CREDIT EDA CASE STUDY

RISK ANALYSIS FOR BANKING AND FINANCIAL SERVICES

Presentation by: Prakhar Newatia

Problem Statement

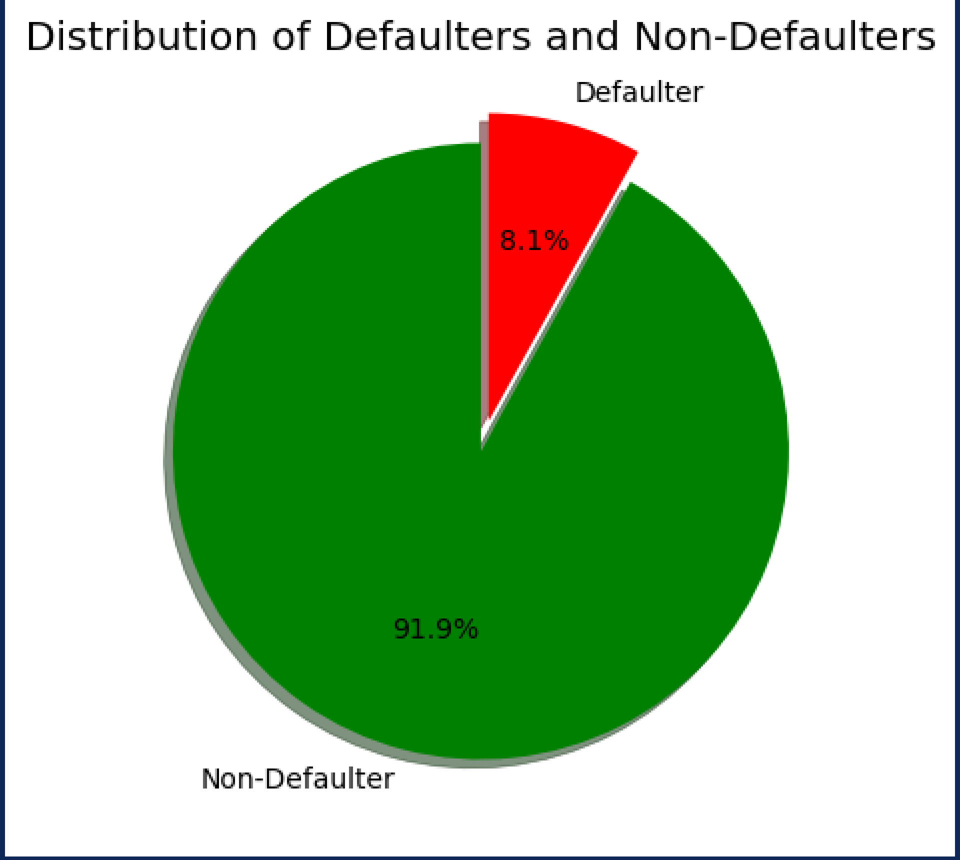
- The objective is to identify and analyze patterns in the data through exploratory data analysis (EDA). This ensures that clients who are capable of repaying loans are not rejected. The bank faces two types of risks:
  - If an applicant is likely to repay the loan, not approving it results in a loss of business for the company.
  - Conversely, if an applicant is unlikely to repay the loan (i.e., likely to default), approving it may lead to a financial loss.

STEPS PERFORMED

- **Data Loading:**
  - We start by loading the two provided datasets into our notebooks for further analysis.
- **Basic Data Analysis:**
  - Conducting initial exploratory analysis helps us understand the information contained in the data.
- **Handling Null Values and Inconsistencies:**
  - We identify null values and inconsistencies within the datasets.
  - Using data cleaning techniques, we remove as many of these issues as possible.
- **Outlier Imputation:**
  - Addressing outliers is crucial. We impute or handle outliers appropriately to ensure robust analysis.
- **Data Imbalance Check:**
  - We assess whether the data suffers from class imbalance.
  - If so, we take necessary steps to address it.
- **Statistical Analysis:**
  - Applying univariate, bivariate, and multivariate analysis techniques provides deeper insights.
  - We also create relevant plots to visualize the data and draw meaningful conclusions.

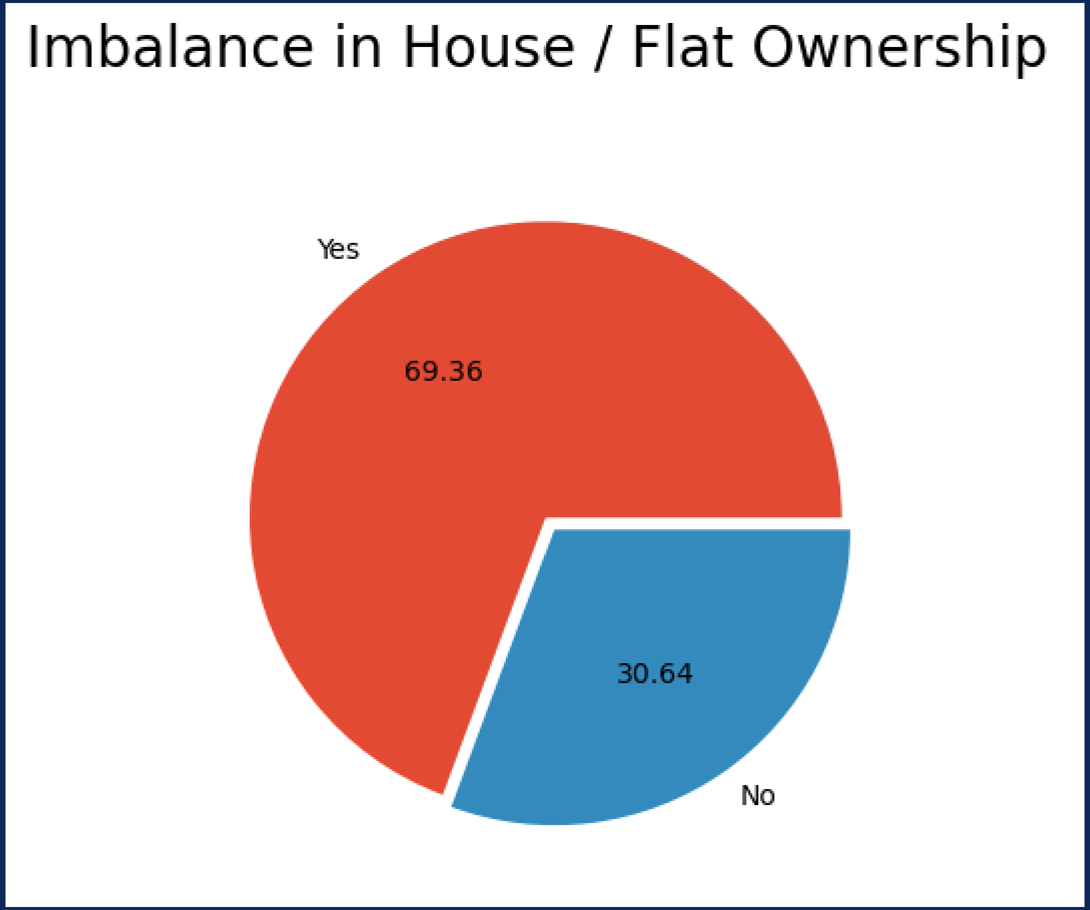
ANALYSING APPLICATION CSV DATA

DATA IMBALANCE : DEFAULTERS VS NON-DEFAULTERS



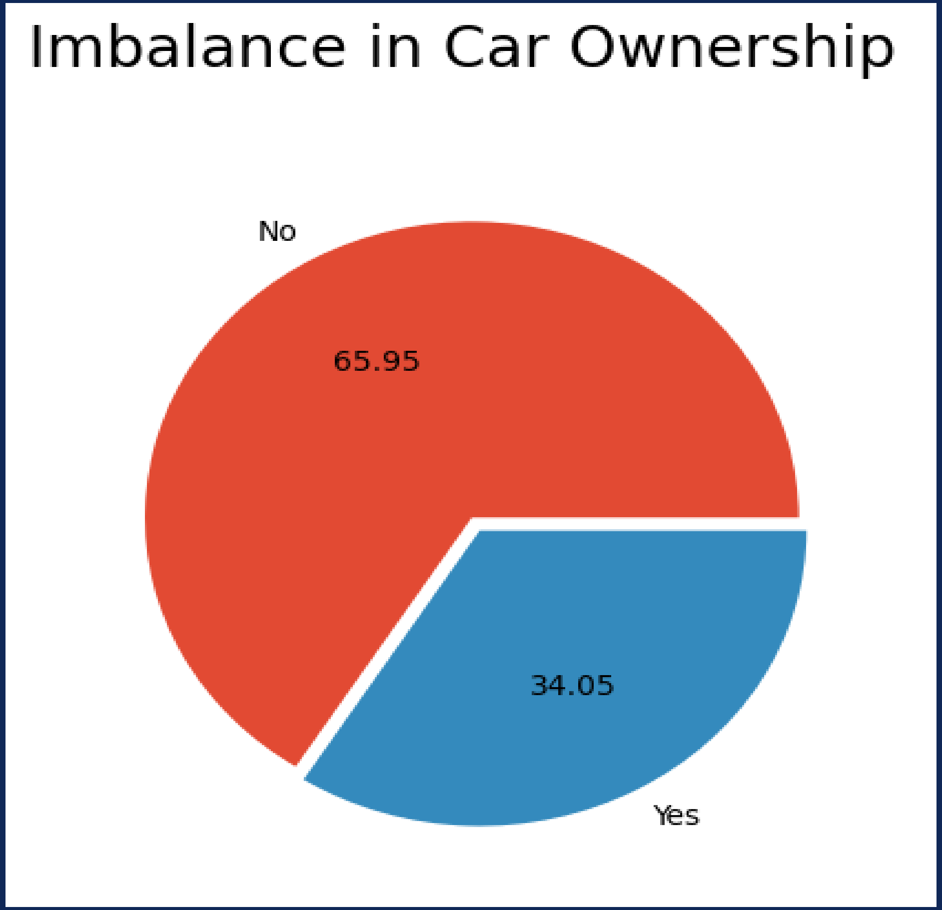
Around 92% people paid their loans on time while 8% defaulted it.

DATA IMBALANCE : HOUSE / FLAT OWNERSHIP



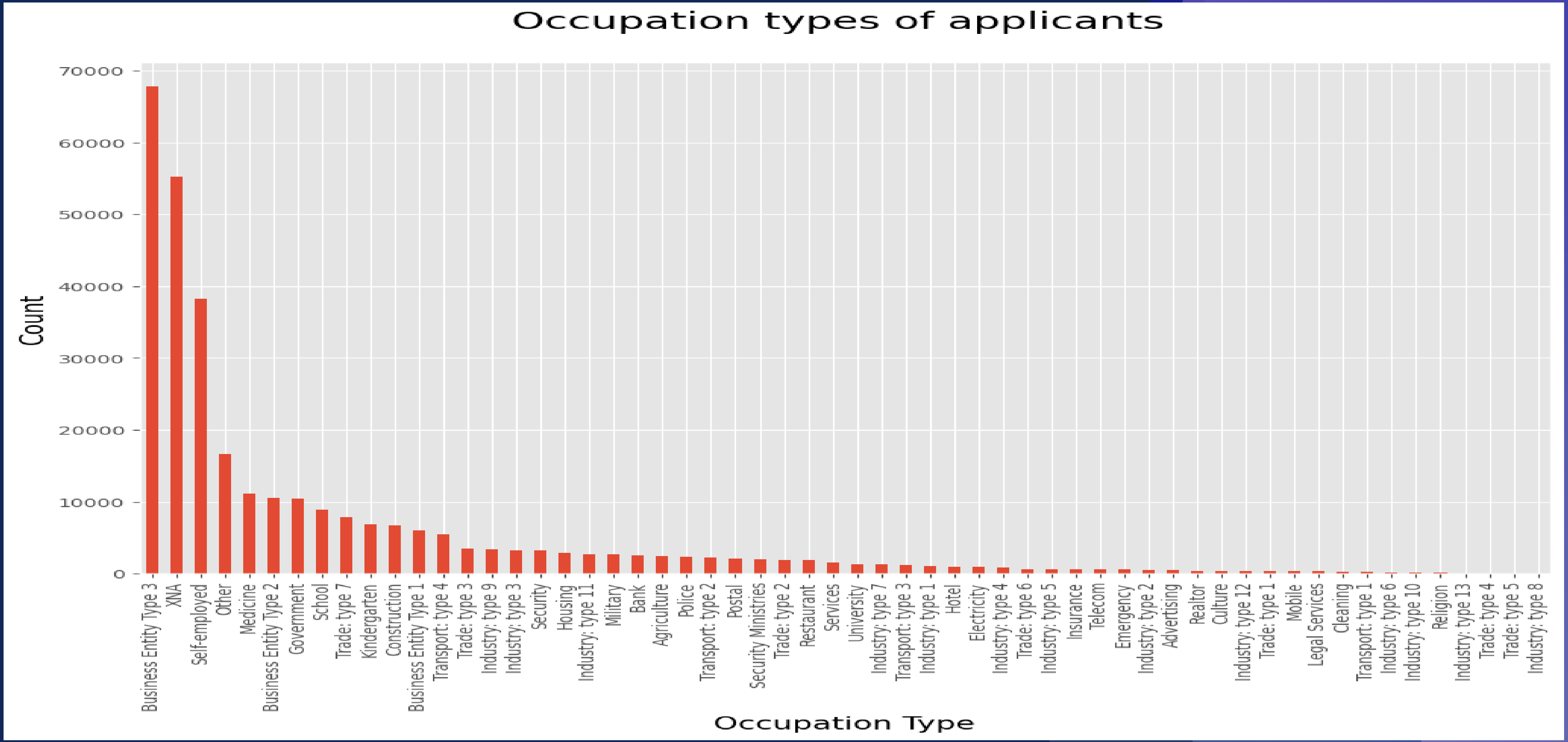
Around 69.36% of clients owns flat while 30.64% of client don't

DATA IMBALANCE : CAR OWNERSHIP



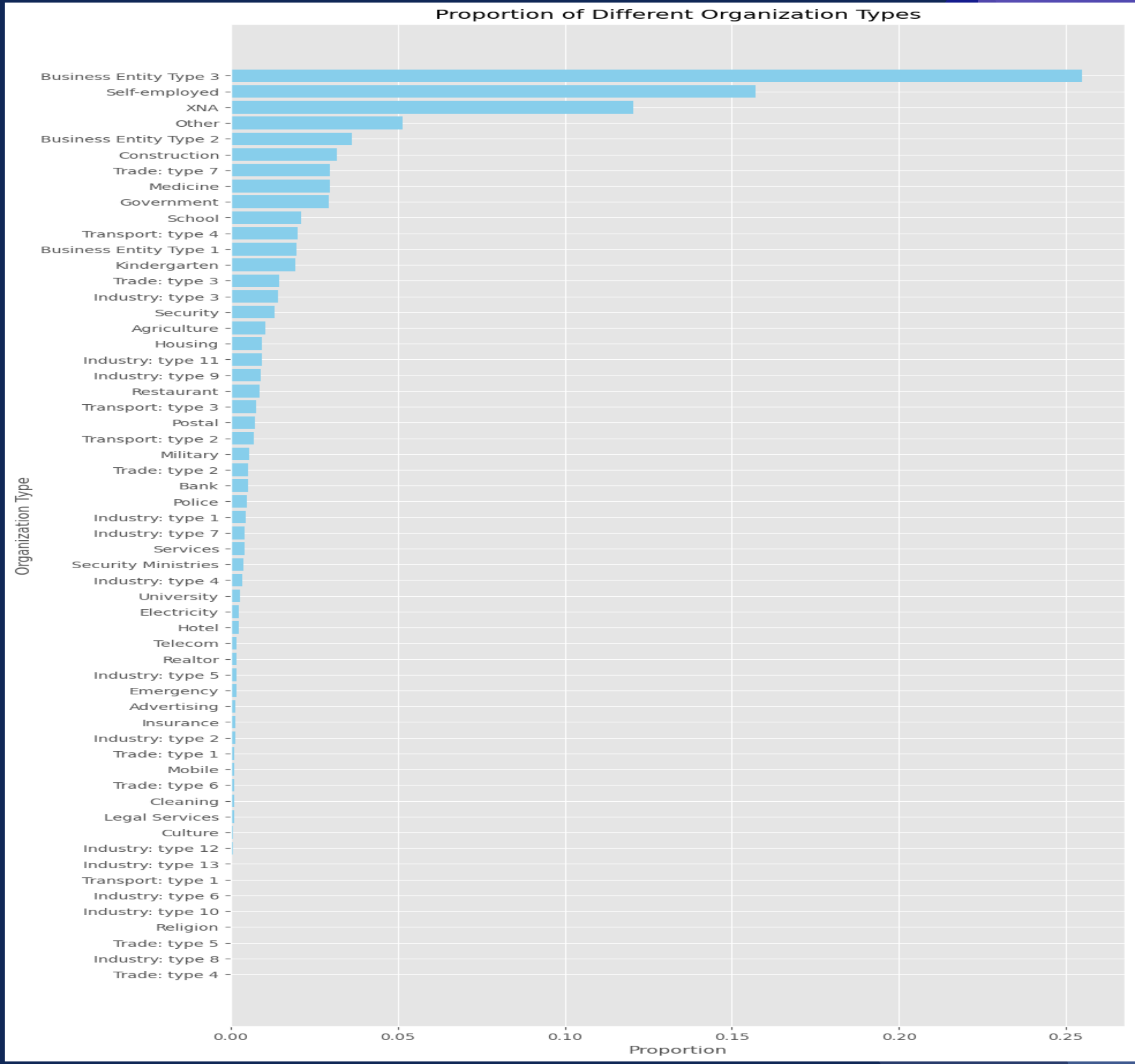
Around 66% of Clients don't own a car while 34% of client does

OCCUPATION TYPES OF APPLICANTS

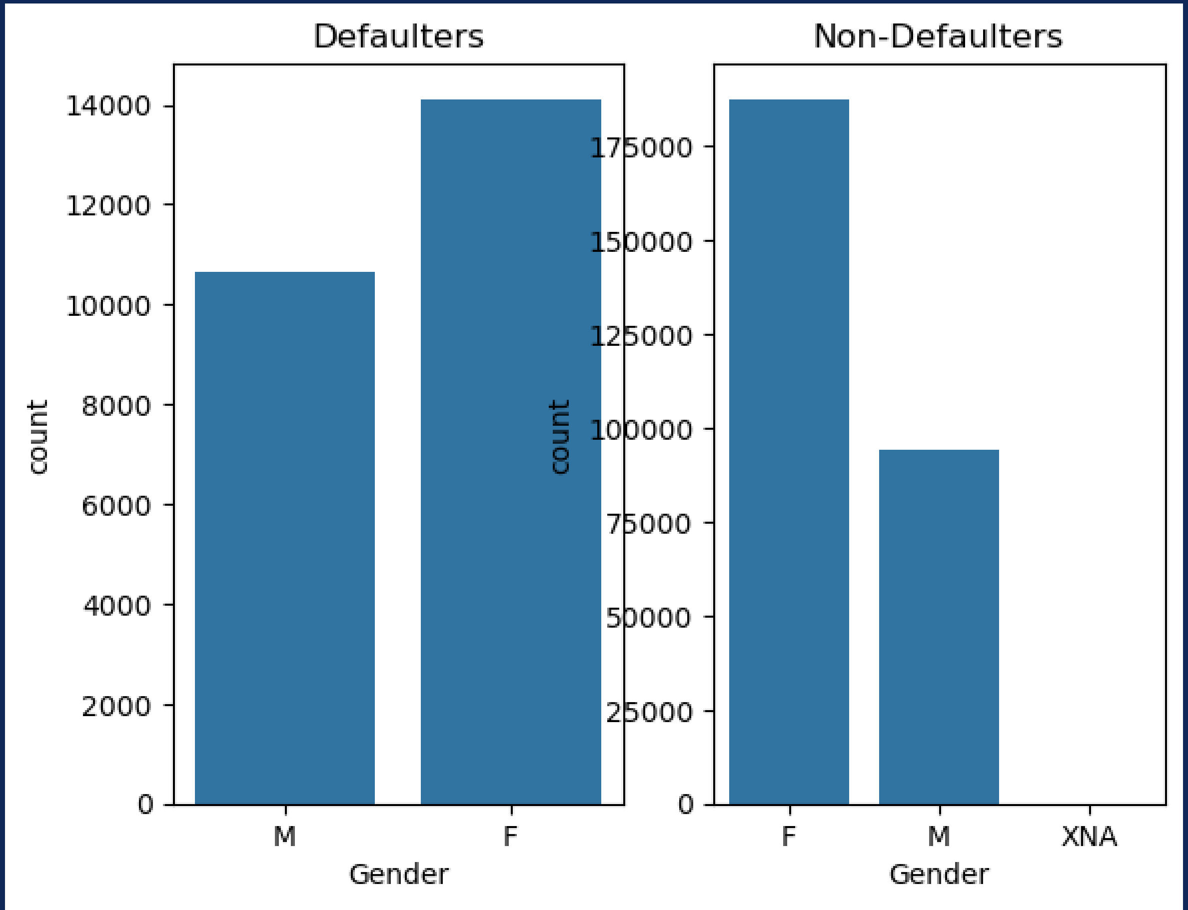


From the plot above, we can see that the majority of loan applications come from 'Business Entity Type 3,' followed by 'Self-employed.' 'Industry: Type 8' has the fewest loan applications.

DIFFERENT ORGANIZATION TYPES



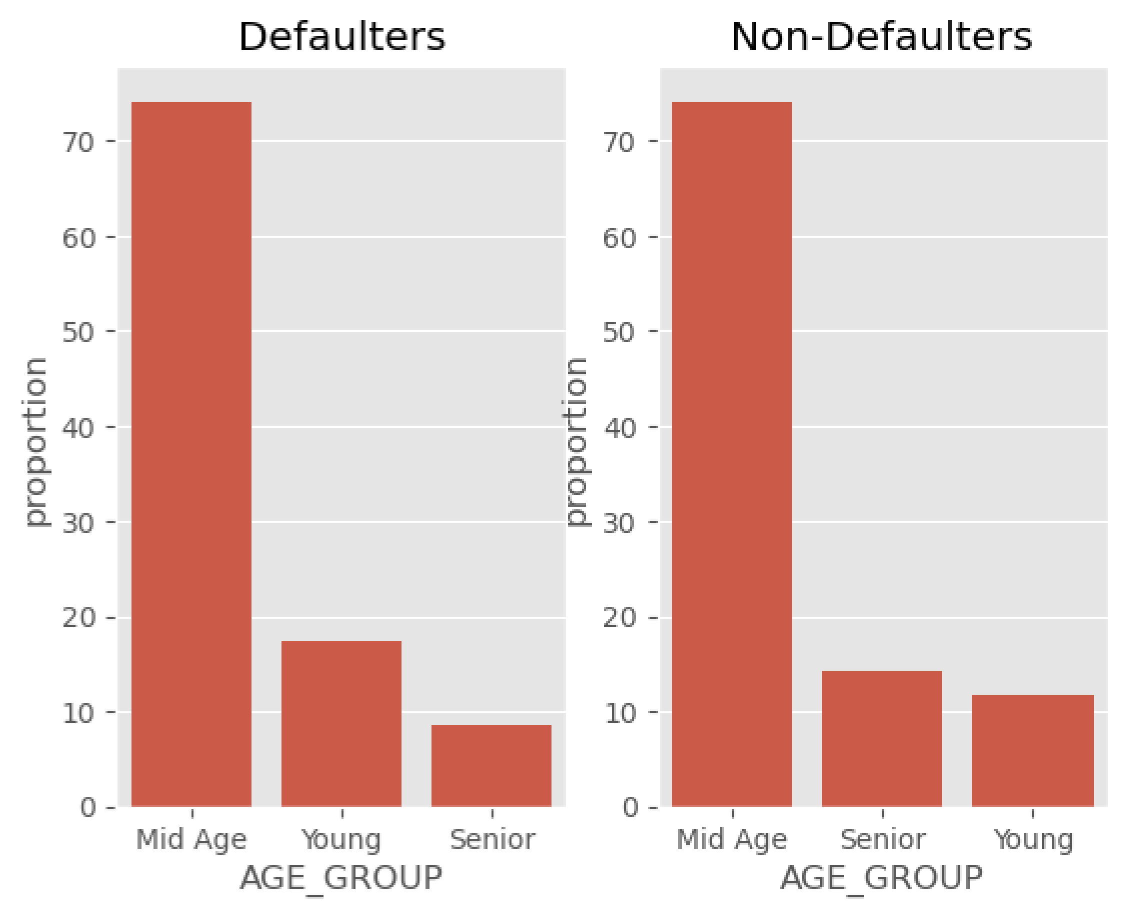
GENDER : DEFAULTERS VS NON-DEFAULTERS



1 Defaulters:
- Among the defaulters, we observe a slightly higher number of females compared to males.
- This suggests that female borrowers are more likely to default on their obligations.

2 Non-Defaulters:
- Similarly, among the non-defaulters, the trend persists: there are more females than males.
- Non-defaulting female borrowers outnumber their male counterparts.

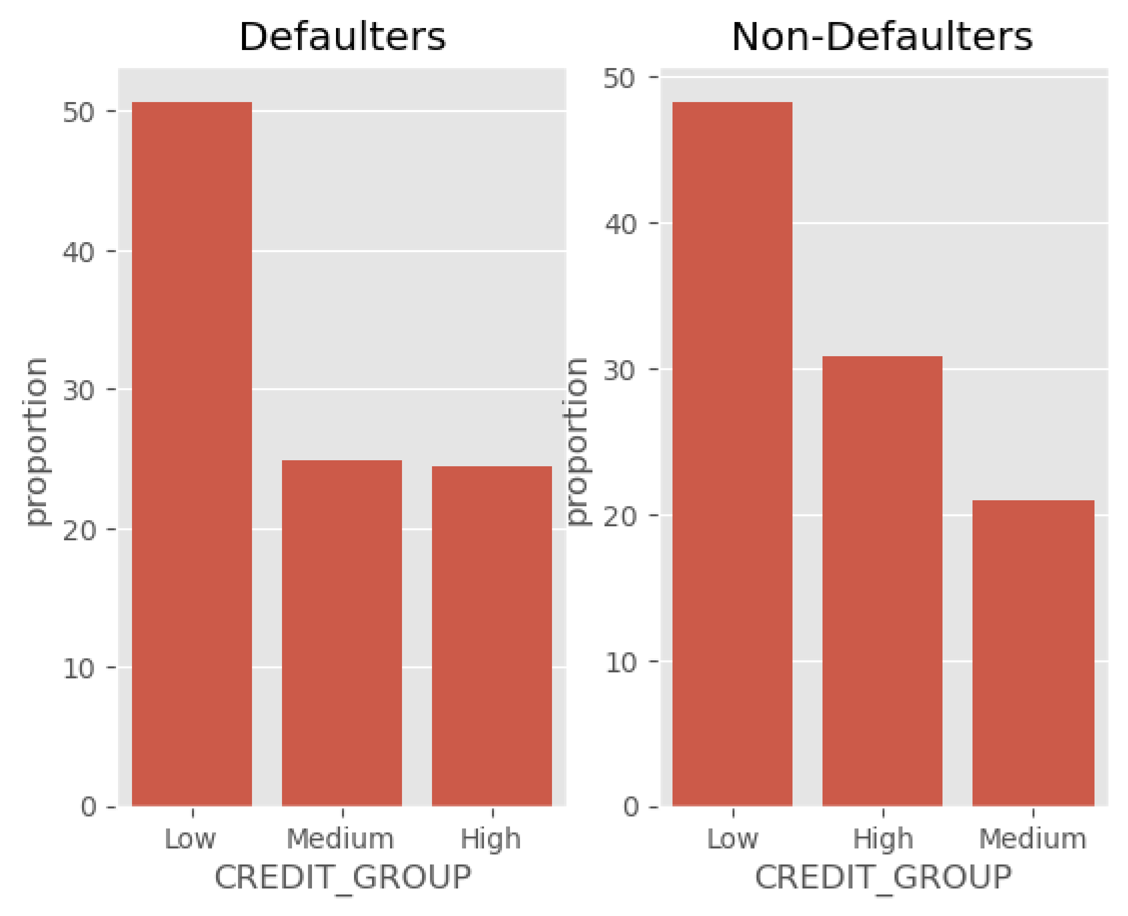
AGEWISE : DEFAULTERS VS NON-DEFAULTERS



1 Defaulters
- Young people are more likely to default compared to the other two age groups. In contrast, senior citizens are less likely to default than others.

2 Non-defaulters
- There is not much difference in the likelihood of non-defaulting across the age groups.

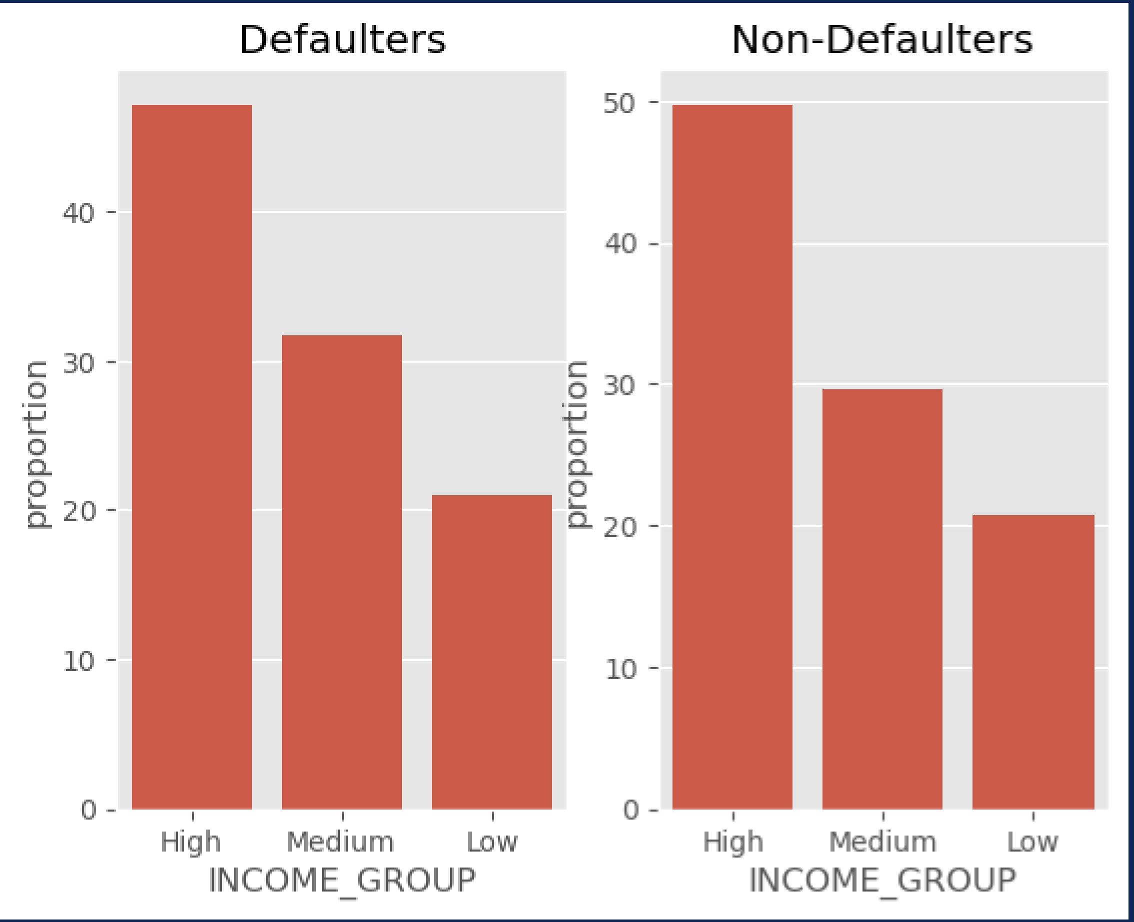
Amount Credit : Defaulters vs Non-Defaulters



1 Defaulters
- People who have AMT_CREDIT lower than ₹100000 have the high chance of defaulting in contrast to other salary brackets, so as the AMT_CREDIT increases there is a high chance of non defaulters.

2 Non-defaulters
- Same pattern seems to have been followed for Non-Defaulters as well.

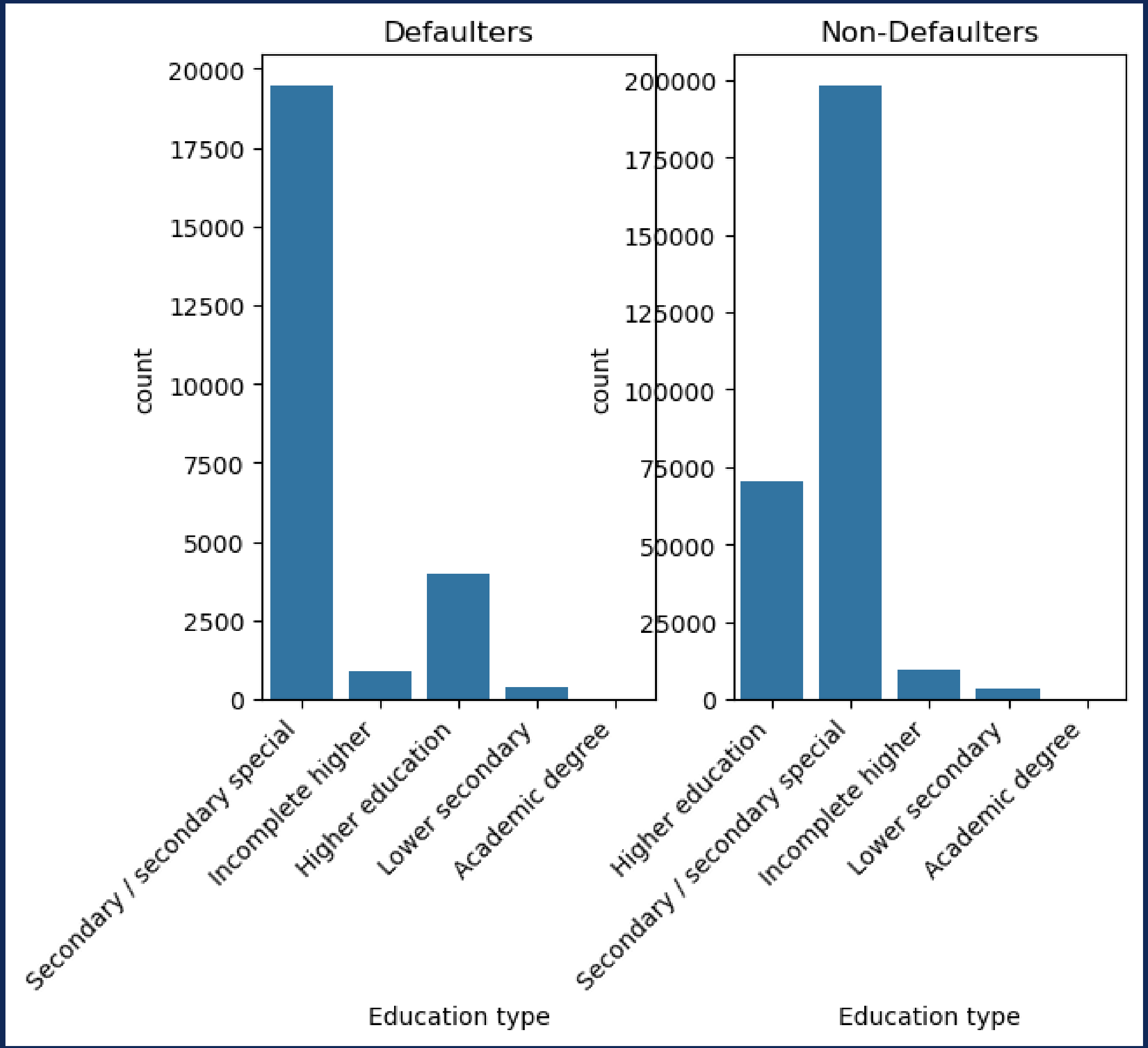
Total Income : Defaulters vs Non-Defaulters



1 Defaulters
- High Salary people have a high chance of defaulting the loan whereas lower the salary, lower is the chance of defaulting.

2 Non-defaulters
- Same pattern seems to have been followed for Non_Defaulters as well.

EDUCATION WISE COMPARISON



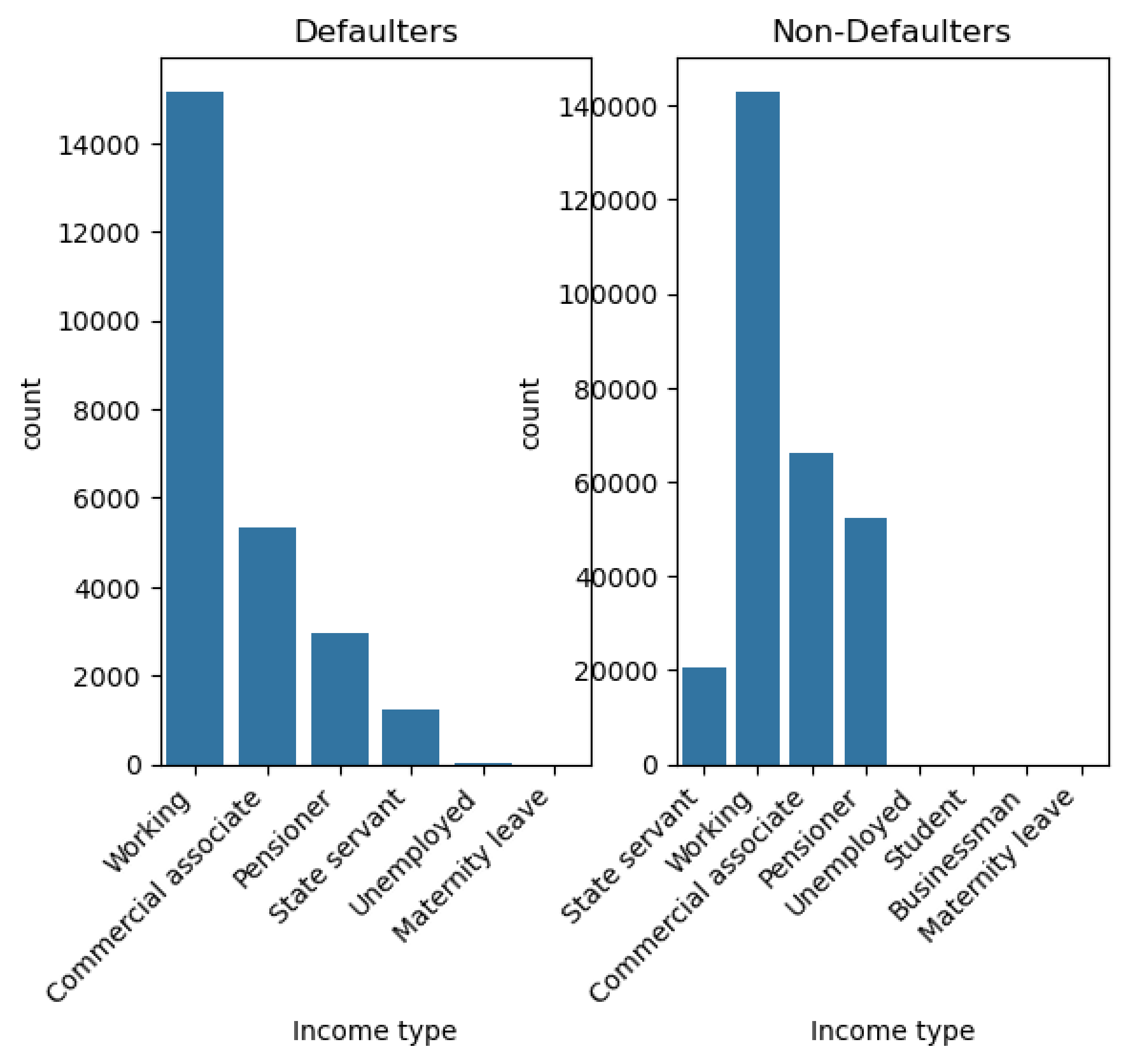
1 Defaulters:

- Among the borrowers who defaulted, a significant proportion had a secondary or secondary special education level.
- This suggests that individuals with lower educational attainment may be more prone to defaulting on their obligations.

2 Non-Defaulters:

- Similarly, among the non-defaulting borrowers, the trend persists: a higher number of individuals with secondary or secondary special education levels.
- Non-defaulting borrowers in this category outnumber those with higher educational qualifications.

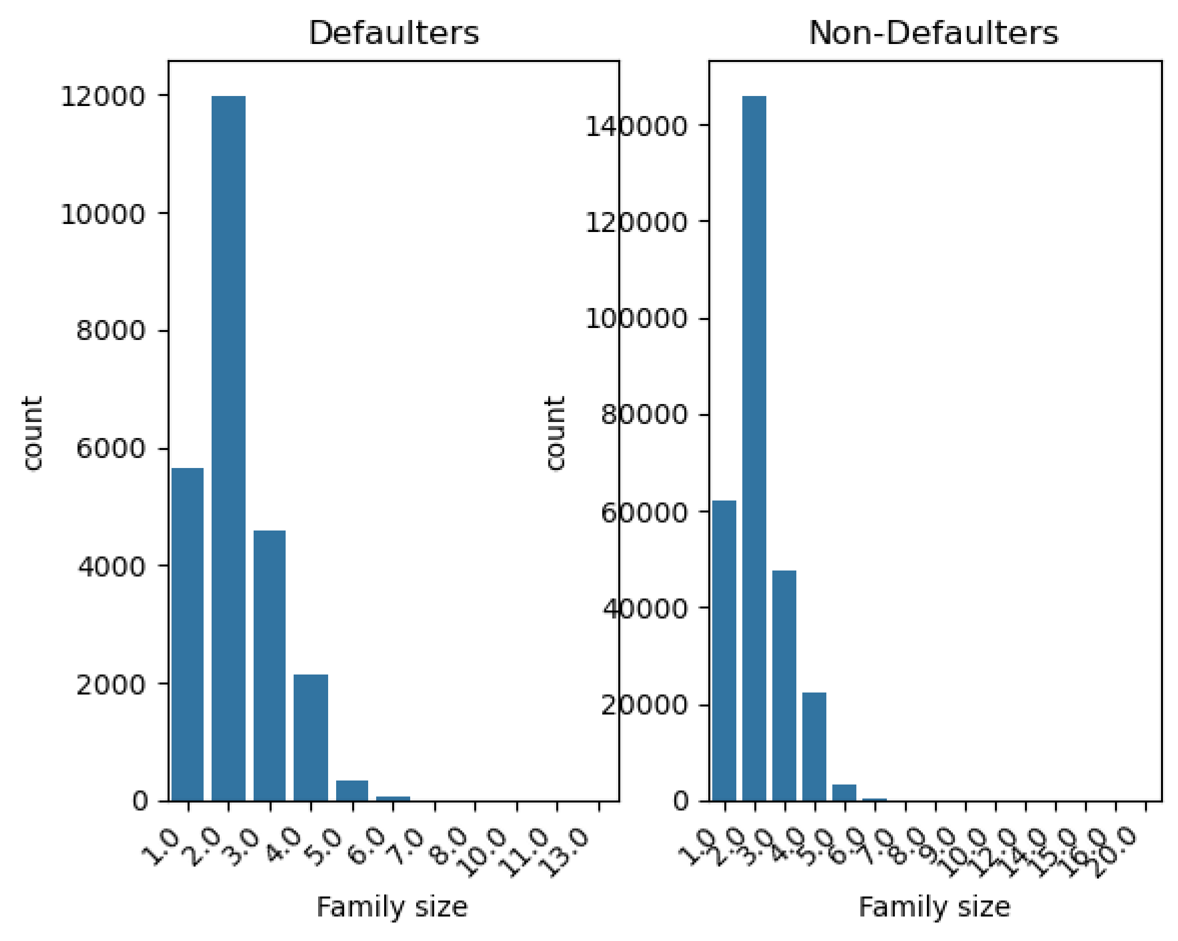
INCOME WISE COMPARISON



1 Defaulters:
- Among the borrowers who defaulted, a significant proportion were working professionals.
- Their higher representation suggests that individuals in the workforce are more likely to default on their obligations.

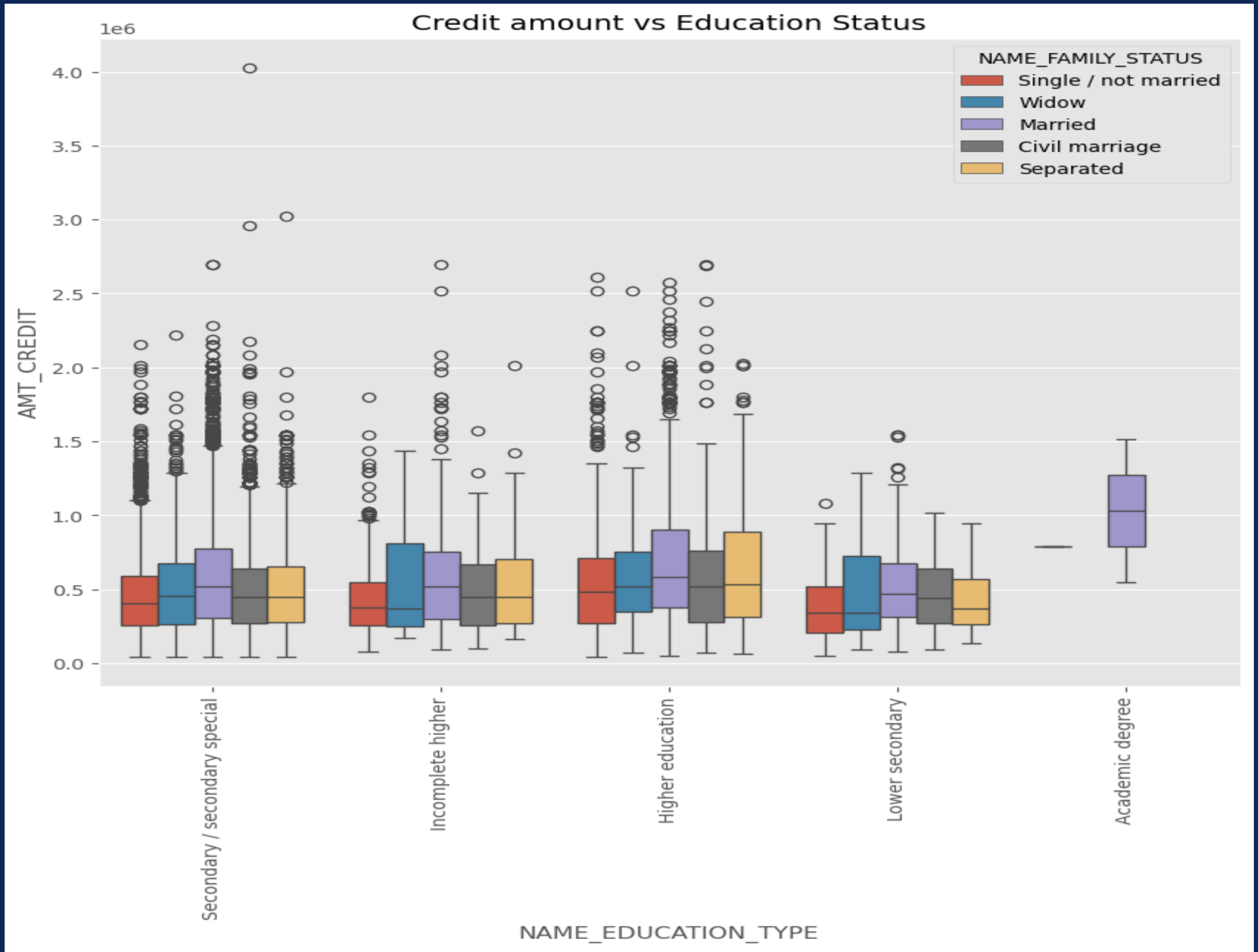
2 Non-Defaulters:
- Similarly, among the non-defaulting borrowers, the trend persists: a larger number of working professionals are defaulters.
- Non-defaulting borrowers in this category outnumber those from other professions.

FAMILY SIZE COMPARISON



Similar kind of pattern seems to be there in both the family sizes of defaulters and non-defaulters
We can see that a family size of 2 has the chance of non-defaulting as well as defaulting the loans.

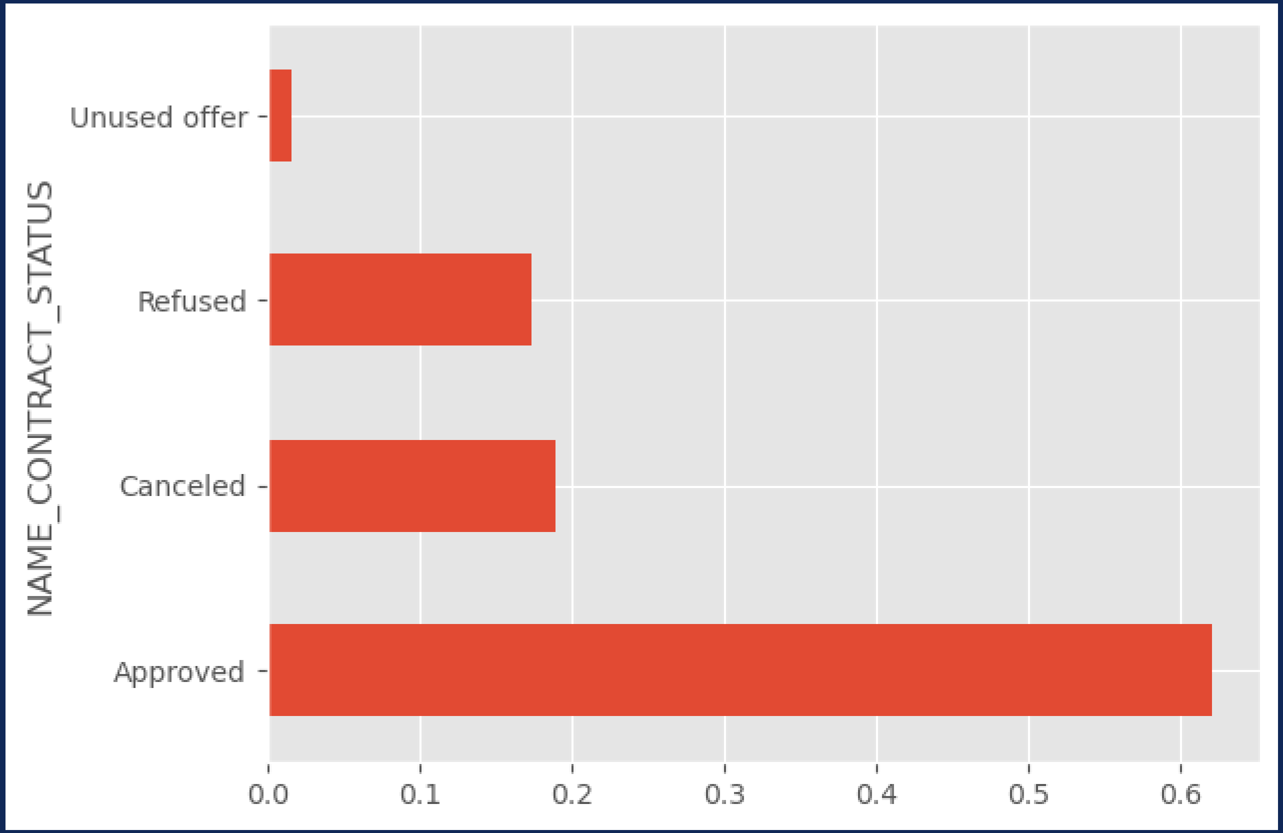
BIVARIATE ANALYSIS: CREDIT AMOUNT VS EDUCATION STATUS



From the box plot, we can conclude that individuals with a family status of "married" who have an academic degree tend to have a higher number of credits than others. Additionally, those with higher education and a family status of 'marriage', 'single', and 'Widow' exhibit more outliers.

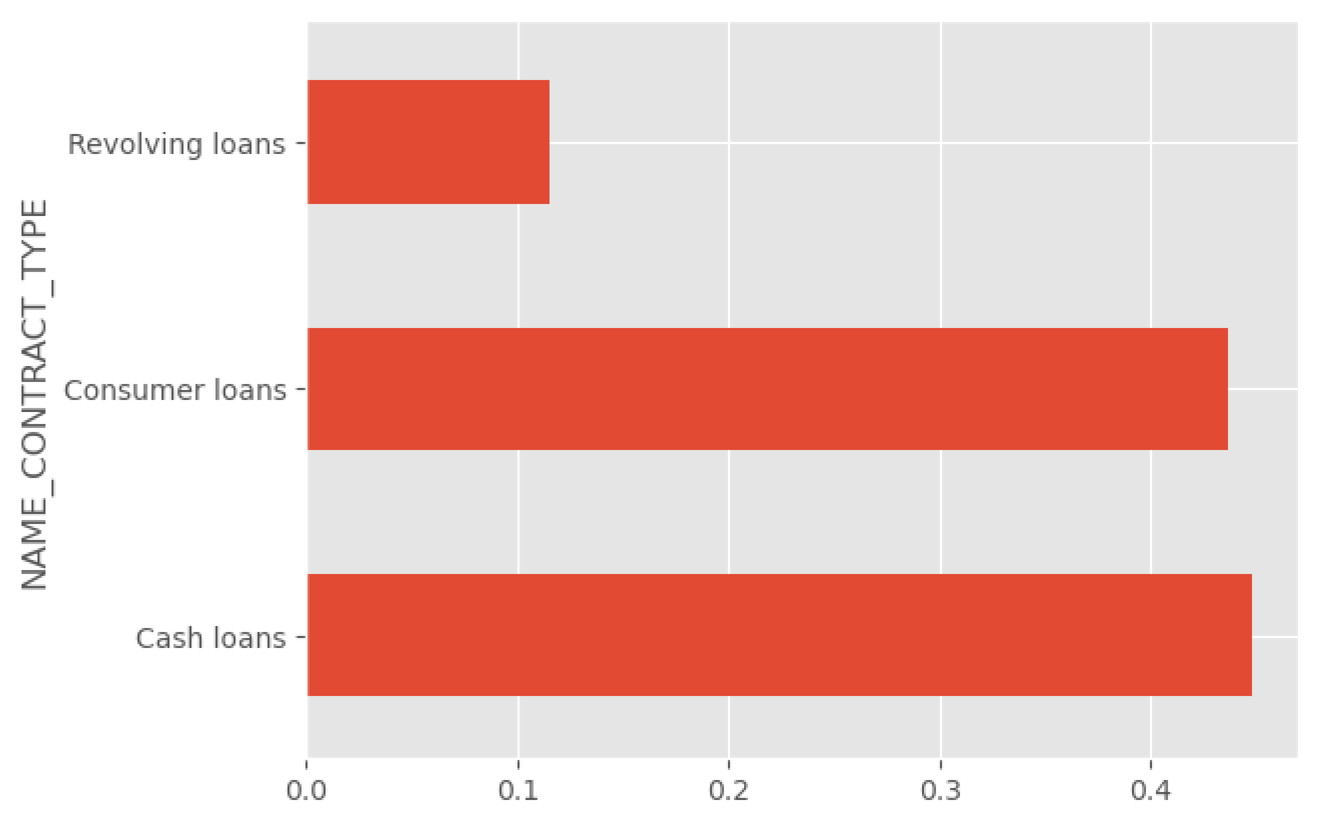
ANALYSING PREVIOUS APPLICATION CSV DATA

ANALYSING NAME CONTRACT STATUS



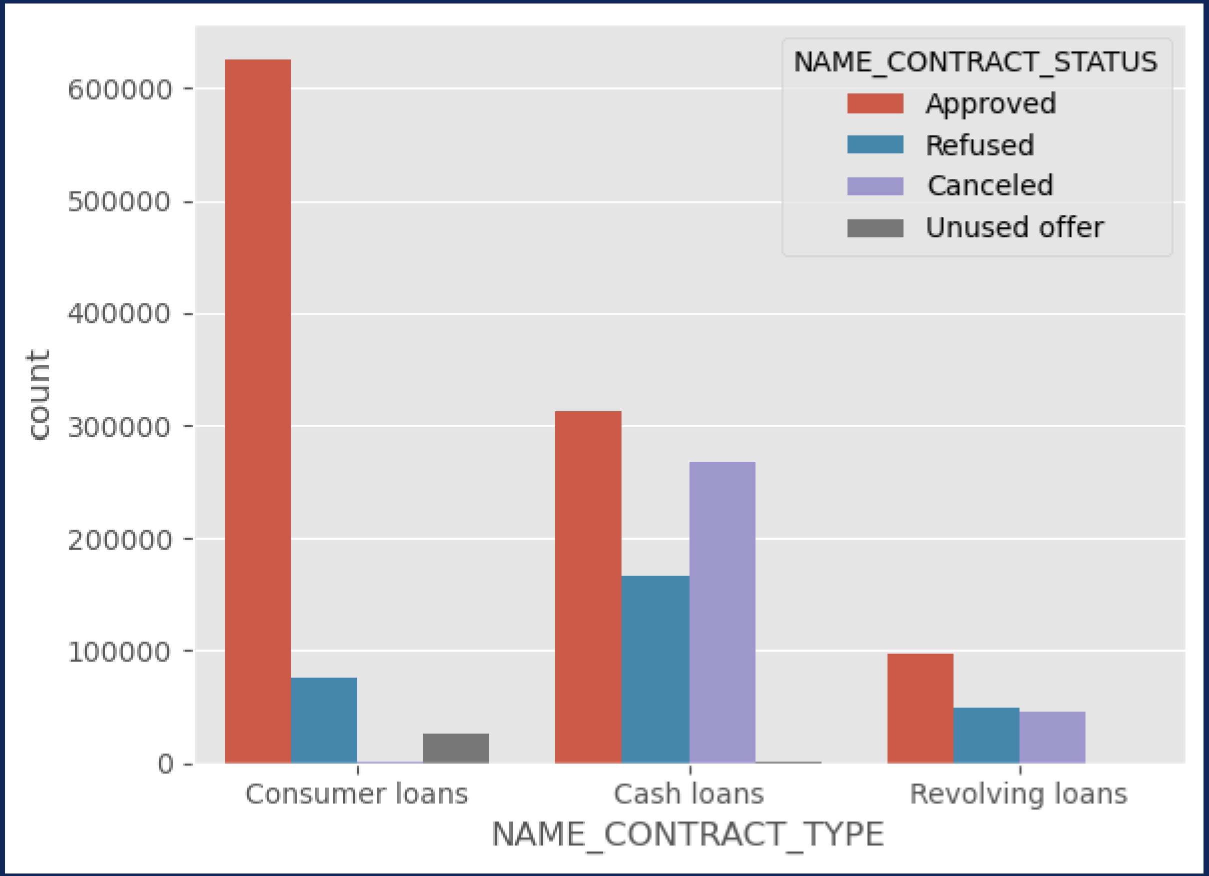
We can see that most of the loans were Approved while the count of Unused offers was extremely low

ANALYSING NAME CONTRACT TYPE



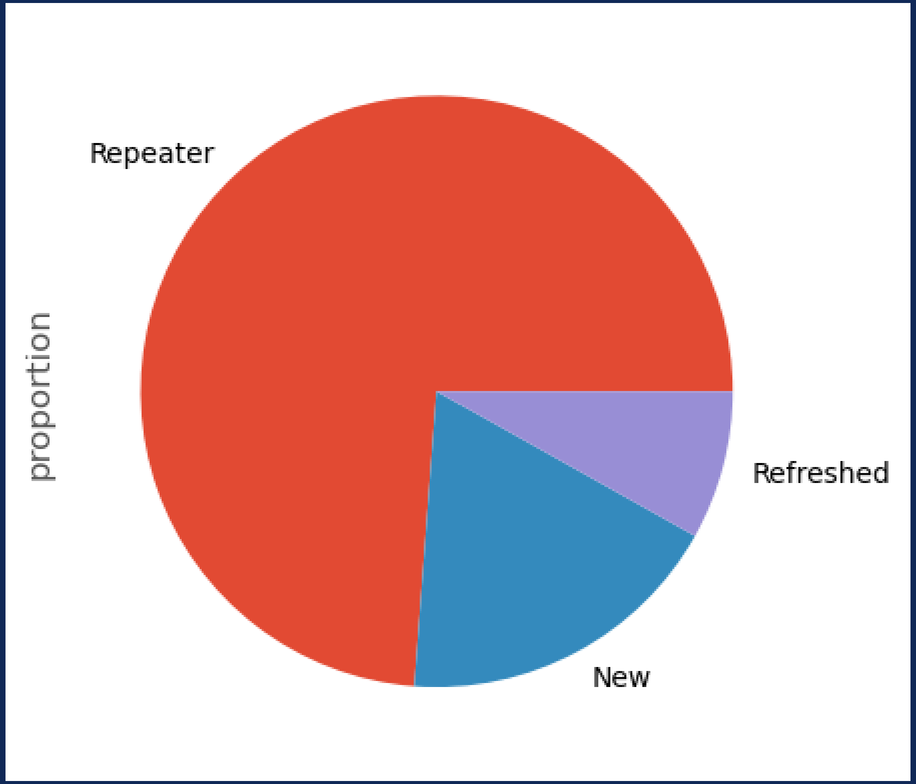
Most of the clients prefer Consumer loans and Cash loans over Revolving loans

ANALYSING NAME CONTRACT TYPE WITH STATUS



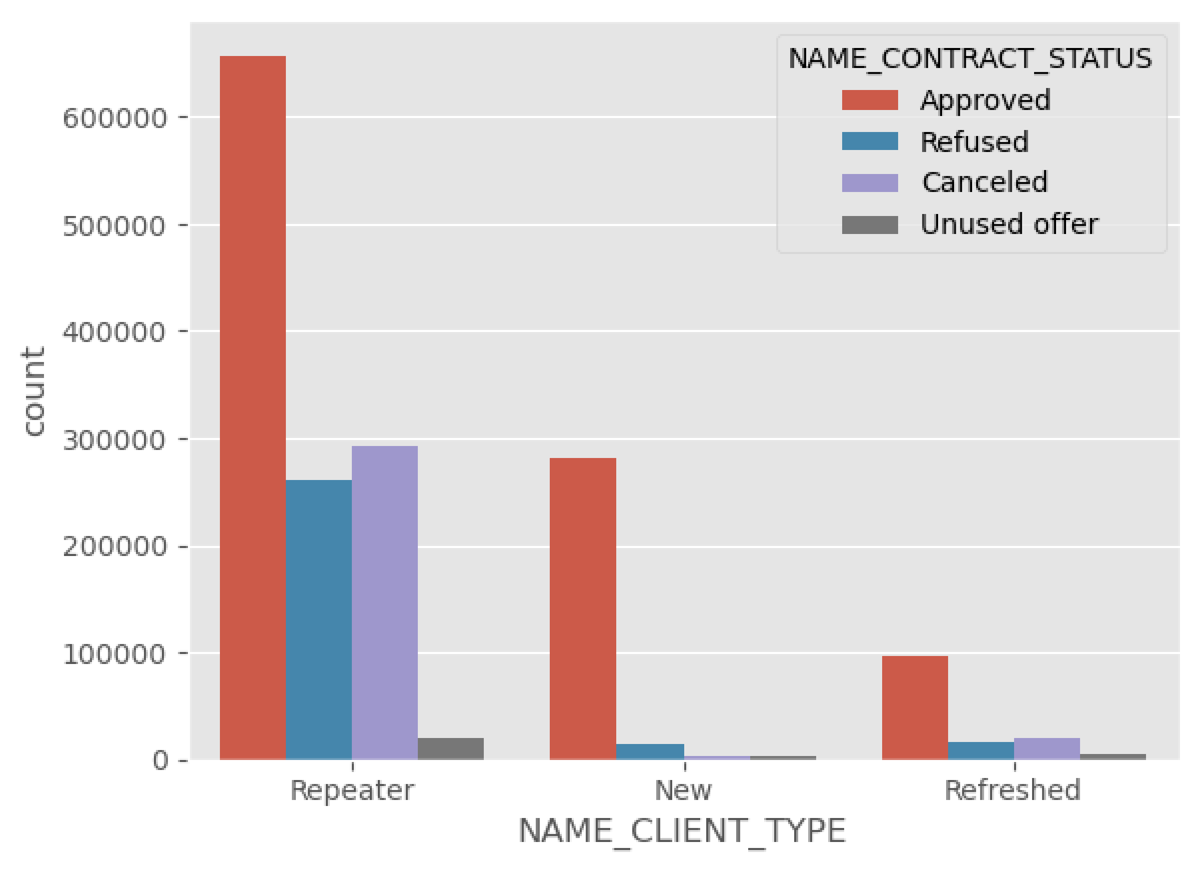
We can analyze that
- Most of the Consumer loans are approved in comparison to Refused loans.
- For Cash loans most of the loans were approved but comparatively loans were cancelled and refused as well.

ANALYSING NAME CLIENT TYPE



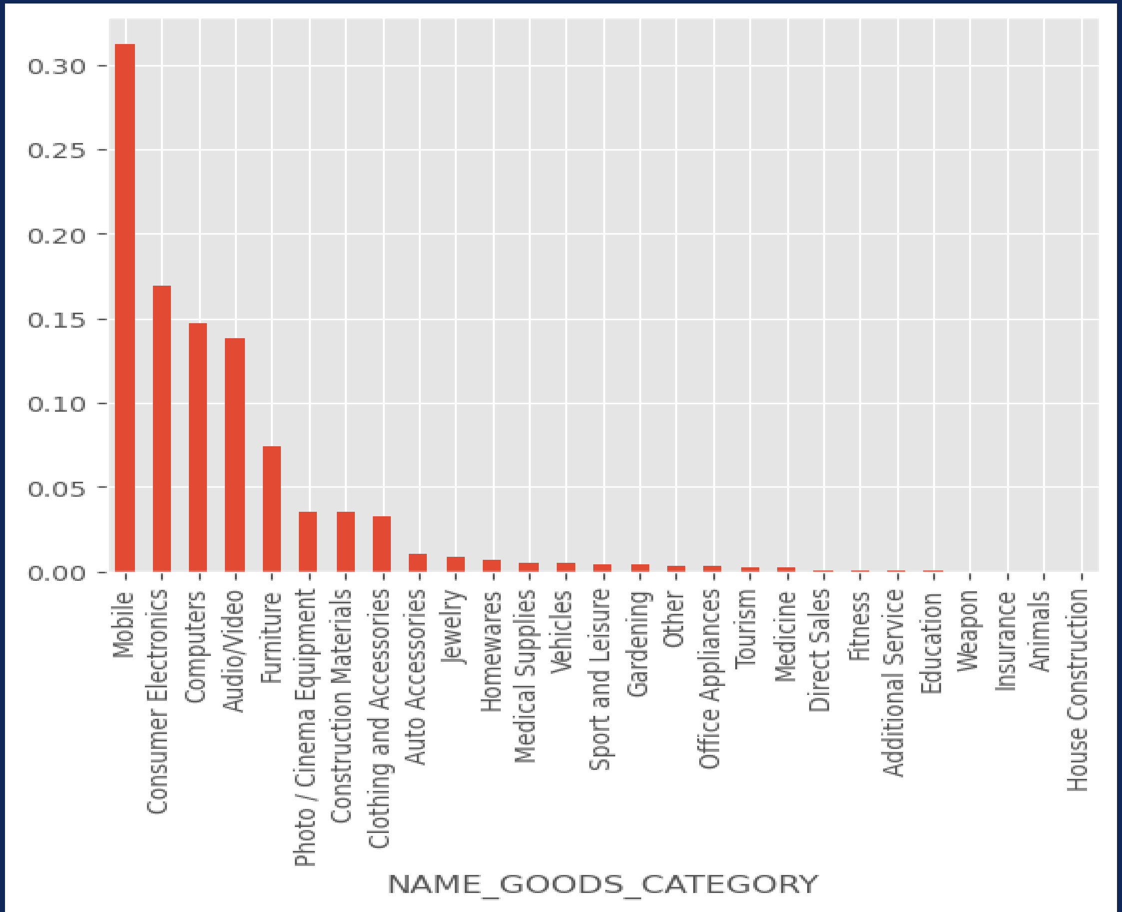
Most of the loan applications are from the Repeater client and the New Clients rather than refreshed clients

ANALYSING NAME CLIENT TYPE VS CONTRACT STATUS



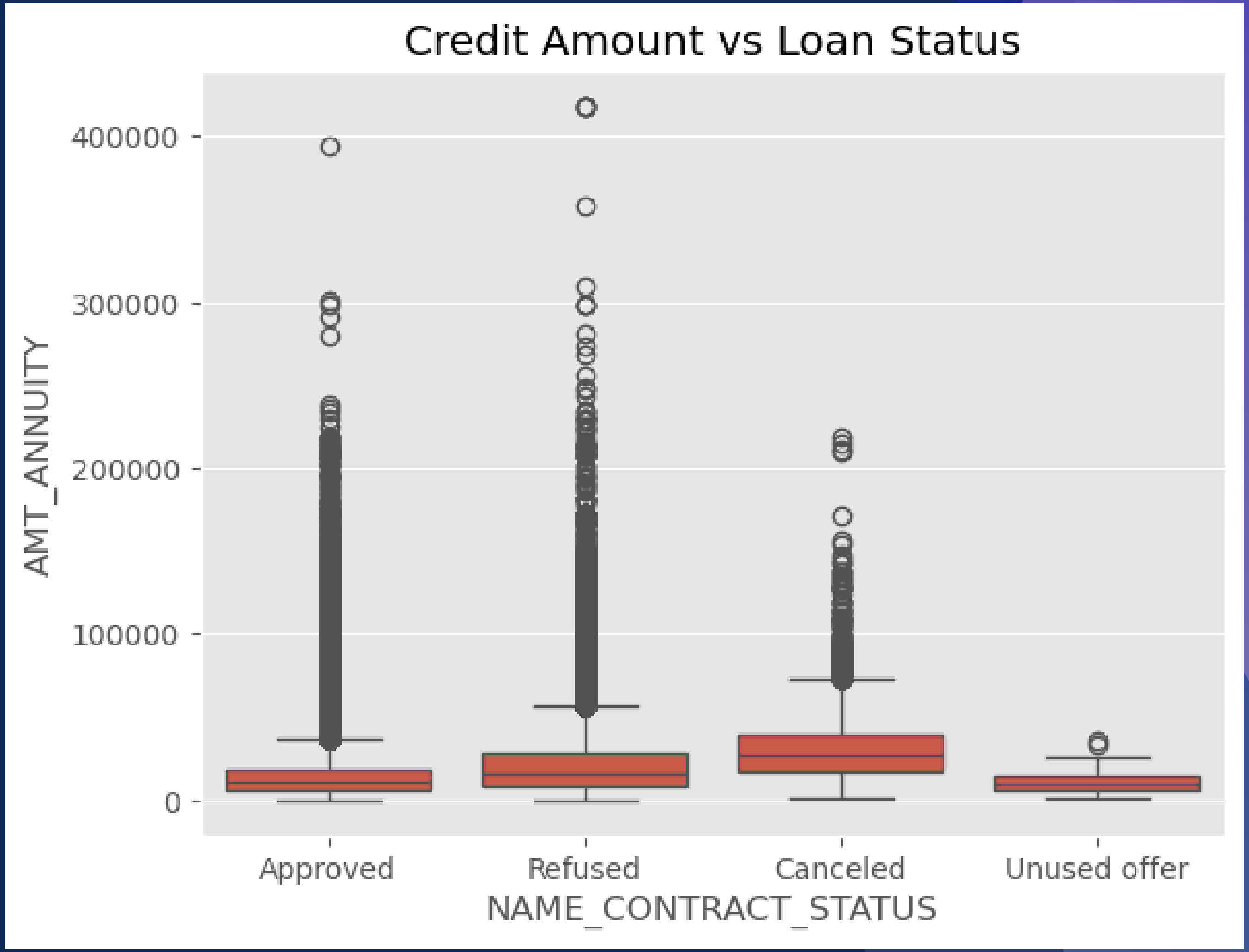
Most of the applications are from Repeater clients get approved, yet some are cancelled or denied. New clients mostly see their applications approved, with only a few denied, cancelled, or unused. This trend is similar for Refreshed clients. In summary, regardless of client type, most applications are approved.

ANALYSING NAME GOODS CATEGORY



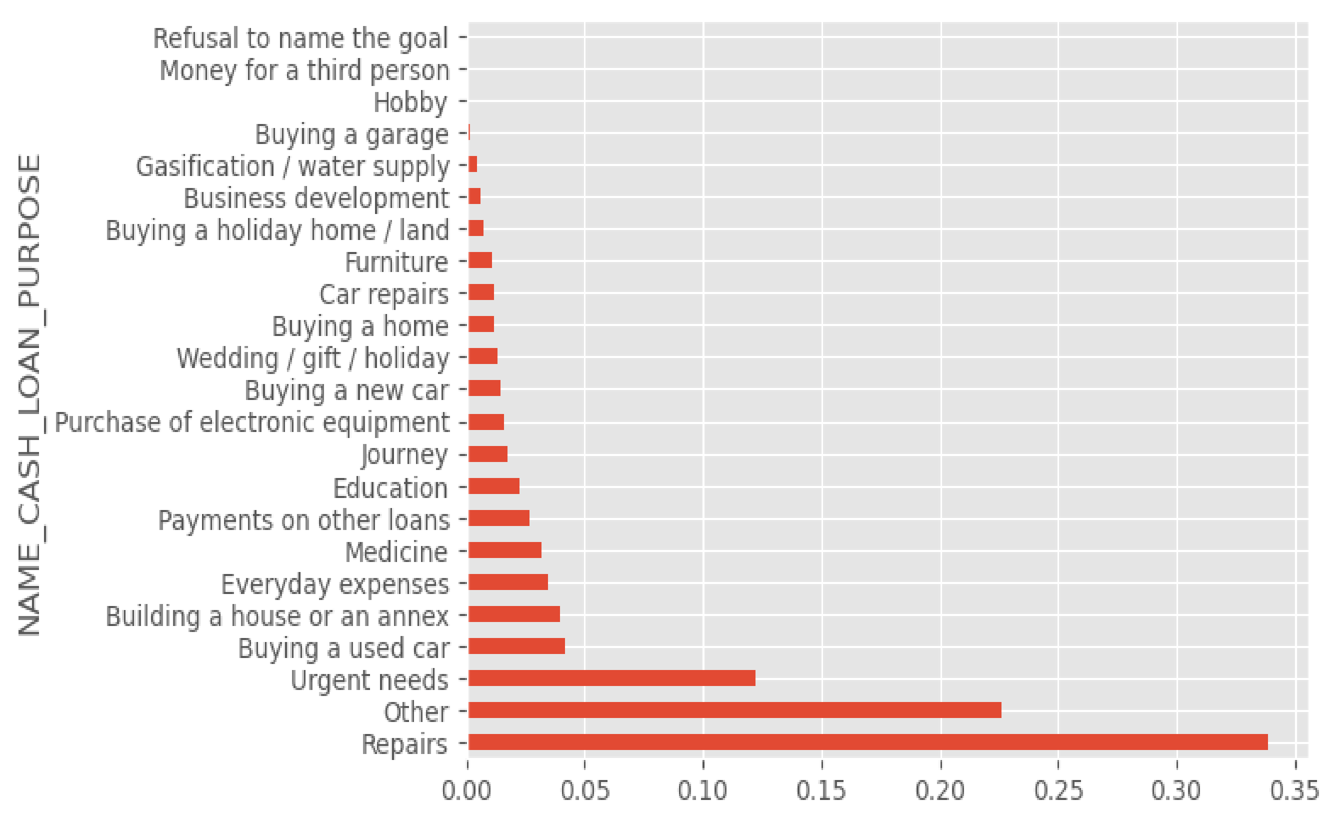
Majority of the loans were approved for the Mobile loans followed by Consumer Electronics and Computers, while Insurance, Animals and House Construction were the lowest.

BIVARIATE ANALYSIS: NAME_CONTRACT_STATUS VS AMT_ANNUITY



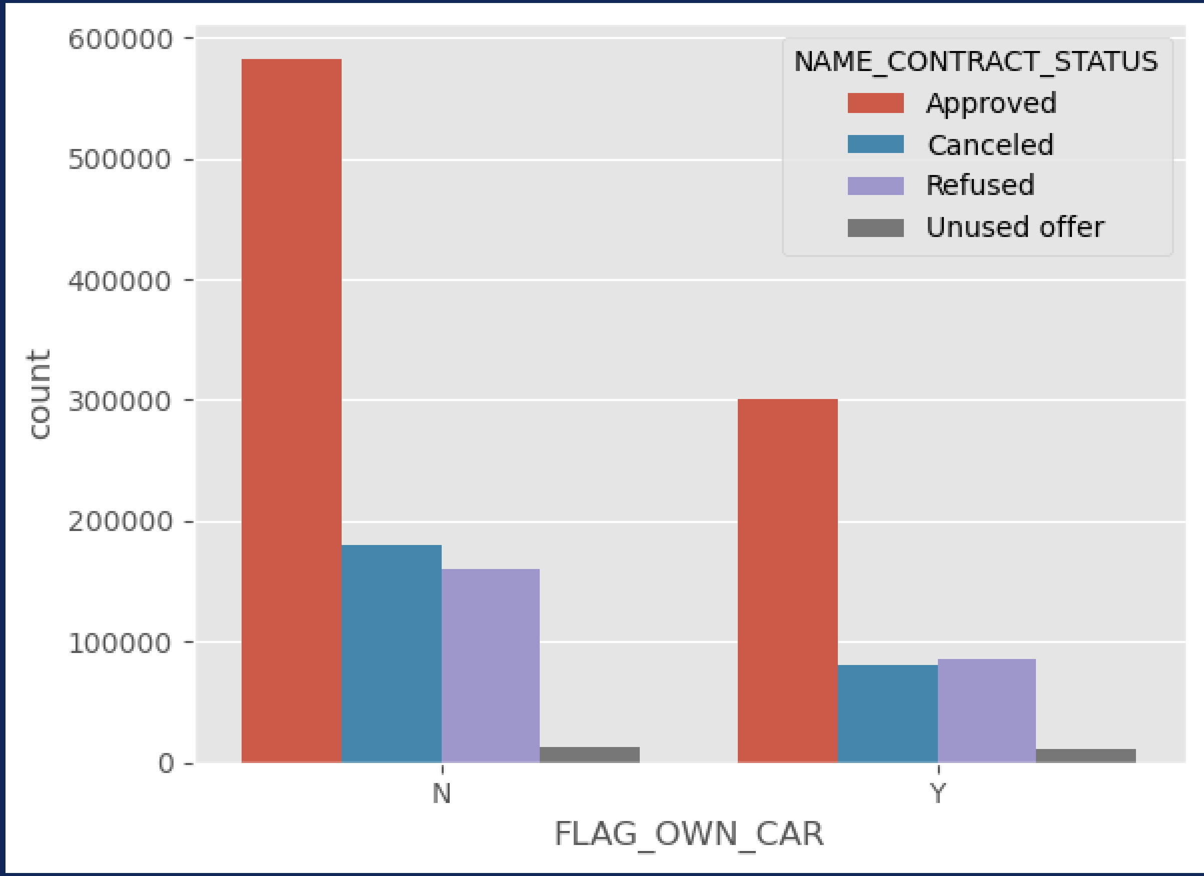
ANALYSING COMBINED CSV DATA

ANALYSING NAME CASH LOAN PURPOSE



Most cash loan applications are submitted for 'Repairs.' The second most common purpose is 'Others,' followed by 'Urgent Needs.' There are relatively few applications where the purpose of the loan is not disclosed. The least common reasons for cash loans are 'Money for a third person' and 'Hobby.'

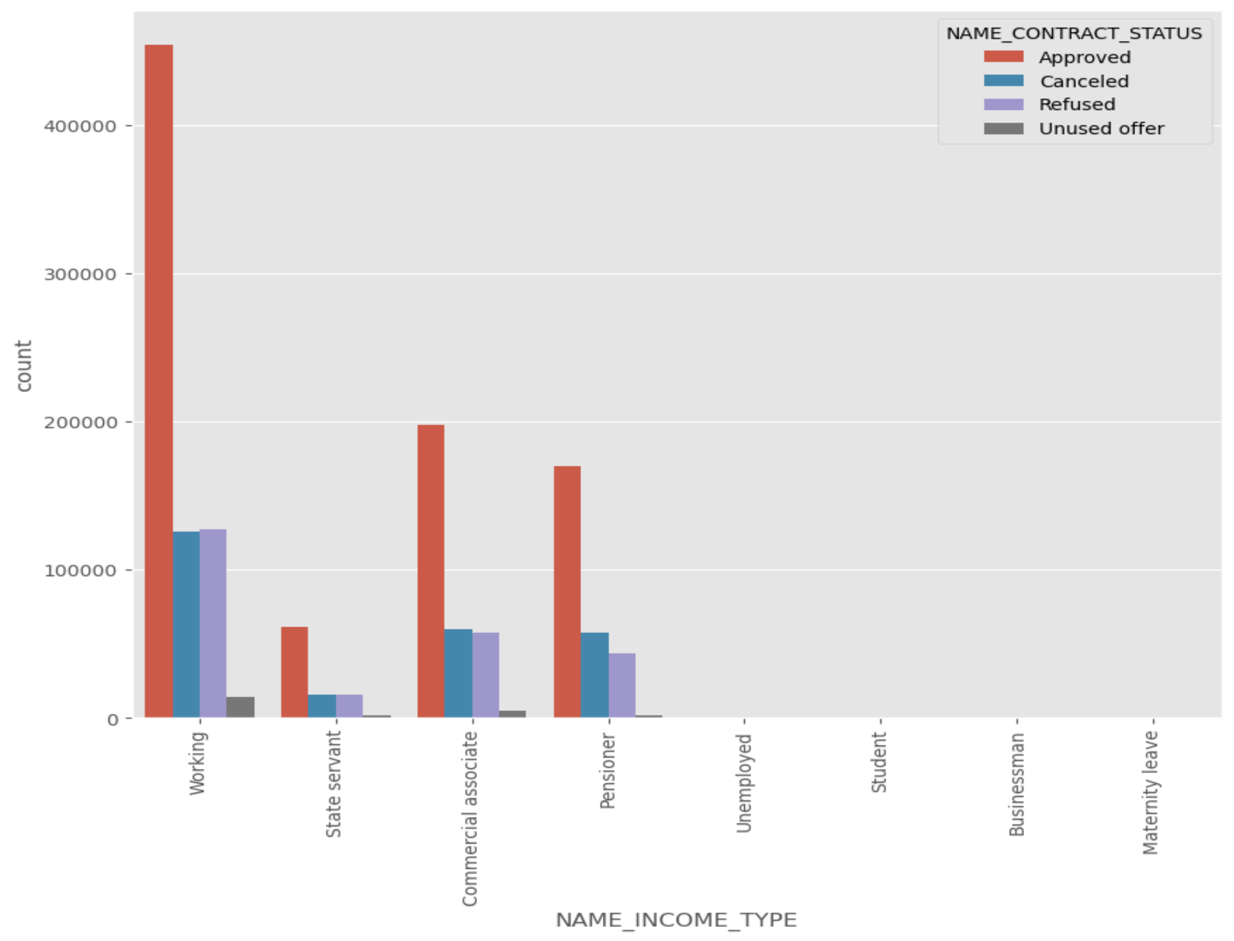
FLAG_OWN_CAR VS NAME_CONTRACT_STATUS



Even if the client owns or not owns the car there is not much difference in terms of Approved loan status

but one thing can be observed that the people who don't own a car seems to apply for loan more than the ones who own the car

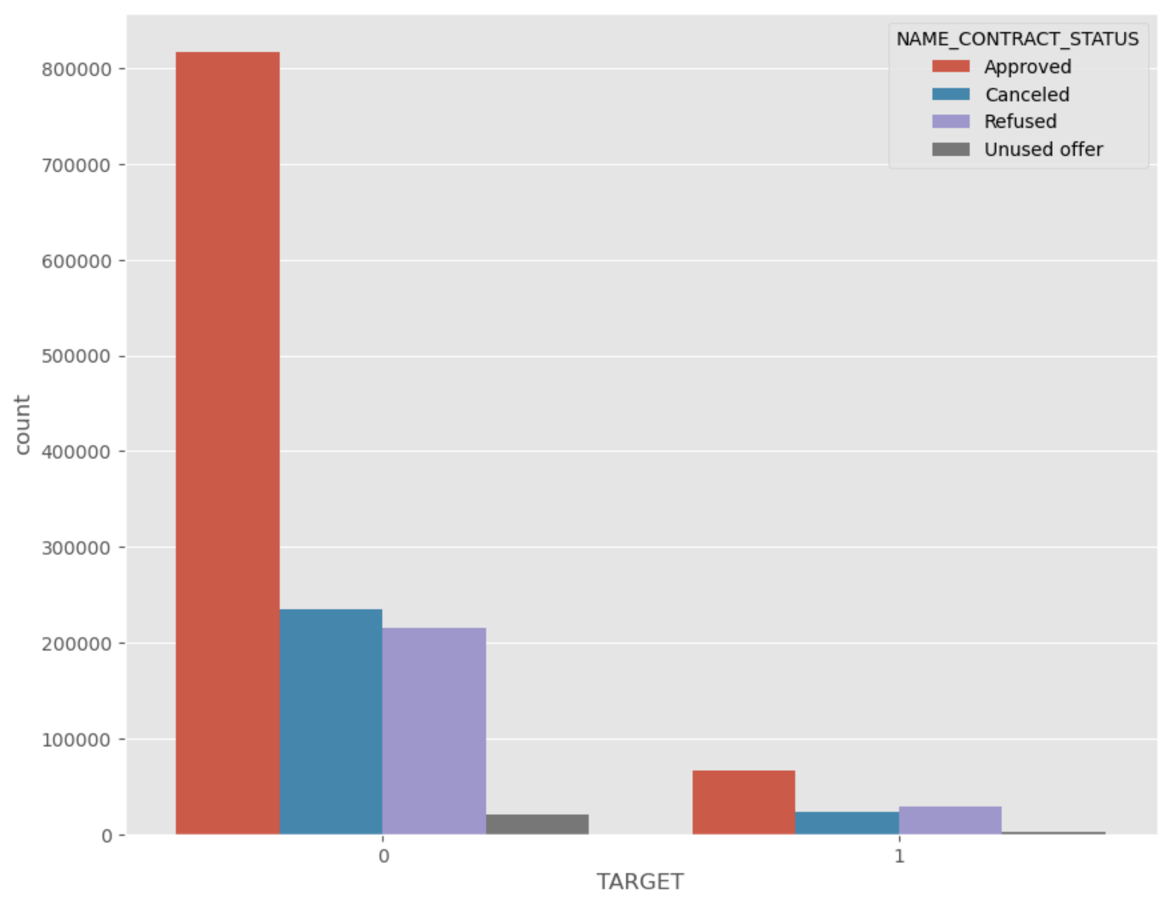
NAME_INCOME_STATUS VS NAME_CONTRACT_STATUS



Couple of inferences can be made from the chart

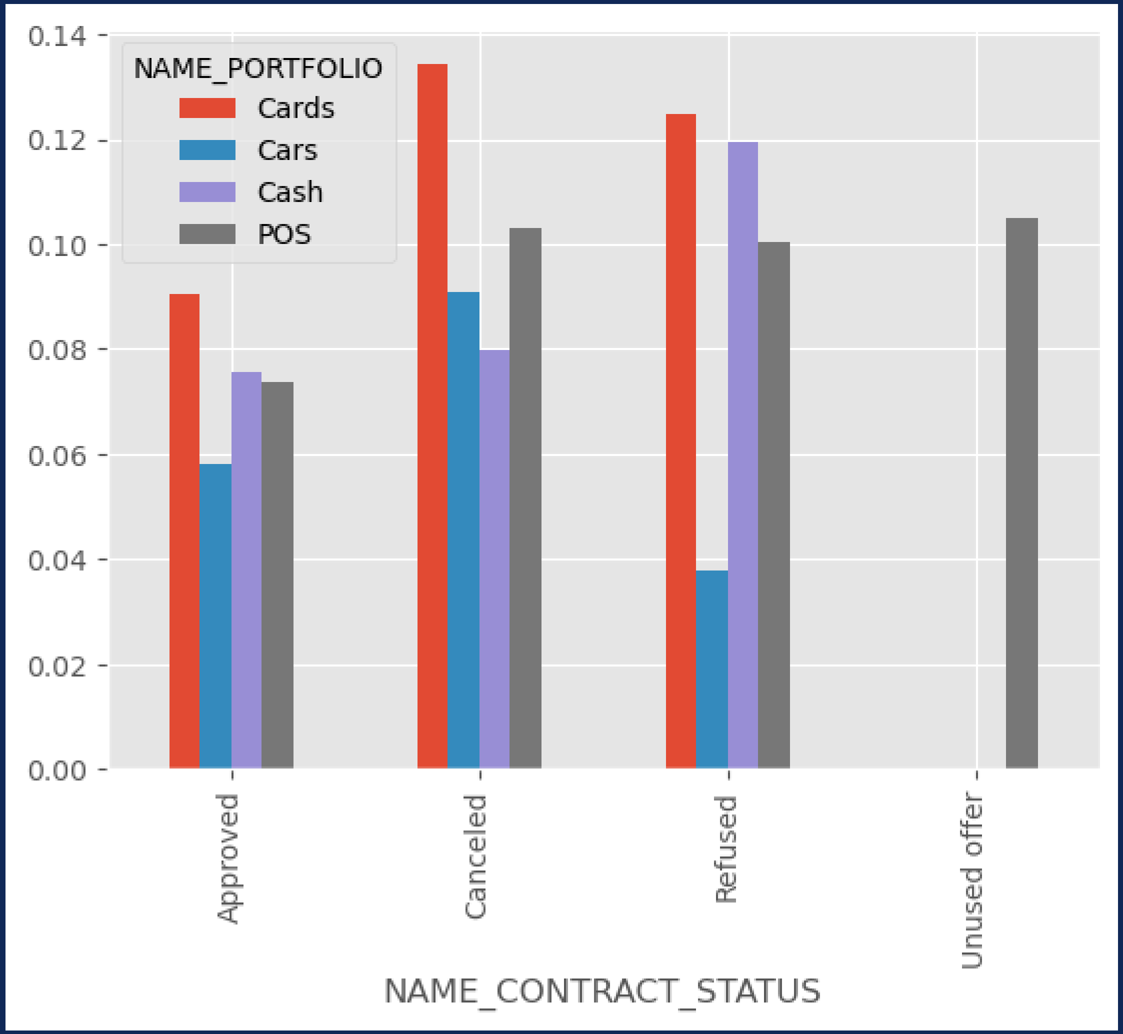
- People who work and earn tends to apply for loans the most and there is a high chance that the loan gets approved
- Unemployed, students, businessman and Maternity leave people don't have the tendency to apply for loans
- People who are earning apply for loans and they get approve there is less chance comparatively that the loan is cancelled or refused.

TARGET VS NAME_CONTRACT_STATUS



Inferences we make from the above chart
- There are a smaller number of defaulters compared to non-defaulters
- Maximum of the loans approved are paid on time and they come under non-defaulters category
- Also we can see a lot of loans for non defaulters are refused or cancelled which ultimately results in loss of business for the company.

MULTIVARIATE ANALYSIS: NAME_PORTFOLIO VS NAME_CONTRACT_STATUS VS TARGET



Inferences made from above chart is
- Most of the client were defaulted who applied previously for the card
- People who applied for car loans and the loan was approved, they were least defaulted
- Also for Refused loans the car loans were least defaulted.

CONCLUSIONS FOR CREDIT EDA ANALYSIS

- Approximately there is a chance of 8.62% of loans being defaulted in the current scenarios.
- When we see the number of loans applied then we notice there are female clients who applied the loans more than males.
- Married client are safer to give loans as their tendency of defaulting the loans is less while single and not married clients are more likely to default loans.
- Client whose previous loans were either rejected or cancelled seems more likely to default again.
- The higher the age, the higher the chance of less defaulting the loan, as we can see younger age people having age less than 30 tends to default loans more.
- People who apply for car loans have a tendency to default loans lesser in comparison to other like Cards.


THANK
YOU