CREDIT RISK ANALYSIS – UPGRAD CASE STUDY

Meenakshi Sharma

02 Mar 2024

BACKGROUND

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. 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.

Objective:

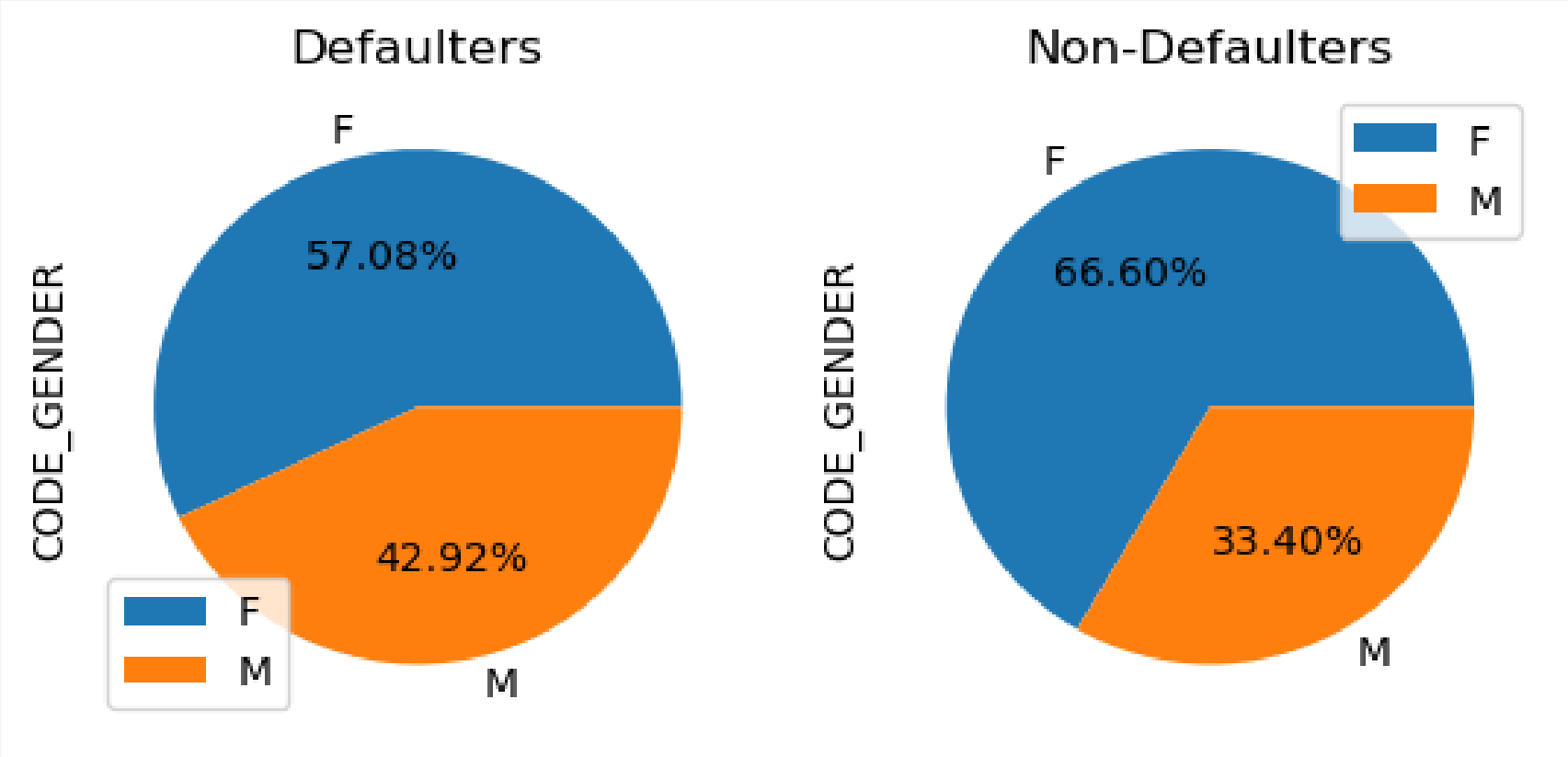
- To identify the driving factors (or driver variables) behind loan default which can be utilized for portfolio and risk assessment.

STEPS REQUIRED FOR ANALYSIS

- Data Understanding
- Data Inspection
- Data Cleaning and Correction
- Check for Data imbalance
- Univariate, Bivariate and Multivariate analysis
- Recommendation and Risks

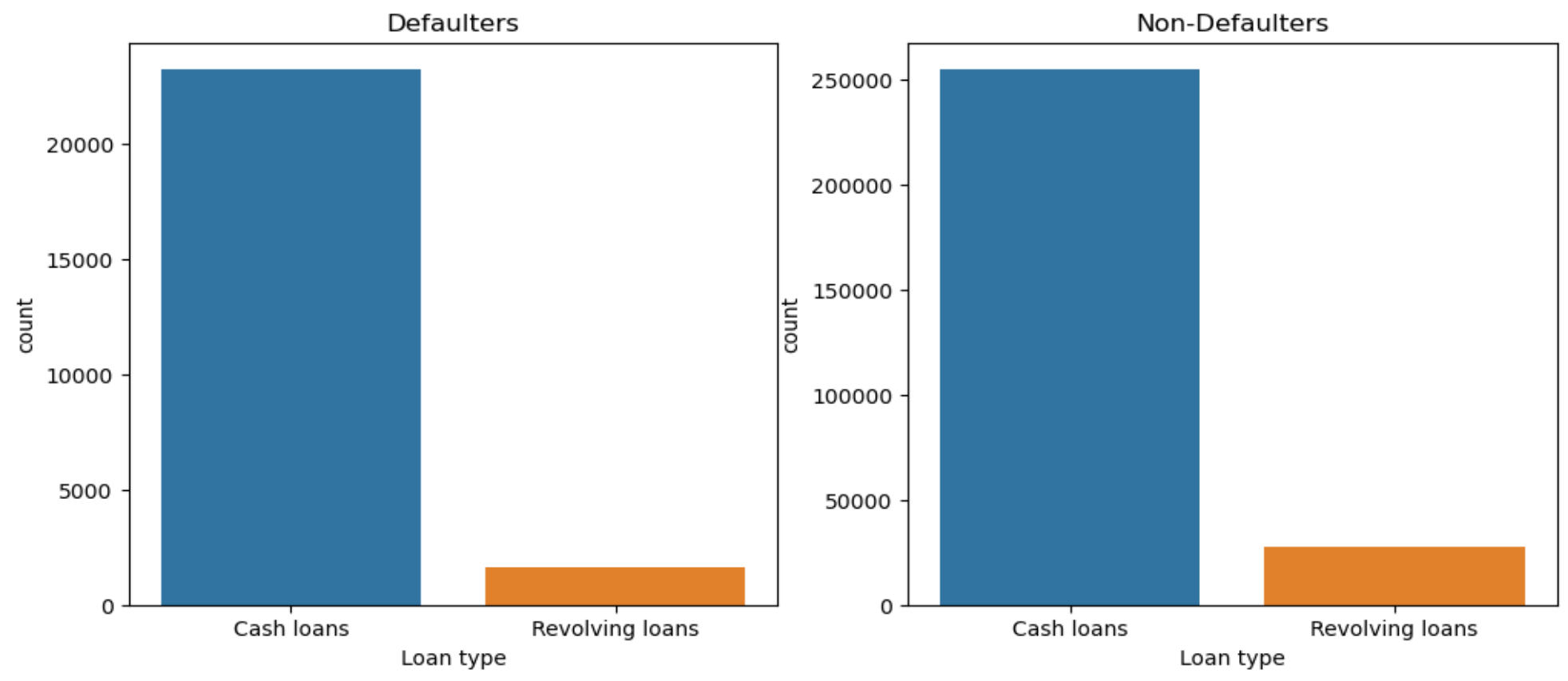
DIVERSITY INDEX OF DATA

- **Female** customers applied **higher than male** customers for loan
- **66.6% Female** customers are **non-defaulters** while **33.4% male** are **non-defaulters**
- **57% Female** customers are **defaulters** while **42% male** customers are **defaulters**



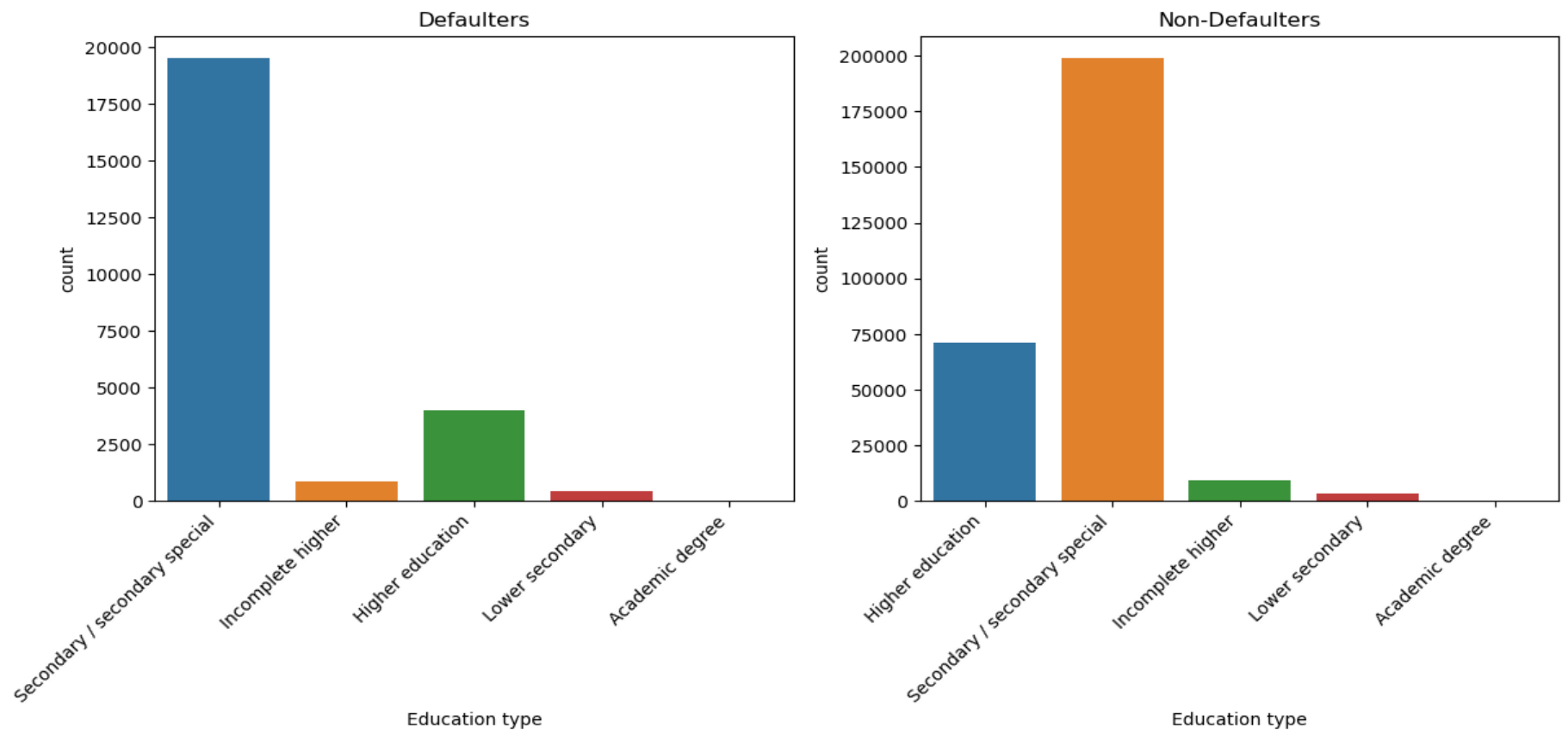
LOAN TYPE

- **Most of the customers** have applied for Cash Loan while **very small proportion** have applied for Revolving loan for both Defaulters as well as Non-defaulters



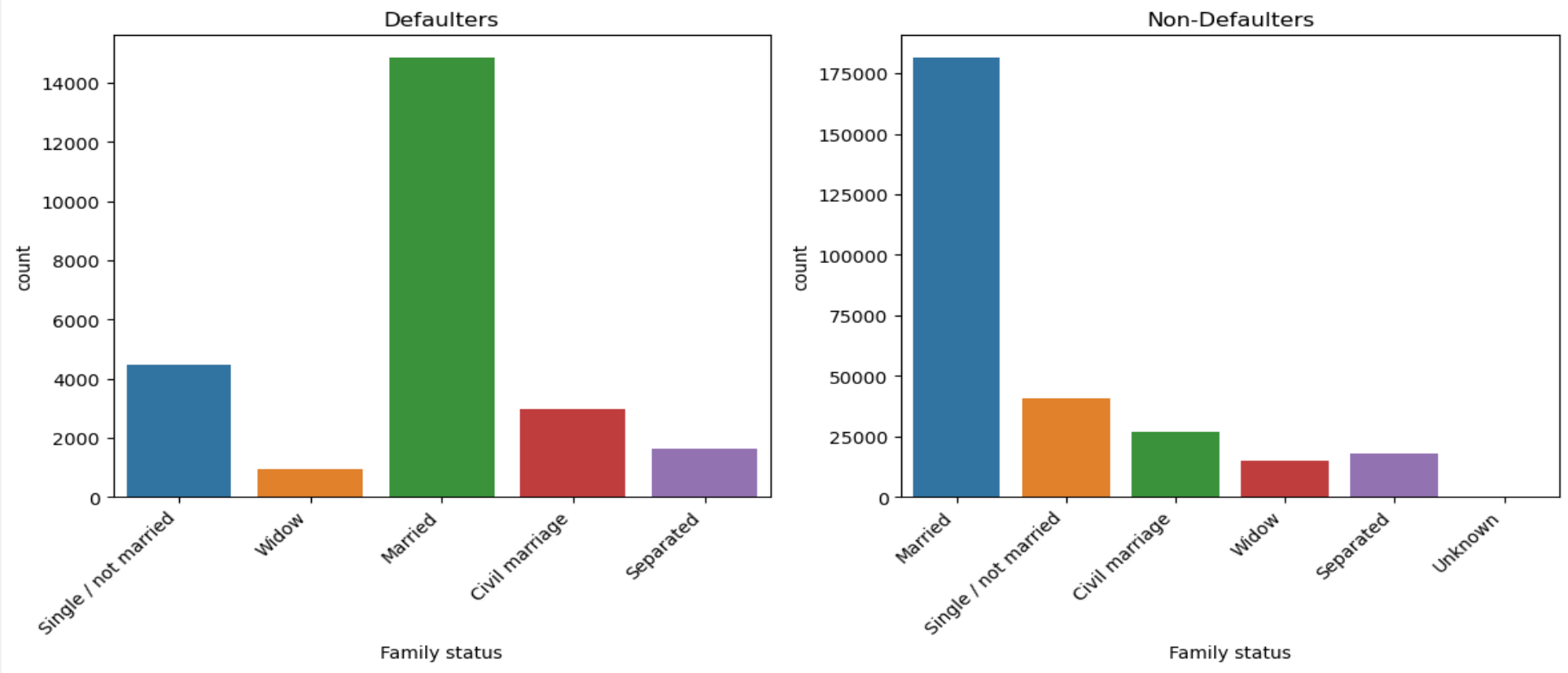
EDUCATION TYPE VS TARGET

- Customers having education Secondary or Secondary Special are more likely to apply for the loan
- Customers having education Secondary or Secondary Special have higher risk to default, Other education types have minimal risk



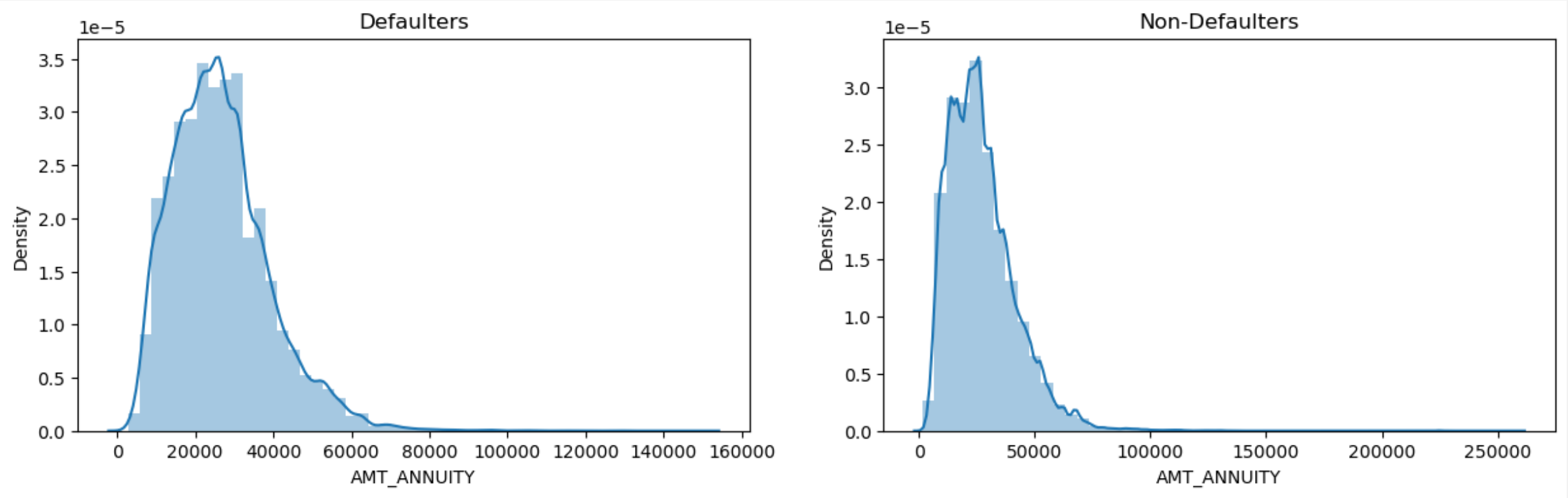
FAMILY STATUS VS TARGET

- **Married customers** seems to be applied most for the loan compared to others for both Defaulters and Non-Defaulters
- In case of Defaulters, customers having single relationship are **less risky**
- In case of Defaulters, Widows shows **Minimal risk**



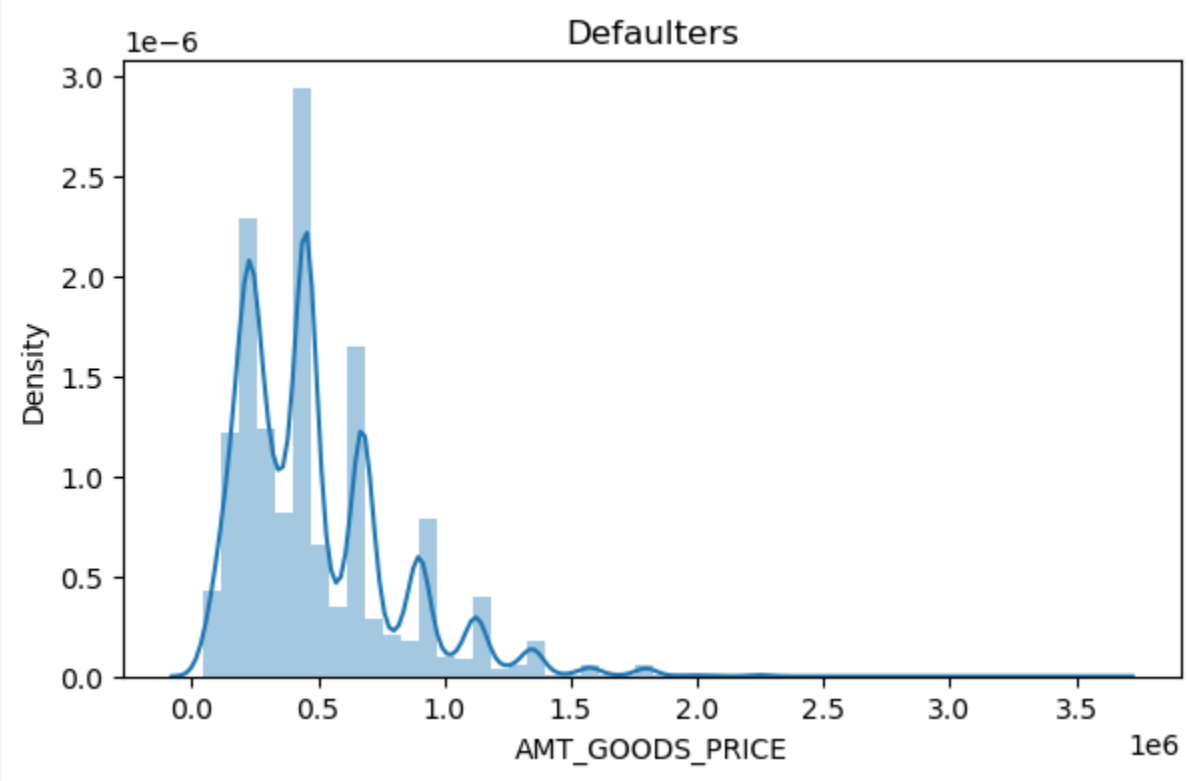
ANNUITY AMOUNT VS TARGET

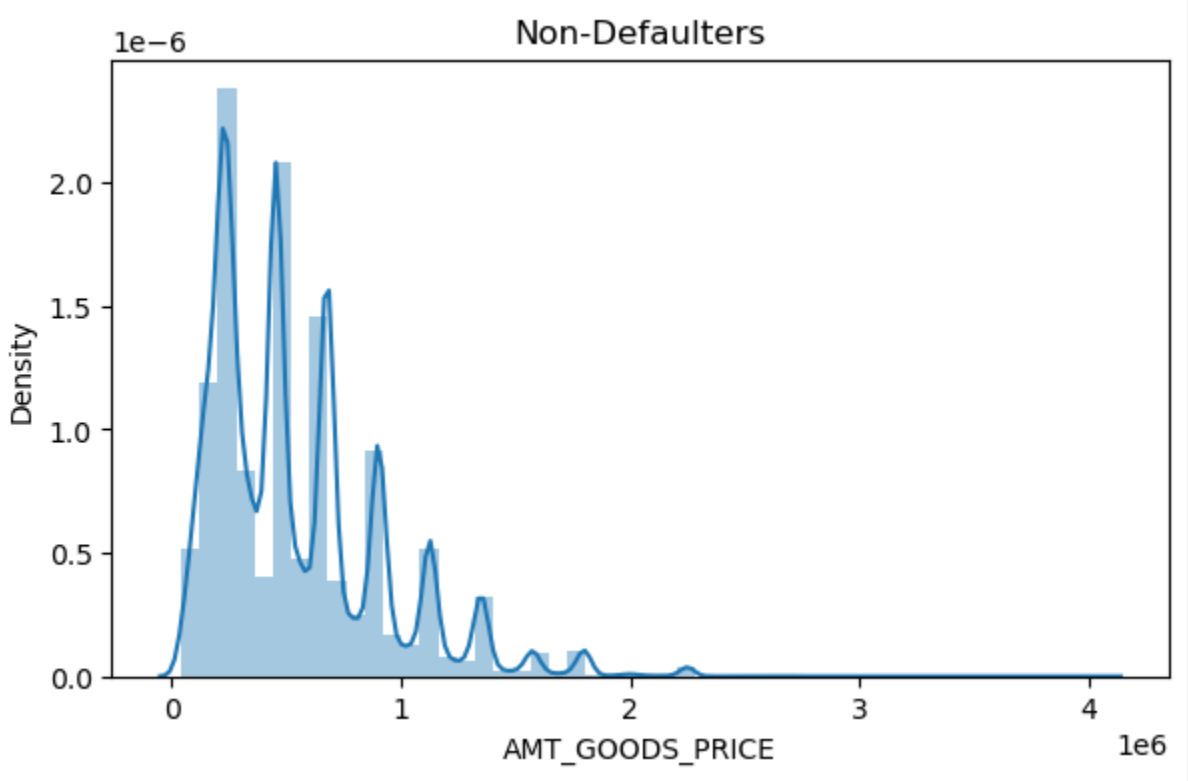
- Below plot clearly shows that the Annuity is similar for defaulters and non-defaulters.



GOODS PRICE AMOUNT VS TARGET

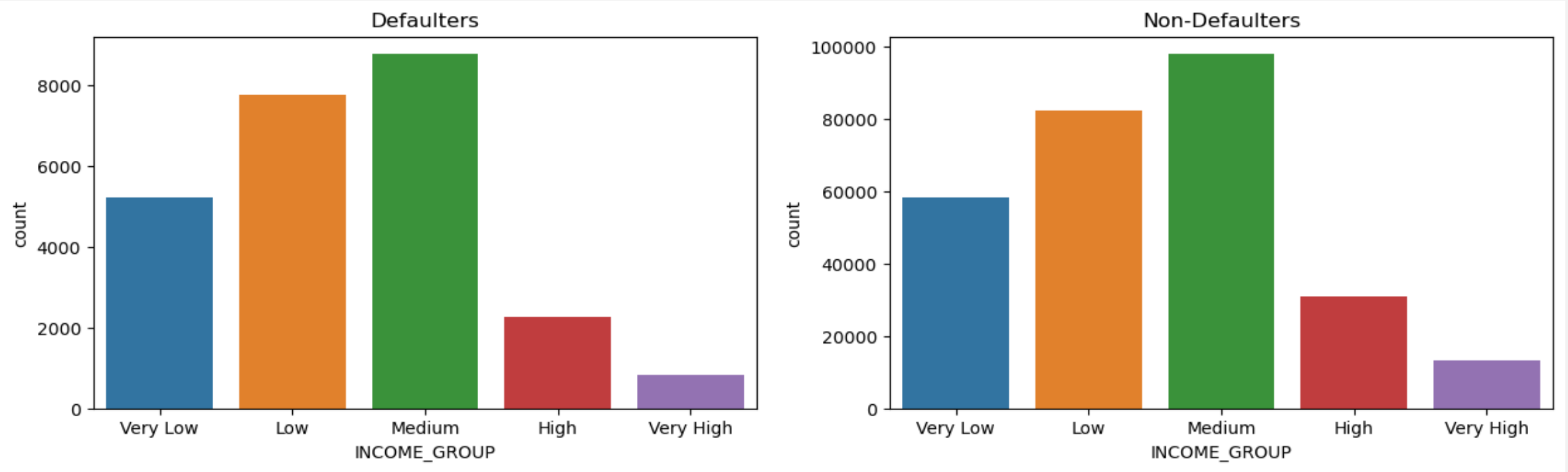
- Below plot clearly shows that the AMT_GOODS_PRICE is similar for defaulters and non-defaulters





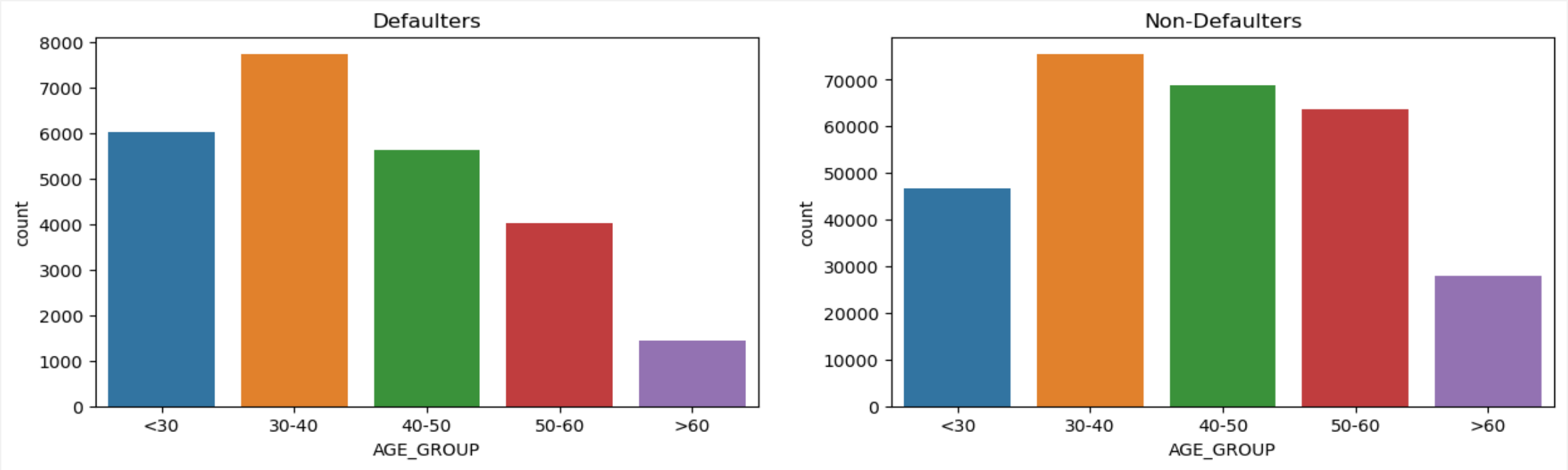
INCOME GROUP VS TARGET

- Customers having **Medium income range** are more likely to apply for the loan for both Defaulters and Non-defaulters
- Customers having **Low** & **Medium** income are at high risk to default



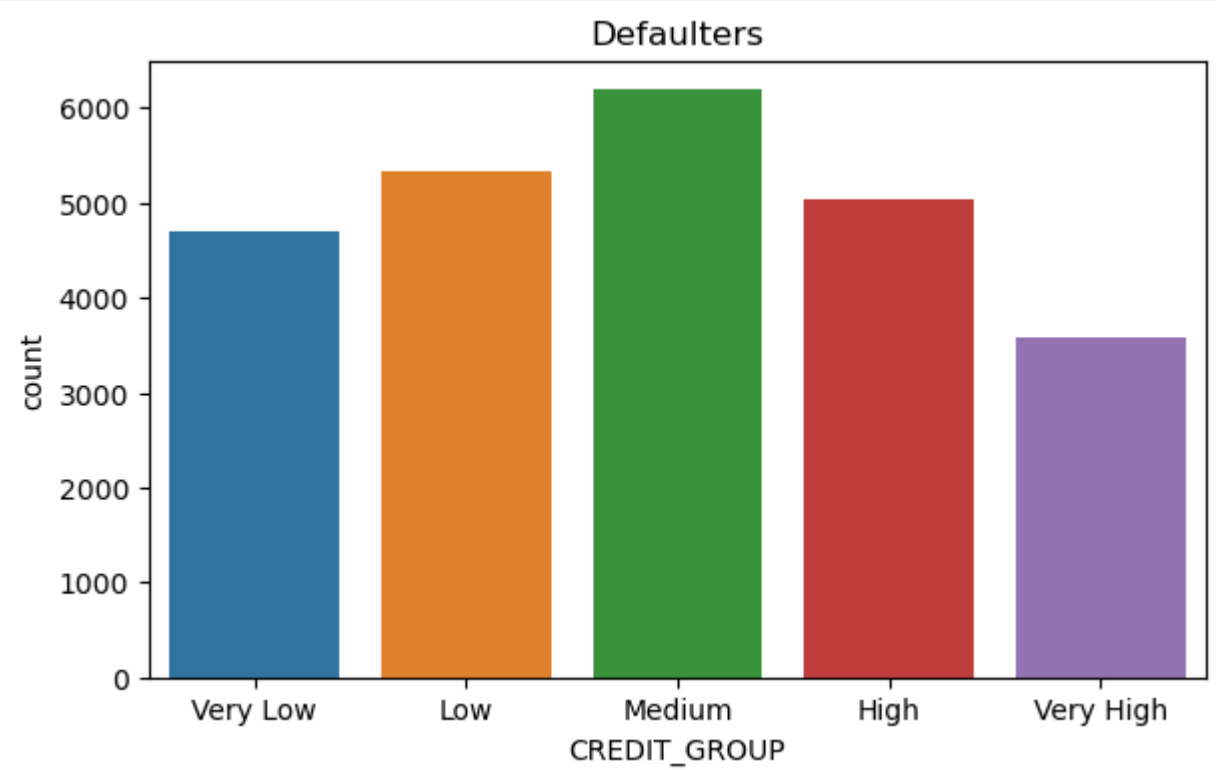
AGE GROUP VS TARGET

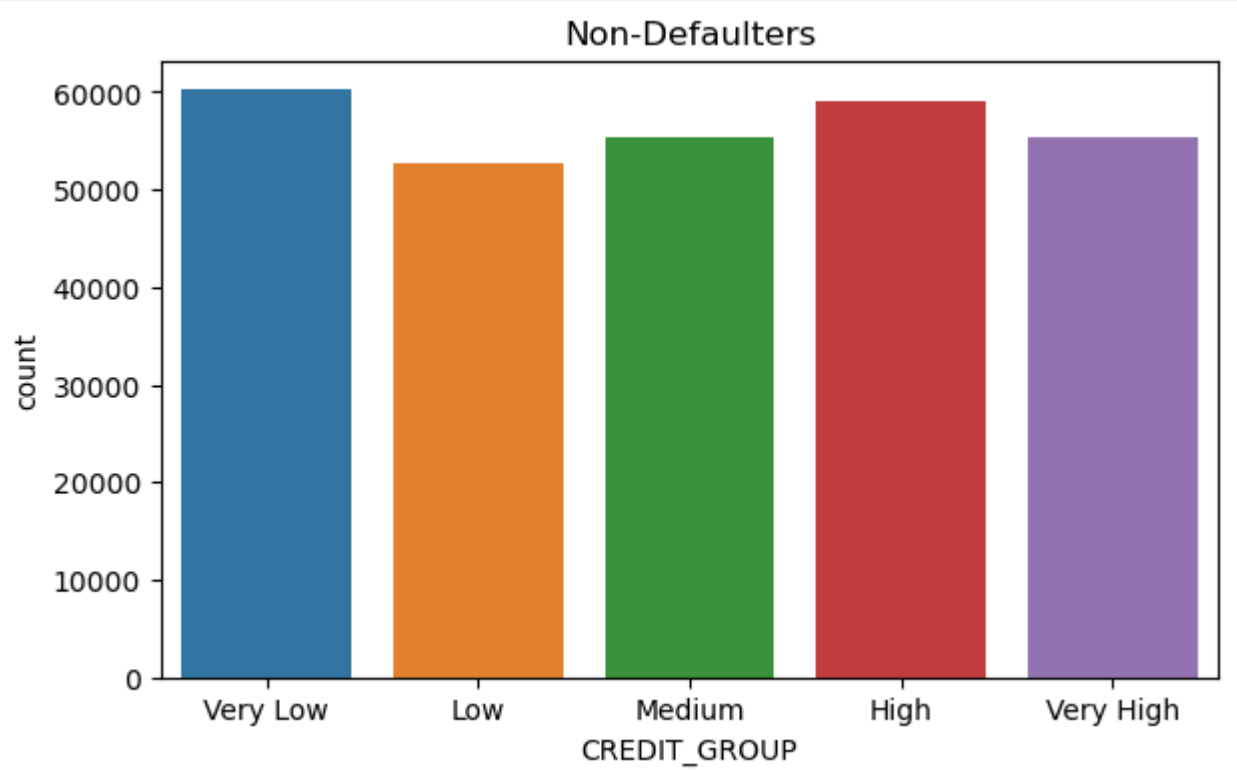
- **Age(30-40)** the group seems to applied higher than any other age group for loans in the case of Defaulters as well as Non-defaulters
- Also, **Age(30-40)** group facing paying difficulties the most
- While **Senior Citizens(60-100)** age group facing paying difficulties less as compared to other age groups



CREDIT GROUP VS TARGET

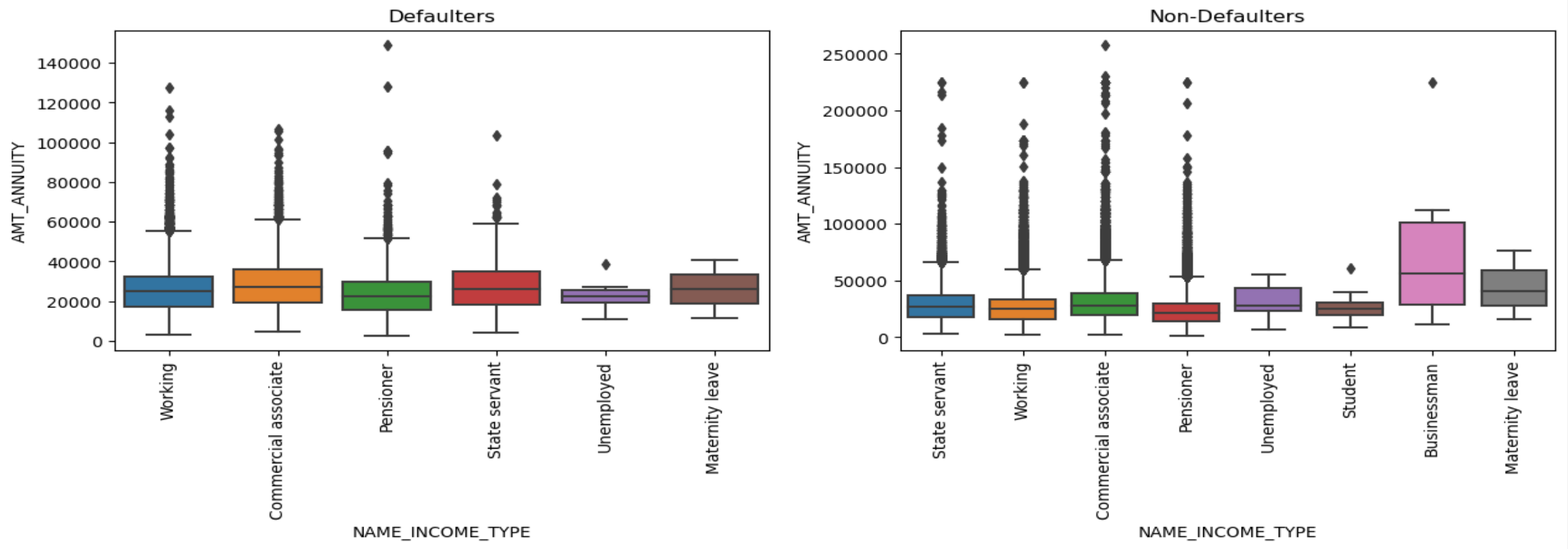
- Customers applying for **Low** and **Medium** credit are at **HIGH** risk of default





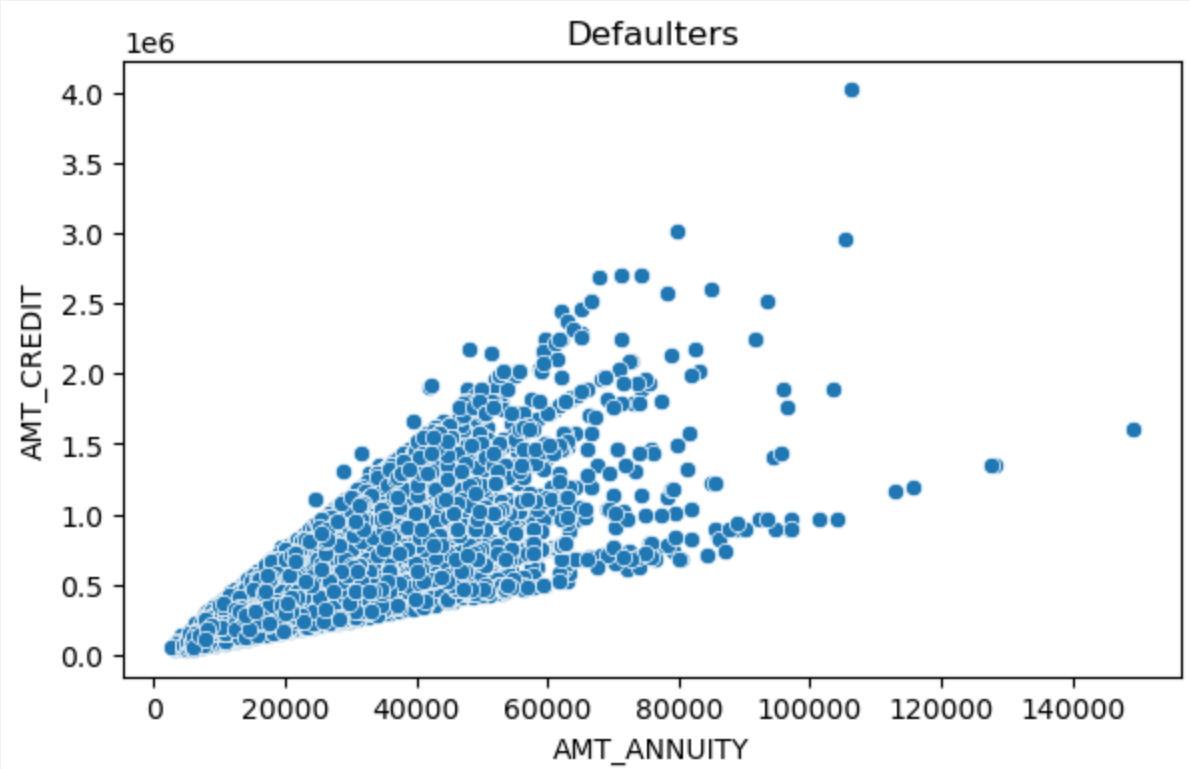
INCOME TYPE VS ANNUITY AMOUNT

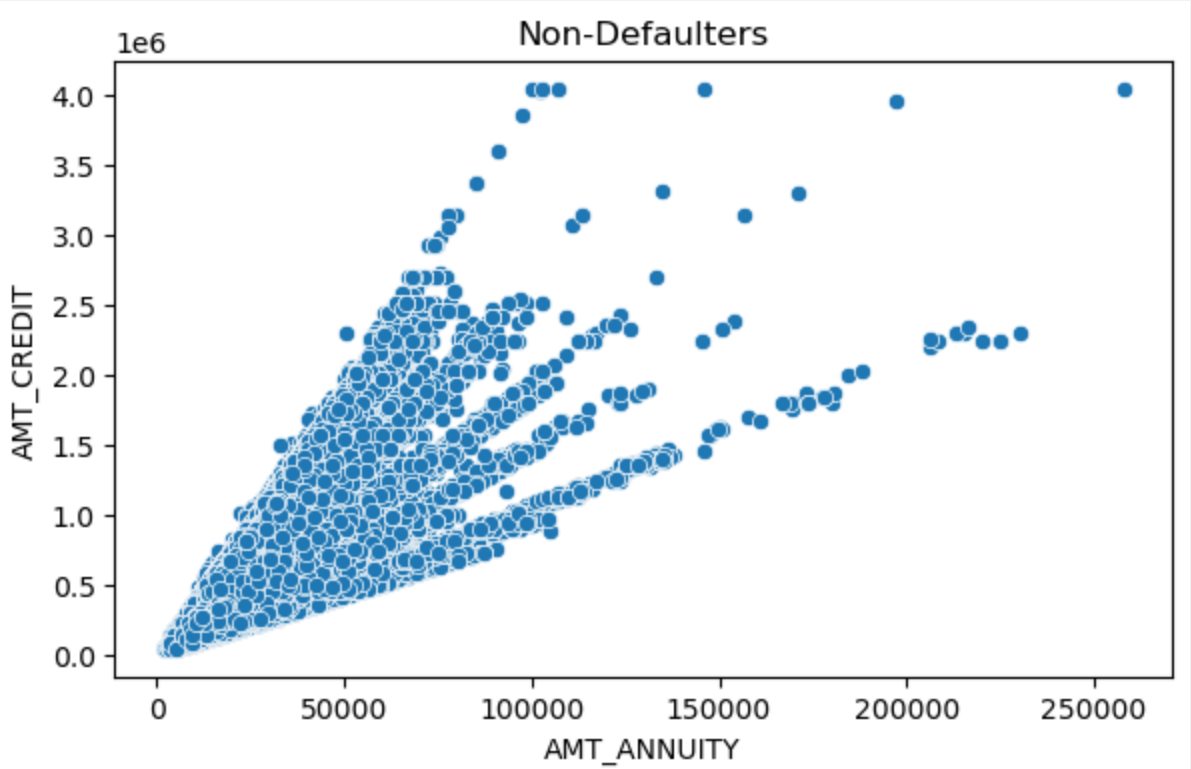
- **Businessman** have higher Annuity amount as compared to other income types
- **Customers** with income type as **Businessman** and **Students** are less likely to default



INCOME AMOUNT VS CREDIT AMOUNT

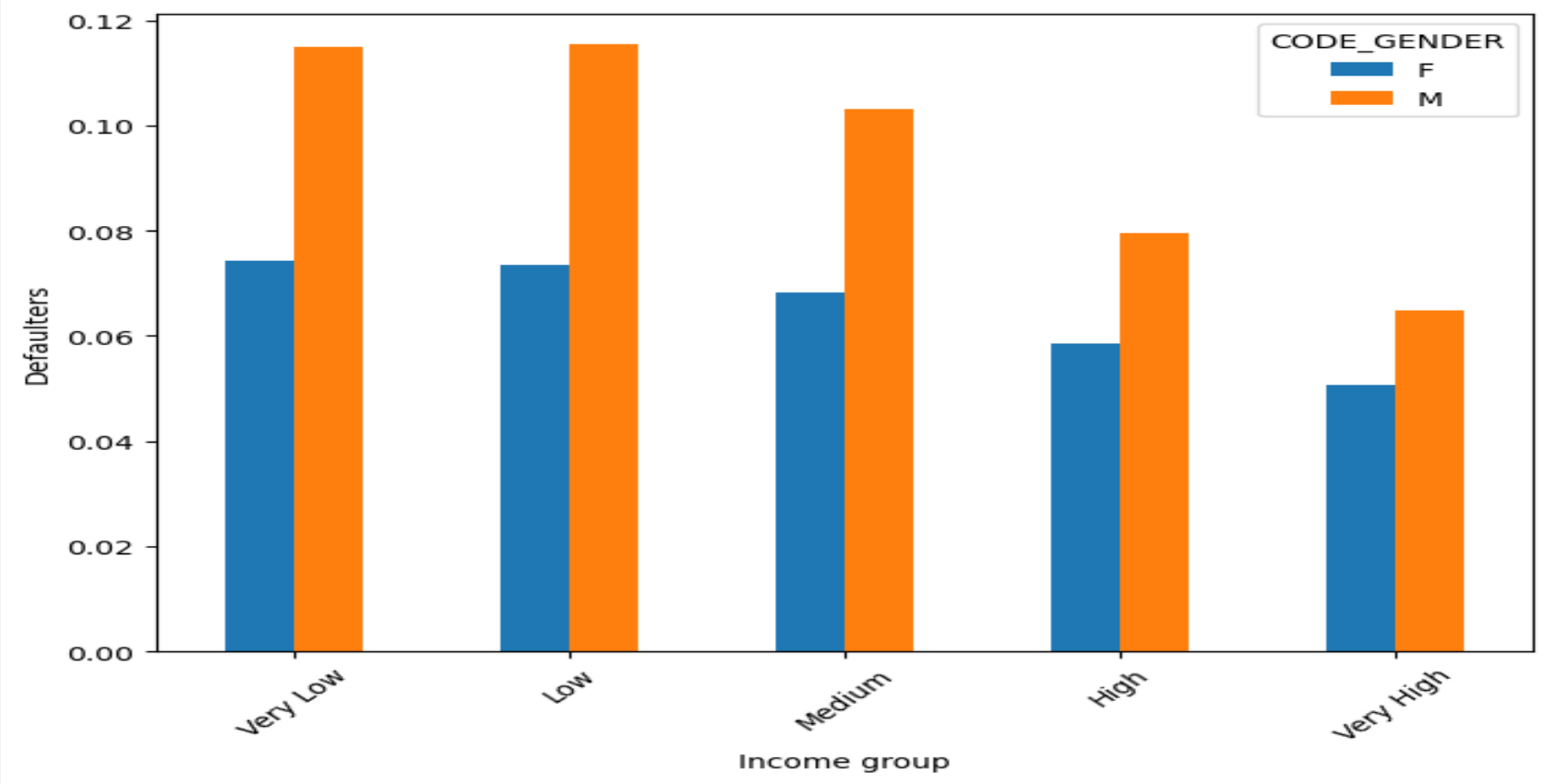
- **AMT_CREDIT** and **AMT_ANNUITY** (EMI) are highly correlated variables for both defaulters and non – defaulters. So as the credit amount of loan increases the EMI amount also increases which is logical



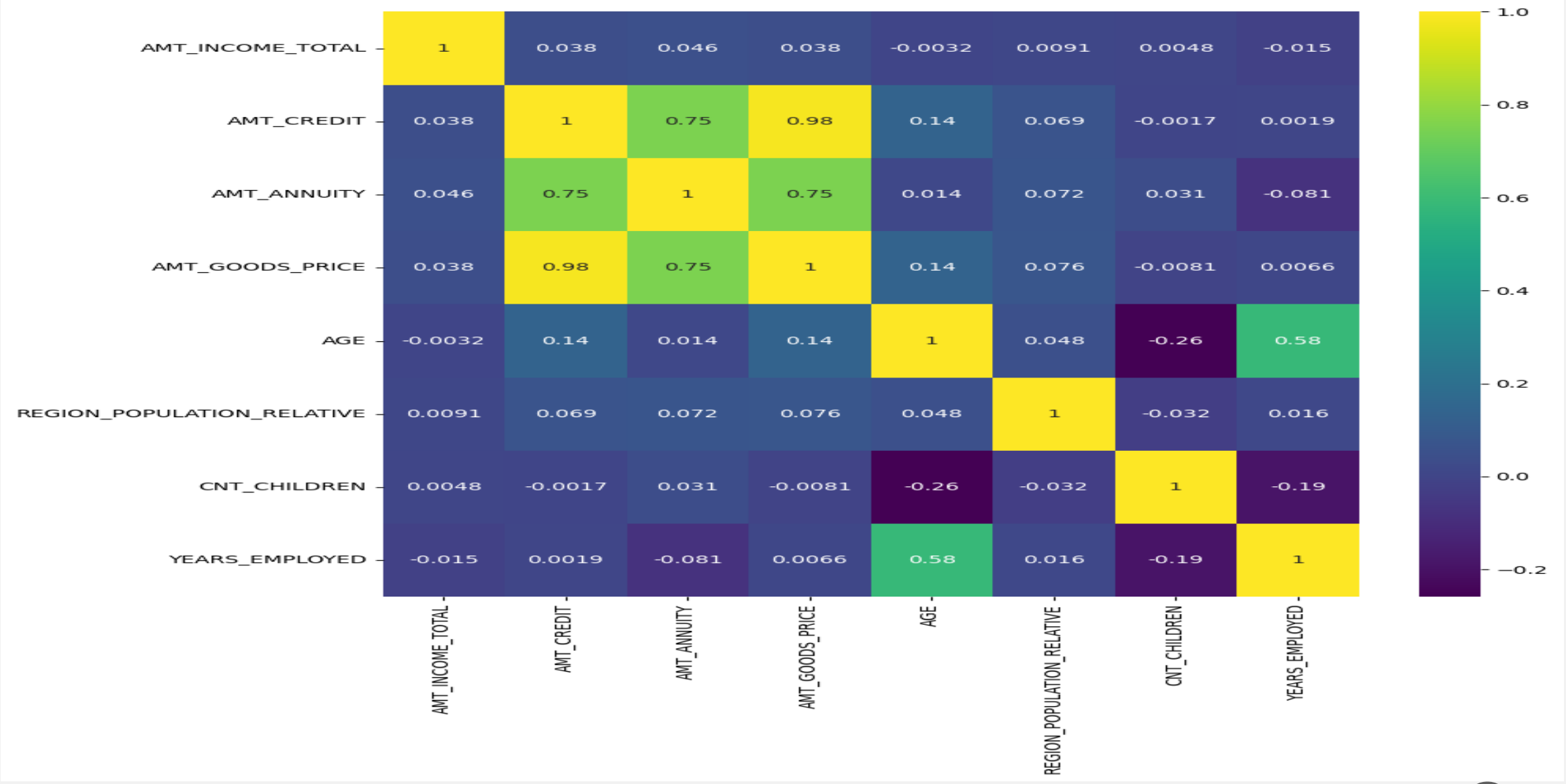


INCOME GROUP AND GENDER

- Men are more likely to default compared to women
- Very low and low income groups are likely to default in male as well as female



DEFAULTERS MULTIVARIATE ANALYSIS



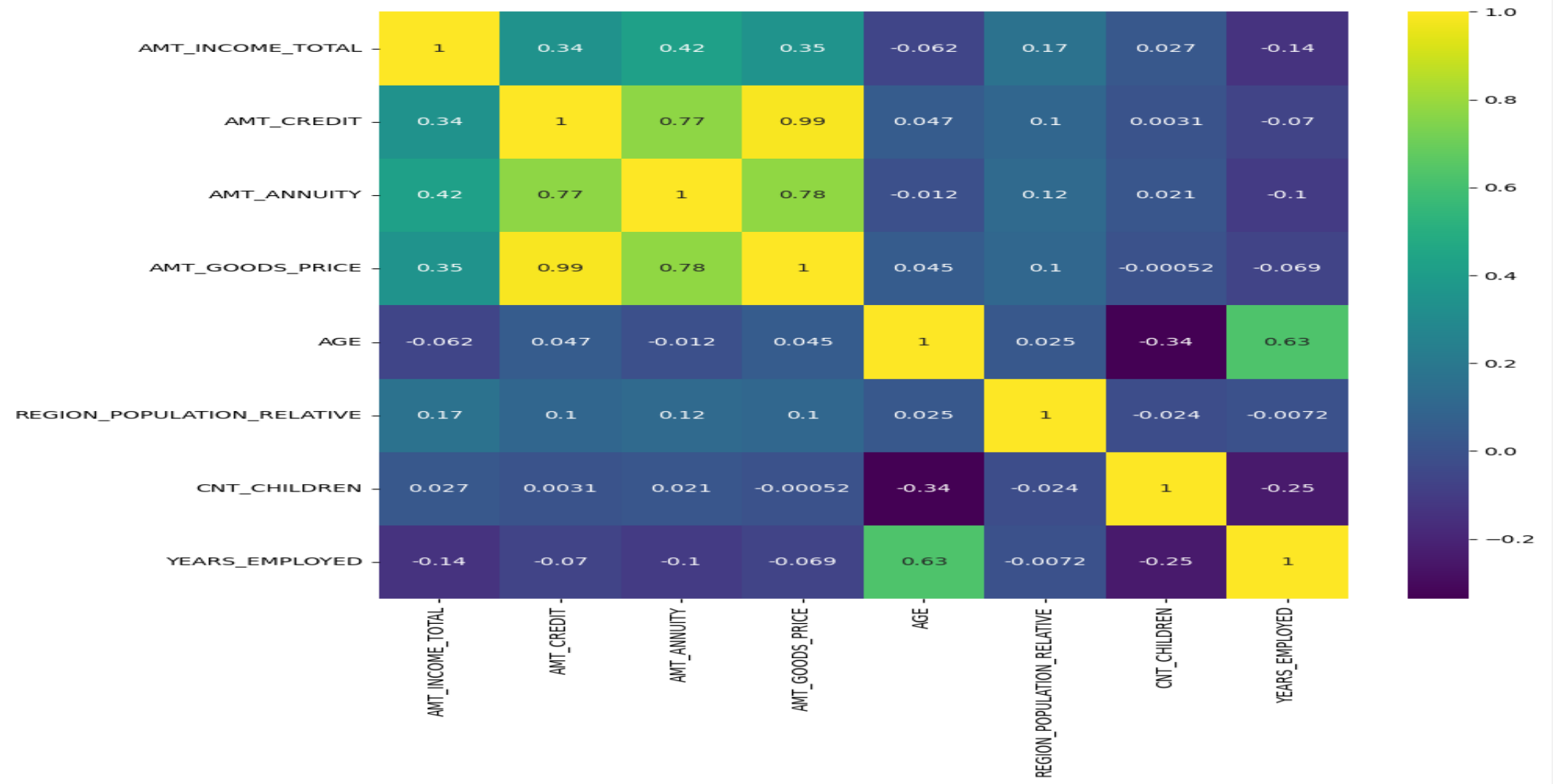
INSIGHTS FROM DEFAULTERS MULTIVARIATE ANALYSIS

- Positive correlations (when one variable increases, the other variable tends to increase) are represented by yellow and green.
- Negative correlations (when one variable increases, the other variable tends to decrease) are represented by blue color.

For Example –

- **AMT_CREDIT** is highly proportional to the **AMT_GOODS_PRICE**, As the price of the goods increases credit also increases
- **AGE** is inversely proportional to the **CNT_CHILDREN**, means older people have fewer children count and vice-versa
- Fewer children customers have in a densely populated area

NON DEFAULTERS MULTIVARIATE ANALYSIS

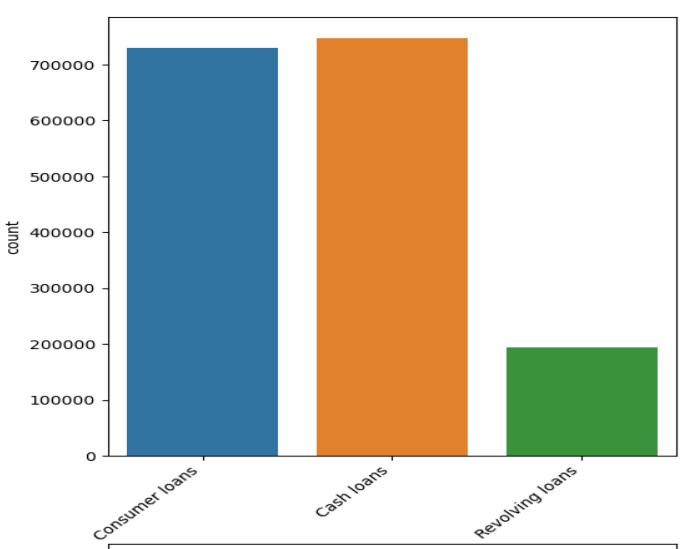


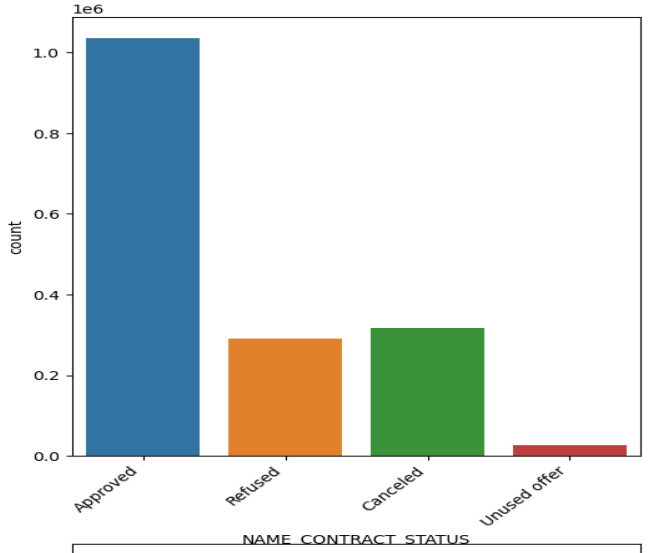
INSIGHTS FROM NON DEFAULTERS MULTIVARIATE ANALYSIS

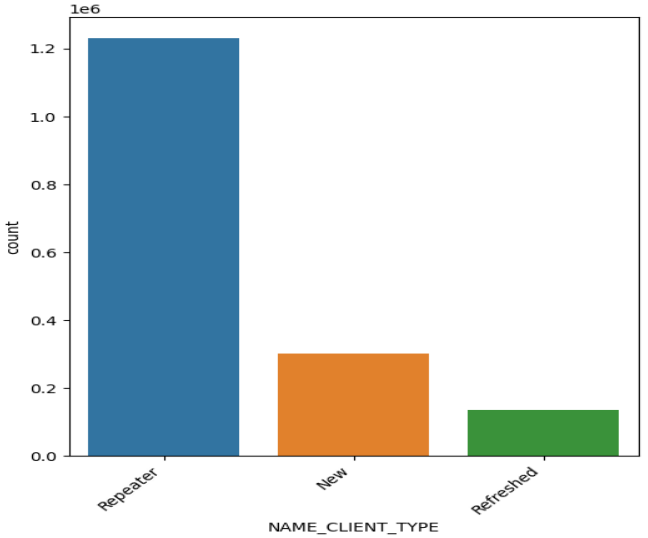
This heat map for Non - defaulters is also having quite the same observation just like defaulters:

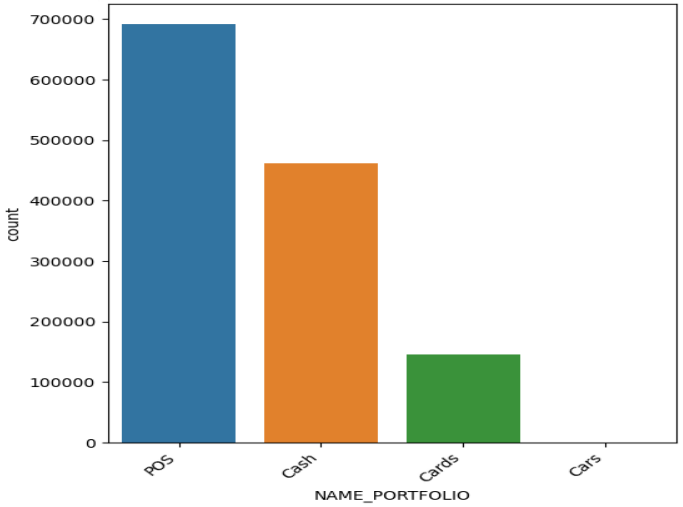
- **AMT_CREDIT** is highly proportional to the **AMT_GOODS_PRICE**, As the price of the goods increases credit also increases
- **AGE** is inversely proportional to the **CNT_CHILDREN**, means older people have fewer children count and vice-versa
- Fewer children customers have in a densely populated area

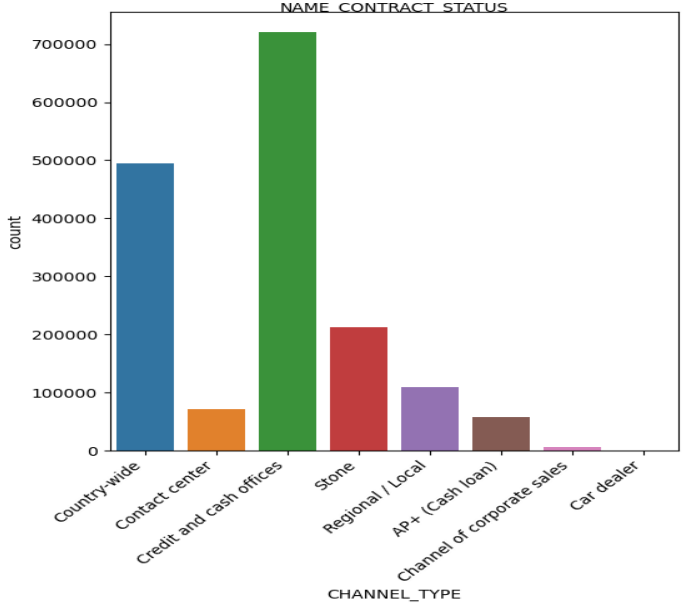
UNIVARIATE ANALYSIS ON PREVIOUS DATA











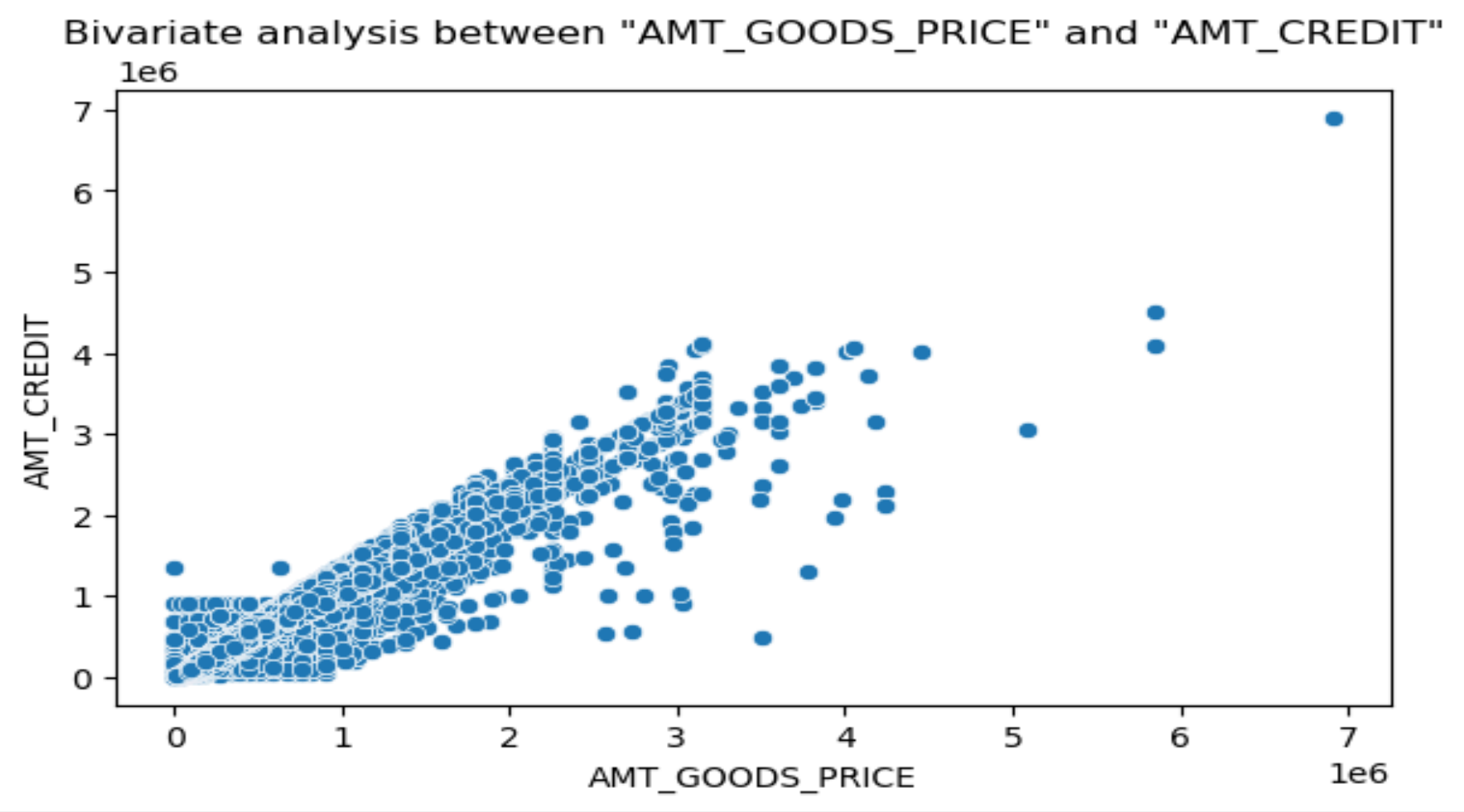
INSIGHTS FROM UNIVARIATE ANALYSIS OF PREVIOUS DATA

We can see that there is data imbalance in below columns:-

- NAME_CONTRACT_TYPE - There are very few Revolving Loans
- NAME_CONTRACT_STATUS - There are very few Refused loans. Almost negligible Canceled loans
- NAME_CLIENT_TYPE - There are very few New applicant. Even fewer Refreshed applicants
- NAME_PORTFOLIO - Very few application for Cards and Cars
- CHANNEL_TYPE - Except Country-Wide, Credit and Cash offices, all other channels are very few in number

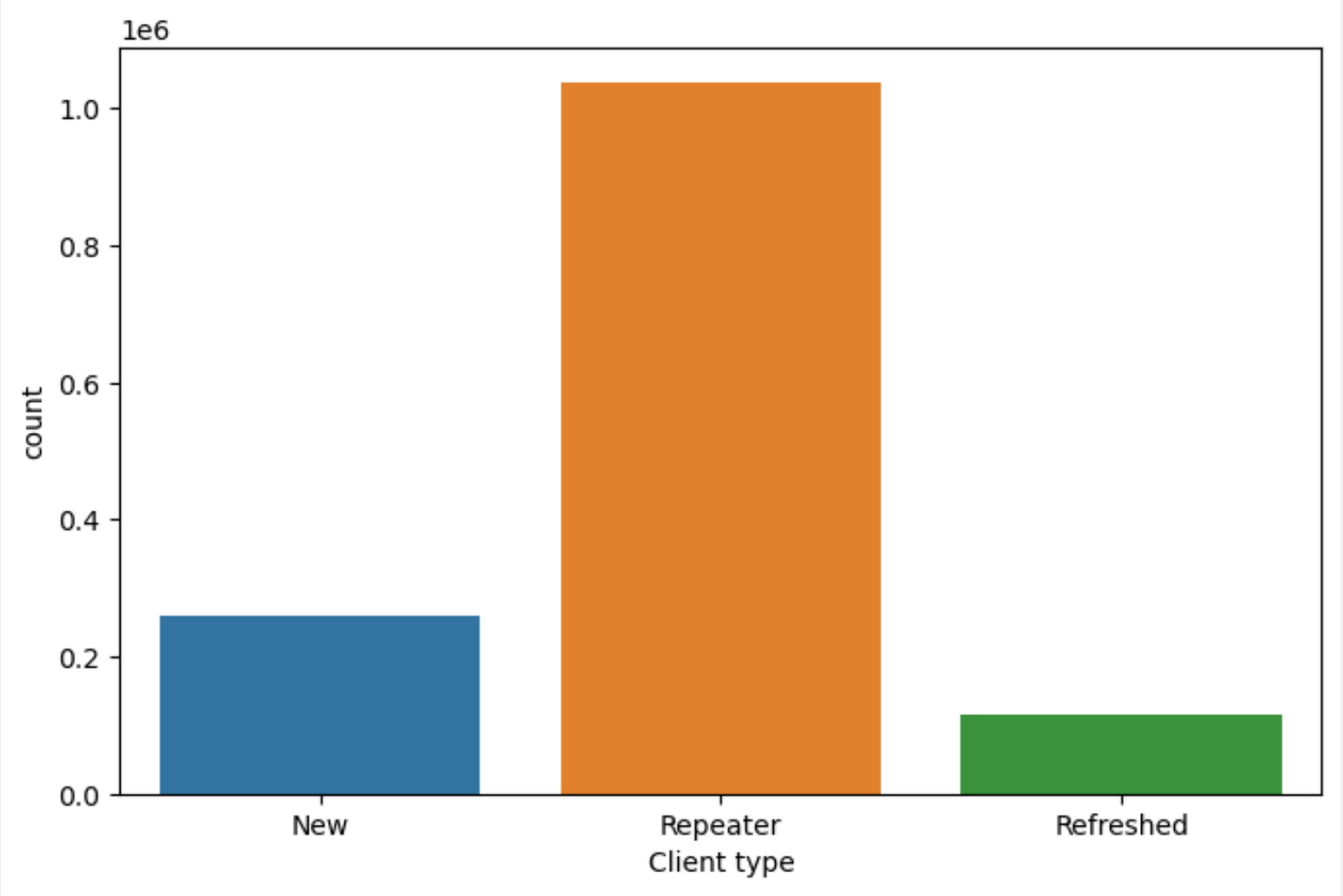
CREDIT AMOUNT VS GOODS PRICE AMOUNT

- AMT_CREDIT and AMT_GOODS_PRICE are in proportion with each other, if one is increase, the other one also increase and vice-versa



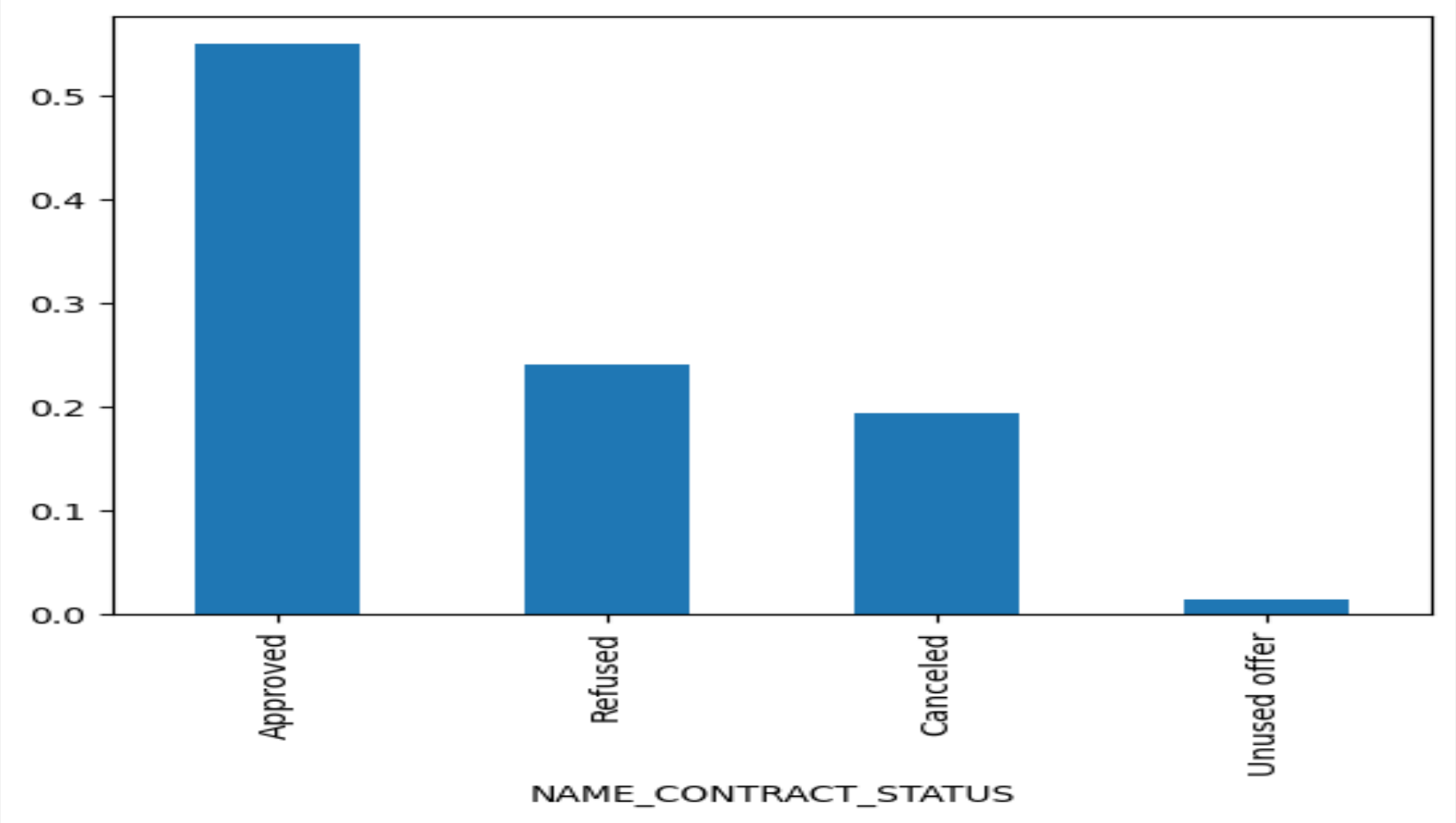
CUSTOMER TYPE

- There are huge number of repeated customers as compared to new customers



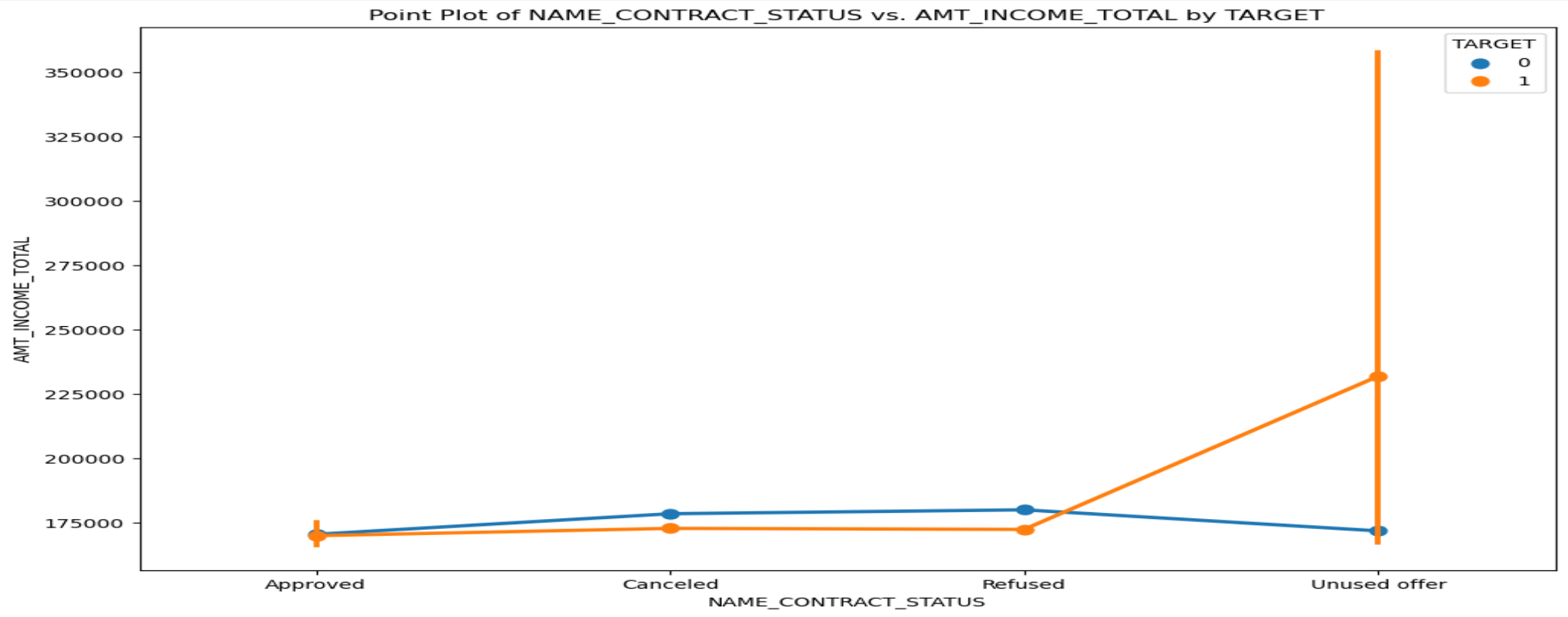
CONTRACT STATUS

- It can be seen that amongst the defaulters 25% of people's applications were rejected in their previous applications and almost 50% applications were approved



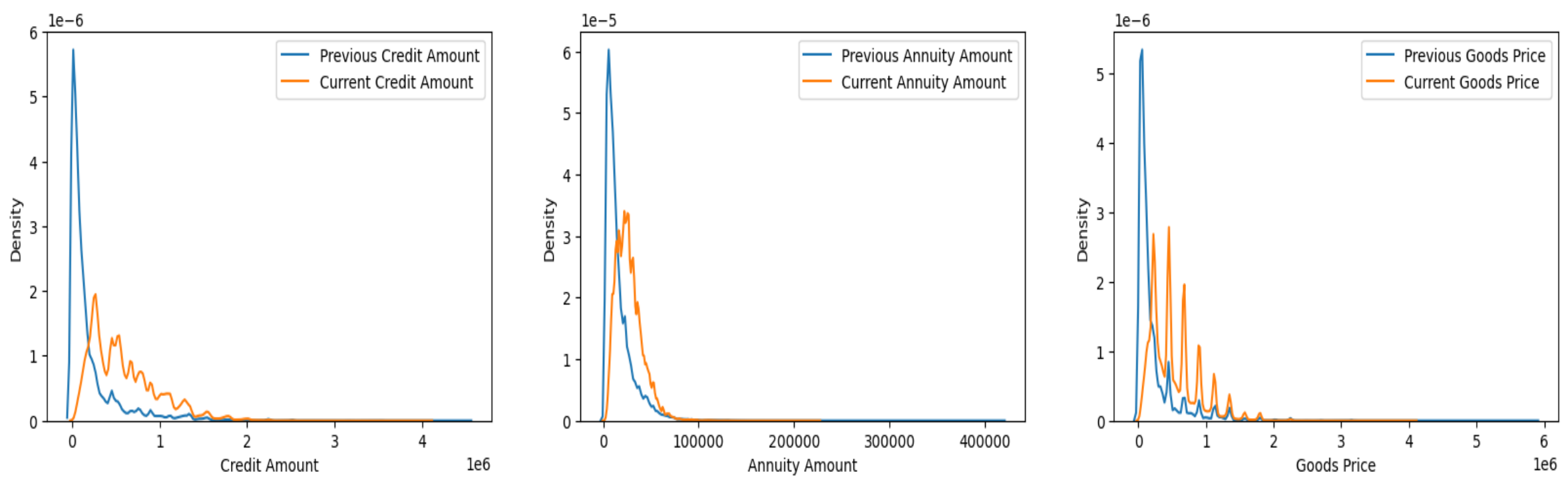
POINT PLOT OF CONTRACT STATUS VS INCOME AMOUNT

- The point plot reveals an interesting trend. It indicates that customers who have not used an offer earlier ("Unused offer" category in NAME_CONTRACT_STATUS) are more likely to default on their loans, even when their average income is higher compared to other contract status categories. This suggests that the history of not using previous loan offers may be a significant factor in predicting loan defaults, potentially even more important than income alone



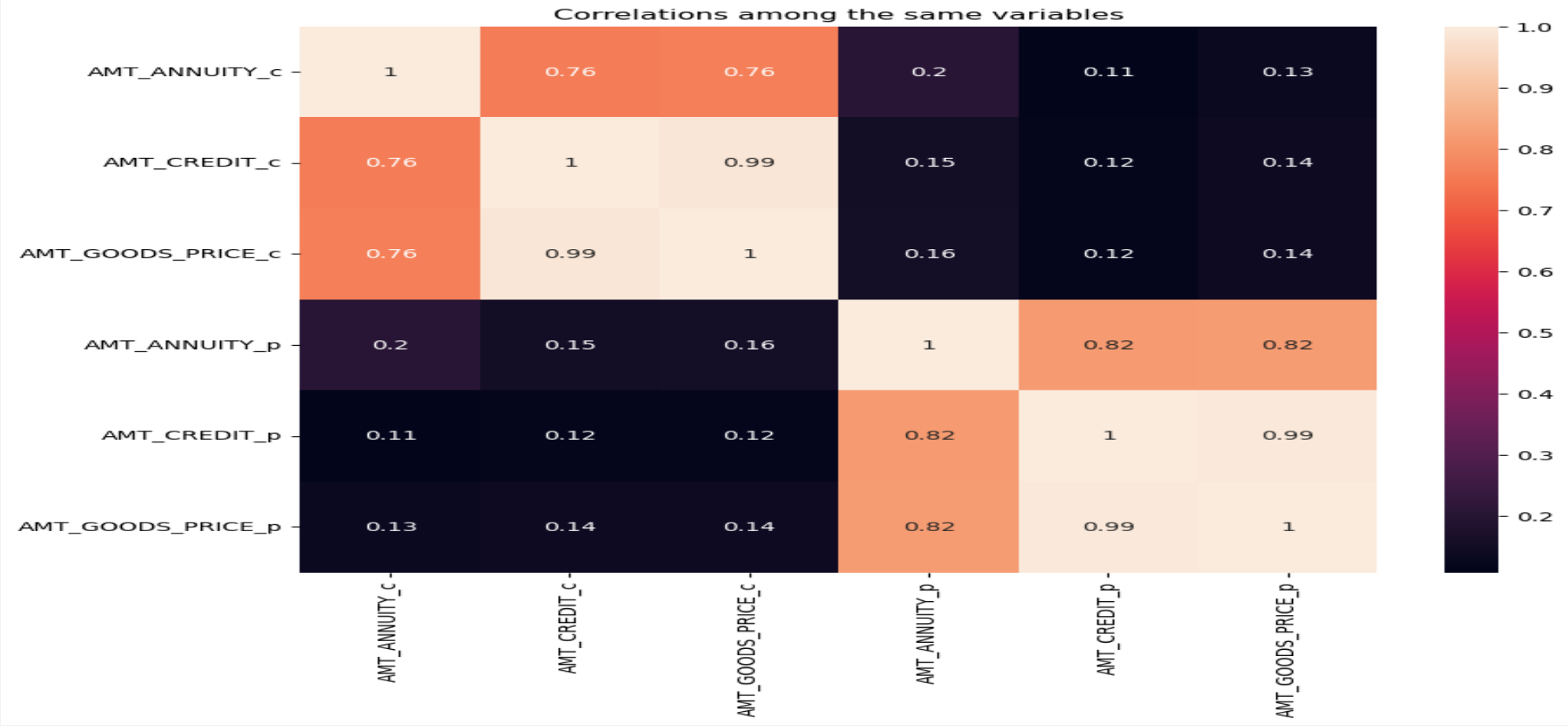
COMPARISON BETWEEN NEW APPLICATION AND PREVIOUS APPLICATION

- The previous applicants are maximum in the credit range of 1L-2L, but a sudden downfall is seen for the credit range of 2L-4L. The current applicants seems to follow a normal trend
- All the graphs clearly depicts that the maximums of each variable in previous applications are less than the maximums of each variable in the current applications



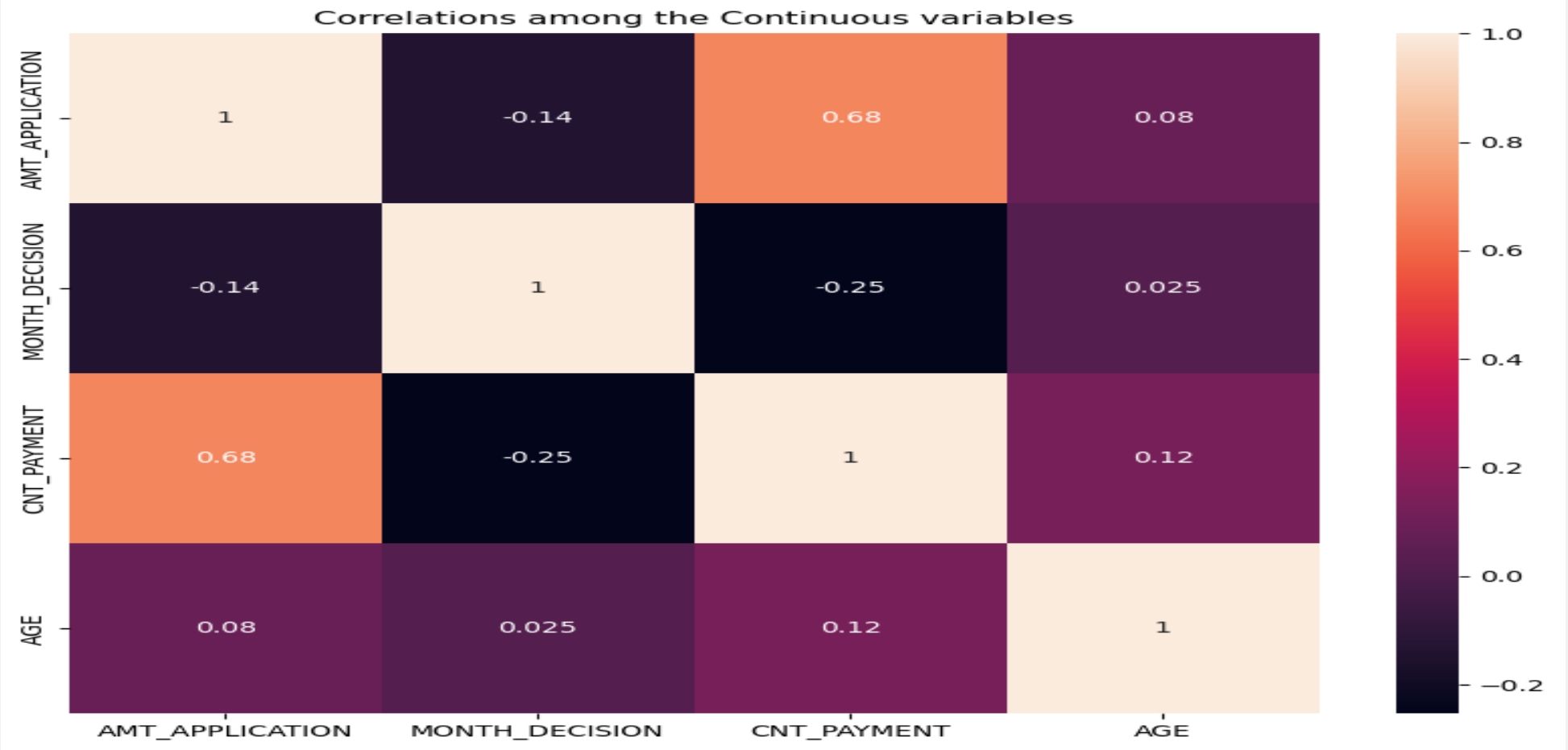
RELATIONSHIP BETWEEN NEW APPLICATION VARIABLES AND PREVIOUS APPLICATION VARIABLE

- Variables of current application are not much in proportion to variables of previous application, means variables of current application doesn't co-relate with previous application variables



CORRELATION AMONG CONTINUOUS VARIABLES

- **CNT_PAYMENT** is highly proportional to the **AMT_APPLICATION**
- **AMT_APPLICATION** is inversely proportional to the **MONTH_DECISION**. Similarly CNT_PAYMENT and MONTH_DECISION are inversely proportional to each other



THANK YOU