CREDIT EDA CASE STUDY
(C62 NOVEMBER BATCH 2023)

BATCH ID : 5674
NAME : MALINI S

INTRODUCTION

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 was learned in the EDA module,

- It develops a basic understanding of risk analytics in banking and financial services

- Also, to understand how data is used to minimize the risk of losing money while lending to customers.were

BUSINESS OBJECTIVES

- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables that are strong indicators of default.
- The company can utilize this knowledge for its portfolio and risk assessment.
- This case study aims to identify patterns that indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of the 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.
- To develop your understanding of the domain, you are advised to independently research a little about risk analytics -understanding the types of variables and their significance should be enough.

ALGORITHM

Step 1: Import necessary Python libraries.
Step 2: Reading the dataset.
    A. Conversion of data into a data frame
    B. Examining the data frame
Step 3: Data cleaning
    1. Calculating the percentage of null values in each column
    2. Finding the number of columns whose percentage of null values more than 40%
    3. Removing/Dropping 40 columns whose percentage of null values greater than 40
    4. Checking the dimensions of the data frame after removing 40 columns
    5. Checking columns whose null values percentages are less than 40%
    6. Handling Missing values:
        A. Segregation of numerical and categorical column
        B. Imputation on the numerical and categorical column

ALGORITHM

7. Checking for any other null values in the data frame other than "OCCUPATION_TYPE"
8. Handing errors:
    1. On thorough inspection of each column, some of the missing values to be handled
        A. Replacing XNA values in CODE_GENDER
        B. Negative values to be replaced
        C. Replacing binary values
        D. Replacing “XNA” values

    2. Standardizing datatypes in each column
    3. Handling Outliers
    4. Binning of continuous variables

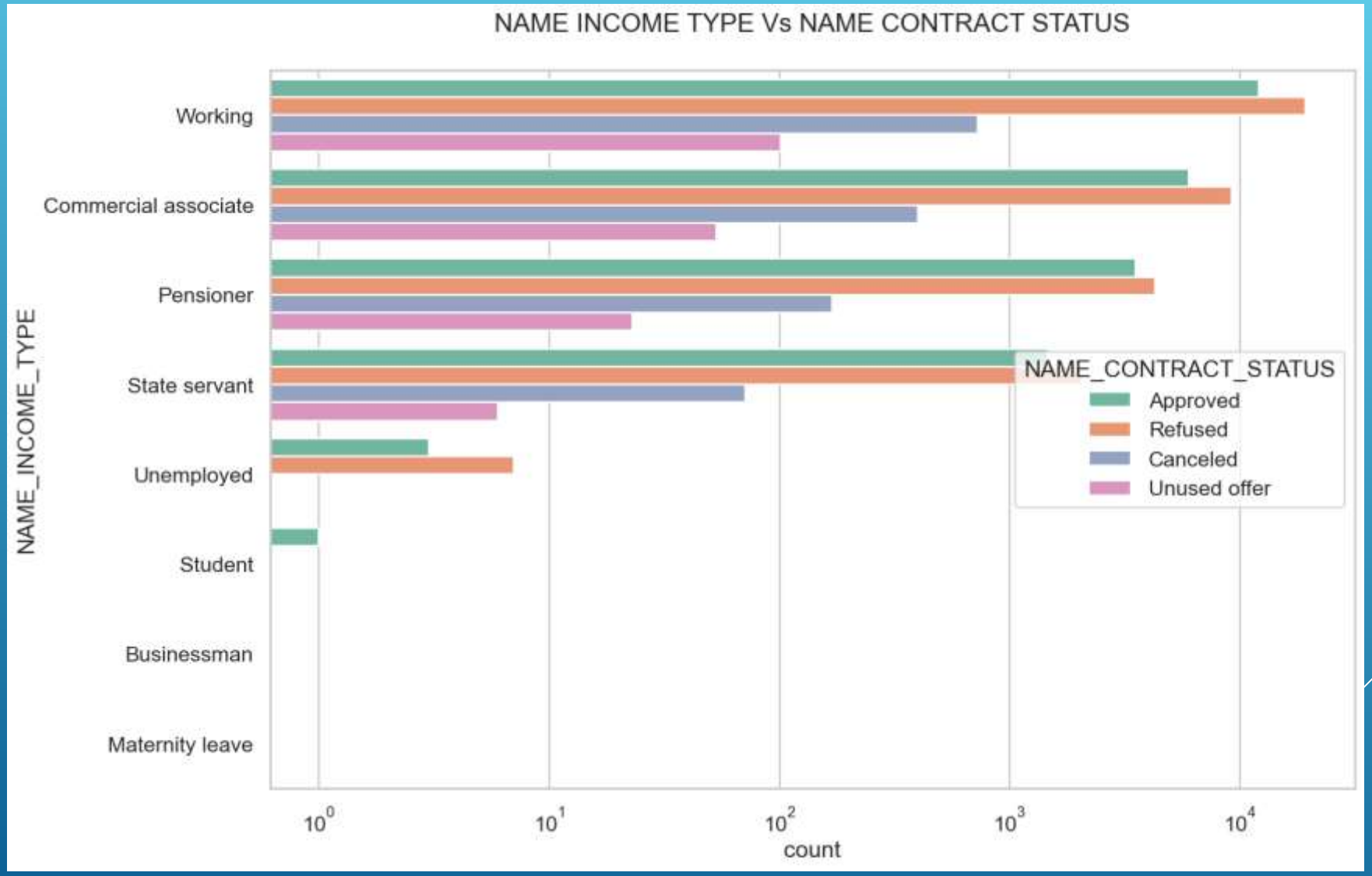
ALGORITHM

Step 4: Checking imbalance percentage
Step 5: Analysis
    1. Univariate Analysis
        a. Categorical columns
        b. Numerical columns
    2. Bivariate Analysis
Step 6: Correlation
    1. pair plot for target0
    2. pair plot for target1
Step 7: Load the previous application data (Repeat the process from 1 to 5)
Step 8: Summary
    1. Inferences
    2. Opinion

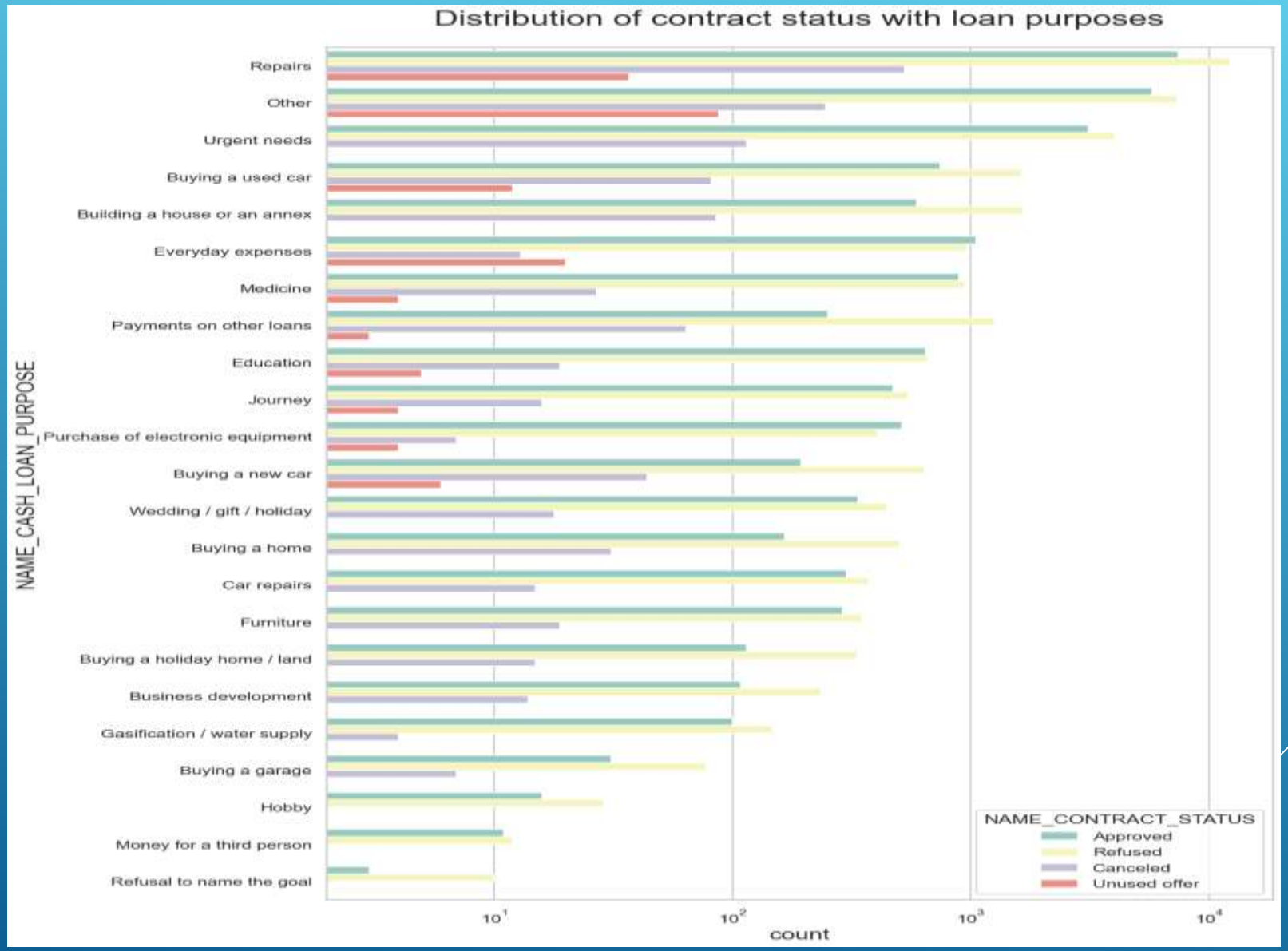
RESULTS OF UNIVARIATE ANALYSIS


NAME_CONTRACT_STATUS
Approved
Refused
Canceled
Unused offer
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE
Children
Family
Group of people
Other_A
Other_B
Unaccompanied
NAME_TYPE_SUITE
Stone
Regional / Local
CHANNEL_TYPE
FRIDAY
WEDNESDAY
MONDAY
TUESDAY
THURSDAY
SATURDAY
SUNDAY
WEEKDAY_APPR_PROCESS_START_PREV
high
low_action
low_normal
middle
XNA
NAME_YIELD_GROUP
New
Repeater
Refreshed
NAME_CLIENT_TYPE

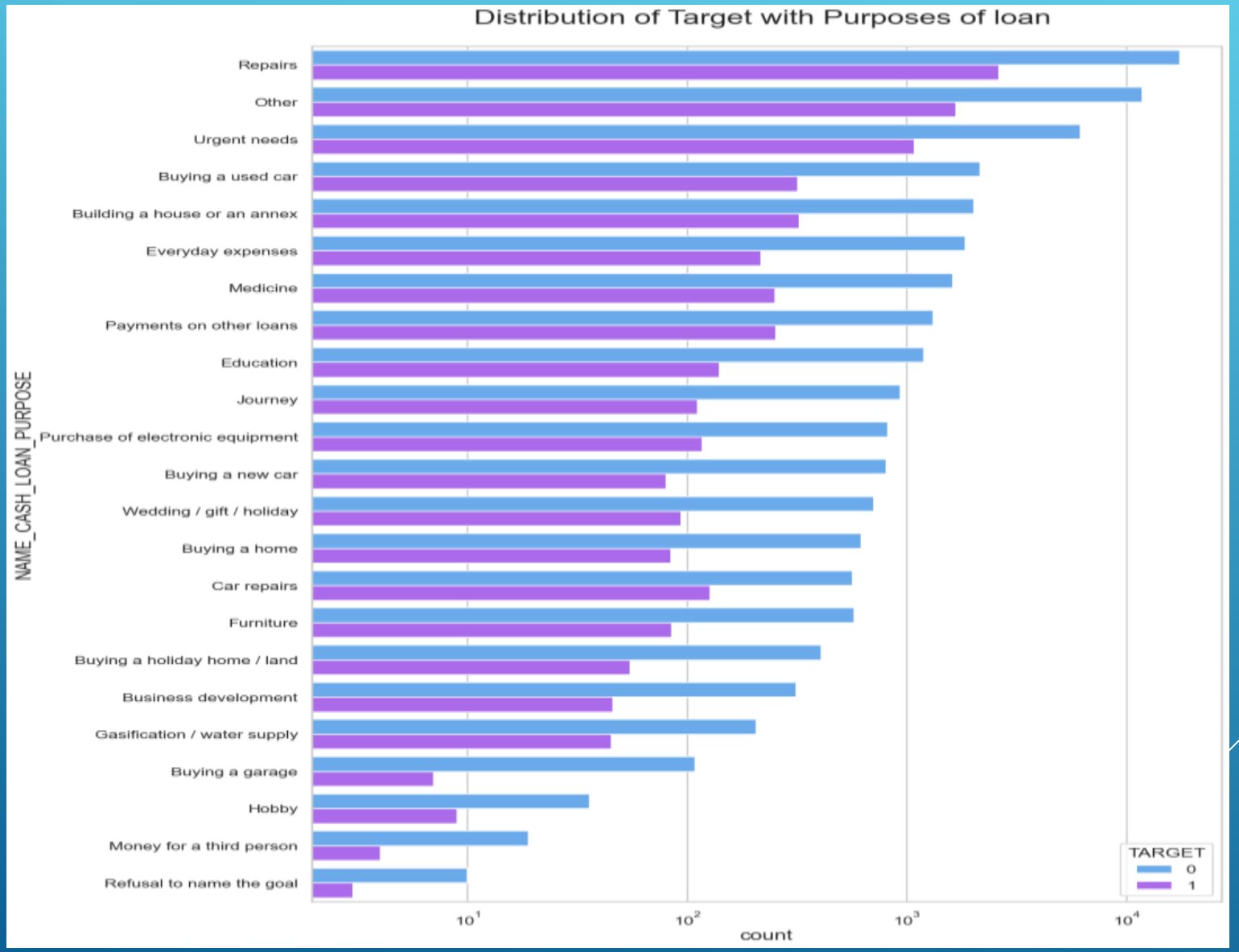
NAME INCOME TYPE VS NAME CONTRACT STATUS



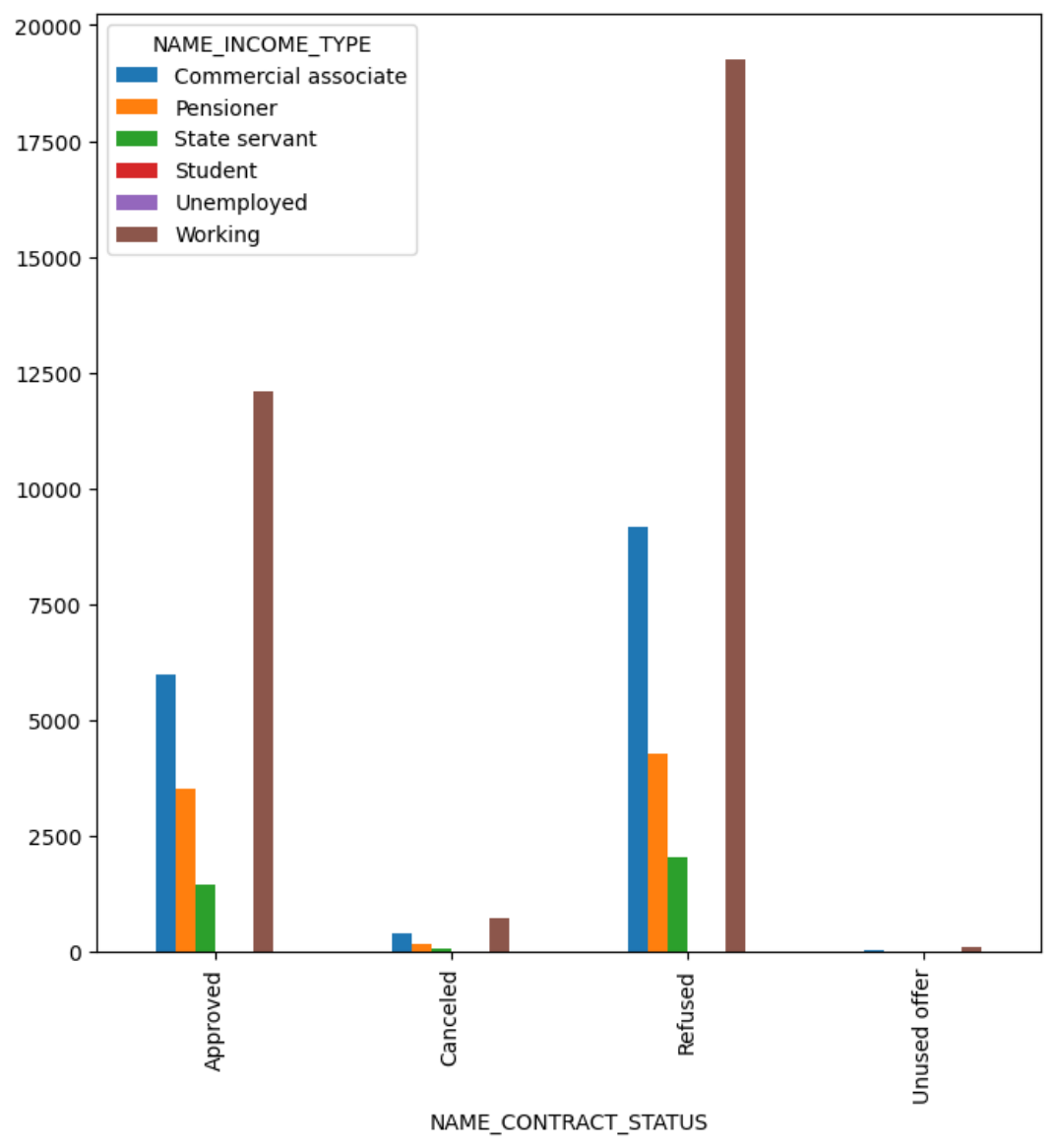
DISTRIBUTION OF CONTRACT STATUS WITH LOAN PURPOSES

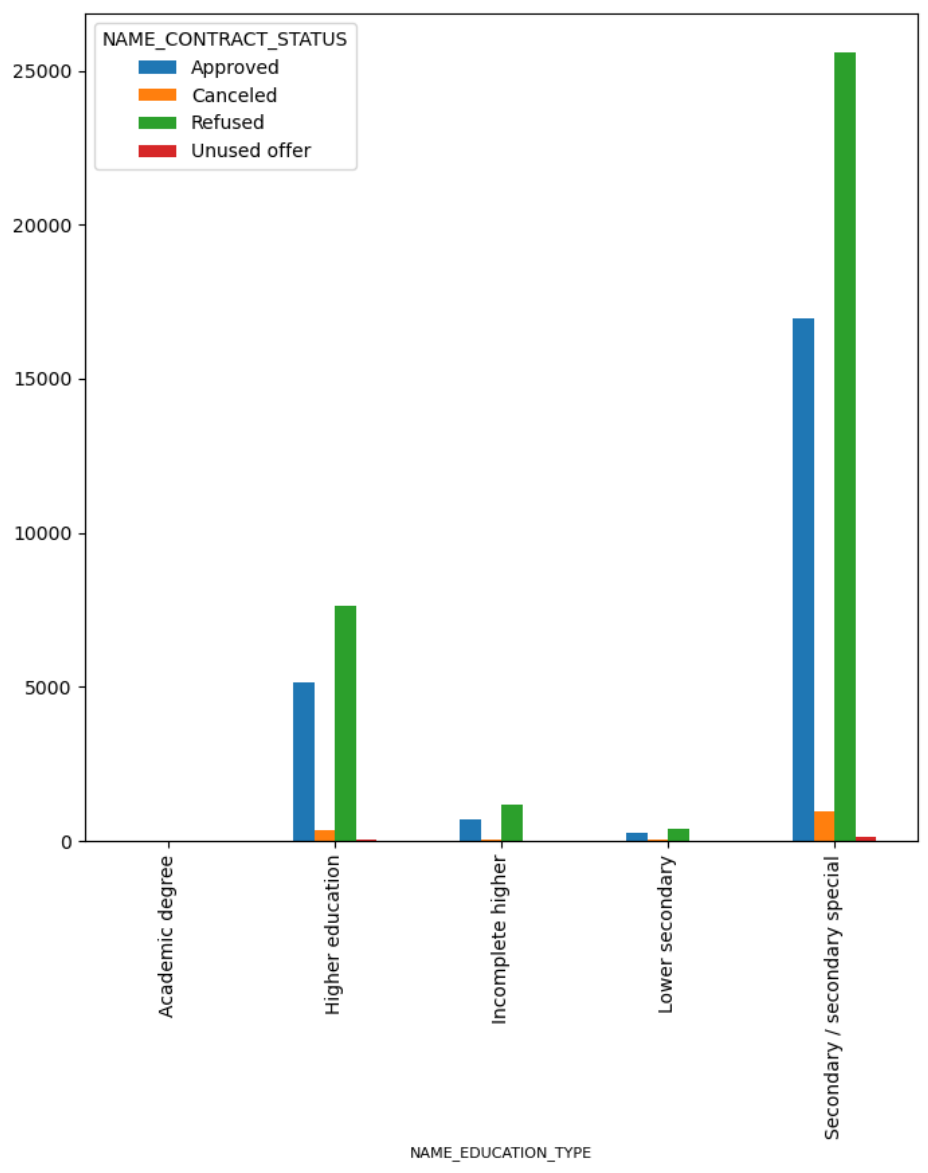


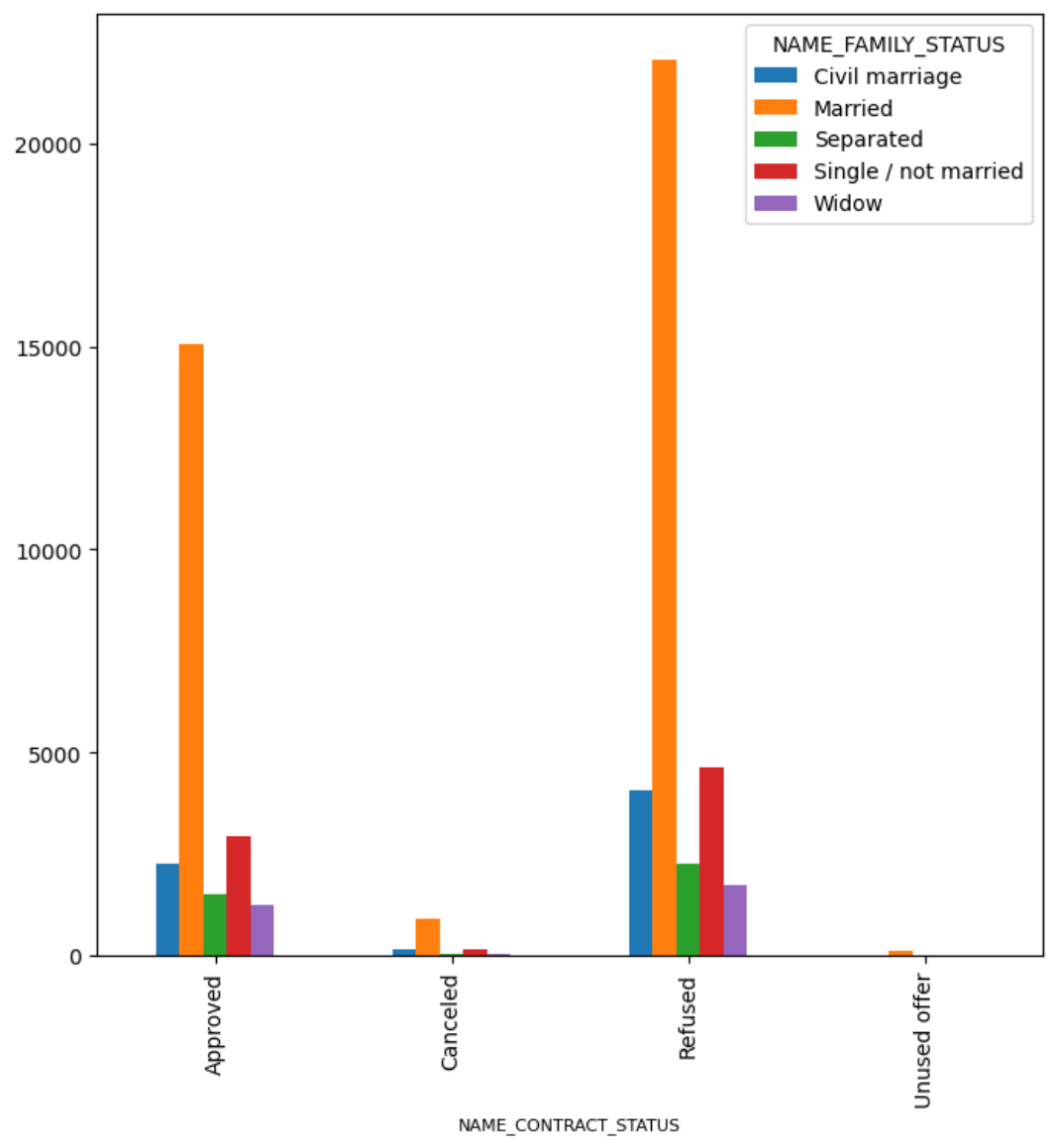
DISTRIBUTION OF TARGET WITH PURPOSES OF LOAN

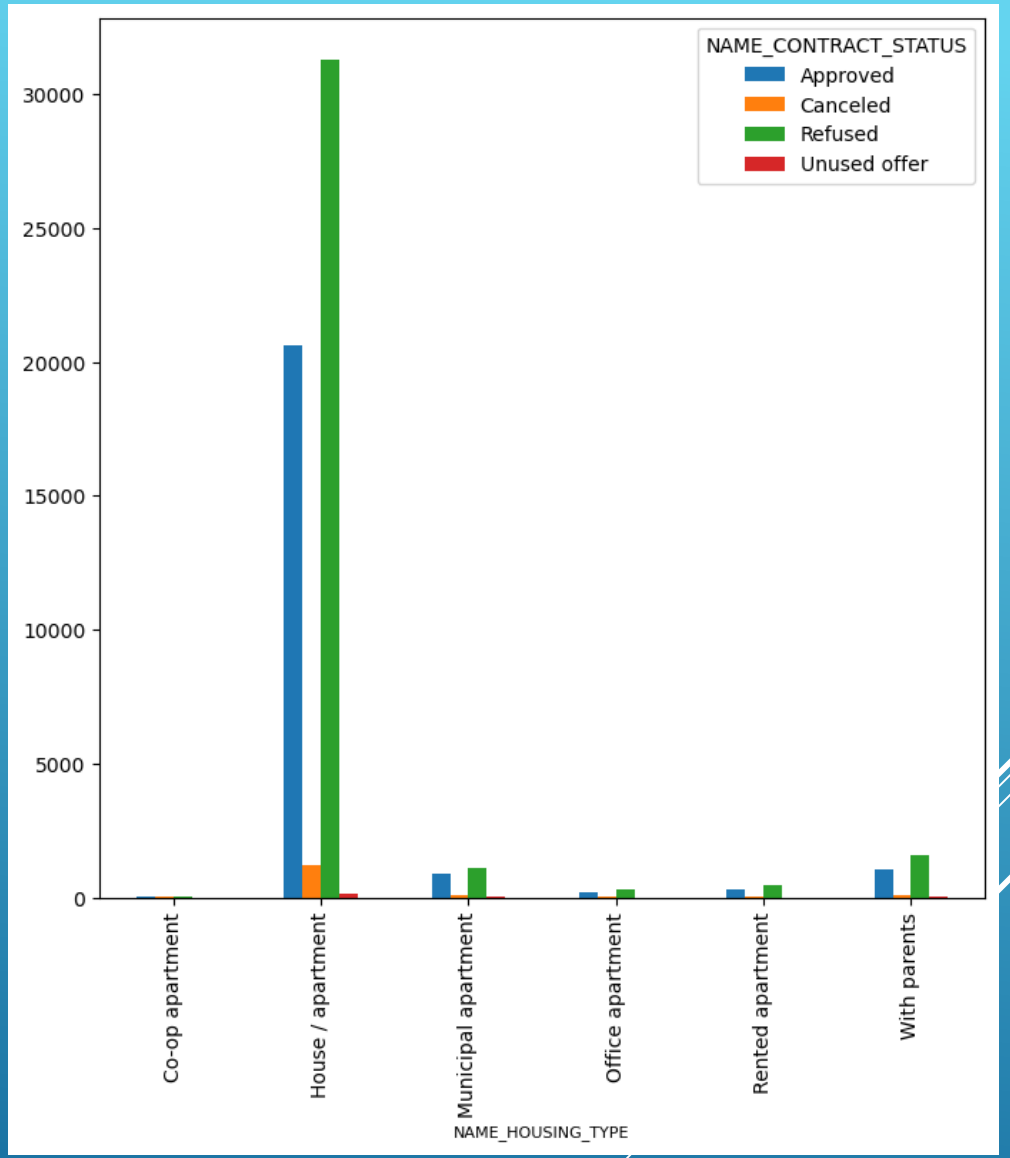


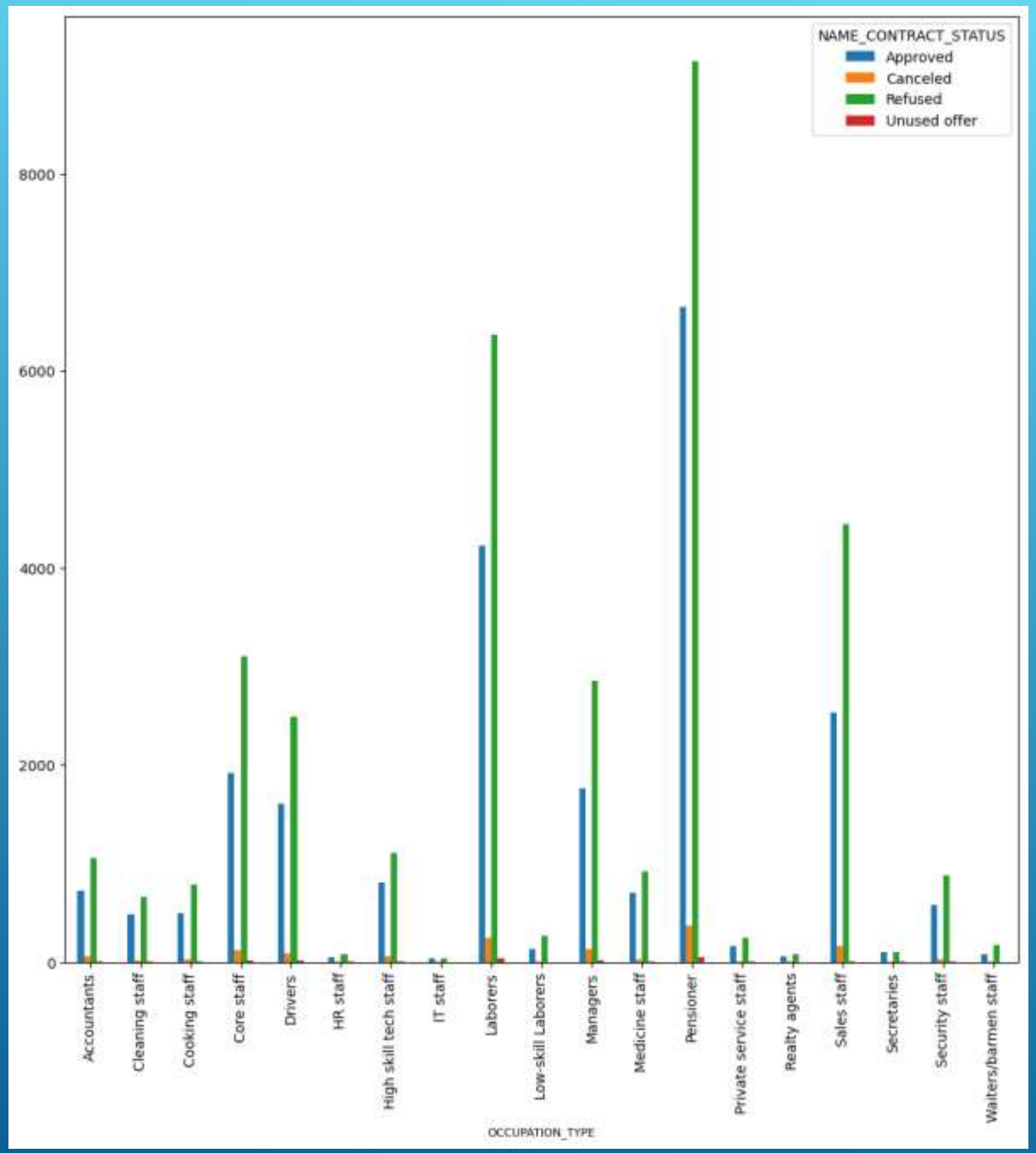
RESULTS OF BIVARIATE ANALYSIS

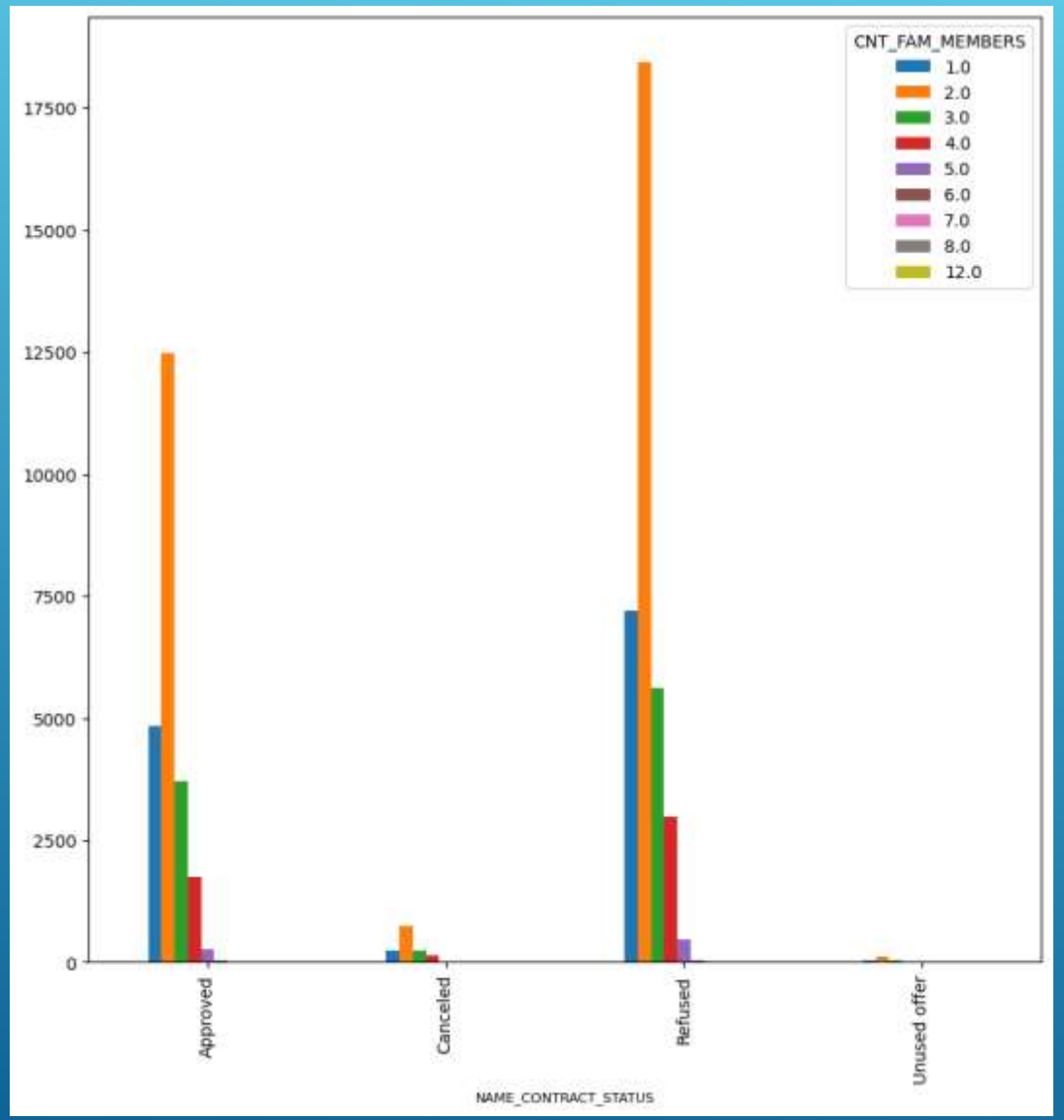

NAME_INCOME_TYPE
Commercial associate
Pensioner
State servant
Student
Unemployed
Working
20000
17500
15000
12500
10000
7500
5000
2500
0
Approved
Canceled
Refused
Unused offer
NAME_CONTRACT_STATUS

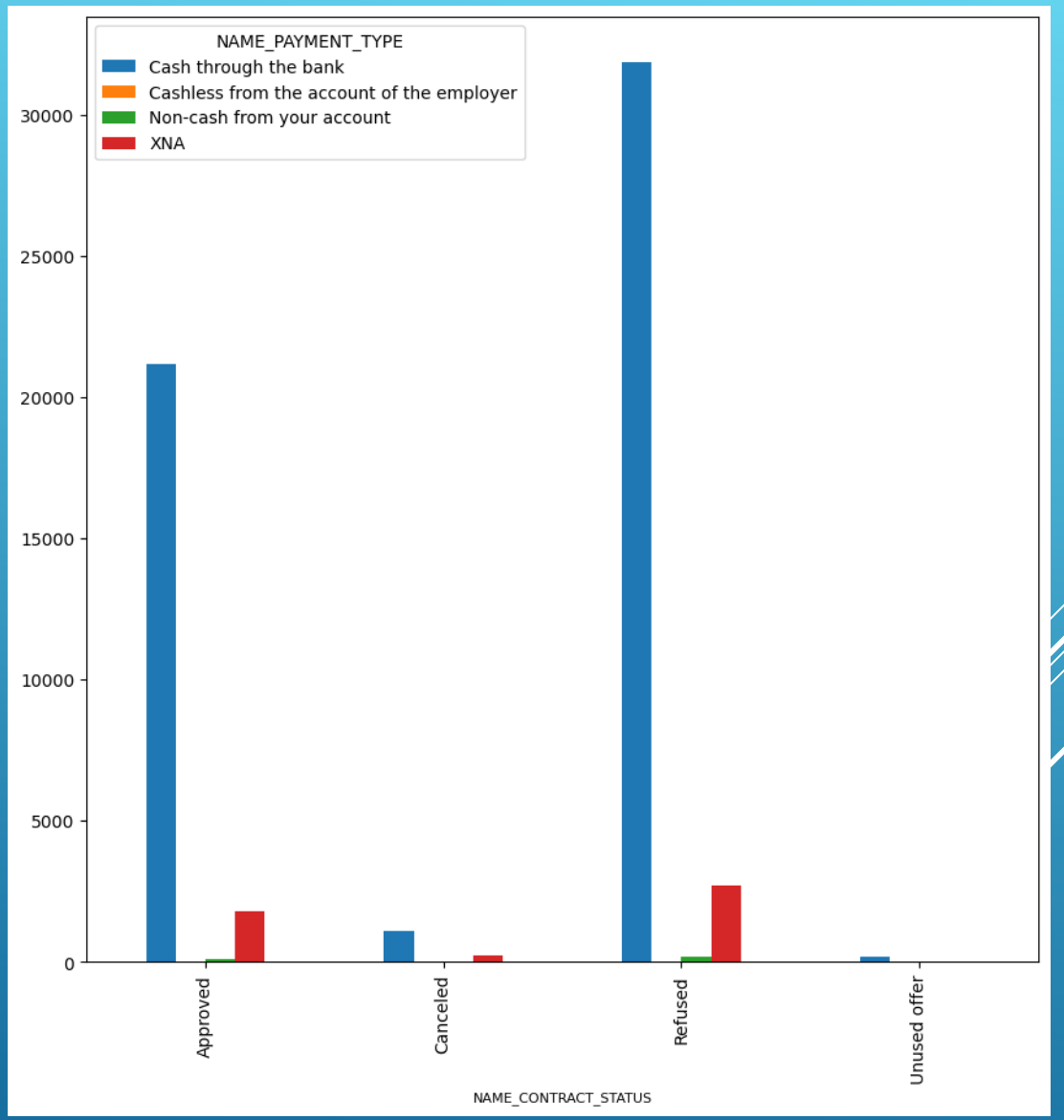

NAME_CONTRACT_STATUS
Approved
Canceled
Refused
Unused offer
25000
20000
15000
10000
5000
0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
NAME_EDUCATION_TYPE


NAME_FAMILY_STATUS
Civil marriage
Married
Separated
Single / not married
Widow
20000
15000
10000
5000
0
Approved
Canceled
Refused
Unused offer
NAME_CONTRACT_STATUS


NAME_CONTRACT_STATUS
Approved
Canceled
Refused
Unused offer
30000
25000
20000
15000
10000
5000
0
Co-op apartment
House / apartment
Municipal apartment
Office apartment
Rented apartment
With parents
NAME_HOUSING_TYPE


NAME_CONTRACT_STATUS
Approved
Canceled
Refused
Unused offer
8000
6000
4000
2000
0
Accountants
Cleaning staff
Cooking staff
Core staff
Drivers
HR staff
High skill tech staff
IT staff
Laborers
Low-skill Laborers
Managers
Medicine staff
Pensioner
Private service staff
Realty agents
Sales staff
Secretaries
Security staff
Waiters/barmen staff
OCCUPATION_TYPE


CNT_FAM_MEMBERS
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
12.0
17500
15000
12500
10000
7500
5000
2500
0
Approved
Canceled
Refused
Unused offer
NAME_CONTRACT_STATUS


NAME_PAYMENT_TYPE
Cash through the bank
Cashless from the account of the employer
Non-cash from your account
XNA
30000
25000
20000
15000
10000
5000
0
Approved
Canceled
Refused
Unused offer
NAME_CONTRACT_STATUS

INFERENCES:

- Cash loans refusal is higher than cash loan approval.
- The client without companions has a higher approval and refusal rate than the clients who came along with their spouse and family members.
- The clients were acquired mostly by credit and cash offices and least by car dealers.
- The clients who applied previously have more refusal rates than approval rates.
- The loan refusal rate is higher than the loan approval rate. Additionally, the loan approval/refusal process is done and higher during weekdays. These processes are done
- during weekends also.
- Working professionals and pensioners have more approval and refusal rates. few student loans also approved may be student loans for studying. Few loans are sanctioned for unemployed.

INFERENCES:

- Car loans, Medicine loans, personal loans, and journey loans are unused.
- Target 1 clients i.e., those with payment difficulties are seen in the repair, urgent needs category.
- The loan refusal rate is larger in the working and commercial associate category.
- Loans are rejected more in secondary/secondary special and higher education.
- Married people applied for more loans.
- The clients who applied for a loan own a house/apartment.
- Pensioners, laborers, and sales staff applied for more loans.
- IT staff has the lowest approval and refusal rate.
- The clients who have a single child applied for the loan are sanctioned than declined.
- The clients who have two children have the most denied loans.
- The client whose loans are sanctioned opted for the cash through the bank

OPINION:

- The bank can concentrate on business and pensioners.
- The people who work in the IT industry have the lowest approval and rejection rate. These people may be targeted.

THANK YOU!