

# CREDIT-EDA CASE STUDY

By

- Prashant Chatterjee
- Joe Paul Freddy Jones

# Table of Content

- Introduction
- Analysis and Inferences
- Final Observations and Recommendation

# Introduction – Problem Statement & Goals

- When a client applies for loan, the company has to decide for loan approval based on the applicant's profile.
- Two types of risks are associated with the company's decision:
  - If the applicant is likely to repay the loan (Non- Defaulter), then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan(Defaulter), then approving the loan to such a client will lead to a financial loss for the company.
- Our goal is to analyze the new and previous application data to study the pattern of defaulters (TARGET) and give recommendations based on the observations which should help the company to make informed decision based on past data, and reduce the number of defaulters in future.

# Understanding the Data

- New Application Data

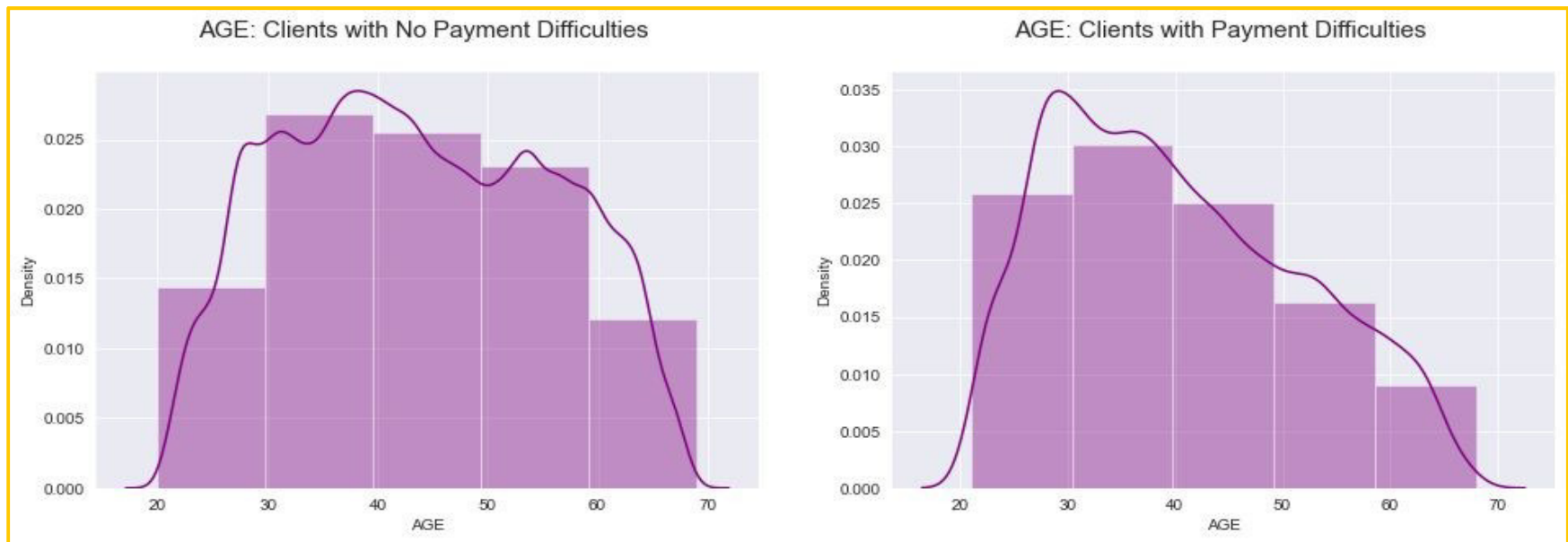
Contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

- Previous Application Data

Contains all the information about the client's previous loan data. It contains the data whether the previous application had been approved, cancelled, refused, or unused offer.

# Analysis and Inferences

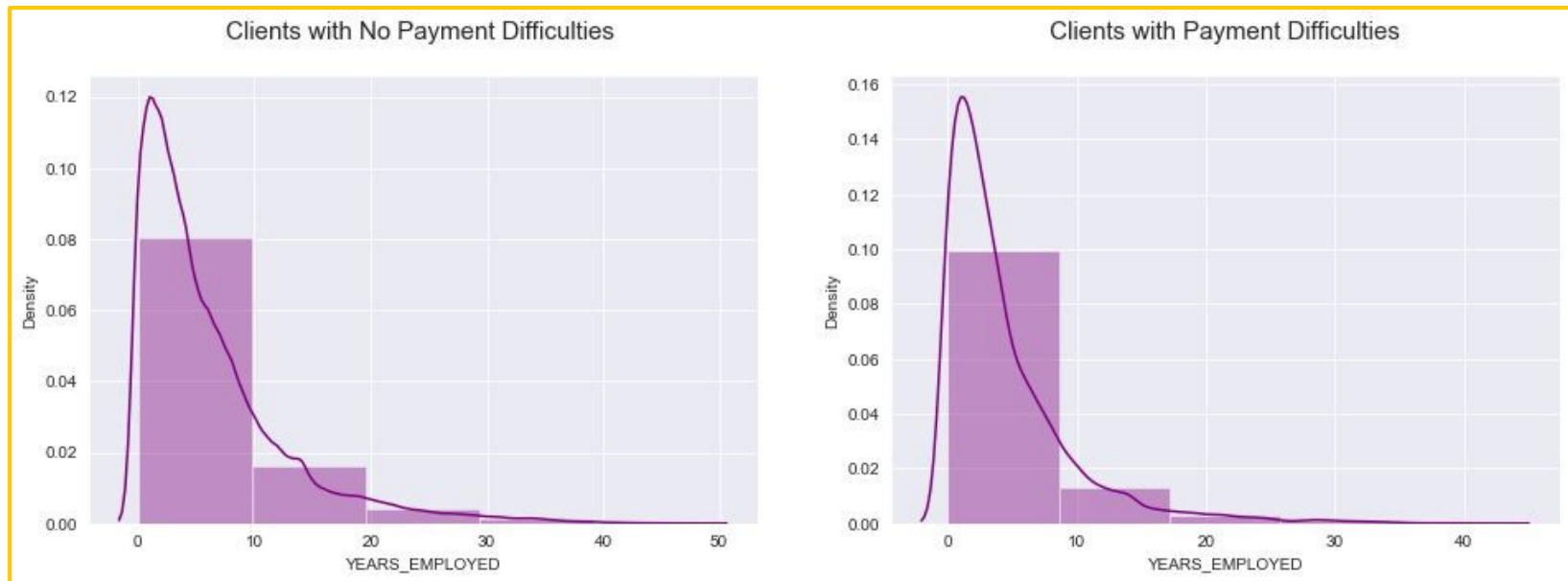
# Age Distribution of the Applicants



- Majority of the clients with payment difficulties are between the age of 30-40 and then it keeps on decreasing.
- Most of the clients with no payment difficulties are in the age group 30-60

# Years Employed

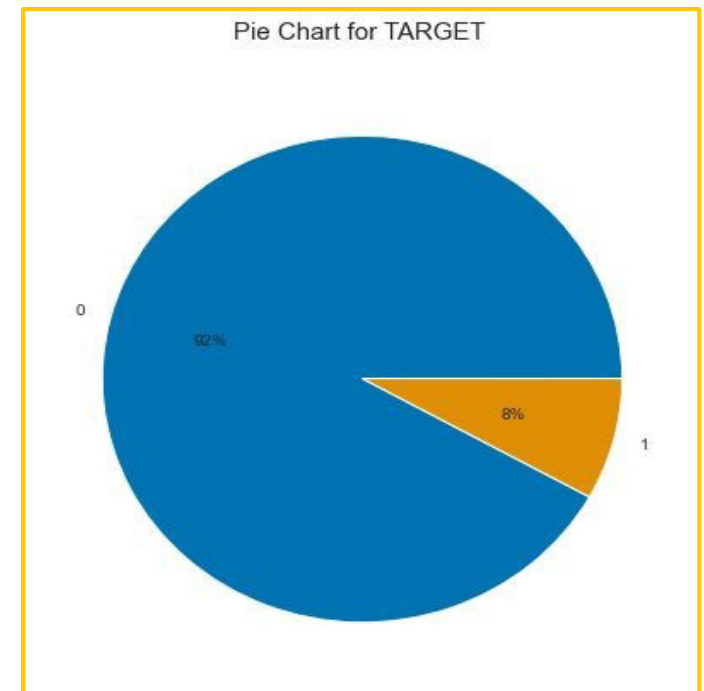
- For both clients with or without payment difficulties, the peak number of applicants are working with the current organization for less than 5 years of employment and drops considerably after that.



The data regarding Days Employed of the client has outliers which were recognized in our analysis and predictably skews the data, the plots were corrected to remove such outliers by adding a reasonable limit to the employment years.

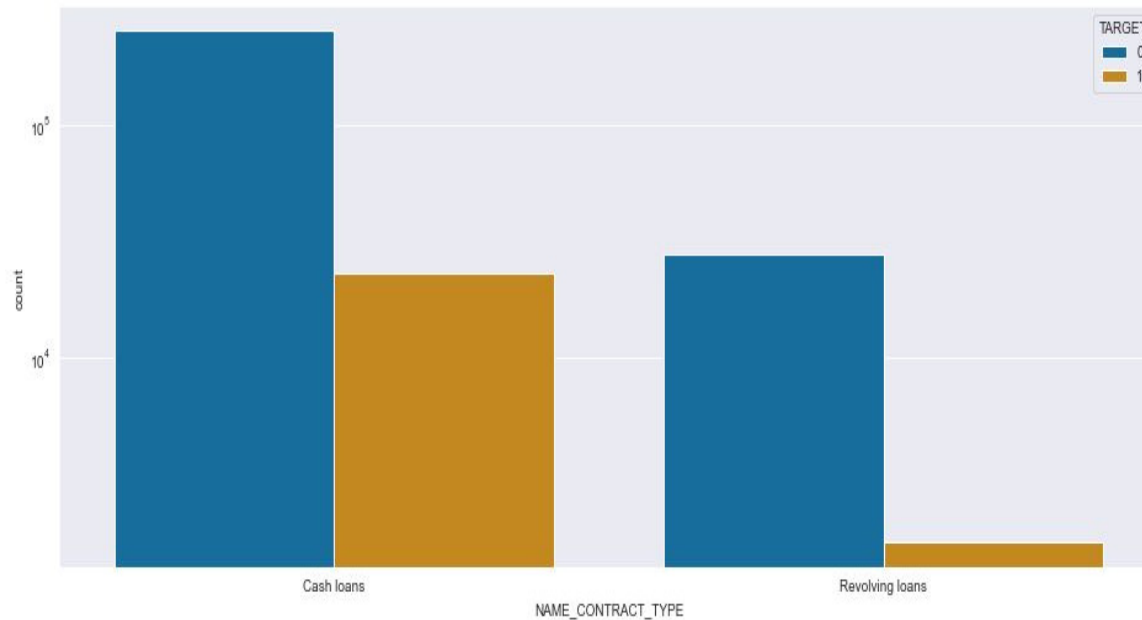
# DATA IMBALANCE

- As we can see that, there is high data imbalance between clients with payment difficulties and the other.
- Approx 92% of clients have paid the loan on time, while approx 8% of clients faced difficulties in paying the loan on time.
- The ratio of data imbalance for TARGET variable is 91.92 : 8.07
- The data is divided into two parts, Clients with Payment Difficulties (TARGET =1) and Clients with No Payment Difficulties (TARGET = 0) for further analysis.





# CONTRACT TYPE

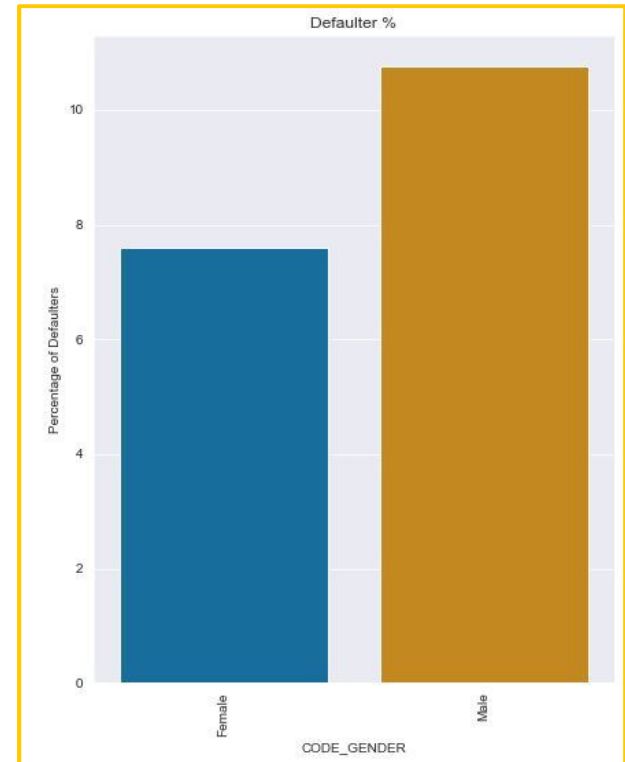
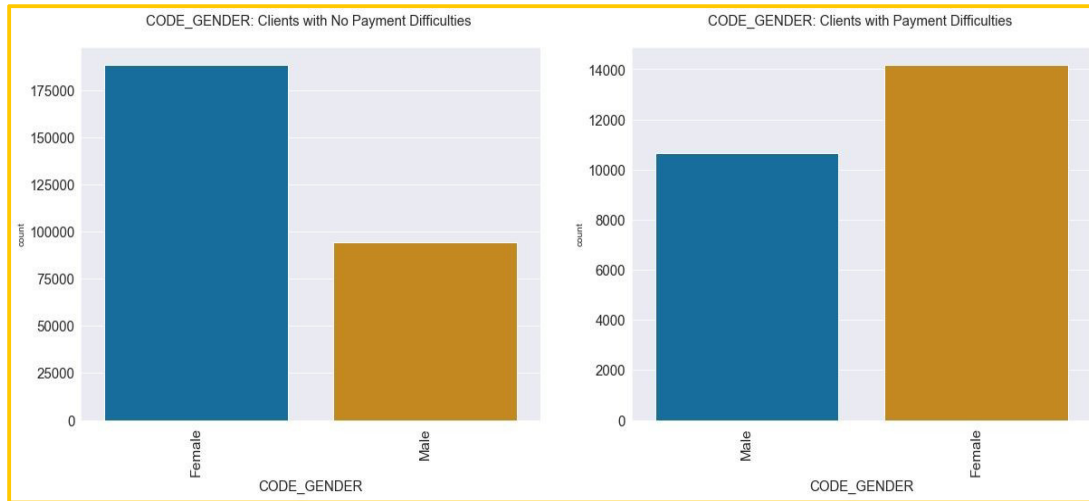


0 = Client with no payment difficulties

1 = Client with payment difficulties

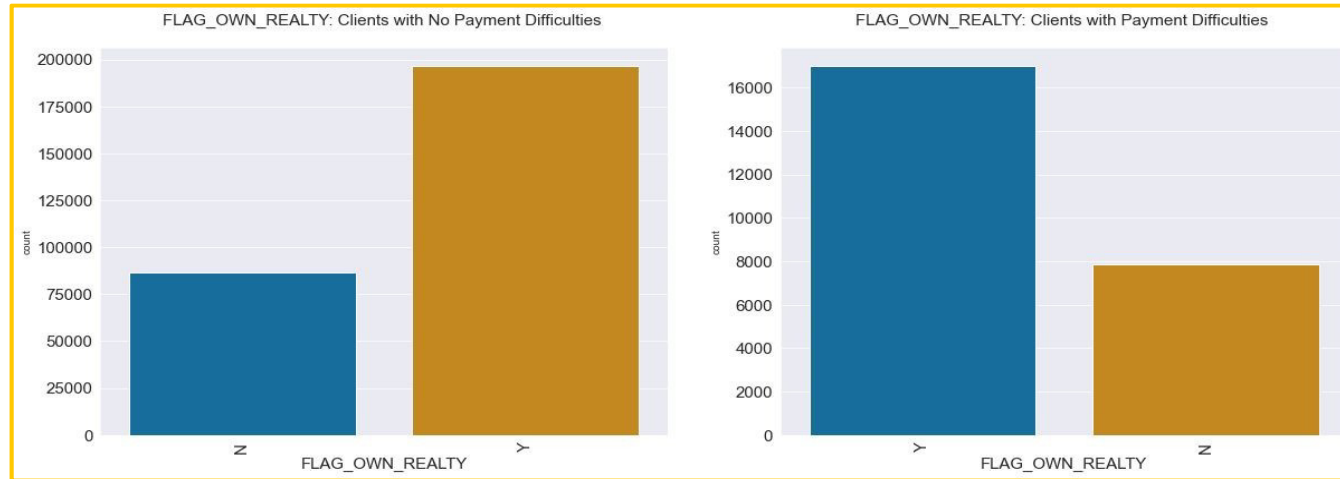
- Cash Loan overall are more popular than revolving loans
- Revolving loan have comparatively more defaulters

# GENDER

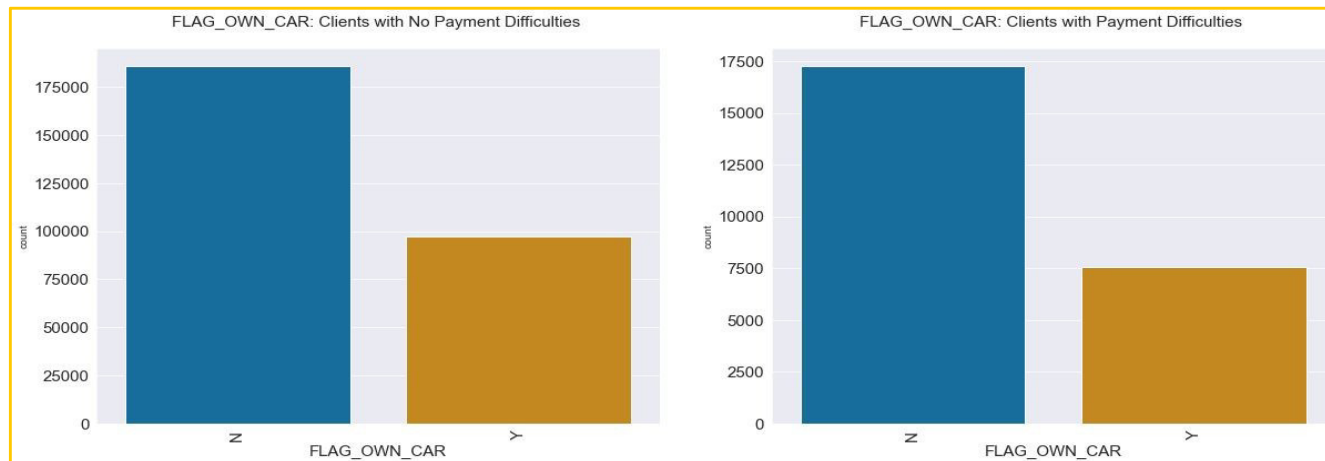


- Males have defaulted more than females, despite of the fact that there are more female clients.

# ASSETS

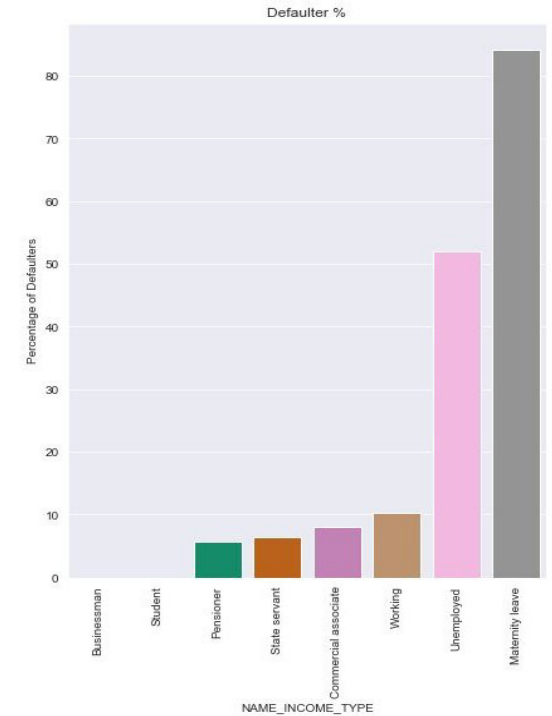
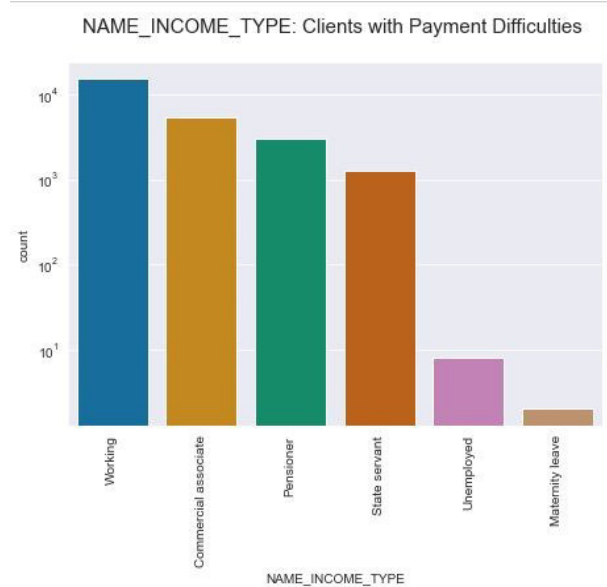
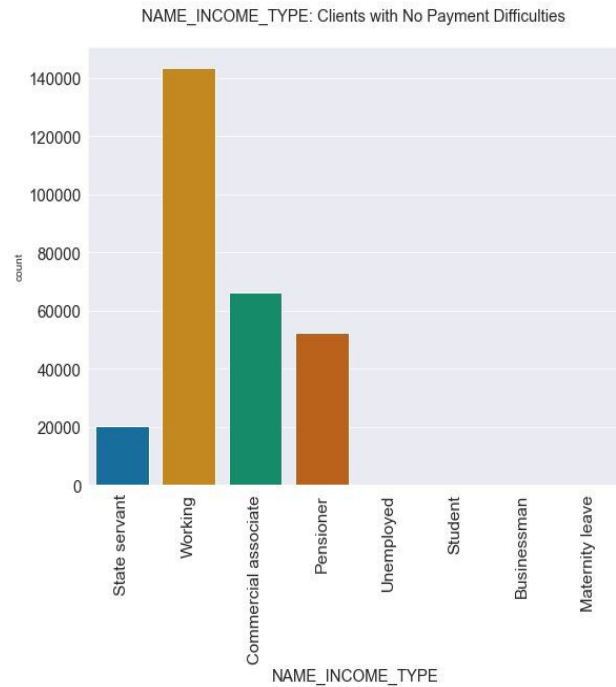


- Considerable number of clients with payment difficulties have realty, which might be the indicator of more ongoing loans and bigger expenses.



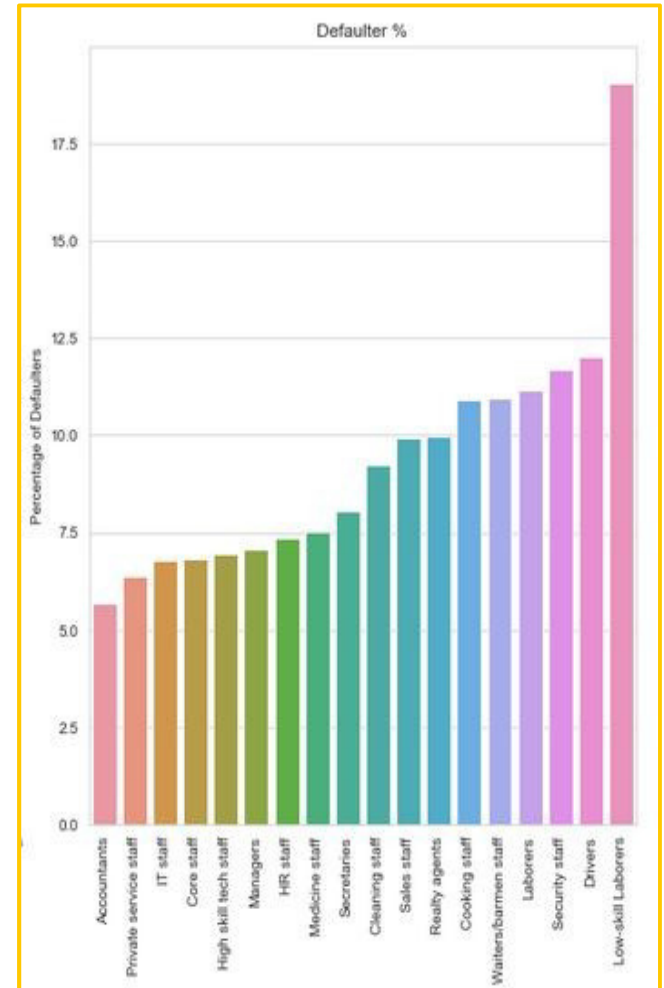
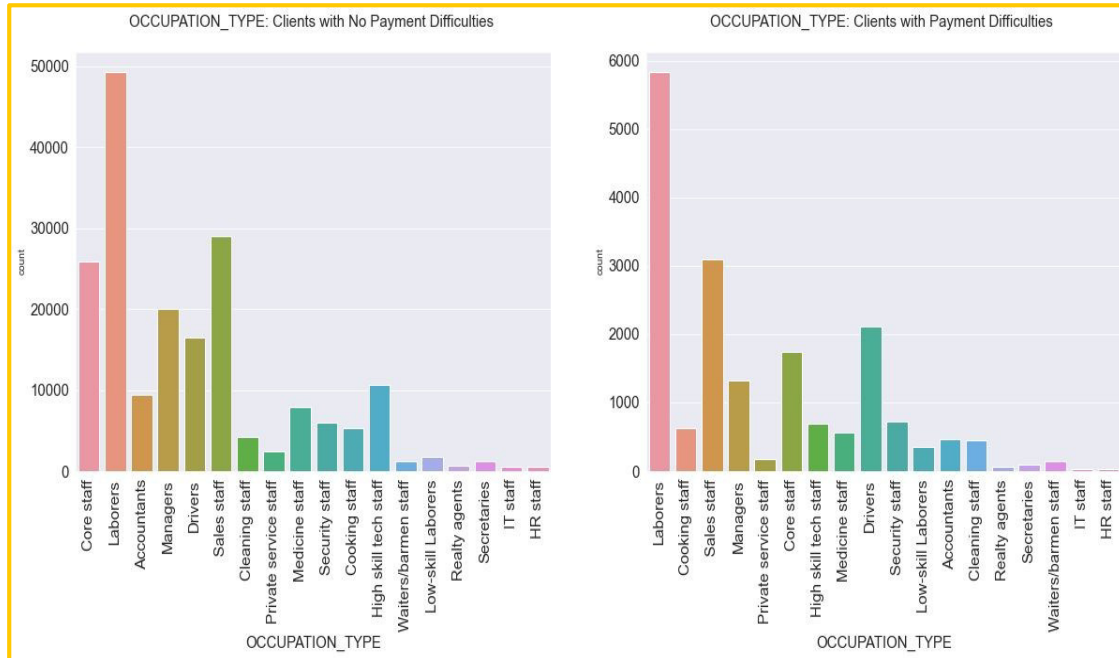
- Same cannot be said about the client owning a car, as those without cars, have defaulted more than those who have cars.

# INCOME TYPE



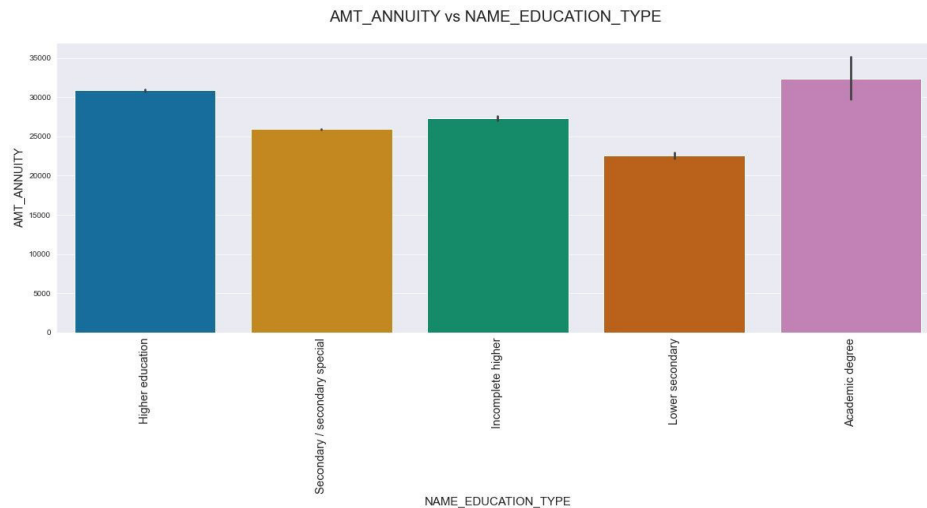
- The defaulter percent is high in maternity leave and unemployed category.
- While the count of non-defaulters is high in working income type. Followed by commercials and pensioners.

# CLIENT OCCUPATION

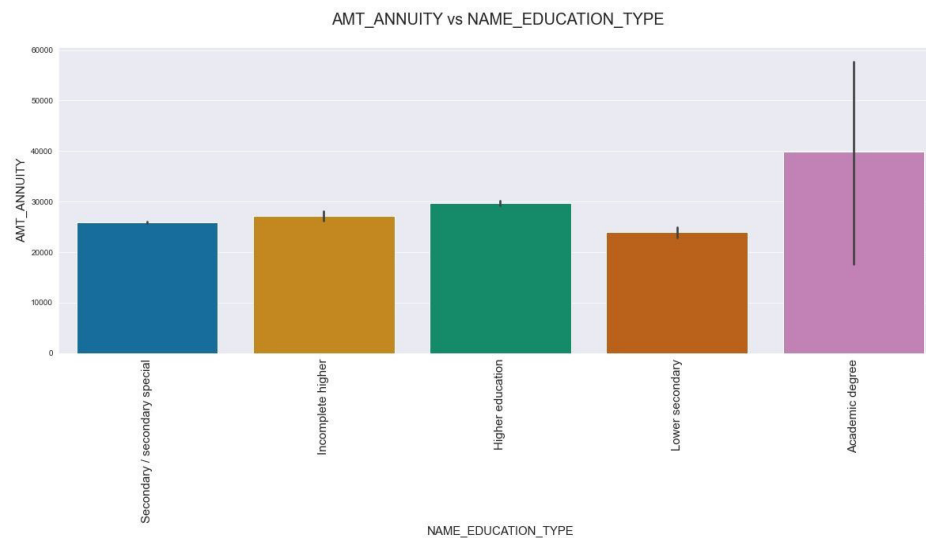


- A big portion of the clients are from occupation of Laborer, Sales Staff and core staff.
- Low skill laborers are by far the biggest defaulters, followed by Drivers and Security staff.

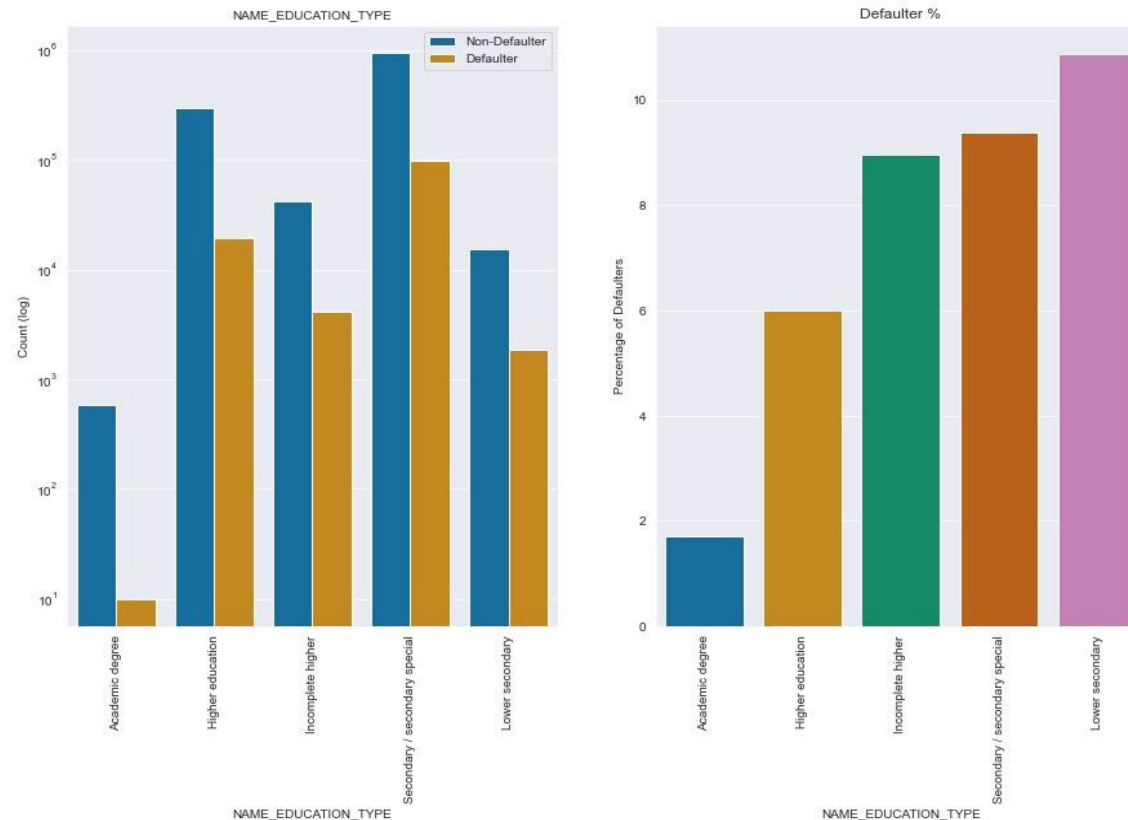
# ANNUITY VS. EDUCATION



- People with higher academic degrees are more likely to pay a much higher annuity, followed by clients with higher education
- Clients with academic degrees and with around 40,000 in annuity seem to have problems.

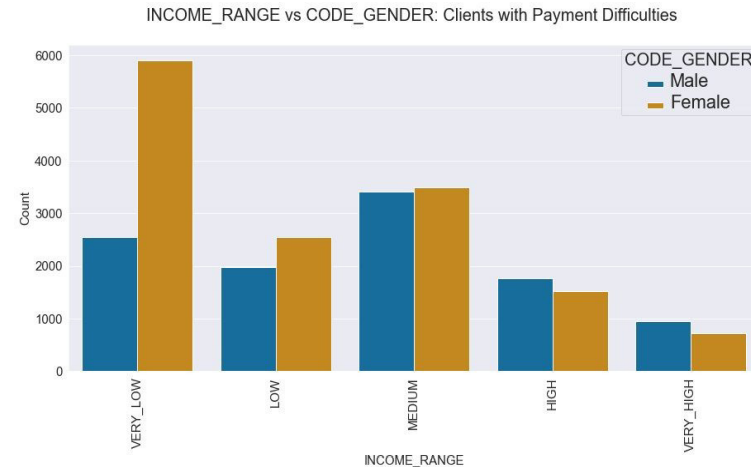


# DEFAULTER AS PER EDUCATION



- Clients having Lower Secondary education have the highest percentage of default followed by secondary education.
- Education is an important factor while approving loans as academic degree holder clients have quite low % of defaults.

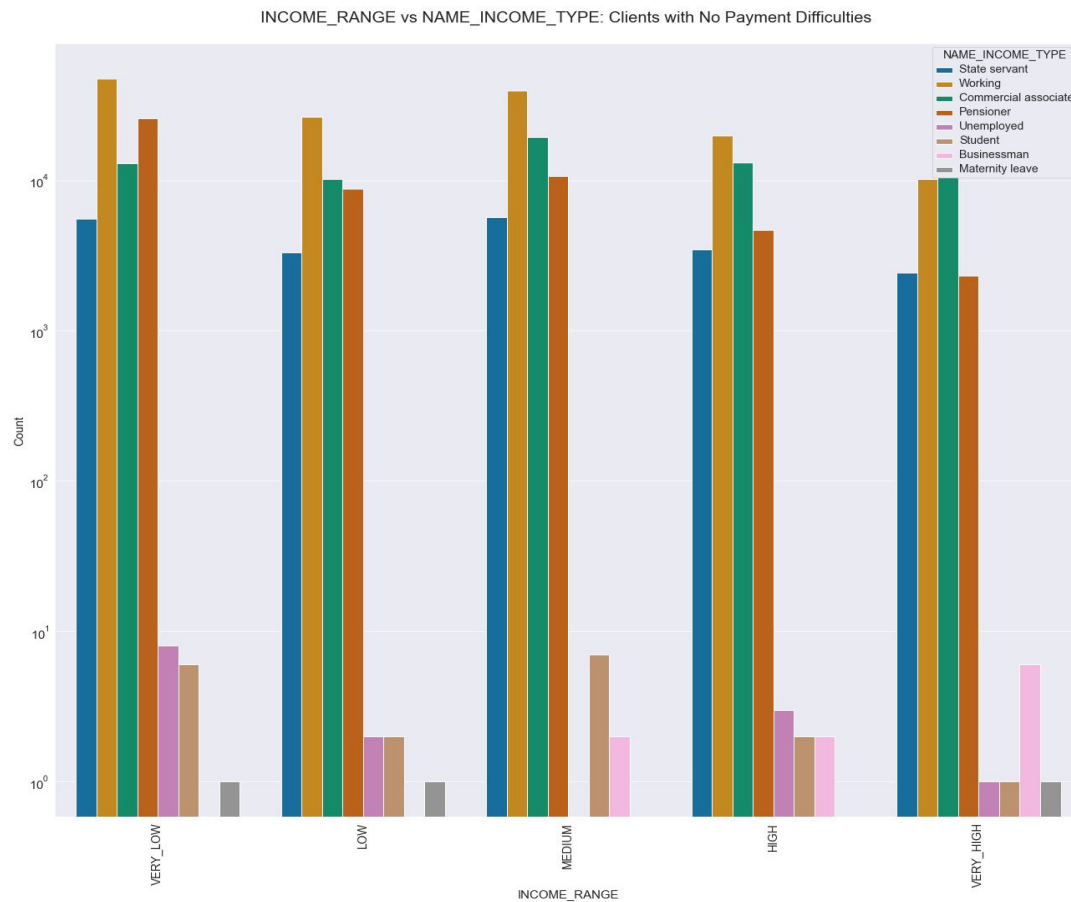
# INCOME VS GENDER



- Clients with No Payment Difficulties- The ratio of females to males in the very low income group is very high, while females are more in number across all income groups (High count of female clients in data) the ratio in every other group is relatively closer.
- Clients with Payment Difficulties- The number of defaulters who are males are higher than the females in the high and very high income groups

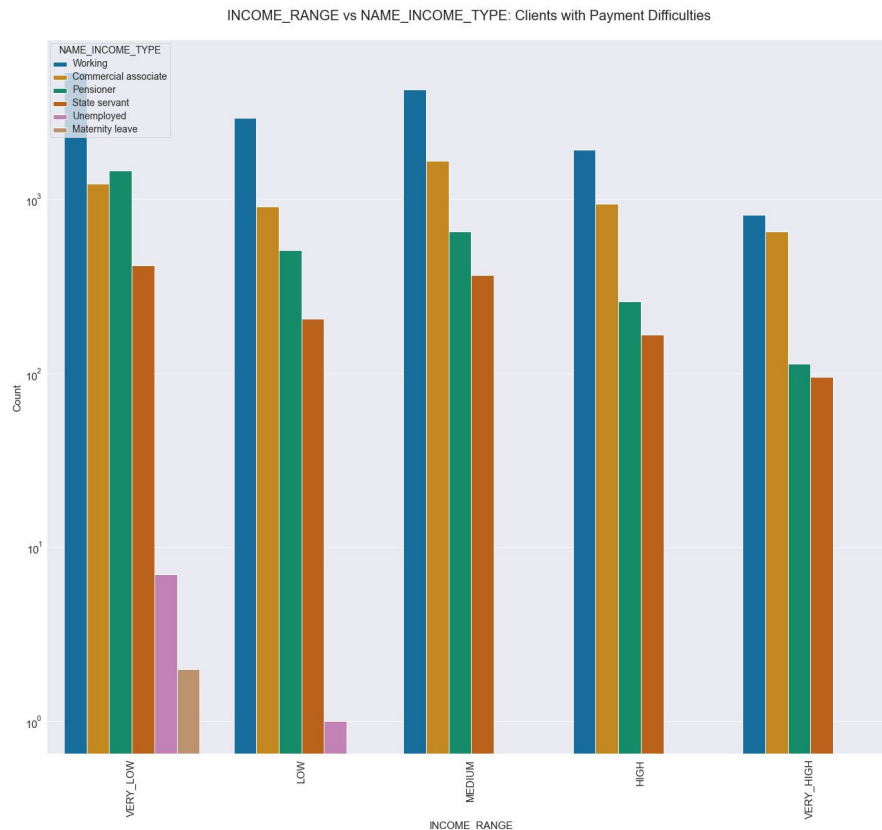


# DISTRIBUTION OF CLIENT BY INCOME AND INCOME RANGE



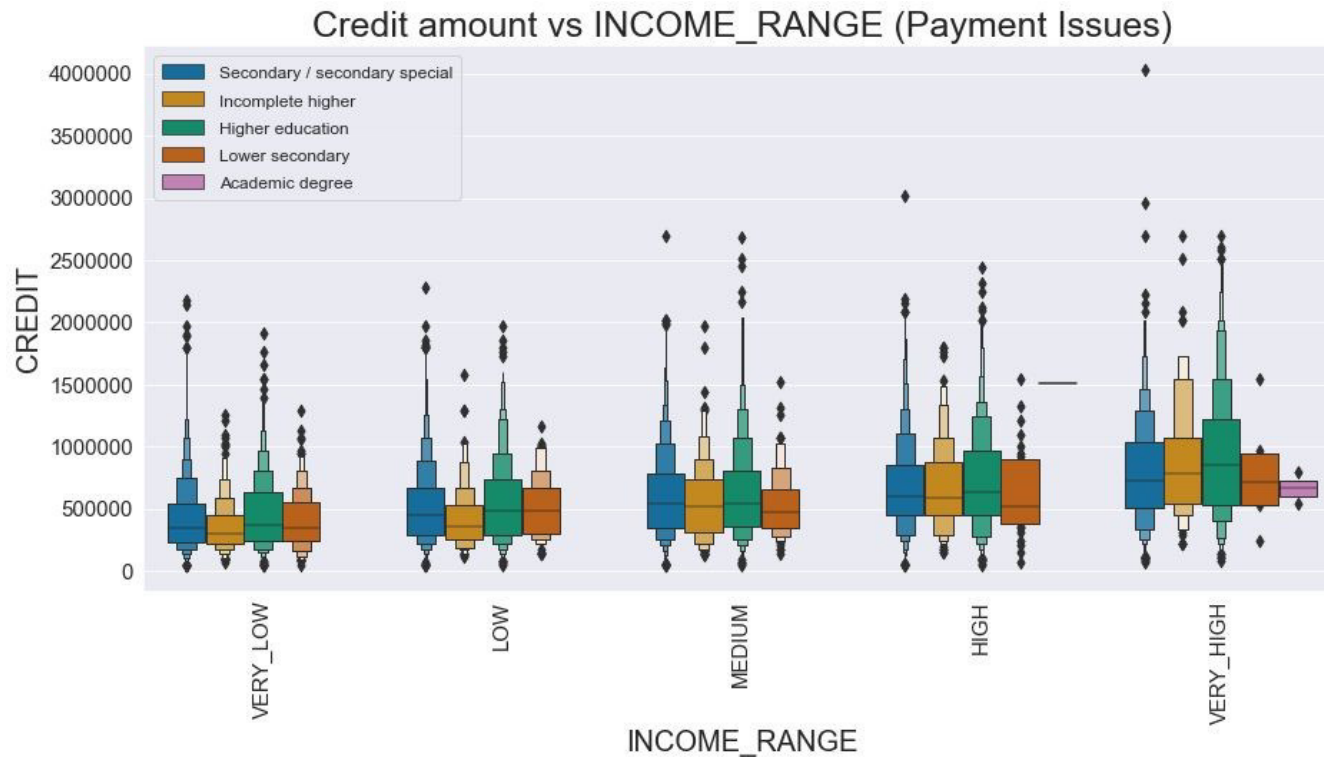
- Clients with No Payment Difficulties, the working and commercial associates, have an equal count in the very high income range. The company should focus more on these clients for business.
- While in the case of very low income range, the working income type has the highest count, followed by pensioners and then the commercial associates.
- The income\_range count of state servant is very close in the category of very low and medium, followed by high and low. The company can approach these customers due to their stable income, but only after thorough investigation.

## DISTRIBUTION OF CLIENT BY INCOME AND INCOME RANGE



- Clients with Payment Difficulties, the working count is slightly more than commercial associates count in the very high income range.
- While in the case of very low income range, the working income type has the highest count, followed by pensioners and then the commercial associates.
- The unemployed are seen only for very low and low income range.

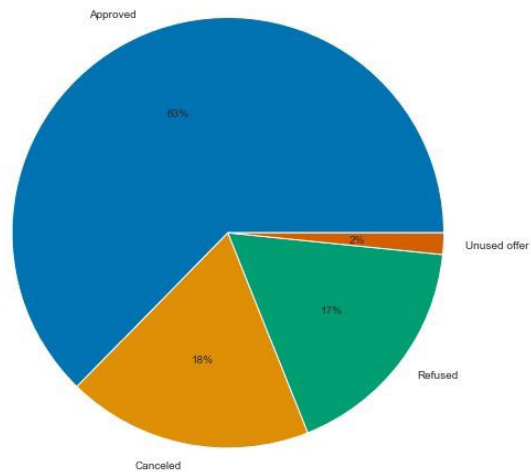
# CREDIT AMOUNT VS INCOME RANGE



- For clients facing payment problems, it is usually with people with higher education or with secondary education who have taken the biggest credits.
- Clients with academic degrees are far less likely to default across all income groups.

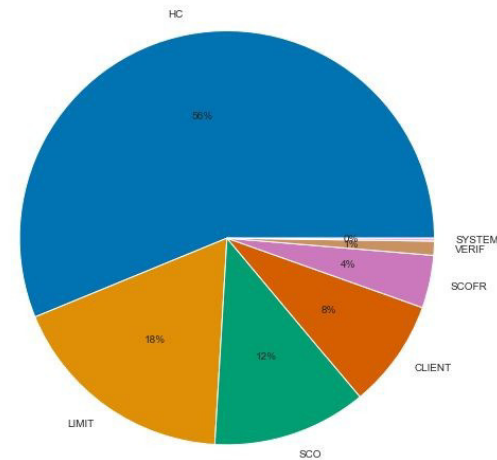
# PREVIOUS CONTRACTS

Pie Chart for NAME\_CONTRACT\_STATUS



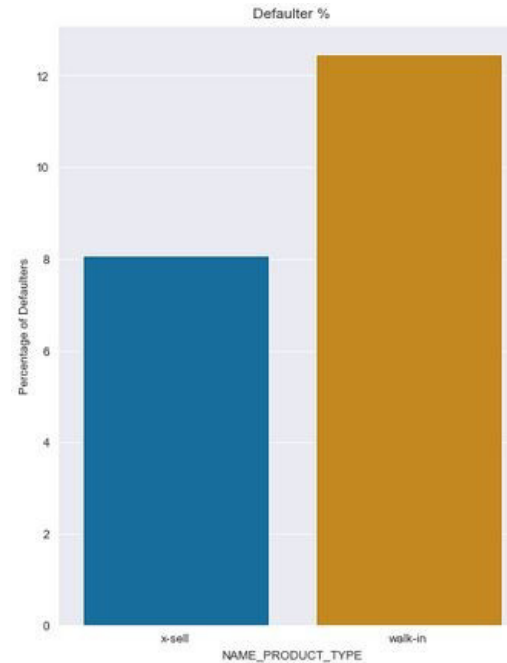
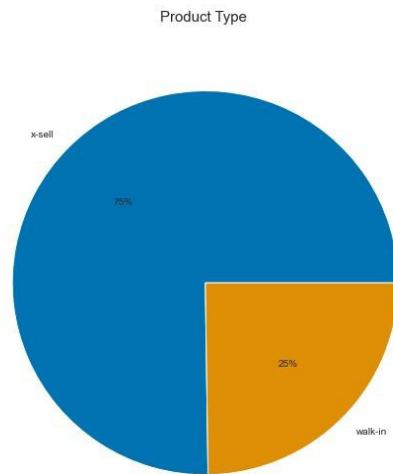
- 63% of the previous applications were approved.
- Only 17% of the applications were refused.

Rejection Reason



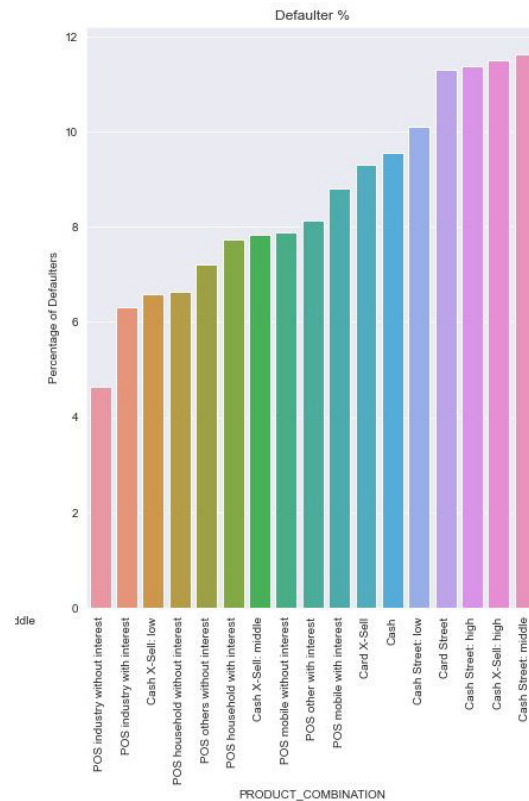
- 56% of the rejections were due to HC
- LIMIT and SCO together accounted for 30% of rejections

# PRODUCT TYPE AND COMBINATIONS



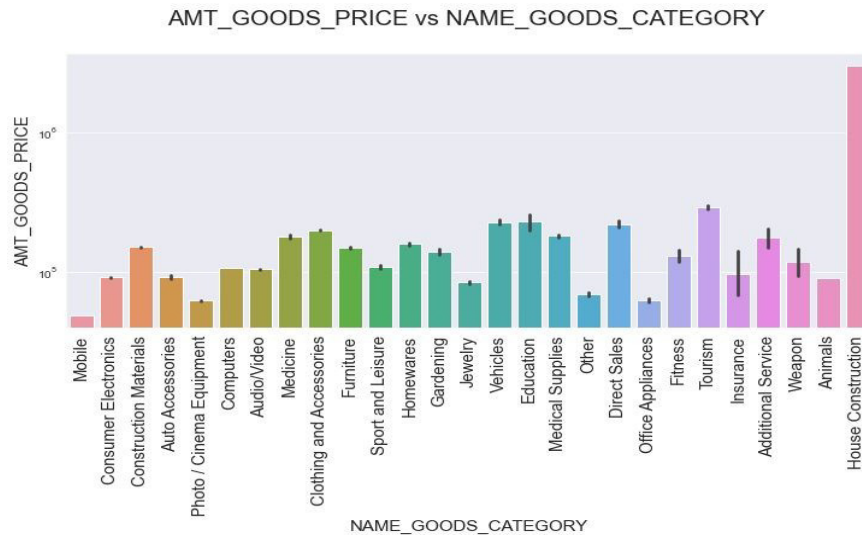
- Cross-Sell (x-sell) accounted for the majority of the previous applications at 75%. While rest were walk-ins.
- Despite accounting for the 75% of Previous applications, x-Sell has a significantly lower default rate than walk-in clients.

# PRODUCT TYPE AND COMBINATIONS

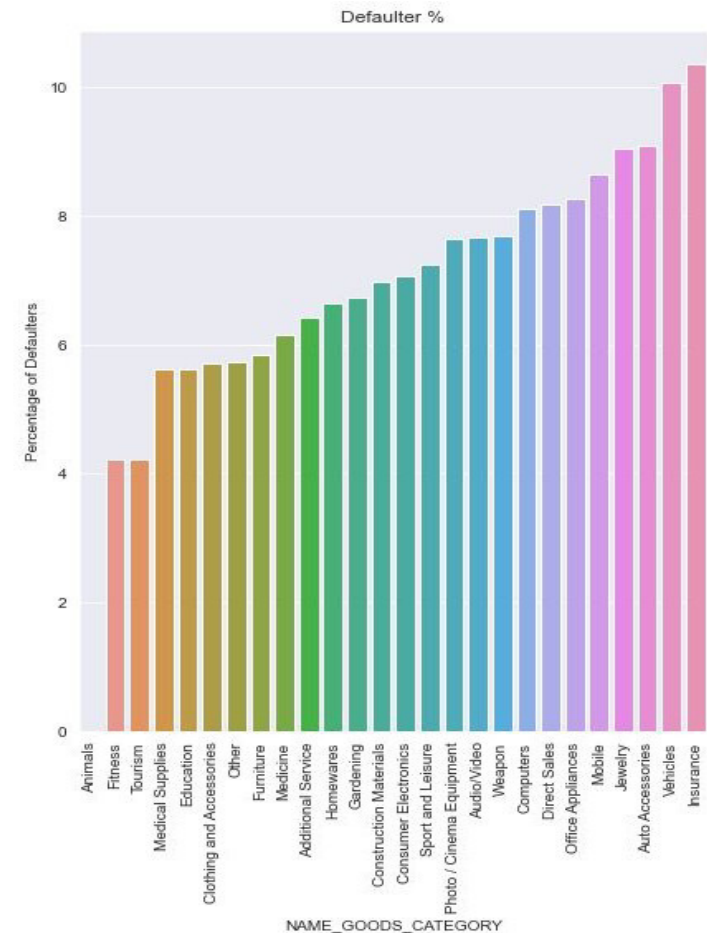


- In case of product combinations Card Street, Cash Street(high & mid), Cash x-sell are all high risk products with high default %.

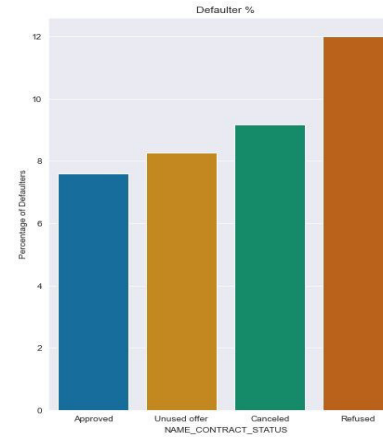
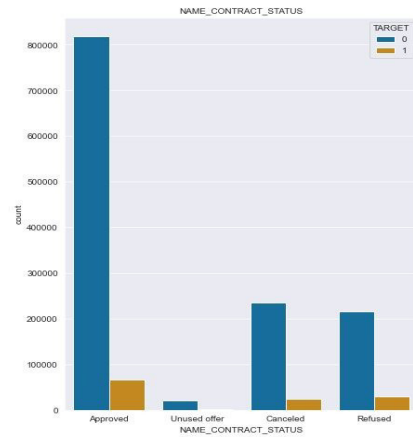
# GOODS PRICE VS CATEGORY



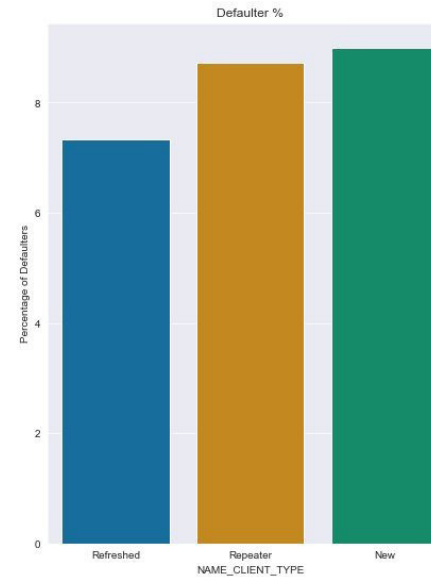
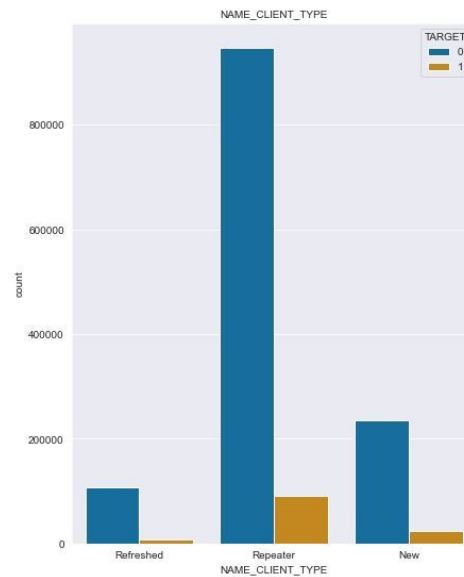
- House construction has been the most expensive applications in the past. Followed by Tourism, Education and Vehicles.
- However, the most defaults are found in categories of Insurance and Vehicles



# CLIENT'S HISTORY



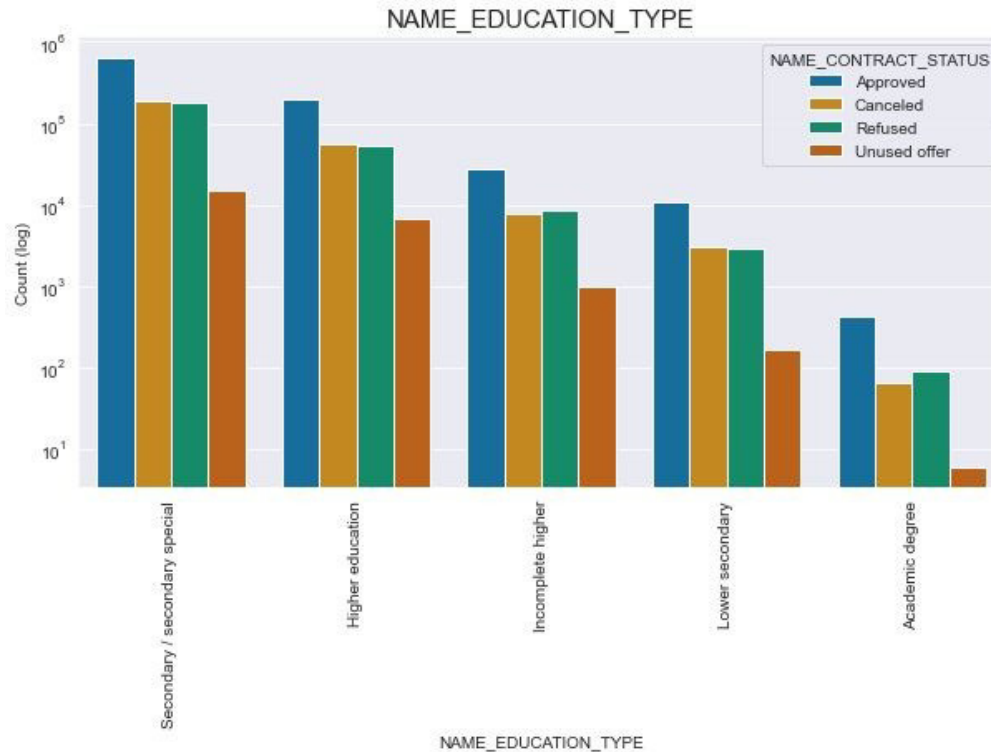
- Clients who were previously refused, have a much higher rate of default.



- The chance of newer clients to default is slightly more than the repeated or refreshed clients.



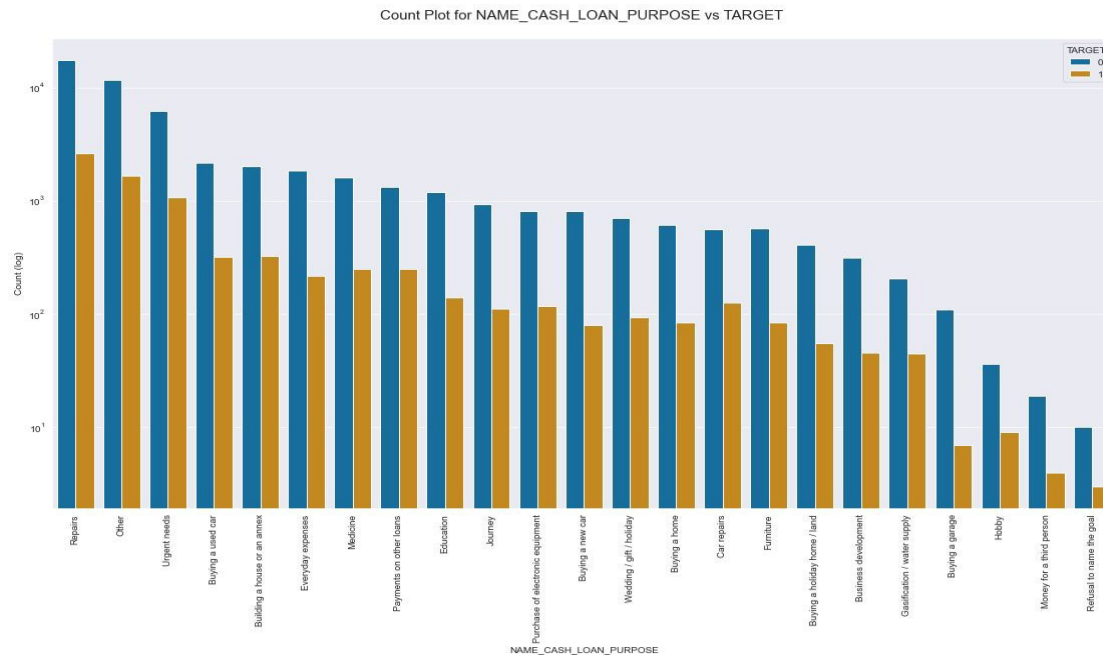
# CLIENT'S WITH ACEDAMICS



Academic degree	Approved	418	72.19%
	Refused	91	15.72%
	Canceled	64	11.05%
	Unused offer	6	1.04%

- Clients with Academic Degrees have the highest approval rate and lowest refusal rate among all the education groups

# LOAN PURPOSE AND DEFAULT



- Loans that were taken for Repairs and Other have the maximum number of defaults.

# OBSERVATIONS

- People who own realty are more likely to default
- Men are more likely to default than females
- Unemployed and those on Maternity Leave are the worst offenders
- Credits given for Insurance and Vehicles have very high default
- People who were employed for less than 5 years at their current job have defaulted the most.
- Clients with better education background are far less likely to default
- Clients who were previously refused, have a much higher rate of default.

# OBSERVATIONS

- Refreshed and repeated clients are safe and have low risk of default
- Loans that were taken for Repairs or Other were most rejected by both company and client. These loans also had the highest number of defaults
- Clients with Academic Degrees have the highest approval rate and lowest refusal rate among all the education group

# RECOMMENDATIONS

- Payment difficulties are more prevalent mainly in the younger demographic of 25-35 age - Checks on past records must be done.
- Cash loans are popular and are safer for the company as revolving loans have considerably high % of defaulters and can become troublesome in the long run.
- Clients who already own a realty may have problems in repaying, which might be due to existing credits. Background check on other loans must be done on such clients to prevent issues in repayment.
- Loan that were taken for Repairs or Others have high default rate, and needs to be assessed.
- Employment status checks for work history with the current employer must be done, along with the client's current work status. As those recruited but are on a leave (eg. Maternity Leaves) are highly likely to default.
- Academic Status of client must be assessed and considered before approval.
- Clients who were previously refused and reapplied, checks should be made not only for their current info/status but also the reason of their previous rejection.
- While gender shouldn't be a determinant factor on who gets a loan, Men have shown tendency to have higher defaults and a stricter approval might reduce those cases.

THANK YOU