CREDIT EDA CASE STUDY

BY BIDHAN CHANDRA ROY

PROBLEM STATEMENT

Urban customers are the target market for a consumer finance company's many forms of loans.When a loan application is received, the company must decide whether to approve the loan based on the applicant's profile.

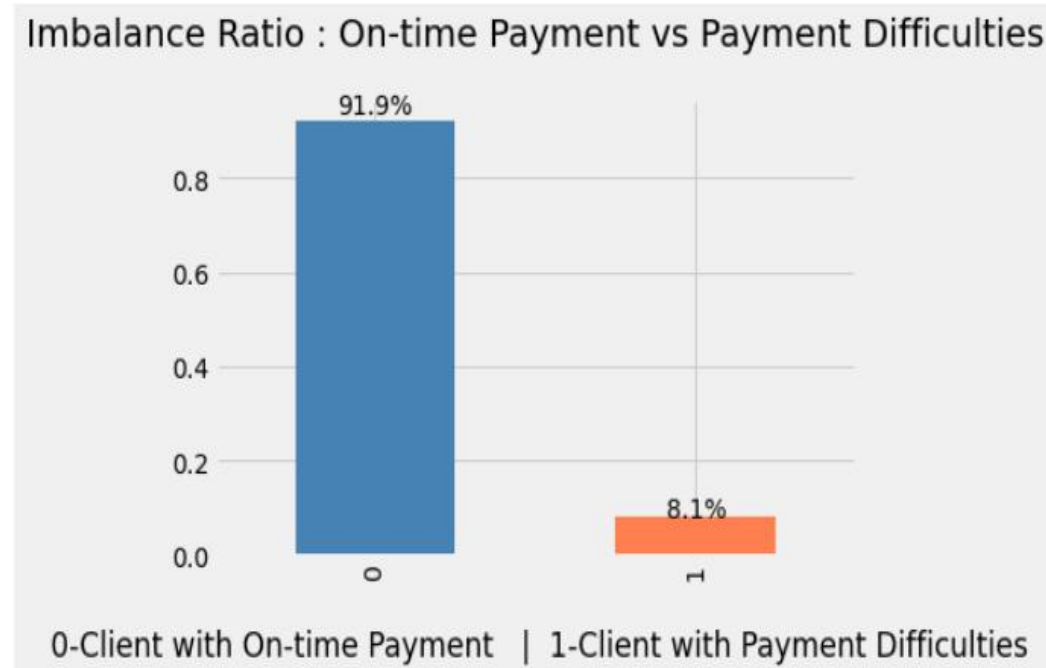The bank's choice is subject to two different kinds of risks:

1. If the borrower is likely to repay the loan, refusing to grant it results in the company losing business.

2. If the borrower is not likely to pay back the loan, or is likely to default, authorising the loan could result in a loss for the business.

The business needs to comprehend the elements that are reliable predictors of loan default, sometimes known as the driving factors (or driver variables) behind default. This information can be used by the business for portfolio management and risk analysis.

STRATEGY FOR ANALISYS

- Ploblem Statement Understanding
- Data Loading and Understanding
- Checking the Quality of data and Missing Values:
  - Inspecting Missing Values
  - Handling Missing Values
  - Missing Value Imputation
  - Checking Data Correctness and Conversion in right Format
- Data Analysis:
  - Data Imbalance Analysis
  - Univariate and Segmented Univariate Analysis
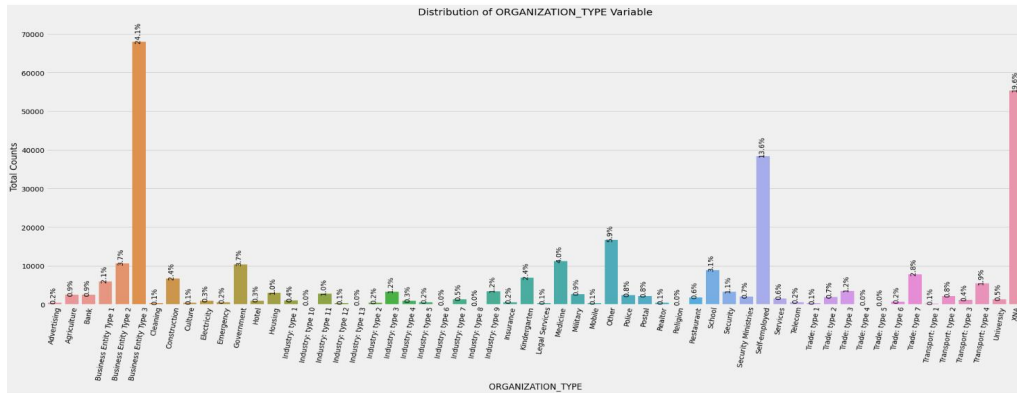 - Bivariate or Multivariate analysis

DATA IMBALANCE

Imbalance Ratio : On-time Payment vs Payment Difficulties

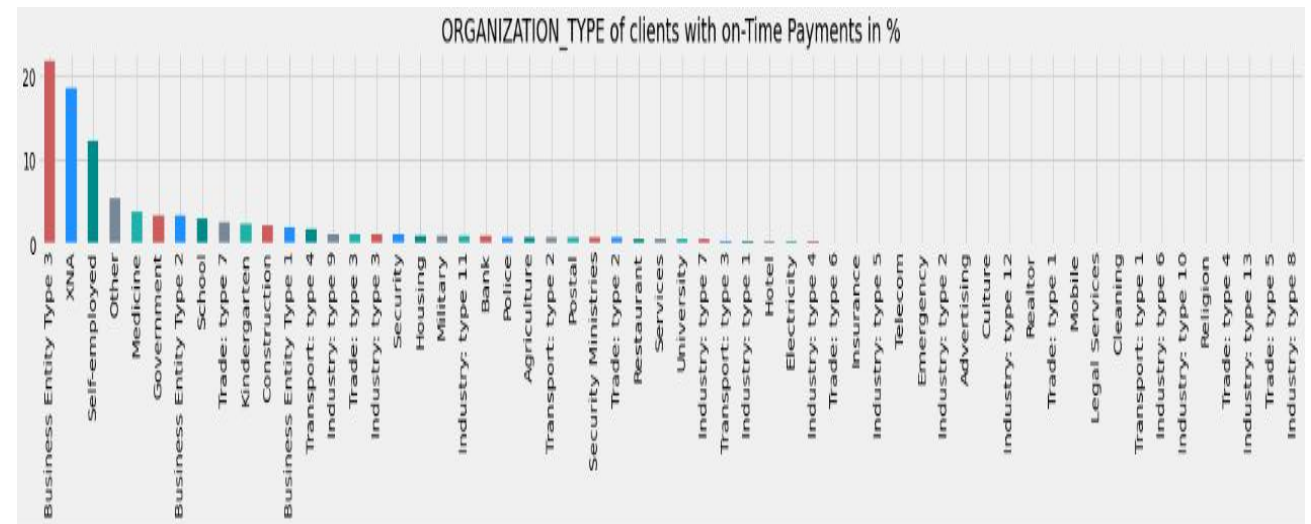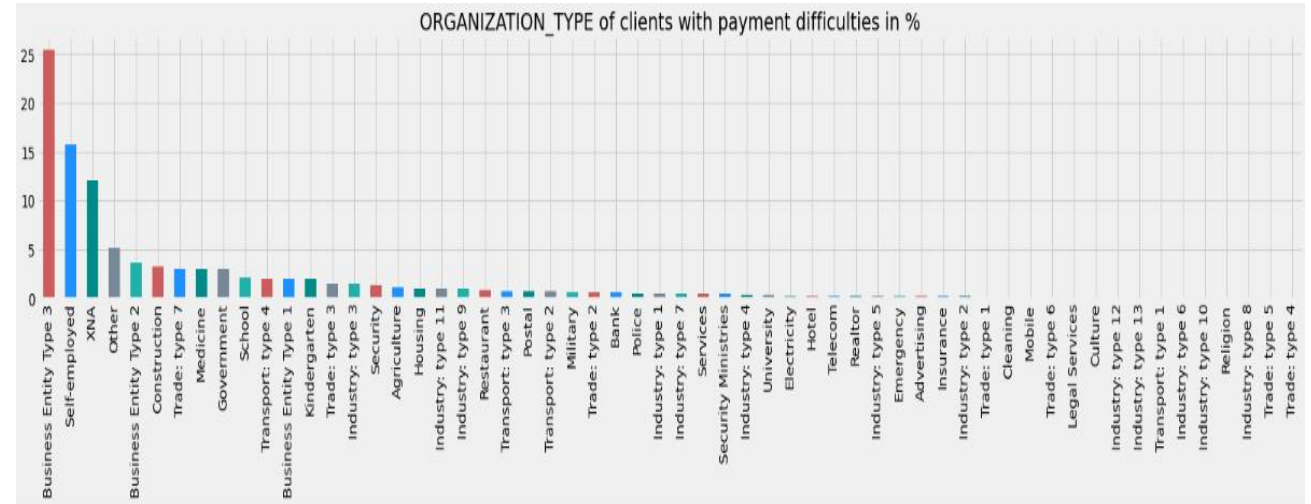0-Client with On-time Payment | 1-Client with Payment Difficulties

- The percentage of non-defaulters and defaulters are 91.9 and 8.1 respectively.
- The imbalance ratio of non-defaulters vs defaulters is 11.38 : 1 that means there are one defaulters of every eleven non-defaulters.
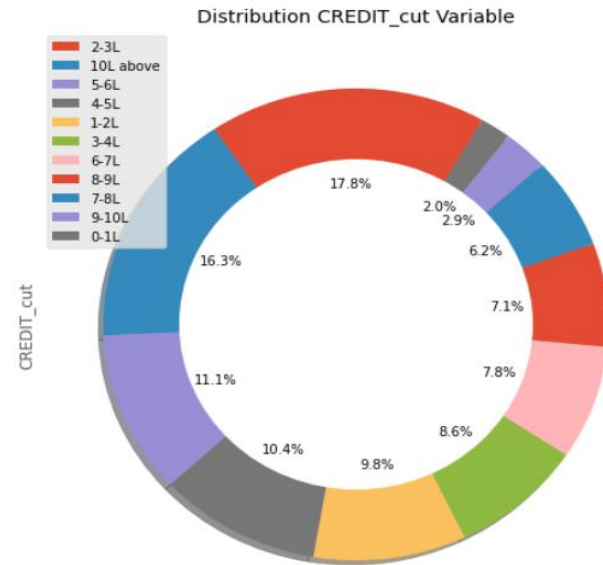
UNIVARITE AND SEGMENTED UNIVARIATE ANALYSIS

ORGANISATION TYPE

Distribution of ORGANIZATION_TYPE Variable

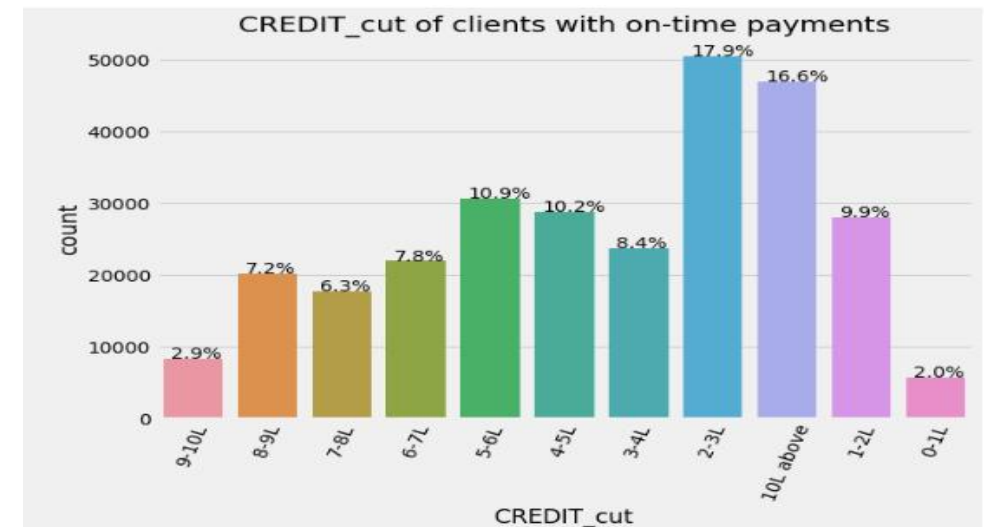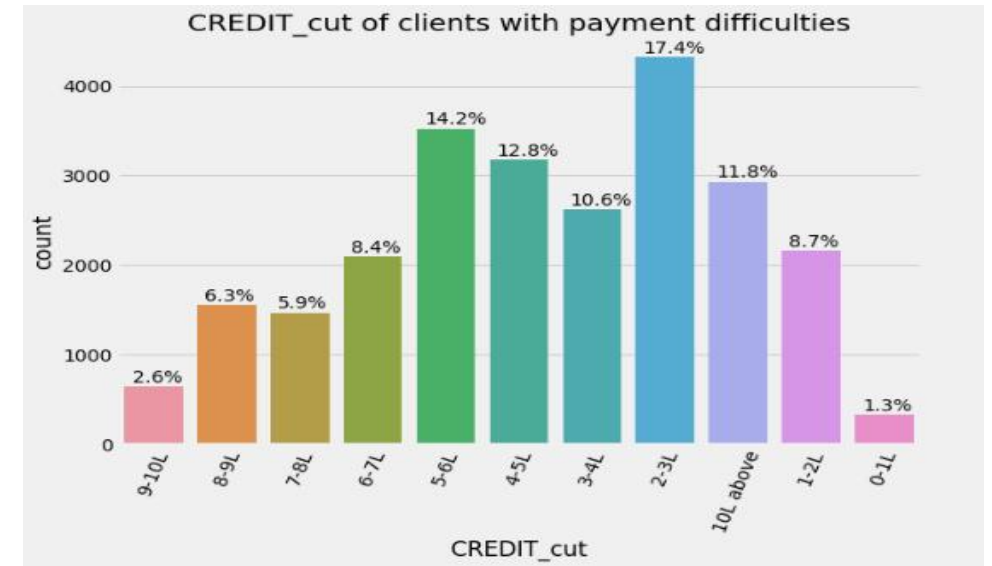ORGANIZATION_TYPE of clients with payment difficulties in %

- Business Entity and self-employed applicants are in great number, according to ORGANIZATION TYPE.

- The self-employed and business entities of types 3 have the most payment issues here.
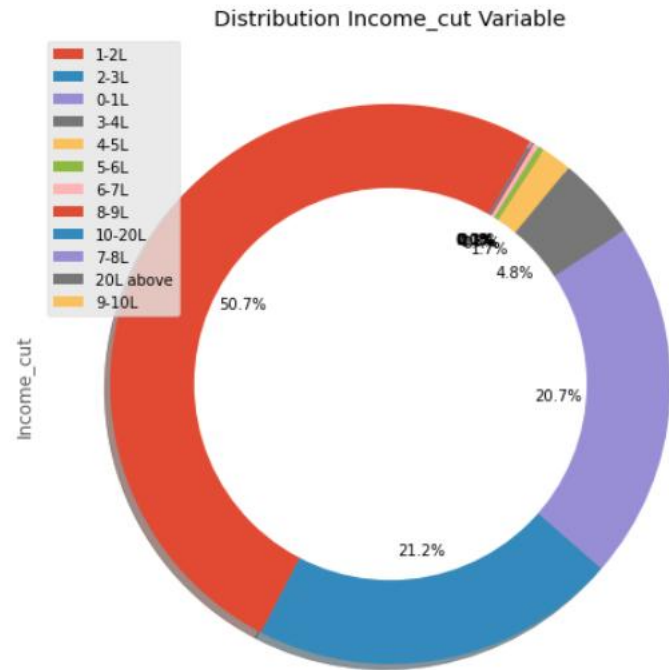

ORGANIZATION_TYPE of clients with on-Time Payments in %

CREDIT AMOUNT

Distribution CREDIT_cut Variable

CREDIT_cut of clients with payment difficulties

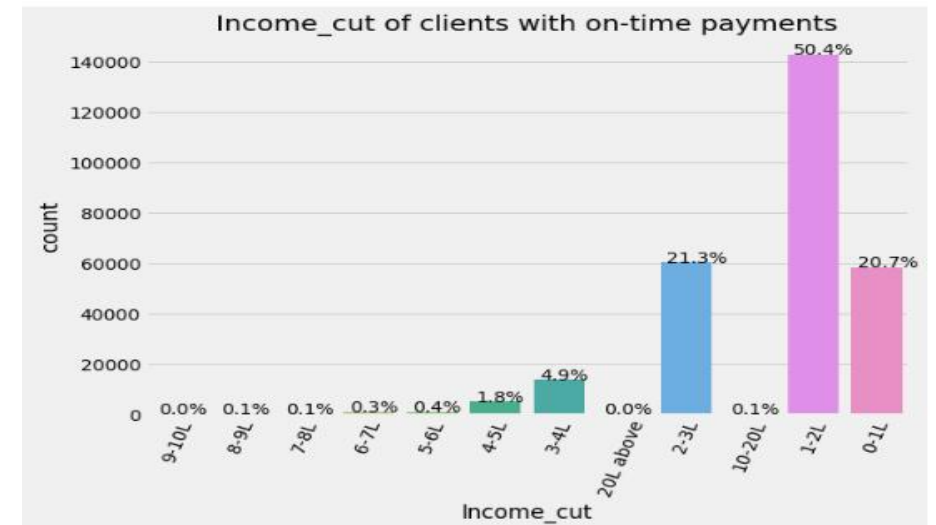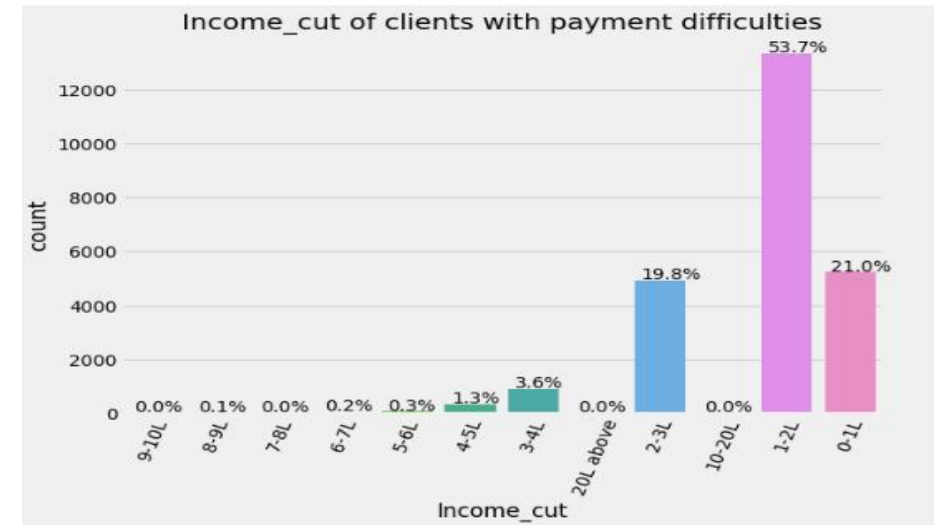CREDIT_cut of clients with on-time payments

- In this instance, Two peak . Client applied loan for 2-3 lacs and above 10 lacs loan are very high.

- Clients who borrowed between 2 to 3 lacs rupees, 4 to 6 lacs rupees, and 10 lacs rupees or more have a harder time repaying the loan.
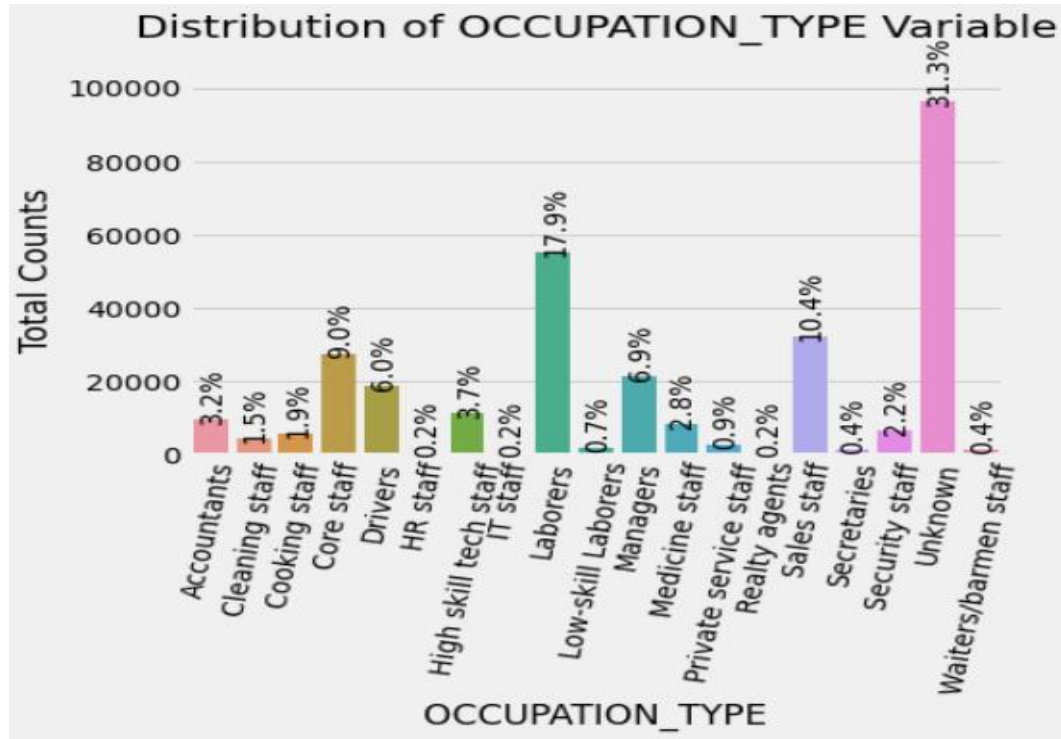
INCOME RANGE

Distribution Income_cut Variable

Income_cut of clients with payment difficulties
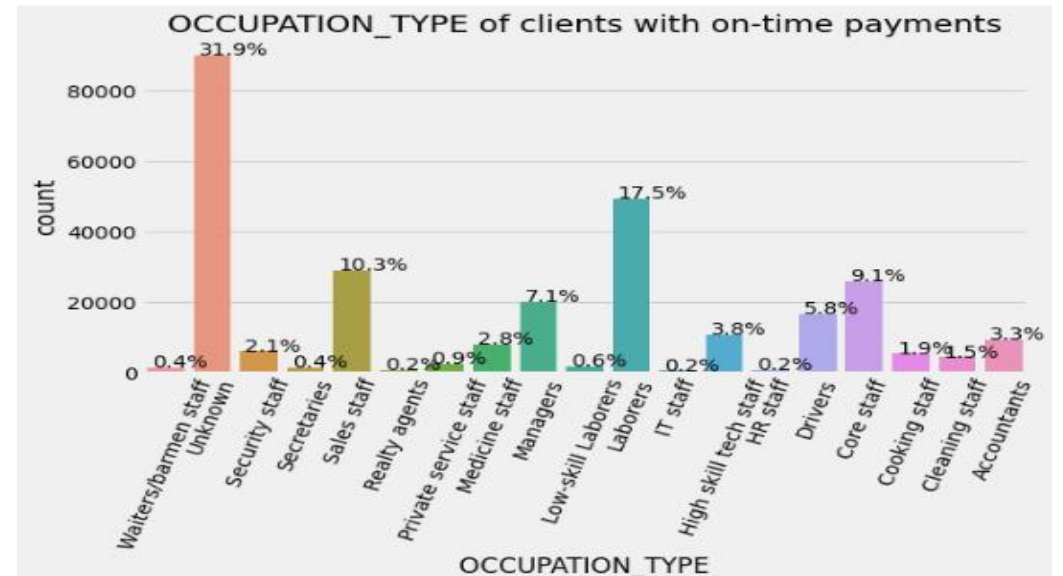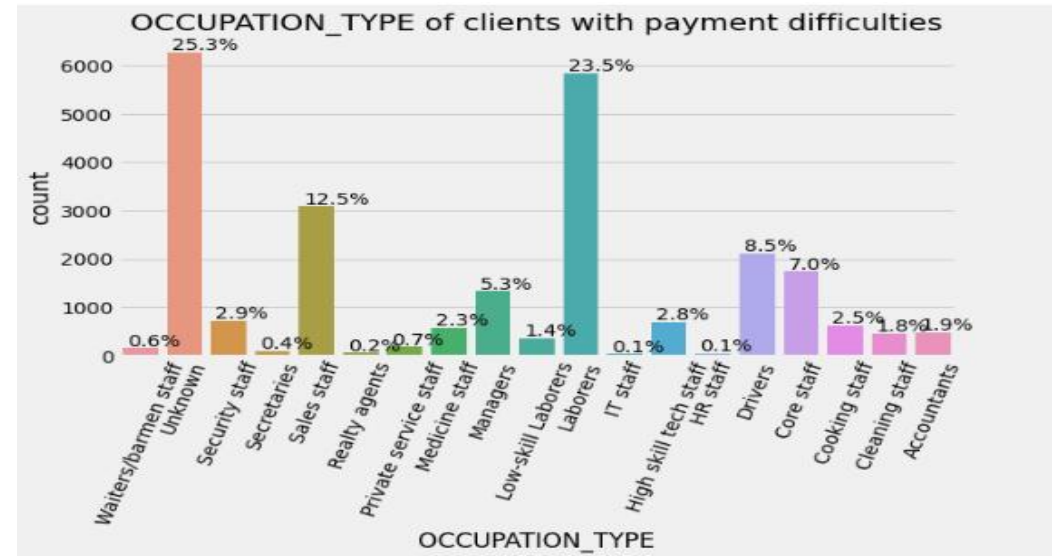

Income_cut of clients with on-time payments

- Of those who applied for loans, 90% had incomes of less 3 lacs rupees.

- Customers that make over 7 lacs are far more likely to make their payments on time.
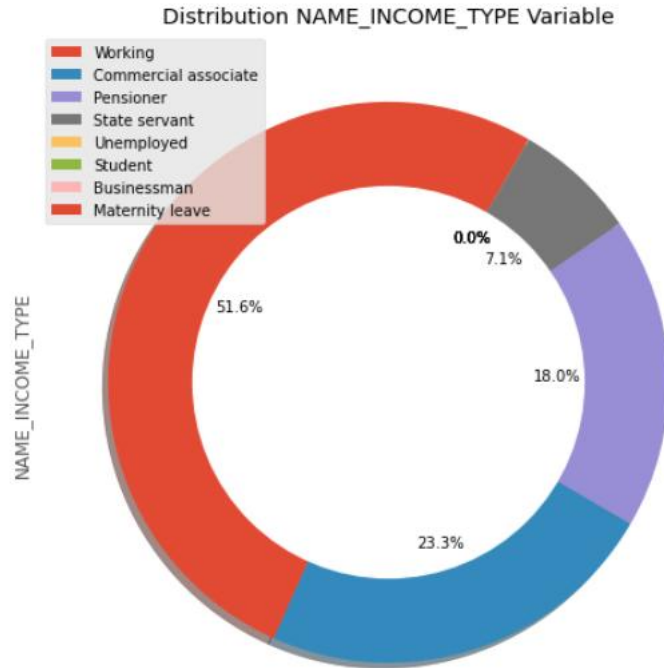
OCCUPATION TYPE

Distribution of OCCUPATION_TYPE Variable

OCCUPATION_TYPE of clients with payment difficulties
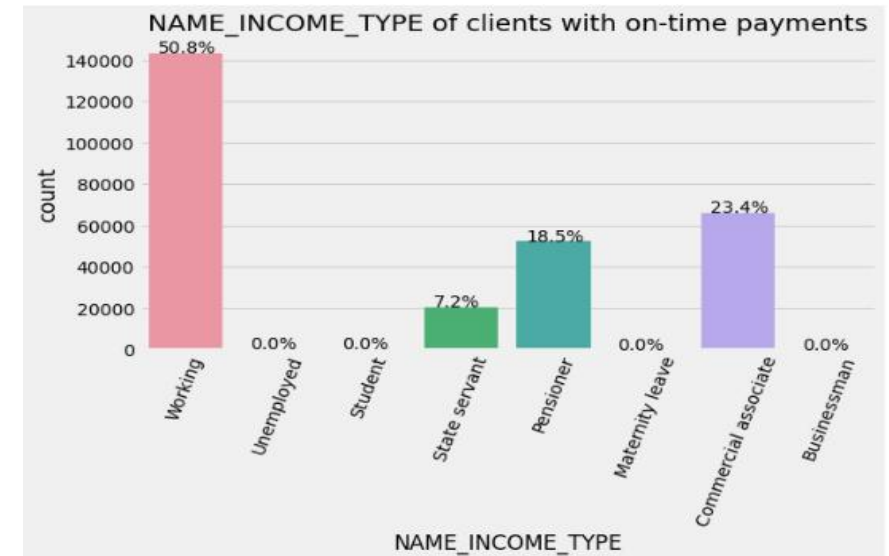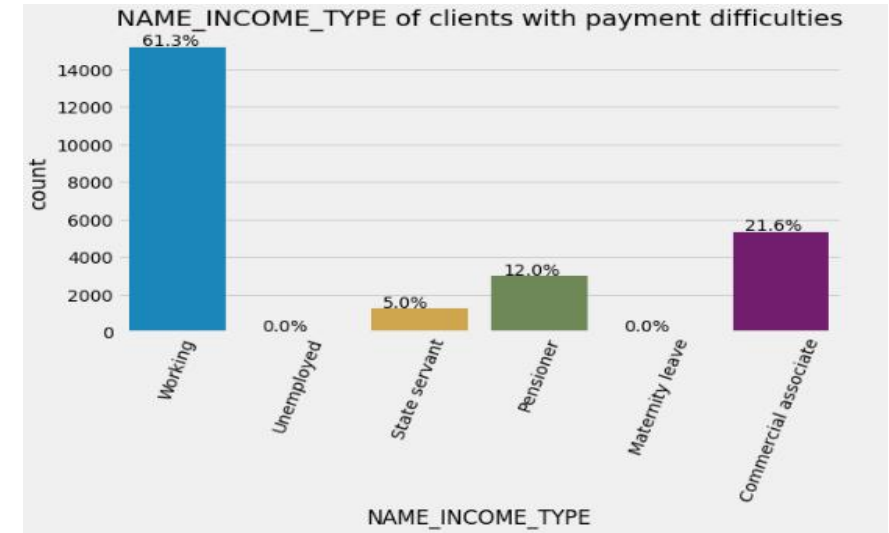

OCCUPATION_TYPE of clients with on-time payments

- Laborers outnumber other known occupation types even though the most specific information is unclear.

- Waiters or barman staff, Low skilled Labourers,Drivers and salespeople in this occupation type had greater records of payment troubles, according to occupation segment analysis
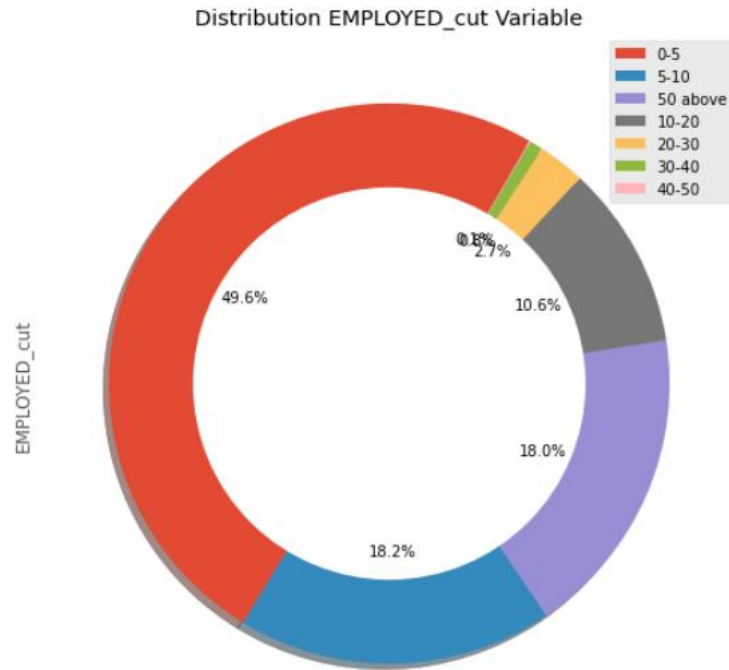
INCOME TYPE

Distribution NAME_INCOME_TYPE Variable

NAME_INCOME_TYPE of clients with payment difficulties

NAME_INCOME_TYPE of clients with on-time payments
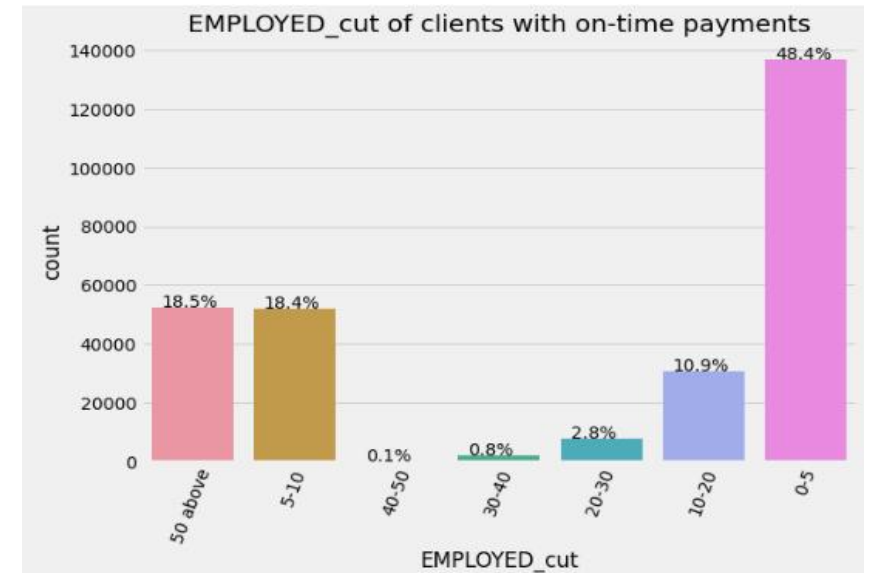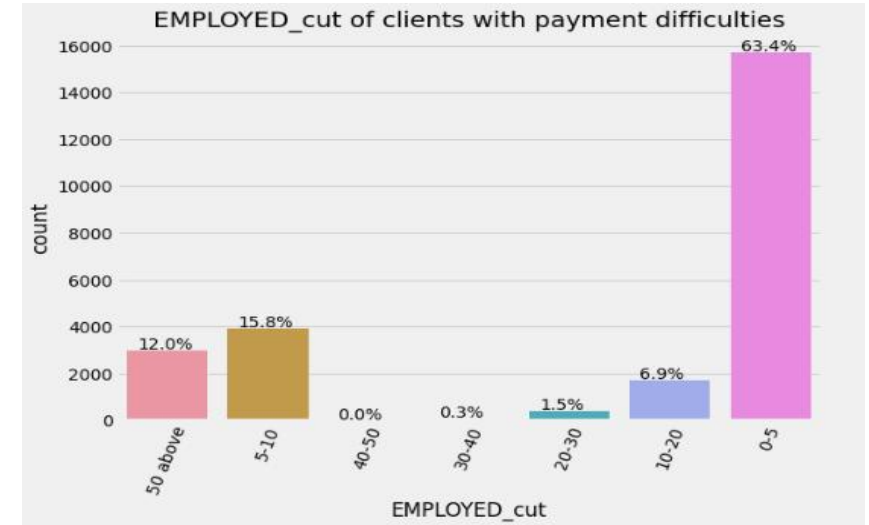
- The majority of applicants (around 51%) are employed, followed by retirees (18%) and business associates (23%).

- While there are many income brackets, working professionals are the ones who borrowed the most.When it comes to on-time payments, commercial associates (23.4%) and retirees (18.5%) perform well, although their numbers are quite low when compared to working professionals

YEARS EMPLOYED

Distribution EMPLOYED_cut Variable

EMPLOYED_cut of clients with payment difficulties
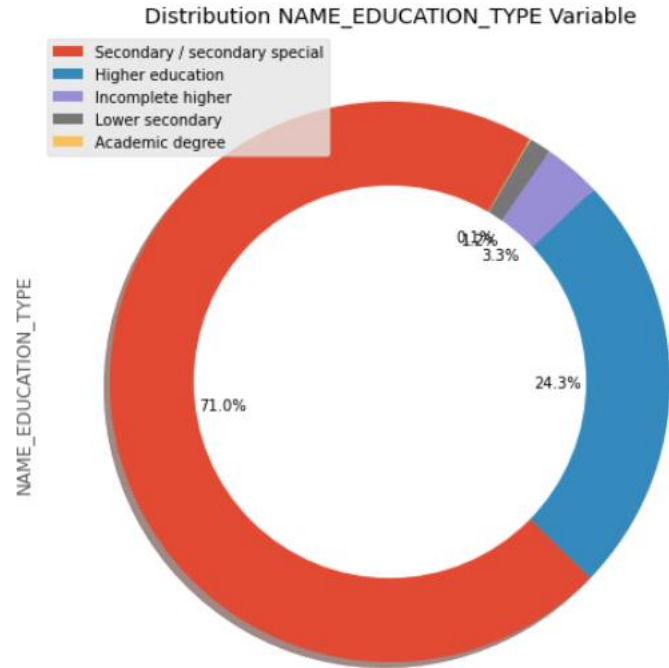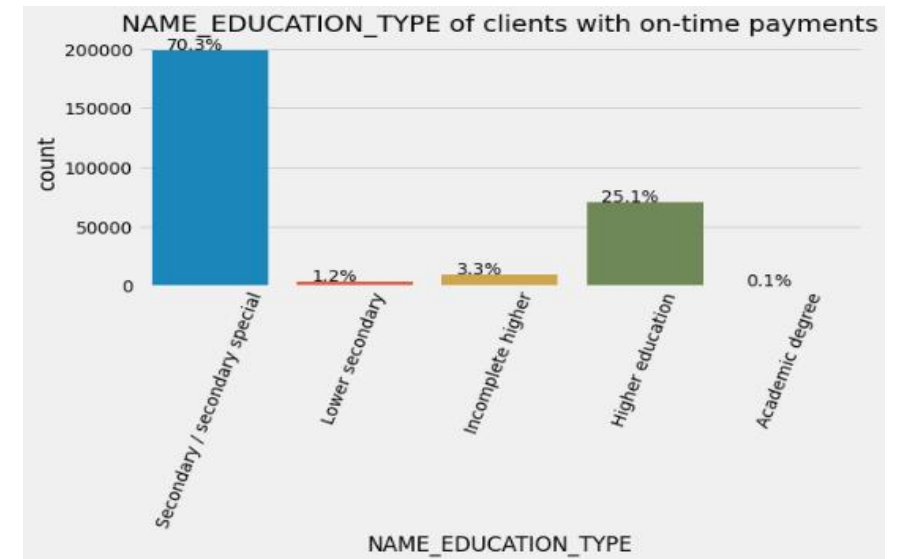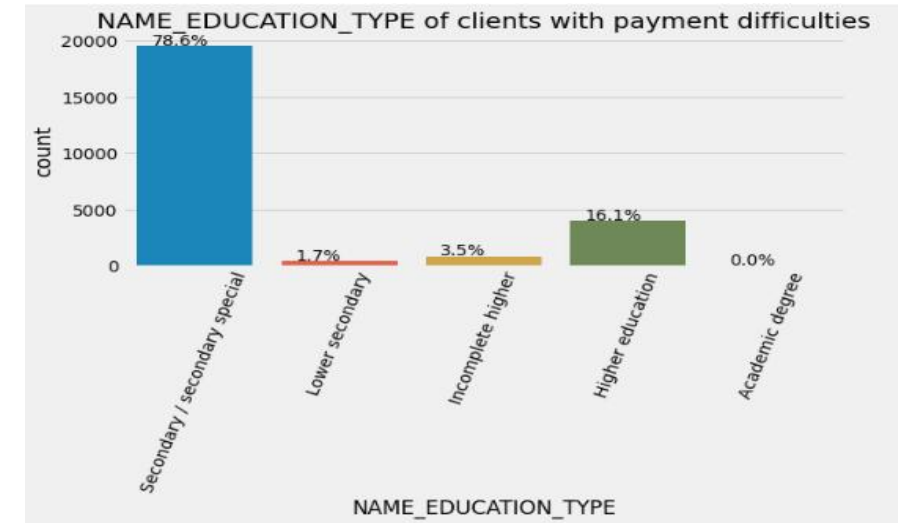

EMPLOYED_cut of clients with on-time payments

- The majority of workers, or 50%, have a history of employment of no more than five years.

- Clients with 5 years or less of employment experience have a higher probability of payment issues.
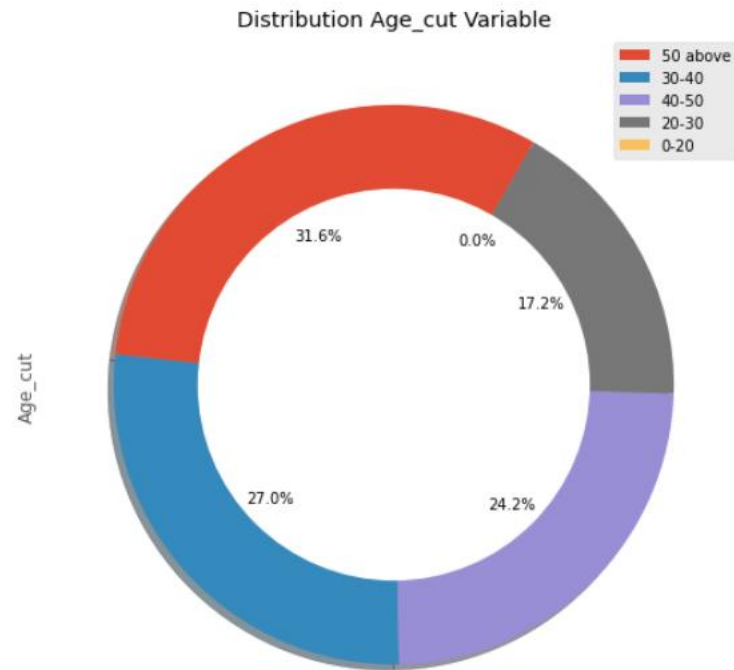
EDUCATION TYPE

Distribution NAME_EDUCATION_TYPE Variable

Legend:
- Secondary / secondary special
- Higher education
- Incomplete higher
- Lower secondary
- Academic degree

71.0% · 24.3% · 3.3% · 0.12%

NAME_EDUCATION_TYPE of clients with payment difficulties

78.6% · 1.7% · 3.5% · 16.1% · 0.0%

NAME_EDUCATION_TYPE of clients with on-time payments

70.3% · 1.2% · 3.3% · 25.1% · 0.1%

- The majority of applicants' educational backgrounds, or 71%, are secondary.

- Secondary/secondary specials have taken on a greater number of loans, and their repayment troubles are very, very significant. Even still, clients with higher education backgrounds have greater on-time repayment rates than problems. Here, the percentage of secondary education is 70.3, and the percentage of higher education is 25.1%. For secondary education, the percentage of defaulters in cases of payment difficulties is relatively high (78.6%).

YEARS BIRTH

Distribution Age_cut Variable

- 50 above
- 30-40
- 40-50
- 20-30
- 0-20

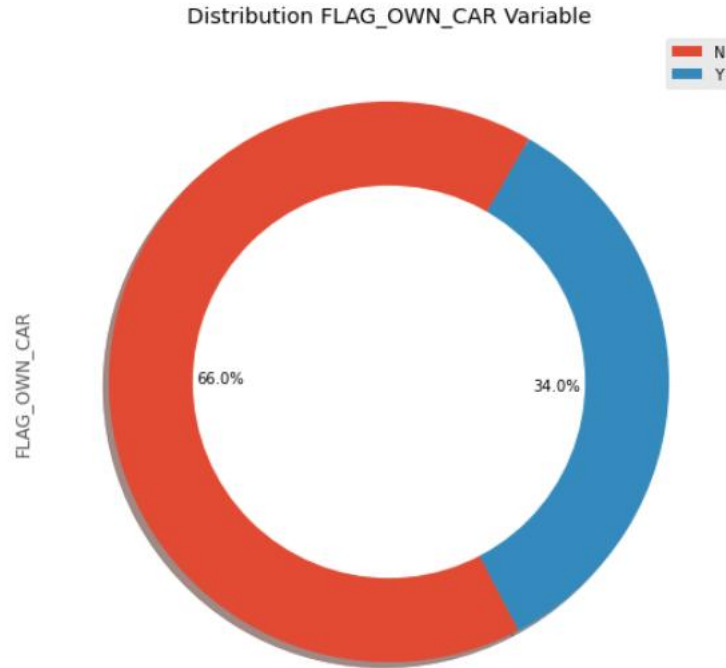Age_cut of clients with payment difficulties

Age_cut of clients with on-time payments

- The majority of borrowers are over 30 years old.

- Customers above the age of 50 pay more in terms of loan payback, whilst customers between the ages of 30 and 40 have trouble making payments.

CAR OWNERS

Distribution FLAG_OWN_CAR Variable
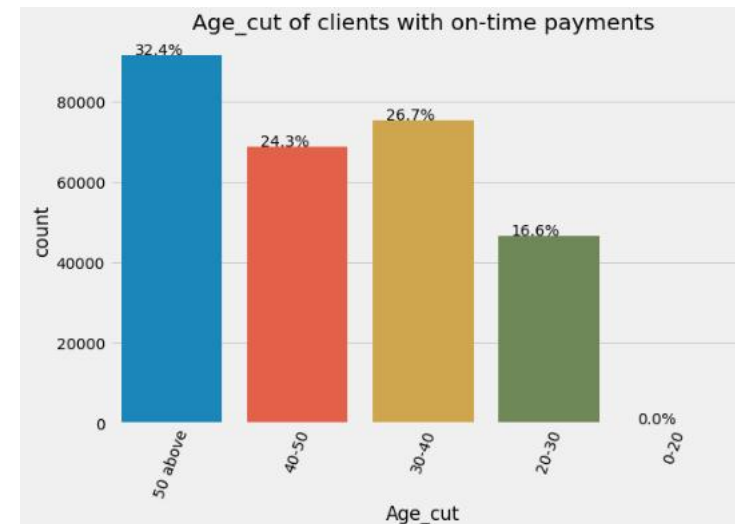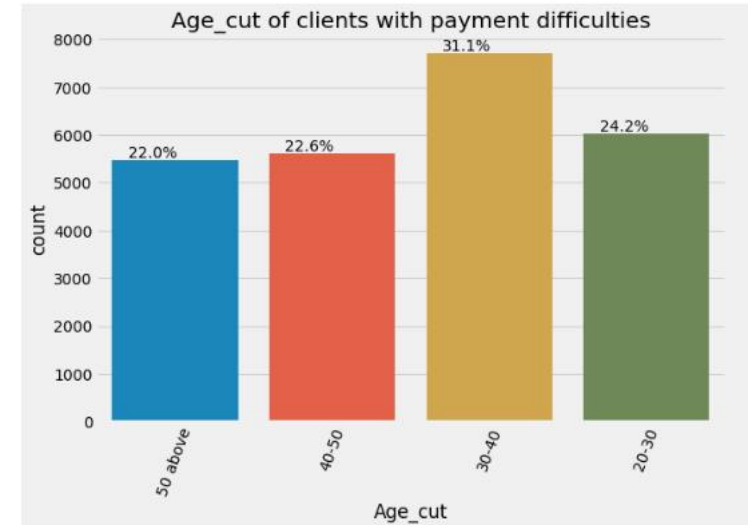
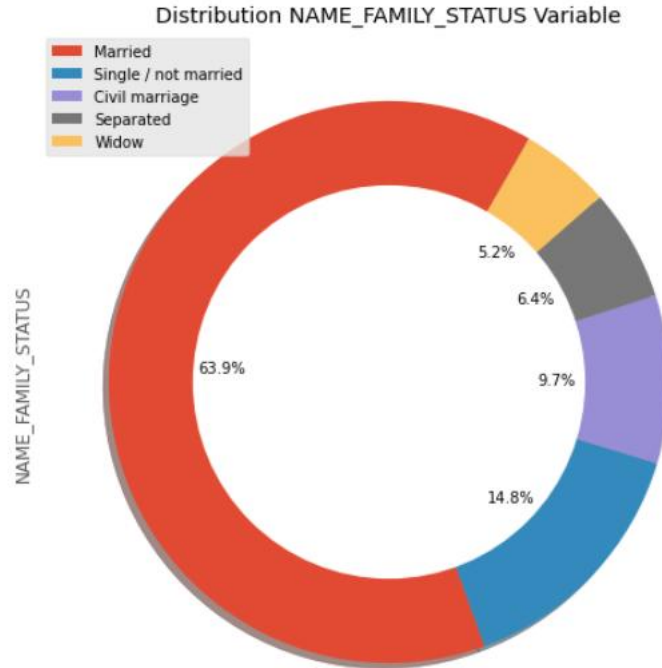- About 66% of applicants don't own a car.

- People who own vehicles have taken out fewer loans than those who don't.When it comes to on-time payments, 34.3% of people have cars, while 65.7% don't.Car owners make up 30.5% of those experiencing payment difficulties, while those without a car make up 69.5%.

FLAG_OWN_CAR of clients with payment difficulties

FLAG_OWN_CAR of clients with on-time payments

FAMILY STATUS

Distribution NAME_FAMILY_STATUS Variable

NAME_FAMILY_STATUS of clients with payment difficulties

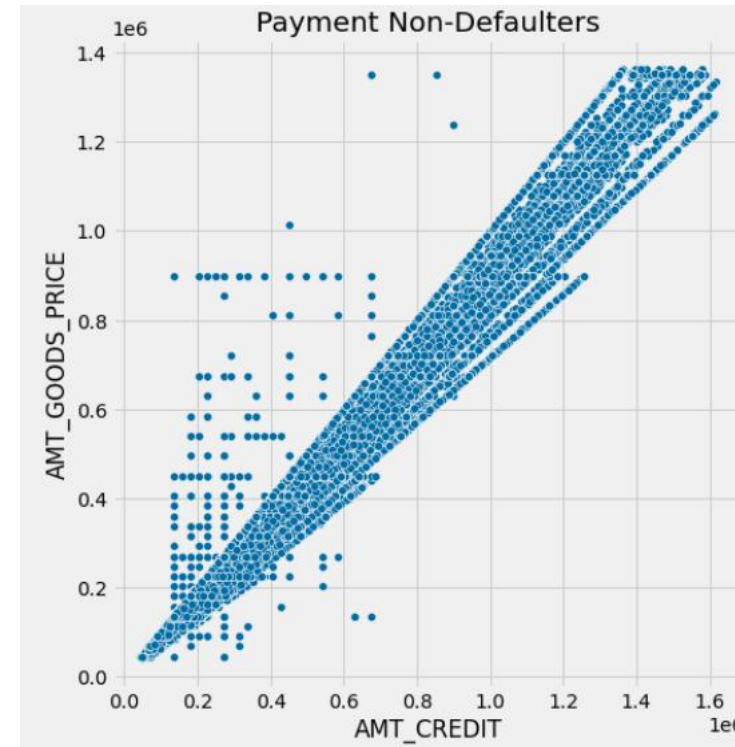NAME_FAMILY_STATUS of clients with on-time payments

- 64 percent of applicants are married.

- Family background analysis reveals that married and widowed clients are more likely to make their payments on time.in the event of a timely payment 64.2% of people are married, 5.4% are widows,
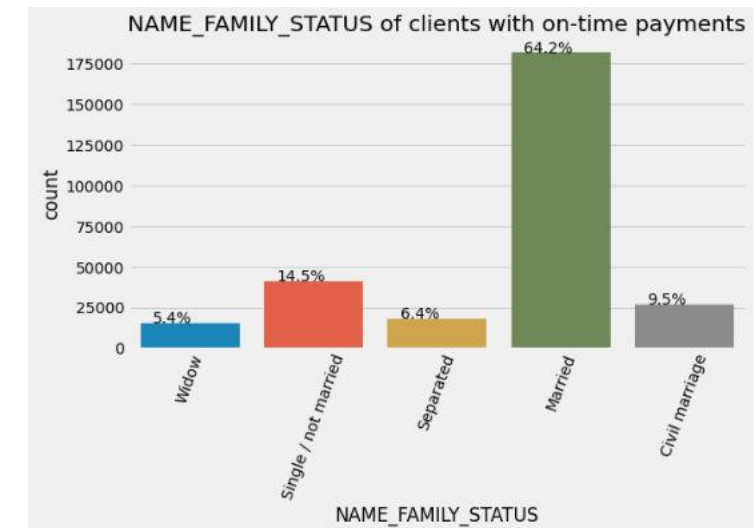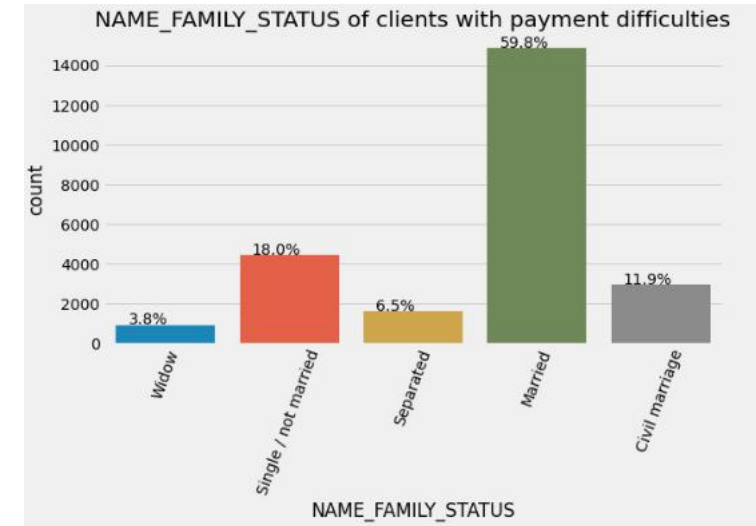
BIVARIATE & MULTIVARIATE ANALYSIS

CREDIT AMOUNT vs GOODS PRICE AMOUNT
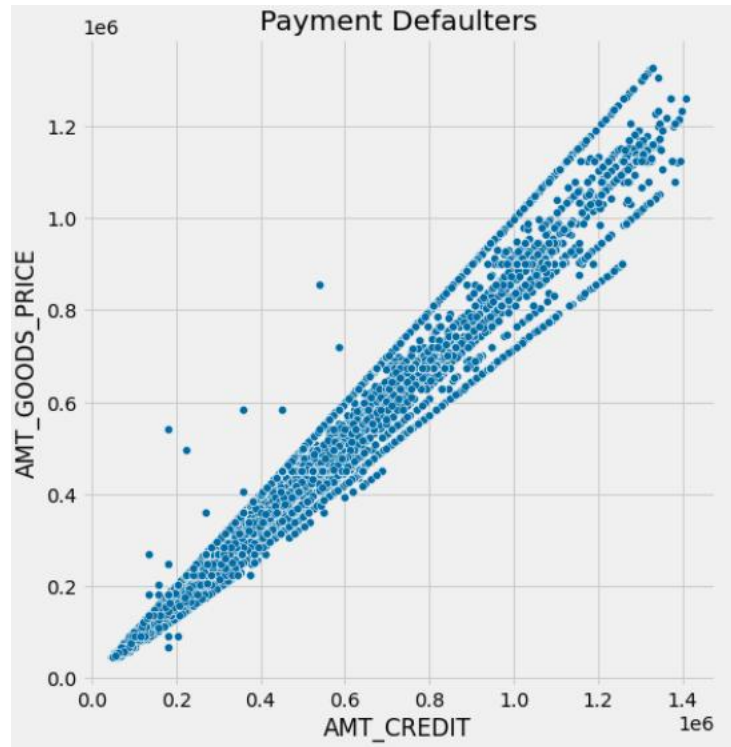
- Amount Credit and amount Goods prices have a close relationship.
- If Amount Credit increases, amount Goods prices also rises. There are numerous correlations.

CREDIT AMOUNT vs ANNUITY AMOUNT
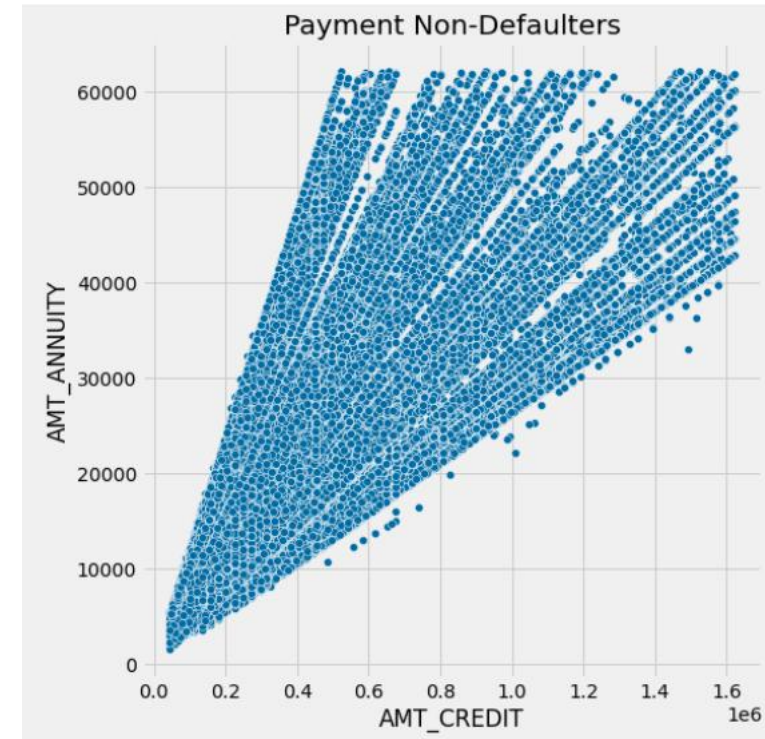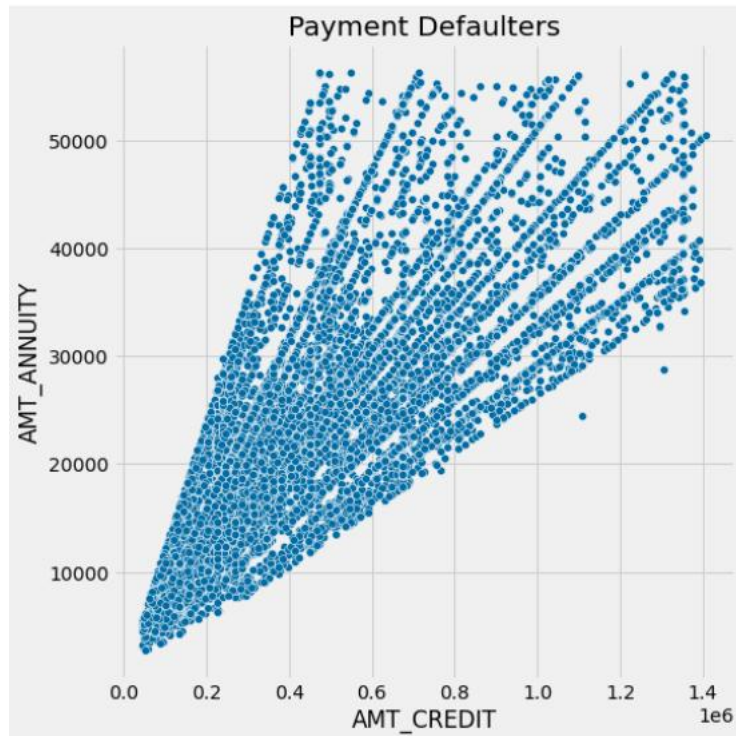
- A close relationship exists between the amounts of Credit and Annuity.
- Amount Annuity rises in proportion to Amount Credit. Numerous relationships exist.

ANNUITY AMOUNT vs GOODS PRICE AMOUNT
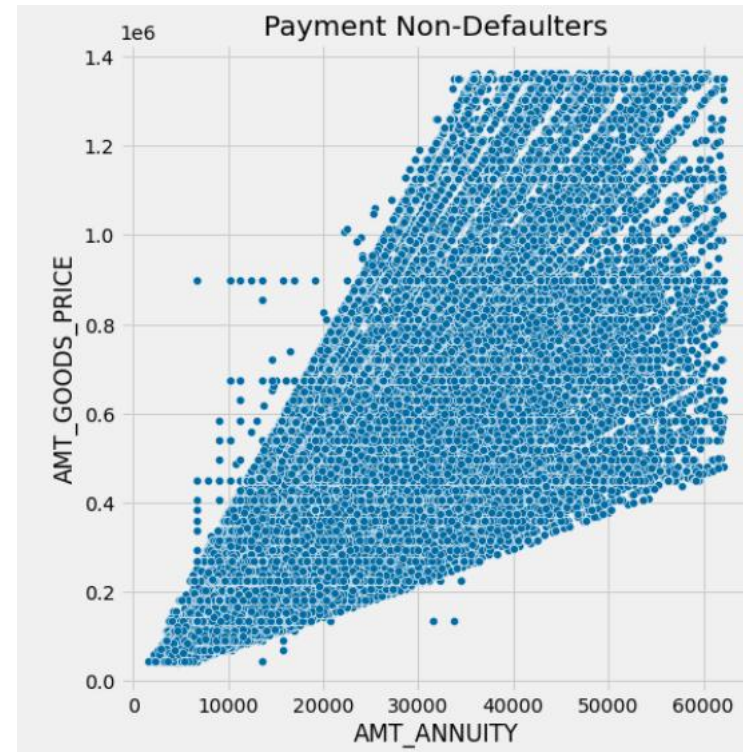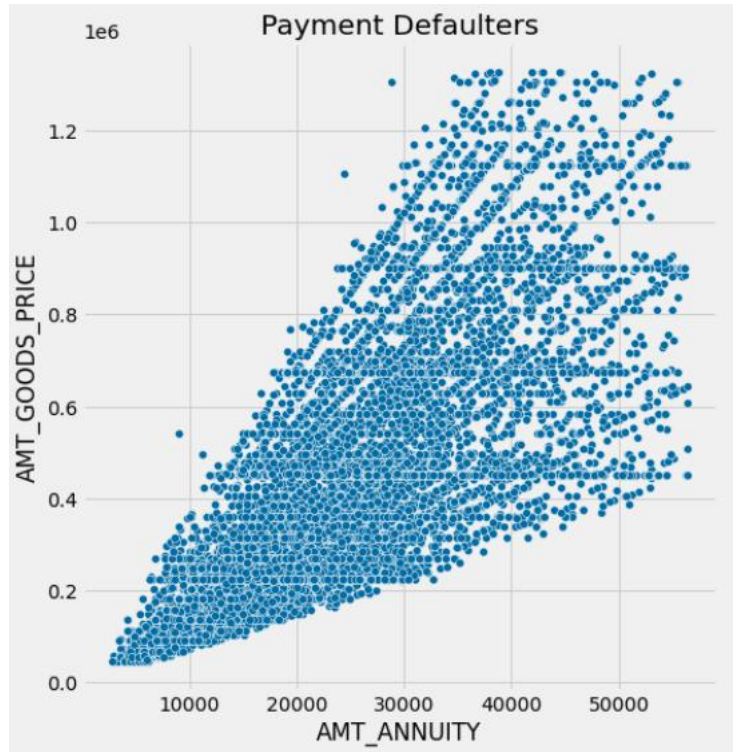
- There is a strong relationship between the variables amount Annuity and amount Goods prices.

- Amount Goods prices rises in response to rising amount Annuity . High correlation exists.

TOP CORRELATIONS FOR DEFAULTERS

Top 10 correlation between continuous column for the clients with payment difficulties:

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998269
AMT_CREDIT	AMT_GOODS_PRICE	0.983103
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956637
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.868994
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.847885
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.778540
AMT_ANNUITY	AMT_GOODS_PRICE	0.752699
AMT_CREDIT	AMT_ANNUITY	0.752195
Years_Age	Years_EMPLOYED	0.582443
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.497937
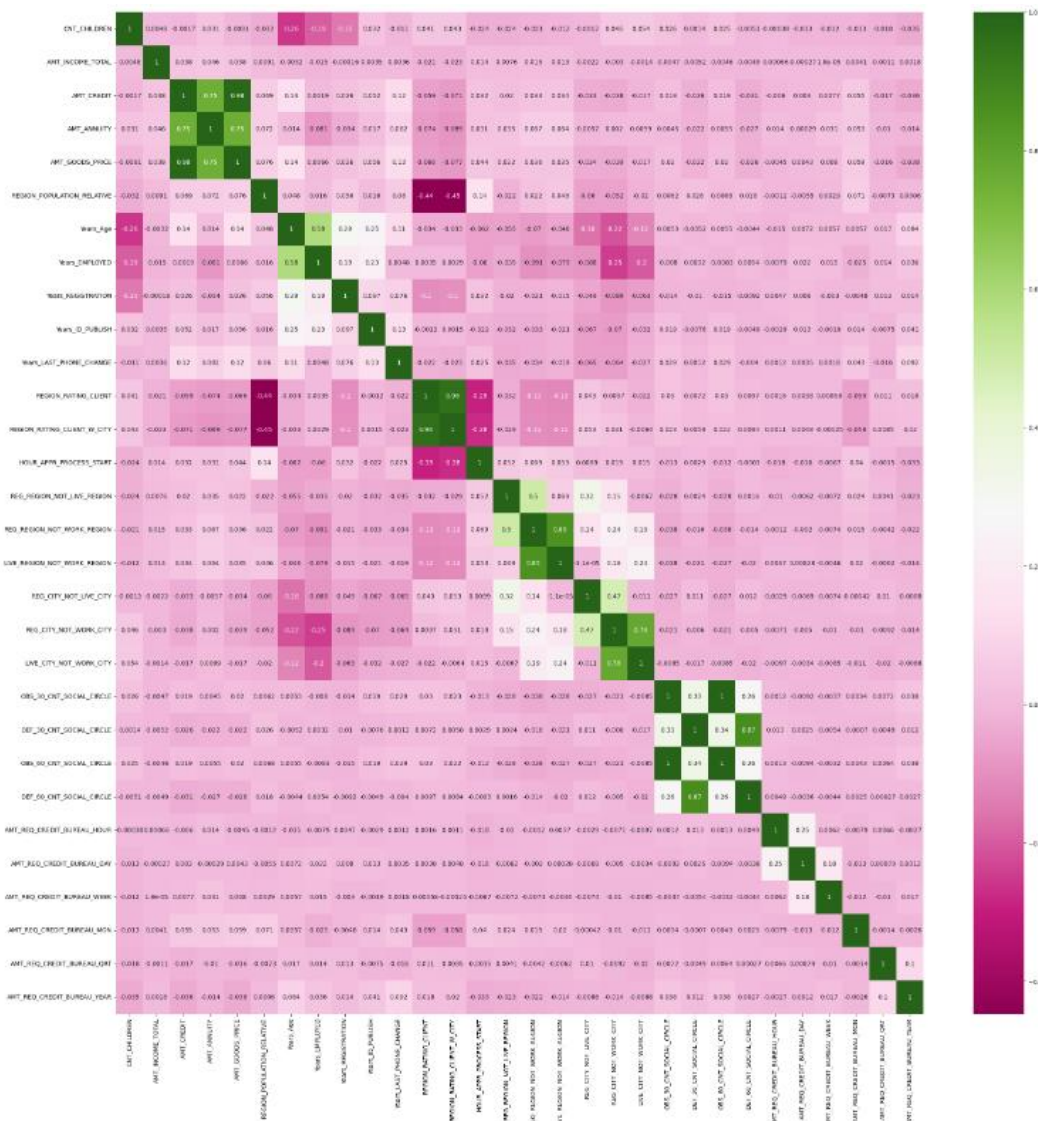
TOP CORRELATIONS FOR NON-DEFAULTERS

Top 10 correlation between continuous column for the clients
with payment non-difficulties:

|		
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998508
AMT_CREDIT	AMT_GOODS_PRICE	0.987250
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.950149
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.861861
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.859332
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830381
AMT_ANNUITY	AMT_GOODS_PRICE	0.776686
AMT_CREDIT	AMT_ANNUITY	0.771309
Years_Age	Years_EMPLOYED	0.626028
REGION_POPULATION_RELATIVE	REGION_RATING_CLIENT	0.539005
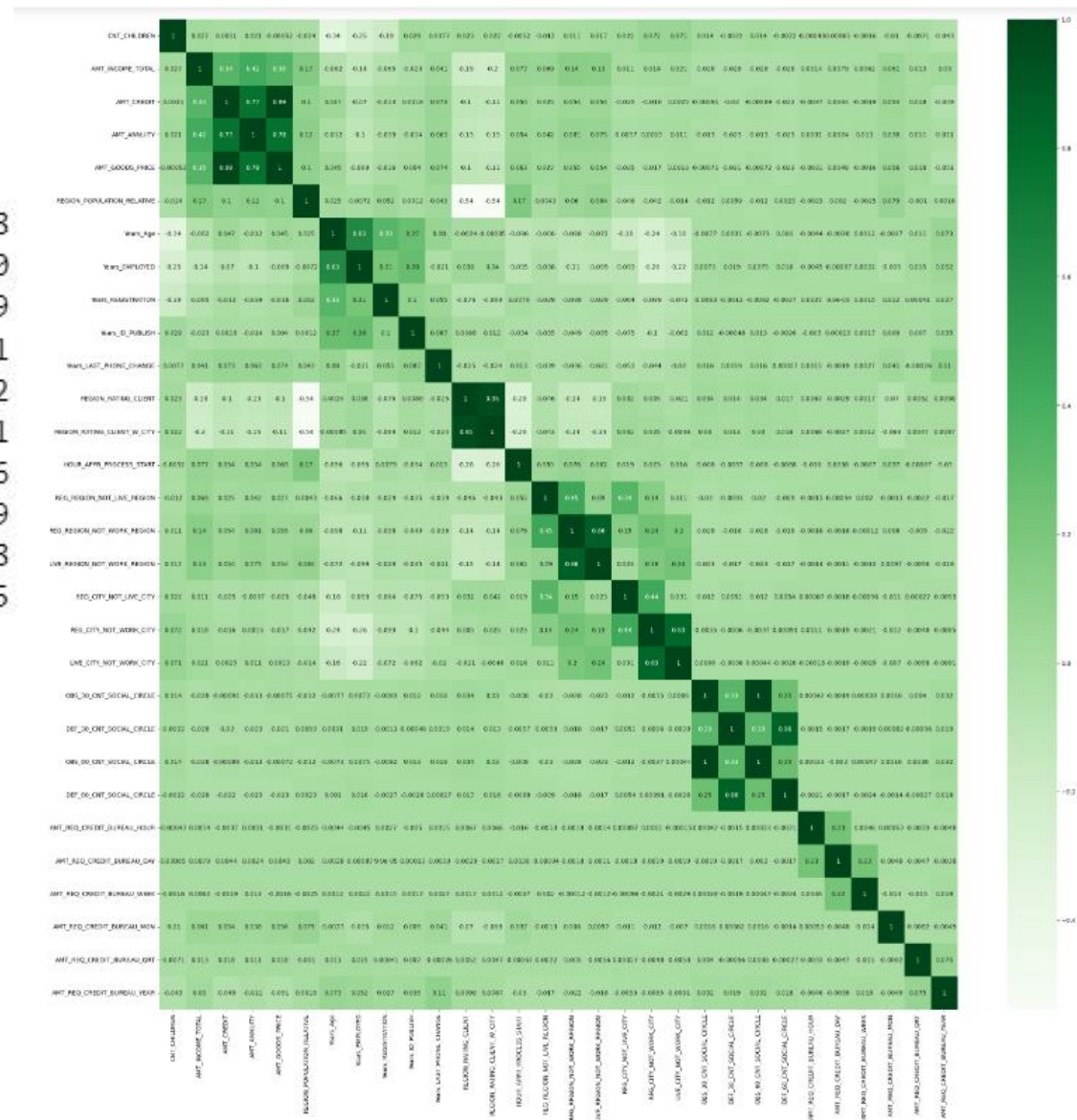
CASH LOAN PURPOSE ANALYSIS

NAME_CASH_LOAN_PURPOSE for Client with Payment Difficulties
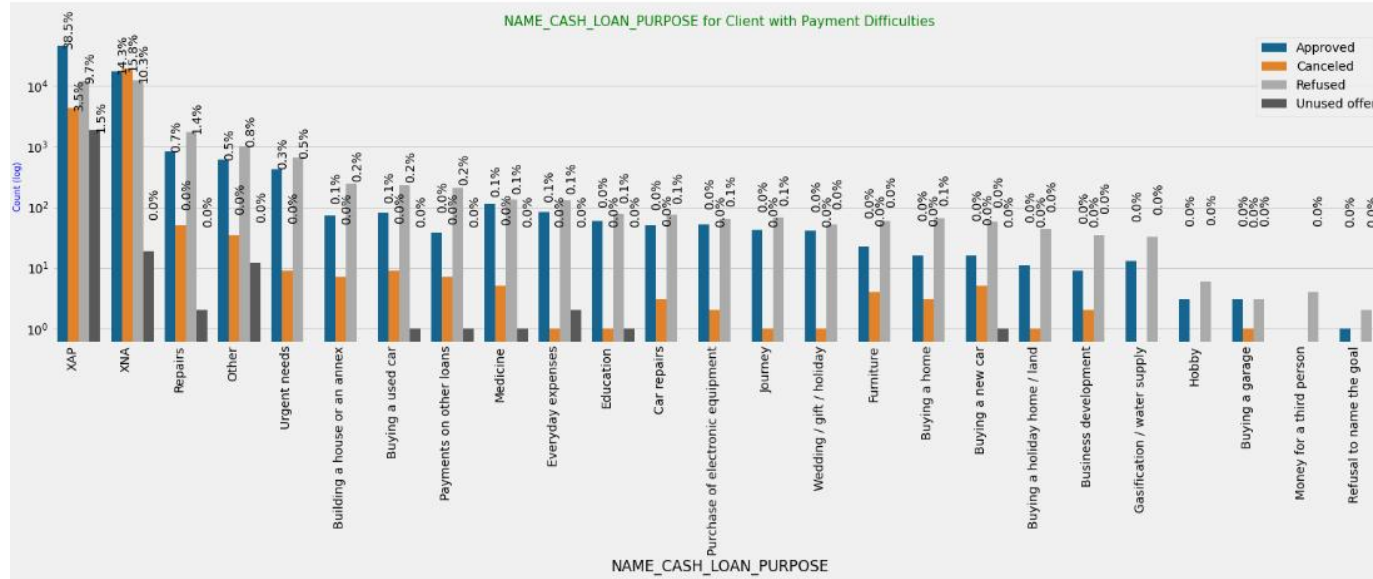
- The cost of loans taken for XNA and XAP is extremely expensive.
- The rate of loan default for the purpose of repair is very high.

- When a client takes out a loan for a hobby or garage purchase, there is a lower likelihood that they would default.

NAME_CASH_LOAN_PURPOSE for Client with On-Time Payment
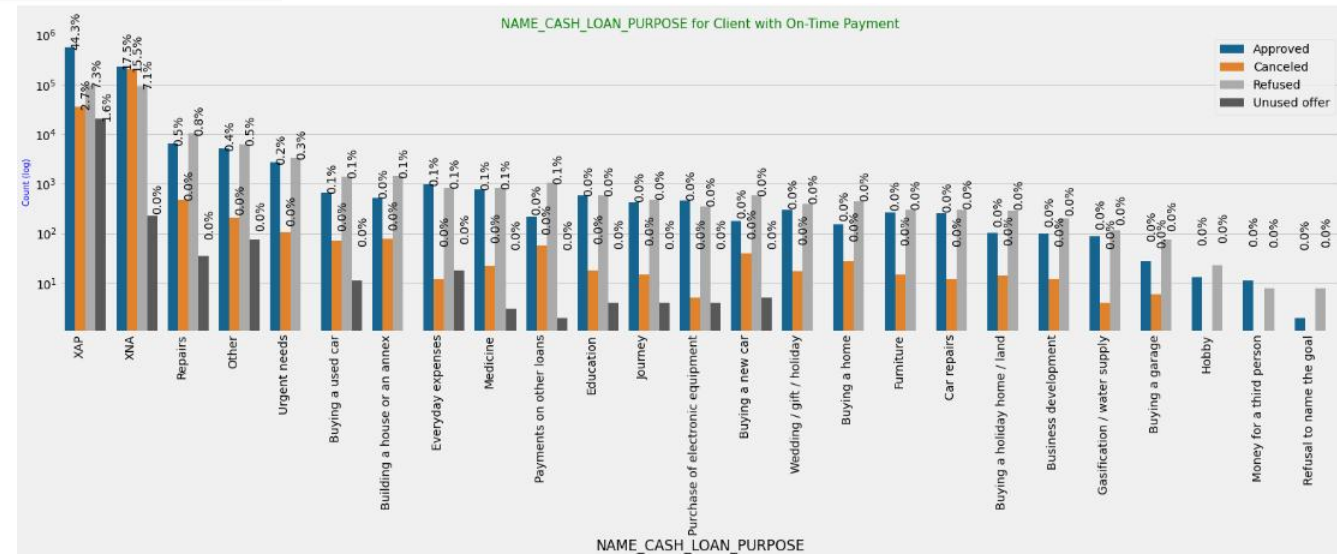
RATING OF THE REGION WHERE CLIENT LIVES
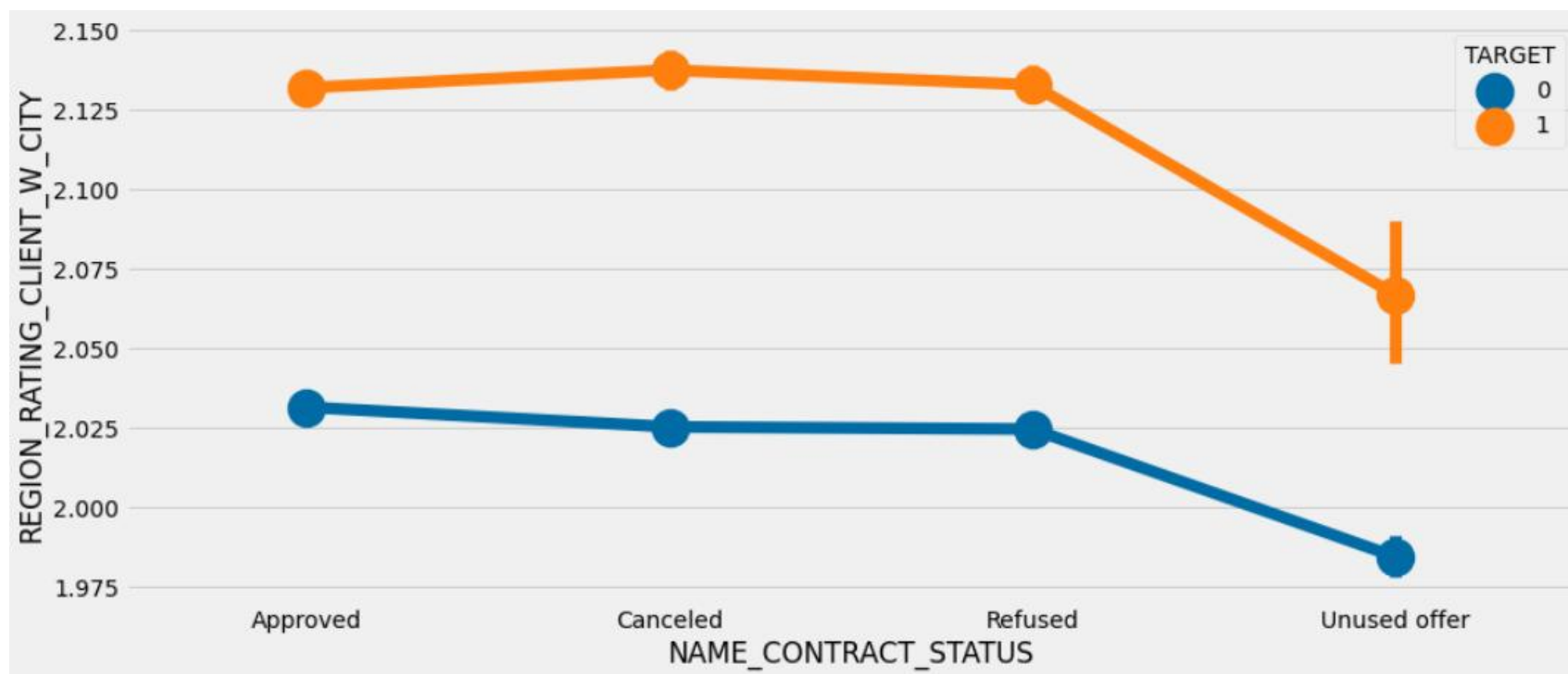
The largest rate of loan defaults is among residents of ratings higher than 2.04.

SUMMARY

Determiners of whether a candidate will be a Repayer:

- **ORGANIZATION TYPE:** Lenders are more likely to grant loans to clients who fall within the Trade Type 4,5 and Industry Type 8 categories.

- **NAME INCOME TYPE:** Individuals with a history of repaying loans based on their income include business owners, students, pensioners, and government workers.

- **DAYS EMPLOYED:** It is very safe to offer loans to customers with a 40–50-year history of employment. Their default proportion is less than 1%.

- **AMT INCOME TOTAL**: Customers that make over 6 lacs are far more likely to make their payments on time.

- **NAME EDUCATION TYPE:** Clients with academic degrees are less likely to default on loans.

- **DAYS BIRTH:** Customers above 50 have a low likelihood of defaulting.

- **FLAG OWN CAR:** Customers who own a car are less likely to make unpaid obligations.

- **NAME FAMILY STATUS:** Clients who are married or widowed are more likely to pay on time.

- **CNT CHILDREN:** People with 0–2 children are more likely to pay back their loans.

Determiners of whether a candidate will be a Defaulters:

- **ORGANIZATION TYPE :** The self-employed and business entities of type 3 experience the most payment problems here. Before granting a loan, the bank ought to reconsider its analysis of that customer's profile.

- **OCCUPATION TYPE**: Low-skill labourers, drivers, waiters/bartenders, security personnel, labourers, and cooks should be avoided due to the high rate of default.

- **DAYS EMPLOYED:** Default rates are higher for those with fewer than five years of work.

- **DAYS BIRTH:** Young persons between the ages of 30 and 40 should be avoided since they are more likely to default.

- **NAME INCOME TYPE:** Customers who are on maternity leave OR who are unemployed frequently default.

- **NAME_EDUCATION_TYPE:** People with lower secondary and secondary education experience greater loan payment challenges, according to NAME EDUCATION TYPE.

- **AMT CREDIT:** The number of defaulters rises when the credit amount exceeds 20 lacs.

- **OBS 60 CNT SOCIAL CIRCLE:** Loan default rates were lower for clients with OBS 60 CNT SOCIAL CIRCLE scores between 1.49 and 1.59.

- **DEF 60 CNT SOCIAL CIRCLE**: It is crucial to take the client's social network into account before approving a loan because customers with average DEF 60 CNT SOCIAL CIRCLE scores of 0.13 or higher default more frequently.

- **REGION RATING CLIENT W CITY:** The largest rate of loan defaults is among residents of ratings higher than 2.04.

THANK YOU!