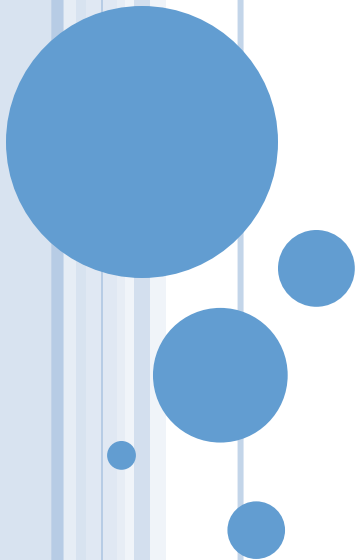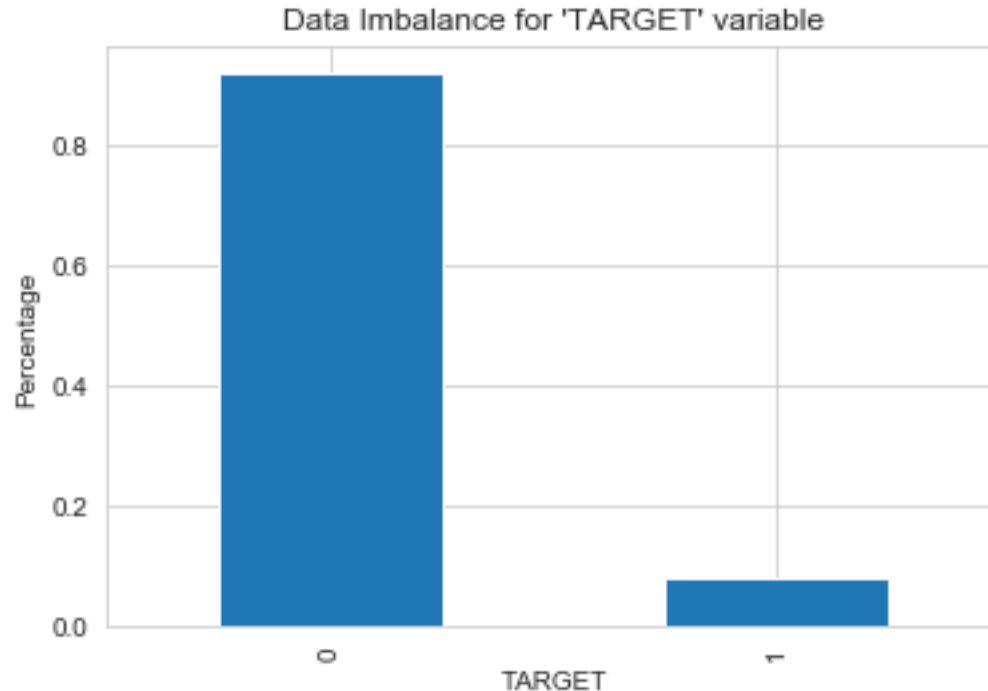# CREDIT EDA CASE STUDY

By Sai Priyesh Chittoori
Mayank Vinocha

ANALYSIS !

- There is data imbalance in the application dataset for TARGET variable.
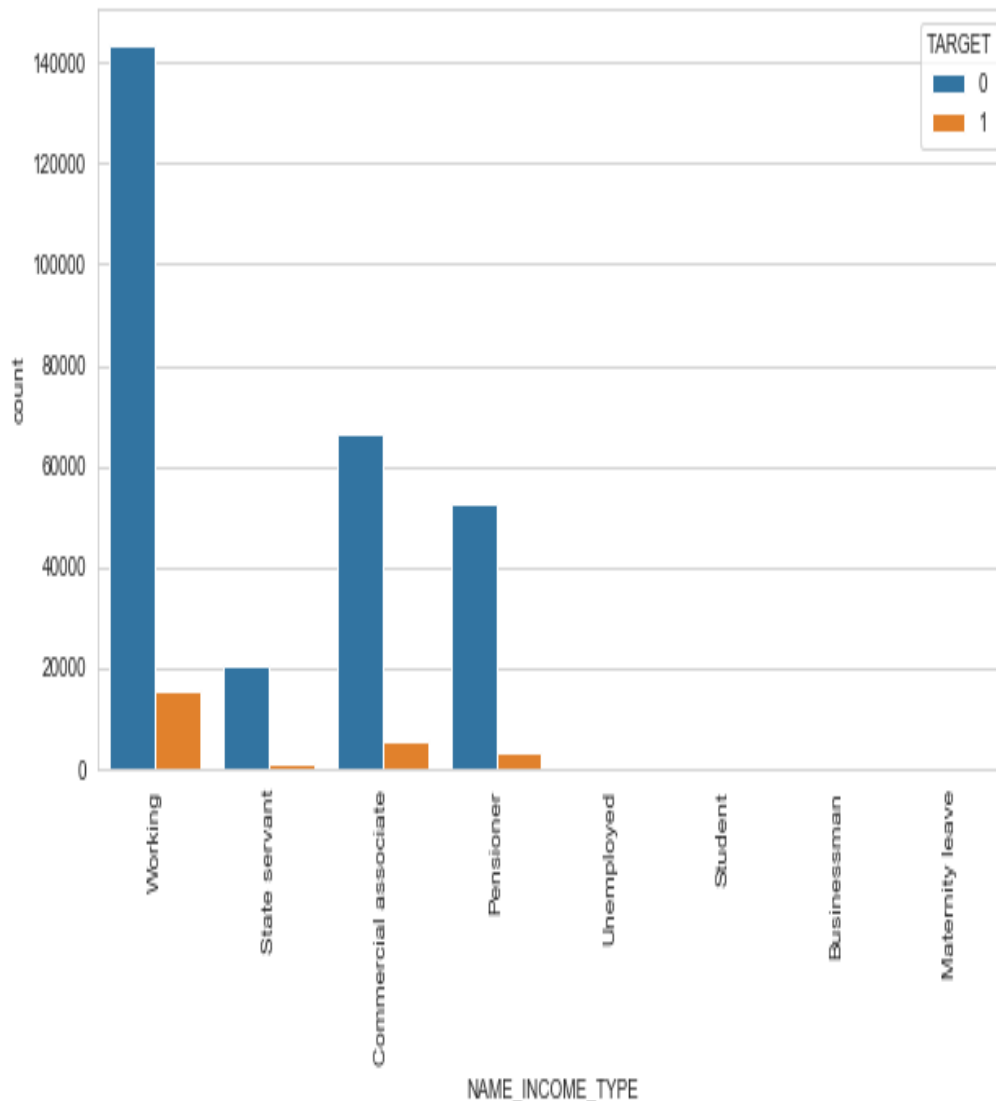
Ratio of data imbalance is **91.9 : 8.1**

Data Imbalance for 'TARGET' variable

'1' is The client with payment difficulties.

'0' is All other cases when the payment is paid on time.

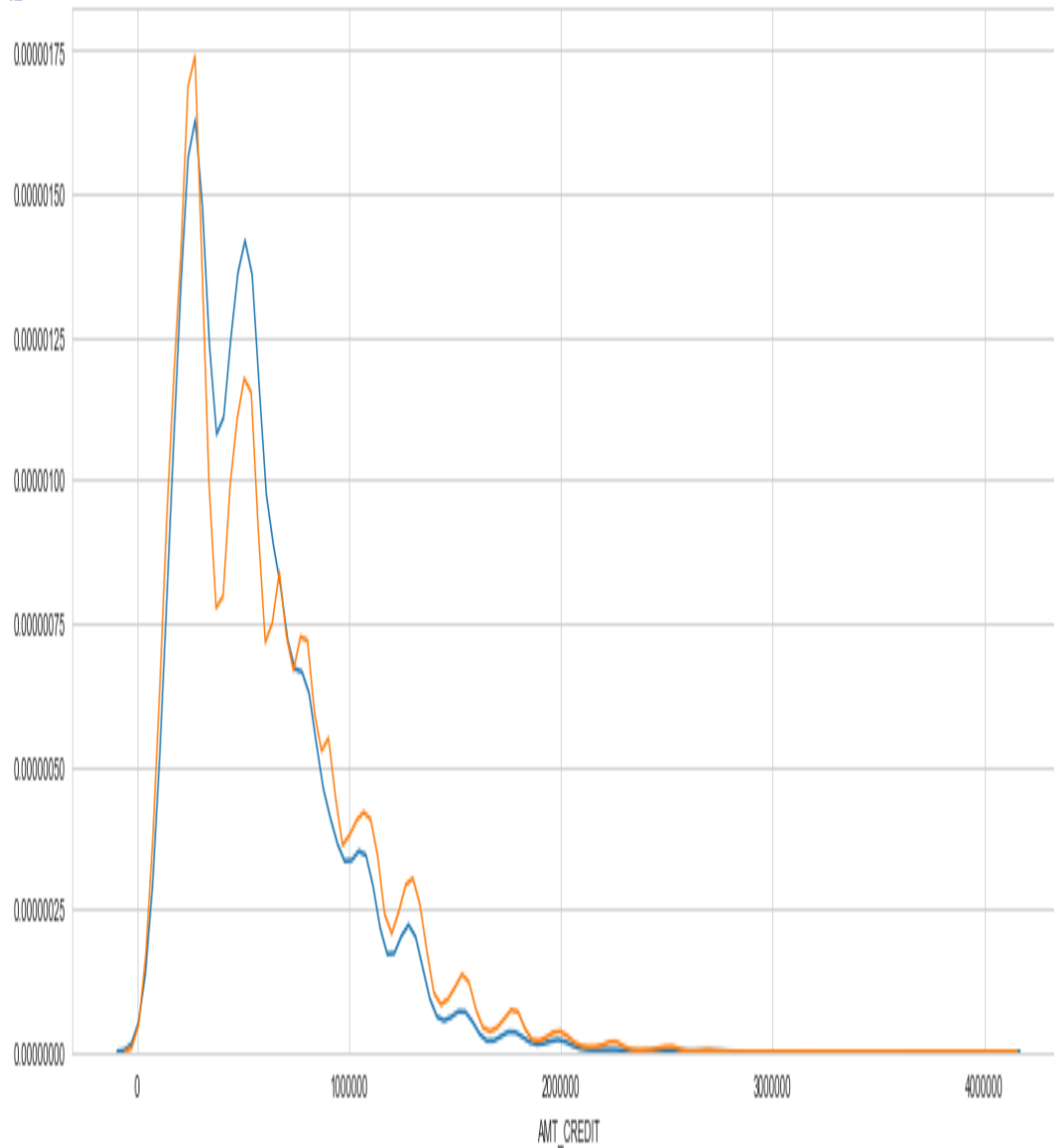# CATEGORICAL UNIVARIATE ANALYSIS WITH BOTH 0 AND 1 TARGET VARIABLE!

Points to be concluded from the graph:

It is a categorical univariate graph for the name income type with Target 0 and Target 1.

- We observed that working professional's are having high count of payment difficulties as well as high count of no payment difficulties than other income types. Where as there is a high ration of no payment difficulties.

- The state servant's, commercial associates and pensions are looking for loans and having high ration of no payment difficulties.

- In the same way students and unemployed persons are not seeking for loans.

# CONTINUOUS UNIVARIATE ANALYSIS WITH BOTH 0 AND 1 TARGET VARIABLE!

Points to be concluded from the graph:

It is a continuous univariate graph for the credit amount with Blue line refers TARGET 1, Orange line refers Target 0.
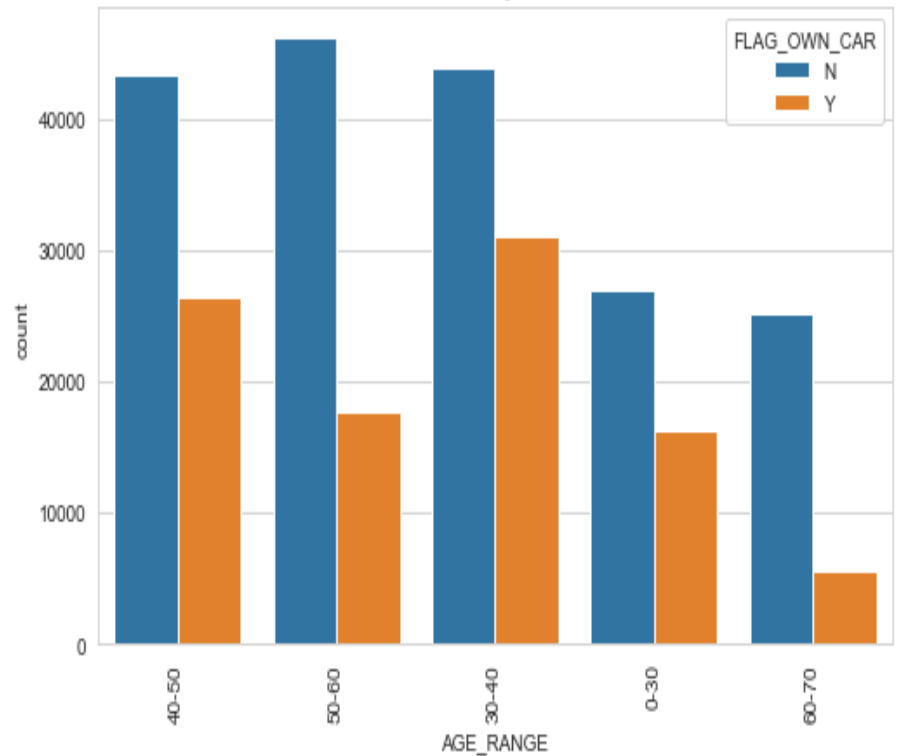
- We observed that the frequency of credit amount between than 25,00,000 - 70,00,000 have more clients are having payment difficulties compared to other clients.

- For the credit amount less than 25,00,000 or more than 70,00,00 more clients who have pay on time compared to clients who have payment difficulties.

CATEGORICAL BIVARIATE ANALYSIS WITH BOTH TARGET 0 & 1!
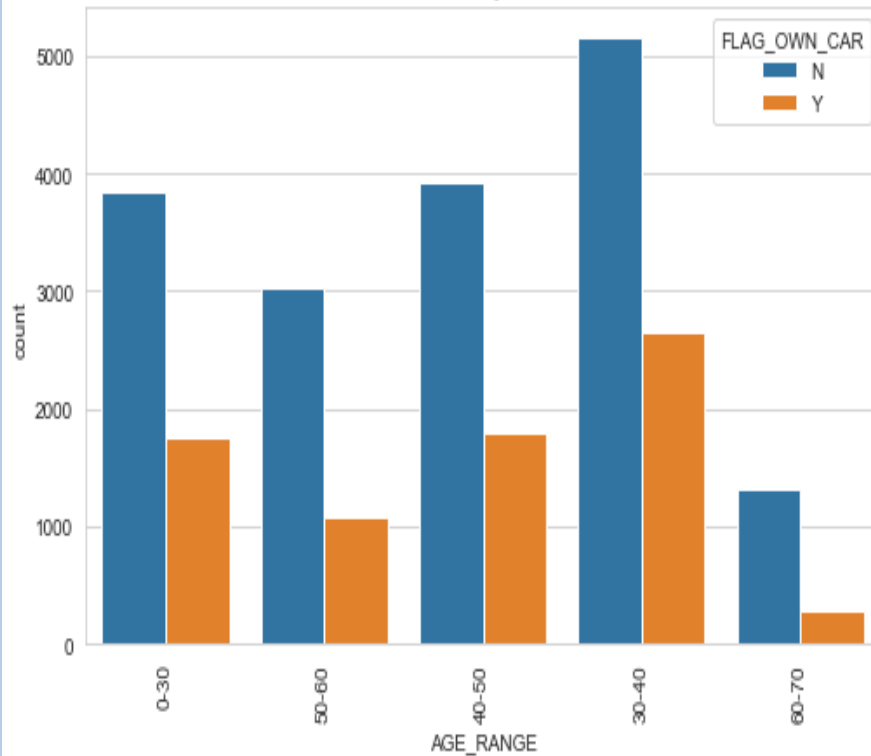
Points to be concluded from the above graphs are:

It is a categorical bivariate graph for the age range with Target 0 and Target 1.

- From the graph with payment difficulties(t-1), we observed that the clients who do not own a car and aged between 30-40 facing more payment difficulties as compared to other age group clients.

- From the graph with no payment difficulties(t-0), we observed that the clients who do not own a car and aged between 50-60 pay on time as compared to other age clients.

CONTINUOUS BIVARIATE ANALYSIS WITH BOTH TARGET 0 AND TARGET 1!
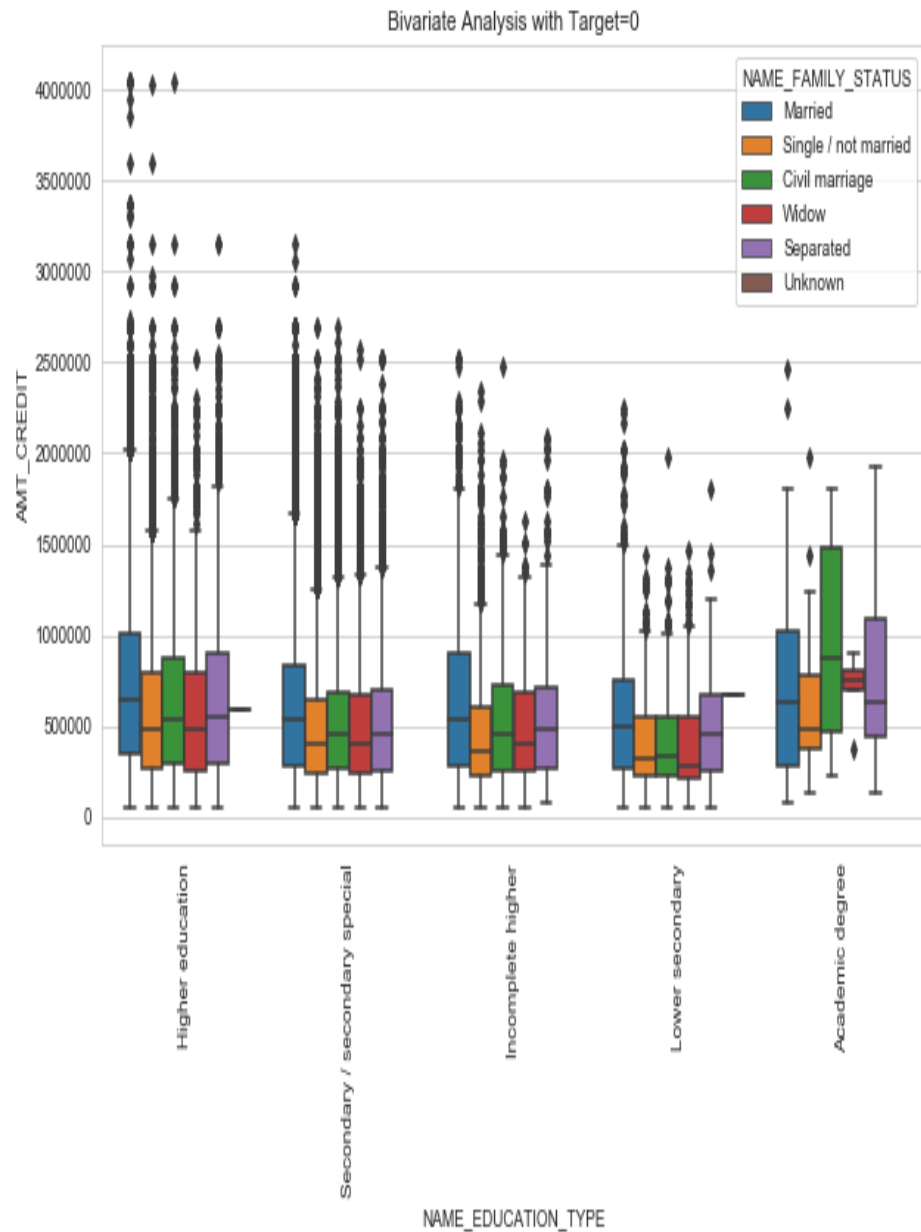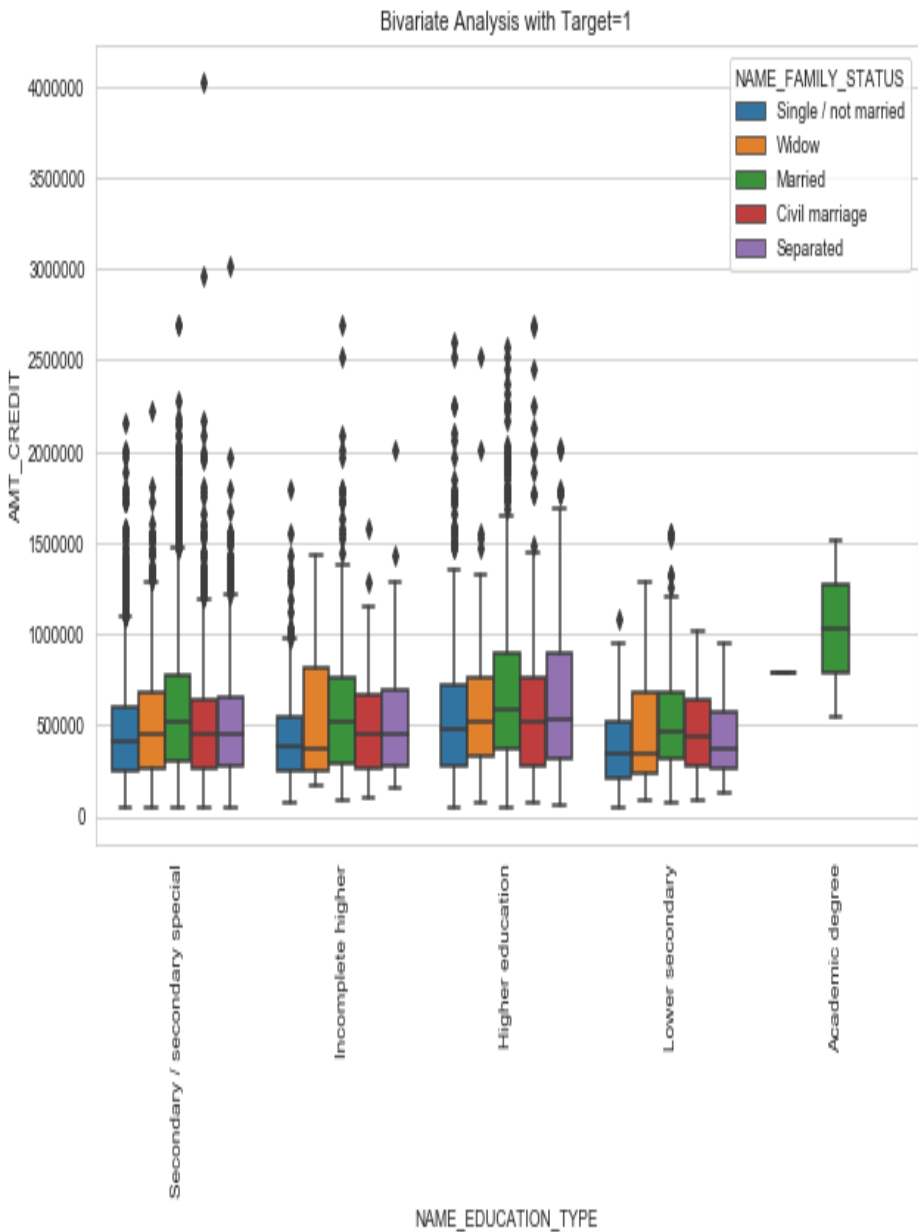
Points to be concluded form the above graphs:

It is a continuous bivariate graph for the name income type vs. amount credit of family status with Target 0 and Target 1.

TARGET 0 (No payment difficulties) :

- Family status of 'married', 'civil marriage', and 'separated' of Academic degree education are having higher number of credits than other clients.

- Higher education of family status on 'married', 'single' and 'civil marriage' are having more credit amount outliers.

- Lower secondary education of 'single', 'civil marriage' and 'widow' are having less number of credits.

- Civil marriage for Academic degree is having most of the credits in the third quartile.

TARGET 1 (Payment difficulties):

- We observed that Family status of 'married' of Academic degree education are having higher number of credits than other clients.

- Most of the credit amount outliers are from education type 'Higher education' and 'Secondary/Secondary special'.

- Lower secondary education of 'single', 'civil marriage' and 'widow' are having less number of credits

- The education type of incomplete higher with the family status of singles are having less credit amount.

CORRELATION FOR THE CLIENT WITH PAYMENT DIFFICULTIES(TARGET=1)!

- There is a correlation between client work phone and how many days before the application the person started current employment.

- There is a correlation between observation of client's social surroundings with observable 60 days past due default and observation of client's social surroundings with observable 30 DPD (days past due) default.

- There is a correlation between the price of the goods for consumer loans and credit amount of the loan.

- There is a correlation between rating of the region where client lives with taking city into account and rating of the region where client live.

- There is a correlation between the family members does client have and number of children that the client have.

- There is a correlation between the client's contact address does not match work address and the client's permanent address does not match work address.

- There is a correlation between the Loan annuity and the credit amount of the loan.

- There is a correlation between the price of the goods for consumer loans and the Loan annuity.

- There is a correlation between the client provide work phone and the client's age in days at the time of application.

- There is a correlation between the number of days before the application the person started current employment and the client's age in days at the time of application.

CORRELATION FOR THE CLIENT WITH NO PAYMENT DIFFICULTIES(TARGET=0)!

- There is a correlation between the client work phone and how many days before the application the person started current employment.

- There is a correlation between client work phone and how many days before the application the person started current employment.

- There is a correlation between the price of the goods for the consumer loans and number of days before the application the person started current employment.

- There is a correlation between rating of the region where client lives with taking city into account and rating of the region where client live.

- There is a correlation between the family members does client have and number of children that the client have.

- There is a correlation between the client's contact address does not match work address and the client's permanent address does not match work address.

- There is a correlation between the observation of client's social surroundings defaulted on 60 days past due and the observation of client's social surroundings defaulted on 30 days past due(DPD).

- There is a correlation between the price of the goods for consumer loans and the Loan annuity.

- There is a correlation between the Loan annuity and the credit amount of the loan.

- There is a correlation between the number of days before the application the person started current employment and the client's age in days at the time of application.

Merging the application file and previous application file!
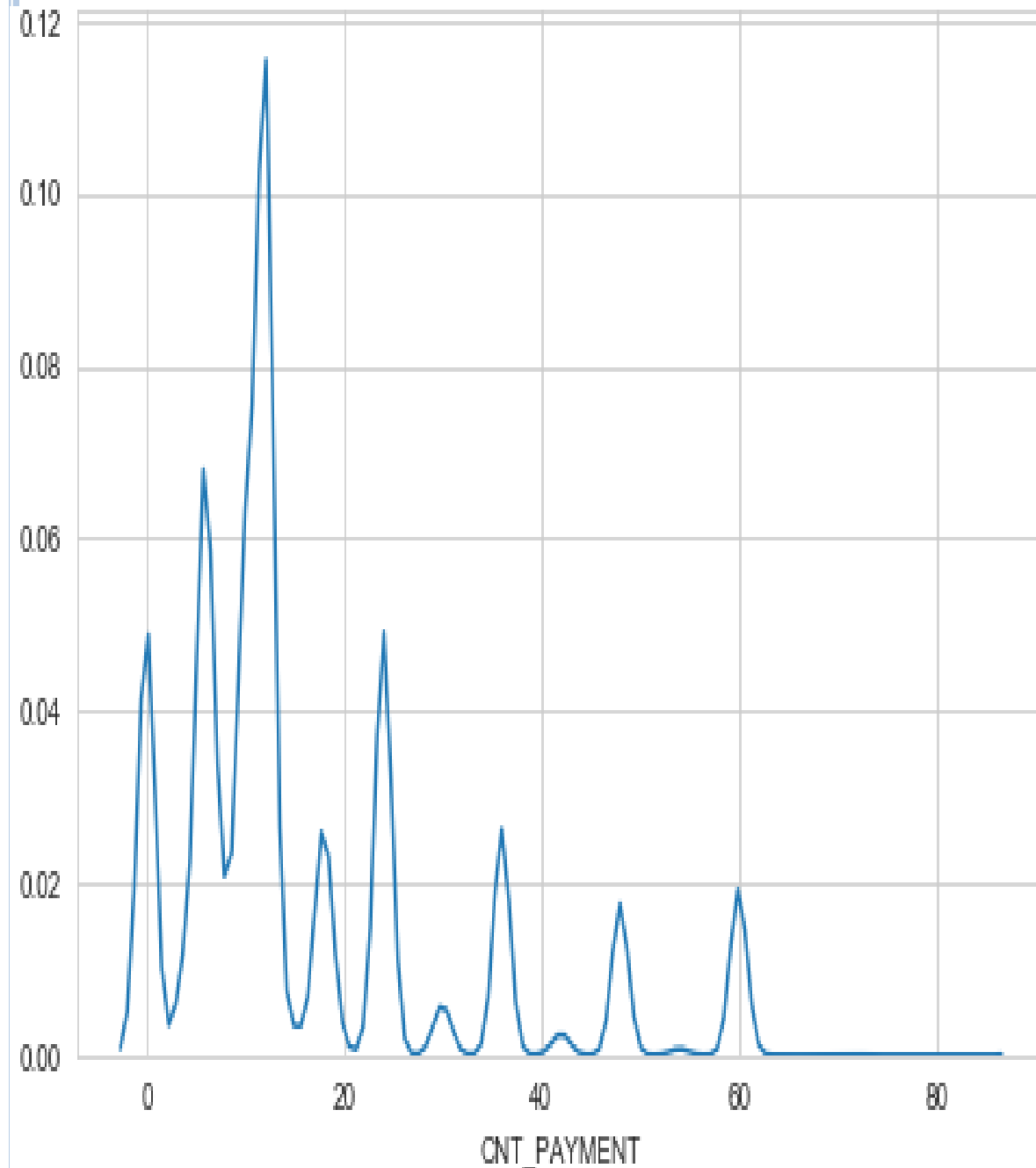
CATEGORICAL UNIVARIATE ANALYSIS!
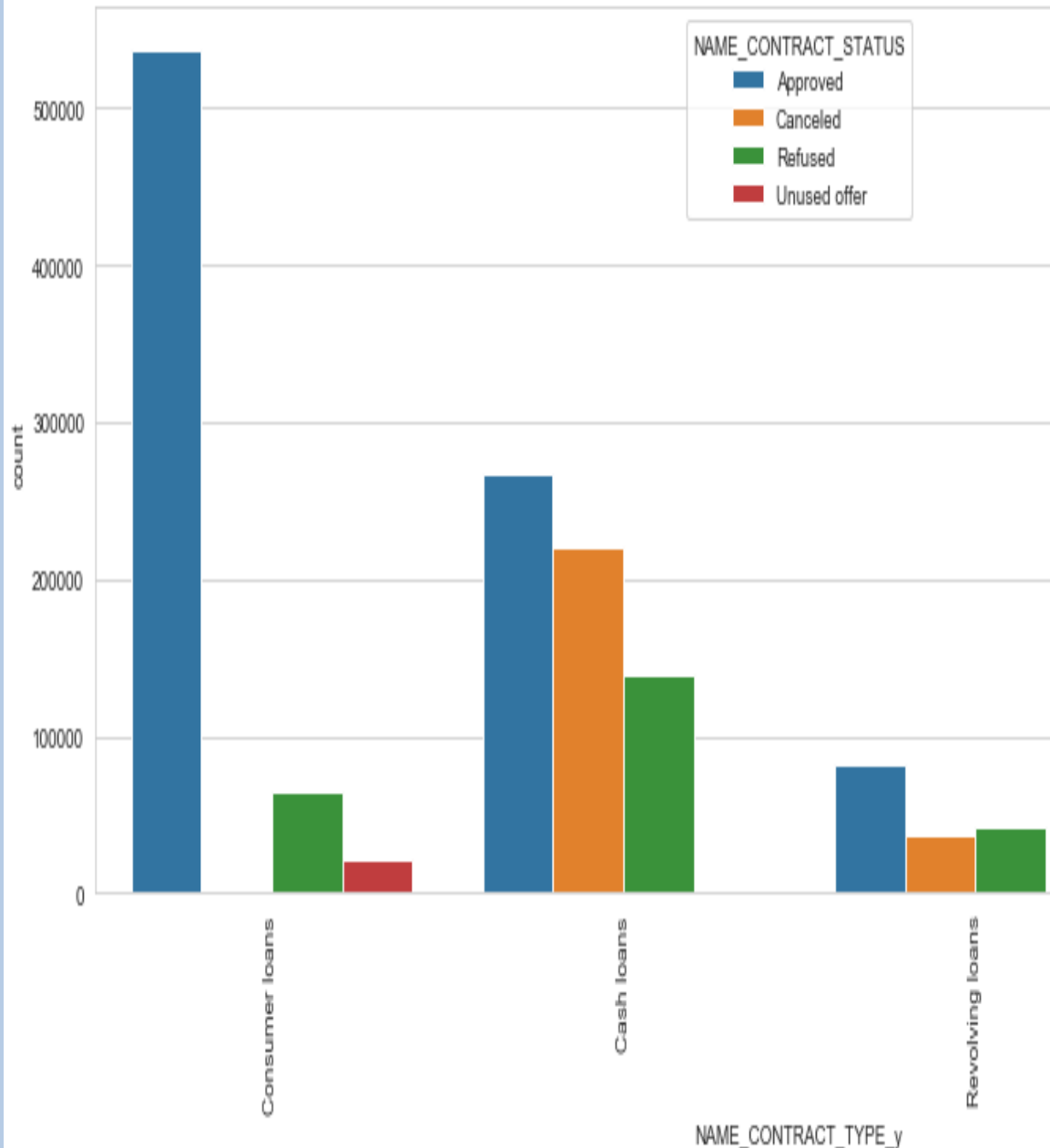
Points to be concluded from the graph:

It is a categorical univariate graph for the previous application contract type with contact status.

- We observed that the highest approved contract status are form consumers loans.

- For the Cash loans, there are more clients with cancelled loans and refused loans.

- Consumer loans are the only one who provides the unused offers.

- The approved ,cancelled and refused are less count in the contract status of revolving loans

CONTINUOUS UNIVARIATE ANALYSIS!

Points to be concluded form the graph:

It is a continuous univariate graph for the previous credit at application of the previous application.

- We observed that the term of previous credit at application of the previous application frequencies are higher at 0 to 20.

- As the term of previous credit at application of the previous application rate increases the frequency of the clients are automatically less.

CONTINUOUS BIVARIATE ANALYSIS WITH NAME_CONTRACT_STATUS!
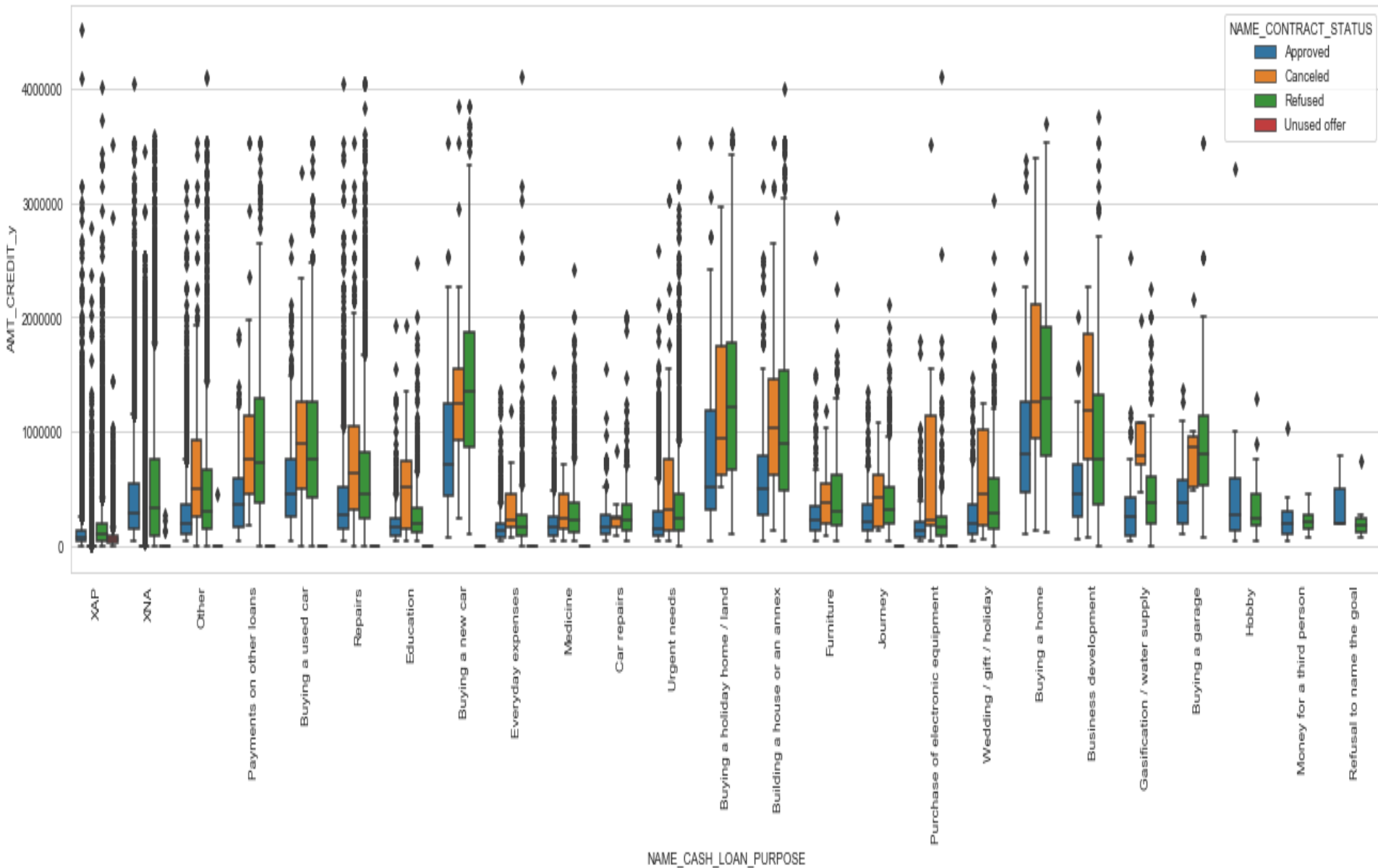
Points to be concluded form the above graph:

It is a amount credit in the previous loan vs. Name cash loan purpose with name contact status.

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

- The application contract status of "cancelled" are high for the loan purpose of buying a home and business development.

- The cash loan purpose of Money for third person or a Hobby is having less credits applied.

- The application contract status of "Refused " are high for the loan purpose of buying a home and buying a new car.

- The name contract status of unused offer are very less amount of credit in the previous application.

RECOMMENDATION!

- Bank should focus more into **Revolving loans** type.
- For the client who **own a car and do not own a flat** are facing less payment issues.
- More chances of payment difficulties for unaccompanied type clients.
- **States servant's** are **good** to provide loans.
- Occupation type of **labours** and **sales staff** are not recommended for loans where as **Accountants** are recommended.
- Provide loan offers for **high Income range** along with **high credit amount** clients are having less payment difficulties .
- Prefer loans for **males over female** clients for the working clients.
- Recommended family status of **'civil marriage', 'marriage'** and **'separated'** of **Academic degree education** are having higher number of credits.
- Prefer the **cash through bank process** for more approved applications.
- Look up for more loan offers for **repeater client** as it takes less decisions time.

CONCLUSION!

❖ Banks should focus more client who **own a car and do not own a flat** are more of successful payments.

❖ Banks should focus less on income type **Working** as they are having most number are **facing payment difficulties**.

❖ Occupation type of **labours** and **sales staff** are not recommended for loans where as **Accountants** are recommended.

❖ Focus on **States servants** are **good** to provide loans.

❖ Get as much as clients from housing type **With parents** as they are having less number of payment difficulties.

THANK YOU..! ☺