

L O A N

Credit EDA Case Study

Compiled By : AKHILESH PRASAD

RIDHAM SHAH

DS C35 DATA SCIENCE

Objective

This case study aims to identify patterns 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.

In other words, the company can understand the driving factors, behind the loan default, the variables which are strong indicators of loan default.

RISKY APPLICANTS

Following are the types of applicants with high risk factor involved:

- Those have been employed for an average of less than 3 years,
- Men with lower secondary Education,
- Age group of <30 earning less than 1Lakh, living in rented apartment,
- Those with region rating 3 and score less than 0.5,
- Unemployed category Pensioners in the age group of <30 are at greater risk of payment difficulty,
- Lower secondary education-Civil Marriage & Single people,
- People living in Rented apartments,
- Age group <30 living in rented apartments shows a greater chance of exhibiting payment difficulty,
- Low-skill laborers,
- Male realty agents,
- Female Waiter/Barmen staff.

Our Overall Observations(According to us) :

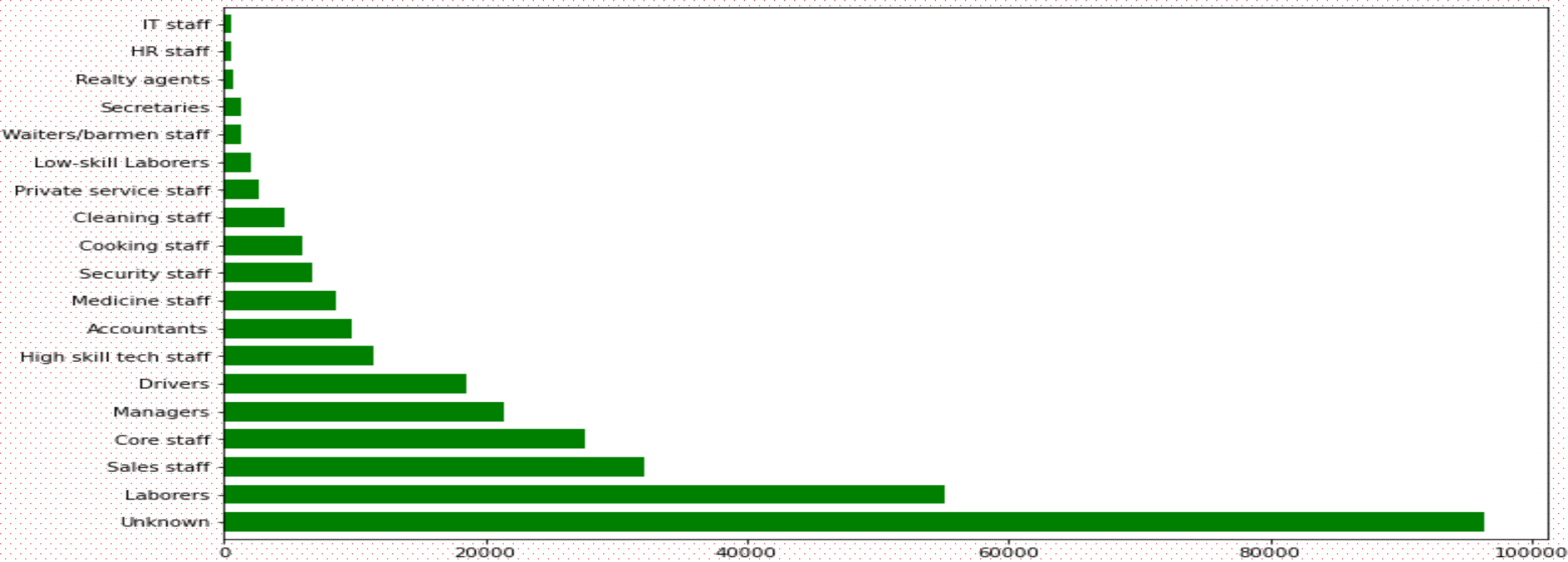
- Customers with the higher rejection ratio and default in the current application (on the lower bottom of graph 1(age group -20 -50) are with high risk profiles hence could be rejected the loans if their external credit score is less and income levels are relatively low and credit amount applied is high,
- More than 25% people who have made no default in the current loan have a lower rejection ratio(higher approval ratio, $>.75$) across all age groups, They are low risk customers who could be given higher credit loans in future applications. One thing to be mindful here is the number of approved loans a particular customer holding till date,
- Customers who had a higher approval ratio but defaulted with the current application could be granted with loans with lesser credit amount since a higher credit amount would attract higher interest charges and that would further stress the customer financially.

Our Overall Observations(According to us) :

- Males exhibited more payment difficulty although there were more female applicants,
- Low skilled labours are more likely to payment default,
- As expected unemployed and lower income group exhibited higher payment default tendency.
- As the employment years of a person is more he is less likely to make default,

Highest percentage of values belongs to unknown group and Second belongs to laborers

Percentage of type of Occupations



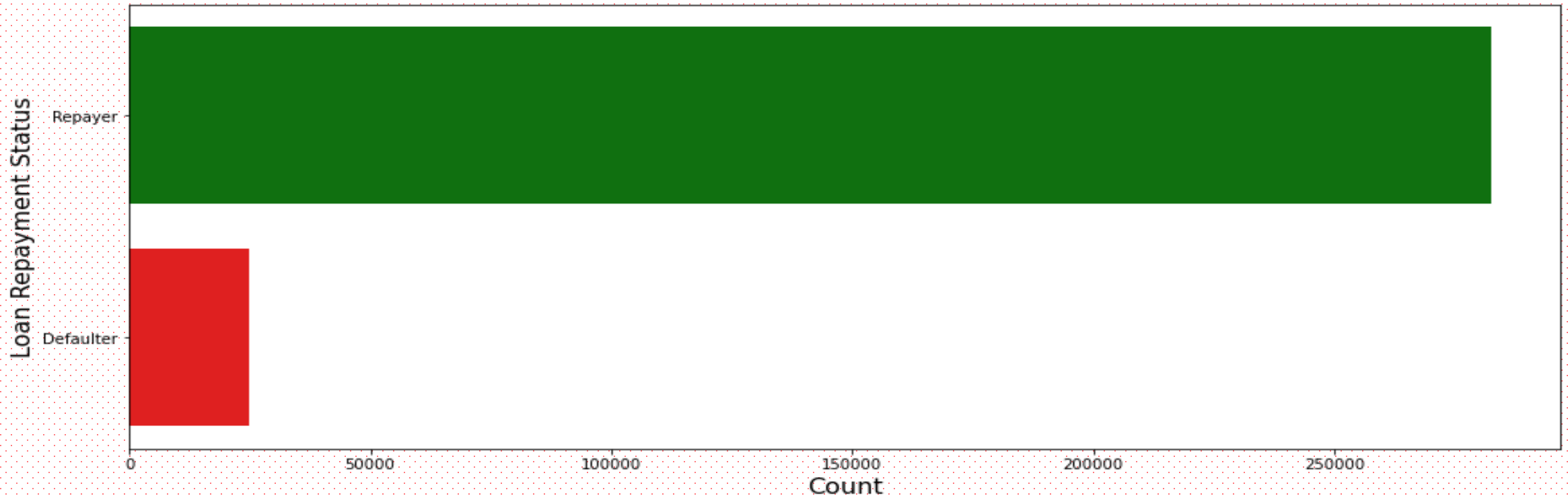
Ratio of imbalance percentage with respect to defaulter and repayer

Repayer Percentage is 91.93%

Defaulter Percentage is 8.07%

Imbalance Ratio with respect to repayer and Defaulter is given: 11.39/1
(approx.)

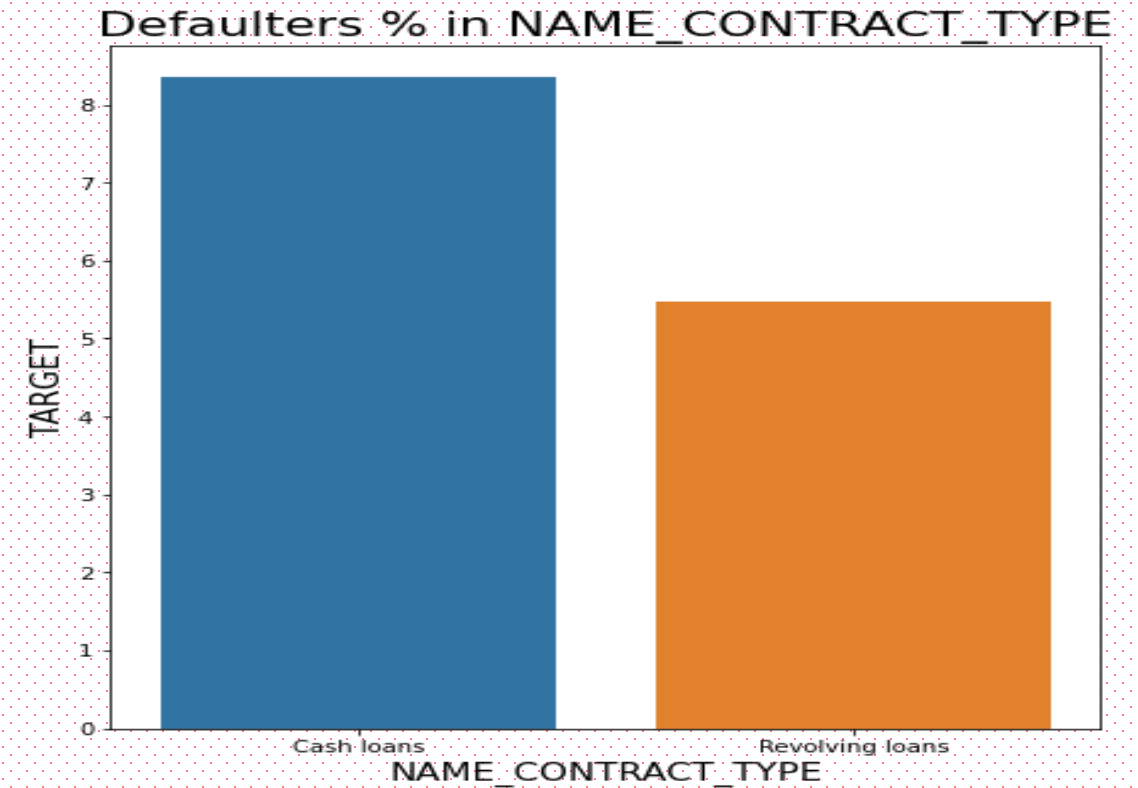
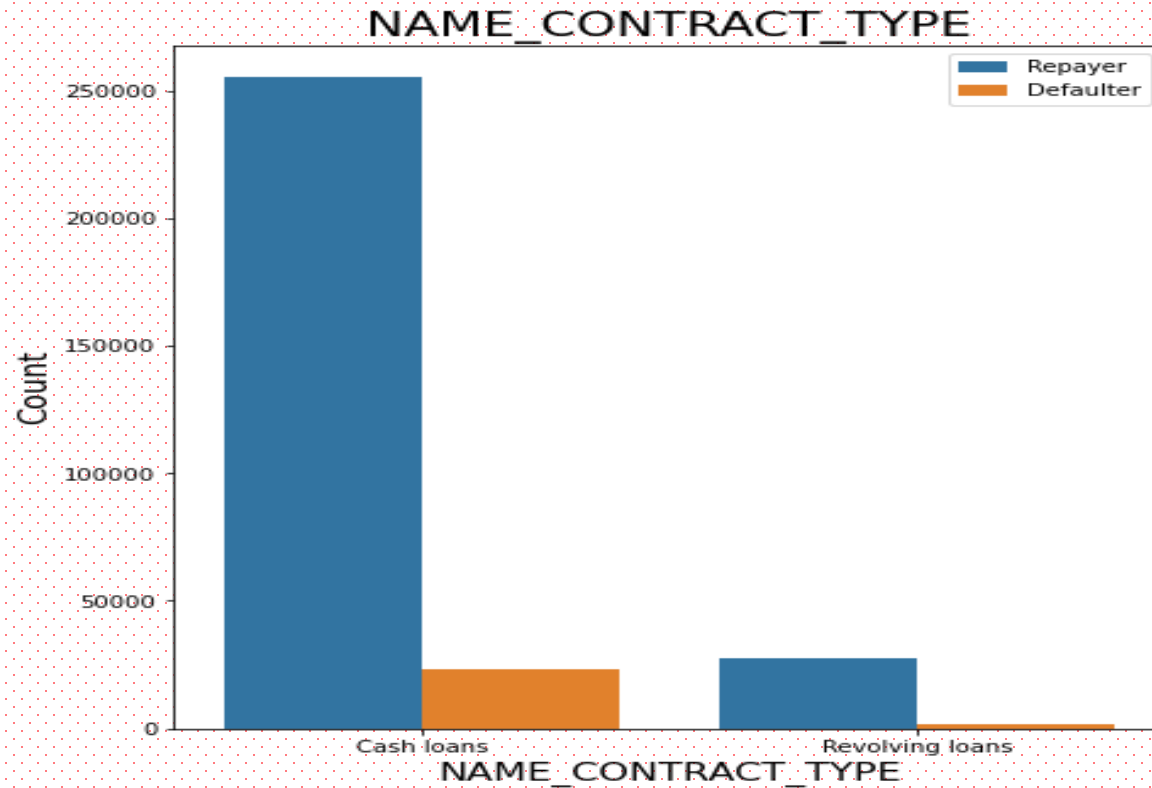
Imbalance Plotting (Repayer Vs Defaulter)



contract type based on loan repayment status

Revolving loans are just a small fraction (10%) from the total number of loans

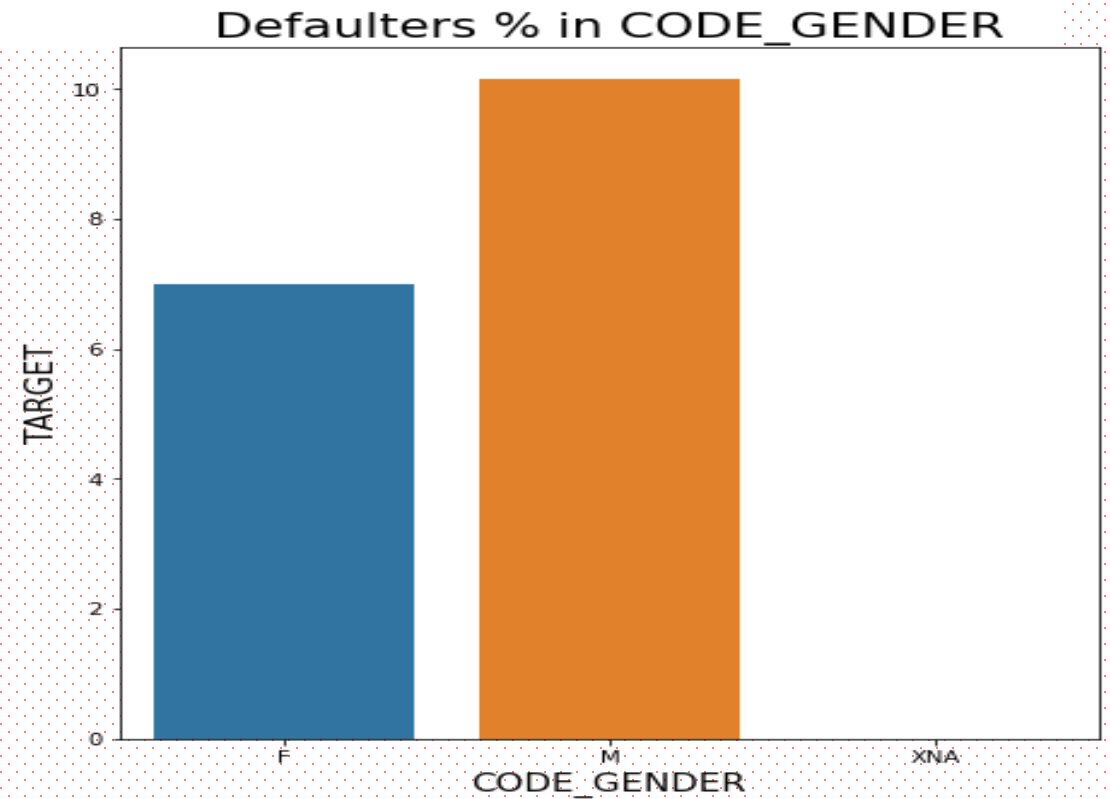
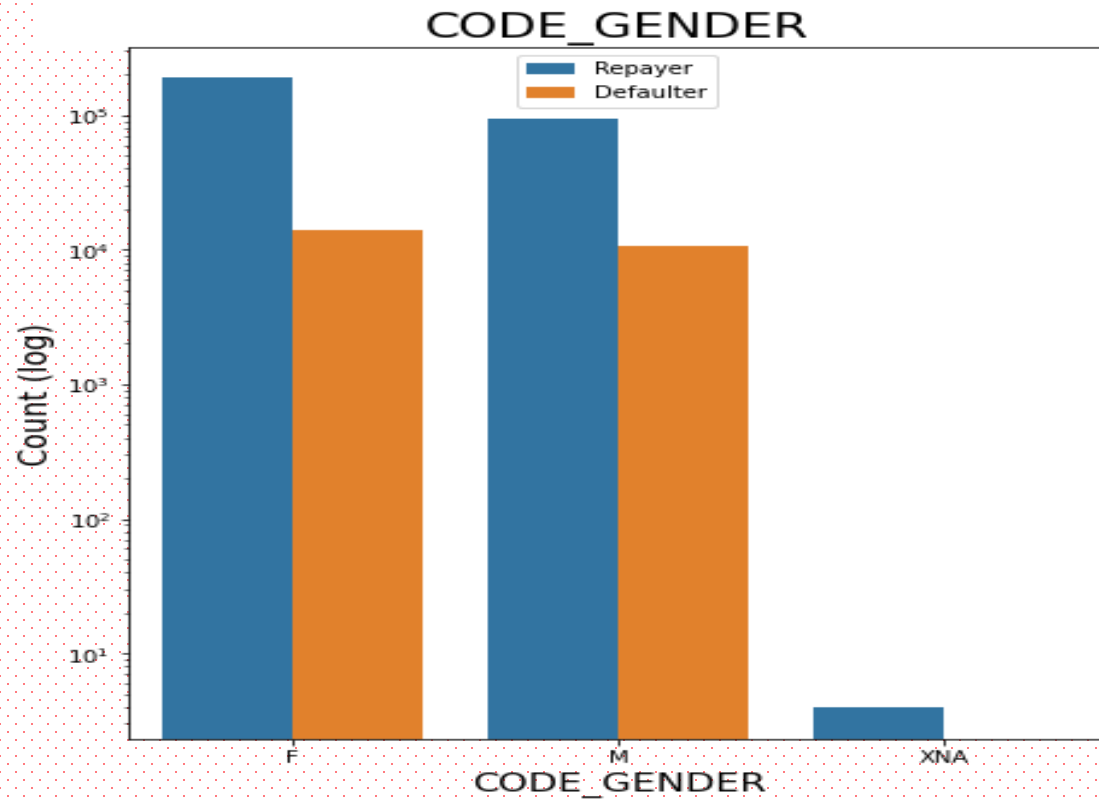
Around 8-9% Cash loan applicants and 5-6% Revolving loan applicant are in defaulters



type of Gender on loan repayment status

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 about 10%, comparing with women about 7%



CONCLUSION

After analyzing the datasets, there are few attributes of a client with which the bank would be able to identify if they will repay the loan or not. The analysis is consisted with the contributing factors and categorization.

A. Decisive Factor whether an applicant will be Repayer:

- REGION_RATING_CLIENT: RATING 1 is safer
- ORGANIZATION_TYPE: Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%
- DAYS_BIRTH: People above age of 50 have low probability of defaulting
- DAYS_EMPLOYED: Clients with 40+ year experience having less than 1% default rate
- NAME_INCOME_TYPE: Student and Businessmen have no defaults
- AMT_INCOME_TOTAL: Applicant with Income more than 700,000 are less likely to default
- NAME_CASH_LOAN_PURPOSE: Loans bought for Hobby, Buying garage are being repaid mostly
- CNT_CHILDREN: People with zero to two children tend to repay the loans.
- NAME_EDUCATION_TYPE: Academic degree has less defaults.

B. Decisive Factor whether an applicant will be Defaulter:

1. CODE_GENDER: Men are at relatively higher default rate
2. NAME_FAMILY_STATUS : People who have civil marriage or who are single default a lot.
3. NAME_EDUCATION_TYPE: People with Lower Secondary & Secondary education
4. NAME_INCOME_TYPE: Clients who are either at Maternity leave OR Unemployed default a lot.
5. REGION_RATING_CLIENT: People who live in Rating 3 has highest defaults.
6. OCCUPATION_TYPE: Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.
7. 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.
8. DAYS_BIRTH: Avoid young people who are in age group of 20-40 as they have higher probability 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 3lakhs, there is an increase in defaulters.

C. Factors that Loan can be given on Condition of High Interest rate to mitigate any default risk leading to 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 3-6 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.
3. **AMT_INCOME**: Since 90% of the applications have Income total less than 3Lakhs and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.
4. **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.
5. **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.

RECOMMENDATION / SUGGESTIONS / PROPSALS :

- 90% of the previously cancelled client have actually repaid the loan. Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customers in future for increase business opportunity.
- 88% of the clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.

THANKYOU

