EDA Credit Assignment

Brindha varadharajan

Problem Statement

This assignment tackles the hurdle of lending to individuals with limited credit history by leveraging Exploratory Data Analysis (EDA).

The objective is to pinpoint patterns revealing client payment challenges in a consumer finance company.

By comprehending the impact of consumer and loan attributes on default tendencies, the aim is to enhance **decision-making, ensuring fair treatment** for **creditworthy applicants** and minimizing the **risk of financial losses.**

Objective

Utilize Exploratory Data Analysis (EDA) to analyze loan application data for a consumer finance company.

***Challenge*:**

Address difficulties in lending, especially for applicants with insufficient credit history.

Mitigate the risk of unfair rejections for creditworthy applicants and financial losses from approving likely defaulters.

***Dataset Overview*:**

Application & Previous applicants information provided in CSV format.

***Analysis Focus*:**

Use EDA techniques to identify patterns influencing loan default tendencies.

Understand the impact of consumer attributes and loan attributes on default probabilities.

***Outcome*:**

Facilitate data-driven and risk-mitigated lending decisions for the company.

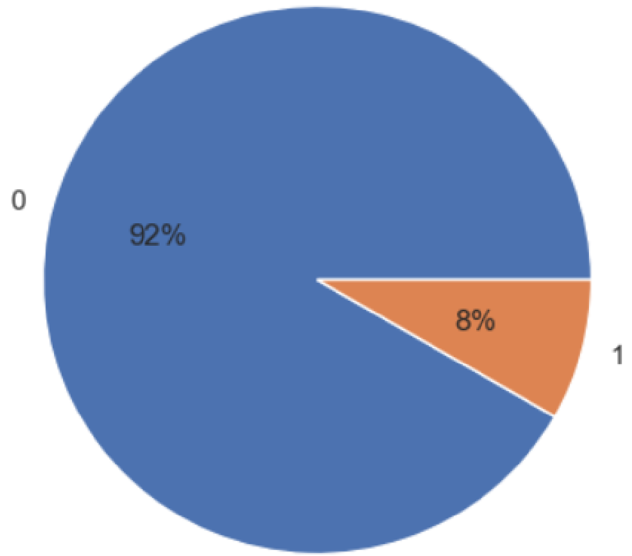
Steps followed in CREDIT EDA Analysis:

1. Data Understanding
    1. Understand the raw dataset and the columns. Shape, size and basic info of the data set.
2. Data Cleaning & Reduction
    1. Check for the missing values, remove the columns if it has more than 40% null values.
    2. Check for values, datatypes mismatch, Impute missing values handle outliers
3. Data Exploration
    1. Univariate Analysis
    2. Bivariate Analysis
    3. Multivariate Analysis

Case Study Insights

Target Variable Imbalance

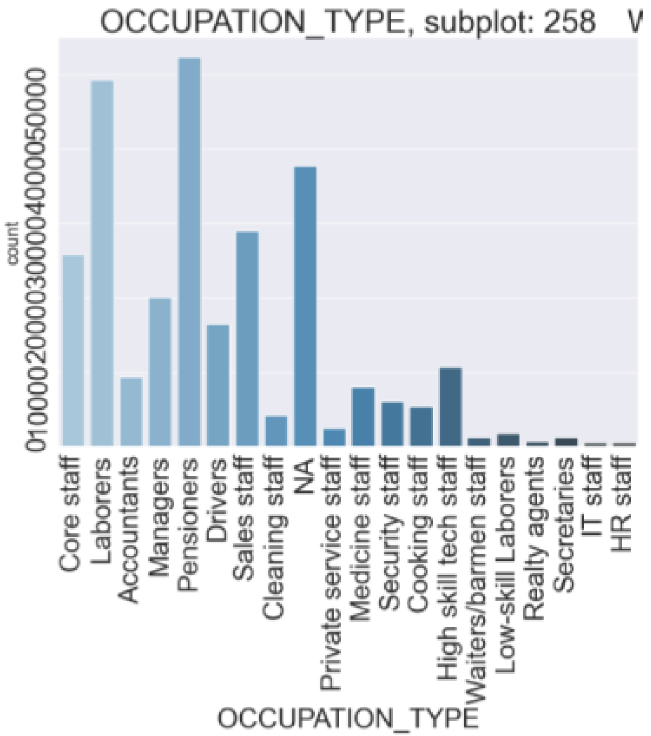
Target Variable denoting defaulters/Non-defaulters is highly imbalanced with 92% Non-defaulters and 8% Defaulters.

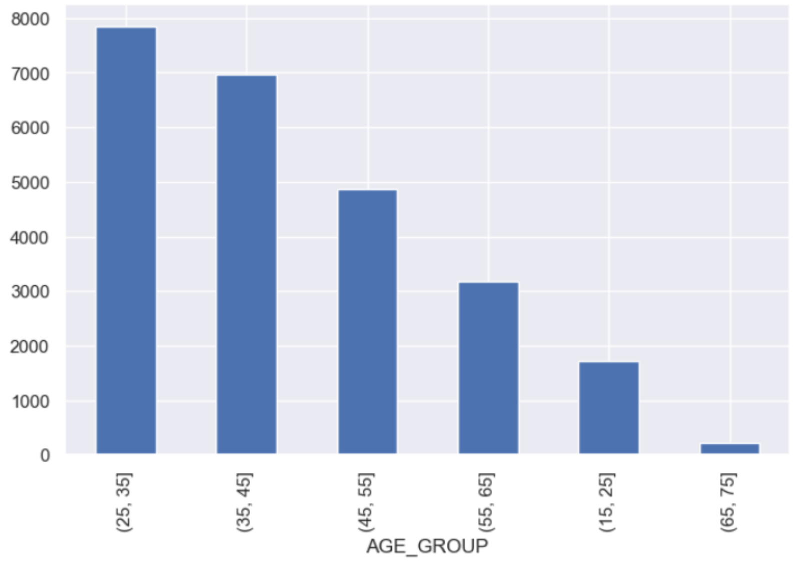


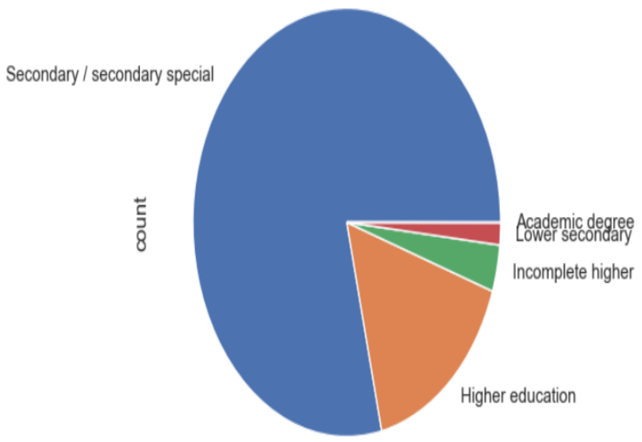
Hence, dividing the dataset to two – with

defaulters and Non-defaulters

Univariate Analysis of Defaulters





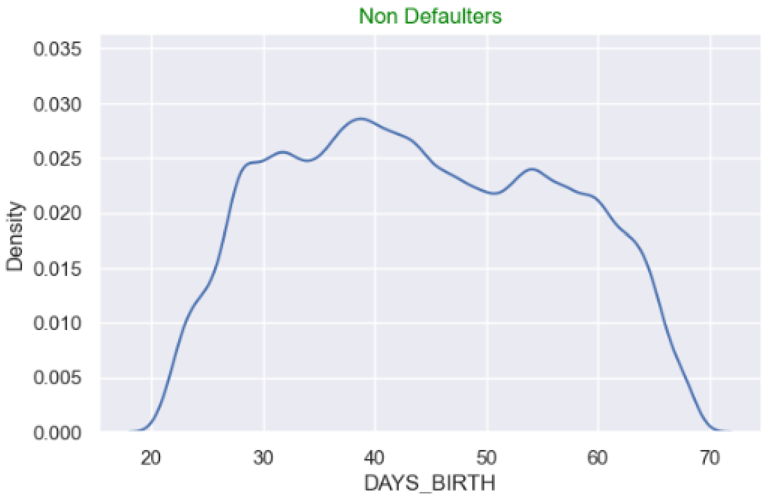


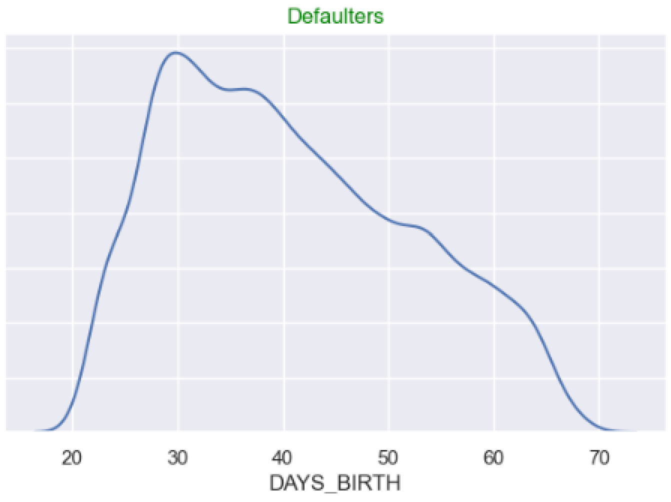
Laborers are more likely to default indicating their financial difficulties

People in their 20s and 30s are more likely to default

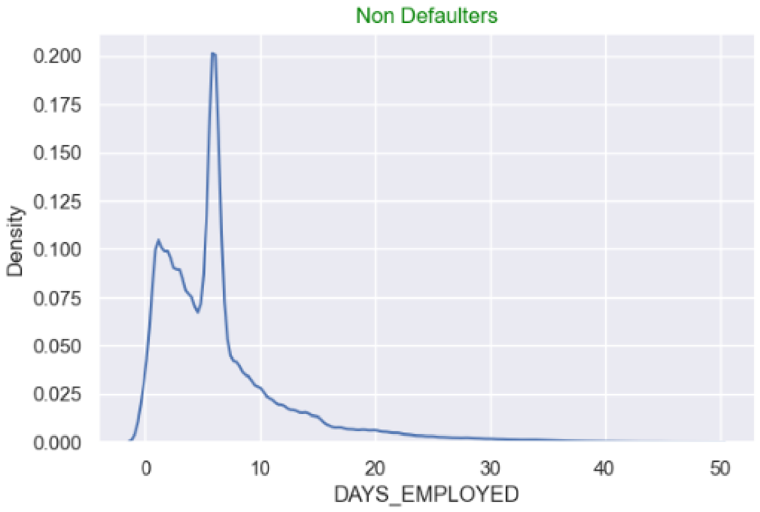
Notably, people with higher education status are defaulting more.

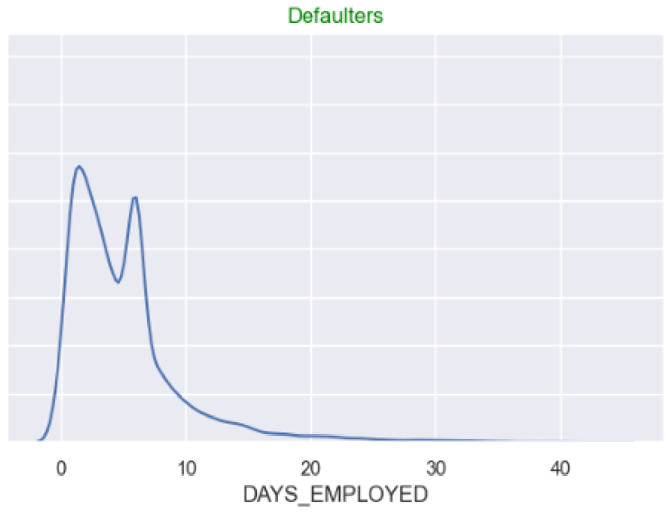
Univariate Analysis of Defaulters – Numerical





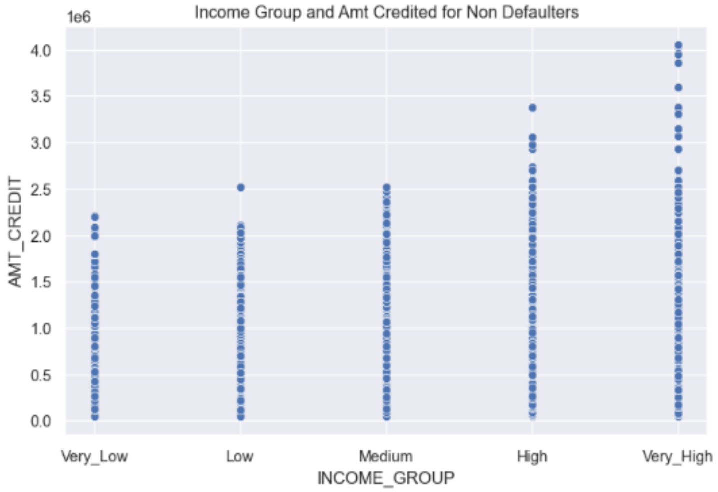
Young people age group density is more in Defaulters

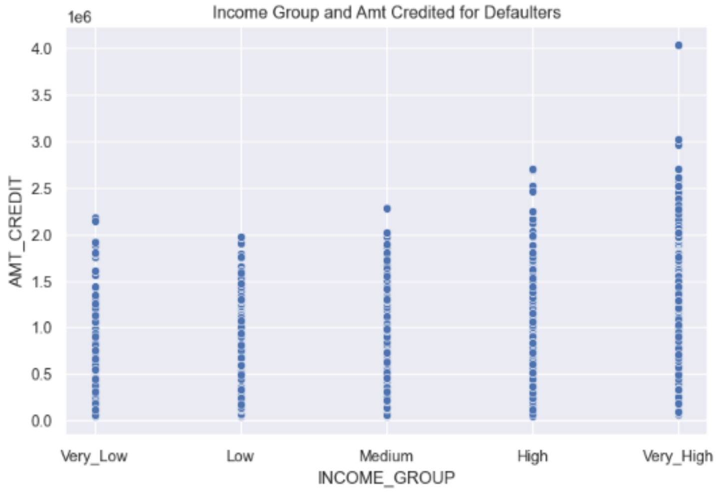




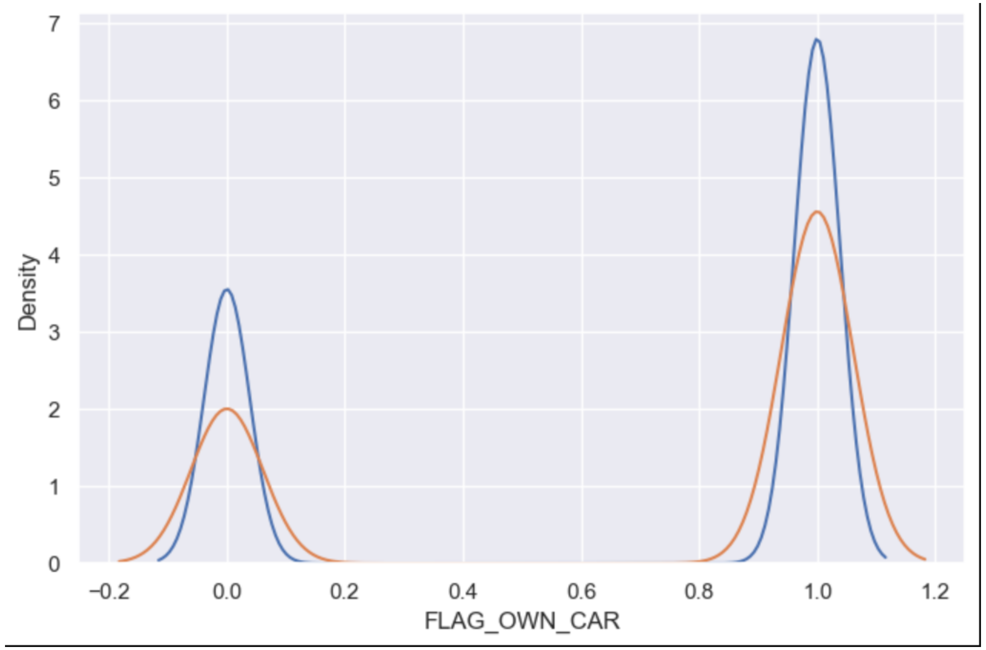
Defaulters have less experience in their work when compared to Non-defaulters

Bivariate Analysis of Defaulters - Numerical



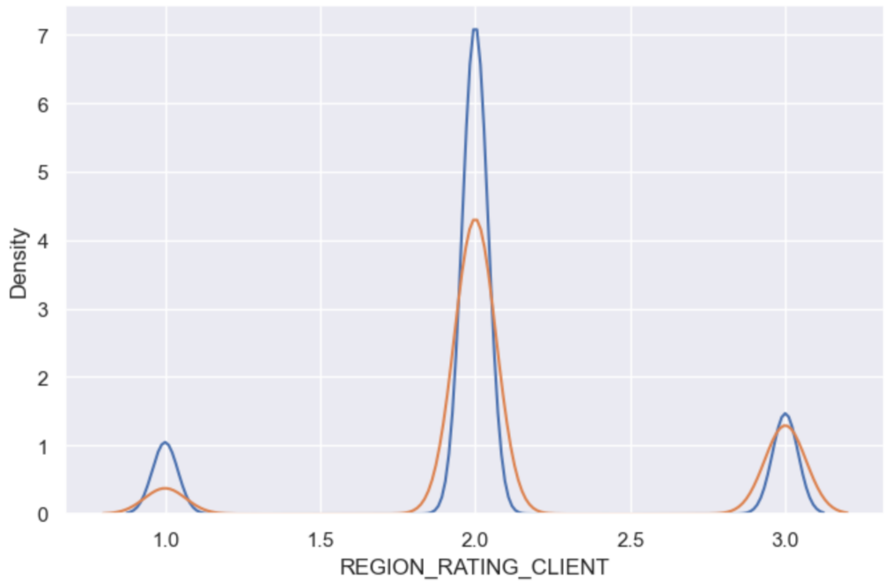


Loan amount for very high value income range seems to be very high(though it seems to be proportionate with the income group)

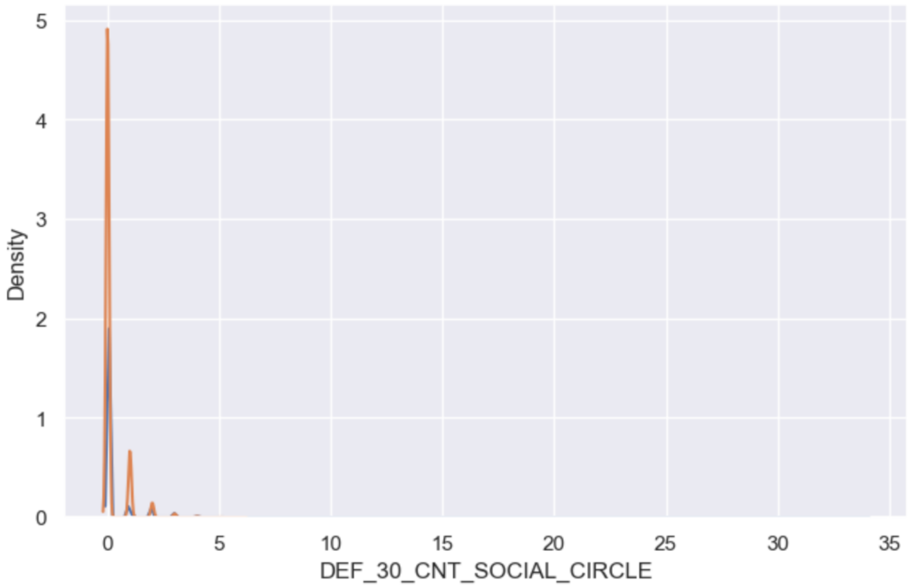


Loan Defaulters seem to own car less than non defaulters

Bivariate Analysis of Defaulters - Numerical

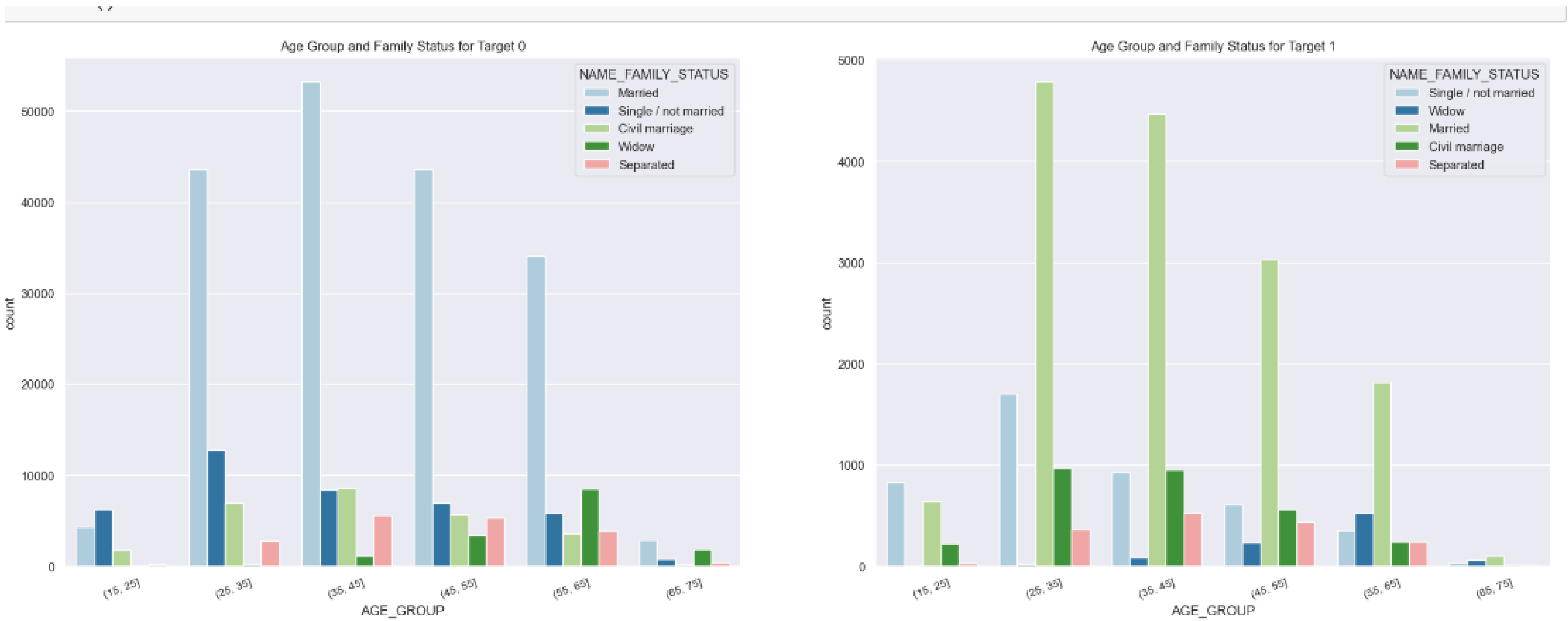


Region Rating of Defaulters see to be low when compared to Non defaulters

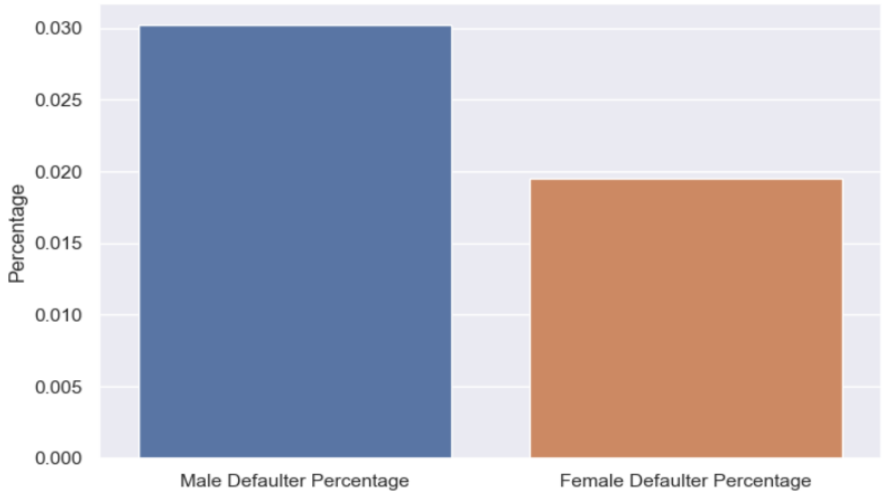


Defaulters social circle graph indicate that people who default come from a same social circle who are also more likely to default.

Bivariate Analysis of Defaulters - Categorical



Married people in the young age group of 25-35 are facing financial difficulties and are likely to default



Even though the female applicants are more, the percentage of male applicants defaulting are more than Female applicants

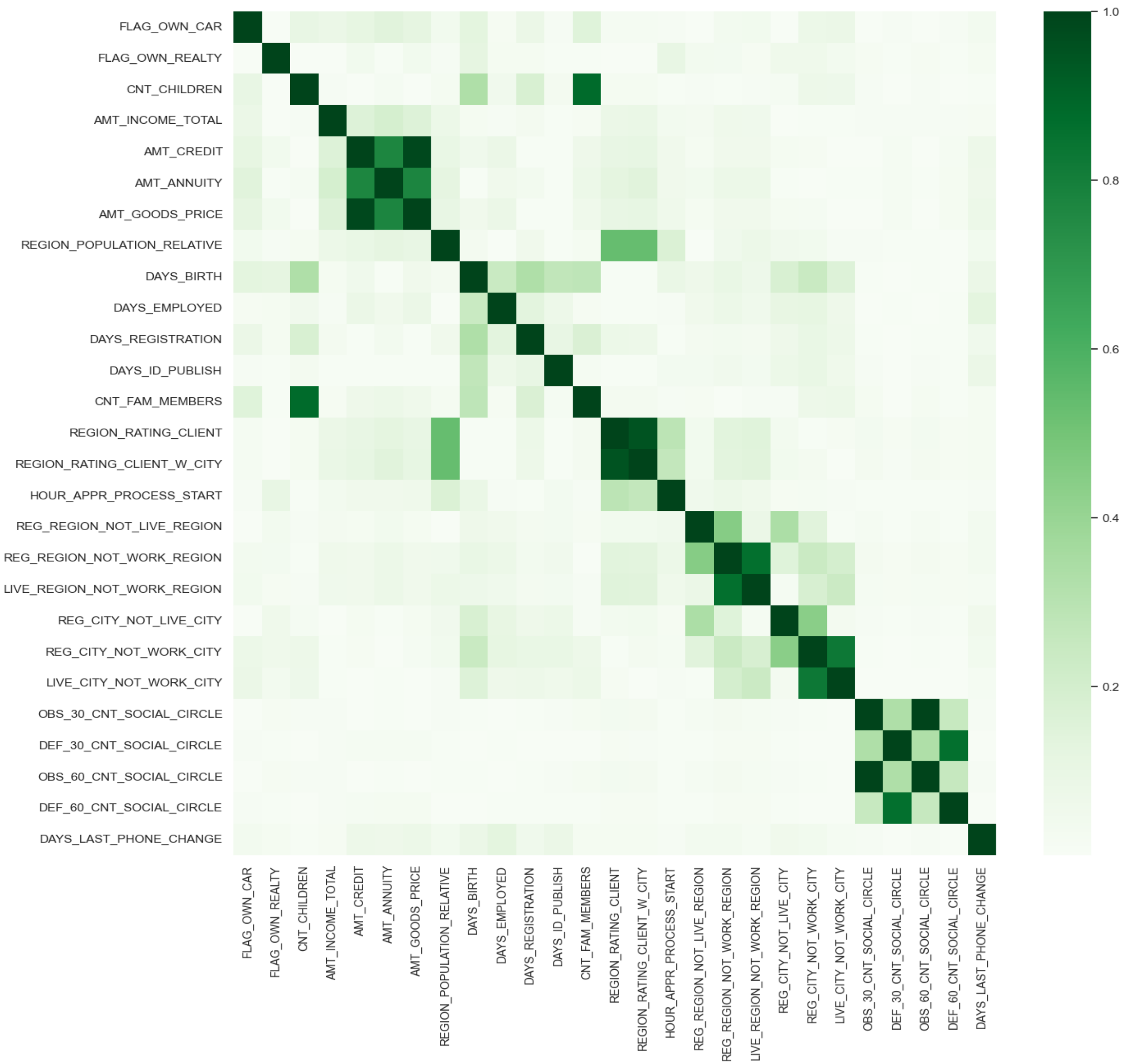
Top 10 Correlation Variables for Defaulters

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.99827
AMT_CREDIT	AMT_GOODS_PRICE	0.982783
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956637
CNT_CHILDREN	CNT_FAM_MEMBERS	0.885484
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.869016
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.847885
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.77854
AMT_ANNUITY	AMT_GOODS_PRICE	0.752295
AMT_CREDIT	AMT_ANNUITY	0.752195
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.497937

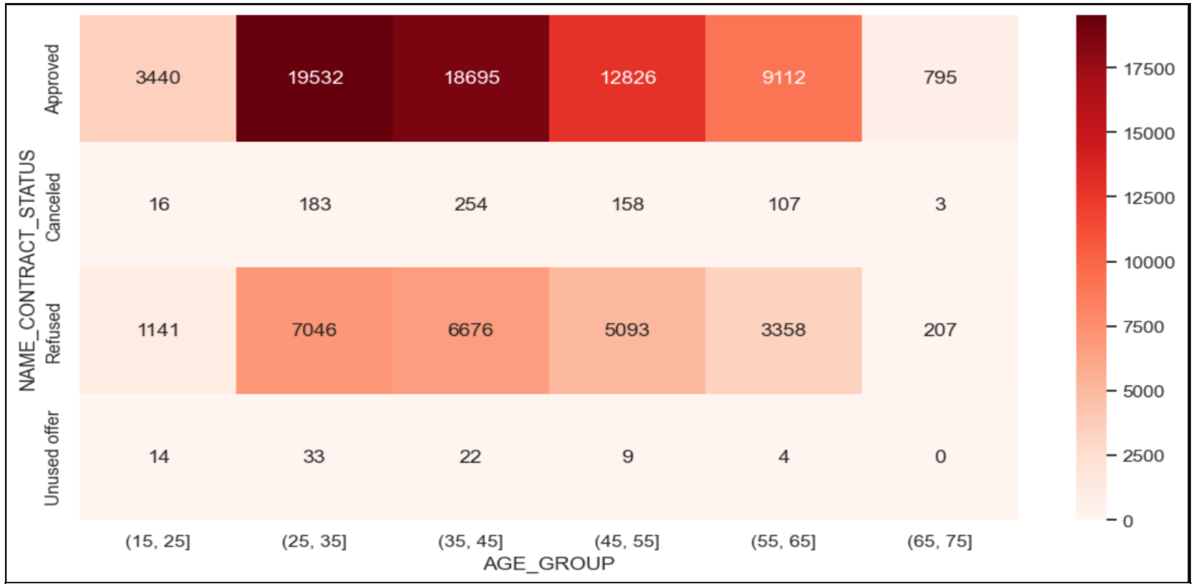
Top 10 Correlation Variables for Non - Defaulters

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.99851
AMT_CREDIT	AMT_GOODS_PRICE	0.987022
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.950149
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878572
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.861851
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.859371
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830379
AMT_ANNUITY	AMT_GOODS_PRICE	0.776423
AMT_CREDIT	AMT_ANNUITY	0.771298
REGION_POPULATION_RELATIVE	REGION_RATING_CLIENT	0.539006

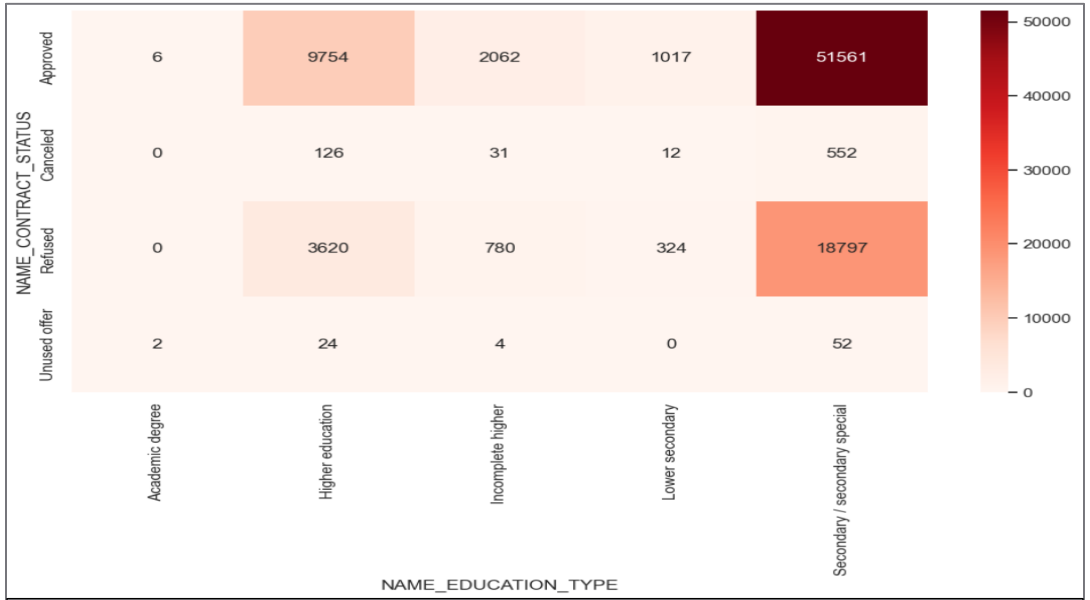
Heatmap showing the corelation of Variables



Previous Application Dataset Insights

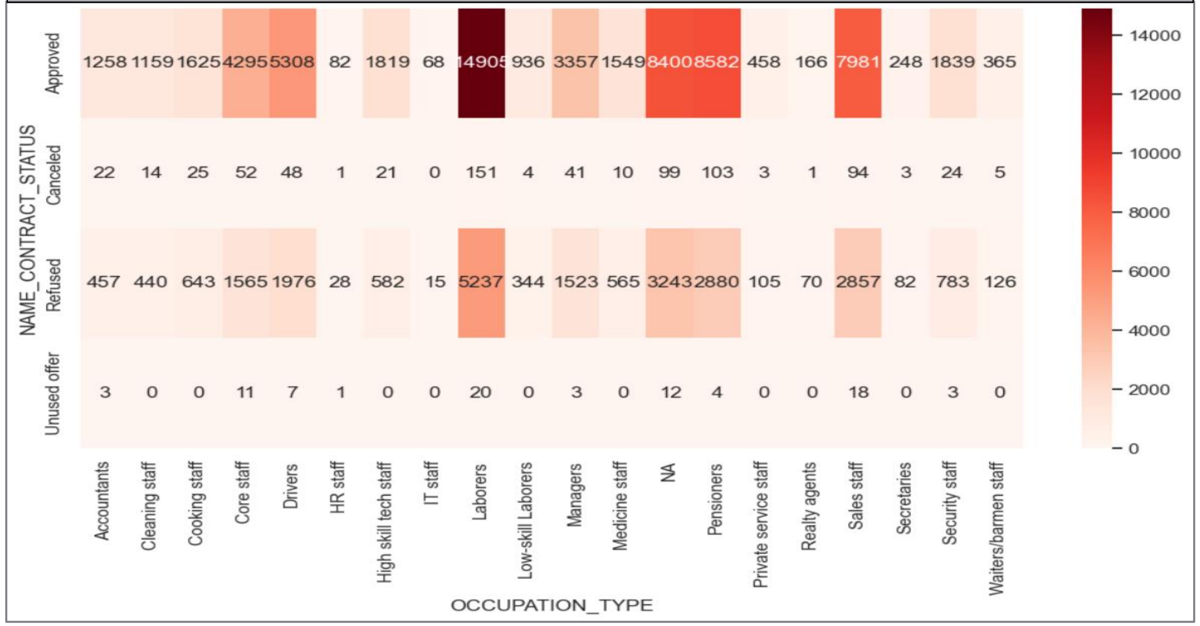


As confirmed earlier, younger age groups are more likely to default

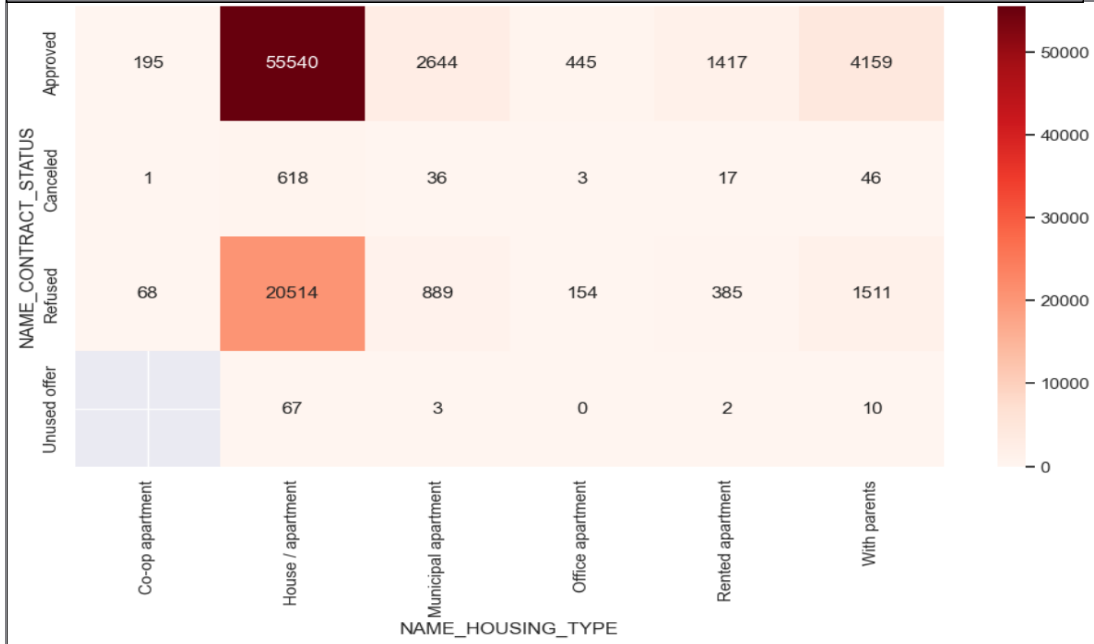


People with secondary education are more likely to default

Labourers face more difficulties in paying on time



Even people with own houses, tend to default more



Summary

Two data sets were considered when performing the EDA which resulted in following insights.

1. Younger age group are likely to default more as they might have lesser financial stability
2. Medium salaried people are likely to default more.
3. Men, in general, are defaulting more than female applicants.
4. Most of the defaulters don’t possess car/realty
5. Social rating provides good indication that people from similar social circle are more likely to default.
6. People with lower external source are showing high trends of defaulting.