



# Course project – credit risk analyser

by-anurag nirwan

## PROBLEM STATEMENT

- **Challenge:** Our consumer finance company, specializing in loans for city dwellers, struggles to approve applicants without a credit history. This can lead to defaults by some borrowers who take advantage of the situation.
- **Solution:** We can leverage Exploratory Data Analysis (EDA) to unearth patterns in our data. This will help us identify creditworthy applicants who might otherwise be rejected due to a lack of traditional credit history.

# APPROACH

- o Loading the data
- o Check for the missing/null values
- o Working on missing/null values as per industry standard
- o Working on modified data
- o Working on graphs
- o Finding out if there is any imbalance in the data
- o Doing Univariate and bivariate analysis
- o Finding out the correlation between the variables

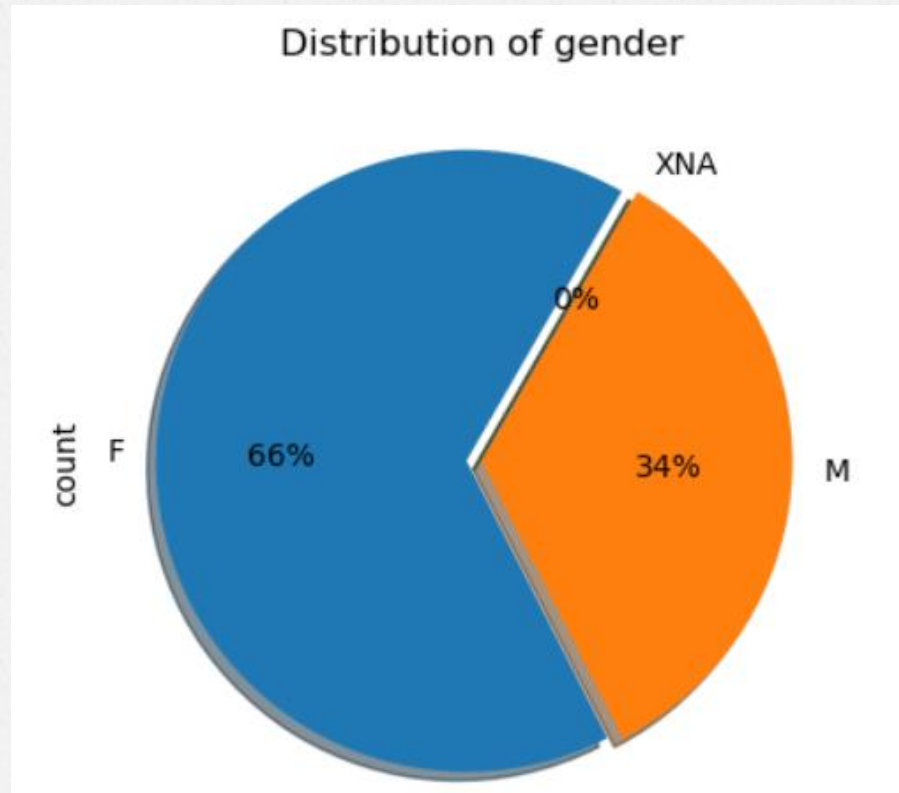


# Missing values

- o Both the data has missing values. As per the industry standard anything between 40% to 50% is threshold limit.
- o In previous application
  - o 'AMT\_DOWN\_PAYMENT', 'RATE\_DOWN\_PAYMENT', 'RATE\_INTEREST\_PRIMARY', 'RATE\_INTEREST\_PRIVILEGED', are having missing values.
- o In application data
  - o 'EXT\_SOURCE\_1', 'EXT\_SOURCE\_2', 'EXT\_SOURCE\_3', 'APARTMENTS\_AVG', 'BASEMENTAREA\_AVG', 'YEARS\_BEGINEXPLUATATION\_AVG', 'YEARS\_BUILD\_AVG', 'COMMONAREA\_AVG', 'ELEVATORS\_AVG', 'ENTRANCES\_AVG', 'FLOORSMAX\_AVG', 'FLOORSMIN\_AVG', 'LANDAREA\_AVG', 'LIVINGAPARTMENTS\_AVG', 'LIVINGAREA\_AVG', 'NONLIVINGAPARTMENTS\_AVG', 'NONLIVINGAREA\_AVG', 'APARTMENTS\_MODE', 'BASEMENTAREA\_MODE', 'YEARS\_BEGINEXPLUATATION\_MODE', 'YEARS\_BUILD\_MODE', 'COMMONAREA\_MODE', 'ELEVATORS\_MODE', 'ENTRANCES\_MODE', 'FLOORSMAX\_MODE', 'FLOORSMIN\_MODE', 'LANDAREA\_MODE', 'LIVINGAPARTMENTS\_MODE', 'LIVINGAREA\_MODE', 'NONLIVINGAPARTMENTS\_MODE', 'NONLIVINGAREA\_MODE', 'APARTMENTS\_MEDI', 'BASEMENTAREA\_MEDI', 'YEARS\_BEGINEXPLUATATION\_MEDI', 'YEARS\_BUILD\_MEDI', 'COMMONAREA\_MEDI', 'ELEVATORS\_MEDI', 'ENTRANCES\_MEDI', 'FLOORSMAX\_MEDI', 'FLOORSMIN\_MEDI', 'LANDAREA\_MEDI', 'LIVINGAPARTMENTS\_MEDI', 'LIVINGAREA\_MEDI', 'NONLIVINGAPARTMENTS\_MEDI', 'NONLIVINGAREA\_MEDI', 'FONDKAPREMONT\_MODE', 'HOUSETYPE\_MODE', 'TOTALAREA\_MODE', 'WALLSMATERIAL\_MODE', 'EMERGENCYSTATE\_MODE', 'OWN\_CAR\_AGE', 'OCCUPATION\_TYPE'

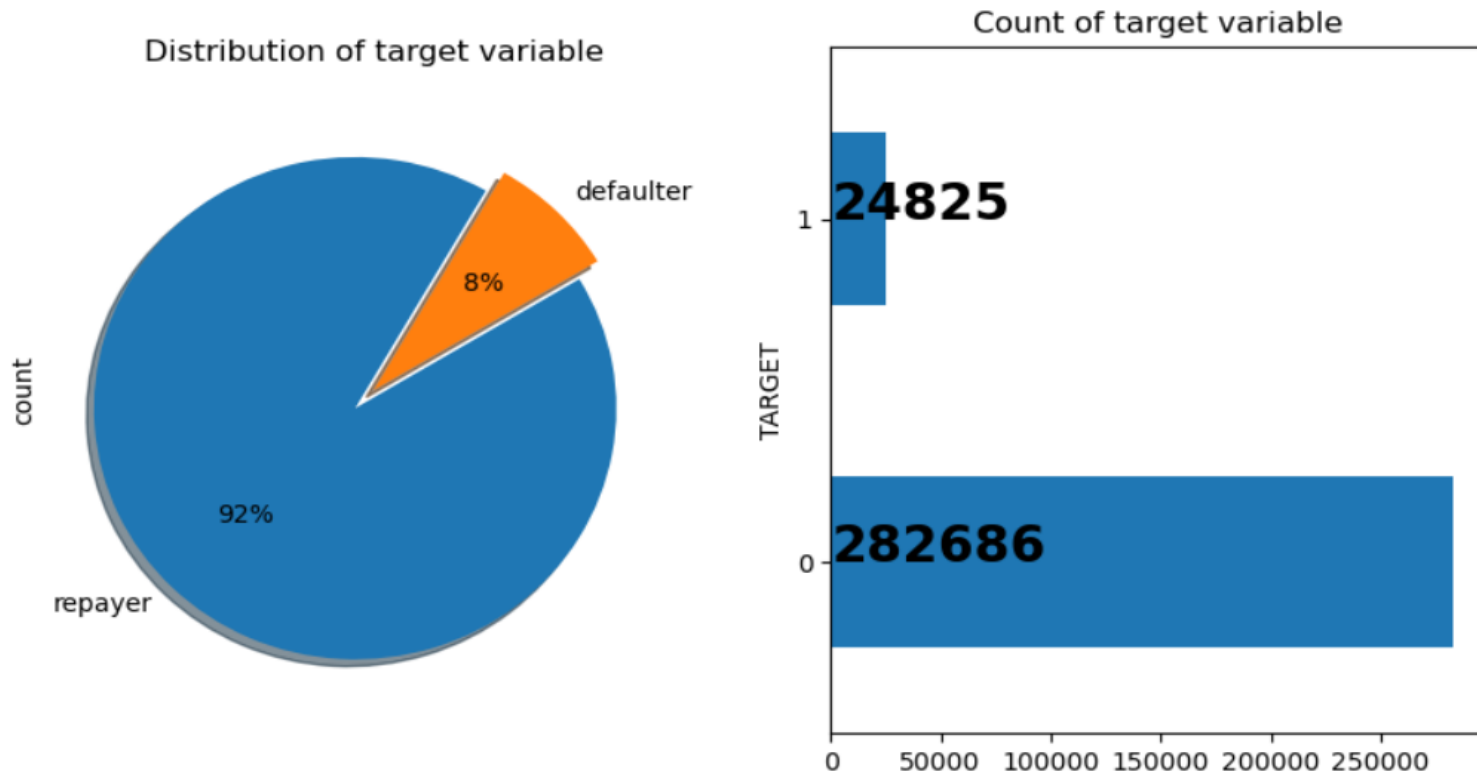
# GRAPHS

imbalance percentage



The application data is imbalance data as there are 66% female count compare to just 34% of male

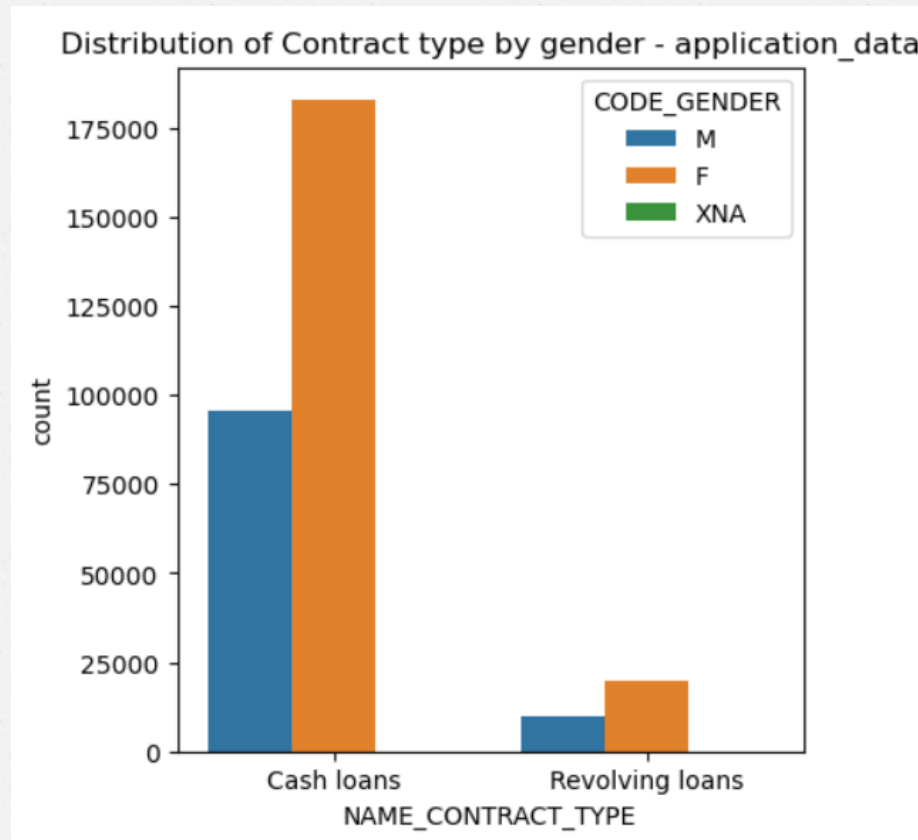
# LOAN DEFAULTER



This graph indicate that a total of 8% of population is loan defaulter



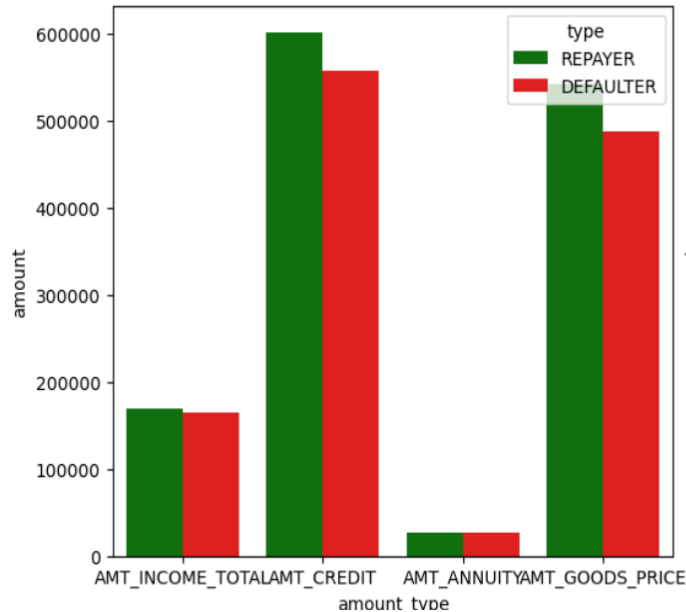
# LOAN DISTRIBUTION BY GENDER



In the above graph it shows that cash loan is always preferred compare to the revolving loans by both female and male

# Summary statistics between defaulters and non defaulters

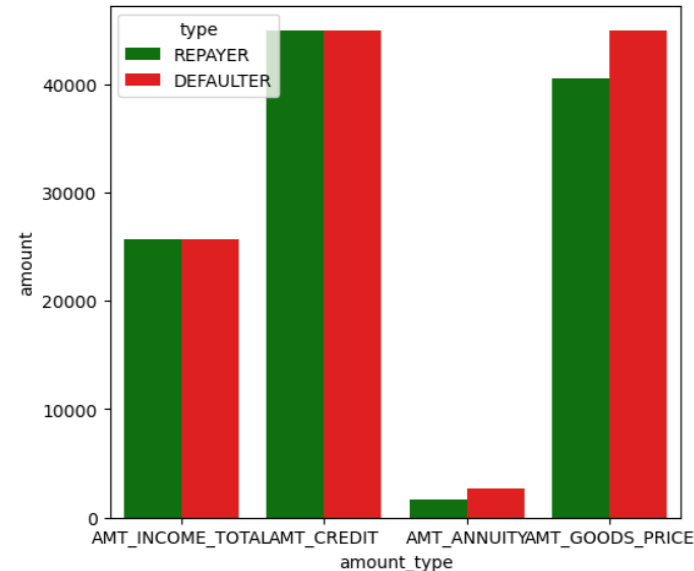
mean -- Defaulters vs Non-defaulters



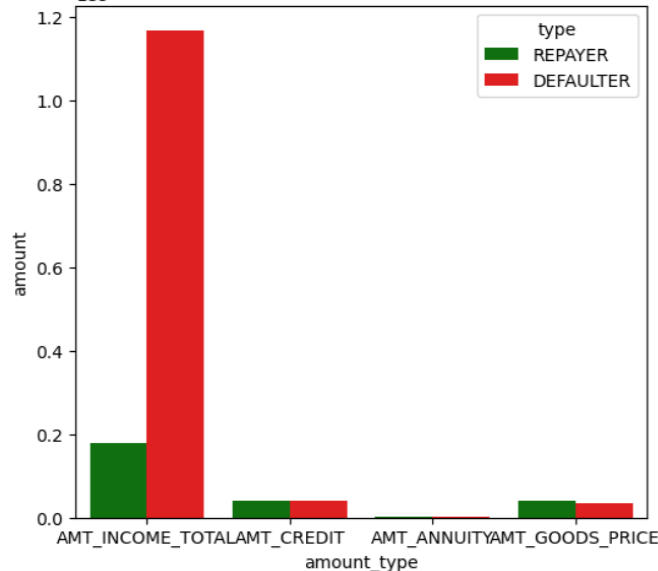
std -- Defaulters vs Non-defaulters



min -- Defaulters vs Non-defaulters



max -- Defaulters vs Non-defaulters



Income of client -

- 1 . Average income of clients who default and who do not are almost same.
- 2 . Standard deviation in income of client who default is very high compared to who do not default.
- 3 . Clients who default also has maximum income earnings

Credit amount of the loan ,Loan annuity, Amount goods price -

- 1 . Statistics between credit amounts, Loan annuity and Amount goods price given to clients who default and who don't are almost similar.