



CREDIT EDA CASE STUDY

Prepared By:
Aditi Mishra
Mayank Gupta



APPLICATION DATA



BUSINESS OBJECTIVE

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

PROBLEM STATEMENT

The loan providing companies want to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. Two types of risks that 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.

UNDERSTANDING THE DATA

Applicant Demographics

AMT_INCOME_TOTAL

CODE_GENDER

NAME_FAMILY_STATUS

NAME_EDUCATION_TYPE

FLAG_OWN_CAR

ORGANIZATION_TYPE

FLAG_OWN_REALTY

NAME_INCOME_TYPE

OCCUPATION_TYPE

AGE_GROUP

Loan Related Information

AMT_CREDIT

AMT_ANNUITY

AMT_GOODS_PRICE

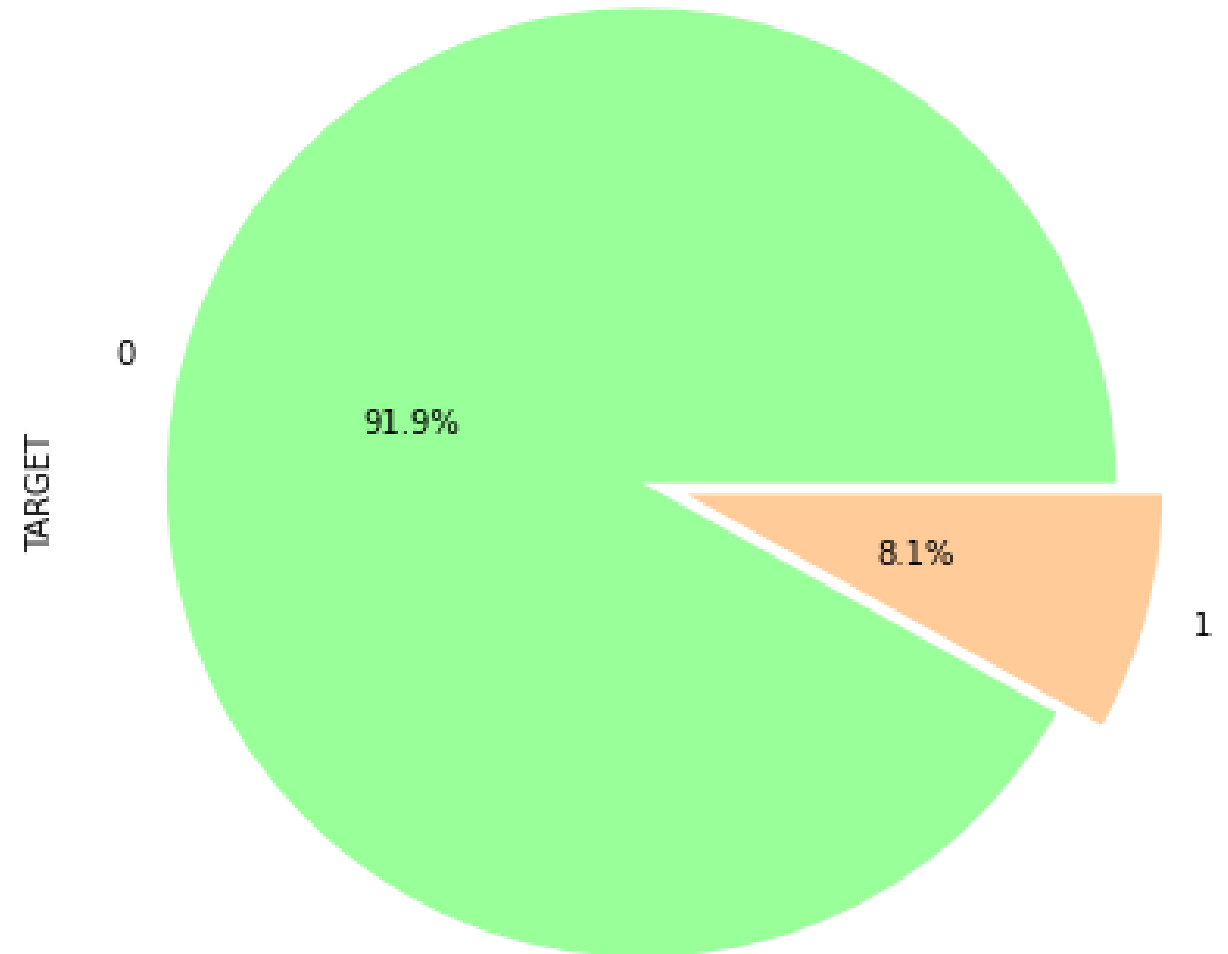
NAME_CONTRACT_TYPE

TARGET

DATA IMBALANCE

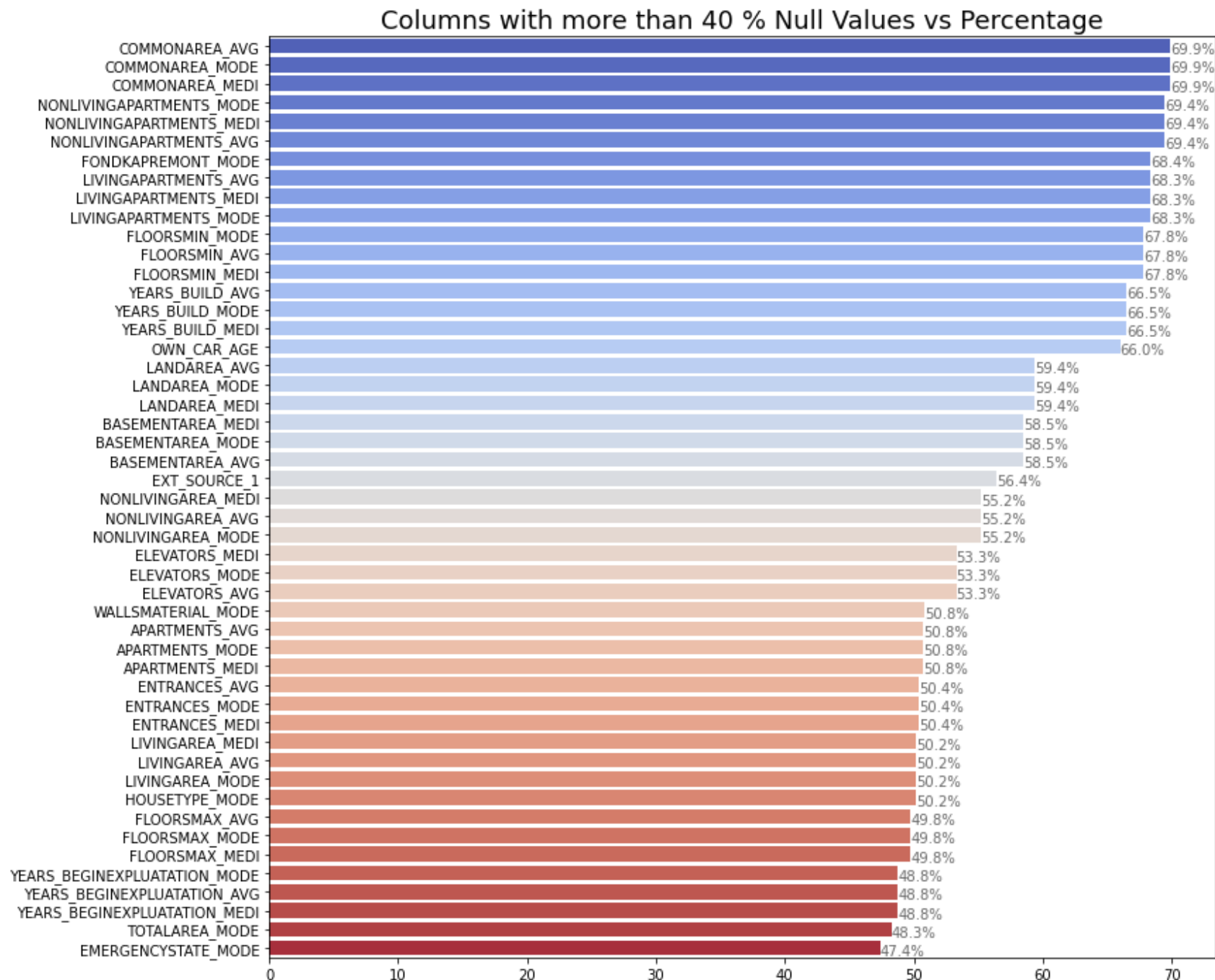
- The Percentage of people who have paid their loan is: 91.93 %
- The Percentage of people who have NOT paid their loan is: 8.07 %
- The Ratio of Data Imbalance is: 11.39

Distribution of TARGET Variable



HANDLING MISSING DATA

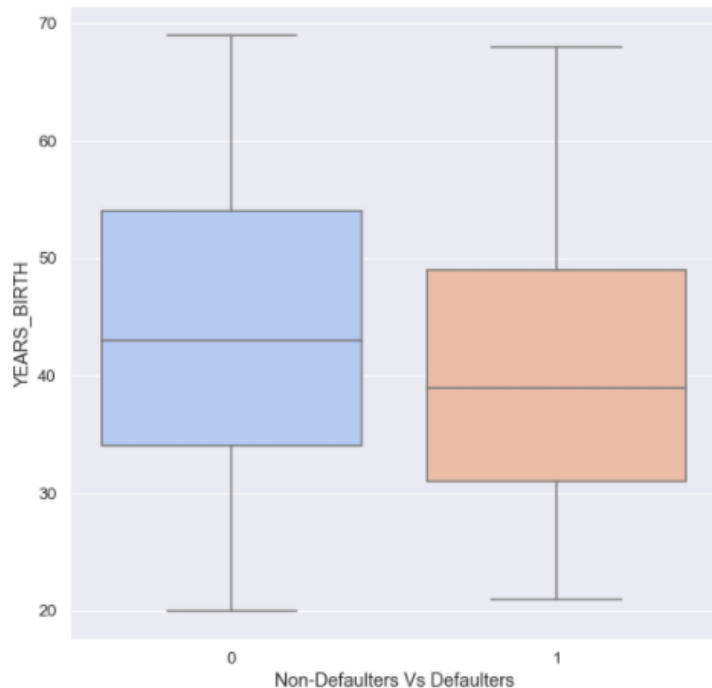
Drop the columns with more than 40% of the Missing values.



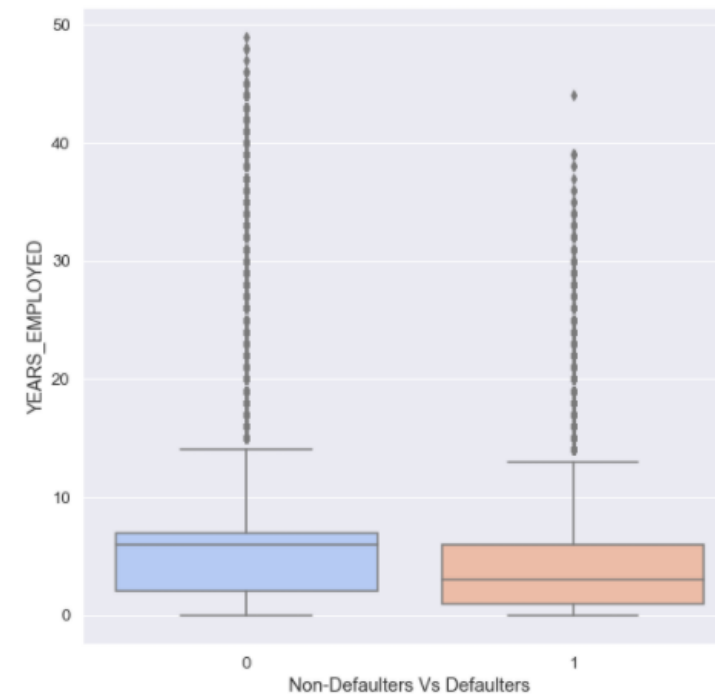
CHECKING OUTLIERS

BOX PLOT FOR YEAR BIRTH & YEARS EMPLOYED

➤ Years Birth: We can state that more than 50% defaulters have age less than 40Y whereas Non-Defaulters have an average age above 40Y.

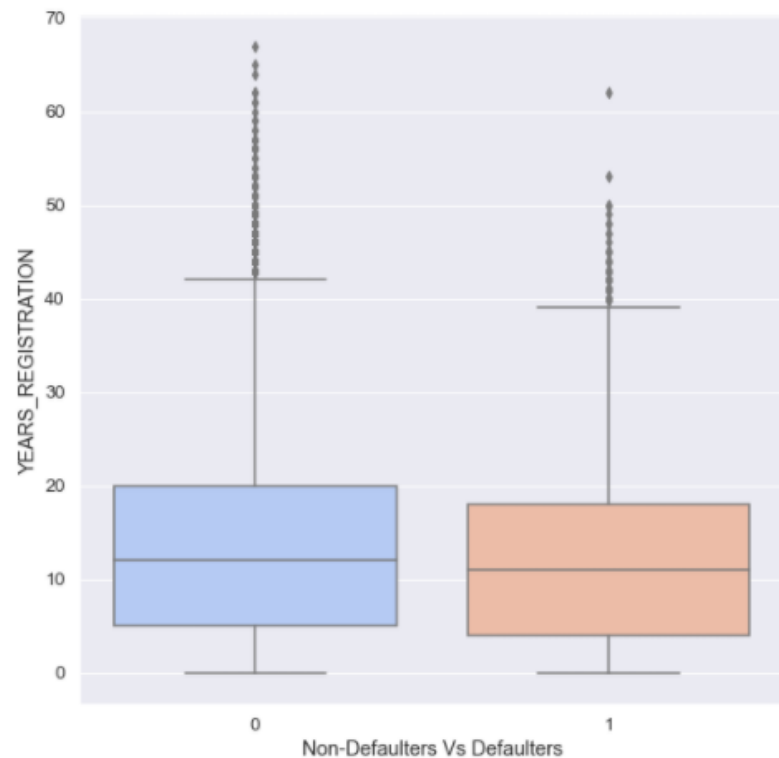


➤ Years Employed: The more the experience the less the Defaulters.

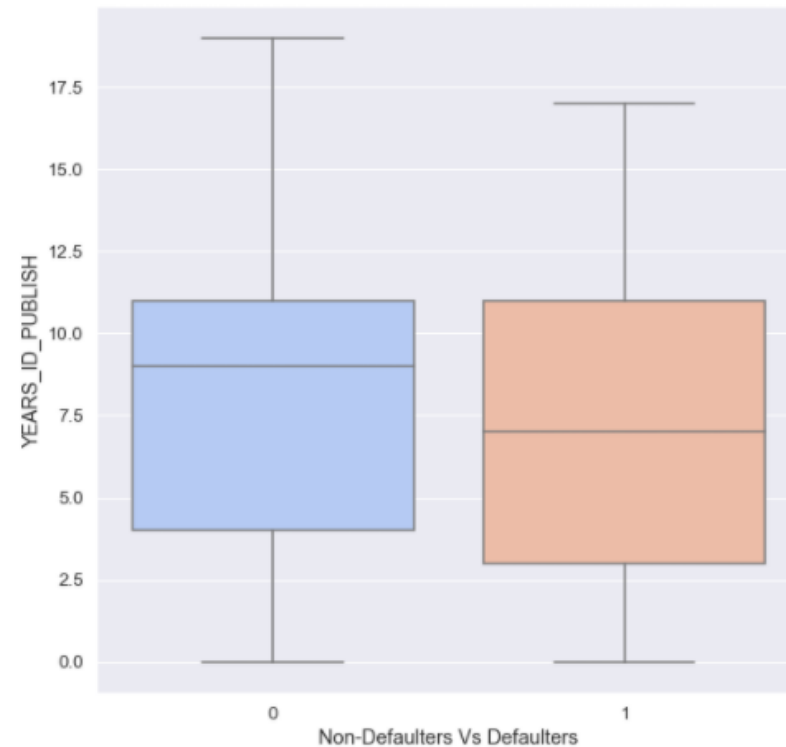


BOX PLOT FOR YEAR REGISTRATION & YEARS ID PUBLISHED

➤ Years Registration: Those who had more years of registration tends to be less Defaulters.

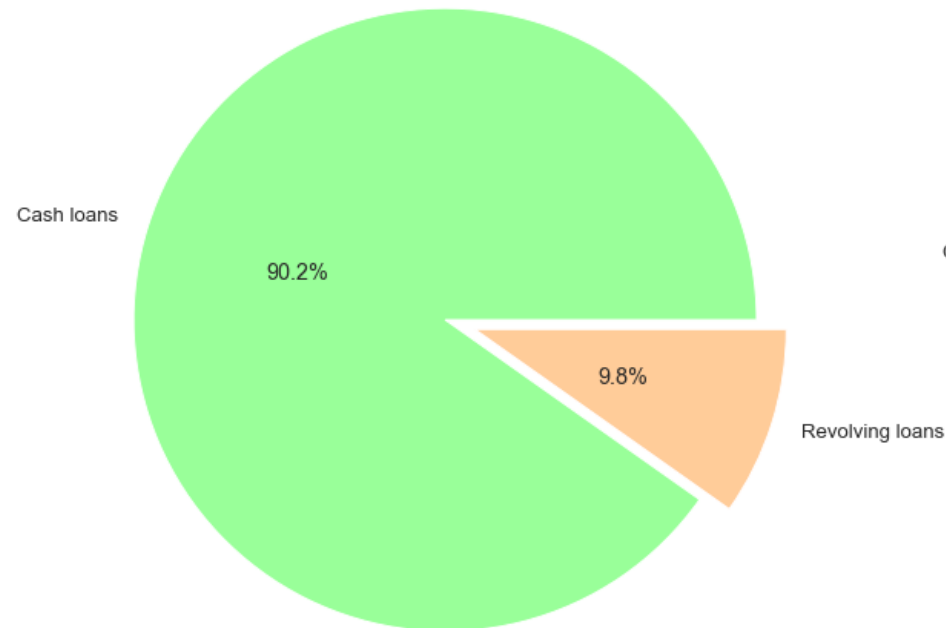


➤ Years ID Published: Defaulters seems to change the identity document more frequently.

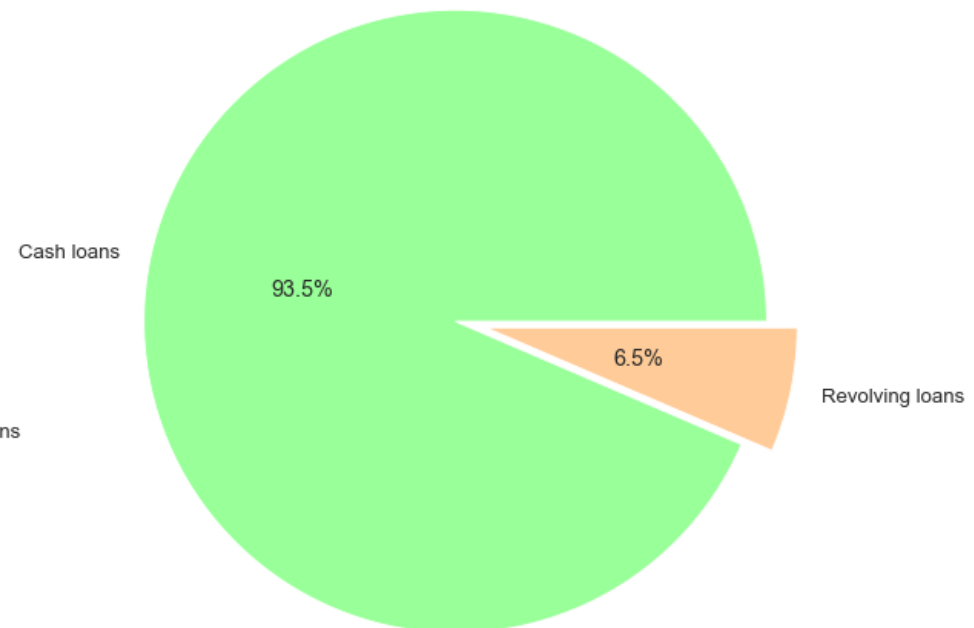


DISTRIBUTION FOR LOAN TYPE

Distribution of Defaulters vs Loan Type



Distribution of Non-Defaulters vs Loan Type



We observe that the number of Cash loans is much higher than the number of Revolving loans for both Defaulters and Non-Defaulters

UNIVARIATE ANALYSIS

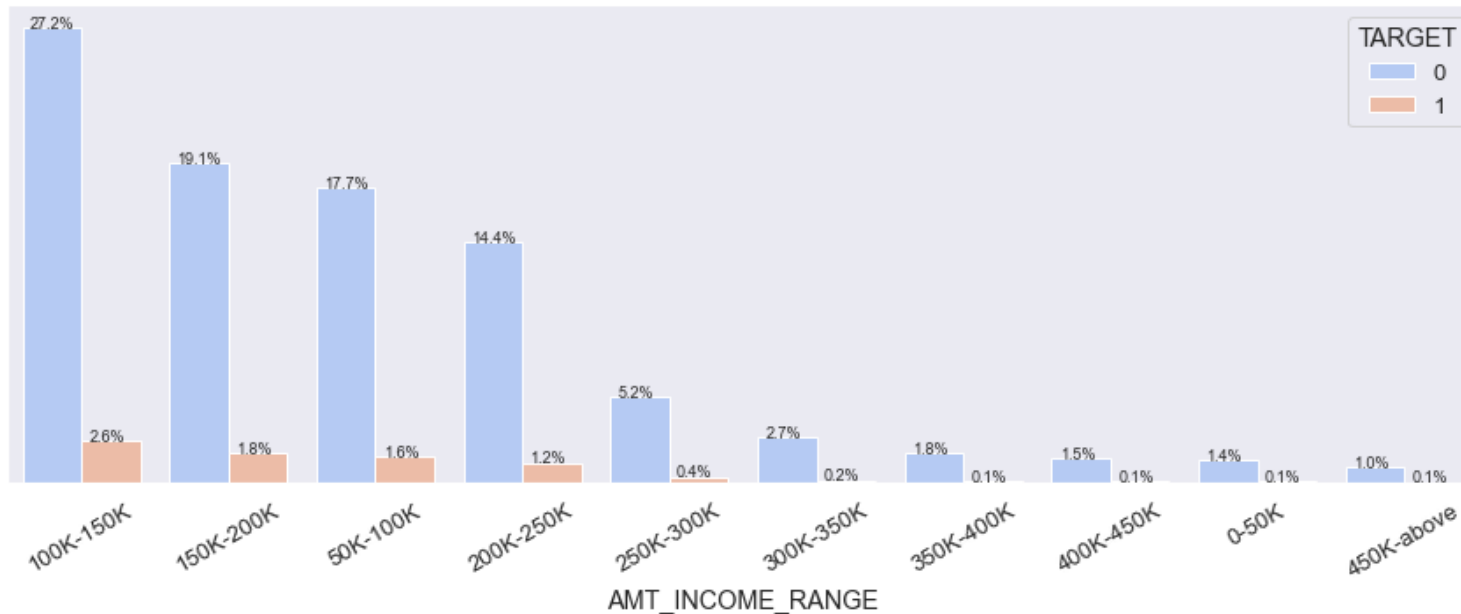
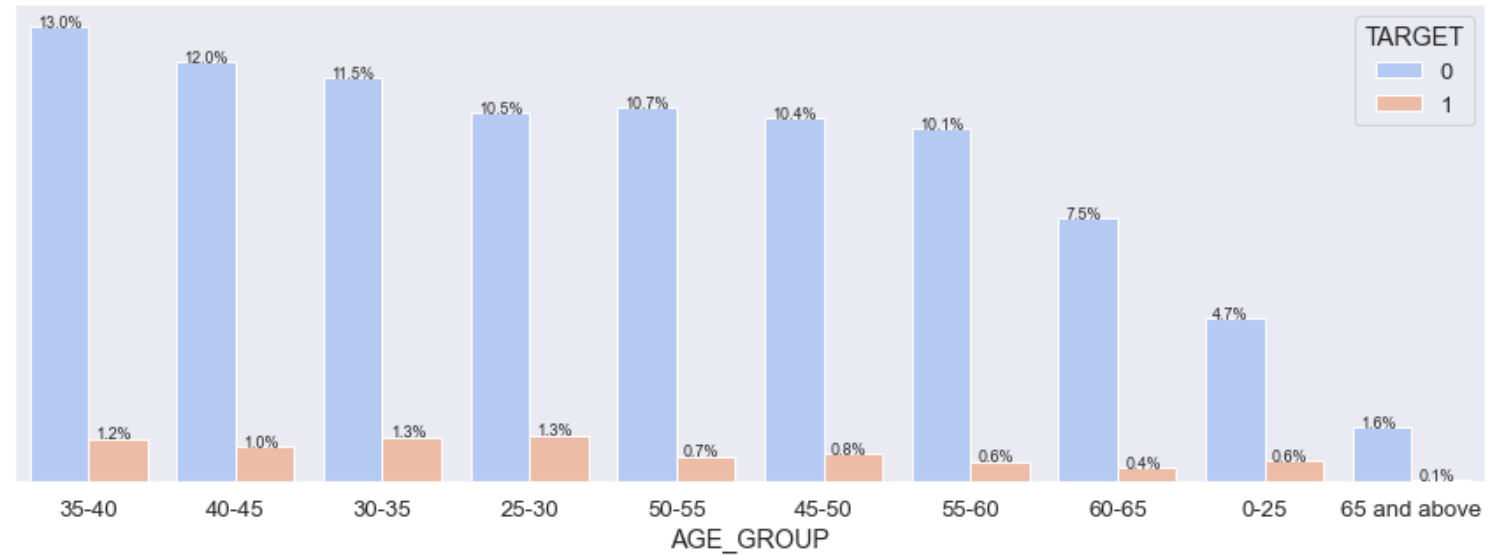
COUNT PLOT FOR INCOME RANGE & AGE GROUP

0 = Non-Defaulters

1 = Defaulters

- A major section of Defaulters belongs between 25-35 Years.

- 35-40 Years of age group seems to repay their loans most



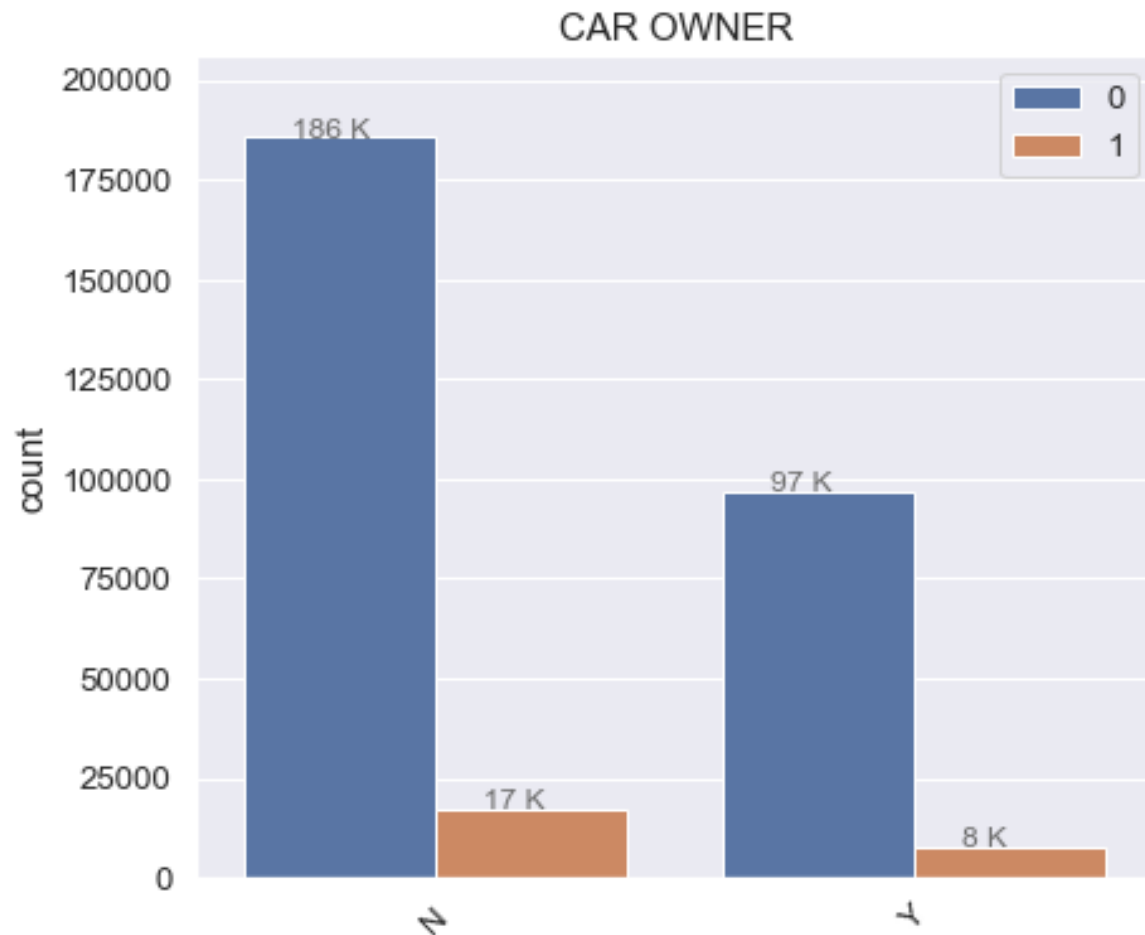
- For Income Range 100K-150K both Defaulters and Non Defaulters holds high chance.

- Clients having income more than 450K seems to repay their loan.

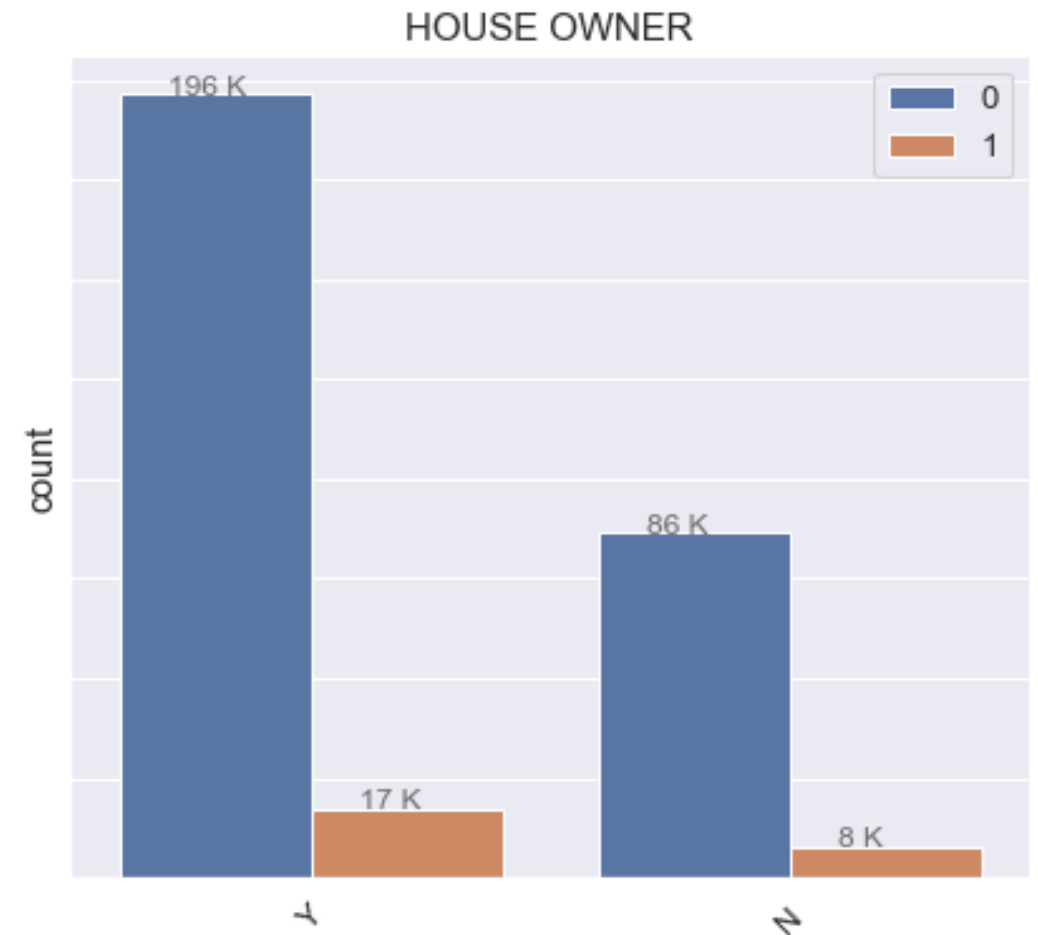
0 = Non-Defaulters

1 = Defaulters

COUNT PLOT FOR NON-DEFAULTERS & DEFAULTERS FOR CAR OWNERS & HOUSE OWNERS



-Majority of the applicants do not have car.

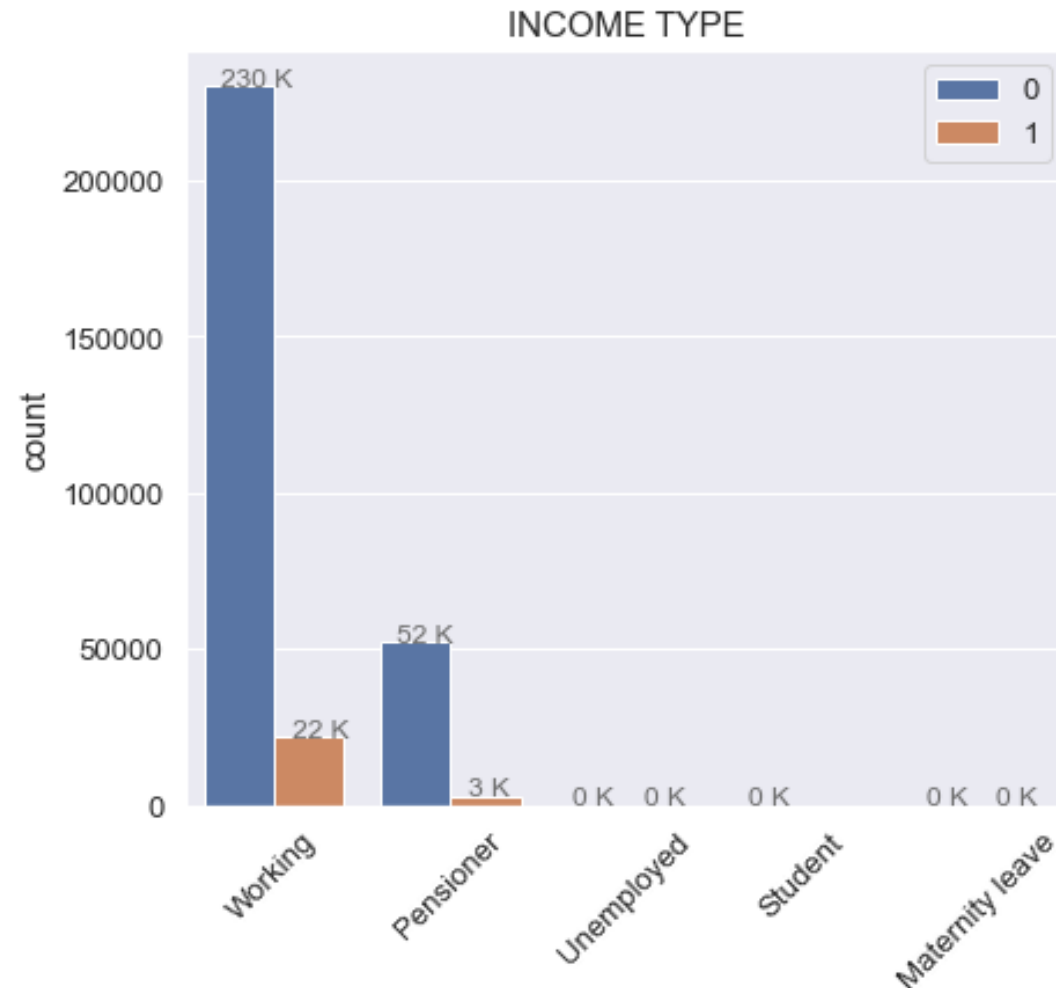


-Majority of the applicants have their own house.

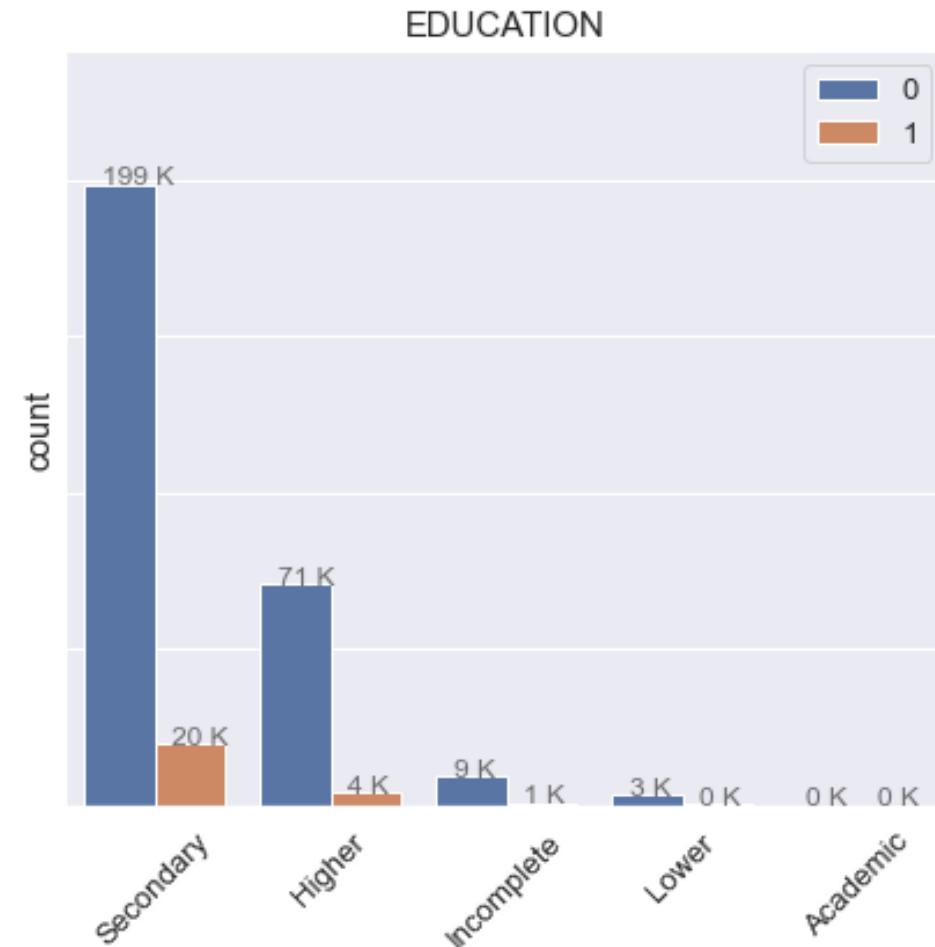
0 = Non-Defaulters

1 = Defaulters

COUNT PLOT FOR NON-DEFAULTERS & DEFAULTERS FOR INCOME TYPE & EDUCATION



-Majority of the applicants are working type.



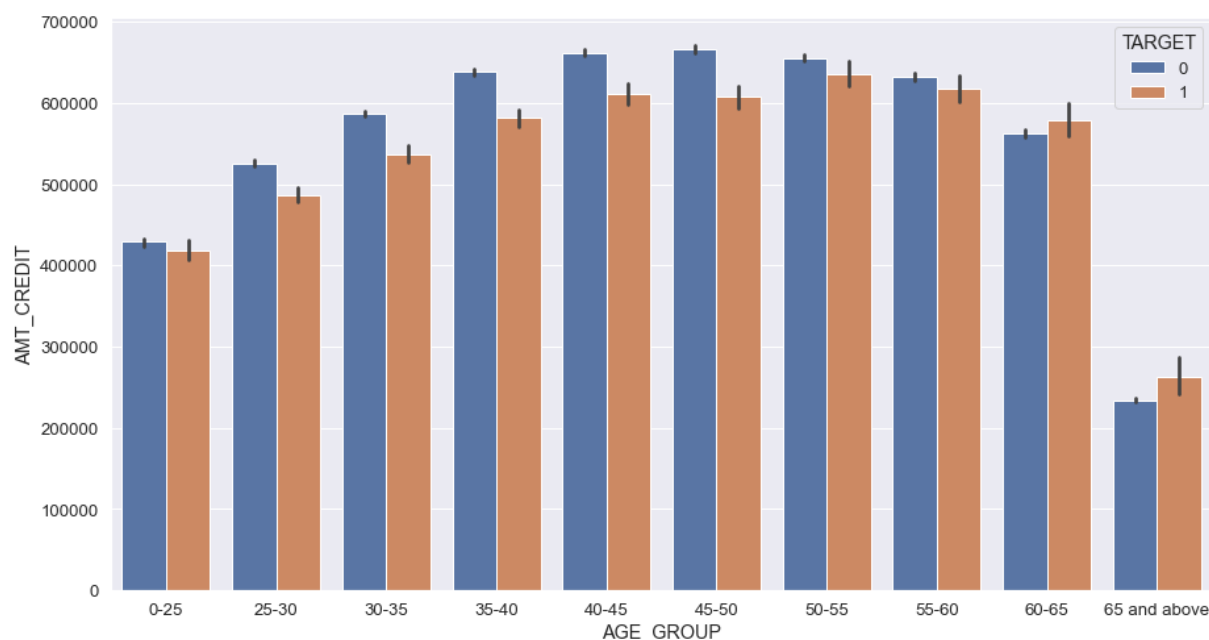
-Majority of the applicants have secondary education.

BIVARIATE ANALYSIS

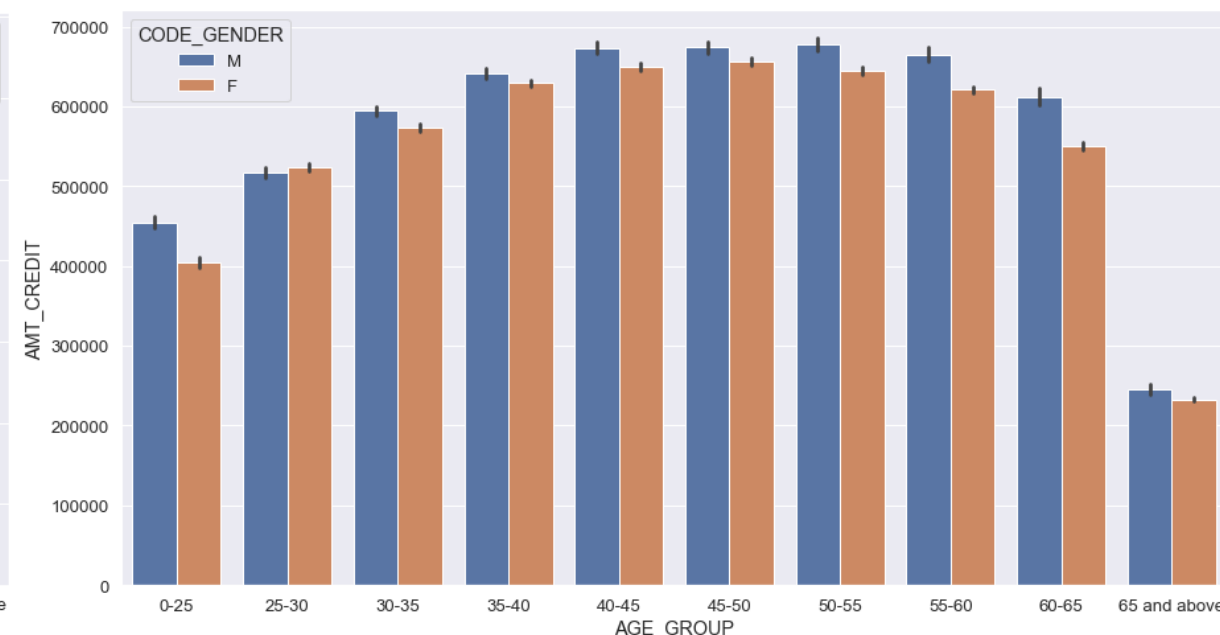
0 = Non-Defaulters

1 = Defaulters

AGE GROUP VS AMOUNT CREDIT FOR TARGET VARIABLE & GENDER

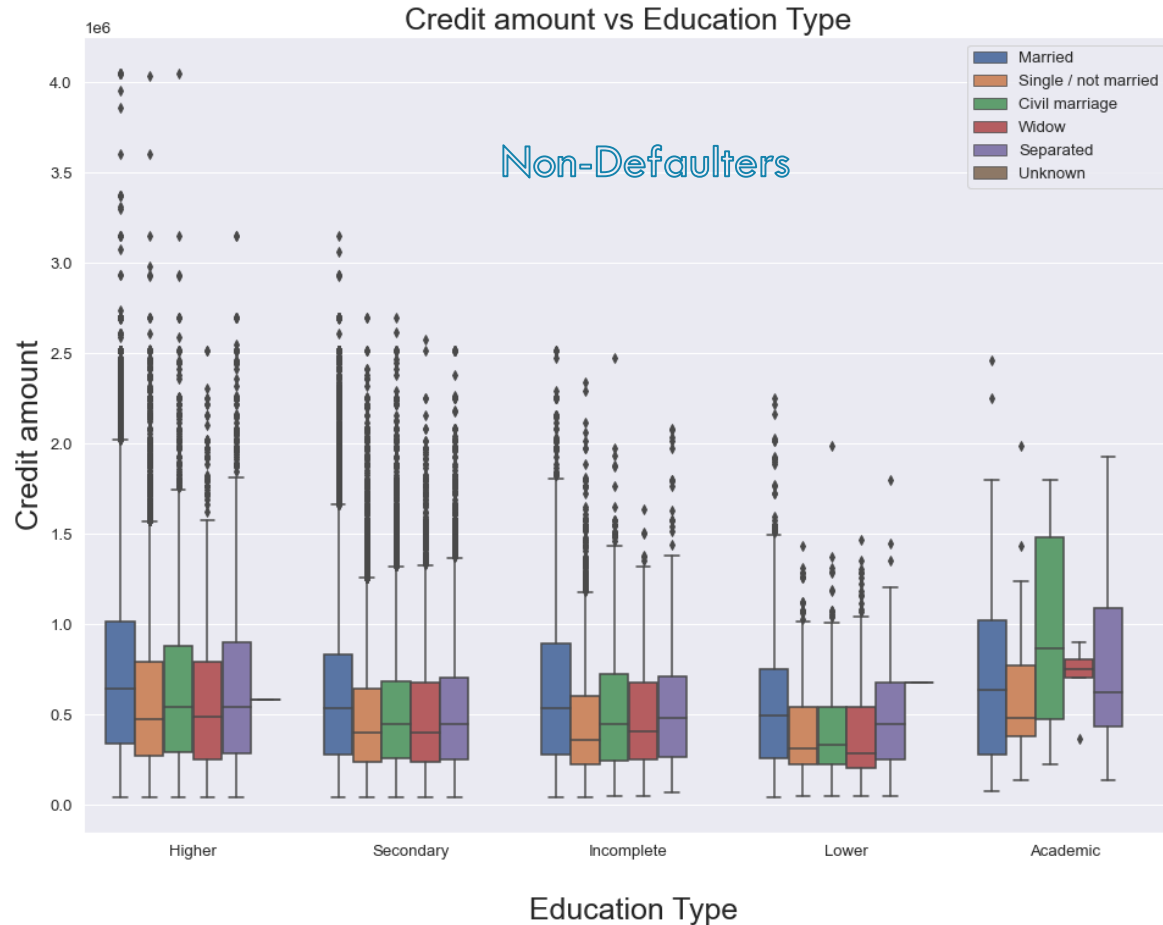


For Credit amount majority of the defaulters have age more than 60

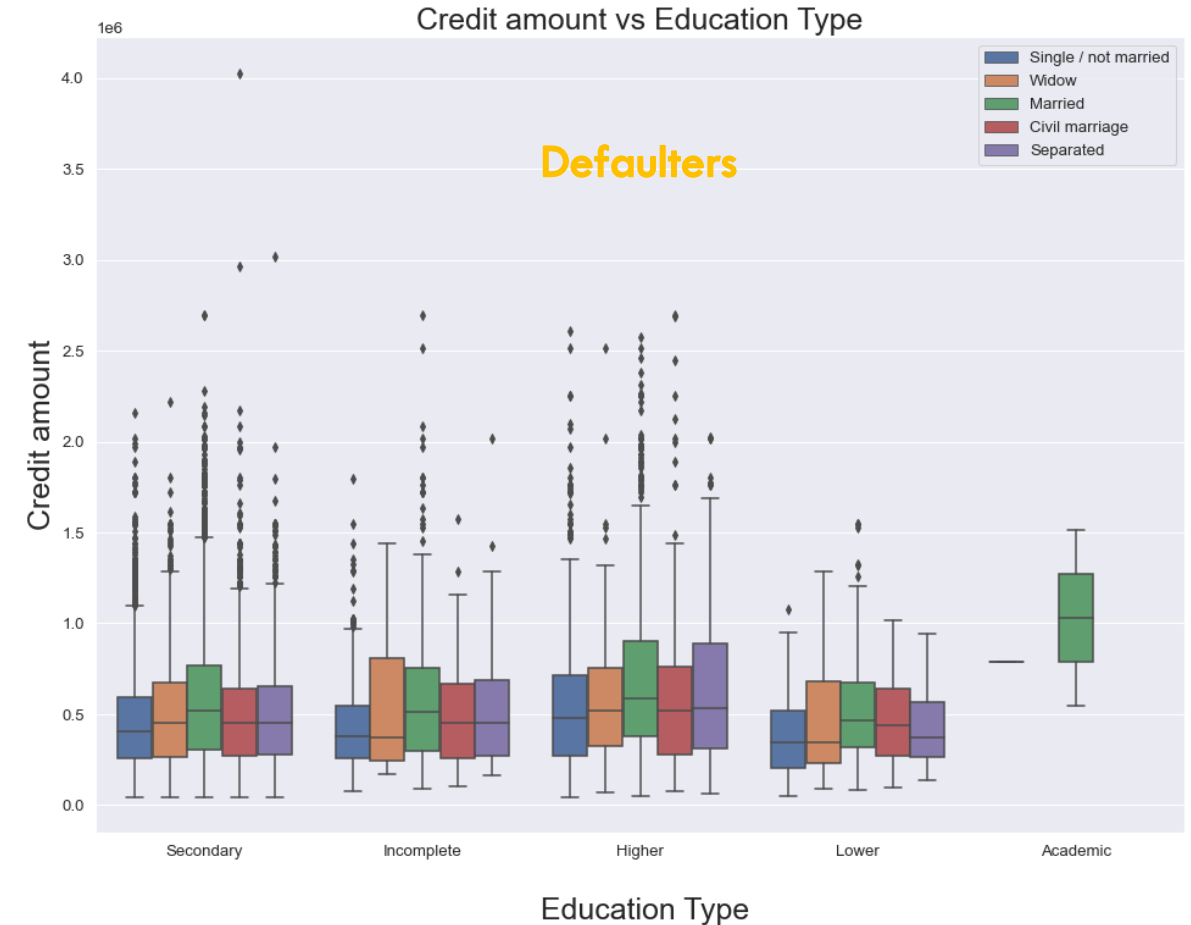


For age group 25-30 the requirement of loan is almost similar for Male & Females

DEPENDENCY OF CREDIT VS EDUCATION STATUS FOR LOAN APPROVALS:

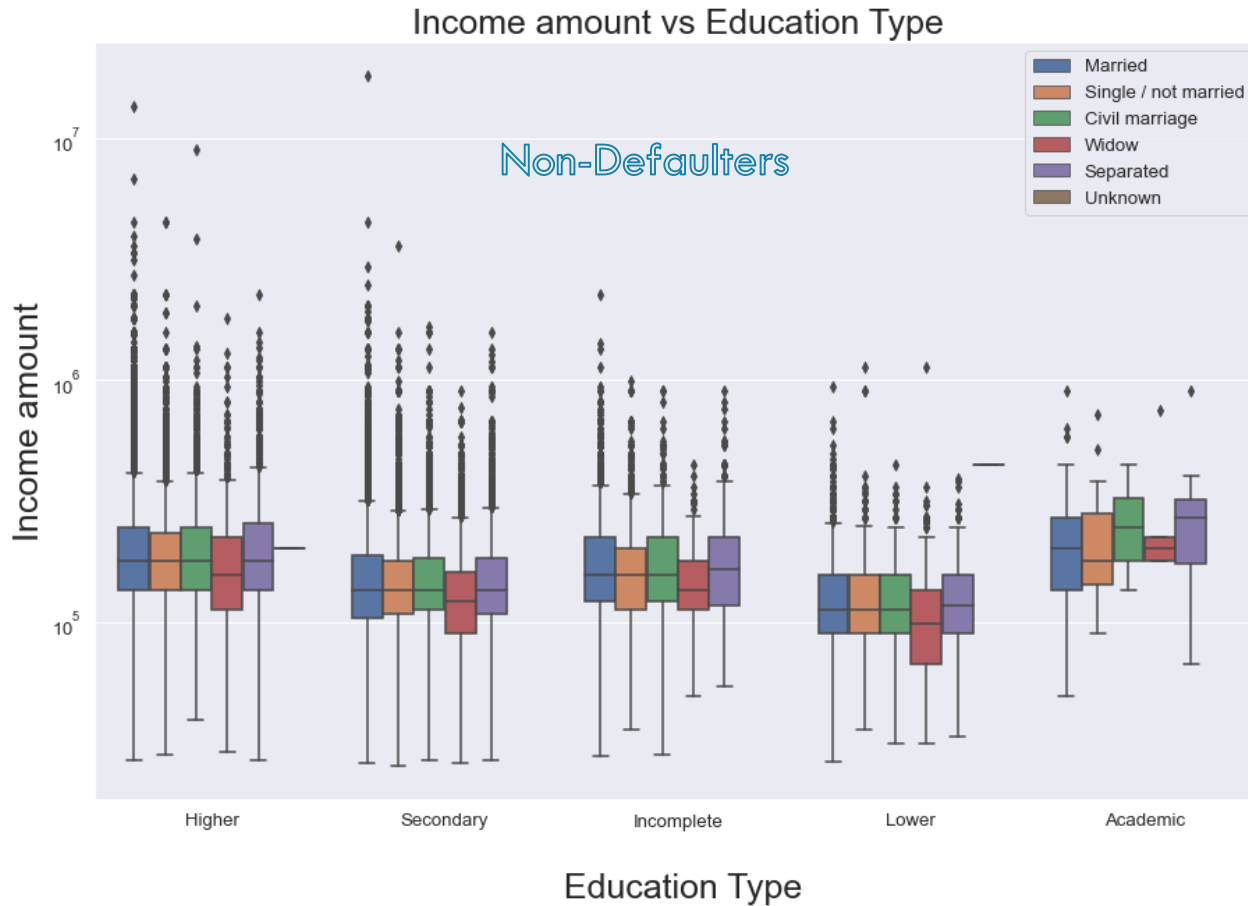


- The Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- The higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers.
- Civil marriage for Academic degree is having most of the credits in the third quartile.

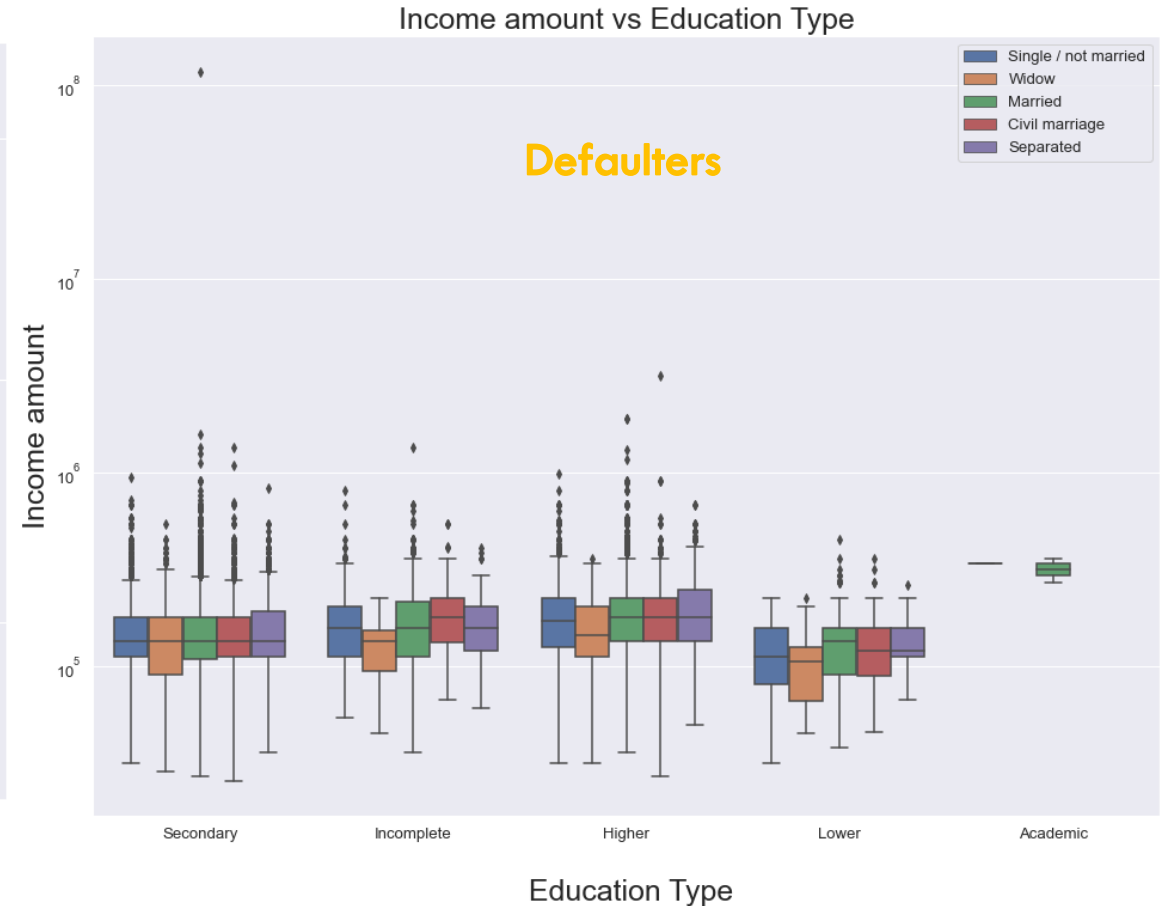


- Quite similar with Non-Defaulters
- Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- Civil marriage for Academic degree is having most of the credits in the third quartile.

DEPENDENCY OF INCOME VS EDUCATION STATUS FOR LOAN APPROVALS:



From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers. Less outlier are having for Academic degree but there income amount is little higher than Higher education. Lower secondary of civil marriage family status have less income amount than others.



Have some similarity with Non-Defaulters, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. Less outlier are having for Academic degree but there income amount is little higher than Higher education. Lower secondary have less income amount than others.

MULTIVARIATE ANALYSIS

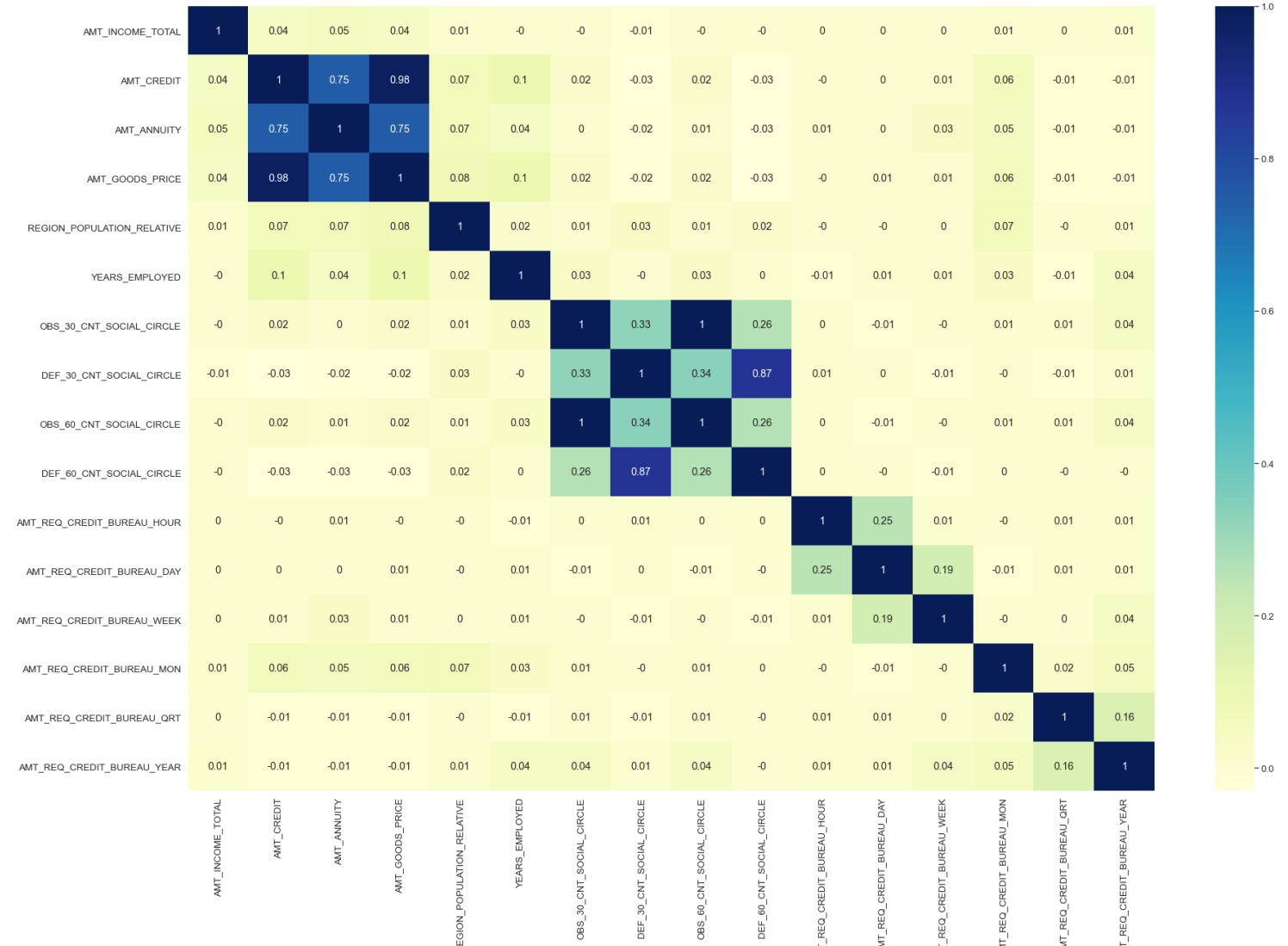
HEAT MAP FOR DEFAULTERS

3 most positive correlations

- 1.AMT_CREDIT - AMT_GOODS_PRICE
- 2.CNT_CHILDREN - CNT_FAM_MEMBERS
- 3.AMT_CREDIT - AMT_ANNUITY

3 most negative correlations

- 1.AMT_CREDIT - AMT_INCOME
- 2.AMT_REQ_CREDIT_BUREAU_YEAR - AMT_CREDIT
- 3.REGION_POPULATION_RELATIVE - AMT_GOODS_PRICE



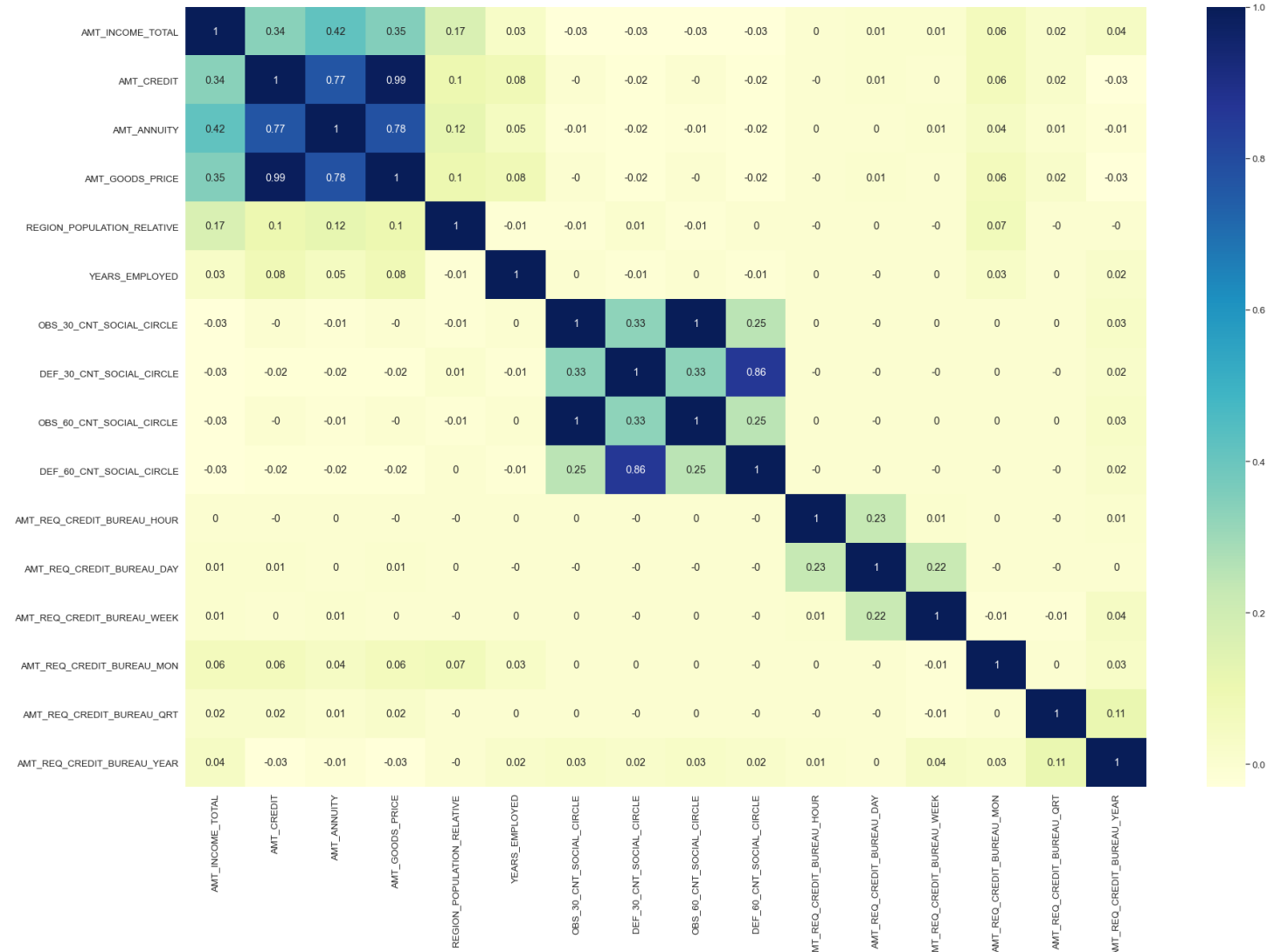
HEAT MAP FOR NON-DEFAULTERS

3 most positive correlations

1. AMT_CREDIT - AMT_GOODS_PRICE
2. CNT_CHILDREN - CNT_FAM_MEMBERS
3. AMT_CREDIT - AMT_ANNUITY

3 most negative correlations

1. AMT_REQ_CREDIT_BUREAU_YEAR - AMT_CREDIT
2. AMT_REQ_CREDIT_BUREAU_YEAR - AMT_GOODS_PRICE
3. REGION_POPULATION_RELATIVE - YEARS_EMPLOYED





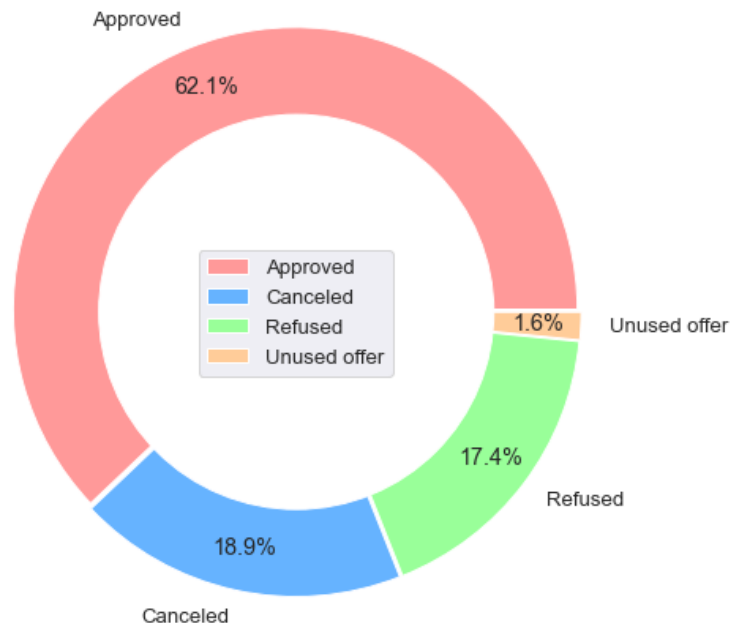
PREVIOUS APPLICATION DATA



UNIVARIATE ANALYSIS OF PREVIOUS APPLICATION

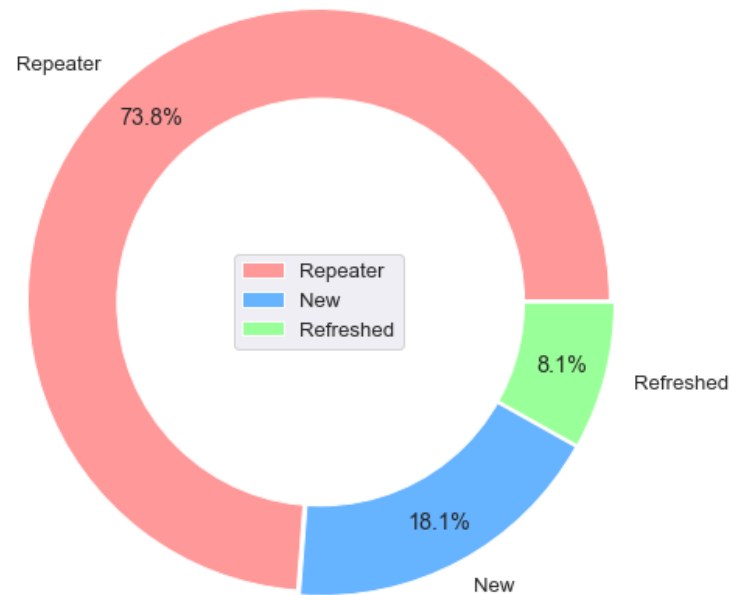
➤ We can observe that majority of loans are approved and very less percentage of loans are unused offer

Contract status of previous application



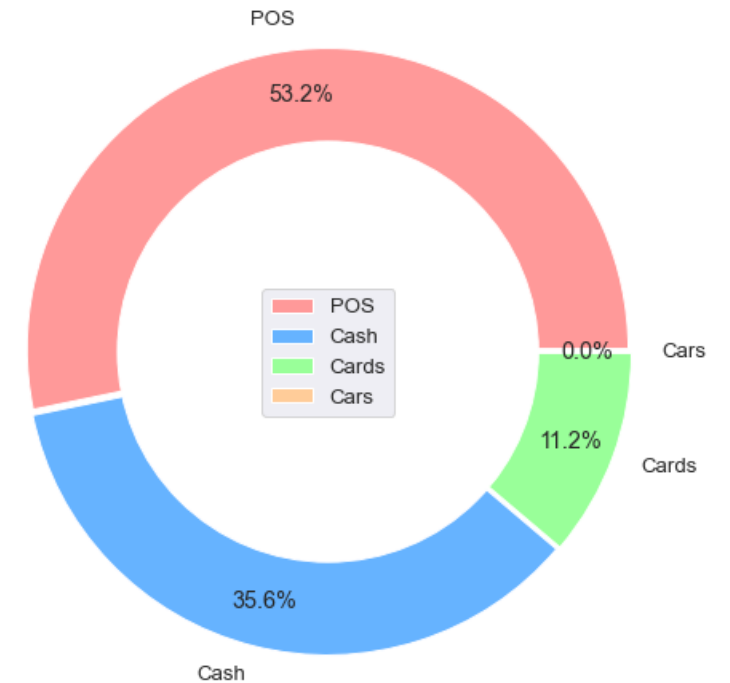
➤ We observe that majority of applicants are repeaters.

Client type for the previous application



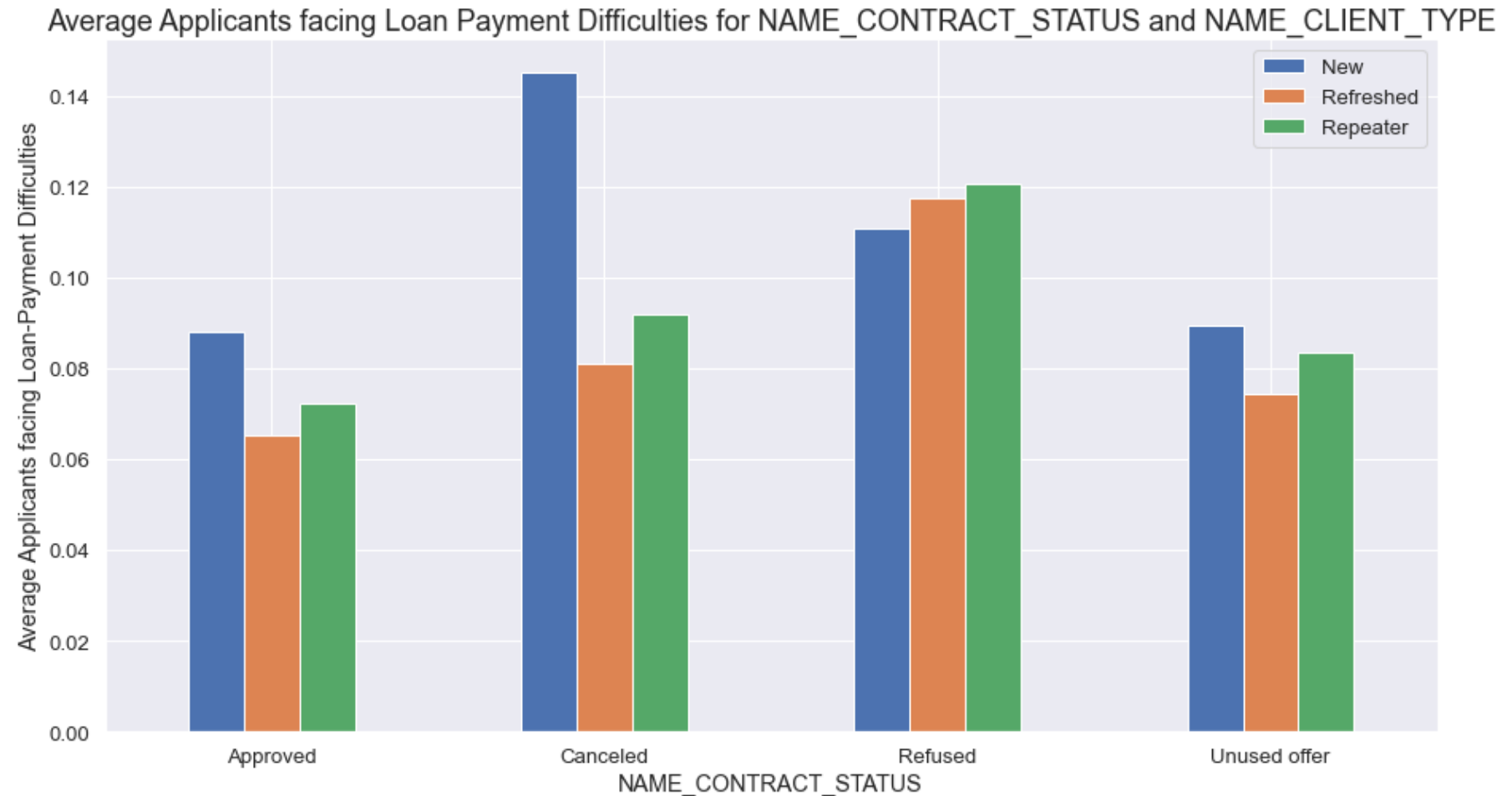
➤ We observe that majority of applicants Portfolio's are for POS and a good amount of it is for cash.

Portfolio for the previous application



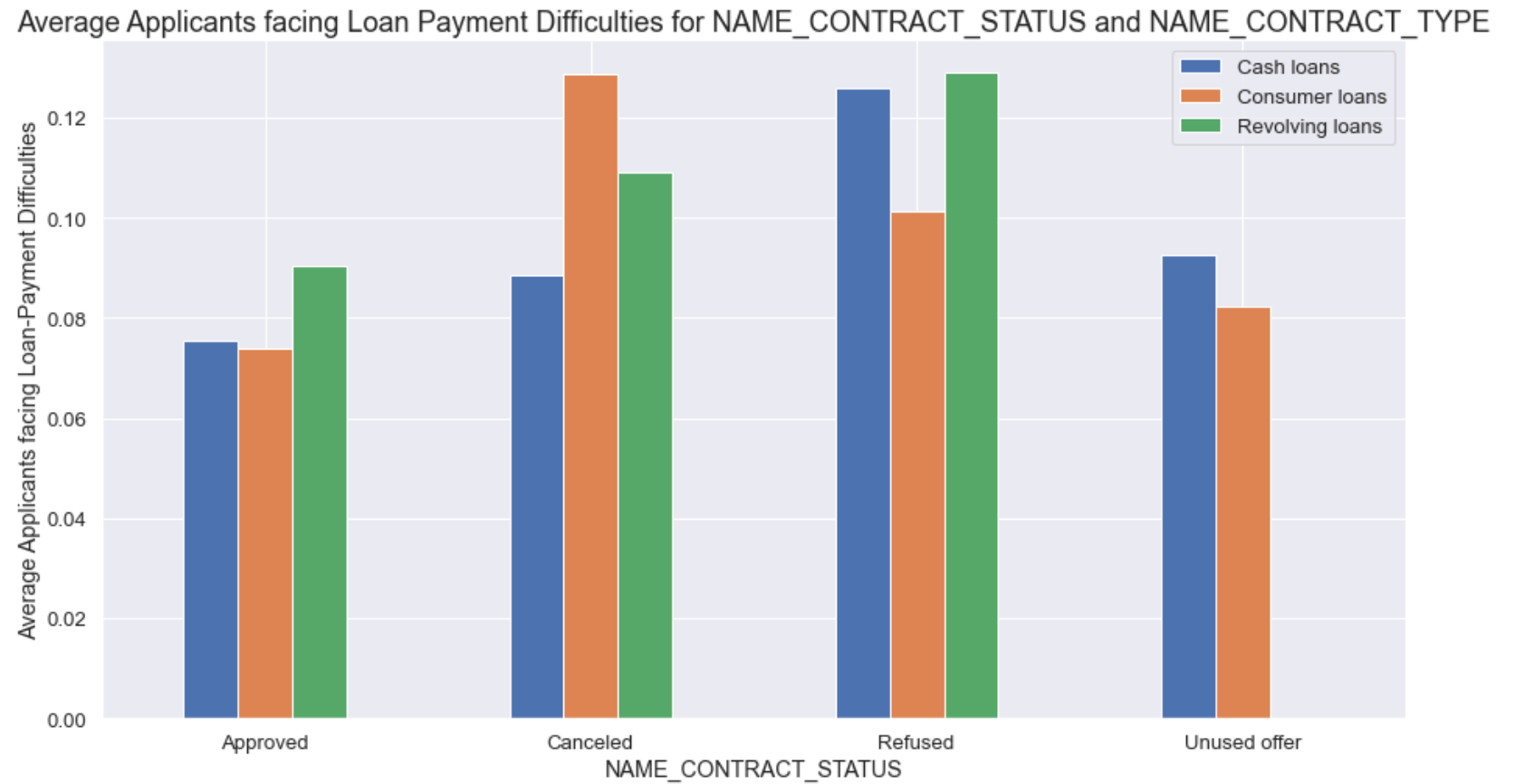
BIVARIATE ANALYSIS OF PREVIOUS APPLICATION

- It can be observed that Client who where 'New' and had 'Cancelled' previous application tend to have more % of Loan-Payment Difficulties in current application.



BIVARIATE ANALYSIS OF PREVIOUS APPLICATION CONT.

- It can be observed that Clients with 'Revolving loans' and with 'Refused' previous application tend to have more % of Loan-Payment Difficulties in current application



CONCLUSION

The company should be cautious of the following flags while providing loans to customers if:

- Applicant's employment experience is less than 10 years.
- Applicant falls in the age group of 21-30.
- Education type is Lower secondary or Secondary/Secondary .
- The applicant is a Low Skilled Laborers.