Risk Analysis of Bank Loan

Credit EDA Assignment

Done By: Anurag Kumar Pal

Content

- Problem Statement
- Business Objective
- Datasets Provided
- Approach Followed
- EDA Results (For Applications, Previous Applications and Merged Dataset)
- Conclusions
- Recommendations

Problem Statement

- Loan approval decisions in the consumer finance industry are often challenged by insufficient or non-existent credit histories, leading to potential risks of defaults.
- Leveraging Exploratory Data Analysis (EDA) techniques within the context of risk analytics for banking and financial services, we need to identify patterns that can distinguish between borrowers capable of repaying loans and those likely to default, thereby mitigating financial losses for the lending company.

Business Objective

- The primary objective is to utilize EDA to uncover meaningful patterns within consumer and loan attributes that correlate with loan repayment difficulties. By doing so, the company aims to achieve the following goals:

  - **Improved Loan Decision Making**: Through the identification of key variables and their significance, the company intends to enhance its loan approval process. This involves distinguishing between clients likely to repay their loans and those at a higher risk of default.

  - **Risk Minimization**: By understanding the driving factors behind loan defaults, the company can take proactive actions, such as adjusting loan amounts, interest rates, or even denying loans to high-risk applicants. This approach helps minimize potential financial losses.

Datasets Provided

- Following are the 3 datasets provided to us-

    - **'application_data.csv':** Contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

    - **'previous_application.csv':** Contains information about the client's previous loan data. It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.

    - **'columns_description.csv':** Data dictionary which describes the meaning of the variables.

Approach Followed

- We have been provided with two datasets-
  - “application_data”: Contains details regarding the current application of a loan
  - “previous_application_data”: Contains details about the same applicant when he/she previously applied for a loan
- Following are the common EDA steps used for the above datasets-
  - Removing the columns that aren’t required
  - Null value treatment
  - Data sanity checks
  - Outlier handling
  - Creation of new categorical and numerical variables
  - Performing Univariate, Bivariate and Multivariate Analysis
- Both the datasets were analyzed individually. Post that, both the datasets were joined together to gain more insights
- From each visualization, few insights were drawn
- Finally, we provided a list of recommendations that would help in risk minimization

EDA of “application_data” dataset

- Following are the common EDA steps used for the above dataset-
  - Removing the columns that aren’t required
  - Null value treatment
  - Data sanity checks
  - Outlier handling
  - Creation of new categorical and numerical variables
  - Performing Univariate, Bivariate and Multivariate Analysis

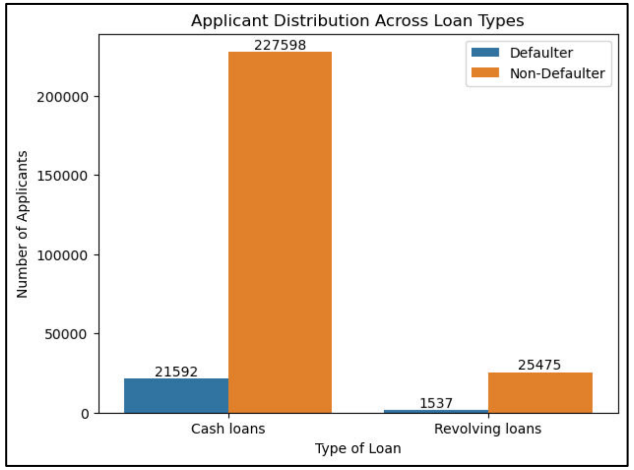
Applicant Distribution Across Loan Types

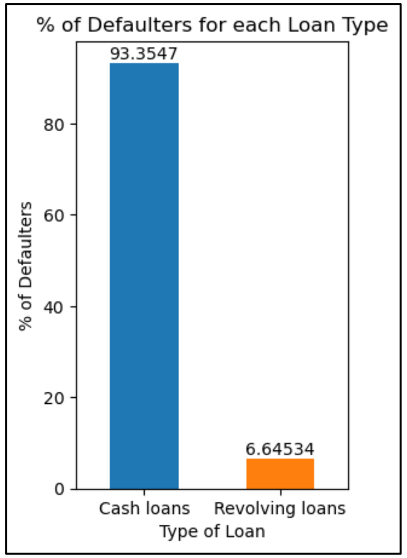
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “Contract Type”
- Type of Analysis: Univariate

➢Insights Drawn

- Cash Loans have higher defaults than Revolving Loans
- Revolving Loans are approx. 10% of Total Loans





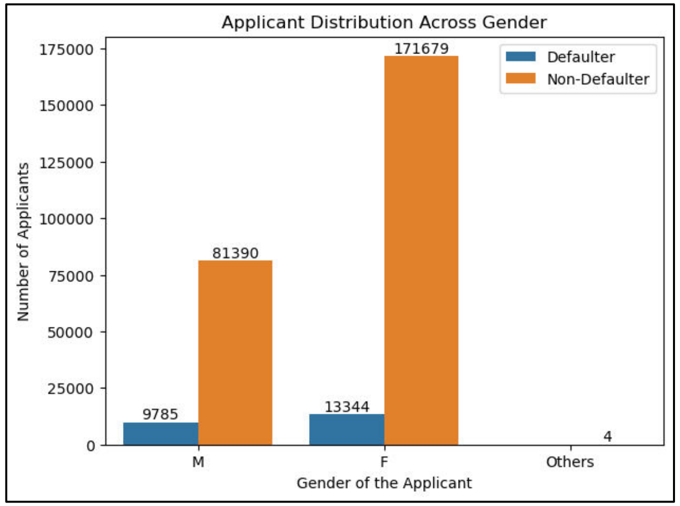
Applicant Distribution Across Gender

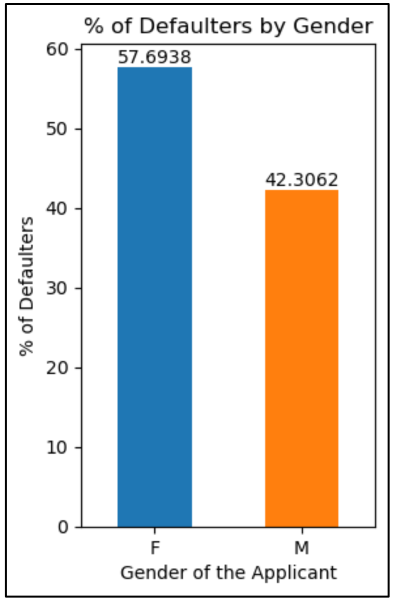
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “Gender”
- Type of Analysis: Univariate

➢Insights Drawn

- Women have higher defaults than men but on an overall level, Men have higher % of default.
- On an average, women tend to take more loans than men.





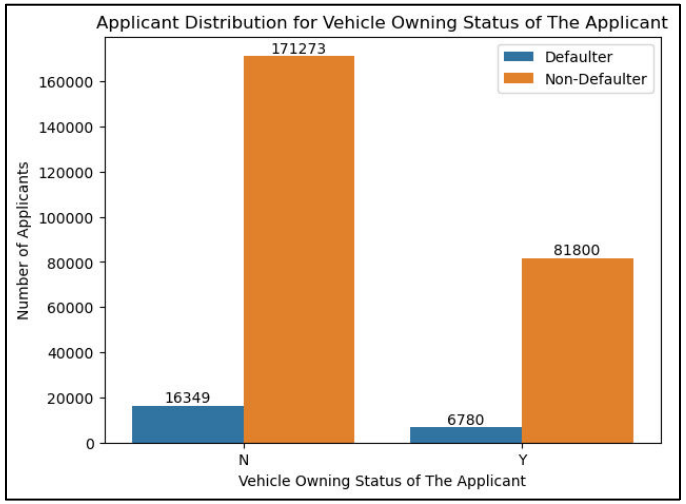
Applicant Distribution for Vehicle Owning Status of The Applicant

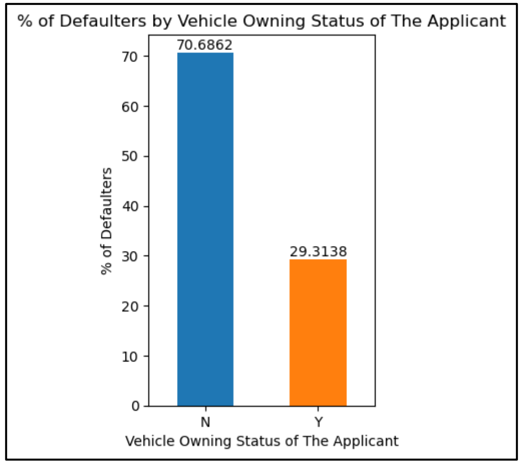
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “FLAG_OWN_CAR”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants who don't own a Vehicle have a higher tendency of being a defaulter
- Majority of the applicants don't own a vehicle of their own





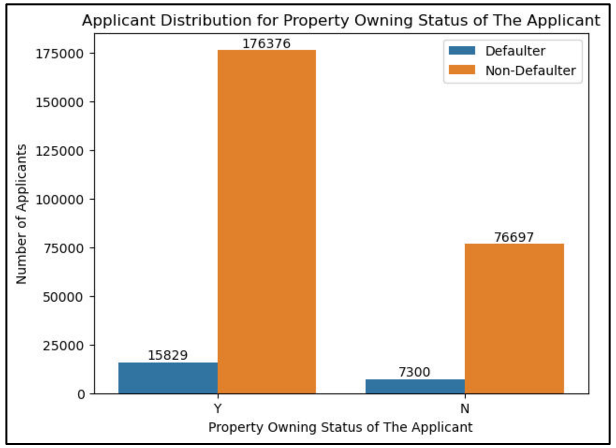
Applicant Distribution for Property Owning Status of The Applicant

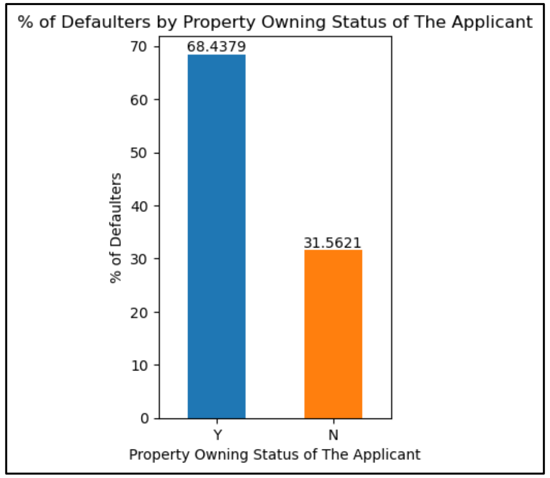
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “FLAG_OWN_REALTY”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants who own a property have a higher tendency of being a defaulter
- Majority of the applicants own a property of their own





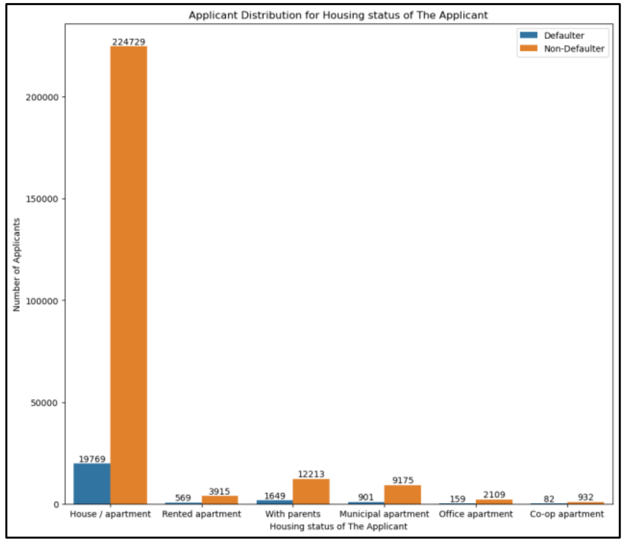
Applicant Distribution for Housing status of The Applicant

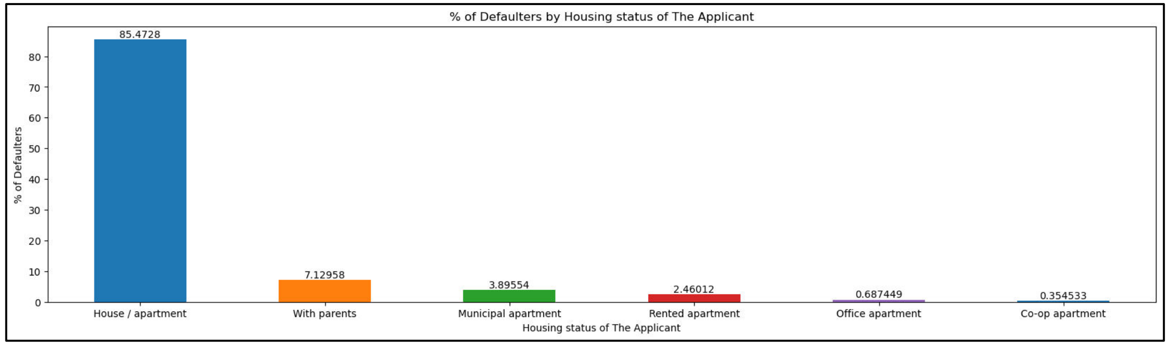
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “NAME_HOUSING_TYPE”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants who live in a "House/Apartment" have a higher tendency of being a defaulter
- Applicants who live in "Co-operate Apartments" are less likely to default

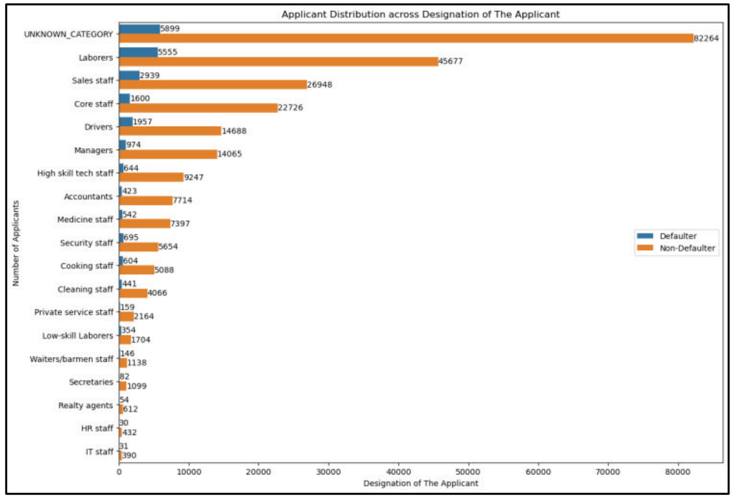




Applicant Distribution across Designation of The Applicant

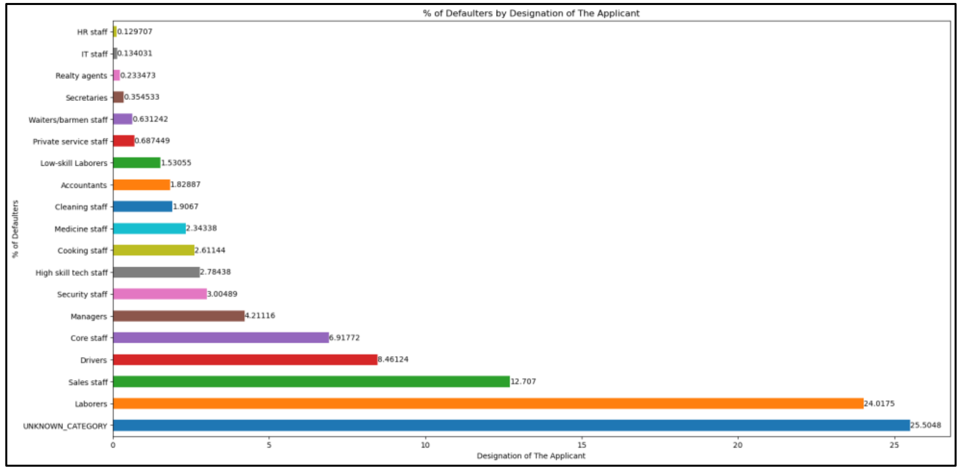
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “OCCUPATION_TYPE”
- Type of Analysis: Univariate



➢Insights Drawn

- Applicants who belong to the "UNKNOWN_CATEGORY" have defaulted the most
- Applicants belonging to the "Laborers" have the second highest tendency to default
- Working applicants from the "Human Resources" and "IT Industry" are least likely to default their loan payment



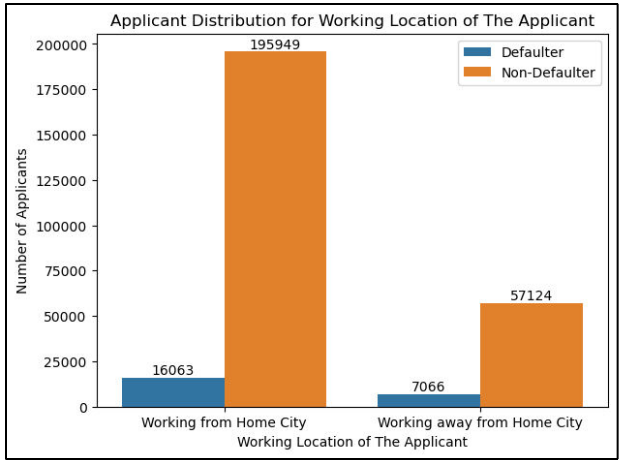
Applicant Distribution for Working Location of The Applicant

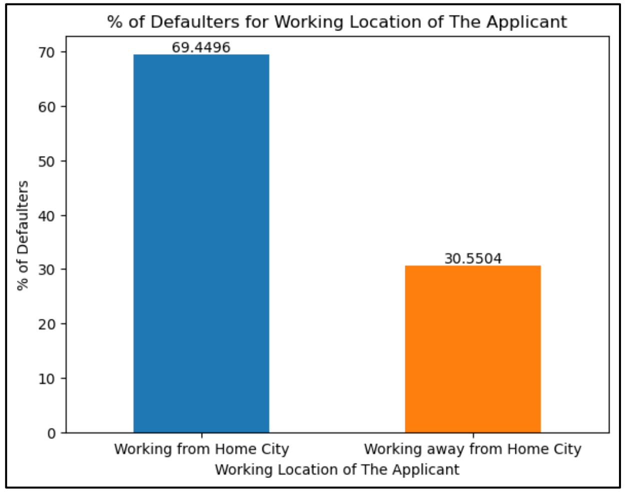
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "EMPLOYEE_JOB_TYPE"
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants who are working from home city have a higher tendency to default
- Applicants who are working from home city have taken more loans than the one living away from home city

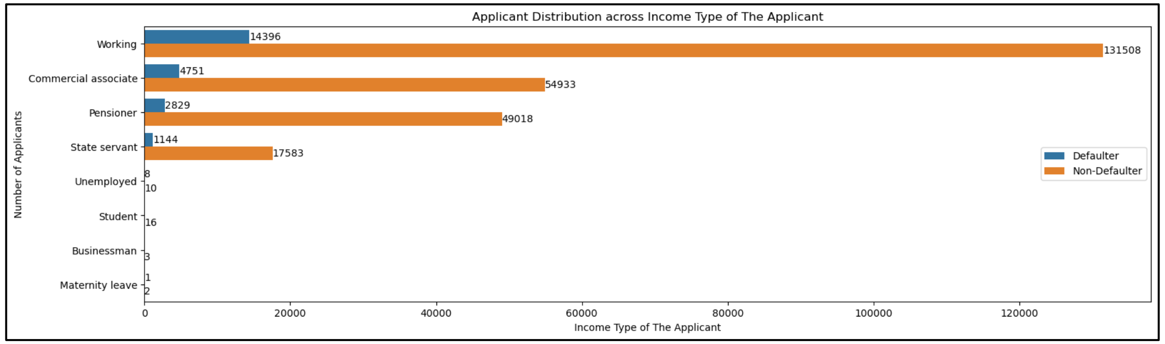




Applicant Distribution across Income Type of The Applicant

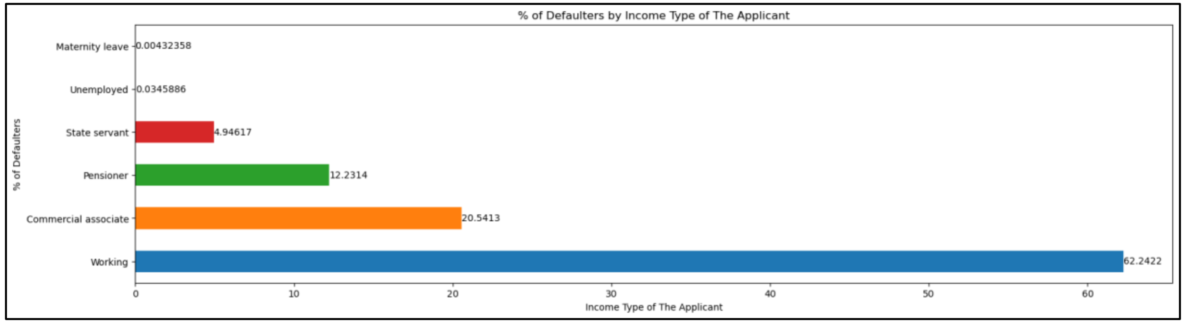
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “NAME_INCOME_TYPE”
- Type of Analysis: Univariate



➢Insights Drawn

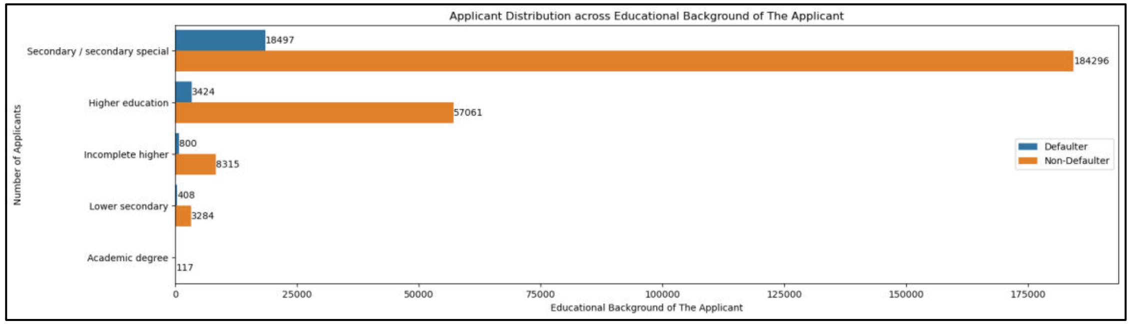
- Applicants who are working have the highest tendency to default
- Applicants who are on maternity leave have the least tendency to default



Applicant Distribution across Educational Background of The Applicant

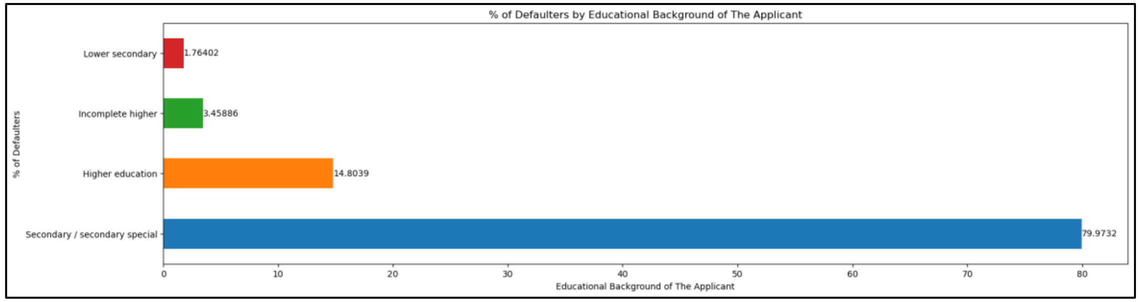
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “NAME_EDUCATION_TYPE”
- Type of Analysis: Univariate



➢Insights Drawn

- Applicants with "Secondary/Secondary Special" educational background have the highest tendency to default
- Applicants who hold an academic degree have made no loan defaults till date



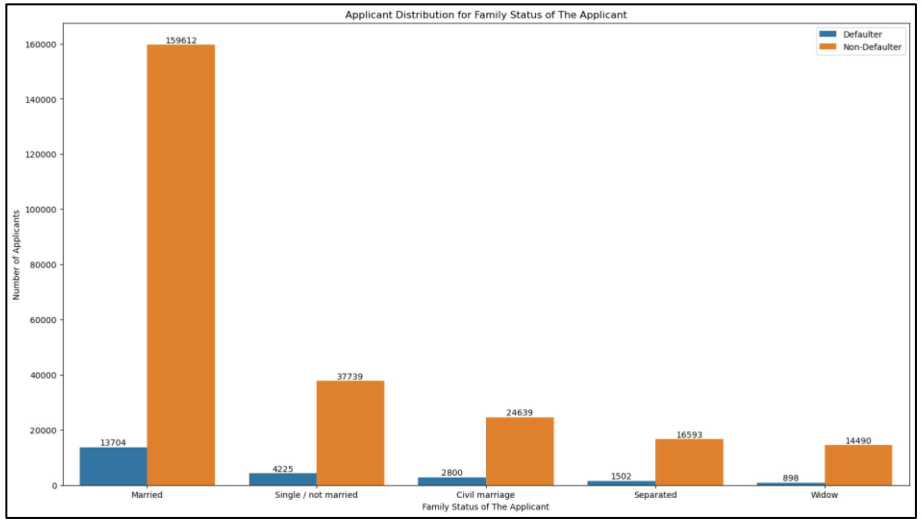
Applicant Distribution for Family Status of The Applicant

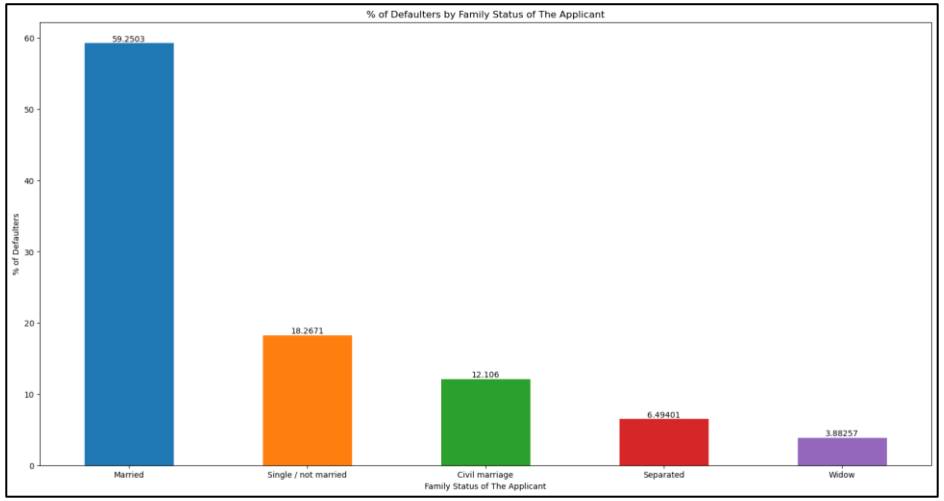
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "NAME_FAMILY_STATUS"
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants who are Married have the highest tendency to default

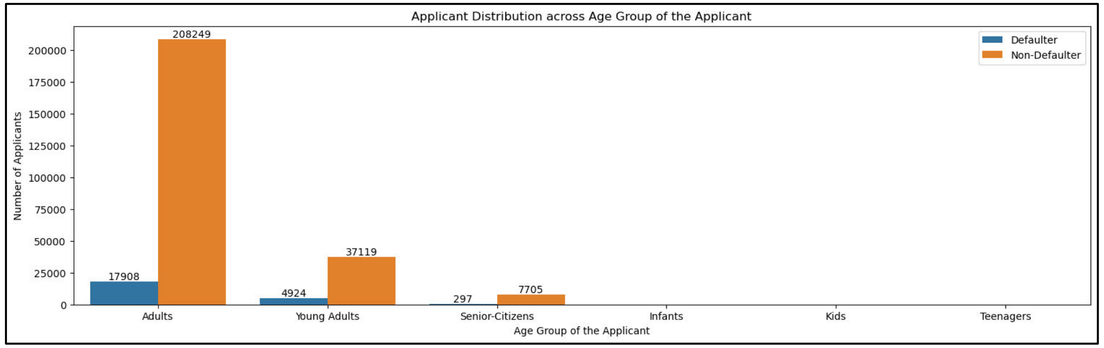




Applicant Distribution across Age Group of the Applicant

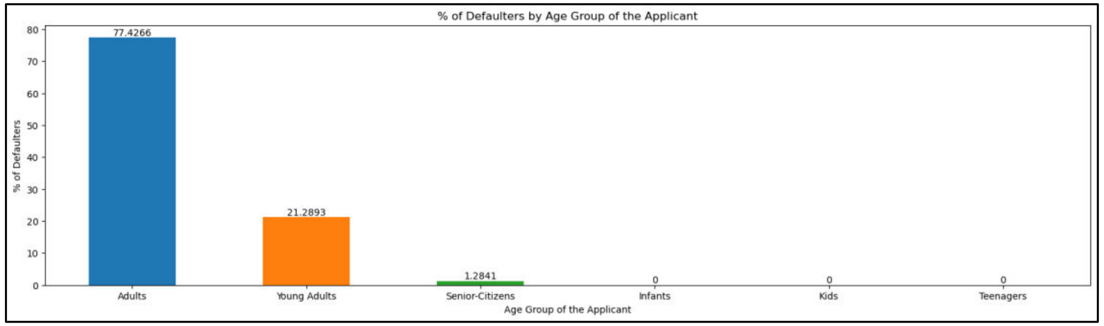
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “AGE_GROUP”
- Type of Analysis: Univariate



➢Insights Drawn

- Applicants belonging to the "Adults" category (31-65 years of Age) have the highest tendency to default
- Applicants from the "Senior-Citizen" category (65+ years of Age) have the least number of default



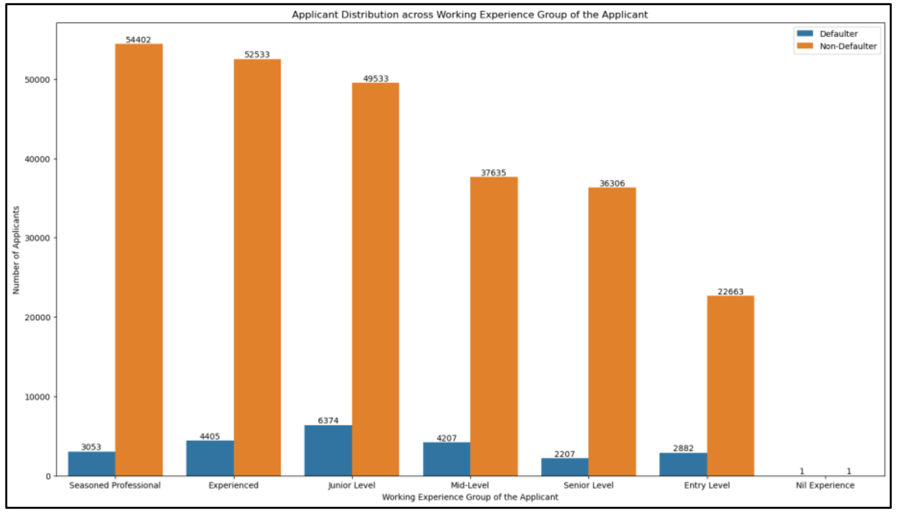
Applicant Distribution across Working Experience Group of the Applicant

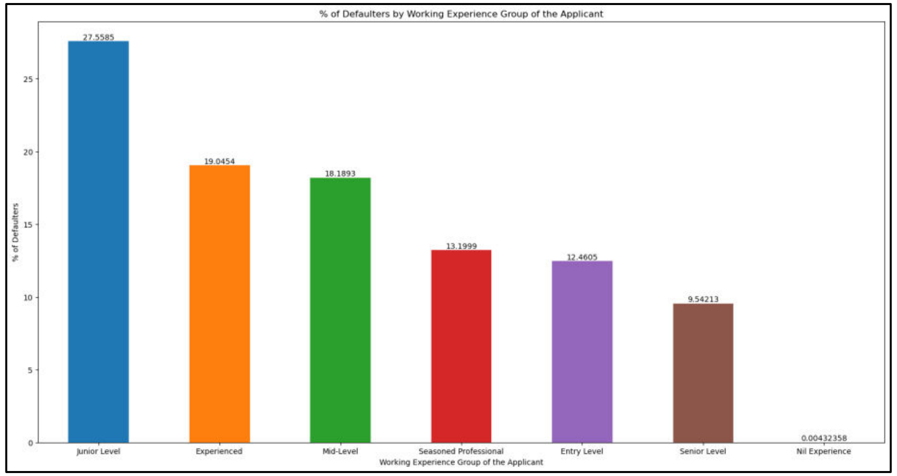
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “WORK_EXPERIENCE_GROUP”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants belonging to the "Junior Level" category (2-3 years of Work-Experience) have the highest tendency to default
- Applicants from the "Senior Level" category (10-25 years of Work Experience) have the second-lowest number of defaults





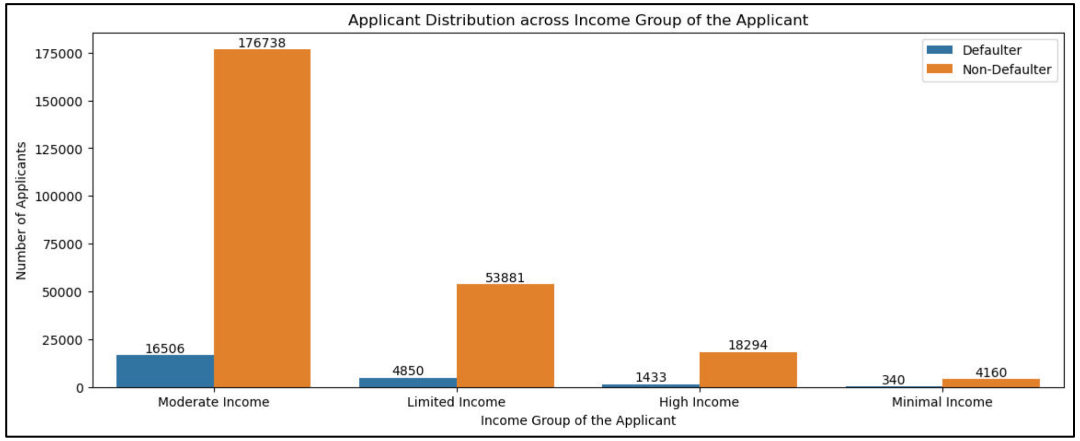
Applicant Distribution across Income Group of the Applicant

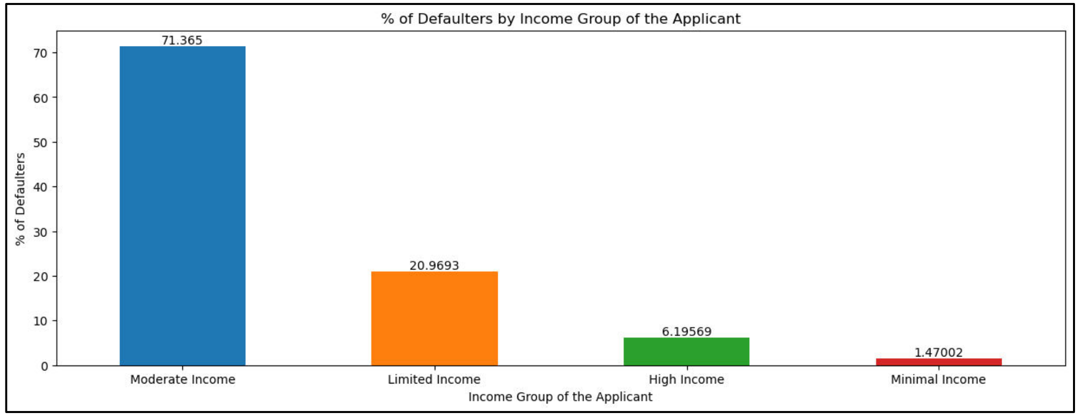
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "APPLICANT_INCOME_GROUP"
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants belonging to the "Moderate Income" category (1-2.5 Lakh Income Bucket) have the highest tendency to default
- Applicants from the "Minimal Income" category (0-0.5 Lakh Income Bucket) have the least number of defaults





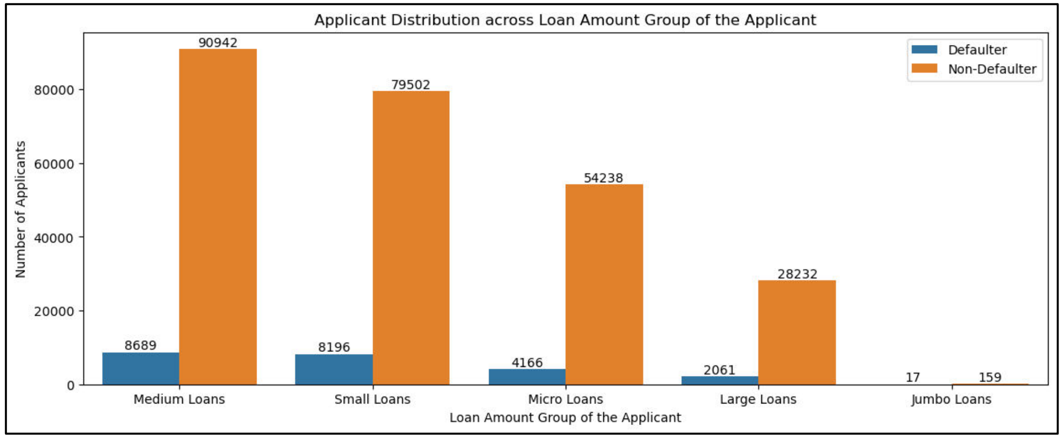
Applicant Distribution across Loan Amount Group of the Applicant

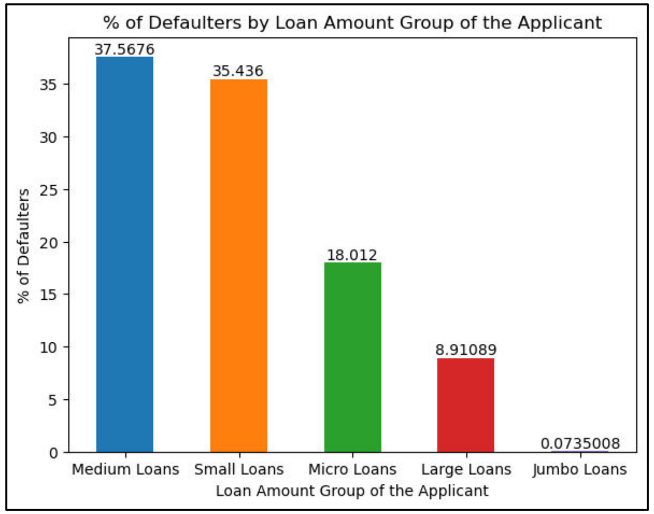
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "LOAN_AMT_GROUP"
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants belonging to the "Medium loans" category (6-10 Lakhs Loan Amount) have the highest tendency to default
- Applicants from the "Jumbo Loans" category (16+ Lakhs Loam Amount) have the least number of defaults





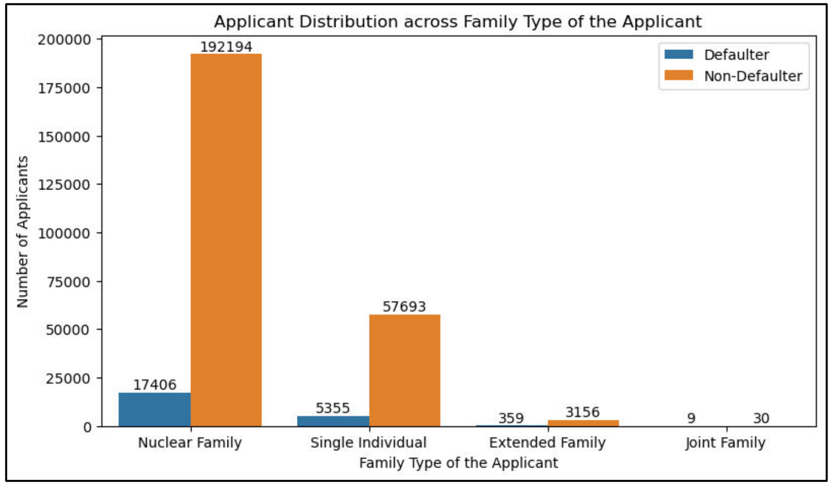
Applicant Distribution across Family Type of the Applicant

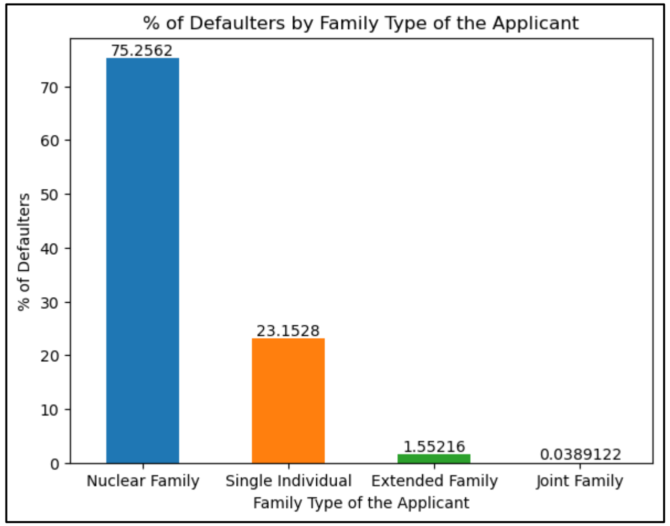
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “FAMILY_TYPE”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants belonging to the "Nuclear Family" category (2-4 Family Members) have the highest tendency to default
- Applicants from the "Joint Family" category (7+ Family Members) have the least number of defaults





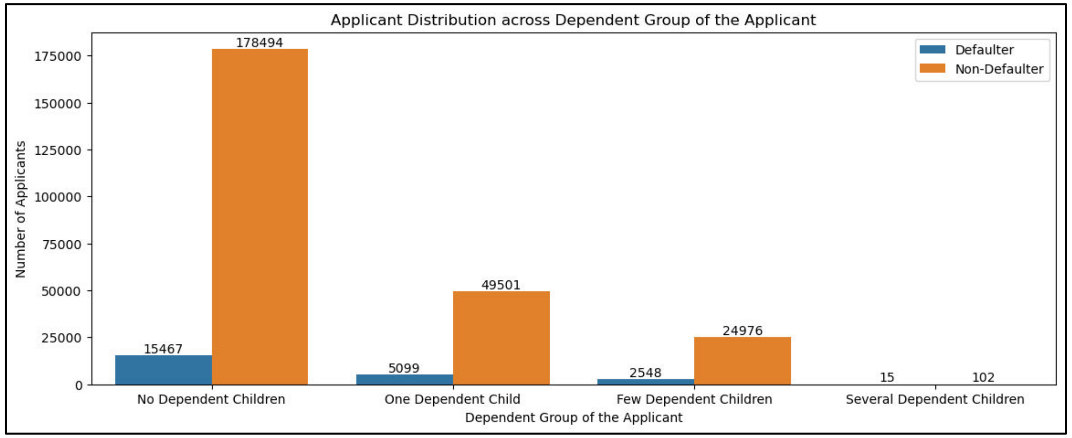
Applicant Distribution across Dependent Group of the Applicant

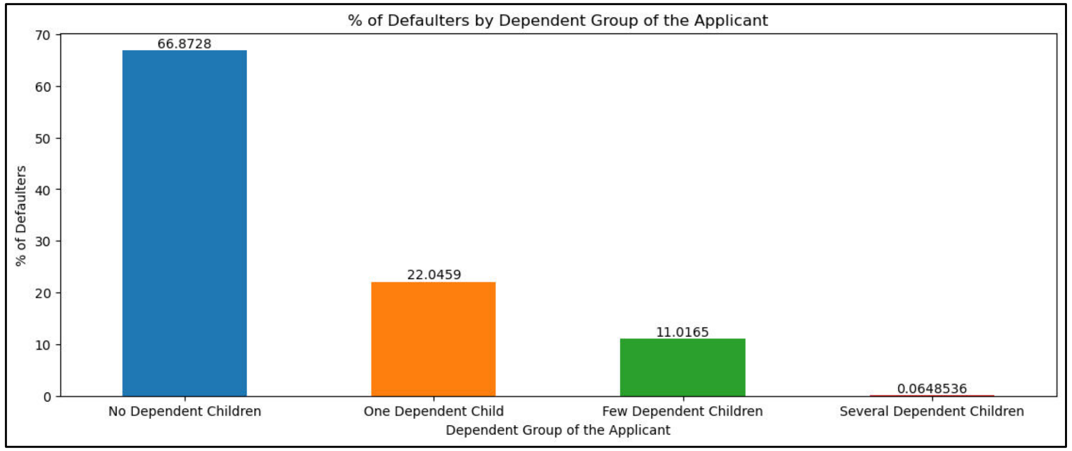
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “DEPENDENT_GROUP”
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants belonging to the "No Dependent Children" category (No children) have the highest tendency to default
- Applicants from the "Several Dependent Children" category (More than 4 Children) have the least number of defaults





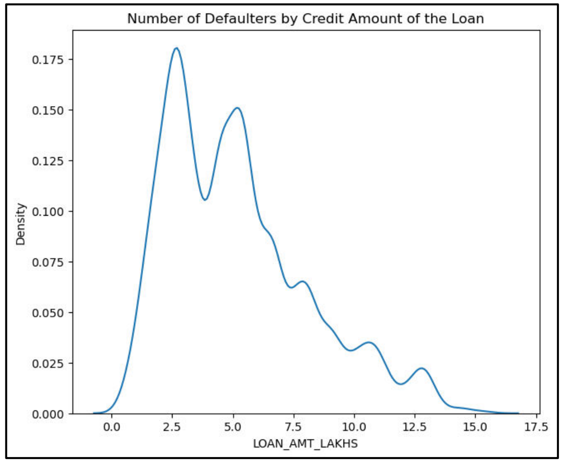
Number of Defaulters and Non-Defaulters by Credit Amount of the Loan

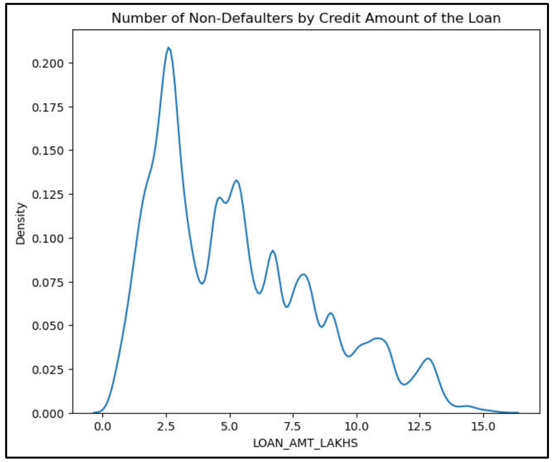
➢Analysis Done

- Number of Defaulters and Non-Defaulters by credit amount of the Loan
- Type of Analysis: Univariate

➢Insights Drawn

- Applicants taking loans b/w 1-2.5 Lakhs have the highest number of defaults





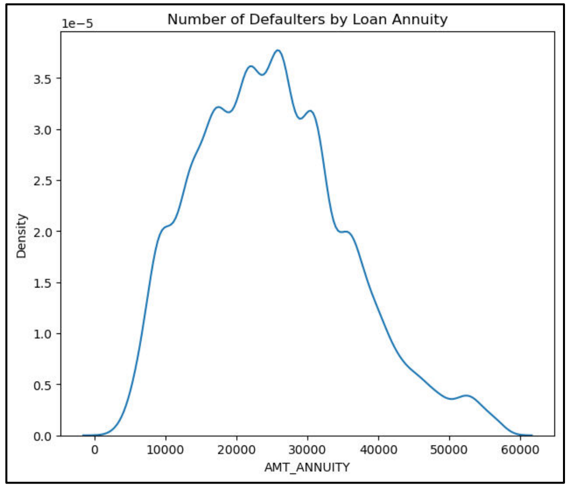
Number of Defaulters and Non-Defaulters by "AMT_ANNUITY"

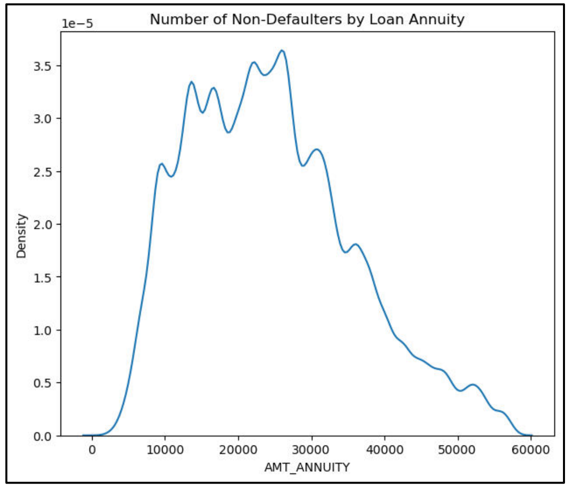
➢Analysis Done

- Number of Defaulters and Non-Defaulters by Loan Annuity
- Type of Analysis: Univariate

➢Insights Drawn

- Loans which have medium range of Loan Annuity amount (between 20k-30k), are the ones that have the highest loan defaults.

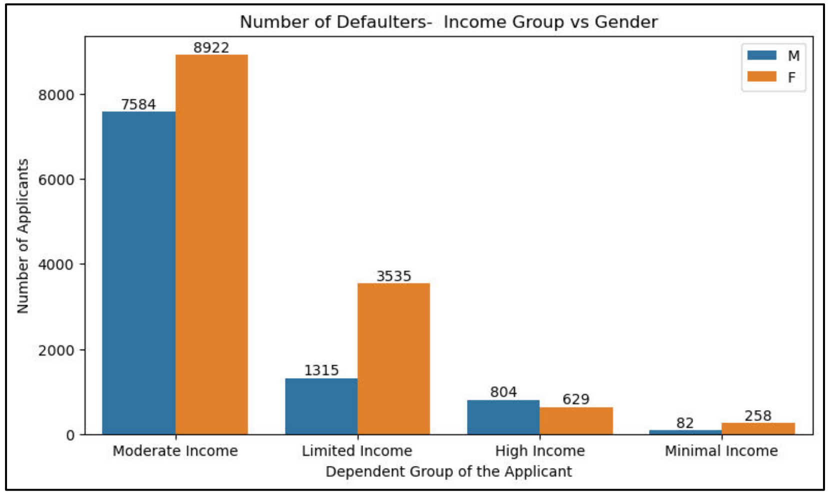




Number of Defaulters and Non-Defaulters- Income Group vs Gender

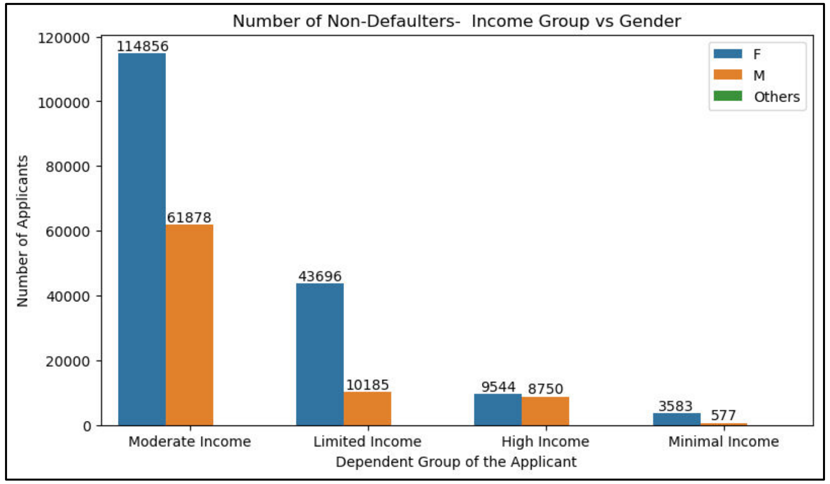
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of “DEPENDENT_GROUP” and “CODE_GENDER”
- Type of Analysis: Bivariate



➢Insights Drawn

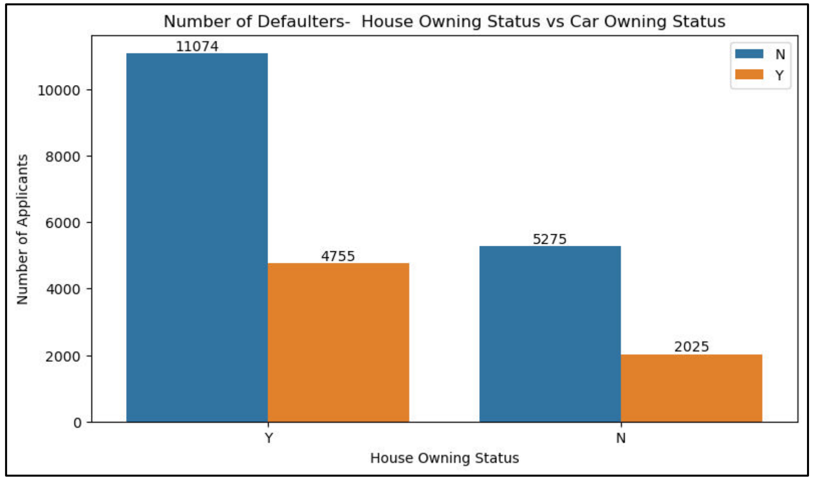
- Female applicants belonging to the "Moderate Income" group have the highest number of defaults



Number of Defaulters and Non-Defaulters- House Owning Status vs Car Owning Status

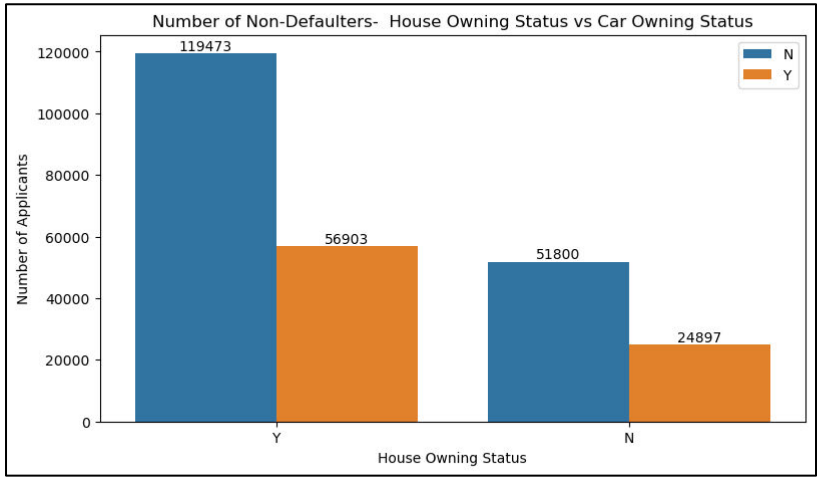
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "FLAG_OWN_REALTY" and "FLAG_OWN_CAR"
- Type of Analysis: Bivariate



➢Insights Drawn

- Applicants who own a house and don't own a car have the highest number of defaults



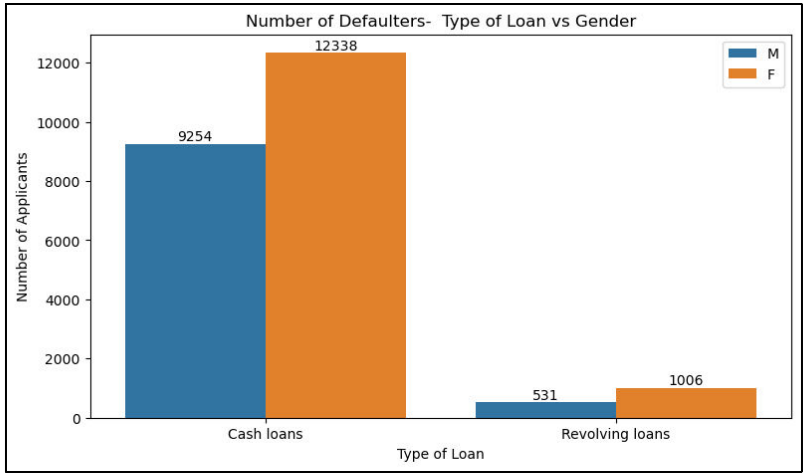
Number of Defaulters and Non-Defaulters- Type of Loan vs Gender

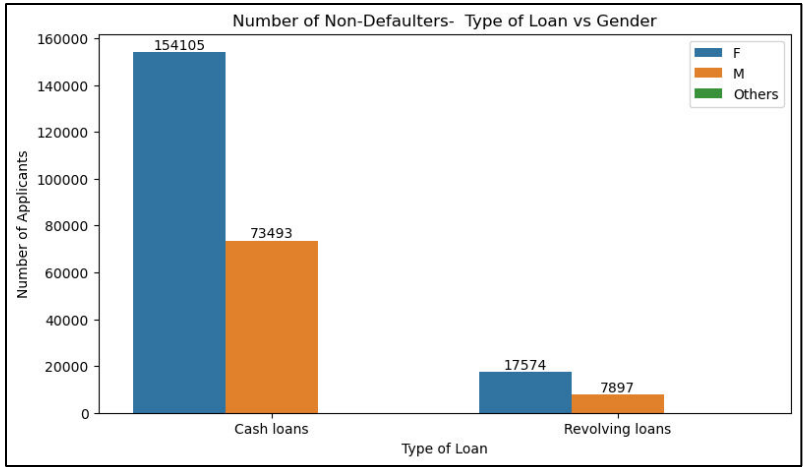
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "NAME_CONTRACT_TYPE" and "CODE_GENDER"
- Type of Analysis: Bivariate

➢Insights Drawn

- Female applicants who have taken "Cash Loans" have the highest number of defaults
- Cash Loans are predominant than Revolving Loans

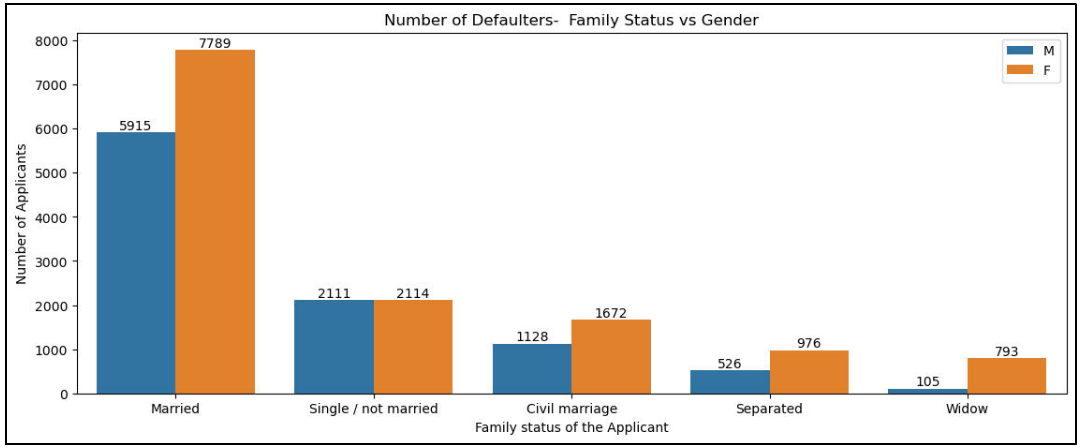




Number of Defaulters and Non-Defaulters- Family Status vs Gender

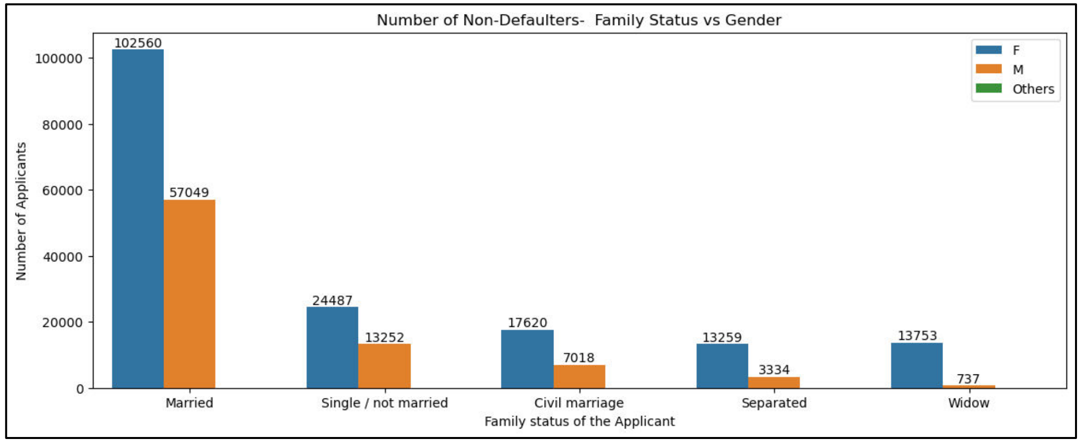
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "NAME_FAMILY_STATUS" and "CODE_GENDER"
- Type of Analysis: Bivariate



➢Insights Drawn

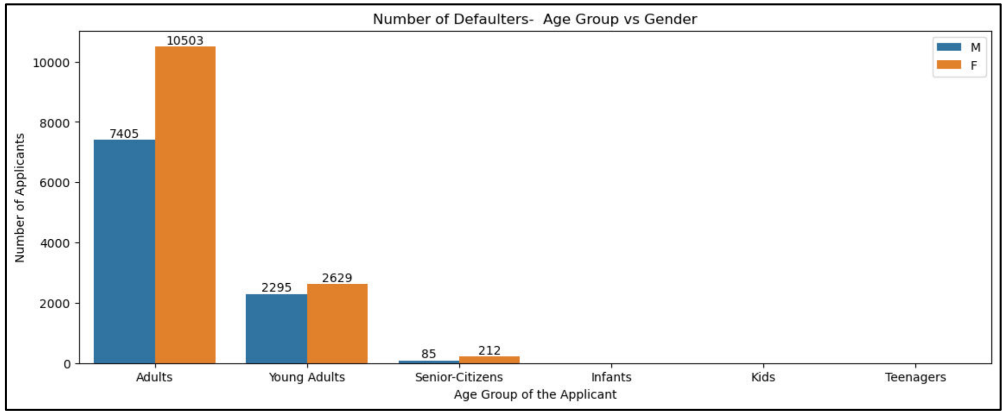
- Female applicants who are married have the highest number of defaults



Number of Defaulters and Non-Defaulters- Age Group vs Gender

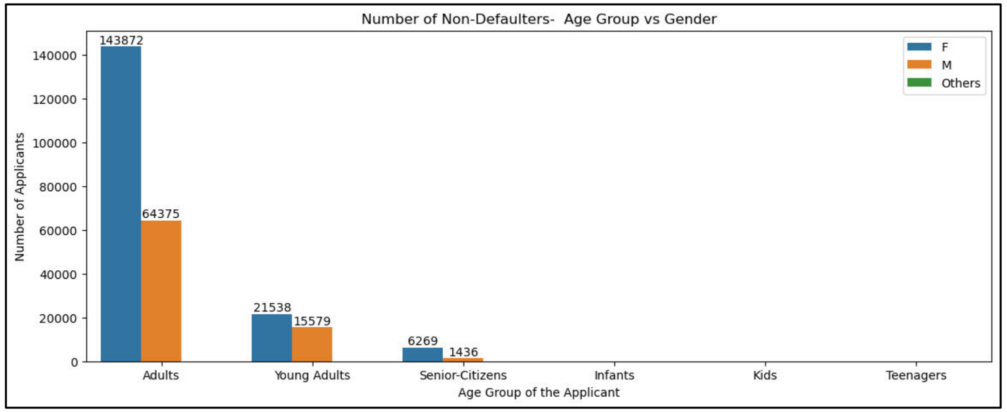
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "AGE_GROUP" and "CODE_GENDER"
- Type of Analysis: Bivariate



➢Insights Drawn

- Female adult applicants have the highest number of defaults



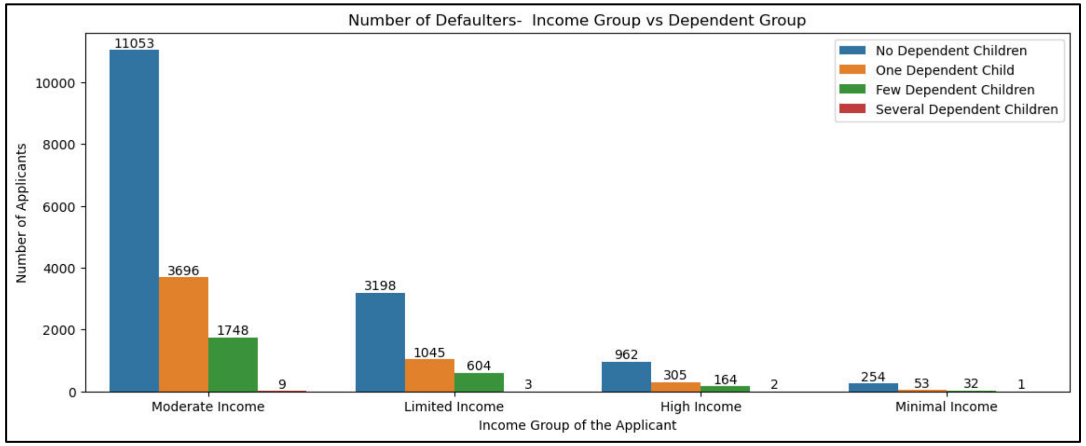
Number of Defaulters and Non-Defaulters- Income Group vs Dependent Group

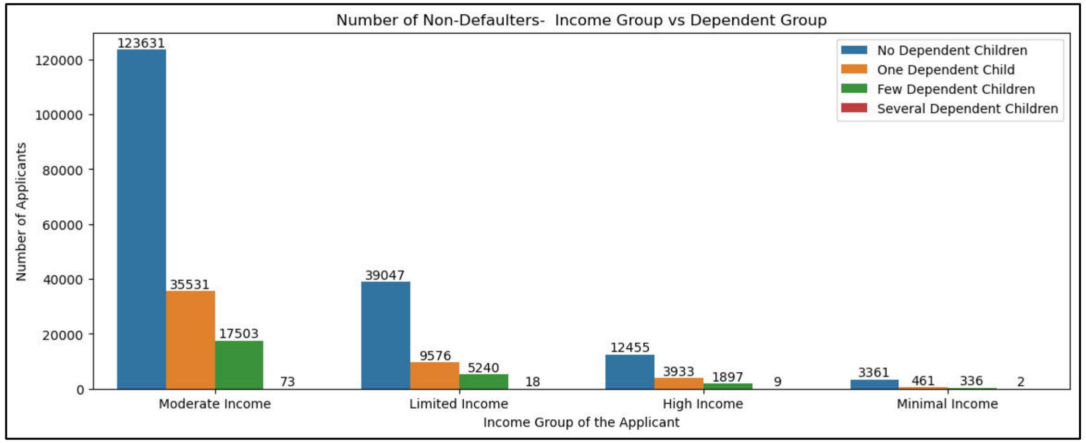
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "APPLICANT_INCOME_GROUP" and "DEPENDENT_GROUP"
- Type of Analysis: Bivariate

➢Insights Drawn

- Applicants belonging to the "Moderate Income" group and that don't have any dependent children have the highest numbers of defaults

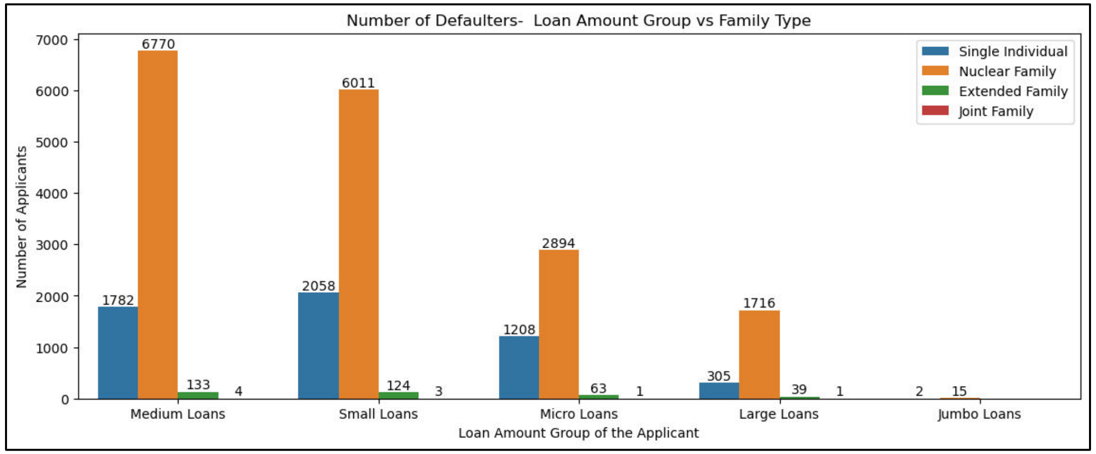




Number of Defaulters and Non-Defaulters- Loan Amount Group vs Family Type

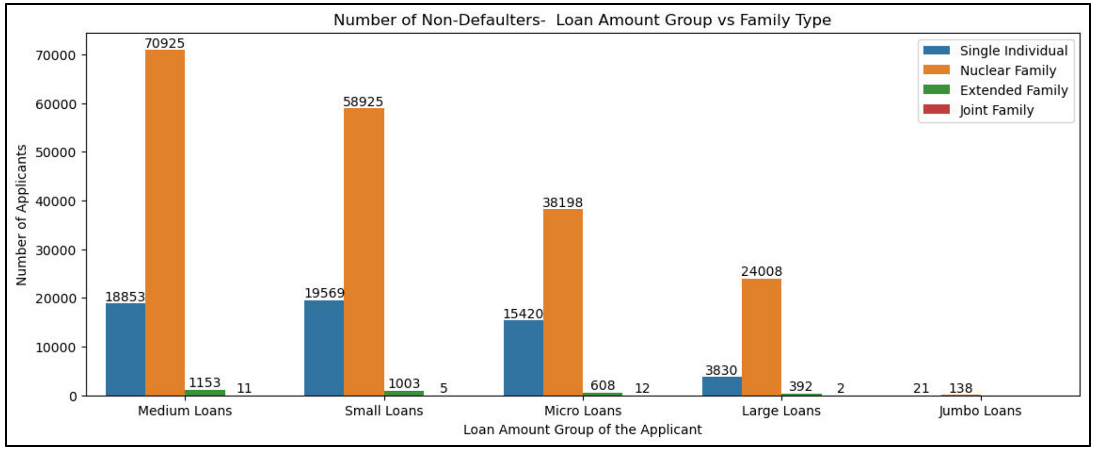
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "LOAN_AMT_GROUP" and "FAMILY_TYPE"
- Type of Analysis: Bivariate



➢Insights Drawn

- Single applicants belonging to the "Medium Loan" group have the highest numbers of defaults



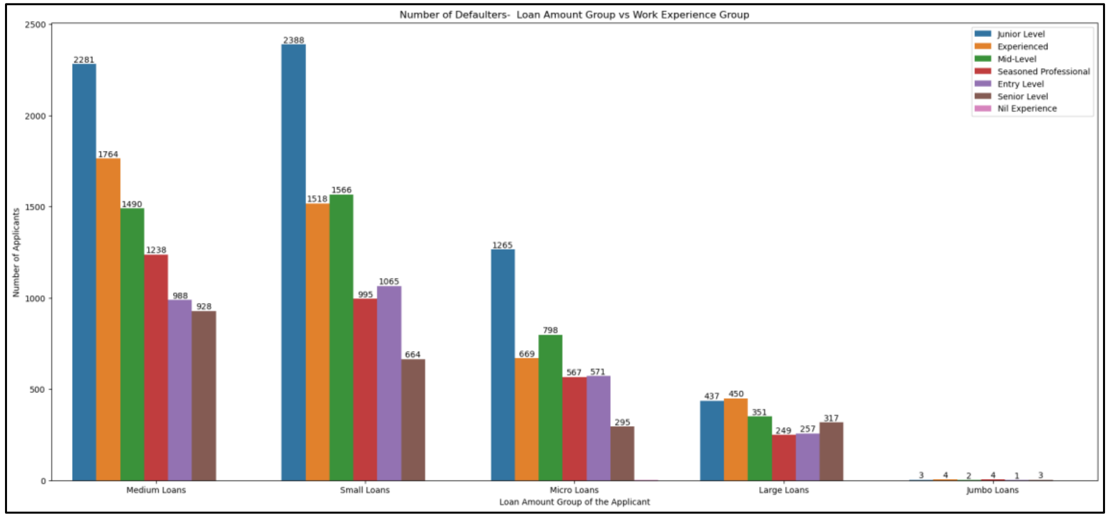
Number of Defaulters and Non-Defaulters-
Loan Amount Group vs Work Experience Group

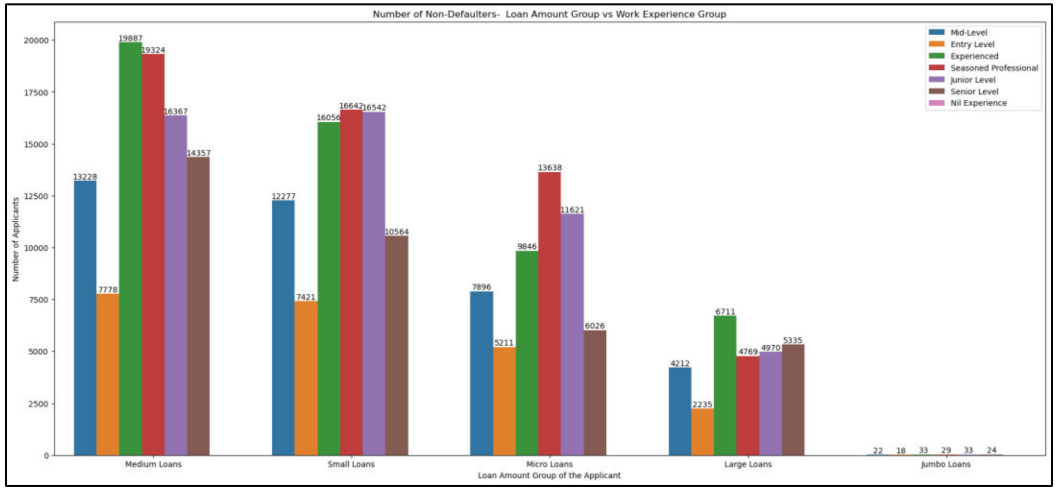
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "LOAN_AMT_GROUP" and "WORK_EXPERIENCE_GROUP"
- Type of Analysis: Bivariate

➢Insights Drawn

- Applicants belonging to the "Medium Loan" and "Entry Level" work experience group have the highest numbers of defaults





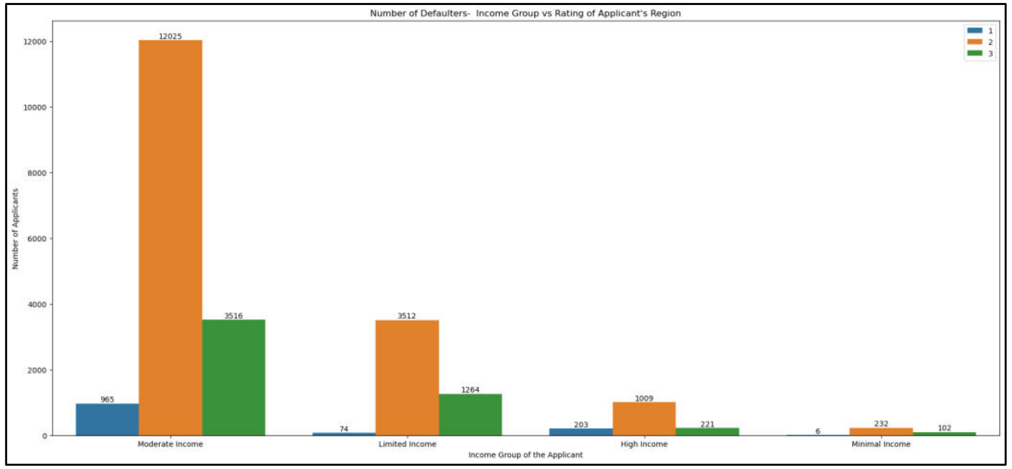
Number of Defaulters and Non-Defaulters- Income Group vs Rating of Applicant's Region

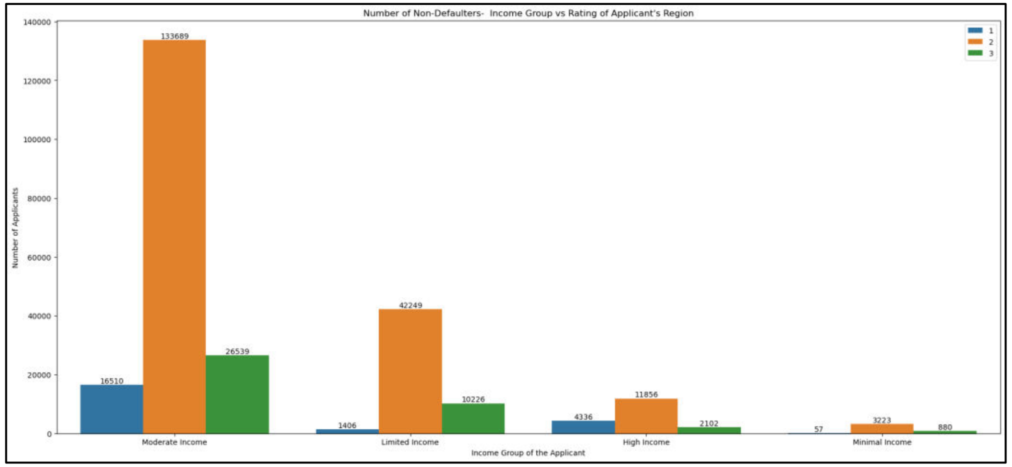
➢Analysis Done

- Plotting two countplots for finding out "Defaulters" and "Non-Defaulters" on the basis of "APPLICANT_INCOME_GROUP" and "REGION_RATING_CLIENT"
- Type of Analysis: Bivariate

➢Insights Drawn

- Applicants belonging to the "Moderate Income" and "Rating 2" region group have the highest numbers of defaults





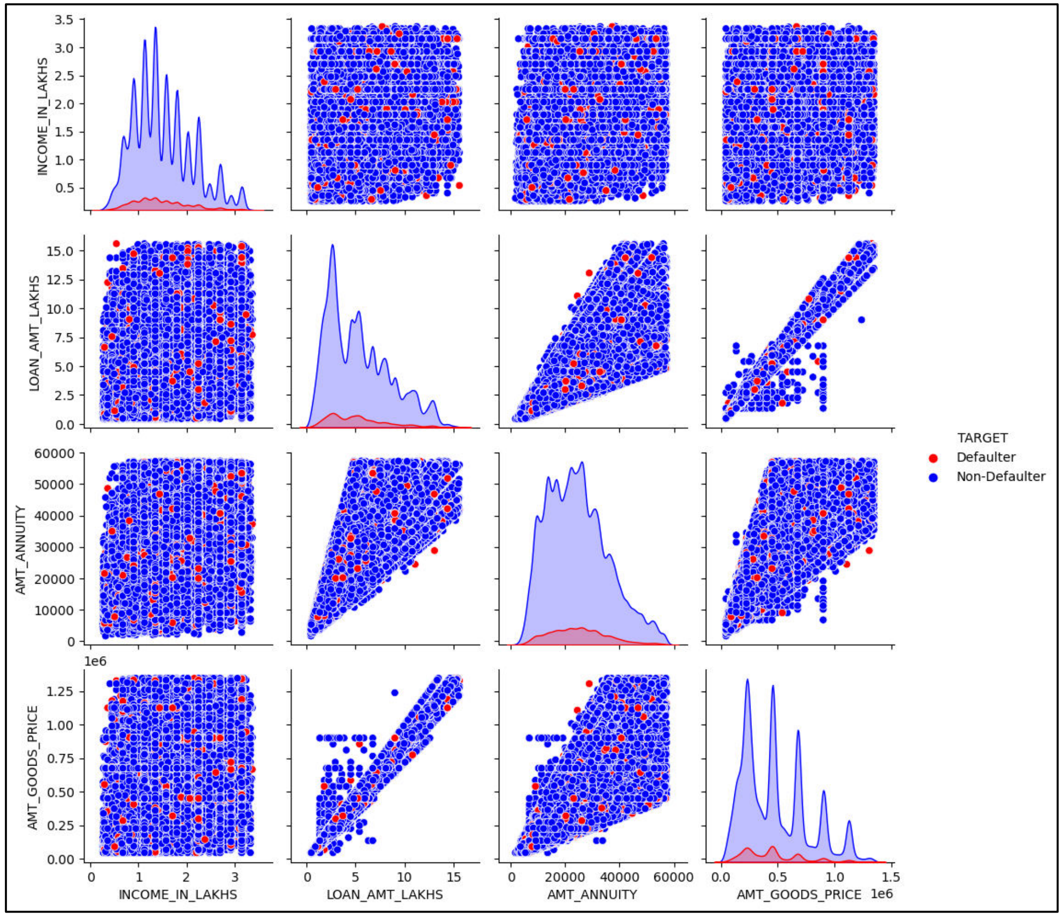
Number of Defaulters and Non-Defaulters- Multiple Columns

➢Analysis Done

- Plotting pair plots for finding out "Defaulters" and "Non-Defaulters" on the basis of 'INCOME_IN_LAKHS’ , 'LOAN_AMT_LAKHS’, 'AMT_ANNUITY’, 'AMT_GOODS_PRICE', 'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'TARGET’
- Type of Analysis: Multivariate

➢Insights Drawn

- 'LOAN_AMT_LAKHS’ and 'AMT_GOODS_PRICE’ show a linear relation
- 'LOAN_AMT_LAKHS’ and 'AMT_ANNUITY’ show a linear relation



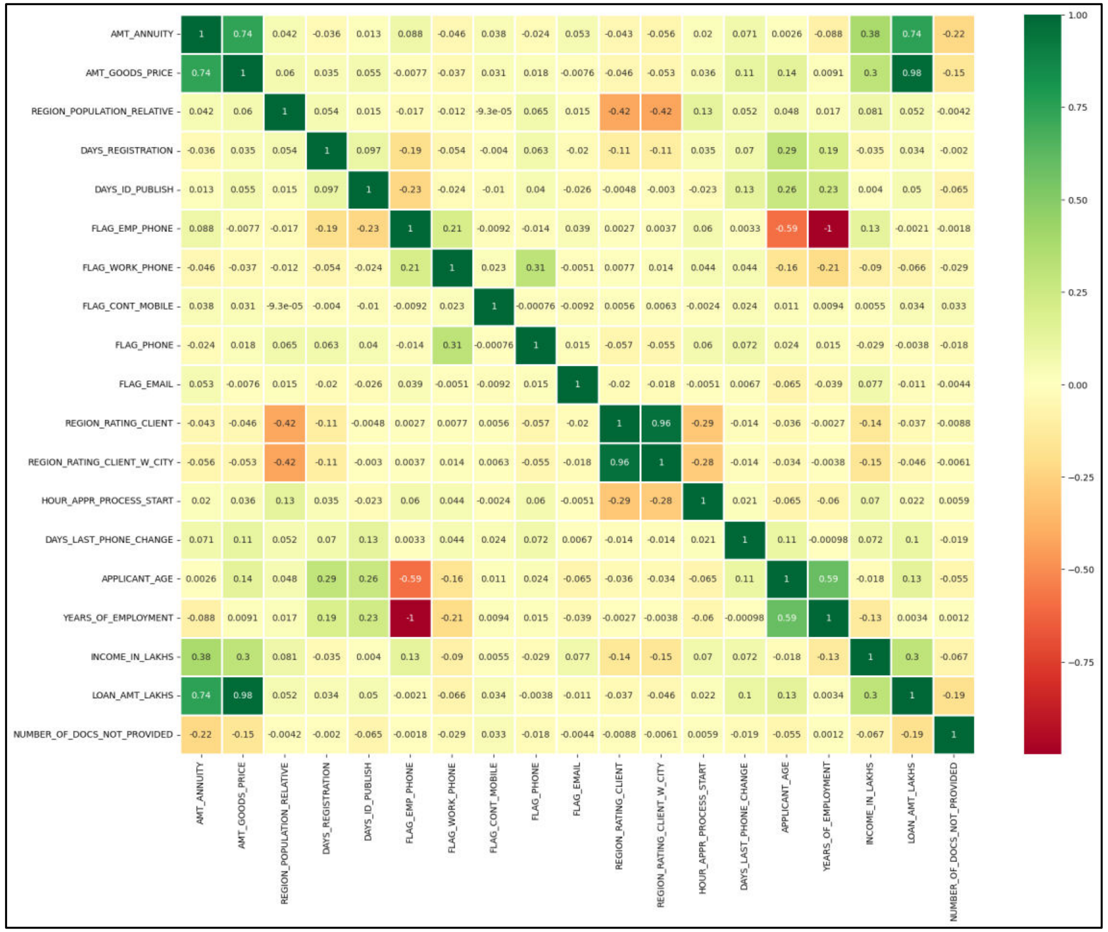
Number of Defaulters- Heatmap

➢Analysis Done

- Plotting a heatmap **<u>only</u>** for the “defaulters” dataset
- Type of Analysis: Multivariate

➢Insights Drawn

- “REGION_RATING_CLIENT_W_CITY” and “REGION_RATING_CLIENT” show a positive correlation
- “APPLICANT_AGE” and “FLAG_EMP_PHONE” show a negative correlation



EDA of “previous_application” dataset

- Following are the common EDA steps used for the above dataset-
  - Removing the columns that aren’t required
  - Null value treatment
  - Data sanity checks
  - Outlier handling
  - Creation of new categorical and numerical variables
  - Performing Univariate, Bivariate and Multivariate Analysis

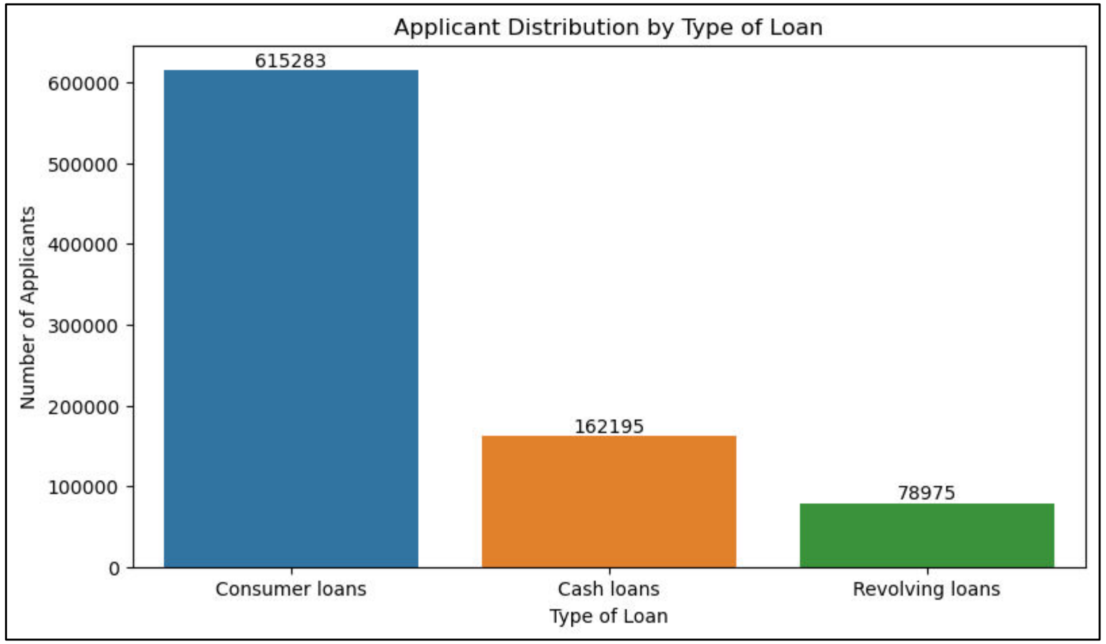
Applicant Distribution by Type of Loan

➢Analysis Done

- Plotting a countplot to show the distribution on the basis of “Type of Loan”
- Type of Analysis: Univariate

➢Insights Drawn

- Majority of the applicants have taken "Consumer Loans"



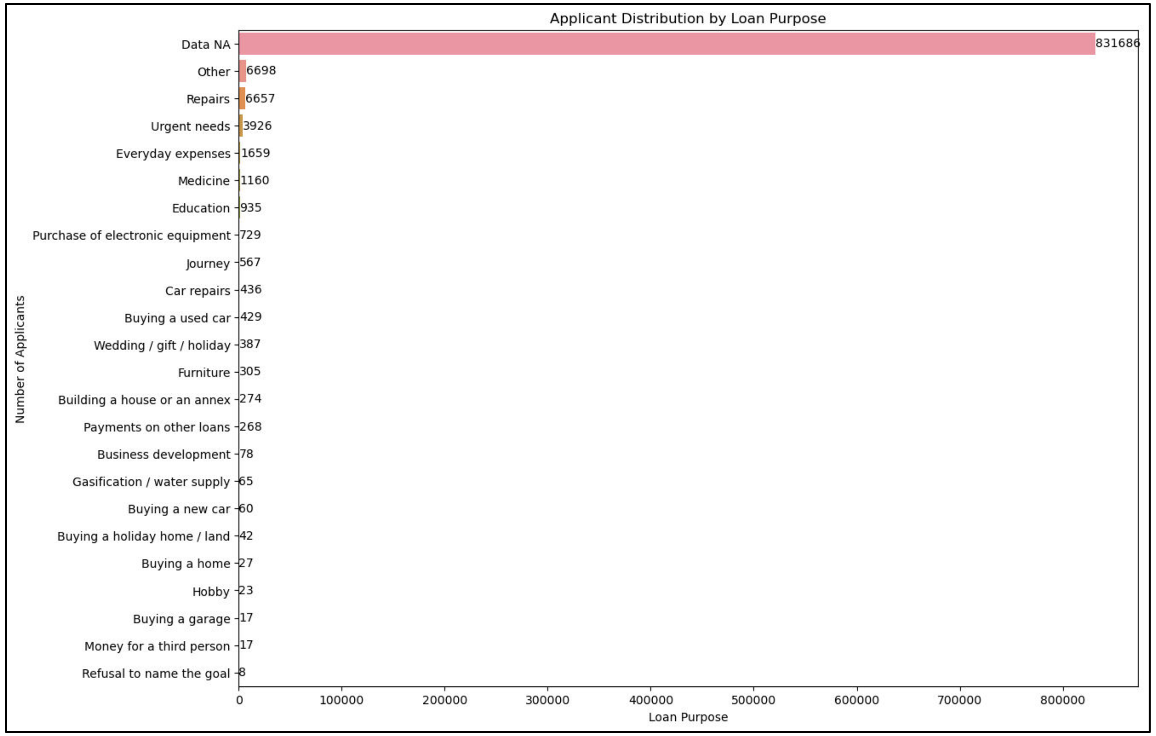
Applicant Distribution by Loan Purpose

➢Analysis Done

- Plotting the countplot to show the distribution on the basis of “Loan Purpose”
- Type of Analysis: Univariate

➢Insights Drawn

- After excluding the "Data NA" and "Other" values, "Repairs" is the main reason why people have taken loans



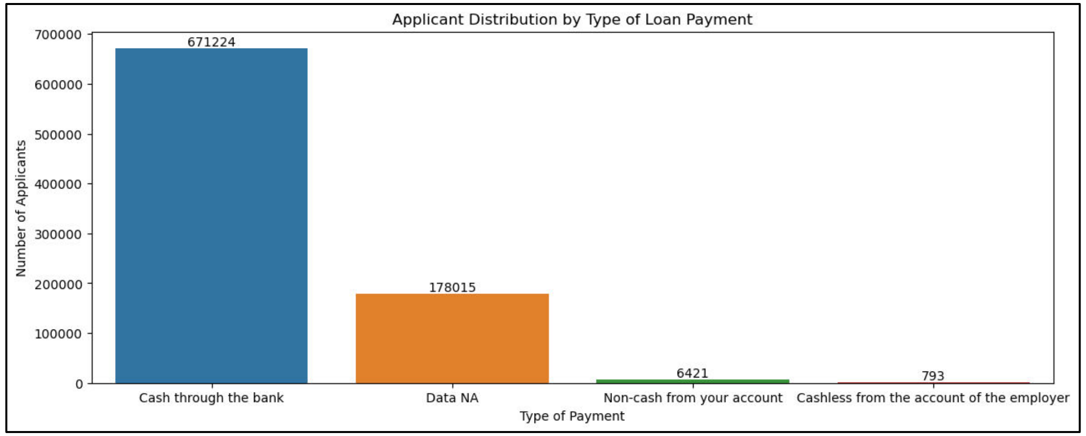
Applicant Distribution by Type of Loan Payment

➢Analysis Done

- Plotting the countplot to show the distribution on the basis of "Type of Payment" used by applicant to pay the loan
- Type of Analysis: Univariate

➢Insights Drawn

- Majority of the applicants used "Cash through the bank" to pay for their loan



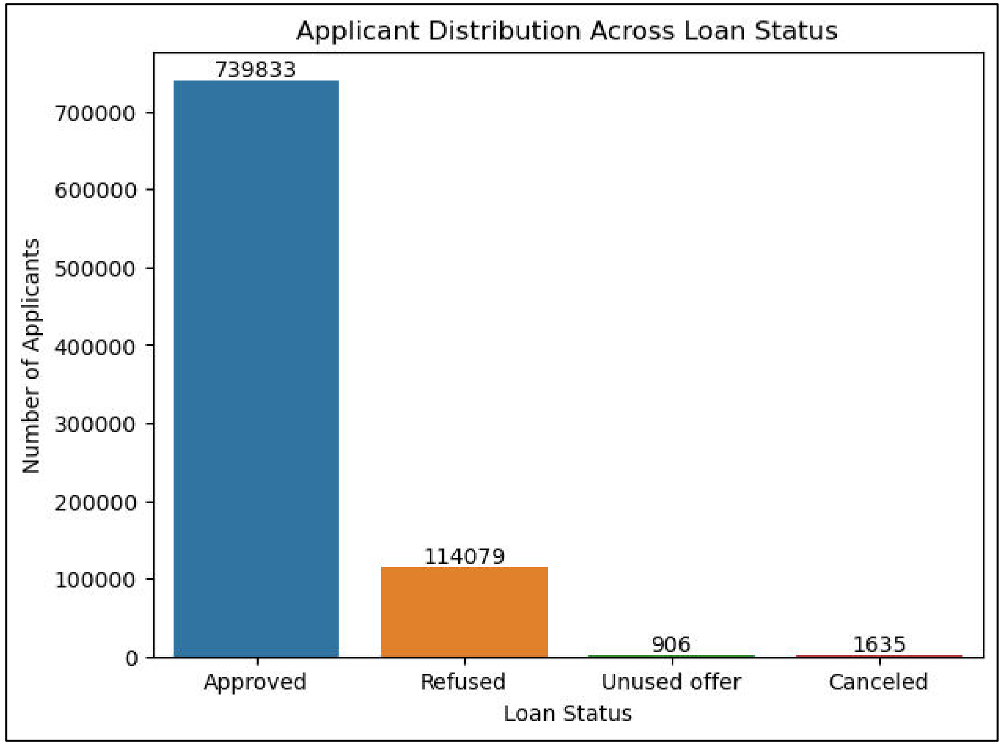
Applicant Distribution Across Loan Status

➢Analysis Done

- Plotting the Countplot to show the distribution on the basis of “Status of the Loan”
- Type of Analysis: Univariate

➢Insights Drawn

- Majority of the Loans are of the "Approved" status



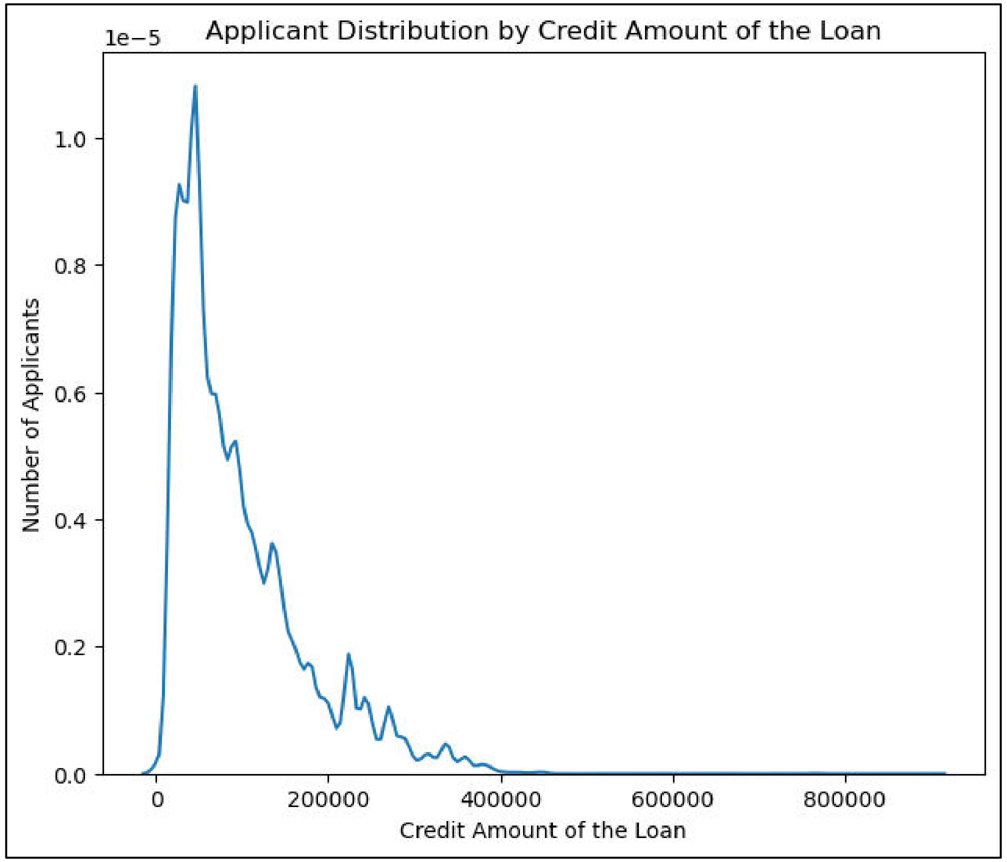
Applicant Distribution by Credit Amount of the Loan

➢Analysis Done

- Plotting distplot for Defaulters on the basis of "AMT_CREDIT"
- Type of Analysis: Univariate

➢Insights Drawn

- Majority of the loans are between 0-2 Lakhs



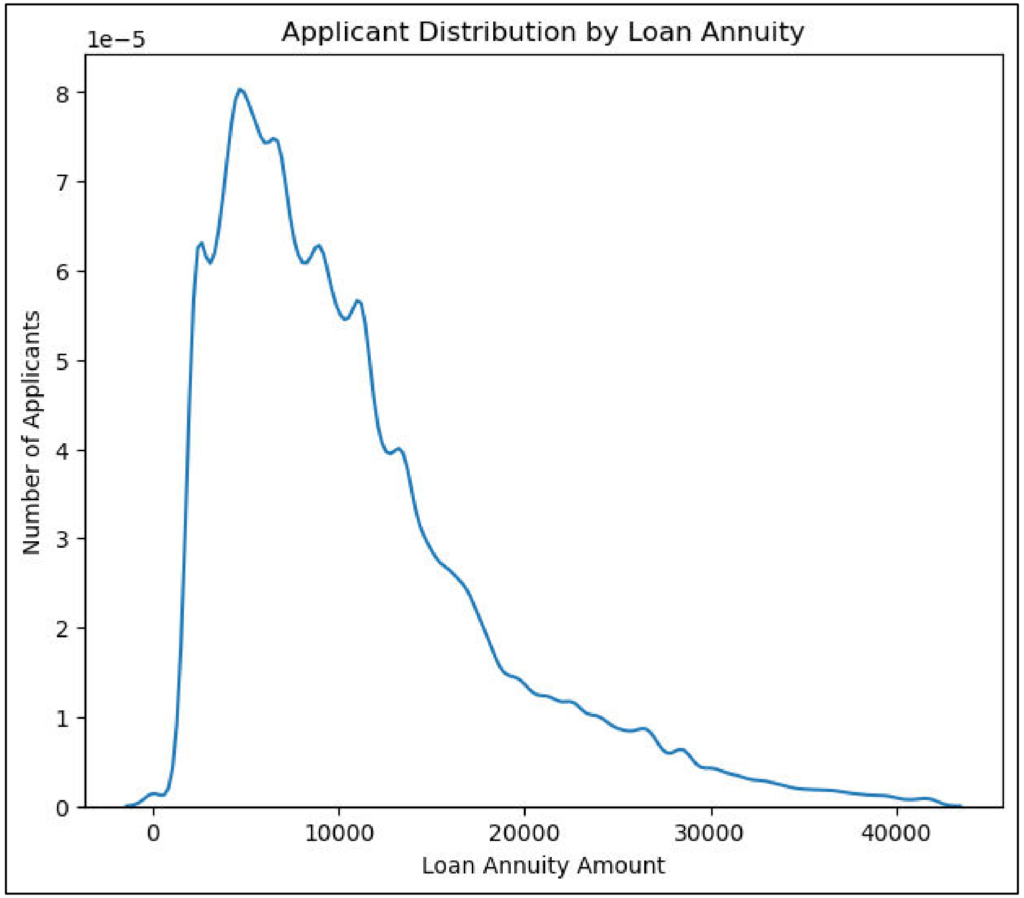
Applicant Distribution by Loan Annuity

➢Analysis Done

- Plotting Distplot for Defaulters on the basis of "AMT_ANNUITY"
- Type of Analysis: Univariate

➢Insights Drawn

- Maximum number of loans have an annuity amount between 0-10k



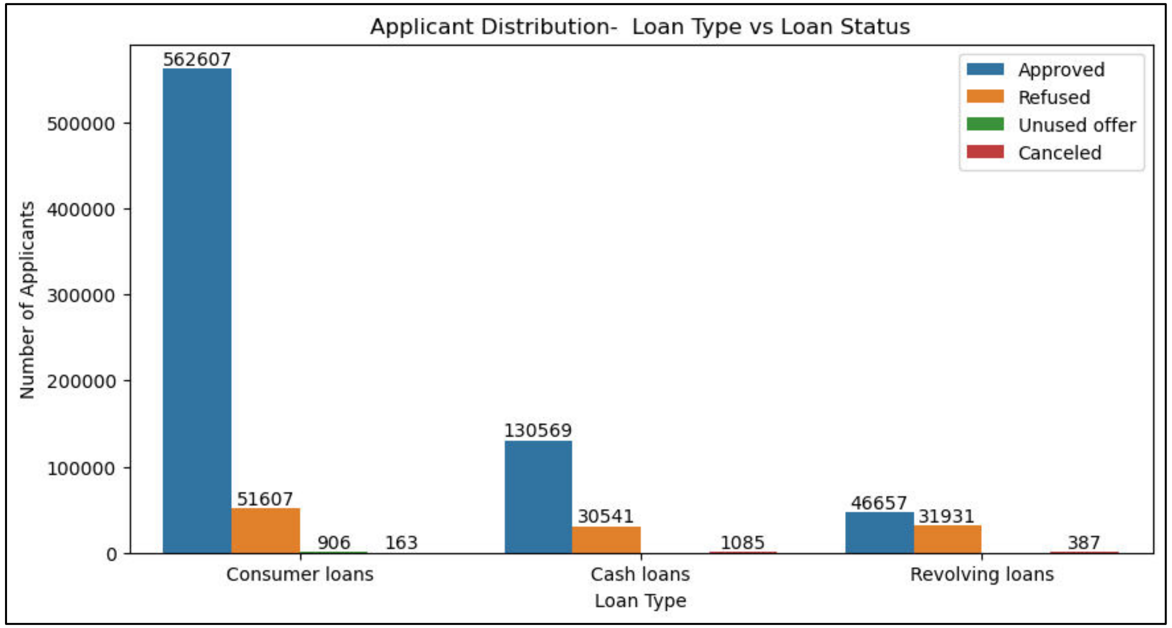
Applicant Distribution- Loan Type vs Loan Status

➢Analysis Done

- Plotting countplot for finding out distribution on the basis of "NAME_CONTRACT_TYPE" and “NAME_CONTRACT_STATUS”
- Type of Analysis: Bivariate

➢Insights Drawn

- "Consumer Loans" are the most preferred type of loans
- Majority of the loans come under the "Approved" status



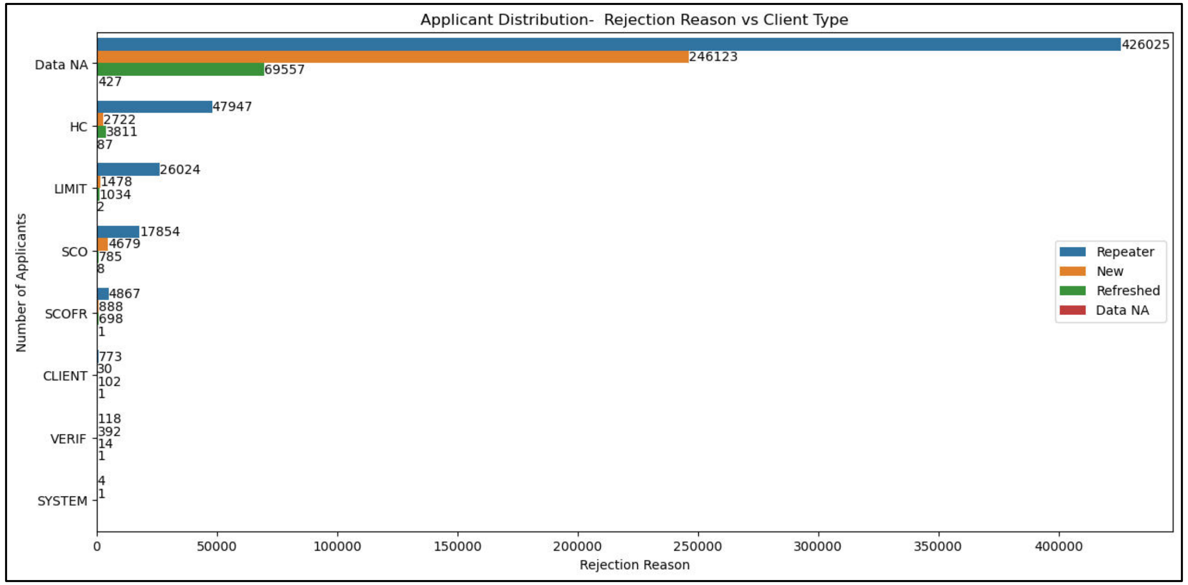
Applicant Distribution- Rejection Reason vs Client Type

➢Analysis Done

- Plotting countplot for finding out distribution on the basis of "CODE_REJECT_REASON" and “NAME_CLIENT_TYPE”
- Type of Analysis: Bivariate

➢Insights Drawn

- Excluding the "Data NA" values, "HC" and "LIMIT" are the major reasons behind loans not being approved
- "Repeaters" have the maximum number of loan rejections



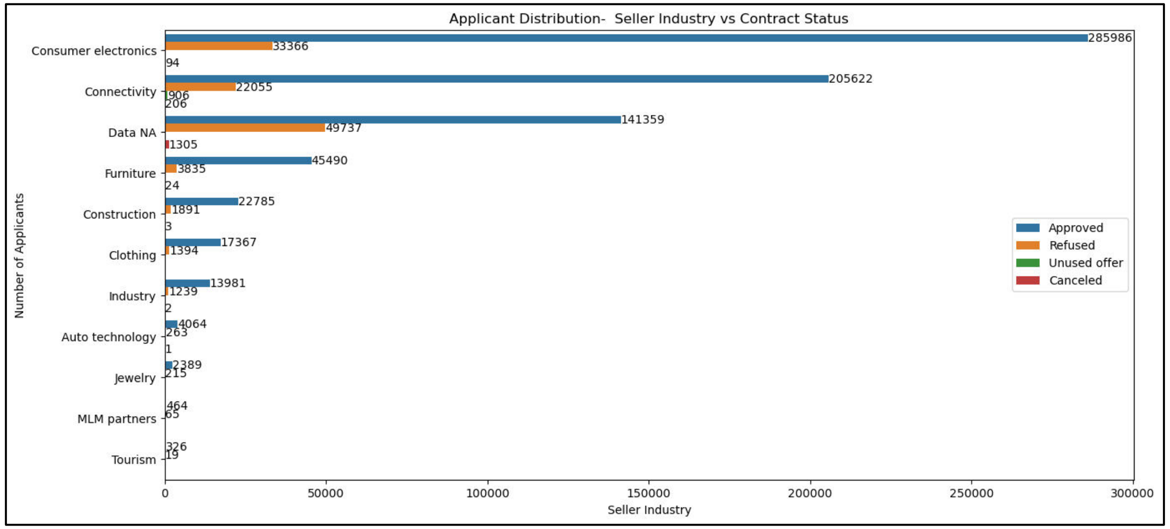
Applicant Distribution- Seller Industry vs Contract Status

➢Analysis Done

- Plotting countplot for finding out distribution on the basis of "NAME_SELLER_INDUSTRY" and “NAME_CONTRACT_STATUS”
- Type of Analysis: Bivariate

➢Insights Drawn

- Sellers from the "Consumer Electronics" industry have provided and approved the maximum number of loans



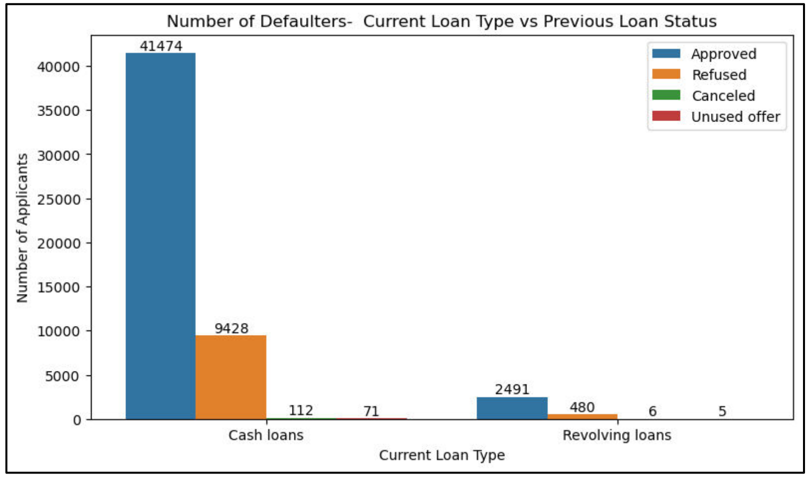
EDA after merging both the datasets

- Following are the common EDA steps used-
  - Removing the columns that aren’t required
  - Null value treatment
  - Data sanity checks
  - Outlier handling
  - Creation of new categorical and numerical variables
  - Performing Univariate, Bivariate and Multivariate Analysis

Number of Defaulters and Non-Defaulters- Current Loan Type vs Previous Loan Status

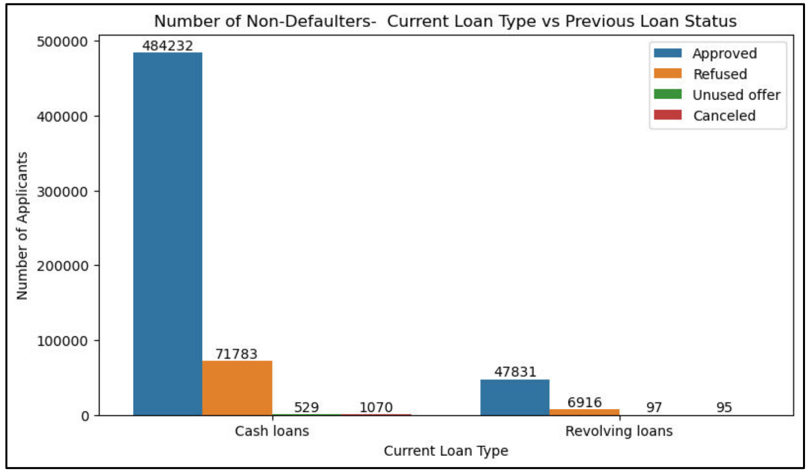
➢Analysis Done

- Plotting countplot for finding out of all active loan application, how many loans were rejected previously
- Type of Analysis: Bivariate



➢Insights Drawn

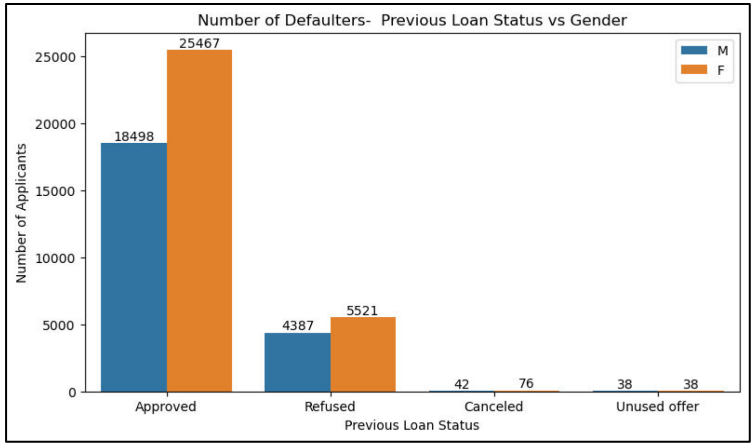
- Maximum number of applicant who have currently applied for "Cash Loans" were refused a loan earlier
- Number of defaulters are highest for "Cash Loans"



Number of Defaulters and Non-Defaulters- Previous Loan Status vs Gender

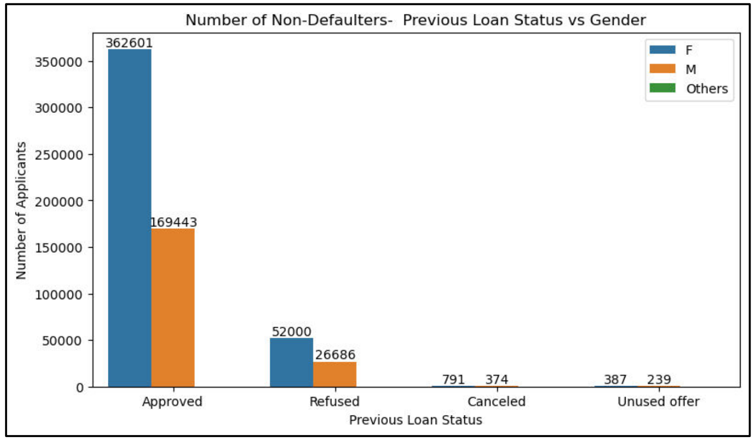
➢Analysis Done

- Plotting countplot for finding out previous loan status on the basis of Gender
- Type of Analysis: Bivariate



➢Insights Drawn

- After getting their loan approved, Female applicants have defaulted the most



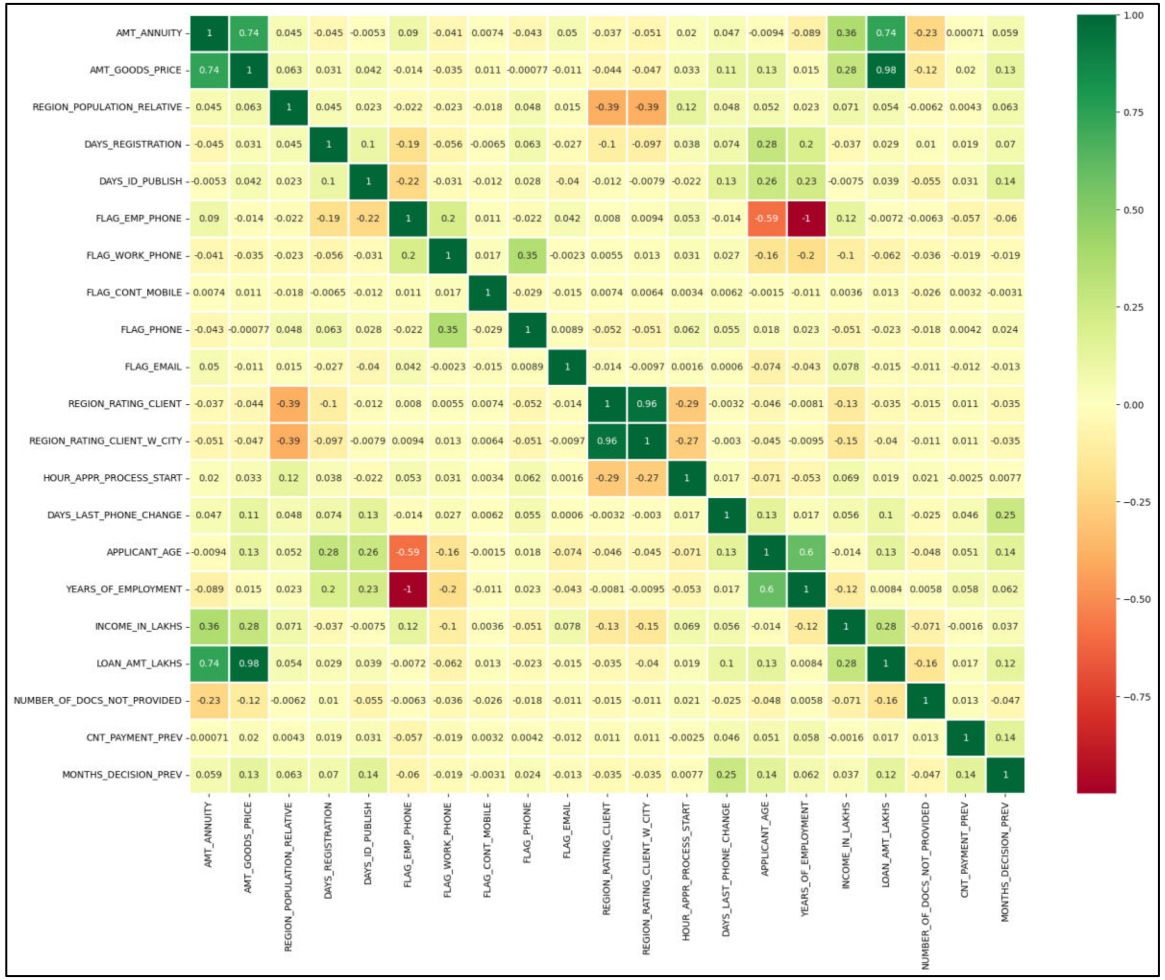
Number of Defaulters and Non-Defaulters- Previous Loan Status vs Gender

➢Analysis Done

- Plotting heatmap to see linear correlation between the applicants who are "Defaulters“
- Type of Analysis: Multivariate

➢Insights Drawn

- "AMT_GOODS_PRICE" is positively correlated to "LOAN_AMT_LAKHS“
- "AMT_ANNUITY" is positively correlated to "LOAN_AMT_LAKHS“
- "FLAG_EMP_PHONE" is negatively correlated to "YEARS_OF_EMPLOYMENT" and "APPLICANT_AGE“
- "HOURS_APPR_PROCESS_START" shows negative correlation with "REGION_RATING_CLIENT" and "REGION_RATING_CLIENT_W_CITY"



Conclusions

- Factors that are an indicator that an applicant is likely to be a **<u>“Defaulter”</u>**:
  - **NAME_CONTRACT_TYPE**: Applicants opting for “Cash Loans” are more likely to default
  - **NAME_EDUCATION_TYPE**: Applicants with "Secondary/Secondary Special" educational background have the highest tendency to default
  - **CODE_GENDER**: Male applicants have higher default rates than women and “Others”
  - **FLAG_OWN_CAR**: Applicants who don't own a Vehicle have a higher tendency of being a defaulter
  - **NAME_HOUSING_TYPE**: Applicants who live in a "House/Apartment" have a higher tendency of being a defaulter
  - **OCCUPATION_TYPE**: "Laborers" have the second highest tendency to default
  - **EMPLOYEE_JOB_TYPE**: Applicants who are working from home city have a higher tendency to default
  - **NAME_INCOME_TYPE**: Applicants who are working have the highest tendency to default
  - **NAME_FAMILY_STATUS**: Applicants who are Married have the highest tendency to default
  - **AGE_GROUP**: Applicants belonging to the "Adults" category (31-65 years of Age) have the highest tendency to default
  - **WORK_EXPERIENCE_GROUP**: Applicants belonging to the "Junior Level" category (2-3 years of Work-Experience) have the highest tendency to default
  - **APPLICANT_INCOME_GROUP**: Applicants belonging to the "Moderate Income" category (1-2.5 Lakh Income Bucket) have the highest tendency to default
  - **LOAN_AMT_GROUP**: Applicants belonging to the "Medium loans" category (6-10 Lakhs Loan Amount) have the highest tendency to default
  - **FAMILY_TYPE**: Applicants belonging to the "Nuclear Family" category (2-4 Family Members) have the highest tendency to default
  - **DEPENDENT_GROUP**: Applicants belonging to the "No Dependent Children" category (No children) have the highest tendency to default
  - **LOAN_AMT_LAKHS**: Applicants taking loans b/w 1-2.5 Lakhs have the highest number of defaults
  - **AMT_ANNUITY**: Loans which have medium range of Loan Annuity amount (between 20k-30k), are the ones that have the highest loan defaults

- Factors that are an indicator that an applicant is likely to be a **<u>“Non-Defaulter”</u>**:
  - **NAME_CONTRACT_TYPE**: Applicants opting for “Revolving Loans” are less likely to default
  - **NAME_EDUCATION_TYPE**: Applicants who hold an academic degree have made no loan defaults till date
  - **CODE_GENDER**: Female applicants have lesser overall default rates than Men
  - **FLAG_OWN_CAR**: Applicants who own a Vehicle are less likely of being a defaulter
  - **NAME_HOUSING_TYPE**: Applicants who live in "Co-operate Apartments" are less likely to default
  - **OCCUPATION_TYPE**: Working applicants from the "Human Resources" and "IT Industry" are least likely to default their loan payment
  - **NAME_INCOME_TYPE**: Applicants who are on “Maternity Leave” / “State Servant” have the least tendency to default
  - **AGE_GROUP**: Applicants from the "Senior-Citizen" category (65+ years of Age) have the least number of default
  - **WORK_EXPERIENCE_GROUP**: Applicants from the "Senior Level" category (10-25 years of Work Experience) have the second-lowest number of defaults
  - **APPLICANT_INCOME_GROUP**: Applicants from the "Minimal Income" category (0-0.5 Lakh Income Bucket) have the least number of defaults
  - **LOAN_AMT_GROUP**: Applicants from the "Jumbo Loans" category (16+ Lakhs Loam Amount) have the least number of defaults
  - **FAMILY_TYPE**: Applicants from the "Joint Family" category (7+ Family Members) have the least number of defaults
  - **DEPENDENT_GROUP**: Applicants from the "Several Dependent Children" category (More than 4 Children) have the least number of defaults

Recommendations

❖ In order to have less number of defaulters, target the applicants who have the below attributes:

- Female applicants
- Applicants who have chosen “Revolving Loans”
- Applicants having an academic degree
- Applicants living in a region whose rating by the client is “1”
- Applicants who own a vehicle but don’t own a house
- Applicants who are married and have more than 2 children
- Applicants who live in a joint family
- Working professionals from the “HR” or “IT” industry who have at least 10+ years of working experience
- Prefer applicants who are older (65+ Years of age)
- Applicants who earn between 0-1.5 Lakhs
- Applicants who have opted for “Jumbo Loans” (16+ lakhs loan amount)