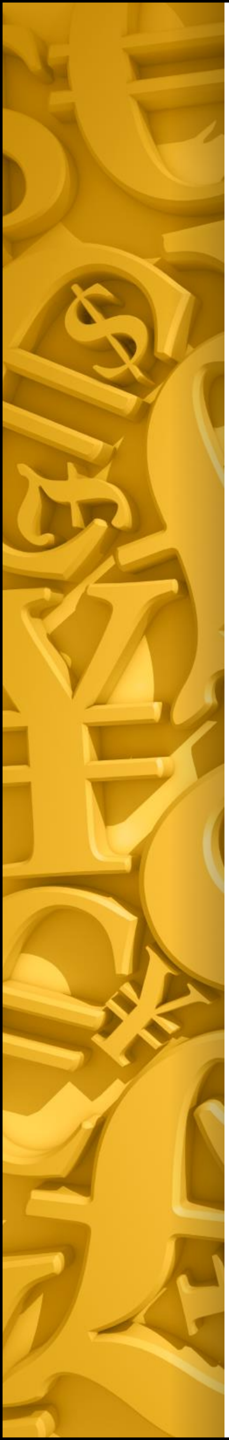


LOAN CREDIT RISK ANALYTICS

Credit Eda Loan Case Study

Submitted By:

GURPREET KAUR

DSC43

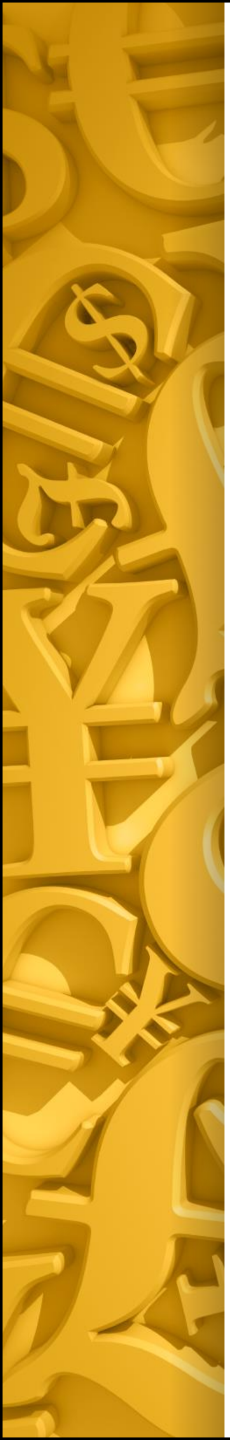
BUSINESS OBJECTIVES

- This **Credit EDA Loan Case Study** aims of applying EDA in a real business scenario. Our Analysis will assist the Financial Organization in making a **LOAN APPROVAL DECISION** for the applicant. This might help the Bank/ Financial company control its losses and avert financial losses.

- By Applying **Credit Risk Analytics Techniques** in banking and financial services to minimize the risk of losing money while lending to customers.

- In this case study, we will use **EDA** to understand how Consumer Attributes and Loan Attributes influence the tendency of default.

- This case study aims to identify Patterns/ **Strong Driving Factors** behind loan default which 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 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.***

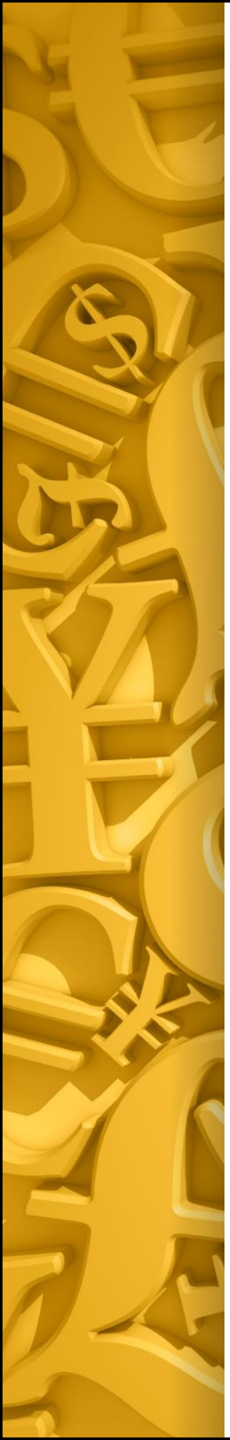


PROBLEM STATEMENT

Two types of risks are associated with the bank's/Financial Organization decision for Loan Approval:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
2. 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.

Identification of such applicants & Strong Defaulter Driving Patterns using EDA is the aim of this case study.



EDA APPROACH

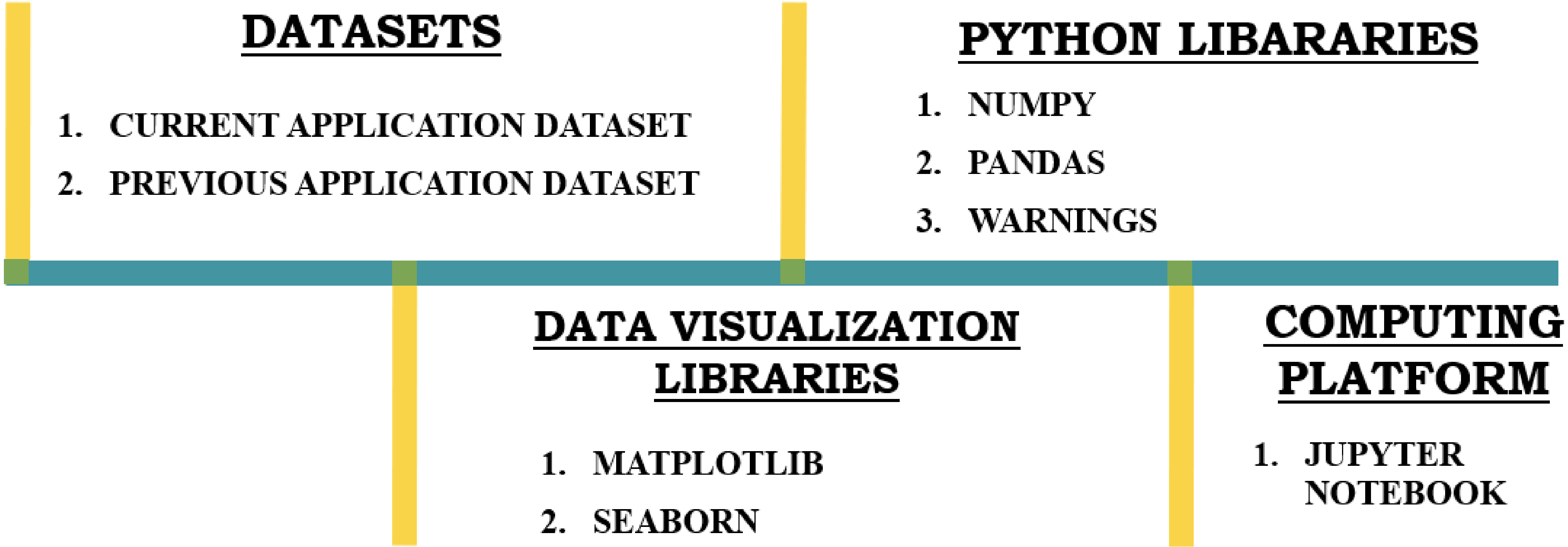
OF "CURRENT APPLICATION" DATASET

1. Dataset Inspection (Data Understanding)
2. Data Imbalance Check
3. Data Cleaning, Manipulation & Analysis
4. Transformation Of Dates (Days, Quarter, Years)
5. Missing Value Analysis
6. Imputation in Missing Value Analysis
7. Outliers Analysis
8. Univariate Analysis Of Categorical Columns
9. Univariate Analysis of Numerical Columns
10. Segmented Univariate Analysis
11. Bivariate Analysis
12. Top 10 Correlated Columns (CORR)

OF "MERGED DATASET" PREVIOUS & CURRENT APPLICATION

13. Data Reading & Dataset Inspection Of Previous Application Dataset
14. Merging Previous-Application dataset with Current-Application dataset
15. Data Cleaning, Manipulation & Analysis
16. Transformation Of Dates (Days, Quarter, Years) of Merged Dataset
17. Missing Value Analysis of Merged Dataset
18. Univariate Analysis Of Categorical Columns Of Merged Dataset
19. Univariate Analysis of Numerical Columns of Merged Dataset.
20. Segmented Univariate Analysis Merged Dataset
21. Bivariate Analysis Of Merged Dataset
22. Conclusion (Summary) With Strong Indicators To Drive Business Decisions

EDA REQUIREMENTS



DATASETS

1. CURRENT APPLICATION DATASET
2. PREVIOUS APPLICATION DATASET

DATA VISUALIZATION LIBRARIES

1. MATPLOTLIB
2. SEABORN

PYTHON LIBARARIES

1. NUMPY
2. PANDAS
3. WARNINGS

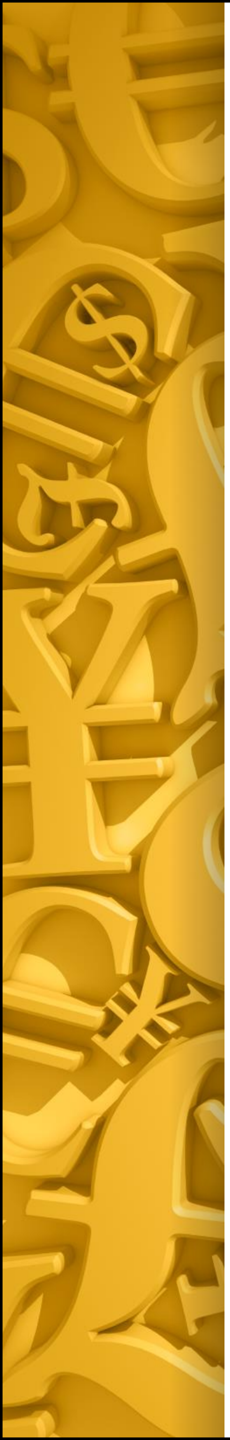
COMPUTING PLATFORM

1. JUPYTER NOTEBOOK

STRONG INDICATORS TO DRIVE BUSINESS DECISIONS

1. Target/focused variable for *"Current-Application" Dataset* : “**TARGET**”
2. Target/focused variable for *"Previous-Application" Dataset* : “**NAME_CONTRACT_STATUS**”
3. **Top Major variables to consider for loan prediction:**
   - NAME_EDUCATION_TYPE
   - DAYS_BIRTH
   - DAYS_EMPLOYED
   - NAME_INCOME_TYPE
   - ORGANIZATION_TYPE
   - NAME_HOUSING_TYPE
   - AMT_INCOME_TOTAL
   - AMT_CREDIT
   - AMT_ANNUITY
   - CODE_GENDER
   - REGION_RATING_CLIENT
   - NAME_CASH_LOAN_PURPOSE

To Minimize risk of loss, the above-mentioned variables should be evaluated before approving an application.



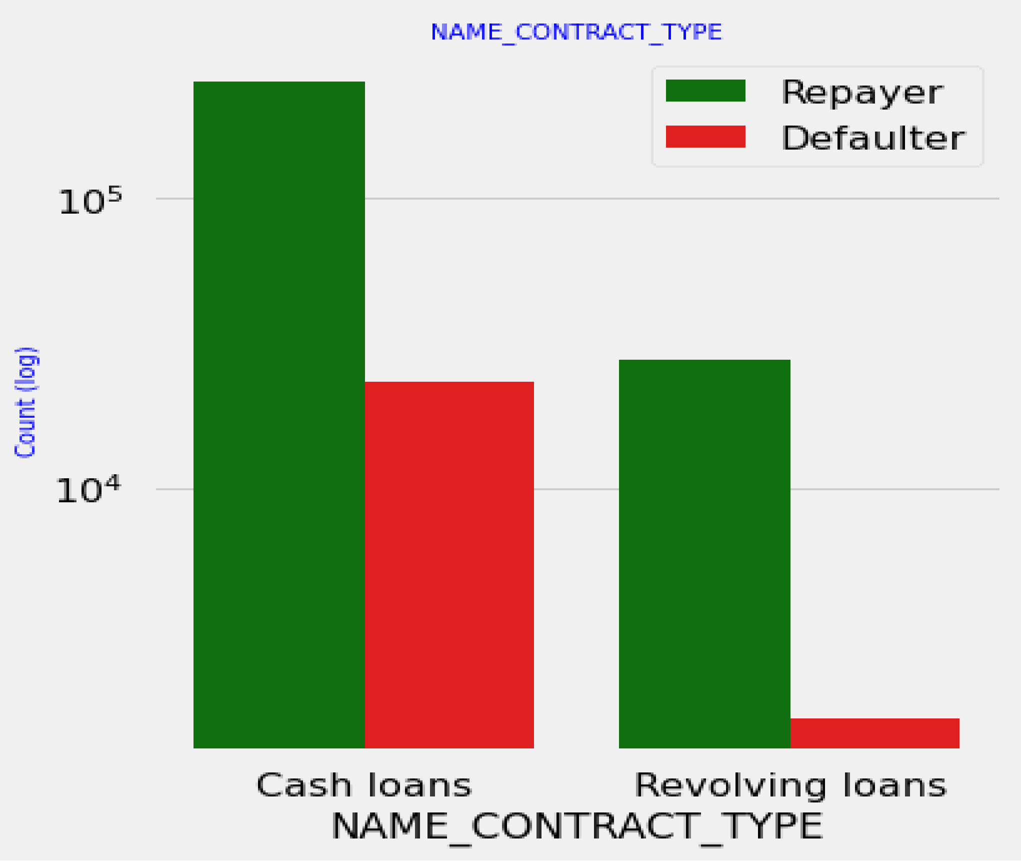
CURRENT ANALYSIS

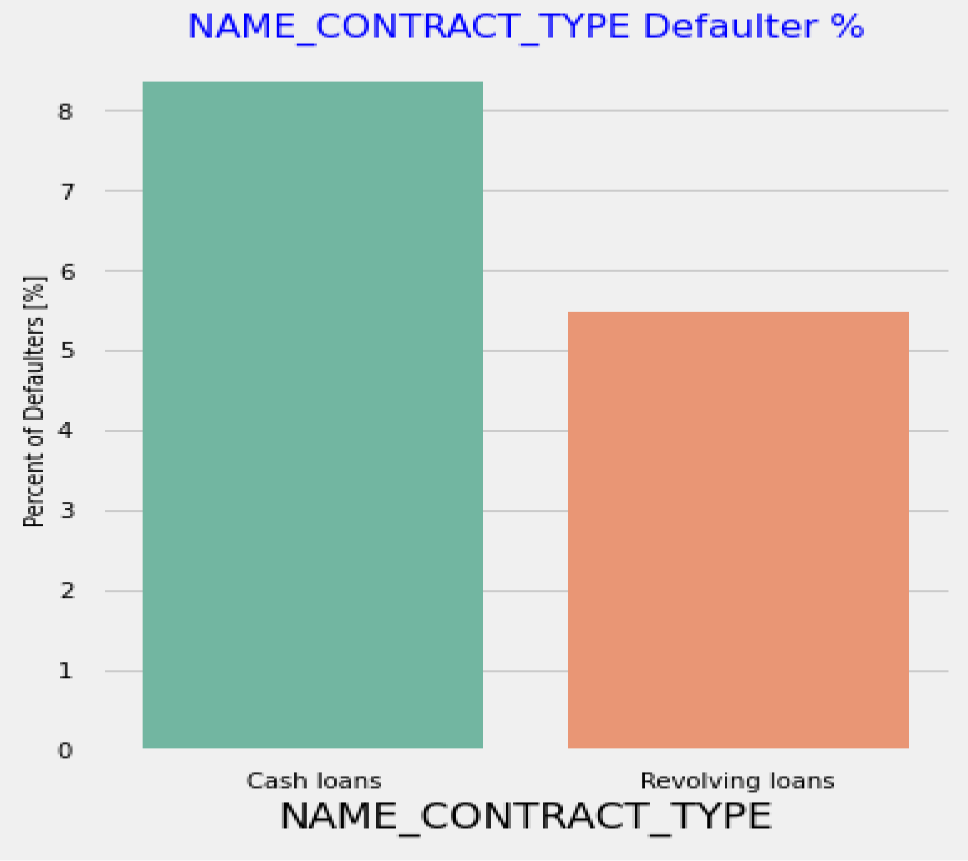
WHO ARE TAKING MORE LOANS?

- **Secondary/Special Educated** People applying for Loans are in Highest No.
- People with **real estate** tends to take more loans
- People tend to take **more cash loans**, and default percentage of revolving loans is less
- People **who don't own a car** tends to take more loans
- **Female** tends to take more loans
- **Married people** tend to take more Loan as compared to other categories.
- People with **House or Apartment** tend to take more loans

UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Contract Type Based On Loan Repayment Status



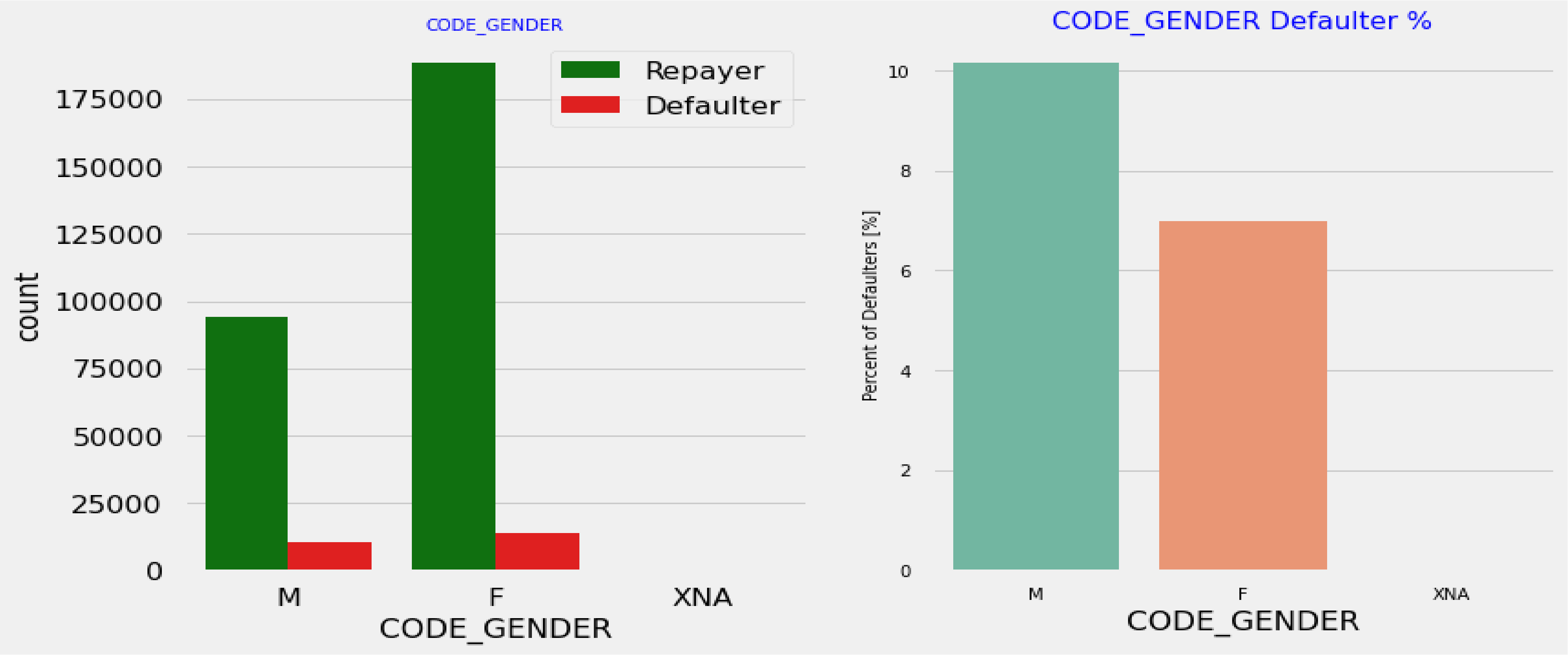


Inferences:

- **Contract type: Revolving loans are just a small fraction (10%) from the total number of loans; in the same time, a larger amount of Revolving loans, comparing with their frequency, are not repaid.**

UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Type Of Gender Based On Loan Repayment Status

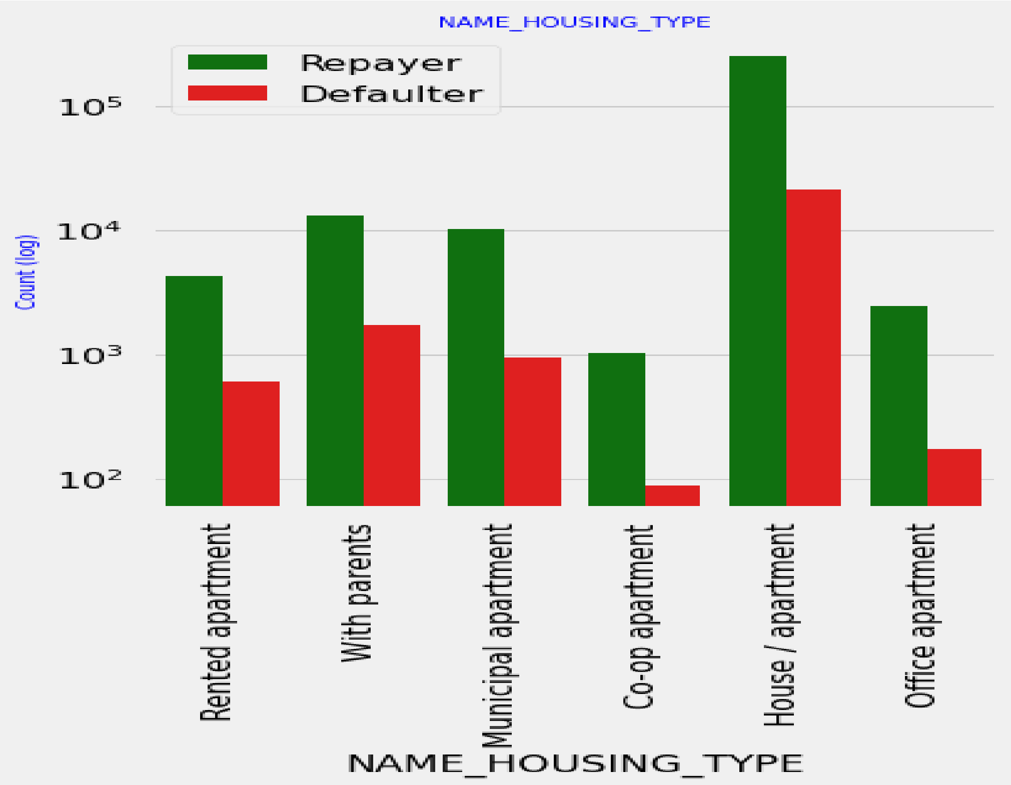


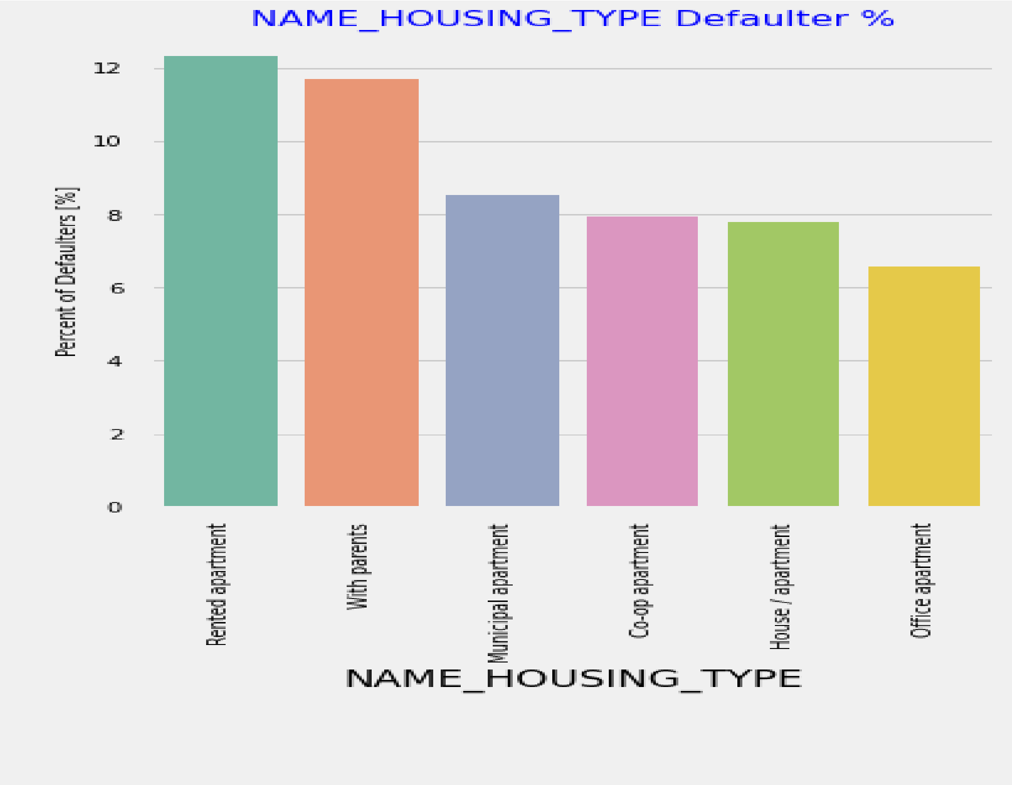
Inferences:

The number of female clients is almost double the number of male clients. Based on the percentage of defaulted credits, males have a higher chance of not returning their loans (~10%), comparing with women (~7%)

UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Housing Type Based On Loan Repayment Status



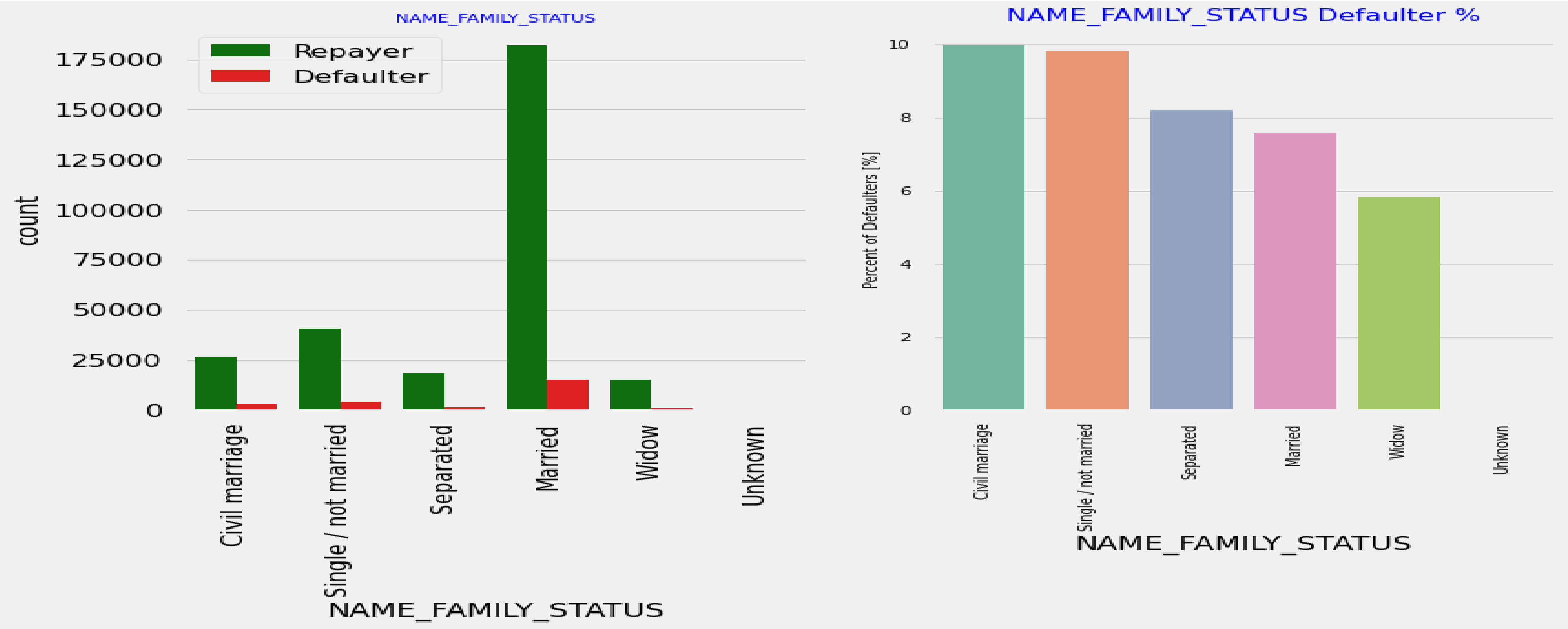


Inferences:

- Majority of people live in House/apartment
- People living in office apartments have lowest default rate
- People living with parents (~11.5%) and living in rented apartments(>12%) have higher probability of defaulting

UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Family Status Based On Loan Repayment Status

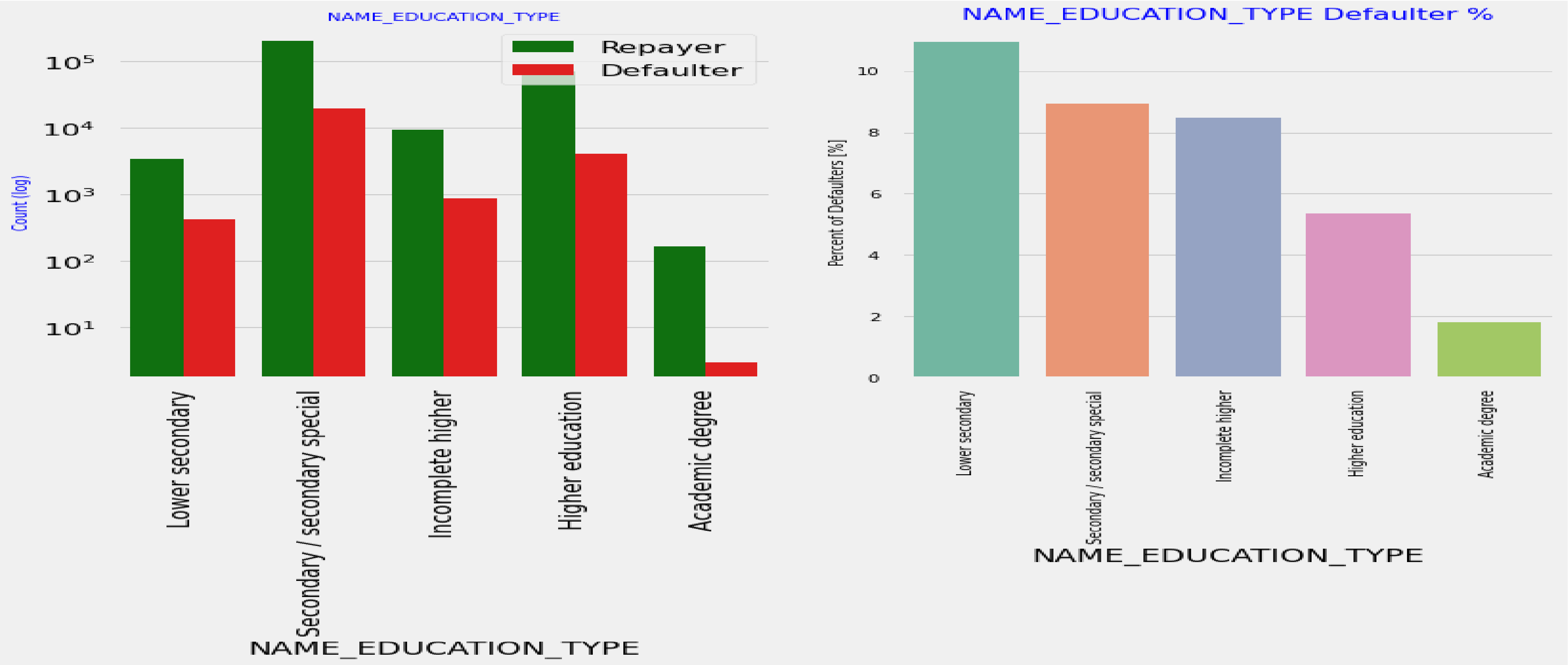


Inferences:

- **Most of the people who have taken loan are married, followed by Single and civil marriage.**
- **In terms of percentage of not repayment of loan, Civil marriage has the highest percent of not repayment (10%), with Widow the lowest (exception being Unknown).**

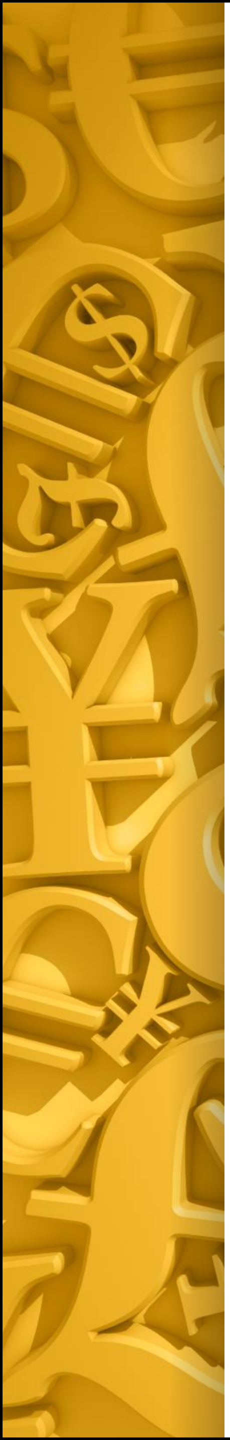
UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Education Type Based On Loan Repayment Status



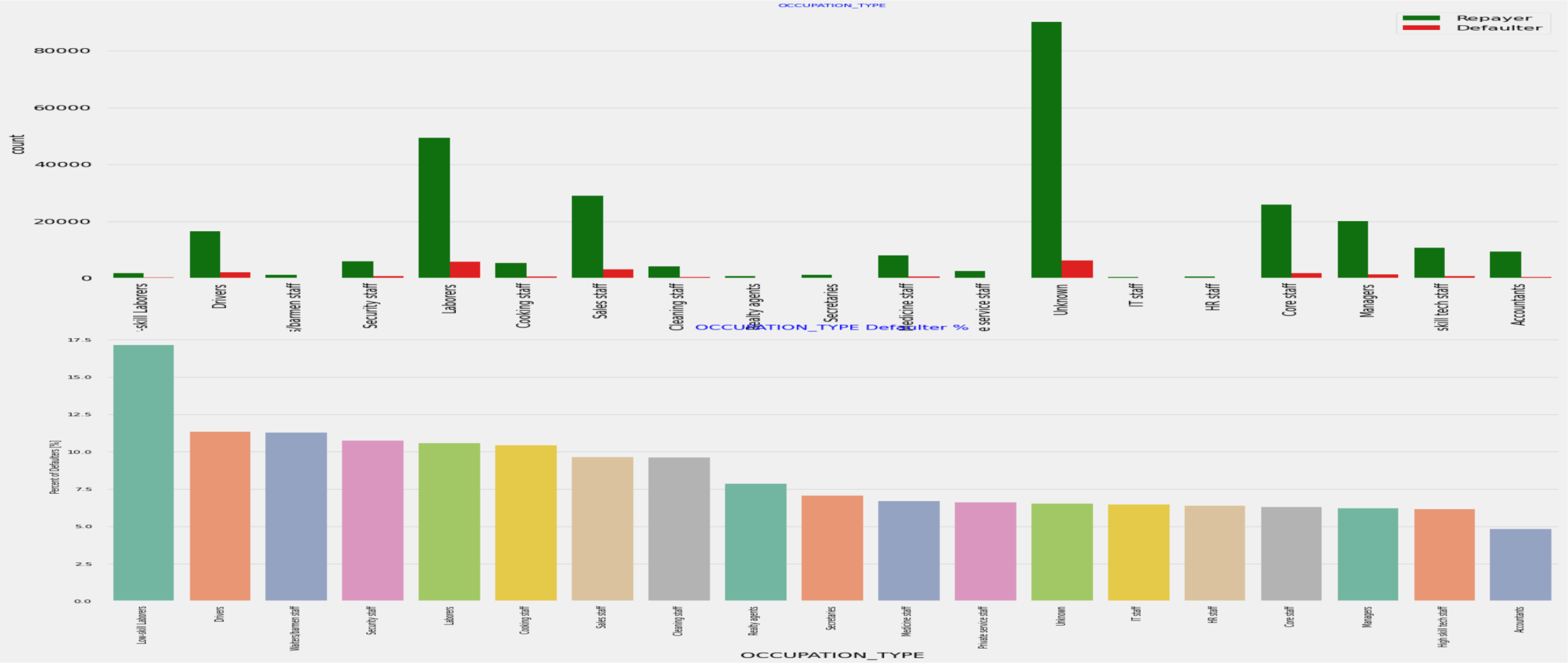
Inferences:

- Majority of the clients have Lower Secondary / secondary special education being Highest Defaulters among all.
- The people with Academic degree have less than 2% defaulting rate.



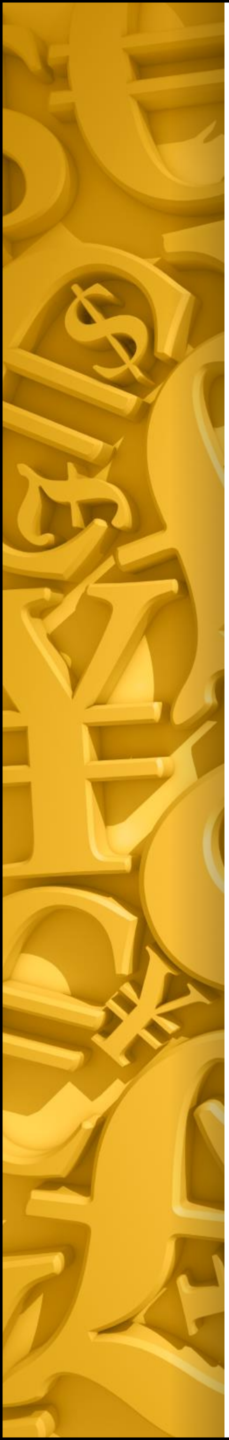
UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Occupation Type Based On Loan Repayment Status



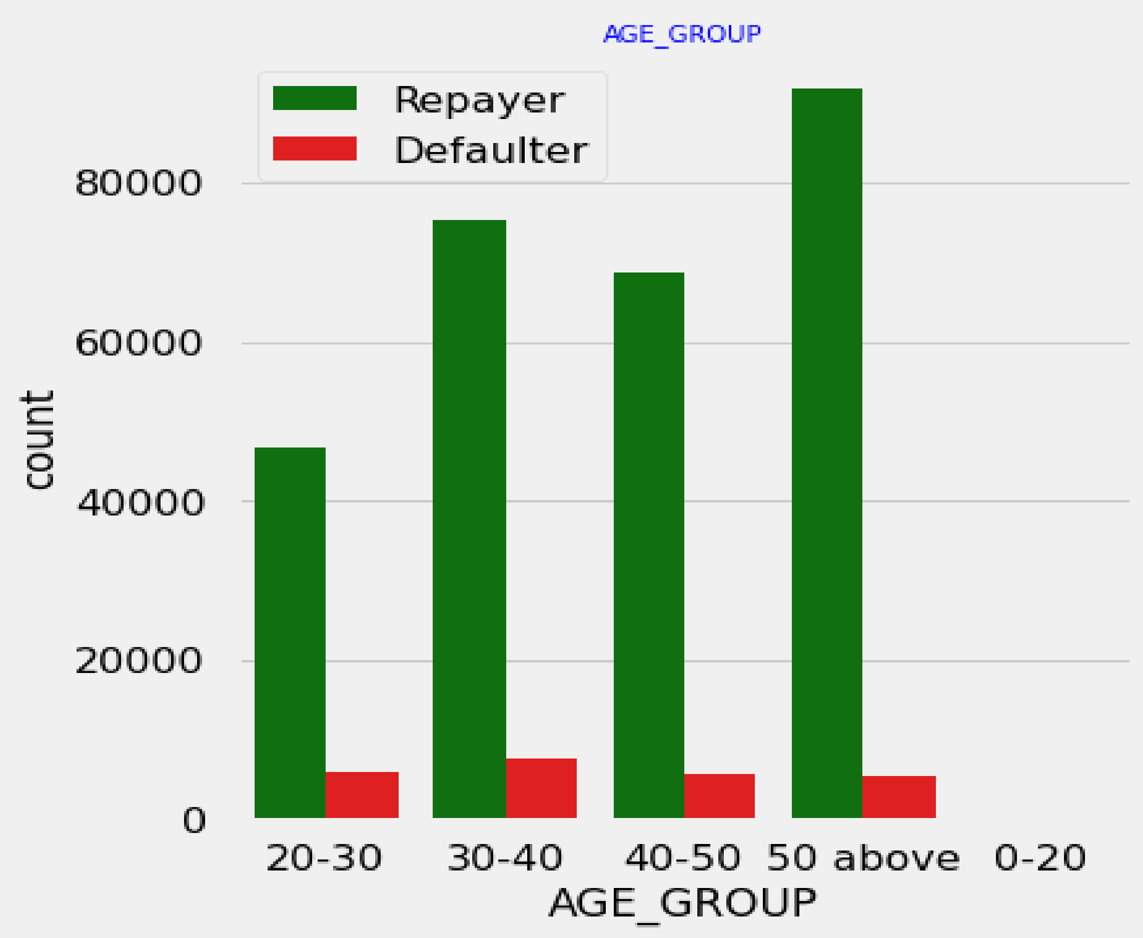
Inferences:

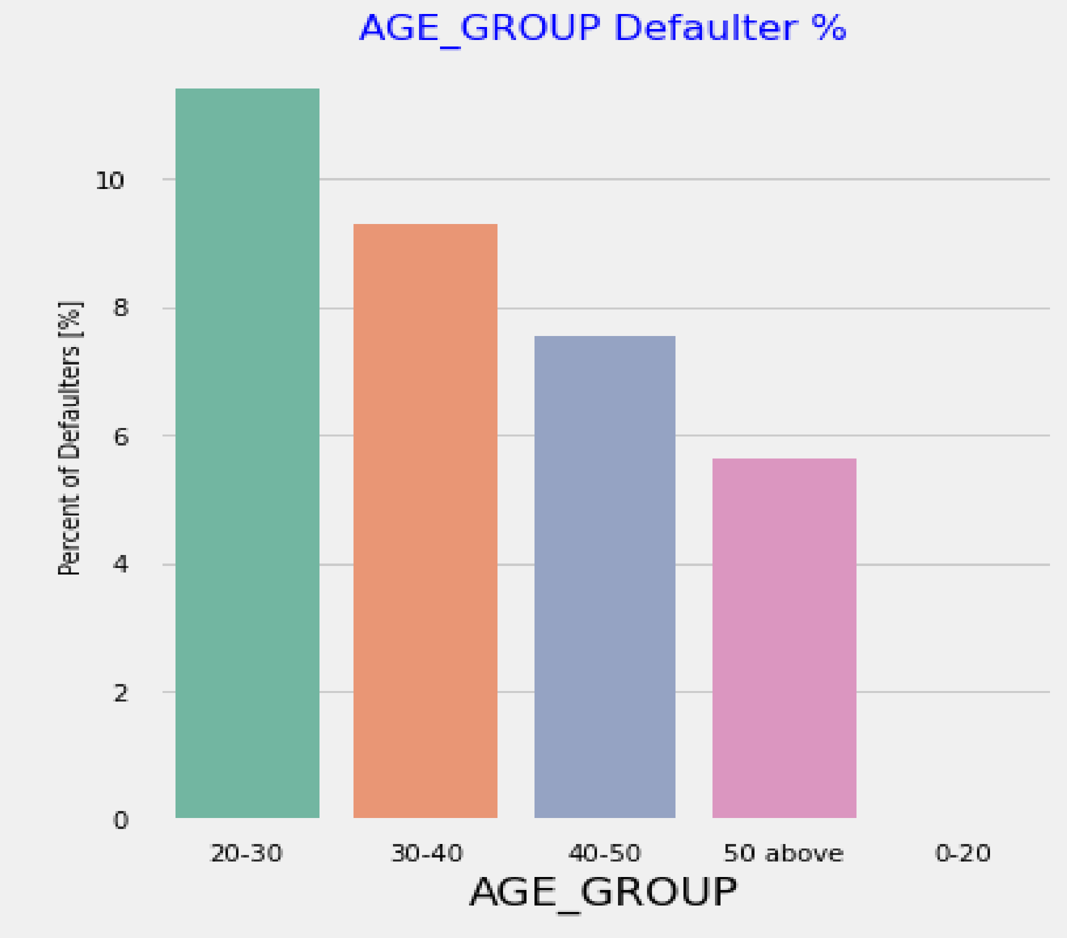
- Most of the loans are taken by Laborers, followed by Sales staff. IT staff take the lowest amount of loans.
- The category with highest percent of not repaid loans are Low-skill Laborers (above 17%), followed by Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff.



UNIVARIATE CATEGORICAL ANALYSIS

Analyzing Age Group Based On Loan Repayment Status



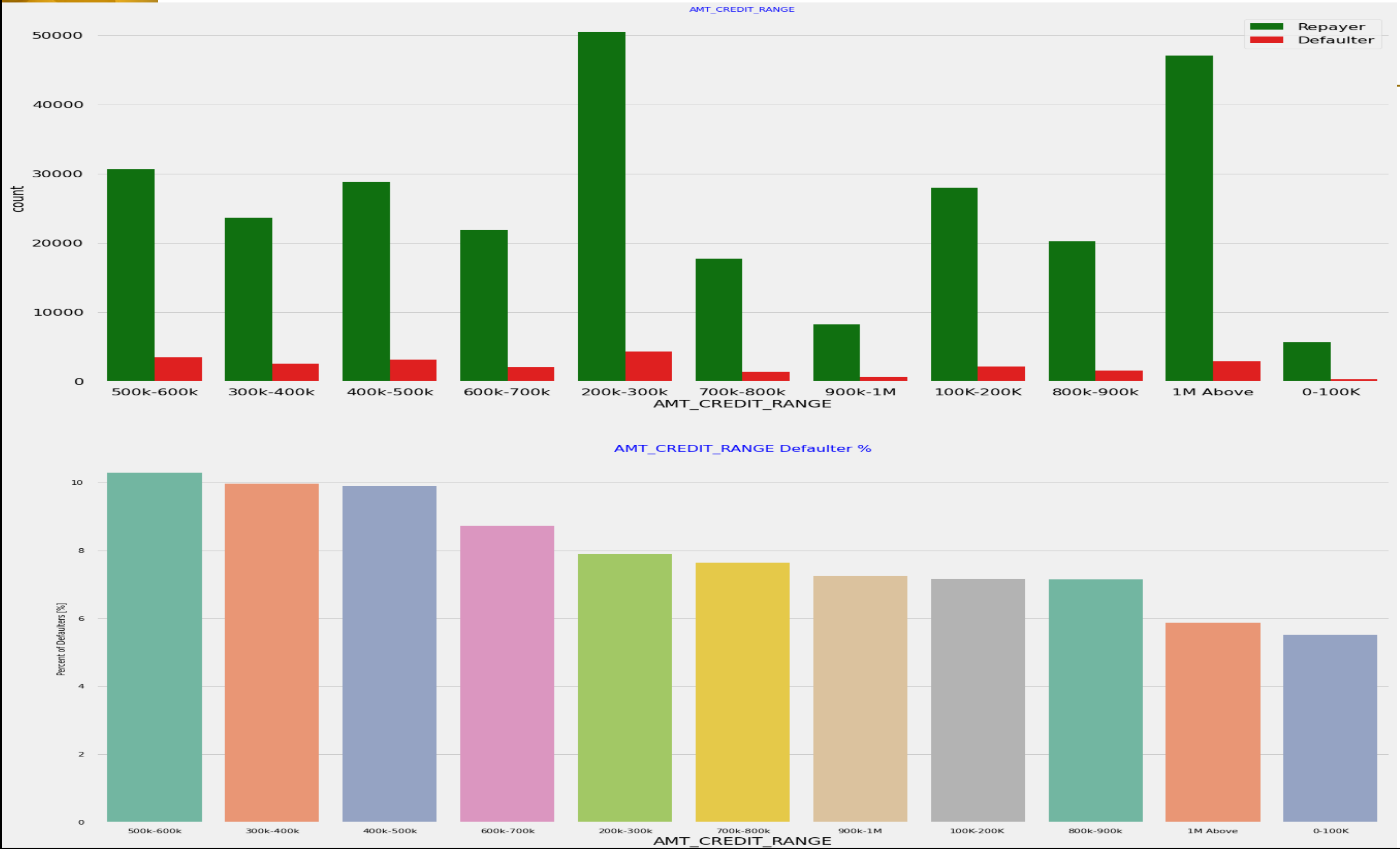


Inferences:

- **People in the age group range 20-40 have higher probability of defaulting**
- **People above age of 50 have low probability of defaulting**

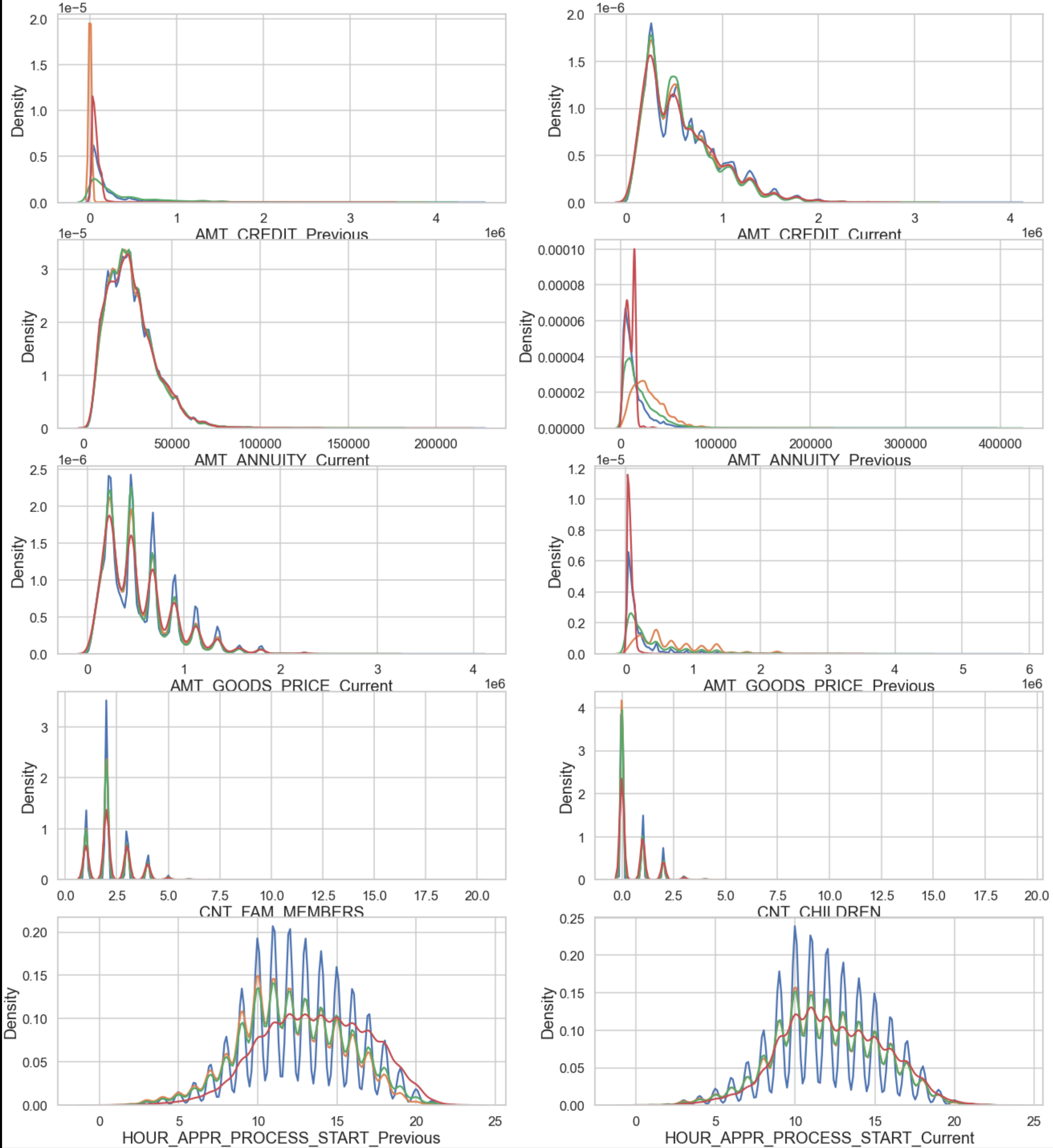
UNIVARIATE NUMERICAL ANALYSIS

Analyzing Amount_Credit Based On Loan Repayment Status



Inferences:

- More than 80% of the loan provided are for amount less than 900,000
- People who get loan for 300-600k tend to default more than others.



UNIVARIATE NUMERICAL CONTINUOUS ANALYSIS

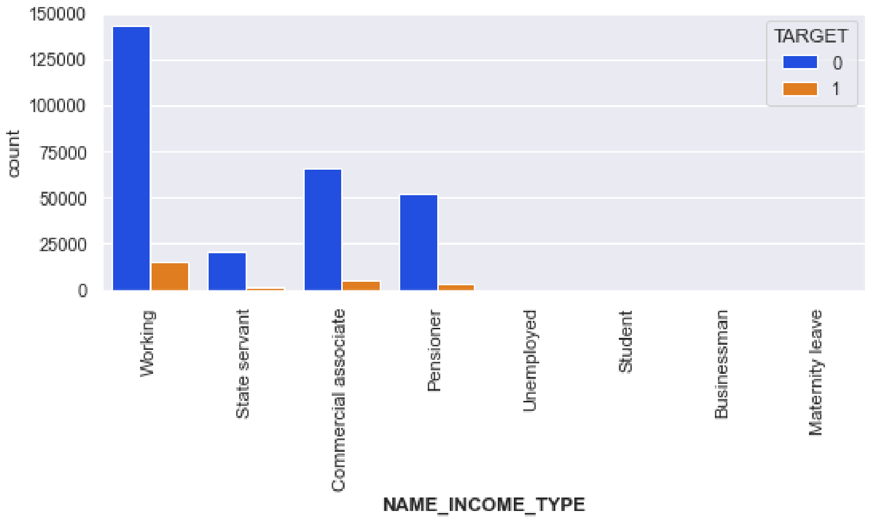
Inferences:

- High number of applications are filed in 9 AM to 2 PM for both Current and Previous data.
- So busiest hours for bank are form 9 AM to 2 PM.
- nuclear family tends to take more loans.
- Previously bank had high unused offers but currently refused is high incase of AMT_GOODS_PRICE.
- Previously bank had high unused offers & currently cancelled/refused offers are similar for AMT_ANNUITY.
- Previously bank had high unused offers & currently high no. of refused offers for AMT_CREDIT.

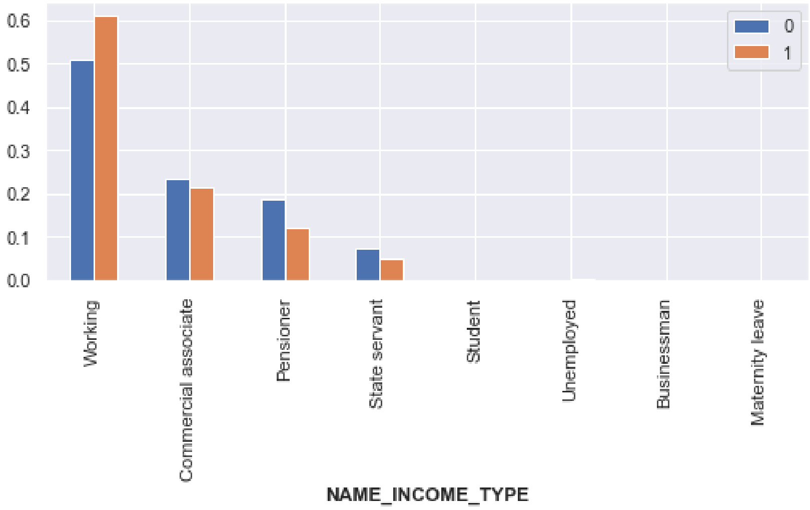
SEGMENTED UNIVARIATE ANALYSIS

(Income Type, Week Day wrt Target Defaulters/Non-Defaulters)

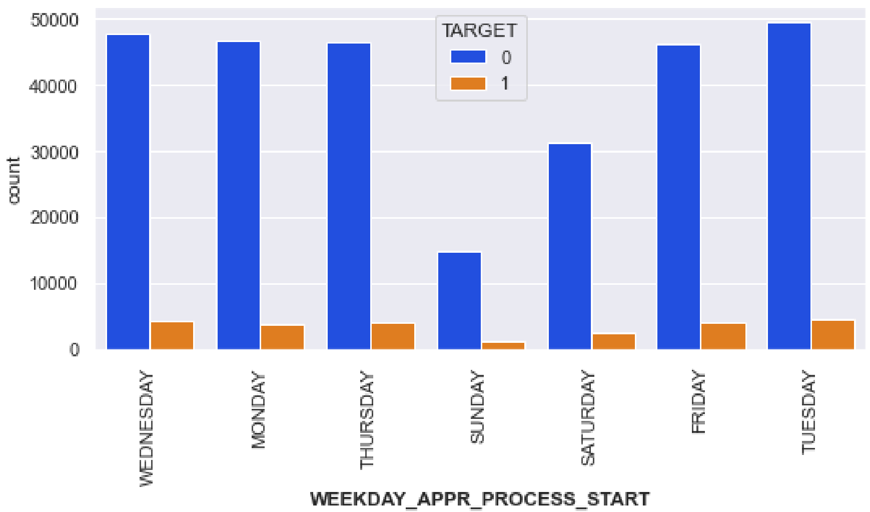
Defaulter_0 & Non-Defaulter_1 Count Chart



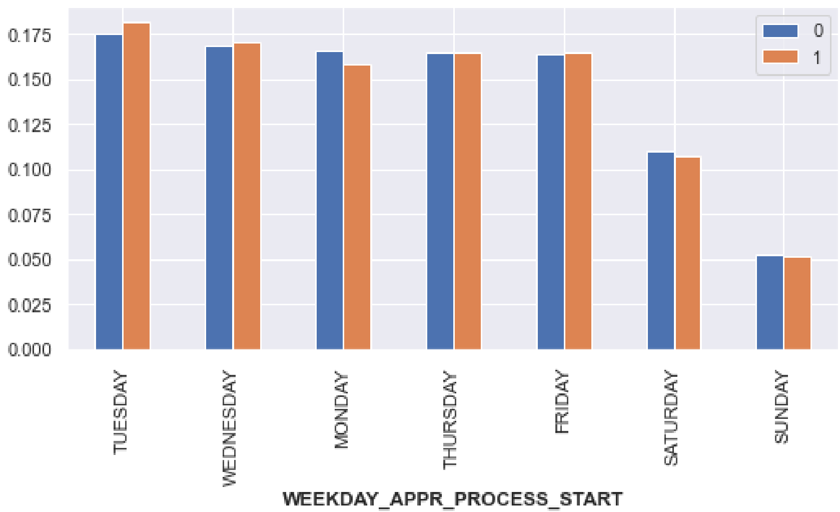
Plotting data for target in terms of percentage



Defaulter_0 & Non-Defaulter_1 Count Chart



Plotting data for target in terms of percentage

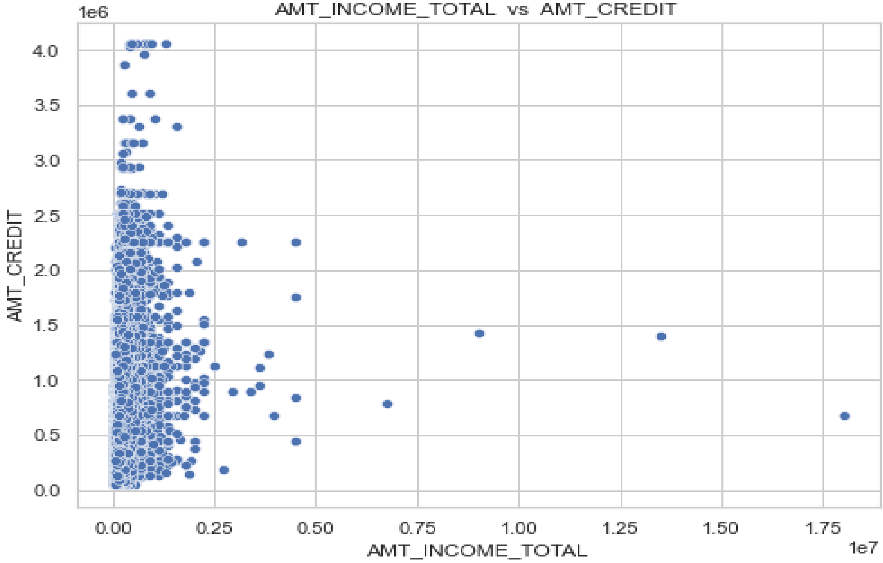


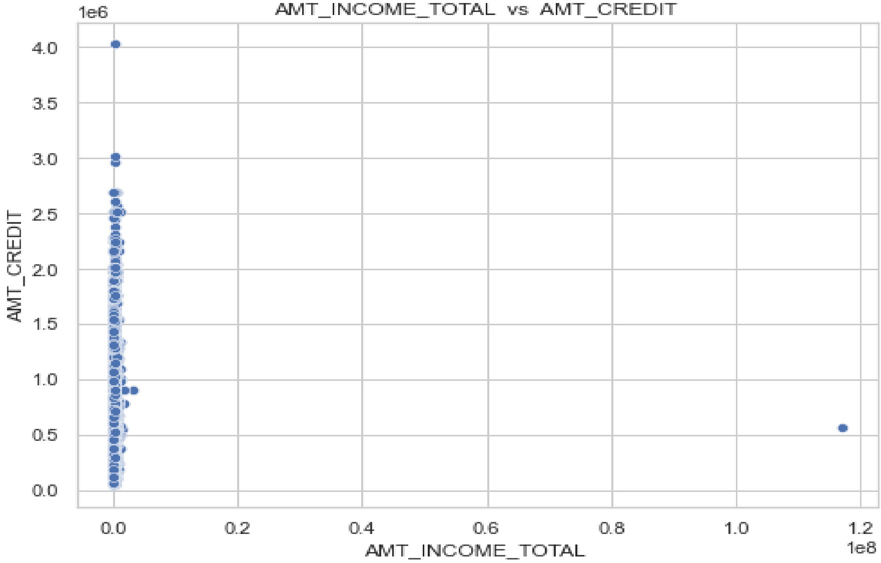
Inferences:

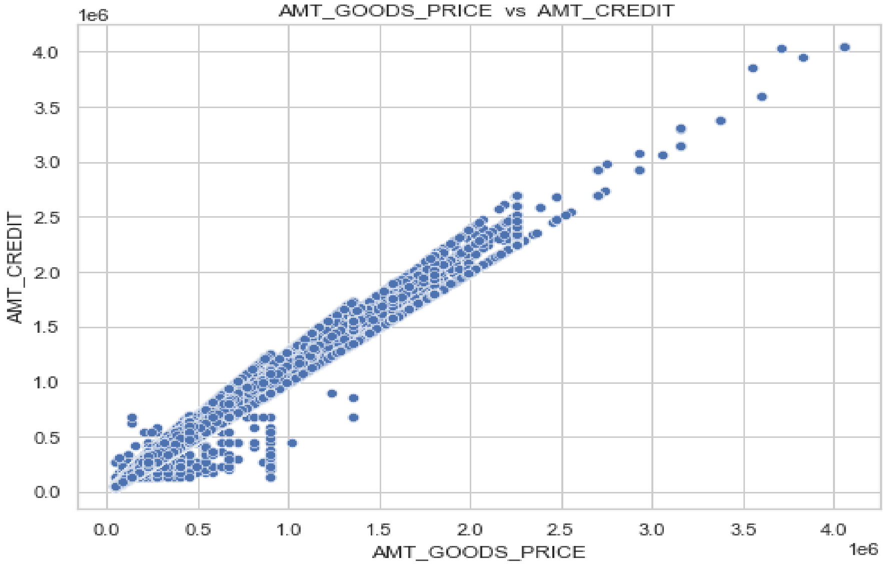
- People having Medium Total Income are more defaulters.
- People having high Credit amount are less defaulters.
- People enrolling for application process on Sunday are less likely to default (min defaulters).
- Saturday and Sunday are least busiest for bank in terms of loan applications.
- Pensioner defaulter is lower than Non-defaulter.

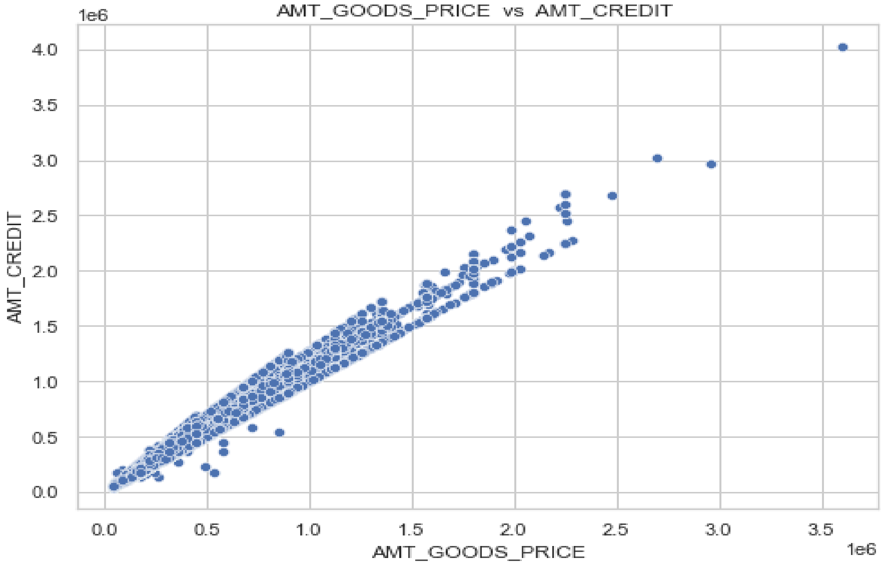
BIVARIATE ANALYSIS OF MERGED DATASET

Analyzing Income vs Credit, Goods Price vs Credit







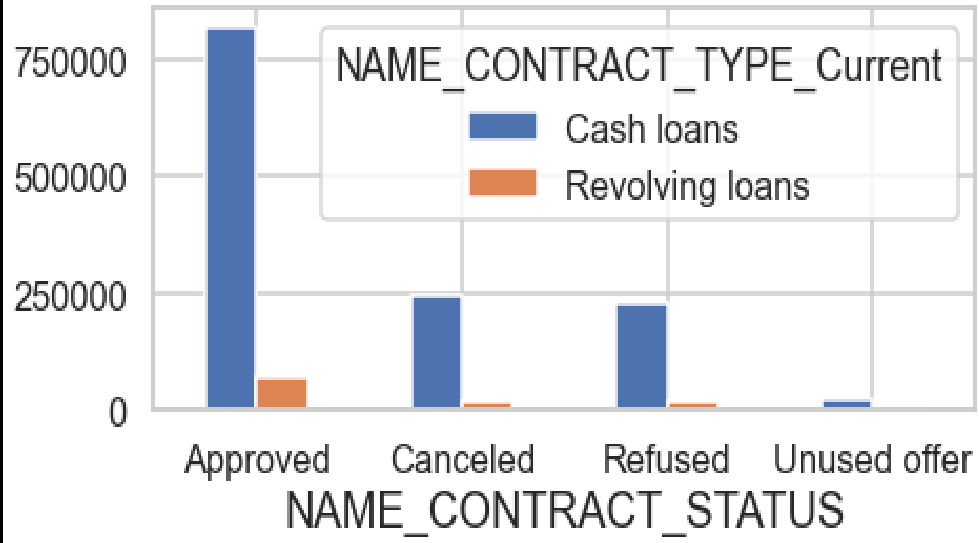


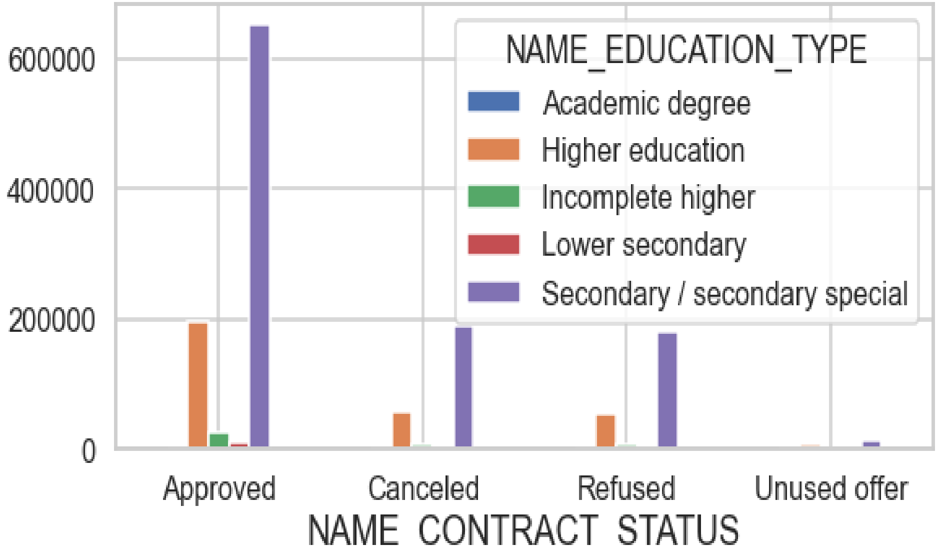
Inferences:

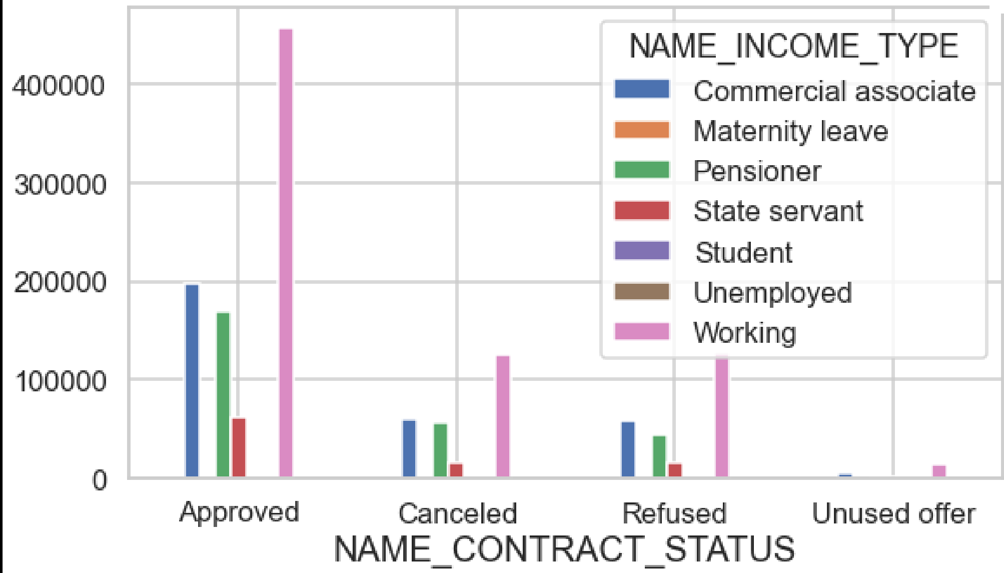
- **AMT_GOODS_PRICE: When the credit amount goes beyond 3M, there is an increase in defaulters.**
- **AMT_CREDIT: People who get loan for 300-600k tend to default more than others and hence having higher interest specifically for this credit range would be ideal.**
- **AMT_INCOME: Since 90% of the applications have Income total less than 300,000 and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.**

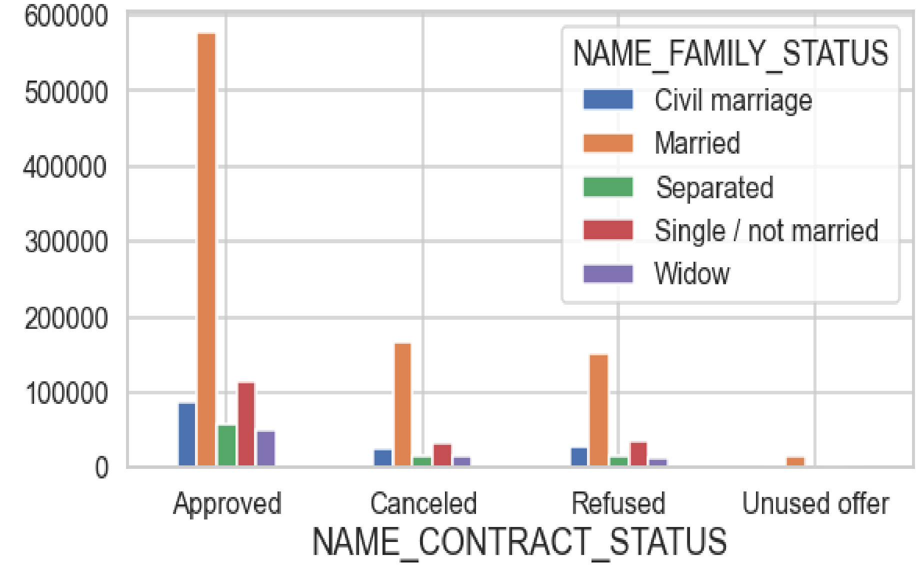
BIVARIATE ANALYSIS OF MERGED DATASET

Analyzing Education, Contract ,Income, Family Type with Contract Status







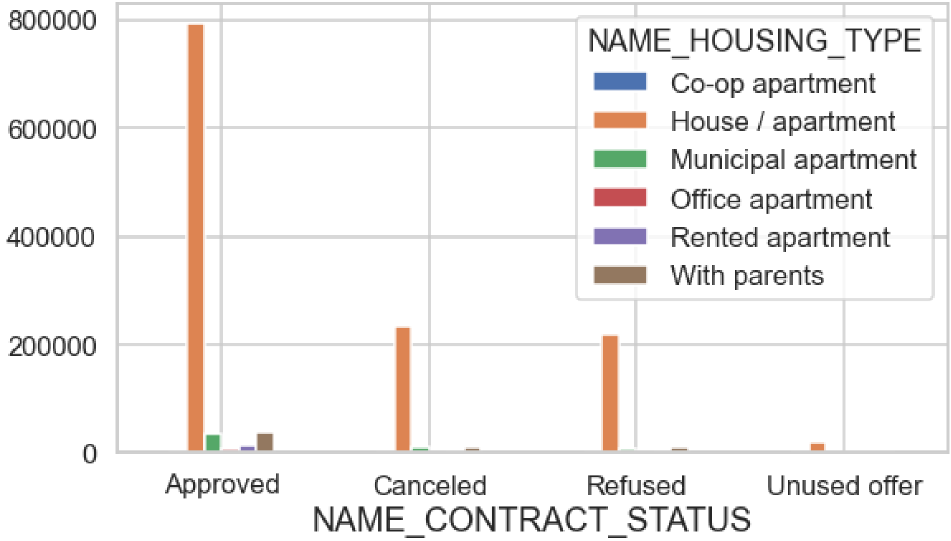


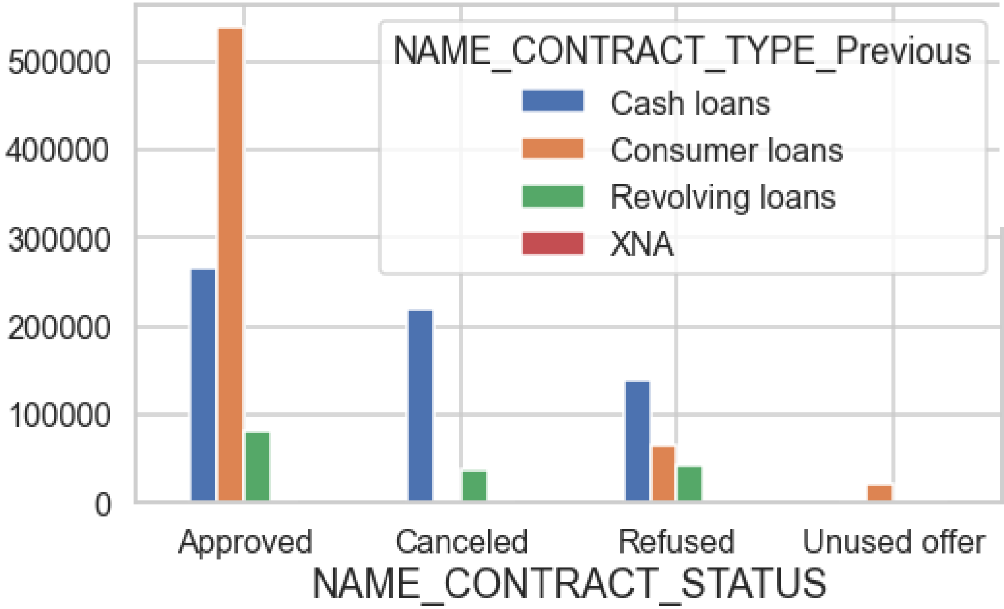
Inferences:

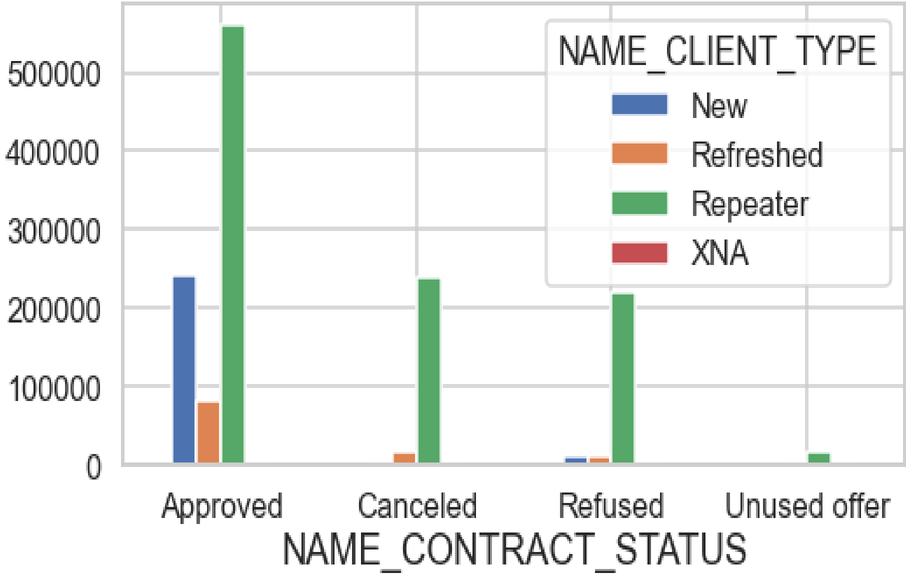
- **Cash loans has highest count of Approved loans**
- **Working applicant have Highest no. of approvals.**
- **Secondary/Secondary -special-educated applicant have Highest no. of approvals.**
- **Married applicant got Highest number of approvals.**

BIVARIATE ANALYSIS OF MERGED DATASET

Analyzing Housing Type, Previous Contract, Client Type with Contract Status



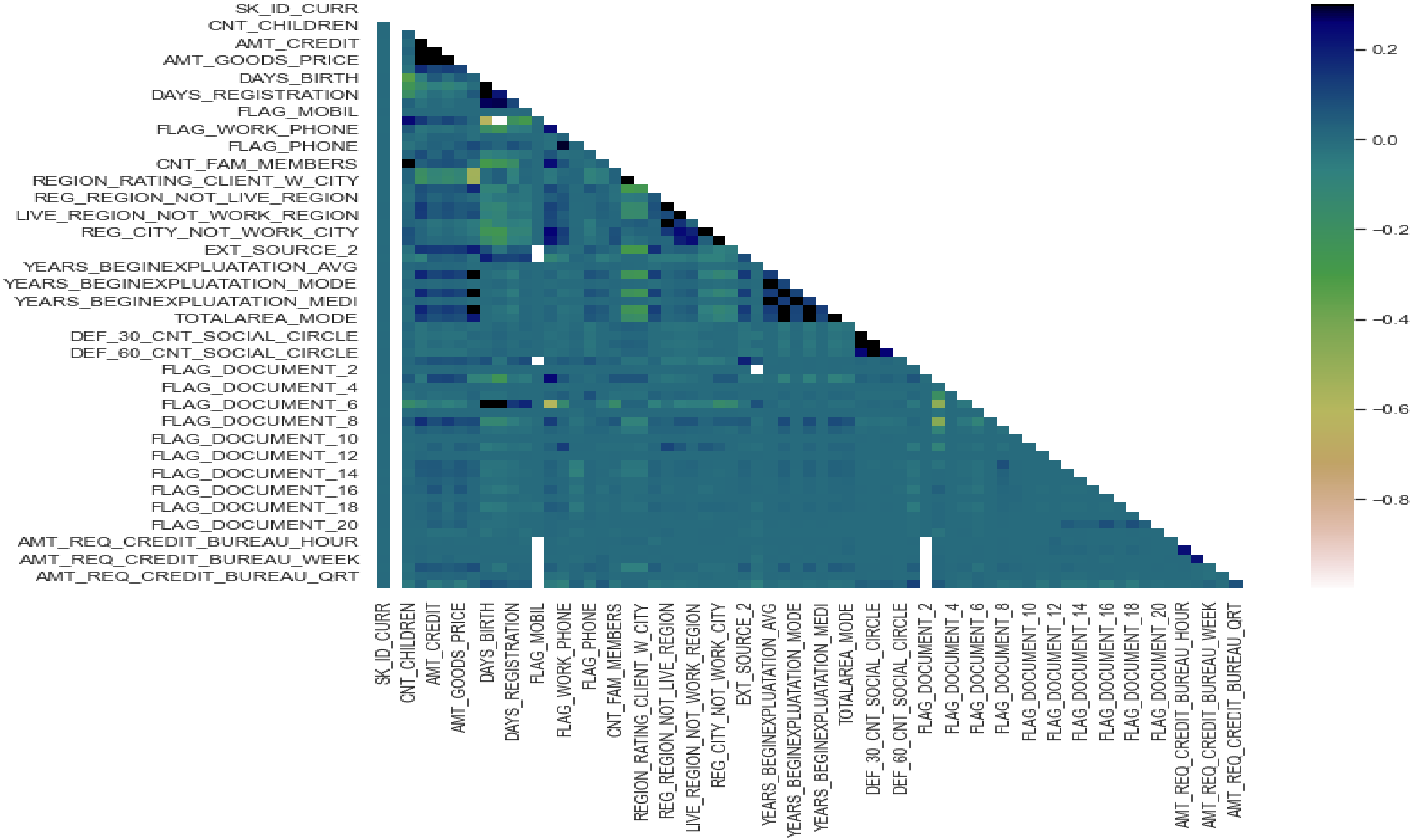




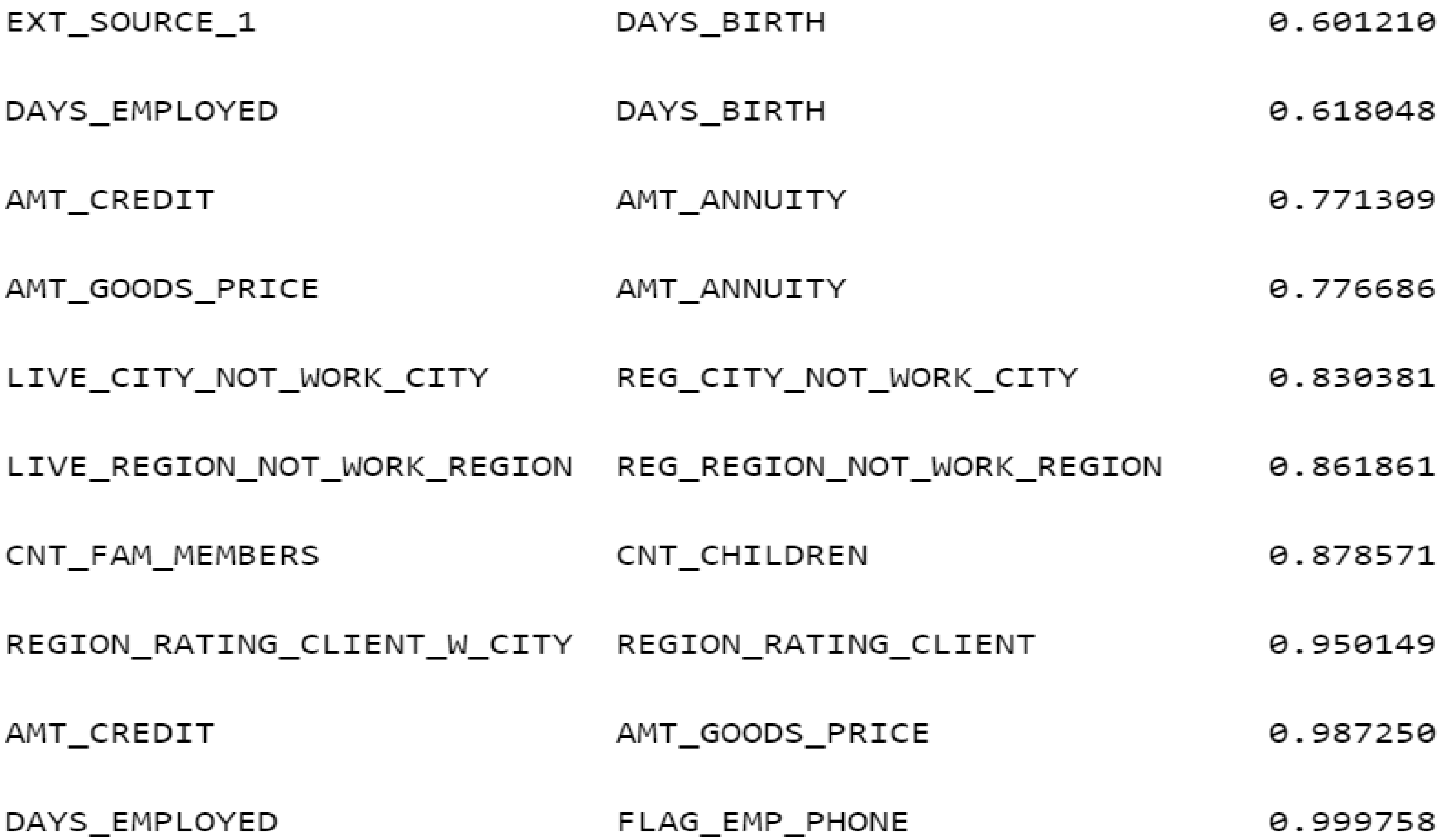
Inferences:

- House/apartment owner have Highest number of approvals
- Consumer Loans have Highest number of approvals.
- Most no. of times Repeated Applications got approved.

NON - DEFAULTER CORRELATION

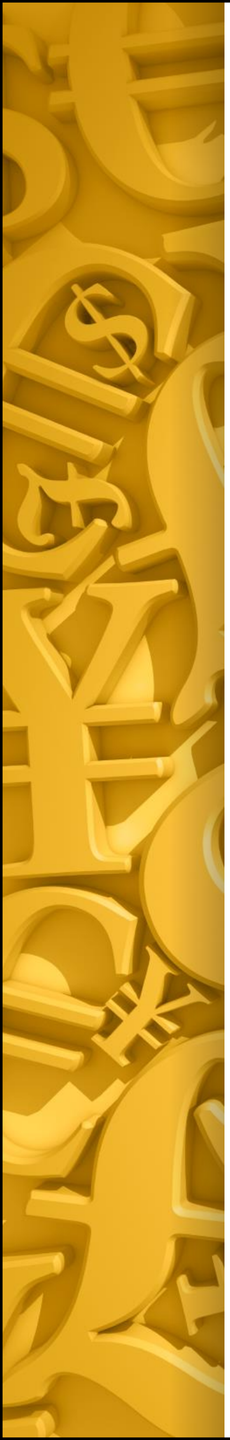


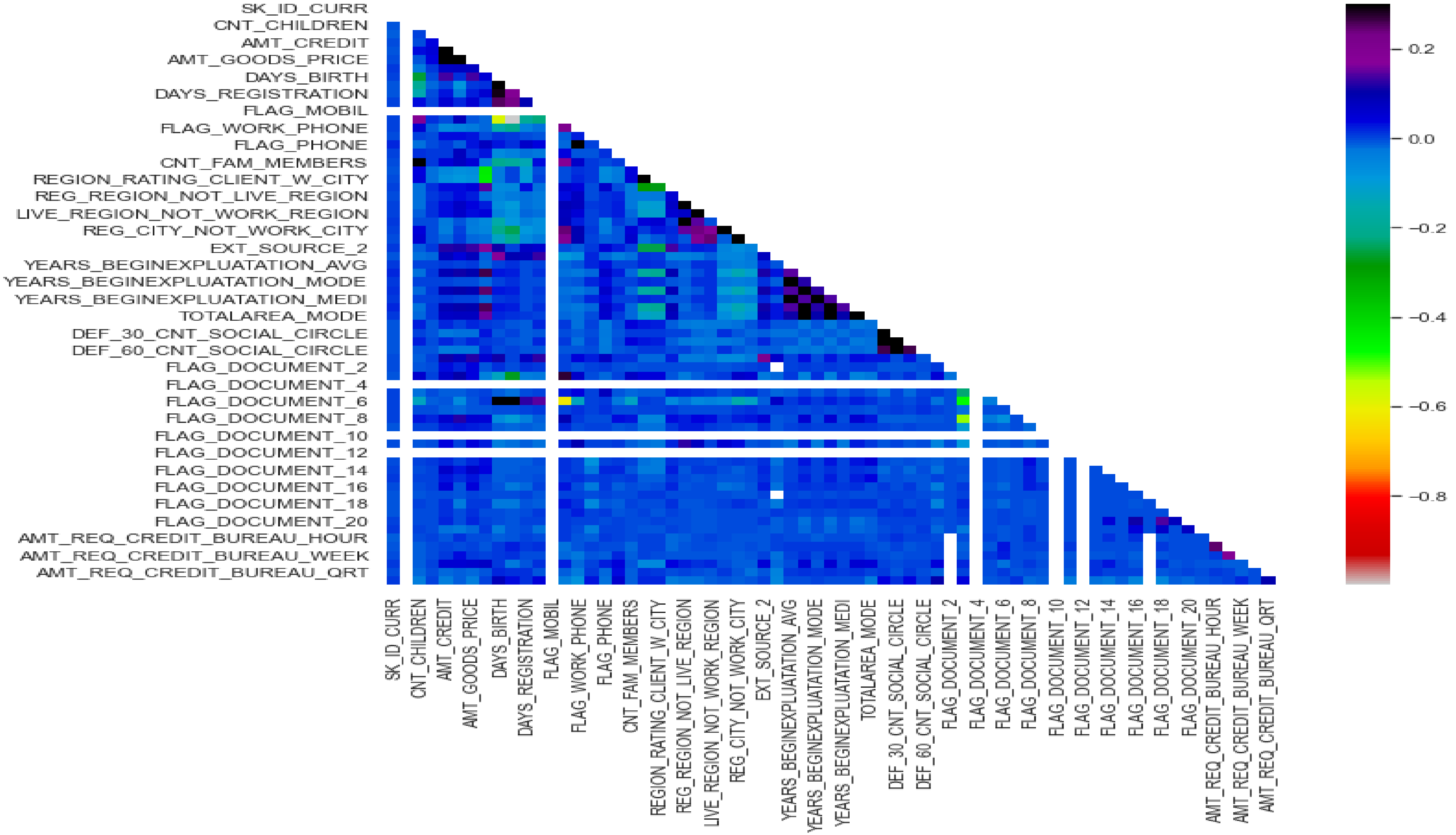
TOP NON - DEFAULTER CORRELATIONS

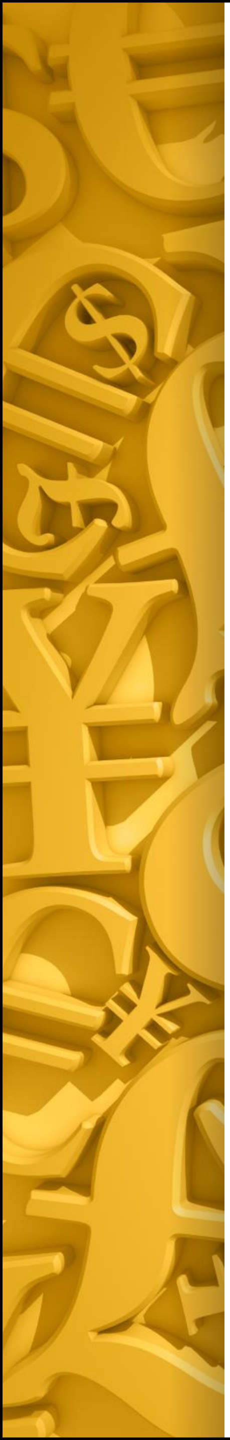


```
EXT_SOURCE_1                 DAYS_BIRTH                   0.601210

DAYS_EMPLOYED                DAYS_BIRTH                   0.618048

AMT_CREDIT                   AMT_ANNUITY                  0.771309

AMT_GOODS_PRICE              AMT_ANNUITY                  0.776686

LIVE_CITY_NOT_WORK_CITY      REG_CITY_NOT_WORK_CITY       0.830381

LIVE_REGION_NOT_WORK_REGION  REG_REGION_NOT_WORK_REGION   0.861861

CNT_FAM_MEMBERS              CNT_CHILDREN                 0.878571

REGION_RATING_CLIENT_W_CITY  REGION_RATING_CLIENT         0.950149

AMT_CREDIT                   AMT_GOODS_PRICE              0.987250

DAYS_EMPLOYED                FLAG_EMP_PHONE               0.999758
```

DEFAULTER CORRELATION

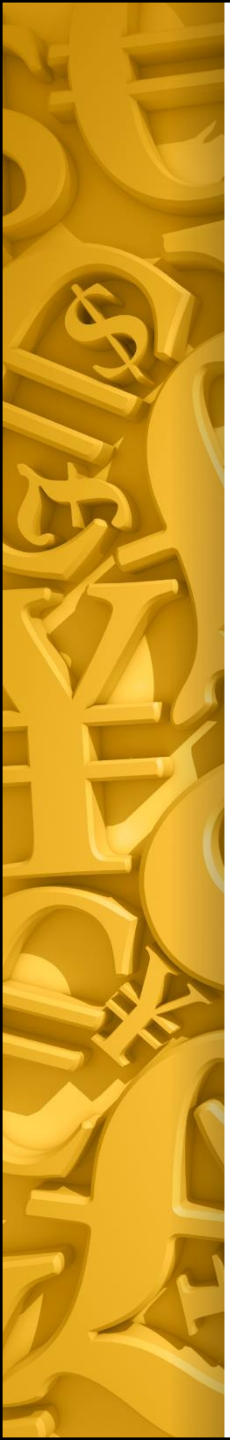






TOP DEFAULTER CORRELATIONS

EXT_SOURCE_1	DAYS_BIRTH	0.570054
DAYS_EMPLOYED	DAYS_BIRTH	0.575097
FLAG_EMP_PHONE	DAYS_BIRTH	0.578519
AMT_CREDIT	AMT_ANNUITY	0.752195
AMT_GOODS_PRICE	AMT_ANNUITY	0.752699
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.778540
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.847885
CNT_CHILDREN	CNT_FAM_MEMBERS	0.885484
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956637
AMT_CREDIT	AMT_GOODS_PRICE	0.983103
DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999702



DECISIVE FACTOR WHETHER AN APPLICANT WILL BE REPAYER

1. **NAME_EDUCATION_TYPE** : Academic degree has less defaults.

2. **NAME_INCOME_TYPE** : Student and Businessmen have no defaults.

3. **REGION_RATING_CLIENT** : RATING 1 is safer.

4. **ORGANIZATION_TYPE** : Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%

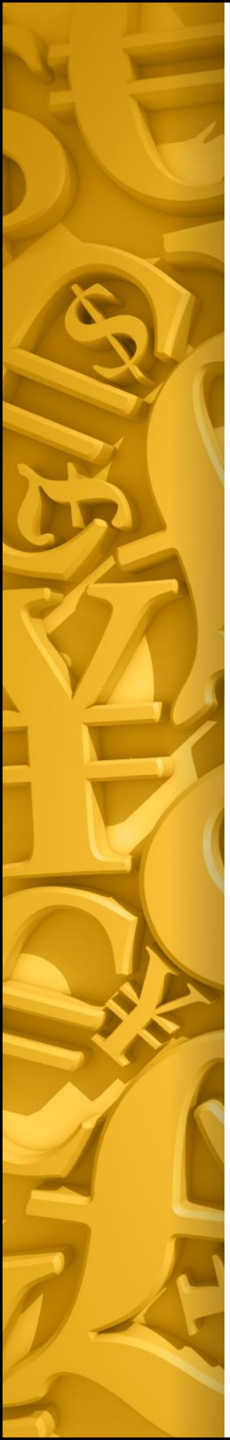
5. **DAYS_BIRTH** : People above age of 50 have low probability of defaulting

6. **DAYS_EMPLOYED** : Clients with 40+ year experience having less than 1% default rate

7. **AMT_INCOME_TOTAL** : Applicant with Income more than 700,000 are less likely to default

8. **NAME_CASH_LOAN_PURPOSE** : Loans bought for Hobby, Buying garage are being repayed mostly.

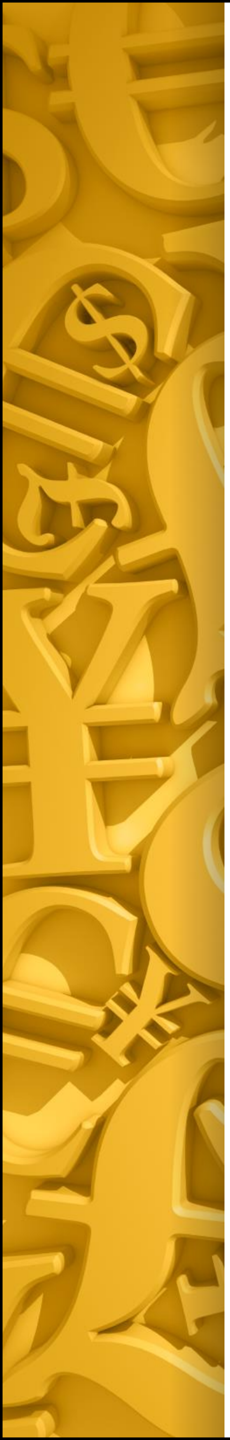
9. **CNT_CHILDREN**: People with zero to two children tend to repay the loans.



DECISIVE FACTOR WHETHER AN APPLICANT WILL BE DEFAULTER

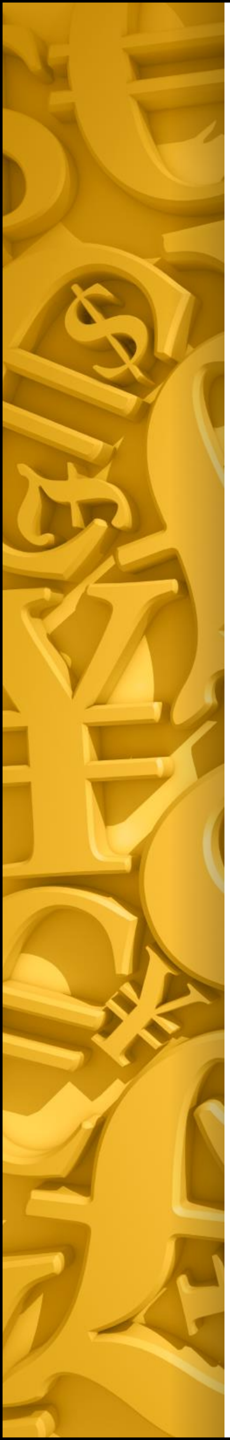
1. **CODE_GENDER** : Men are at relatively higher default rate
2. **NAME_INCOME_TYPE** : Clients who are either at Maternity leave OR Unemployed default a lot.
3. **NAME_EDUCATION_TYPE** : People with Lower Secondary & Secondary education
4. **REGION_RATING_CLIENT** : People who live in Rating 3 has highest defaults.
5. **OCCUPATION_TYPE** : Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as the default rate is huge.
6. **NAME_FAMILY_STATUS** : People who have civil marriage or who are single default a lot.

Continued.. ⏭



DECISIVE FACTOR WHETHER AN APPLICANT WILL BE DEFAULTER

7. **DAYS_BIRTH** : Avoid young people who are in age group of 20-40 as they have higher probability of defaulting
8. **ORGANIZATION_TYPE** : Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.
9. **DAYS_EMPLOYED**: People who have less than 5 years of employment have high default rate.
10. **CNT_CHILDREN & CNT_FAM_MEMBERS** : Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
11. **AMT_GOODS_PRICE** : When the credit amount goes beyond 3M, there is an increase in defaulters.



The following **ATTRIBUTES** indicate that people from these **category tend to default** but then due to the number of people and the amount of loan, the bank could provide loan with **HIGHER INTEREST TO MITIGATE ANY DEFAULT RISK THUS PREVENTING BUSINESS LOSS** :

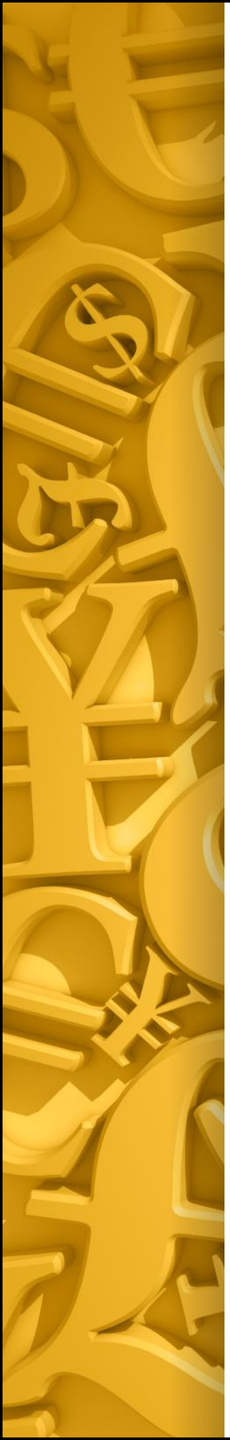
1. **NAME_HOUSING_TYPE :** High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.
2. **AMT_CREDIT :** People who get loan for 300-600k tend to default more than others and hence having higher interest specifically for this credit range would be ideal.
3. **CNT_CHILDREN & CNT_FAM_MEMBERS :** Clients who have 4 to 8 children has a very high default rate and hence higher interest should be imposed on their loans.
4. **NAME_CASH_LOAN_PURPOSE :** Loan taken for the purpose of Repairs seems to have highest default rate. A very high number applications have been rejected by bank or refused by client in previous applications as well which has purpose as repair or other. This shows that purpose repair is taken as high risk by bank and either they are rejected, or bank offers very high loan interest rate which is not feasible by the clients, thus they refuse the loan. The same approach could be followed in future as well.
5. **AMT_INCOME :** Since 90% of the applications have Income total less than 300,000 and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.

CONCLUSION

TO WHOM SHOULD LOAN CAN BE DISAPPROVED?

(Risky Group Category Of DEFAULTERS)

- **Lower Secondary Educated Clients** are the Highest & Most Riskiest Clients in no. to be Defaulters, especially when their Previous Loans were Cancelled Or Refused.
- **Male** Clients with **Civil Marriage**
- **Previously Refused** Loan Status **Group**.
- Banks should focus less on the **'Working' Income Type**, which has the highest number of failed payments.
- People with **Medium** total income are more likely to default
- Considering **Loan purpose 'Repair'** is having Greater No. of Unsuccessful Payments on time.

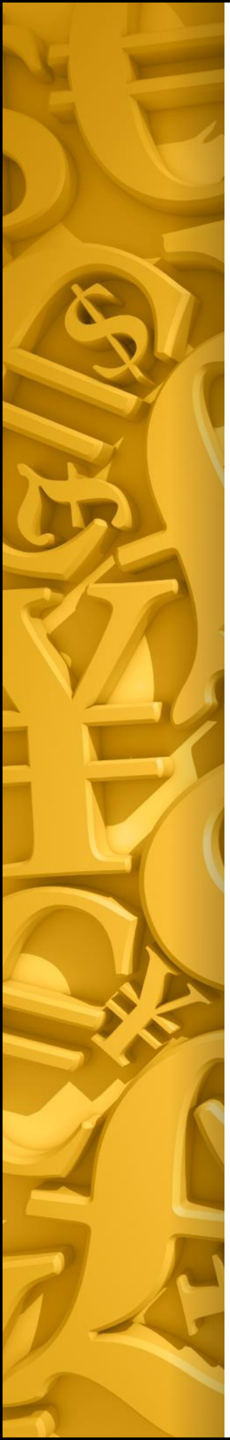


CONCLUSION

TO WHOM SHOULD LOAN CAN BE APPROVED?

(Recommended Group Category To Have Non-Defaulter)

- Banks should focus on **'Students' ,'Pensioner**' for successful repayments.
- Client with the **Higher Income Category**.
- **FEMALE** Client with **Higher Education**.
- Clients who are working as **STATE Servant**.
- **Old People** of any Income Group.
- Client’s who’s **Previous Loan** was **Approved**.
- **Old Female** Client
- For successful payments, banks should focus more on contract types such as "**businessman"**, "pensioner" and "student" with **housing types** other than "co-op unit.“
- Get as much as clients from housing type ‘**With parents’** as they have the Fewest Failed Payments.
- **Repeater** has highest number of approved loans.



SUBMITTED BY:

GURPREET KAUR

DSC43