

Credit EDA Case Study Assignment

Submitted By
Shreyosi Chattopadhyay

Problem Statement

- In this case Study we aim to analyse that when bank receives an loan application, they can decide for loan approval based on the applicant's profile.
- Two types of risks are involved-
- If applicant is likely to repay the loan and not approving the loan results in loss of business to company.
- If applicant is not likely to repay the loan payment still approving the loan may lead to financial loss to company.

Tools Used in Assignment

- Numpy
 - Pandas
 - Matplotlib
 - Seaborn
-
- Jupiter Notebook is used to showcase the code

Approach Involved

- There are two datasets provided-
- "application_data.csv"- Denotes current application
- "previous_application.csv"- Denotes previous application history

Steps Involved-

Data Cleaning of current application and analyzing it individually

Data Cleaning of previous application and analyzing it individually

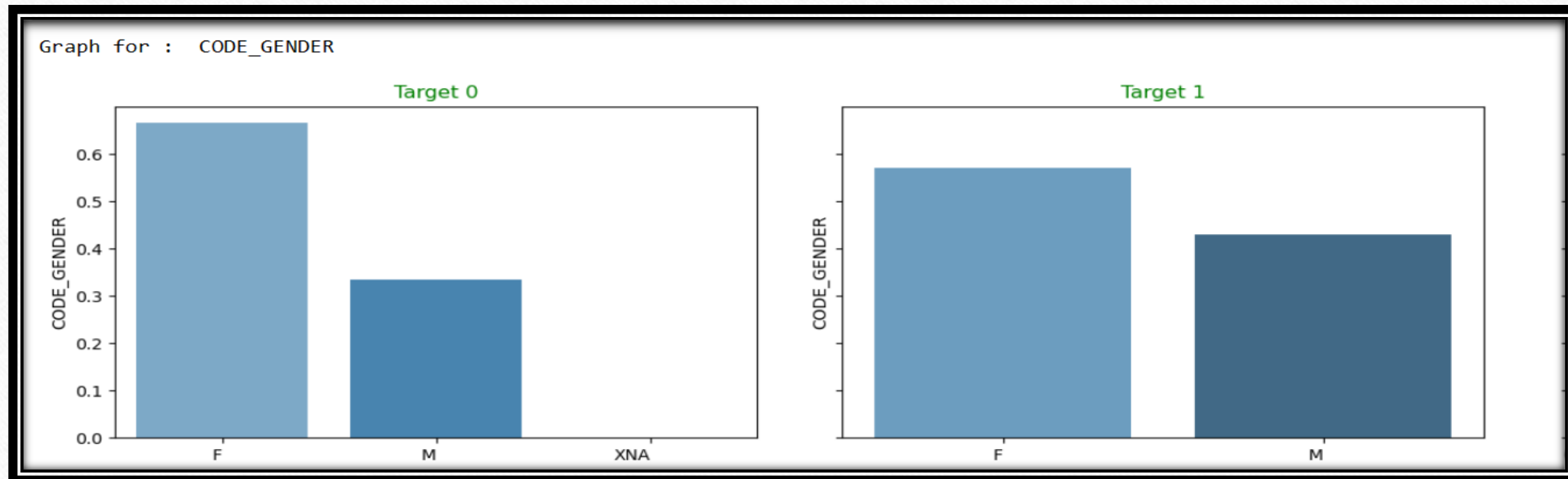
Merging both Datasets and then analyzing it together

Steps Involved

- Data Loading and Data Inspection
- Data Cleaning
- Handling Null Values
- Handling Outliers
- Data Visualization using Univariate, Bivariate and Multivariate analysis

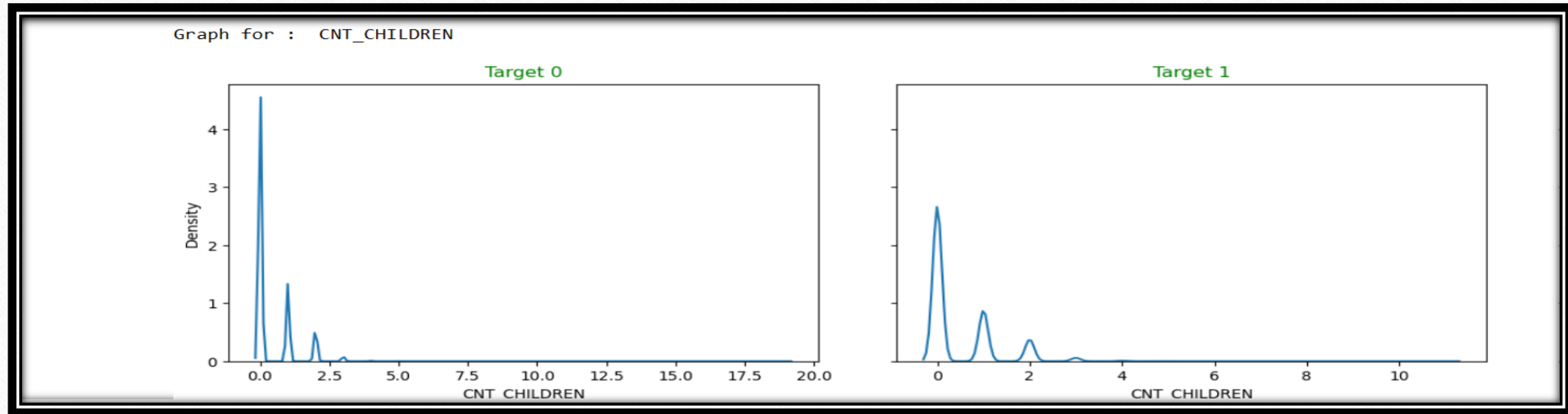
Analysis on Application Data

Gender Wise Analysis



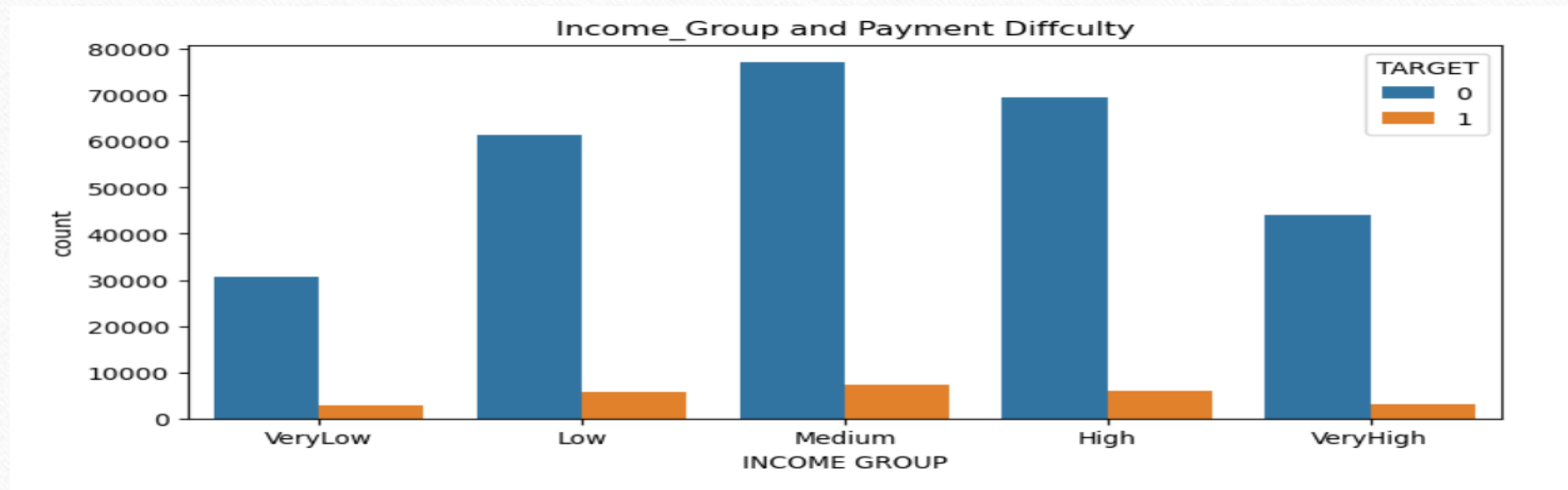
According to above graph we can clearly see that Males percentage are more likely to be defaulter than Females, so we can provide loans to Females easily

Defaulter VS Count of Children



According to the above graph, we can clearly see that who has more number of children they are less likely to be defaulters

Income Group VS Payment Difficulty



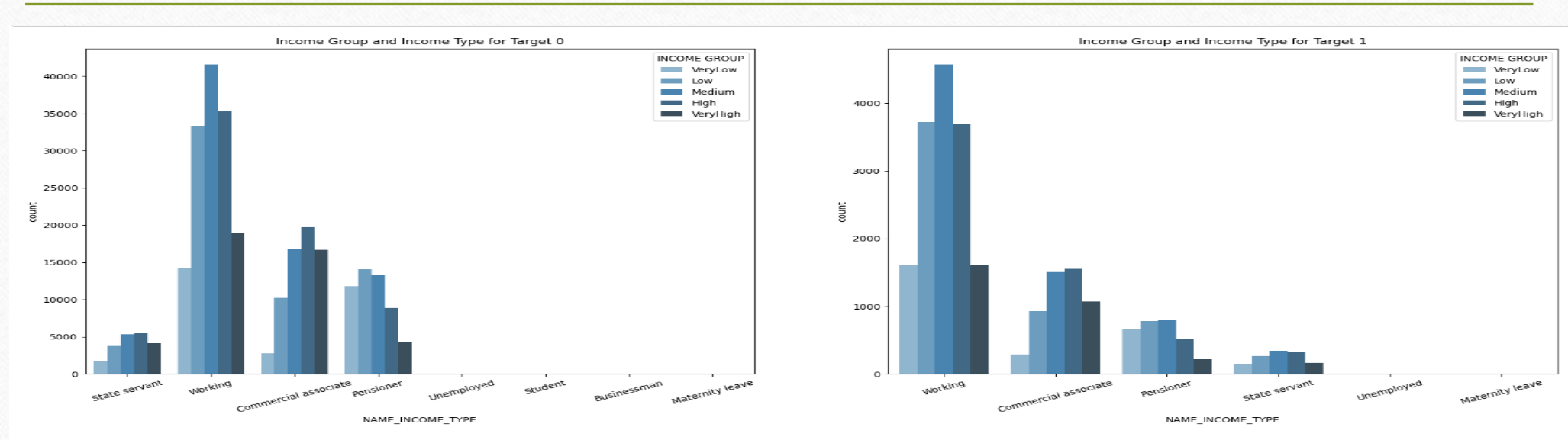
As per above graph it clearly states that Medium Income group are given maximum number of loan amount. Default value per loan is highest in High Income group as AMT_CREDIT is higher.

Contract Type VS Gender

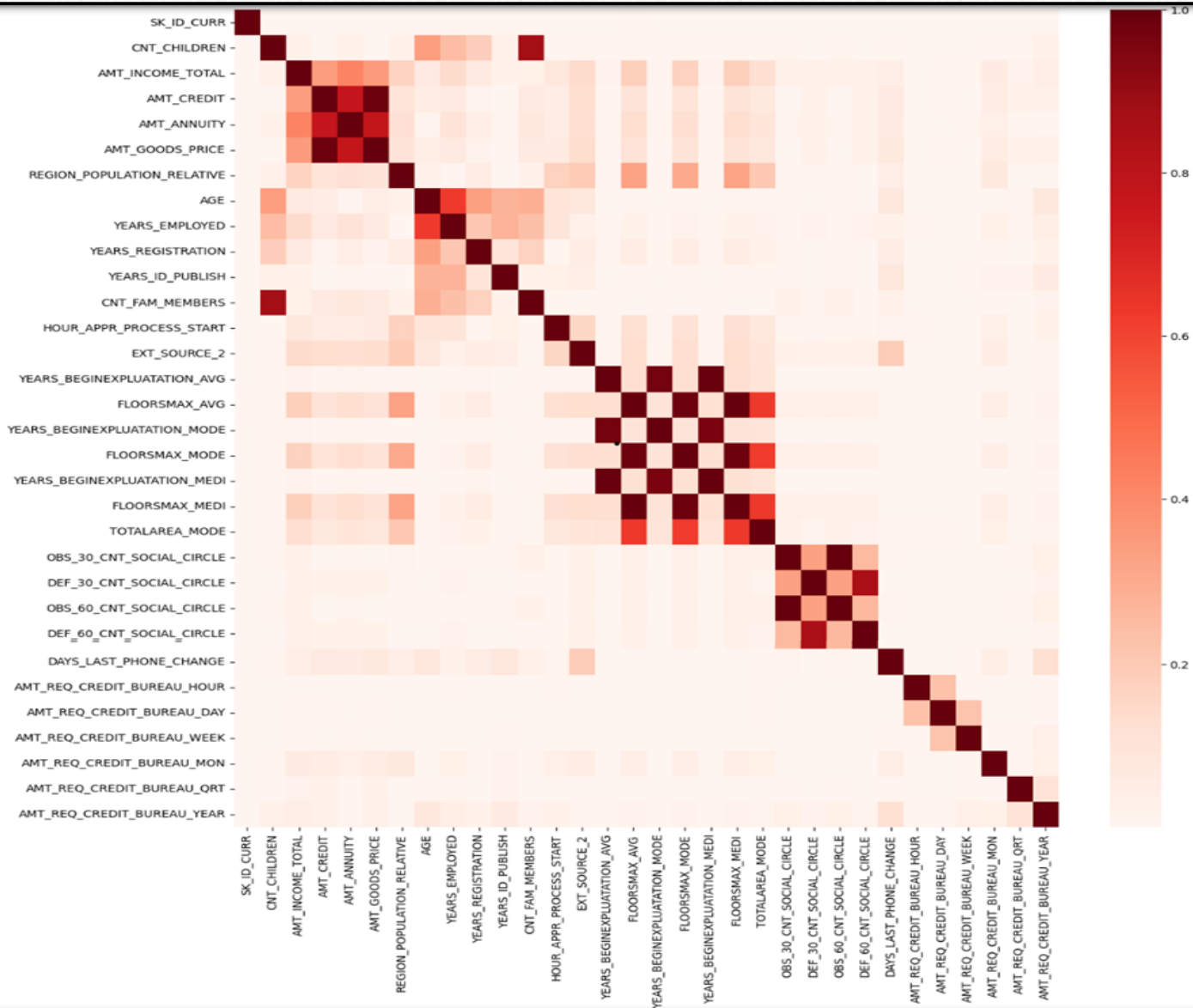


As per above graph Females are more loan applicant, ratio of male applicant defaulting are higher

Income Group VS Income Type



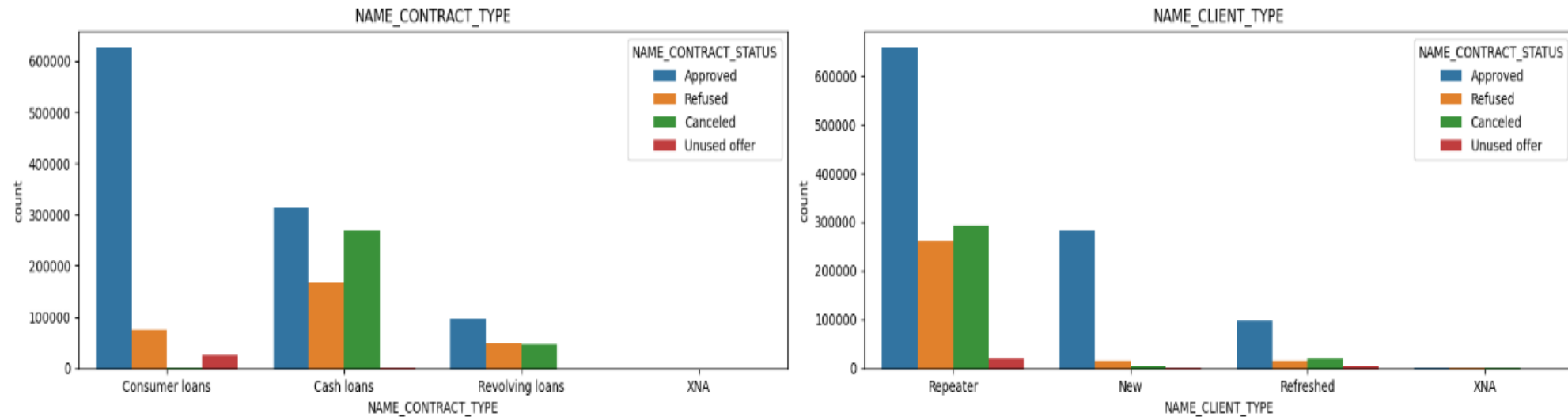
As per the above graph, Medium Income Working Group has most defaults



Heatmap of
two correlation
for Target 0

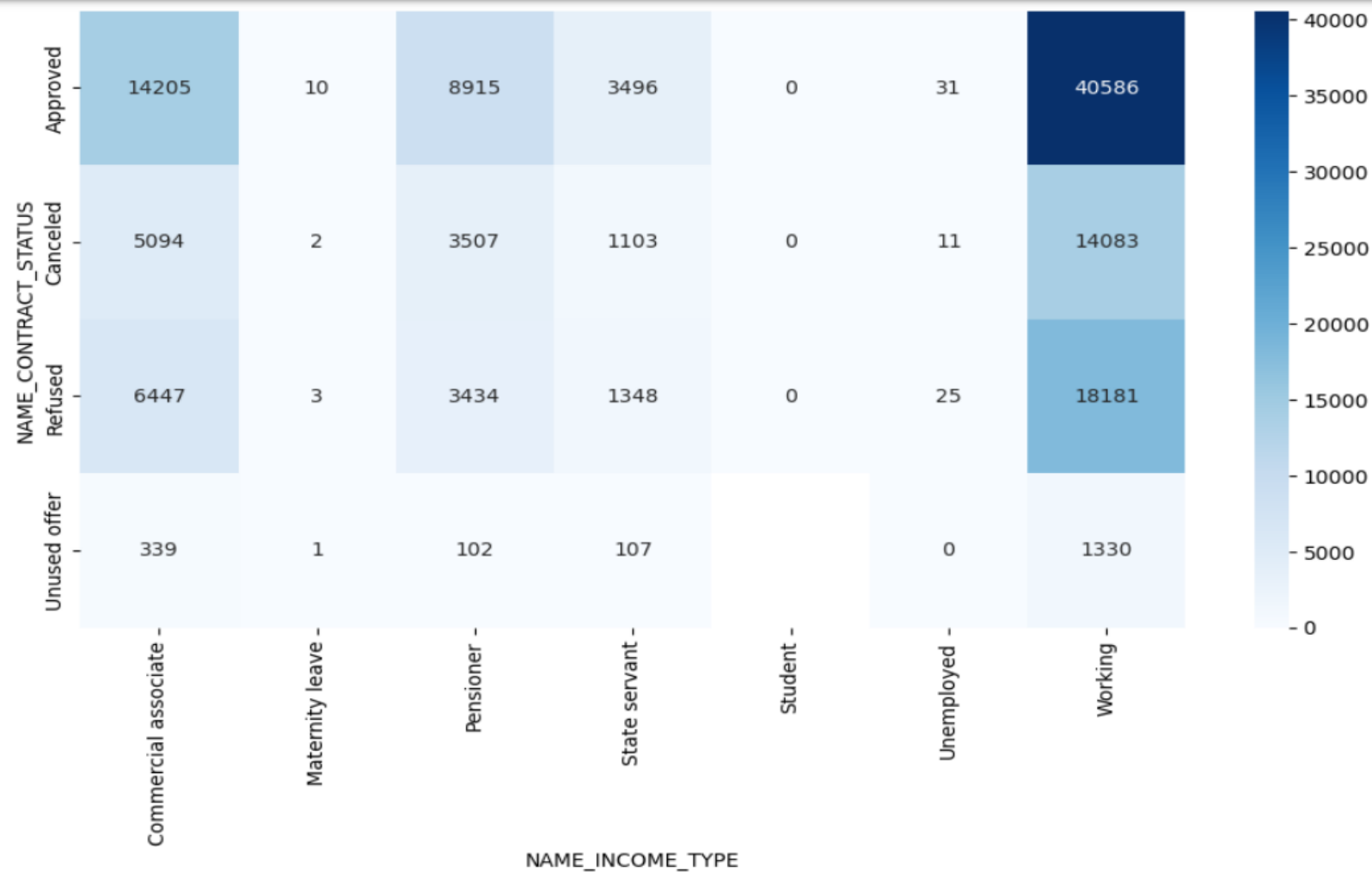
Analysis on Previous Application Data

Contract Type and Client Type different Contract Status



As per the above graph it clearly states that consumer loans has highest number of applicants with most cancelled loans. More Cash loans have been refused than consumer loans. Bank has more approved repeaters in all categories.

Analysis on Merged Data Analysis



Heatmap of correlation of Merged Data – Income Type VS Contract Status

Conclusions

- Males are more likely to be defaulters, so females would be ideal candidate to offer loans
- Client with more number of children are likely to be defaulters
- Working people pay off their debts
- Higher Income Range likely pay off their debts
- Repeaters are the ones who got loan approved previously
- Most defaulters are the ones whose loans was previously refused

Thank You