CREDIT EDA ASSIGNMENT

Aditya Dhir

Problem Statement

In this credit analysis, the idea is to find different patterns which can indicate if client would have some difficulty in paying the installments that can be further used to take action such as reducing the amount of loan or lending at higher interest rates. In simple terms the company wants to understand the driving factors behind the loan default.

Business Objective.

Most of the times it happens that companies find it very difficult to identify the genuine applicants to whom loan should be given. It depends on various factors. In this Presentation, we will discuss about some important factors using EDA for the applications.

When a bank receives the application for loan, the company needs to decide whether it should be given a loan or not based on the profile of theirs. There can be two scenarios.

Scenario 1 : the applicant has great profile but bank is not able to provide them loan, in that case the probability is that the bank will face some loss.

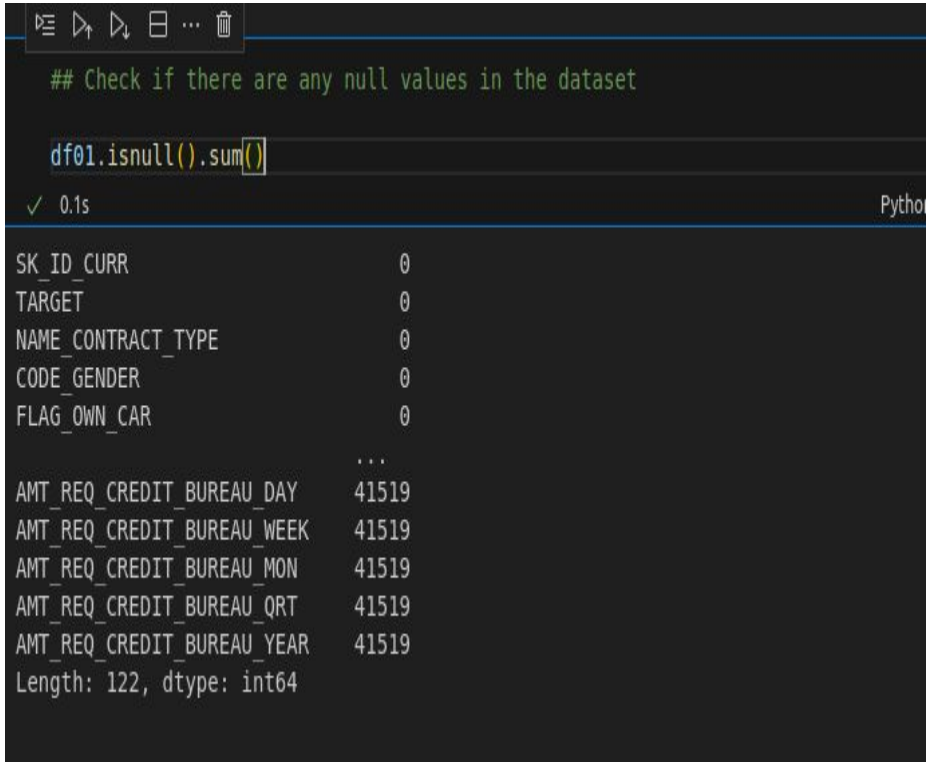
Scenario 2: the applicant has not a great profile but banks provides them loan, in that case the bank can face huge loss.

Approach followed

For the analysis, the first step that I followed was to do data cleaning. Most of the times the data is in the raw format, so it is quite important to know if there are any null values, if yes, then we can either fill them with mean or mode based on the column data. If that data is missing more than 50 percent of rows then we can just remove those columns.

After the data cleaning was, done, I started performing data analysis. There were different scenarios that i tried the analyzed. Few of them are listed in this PPT.

In this image we can see that there are lot of columns whose value is missing, for example “AMT_REQ_CREDIT_ BUREAU_WEEK”. We can fill this with the mean or mode.

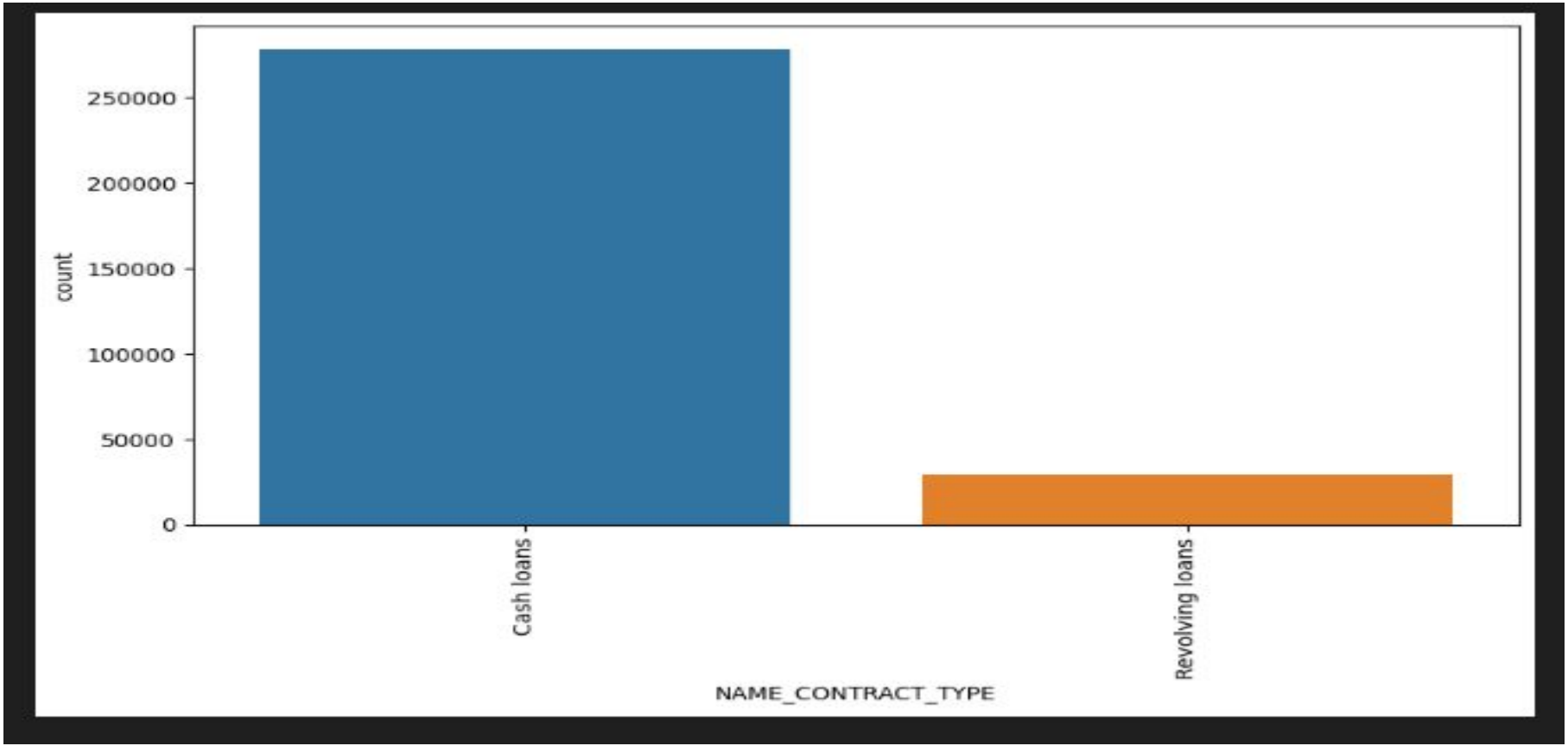


Data Analysis

As per the dataset, there are two types of loans.

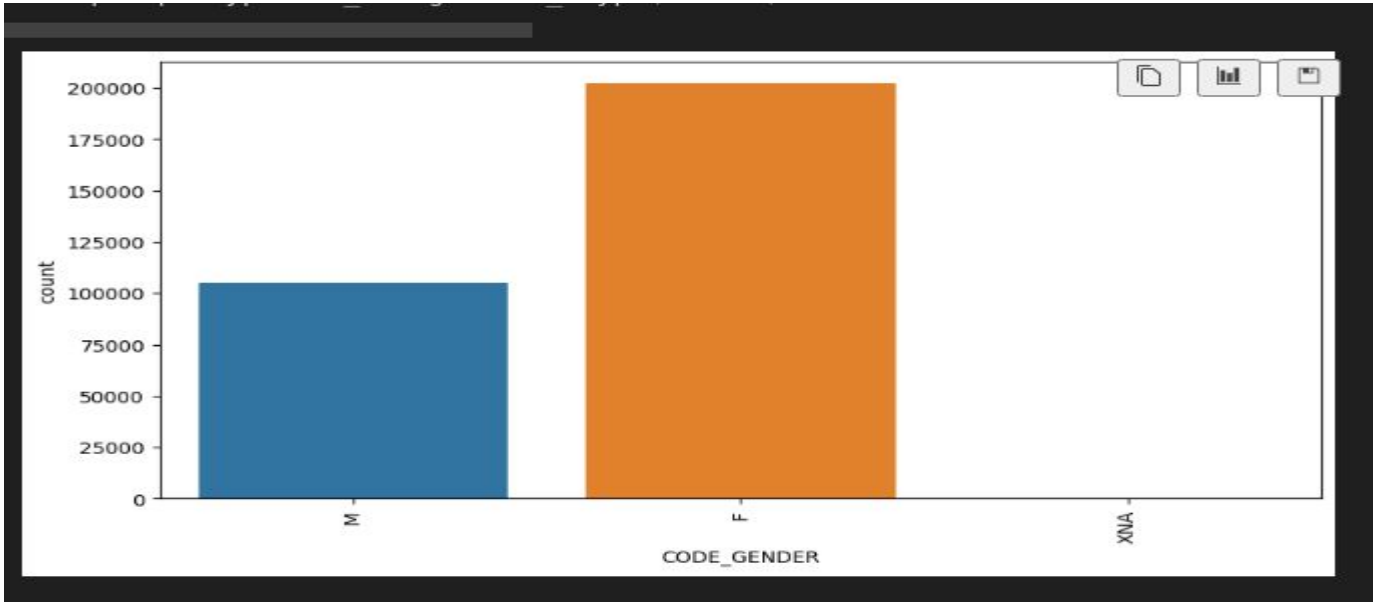
- Cash Loans
- And Revolving loans.

Majority of the applications are for Cashloans where as minority loans are Revolving loans.

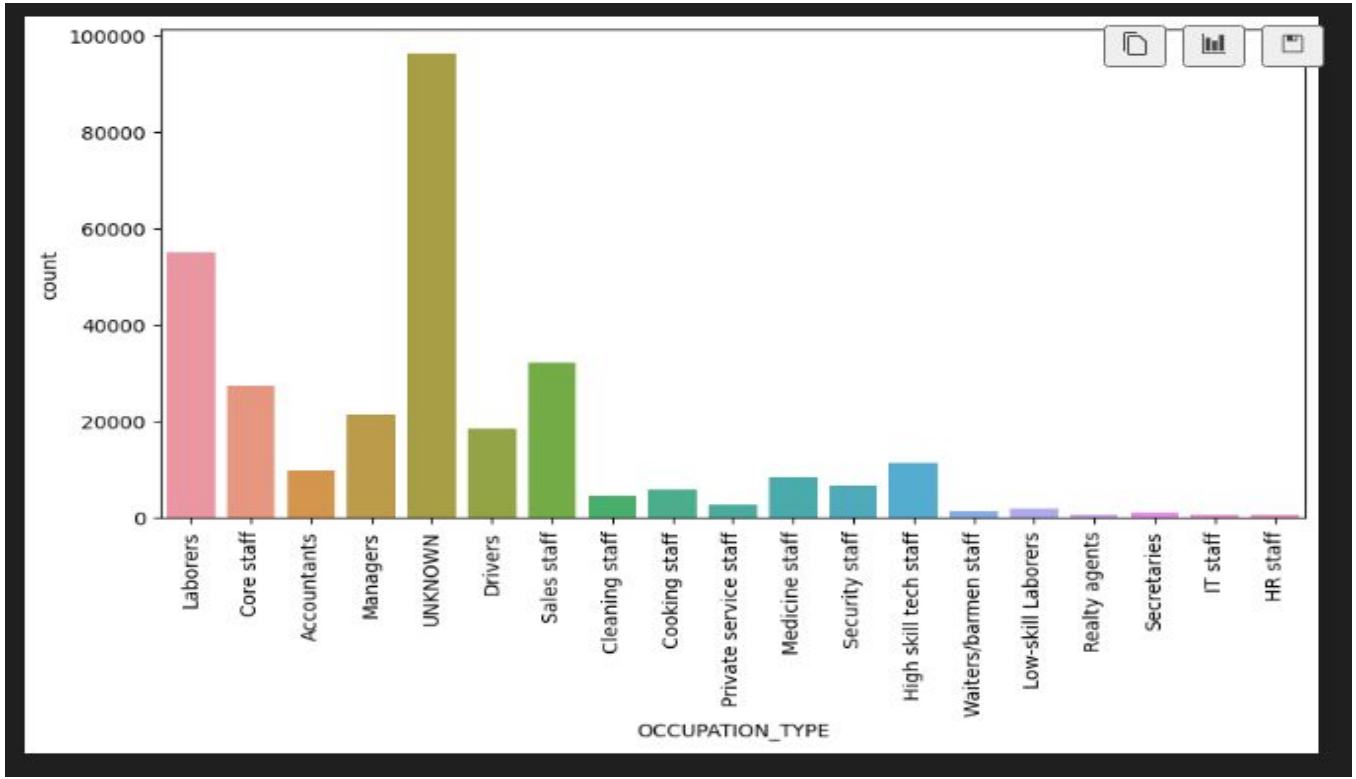

250000
200000
150000
100000
50000
0
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

Gender analysis

A lot of applicants who have applied for the loan are Female.



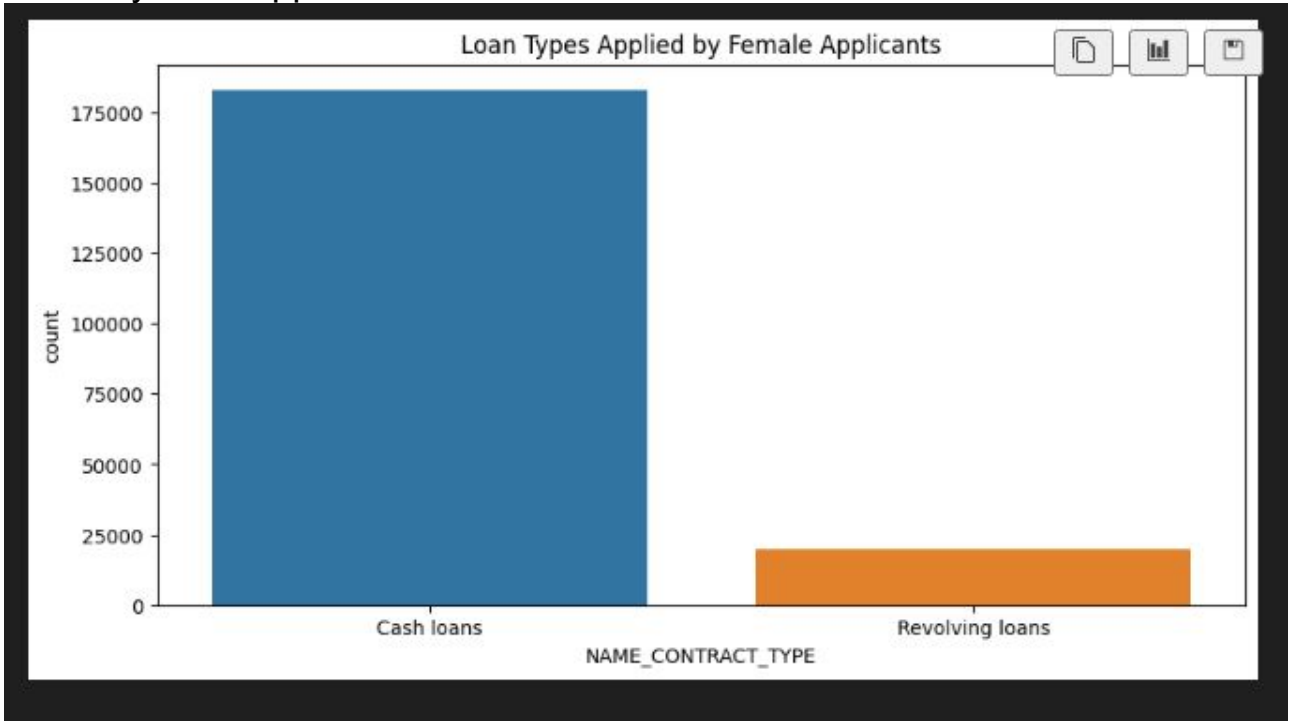
Profession analysis.



If we look in image slide, there are lot of applicants who are laborers whereas for many applications, the profession is Unknown. Giving loan to people whose profession is not known can be quite risky.

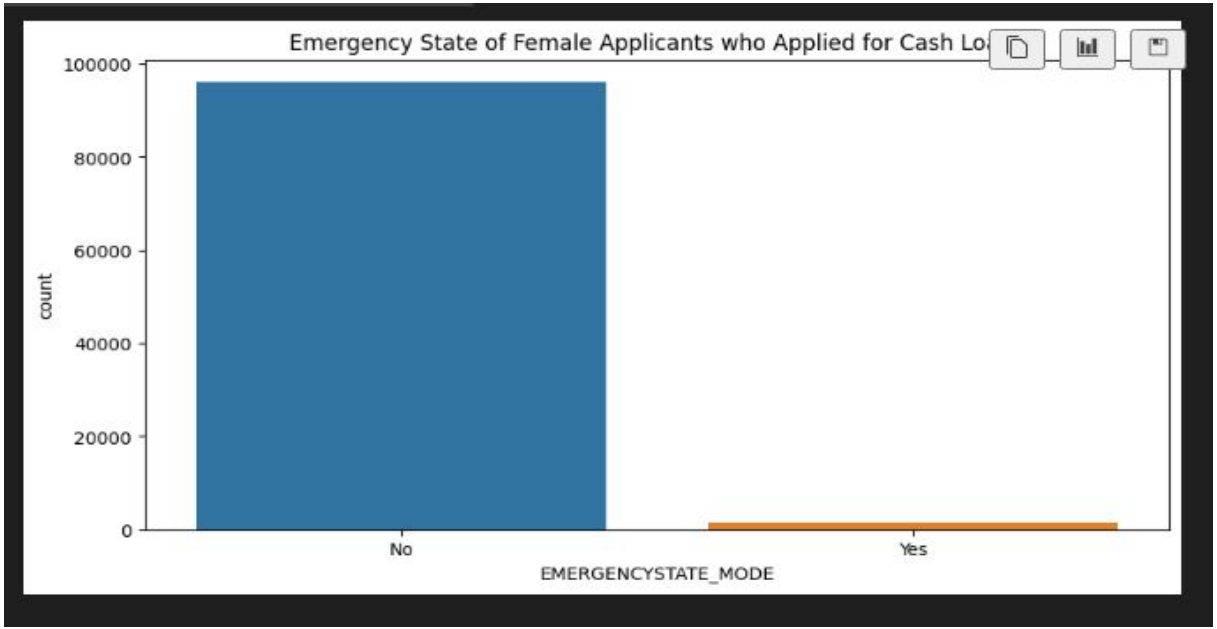
Types of loan for maximum applicants have applied for.

Like in previous few slides we have seen that the majority are female applicants, we will now look at what kind of loan does they have applied for.



Applications, that have applied loan for any emergency.

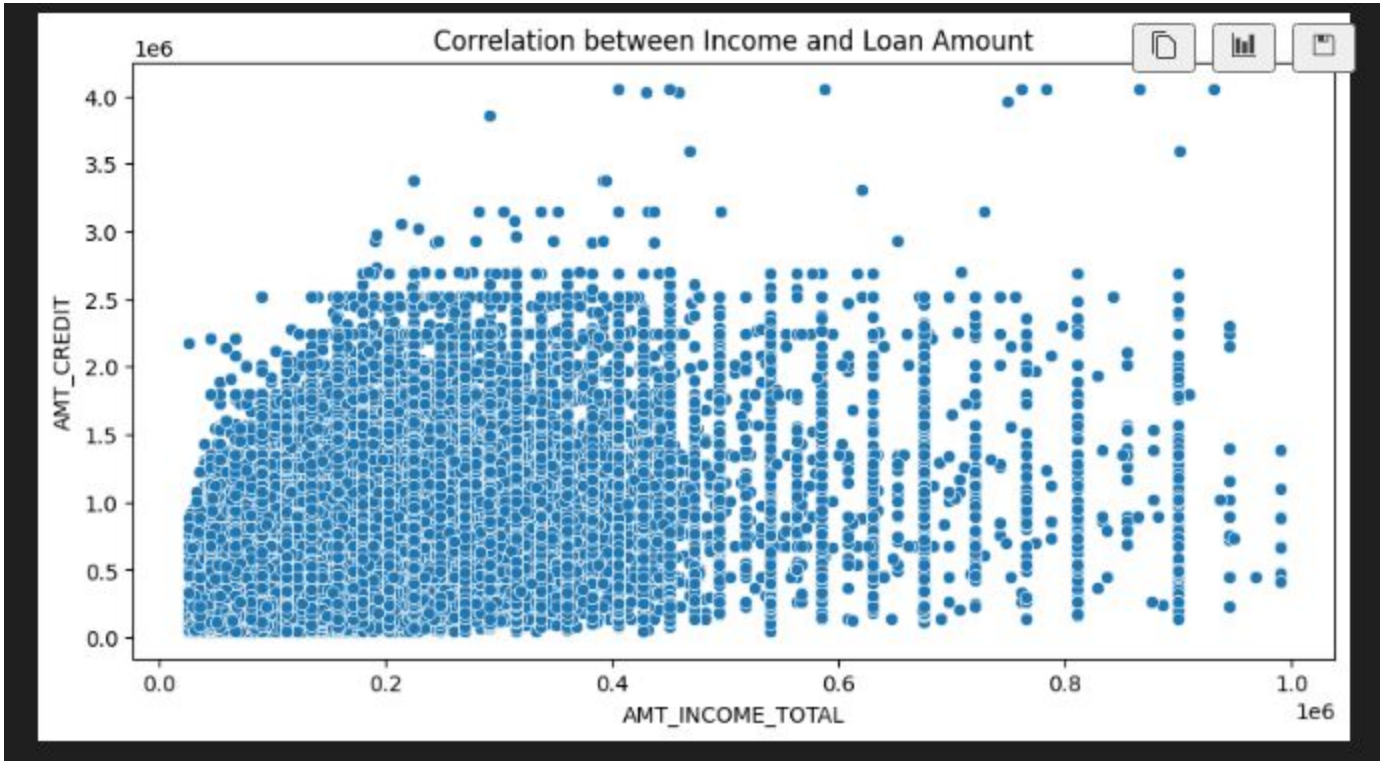
According to the data set most of the applicants have not applied for anykind of emergency loan. From this we can assume that they have good credit histories of having financial stabilities and do not find themselves in urgent need of loan.



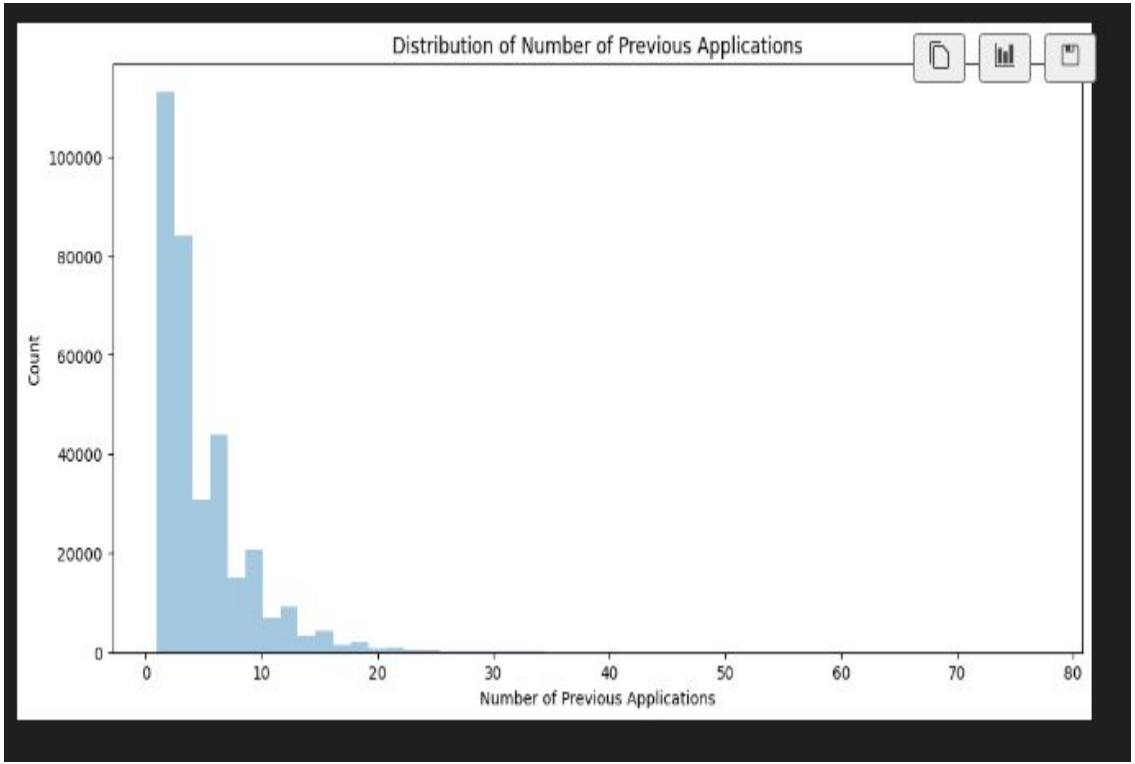
Income and Amount Credit Correlation.

	AMT_INCOME_TOTAL	AMT_CREDIT
AMT_INCOME_TOTAL	1.000000	0.397647
AMT_CREDIT	0.397647	1.000000

From the image we can see that, the correlation between AMT_INCOME_TOTAL and AMT_CREDIT is positive (0.39). From this we can conclude that applicants with higher income are tend to receive more credit amount, but there is no causation as it is not always likely that people with high income can get high credit.


Correlation between Income and Loan Amount
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
AMT_CREDIT
0.0
0.2
0.4
0.6
0.8
1.0
1e6
AMT_INCOME_TOTAL

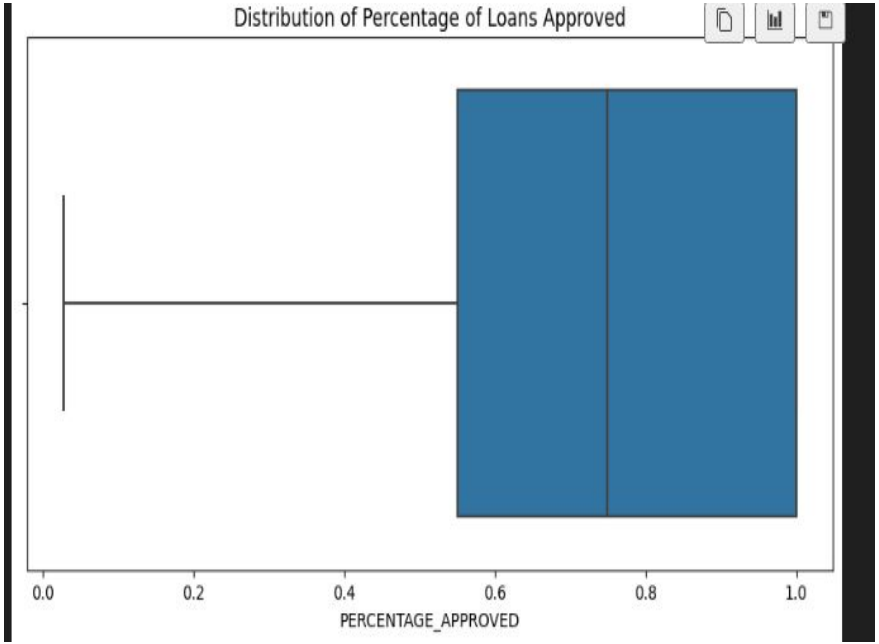
Applicants with previous applications.



There are lot of applications who have applied for the different loans. For most of them there are only 1-2 applications but few candidates have large number of applications. It might not be feasible to offer loans who have applied quite a lot of times.

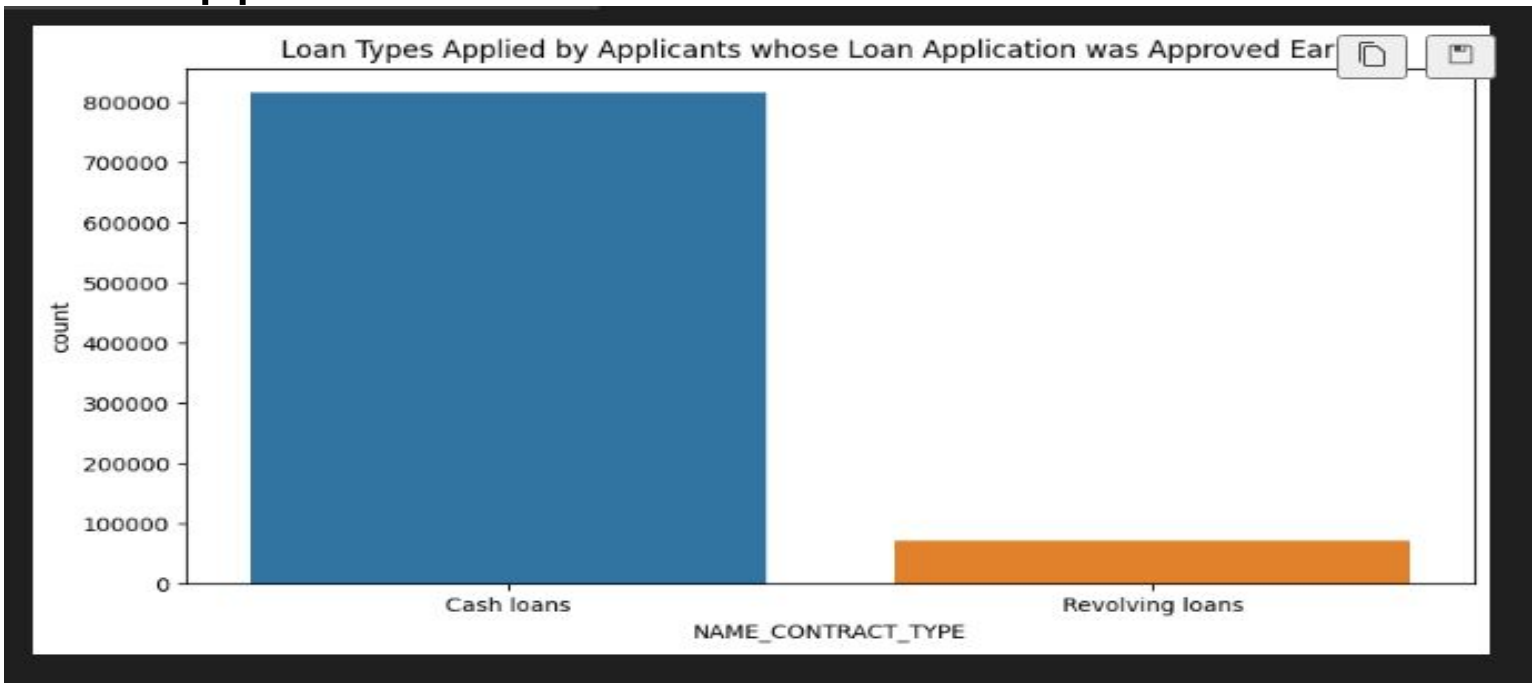
Percentage of Applications approved for each applicant.

Based on the business logic, there can be the certain number of applications which might be considered safe but we need to also check if for all the applications the loan was approved. If higher number of applications for same applicant were approved then we can consider giving them a new loan.



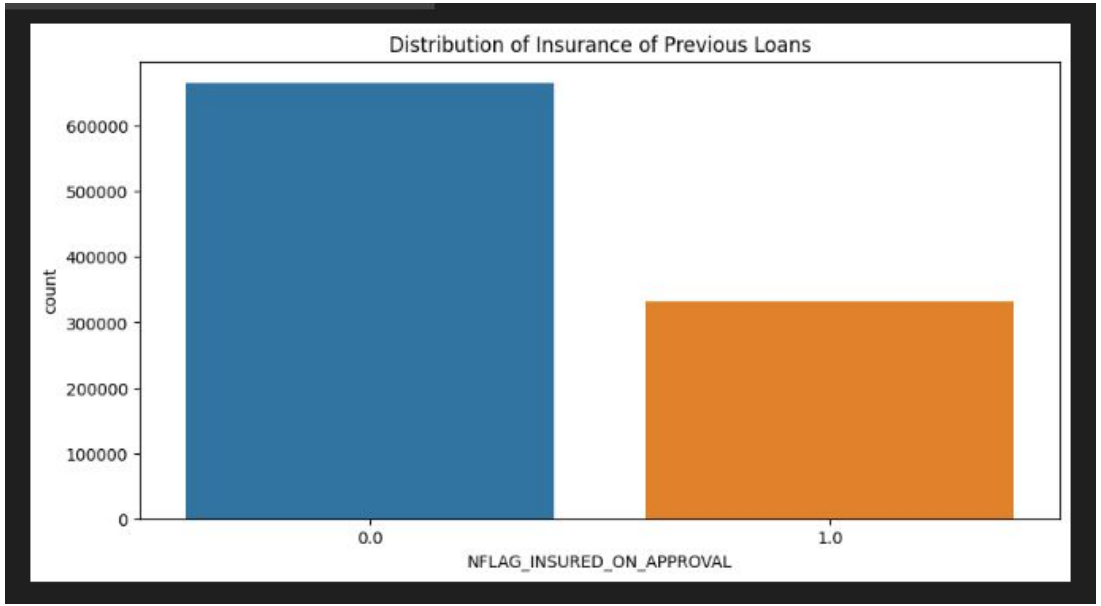
As we can see for majority of the applicants a large number of loans were approved. We can further look for each candidate. And consider various other factors.

Loan types applied by applicant whose loan application was approved earlier.



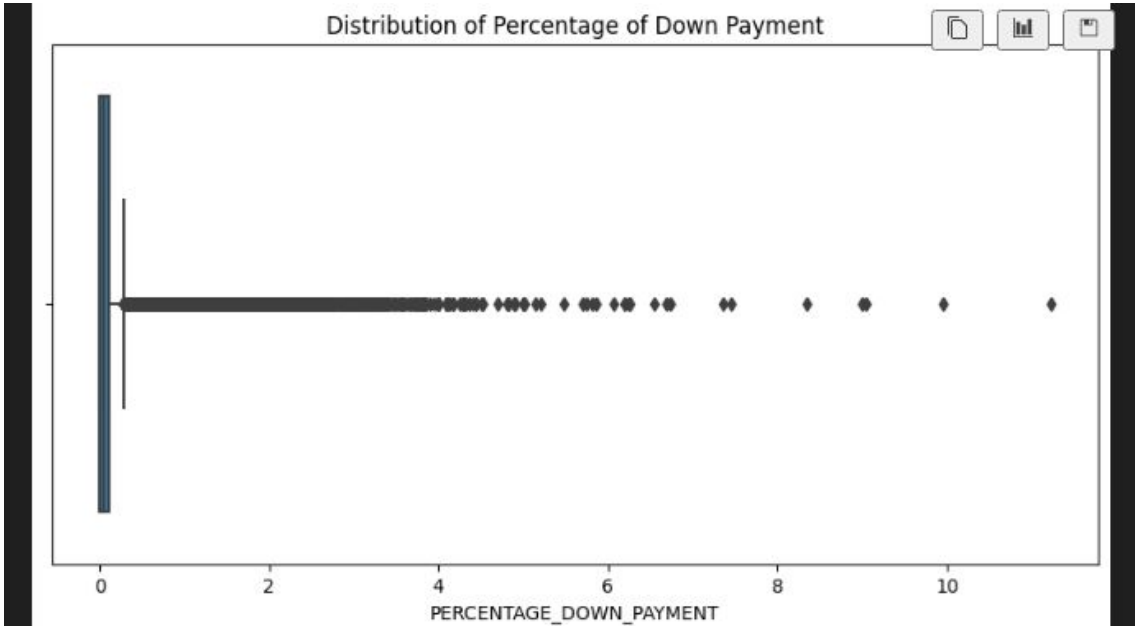
People with previous applications did not take insurance.

A lot of applicants did not take any insurance for the loan they they took. Not having a insurance can be quite riskier for the lender and borrowers. For lender it is quite riskier because they have less safeguards.



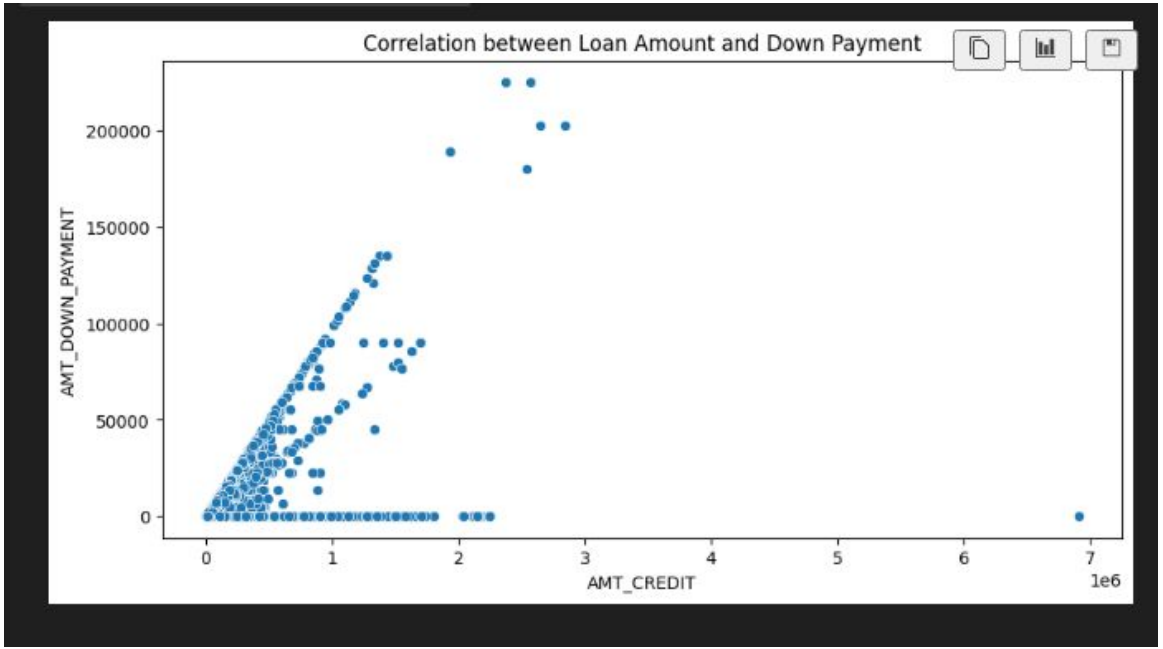
Applicants who have paid down payment.

Down payment plays a major role in deciding whether the applicant should be provided the loan or not.



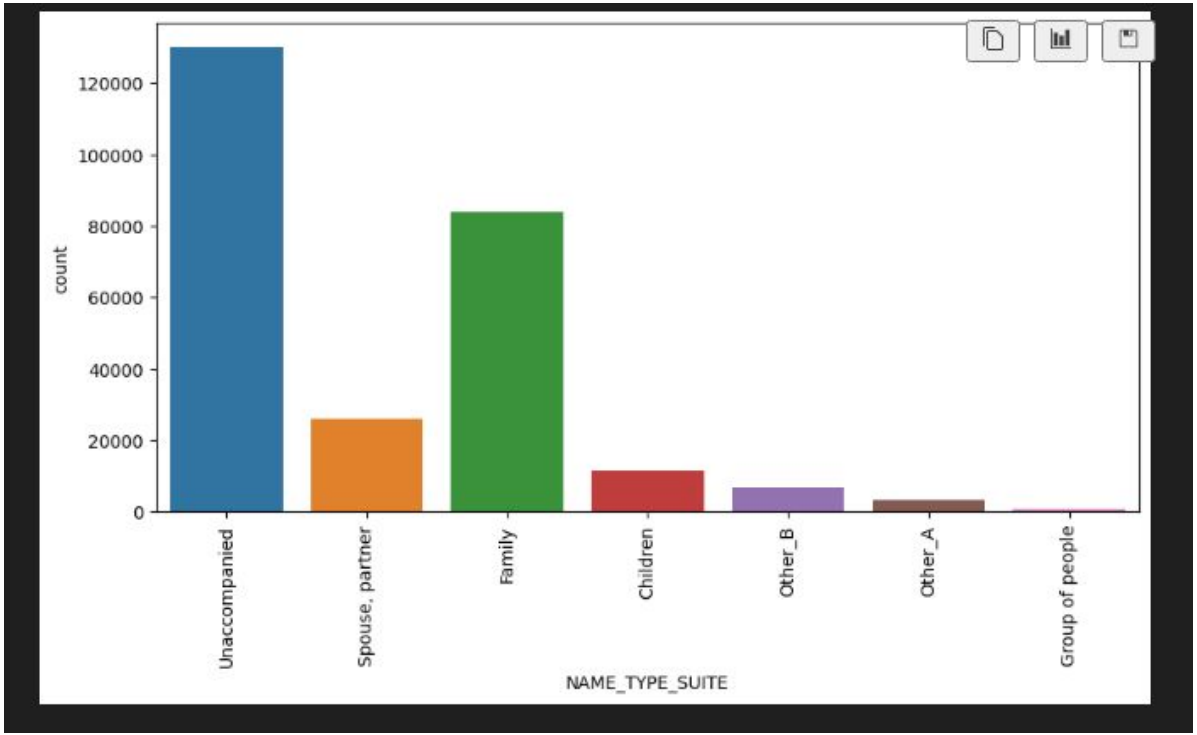
Correlation between Amount Credit and Amount Downpayment.

We can see that the downpayment made and the amount asked for loan has quite low correlation. It is not necessary that people asking for higher loans have made downpayment.



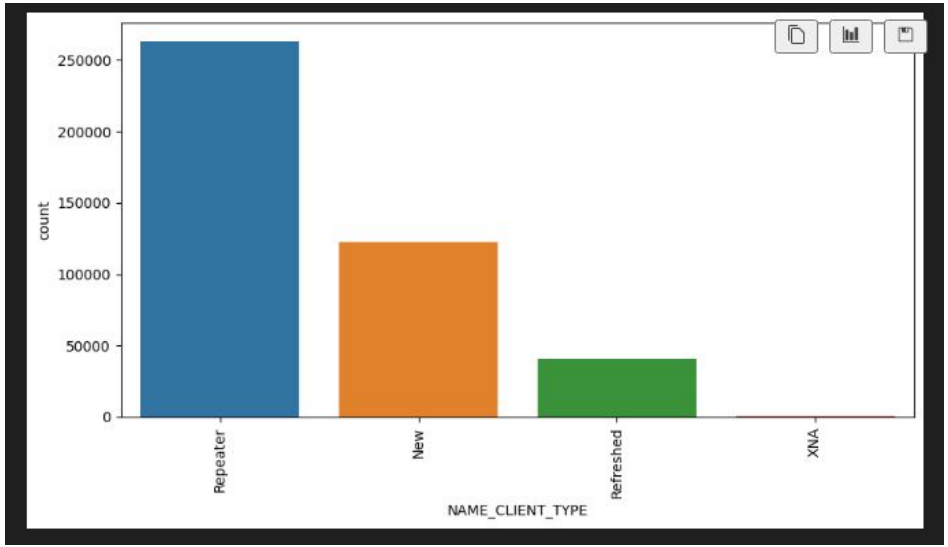
Nature of applicants, whether they are applying alone or with coapplicant.

Majority of the loan applicants are single. It is quite better to give them loan, as they dont have any dependents.

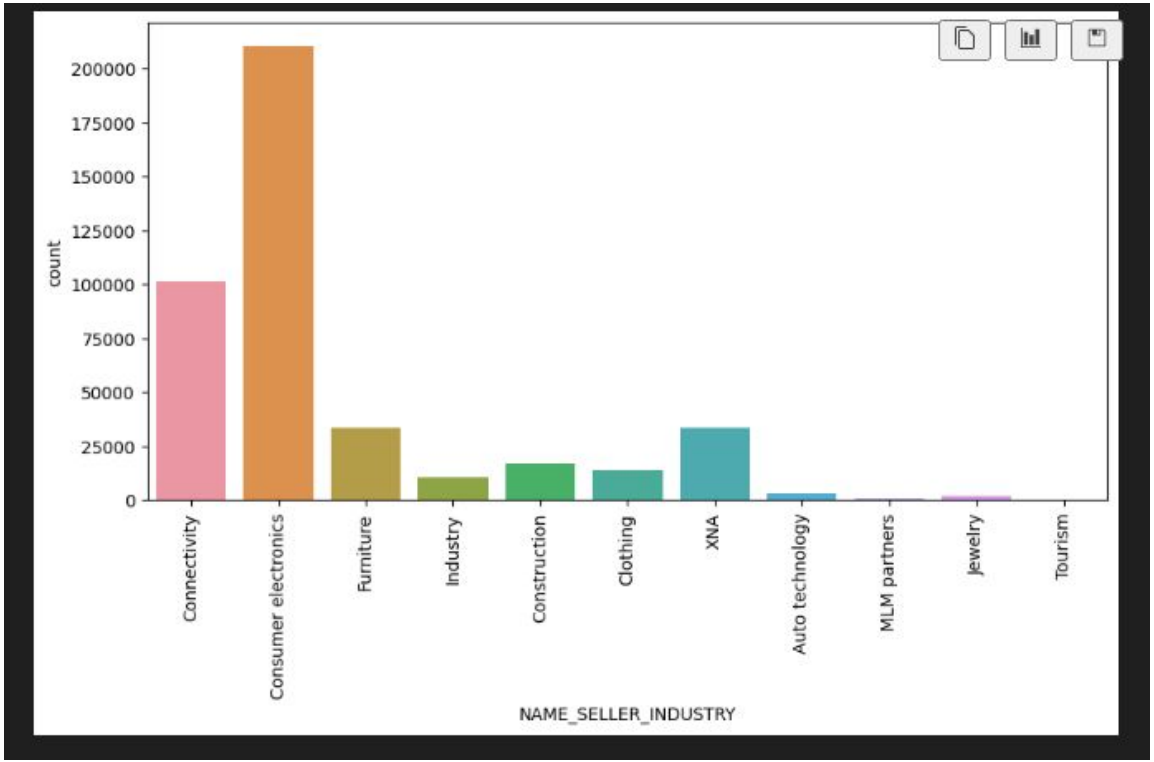


Are there any new applicants or repeaters.

Majority of the people who have applied for the loan are the repeater. It is quite obvious that, priority should be given to the existing customers as there is credit history of them and we know how well they have paid the installments.



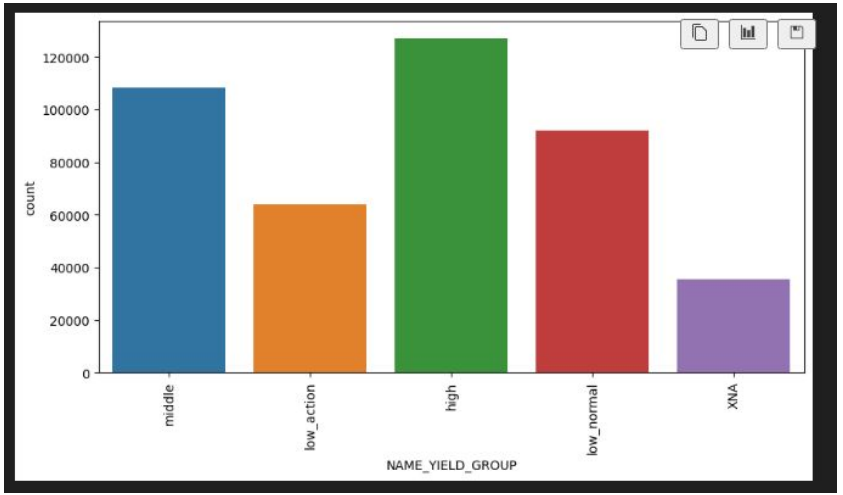
Kinds of products for which people have applied loan for.



Interest Rate yielded from the people

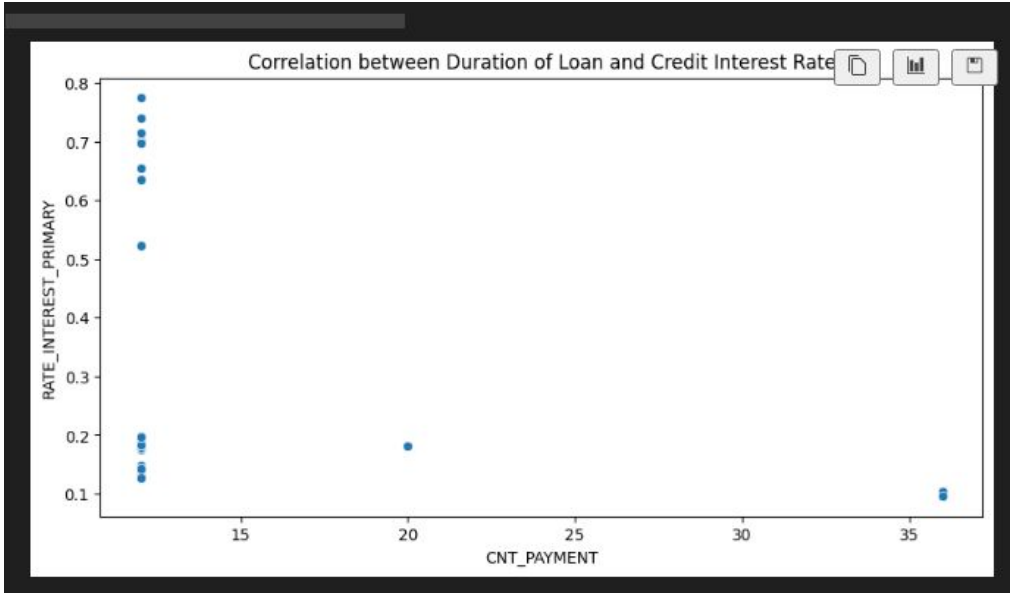
From majority of people a high interest rate

Was yielded.

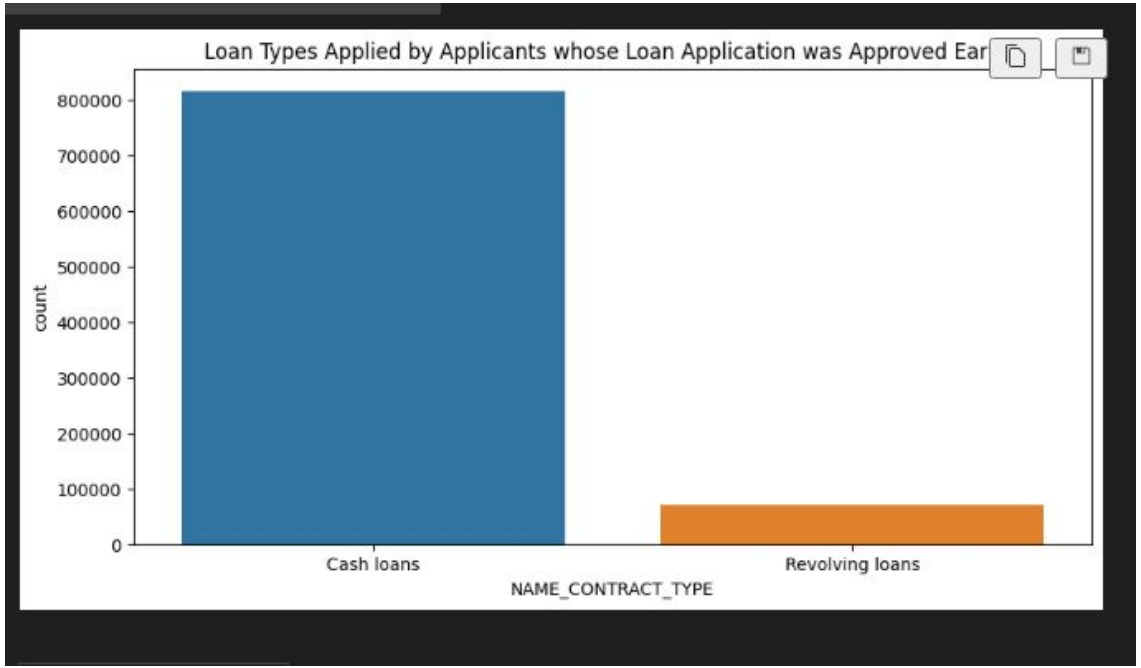


Relationship between CNT_PAYMENT and RATE_INTEREST_PRIMARY

People who have paid the more installements on time had to pay the less interest rate whereas people who less pay the installments on time had to pay more interest rate.



Number of applicants who got loan approved earlier and have applied again.



In previous few slides we saw that

- Majority of applicants are female, who are looking to buy household electronics items. For few of them the profession is UNKNOWN due to which there is some risk associated with providing them loan.
- Most of the applicants tends to buy items using POS with high interest rate.
- Majority of loan applicants have previously applied and looking for another loan, giving them loan could be the first priority.
- A lot of people who have taken loan for maximum duration previously has got lower interest rate whereas people who have taken loan for lower duration had to pay more interest rate. Which is expected behaviour.
- A lot of applicants are the previous customers who have already taken loan and have more than one application approved.