

# Credit EDA Case Study

Harsha Tamma



# The problem

## Context

The data given below contains the information about the loan application at the time of applying for the loan.

It contains two types of scenarios:

- The client with payment difficulties
- All other cases: All other cases when the payment is paid on time.

## Problem statement

The company can utilise the analytical data for its portfolio and risk assessment while approving / choosing the customers in providing the loan



# Assumptions

- Values like XNA and XAP could be Missing Not A Random (MNAR). These are left as it is without any handling.



# Overall Solution Approach

Step 1

## Dataset Loading

- Import necessary Libraries.
- Load the datasets.

Step 2

## Dataset Cleaning

- Cleansing for columns wherever necessary and dropped the columns with majority of data as NA.
- Outliers are identified and handled wherever possible.
- Created any new columns as and when required.

Step 3

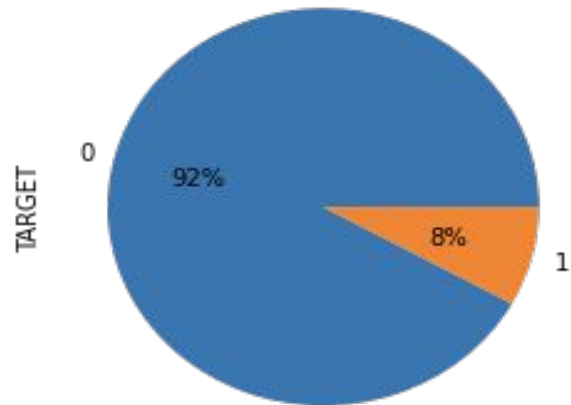
## Dataset Analysis

- Data imbalance is checked.
- Performed Univariate/Bi-variate/Segmented Univariate analysis of the processed datasets.
- Analysis on Current and Previous application data is done to derive insights based on Approval status of the bank loan.



# Data Imbalance

- **92%** of clients are **not having** difficulties in re-paying the loans on time.
- **8%** of the clients are **having** difficulties in re-paying loans on time.
- Ratio of Imbalance : **11.5**





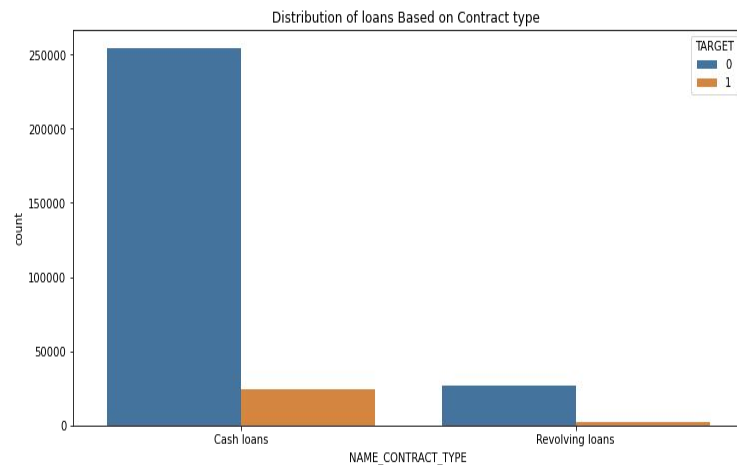
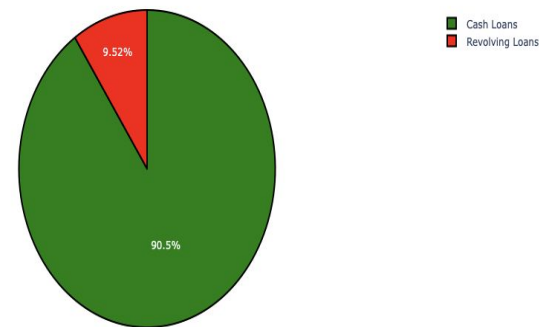
# Univariate/Bi-variate analysis



# Loan Type

- **Inference:**
  - Majority of Loans (90%) taken from banks are Cash Loans.
- **Analysis w.r.t Target Variable:**
  - Customers taking Cash loans are having more difficulties in repaying the loan when compared to Revolving loans.

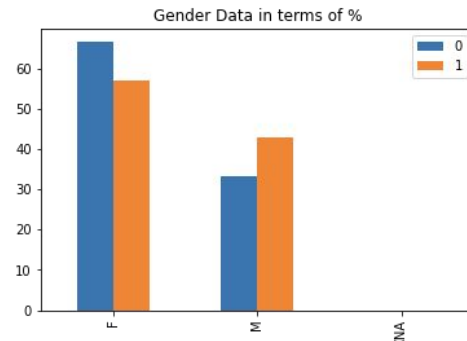
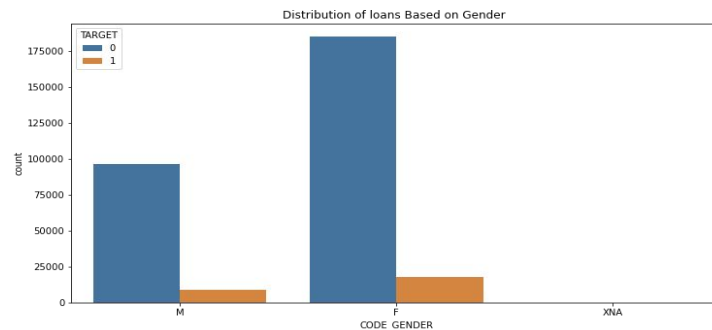
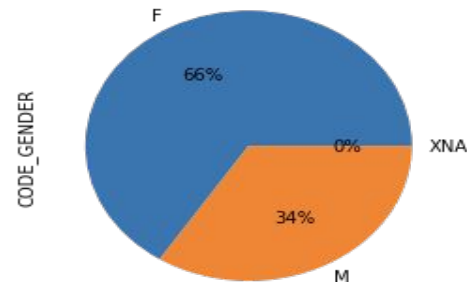
Percentage of Loan Type





# Gender

- **Inference:**
  - There are more Female applicants when compared to Male applicants
- **Analysis w.r.t Target Variable:**
  - Male applicants are having difficulties in repaying their loans when compared to Female applicants.







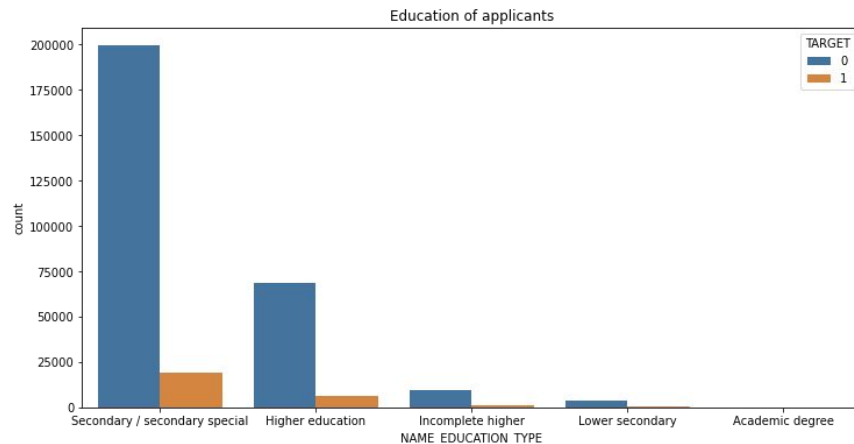
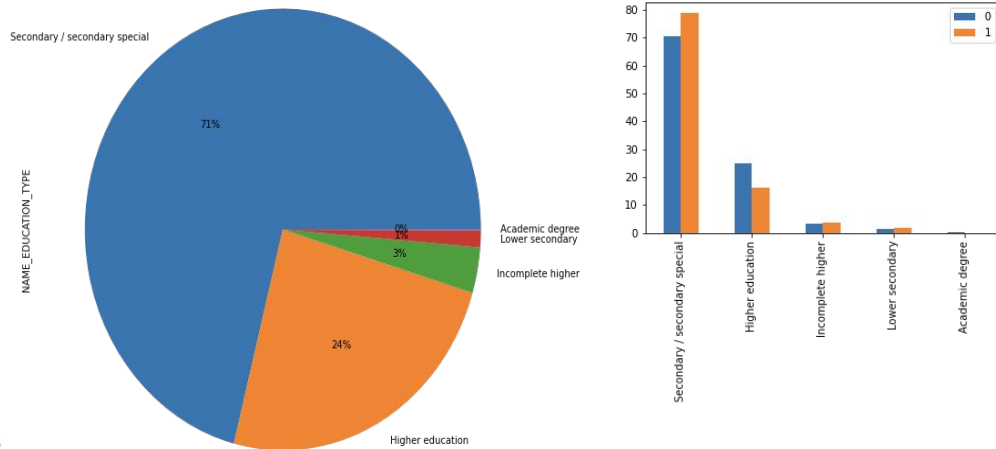
# Education

- **Inference:**

- Customers with Secondary/Secondary special, Higher Education are the ones that receive most of the loans.

- **Analysis w.r.t Target Variable:**

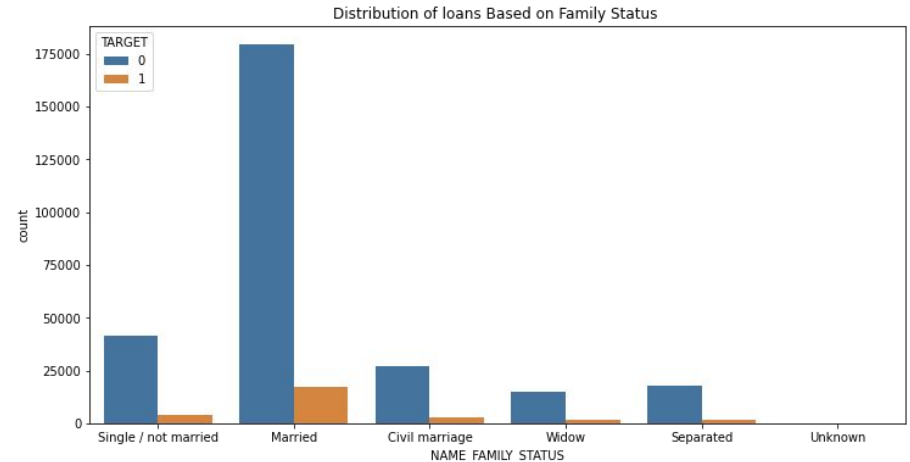
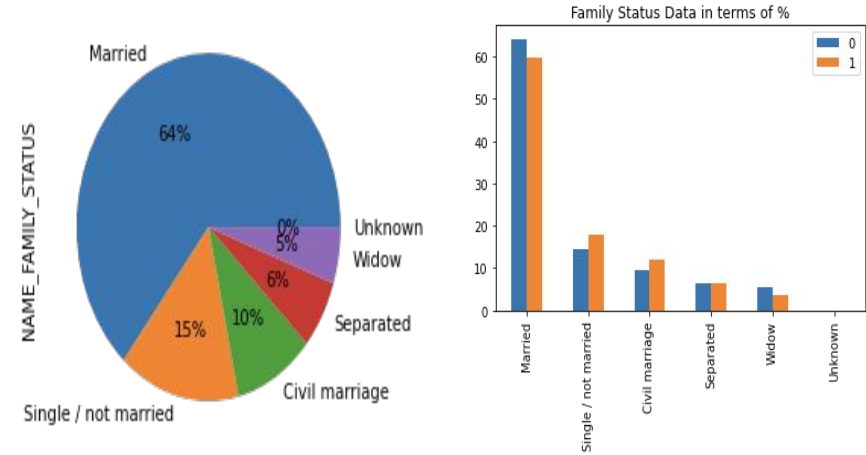
- Higher education customers have less difficulties in re-paying the loans.
- This could be due to better jobs and better living standards with higher education.





# Family Status

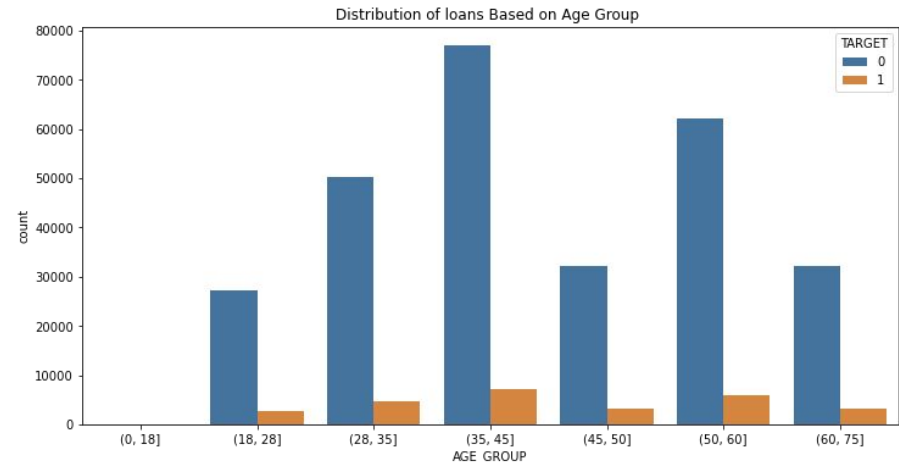
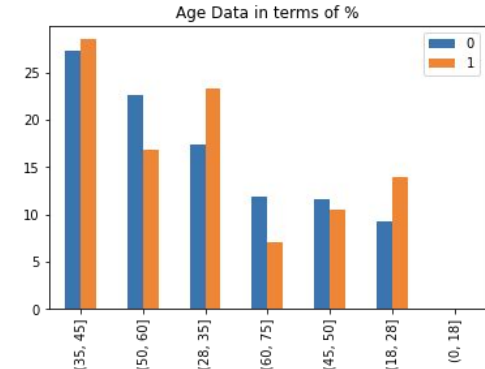
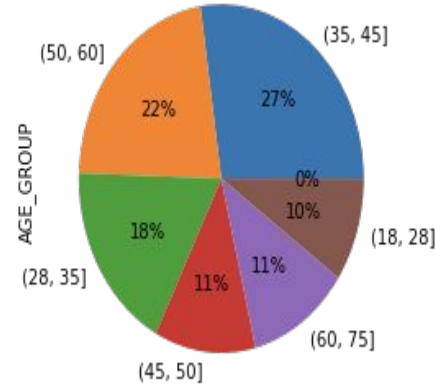
- **Inference:**
  - Most of the loans are take by married people
- **Analysis w.r.t Target Variable:**
  - Single/not-married people are facing difficulties in re-paying the loans compared to others.
  - This could be due to dual income sources.





# Age Group

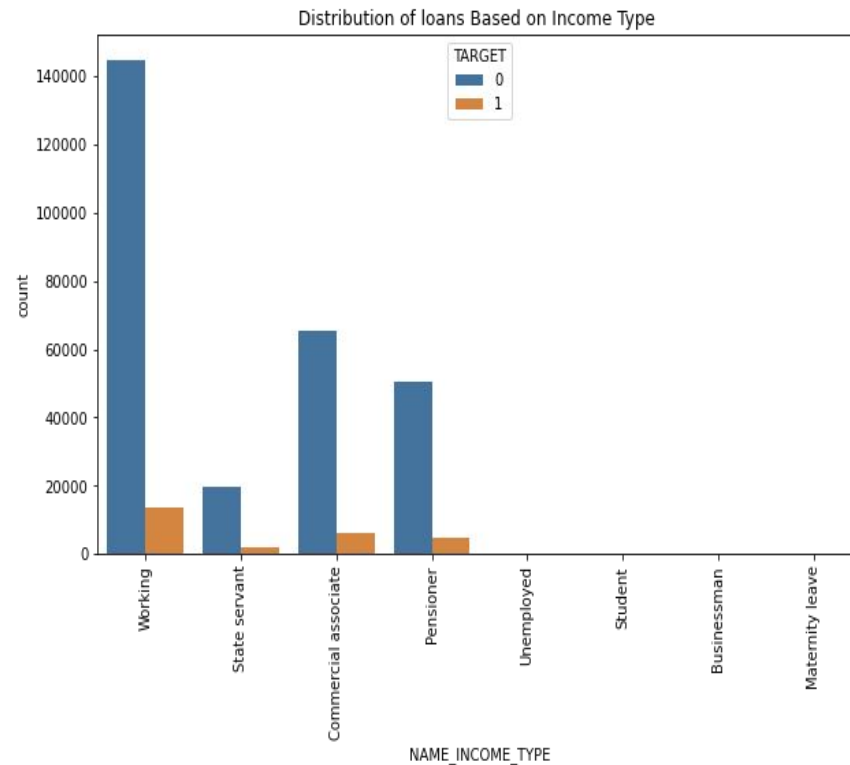
- **Inference:**
  - Almost all age groups (except youngest age) have taken loans.
- **Analysis w.r.t Target Variable:**
  - People between ages of 28-35 are having difficulties in repayment of loans.
  - People above age of 50 are having less difficulties in paying the loans.





# Income Type

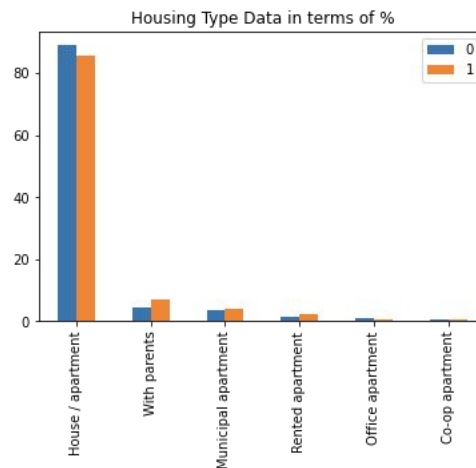
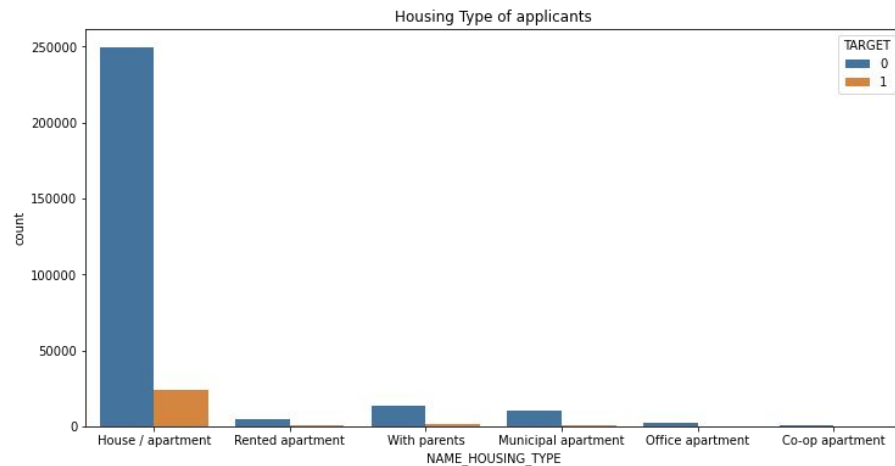
- **Inference:**
  - Working Professionals are taking most of the loans.
- **Analysis w.r.t Target Variable:**
  - Working professionals are having difficulties in re-paying the loans followed by Commercial associate.





# Housing Type

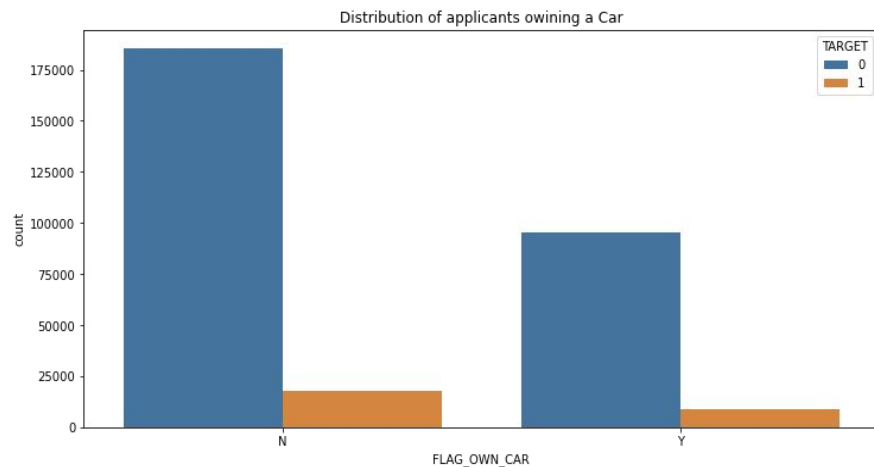
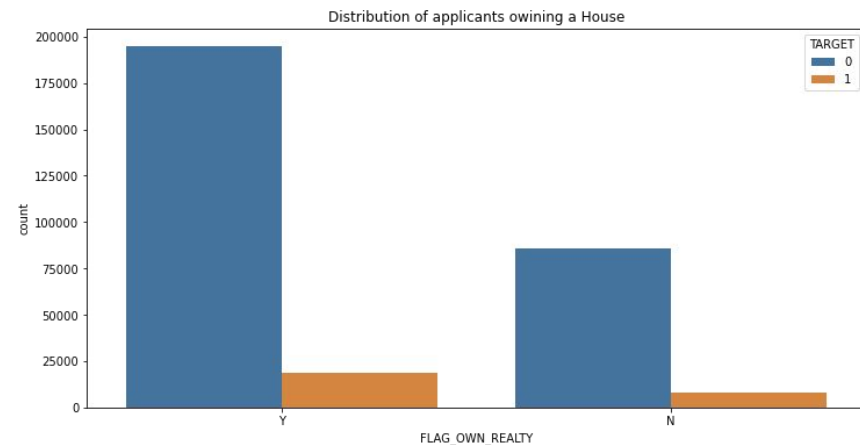
- **Inference:**
  - Most of the loans are taken by customers living in Apartments/Houses
- **Analysis w.r.t Target Variable:**
  - Customers living with parents are having difficulties in re-paying the loans.





# Own Car/House

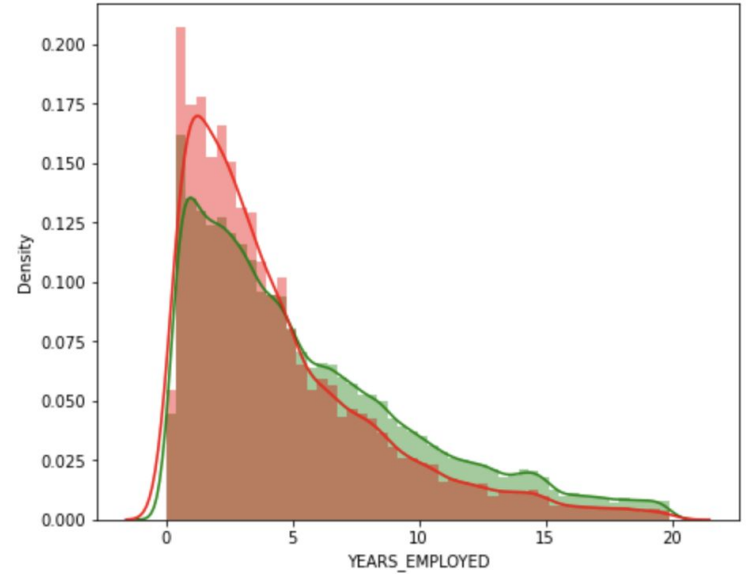
- Analysis w.r.t Target Variable:
  - Customers owing a car/house have less difficulties in repaying their loans.





# Work Experience

- **Analysis w.r.t Target Variable:**
  - Customers with work experience of less than 5 years are having difficulties in re-paying the loans





# Multivariate analysis

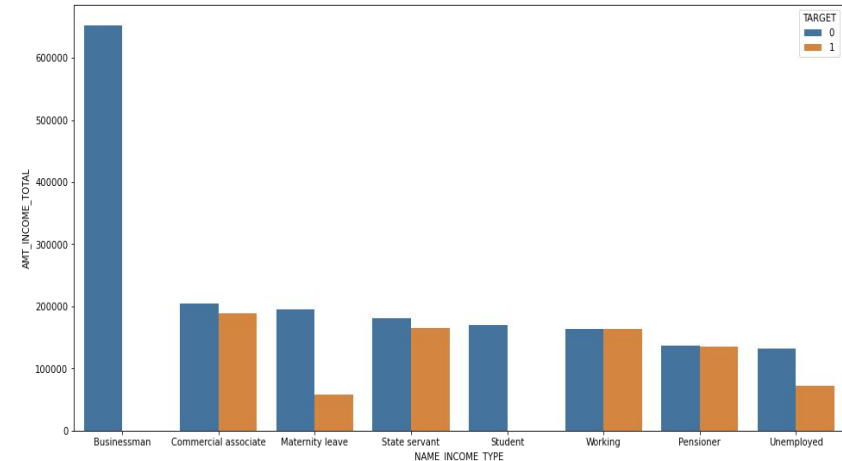
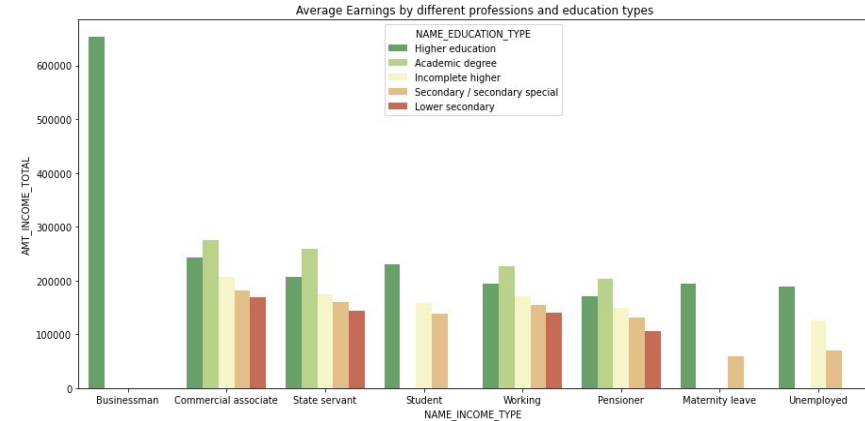






# Total Income, Education and Income Type

- Analysis w.r.t Target Variable:
  - Businessmen have the higher education and highest income. They also the safest customer to grant loan (as they don't default)
  - Working and Pensioner groups have high risk of defaulting on the loan.



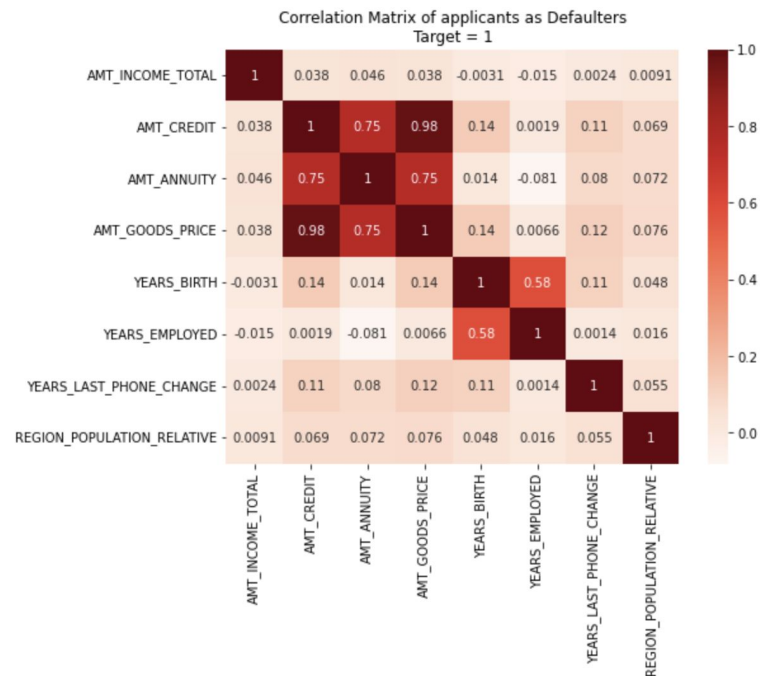


# Correlation



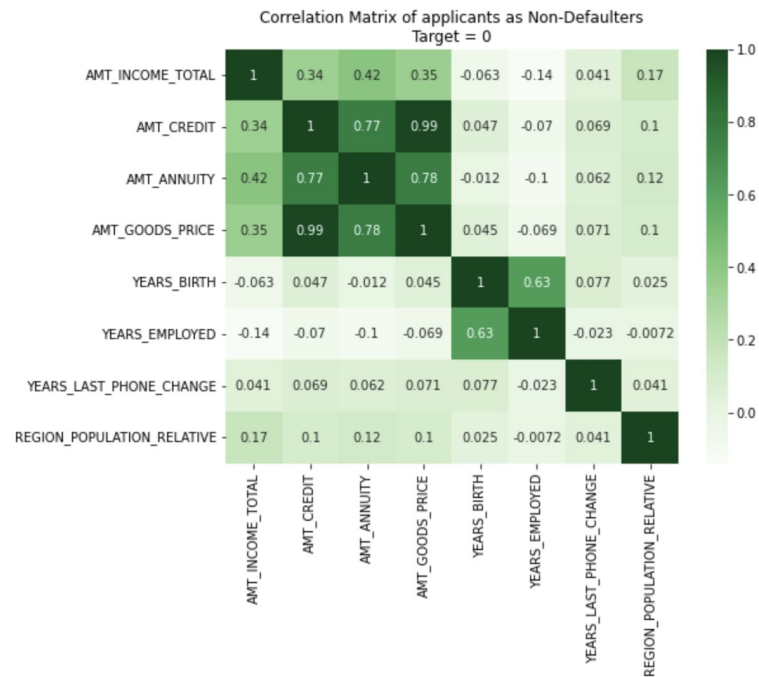
## Defaulters (>10%)

- AMT\_CREDIT – AMT\_GOODS\_PRICE : **98%**
- AMT\_ANNUITY – AMT\_CREDIT : **75%**
- AMT\_ANNUITY – AMT\_GOODS\_PRICE : **75%**
- YEARS\_BIRTH – YEARS\_EMPLOYED : **58%**
- AMT\_CREDIT – YEARS\_BIRTH : **14%**
- AMT\_GOODS\_PRICE - YEARS\_BIRTH: **14%**
- AMT\_GOODS\_PRICE – YEARS\_LAST\_PHONE\_CHANGE : **12%**
- AMT\_CREDIT – YEARS\_LAST\_PHONE\_CHANGE : **11%**
- YEARS\_BIRTH - YEARS\_LAST\_PHONE\_CHANGE : **11%**



## Non-Defaulters (>10%)

- AMT\_CREDIT – AMT\_GOODS\_PRICE : **99%**
- AMT\_ANNUITY – AMT\_GOODS\_PRICE : **78%**
- AMT\_ANNUITY – AMT\_CREDIT : **77%**
- YEARS\_BIRTH – YEARS\_EMPLOYED : **63%**
- AMT\_INCOME\_TOTAL – AMT\_ANNUITY : **42%**
- AMT\_GOODS\_PRICE - AMT\_INCOME\_TOTAL: **35%**
- AMT\_INCOME\_TOTAL – AMT\_CREDIT : 34%
- AMT\_INCOME\_TOTAL –  
REGIONAL\_POPULATION\_RELATIVE : **17%**
- AMT\_ANNUITY –  
REGIONAL\_POPULATION\_RELATIVE : **12%**



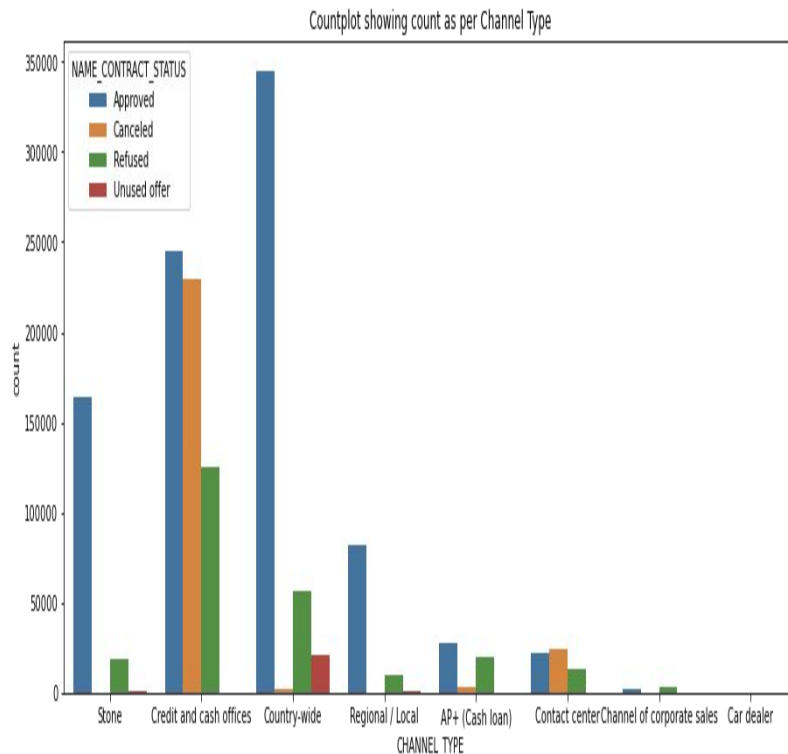


# Merged Data analysis



## Status, Channel Types and Contract status

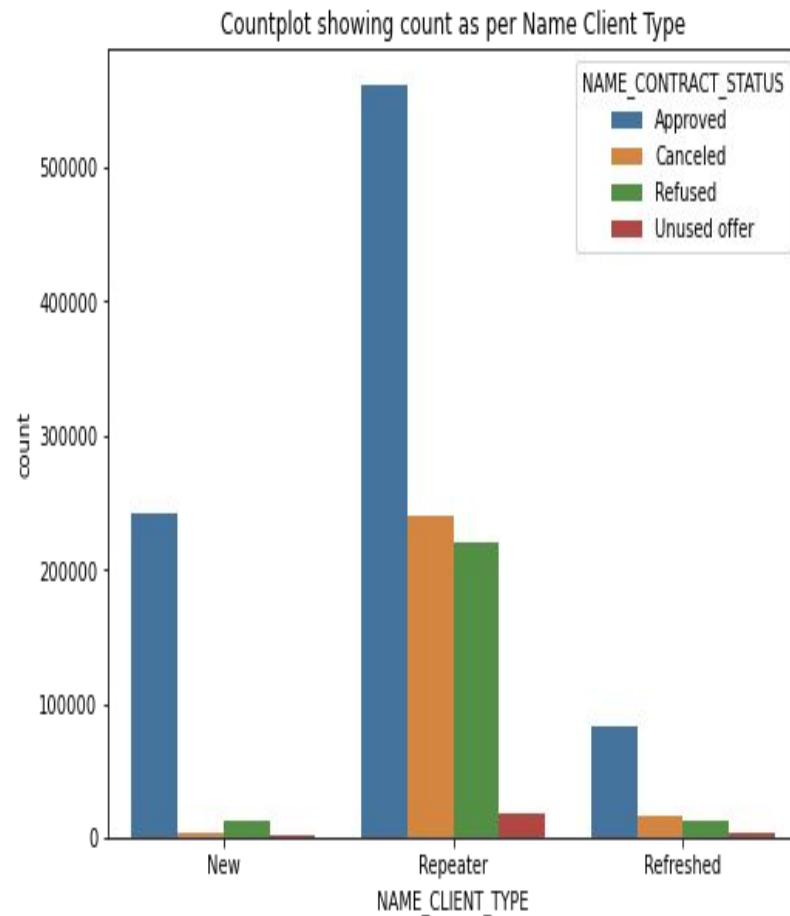
- Loan approvals are significantly higher than Cancelled, Refused and Unused.
- Country wide channel has the highest loan approvals.
- Stone Channels have 0% loan cancellations.





## Client Type and Contract status

- Repeater Client seems to have more number of refusals and cancellations.





# Conclusion

- Banks should focus on clients with age of more than 50.
- Banks should prefer customers who live and own a house/apartment/car
- Banks should focus on Country wide channel type as it sees more approvals, where as credit and cash sees more refusals and cancellations.
- Banks should focus on clients with education type “Higher education” and avoid Secondary/secondary special, incomplete higher or lower secondary as they face paying difficulties.
- Banks should focus on Business professionals since there are no defaults. They should avoid “working” clients as they have high percentage of defaults.

