CREDIT RISK ANALYSIS

EDA-PYTHON

MEERA RADHISH

Supervised by
ARITRI DEBNATH

Introduction

When a loan is given out by the lender, it also causes a credit risk. A credit risk is the risk of default on a loan that may arise from a debtor failing to make necessary payments. The risk to the lender comprises of lost principal and interest, disorder in cash flows, and an increase in collection charges.

To determine the credit risk accompanying a loan portfolio, it would be worthwhile to forecast whether a borrower will default. Consequently, actions can be taken to mitigate possible loss to the creditor.

In this project, Probability of Default model has been developed with the help of past credit information about existing borrowers and their behavior with the bank to calculate and measure the risk of each borrower.

Overview

Problem Statement

Exploratory Analysis

Conclusion

Problem Statement

Using EDA we are helping the customer finance company who lends various types of loans to ensure the applicant is capable of loan repayment as well as finding the defaulters. On recieval of a loan application , the approval of the same depends on applicant‘s profile like age, salary . Assets, previous loan history etc.
Major risks asssociated with the decision are:

1. If the applicant is ontime payer, then rejecting the loan will result big loss to the company
2. If the applicant is defaulter, then loan approval will also generate financial loss to the company

Objective

Our objective is to recognize the patterns which demonstrate which group of clients have the payment difficulties and which group don’t have payment difficulties. Thus help the company to

Take precautions while approving the loan to these groups.

Understanding of the dataset

The data set has 3 files as follows:

- *application_data.csv* : all information about the applicant at the time of application. It‘s all about to know whether applicant has payment difficulties.

- *previous_application.csv* : Information about the applicant‘s previous loan history. It contains the information on whether the loan has been approved, canceled or rejected.

- *columns_description.csv* : It‘s a data dictionary which gives the meanings of the variables used in the dataset

Steps involved

- Understand the domain and variables with the help of data dictionary provided.
- Import the *application_data.csv*
- Check the structure of the data
- Data cleaning- removing unwanted columns, replace the negative values of the columns with absolute values
- Created new column for applicant‘s age based on *days_birth* column
- Dropped the null values of applicant‘s dataset
- Import the *previous_application.csv.*
- Checked the strucure of the data as well as the null values.
- Data cleaning- removing unwanted columns as well as null values, renamed certain columns with suffix ‚pre‘.
- Observed common column ‚*SK_ID_CURR‘ in both the data sets*
- *Merged both datasets*
- *Replaced NaN values of the data set with 0*

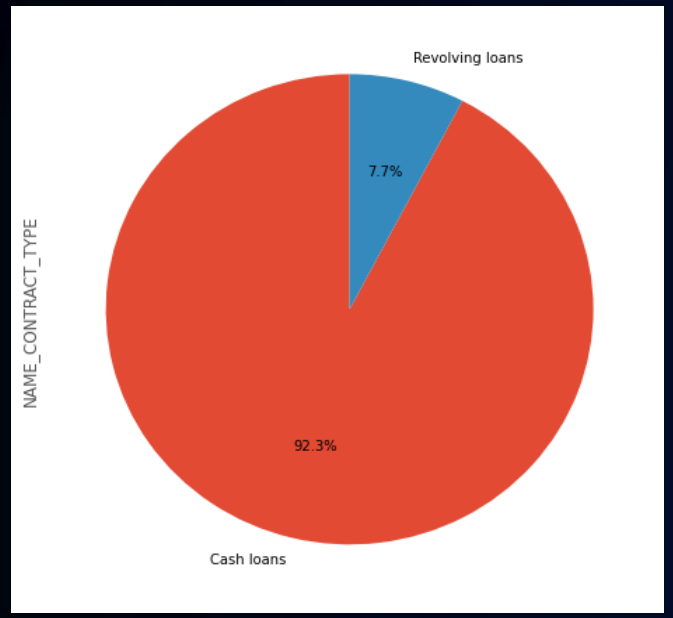
Steps involved

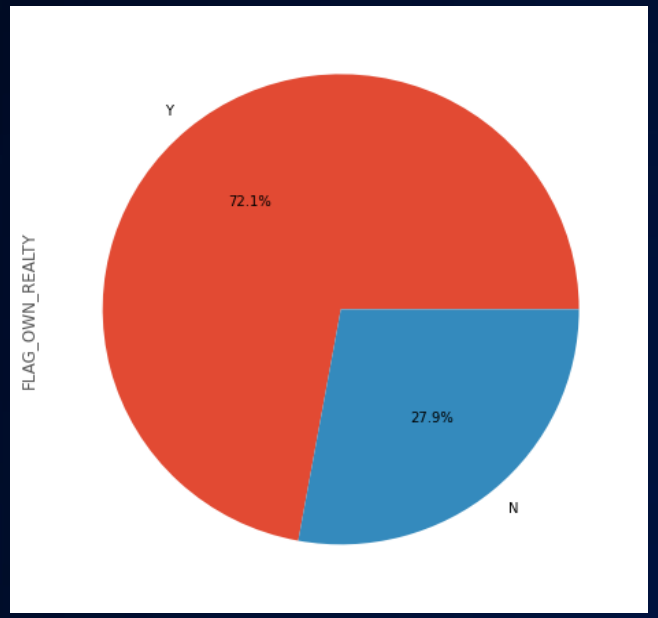
DATA ANALYSIS

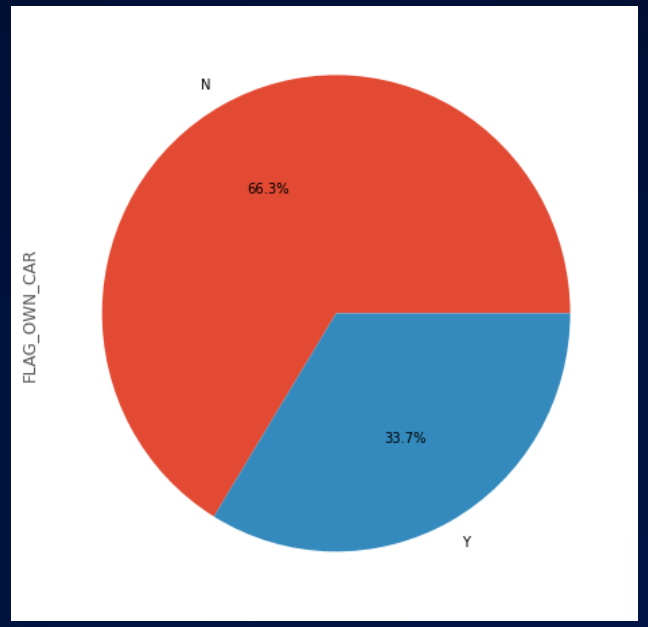
- *Used pie chart for finding contract type distribution,property status,car ownership status*
- *Used bar chart for finding occupation counts of applicants*
- *Outlier Analysis- Income,credit,age,years of employment*
- *Target Analysis*
- *Univariate Analysis*
- *Segmented Univariate analysis*
- *Bivariate Analysis*

APPLICATION DATA ANALYSIS

Visualization

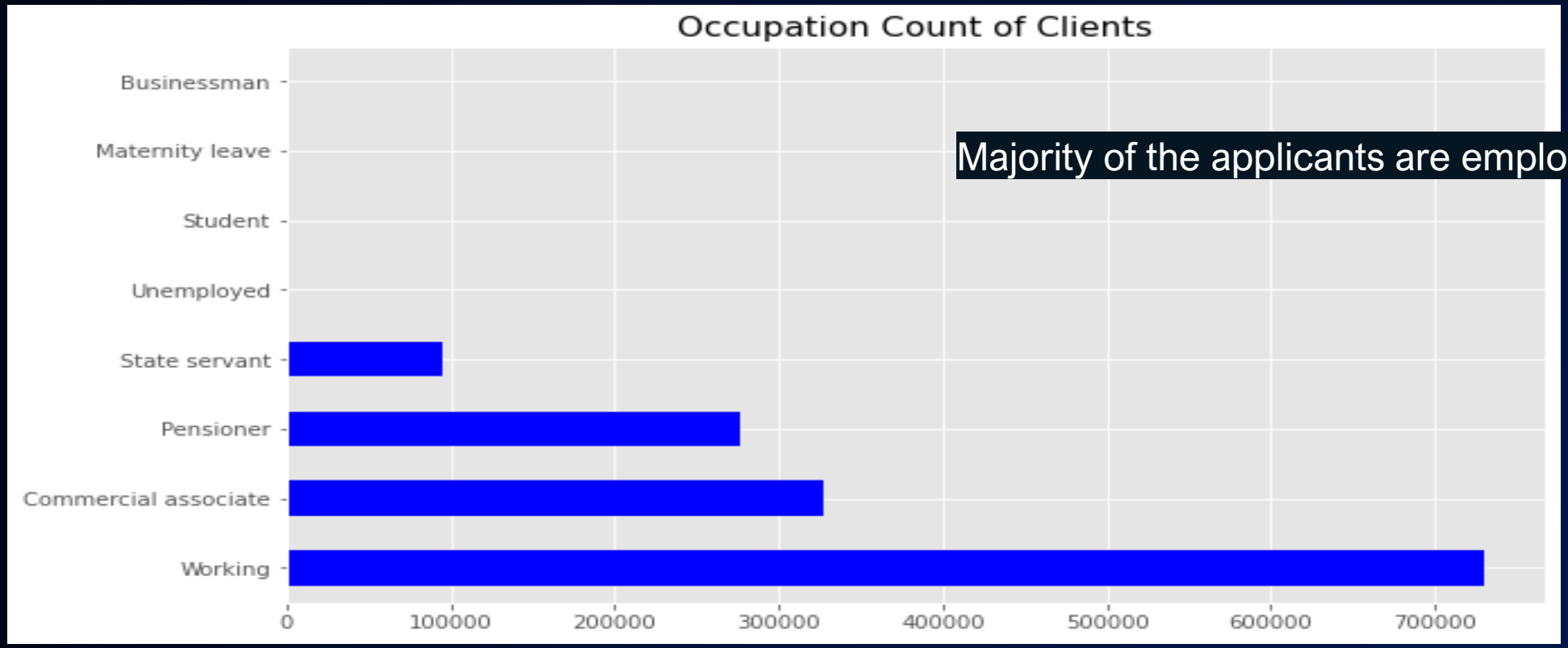






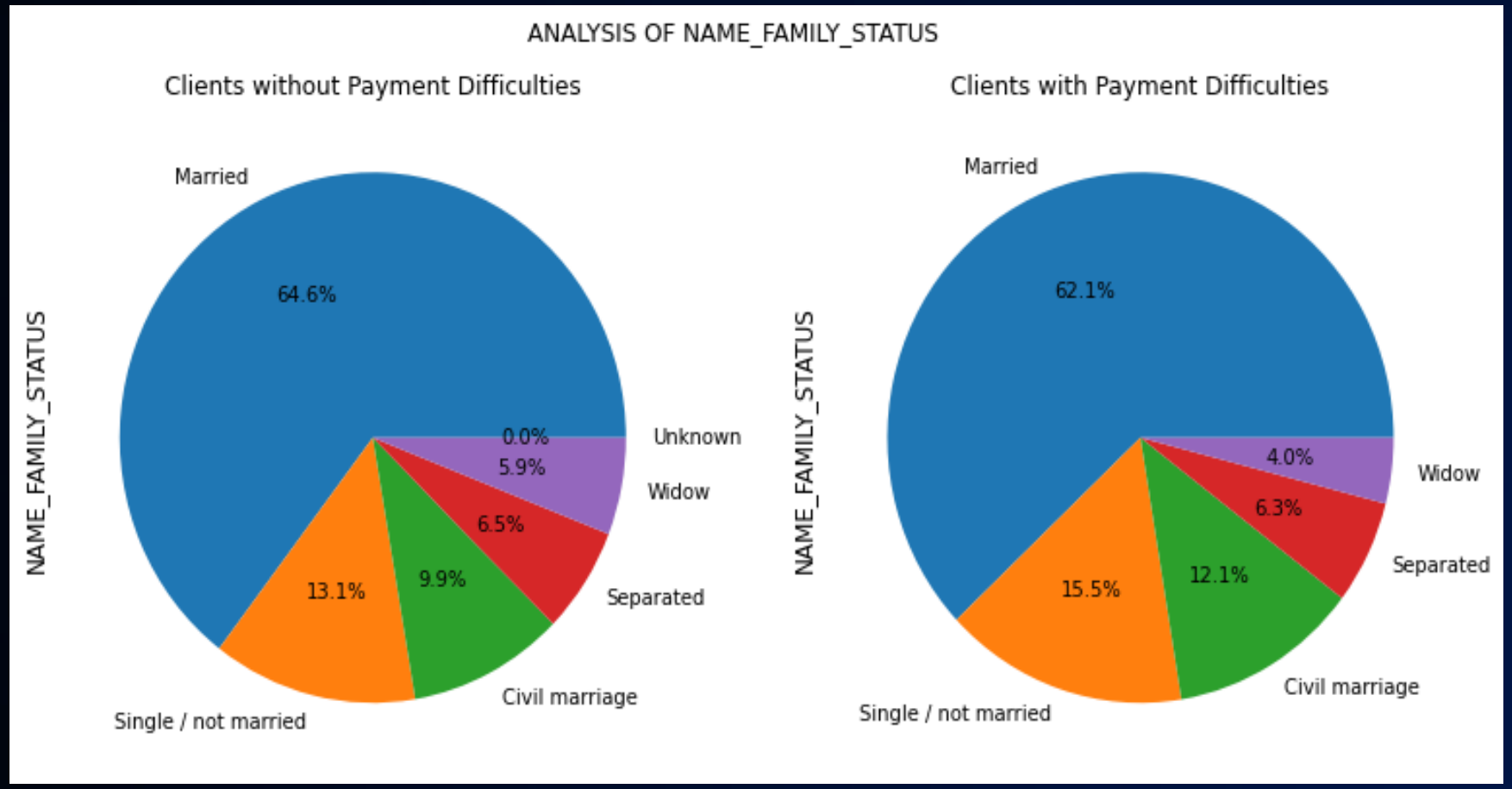
- Applicants are more rely on Cash loans than revolving loans.
- Most of the applicants own property
- Majority of the applicants doesn‘t own a car

Visualization

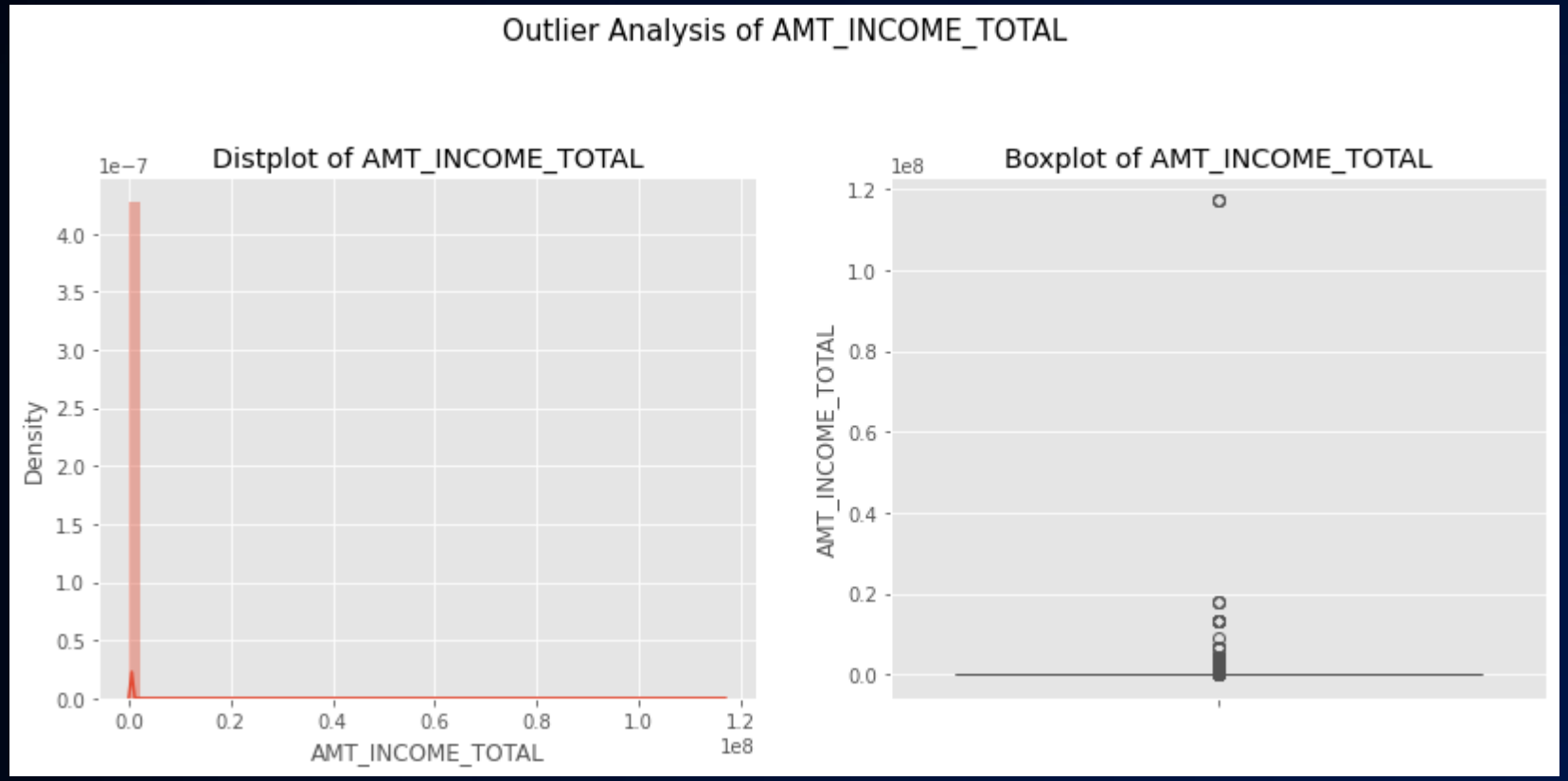


Majority of the applicants are employed

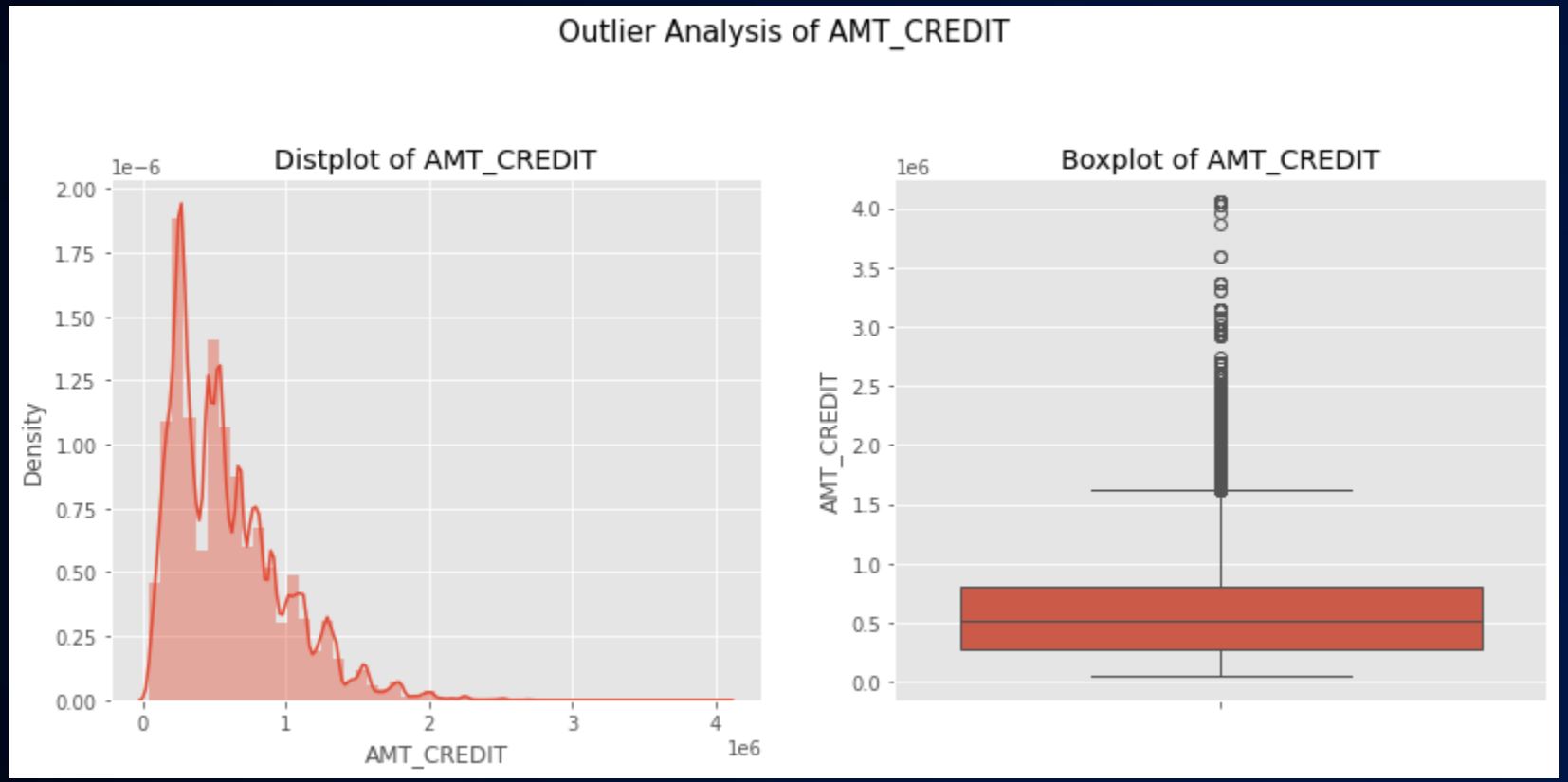
Visualization



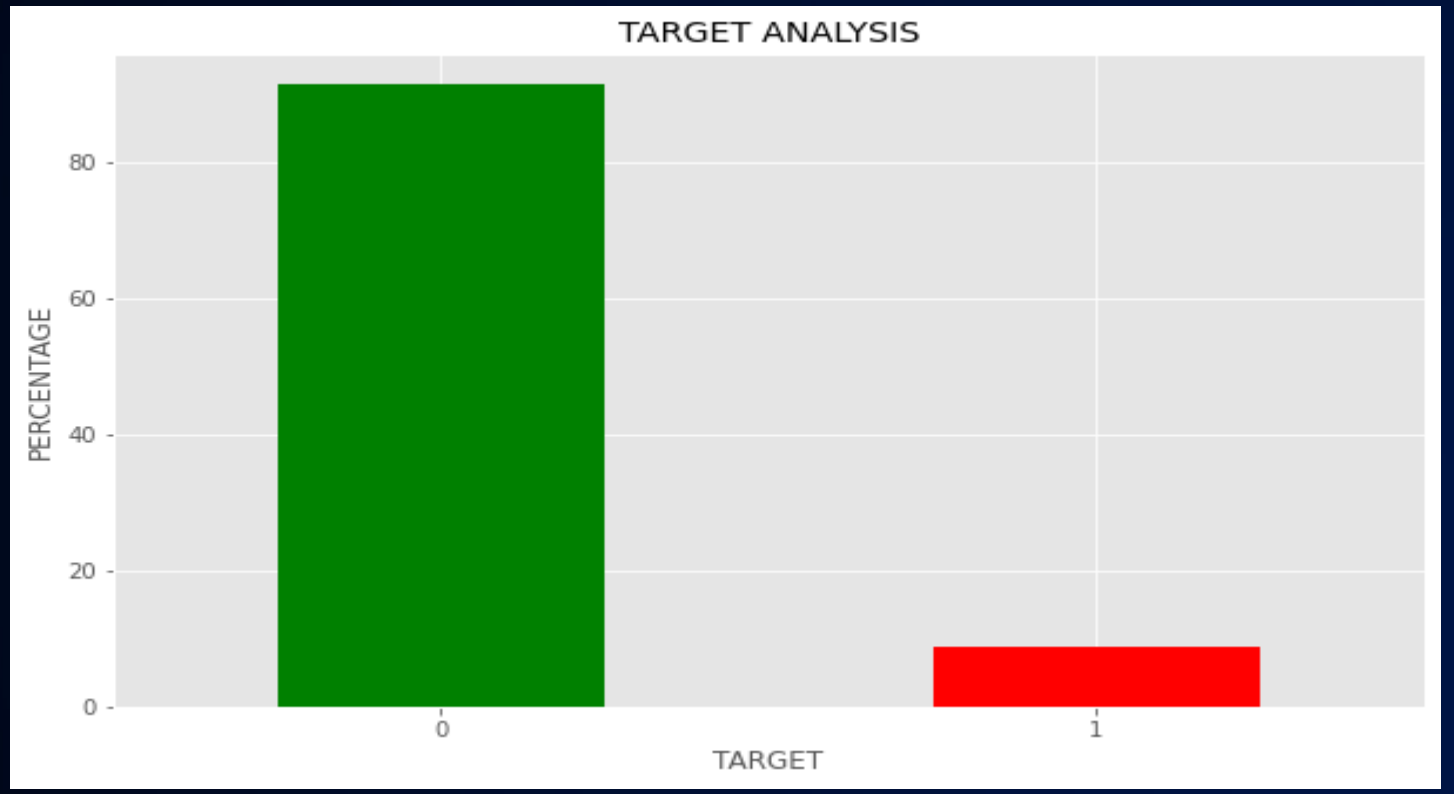
Visualization



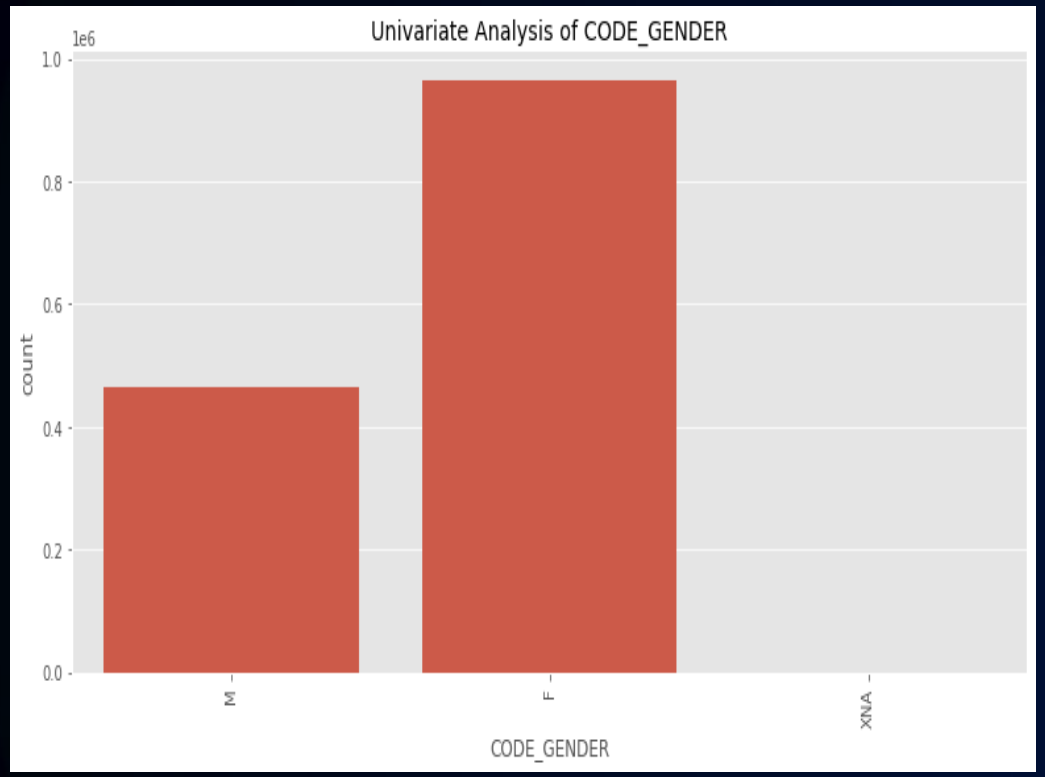
Visualization

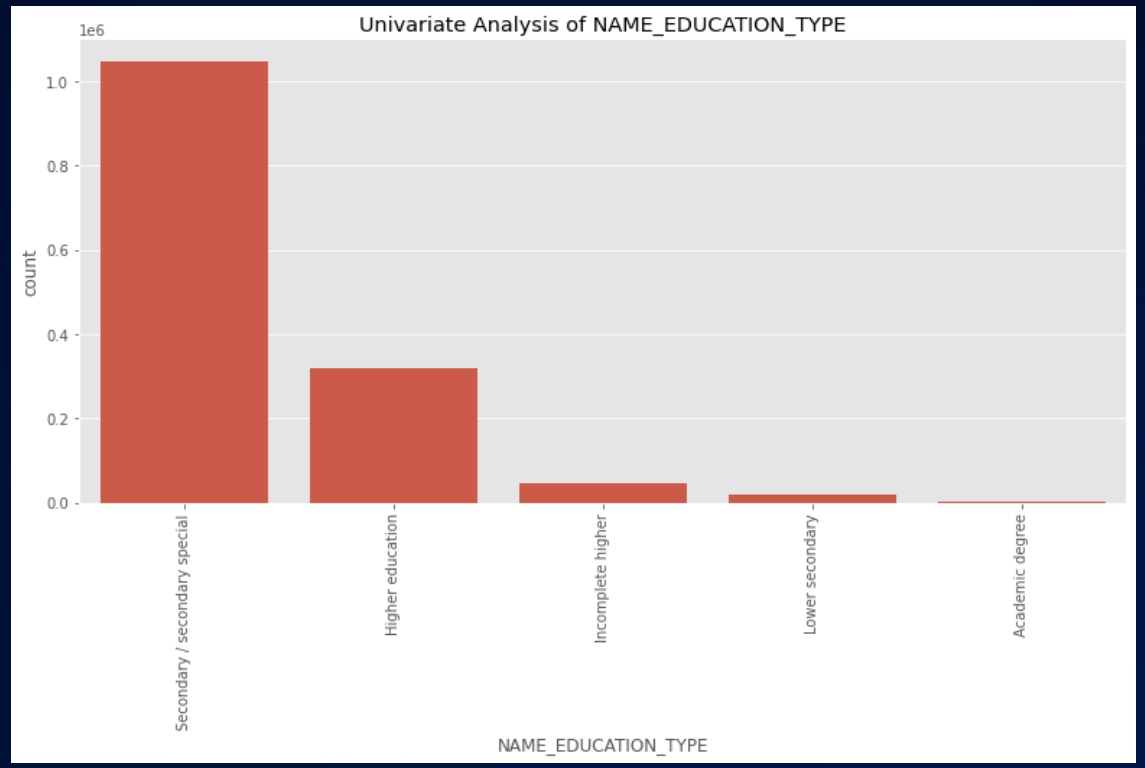


Visualization

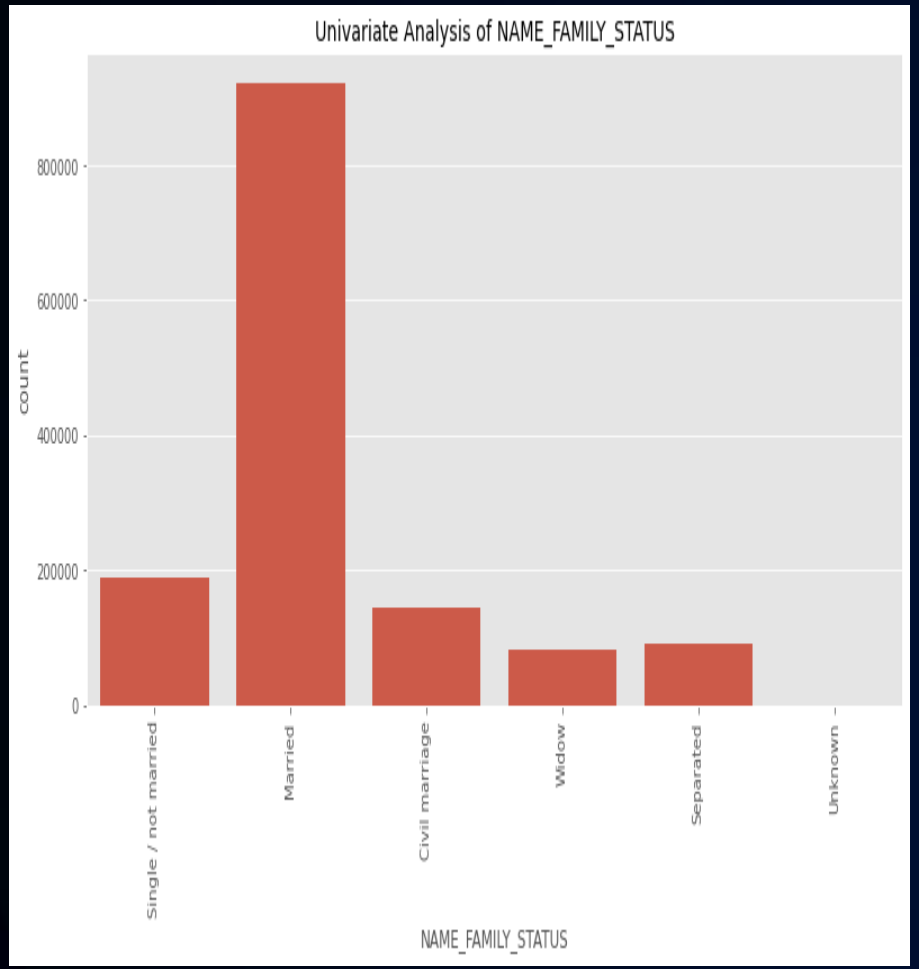


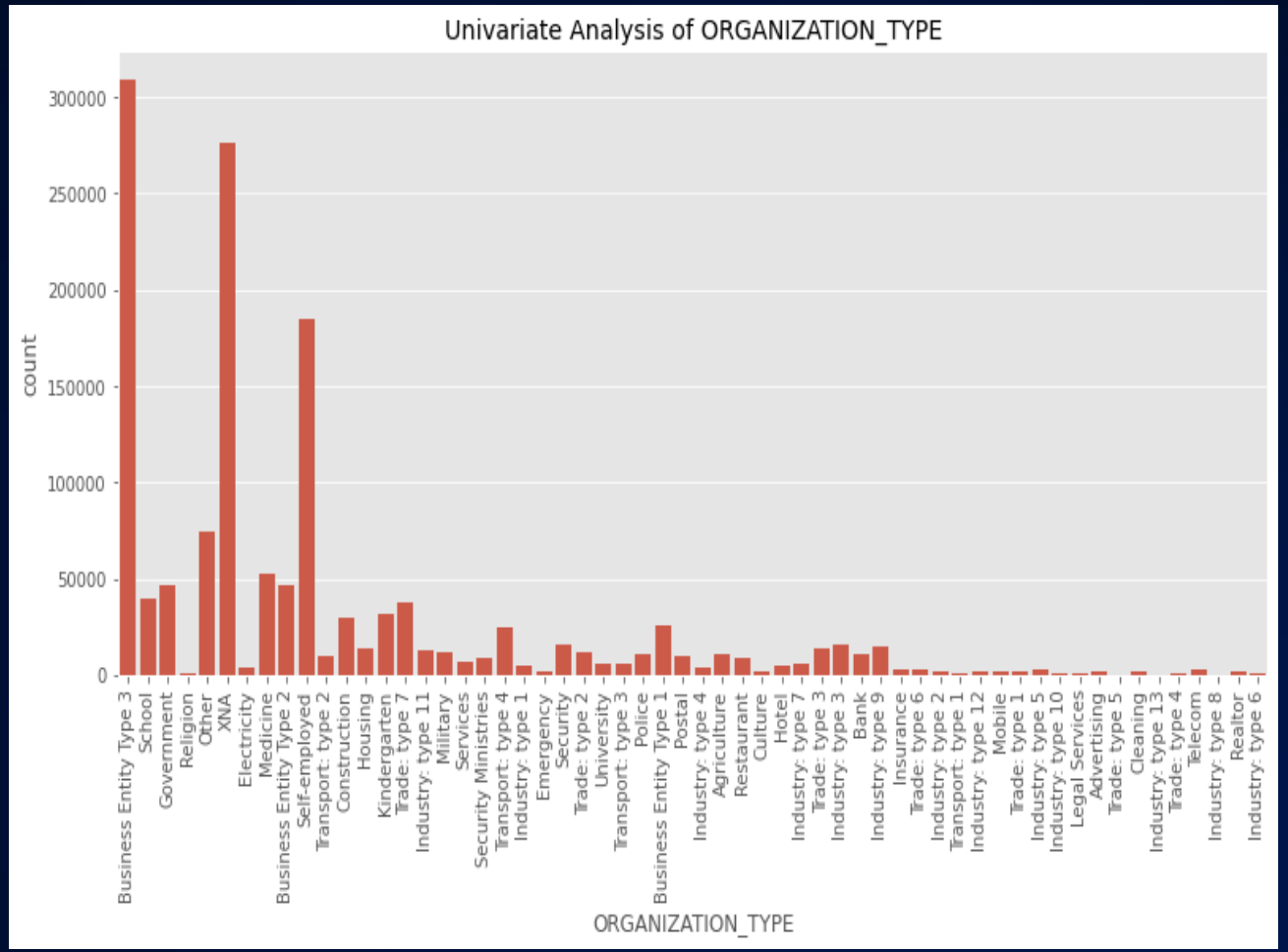
Visualization



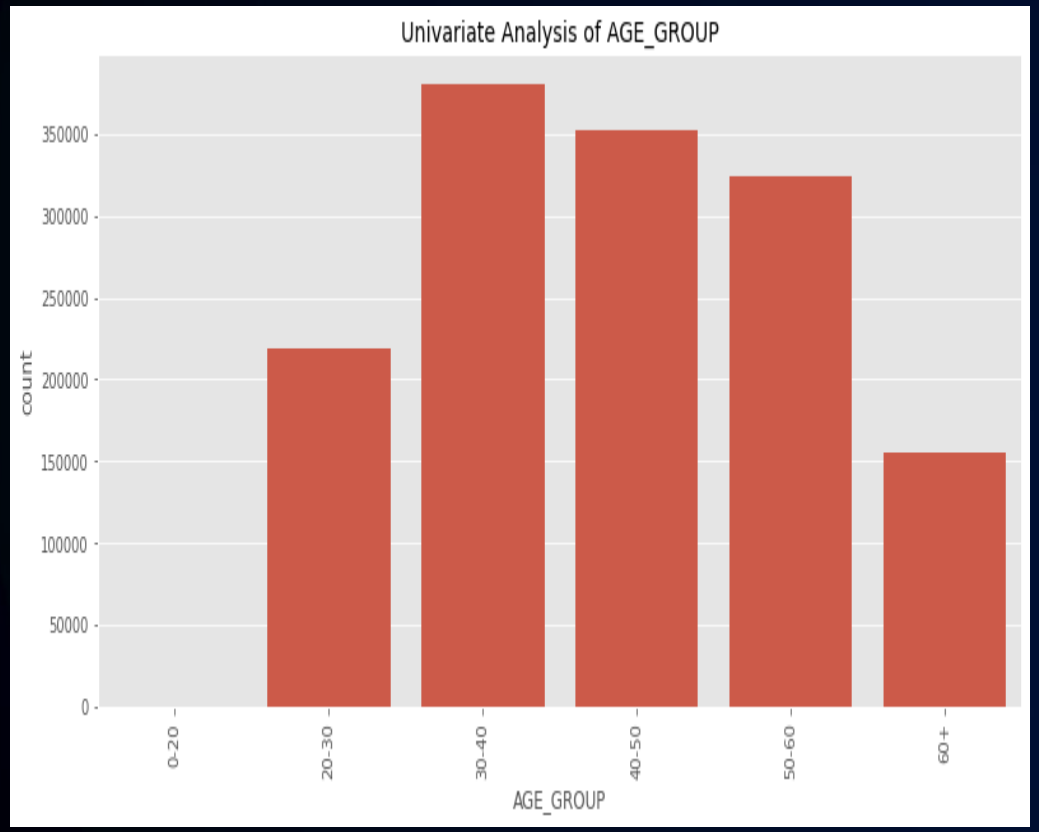


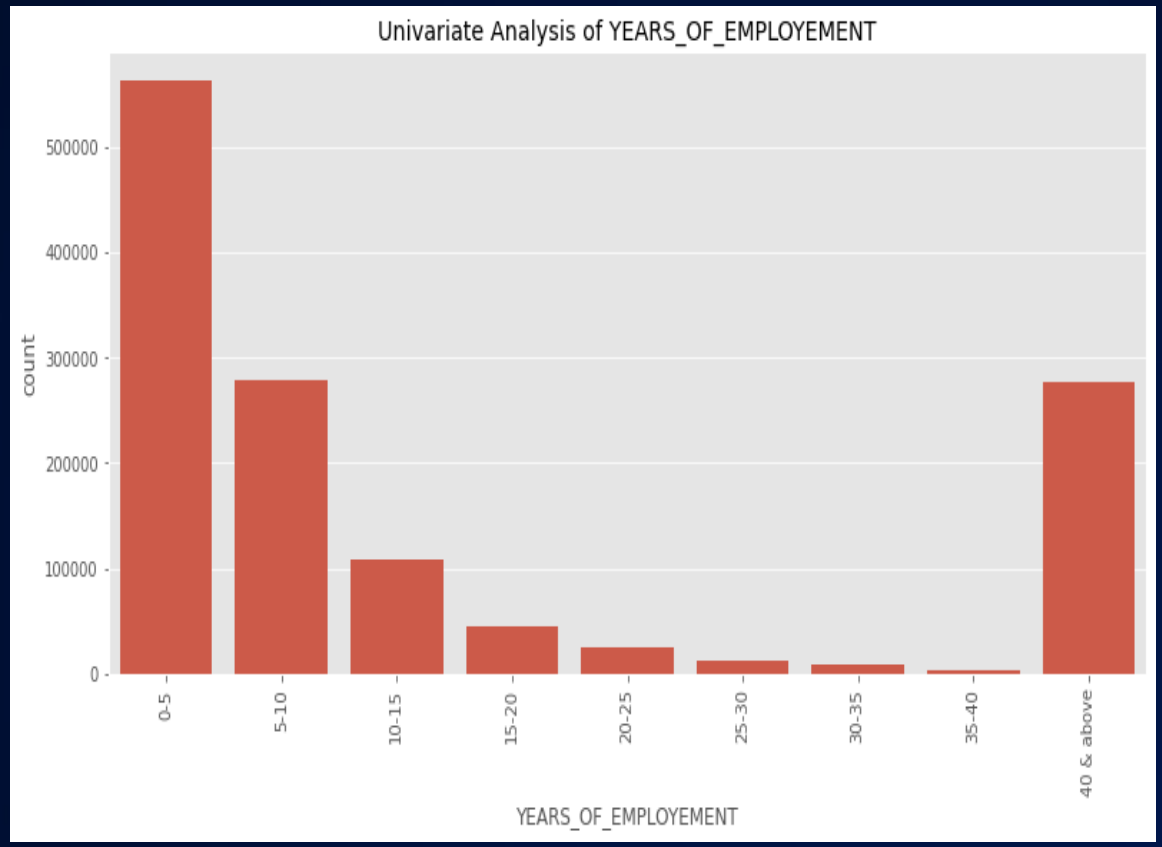
Visualization



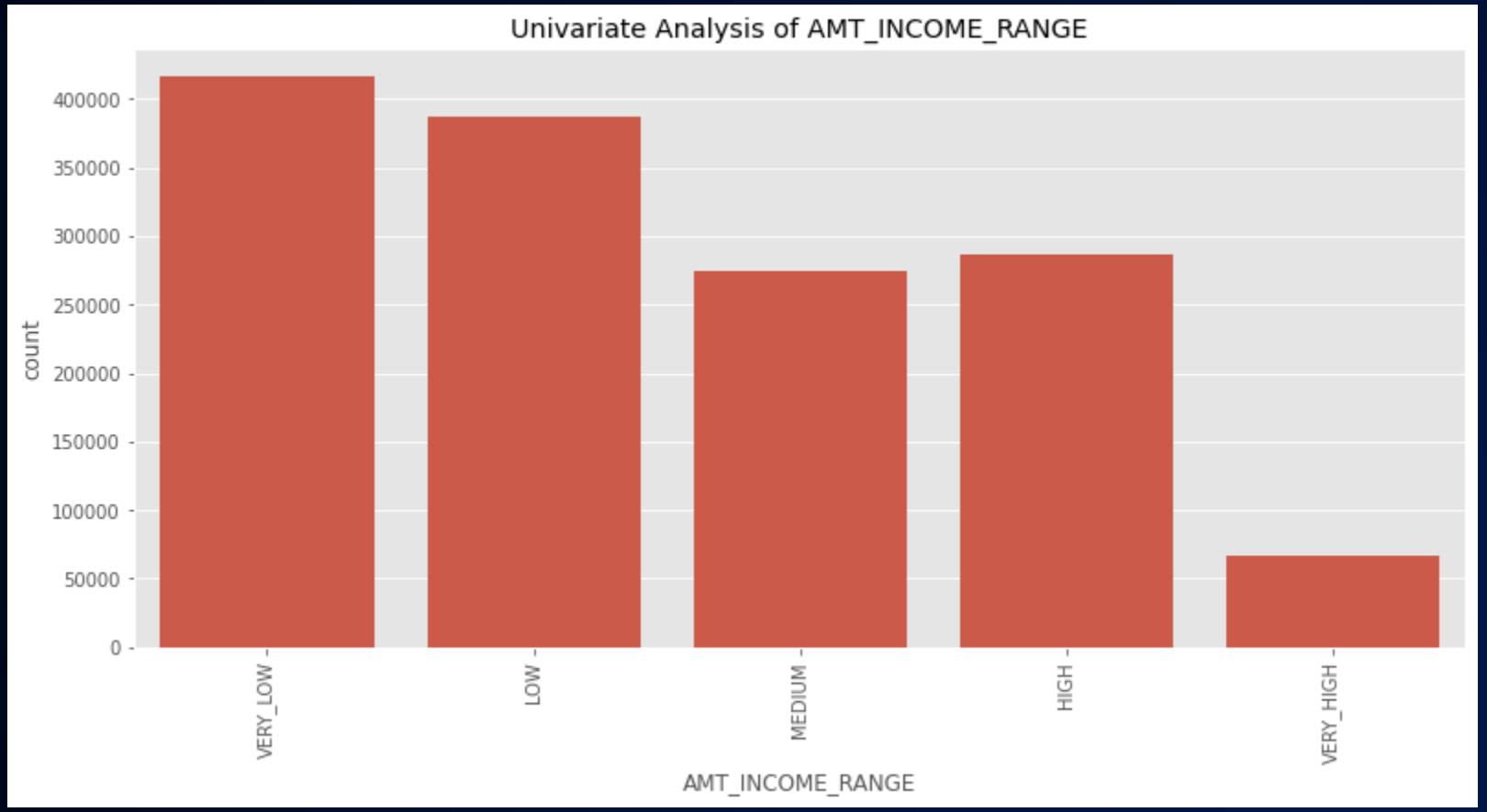


Visualization

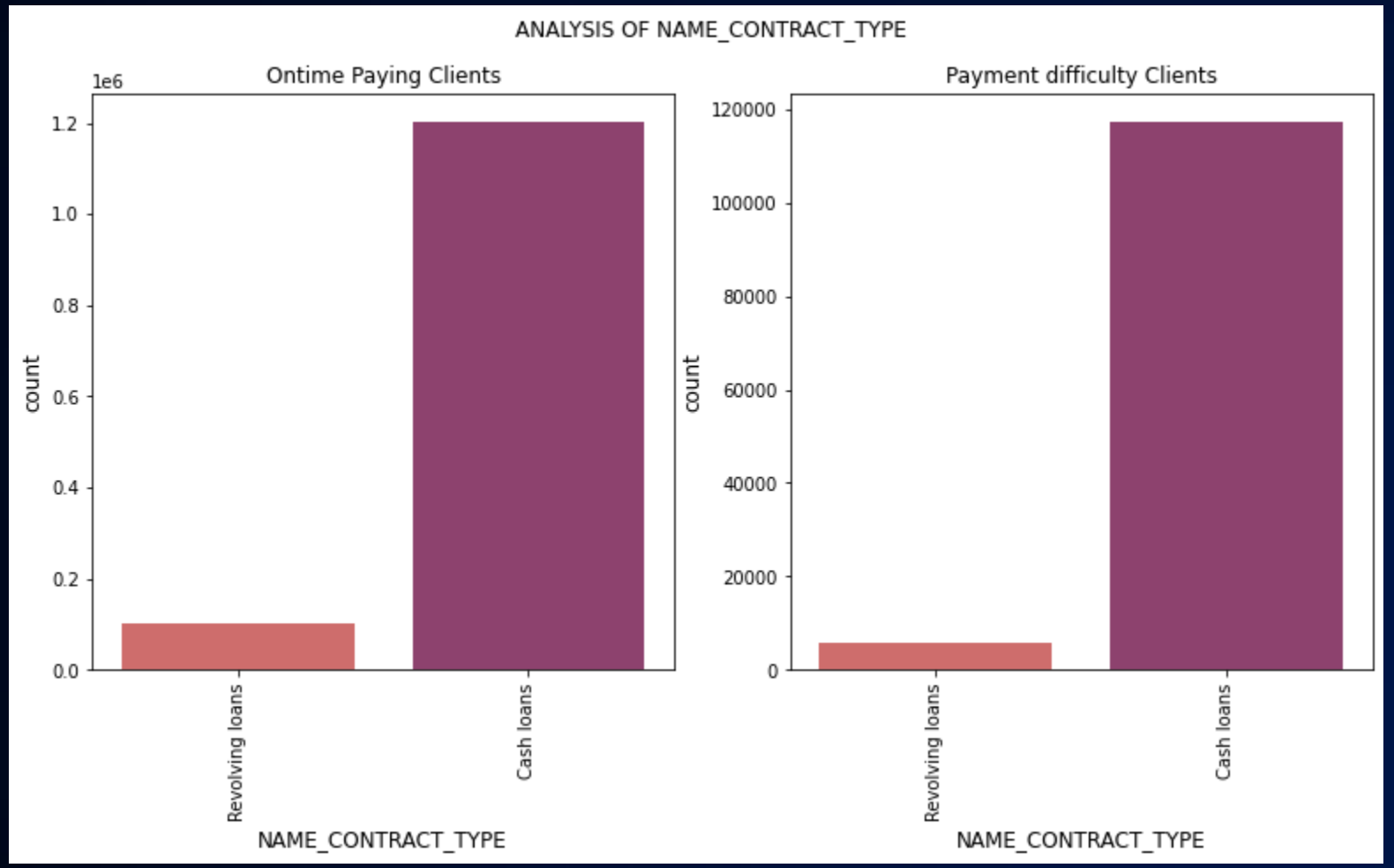




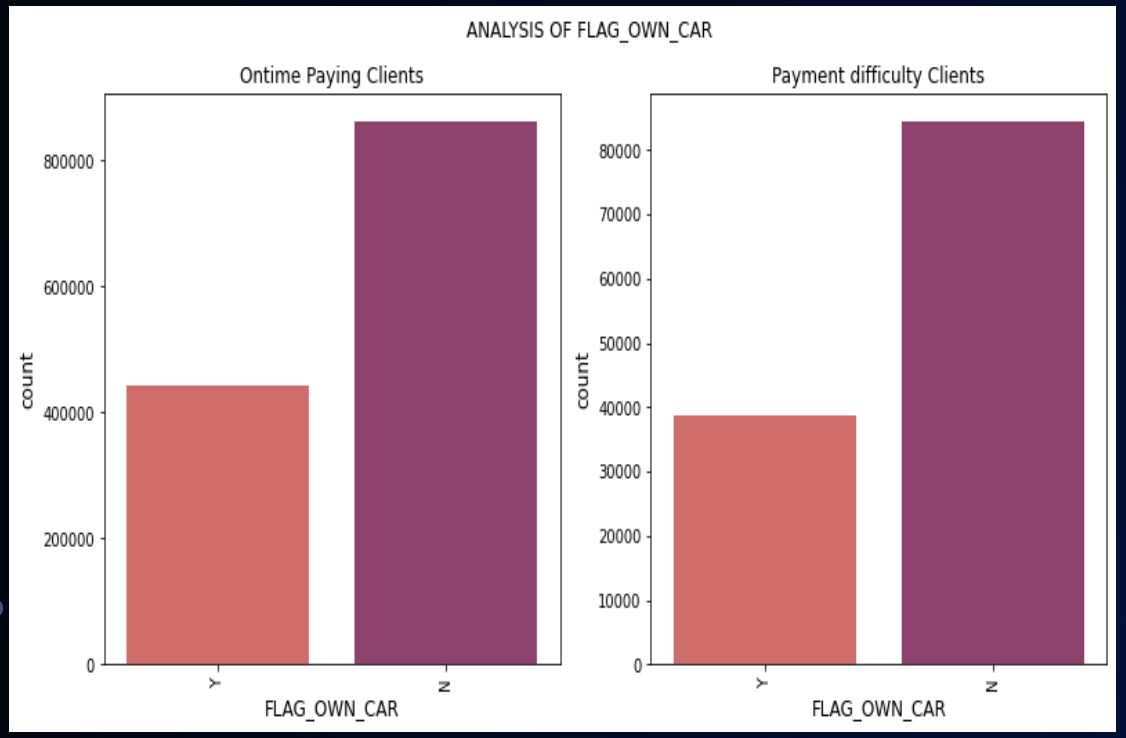
Visualization

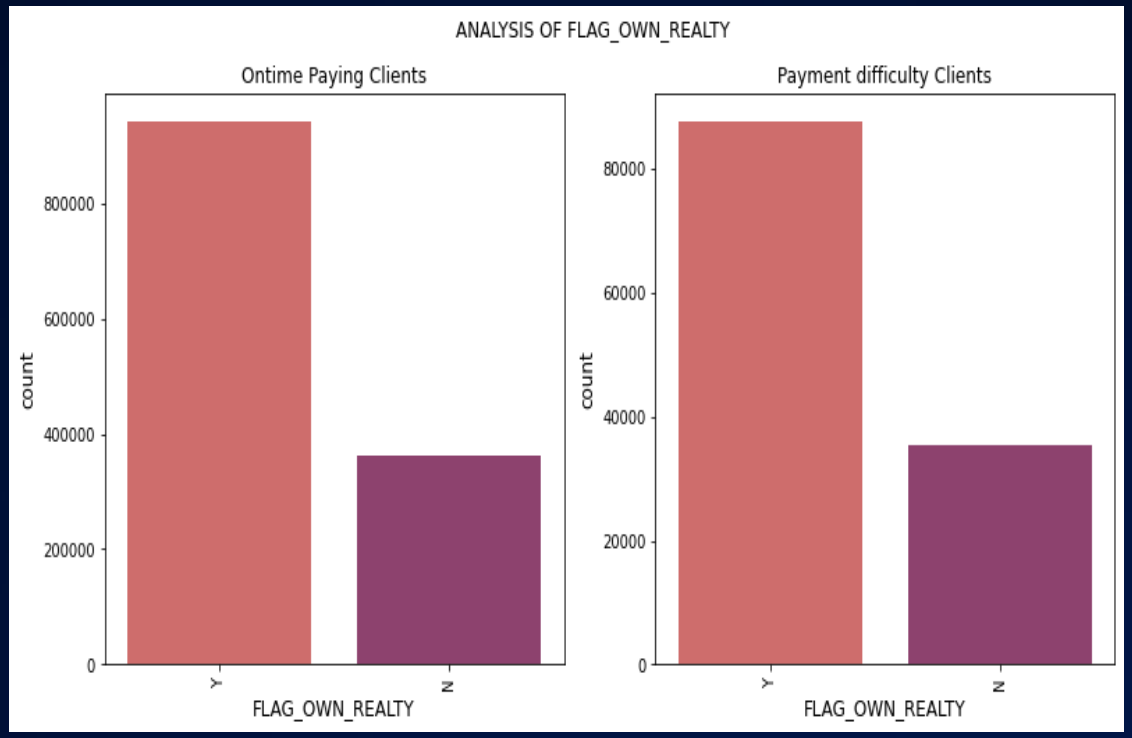


Visualization

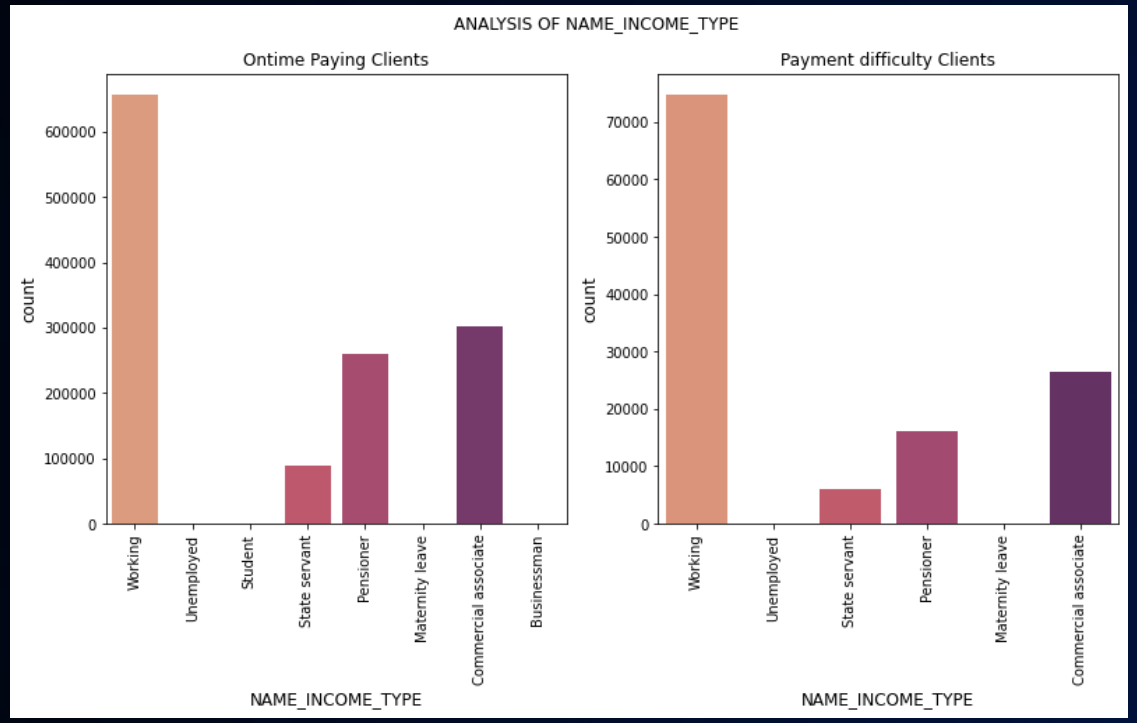


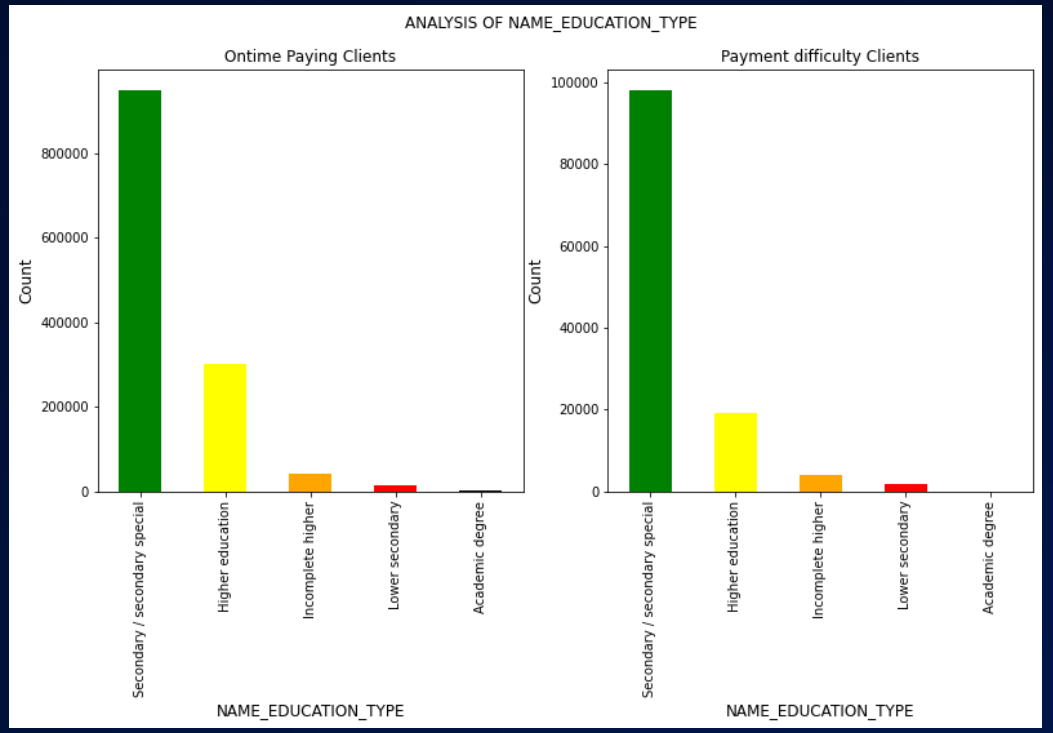
Visualization



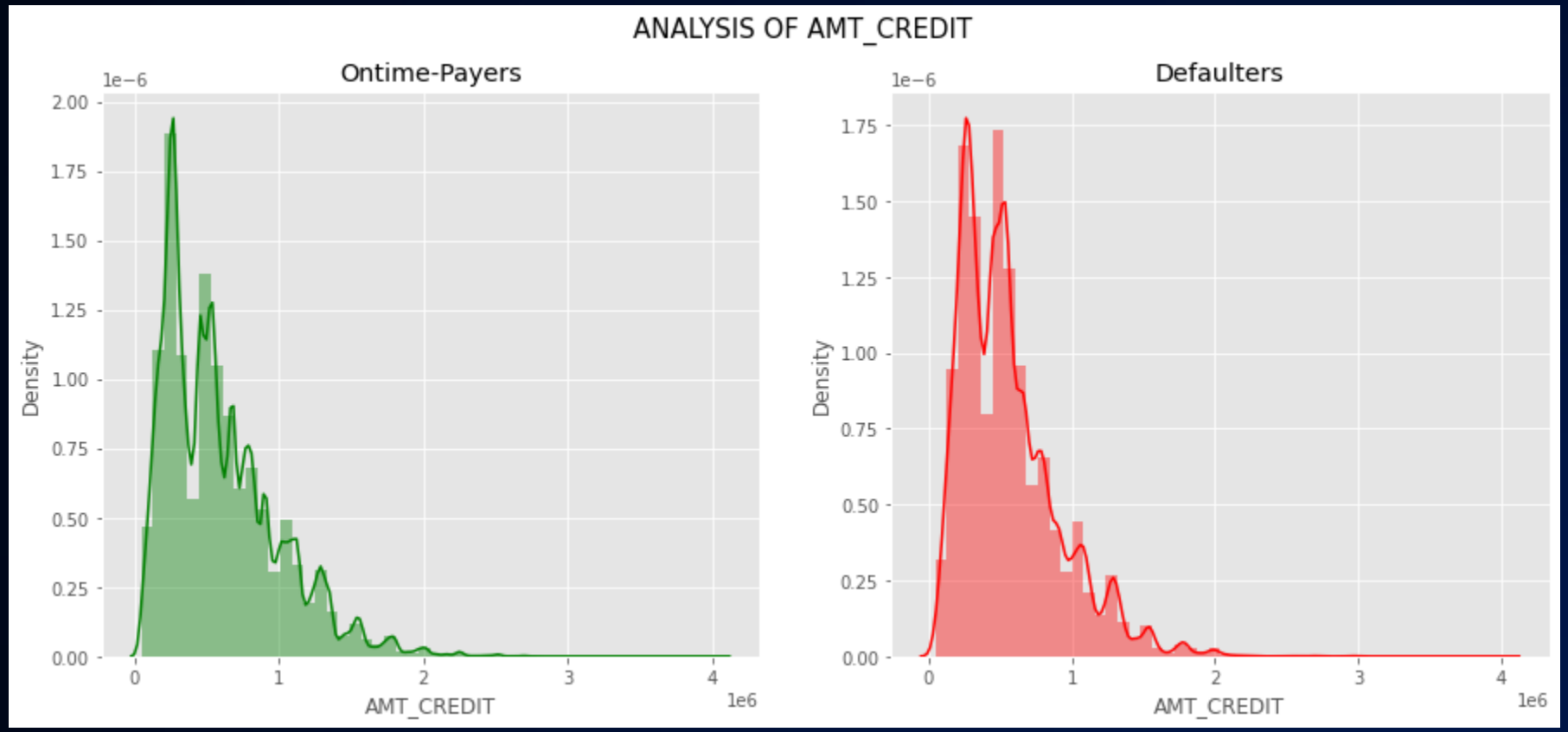


Visualization

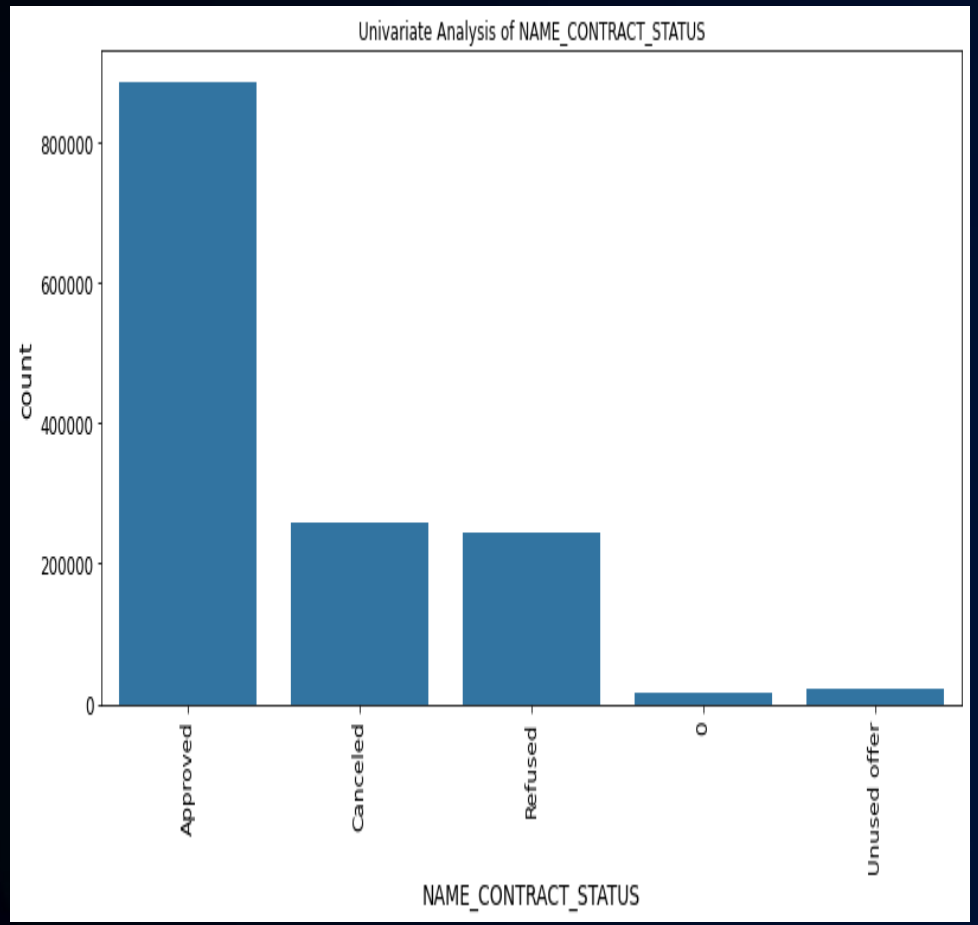


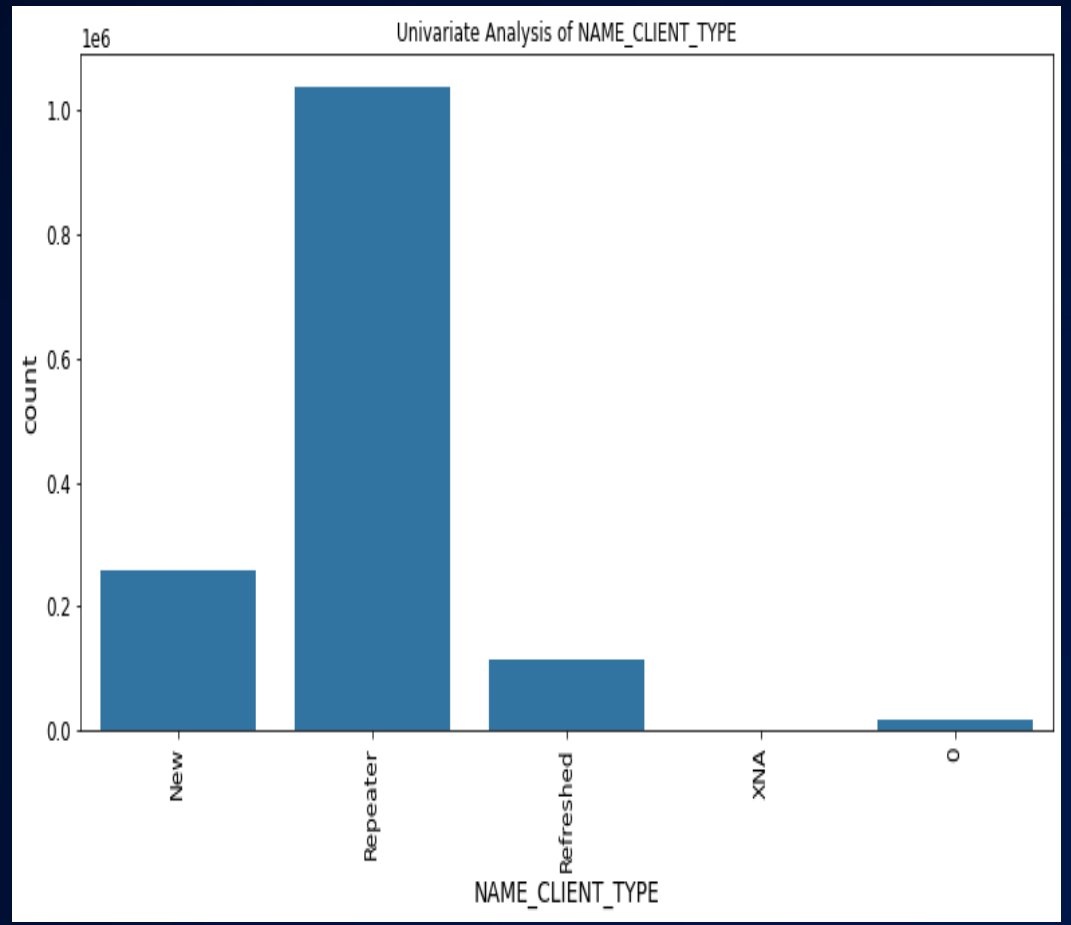


Visualization

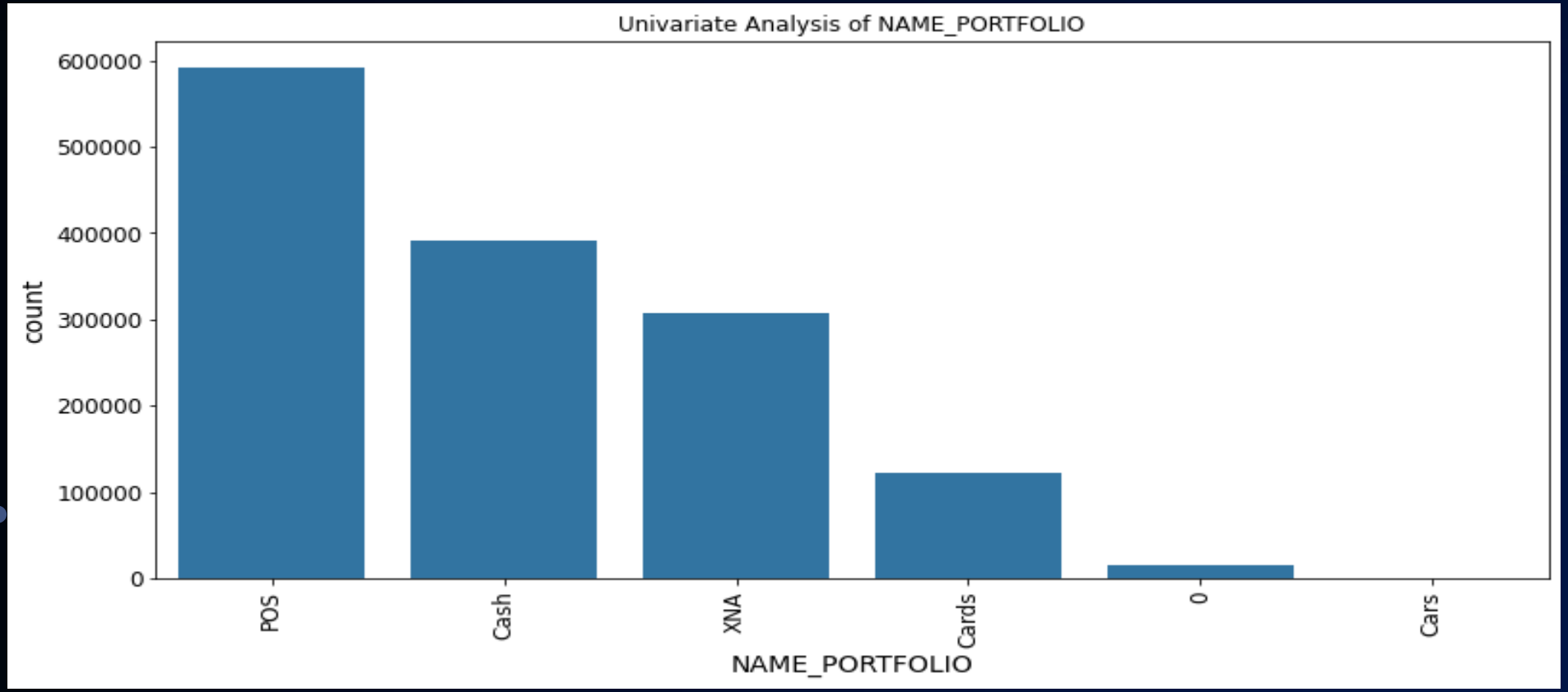


Visualization

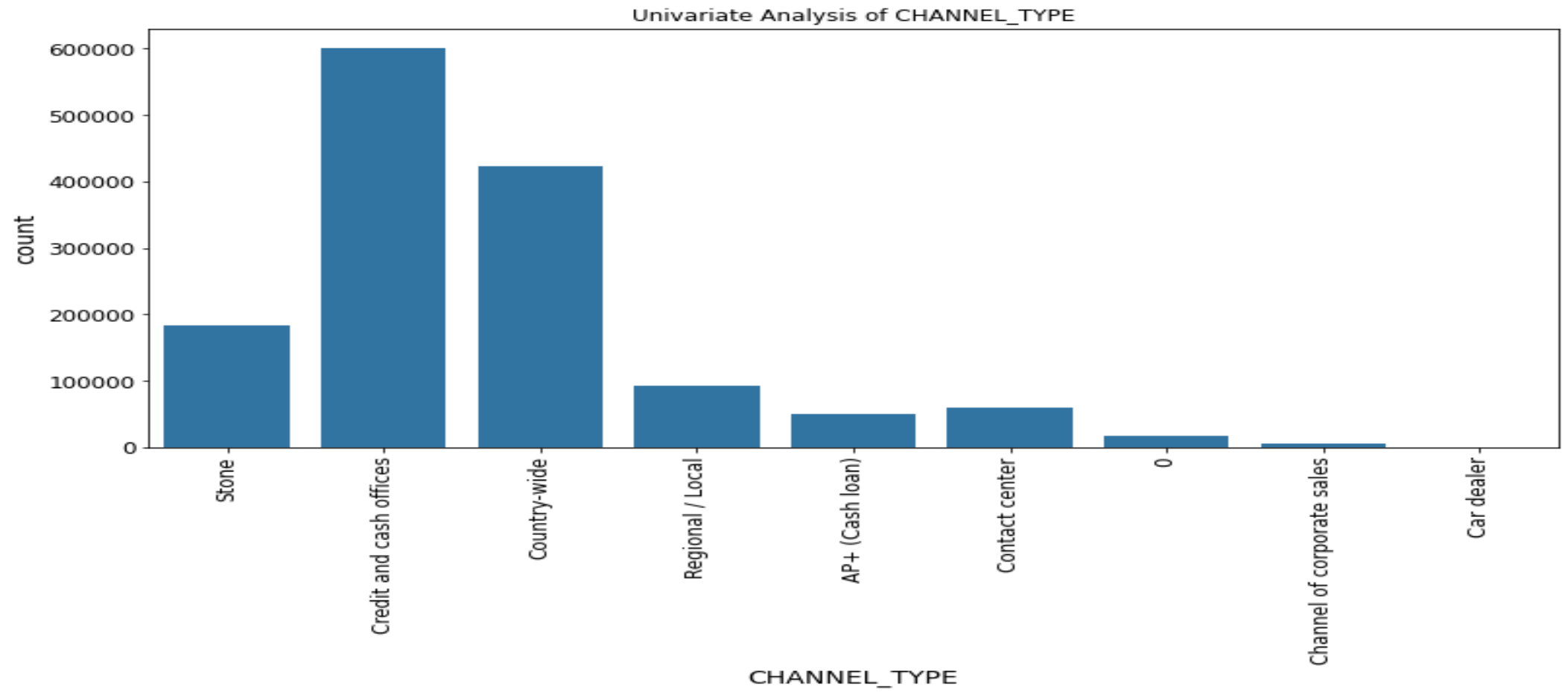




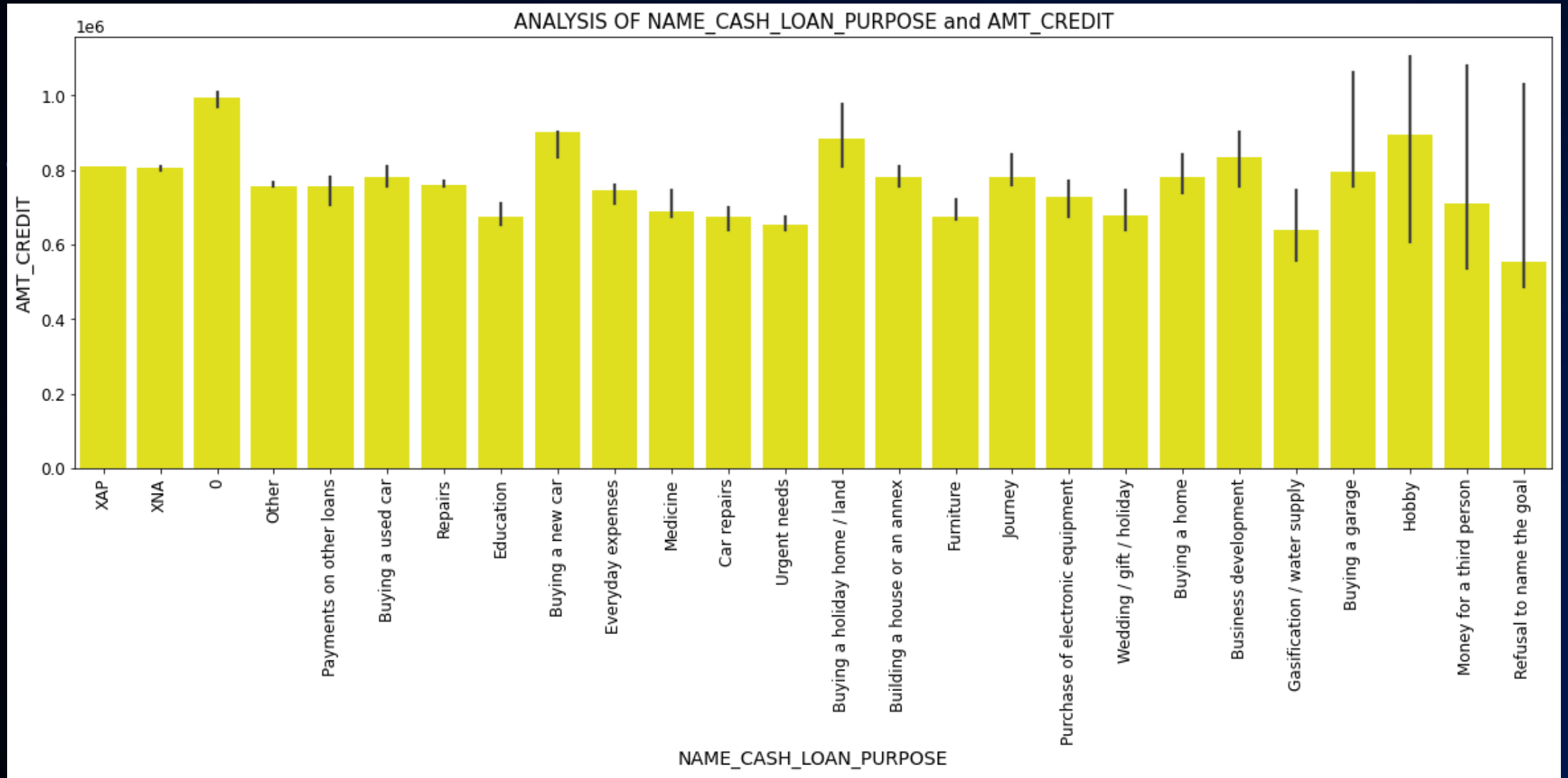
Visualization



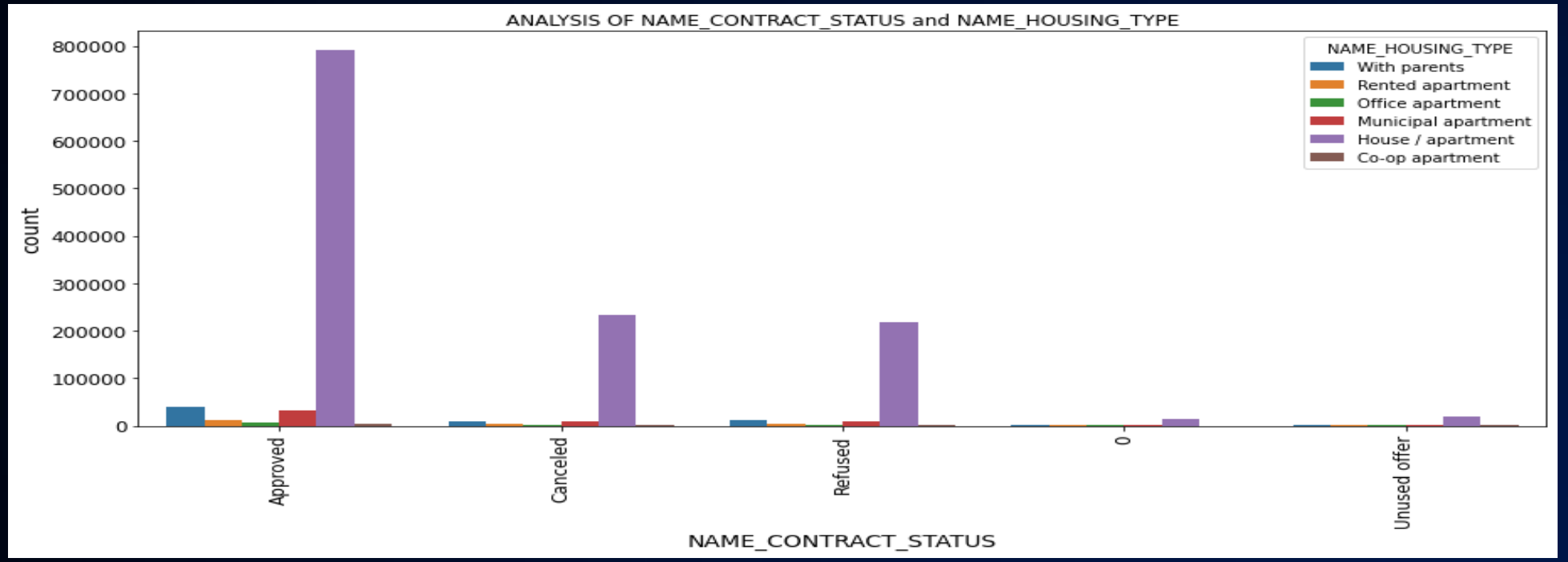
Visualization



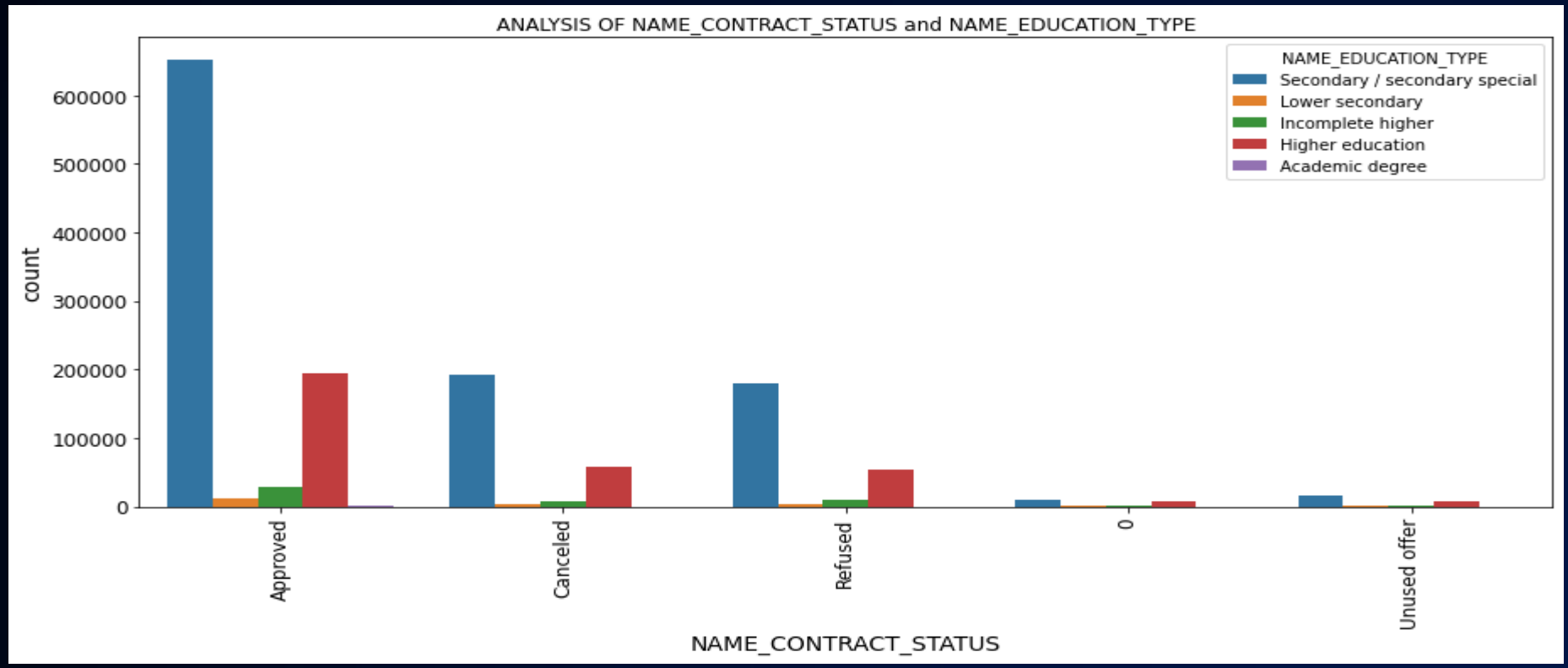
Visualization



Visualization

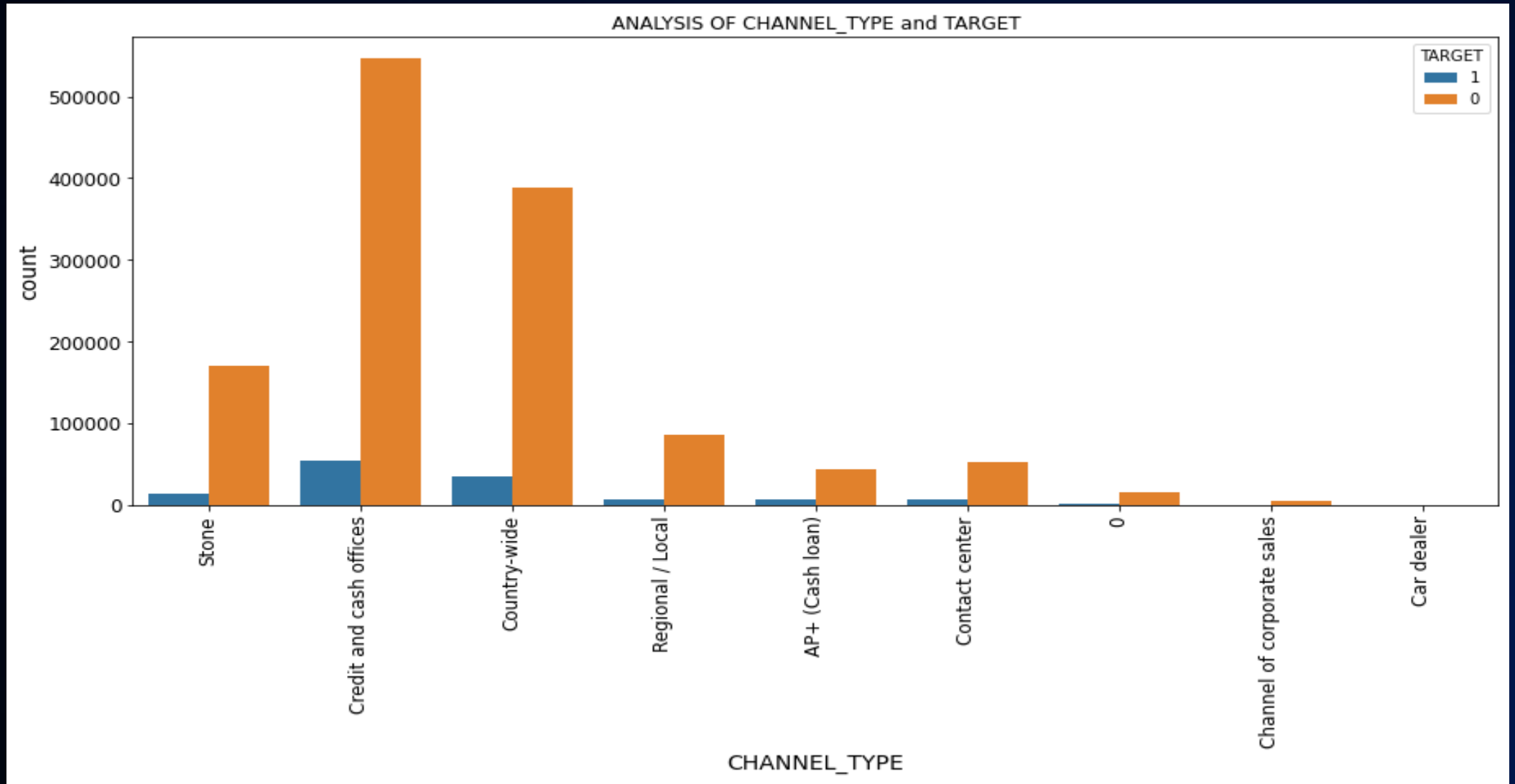


Visualization

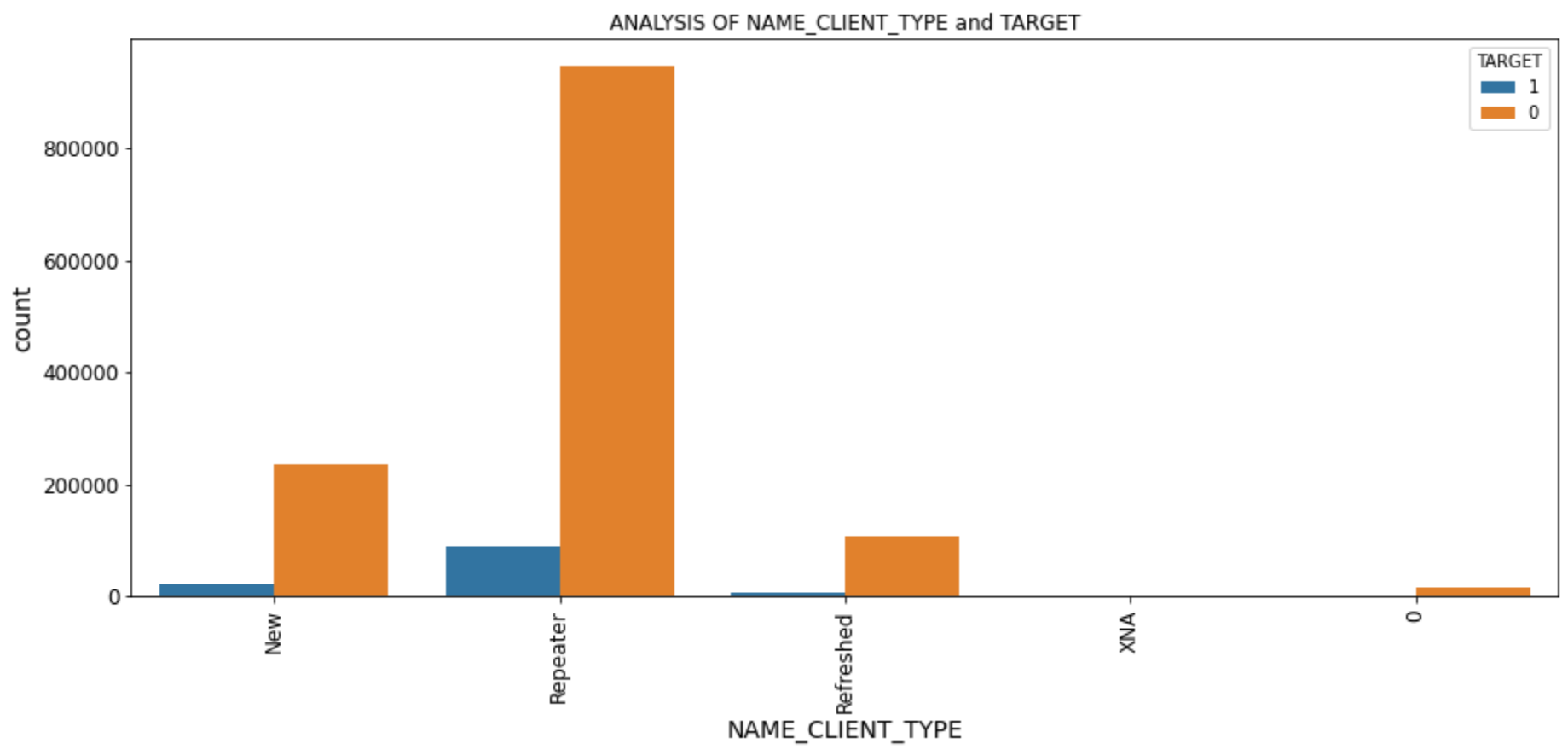


Secondary/secondary special has got higher loan approvals

Visualization



Visualization



Generally repeaters or old applicants are ontime payers than new applicants.

Conclusion

- Default rate for females are less compared to males
- Secondary or low level education has high tendency to default compared to higher education levels. But their loan approval rate was high.
- Occupation plays a key role in deciding whether the applicant is capable of repayment.^Applicants with good jobs tend to default less.
- Applicants with very low or low income tend to default even for smaller credit but their loan approvals are high in rate
- Married applicants have less default rate compared to their ontime payment rate. Single clients have more difficulty in repayment
- Repeater or old applicants tends to pay ontime compared to new applicants

Recommendation

- Company should focus more on females while lending loans as the have higher tendency to repay.
- Promote loans for higher education levels instead of lowe reducational levels.
- As businessman and student category default less ,promoting loans for lesser rate enhance the profit
- Reduce the focus on working class as default is more there
- More preference should be given to existing clients as default rate is less
- Be vigilant on providing loans to default income applicants as there is more default rate

Thank you