Credit EDA Case Study

By:

Ritesh Reddy G
Prahalad RK

Problem Statement

▶ Banks/ Loan providing companies face a real challenge while approving or rejecting a loan application.

▶ When a company received a loan application, it has to decide to approve or reject the loan based on the applicant's profile.

▶ Two types of risks are associated with the bank's decision:

 ▶ Loss of business when loan is not approved and applicant is likely to repay the loan

 ▶ If applicant is likely to default, then the bank may incur a financial loss once the loan is approved.

Types of Decisions

- On receiving a loan application, the below four types of decisions could be possible from client or company's side.

- Types of Decisions:
  - Approved: When the loan lending company approves the loan.
  - Cancelled: Client cancelling the loan application during loan processing.
  - Refused: Company rejects the loan as requirements are not met.
  - Unused Offer: Client refuses to utilise the loan even after receiving approval.

Analysis for Target column

Inference from the plot for the target variable:

- The target variable has two categories.
    1 - Client with payment difficulties &
    0 - All other cases.

- We can clearly see that there's a major **imbalance** between both the categories.

- 8% people have had difficulties in repaying the loan while 92% people can be considered as majority of people repaying the loan without difficulties.
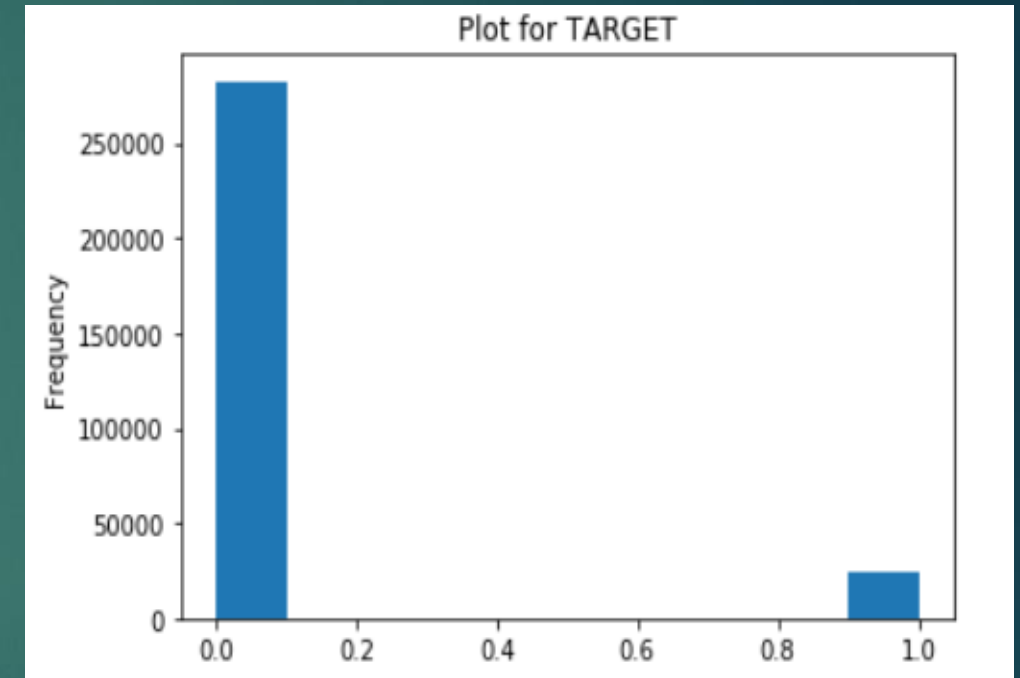
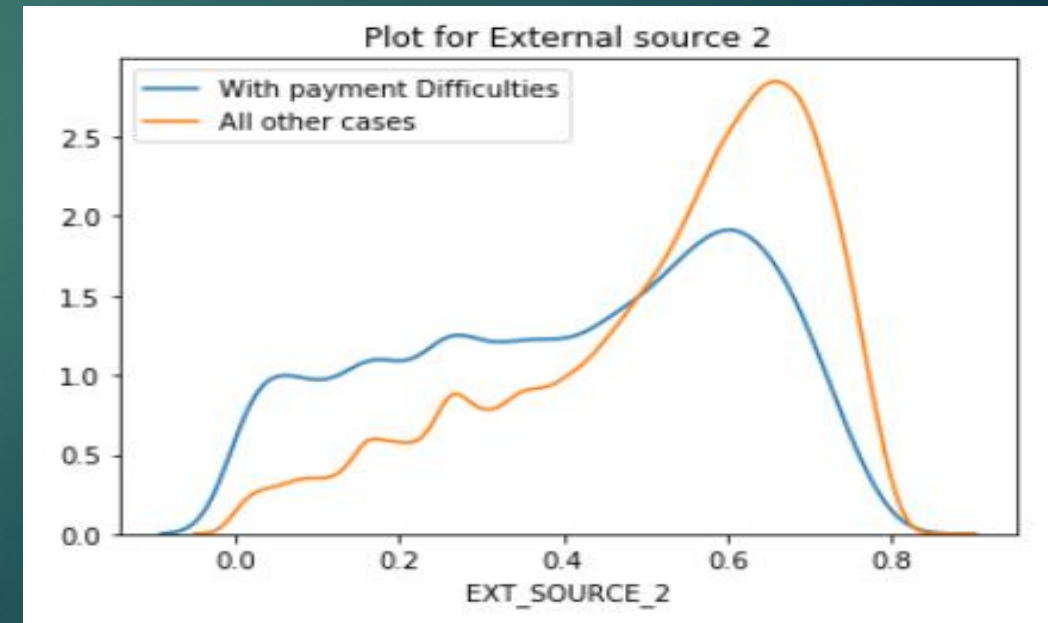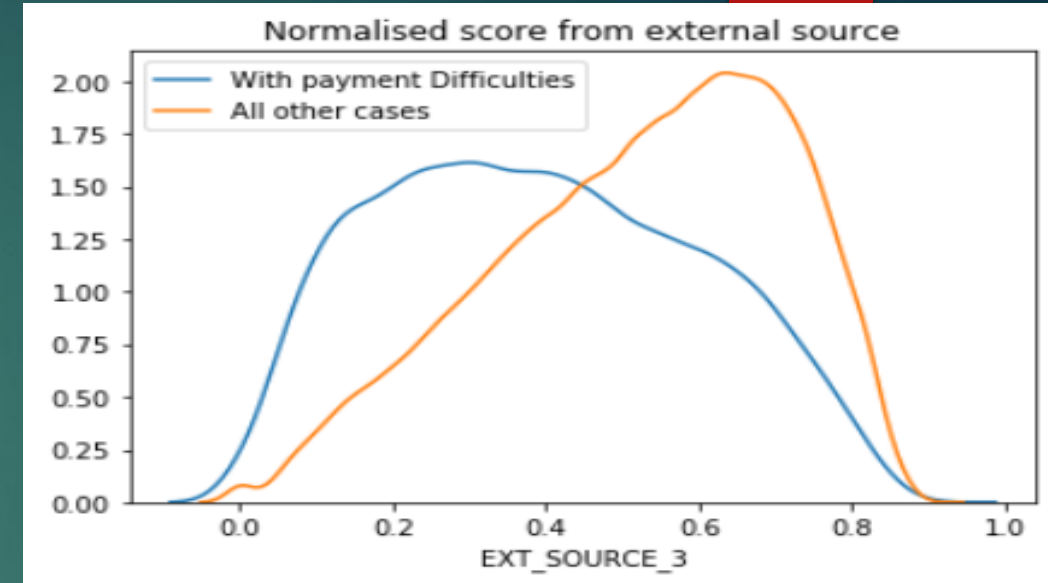# Distribution for Clients age

Inference from the plot:

- Clients under the age range 20-30 are more likely to face payment issues more compared to other ages.

- Clients having age greater than 50 are likely not to face any payment issues. Hence less chance of defaulting.

- We can see that as age increases the tendency to face payment issues drops as shown by the plot.

AGE of the client

Influence of External Sources on Target

Points we can infer from the plots:

- Clients with external source 3 values less than ~0.45, the possibility of client facing payment difficulties is more compared to other cases.

- On the other hand, external source3 values greater than ~0.45 the client is most probable of not facing payment difficulties

- For external source 2, (0-0.4) value has a trend of client facing payment difficulties more while value > ~ 0.5 shows other cases are more.
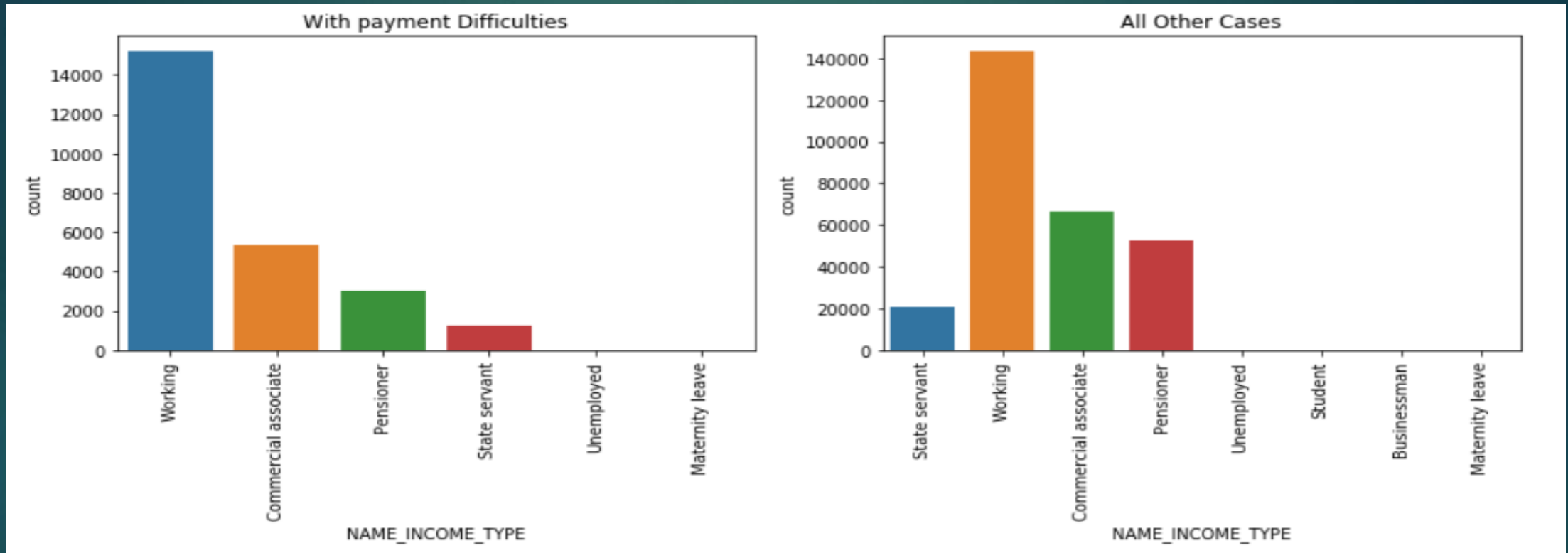
Loan type distribution
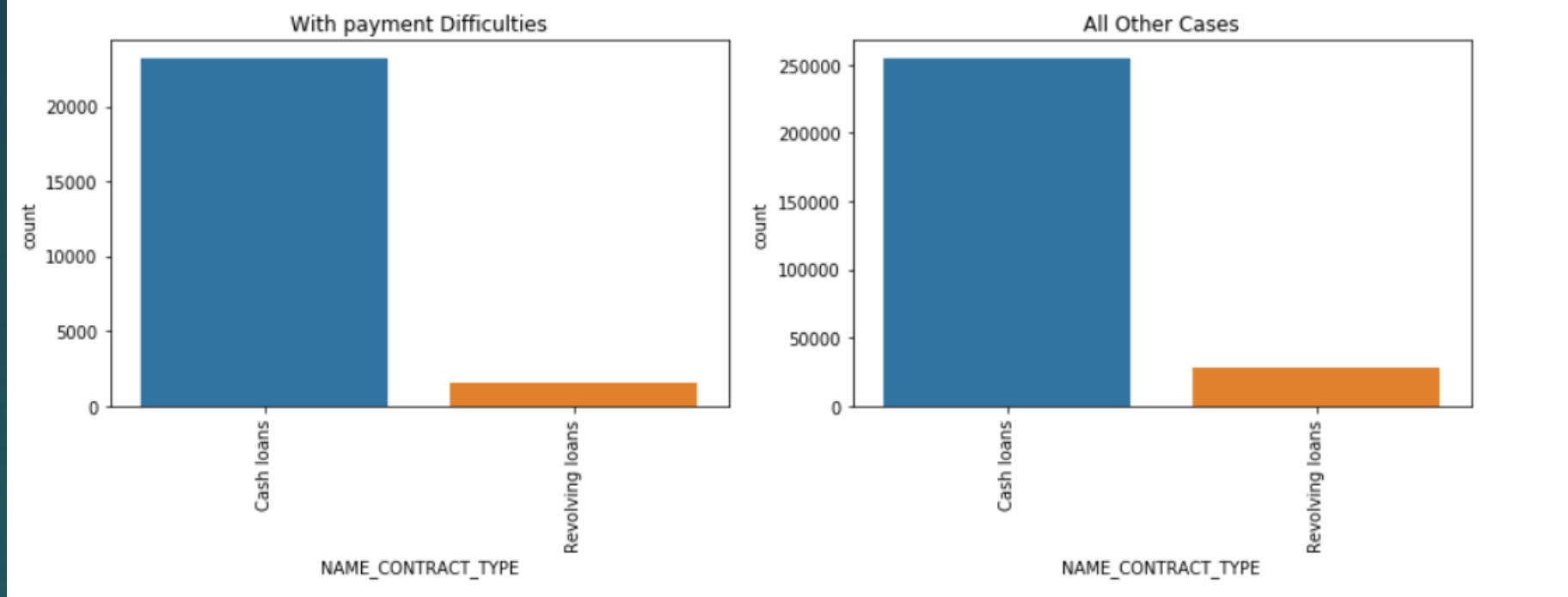
Inference from the above graphs:

Most of the client are preferring Cash loans to Revolving loans but the loan type doesn't show any relation with clients payment difficulties.
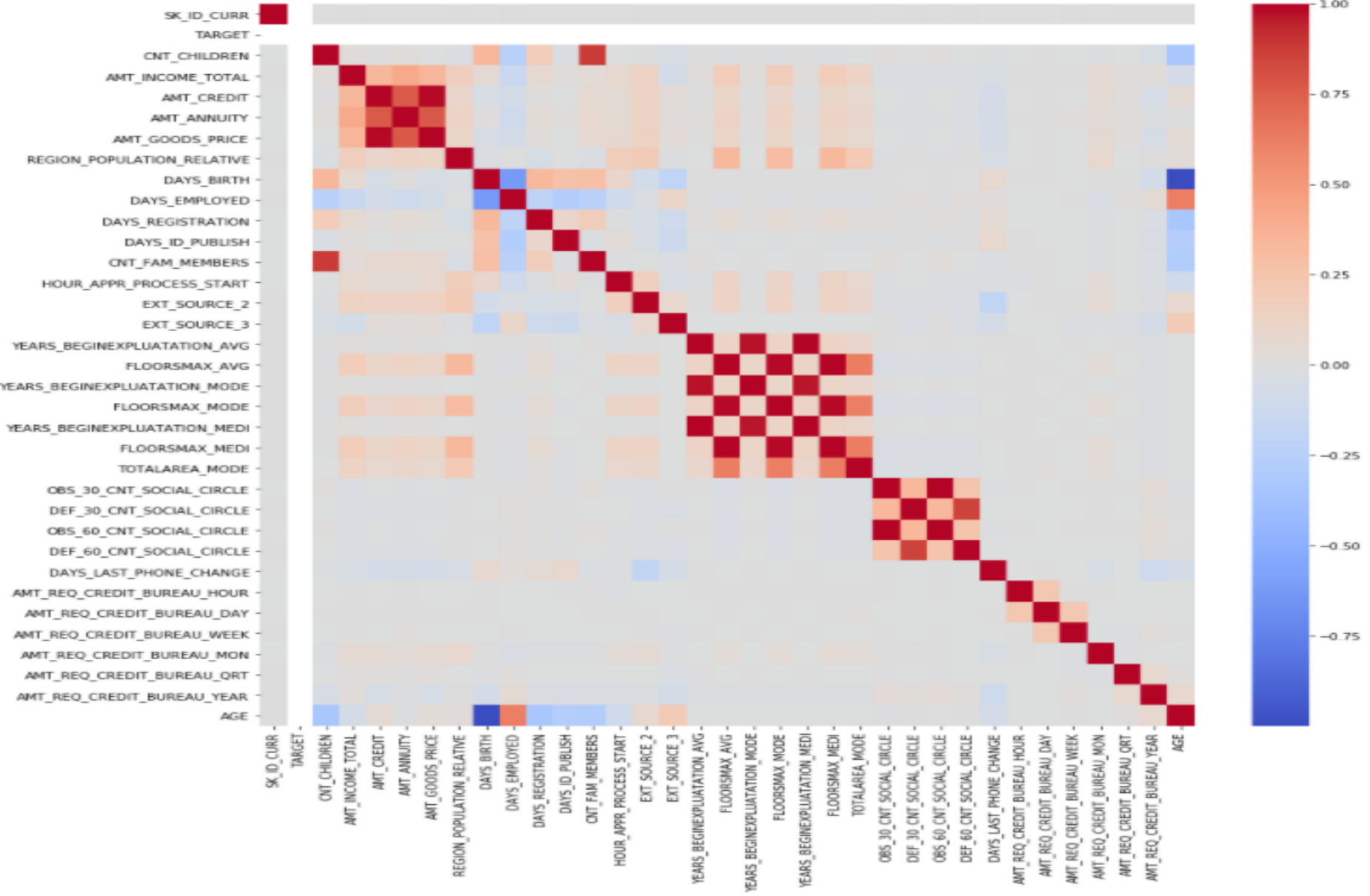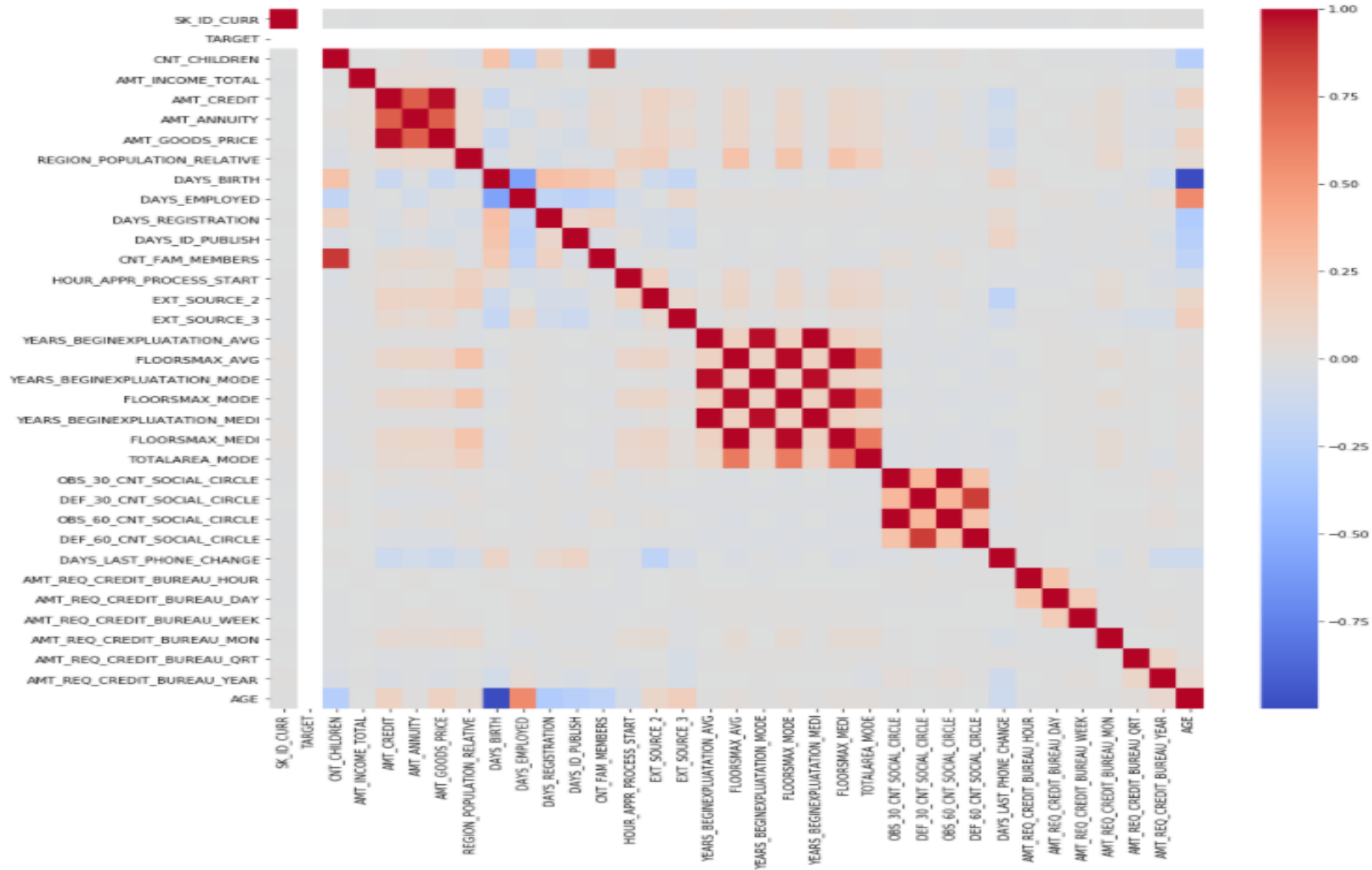
Distribution for client's income type

From this plot , we can observe the following:

- The working clients have the high frequency of applying loans followed by commercial associate and pensioner.

- The least income types are state servant, unemployed and Maternity leave.

Correlation for Target 0

Correlation for Target 1

Inferences from correlation heat maps

- Income of the client has shown high correlation with credit amount of the loan and the price of the goods for which loan is taken.

- As client's age increases, the probability of client facing payment issues decreases.

- External sources have shown high correlation with target variable.

- All the correlating pairs are almost same for client facing difficulties and clients with all other cases.

Highly correlated features with target variable

Features that show highly negative correlation

    External Source 3

    External Source 2

    Age of the client

    Number of days the client was employed before he applied for loan

    FLOORSMAX AVG

Features that show highly positive correlation

    Days Registration

      (Days before the application client change his registration)

    DAYS_ID_PUBLISH

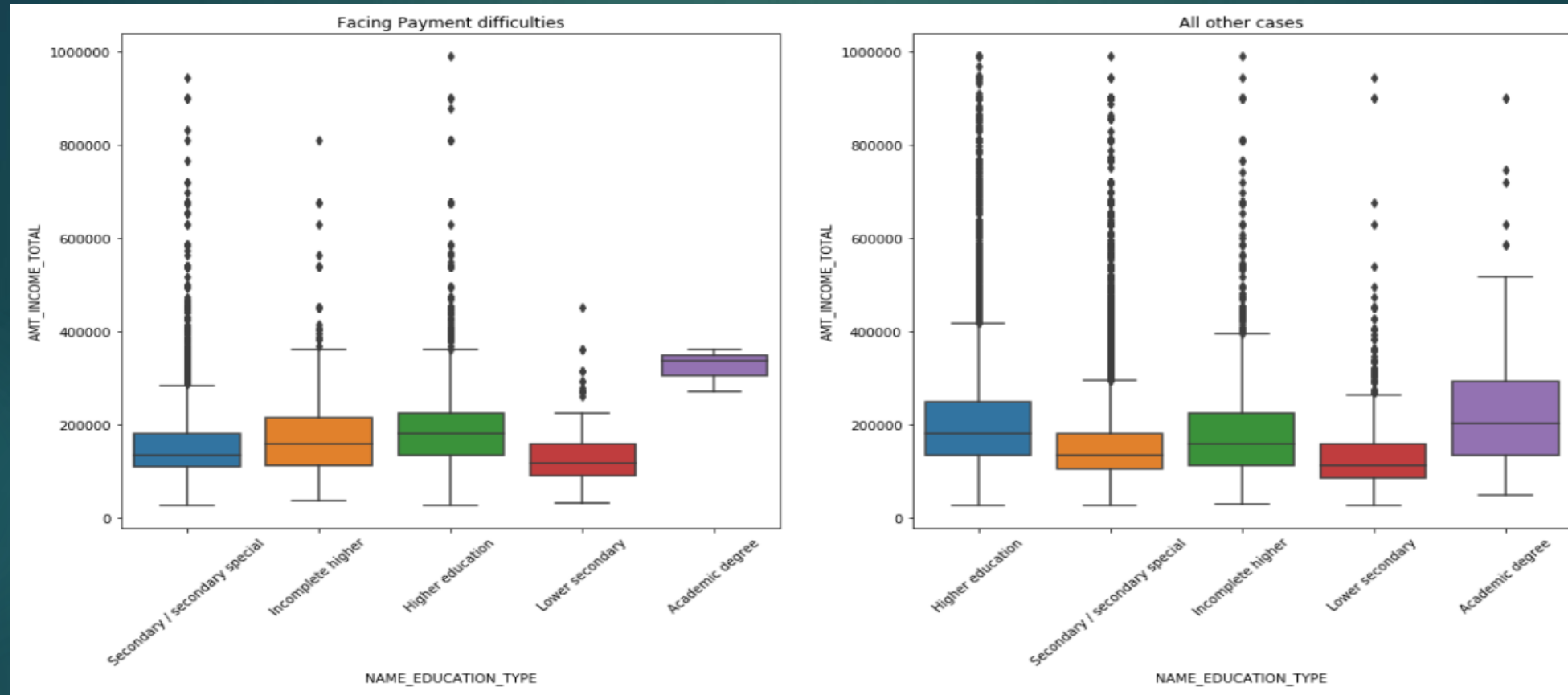   (Days before the application client changed the ID with which he applied for the loan)

    DAYS_LAST_PHONE_CHANGE

      (Days before the application when client changed the phone)
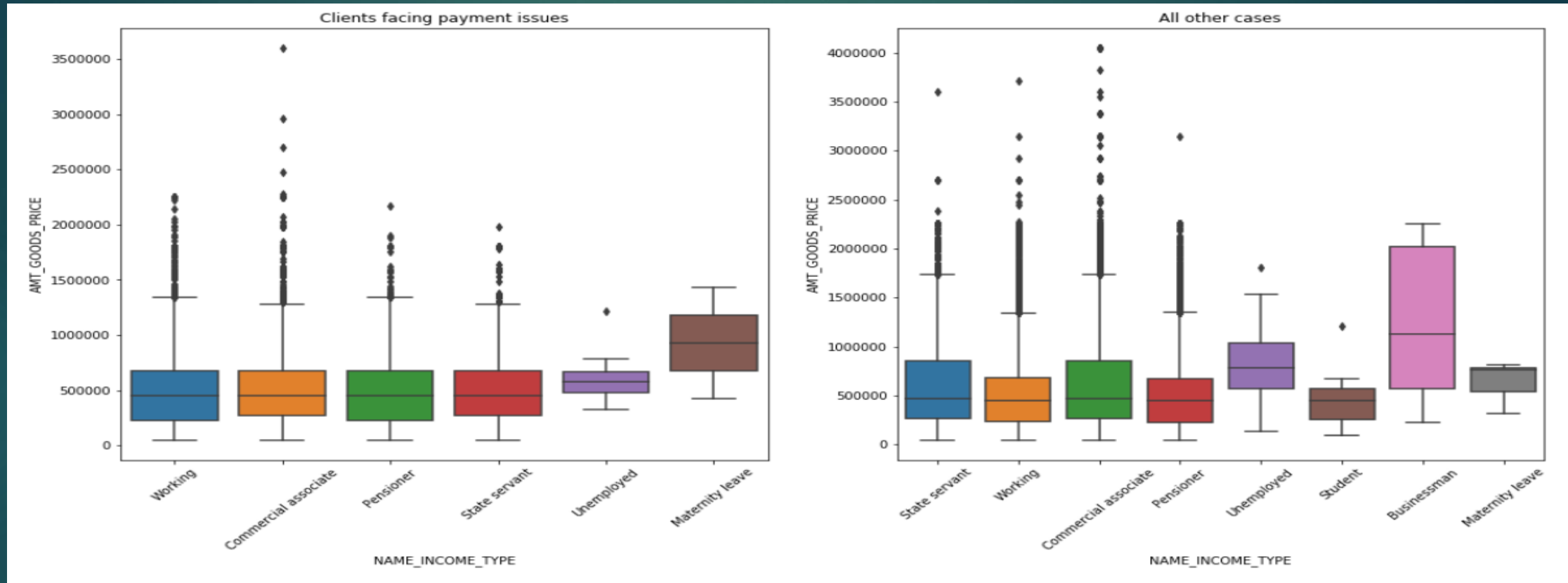
    DAYS_BIRTH

      (Client's age in days)

# Observation from Education type of the client

Academic degree has a wider variance with income of the client in the case of "all other cases " as compared to "Client facing payment difficulties".

Inference from Income type of the client

From the plot, we can infer that Businessmen will not face any payment difficulties compared to all other income types.

# Analysis of Age group of the client with client owning a flat or not
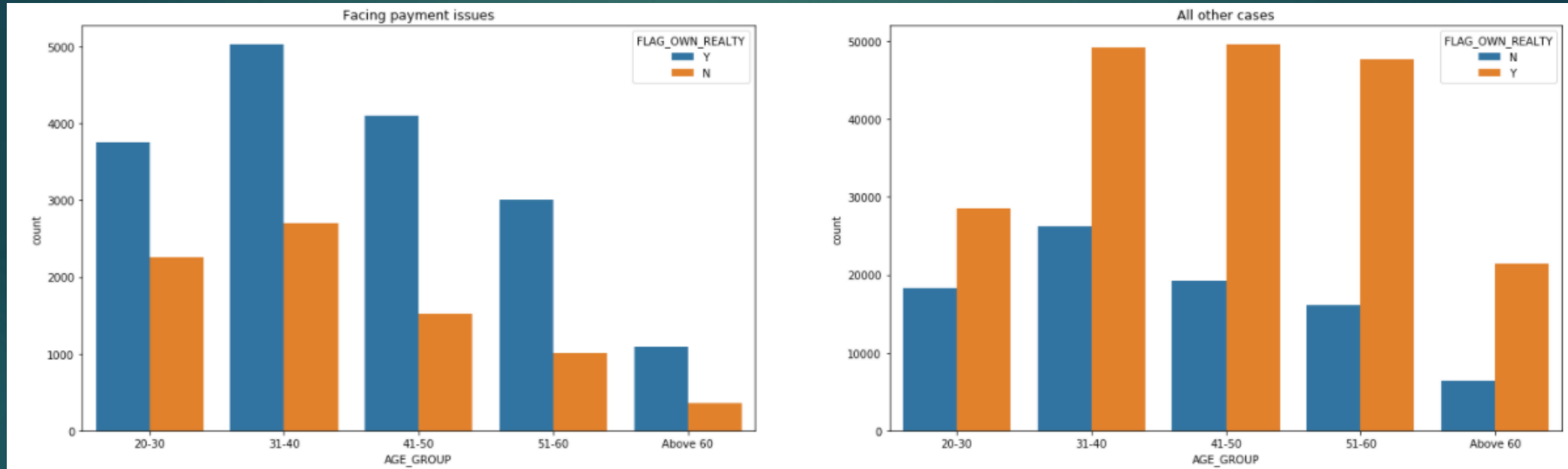
- ▶ Clients Under age group 20-30 , owning a realty flat are more likely to face payment issues.
- ▶ Clients in the age group 40 and above owning a realty are most likely not to face any payment issues.
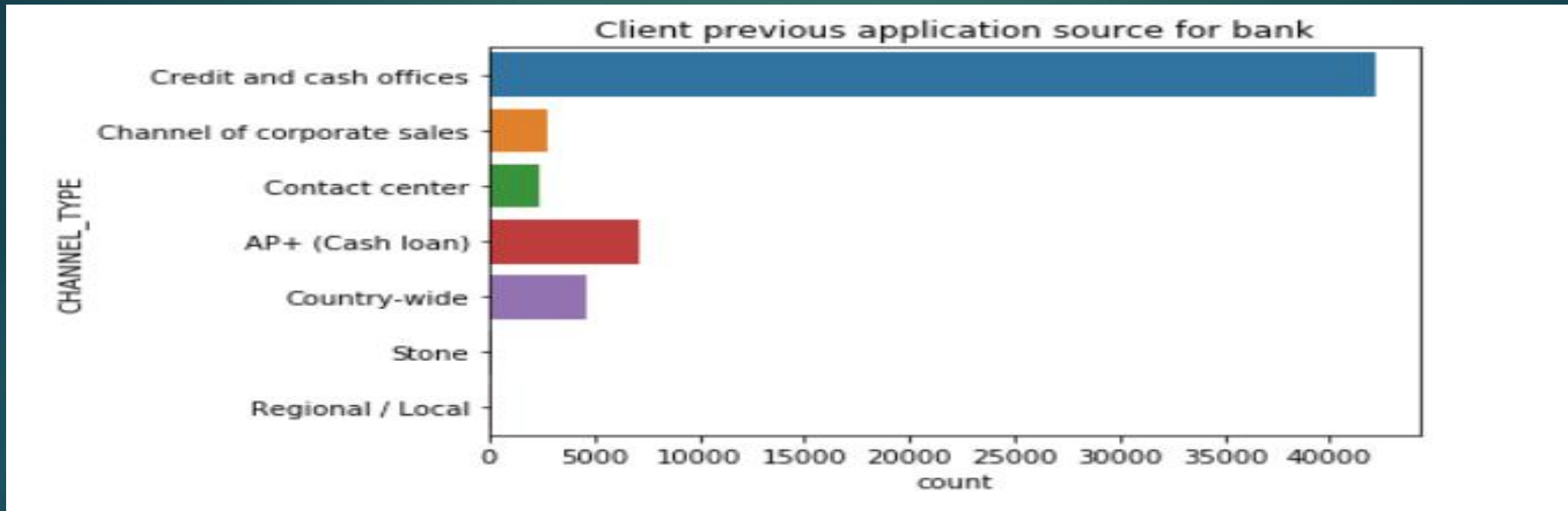
Status of client's previous application
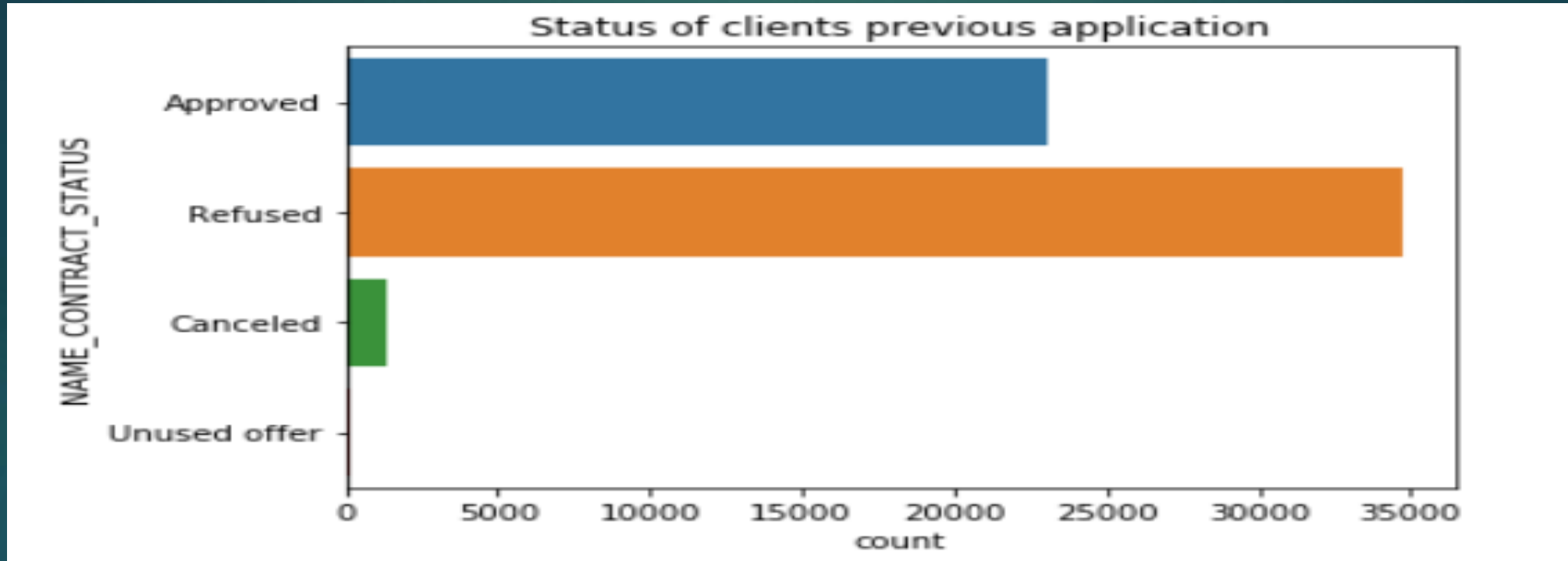
From the above graph, we can infer that there are a lot of clients who have their previous applications refused from the bank.

Reasons of Refusal by the bank


Reasons for loan refusal by bank
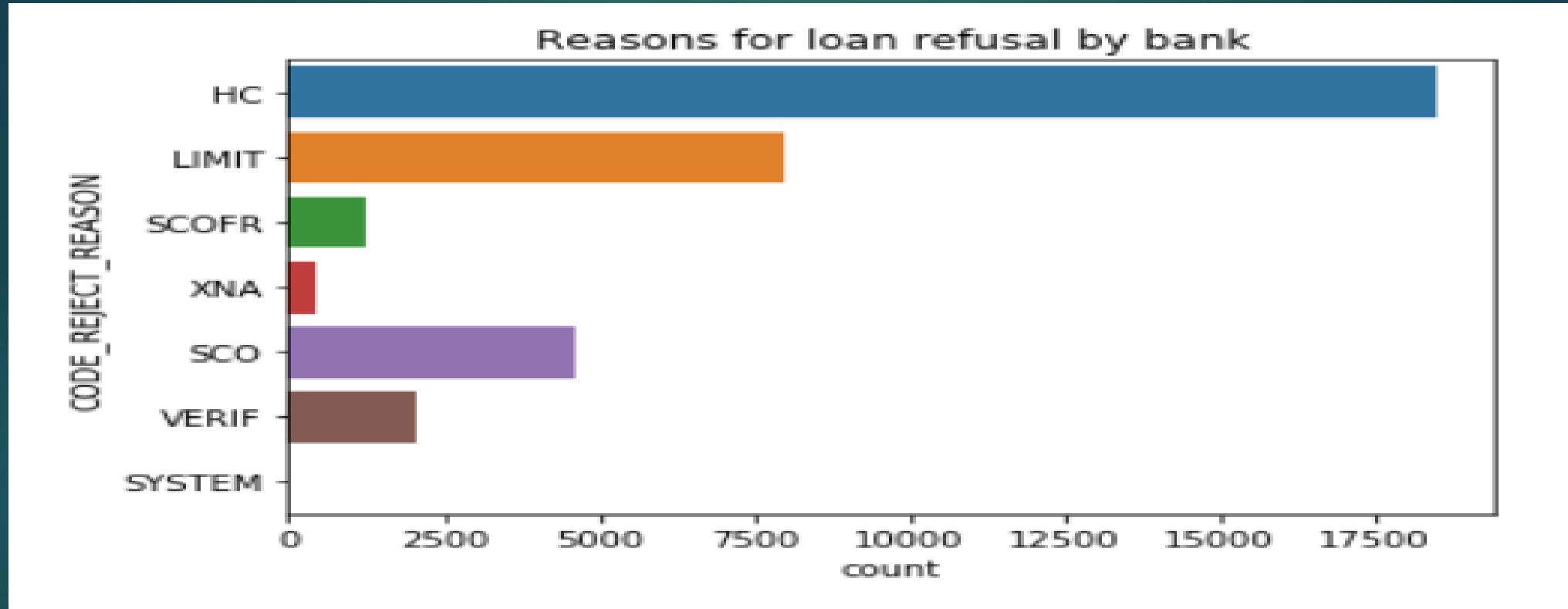
The insight we can derive from this graph is that bank/loan providers reject the client's application if it is "HC" followed by the client's credit limit.

Client's Source for the bank

Client previous application source for bank

Credit and cash offices are the major channels followed by AP+ for clients previous application for loan.
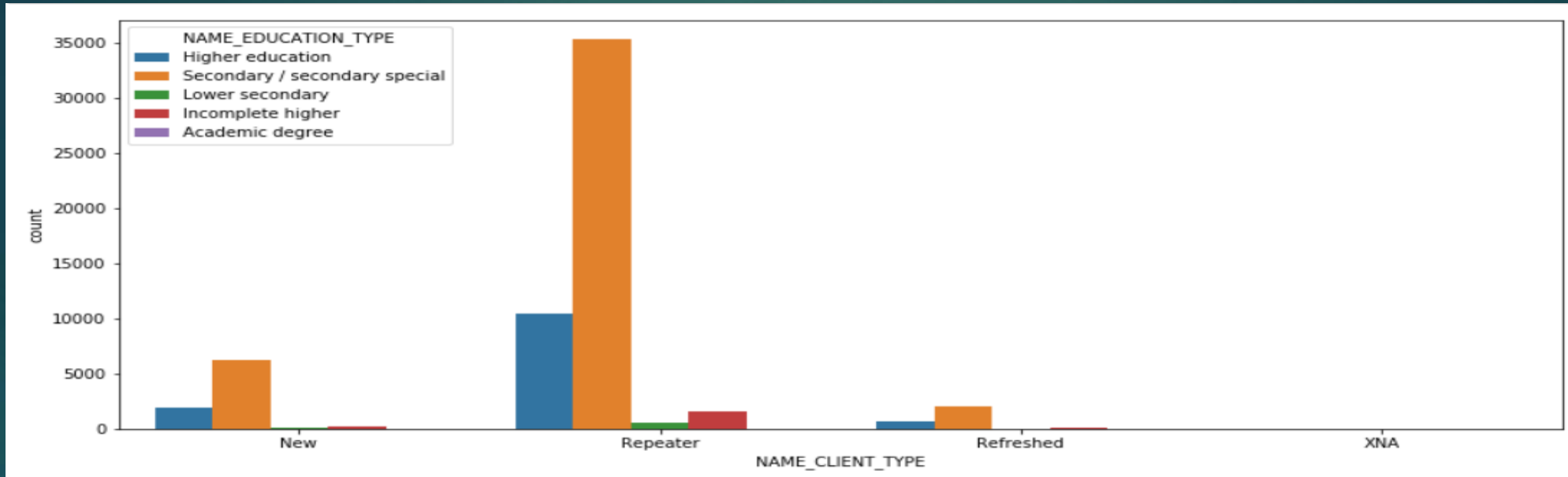
Occupation type vs Loan Application
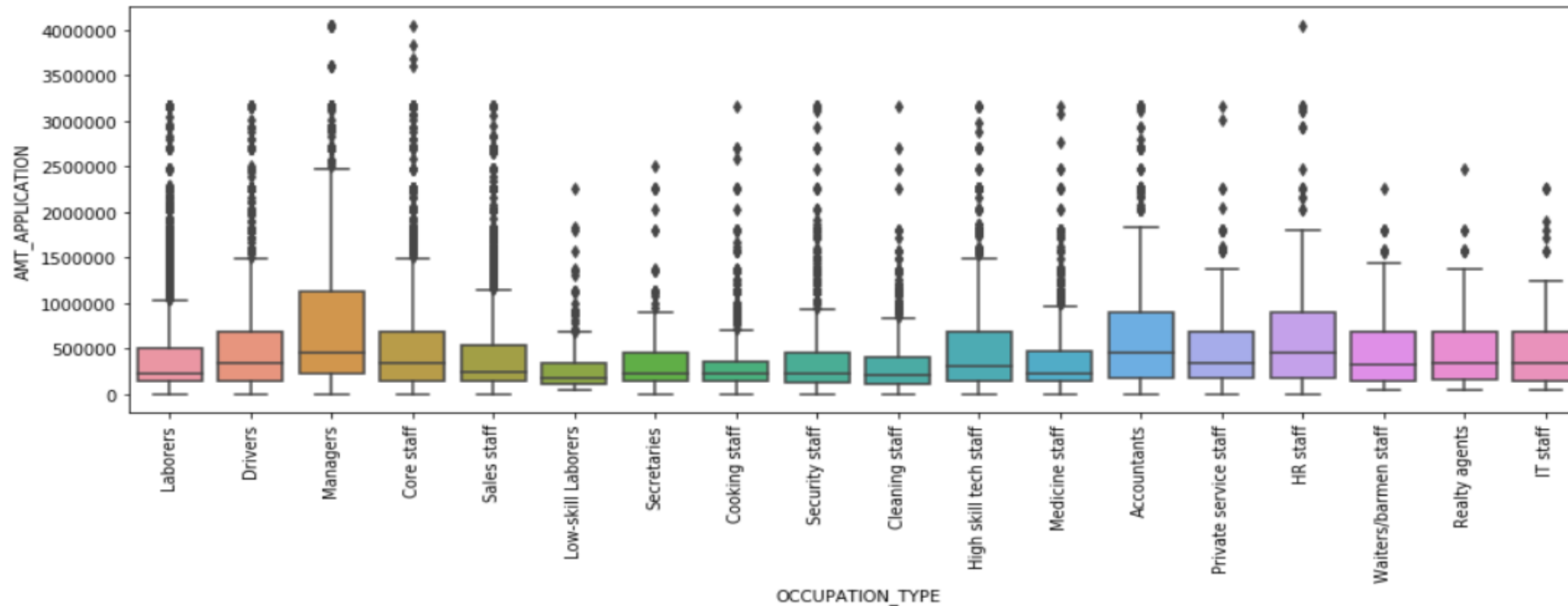
Managers have requested for a loan amount of more than 10Lacs followed by accounts.
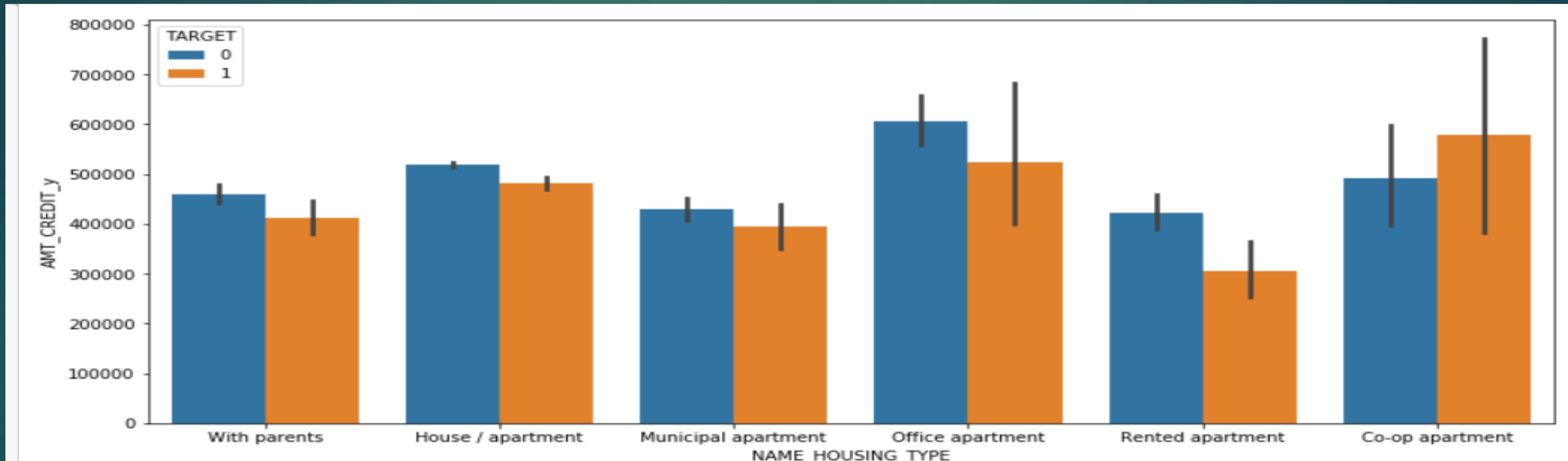
Low-skilled labourers and cooking staff have requested for very low amount of loan in their previous applications.

Type of client vs Client Education

Clients with secondary education are more likely to return to the bank for applying loan, followed by New clients.

Distribution of Housing type of the client

Clients living in Co-op apartment have high chance of having payment difficulties compared to people living in other housing types

Distribution of Client type vs payment difficulties
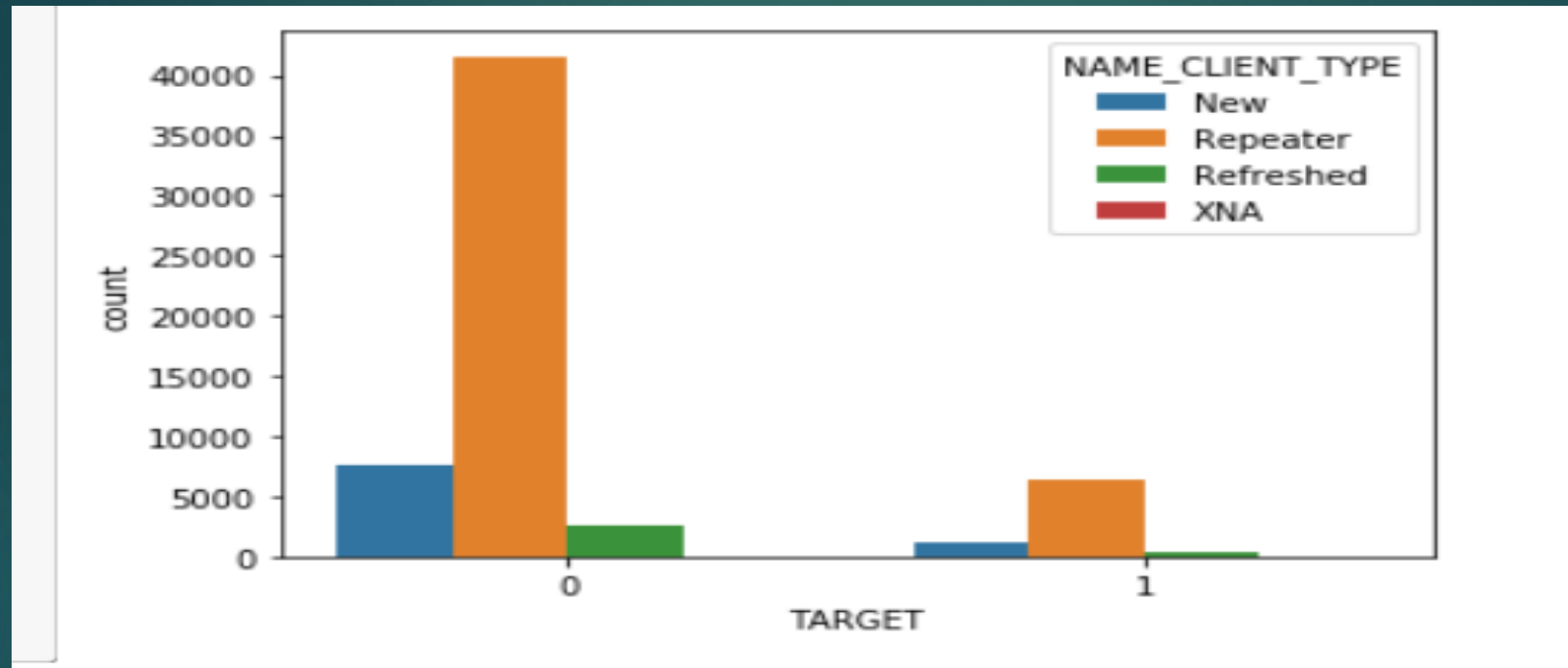
We can infer that the repeating clients are less likely to face payment difficulties compared to others.

Conclusions from the study

- As the clients age increases the tendency to face payment issues drops significantly. Hence banks can prefer to choose clients with age above 45 years to make a good business.

- External sources have shown a good correlation with target variable. So while approving to a client, banks can verify the clients external source score and make the decision.

- Clients with secondary education are more likely to return to the bank for applying loan.

- Banks can approach the client who are Businessmen as they are likely to repay the loan and should not reconsider other while approving other income type.

- Clients living in Co-op apartment have high chance of having payment difficulties compared to people living in other housing types.

Thank you