CREDIT EDA CASE STUDY

Arjun Pujari

Problem Statement

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse 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 Objective

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to 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 utilise this knowledge for its portfolio and risk assessment.

Analysis Approach

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Univariate Analysis

Analysing each column, plotting the distributions of each column.

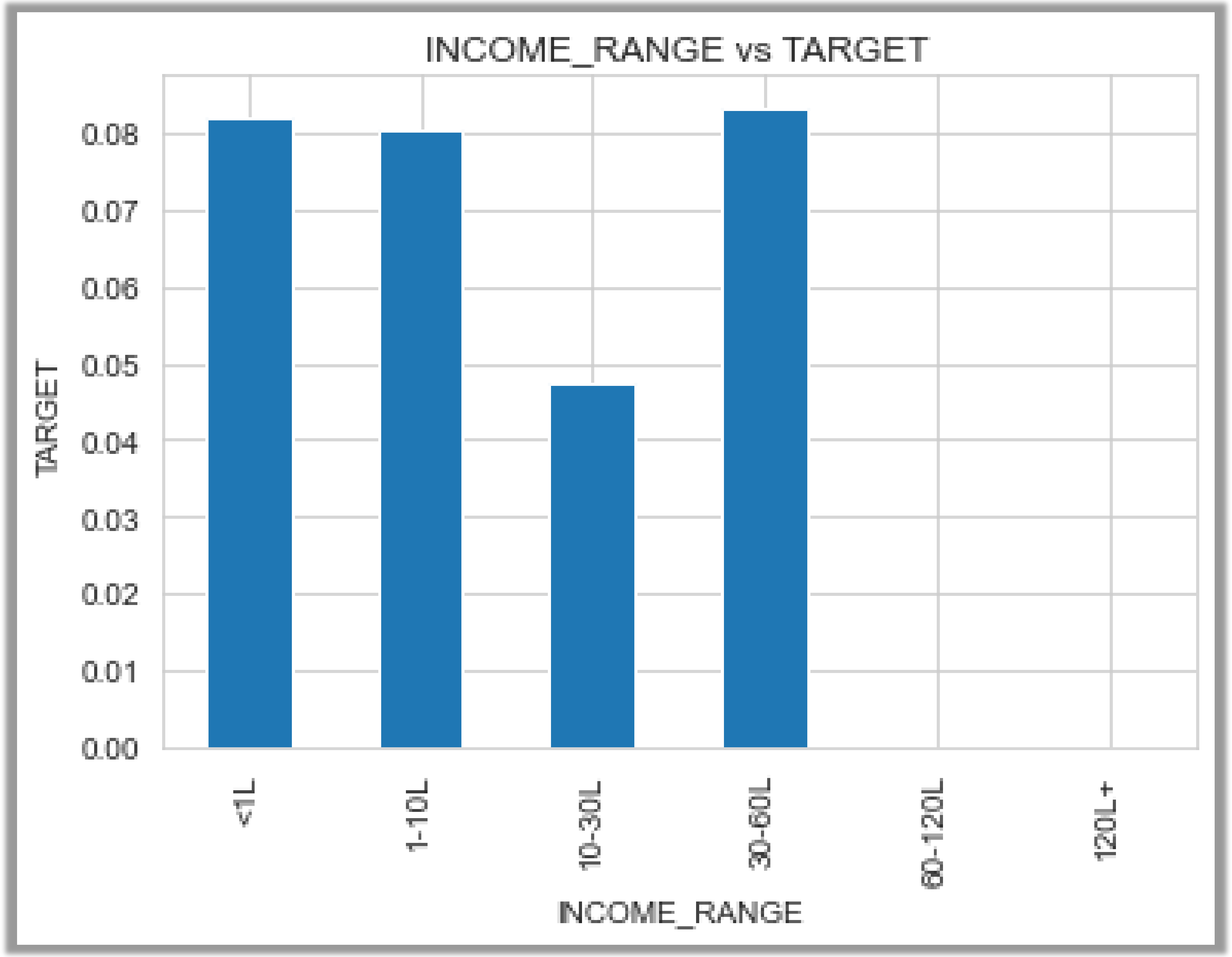
Bivariate Analysis

Analysing the each variable behaviour with respect to target variable

Recommendations

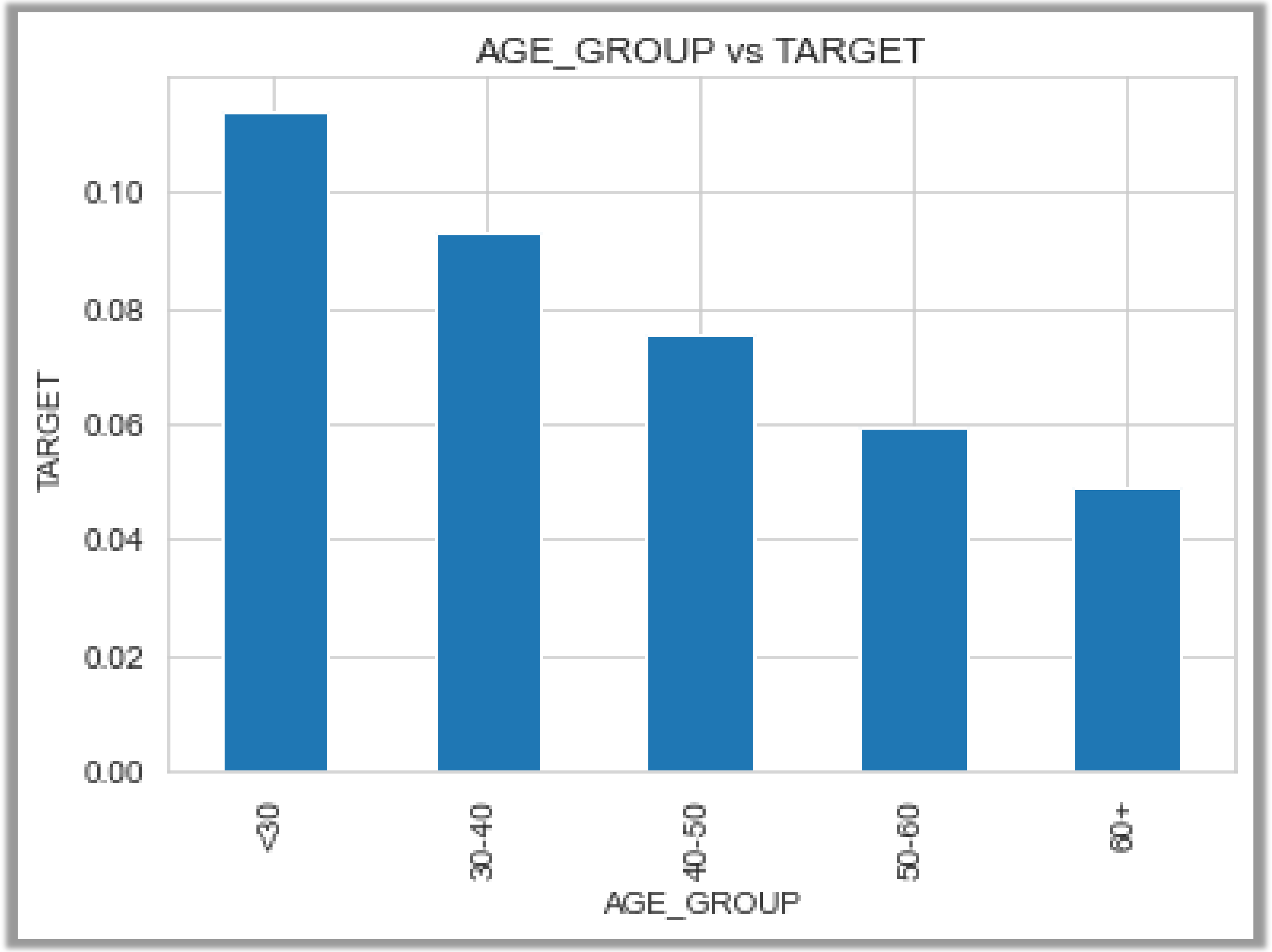
Analysing all plots and recommendations for reducing the loss of Business or Risks by detecting columns best which contribute to loan defaulters.

Analysis - Income Group & Target



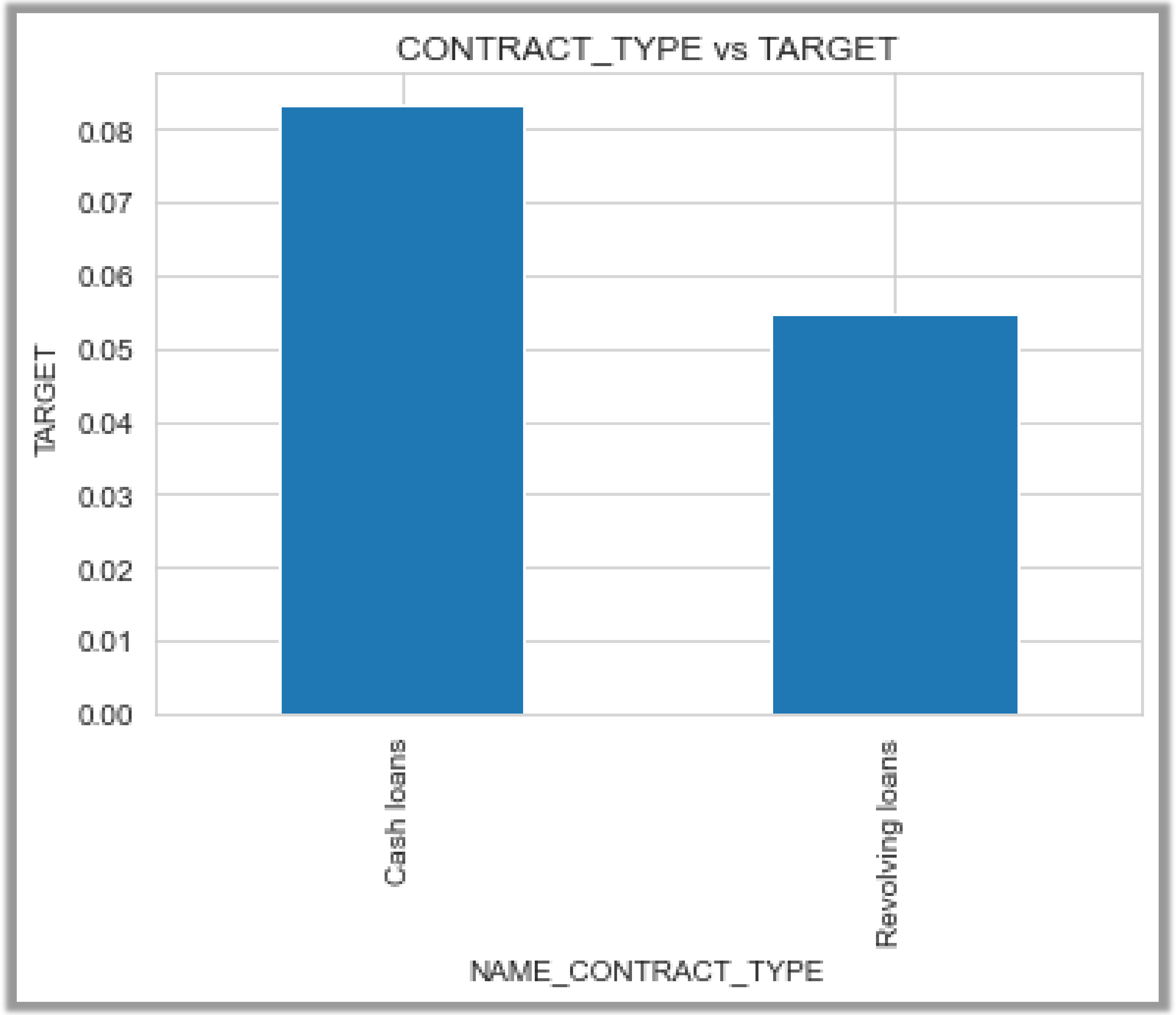
Observation: Clients with income 60 lakhs and above are having less difficulty in payment

Analysis – Age Group & Target



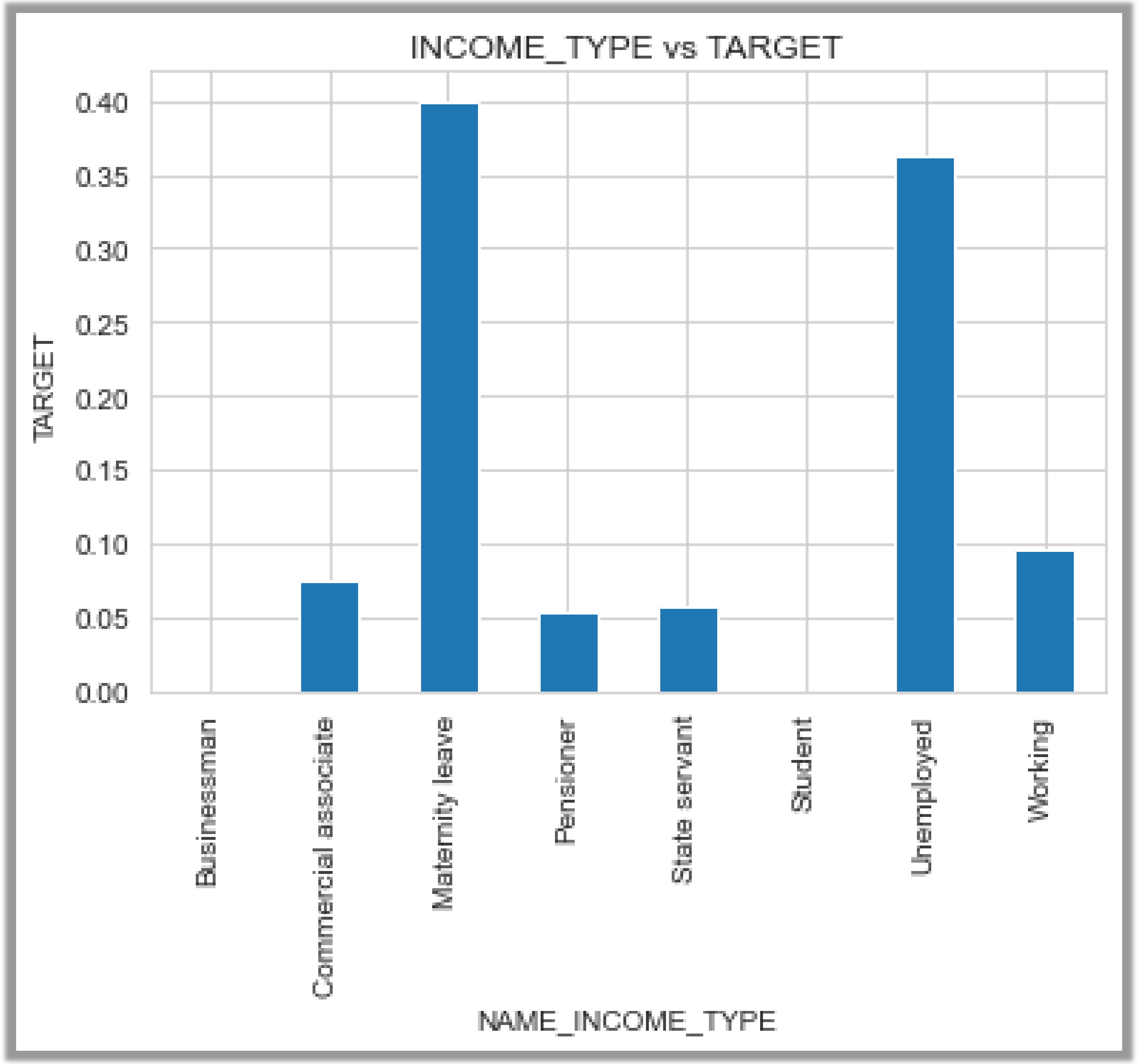
Observation: Clients with age less than 30 years seems to have more difficulty in payment

Analysis – Contract type & Target



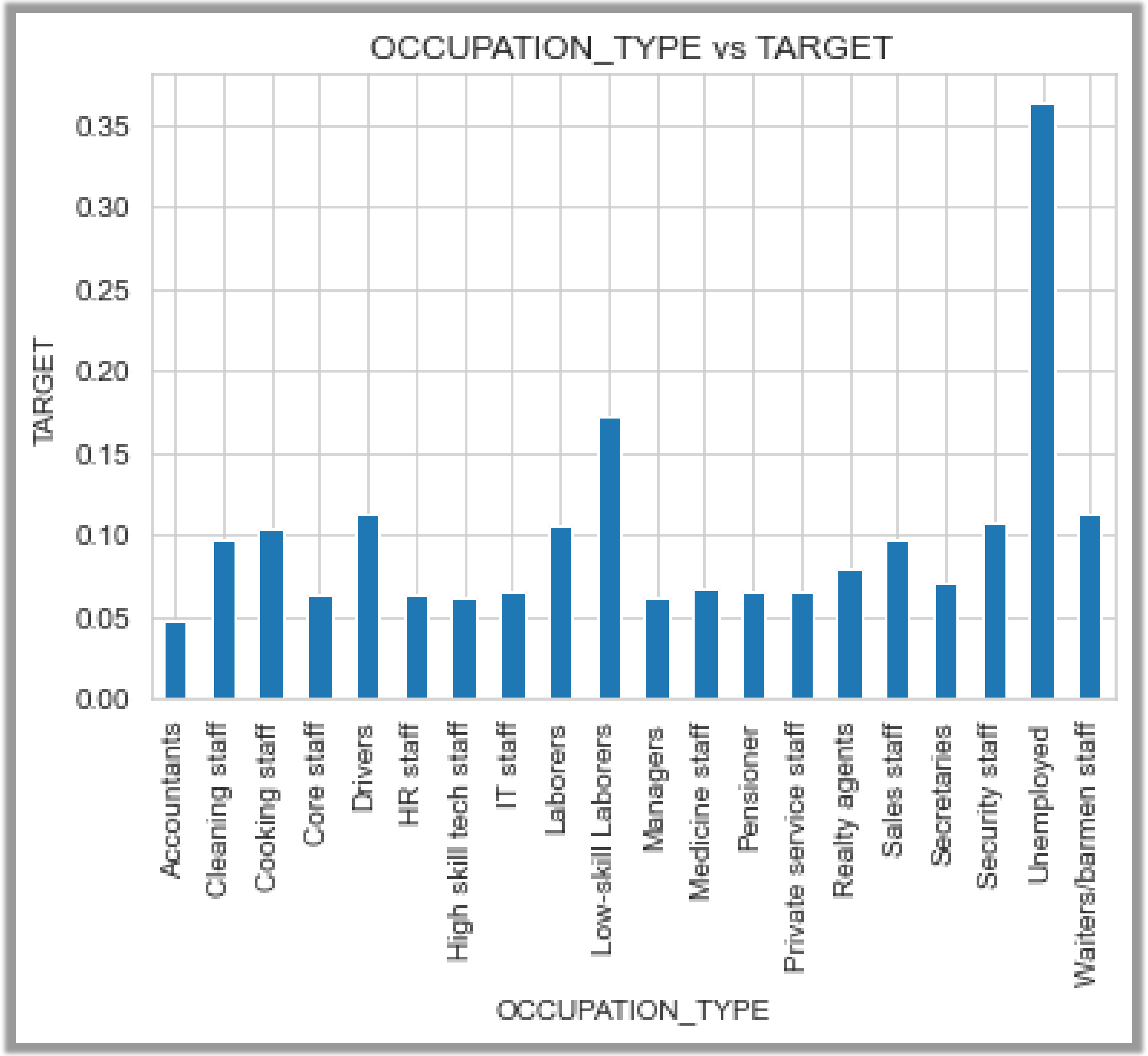
Observation: Clients with cash loans have more difficulty in payment

Analysis – Income & Target



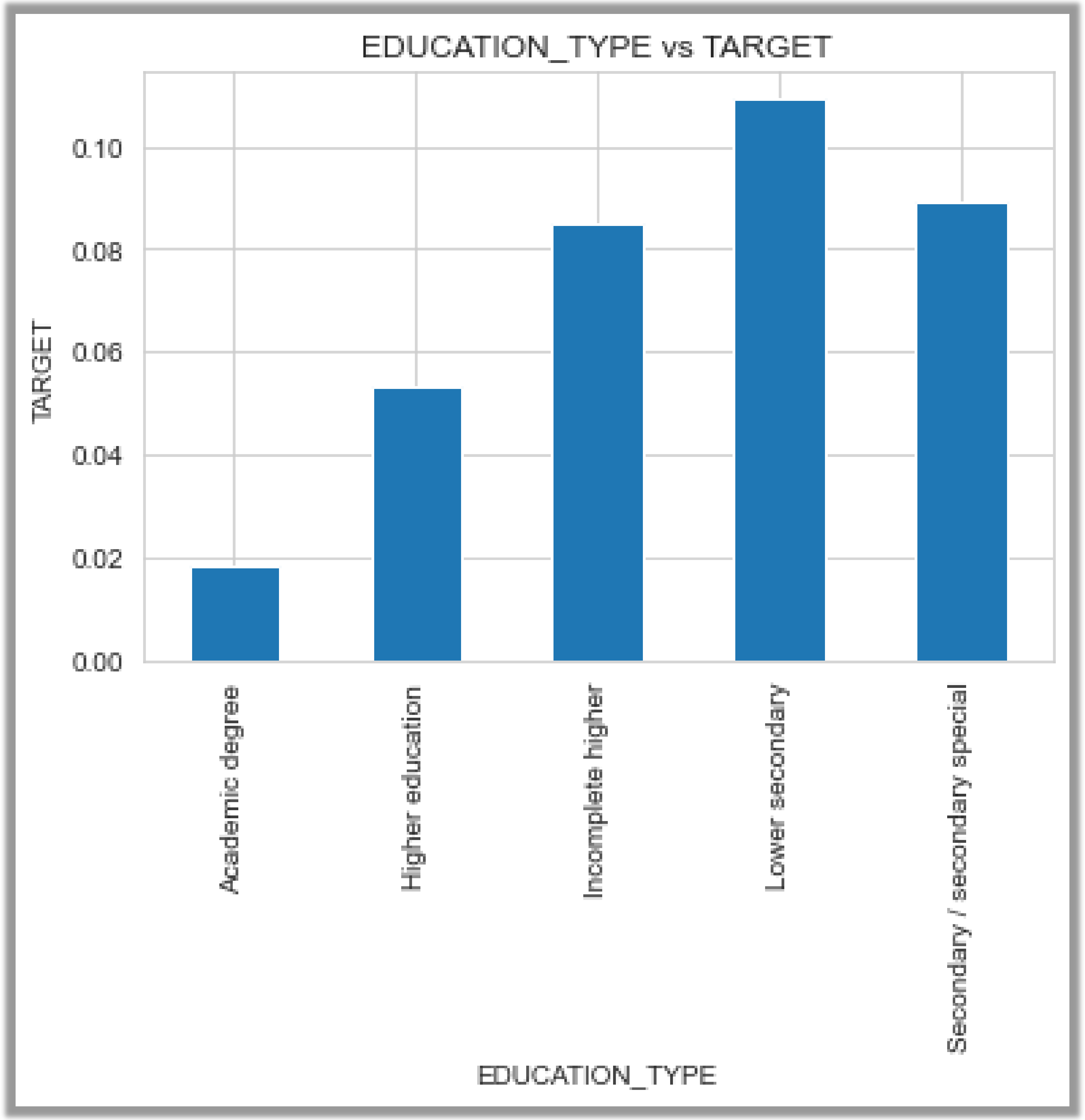
Observation: Clients unemployed or on Maternity leave have more difficulty in payment

Analysis – Occupation & Target



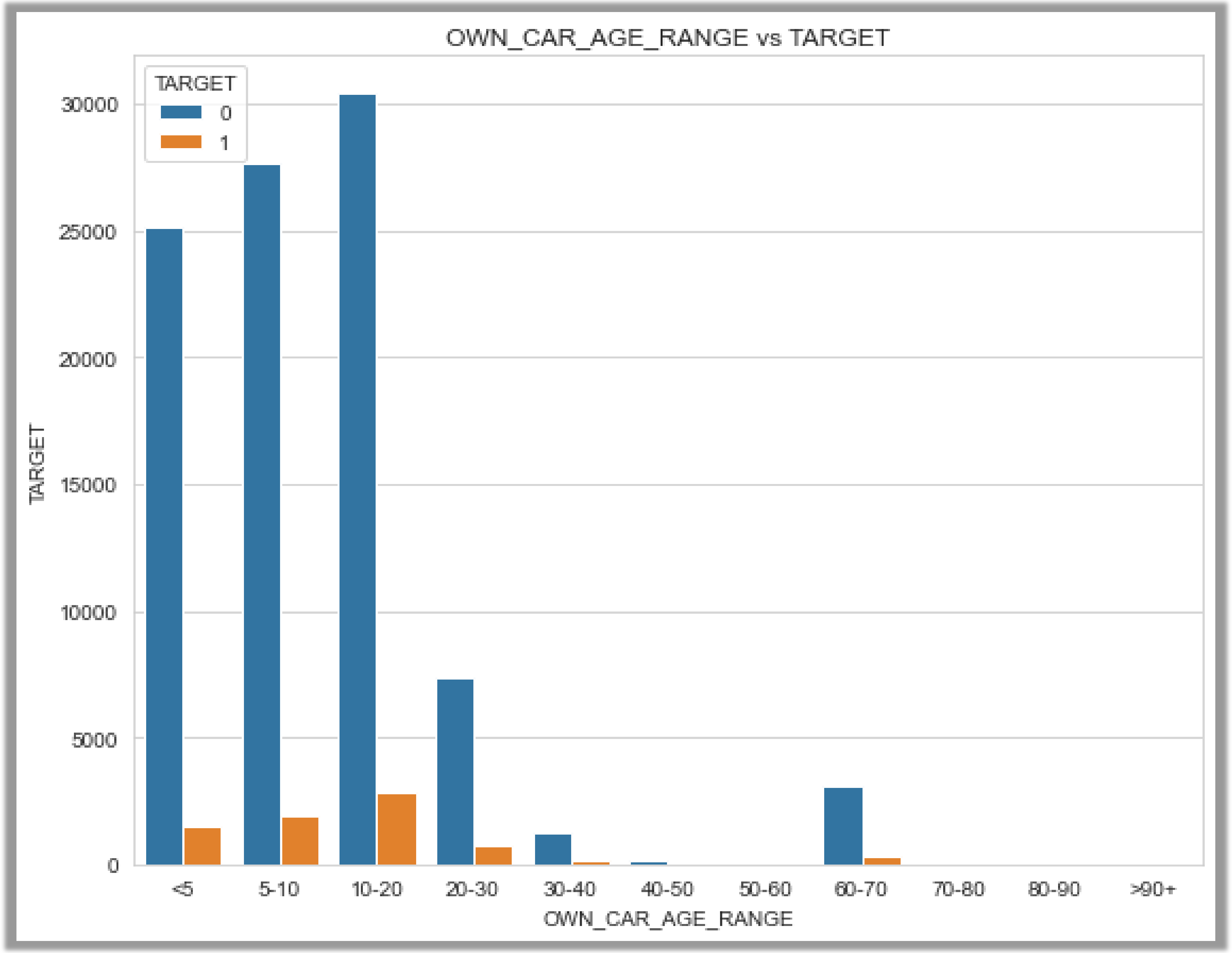
Observation: Clients unemployed or Low Skilled labourers have more difficulty in payment

Analysis – Education & Target



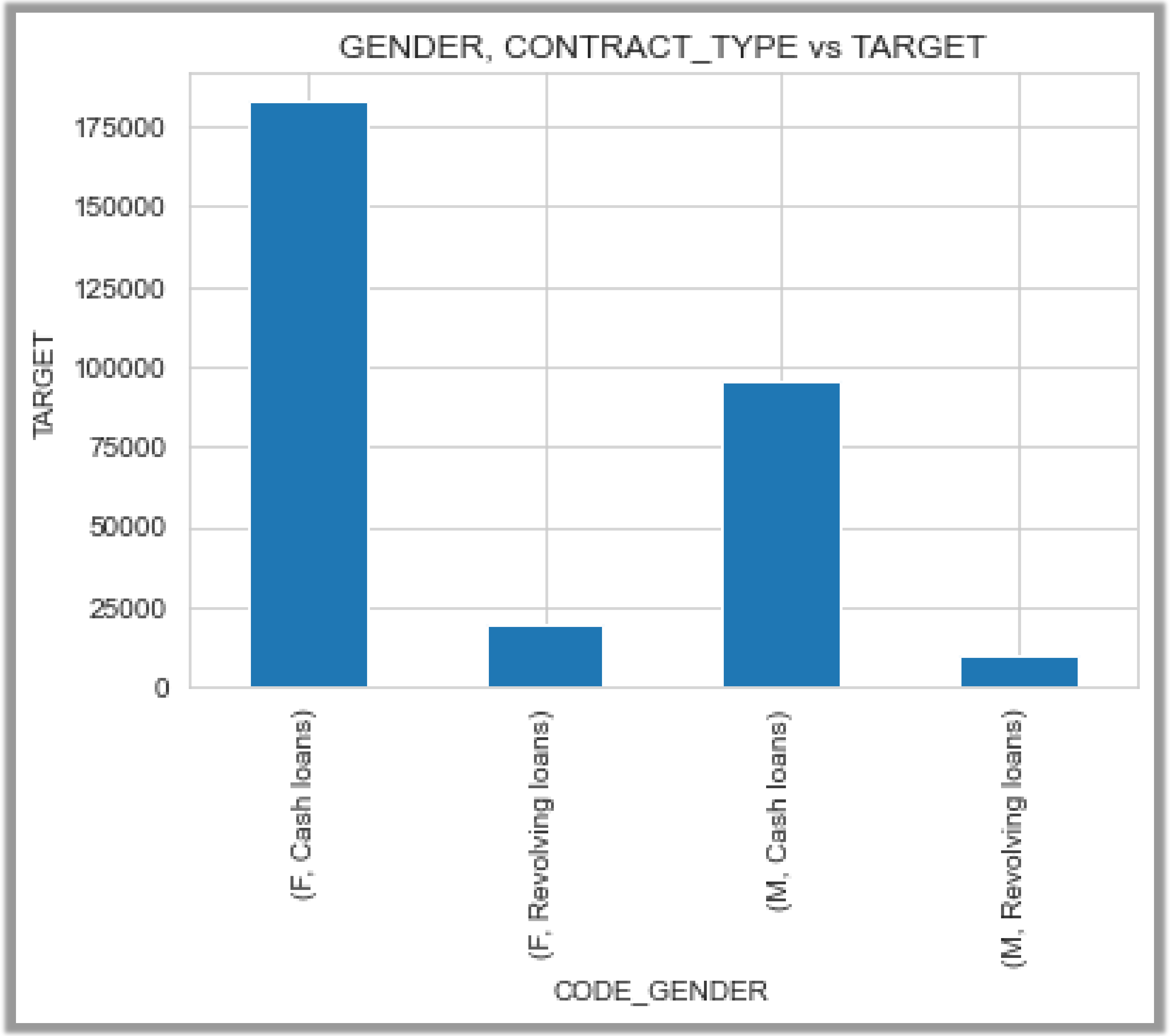
Observation: Clients with education level Lower Secondary has Difficulty in payment

Analysis – Own Car Age Group & Target



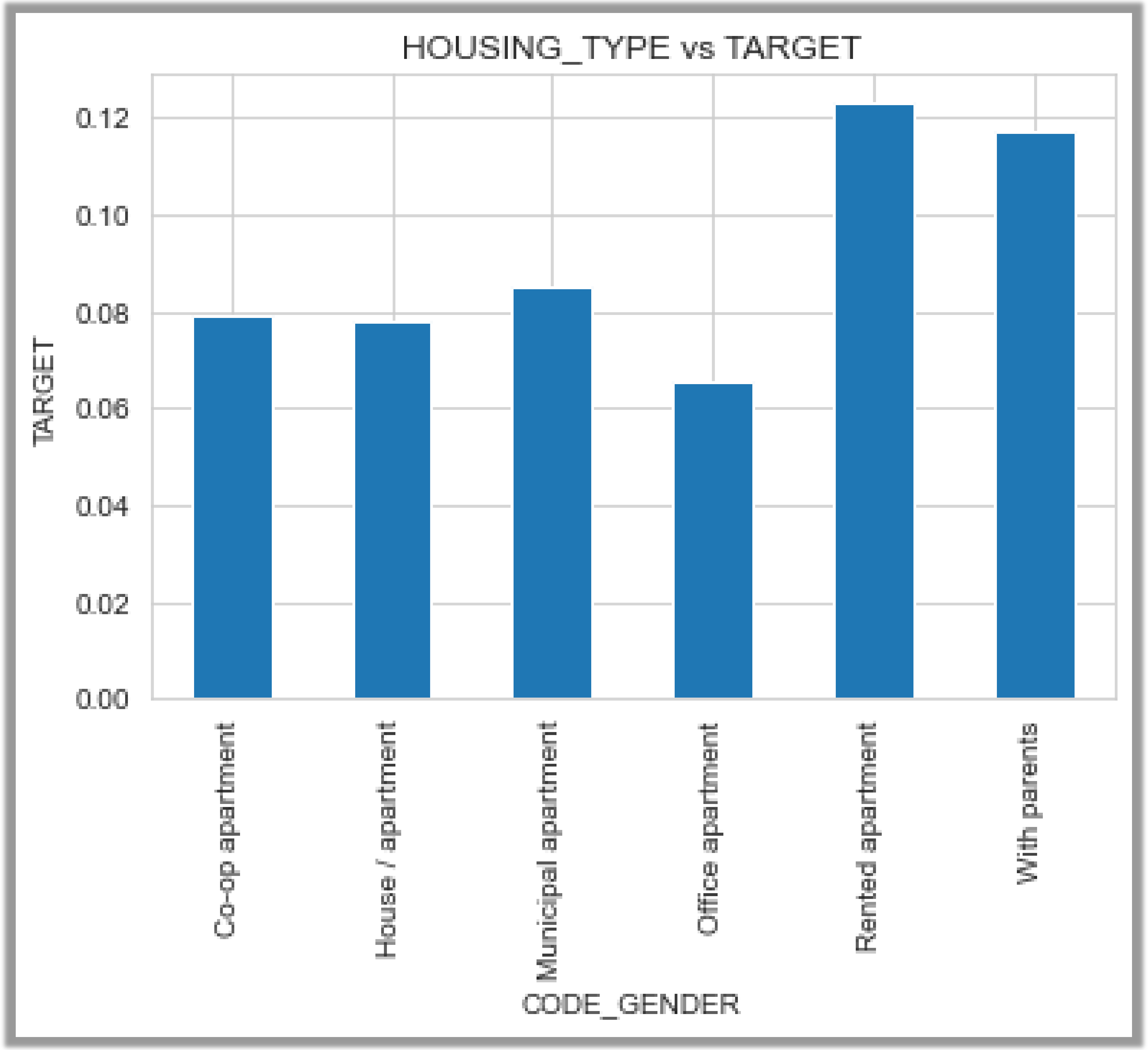
Observation: Clients owning car aged less than 30 days have Difficulty in payment

Analysis – Gender, Contract type & Target



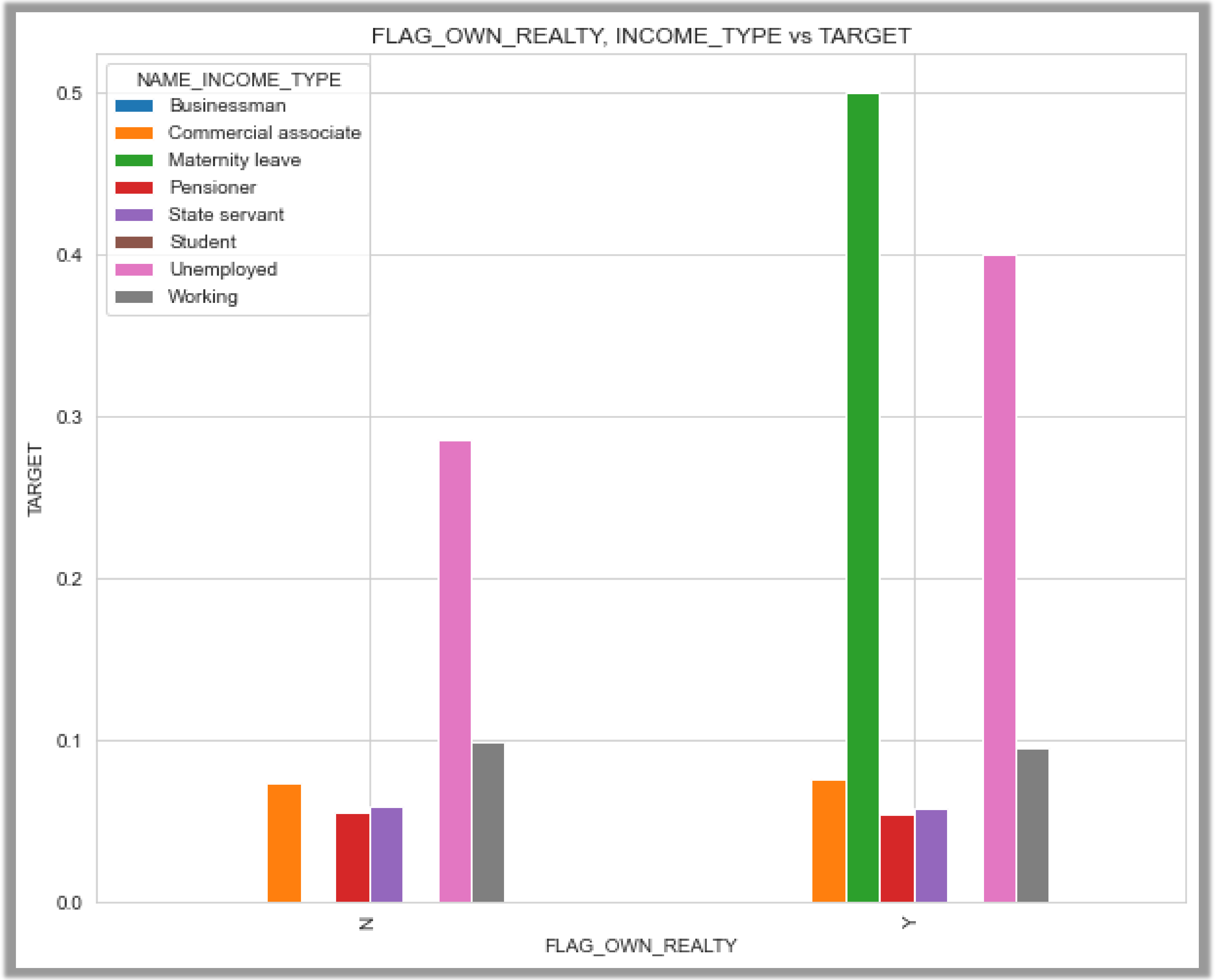
Observation: Female Clients have more Difficulty in payment in both cash loans and revolving loans type

Analysis – Housing type & Target



Observation: Clients staying in rented apartments or staying with parents have more Difficulty in payment

Analysis – Own Realty, Gender & Target

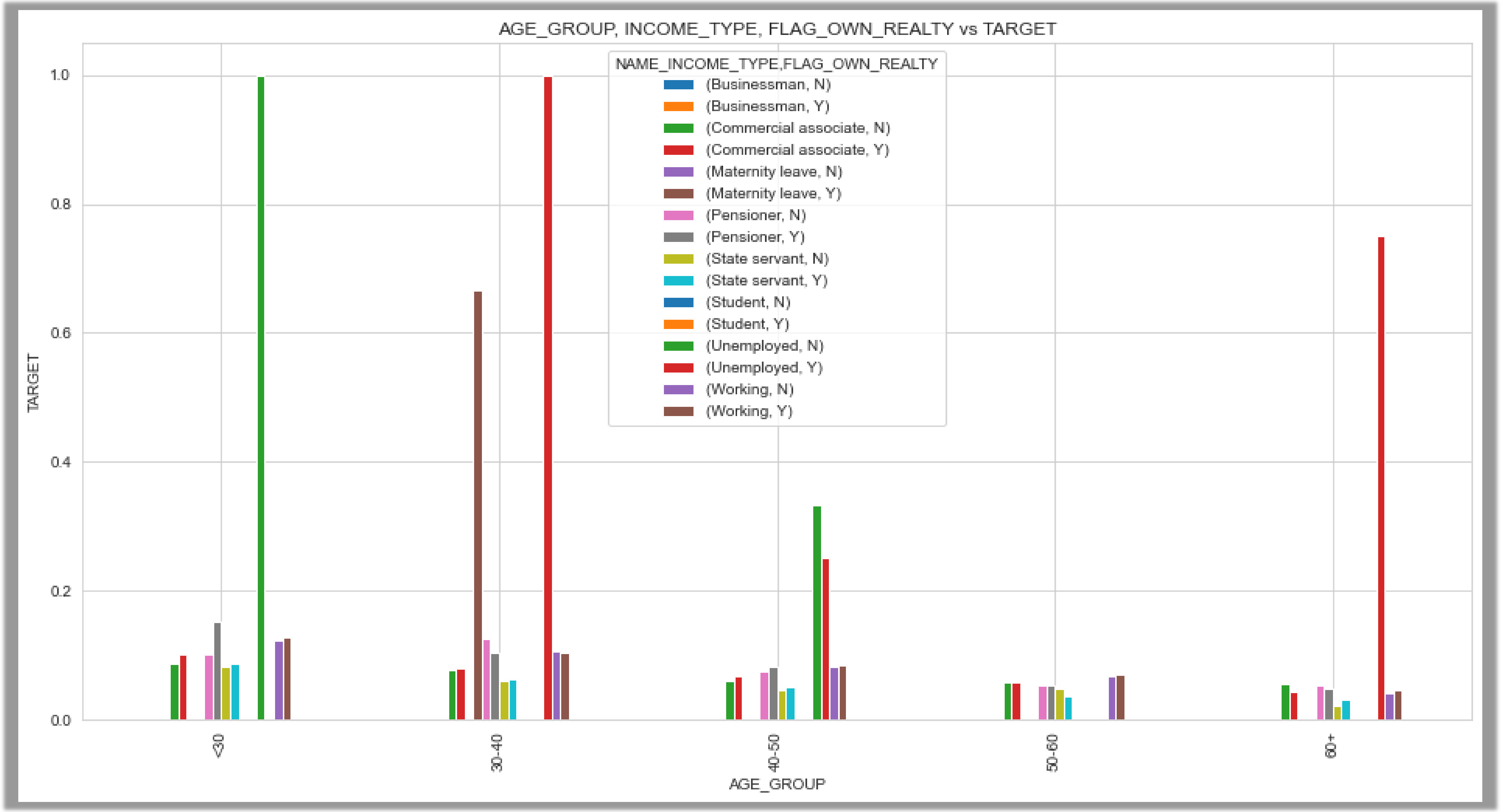


Observation:

1) Clients on maternity leave & own realty

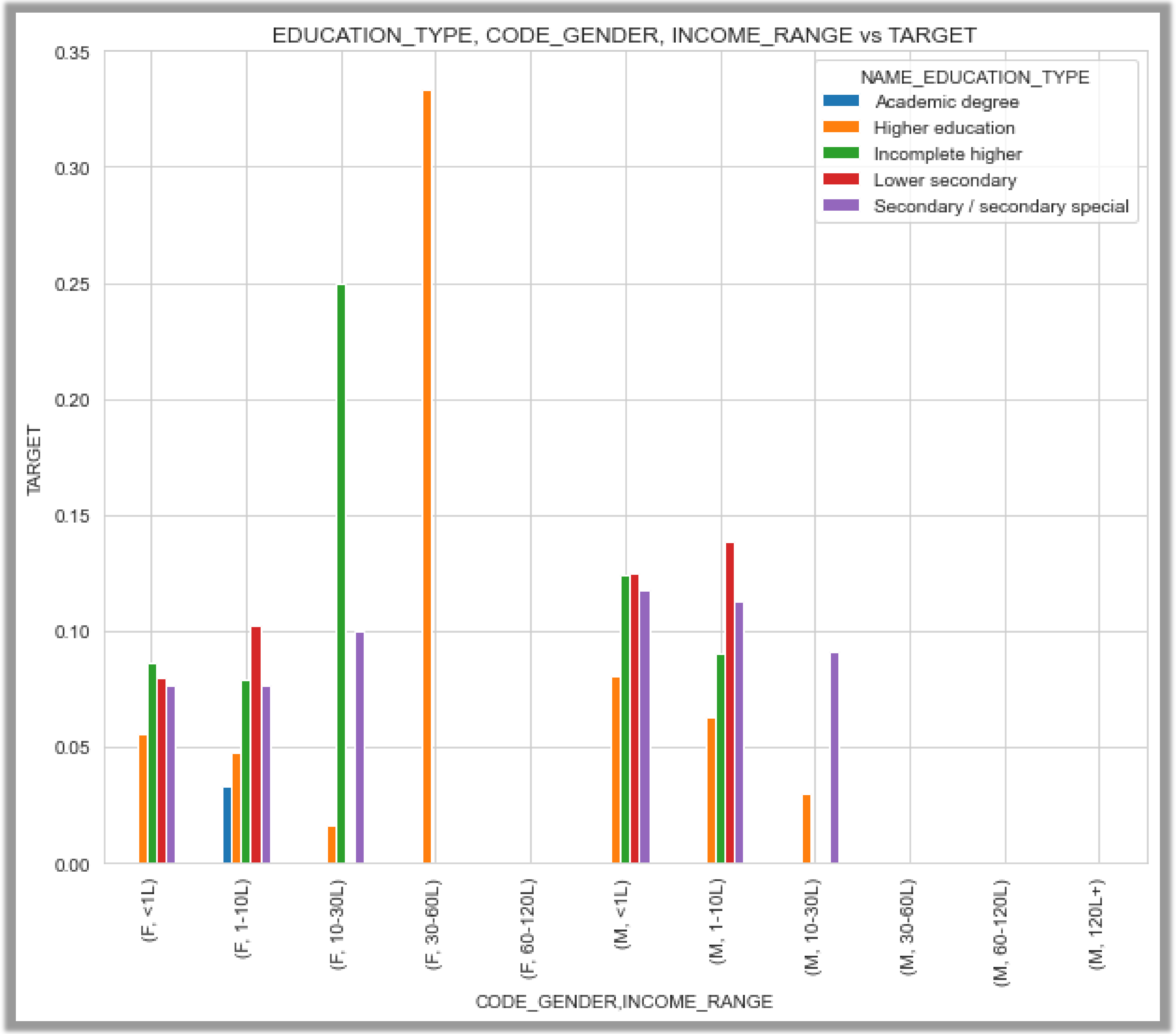
2) Unemployed Clients with or without own realty are having difficulty in paying

Analysis – Age Group, Own Realty & Target



Observation: 1) Unemployed clients aged less than 30 without own realty
2) Clients aged 30-40 & unemployed 0r on maternity leave are having difficulty in paying

Analysis – Education, Gender, Income & Target



Observation:

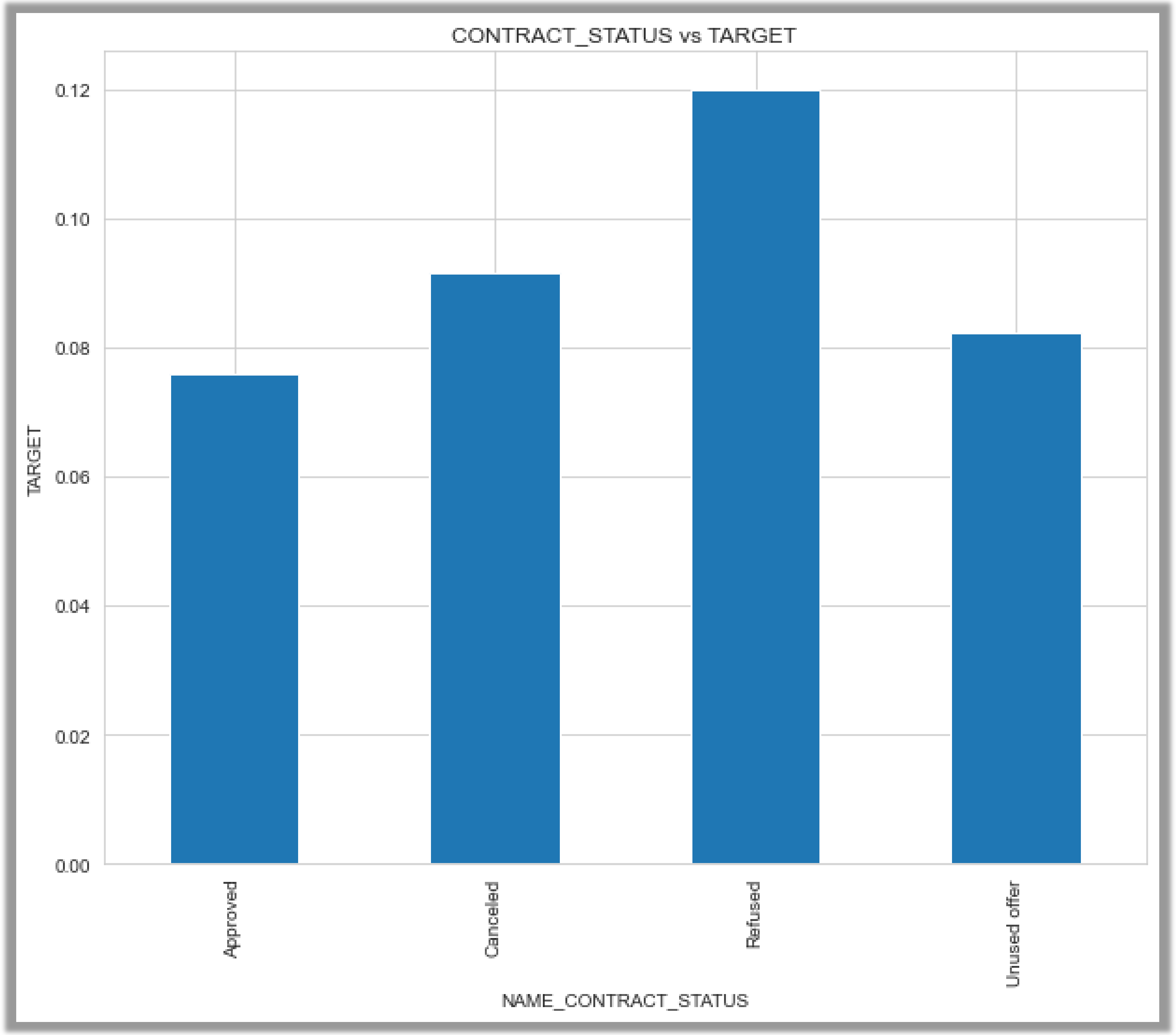
1) Female clients with Higher education & income range between 30-60 lakhs have difficulty in payment

2)Female clients with incomplete higher with income 10-30 lakhs have difficulty in payment

3)Male clients with Lower secondary education with income less than 10 lakhs have difficulty in payment

Analysis – Merged Dataset
(Previous and Current applications)

Analysis – Contract Status & Target

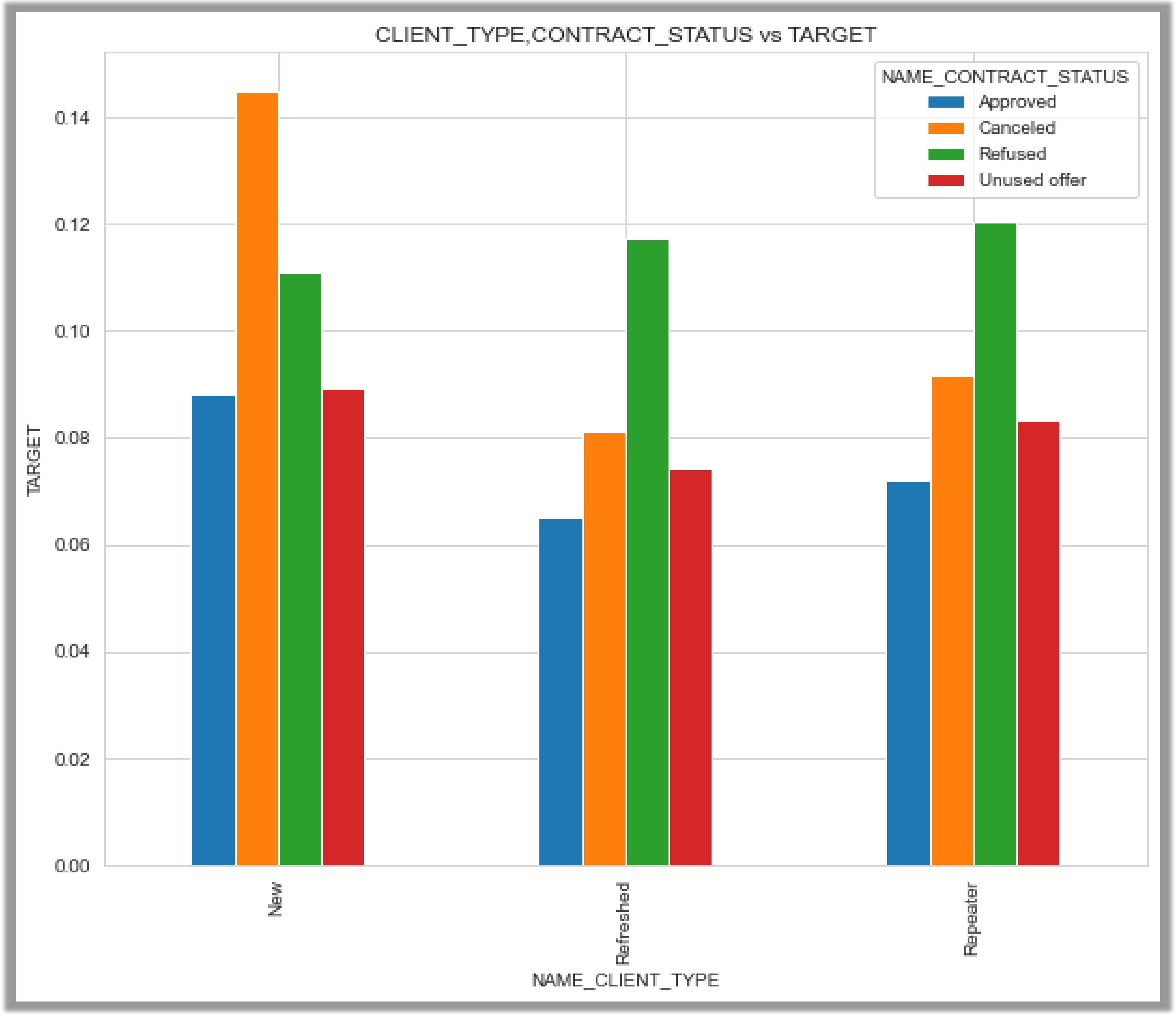


Observation:

1)Refused contracts from previous application have maximum Loan-Payment Difficulties .

2)Clients whose previous application was approved have less % with Loan-Payment Difficulties.

Analysis – Client Type, Contract Status & Target

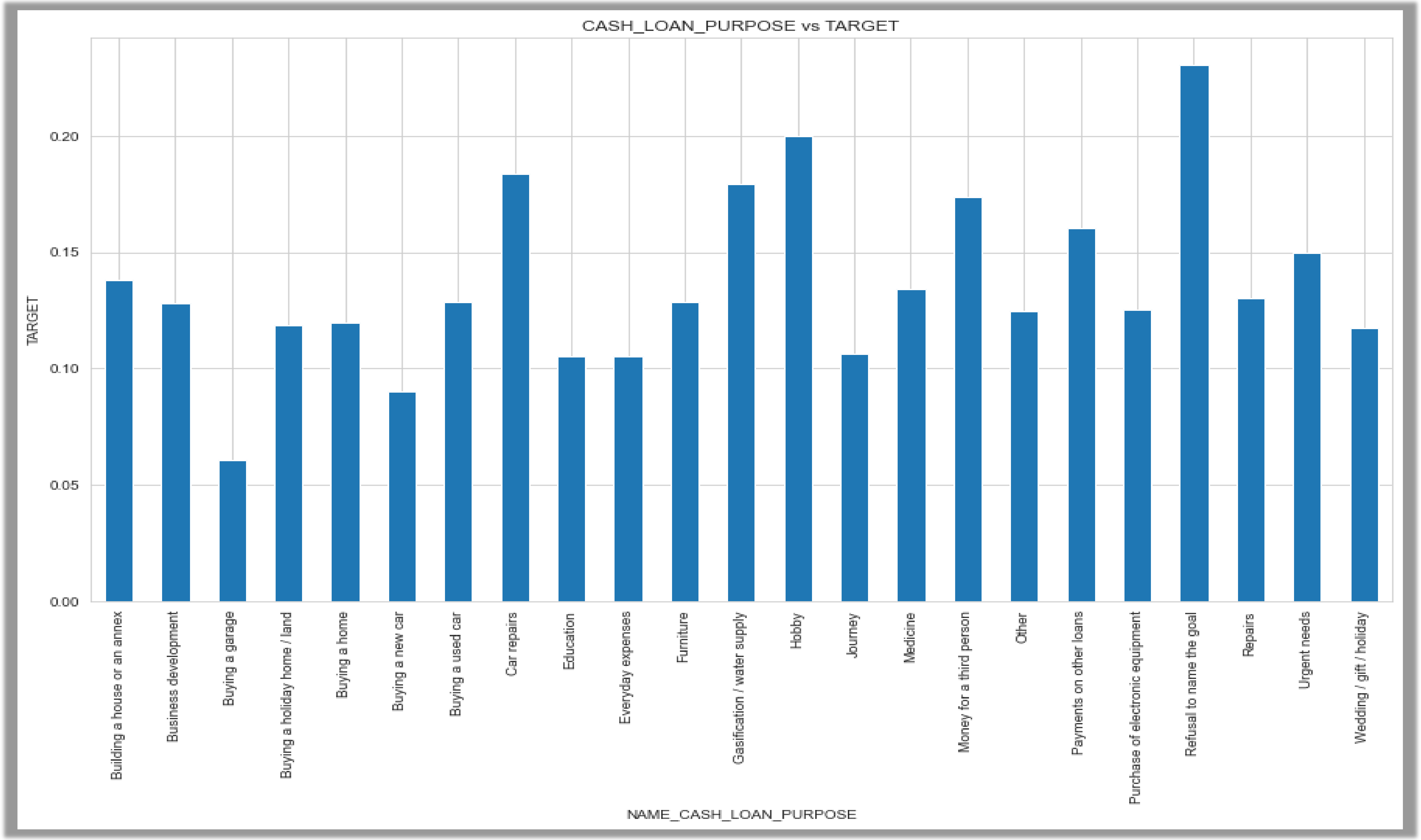


Observation:

1) More % of cancelled and refused application previously are observed for new clients and are having difficulty in payment

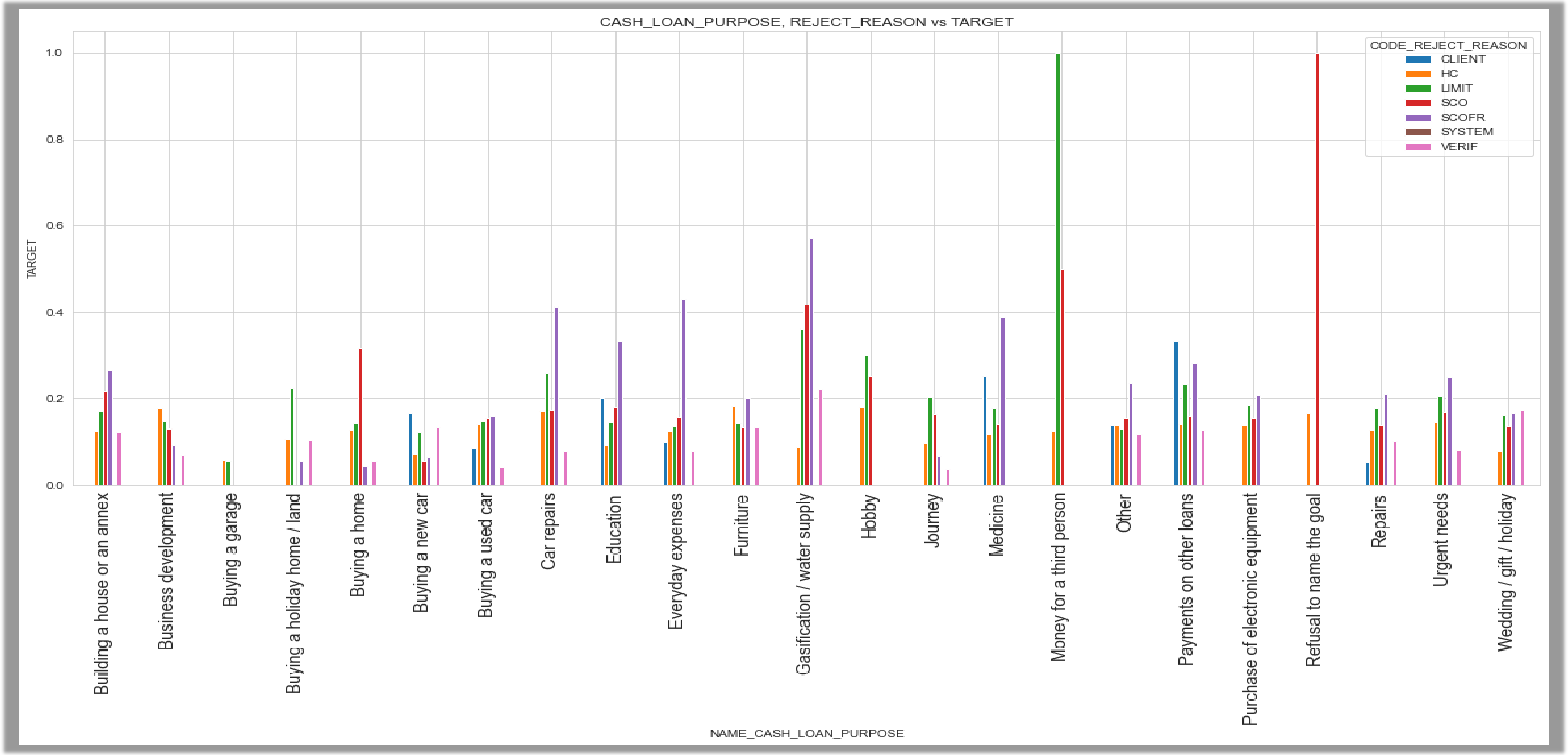
2) Number of refused requests is same across all client types

Analysis – Cash Loan Purpose & Target



Observation: Clients with loan purposes stated as Refusal to name the goal, Hobby and Car repairs have payment difficulties

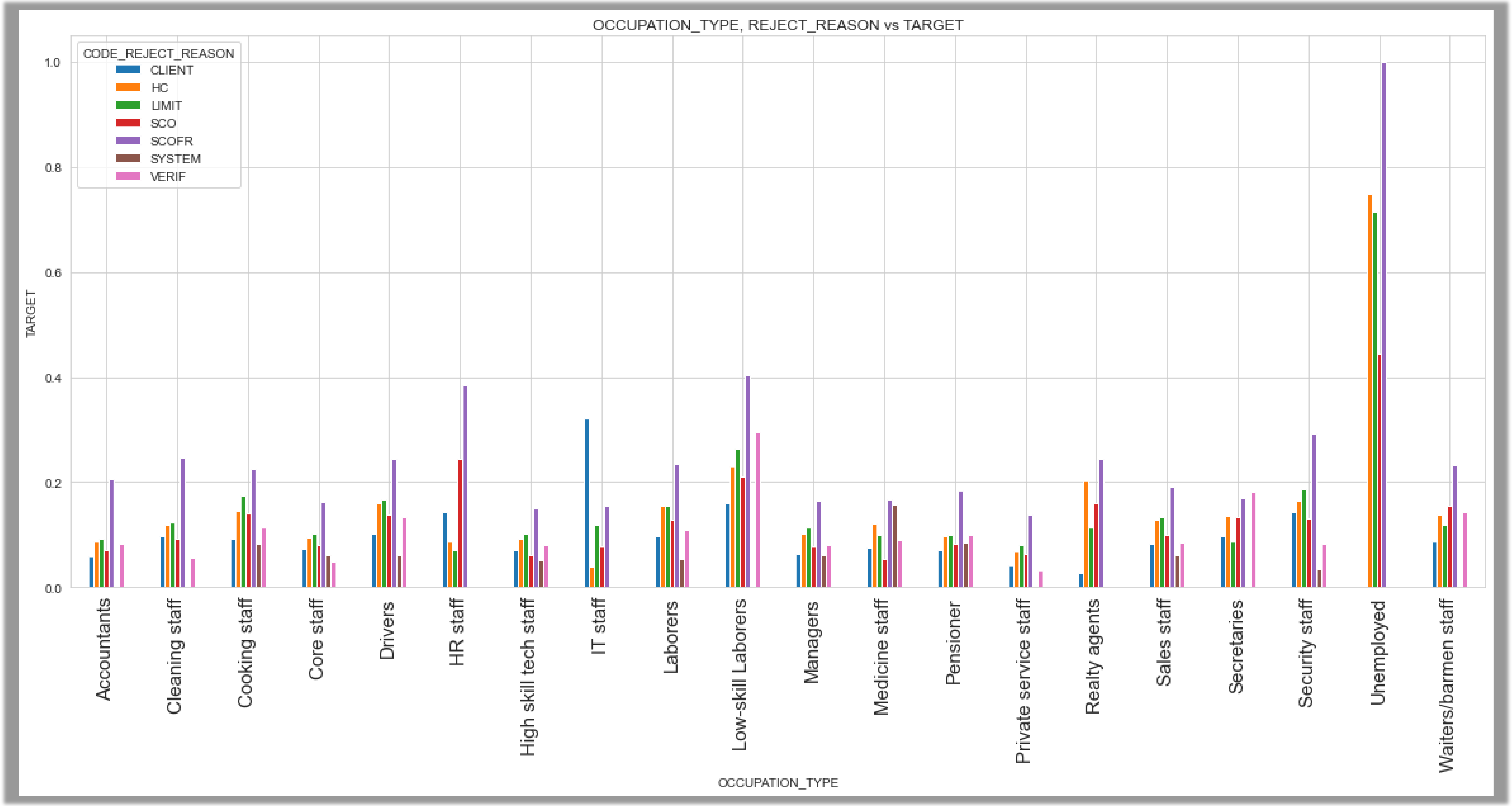
Analysis – Cash Loan Purpose, Reject reason & Target



Observation: Clients with loan purposes stated as Refusal to name the goal, with rejected application for SCO previously have payment difficulties

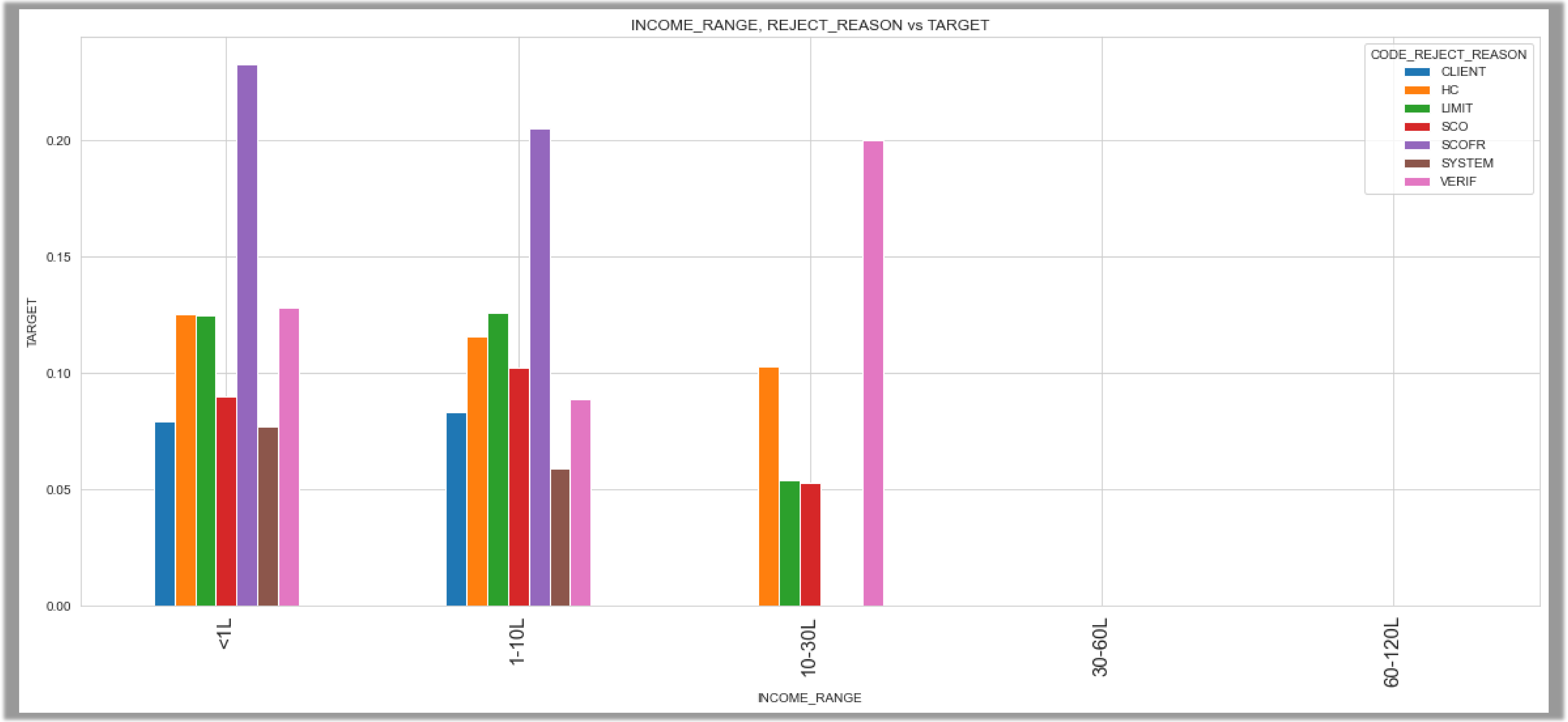
Clients with loan purposes stated as Money for third person, with Rejected application for LIMIT previously have payment difficulties

Analysis – Occupation Type, Reject reason & Target



Observation: Unemployed Clients, with rejected application for SCOFR or HC previously have payment difficulties

Analysis – Income Group, Reject reason & Target



Observation: 1)Clients in income range less than 1lakh with rejected application for SCOFR previously have payment difficulties
2)Clients in income range 1-10lakh with rejected application for SCOFR previously have payment difficulties¶
3)Clients in income range 10-30lakh with rejected application for VERIF previously have payment difficulties

Conclusion

Recommended groups :

- Approved clients in their previous applications.
- Highly educated clients.
- Clients with higher income.
- Senior citizens in all categories.
- Male clients are comparatively favourable than females.

Risky groups :

- Previously refused, cancelled or unused offer clients.
- Low income groups with previously refused status.
- Unemployed clients.
- Young clients are comparatively riskier than mid age clients and senior citizens.
- Lower secondary and secondary educated clients.
- Clients with loan purpose as 'Refusal to name the goal', 'Hobby', 'Money for a third person' and 'Car repairs'.
- Clients having lesser days of employment .