

CREDIT EDA ASSIGNMENT



By: Anindya Ghatak
DSC-56

PROBLEM STATEMENT

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments 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.

To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough.

APPROACH FOLLOWED

Each dataset(provided)* is analysed as per the following steps:

- Data inspection.
- Data cleaning and fixing anomalies.
- Handling null values
- Dropping the extra/unnecessary columns
- Imputing (If needed)
- Handling outliers
- Typecasting data types
- Data visualization by univariate, bivariate and multivariate

DATA SET PROVIDED:

1. APPLICATION DATASET= contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
2. PREVIOUS APPLICATION DATASET= contains information about the client's previous loan data. It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.



DATA ANALYSIS



BASIC VISUAL UNDERSTANDING OF APPLICATION DATA

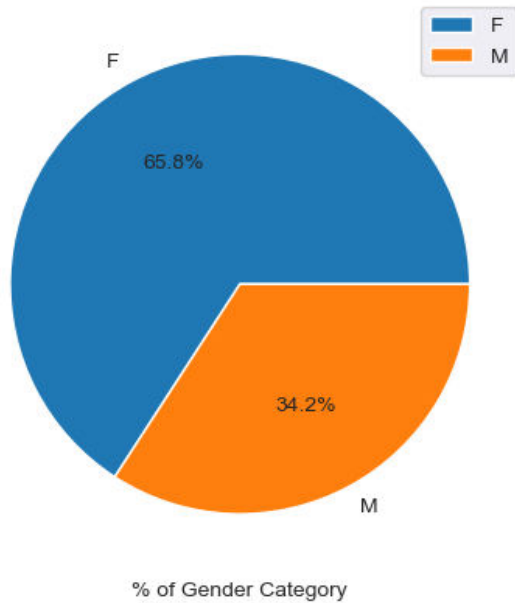


FIG-01

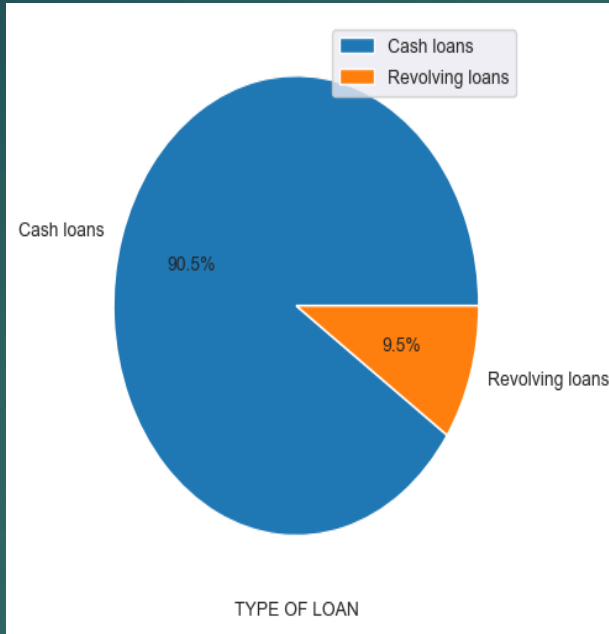


FIG-02

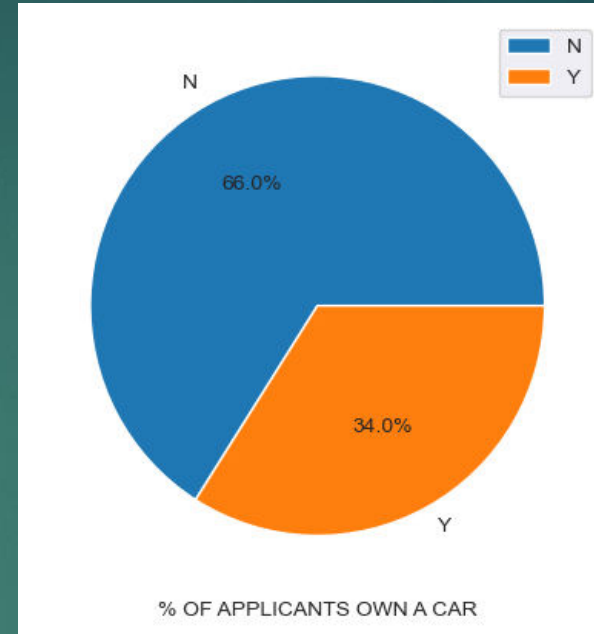


FIG-03

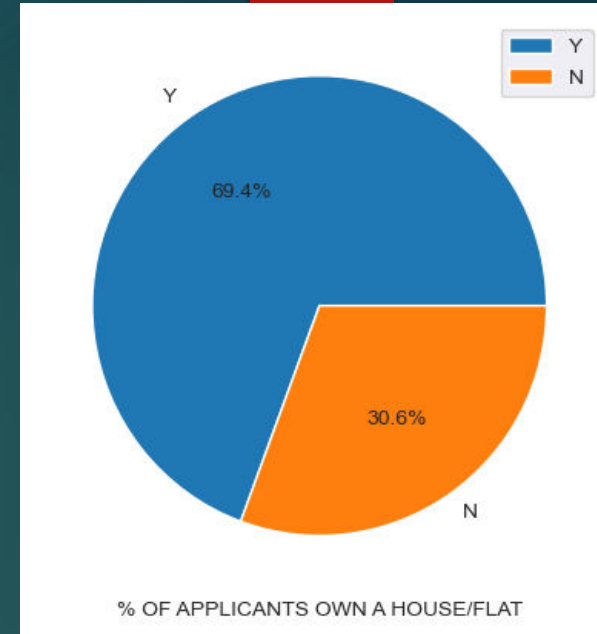


FIG-04

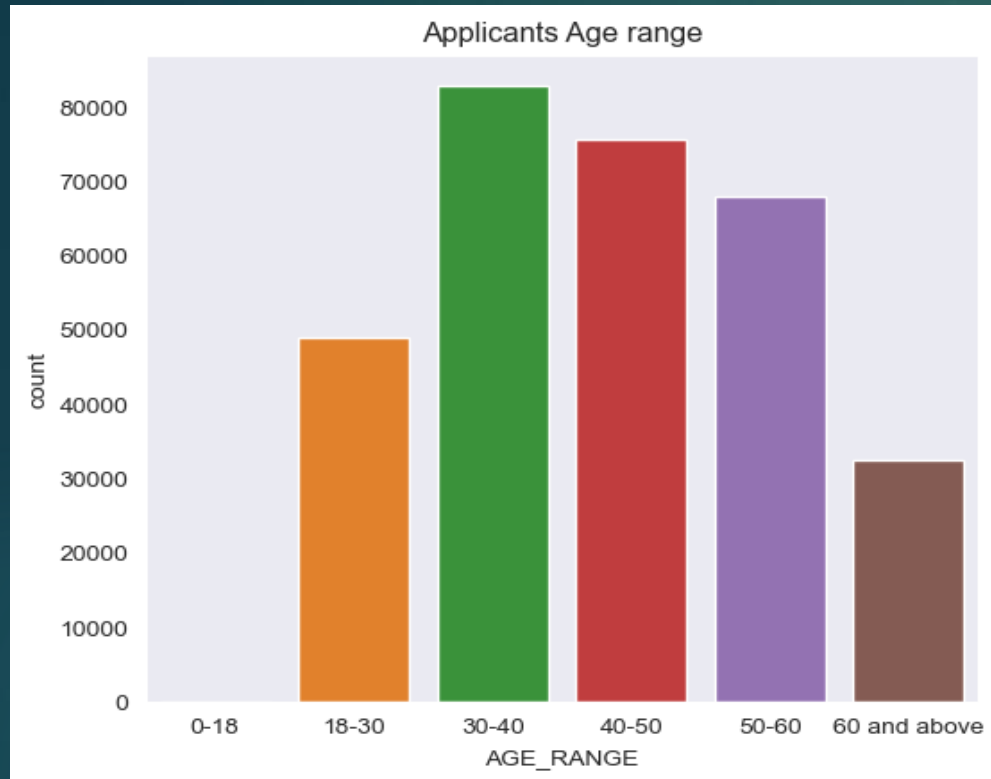
FIG-01: From the graph and data above, we can conclude that most of the applicants are females(F) which is around 65.8%.

FIG-02: From the above graphical representation we can say most of the loan applicants has applied for is Cash Loan(90.5%) as compared with Revolving Loans(9.5%)

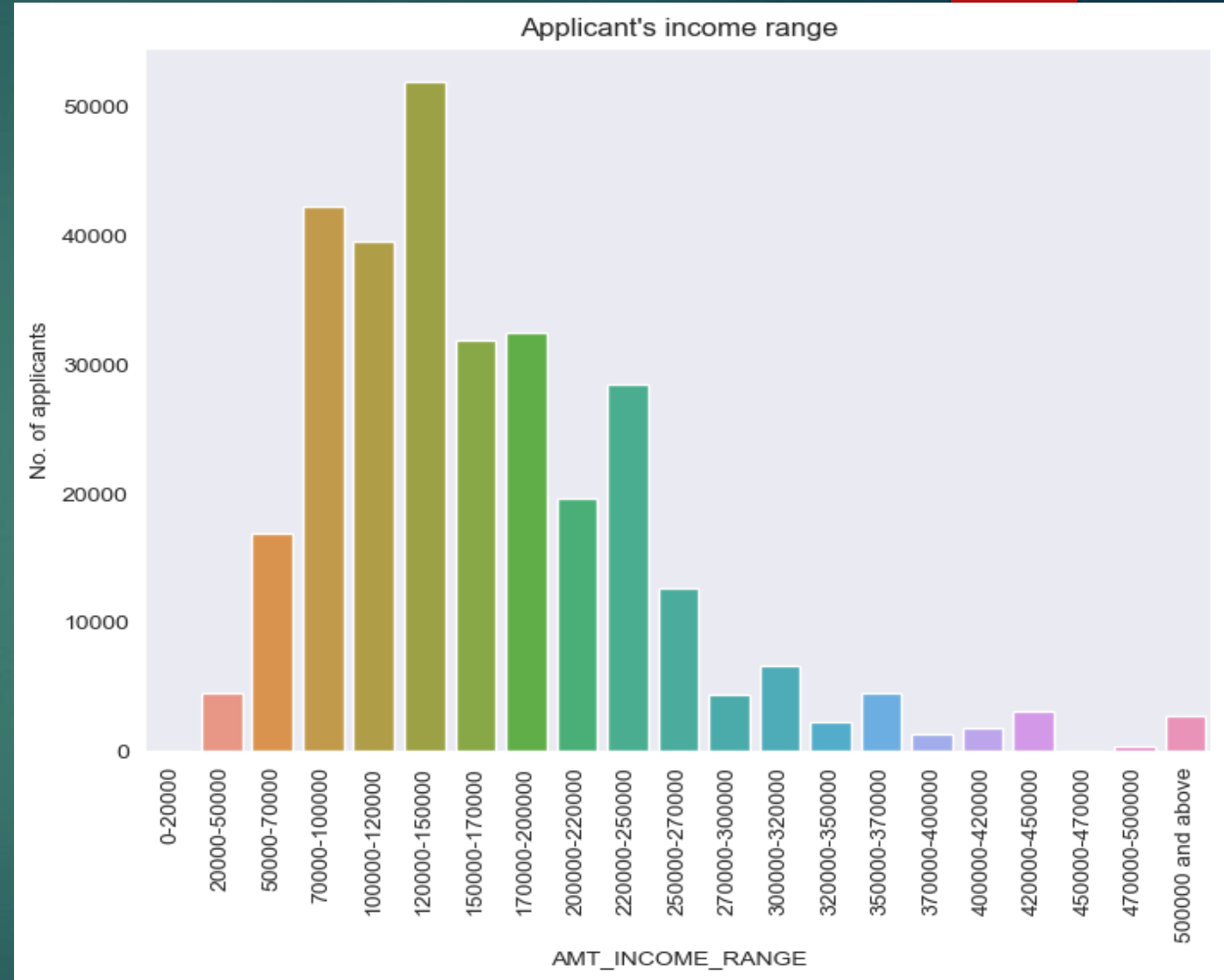
FIG-03: From the above graph we can interpret that most of the applicant does not own a car.

FIG-04: From the above graph we can interpret that most of the applicant does own a house/flat which is at (69.4%).

BASIC VISUAL UNDERSTANDING OF APPLICATION DATA

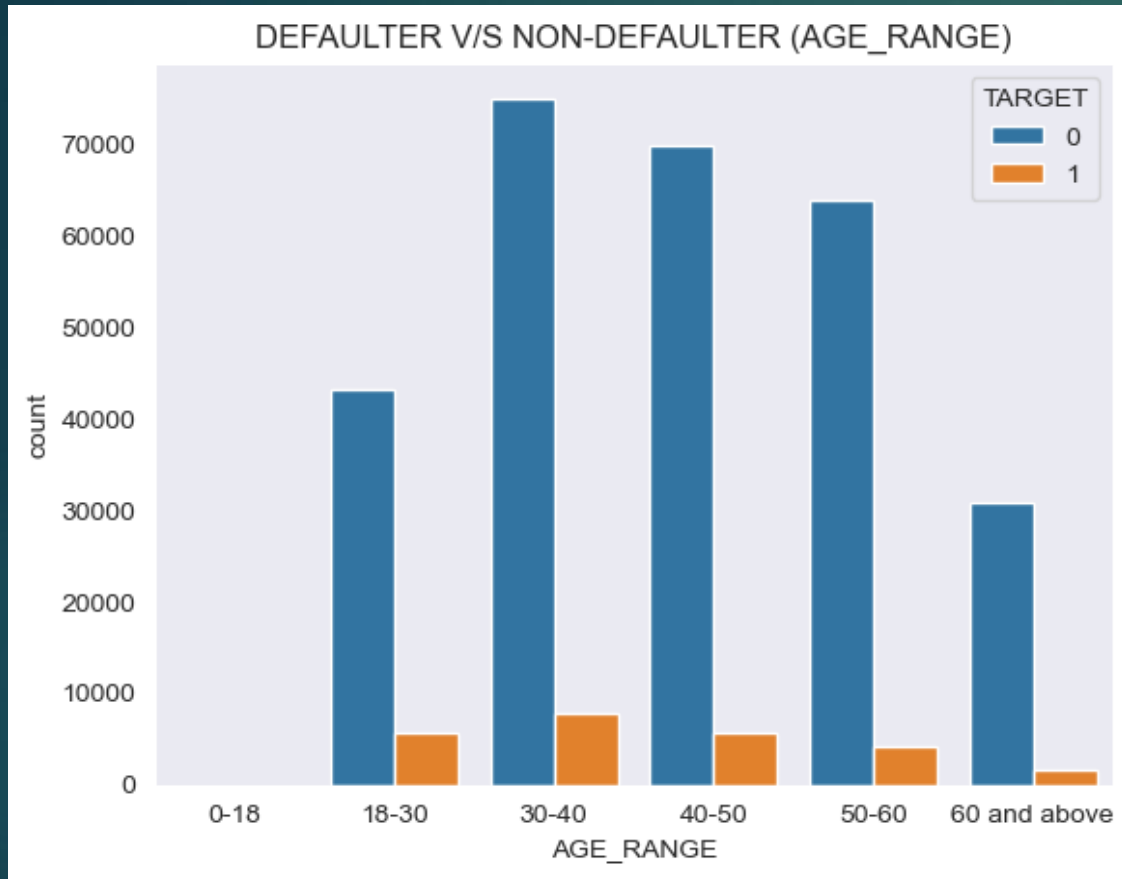


MOST OF THE APPLICANTS APPLYING FOR LOAN BELONG TO THE AGE GROUP OF 30-40

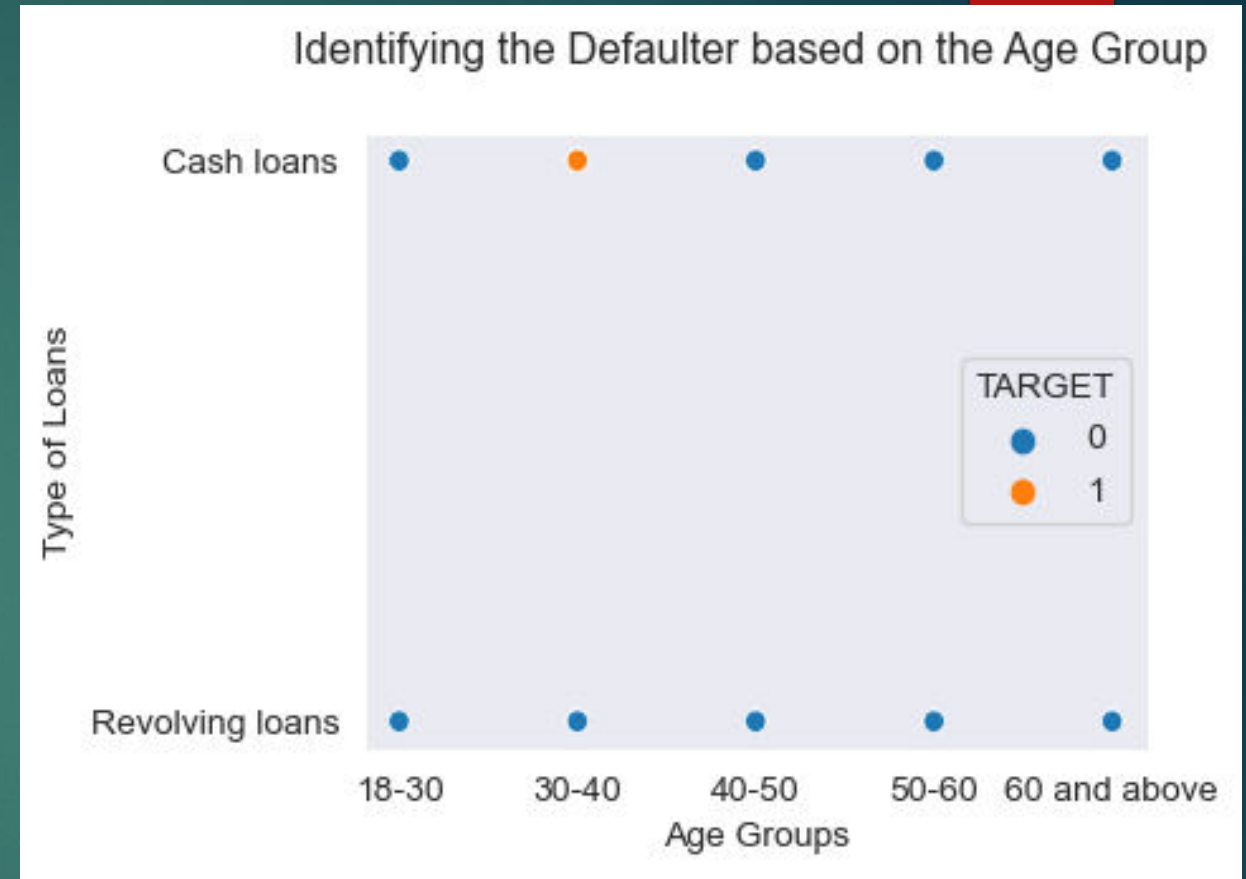


MOST OF THE APPLICANTS BELONG TO THE INCOME RANGE OF 120000-150000

VISUALISING THE DEFAULTERS V/S NON-DEFAULTERS

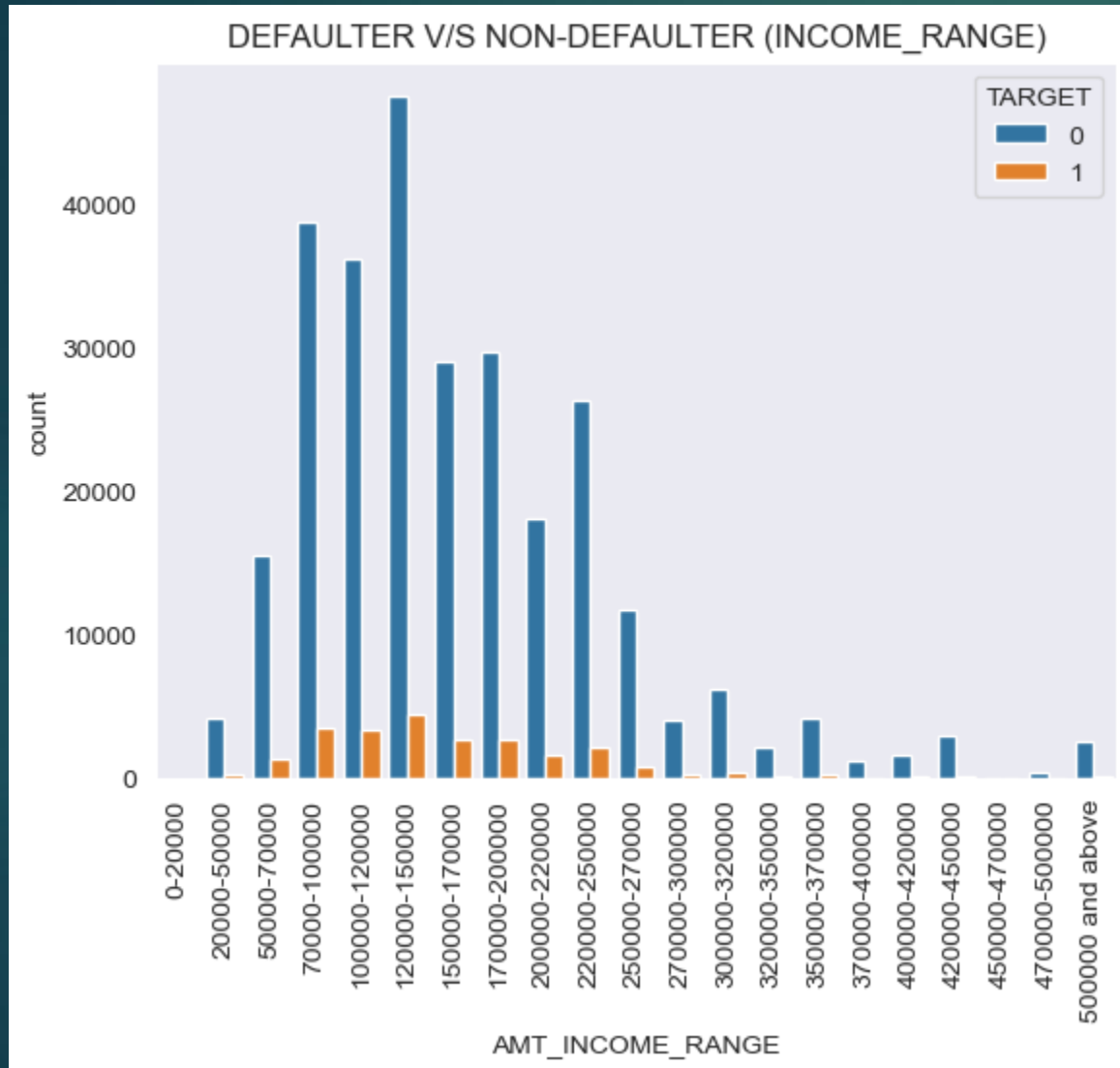


- AS WE HAVE SEEN MAJORITY OF THE APPLICANTS BELONG TO THE AGE GROUP OF 30-40. SIMILARLY WE DO SEE THE MAJORITY OF NON-DEFAULTER IN THE SAME AGE GROUP



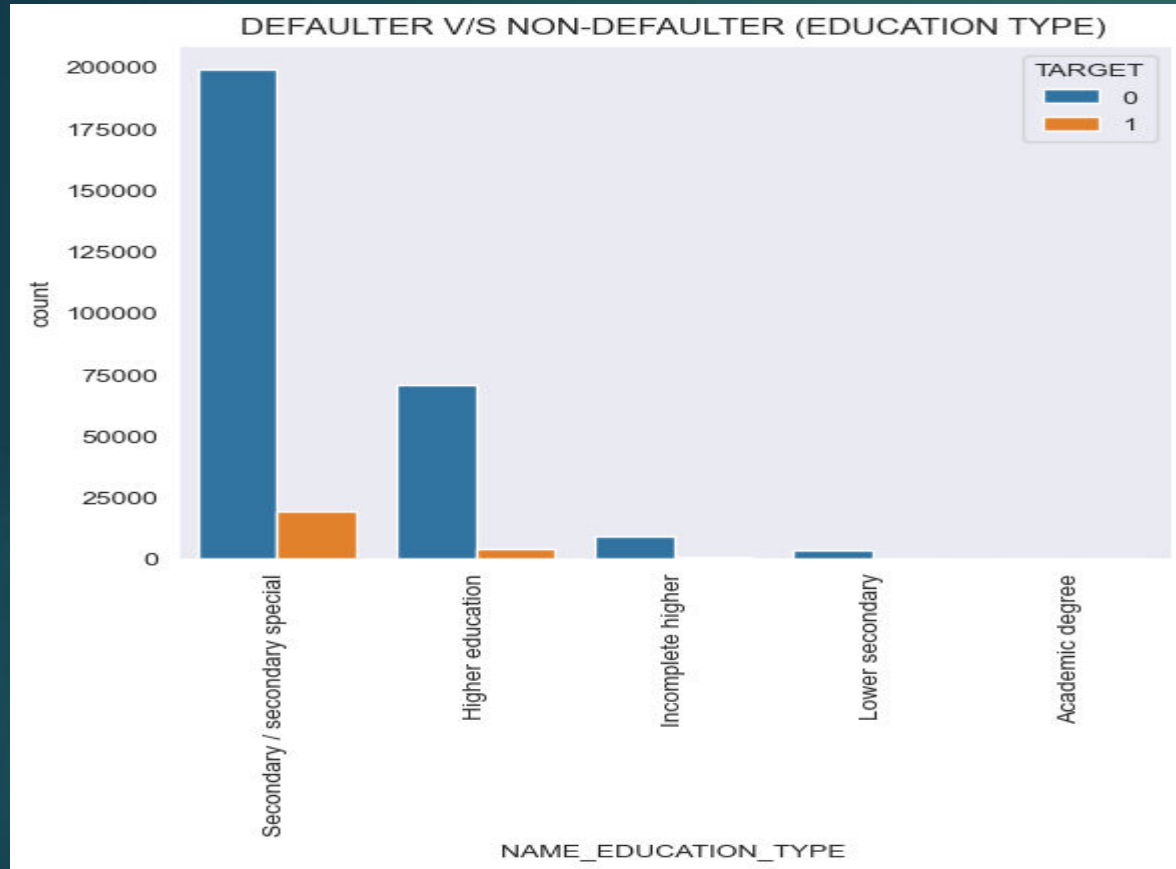
- ABOVE PLOT GIVE MORE INSIGHTS FOR THE DEFAULTER IN THE AGE GROUPS.
- WE SEE DEFUALTERS ARE MOSTLY IN THE AGE GROUP (30-40) CONCERNING CASH LOAN.
- THIS COULD BE DUE TO EXCESS SPENDING/ IMPROPER FINANCIAL PLANNING/LIFESTYLE INFLUENCE.

VISUALISING THE DEFAULTERS V/S NON-DEFAULTERS

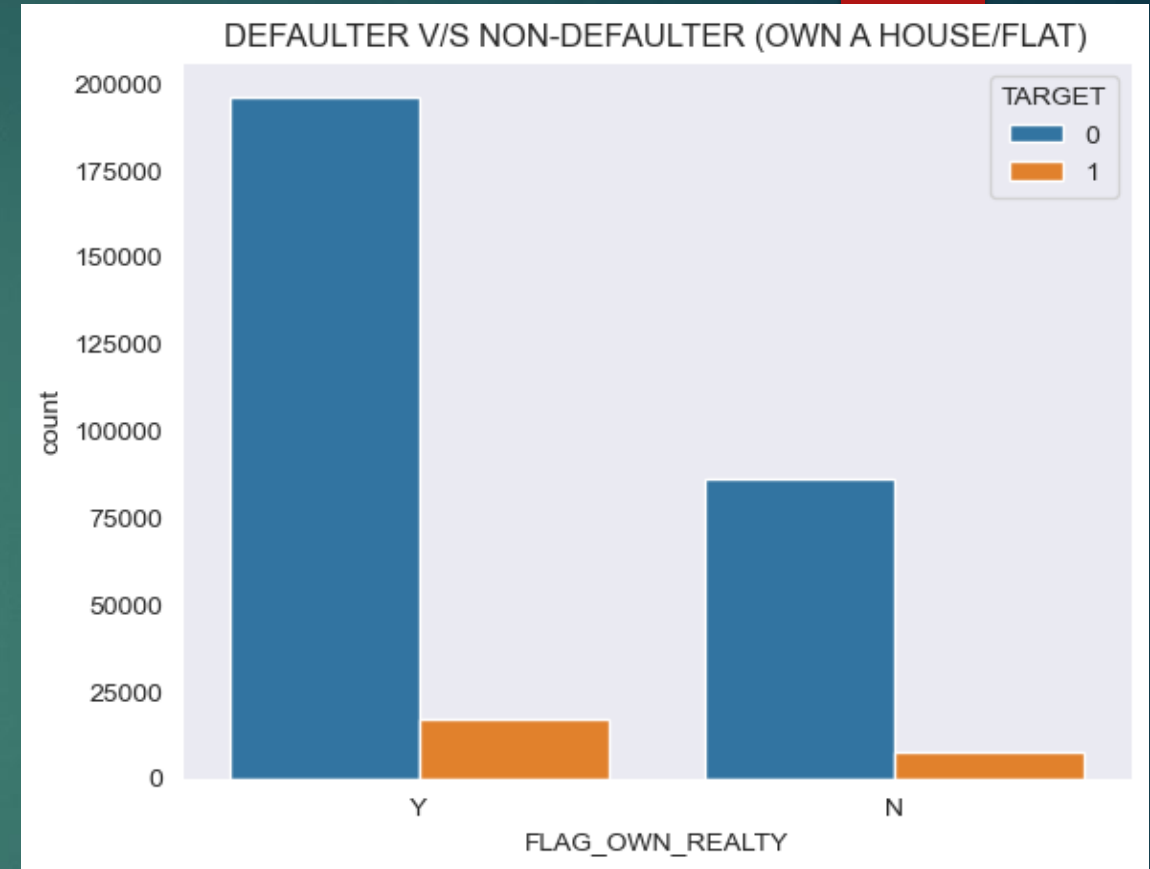


- WE DO SEE THAT THE INCOME GROUP (120000-150000) IS A TARGET INCOME GROUP AS THEY HAVE MORE NUMBER OF NON-DEFAULTER APPLICANTS.
- ALSO, WE CAN SEE THAT APPLICANTS WITH INCOME RANGE MORE THAN 3,00,000 ARE LESS LIKELY TO MAKE ANY DEFAULT IN PAYING BACK THE MONEY HENCE, THEY ARE ALSO A GOOD TARGET TO ANALYSE

VISUALISING THE DEFAULTERS V/S NON-DEFAULTERS

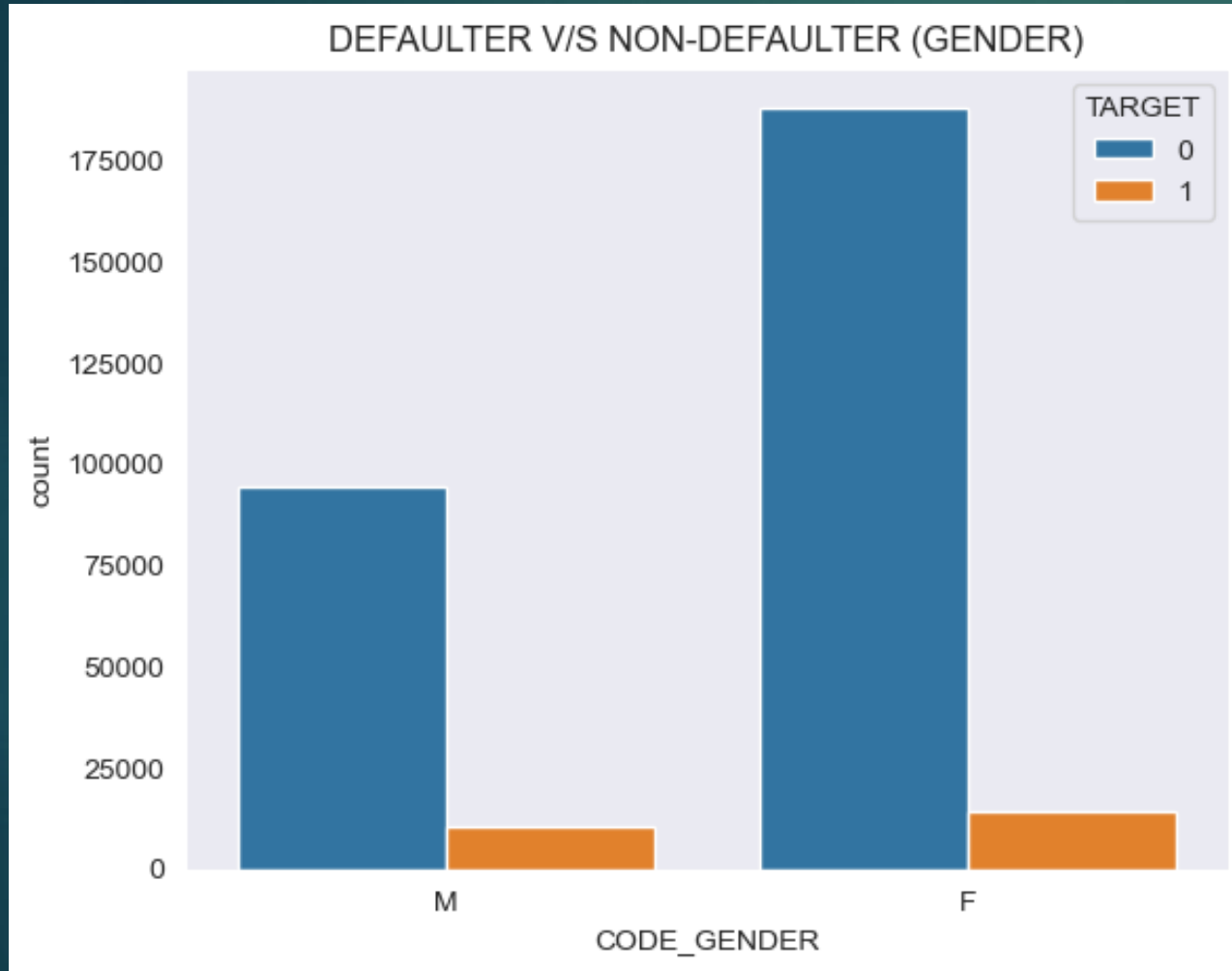


WE DO SEE THAT THE APPLICANTS BELONGING TO THE EDUCATION CATEGORY OF SECONDARY/SECONDARY SPECIAL ARE MORE LIKELY TO BE TARGETED AS THEY ARE DISCIPLINED IN PAYING BACK THE LOAN WITHOUT ANY MISS.



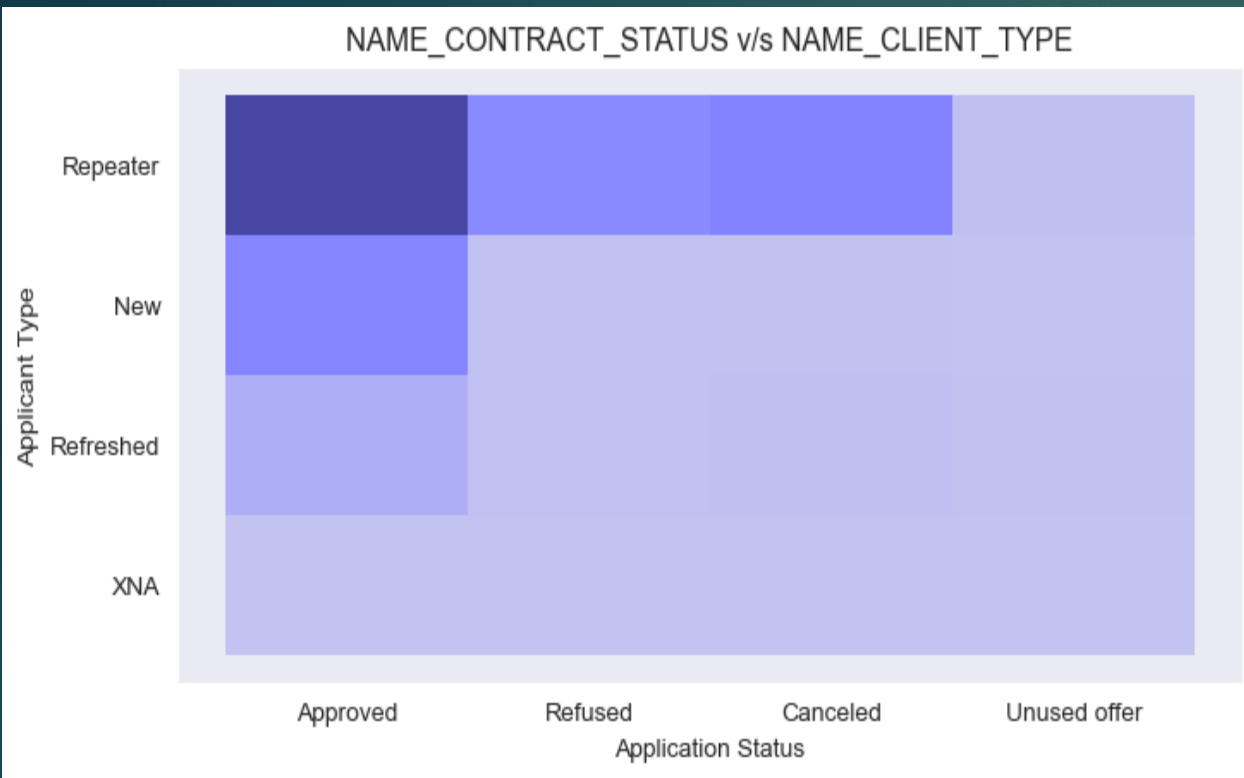
WE SEE THAT THAT APPLICANT OWNING A HOUSE/FLAT ARE MORE DISCIPLINED IN PAYING BACK THE LOAN AMOUNT ON TIME AND HENCE WE SEE A HIGH AMOUNT OF NON-DEFAULTER AS COMPARED TO THE APPLICANTS WHO DOES NOT OWN A HOUSE/FLAT.

VISUALISING THE DEFAULTERS V/S NON-DEFAULTERS



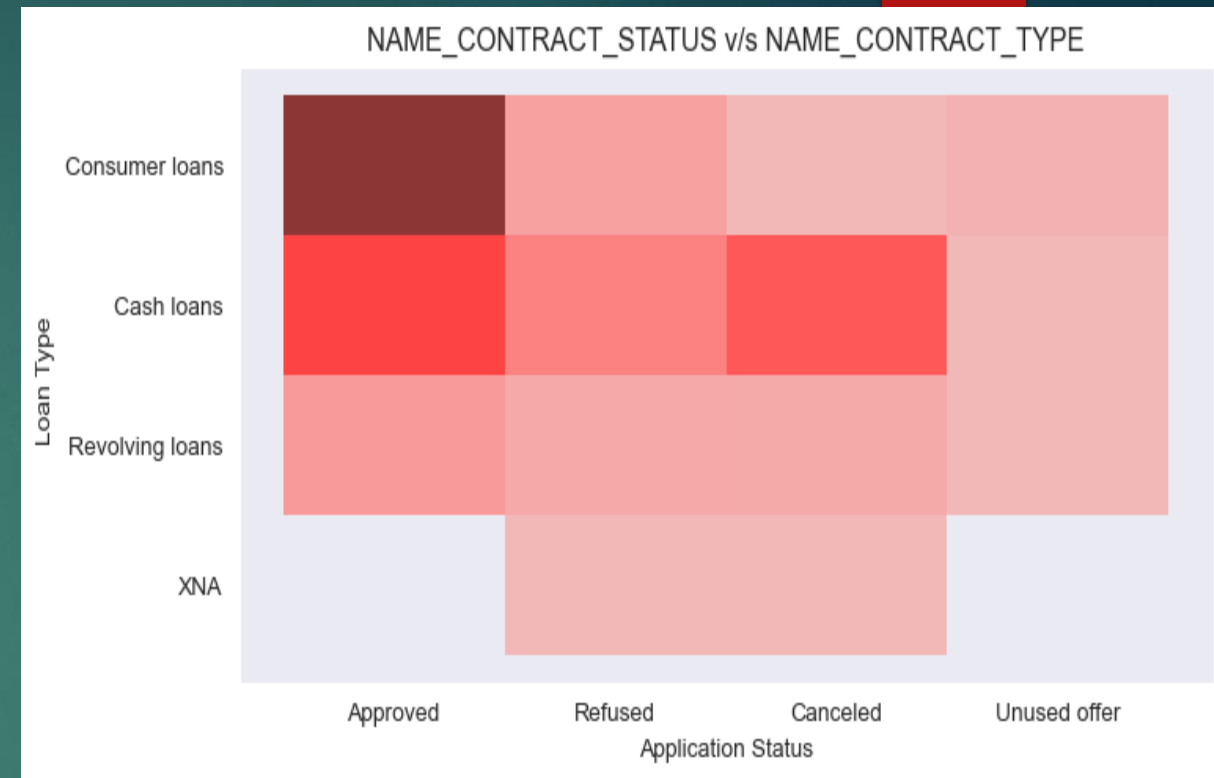
WE SEE THAT FEMALES ARE HIGHLY DISCIPLINED IN PAYING BACK THE LOAN ON TIME AS COMPARED TO MALES HENCE, THEY ARE A GOOD APPLICANTS TO APPROVE.

VISUALISING THE APPLICANT TYPE V/S LOAN STATUS



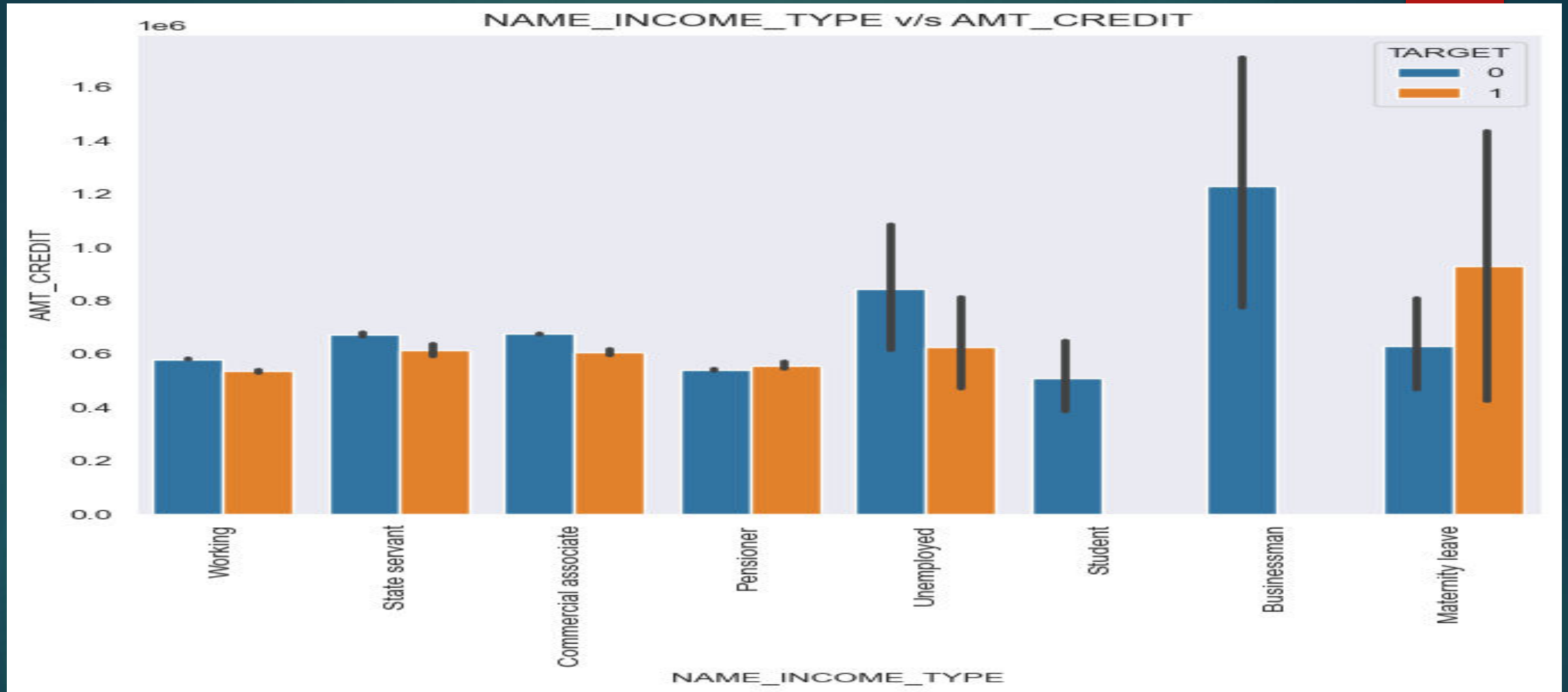
WE SEE THAT MOST OF THE LOAN APPLICATION WERE APPROVED FOR THE REPEATER APPLICANTS AS THEY MIGHT HAVE A CLEAR BACKGROUND AND A GOOD CREDIT SCORE OR HISTORY OF NO DUE PAYMENT

VISUALISING THE LOAN TYPE V/S LOAN STATUS



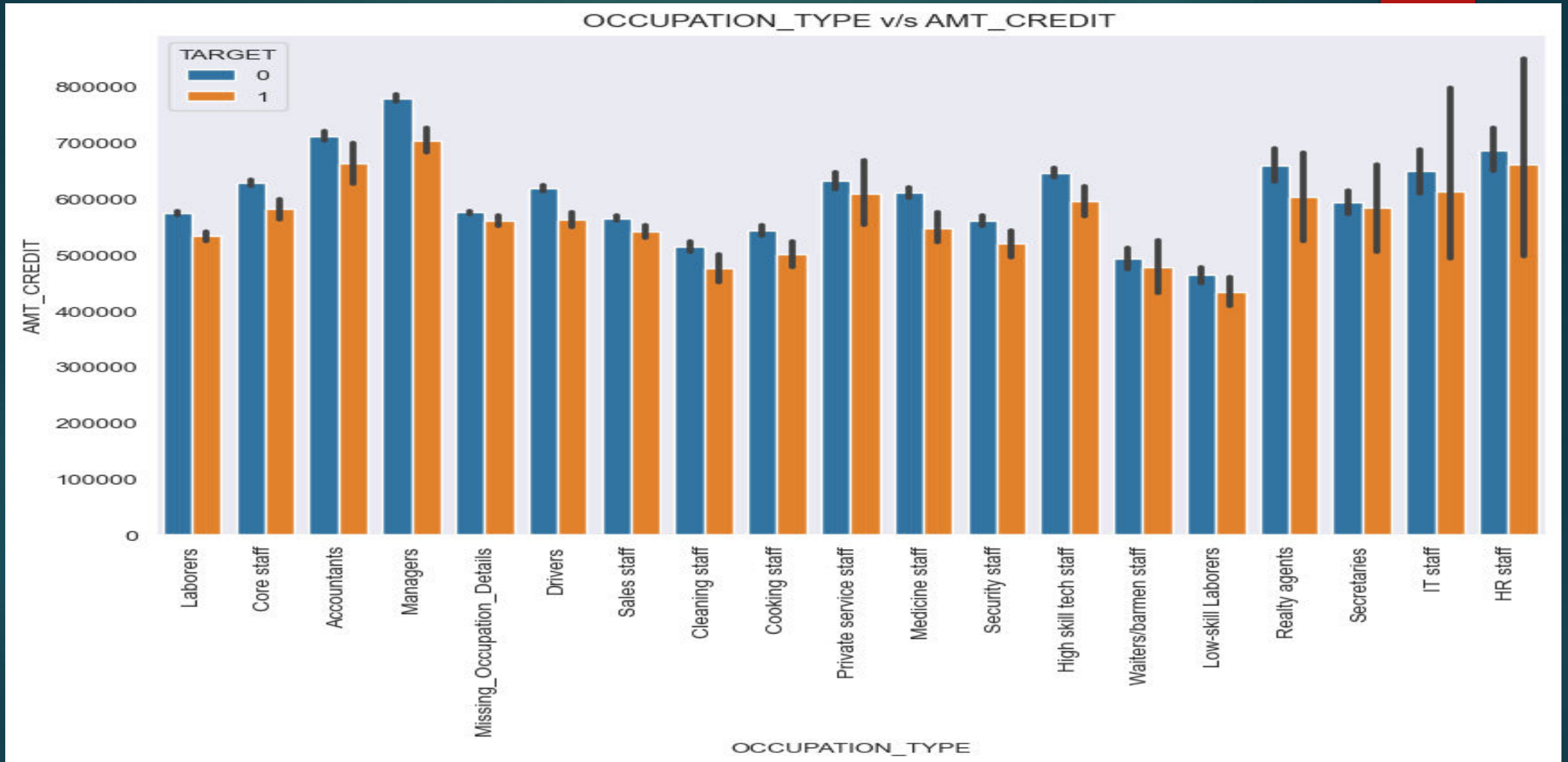
- HERE WE SEE THAT MOST THE LOAN APPLICATION APPROVED ARE CONSUMER LOAN FOLLOWED BY CASH LOAN.
- SECONDLY, WE DO SEE THAT THE NUMBER OF CANCELLED LOAN IS QUITE HIGH FOR CASH LOAN

VISUALISING THE INCOME TYPE V/S AMOUNT TAKEN FOR CREDIT



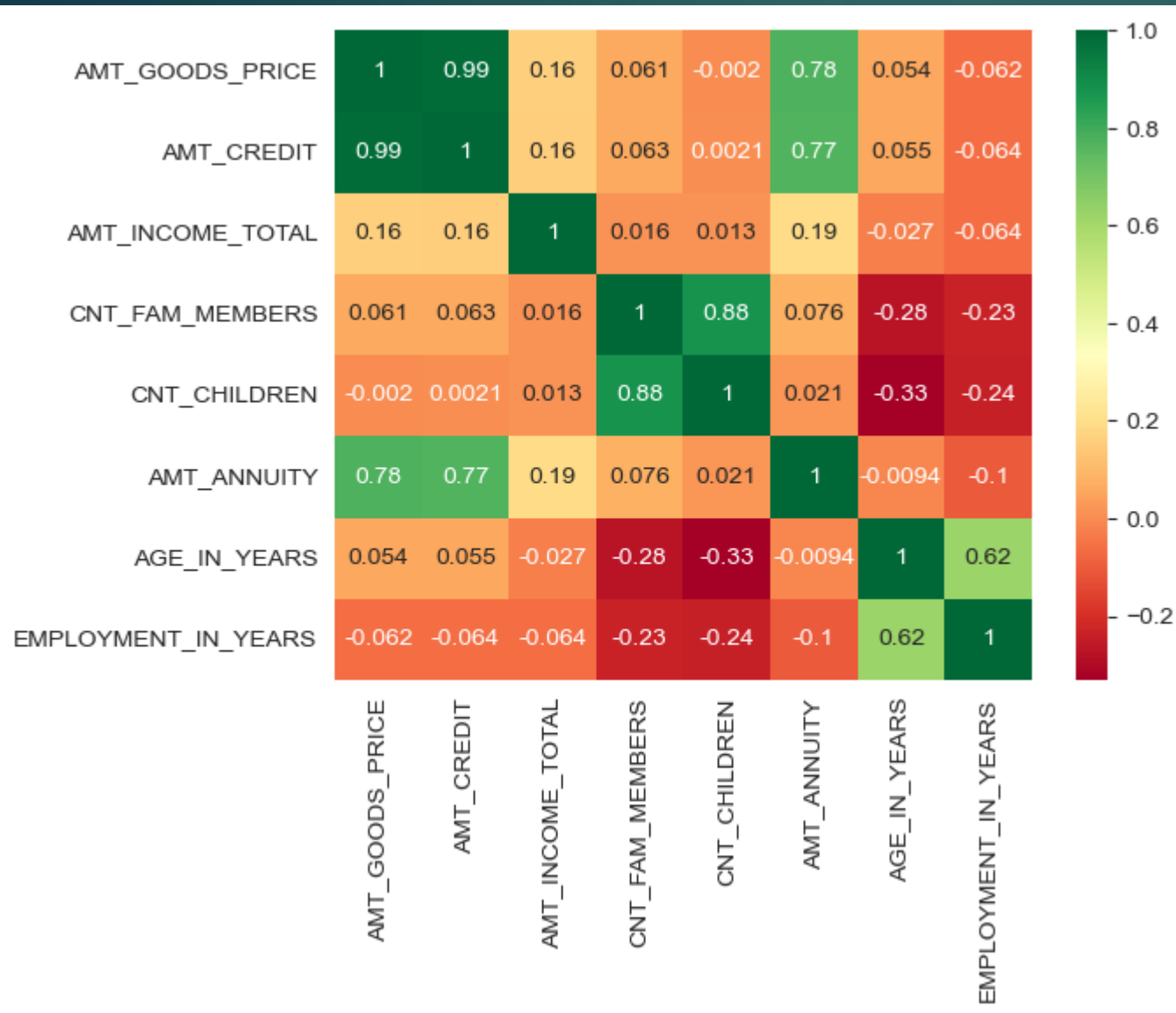
- Based on the above visual, we can say that the majority of applicants who are on Maternity leave are not able to pay their loan back and hence we see high defaulter following for those belonging to Unemployed category are defaulter as they fall under low income range.
- Again we see that for applicants belonging to Students and Businessman do not have any defaulter.
- We also see that business man take highest credit (loan amount) as compared to all.

VISUALISING THE OCCUPATION TYPE V/S AMOUNT TAKEN FOR CREDIT



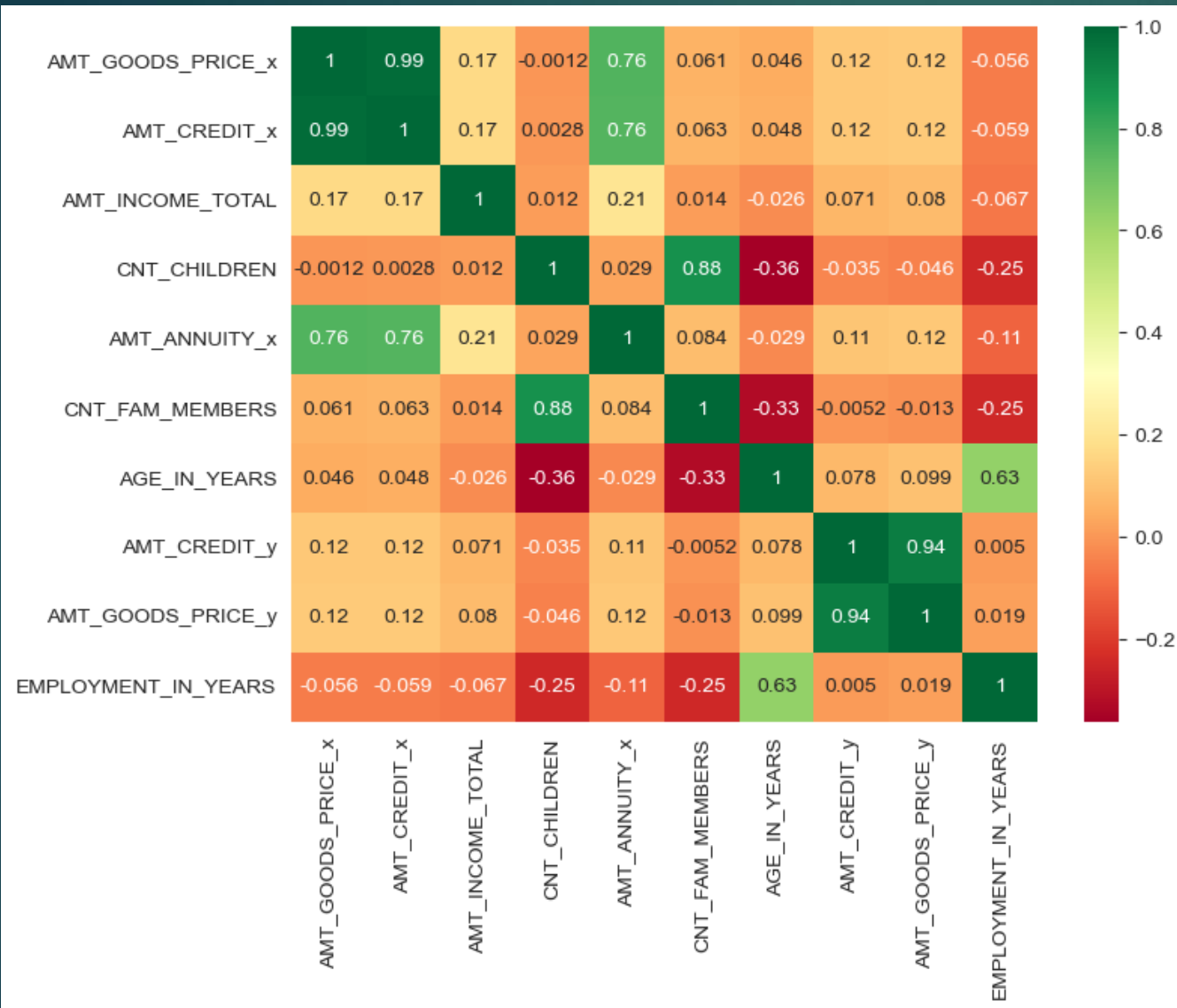
From the above graph, we can interpret that Applicants with occupation category as Manager has applied for maximum credit(loan)

VISUALISING THE CORRELATION OF THE APPLICATION DATA



- Based on the above visual representation, firstly we can interpret that the amount taken for credit i.e 'AMT_CREDIT' has a high correlation factor with the amount goods price i.e 'AMT_GOODS_PRICE' which is understandable as more the goods price more is the loan amount.
- Secondly, We do see a good correlation between 'AMT_ANNUITY', 'AMT_CREDIT' and 'AMT_GOODS_PRICE'.
- Thirdly, we also see a good correlation between 'CNT_CHILDREN' and 'CNT_FAM_MEMBERS'.
- We also see that a good correlation between 'AGE_IN_YEARS' and 'EMPLOYMENT_IN_YEARS'.
- lastly, we do see negative correlation with more time being employed(EMPLOYMENT_IN_YEARS) and the loan amount being issued. which is again understandable as bank would not give loan for senior people who are in verge of retirement as they might be leading to be as defaulter.

VISUALISING THE CORRELATION OF THE COMBINED DATA



- Based on the above visual we can interpret, that loan amount taken('AMT_CREDIT_X','AMT_CREDIT_Y') are both highly correlated with the goods price against which the loan was taken ('AMT_GOODS_PRICE_x', 'AMT_GOODS_PRICE_y').
- Followed by, We can also interpret that the 'CNT_FAM_MEMBERS' has a good correlation with 'CNT_CHILDREN'.
- Followed by, we can see that 'AMT_ANNUITY' is highly correlated with 'AMT_GOODS_PRICE' and 'AMT_CREDIT', as more/less the amount taken for credit more/less is the amount annuity.
- Followed by, we can say that the 'AGE_IN_YEARS' is also good correlated with the 'EMPLOYMENT_IN_YEARS' column.

CONCLUSION FROM THE LOAN CASE STUDY

- BASED ON THE STUDY, WE SEE THAT THE APPLICANTS BELONGING THE AGE GROUP OF 30-40
- BASED ON THE STUDY, WE SEE THAT THE APPLICANTS BELONGING TO THE INCOMING RANGE OF 120000-150000 ARE THE DISCIPLINED APPLICANTS AS THEY PAY THEIR LOAN ON TIME WITHOUT ANY DUE.
- WE ALSO SEE THAT APPLICANTS BELONGING TO HIGH INCOME GROUP ARE ALSO GOOD APPLICANTS AS THEY DO NOT HAVE ANY DEFAULTERS AND LIKELY TO REPAY THE LOAN BACK.
- WE SEE THAT APPLICANTS WITH SECONDARY/SECONDARY SPECIAL EDUCATION HAS HIGH NON-DEFAULTER APPLICANTS HENCE, THEY ARE A GOOD APPLICANTS TO TARGET.
- WE SEE THAT APPLICANTS WHO OWNS A FLAT/HOUSE ARE A GOOD APPLICANT AS THEY ARE LIKELY TO REPAY THE LOAN AMOUNT BECOMING NON-DEFAULTER.
- WE SEE THAT FEMALES ARE HIGHLY DISCIPLINED IN PAYING BACK THE LOAN AMOUNT ON TIME WITHOUT ANY DEFAULT.
- BASED ON THE APPLICANT TYPE, WE SEE THAT REPEATER ARE THE ONES WHOSE LOAN APPLICATIONS ARE MOSTLY APPROVED.
- BASED ON THE LOAN TYPE, WE SEE THAT CONSUMER LOAN IS MOSTLY APPROVED FOLLOWING CASH LOAN WHICH IS MORE OPTED BY THE APPLICANTS.
- WE ALSO SEE A GOOD AMOUNT OF APPLICATION BEING CANCELLED FOR CASH LOAN.
- BUSINESSMAN AND STUDENT ARE A GOOD APPLICANTS AS THEY DO NOT HAVE ANY DEFAULTER, AND ALSO BUSINESSMAN ALSO TAKES MAXIMUM AMOUNT OF CREDIT.
- MANAGERS ARE THE ONE WHO HAS TAKEN MAXIMUM CREDIT AMOUNT.

KEY POINTS & RECOMMENDATIONS

RECOMMENDED GROUP/APPLICANTS

1. FEMALE APPLICANTS ARE MORE RECOMMENDED THAN MALE APPLICANTS.
2. APPLICANTS WITH SECONDARY/SECONDARY SPECIAL ARE MORE FAVOURABLE.
3. APPLICANTS HAVING FLAT/HOUSE ARE MORE FAVOURABLE APPLICANTS.
4. APPLICANTS BELONGING TO HIGH INCOME GROUP ARE MORE FAVOURABLE.
5. REPEATER APPLICANTS ARE MORE FAVOURABLE TO GET THEIR APPLICATIONS APPROVED AS THEIR CREDIT HISTORY MIGHT BE GOOD.

NOT/LESS RECOMMENDED GROUP/APPLICANTS

1. APPLICANTS NOT HAVING A HOUSE/FLAT ARE NOT FAVOURABLE.
2. APPLICANTS WITH LOWER SECONDARY EDUCATION ARE LESS FAVOURABLE.
3. APPLICANTS BELONGING TO LOW INCOME GROUP ARE LESS FAVOURABLE.
4. APPLICANTS BEING UNEMPLOYED OR IN MATERNITY LEAVE ARE LESS FAVOURABLE AS THEY HAVE GOOD AMOUNT OF DEFAULTER.
5. APPLICANTS OLDER IN AGE & HIGH EMPLOYEMENT IN YEARS ARE LESS FAVOURABLE TO GET THE LOAN.