EDA ASSIGNMENT

BY

ANEESA AYAZ

C10 GLOBAL BATCH

OBJRCTIVES

- 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 utilize this knowledge for its portfolio and risk assessment.

DATASET

TWO DATASET

- *'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.**

- *'previous_application.csv'* contains information about the client's previous loan data. It contains the data on whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**

PREVIOUS DATA

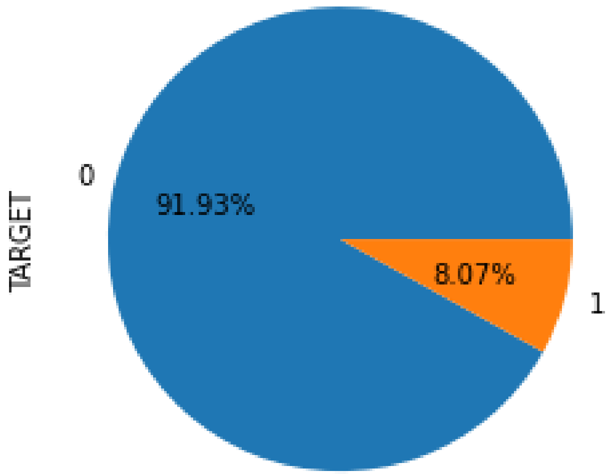
- Handled all missing values
- Drop all columns greater than 40% missing values.It will effect the further analysis
- Missing value replacemet with mean median and mode values
- If outliers are exist median value should be preferred
- Otherwise mean and mode values are good to replace

APPLICATION DATA

- Drop all columns with more than 40% missing values
- Columns with less than 40% missing values dealt with two ways
  - some column have more missing values should be leave it as it is
  - Other columns replace with mean median and mode values

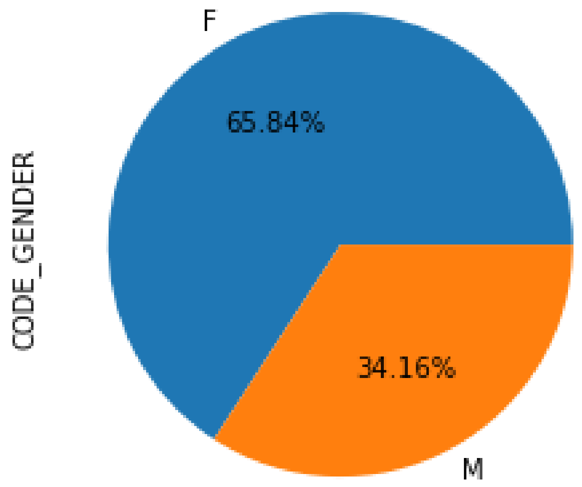
UNIVARIATE ANALYSIS FOR APPLICATION DATA

- TARGET COLUMN
- 8% CUSTOMERS TENDS TO BE DEFAULTERS



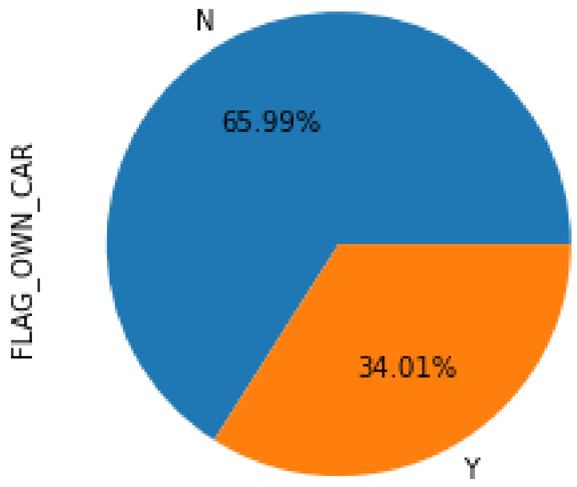
CODE _GENDER COLUMN

- 65 % FEMALE CUSTOMERS
- 34 % MALE CUSTOMERS



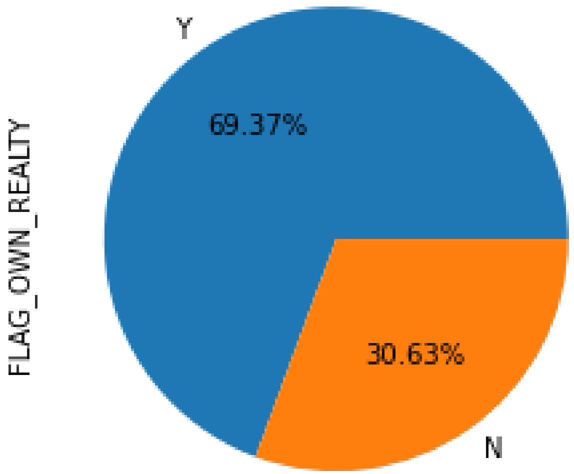
FLAG_OWN_CAR COLUMN

- 65% CUSTOMERS DOES NOT HAVE CAR
- 34% OWNS CAR



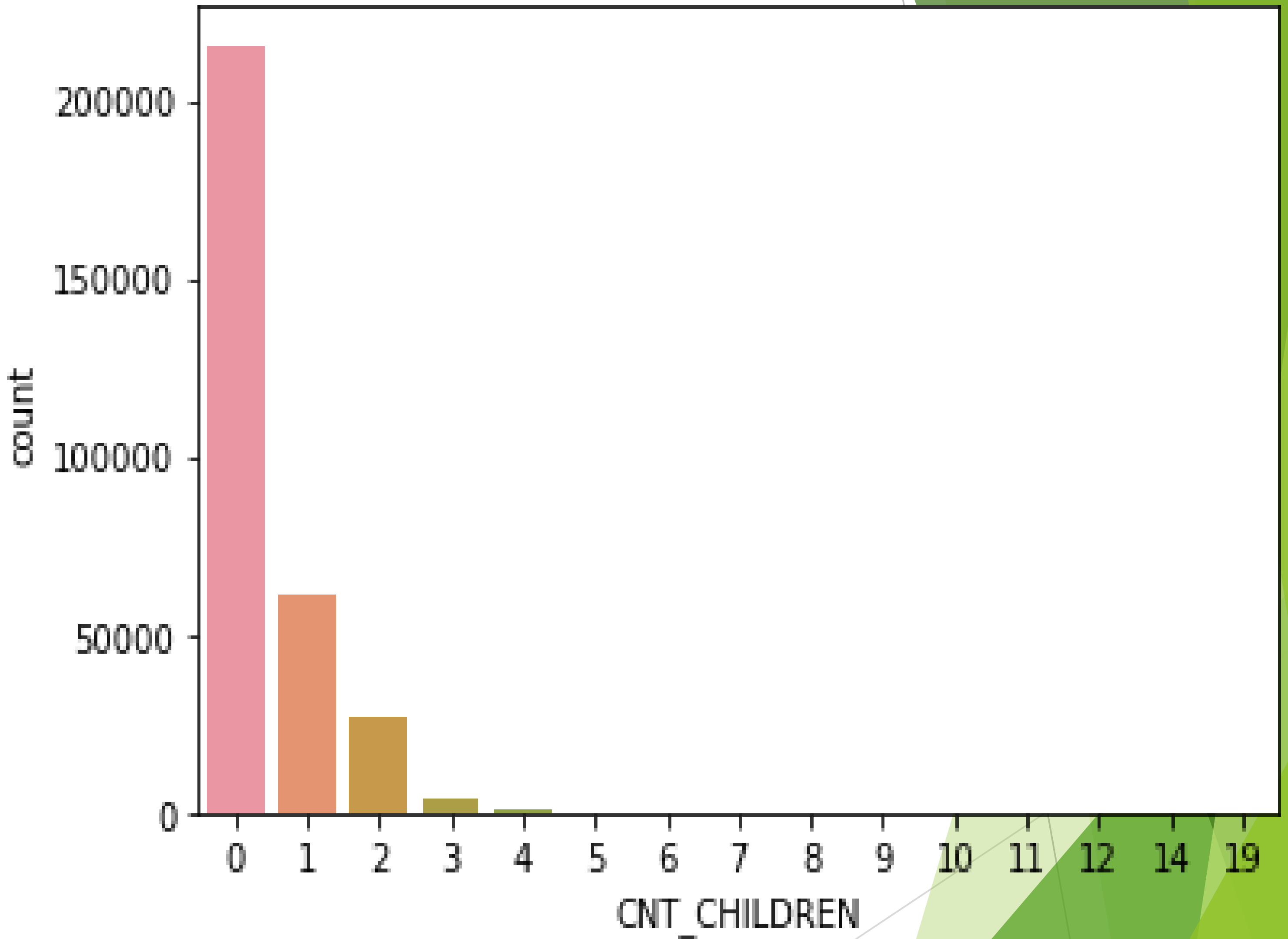
FLAG_OWN_REALTY

- 69% OWN PROPERTY
- 31% DOES NOT OWN PROPERTY



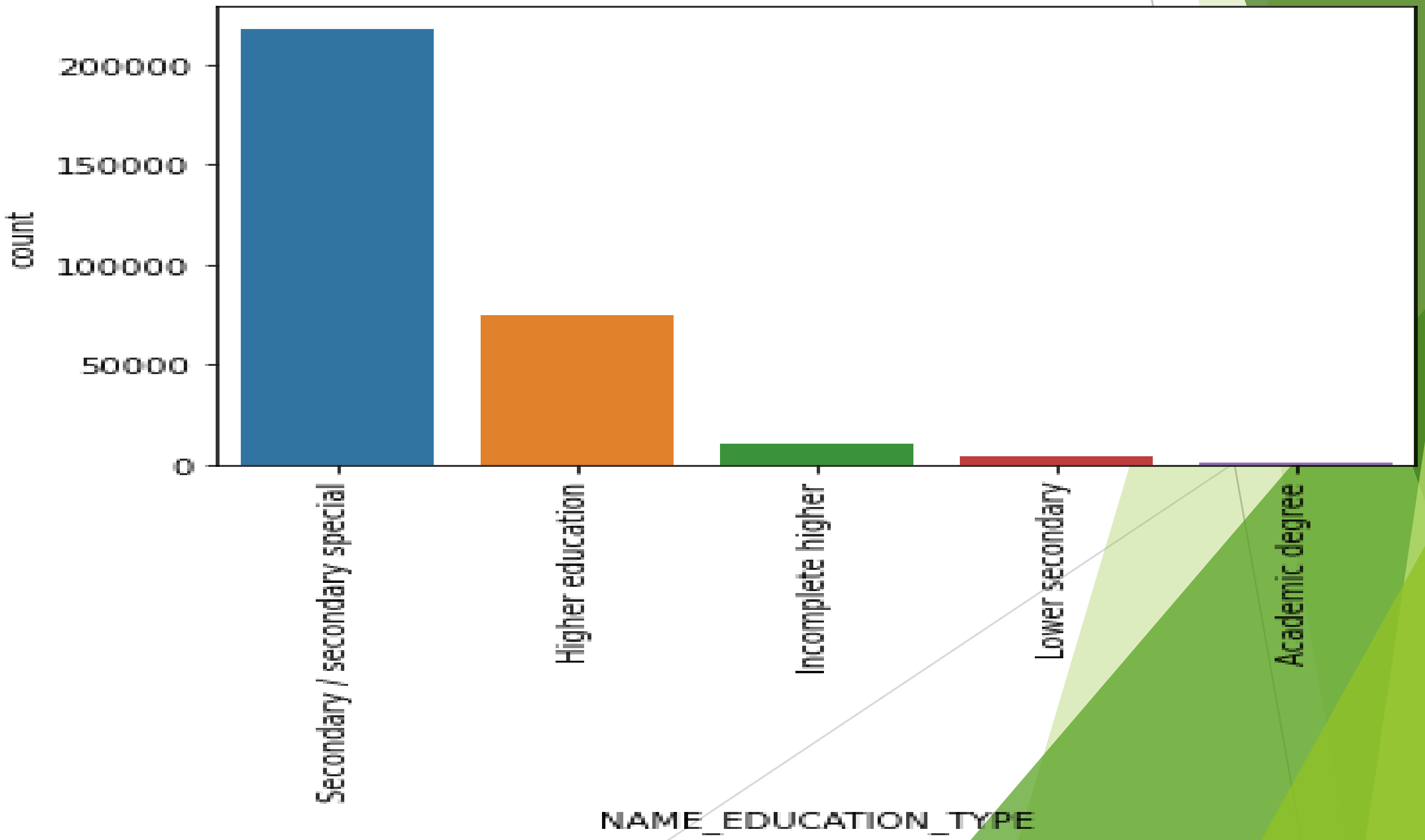
CNT_CHILDREN

- 70% CUSTOMERS HAVE NO KIDS
- CATEGORICAL ORDERED DATA



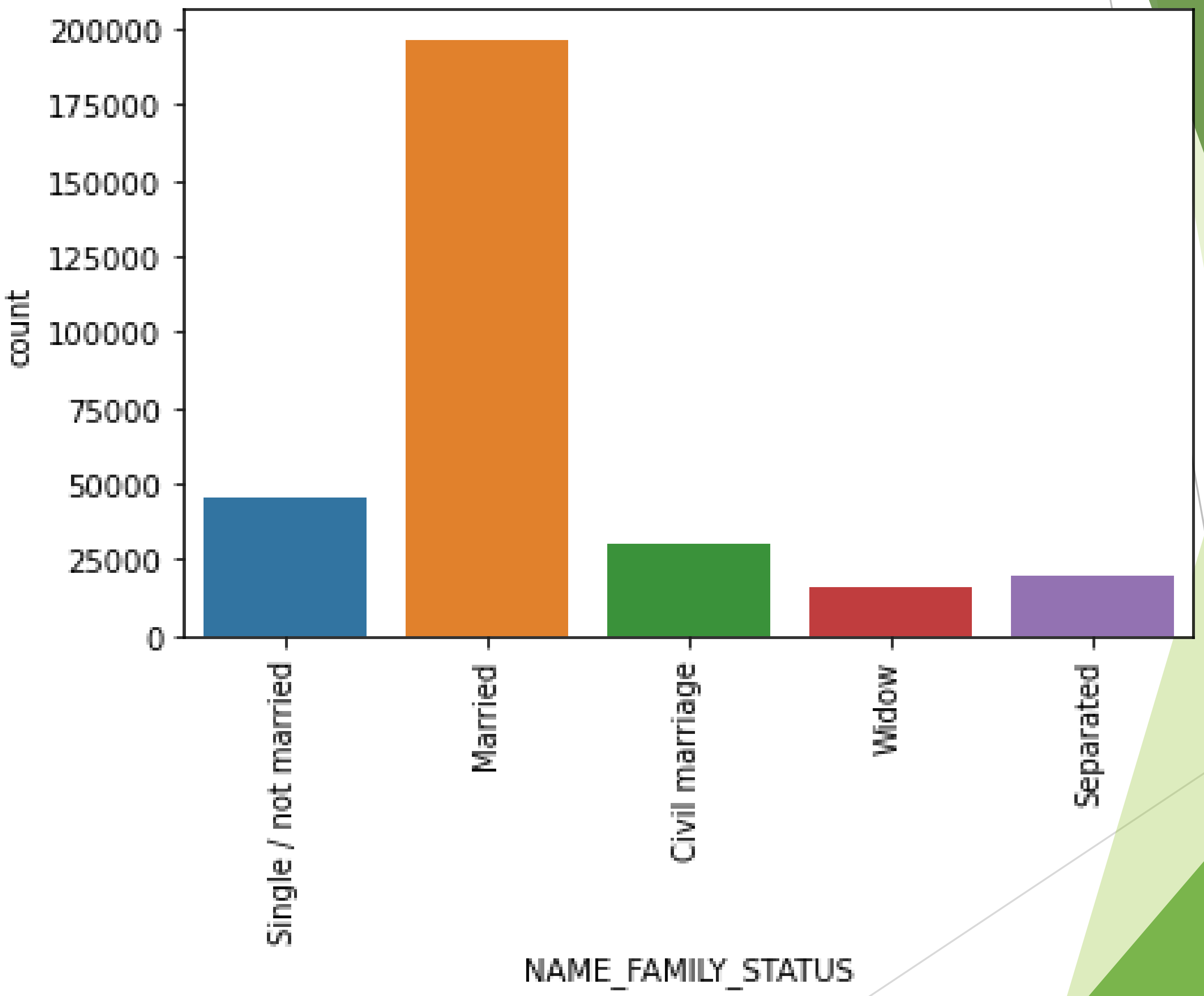
NAME_EDUCATION_TYPE

- MOST OF THE CUSTOMERS ARE SECONDARY / SECONDARY SPECIAL HOLDERS
- CATEGORICAL ORDERED DATA



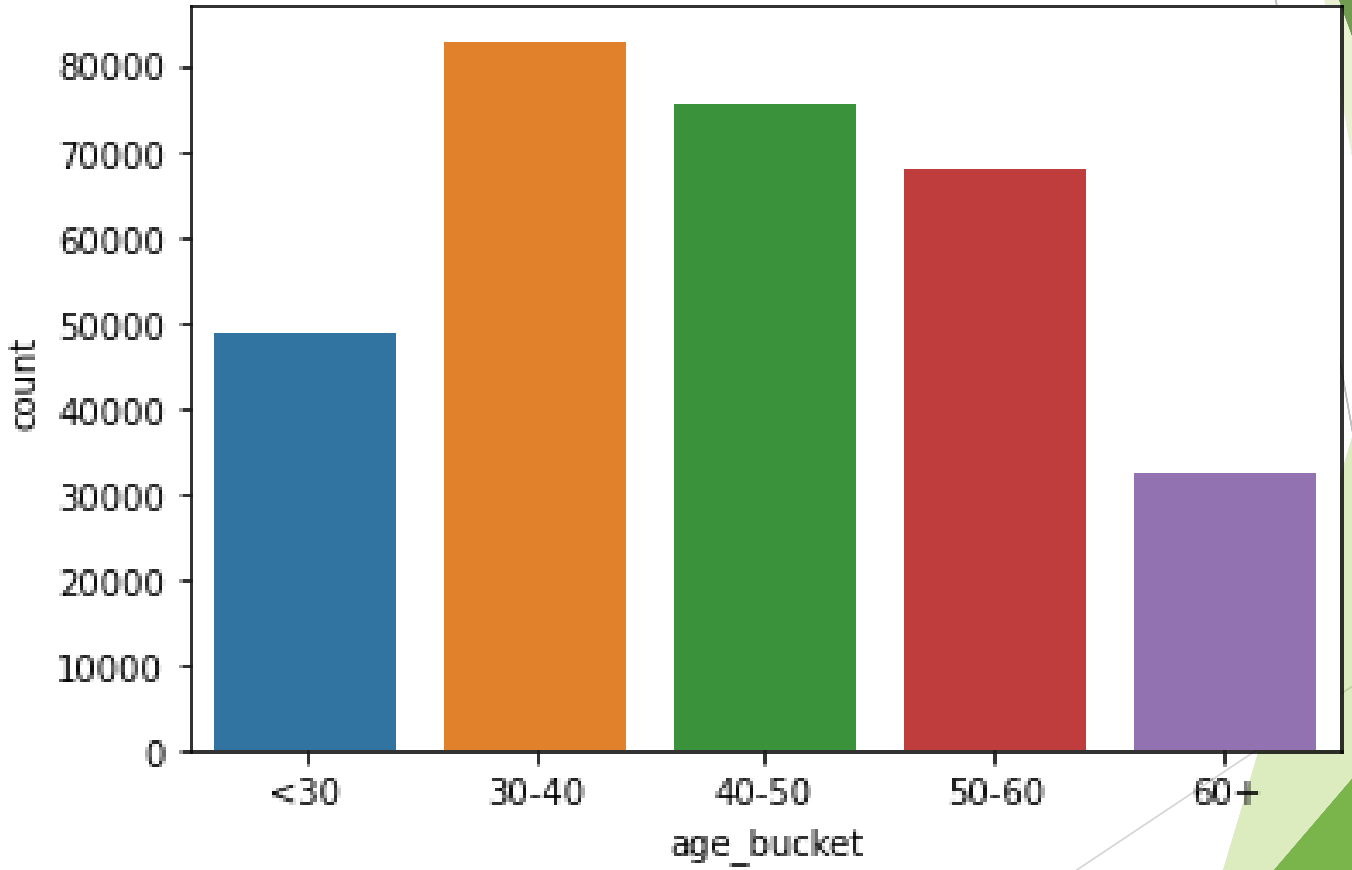
FAMILY_STATUS

- MOST CUSTOMERS ARE MARRIED



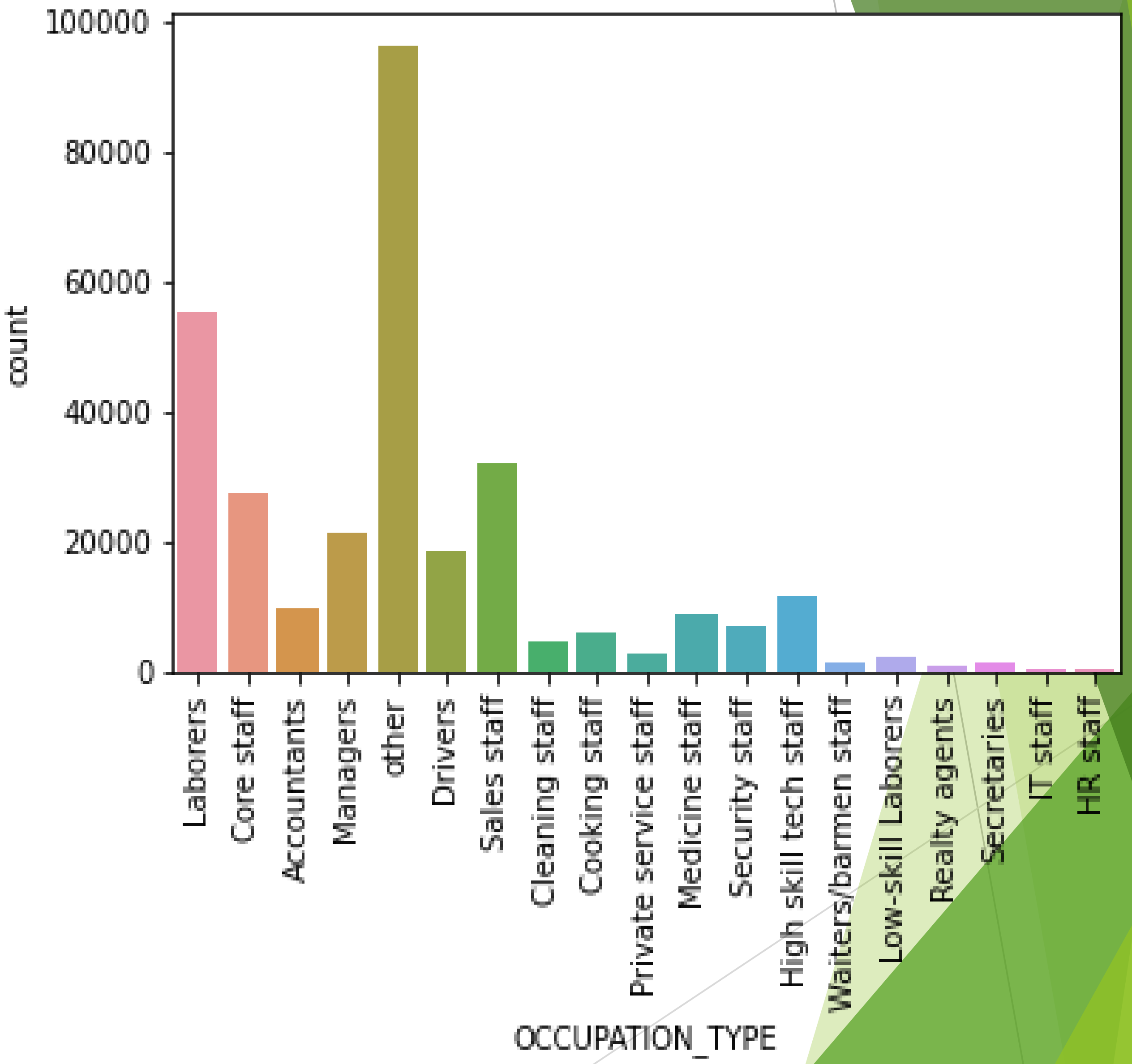
AGE BUCKET

- 30-40 AGE BUCKET HAS MORE VALUES



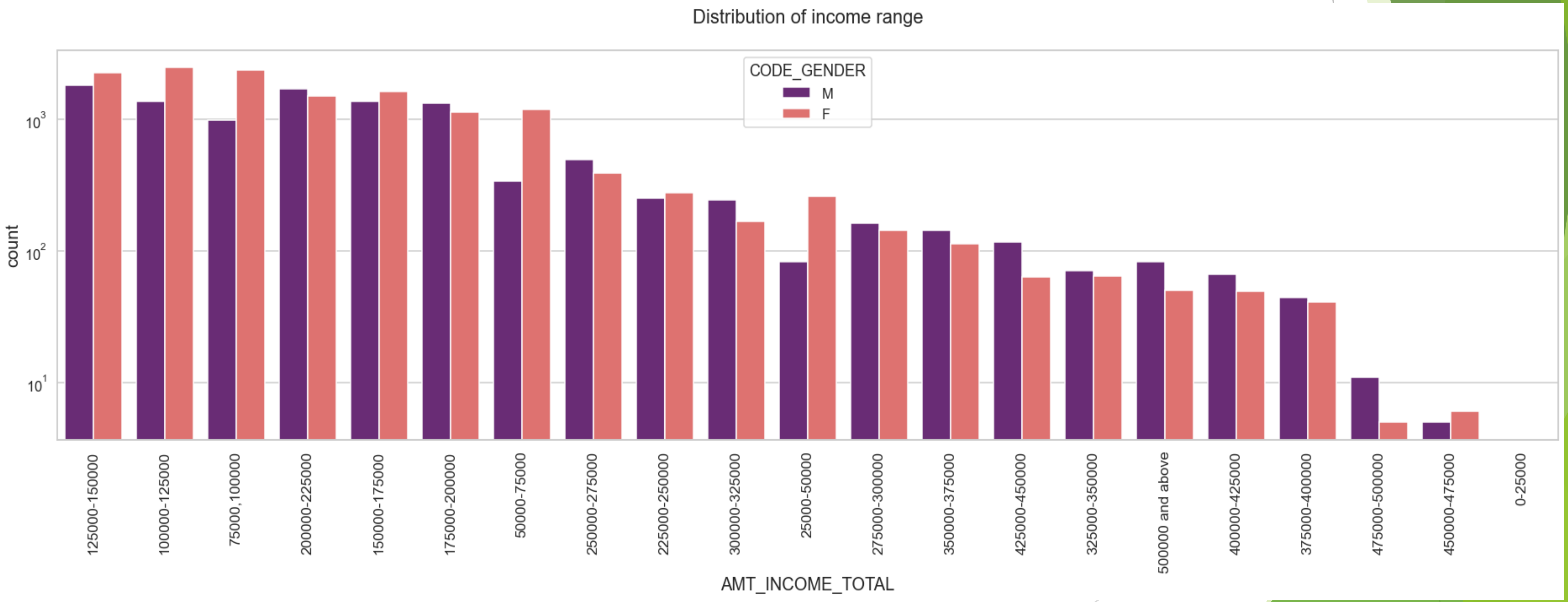
OCCUPATION_TYPE

- OTHER CATEGORY HAS MORE VALUES

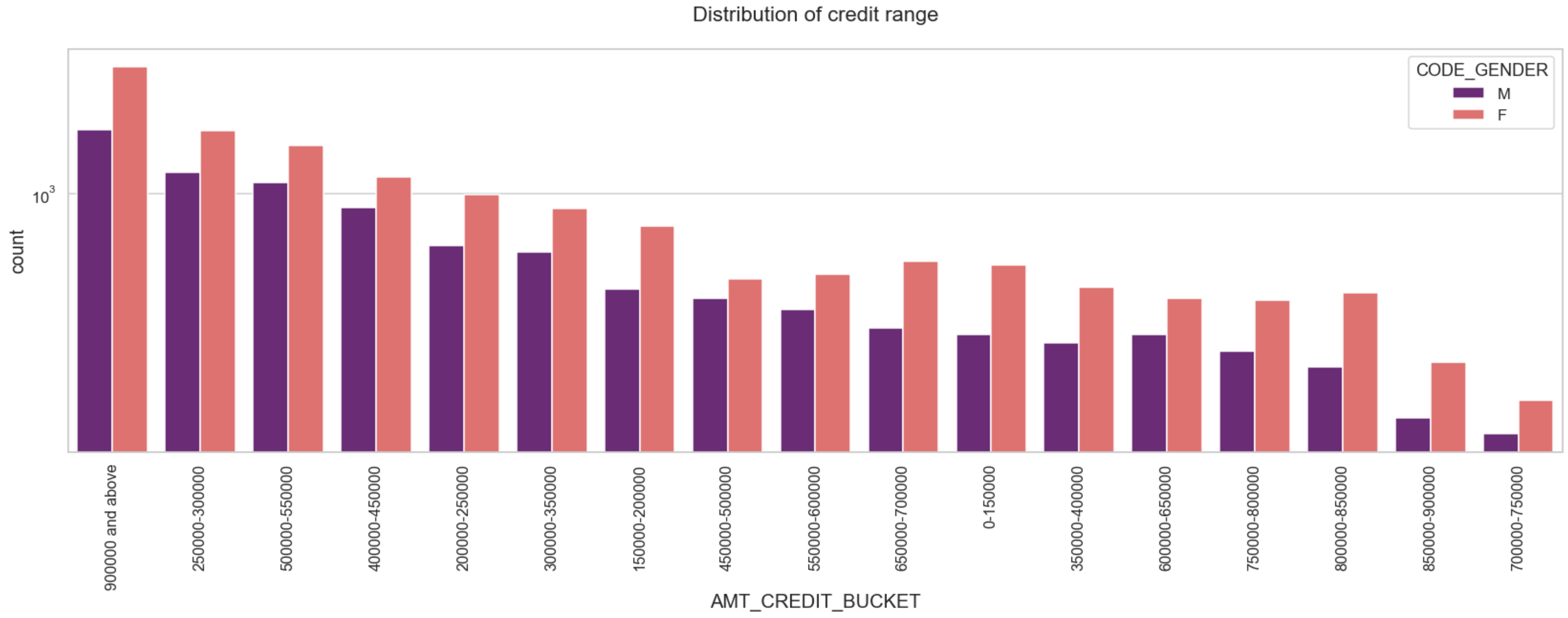


BIVARIATE ANALYSIS

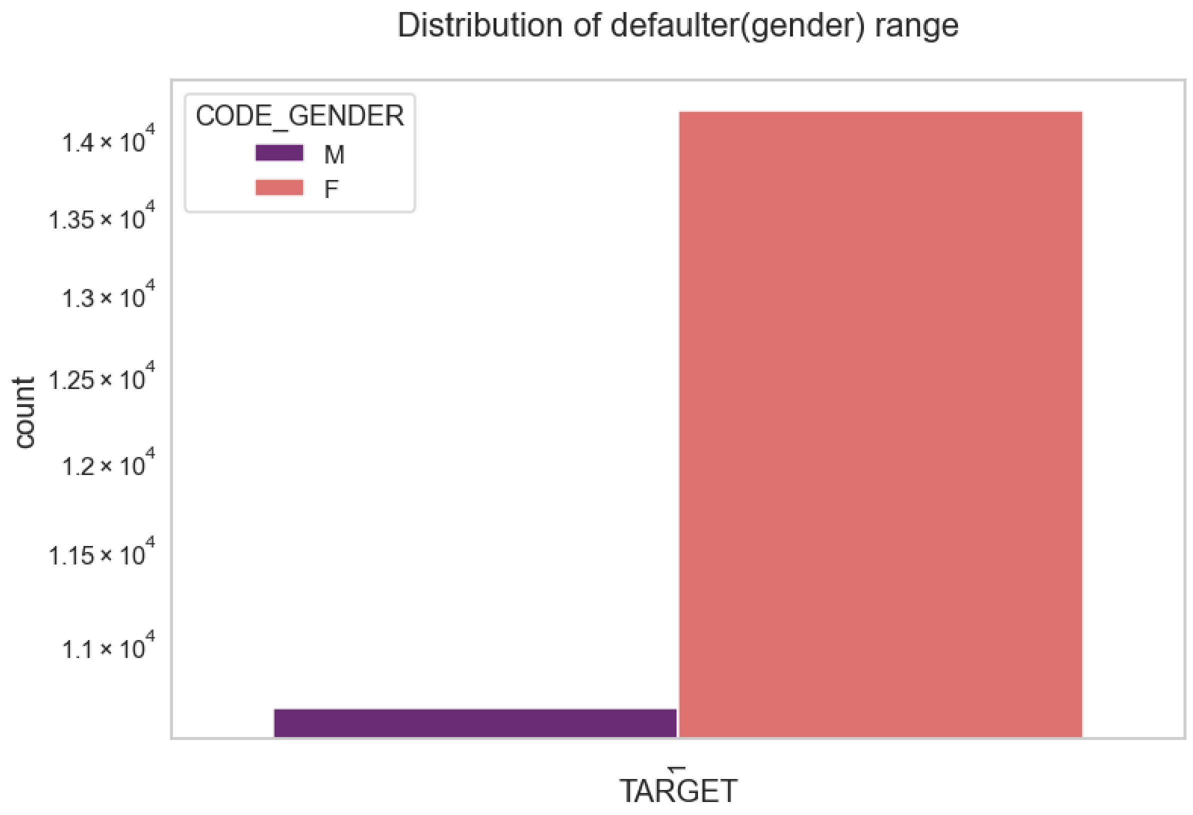
- TARGET_1 DATAFRAME(DATAFRAME WHICH HAS DEFAULTERS DETAILS)
- INCOME VS GENDER



CREDIT AMOUNT VS GENDER

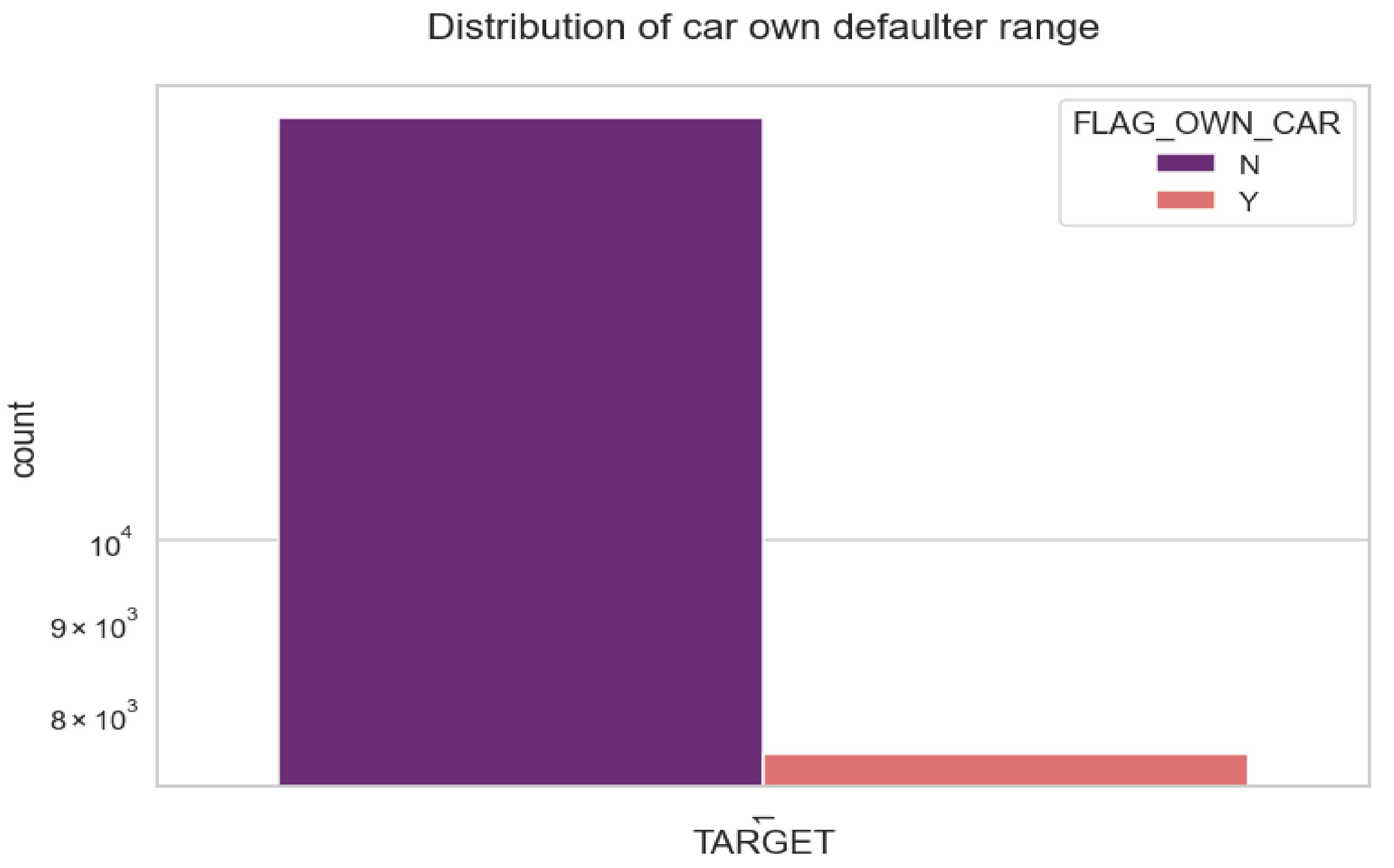


TARGET VARIABLE VS GENDER

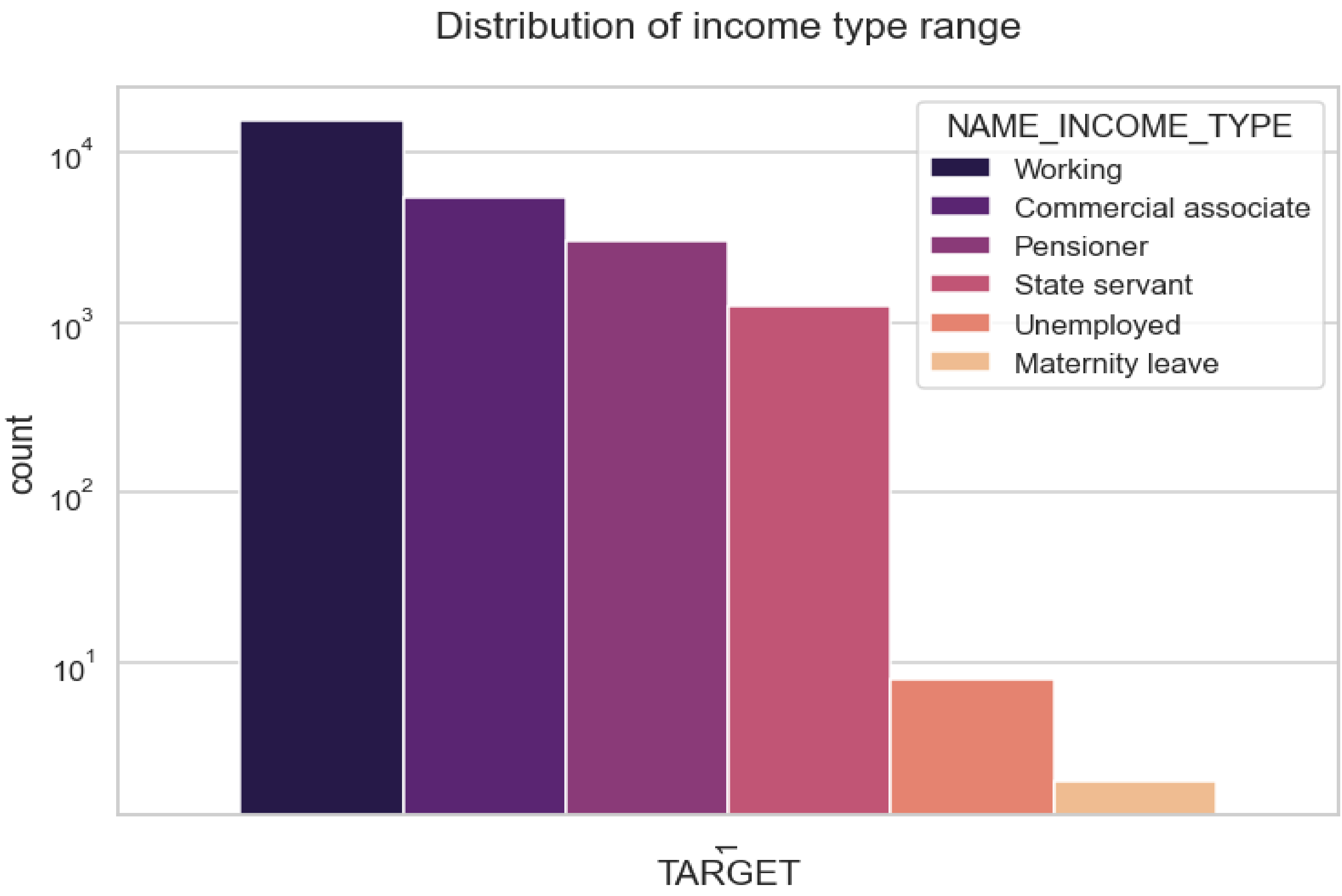


TARGET VS CONTRACT TYPE

OWNS CAR VS TARGET



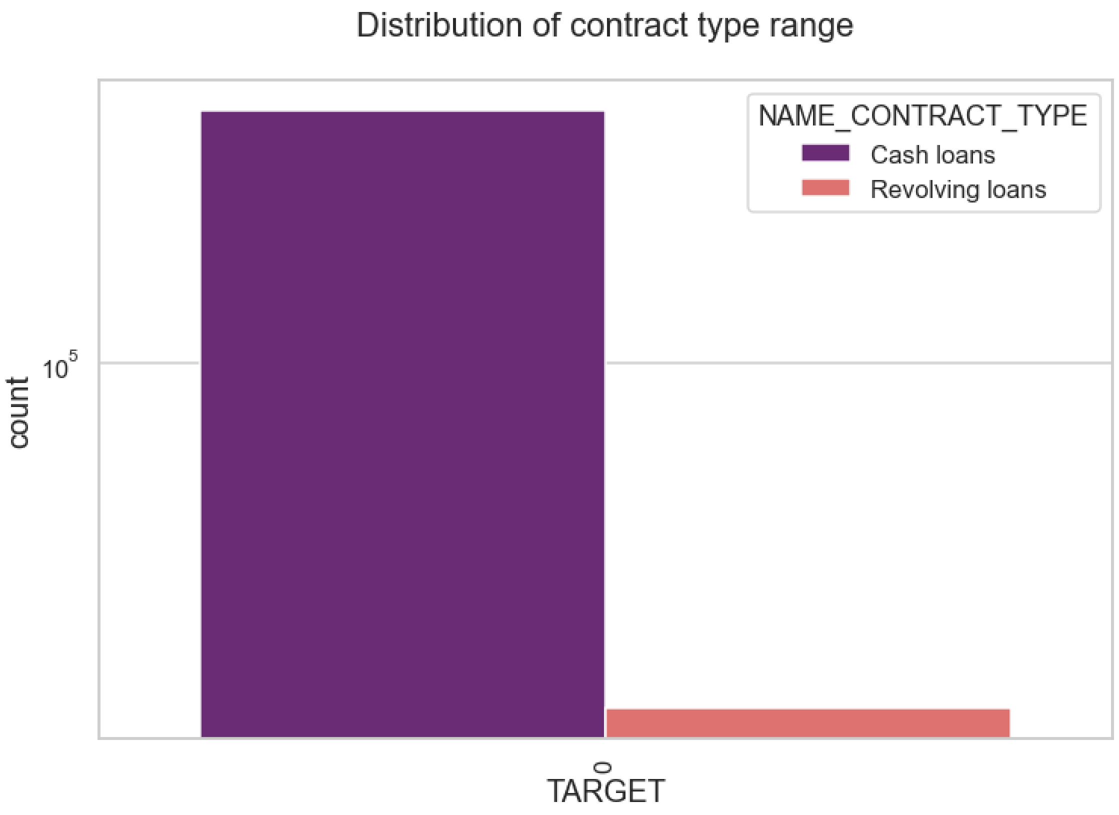
INCOME TYPE AND TARGET



TARGET 0(DATAFRAME FOR NON DEFAULTERS)

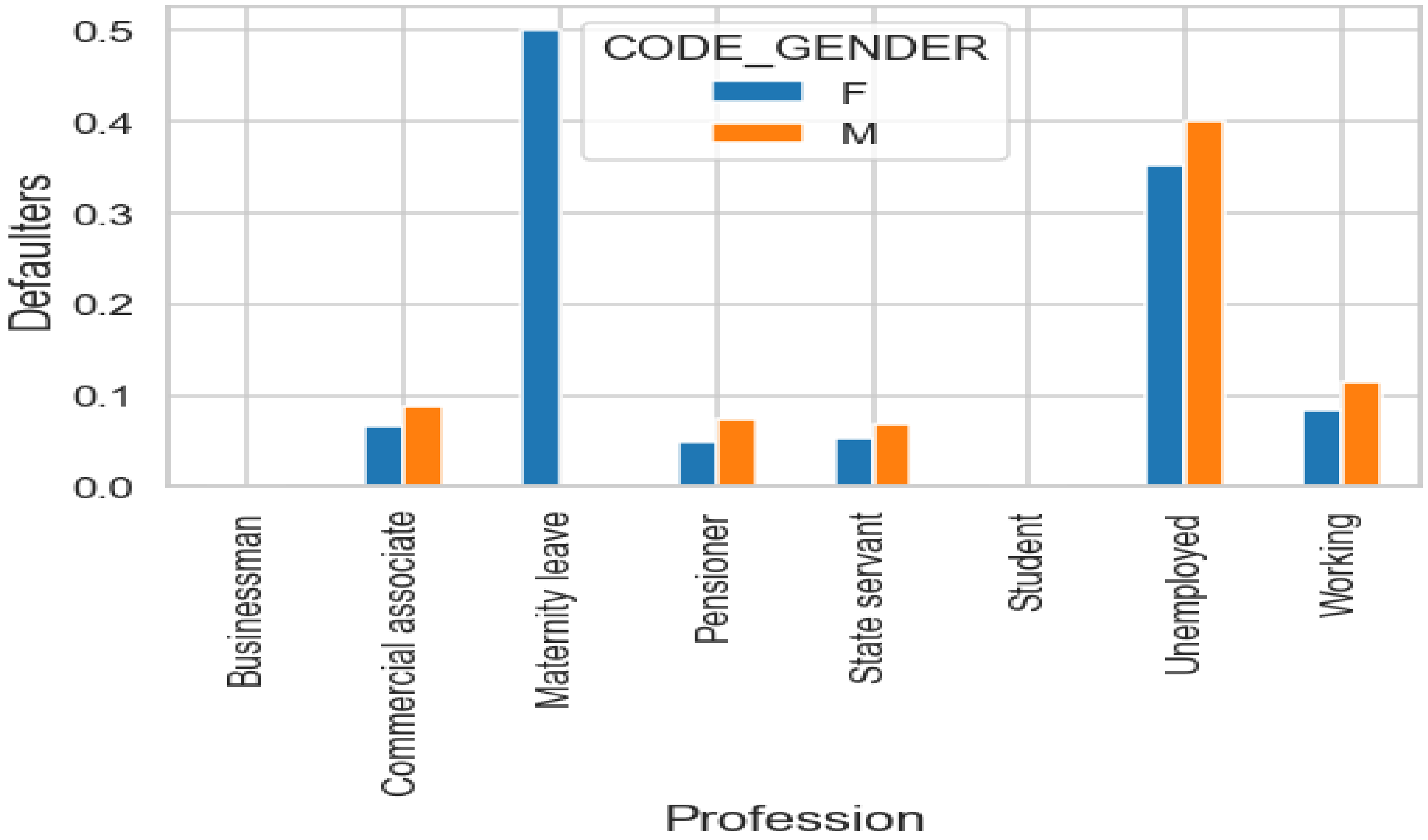
- INCOME DATA VS GENDER

TARGET VARIABLE VS CONTRACT TYPE

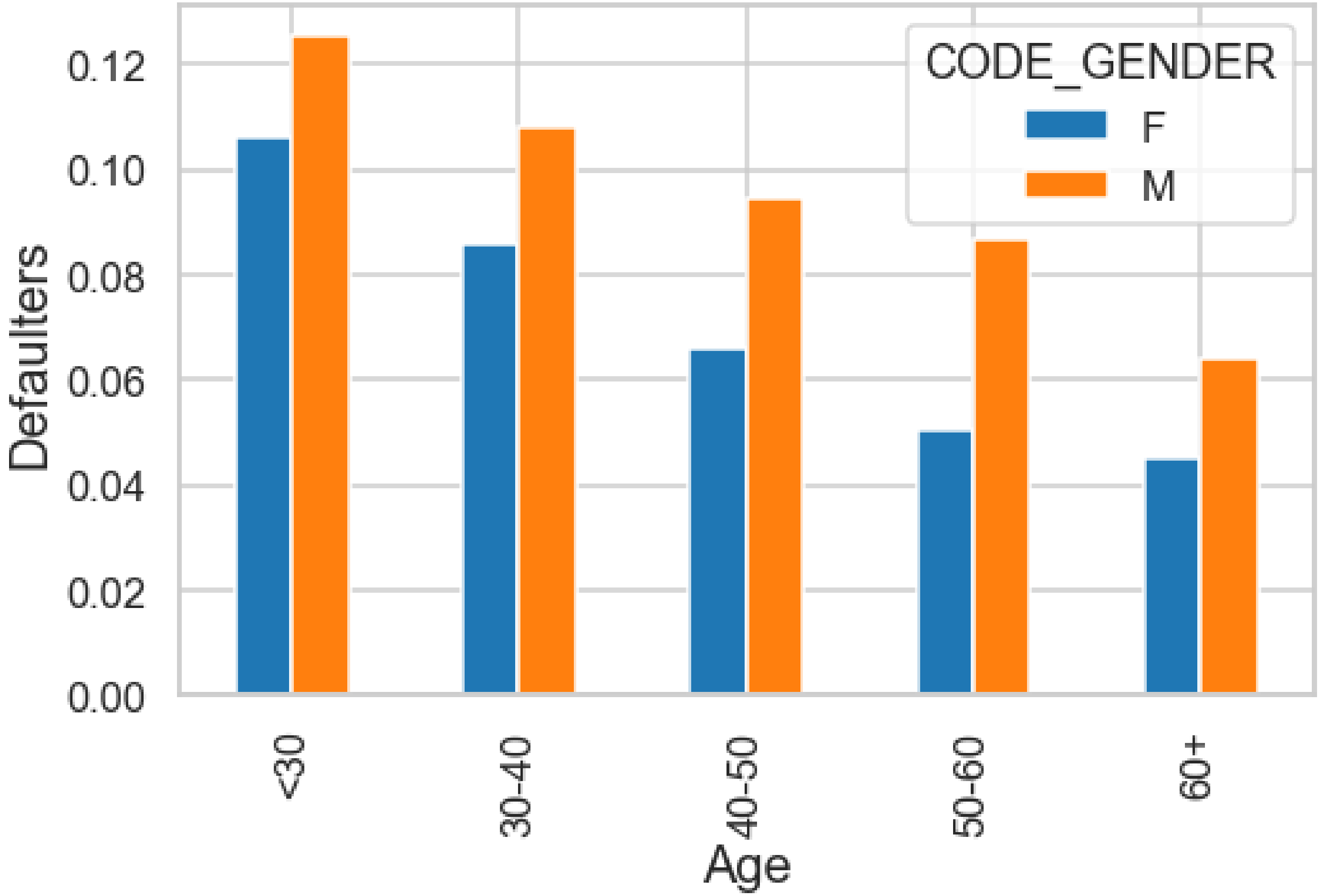


MULTI VARIATE ANALYSIS

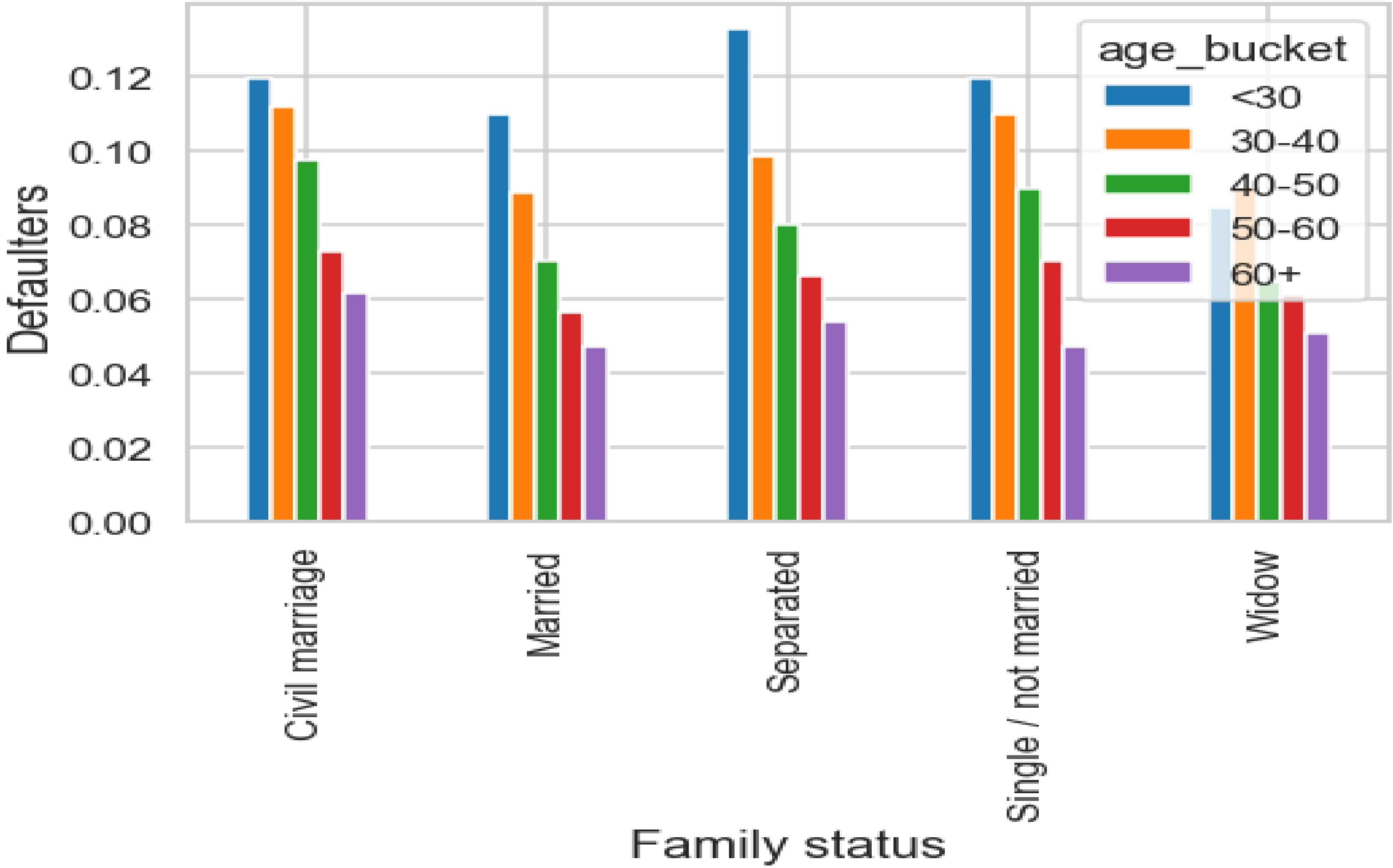
- TARGET VS INCOME TYPE VS GENDER



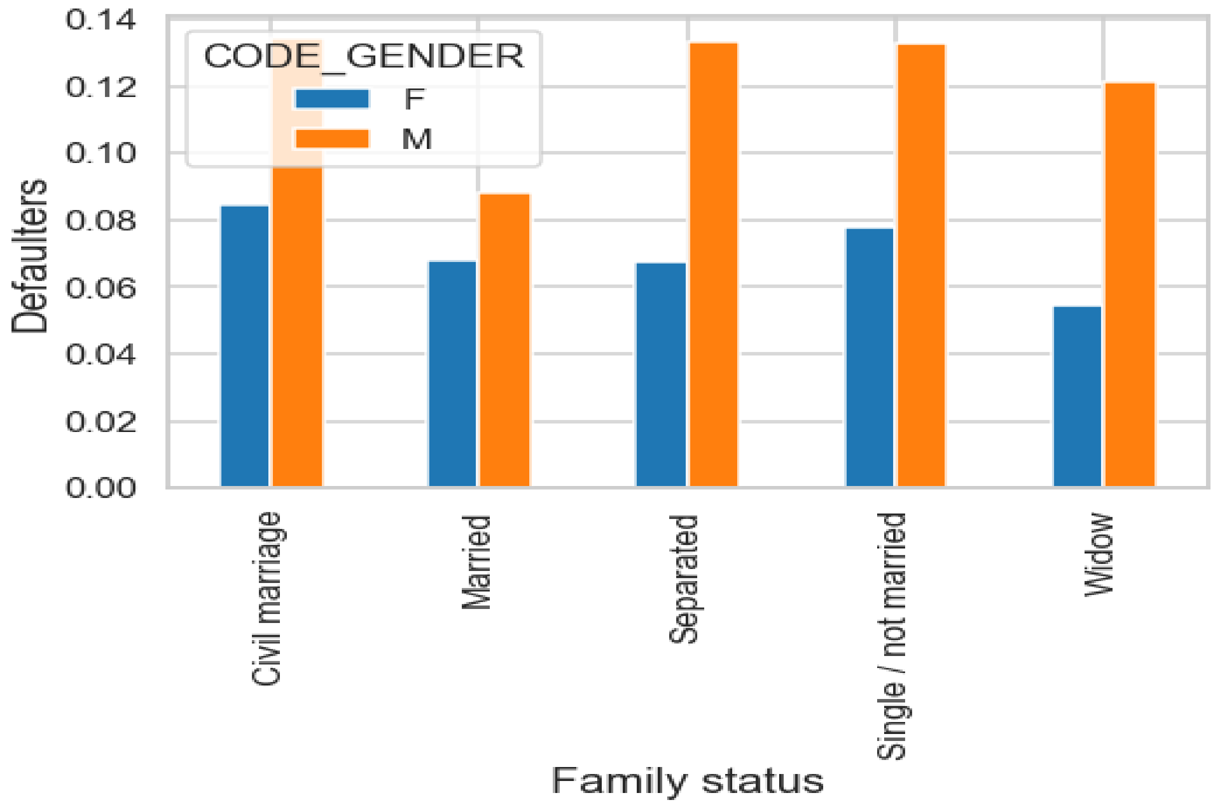
TARGET VS AGE BUCKET VS CODE GENDER



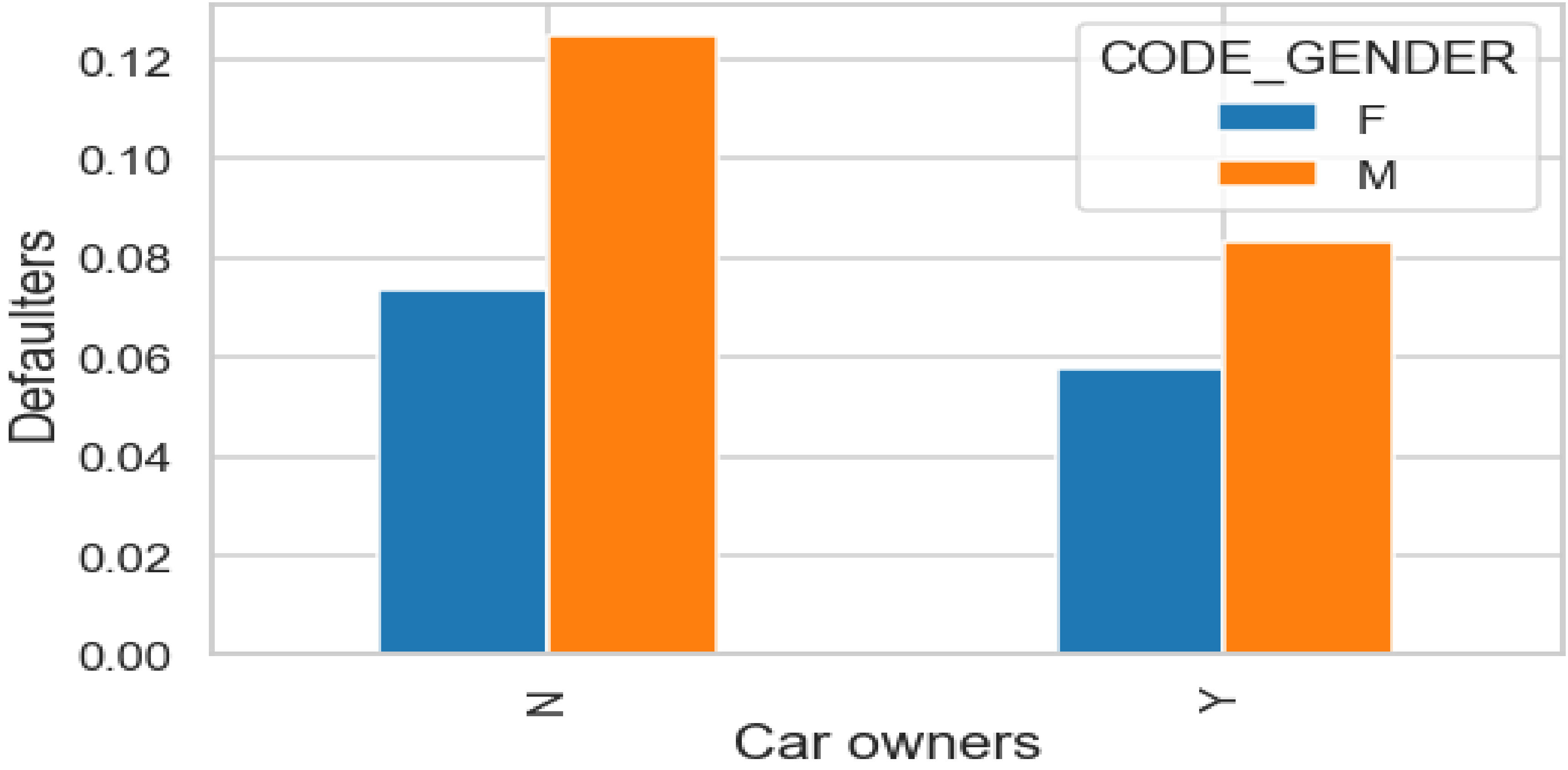
TARGET VS FAMILY STATUS VS AGE BUCKET



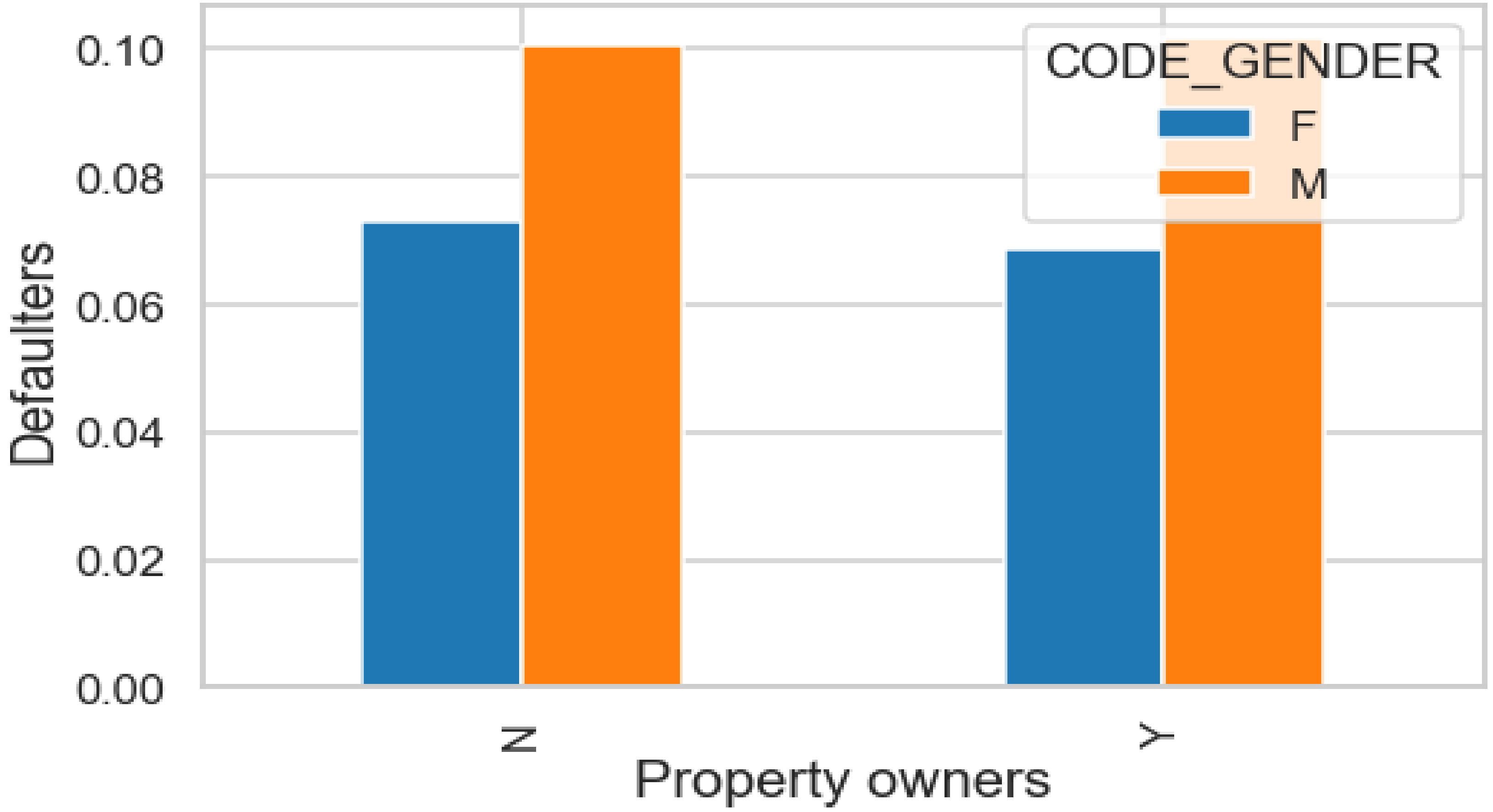
TARGET VS INCOME TYPE VS CODE GENDER



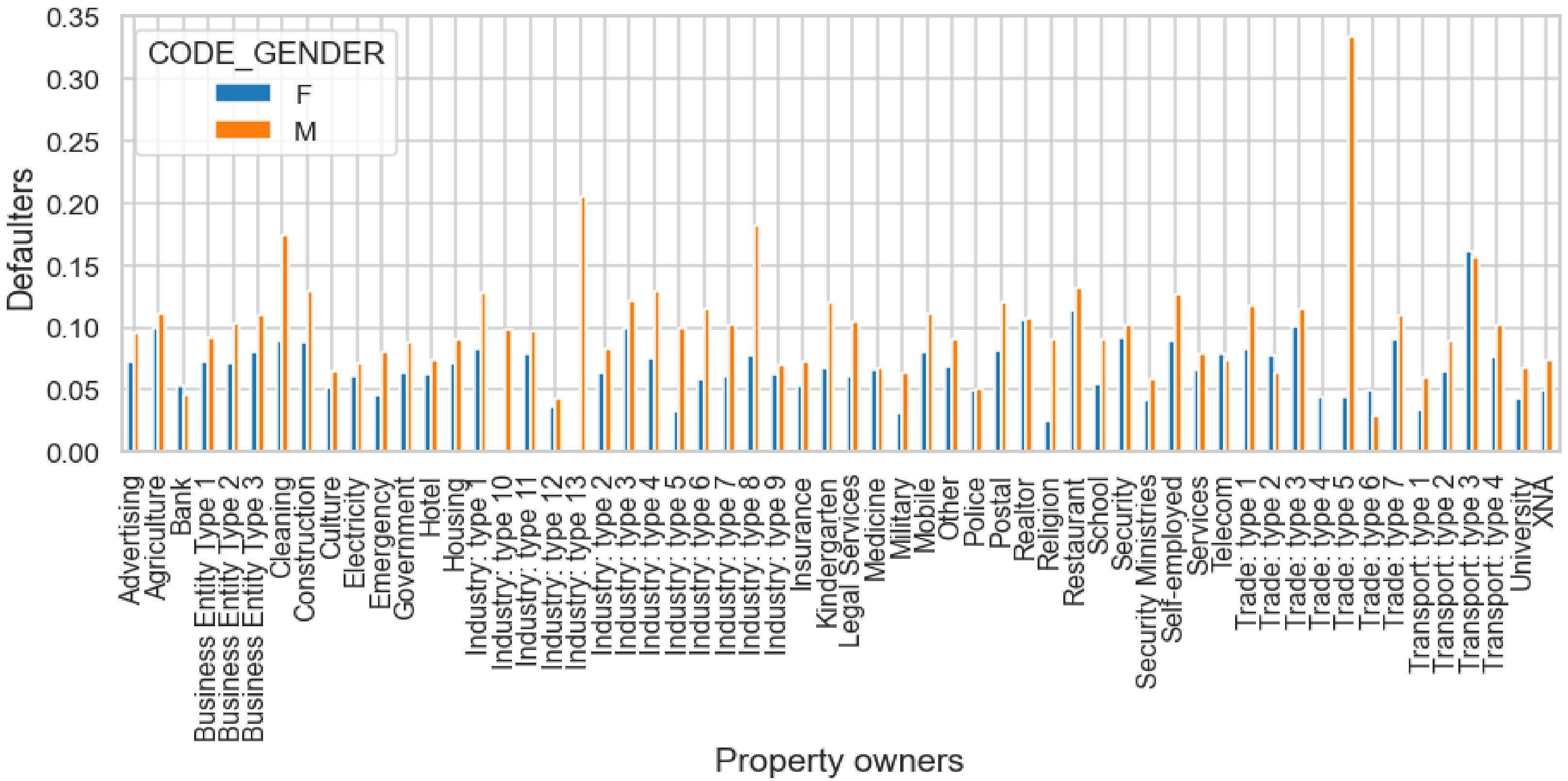
TARGET VS FLAG OWN CAR VS CODE GENDER



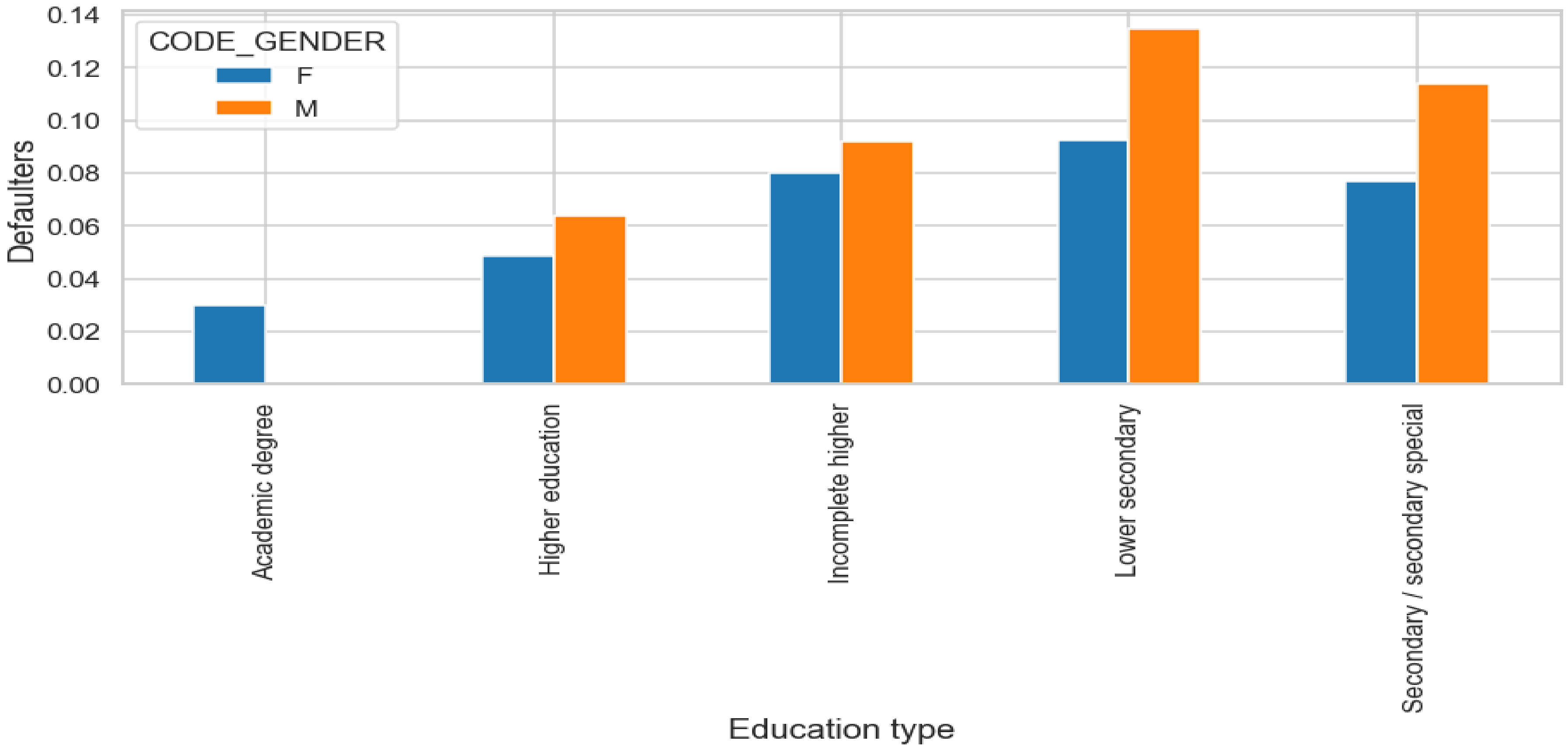
TARGET VS FLAG OWN REALTY VS CODE GENDER



TARGET VS ORGANIZATION TYPE VS CODE GENDER



TARGET VS EDUCATION TYPE VS CODE GENDER



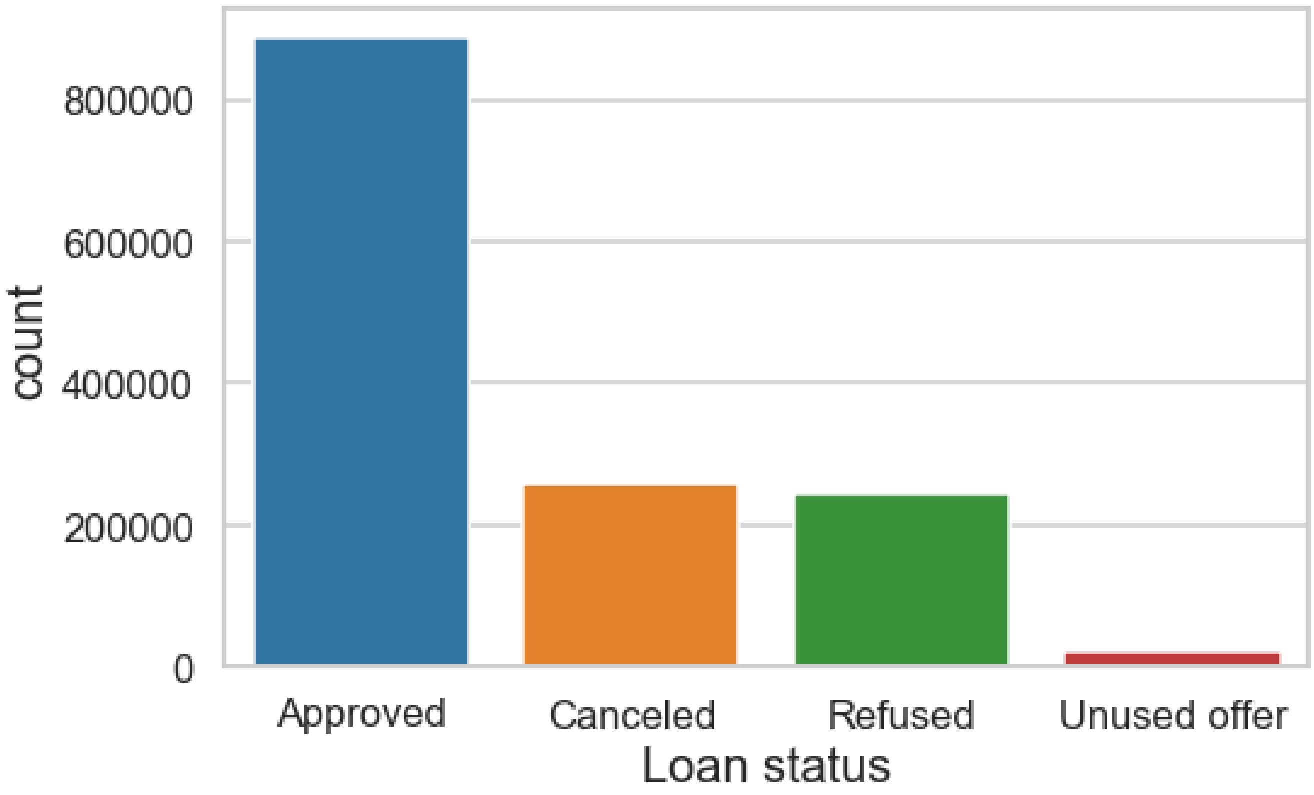
CORRELATION BETWEEN NUMERICAL VALUES

- FOR DEFAULTER

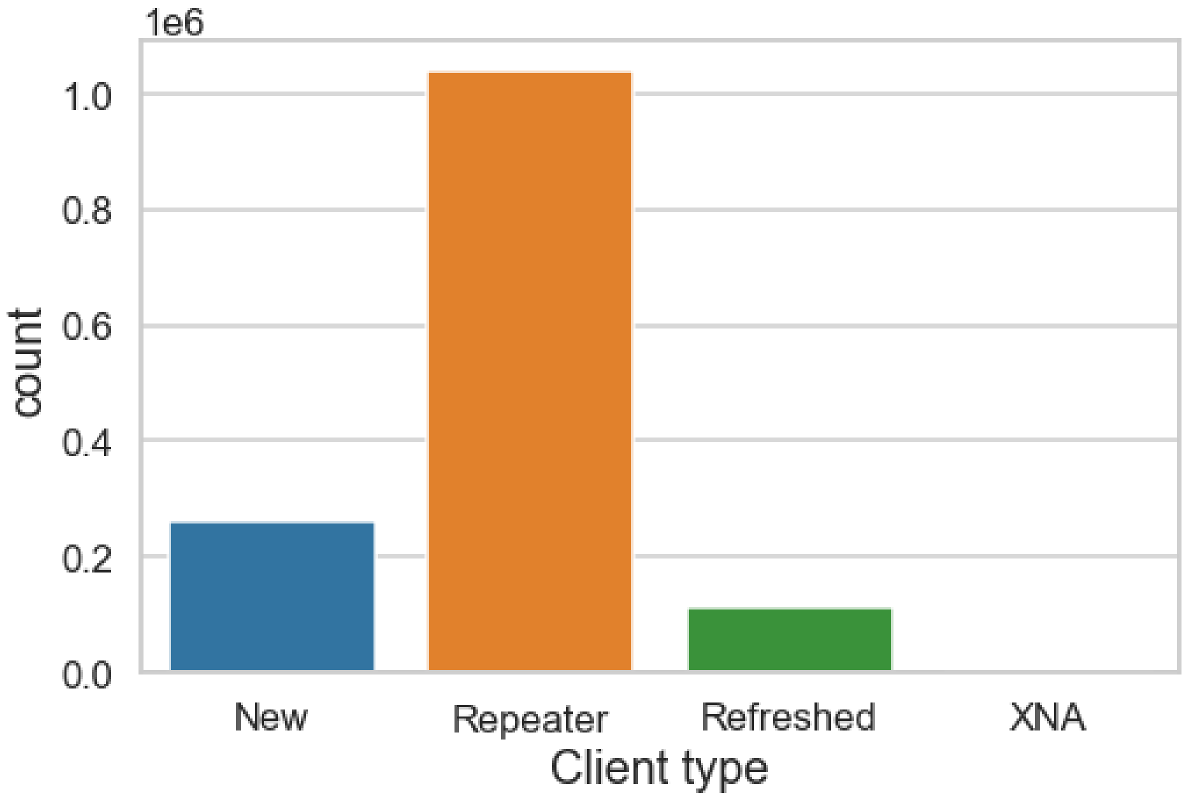
FOR NON DEFAULTERS

AFTER MERGING APPLICATION DATA AND PREVIOUS DATASET

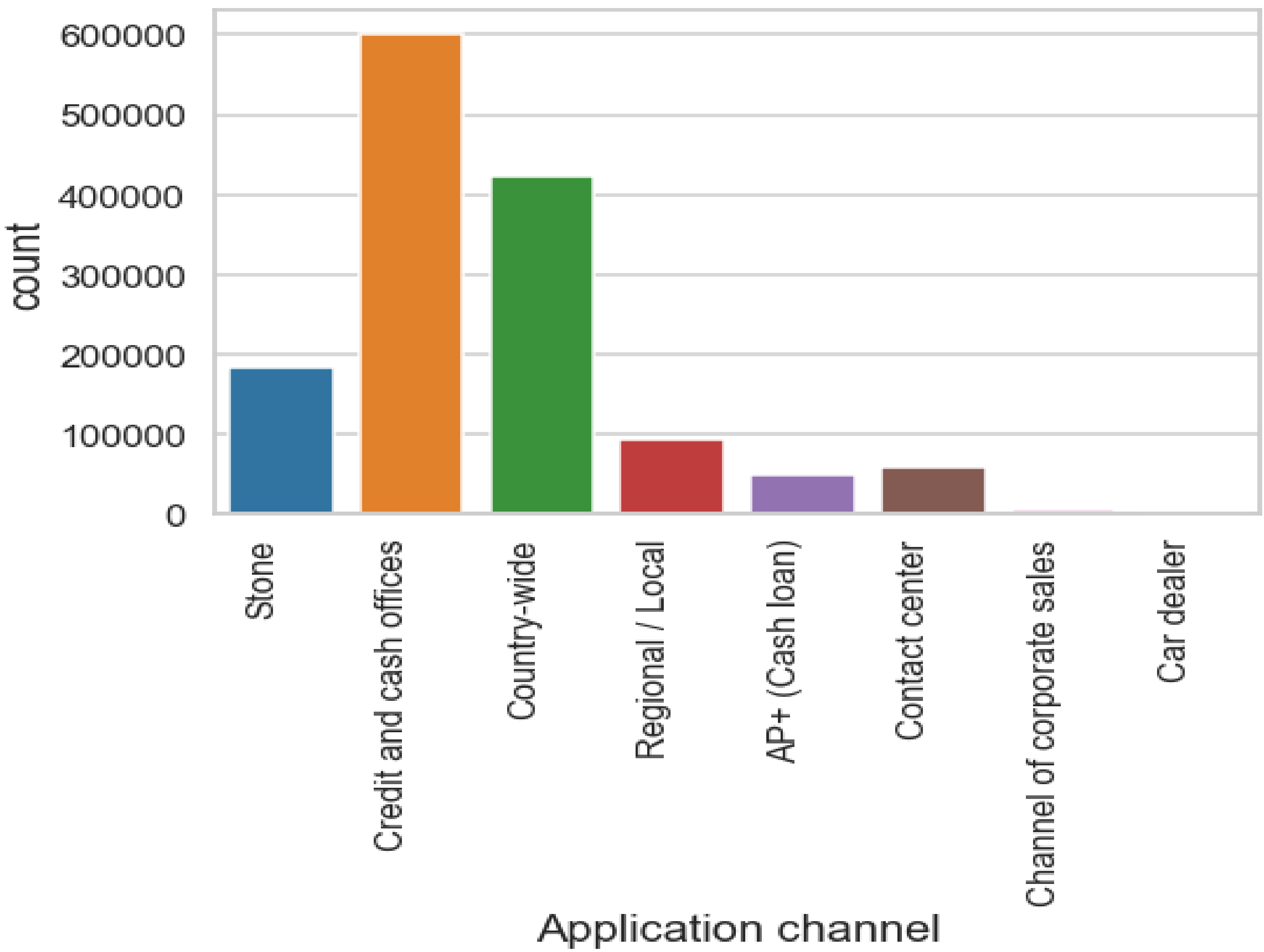
NAME CONTRACT STATUS



NAME CLIENT TYPE



CHANNEL TYPE



CONCLUSION

- BANK SHOULD FOCUS LESS ON INCOME TYPE “WORKING” AS THEY ARE HAVING MOST NUMBER OF UNSUCCESSFUL PAYMENT
- Also the loan purpose repair having higher number of unsuccessful payments
- Female clients on maternity leave should not be targeted they have no records of payments
- Clients should not be targeted based on their education
- Bank should not focus on students have high rate of payment issues
- Bank should not target clients who have car
- Bank should not target female clients as they are highest repayers

THANK YOU