CREDIT EDA CASE STUDY

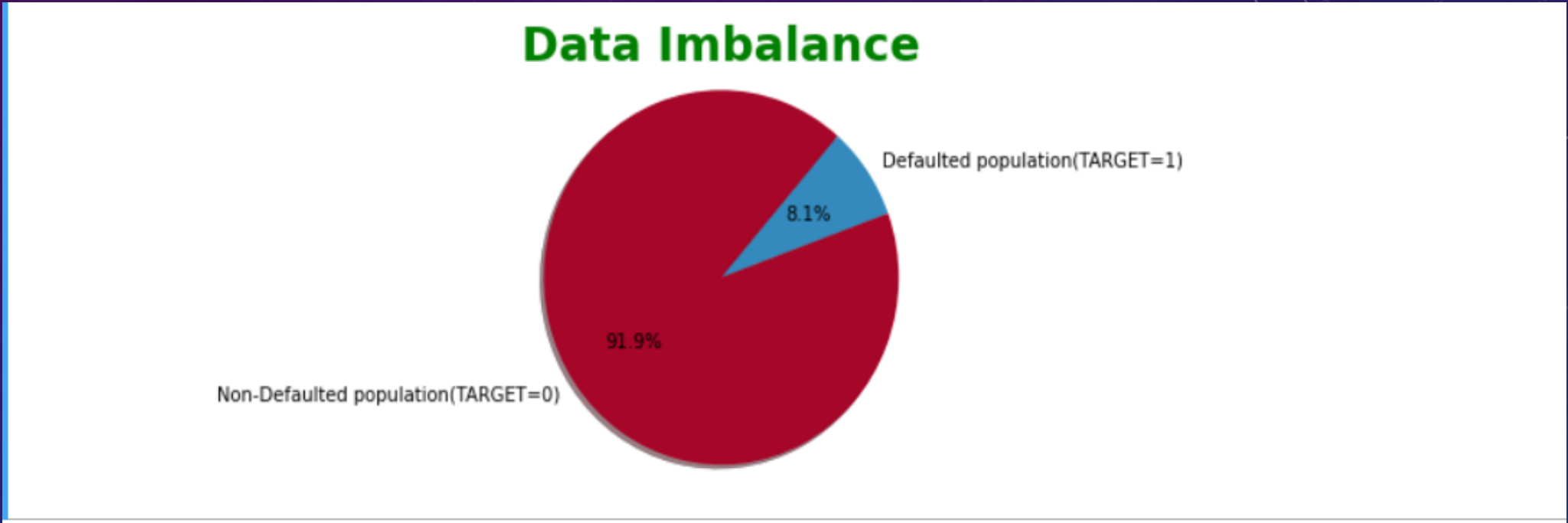
BY- AYUSH MISHRA

STEPS FOR DATA ANALYSIS:

- Data Understanding and Sourcing
- Check the data Quality and Data Cleansing
- Univariate Analysis
- Bivariate and Multivariate Analysis
- Merging of applicant data with previous applicant
- Data Analysis and correlation

APPLICATION DATA

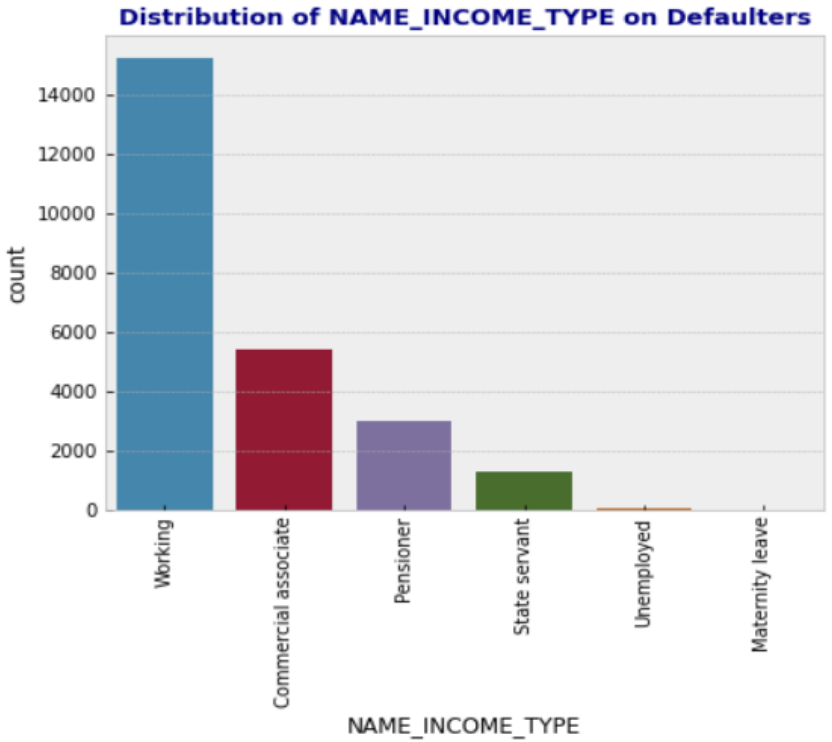
CHECKING DATA IMBALANCE

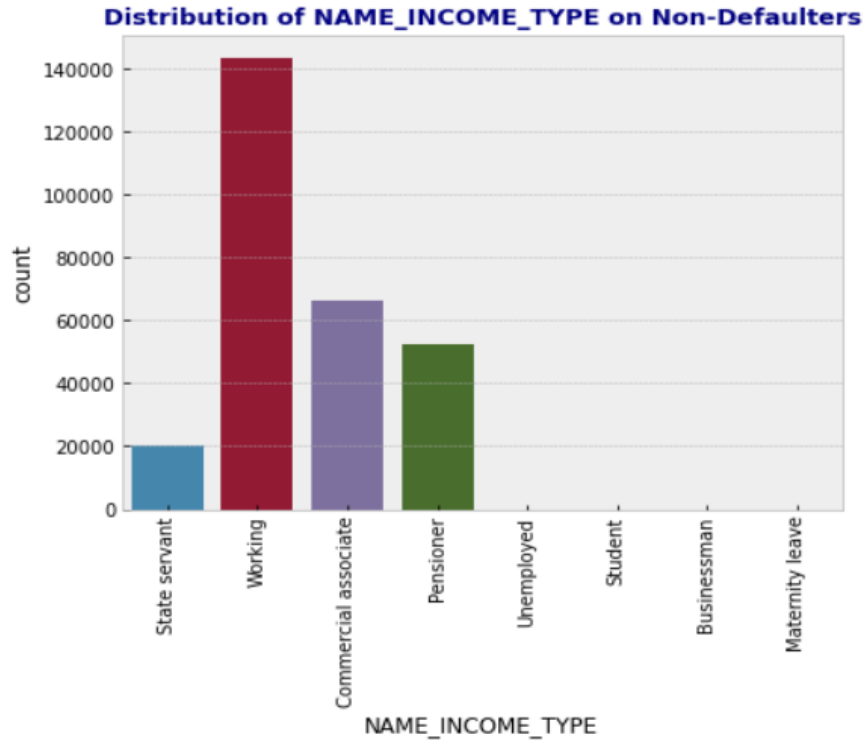


As per the pie chart 91.9% population are non-defaulted and 8.1% population are defaulted. Ratio of the imbalance is 11.4.

UNIVARIANT ANALYSIS FOR CATEGORICAL AND NUMERICAL

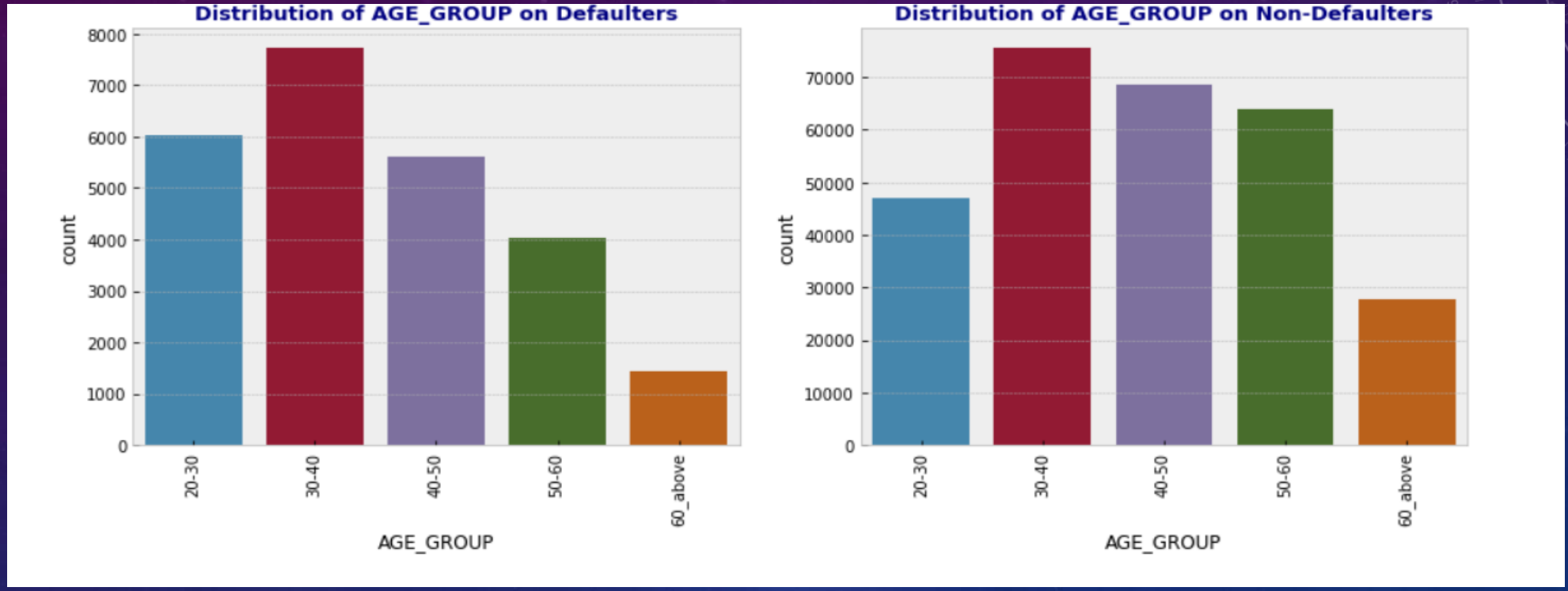
INCOME TYPE





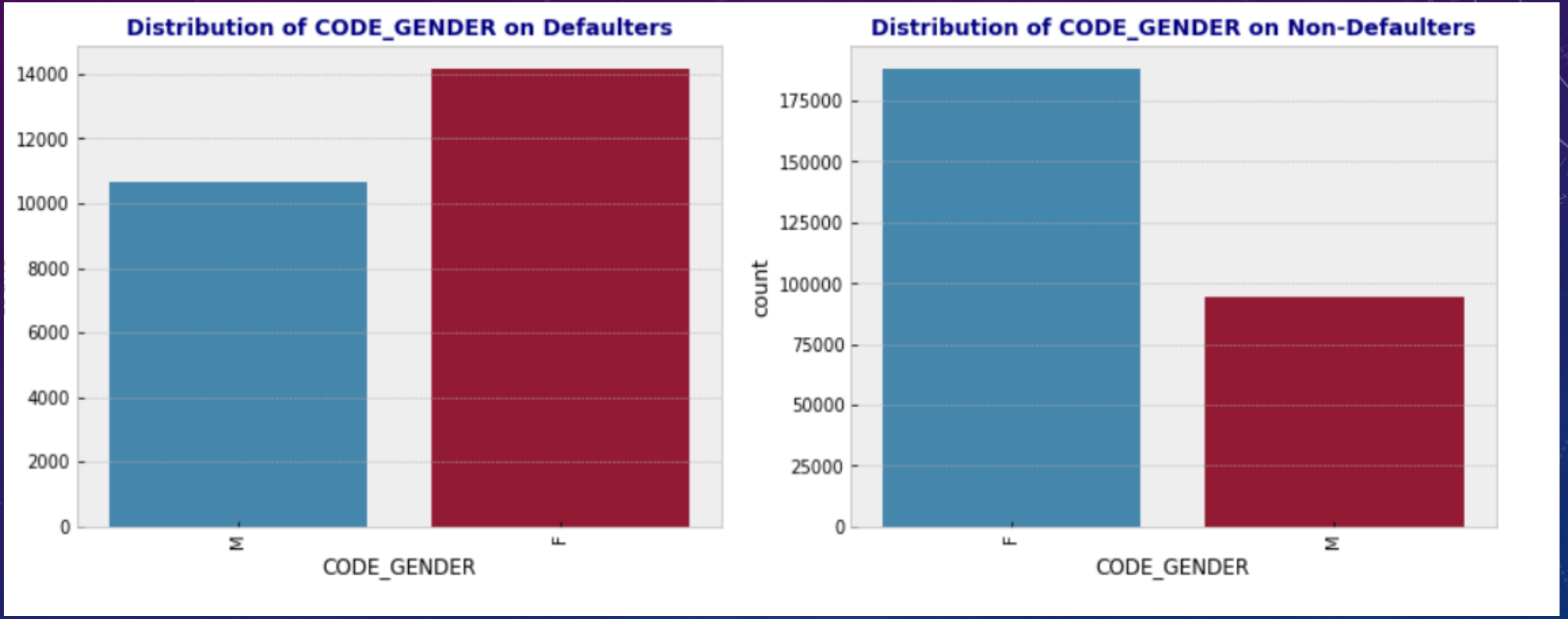
It seems like most of the defaulters are the working people, so as in Non-Defaulter. And there seems to be a moderate rate in Commercial associate, Pensioner, and state servant. Least count to be seen in Unemployed, Maternity Leave, Businessman and Student.

AGE GROUP



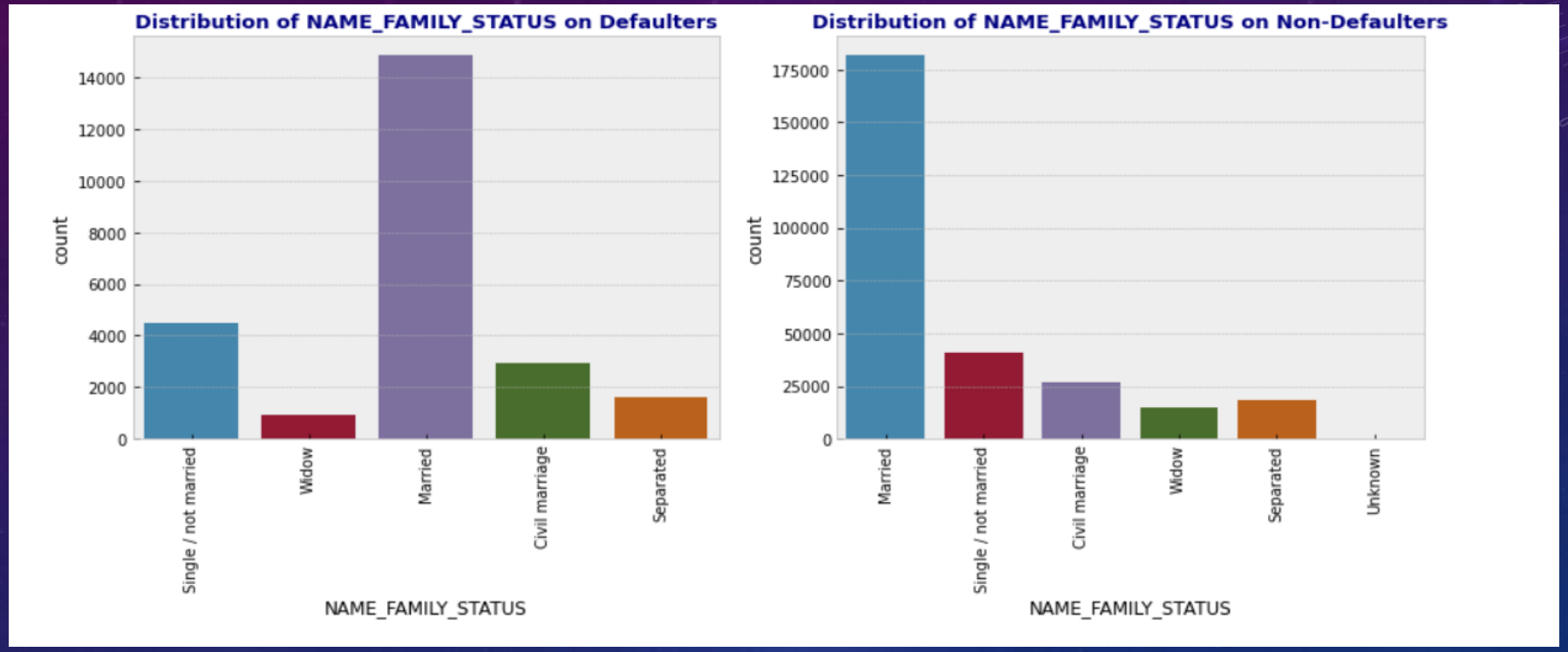
As per the above plot, there seems to be an increase in number of people of age group 20-30 that to defaulters (facing payment difficulties in loan) while comparing with non-defaulters.

CODE GENDER



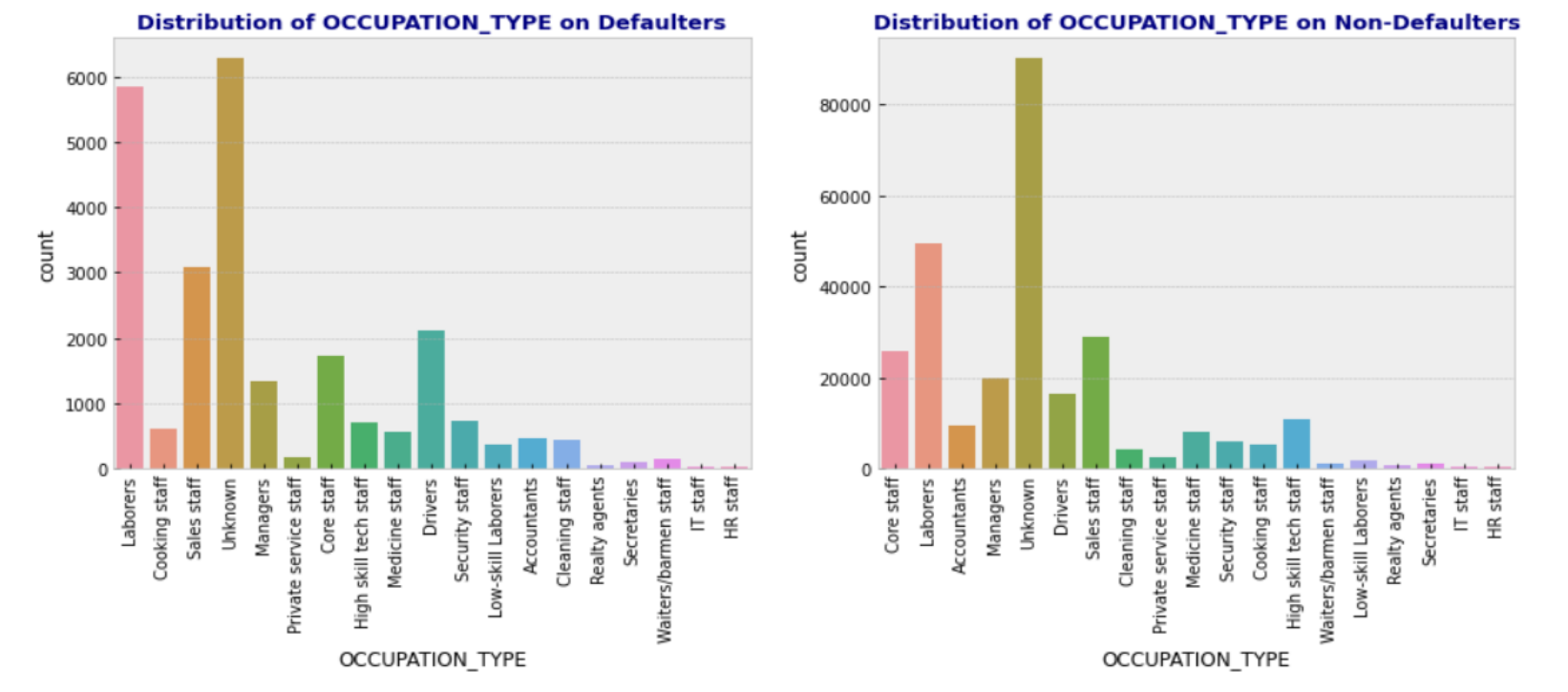
Females are the majority in both Defaulters and Non-Defaulters. Undoubtedly, most females have applied for loan than males.

FAMILY STATUS



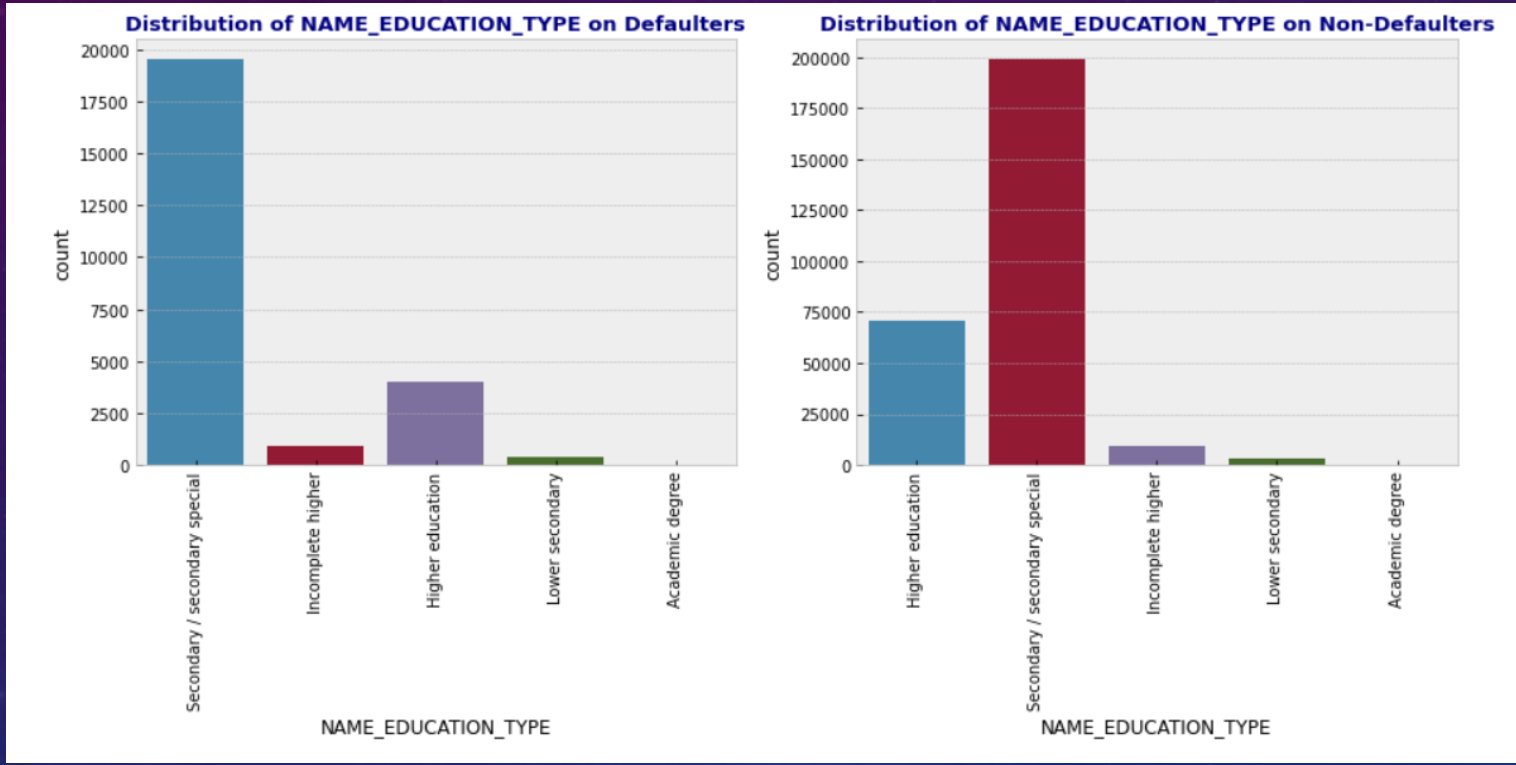
Most of pepoles are Married. Single / not married is proportionally higher in dafaulted population as compared to non defaulted population. This concludes that Single applicants have higher defaults.

OCCUPATION TYPE



In the above plot, Sales ,core staff has highest number of loan applications but when it comes to Non repayment of loan then the highest number can be seen in Low-skill laborer's, Drivers, Sales staff and so on.

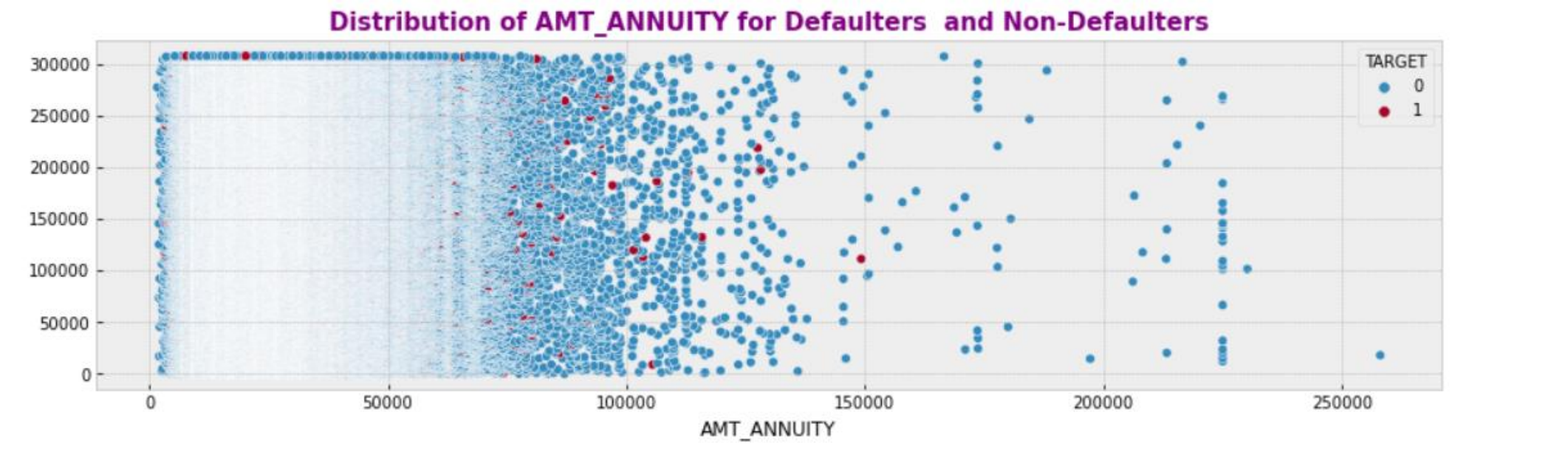
EDUCATION TYPE



Higher education count is proportionally lesser in dafualted population as compared to non defaulted population. Hence higher the education level, lower the default rate.

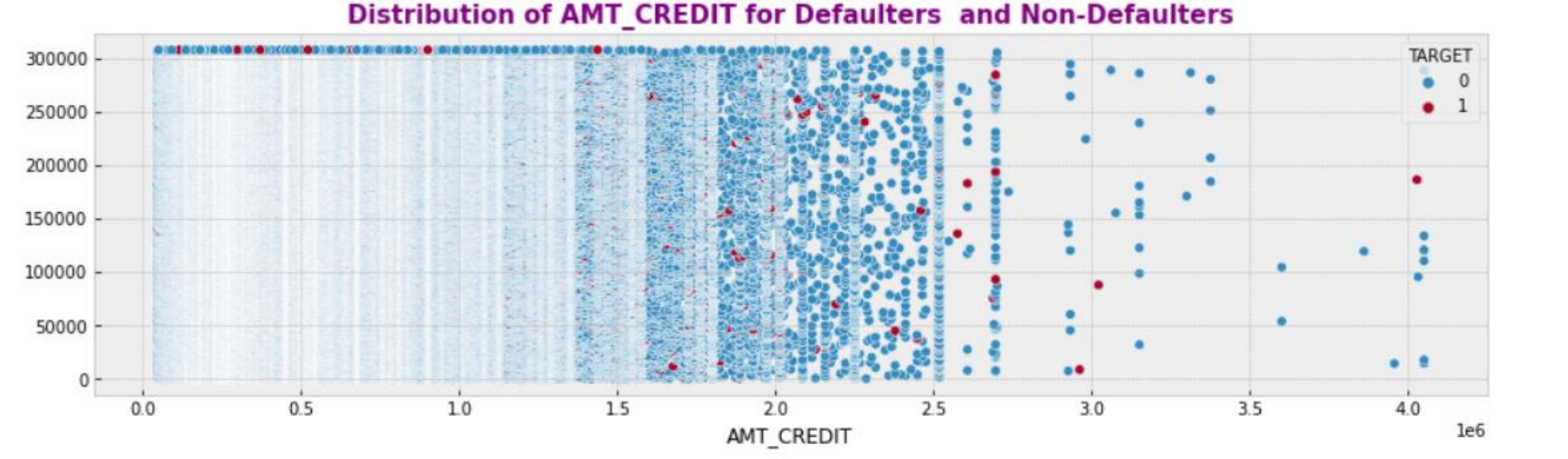
UNIVARIATE ANALYSIS FOR NUMERICAL

AMT_ANNUITY



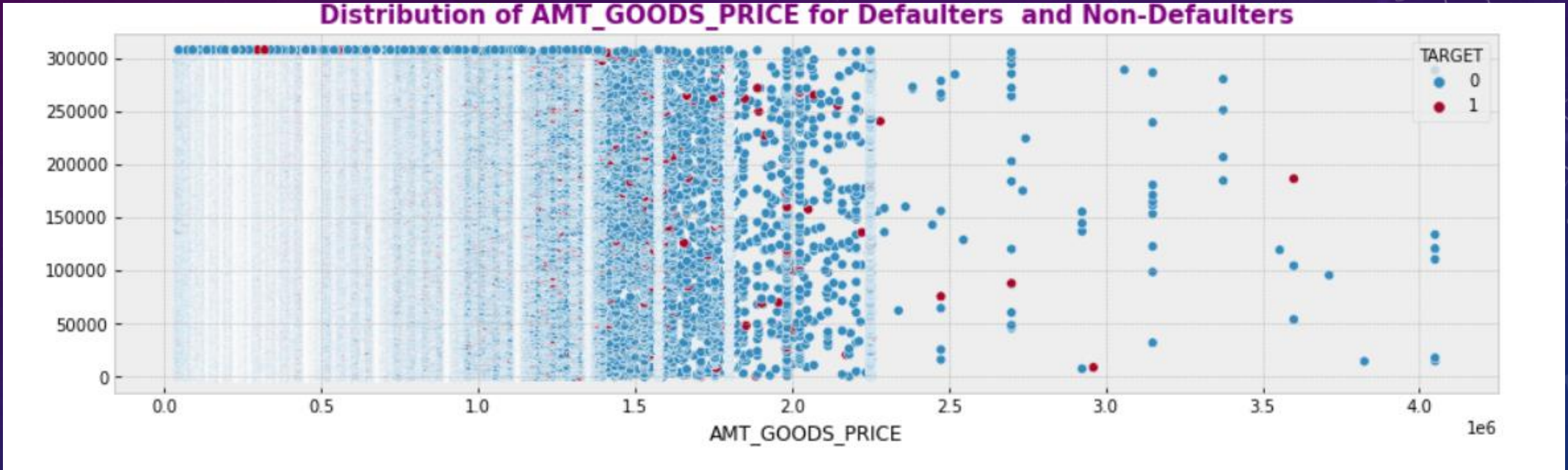
In the above scatterplot,outliers are visisble in annuity amount and most of values are upto 75k.

AMT_CREDIT



In the above scatterplot,there is a visisbility of outliers and most of the values are in between 0.1 to 2 lakhs.

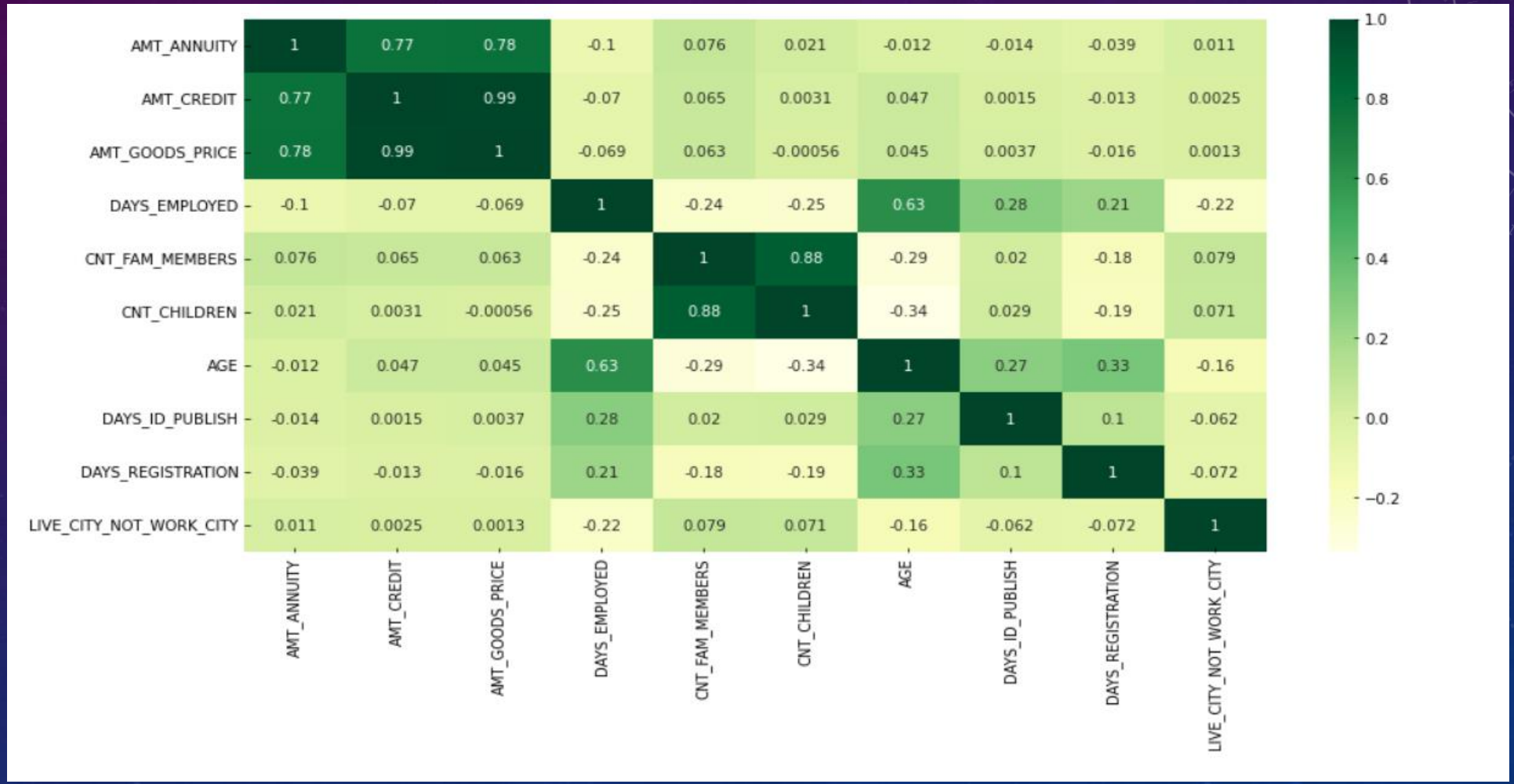
AMT_GOODS_PRICE



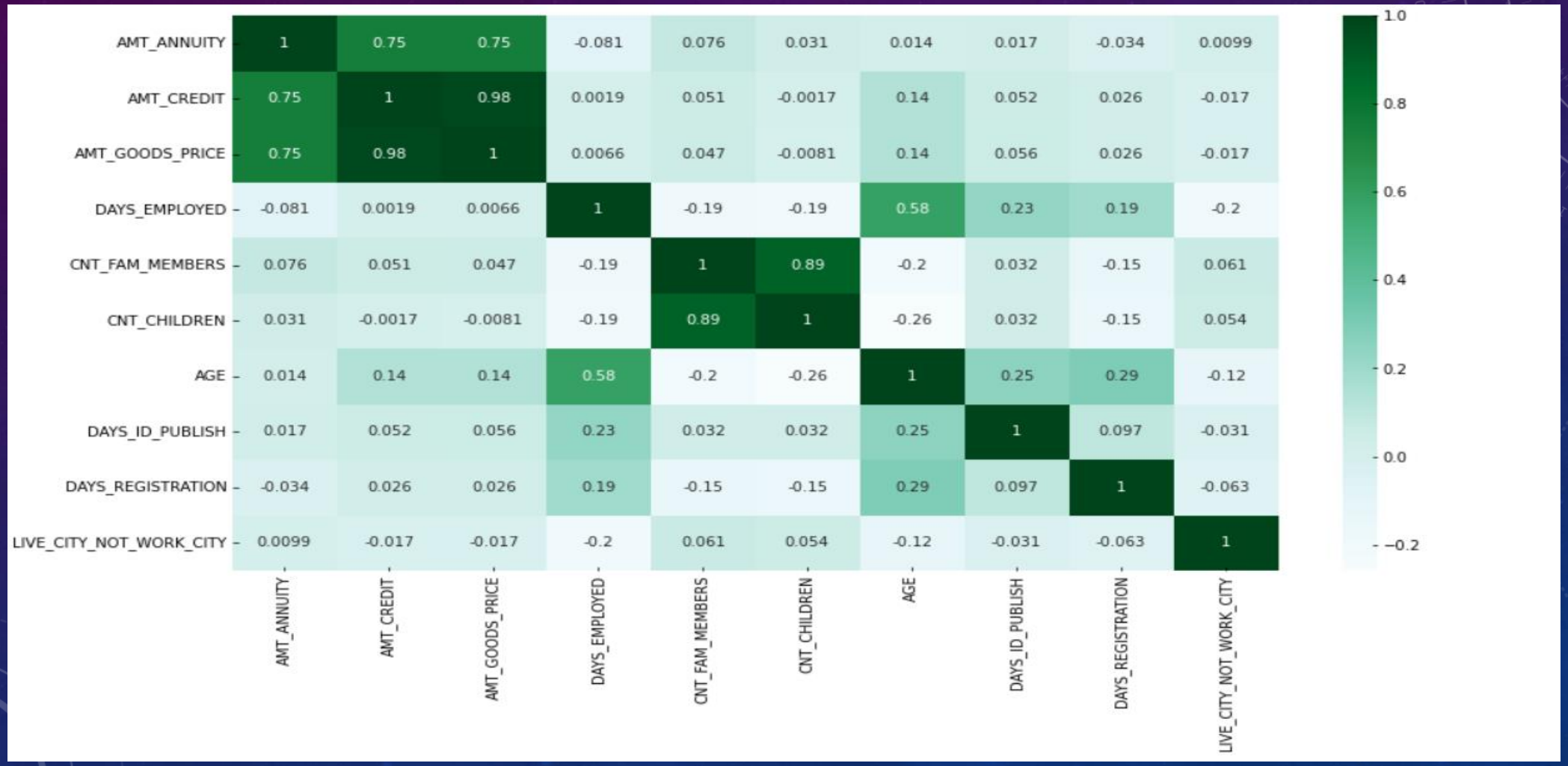
In the above scatterplot,there is a visibility of outliers and most of values are upto 1.5 lakh.

BIVARIATE ANALYSIS AND CORRELATION

CORRELATION FOR TARGET = 0

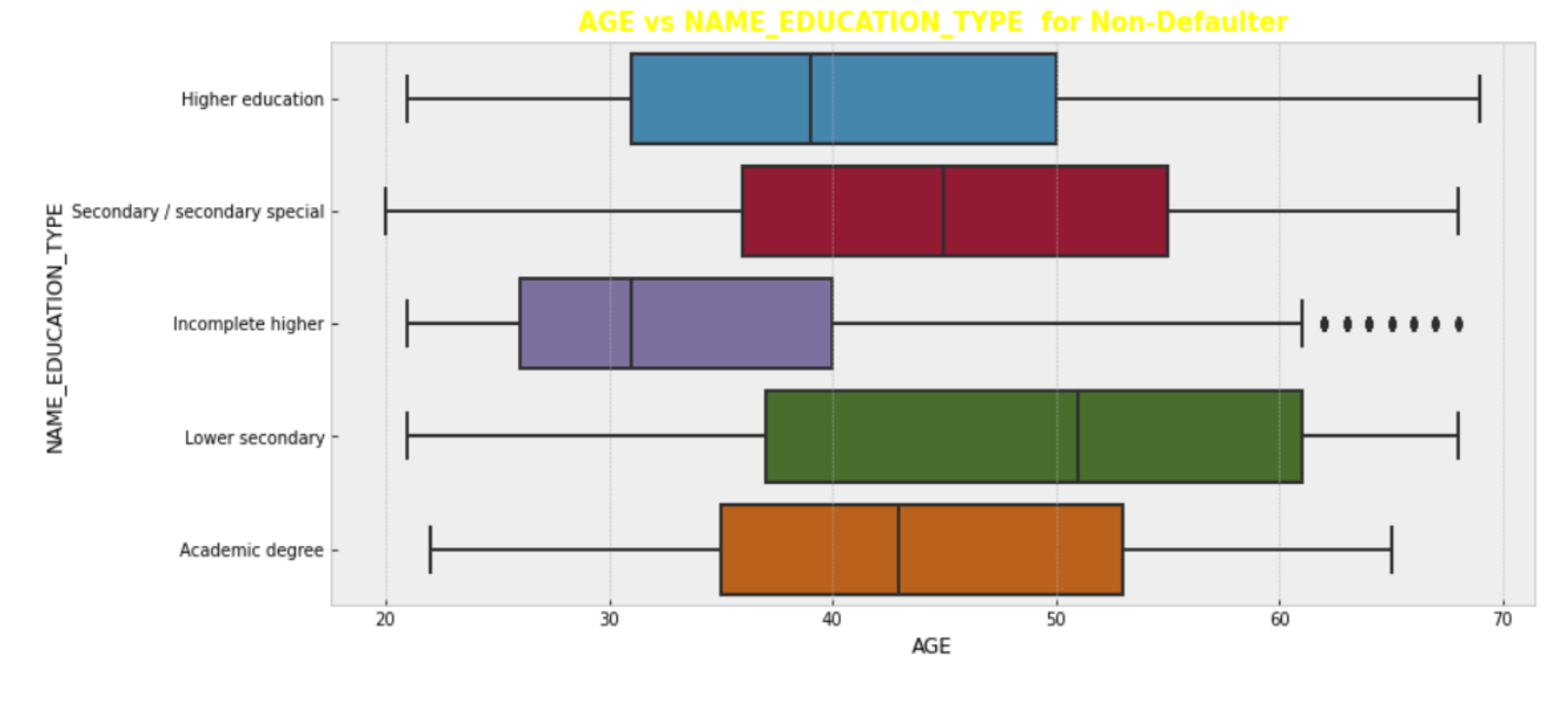


CORRELATION FOR TARGET = 1

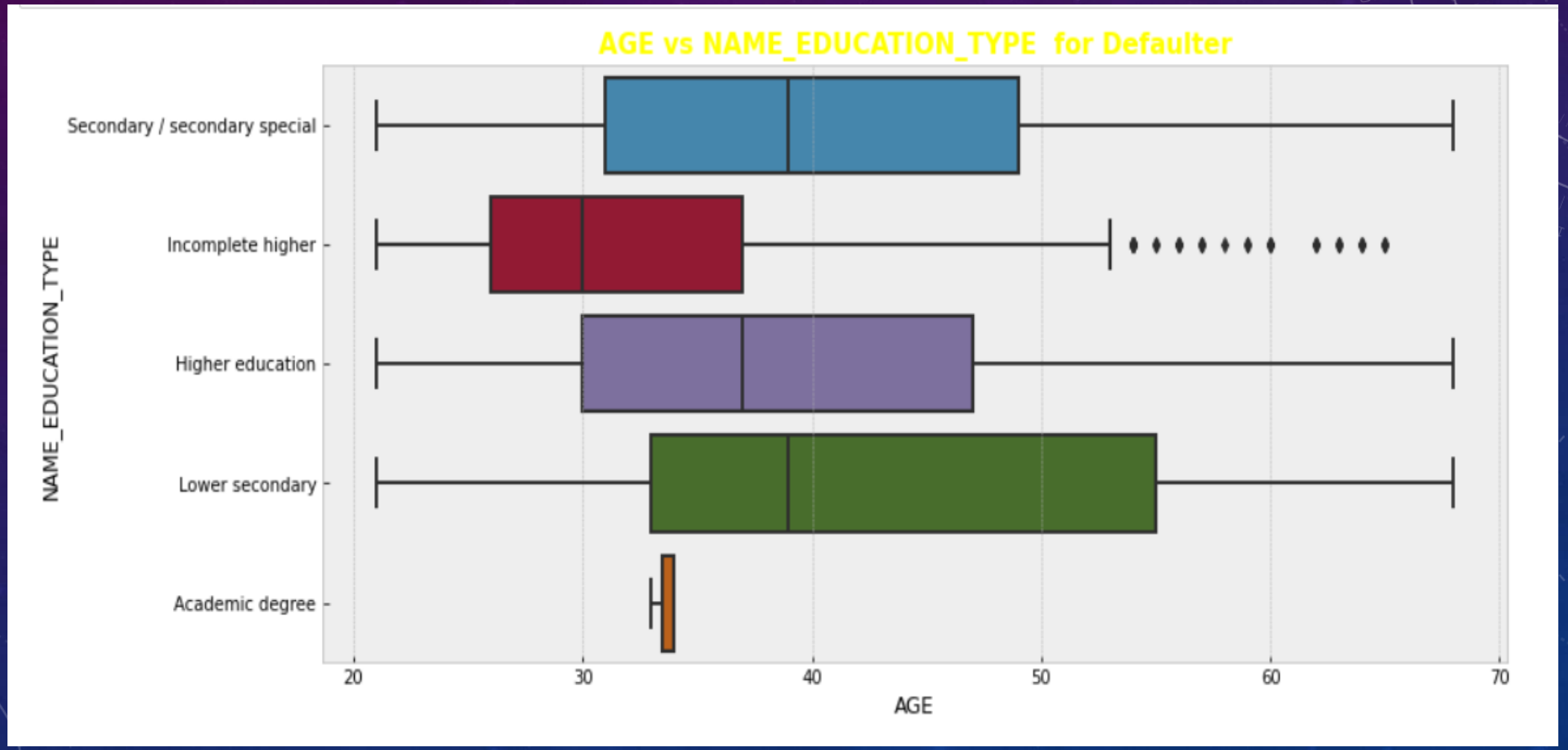


CATEGORICAL VS NUMERICAL

AGE VS EDUCATION TYPE FOR NON-DEFAULTER



AGE VS EDUCATION TYPE FOR DEFAULTER



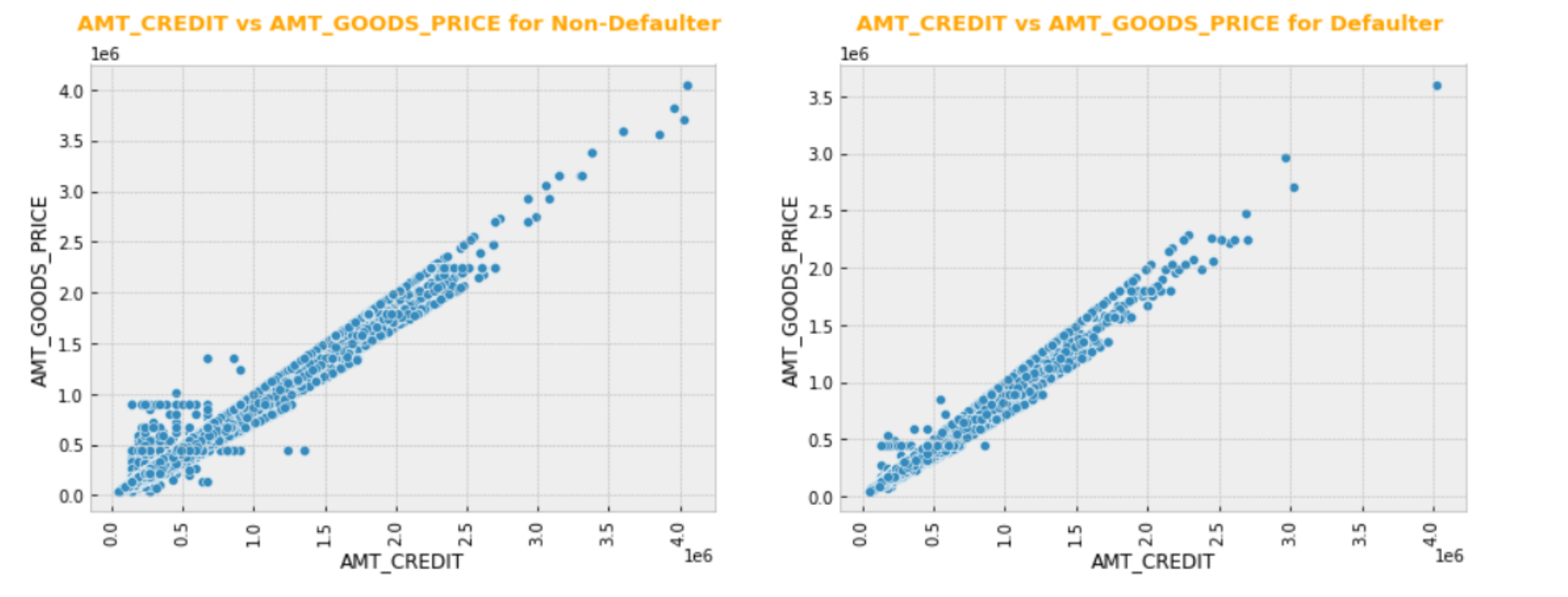
For Defaulters:

- There is the visibility of an outliers in Incomplete higher from the age group between 50-70. People with an age group 30-40 and education type as Academic degree and Incomplete higher seem to be facing difficulties in loan repayment.

For Non-Defaulters:

- There is an outliers in Incomplete higher in between age 60-70. Apart from Incomplete higher where the first quartile lies at age 40, other education types aren't facing much difficulties in loan repayment.

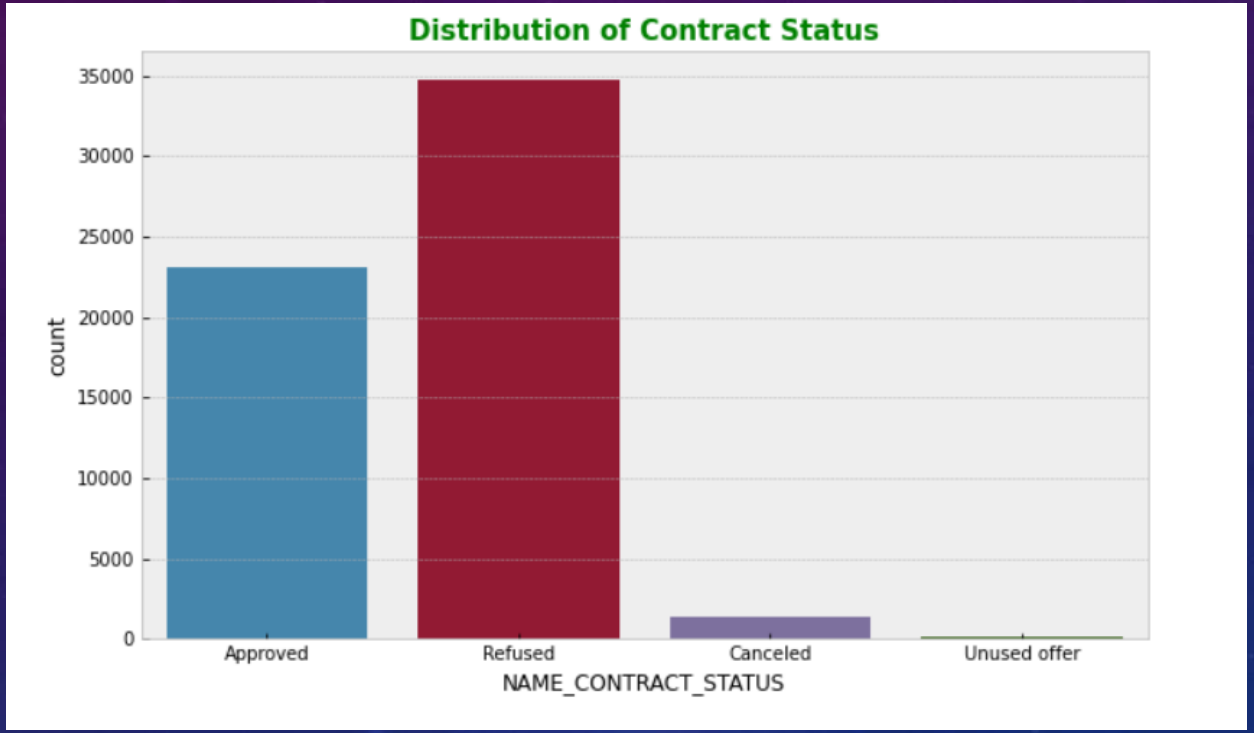
BIVARIATE_NUMERICAL('AMT_CREDIT', 'AMT_GOODS_PRICE')



The defaulter seems to be in less number whenever the price of goods is upto 50k and credit amount is even lesser than 50k.

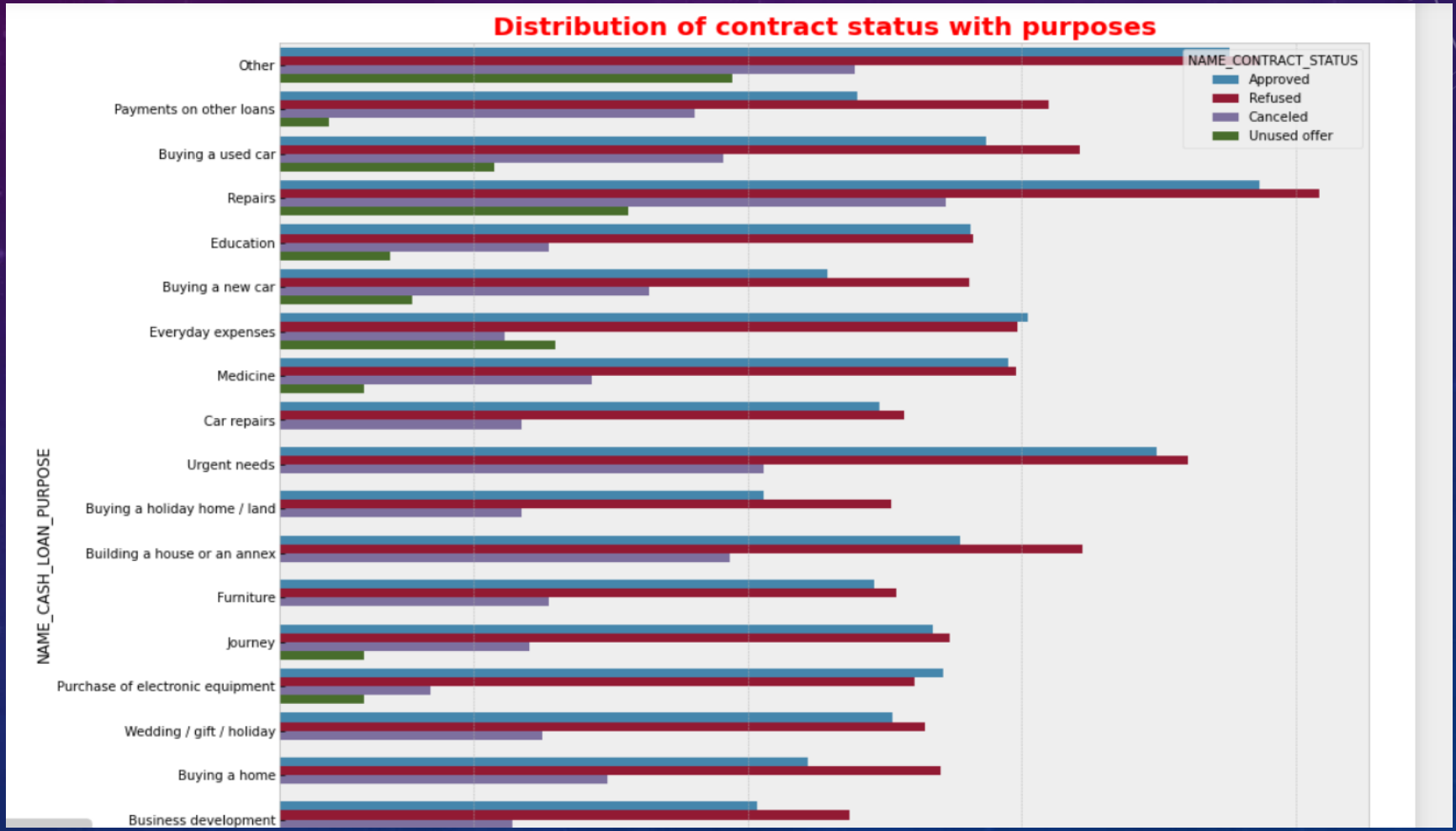
PREVIOUS APPLICATION

DISTRIBUTION OF CONTRACT STATUS



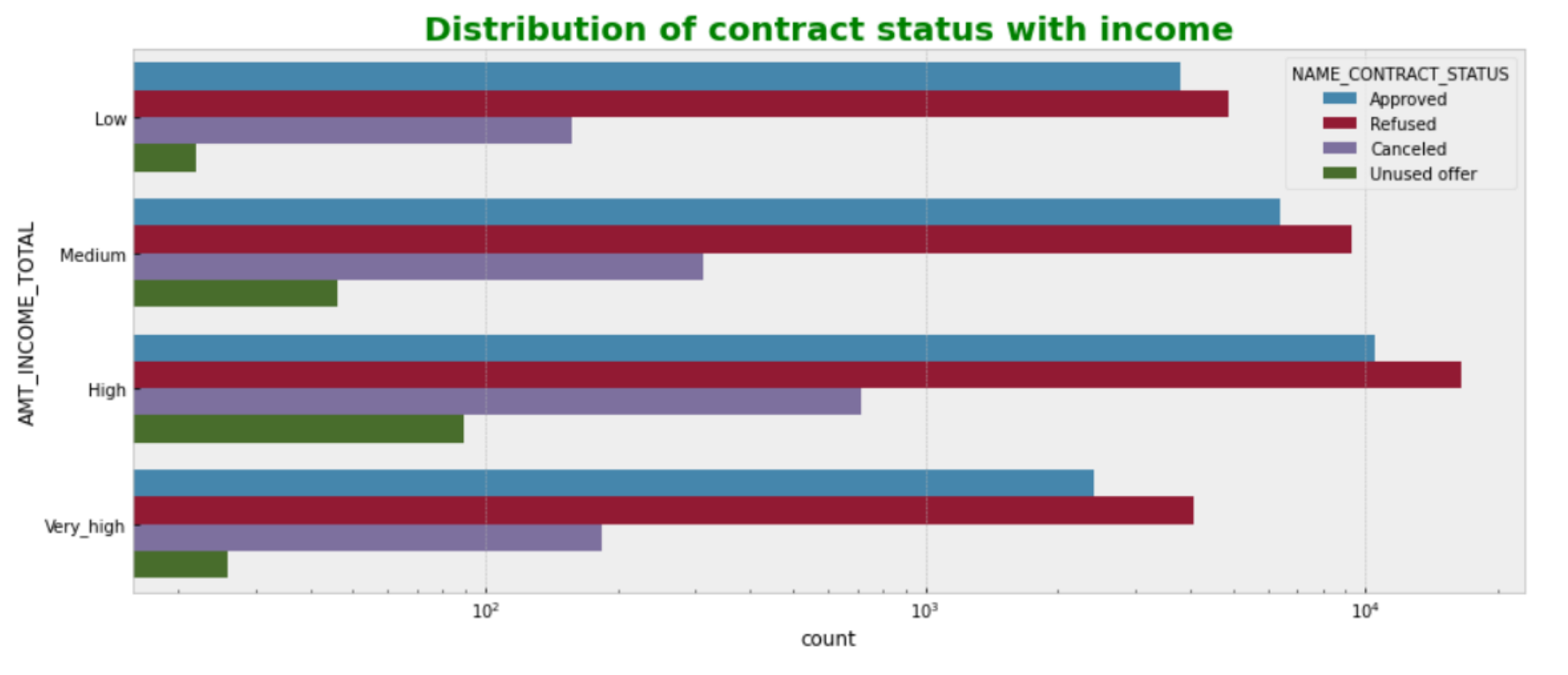
In the above plot, the refused application seems to be more than other status of applications.

DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

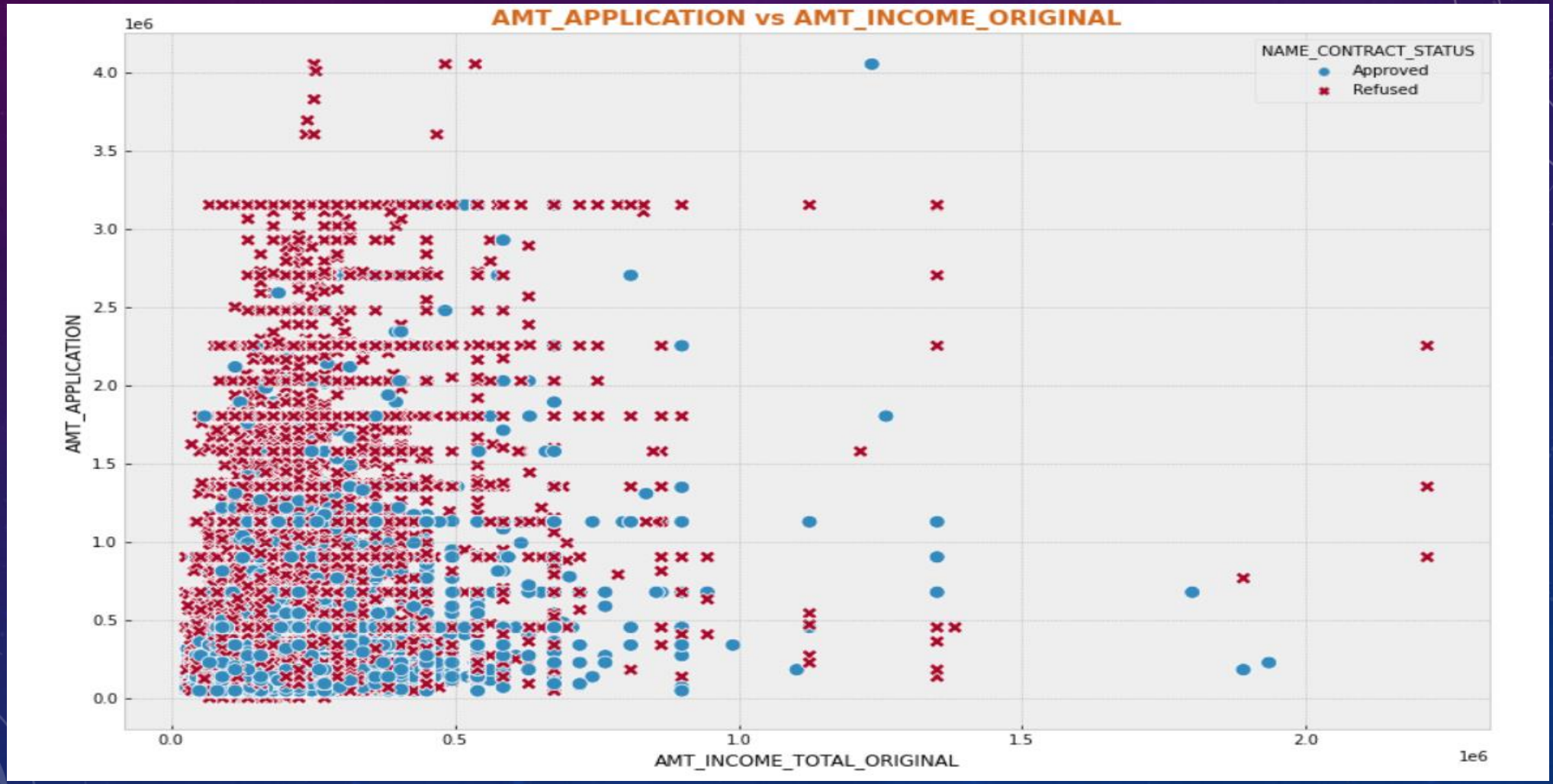


- Most of the refused applications is from Repairs.
- In purchase of electronic equipment and everyday expenses, approved application status is more than refused one.
- Approved and Refused status seems to be almost equal in education.

DISTRIBUTION OF CONTRACT STATUS WITH INCOME



APPLICATION AMOUNT VS TOTAL INCOME



CONCLUSION

- Bank should refuse loan application if loan amount is higher than 300k and income is less than 100k. Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
- Loan purpose ‘Repair’ is having higher number of unsuccessful payments on time. Bank should refuse loan application of Labors because they found as most defaulter. Banks should focus more on contract type ‘Student’ , ’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.

THANK YOU