

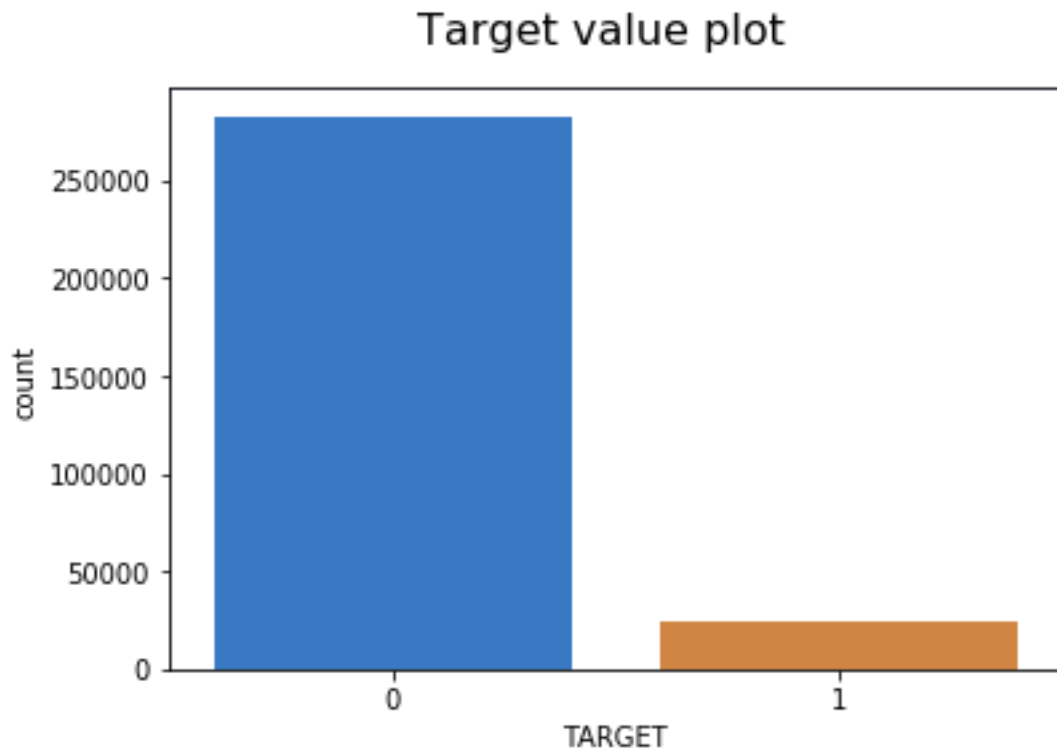
EDA Case Study

By:
Venkata Nagarjuna T H

Target Variable:

First we have divided data frame after cleaning , imputing values when ever is required, binning variables as mentioned in python notebook, into defaulter and non defaulter category based on the target variable.

The plot between the defaulter and non defaulter as shown below:



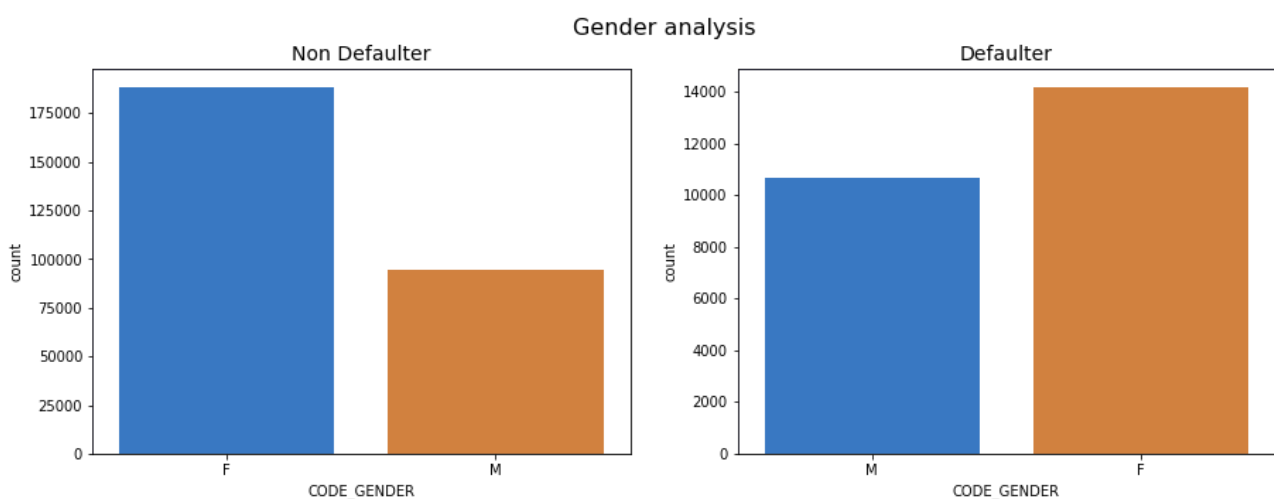
Here 0 represents non defaulter and 1 represents defaulter category and the ratio of defaulter to non defaulter is 8.45 and 91.5 respectively.

Data Analysis:

We have performed Data analysis based on the above data frames divided on target variables.

Univariate Analysis of Categorical Variables:

Gender

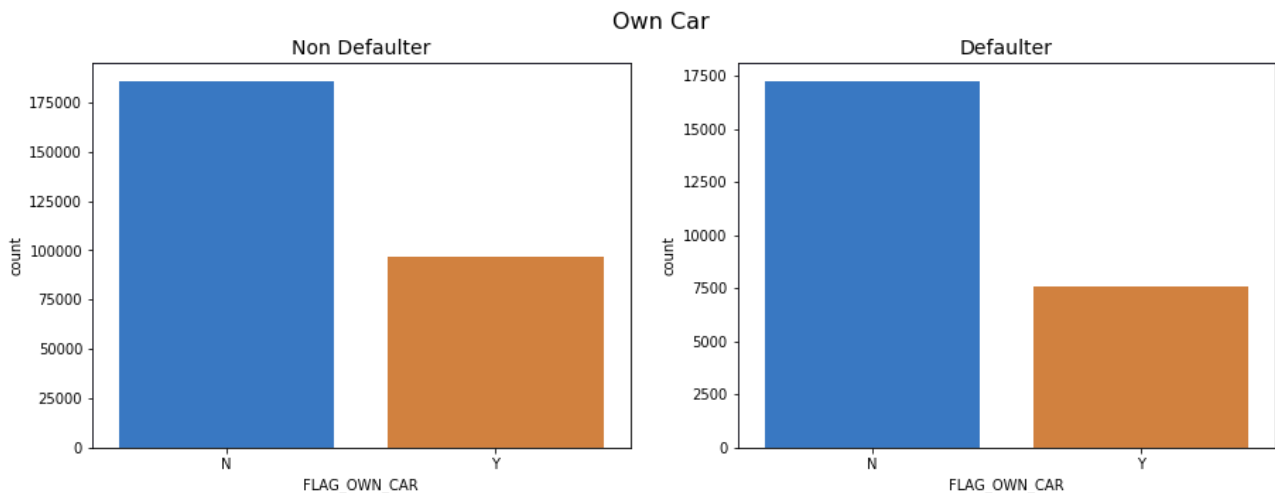


As shown in the above we have plotted the gender count between defaulter and non defaulter categories.

We found that female clients almost the double to the male clients in non defaulter category where as males clients are just less than to female clients in defaulter category.

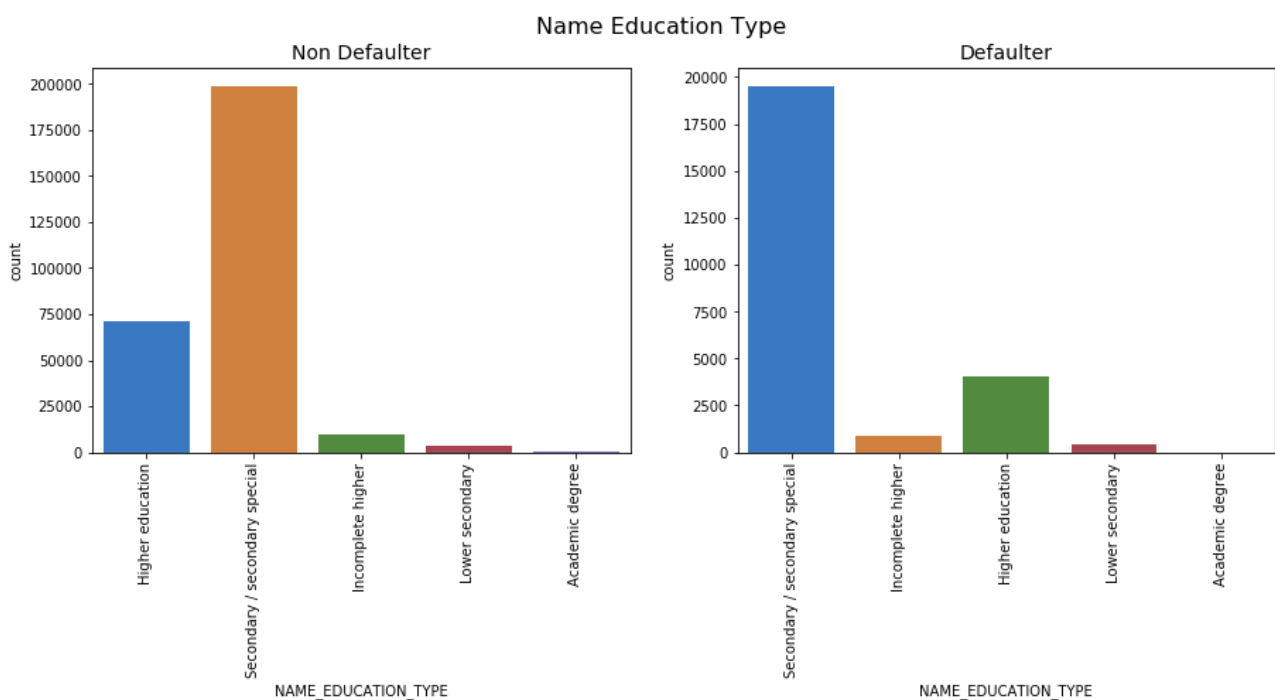
In the above, female category is more defaulter when compared with male category.

Flag own Car



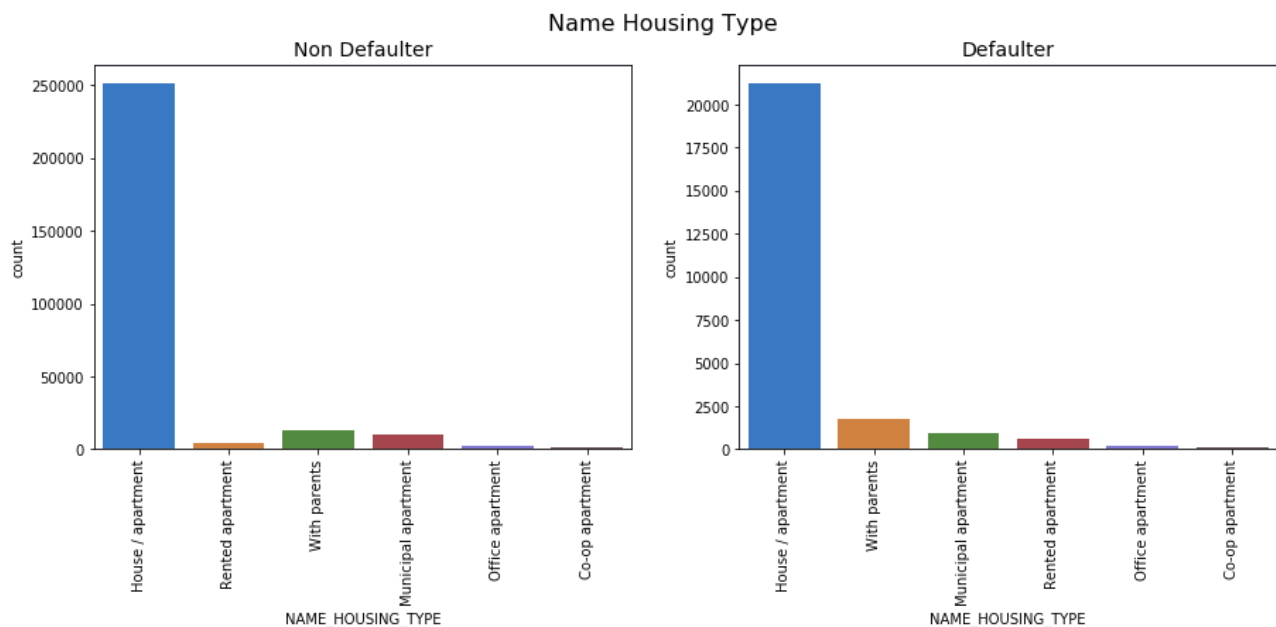
The clients that owns a car are almost a half of the ones that doesn't own one. The clients that owns a car are less likely to not repay a car that the ones that own.

Education Type:



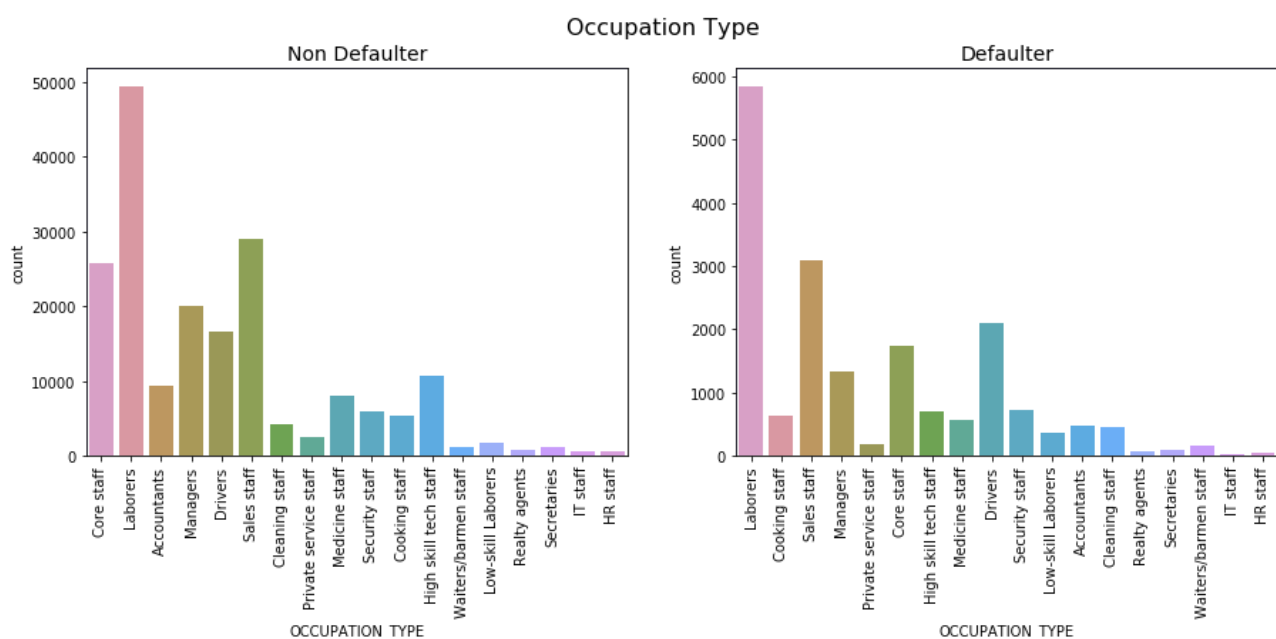
A large number of applications are filed by people having secondary education followed by people with Higher Education applications. we see that the applicants with Lower Secondary education status has the highest percentage of payment related problems.

Housing Type:



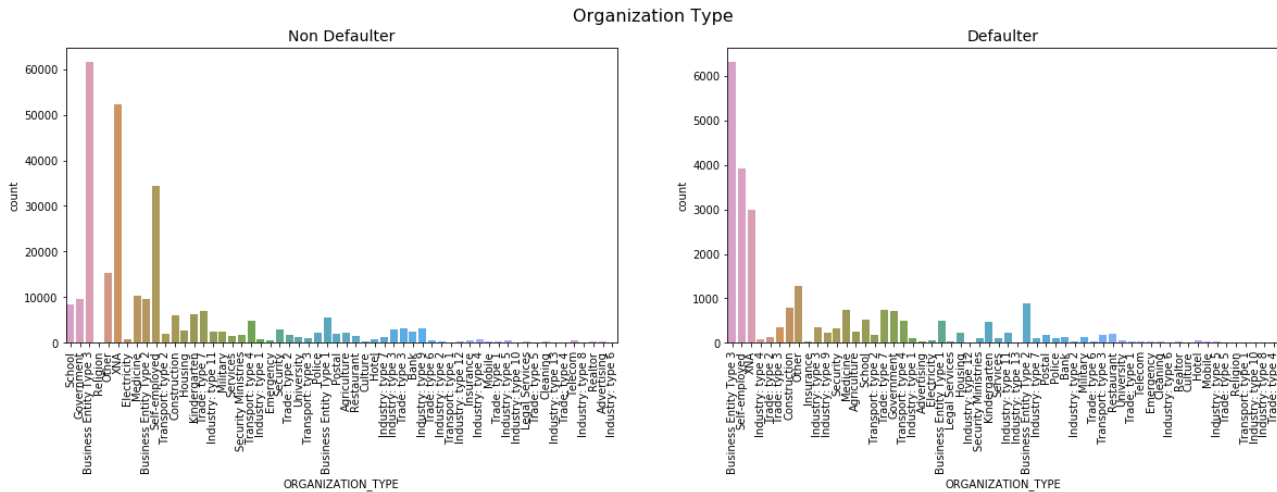
Applicants living in House / apartments has the highest number of loan applications equal While we see that Rented apartment and applicants living with parents has highest percentage having more payment problems

Occupation Type



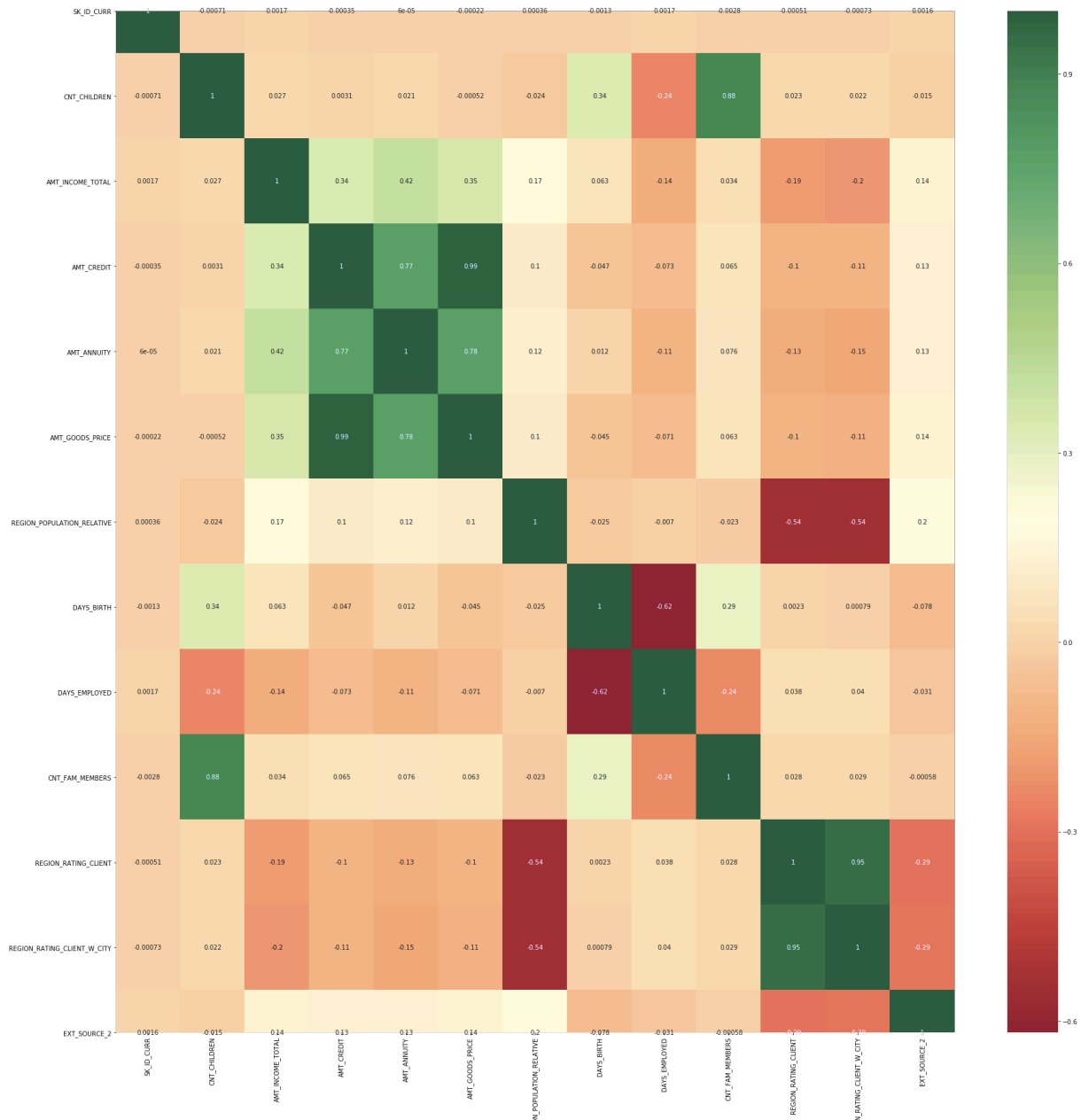
From the above, We can see that laborer's , Sales ,core staff has highest number of loan applications but wen it comes to Non loan repayment highest number can be seen in Low-skill laborer's,Drivers,Waiters,cookin staff and so on...

Organisation Type



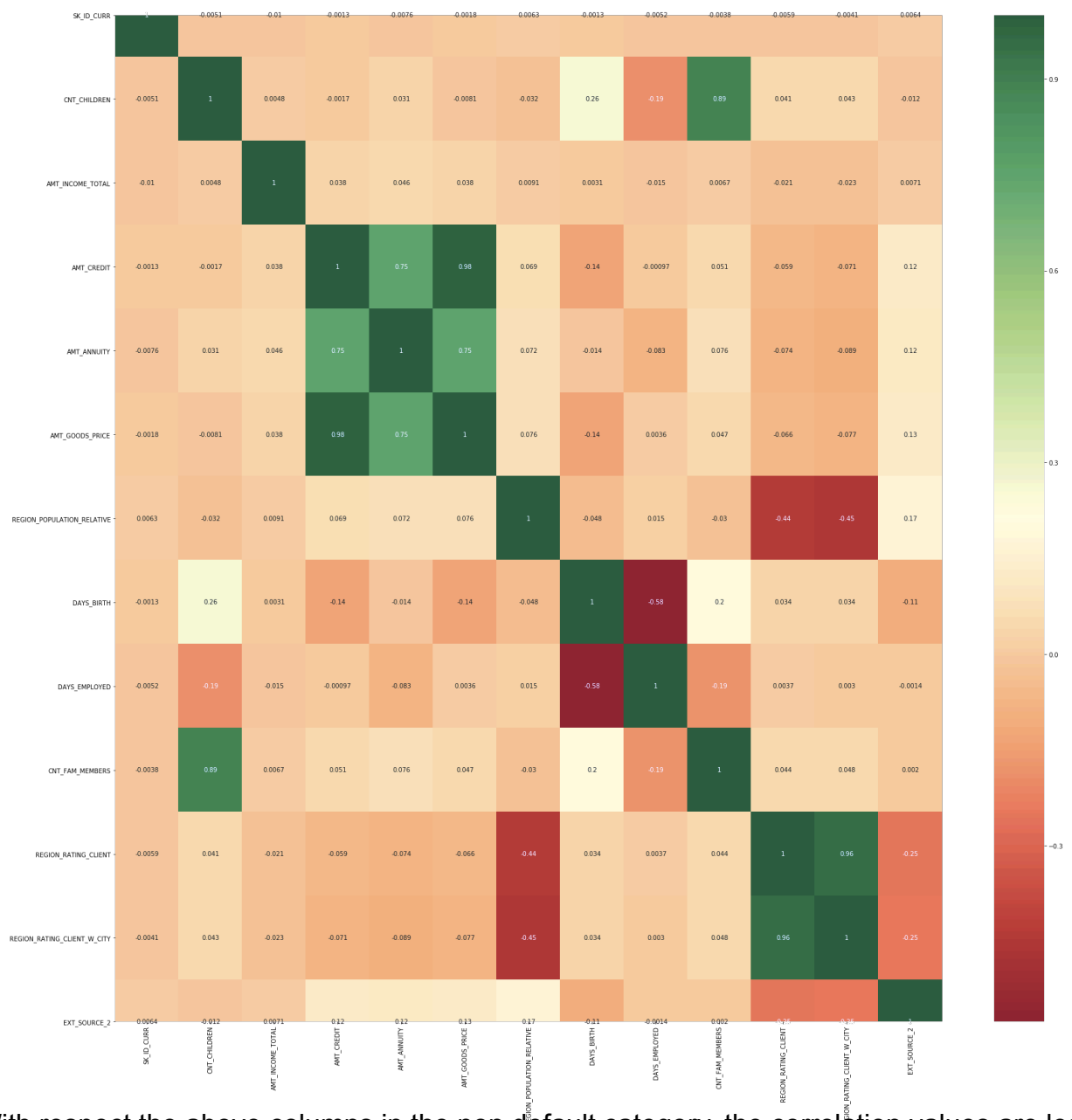
From the above, we can observe that Applicants works mostly in an organization of Business Entity Type 3, second number is of the Self-employed. But when it comes to Non loan repayment highest number can be seen in Applicants who work in Transport Type3

Heat Map of Non Defaulter Category:



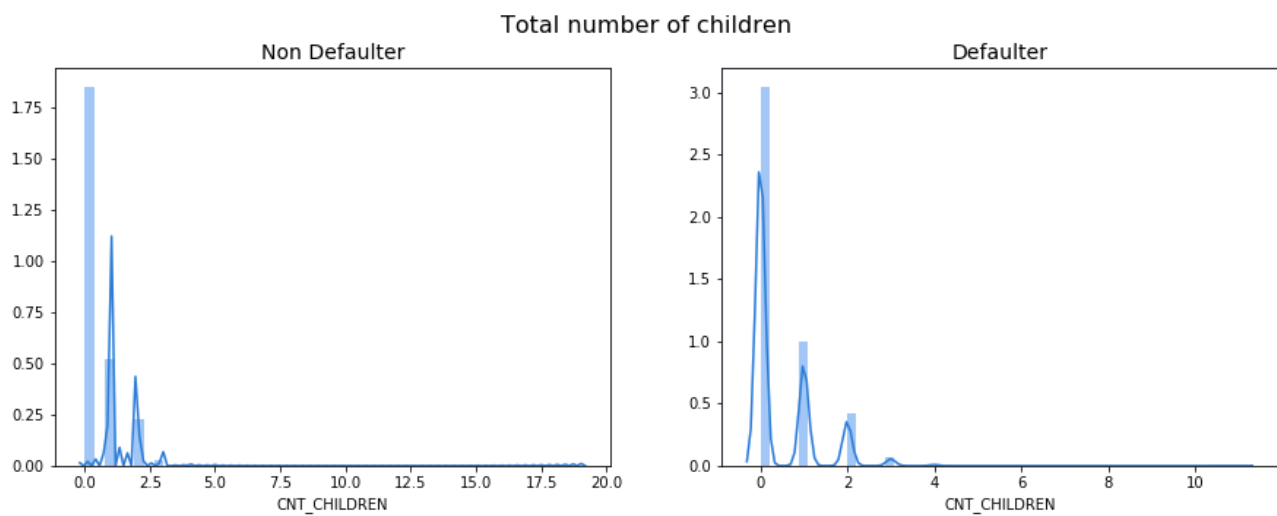
From the above, we can observe that the few columns like AMT_credit, Amt_Annuity, Amt_goods_price are related so closely in the green shaded area.

Heat Map of Defaulter Category:



With respect the above columns in the non defaulter category, the correlation values are less when compared with the defaulter category of the variables mentioned in defaulter category.

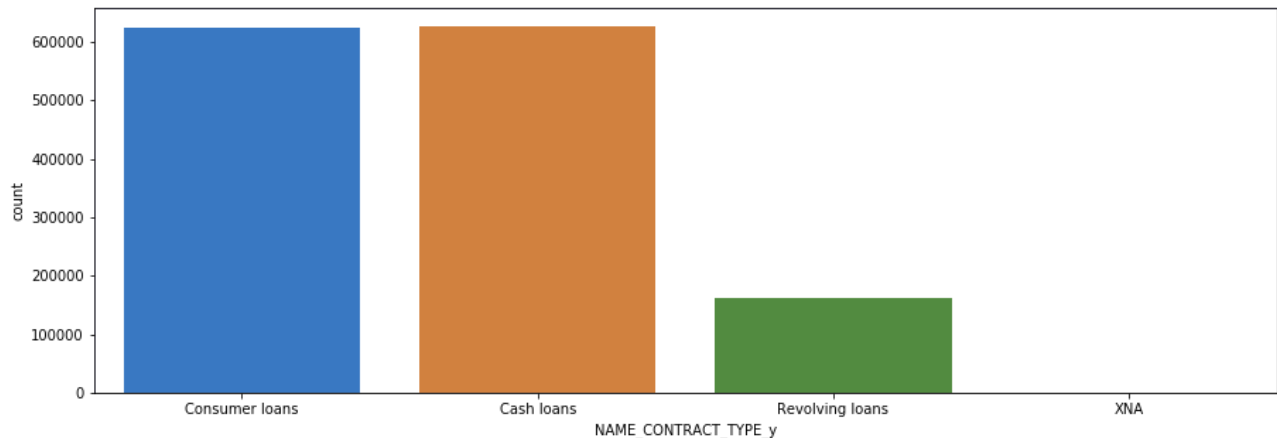
Univariate Analysis of continuous variables:



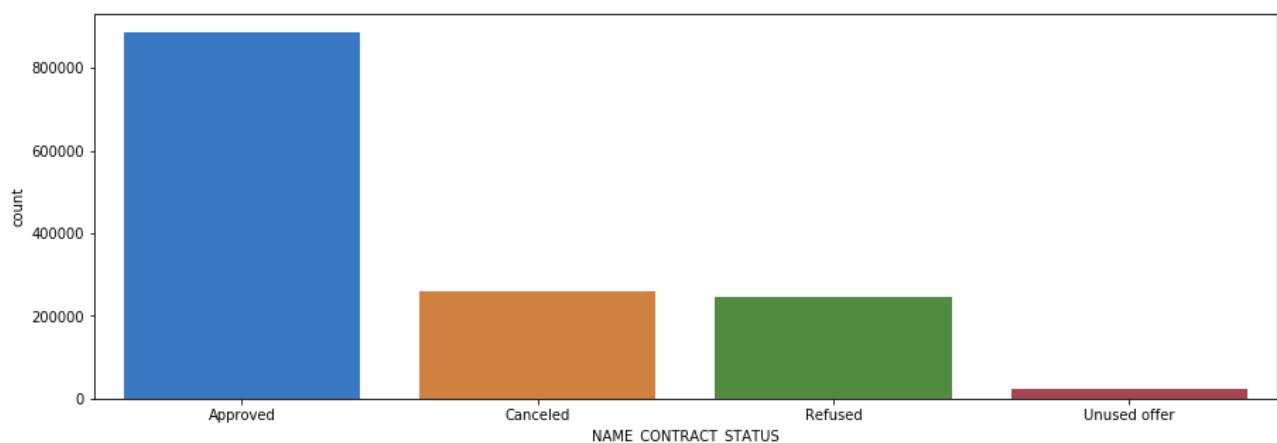
From the above picture, When a person is having less number of kids he is less likely to be defaulter, when the Number of children is more, then the person is more likely to be defaulter.

Merging the Previous Data

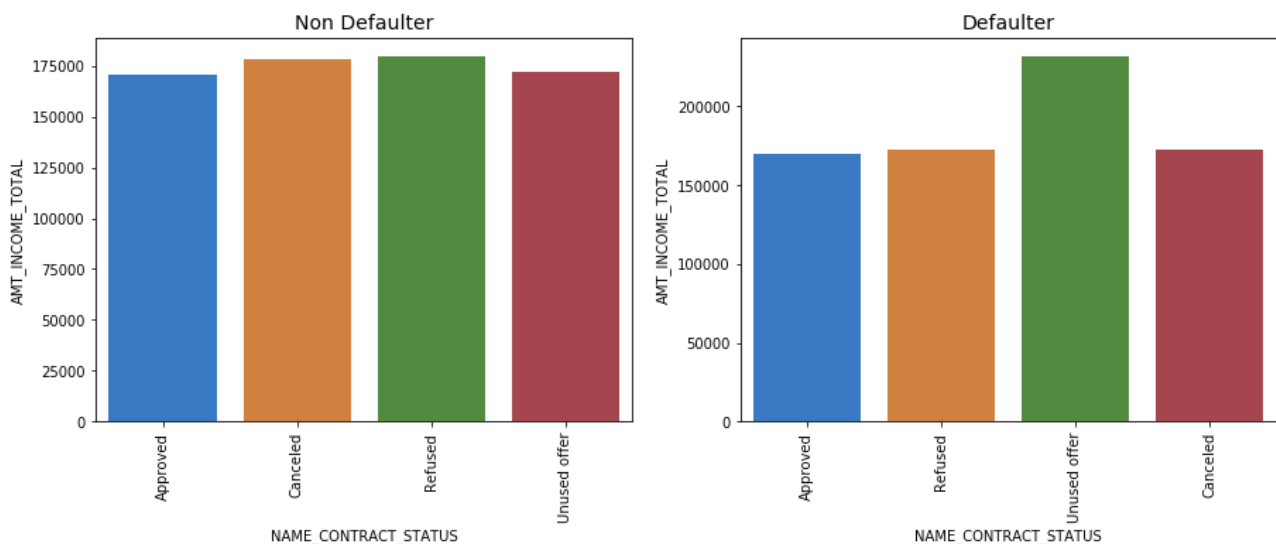
With the merging the previous Data, The major portion of the loans includes, Consumer and Cash loans where as very less number of revolving loans.



With the below plot, we can see most of the loans are approved, and canceled and refused loans occupies the same percentage of amount.



With the below plot we plotted between the Status of the loan and total income, There are many loans where non defaulter is taking place , when the total income is more then less likely to become defaulter of the loan.



Recommendations:

From the above plotted graphs between various categorical and numerical variables, I have drawn the following recommendations to the agency before considering to approve / reject the loan submitted by person.

1. When Person is having less Number of kids, then he is less likely to be a defaulter.
2. When compared with the Gender, Females are having most defaulters percentage when compared with the male category.
3. When Person owns a car he is less likely to be a defaulter
4. When person Education type is under graduate / secondary he is more likely to be a defaulter.
5. When person is living in rented house / apartment, with parents are more likely to be defaulter category.
6. When person occupation type is Business type 3, core staff, labour they are more likely to be a defaulter.
7. Above of all, If the correlation between total annual income and amount annuity is good and more he is less likely to be a defaulter.
8. If the correlation between the four categories are good, he is less likely to be defaulter.