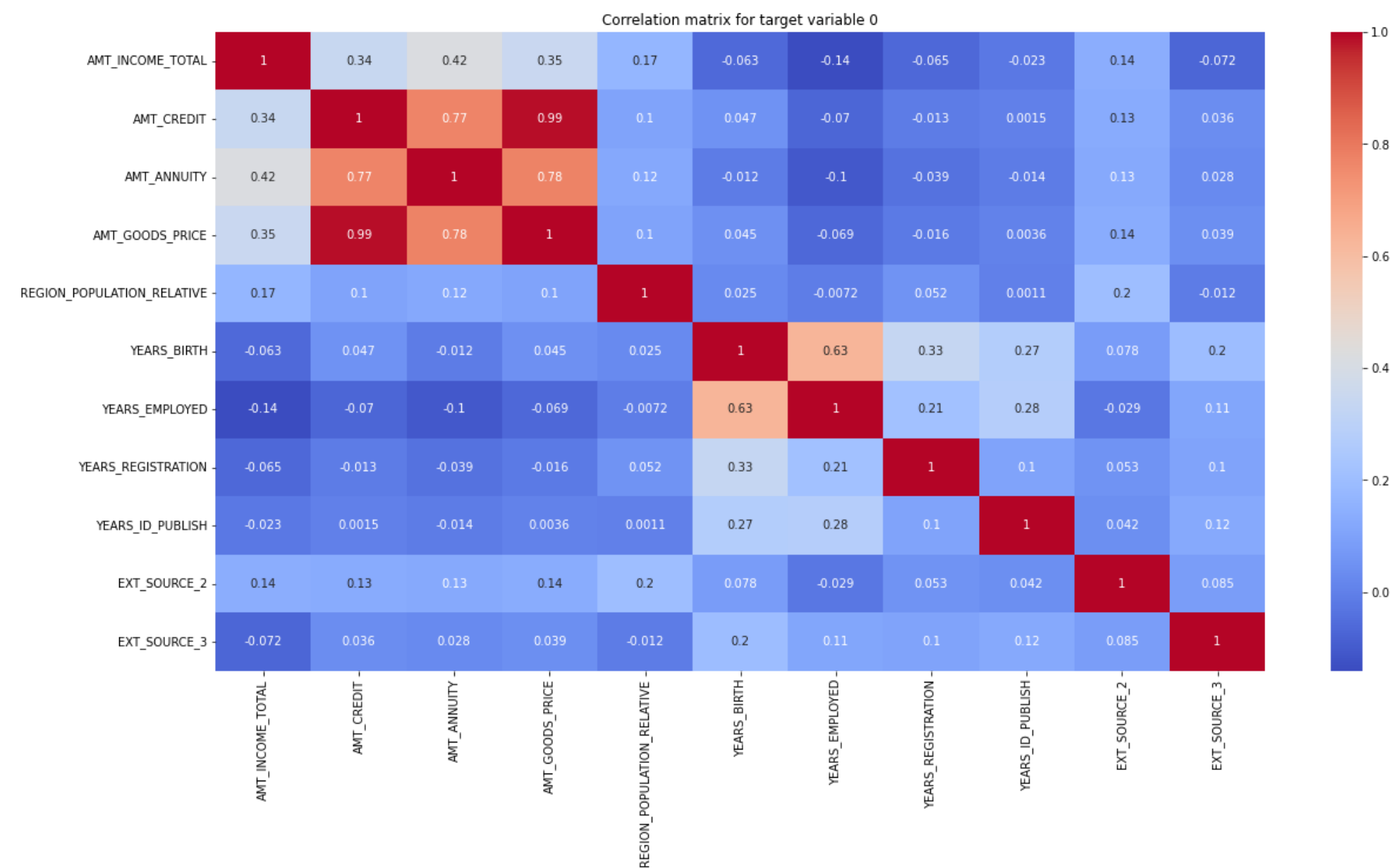
Merged the dataset with the previous application dataset and made some more analysis with the help of subplot to find the different relations between Approved, Refused, Cancelled and Unused contract status. This helped gain further insights on which sectors to target without facing more payment difficulties by the customers.



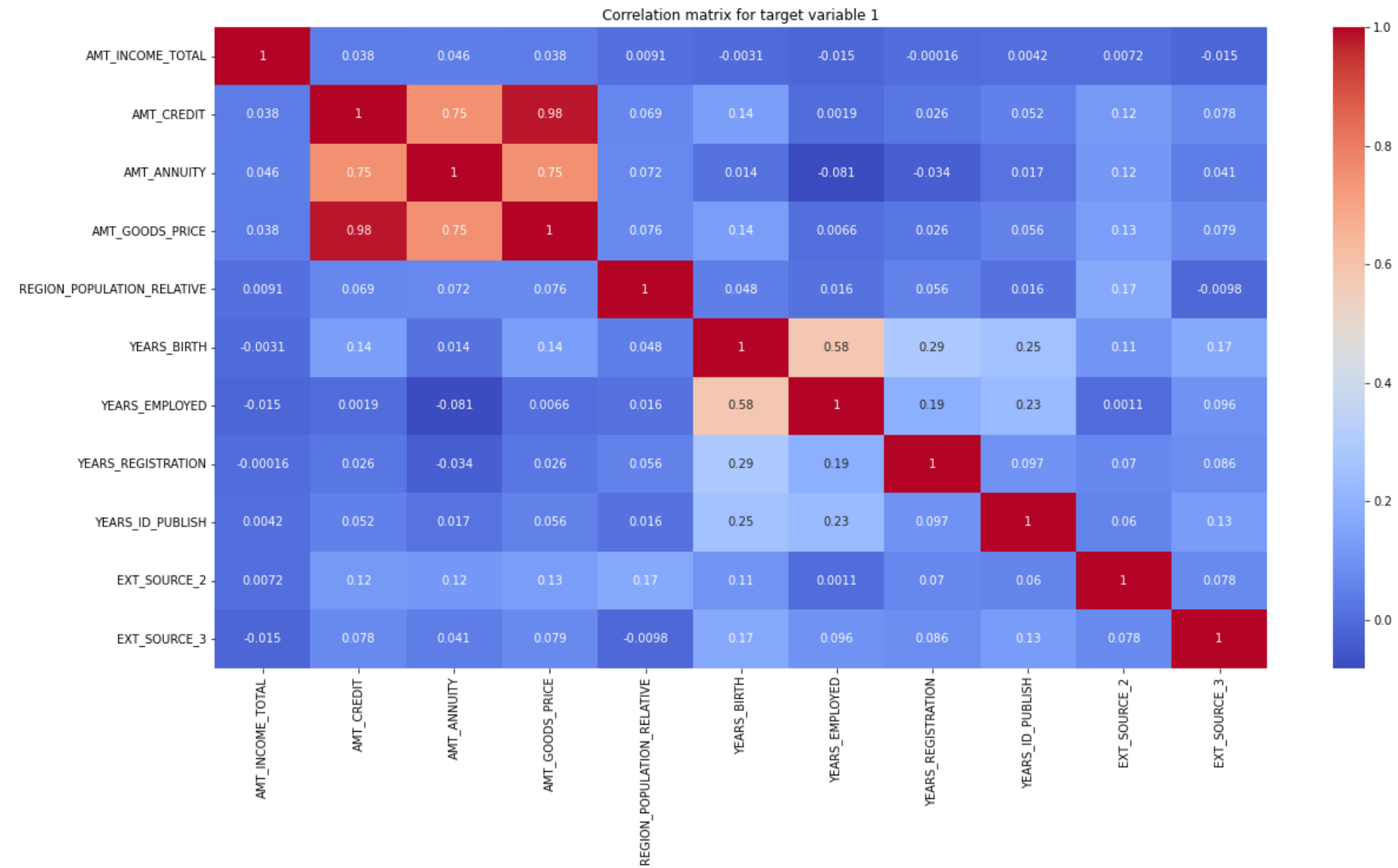
Visualization

01

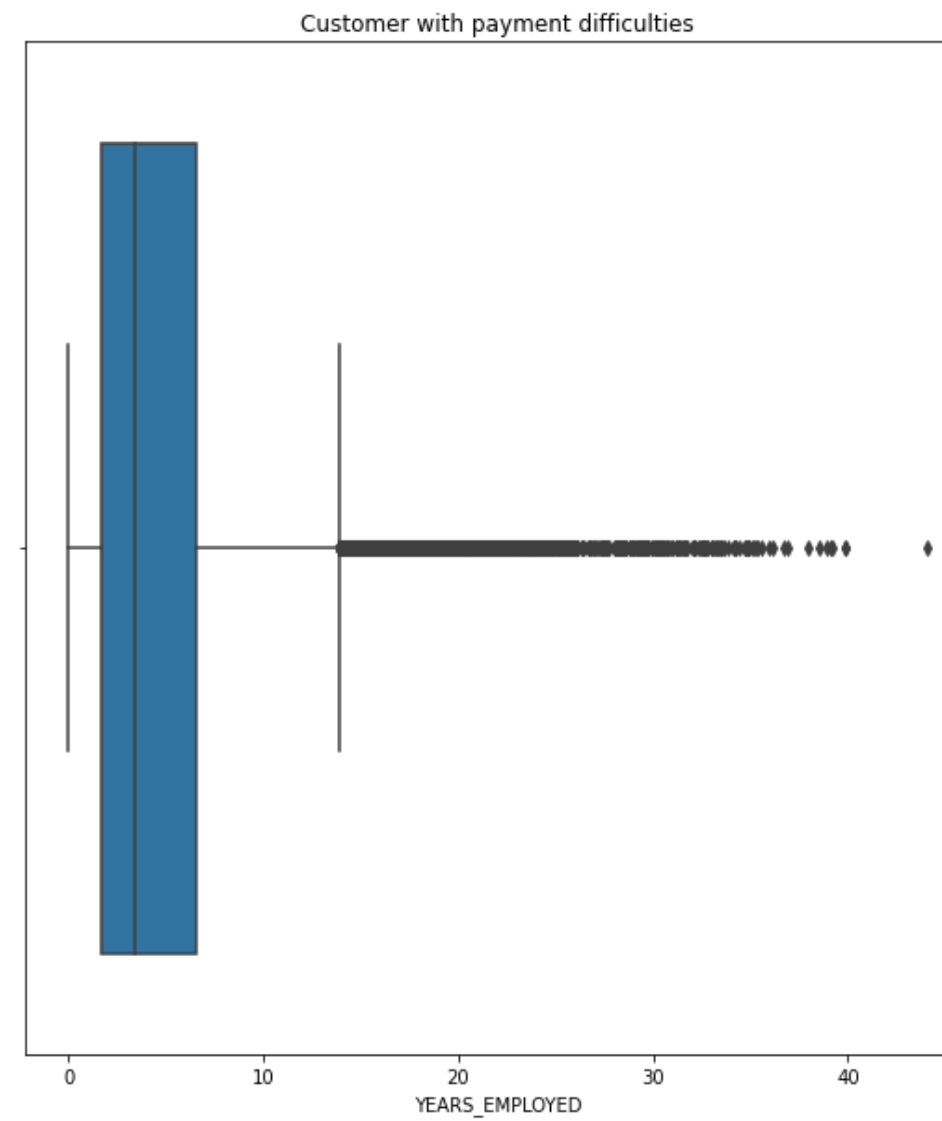
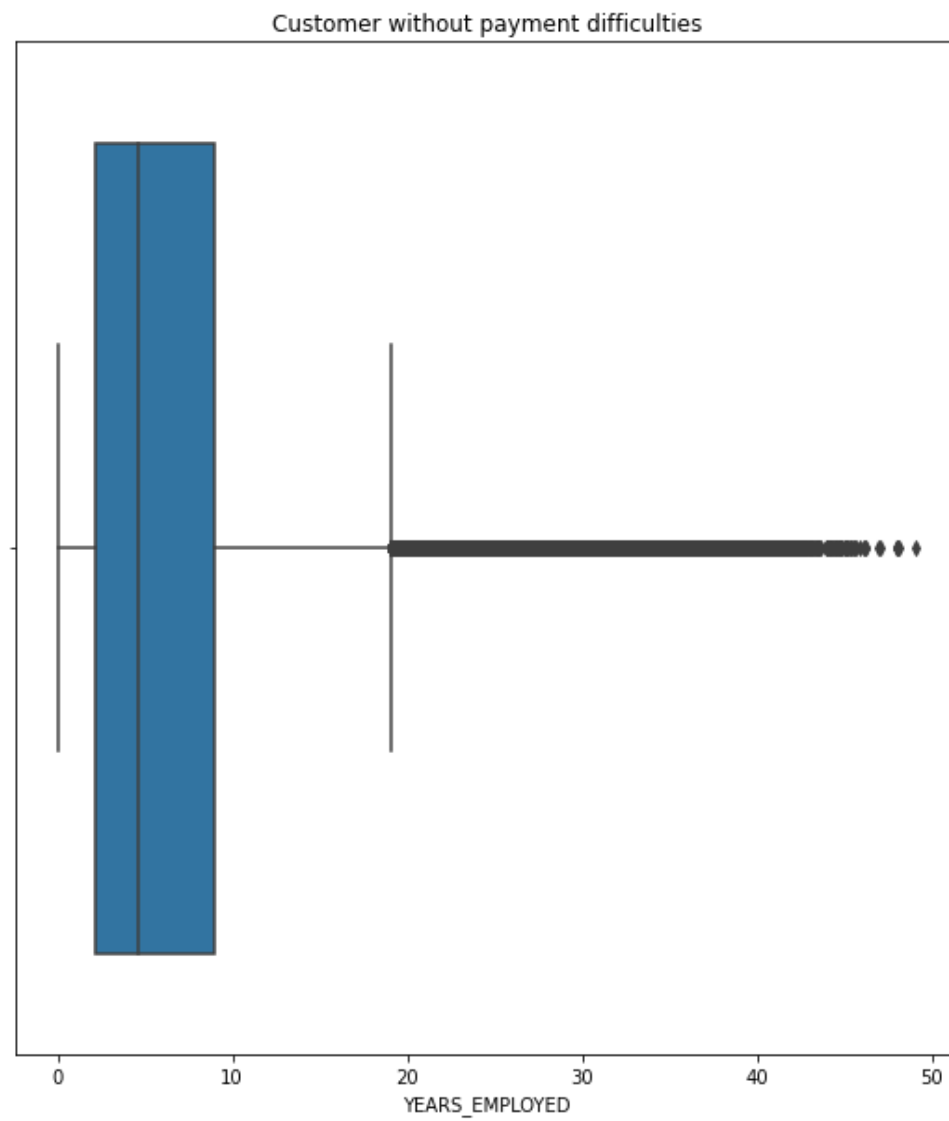
Correlation Matrix for Target Variable 0



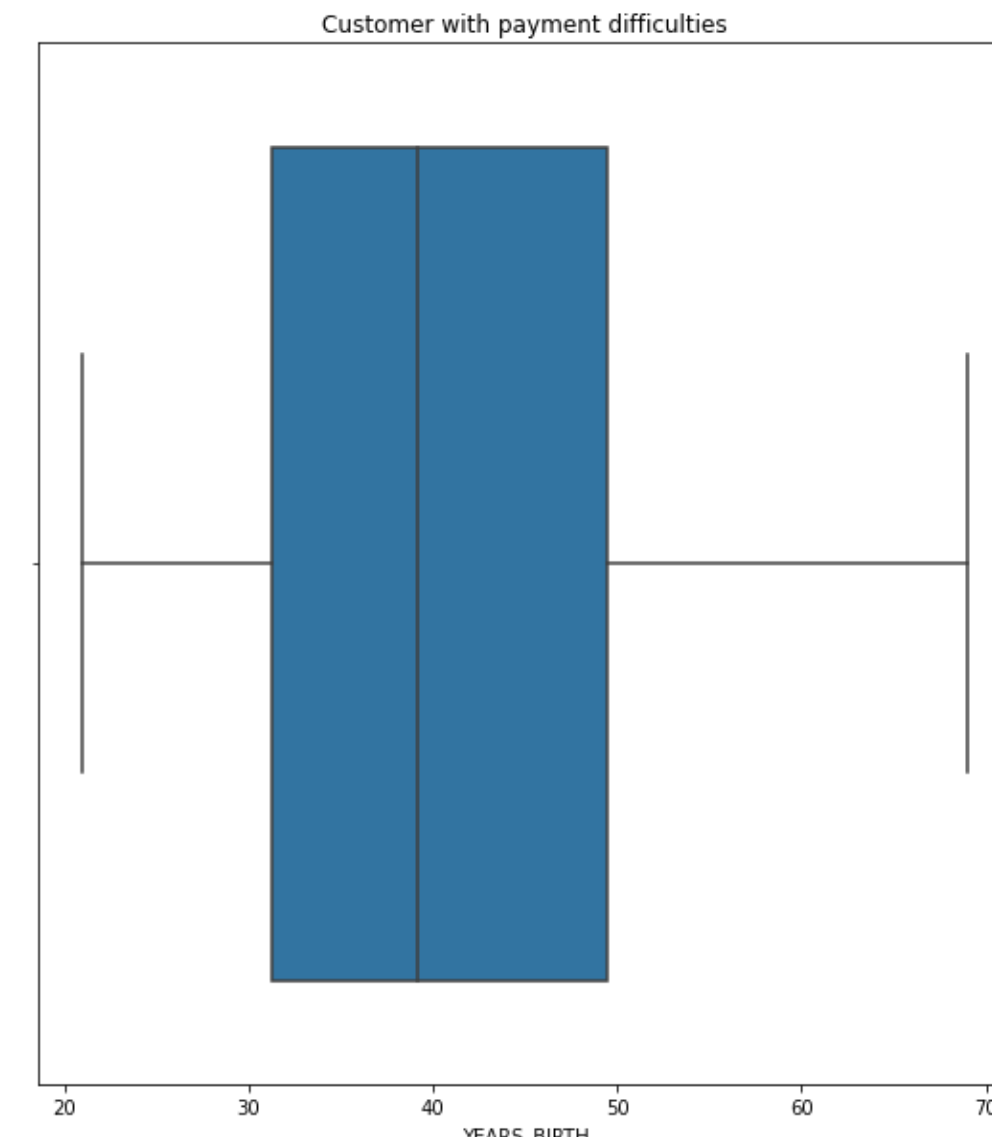
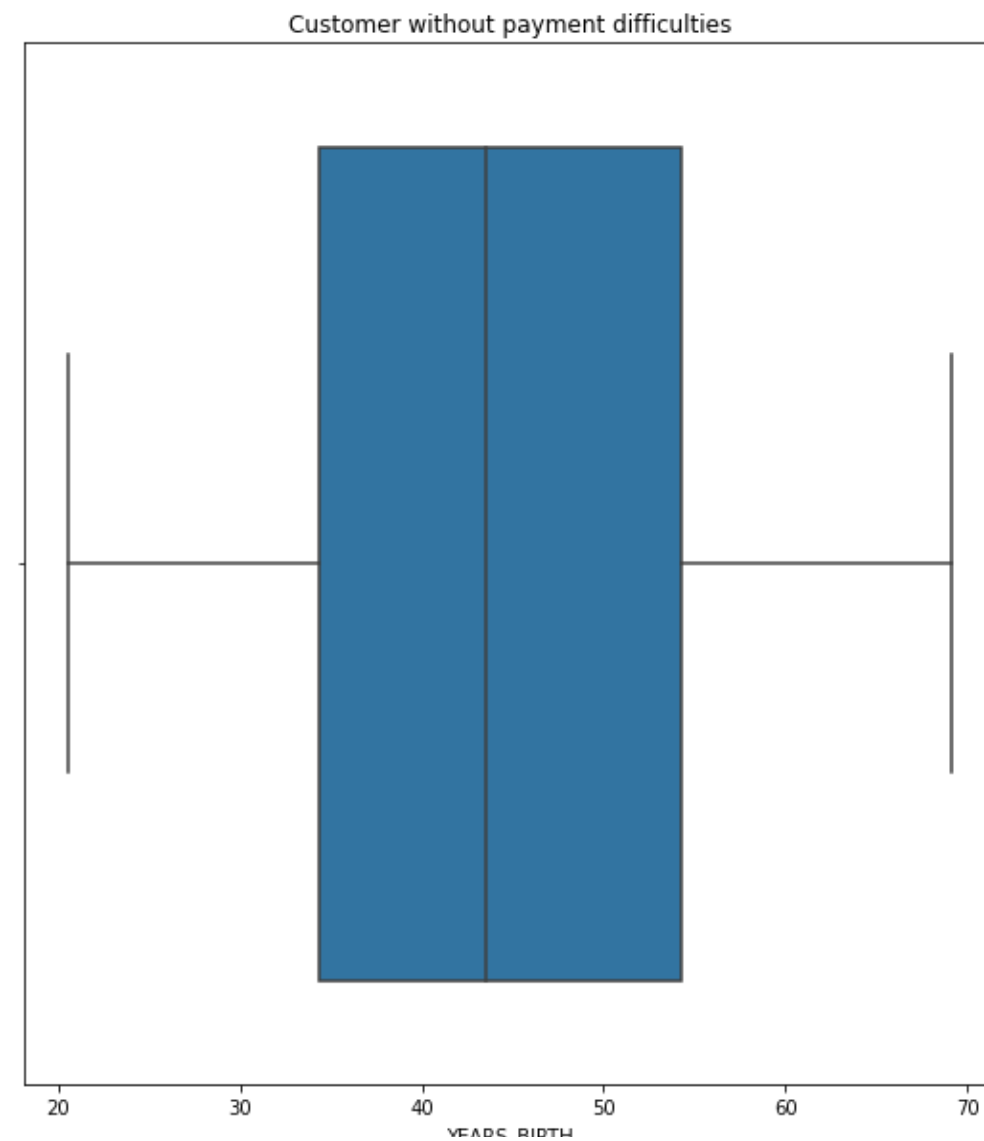
Correlation Matrix for Target Variable 1



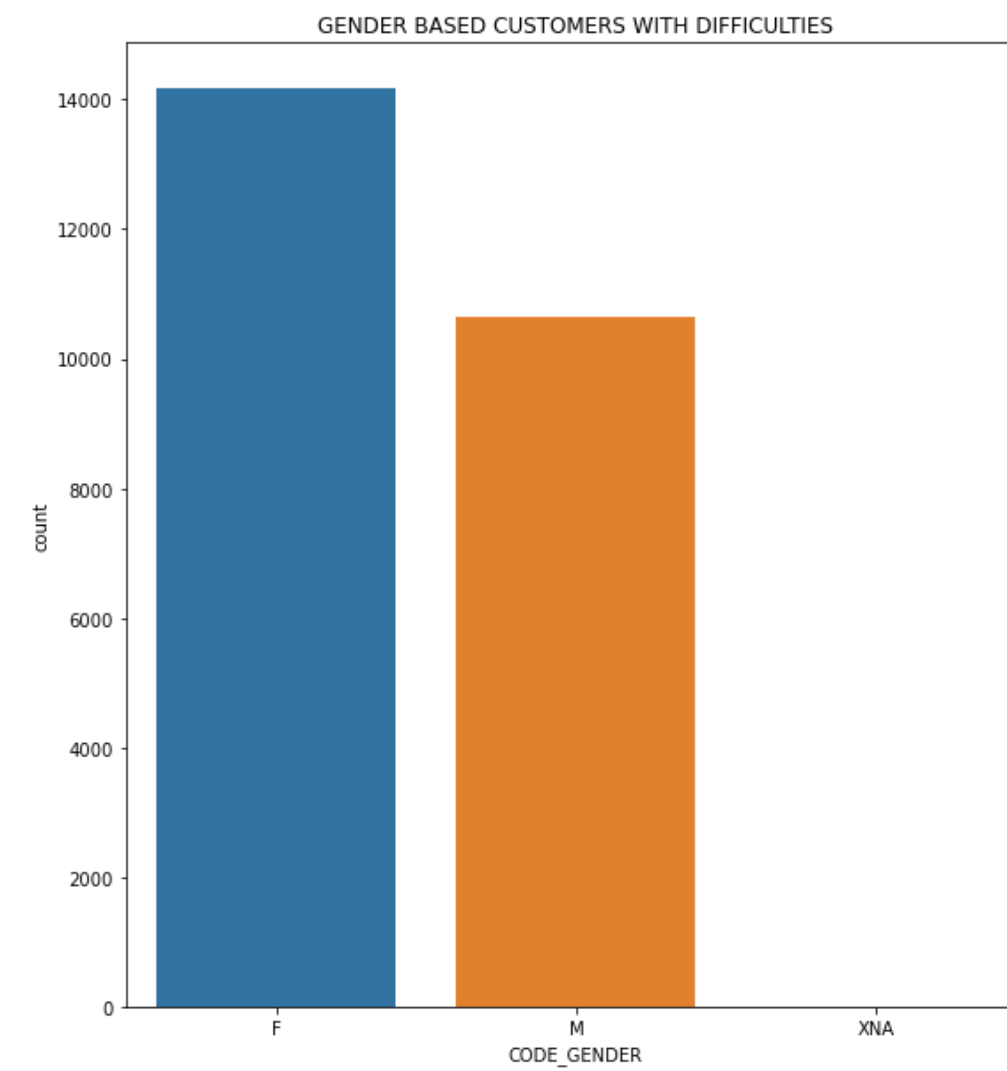
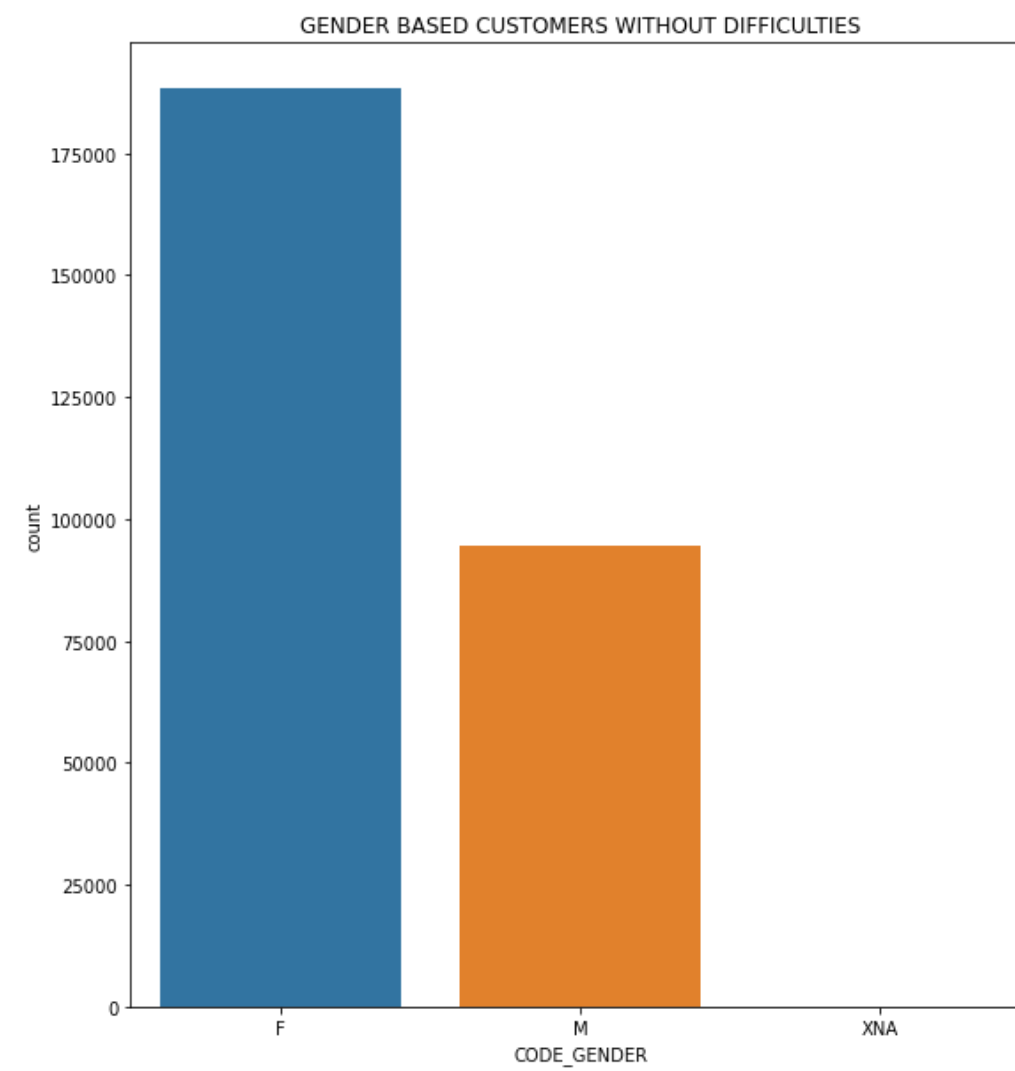
Univariate Analysis on YEARS_EMPLOYED



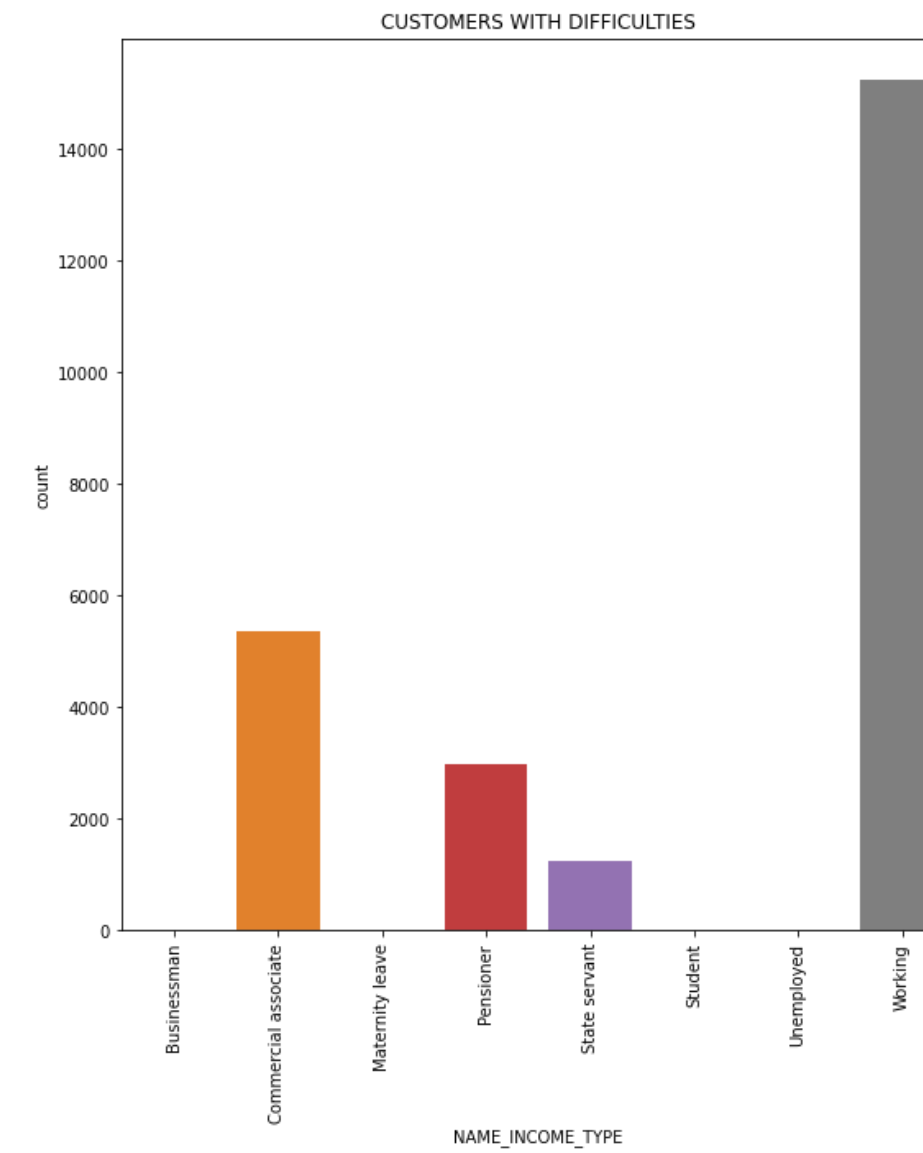
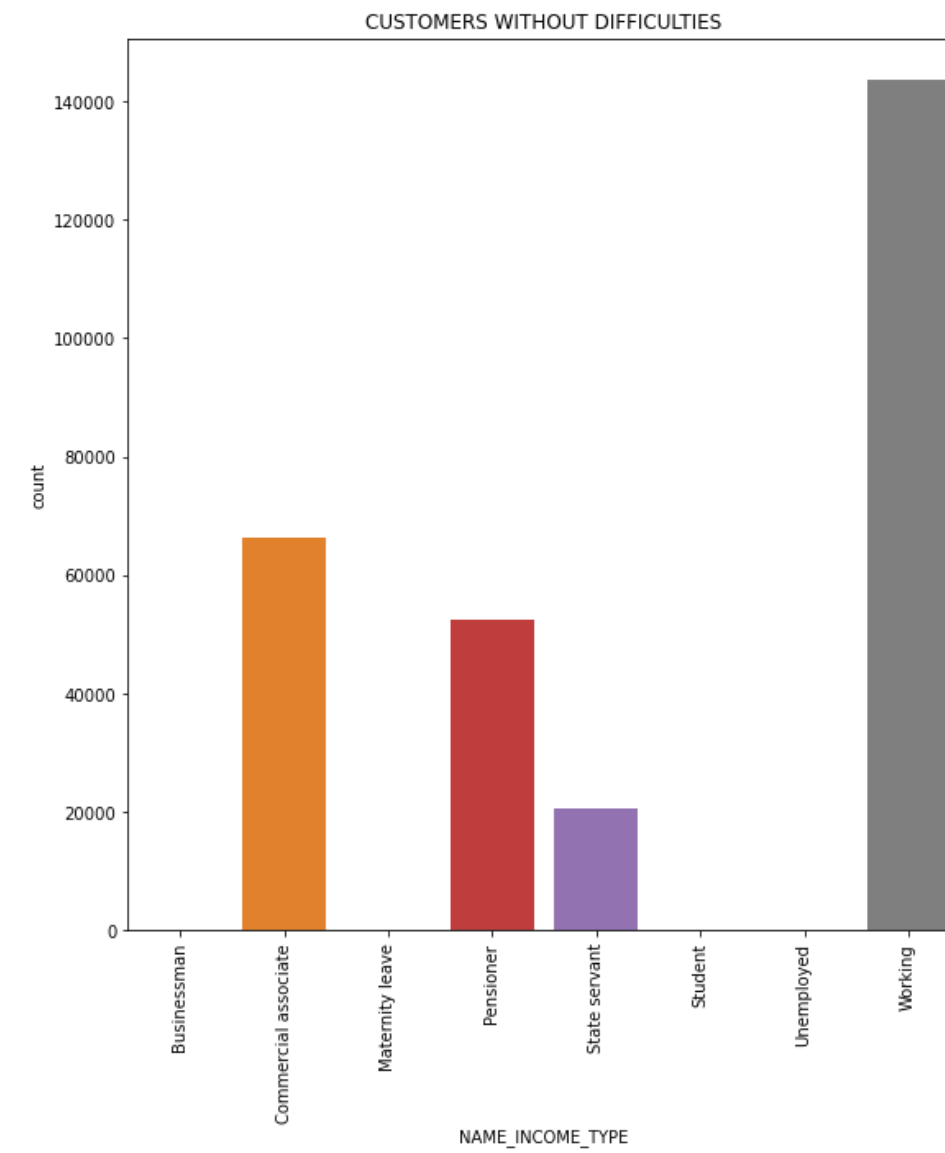
Univariate Analysis on YEARS_BIRTH



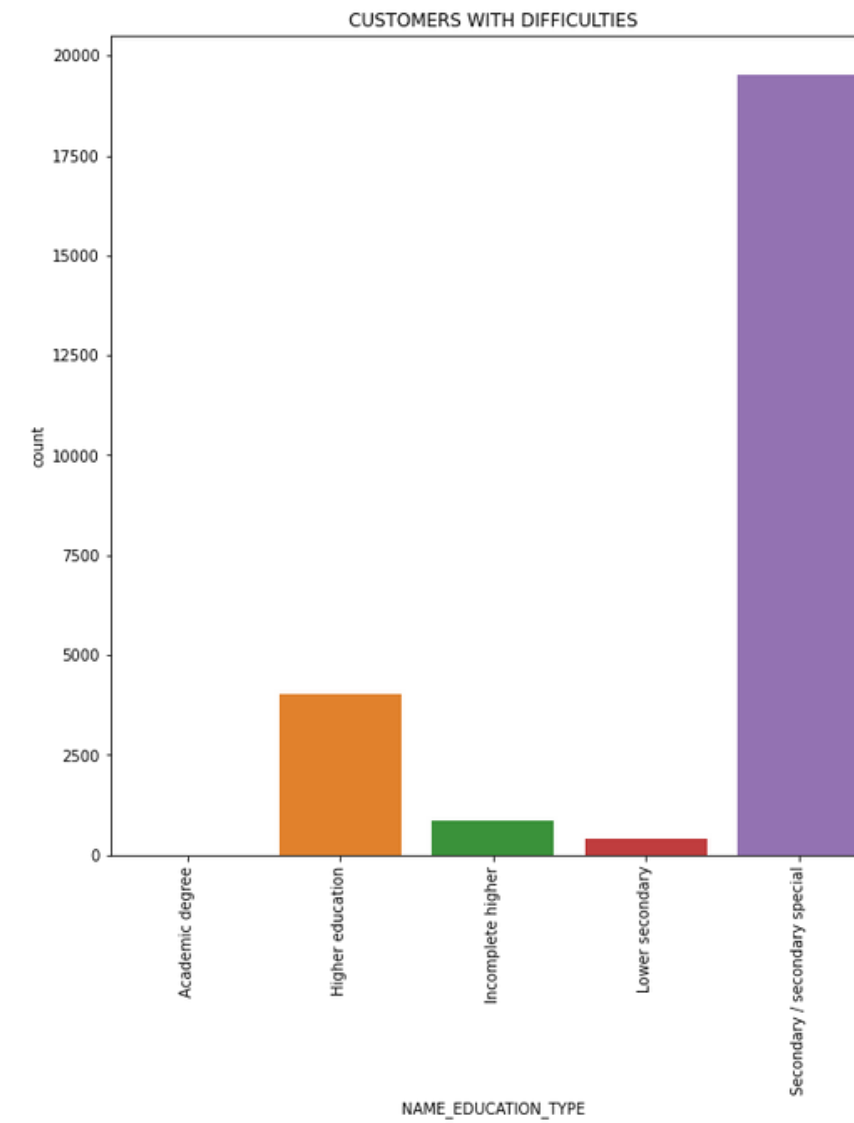
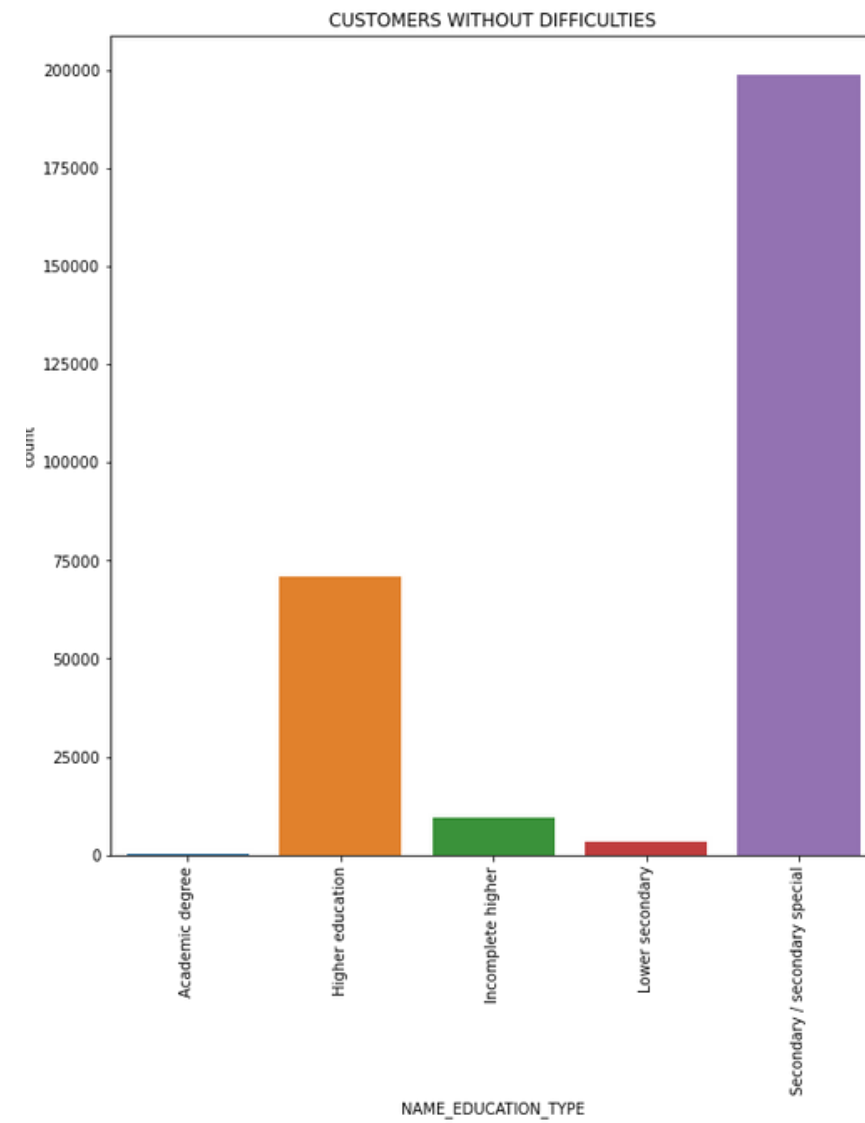
Univariate Analysis on CODE_GENDER



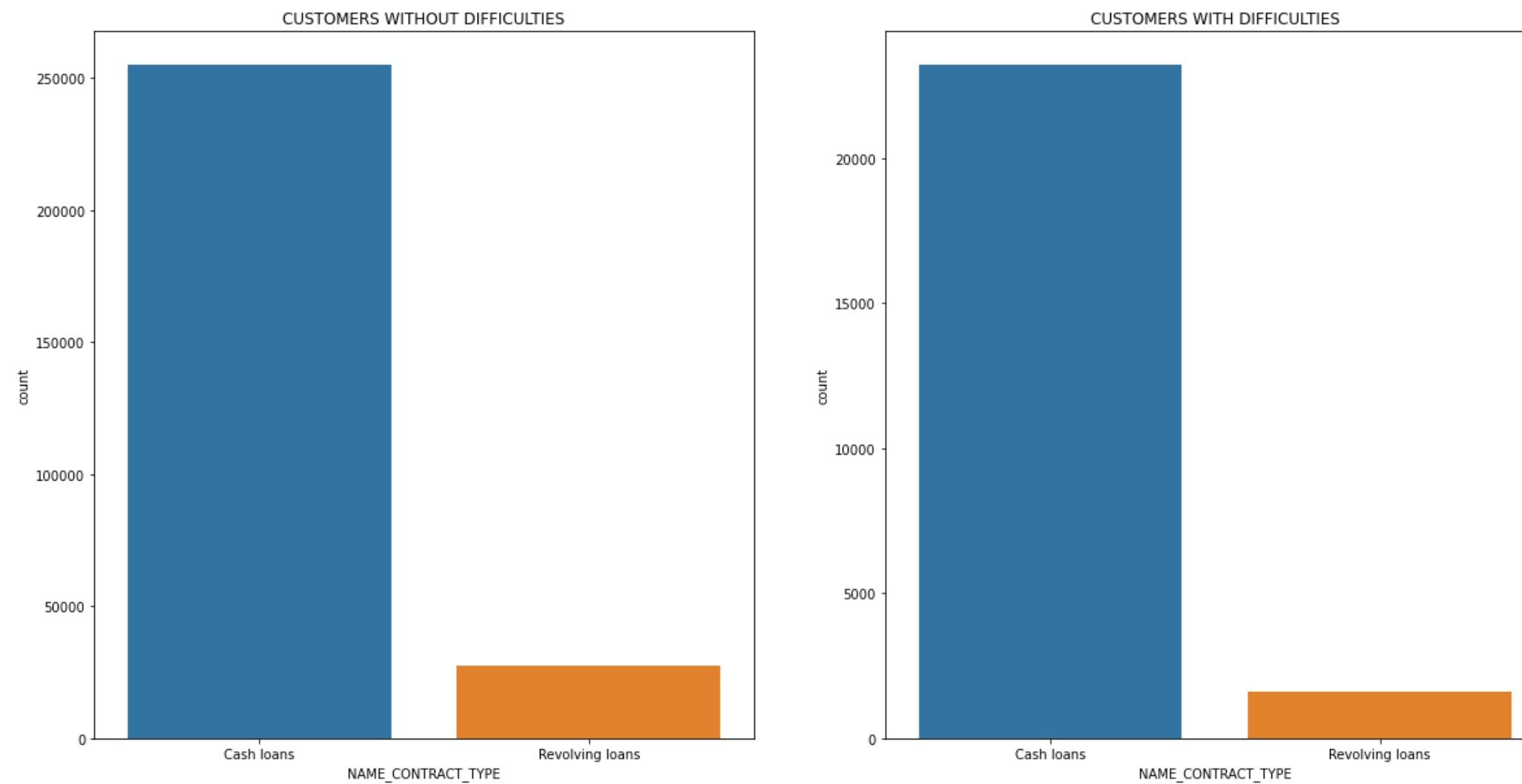
Univariate Analysis on NAME_INCOME_TYPE



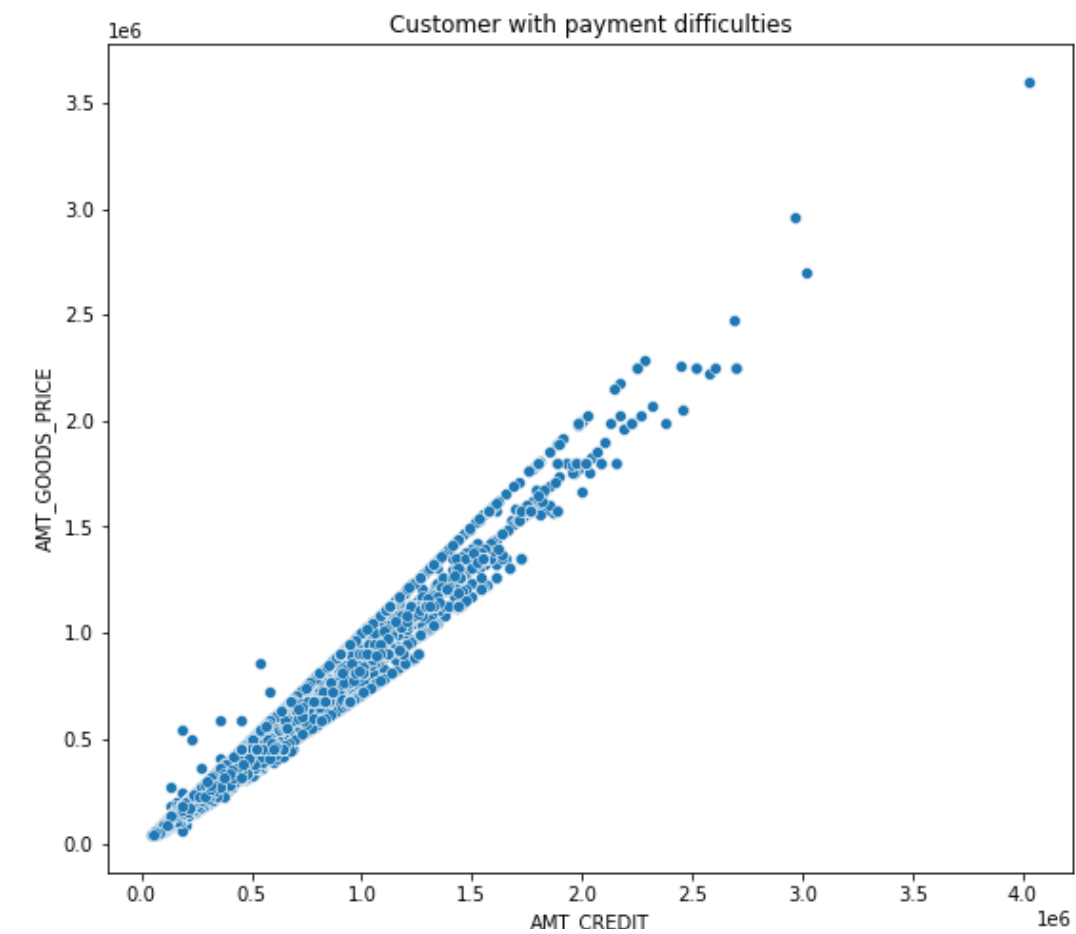
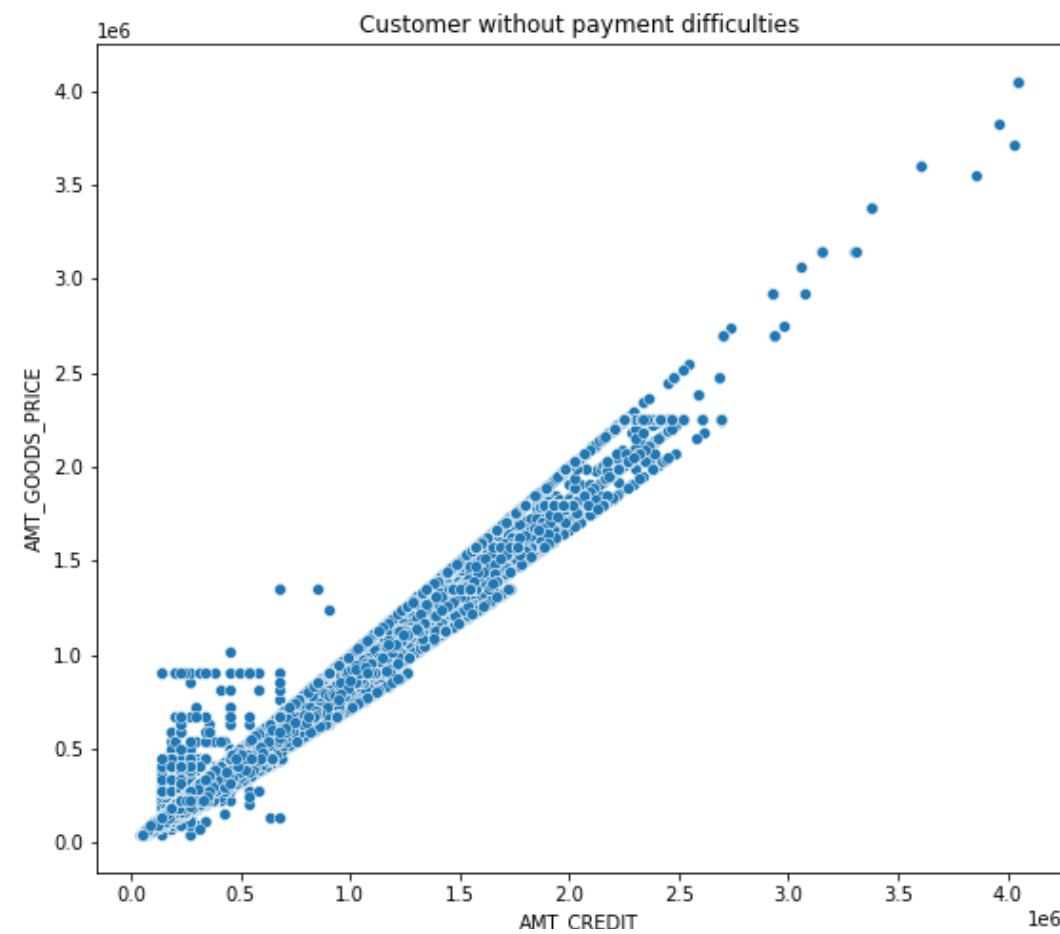
Univariate Analysis on NAME_EDUCATION_TYPE



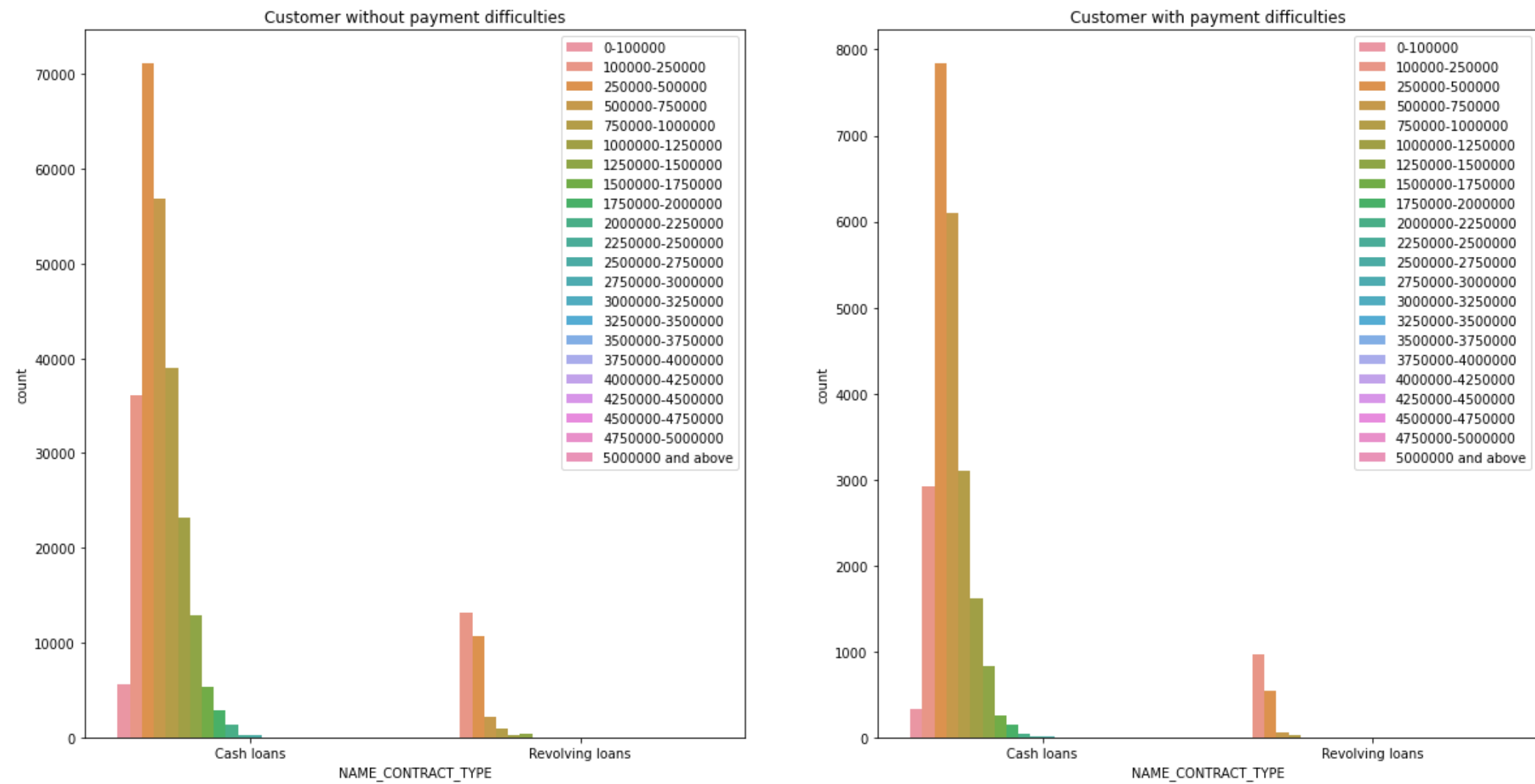
Univariate Analysis on NAME_CONTRACT_TYPE



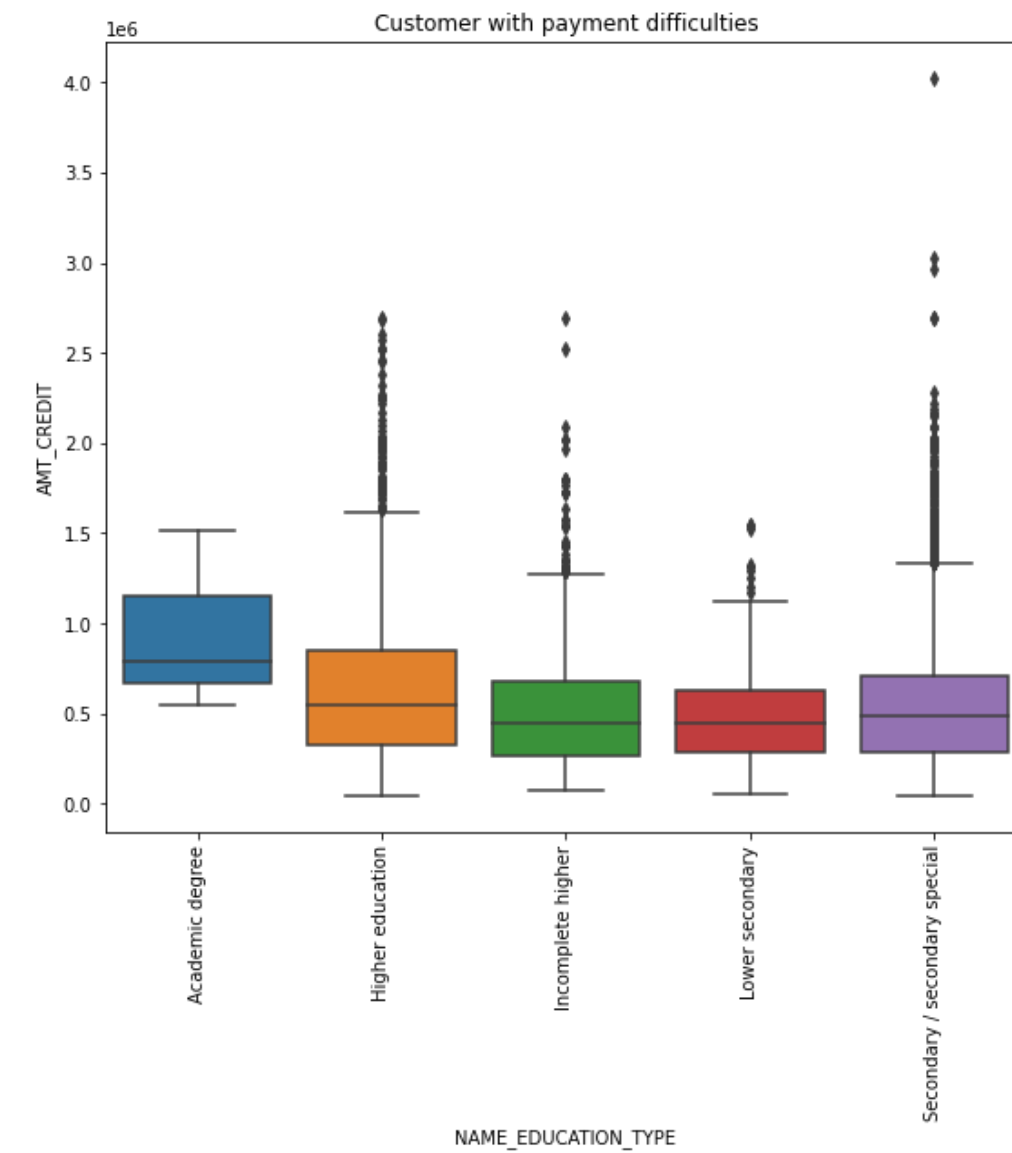
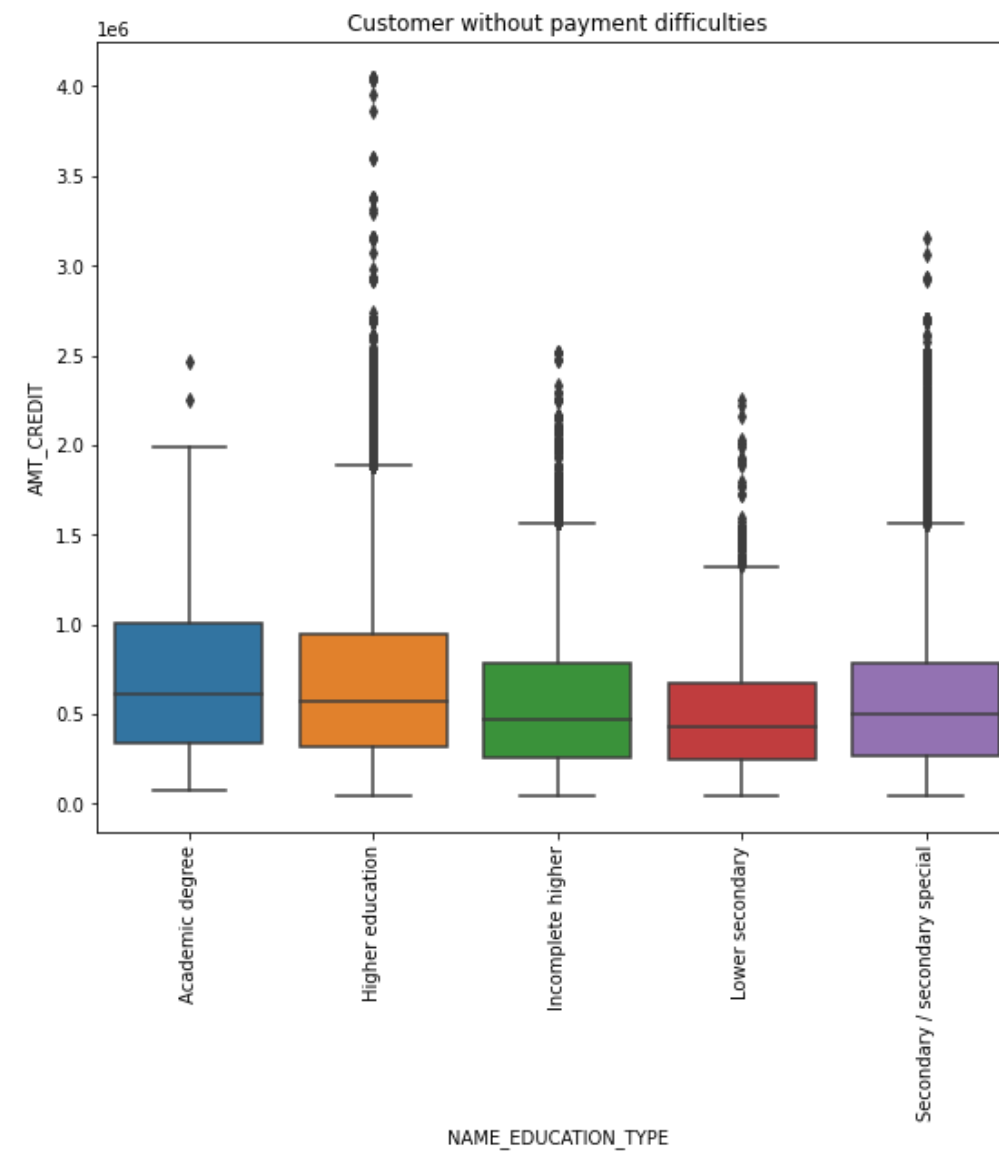
Bivariate Analysis for AMT_CREDIT to AMT_GOOD_PRICE



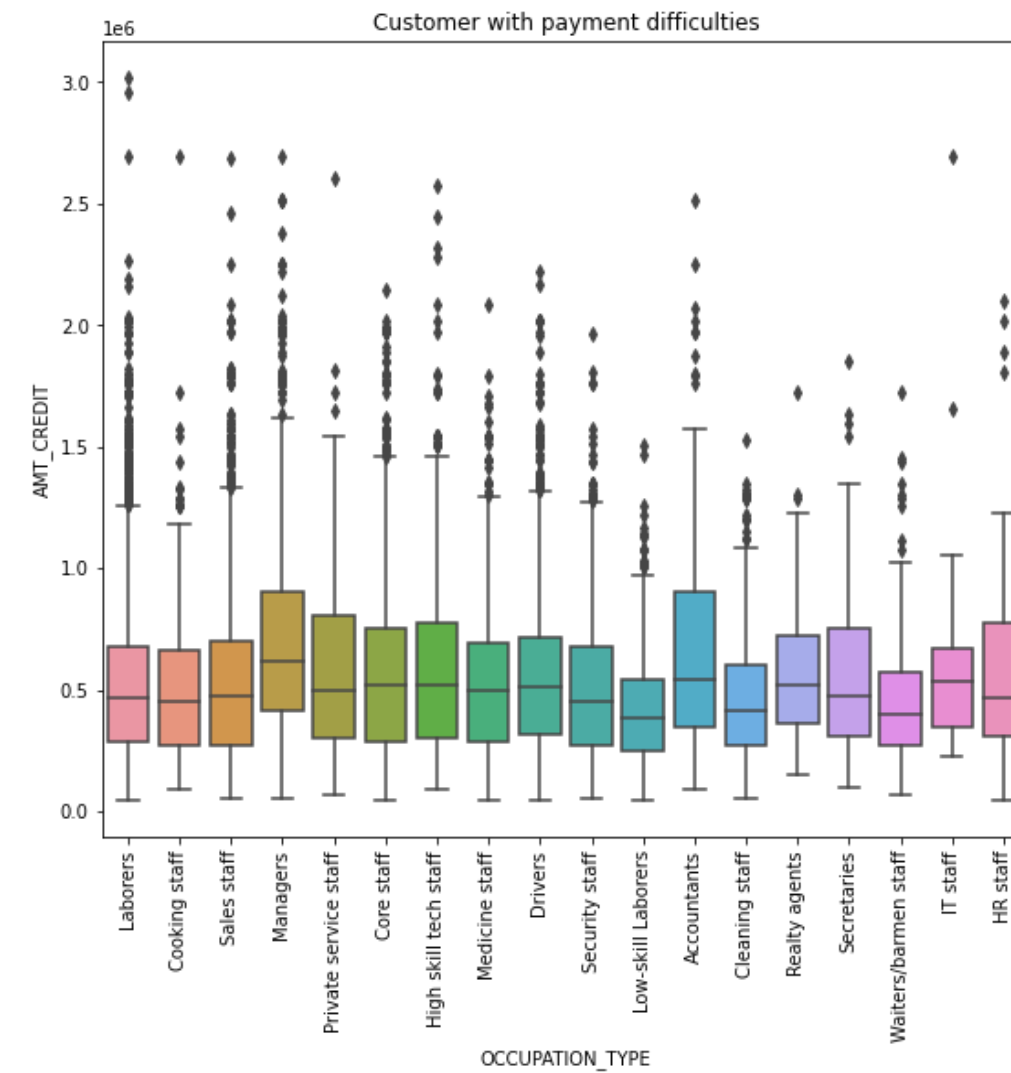
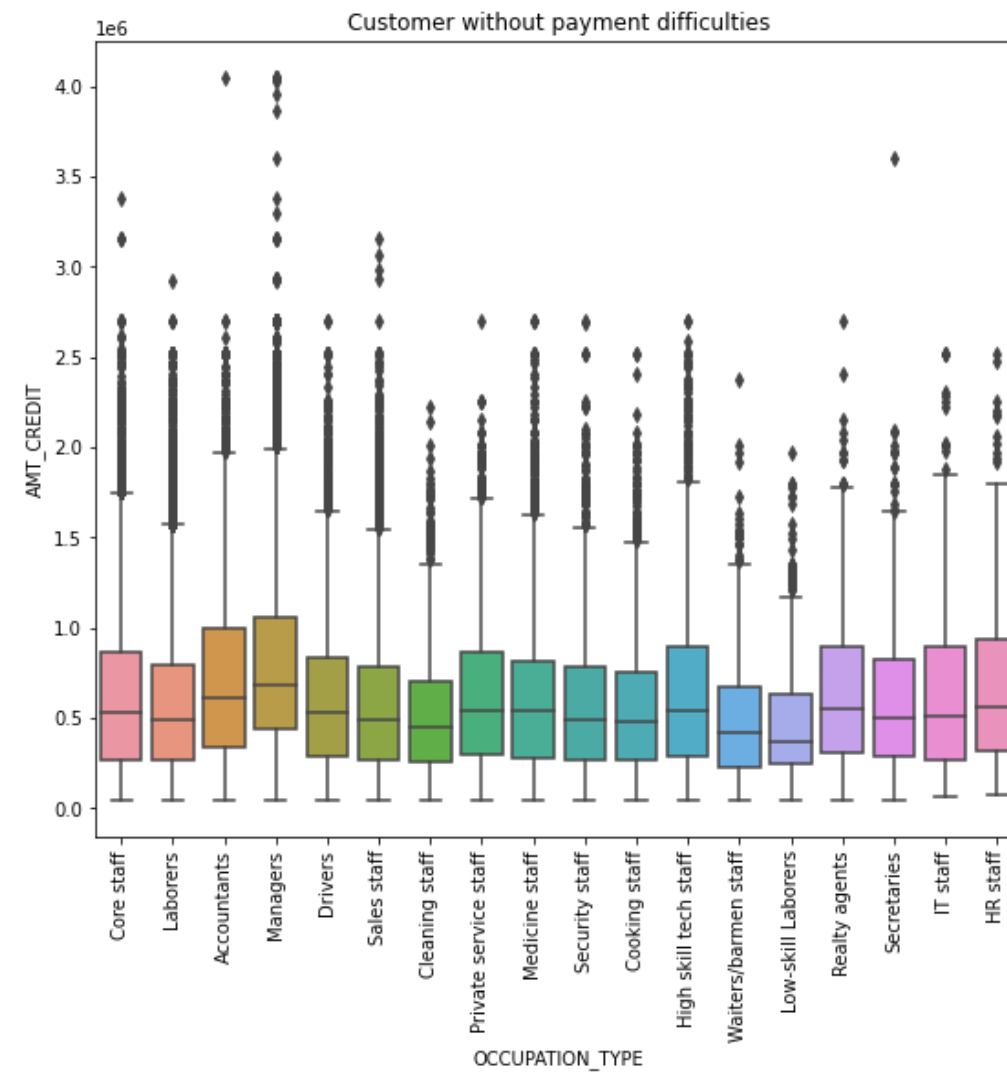
Bivariate analysis for NAME_CONTRACT_TYPE to CREDIT_RANGE



Bivariate analysis for AMT_CREDIT to NAME_EDUCATION_TYPE



Bivariate analysis for AMT_CREDIT to OCCUPATION_TYPE



OBSERVATIONS

01

We observe that the Working Class customers face more difficulties in payments compared to the other classes.

02

Females face more payment difficulties and also apply for more credit and is also approved

03

Customers with Repairs as reason for Credit tend to face more difficulties with Credit Repayment

04

Married Customers apply for most Credit amongst the other classes and also face difficulties in Payment

05

Married Customers apply for most Credit amongst the other classes and also face difficulties in Payment

RECOMMENDATIONS

01

Banks should focused on Single persons as they have the least payment difficulty and also low refused rate and may help with succesful payments.

02

Banks should focus less on 'Married Persons' as they have the most difficulties with payment and also has the highest approval. Which may lead to higher payment difficulties.

03

Banks should not focus on Education status 'Academic Degree' as they have the most difficulties with payment and can focus on Lower Seconday as they have least difficulties with payment.

04

Banks can focus on 'Realty Agents' as they have somewhat lower difficulties with payments then 'Managers' and 'Accountants' but a low approval rate. So it might be beneficial to focus on that Ocuupation type.

05

Banks rather then focusing on 'Working' Income Type should focus more on 'Commercial Associate' , 'Pensioners' and 'State Servants' as they have the less difficulties with payment compared to 'Working' class.