CREDIT EDA CASE STUDY

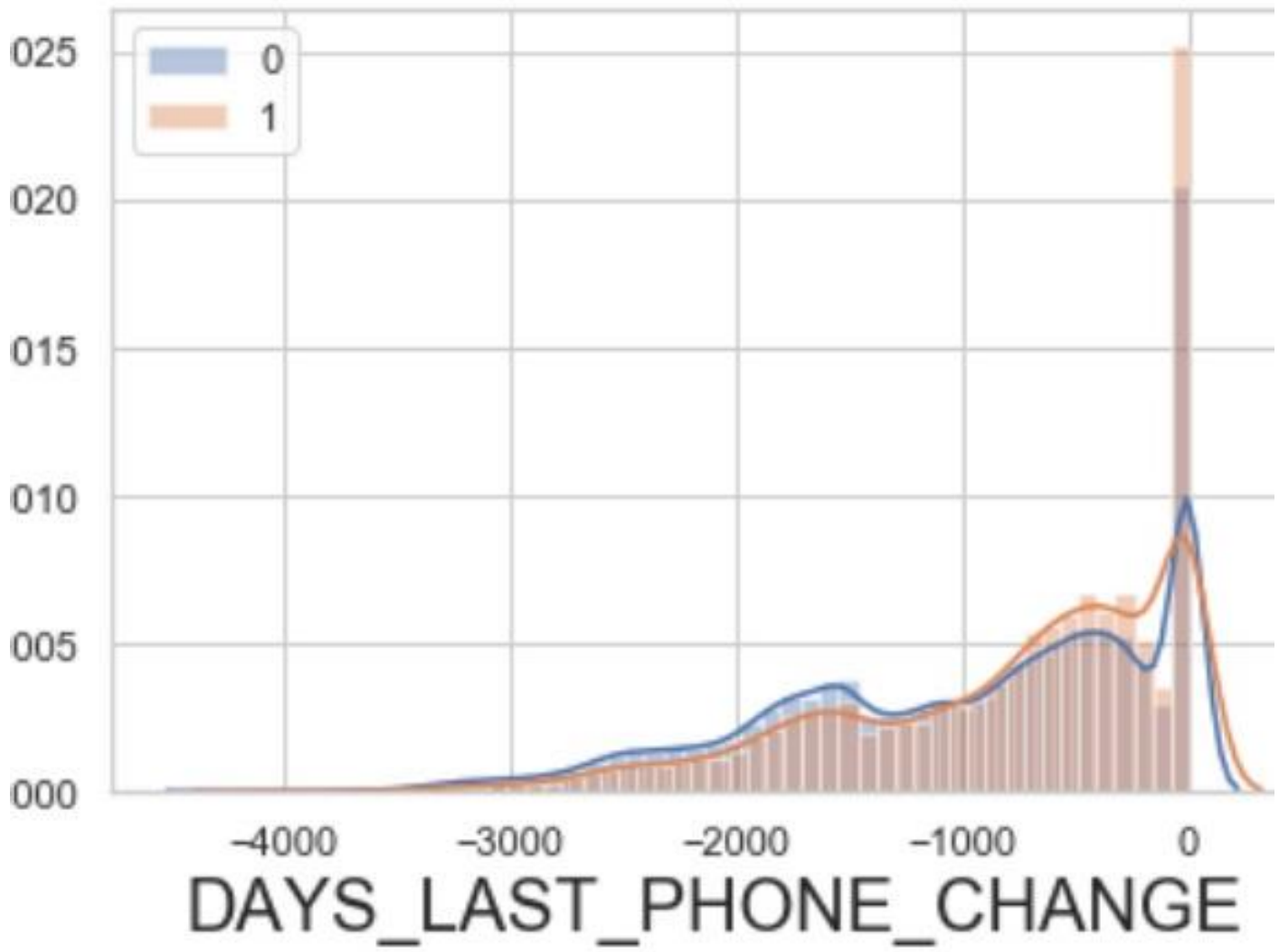
By:- HARSH SHARMA

SHARDOOL PAL

Univariate analysis

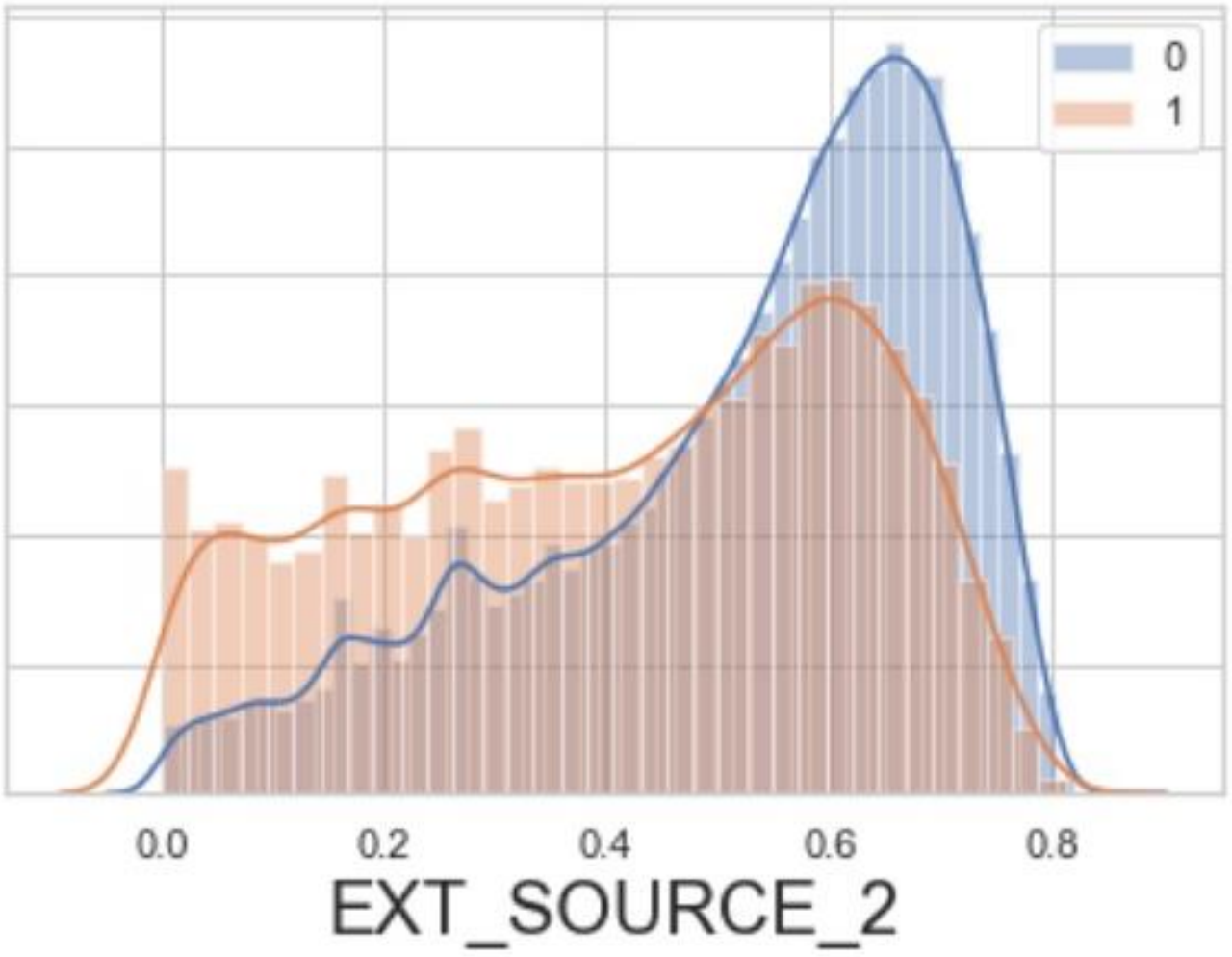
Inferences

- The dist-plot is plotted for target variable 0 and 1 giving high number of phone change for defaulters within few days of placing an application
- It also states that the numbers are negative that means from the date of application the numbers were changed last updated.



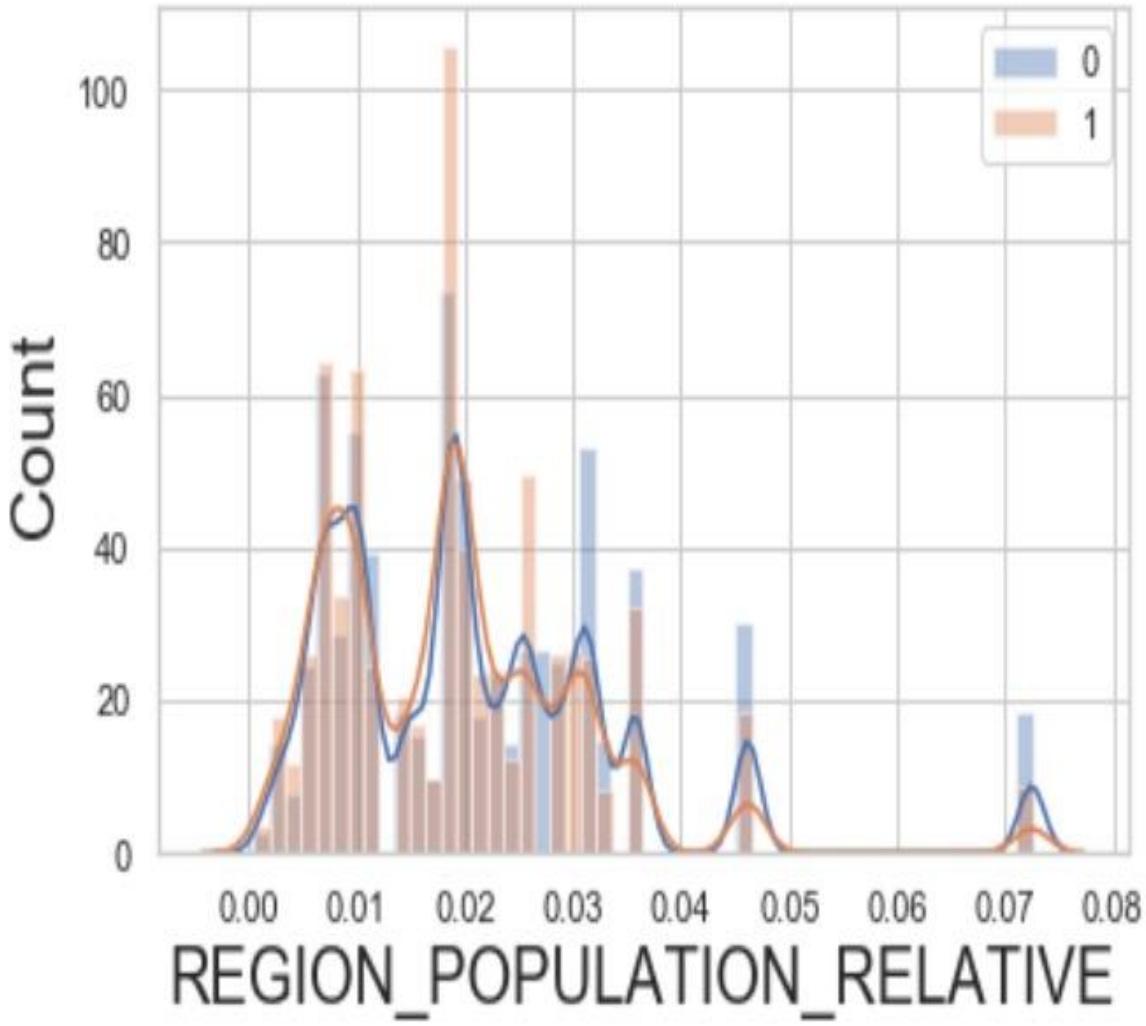
Inferences

- As observed in the graph we can see Target 1 and target 0 are inversely affected with EXT_SOURCE_2 rating
- It can be stated that EXT_SOURCE_2 normalized score are more reliable to predict defaulters



Inferences

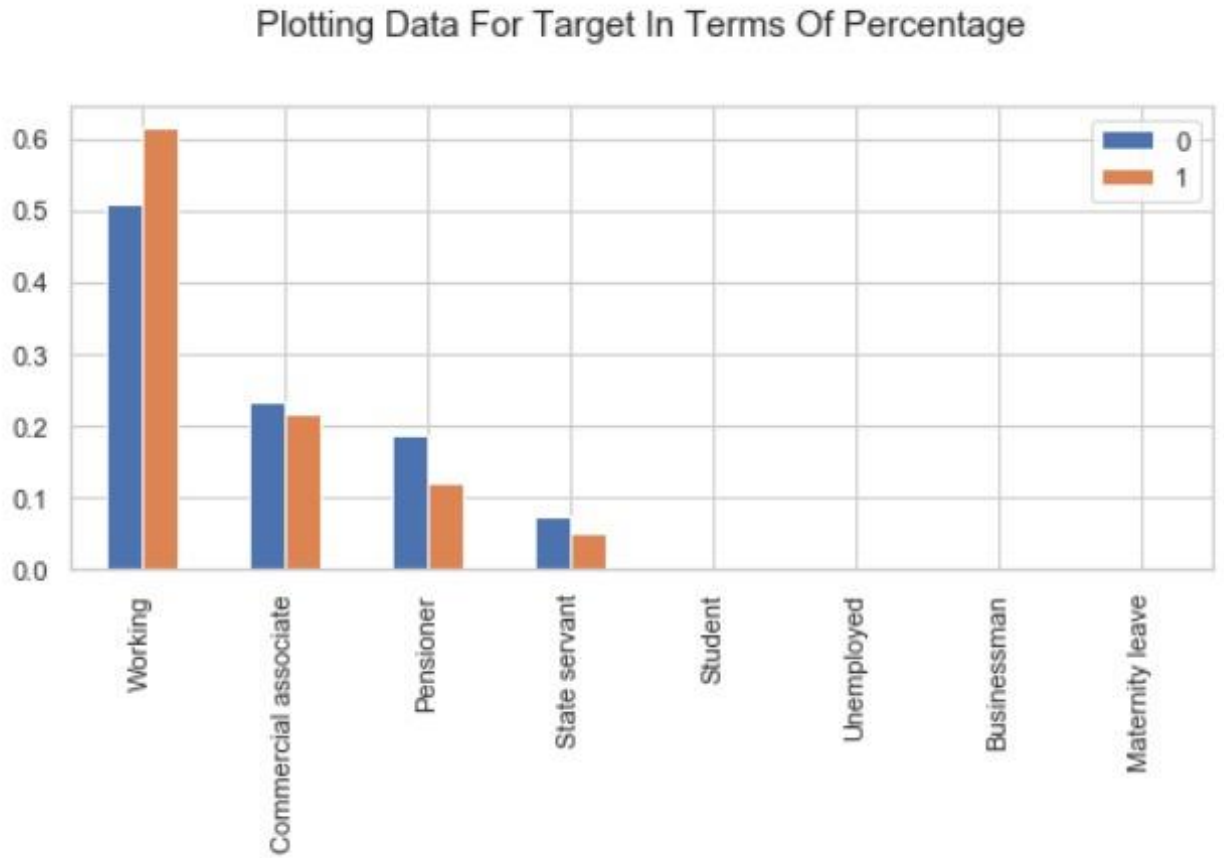
- It can be observed that the less-populated areas have more clients.
- Defaulters have a high probability of being from 0.02 REGION_POPULATION_RELATIVE
- The defaulter and non-defaulter ratio increases with density towards more number of non-defaulters



Categorical Univariate analysis

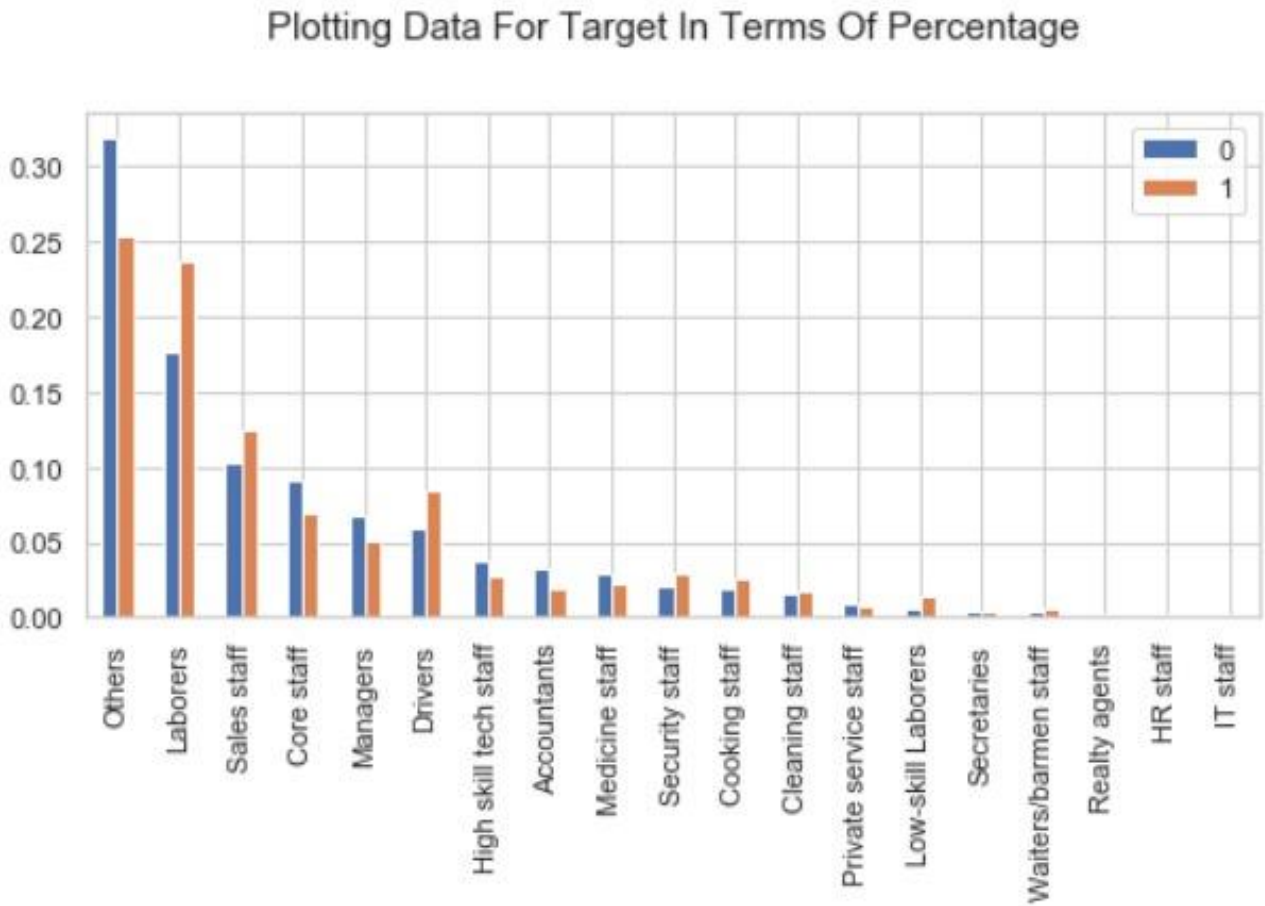
Inferences

- Majority of clients for loans are from working class
- It can be observed the rate of defaulters is high in working class
- Pensioners have high ratio of non-defaulters. Hence they can be considered when taking an application.



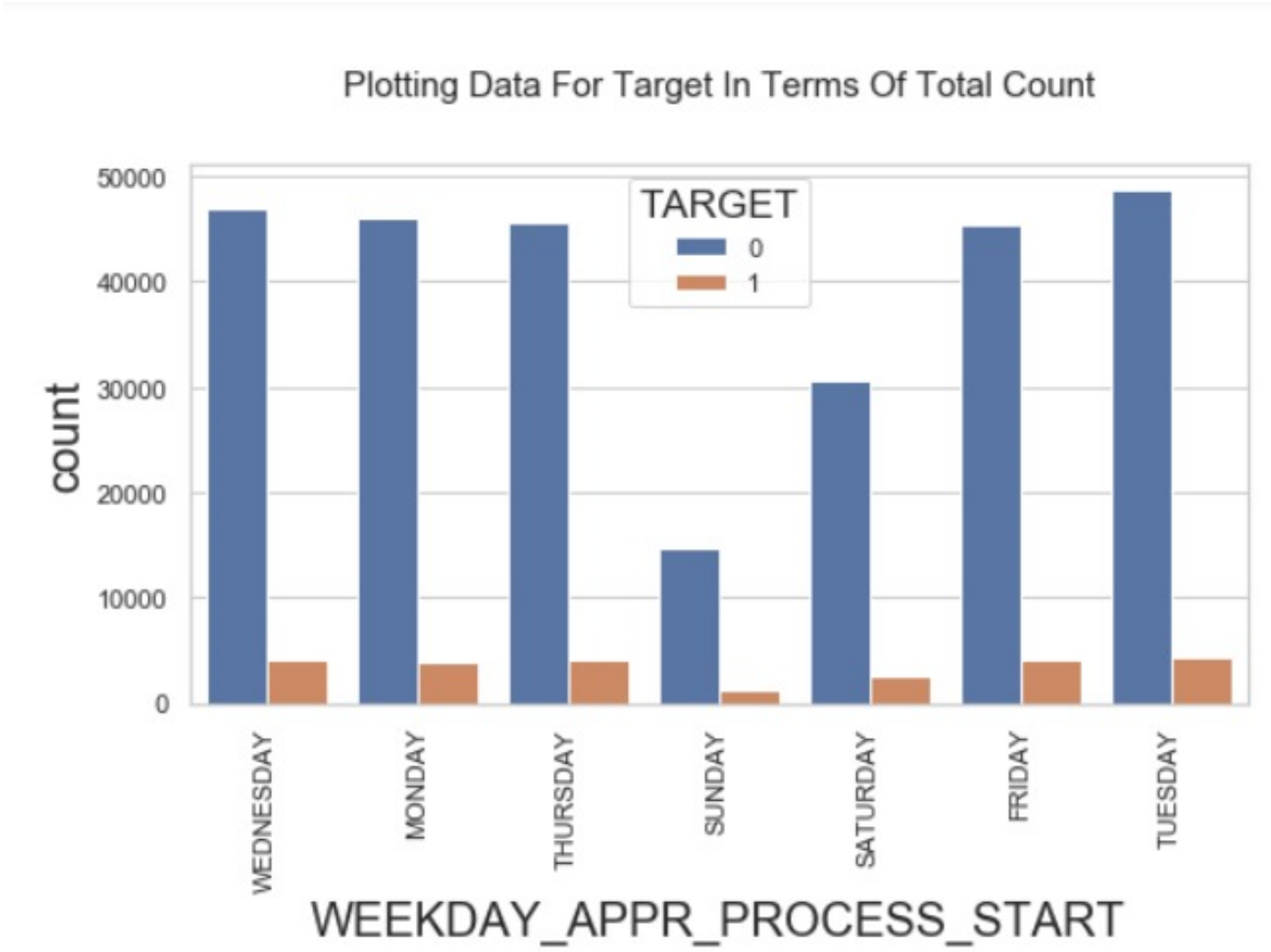
Inferences

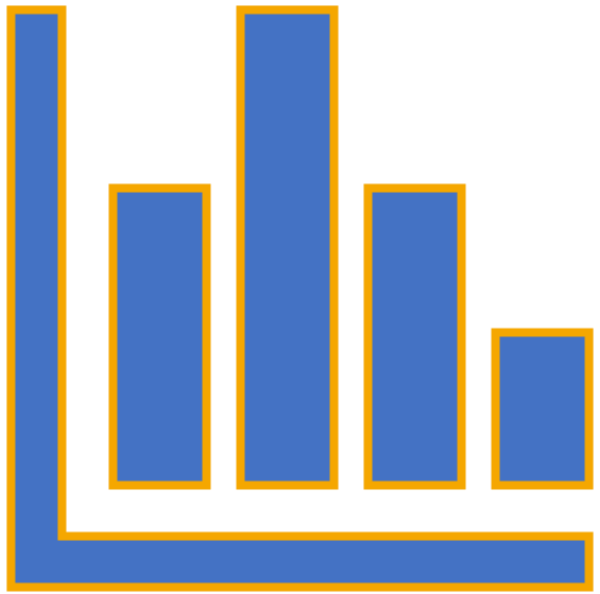
- It can be observed that defaulter's percentage is high in laborer's, sales staff and Drivers.
- Managers and Core Staff which are considered to be high paying jobs have less defaulters.



Inferences

- Sundays and Saturdays the loans applied are the least due to Bank closure or holidays.
- We can see Tuesdays have maximum application days as people in India consider Tuesdays to be auspicious.

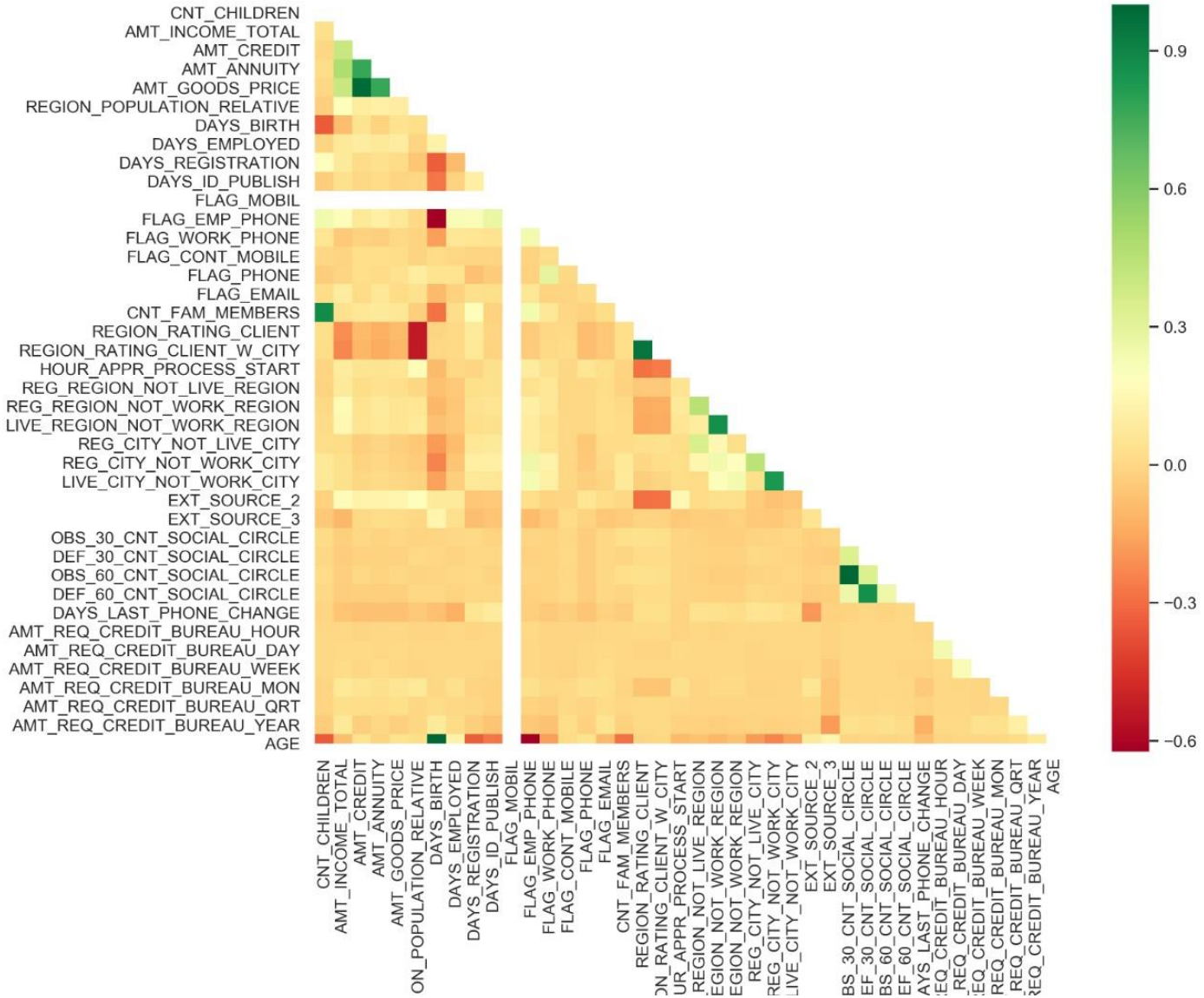


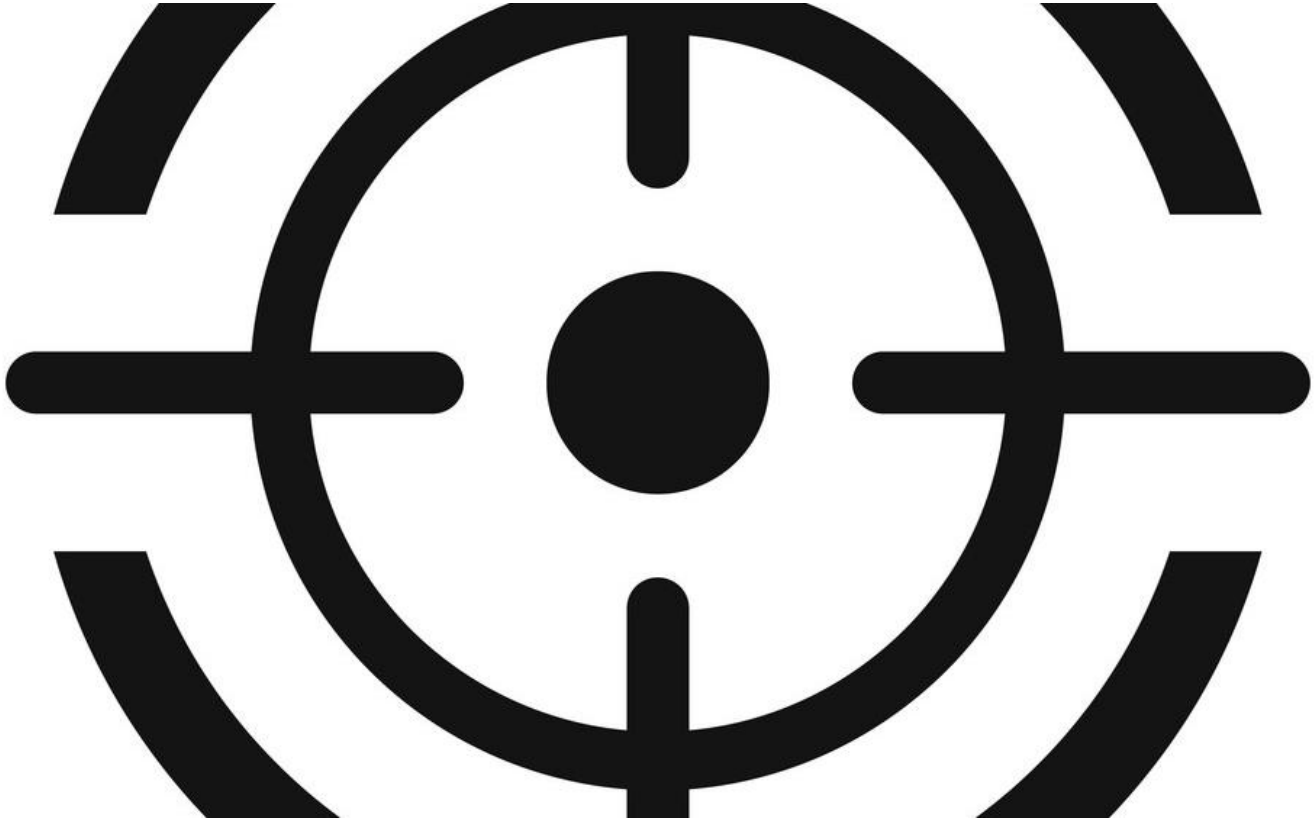


Correlation of target 0

Inferences:

- Credit amount is inversely proportional to the date of birth implies Credit amount is higher for low age and vice-versa.
- Age and flag_emp_phone is inversely proportional, that means people with less age provide phone number of work phone more
- 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.

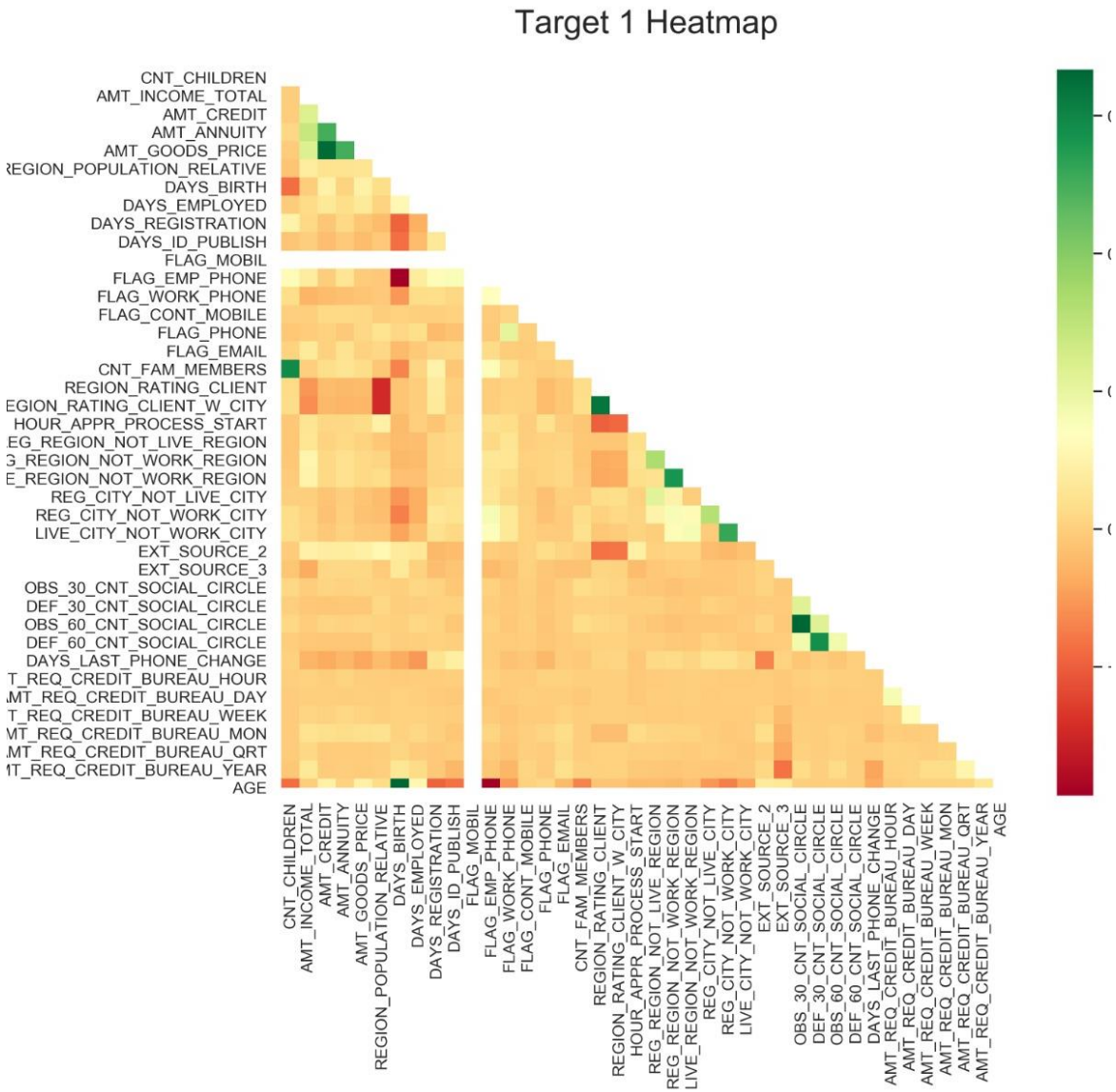


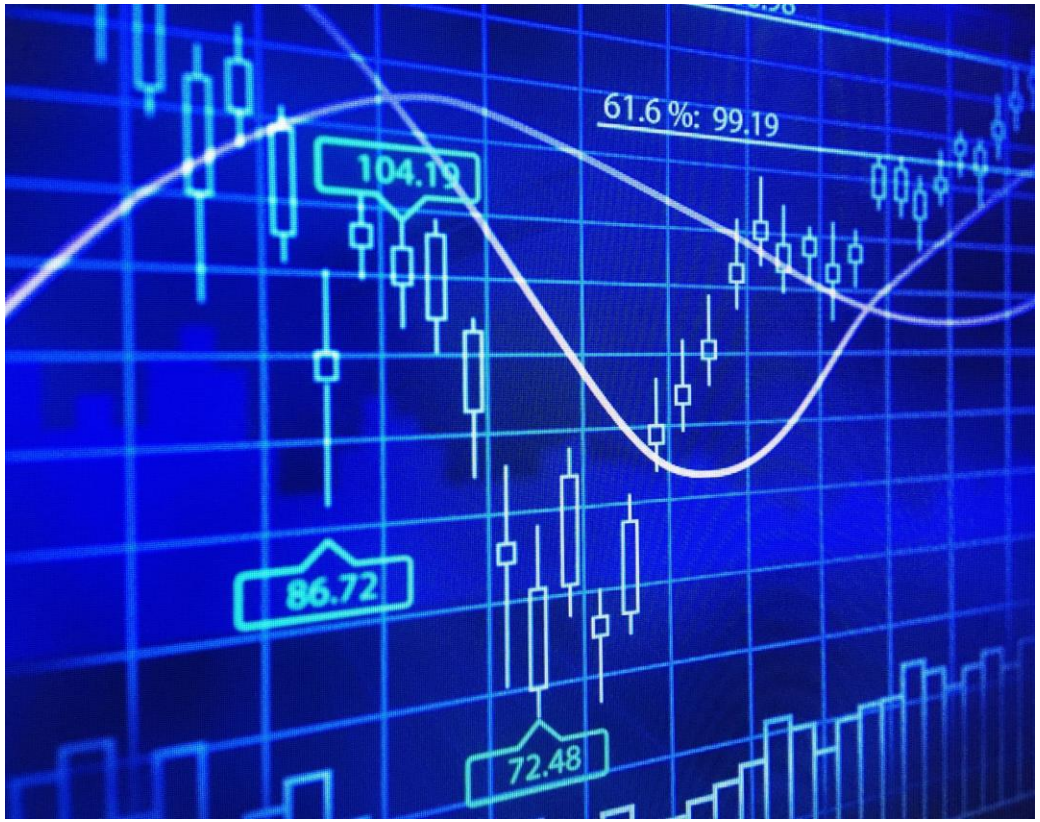


Correlation of target 1

Inferences:

- Credit amount is higher in densely populated area.
- The Income is also higher in densely populated area.
- Correlation matrix are similar for both the cases with slight difference in the same.

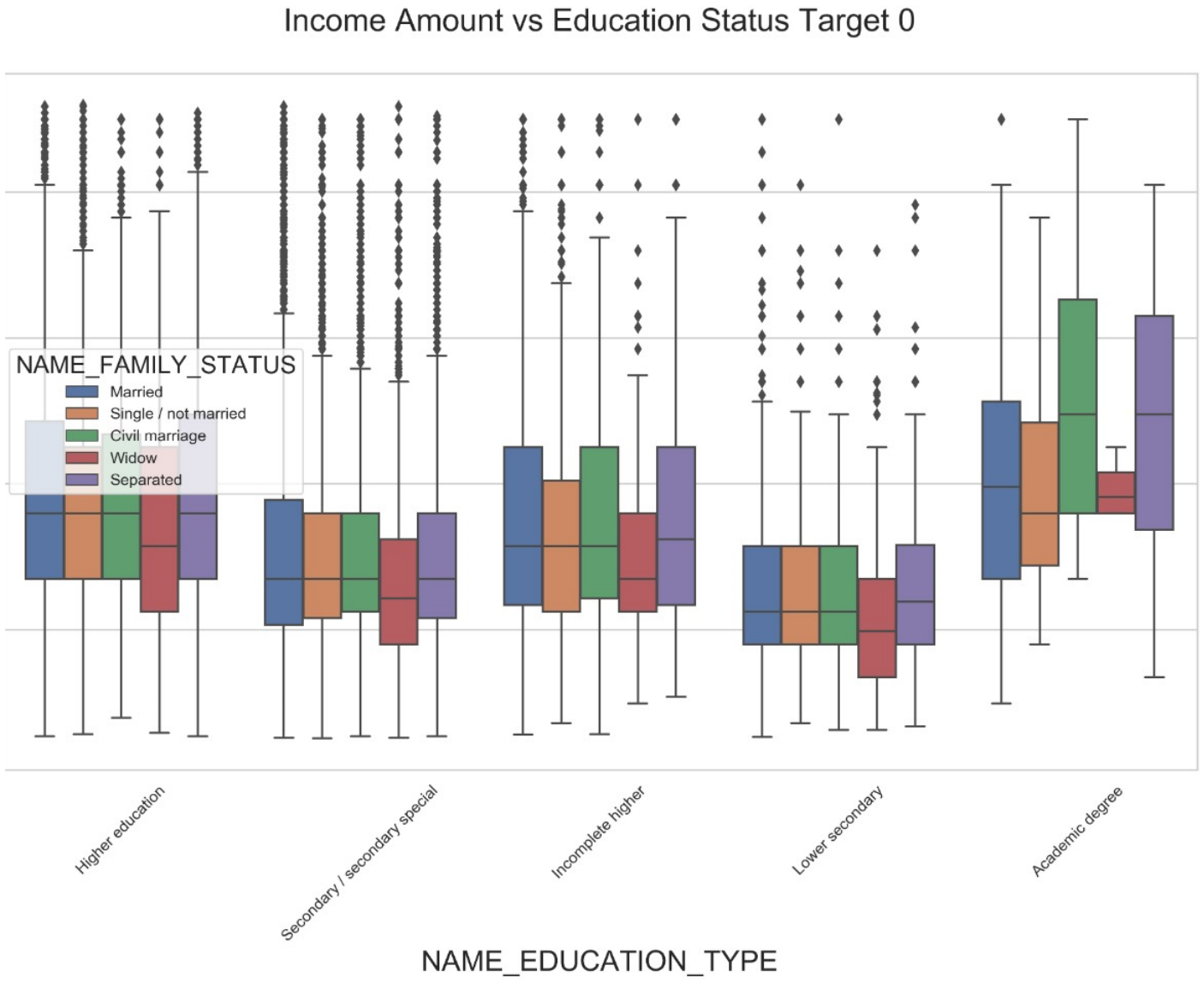




BIVARIATE ANALYSIS TARGET 0

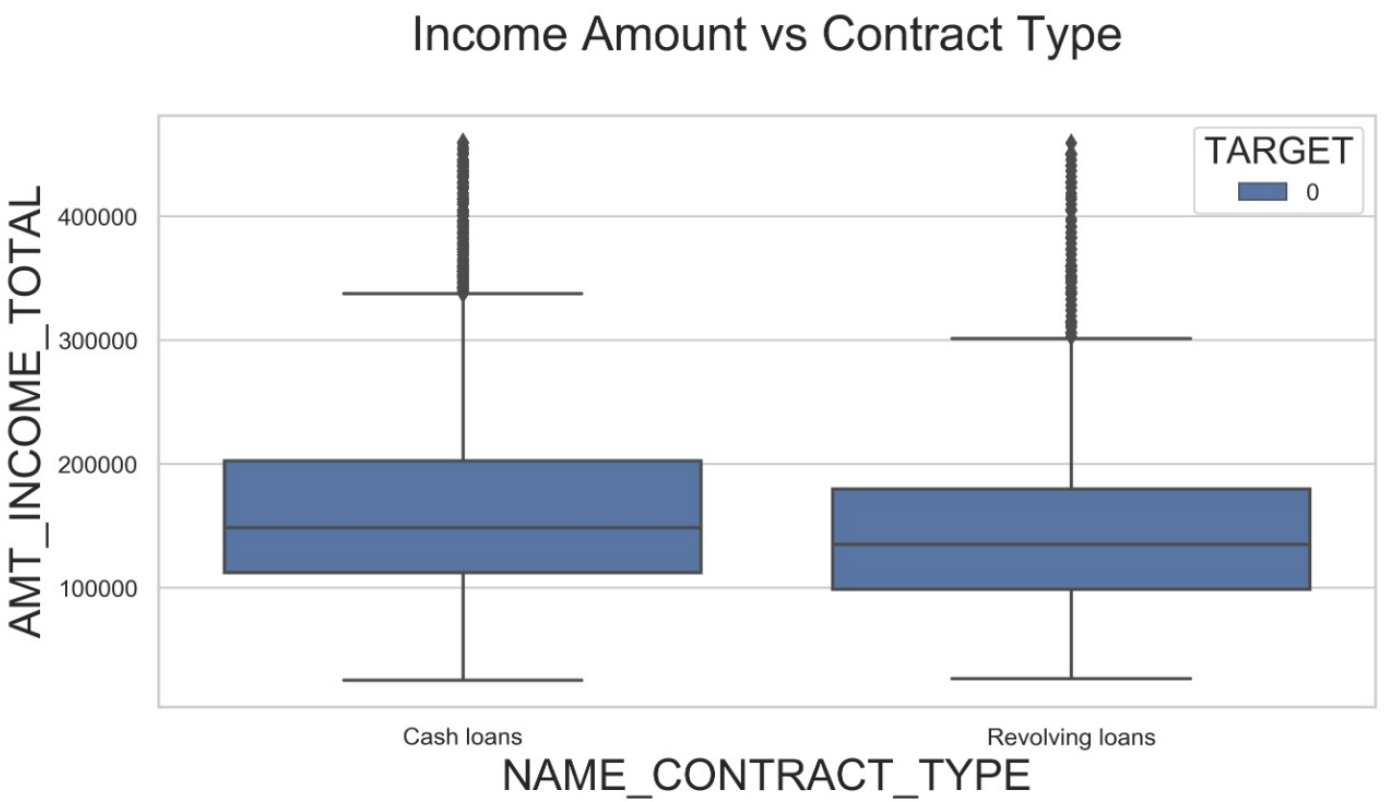
Inference

- Clients with academic degree have higher income compared to all others.
- Widowed category have low income when compared to all other family statuses
- Based on this inference and comparing it with Target 1 values we get insight as to which Education type should be preferred when giving loans



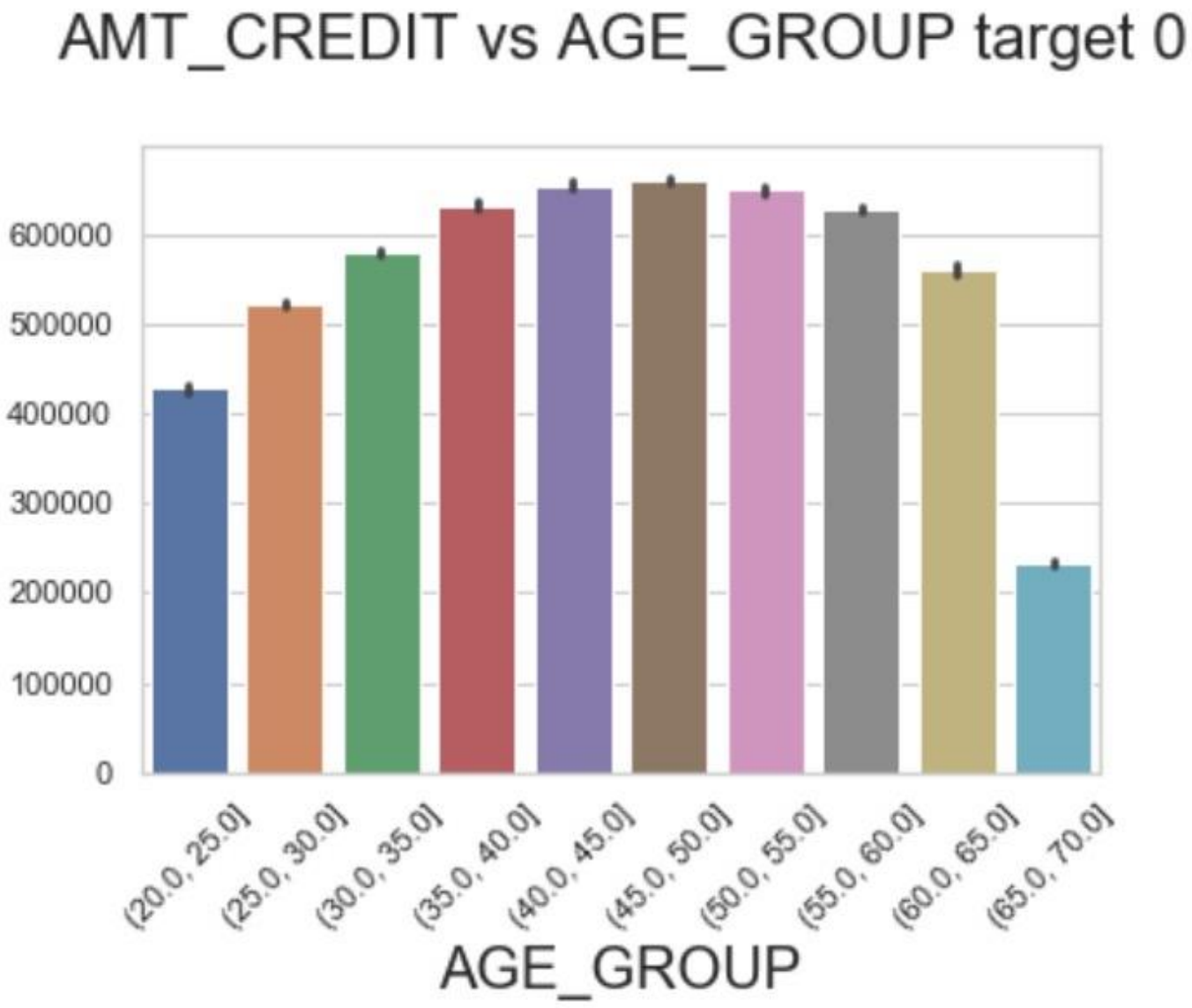
Inference

- We can observe Cash loans are more preferred type of loans
- Range of both the loans are similar
- When compared to Target 1 data defaulters comprise more of Cash Loans



- Inference

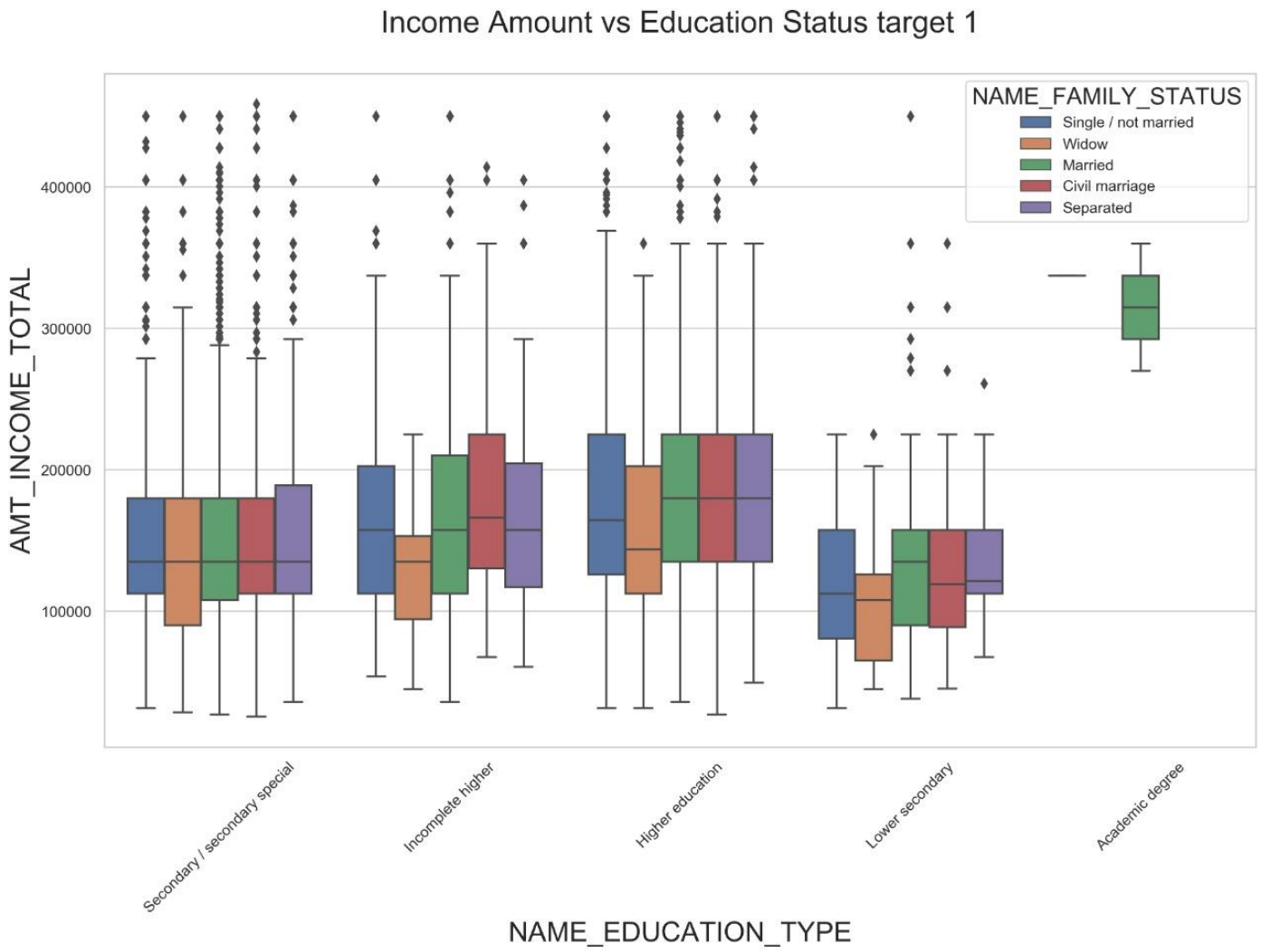
- Age group 45-50 comprises of high credit amount
- People with 65 are majorly considered as pensioners they have low amt_credit as the value of pensions are low.
- It forms a parabola opening towards x axis.



BIVARIATE ANALYSIS TARGET 1

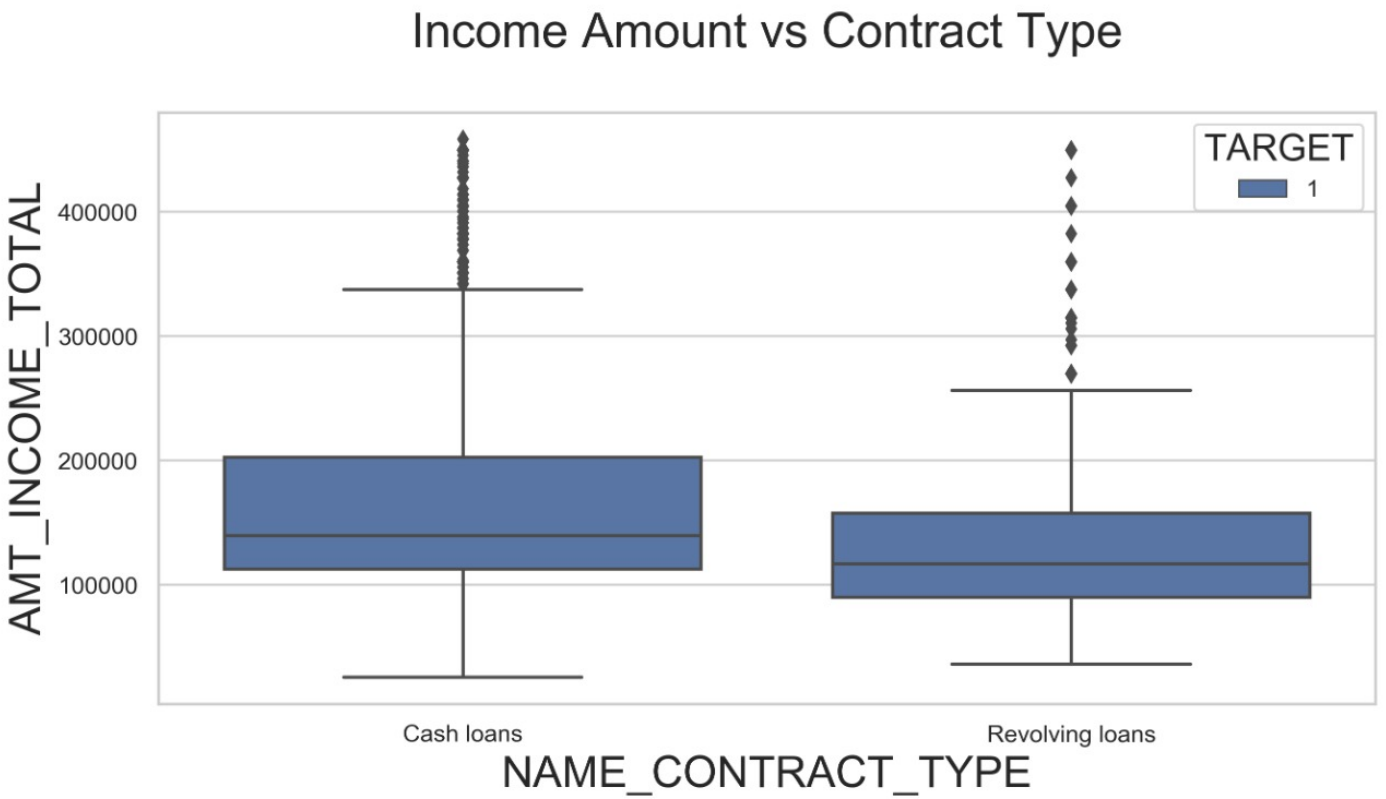
Inference:

- Married People with academic degree has high income and are in defaulters category.
- Married people in lower secondary qualification category are more prone to being a defaulter
- The outliers in incomplete higher education are less stating that the people stick to norms of categorizing

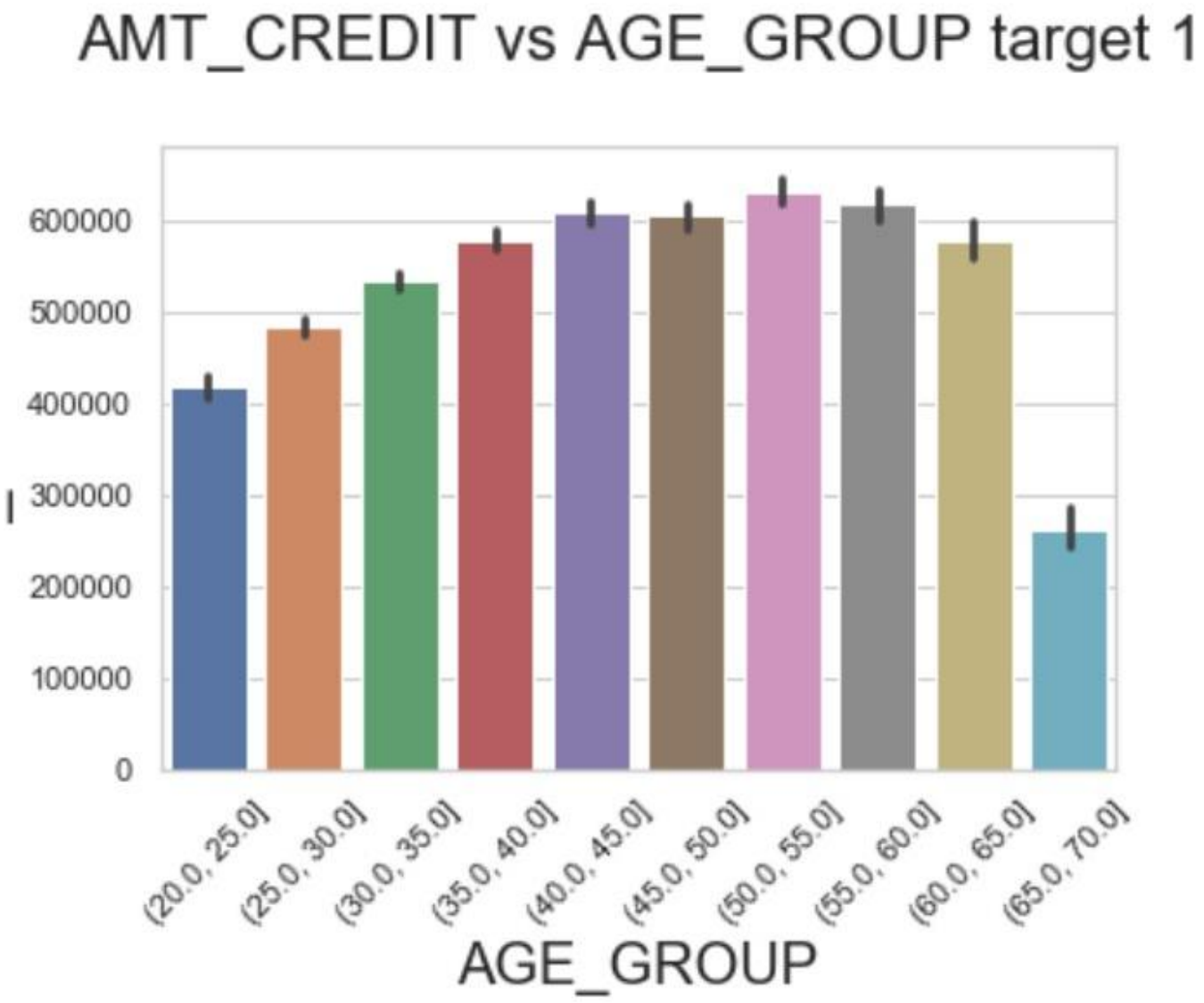


Inference:

- Revolving loans people have comparatively low AMT_INCOME_TOTAL.
- Defaulters comprise of less revolving loans people and vice versa
- Bank can prefer revolving loans to reduce defaulters



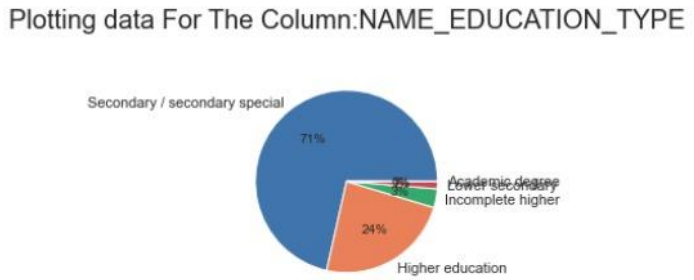
- Inference
- Defaulters are more in 50-55 category with high income.
- Pensioners remain at low salary and same ratio of target 0 or 1

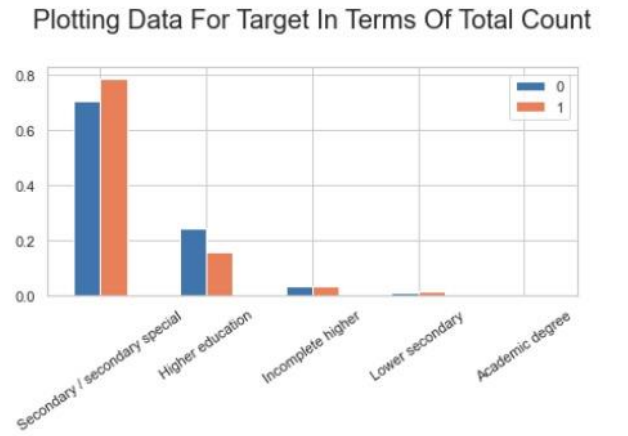


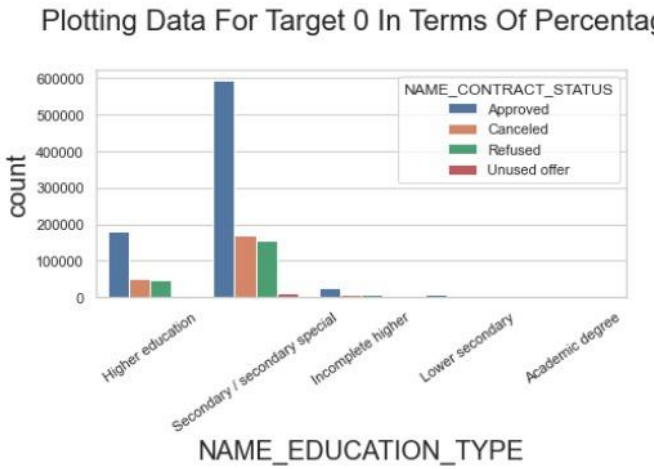
Bivariate Analysis after Merging Previous Application Data

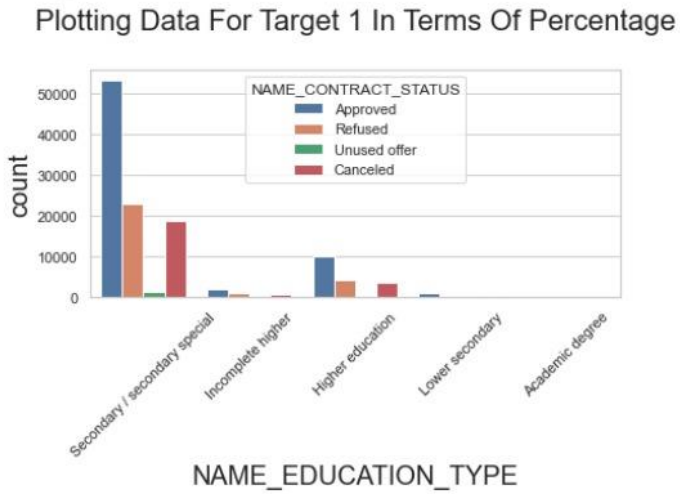
Inference

- Approved status of loans are comparatively very high in secondary/secondary school category
- Defaulters are less for higher education
- Secondary education group refuse





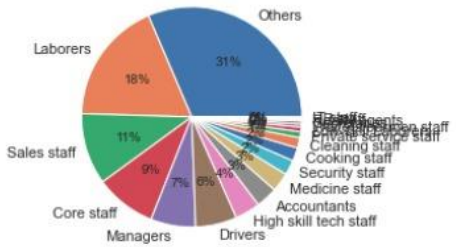




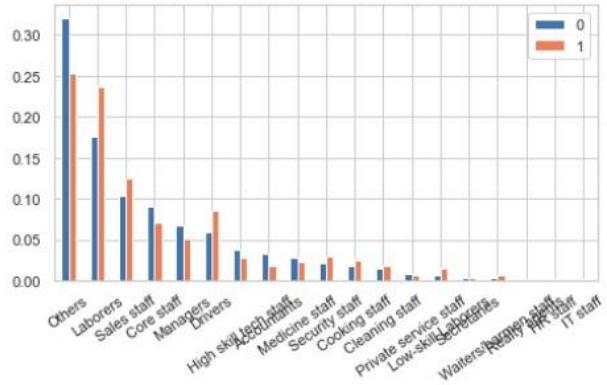
Inference

- Married Labourers are non-defaulters majorly
- Married clients have less chances of defaulting
- Single people who are laborers have less chances of defaulting

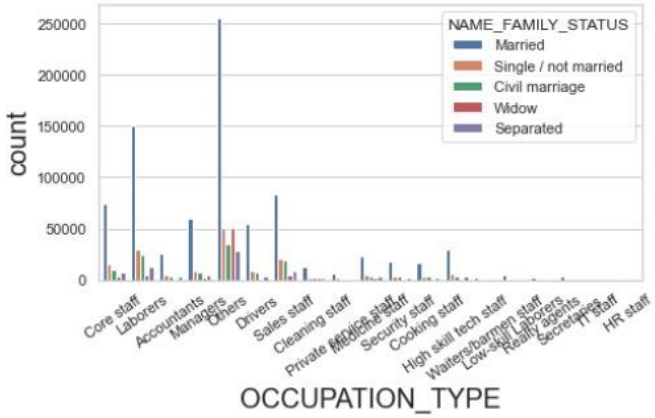
Plotting data For The Column:OCCUPATION_TYPE



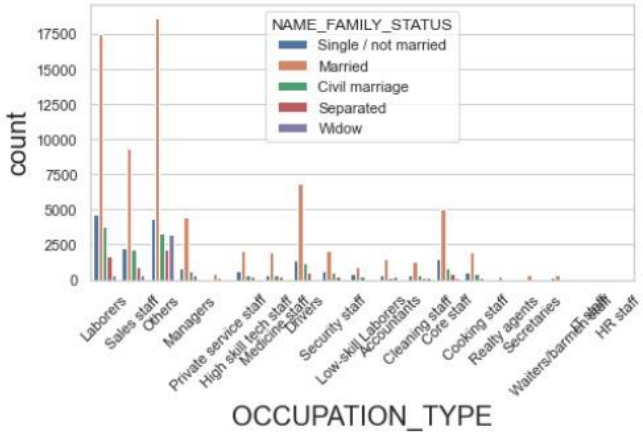
Plotting Data For Target In Terms Of Total Count



Plotting Data For Target 0 In Terms Of Percentage

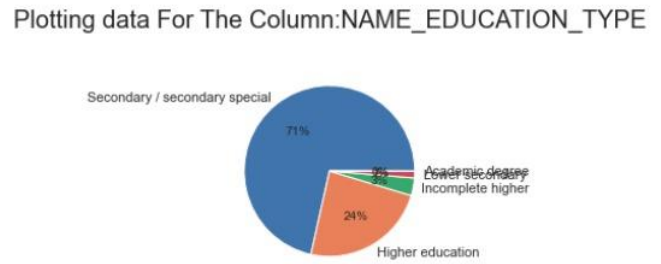


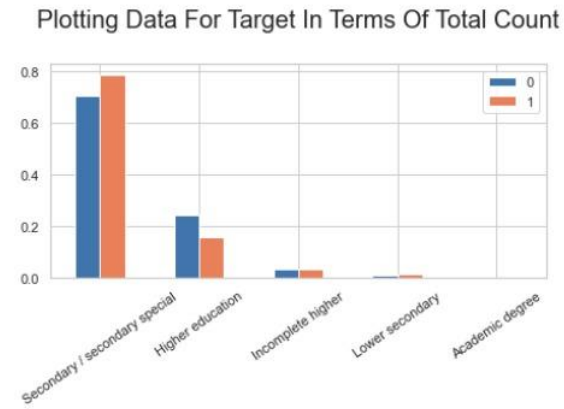
Plotting Data For Target 1 In Terms Of Percentage

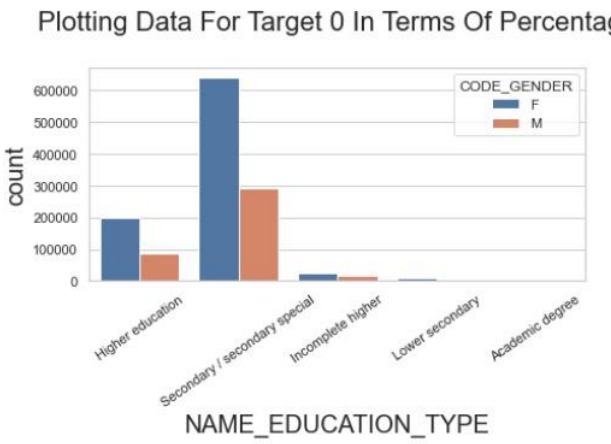


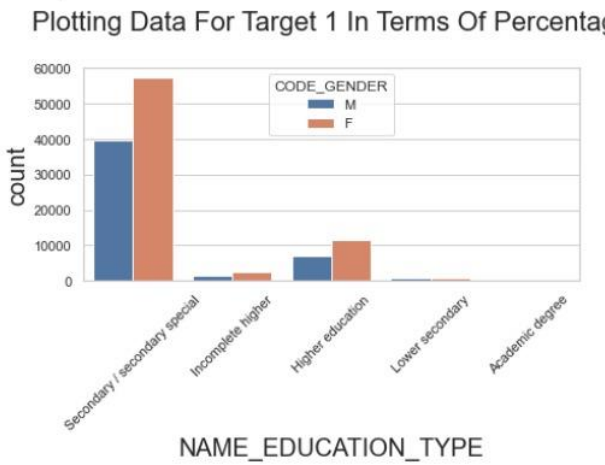
Inference

- Female should be preferred overall when giving loans
- It can be inferred that married women have support from their husbands hence less chances of defaulting
- Women in secondary education category have high chances of defaulting.









CONCLUSION

- Majority of Clients for loan are from working class
- Married Clients should be preferred as they pay debt in time.
- Managers and core staff which are considered to be high paying jobs should be preferred with the clients
- Academic education people have less defaulters. And outliers are less in the academic education.

THANK YOU