CREDIT EDA ASSIGNMENT

SHYAM RAJAGOPAL

CONTENTS

PROBLEM STATEMENT

GOAL:

The goal is to help the bank to identify if the client has any problems in payment, which gives the bank the option to:

- Approve the loan or not
- Plan new lending schemes
- Reject/deny the loan

WORK PLAN

- Import warnings and libraries
- Import data files
- Identify missing values and null values
- Eliminate the said values
- Check and validate data types
- Handle outliers
- Bin variables
- Univariate analysis
- Bivariate analysis
- Conclusion

IMPORTING WARNINGS AND LIBRARIES

IMPORTING WARNINGS:

Ignores the warnings but highlights them and helps us to run the program

IMPORTING LIBRARIES

Importing numpy, pandas. Matplotlib and seaborn are very effective for data loading and visualization

READING DATASET

- The flag variable is our target variable which enables us to check if the client will pay installments on time or not.
- Two main data files were extracted from the given dataset. namely - 'application_data.csv' and 'previous_data.csv'
- Datafile description, shape etc., has been highlighted in the notebook for elaborated experience in reading the data.

HANDLING DATA, NULL & MISSING VALUES

- Checked for null values in aapplication_data.csv and eliminated 49 columns which had null values more than 40%
- Post that, AMT_ANNUITY, AMT_GOODS_PRICE, EXT_SOURCE_2, NAME_TYPE_SUITE, had less than 1% of null (& numeric) values. Hence, identified outliers and imputed using the best approach available.
- checked for unique values in columns by the following condition:

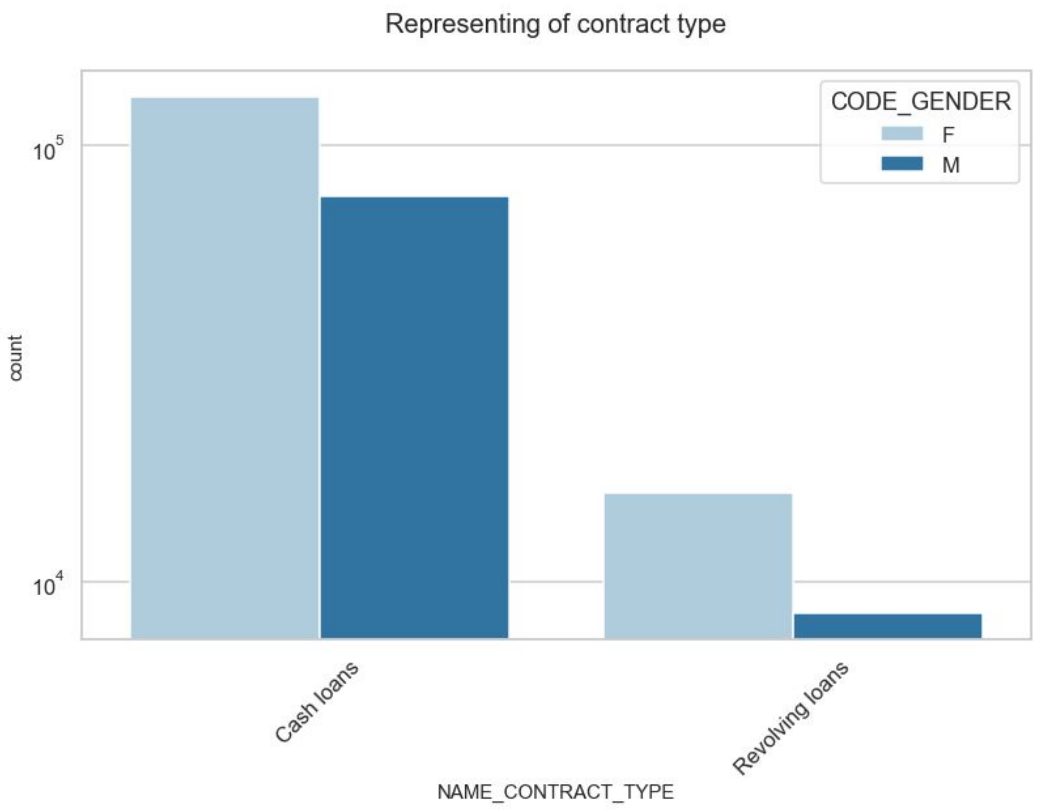
  1) If the count of unique values <=40, it's a categorical column

  2) If the count of unique values > 50, it's a continuous column

UNIVARIATE ANALYSIS

CONTRACT TYPE REPRESENTATION (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0)

- Females are highest in applying Credits
- Revolving loans have lesser Credits than cash loans Contract type

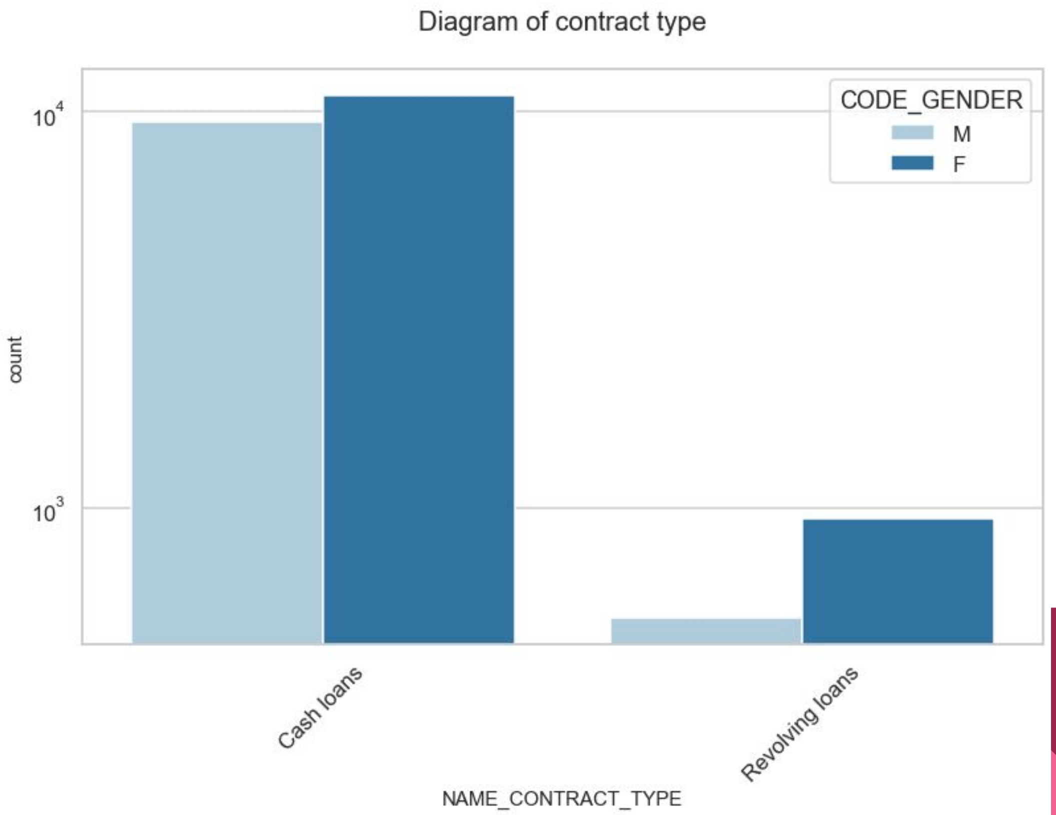


CONTRACT TYPE REPRESENTATION (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1)

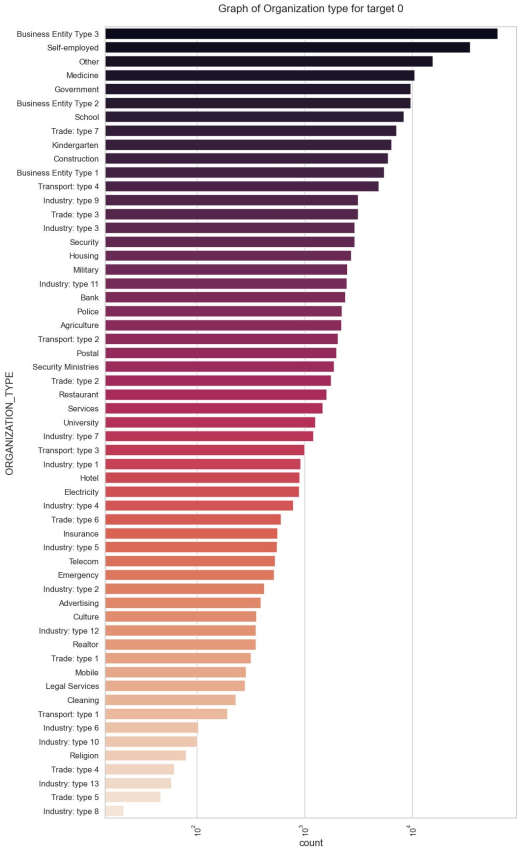
- In this data, Female is once

  Again leading for applying credits
- ‘Cash loans’ is having more number of

  credits than ‘Revolving loans’ .



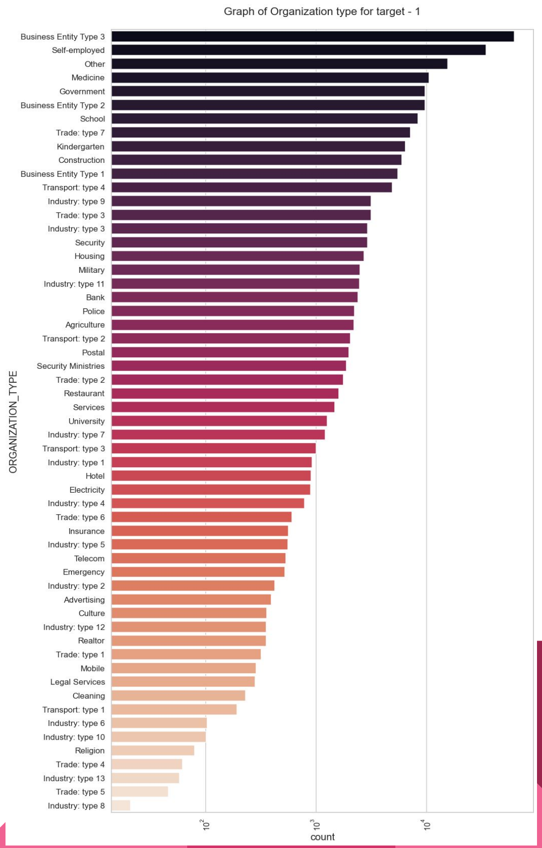
REPRESENTATION OF ORGANIZATION TYPE(CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0)



•Most clients who applied for credits are from ‘Government’, ‘Business entity Type 3’ , ‘Self employed’ , ‘Medicine’ and ‘Other’ .

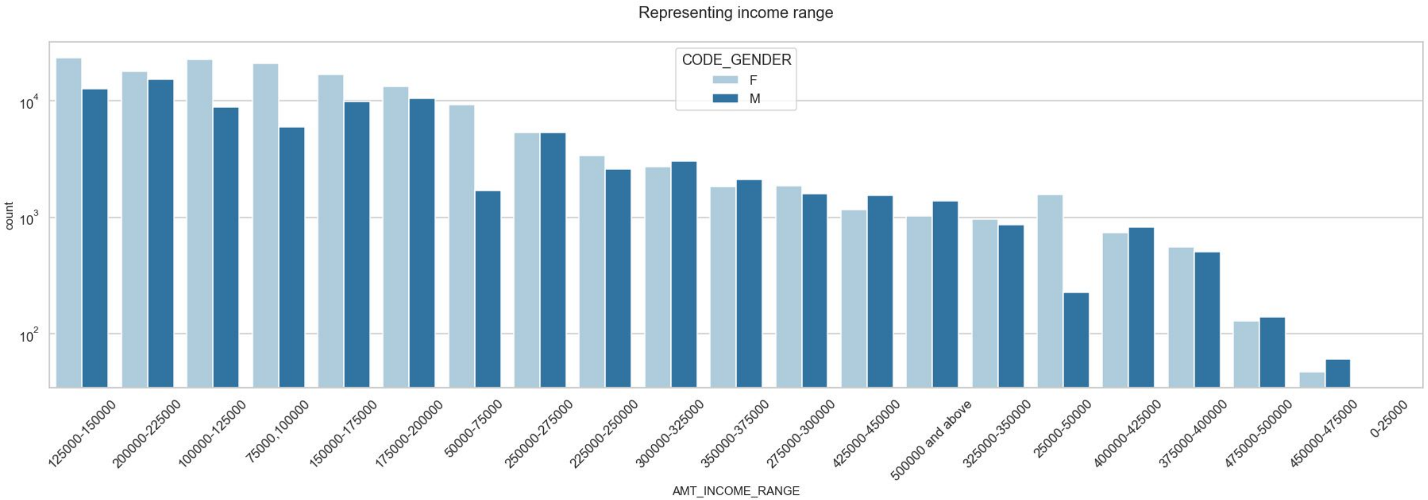
REPRESENTATION OF ORGANIZATION TYPE(CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1)

- Just like Target 0, most clients who applied for credits are from ‘Government’, ‘Business entity Type 3’ , ‘Self employed’ , ‘Medicine’ and ‘Other’ .



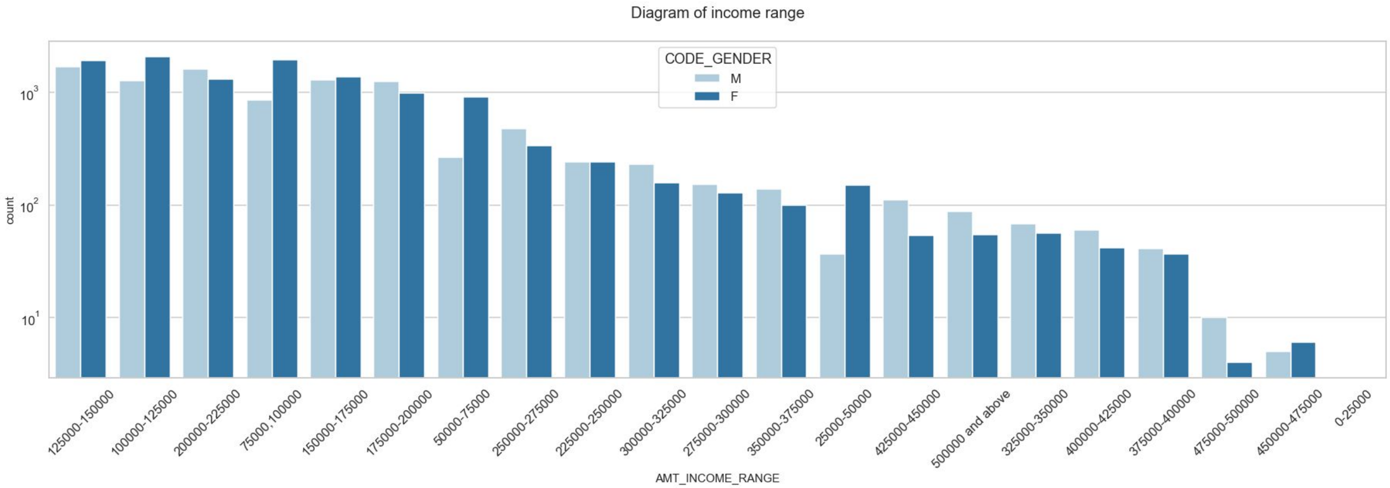
Graph of Income range (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0)

- Male counts are lesser compared to female.
- Income range from 100000 to 200000 has more number of credits.
- Count for income range 400000 and above is very less



Graph of Income range (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1)

- Female counts are lesser than male.
- Income range from 100000 to 200000 is has more credits.
- Count for income range 400000 and above is very less

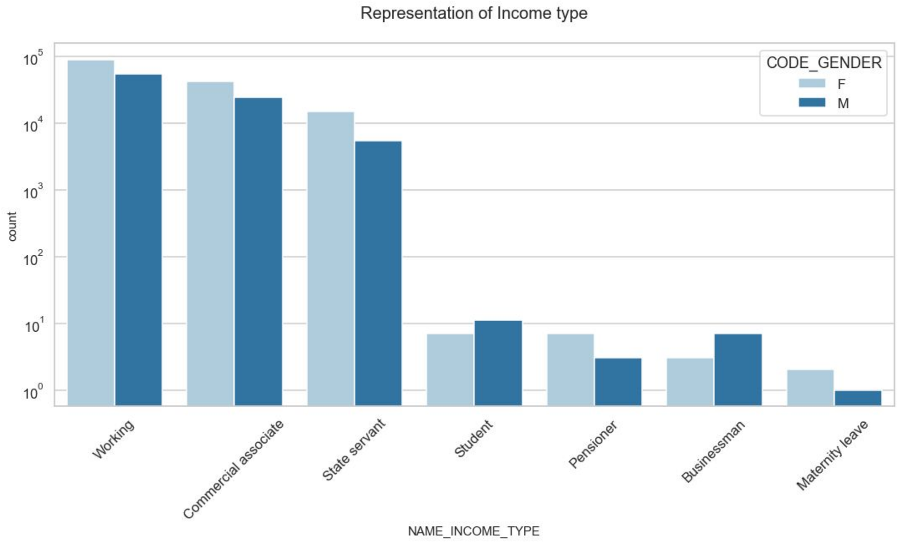


Graph of Income type (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0)

•The number of credits are higher for income type ‘working’, ’commercial associate’, and ‘State Servant’.

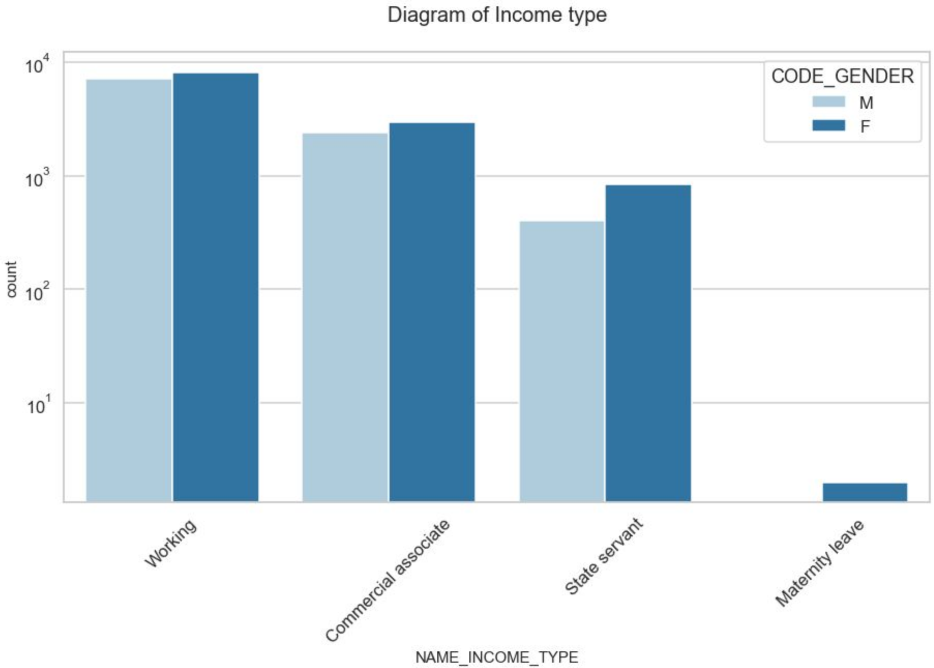
• ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’ has less credits.

•Females have more number of credits than males.



Graph of Income type (CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1)

- ‘Working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher compared to others
- Females are again, having more number of credits than males.
- For type 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.

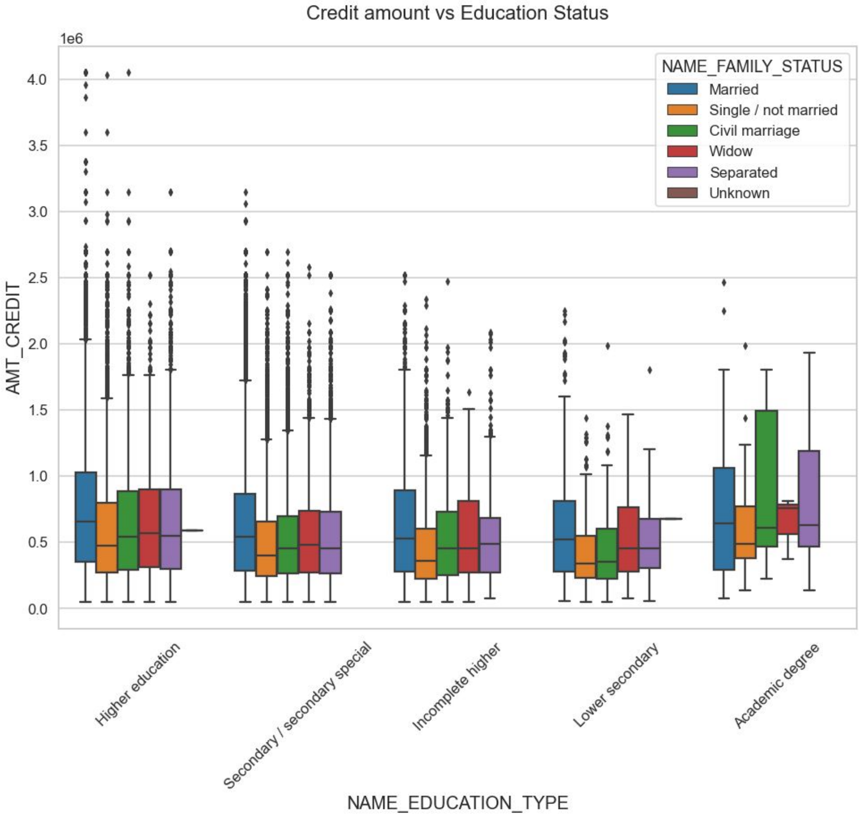


BIVARIATE ANALYSIS FOR TARGET 0 & 1

REPRESENTATION OF CREDIT AMOUNT vs EDUCATION STATUS (TARGET 0)

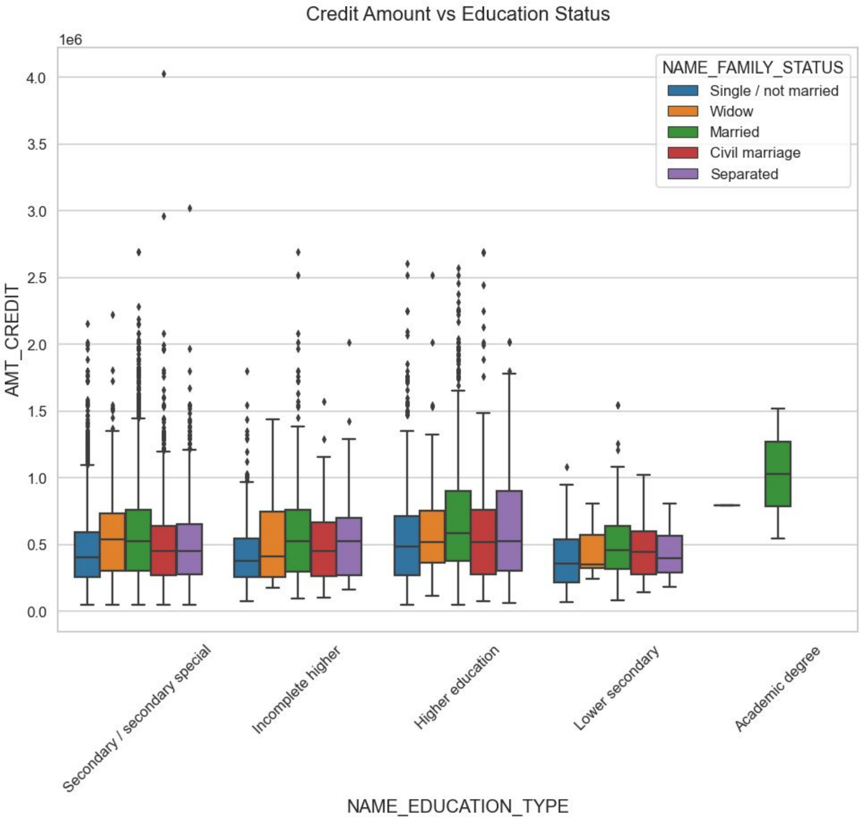
•In the graph, high credits are shown represented to Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education

•Higher education of family status of 'marriage', 'single' and 'civil marriage' has more outliers.

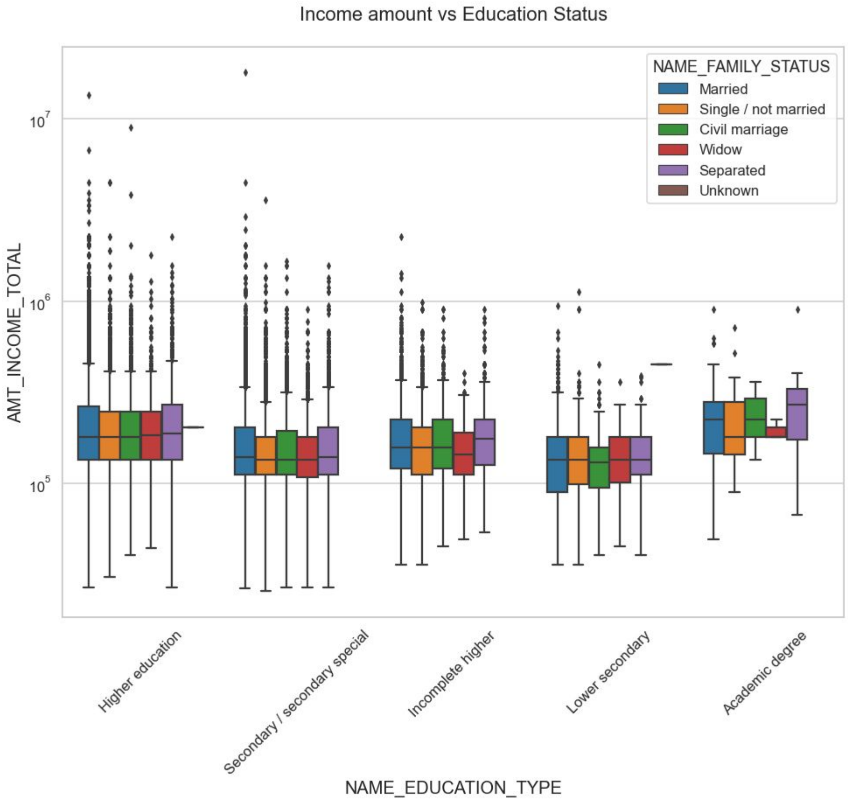


REPRESENTATION OF CREDIT AMOUNT vs EDUCATION STATUS (TARGET 1)

- Civil marriage has an outlier
- Married academic degree seems to have the highest 50%

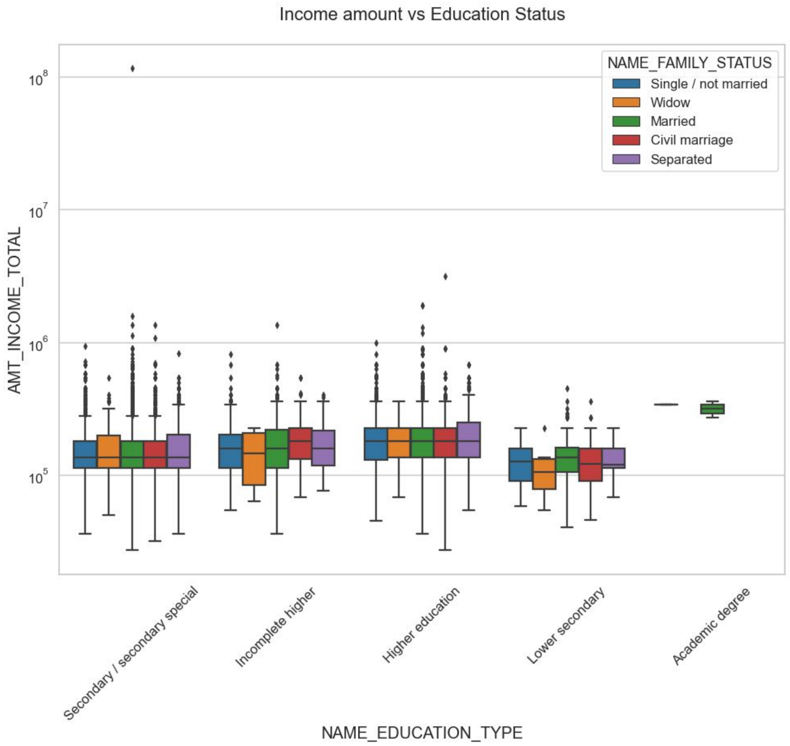


REPRESENTATION OF INCOME AMOUNT vs EDUCATION STATUS (TARGET 0)



- Education type 'Higher education' does contain many outliers.
- Academic degree has less outliers
- Lower secondary of civil marriage family status has less income amount than others.

REPRESENTATION OF INCOME AMOUNT vs EDUCATION STATUS (TARGET 1)



- "Married" has the least lower and upper Academic degree values
- "Civil Marriage" has outliers in higher Education.

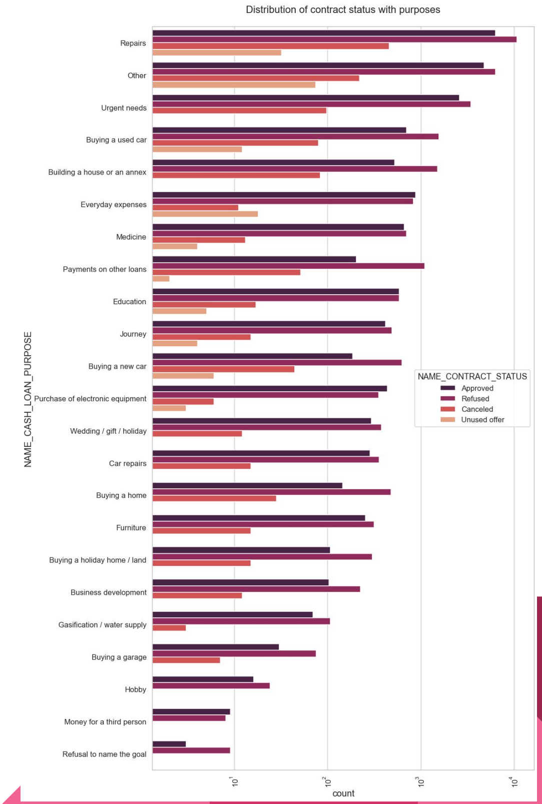
UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

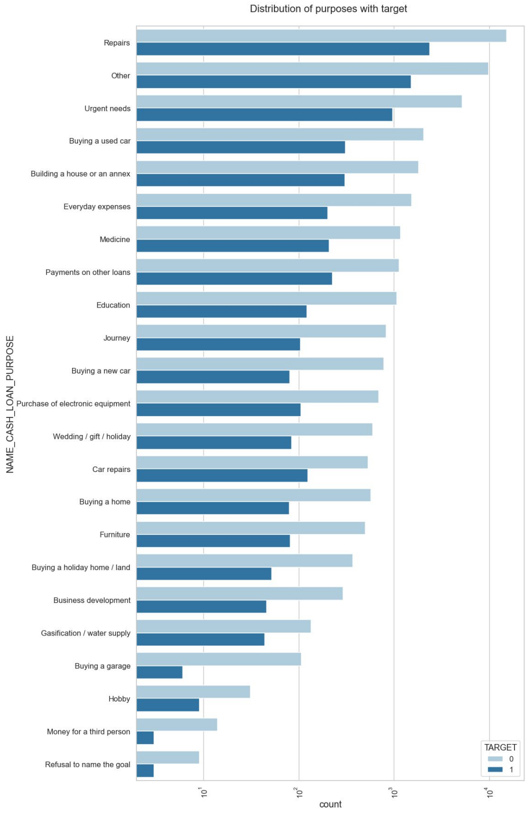
- 'Repairs' has most rejection of loans

- For education purposes we have equal amount of approvals and rejection

- Paying other loans and buying a new car is having significant higher rejection than approves.



DISTRIBUTION OF PURPOSES WITH TARGET

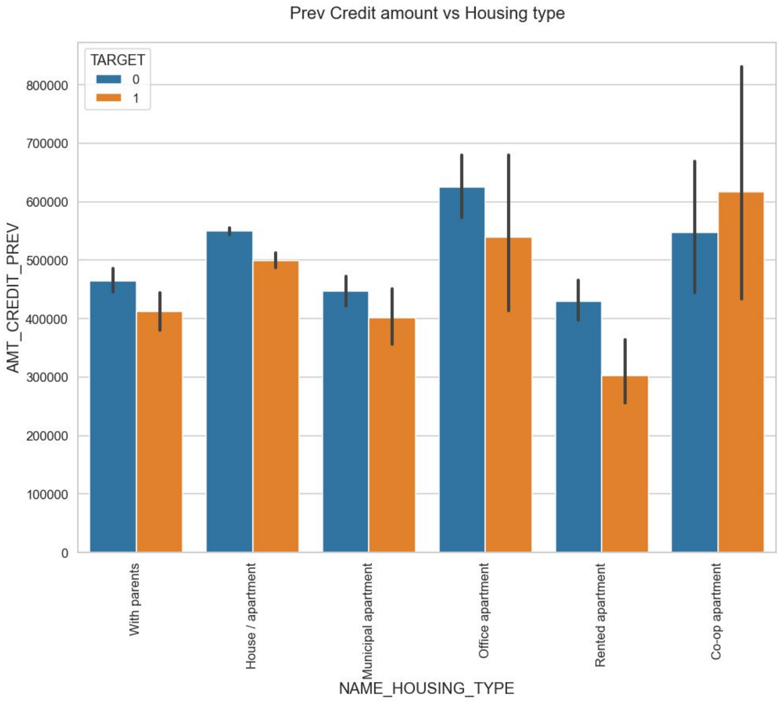


•Loan purposes with 'Repairs' faces difficulties paying on time.

•'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education' have significantly higher loan payment.

BIVARIATE ANALYSIS

GRAPH FOR PREV CREDIT AMOUNT VS HOUSING TYPE



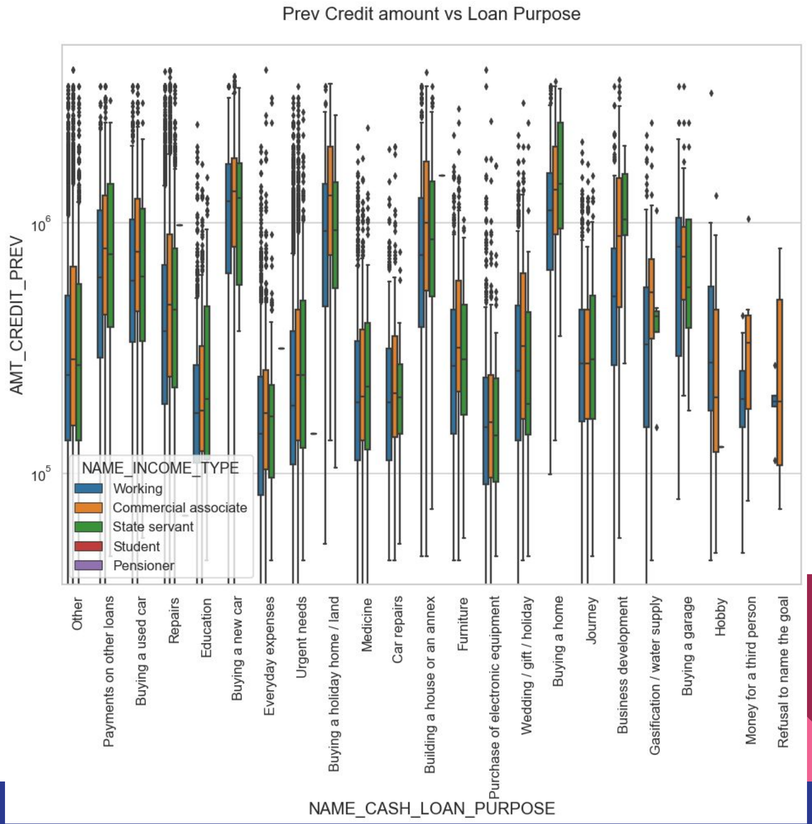
- Here for Housing type, office apartment has higher credit of target 0 and co-op apartment is having higher credit of target 1.
- Bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.
- Bank can focus mostly on housing type with parents or House or apartment or municipal apartment for successful payments.

GRAPH FOR PREV CREDIT AMOUNT vs LOAN PURPOSE

•The credit amount of Loan purposes is higher for ‘Building a house', 'Buying a home’, ’Buying a land’, and 'Buying a new car'

•Income type of state servants have a high amount of credit applied

•Money for hobby o is having less credits applied for.



CORRELATION

CORRELATION FOR TARGET 0

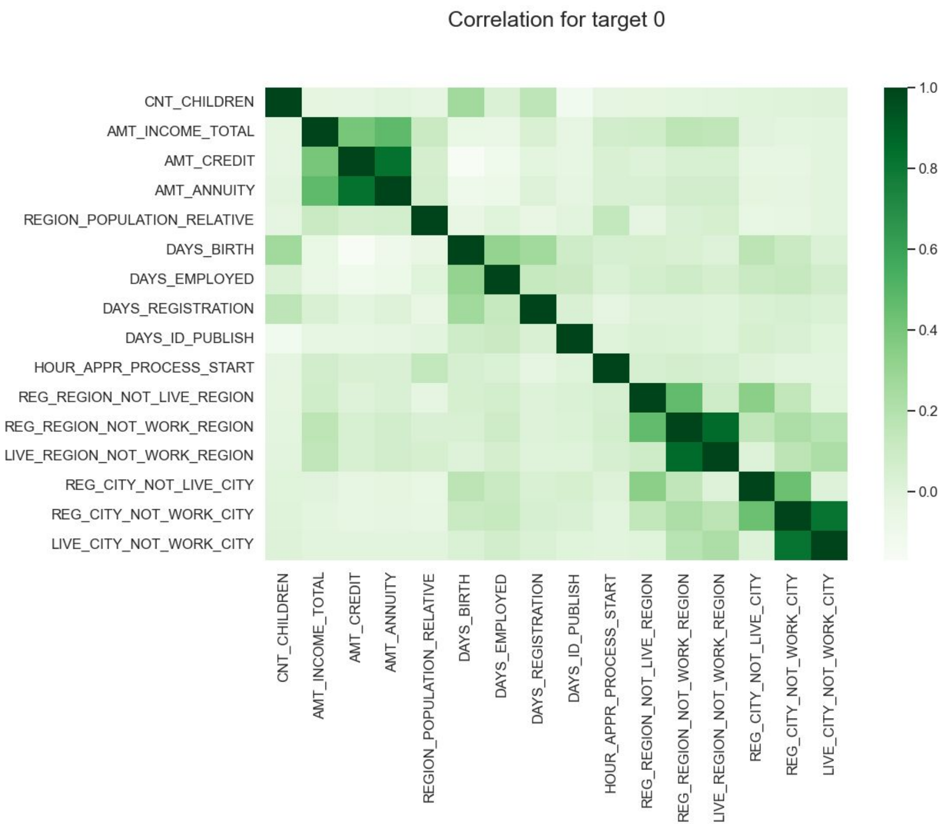
- Credit amount is low for high age and vice-versa.

- Credit amount is lesser for high children count client have and vice-versa.

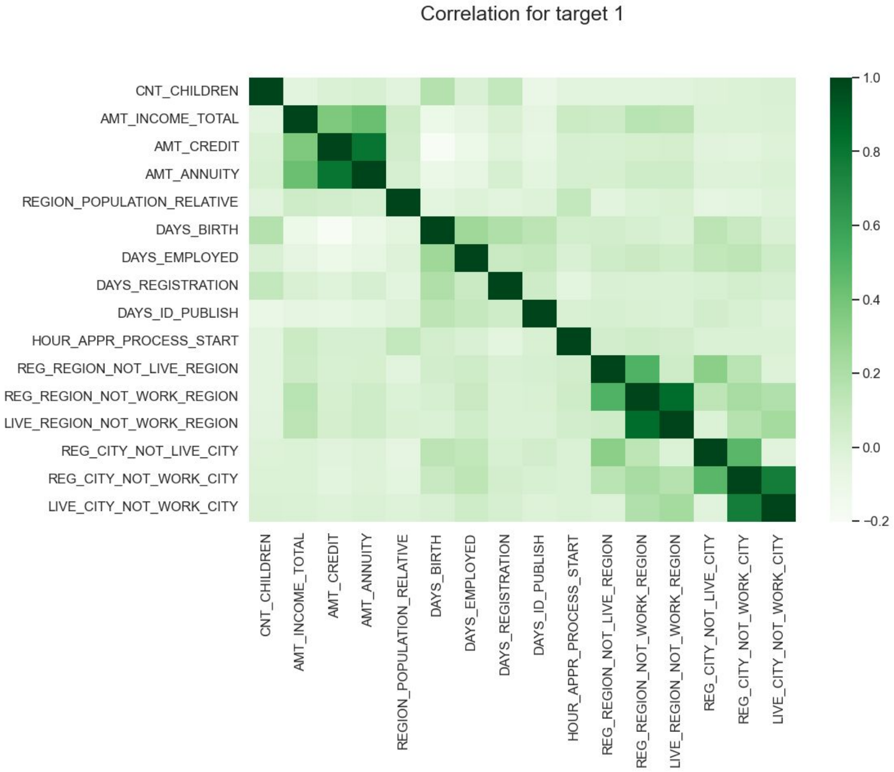
- Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.

- less children client have in densely populated area.

- Credit amount and income is higher to densely populated area.



CORRELATION FOR TARGET 1



This heat map is similar to the previous heat map.

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

CONCLUSION

CONCLUSION

After all these analysis, we have come to the conclusion that:

- Banks should spend their time focusing more on contract types such as ‘Student’ ,’pensioner’ and ‘Businessman’ with housing type other than ‘Co-op apartment’ for more successful payments.

- ‘Working’ income type are having most number of unsuccessful payments, therefore banks should focus less on them

- Bank should focus on clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.

- Loan purpose ‘Repair’ has higher number of unsuccessful payments on time.

THANK YOU