



# NIKHIL KUMAR SINGH

## **EXPLORATORY DATA ANALYSIS**

# ASSIGNMENT



# PROBLEM STATEMENT:

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter.

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.

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 utilise this knowledge for its portfolio and risk assessment.

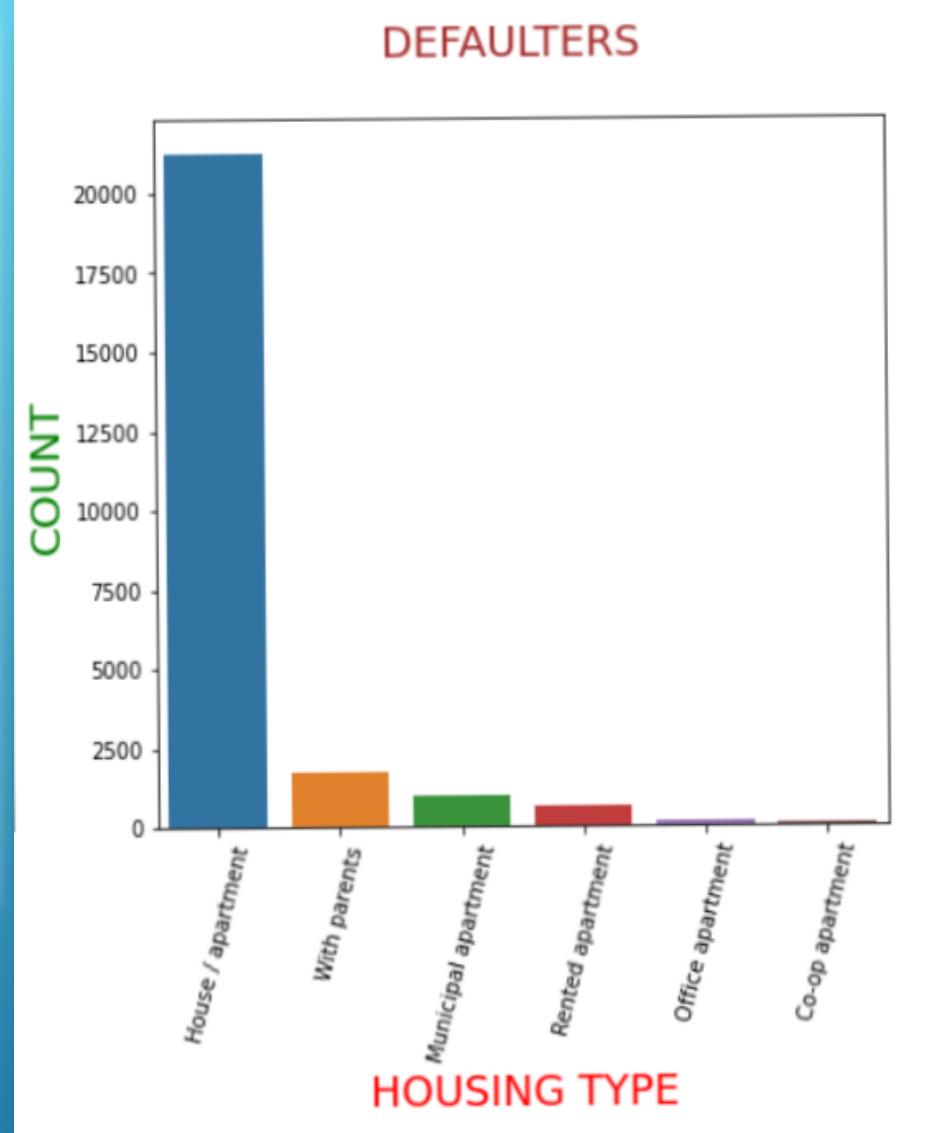
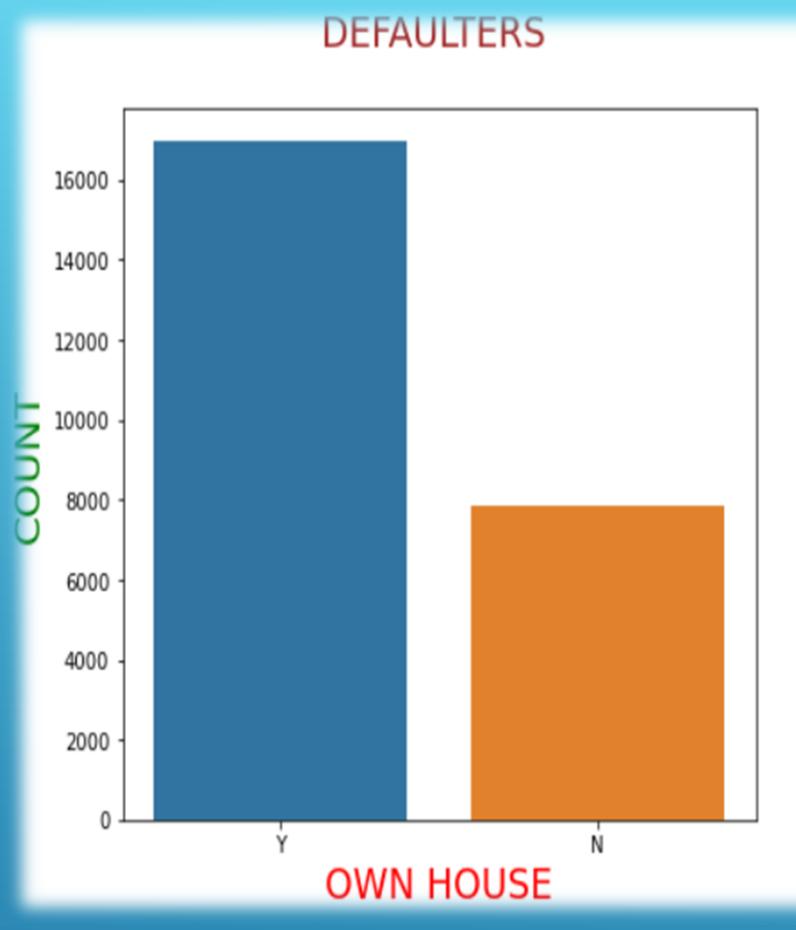
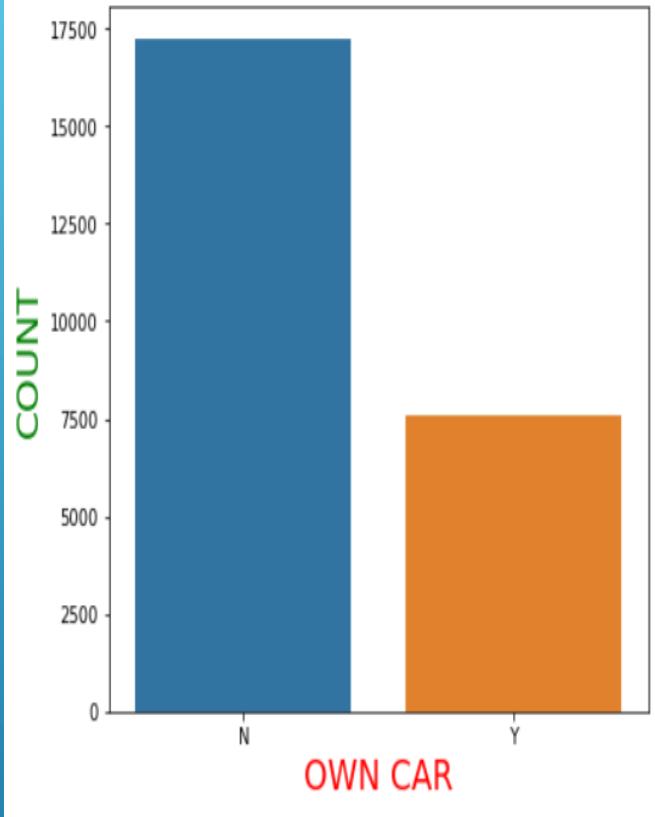
# OVERALL ANALYSIS APPROACH:

- ❖ UNDERSTANDING PROBLEM STATEMENT
- ❖ UNDERSTANDING DATA
- ❖ DATA IMPUTATION
- ❖ DATA CLEANING :-
  1. HANDLING & IMPUTING MISSING VALUES
  2. HANDLING ERRORS IN COLUMNS
- ❖ OUTLIERS ANALYSIS
- ❖ CHECKING DATA IMBALANCE
- ❖ DATA ANALYSIS :-
  1. UNIVARIATE ANALYSIS
  2. BIVARIATE ANALYSIS
  3. MULTIVARIATE ANALYSIS

# CURRENT APPLICATION DATA-SET

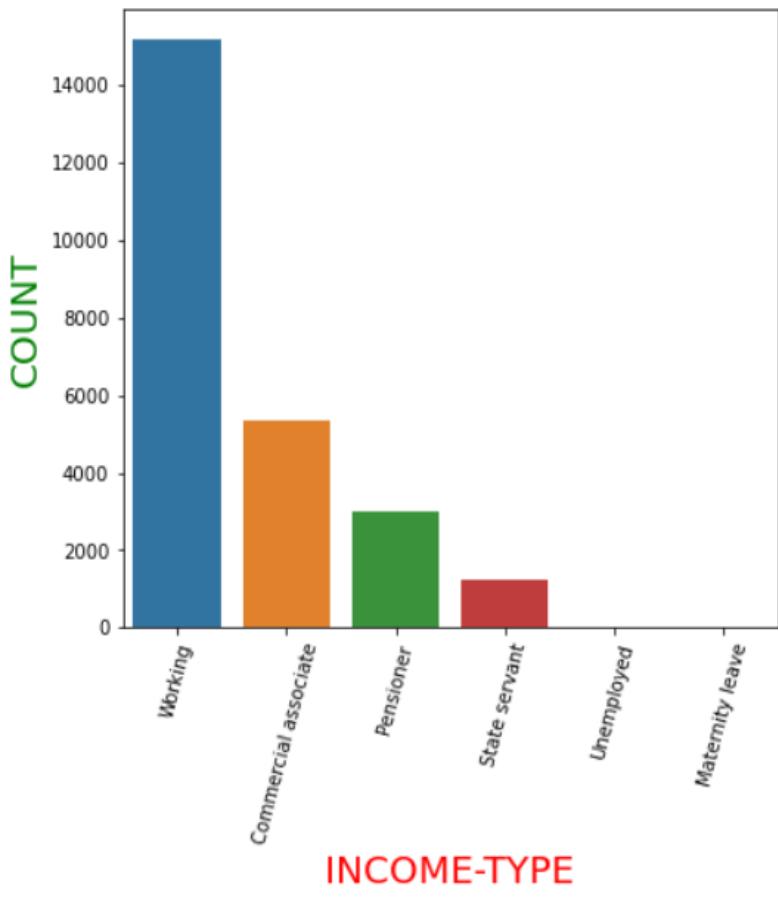


## DEFULTERS

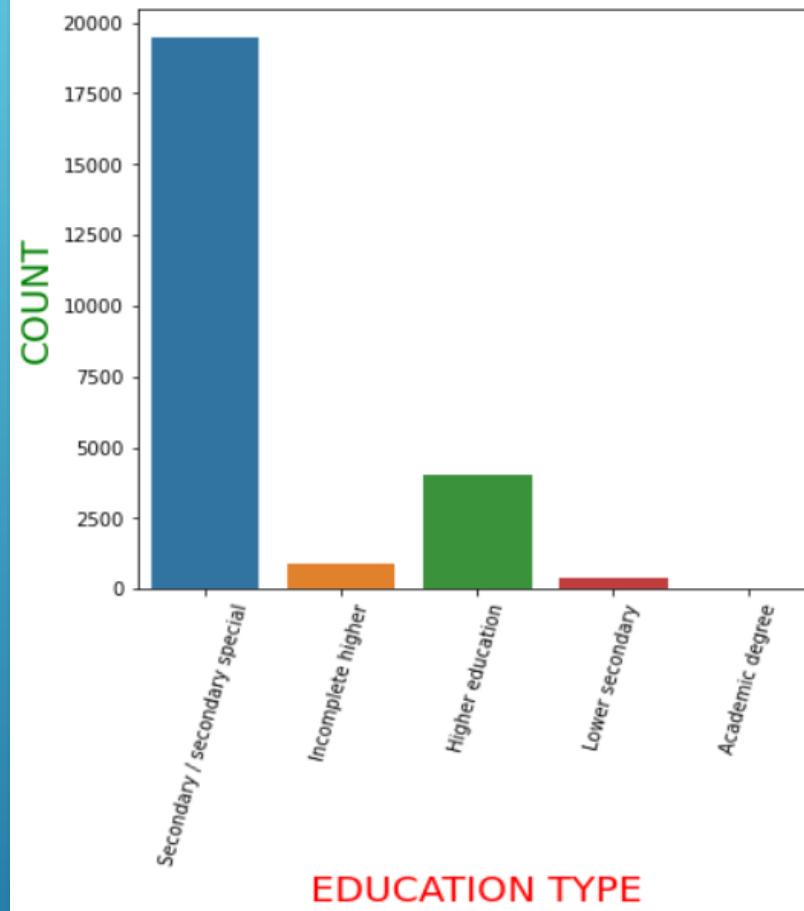


- MOST OF THE DEFULTERS DON'T HAVE THERE OWN CAR.
- MOST DEFULTERS HAVE THERE OWN HOUSE.
- MOST OF THE DEFULTERS LIVE IN HOUSE/APARTMENT.

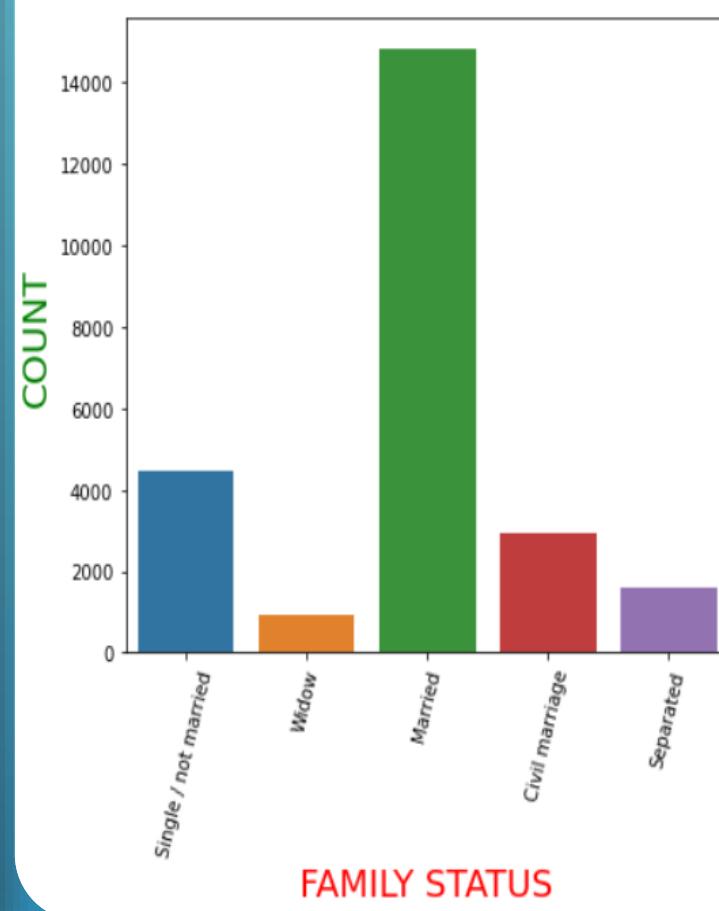
## DEFULTERS



## DEFULTERS

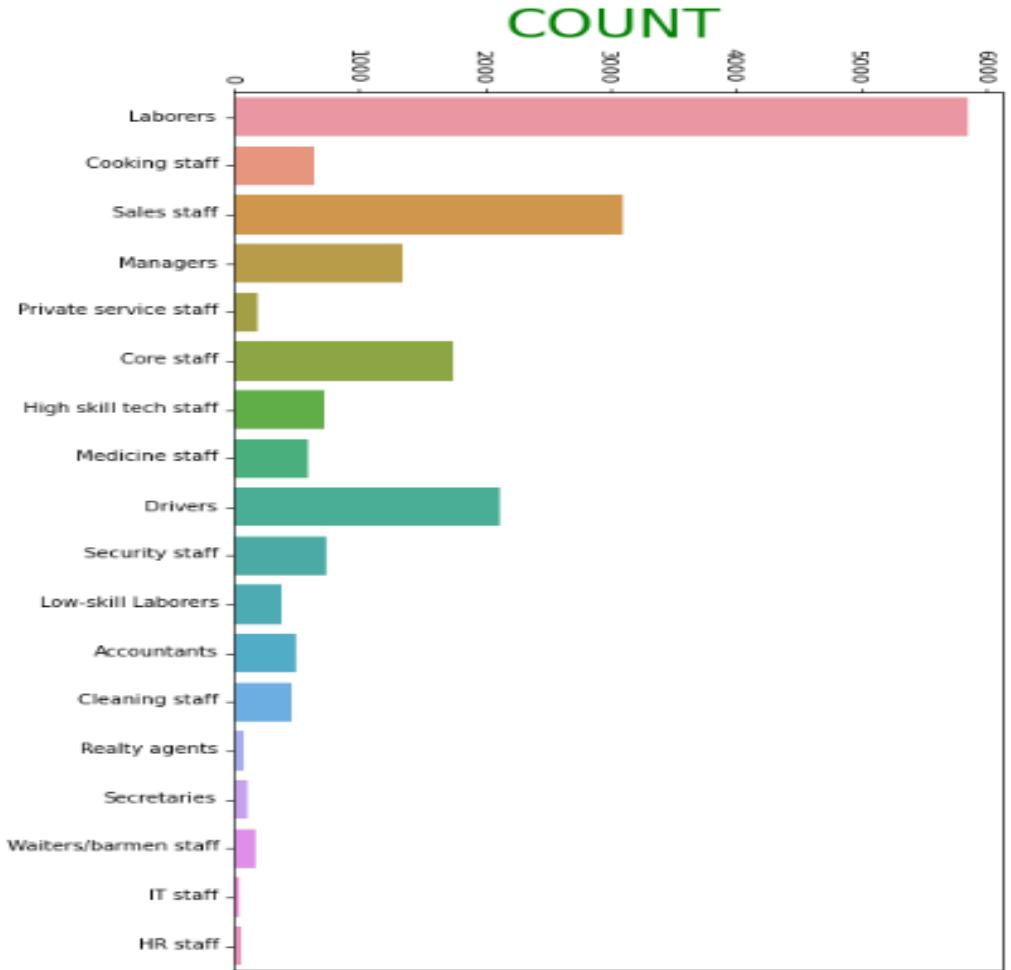


## DEFULTERS

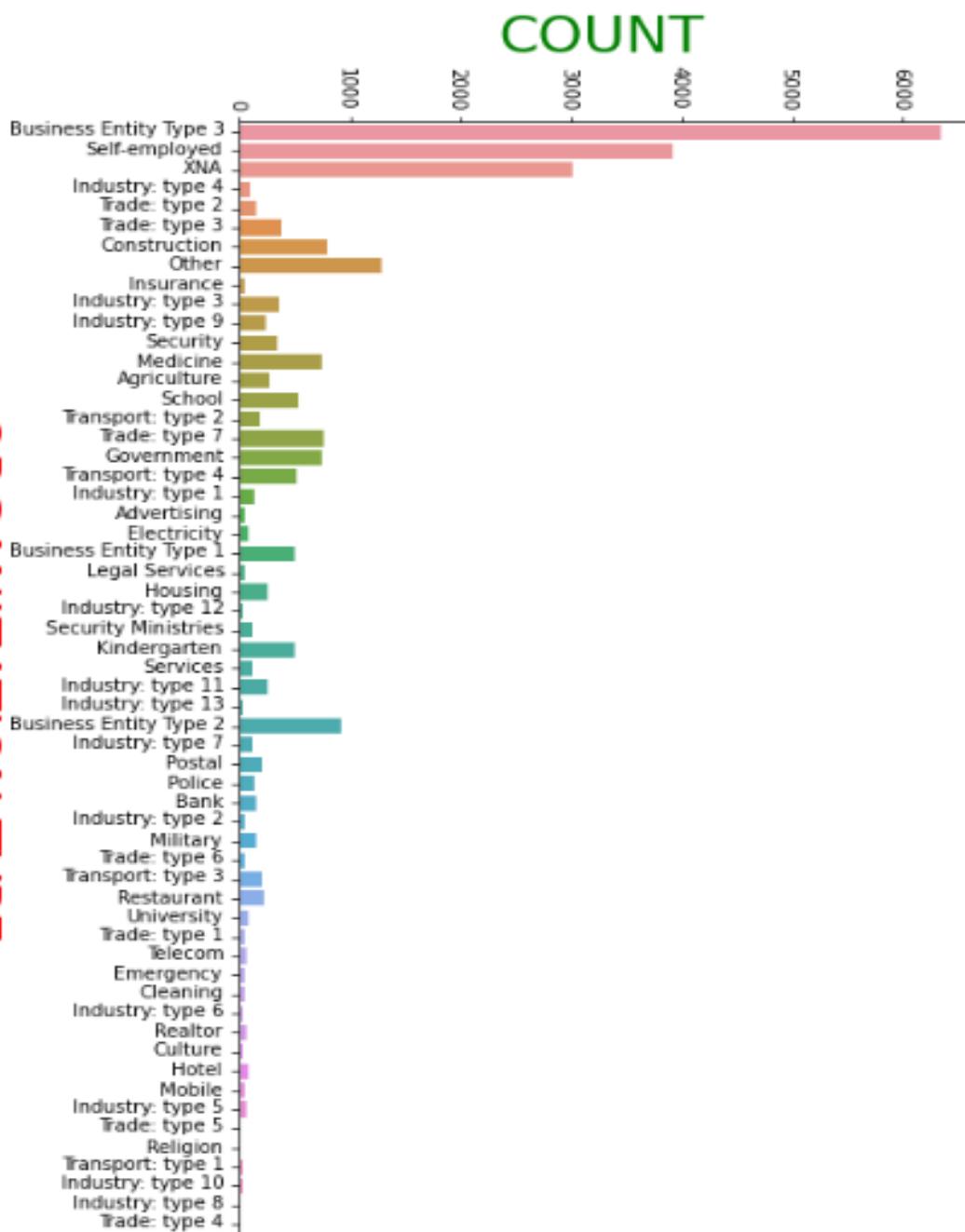


- MOST OF THE DEFULTERS ARE WORKING PEOPLE FOLLOWED BY COMMERCIAL ASSOCIATE.
- MOST DEFULTERS HAVE DONE THEIR SECONDARY/SECONDARY SPECIAL EDUCATION.
- MARRIED PEOPLE ARE THE MOST DEFULTERS IN ALL FAMILY STATUS FOLLOWED BY SINGLE PEOPLE.

## OCCUPATION TYPE



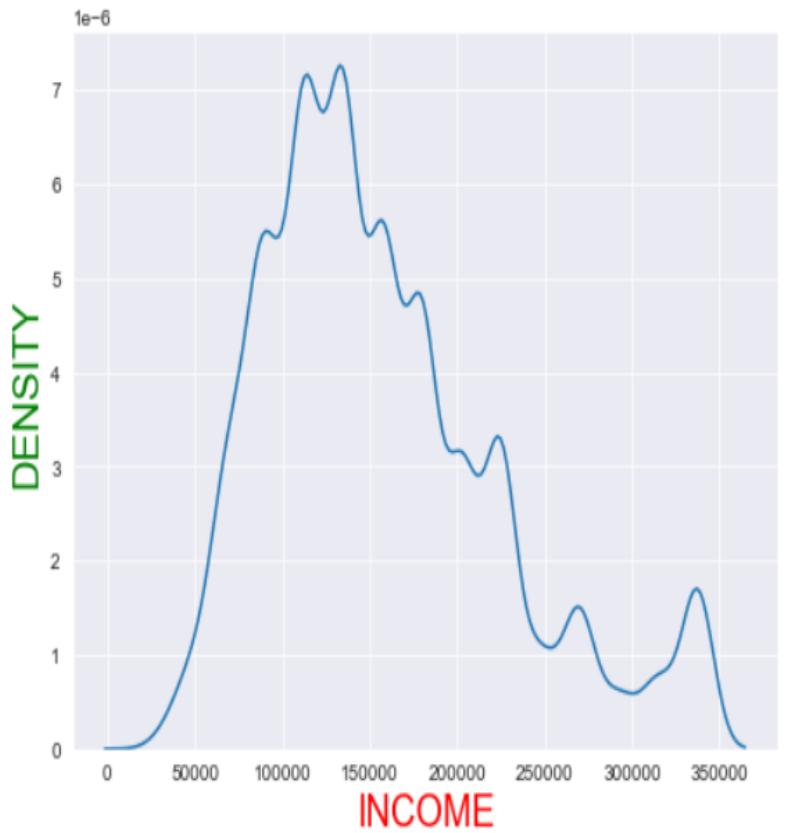
## DEFAULTERS



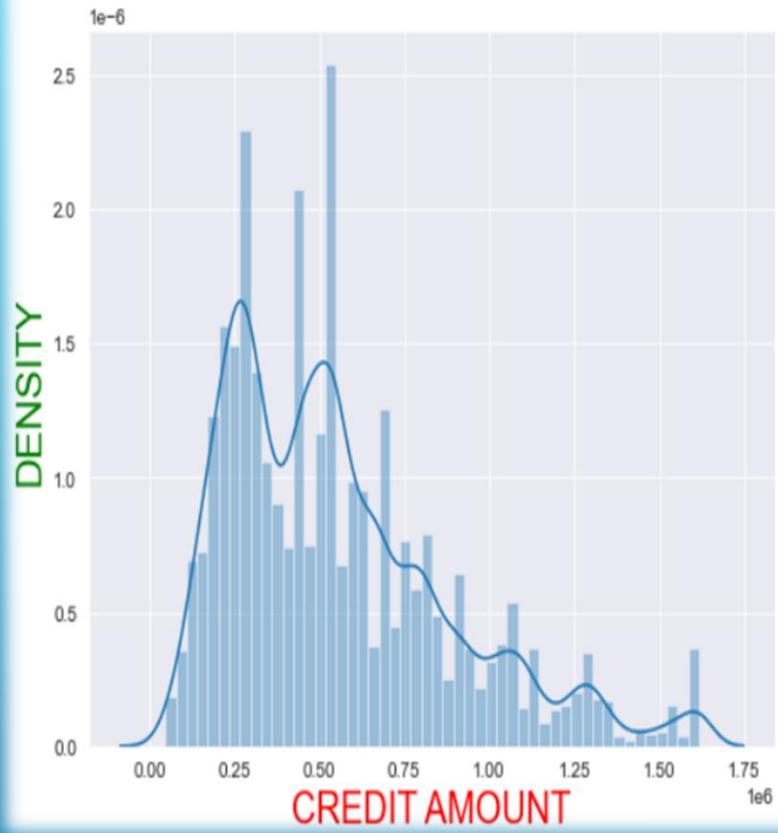
## DEFAULTERS

- LABORERS ARE THE MOST DEFAULTERS FOR ALL THE OCCUPATION TYPE.
- MOST DEFAULTERS ARE WORKING IN BUSINESS ENTITY TYPE 3.

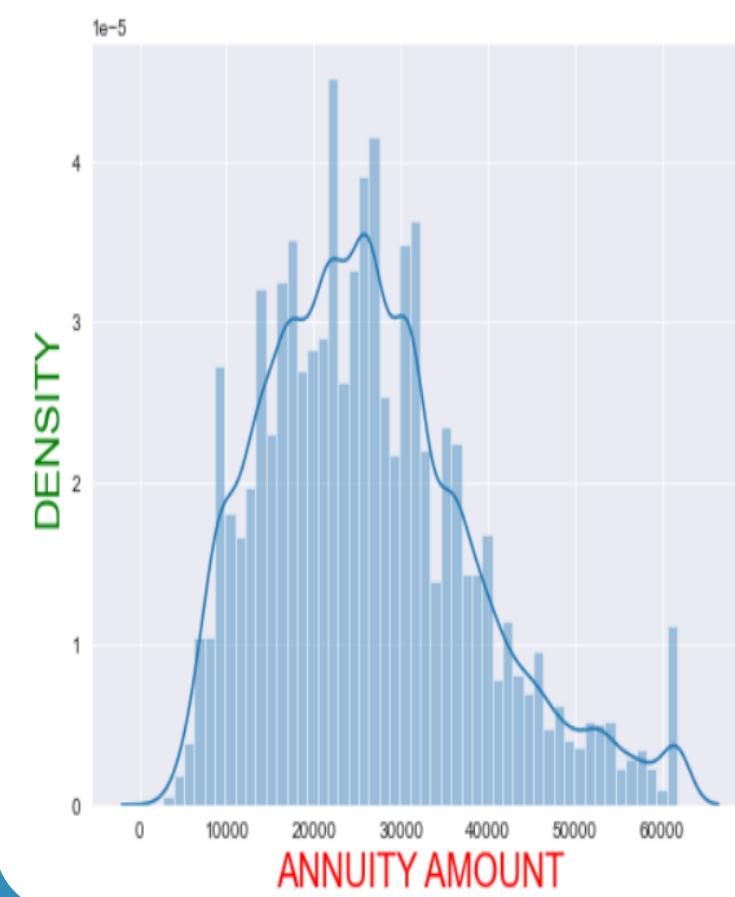
## DEFULTERS



## DEFULTERS

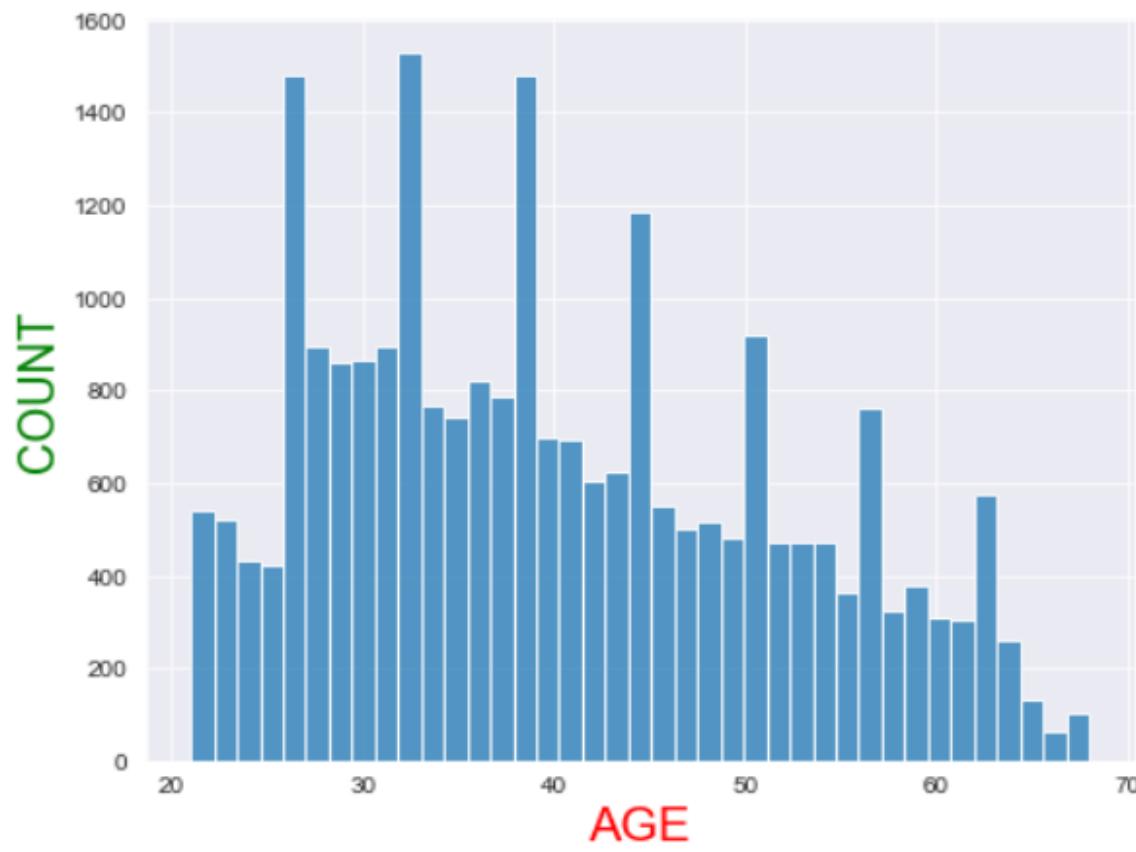


## DEFULTERS

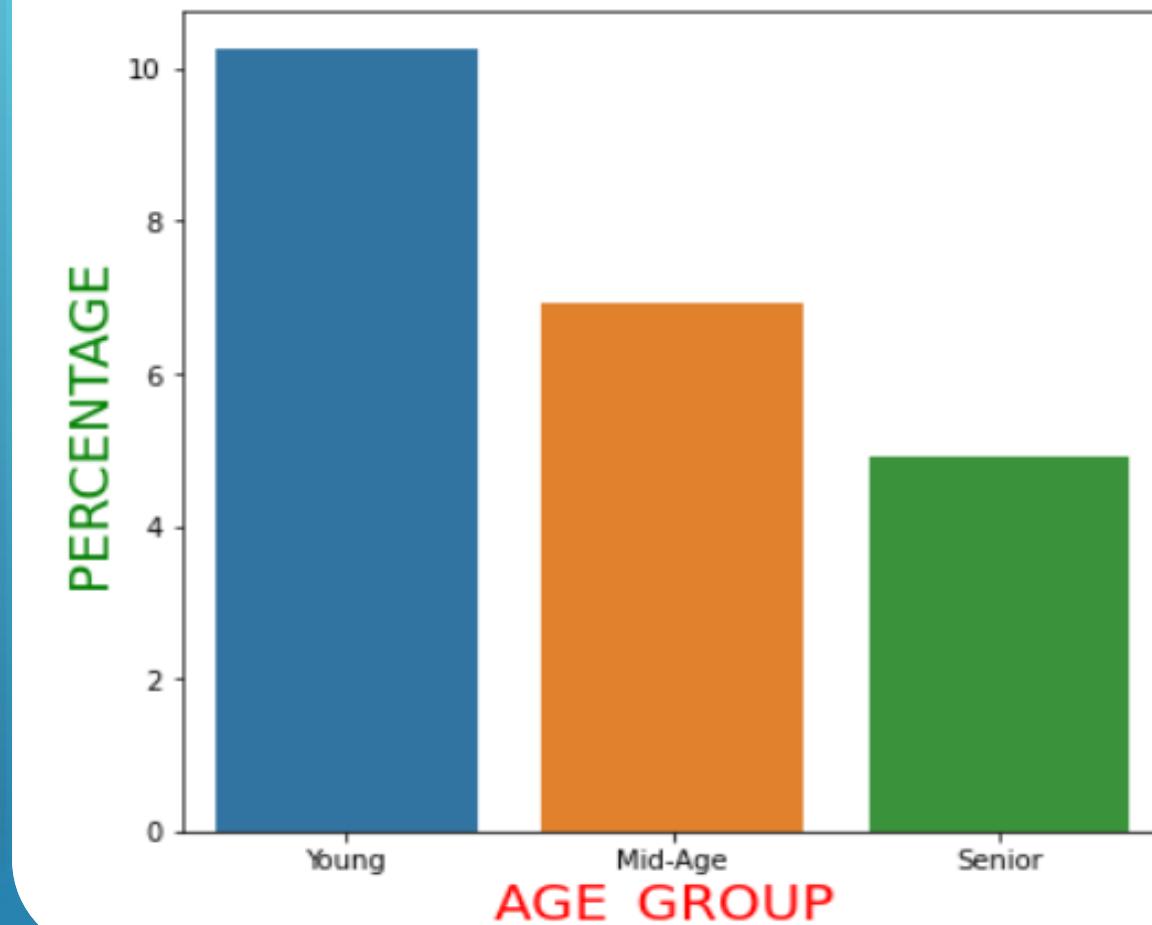


- MOST OF THE DEFULTERS ARE EARNING BETWEEN 1LAKHS TO 1.6LAKHS ANNUALLY.
- THE LESSER THE AMOUNT OF CREDIT LOAN MORE THE CHANCES OF BEING DEFULTERS.
- DEFULTERS ARE HAVING LOAN ANNUITY BETWEEN 10000 TO 30000.

## DEFALTERS

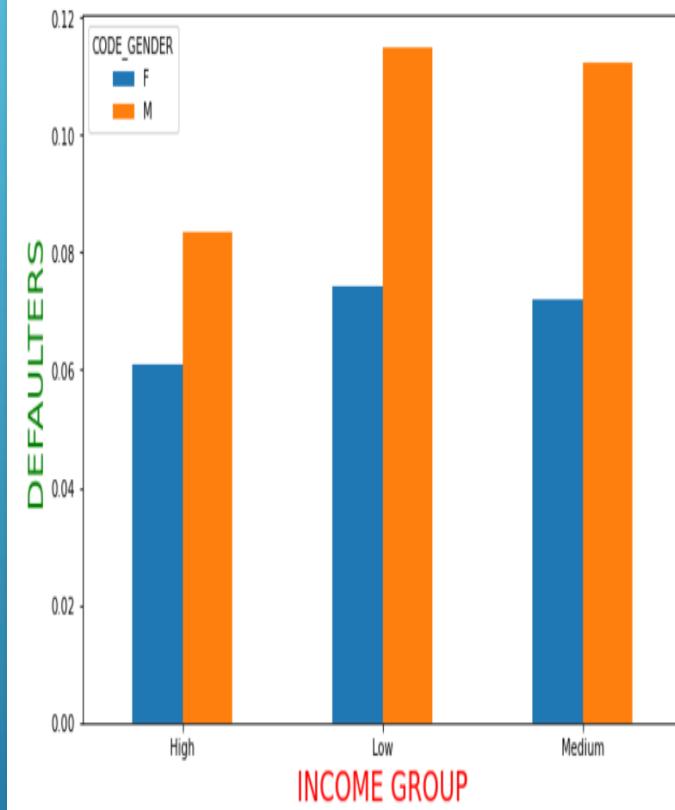


## DEFALTERS

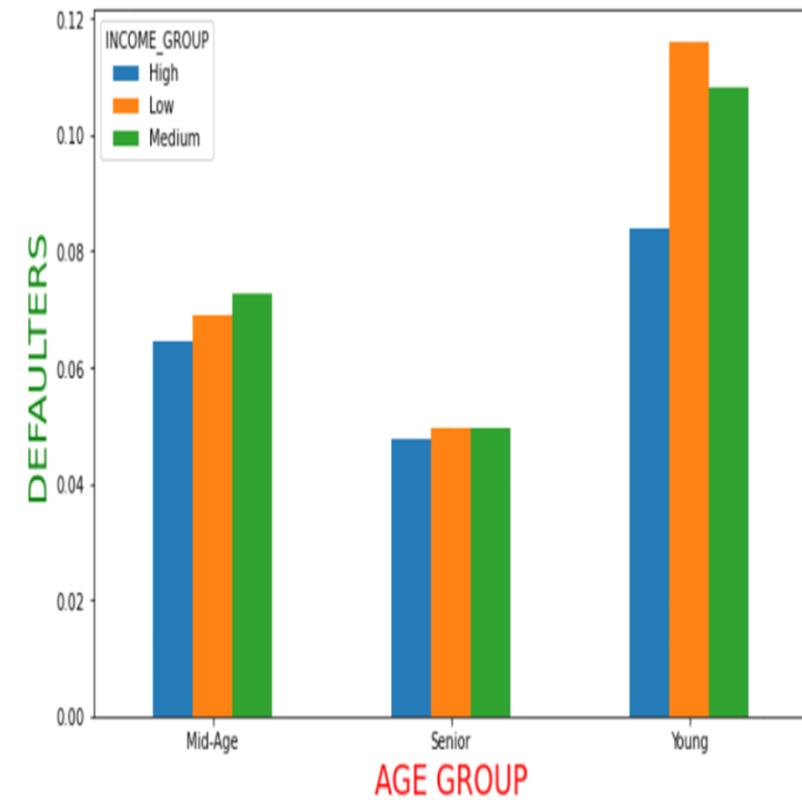


- DEFALTERS AGE MOSTLY LIES BETWEEN 20-25 YEARS TO 40-45 YEARS.
- YOUNG PEOPLE ARE MORE LIKELY TO BE DEFALTERS THAN OTHER's.

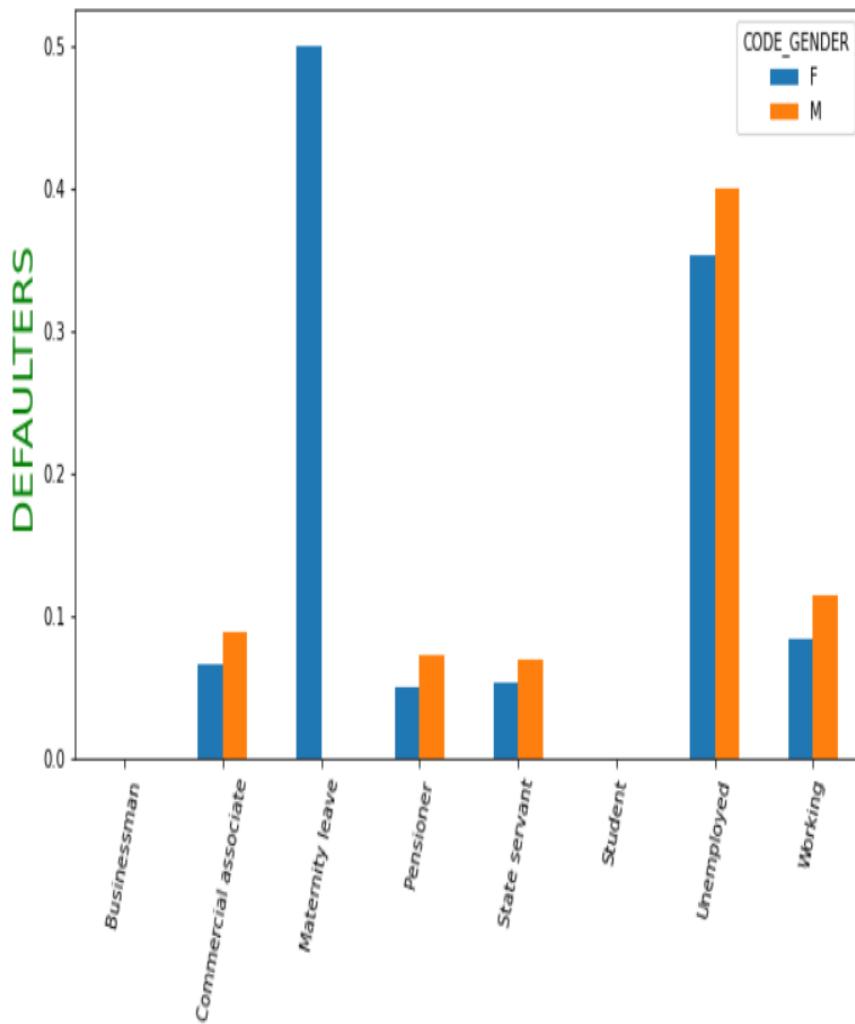
INCOME-GROUP vs GENDER w.r.t. TARGET



AGE-GROUP vs INCOME-GROUP w.r.t. TARGET

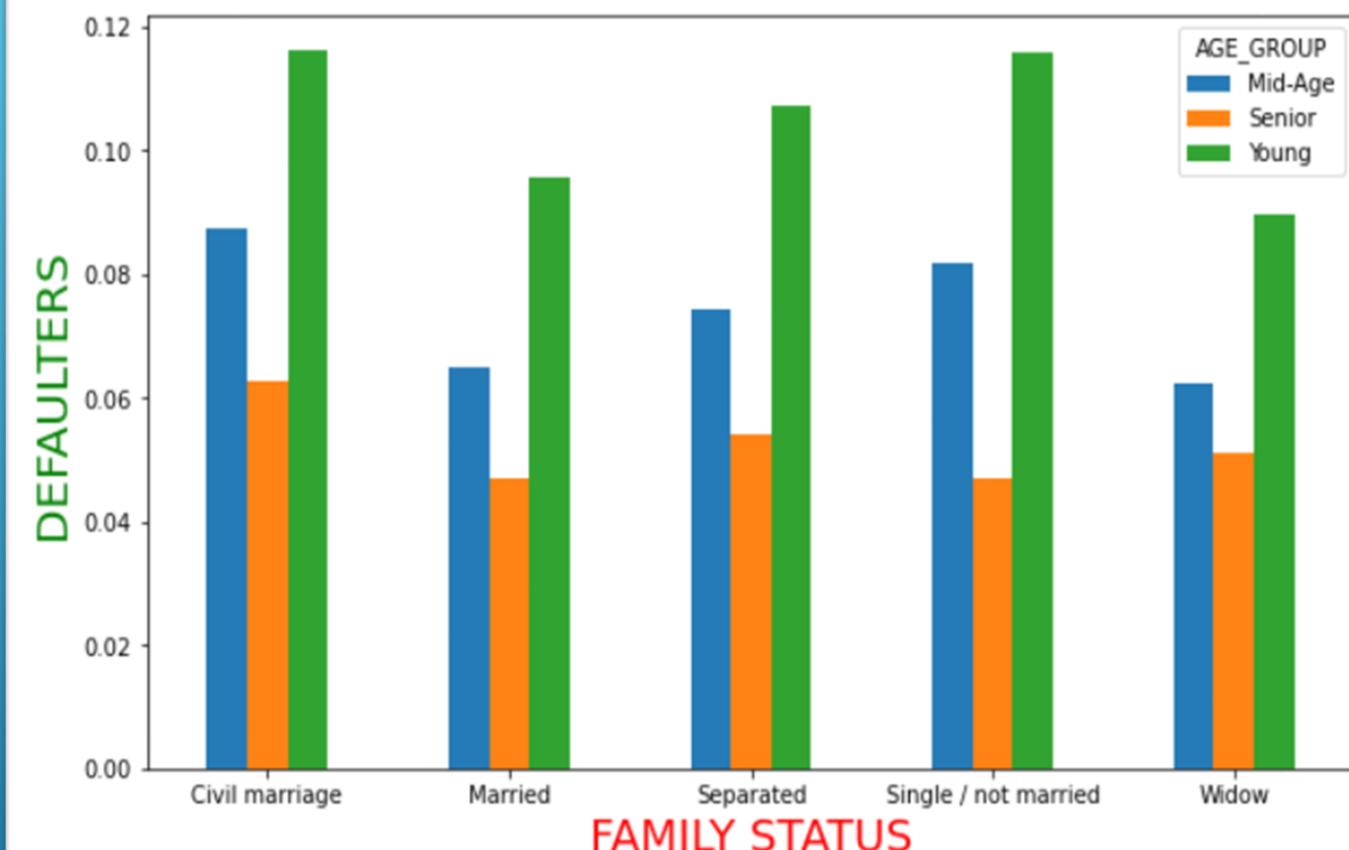


INCOME-TYPE vs GENDER w.r.t. TARGET

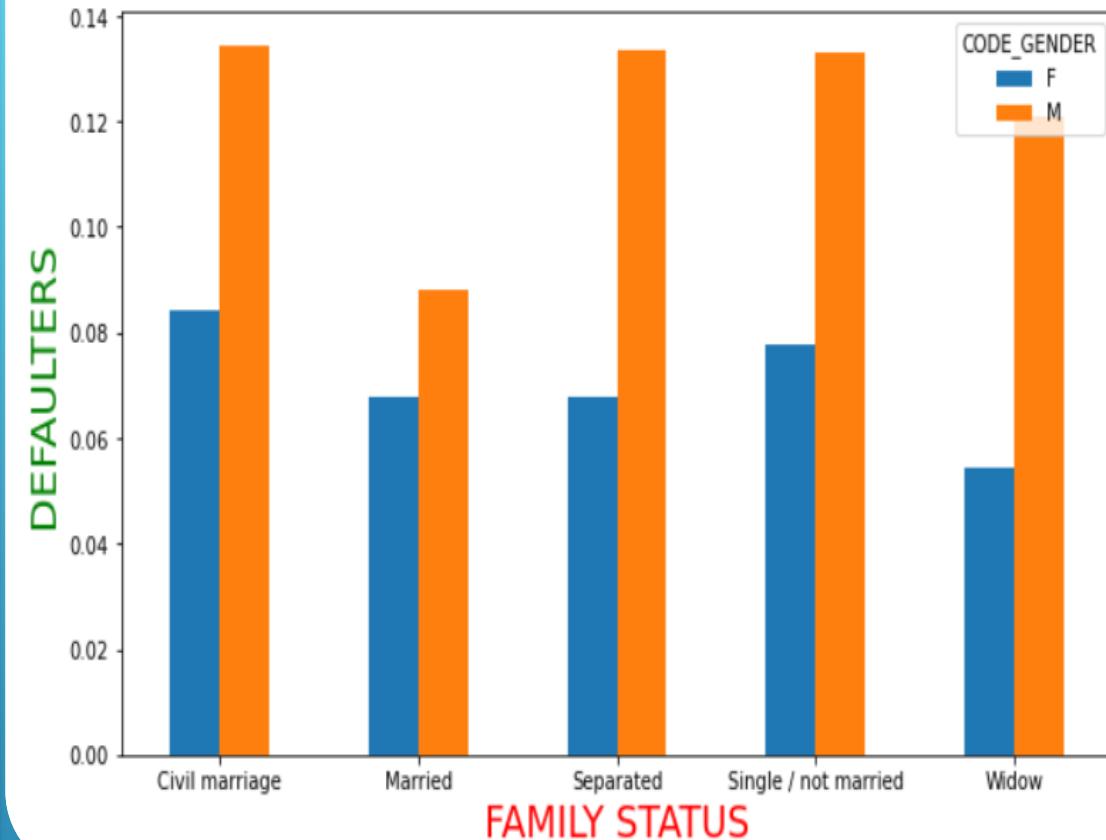


- MALES ARE MORE LIKELY TO BE DEFAULTERS THAN FEMALES IN ALL INCOME GROUP.
- IN YOUNG AGE GROUP LOW INCOME PEOPLE ARE MORE DEFAULTERS THAN OTHER's.
- MALES( ESPECIALLY UNEMPLOYED ) ARE MORE DEFAULTERS THAN FEMALES IN ALL THE PROFESSION EXCEPT MATERNITY LEAVE.

## FAMILY STATUS vs AGE-GROUP w.r.t. TARGET

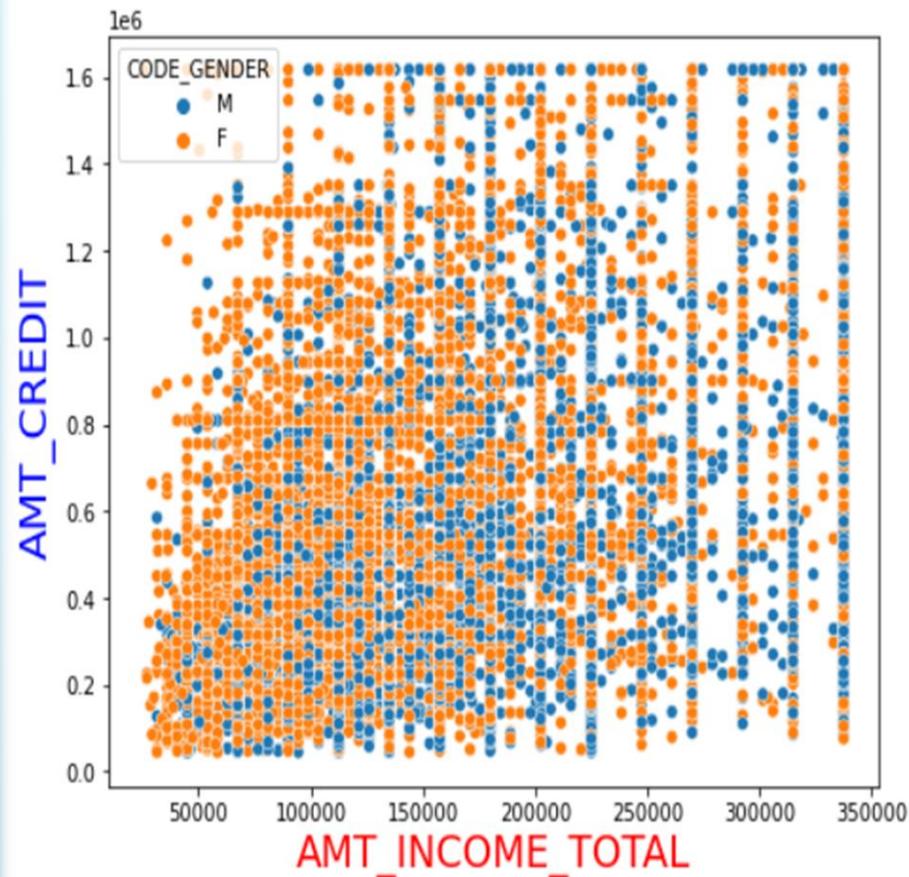


## FAMILY STATUS vs GENDER w.r.t. TARGET

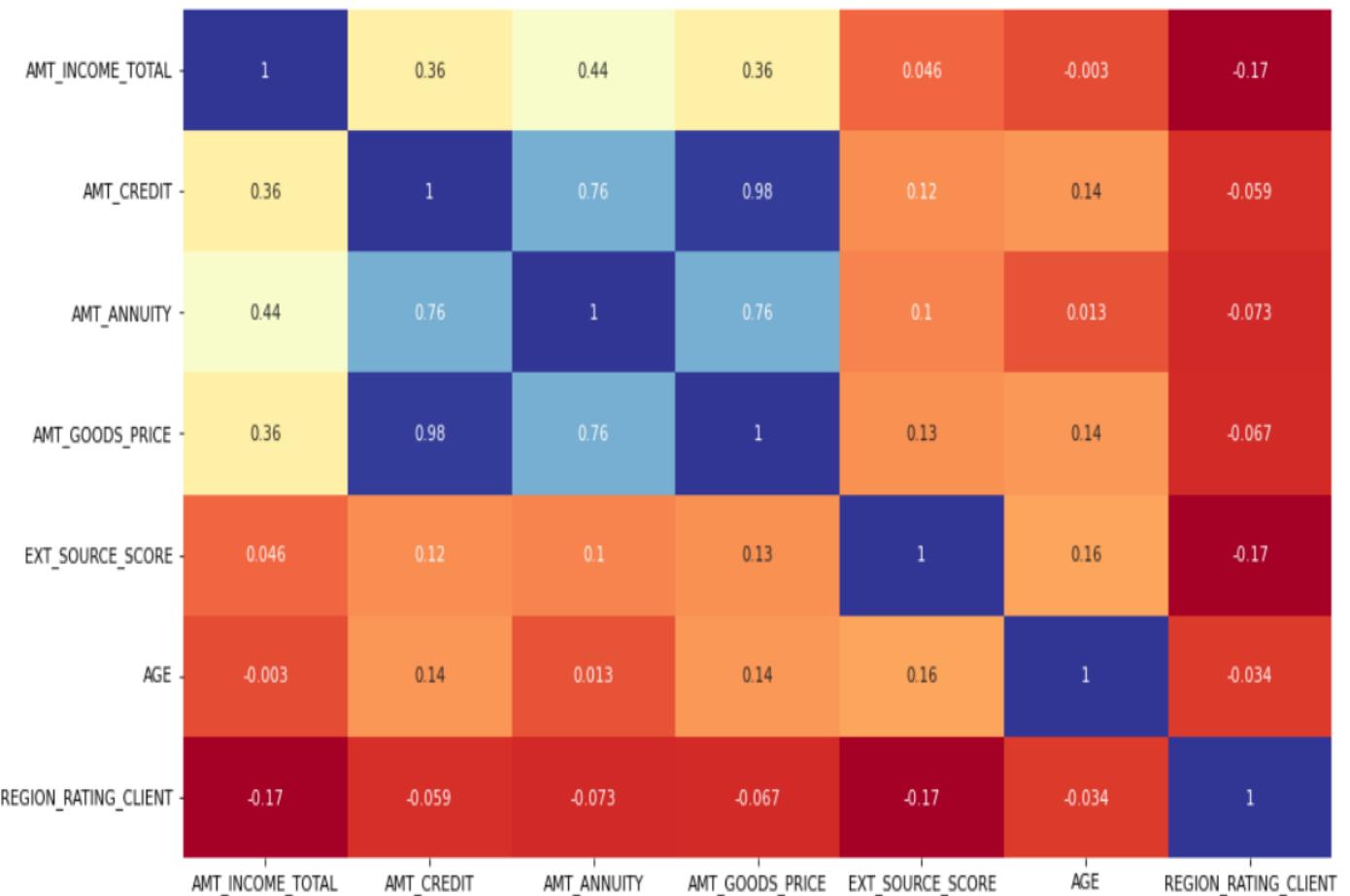


- YOUNG CLIENT's ARE MORE DEFAULTERS THAN OTHER's IN ALL THE FAMILY STATUS TYPE.
- SENIOR CITIZEN's ARE LESS DEFAULTERS.
- MALES ARE MORE DEFAULTERS THAN FEMALES IN ALL THE FAMILY STATUS TYPE.

## DEFALUTERS



## DEFALUTERS



- AS THE INCOME INCREASES AMOUNT OF CREDIT LOAN ALSO INCREASES FOR BOTH GENDERS.
- VERY HIGH CORRELATION BETWEEN AMT\_CREDIT & AMT\_GOODS\_PRICE (0.98%) MEANS :-  
APPLICANTS OWNING GOODS OF HIGH VALUE CAN TAKE LOANS OF HIGHER AMOUNTS.

# DATA-FRAME MERGED

APPLICATION DATA-SET &  
PREVIOUS APPLICATION  
DATA-SET

### NAME\_CONTRACT\_STATUS

COUNT

800000  
600000  
400000  
200000  
0

Approved Refused Unused offer Canceled

LOAN STATUS

### NAME\_CLIENT\_TYPE

COUNT

800000  
700000  
600000  
500000  
400000  
300000  
200000  
100000  
0

New Repeater Refreshed

CLIENT TYPE

### NAME\_CONTRACT\_TYPE

COUNT

600000  
500000  
400000  
300000  
200000  
100000  
0

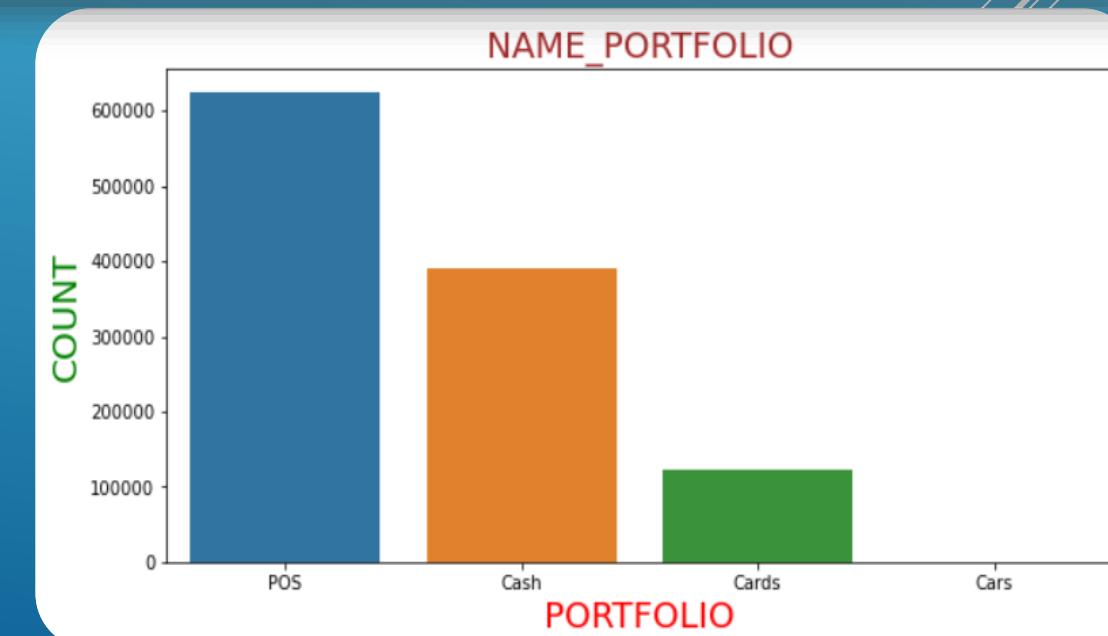
Consumer loans Cash loans Revolving loans

LOAN TYPES

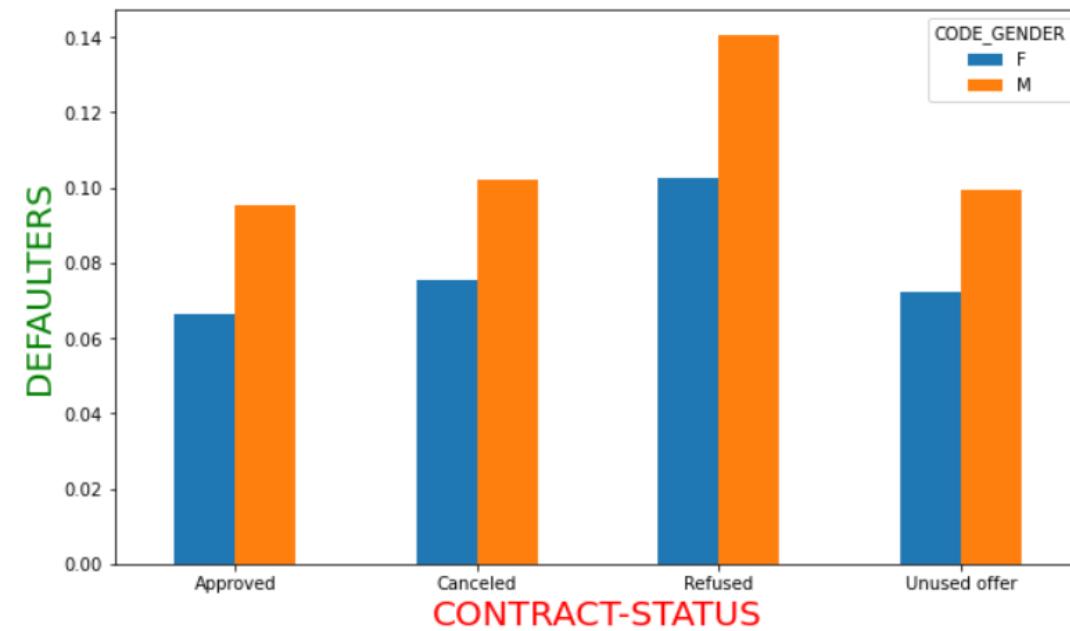
- APPROVED LOANS ARE THE MOST FOLLOWED BY REFUSED LOANS.
- MOSTLY THE APPLICANT's WERE REPEATER FOLLOWED BY NEW ONE's.
- CONSUMER LOANS ARE THE MOST FOLLOWED BY CASH LOANS.



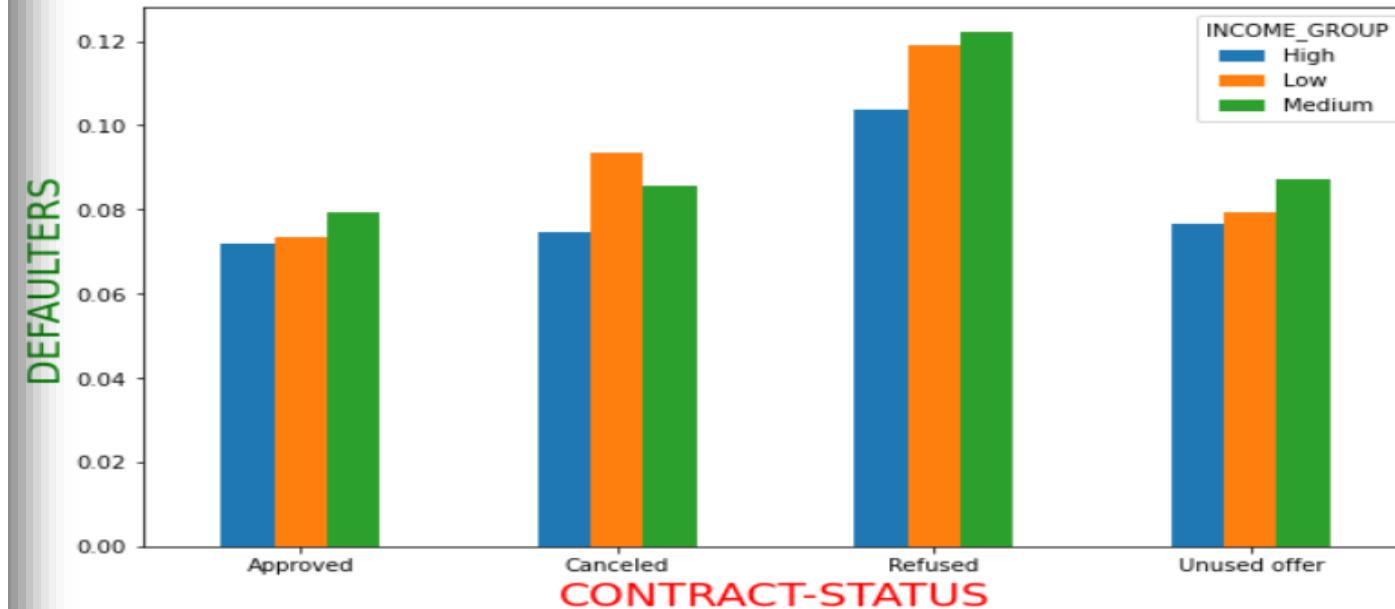
- SECONDARY/SECONDARY SPECIAL ARE VERY HUGE IN NUMBER FOLLOWED BY HIGHER EDUCATION.
- COUNTRY-WIDE WAS HEAVILY USED FOLLOWED BY CREDIT & CASH OFFICES, STONE, REGIONAL/LOCAL.
- HIGHEST NUMBER WAS FOR POS FOLLOWED BY CASH FOLLOWED BY CARDS.



CONTRACT-STATUS vs GENDER w.r.t. TARGET

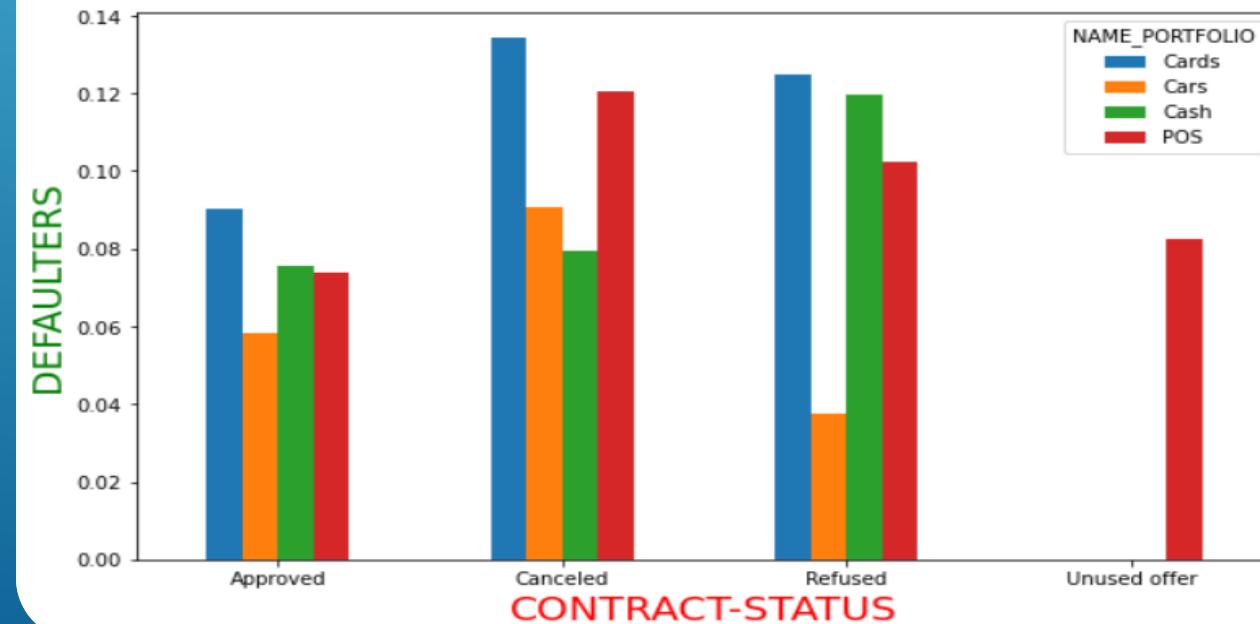


CONTRACT-STATUS vs INCOME-GROUP w.r.t. TARGET

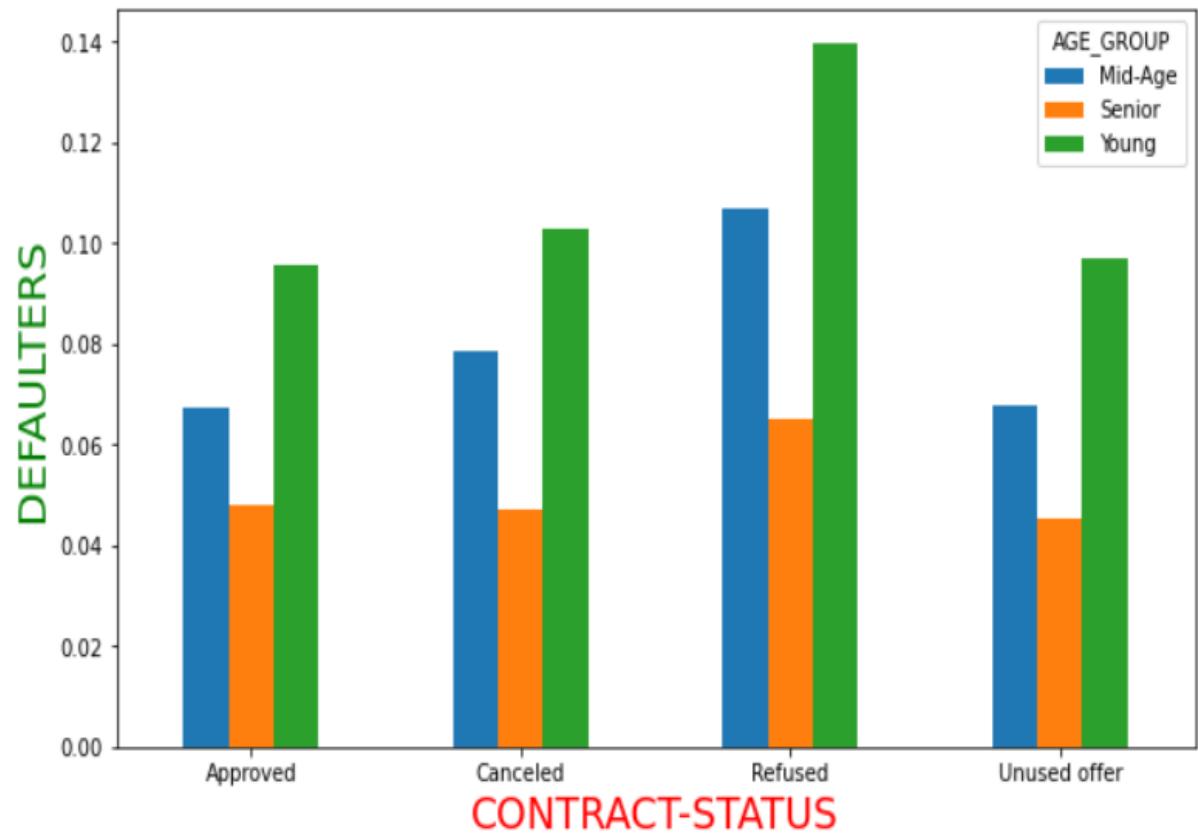


- REFUSED CLIENT's & MALES ARE THE MOST DEFAULTERS FOR ALL THE LOAN STATUS.
- ALL THE INCOME GROUPS ARE ALMOST EQUALLY DEFAULTERS FOR ALL THE LOAN STATUS ESPECIALLY REFUSED ONE's.
- PEOPLE WHO PREVIOUSLY APPLIED LOAN FOR CARDS ARE MORE DEFAULTERS THAN OTHER's.

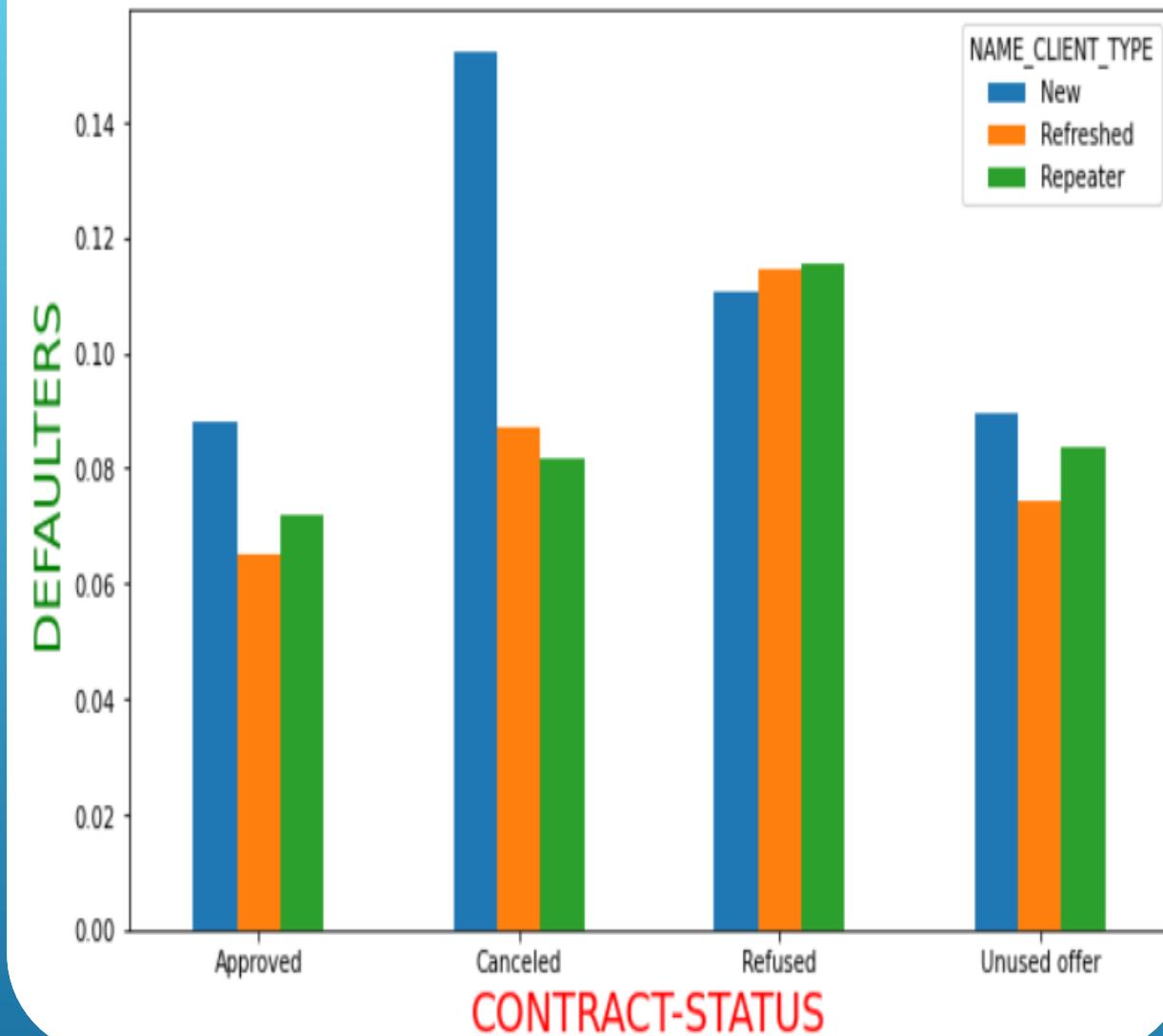
CONTRACT-STATUS vs PORTFOLIO w.r.t. TARGET



## CONTRACT-STATUS vs AGE-GROUP w.r.t. TARGET

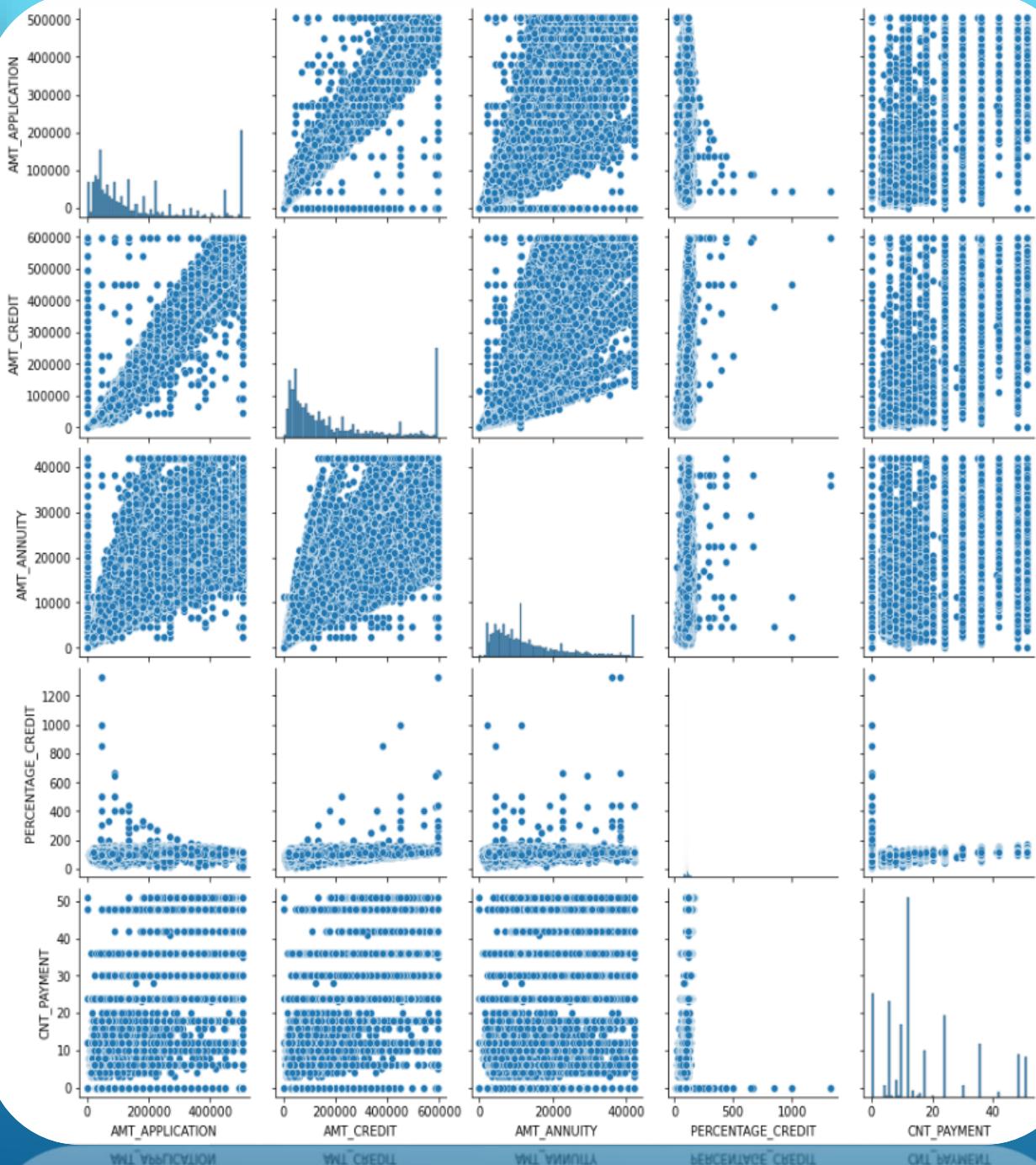
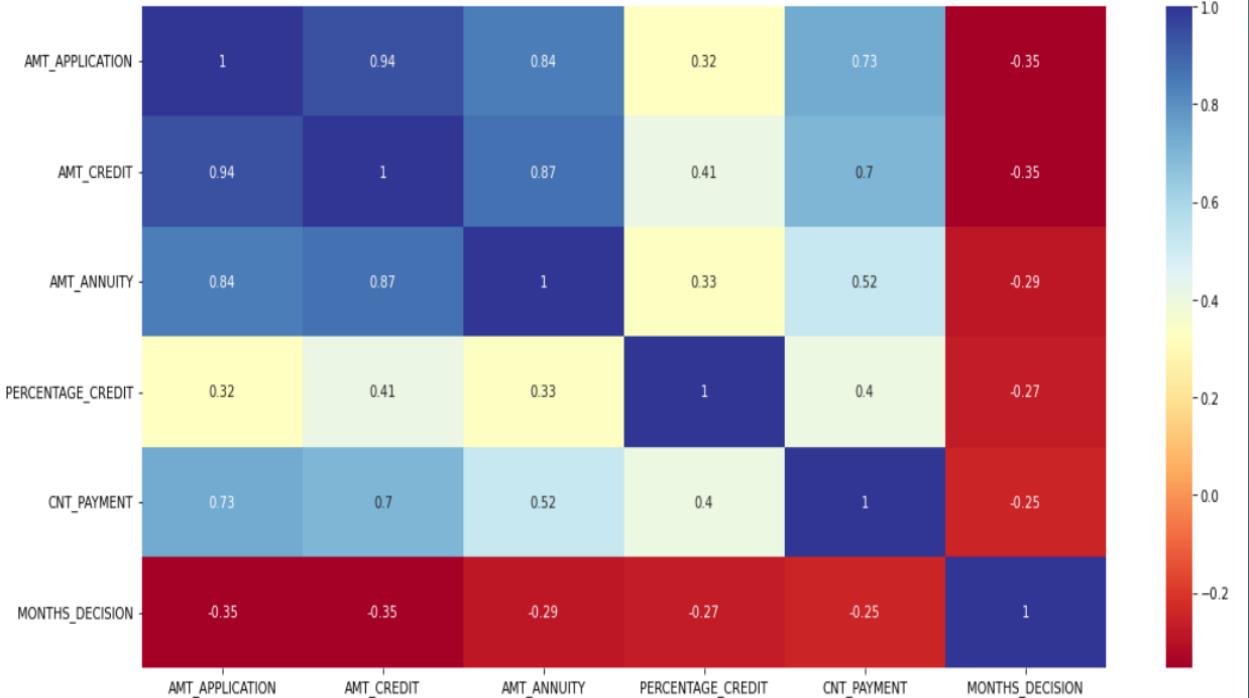


## CONTRACT-STATUS vs CLIENT-TYPE w.r.t. TARGET



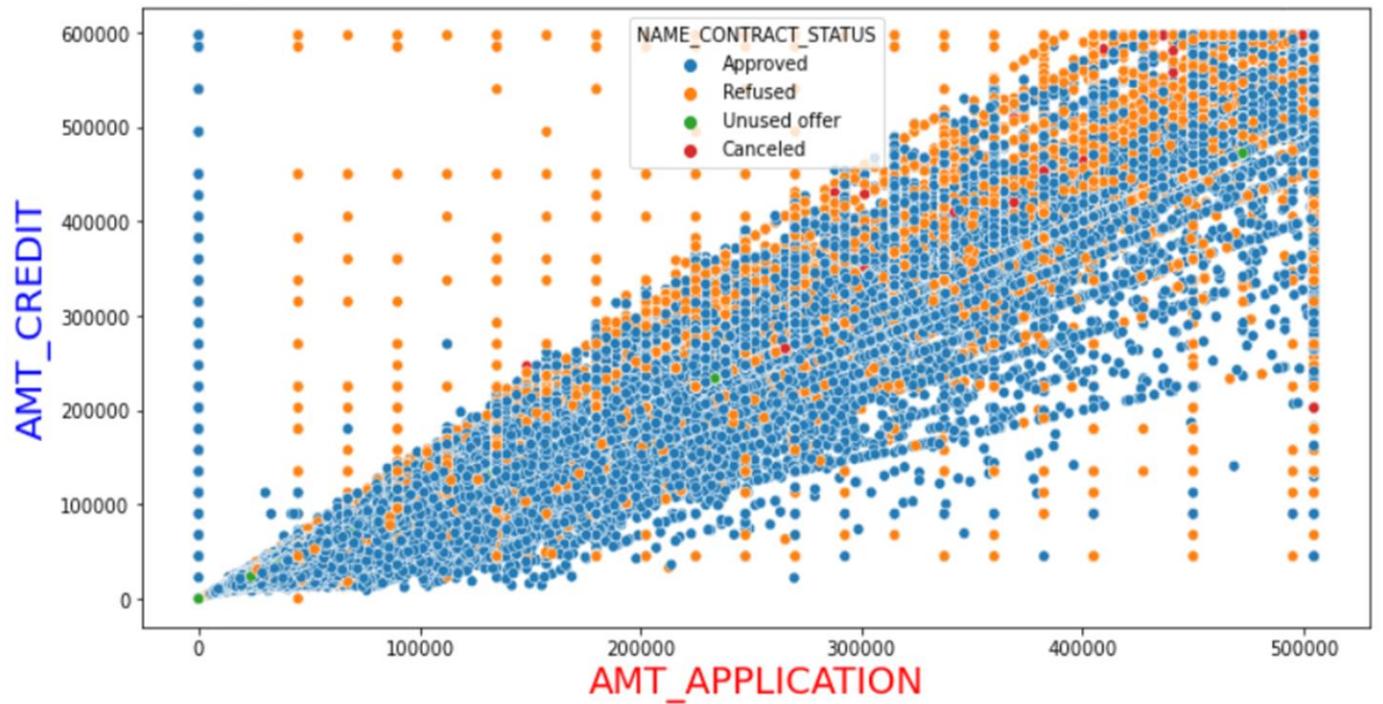
- YOUNG PEOPLE ARE THE MOST DEFAULTERS FOR ALL THE PREVIOUS LOAN STATUS.
- SENIOR CITIZEN ARE LESS DEFAULTERS.
- DEFAULTERS ARE MORE FOR PREVIOUSLY CANCELED LOAN STATUS WHO ARE NEW CLIENT's.

## DEFAULTERS

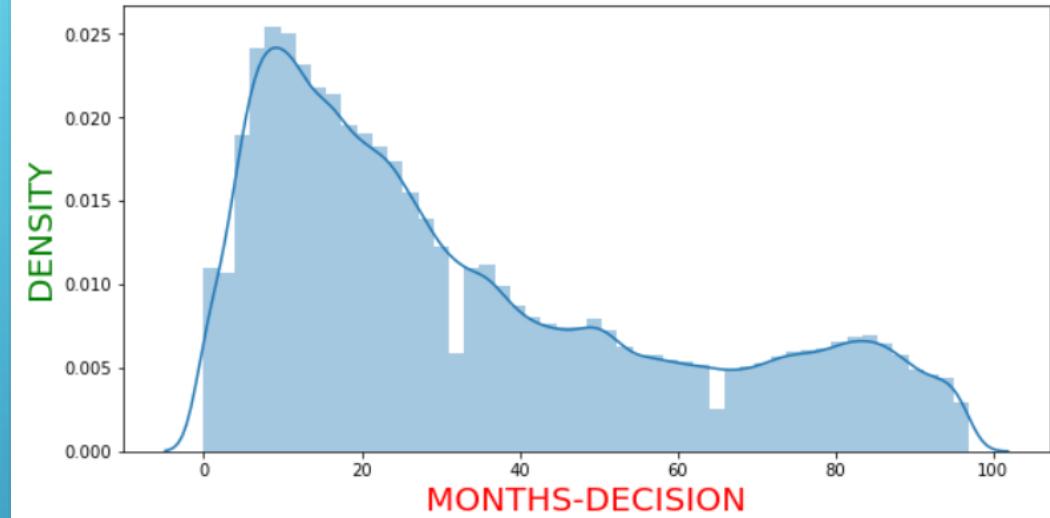


- VERY HIGH CORRELATION BETWEEN SOME VARIABLES MEANS :- AS ONE VARIABLE INCREASES THE OTHER VARIABLE ALSO INCREASES W.R.T. IT.
- VARIABLES ARE :- AMT\_APPLICATION & AMT\_CREDIT(0.94%) , AMT\_APPLICATION & AMT\_ANNUITY(0.84%) , AMT\_CREDIT & AMT\_ANNUITY(0.87%).

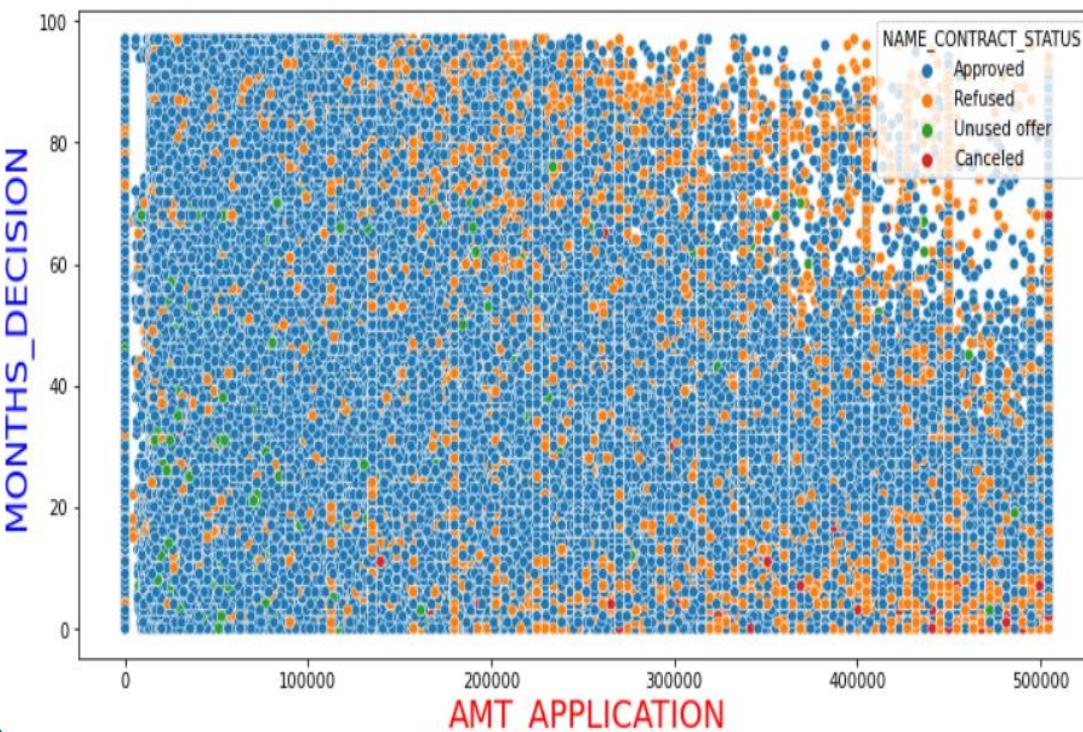
### APPLICATION vs LOAN CREDIT



### MONTHS-TAKEN



### APPLICATION vs MONTHS



- CREDITED AMOUNT INCREASES WITH RESPECT TO APPLICATION AMOUNT.
- MOST OF DECISION TAKEN BETWEEN 15 TO 30 MONTHS BUT SPREAD IS UPTO 100 MONTHS.
- THE MORE THE APPLICATION AMOUNT OF LOAN THE LESS THE MONTHS TAKEN FOR MAKING DECISIONS.

# **CONCLUSION & RECOMMENDATION**

## **RECOMMENDED GROUP :**

- SENIOR CITIZEN's OF ALL CATEGORY.
- MALES WITH ACADEMIC DEGREE.
- CLIENT's WITH HIGH INCOME GROUP.
- MARRIED PEOPLE.
- CLIENT WITH MEDIUM & HIGH EXTERNAL SOURCE SCORE.
- HIGH SKILL TECH STAFF, MANAGERS, SALES STAFF.
- FEMALES ARE LESS DEFAULTERS THAN MALES.
- ANY CLIENT WHOSE PREVIOUS LOAN WAS APPROVED.
- APPROVED CLIENT's IN THEIR PREVIOUS APPLICATIONS.

## **NOT RECOMMENDED GROUP :**

- YOUNG CLIENT's OF ALL CATEGORY.
- CLIENT's WITH LOW INCOME GROUP.
- UNEMPLOYED PEOPLE.
- MALES ARE MORE DEFAULTERS THAN FEMALES.
- LABORERS, DRIVERS, CORE STAFF.
- PEOPLE WORKING IN BUSINESS ENTITY TYPE 3 & SELF-EMPLOYED PEOPLE.
- FEMALES WHO ARE AT MATERNITY LEAVE.
- APPLICANT's WHO ARE AT SECONDARY/SECONDARY SPECIAL.
- PREVIOUSLY REFUSED CLIENT's.
- POOR EXTERNAL SOURCE SCORER.