CREDIT EDA CASE STUDY

BY: PRATHAM KUMAR

INTRODUCTION

- This case study aims to give an idea of applying EDA in a real business scenario. In this case study, we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

PROBLEM STATEMENT

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Suppose we work for a consumer finance company which specialises in lending various types of loans to urban customers. We have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
- 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.

ANALYSIS APPROACH

1-Data Sourcing: (Provided in the Assignment)

2-Data Loading and Data Cleaning

- Fixing Rows and Columns
- Impute/Remove Missing Values
- Handling Outliers
- Standardising Values

3-Univariate Analysis

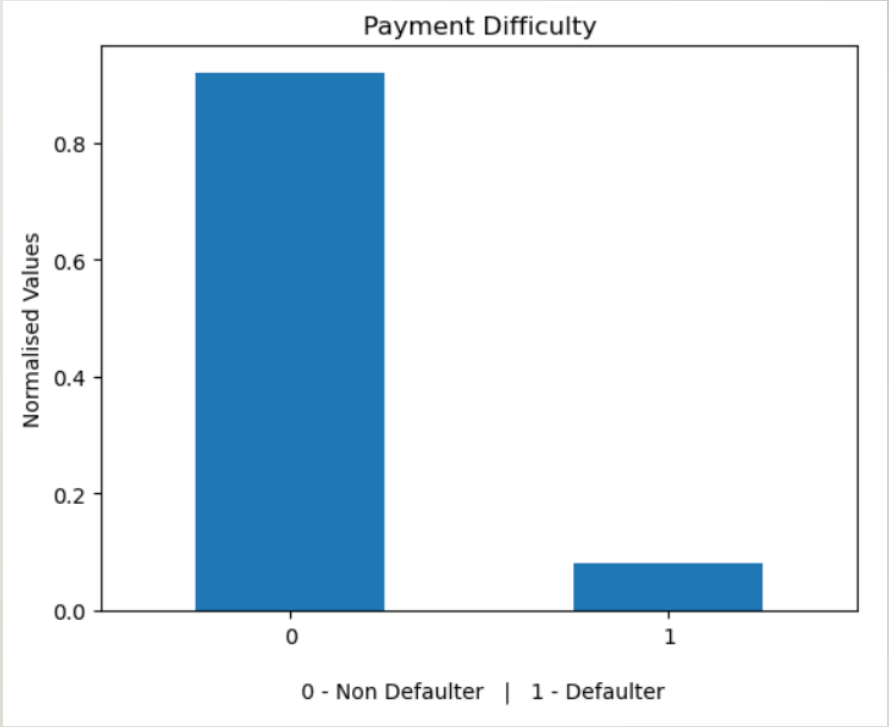
4-Bivariate and Multivariate Analysis

5-Conclusion

DATA IMBALANCE ANALYSIS OF TARGET VARIABLE

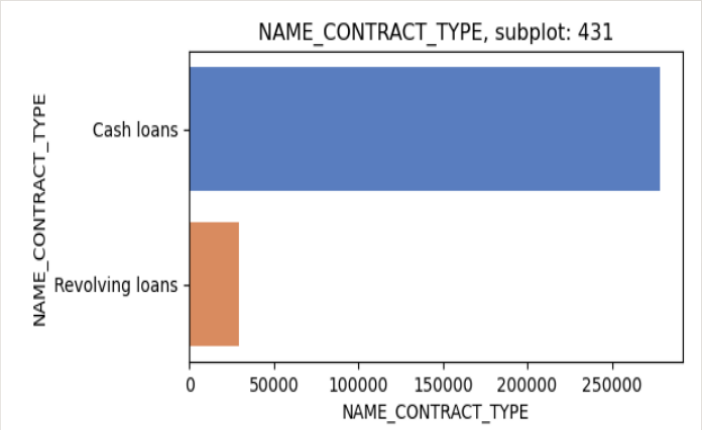
Inferences

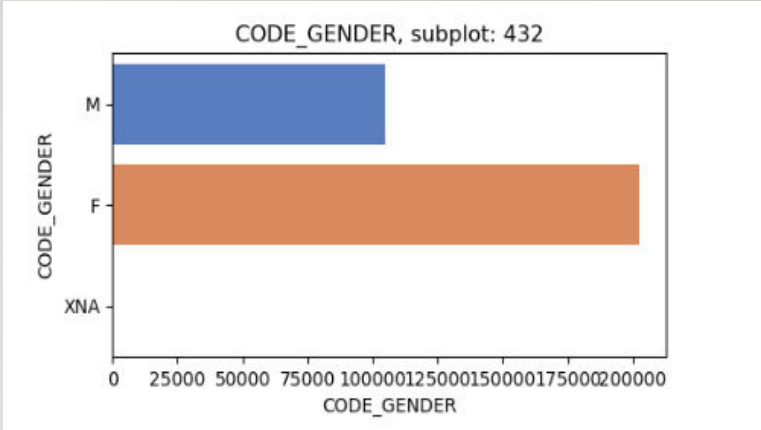
- 0 represents Non Defaulter.
- 1 represents as Defaulter.
- 1 in every 11 applicant has payment difficulty or is a defaulter.

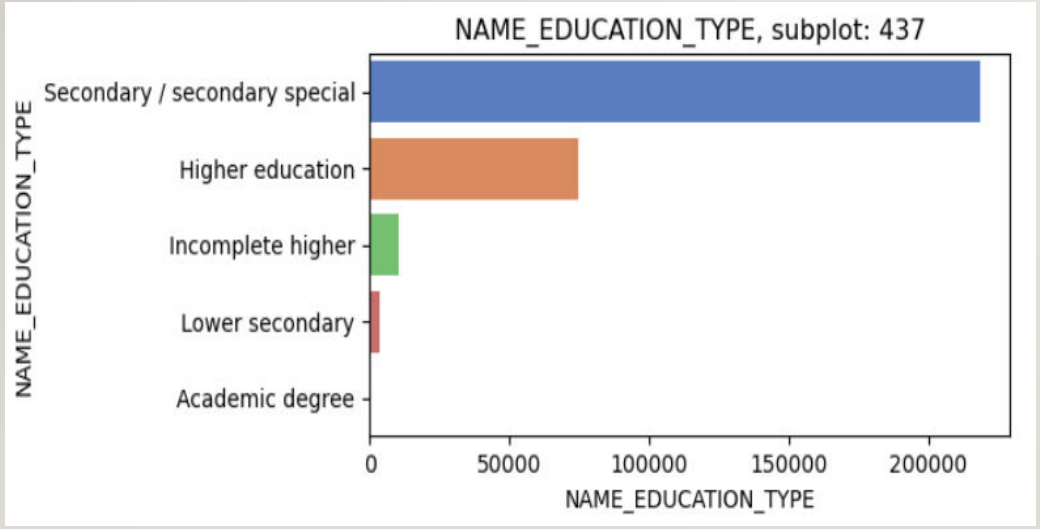


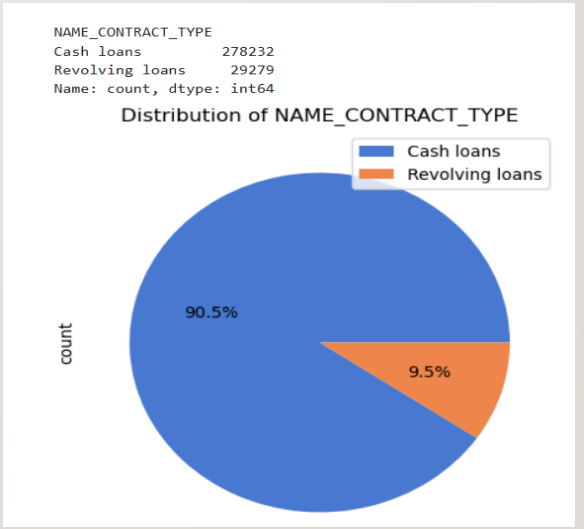
UNIVARIATE ANALYSIS (CATEGORICAL VARIABLE)

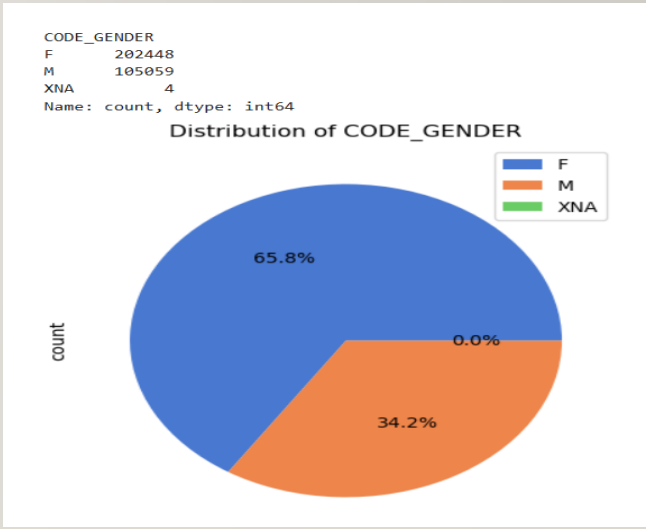
- Cash loans are more prevalent than revolving loans, accounting for 90% of the total.
- A notable 65% of loan applicants are female, compared to 34% male.
- 71% of applicants have completed secondary education.





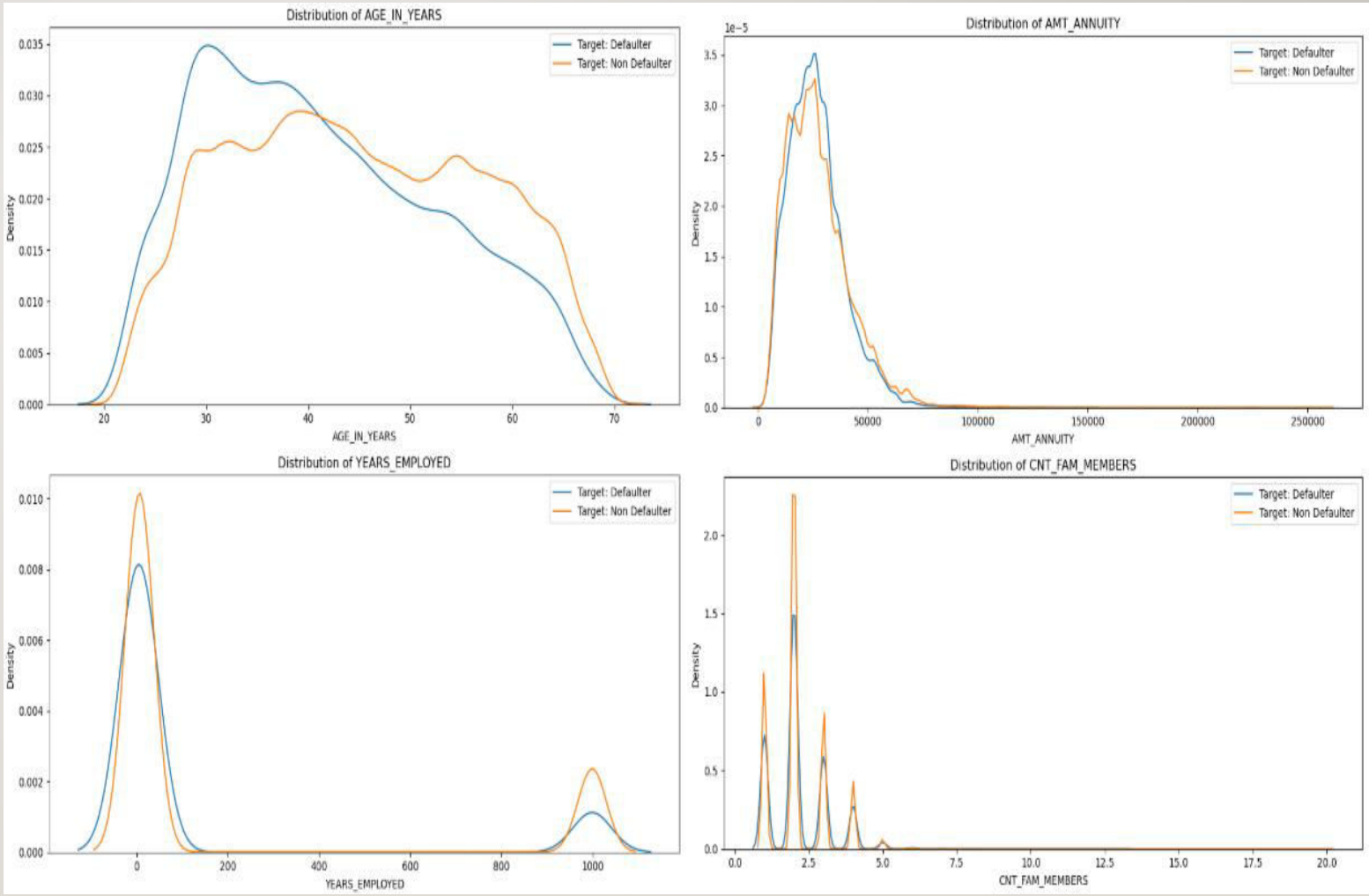






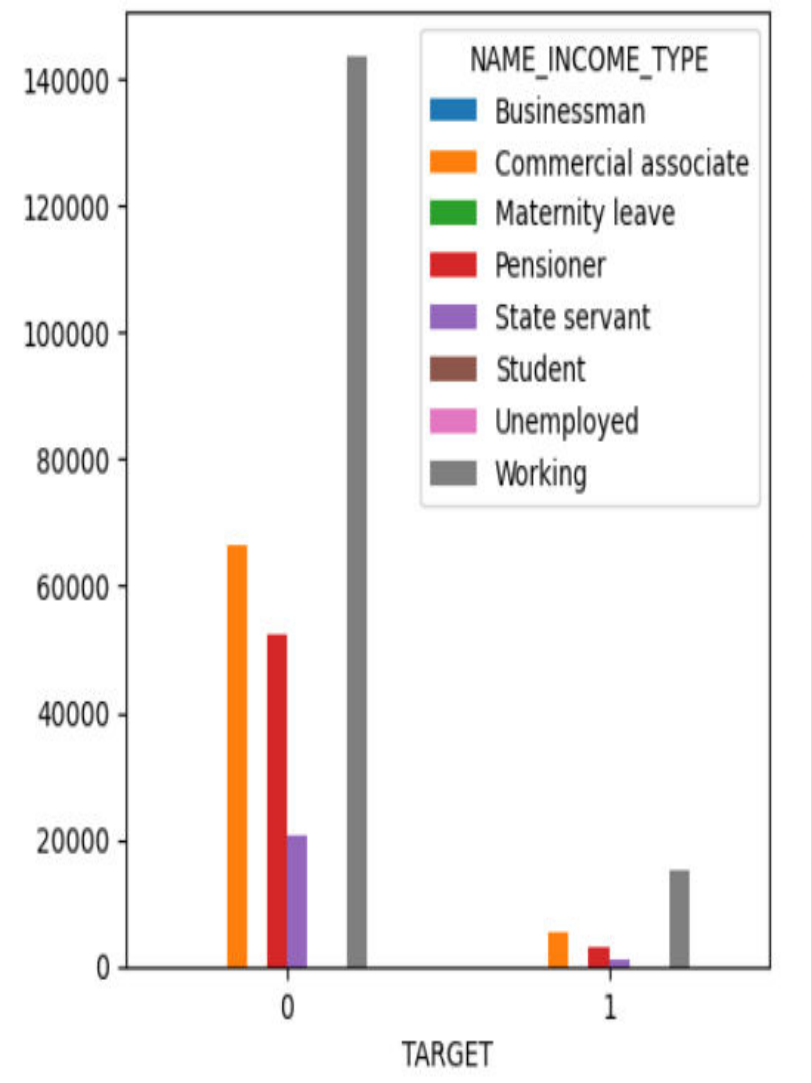
UNIVARIATE ANALYSIS (NUMERICAL VARIABLE)

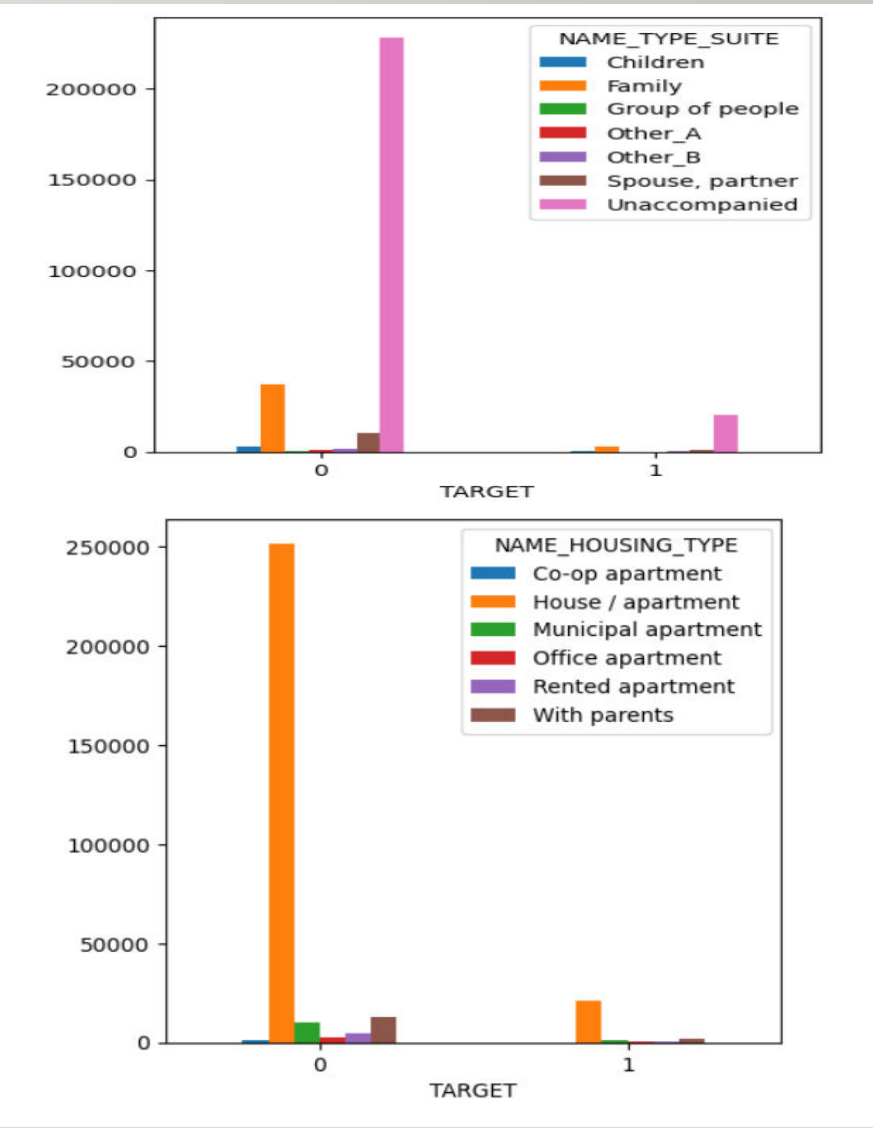
- A higher number of loans are associated with lower annuity amounts.
- Nuclear families are more likely to take out loans.
- People aged between 27 years (10,000 days) and 41 years (15,000 days) are more inclined to take out loans.
- Retired individuals are more likely to take out loans.



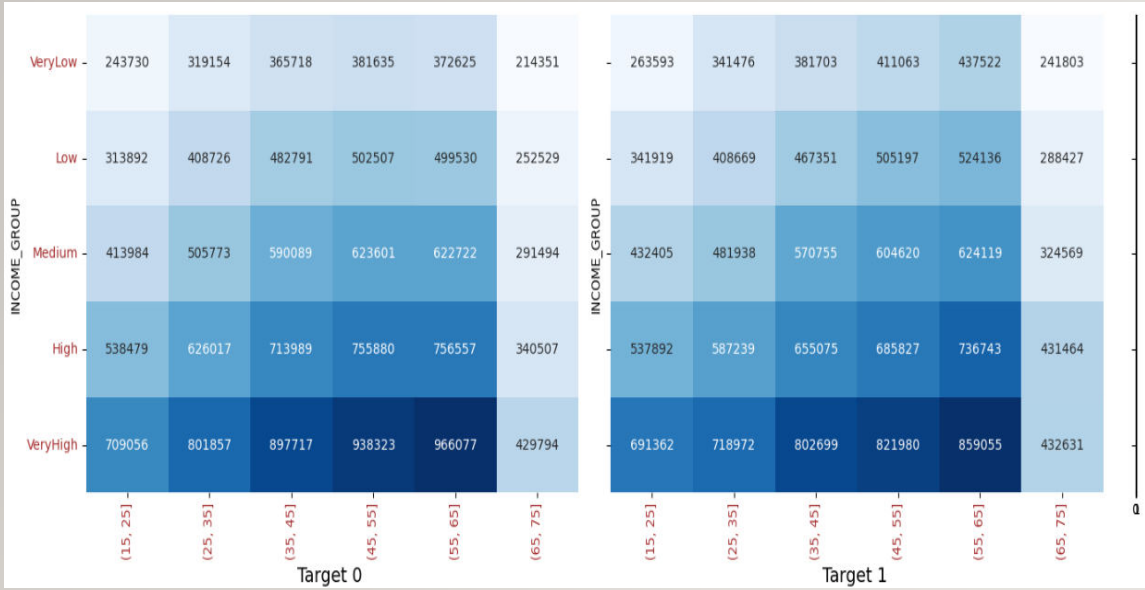
BIVARIATE ANALYSIS (CATEGORICAL)

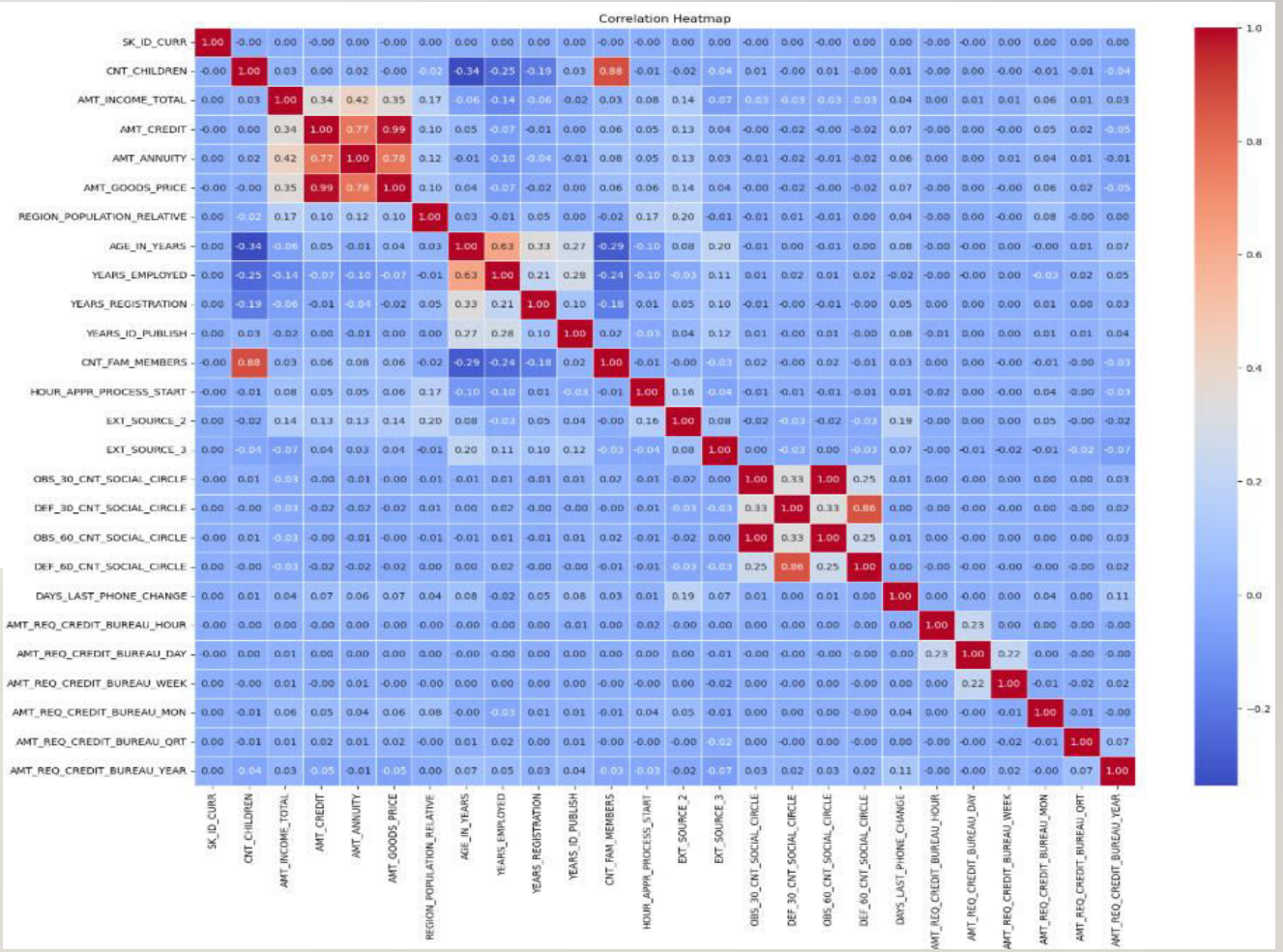
- Most of the people come alone when taking a loan
- Working people take more loans.
- People having house/appartment tend to take more loans





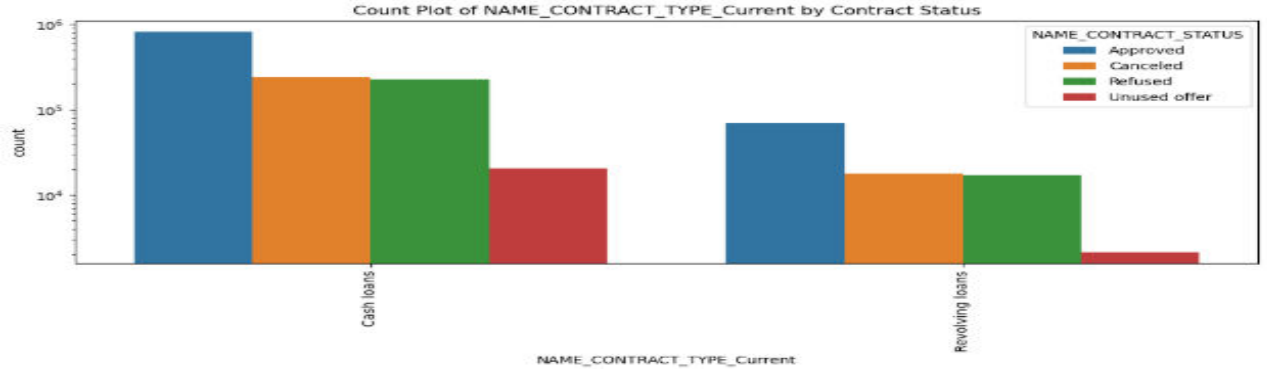
MULTIVARIATE ANALYSIS (CORRELATION)

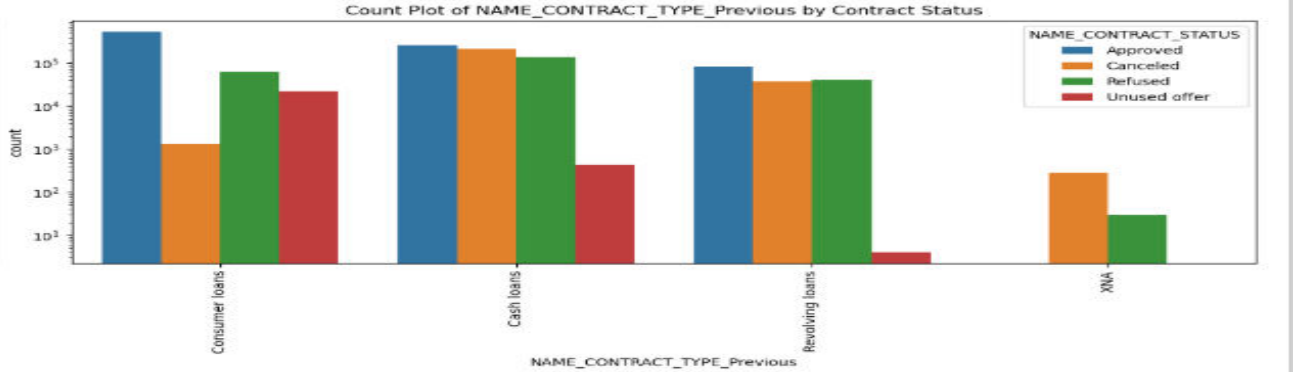


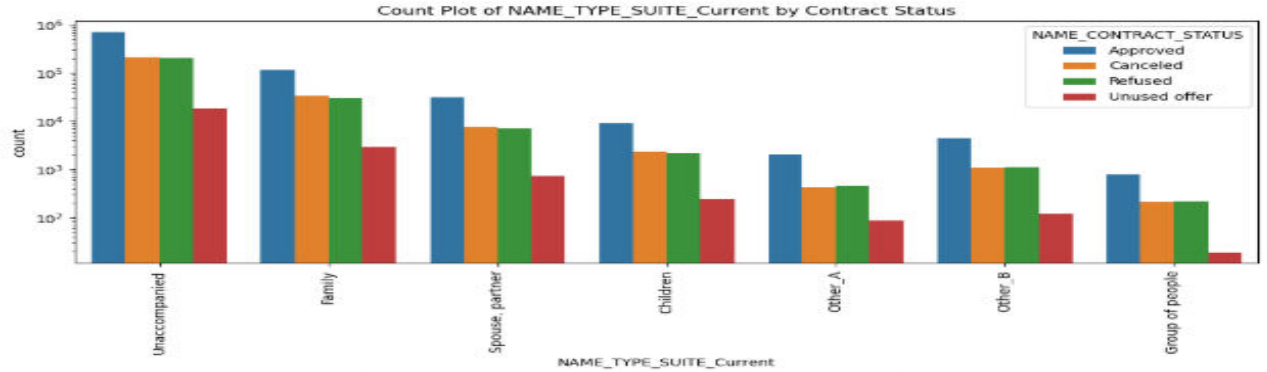


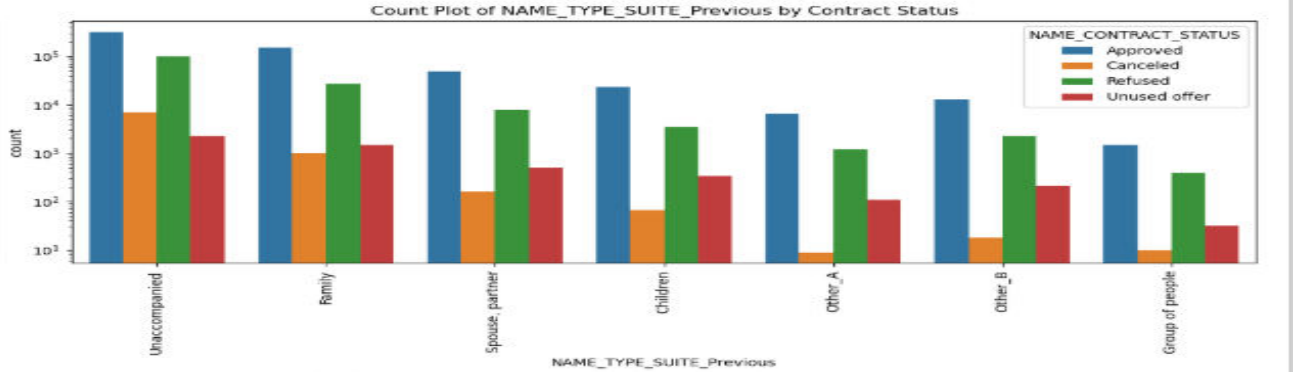
Age Group 55-65 in very high-income group has high amount credit.

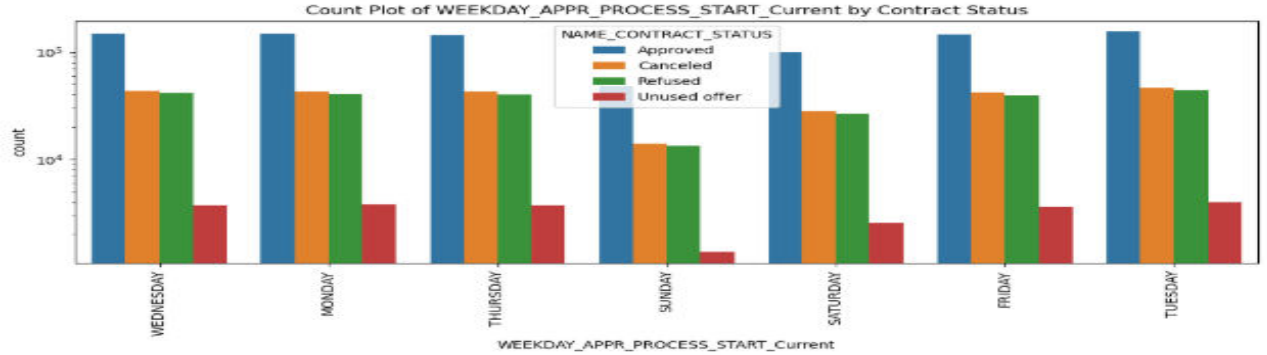
AFTER MERGING TWO APPLICATIONS

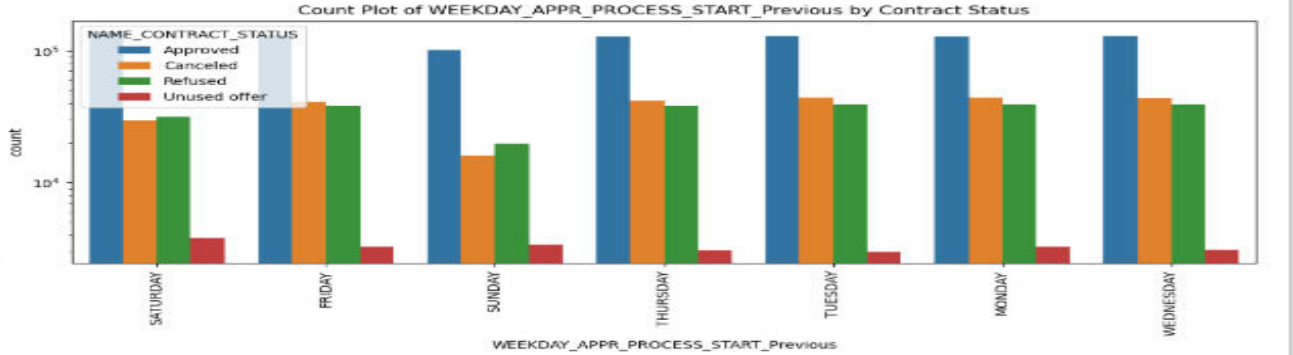


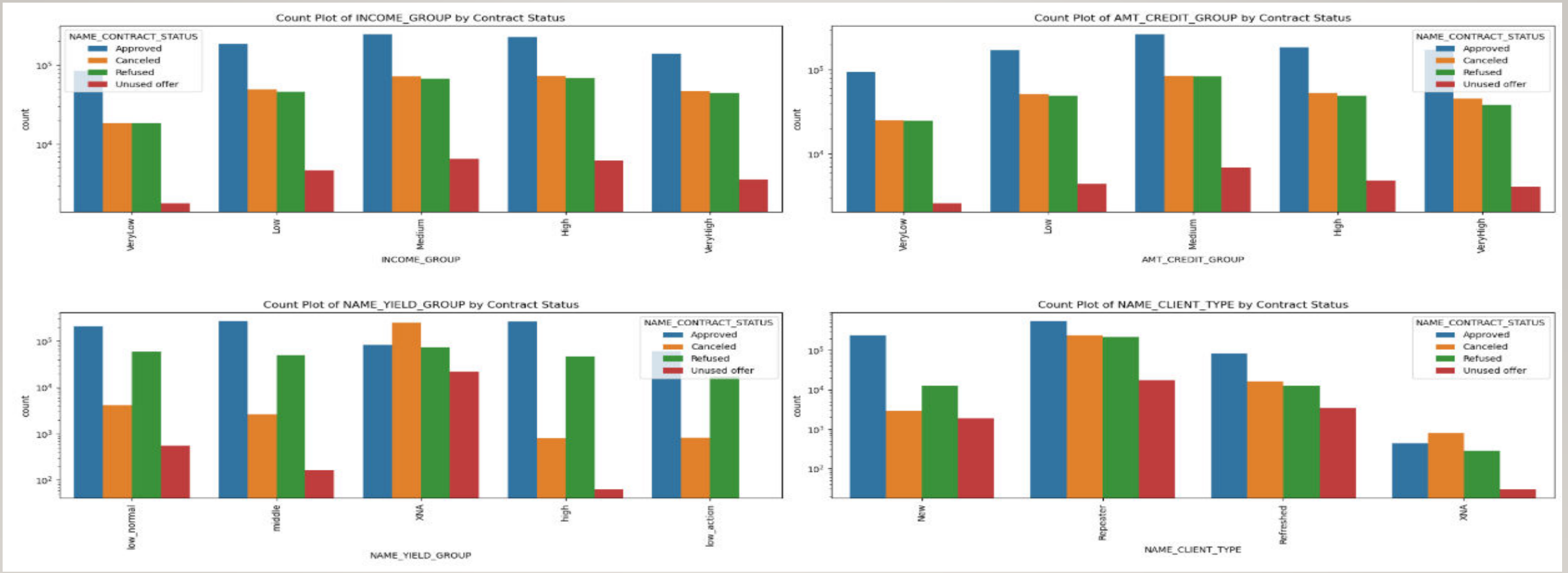




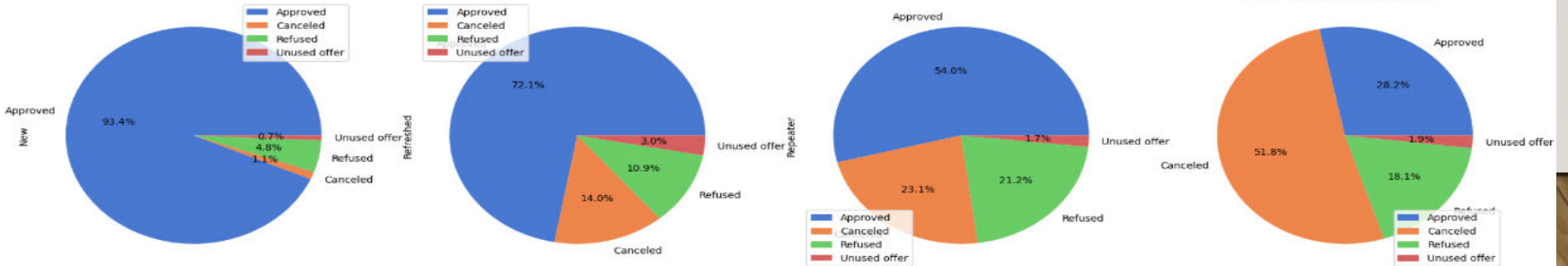




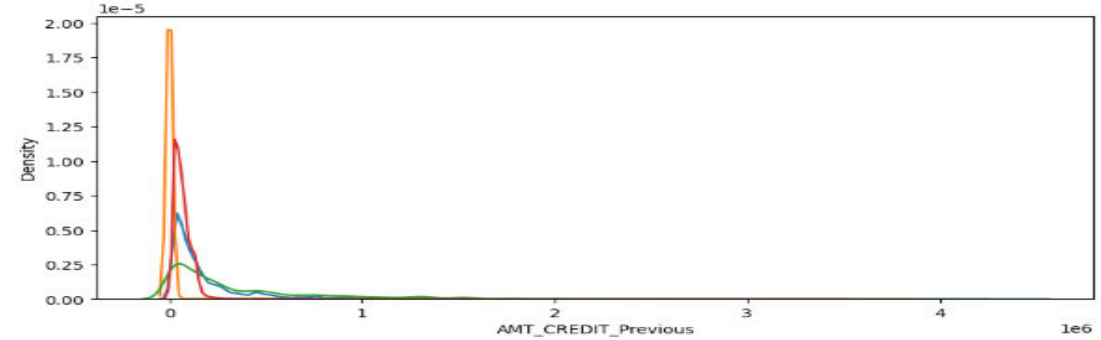


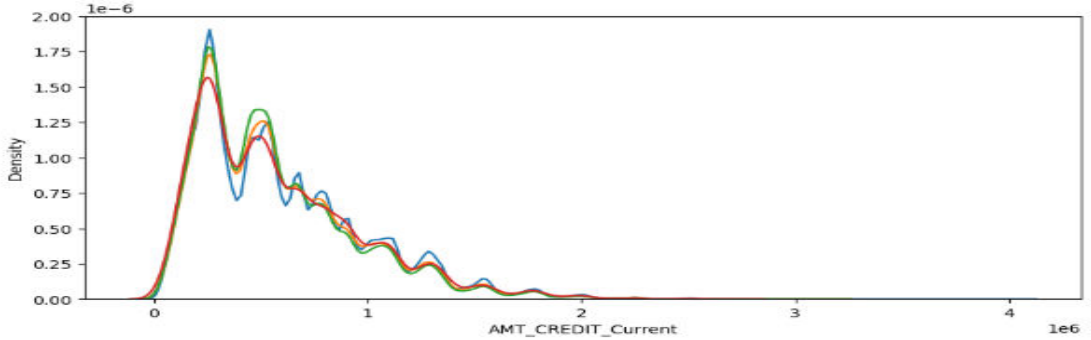


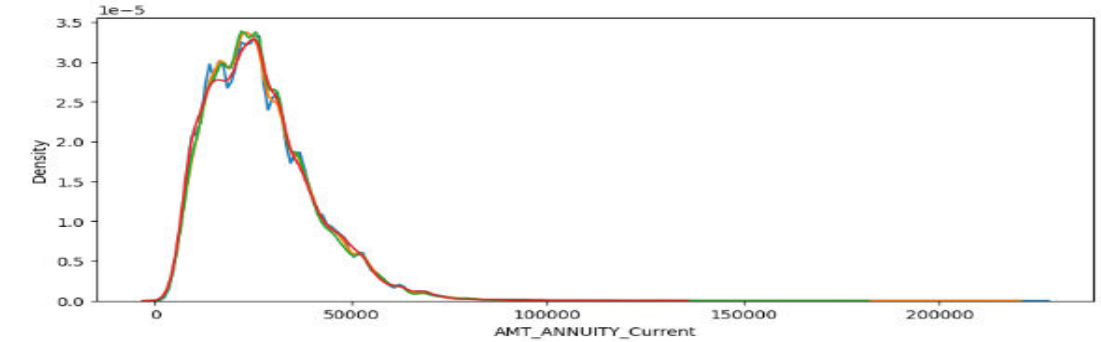
NAME_CLIENT_TYPE	New	Refreshed	Repeater	XNA
NAME_CONTRACT_STATUS				
Approved	242395	82891	560378	435
Canceled	2920	16104	239618	799
Refused	12358	12537	220215	280
Unused offer	1867	3404	17471	29

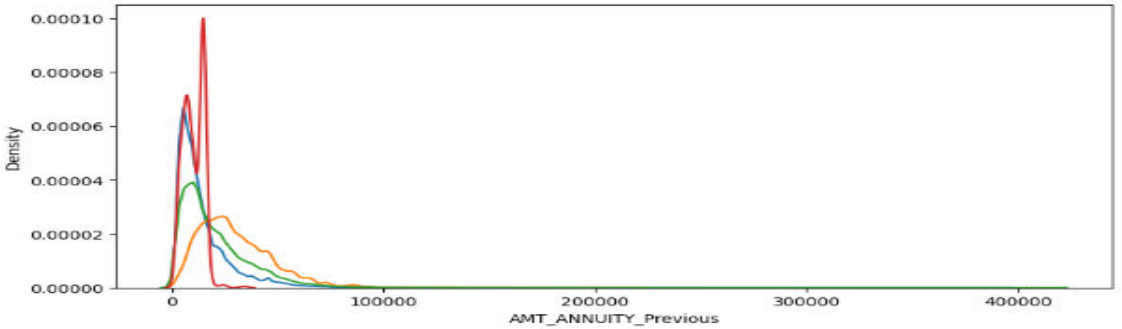


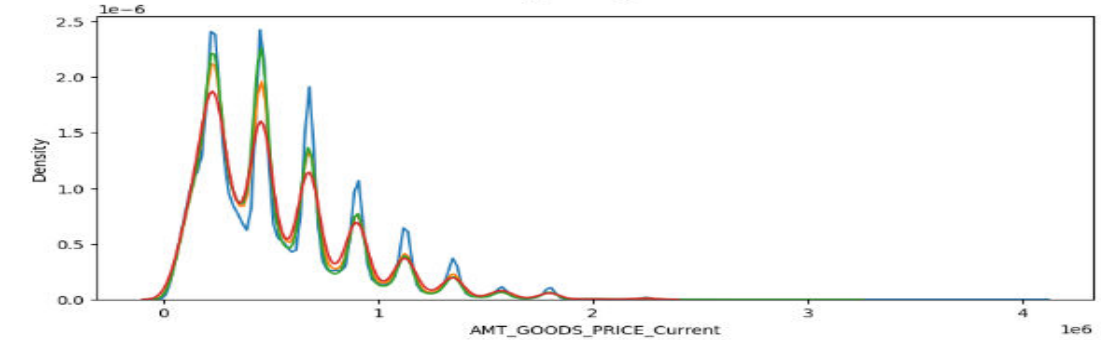
NUMERICAL VARIABLE

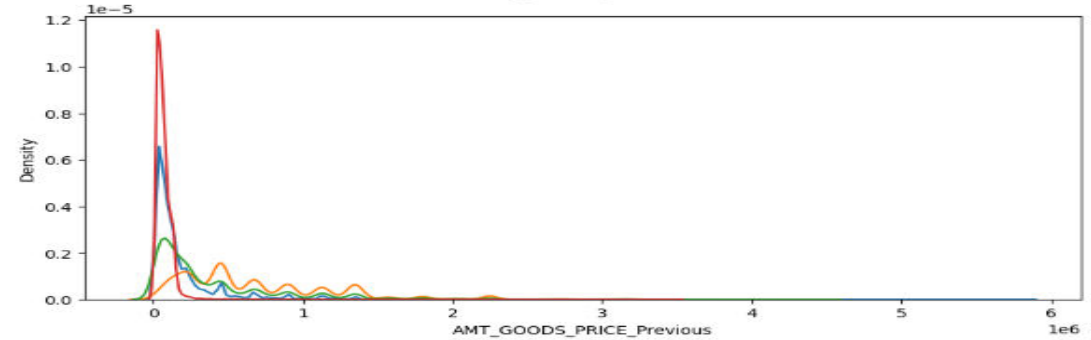












CONCLUSION

- The number of consumer loans was highest previously, but now the highest number is Cash loans.
- Currently, the bank is only offering two types of loans: Cash and Revolving Loans. Previously, the bank provided Cash, Revolving, and Consumer loans.
- Repeat applicants have the highest number of approved loans.
- Previously, the bank had a high number of unused offers, but currently, the number of refused offers is high for AMT_GOODS_PRICE.
- Previously, the bank had a high number of unused offers, and currently, the number of cancelled/refused offers is similar for AMT_ANNUITY.
- Previously, the bank had a high number of unused offers, and currently, there is a high number of refused offers for AMT_CREDIT.
- **Target/focused variable for Application dataset - TARGET**
- **Target/focused variable for Previous dataset - NAME_CONTRACT_STATUS**
- **Top Major variables to consider for loan prediction:**

1- NAME_EDUCATION_TYPE

2- AMT_INCOME_TOTAL

3- DAYS_BIRTH(AGE_IN_YEARS)

4- AMT_CREDIT

5- DAYS_EMPLOYED(YEARS_EMPLOYED)

6- AMT_ANNUITY

7- NAME_INCOME_TYPE

8- CODE_GENDER

9- NAME_HOUSING_TYPE

- **The above mentioned variables are to be considered before approving application to minimize risk of loss.**