CREDIT EDA CASE STUDY

BANK CUSTOMERS ATTRIBUTES ANALYSIS

KEERTHI P

CREDIT EDA CASE STUDY

INTRODUCTION

   This project is about understanding the factors which drives a customer of a bank to become a defaulter. Here,all the available attributes of the loan applicant is analysed to conclude whether he/she is credible enough to be given a loan.The major deciding criteria is based on two factors.

▶ 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.

APPROACH

- DATA CLEANING
- CHECKING NULL VALUES
- IMPUTING MISSING VALUES
- HANDLING OUTLIERS
- UNIVARIATE ANALYSIS
- BIVARIATE ANALYSIS
- MULTIVARIATE ANALYSIS
- CORRELATION
- DATA IMBALANCE
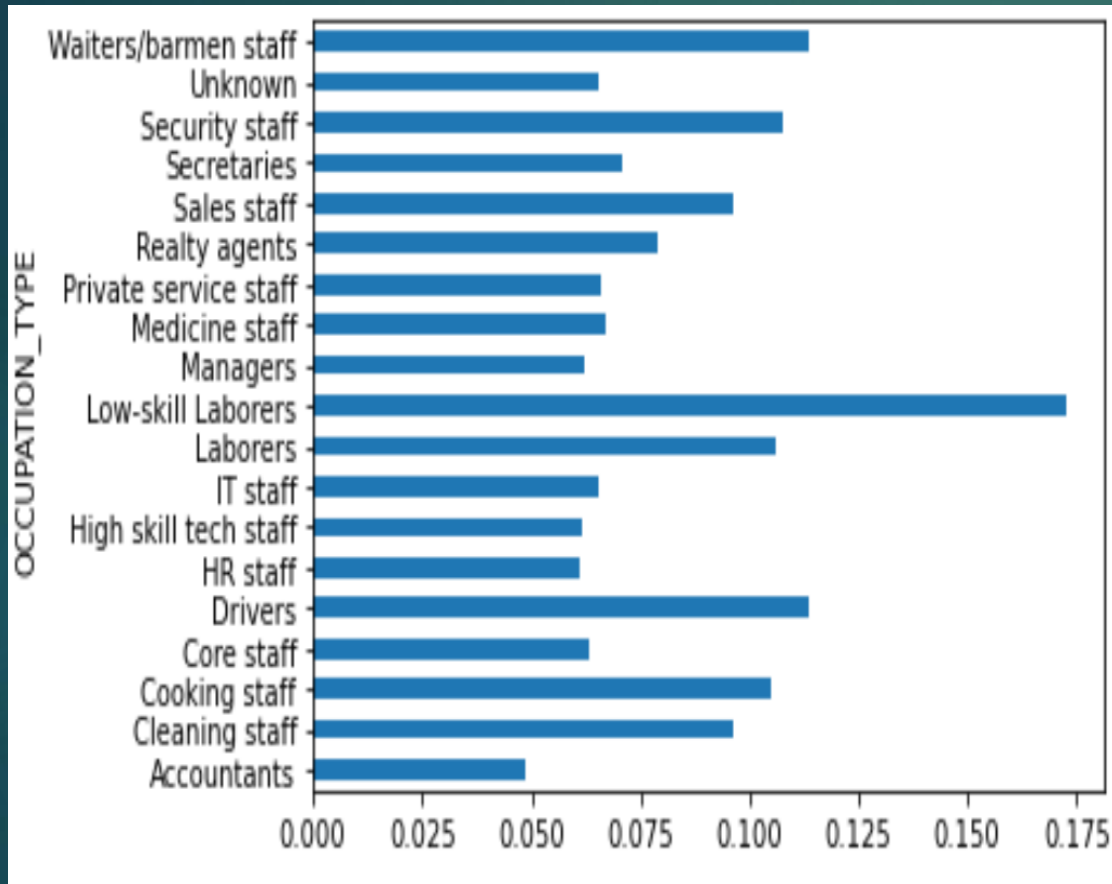- MERGING DATAFRAMES
- CONCLUSION

TARGET VS MARITAL STATUS
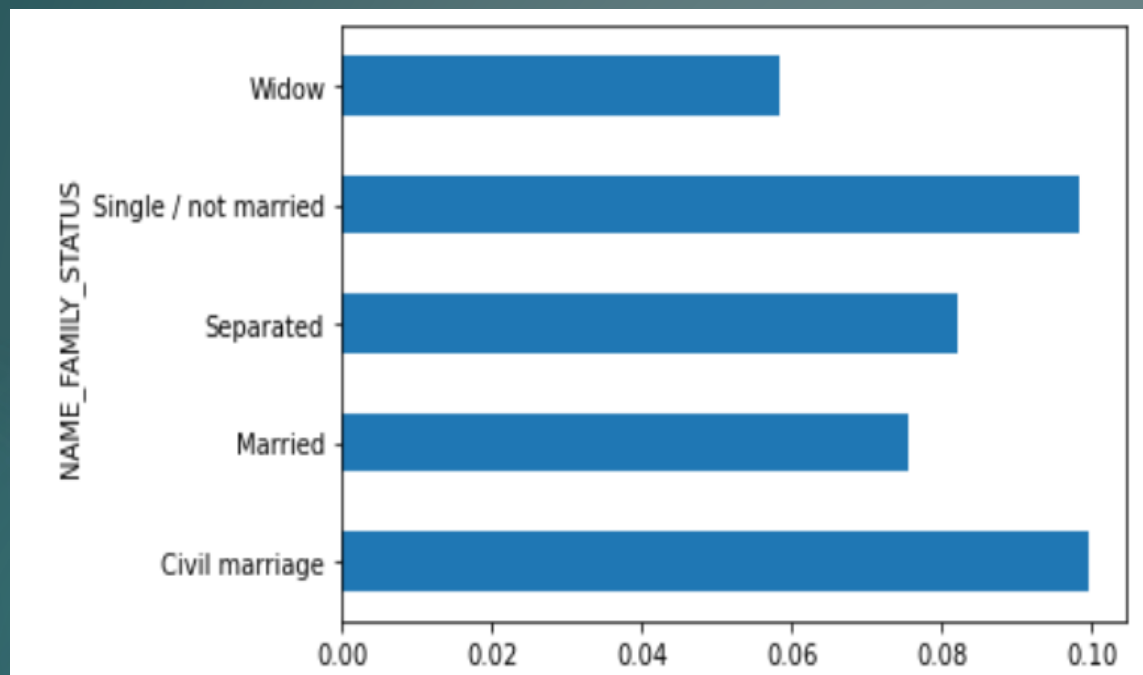
People in civil marriage and those who are Single/not married has higher default,whereas widows default the lowest.

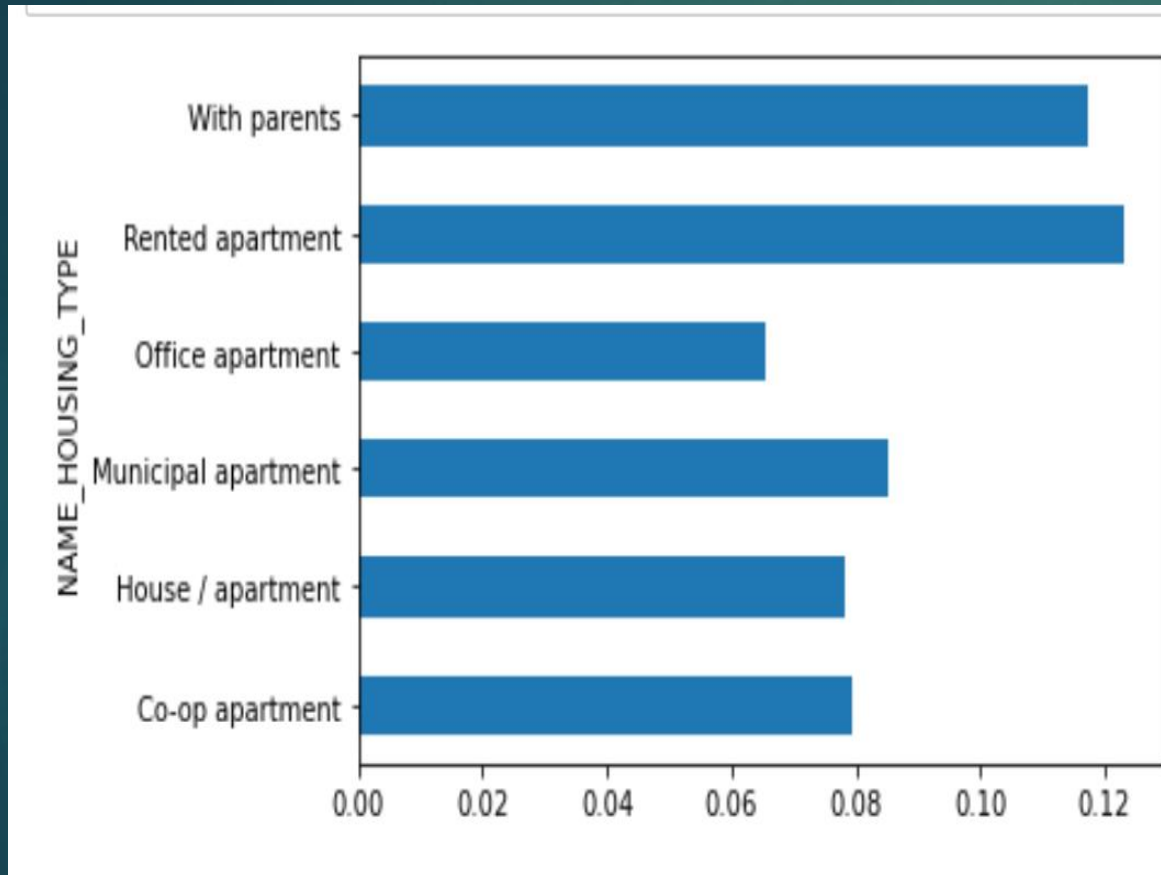The pie chart shows that majority of customers who got loan are married people.
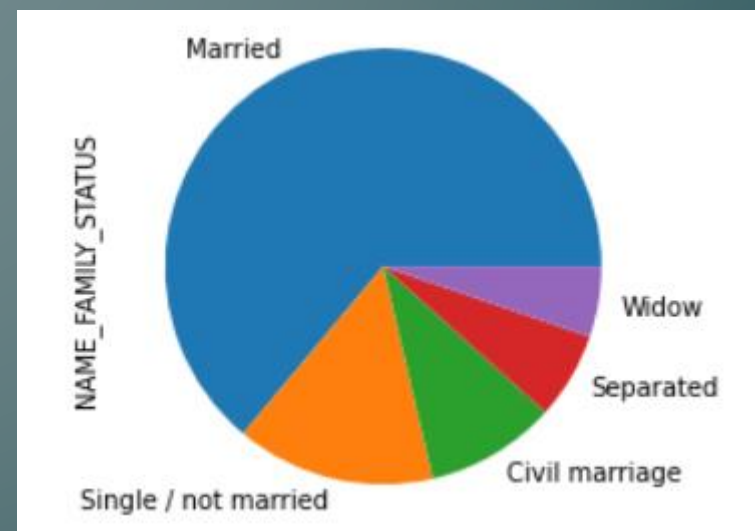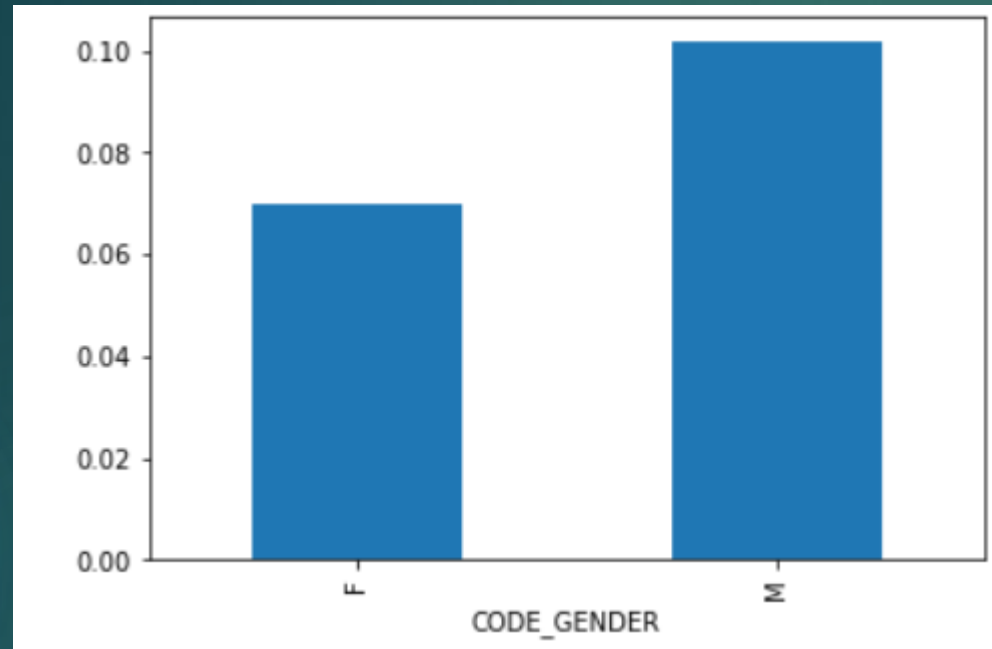
TARGET VS OCCUPATION

The low skill labourers have the highest percentage of default, the Accountants have the lowest default percentage
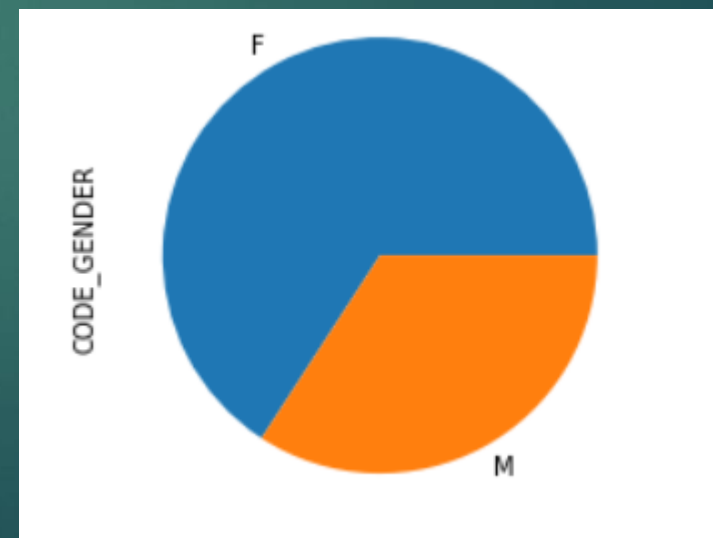
TARGET VS HOUSETYPE

Customers living with parents and living in Rented Apartment has higher default rate.
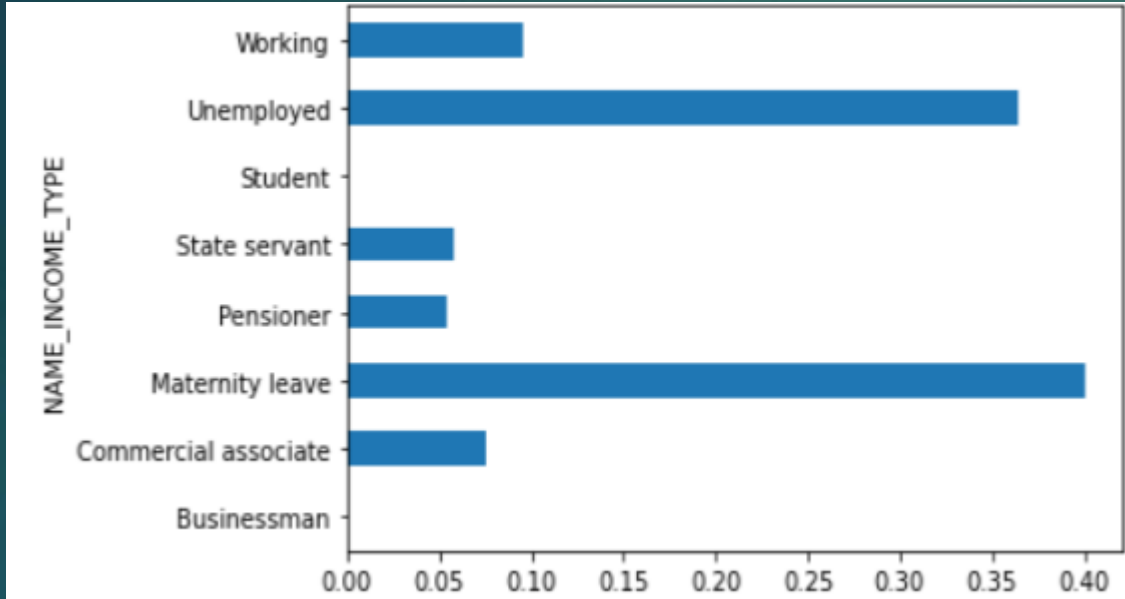
TARGET VS GENDER

Female have lesser default rate than male

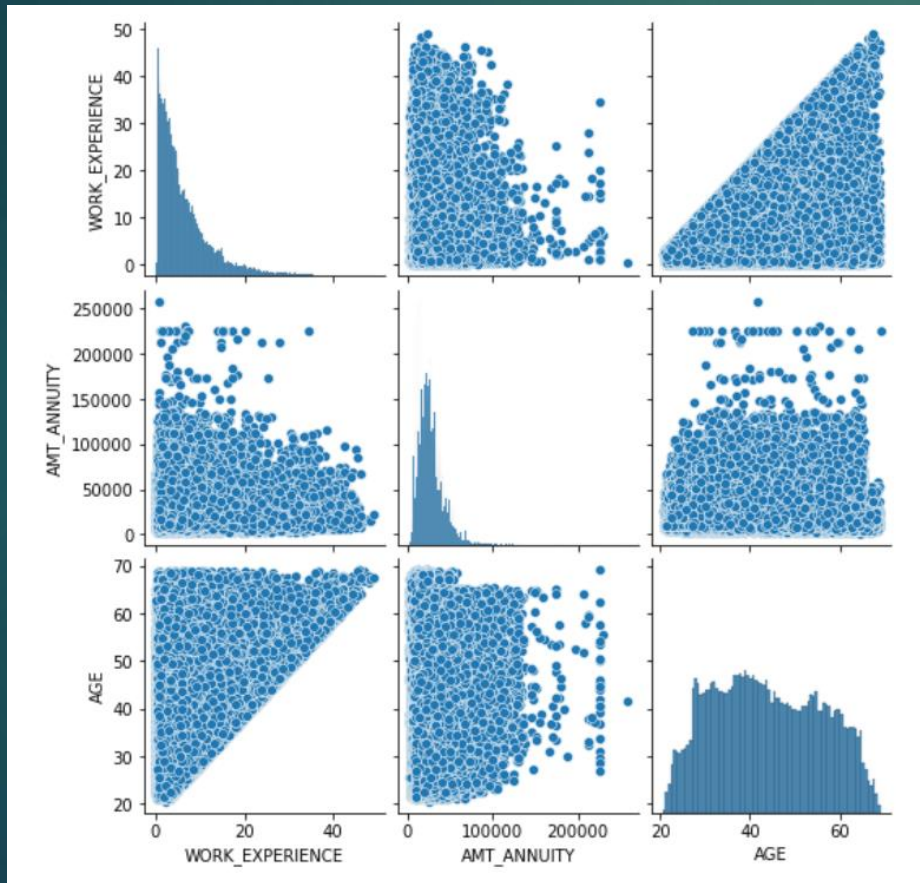But we see that females got more loans than male from the pie chart.

TARGET VS INCOME TYPE

Those Customers who are Students or Businessman do not have missed payments at all

Unemployed customers and those who are on Maternity Leave have missed payments

# AGE VS WORK EXPERIENCE VS AMOUNT ANNUITY

The pair plot doesn't show any significant relation between the variables expect higher the age,higher the work experience.

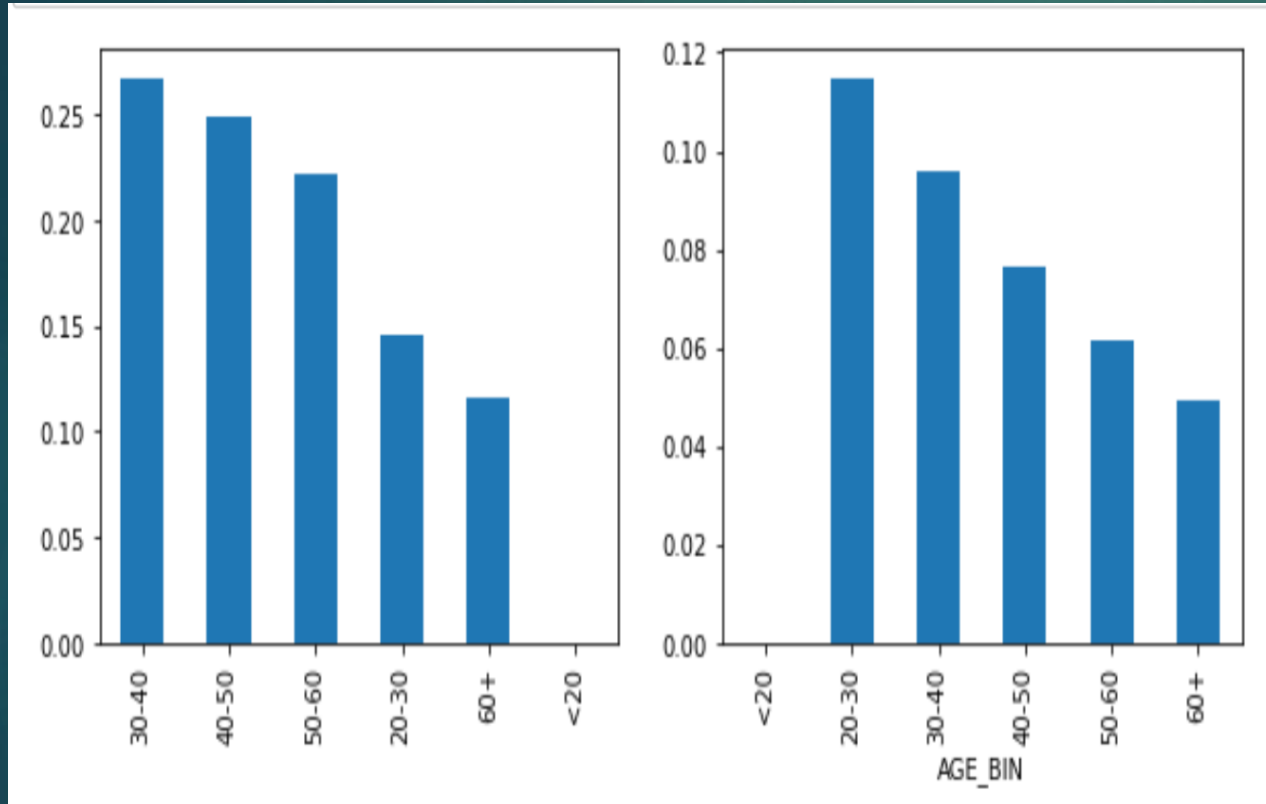# ANNUAL INCOME VS COUNT OF FAMILY MEMBERS VS LOAN ANNUITY
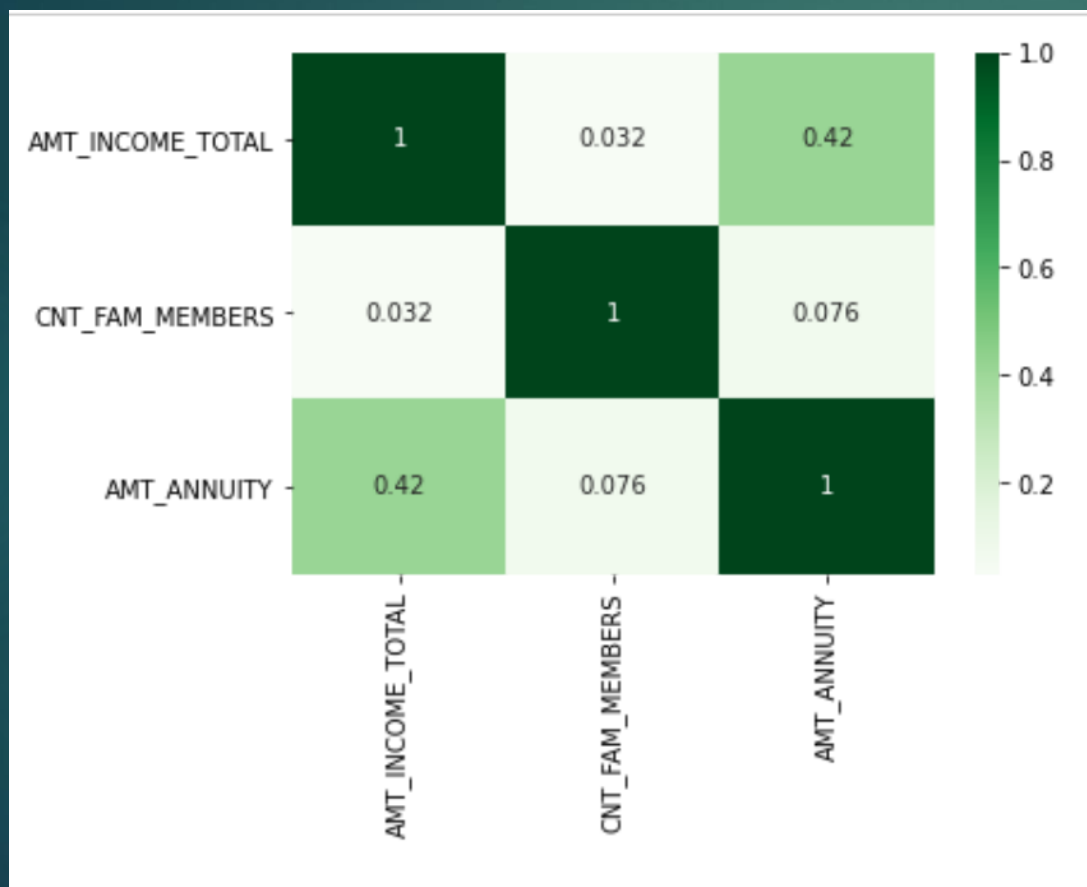
The correlation matrix shows a positive correlation of .42 between Annual Income and Loan Annuity

ANNUAL INCOME VS COUNT OF FAMILY MEMBERS VS LOAN AMOUNT TAKEN
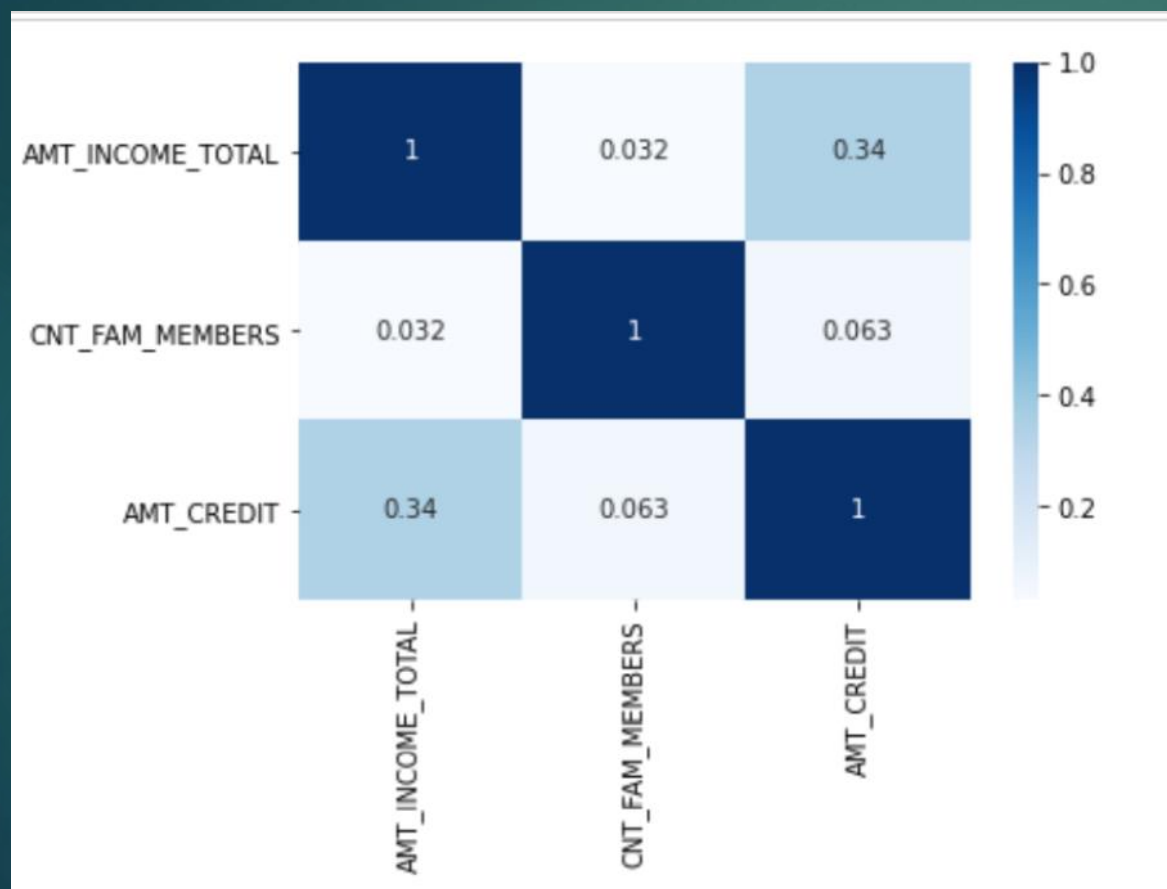
The correlation matrix shows a positive correlation of .34 between Annual Income and Loan Amount Obtained from the bank.
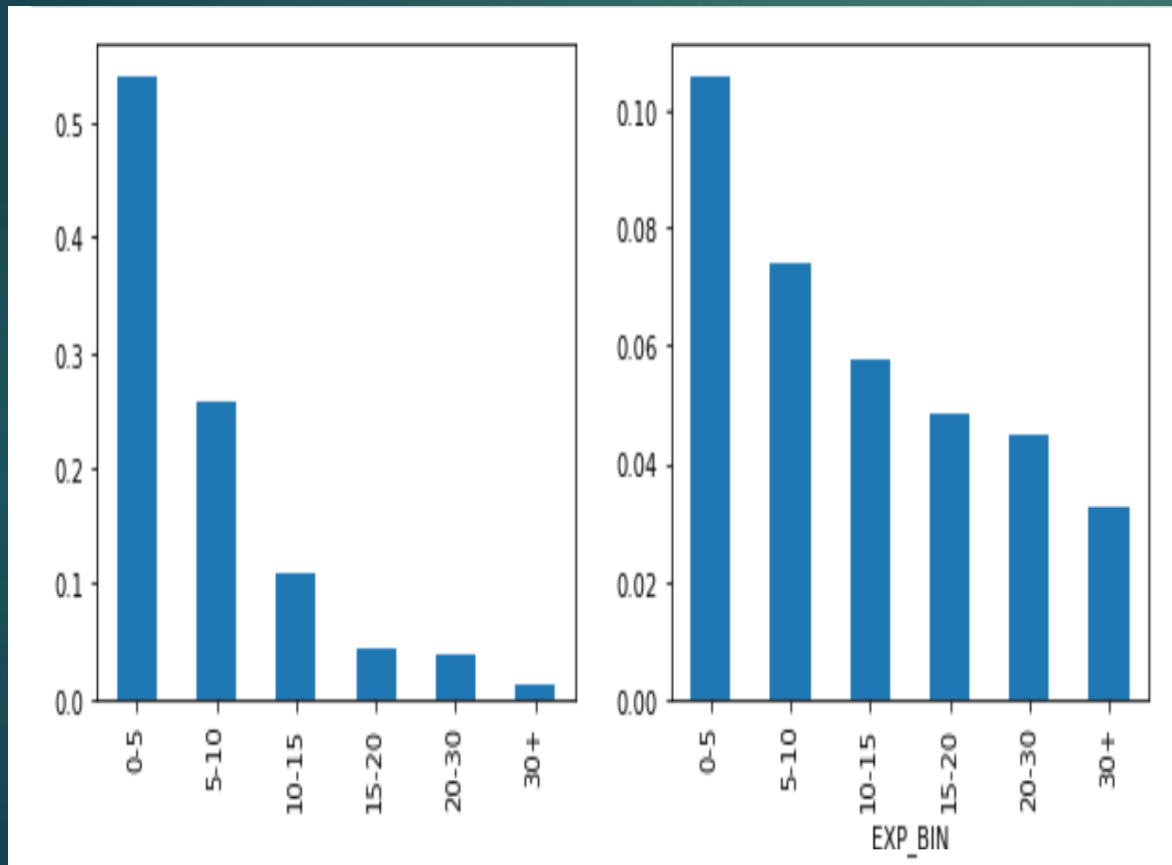
AGE BUCKETS VS TARGET

The first graph shows normalized age group of people who got loan from highest to lowest.30-40 being the highest and 60+ being the lowest.

Not one person under 20 got loan.

The second graph shows the age group of people who defaulted the most highest to lowest which clearly indicates most of the defaulters are in range 20-30.

WORK EXPERIENCE BUCKETS VS TARGET

The first graph shows normalized age group of people who got loan from highest to lowest. 0-5 years being the highest and 30+ years of experience being the lowest.

The second graph shows the group of people who defaulted the most highest to lowest which states that most defaulters have a work experience of 0-5 years.

Defaulting rate and working experience seems to be inversely proportional.

INCOME BUCKETS VS TARGET
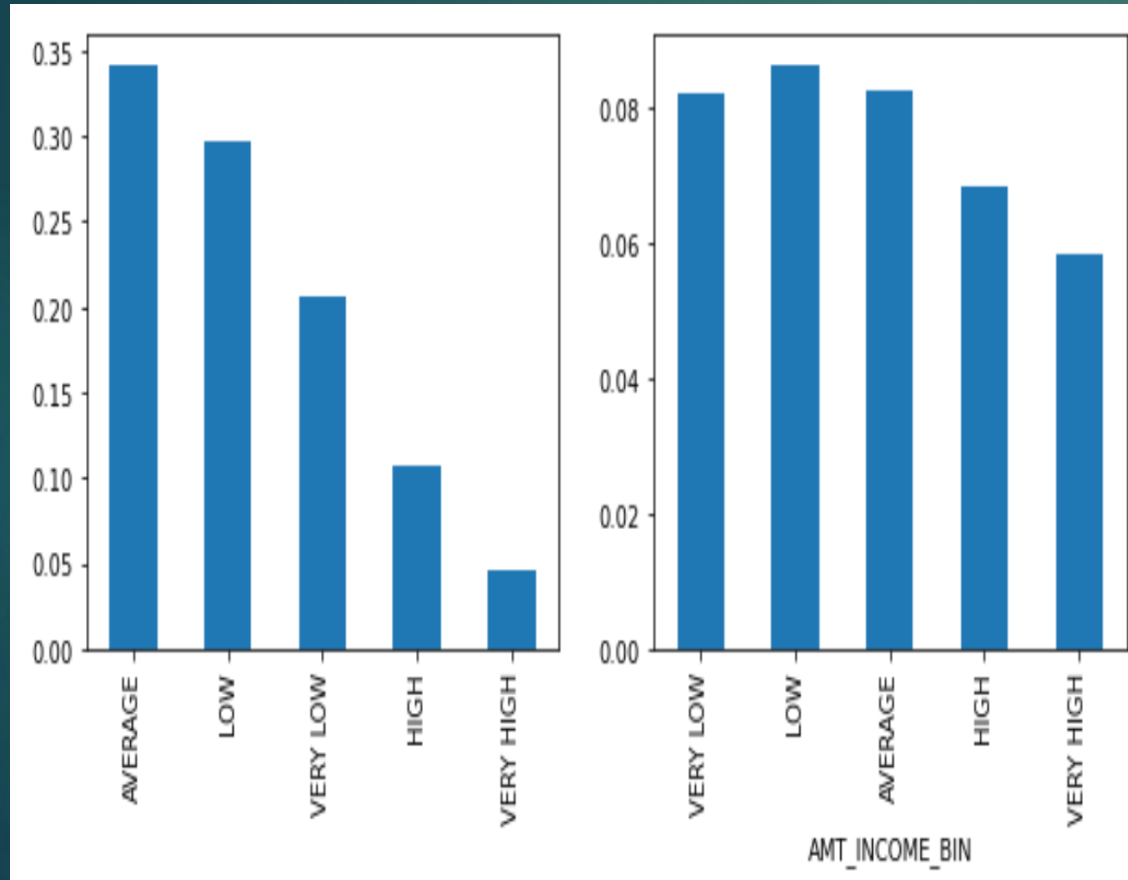
The first graph shows normalized income of people who got loan from highest to lowest.Average category got the most number of loans.Very High category got the least number of loans.
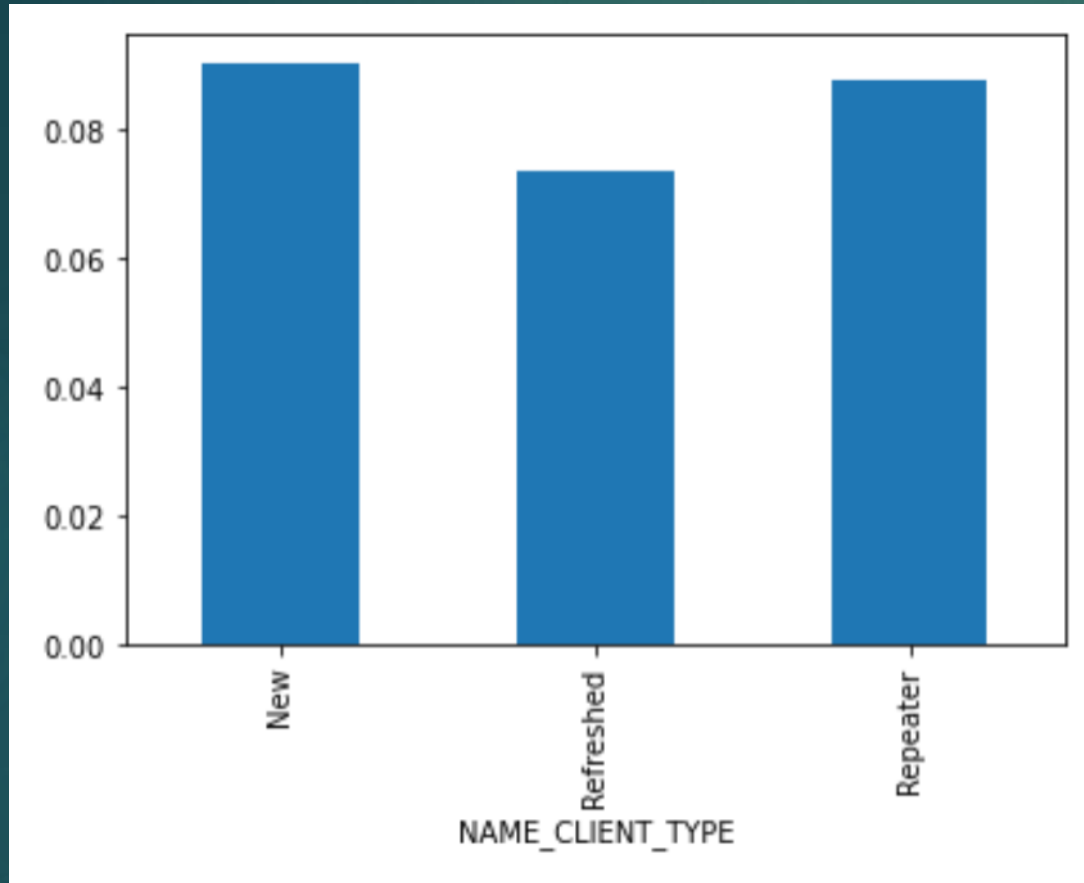
The second graph shows the income range of people who defaulted the most highest to lowest which clearly indicates most of the repayers are in Very High Category.
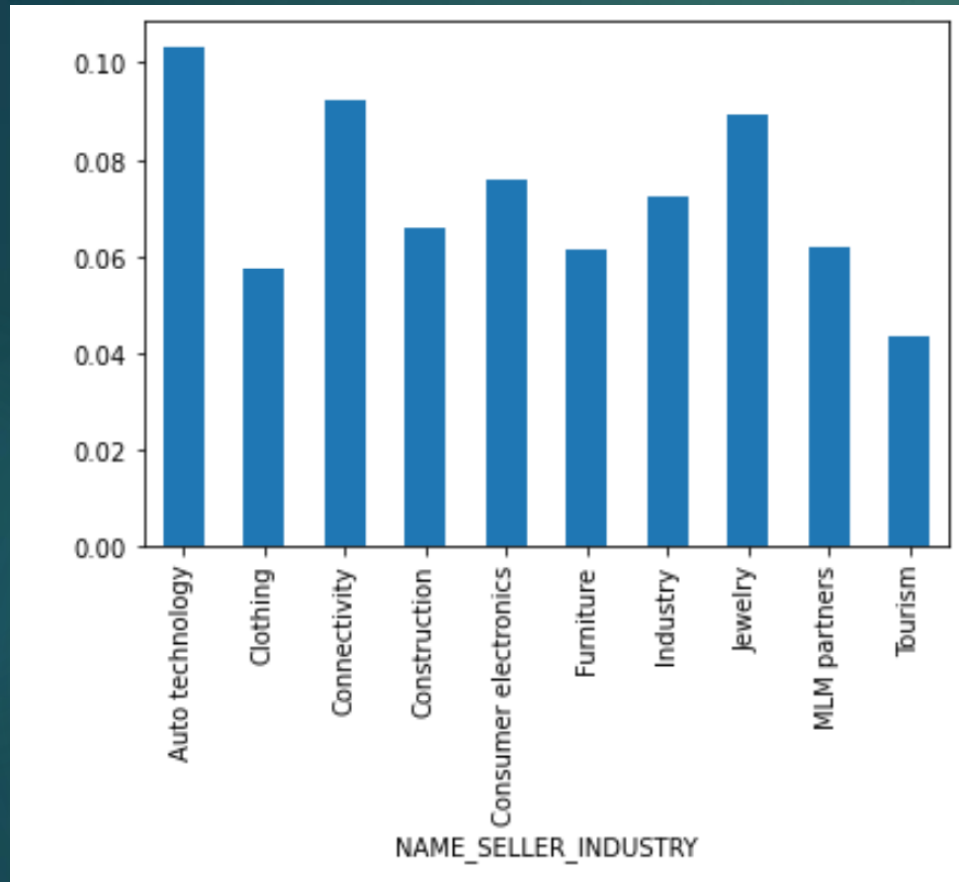
TYPE OF LOAN VS TARGET
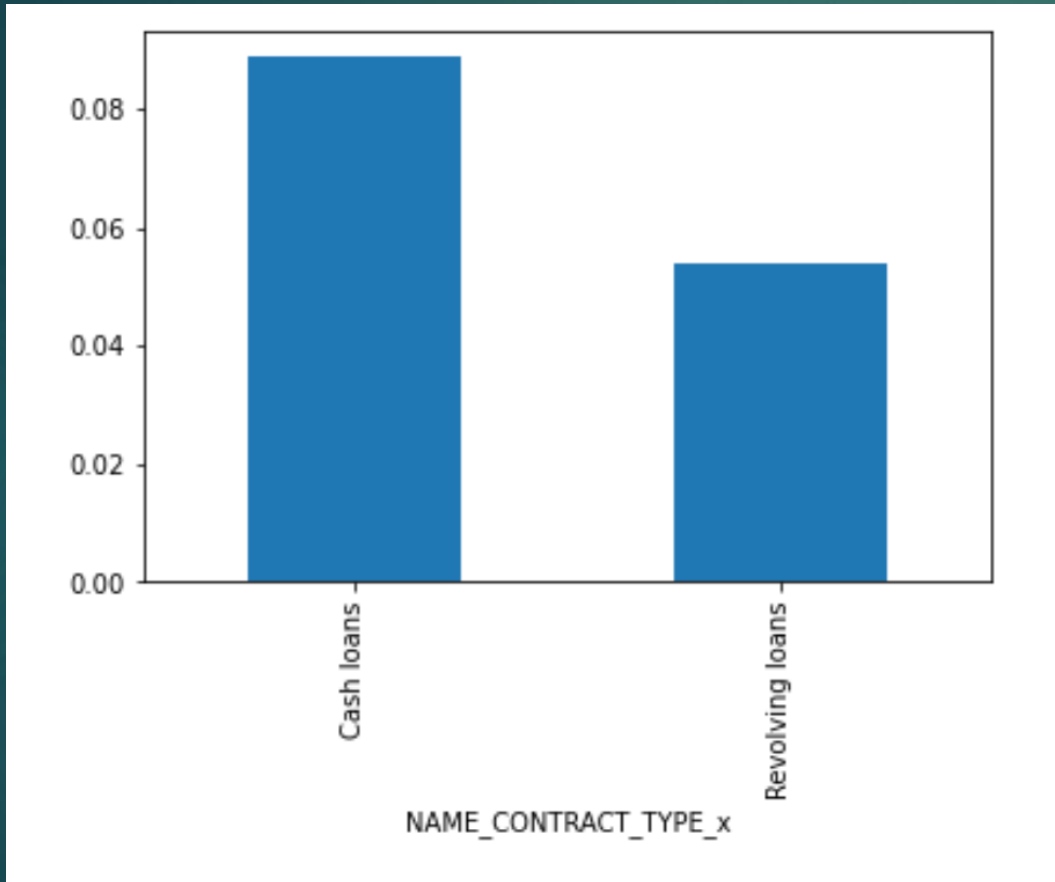
Cash loans are defaulted more than Revolving loans.

TYPE OF CLIENT VS TARGET

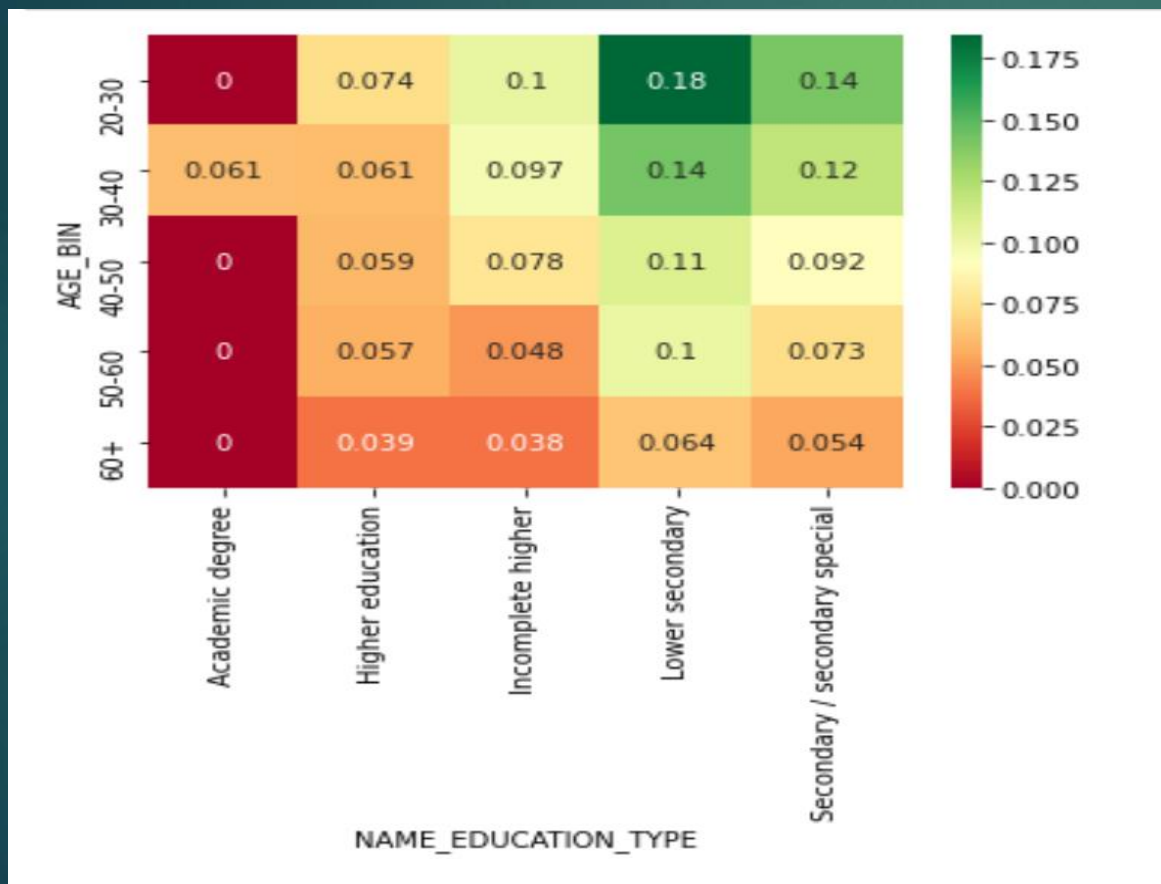Repeaters and Newbies default more than Freshers.

INDUSTRY OF CLIENT VS TARGET

Tourism is the least defaulted Industry,

The most defaulted is Auto Technology followed by Connectivity and Jewelry
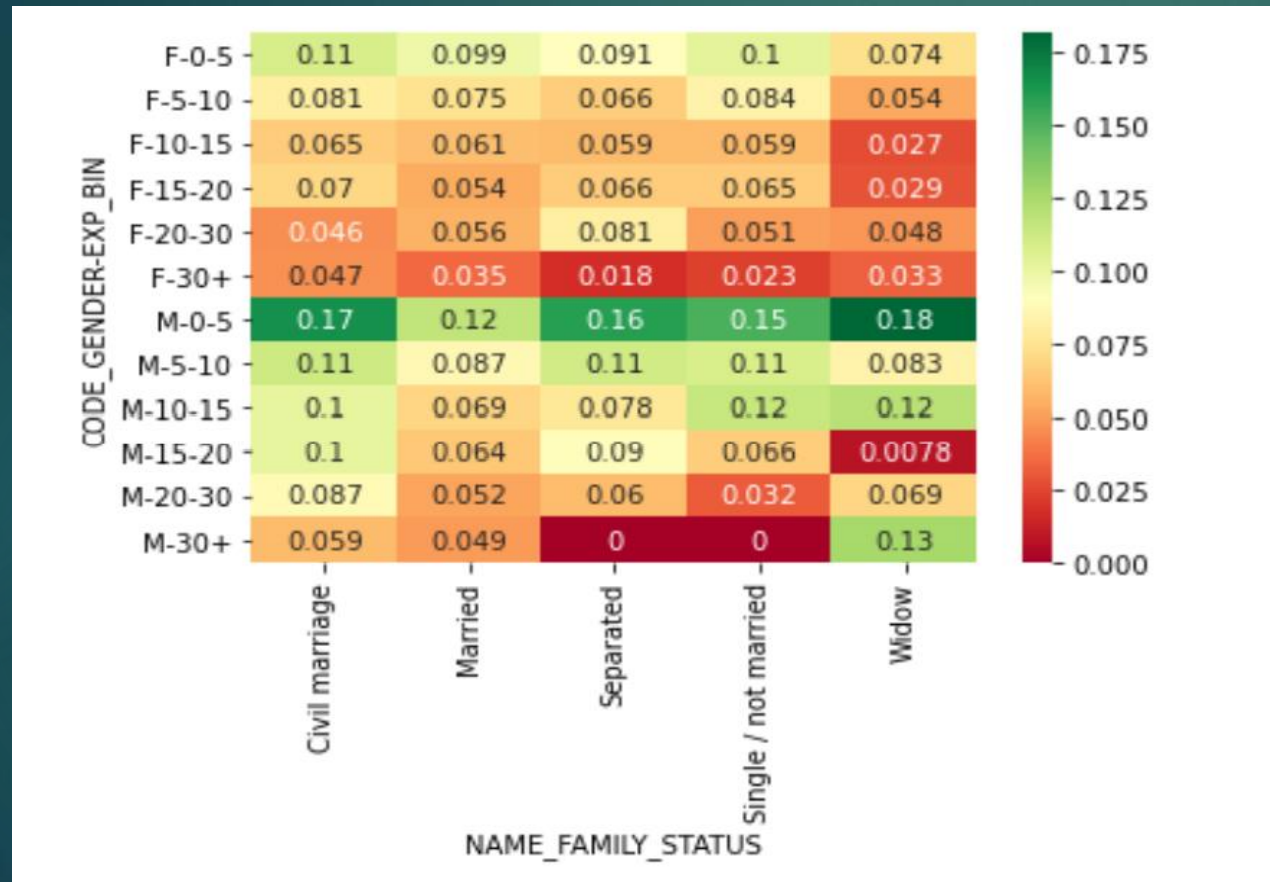
HEATMAP OF AGE BUCKETS VS EDUCATION

The highest defaulters are in the age group of 20-30 and have Lower Secondary Education.

Those with Higher Education are less likely to default.

Defauting is very rare in customers with Academic Degree

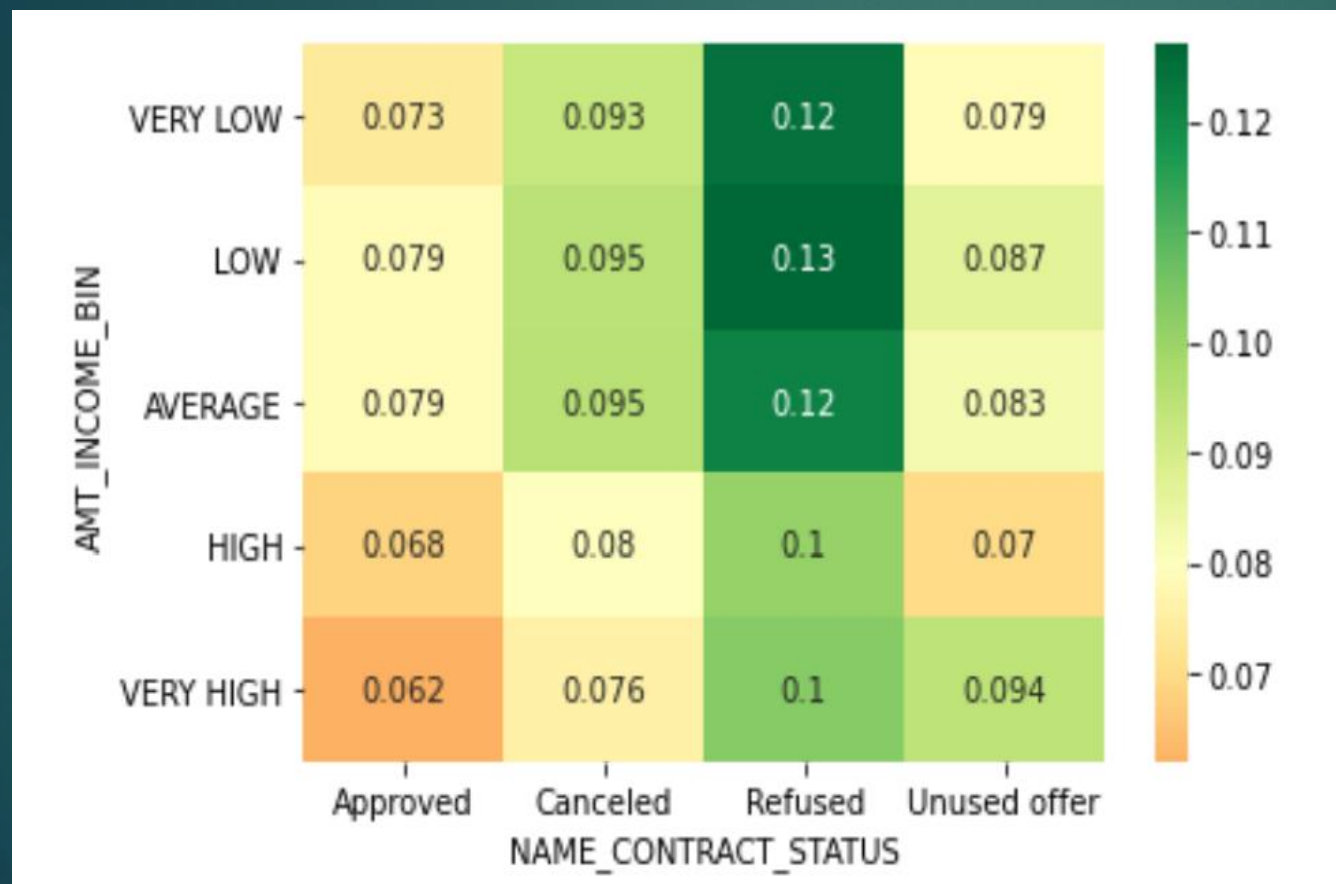# HEATMAP OF MARITAL STATUS VS EXPERIENCE BUCKETS VS GENDER

We can infer from the heatmap that

The Male Category with 0-5 years of Work Experience tend to default a lot.

The Male with 15-20 years of experience who are widowed are less likely to default.

HEATMAP OF INCOME BUCKETS VS PREVIOUS LOAN STATUS

The heatmap shows that the applicants who have refused the previous loan seem to have a higher default rate.

Those applicants who were already approved a loan and has Very High income level are less likely to default

CONCLUSION

▶ The percentage of women repayers are higher making them better clients for bank.

▶ Higher the age and work experience, lesser the default percentage,is a sign for potential repayers.

▶ The work group to invest on are Students,Businessman,Accountants and Managers.

▶ Low skill labourers, persons on Maternity Leave and Unemployed need to have a check.

▶ Single people or people in civil marriage has higher default rate.

▶ The applicants who were already approved a loan and has Very High income level are less likely to default.

▶ Academic Degree Holders default rarely whereas people with Lower secondary/Secondary education default more