CREDIT EDA CASE STUDY

- By POOJA KUMARI

EDA(Exploratory Data Analysis)

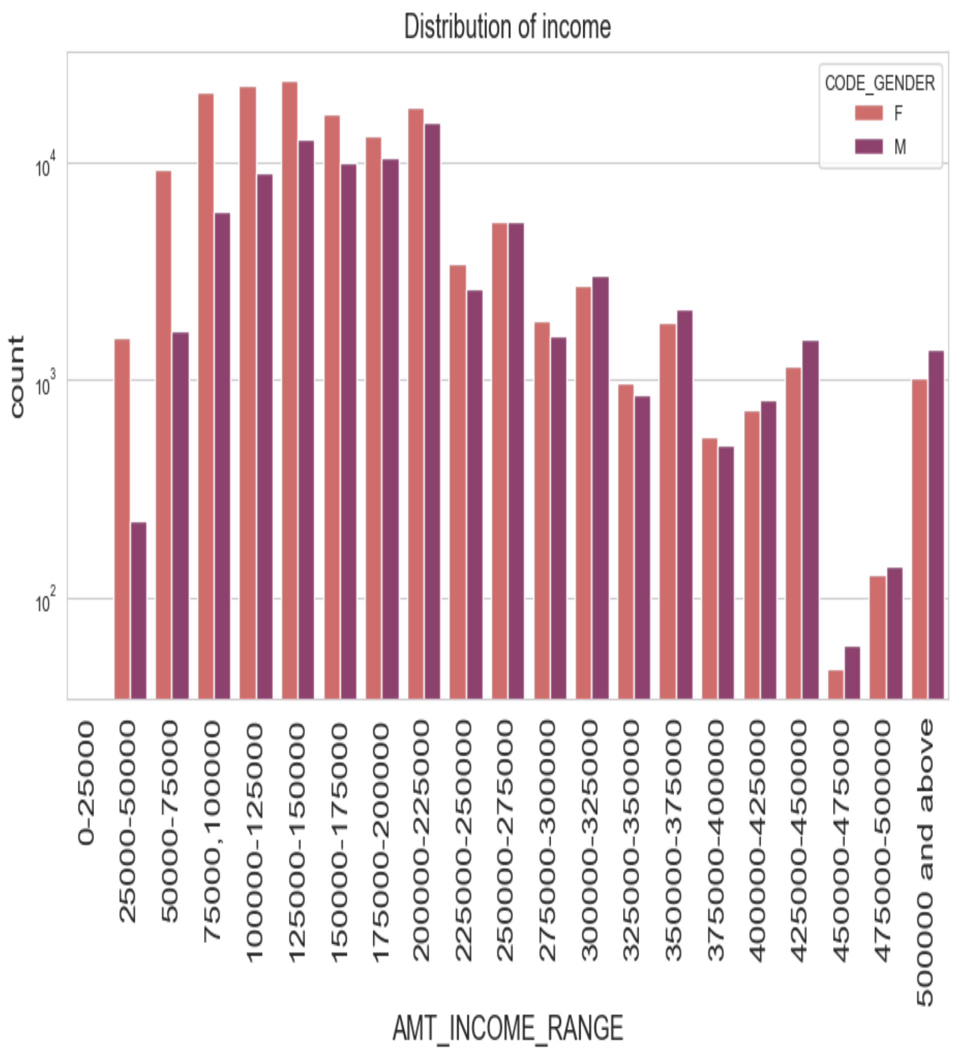
- EDA is an important part of any data analysis, in EDA we always need to investigate the quality of our data. Data cleaning is just one application of EDA: we ask questions about whether our data meets our expectations or not. To do data cleaning, we'll need to deploy all the tools of EDA: visualisation, transformation, and modelling.
- Our goal during EDA is to develop an understanding of our data. The easiest way to do this is to use questions as tools to guide our investigation. When we ask a question, the question focuses our attention on a specific part of our dataset and helps us decide which graphs, models will be more usefull.

CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_0

Distribution of income range

Points to be concluded from the graph.

1. Female counts are higher than male.
2. Income range from 750000 to 200000 is having more number of credits.
3. Females are more than male is having credits.
4. Very less count for income range between 425000 and 475000.



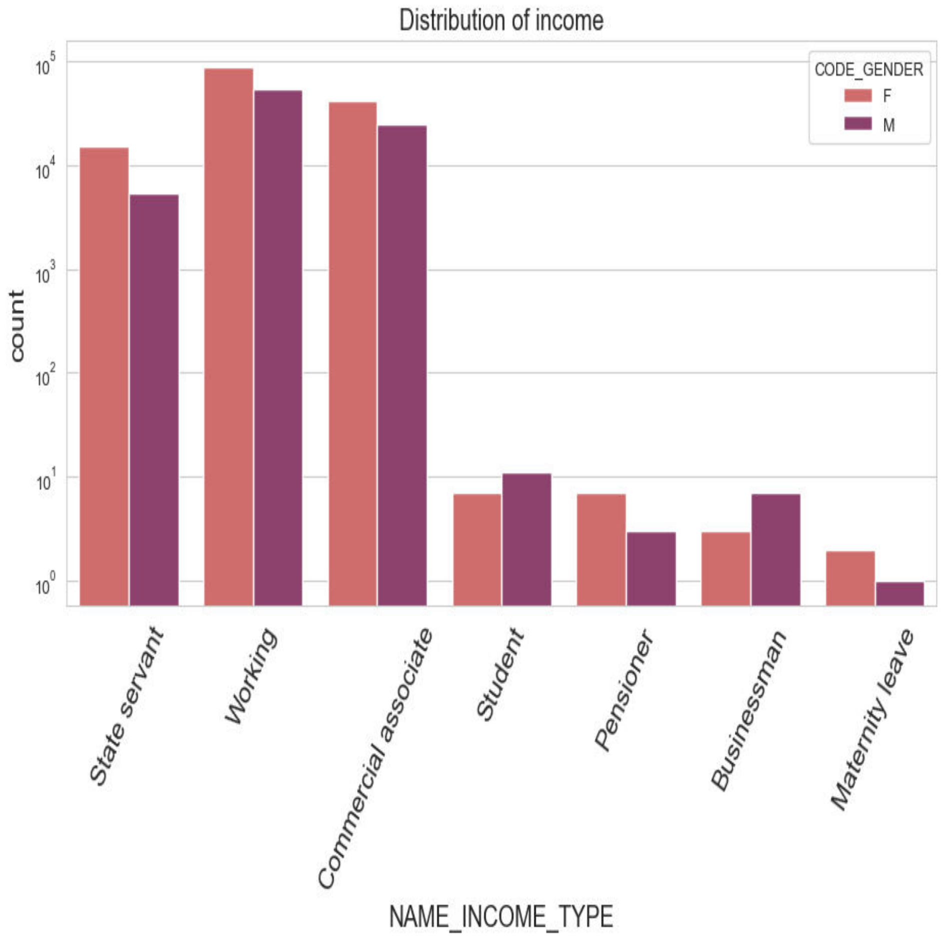
Distribution of income type

Points to be concluded from the graph.

1. Working Female have more credit counts than others.

2. Student, Pensioner, Businessman and Maternity leave have less number of credit counts.

3. State Servant, Working and Commercial Associate have more credit counts than others.

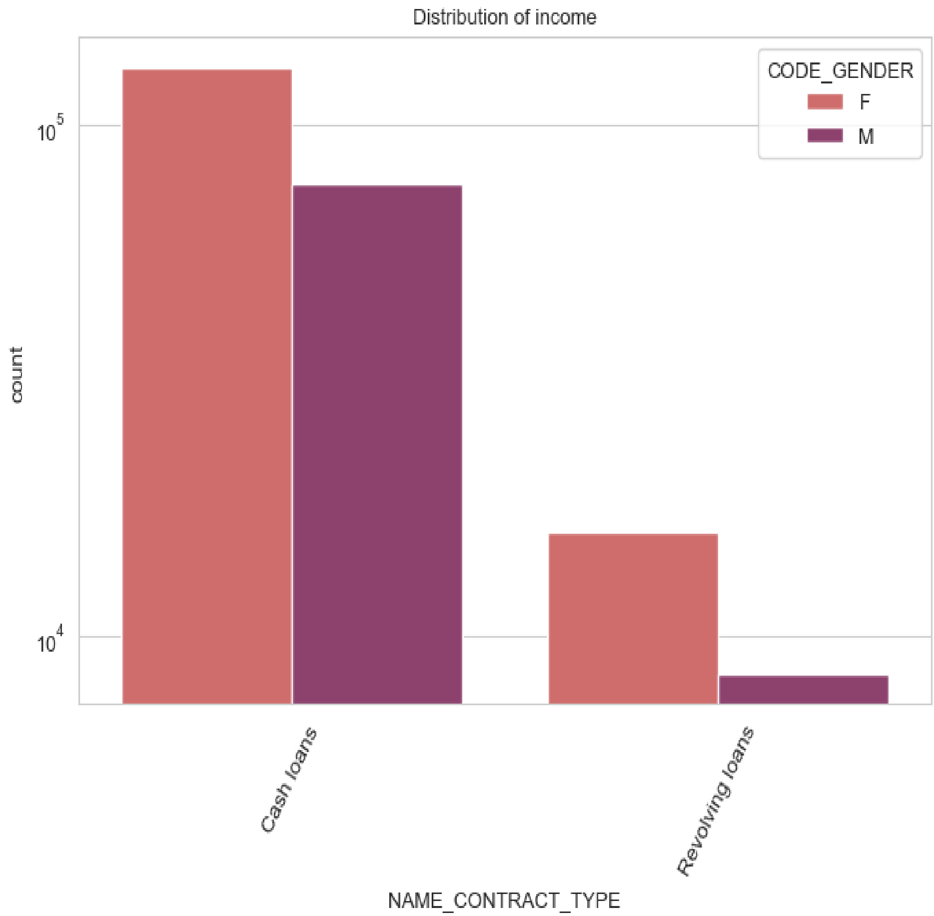


Distribution of Contract type

conclusion from the graph.

1. Here Cash loans is having higher number of credit counts than Revolving loans in contract type graph.

2. In this Female has higher counts than Males for applying credits.



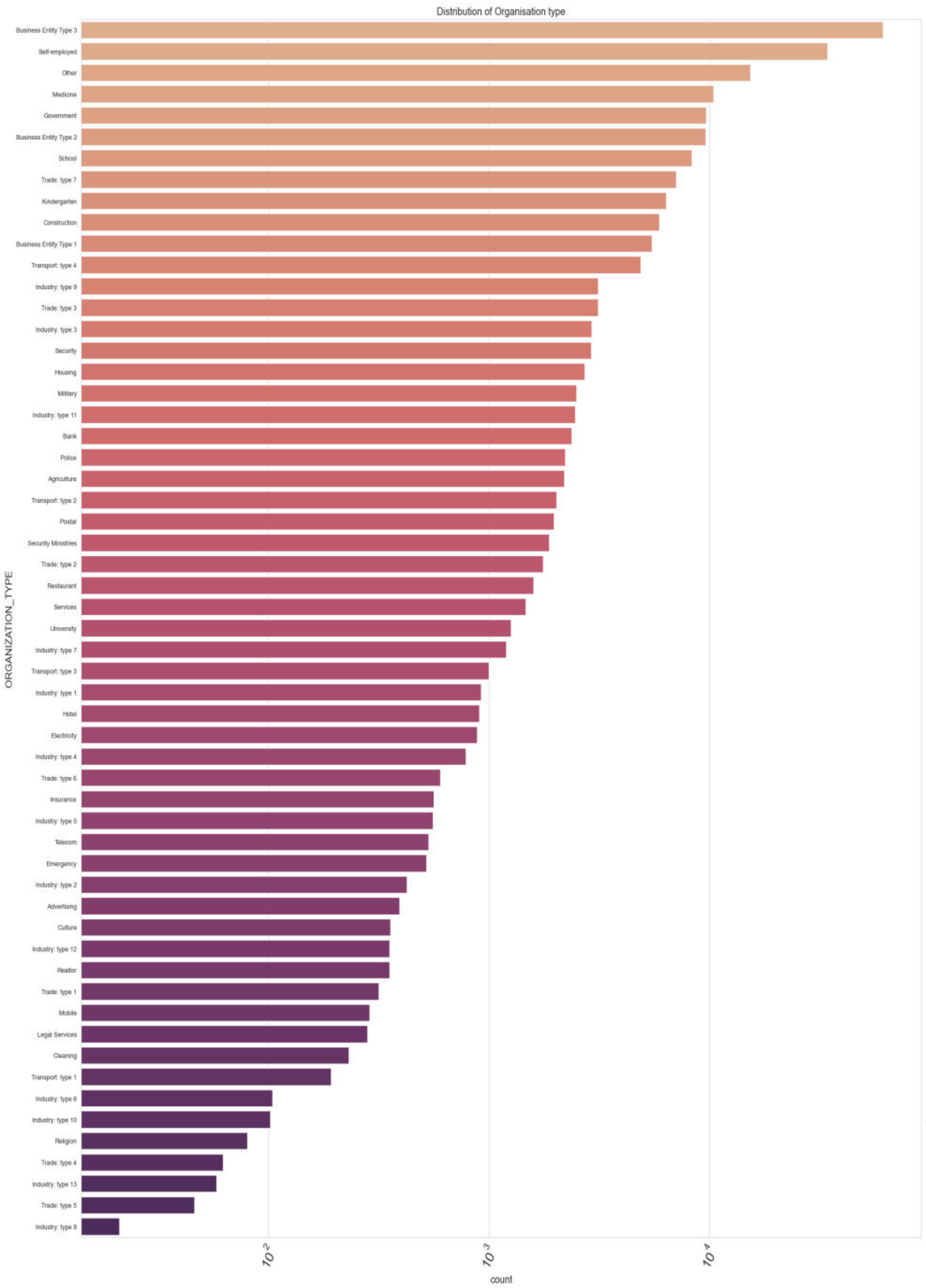
Categorical Univariate Analysis for target1

Distribution of Organisation type

Conclusion from graph.

1. Less clients are from Industry type 8, type 6, type 10, religion and trade type 5, type 4.

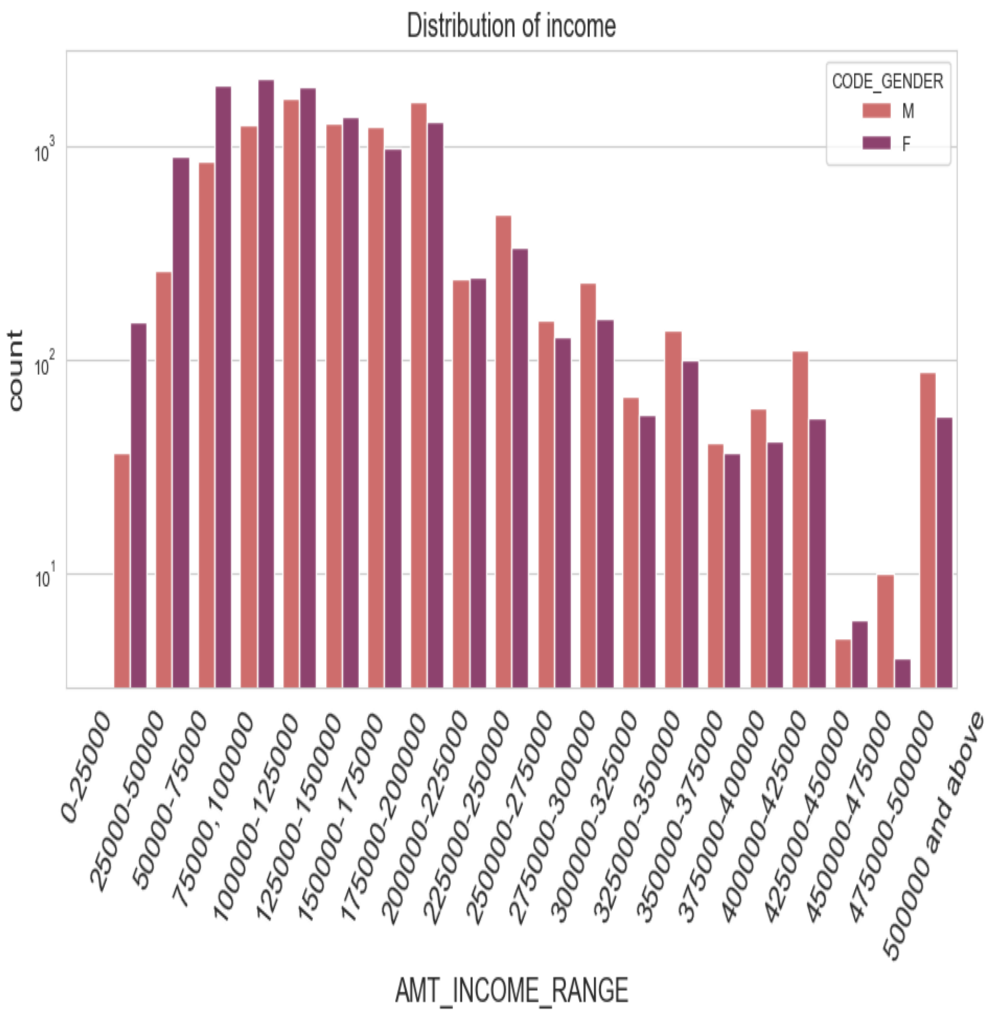
2. Clients which have applied for credits are mostly from the organization type Business entity Type 3 , Self employed , Other, Medicine and Government.



Distribution of income range

Conclusion from graph.

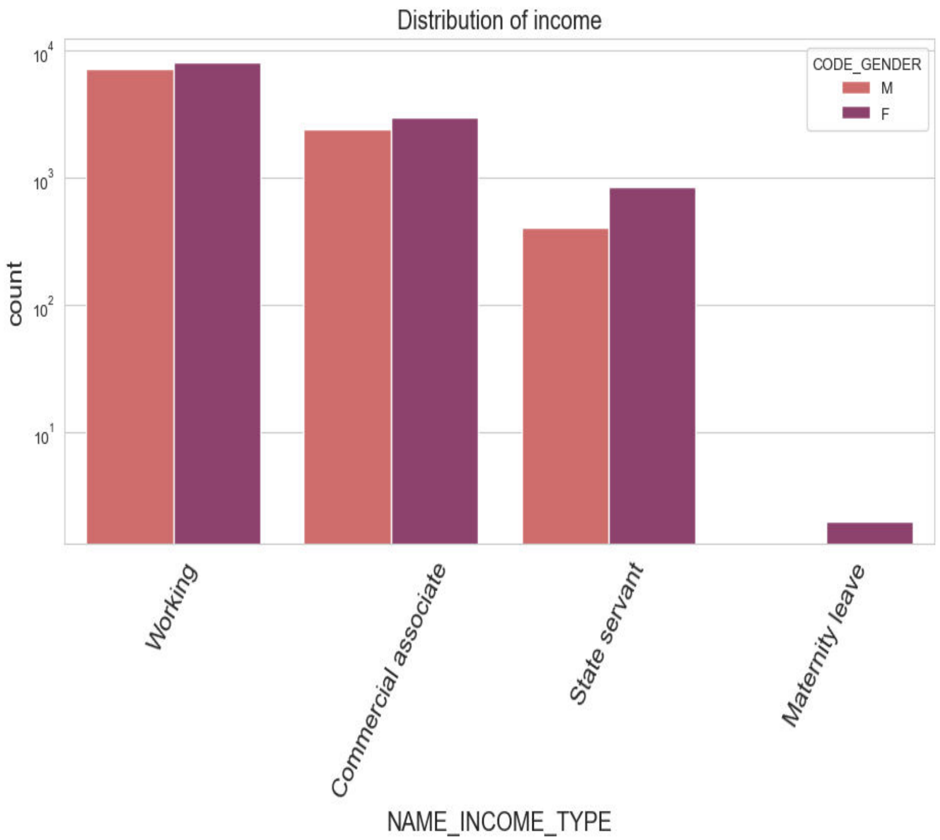
1. Male counts are higher than female.
2. Income range from 75000 to 200000 having more number of credits.
3. This graph shows that males having more credits than female.
4. Very less count for income range 425000 and above.



Distribution of income type

conclusion from above graph.

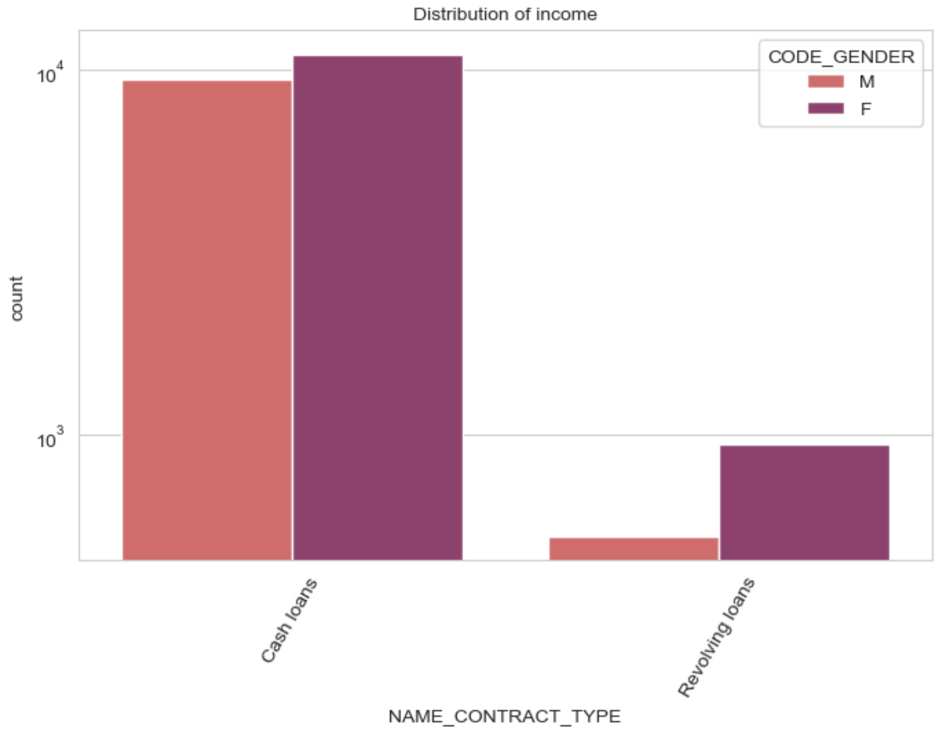
1. In income type working, commercial associate, and State Servant credits are higher than other.
2. Less number of credits for Maternity leave.
3. Females are having more number of credits than male.



Distribution of Contract type

Conclusion from above graph.

1. Female are more applying for credits.
2. Cash loans having higher credits than Revolving loans.

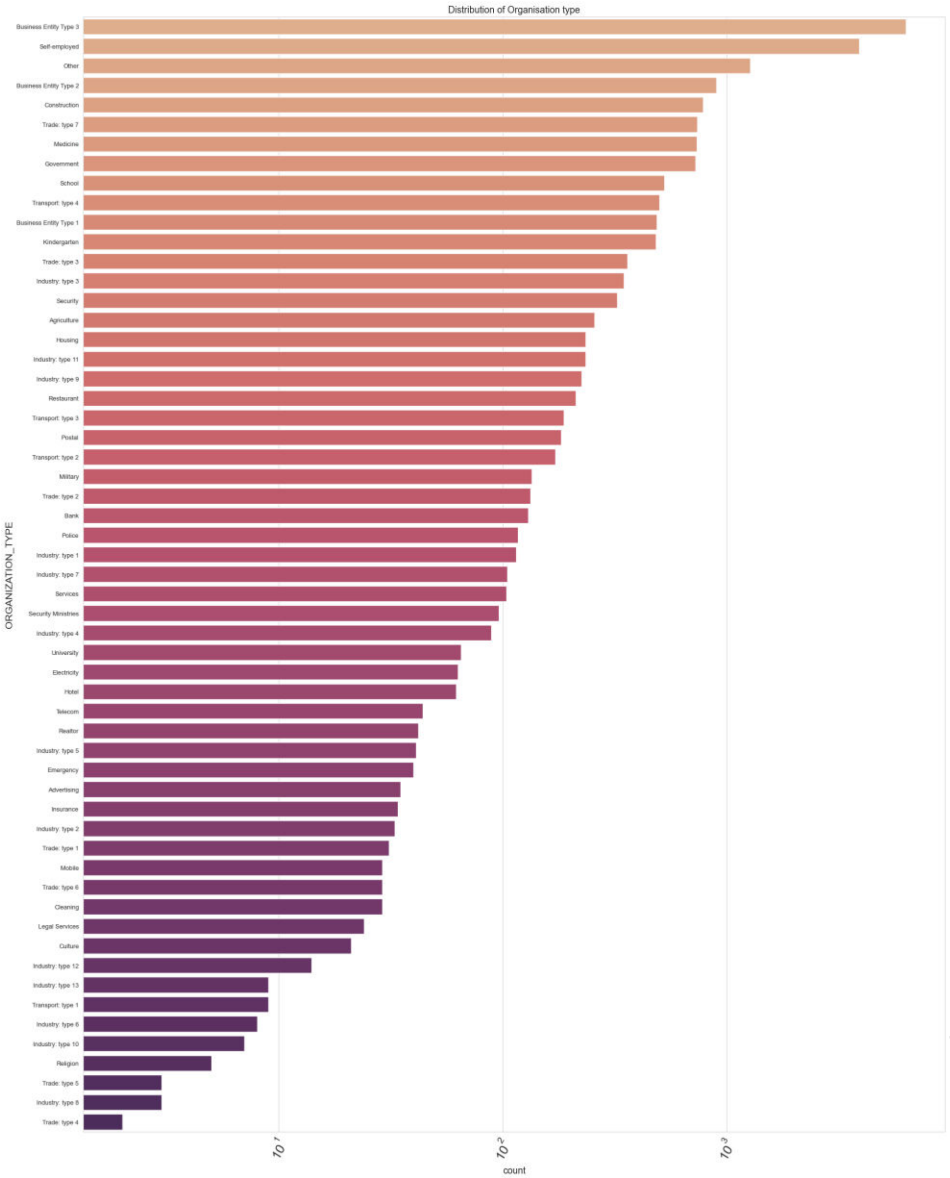


Distribution of Organisation type

conclusion from the above graph.

1. Clients which have applied for credits are most of the from Business entity Type 3, Self employed, Other, Medicine and Government.

2. Clients are less from Industry type 8, type 6, type 10, religion and trade type 5, type 4.



Correlation of target_0

Conclusion from above correlation heatmap

1. Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.

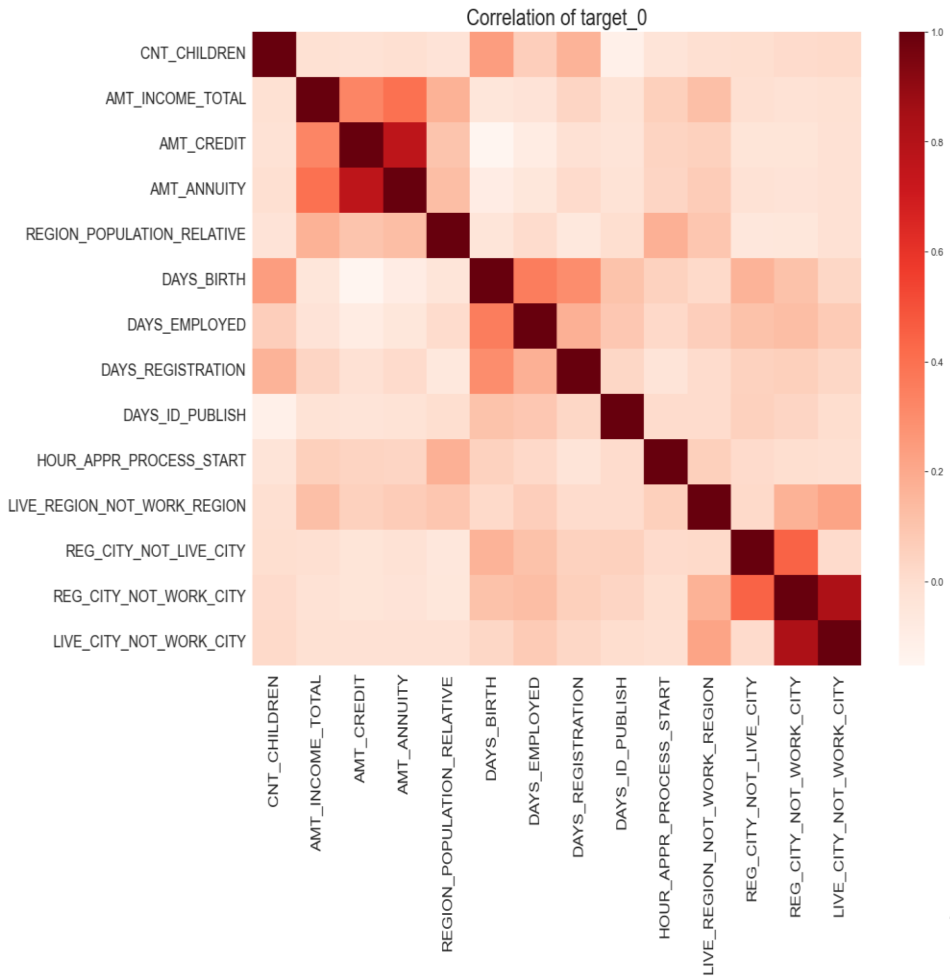
2. Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.

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. less children client have in densely populated area.

5. Credit amount is higher to densely populated area.

6. The income is also higher in densely populated area.

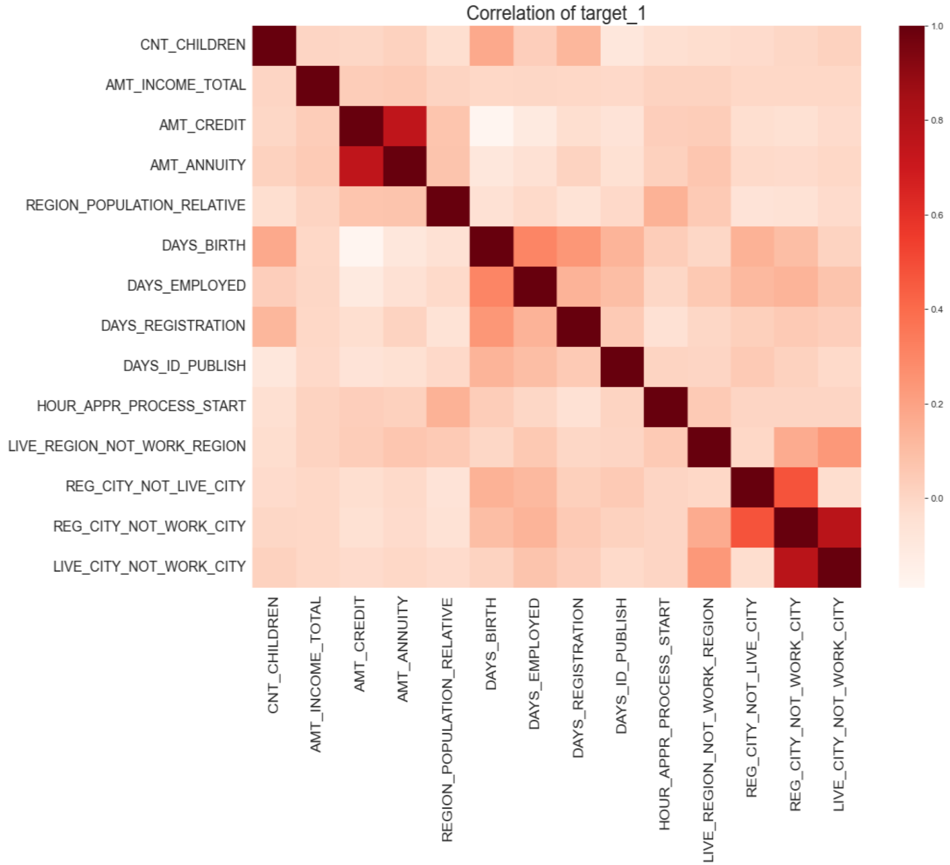


Correlation of target_1

Conclusion from above correlation heatmap

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

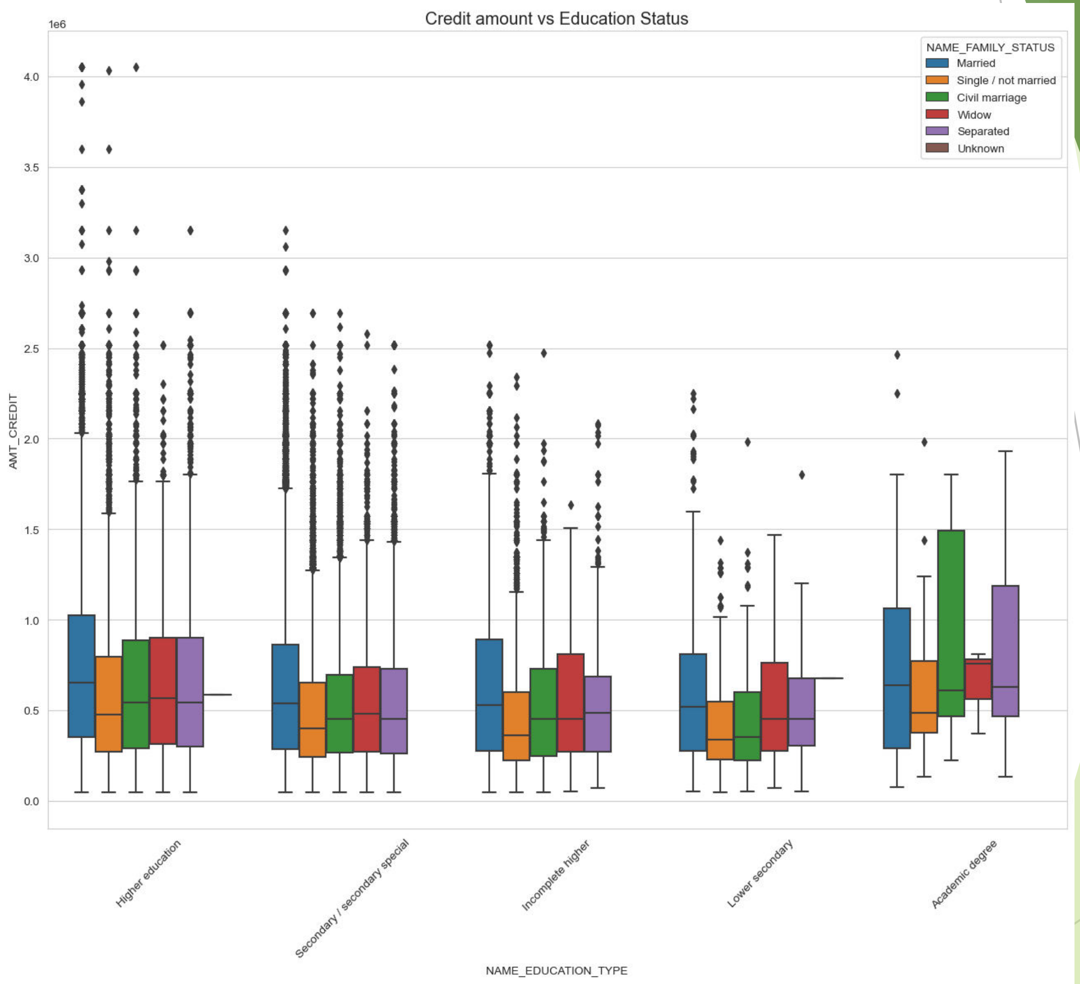
2. the client's permanent address does not match work address are having less children and vice-versa



Bivariate analysis for numerical variables

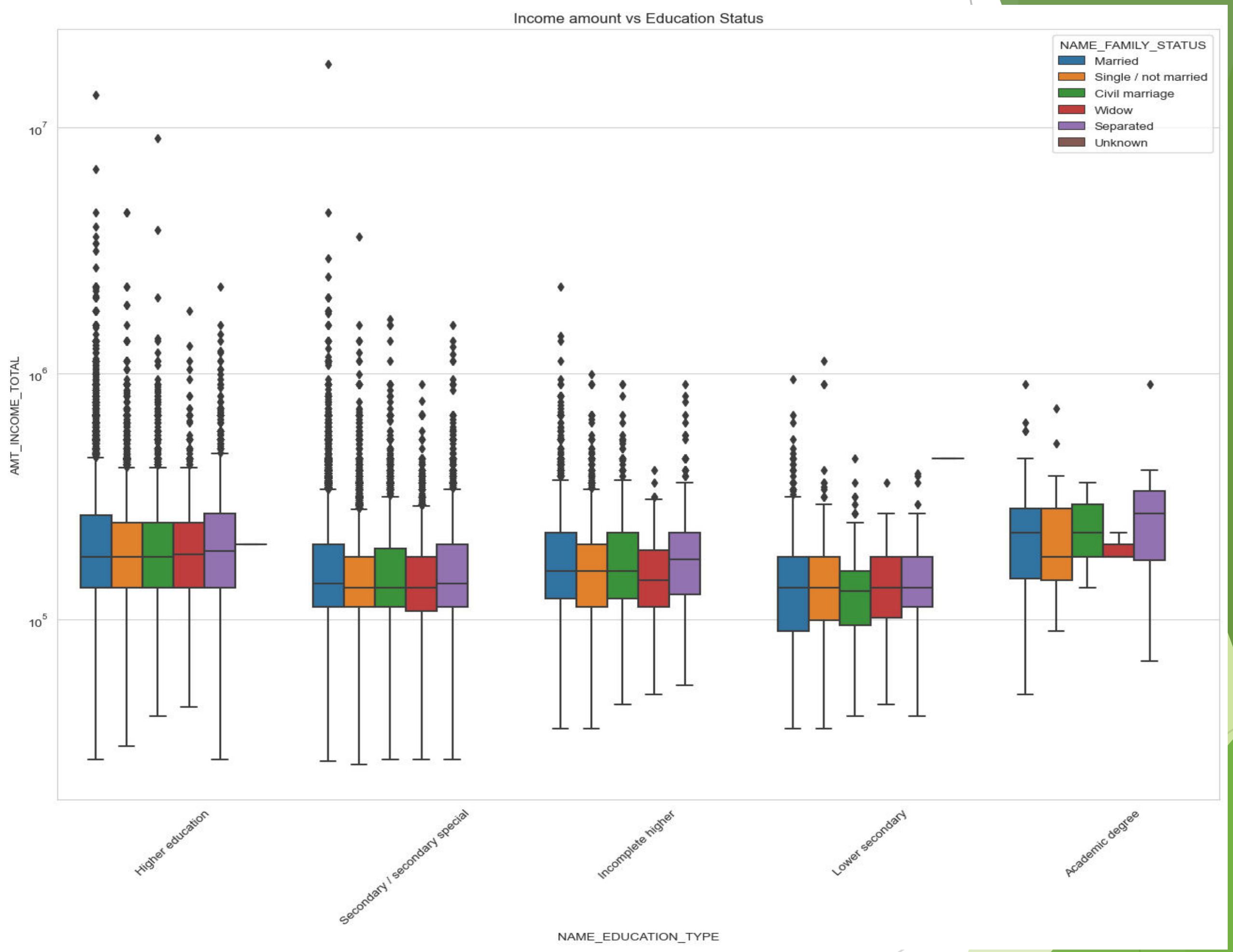
Credit amount vs Education Status

From the box plot we can conclude that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others. Also, higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile.



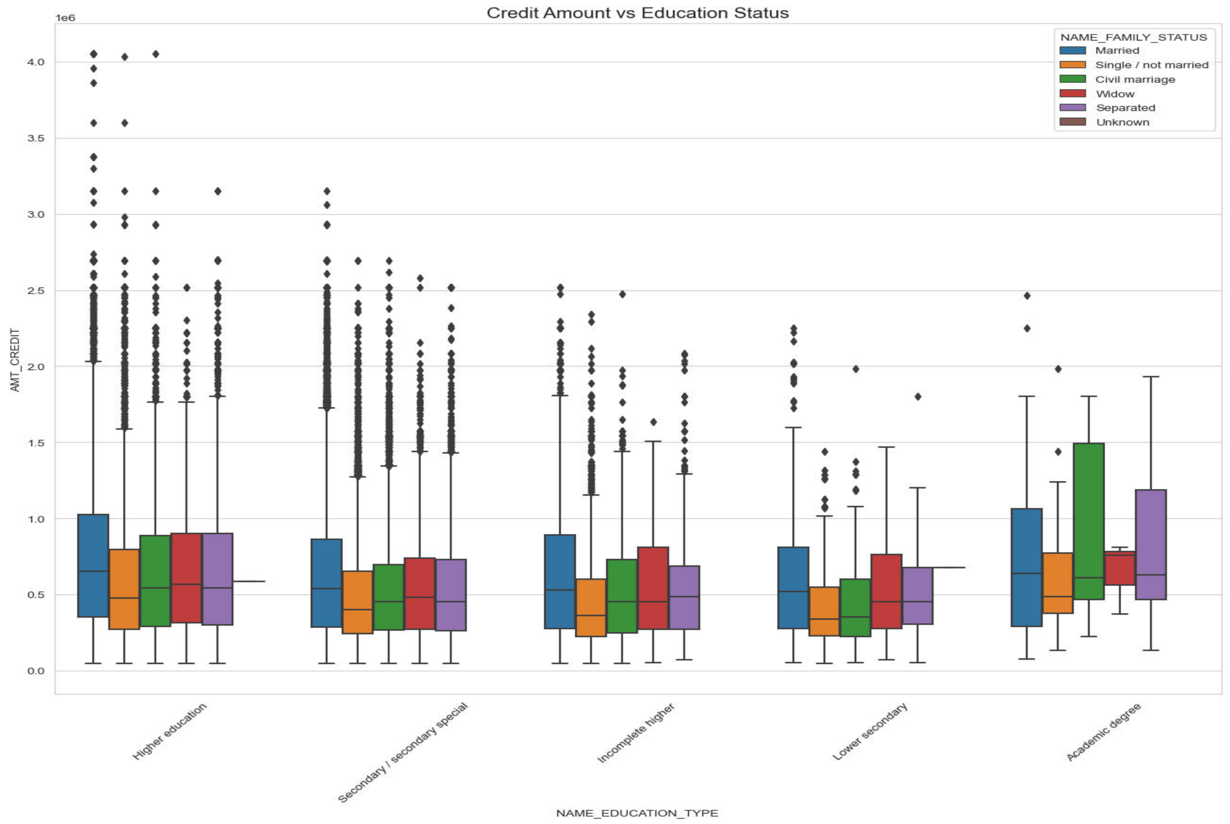
Income amount vs Education Status

From the boxplot for Education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers. Less outlier are having for Academic degree but there income amount is little higher that Higher education. Lower secondary of civil marriage family status are have less income amount than others.



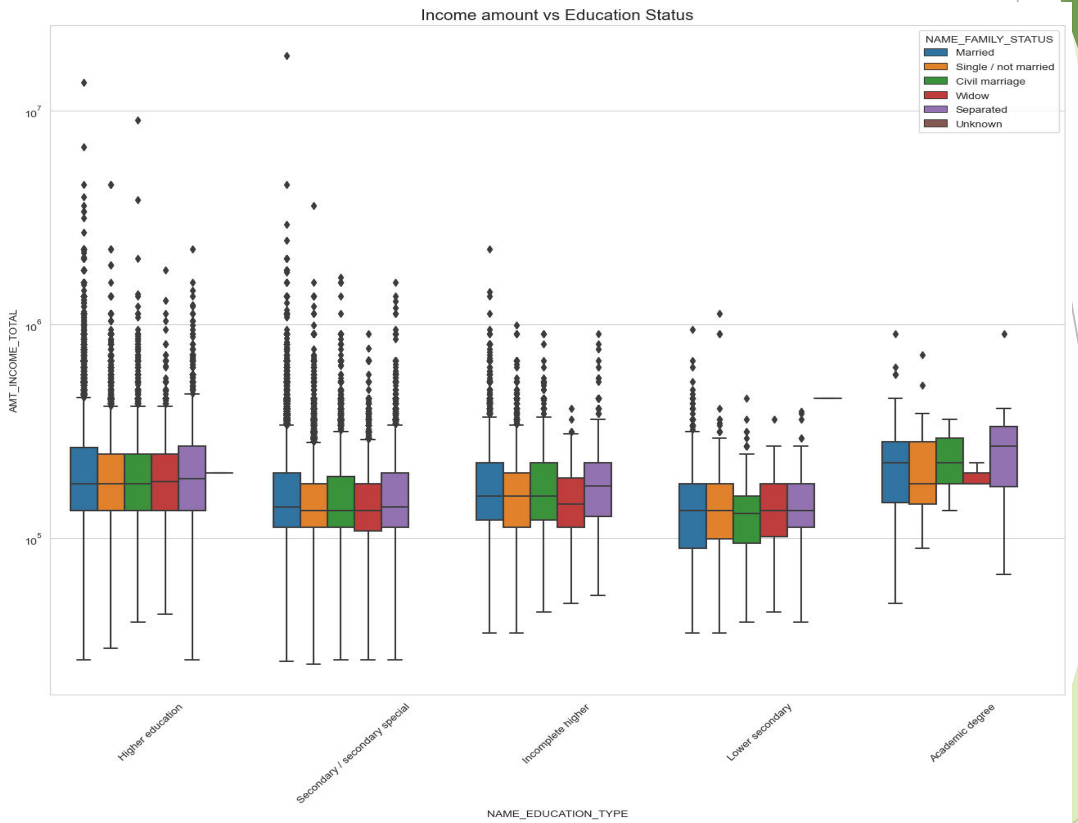
Credit Amount vs Education Status

It is similar with Target 0 From the box plot we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others. Most of the outliers are from Education type 'Higher education' and 'Secondary'. Civil marriage for Academic degree is having most of the credits in the third quartile.



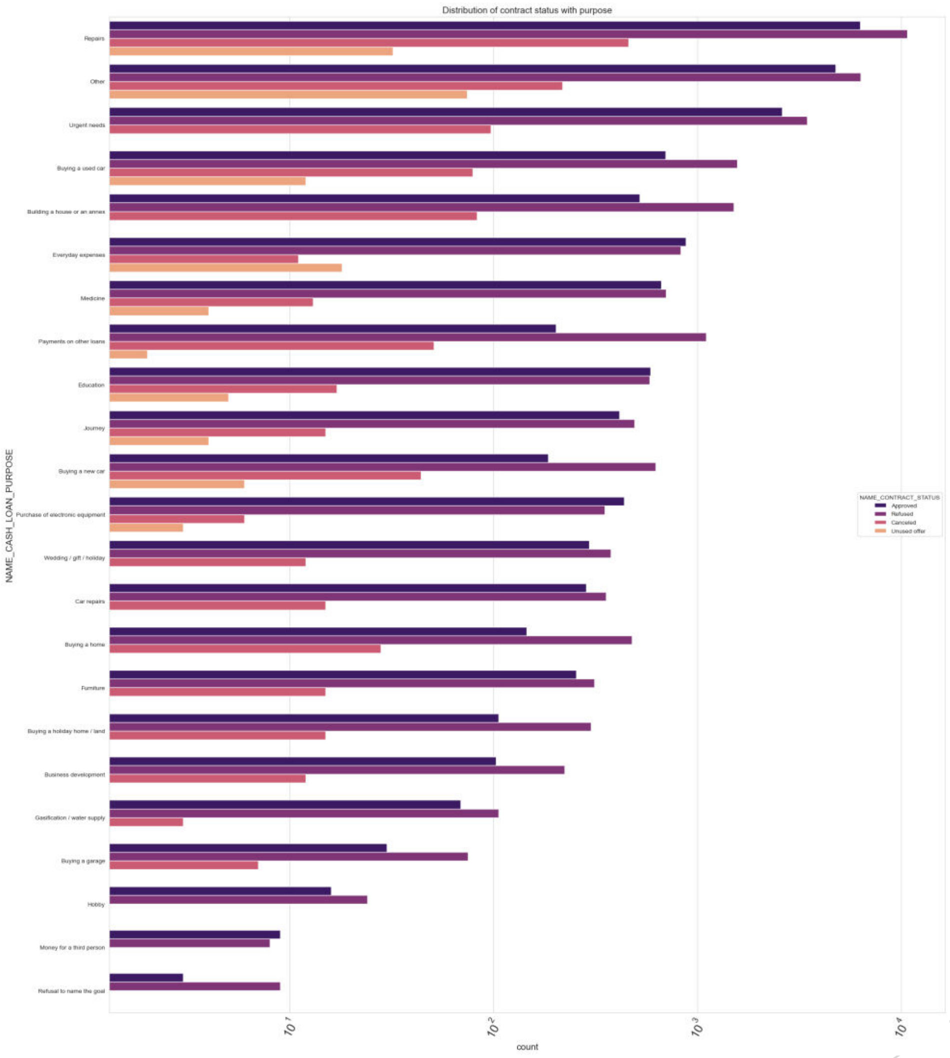
Income amount vs Education Status

Have some similarity with Target0, From boxplot for Education type 'Higher education' the income amount is mostly equal with family status. Less outlier are having for Academic degree but there income amount is little higher that Higher education. Lower secondary are have less income amount than others.



Univariate analysis after merging data

Distribution of contract status with purpose



Conclusion 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.

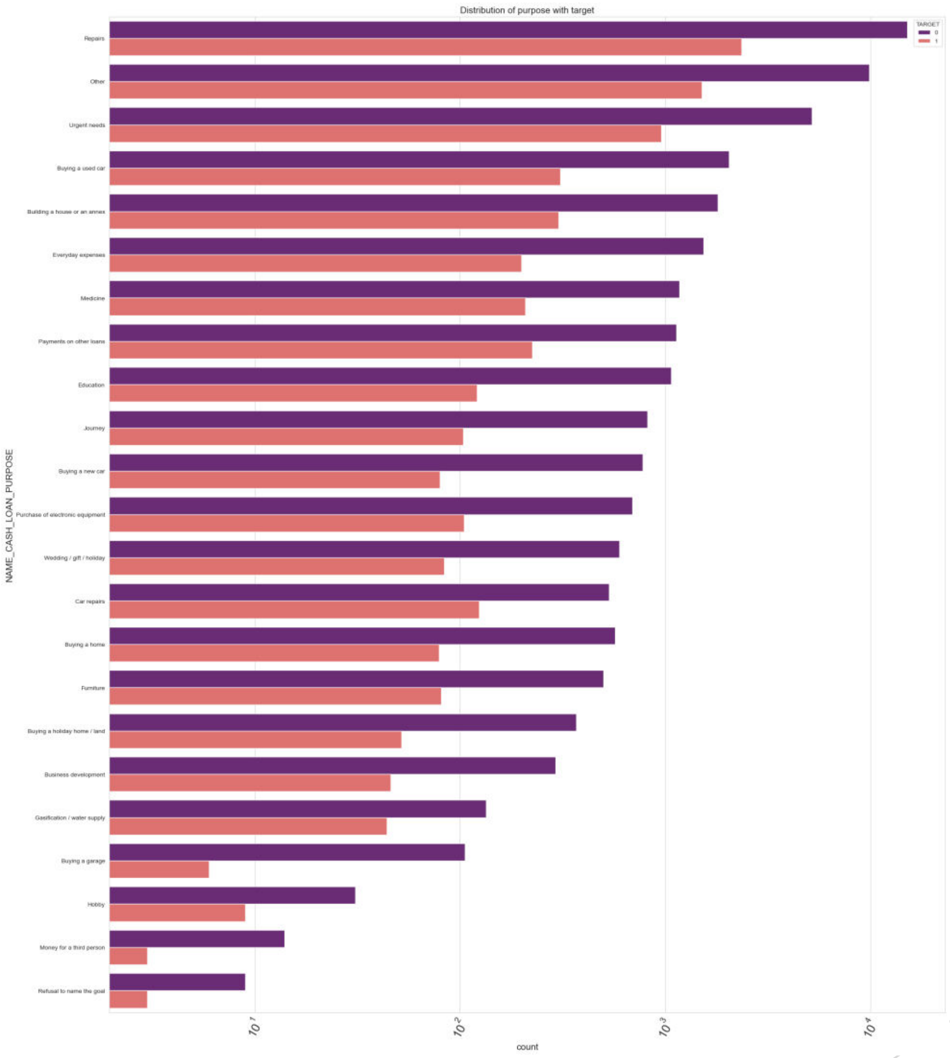
Distribution of purpose with target

Conclusion from Above graph

1. Loan purposes with 'Repairs' are facing more difficulites in payment on time.

2. There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business developemt', 'Buying land','Buying a new car' and 'Education'

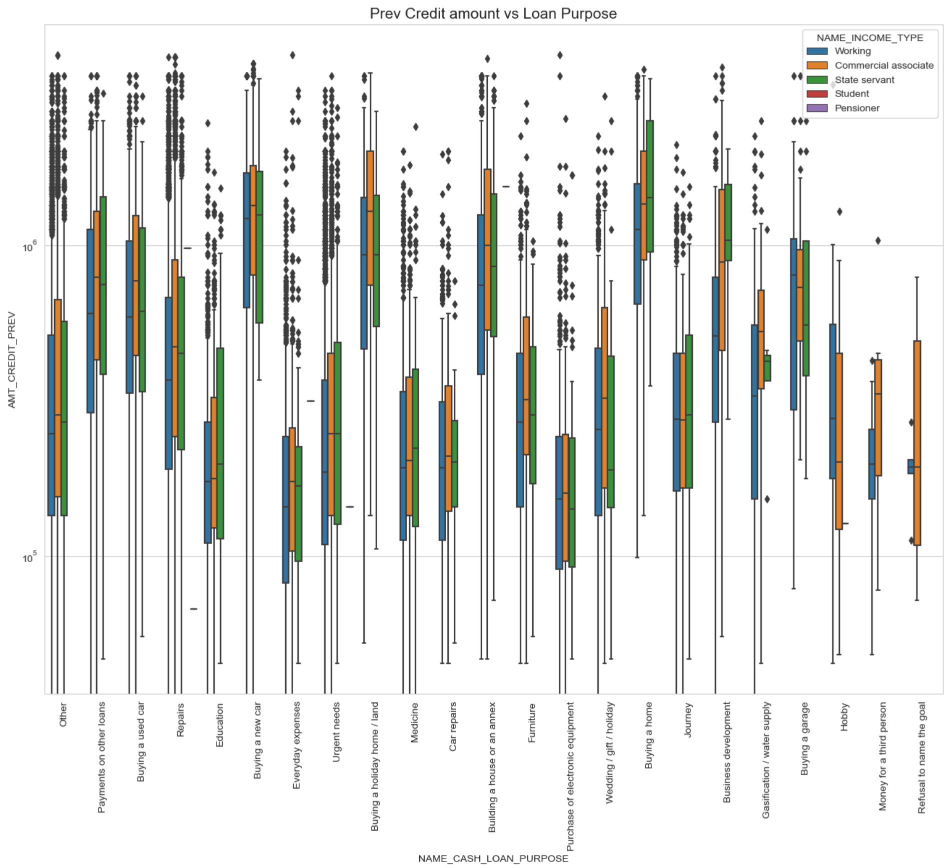
3. Hence we can focus on these purposes for which the client is having for minimal payment difficulties.



Performing bivariate analysis

Previous Credit amount vs Loan Purpose

Conclusion from Above graph



1. The credit amount of Loan purposes like Buying a home, Buying a land, Buying a new car and Building a house is higher.

2. Income type of state servants have a significant amount of credit applied

3. Money for third person or a Hobby is having less credits applied for.

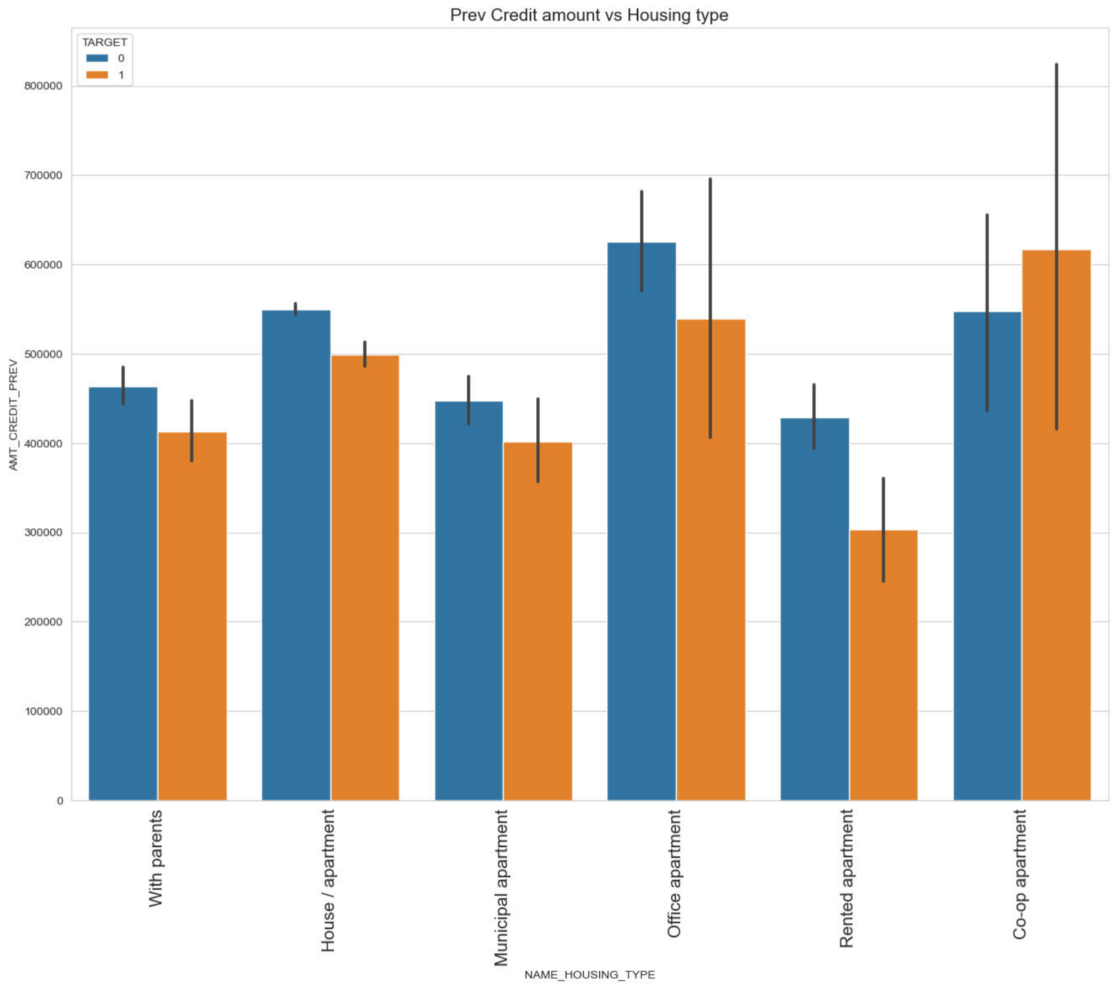
Previous Credit amount vs Housing type

Conclusion from above graph

For Housing type

1. Office appartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1. So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.

2. Bank can focus mostly on housing type with parents or House\appartment or municipal appartment for successful payments.



Conclusion derived from Bank Dataset

- CONCLUSION
- 1. Banks should target more on contract type Student, pensioner and Businessman with successful payments.
- 2. Bank should target on the customers working in other business entity type such as self employed, Accountant, laborers, Managers.
- 3. Banks should target married customer having children not more than 2.
- 4. Banks should target mostly female as they pay loan on time.
- 5. Banks should target People having house/ appartment are safe to give the loan.

THANK YOU