Credit EDA Assignment

Anmol Bhandari

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 you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

BUSINESS UNDERSTANDING

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 to 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.

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.

STEPS FOR DATA UNDERSTANDING

Data Collection - Acquire the data from various sources

Data Profiling - Examine the basic structure of the data. This includes understanding the number of records, features, identifying primary keys, and looking at data type

Basic Statistics - Generate descriptive statistics (mean, median, mode, standard deviation, etc.) to understand the central tendency and dispersion of numerical

STEPS TO PERFORM EDA

1. Basic Information - Use methods like .describe(), .info() in Python to get basic information about the data (size, types of variables, initial statistics).

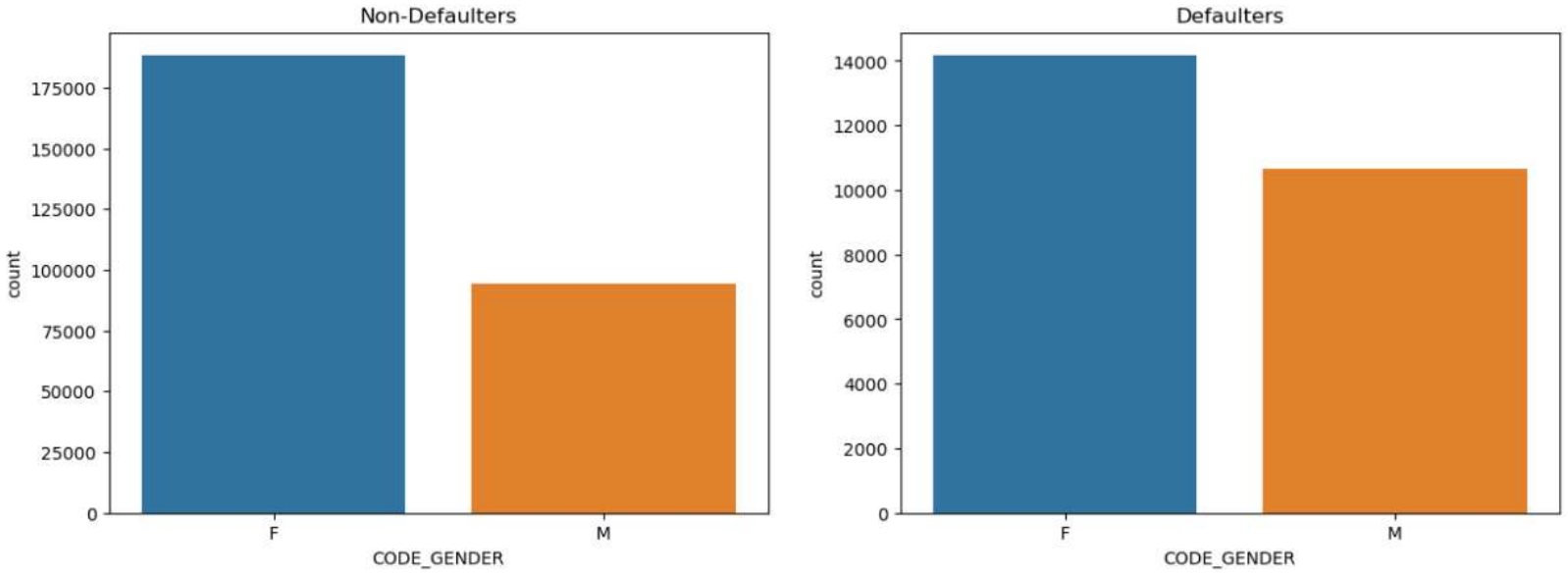
2. Missing Values: Identify and handle missing data through imputation or removal.

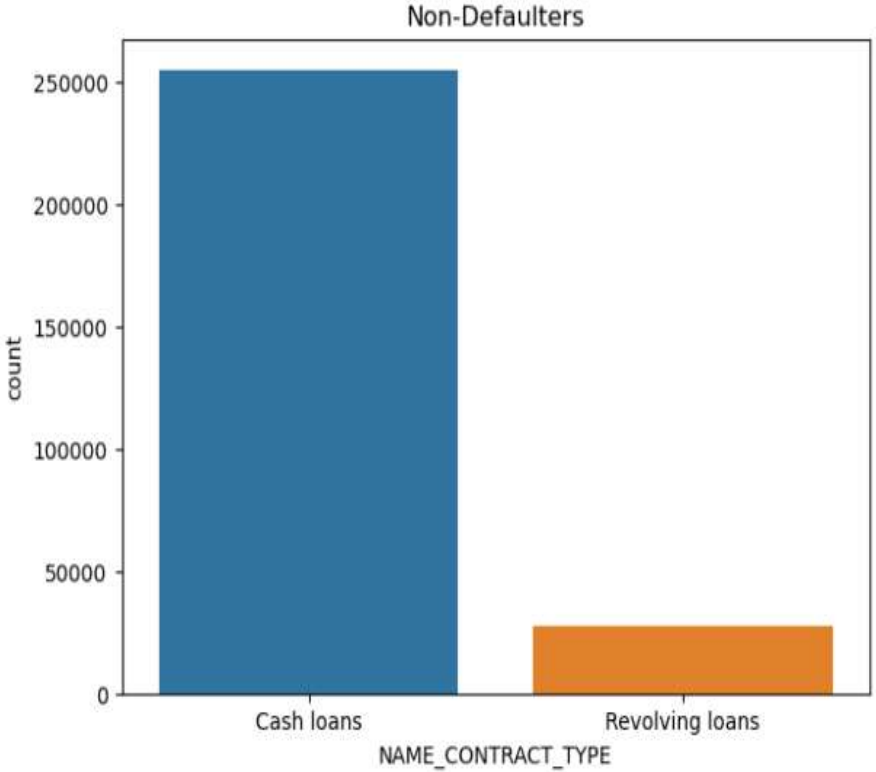
3. Outliers: Detect and handle outliers, if appropriate.

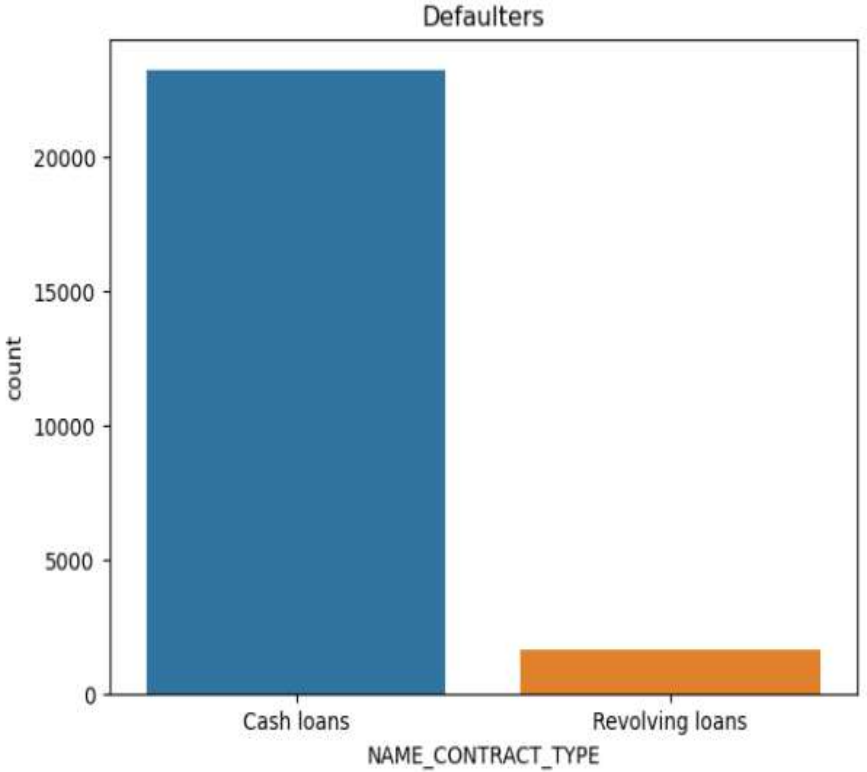
4. Inconsistencies: Correct any inconsistencies in data

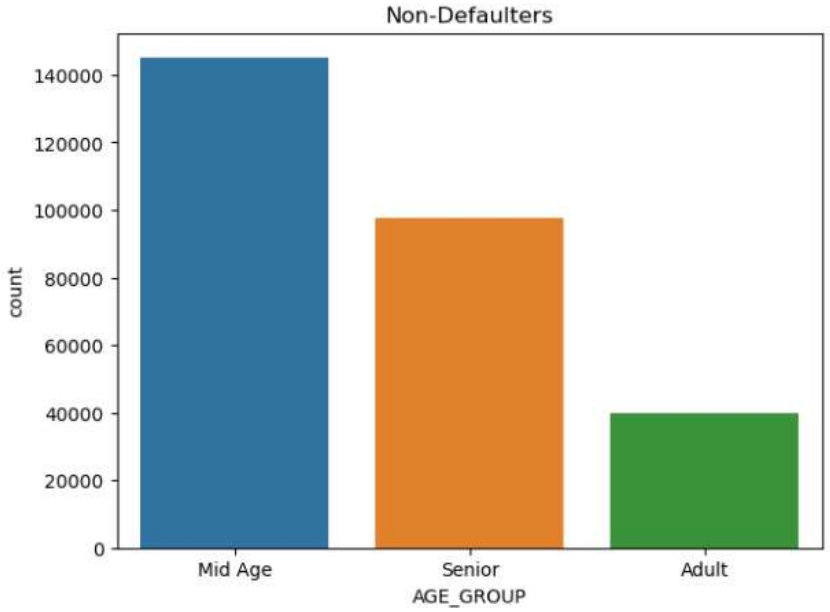
5. Data Types: Ensure each column is of the correct data type

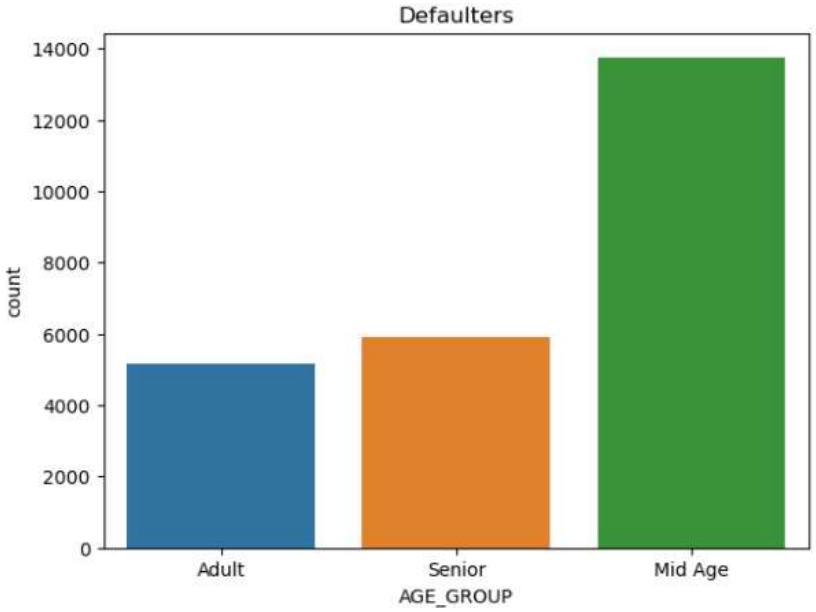
UNIVARIATE ANALYSIS FOR CATEGORICAL VARIABLES

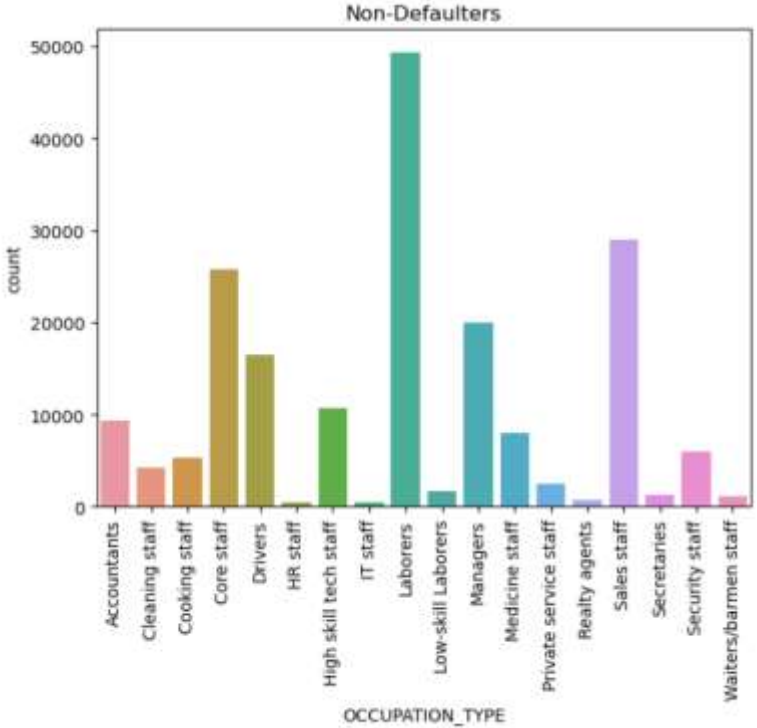


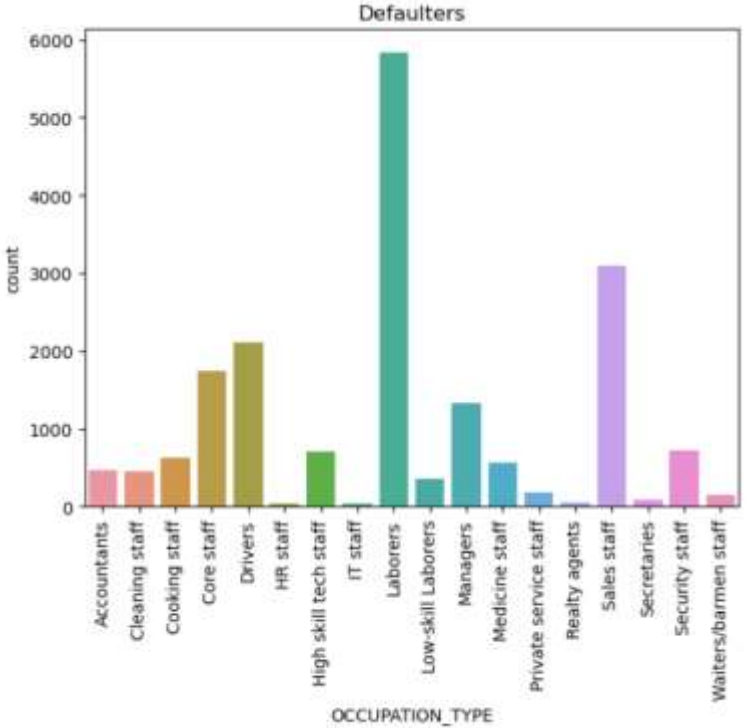

Non-Defaulters
250000
200000
150000
100000
50000
0
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

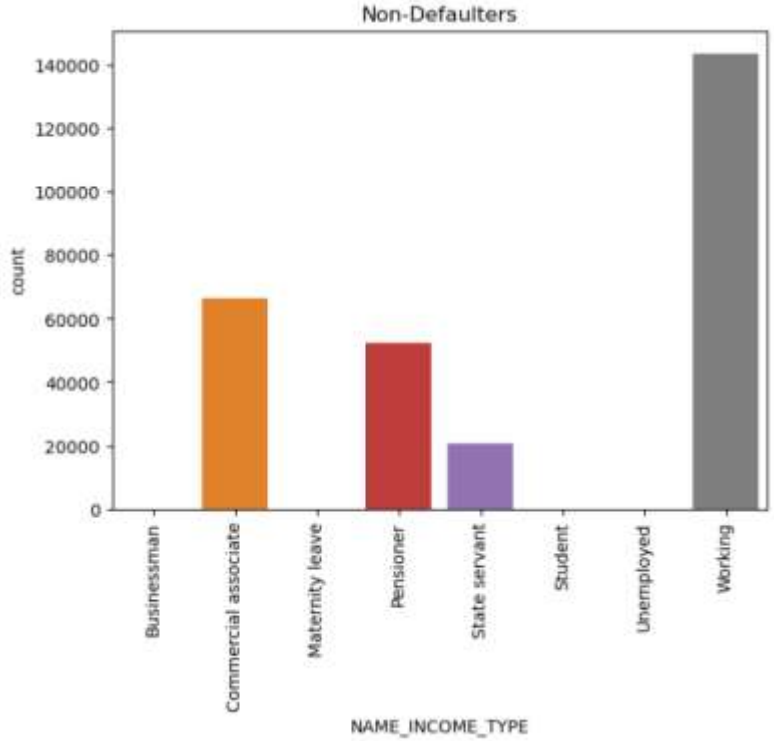

Defaulters
20000
15000
10000
5000
0
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

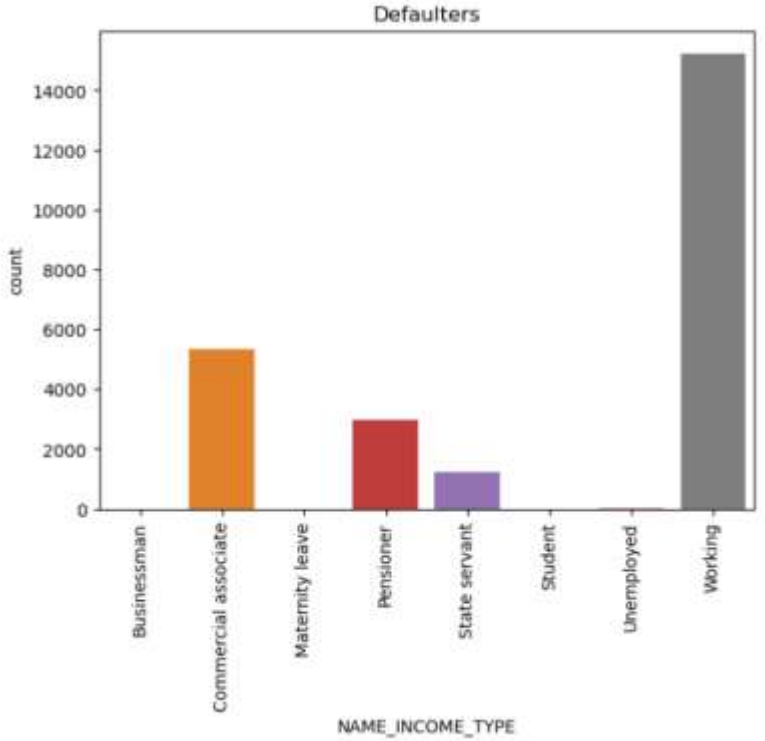

Non-Defaulters
140000
120000
100000
80000
60000
40000
20000
0
count
Mid Age
Senior
Adult
AGE_GROUP

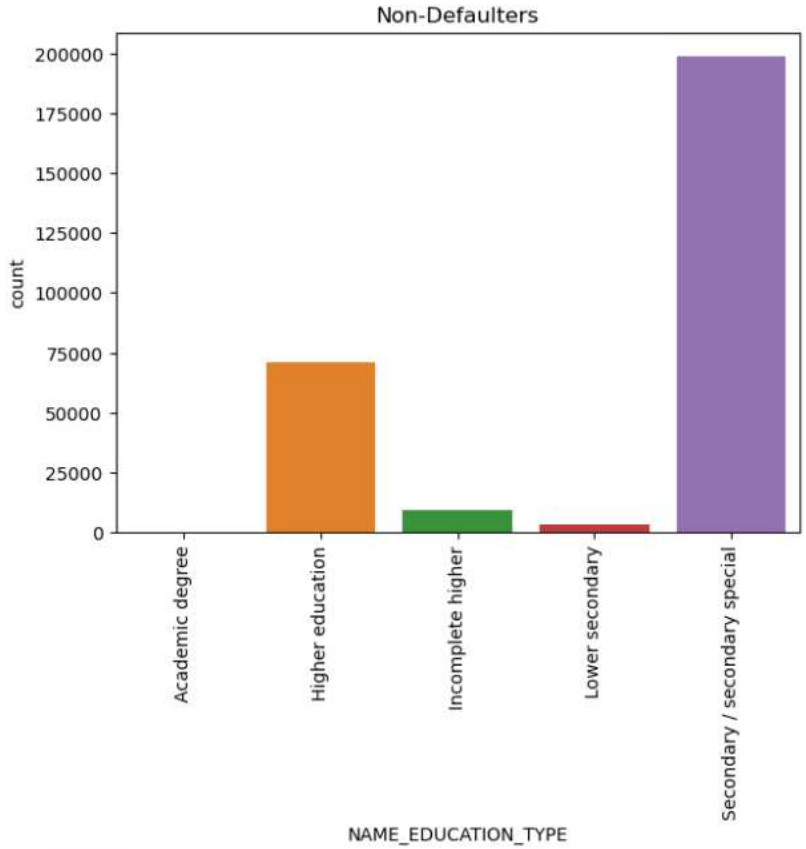

Defaulters
14000
12000
10000
8000
6000
4000
2000
0
count
Adult
Senior
Mid Age
AGE_GROUP

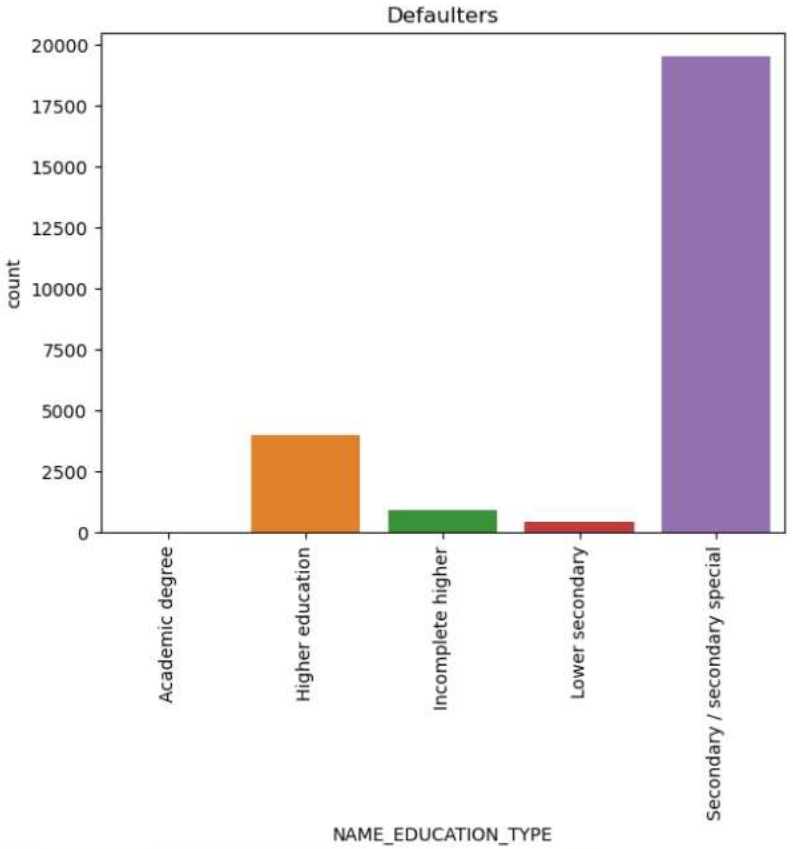

Non-Defaulters
count
50000
40000
30000
20000
10000
0
Accountants
Cleaning staff
Cooking staff
Core staff
Drivers
HR staff
High skill tech staff
IT staff
Laborers
Low-skill Laborers
Managers
Medicine staff
Private service staff
Realty agents
Sales staff
Secretaries
Security staff
Waiters/barmen staff
OCCUPATION_TYPE


Defaulters
count
6000
5000
4000
3000
2000
1000
0
Accountants
Cleaning staff
Cooking staff
Core staff
Drivers
HR staff
High skill tech staff
IT staff
Laborers
Low-skill Laborers
Managers
Medicine staff
Private service staff
Realty agents
Sales staff
Secretaries
Security staff
Waiters/barmen staff
OCCUPATION_TYPE

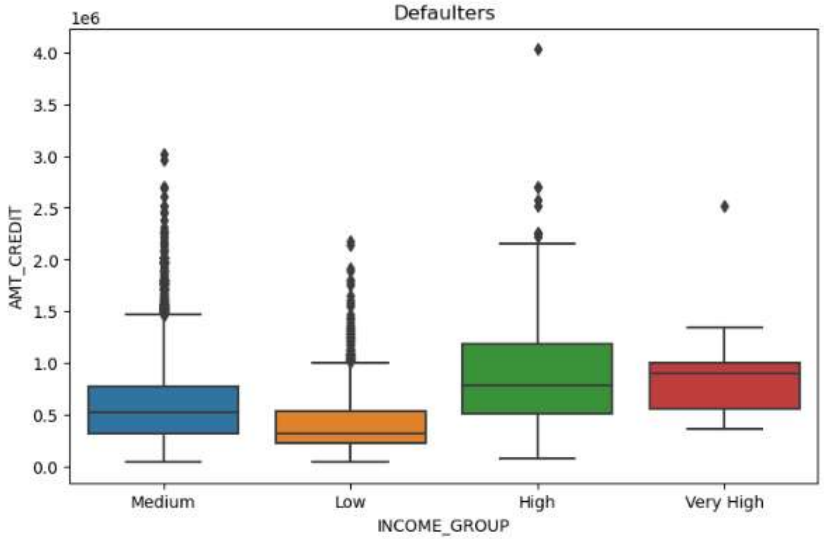

Non-Defaulters
count
140000
120000
100000
80000
60000
40000
20000
0
Businessman
Commercial associate
Maternity leave
Pensioner
State servant
Student
Unemployed
Working
NAME_INCOME_TYPE

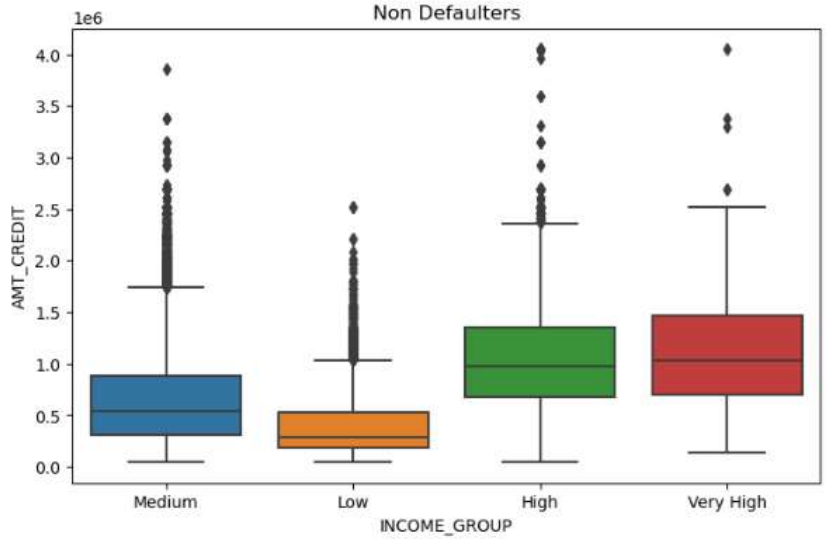

Defaulters
count
14000
12000
10000
8000
6000
4000
2000
0
Businessman
Commercial associate
Maternity leave
Pensioner
State servant
Student
Unemployed
Working
NAME_INCOME_TYPE

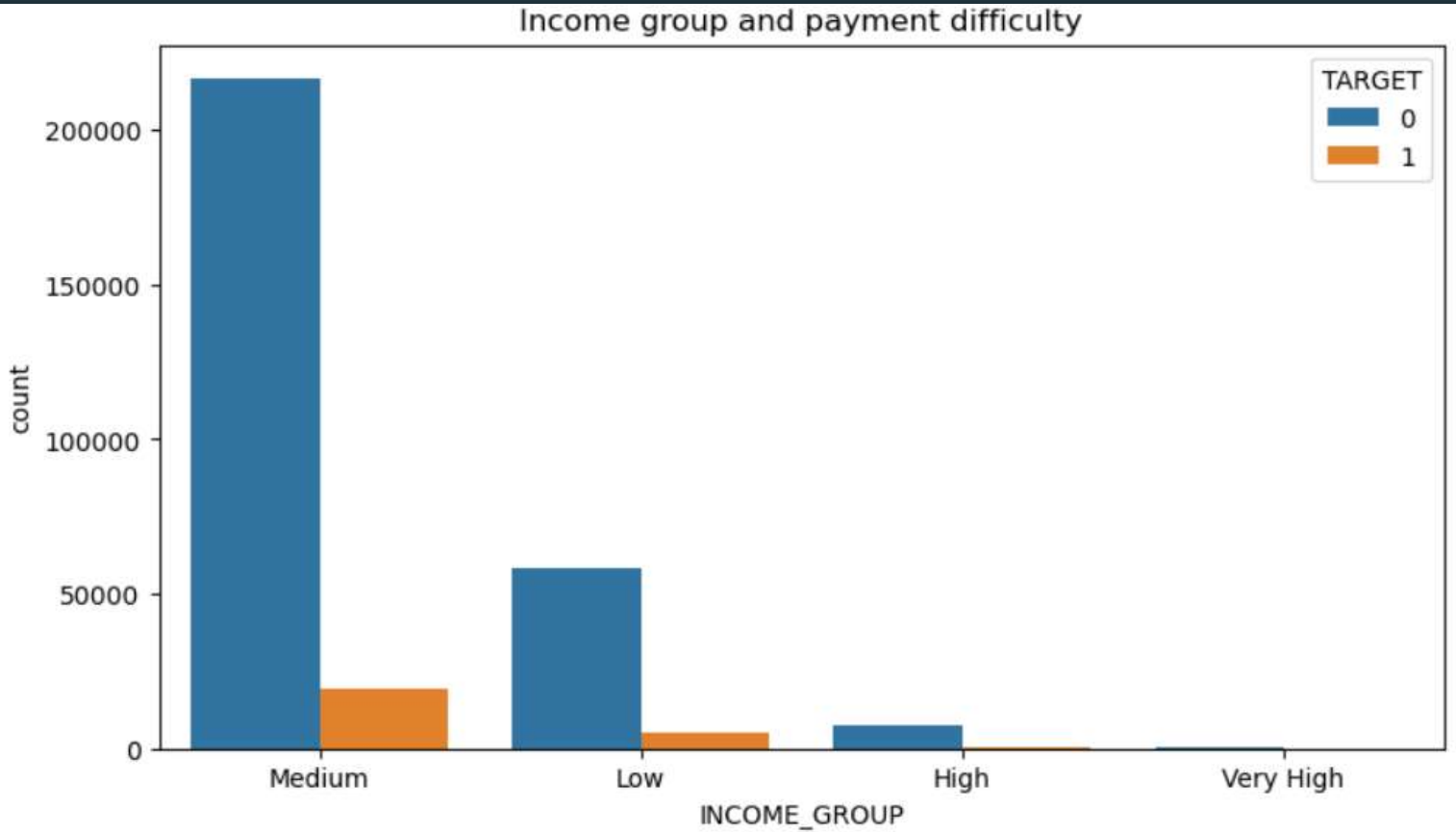

Non-Defaulters
count
200000
175000
150000
125000
100000
75000
50000
25000
0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
NAME_EDUCATION_TYPE

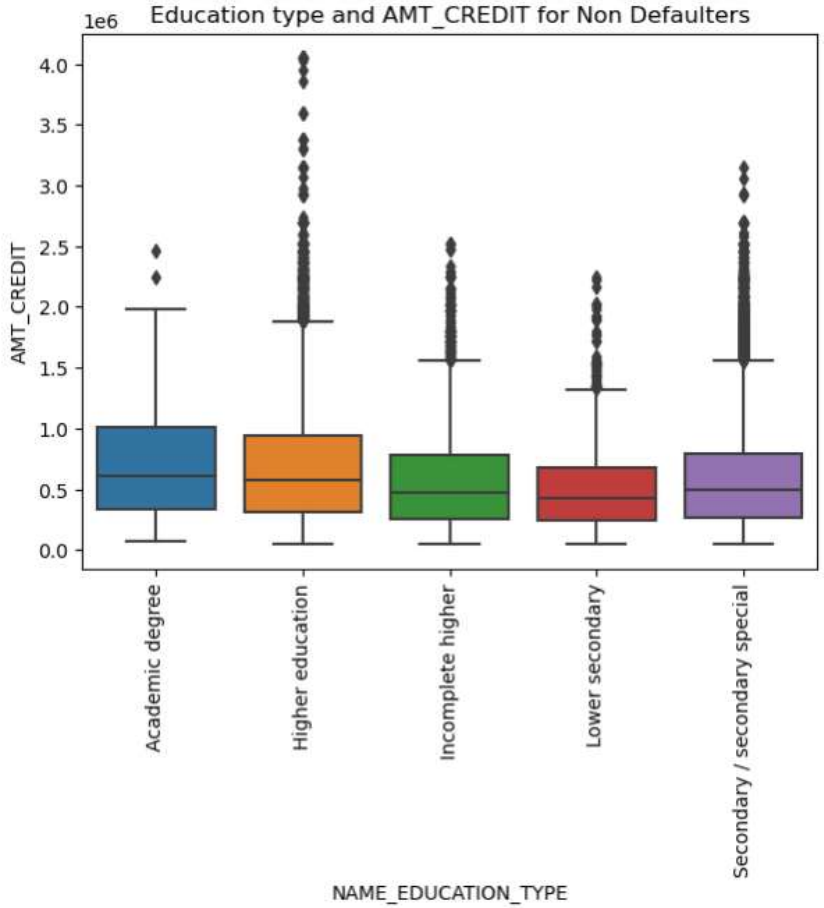

Defaulters
count
20000
17500
15000
12500
10000
7500
5000
2500
0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
NAME_EDUCATION_TYPE

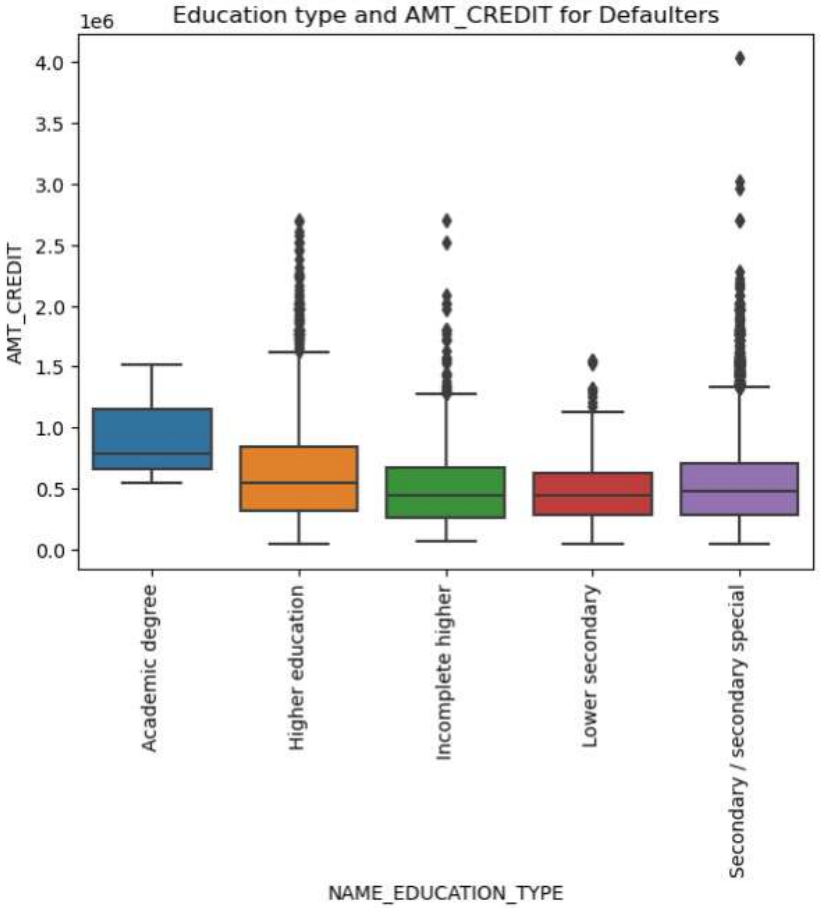
BIVARIATE ANALYSIS

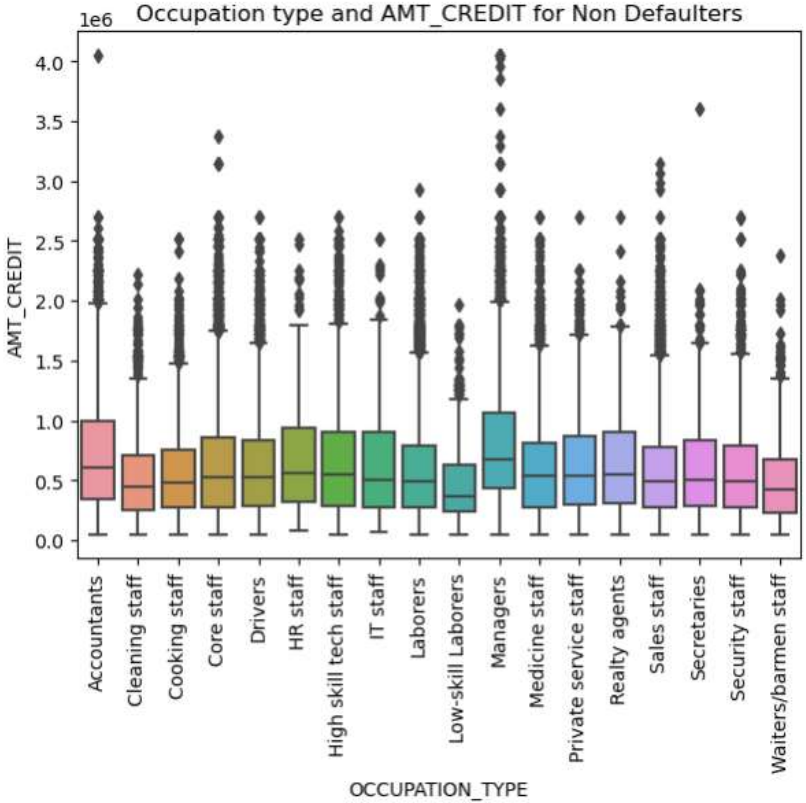


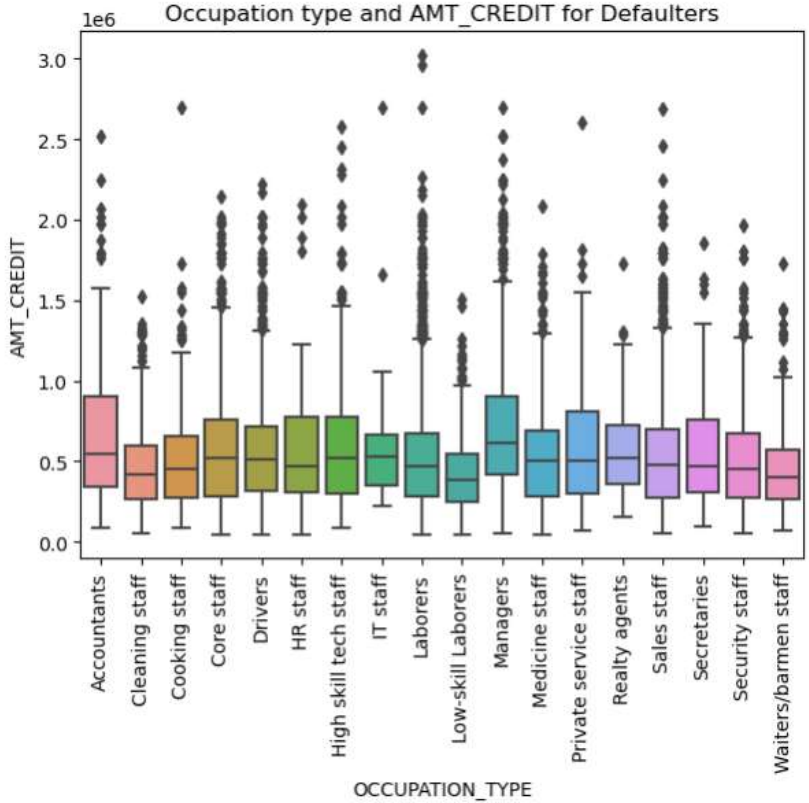



Income group and payment difficulty
TARGET
0
1
200000
150000
100000
50000
0
count
Medium
Low
High
Very High
INCOME_GROUP

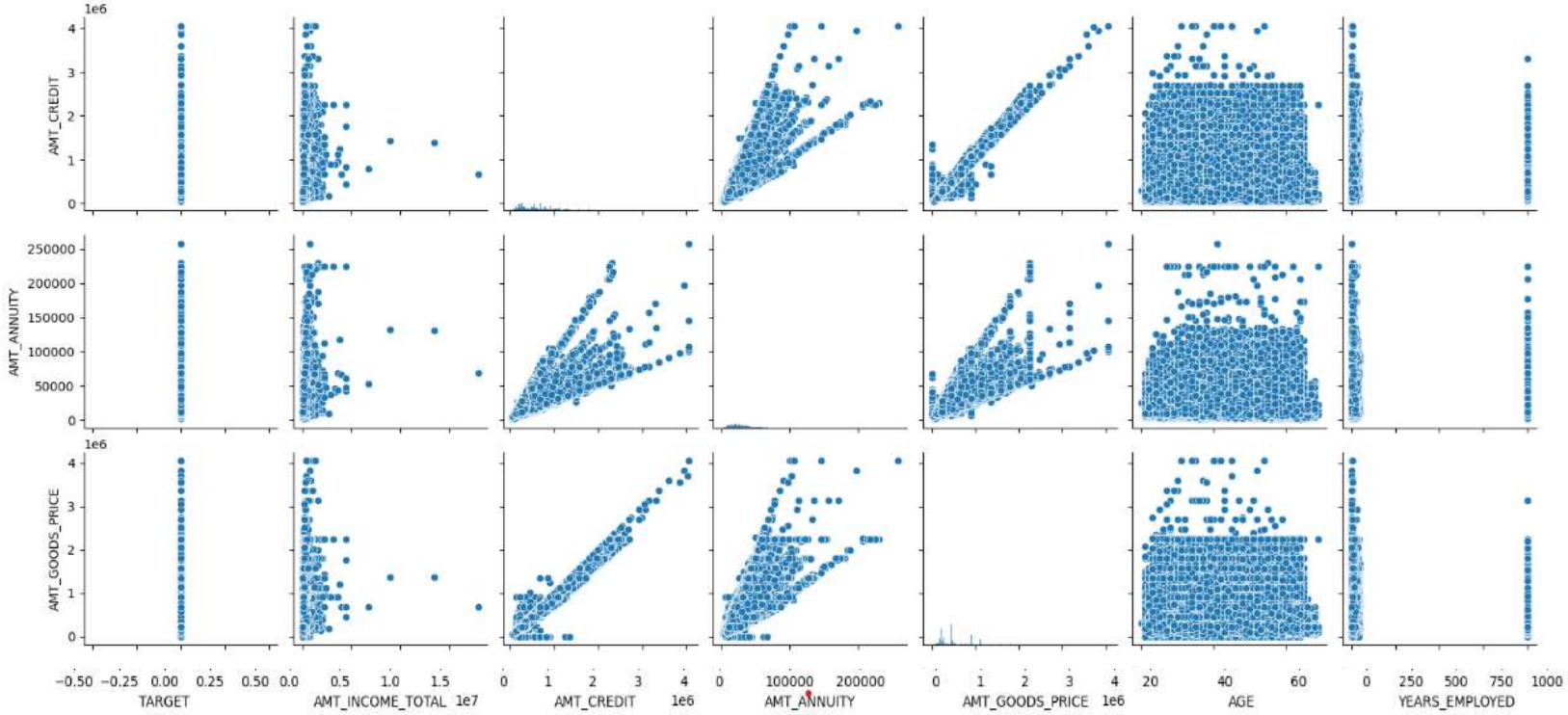

Education type and AMT_CREDIT for Non Defaulters
1e6
AMT_CREDIT
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
NAME_EDUCATION_TYPE


Education type and AMT_CREDIT for Defaulters
1e6
AMT_CREDIT
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Academic degree
Higher education
Incomplete higher
Lower secondary
Secondary / secondary special
NAME_EDUCATION_TYPE


Occupation type and AMT_CREDIT for Non Defaulters
1e6
AMT_CREDIT
OCCUPATION_TYPE
Accountants
Cleaning staff
Cooking staff
Core staff
Drivers
HR staff
High skill tech staff
IT staff
Laborers
Low-skill Laborers
Managers
Medicine staff
Private service staff
Realty agents
Sales staff
Secretaries
Security staff
Waiters/barmen staff


Occupation type and AMT_CREDIT for Defaulters
1e6
AMT_CREDIT
OCCUPATION_TYPE
Accountants
Cleaning staff
Cooking staff
Core staff
Drivers
HR staff
High skill tech staff
IT staff
Laborers
Low-skill Laborers
Managers
Medicine staff
Private service staff
Realty agents
Sales staff
Secretaries
Security staff
Waiters/barmen staff

CORRELATION METRICS




TARGET
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
AGE
YEARS_EMPLOYED
1 0.34 0.42 0.35 -0.062 -0.14
0.34 1 0.77 0.99 0.047 -0.07
0.42 0.77 1 0.78 -0.012 -0.1
0.35 0.99 0.78 1 0.045 -0.068
-0.062 0.047 -0.012 0.045 1 0.63
-0.14 -0.07 -0.1 -0.068 0.63 1
1.0
0.8
0.6
0.4
0.2
0.0

RECOMMENDATIONS

- Old people of any income group can be given loans.
- Old female clients could have less chances of defaulting a loan.
- Clients with higher education can be at less risk of default
- Any client who had an approved loan prior to the current loan

Risk analysis

- Clients with lower secondary education are at high risk of default.
- Business type 3 clients are at a high risk of default and lending them should be checked.
- Clients who have been previously refused loans should be considered risky.