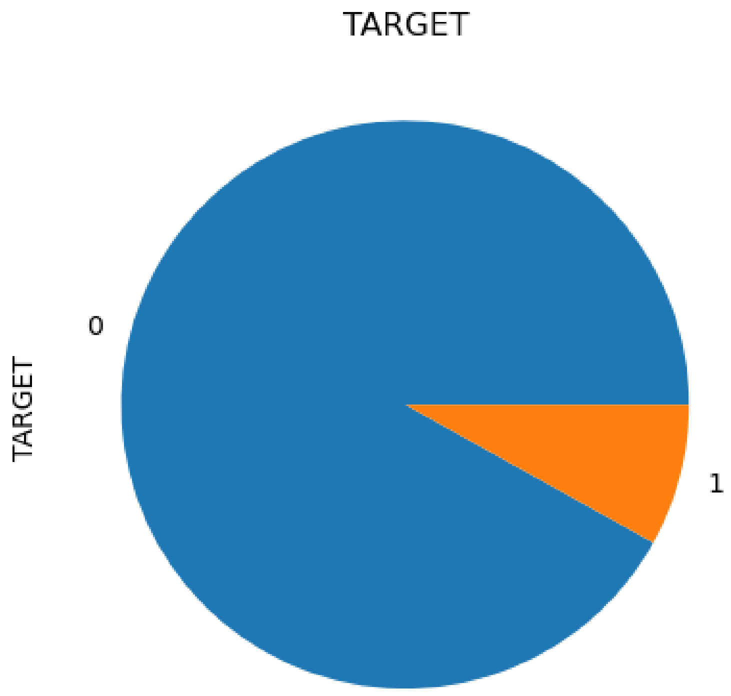
CREDIT EDA CASE STUDY

BHUMIKA JAIN

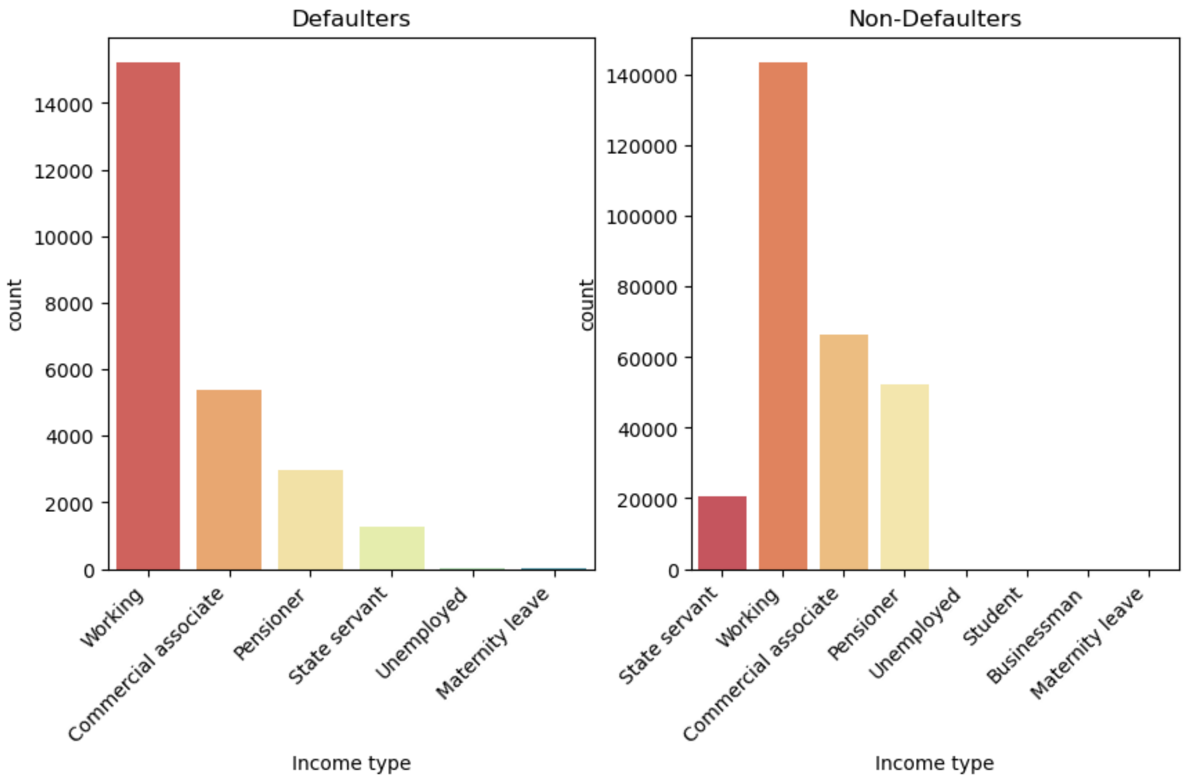
CURRENT APPLICATION DATASET

TARGET VERIABLE ANALYSIS



As per this pie chart, we can observe that there are very few people who are facing difficulty in paying installments all others are not having any difficulty in payment.

Defaulters and non-defaulters on the basis of Income type

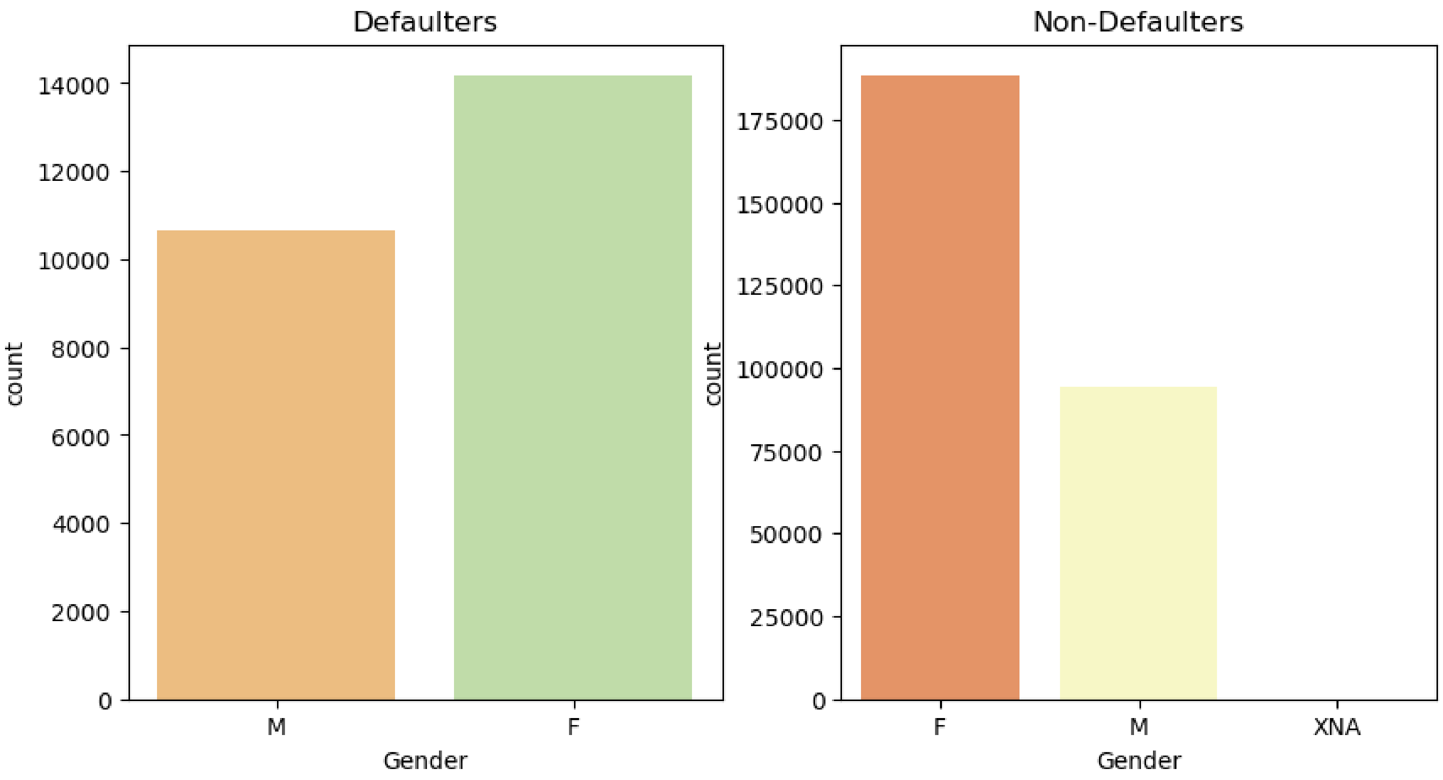


****Analysis****

1. Defaulters: As per the analysis 'Working' people are facing most difficulty in the payment.

2. Non-Defaulters: Most higher percentage of Non-Defaulter is also a 'Working' Segment.

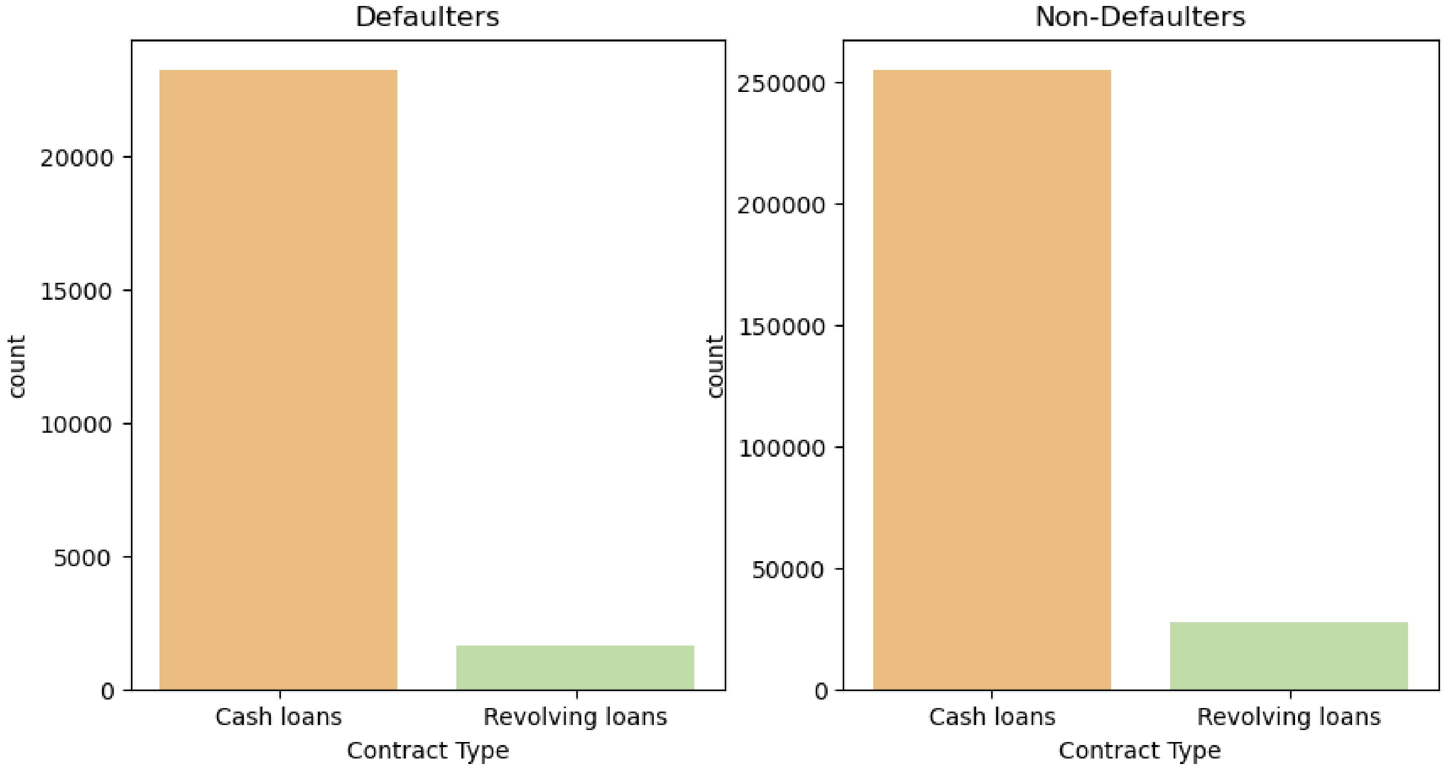
Defaulters and non-defaulters on the basis of Gender



Analysis

1. Defaulters: As per the analysis `'F'` Females are facing most difficulty in the payment.
2. Non-Defaulters: Most higher percentage of Non-Defaulter is also a `'F'` Segment.

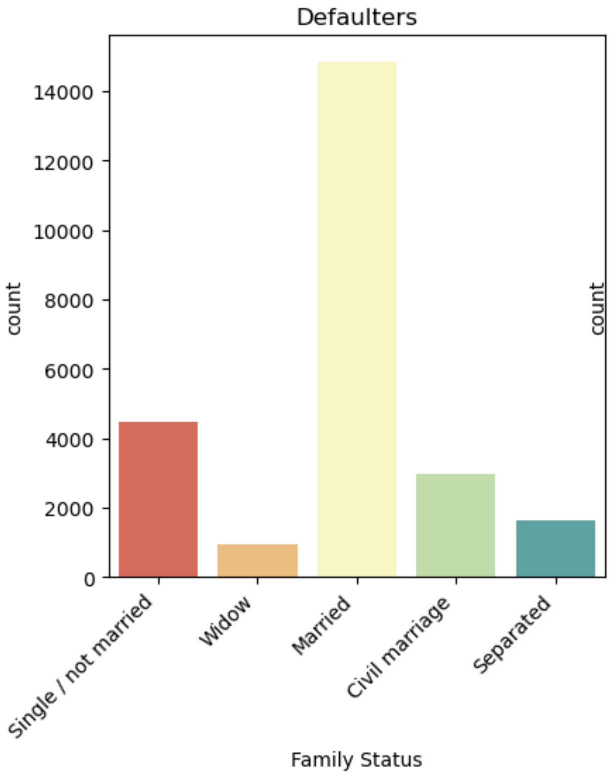
Defaulters and non-defaulters on the basis of Contract Type

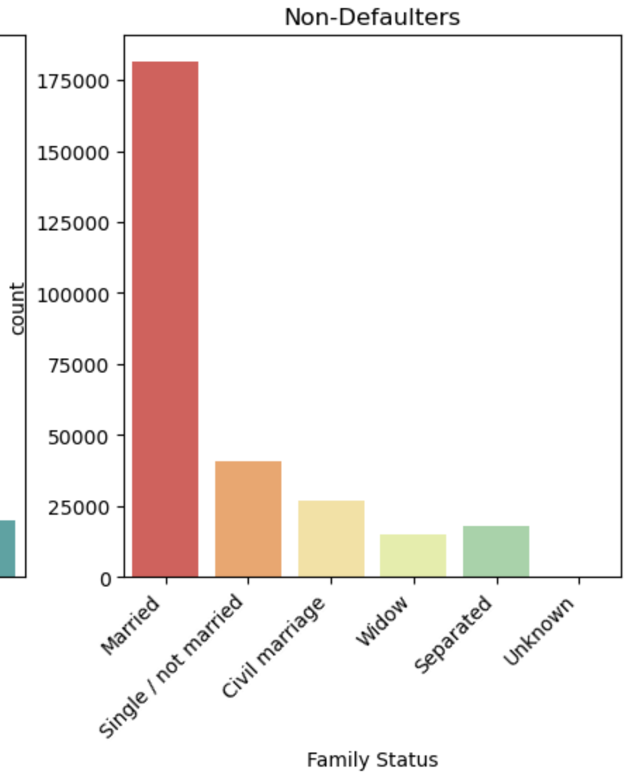


****Analysis****

1. Defaulters: As per the analysis people with `Cash Loans` are facing most difficulty in the payment.
2. Non-Defaulters: Most higher percentage of Non-Defaulter is also a `Cash Loans` Segment.

Defaulters and non-defaulters on the basis of Family Status

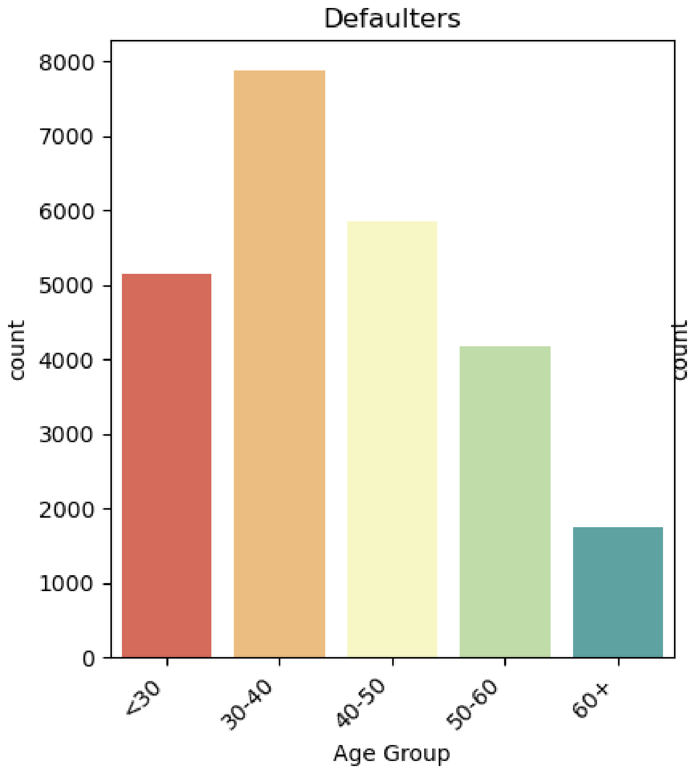


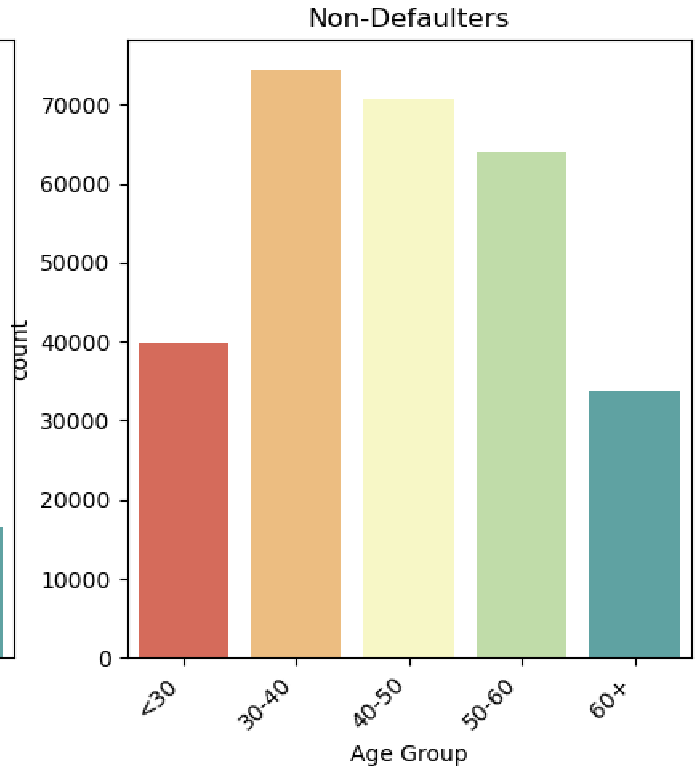


Analysis

1. Defaulters: As per the analysis people with `Married` status are facing most difficulty in the payment.
2. Non-Defaulters: Most higher percentage of Non-Defaulter is also a `Married` Segment.

Defaulters and non-defaulters on the basis of AGE_GROUP

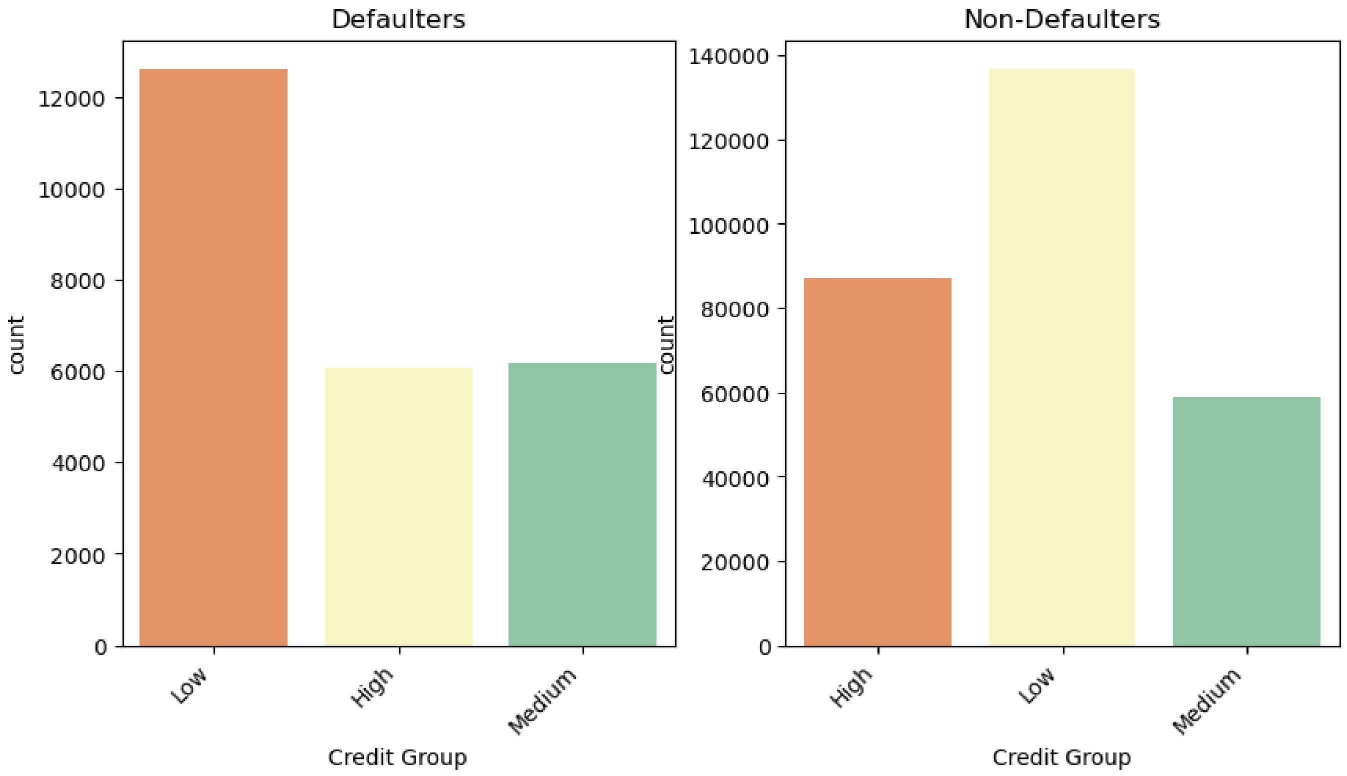




****Analysis****

1. Defaulters: As per the analysis people with age `30-40` are facing most difficulty in the payment.
2. Non-Defaulters: Most higher percentage of Non-Defaulter is also a `30-40` Segment, but age group of `40-50` and `50-60` are also paying the amounts on time.

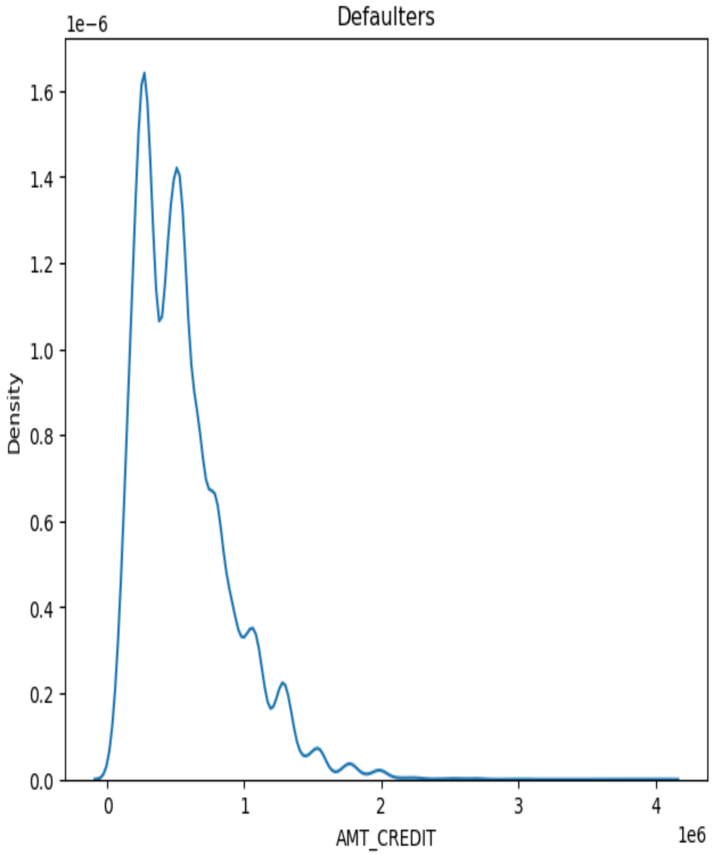
Defaulters and non-defaulters on the basis of CREDIT_GROUP

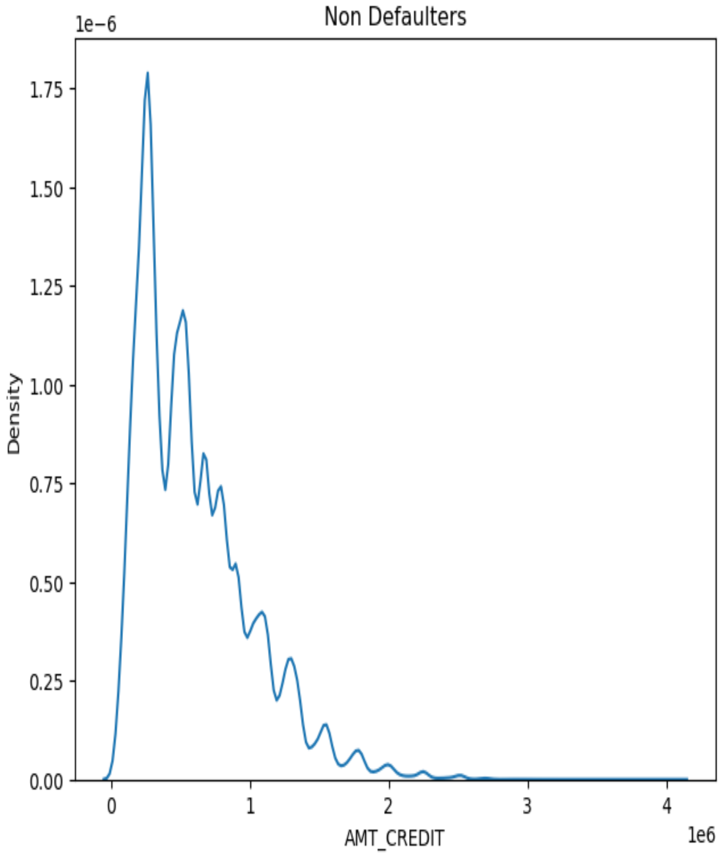


****Analysis****

1. Defaulters: As per the analysis people with age `Low` income are facing most difficulty in the payment.
2. Non-Defaulters: But also the people who are paying on time belongs to the `Low` income group.

Defaulters and non-defaulters on the basis of AMT_CREDIT

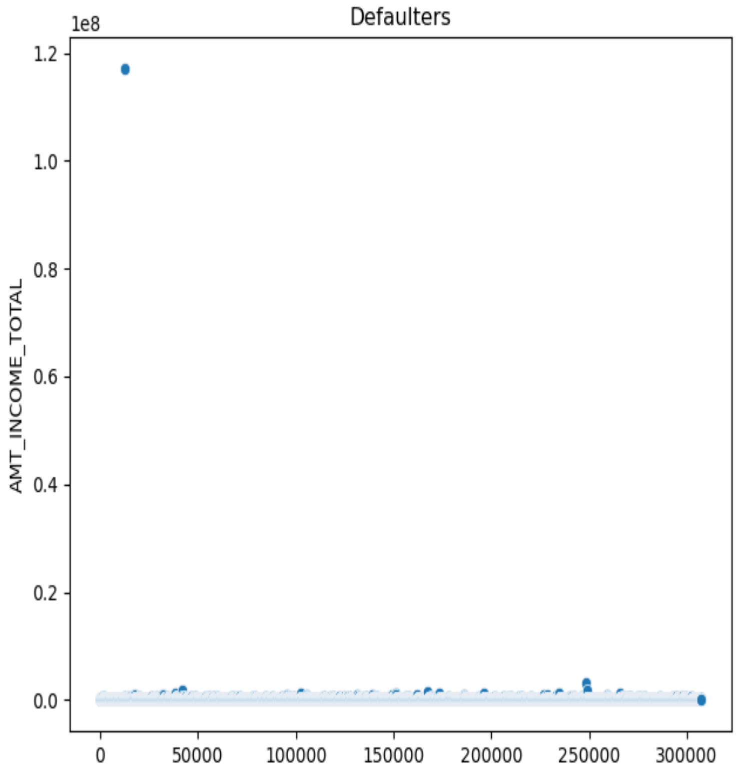


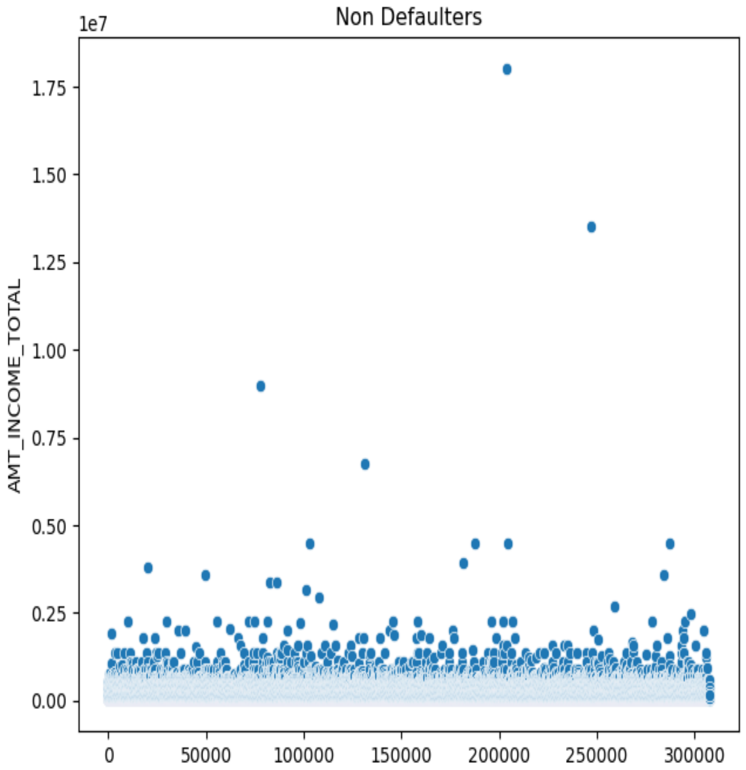


****Analysis****

1. Defaulters: The lesser the amount of loan the more chance of being defaulter, here the spike is till 500000.
2. Non-Defaulters: Same goes here as well the lesser the amount of loan, more chances of being non-defaulter.

Defaulters and Non-Defaulters on the basis of their Total Income

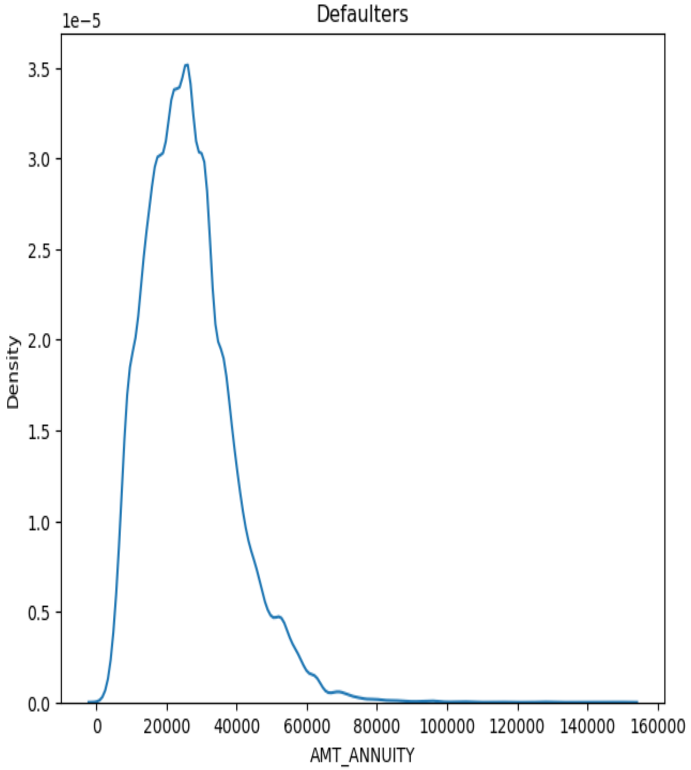


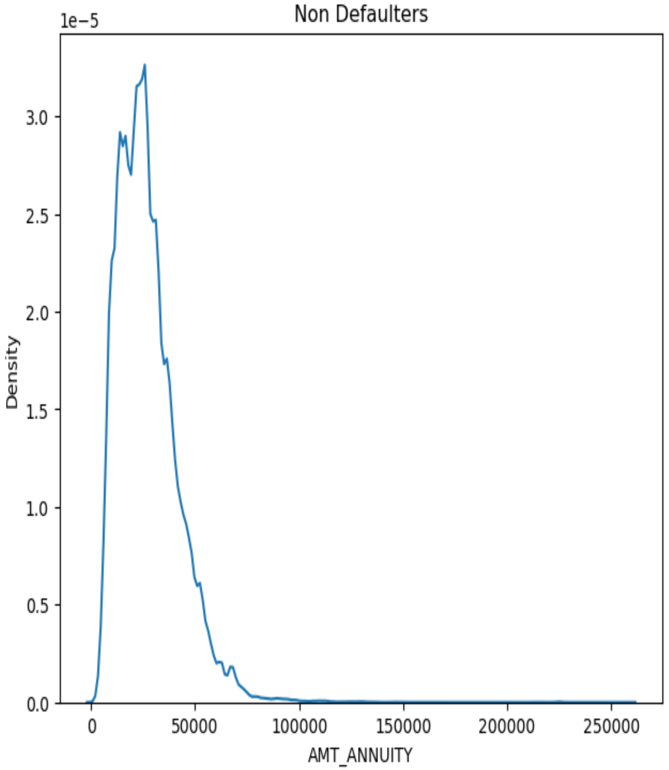


****Analysis****

1. Defaulters: As per observation most of the people with low income are defaulters.
2. Non-Defaulters: Same goes here as well the lesser the amount of income, more chances of being non-defaulters.

Defaulters and Non-Defaulters on the basis of their Annuity amount

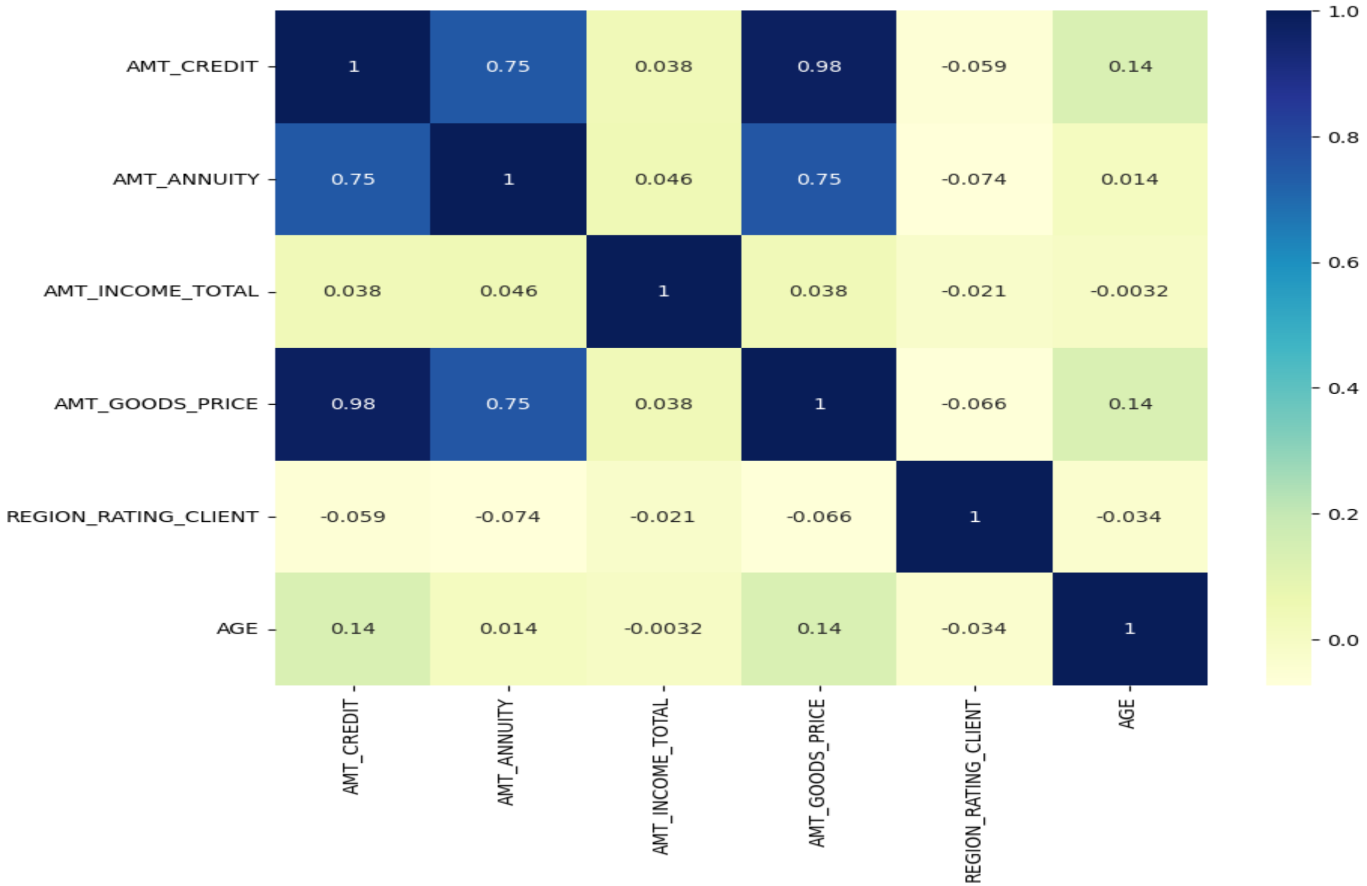




****Analysis****

1. Defaulters: As per observation most of the people with annuity amount between 2000-4000 makes more defaults.
2. Non-Defaulters: Where as the people with annuity amount of 0-5000 are not facing any difficulties in the payment.

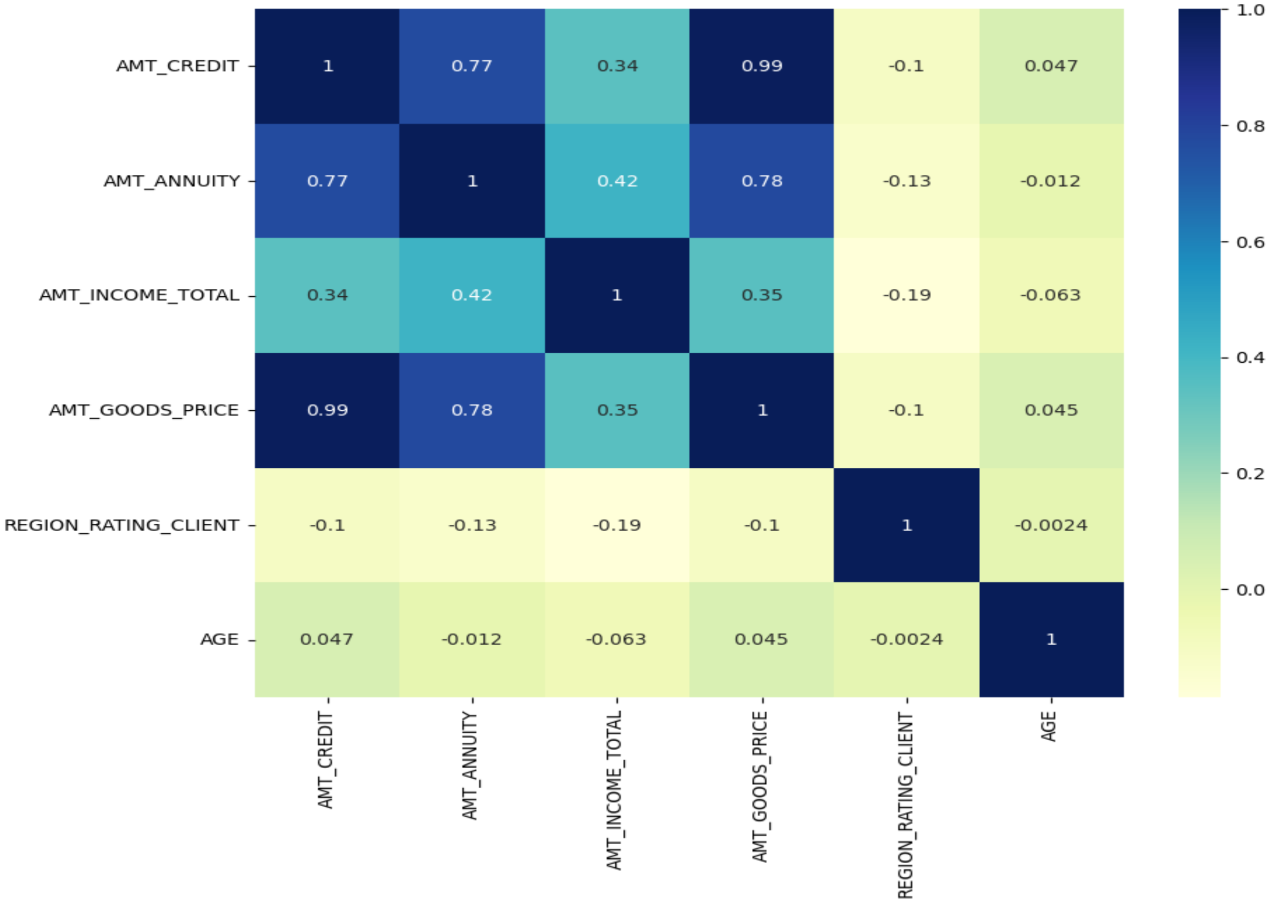
Heatmap for analysing all the numeric variables on the basis of Defaulters



****Analysis (Highly Correlted Columns)****

1. *'AMT_CREDIT'* & *'AMT_ANNUITY'* are highly correlated (0.74)
2. *'AMT_ANNUITY'* is highly correlated with *'AMT_CREDIT'* & *'AMT_GOODS_PRICE'* (0.75 each)
3. *'AMT_GOODS_PRICE'* is highly correlated with *'AMT_CREDIT'* (0.98) & *'AMT_ANNUITY'* (0.75)

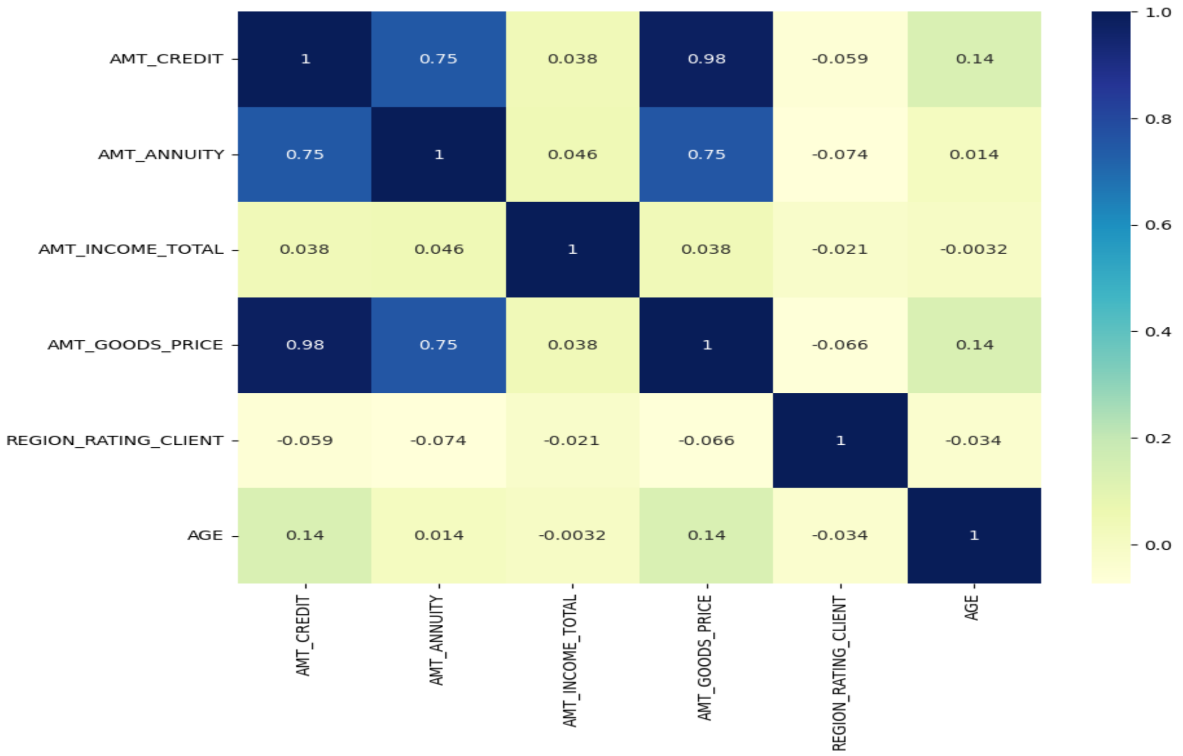
Heatmap for analysing all the numeric variables on the basis of Non-Defaulters

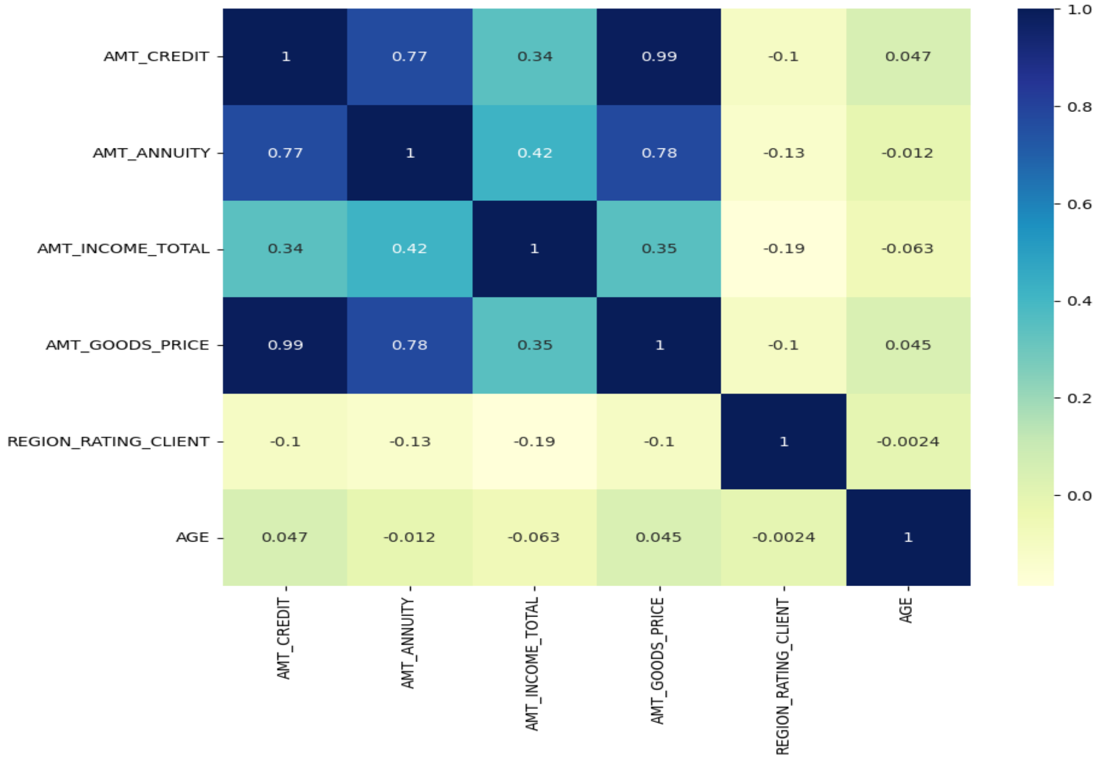


****Analysis (Highly Correlted Columns)****

1. *'AMT_CREDIT'* & *'AMT_GOODS_PRICE'* (0.99) are highly correlated
2. *'AMT_ANNUITY'* is highly correlated with *'AMT_CREDIT'*(0.77) & *'AMT_GOODS_PRICE'* (0.78)
3. *'AMT_GOODS_PRICE'* is highly correlated with *'AMT_CREDIT'* (0.99) & *'AMT_ANNUITY'* (0.78)

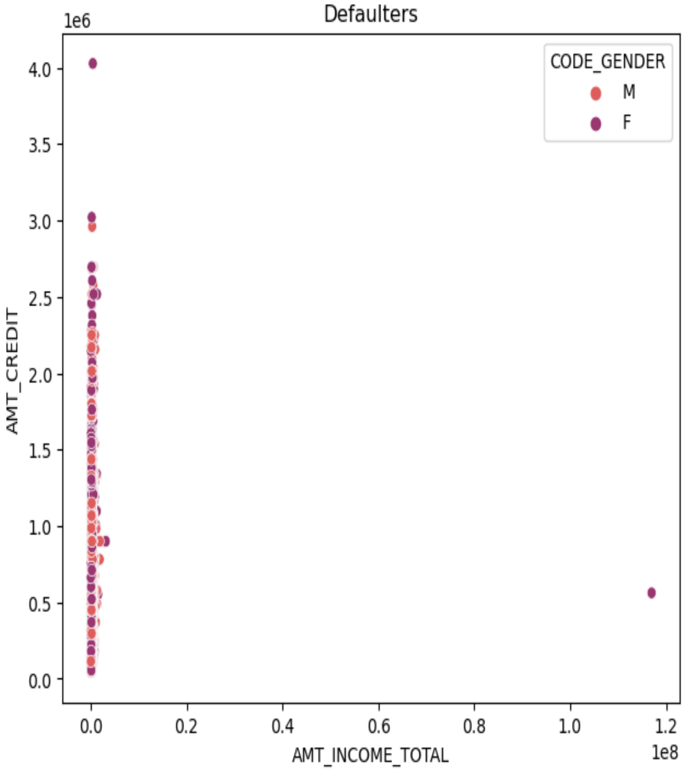
Heatmap for Defaulters and Non-Defaulters

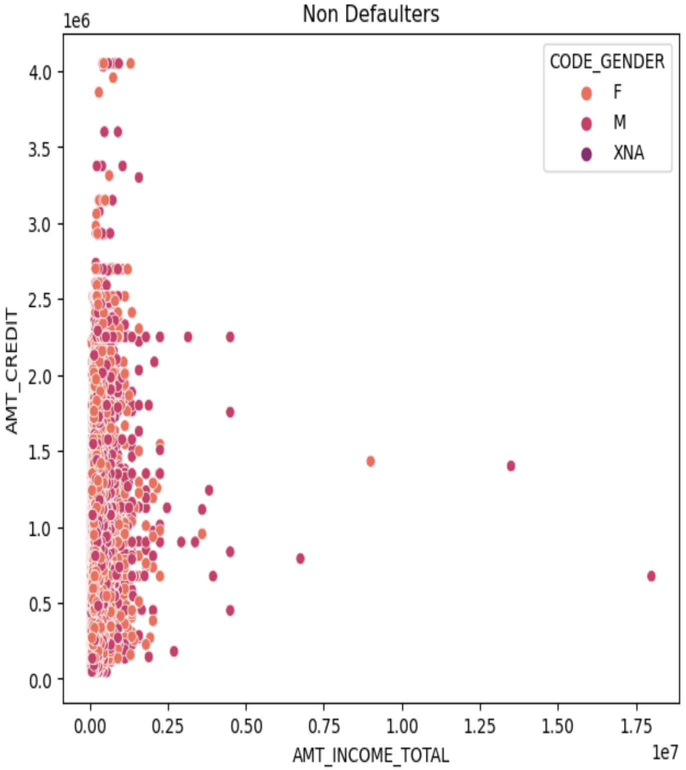




****Conclusion :****
Both defaulter and non-defaulter are highly correlated with same variables.

Credit analysis of loan on the basis of income based on gender

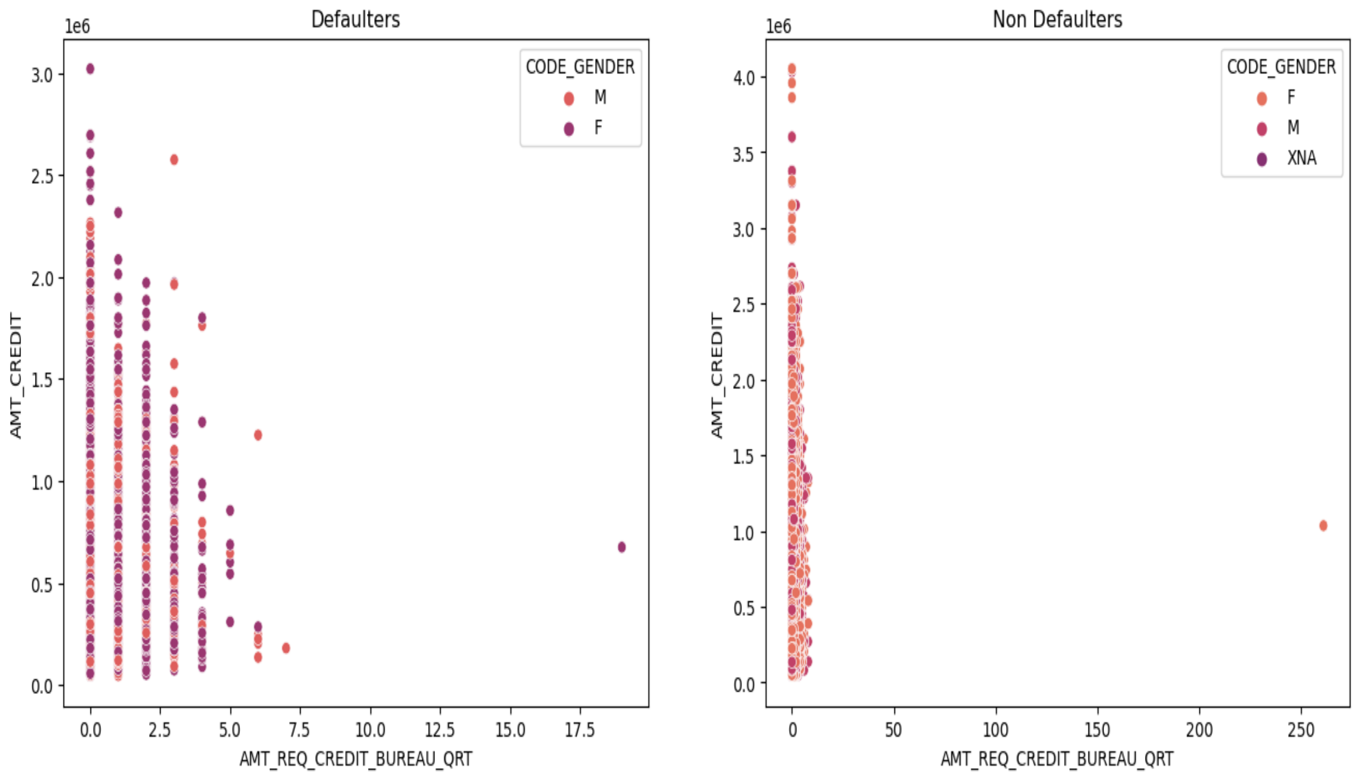




Analysis

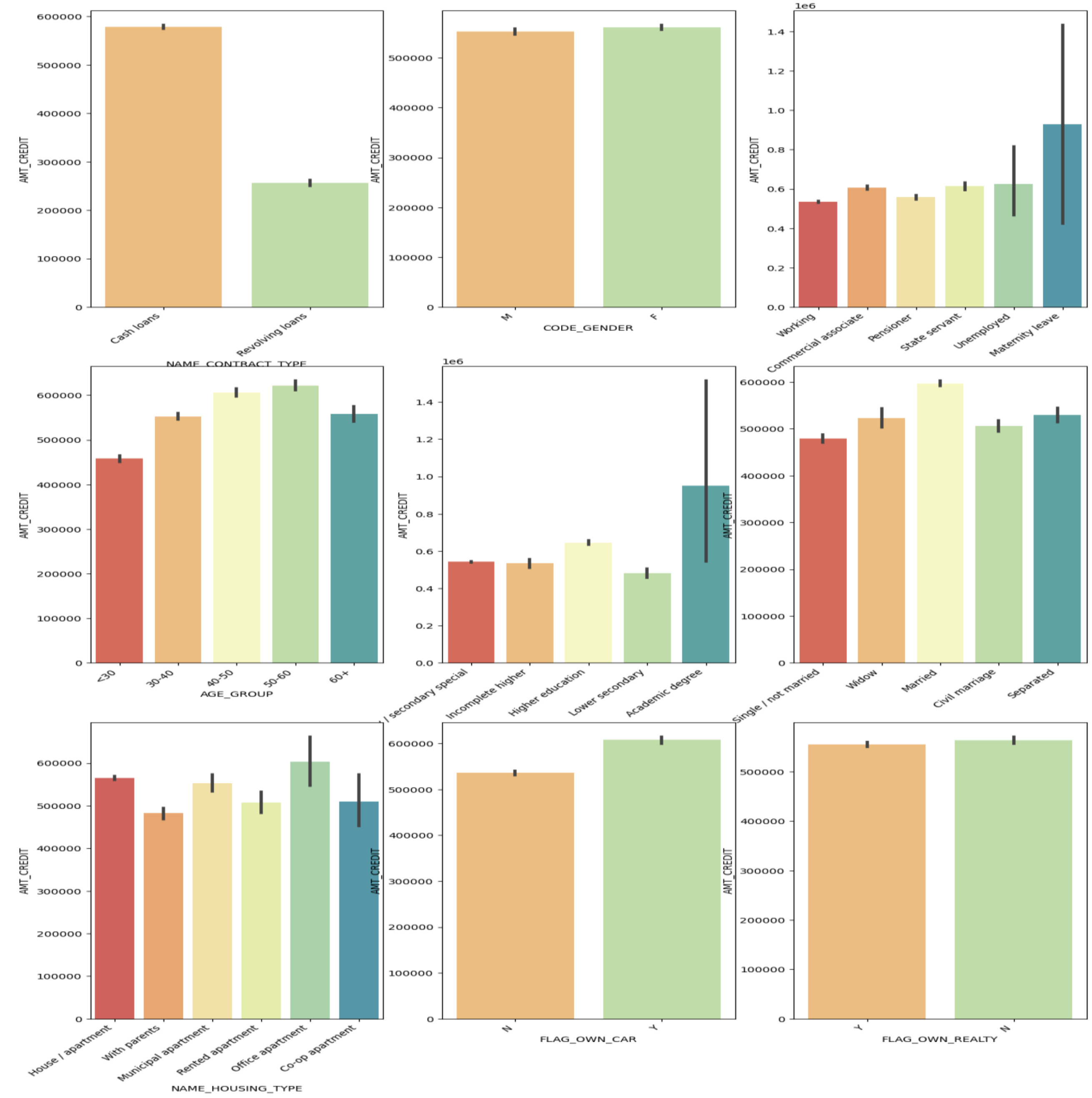
1. Defaulters: As we can observe there is no such pattern found in the data.
2. Non-Defaulters: Here we can see as that as the income increases, amount of loan also increased.

Credit amount of the loan on the basis of No. of inquires for credit score checks



Analysis

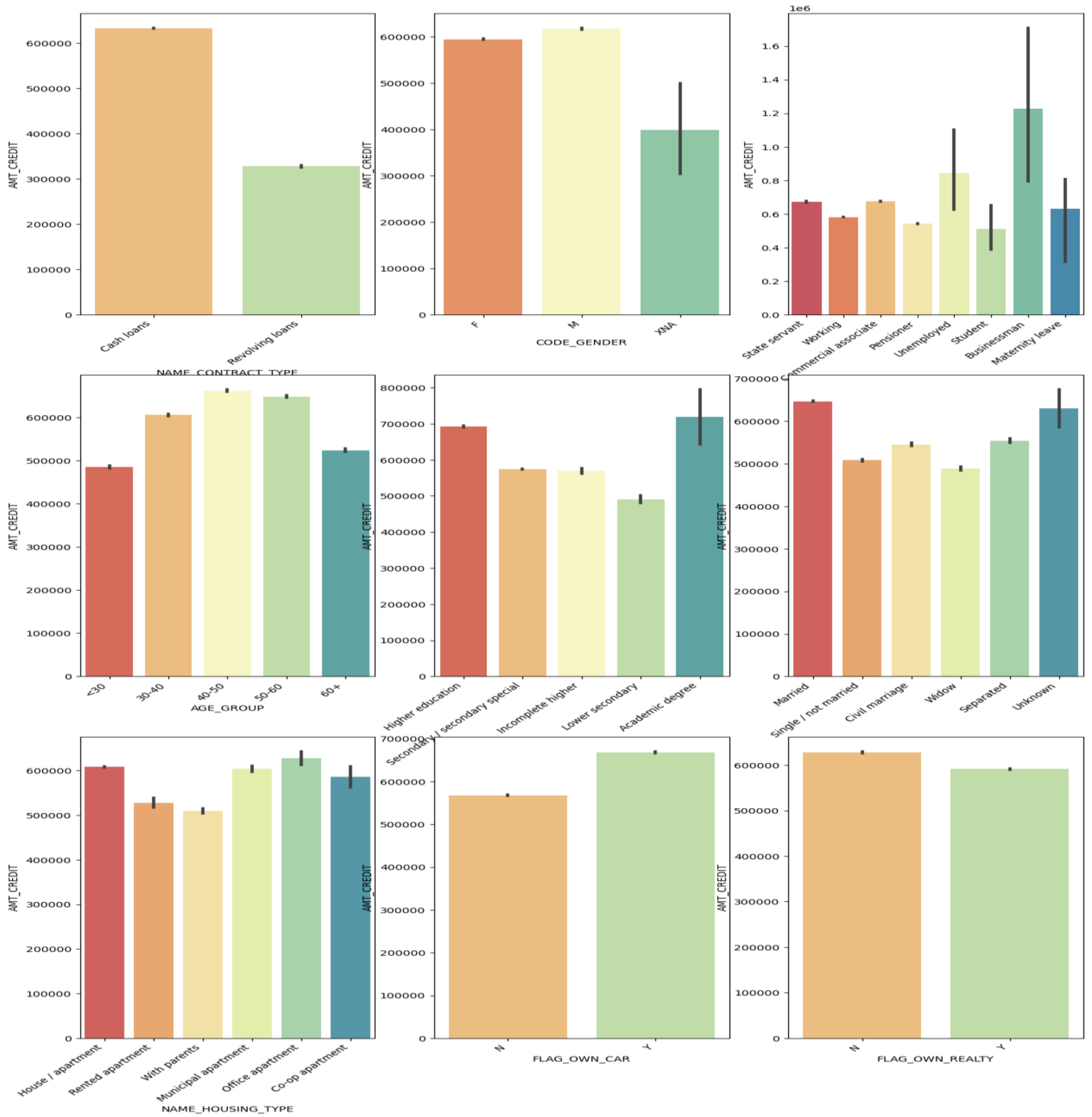
We can observe here in both the cases of Defaulters and Non Defaulter more the no. of inquiries less the credit amount.



Impact of all the categories on credit amount (Defaulters)

Analysis (Defaulters)

1. Credit amount of loan is low for revolving loans i.e. more defaults are happening in cash loans.
2. In the `CODE_GENDER` there is no major difference but females got slightly more credit amount.
3. In the `NAME_INCOME_TYPE` people with maternity leaves got more credit.
4. In the `AGE_GROUP` younger people got lesser amount of loan.
5. In `NAME_EDUCATION_TYPE` people with Academic degree got more amount of loan.
6. Married people got more loan in `NAME_FAMILY_STATUS`.
7. People with office apartment in `NAME_HOUSING_TYPE` got more credit amount.
8. clients with own car and no realty got more credit amount.



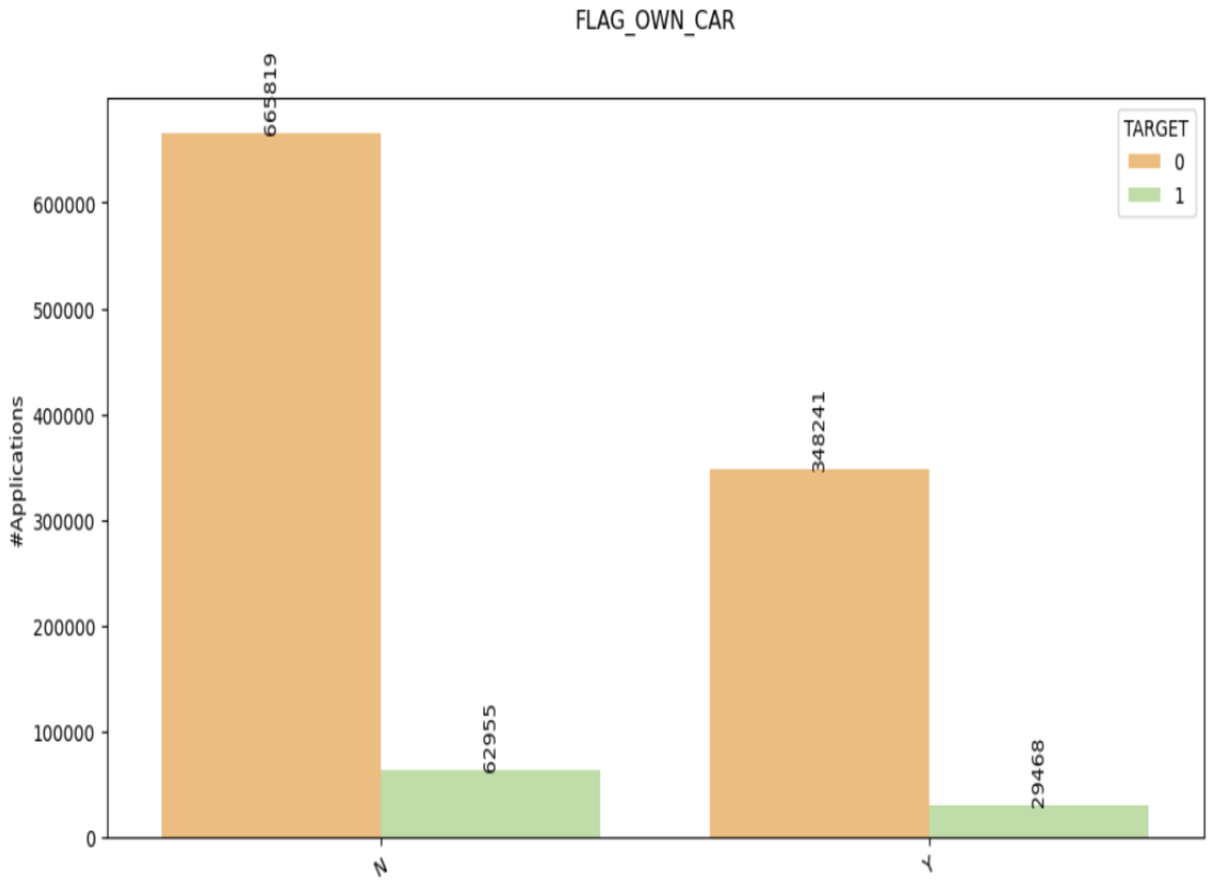
Impact of all the categories on credit amount (Non-Defaulters)

Analysis (Non -Defaulters)

1. Credit amount of loan is low for revolving loans.
2. In the `CODE_GENDER` there is no major difference but males got slightly more credit amount.
3. In the `NAME_INCOME_TYPE` businessman got more credit.
4. In the `AGE_GROUP` younger people got lesser amount of loan and people with the age of '40-50' got the highest credit amount.
5. In `NAME_EDUCATION_TYPE` people with Academic degree got more amount of loan.
6. Married people got more loan in `NAME_FAMILY_STATUS`.
7. People with office apartment in `NAME_HOUSING_TYPE` got more credit amount.
8. clients with own car and no realty got more credit amount.

Current and Previous Application Dataset Merged Analysis

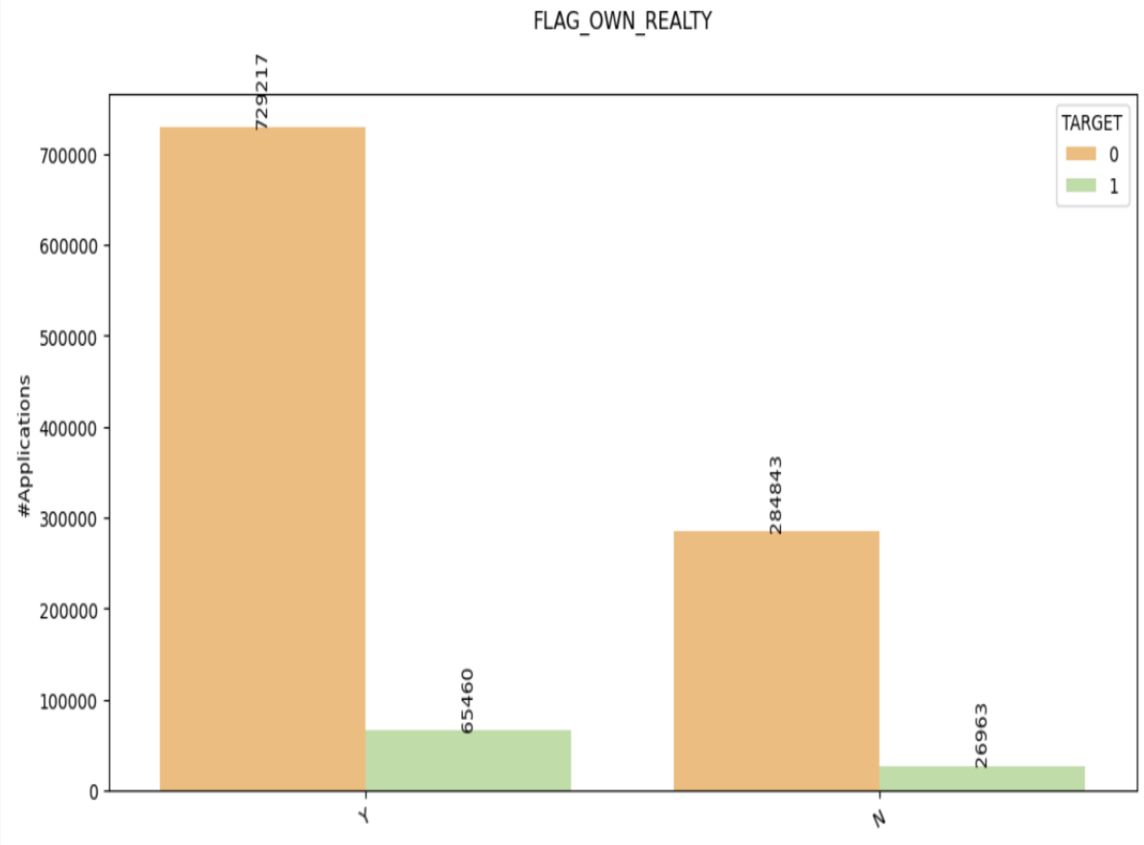
Defaulters and non-defaulters on the basis of own car



Analysis

People who own car and does not own car does not give any conclusion about payment difficulties.

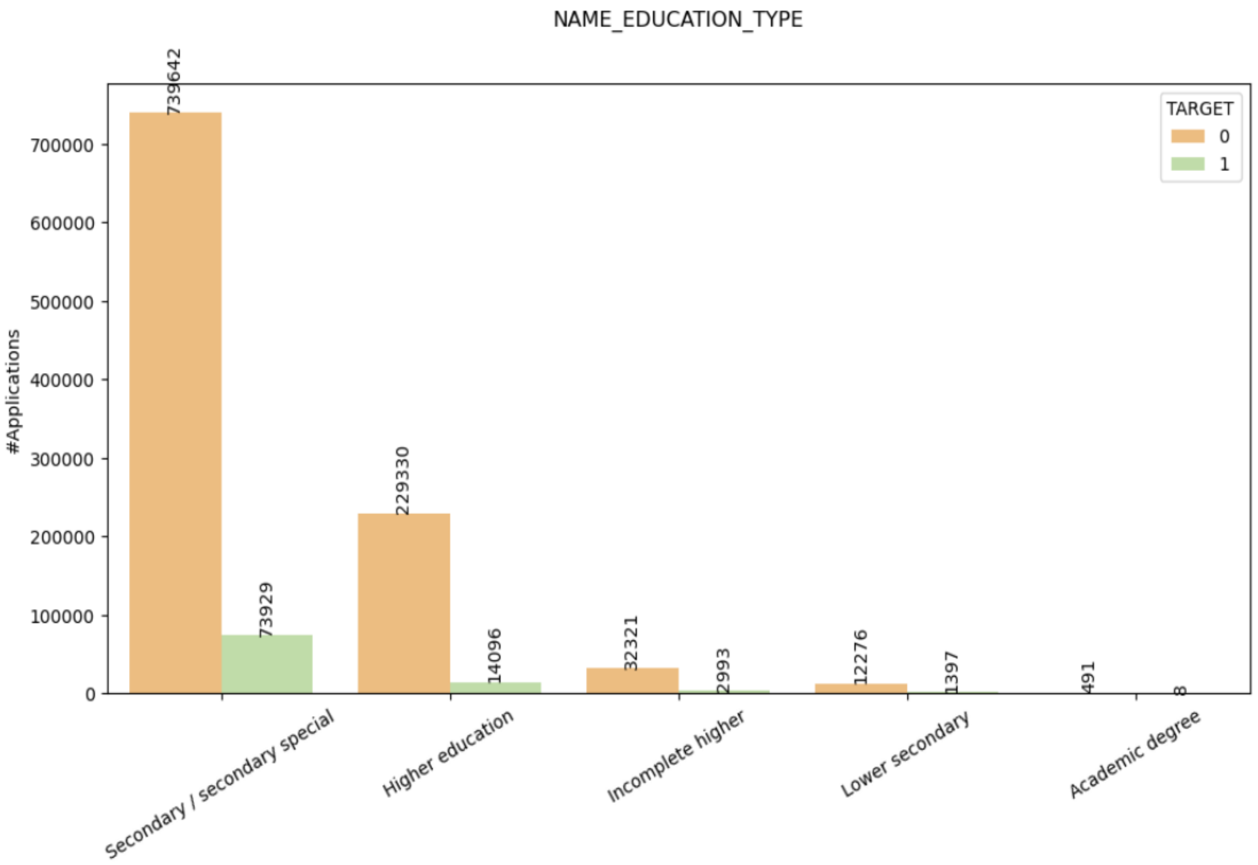
Defaulters and non-defaulters on the basis of own realty



Analysis

Here also the people who own realty or does not own realty are not giving any conclusion regarding payment difficulties.

Defaulters and non-defaulters on the basis of education type

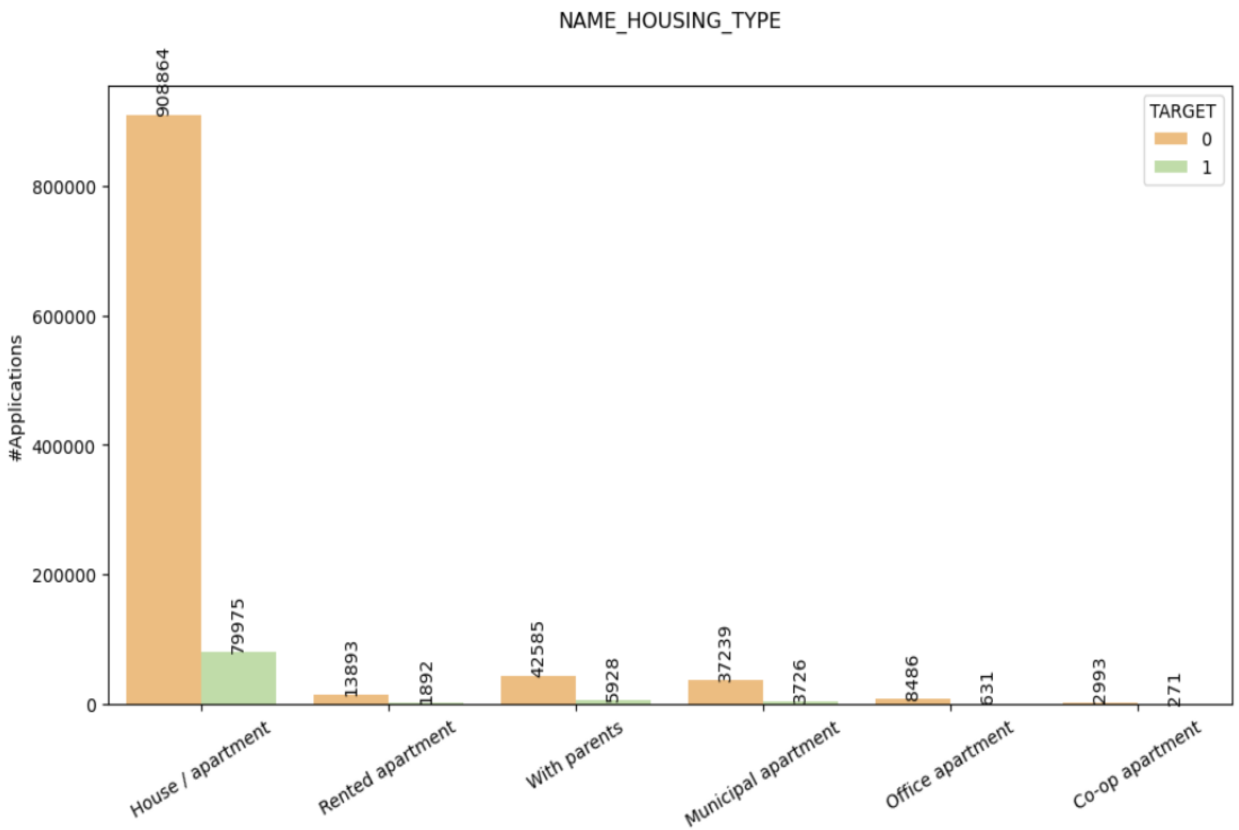


****Analysis****

1. Defaulters: People who studied secondary and secondary special are paying on time.
2. Non-Defaulters: The people who are facing difficulties also fall in the secondary and secondary special segment.

Hence it is not that conclusive.

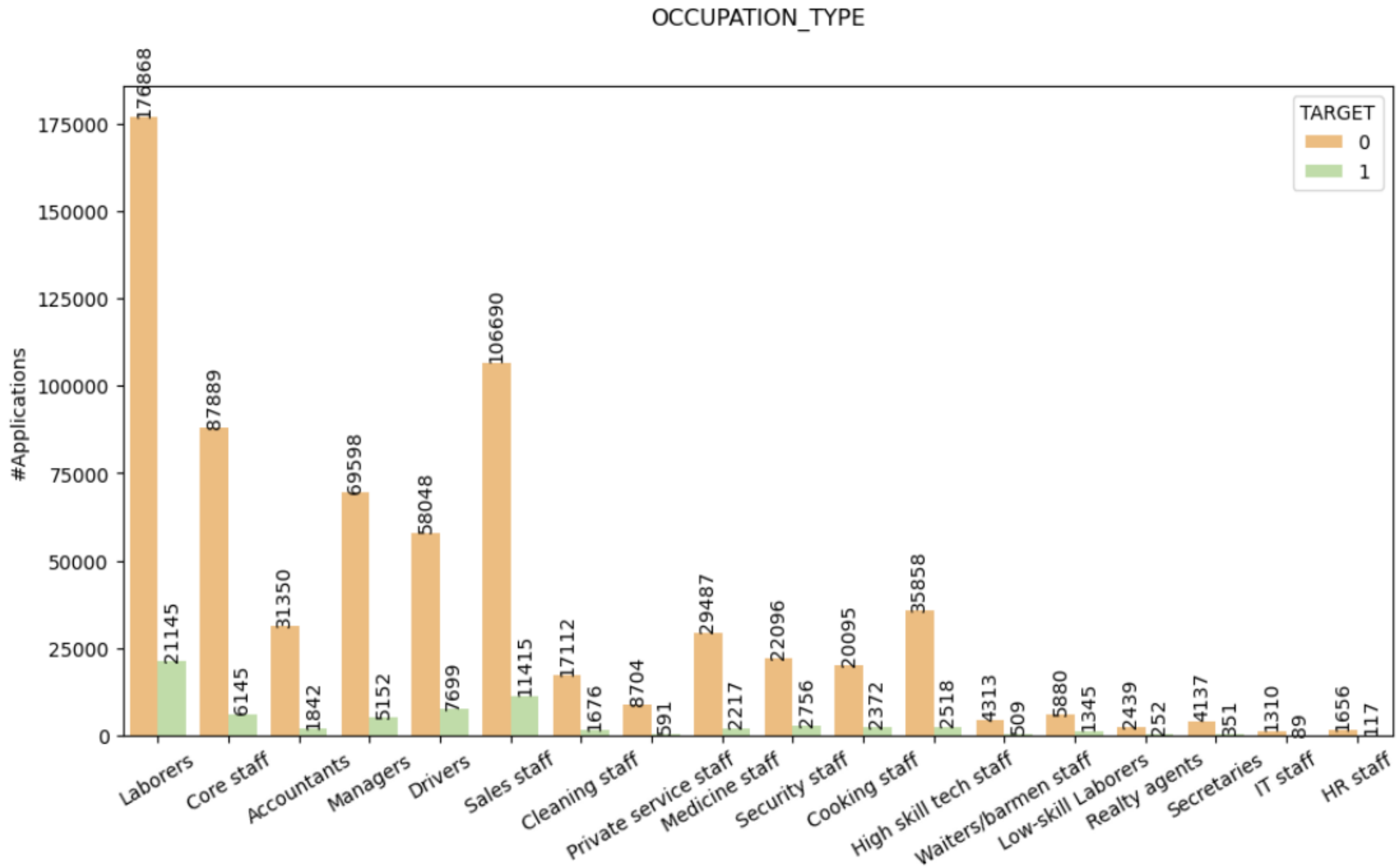
Defaulters and non-defaulters on the basis of Housing type



****Analysis****

People with Housing type of House/Apartment both are high in case of defaulters and non-defaulters.

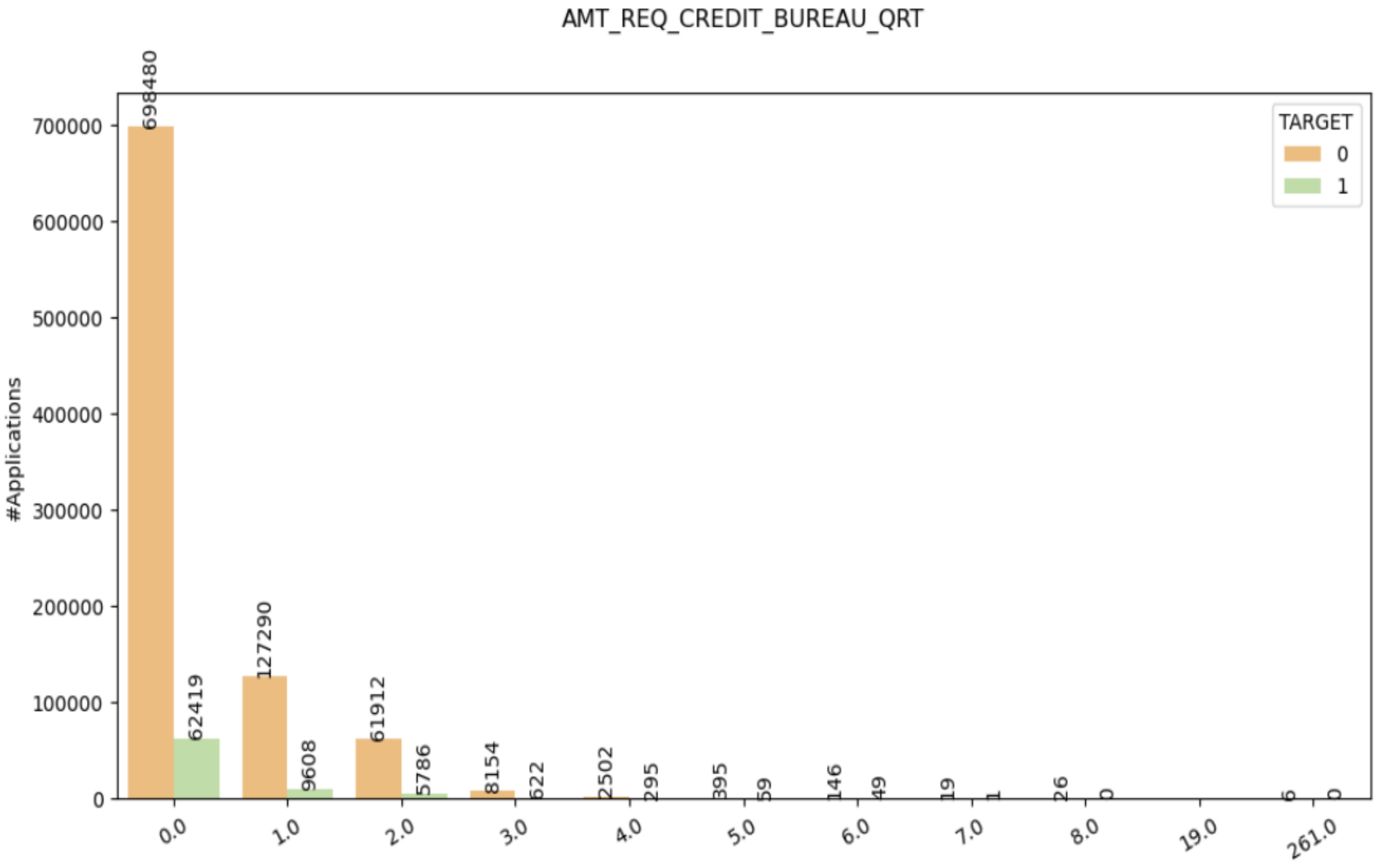
Defaulters and non-defaulters on the basis of occupation type



Analysis

1. Defaulters: Labors and Drivers both pays on time are tend to be less defaulters.
2. Non-Defaulters: Here as well the Labors are the one who faces most difficulty in paying back.

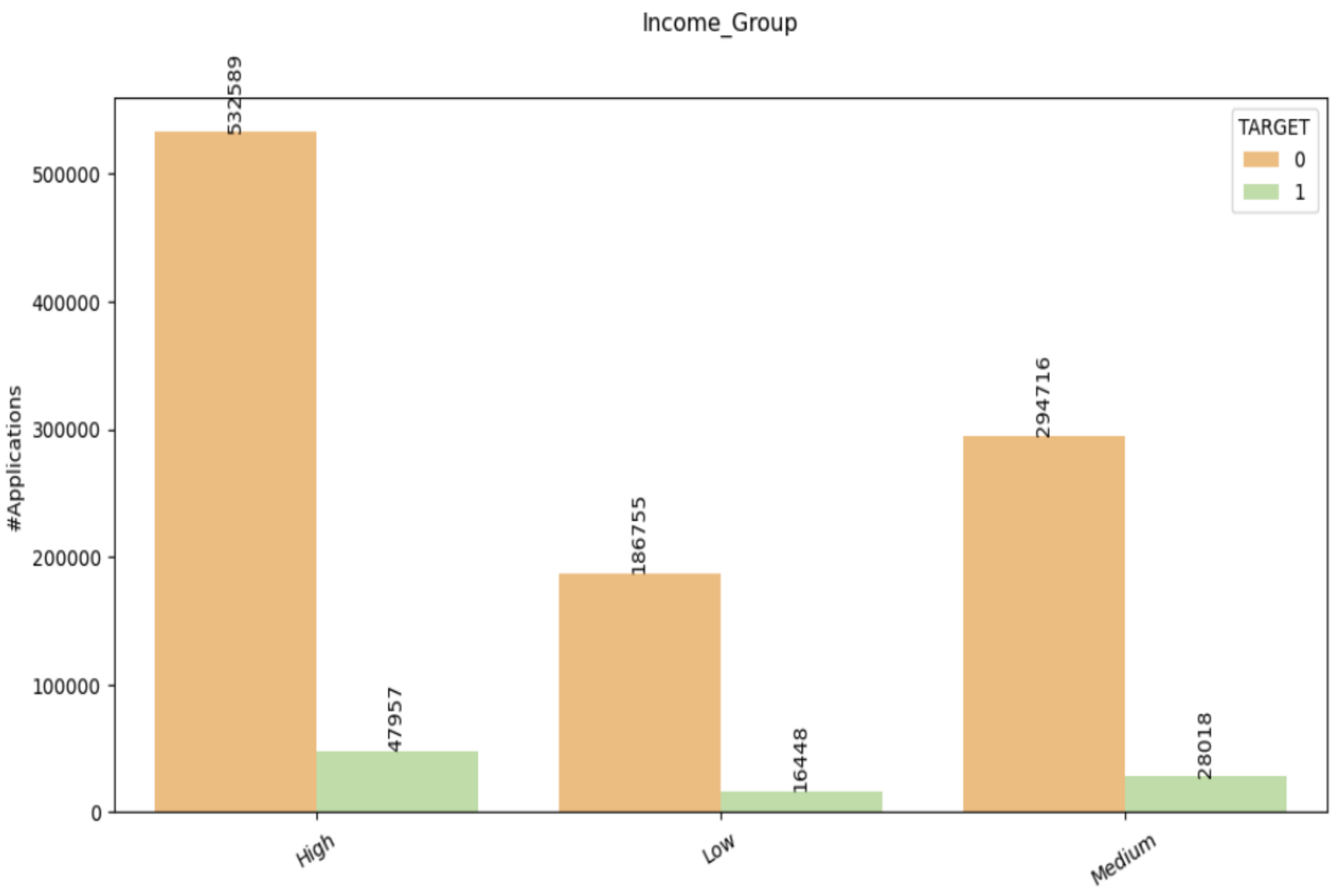
Defaulters and non-defaulters on the basis of credit bureau quarterly report



Analysis

1. Defaulters: People who never checked their credit score were paying on time.
2. Non-Defaulters: People who are facing difficulties in the payment also didn’t checked their credit score.

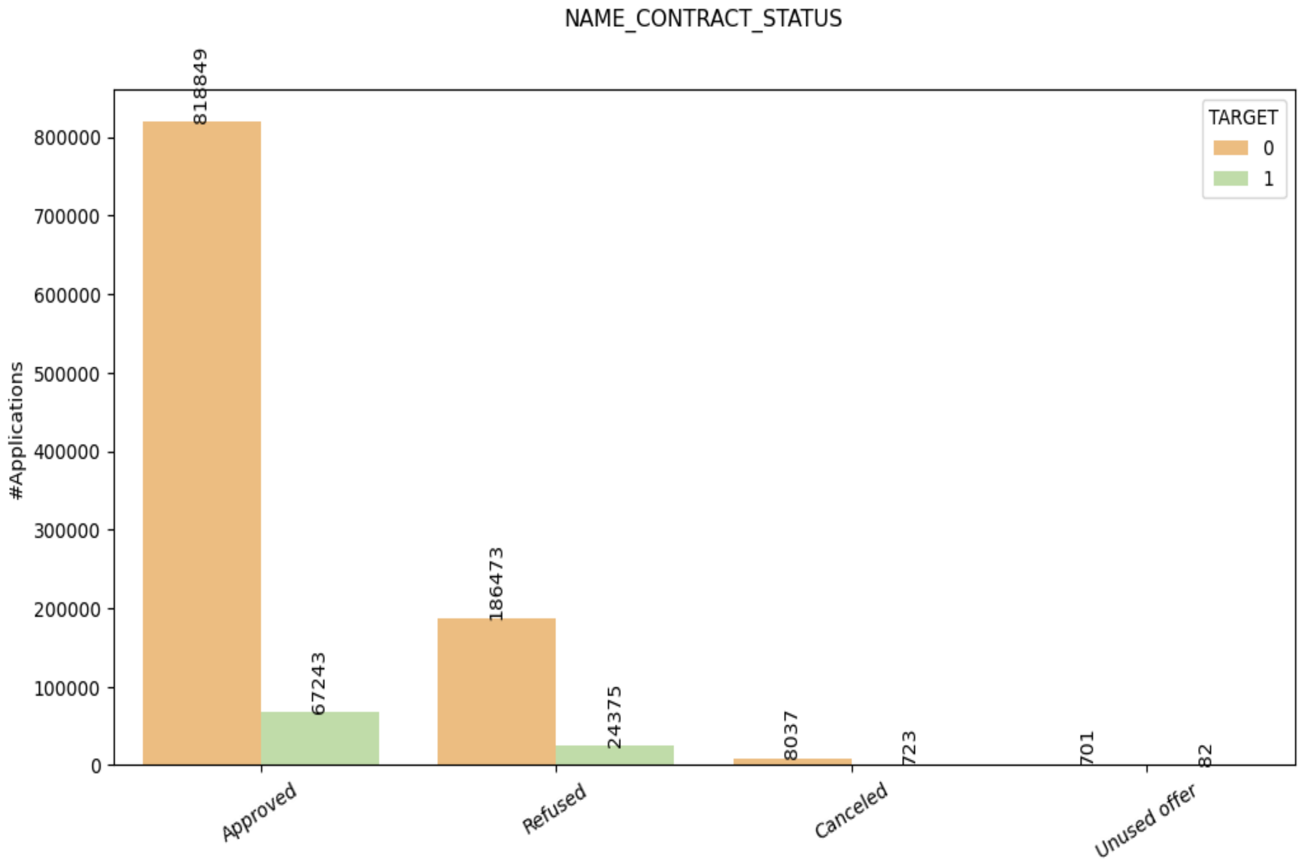
Defaulters and non-defaulters on the basis of income group



Analysis

People with high income are falling in both the segments defaulter and non-defaulters.

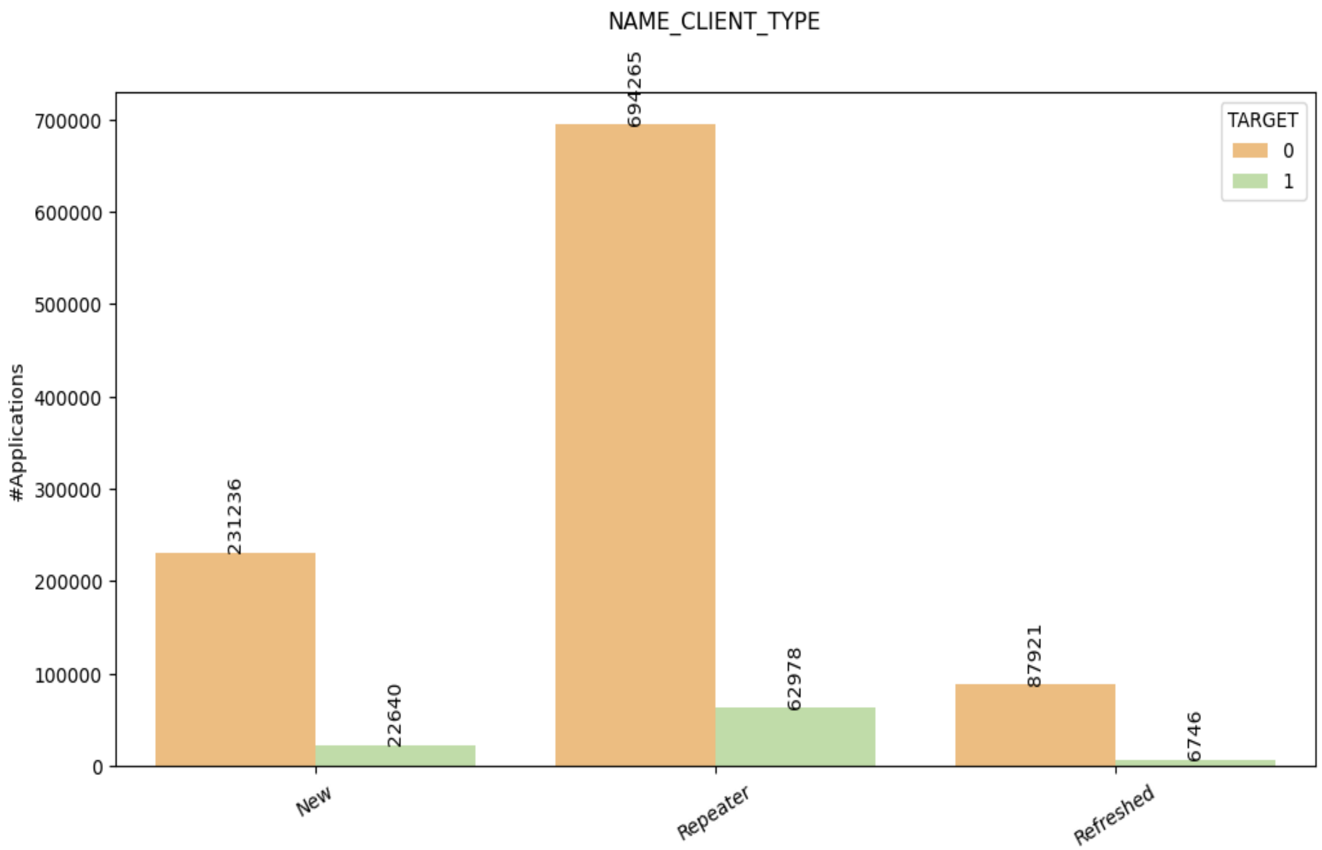
Defaulters and non-defaulters on the basis of contract status



****Analysis****

The people whose loan got approved were more defaulters and this same segment was paying on time as well.

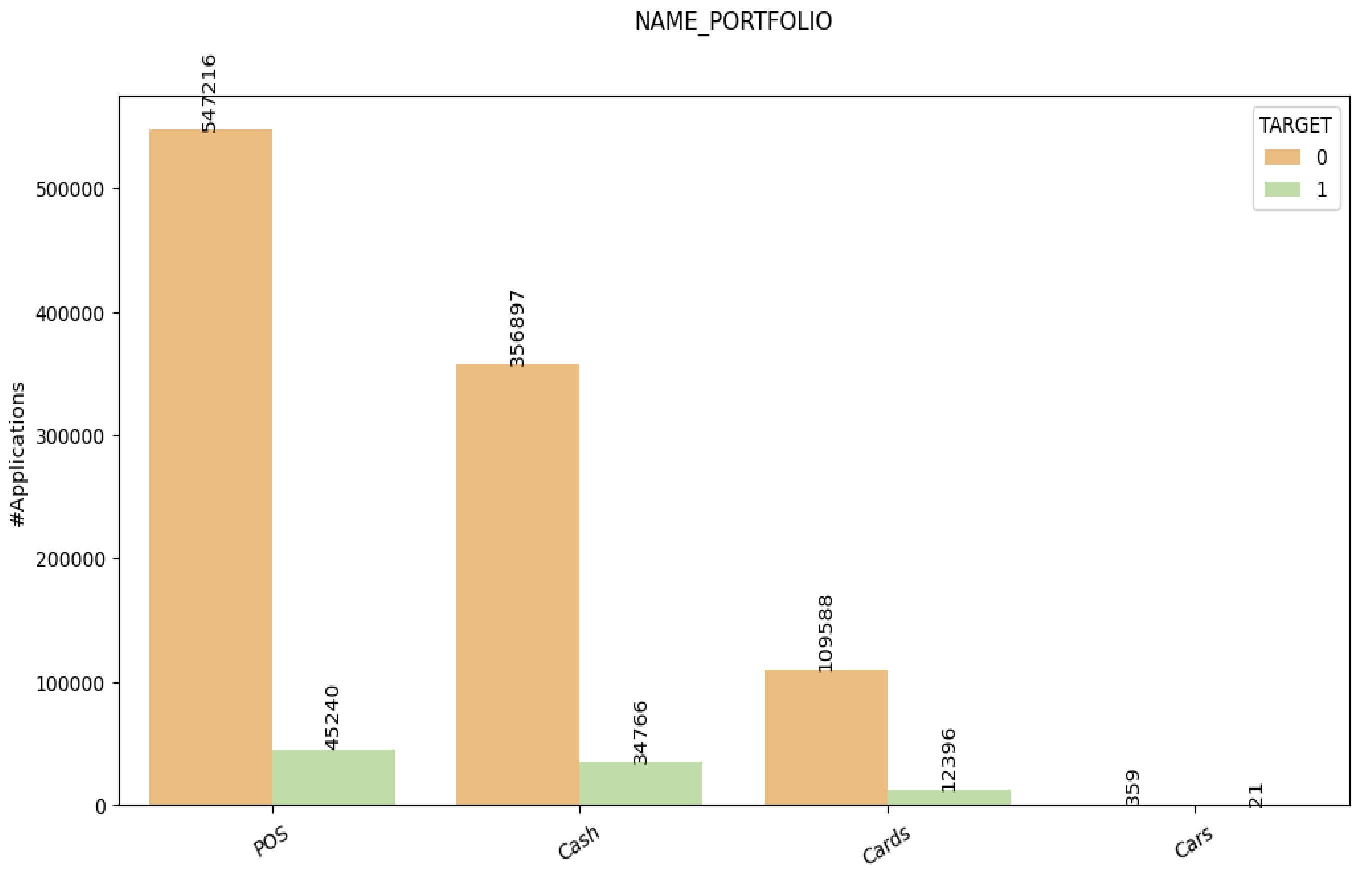
Defaulters and non-defaulters on the basis of client type



Analysis

The people who were repeater were defaulters as well as non-defaulters, for this the banks can check their previous loan history whoever is paying on time and having a good credit history tend to be non-defaulters.

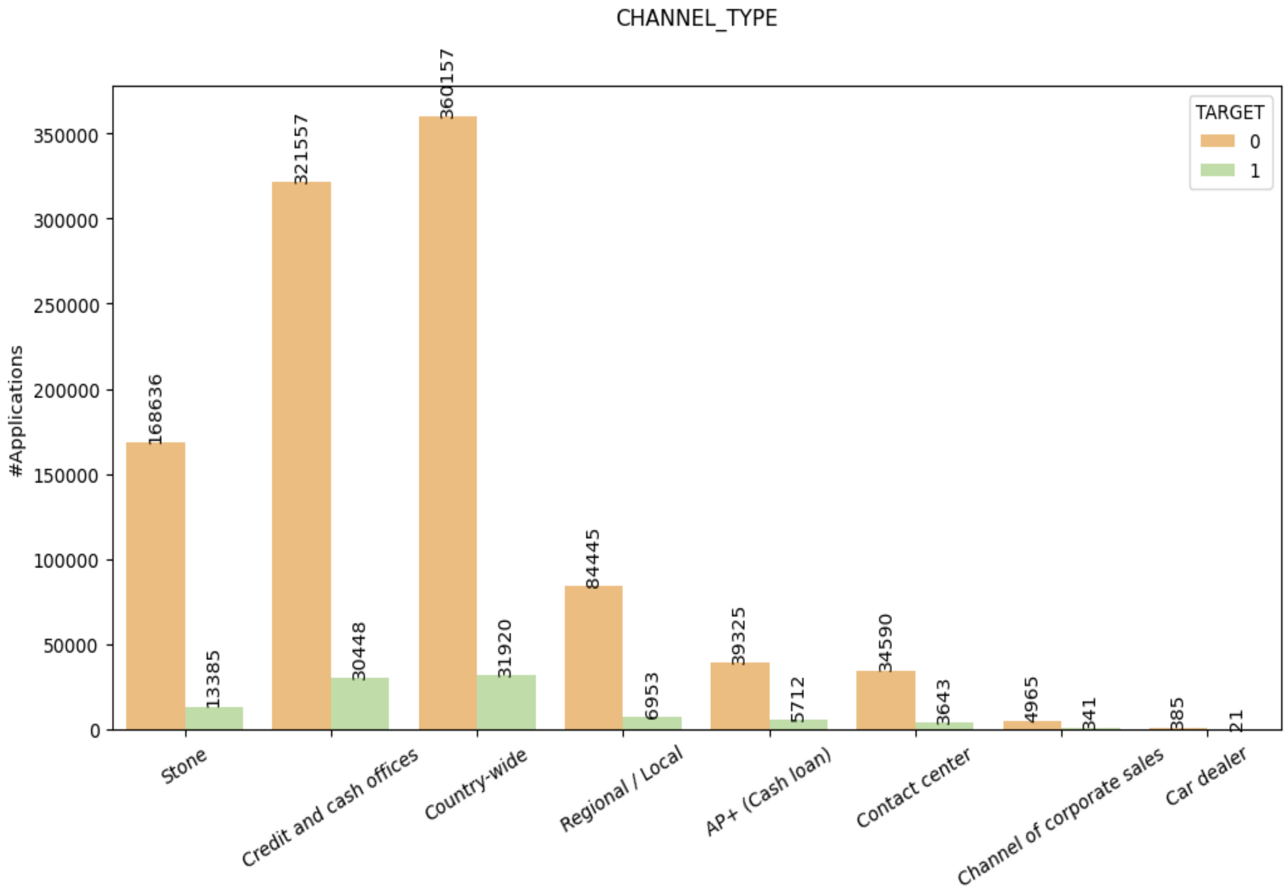
Defaulters and non-defaulters on the basis of portfolio



Analysis

People with POS were more defaulter and non defaulters.

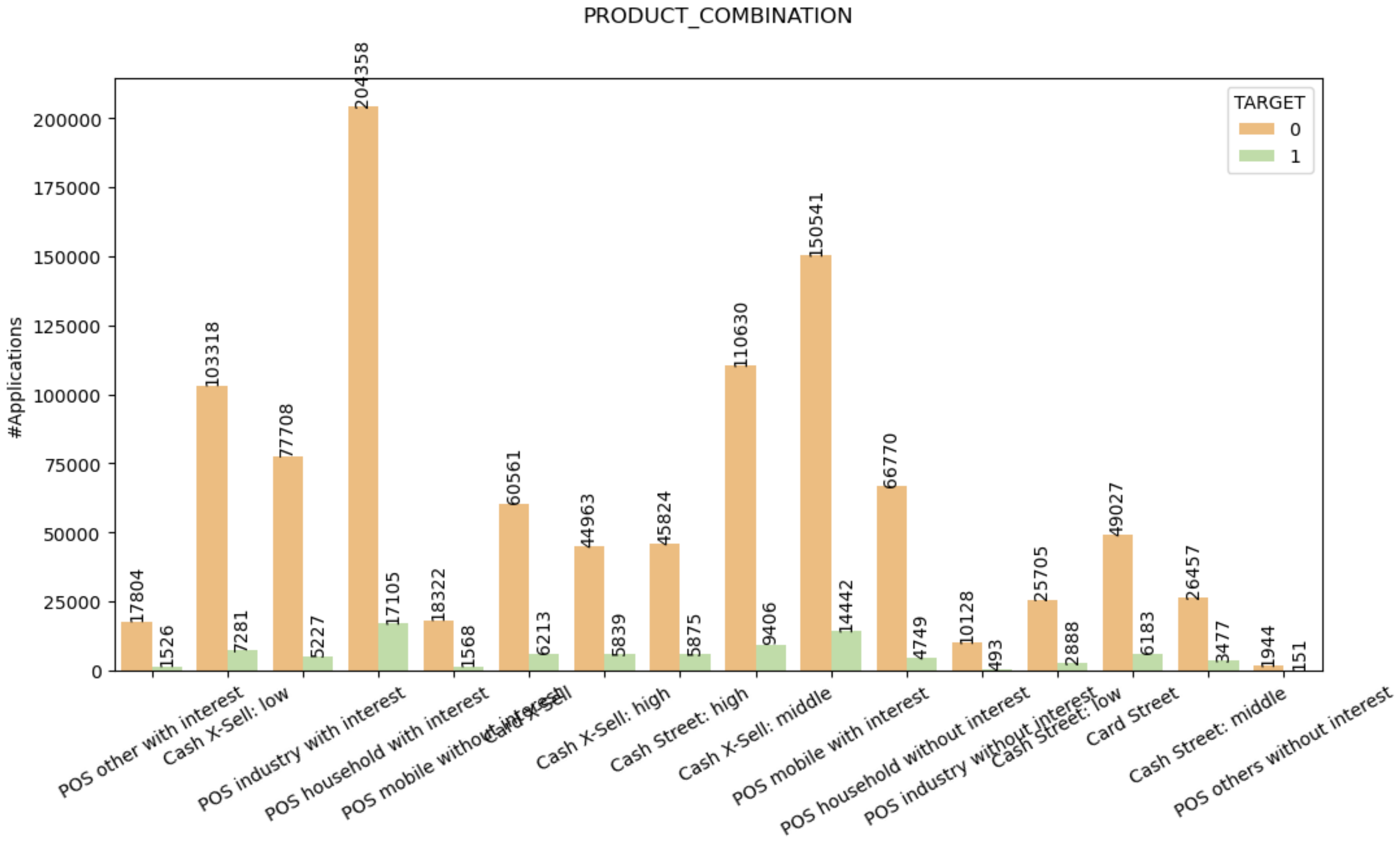
Defaulters and non-defaulters on the basis of channel type



****Analysis****

1. Defaulters: As we can observe country wide channel has higher applications for loan and also tend to be more defaulters in this.
2. Non-Defaulters: The people who are not facing difficulties in the payment belongs to country wide channel.

Defaulters and non-defaulters on the basis of product combination



****Analysis****

People with POS industry and interest are falling in the high segment of both defaulters and non-defaulters.

THANK YOU