CREDIT EDA ASSIGNMENT

Nikita Mundhra | DS 60 ACP | 28-11-2023

PROBLEM STATEMENT

➤ Loan Providers Find it Hard to Decide The Right Person Whom They Can Loan Money.

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

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

   ➤ 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.

➤ The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
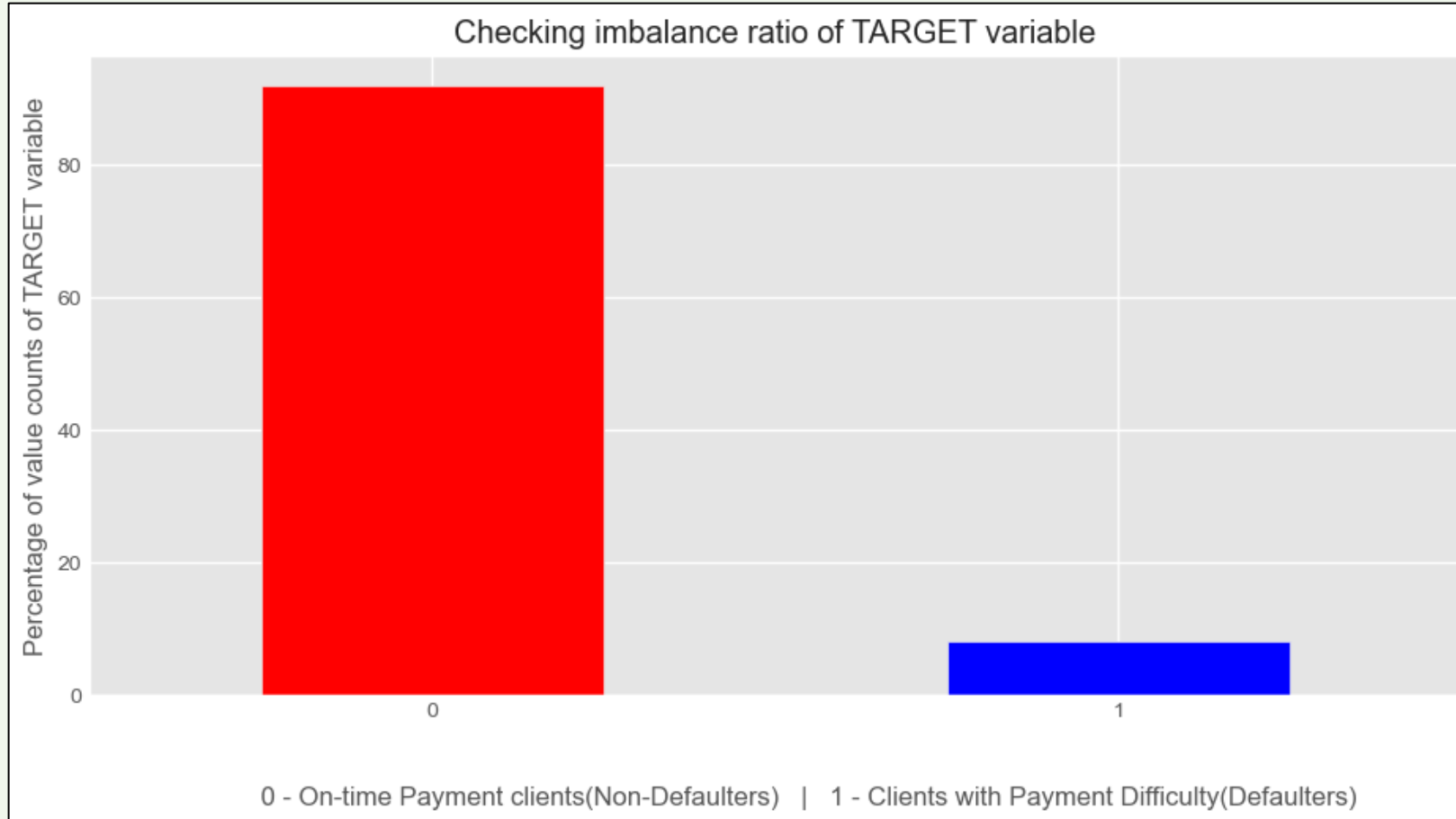
PROBLEM STATEMENT

➢ For This We Are Provided With 2 Data Sets.

    ➢ application_data

    ➢ previous_application

➢ In Application Data Current Application Details is Given While in Previous Data Dataset Customers Previous Record is Mentioned; if he/she took loan in the past & his/her Loan Return Data.

➢ We Are Supposed To Analyze These Data Sets & Come Up With Solution For Company Whether It Is Safe To Give Loan To a Certain Customer, Or Need To Reduce Amount Or lending (to risky applicants) at a higher interest rate, etc.

ANALYSIS APPROACH

➤ First I had to go through the data sets given and understand the columns and their description.

➤ I started with 'application_data.csv' data set.

➤ Then very first step was to clean the data by getting rid of all null values and deleting unwanted columns.

➤ After erasing the unwanted columns, I treated the outliers one by one for each column.

➤ Next was to do the analysis for entire data set starting with Univariate Analysis – How factors are deciding the loan application like which gender clients are applying for more loans, etc

➤ Then I did Bivariate Analysis with respect to Target Variable - What is record of Application Against Target Data or Should I Say Whose Application is Approved More Often Male Or Female etc.

➤ The same approach was followed for 'previous_application.csv' data set.

CHECKING THE IMBALANCE RATIO

Checking imbalance ratio of TARGET variable

0 - On-time Payment clients(Non-Defaulters) | 1 - Clients with Payment Difficulty(Defaulters)

➢ Imbalance Ratio with respect to Repayer (0) and Defaulter(1) is given: 11.39.

➢ 1 in every 11 applicant has payment difficulty.

APPLICATION DATASET - ANALYSIS
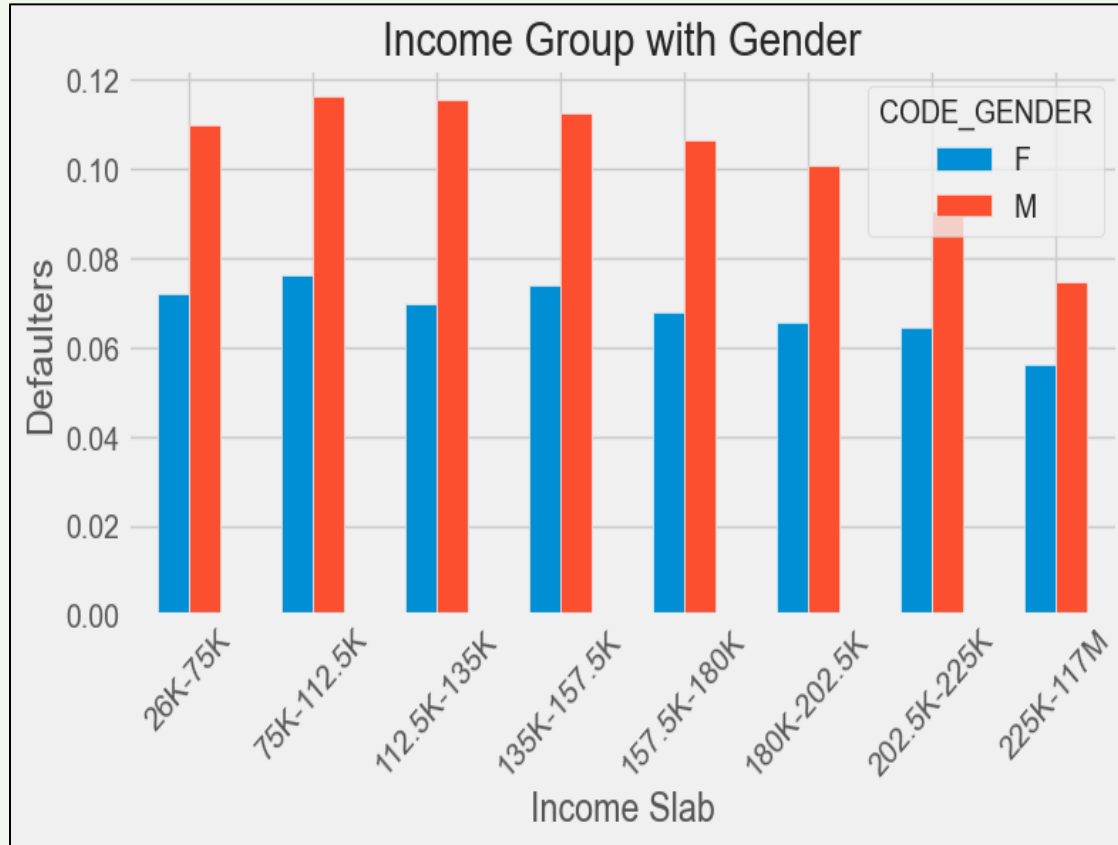
- Male clients earn more than female clients, however they have higher default rates.

- Clients with less income tend to have payment difficulties than clients earning a high income.

APPLICATION DATASET - ANALYSIS
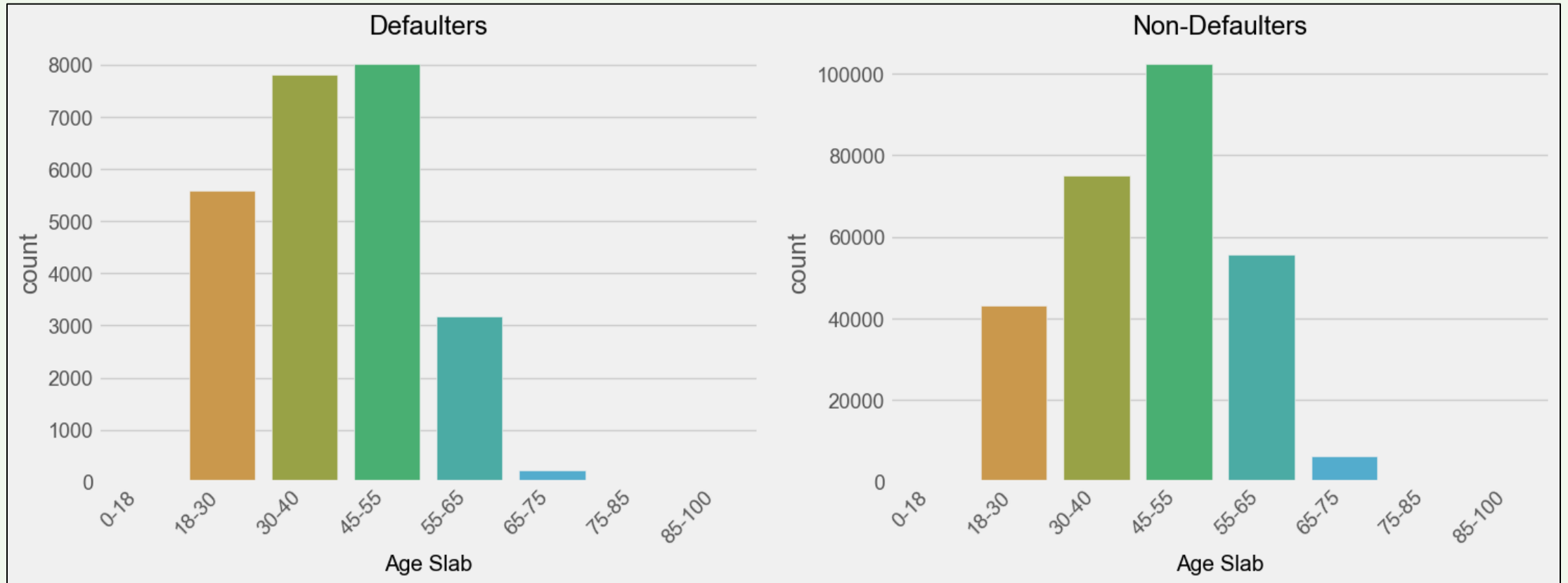
➢ Majority of the clients lie in the age group 30-55 years of age tend to default more.

➢ Senior citizens clients tend to pay back their loan on time.

APPLICATION DATASET - ANALYSIS

Education Type with Family Status
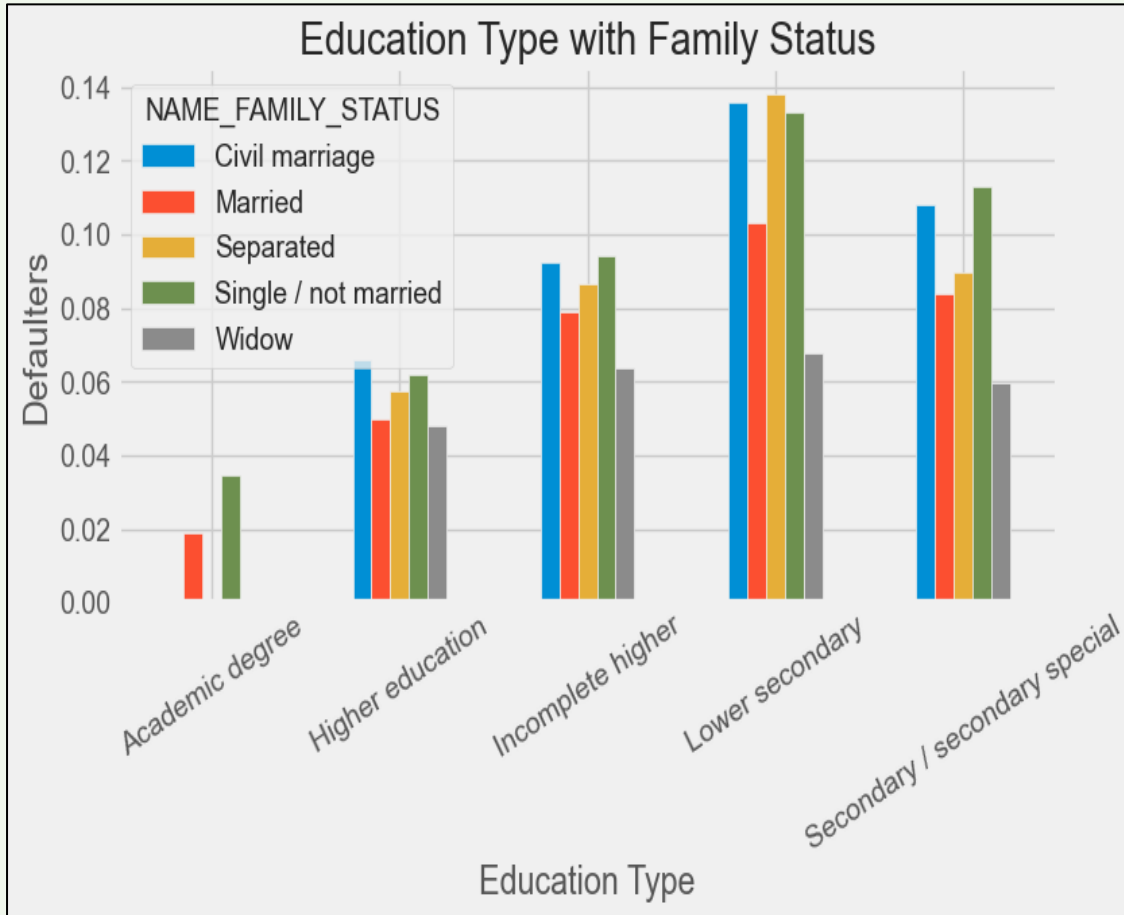
> ➤ Clients with family statuses like 'civil marriage,' 'not marriage,' and 'separated' who have achieved an 'Lower Secondary' education tend to have payment with Difficulties.

> ➤ Clients having an Academic degree have lower default rates.

> ➤ Widows are having pretty low default rates as compared to other family status.
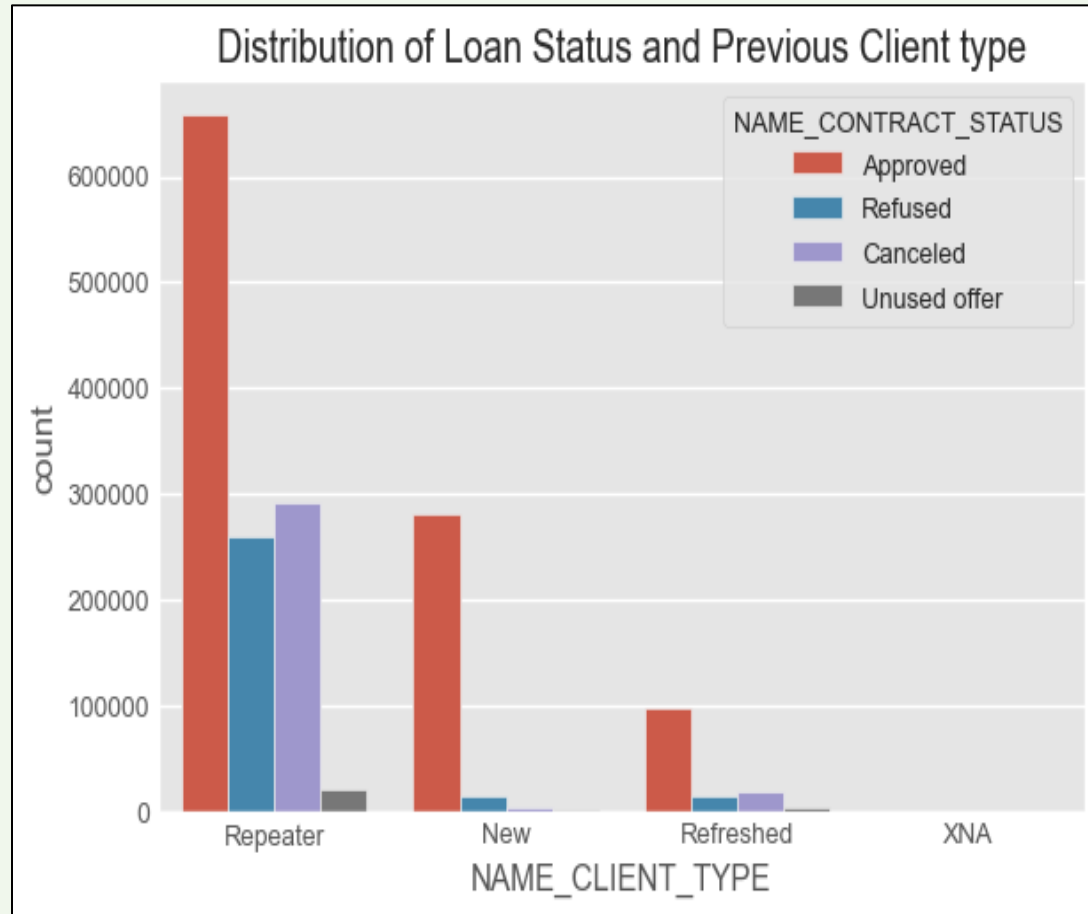
APPLICATION DATASET - ANALYSIS
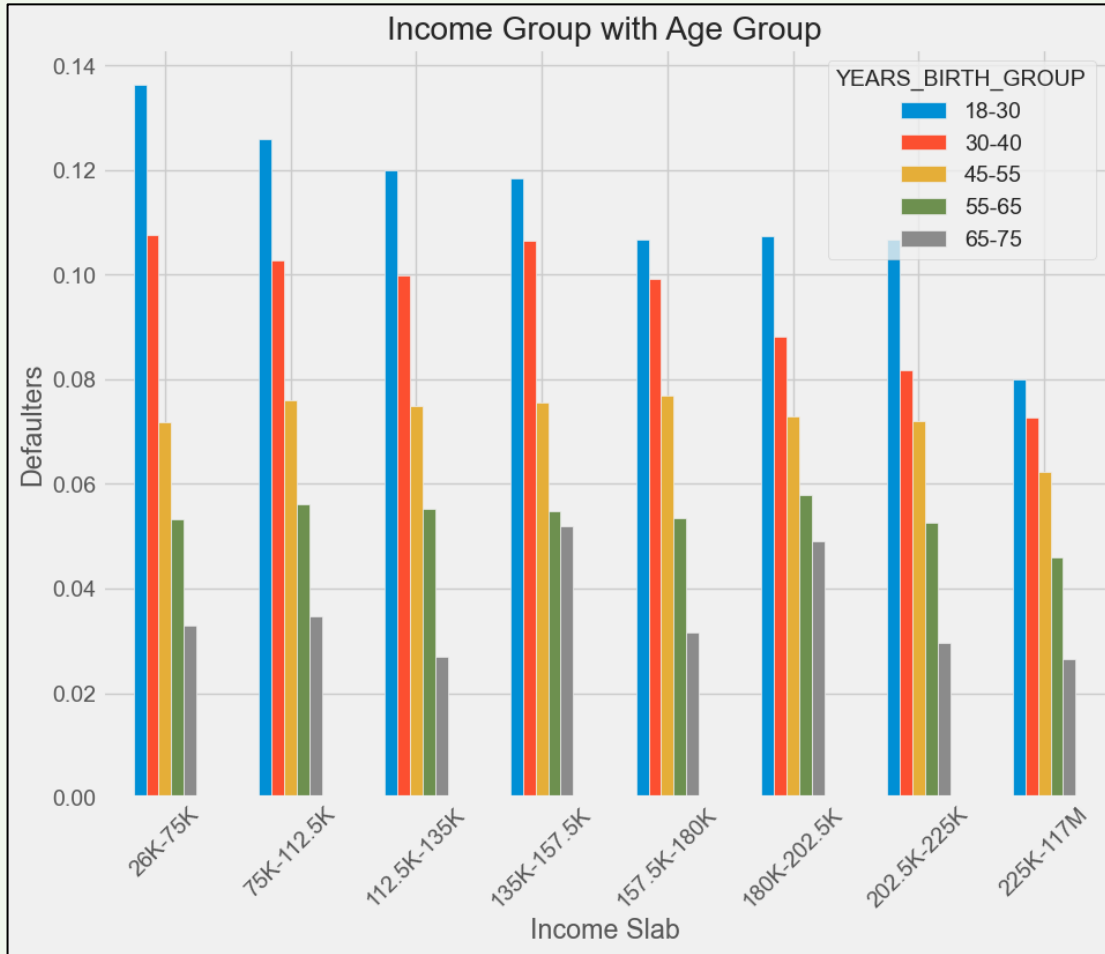
➤ High income groups are less defaulter than comparatively lower income groups.

➤ Young clients having low income tend to have more difficulty in payment of loans.

➤ Senior clients with high income are less likely to be defaulted.

# PREVIOUS APPLICATION DATASET - ANALYSIS

Distribution of Loan Status and Previous Client type

➢ Banks tend to approve loans for repeat customers.

➢ For new customers, Banks generally tend not to cancel the loans.

PREVIOUS APPLICATION DATASET - ANALYSIS

Distribution of Loan Status wrt Gender
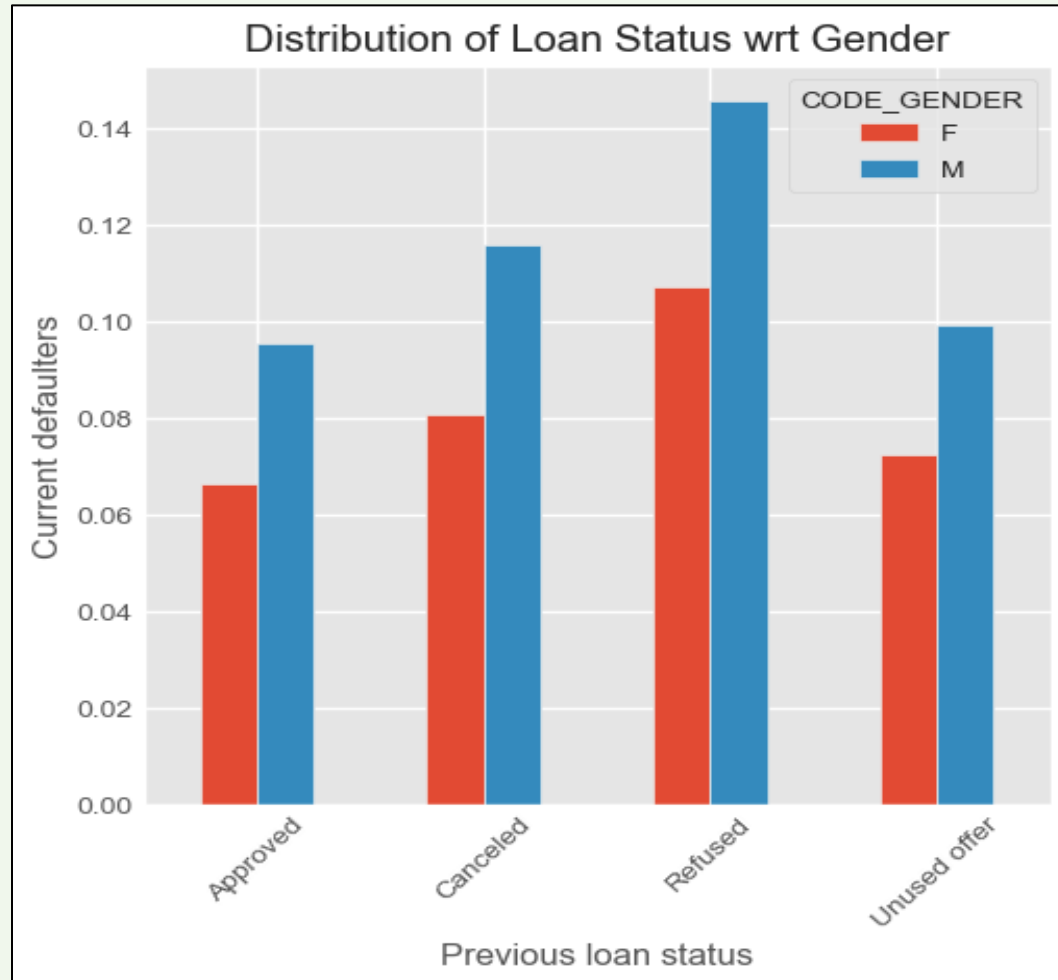
> Previously also Banks have refused the loans for Male clients as they have high default rates.

> Previously, Banks approved lesser loans of female clients than male clients.

> Male clients are the one who have cancelled / not used the loans more than Female clients.

MERGED DATASET - ANALYSIS
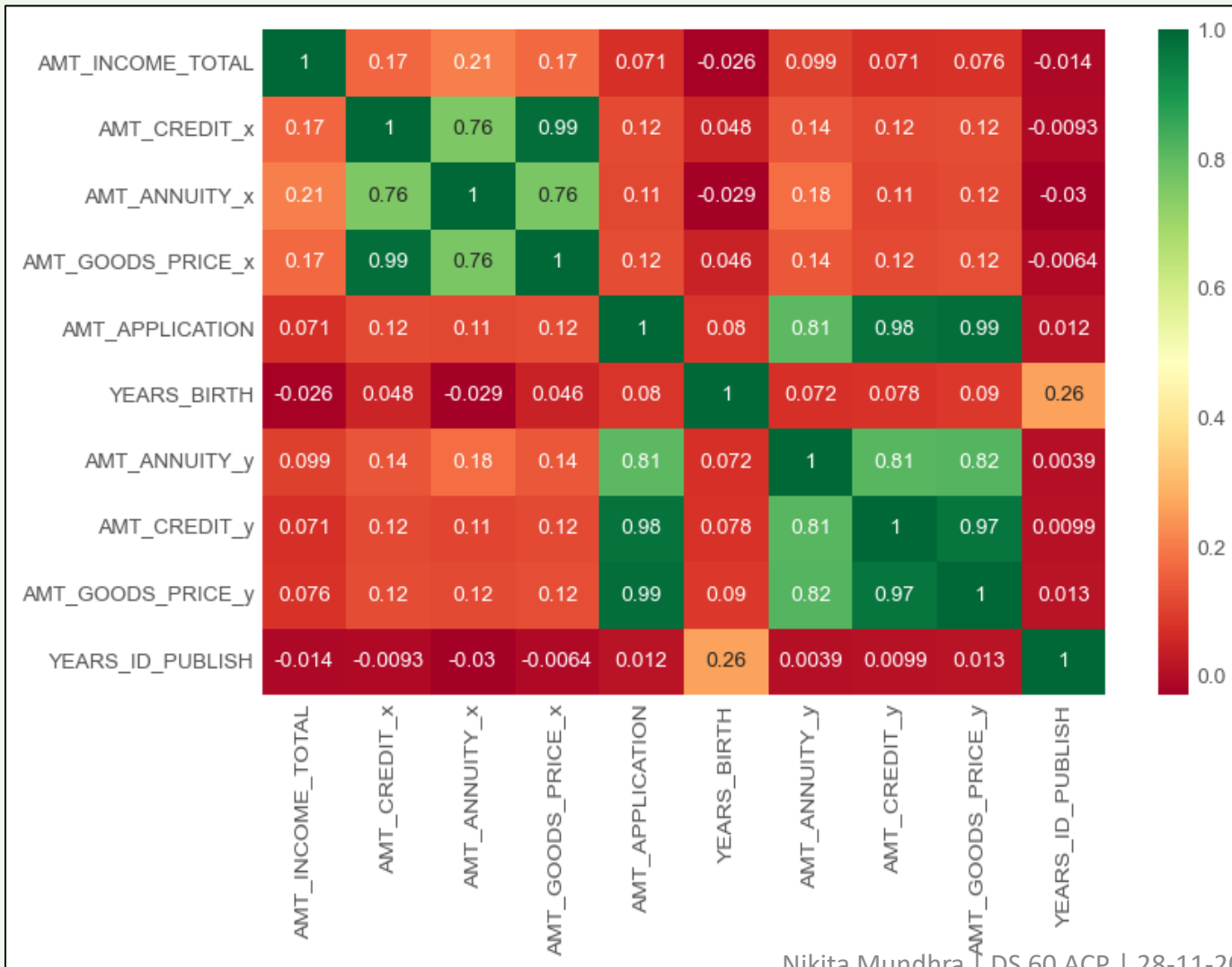
- ➤ There is Very High Correlation Between Amount Goods Price & Amt Credit; Meaning People Having Higher Goods Price Opt For Higher Amount Of Loan

- ➤ Amt Annuity Has Good Correlation With Amt Credit & Amt Goods Price - HIgher the credit value, higher the loan annuity value.

- ➤ Age is having negative correlation with Total Income indication no robust relation between age and income.

INFERENCES

➤ Banks should focus on female clients as they have less chances of payment difficulty.

➤ Banks should be careful while providing loans to clients with high income as they are ones who have a high chance of becoming defaulters.

➤ Females Ask For Loan More Than Males But Yet Males % Of Approval is Higher Than Females.

➤ There is an imbalance ratio of 11%, ie 1 in every 11 people have payment difficulties.

➤ Banks should focus on Students and Business as they are less likely to default.

➤ Clients with fewer dependents, ranging from zero to two children, are the ones who are not having payment difficulty.

➤ Individuals in civil marriages or single status exhibit a higher default rate.

➤ Majorly 30-45 Years Age People Apply For Loan & Get Approval.

➤ Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.

➤ Certain job categories, such as low-skill laborers, drivers, waiters/barmen staff, security staff, laborers, and cooking staff, carry a higher risk of default.

Thank You !