

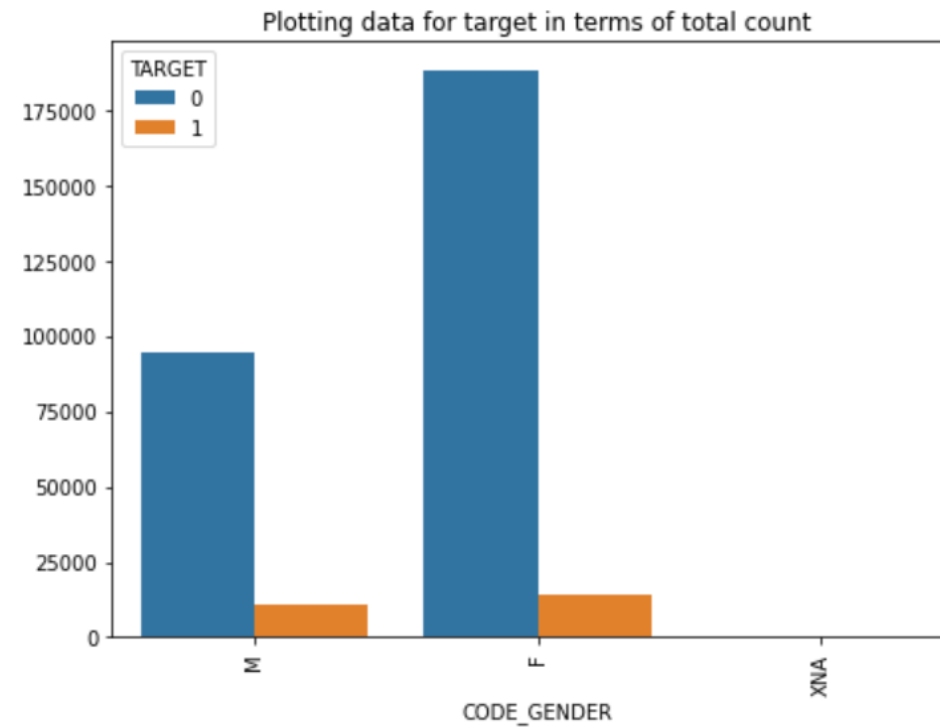
CREDIT EDA CASE STUDY

BY DHARMESHNATH MISHRA

UNIVARIATE ANALYSIS ON APPLICATION DATA

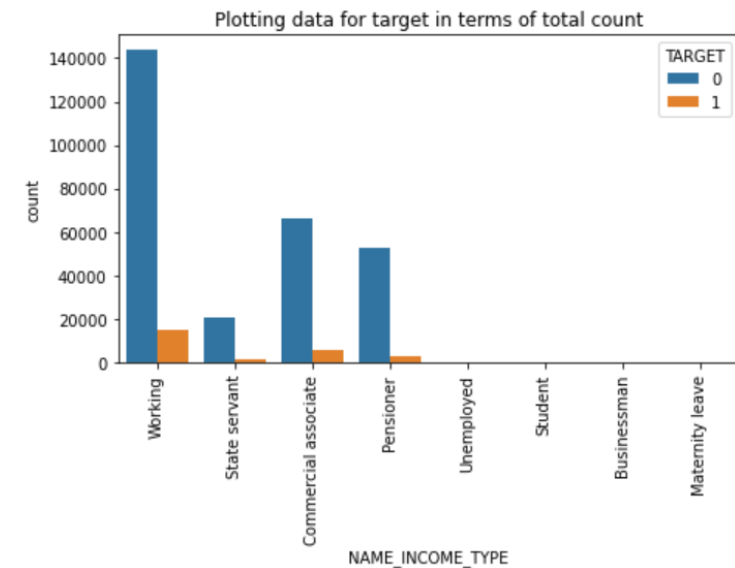
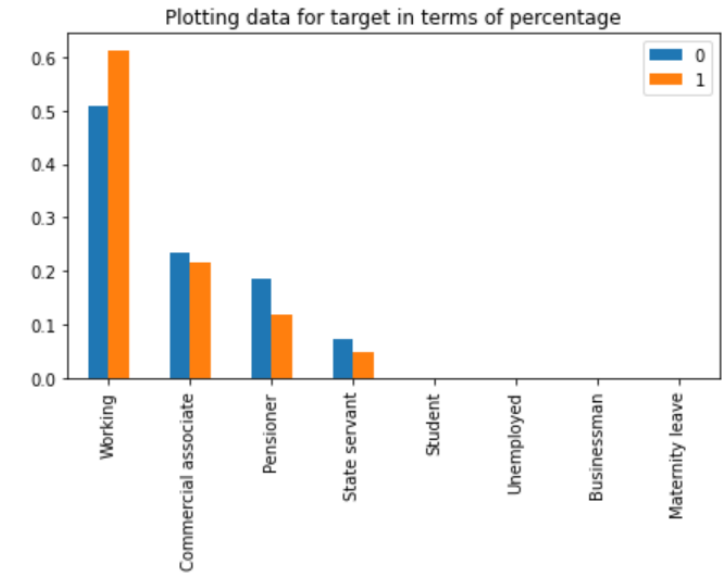
CODE_GENDER

- Though females are taking most of the loan, but the rate of defaulting is much lower than their male counterparts.



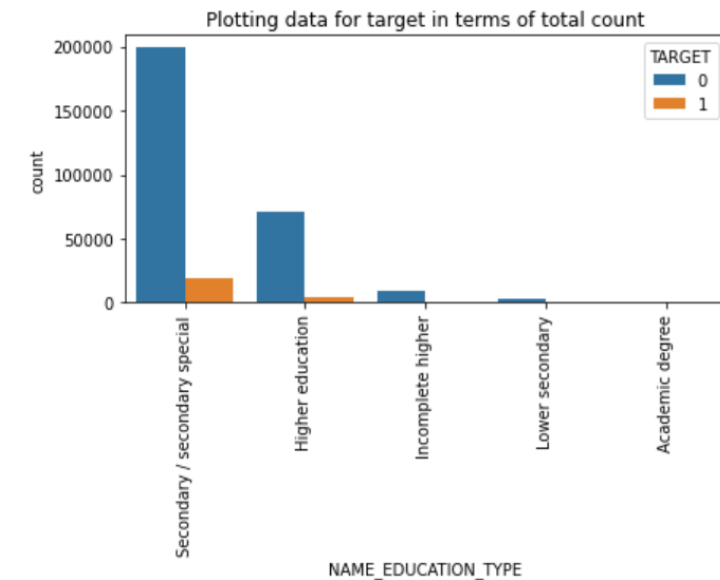
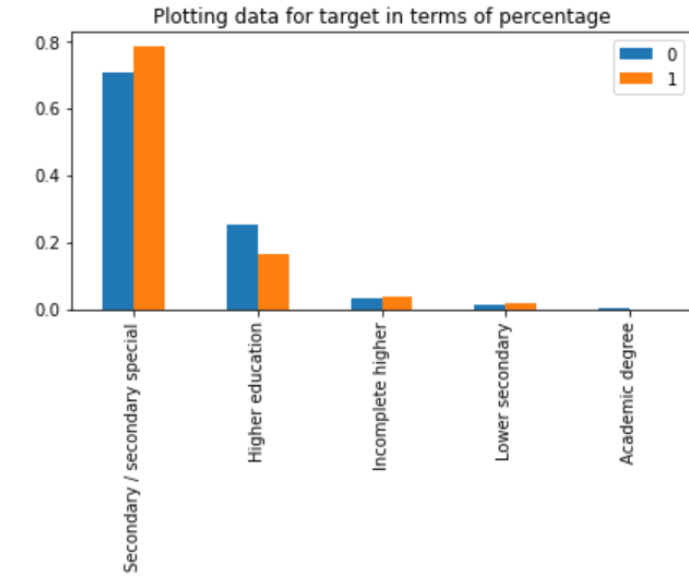
NAME_INCOME_TYPE

- Since working people are highest amongst all in that 50% working people are defaulter.
- Interesting thing to note here is 3rd highest loan applicants are from pensioner category.



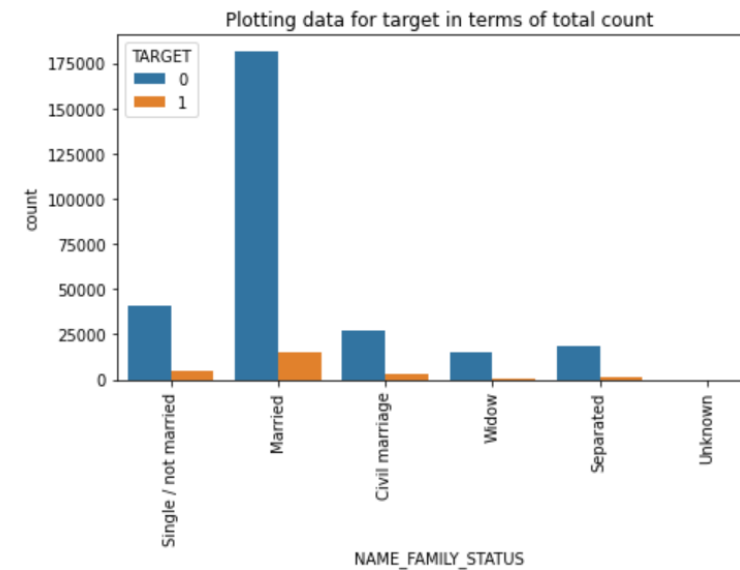
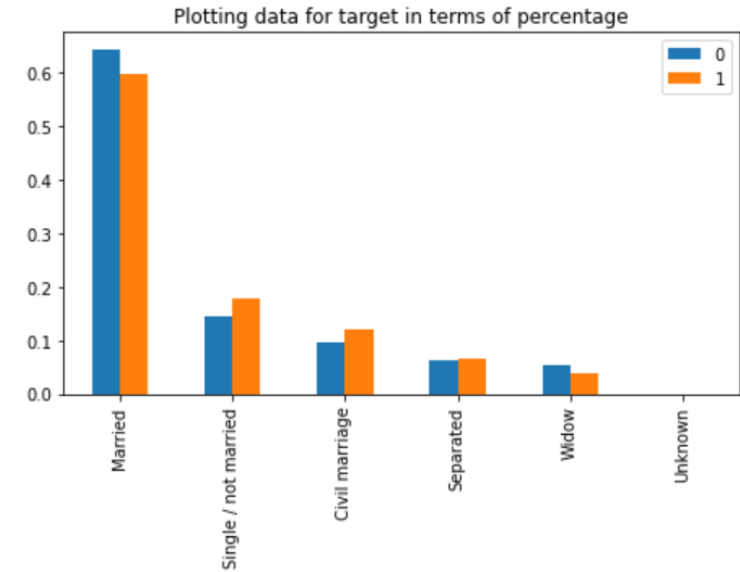
NAME_EDUCATION_TYPE

- In both Target 0 and 1, applicants with Secondary Education has applied for loans more than others. 90% of defaulting payments are from applicants with secondary education.
- further analysis needed



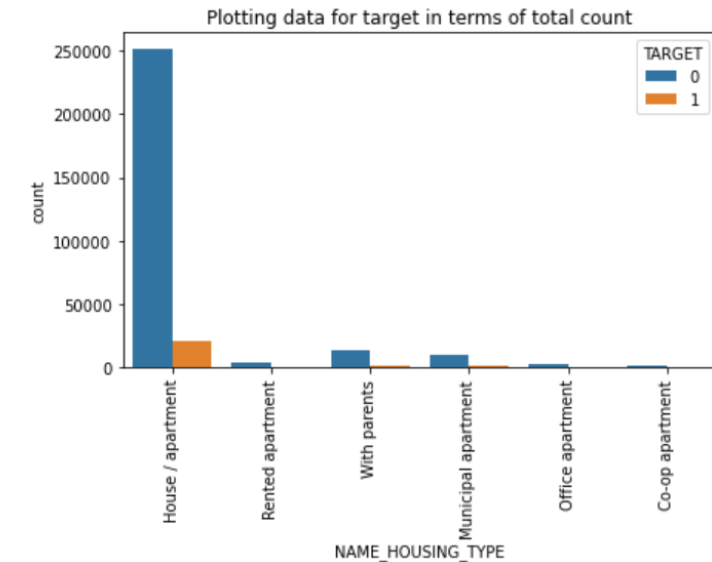
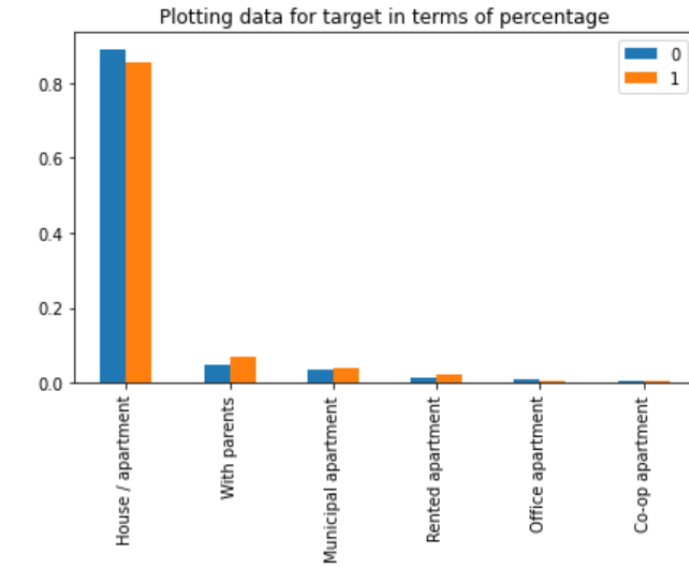
NAME_FAMILY_STATUS

Almost 60% of married individuals have defaulted on payments



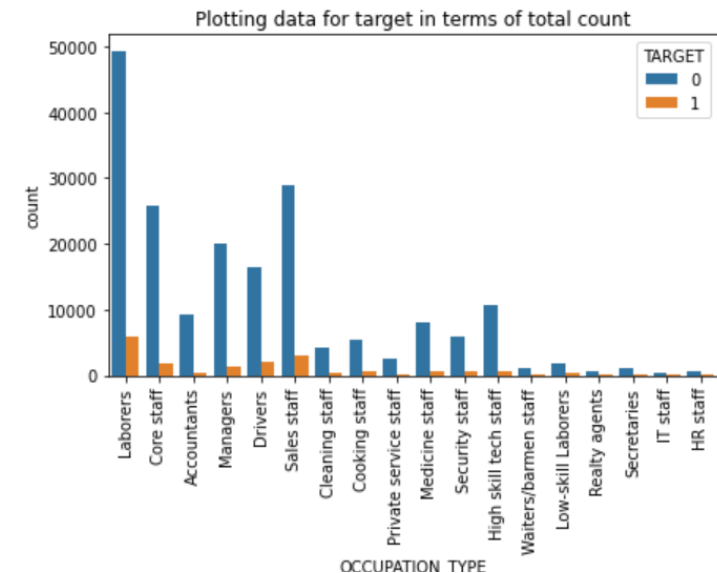
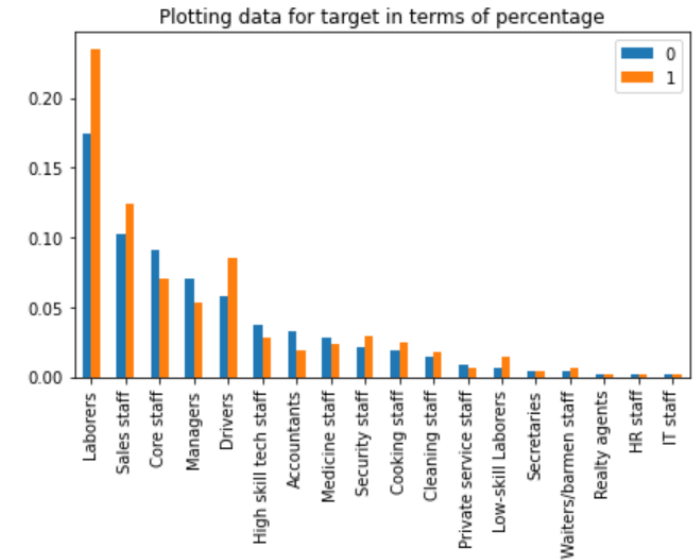
NAME_HOSUING_TYPE

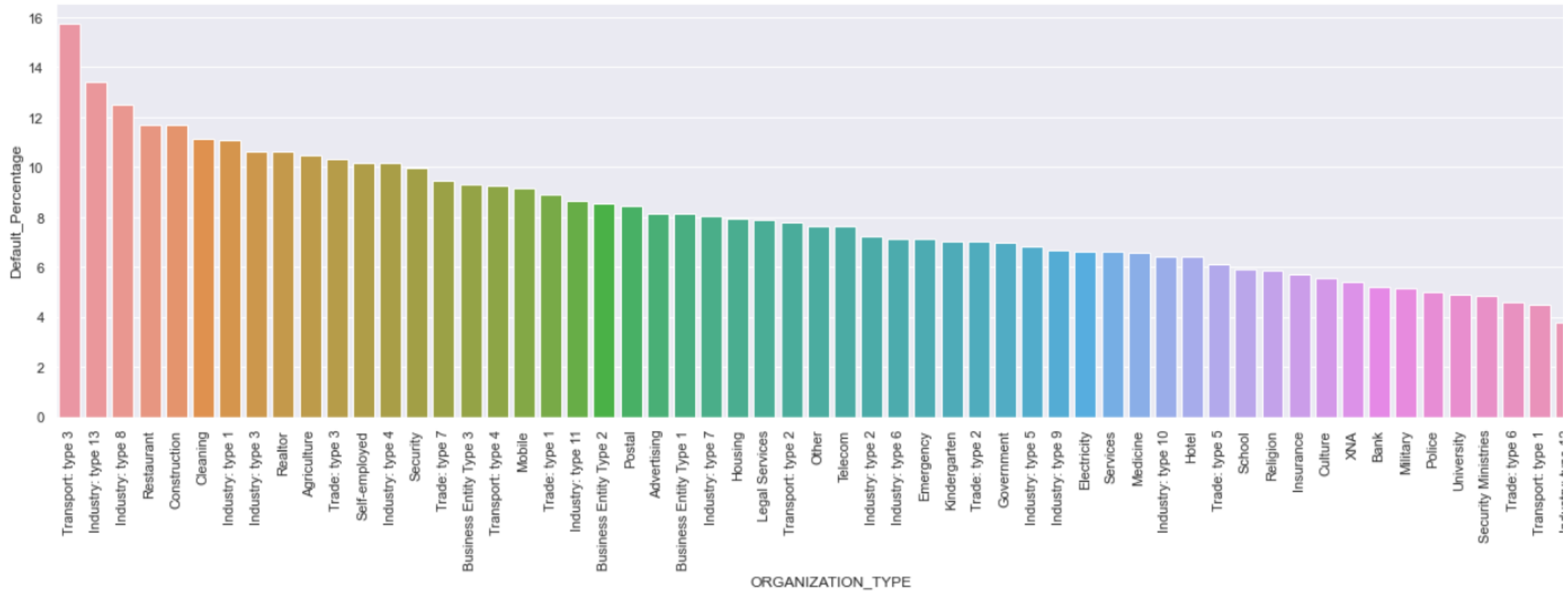
- Since 85-90% in Target 0 and Target 1 applicants are staying in "House/apartment". Indicating, this is not a parameter that can influence payment default.



OCCUPATION_TYPE

- Labourers, sales staff, core staff, drivers accounts for 50% of defaulters. Out of which Labourers has the highest percentage of applicants too, followed by sales staff and Drivers, respectively.



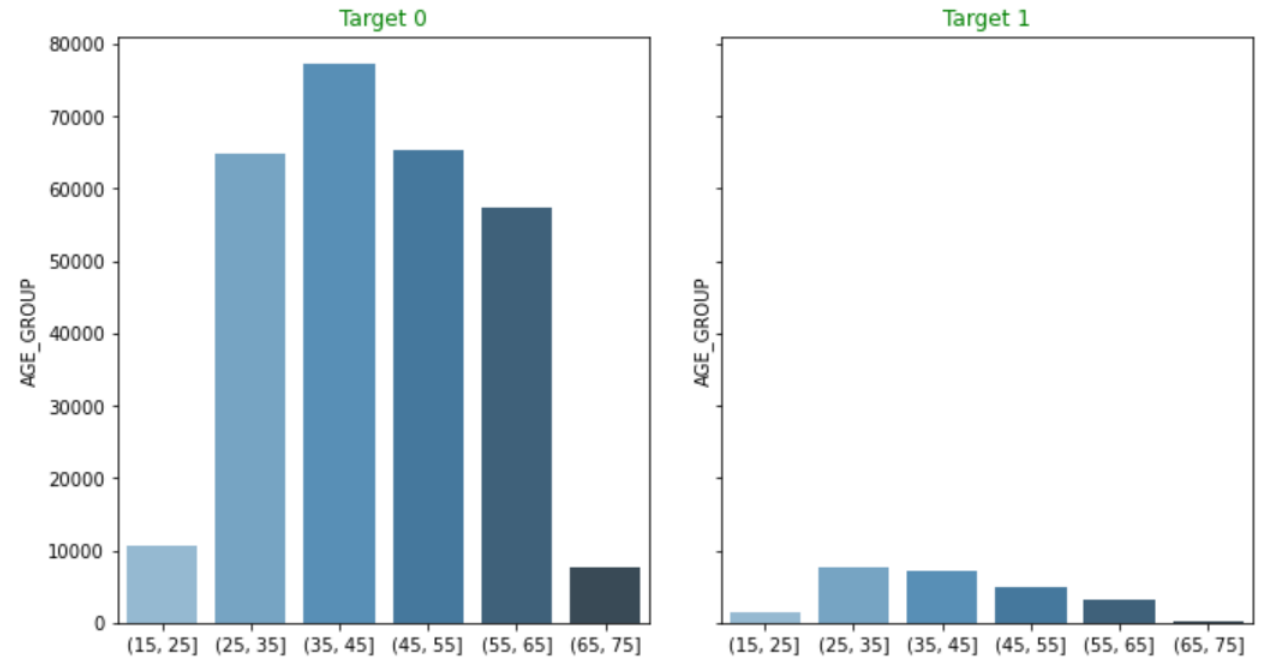


ORGANIZATION_TYPE

- ORGANIZATION_TYPE - From above graph, highest percentage of defaulter can be seen in Applicants who work in Transport type3 followed by Industry type13 and industry type8.

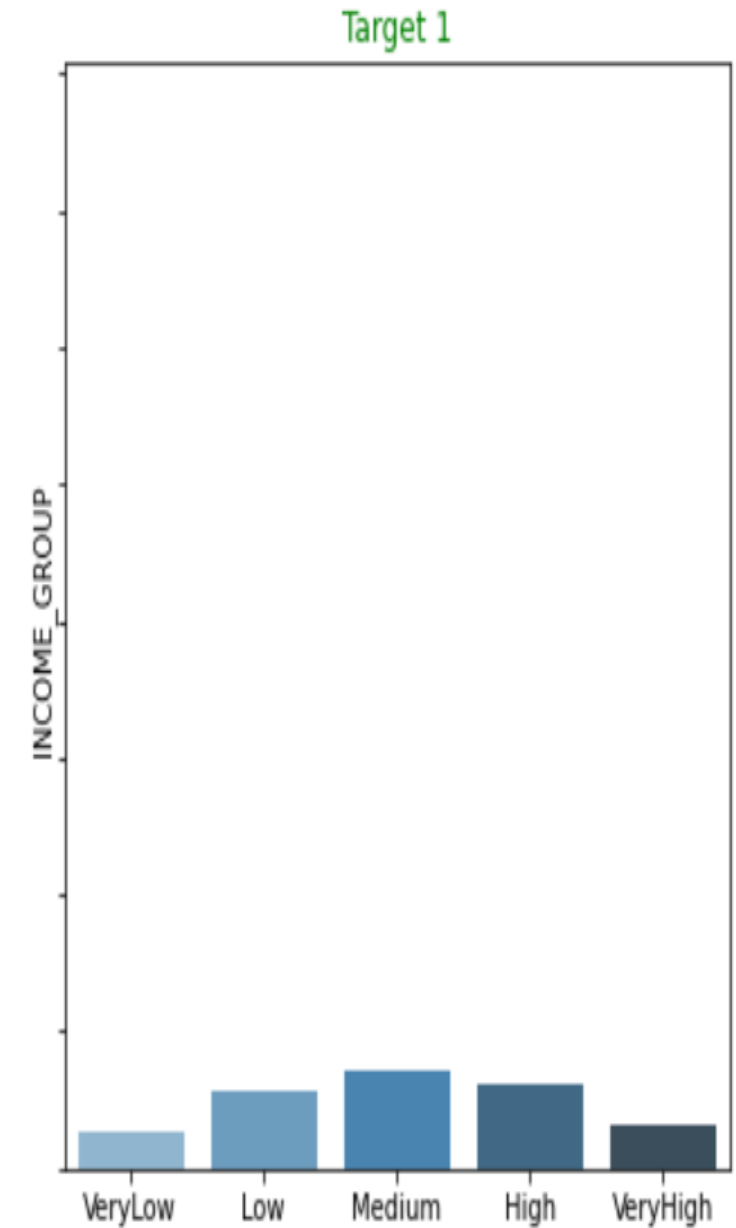
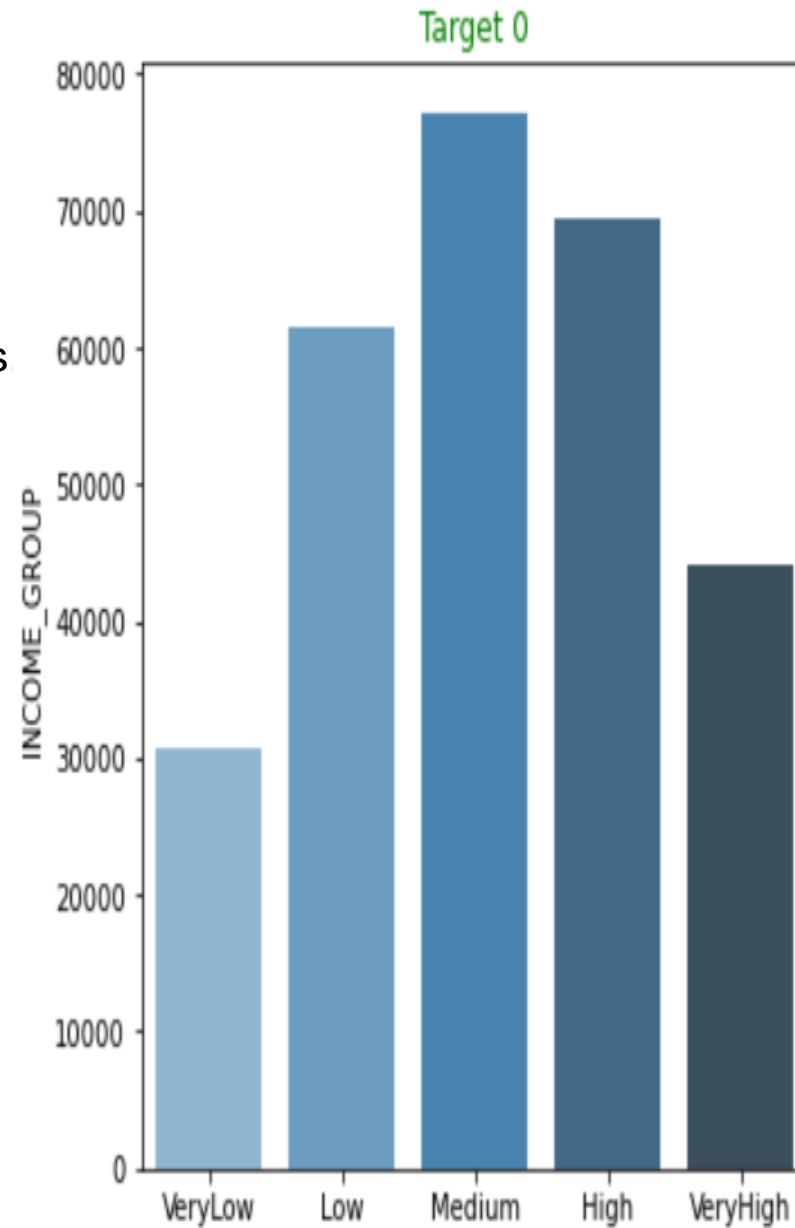
AGE_GROUP

- Since 25-35 age group tend to default more often. Company can be extra cautious before approving their loan. Though as age increases number of defaulters are decreasing, this might be because of increase in salary with as their year of experience increases.



INCOME_GROUP

Medium income group has largest number of defaulters and non-defaulters

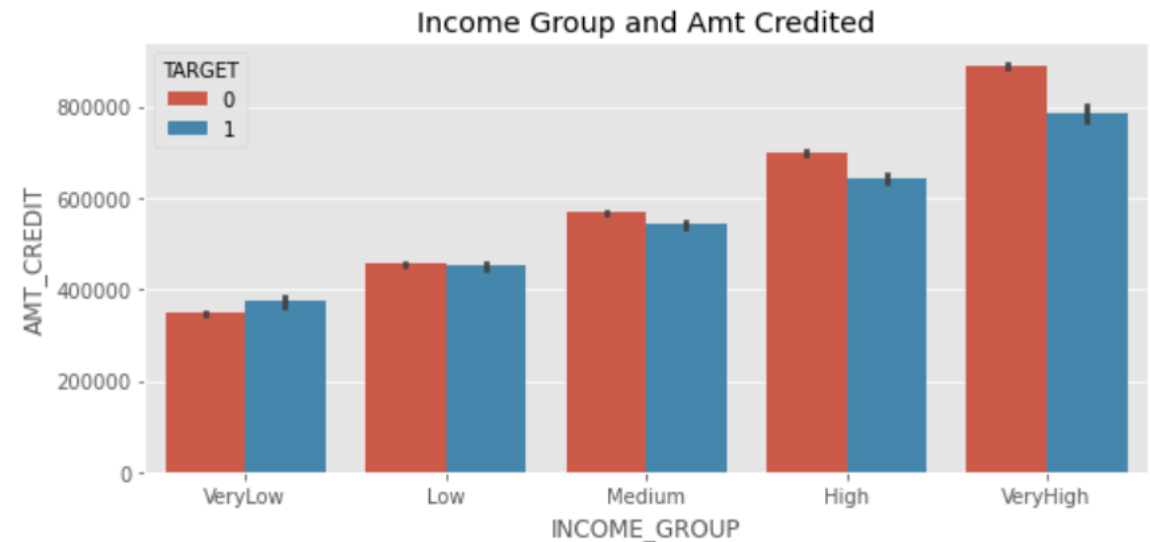


BIVARIATE ANALYSIS FOR TARGET 1

INCOME_GROUP vs AMT_CREDIT

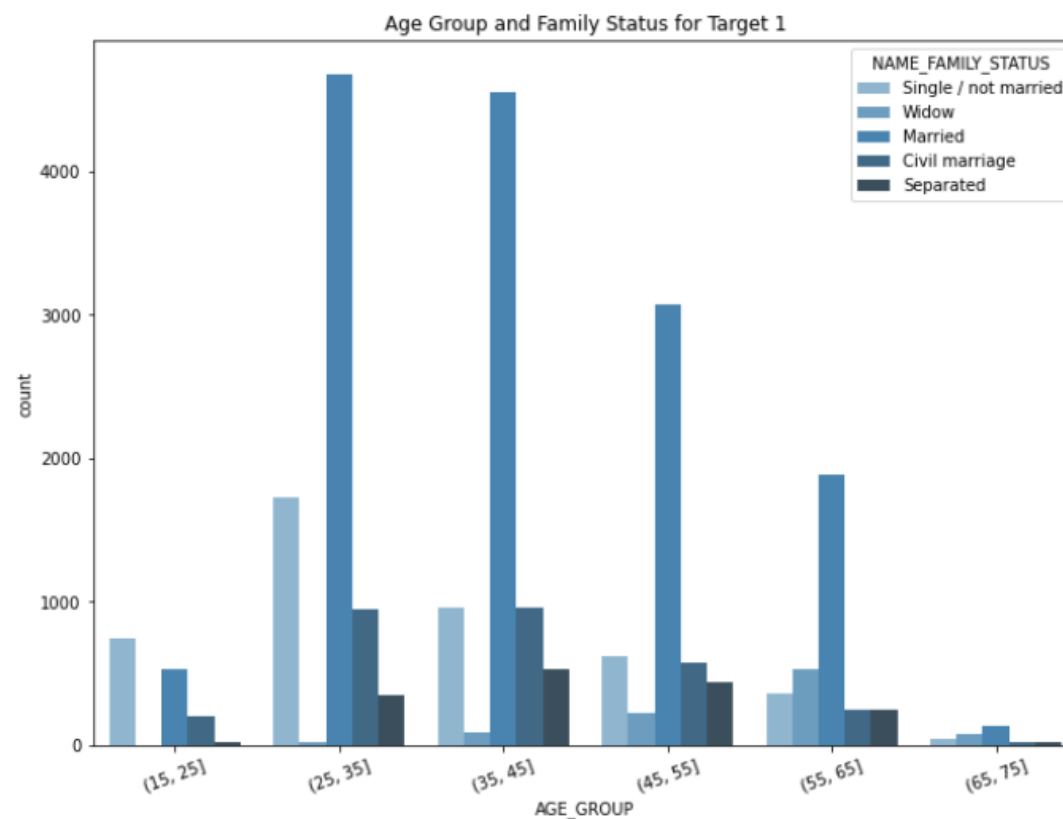
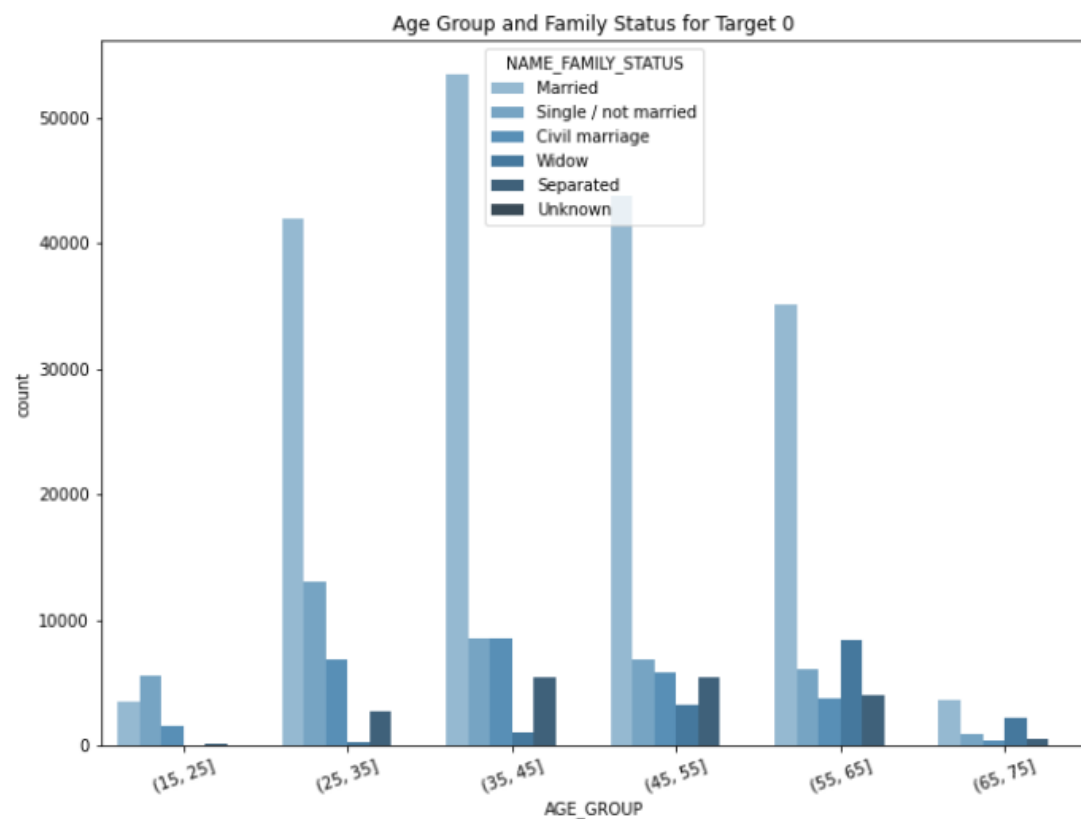
- Though the maximum no of loans is given to Medium income group. Default value per loan is highest in High income group as the AMT_CREDIT is higher too.
- To protect themselves from these huge losses company needs to make policies and define some strict rules before approving higher incomer category loans.

	TARGET	0	1
INCOME_GROUP			
VeryLow		347868.46	373460.40
Low		456525.23	451343.70
Medium		568861.72	542580.16
High		699628.82	643297.93
VeryHigh		890186.33	786021.83

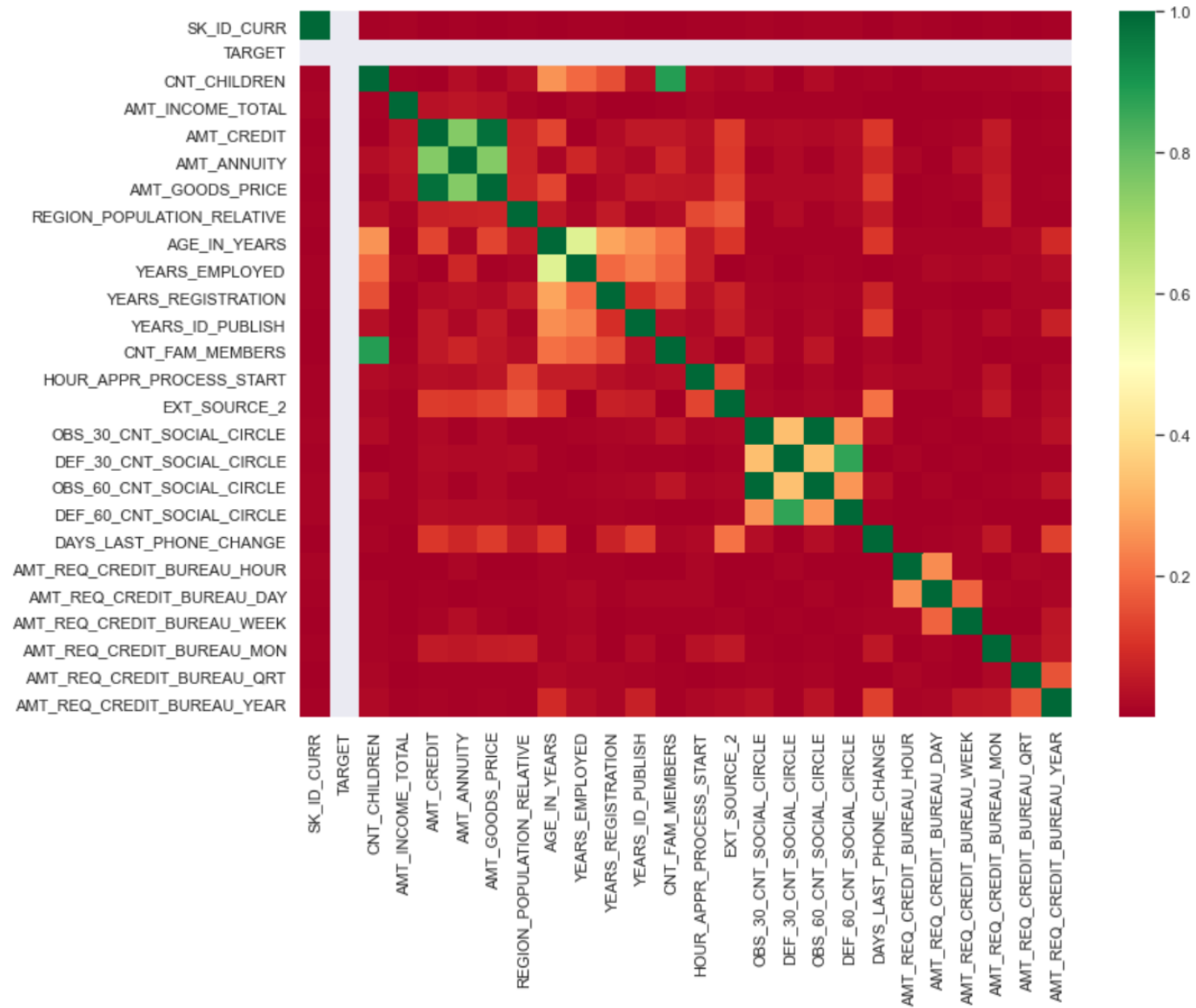


AGE_GROUP VS NAME_FAMILY STATUS

- The largest number of defaulter in Married applicant belong to age group of 25-35 and 35-45.



CORRELATION FOR TARGET 1

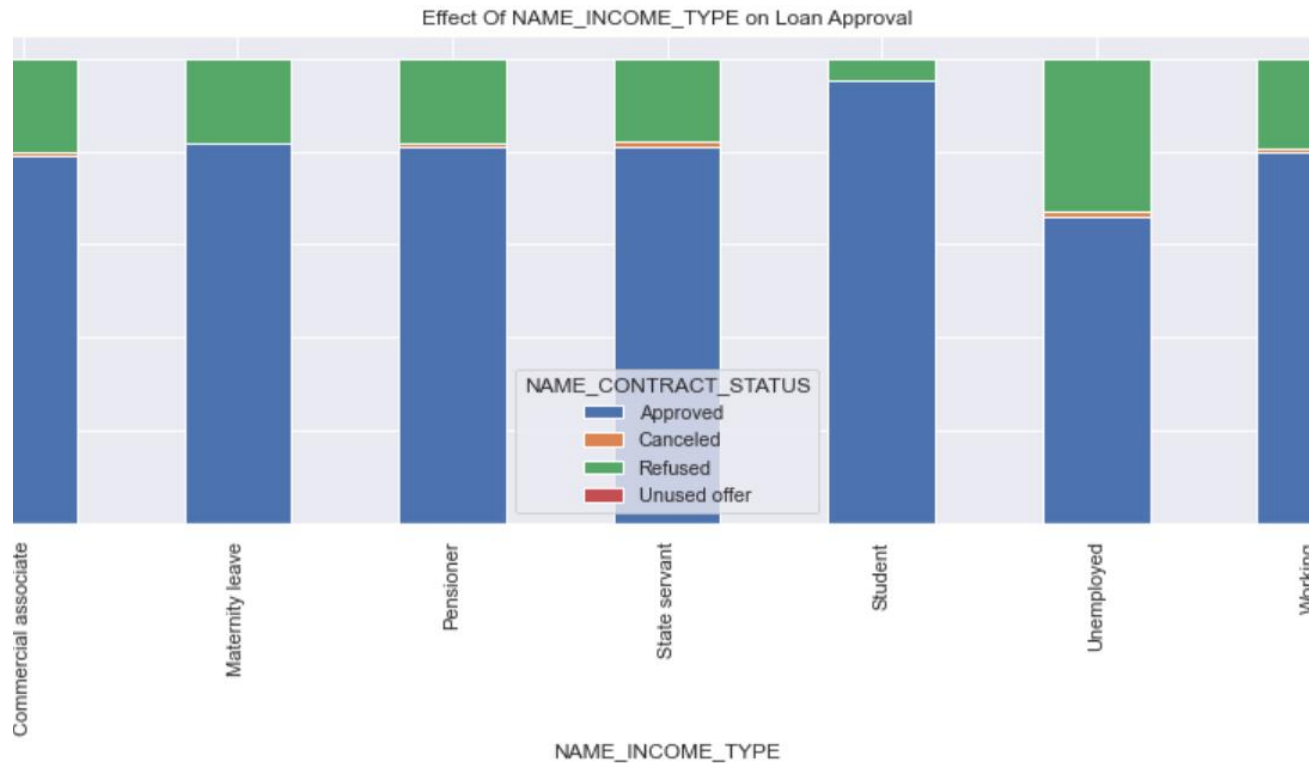


OBSERVATIONS

- 1.OBS_30_CNT_SOCIAL_CIRCLE',OBS_60_CNT_SOCIAL_CIRCLE' - denote the client's social surroundings with observable 30/60 DPD. These are definitely correlated for Target 1, signifying that in approval process this parameter must be strongly looked into.
- 2.DEF_30_CNT_SOCIAL_CIRCLE is highly correlated with DEF_30_CNT_SOCIAL_CIRCLE for Target 1.
- 3.Years employed has an outlier value of 999 and this is skewing the graph
- 4.AMT_CREDIT is highly correlated with AMT_GOOD_PRICE and AMT_ANNUITY for Target 1. Hence bank should look into these variable before approving the loan.

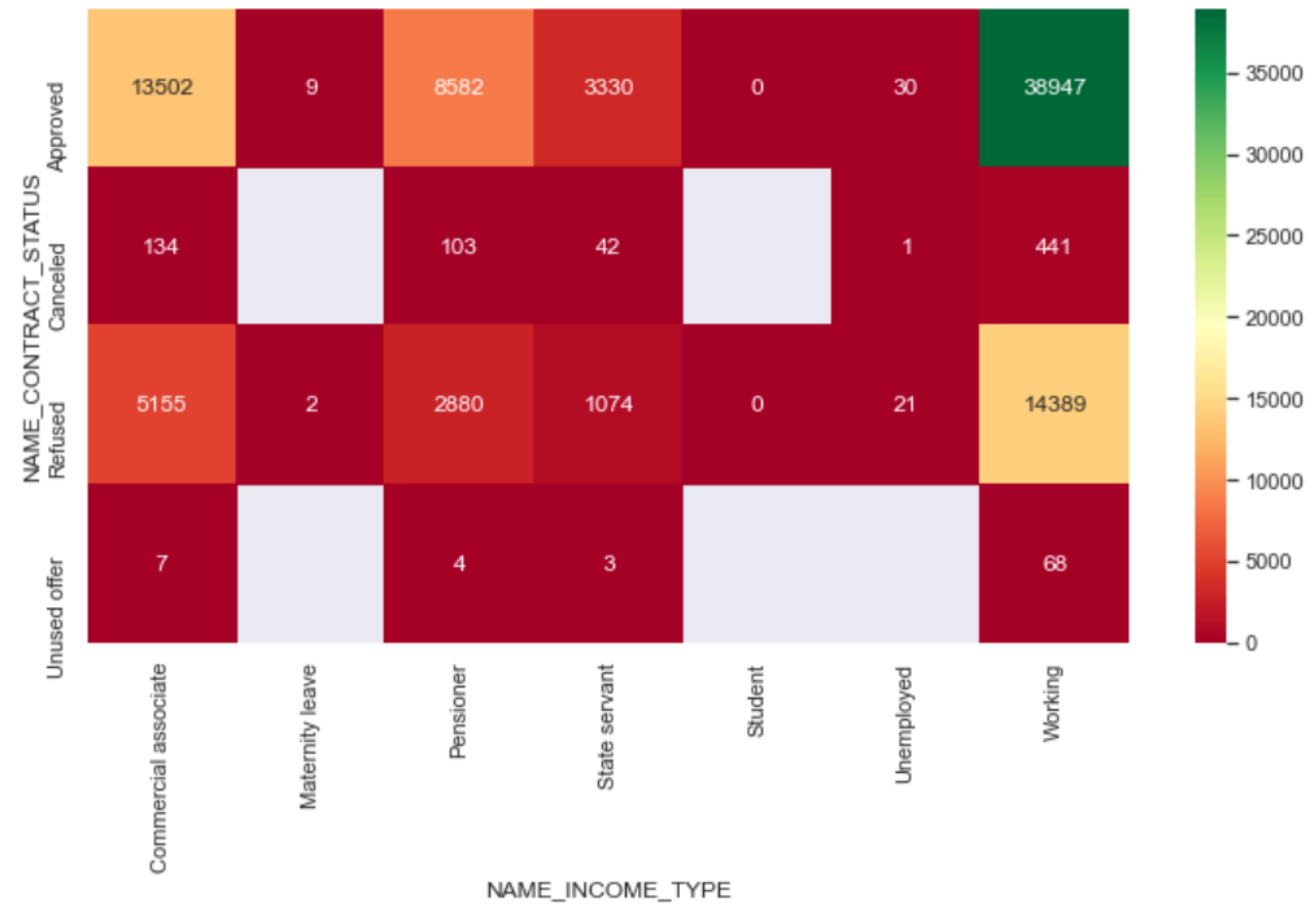
ANALYSIS OF MERGED DATA

EFFECT OF NAME INCOME TYPE ON LOAN APPROVAL



- We can see that the highest percentage of loan approved is in the student category. Which make sense too, since student apply for education loan to study abroad or to do higher studies.

NAME_CONTRACT_STATUS vs NAME_INCOME_TYPE on Target

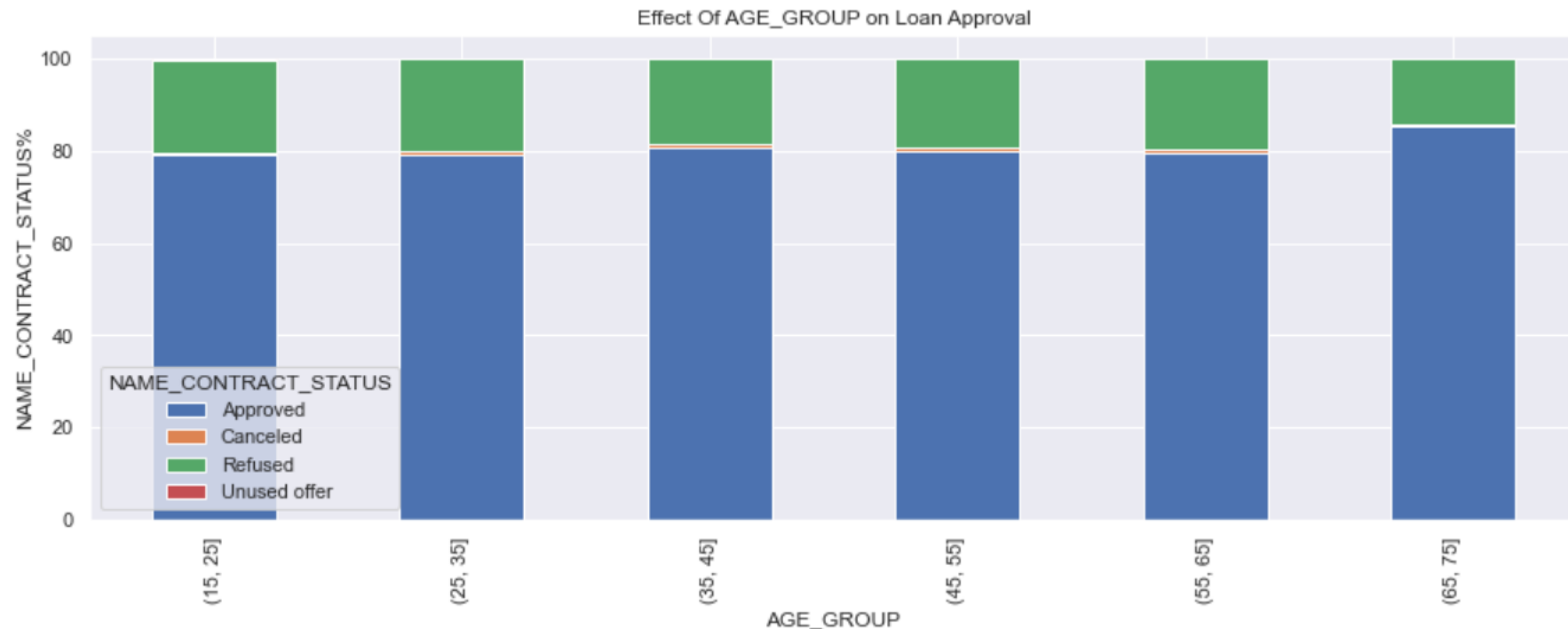


OBSERVATIONS

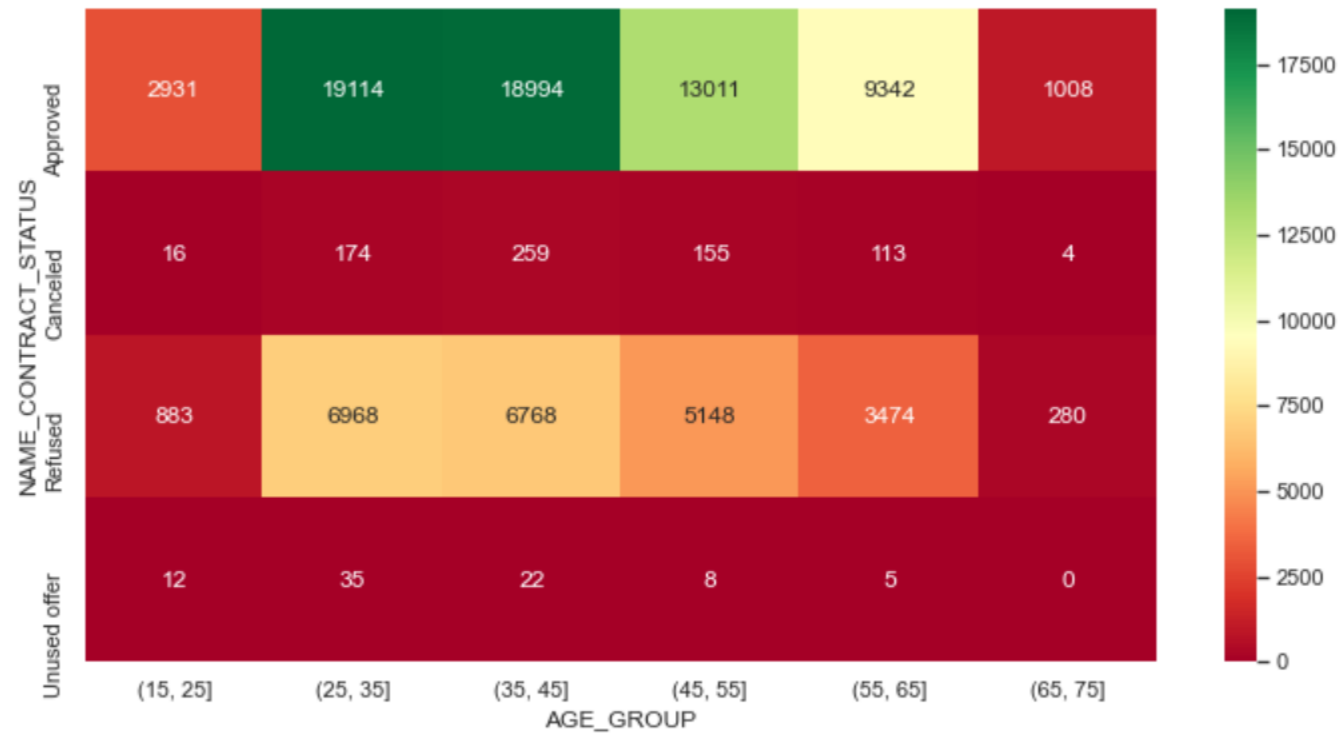
1. We can see that the highest number of defaulter applicant belongs to Working category.
2. Working applicant with Approved status have defaulted in highest numbers
3. The matter of concern should be, 14,389 applicants of working class who were REFUSED earlier and now have defaulted. As mentioned above company should look into this to find out what changed that these applicant got their loan approved at the first place.

EFFECT OF AGE_GROUP ON LOAN APPROVAL

- Interesting thing to not here is approval rate is higher amongst the age group of 65-75.



NAME_CONTRACT_STATUS vs AGE_GROUP ON TARGET



OBSERVATIONS

- 1.As we have seen earlier that loans approved of age group 25-55 have defaulted more often amongst all the age groups.
- 2.Company should look into their policies which has approved people with Refused or cancelled category in previous application, since they are in current defaulters.

ANALYSIS ON APPROVED DEFAULT DATAFRAME

data extracted from merged data (approved and defaulted)

(25, 35]	0.296801	F	0.59222	Working	0.604767
(35, 45]	0.294938	M	0.40778	Commercial associate	0.209658
(45, 55]	0.202034	Name: CODE_GENDER, dtype: float64		Pensioner	0.133261
(55, 65]	0.145062	Medium	0.291770	State servant	0.051708
(15, 25]	0.045512	High	0.261304	Unemployed	0.000466
(65, 75]	0.015652	Low	0.214581	Maternity leave	0.000140
Name: AGE_GROUP, dtype: float64		VeryHigh	0.137997	Name: NAME_INCOME_TYPE, dtype: float64	
Business Entity Type 3	0.246056	VeryLow	0.094348		
Self-employed	0.155016	Name: INCOME_GROUP, dtype: float64		Laborers	0.314332
XNA	0.133727			Sales staff	0.168312
Other	0.050248	0	0.685606	Drivers	0.111941
Business Entity Type 2	0.036351	1	0.314394	Core staff	0.090577
Medicine	0.031988	Name: OWN_CAR_flag, dtype: float64		Managers	0.070796
Trade: type 7	0.030652			Security staff	0.038783
Construction	0.029876	1	0.706925	High skill tech staff	0.038361
Government	0.029457	0	0.293075		
School	0.021460	Name: OWN_REALTY_flag, dtype: float64			
Business Entity Type 1	0.020295				
Transport: type 4	0.019938				
Kindergarten	0.019363				

FINDINGS FORM THE ABOVE SLIDE (HIGH PROBABILITY OF DEFALUT)

- 1.In 'INCOME_GROUP', Medium income applicant
- 2.In 'AGE_GROUP' applicant between age group of 25-45
- 3.In 'NAME_INCOME_TYPE' applicant from Working category
- 4.In 'OCCUPATION_TYPE' applicant from Labourers category
- 5.In 'ORGANIZATION_TYPE' applicant working in Business type3 organization
6. In 'OWN_CAR_flag', 31% don't have car defaulted
- 7.In 'OWN_REALTY_flag', 29% don't own home defaulted

SUMMARY

After doing our analysis on all the data frame , we can confirm that company should be more careful before accepting the loan application of :

- Medium income applicant
- Applicant between the age group of 25-45
- Male applicant since the rate of default is high amongst them
- applicants who works as Labourers, Salesman or Drivers
- applicants who works in Business type 3 organization
- applicants who doesn't Own House because they already have a liability