

# BANK LOAN CASE STUDY

SALMAN BAREESH

# Project Description

- ★ To analyse data on loans and find pattern that predicts of client having difficulty of payment in the future
- ★ There are 2 types of risks associated while providing loan to a client
  - 1.If the loan is not provided to the customers who can pay back then it will be loss for the company
  - 2.If the loan is provided to the customer who can not pay then also it will be the loss for the company.

# Approach



## Data understanding

- Importing correct libraries
- Checking each column, Index, header, footer etc
- Identifying data quality issues

## Data Cleaning and Manipulation

- Missing value imputation analysis
- Checking the structure and the metadata
- Changing datatypes to date, time, string, int, bool, etc for ease of analysis

## Outlier check and data imbalance check

- Checking the data for outliers that would cause the analysis to be biased
- Checking for imbalances, ratio, percentage of imbalance

## Data analysis

- Business requirement oriented analysis
- Correlation between columns
- Univariate analysis
- Bivariate analysis
- Creating plots to understand the data better and find insight

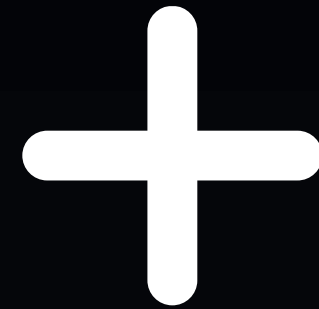
## Conclusion and Presentation

- Explains Business-driven, type-driven and data-driven metrics created in previous steps
- Helps in getting better understanding about the data

# Tech-Stack Used



Google Colab python used  
to do Exploratory Data  
Analysis



Canva used to do  
Presentation

# Identifying Missing Data

There are 66 columns that contains missing values out of 122 columns and deleted columns which have blanks percentage more than 30%

## Missing values Imputation

Median - AMT\_ANNUITY,AMT\_GOODS\_PRICE,AMT\_REQ\_CREDIT\_BUREAU columns

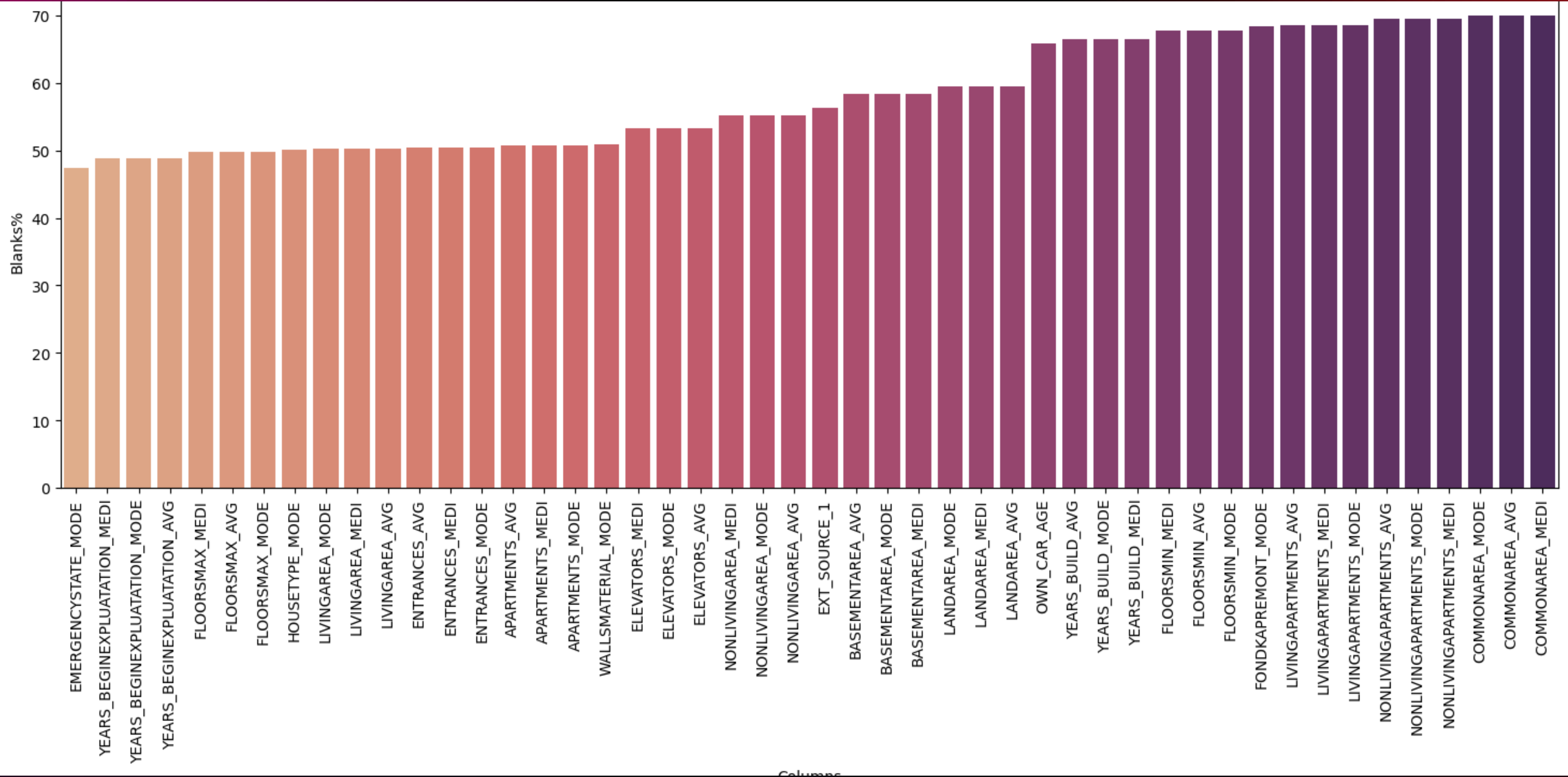
Mode - NAME\_TYPE\_SUITE, OCCUPATION\_TYPE

OCCUPATION\_TYPE : imputing missing values with 'Others' in OCCUPATION\_TYPE as it is categorical data

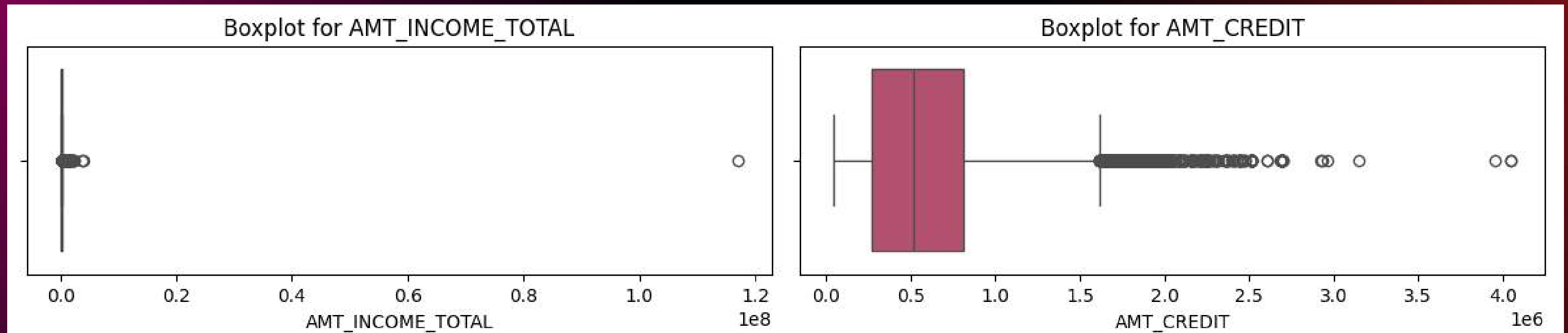
EXT\_SOURCE\_1,EXT\_SOURCE\_2,EXT\_SOURCE\_3 : joined three columns by taking average and set name as EXT\_SOURCE

CNT\_FAM\_MEMBERS : deleting the row since the count of null values is only 2, and it wouldnt have much impact on overall data

# Identifying Missing Data



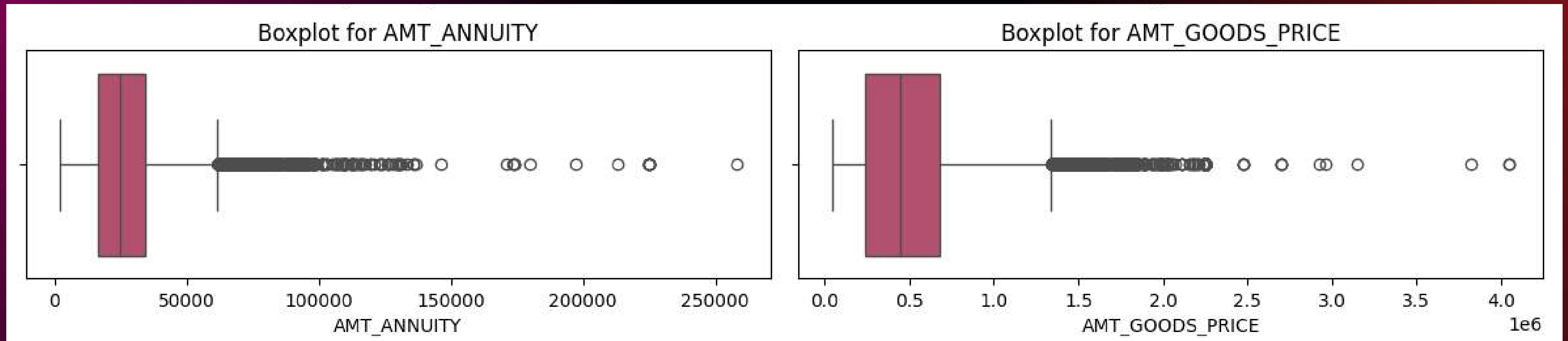
# Identifying Outliers in the Dataset



AMNT\_INCOME\_TOTAL - We are deleting rows of income more than 100M

AMNT\_CREDIT - As outliers are valid so we are not removing

# Identifying Outliers in the Dataset

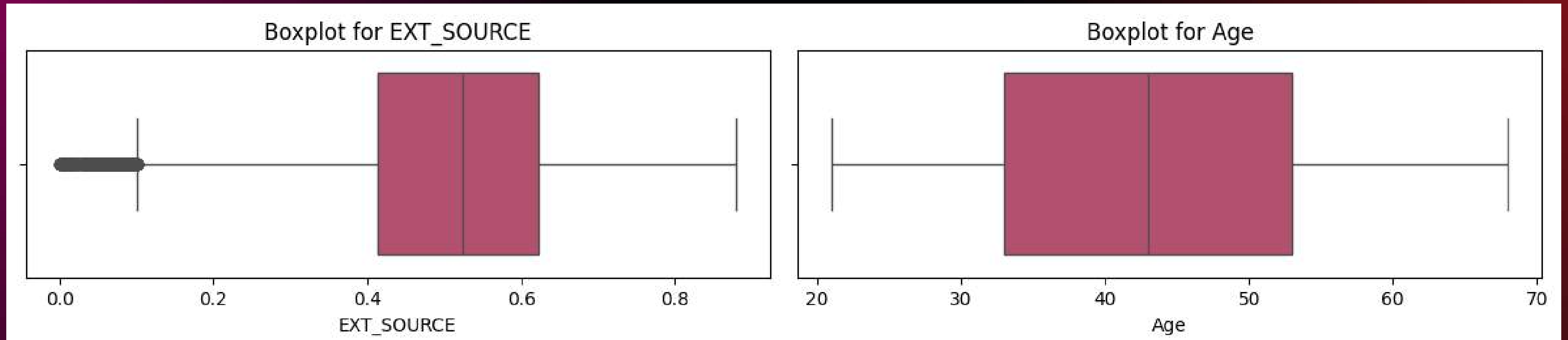


AMNT\_ANNUIITY - As outliers are valid so we are not removing

AMNT\_GOODS\_PRICE- As outliers are valid so we are not removing



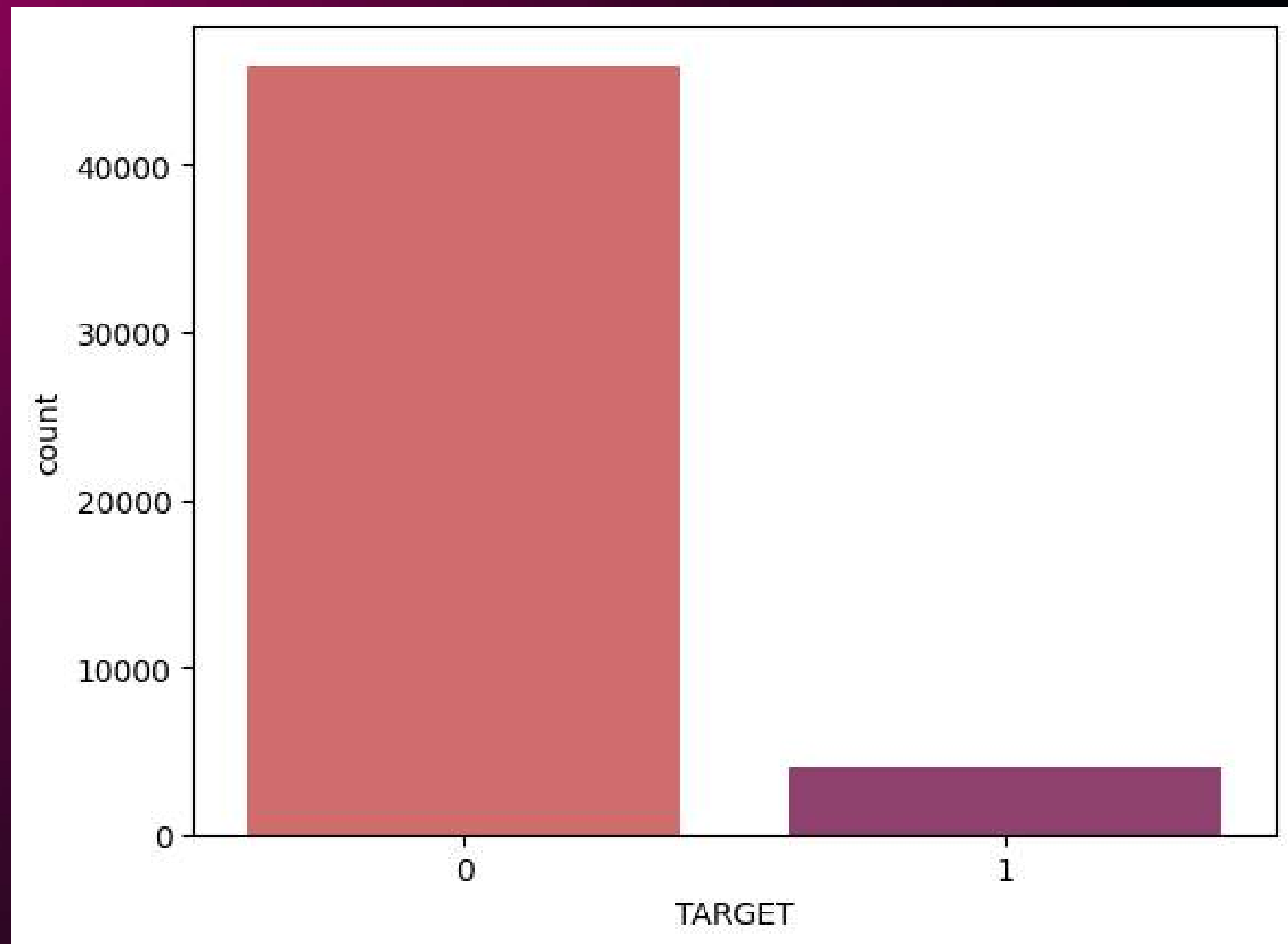
# Identifying Outliers in the Dataset



EXT\_SOURCE - As outliers are valid so we are not removing

Age- There are no outliers

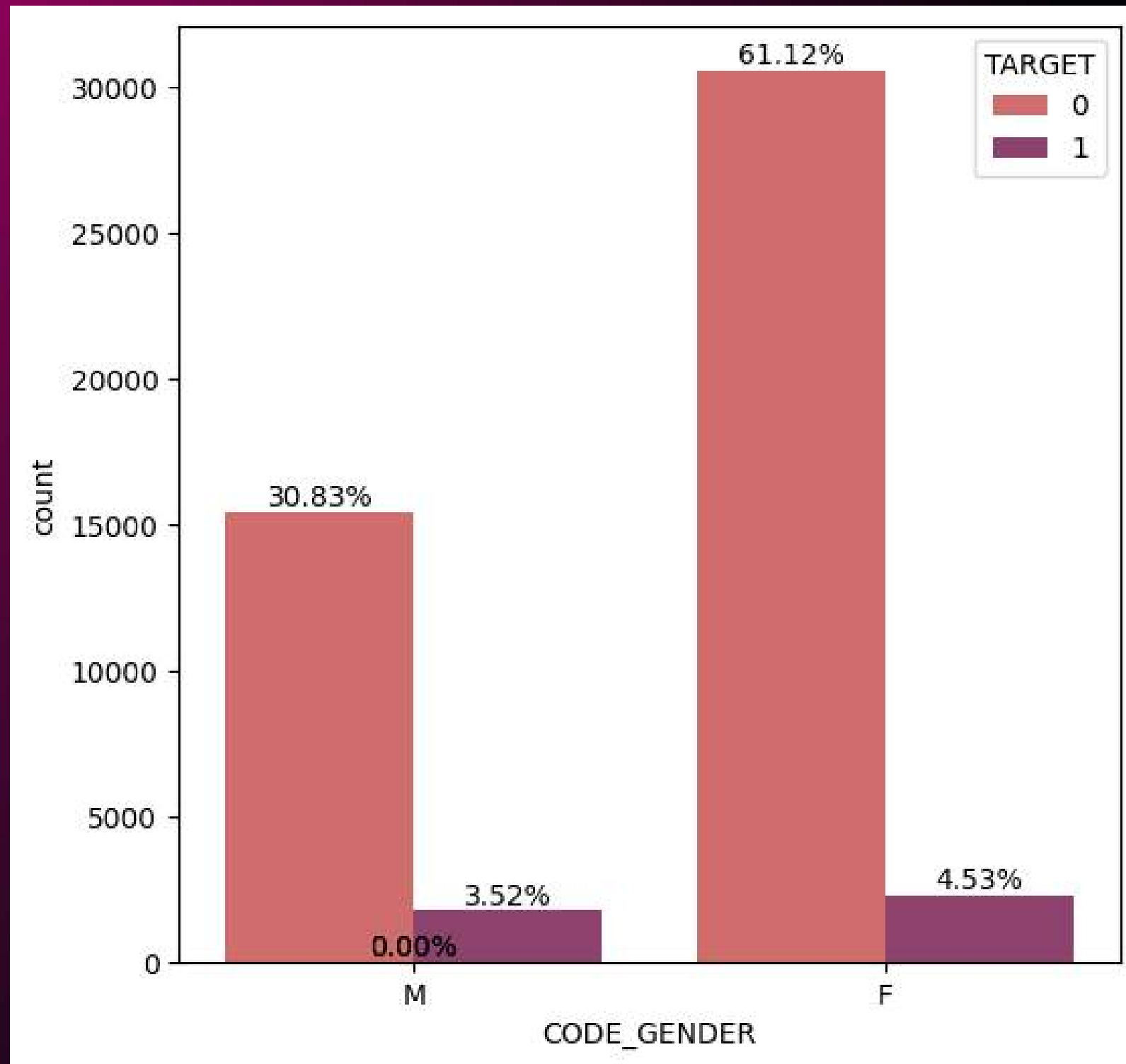
# Analyzing Data Imbalance



The ratio for people who have difficulty paying to on-time payers is 11.387150050352467

1 in every 11 applicant has payment difficulty.

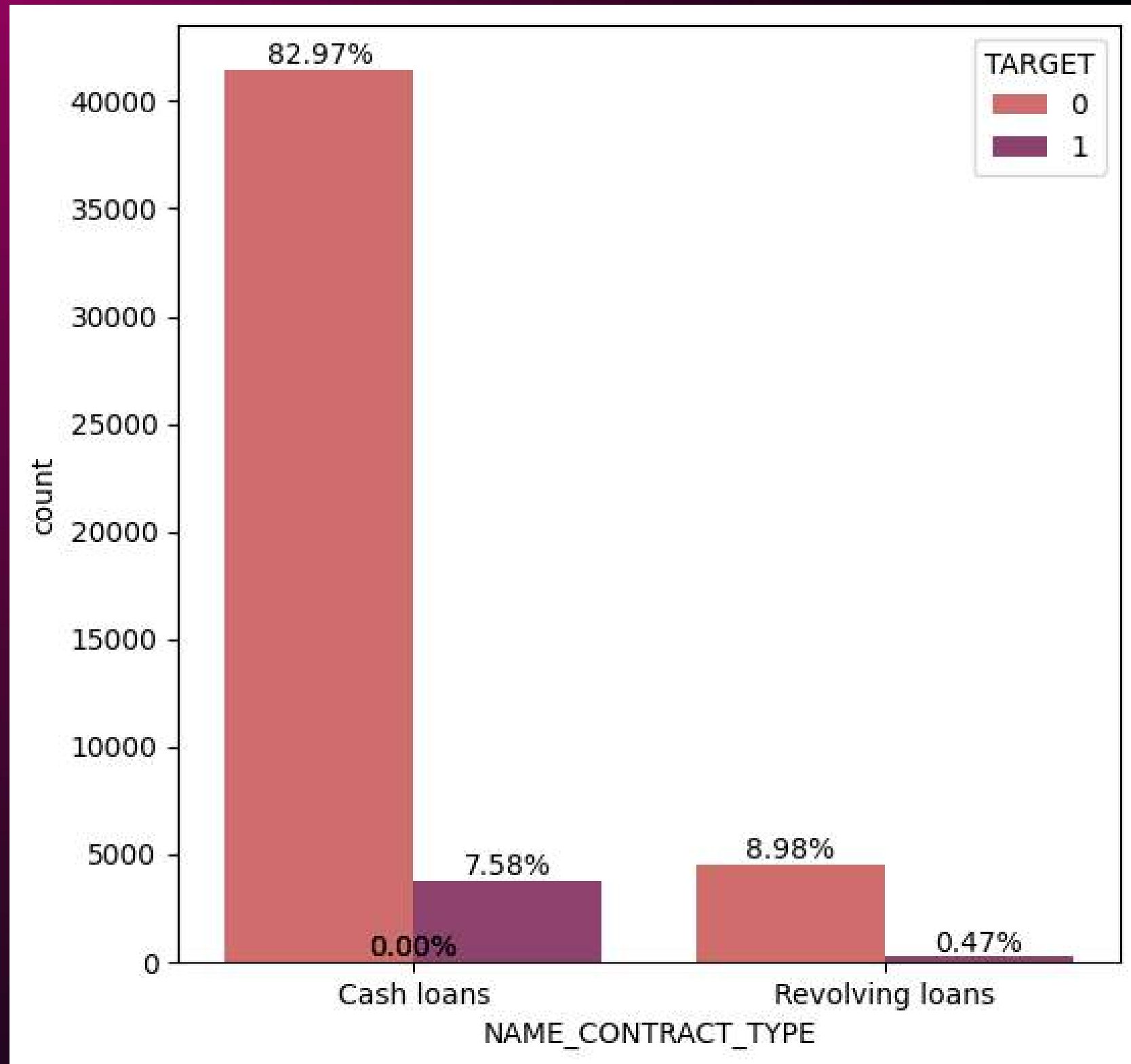
# Univariate Analysis- Gender



Females have processed more loans in comparison to males.

There is a 9.4% decrease in "Male" values from CoDE\_GENDER b/w clients with payment difficulties to on-time payments.

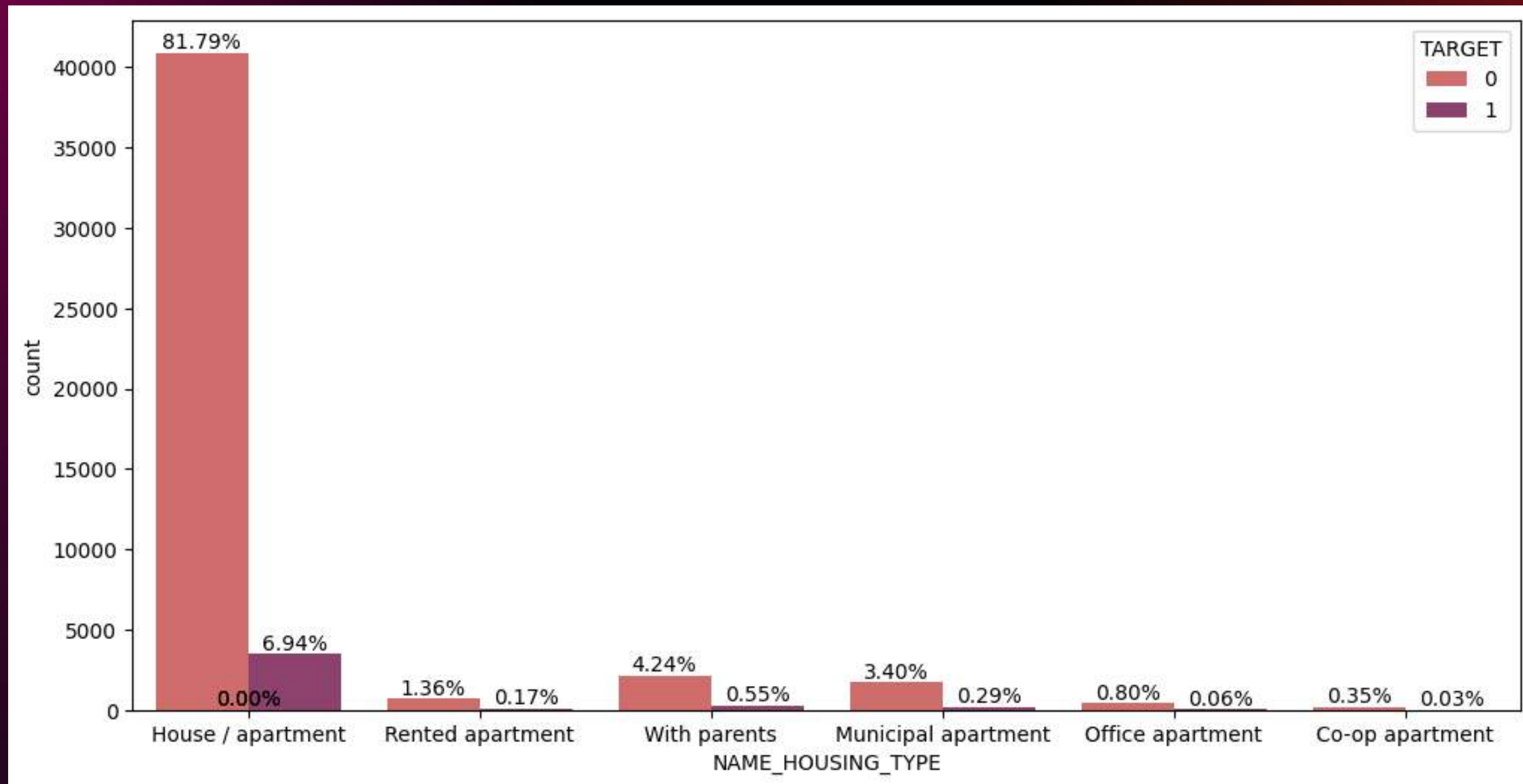
# Univariate Analysis- Type of loans



More Cash loans have been processed compared to revolving loans

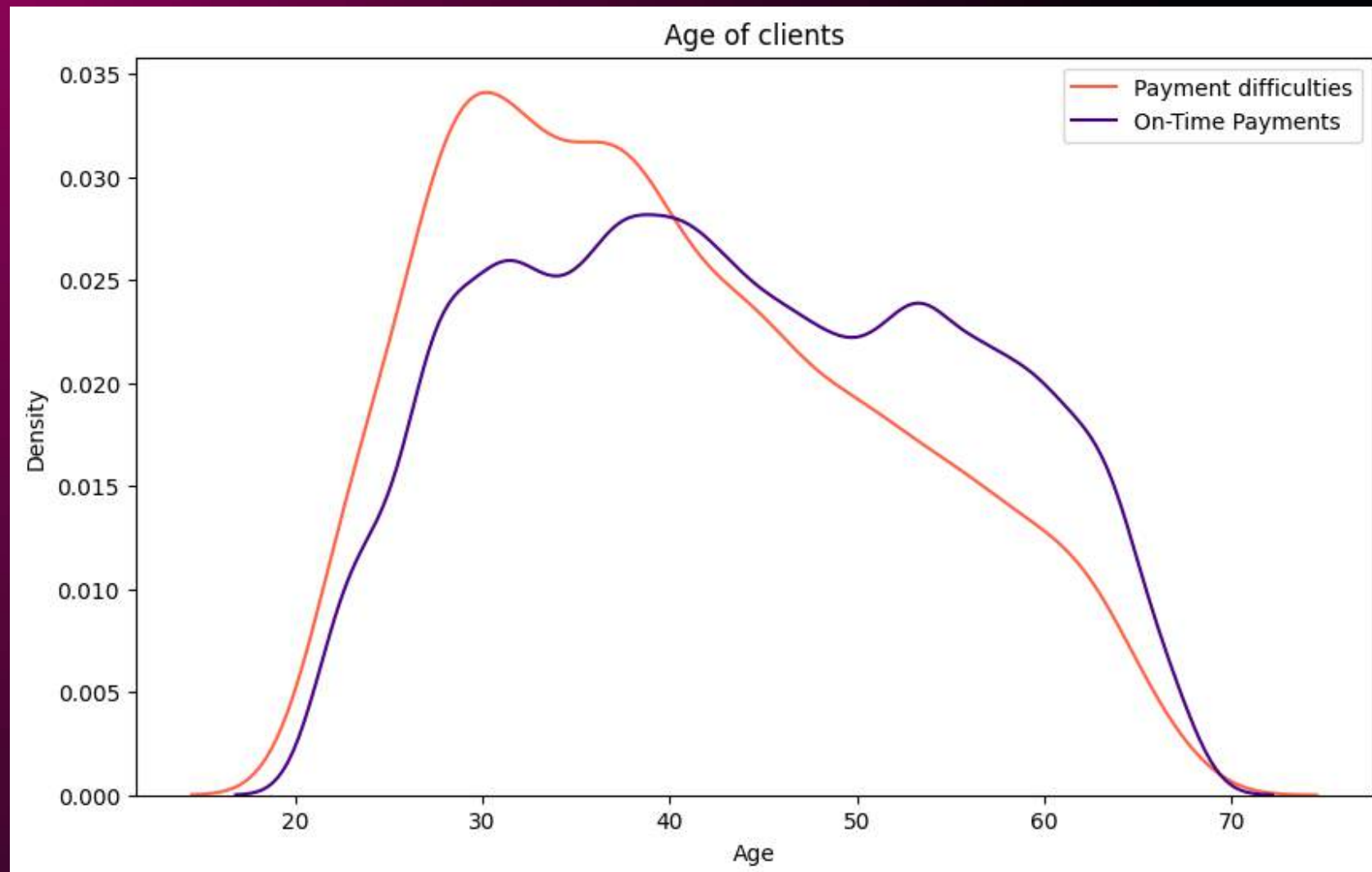
The most approved type of loan is consumer loan, followed by cash loan and Revolving loan has a higher refusal rate

# Univariate Analysis- Type of loans

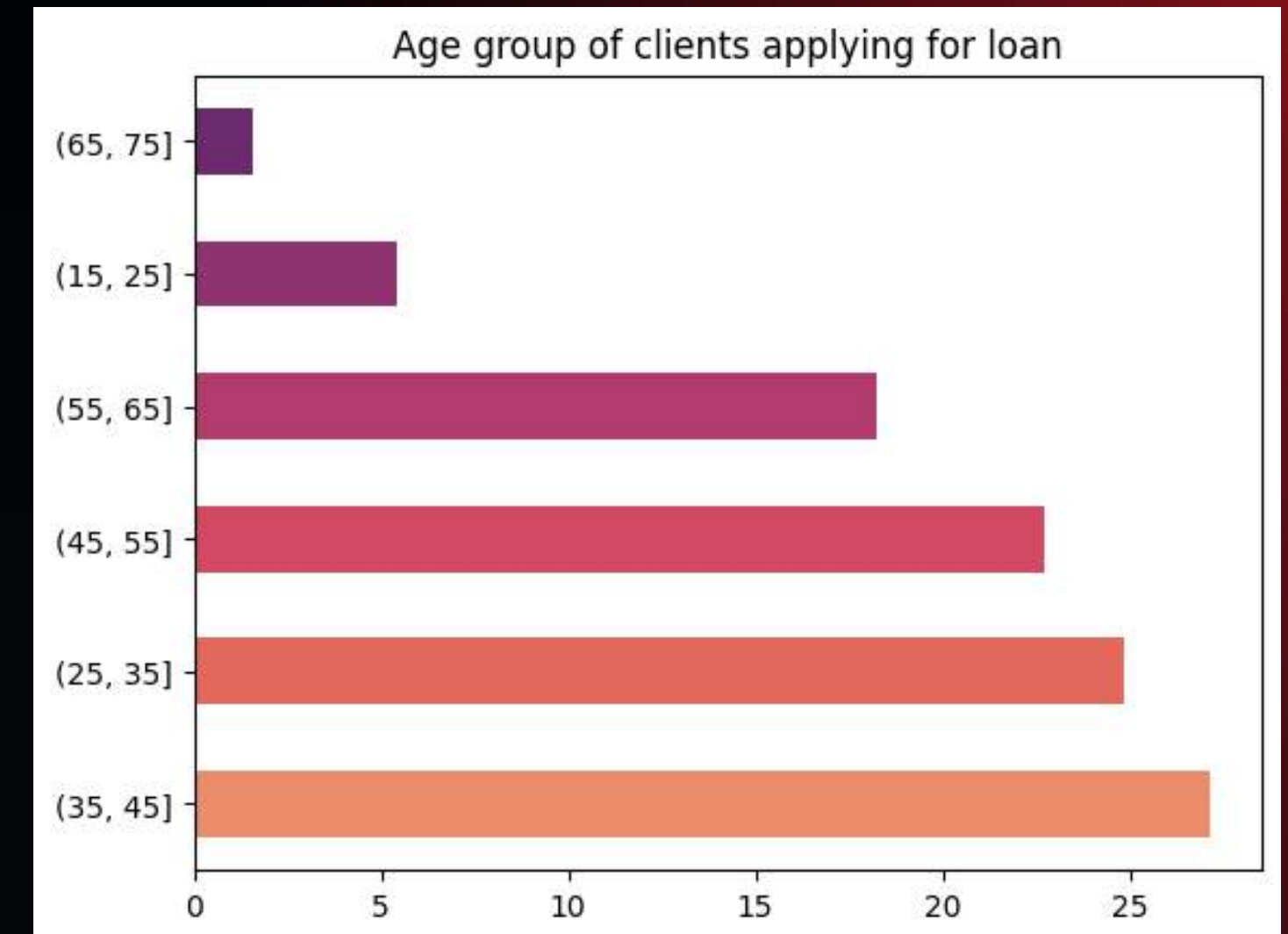


People who live in rented flat or with parents tends to default more on loan payments than people living in House/Flat or office flats.

# Univariate Analysis-Age

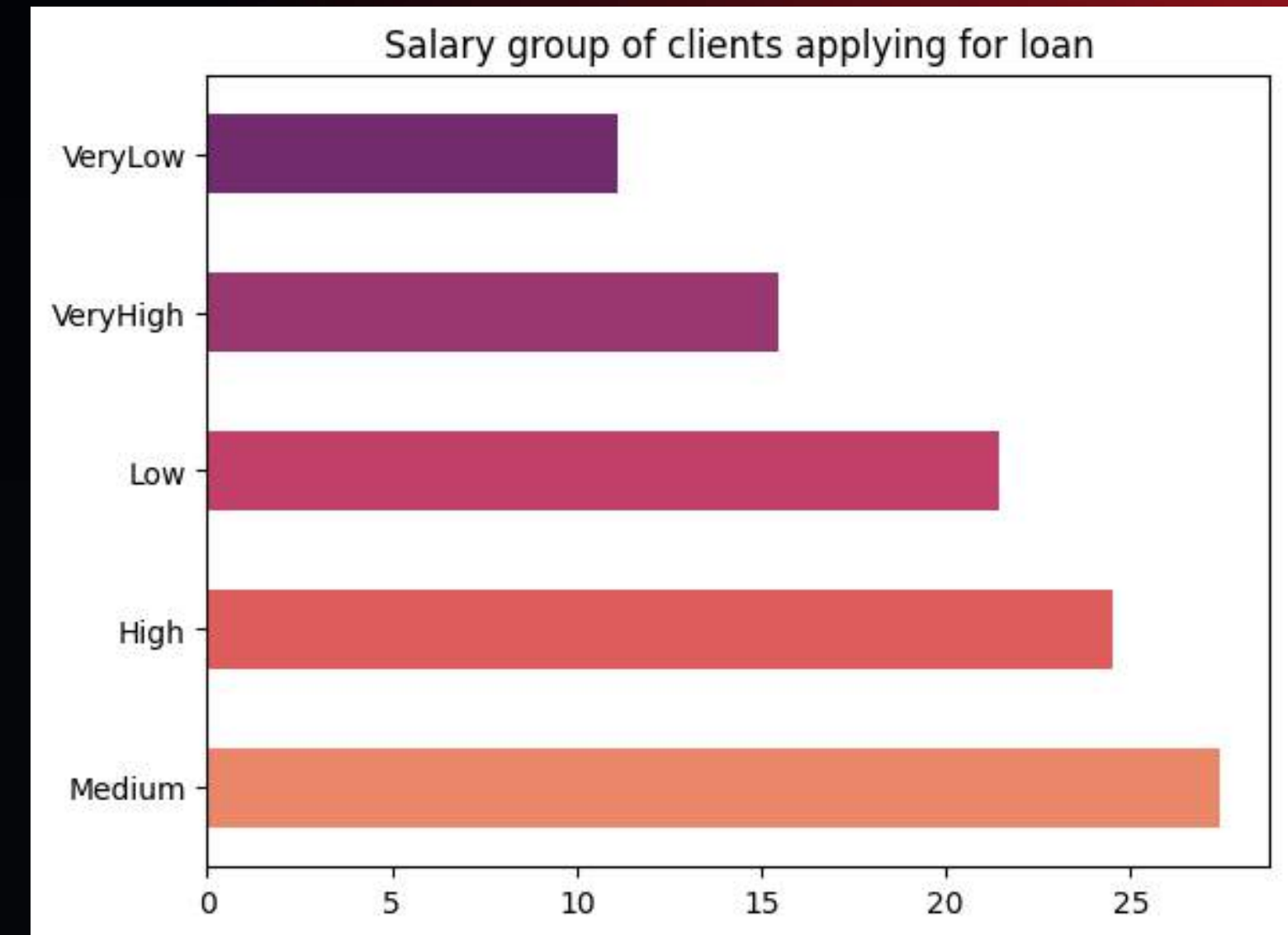
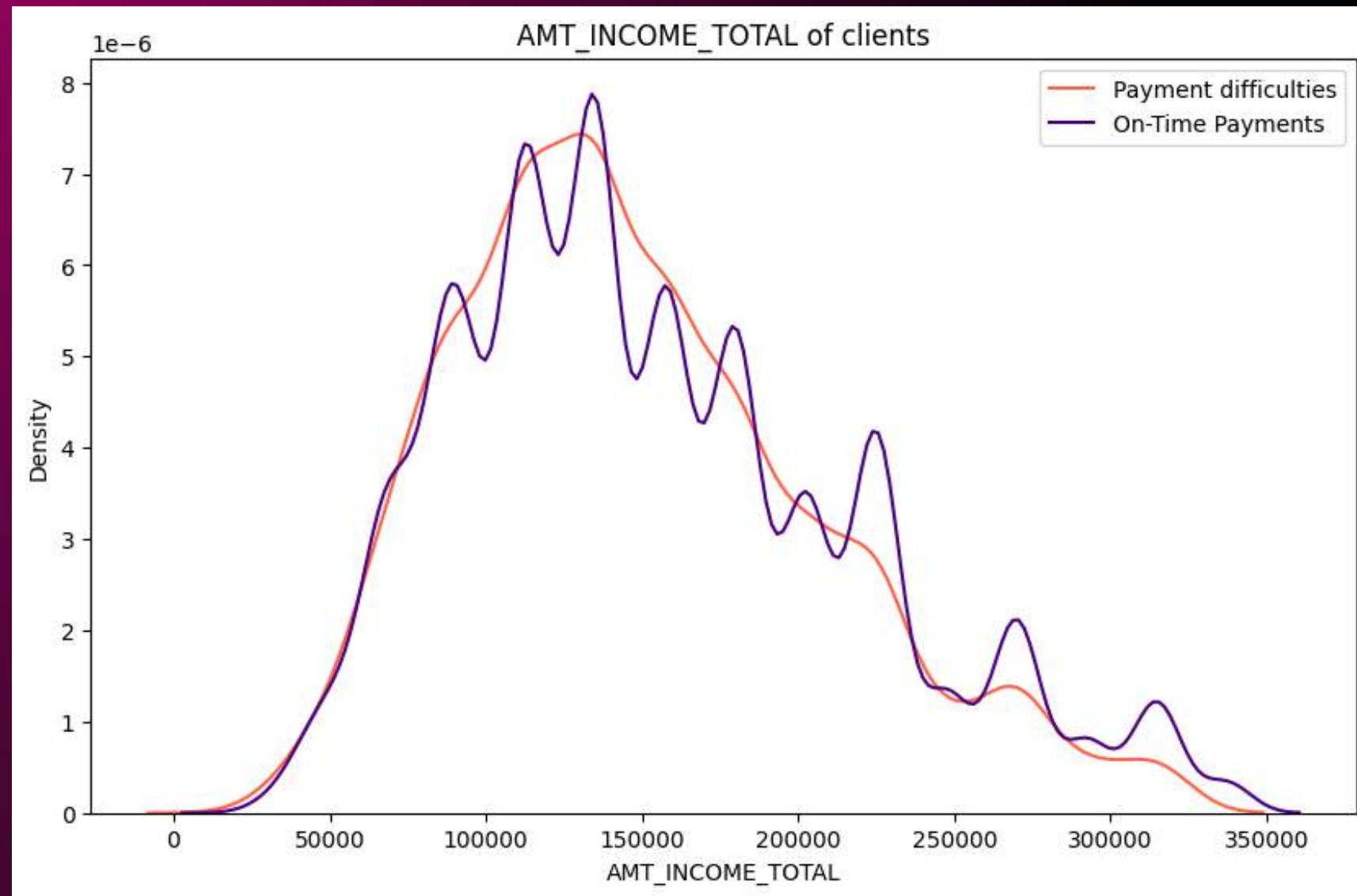


For YEARS\_BIRTH between 20 and 40, there are more clients with Payment difficulties



35-45 Age group is the largest Group of Age applying for loans.

# Univariate Analysis-Income

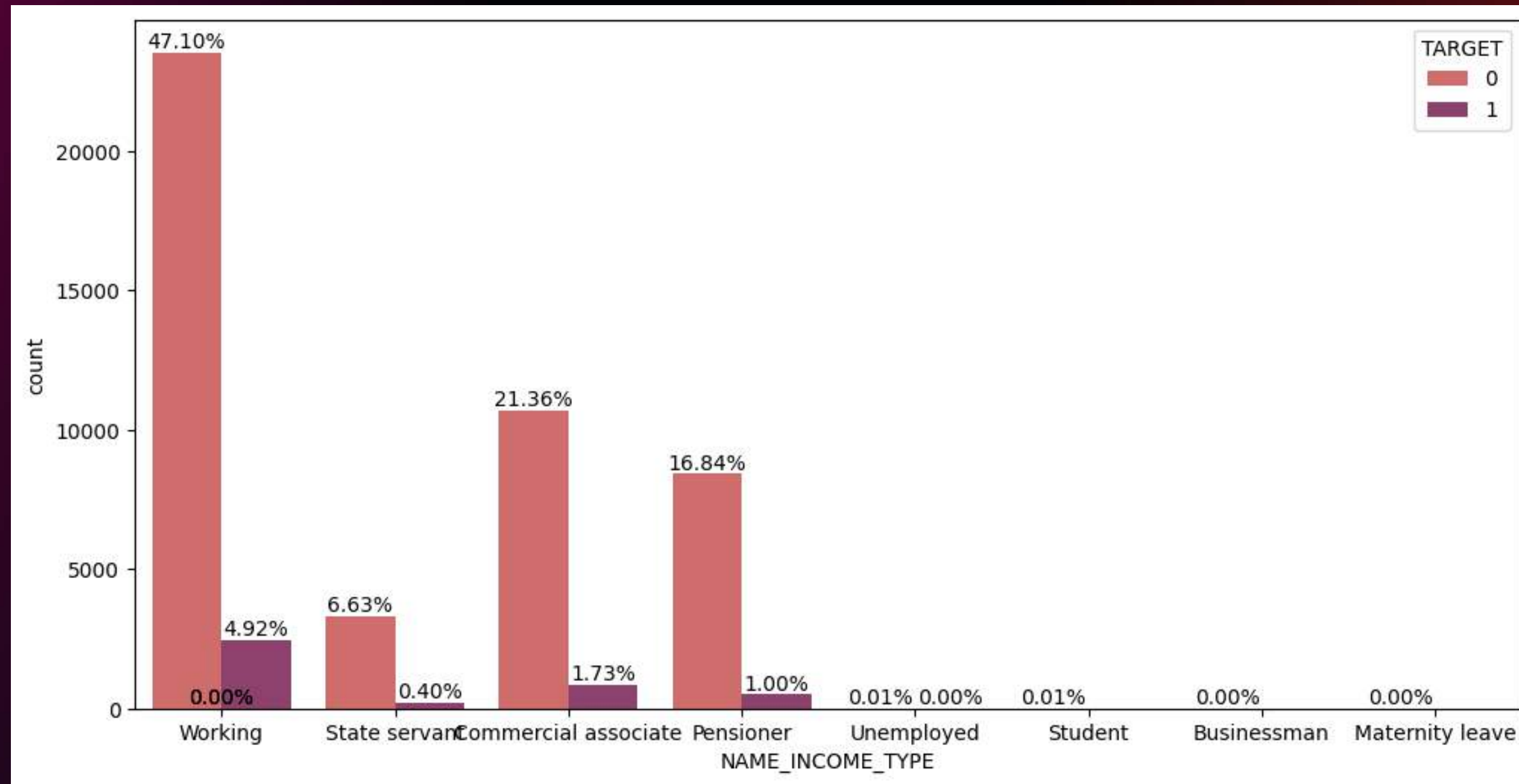


'Medium' Income group is the largest group applying for loans, followed by 'High' income group. 'VeryLow' income group is the smallest group applying for loan.

Medium income groups defaults a lot more compared to other groups Amount Income 'Low' and 'High' is a concern as well



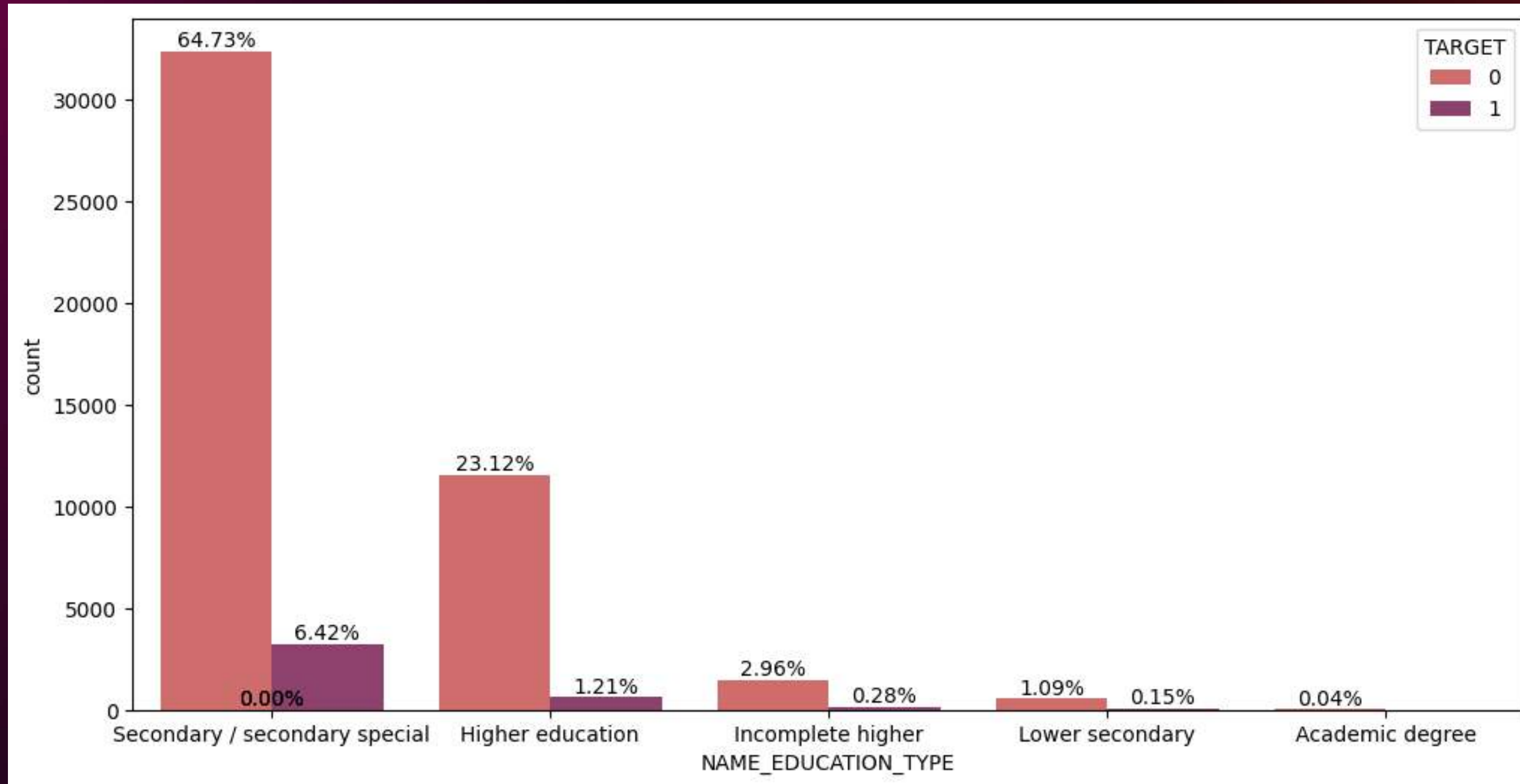
# Univariate Analysis-Income Type



Working' class have processed more loans in comparison to other categories and Pensioners have better on-time payment

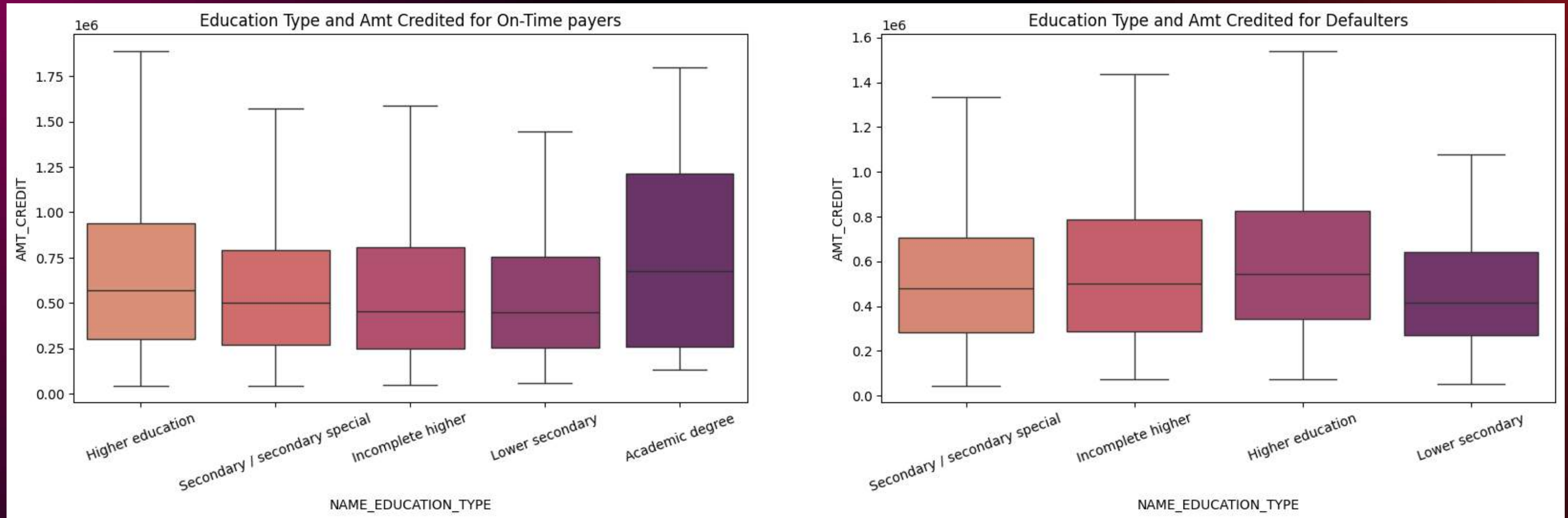


# Univariate Analysis-Education Type



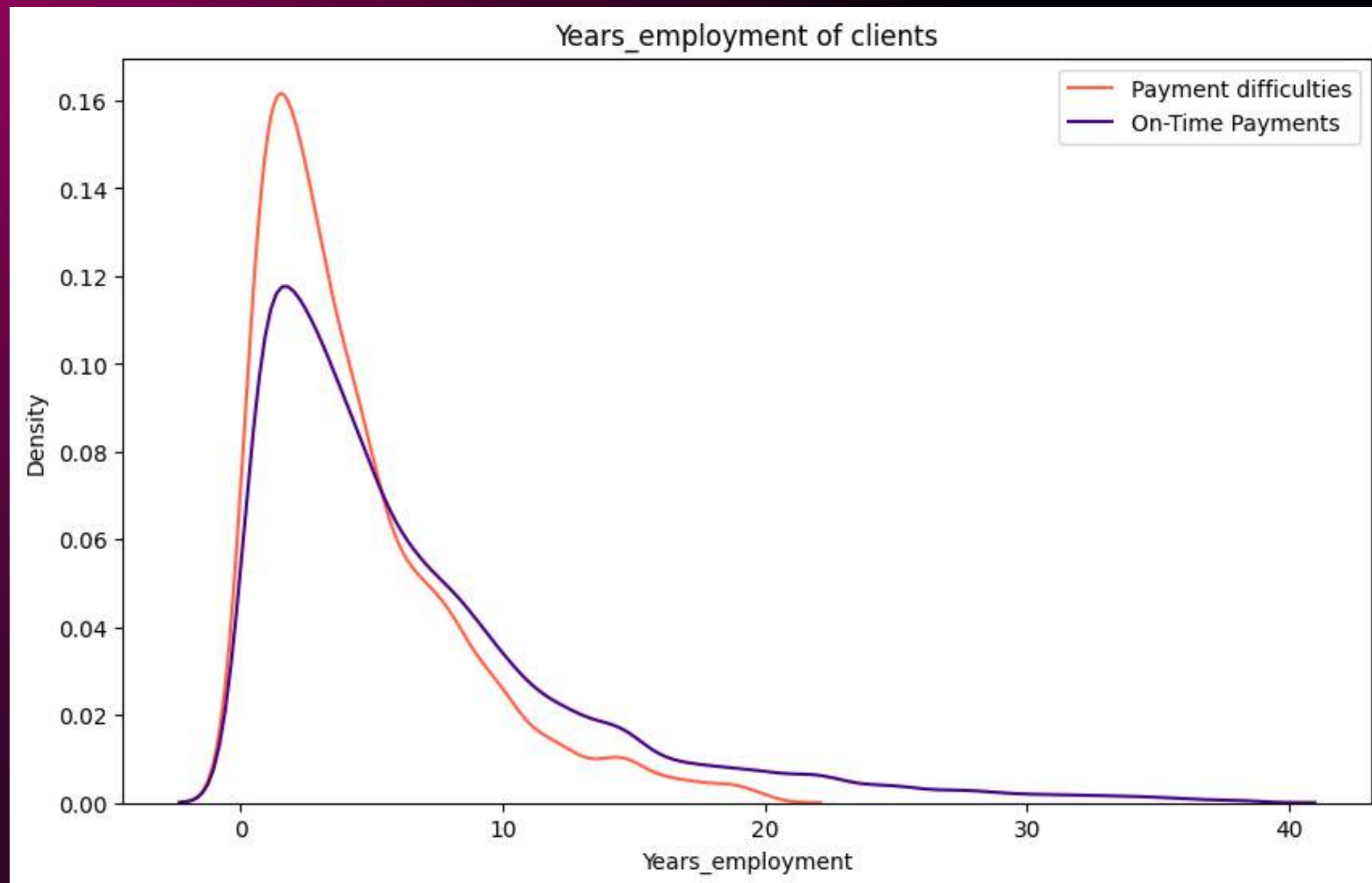
Clients with 'Higher education' have better on-time payments than others and have less payment difficulties.

# Univariate Analysis-Education Type



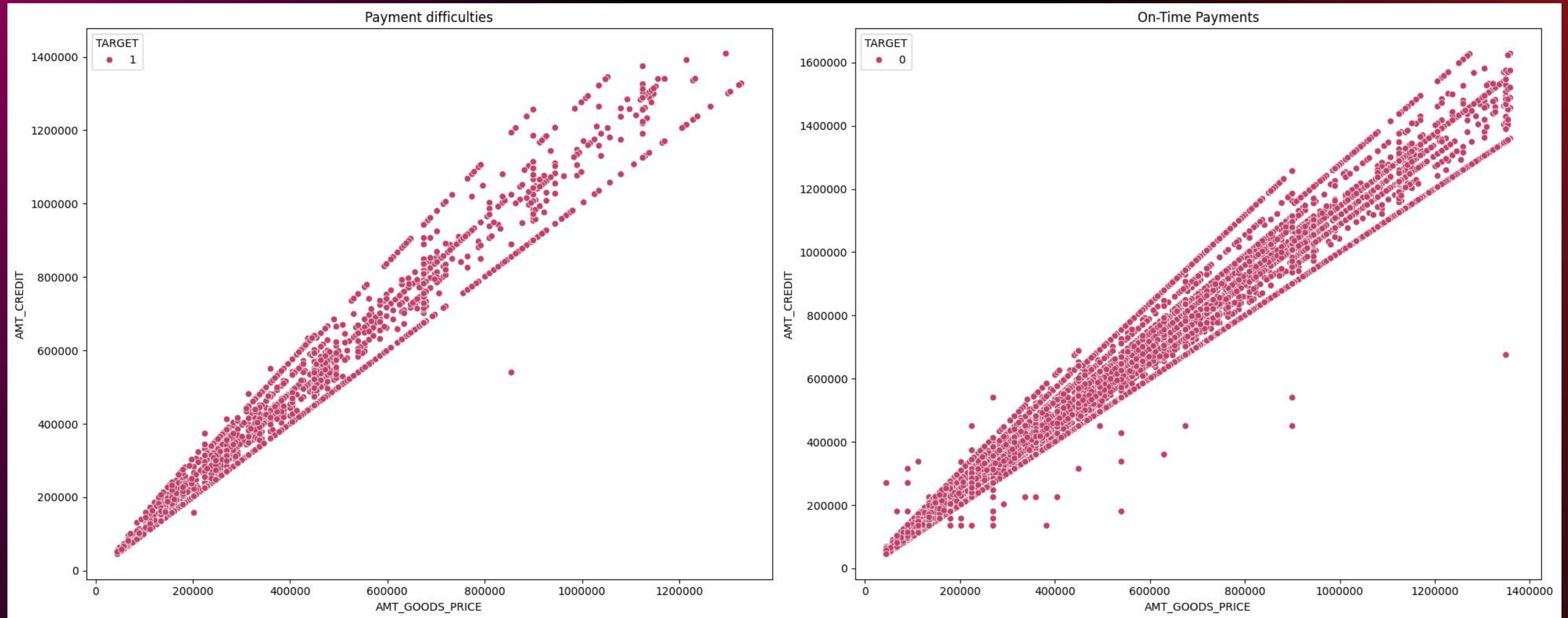
Median of Loan values defaulting for Applicants with Academic degree is higher and Clients with academic degrees should get higher preference

# Univariate Analysis-Years\_Employment



Clients who are employed for a long time days are making their payments on-time but these category of clients do not exist in Payments difficulties group

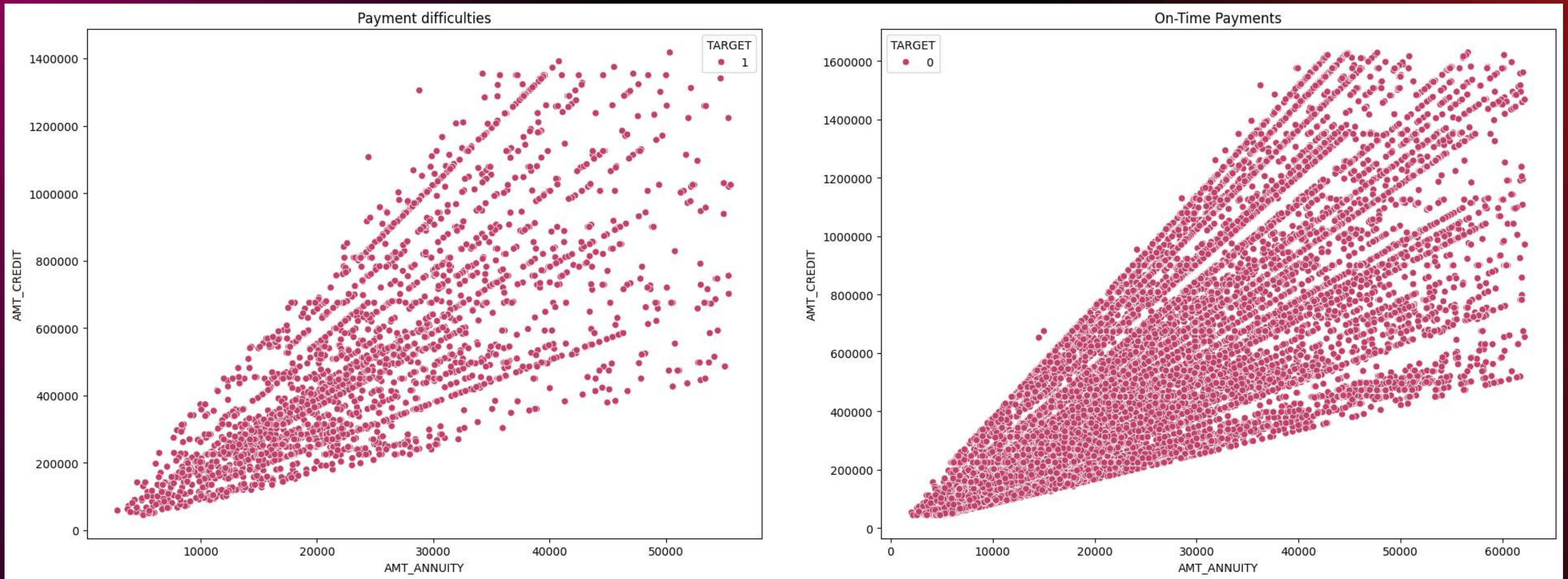
# Bivariate Analysis-Goods & Credit Amount



There is a strong correlation between Goods amount and Credit amount



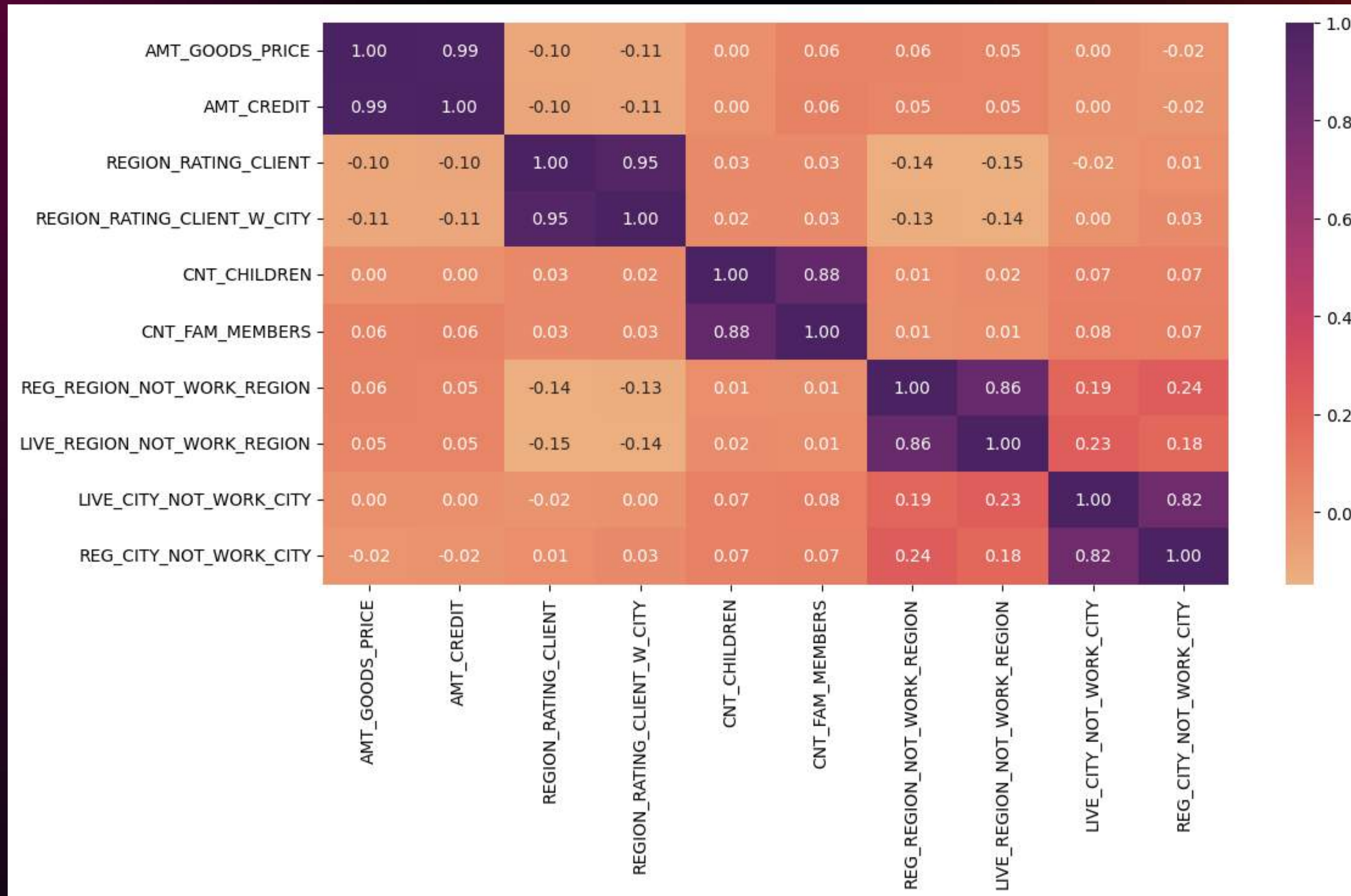
# Bivariate Analysis-Annuity & Credit Amount



There is a strong correlation between Annuity amount and Credit amount



# Top 10 Correlations



# Top 10 Correlations

- AMT\_GOODS\_PRICE AMT\_CREDIT - 0.98
- REGION\_RATING\_CLIENT REGION\_RATING\_CLIENT\_W\_CITY - 0.96
- CNT\_FAM\_MEMBERS CNT\_CHILDREN - 0.89
- DEF\_60\_CNT\_SOCIAL\_CIRCLE DEF\_30\_CNT\_SOCIAL\_CIRCLE - 0.87
- REG\_REGION\_NOT\_WORK\_REGION LIVE\_REGION\_NOT\_WORK\_REGION - 0.85
- LIVE\_CITY\_NOT\_WORK\_CITY REG\_CITY\_NOT\_WORK\_CITY - 0.78
- AMT\_ANNUITY AMT\_GOODS\_PRICE - 0.75
- AMT\_ANNUITY AMT\_CREDIT - 0.75
- Years\_employment FLAG\_DOCUMENT\_6 - 0.62
- Age Years\_employment - 0.58

# Result

**People having payment difficulty are from the following group:**

- Medium income
- 25-35 years old age group, followed by 35-45 years age group
- Male
- Unemployed
- Labourers, Salesman, Drivers
- Don't own house, stay with parents, rented flats
- Have medium or high income

**Click Here - [Google colab notebook of Bank Loan Case Study](#)**



# Result

## People to focus on:

- Students & Businessmen have no problem in repayment of the loan
- Repeater clients
- Academic degree clients
- Clients who are employed for more than 19 years
- Clients who are Married
- Clients who own a house
- Pentioners
- Female applicants should be given extra weightage as defaults are lesser.

**Click Here - [Google colab notebook of Bank Loan Case Study](#)**

**THANK YOU**