

# CREDIT EDA ASSIGNMENT

Sheelu Singh



# CONTENTS

- PROJECT DESCRIPTION
- TASK
- APPROACH
- INSIGHTS AND RESULT
- CONCLUSION



# Content Description:

This assignment aims to give you an idea of applying EDA in a real business scenario. In this assignment, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.



# Task!

- Identify Missing Data and Deal with it
- Identify Outliers in the Dataset
- Analyse Data Imbalance
- Perform Univariate, Segmented Univariate, and Bivariate Analysis
- Identify Top Correlations



# Approach

- Data Acquisition
- Data Cleaning
- Data Analysis
- Data Visualization
- Project Report



# Insights and Result!

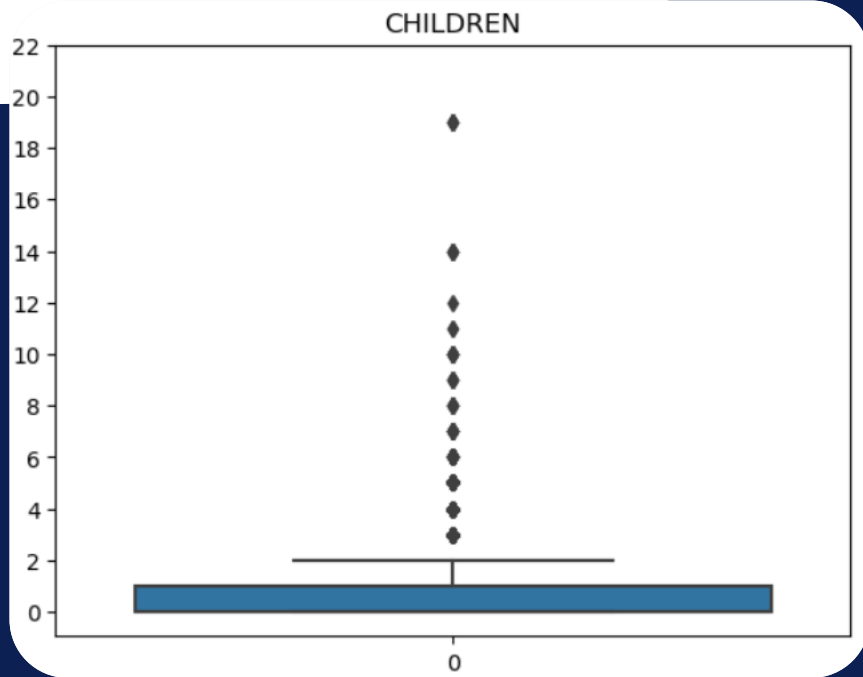


01

# Identifying outliers in Dataset



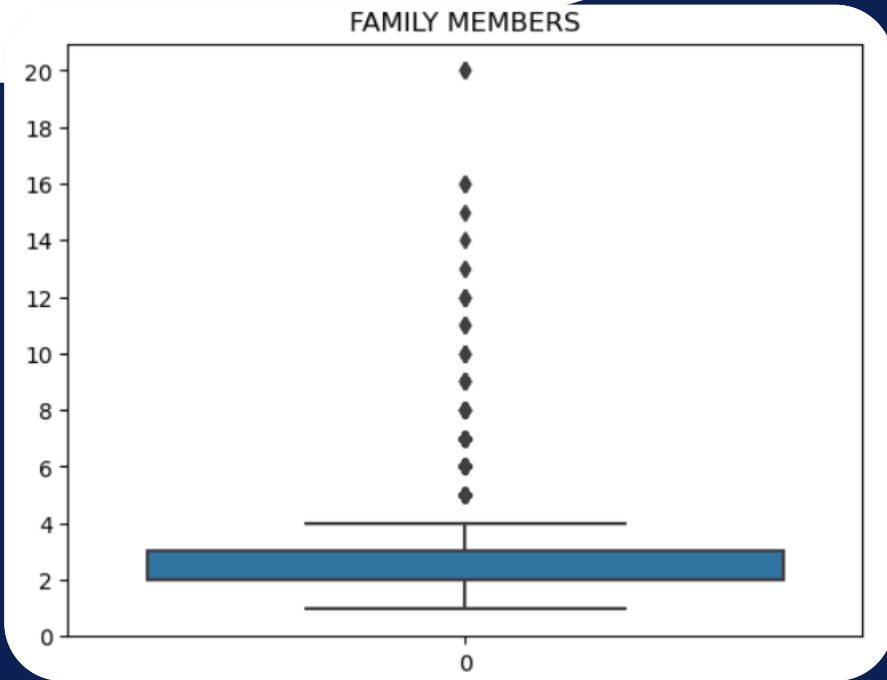
# COUNT CHILDREN



Count of children has outliers with 19-children families, while the majority have 2 or 3 children. These outliers tell us that there may be unique family structures or cultural dynamics. By understanding these outliers we can understand specific loan needs of larger families.

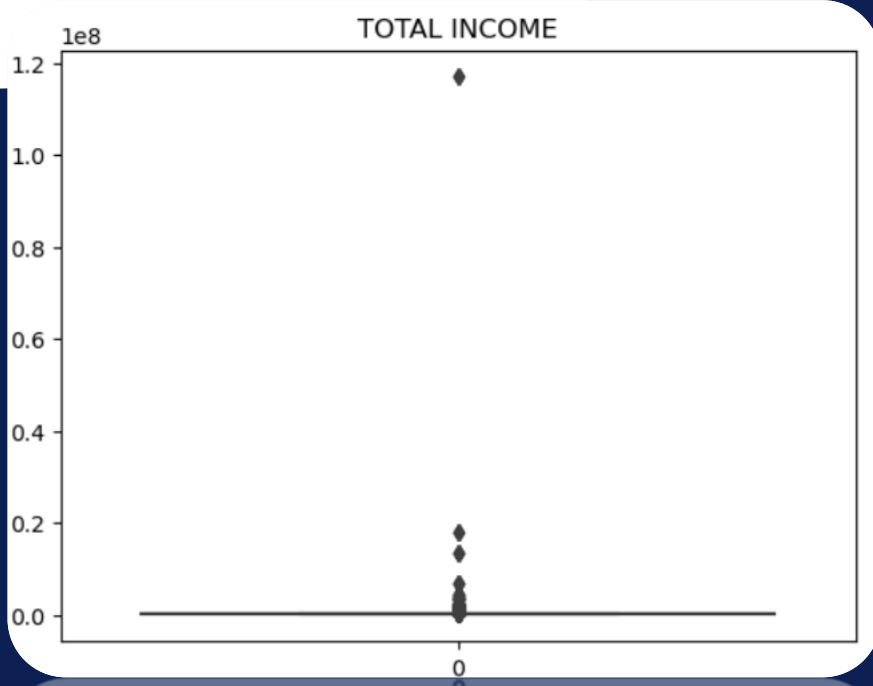


# COUNT FAMILY MEMBERS

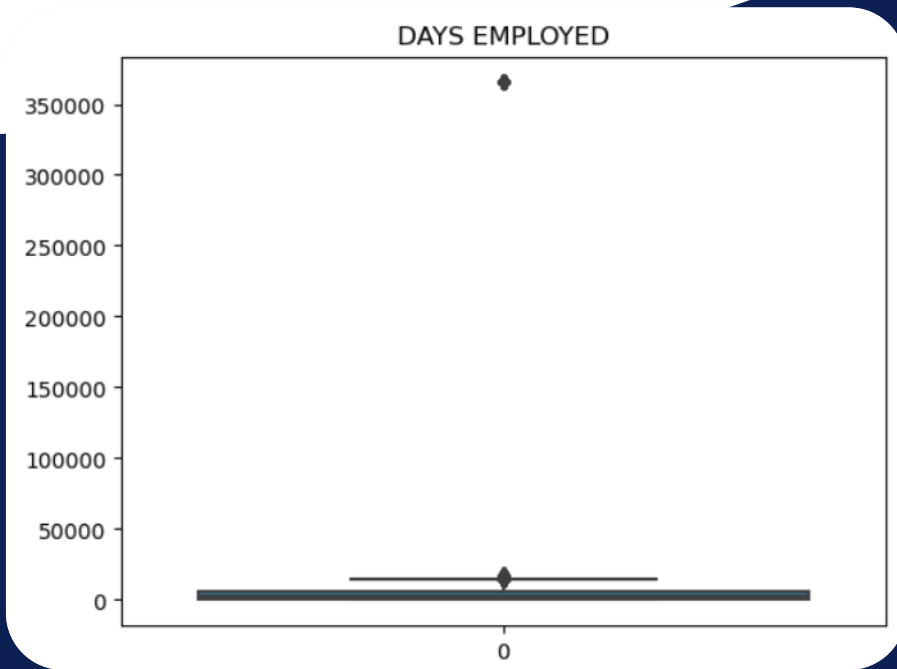


Count of family members exhibits outliers with 20-member families, while the majority have 3 or 4 members. These outliers may represent joint families. Knowing these outliers helps us to understand financial dynamics of larger households.

# TOTAL INCOME



Total Income have outliers indicates a presence of individuals with higher incomes compared to the majority of applicant. Understanding this distribution can help tailor financial products and assess borrower risk.



# DAYS EMPLOYED

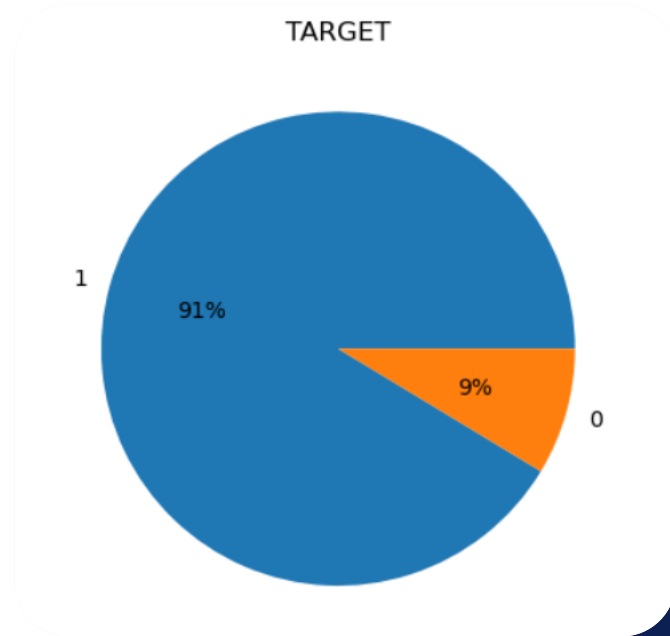
Days employed show outliers which indicates that there are some individuals which have been employed for an long time. This may indicate stable employment and a lower risk of job loss. Recognizing these outliers informs risk assessment and enables tailored loan products for long-tenured employees.



02

# Analyse Data Imbalance





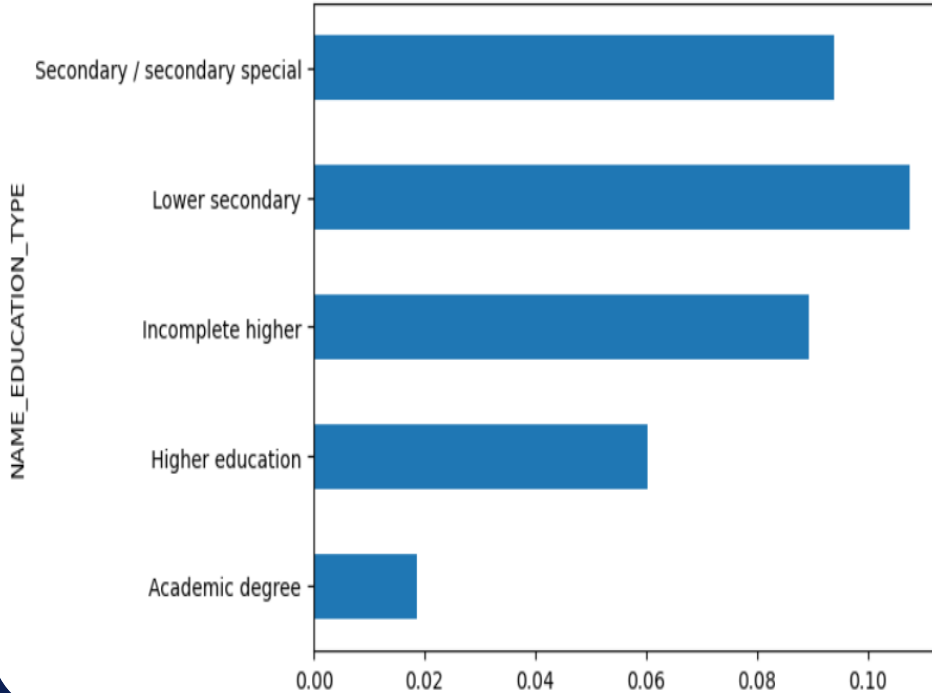
The dataset shows a significant class imbalance, with a ratio of approximately 1:11 between non-defaulters and defaulters. Such a skewed distribution indicates that the number of default clients is relatively small compared to non-default clients.

**03**

# **UNIVARIANT AND BIVARIANT ANALYSIS**



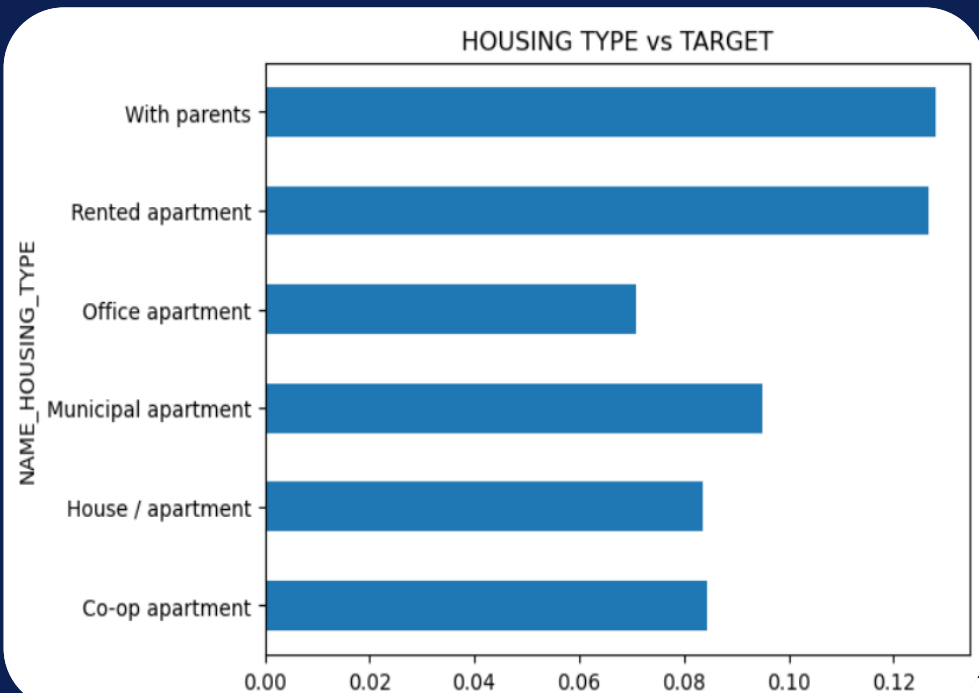
Education type vs Target



## EDUCATION TYPE Vs TARGET

When considering the applicants education type. Clients with secondary/secondary Special, Incomplete higher and lower secondary education are the ones who face payment difficulties.

# HOUSING TYPE vs TARGET

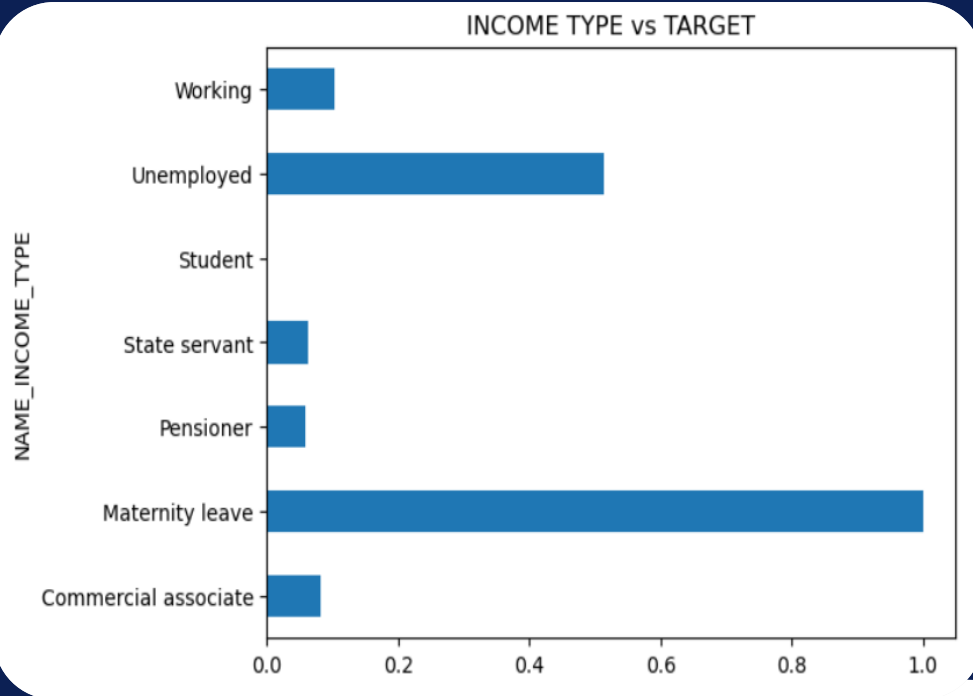


When considering the applicants housing type. Clients with parents, rented apartment and municipal Apartment are the ones who face payment difficulties

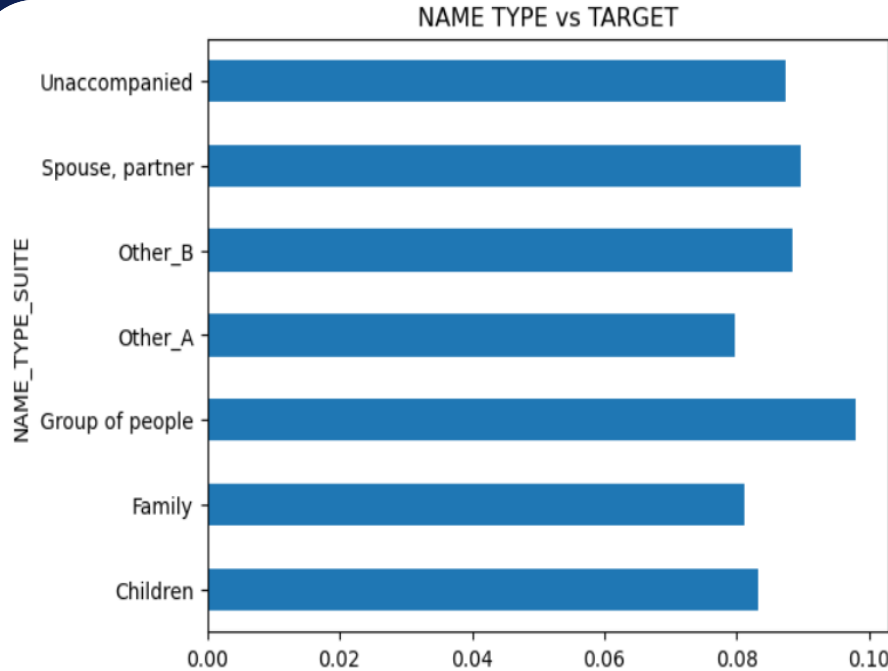


# INCOME TYPE vs TARGET

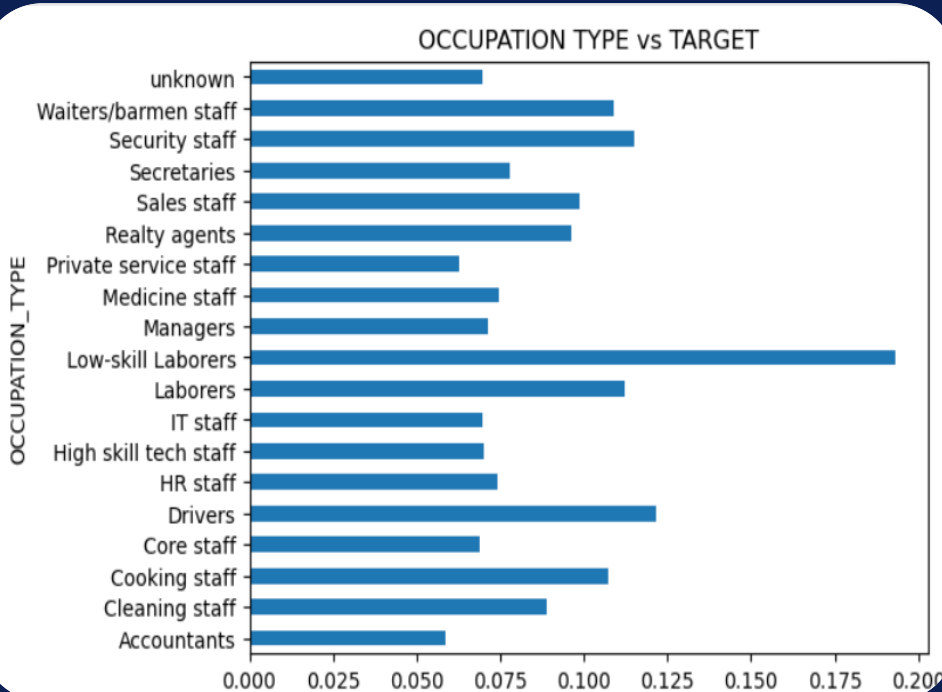
When considering the applicant's income type. Clients with income type maternity leave and unemployed are the ones who face payment difficulties



# NAME TYPE vs TARGET



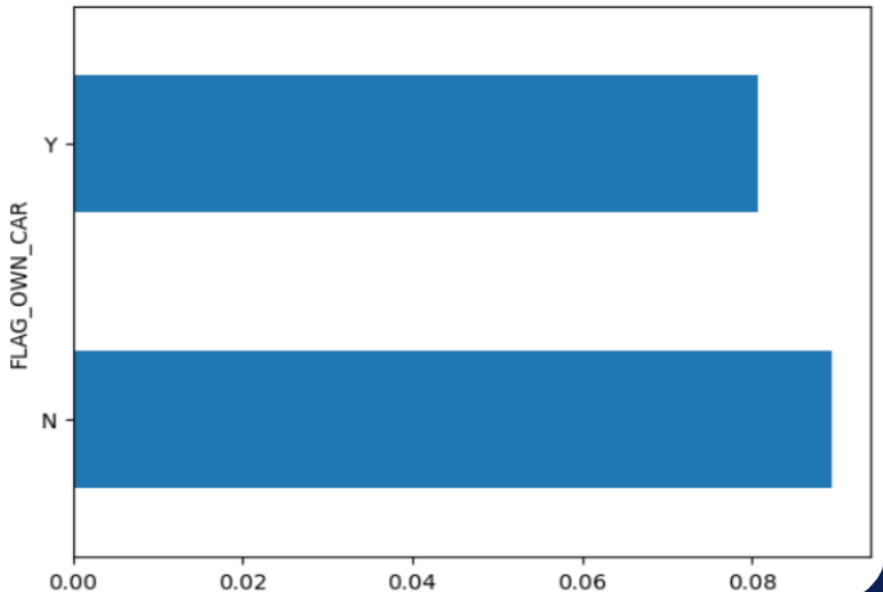
When considering the people who unaccompanied and group of people clients are the one who lies in defaulter list. Some clients are accompanied by family members and most of the time These clients have difficulties in paying back the loan.



# OCCUPATION TYPE vs TARGET

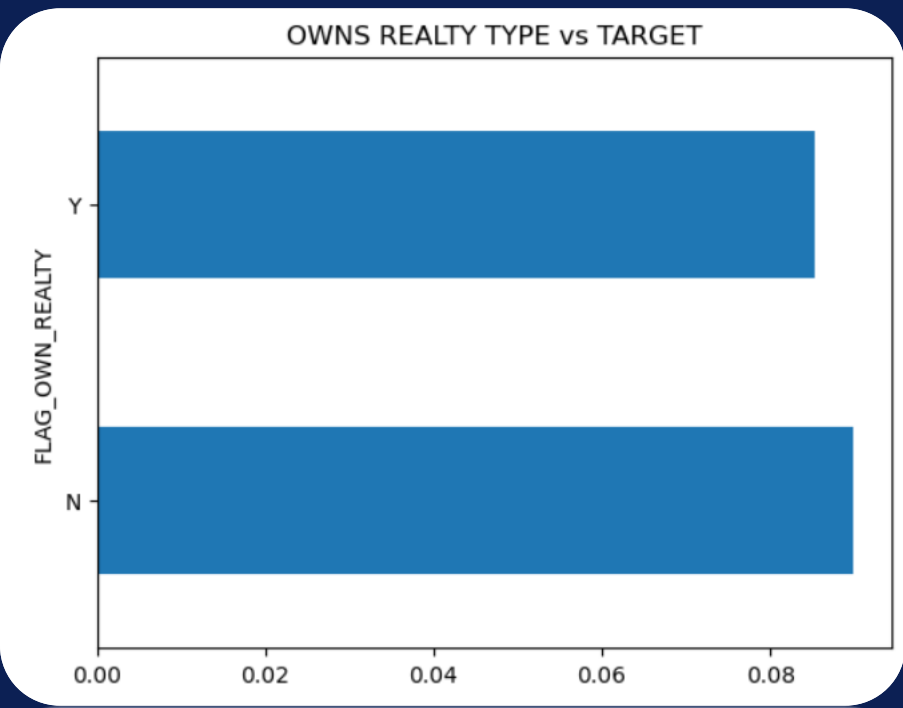
When considering the applicant's occupation type security staff, drivers, Cooking staff, low-skill labor, and waiter/bartender staff are the clients who face payment difficulties.

OWNS CAR vs TARGET



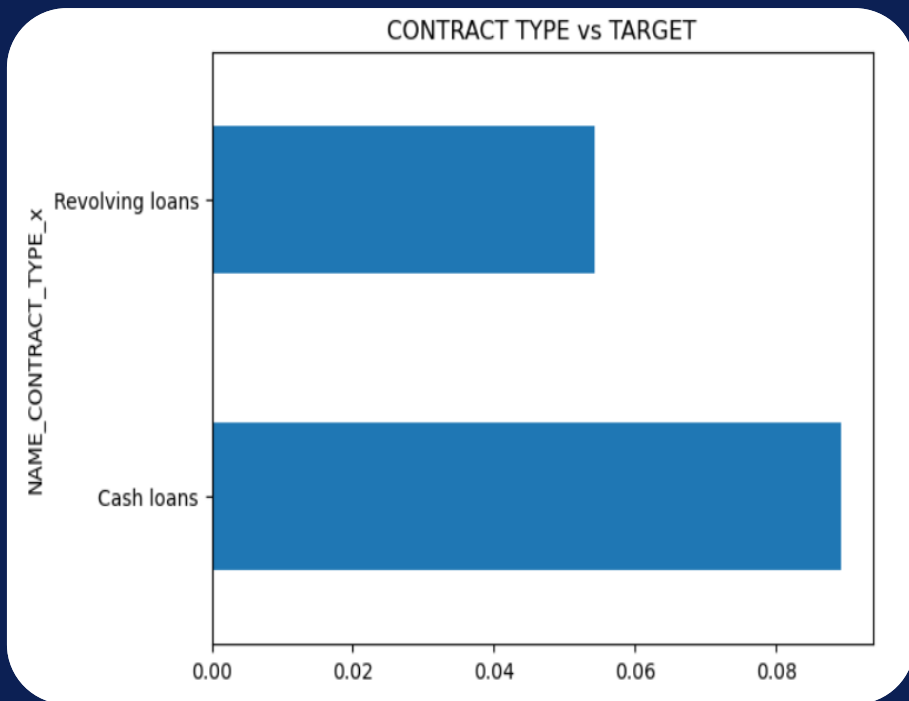
# OWNS CAR vs TARGET

Clients who have payment difficulties include both those who own car and those who don't. But the client who don't owns car faces slightly more payment issue as compared to the client who owns.



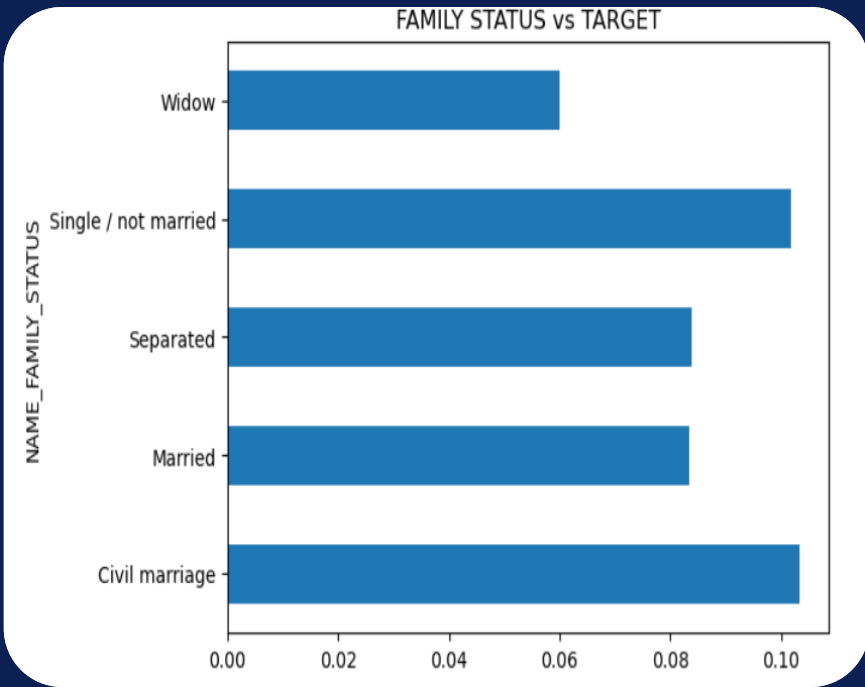
# OWNS REALTY vs TARGET

Clients who have payment difficulties include both those who own realty and those who don't



# CONTRACT TYPE vs TARGET

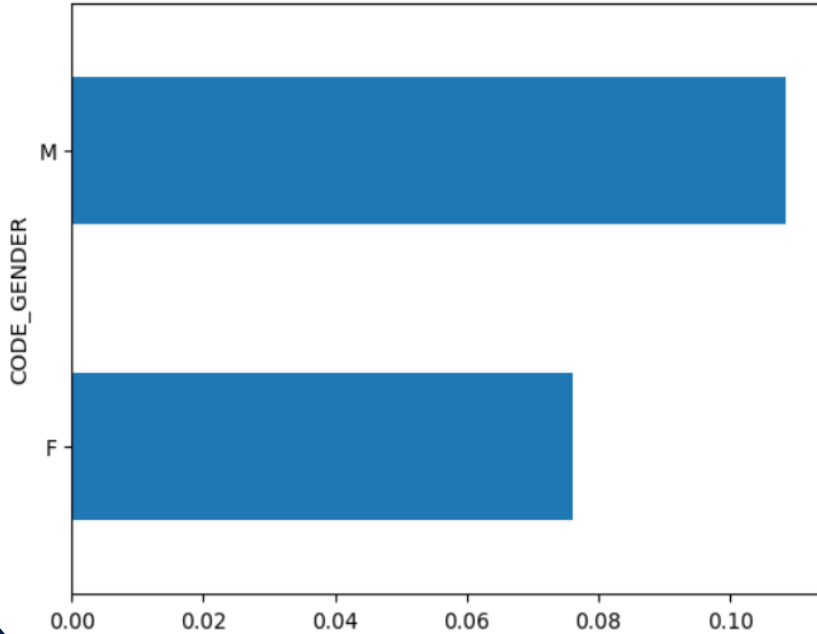
When considering the clients contact type, Clients who take cash loans often face payment difficulties compared to those who choose revolving loans.



# FAMILY STATUS vs TARGET

When considering the clients Family status, Clients who's family status is civil marriage and single/not married are less likely to pay back as compered to others.

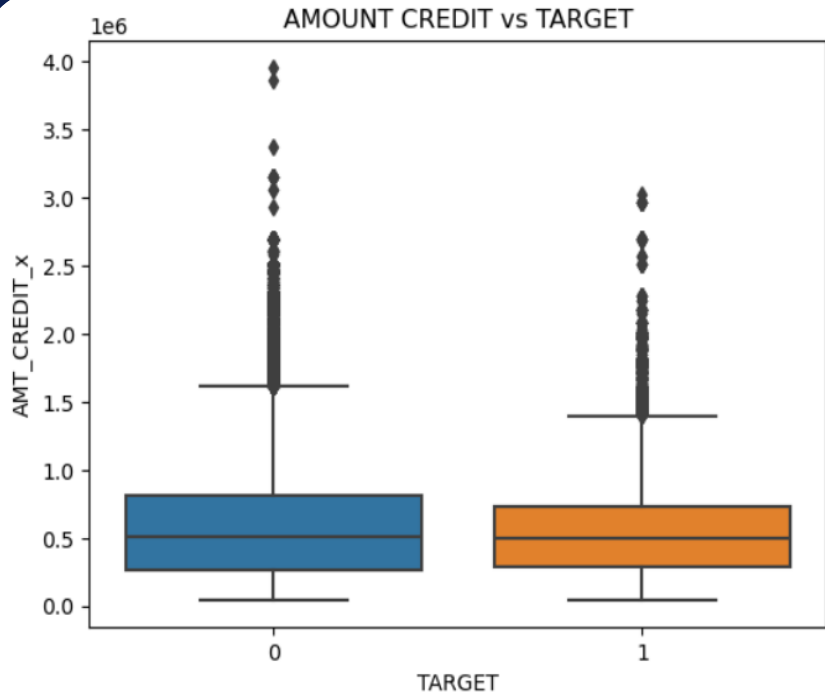
GENDER vs TARGET



# GENDER vs TARGET

When considering the clients gender, Clients who gender is male often face payment difficulties compared to those who's gender is female.

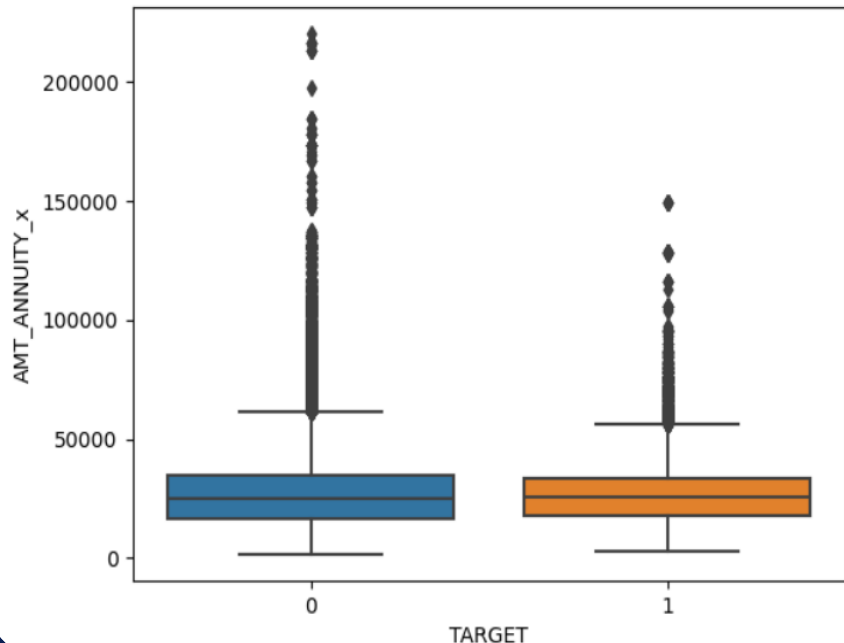




# AMOUNT CREDIT vs TARGET

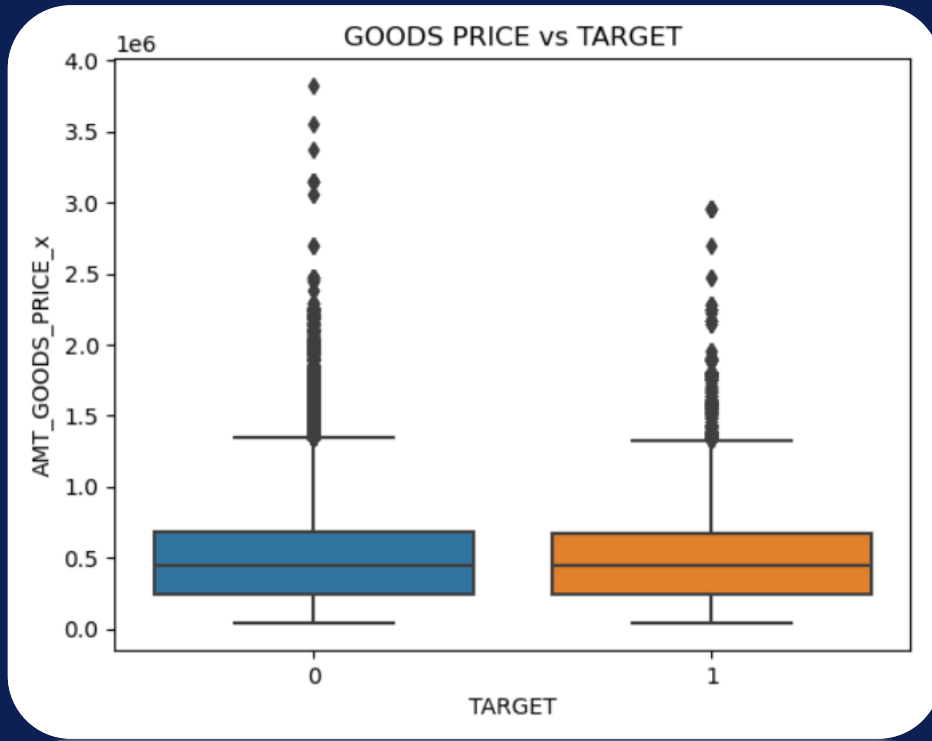
When considering the clients' credit amount, most clients with payment difficulties have loans ranging from 45,000 to 2,000,000.

AMOUNT ANNUITY vs TARGET



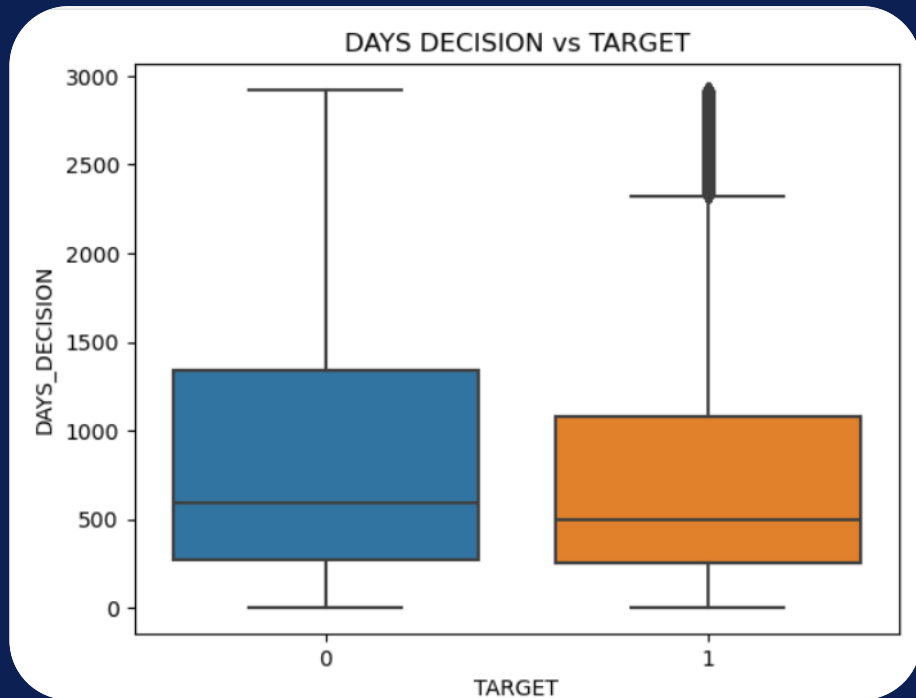
# AMOUNT ANNUITY vs TARGET

When considering the annuity amount, clients with payment difficulties mostly have annuity amounts between 5,000 and 150,000.



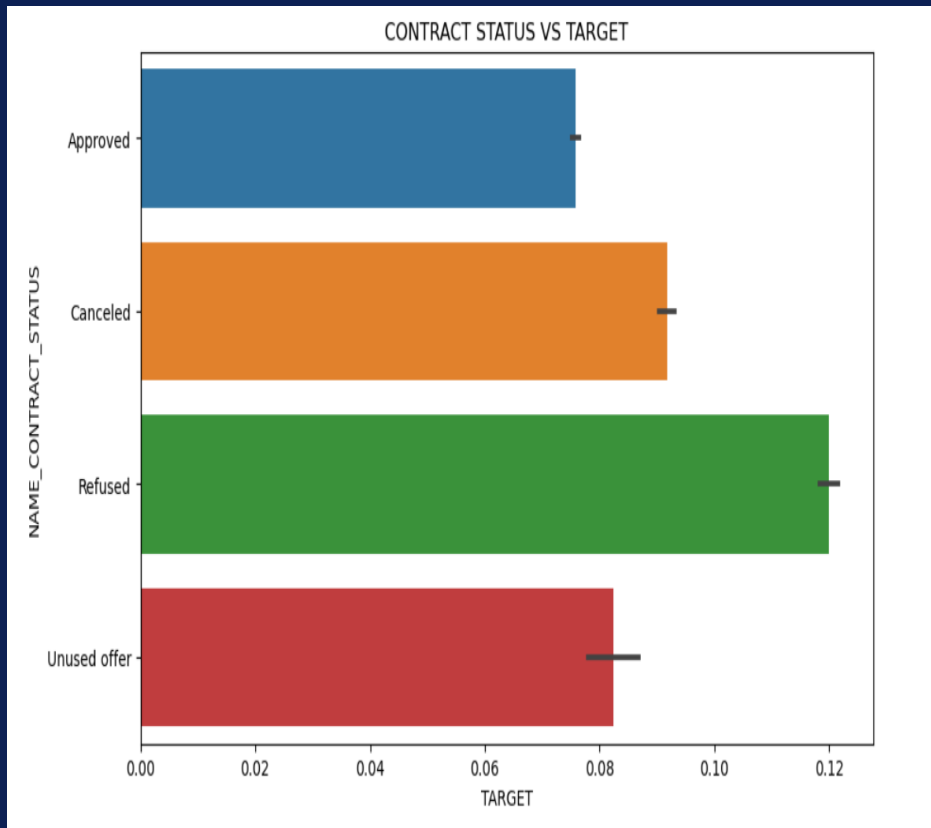
# GOODS PRICE vs TARGET

When considering the goods price, clients who have payment difficulties tend to purchase goods within the range of 45,000 to 2,000,000.



# DAYS DECISION vs TARGET

When considering the days taken for decision, whether it is for loan approval or other purposes, the minimum and maximum time observed were 2 days and 2920 days, respectively.



# CONTRACT STATUS vs TARGET

When considering the contract status, most of the loans are approved. However, it is noteworthy that some clients facing payment difficulties initially experience loan refusal for specific reasons. Subsequently, after resolving the identified issues, their loans are reevaluated and ultimately approved.

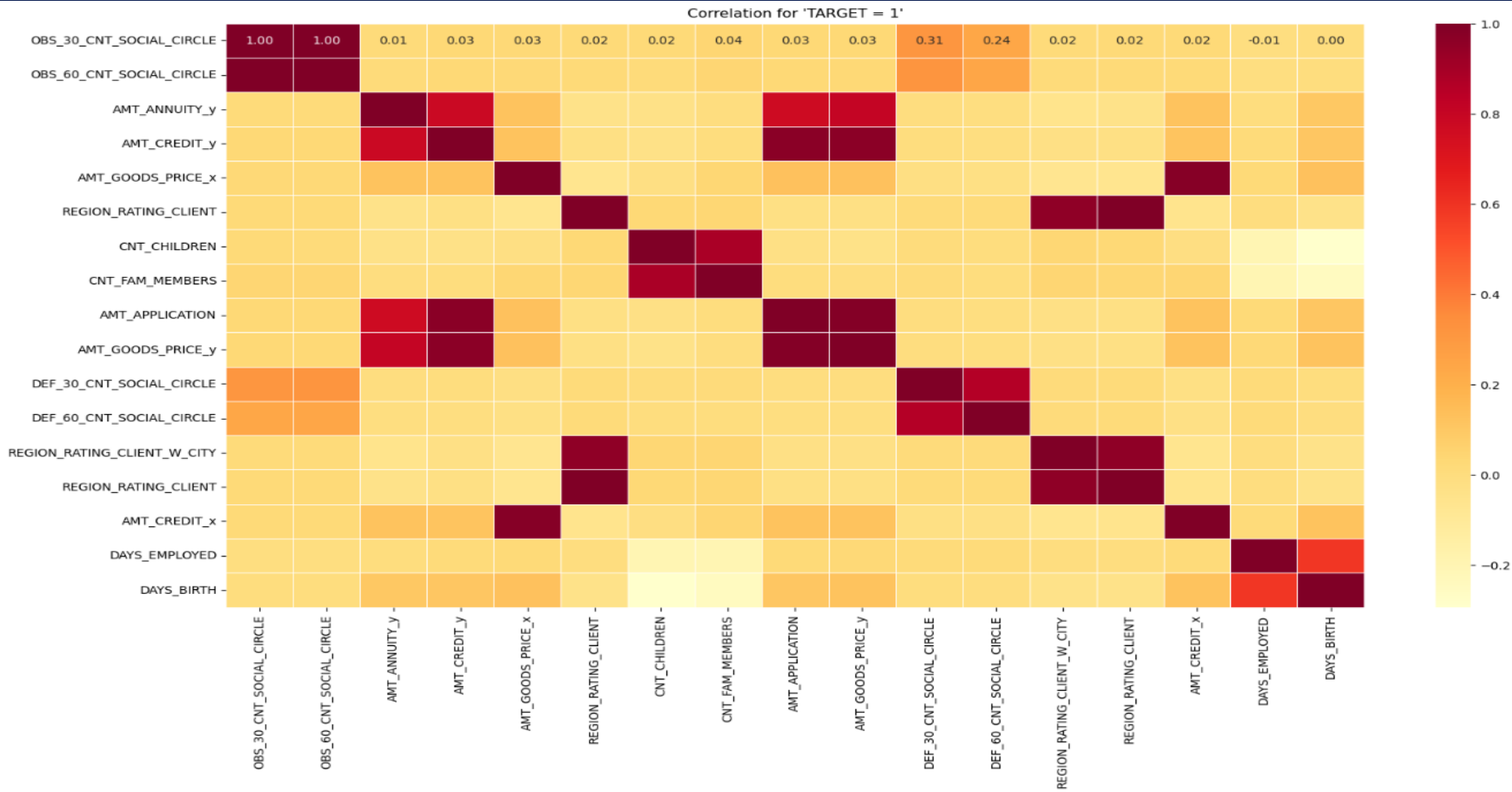


04

# IDENTIFYING TOP CORRELATION



## Top 10 Correlation For TARGET = 1

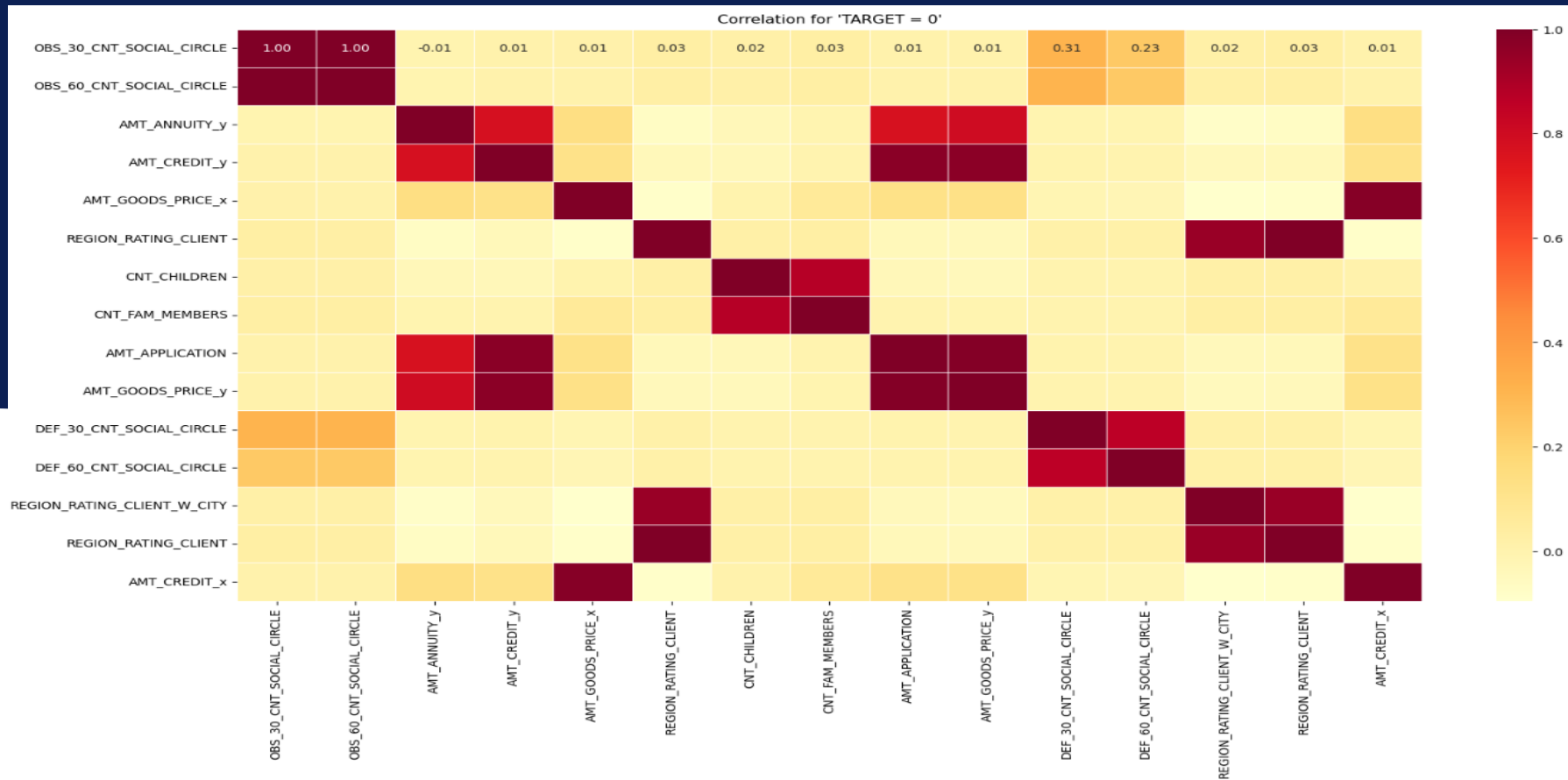


# Top 10 Correlation For TARGET = 1

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998403
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998403
AMT_APPLICATION	AMT_GOODS_PRICE_y	0.985811
AMT_GOODS_PRICE_y	AMT_APPLICATION	0.985811
AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.982256
AMT_CREDIT_x	AMT_GOODS_PRICE_x	0.982256
AMT_APPLICATION	AMT_CREDIT_y	0.974656
AMT_CREDIT_y	AMT_APPLICATION	0.974656
AMT_GOODS_PRICE_y	AMT_CREDIT_y	0.968599
AMT_CREDIT_y	AMT_GOODS_PRICE_y	0.968599
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956513
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956513
CNT_FAM_MEMBERS	CNT_CHILDREN	0.885400
CNT_CHILDREN	CNT_FAM_MEMBERS	0.885400
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.860910
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.860910
AMT_ANNUITY_y	AMT_GOODS_PRICE_y	0.806837
AMT_GOODS_PRICE_y	AMT_ANNUITY_y	0.806837
AMT_ANNUITY_y	AMT_CREDIT_y	0.778721
AMT_CREDIT_y	AMT_ANNUITY_y	0.778721



# Top 10 Correlation For TARGET = 0



# Top 10 Correlation For TARGET =0

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998552
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998552
AMT_GOODS_PRICE_y	AMT_APPLICATION	0.987610
AMT_APPLICATION	AMT_GOODS_PRICE_y	0.987610
AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.986254
AMT_CREDIT_x	AMT_GOODS_PRICE_x	0.986254
AMT_CREDIT_y	AMT_APPLICATION	0.975786
AMT_APPLICATION	AMT_CREDIT_y	0.975786
AMT_GOODS_PRICE_y	AMT_CREDIT_y	0.971807
AMT_CREDIT_y	AMT_GOODS_PRICE_y	0.971807
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.944131
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.944131
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878419
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878419
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.862324
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.862324
AMT_ANNUITY_y	AMT_GOODS_PRICE_y	0.799748
AMT_GOODS_PRICE_y	AMT_ANNUITY_y	0.799748
AMT_CREDIT_y	AMT_ANNUITY_y	0.775093
AMT_ANNUITY_y	AMT_CREDIT_y	0.775093

# CONCLUSION

In conclusion, the insights gathered from this project shed light on various aspects of loan applicants' characteristics and behaviors. The analysis of income totals revealed the presence of outliers, indicating individuals with significantly higher incomes compared to the majority. This highlights the importance of considering income disparity when assessing borrower risk and tailoring financial products to cater to different income levels.