

# Credit EDA Case Study

Submitted By: SAHIL JAIN

# PROBLEM STATEMENT :

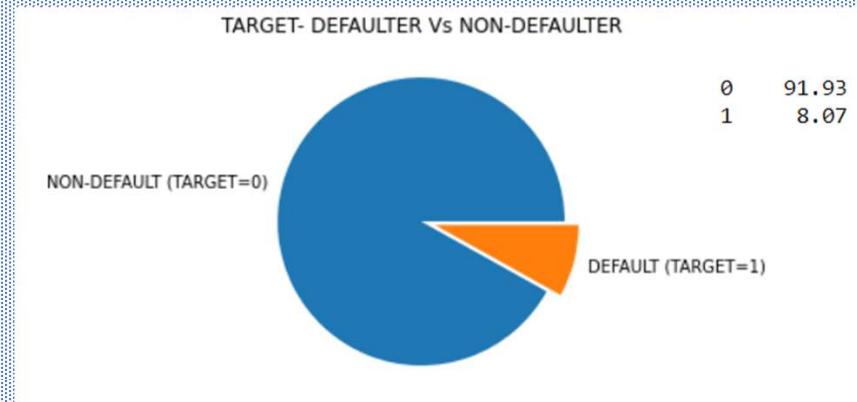
- Use EDA to analyze the patterns present in the data.
- Ensure that the applicants capable of repaying the loan are not rejected.
- Understand how consumer attributes and loan attributes influence the tendency of default.
- Identify patterns which indicate if a client has difficulty paying their installments 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.

# Approach to EDA:

1. Understanding the domain/variables which includes understanding the data using data dictionary and getting domain knowledge.
2. Import the libraries and Load the datasets.
3. Check the structure/metadata of the data which includes shape, datatypes, info, rows/columns, describe etc.
4. Missing value check
5. Outlier Check, imputation and binning
6. Perform Univariate Analysis.
7. Perform Segmented Univariate.
8. Perform Bivariate Analysis.

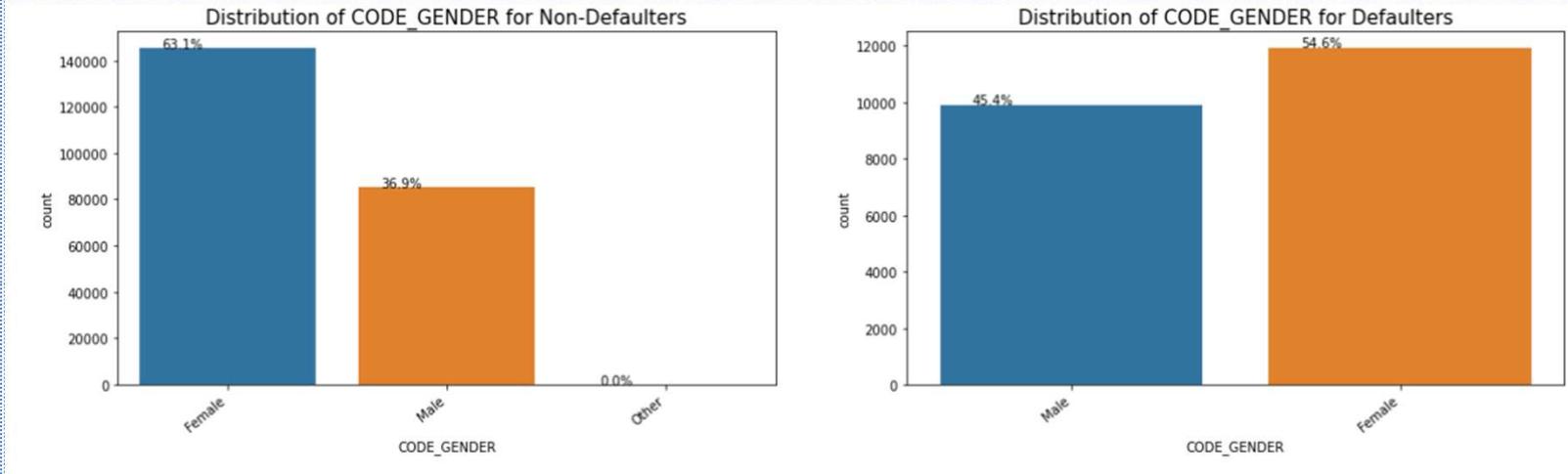
# Exploratory Data Analysis Results:

# Data Imbalance:



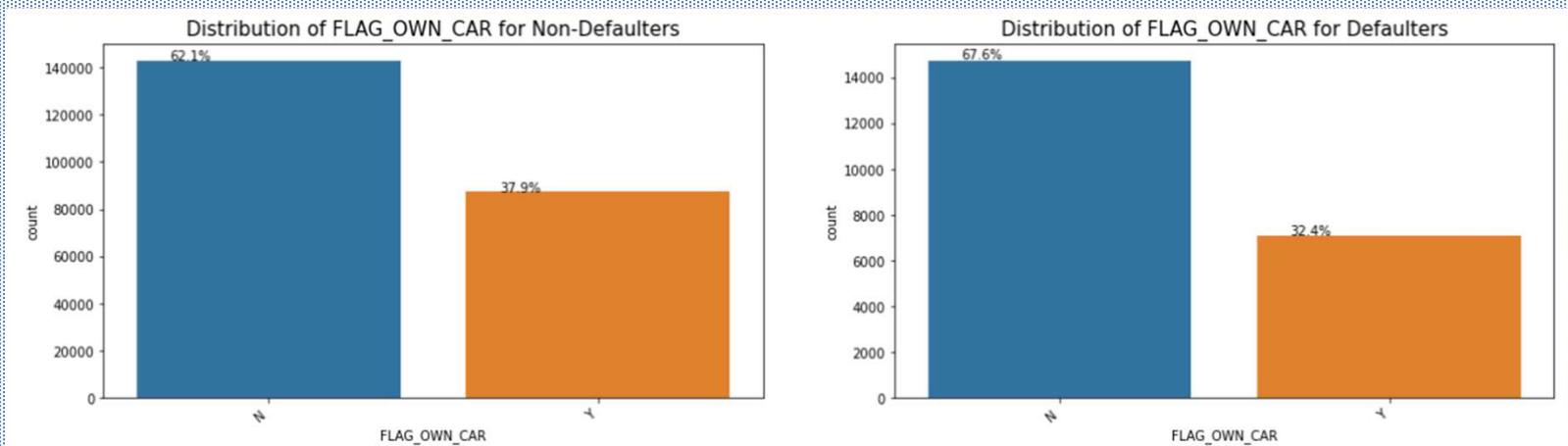
- This pie diagram shows the data imbalance basis our target of analysis i.e., Defaulters VS Non-Defaulters
- Here 1 represents client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample and 0 represent all other cases.
- We can see that in our data, we have approx. 92% people in non defaulter category and remaining approx. 8% in defaulter category.

# GENDER:



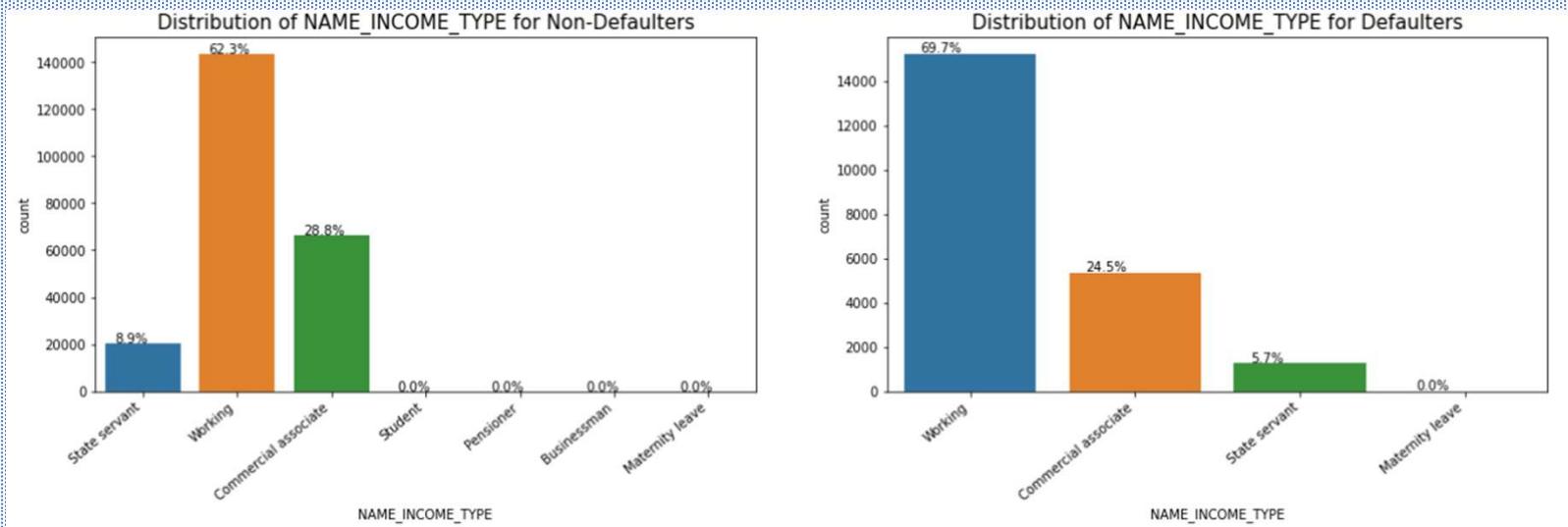
- This diagram shows the Gender of the client
- We can clearly see that the female applicants is more than male applicants
- Also, we can see that female defaulters is more than male defaulters probably because of more number of applications
- However, the rate of default of females is lower than the rate of default of males.
- Therefore, more females must be targeted for loans.

# CAR OWNERS:



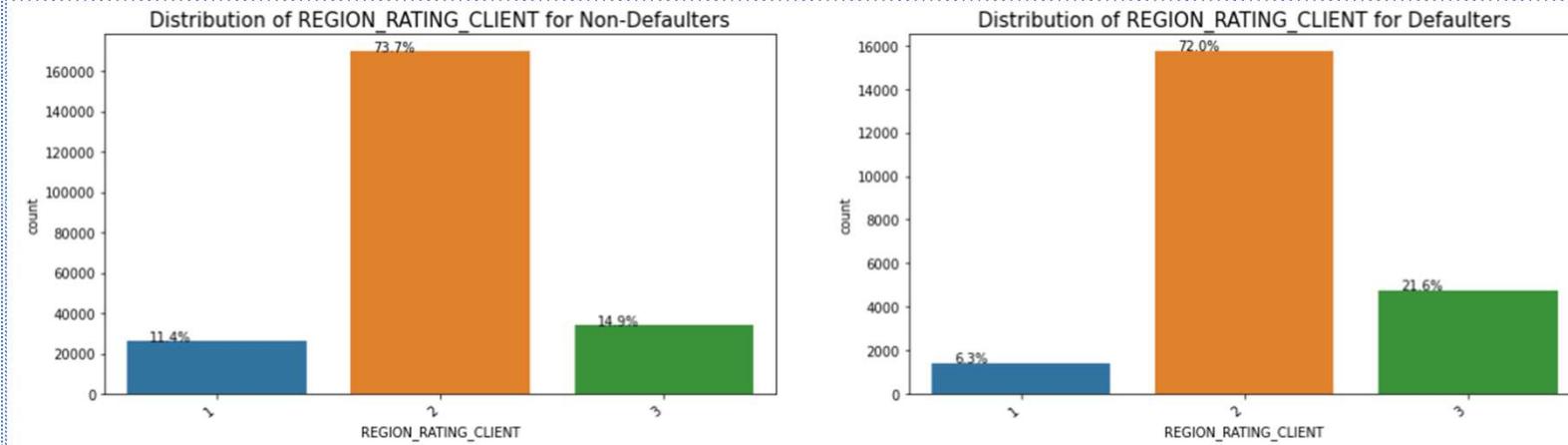
- This diagram shows if the client owns a car
- We can clearly see that the majority of applicants do not own a car
- Also, we can see that people without car default more often probably because of more number of applicants in that category
- However, the rate of default is similar for both the categories
- Hence, default on loan does not depend on whether the person owns a car or not it just tells about the social status

# INCOME TYPE:



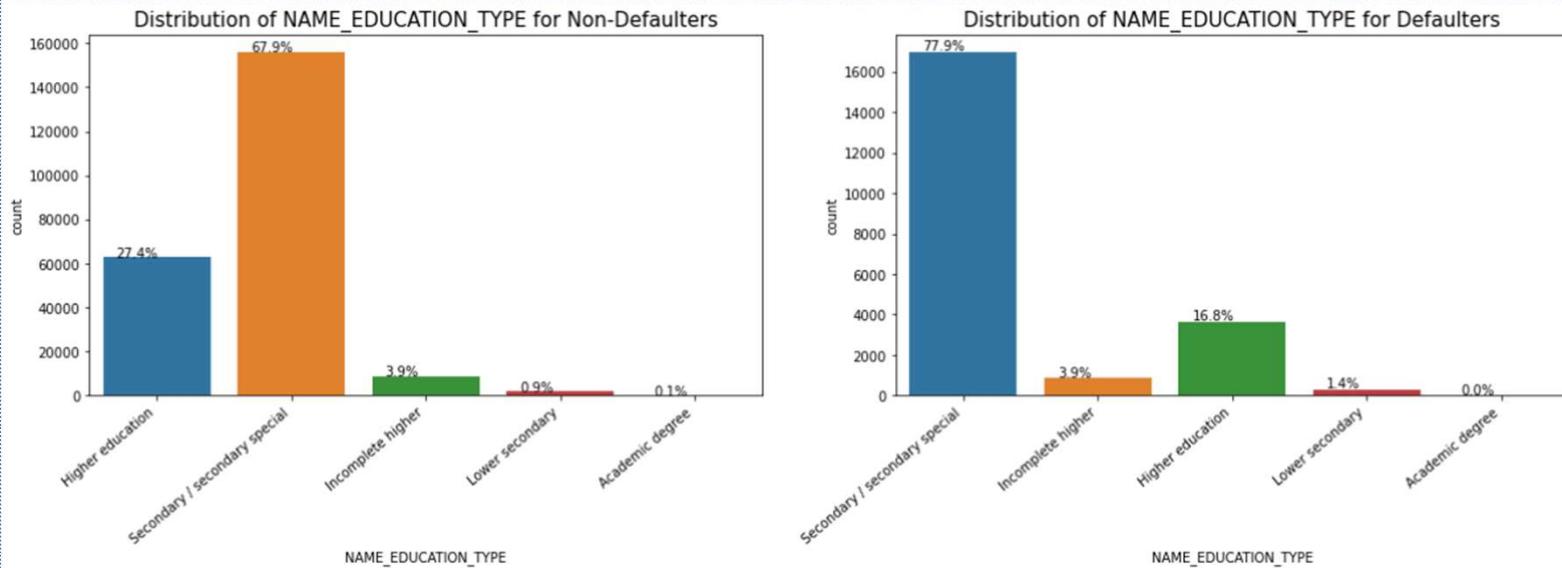
- This diagram shows the clients income type (businessman, working, maternity leave,...)
- We can clearly see that Businessmen and Students donot default
- Most of the loans are taken by Working Class and hence the maximum number of defaults
- Therefore, risk free target audience will be students and Businessmen

# REGION RATING:



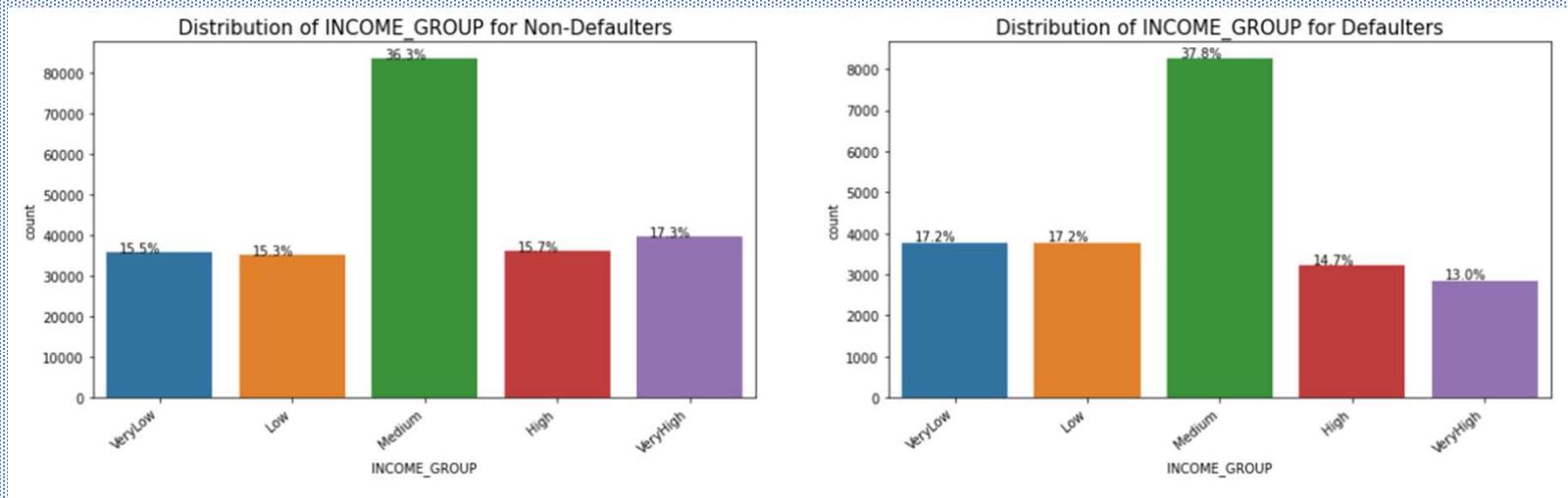
- This diagram shows our rating of the region where client lives (1,2,3)
- More people from second tier regions tend to apply for loans and hence more defaulters
- We can infer that people living in better areas(Third tier) tend contribute more to the defaulters by their weightage
- People living in first tier rated areas are comparatively safest target audience

# EDUCATION:



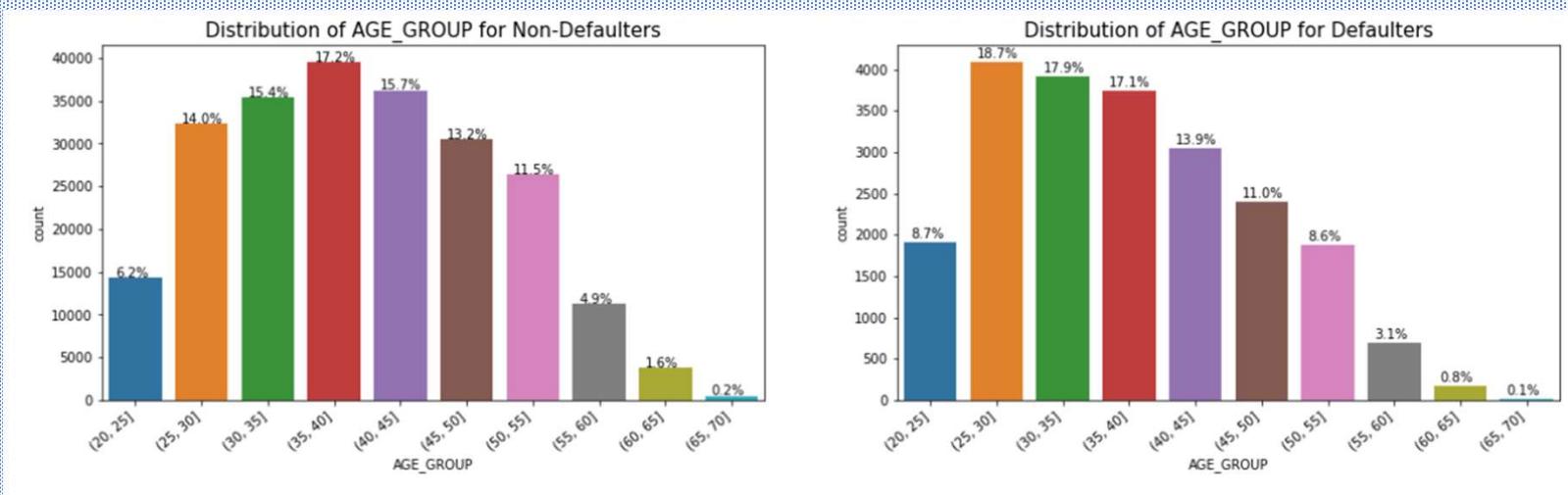
- This diagram shows Level of highest education the client achieved
- Higher educated ones are less likely to default probably due to better salaries and income due to education
- Secondary educated people are more likely to default probably due to high number of applicants

# INCOME GROUP:



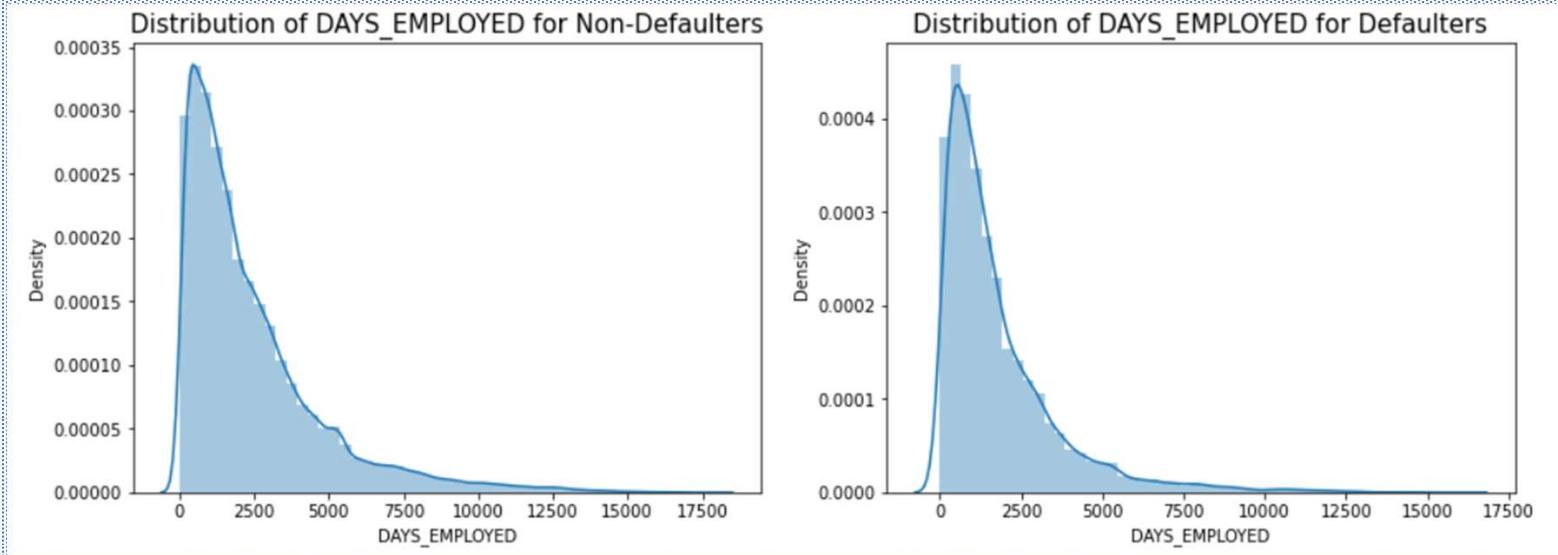
- We grouped the clients on the basis of their incomes
- This diagram shows income level of the clients
- As expected very high income group is the safest as they default the least

# AGE GROUP:



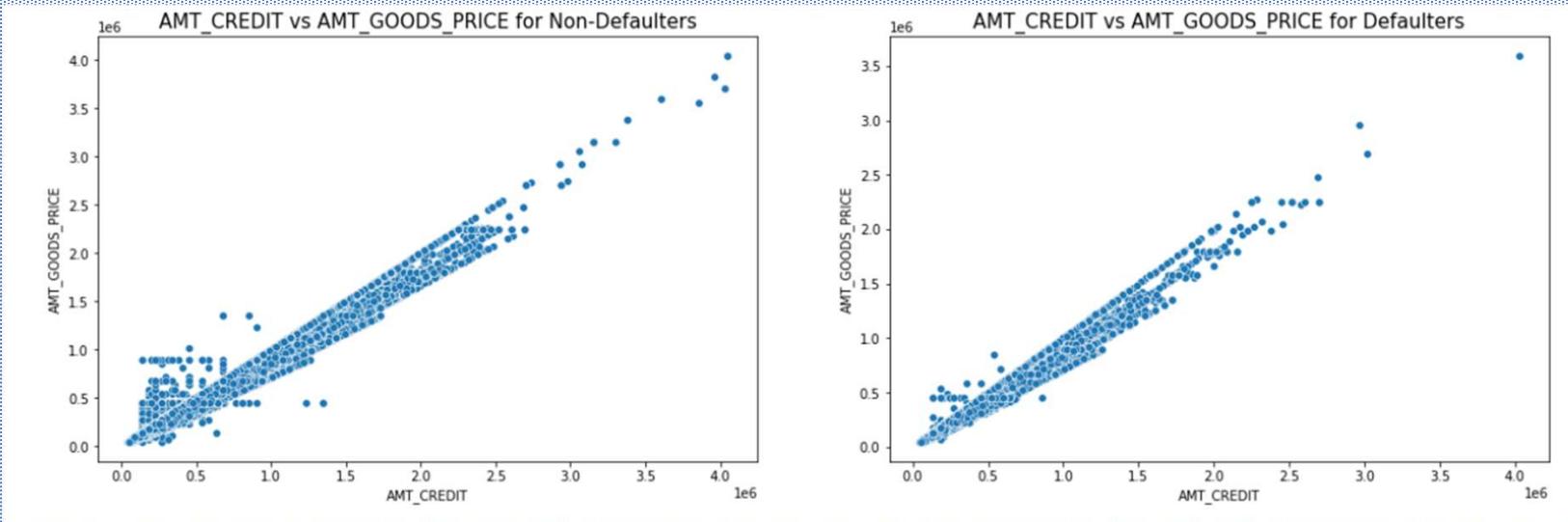
- We grouped the clients on the basis of their age
- This diagram shows age groups of the clients
- We see that (25,30) age group tend to default more.
- With increasing age group, people tend to default less starting from the age 25. One of the reasons could be they complete their education and get employed around that age and with increasing age, their experience and salary also increases.

# EXPERIENCE:



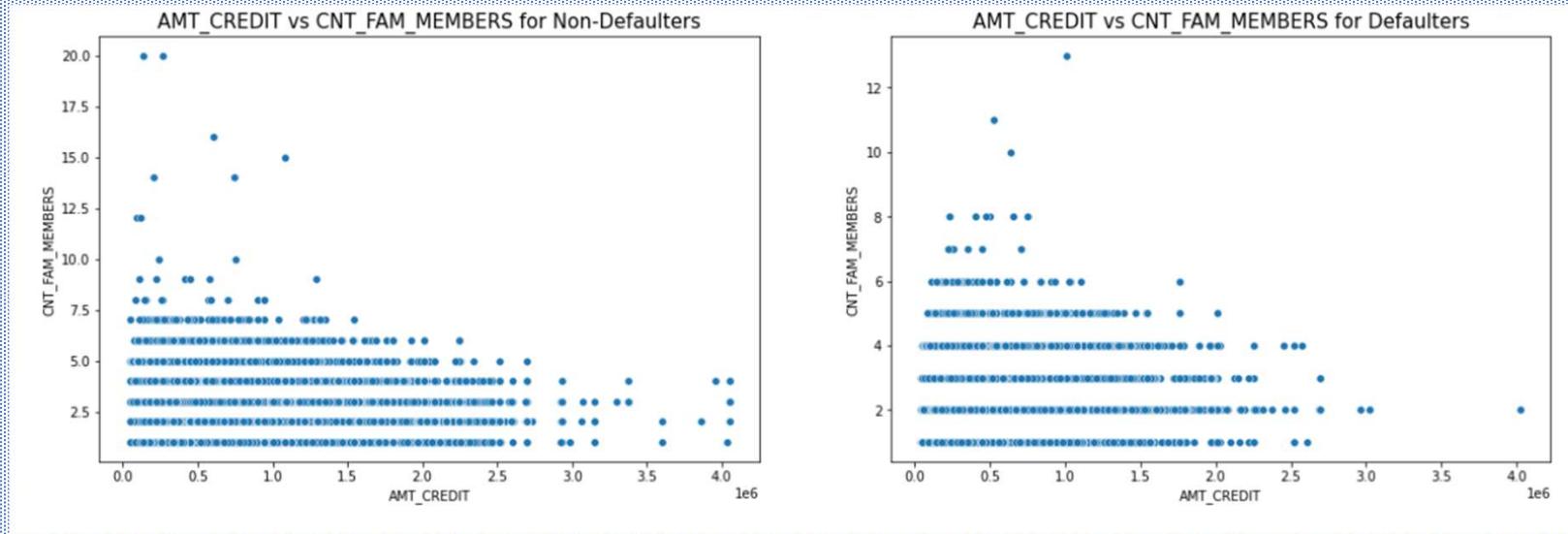
- This diagram shows how many days before the application the person started employment
- From the above plotting, we can say that most of the population is employed for less than 5000 days i.e less than 13 years who are applying for loan and hence more number of defaulters in the same category.
- As the number of years of experience increase, they get better salaries/income and hence do not apply and default on loans.

# CREDIT AMOUNT & PRICE OF GOODS:



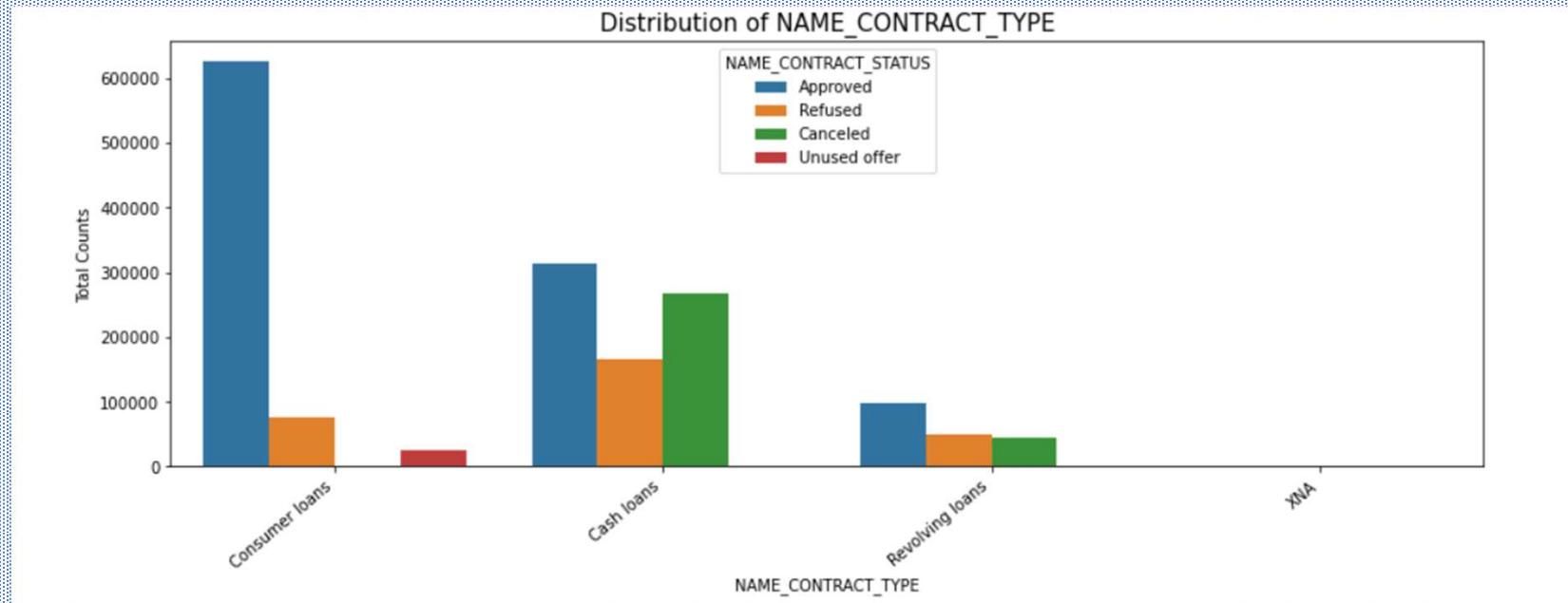
- This diagram shows the relation between final credit amount on the previous application and the price of the goods for which the loan is given for the consumer loans.
- Since they are linearly scattered, we can say that they are directly proportional.
- As the price of goods increase, the credit amount increases which indicates that the goods act as security of the amount loaned
- And hence we can say that bank must target consumer loans as they are secured

# CREDIT AMOUNT & FAMILY SIZE:



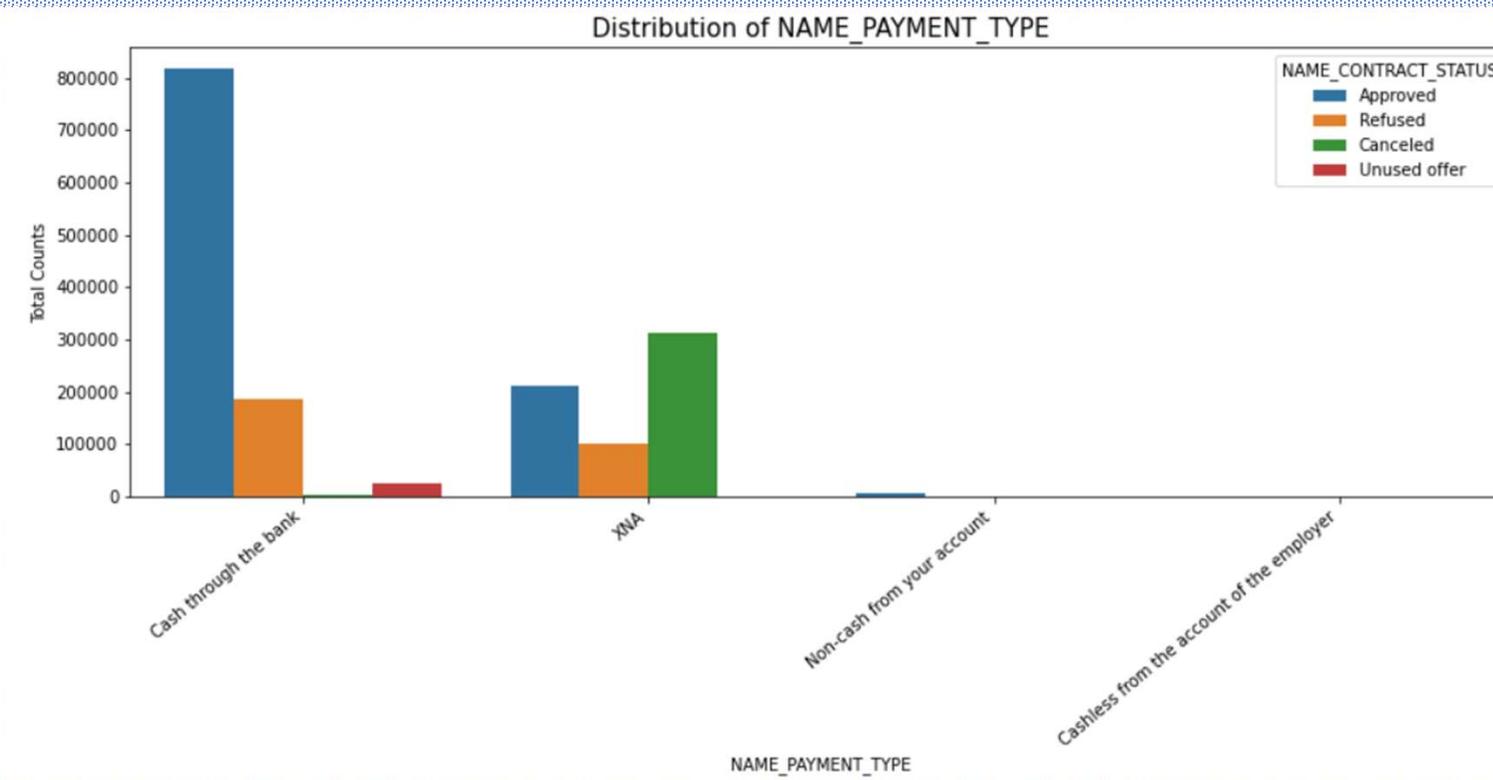
- This diagram shows the relation between final credit amount on the previous application and the number of family members.
- We can see that the density in the lower left corner is similar in both the case, so the people are equally likely to default if the family is small and the credit amount is low.
- We can observe that larger families and people with larger credit amount default less often
- Therefore, people who take small loans tend to take loans for survival and hence default more.

# PREVIOUS CONTRACT TYPE:



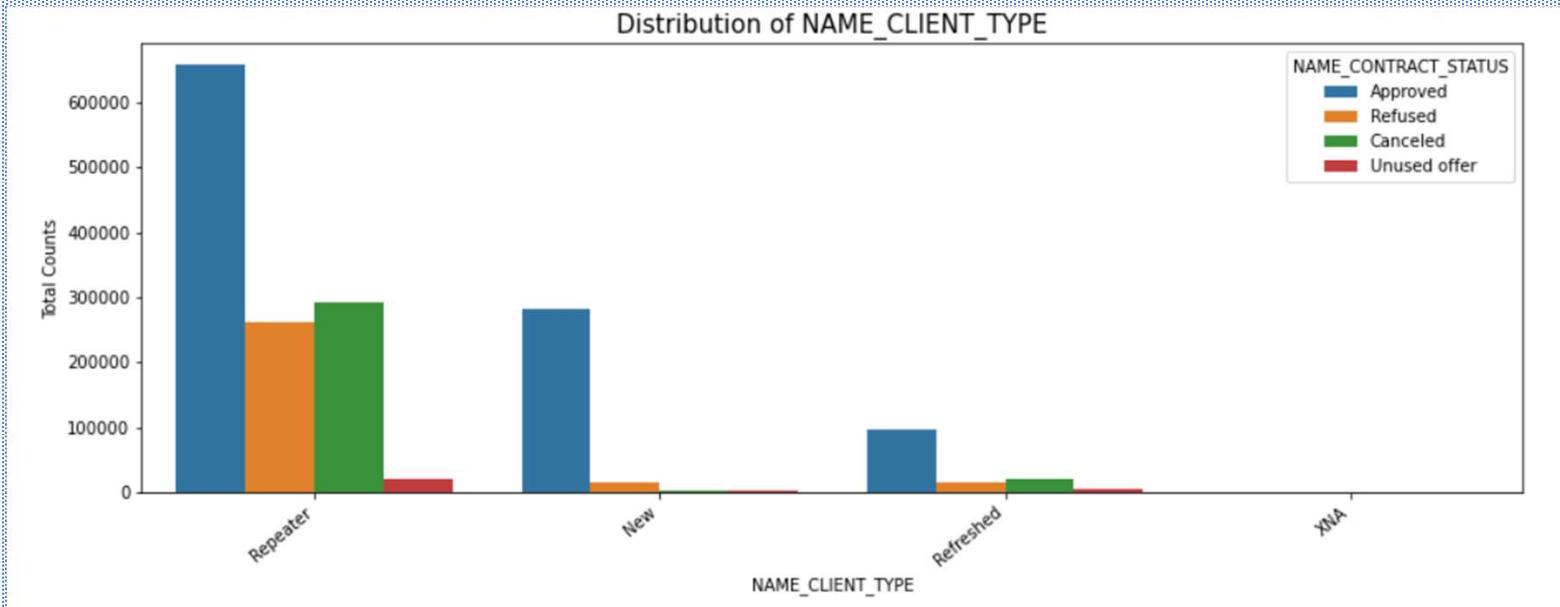
- This diagram shows the contract product type (Cash loan, consumer loan [POS] ,....) of the previous application
- We can see that max number of applications were for consumer loans and cash loans.
- Also, cash loans were refused the most and consumer loans were approved the most.

# PREVIOUS PAYMENT TYPE:



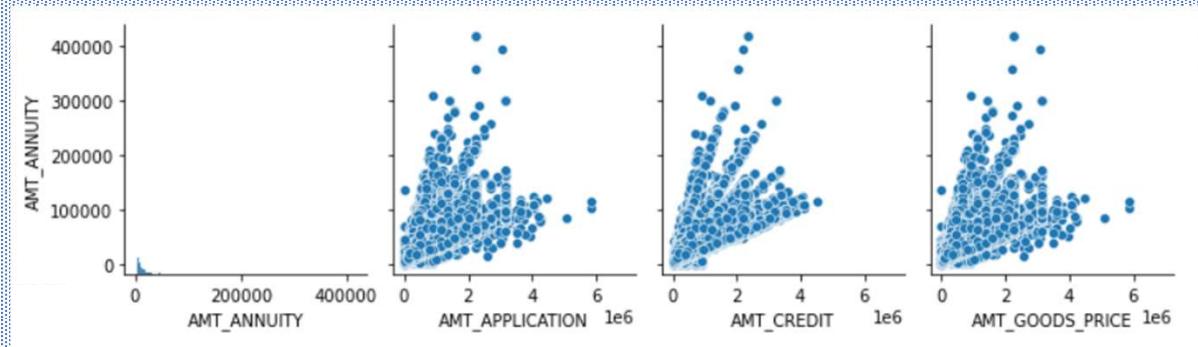
- This diagram shows the payment method that client chose to pay for the previous application
- From the above chart, we can infer that most of the clients chose to repay the loan using the 'Cash through the bank' option
- We can also see that 'Non-Cash from your account' & 'Cashless from the account of the employee' options are not at all popular in terms of loan repayment amongst the customers.

# PREVIOUS CLIENT TYPE:



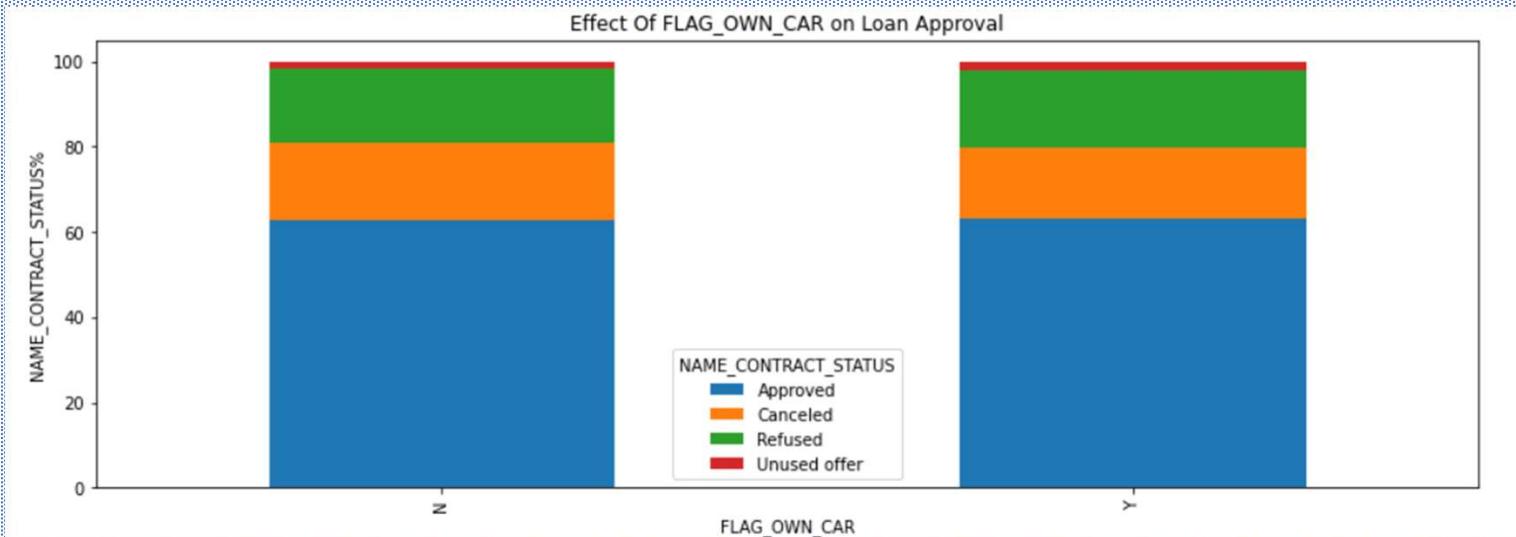
- This diagram shows whether the client was old or new client when applying for the previous application
- Out of the total applications 70% of customers are repeaters. They also get refused most often.
- We need to capture more new clients.

# PREVIOUS CORRELATION ANALYSIS:



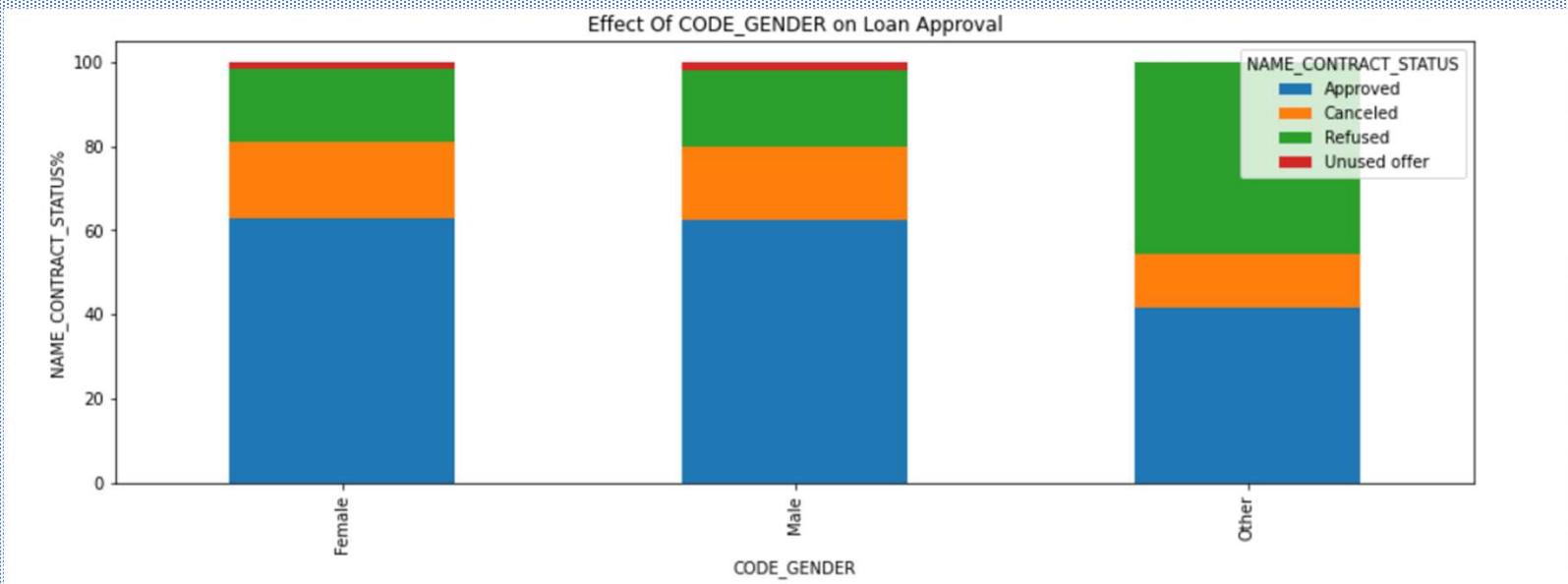
- This diagram shows relationship between Annuity of previous application, For how much credit did client ask on the previous application, Final credit amount on the previous application and Goods price of good that client asked for (if applicable) on the previous application
- Annuity of previous application has a very high and positive influence over:
  1. How much credit did client asked on the previous application
  2. Final credit amount on the previous application that was approved by the bank
  3. Goods price of good that client asked for on the previous application.

# EFFECT OF CAR OWNERSHIP:



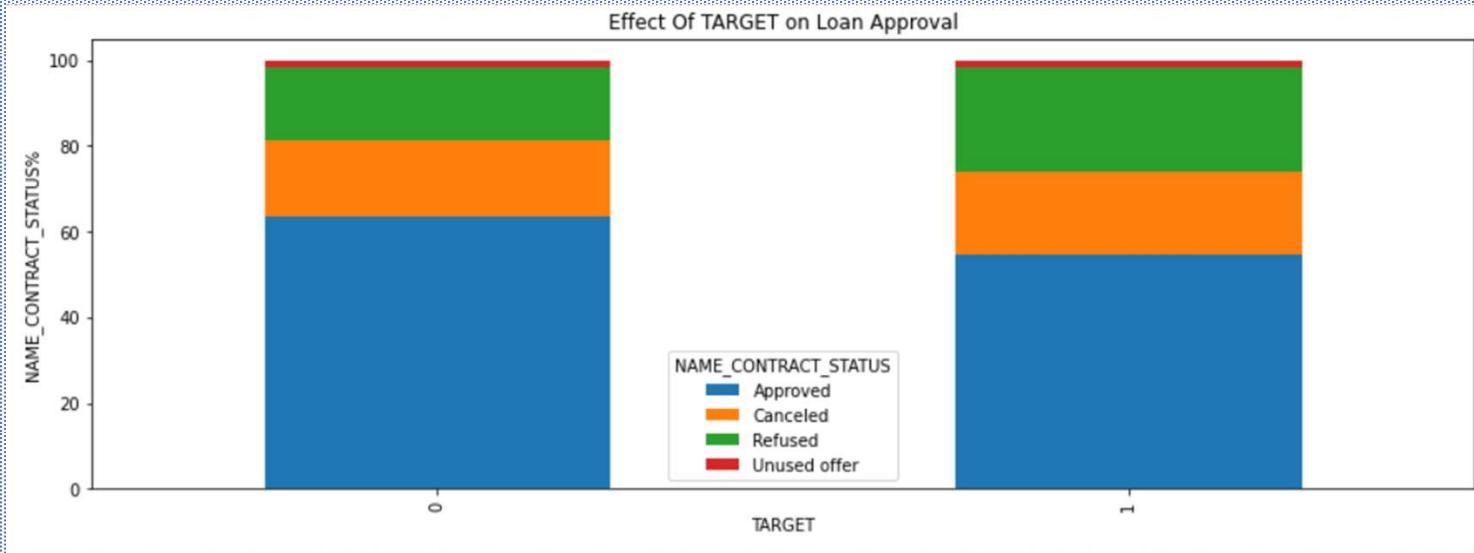
- This diagram shows relationship between the client owning a car and contract status (approved, cancelled, ...) of previous application
- We see that car ownership doesn't have any effect on application approval or rejection.
- We saw earlier that the people who have a car have lesser chances of default.
- It is generally observed that people who have a car are more well off.
- The bank can add more weightage to car ownership while approving a loan amount

# EFFECT OF GENDER:



- This diagram shows relationship between the gender or the client and contract status (approved, cancelled, ...) of previous application
- We see that code gender doesn't have any effect on application approval or rejection.
- We saw earlier that female have lesser chances of default compared to males.
- The bank can add more weightage to female while approving a loan amount.

# EFFECT OF LATE PAYMENTS:



- This diagram shows relationship between the repayment difficulties to clients and contract status (approved, cancelled, ...) of previous application
- We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.
- Therefore, Bank must analyze the reason for previous refusal and take decision accordingly