Problem Statement

To analyse data on loans and find pattern that predicts of client having difficulty of payment in the future

---

There are 2 types of risks associated while providing loan to a client

1. If the loan is not provided to the customers who can pay back then it will be loss for the company
2. If the loan is provided to the customer who can not pay then also it will be the loss for the company.

Steps followed for analysis on the data

Data understanding

- Importing correct libraries
- Checking each column, Index, header, footer etc
- Identifying data quality issues

Data Cleaning and Manipulation

- Missing value imputation analysis
- Checking the structure and the metadata
- Changing datatypes to date, time, string, int, bool, etc for ease of analysis

Outlier check and data imbalance check

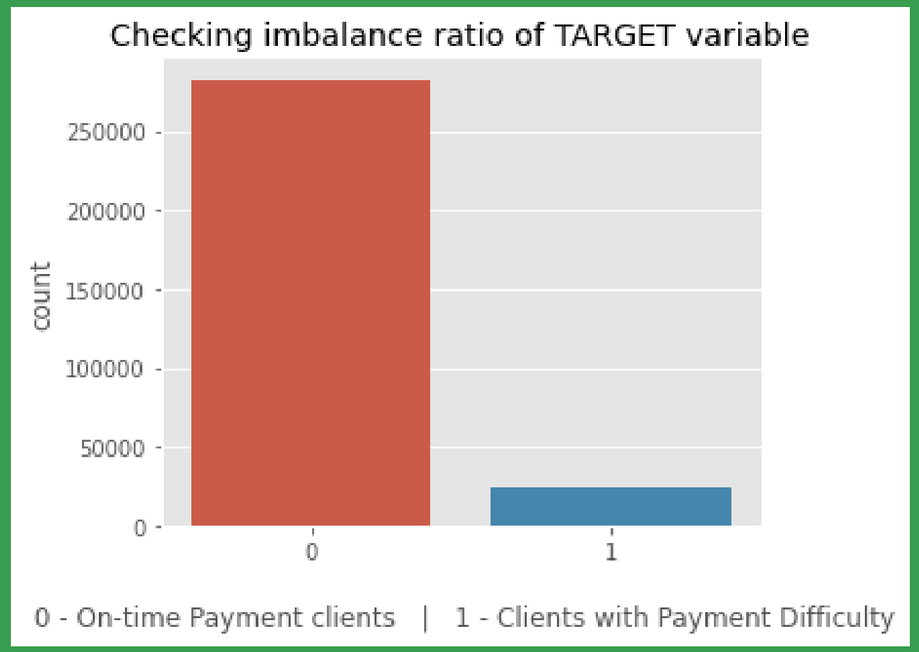
- Checking the data for outliers that would cause the analysis to be biased
- Checking for imbalances, ratio, percentage of imbalance

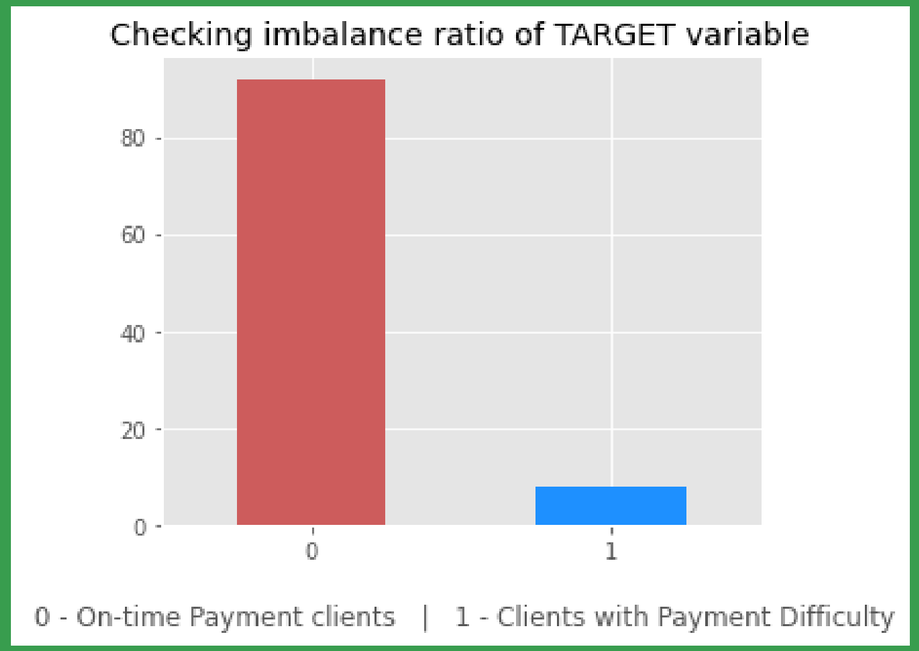
Data analysis

- Business requirement oriented analysis
- Correlation between columns
- Univariate analysis
- Bivariate analysis
- Creating plots to understand the data better and find insight

Conclusion and Presentation

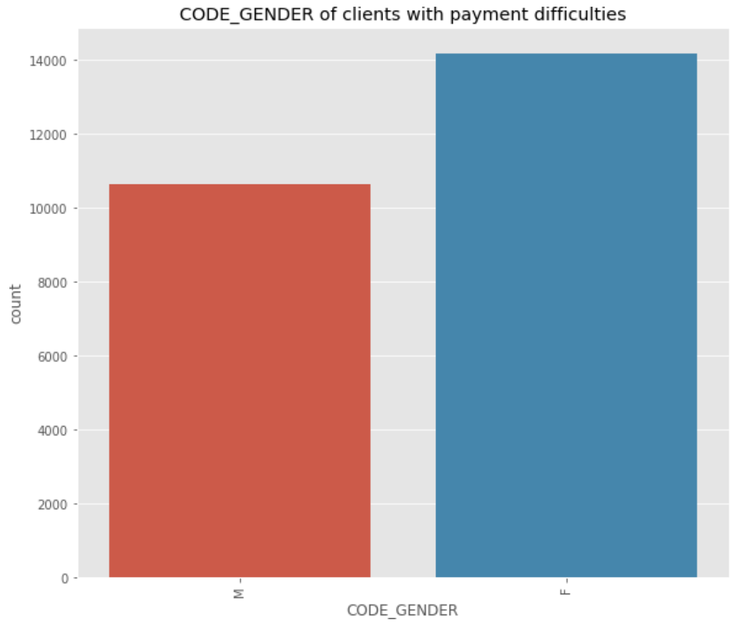
- Explains Business-driven, type-driven and data-driven metrics created in previous steps
- Helps in getting better understanding about the data

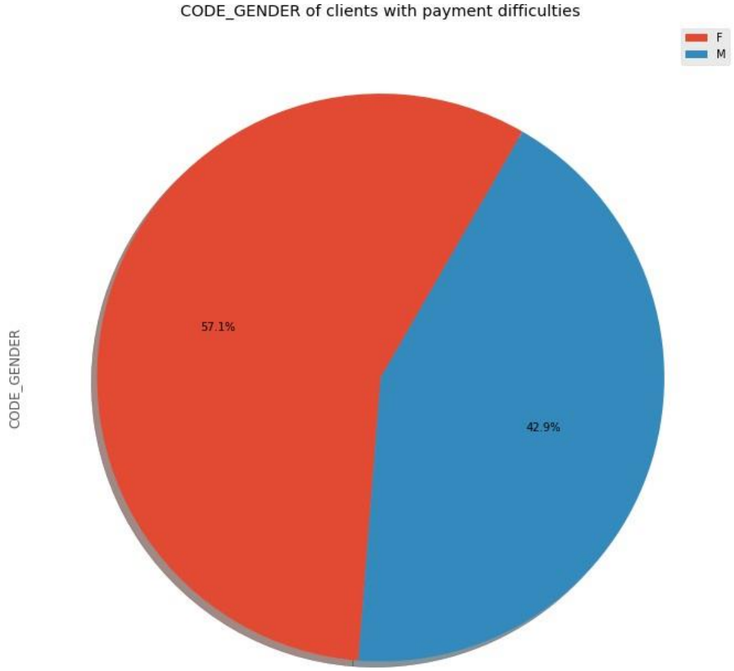


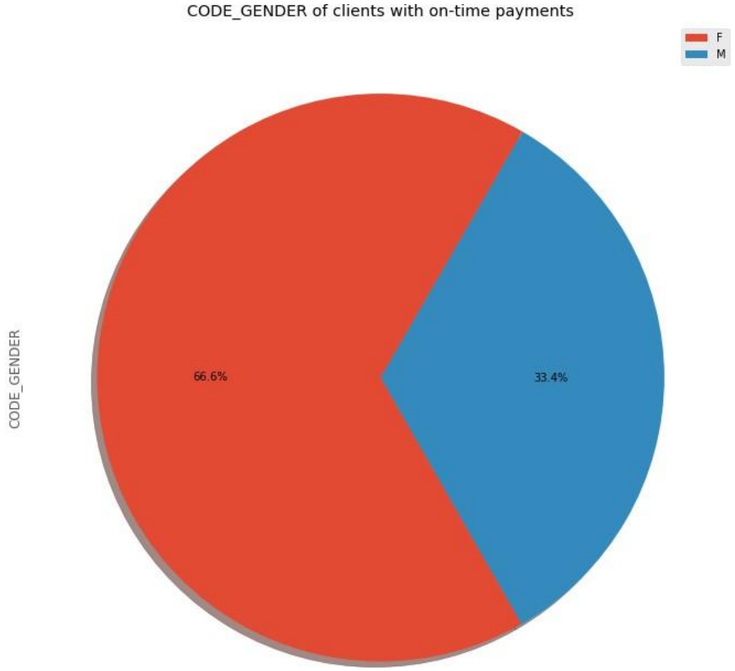


Checking Imbalance for target column 'TARGET'

- **The ratio for people who have difficullty paying to on-time payers is 11.387150050352467**
-
- **1 in every 11 applicant has payment difficulty.**

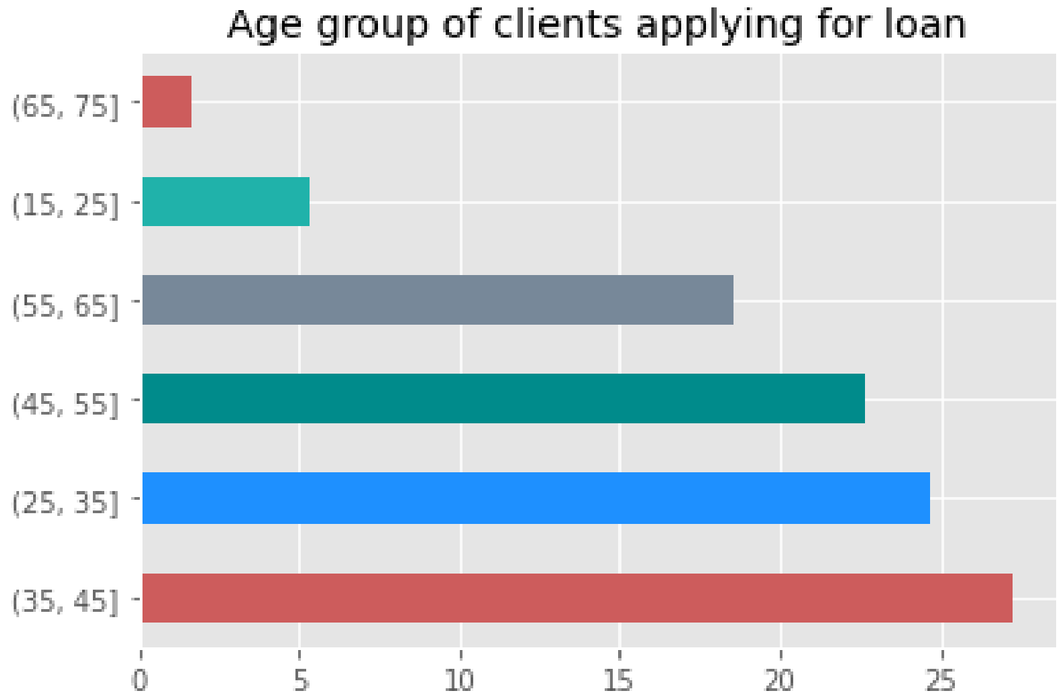


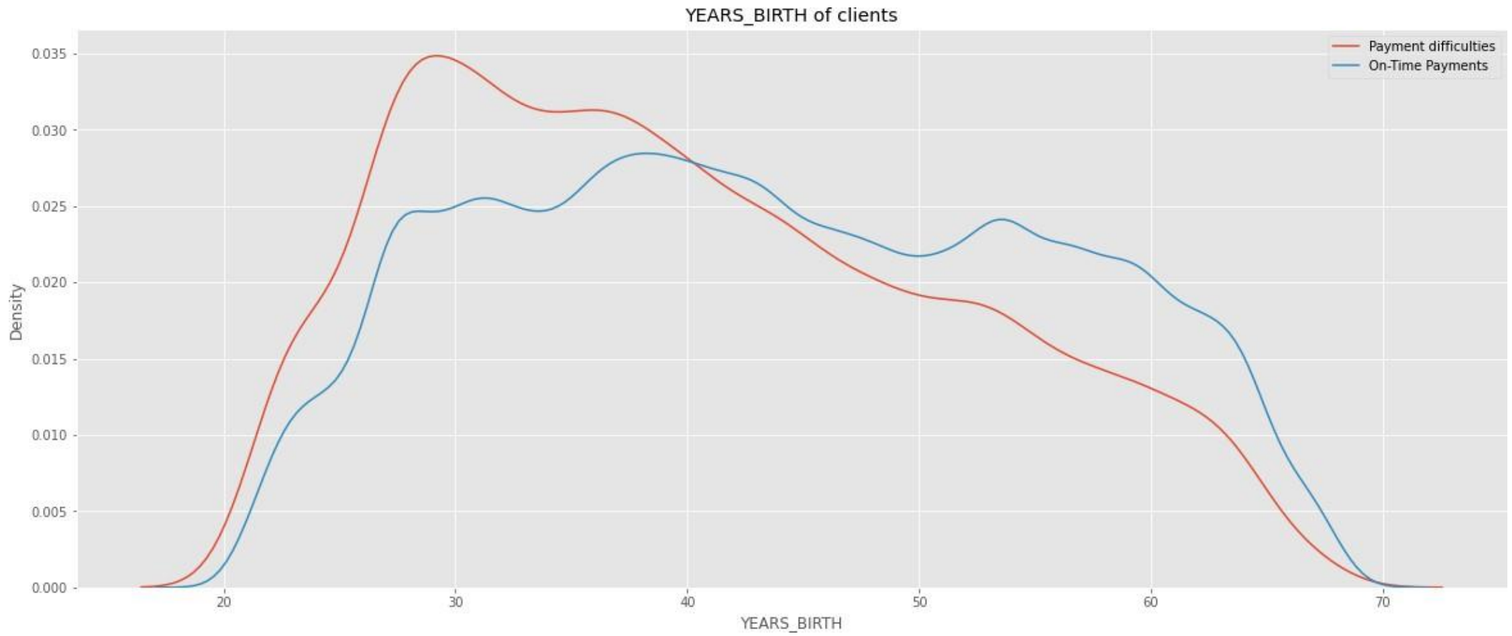




Gender is a factor when determining whether client can pay on-time

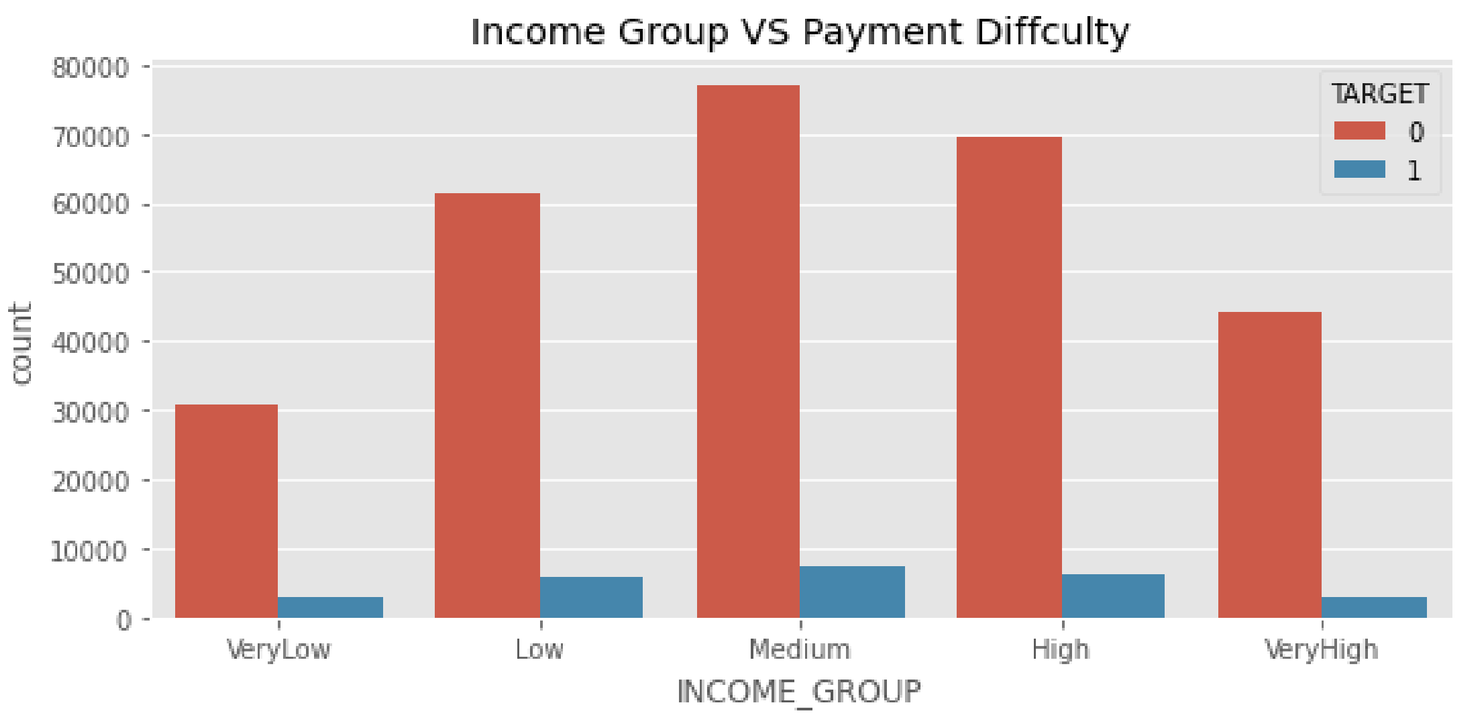
- Females have processed more loans in comparison to males.
- There is a 9.4% decrease in "Male" values from CoDE_GENDER b/w clients with payment difficulties to on-time payments.
- Male applicants are defaulting more that female applicants
- Female clients who are working are more likely to make on-time payment
- Female applicants should be given extra weightage as defaults are lesser.
-

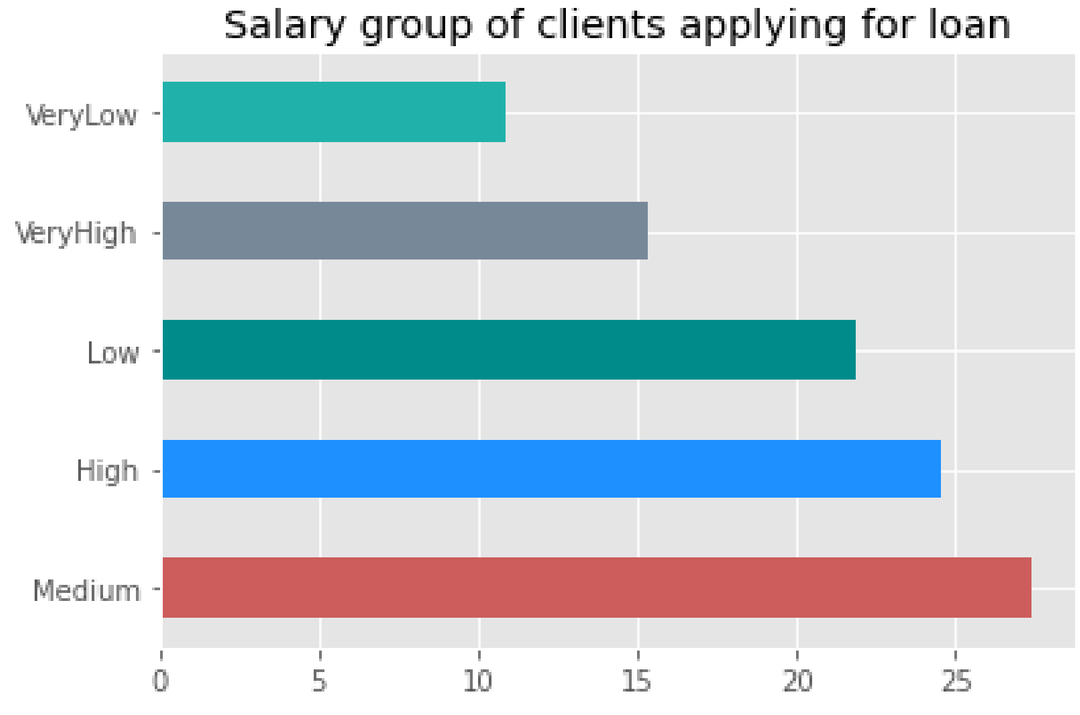




Age is a factor when determining whether client can pay on-time

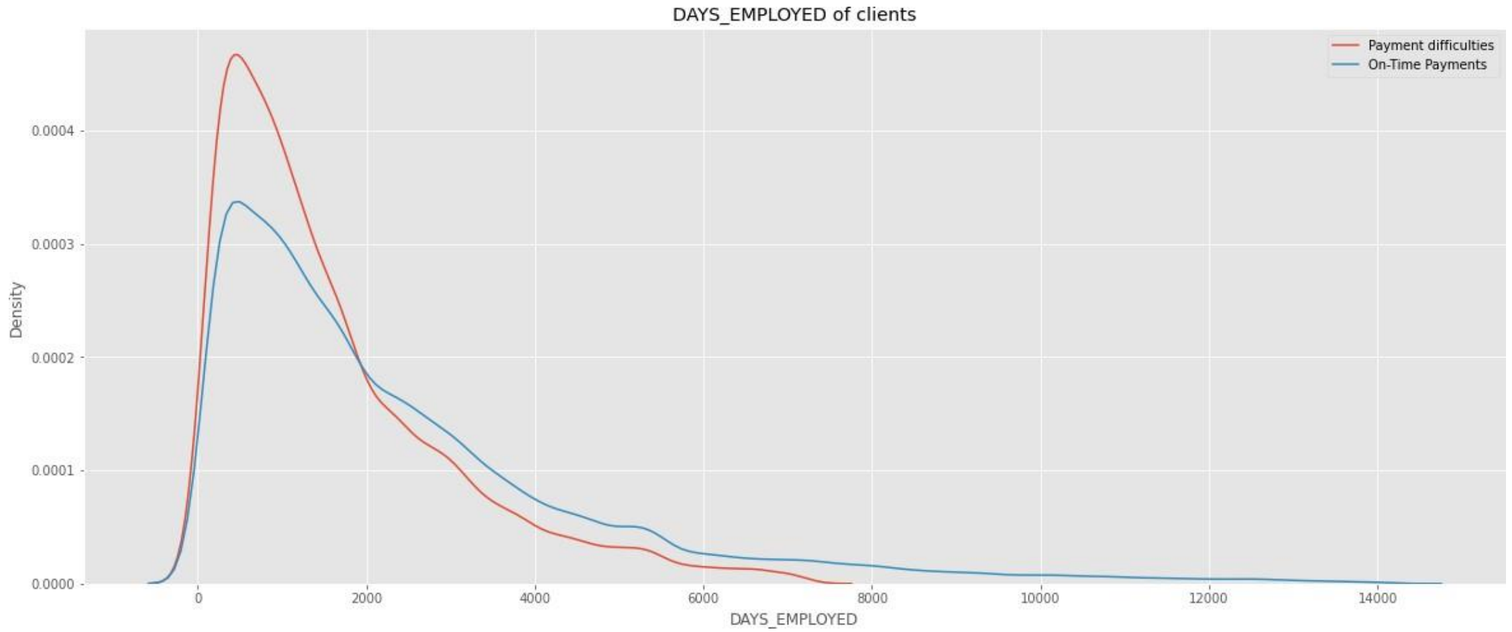
- For YEARS_BIRTH between 20 and 40, there are more clients with Payment difficulties
- For YEARS_BIRTH > 40 , there are more clients with on-Time Payments
- 35-45 Age group is the largest Group of Age applying for loans.
- People having difficulty repaying are in the group 25-35 years olds, followed by 35-45 years age group
- Clients in the age range 30-40 and 40-50 is ideal target for loans

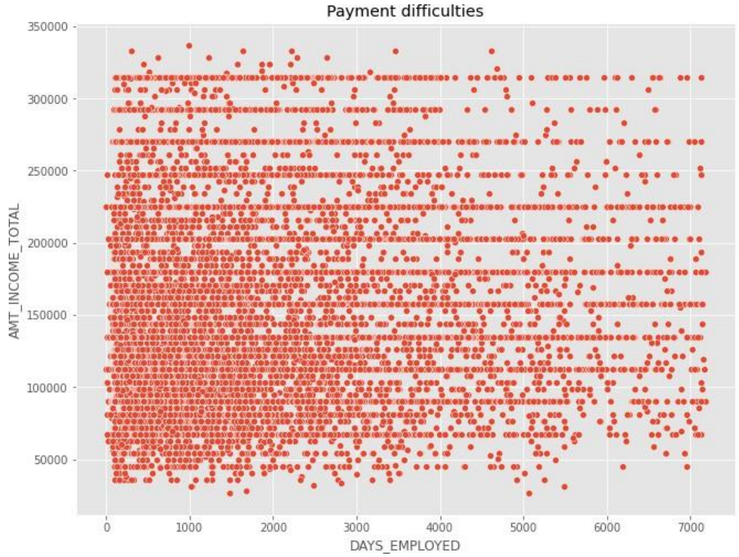


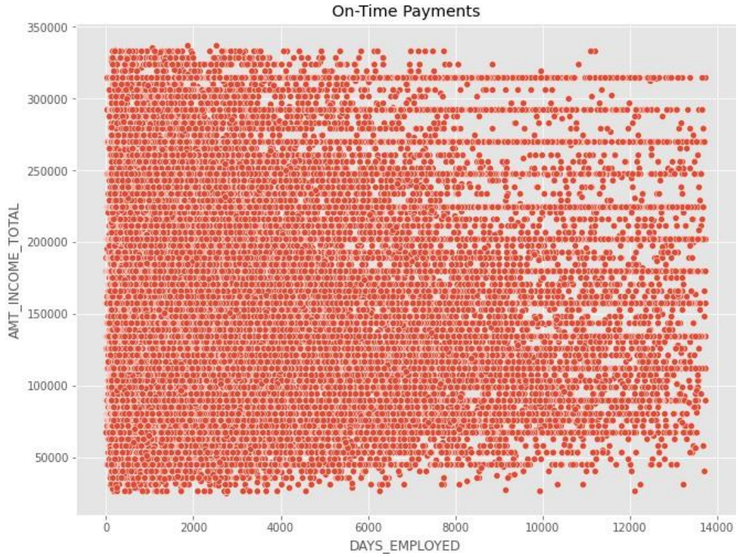


Income is a factor when determining whether client can pay on-time

- **'Medium' Income group is the largest group applying for loans, followed by 'High' income group. 'VeryLow' income group is the smallest group applying for loan.**
- **Medium income groups defaults a lot more compared to other groups**
- **Amount Income 'Low' and 'High' is a concern as well**
- **Default value per loan is highest in High income group as the AMT_CREDIT is higher too.**

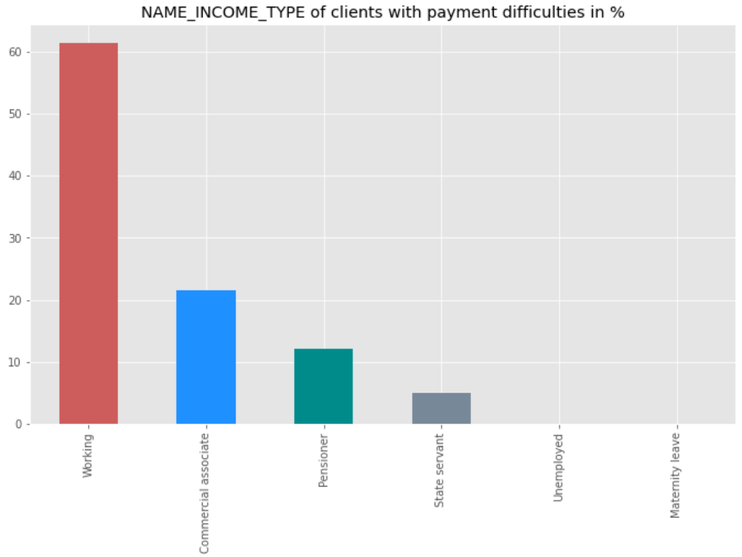


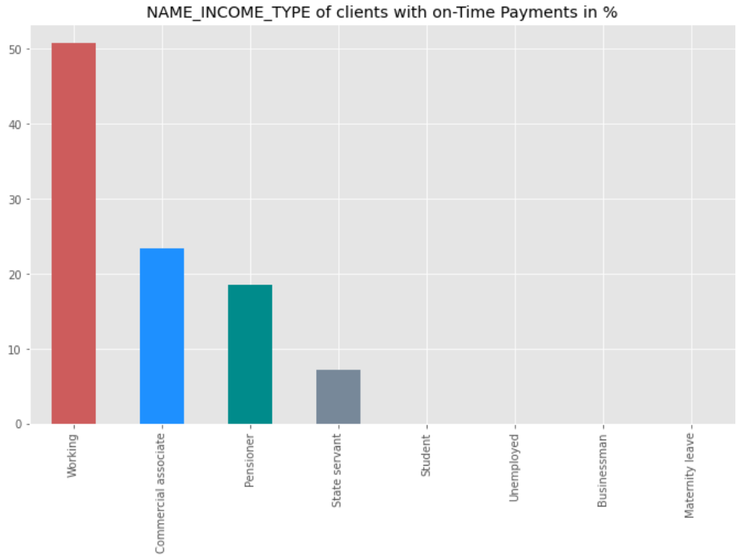


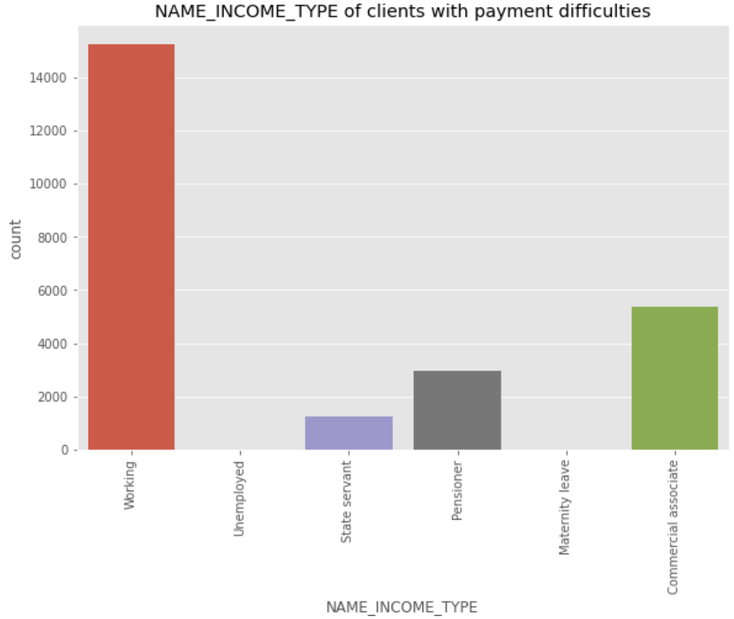


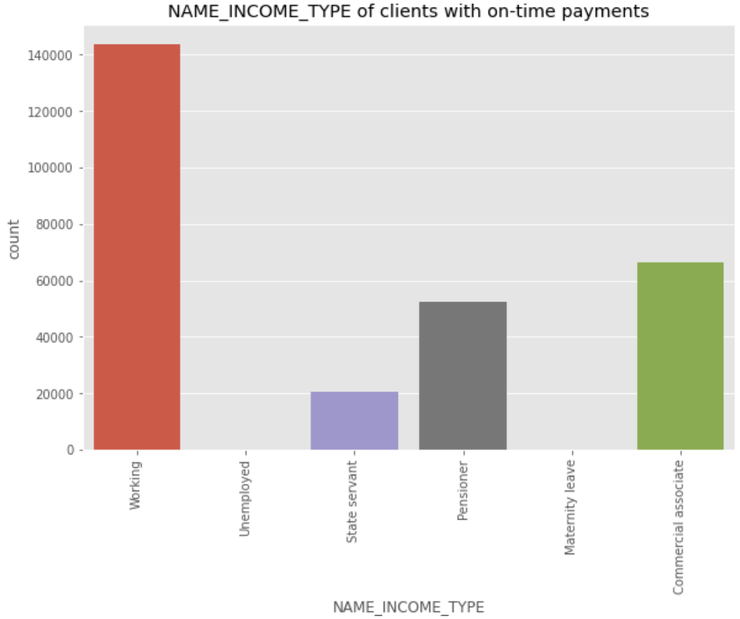
Employment years is a factor when determining whether client can pay on-time

- Clients who are employed for a long time days are making their payments on-time but these category of clients do not exist in Payments difficulties group
- Even looking at Payment difficulties group, clients with more than 4000 days of employment are sparse
- For DAYS_EMPLoYED > 2000 , there are more clients with on-Time Payments, impluing that those who are employed longer have better chances of repaying the loan



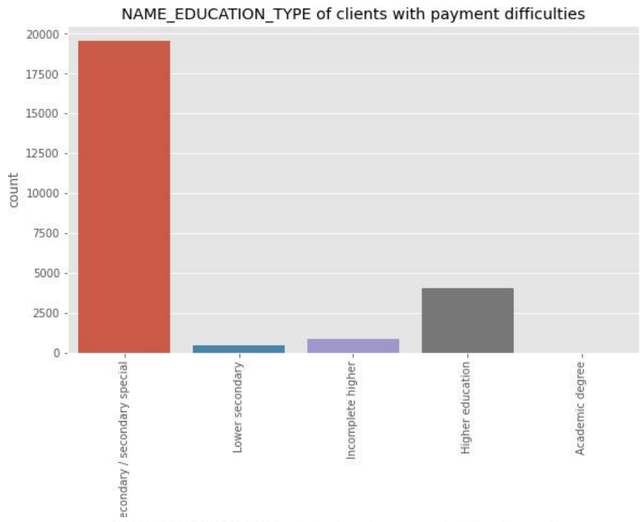


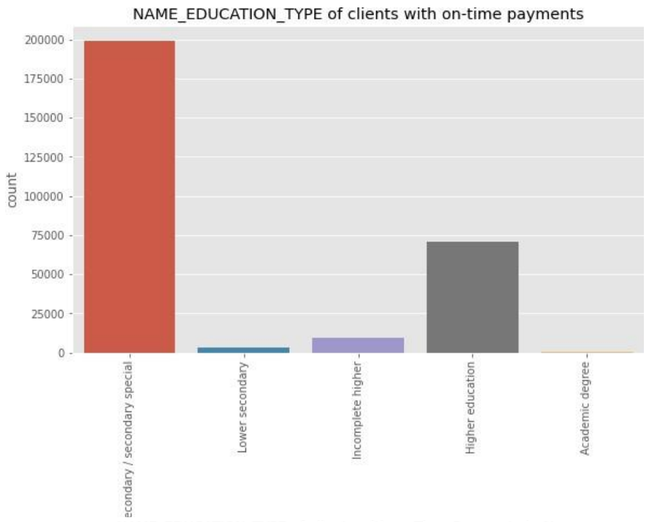


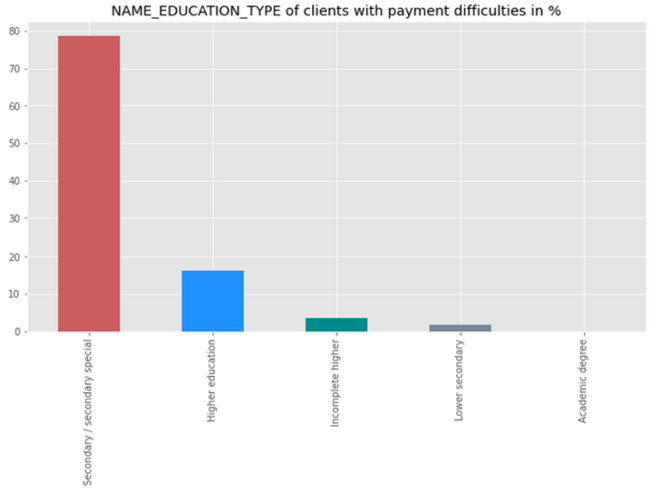


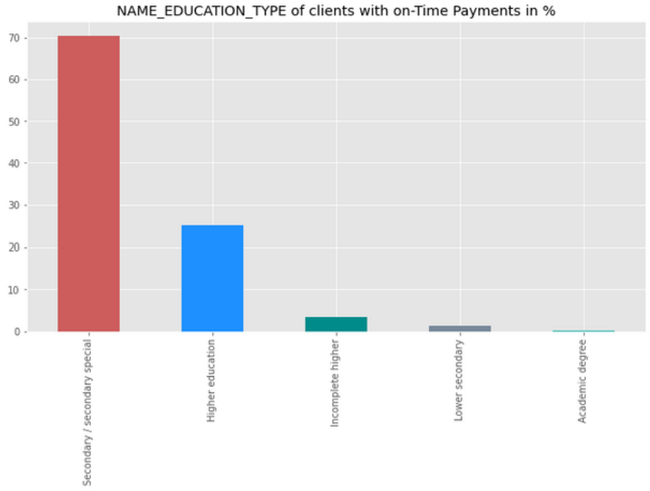
Job type is a factor when determining whether client can pay on-time

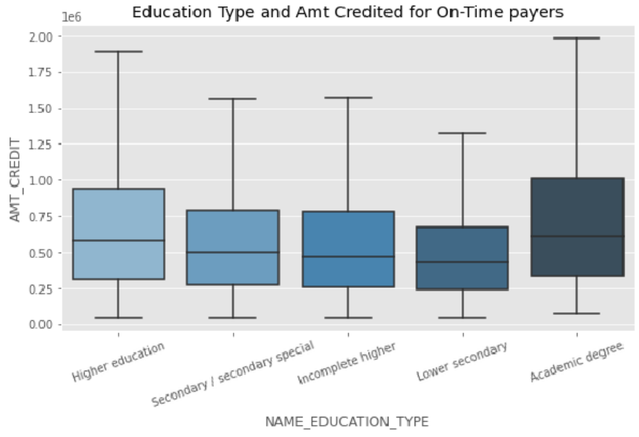
- 'Working' class have processed more loans in comparison to other categories.
- Pensioners have better on-time payments
- Students don't have Payment difficulties
- Businessmen don't have Payment difficulties

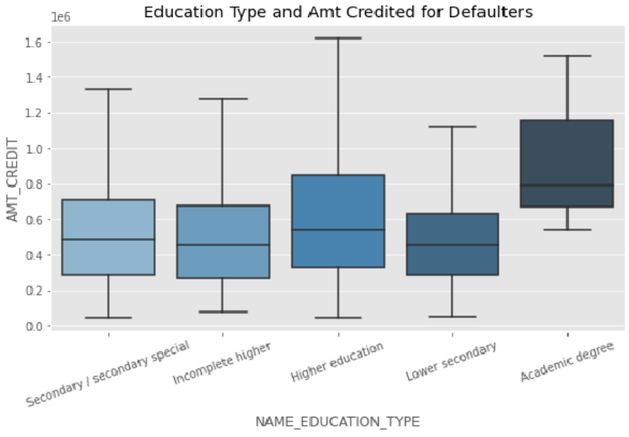






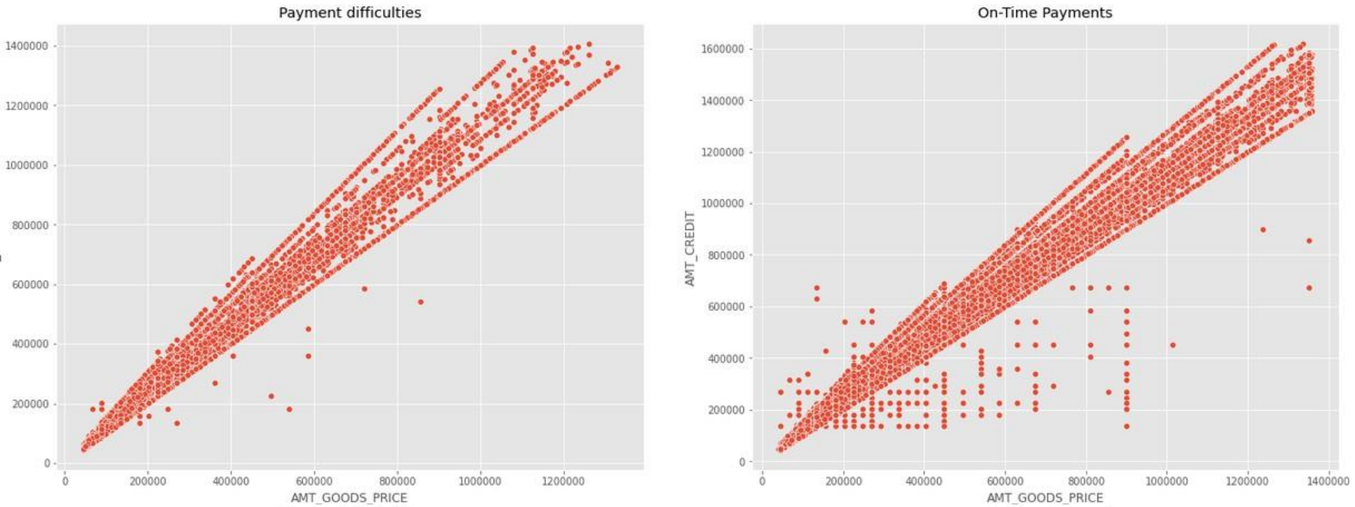


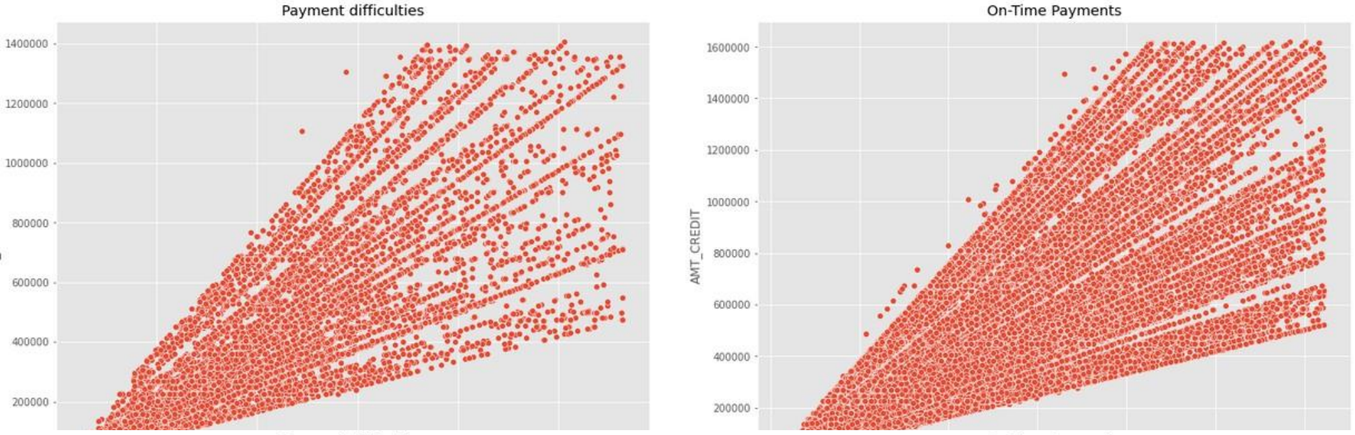


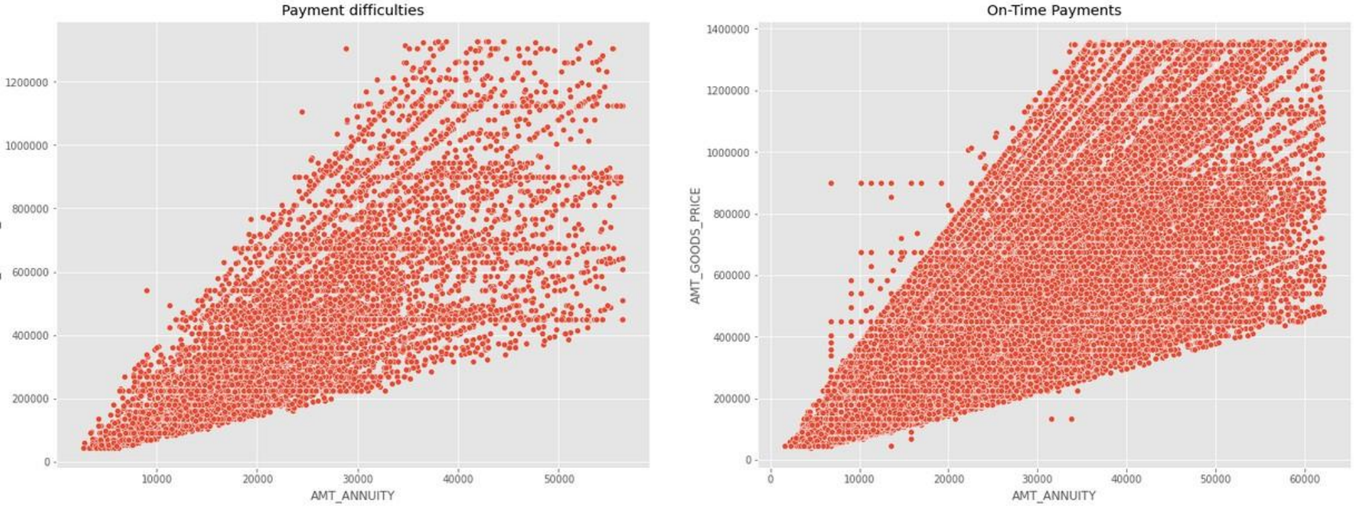


Education level is a factor when determining whether client can pay on-time

- Clients with 'Higher education' have better on-time payments than others and have less payment difficulties.
- Median of Loan values defaulting for Applicants with Academic degree is higher.
- Clients with academic degrees should get higher preference



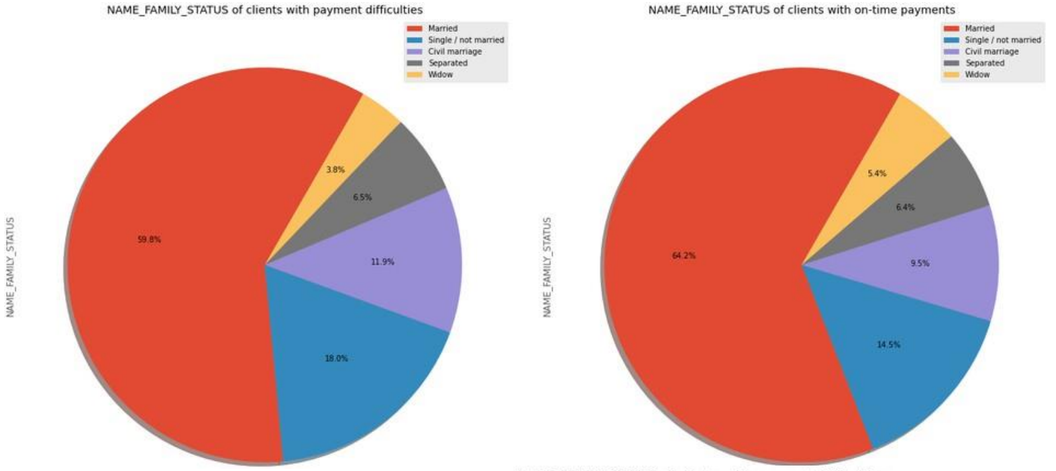


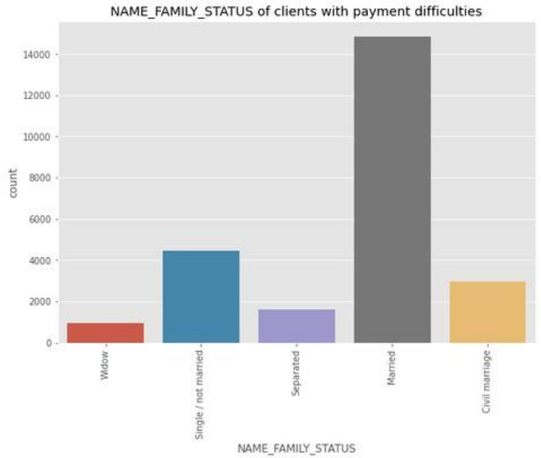


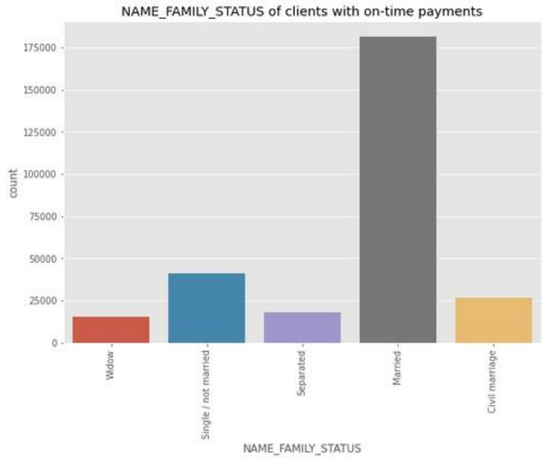
Correlation between different variables

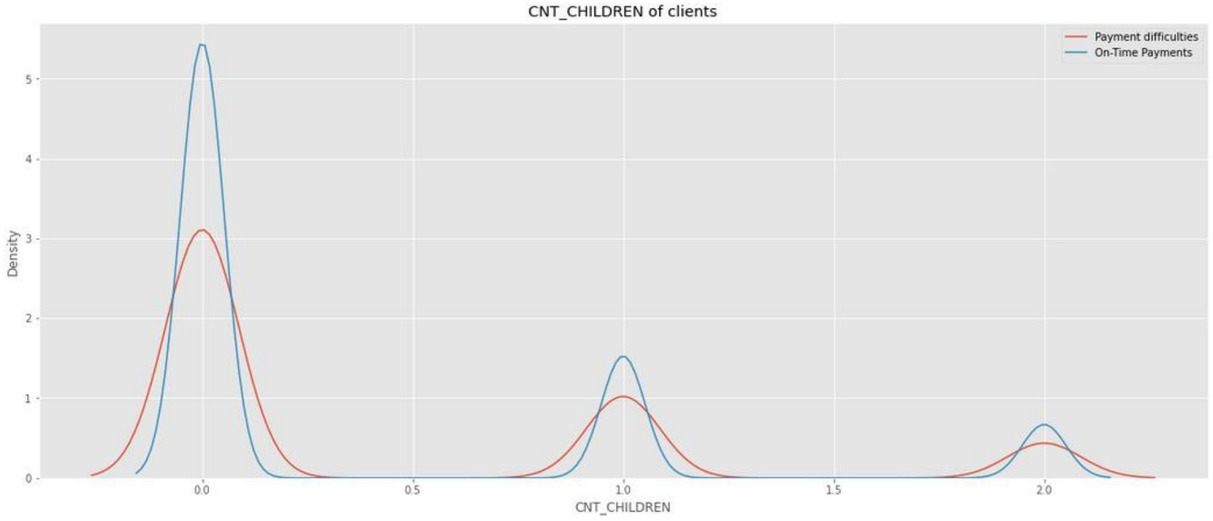
- **There are strong positive correlation found among the following variables, meaning a increase in one will cause increase in other variable**
- **AMT_ANNUITY and AMT_GOODS_PRICE**
- **AMT_ANNUITY and AMT_CREDIT**
- **DAYS_EMPLOYED and AMT_INCOME_TOTAL**
- **AMT_GOODS_PRICE and AMT_CREDIT**
- **CNT_FAM_MEMBERS and CNT_CHILDREN**

1. **DAYS_BIRTH and DAYS_EMPLOYED**



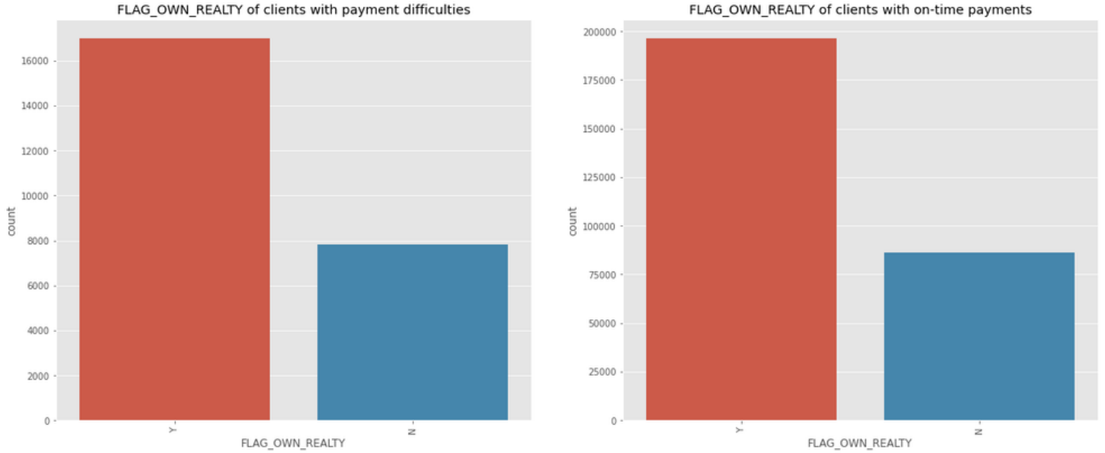


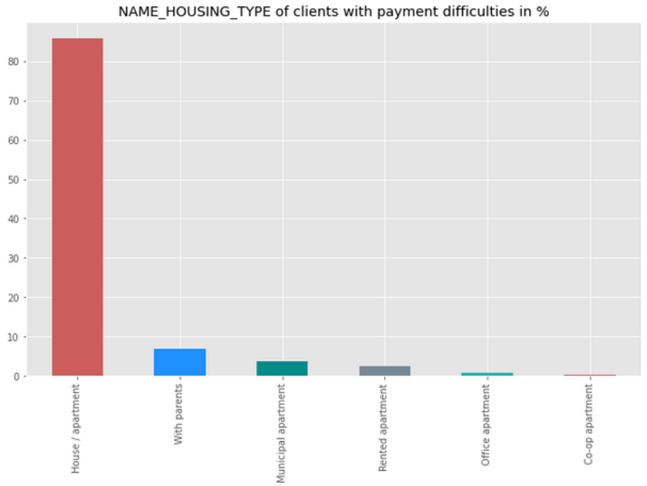


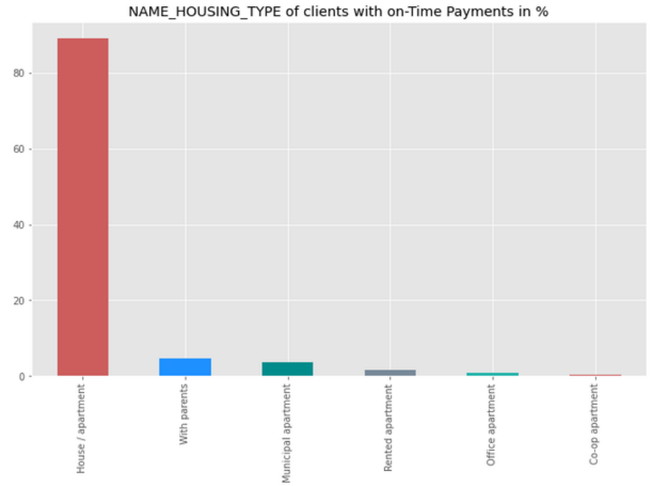


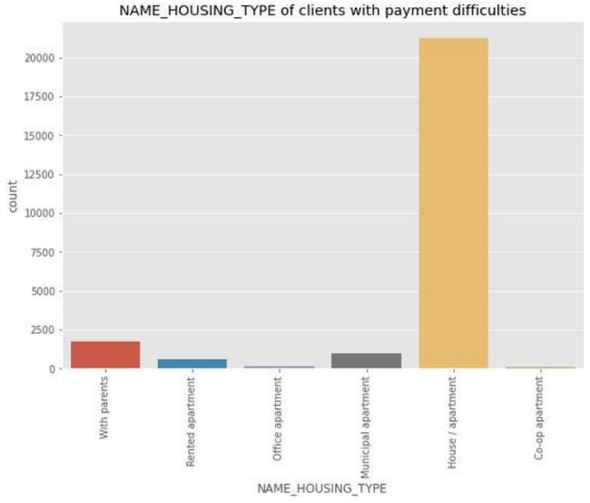
Family status is a factor when determining whether client can pay on-time

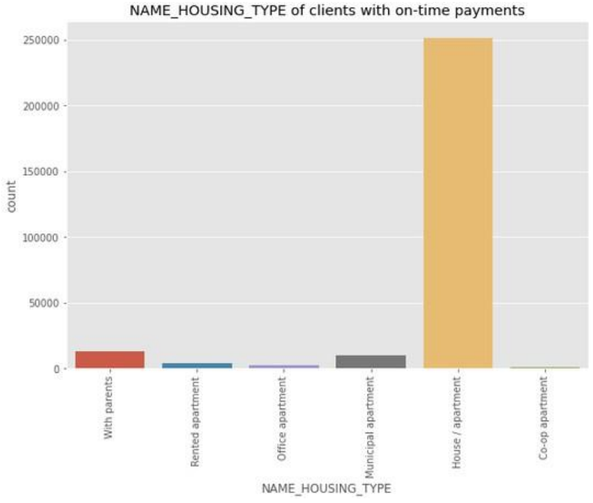
- People with 0 children are more prone to make on-time payments
- Clients who are 'Married' oR 'Widow' do on-time payments better comparatively.
- Clients who are 'Single/not married' have more difficulties with on-time payments comparatively.
- Married people are the ones mostly applying for loan followed by Single or Not married
- Widows are the ones who have less number of application for a loan







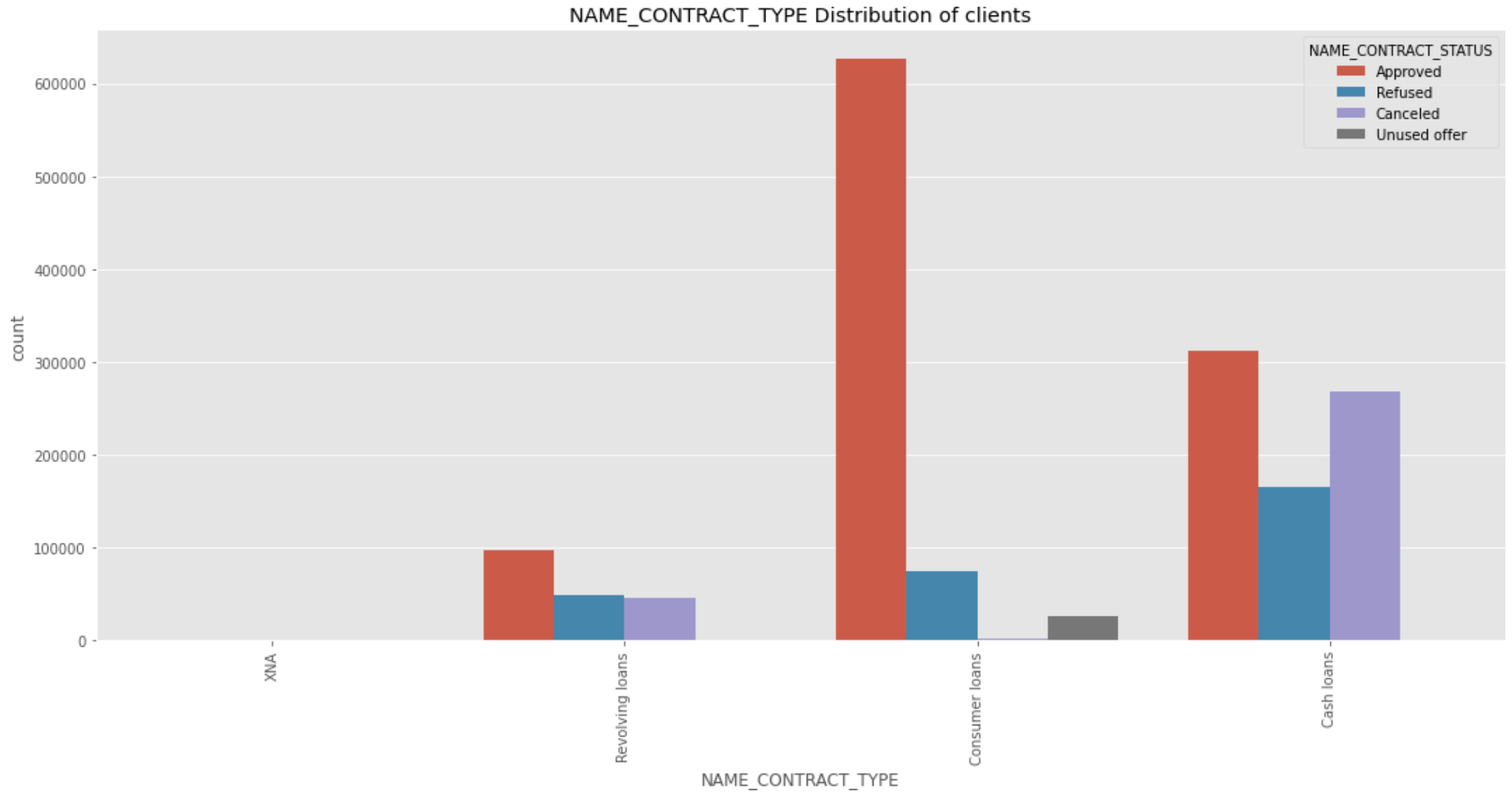




Housing type is a factor when determining whether client can pay on-time

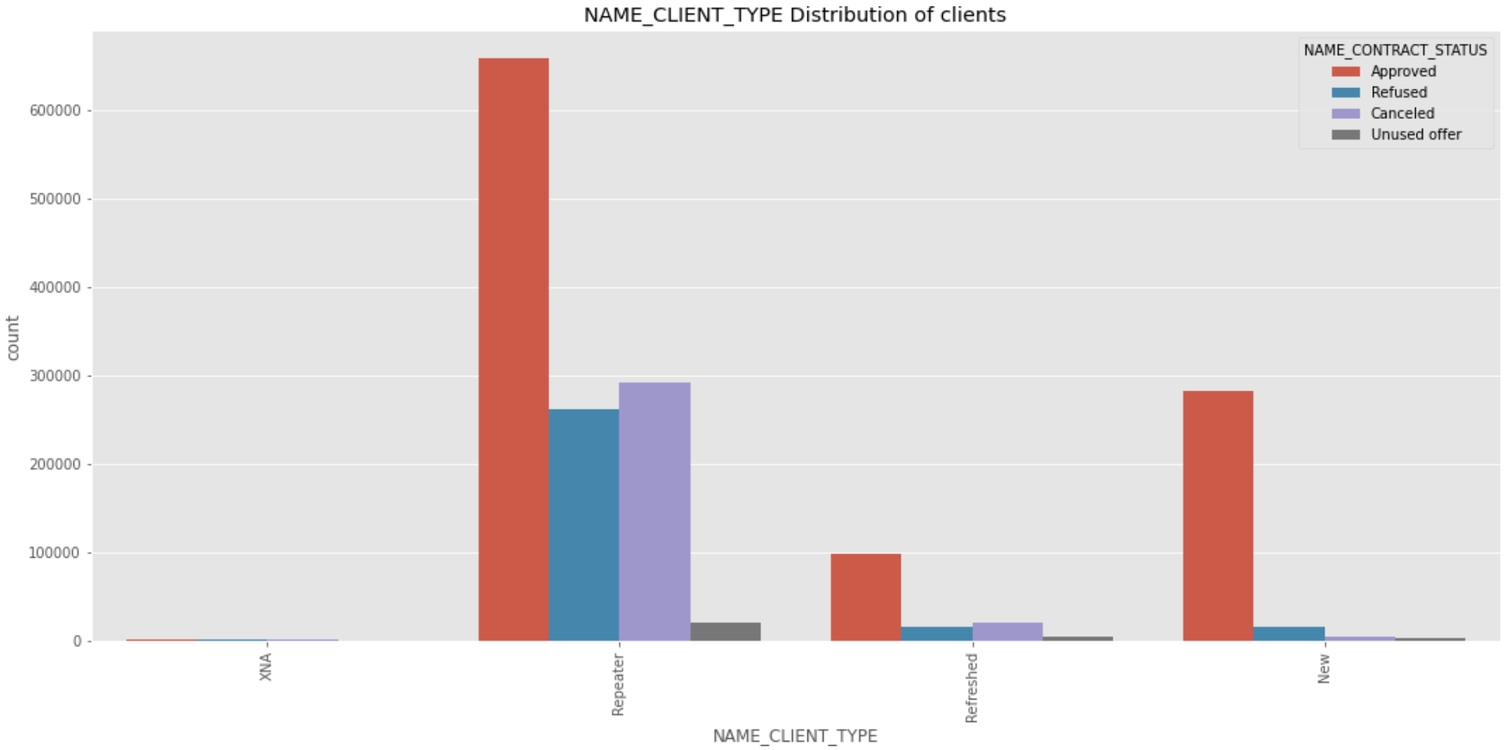
- Applicants who own a house have processed more loans in comparison to those who don't
- People who own house have better on-time repayment record
- People who live in rented flat or with parents tends to default more on loan payments than people living in House/Flat or office flats.

Types of loans requested by customer

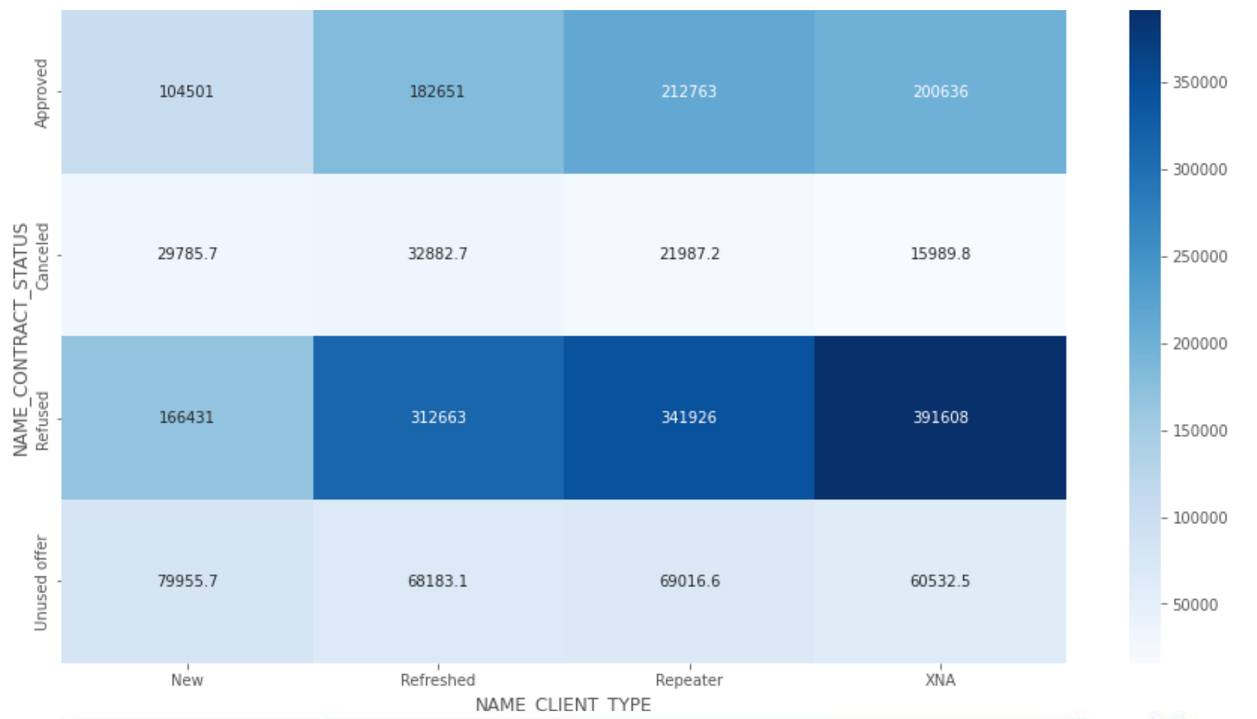


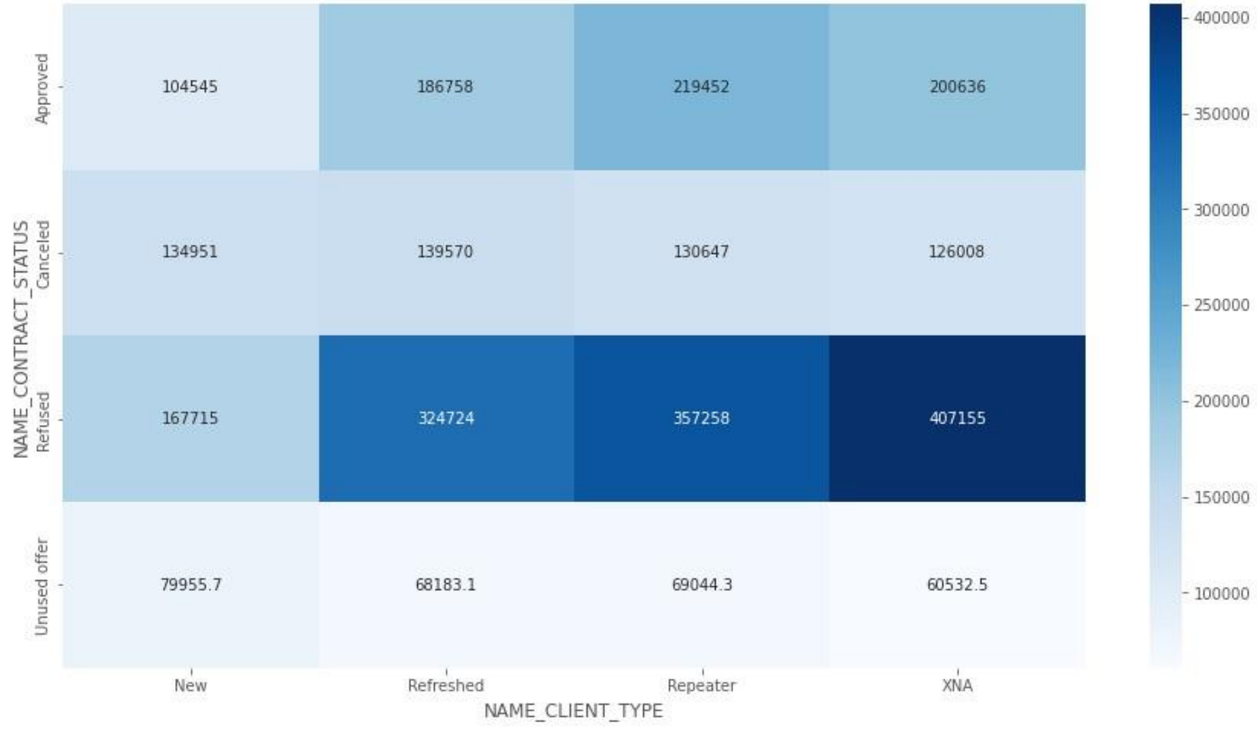
- More Cash loans have been processed compared to revolving loans
- The most approved type of loan is consumer loan, followed by cash loan
- Revolving loan has a higher refusal rate
- There seem to be no cancelled loans in cash loan category than consumer loan.
- More cash loans have been refused than consumer loans.
-

Types of clients requesting for a loan



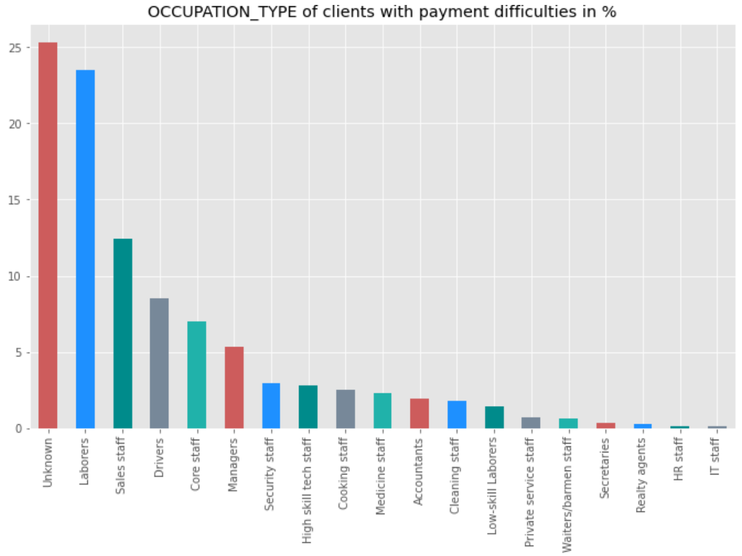
- The company has more repeaters in all approved, refused, unused, cancelled categories
- PoS transactions seem to be more for repeaters, and more loans have been refused for repeaters than other group.
- Client cancelling the loan application is high in the case of repeaters

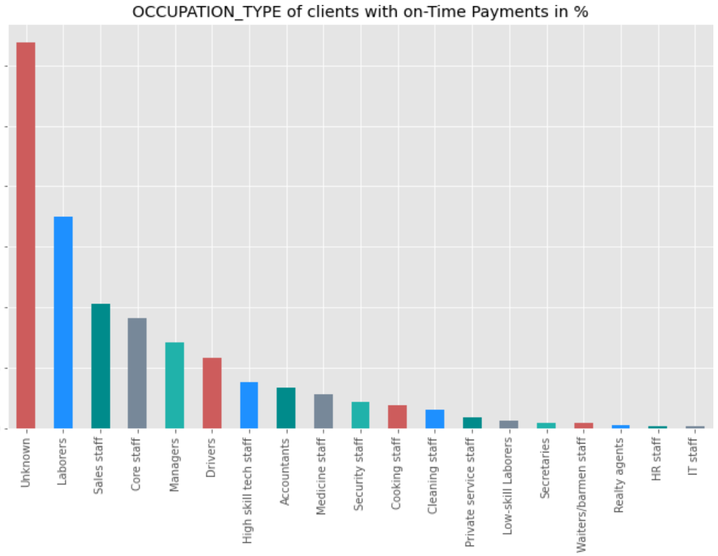


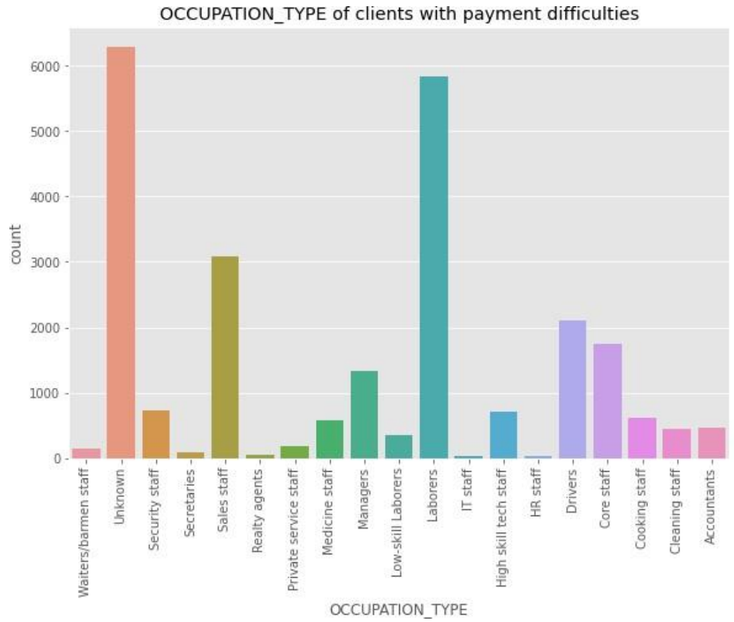


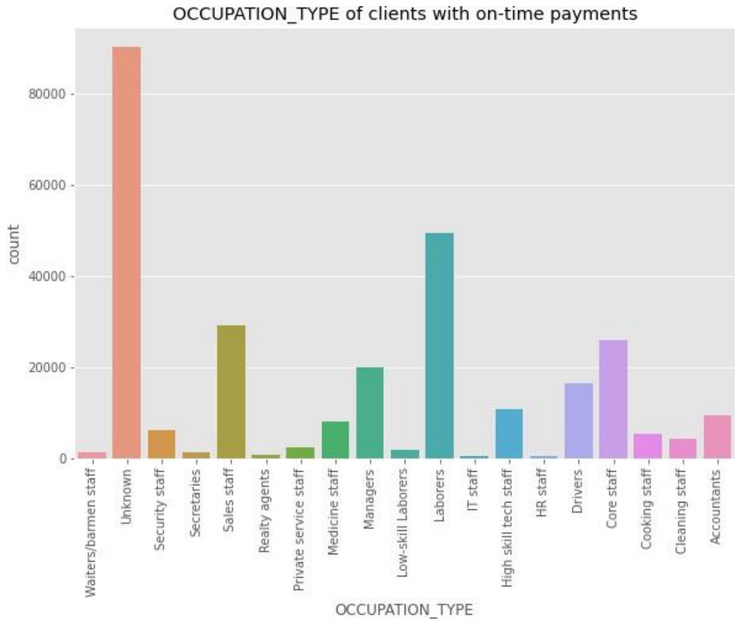
Reason for repeaters cancelling loan

- **Unused offer CREDIT AMoUNT is low. This may be the reason for customer not using**
- **Need to investigate why there is credit amount for cancelled and refused loan**
- **All cancelled and refused cases have higher value of goods than other categories**



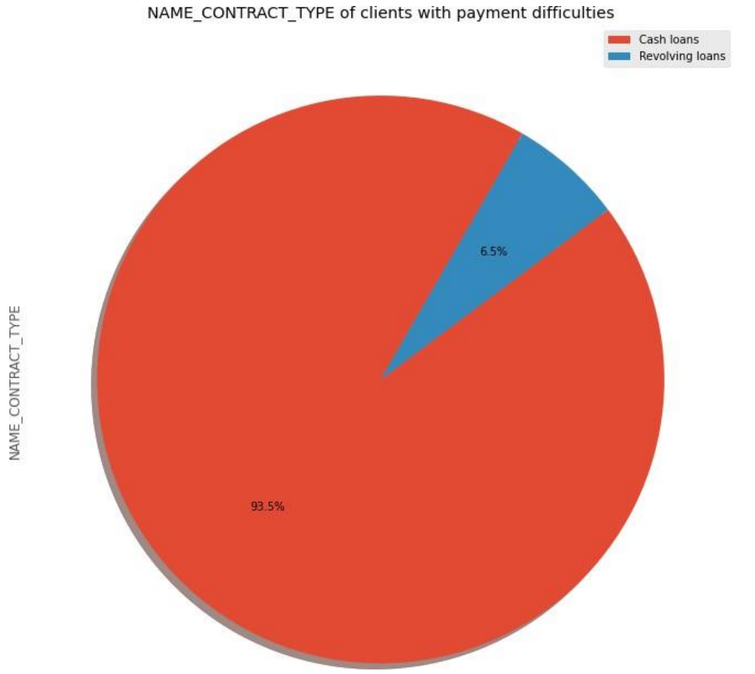


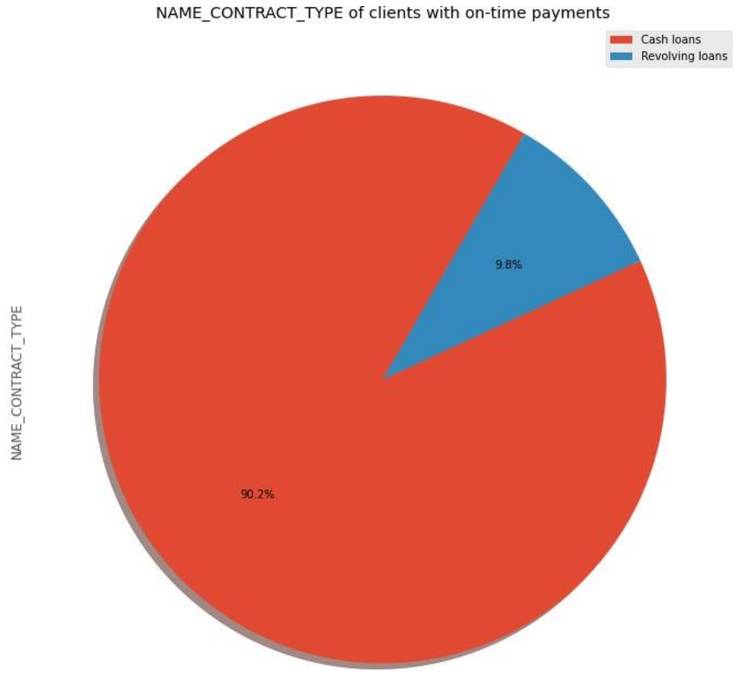


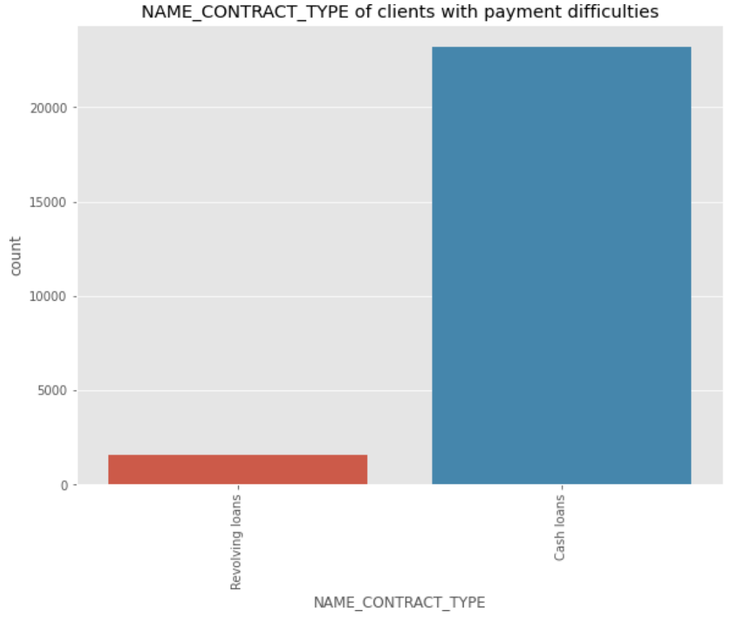


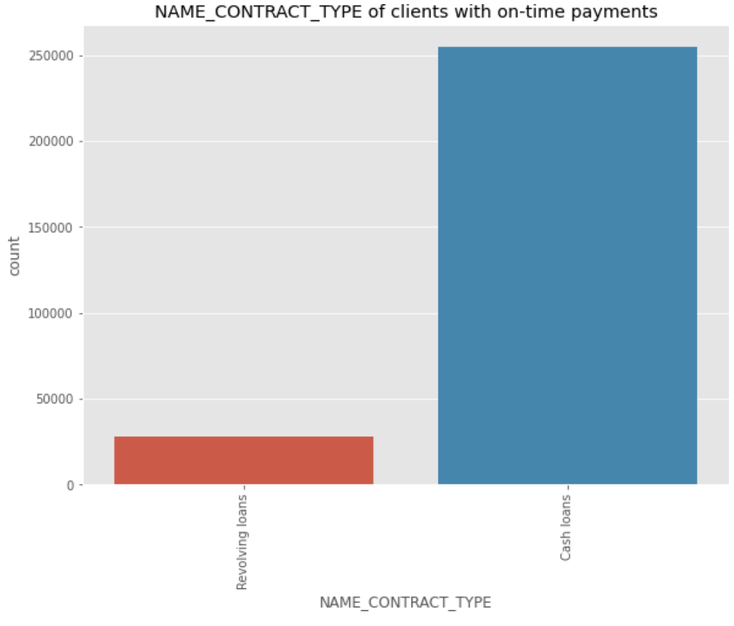
Client Occupation type

- If we talk about the occupation of the people applying for loan, labourers are on top followed by sales staff and core staff.
- Maximum percent of people defaulting on payments are from the occupations :Low skilled labourers, Drivers, waiters/barmen staff
- Accountants, High skill tech staff and manager are veryless likely to default.









Types of loan disbursed

- More people take cash loans then revolving loans
- More people default on cash loan then revolving loans

Conclusion from the case study

People having payment difficulty are from the following group:

- **Medium income**
- **25-35 years old age group, followed by 35-45 years age group**
- **Male**
- **Unemployed**
- **Labourers, Salesman, Drivers**
- **Don't own house, stay with parents, rented flats**
- **Have medium or high income**

People to focus on:

- **Previously cancelled, refused, unused loan cases**
- **Students & Businessmen have no problem in repayment of the loan**
- **Repeater clients**
- **Academic degree clients**
- **Clients who are employed for more than 19 years**
- **Clients who are Married**
- **Clients who own a house**
- **Pentioners**
- **Female applicants should be given extra weightage as defaults are lesser.**