Presentation

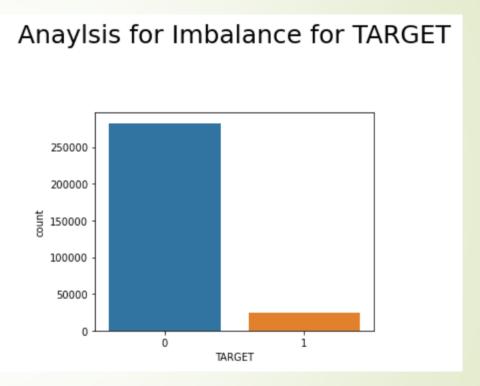
On
Credit EDA Case
study by
Viniba D Ponraj

Purpose

Credit risk analysis will help the company to make a decision for loan approval based on the applicant profile. Which controls the loss of business to the company and avoid financial loss for the company

Anaylsis for Imbalance in Target variable

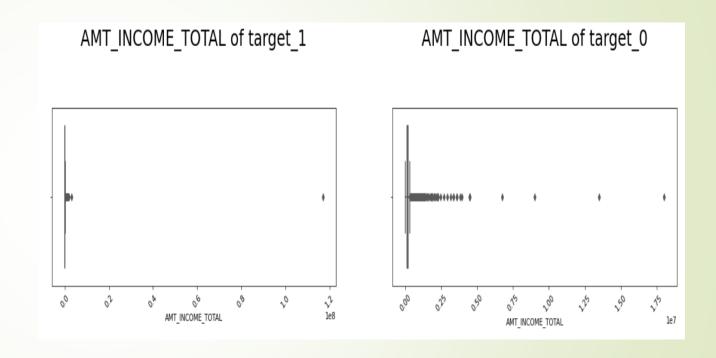
- Target 1 represents payment difficulties
- Target 0 represent
 remaining customer have no difficulties



Continuous Univariate analysis

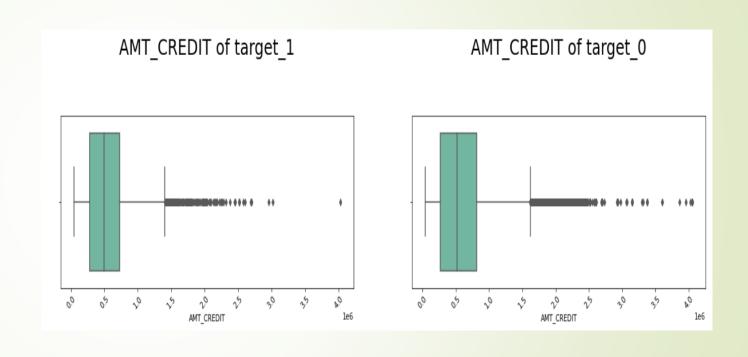
Frequency Distribution of Income amount

- There is a huge difference between the maximum value and the 99th quantiles.
- Total income vary from person to person.
- Some outliers are there in amount target.



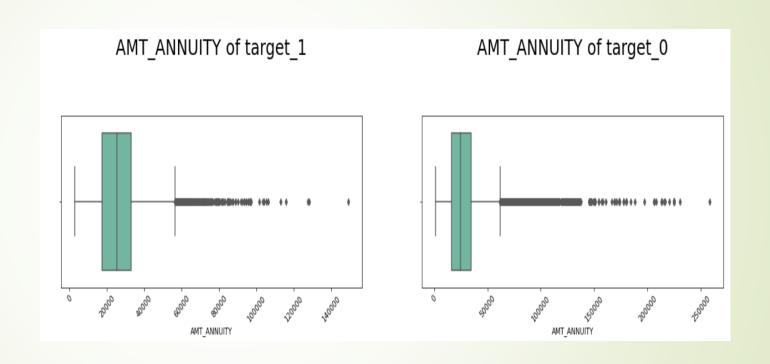
Frequency Distribution of Amount Credit

- Amount credited varies from person to person.
- Low quantile is smaller than middle quantile
- Middle quantile is smaller than upper quantile.
- Outliers are there above 99% quantile



Frequency of Amount Annuity

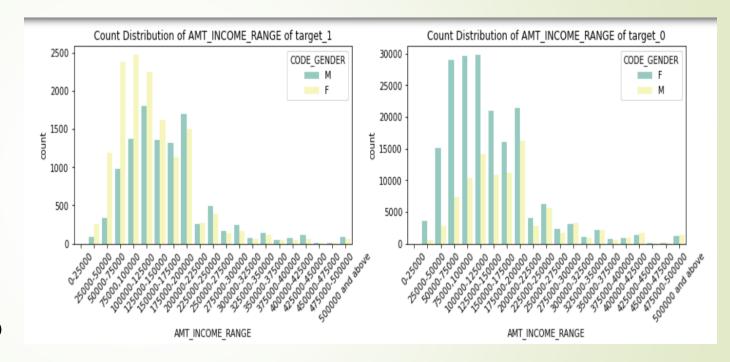
- Some outliers are there in annuity.
- Most of the customers annuity approximaly 60K



Continuous Univariate analysis for segmented variable

Distribution of Income range

- Very less count for income range 400000 and above.
- High income starts from 50000-75000 to 175000 to 200000
- Female have high income range in target 1 and target 0



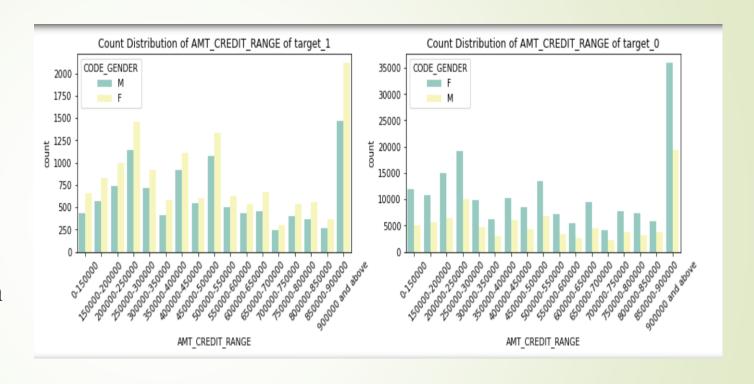
Distribution of Age range

- High age range of female from 30-40 in target 1
- High age range of female from 30-40,40-50,50-60 in target 0



Distribution of Amount Credit range

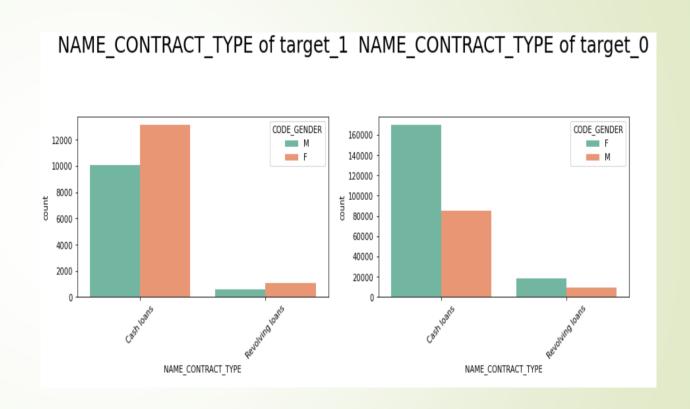
- High credit amount range is 900000 and above for female for both target 1 and target 0
- Minimum amount credit range 650000 to 700000 in both target 1 to target 0



Categorical univariate analysis for target 0 and target 1

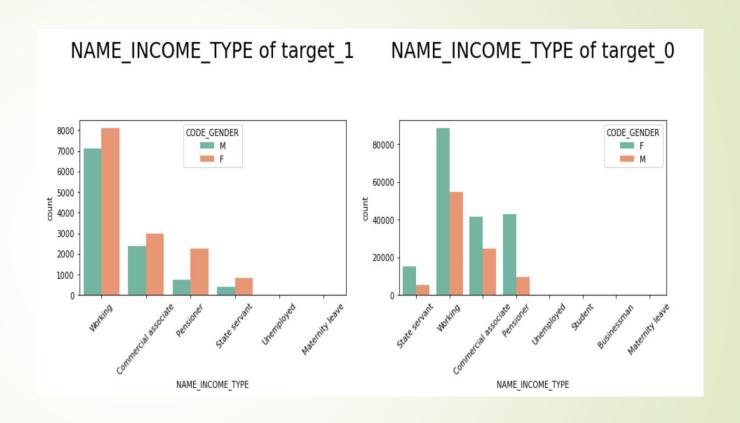
Count distribution of NAME_CONTRACT_TYPE

- There is higher cash loan than revolving loan
- Female customers are leading in applying loans



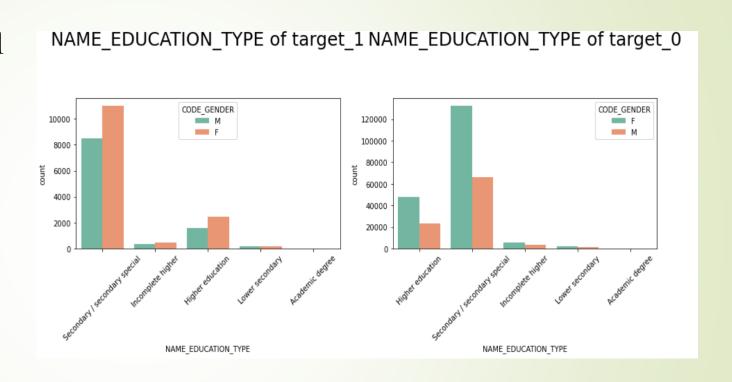
Count distribution of NAME_INCOME_TYPE

- Female have higher income type for working
- Least income type for state servant
- There is no income type for unemployed, student and Maternity leave



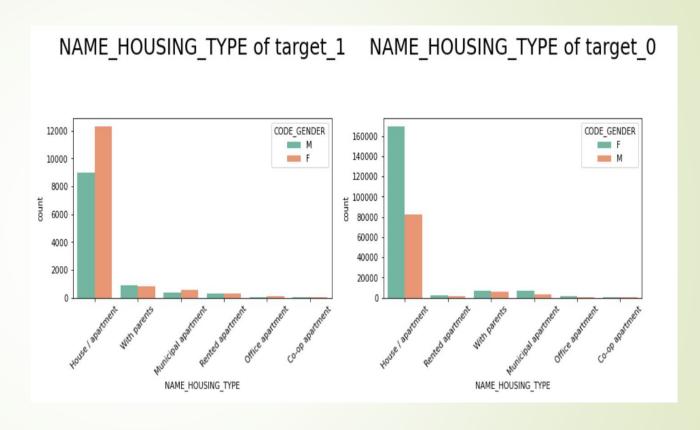
Count distribution for NAME_EDUCATION_TYPE

- Secondary/Secondary special is higher than any other education type
- Female are leading than men in education type
- Lower secondary and Academic degree have least in Education type



Count Distribution of HOUSING_TYPE

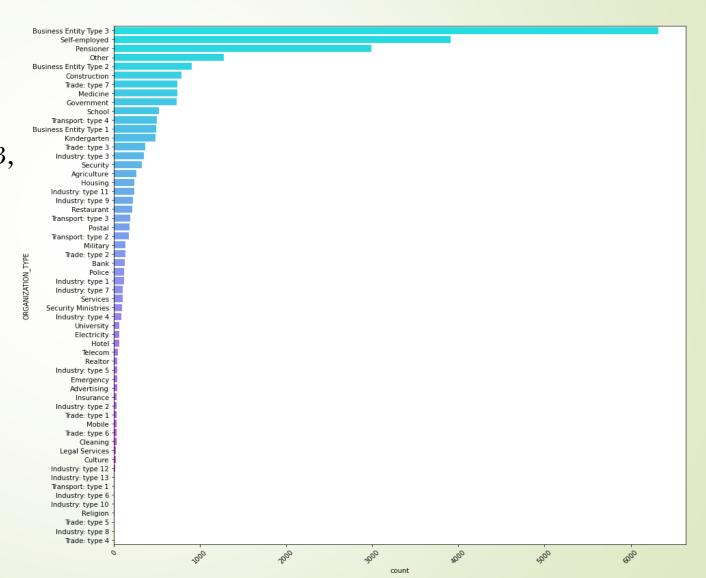
- House/Apartment is leading than other housing type
- Co op apartment and office apartment have least in number



Count Distribution of ORANIZATION_TYPE

Count Distribution of ORGANIZATION_TYPE of target_1

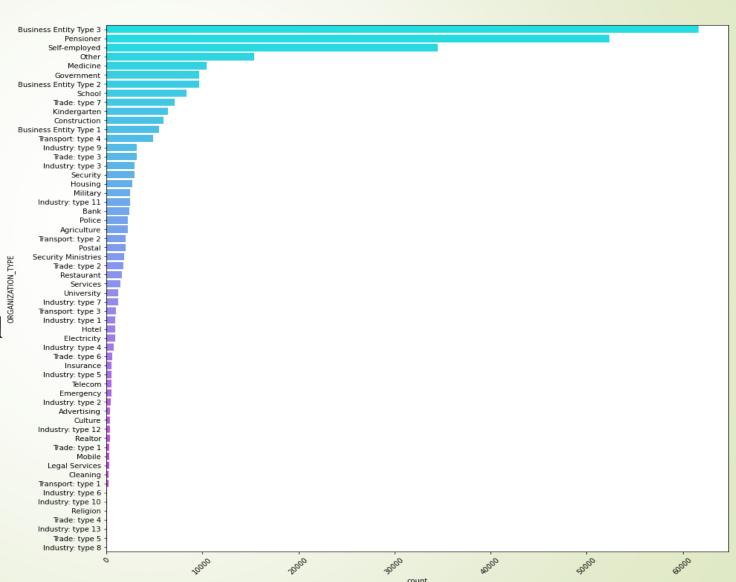
- Business Entity Type 3 have highest Organization type
- Industry type 12,Industry 13,
 Transport type 1, Industry
 type 6,Industry type
 10,Religion,Trade type 5,
 Industry type 8 and Trade
 type 4 have lease in
 Organization type



Count distribution of ORGANIZATION TYPE

Count Distribution of Organization type of target_0

- Business Entity Type 3
 have highest Organization
 type
- Industry type 12,Industry
 13, Transport type 1,
 Industry type 6,Industry
 type 10,Religion,Trade
 type 5, Industry type 8 and
 Trade type 4 have lease in
 Organization type



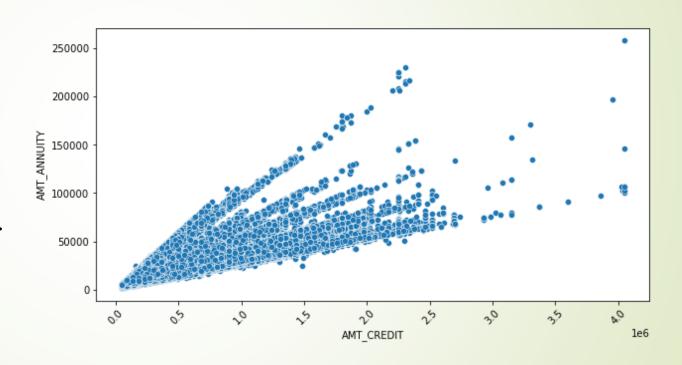
Bivariante Analysis

Continuous Vs Continuous Variables

Relationship between AMT_CREDIT VS AMT_ANNUITY

Relationship of AMT_CREDIT Vs AMT_ANNUITY

- Amount credit and amount annuity have strong positive coorelation.
- If Amount credit increases amount annuity also increases.

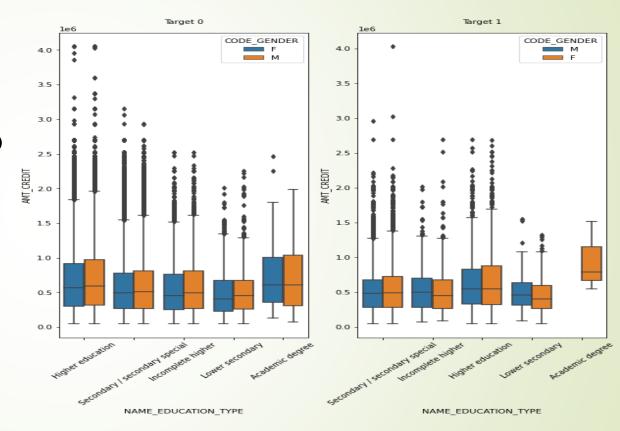


Bivariante Analysis

Continuous Vs Categorical Variables

Analysis of NAME_EDUCATION_TYPE VS 'AMT_CREDIT VS 'CODE_GENDER

- Clients with `Academic Degree`
 have a wide range of credits for
 Target 0 whereas the range is
 much lower for ones with Target 0
- Clients with `Academic Degree` and Payment difficulties take mean and median credit at a much higher range than On-Time Payment clients
- / `Male` clients with `Academic Degree` always pay the loan ontime

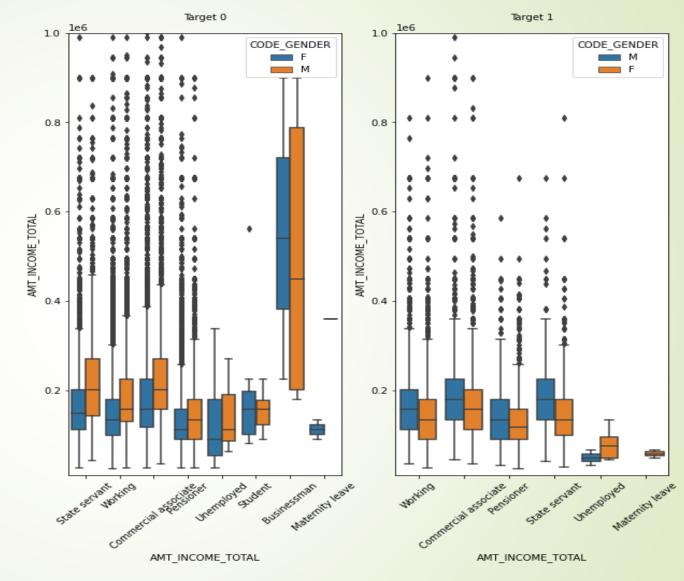


Analysis of `NAME_INCOME_TYPE` V/S `AMT_INCOME_TOTAL`

V/S 'CODE_GENDER'

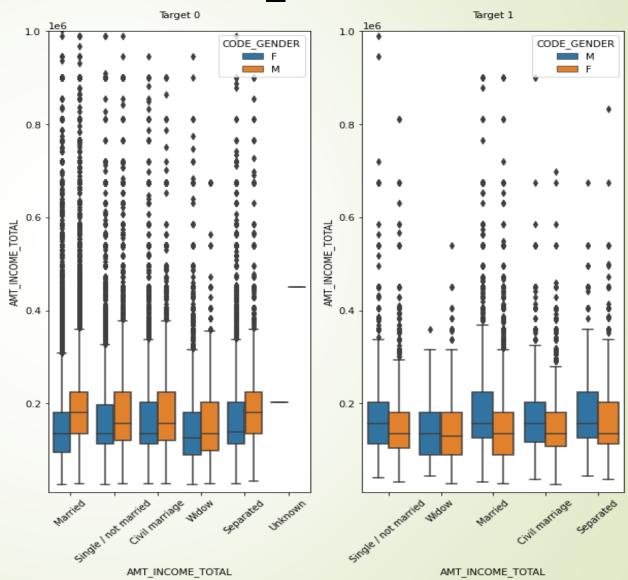
Businessman clients do their payments on time

- Student client do their payments on time
- Clients who are `Unemployed` and `Male` have a very high income in Target 0 than Target 1
- Chents who are in `Maternity Leave` and `Female` have a very high income in Target 0 than Target 1



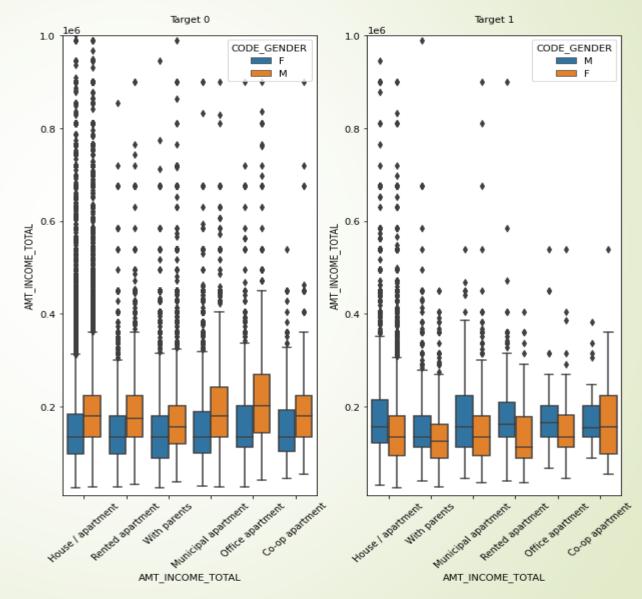
Analysis of `NAME_FAMILY_STATUS` V/S `AMT_INCOME_TOTAL` V/S `CODE_GENDER`

 `Married` clients have a slightly higher mean/median income with Target 0 than Target 1



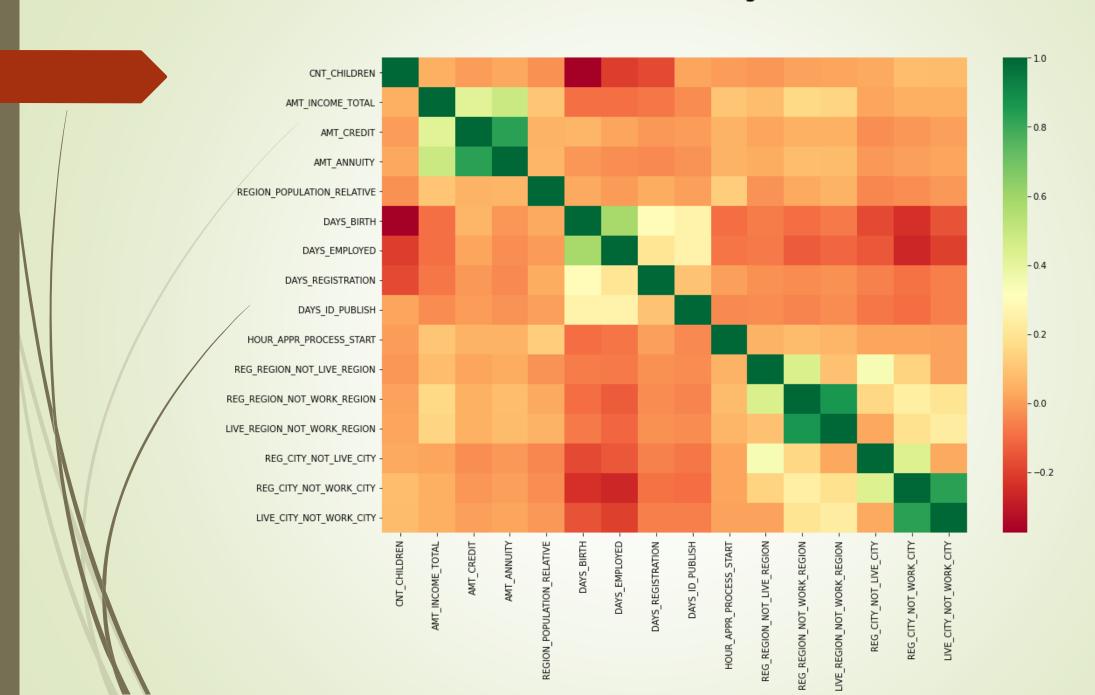
Analysis of `NAME_FAMILY_STATUS` V/S `AMT_INCOME_TOTAL` V/S `CODE_GENDER

- Client who in office apartment pay loan on time
- Client who in Co op apartment have high range in Target 1 than target 0



Correlation of Target 0

Correlation for target 0



- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
- 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.
- 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 is higher to densely populated area.
- The income is also higher in densely populated area.

Correlation of Target 1

Correlation for target 1



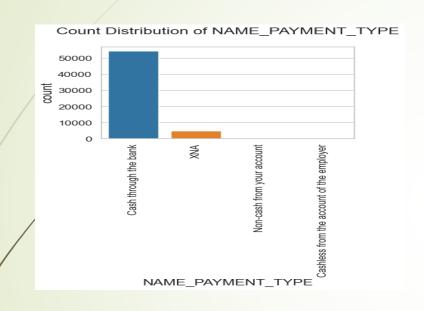
This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.

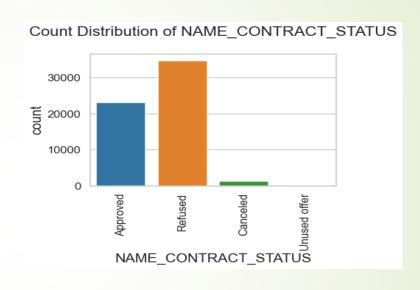
- The client's permanent address does not match contact address are having less children and viceversa
- The client's permanent address does not match work address are having less children and viceversa

Analysis of merged information about the client's previous loan data and current loan application

Univariate analysis of categorical variable

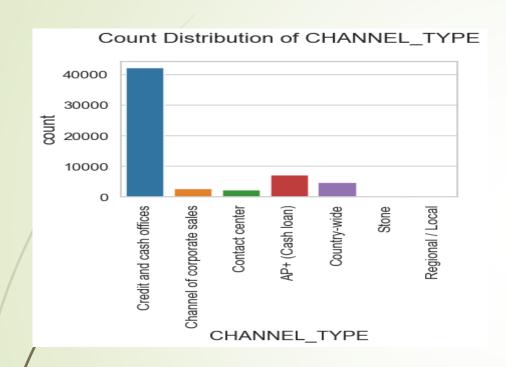
Count distribution of 'NAME_CONTRACT_STATUS' and NAME_PAYMENT_TYPE

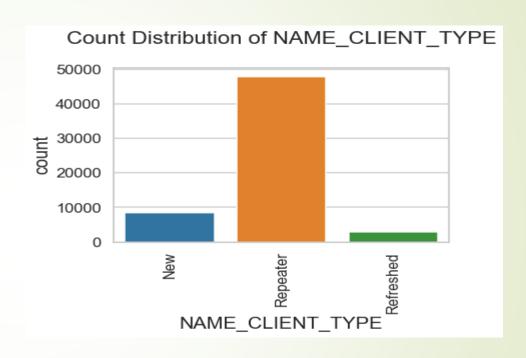




- Refused contract status is highest among all.
- Cash through the bank is highest among all.
- Approved contract status is second highest of contract status

Count distribution of CHANNEL_TYPE AND NAME CLIENT TYPE

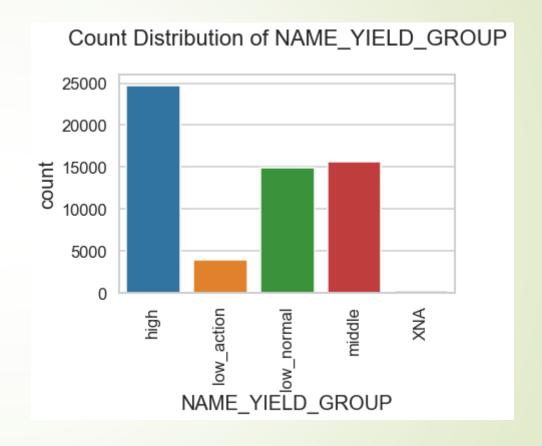




- Credit and cash offices in channel type are highest among all.
- Regular client are applying for bank loan than new client

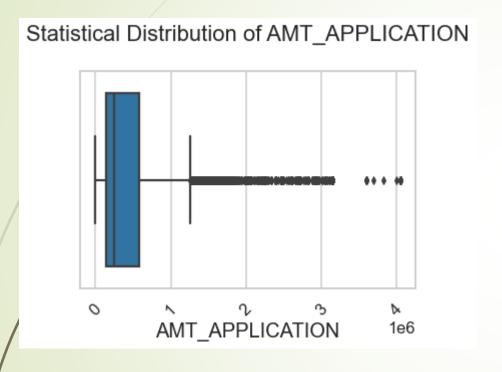
Count distribution of Name yield group

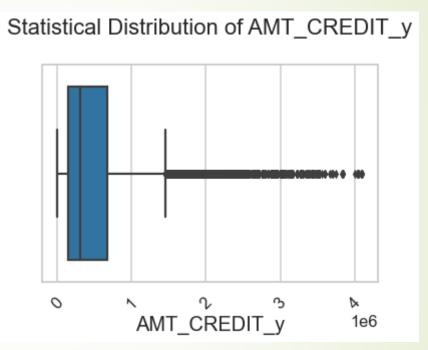
- XNA` interest rate is the highest among all loan applications
- 'middle' and 'high' interest rates are the second and third highest among all loan application



Univariate analysis of Continuous variables

Count Distribution of AMT APPLICATION AND AMOUNT CREDIT





- This distribution very closely resembles that of AMT_APPLICATION. This means that most people received the loan amount that they applied for.
- Most of the loan amount applied by the clients initially seems to be very small as can be seen from the huge spike at the beginning of the distribution

Correlation matrix

- `AMT_APPLICATION` has a high correlation with `AMT_ANNUITY`, `AMT_CREDIT_y`.
- `AMT_ANNUITY` has decent correlation with 'AMT_INCOME_TOTAL'

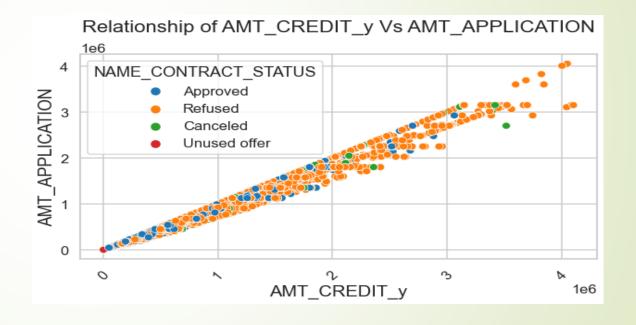


Bivariant Analysis

Continuous Vs Continuous Variable

Analysis of `AMT_APPLICATION` V/S `AMT_CREDIT_y` V/S `NAME_CONTRACT_STATUS`

 Application amount has strong positive correlation with Credit amount



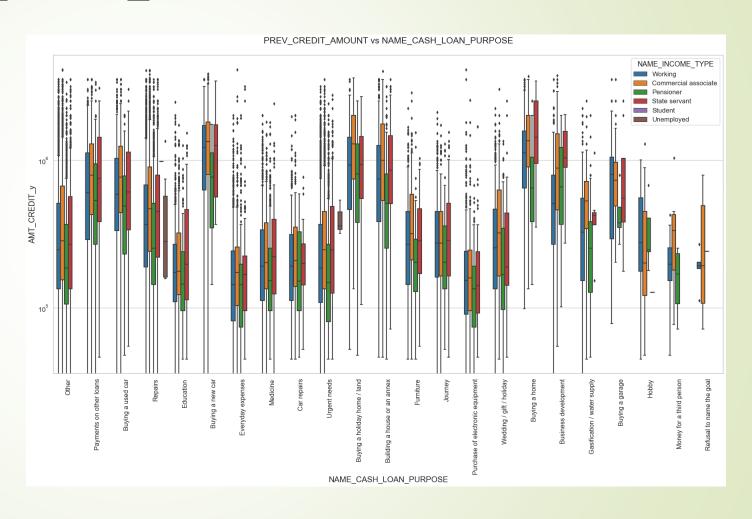
Continuous V/S Categorical variable

CREDIT AMOUNT VS NAME_CASH_LOAN_PURPOSE

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

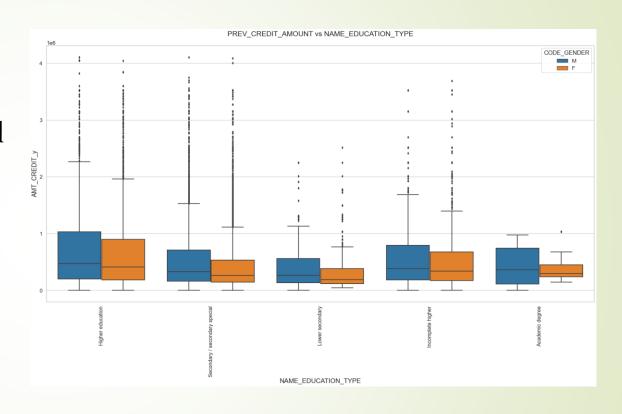
Income type of state servants have a significant amount of credit applied

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



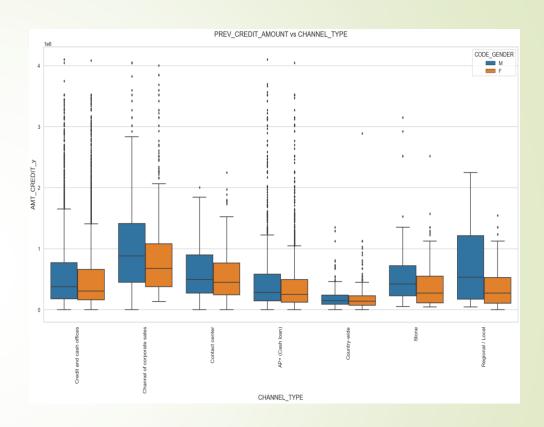
CREDIT AMOUNT VS NAME EDUCATION TYPE

- Credit amount for client with higher education and academic degree are high in range compared to other education status
- Credit amount for academic degree for female is comparatively less with male client
- Credit amount for lower secondary is least than other education type



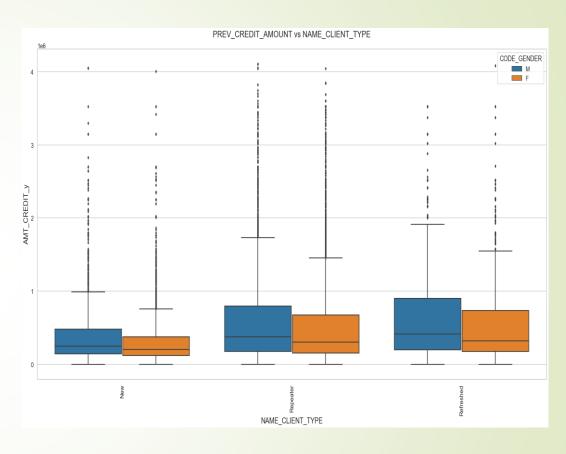
CREDIT AMOUNT VS CHANNEL_TYPE

- Credit amount for channel of corporate sales and Regional/Local are high range comparatively others
- Credit amount for Country wide is least
- Male has highest credit amount range than female



CREDIT AMOUNT VS NAME CLIENT TYPE

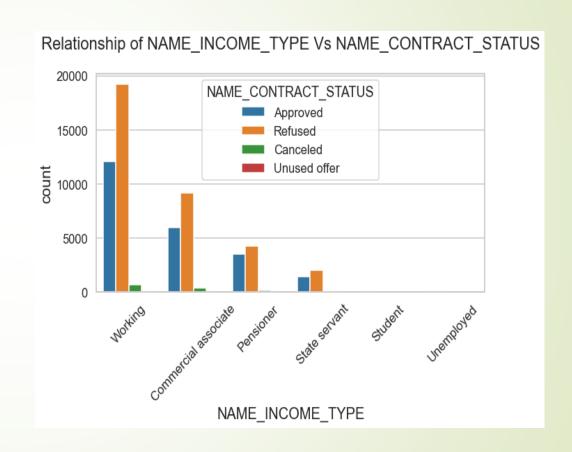
• Credit amount for new client will be least than refreshed and repeater client.



Categorical Vs Categorical Variable

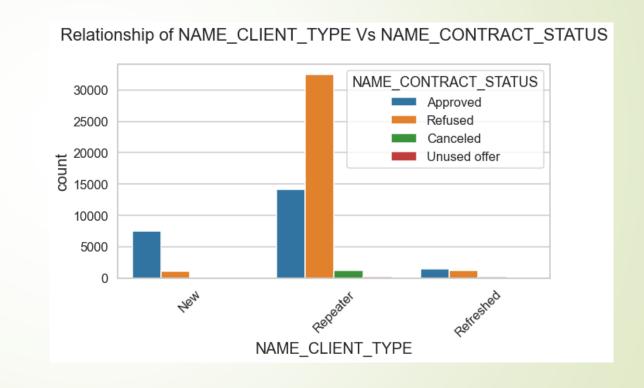
RELATIONSHIP OF NAME INCOME TYPE VS NAME CONTRACT STATUS

• Client who are working have most approval and refusal contract.



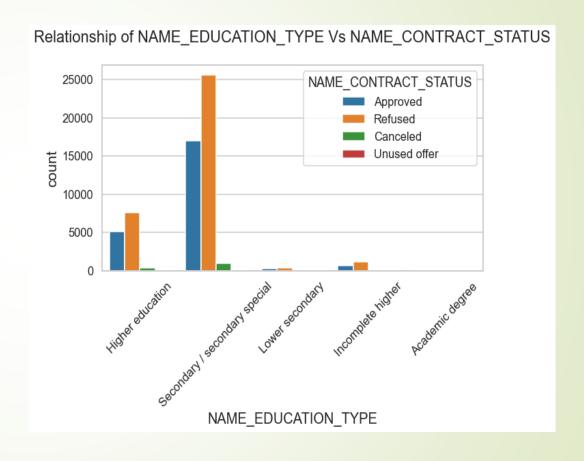
RELATIONSHIP OF NAME CLIENT TYPE VS NAME CONTRACT STATUS

- Repeater client have most approved contract status.
- Refreshed have least approval status
- New client have highest approval than refreshed



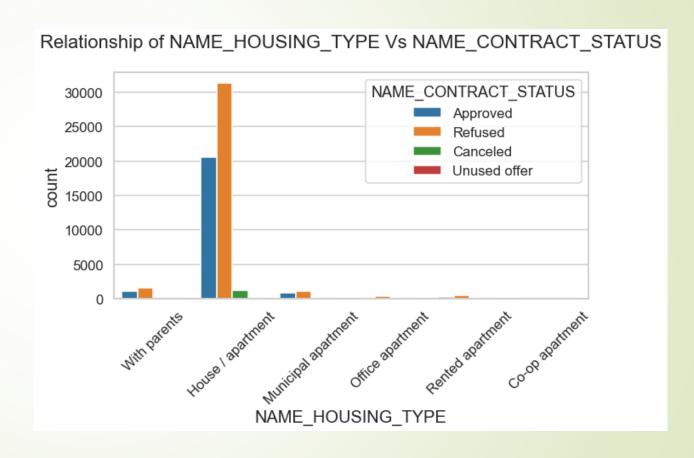
RELATIONSHIP OF NAME EDUCATION TYPE VS NAME CONTRACT STATUS

- Client with secondary/secondary special have highest approval and highest refusal contract status.
- Academic degree have lease contract status.



RELATIONSHIP OF NAME HOUSING TYPE VS NAME CONTRACT STATUS

- Co-op apartment has least contract status.
- House /apartment has highest approval and refusal contract status



CONCLUSION

Client categories to be targeted for providing loan:

- Banks should focus more on contract type 'Student', 'pensioner' and 'Businessman'
- Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.
- Clients who are married
- Repeater clients
- Male clients with academic degree.
- \$tudents and businessman