Capstone Project

Credit Card Default Prediction

Content

1. **Checking the data**

- Checking Defaulters ratio
- Gist of Age and Credit limit
- Clients in each Age group
- Defaulters with different category wise

2. **Implementing Classification techniques**

Problem Statement

Predicting if a customer will default the payment

Data Summary

Data set name – default of credit card clients

Shape of combined Dataset- 30000 rows, 26 columns

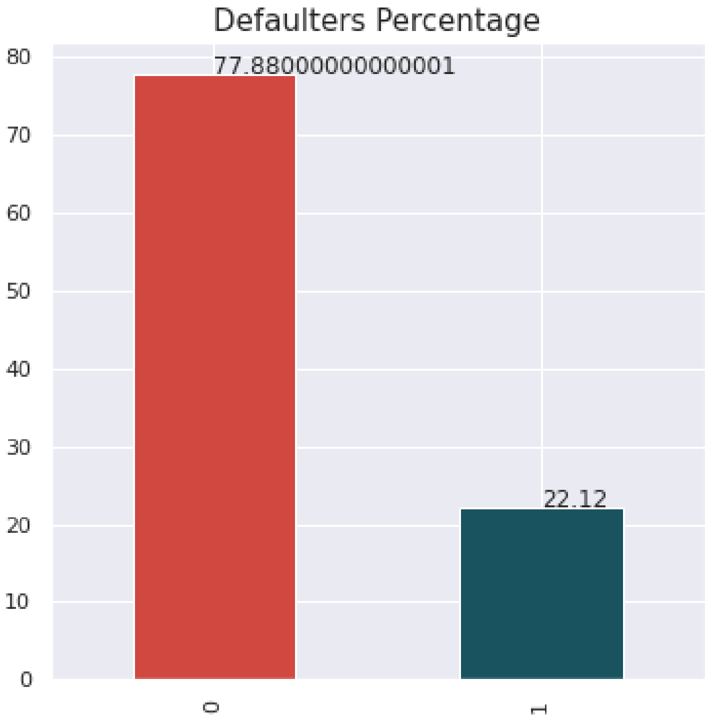
Columns - 'ID', 'LIMIT_BAL', 'SEX', 'EDUCATION', 'MARRIAGE', 'AGE', 'PAY_0', 'PAY_2', 'PAY_3', 'PAY_4', 'PAY_5', 'PAY_6', 'BILL_AMT1', 'BILL_AMT2', 'BILL_AMT3', 'BILL_AMT4', 'BILL_AMT5', 'BILL_AMT6', 'PAY_AMT1', 'PAY_AMT2', 'PAY_AMT3', 'PAY_AMT4', 'PAY_AMT5', 'PAY_AMT6', 'defaulters', 'AGE_BIN'

Cleaning dataset

All the values were already non null

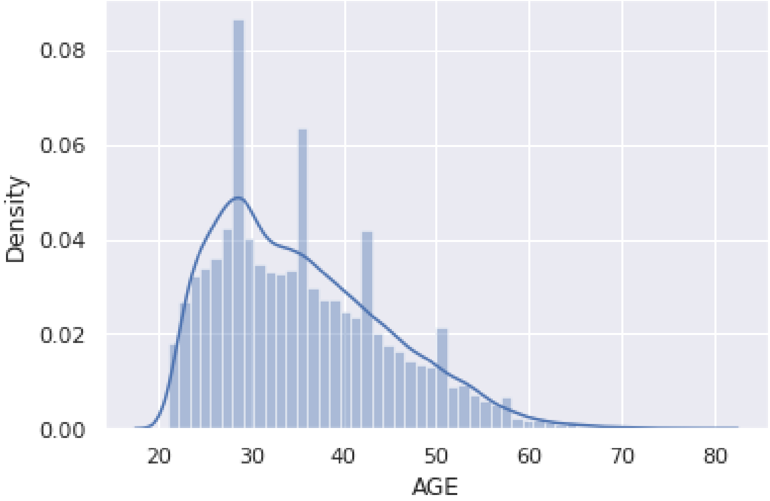
Defaulter’s Ratio

so we have 22% defaulters in our dataset and 77% persons are non defaulters

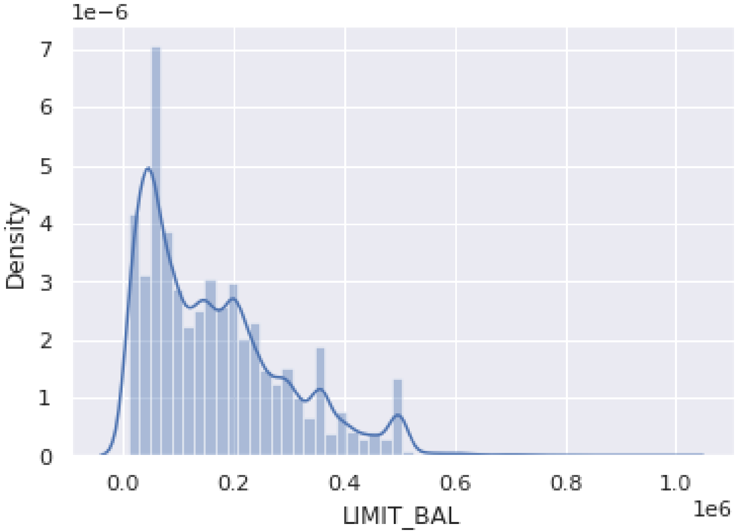


Gist of Age and Credit limit

The data shows that most people are of age range 20-40 and a few only from 50-60 age group

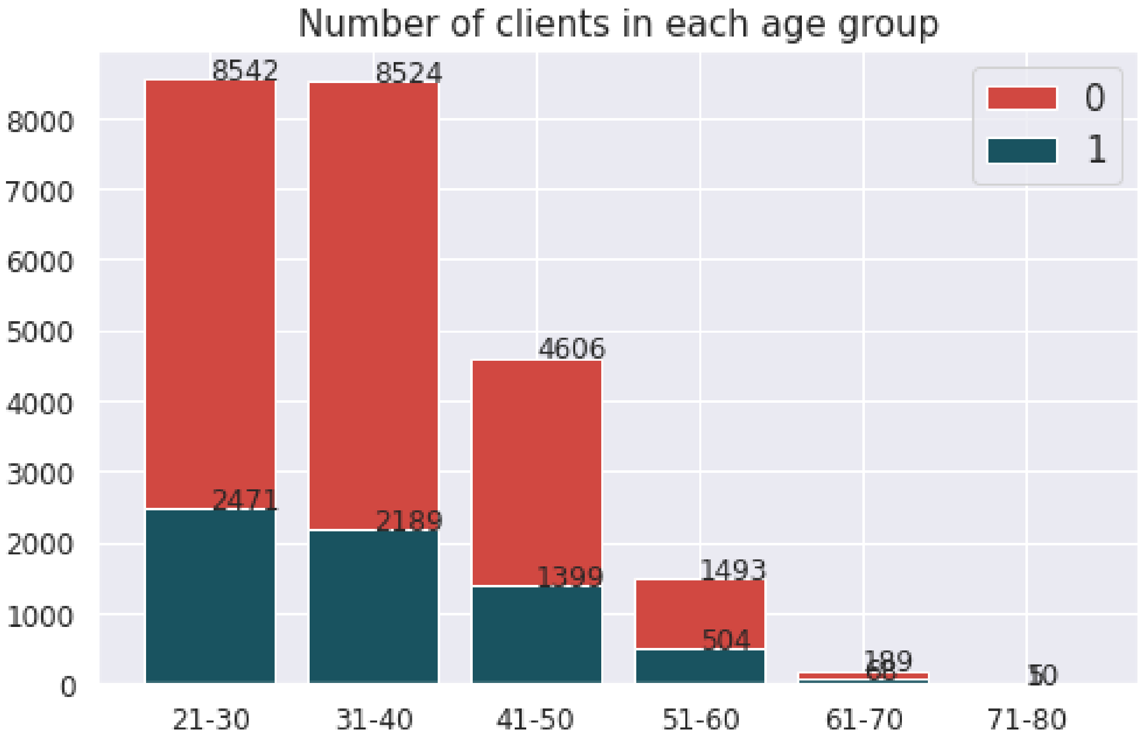


The data shows that most people are with 10-20K of credit limit



Clients in each Age group

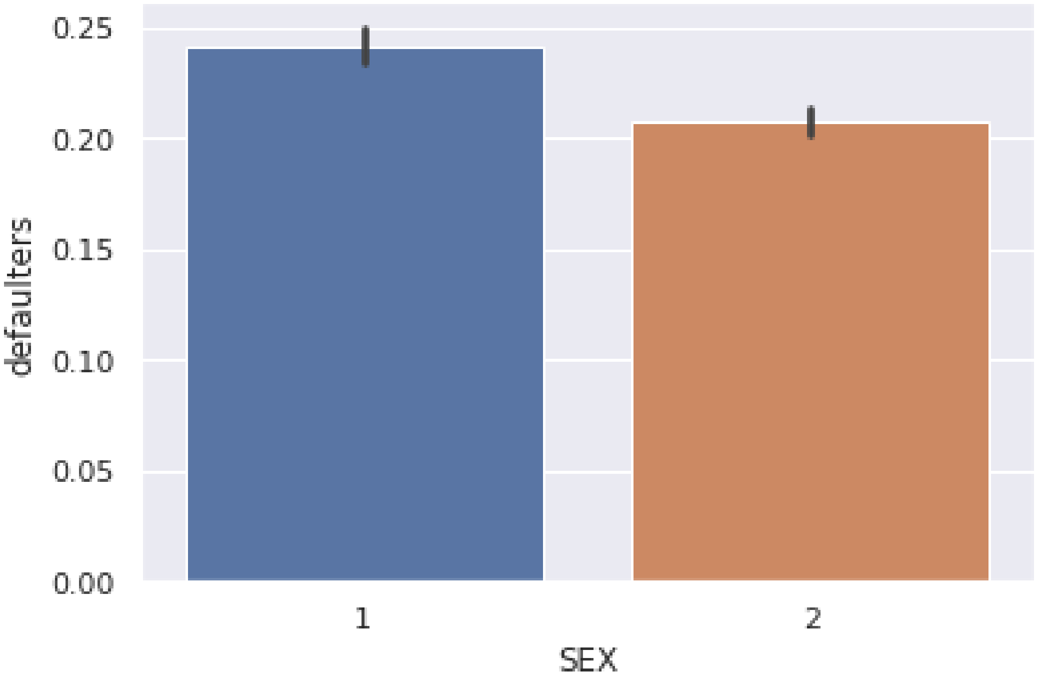
We have maximum clients from 21-30 age group followed by 31-40.
Hence with increasing age group the number of clients that will default the payment next m onth is decreasing.



Defaulters with different category

- SEX

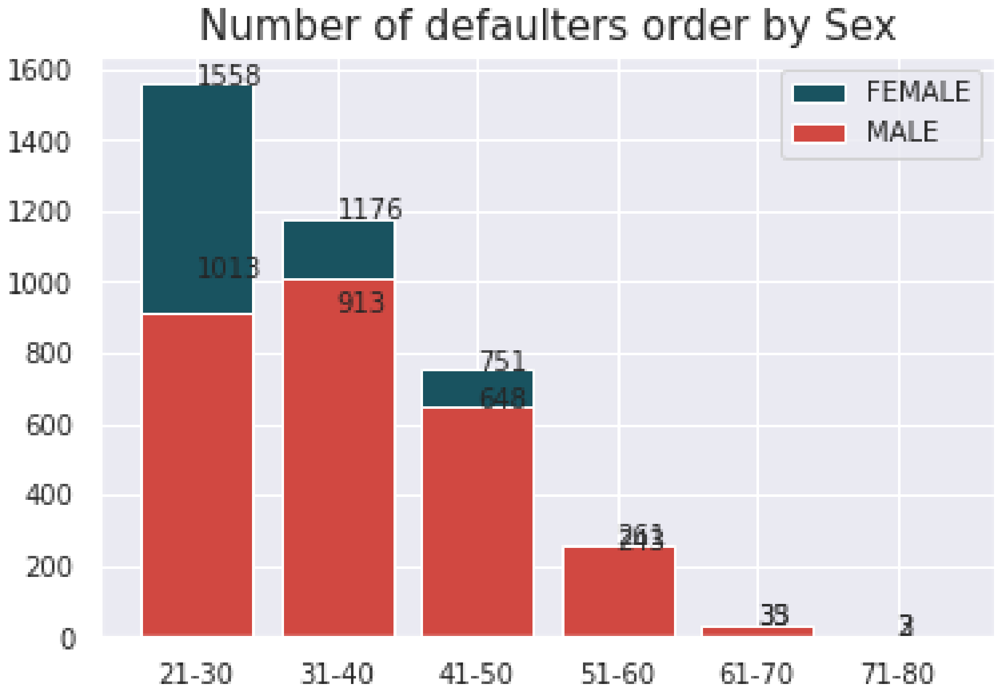
So we have more male defaulters



Defaulters with different category

- SEX

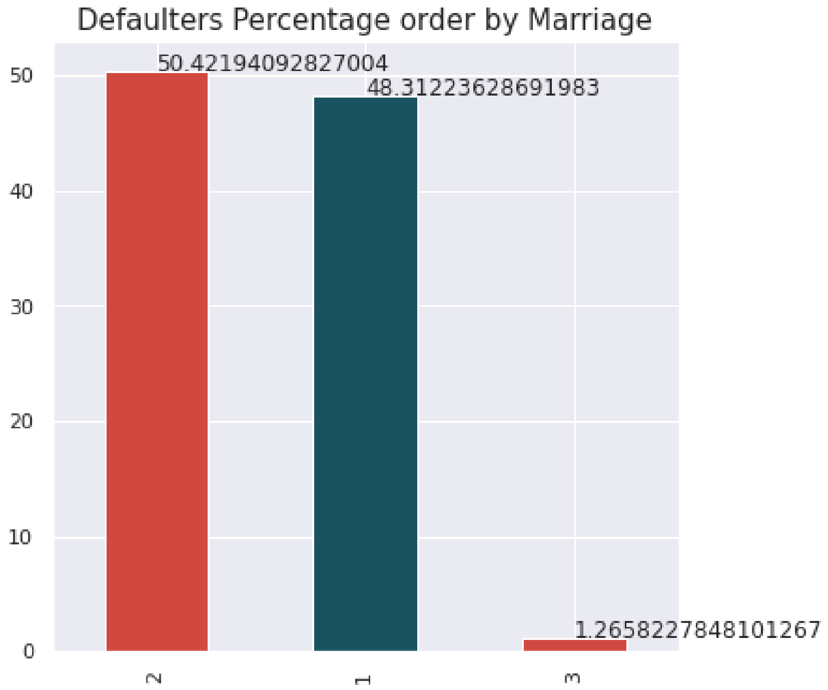
we have female defaulters more than males in some age groups ranging 21-50 years



Defaulters with different category

- MARRIAGE

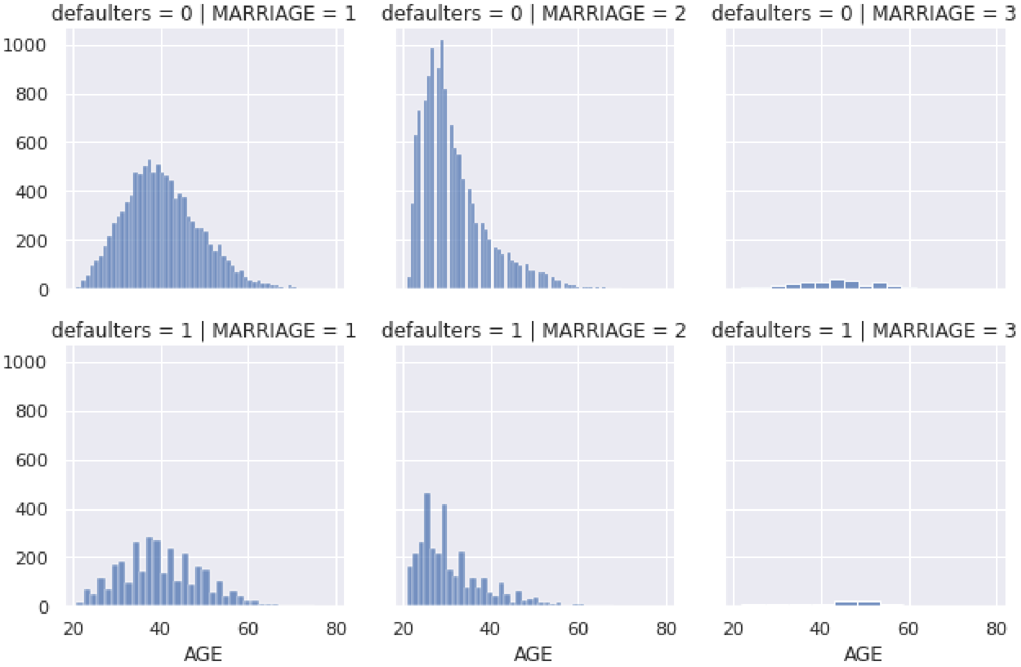
We can see there is no trend or behavior of married or unmarried people as a defaulter.



Defaulters with different category

- MARRIAGE

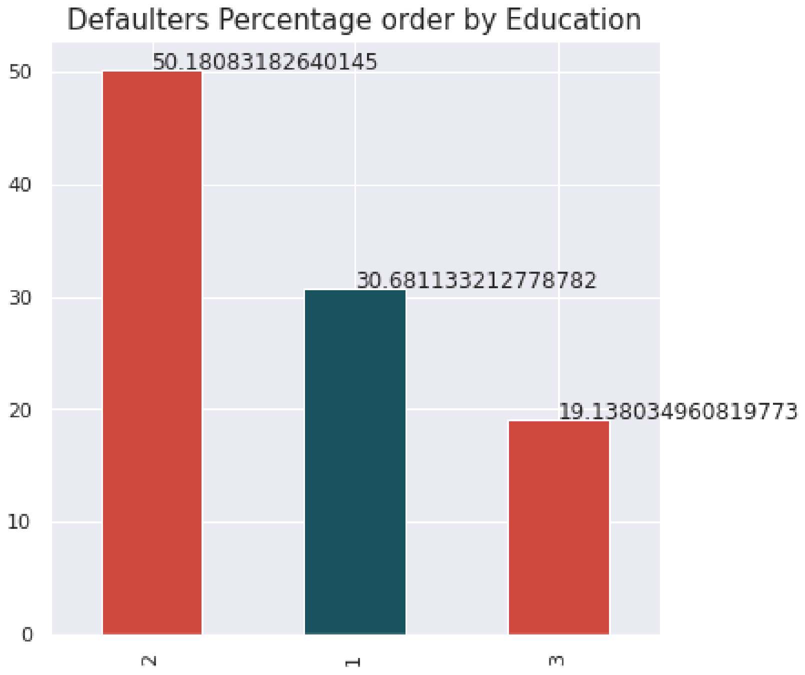
Married people between the age of 30-35 have maximum chances of being defaulters, same for unmarried, which confirms marriage is not the case, Age is.



Defaulters with different category

- EDUCATION

University level student tend to default more followed by graduate and high school students



Classification

With unbalanced Dataset

(Recall is more imp in this problem case)

Random Forest:-

- Recall– 37%
- AUC score – 66%

KNN:-

- Recall– 9%
- AUC score – 53%

XGBoost

- Recall– 38%
- AUC score – 66%

With balanced Dataset

(Recall is more imp in this problem case)

Random Forest:-

- Recall – 83%
- AUC score – 87%

KNN:-

- Recall – 82%
- AUC score – 80%

Conclusions

- we have 22% defaulters in our dataset and 77% persons are non defaulters
- The data shows that most people are of age range 20-40 and a few only from 50-60 age group
- Most people are with 10-20K of credit limit
- We have maximum clients from 21-30 age group followed by 31-40.
- With increasing age group the number of clients that will default the payment next month is decreasing
- There is no trend or behavior of married or unmarried people as a defaulter.
- we have overall more male defaulters but female defaulters are more than males in some age groups ranging 21-50 years
- Recall is the best accuracy metrics here, because if the algorithm will not detect the defaulters, that will encounter more loss to the bank
- Random Forest with SMOT gives the maximum Recall of 83% in this case