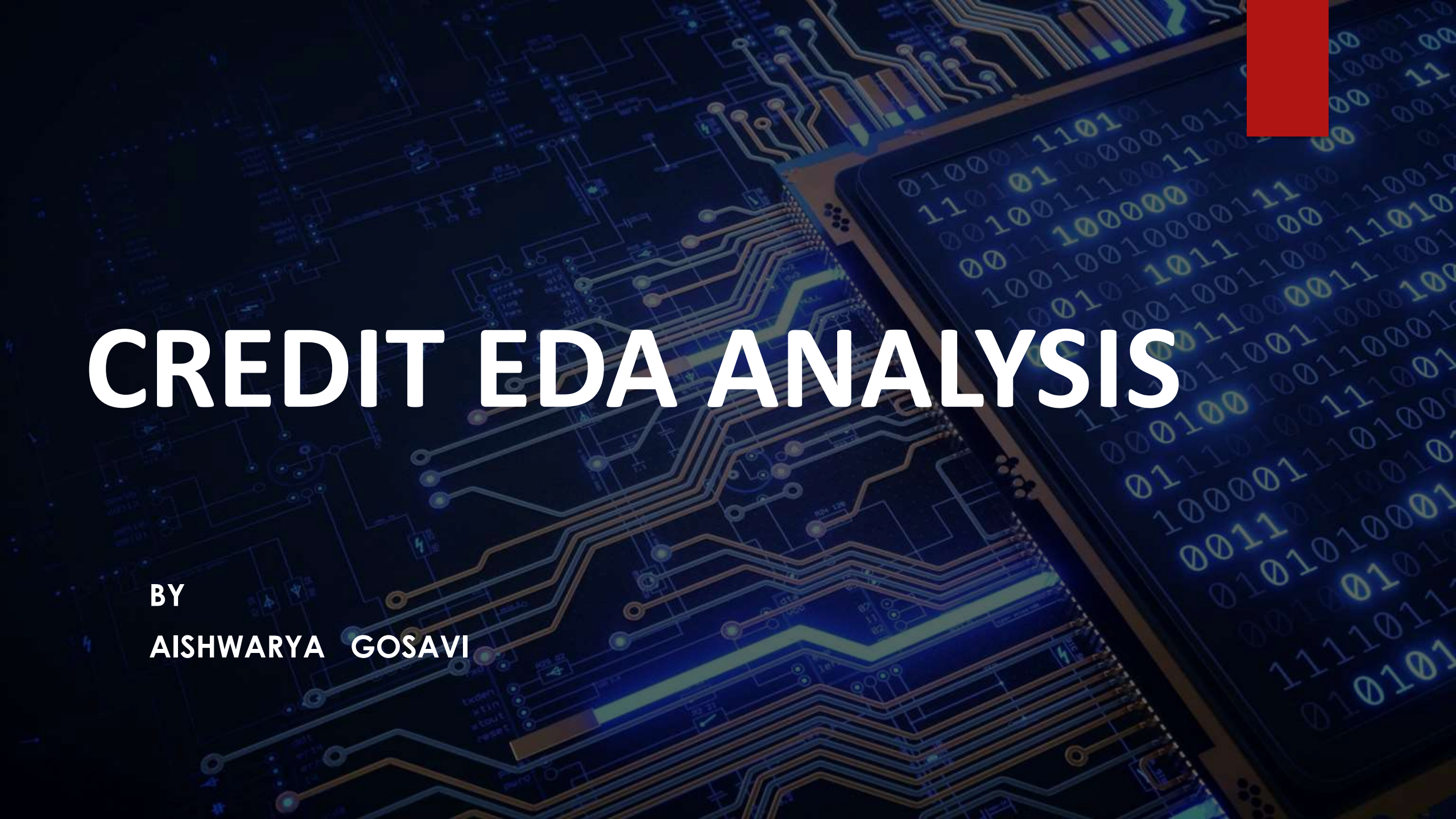


CREDIT EDA ANALYSIS

BY

AISHWARYA GOSAVI

PROBLEM STATEMENT :

- The loan providing companies find it hard to give loans to people due to their insufficient or non-existent credit history.
- When the company receives a loan application, the company has to decide for loan approval based on the application profile.
- There are 2 types of risks-

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

2. 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.

BUSINESS OBJECTIVE :

- This case study aims to identify patterns which indicates if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan ,lending(to risky applicants)at a higher interest rate, etc.This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study
- The company wants to understand the driving factors(or driver variables)behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

STEPS INVOLVED :

1. **DATA SOURCING (Provided already in assignment)-**
    1.Application data CSV file
    2.Previous data CSV file
    3.Application Description

2. **DATA UNDERSTANDING –**
    1. Importing necessary libraries.
    2. Checking each column,Index,header etc.
    3. Identifying data Quality issues.

3. **DATA MANIPULATION -**
    1. Checking the structure of the data.
    2.Missing value imputation analysis.
    3.Checking datatypes to date,time etc for ease case of analysis.

4. OUTLIERS CHECK &DATA IMBALANCE CHECK -

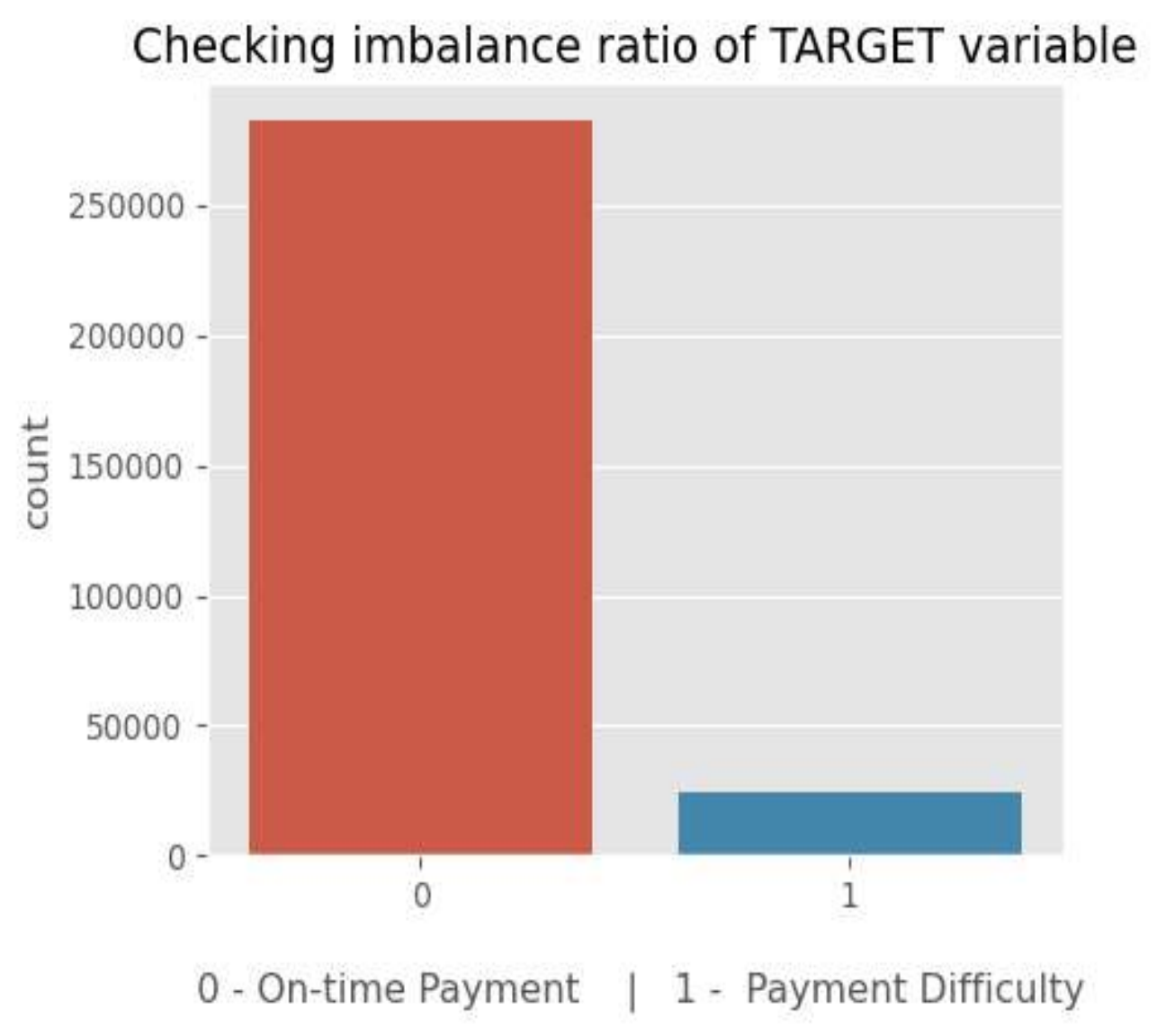
- Checking the data for outliers that would cause the analysis to be biased.
- Checking for imbalances, ratio, percentage of imbalance.

5.DATA ANALYSIS -

- Business requirement oriented analysis.
- Correlation between columns
- Univariate Analysis
- Bivariant Analysis
- HEATMAP

6.CONCLUSION-

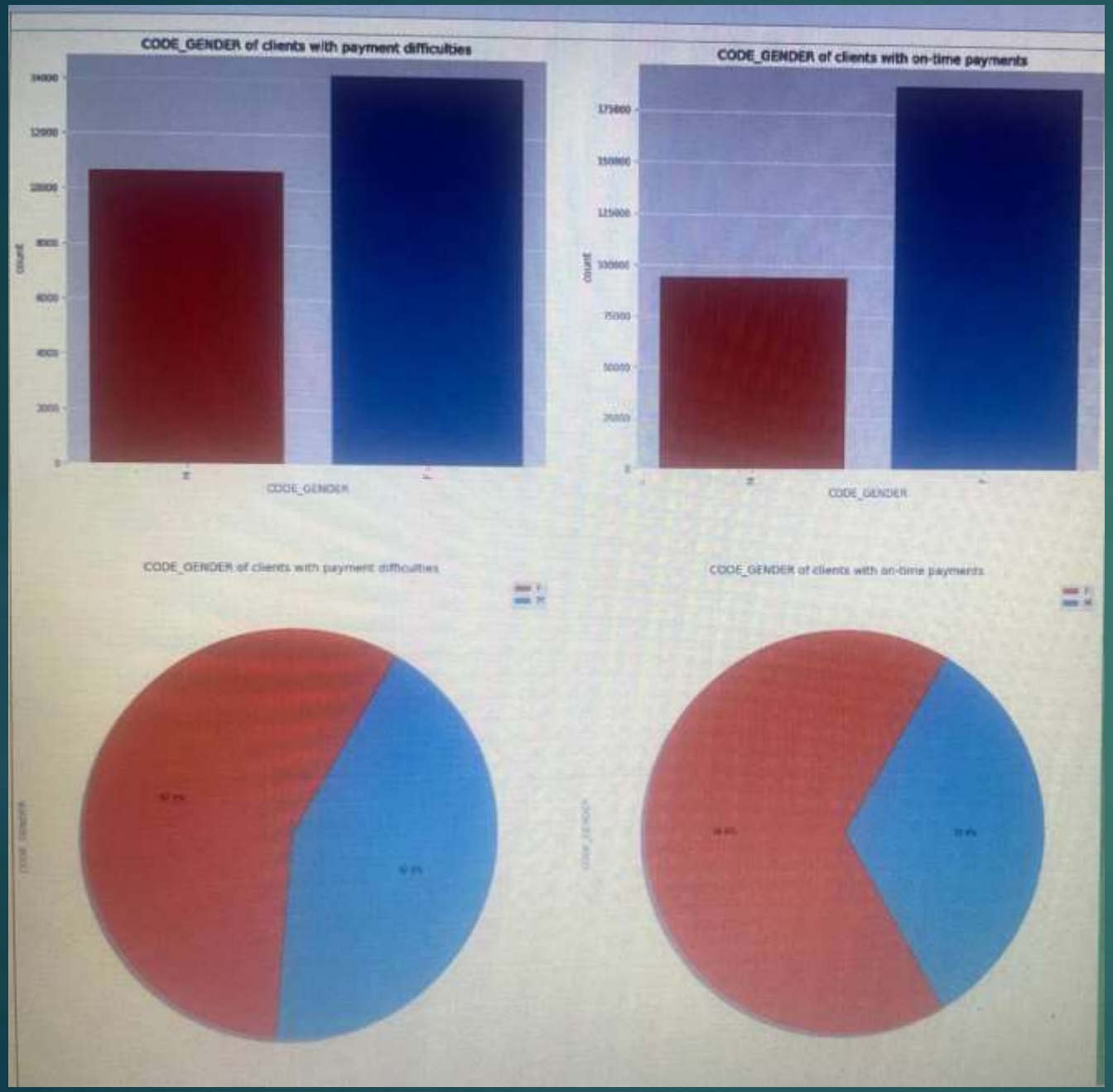
- How EDA is helpful for taking future course of action for business.



CHECKING IMBALANCE FOR TARGET COLUMN 'Target'-

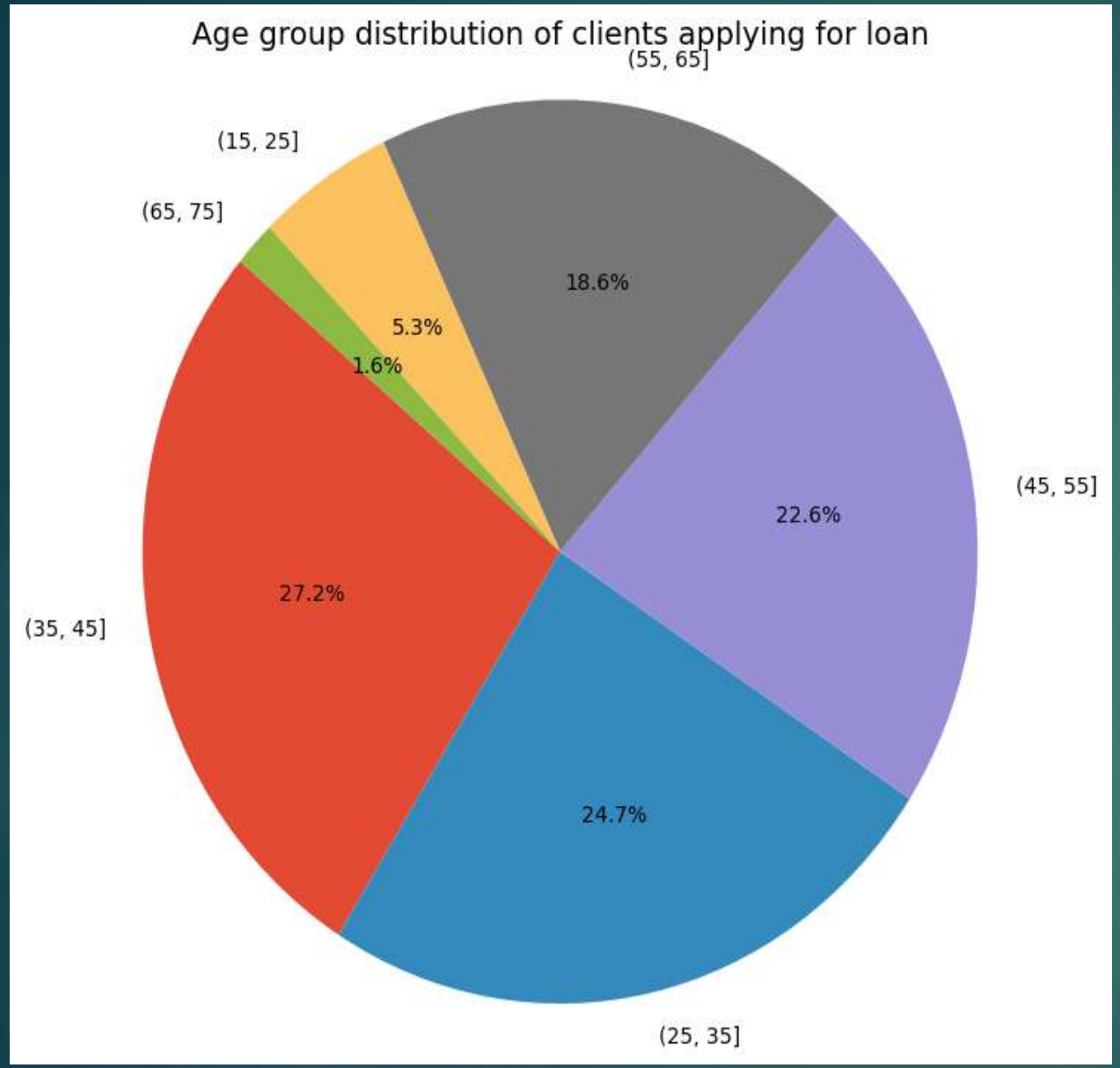
- The ratio for people who have difficulty paying to on-time payers is 11.38.
- 1 in every 11 applicant has payment difficulty.
- print(df2['TARGET'].value_counts(normalize=True)) shows

	TARGET
0	0.919271
1	0.080729



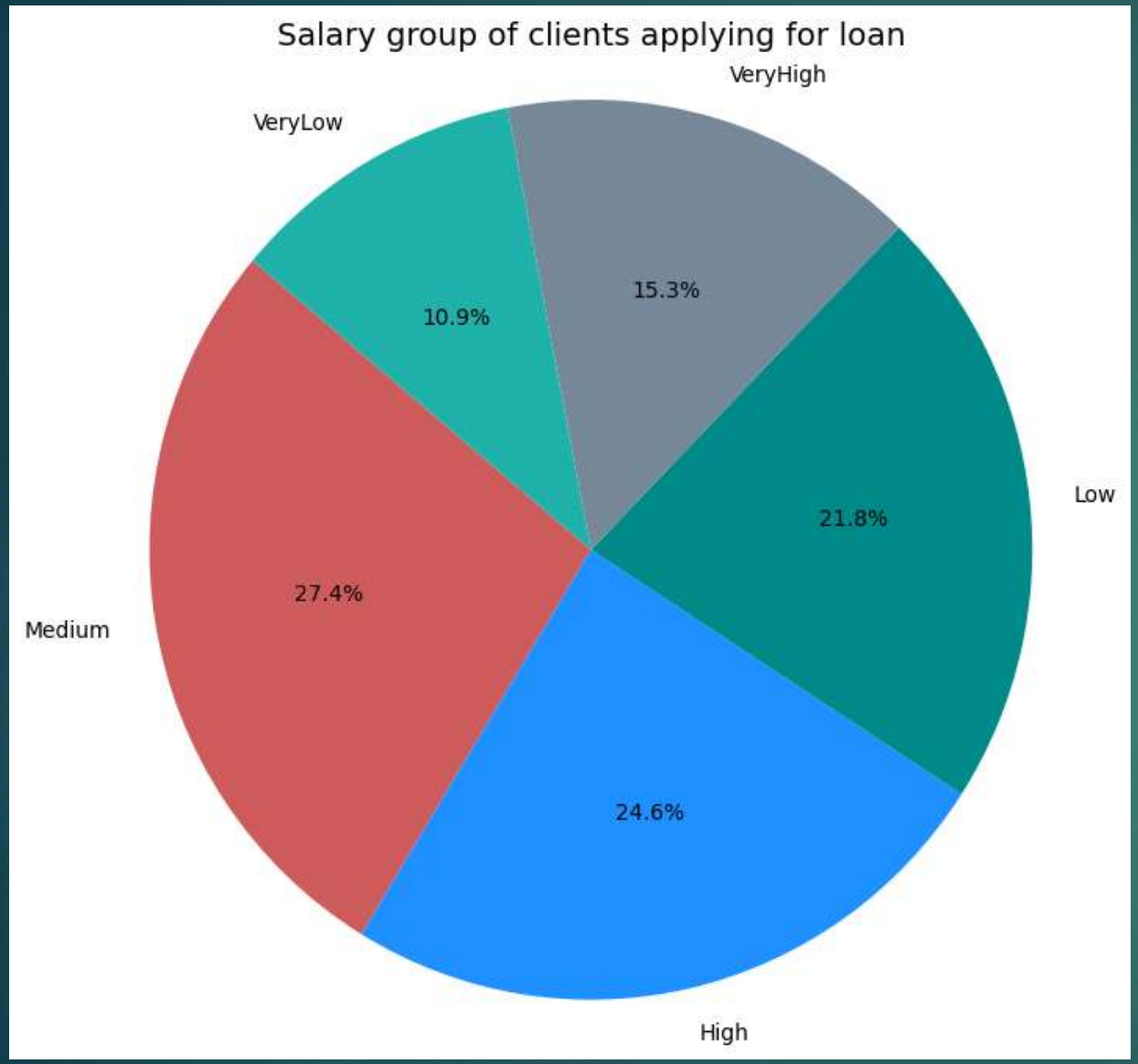
GENDER FACTOR

- Male applicants are more defaulting than Female factor.
  - Female have processed more loans comparison to males.
  - Female clients who are working are more likely to make on time payment.



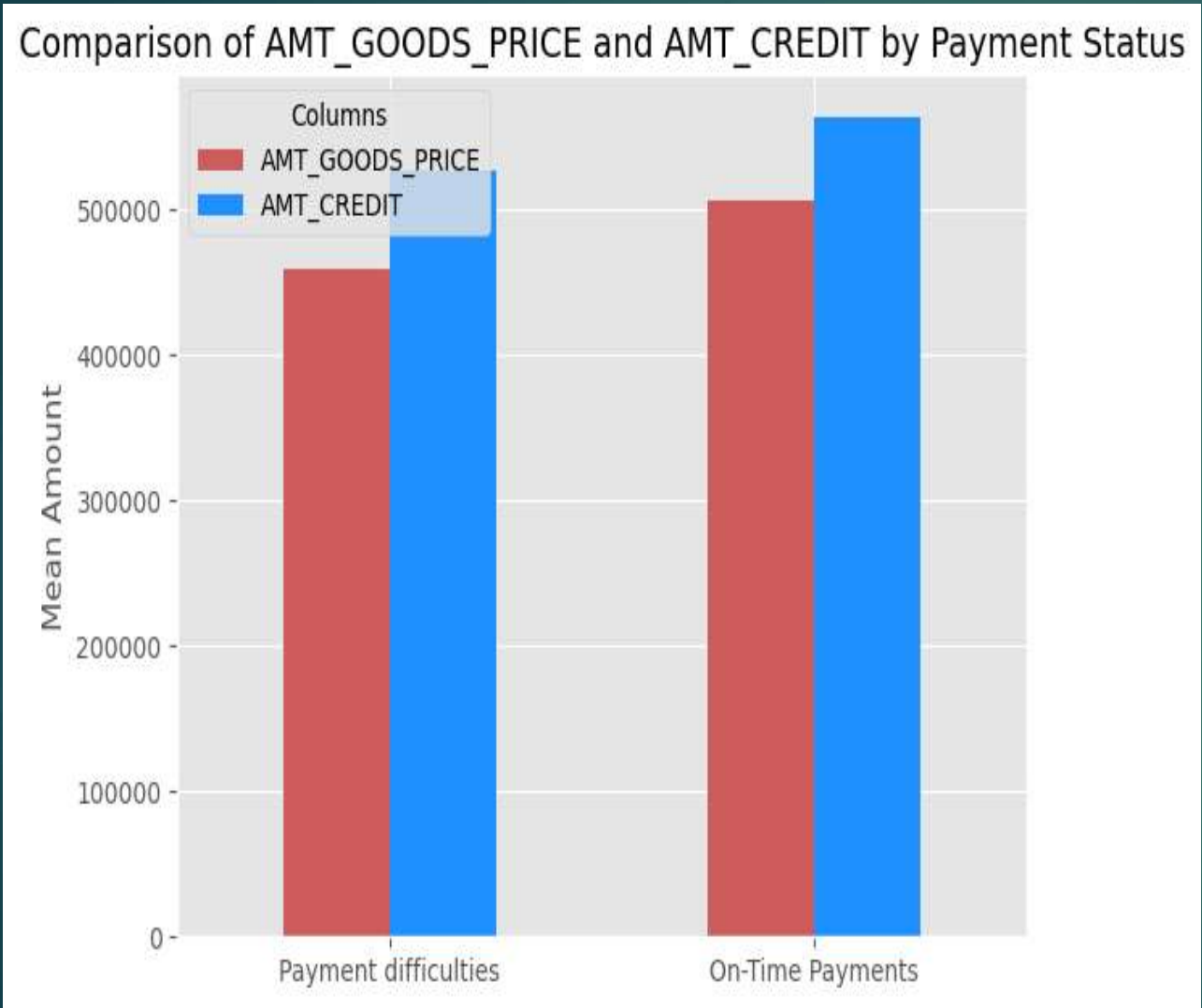
Age is factor when determining whether client can pay on time

- For YEARs_BIRTH between 20 and 40,there are more clients with payment difficulties.
- 35-40 age group is the largest group of age applying loans.
-Clients in the age range 30-40 and 40-50 is ideal target for loans.

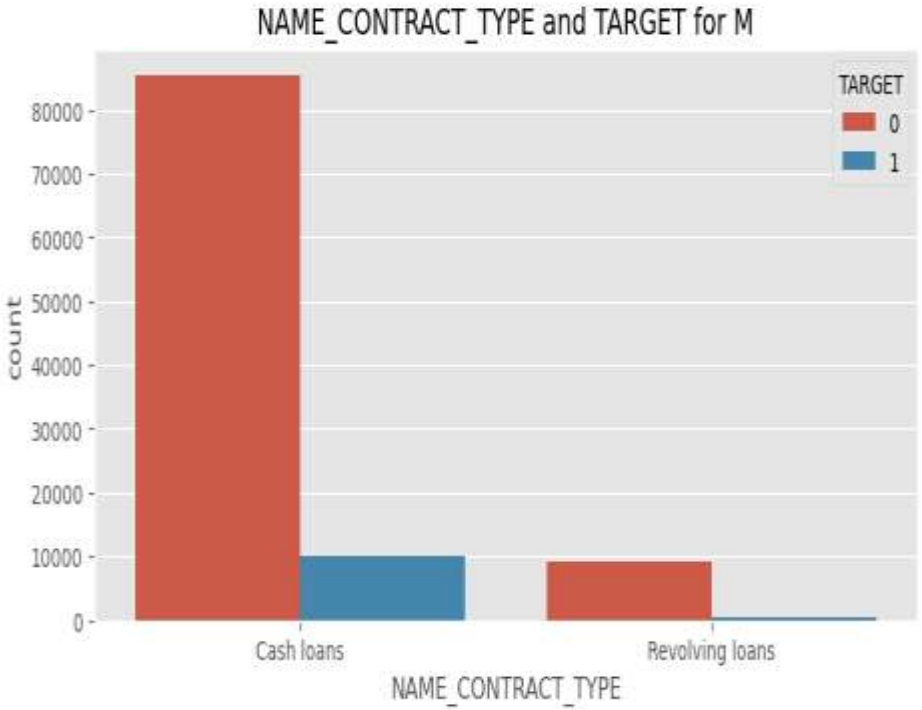


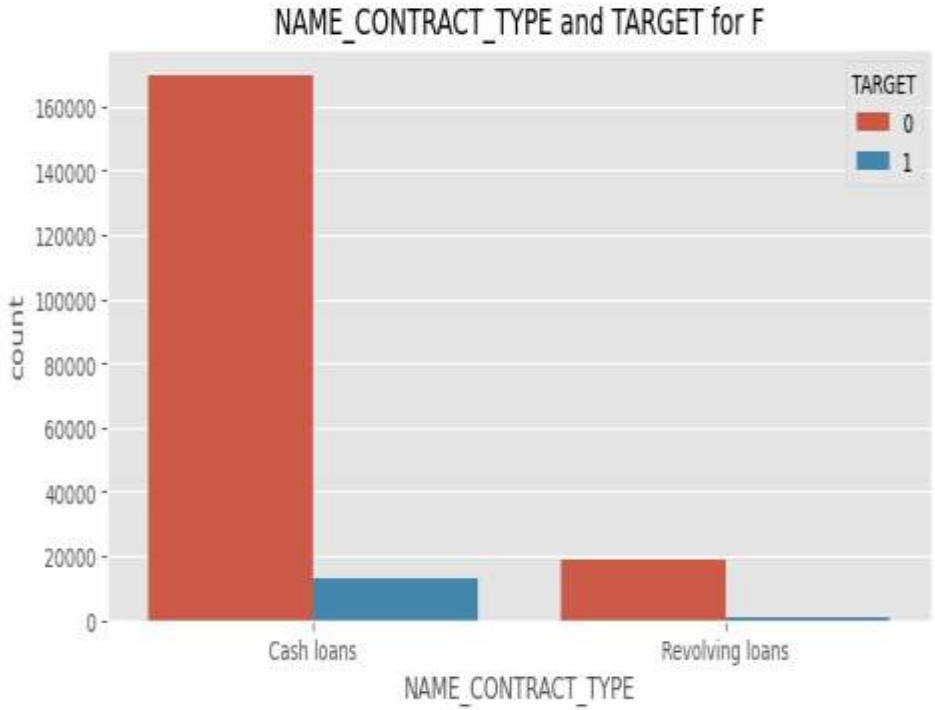
Salary group of clients applying for loan-

'Medium' income group is the largest group applying for loans. Followed by "High" income group. "Very Low" income group is the smallest group applying for loan.

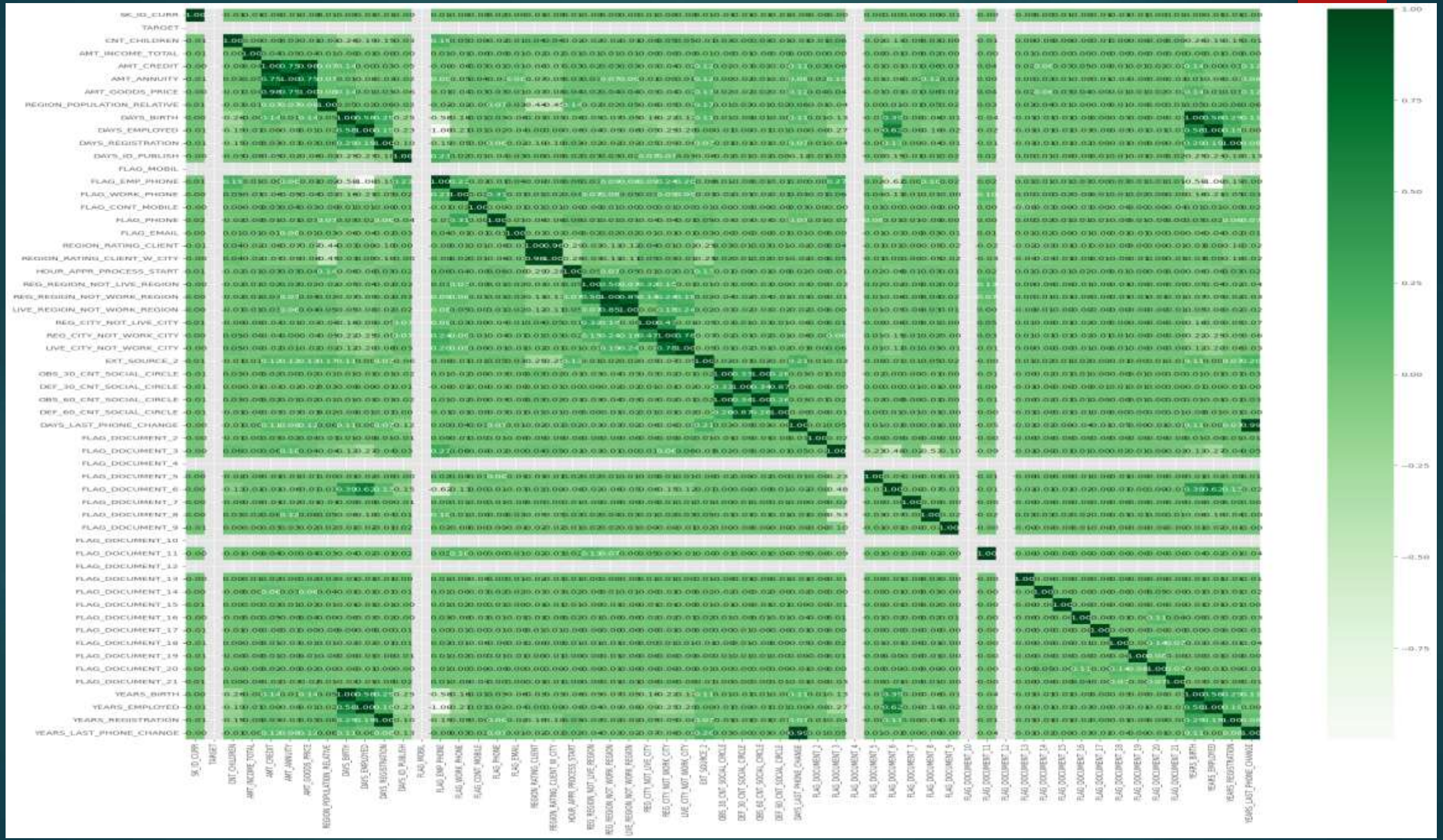


- On average, individuals with on-time payments have both higher AMT_GOODS_PRICE and AMT_CREDIT amounts compared to those with payment difficulties.
- The gap between AMT_GOODS_PRICE and AMT_CREDIT is more pronounced for those with payment difficulties, indicating a potential risk factor where higher credit amounts relative to goods prices might correlate with payment issues.





- **Cash Loans Dominance**: Both males and females predominantly have cash loans, with a significantly higher count compared to revolving loans.
- **Default Rates**: For both genders, the proportion of defaults (TARGET=1, shown in blue) is higher in cash loans compared to revolving loans.
- **Gender Comparison**: The overall count of loans (both cash and revolving) is higher for females than males, indicating a higher loan acquisition rate among females.


SK_ID_CURR
TARGET
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
FLAG_MOBIL
FLAG_EMP_PHONE
FLAG_WORK_PHONE
FLAG_CONT_MOBILE
FLAG_PHONE
FLAG_EMAIL
REGION_RATING_CLIENT
REGION_RATING_CLIENT_W_CITY
HOUR_APPR_PROCESS_START
REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
EXT_SOURCE_2
OBS_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE
FLAG_DOCUMENT_2
FLAG_DOCUMENT_3
FLAG_DOCUMENT_4
FLAG_DOCUMENT_5
FLAG_DOCUMENT_6
FLAG_DOCUMENT_7
FLAG_DOCUMENT_8
FLAG_DOCUMENT_9
FLAG_DOCUMENT_10
FLAG_DOCUMENT_11
FLAG_DOCUMENT_12
FLAG_DOCUMENT_13
FLAG_DOCUMENT_14
FLAG_DOCUMENT_15
FLAG_DOCUMENT_16
FLAG_DOCUMENT_17
FLAG_DOCUMENT_18
FLAG_DOCUMENT_19
FLAG_DOCUMENT_20
FLAG_DOCUMENT_21
YEARS_BIRTH
YEARS_EMPLOYED
YEARS_REGISTRATION
YEARS_LAST_PHONE_CHANGE
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

Last slide of Heatmap insight

- The heatmap displays a correlation matrix, with dark green indicating strong positive correlations and white indicating no correlation.
- Notable strong correlations include **OBS_30_CNT_SOCIAL_CIRCLE** with **DEF_30_CNT_SOCIAL_CIRCLE**, and **AMT_GOODS_PRICE** with **AMT_CREDIT**.
- Weak correlations are observed between **FLAG_OWN_CAR** and **CNT_CHILDREN**, as well as the **TARGET** variable with most features.

Conclusion

- Chances of client having payment difficulty All the below variables were established in analysis of Application data frame as leading to default.
- Checked these against the Approved loans which have defaults, and it proves to be correct Medium income 25-35 years old s, followed by 35-45 years age group Male Unemployed Laborers, Salesman, Drivers Own House - No Other IMPORTANT Factors to be considered No of Bureau Hits in last week. Month etc – zero hits is good Amount income not correspondingly equivalent to Good Bought – Income 'Low' and 'High' is a concern Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern.
- This indicates that the financial company had Refused/Cancelled previous application but has approved the current and is facing default on these. Credible Applications refused Female applicants should be given extra weightage as defaults are lesser.
- Students and Business mean have no problem in repayment of the loan Previous applications with Refused, Cancelled, Unused loans also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.