CREDIT RISK ANALYSIS

EDA Case Study

SIDDHARTH SUMAN RATH
IIIT BANGLORE
Data Science

Problem Statement

- Minimise the loan approval of those who are more likely to default the loan
- Minimise of loan rejection who are more likely to repay the loan.

- Analysis to be done in python.

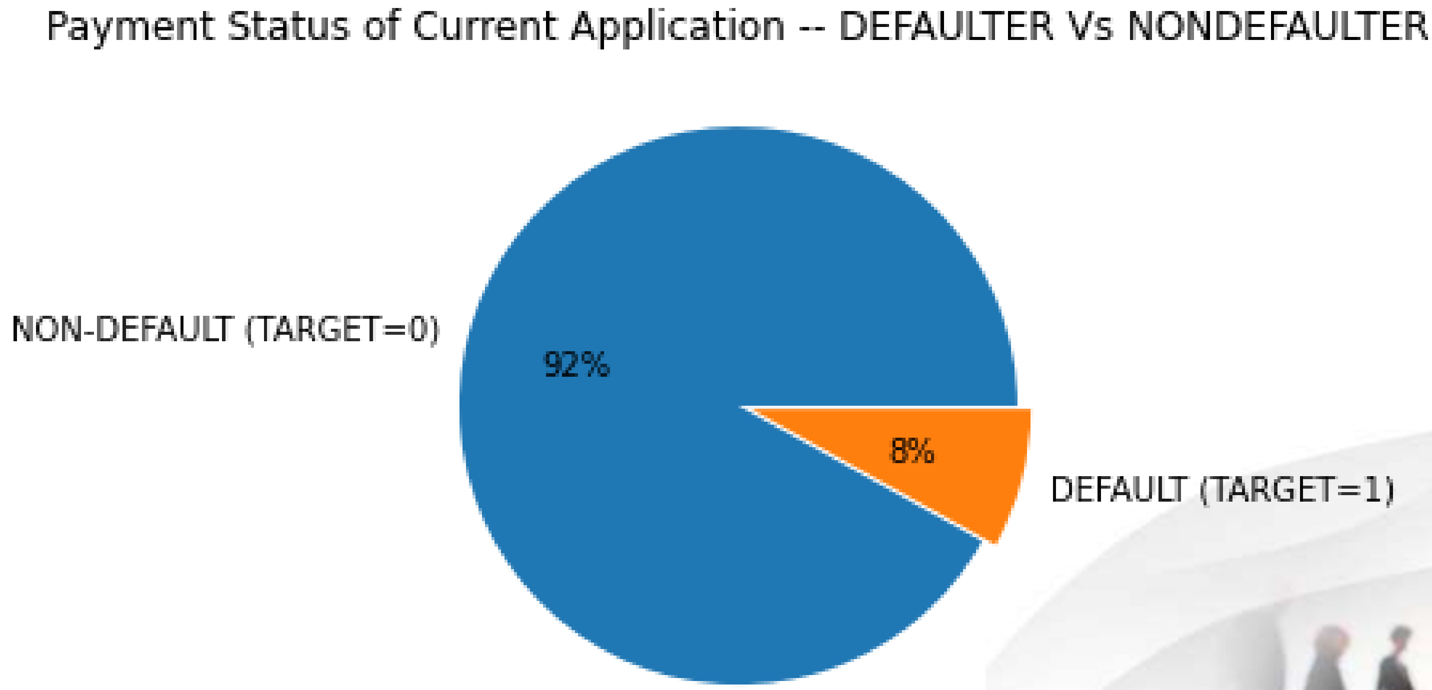
Analysis Steps

- Sanity check for both the data set.
- Identification of missing values from all coulumns
- Required action taken for the null values (e.g replacing it with mean, median, mode or zero etc.)
- Identification of outliers
- Required action taken to deal with outliers
- Creation of functions to perform similar operations on various columns
- Default percentage is calculated for different columns

- Data imbalance is checked
- Univariate analysis is done on few columns
- Bivariate analysis is done on few columns
- Correlation analysis done on all numeric columns
- Top 10 correlation for defaulters and non-defaulters are found out
- Graphs are plotted for visual analysis

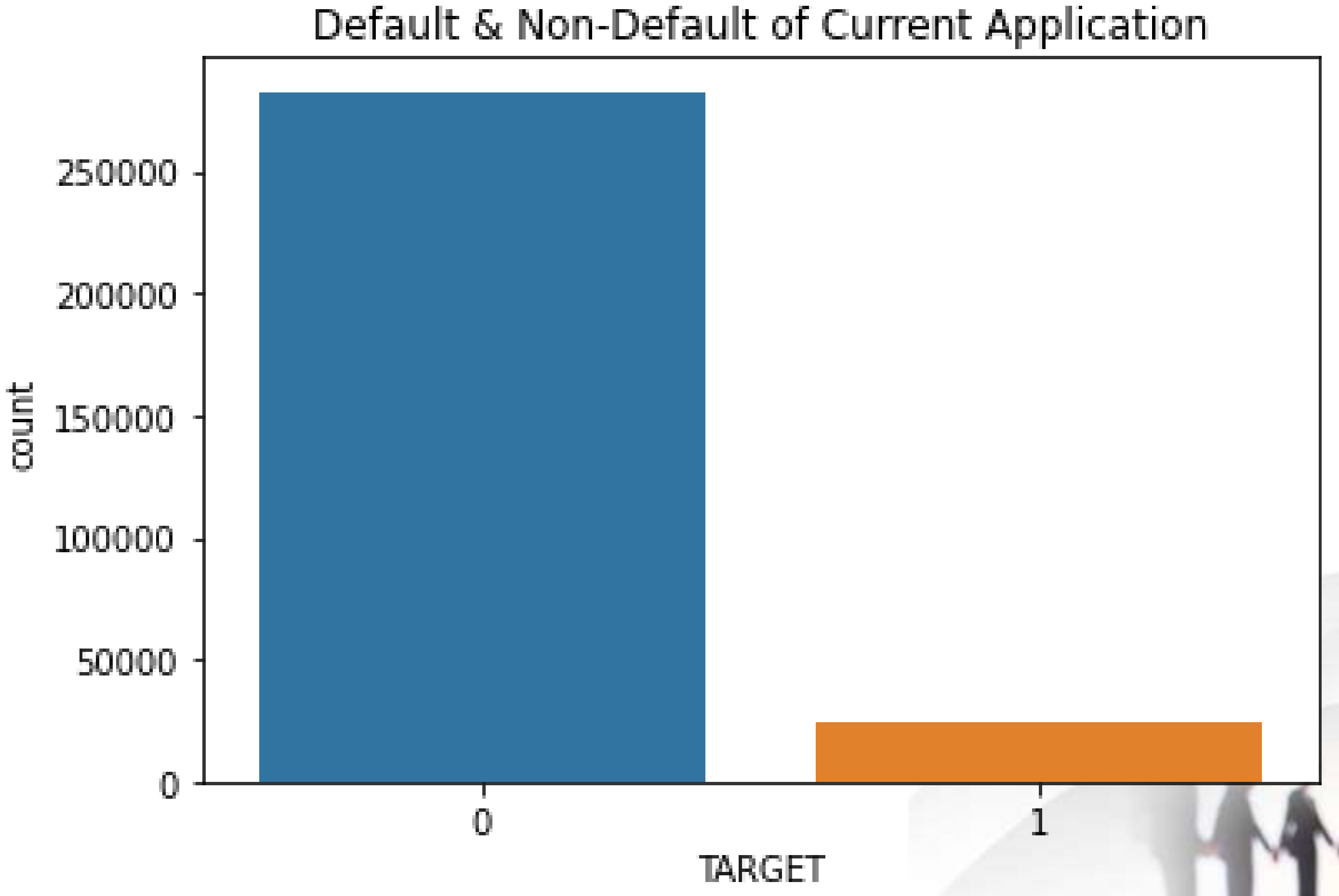
Defaulter & Non-Defaulter Ratio

- Defaulters are are around 8 %



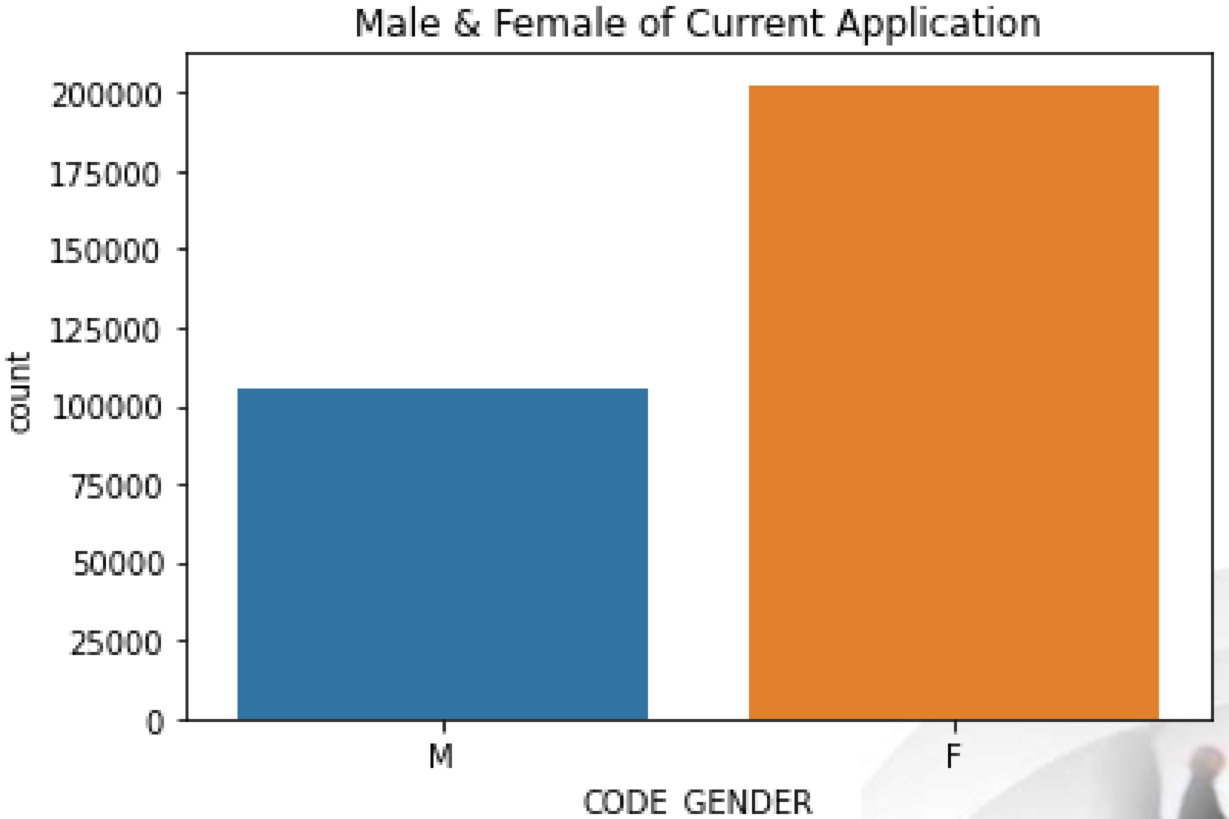
Imbalance in Target

Graph showing imbalance in Target column where defaulters are around 22500

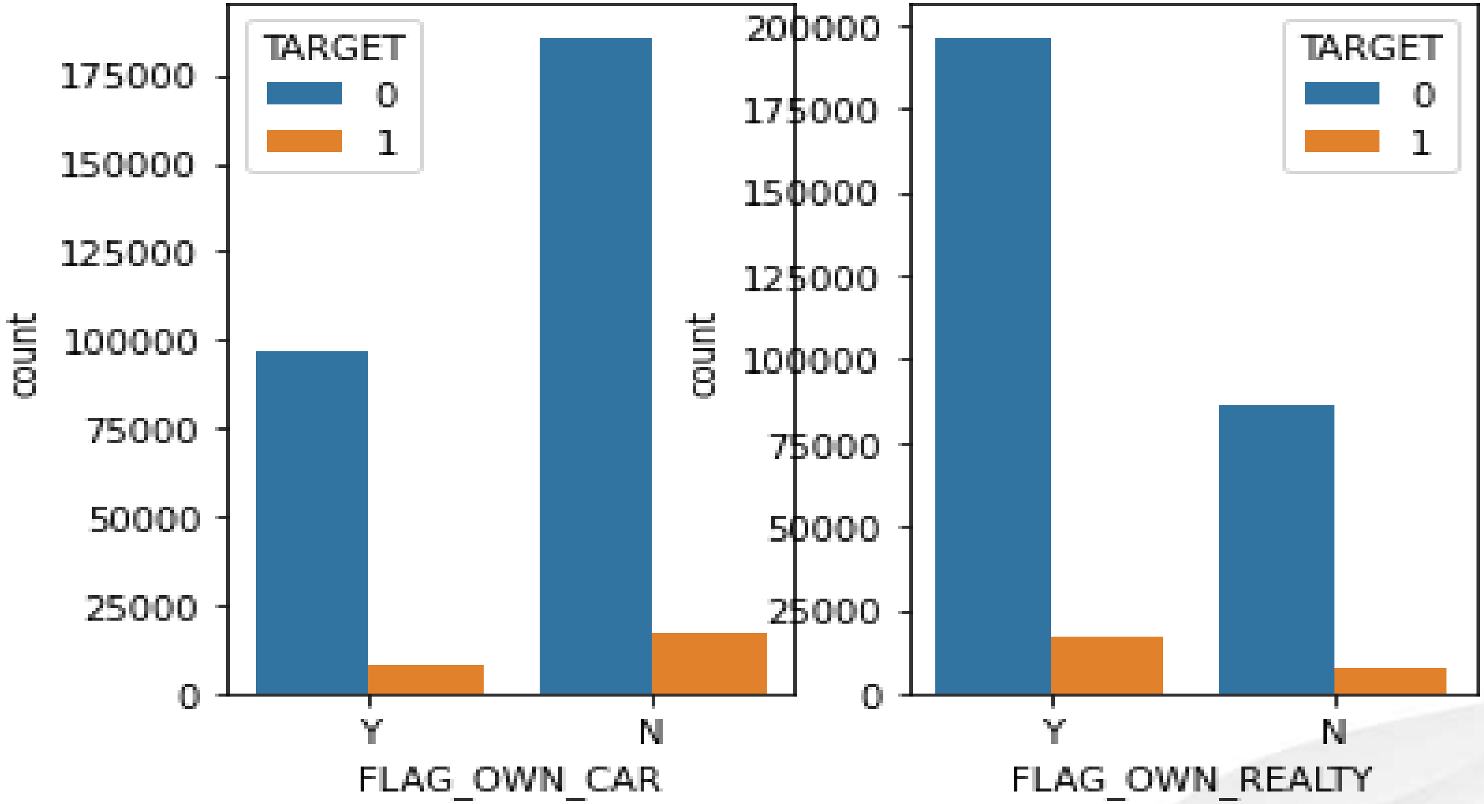


Imbalance in Gender

2 lakhs female and 1.1 lakhs male applied for loan

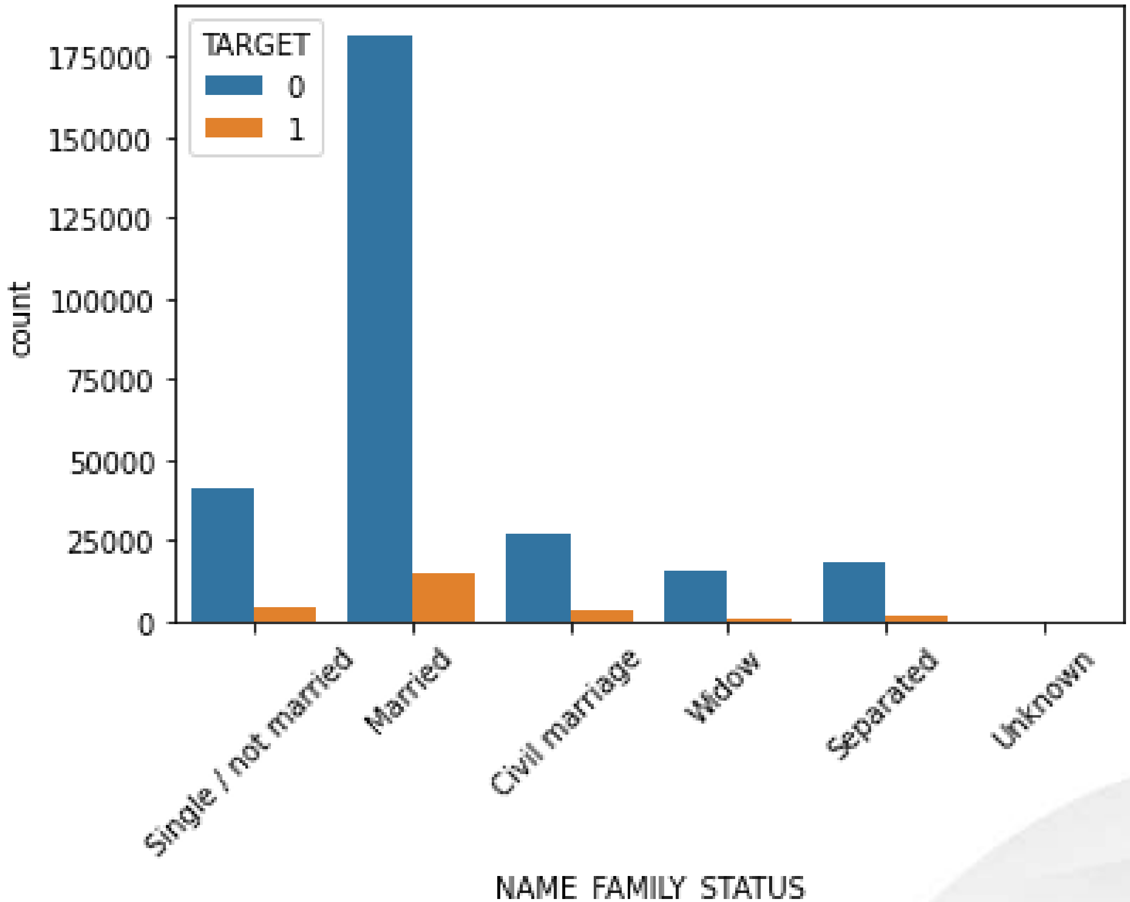


Defaulters Ratio by Cars & Realty



- People not owning reality and car and have a slightly higher default rate than the people who own reality and car

Defaulter & Non-Defaulter Proportion (Family Status)

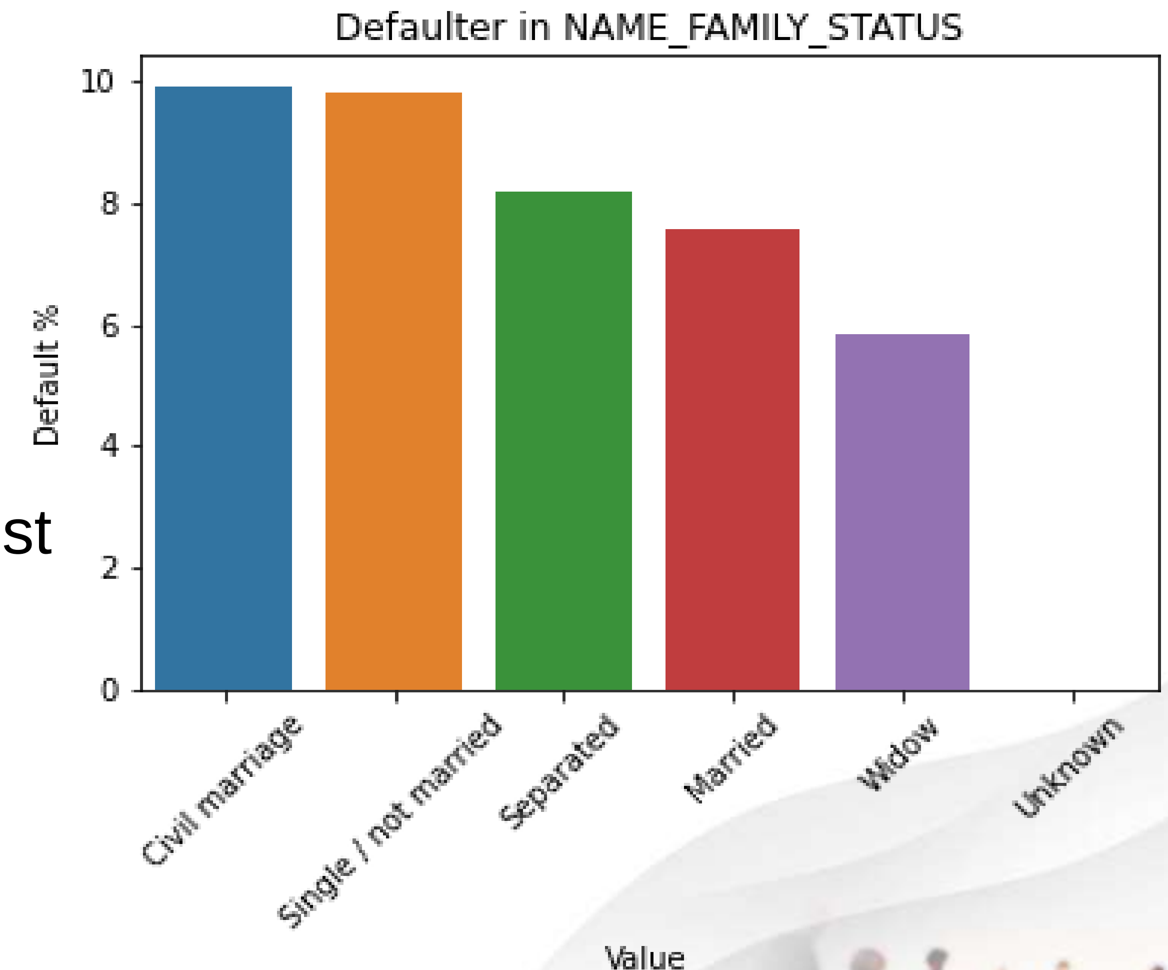


Graph showing defaulters and non-defaulters in each category of family status column

Defaulter Percentage

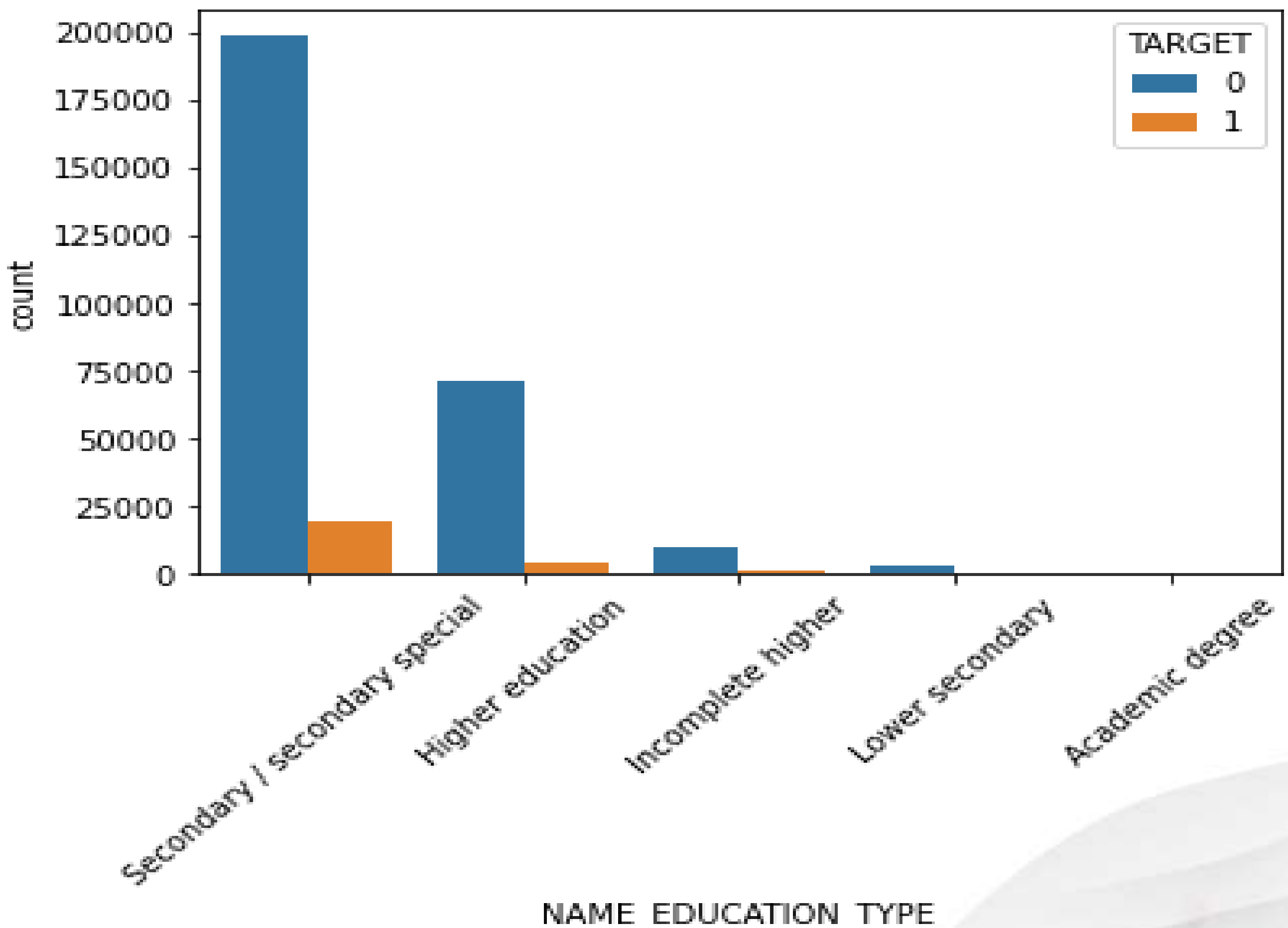
- Default rate is highest for Civil Marriage and Single applicants

- Default rate is lowest for widow and married applicants



Graph showing defaulter percentage in each category of family status

Defaulter & Non-Defaulter Proportion (Education Type)

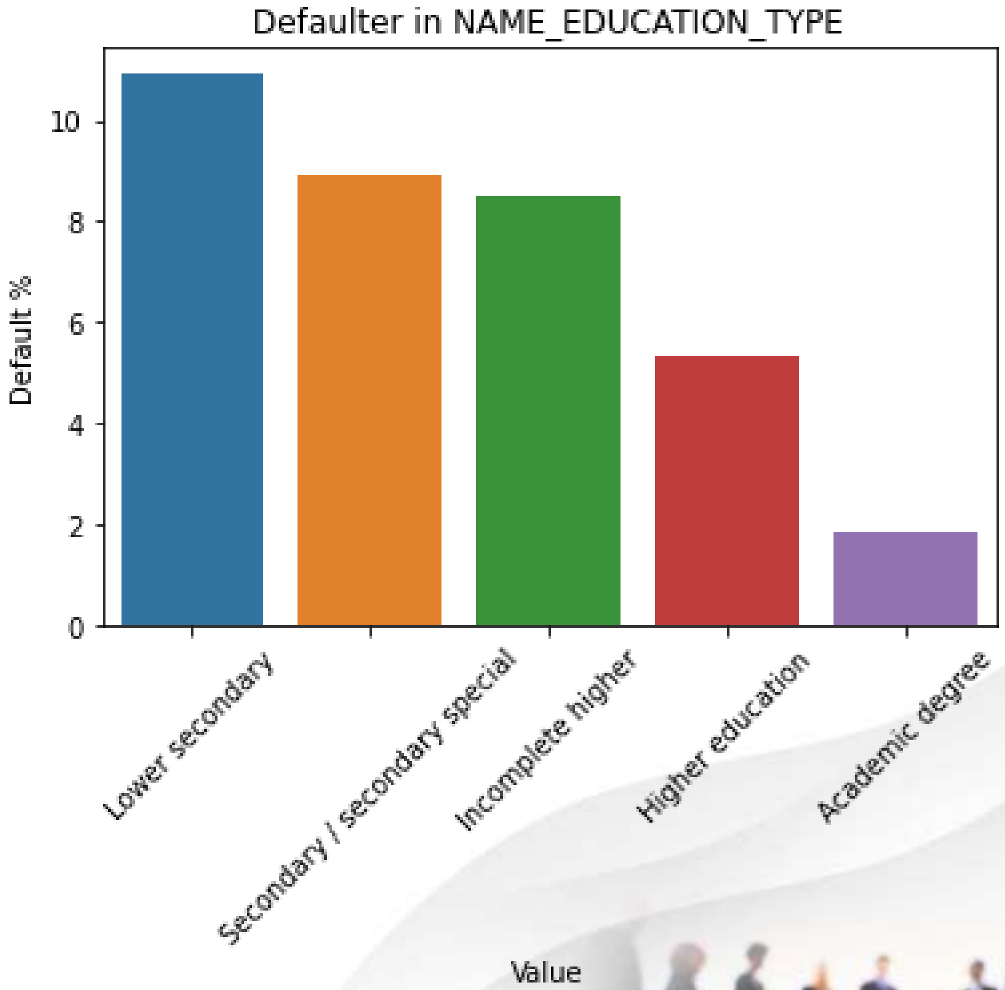


Graph showing defaulters and non-defaulters in each category of education type column

Defaulter Percentage

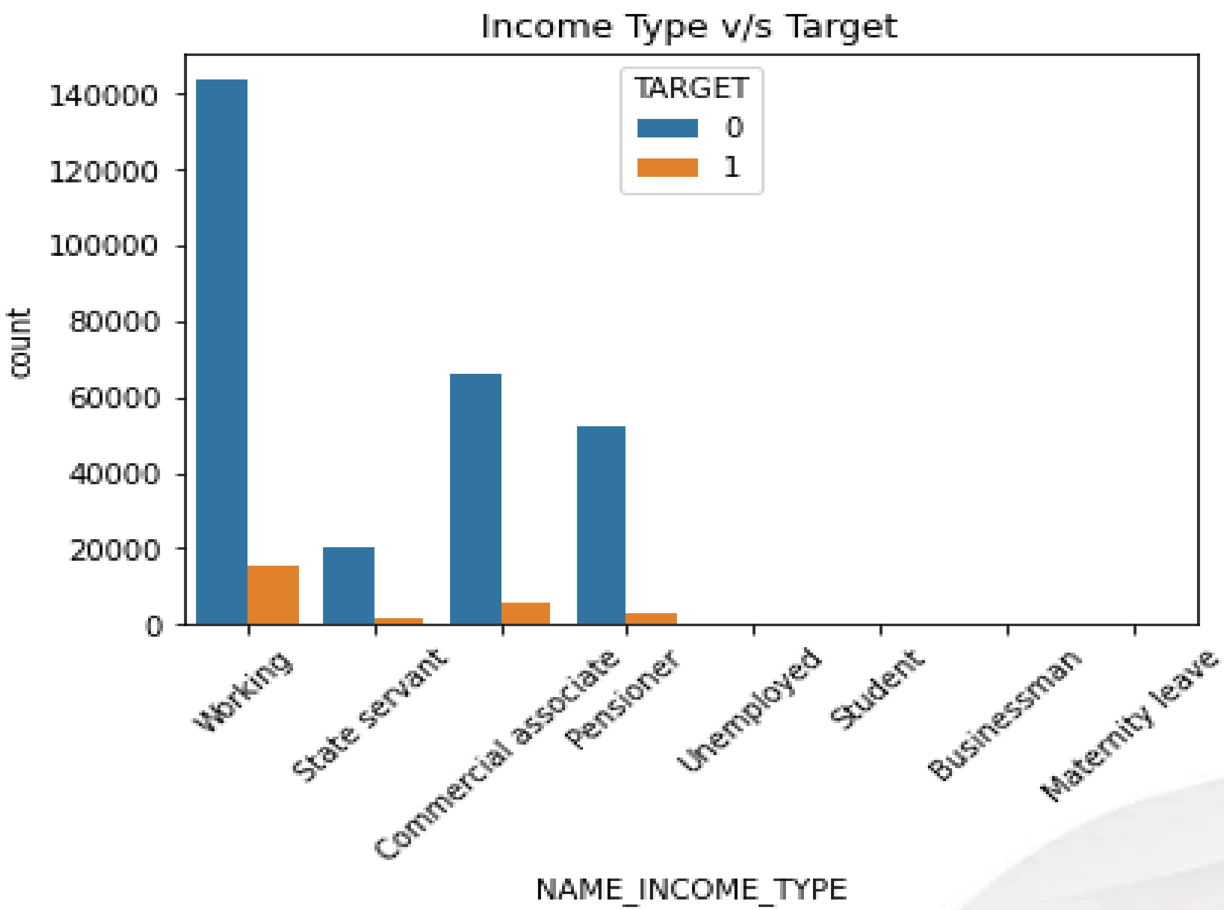
- Default rate is highest for Lower secondary

- As education label increases the default percentage also decreases



Graph showing defaulter percentage in each category of education type

Defaulter & Non-Defaulter Proportion (Income Type)

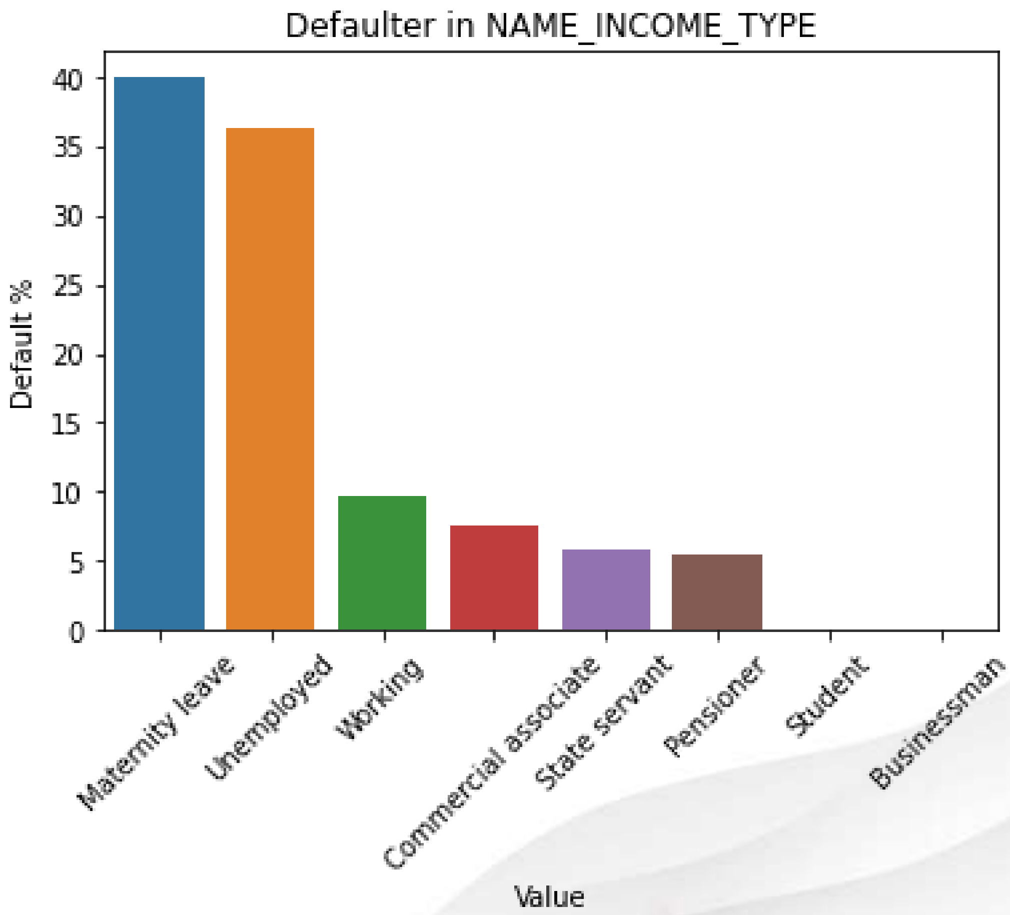


Graph showing defaulters and non-defaulters in each category of income type column

Defaulter Percentage

- Default rate is highest for Maternity Leave and Unemployed

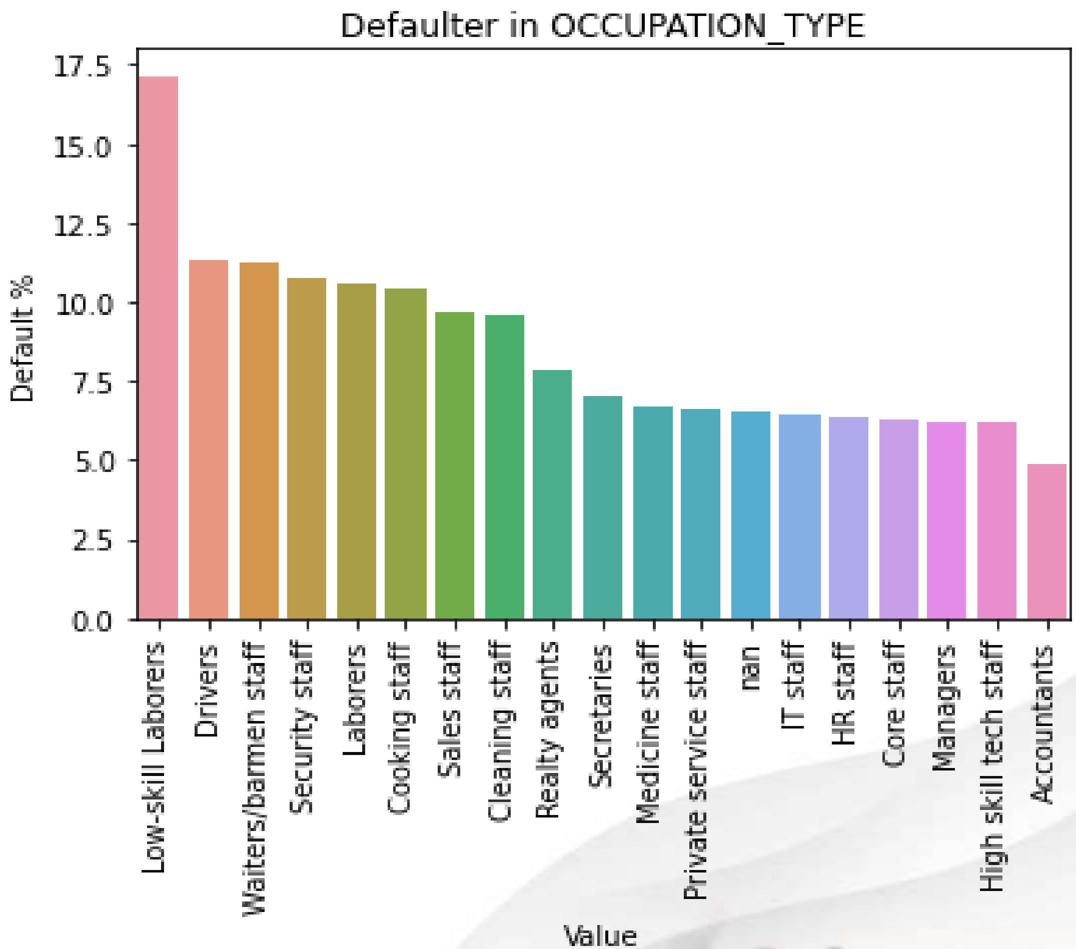
- Default rate is nearly zero for Businessman and Student



Graph showing defaulter percentage in each category of income type

Defaulter Percentage

- Low skilled labourers have very high rate of defaulters in comparison to other occupations

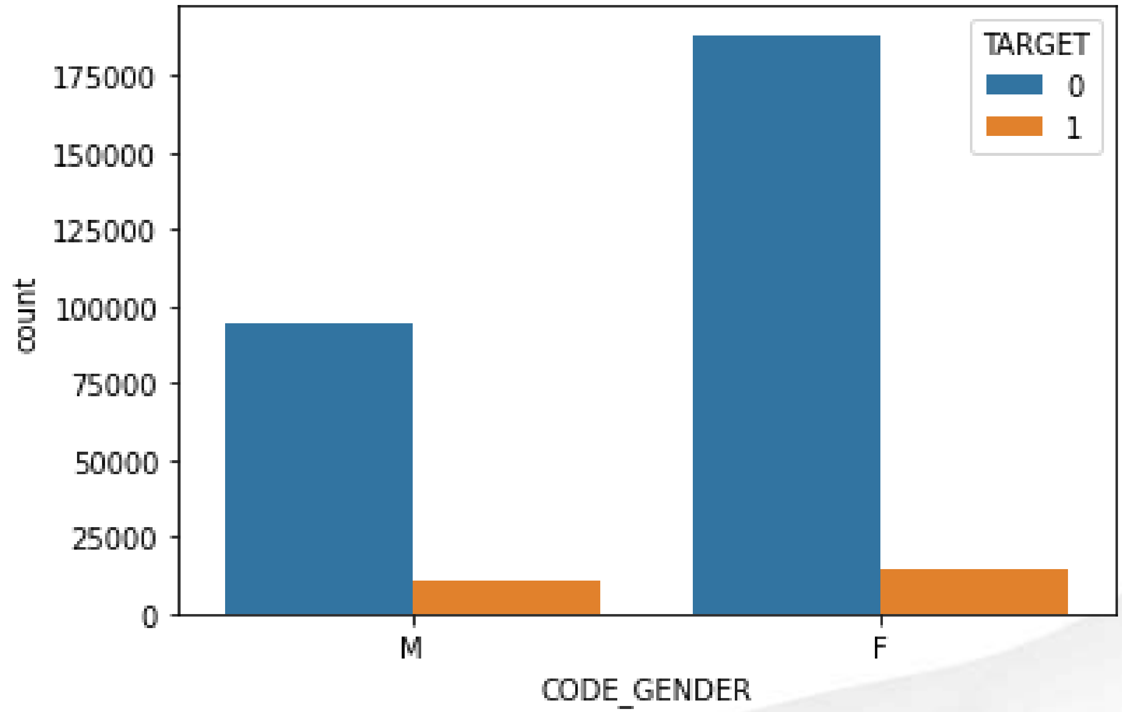


Graph showing defaulter percentage in each category of occupation type

Defaulters Proportion by Gender

- Female applicants are more than male applicants

- Defaulter percentage is higher for male applicants



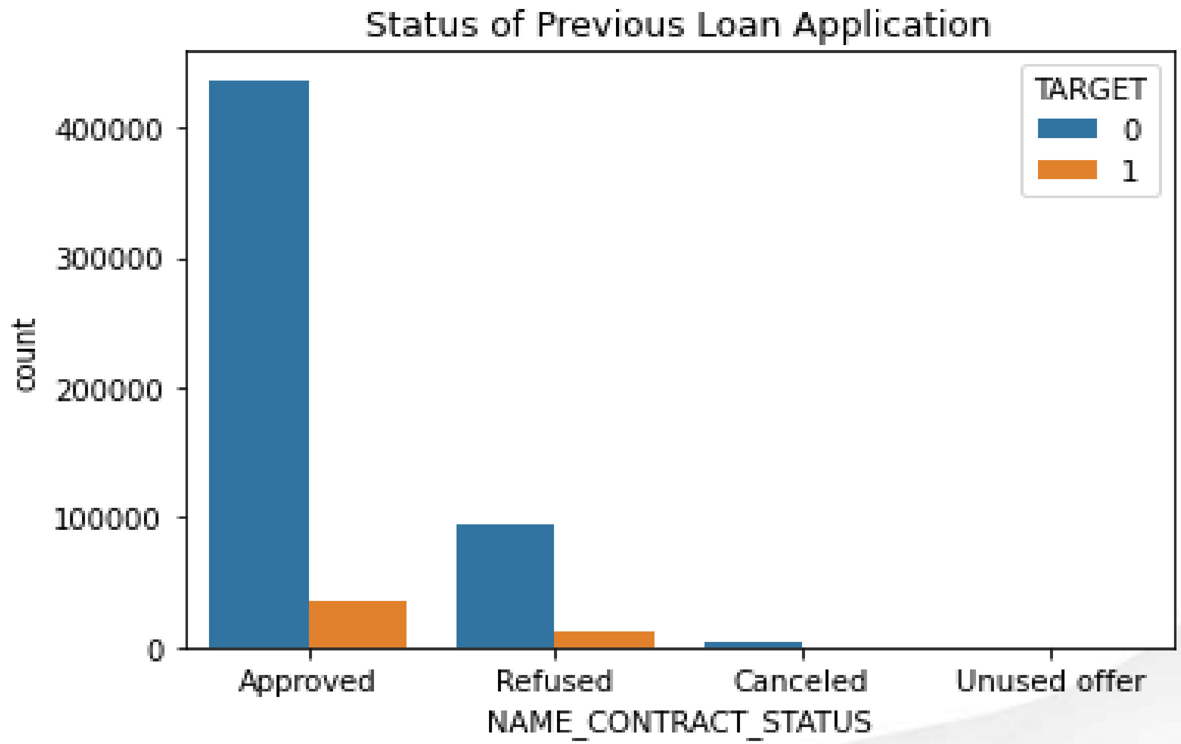
Graph showing defaulters and non-defaulters proportion by gender

Defaulters Ratio by Contract Status

- The applicants whose previous loans were approved are more likely to pay current loan in time, than the applicants whose previous loans were rejected.

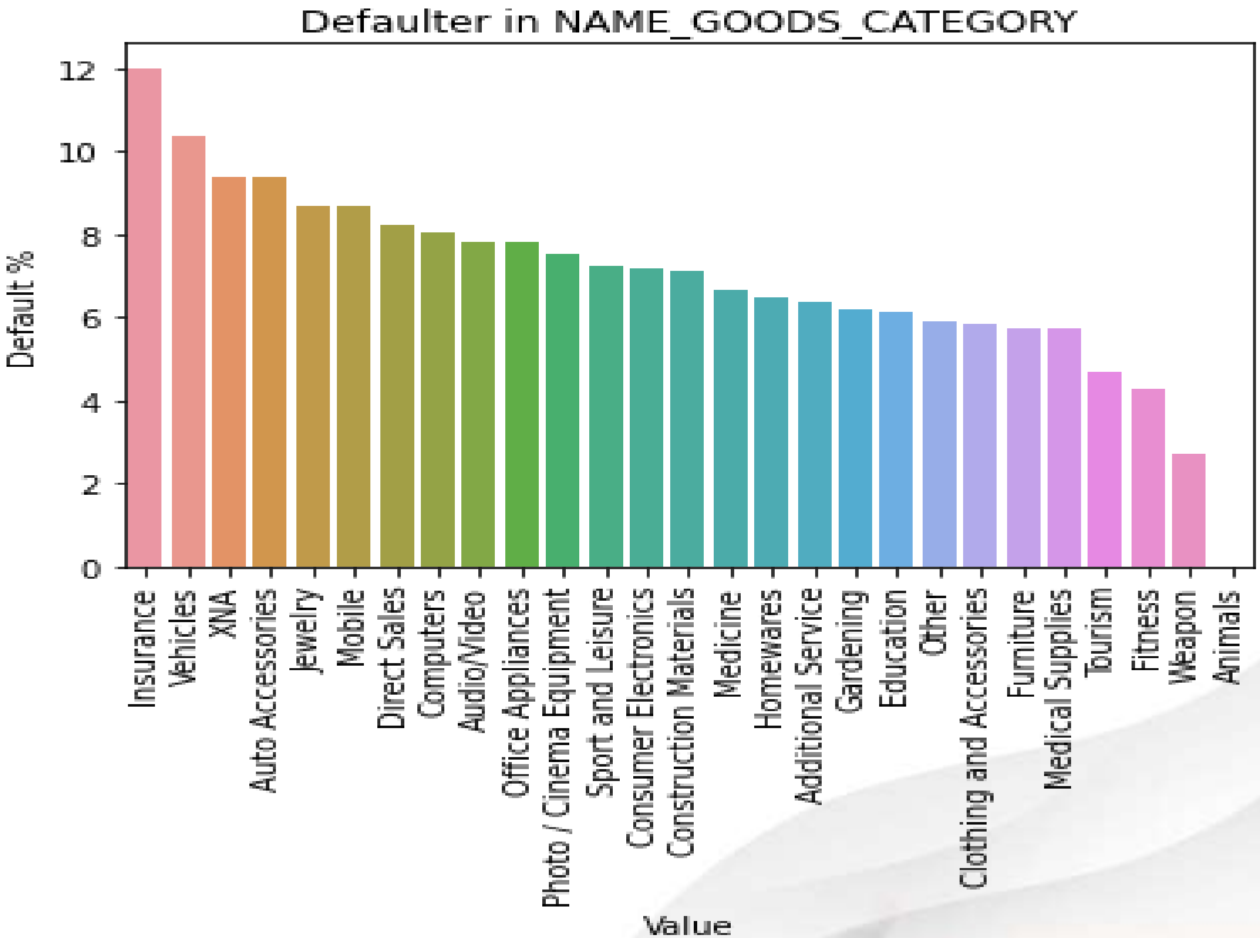
- 7% of the previously approved loan applicants that defaulted in current loan

- 90 % of the previously refused loan applicants that were able to pay current loan



Defaulter Percentage

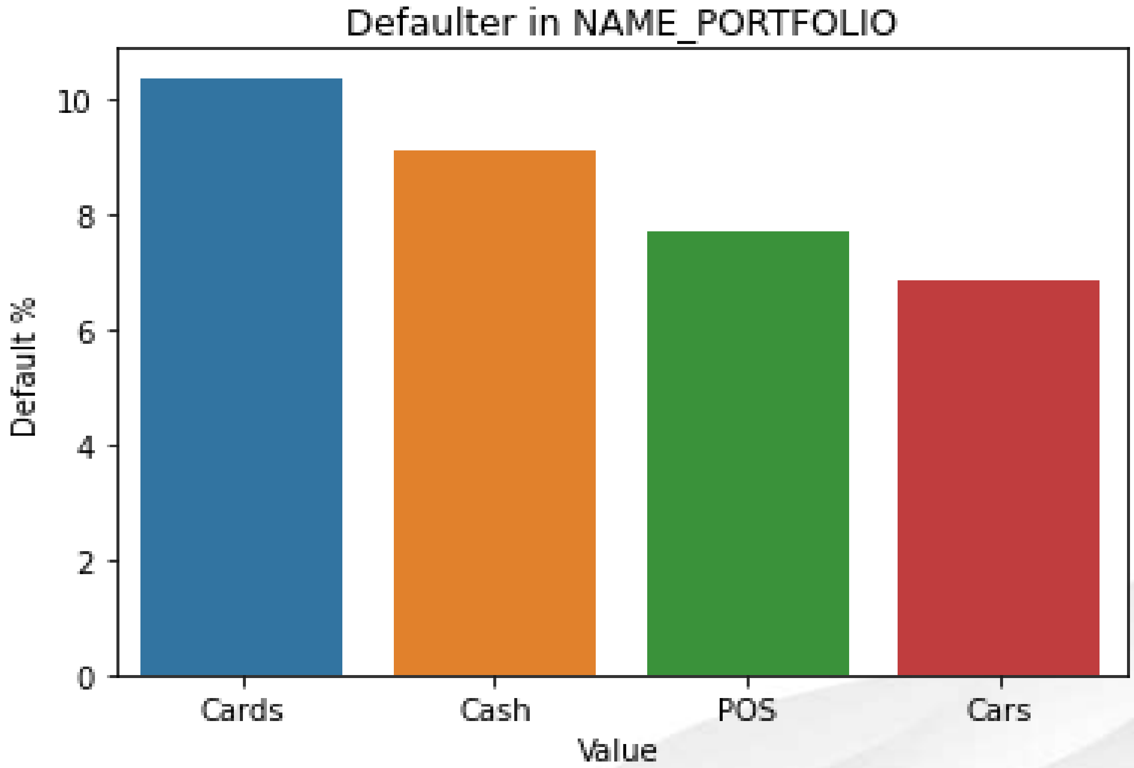
- People with insurance and vehicles are among highest defaulters



Graph showing defaulter percentage in each category of goods category

Defaulter Percentage

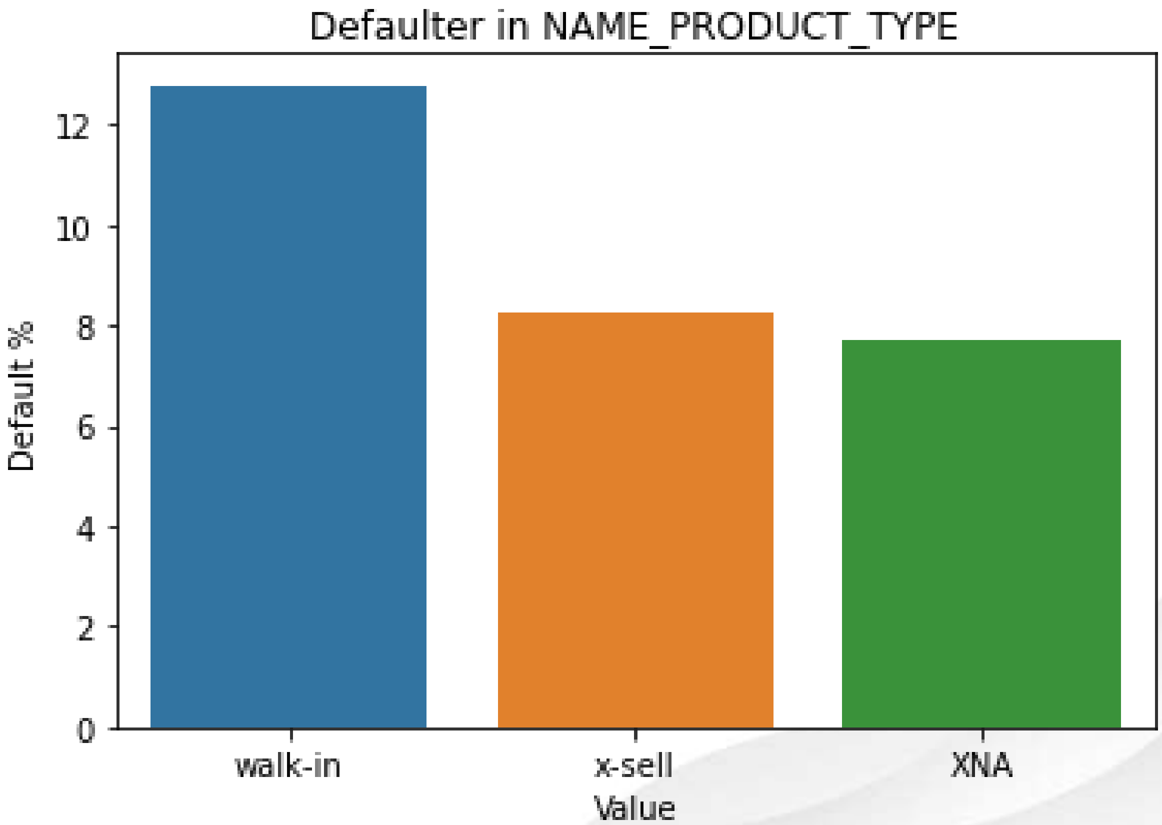
- People with cards are among highest defaulters



Graph showing defaulter percentage in each category of portfolio

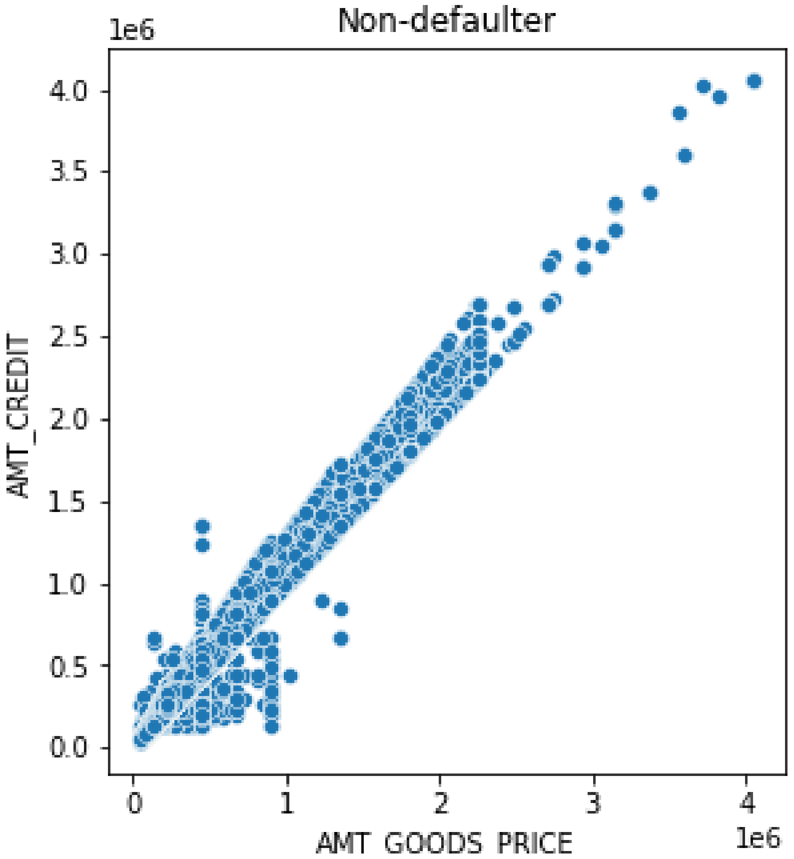
Defaulter Percentage

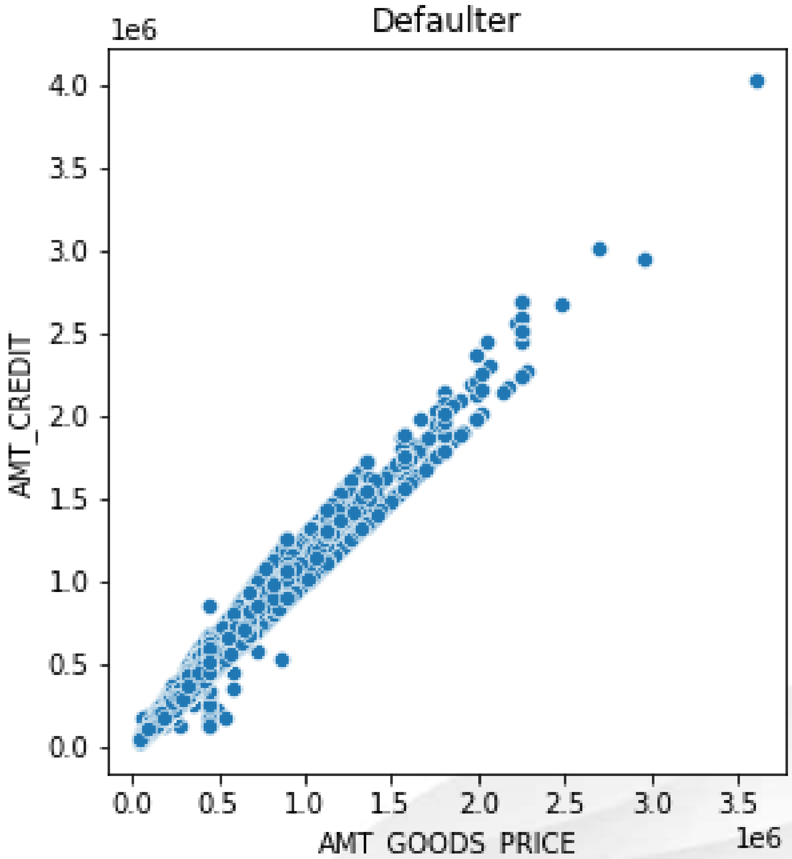
- People with walk-in are among highest defaulters



Graph showing defaulter percentage in each category of product type

Bivariate analysis between goods price & amt credit



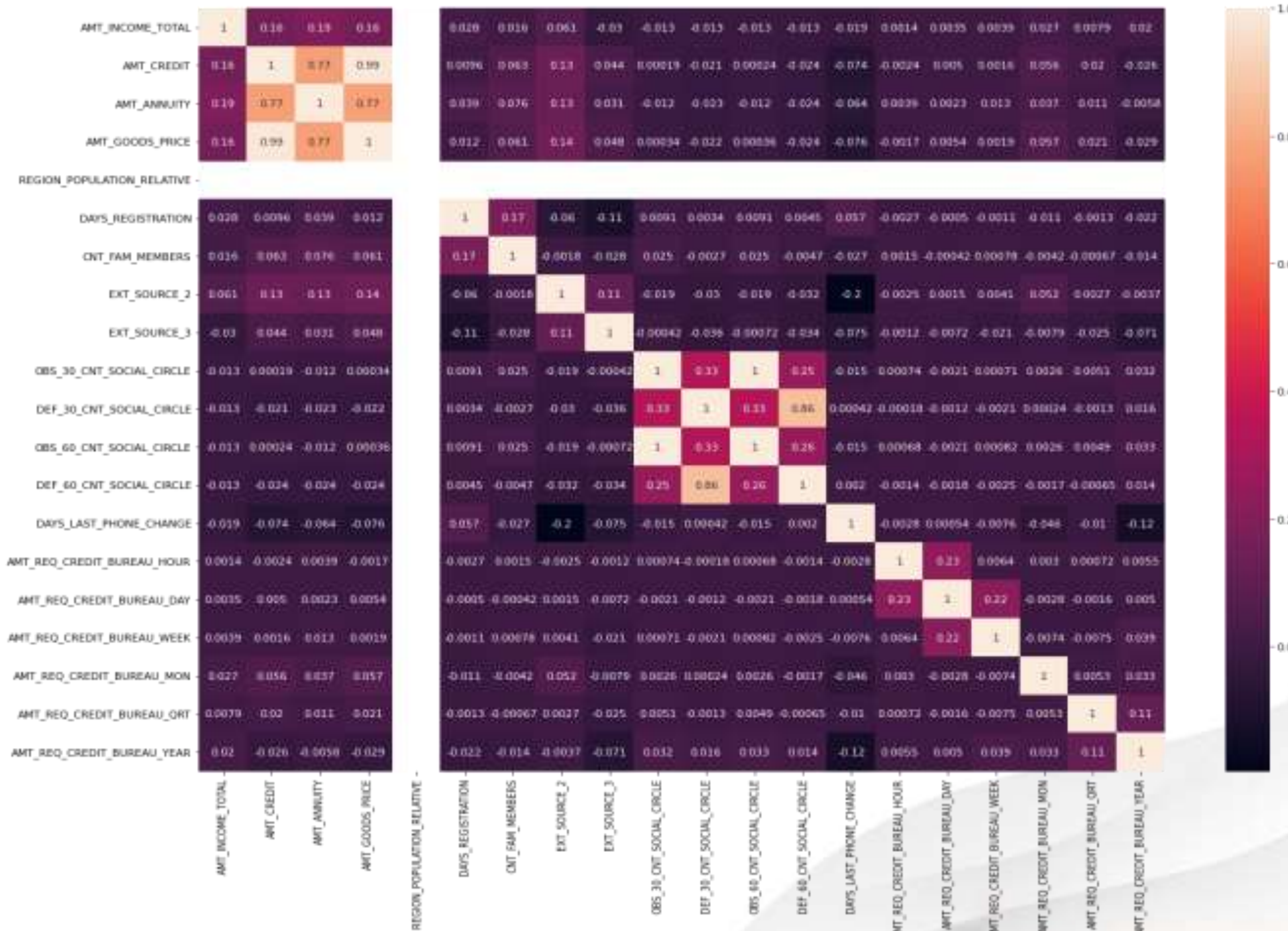


- Defaulters are less for lower range of credit amount and goods price amount

Correlation Graph

- Light colour represent Positive correlation

- Dark colour represent Negetive correlation



Graph showing correlation between all numeric columns

Top Correlation for Defaulters & Non-Defaulters

- Top 10 correlated columns for defaulters

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998269
AMT_GOODS_PRICE	AMT_CREDIT	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	0.868994
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.847885
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.778540
AMT_ANNUITY	AMT_GOODS_PRICE	0.752296

- Top 10 correlated columns for non-defaulters

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998508
AMT_CREDIT	AMT_GOODS_PRICE	0.987024
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950148
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878570
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859289
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830381
AMT_ANNUITY	AMT_GOODS_PRICE	0.776421
AMT_CREDIT	AMT_ANNUITY	0.771296

Conclusion

Hence, we conclude that in EDA of banking Data :

- There are 8% defaulters in current application
- Bank lends more loan to female and it is safer option as male defaulter percentage is more than female
- Approving loans who have cars & realty is safer
- Approving loans for widow & married people is safer
- As education lebel increases lending loans are safer
- It is risker to approve loan for people with maternity leave and unemployed
- It is risker to approve loan with low skills
- It is risker to approve loan for those have rejected loan history

Thank you !

Relationships beyond Banking