EDA Group Project

Topic: Analyze the Risk for Loan Application Approval & Rejection

Date: 5th April'21

Team Members:
Vishal Kale
Abhinandan Gupta

EDA Group Project

Problem Statement:

Understanding if the consumer will repay the loan on time or will not repay the loan on time and become a defaulter before taking decision to pass the loan application. Though the company has insufficient or non-existent credit history , there are certain risks involved before accepting/rejecting the loan application:

- **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.**

EDA Group Project

Data Cleaning & Outlier Analysis:

- ***Eliminating the columns having more than 50% null values :*** *There has been 41 no of columns that were eliminated.*

- ***Removing the unwanted columns as these columns does not represent any specific information about the user:*** *There has been 27 no of columns that were eliminated.*

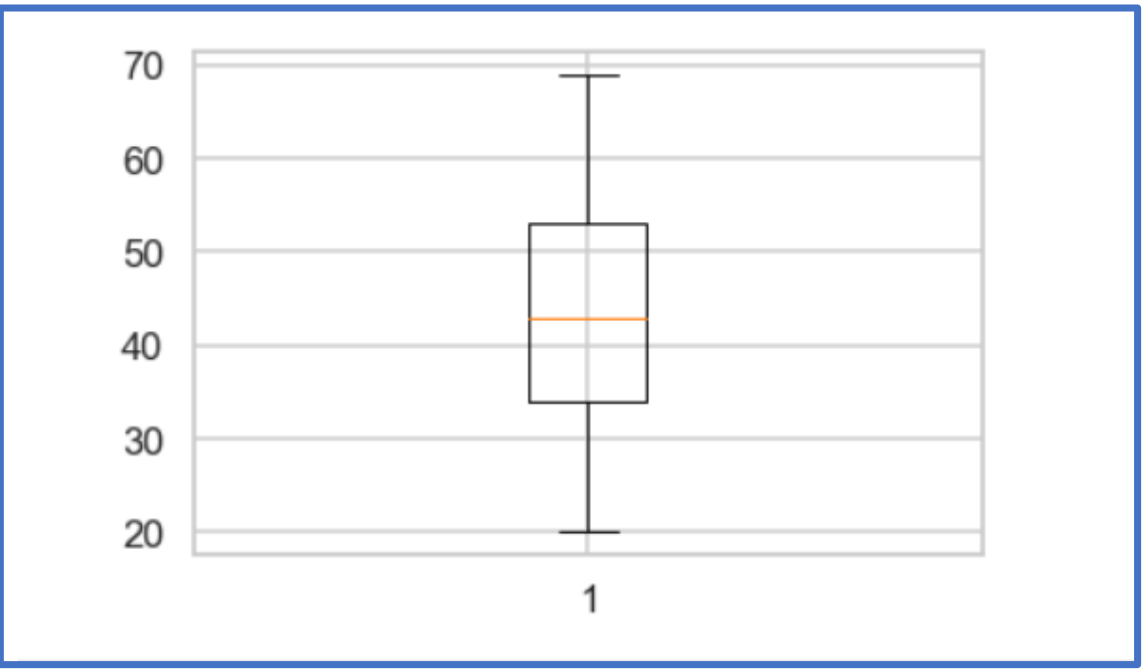
- ***There are four rows having XNA which means "Not Avaiable":*** *Since, Female is having the majority and only 4 rows are having NA values, we can update those columns with Gender 'F' as there will be no impact on the dataset.*

EDA Group Project

Data Cleaning & Outlier Analysis:

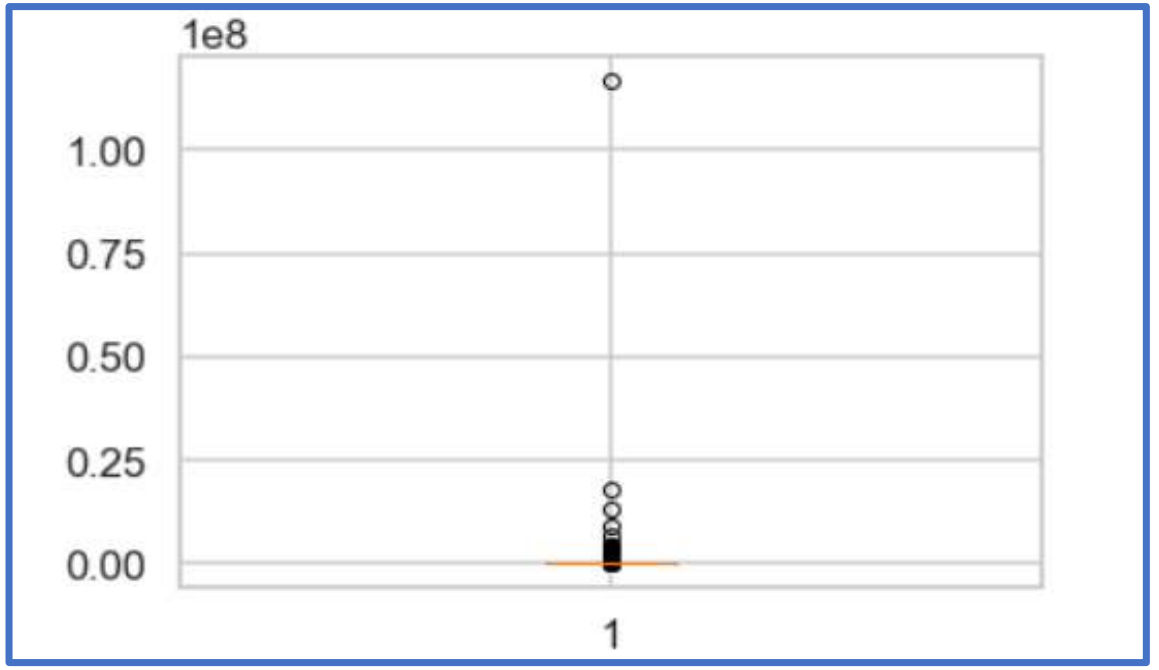
The outlier analysis for done for the vide range of data to check for any present outliers that can be eliminated ,cleaned and further analysis of the dataset can be carried out. Analysis & Identification of outliers done using box plot shown below:

Checking For Age:



No outliers as most of the applicants are in age group of 35 to 55:

Checking For Total Income:

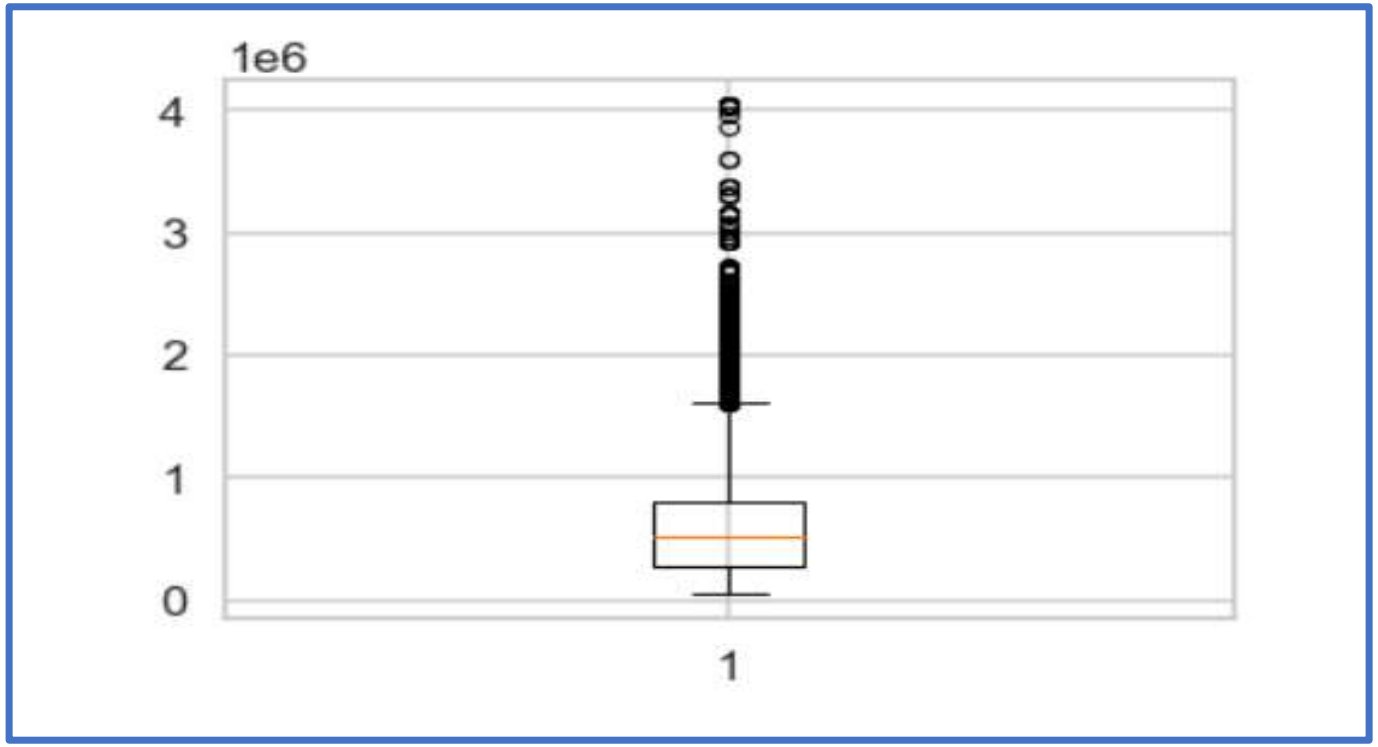


Some Outliers can be seen.

EDA Group Project

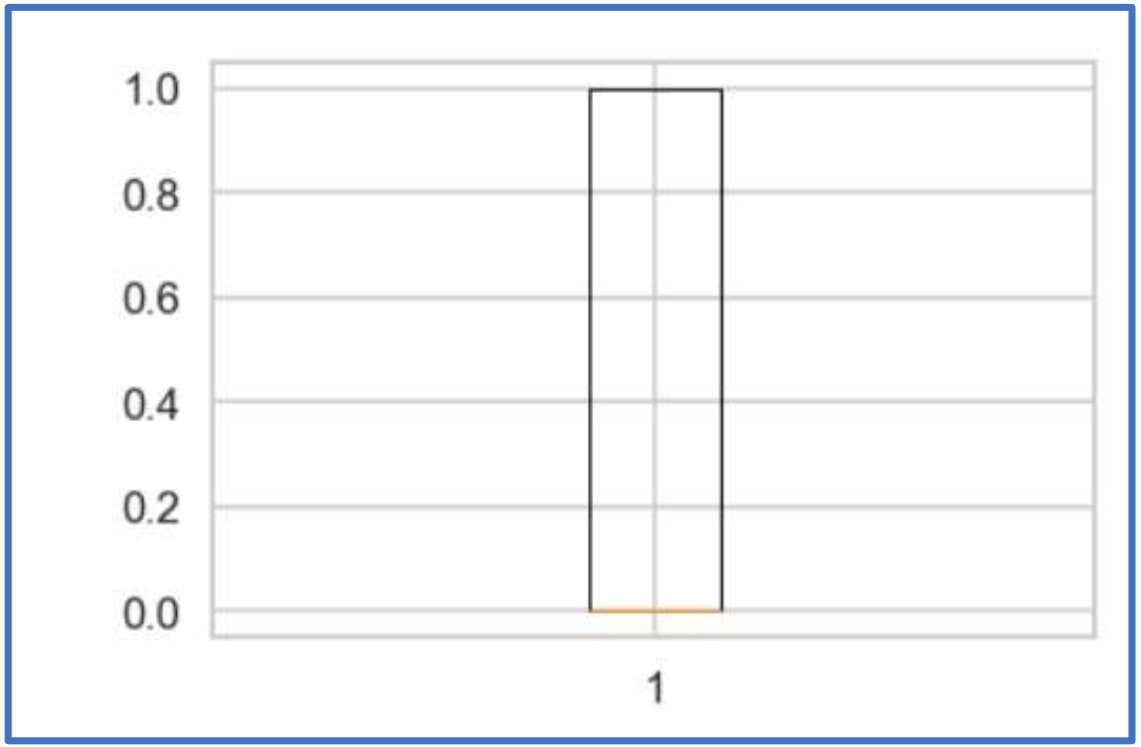
Data Cleaning & Outlier Analysis:

Checking For Credit Amount:



High number of outliers can be seen , that indicates most of the credit lies in outliers

Checking flag phone:



All applicants phone were reachable.

EDA Group Project

Binning

Bin created for “Amount Income total” & “Amount Credit”

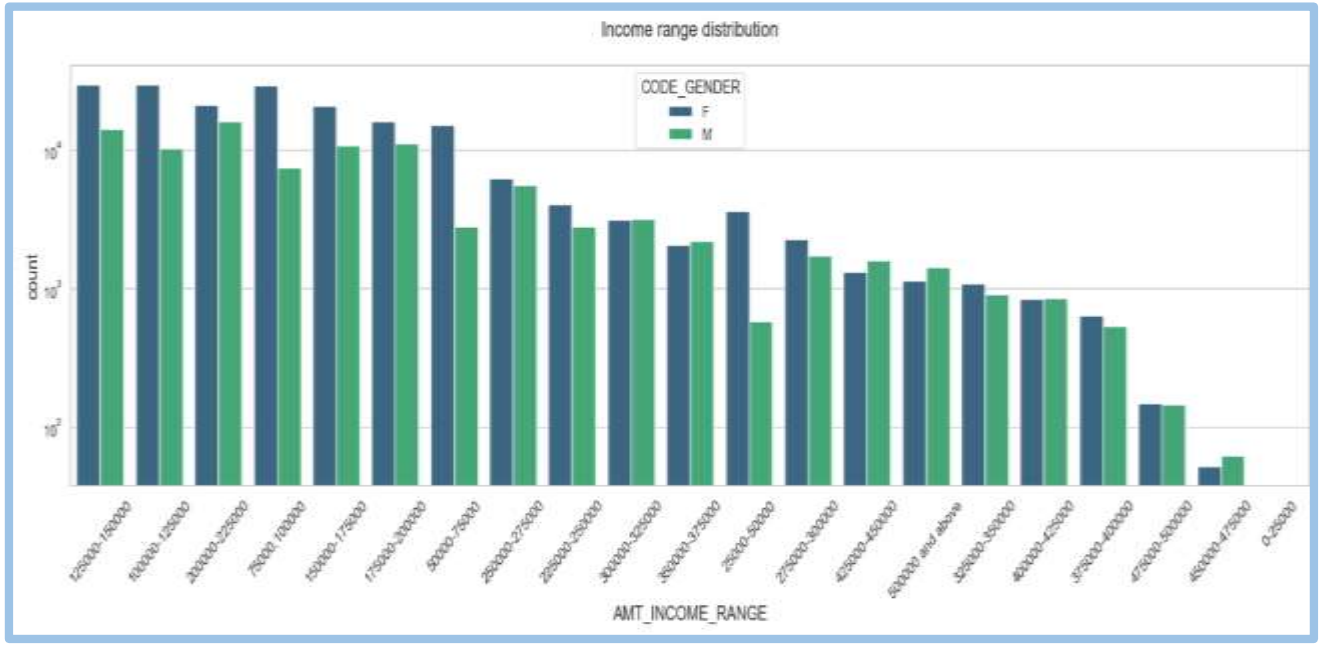
In [1]:
```
# AMT_INCOME_TOTAL

bins = [0,25000,50000,75000,100000,125000,150000,175000,200000,225000,250000,275000,300000,325000,350000,375000,400000,4
slot = ['0-25000', '25000-50000','50000-75000','75000,100000','100000-125000', '125000-150000', '150000-175000','175000-
       '200000-225000','225000-250000','250000-275000','275000-300000','300000-325000','325000-350000','350000-375000',
       '375000-400000','400000-425000','425000-450000','450000-475000','475000-500000','500000 and above']

application_data['AMT_INCOME_RANGE']=pd.cut(application_data['AMT_INCOME_TOTAL'],bins,labels=slot)
```

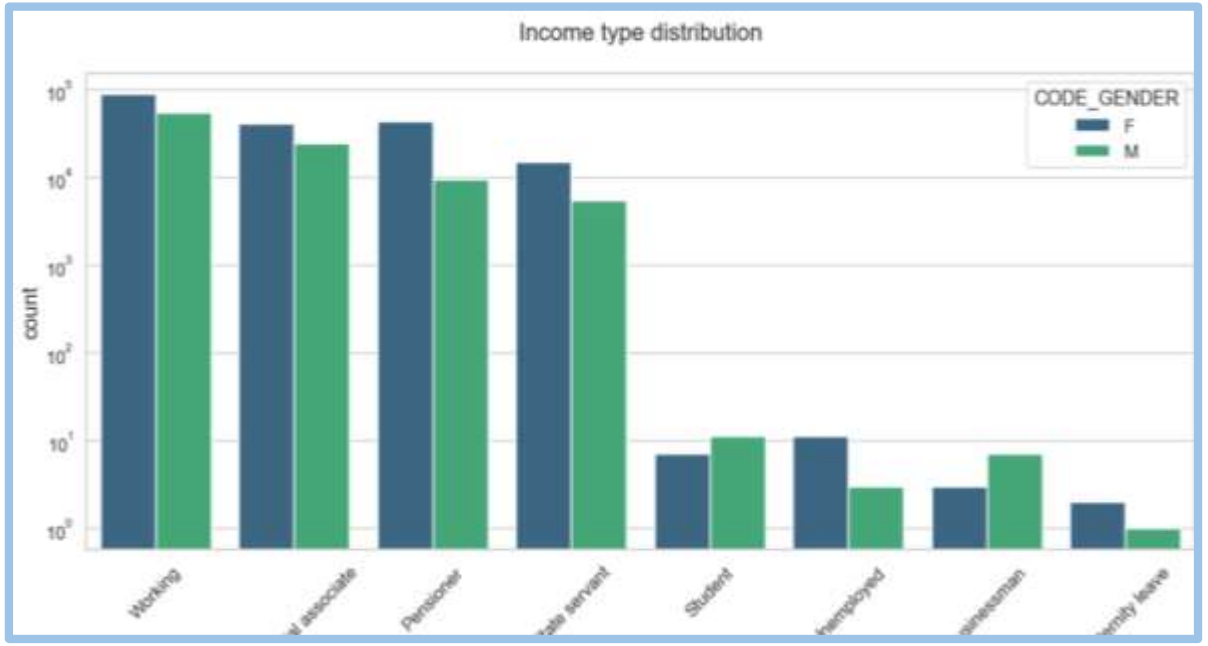
In [25]:
```
# AMT_CREDIT

bins = [0,150000,200000,250000,300000,350000,400000,450000,500000,550000,600000,650000,700000,750000,800000,850000,9000(
slots = ['0-150000', '150000-200000','200000-250000', '250000-300000', '300000-350000', '350000-400000','400000-450000',
        '450000-500000','500000-550000','550000-600000','600000-650000','650000-700000','700000-750000','750000-800000',
        '800000-850000','850000-900000','900000 and above']

application_data['AMT_CREDIT_RANGE']=pd.cut(application_data['AMT_CREDIT'],bins=bins,labels=slots)
```

EDA Group Project

Univariate Analysis:

*The univariate analysis done for consumers with **no payment difficulties(Target 0)**:*





Points that can be concluded from above graph:

1. **As compared to Male, Female count is more.**
2. **Most of the credits are in the income range of 50000 to 225000.**
3. **For income range 450000 to 475000 count is very less**

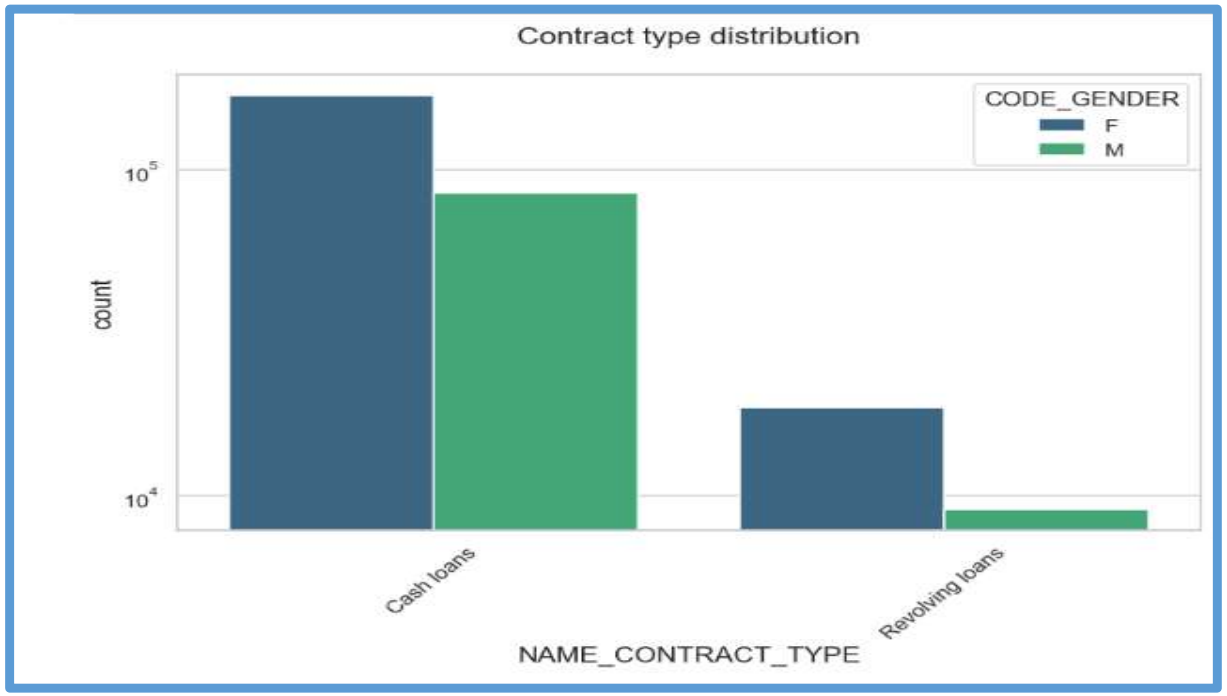
Points that can be concluded from above graph:

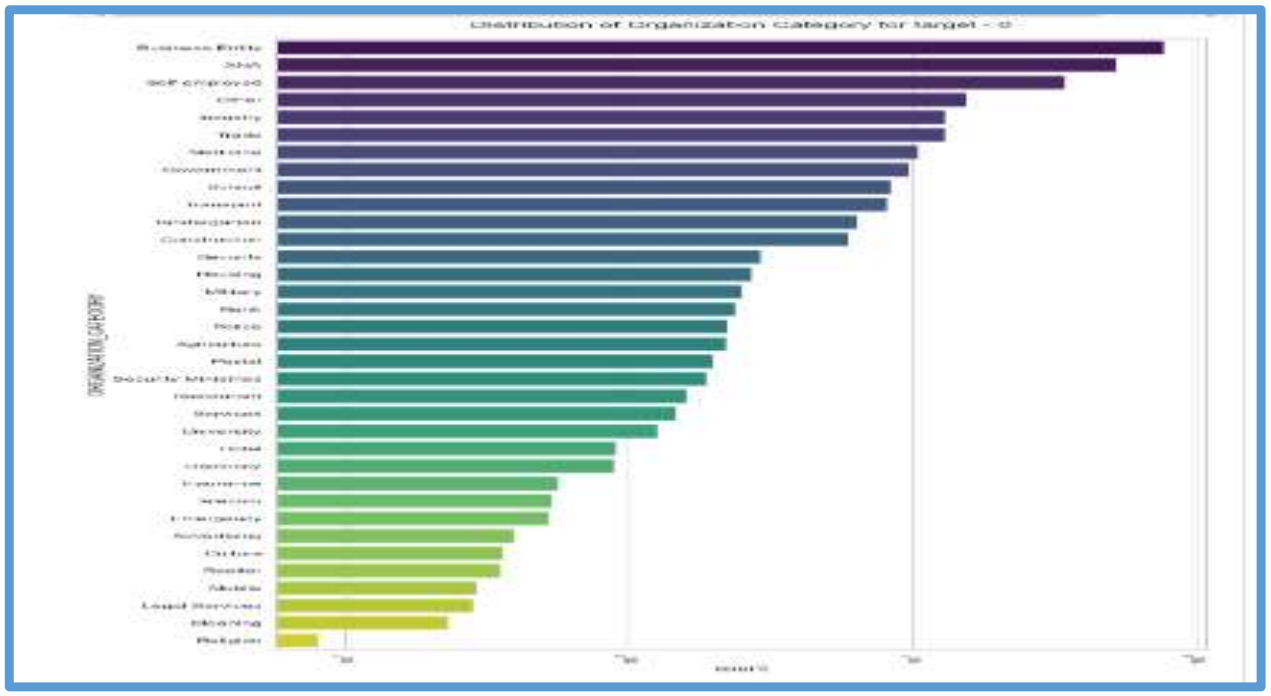
1. **For income type working, commercial associate, Pensioner and State Servant the number of credits are higher than others.**
2. **Less number of credits for income type Student ,Unemployed, Businessman and Maternity leave.**
3. **Mostly number of female credits is more than men in income type.**

EDA Group Project

Univariate Analysis:

The univariate analysis done for consumers with ***no payment difficulties(Target 0):***





Points that can be concluded from above graph:

1. **For contract type 'cash loans' number of credits is much higher than 'Revolving loans' contract type.**
2. **Number of female credits is more than men.**

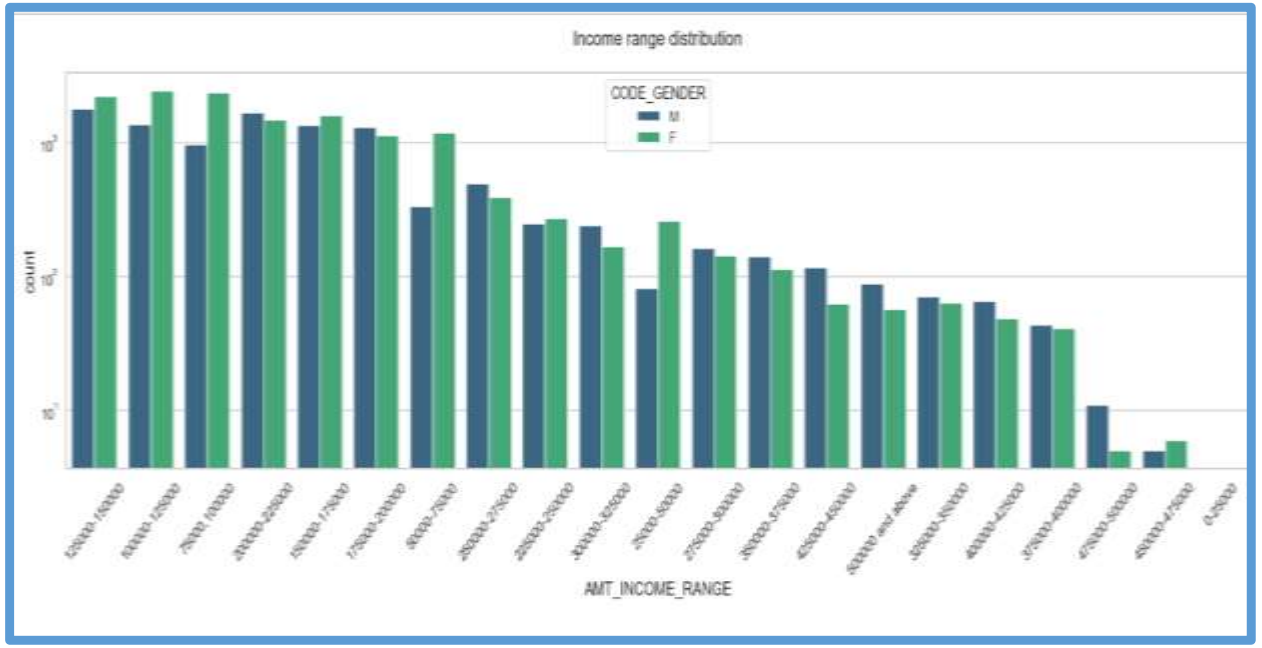
Points that can be concluded from above graph:

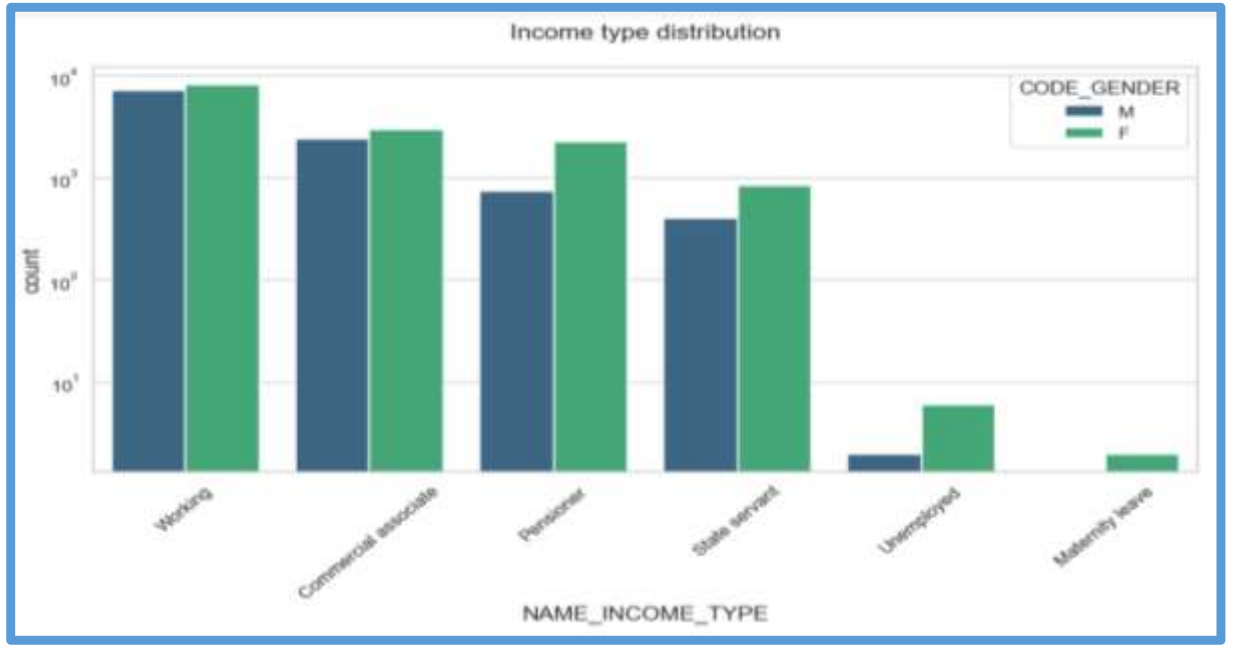
1. **Most the applicants are from Business, Self employed, Other, Industry and Trade category.**
2. **Less number of applicants are from Religion, Cleaning, Legal Services.**

EDA Group Project

Univariate Analysis:

*The univariate analysis done for consumers with **payment difficulties(Target 1)**:*





Points that can be concluded from above graph:

1. **As compared to Male, Female count is more.**
2. **Most of the credits are in the income range of 50000 to 225000.**
3. **For income range 450000 to 500000 count is very less.**

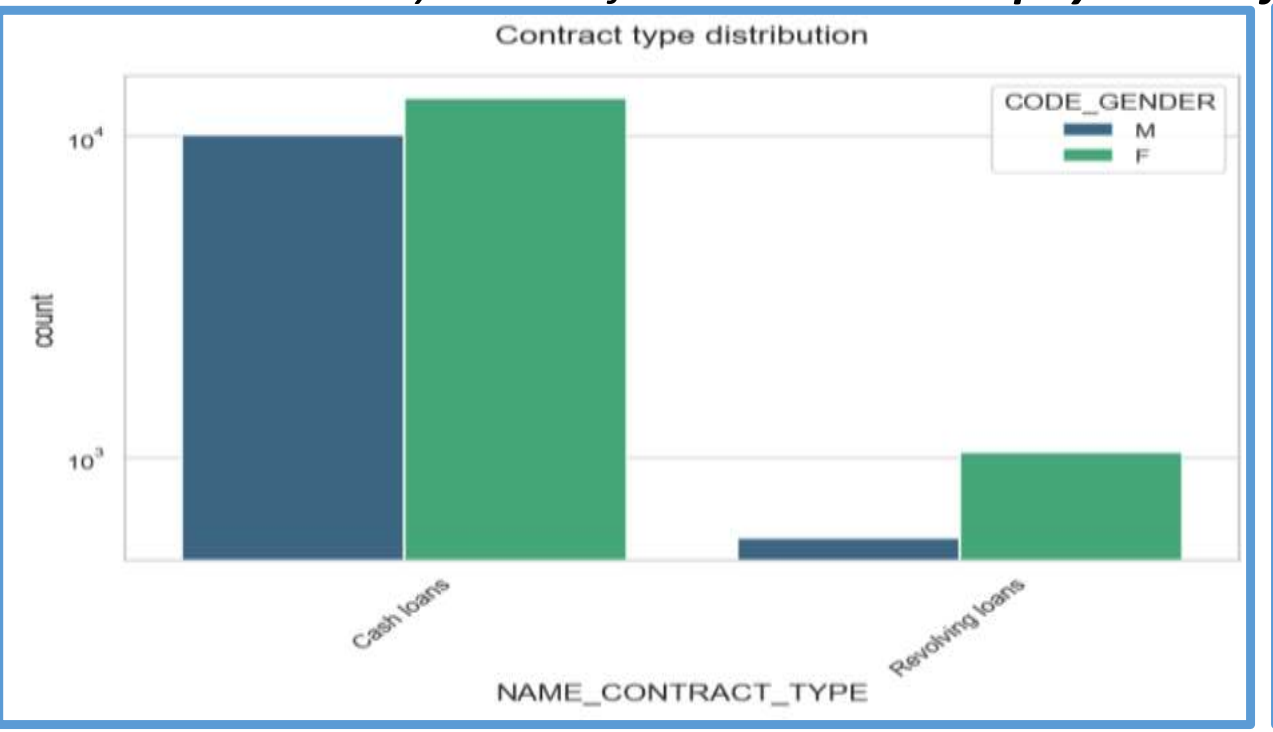
Points that can be concluded from above graph:

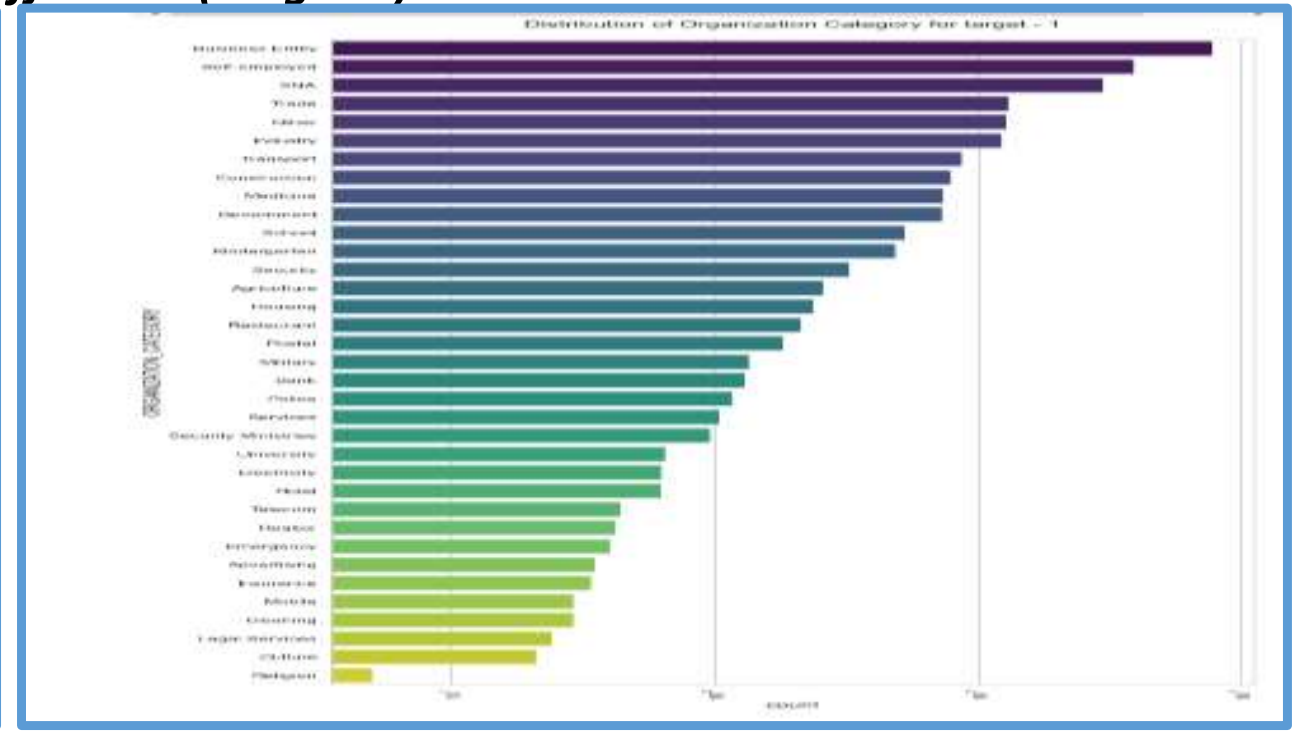
1. **For income type working, commercial associate, Pensioner and State Servant the number of credits are higher than others.**
2. **Less number of credits for income type Unemployed and Maternity leave.**
3. **Most important thing to notice here that Student & Businessman are not defaulters and very less difficulty in payment.**

EDA Group Project

Univariate Analysis:

*The univariate analysis done for consumers with **payment difficulties(Target 1)**:*





Points that can be concluded from above graph:

1. **For contract type 'cash loans' number of credits is much higher than 'Revolving loans' contract type.**
2. **Number of female credits is more than men**

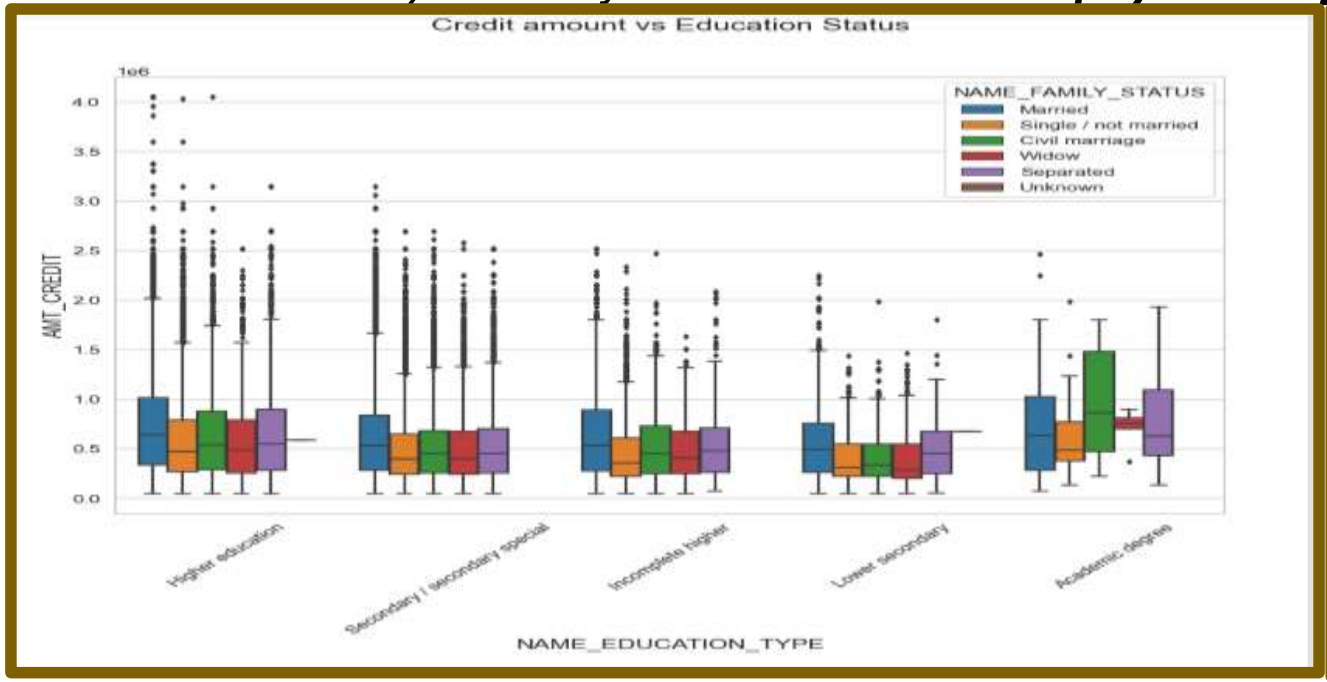
Points that can be concluded from above graph:

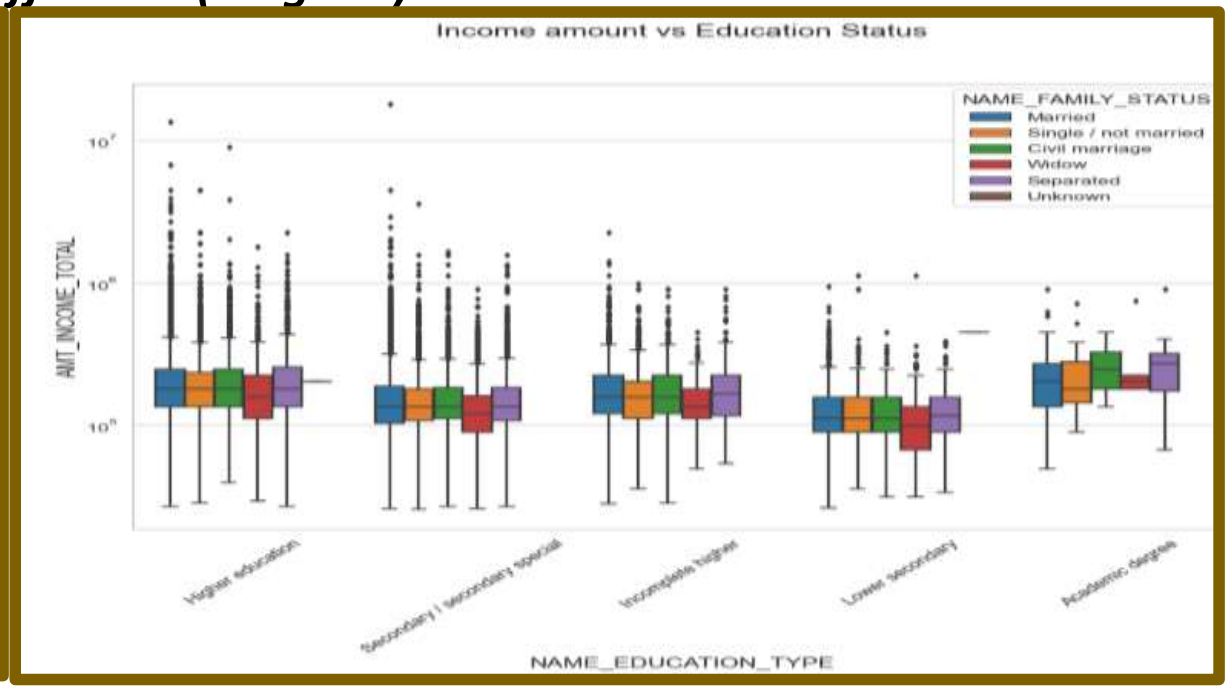
1. **Most the applicants are from Business, Self employed, Other, Industry and Trade category.**
2. **Less number of applicants are from Religion, Culture, Legal Services**

EDA Group Project

Bivariate Analysis:

*The univariate analysis done for consumers with **no payment difficulties(Target 0)**:*





Points that can be concluded from above graph:

1. **Family status of 'civil marriage', 'married' and 'separated' of Academic degree education are having higher number of credits than others.**
2. **Higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers.**
3. **Lower secondary and Academic degree have very less outliers.**
4. **For mostly all education type, Marries family status have high number of credits**

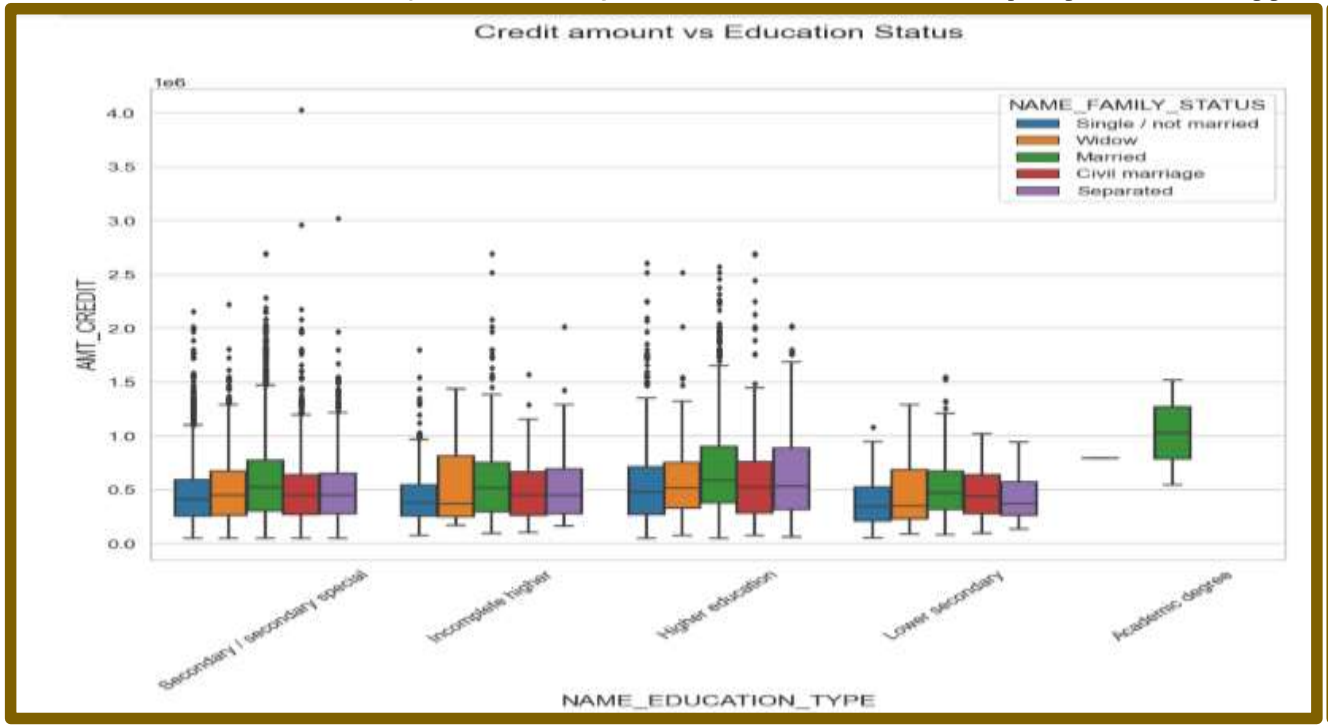
Points that can be concluded from above graph:

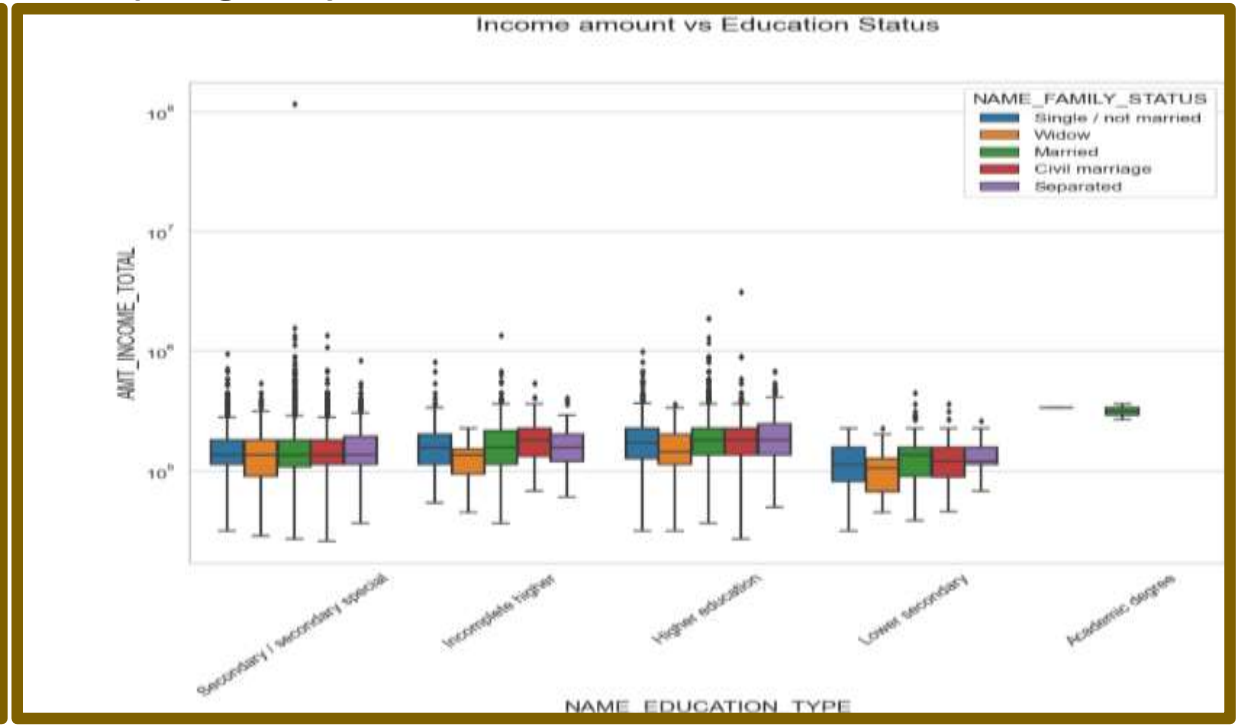
1. **For education type 'Higher education' the income amount is mostly equal with family status.**
2. **Higher education have more outliers while Academic degree have very less.**
3. **Academic degree have high income except family status Widow**

EDA Group Project

Bivariate Analysis:

*The univariate analysis done for consumers with **payment difficulties(Target 1)**:*





Points that can be concluded from above graph:

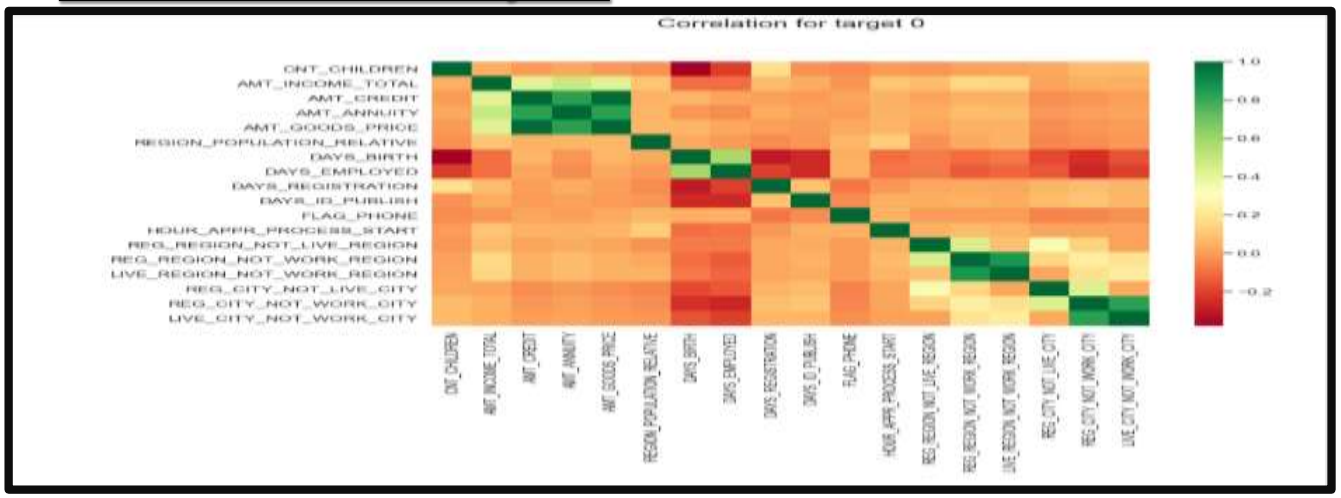
1. **Highest credits is for family status Marries for education type Academic degree.**
2. **Most of the outliers are from Education type 'Higher education' and 'Secondary'.**

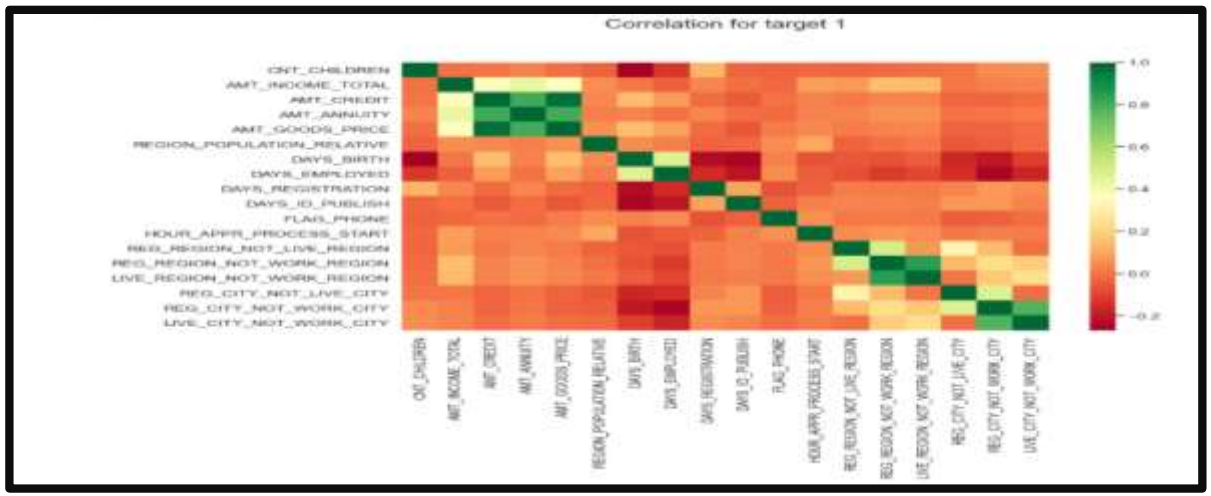
Points that can be concluded from above graph:

1. **Education type 'Higher education' the income amount is mostly equal with family status**
2. **For education type Academic degree, majorly. Married family status have high credit. Income is also high.**

EDA Group Project

Correlation Analysis :





Points that can be concluded from above graph (With no payment difficulties Target 0):

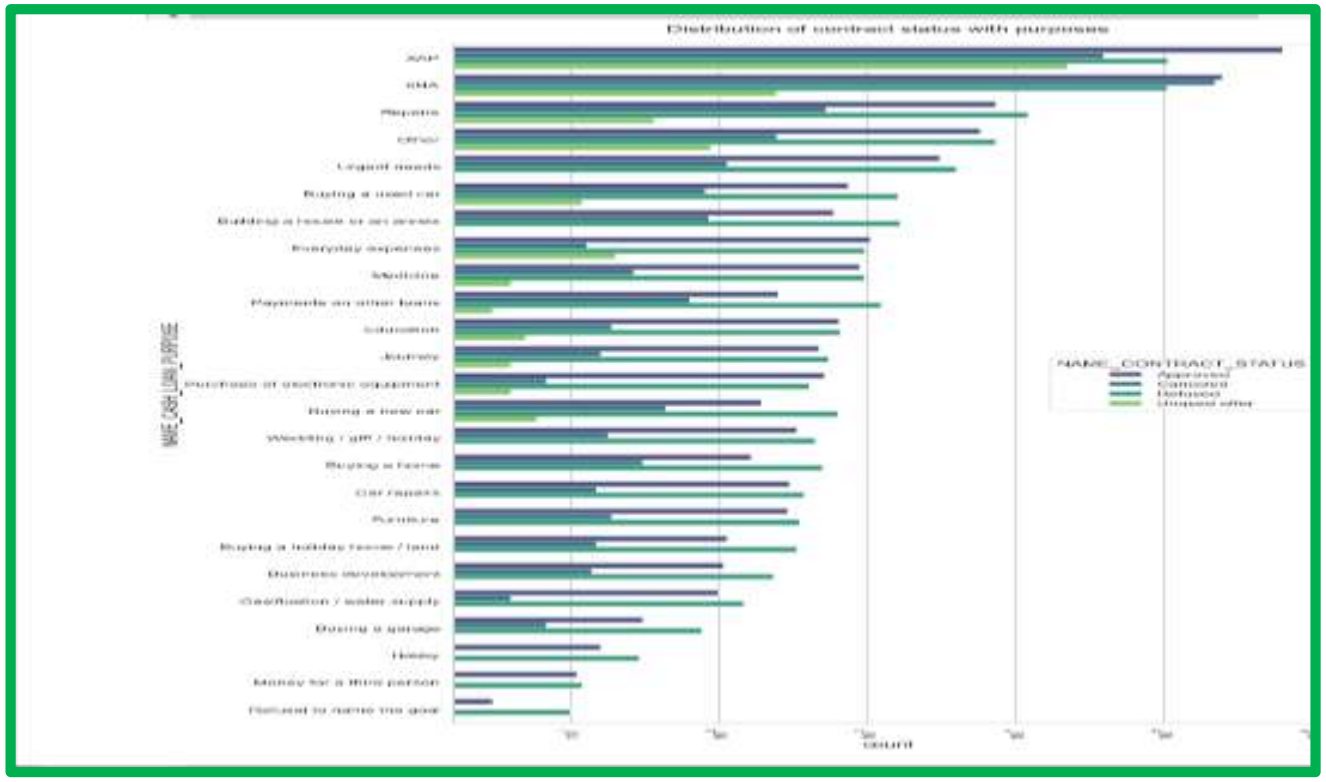
1. **Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.**
2. **Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.**
3. **Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.**
4. **Credit amount is higher to densely populated area.**
5. **The income is also higher in densely populated area**

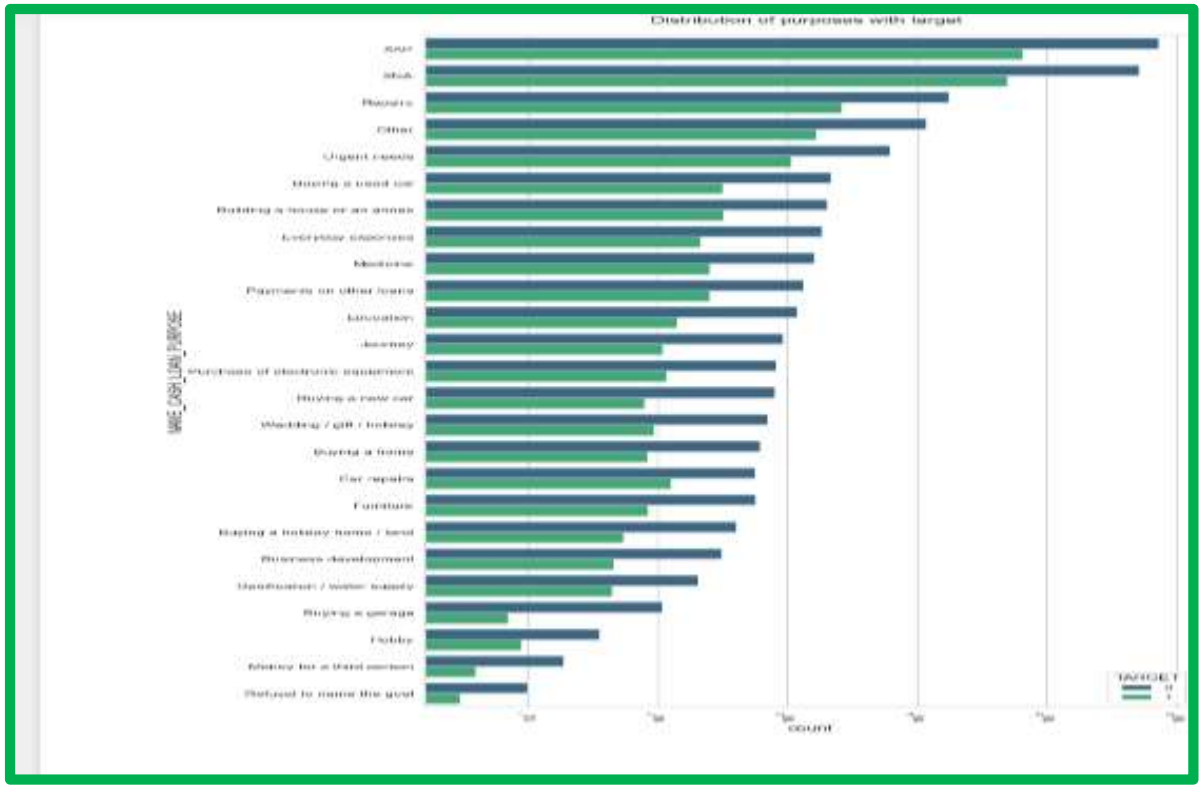
Points that can be concluded from above graph (With payment difficulties Target 1):

1. **The client's permanent address does not match contact address are having less children and vice-versa.**
2. **The client's permanent address does not match work address are having less children and vice-versa.**

EDA Group Project

Merging: Univariate Analysis





Points that can be concluded from above graph:

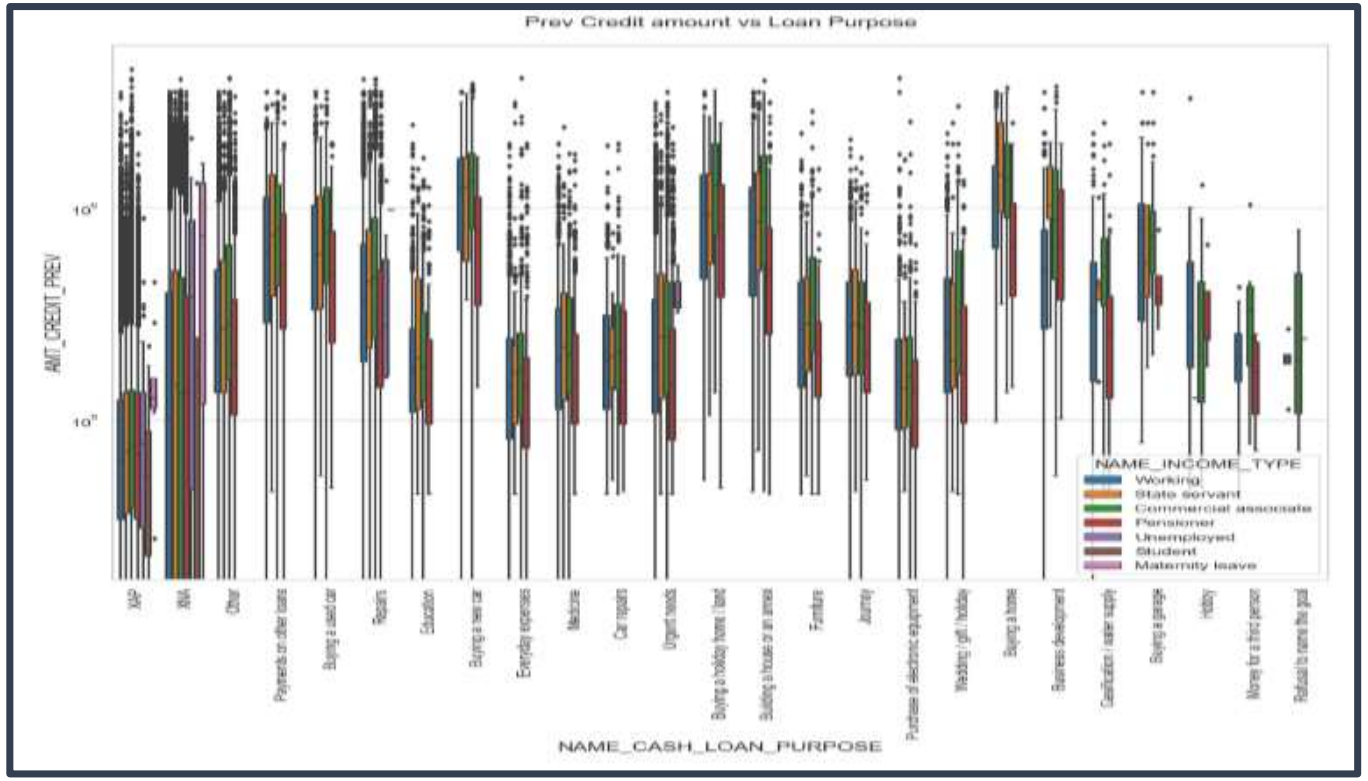
1. Most rejection of loans came from purpose 'repairs'.
2. For education purposes we have equal number of approves and rejection.
3. Paying other loans and buying a new car is having significant higher rejection than approves.

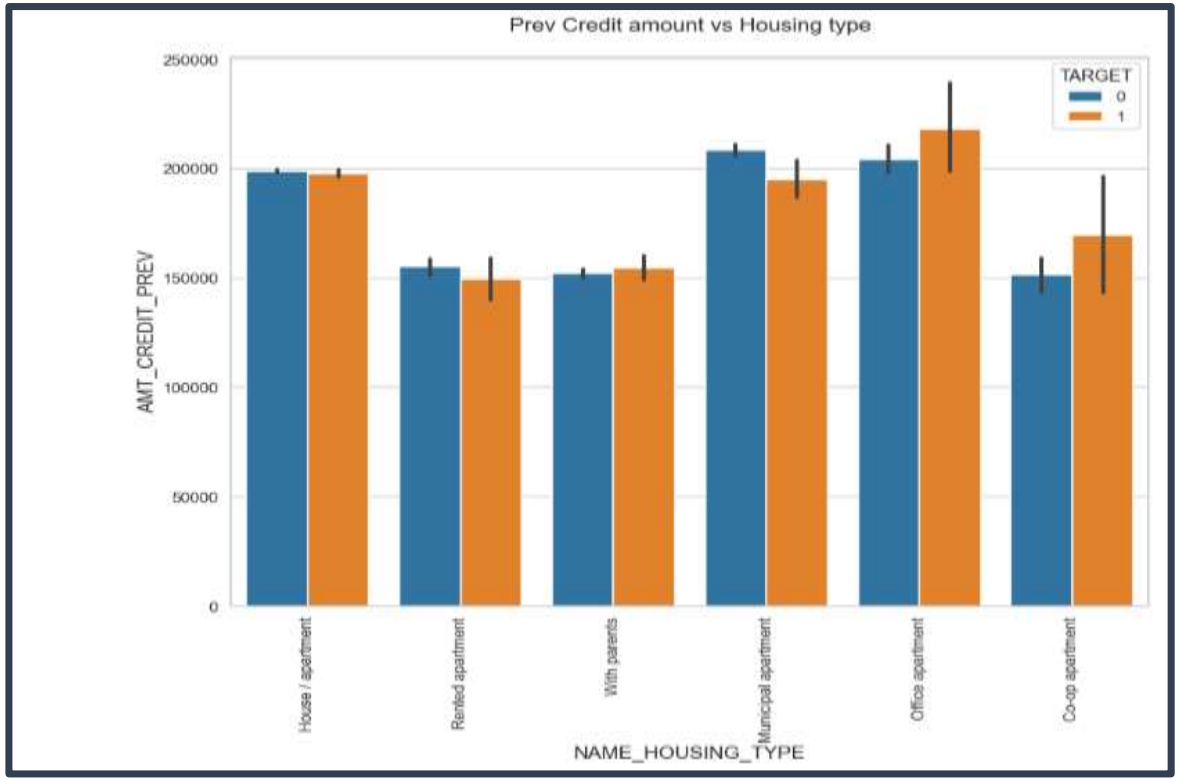
Points that can be concluded from above graph:

1. Loan purposes with 'Repairs' are facing more difficulties in payment on time..
2. There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education' Hence we can focus on these purposes for which the client is having for minimal payment difficulties.

EDA Group Project

Merging: Bivariate Analysis





Points that can be concluded from above graph:

1. **Income type of state servants have a significant amount of credit applied'.**
2. **The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.**

Points that can be concluded from above graph:

1. **For housing type House apartment and Municipal apartment have higher credit of target 0(no payment difficulties). It means less number of defaulters..**
2. **For housing type Office apartment and co-op apartment have more credits in target 1(with payment difficulties). It means more number of defaulters..**

EDA Group Project

Recommendations:

- More focus should be on clients from housing type ‘Rented apartment’, ‘Municipal apartment' as they are having least number of unsuccessful payments.
- Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ and 'Office apartment' for successful payments.
- Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
- Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.

Thank You