BANK CREDIT EXPLORATORY DATA ANALYSIS

Mathankumar Selvaraj | 30th July,2021

CONTENTS

OBJECTIVE

About Data Set:

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:

1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2) 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.

Two type of data provided : Application Data and Previous Application.
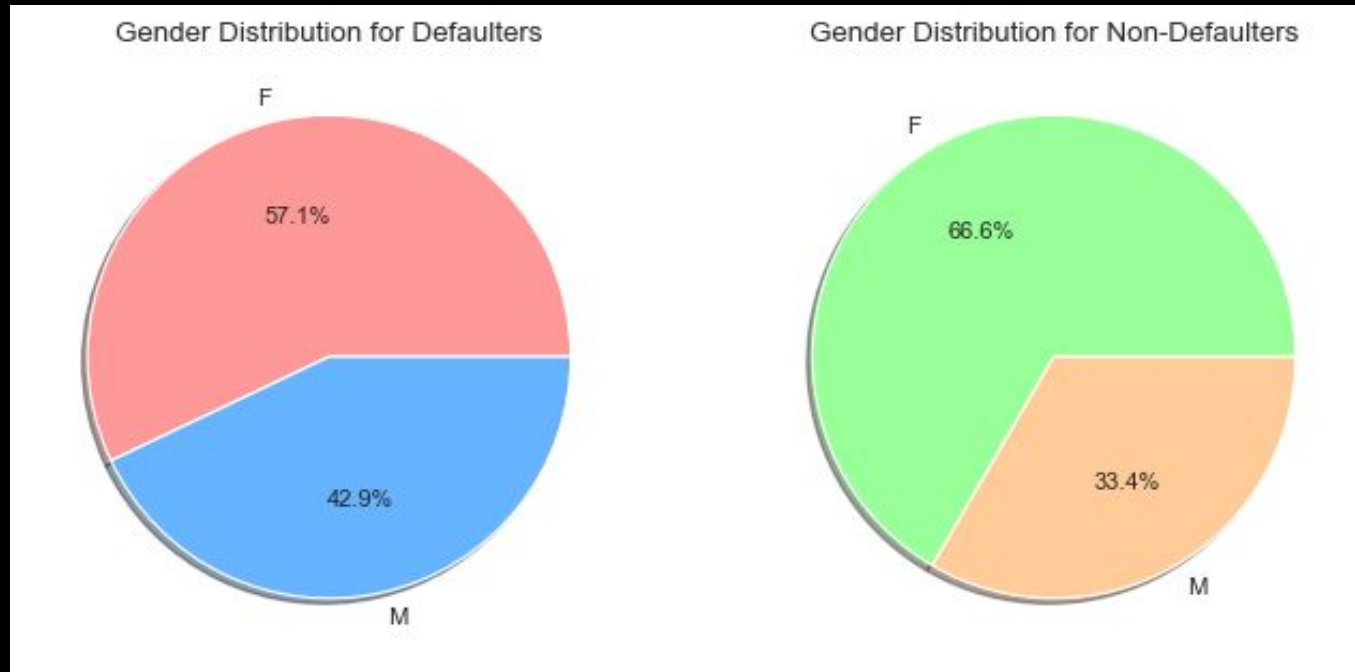Data Source: Bank database Timeframe: Not mentioned

The proportion of Defaulter and Non-Defaulter in Application Data
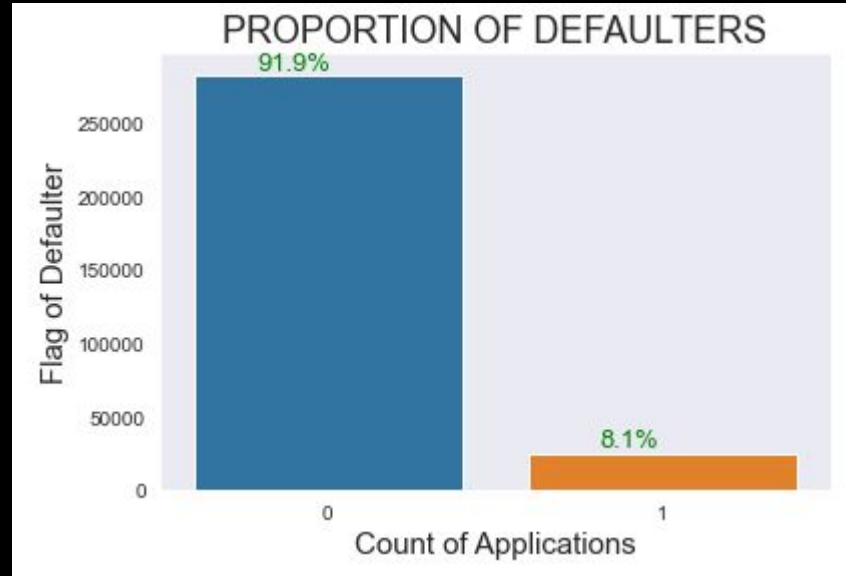
1) In the Application data 8.% applicants are not able to pay on time

The distribution of gender for Defaulter and Non-Defaulter

Gender Distribution for Defaulters

F
57.1%

M
42.9%

Gender Distribution for Non-Defaulters
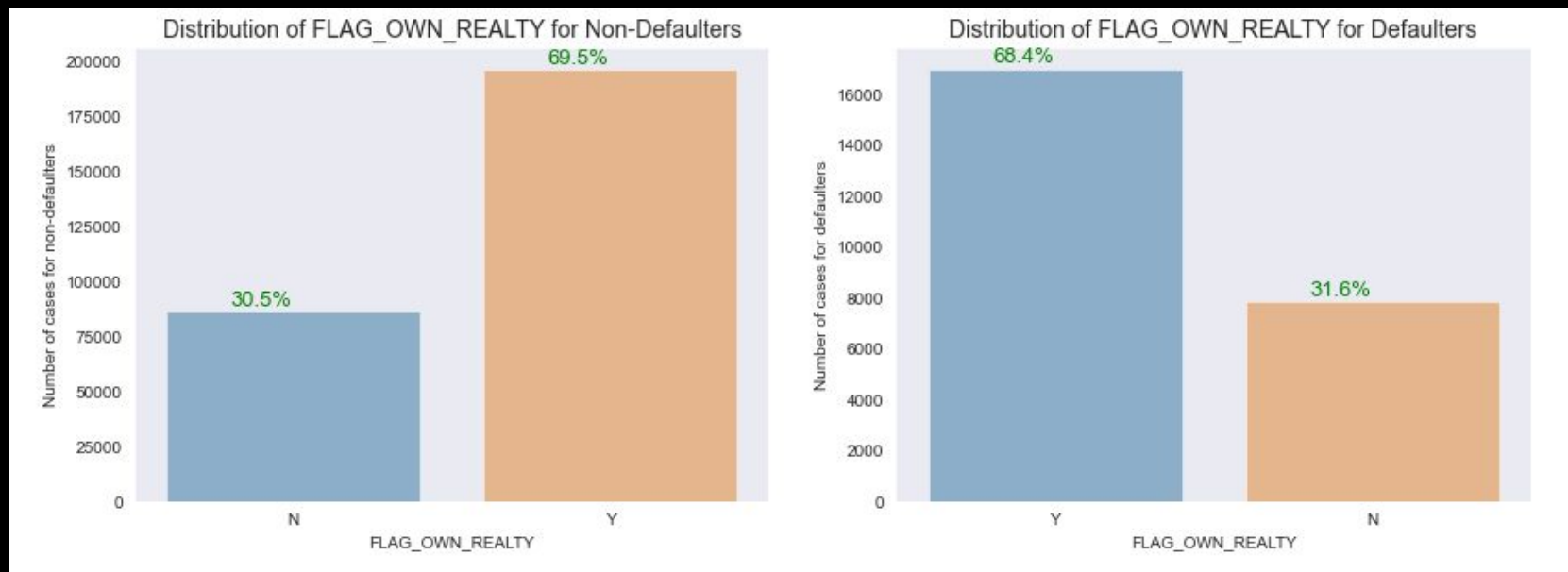
F
66.6%

M
33.4%

57% female are defaulter. Which is slightly higher than the male defaulter. Also, 66% female not defaulters

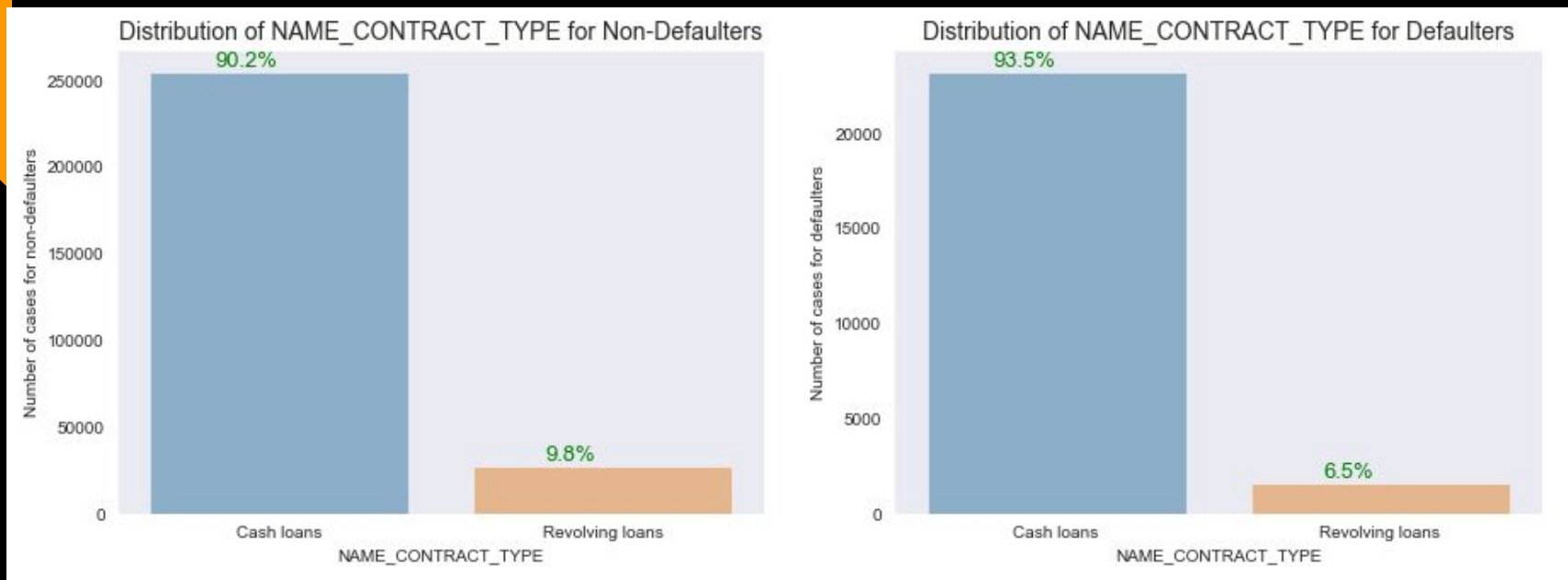# The count of applicants who has car for both case
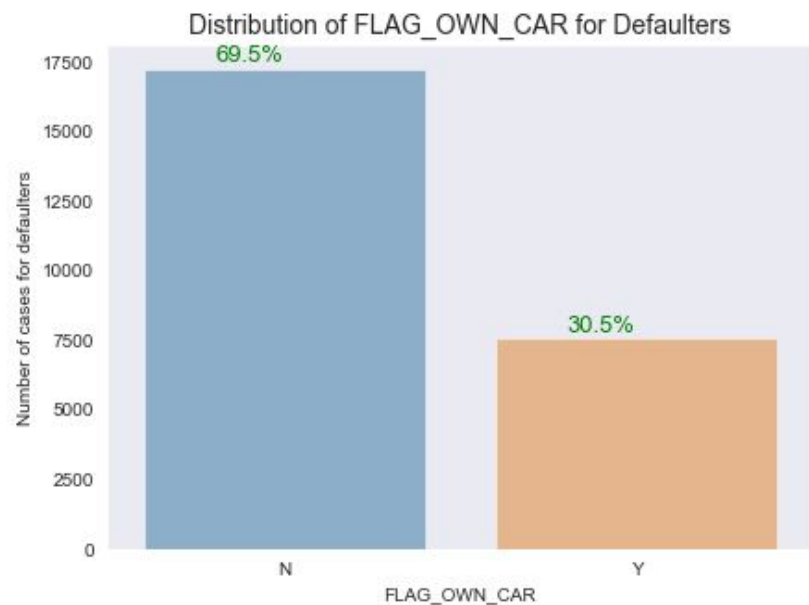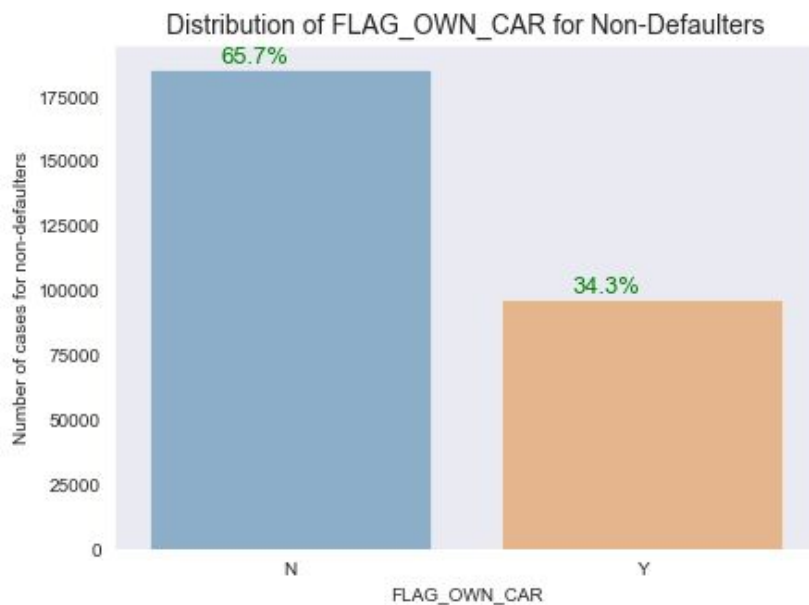
Almost,65% of applicants are in the both case. Which indicates owning car doesn't affect the repay of loan

# The count of applicants who has realty for both case

The 68% applicants were defaulter who has realty.

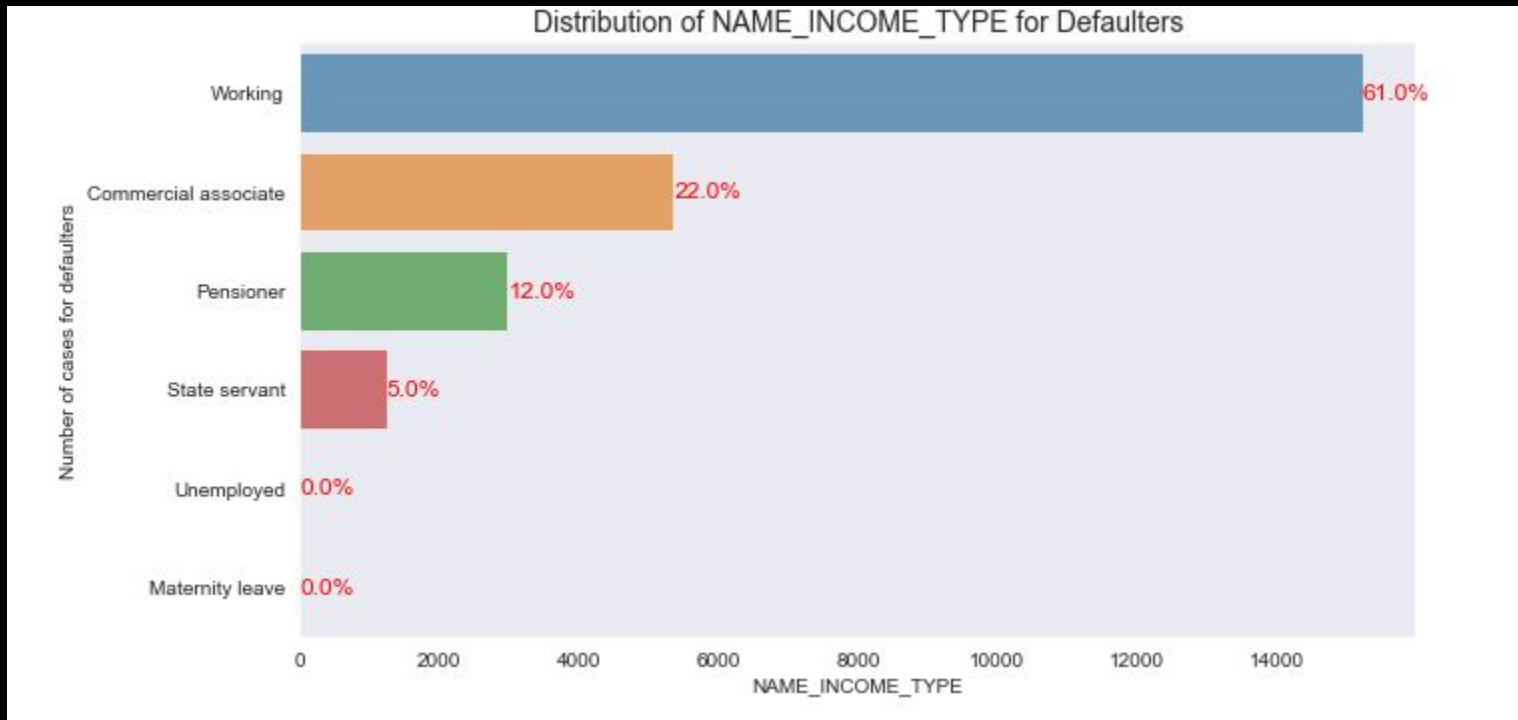# The count of applicants based on the contract type for both case

For contract type 'cash loans' is having higher number of credits than 'Revolving loans' contract type.

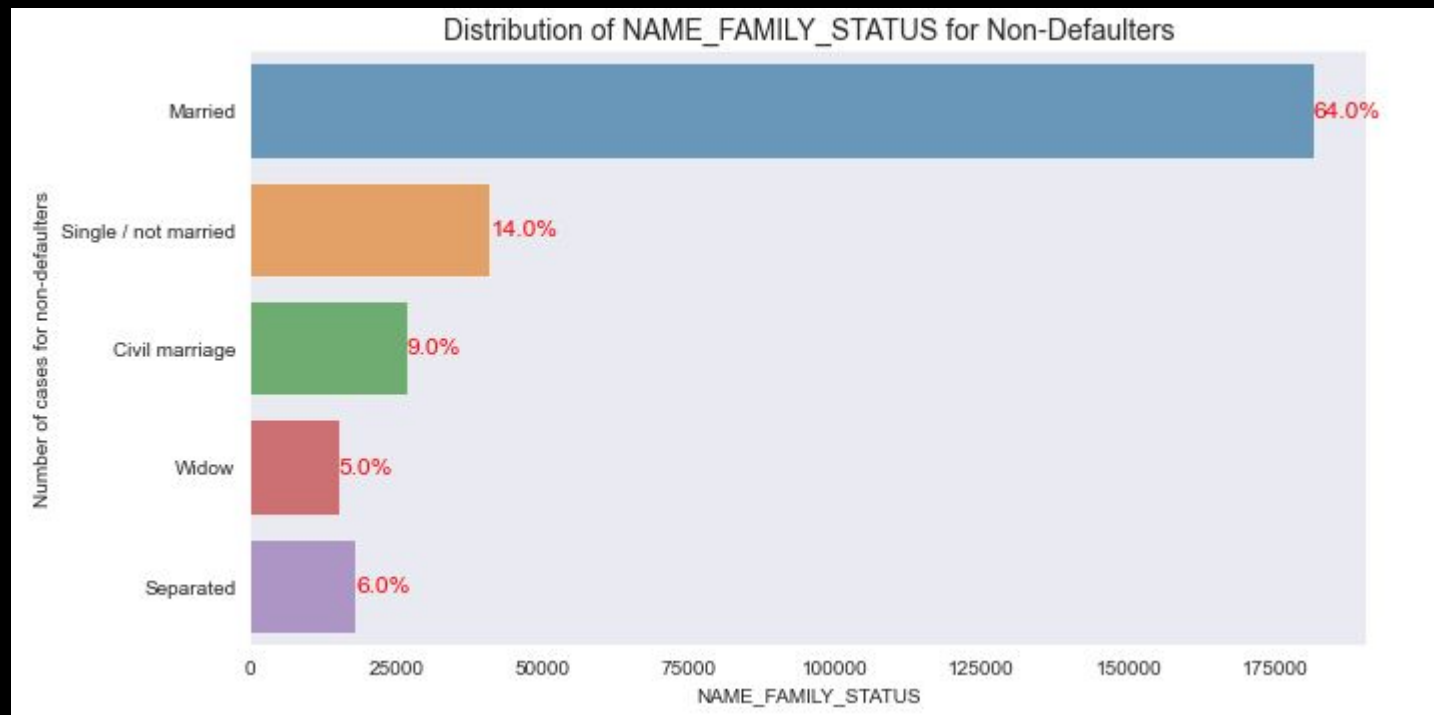# The count of applicants based on the Income type for type 0

Distribution of NAME_INCOME_TYPE for Non-Defaulters

The count of applicants based on the Income type for type 1

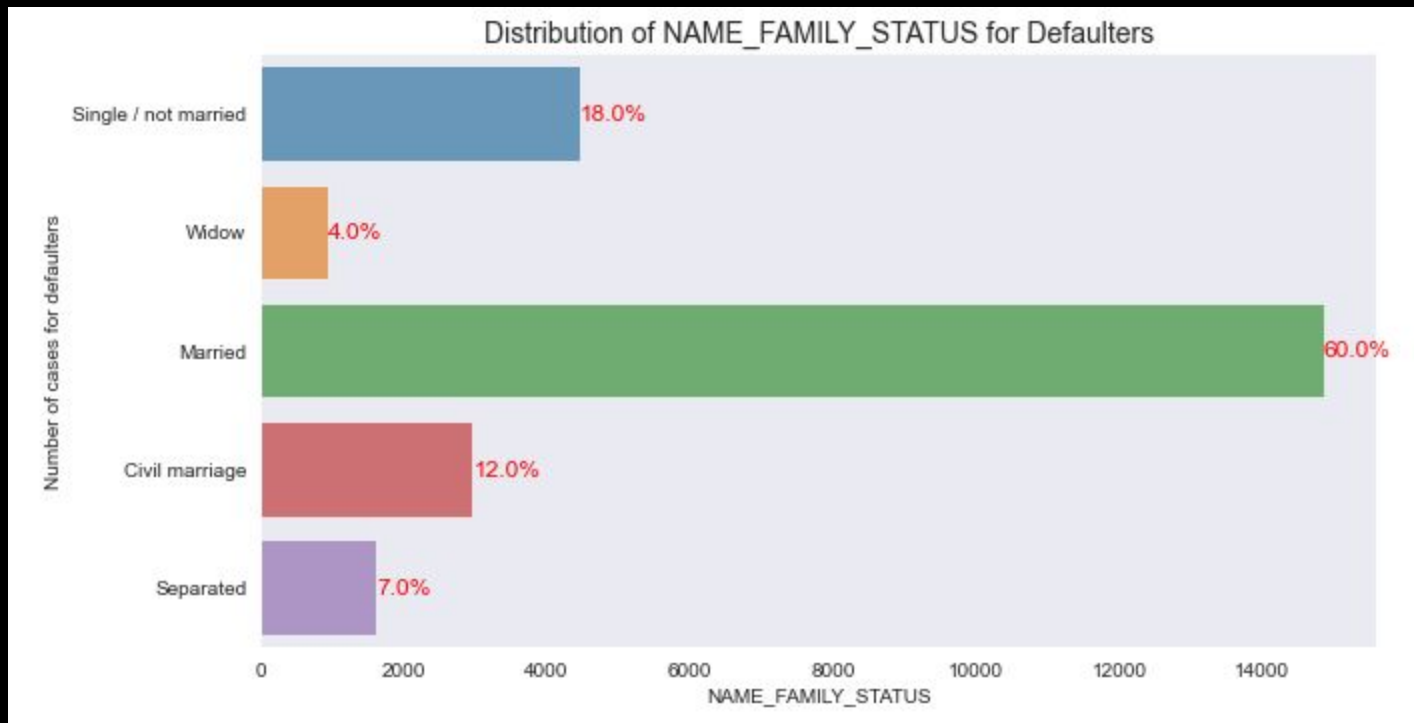Distribution of NAME_INCOME_TYPE for Defaulters
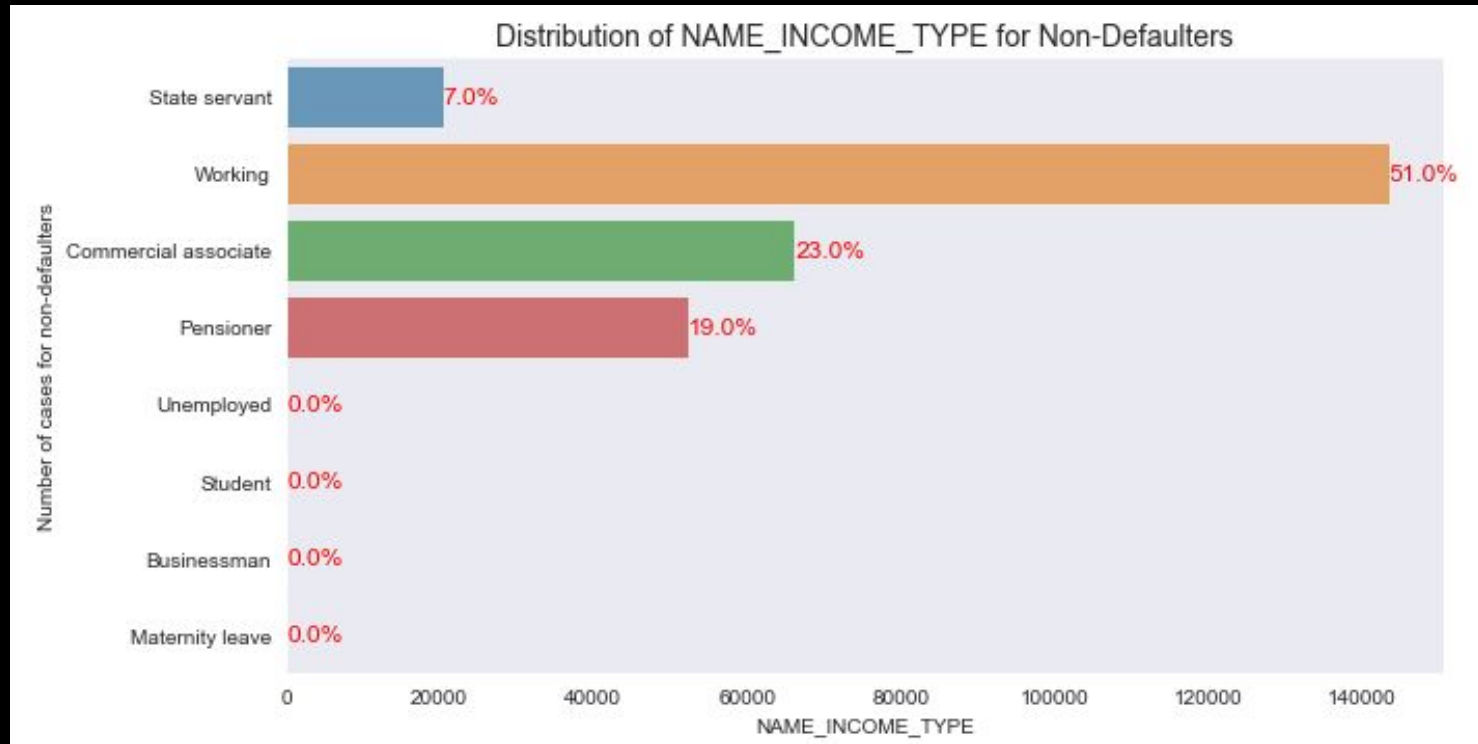
Working people are more in numbers who are defaulters. • Also found that, client with Secondary/ secondary special education are more defaulters in number.
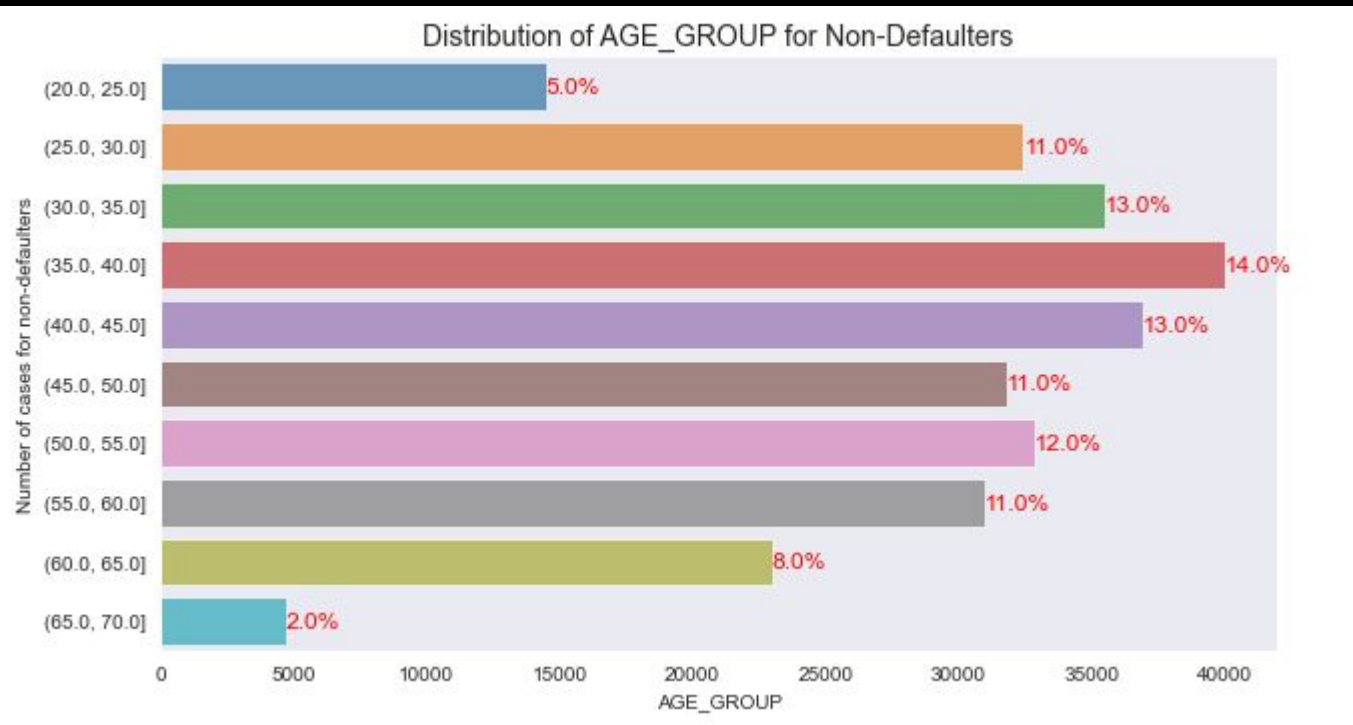
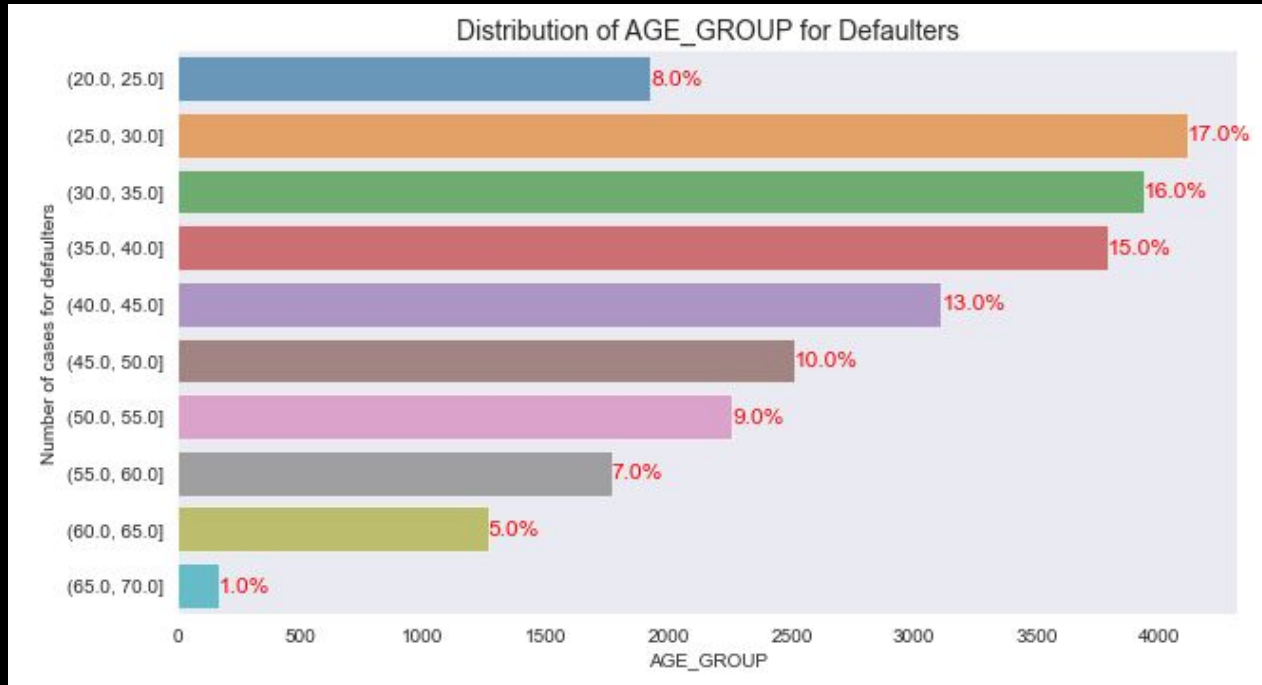# The count of applicants based on the Family Status for type 0

Distribution of NAME_FAMILY_STATUS for Non-Defaulters

The count of applicants based on the Family Status for type 1

Distribution of NAME_FAMILY_STATUS for Defaulters

# The count of applicants based on the age group for type 0

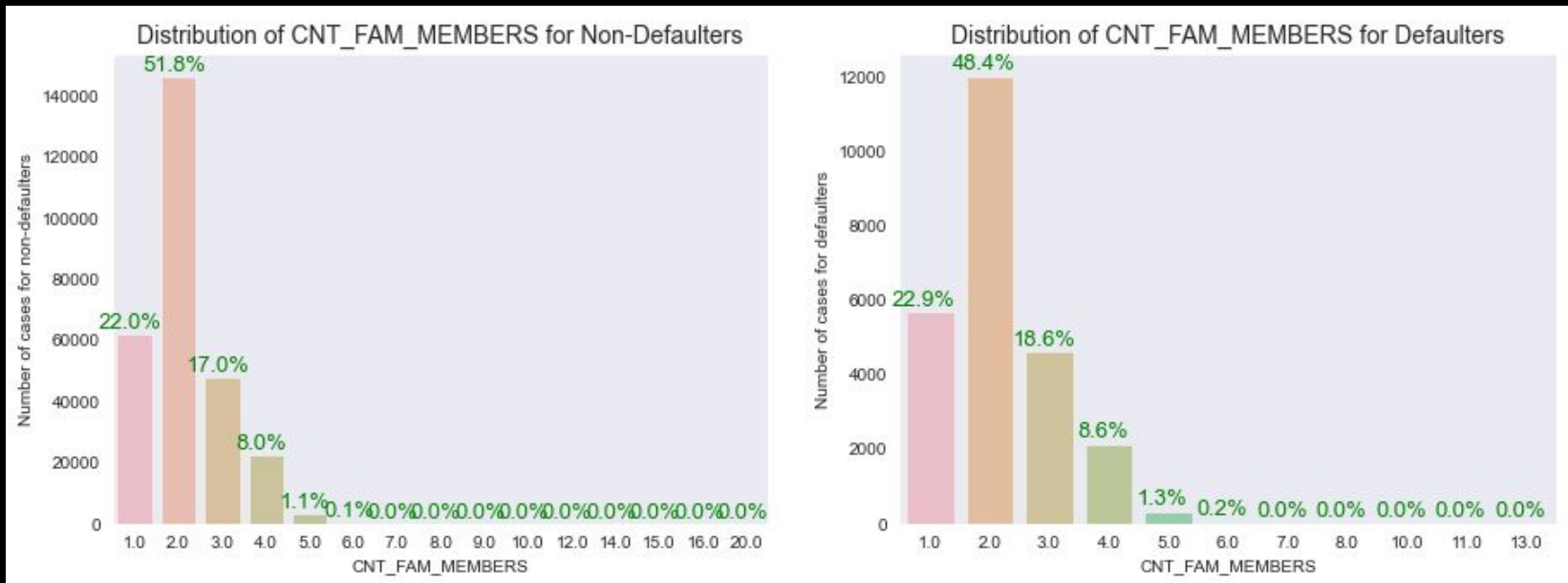# The count of applicants based on the age group for type 1

Young people are more likely to be defaulted than mid age or old people. • Low income group clients, no surprisingly, are more defaulted followed by medium income group. • Number of defaulters who are old, are same across the income group.
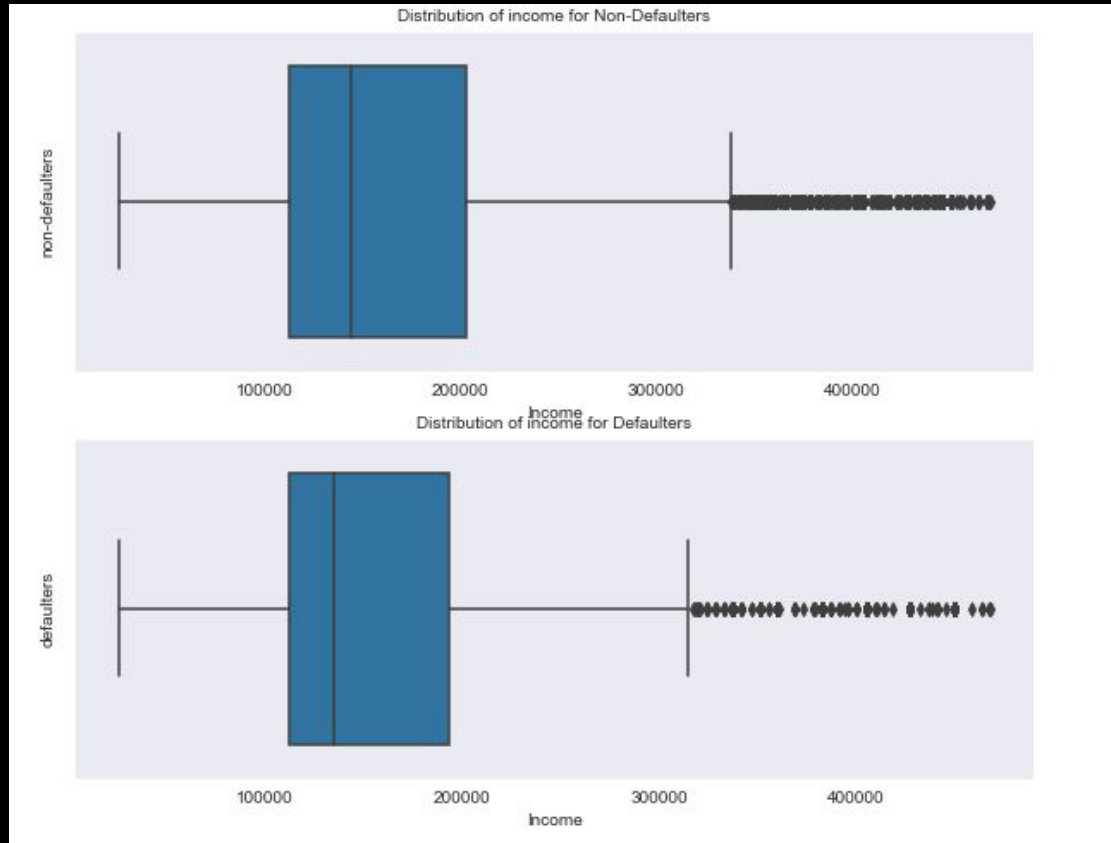
The histogram of car age for both case

Distribution of car age for Non-Defaulters

Distribution of car age for Defaulters
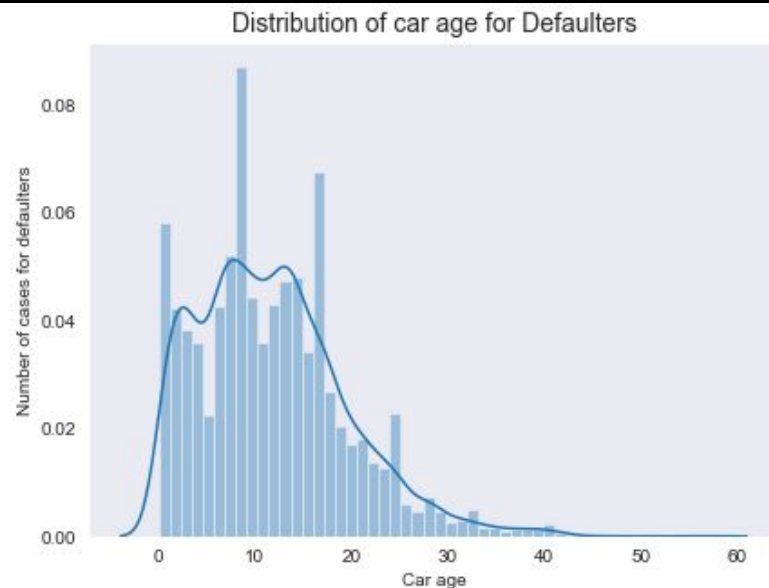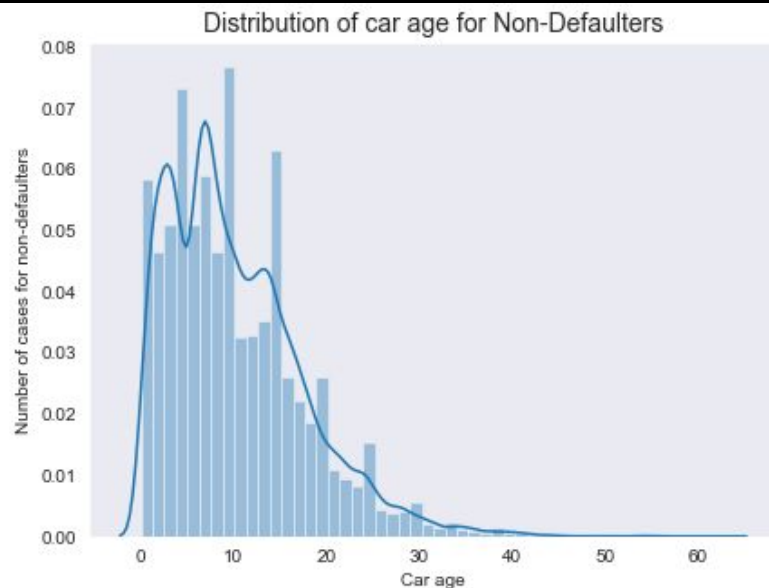
The distribution of family members for both case

Distribution of CNT_FAM_MEMBERS for Non-Defaulters

Distribution of CNT_FAM_MEMBERS for Defaulters

The distribution of Income for defaulters and non-defaulters
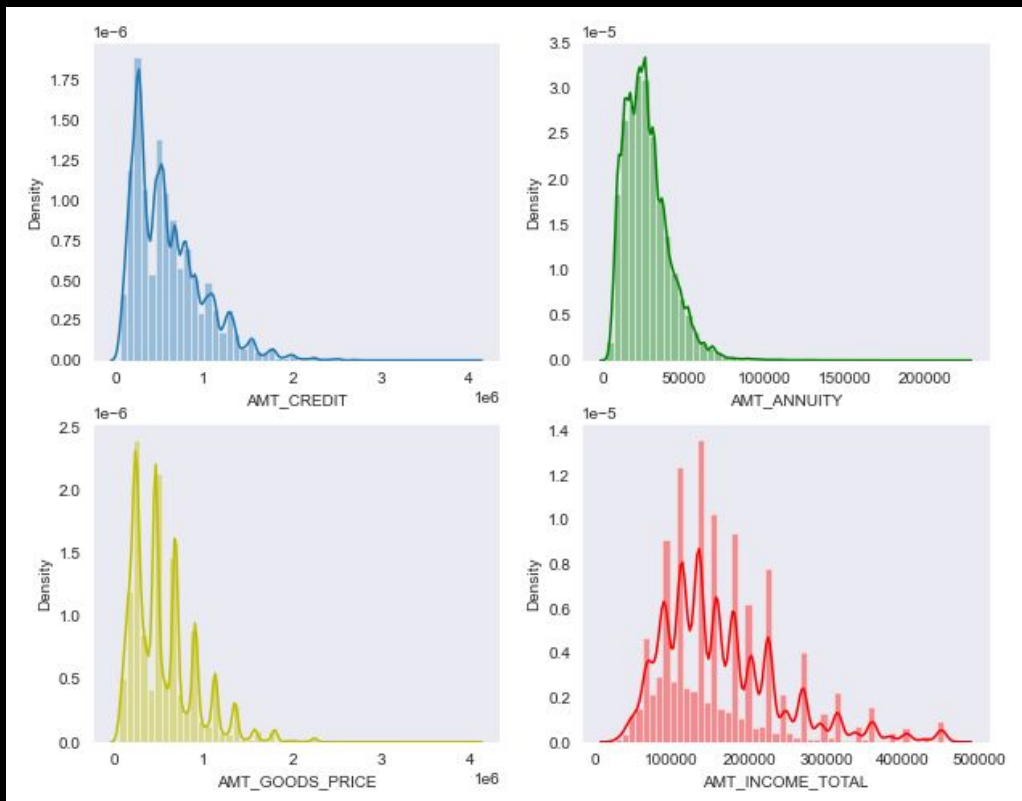
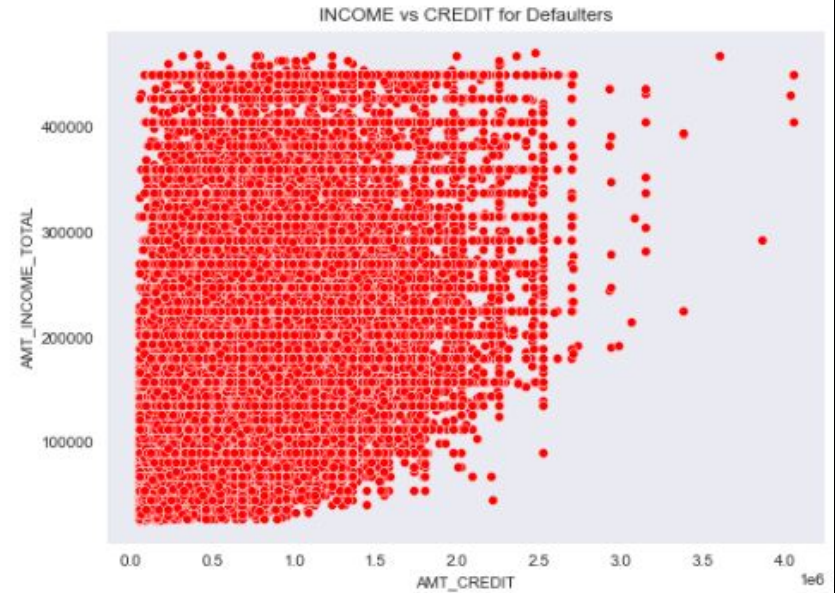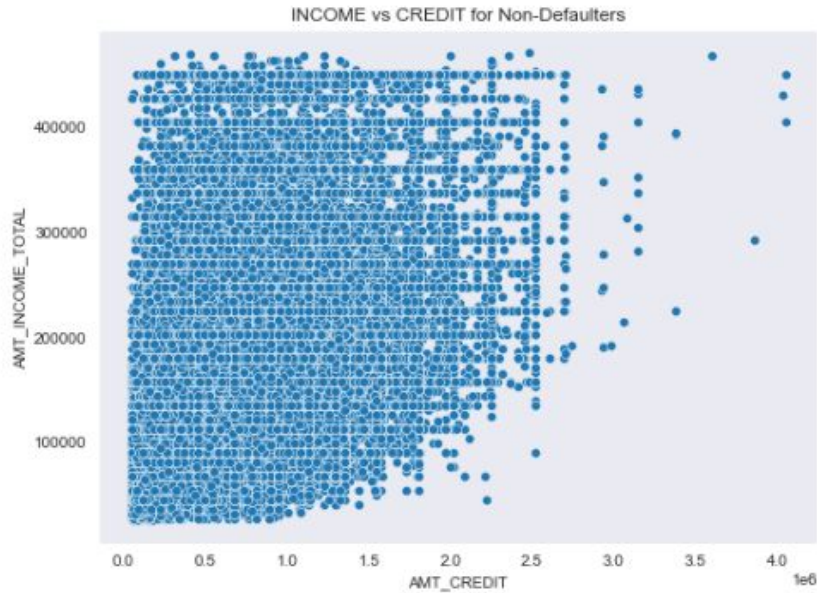# The distribution of income range bins
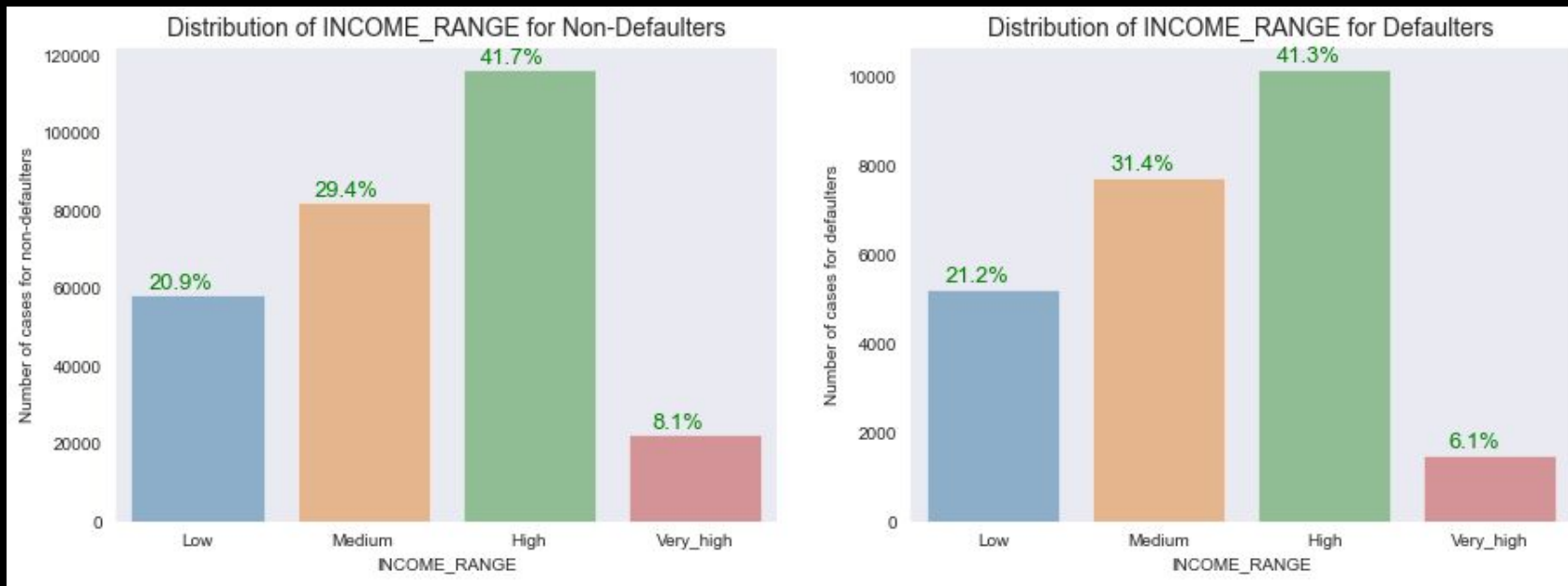
Medium amount of loan credited people are more defaulted. · High loan amount people are less defaulters across the income group. · Medium income group people with all buckets of credit category have high default rate.
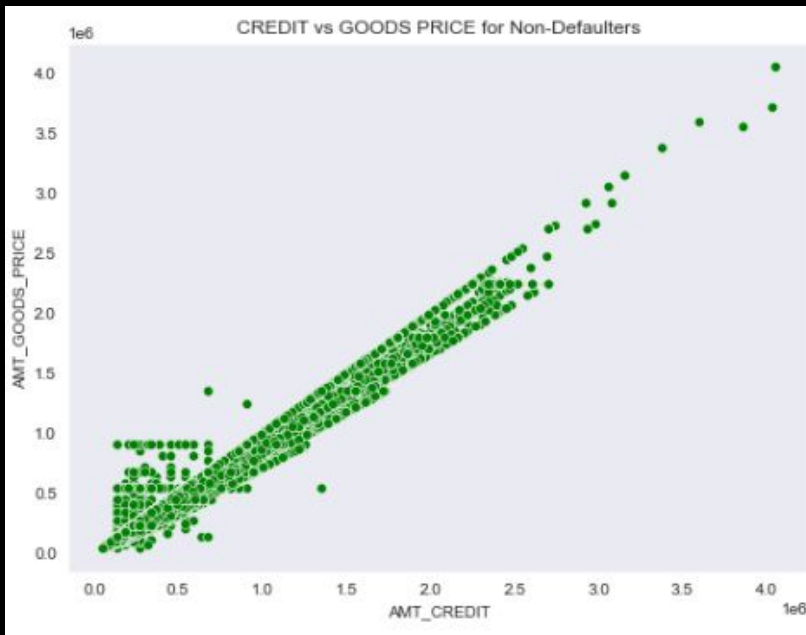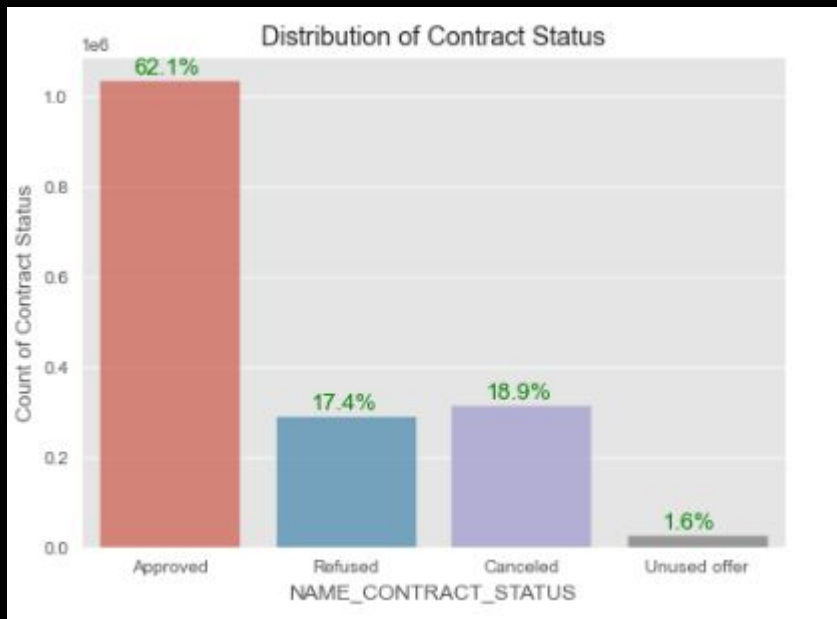
The distribution of financial attributes
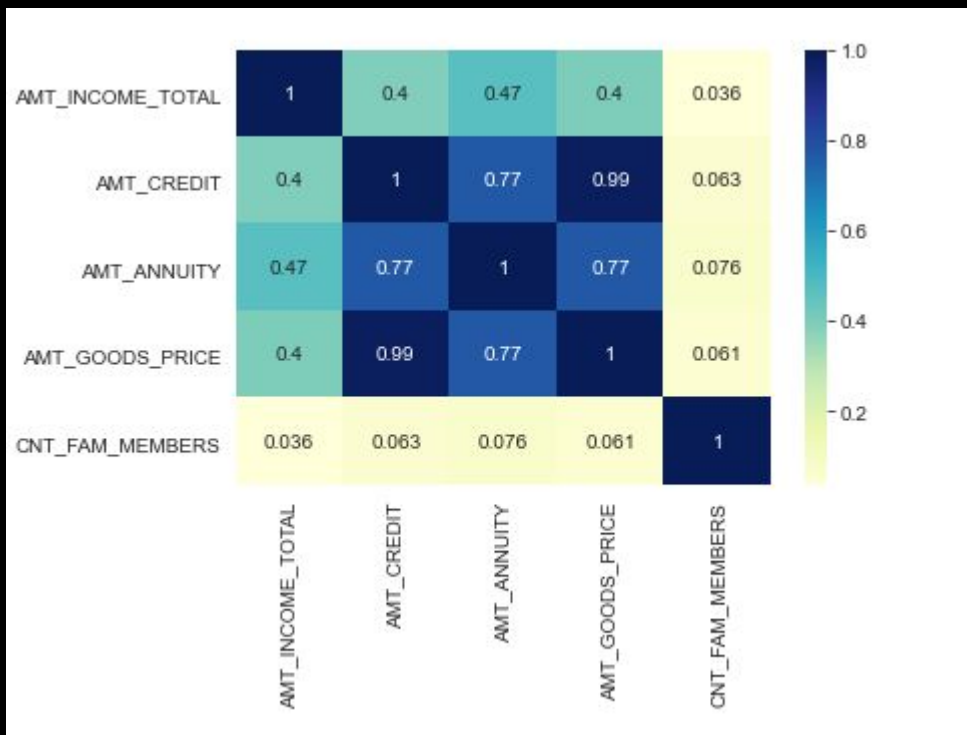
The relationship plots for financial attributes
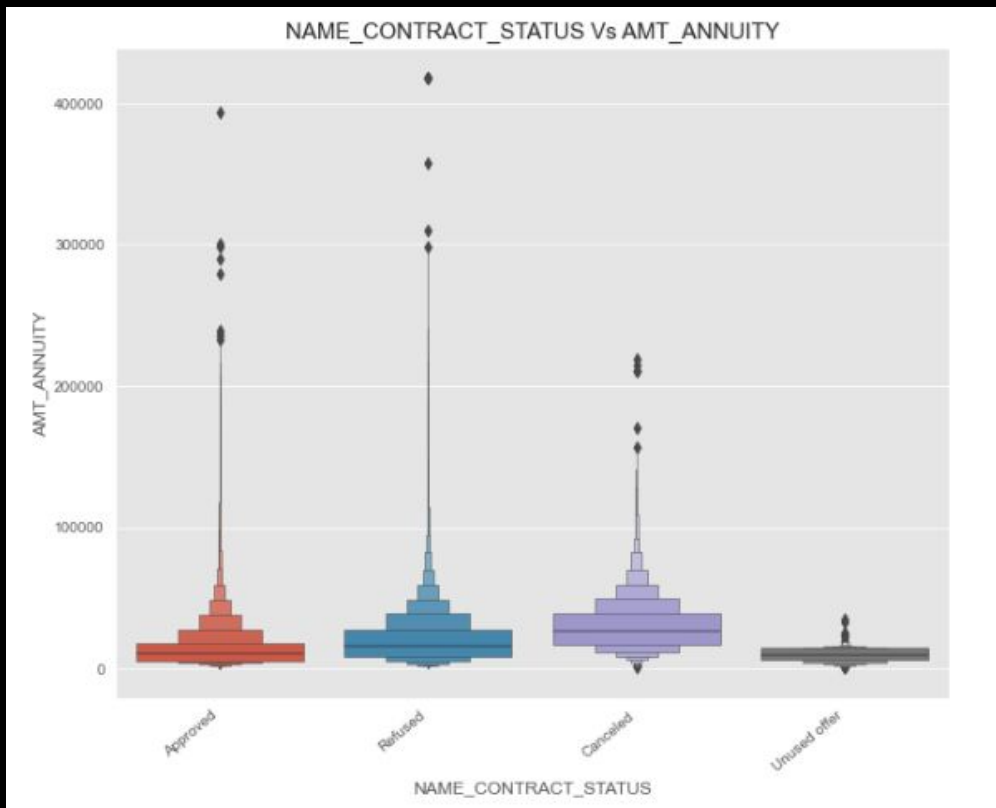
The relationship plots for financial attributes

The count of Contract status after merging the data
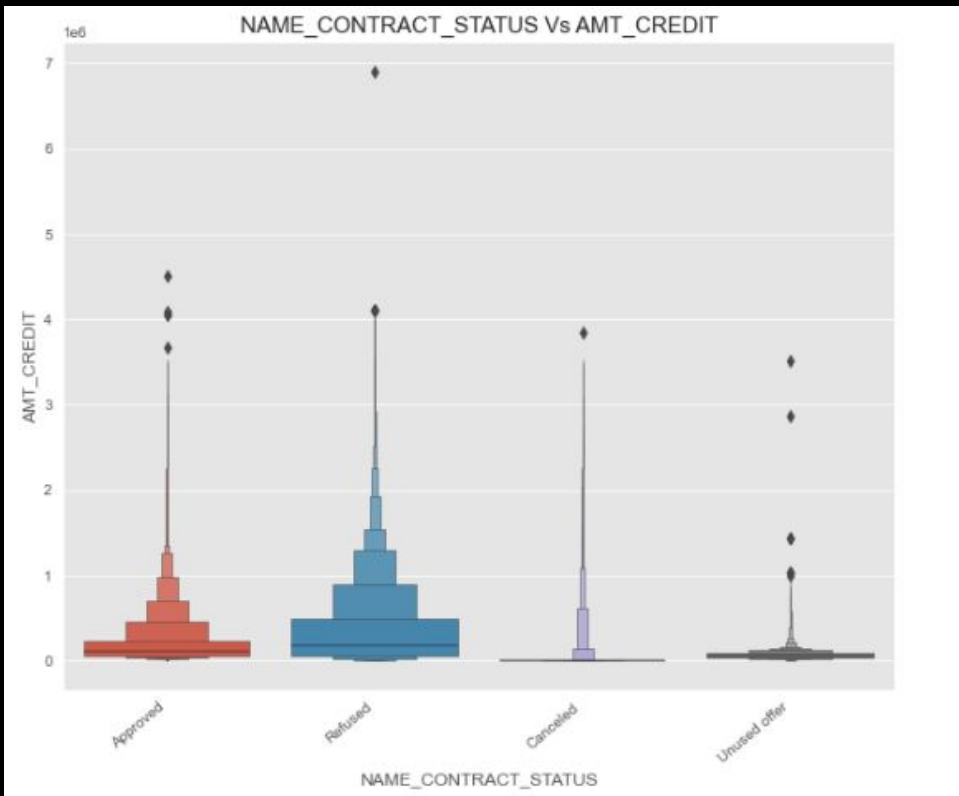

Distribution of Contract Status

The correlation heatmap of numerical attributes after merging application and previous application data
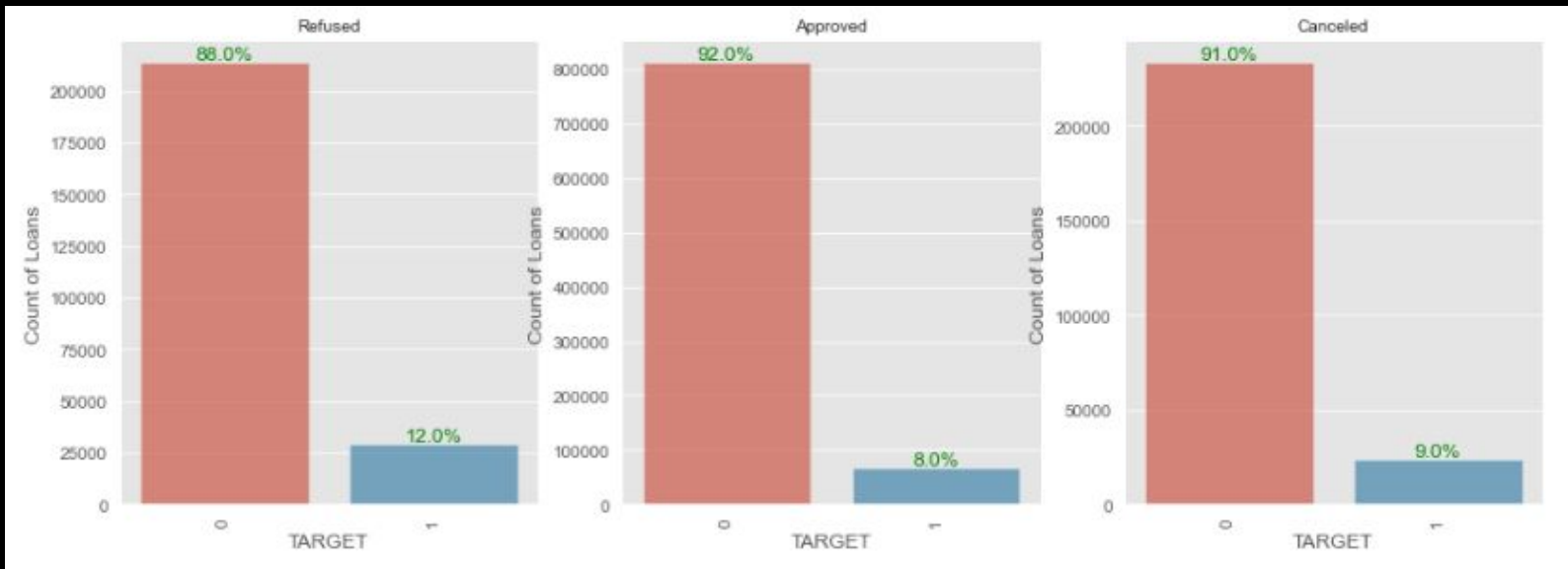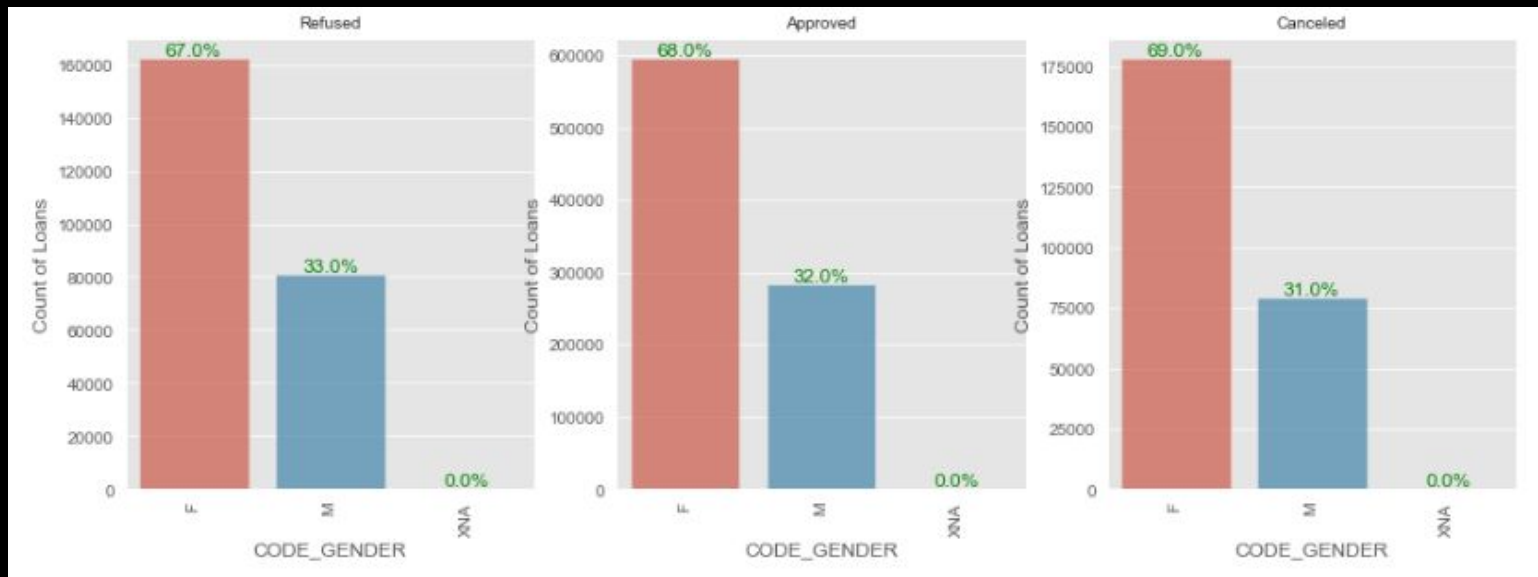
The distribution of amount of annuity by contract status

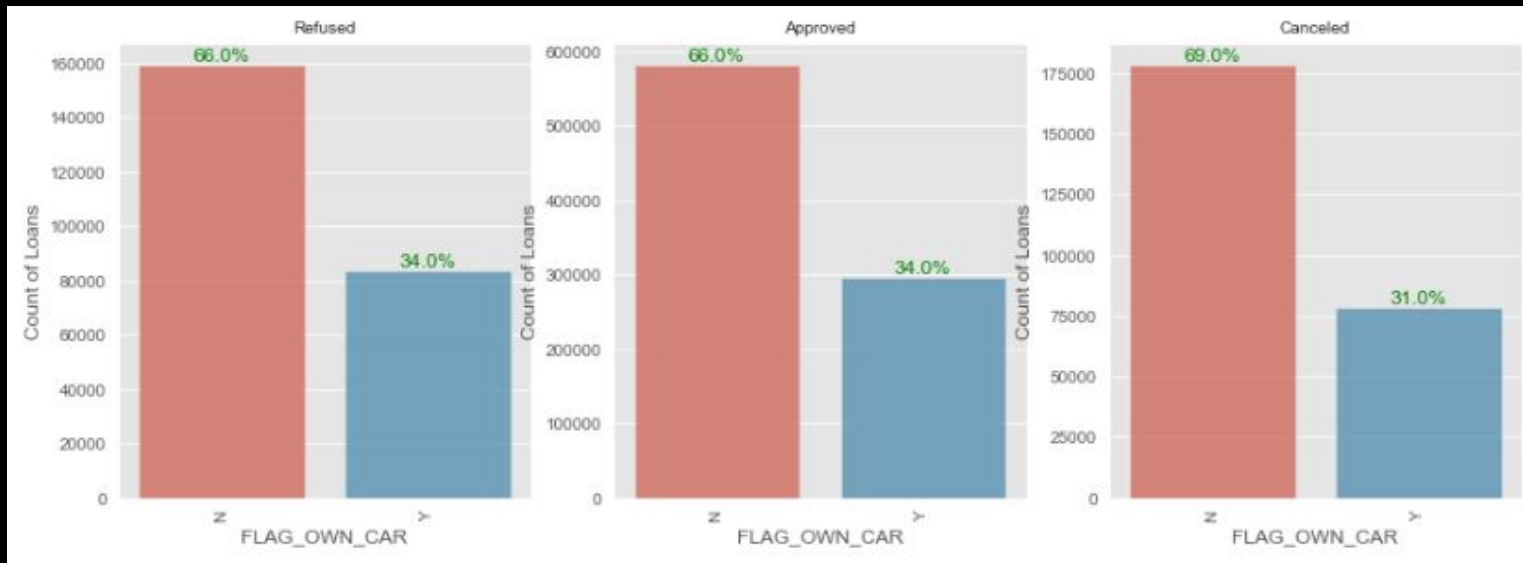# The distribution of amount credit by contract status

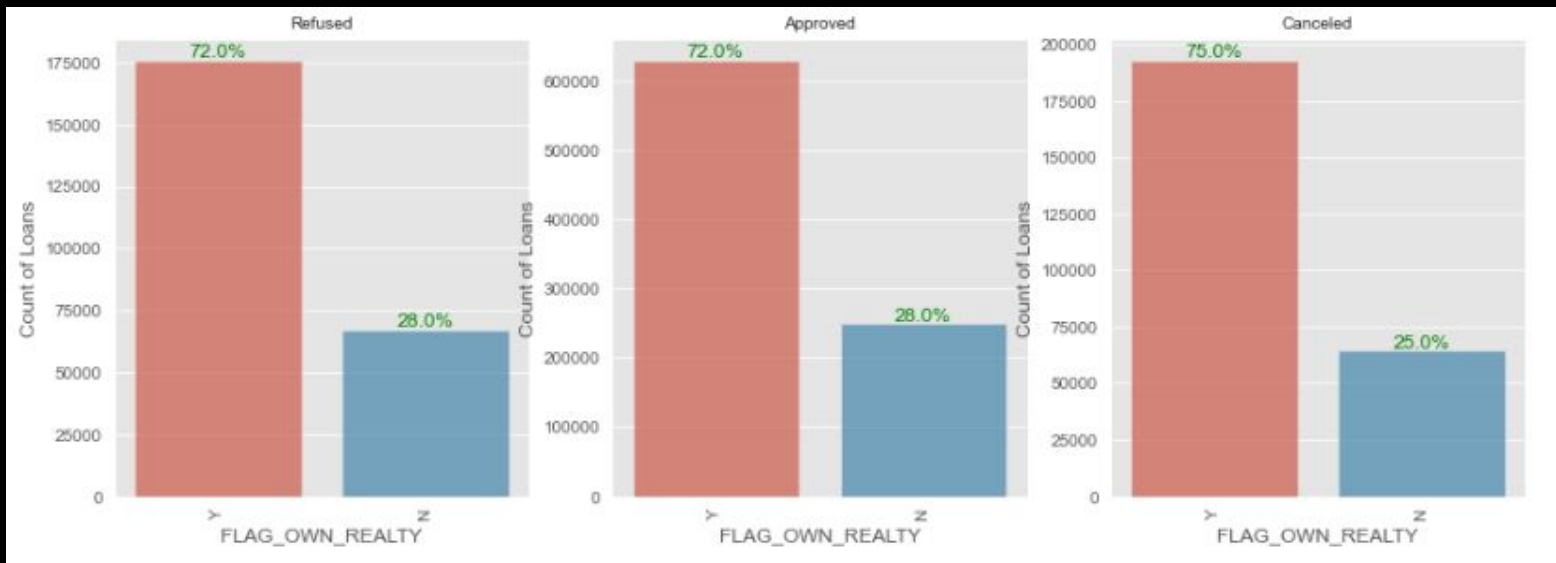# The distribution of defaulter and Non-Defaulter by contract status

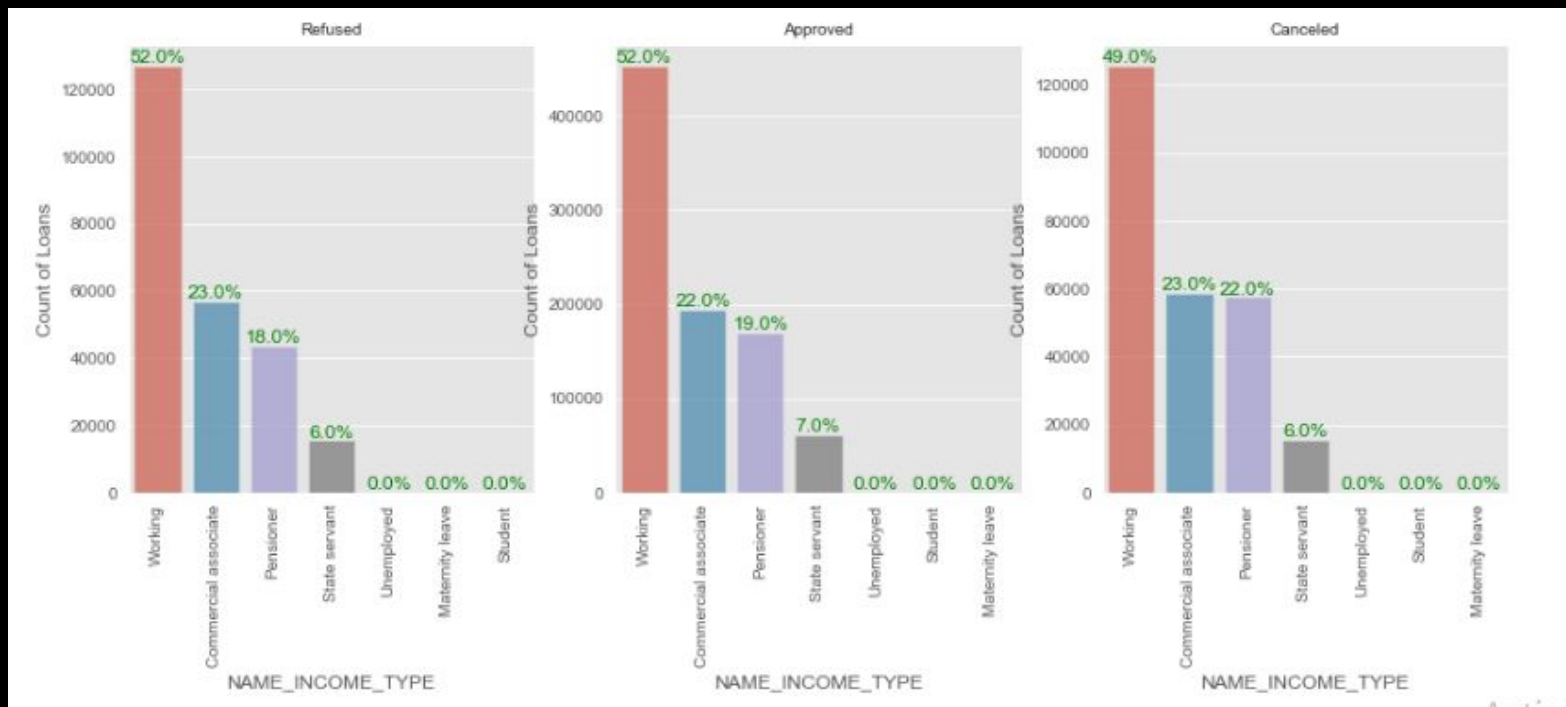# The distribution of gender by contract status

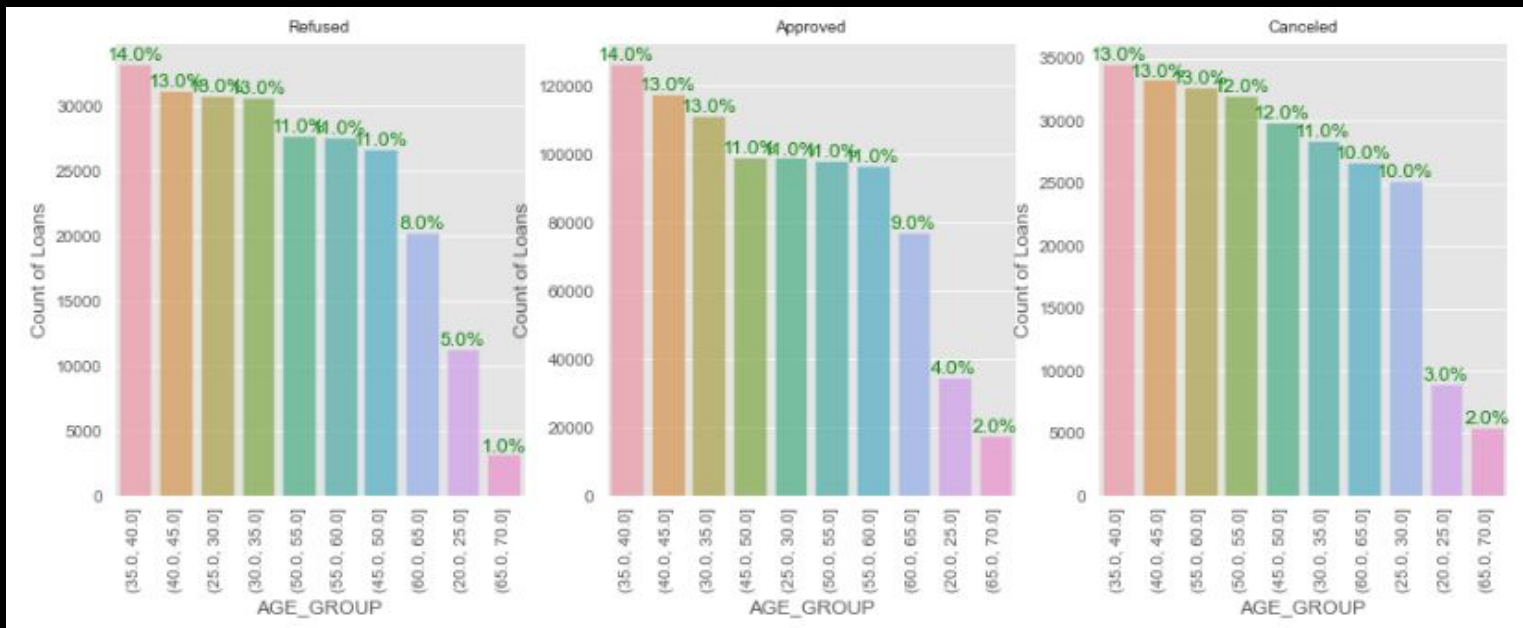# The distribution of applicants who has a car by contract status

The distribution of applicants who has a realty by contract status
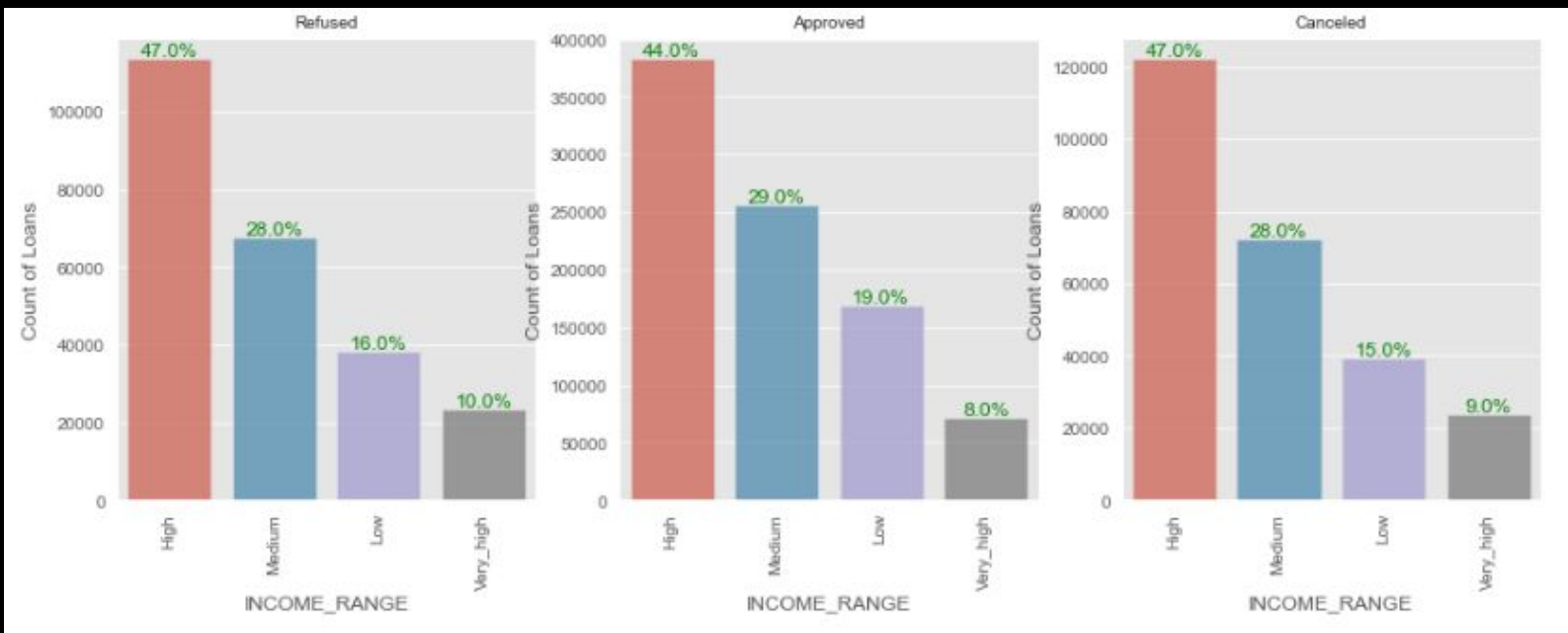
The distribution of income type by contract status

# The distribution of age group by contract status

The distribution of income range by contract status

Insights

- Lower secondary male clients has high default rate. • Young male clients has high default rate followed by mid age male. • Female old client are the safe client to credit loan. • Also female clients with higher education are less likely to be defaulted

- Approved loans are more in numbers. • Repeater clients applied for more number of loans previously. • Previously more loans credited as POS. • Through country-wise channel more loans were processed followed by credit and cash offers.
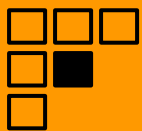
RECOMMENDATION

•Banks should focus more on contract type 'Student' ,'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.

•Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.

. State servant, Old people,high income category, female with higher education, who's previous loan approved are less likely defaulter

. Lower secondary educated clients are the most in number to be defaulted when their previous loans were cancelled or refused and male clients with civil marriage most likely defaulter