



# CREDIT EDA CASE STUDY

AJAY SARVAJEETH

# PROBLEM STATEMENT

- Consider a company receives an application for loan, based on the applicant's profile or credit history bank decides on the loan approval
- 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 to their advantage by becoming a defaulter.
- Two types of risks are 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.

# GIVEN DATA

- The data given contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:
  1. The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
  2. All other cases: All other cases when the payment is paid on time.
- When a client applies for a loan, there are four types of decisions that could be taken by the client/company):
  1. Approved
  2. Cancelled
  3. Refused
  4. Unused offer.

# APPLIED APPROACHES

## Data Scrubbing:

- In the 'application\_data' columns that had more than 40% null values were deleted and in the 'previous\_application' columns that had more than 30% null values were deleted
- Columns that were irrelevant were identified and deleted to simplify the data.

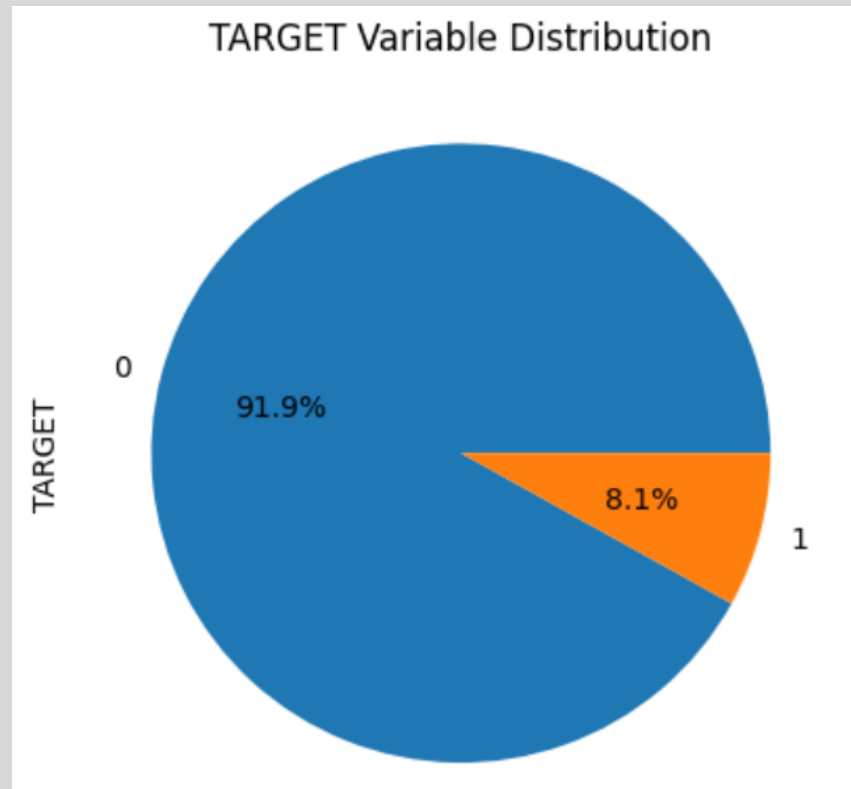
## Missing Values Treatment:

- Numerical columns were imputed with median wherever the outliers were detected.
- Category columns were imputed with mode (most frequent values).

## Outliers and Variation:

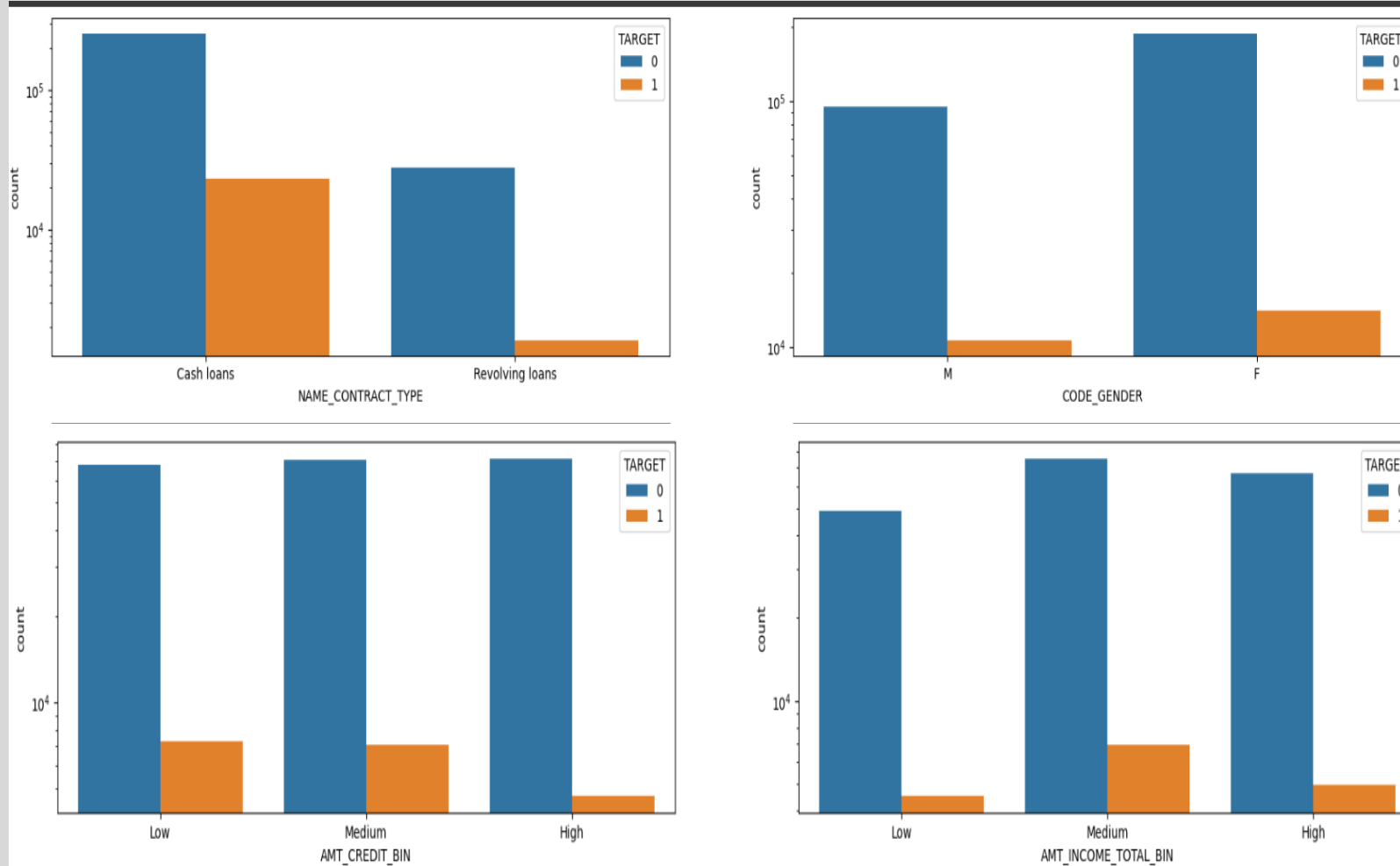
- Outliers were detected in most of the numerical columns, that could be due to the error in the data entry
- As there was Outliers and Variations used binning based on the quantiles of the data so that analysis could be done easily
- Negative values were converted to positive values

# DATA IMBALANCE



- Target column is not balanced, as it shows only 8.1% of the clients are defaulters and 91.9% are non-defaulters.
- Data imbalance ratio is 1 : 11.38.
- We can divide this data into 2 parts and analyze further for more insights as there is imbalance of data.

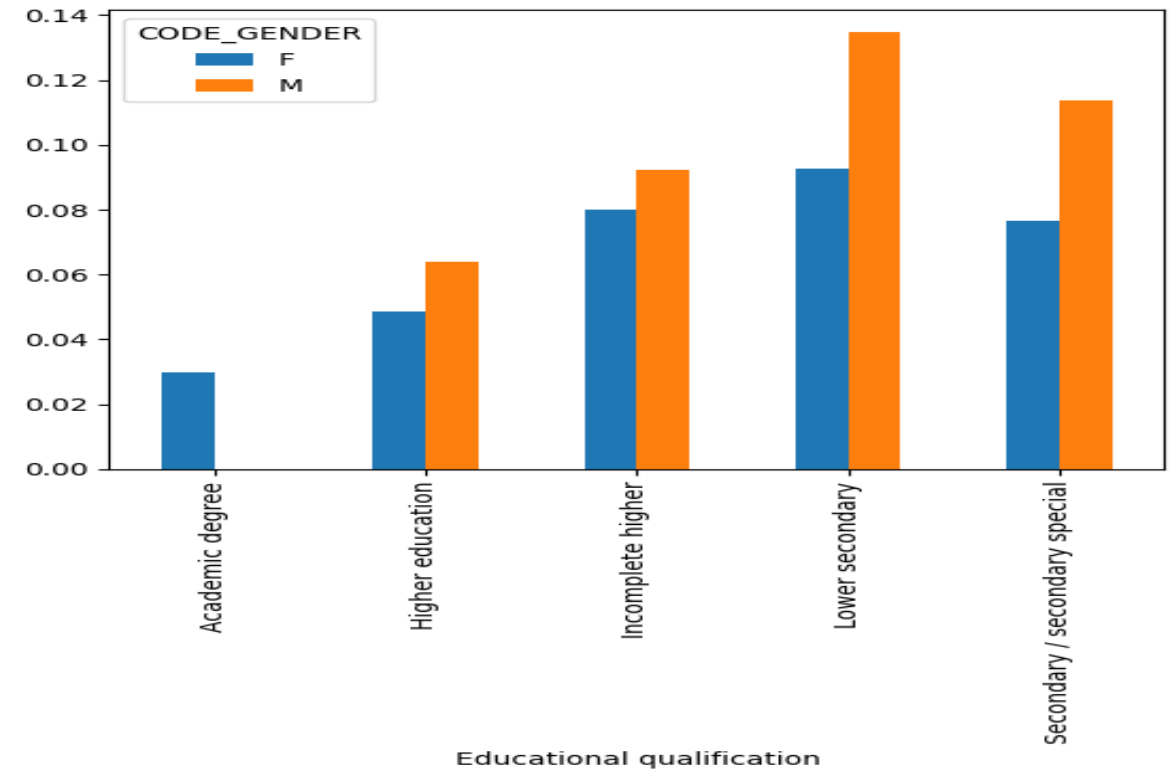
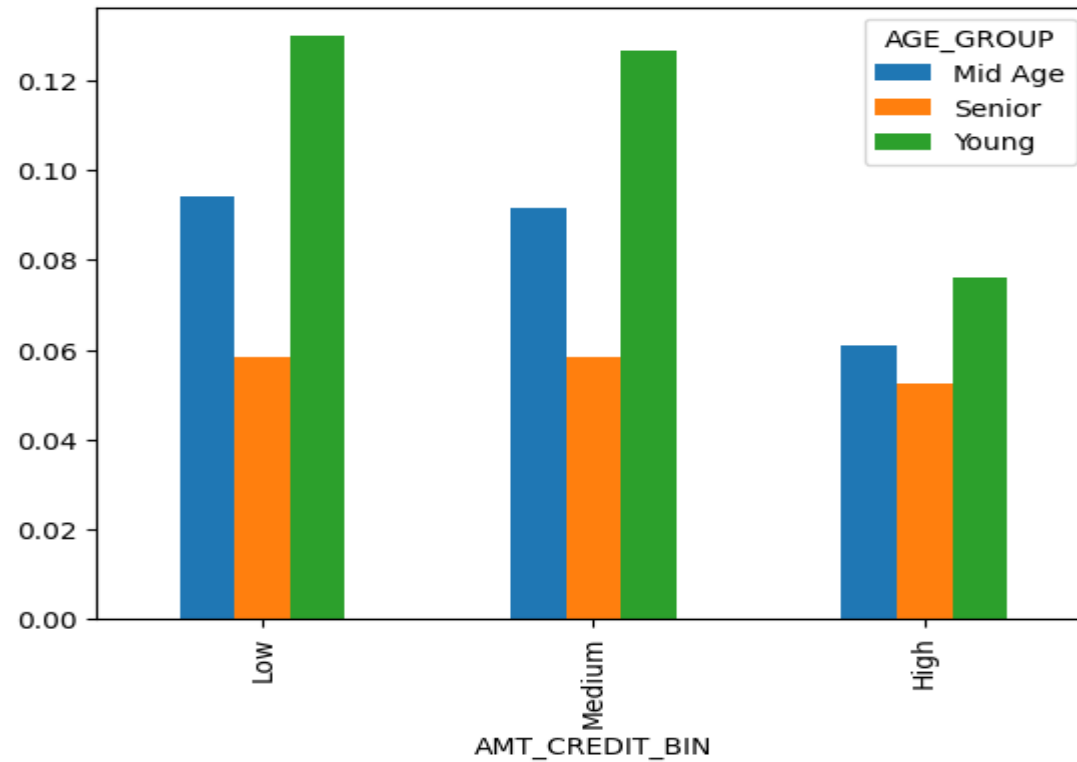
# Analysis



- Cash loans have more defaulters than Revolving loans
- Female applicants are more compared to male
- Low and Medium Credit amount applicants are more likely to not repay the low
- Medium income applicants are more likely to be the defaulters here

- DAYS\_EMPLOYED and DAYS\_REGISTRATION is inversely proportional (-ve correlation)
- AMT\_CREDIT, AMT\_ANNUITY and AMT\_GOODS\_PRICE are directly proportional (+ve correlation)

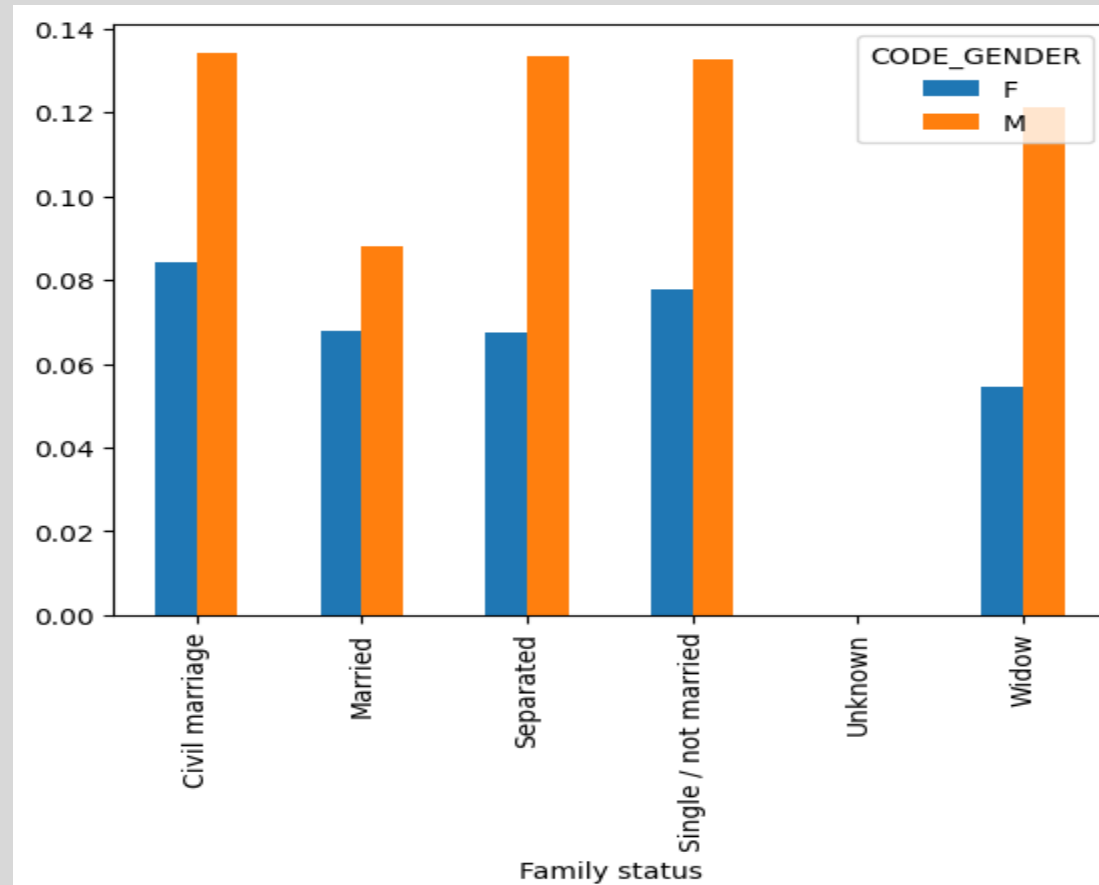




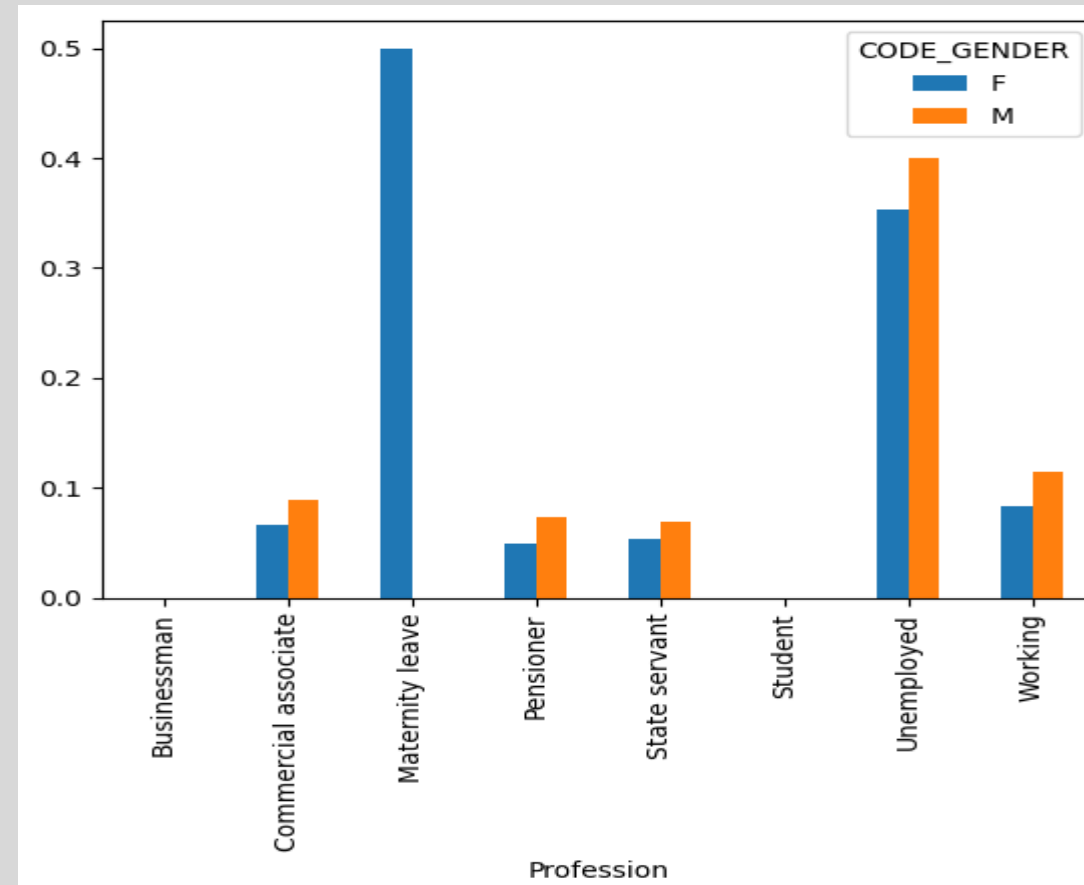
- Young applicants with medium and low credit amount group are high defaulters.
- Senior category applicants are less likely to default

- Lower secondary educated clients are high defaulted trailed by Secondary and Incomplete higher educated clients.
- The Degree level educated people are less defaulters.
- Females are less defaulted than male.





- Males are more defaulters across all family status

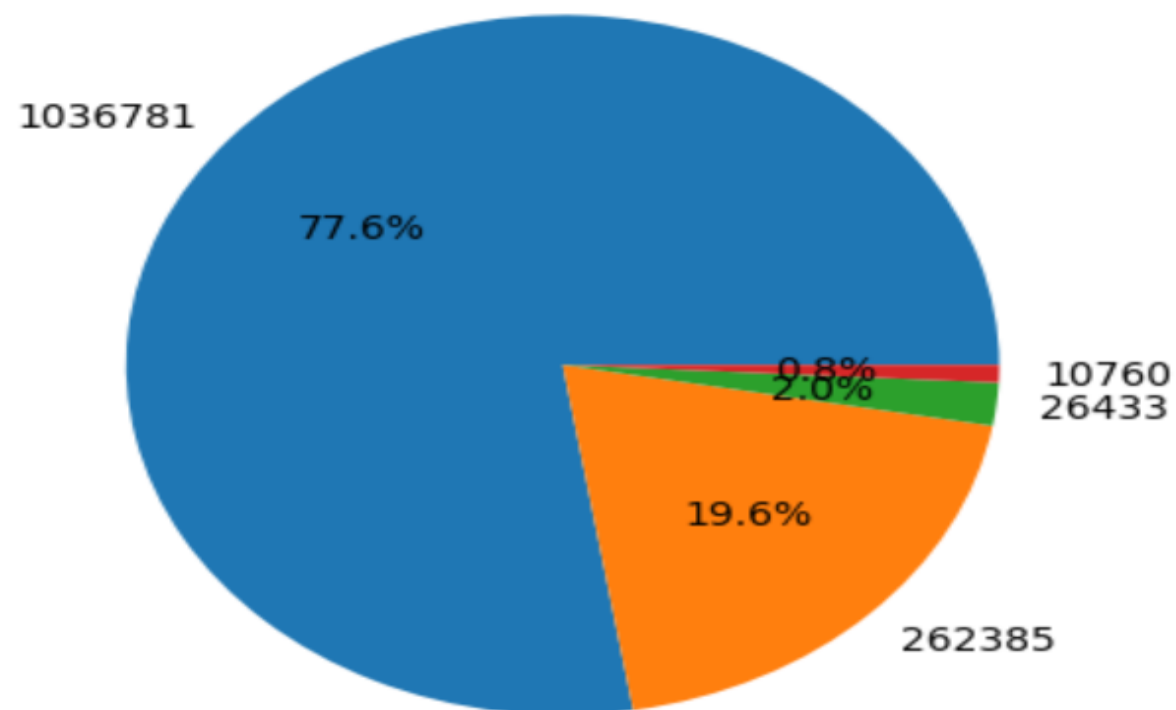


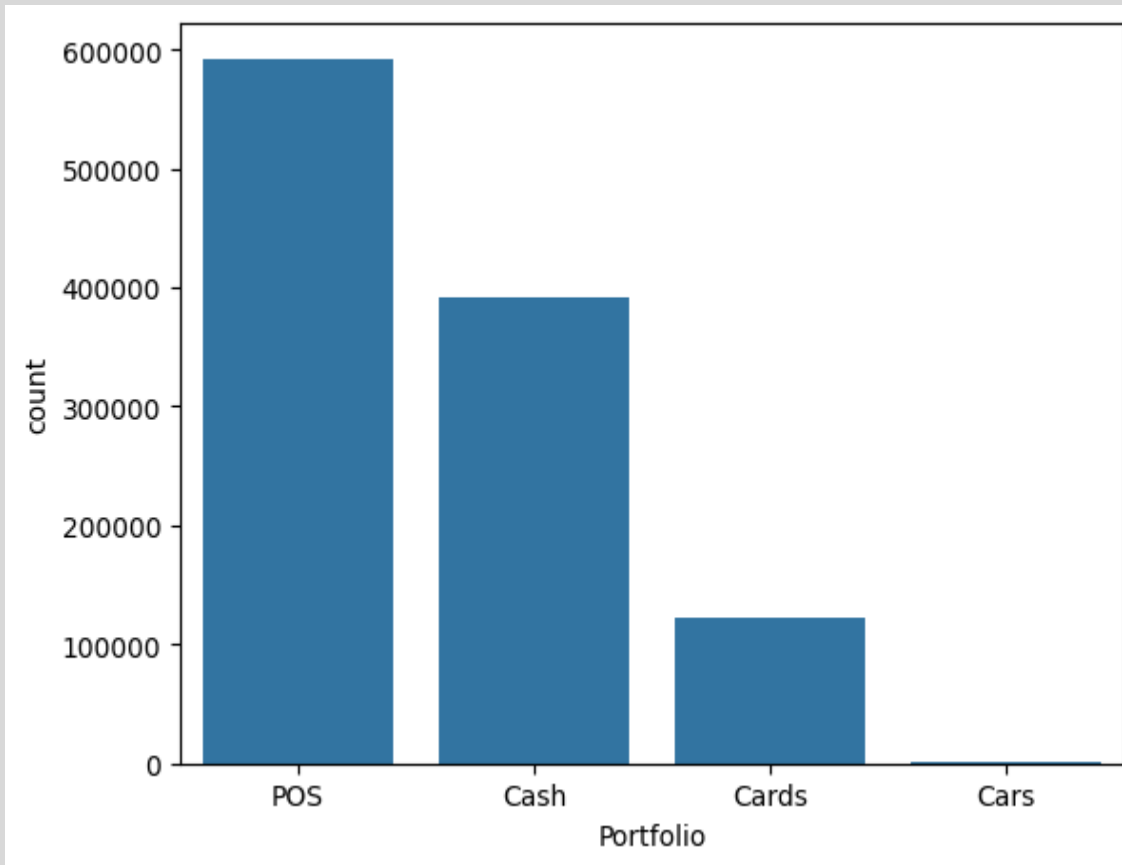
- Unemployed clients are more the defaulters.
- Maternity leave applicants are likely to be defaulted more.
- The default percentage is less in other professions.

- In the previous data set, most of the loans are Approved than Refused.
- Unused and Cancelled status are almost negligible

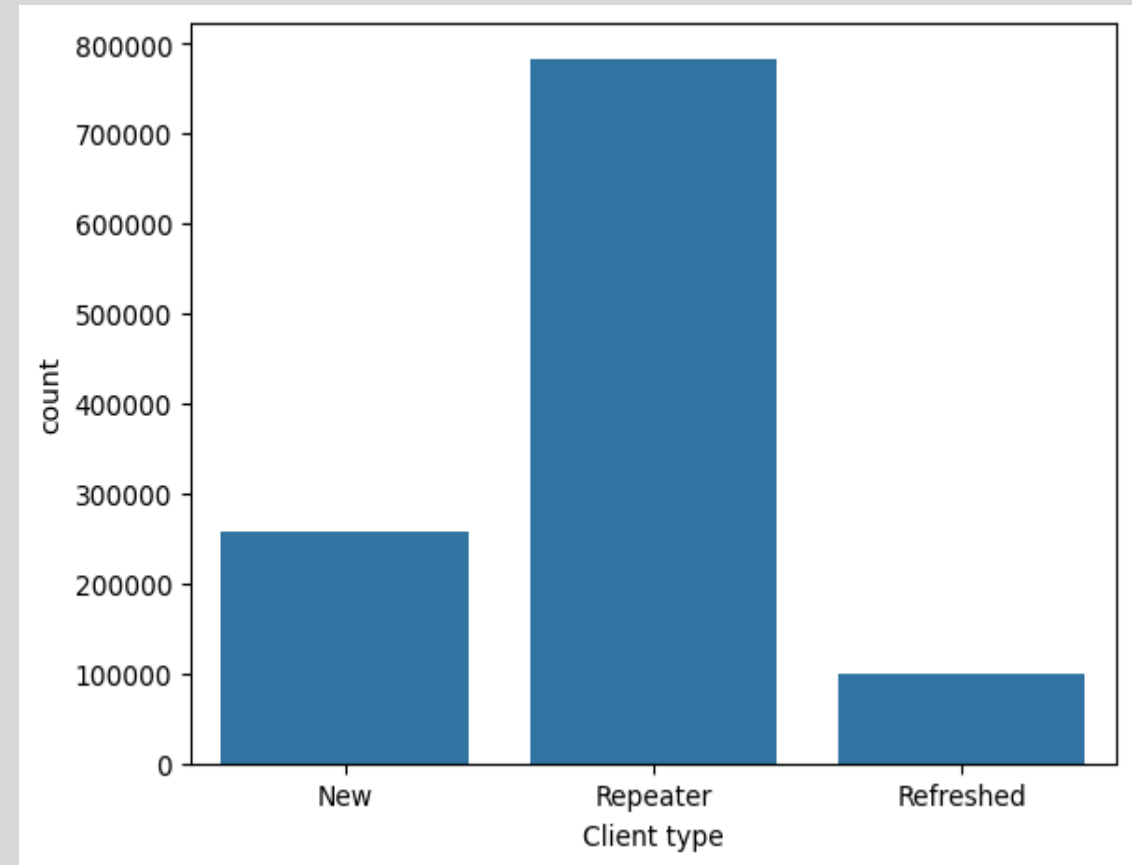
```
Approved      77.582521
Refused       19.634320
Unused offer   1.977986
Canceled       0.805173
Name: NAME_CONTRACT_STATUS, dtype: float64
```

Proportion of Previous Application Status





- The highest number of the previous applicants was for POS.
- Cash applicants has decent number.
- Cards applicants were very few.



- Most of the applicants are repeaters
- Followed by New applicants

# SUMMARY

Factors that indicate possibility of loan default:-

- Male applicants are more likely to default
- Individuals in civil marriages/single status has high default rate than the other.
- Less educated people are more prone to default their loan.
- Applicant's whose previous loans are rejected/cancelled are more likely to be the defaulters.
- Young applicants have high probability of default.
- Applicants who are on maternity leave or unemployment have higher default rate.
- Applicants with more children tend to be defaulters.
- Banks should focus more on contract type 'Student', 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for non-defaulting payments

THANK YOU!