

# **CREDIT EDA ASSIGNMENT**

**By – RUDRARUP GOSWAMI**  
**Date – 28/11/2023**

## **BUSINESS OBJECTIVE**

Understand the driving factors /driver variable behind loan defaults  
Identifying client loan payment patterns

## **DATA UNDERSTANDING**

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

Previous\_Application.csv - Contains information about the client's previous loan data.  
It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.

Columns\_Description.csv - Data dictionary which describes the meaning of the variables.

# PROCESS

## DATA CLEANING

- Columns with 40% null values dropped (Eg – ‘Rate Down Interest Payment’, ‘Amt Down Payment’)
- Dropped columns with no relevance (Eg – Common Area Mode, Living Area Avg)

## DATA STANDARDIZATION

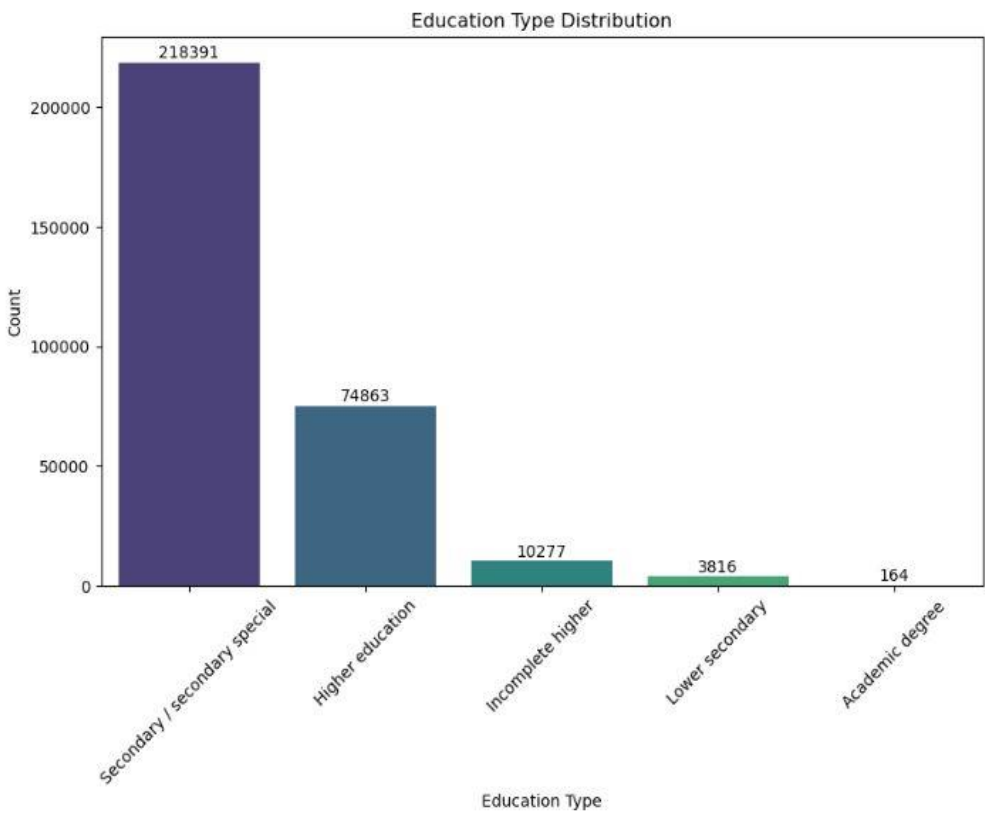
- Converted (negative) days column to positive (Eg – ‘Days Birth’, ‘Days Employed’,)
- Replaced XNA values with NaN – (Eg – ‘Name Product Type’, ‘Code Gender’, ‘Occupation Type’)
- Converted object type column to categorical type column

## DATA IMPUTATION

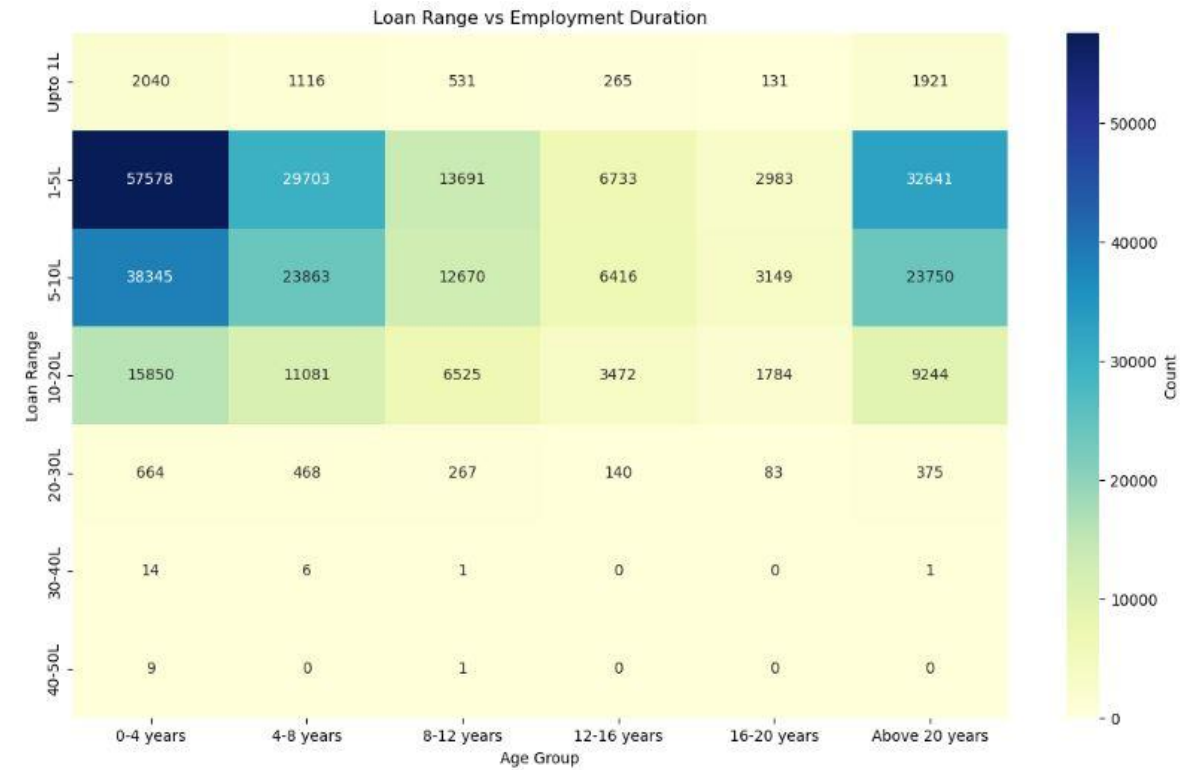
- Filled missing values in numerical type column with median. (Eg – ‘Amt Req Credit Bureau Hour’,)
- Filled missing values in categorical type column with mode. (Eg – Name Type Suite)
- AMT\_APPLICATION, AMT\_CREDIT, AMT\_GOODS\_PRICE columns have certain outliers which are normal as there can be high loan amount given out
- Created bins on numerical columns for better analysis – (Eg – Amt Credit, Amt Income Total, Decision Days)

# DATA ANALYSIS

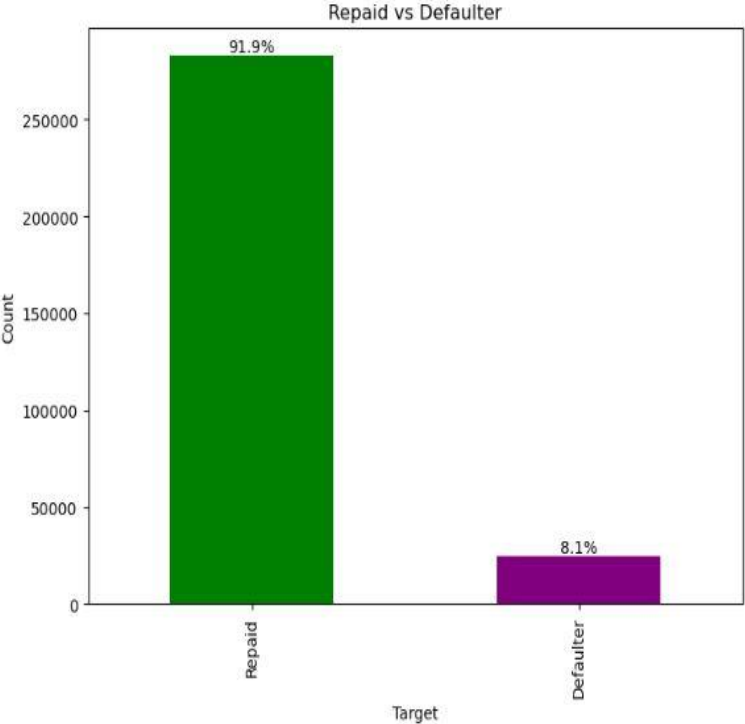
UNIVARIATE ANALYSIS – Count Plot and Bar Graph plotted



BIVARIATE ANALYSIS – Heat map, Stacked column

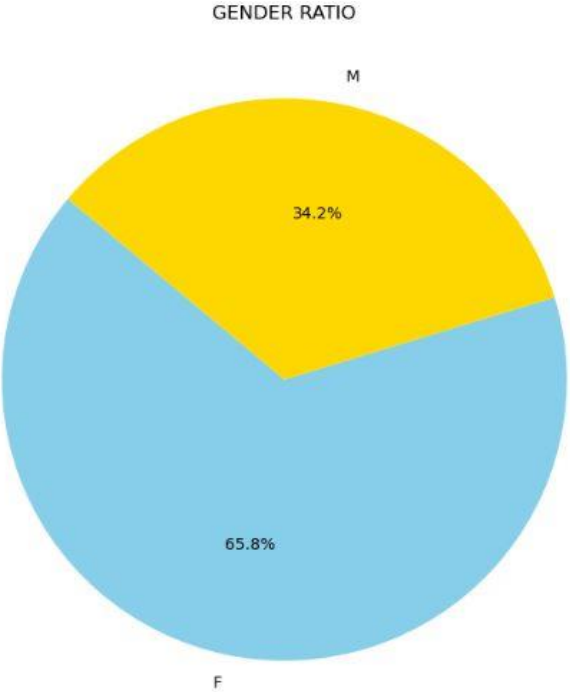


# Default rate



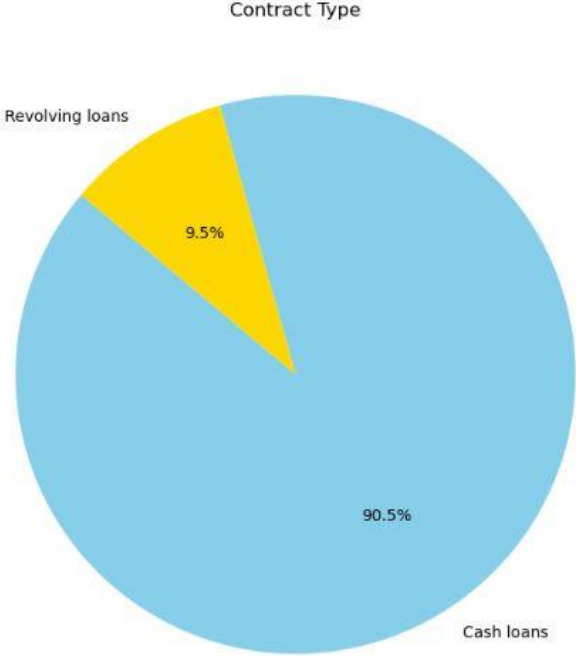
92% of the loan has been repaid  
8% of the loans are defaulter

# Gender Ratio



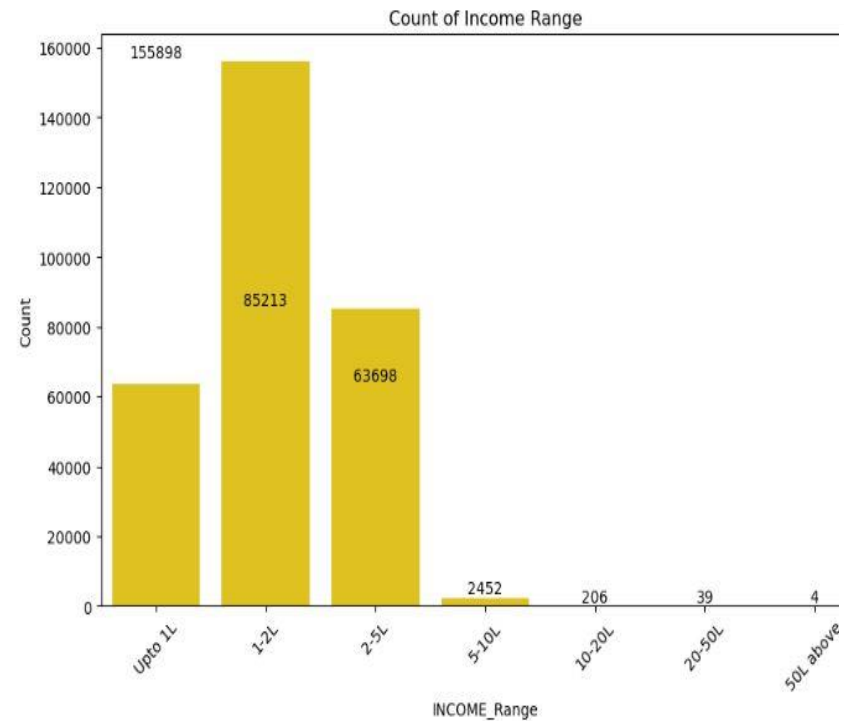
66% of the loan applicants are Female  
34% of the loan applicants are Male

# Contract Type

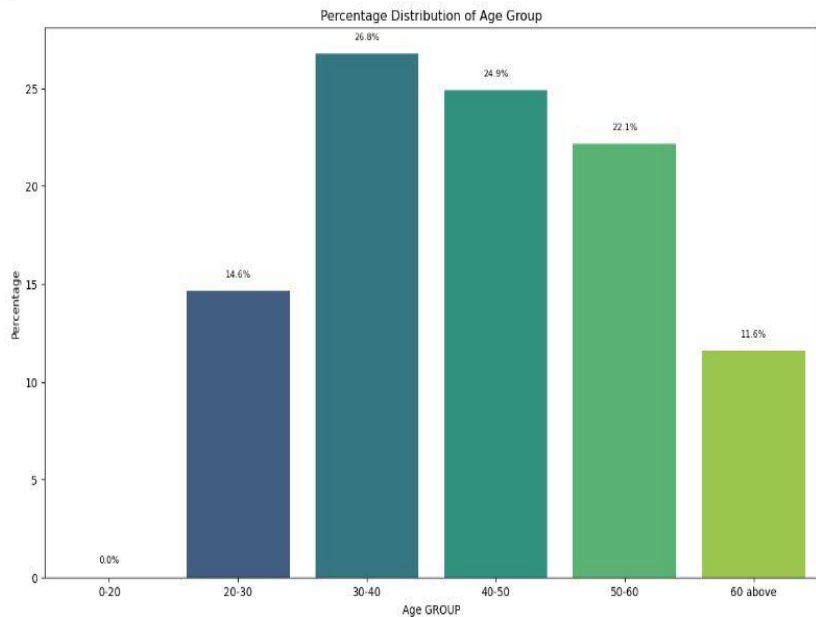


91% of the loan applications are Cash Loans  
9% of the loan applications are Revolving Loans

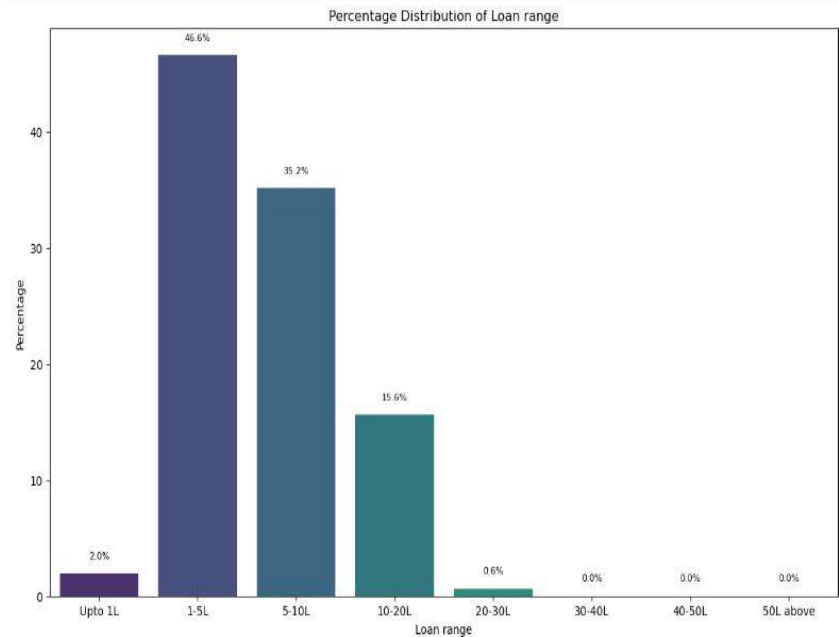
# BIN DISTRIBUTION CHARTS



Most of the loan applicants have an income range of 1-2 lakh.  
The second most is 2-5 lakh

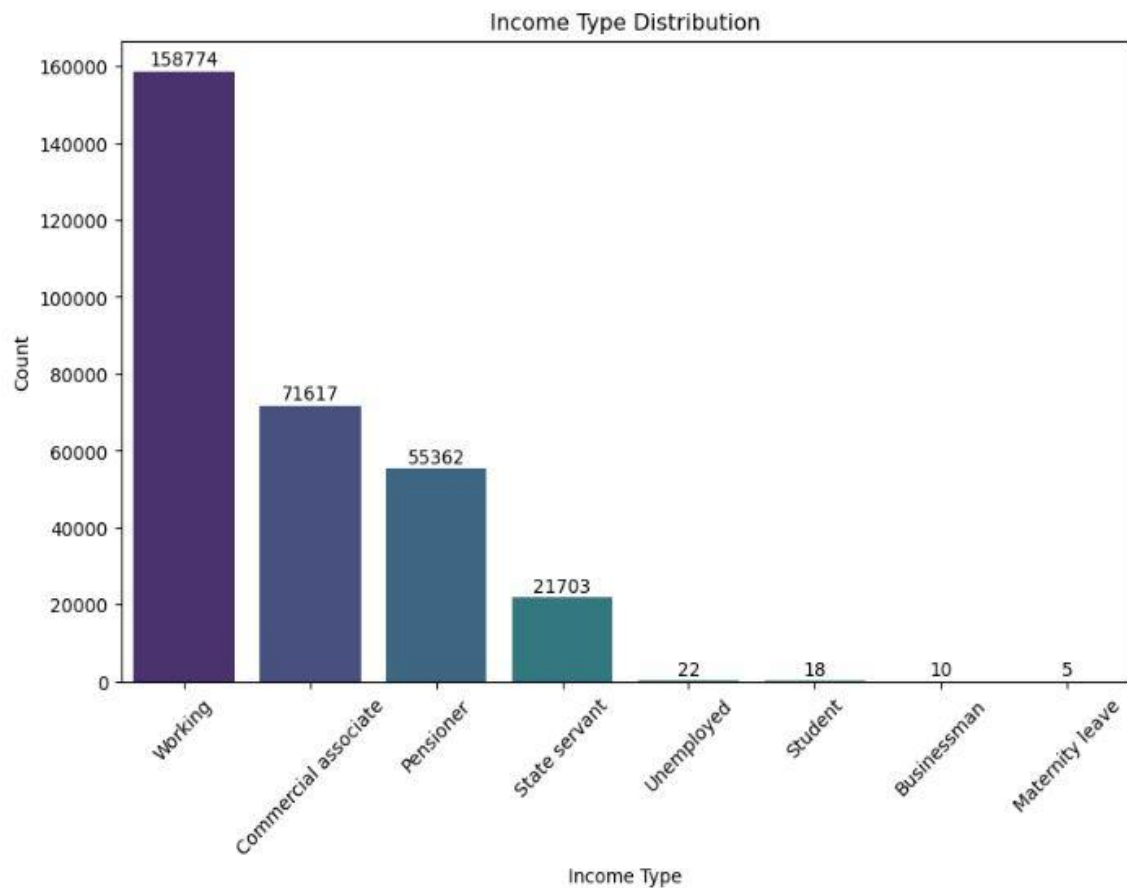


27% of the loan applicant is from the age group 30-40 years  
12% of the loan applicants are senior citizen (60+)

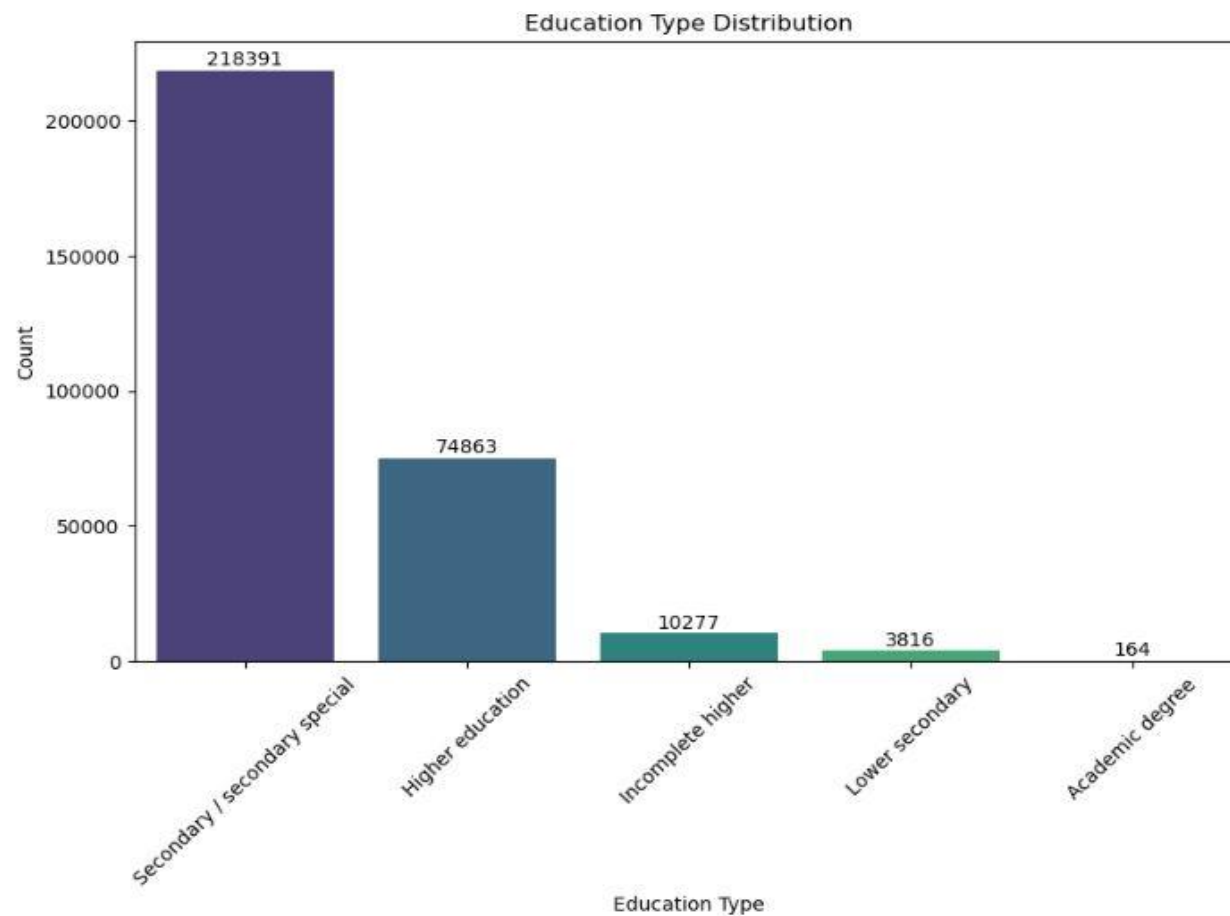


47% of the loan amount is between 1-5 lakhs.  
There are very few high end loans above 40 lakhs

## CATEGORY CHARTS

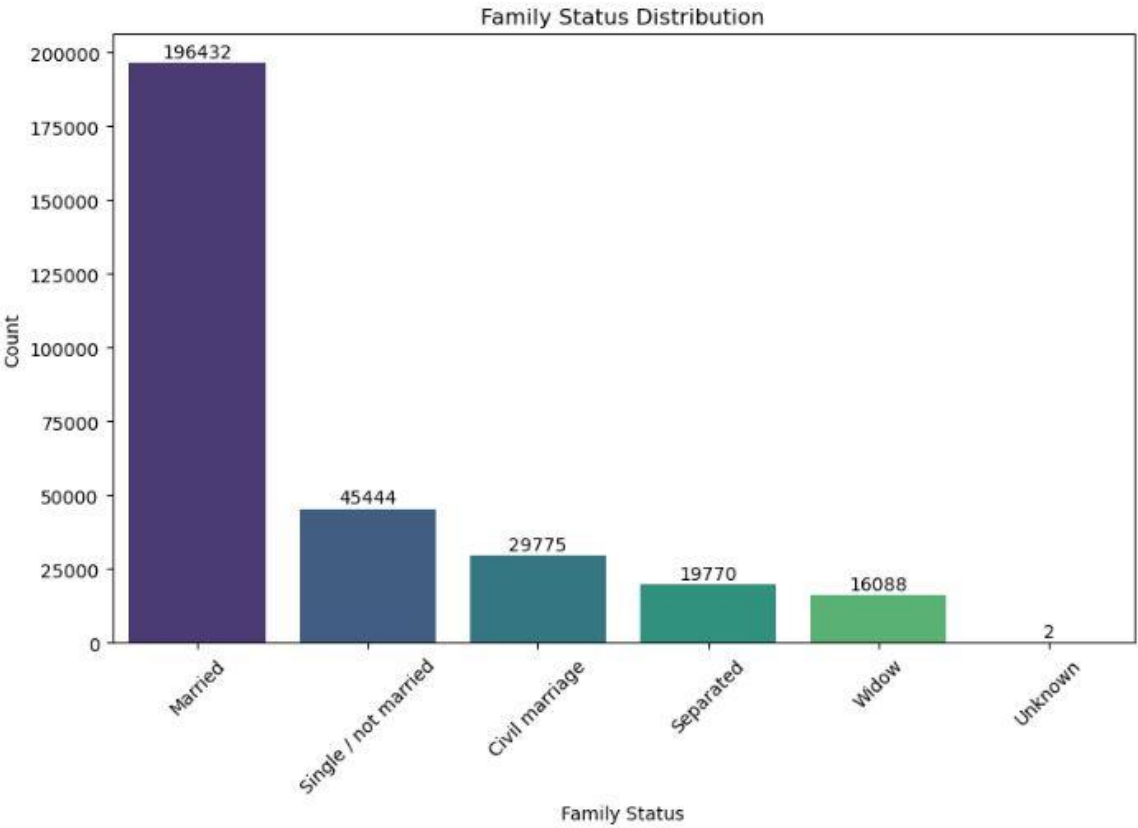


Working people has applied for the most number of loans  
Commercial Associate's are the second most applicants

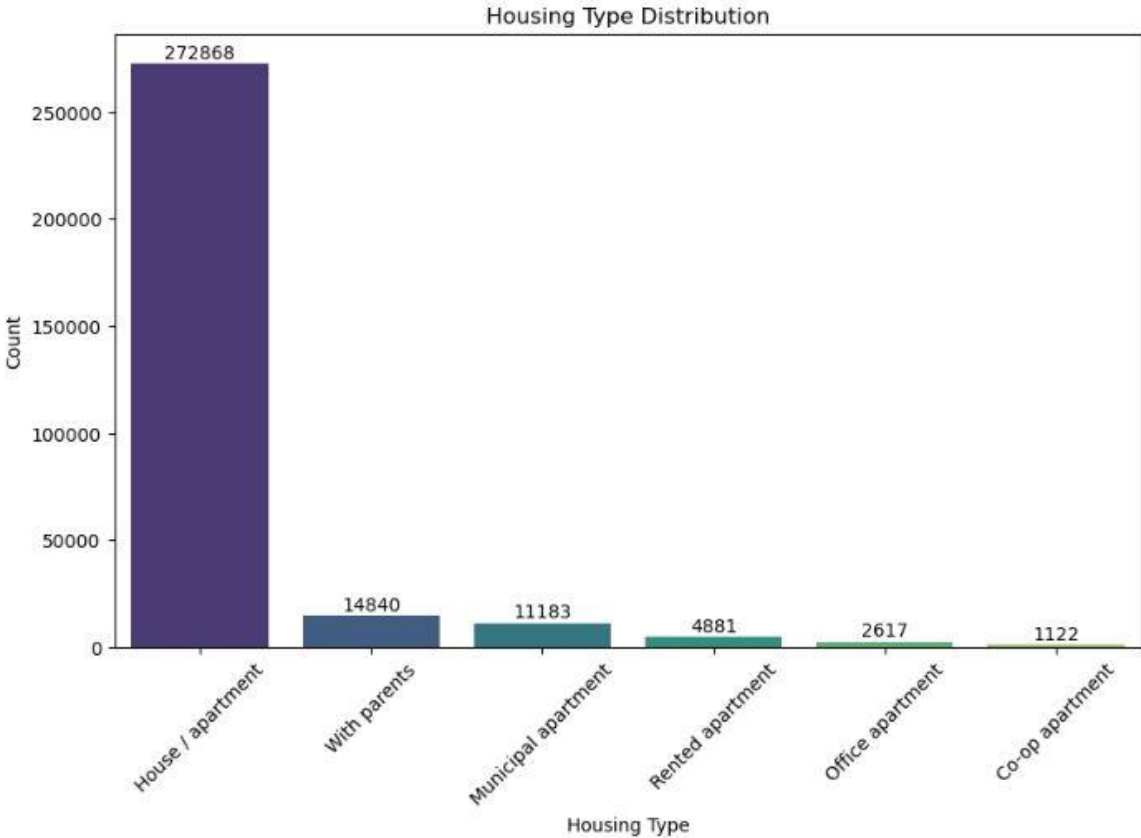


Secondary qualified client has applied for the most number of loans  
Higher education qualified are the second most applicants

# CATEGORY CHARTS



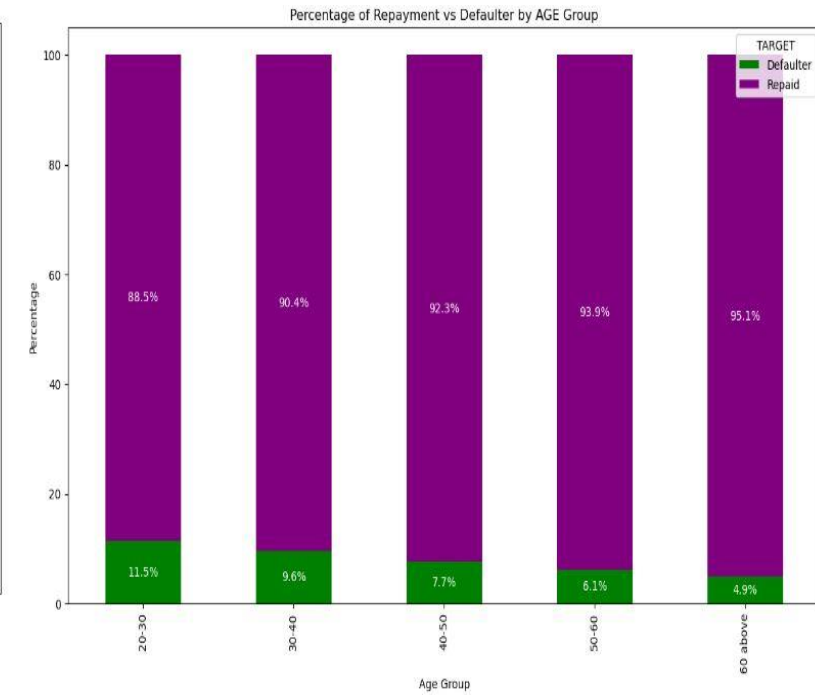
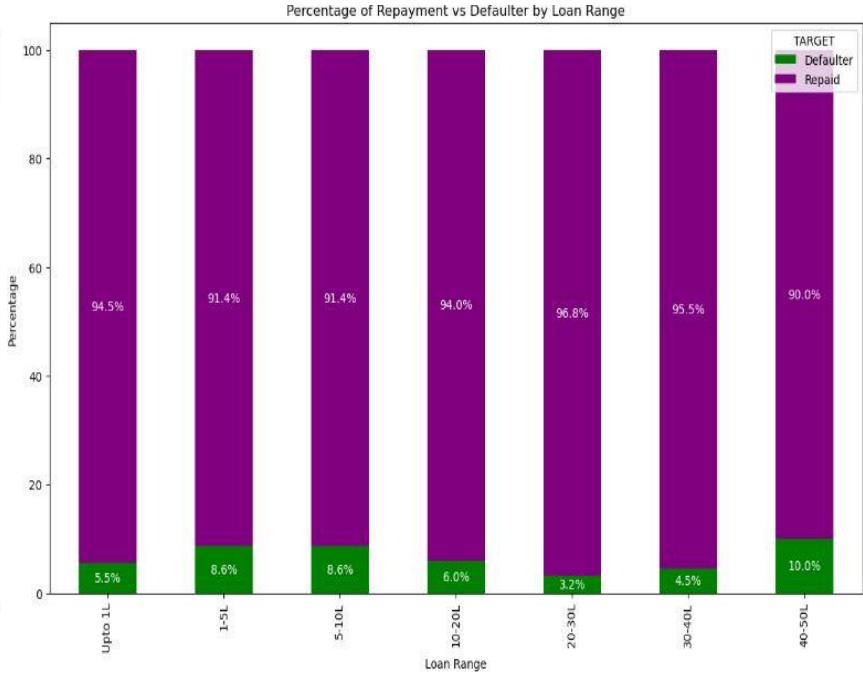
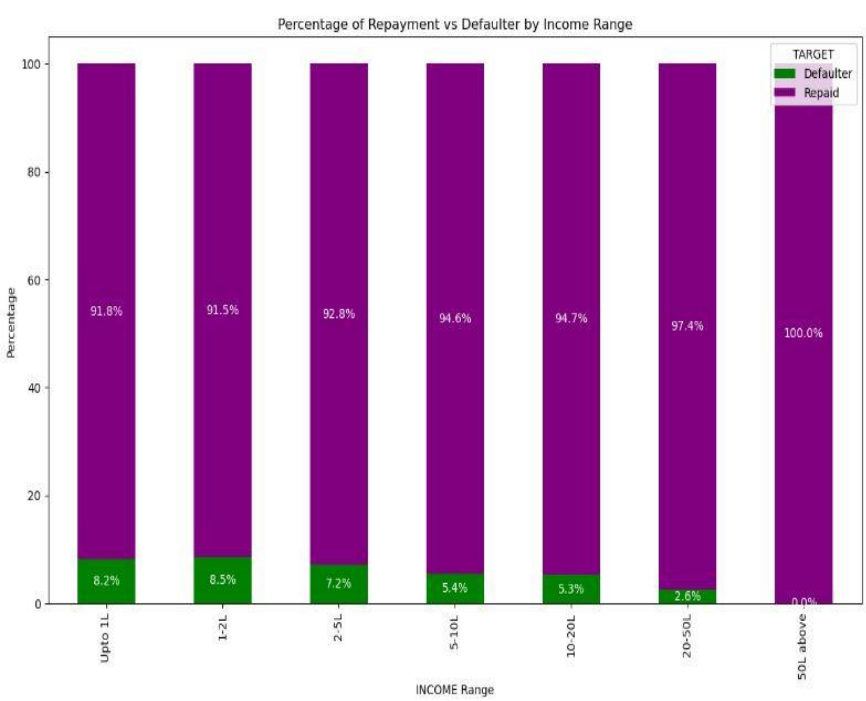
Married people have applied for the most number of loans



People living in House/Apartment have applied for the most number of loans



# STACKED COLUMN CHART

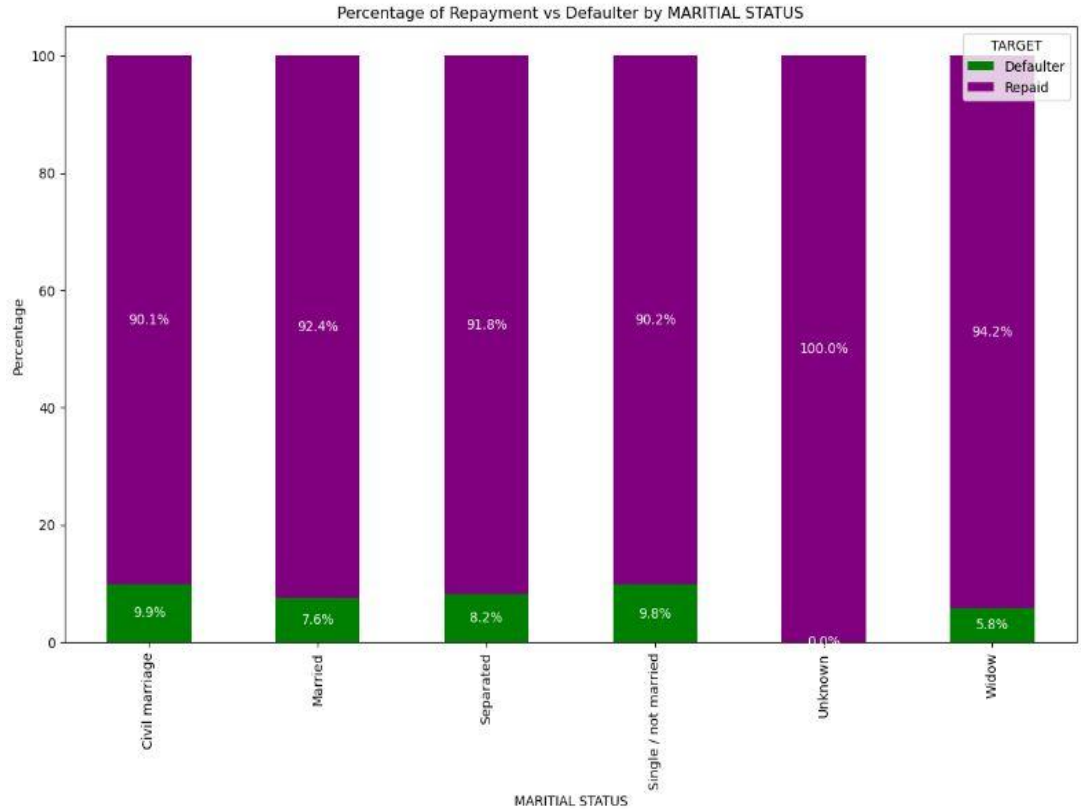


Clients with Income less than 2 lakh are more likely to default  
Clients with income of more than 50 lakh are more likely to repay the loan

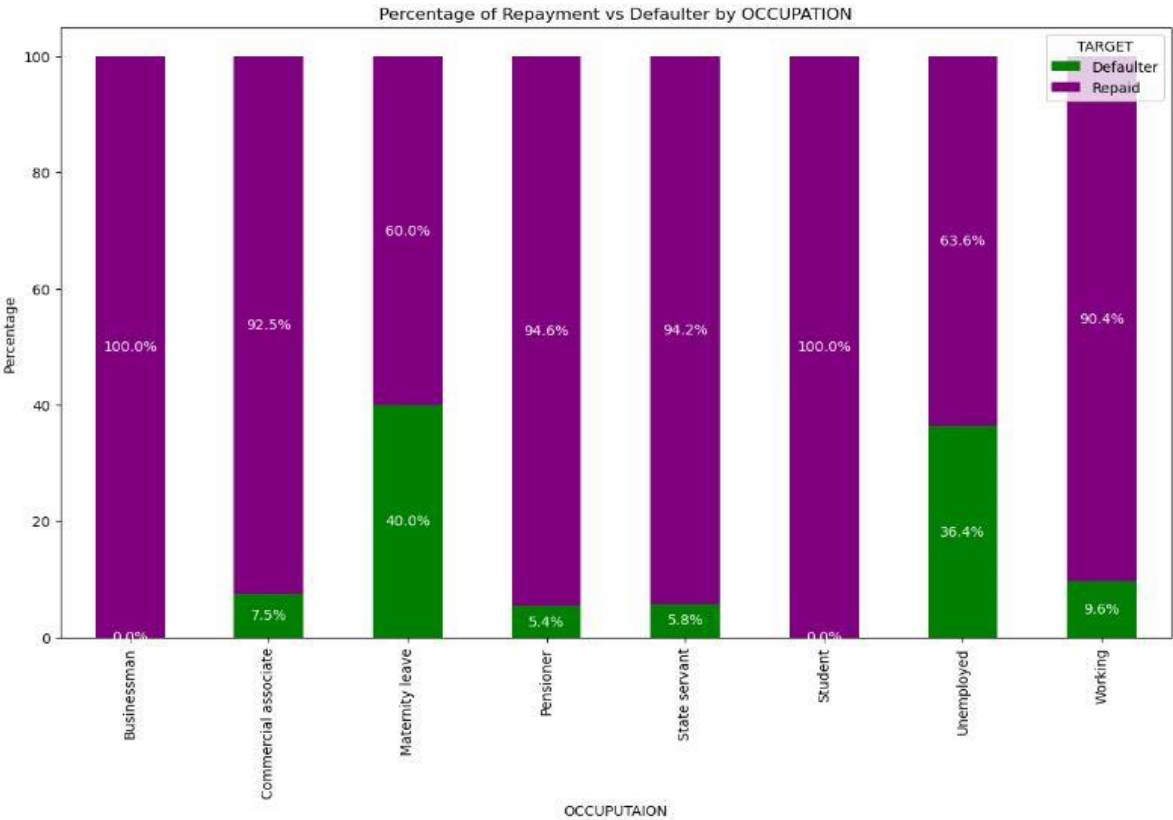
Loan amount less than 10 lakh have more default rate than high loan value

People aged below 40 have a 20% default rate.  
Senior citizen (60+) has the lowest default rate

# STACKED COLUMN CHART

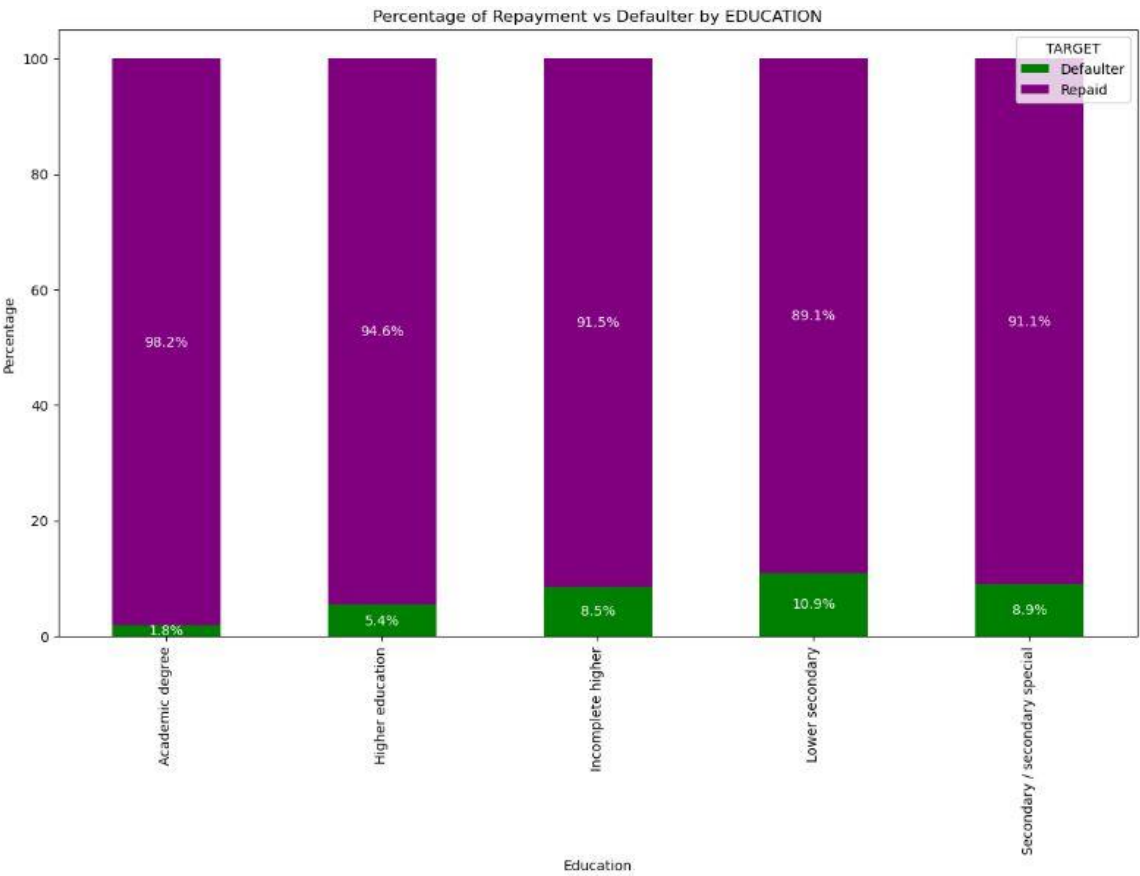


Separated and Single people have a 18% default rate  
Clients with civil marriage has a 10% default rate

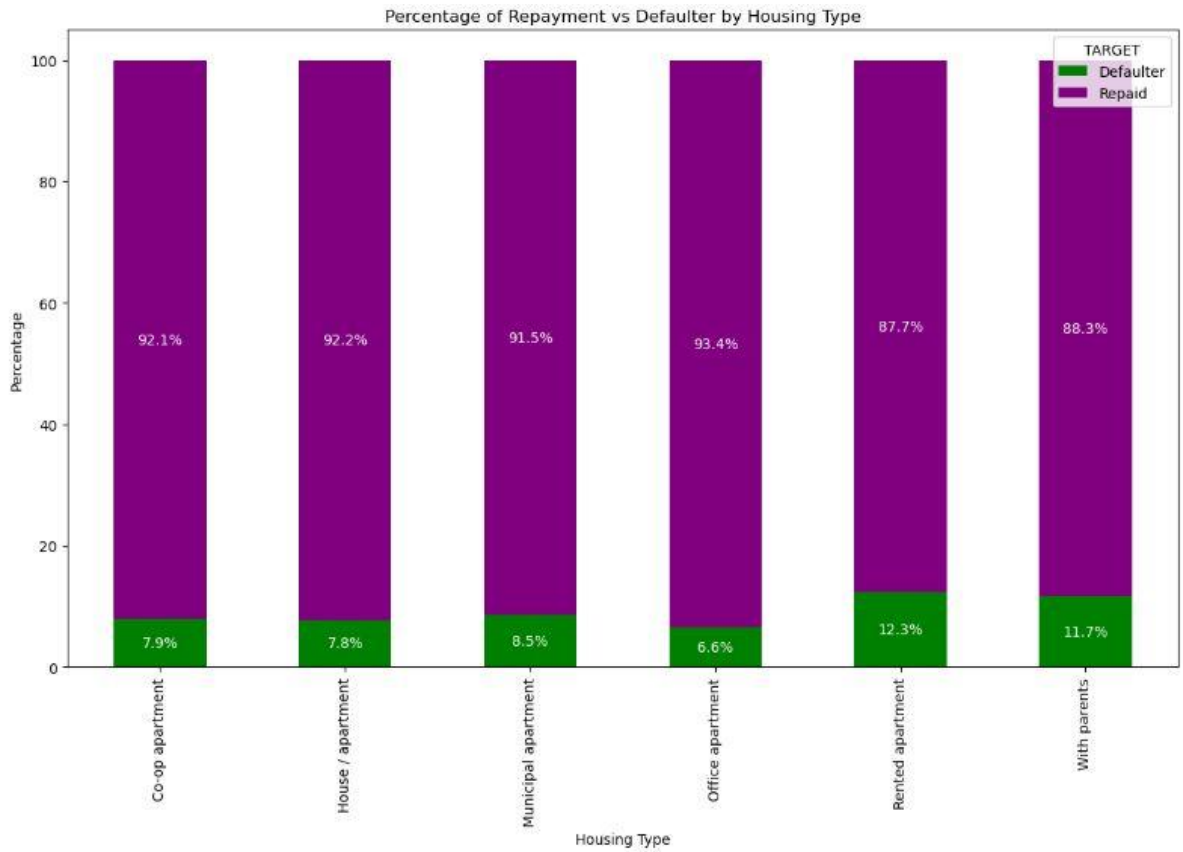


Clients on Maternity Leave and Unemployed people have a 76% default rate  
Businessmen have the best repayment rate

# STACKED COLUMN CHART

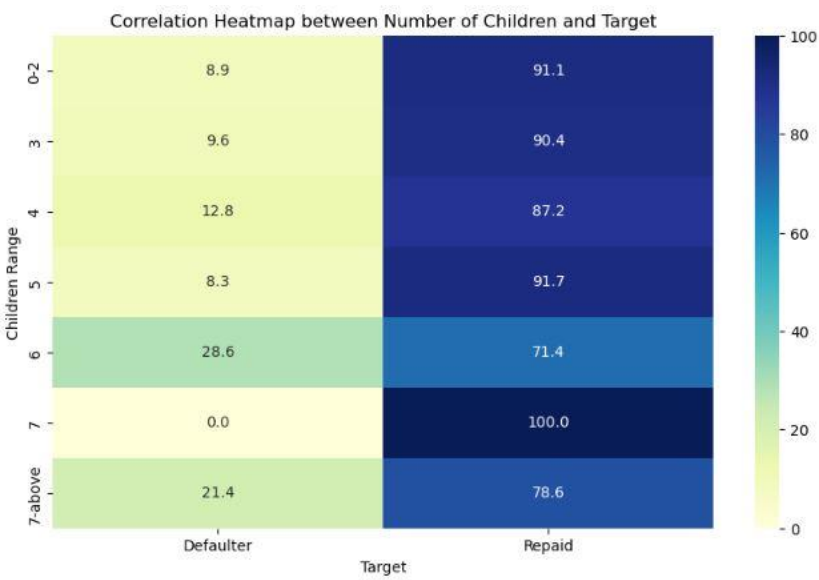


People with education qualification of secondary and less than secondary has a 20% default rate  
People with Academic range has the highest Repayment rate

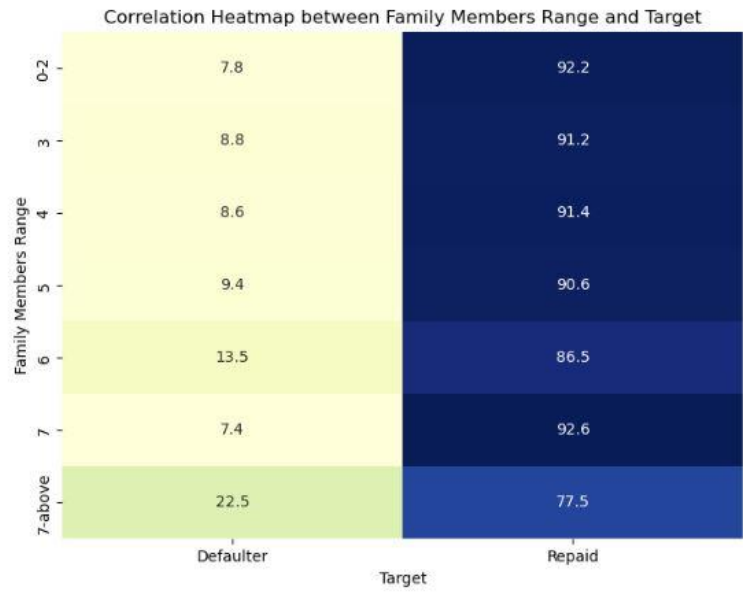


Clients living in Rented apartments have the highest default rate  
Clients living in Office apartments have the highest repayment rate

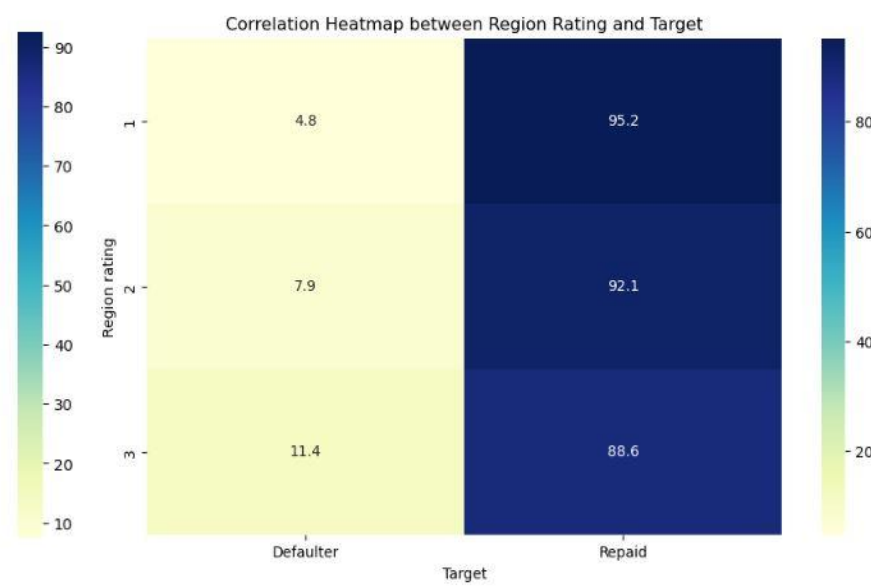
# HEAT MAPS



Clients with more children has a higher chance of defaulting

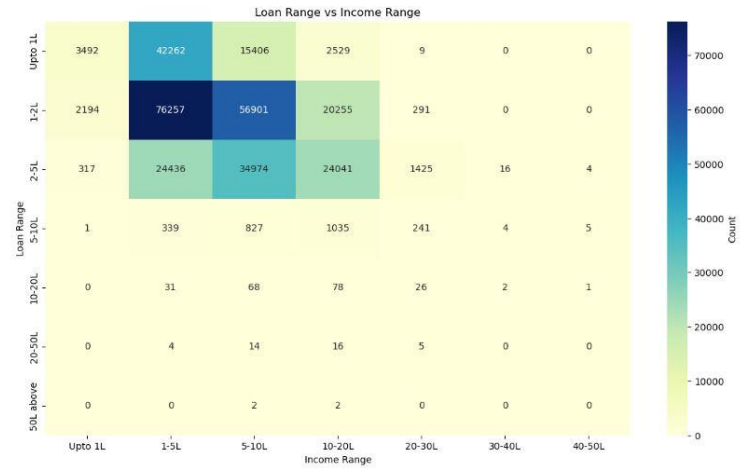


Clients with more family members has a higher chance of defaulting



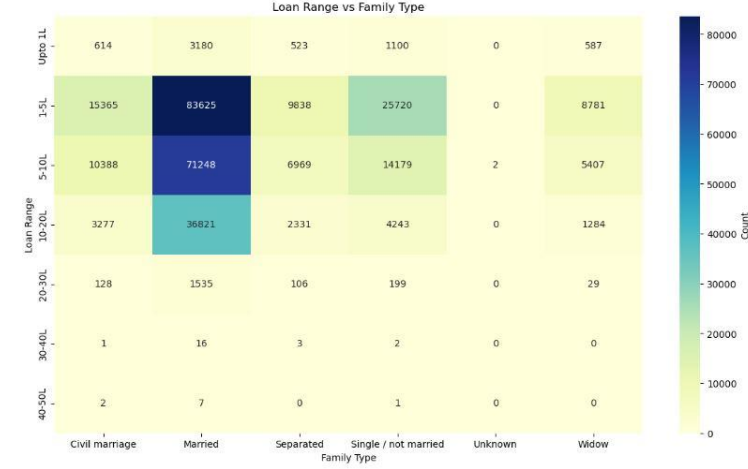
City rating 1 has a low default rate in compared to the other

# CATEGORIES vs LOAN RANGE HEATMAPS



Clients with Income below 5 lakhs takes credit below 5 lakh.

Client with income greater than 50 lakh takes credit between 20-50 lakh

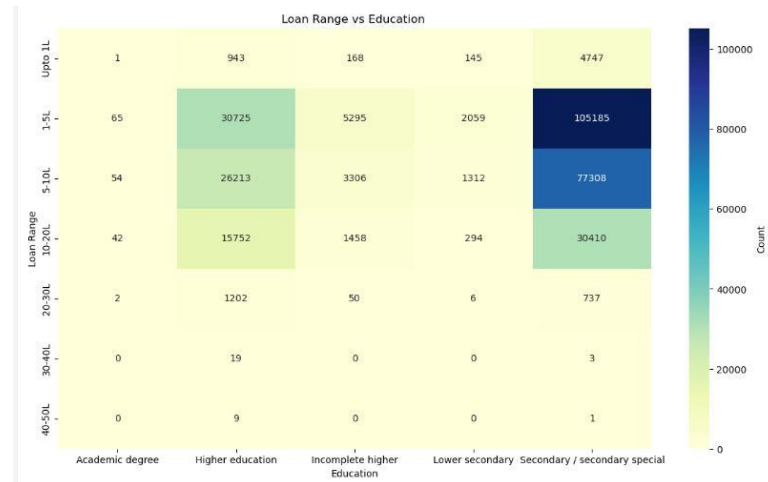


Married People take more credit, across all loan range



Working Income type people takes credit less than 10 lakh

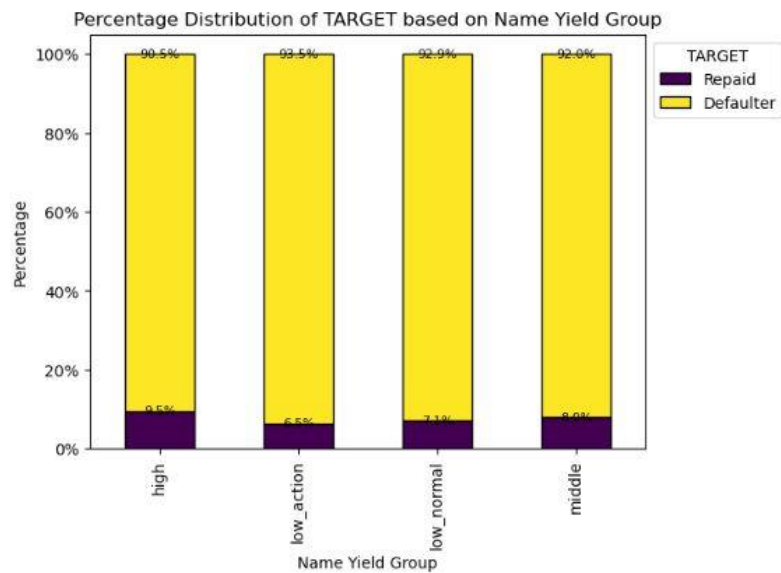
Commercial Associates takes high credit more than 40 lakh



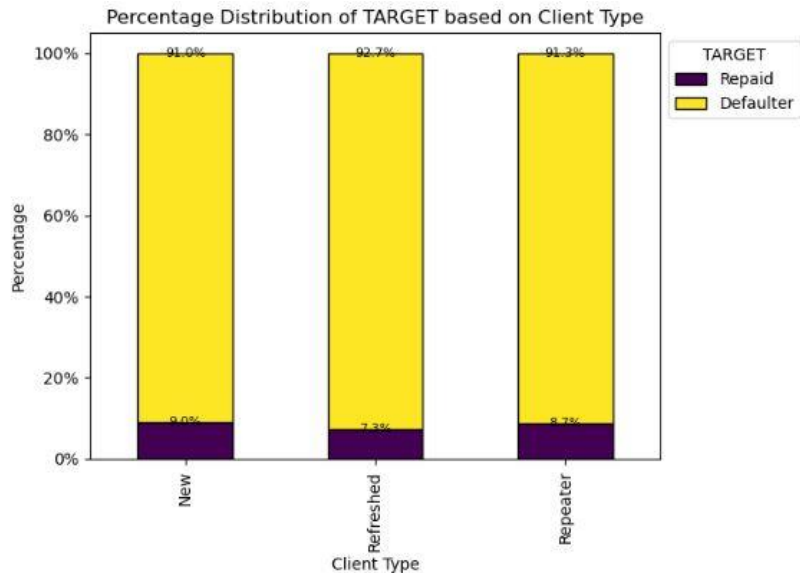
Lower Qualified people takes more credit

Higher Qualified are people who takes loan more than 40 Lakh

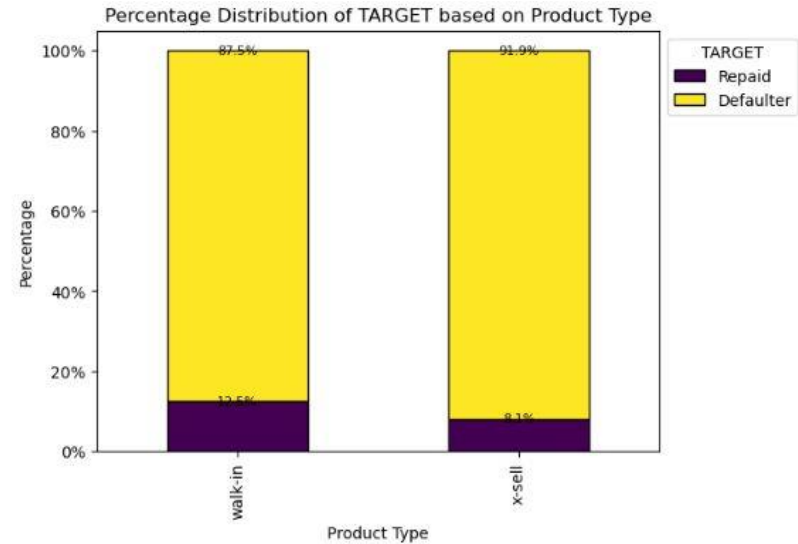
# PREVIOUS APPLICATION DEFAULT



High interest rate has high default rate



ALL Client type has same default rates



Walk in has less default rates than sell in

# INFERENCE

- 92% of the loans are repaid. Only 8% are defaulters
- 66% of loan applicants are Female and 34% loan applicants are Male
- Clients with more family members and higher children count has a higher default rate
- Clients with low education qualification has a higher default rate than clients with higher qualification
- Clients people has a lower default rate in comparison to single and separated clients
- Clients living in rented apartments have a higher default rate
- Unemployed and Maternity clients have a higher default rate
- Clients with income less than 5 lakh rupees have a higher default rate
- Clients aged 40 years and less have a higher default rate
- High loan amount has low default rate in comparison to low loan amount
- High yielding loans have a higher default rate
- Walk in loans have a lower default rate

THANK YOU