

# CREDIT EDA CASE STUDY

## **Business Objectives :**

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.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

**Prepared by:**  
**Mr. Akshay Bugade**

Date: 01/12/2021

# SUMMARY OF DATA: FOR EDA ANALYSIS

Given Dataset :

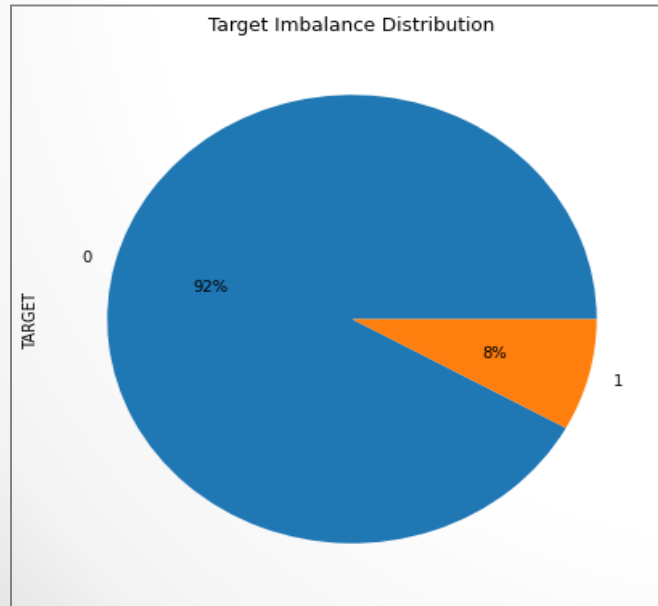
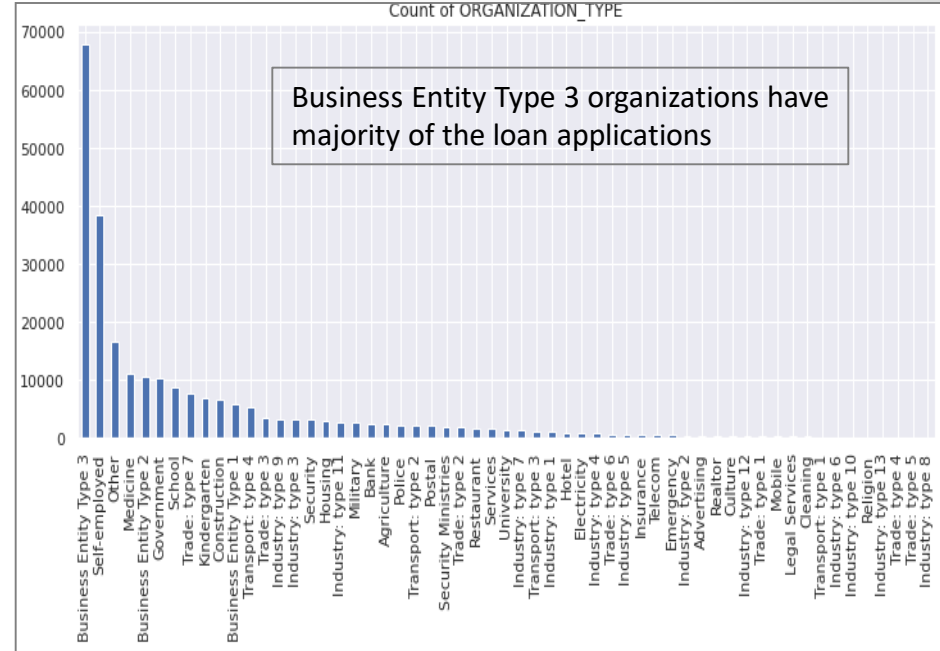
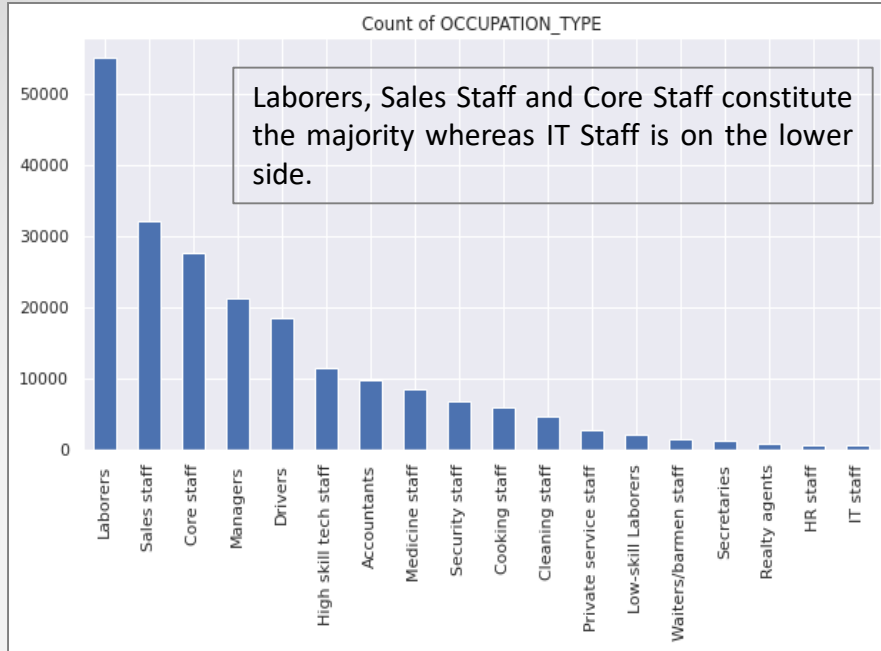
- **Application Data**

- Existing No.of entries- 3,07,511 (rows) & Features – 122(columns)
- Useful No.of entries- 3,07,511 & Features – 73 (after removing of more than 45% missing values column)
- TARGET Column's categories on which we have to work for insight
  - **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample –indicated as “ 1 ”
  - **All other cases:** All other cases when the payment is paid on time –indicated as “ 0 ”

- **Previous Application Data**

- Existing No.of entries- 16,70,214 (rows) & Features – 37 (columns)
- TARGET Column's categories on which we have to work for insight
  - **Approved:** The Company has approved loan Application
  - **Cancelled:** The client cancelled the application sometime during approval ( due various reasons)
  - **Refused:** The company had rejected the loan (Client doesn't meet banks requirement)
  - **Unused offer:** Loan has been cancelled by the client but on different stages of the process.

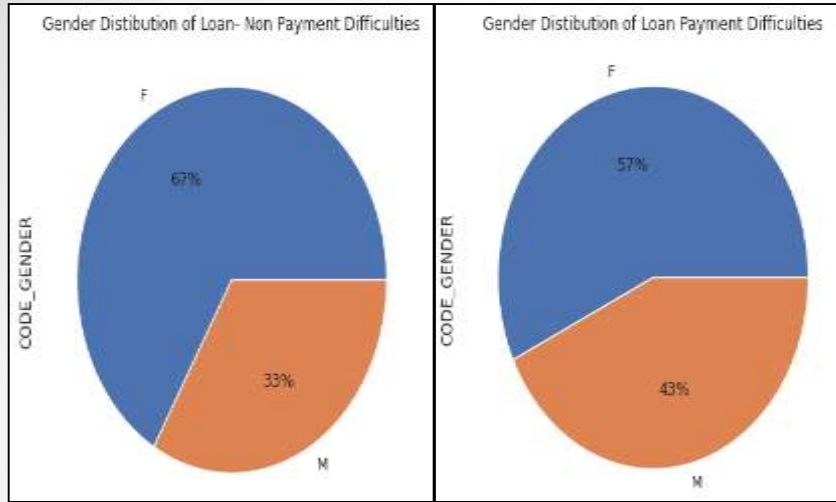
# Count of Applicants :Application Data



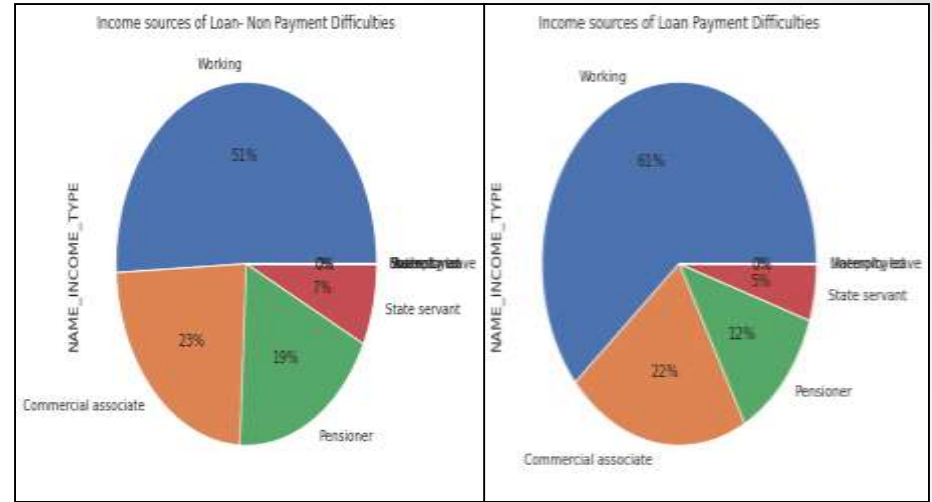
**Out of 3,07,511 Applicants:**

- 2,82,910 Applicants made payment on time (92%)
- 24,600 Applicants with payment difficulties (8%)

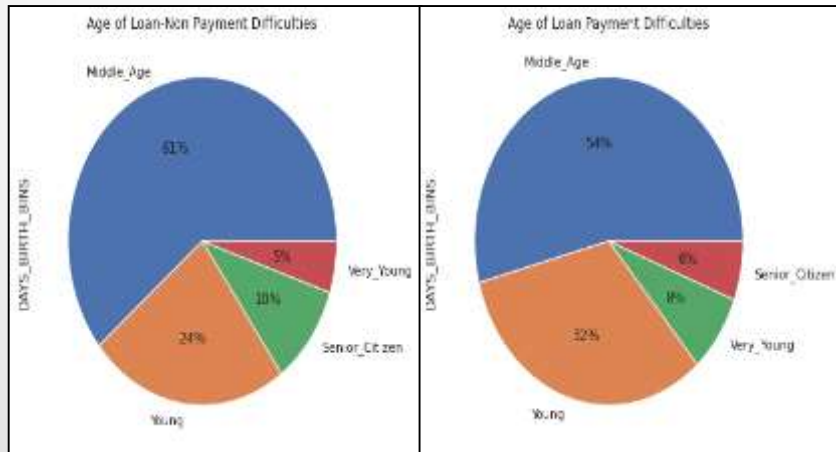
# Loan Payment Defaulter categories w.r.t applications



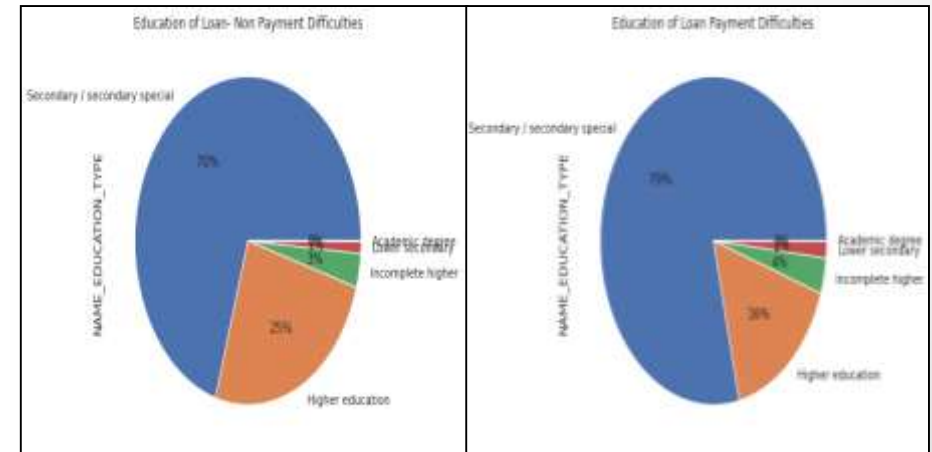
CODE\_GENDER :Male has more Payment Difficulties than female



INCOME\_TYPE: Working has more Payment Difficulties than others

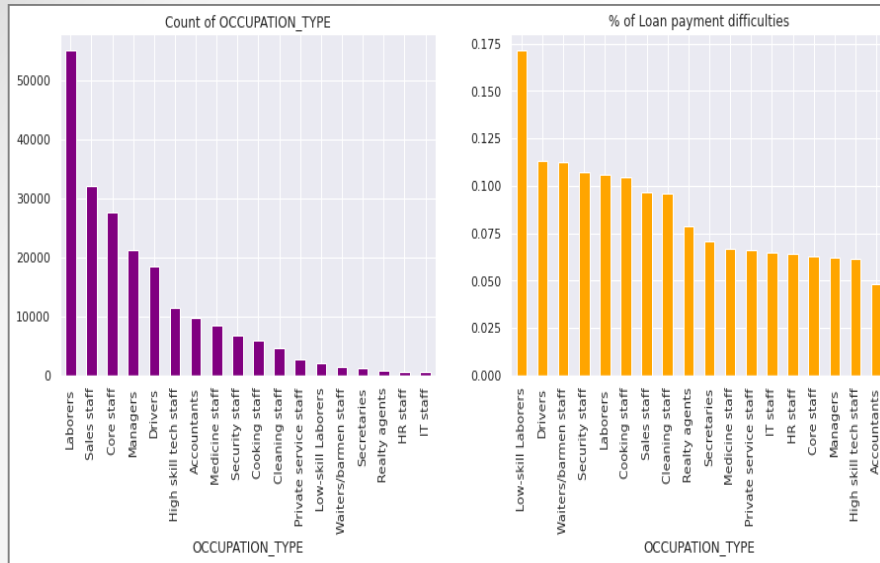


DAY\_BIRTH\_BINS : Young has more Payment Difficulties than others



NAME\_EDUCATION\_TYPE: Secondary has more Payment Difficulties than female

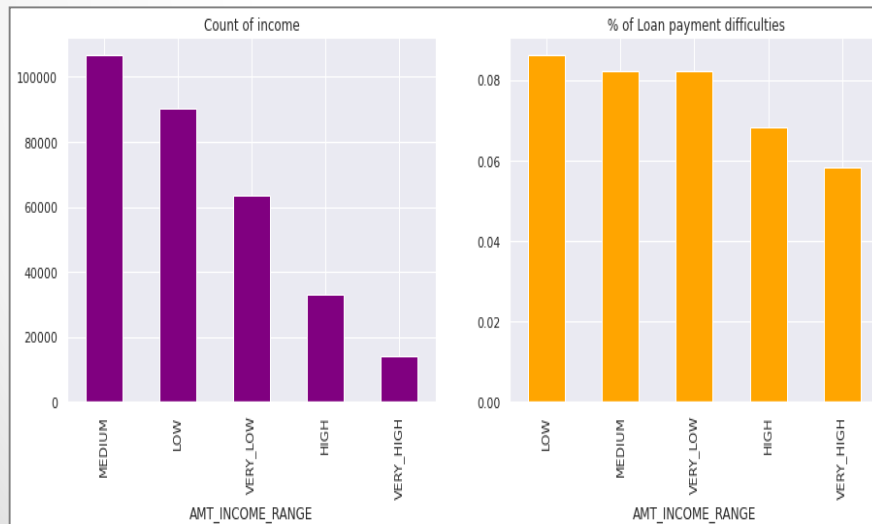
# Major categories : Loan Defaulter



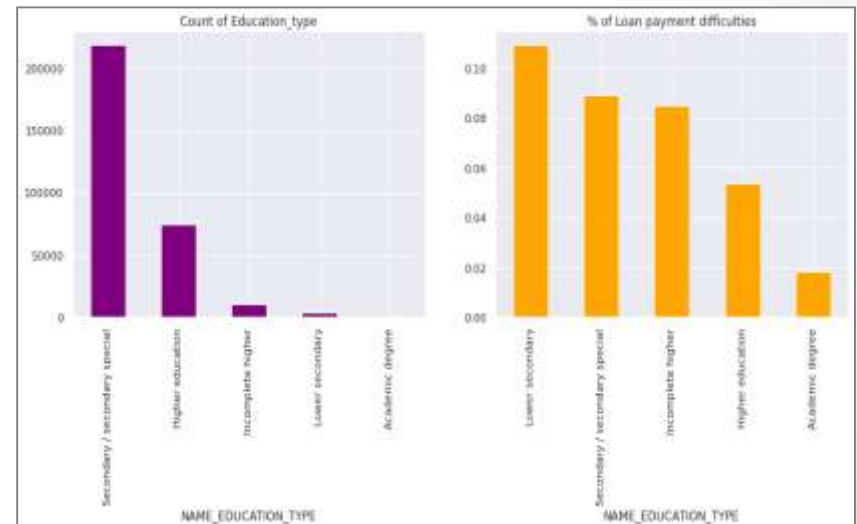
clients with '**Lower skill Laborers**' occupation type have maximum % of Loan-Payment Difficulties



Clients with '**Maternity leave**' Income type have maximum % of Loan-Payment Difficulties



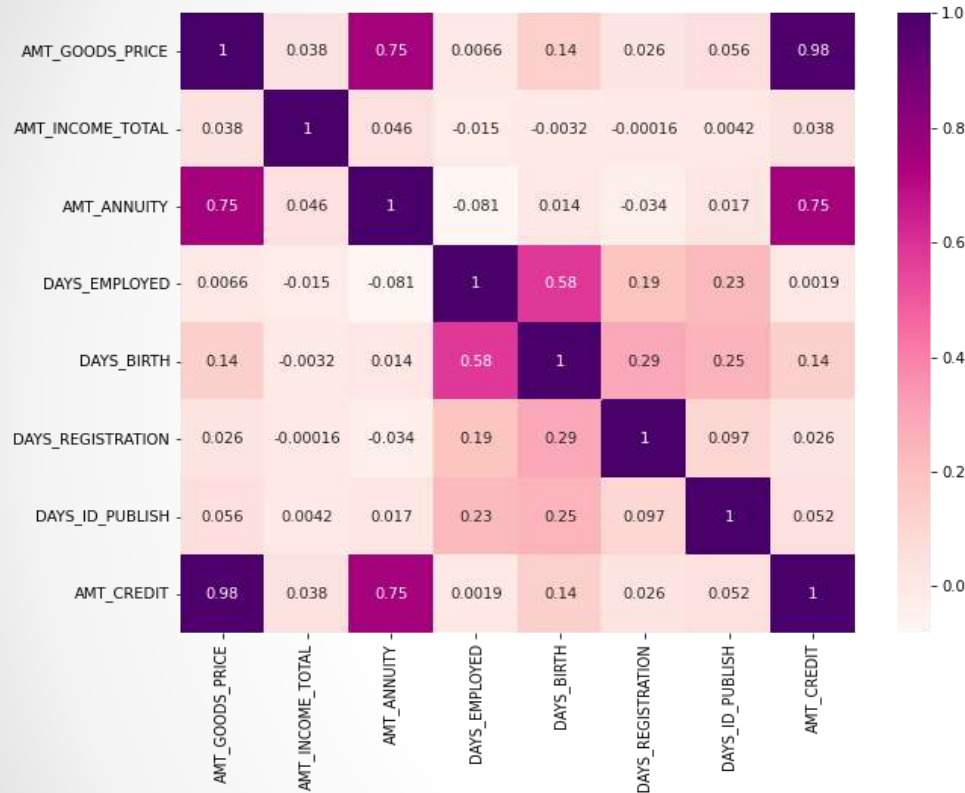
Clients with '**LOW**' Income range have maximum % of Loan-Payment Difficulties



Clients with '**Lower secondary**' education type have maximum % of Loan-Payment Difficulties

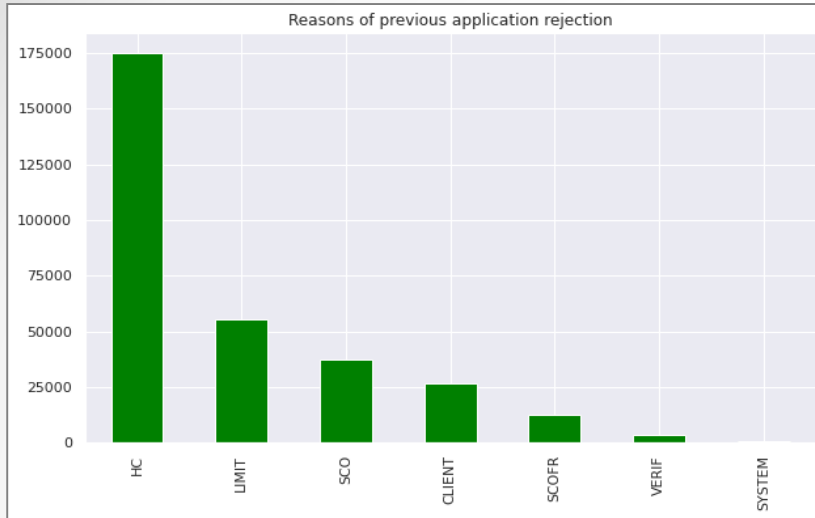
# Top 10 Correlation:

- AMT\_CREDIT & AMT\_GOODS\_PRICE has Maximum positive Correlation i.e 0.98

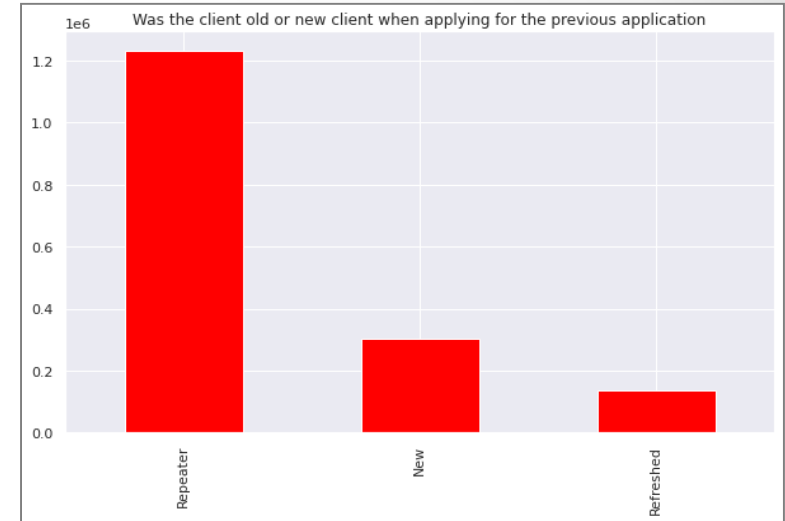


SR.NO.	VAR1	VAR2	CORRELATION	CORR_ABS
1	AMT_CREDIT	AMT_GOODS_PRICE	0.98	0.98
2	AMT_ANNUITY	AMT_GOODS_PRICE	0.75	0.75
3	AMT_CREDIT	AMT_ANNUITY	0.75	0.75
4	DAYS_BIRTH	DAYS_EMPLOYED	0.58	0.58
5	DAYS_REGISTRATION	DAYS_BIRTH	0.29	0.29
6	DAYS_ID_PUBLISH	DAYS_BIRTH	0.25	0.25
7	DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.23	0.23
8	DAYS_REGISTRATION	DAYS_EMPLOYED	0.19	0.19
9	DAYS_BIRTH	AMT_GOODS_PRICE	0.14	0.14
10	AMT_CREDIT	DAYS_BIRTH	0.14	0.14

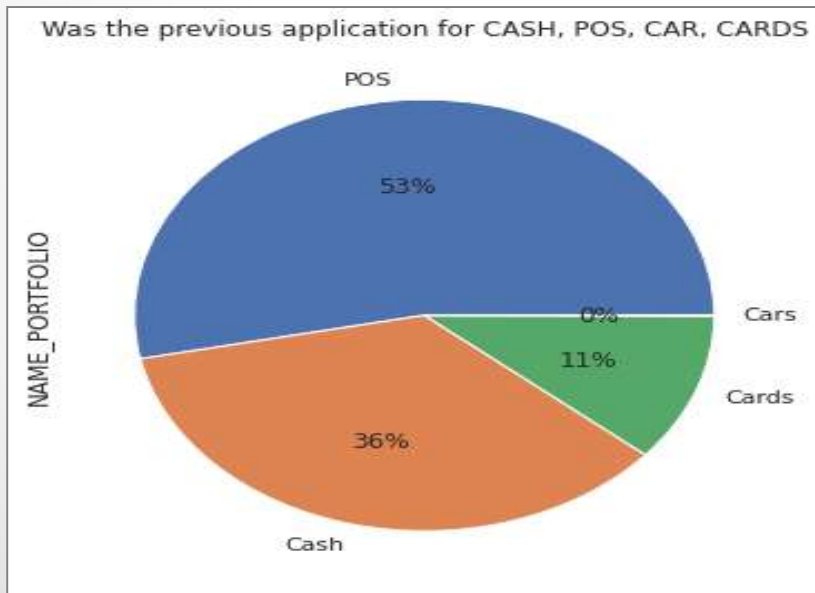
# Previous Application Summary:



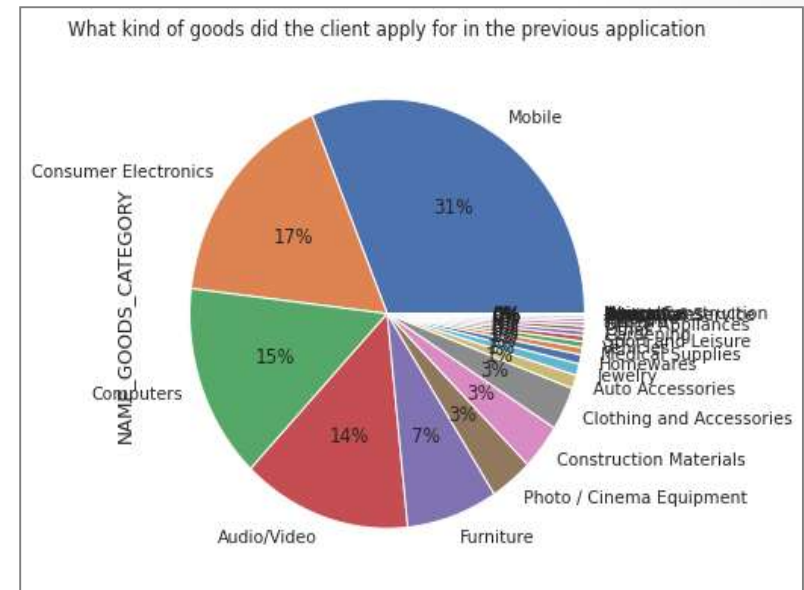
HC is the reason majority of applications got rejected



Majority of the clients are repeaters.

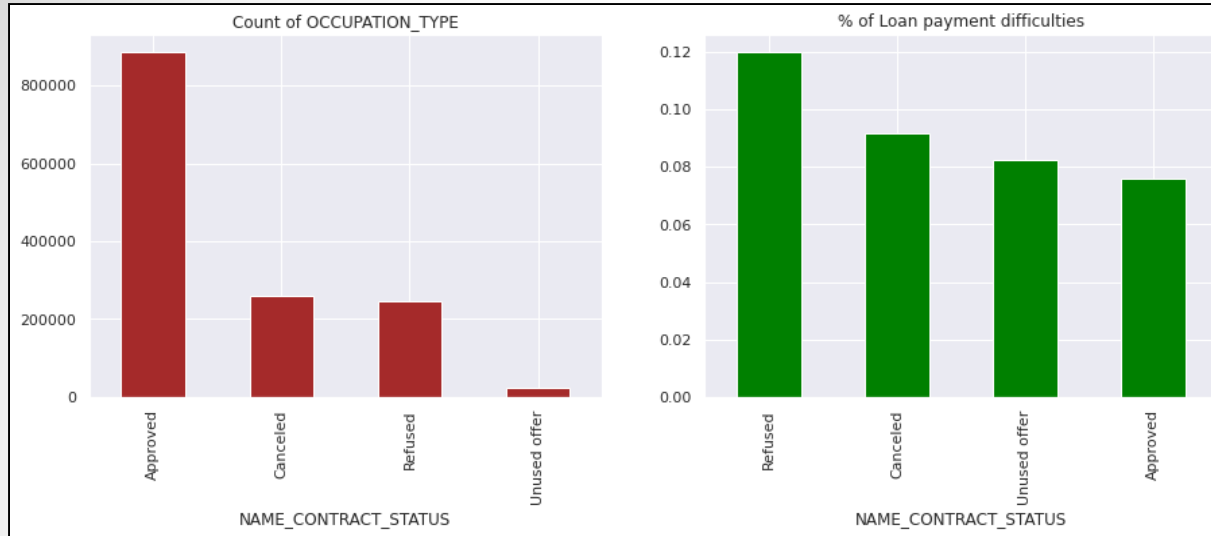


Majority of previous application is for POS and a good amount of it is for cash.

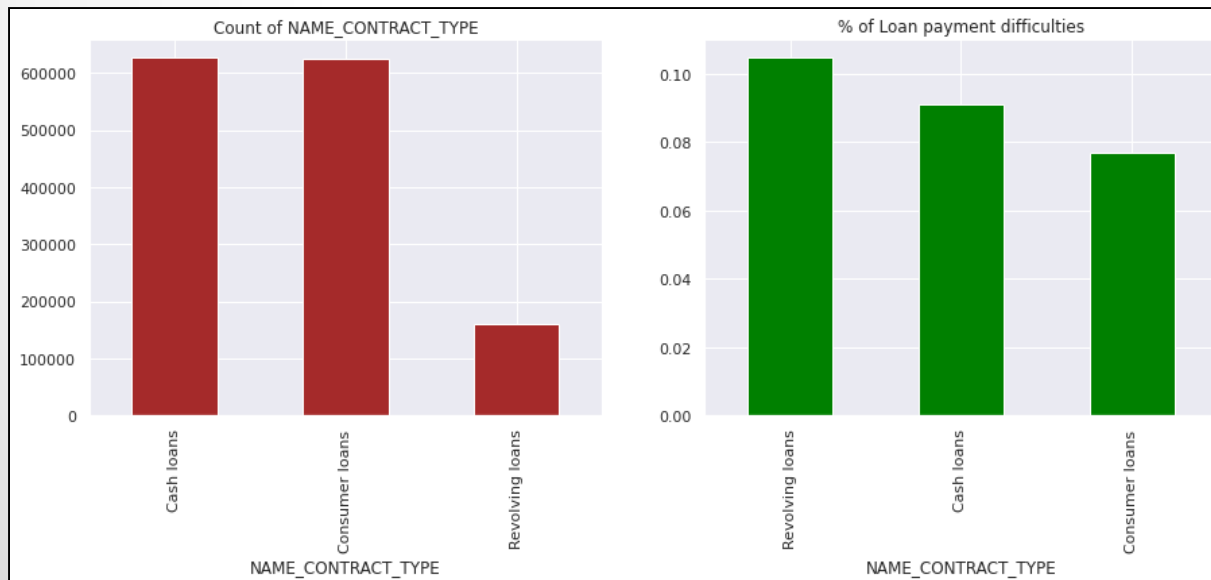


Majority of loans are for mobiles, consumer electronics, computers and Audio/Video

# Loan Payment Defaulter categories w.r.t applications



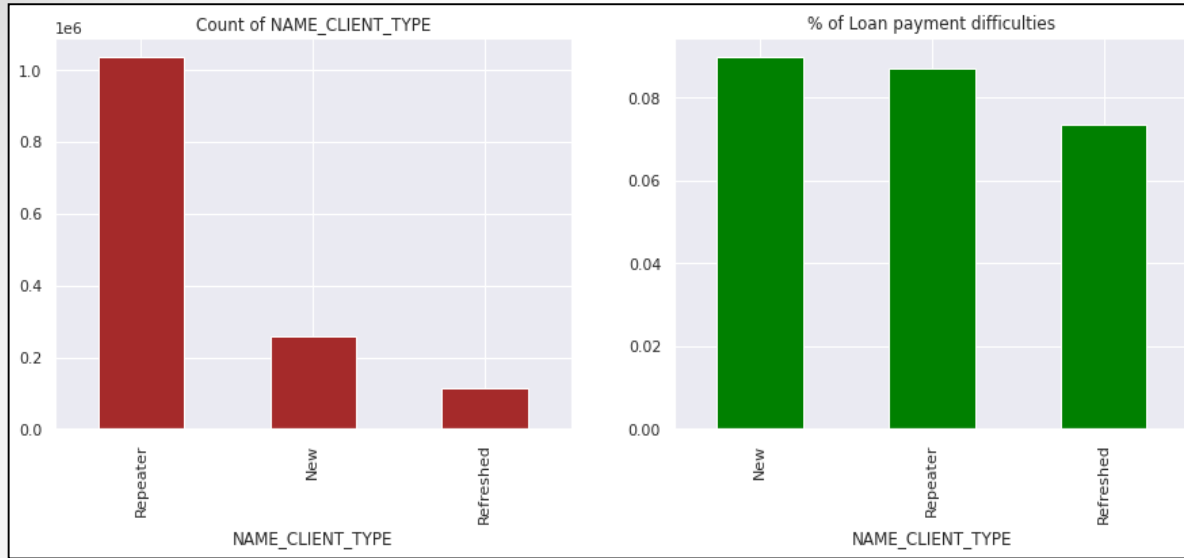
- 'Refused' contracts from previous application are the ones who have maximum % of Loan-Payment Difficulties from current application.
- 'Approved' contracts from previous application are the ones who have minimum % of Loan-Payment Difficulties from current application.



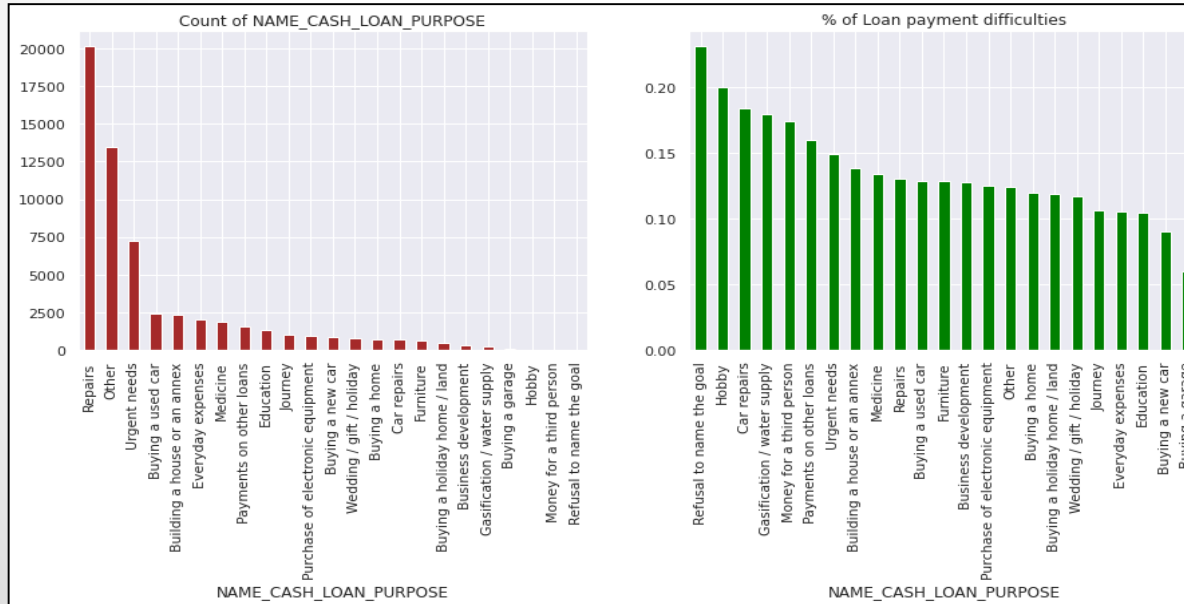
- most of the contract type from previous application was 'Cash loans'



# Loan Payment Defaulter categories w.r.t applications



- 'New' clients from previous application are the ones who have maximum % of Loan-Payment Difficulties from current application.



- Refusal to name the goal' for cash loan from previous application are the ones who have maximum % of Loan-Payment Difficulties from current application.

## Driving Factors for Loan Payment Defaulter:

Count of subcategory Applicants from below tables are comparatively very less in its Category still have Maximum % of payment difficulties

Sr. No.	Category	Subcategory	Maximum % of Payment Difficulties
1	NAME_INCOME_TYPE	Maternity leave	40
2	OCCUPATION_TYPE	Low skilled Laborers	17
3	NAME_EDUCATION_TYPE	Lower Secondary	11
4	NAME_CASH_LOAN_PURPOSE	Refusal to name the goal	23
5	NAME_CONTRACT_STATUS	Refused	12
6	NAME_CONTRACT_TYPE	Revolving Loans	10

Thank You