



CREDIT EDA

CREDIT RISK ANALYSIS

Rajeev Ranjan Sinha and Karthick VG

PROBLEM STATEMENT

◆ There are two types of risks associated with any loan requests:

- **H0** If the application is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- **H1** 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.

Analysis of the dataset has been done in python on jupyter Notebook

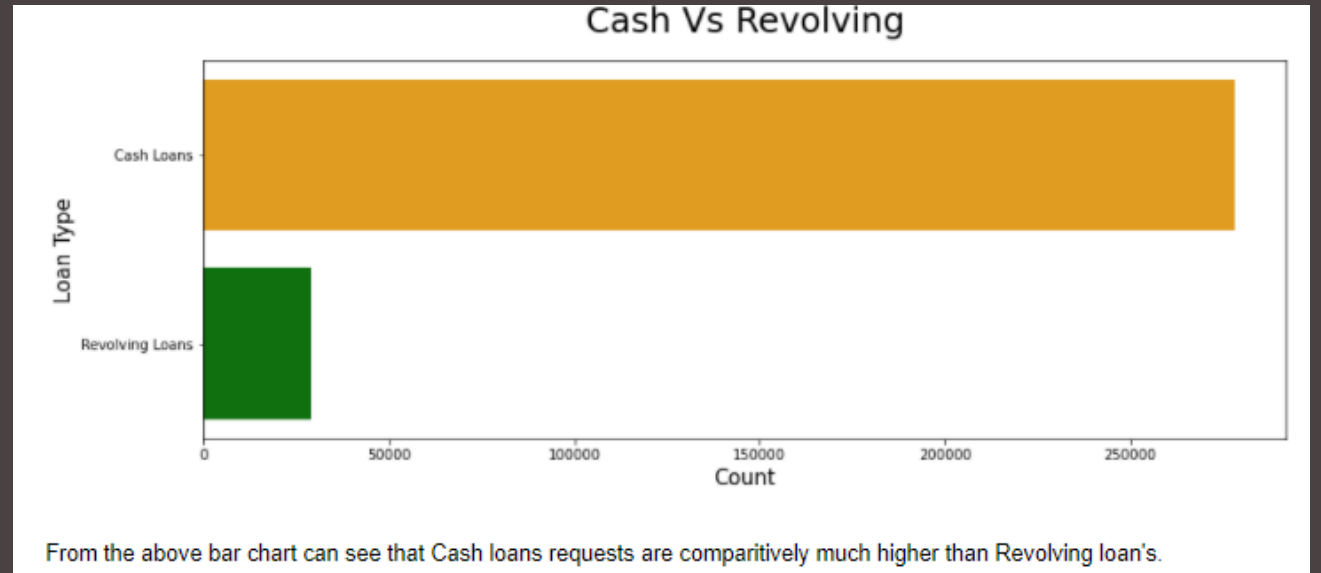
ANALYSIS DONE

◆Steps:

- Check and handle Missing Values.
- Check and treat outlier for data imbalance.
- Top 10 correlation for the client with payment difficulties other variables within Application DF and Previous App df.
- Check which correlation is most relevant.

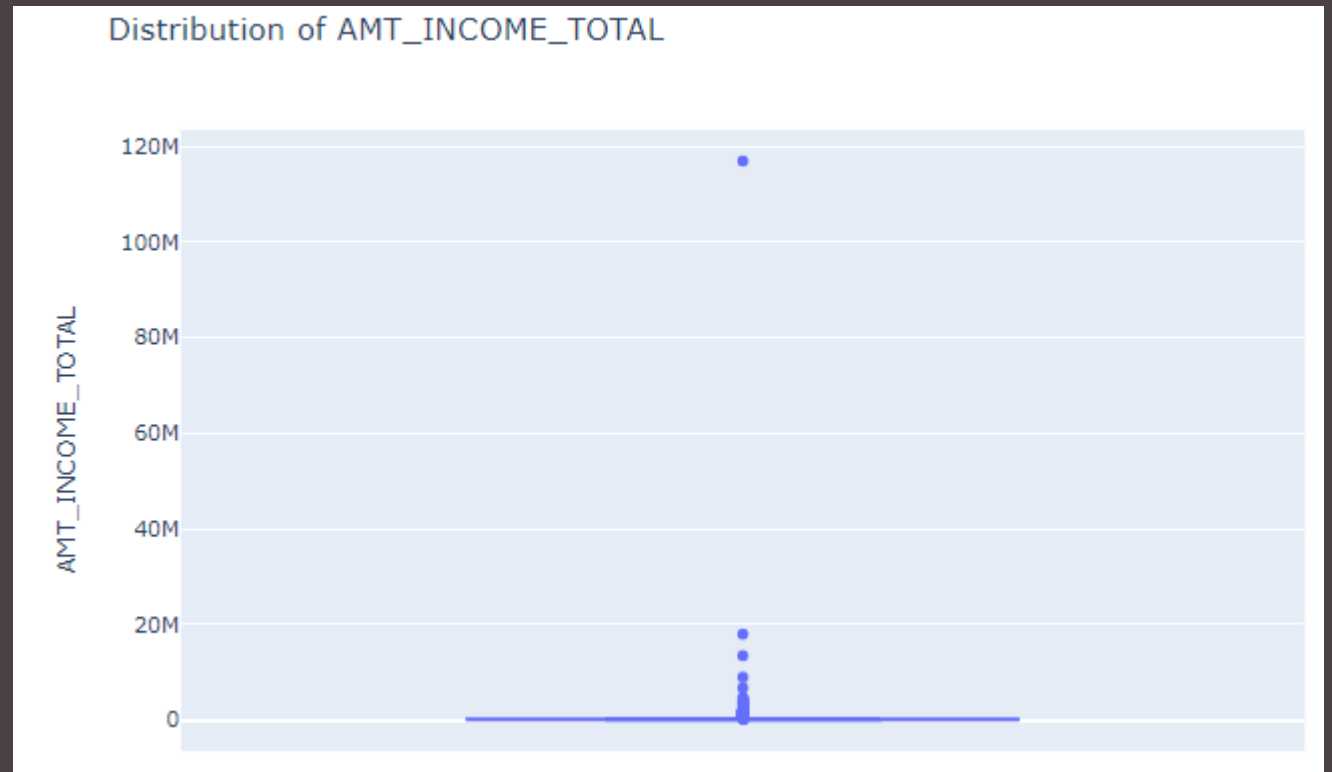
CASH VS REVOLVING REQUESTS

We can see there is clear most request is for the Cash Loan (6-7 times more) as compared to Revolving loan.



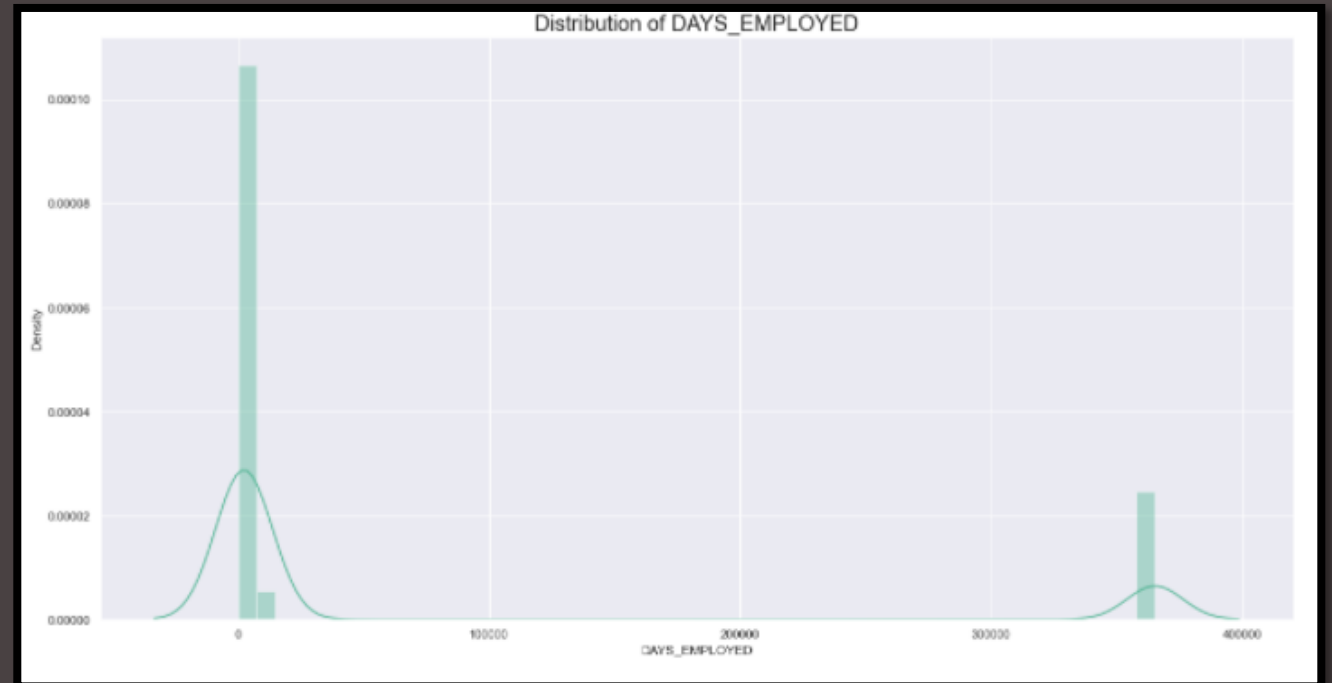
DISTRIBUTION OF AMOUNT OF TOTAL INCOME

We observe a value of 117M which is surely an outlier.



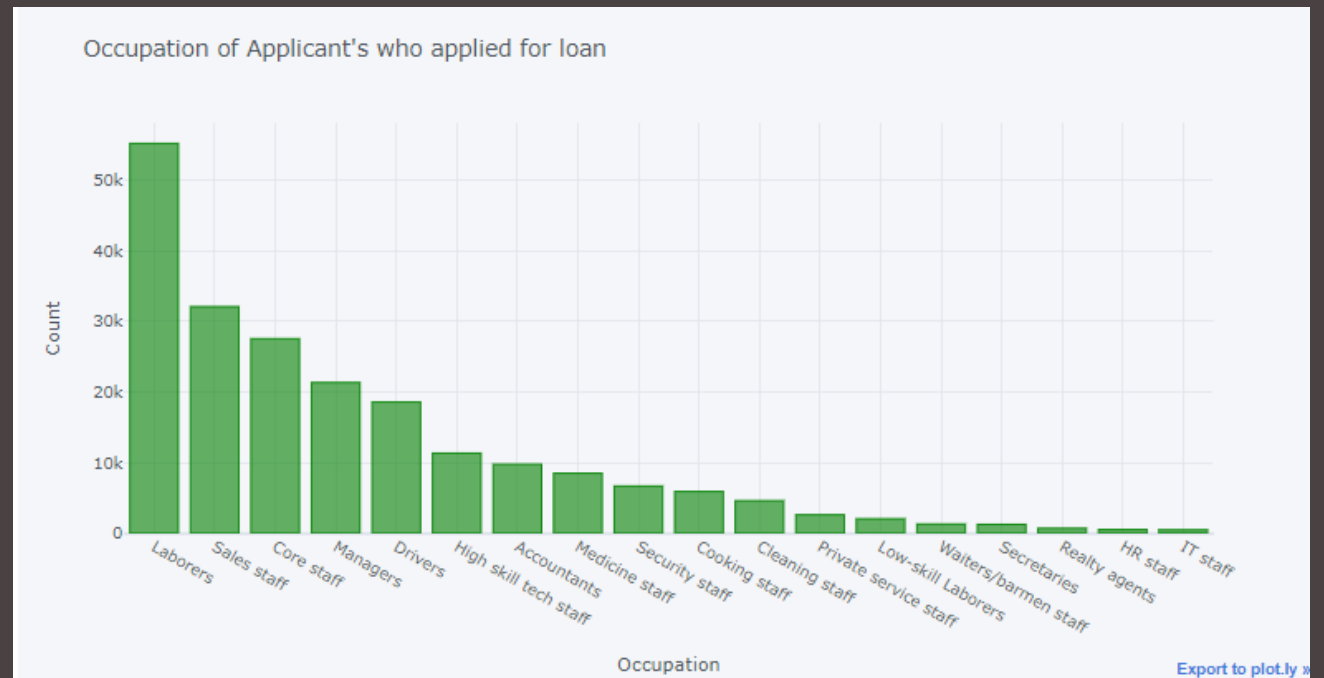
DISTRIBUTION OF DAYS EMPLOYED

Considering that a person started working at age of 21, the person will be 21+54 will turn out to be 75 years old.



DISTRIBUTION OF 'OCCUPATION_TYPE'

We can observe that Laborers, Sales Staff and Core Staff constitute the majority whereas IT Staff is on the lower side.



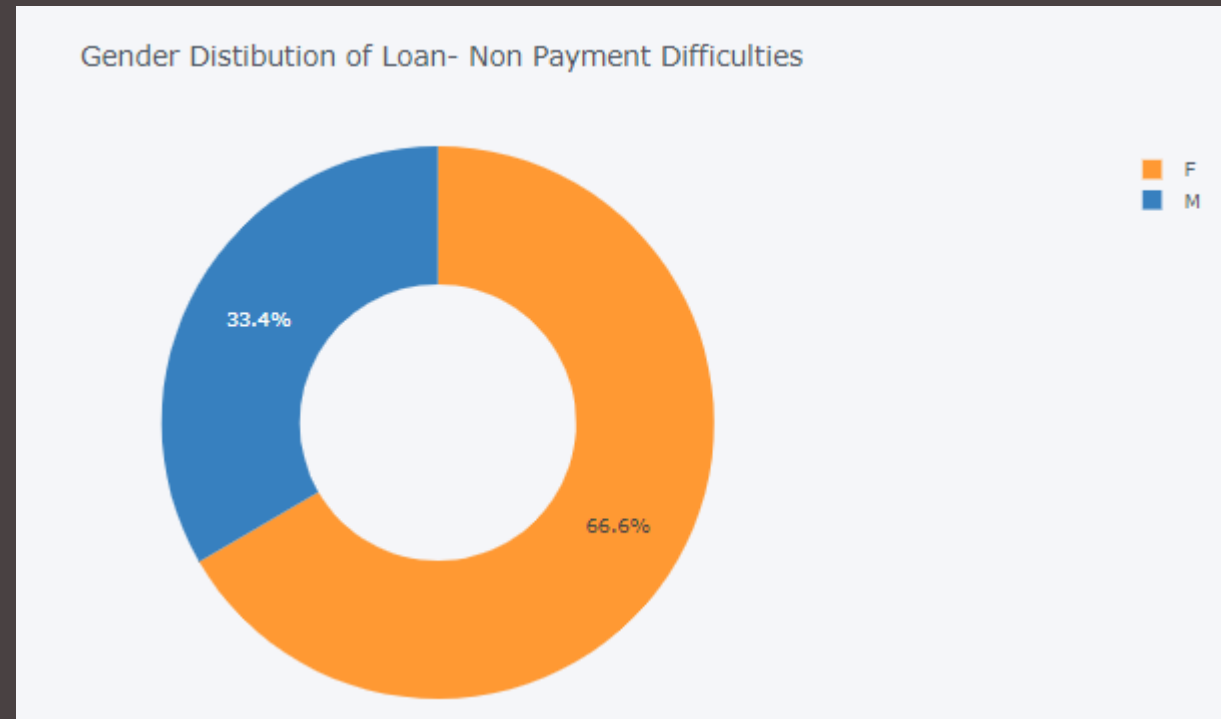
DISTRIBUTION OF 'ORGANIZATION_TYPE'

We can observe that Business Entity Type 3 organizations have majority of the loan applications.



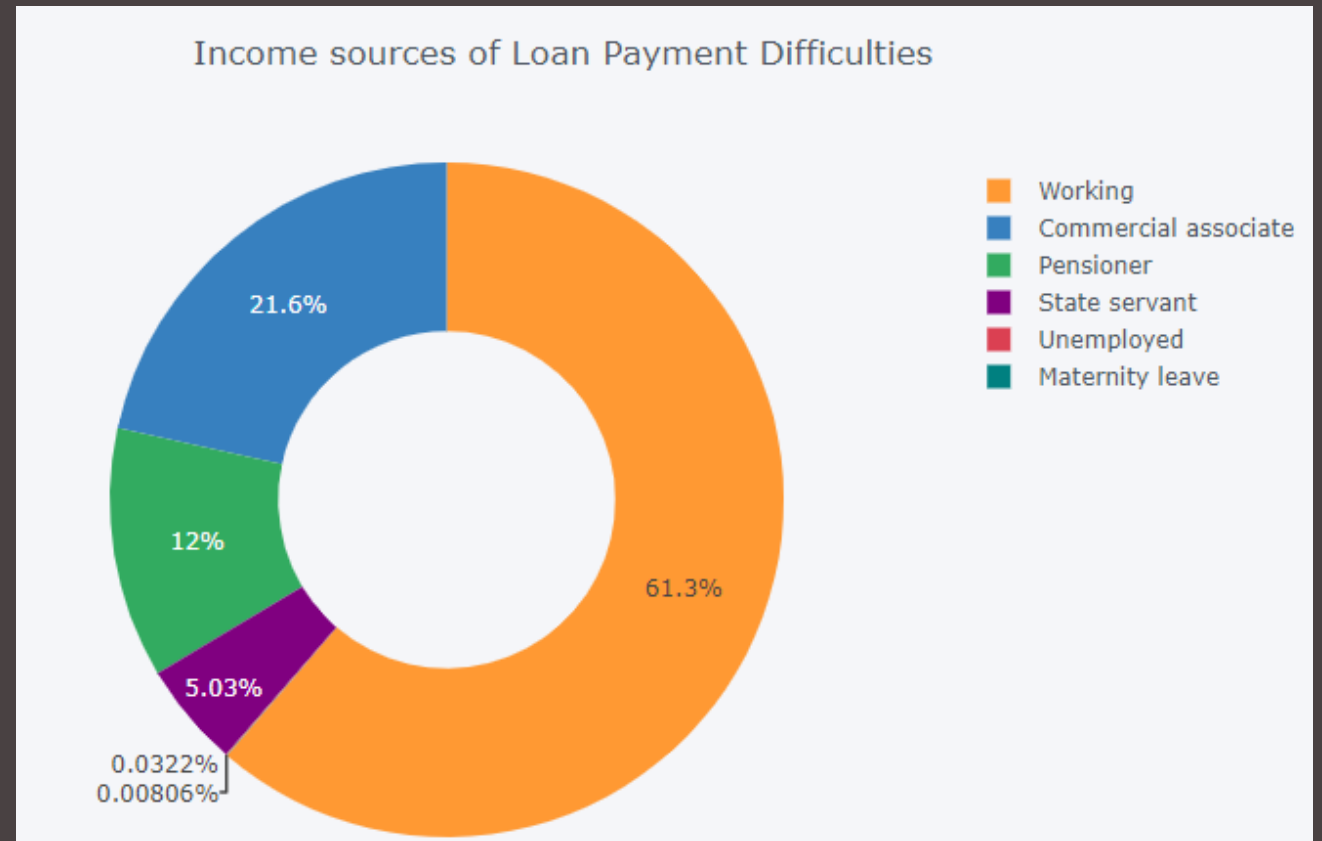
GENDER DISTRIBUTION OF LOAN

This pie chart shows that 2/3 or Female applicant has non-payment difficulties than Male applicant.



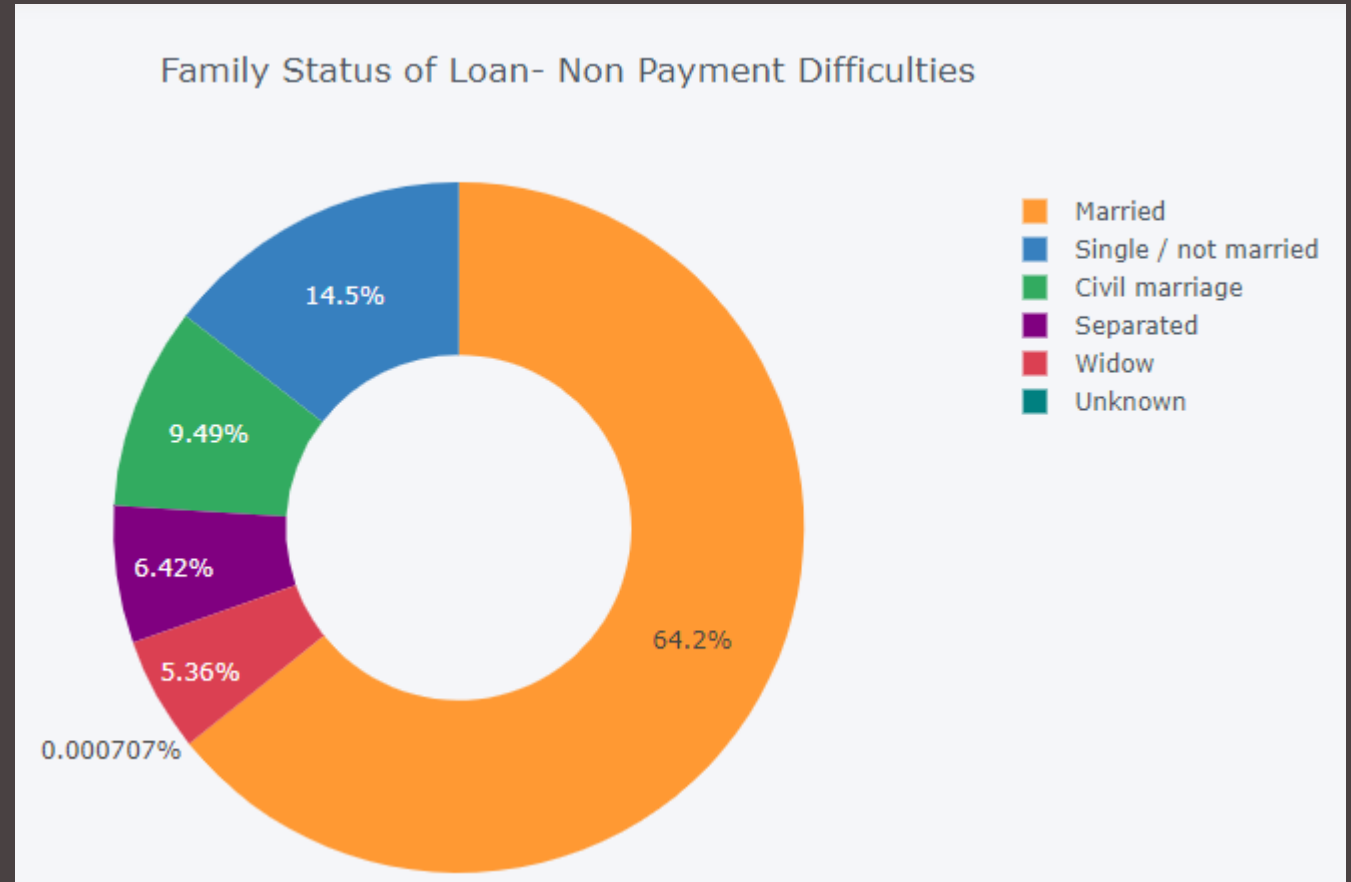
INCOME SOURCES OF LOAN- NON PAYMENT DIFFICULTIES

61% Working applicant
followed by 22%
Commercial associate
are Income sources of
loan payment
difficulties



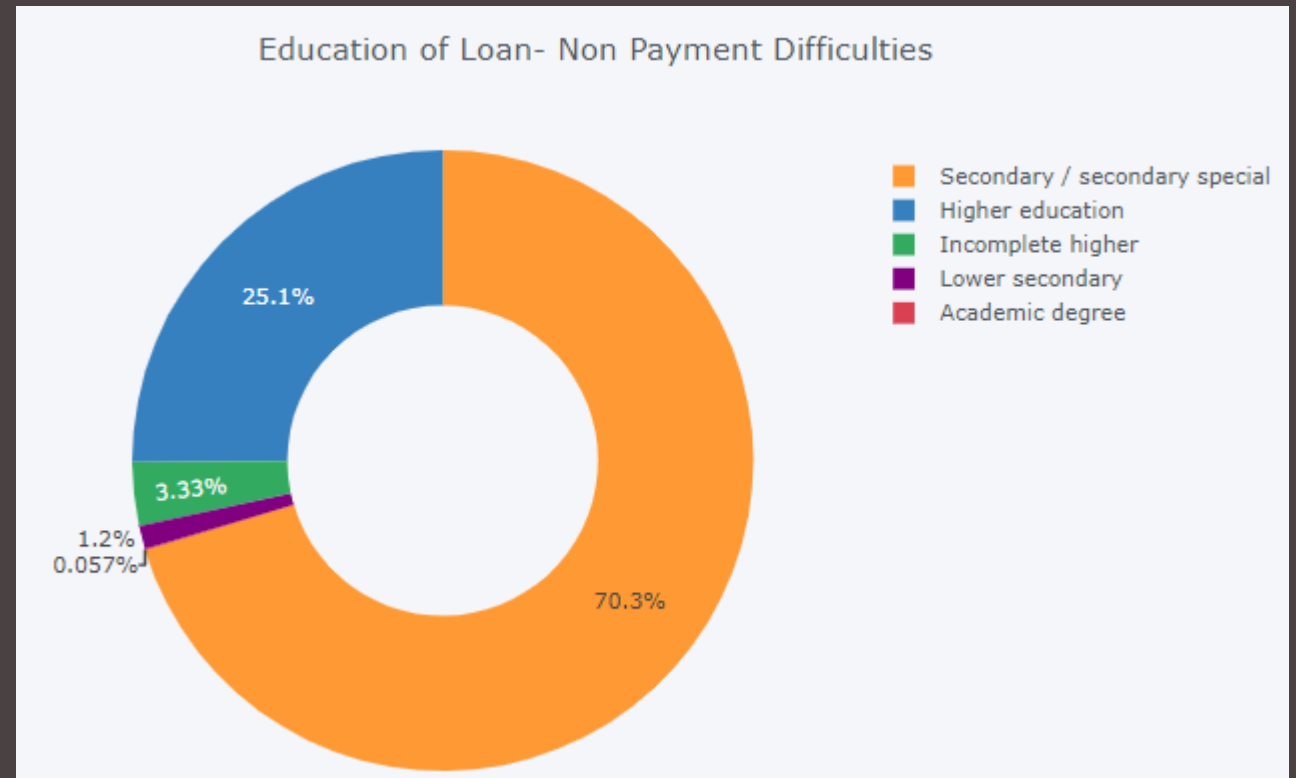
FAMILY STATUS OF LOAN – NON PAYMENT DIFFICULTIES

Married applicant are among the highest 65% with loan –non payment difficulties.



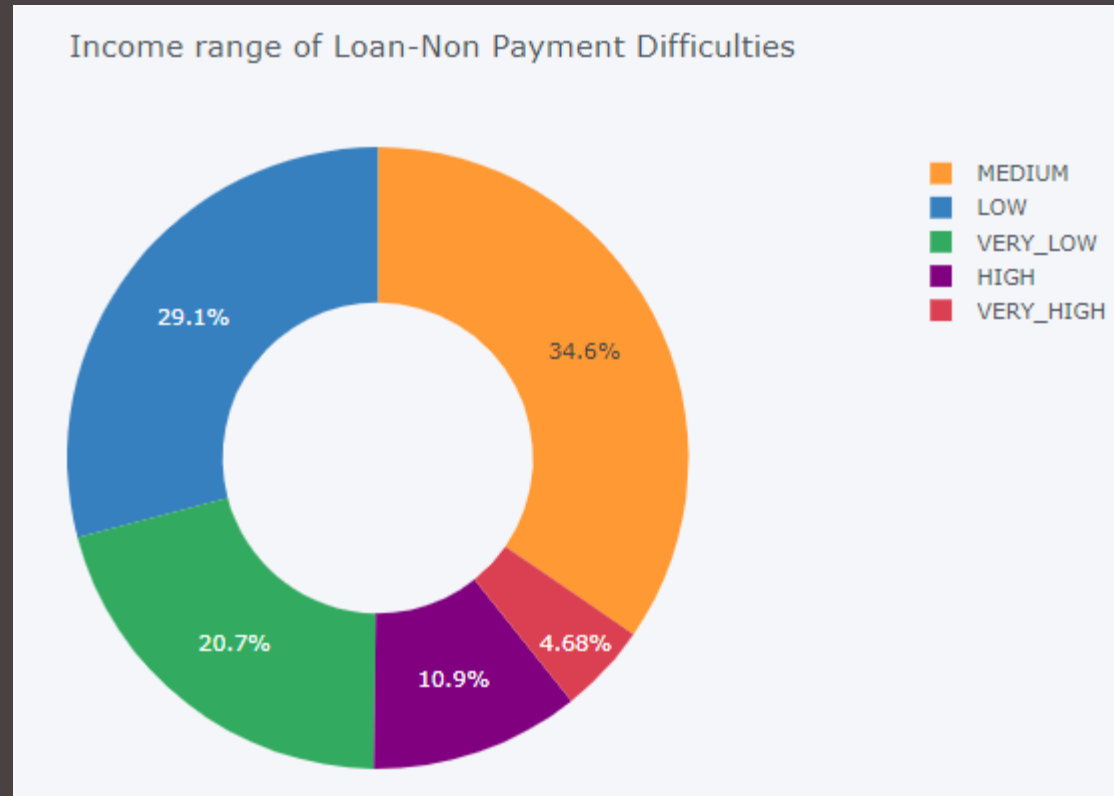
EDUCATION OF LOAN- NON PAYMENT DIFFICULTIES

Around 70% of secondary passed population are among the highest for loan – non payment difficulties.



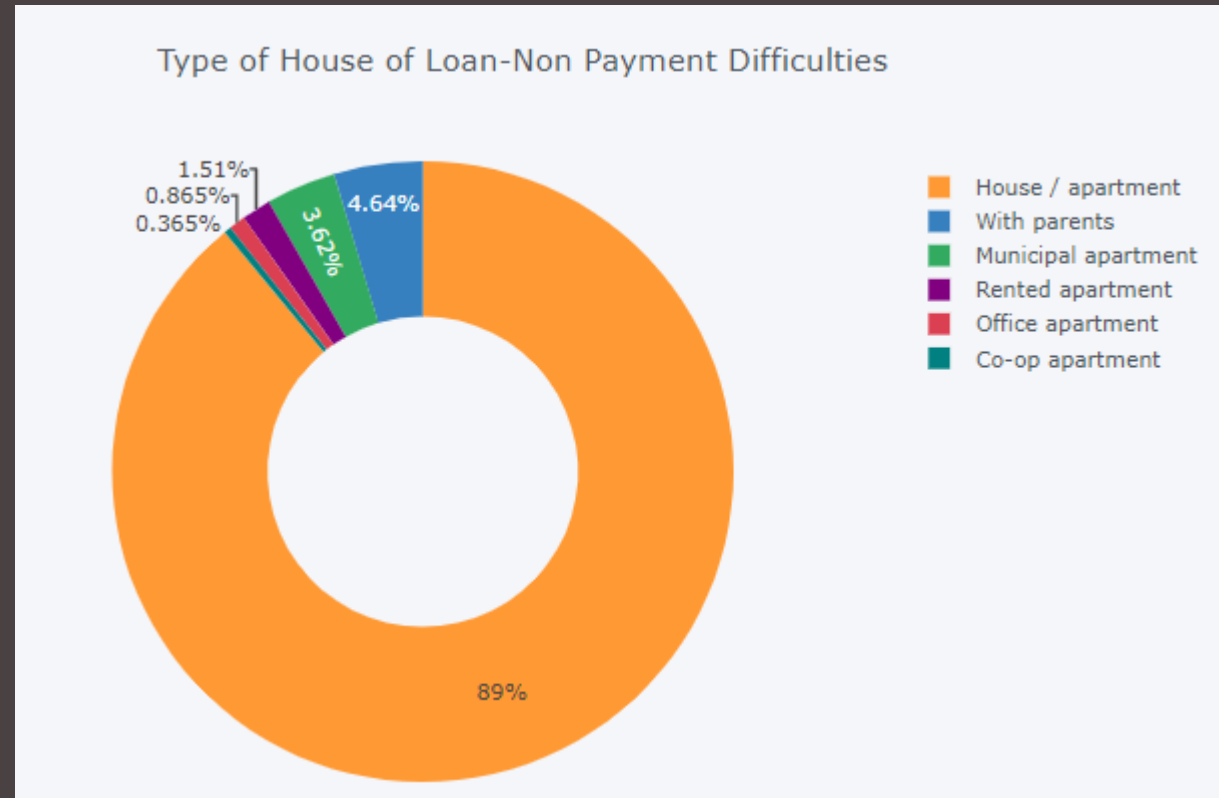
INCOME RANGE OF LOAN-NON PAYMENT DIFFICULTIES

Around 35% of population in medium income range are among the highest for loan –non payment difficulties.



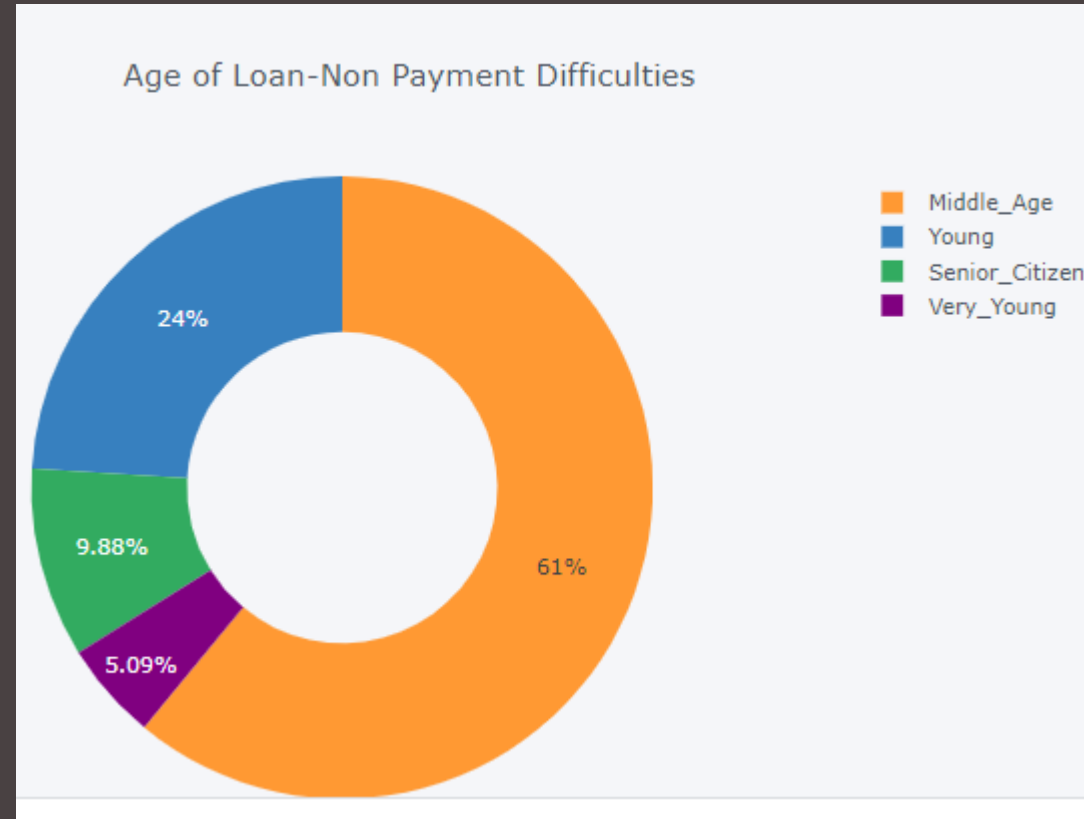
TYPE OF HOUSE OF LOAN- NON PAYMENT DIFFICULTIES

Around 90% of population possess their own house/apartment are among the highest for loan –non payment difficulties.



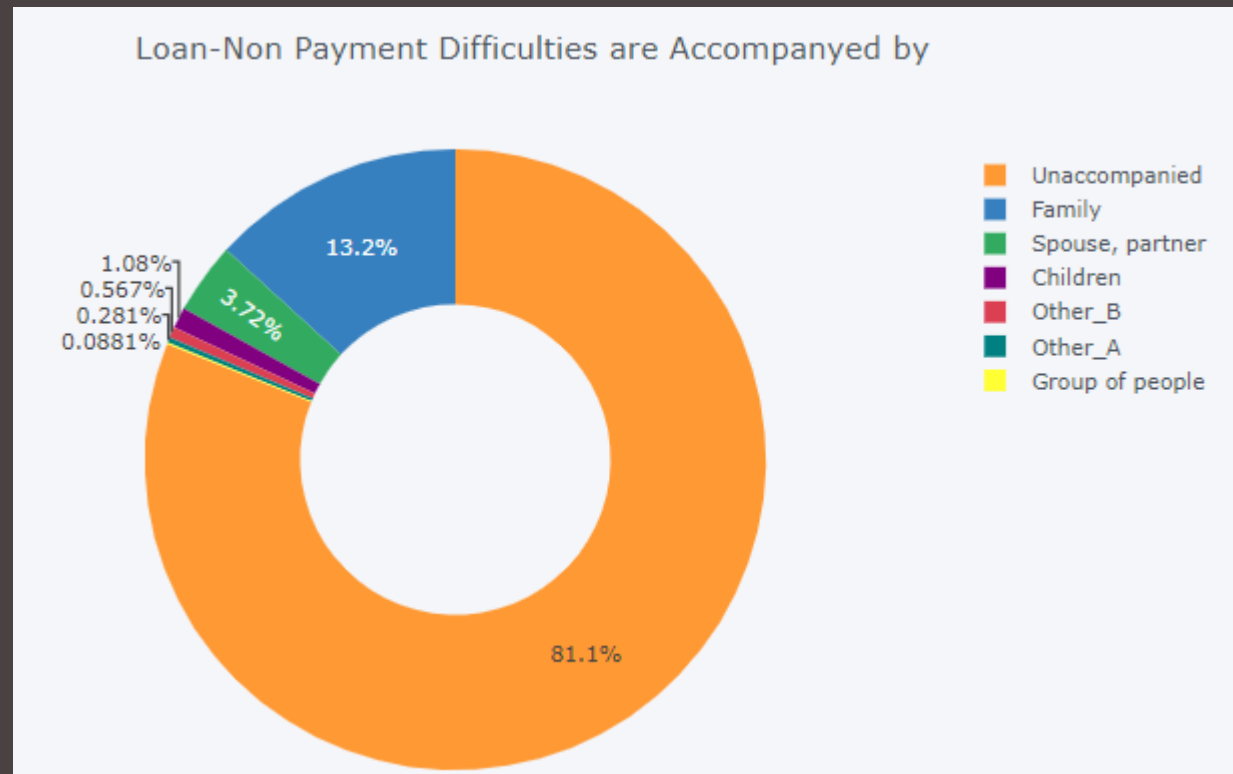
AGE OF LOAN-NON PAYMENT DIFFICULTIES

Around 61% of population in mid age group are among the highest for loan –non payment difficulties.



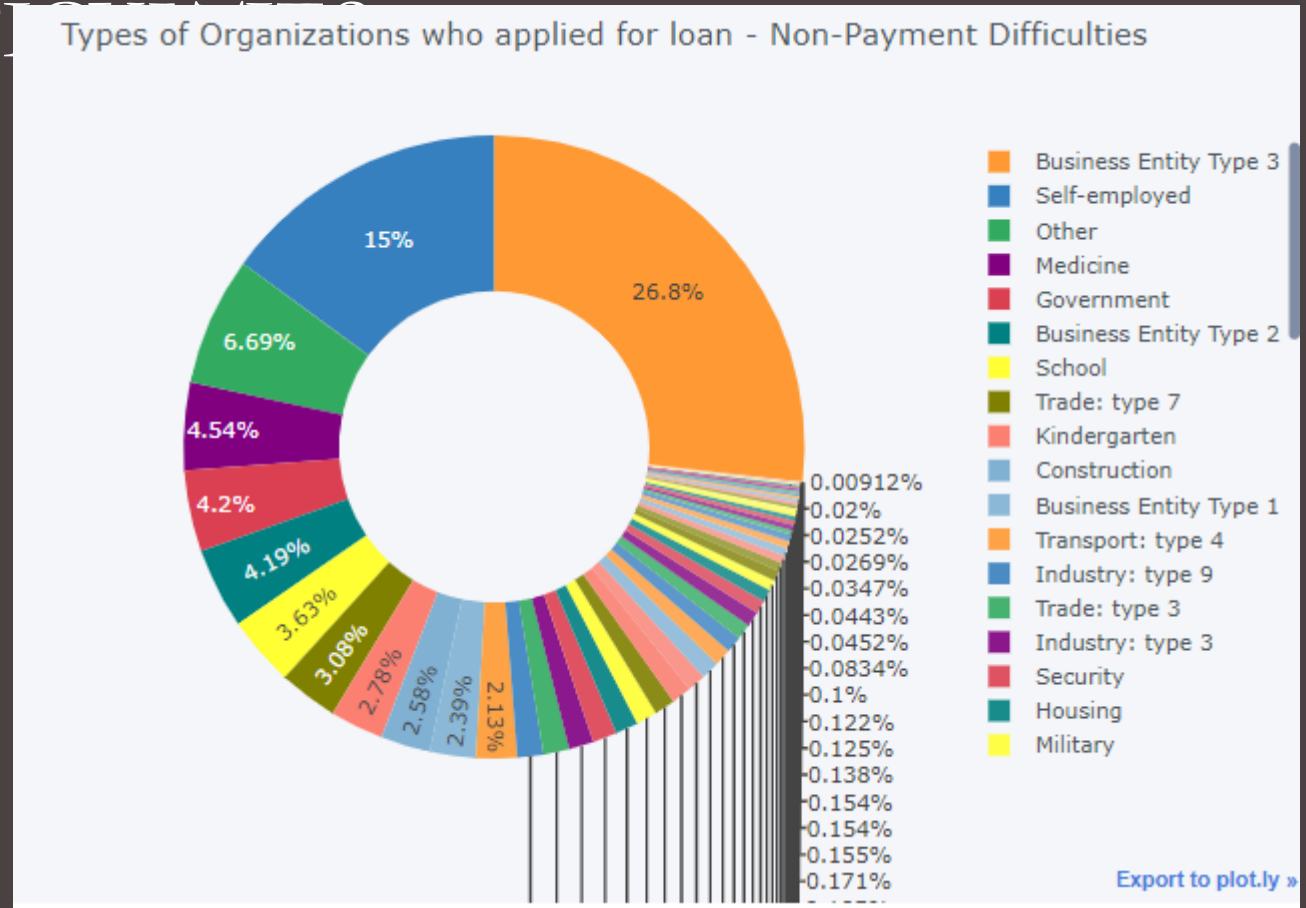
LOAN-NON PAYMENT DIFFICULTIES ARE ACCOMPANIED BY

Around 11% of population in unaccompanied group are among the highest for loan –non payment difficulties.



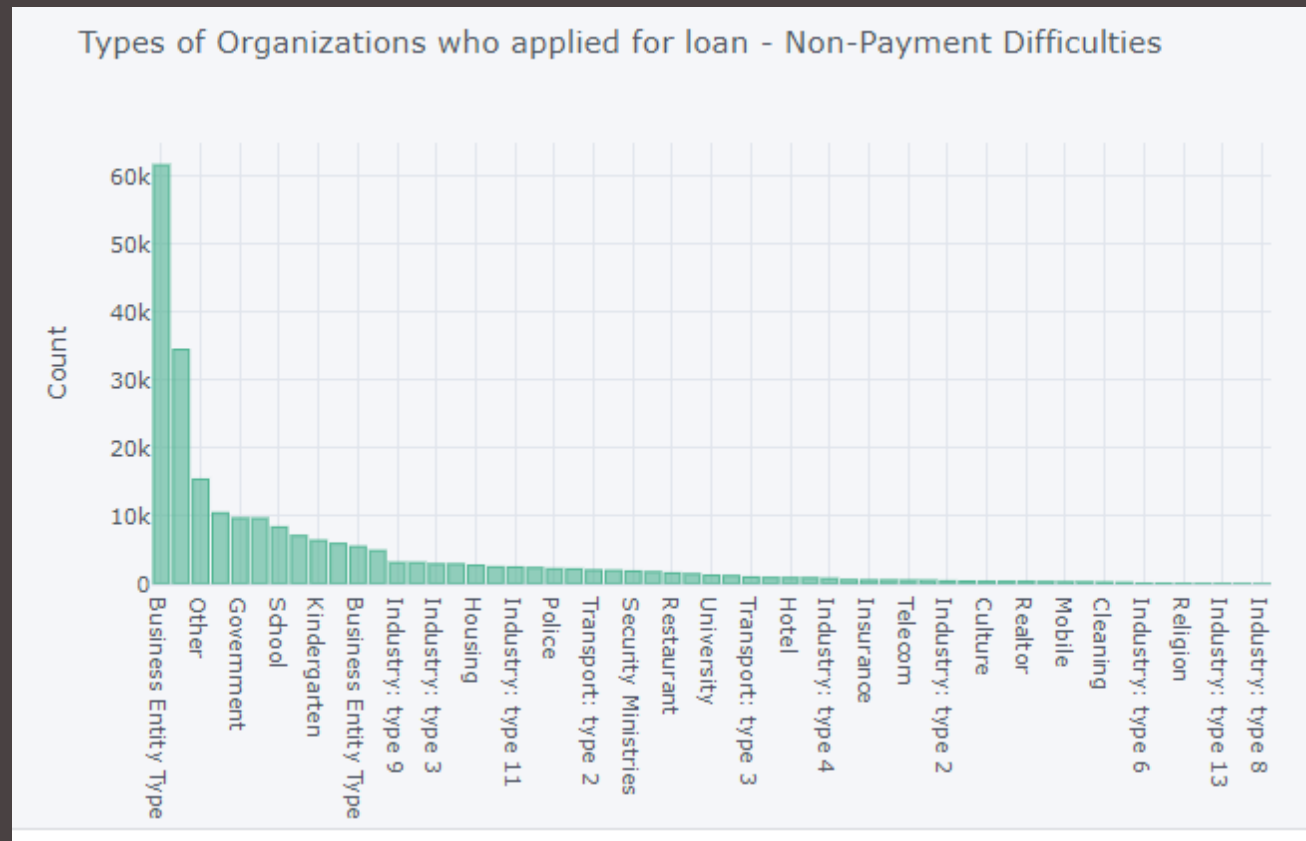
TYPES OF ORGANIZATIONS WHO APPLIED FOR LOAN - NON-PAYMENT DIFFICULTIES

Around 28 % of population from Business entity and 15% from self employed are among the highest for loan – non payment difficulties.



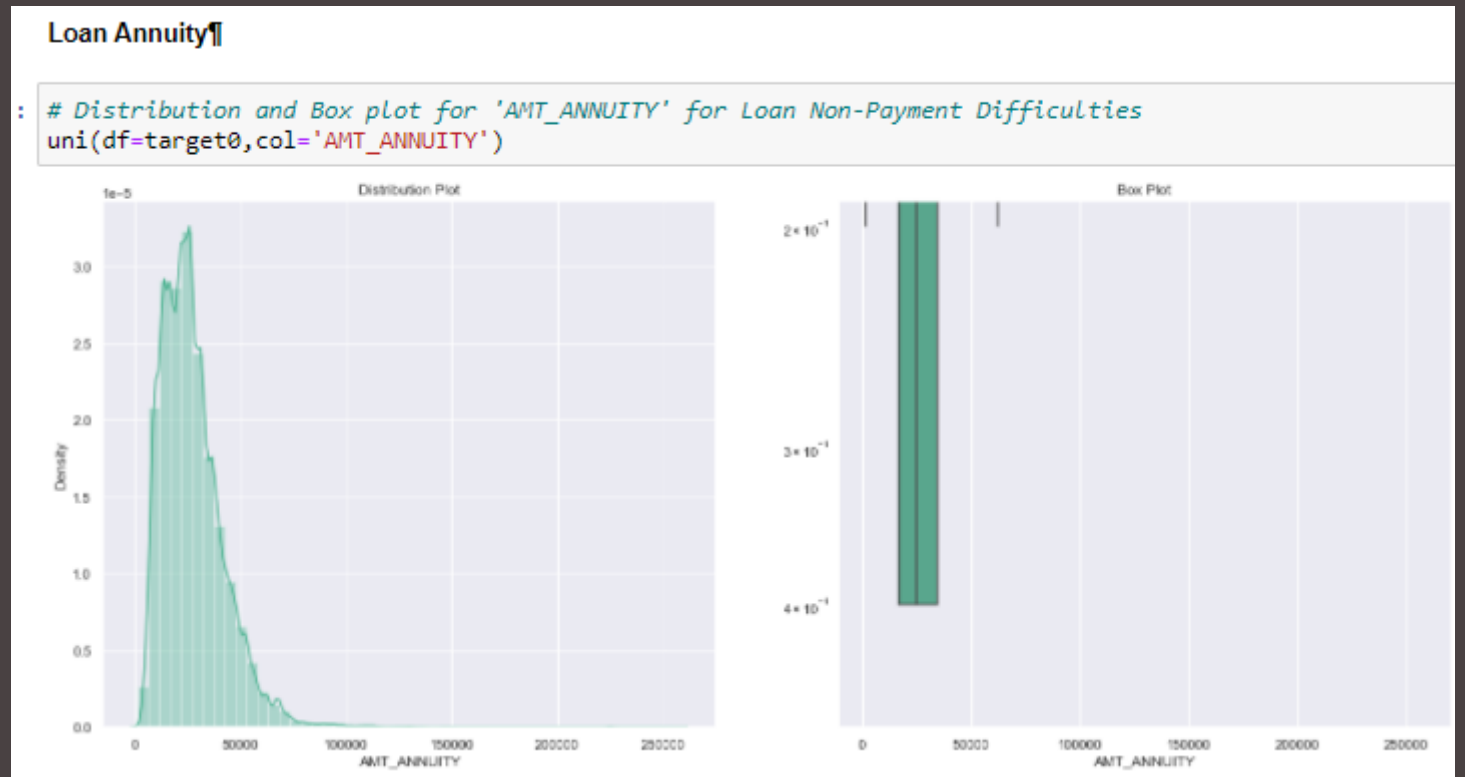
TYPES OF ORGANIZATIONS WHO APPLIED FOR LOAN - NON-PAYMENT DIFFICULTIES

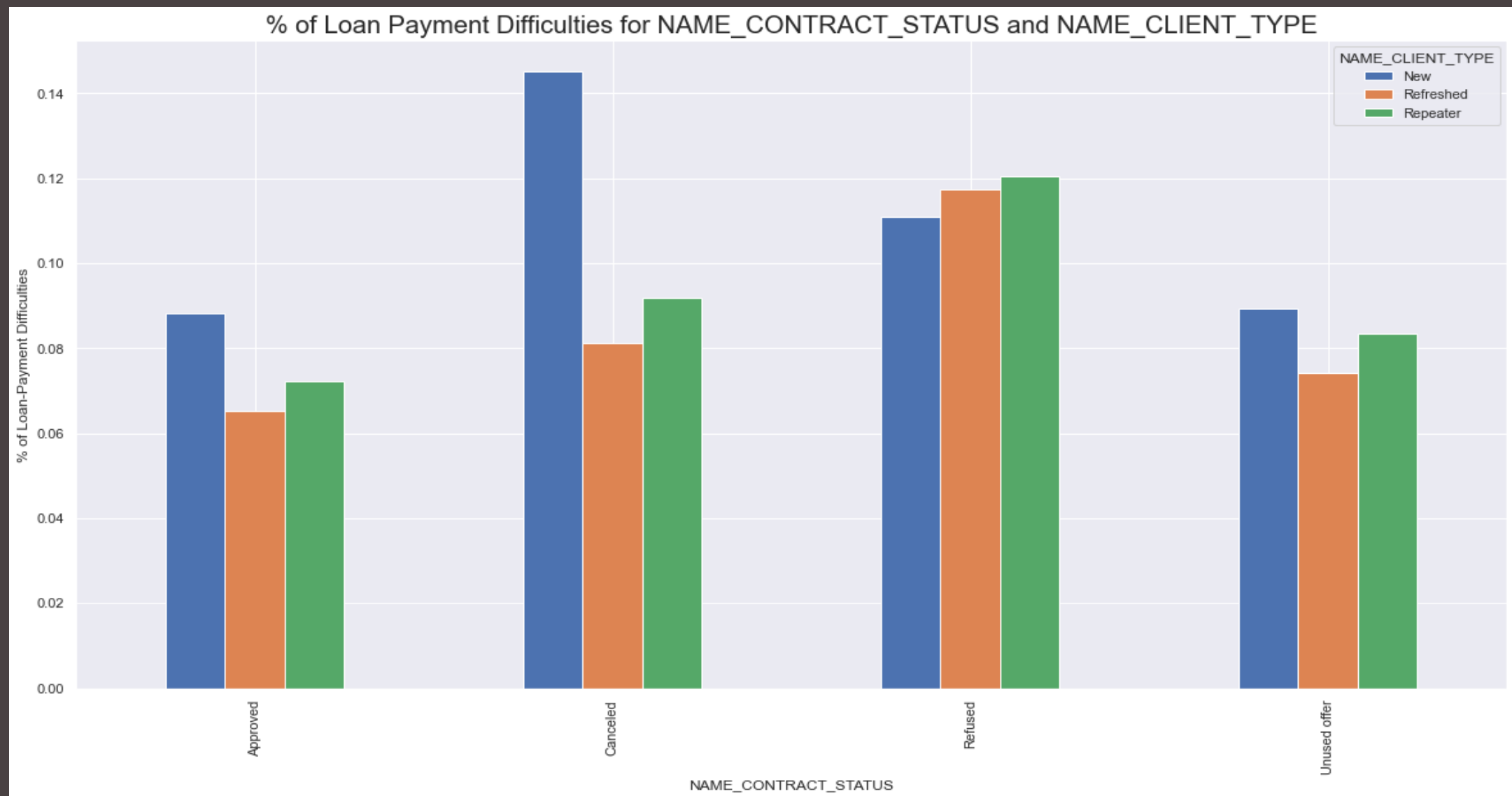
Business entity group
are the top applicant
with 60k application,
followed by others and
Government.



DISTRIBUTION AND BOX PLOT FOR 'AMT_ANNUIITY' FOR LOAN NON-PAYMENT DIFFICULTIES

Loan





It can be observed from the above graph that Client who where 'New' and had 'Cancelled' previous application tend to have more % of Loan-Payment Difficulties in current application

CONCLUSION (1/2):

Insights

1 Application Data

- The count of 'Maternity Leave' in 'NAME_INCOME_TYPE' is very less and it also has maximum % of payment difficulties- around 40%. Hence, client with income type as 'Maternity leave' are the driving factors for Loan Defaulters.
- The count of 'Low skilled Laborers' in 'OCCUPATION_TYPE' is comparatively very less and it also has maximum % of payment difficulties- around 17%. Hence, client with occupation type as 'Low skilled Laborers' are the driving factors for Loan Defaulters.
- The count of 'Lower Secondary' in 'NAME_EDUCATION_TYPE' is comparatively very less and it also has maximum % of payment difficulties- around 11%. Hence, client with education type as 'Lower Secondary' are the driving factors for Loan Defaulters.

CONCLUSION (2/2):

Insights

2 Previous Application Data

- The count of 'Refusal to name the goal' in 'NAME_CASH_LOAN_PURPOSE' is comparatively very less and it also has maximum % of payment difficulties- around 23%. Hence, clients who have 'Refused to name the goal' for cash loan in previous application are the driving factors for Loan Defaulters.
- The count of 'Refused' in 'NAME_CONTRACT_STATUS' is comparatively less and it also has maximum % of payment difficulties- around 12%. Hence, client with contract status as 'Refused' in previous application are the driving factors for Loan Defaulters.
- The count of 'Revolving Loans' in 'NAME_CONTRACT_TYPE' is comparatively very less and it also has maximum % of payment difficulties- around 10%. Hence, client with contract type as 'Revolving loans' in previous application are the driving factors for Loan Defaulters.
- It can be observed from the graph that Clients with 'Revolving loans' and with 'Refused' previous application tend to have more % of payment difficulties in current application. Since the count of both 'Revolving loans' and 'Refused' is comparatively less(from the graphs in previous slide), clients with 'Revolving Loans' and 'Refused' previous application are driving factors for Loan Defaulters

