

EDA CASE STUDY

Prerna Prakash
Murali Krishna Kalla

Univariate analysis for application_data.csv

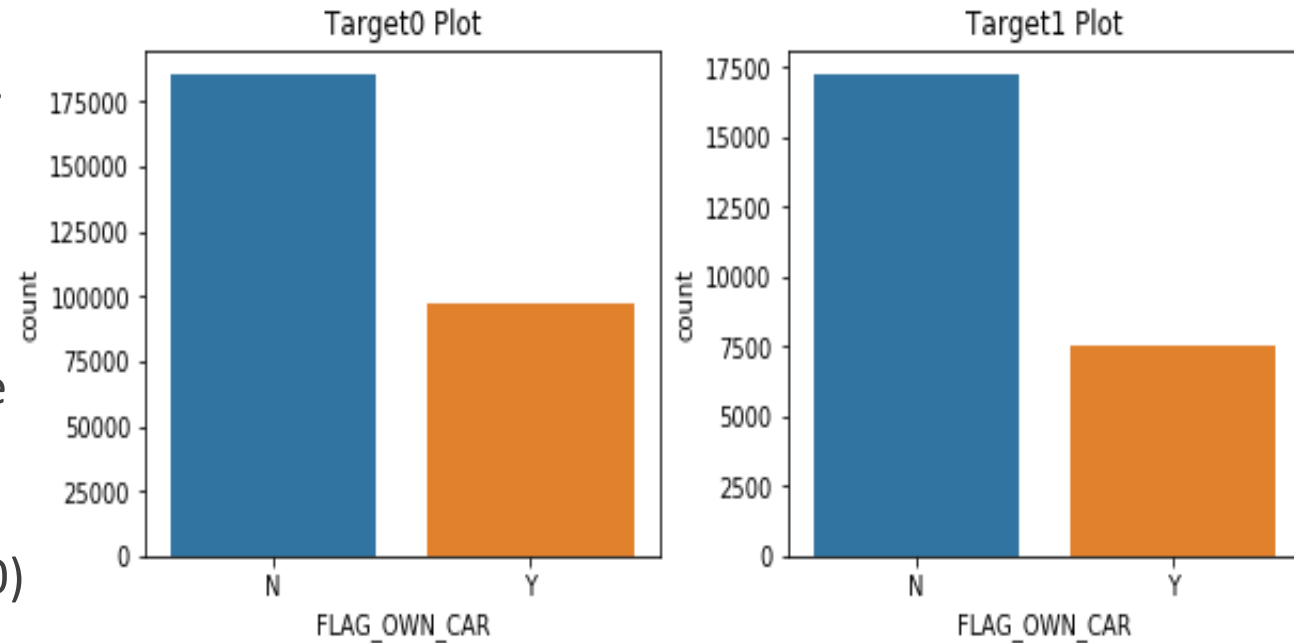
CATEGORICAL:

FLAG_OWN_CAR - Flag if the client owns a car

Comment:

Non defaulted clients(Target=0) with no car(x=N) are more in number compared to the customers who are defaulted(Target=1)

Majority of the Non defaulted clients(Target=0) own a car(x=Y) in comparison to the defaulted customers(Target=1).



REG_CITY_NOT_WORK_CITY - Flag if client's permanent address does not match work address
(1=different, 0=same, at city level)

Comment:

Number of people who are not defaulted
(Target =0) have high number of cases where
client's work address and permanent address
are same, as compared to people who have
Defaulted.



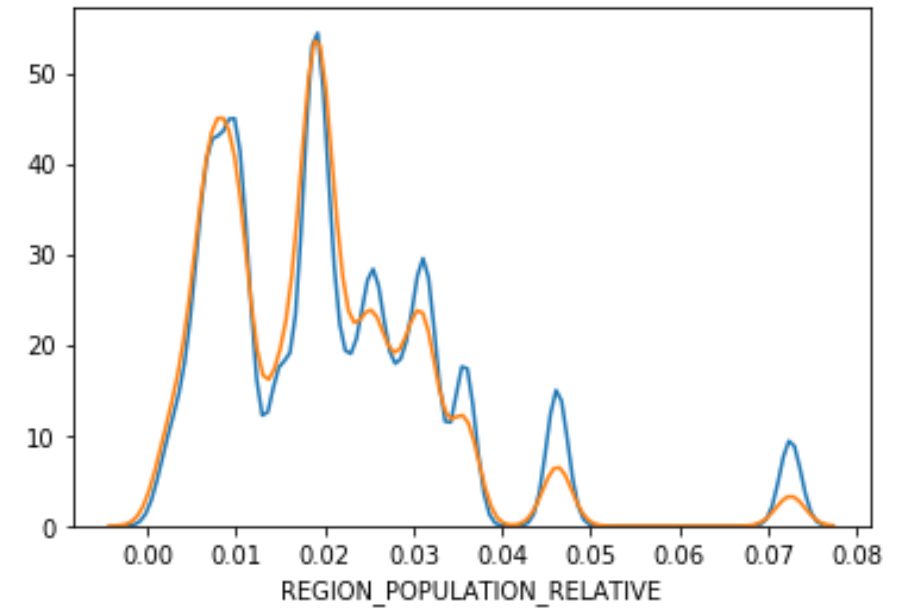
Similarly we can also observe that the people who stay in different city and are not defaulted are still higher in number than defaulted customers in different cities

Continuous:

REGION POPULATION RELATIVE - Normalized population of region where client lives (higher number means the client lives in more populated region)

Comment:

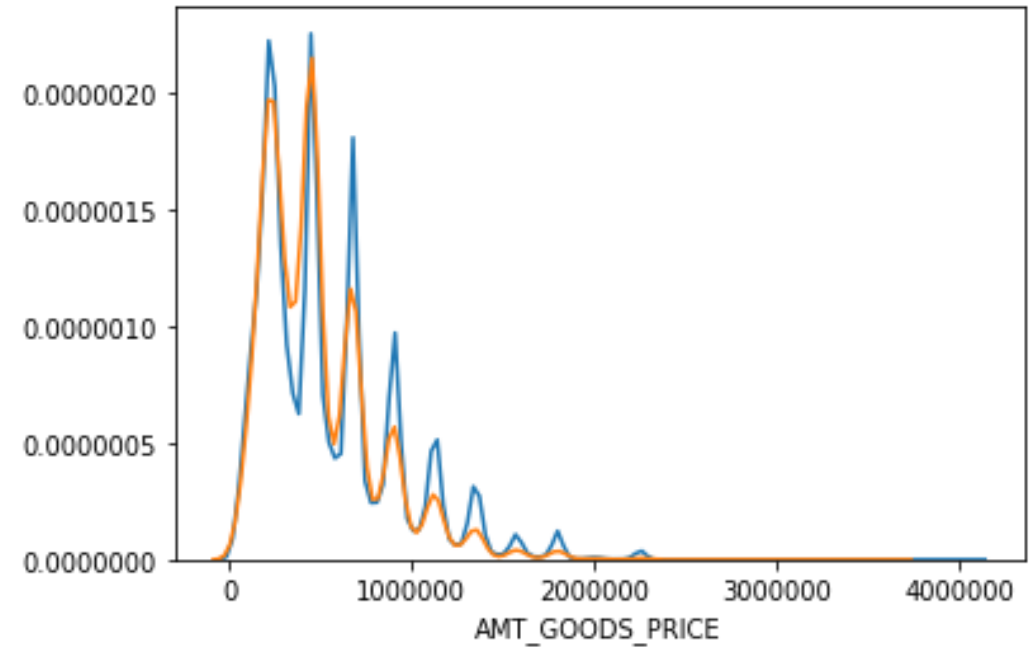
As the density of the population increases the number of defaulted clients are lesser compared to non-defaulted clients.



AMT_GOODS_PRICE: Goods price of good that client asked for (if applicable) on the previous application

Comment:

For goods price ranging from low to high, defaulted customers are lesser in comparison to non-defaulted customers



Bi-variate analysis for application_data.csv

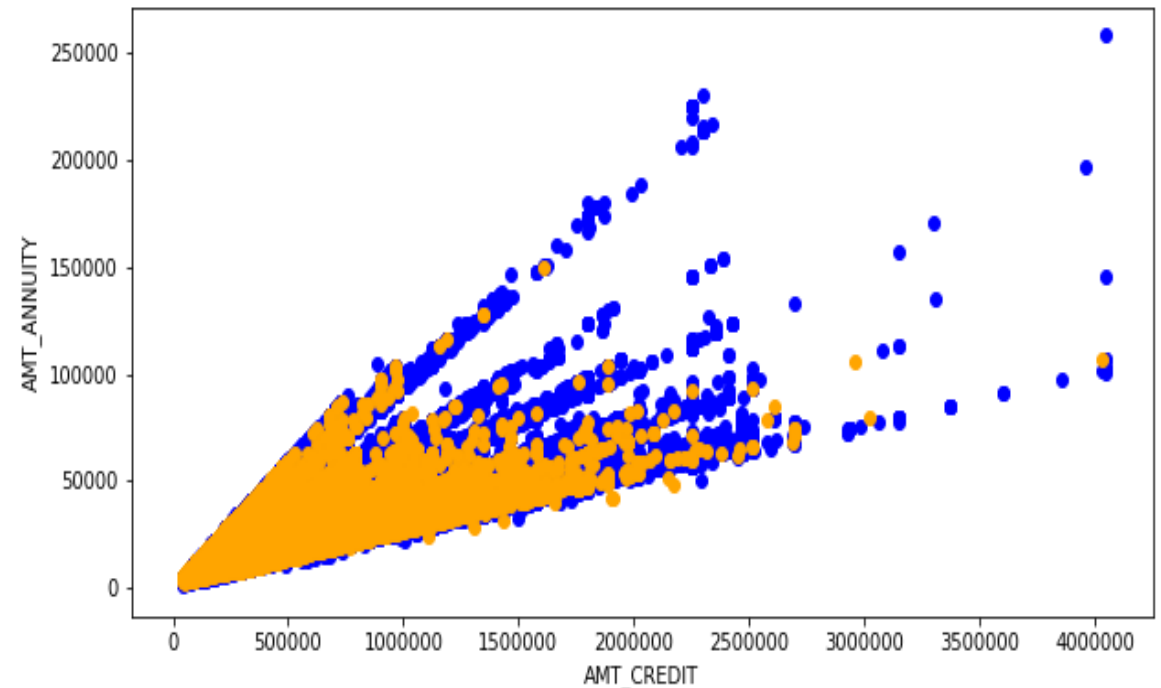
Continuous - Continuous

AMT_CREDIT: Credit amount of the loan

AMT_ANNUIITY: Loan annuity

Comment:

Loan annuity growth is proportional to credit amount of the loan in majority of the cases against the defaulted and non- defaulted clients.



Continuous-Categorical

CODE_GENDER: Gender of the client

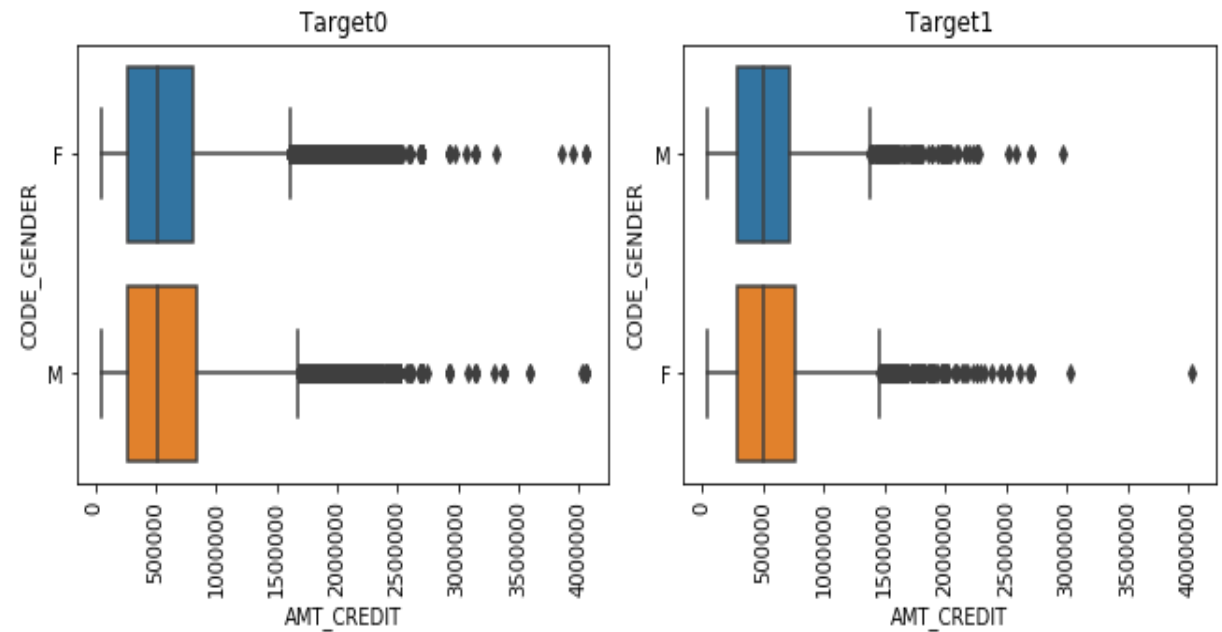
AMT_CREDIT: Credit amount of the loan

Comment:

Credit amount of the loan is equal for both male and female non-defaulted customers.

Defaulted female customers have a higher credit amount of the loan than defaulted male customers.

Non defaulted customers have higher credit amount of the loan than the defaulted customers.



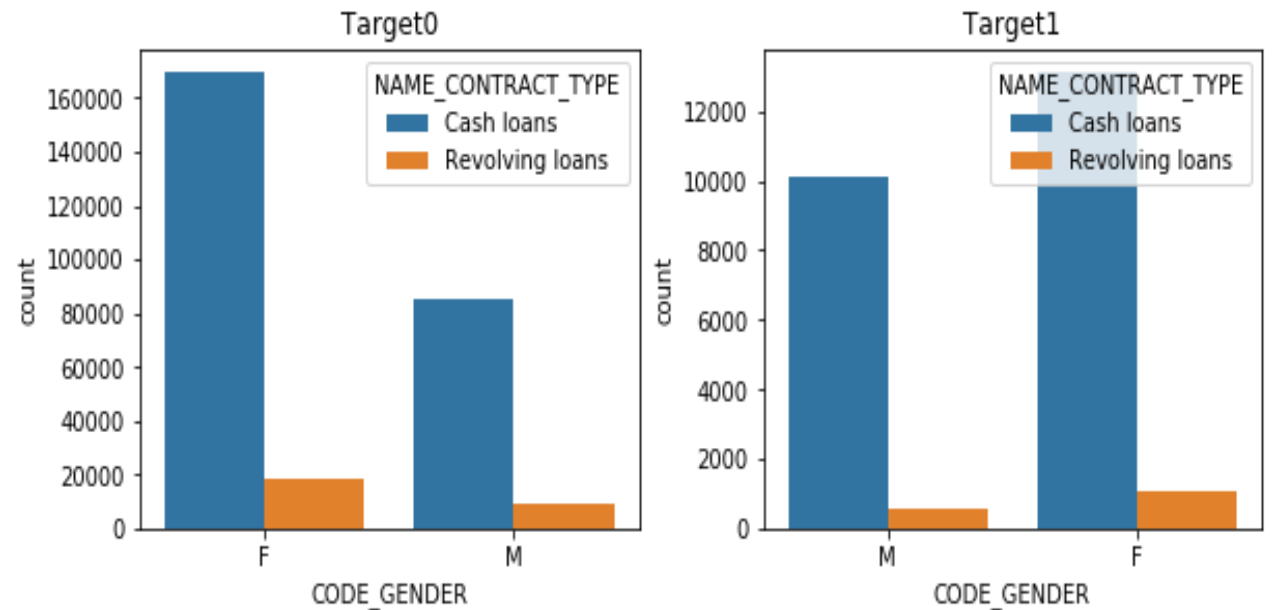
Categorical-Categorical

CODE GENDER: Gender of the client

NAME CONTRACT TYPE: Contract product type (Cash loan, consumer loan [POS] ,...) of the previous application

Comment:

There are more number of Female customers who have taken cash loans and Revolving loans compared to Male customers in both cases (i.e., Non-Defaulted(Target = 0) and Defaulted (Target = 1))



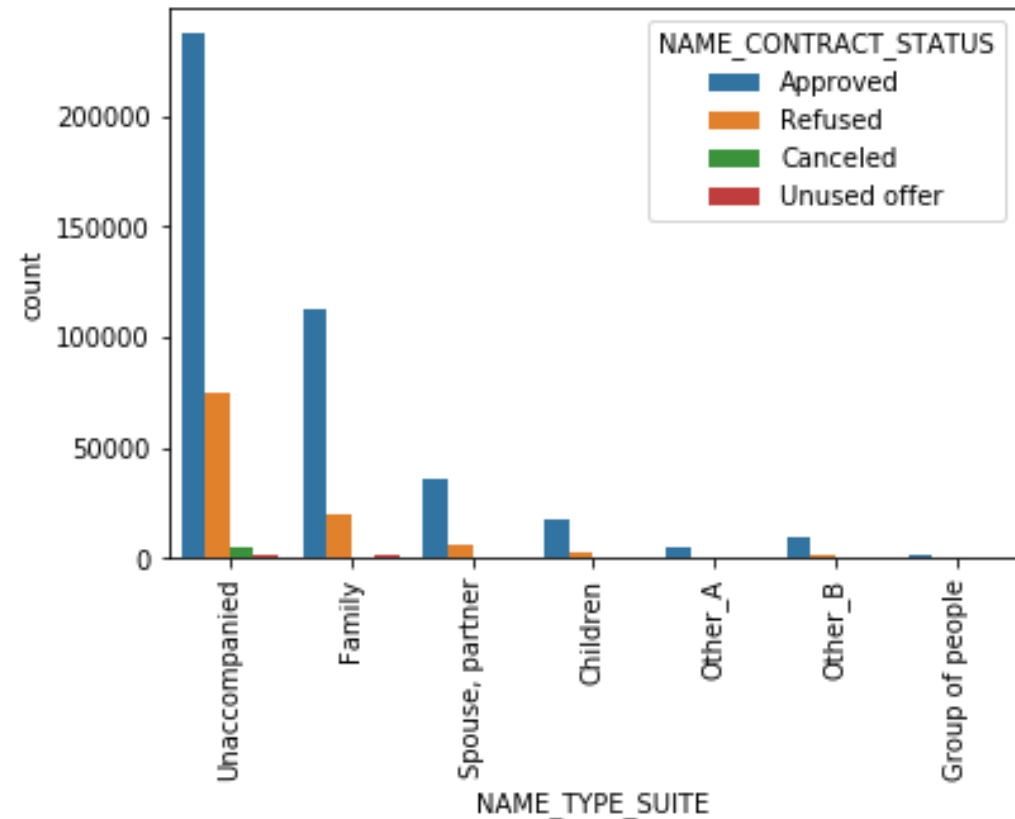
Analysis for application data.csv + previous application.csv

NAME TYPE SUITE: Who was accompanying client when he was applying for the loan

Comment:

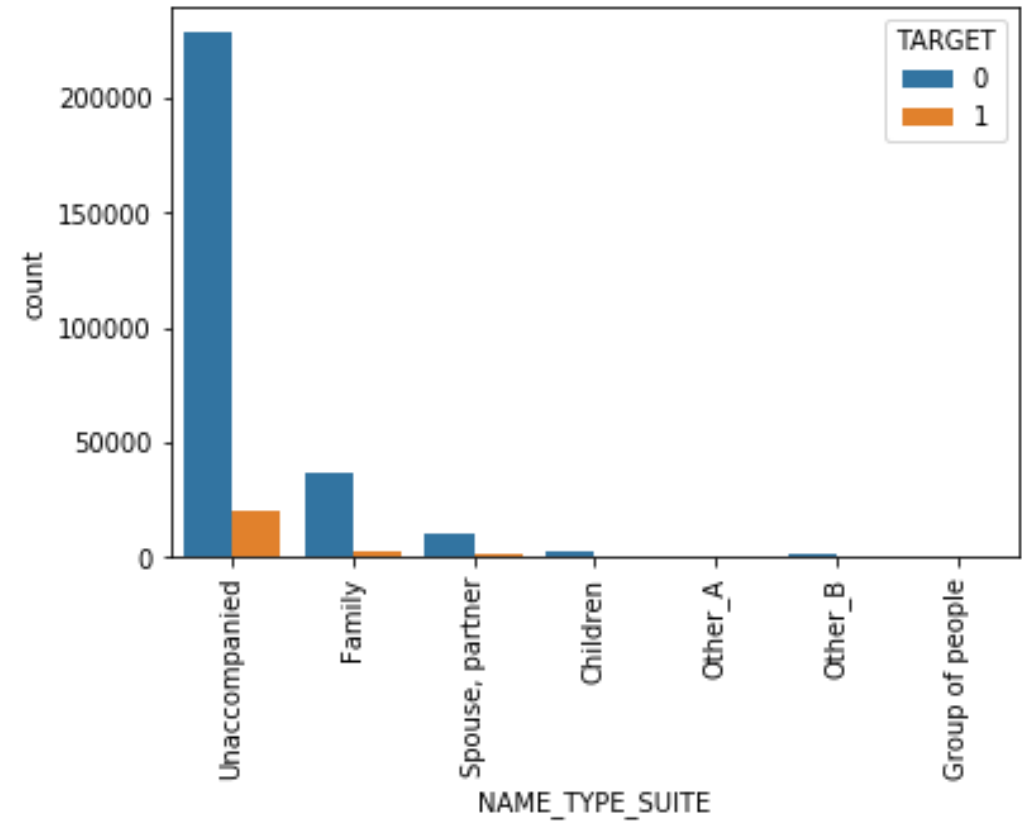
Population who are unaccompanied have high number of approved loans compared to others who had company.

The same set of people who are unaccompanied have high number of refused, cancelled and unused offers compared to others who had company.



Number of Approved loans for unaccompanied people is much higher compared to other categories.

Non-defaulted people have high number of approved loans compared to defaulted people.



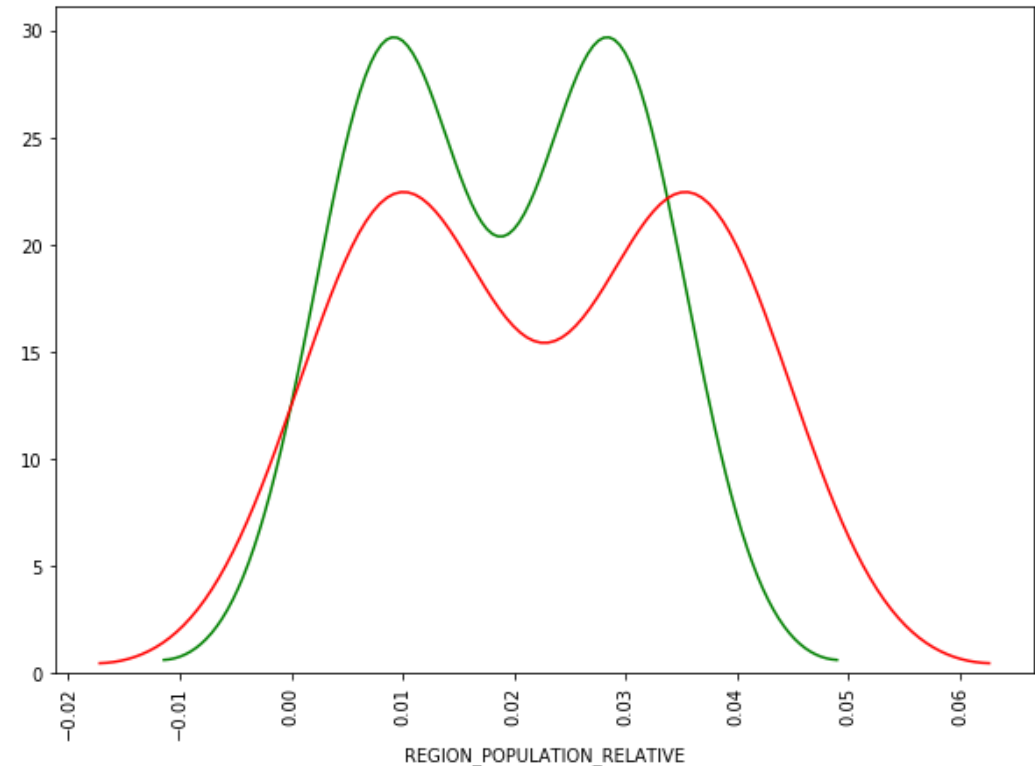
Univariate analysis for application data.csv + previous application.csv

Continuous:

REGION_POPULATION_RELATIVE - Normalized population of region where client lives (higher number means the client lives in more populated region)

Comment:

As we can observe from the above plot client with Status as approved are lesser at highly dense population region as compared to client with status as refused



Bi-variate analysis for application data.csv + previous application.csv

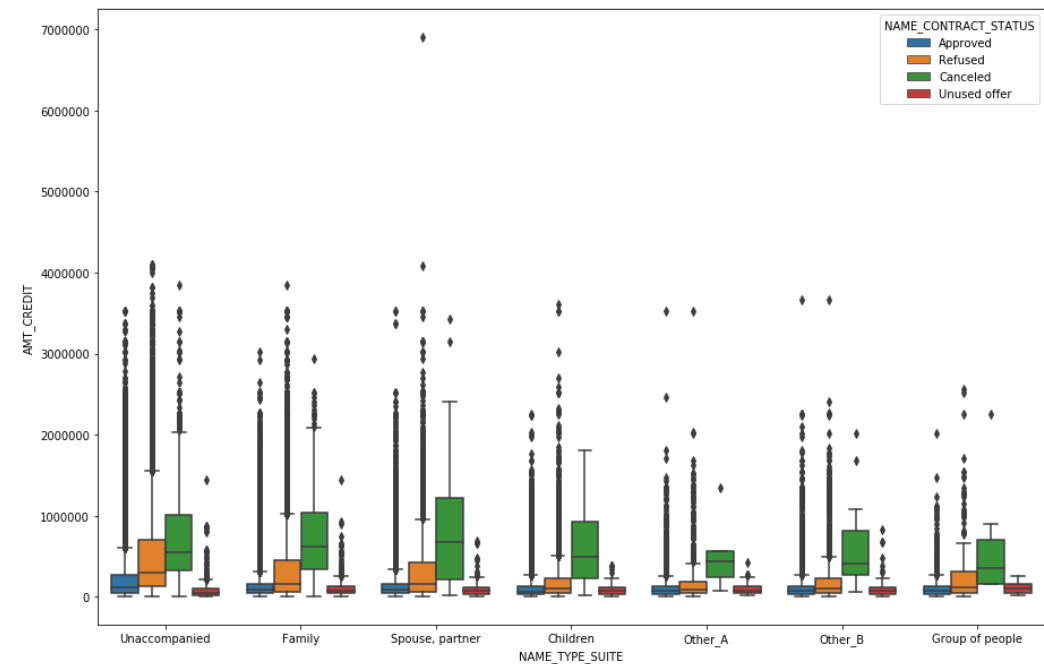
Continuous - Categorical

NAME CONTRACT STATUS: Contract status (approved, cancelled, ...) of previous application

AMT CREDIT: Credit amount of the loan

Comment:

Loan status is Refused for the clients who have high credit amount of the loan as compared to Approval, Cancellation and Unused offers



Continuous – Continuous

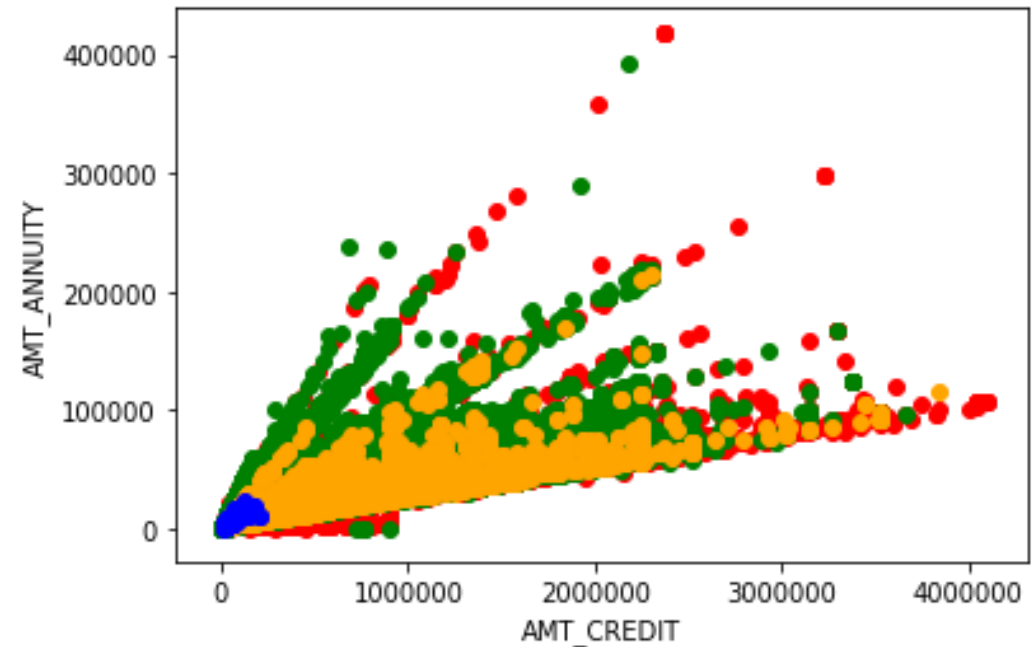
AMT_CREDIT: Credit amount of the loan

AMT_ANNUITY: Loan annuity

COMMENT:

Loan annuity growth is proportional to credit amount of the loan in majority of the cases where the loan is approved, refused and cancelled.

Loan annuity for unused offers has a very small amount for credit



Categorical – Categorical

WEEKDAY_APPR_PROCESS_START: On which day of the week did the client apply for the loan
NAME_CONTRACT_STATUS: Contract status (approved, cancelled, ...) of previous application

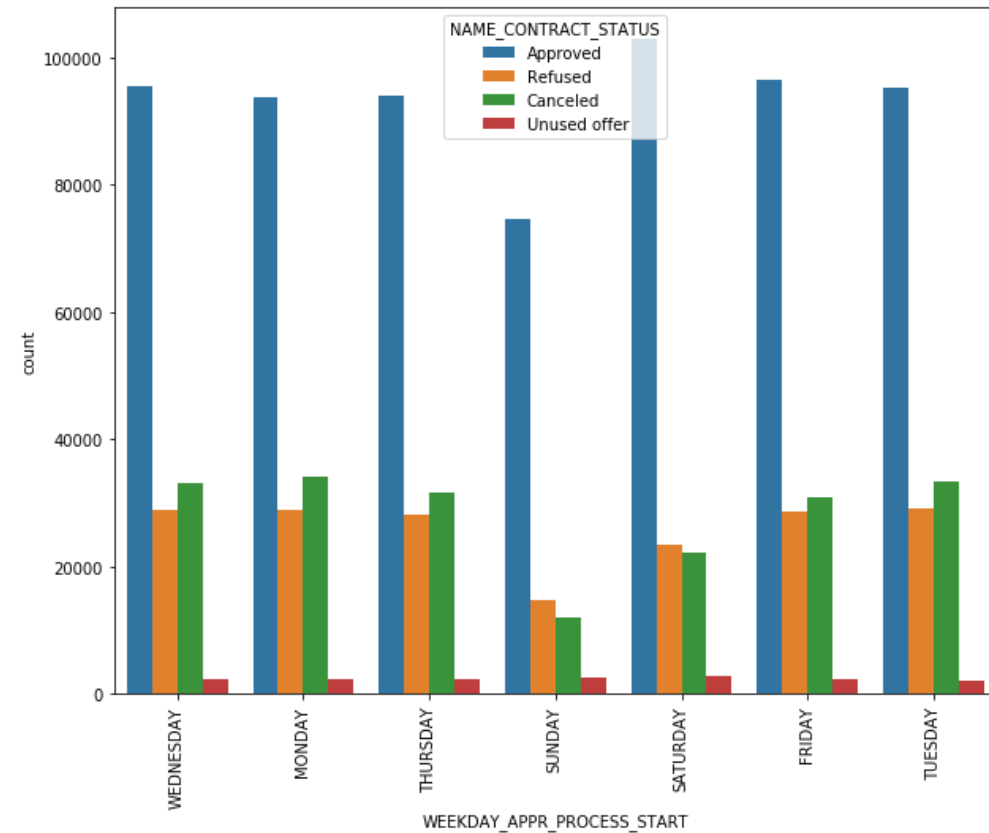
COMMENT:

Approved cases are more in number on Saturday compared to all other days.

More number of loans were canceled on Monday and Tuesday.

On Saturday and Sunday, lesser number of loans were refused compared to other days.

On all days number of unused offers were constantly the same.



Thank you.

