

Assignment: Credit EDA Case Study

BY : TUSHAR GUPTA

Problem Statement

Whenever the company receives a loan application, it needs to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- I. If the applicant is likely to repay the loan, then not approving the loan would result in a loss of business to the company.
- II. If the applicant is not likely to repay the loan, which means he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Business Objectives

- I. This case study helps to identify various patterns which indicate whether 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 applicants capable of repaying the loan are not rejected. Identification of such applications using EDA is the aim of this case study.
- II. In other words, the company wants to understand the driving factors (or driver variables) behind loan default, which means the variables which are strong indicators of default. The company can use this knowledge for its portfolio and risk assessment.

Analysis Steps and Approach

Understanding Problem Statement

Exploring the data

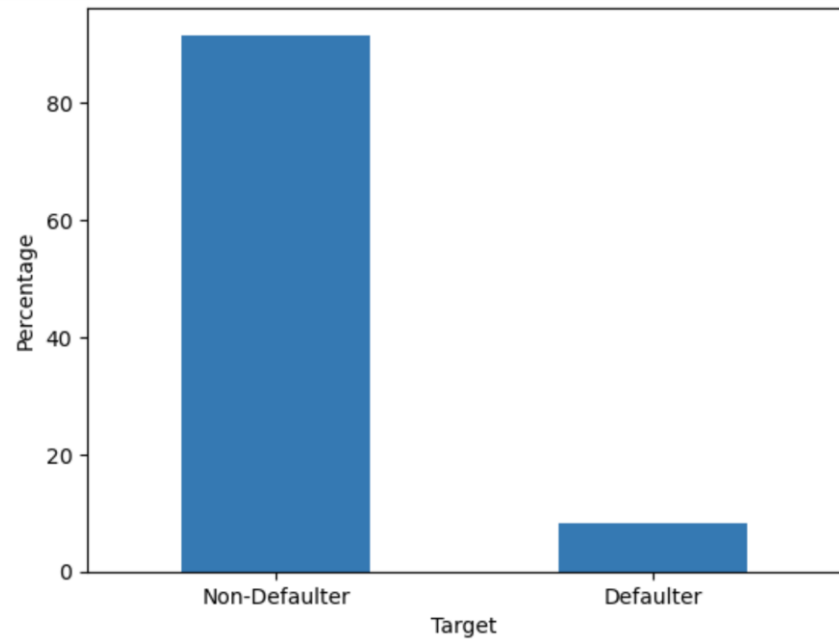
Data Cleaning

Outlier Analysis

Data Analysis



Data Imbalance Ratio



► The Data imbalance Ratio Between Non-Defaulters and Defaulters In target Variable is 91.626 and 8.374

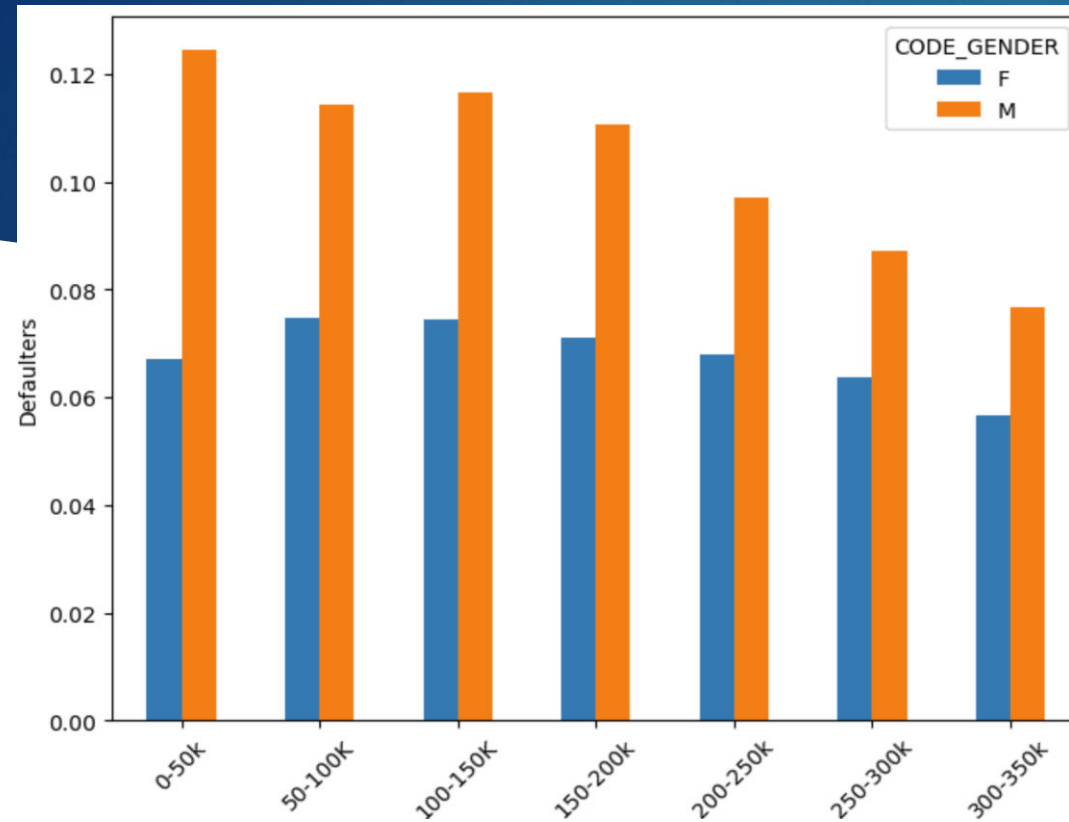
► Making a ratio of 10.94



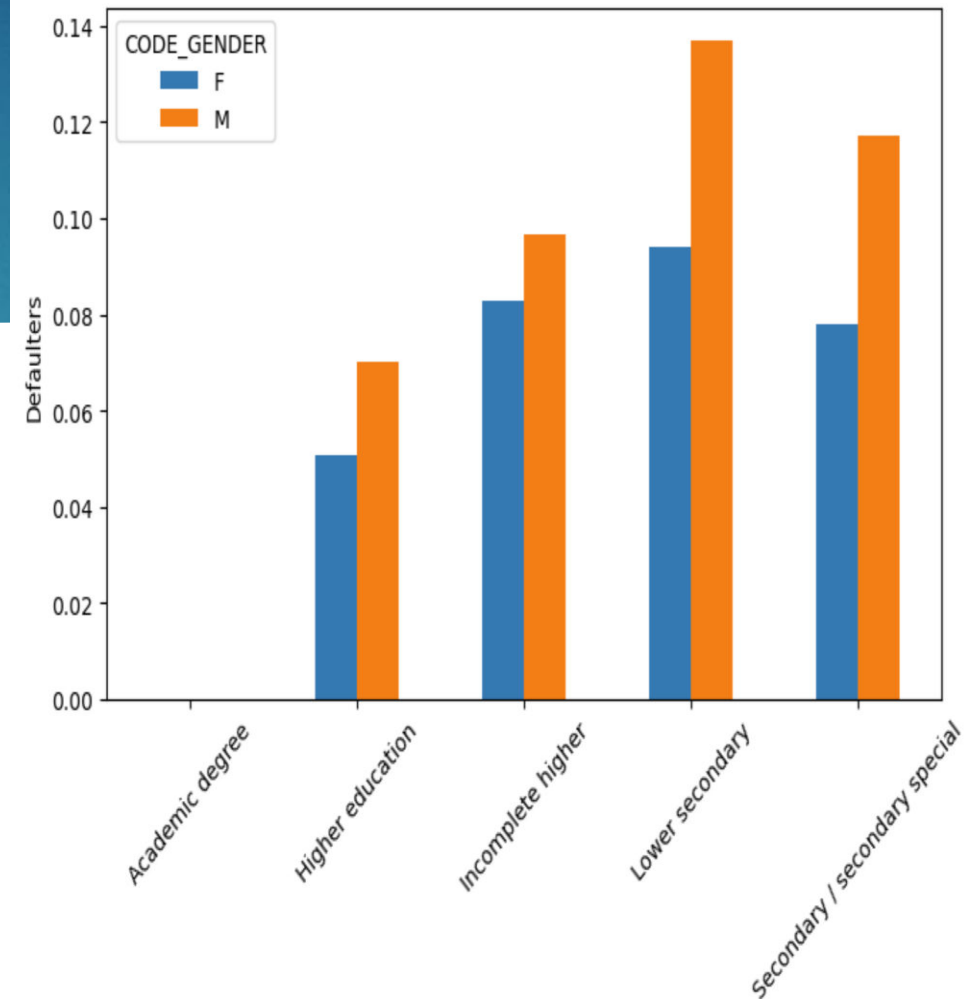
Current Application

ANALYSIS OF SEGMENTED VARIABLES(TWO AT A TIME) FOR
DEFAULTERS

Income group and Gender

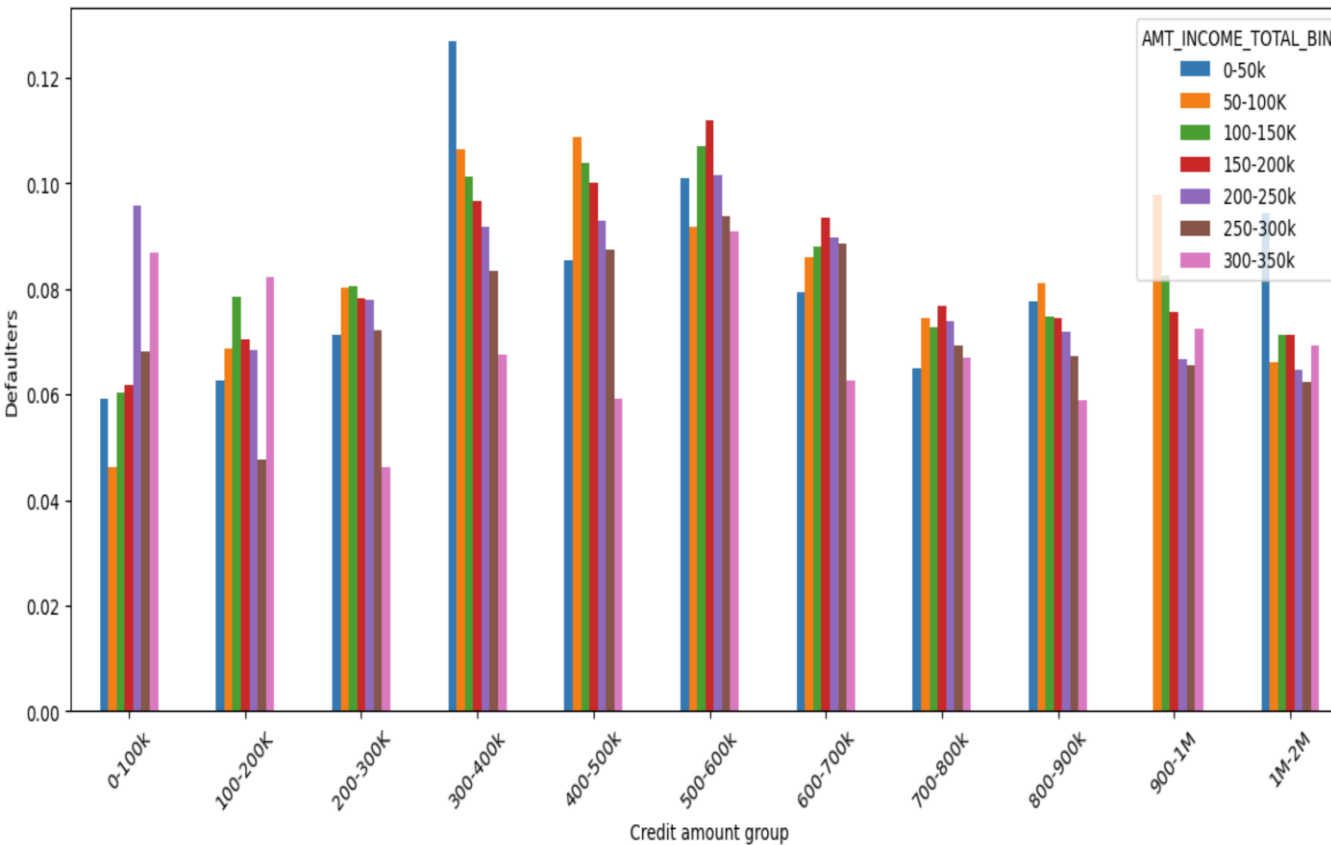


Education and Gender

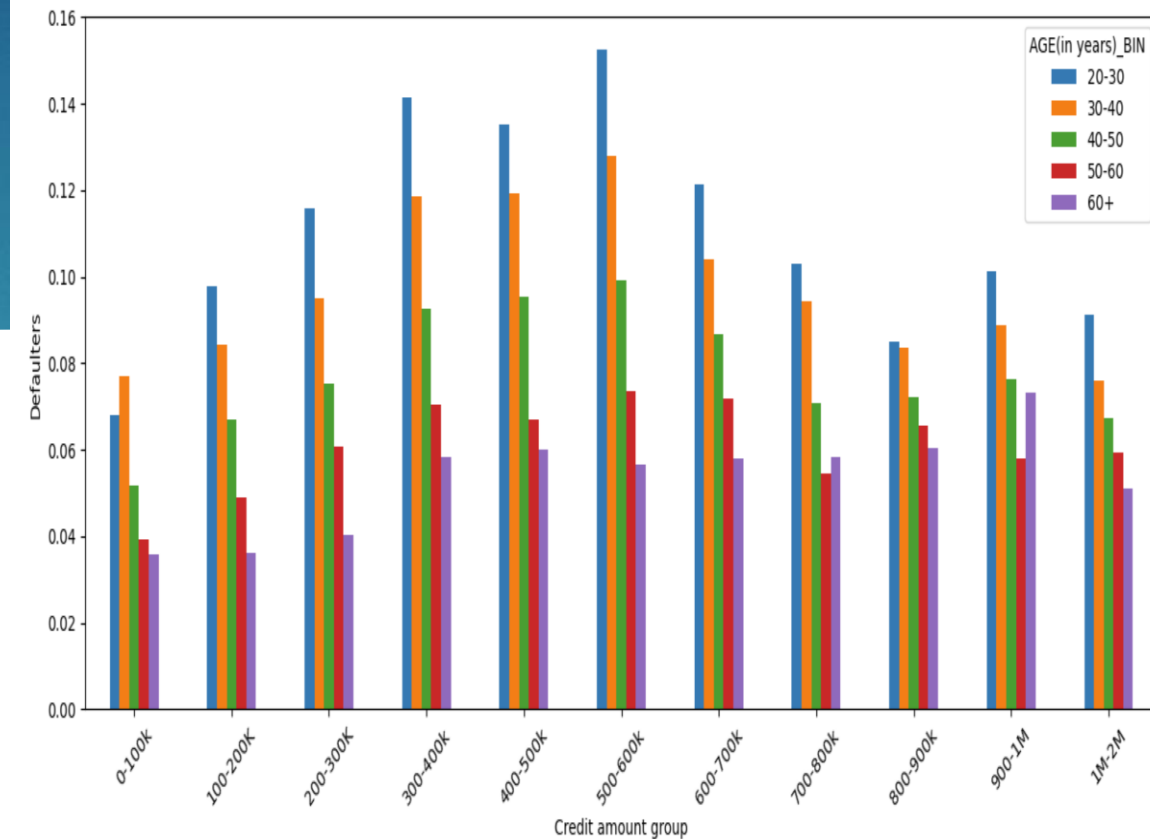


1. Males are more likely the defaulters than Females across all income groups.
2. Lower secondary educated clients are the most as defaulters followed by Secondary and Incomplete higher educated clients while the Higher education clients are among the least defaulters.
3. Across all educated levels, Males are more likely the defaulters than Females.

Credit Amount group and Income group

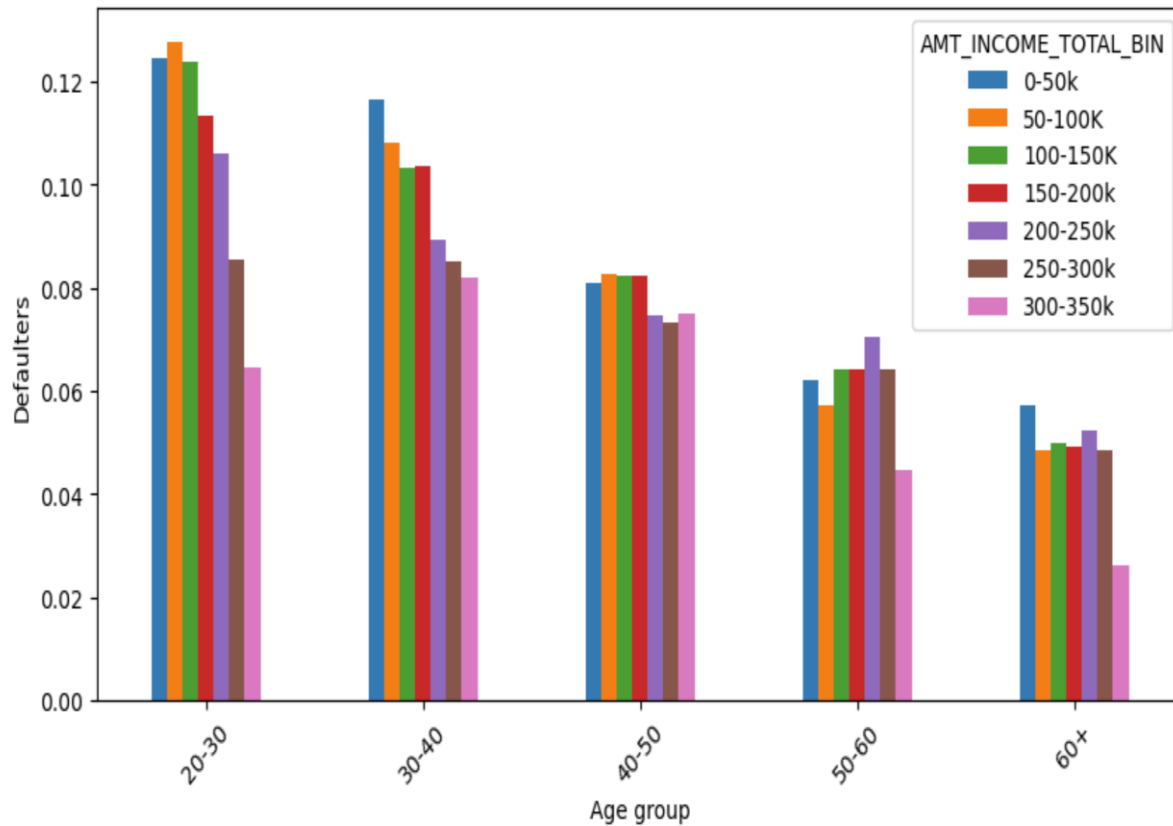


Credit Amount group and Age group

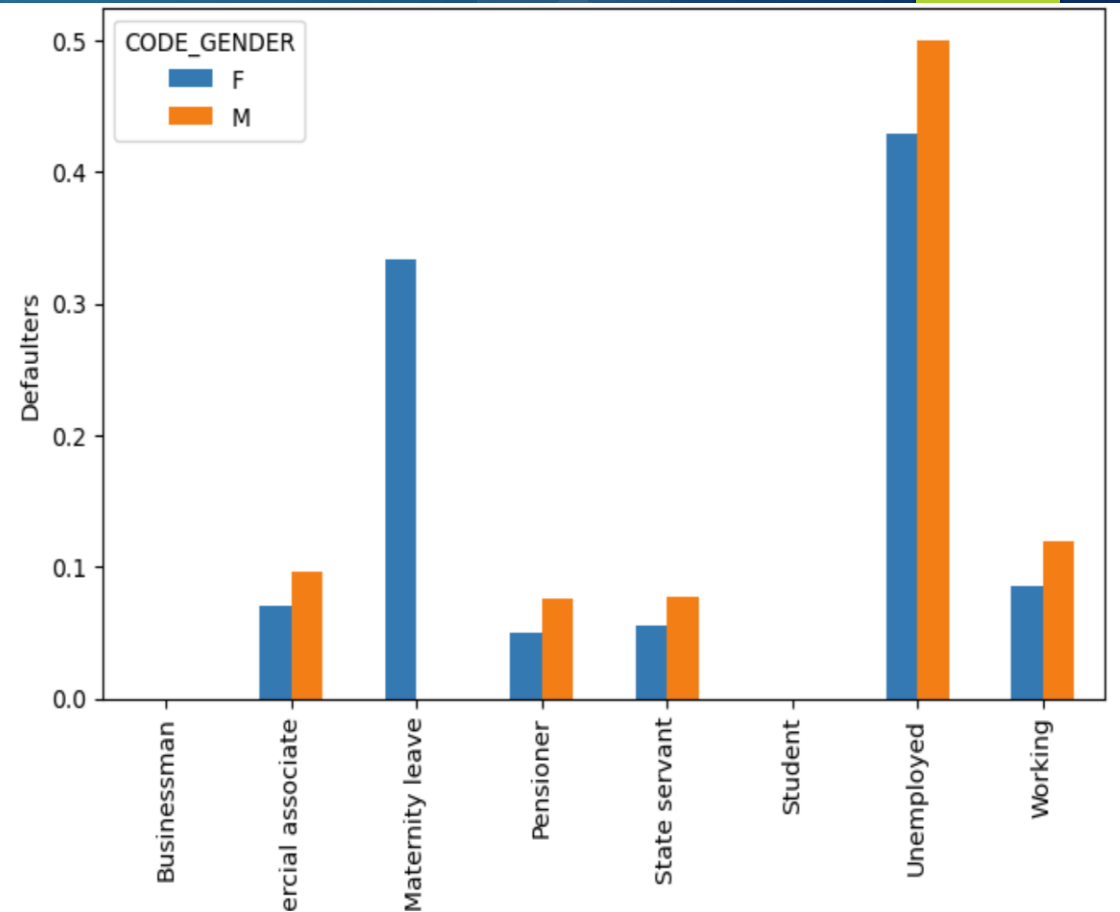


1. 100-250k credit amount groups have the highest defaulters amongst all income groups.
2. 300-350k credit amount group has the least defaulters amongst most income groups.
3. 20-40 yrs age groups with 200-700k credit amount groups are having most defaulters.
4. 60+ age group across all credit amount groups have the least defaulters.

Age group and Income group

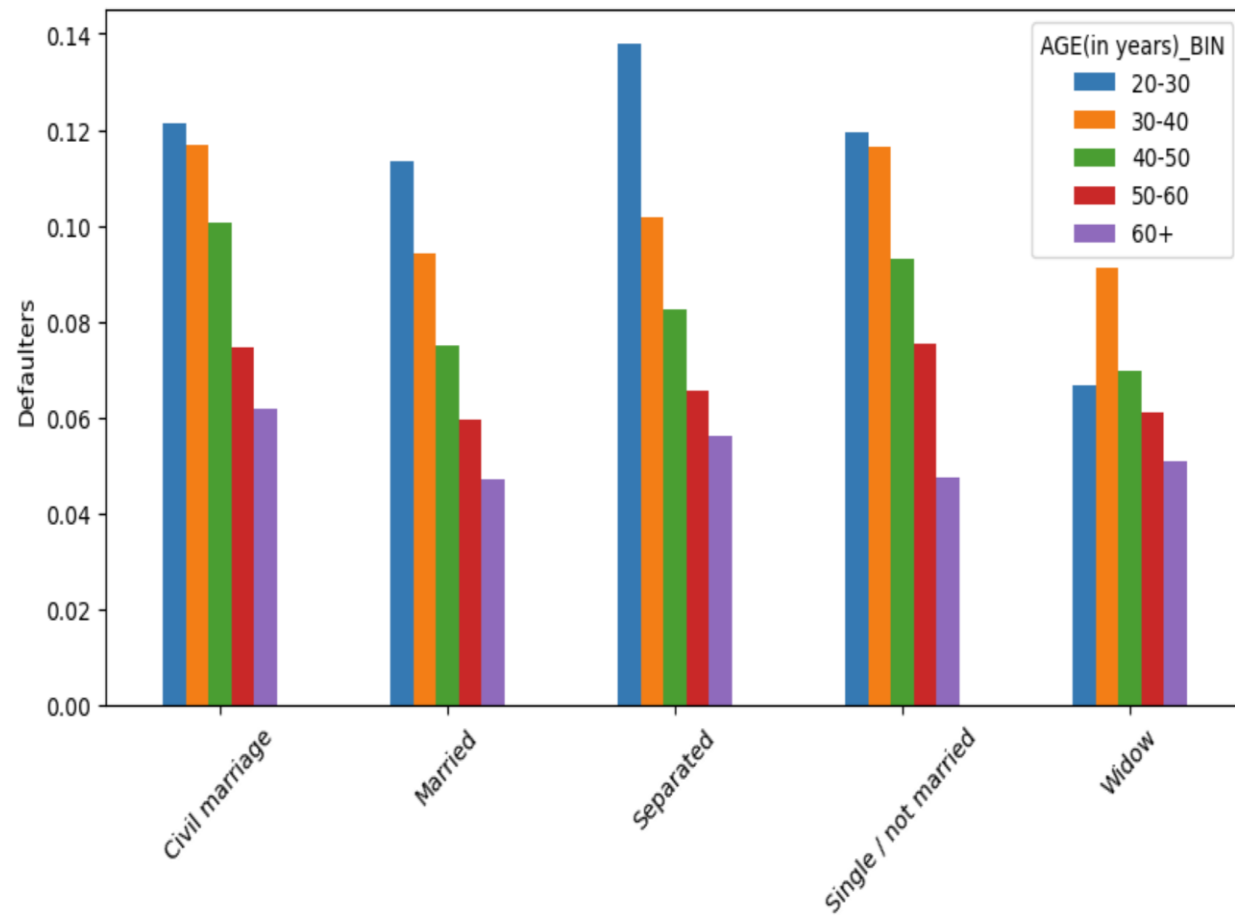


Profession and Gender

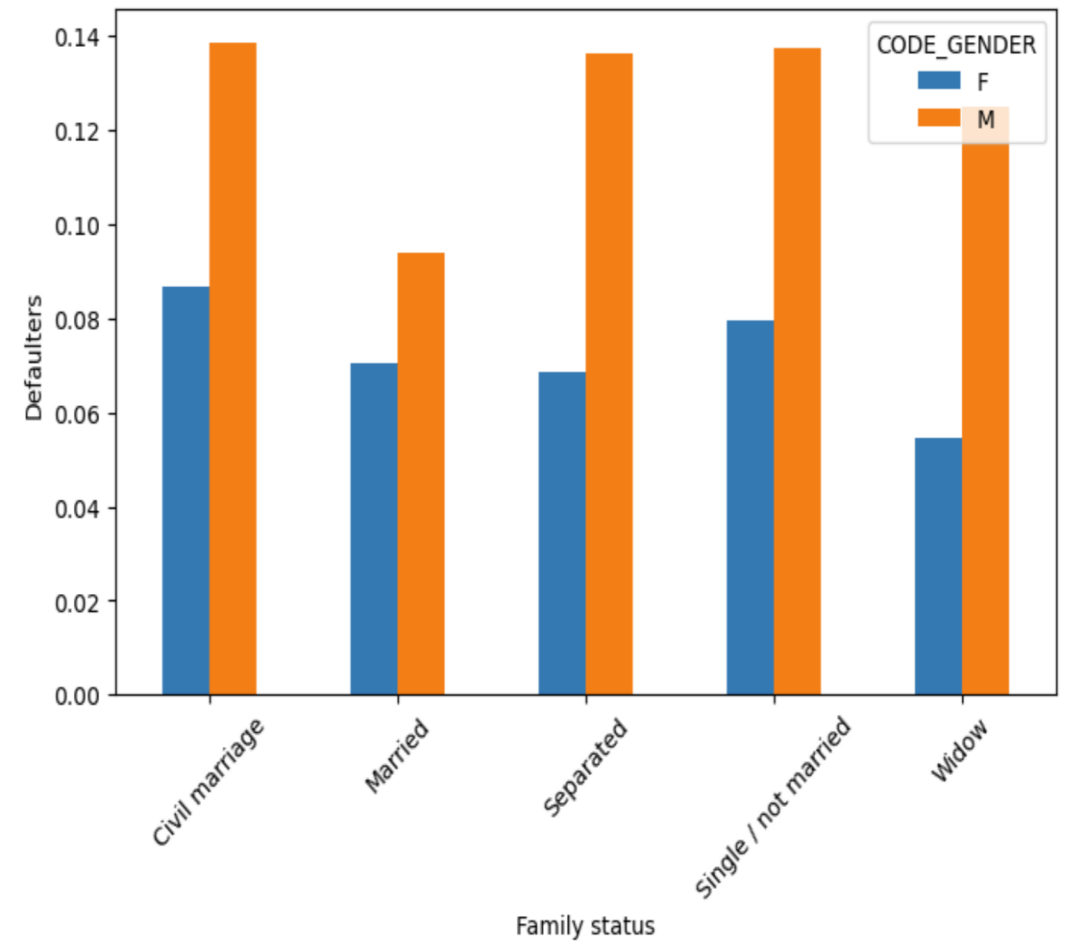


1. 20-40 yrs age group of clients are having more defaulters than other age groups and amongst those, 0-200k income people are most of the defaulters.
2. For 40-60+ age groups, the rate of defaulters is almost the same in all income groups.
3. Unemployed clients and the clients with Maternity leave are the most defaulted groups among all professions but the default rate is quite lesser in all other professions.
4. Males are having more defaulters than Females in all professions.

Family status and Age group



Family status and Gender

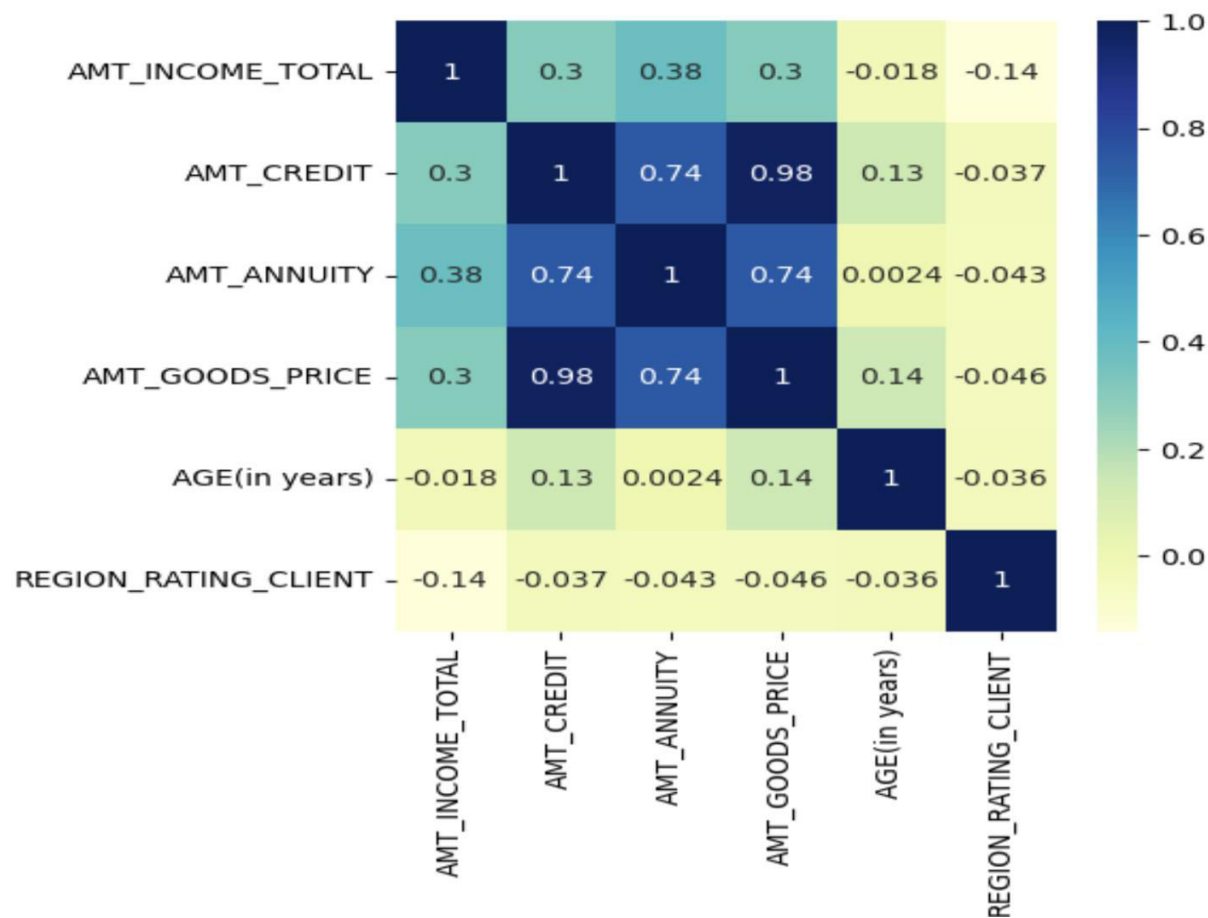


1. 20-40 years age groups have the most defaulters and 60+ age group has the least defaulters across all the family statuses.
2. Male clients are having more defaulters than Female clients across all family statuses.

Current Application

BIVARIATE ANALYSIS

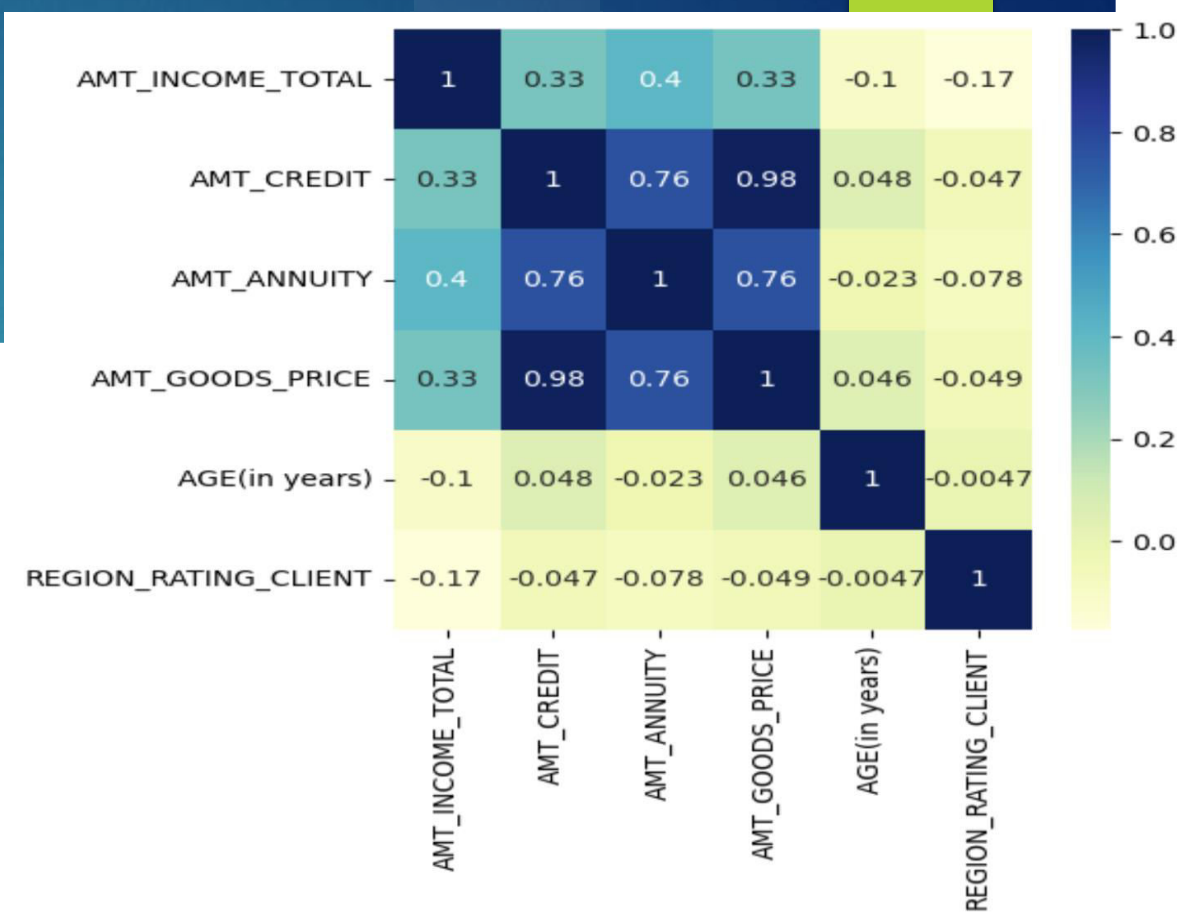
Defaulters



Highly correlated columns for defaulters

1. AMT_CREDIT and AMT_ANNUITY (0.74)
2. AMT_CREDIT and AMT_GOODS_PRICE (0.98)
3. AMT_ANNUITY and AMT_GOODS_PRICE (0.74)

Non-Defaulters



Highly correlated columns for non-defaulters

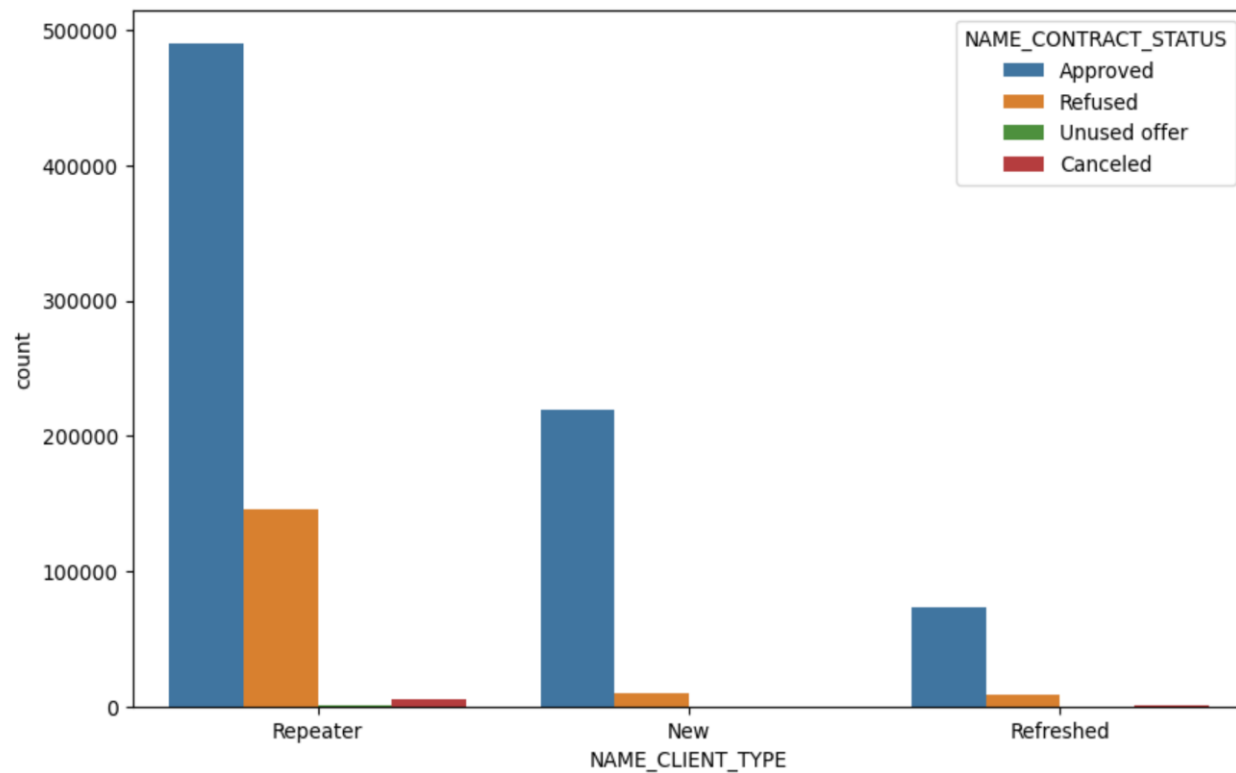
1. AMT_CREDIT and AMT_ANNUITY (0.76)
2. AMT_CREDIT and AMT_GOODS_PRICE (0.98)
3. AMT_ANNUITY and AMT_GOODS_PRICE (0.76)



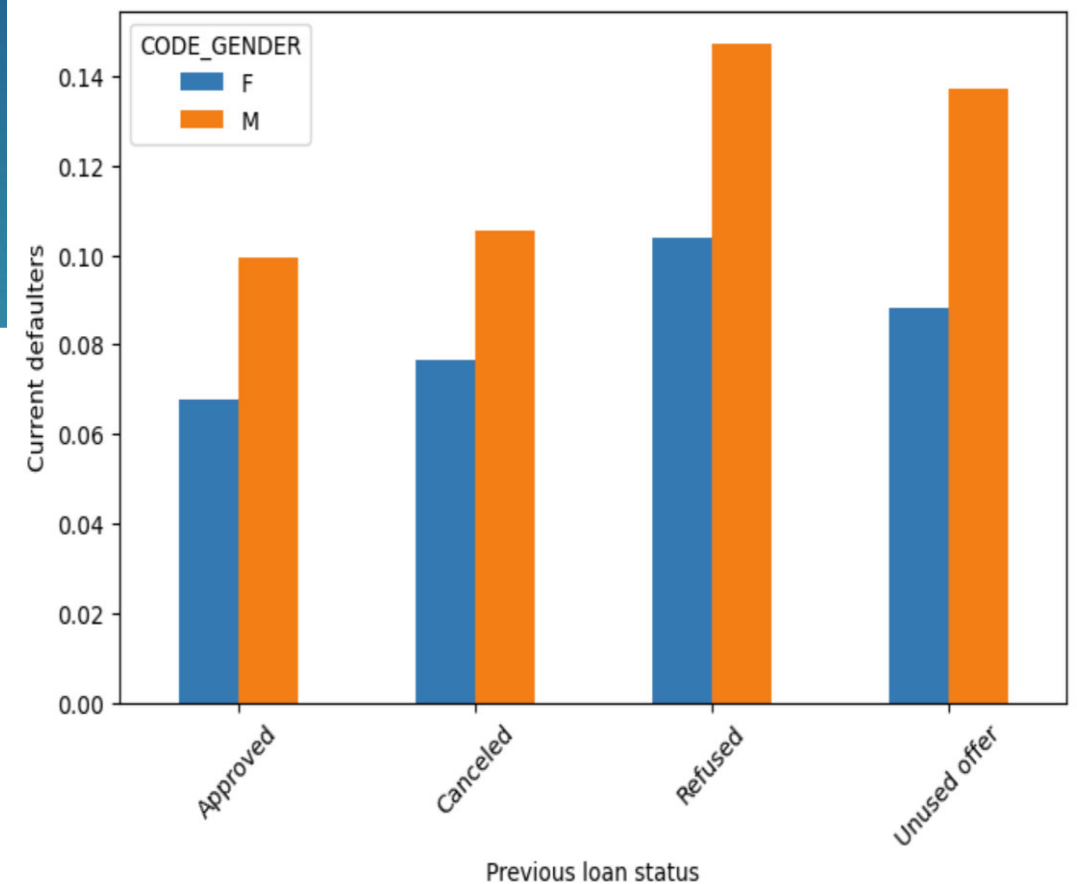
Current and Previous Application

ANALYSIS OF SEGMENTED VARIABLES(TWO AT A TIME)

Status and Client type

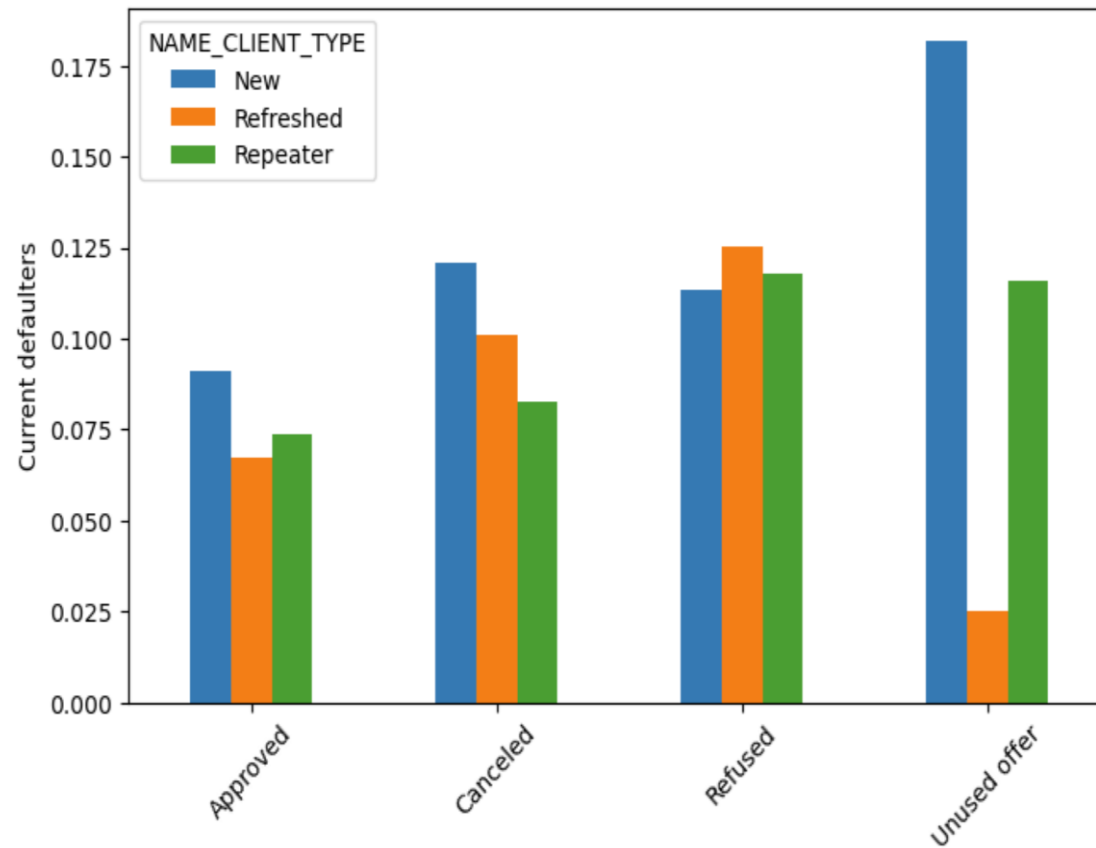


Previous Loan status and Gender

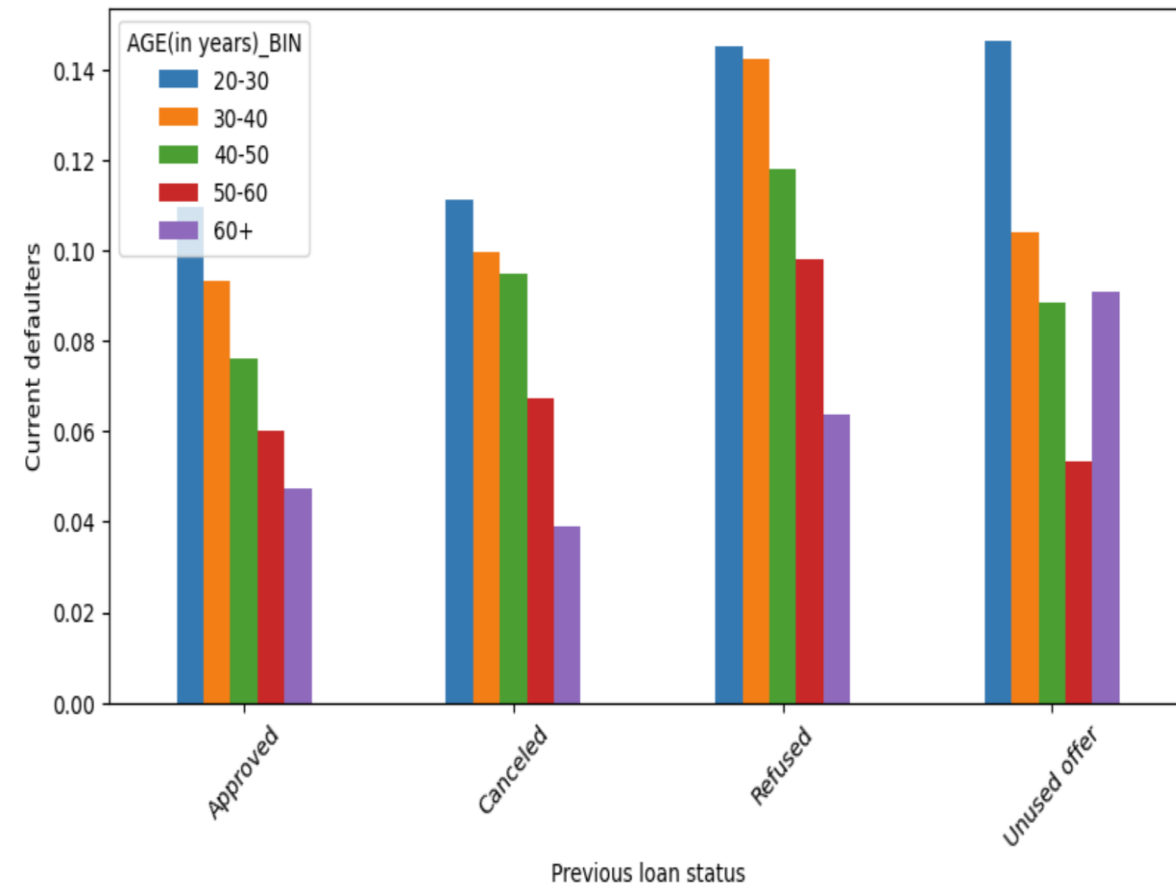


1. Repeater clients have more approved loans than New and Refreshed clients. Same is valid for refused loans as well.
2. Previously Refused clients are more defaulted than previously Approved clients. Also, in all loan statuses, the Males are more defaulted than Females.

Previous loan status and Client type

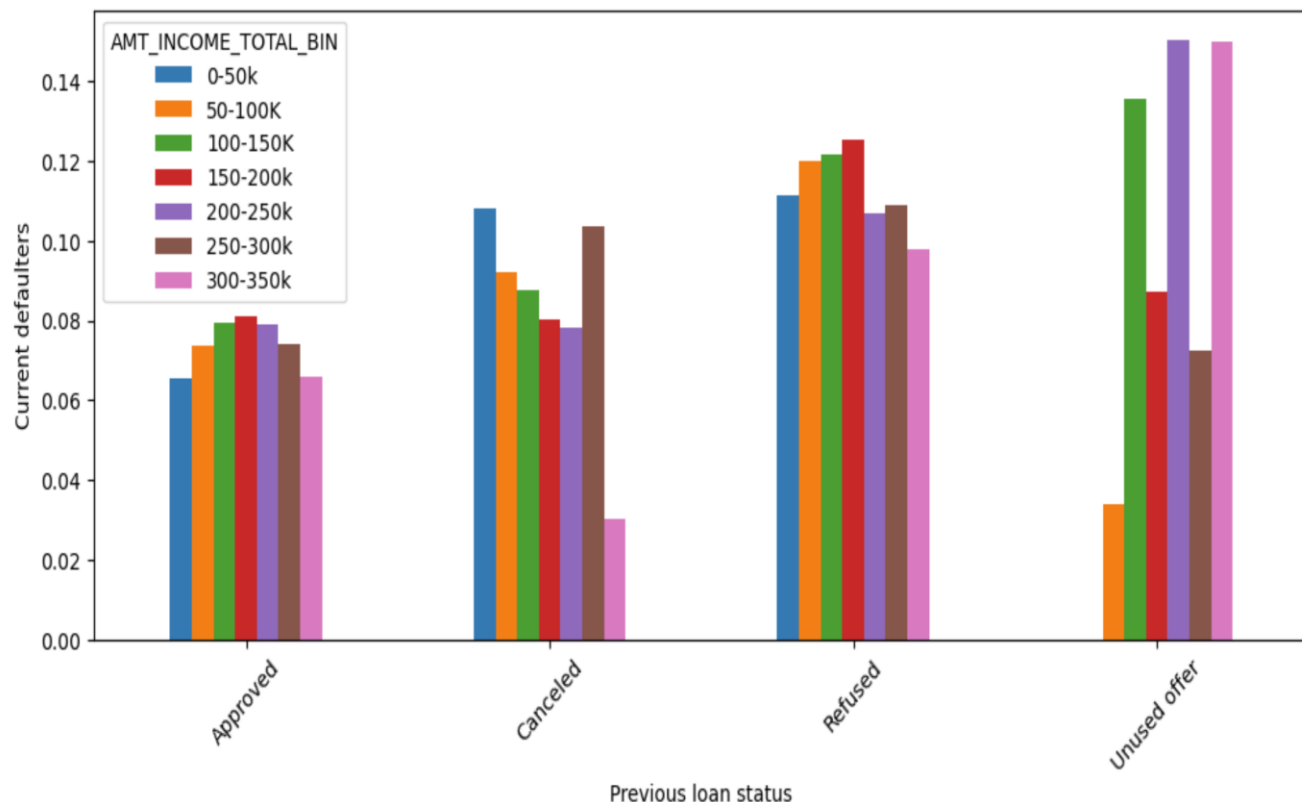


Previous loan status and Age group

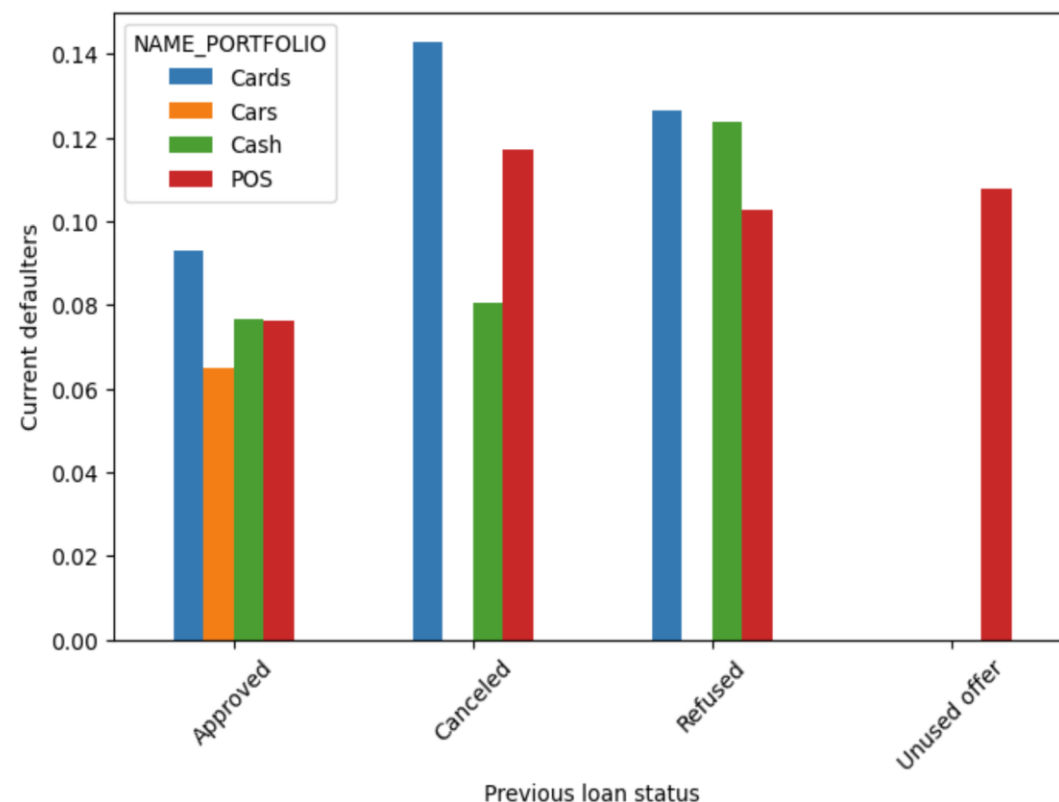


1. Defaulters are more for previously Unused offers loan status clients, who were New clients as well.
2. For previously Approved loan status, the New clients were more defaulted followed by Repeater clients.
3. For previously Refused loan status, the Refreshed clients were more defaulted followed by Repeater clients.
4. For previously Canceled loan status, the New clients were more defaulted followed by Refreshed clients.
5. For all the Previous loan statuses, 20-30 yrs age group applicants are the most defaulted among all other age groups.
6. For all the Previous loan statuses, 60+ yrs age group applicants are the least defaulted among all other age groups.

Previous loan Status and Income group



Previous loan status and Portfolio

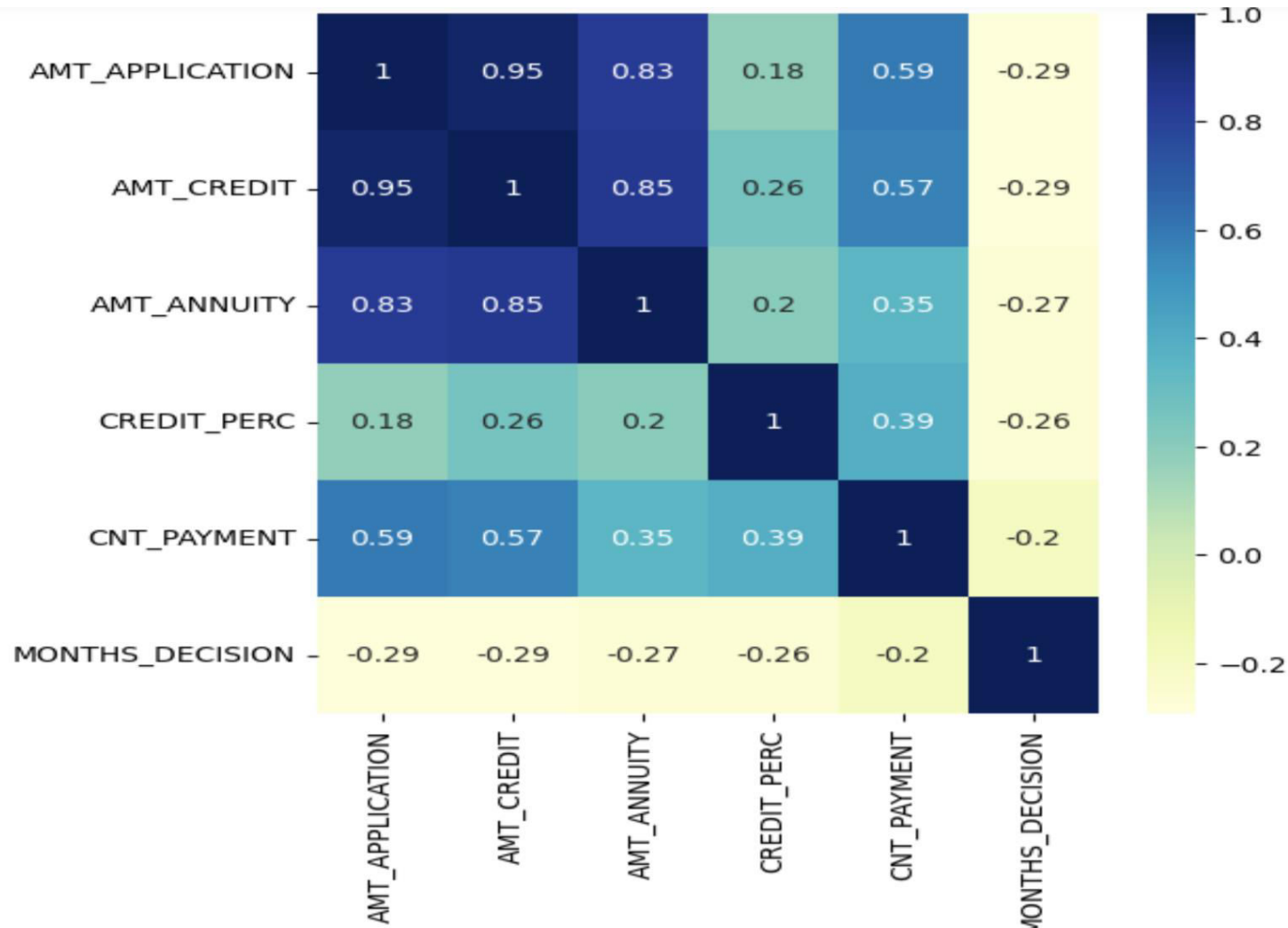


1. For previously Unused offer loan status, the 100-350k income groups are most defaulted and 50-100k income group is the least defaulted.
2. For all other previous loan statuses, all income groups are almost equally defaulted.
3. Clients who previously applied loan for Cards are most defaulted in all previous loan statuses.
4. Clients who previously applied loan for Cars are least defaulted for Approved loan status.
5. Clients who previously applied loan for Cash are least defaulted for Canceled loan status.
6. Clients who previously applied loan for POS are least defaulted for Refused loan status.



Current and Previous Application

BIVARIATE ANALYSIS



Highly correlated columns

1. AMT_APPLICATION and AMT_CREDIT (0.95)
2. AMT_APPLICATION and AMT_ANNUITY (0.83)
3. AMT_CREDIT and AMT_ANNUITY (0.85)

Moderately correlated columns

1. AMT_APPLICATION and CNT_PAYMENT (0.59)
2. AMT_CREDIT and CNT_PAYMENT (0.57)

Recommendations

Current Applications

- ▶ Safer to grant loans for 40 to 60+ years age group with higher income.
- ▶ Risky to grant loans with less than 40 years age group with lower income groups.
- ▶ Best to grant loans to 60+ years age group of all family statuses.
- ▶ Risky to grant loan for Single, Separated and Civil marriage men with less than 40 years of age.
- ▶ Grant slightly higher amount of loan to all age groups
- ▶ Risky to grant medium and low loan amounts to clients with <40 years of age.
- ▶ Best to grant loans to higher educated clients across all professions except when they are unemployed(or women with maternity).

Current and Previous applications

- ▶ Recommended to provide loans to previously approved females.
- ▶ Risk to provide loans to clients whose applications were refused or previously unused.
- ▶ Safer to grant loans for 60+ years age group.
- ▶ Least risk to grant loans for approved clients of all income groups.
- ▶ Safer to provide loans for any portfolio for previously approved applicants.

Final Conclusion

Top Recommended groups

- ▶ Approved clients in their previous applications.
- ▶ Highly educated clients with higher income.
- ▶ 60+ years age group in all categories.
- ▶ Married clients when compared to other family statuses.
- ▶ Females are favorable than males comparatively in all categories.

Top Risk groups

- ▶ Previously refused, canceled or unused offer clients.
- ▶ Lower income groups with previously refused status.
- ▶ Unemployed clients.
- ▶ Below 30 years age group is riskier than other age groups comparatively.
- ▶ Lower secondary and secondary educated clients.



Thank You!!!