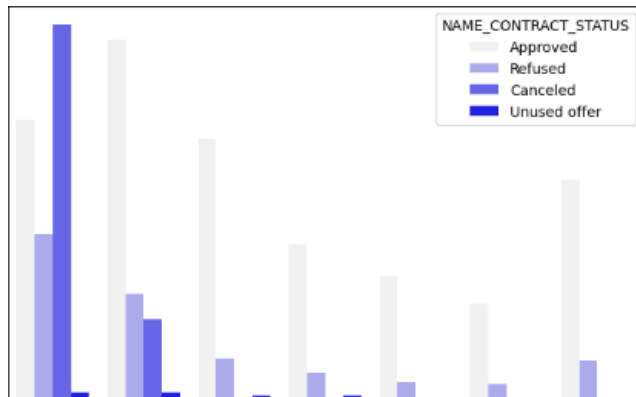


Credit EDA Assignment

Amos Jaimes



Problem Statement

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.

Approach & Methodology Used

- ◆ Understand Data
- ◆ Clean Data
- ◆ Handle outliers If any
- ◆ Univariate
- ◆ Bivariate and Multi Variate Analysis
- ◆ Recommendations

Data Cleaning

Dropped column

The columns which has more than 40% of the missing values are dropped

Standardization

There are some negative values in both datasets, so converted them to absolute values using `abs()`

Imputing

Used mean, median and mode data to impute missing values for some columns.

Missing Value Imputation/Dropping

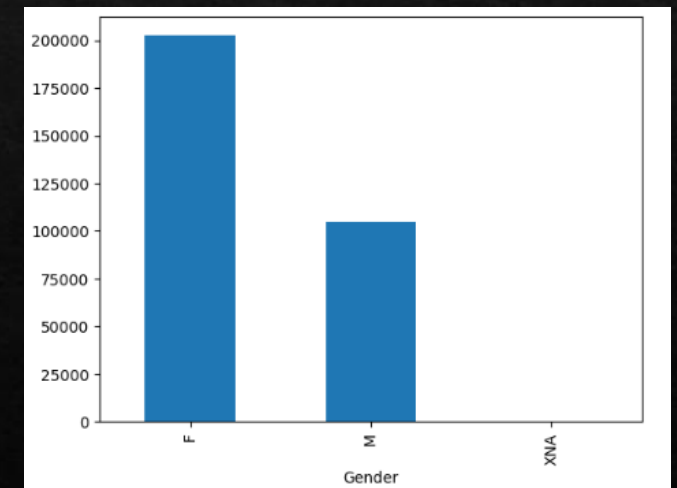
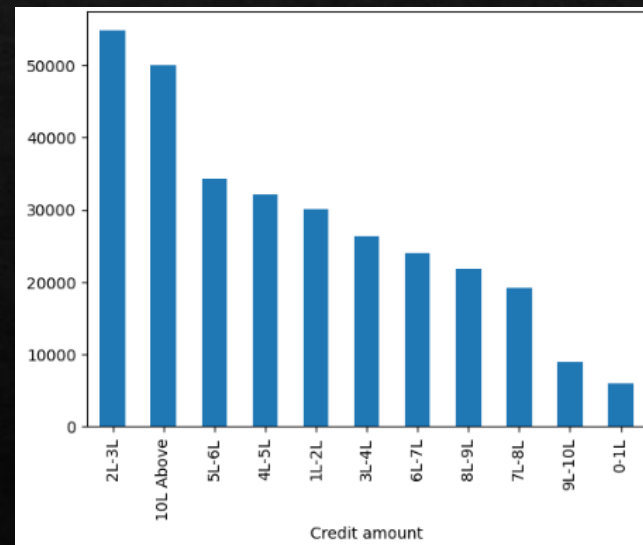
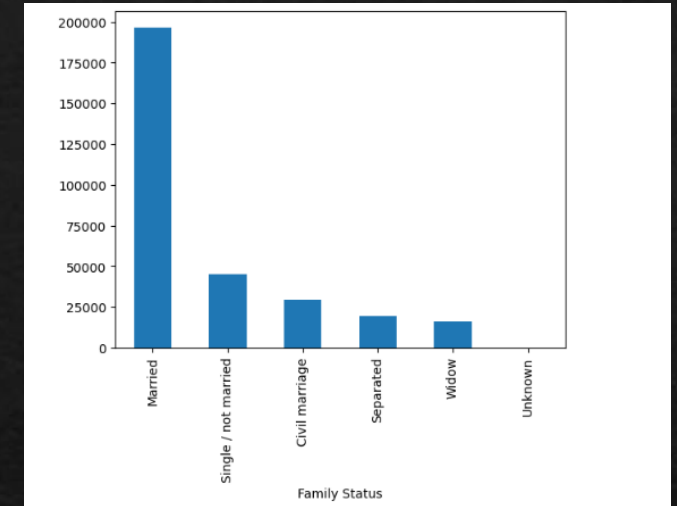
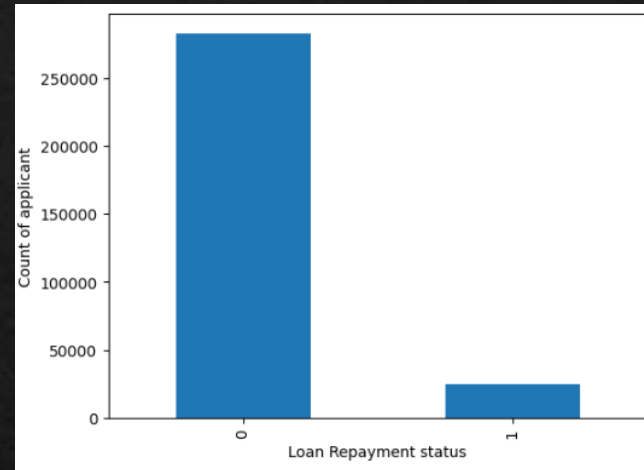
Kept 46 columns in `application_data`

and dropped the rest which had more than 40 percent missing values and

kept 22 columns in `previous_application_data` by following the latter criteria

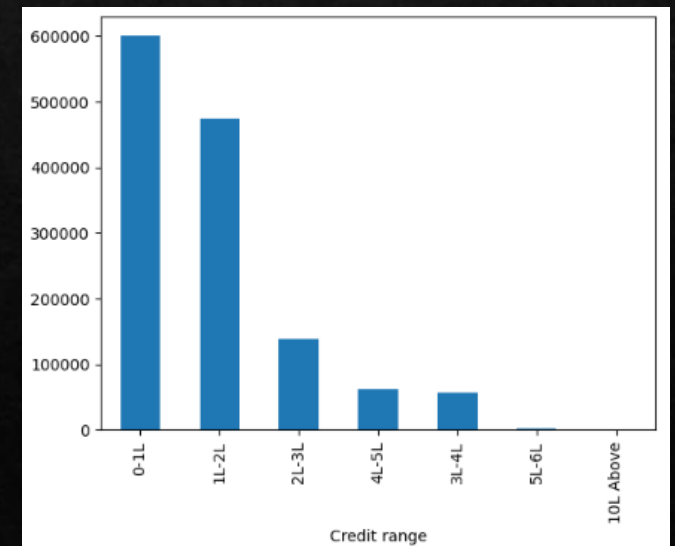
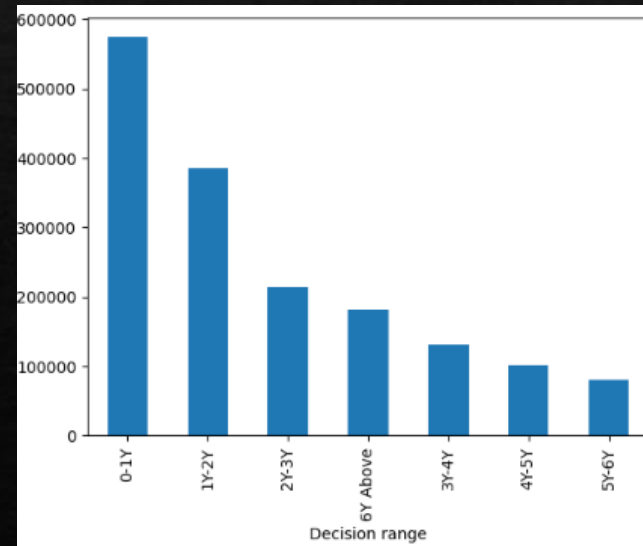
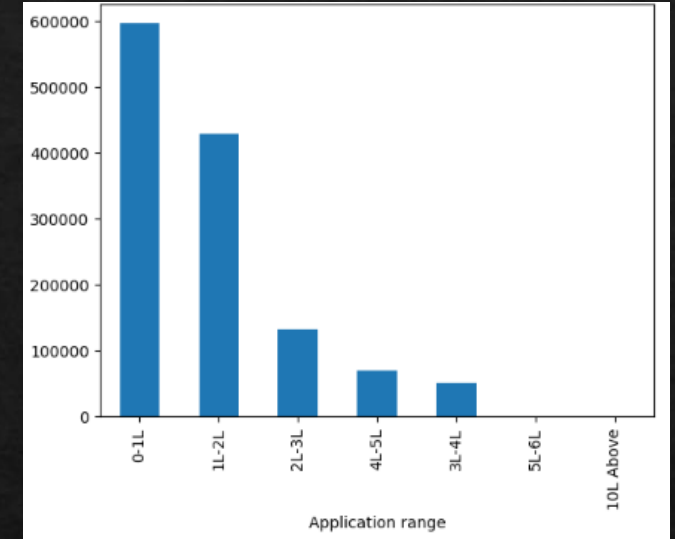
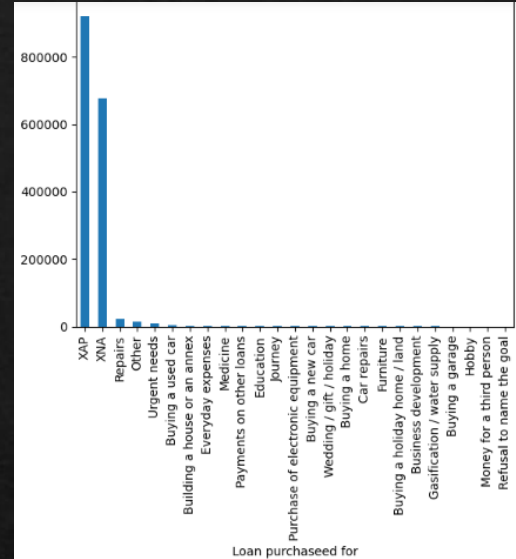
Univariate Analysis for Application data and it's Inference

- ◇ Graph 1: Most of the applicant are paying the loan amount on time
- ◇ Graph 2: Most loans are taken are by married peoples
- ◇ Graph 3: Most of the applicant whose loans credit are between 2L-3L followed by 10L and above
- ◇ Graph 4: Most of the loan applicant are females



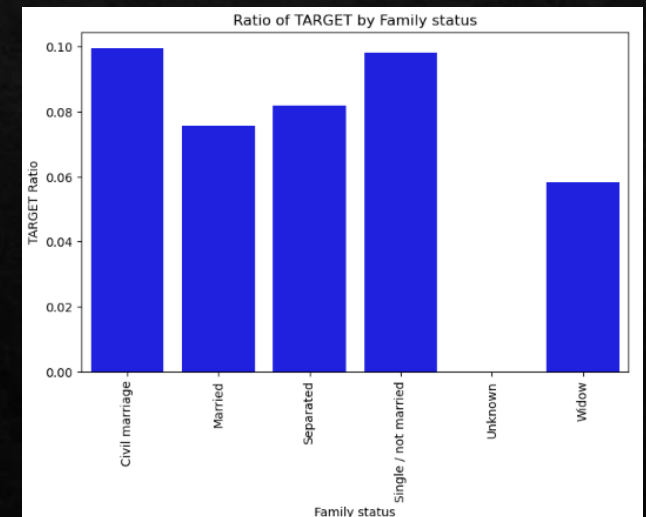
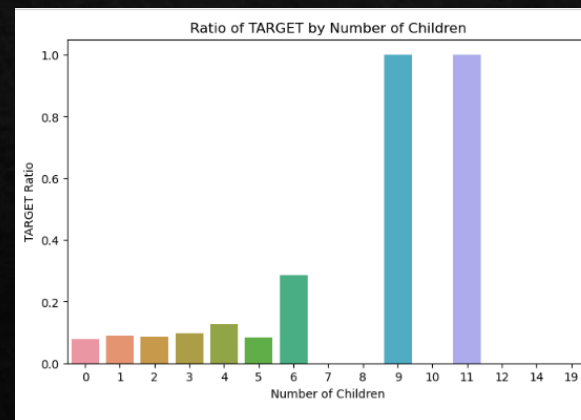
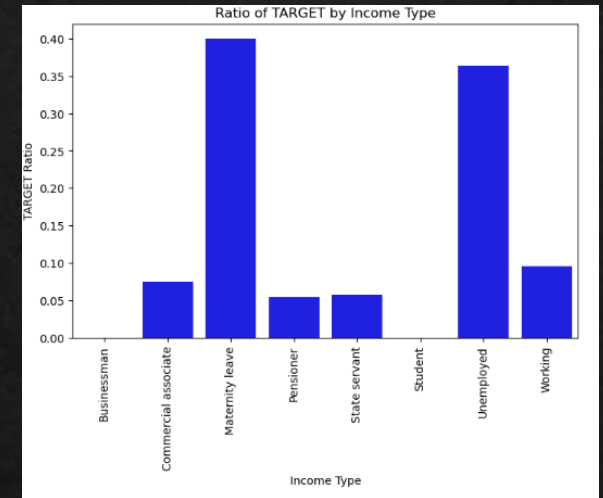
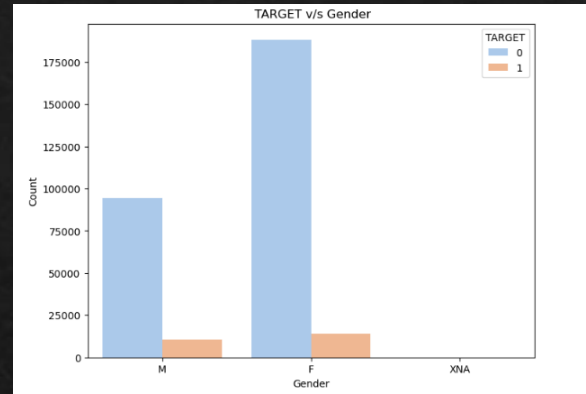
Univariate Analysis for Previous data and it's Inference

- ❖ Graph 1: Most of the client had not informed about the purpose of the loan amount
- ❖ Graph 2: Most of the applicant are asked the credit amount around 0-1L
- ❖ Graph 3: Most client took around 0-1 years to decide whether to take loan or not
- ❖ Graph 4: Most of the applicant got the credit in a range on 0-1L



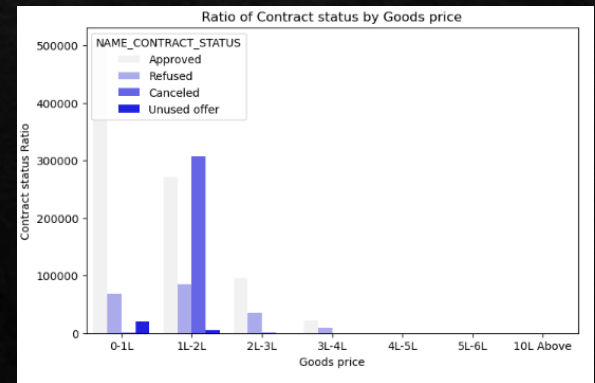
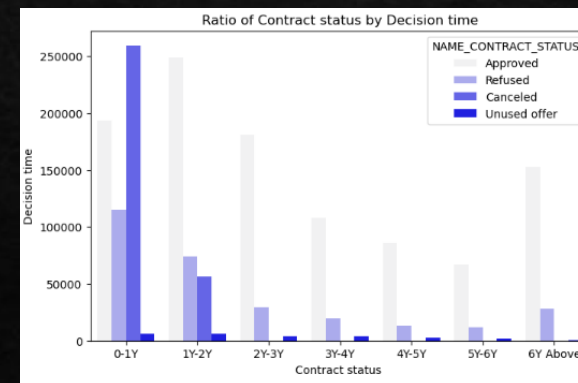
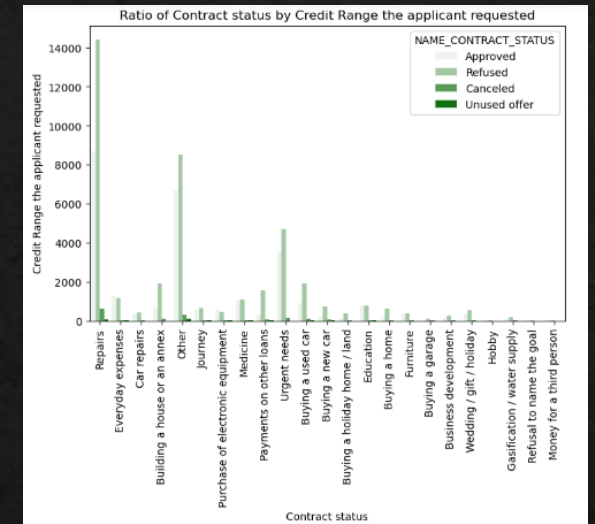
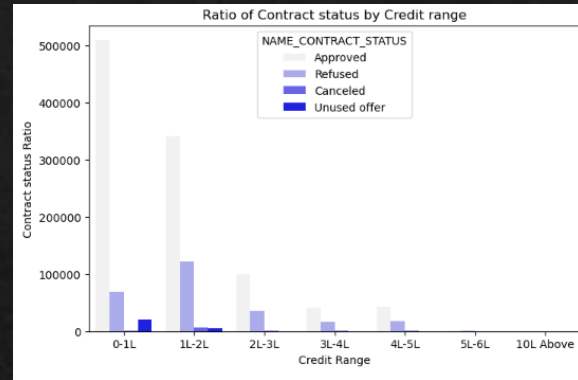
Bi Analysis for Application data and it's Inference

- ◇ Graph 1: The most female client are paying the loan on time
- ◇ Graph 2: The client who are in maternity leave and unemployed are the most defaulters while working professional pay the loans on time
- ◇ Graph 3: The applicant with no children are paying the loan on time # and the most defaulters are the applicant who has children 9 and 11
- ◇ Graph 4: Singles\ not married and Civil marriage category applicant have more defaulters while other are of not that much Married client are paying the loans on time



Bi Analysis for Previous data and it's Inference

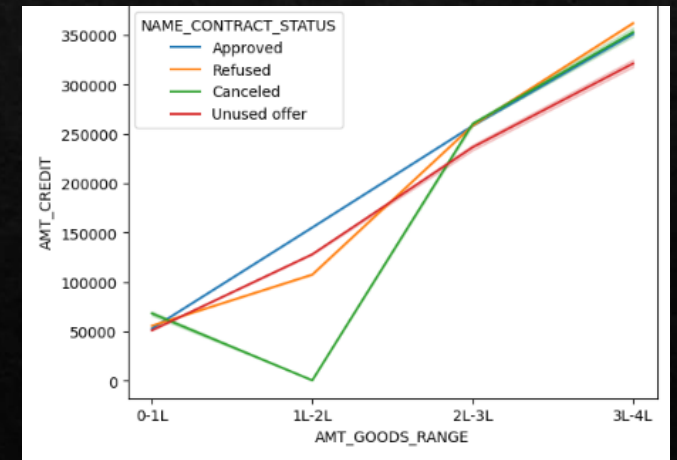
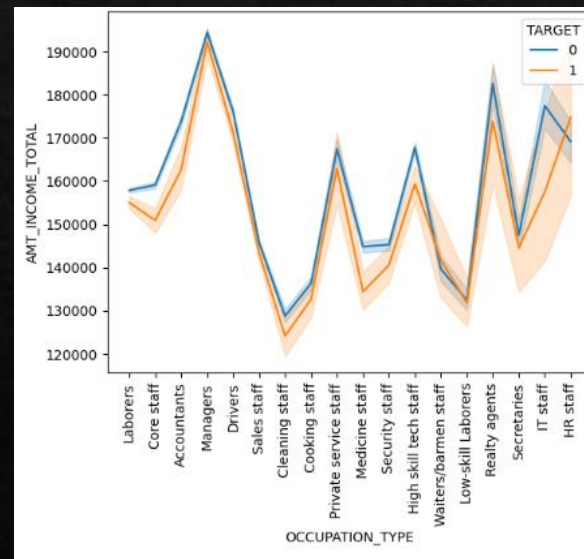
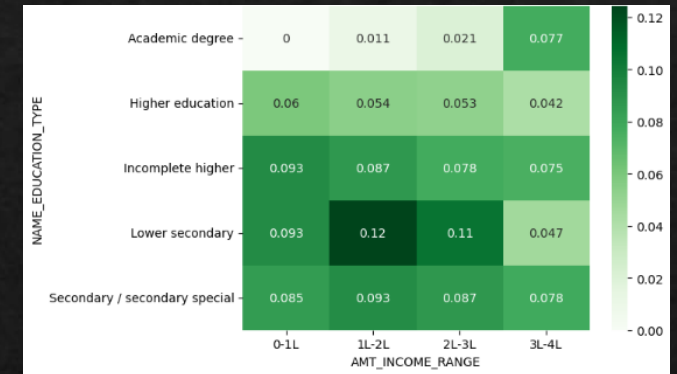
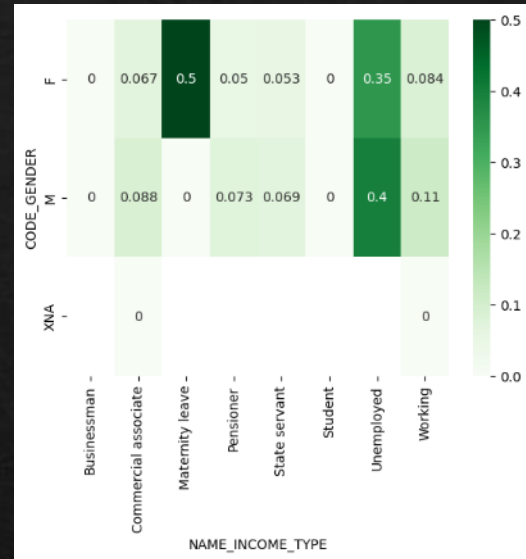
- ◇ Graph 1: Most approved credit range for the loan is between 0-1L
- ◇ Graph 2: Most approved application is for the loans which are requested for repairs purposes
- ◇ Graph 3: Most time taken for applicant to decide to apply for the loan is around 1-2L
- ◇ Graph 4: Most approved application for the loan are those good price is around 0-1L



Multivariant Analysis for Application data, Previous data and it's Inference

- Graph 1: The more defaulters are women in maternity leave followed by Men & Women who are unemployed
- Graph 2: The applicant who have lower secondary as the education and income range between 1-2L is not safe to give loan as they are the most defaulters followed by the applicant who has lower secondary and income range between 2-3L.
- Graph 3: the Waiter/barman staff and the HR staff are the most defaulters which says loans giving to them is of risk
- Graph 4: The client who have offered credit between 0-1L and the with goods price between 1-2L have the most unused offers

Conclusion: From the multivariant analysis we can conclude that there is a risk in giving loans to Female who are in maternity leave and the client who has lower secondary the education type.

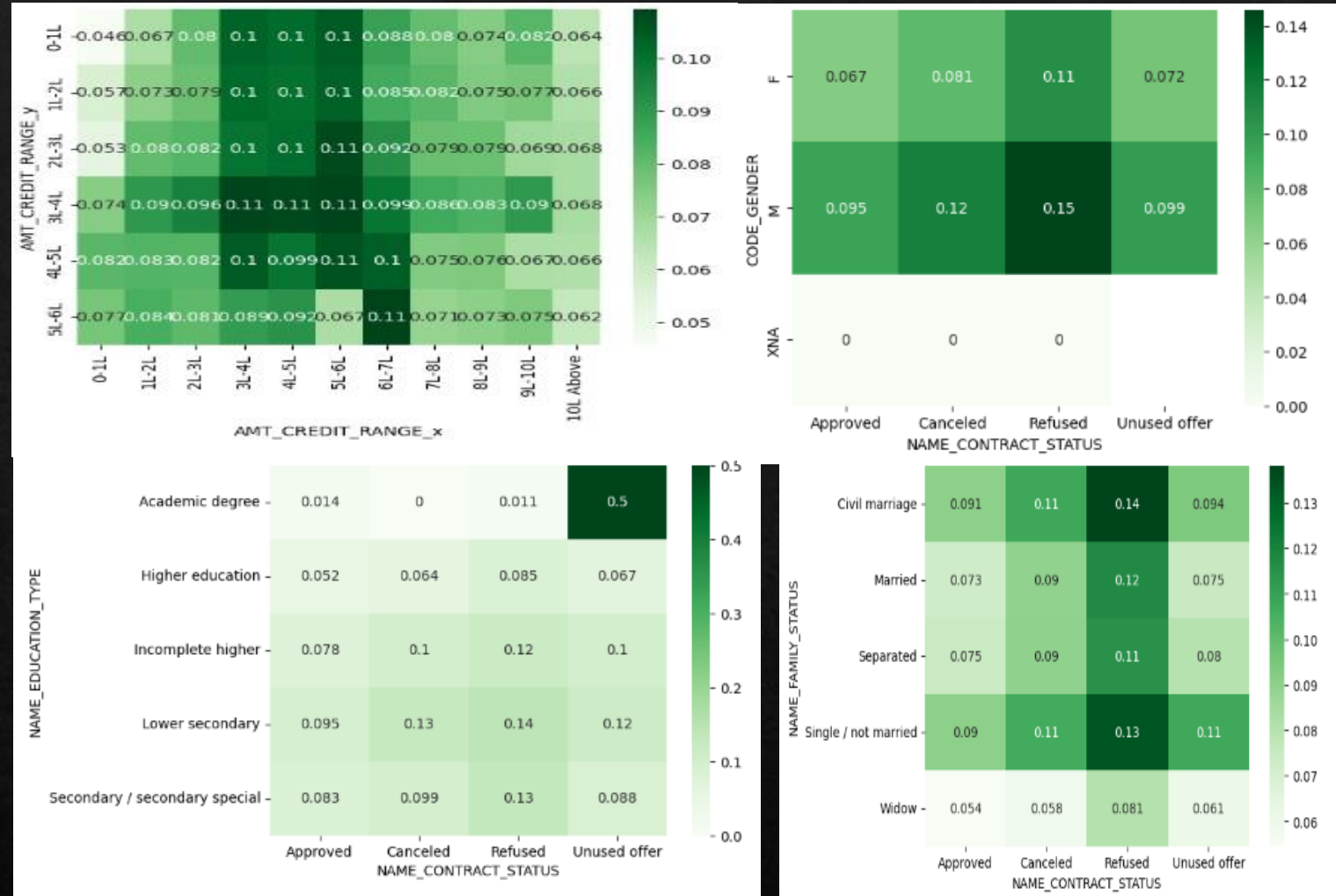


Merged data

- ◇ Graph 1: The client who had taken the loan for amount between 5-6L and the next time took loan for 6-7L had the most default rate
- ◇ Graph 2: The client who had refused the loan before and are male have the most default rate
- ◇ Graph 3: The client who has not used the offer first but later taken the loan with an Academic degree has the most default rate
- ◇ Graph 4: The client who had refused the loan before and had a civil marriage have most of the defaulters

Conclusion: The is highly noted that the client who had refused the loan at the first time and currently given loan have the most default rate.

Also, we had seen that the clients who had refused the loan before and have a credit amount around 4-5L have the more default rate when loan we given the next time



Recommendations

- ◆ The loan can be given to cancelled the first time and is having an Academic.
- ◆ It is recommended not to give loans to the Female clients who are on Maternity leave.
- ◆ It is recommended not to give loans to the client who have own realty and have more children.
- ◆ It is recommended to give loans to businessman and working professionals.
- ◆ It is recommended not to give loans to HR staff and waiter/barmen staffs.
- ◆ It is recommended that the loans are given to the clients between 35-45 years of age.
- ◆ It is recommended to give loans with a low credit amount between 0-2L.
- ◆ It is recommended not to loans to the client who had cancelled the loans first time and took less than 1 year to decided.