

The background of the slide features a collage of financial data visualizations. At the top, a bar chart shows values for the years 2001 through 2006, with a y-axis ranging from 0 to 1,000,000. Below this, a line graph tracks 'Total assets' (yellow bars), 'Total debt' (green bars), and 'Net income' (red bars) from 2001 to 2006, with a y-axis from 0 to 8,000,000. A silver pen with a black grip is positioned diagonally across the center, its tip pointing towards the line graph. On the far left, a portion of a pie chart with yellow, green, and orange segments is visible. The entire scene is set against a dark blue background with a subtle grid.

# Risk Analysis on Bank Credits

BY – P Anand Rao

# Introduction

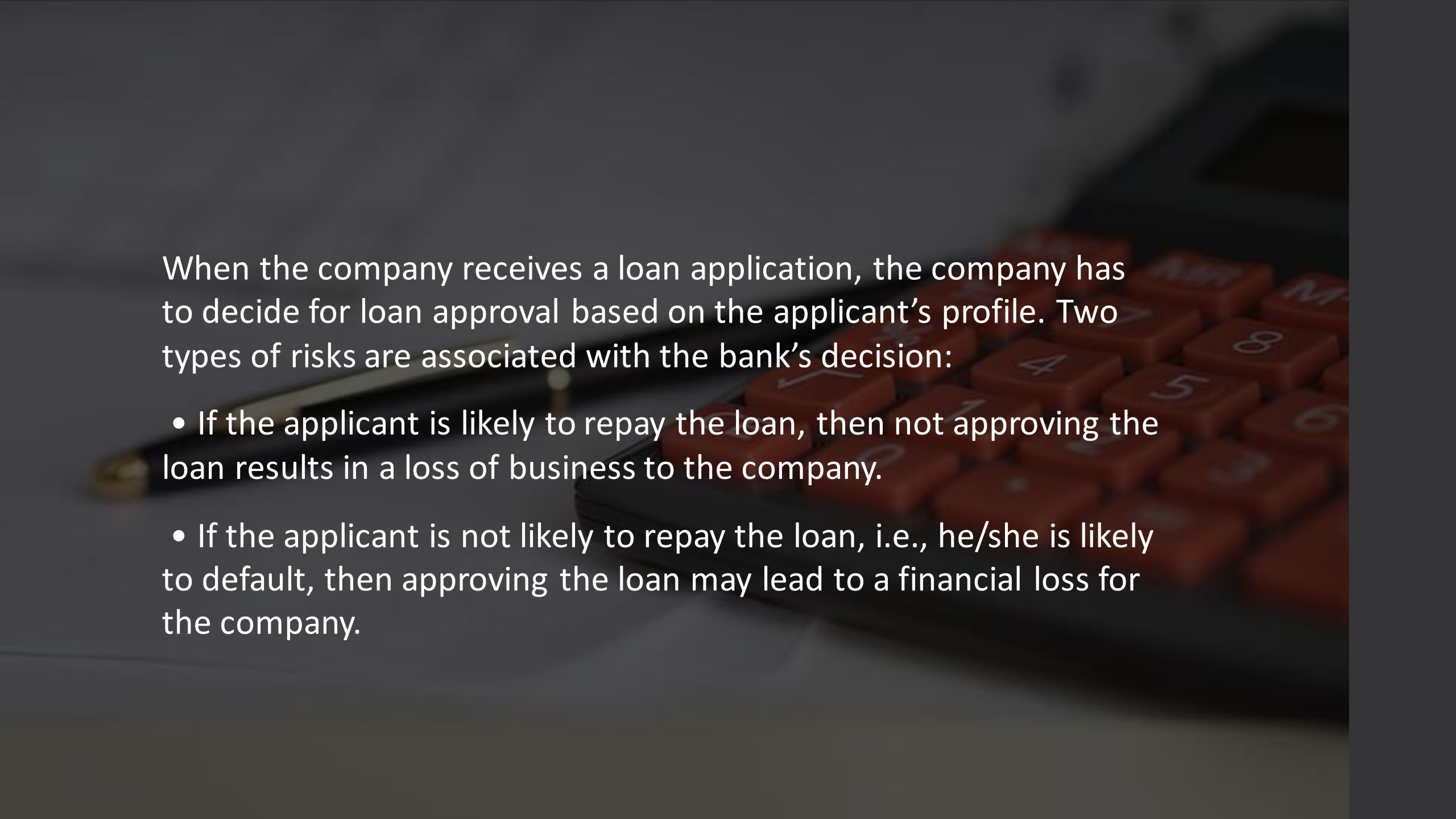
When the company receives a loan application, the company has to decide on loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business for the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

# Business Understanding - 1

The loan-providing companies find it hard to give loans to people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Suppose you work for a consumer finance company specializing in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.





When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.



# Business Understanding – 2

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- All other cases: All other cases when the payment is paid on time.



When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

1. **Approved:** The Company has approved loan Application.
2. **Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
3. **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.
4. **Unused offer:** Loan has been cancelled by the client but on different stages of the process. In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

# Business Objectives

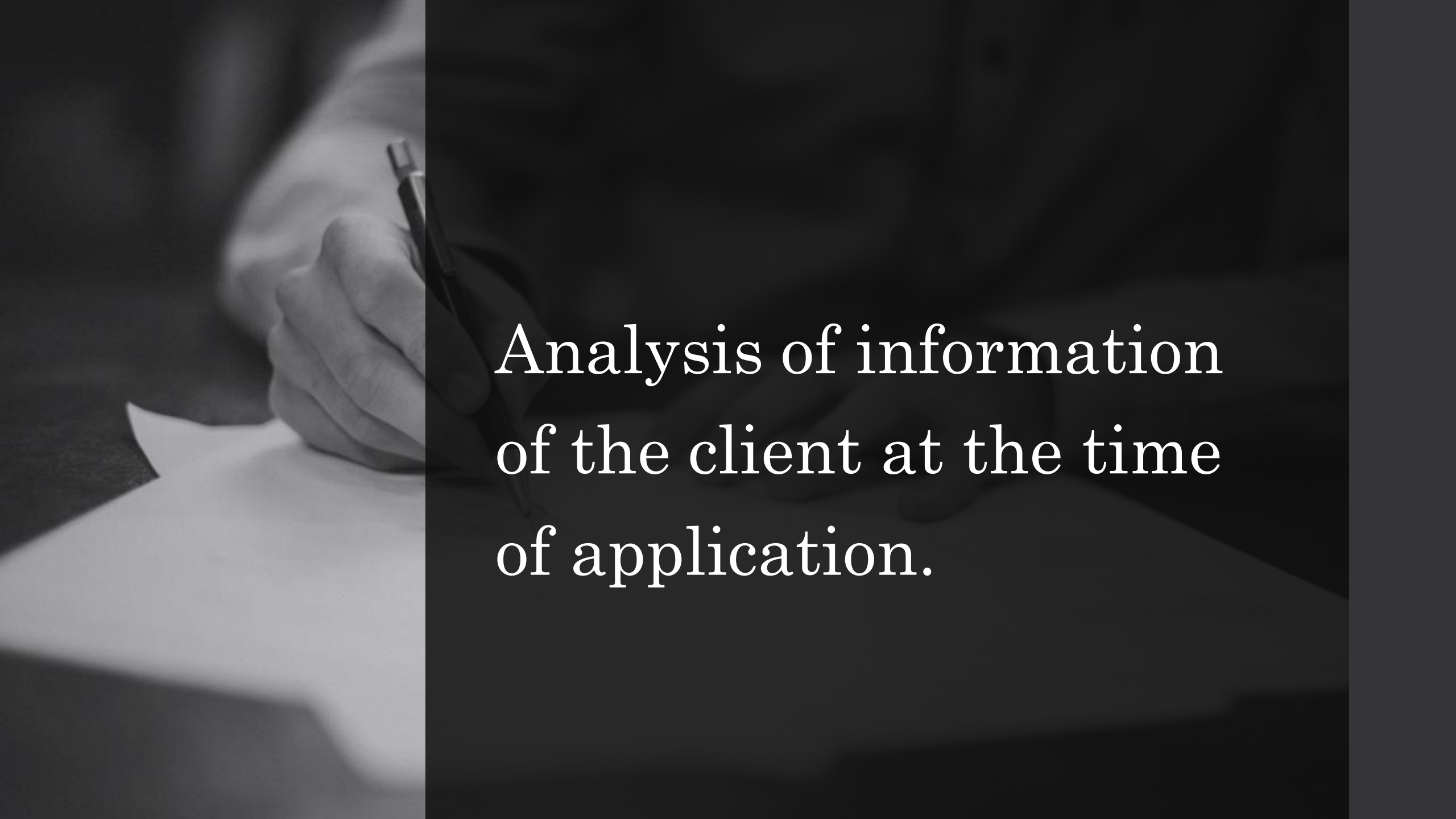
- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (too risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.
- To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough)

# Data Understanding

This dataset has 3 files as explained below:

- 1. 'application\_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
- 2. 'previous\_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
- 3. 'columns\_description.csv' is data dictionary which describes the meaning of the variables.



A black and white photograph of a hand holding a pen, writing on a document. The image is partially obscured by a dark vertical bar on the right side, which contains the text.

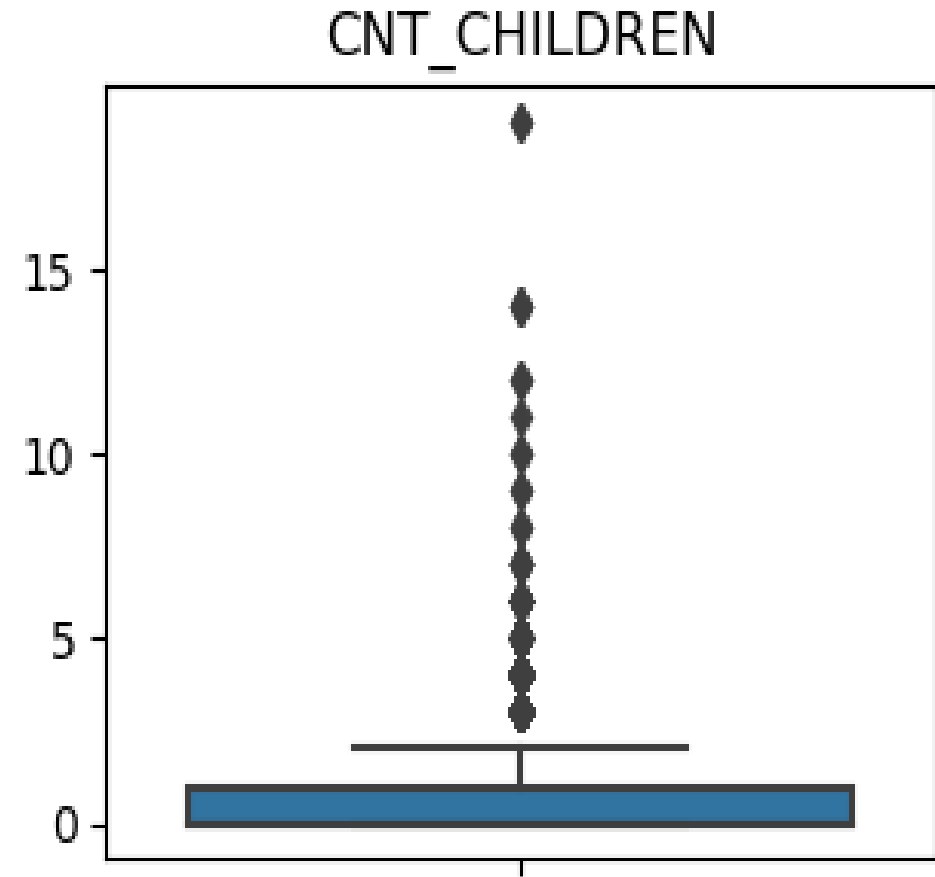
Analysis of information  
of the client at the time  
of application.

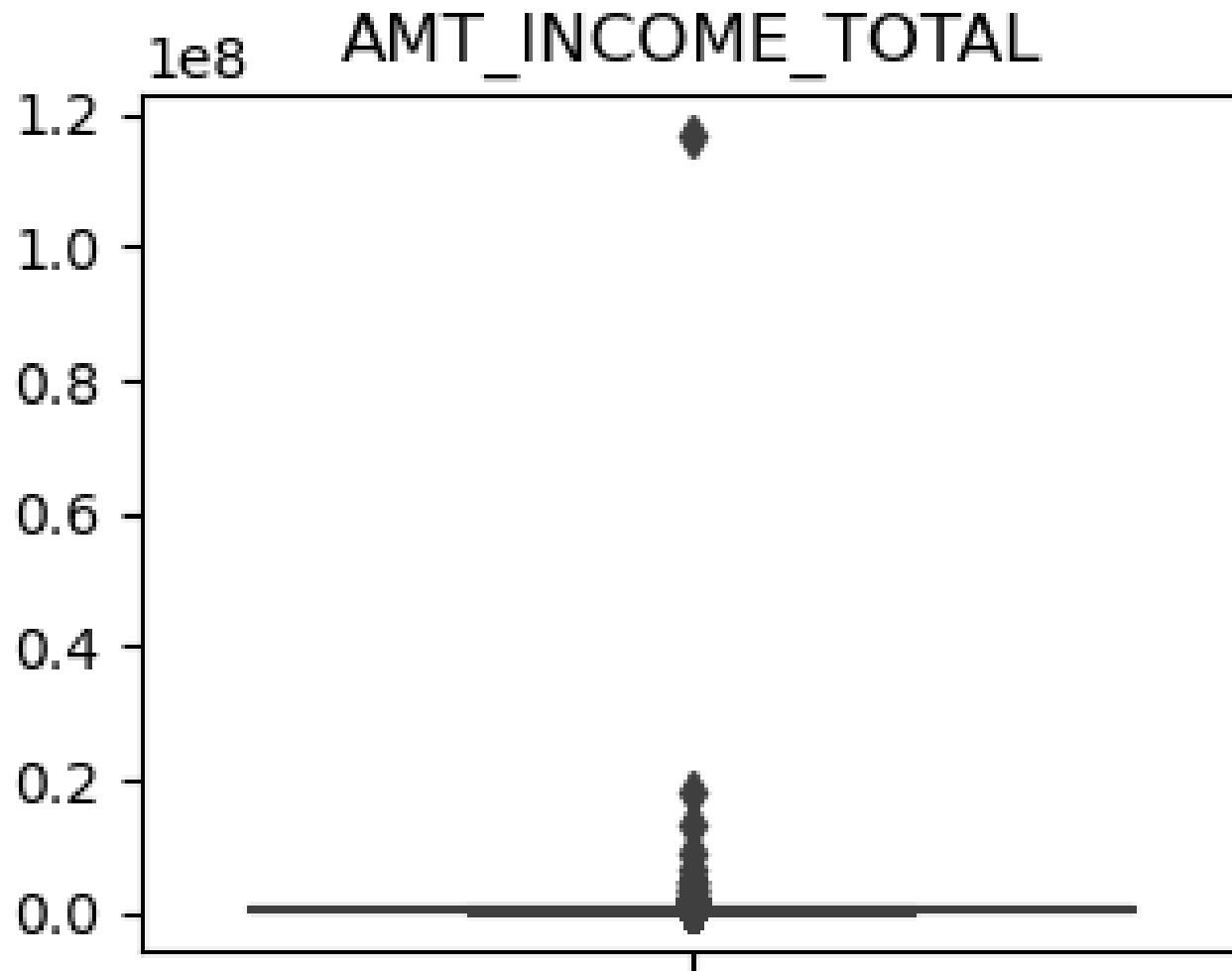


Outlier  
Analysis of  
Analysis  
Dataset

# Analysis of 'CNT\_CHILDREN'

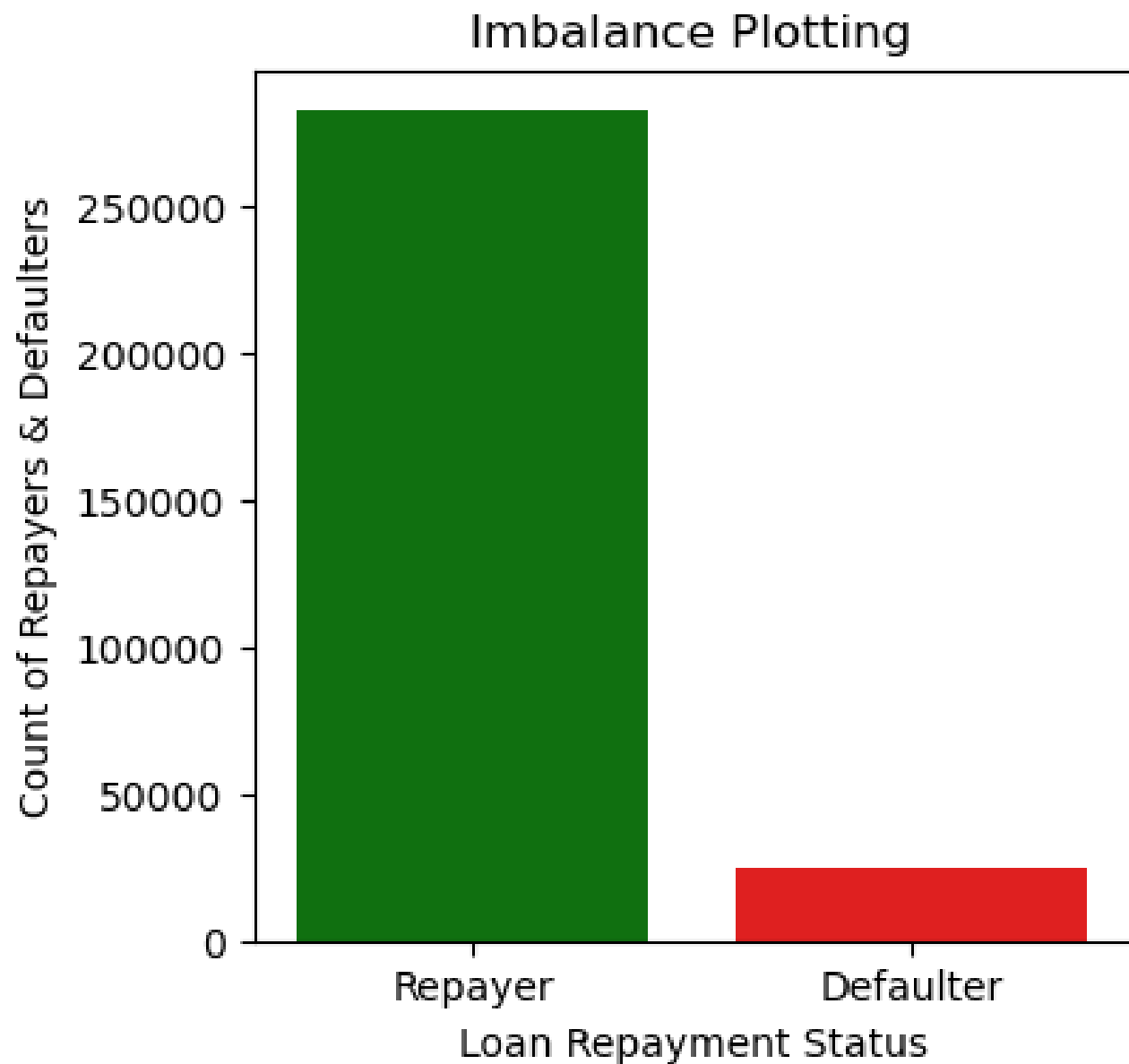
- Looking at the data, we can see that above 7 children, the count of applicants are very minimal (2 or 3 in each category)
- Also, looking at the observation data for applicants with 10 children, the applicants are only 31 and 41 years respectively. This seems a one-off scenario and can be treated as an outlier
- Both distplot and boxplots clearly show the values above 2.5 as being outliers
- Applicants with 3 or more children are outlier cases





## Analysis of AMT INCOME TOTAL

Applicants with  
Income above 900K  
(99.9% value) are  
outlier



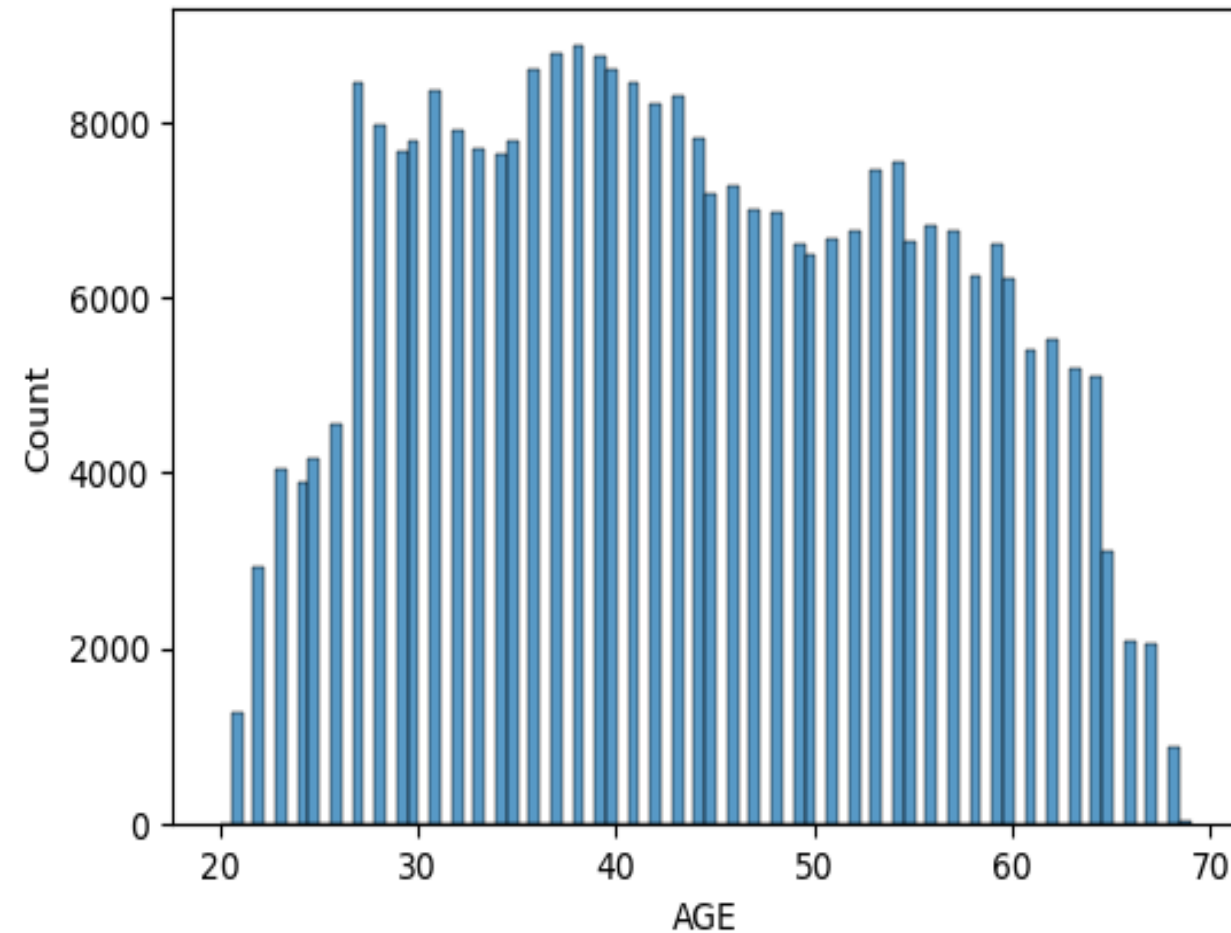
## Analysis of Imbalance Plotting

No of defaulters are more than 7 times less than repayer.

# **Univariate Analysis**

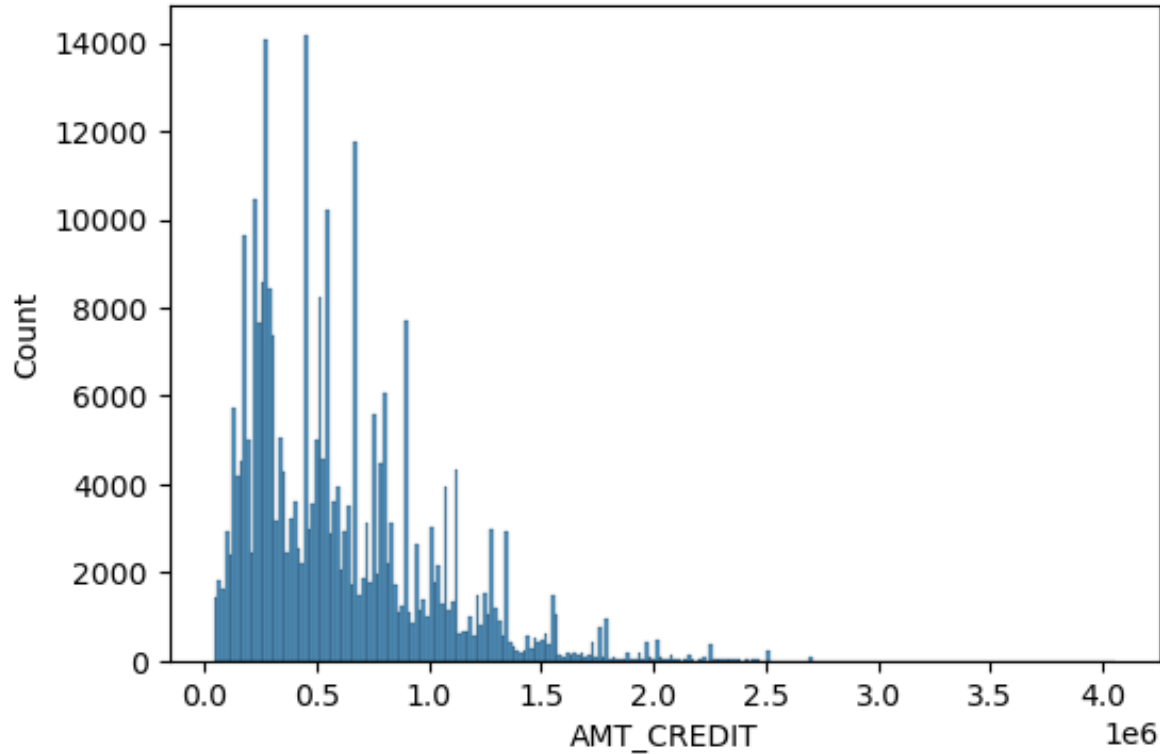


# Analysis of AGE column



- Most of the applicants age lies between 28 to 45.

# Analysis of AMT\_CREDIT

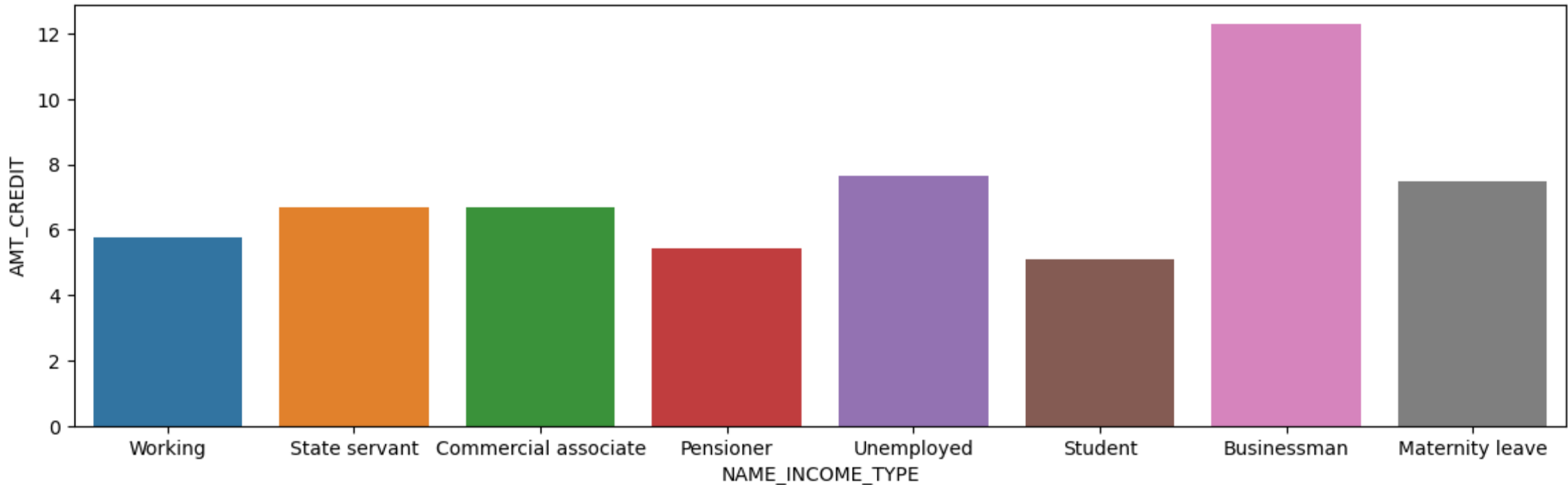


- Over 1M amount of loan taken by only 16% applicants and all of them are Businessman

# Bivariate Analysis

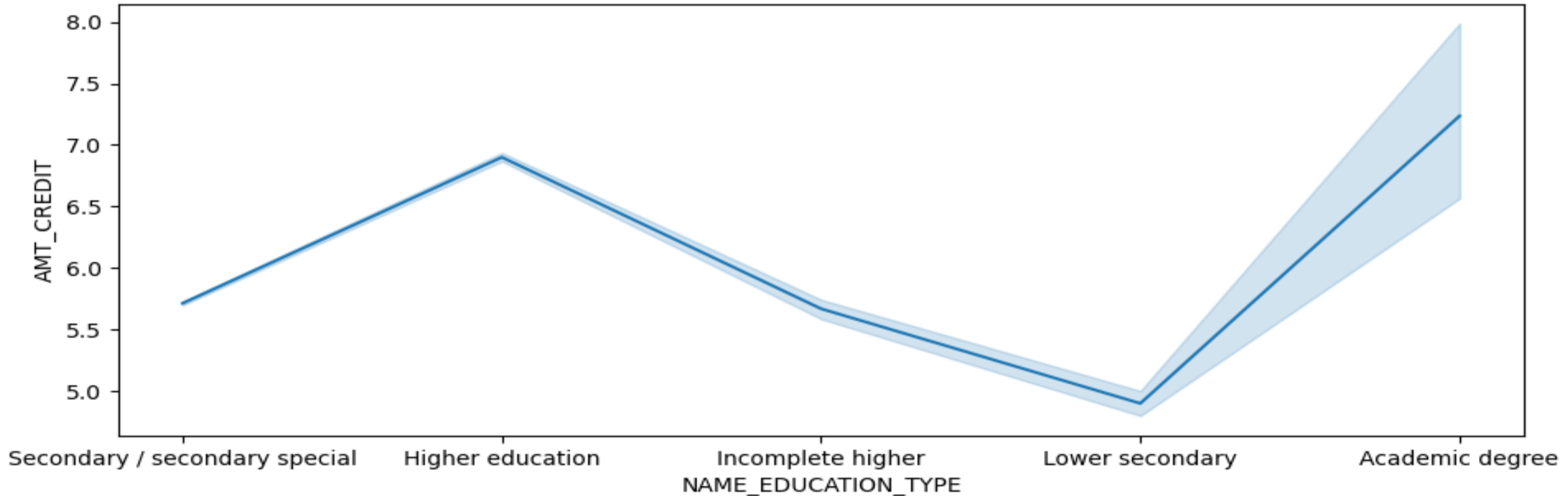


# Analysis of 'NAME\_INCOME\_TYPE' Vs 'AMT\_CREDIT'



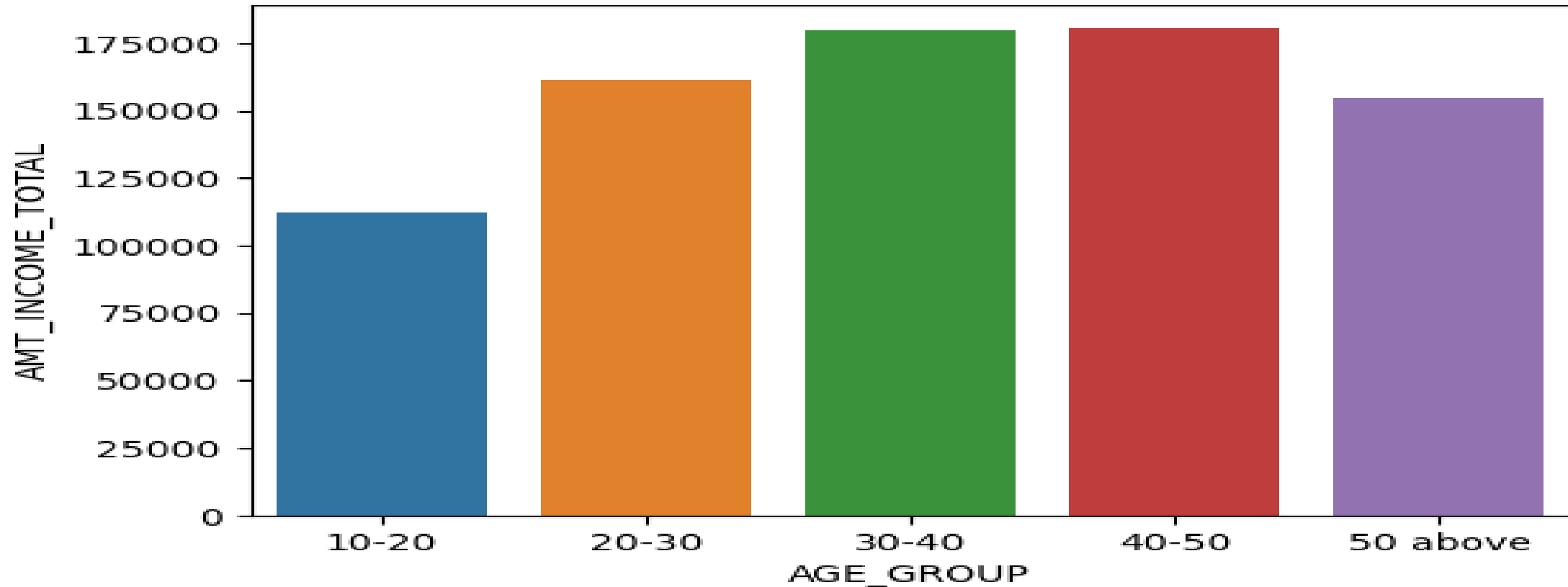
- Businessman take loan 2 times more than than other Income Type

# Analysis of NAME EDUCATION\_TYPE Vs AMT\_CREDIT



- It clearly state that Higher Education and Academic Degree applicants lending loan more than the rest of all.

# Analysis of AGE GROUP Vs AMT INCOME TOTAL



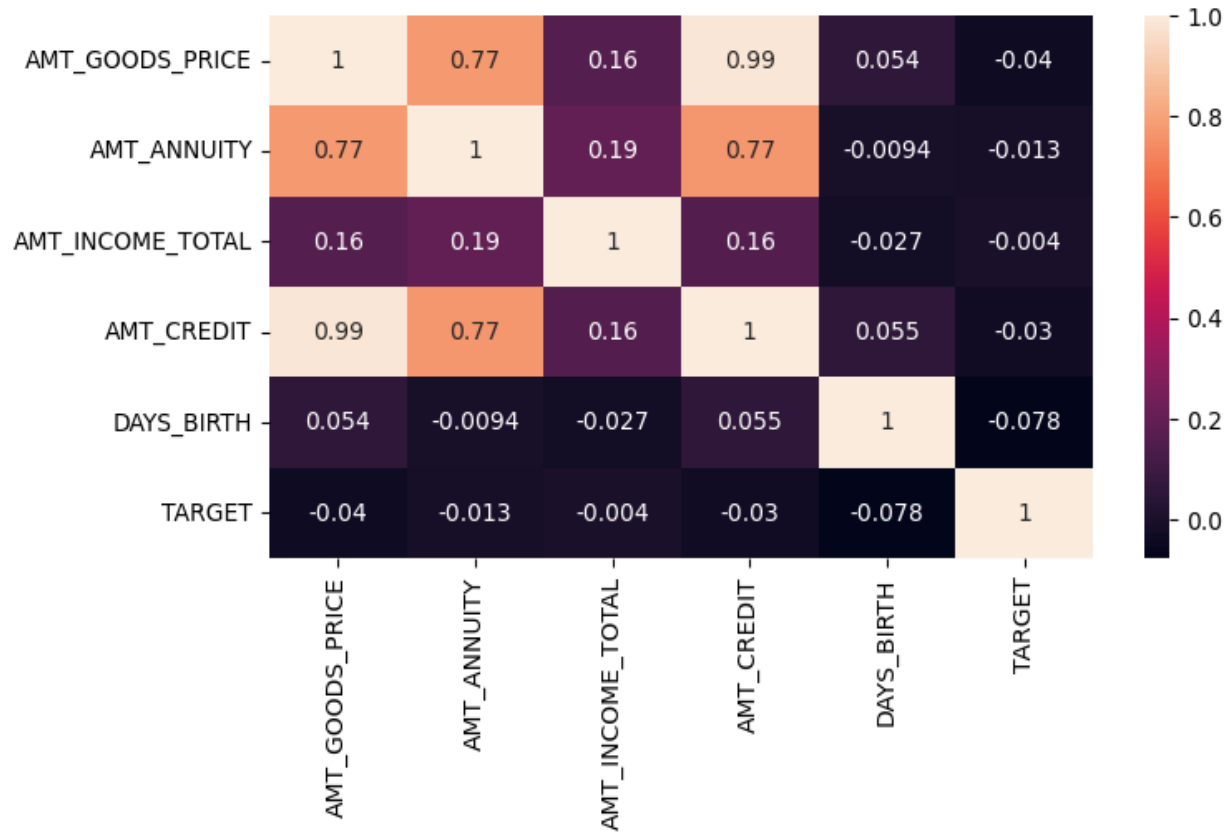
- Applicants age between 30 - 50 having total Income more than 1,75,000



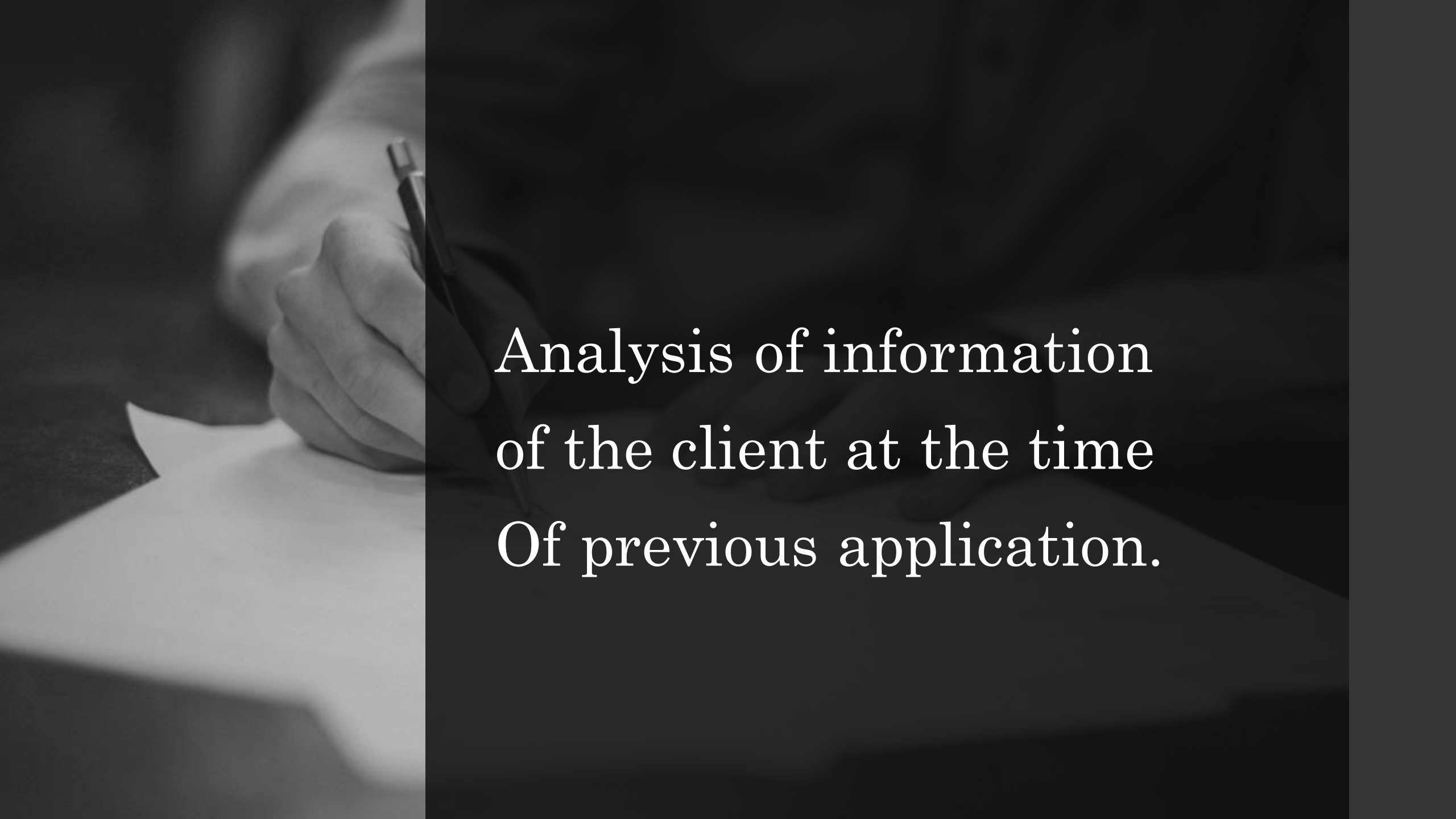
# Multivariate Analysis

A magnifying glass is positioned over a line graph. The graph features several data series represented by different colored lines: red, blue, green, and pink. The lines show various trends, with some peaking and others dipping. Numerical values are printed near the lines, such as 0.86, 0.89, 1.32, 1.36, 1.29, and 1.48. The background of the graph is white with a grid of dashed lines. The magnifying glass has a silver handle and a clear lens that focuses on the central part of the graph.

# Analysis of Correlation matrix



- Credit amount is highly correlated with:-
- Amount of goods price - Loan annuity
- Total income With 0.77 of positive correlation loan annuity and amount of goods price is correlated.
- Total amount shows a negative correlation with defaulters of -0.004.



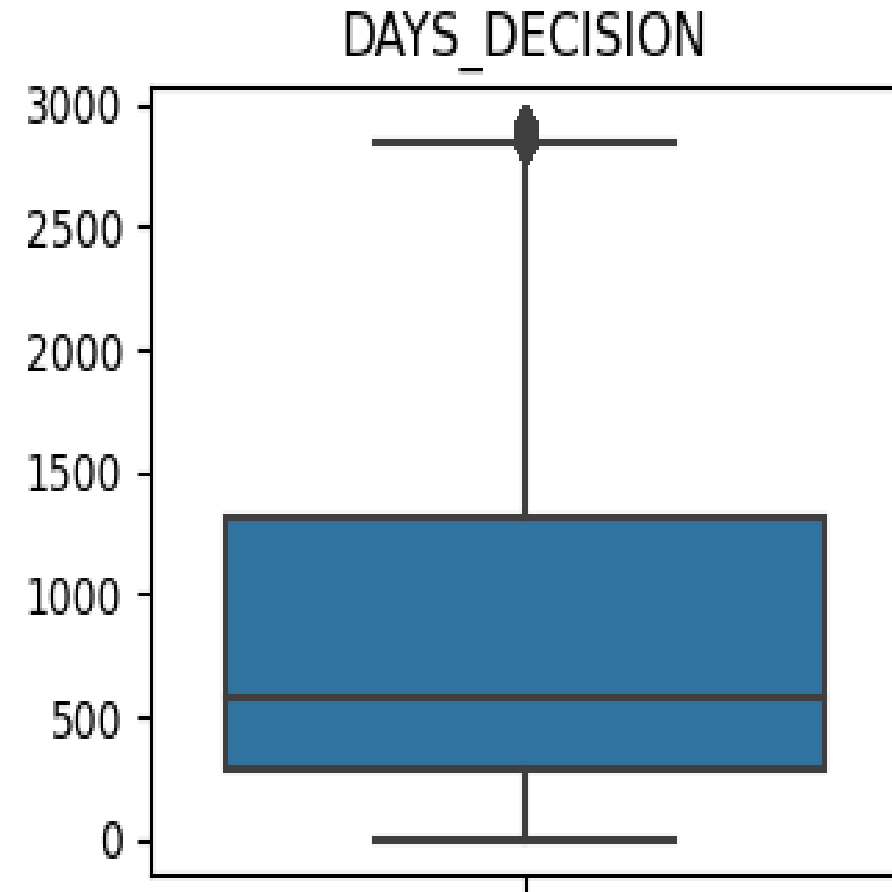
Analysis of information  
of the client at the time  
Of previous application.

A 3D bar chart with several bars of increasing height from left to right. The chart is rendered in a dark, monochromatic style. A large, semi-transparent dark gray rectangular overlay covers the right half of the image, serving as a background for the text.

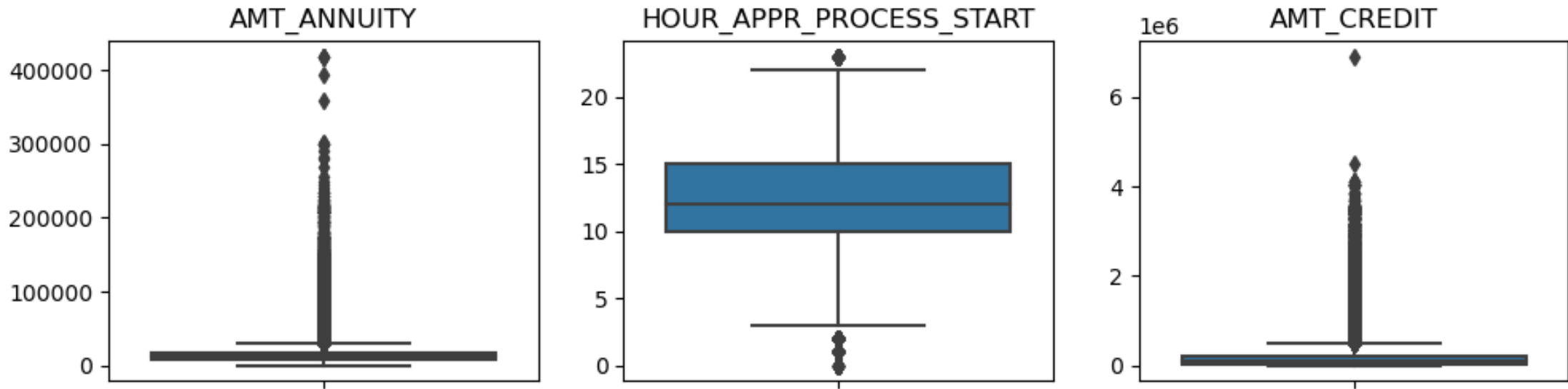
# Outlier Analysis of Previous Dataset

# Analysis of 'AMT\_GOODS\_PRICE'

- The average time taken by bank to process is 600 days.
- This Graph shows very less amount of Outliers which is between the range of 2800 to 3000.



# Analysis of AMT ANNUITY, HOUR\_APPR\_PROCESS\_START & AMT\_CREDIT



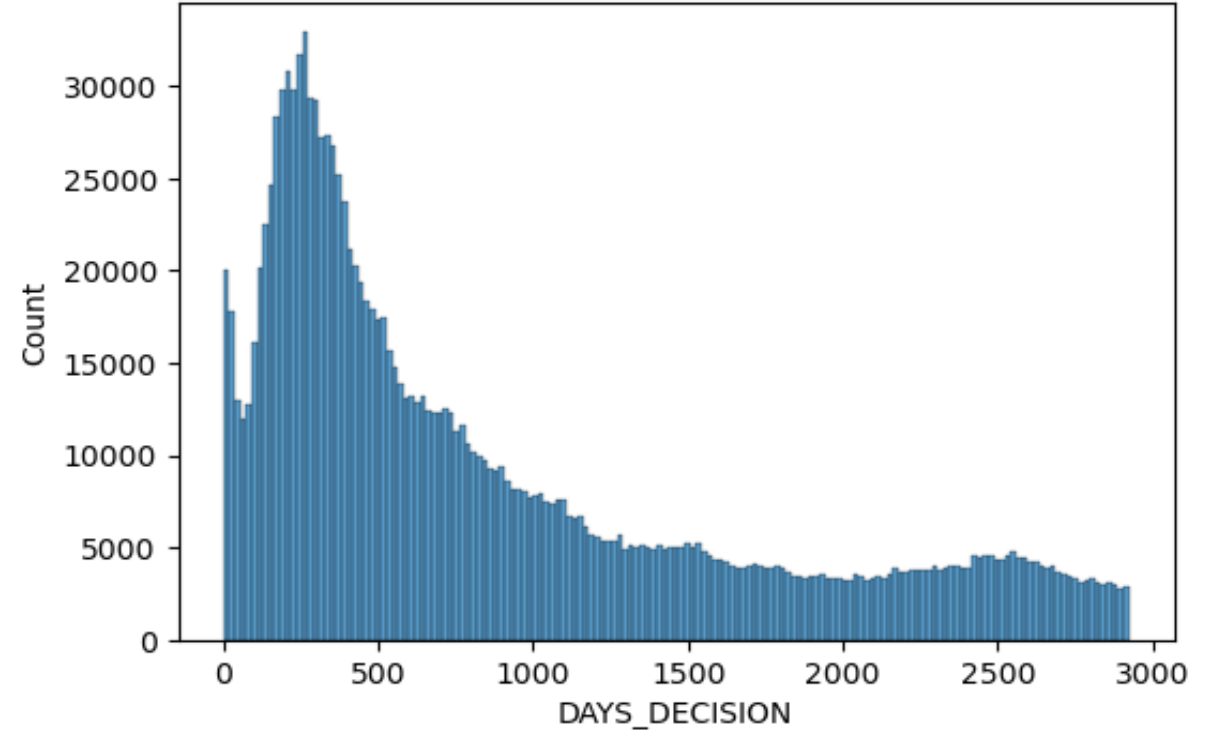
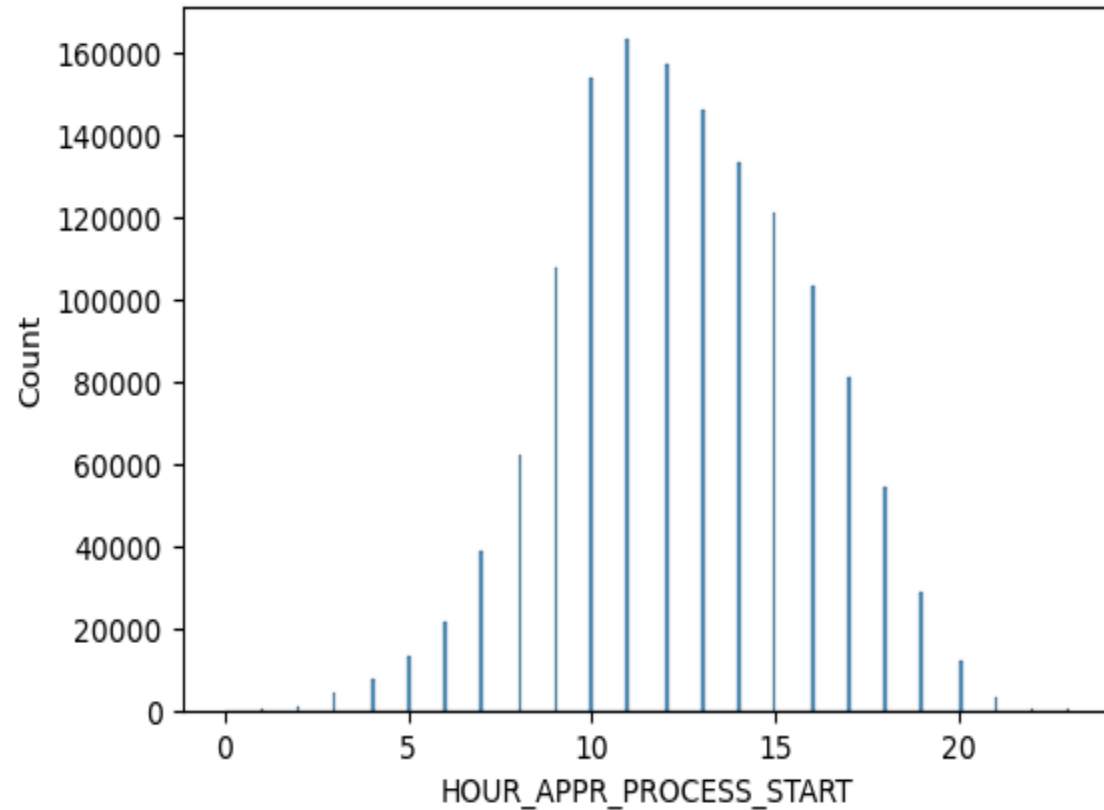
- The average amount credit is  $< 1$
- The amount annuity is  $> 4$  lakh





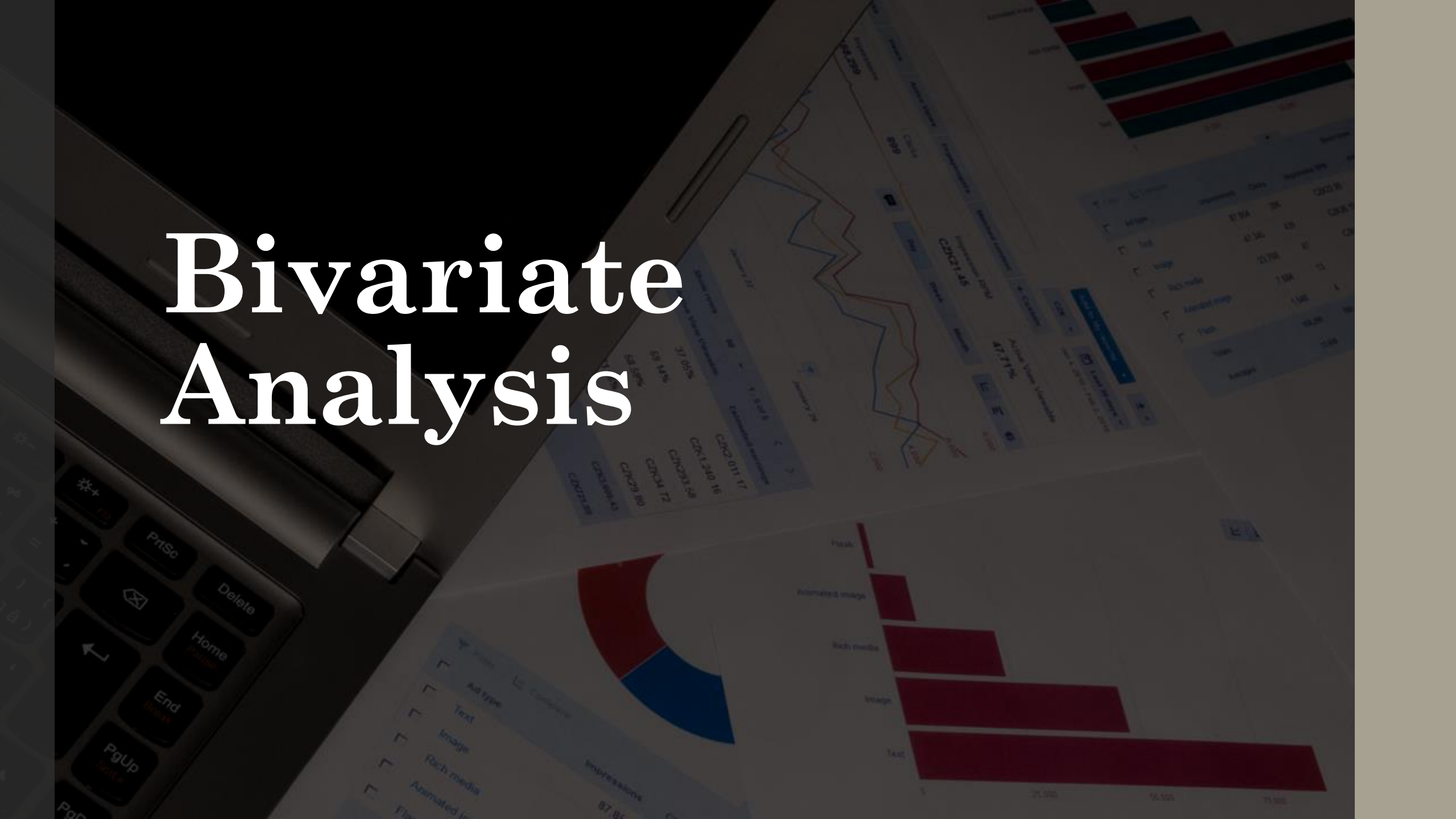
# Univariate Analysis

# Analysis of 'HOUR\_APPR\_PROCESS\_START' & 'DAYS\_DECISION'

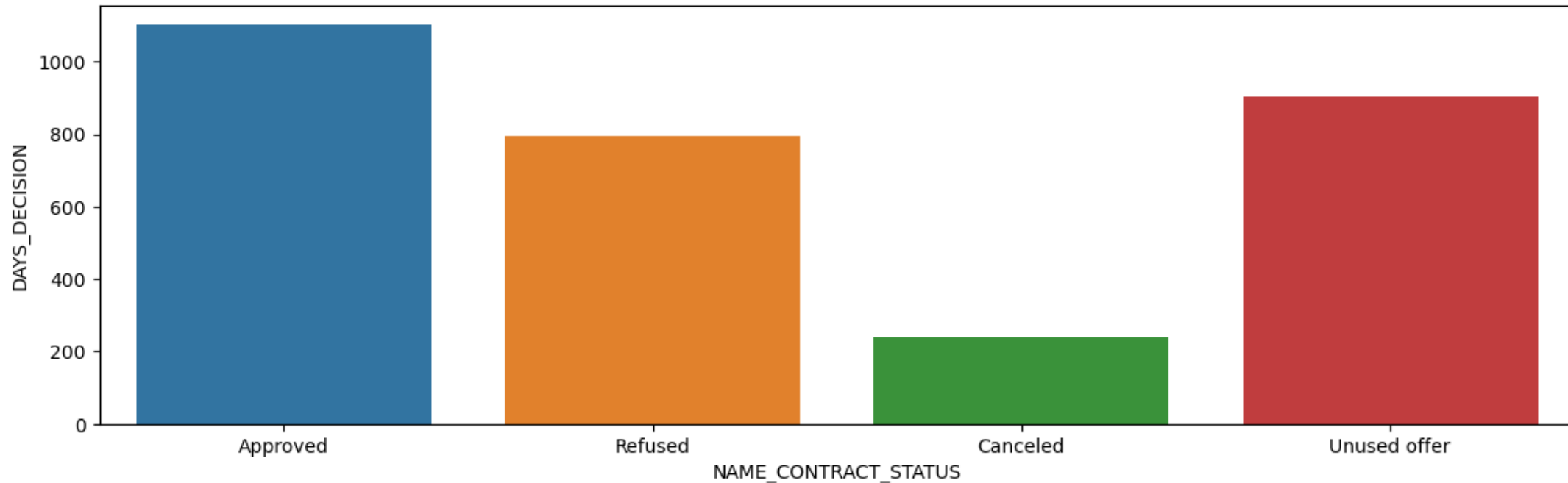


- The average application processing starts between the range of 10:00 A.M to 15:00 P.M.
- Most of the previous data decision were taken between 250 to 700 days.

# Bivariate Analysis



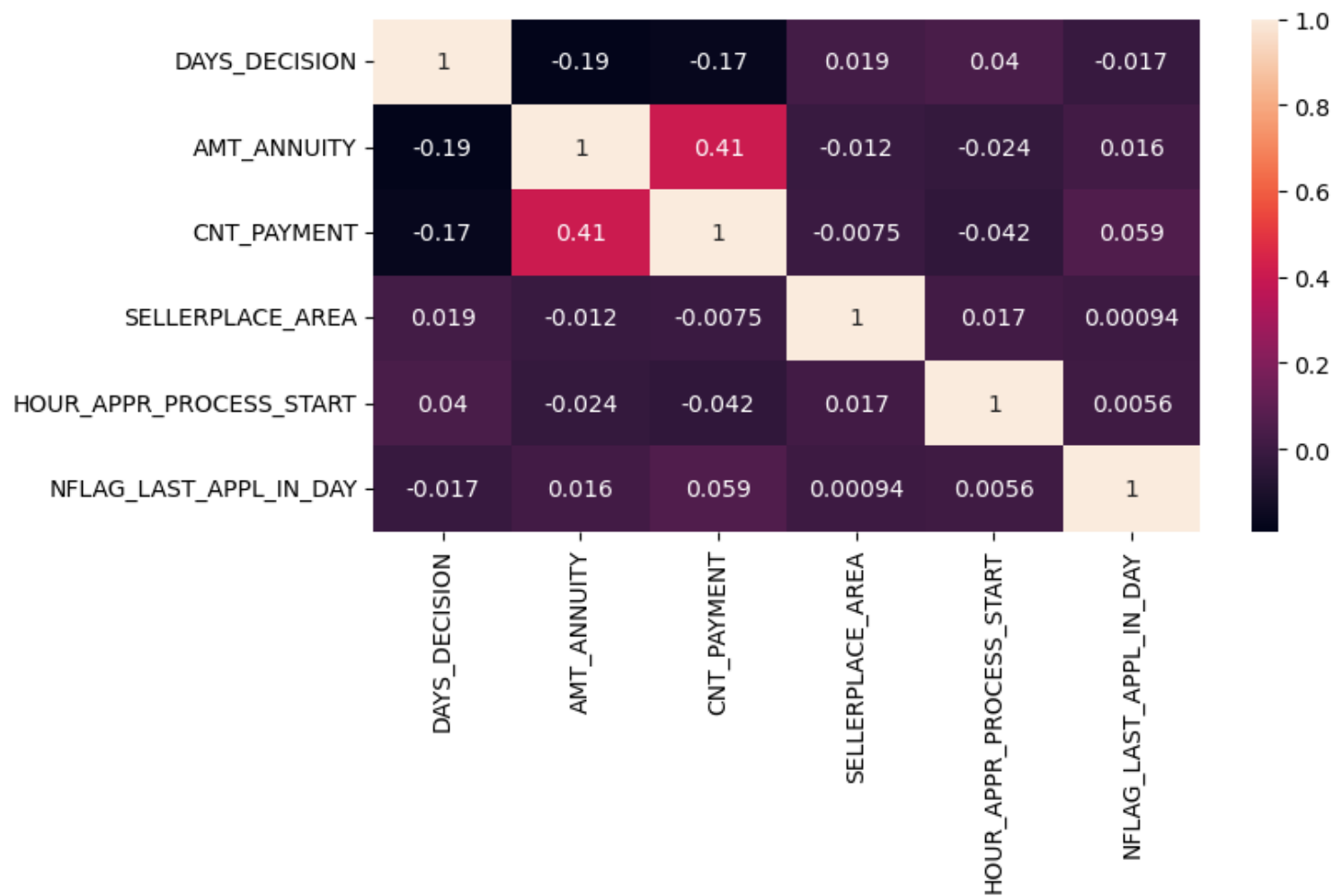
# Analysis of NAME CONTRACT STATUS & DAYS DECISION



- Application take more than 1000 days to get approved, from the previous application
- Application get refused in less than 800 days
- Application cancelled in less 300 days



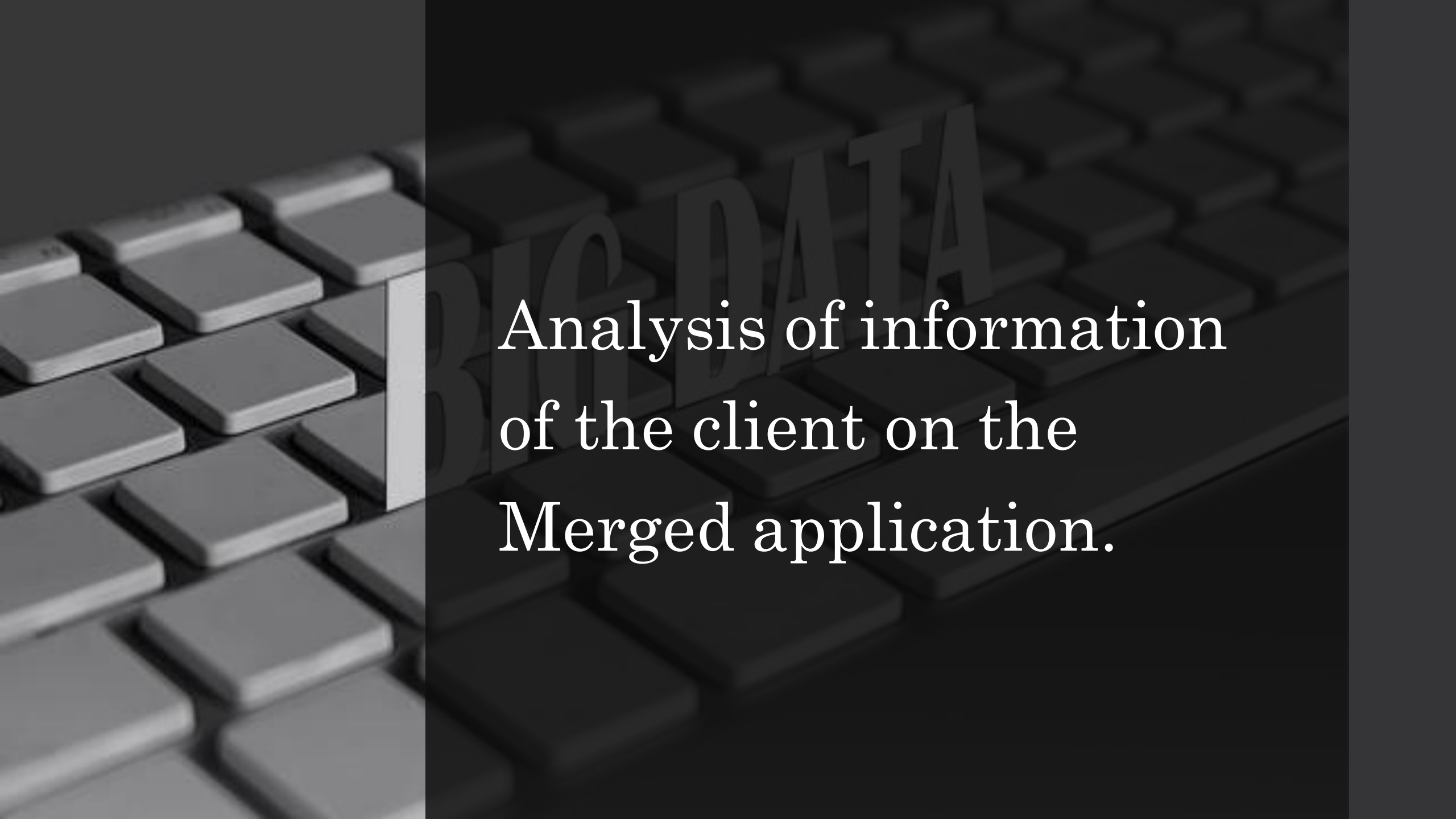
# Multivariate Analysis



## Analysis of Correlation matrix

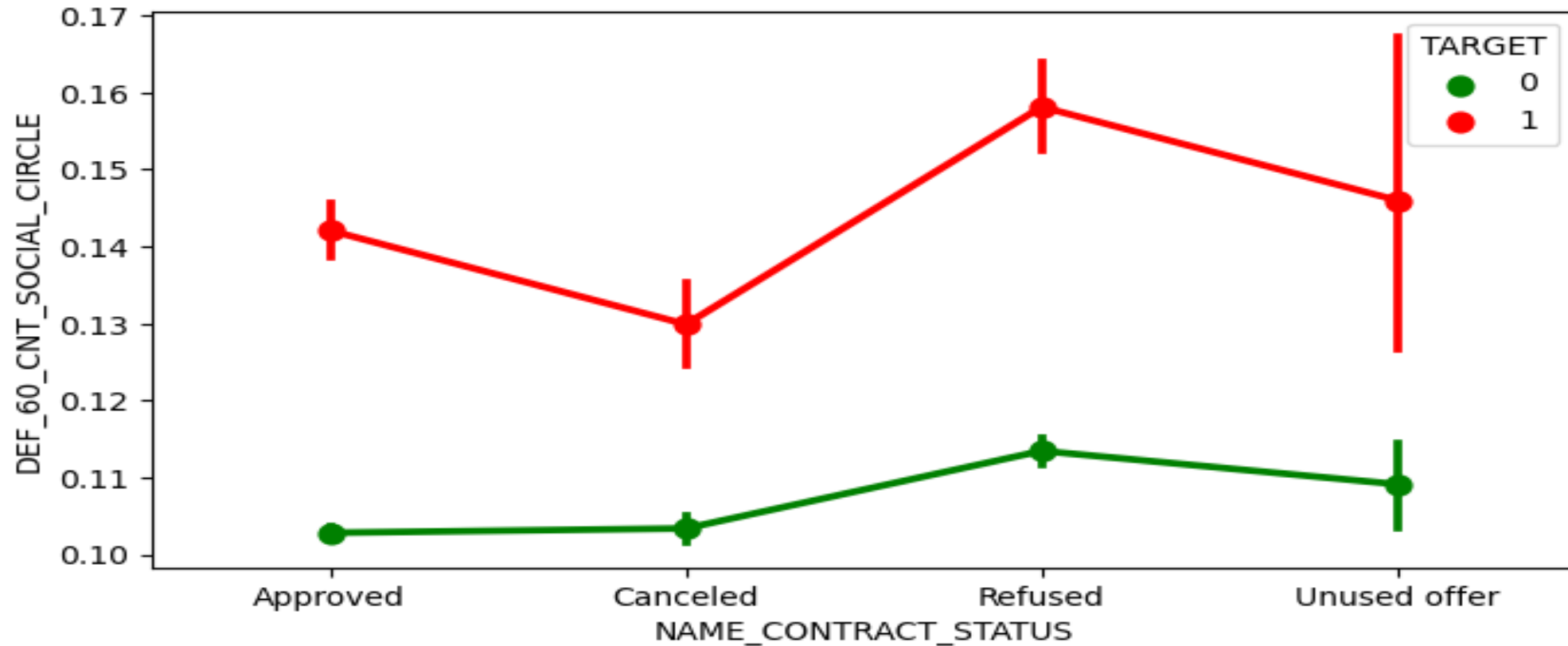
Amount Annuity is (0.41) positively correlated with Count Payment.





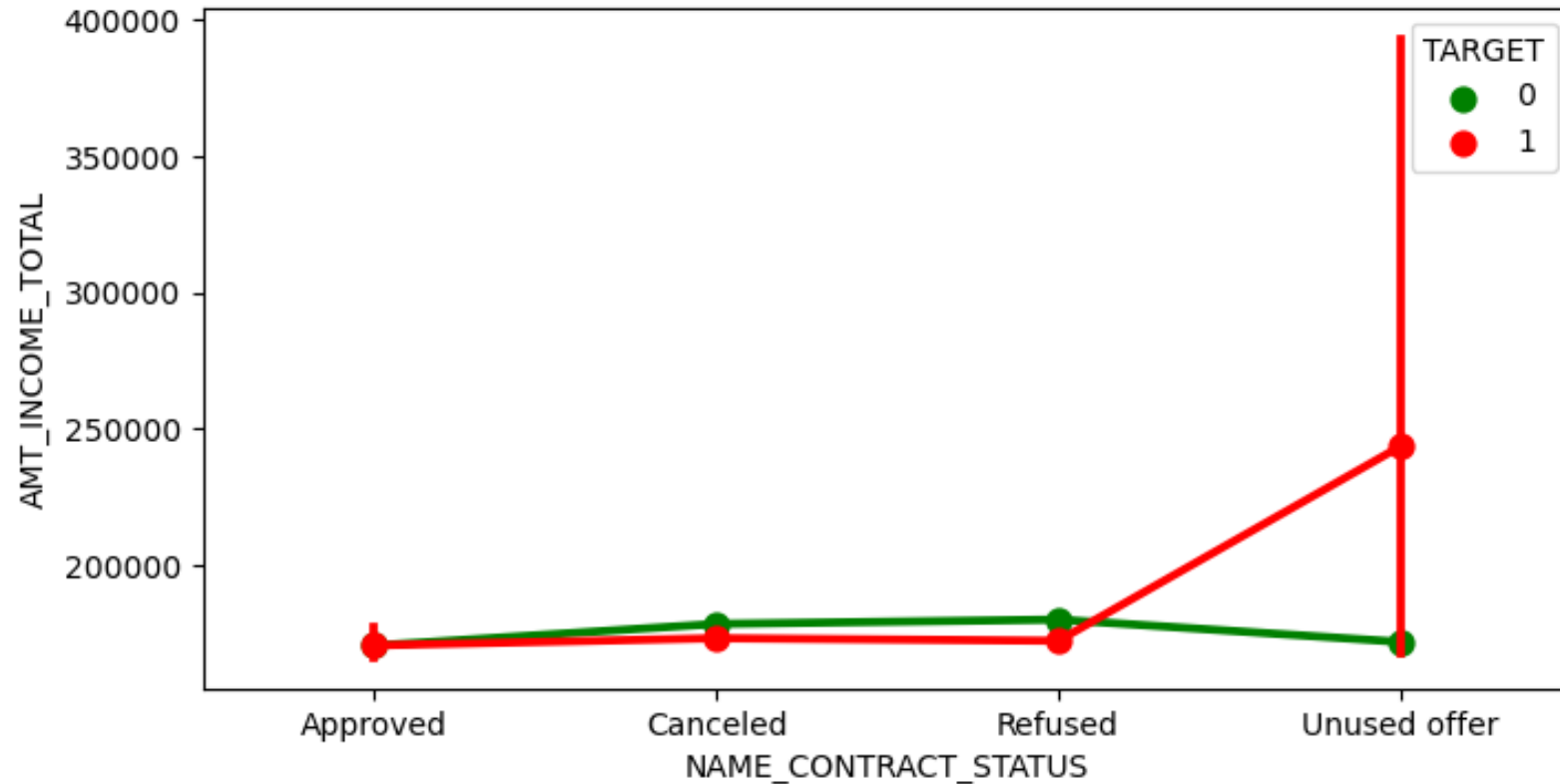
Analysis of information  
of the client on the  
Merged application.

# Analysis of NAME CONTRACT STATUS AND DEF 60 CNT SOCIAL CIRCLE



- The applicants having DEF\_60\_CNT\_SOCIAL\_CIRCLE more than average of 0.13 are tends to defaults more. So, analyzing social circle will help in disbursement of loan.

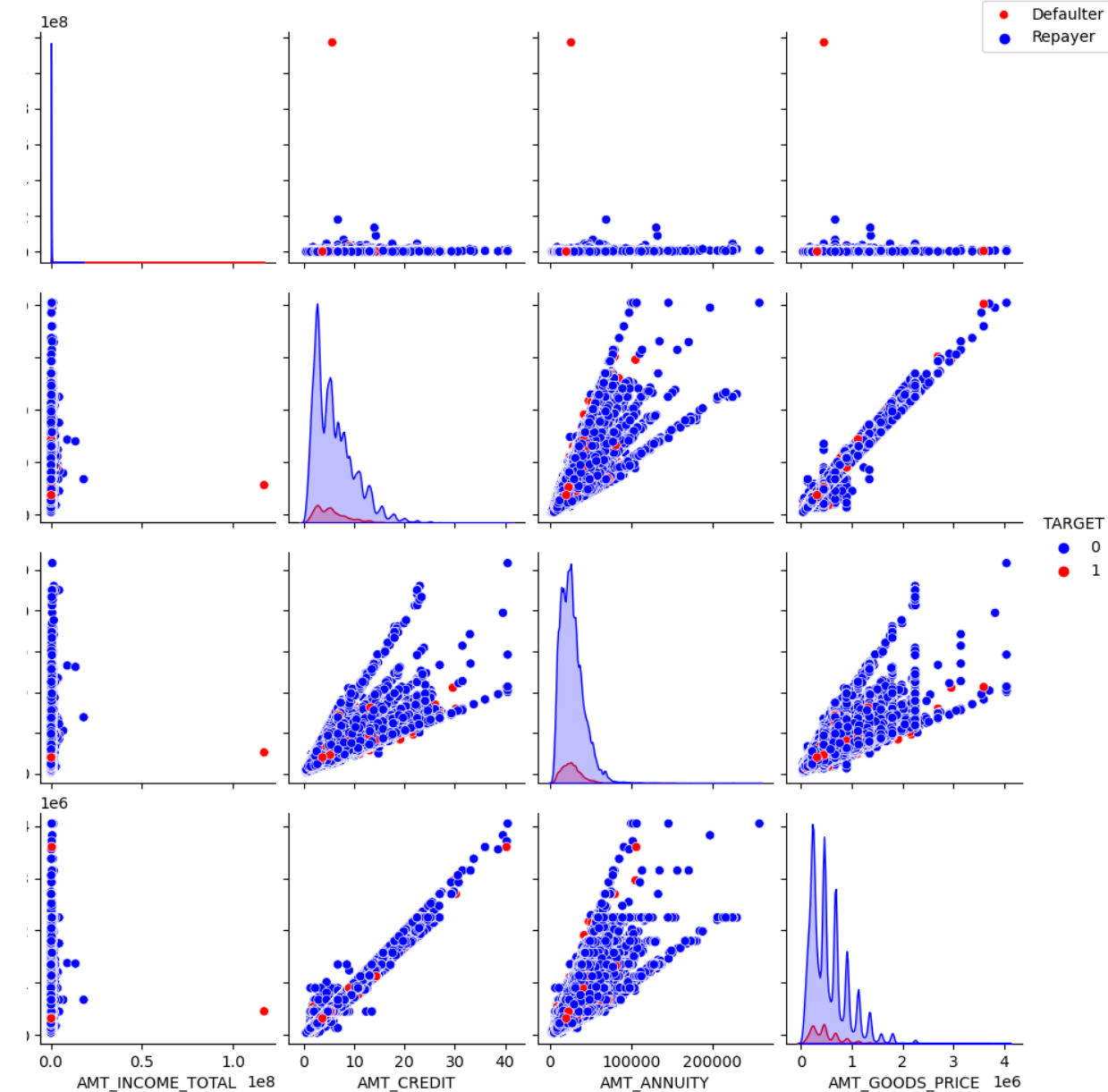
# Analysis of NAME CONTRACT STATUS AND AMT INCOME TOTAL



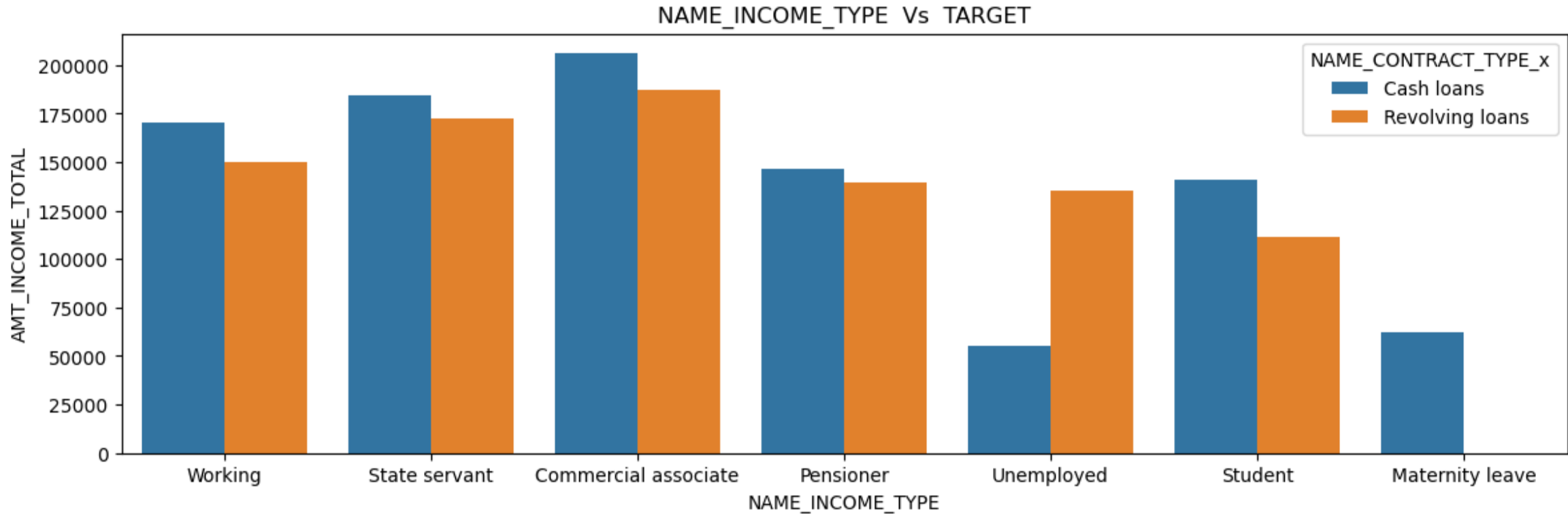
- The line plot shows that people who has total income more than the average get defaulted because they didn't used the offer.

# Analysis of Merged Column

- When Annuity Amount > 15K and Good Price Amount > 20 Lakhs, there is a lesser chance of defaulters
- Loan Amount(AMT\_CREDIT) and Goods price(AMT\_GOODS\_PRICE) are highly correlated as based on the scatterplot where most of the data are consolidated in form of a line
- There are very less defaulters for AMT\_CREDIT > 20 Lakhs

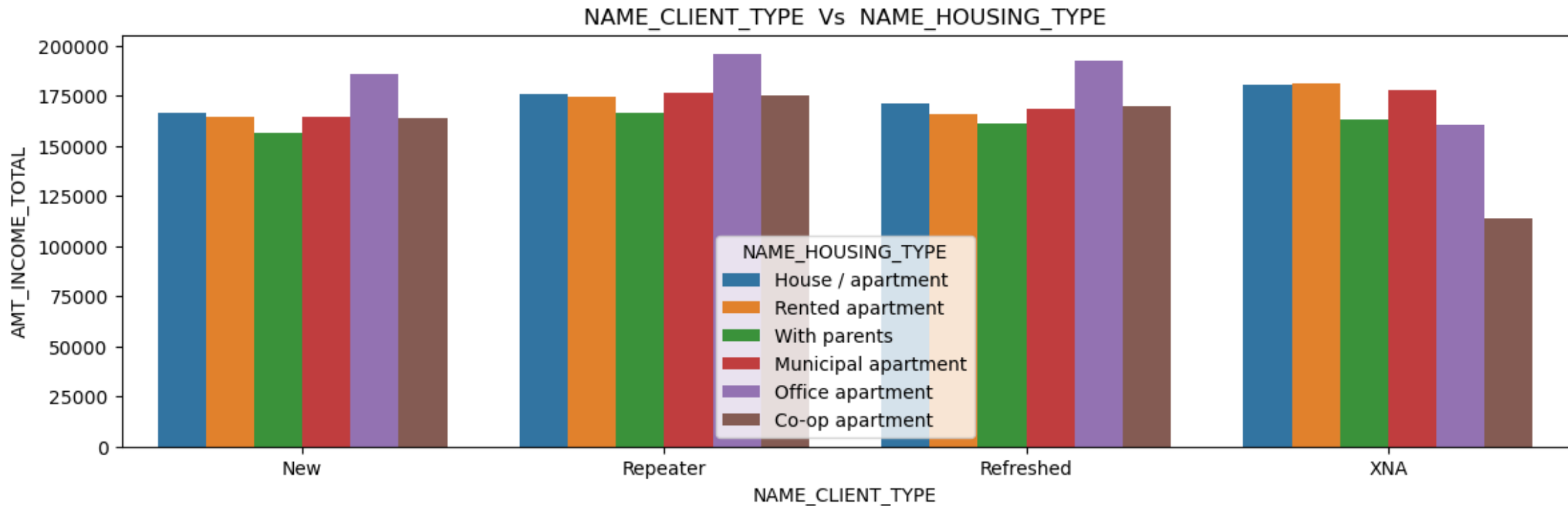


# Analysis of NAME INCOME TYPE AND AMT INCOME TOTAL



- Graph shows that most of the people prefer to take cash loans but more over the unemployed applicants get loan as revolving.

# Analysis of NAME\_CLIENT\_TYPE AND AMT\_INCOME\_TOTAL



- It state that most the applicants who applied for loan are from office apartment.
- And rest of all are slightly closer to each other in term of total income.
- Number of Repeater are little bit higher than rest of all but the applicants refreshed ratio is also more than 95%. It means the new repeaters were less than 5%.

# Conclusion

- Banks must target more on contract type 'Student' , 'Pensioner' and 'Businessman' for profitable business and housing 'type other than 'Co-op apartment' for successful payments.
- Banks must focus less on income type 'Working' as it is has most number of unsuccessful payments in order to get rid of financial loss for the organization.
- Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.
- Banks has to pay more attention towards the days taken for approval, because it may cause decrease in loan applicants if days decision increases.



THANK YOU