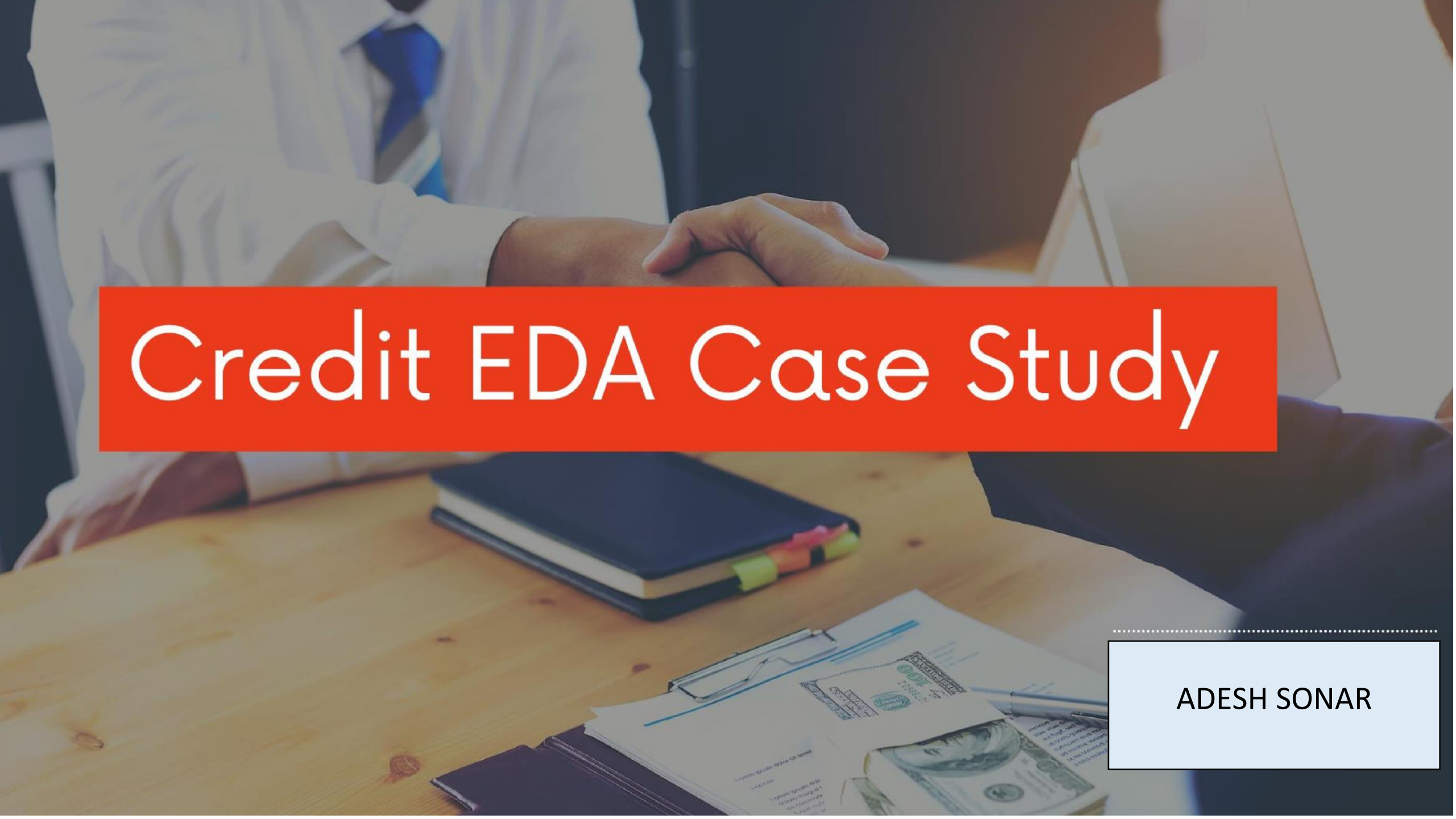


Credit EDA Case Study

ADESH SONAR

Problem Statement

A consumer finance company specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

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

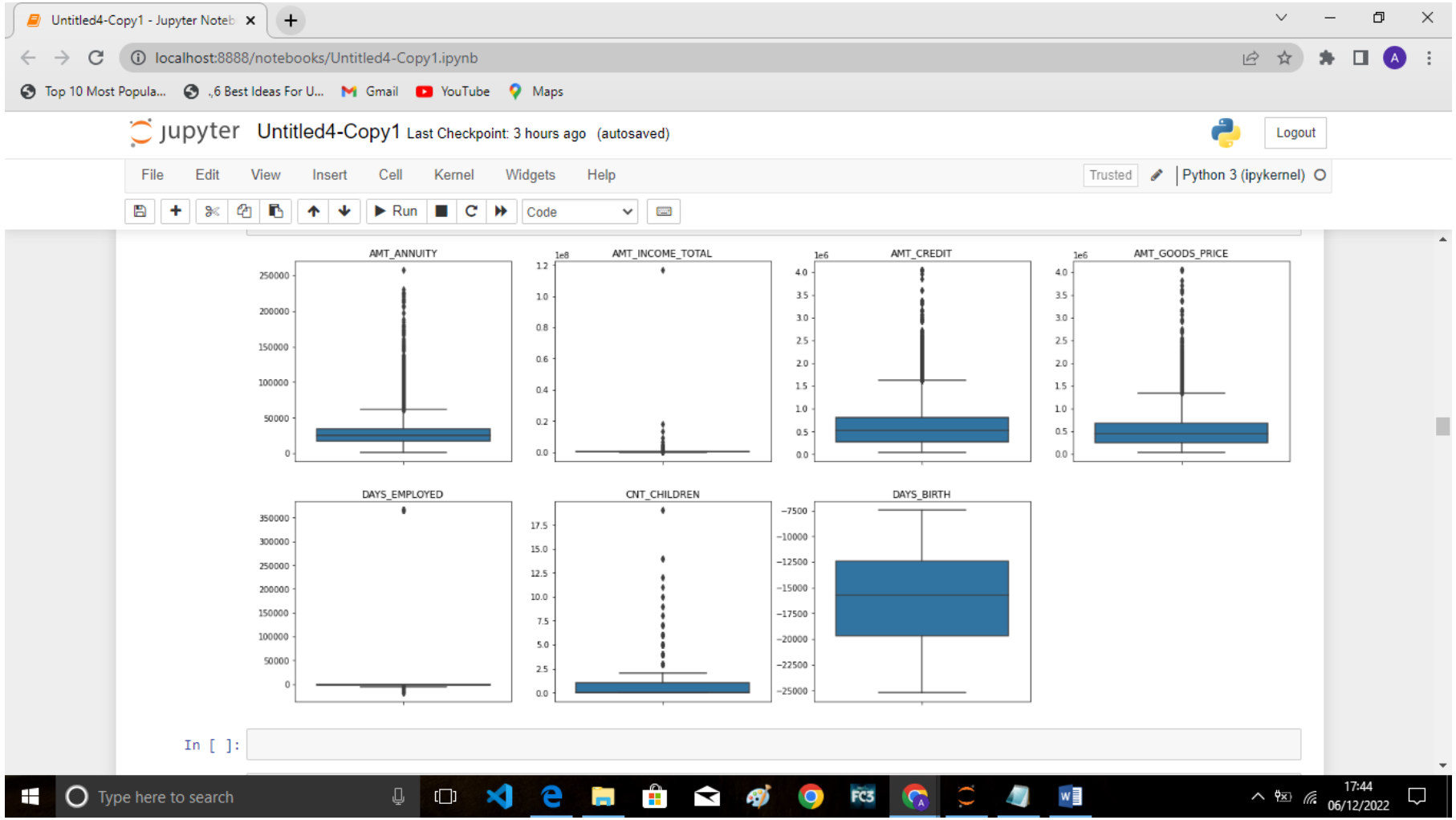
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

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.

Analysis Approach

1. Understanding problem statement
2. Understanding data
3. Data cleaning
4. Outlier Analysis
5. Performing different types of Analysis
6. Conclusion

• OUTLIERS ANALYSIS

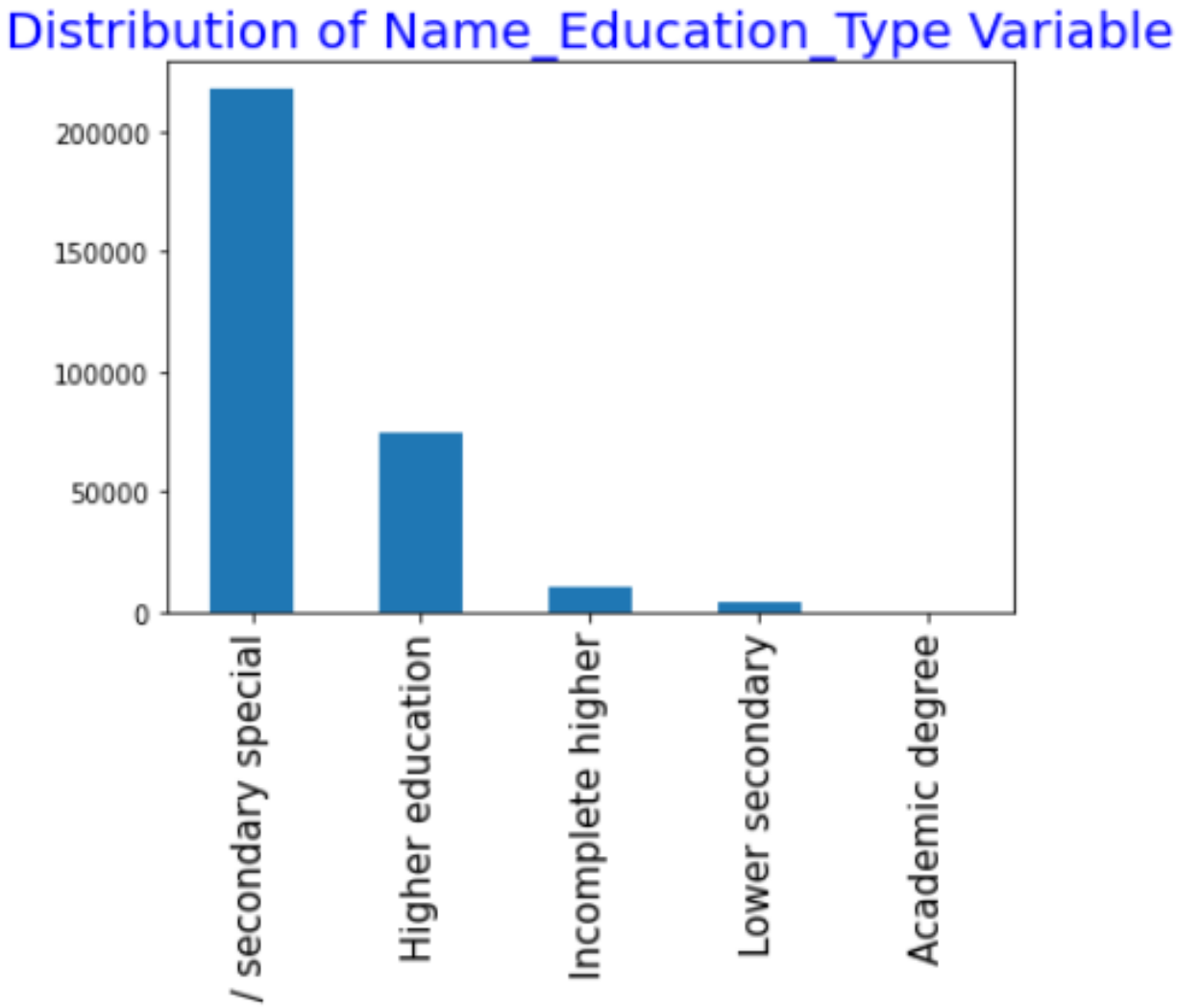


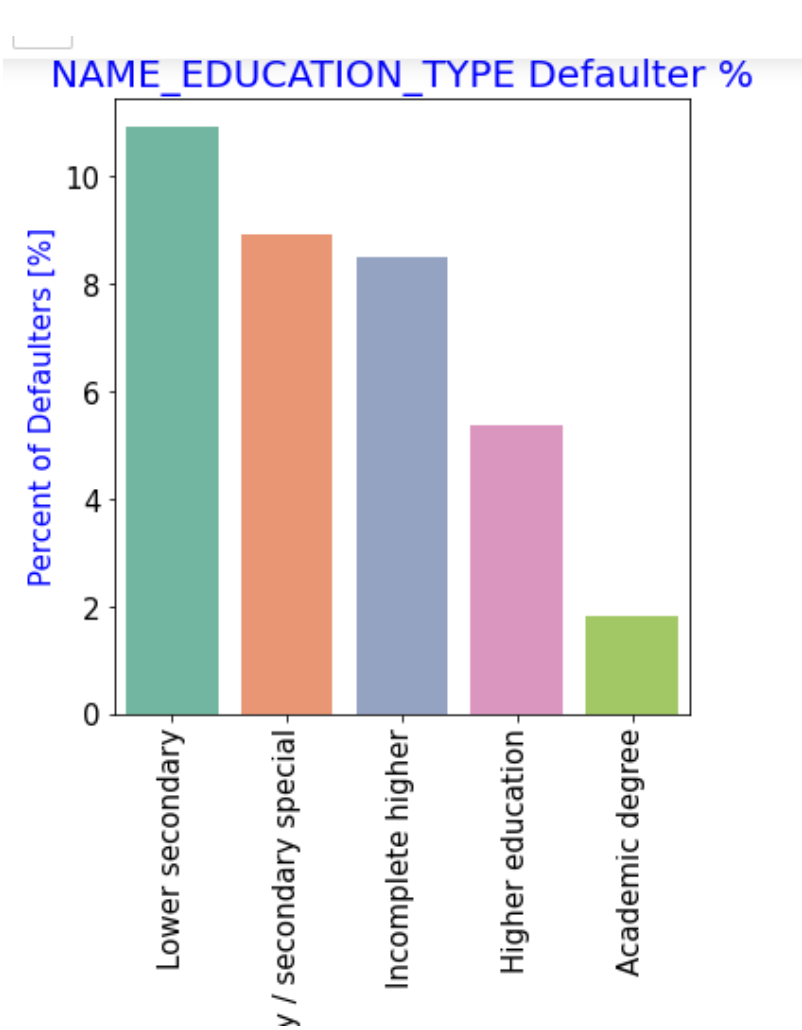
Analysis

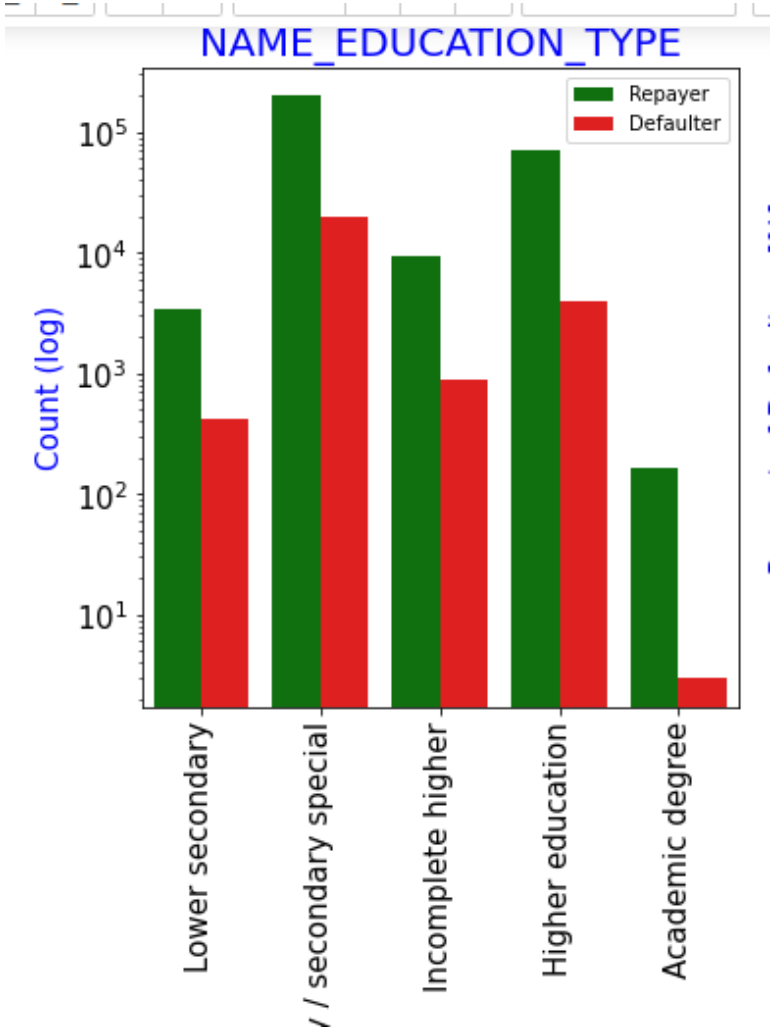
- AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE,CNT_CHILDREN have some number of outliers.
- AMT_INCOME_TOTAL has huge number of outliers it mean few people have high income.
- DAYS_BIRTH has no outliers .
- DAYS_EMPLOYED has high number of outlier which means data can be incorrect.

Types of Analysis

1. Univariate analysis

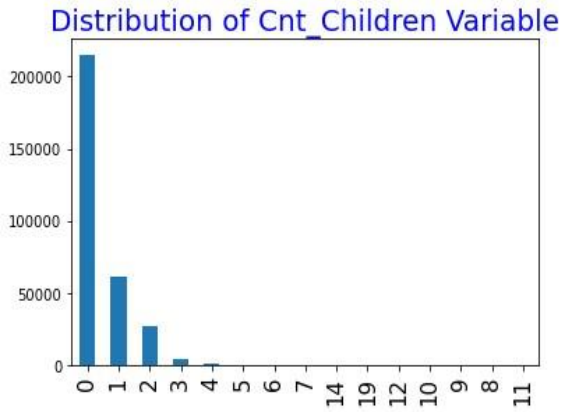


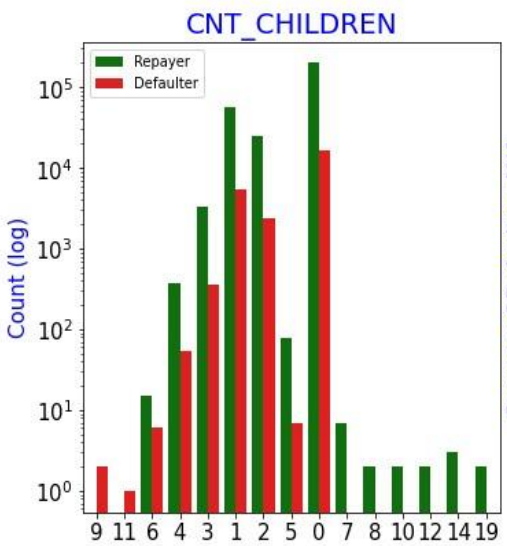

NAME_EDUCATION_TYPE Defaulter %
Percent of Defaulters [%]
0
2
4
6
8
10
Lower secondary
y / secondary special
Incomplete higher
Higher education
Academic degree

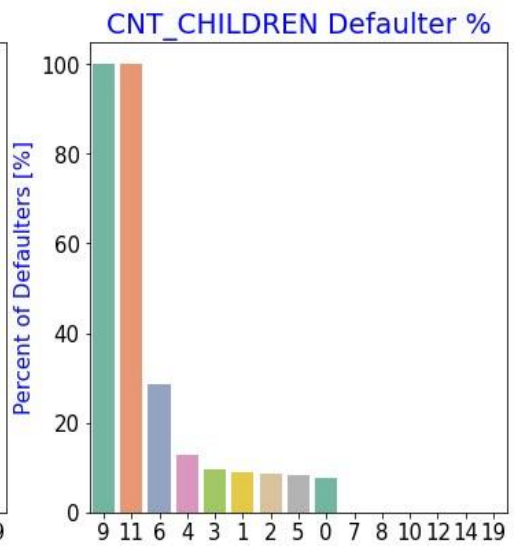

NAME_EDUCATION_TYPE
Repayer
Defaulter
Count (log)
10^1
10^2
10^3
10^4
10^5
Lower secondary
/ / secondary special
Incomplete higher
Higher education
Academic degree

Analysis

- Majority of the defaulters 12 % have secondary special education or Higher education.
- Small number of defaulters have an academic degree with 2% .

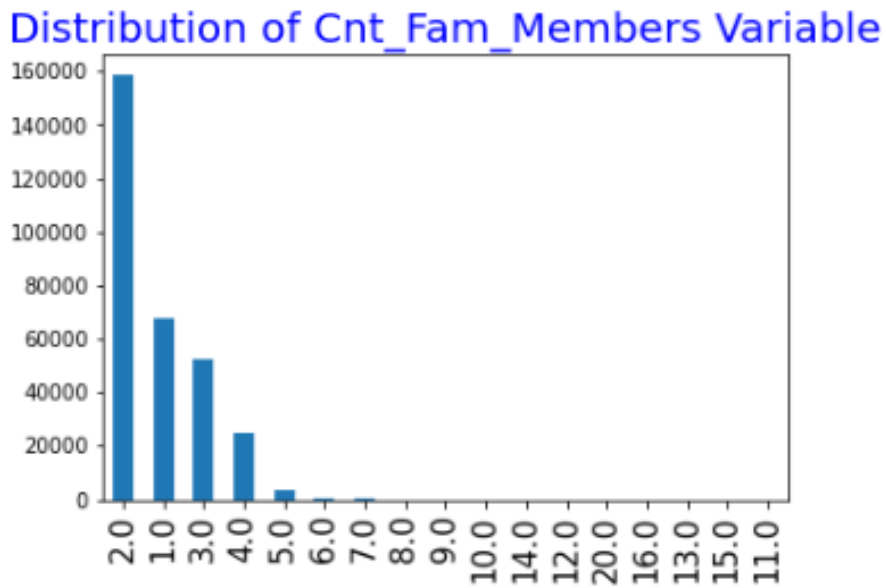


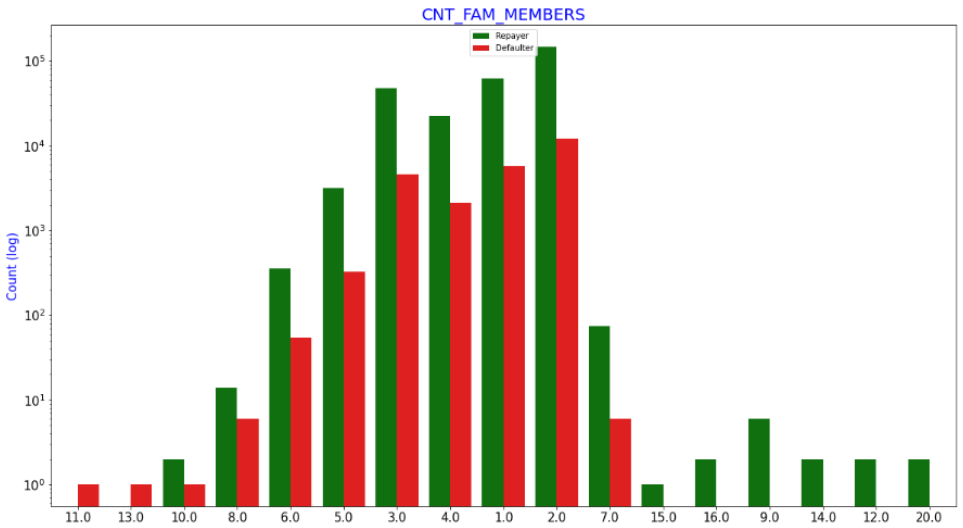


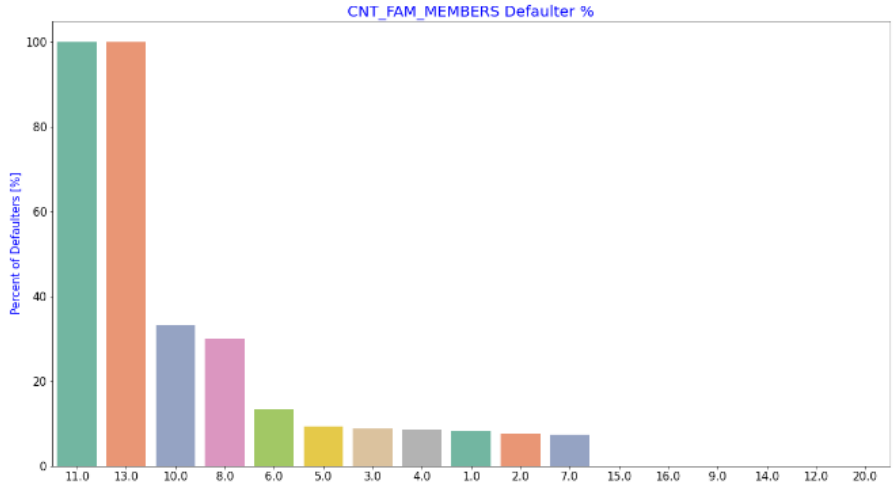


Analysis

- Clients with zero to two children tend to repay the loans.

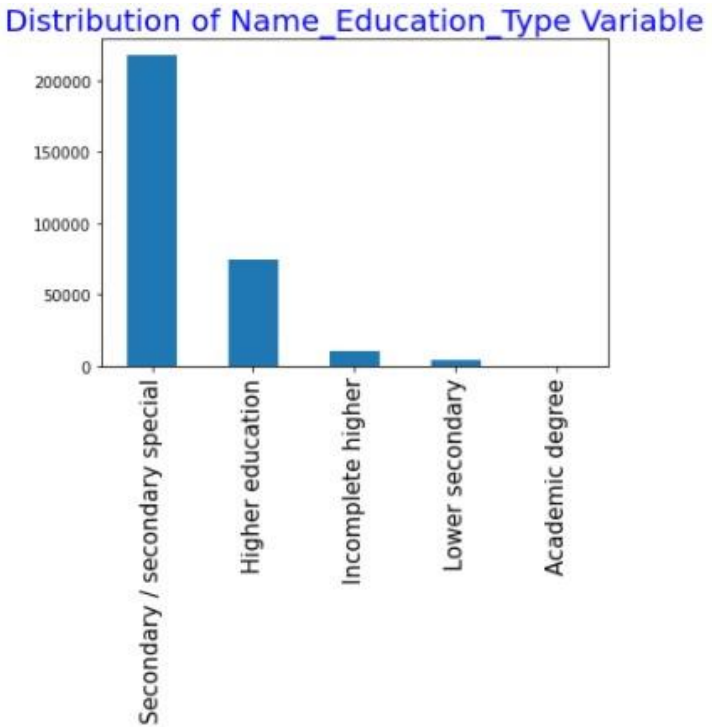

Distribution of Cnt_Fam_Members Variable
160000
140000
120000
100000
80000
60000
40000
20000
0
2.0
1.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
14.0
12.0
20.0
16.0
13.0
15.0
11.0

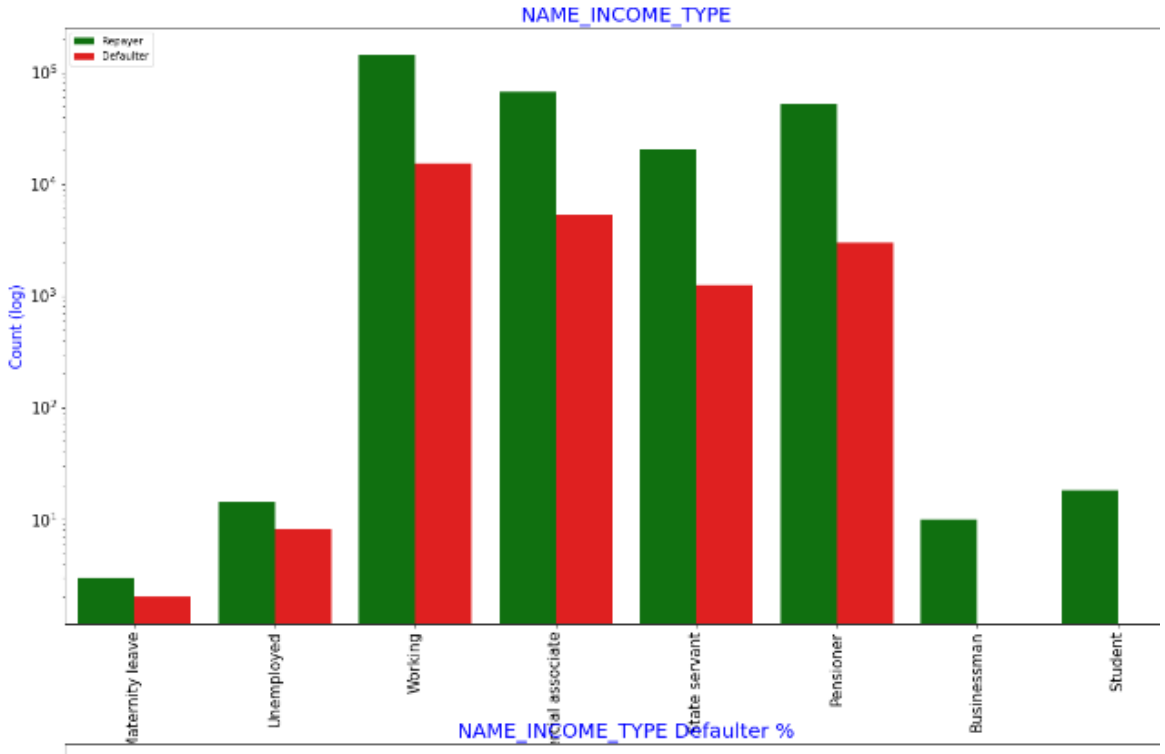


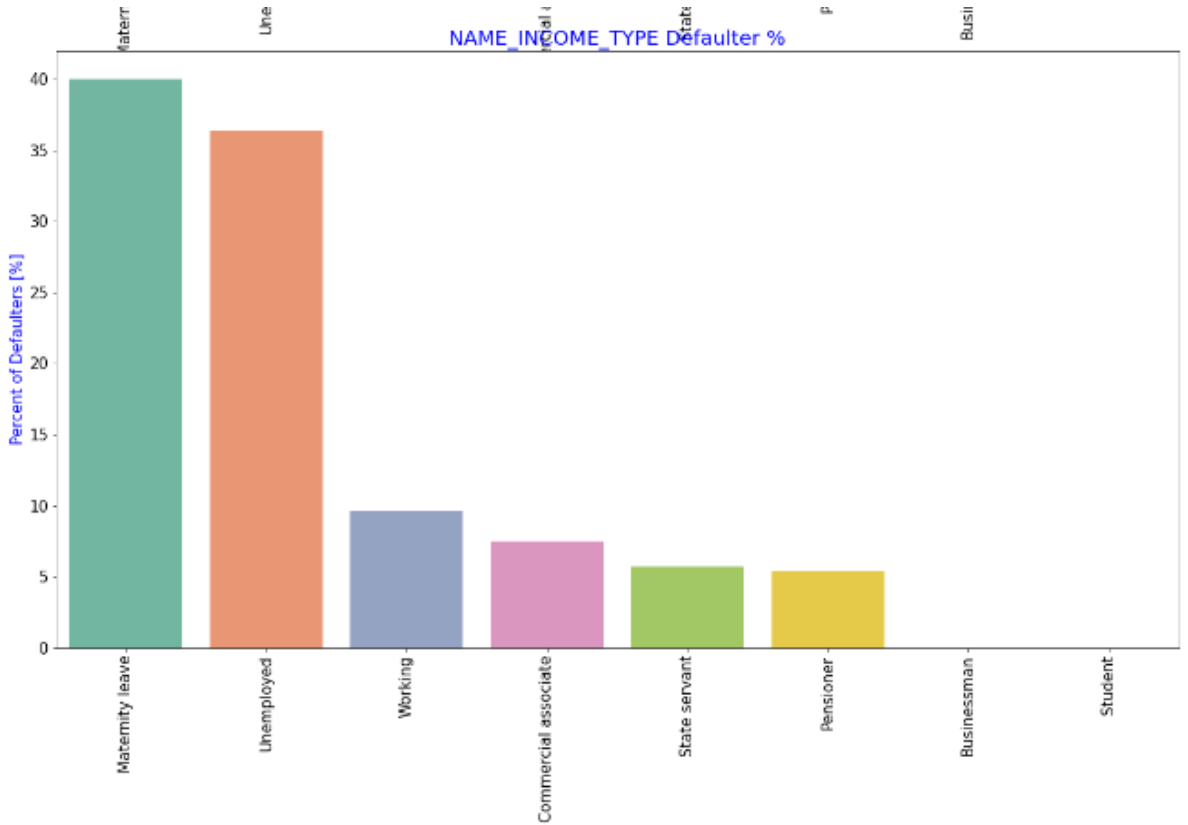


Analysis

- Family member follows the same trend as children where having more family members increases the risk of defaulting.

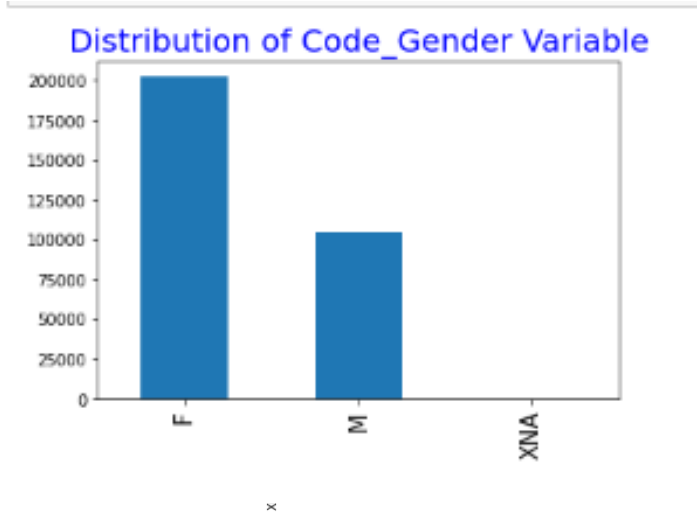

Distribution of Name_Education_Type Variable
200000
150000
100000
50000
0
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree

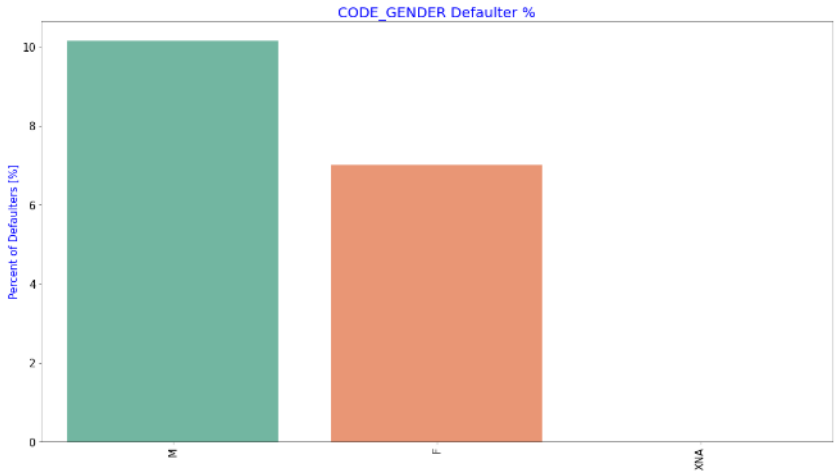

NAME_INCOME_TYPE
Repayer
Defaulter
Count (log)
Maternity leave
Unemployed
Working
Pensioner
Businessman
Student
NAME_INCOME_TYPE Defaulter %

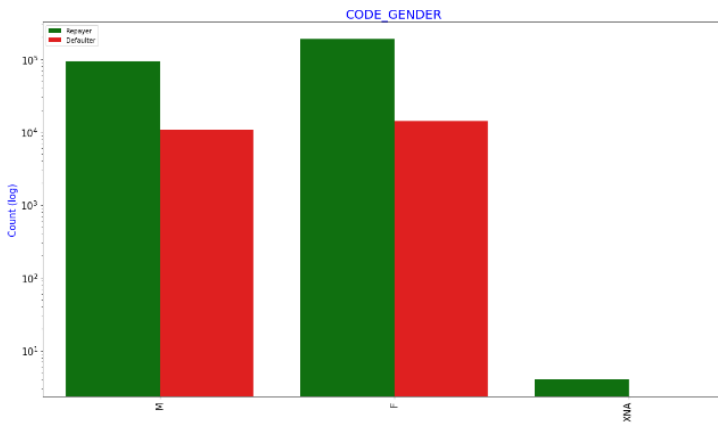


Analysis

- Applicants who are Academic degree holder have lower default rate

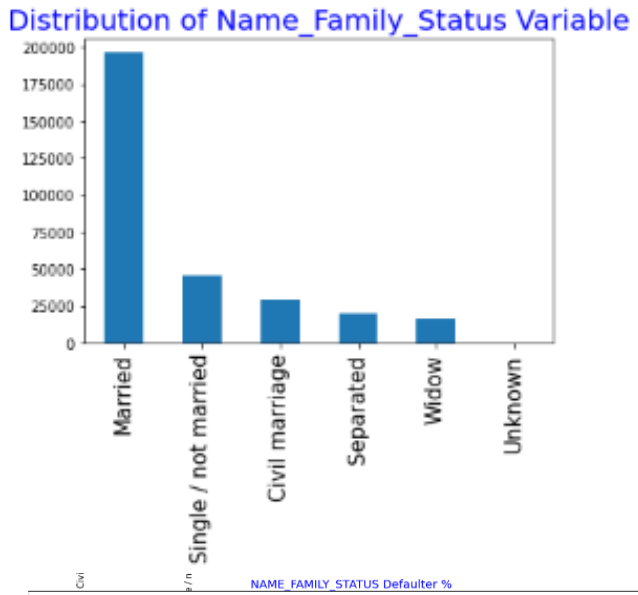

Distribution of Code_Gender Variable
200000
175000
150000
125000
100000
75000
50000
25000
0
F
M
XNA
X

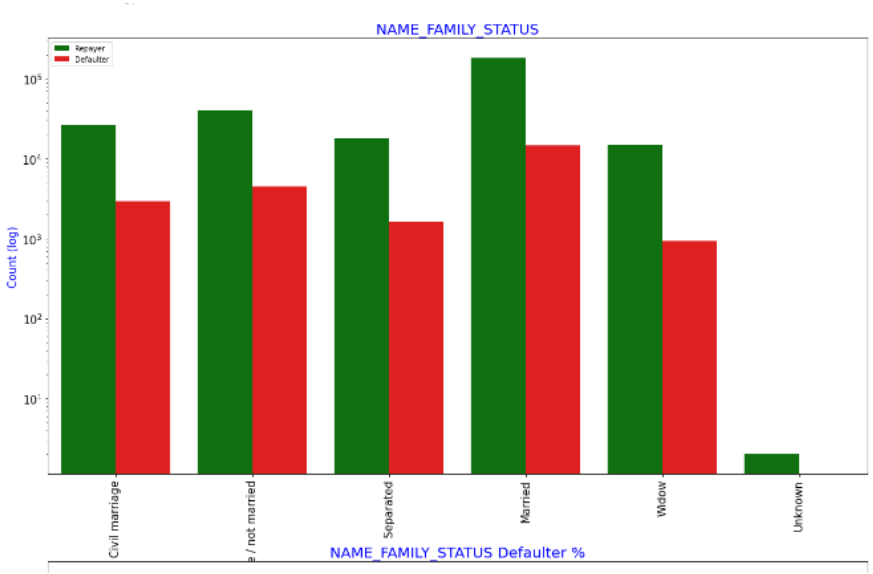

CODE_GENDER Defaulter %
Percent of Defaulters [%]
10
8
6
4
2
0
M
F
XNA

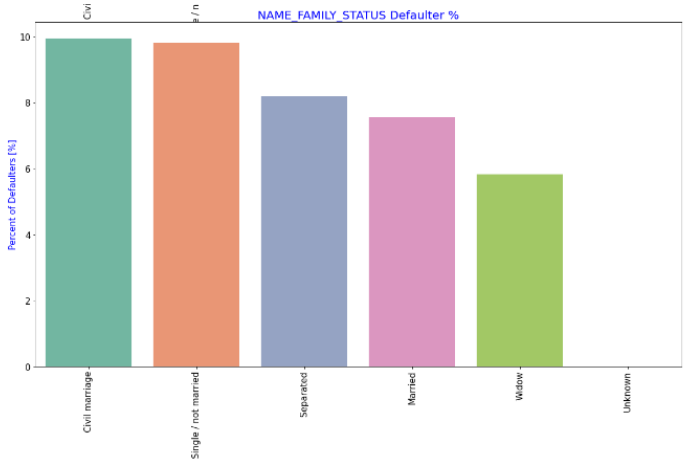

CODE_GENDER
Repayer
Defaulter
Count (log)
M
F
XNA

Analysis

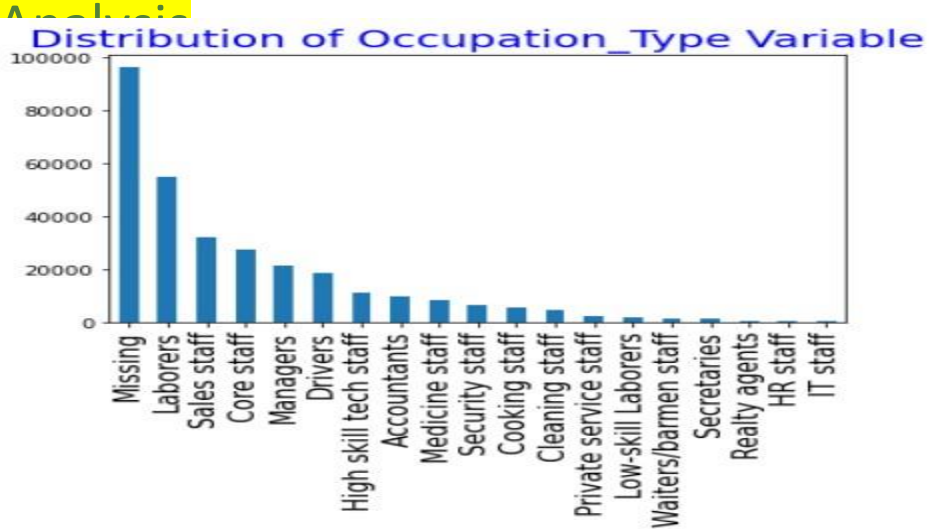
- Males have a higher chance of not returning their loans (10%), comparing with women (7%).

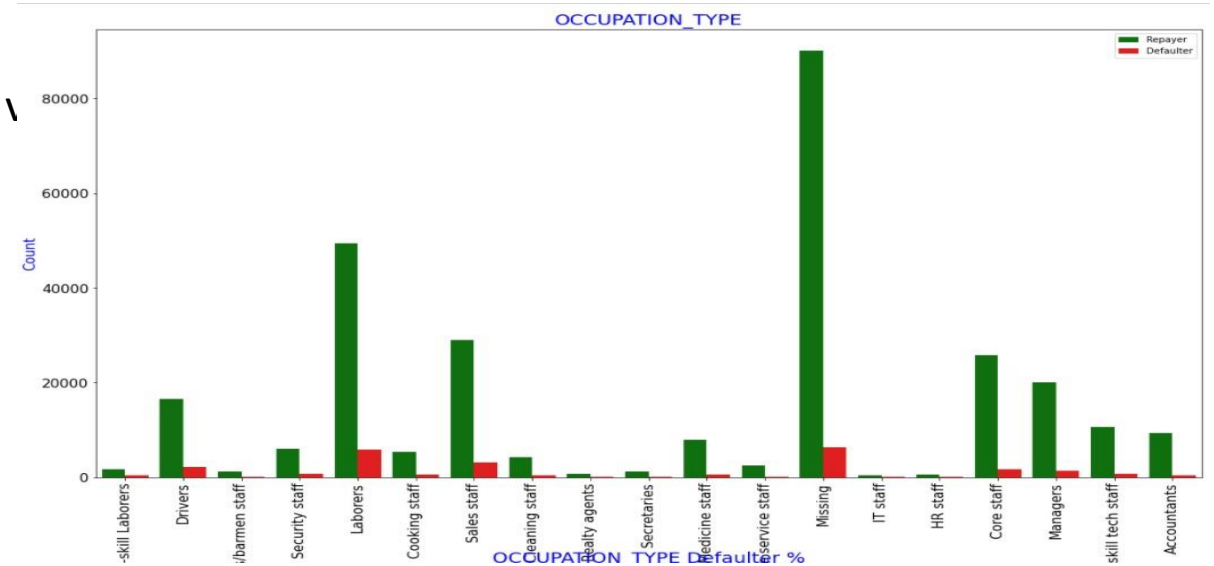

Distribution of Name_Family_Status Variable
200000
175000
150000
125000
100000
75000
50000
25000
0
Married
Single / not married
Civil marriage
Separated
Widow
Unknown

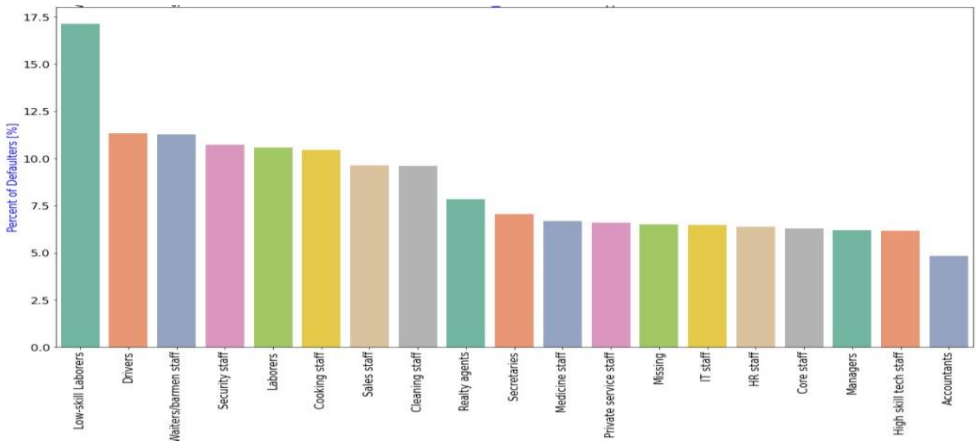

NAME_FAMILY_STATUS
Repayer
Defaulter
Count (log)
Civil marriage
Separated
Married
Widow
Unknown
NAME_FAMILY_STATUS Defaulter %


NAME_FAMILY_STATUS Defaulter %
Percent of Defaulters [%]
Civil marriage
Single / not married
Separated
Married
Widow
Unknown

Analysis

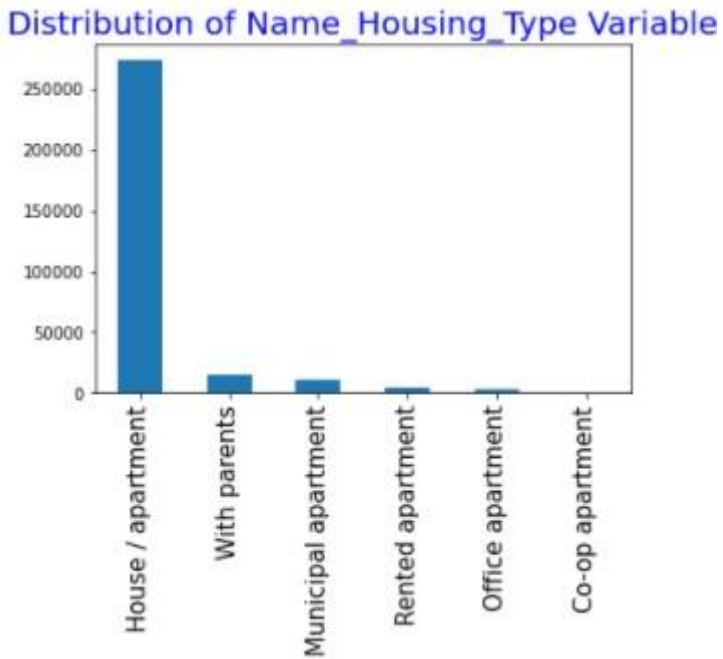

Distribution of Occupation_Type Variable
100000
80000
60000
40000
20000
0
Missing
Laborers
Sales staff
Core staff
Managers
Drivers
High skill tech staff
Accountants
Medicine staff
Security staff
Cooking staff
Cleaning staff
Private service staff
Low-skill Laborers
Waiters/barmen staff
Secretaries
Realty agents
HR staff
IT staff

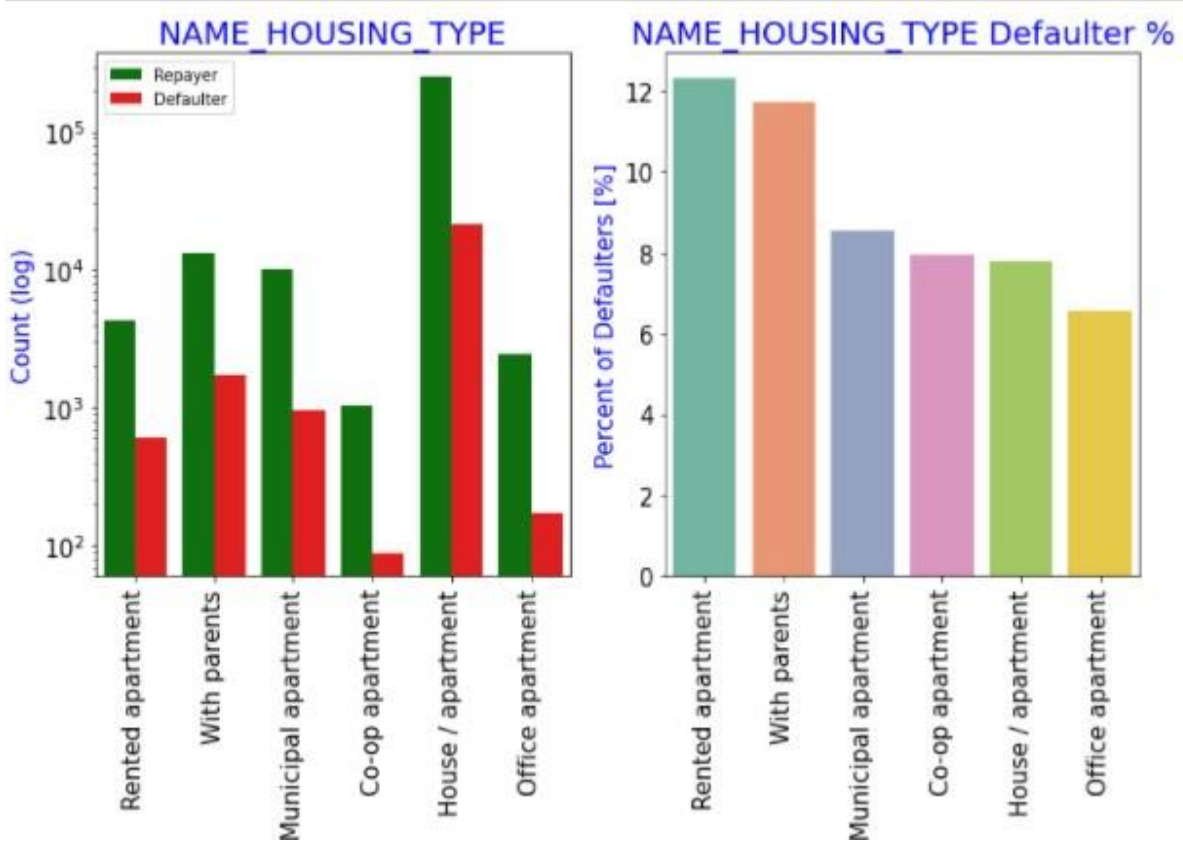

OCCUPATION_TYPE
Repayer
Defaulter
Count
80000
60000
40000
20000
0
-skill Laborers
Drivers
/barmen staff
Security staff
Laborers
Cooking staff
Sales staff
Cleaning staff
Realty agents
Secretaries
Medicine staff
service staff
Missing
IT staff
HR staff
Core staff
Managers
skill tech staff
Accountants
OCCUPATION_TYPE Defaulter %


Percent of Defaulters [%]
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Low-skill Laborers
Drivers
Waiters/barmen staff
Security staff
Laborers
Cooking staff
Sales staff
Cleaning staff
Realty agents
Secretaries
Medicine staff
Private service staff
Missing
IT staff
HR staff
Core staff
Managers
High skill tech staff
Accountants

Analysis

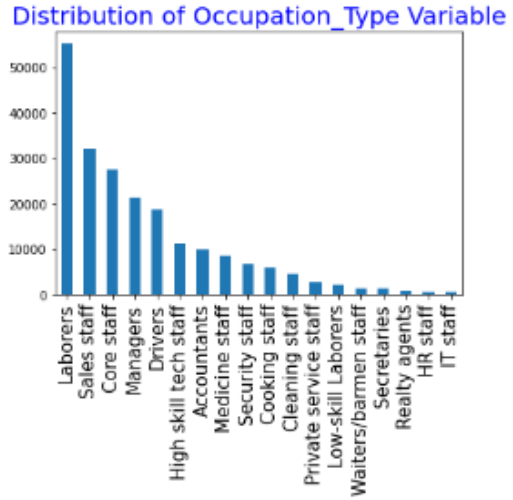
- Clients who are low skilled laborers, drivers , waiters / barmen staff, security staff,laborers and cooking staff have high default rate.

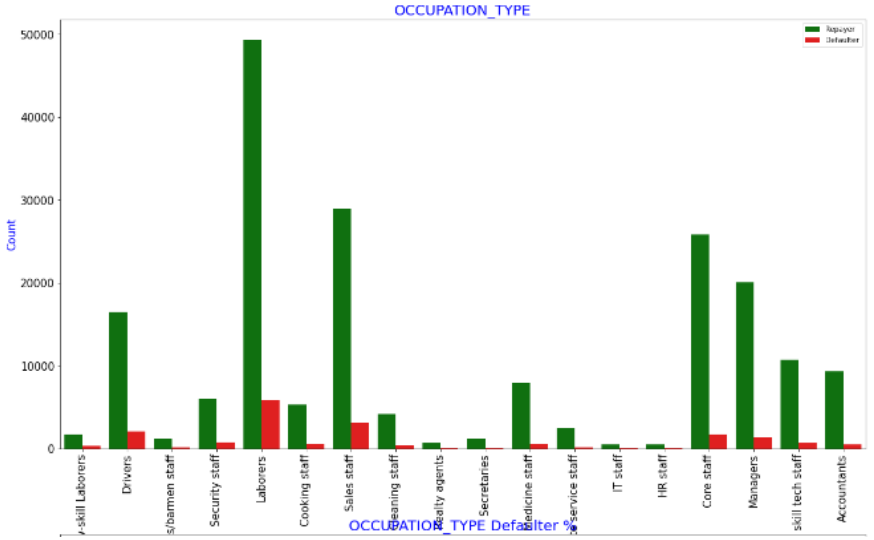

Distribution of Name_Housing_Type Variable
250000
200000
150000
100000
50000
0
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment

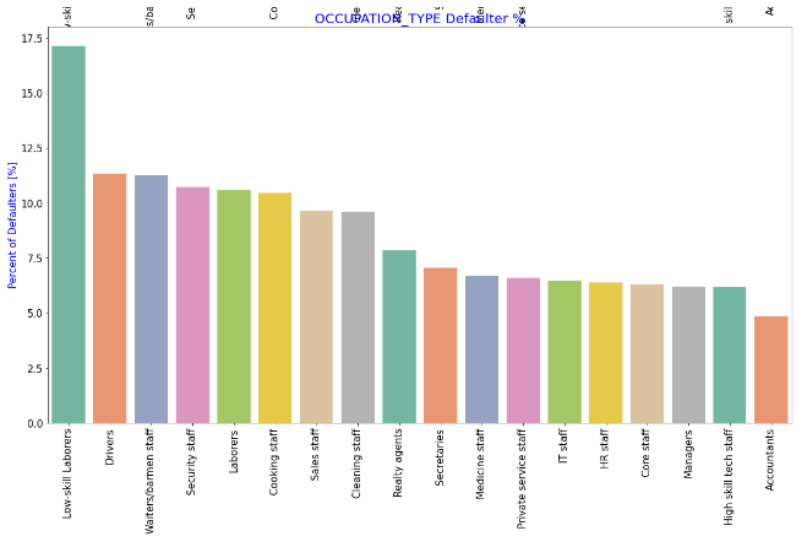

NAME_HOUSING_TYPE
Repayer
Defaulter
Count (log)
10^5
10^4
10^3
10^2
Rented apartment
With parents
Municipal apartment
Co-op apartment
House / apartment
Office apartment
NAME_HOUSING_TYPE Defaulter %
Percent of Defaulters [%]
12
10
8
6
4
2
0
Rented apartment
With parents
Municipal apartment
Co-op apartment
House / apartment
Office apartment

Analysis

- High number of loan applications are from the category of people who live in Rented apartments &
  living with parents so offering the loan would mitigate the loss.



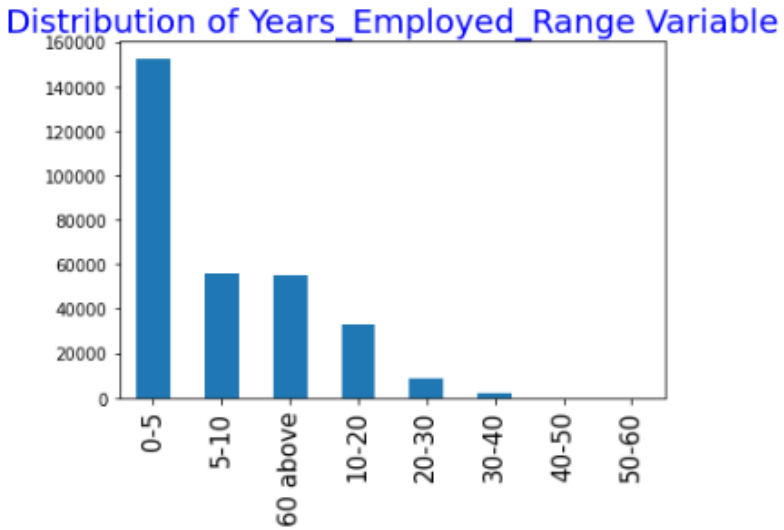


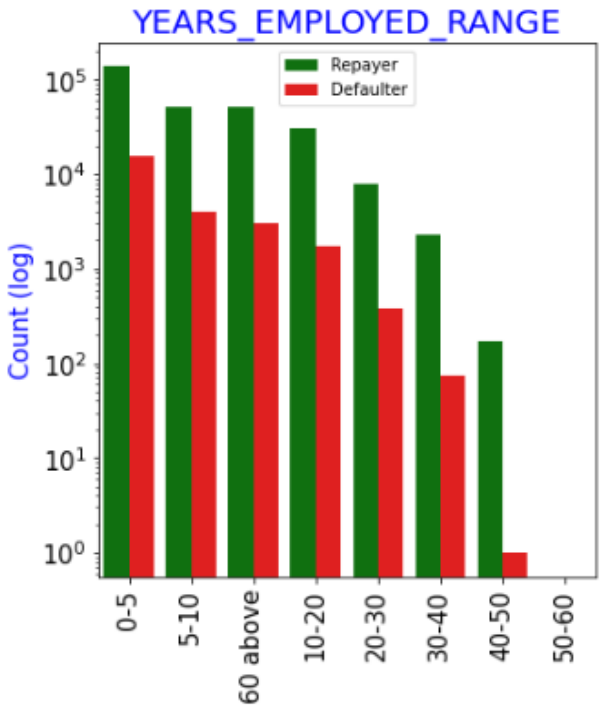


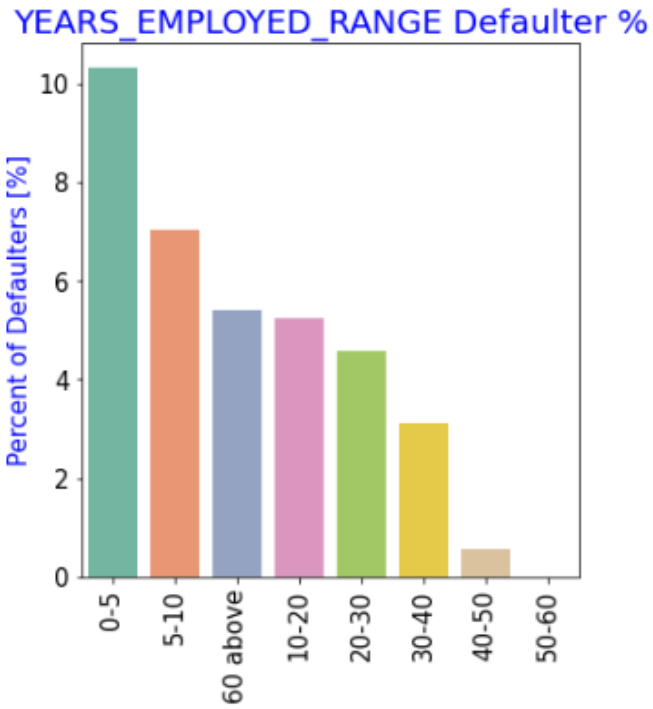
Analysis

- Most of the applications for loan are from people working in Business Entity Type 3 organization

- Organizations with highest percent of loans not repaid are Transport: type 3 (around 16%), Industry: type 13 (13.5%), Industry: type 8 (around
- Self employed people have relative high defaulting rate (10%), and thus should be thoughly scrutized before being approved for loan or provide
- For a very high number of applications, Organization type information is unavailable(XNA)
- It can be seen that following category of organization type has lesser defaulters thus safer for providing loans: - Trade Type 4
- Industry type 122%).


Distribution of Years_Employed_Range Variable
160000
140000
120000
100000
80000
60000
40000
20000
0
0-5
5-10
60 above
10-20
20-30
30-40
40-50
50-60


YEARS_EMPLOYED_RANGE
Repayer
Defaulter
Count (log)
10^5
10^4
10^3
10^2
10^1
10^0
0-5
5-10
60 above
10-20
20-30
30-40
40-50
50-60

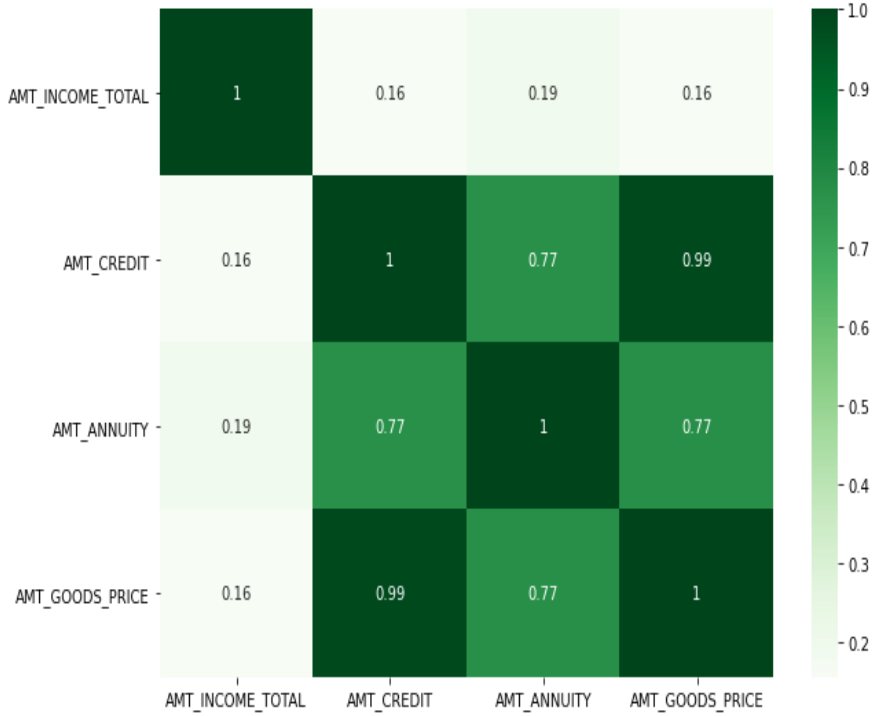

YEARS_EMPLOYED_RANGE Defaulter %
Percent of Defaulters [%]
10
8
6
4
2
0
0-5
5-10
60 above
10-20
20-30
30-40
40-50
50-60

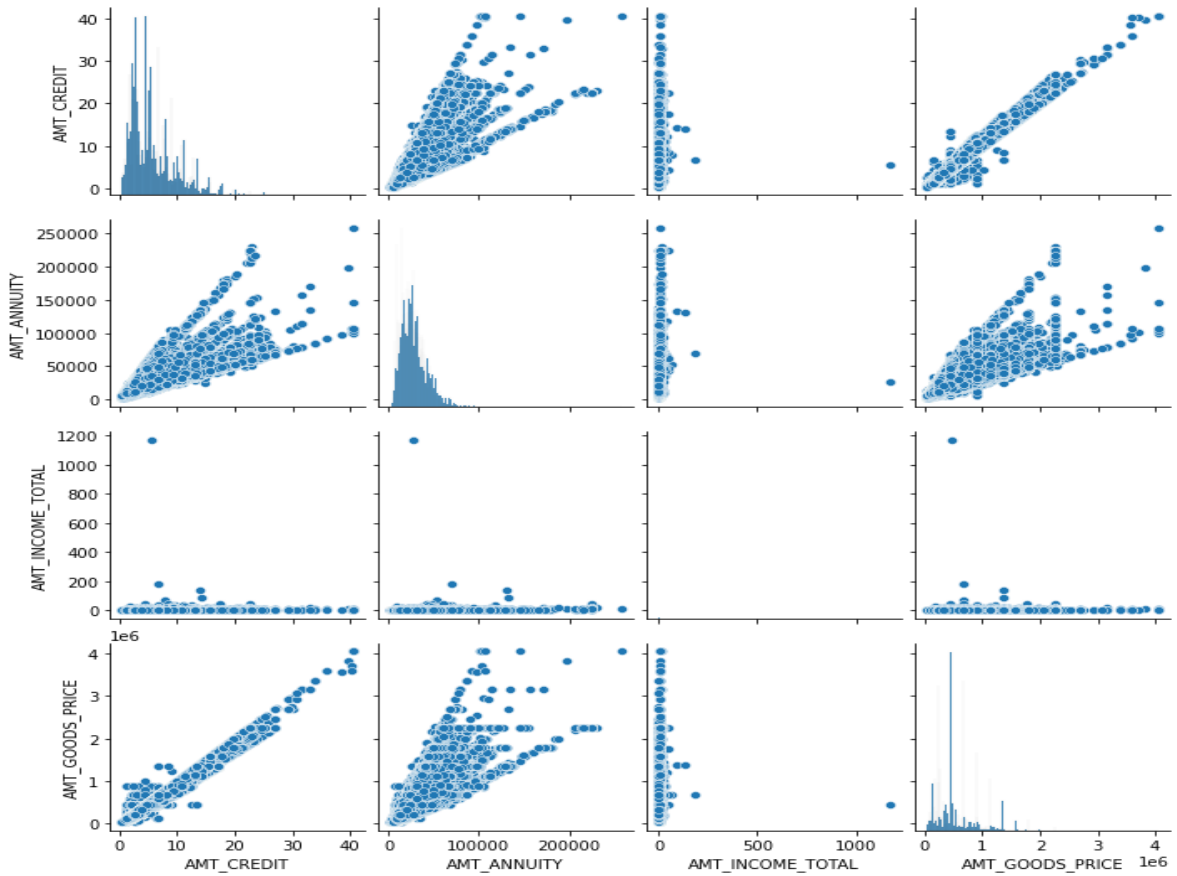
Analysis

- Majority of the applicants have been employed in between 0-5 years. The defaulting rating of this group is also the highest which is 10%.
- With increase of employment year, defaulting rate is gradually decreasing with people having 40+ year experience having less than 1% default rate.

Bivariate & Multivariate

AMT_CREDIT, AMT_ANNUITY, AMT_INCOME_TOTAL , AMT_GOODS_PRICE





Analysis

- Very high correlation between AMT_CREDIT and AMT_GOODS_PRICE - Applicants owning goods of high value can take loans of higher amounts.

Summary

Decisive Factors for an applicant to be safe borrowers.

- Applicants with 40+ year experience having less than 1% default rate.
- Applicants with zero to two children tend to repay the loans.
- Academic degree has less defaults.
- Applicants with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%.
- Student have no defaults.
- Applicants who have children equal to or more than 9 default 100% and hence their applications can to be rejected
- Male applicants have relatively higher default rate
- Applicants with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%.

Decisive Factors for an applicant to be a potential Defaulter

- Applicants who have children equal to or more than 9 default 100% and hence their applications can to be rejected.
- Avoid young applicants who are in age group of 20-40 as they have higher probability of defaulting.
- Applicants who have higher family members (>=11) have higher default rate and their applications can be rejected.
- Applicants who have less than 5 years of employment have high default rate.
- Applicants who are either at Maternity leave or Unemployed have higher default rate.
- Applicants with Lower Secondary education , ,incomplete education have higher default rate.
- Applicants in civil marriage or who are single have higher default rate.
- Low-skill Laborers, drivers and Waiters/barmen staff, Security staff have huge default rate.

- Applicants who have 4 to 8 children have a very high default rate and hence higher interest should be imposed on their loans.
- Applicants with family members between 8 to 10 have a very high default rate and hence higher interest should be imposed on their loans.
- People living in rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.
- Industry type 3 , type 13 and type 8 have high defaulting rate

Thankyou !