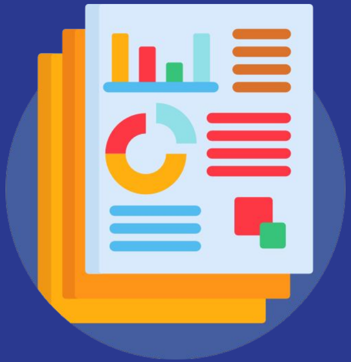
Credit EDA Assignment

DS-38 | Sagarmoy Sarkar



Problem statement

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. In this case study we will use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected so that there is minimum risk and the company can avoid financial losses.

Addressing Problem Statement

Risks associated with the bank's decision:

- 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.

The company wants to understand the driving factors(or driver variables) behind loan default, i.e. the variables which are strong indicators of defaulter.

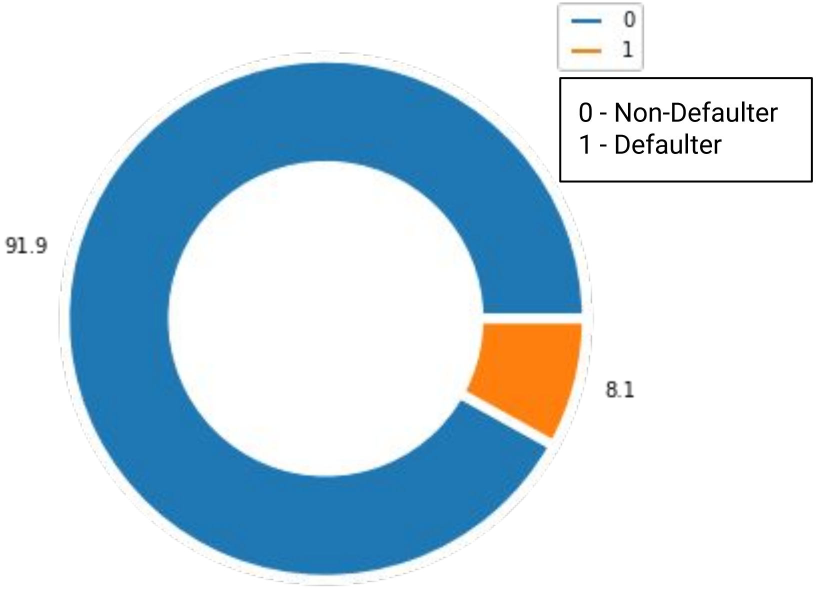
Aim of the Case Study

- This case study aims to 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. Identification of such applicants using EDA is the aim of this case study.

Initial Steps of EDA

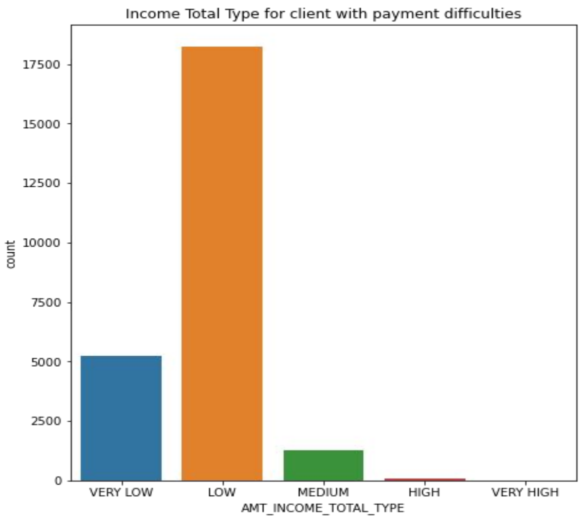
- Importing the important libraries
- Reading the Dataset
- Check for missing values and use correct approach to handle them
- Checking Outliers
- Finding Imbalance in data
- Finding the most relevant correlations between important variables

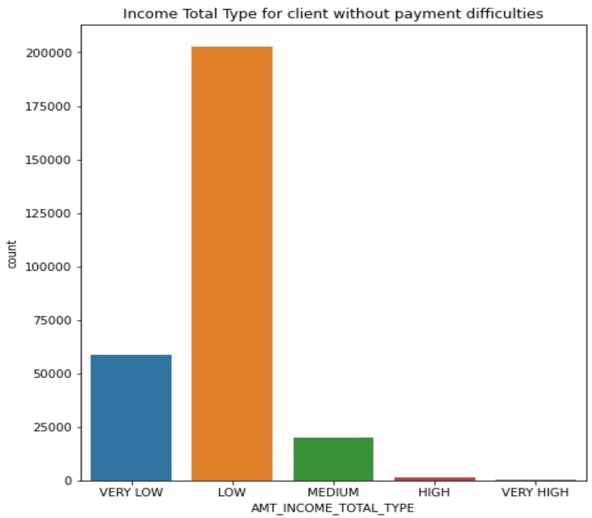
Defaulter vs Non-Defaulter Ratio



Almost 92% of people/applicants do not have payment difficulties and are non-defaulters but remaining 8% of people/applicants have payment difficulties and are defaulters.

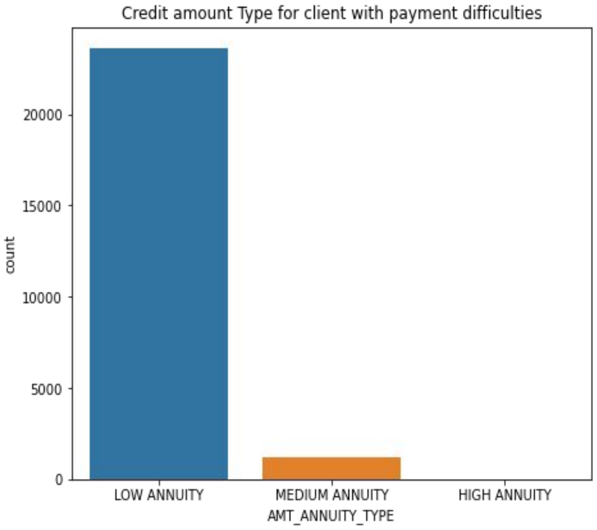
Comparing the Income Total Type of Defaulter vs Non-Defaulter

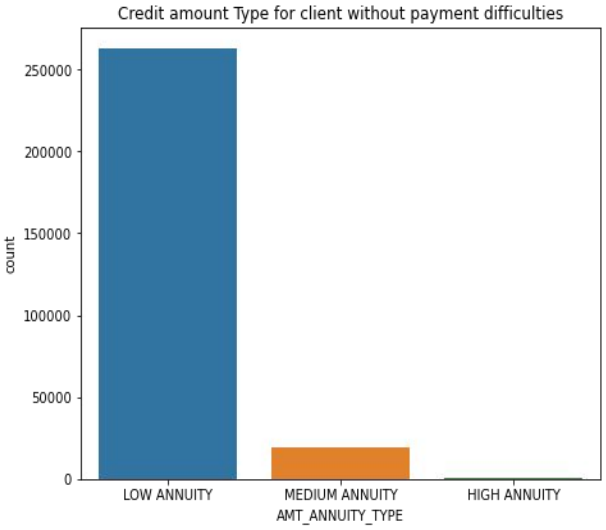




Clients with very high income has no difficulty regarding the payments of the loan, whereas clients with lower income has difficulty regarding the payments of the loan.

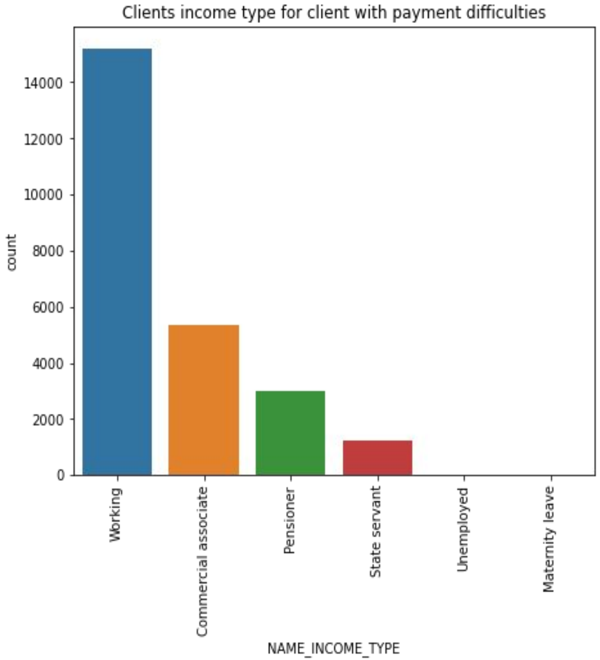
Comparing the Credit Amount Type of Defaulter vs Non-Defaulter

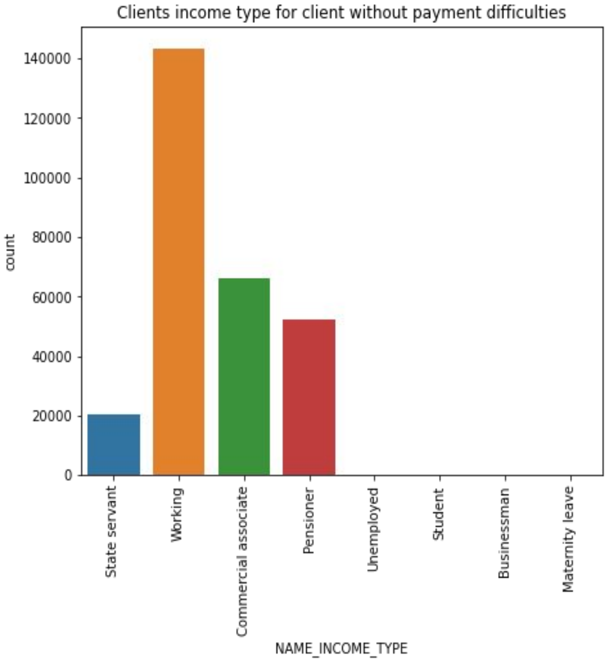




Higher the amount of loan is taken by the client, in those cases the payment difficulties are less.

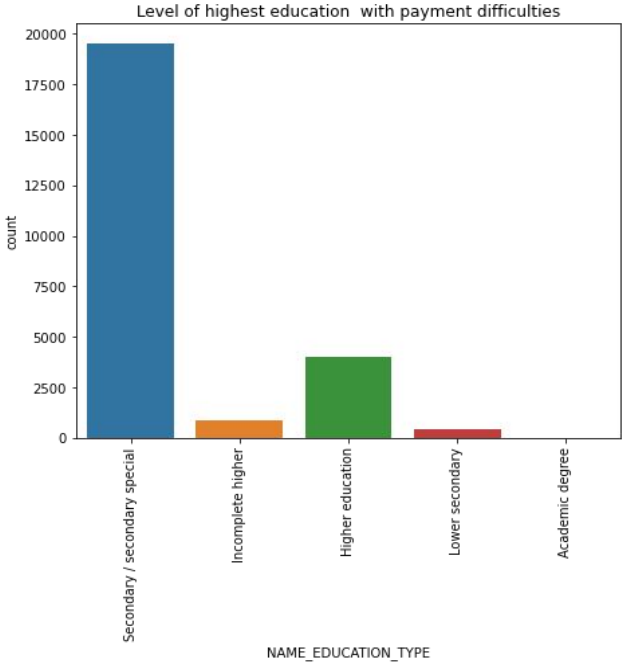
Comparing the Clients Income Type of Defaulter vs Non-Defaulter

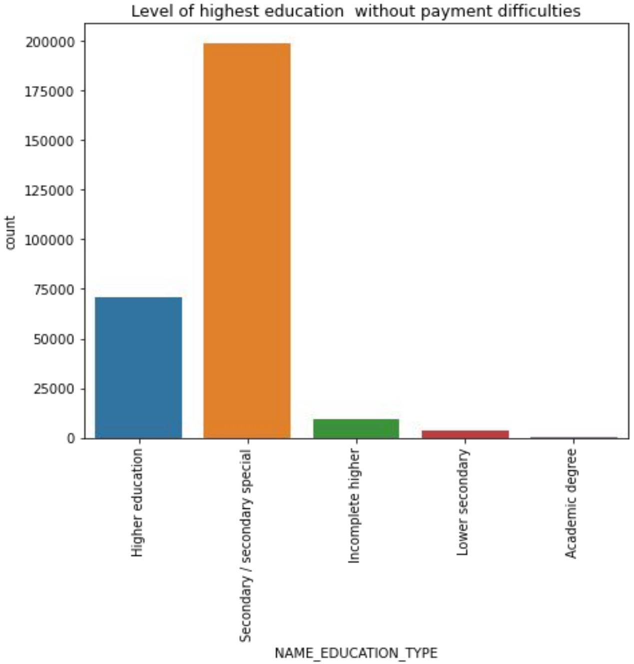




- Students and Businessman don't have any difficulties in the payment.
- Pensioner, Commercial Associate and State servants have a better chance of paying the loan without difficulty.
- For Working, Unemployed and Maternity leave clients other variables should be taken into account before approving the loan, as they have a 50-50 chance with payment difficulties.

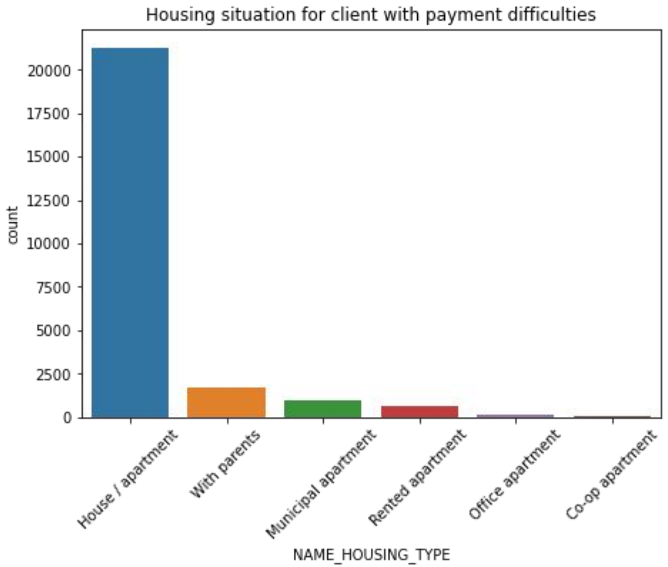
Comparing the Level of highest education of Defaulter vs Non-Defaulter

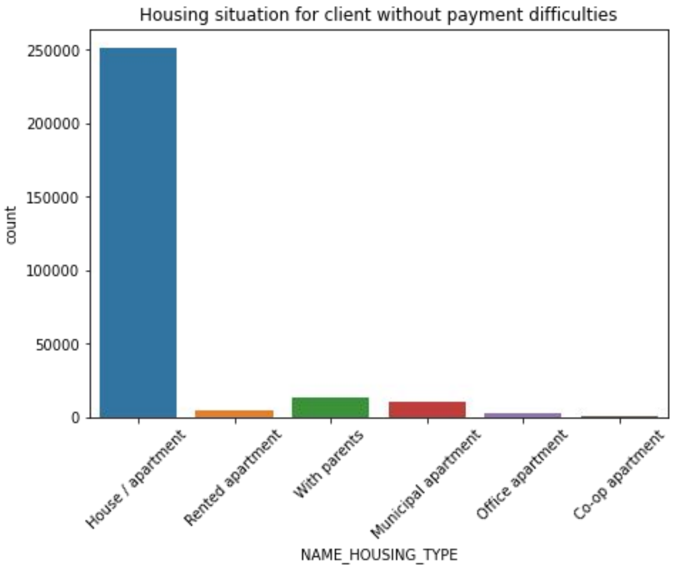




Clients with Academic degree and Higher education have less difficulty in payments, while those with Secondary or Incomplete Higher education have difficulty in payments.

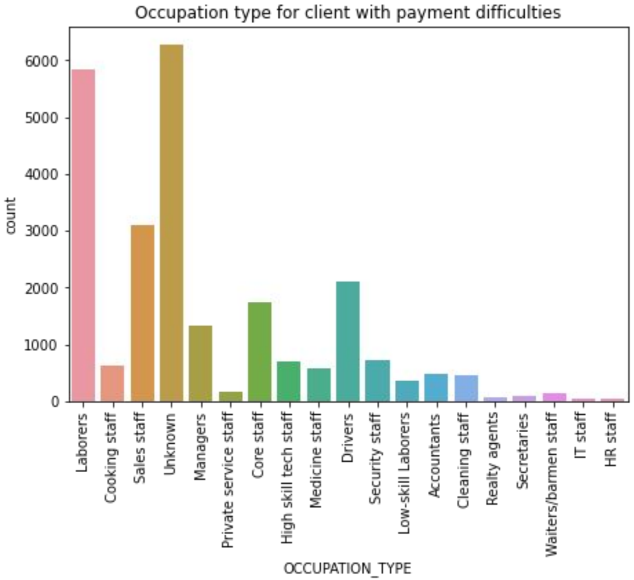
Comparing the housing situation of Defaulter vs Non-Defaulter

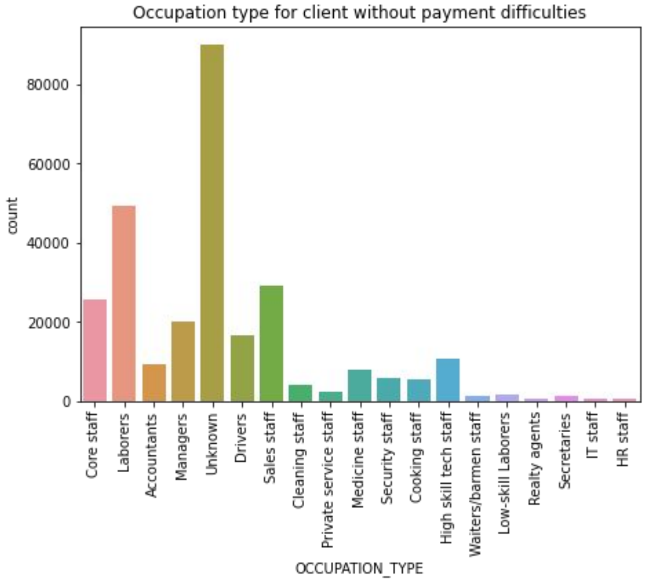




Clients who live in rented apartments or live with parents are more likely to have payment difficulties.

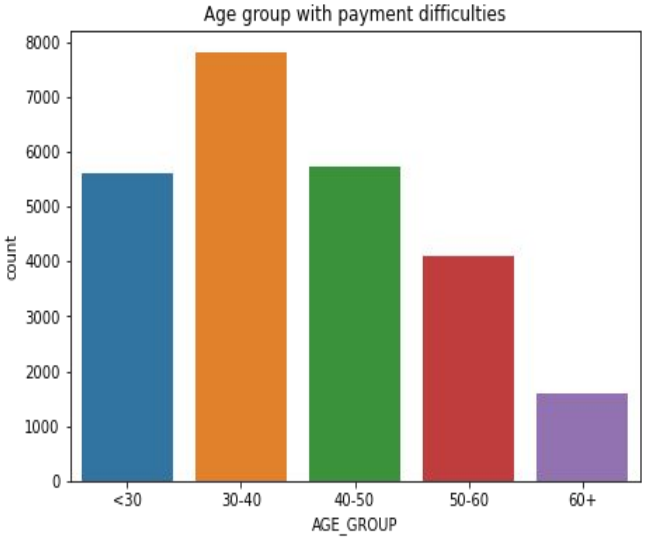
Comparing the occupation of Defaulter vs Non-Defaulter

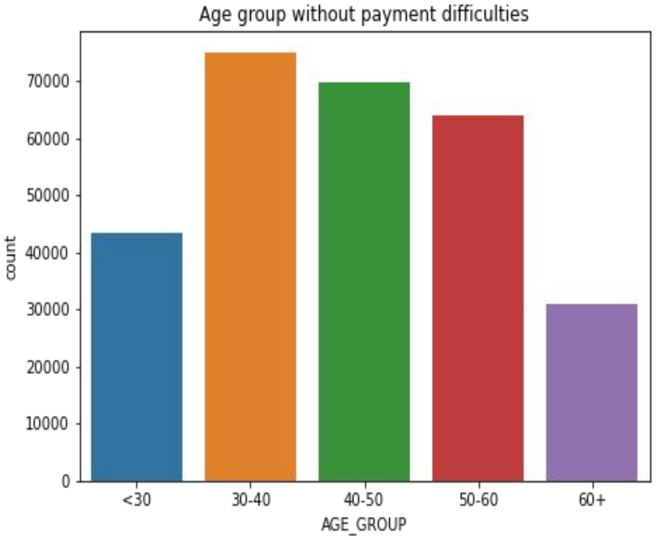




- Labourers, Drivers, Low-skill labourers, Cooking staff, Cleaning staff are more likely to have payment difficulties. Accountants, Managers, IT-staff, HR-staffs are less likely to have payment difficulties.
- As, Labourers and staffs earn less they have more payment difficulties rather than Accountants or Managers who earn more and have less payment difficulties.

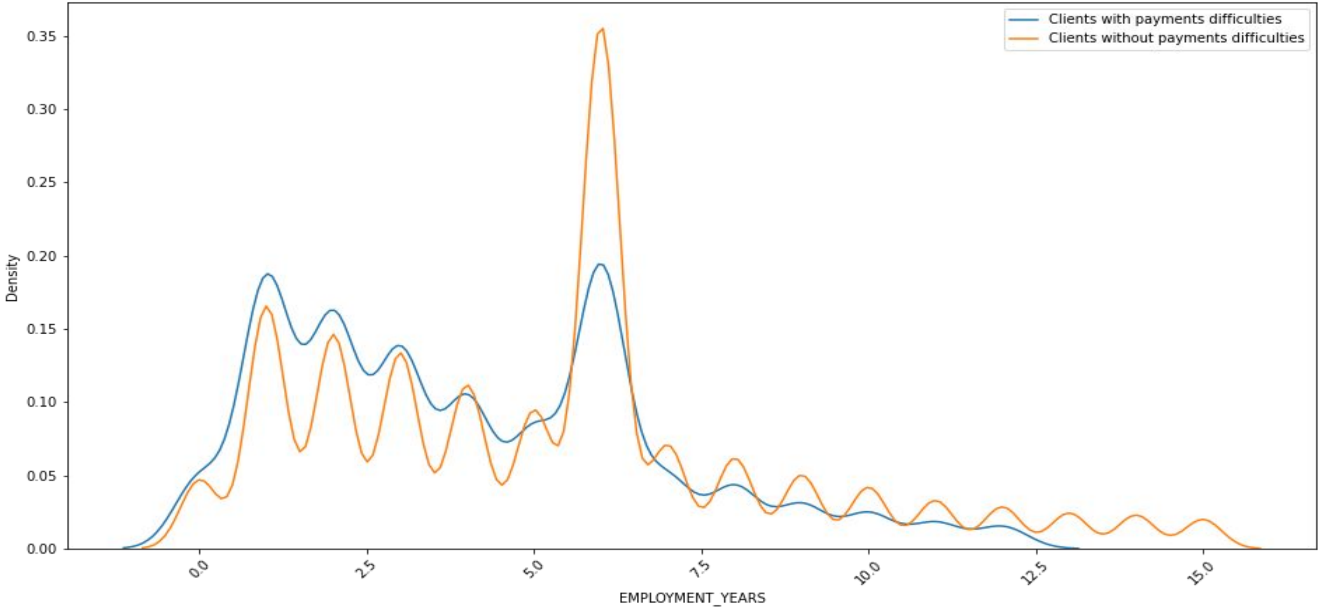
Compare clients from age group of Defaulter vs Non-Defaulter





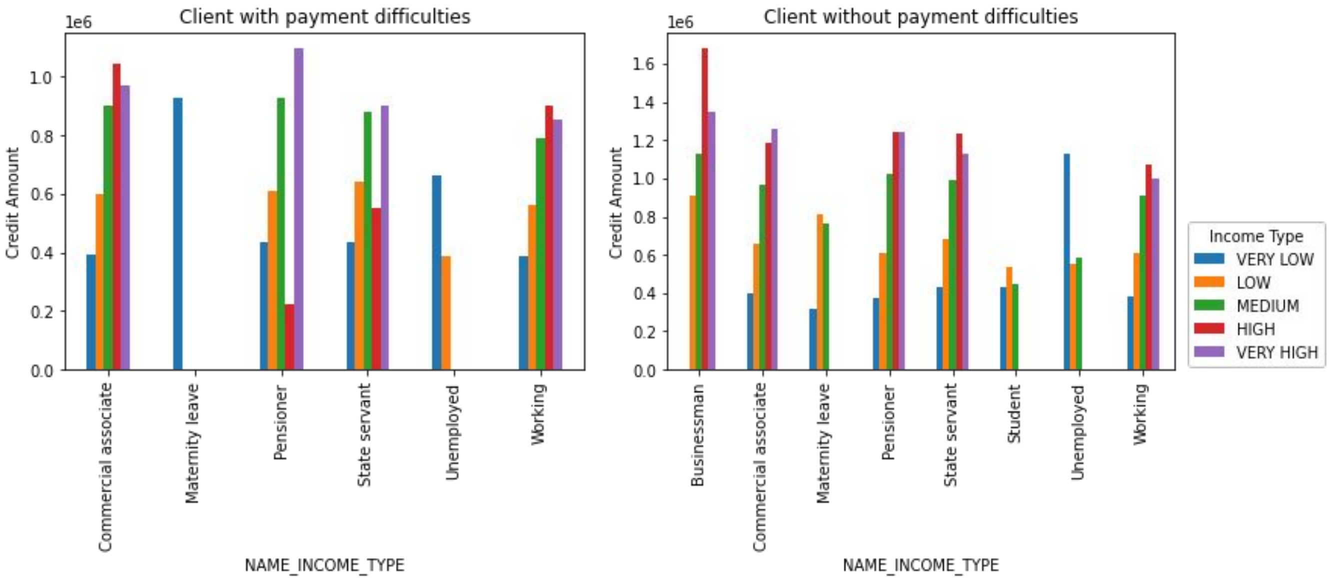
Age group of less than 30 and 30-40 has more difficulties to pay the loans. Age group of 40-50, 50-60 and 60+ has less difficulties to pay the loans.

Comparing Years of employment for Defaulter vs Non-Defaulter



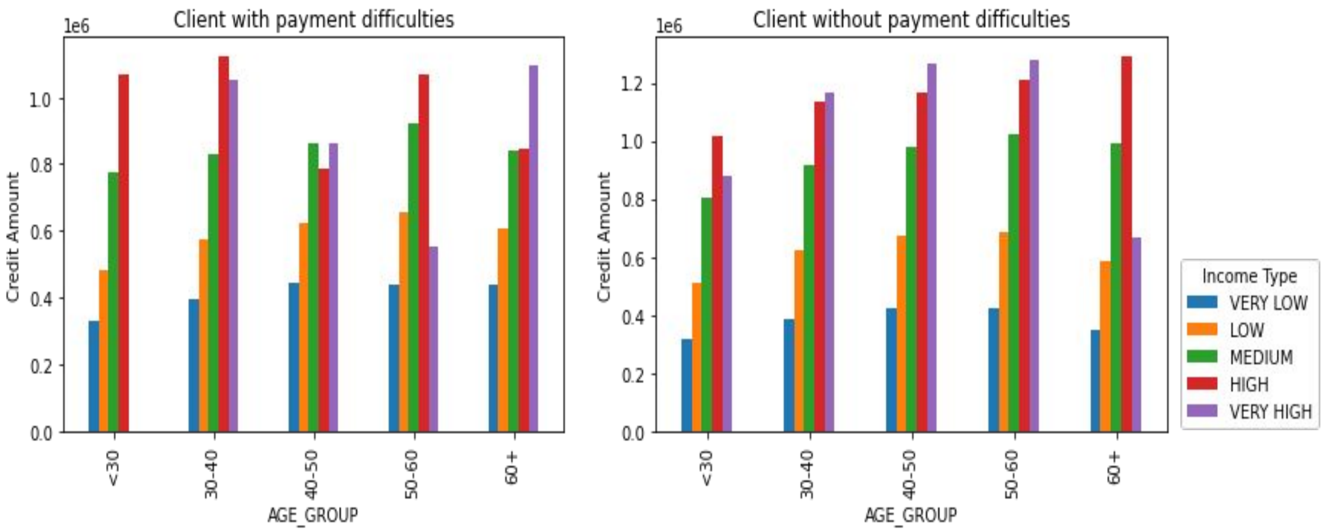
Clients who are employed more number of years do not face payment difficulties, whereas clients who are employed less number of years faces payment difficulties.

Checking how client Income Type gets Loan Credit on basis of their Income Total for Defaulter vs Non-Defaulter



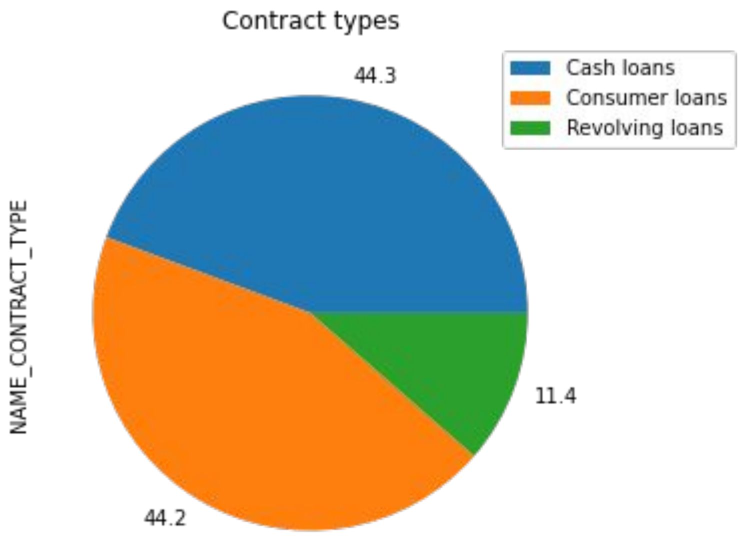
- Clients who have payment difficulty, their credit amount is also lower rather than Clients who don't have payment difficulty.
- Businessman who has a "HIGH" Income gets most Loan Credit, whereas clients who are on Maternity leave and have "VERY LOW" Income gets the least Loan Credit.
- Income type of Maternity leave with low income tends to be a good indication of defaulter
- Businessman and students can be a good client for the company as they don't have such payment difficulties.

Checking which age group gets how much Loan Credit on basis of their Income Total for Defaulter vs Non-Defaulter



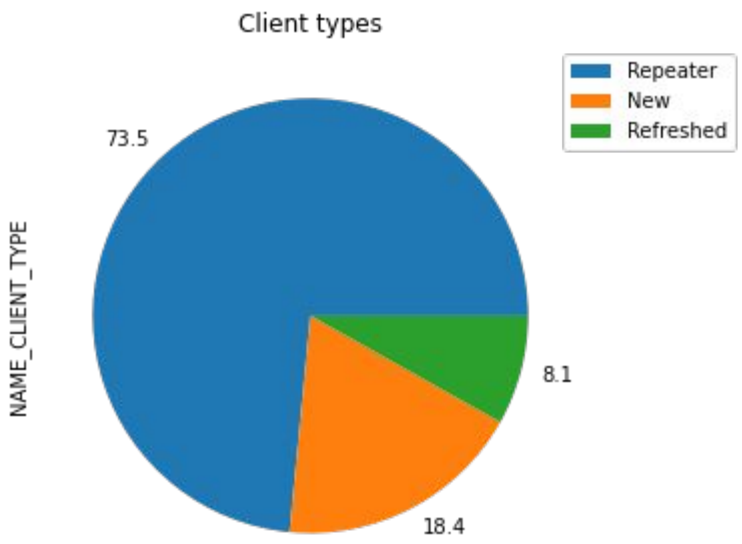
- Clients who have difficulty in payments, if their age is less than 30 or between 30-40 and Income is low then they get less loan credit as they clearly shows indication of becoming a defaulter
- Clients who don't have difficulty in payments, if they are in age group of 40-50, 50-60 or more than 60 and have a high Income gets more loan credit.

Checking Contract product type for previous application data



More than 40% of the loans taken are either Cash loans or Consumer loans and more than ~11% loans taken are Revolving loans.

Was the client old or new client when applying for the previous application

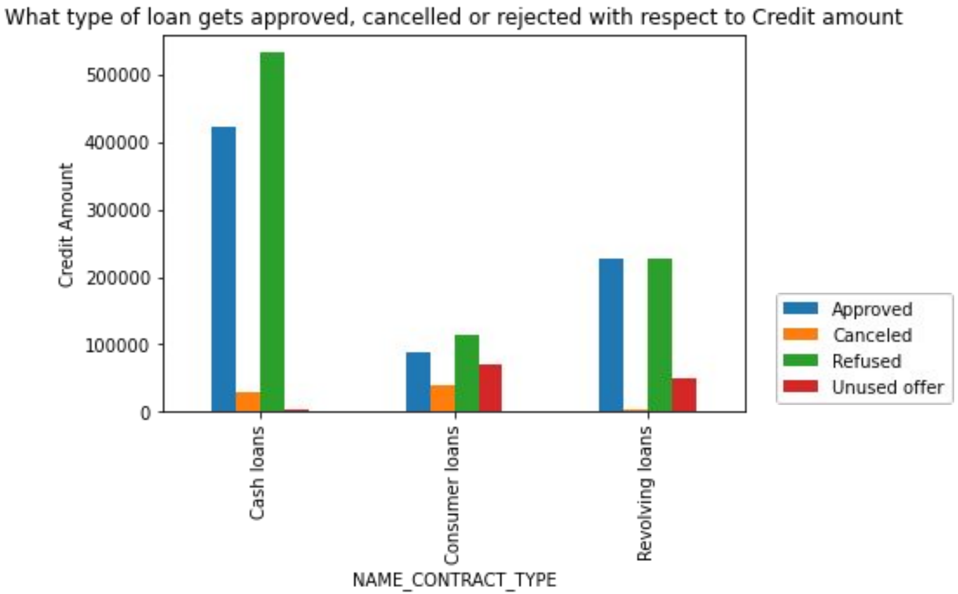


Most of the previous loan application received are from repeater customers/clients. New clients and refreshed clients are comparatively less.

Correlation between Goods price, Application amount, Credit amount and Annuity amount

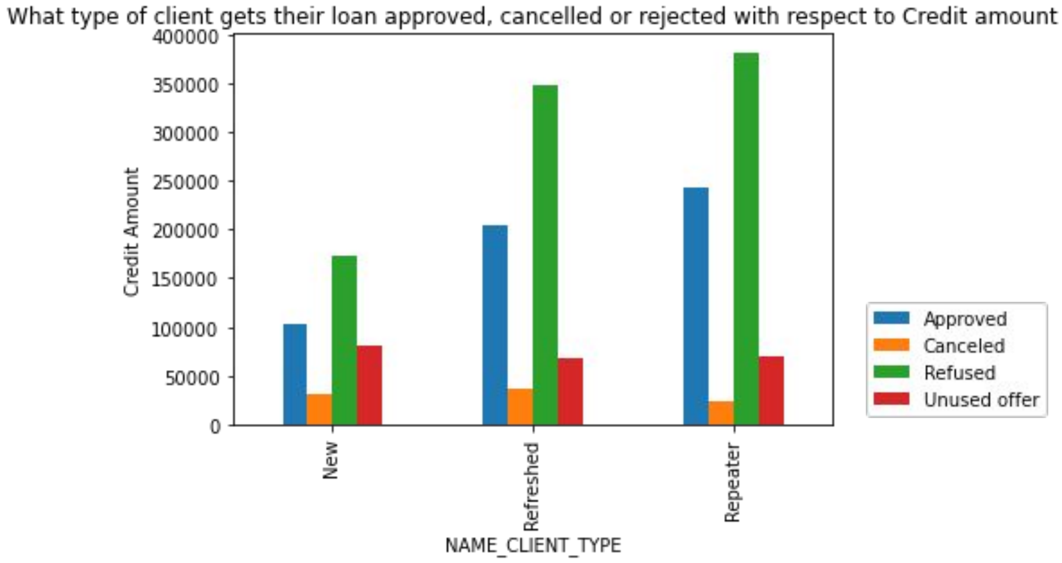
Goods price amount, Application amount, Credit amount and Annuity amount has a very high positive correlation among themselves. If any of them increases then others increases as well.
Let's say if a client wants to buy something whose price is 25,000 then application amount and credit amount may lie between 25,000-30,000(more or less). But, if a client wants to buy something whose price is 100,000 then application amount and credit amount may lie between 100,000-120,000(more or less). Similarly, the annuity also increases with the credit amount.

Checking how types of loans gets approved, cancelled or rejected with respect to Credit amount



- Cash loans delivers the most credit amount
- Cash loans of high credit amounts gets approved more rather than Consumer and Revolving loans.

Checking how client types gets approved, cancelled or rejected with respect to Credit amount



- If the client is a repeater and his application amount is high, then most likely their loan gets approved rather than new and refreshed clients.
- Overall loan approval rate is more on less credit amount and are rejected if credit amount is more.

Conclusion

- All the points can make it clear that to which client/customer can the bank give loan credit so that they are not a defaulter in the future.
- The bank/finance company can tackle the risks of controlling the loss of business to the company and avoid financial loss for the company.

References

- Download the dataset from the link below. https://drive.google.com/open?id=16RQztUqCfJOlbooHqYlJrp6Q7iL65uZB
- Jovian link- https://jovian.ai/sarkar-sagarmoy/credit-eda-assignment