Credit EDA Case Study

– Ashok G & Devendar G

Credit EDA Case Study:

Problem Statement - I

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

1. *If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company*
2. *If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.*

Business Objectives

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

Case Study Understanding:

The expectation from the case study is to find the driving factors or strong indicators of the loan defaults.

1. Understanding the data provided and developing risk strategy by analyzing different variables in the data.
2. Understanding the influences of different parameters on the loan default.
3. Visually depicting multiple variables influence on loan defaults.

Case Study Analysis:

1. From the data sets provided, we have taken the ‘application_data.csv’ data set for our analysis as it contains current loan applicants data. Moreover, we’ve carefully examined and identified some columns which might represent pattern which could be extremely useful in depicting loan defaults if analyzed visually.

2. Secondly, we have analyzed and calculated the percentage of missing columns and did the missing columns,outliers treatment for the provided variables.

3. Next, we have divided the data into two segments on ‘TARGET’ variable as it has 0 and 1’s.
   1 – For loan applicants who have payment difficulties.
   0 – For other loan applicants.

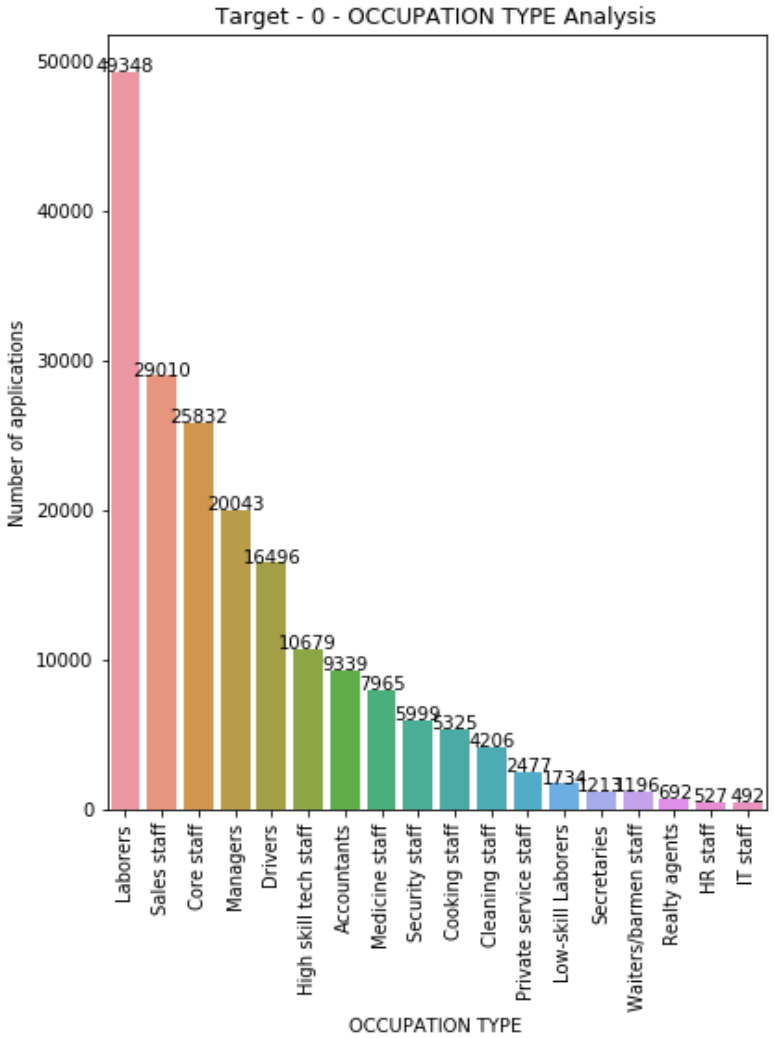
4. After analyzing ,cleaning and meticulously refining the data set we have more accurate and relevant data for our further analysis.

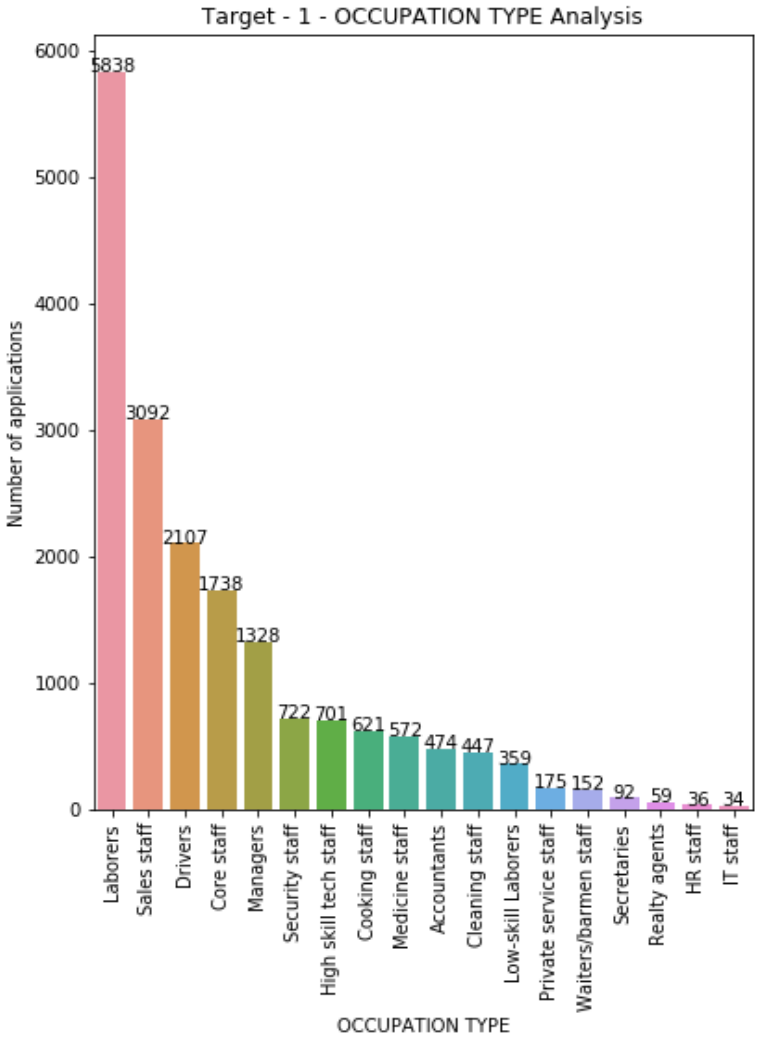
5. We planned to conduct ‘Univariate’, ‘Bi-Variate’ analysis to see the patters of loan defaults which you can see next.

2. Univariate analysis for Occupation type:

The below bar plot represents trend of loan applications across ‘Occupation type. From an overall perspective it is evident that top 5 loan applicants are ‘Laborers’, ’ sales staff’, ’Core Staff’, ’Managers’ and ‘Drivers’. In contrast, ‘IT Staff’, ‘HR Staff’ and ‘Reality agents’ are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as Laborers’, ’ sales staff’ and ‘Drivers’ close to ~10000 where as all other fields are ~8000.

Observation: that Laborers’, ’ sales staff’, ’Core Staff’, who are not highly skilled labors could be potential loan defaulters when compared with other occupation types w.r.t total applicants.

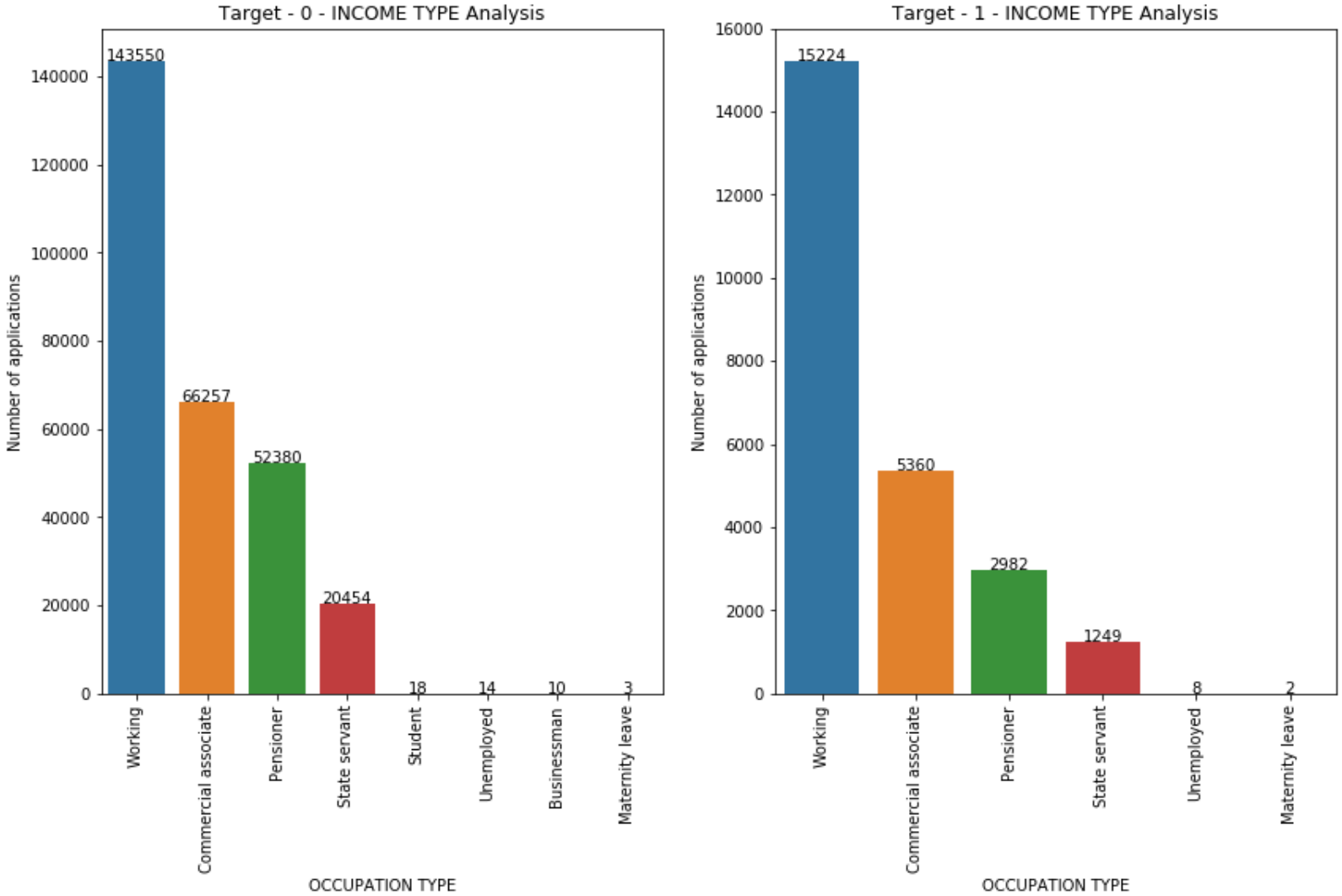




3. Univariate analysis for Income type:

The below bar plot represents trend of loan applications across 'Income Type'. From an overall perspective it is evident that top 3 loan applicants are 'Working', 'Commercial staff', 'Pensioner', and 'State servant'. In contrast, 'Maternity, 'Business man' are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as 'Working', 'Commercial staff', 'Pensioner', close to ~23000 where as all other fields are ~1200.

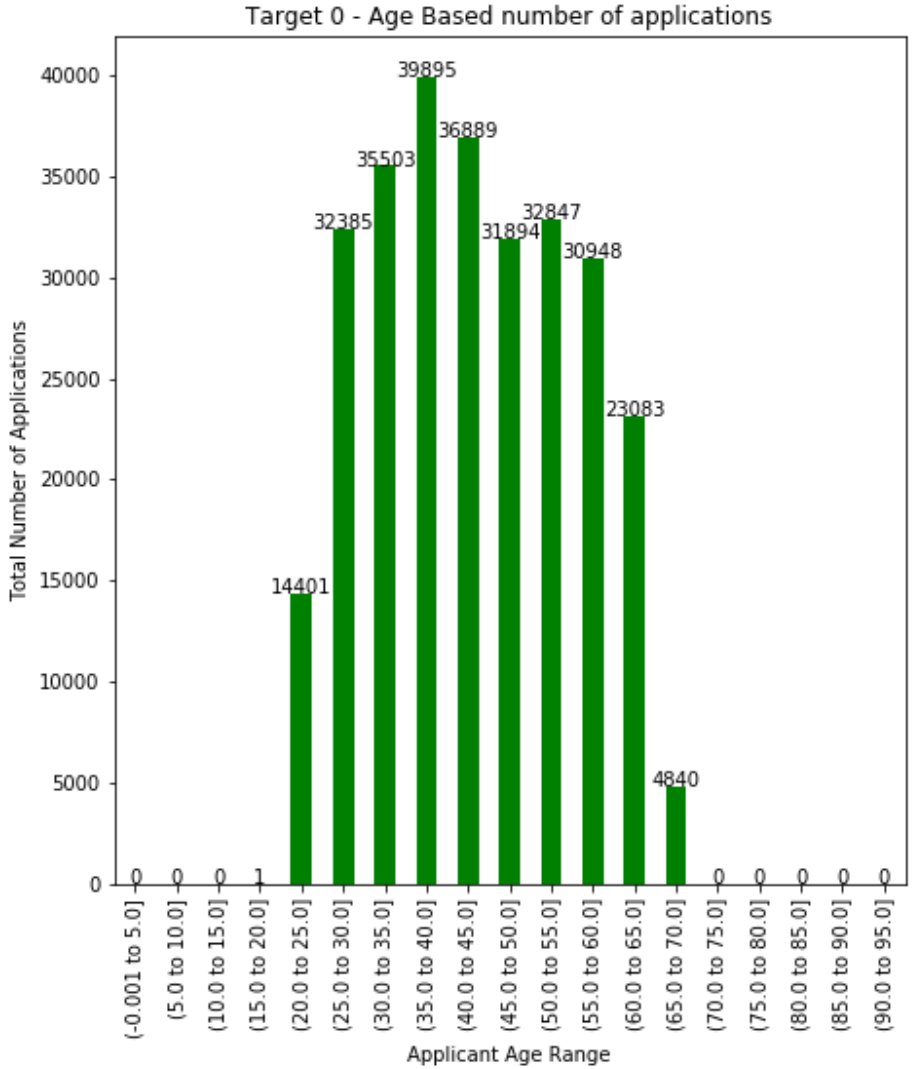
Observation: Interesting insight with the below plot we can say that Working', 'Commercial staff', 'Pensioner', who receive low income are likely to be potential loan defaulters when compared with other income types w.r.t total applicants.

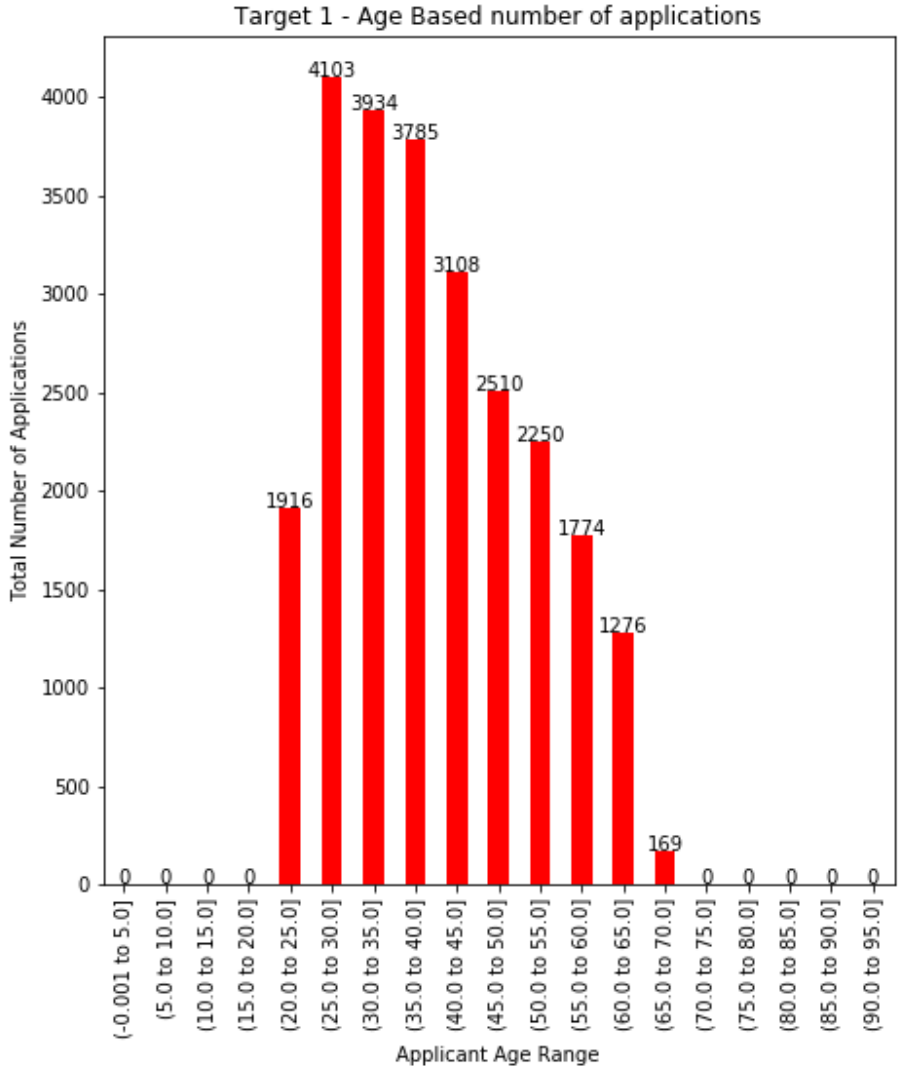


4. Univariate analysis for Age type:

The below bar plot represents trend of loan applications across 'Age Groups'. From an overall perspective it is evident that top 4 loan applicants are in the age group '25-30', '30-35', '35-40', and '40-45' ~14000. In contrast, '20-25, '65-70' are the least loan applicants ~18000. Similarly, the loan applicants **with payment difficulties** are clearly depicted as '20-25', 25-30', '35-40', close to ~11000 where as all other fields are ~8000.

Observation: that, though age group 20-25' are less number of applicants, they have more difficulties to pay the loan compared with age groups 25-30', '35-40' and are prone to risk of potential loan defaulters when compared with other age groups w.r.t total applicants.

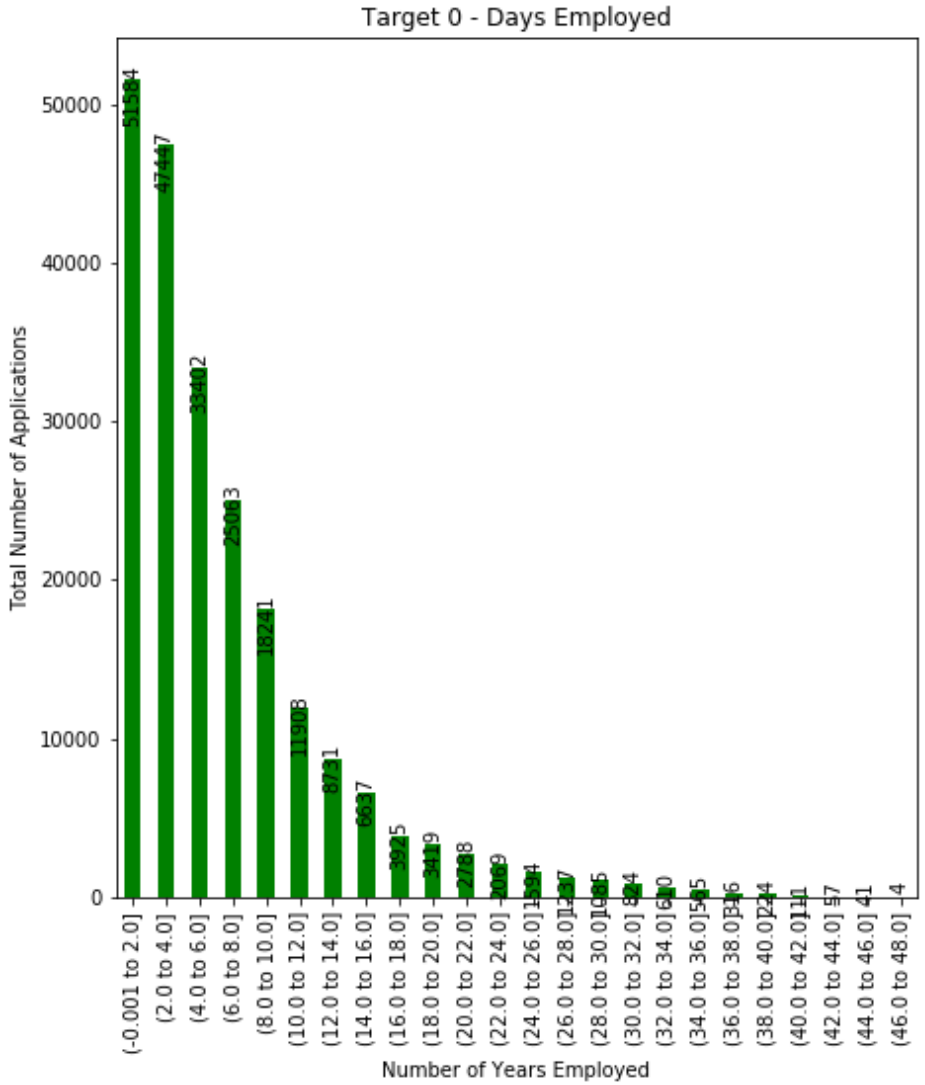


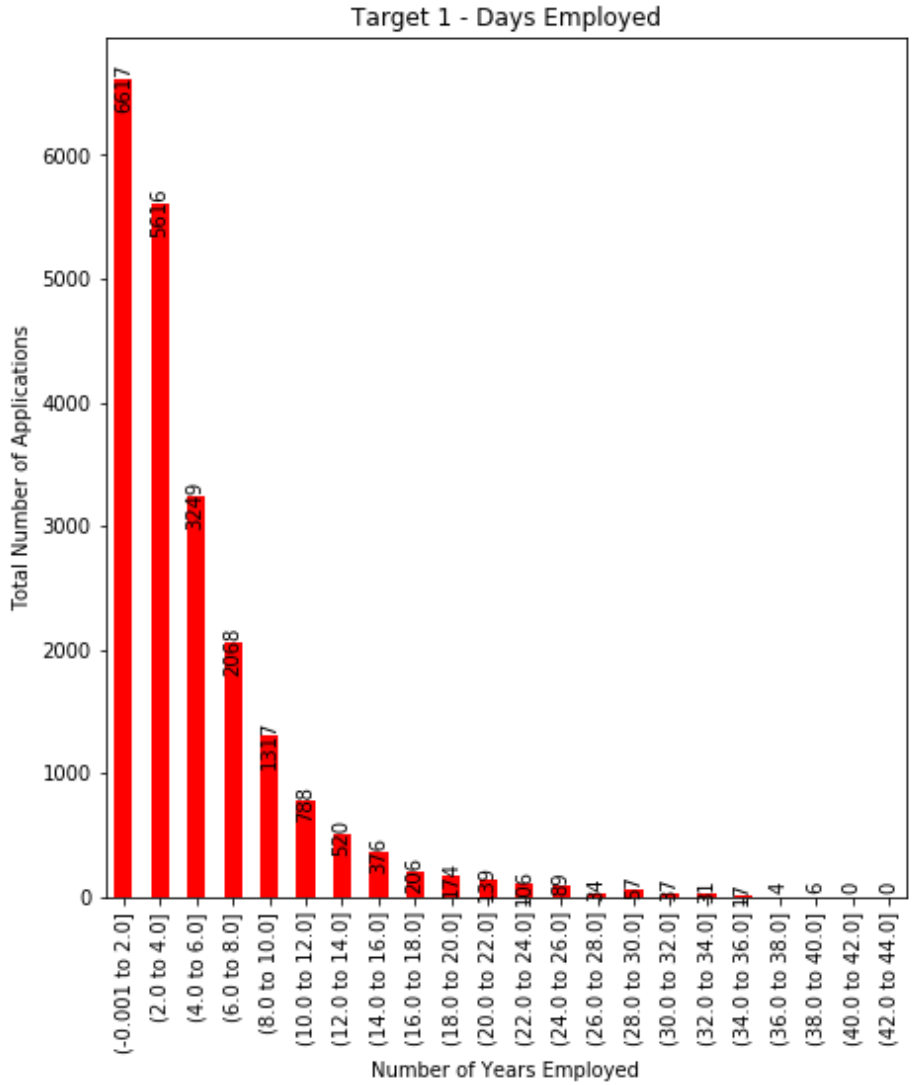


1. Univariate analysis for Days Employed type:

The below bar plot represents trend of loan applications on 'Days Employed'. From an overall perspective it is evident that top 5 loan applicants are employed '0-2 yrs.', '2-4 yrs.', 6-8 yrs.' are ~13000. In contrast, '44-46 yrs.', '46-48 yrs.' are the least loan applicants are ~50. Similarly, the loan applicants **with payment difficulties** are clearly depicted as 0-2 yrs. and 2-4 yrs.

Observation: that loan applicants who **are employed less number** of days are likely to be potential loan defaulters when compared with total number of applicants with more experience w.r.t total applicants.

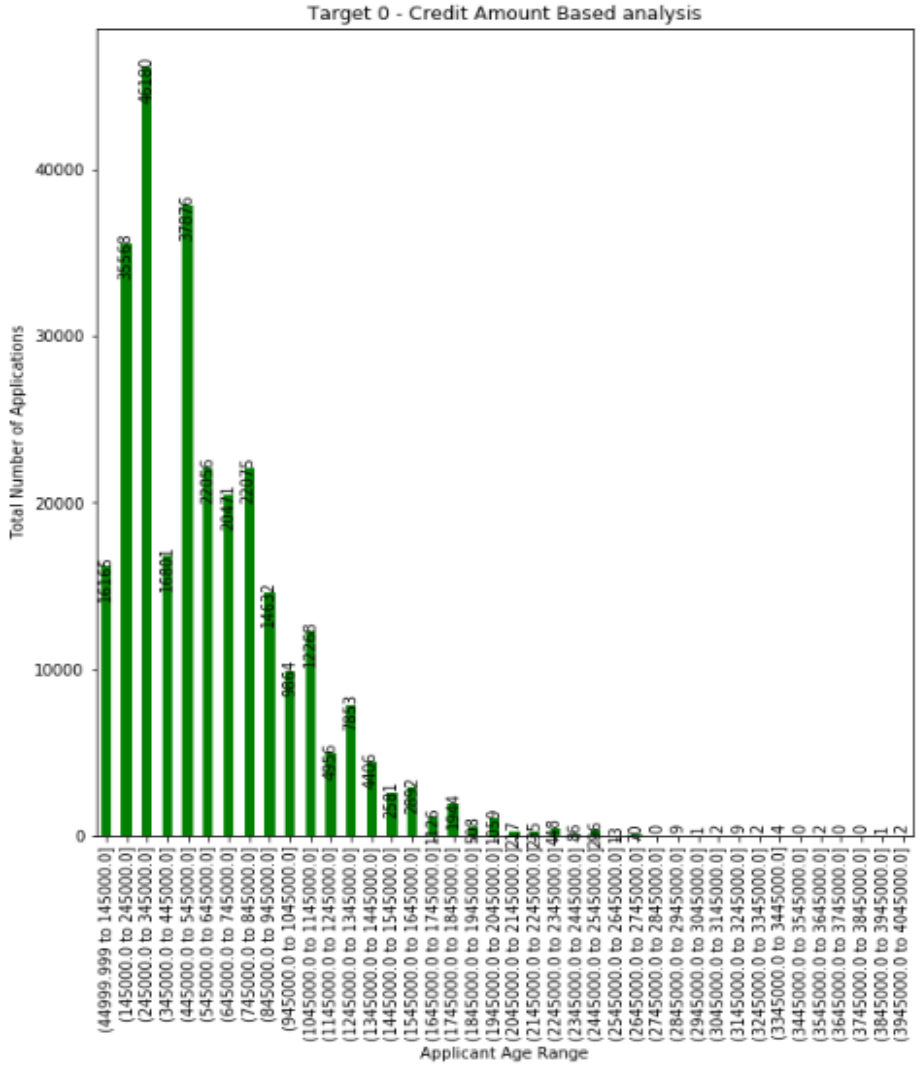


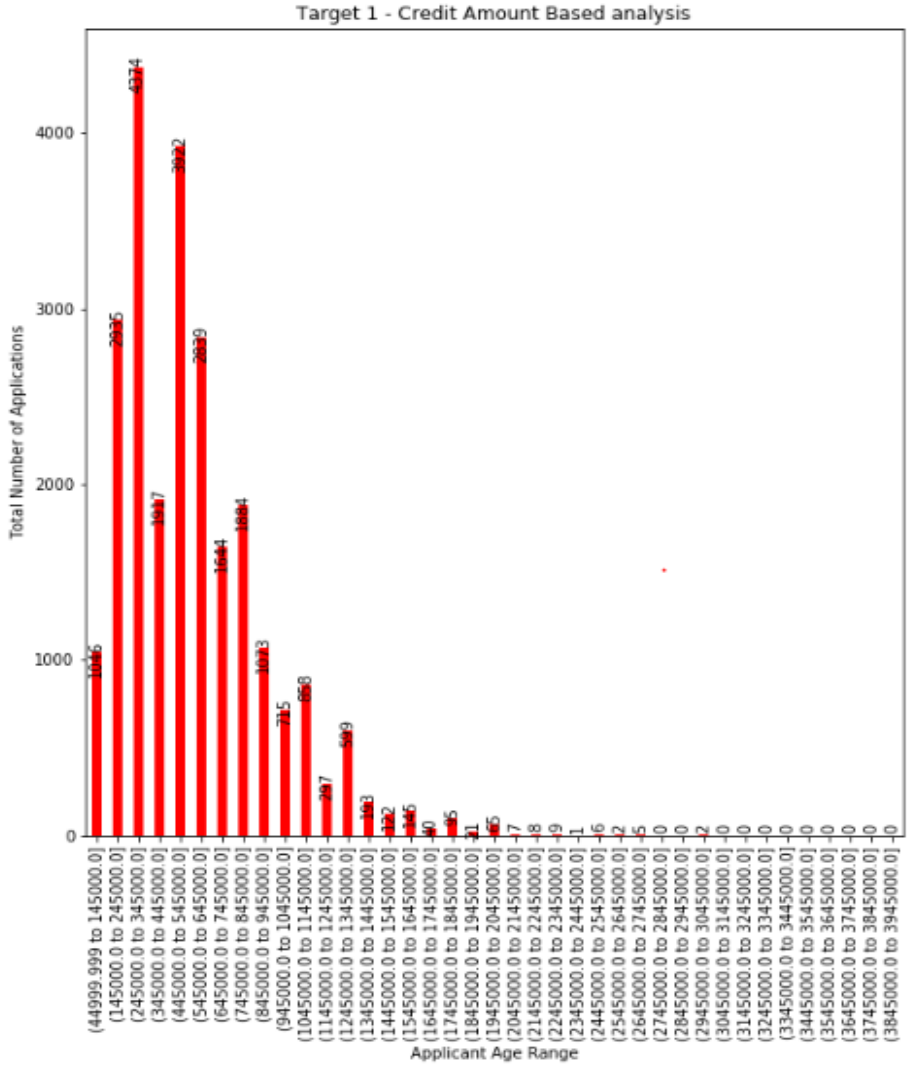


5. Univariate analysis for Credit Amount type:

The below bar plot represents trend of loan applications across ‘Credit amount of the loan’. From an overall perspective it is evident that top 3 loan applicants are in the credit group ’145000-245000-’, ’ 245000--345000-’, and ’445000--545000’. In contrast, ’1645000-1745000, ’ 1745000-1445000’ are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as 1345000-1445000’, 144500-1545000’, .

Observation: : Interesting insight with the below plot we can say that credit buckets 245000—345000 and ’445000--545000’, are likely to be potential defaulters when compared with other credit amount range groups w.r.t total applicants.

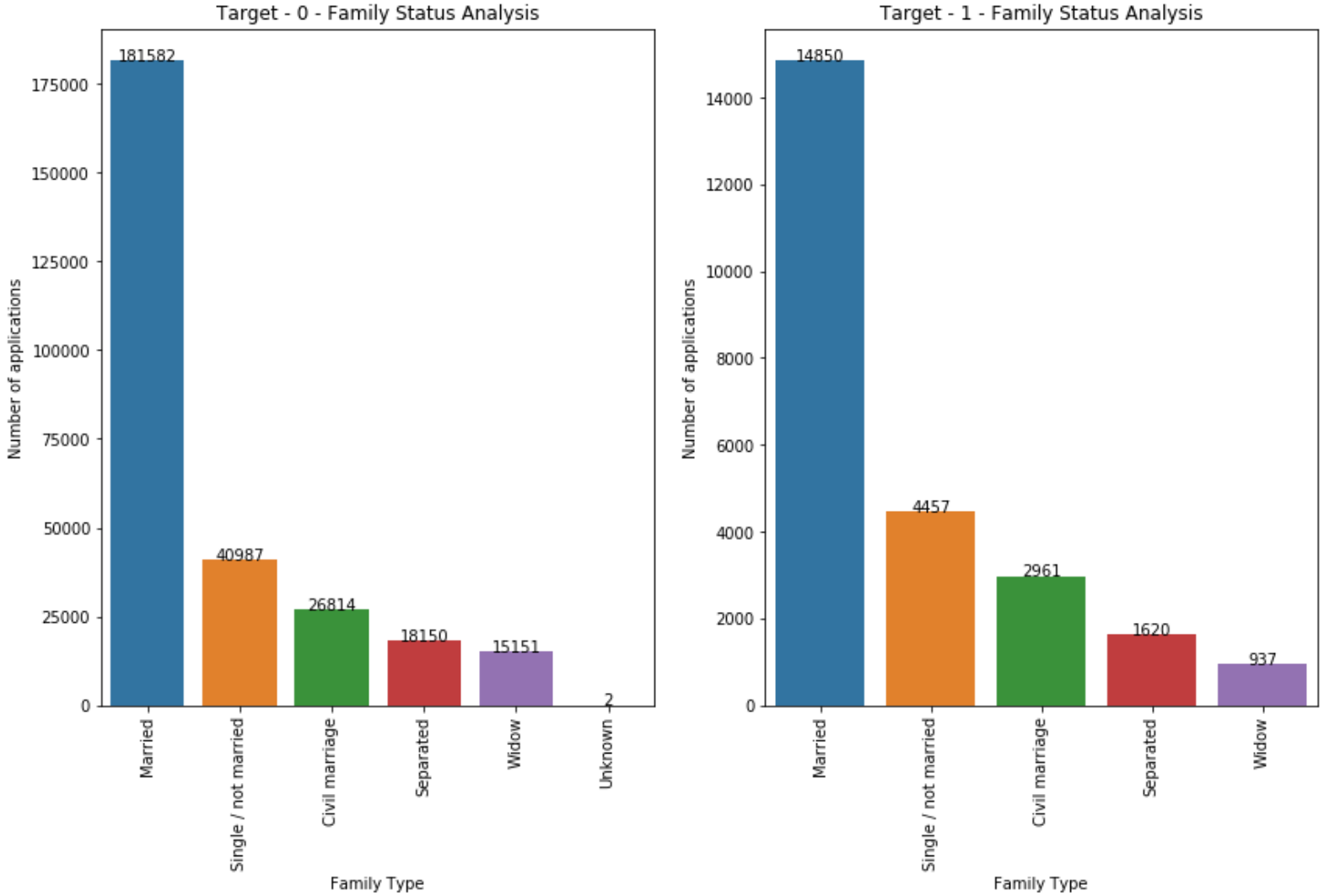




6. Univariate analysis for Family status type:

The below bar plot represents trend of loan applications on 'Family Status'. From an overall perspective it is evident that top 2 loan applicants are 'Married', and 'Single/not Married''. In contrast, 'Civil Marriage', 'Separated' and 'Widowed' are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as 'Married', and 'Single/not Married'.

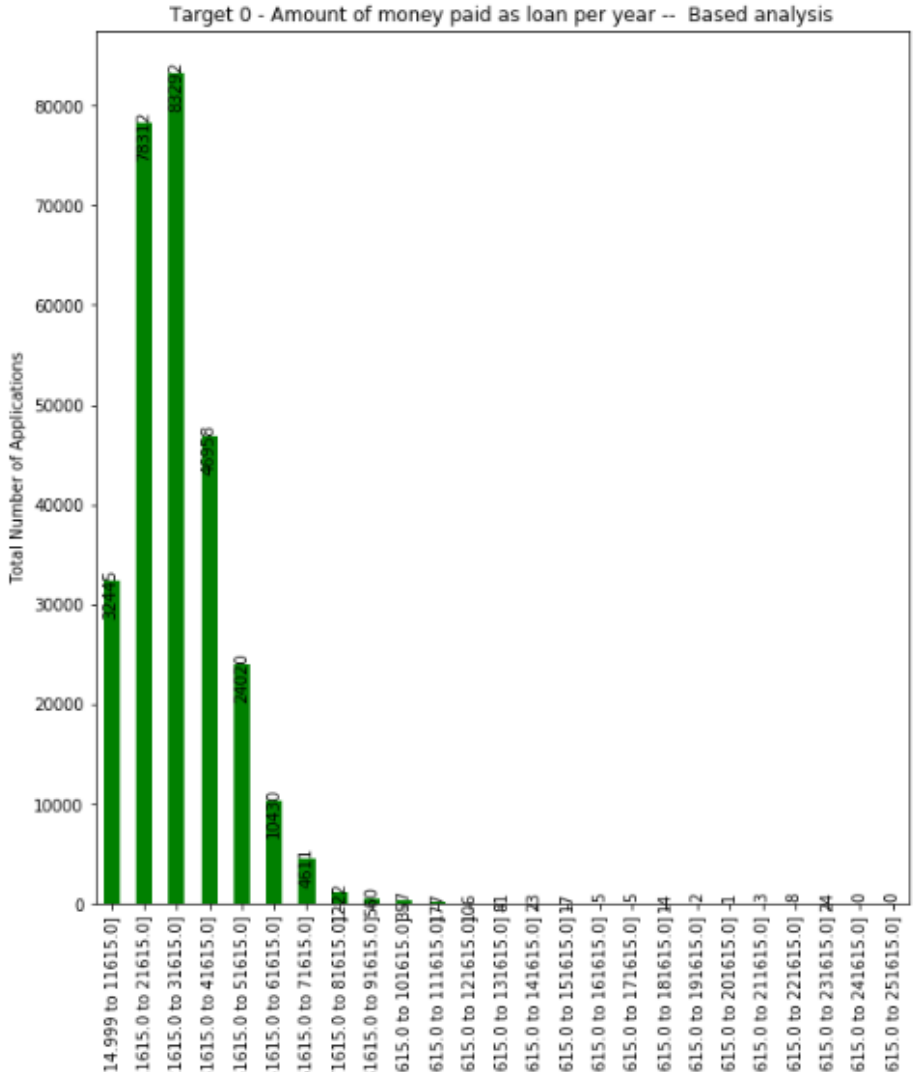
Observation: : that a significant number of Married and single/Not married applicants are likely to be be potential loan defaulters when compared with other Family status type w.r.t total applicants.

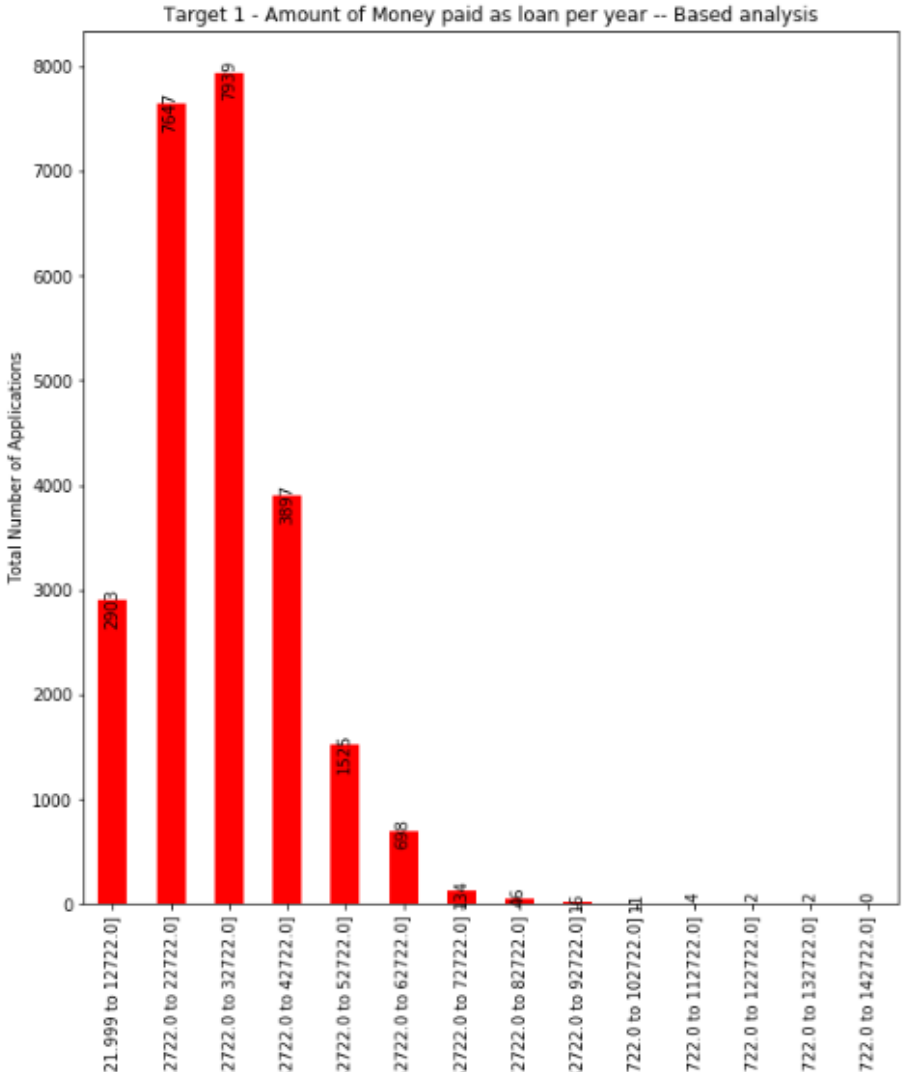


7. Univariate analysis for loan Annuity:

The below bar plot represents trend of loan applications on 'Annuity' or loan amount paid annually. From an overall perspective it is evident that top 2 annuity groups are '11615 -21615', and ' 21615 - 31615'. In contrast, '62722 - 72722, and '72722 - 82722' are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted 11615 -21615', and ' 21615 - 31615.

Observation: : that annuity groups 11615 -21615', and ' 21615 - 31615 applicants are likely to be potential loan defaulters when compared with other loan annuity groups w.r.t total applicants.

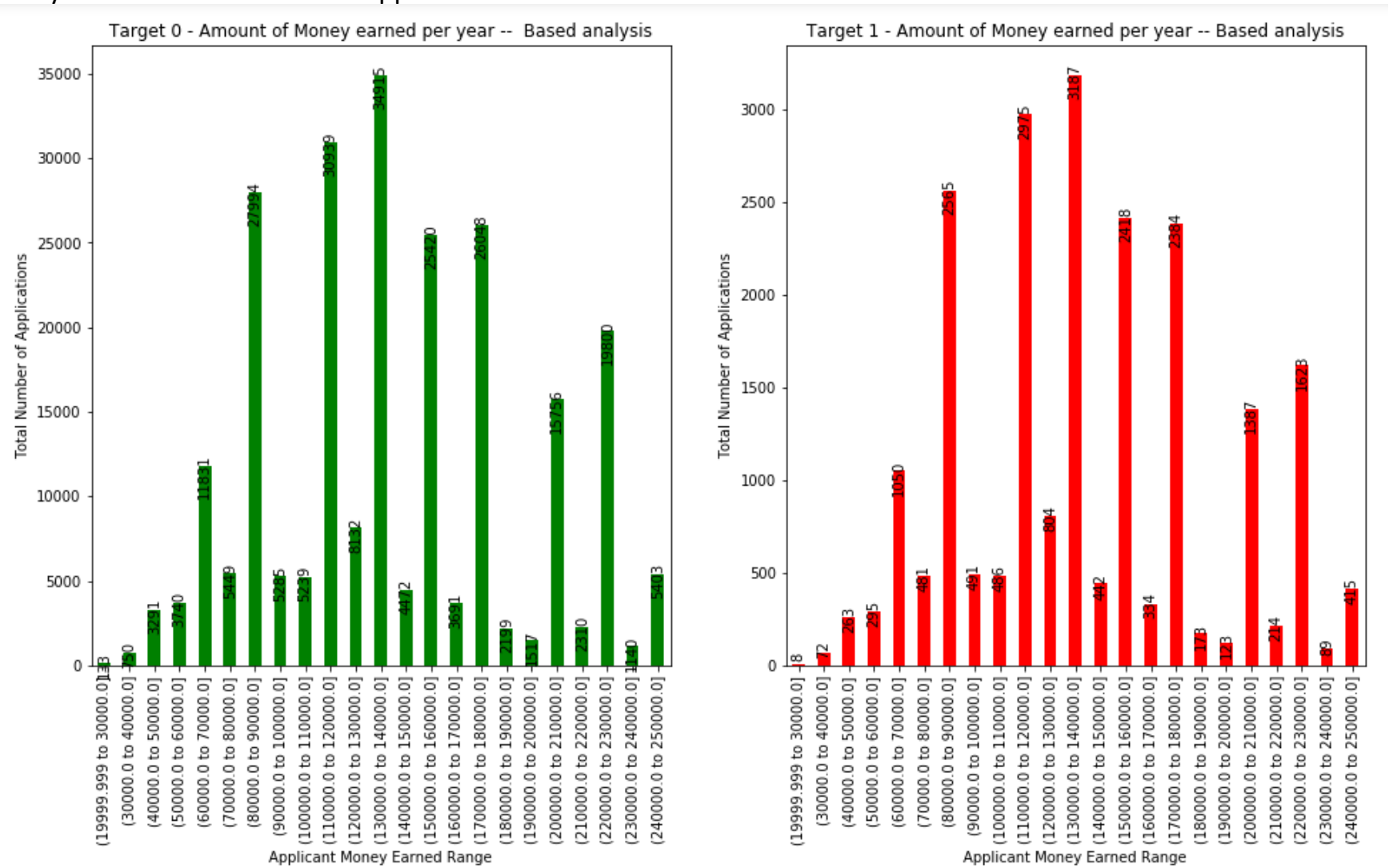




8. Univariate analysis for Income Total type:

The below bar plot represents trend of loan applications on 'Income earned annually. From an overall perspective it is evident that top 3 annuity groups are '80000-90000', and ' 110000 - 120000', '130000 - 140000'. In contrast, '190000 - 200000', and ' 210000 - 220000'' and 230000 - 240000' are the least loan applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted in group 110000 - 120000', and '130000 - 140000'.

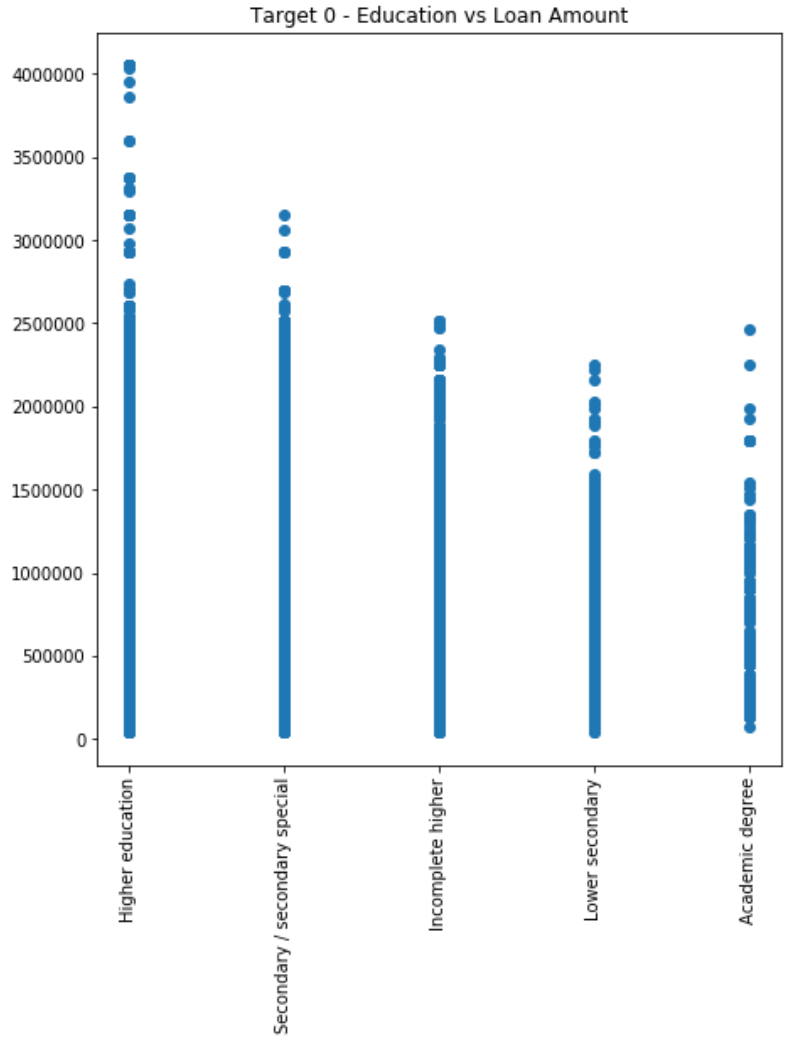
Observation: : that a income groups 110000 - 120000', and '130000 - 140000' applicants are likely to be potential loan defaulters when compared with other income earned total for an year w.r.t total number of applicants.

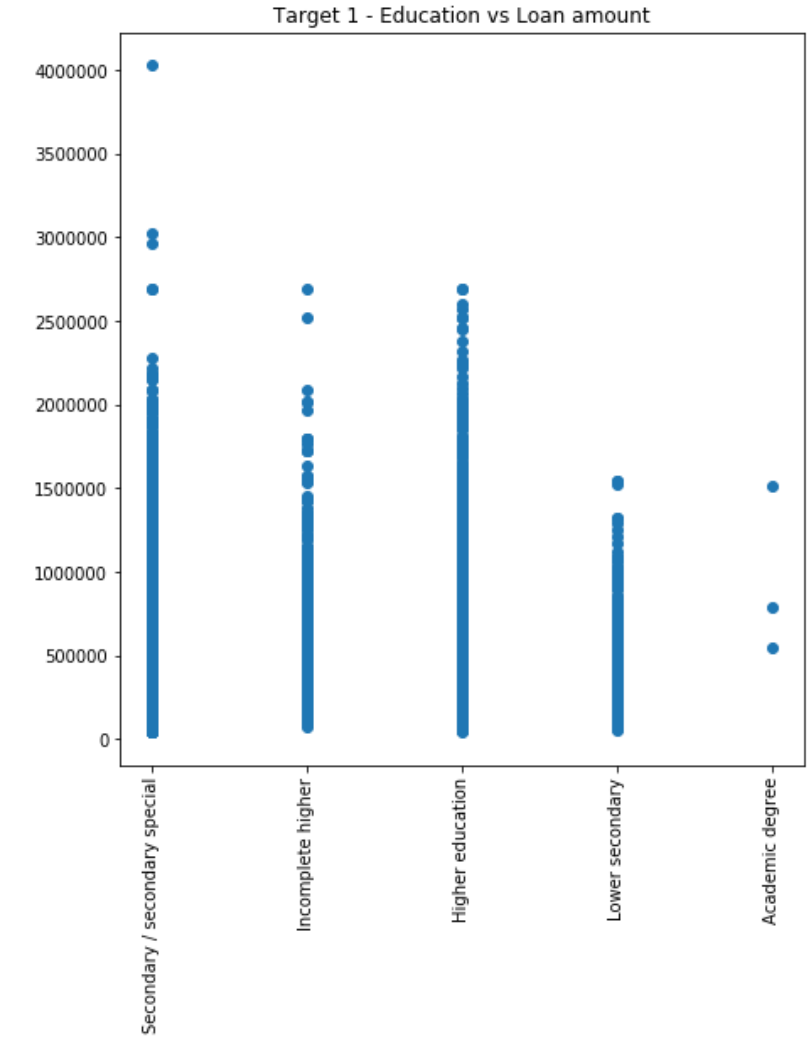


9. Bivariate analysis for Education Vs Loan Amount:

The below bar plot represents trend of loan applications against ‘Education’ vs ‘Loan amount’. From an overall perspective it is evident that top 3 applicants with higher loan amounts has education qualification as ‘Higher Education’, ‘secondary/Special’ and ‘incomplete higher’. In contrast, ‘Academic Degree’ and ‘lower secondary 'are the least loan amount applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as ‘Higher Education’, ‘secondary/Special’ and ‘incomplete higher.

Observation: : that a significant number of ‘Higher education’ and ’ Secondary/special Education’ applicants are likely to be potential loan defaulters when compared with other education types w.r.t total applicants.

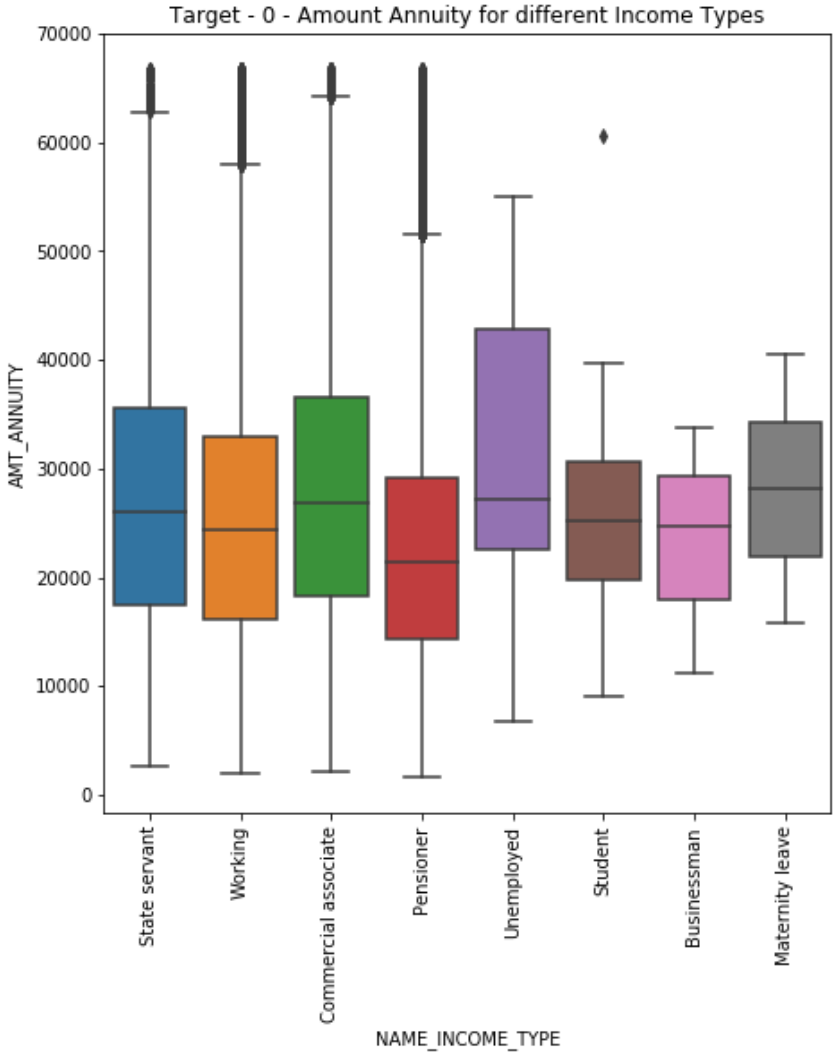


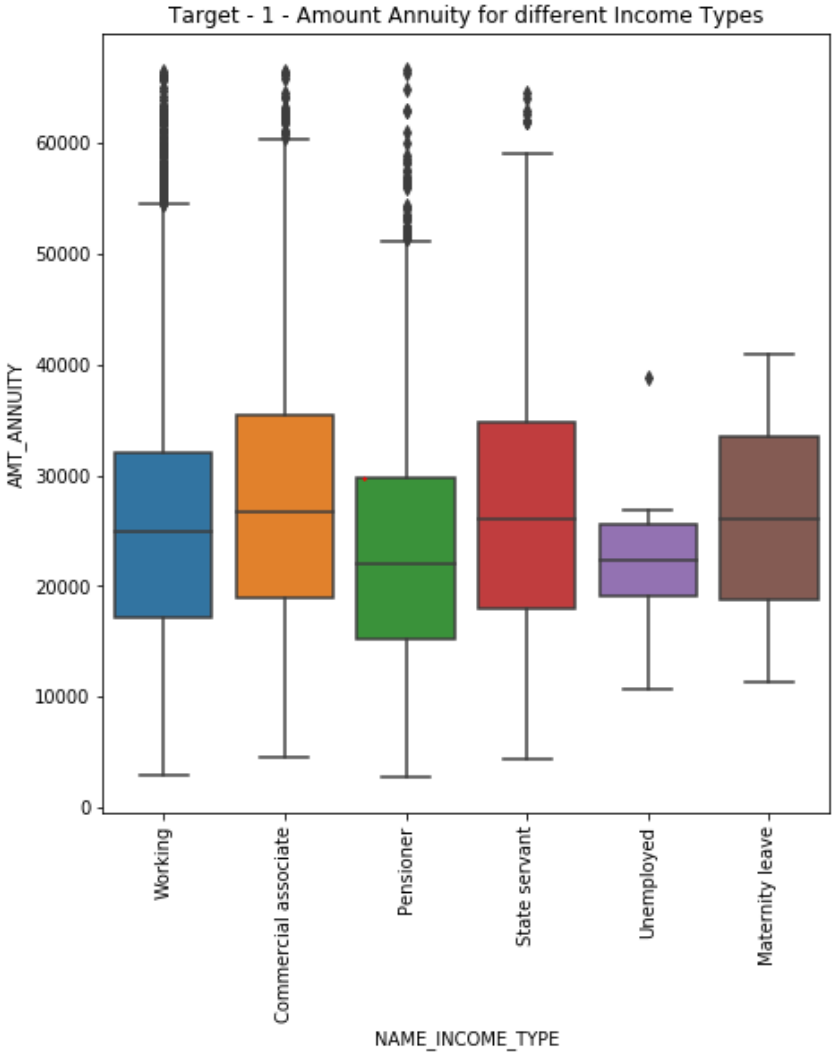


10. Bivariate analysis for Loan Annuity Vs Income type:

The below bar plot represents trend of loan applications ‘Loan Annuity’ vs ‘Income type’. From an overall perspective it is evident that top 3 income types with higher loan annuity are ‘Unemployed’, ‘State Servant’ and ‘Commercial associate’ . In contrast, ‘Business Man’ ‘Maternity Leave’ and ‘pensioner’ are the least loan amount applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as ‘State Servant’, ‘commercial associate’ and ‘Maternity leave.

Observation: : that some loan applicants of type ‘State Servant’ and ‘Commercial associate’ are likely to be top potential loan defaulters when compared with other income types w.r.t loan annuity amount .

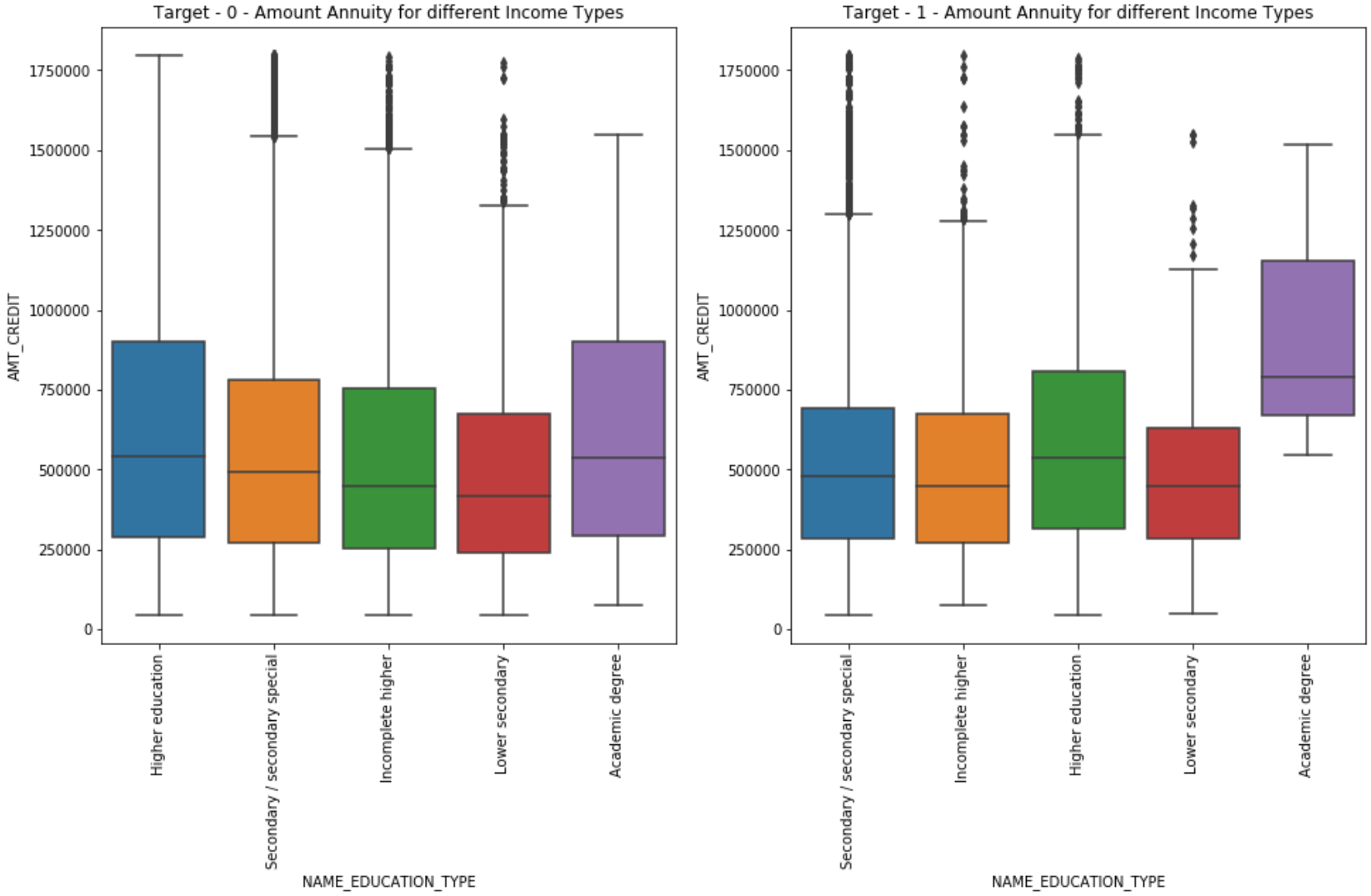




11. Bivariate analysis for Education Vs Loan Amount:

The below box plot represents trend of loan applications against 'Education' vs 'Loan Credit'. From an overall perspective it is evident that top 2 applicants education qualification w.r.t higher loan credit is 'Higher Education', 'Academic Degree'. In contrast, 'Lower Education' and 'Incomplete higher' are the least loan amount applicants. Similarly, the loan applicants **with payment difficulties** are clearly depicted as 'Academic Degree', 'Higher Education'.

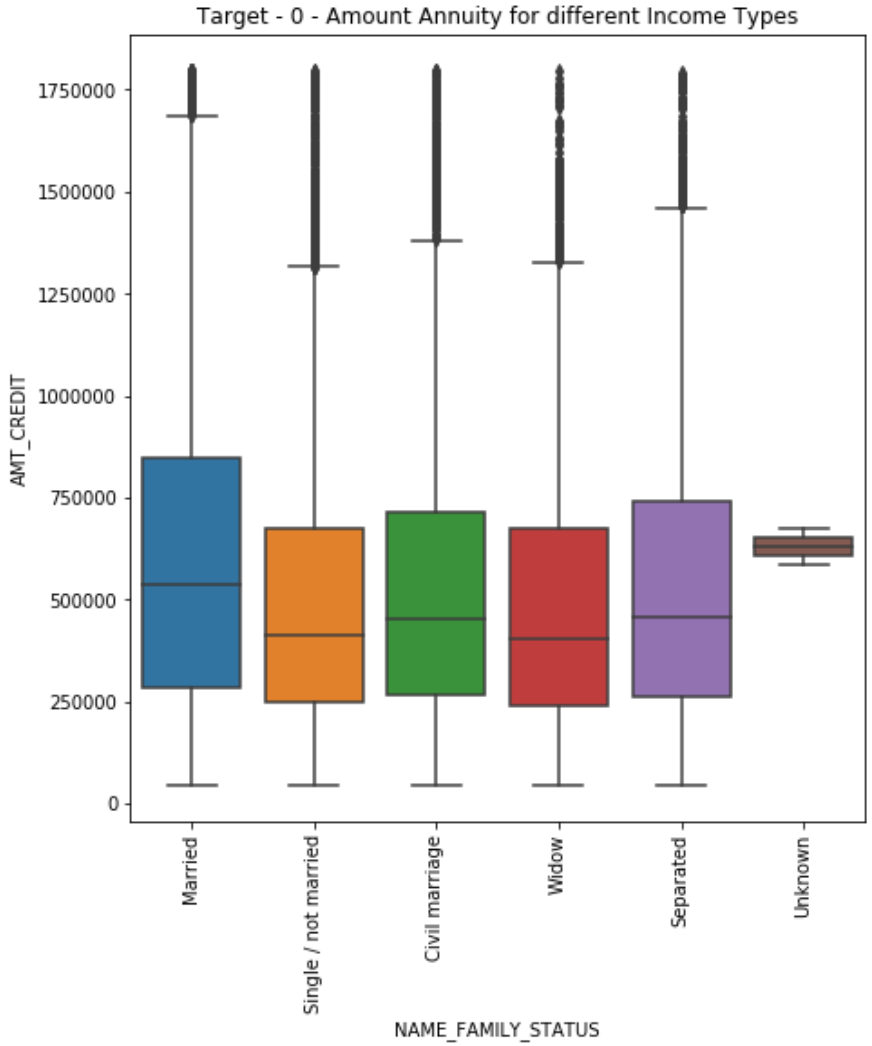
Observation: : that a significant number of 'Academic Degree', 'Higher Education' applicants are likely to be potential loan defaulters when compared with other Education types w.r.t loan credited.

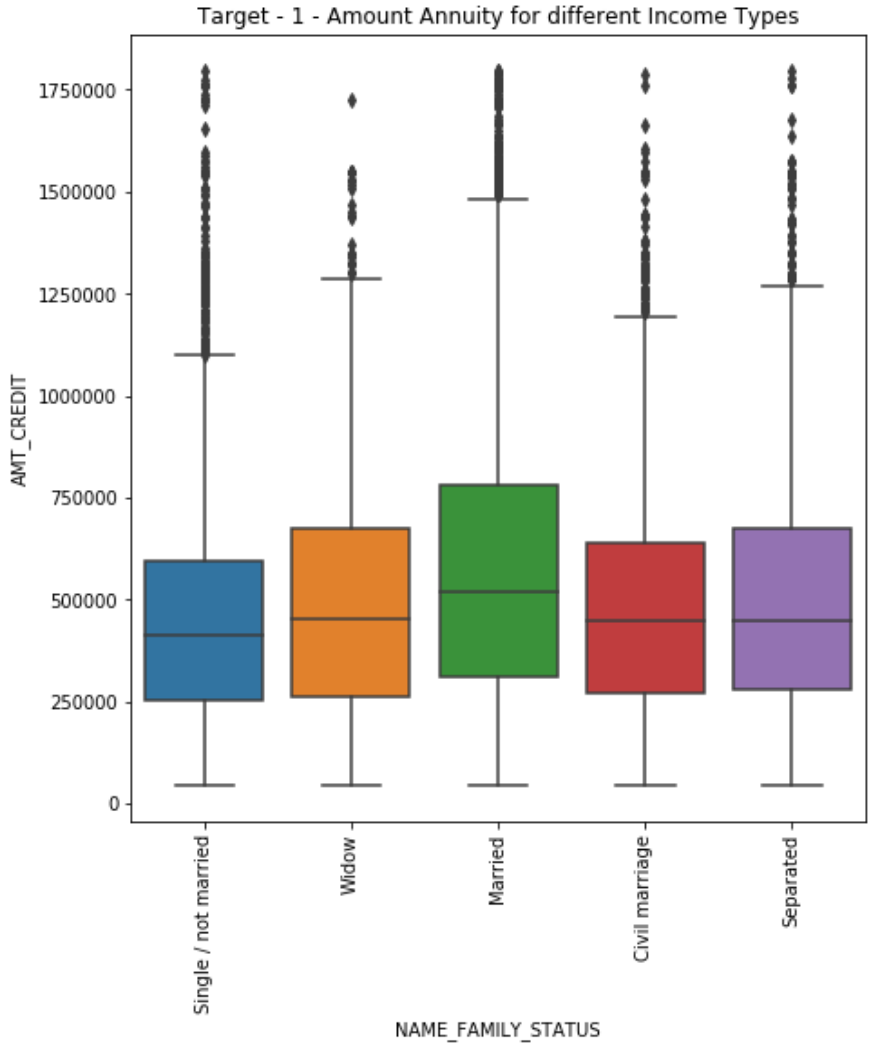


12. Bivariate analysis for Education Vs Loan Annuity:

The below box plot represents trend of loan applications against 'Loan annuity' vs 'Income Type'. From an overall perspective it is evident that top 2 applicants income type w.r.t family status is 'Married' and 'Separated'. In contrast, 'Single' and 'Widow' are the least loan amount applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Married' and 'Separated'.

Observation: : that a significant number of loan applicants 'Married' and 'Separated' are more likely to be potential defaulters when compared with other family status types w.r.t loan credit amount.

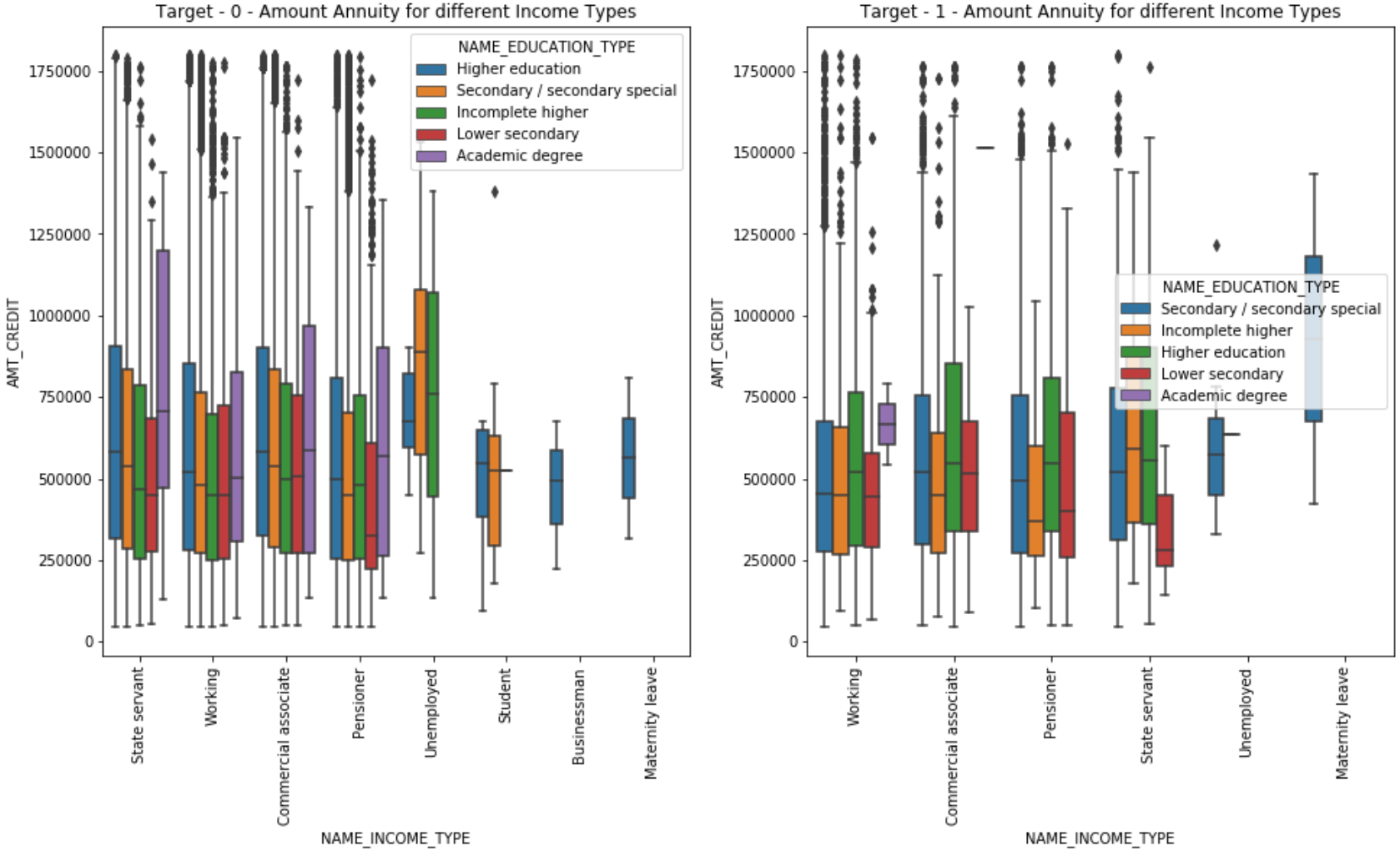




13. Multivariate/Bivariate analysis for Education Vs Loan credit Vs Income Type:

The below box plot represents trend of loan applications against 'Loan credit' vs 'Income Type' vs 'Education type'. From an overall perspective it is evident that top 2 loan credit applicants are 'State Servants' with 'Academic Degree' and 'Unemployed' with 'incomplete higher' education. In contrast, 'Business men' and 'Student' with 'Higher Education' are the least loan credit applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Maternity Leave' applicants with 'Secondary/special' and ' State Servants' with 'Higher education'.

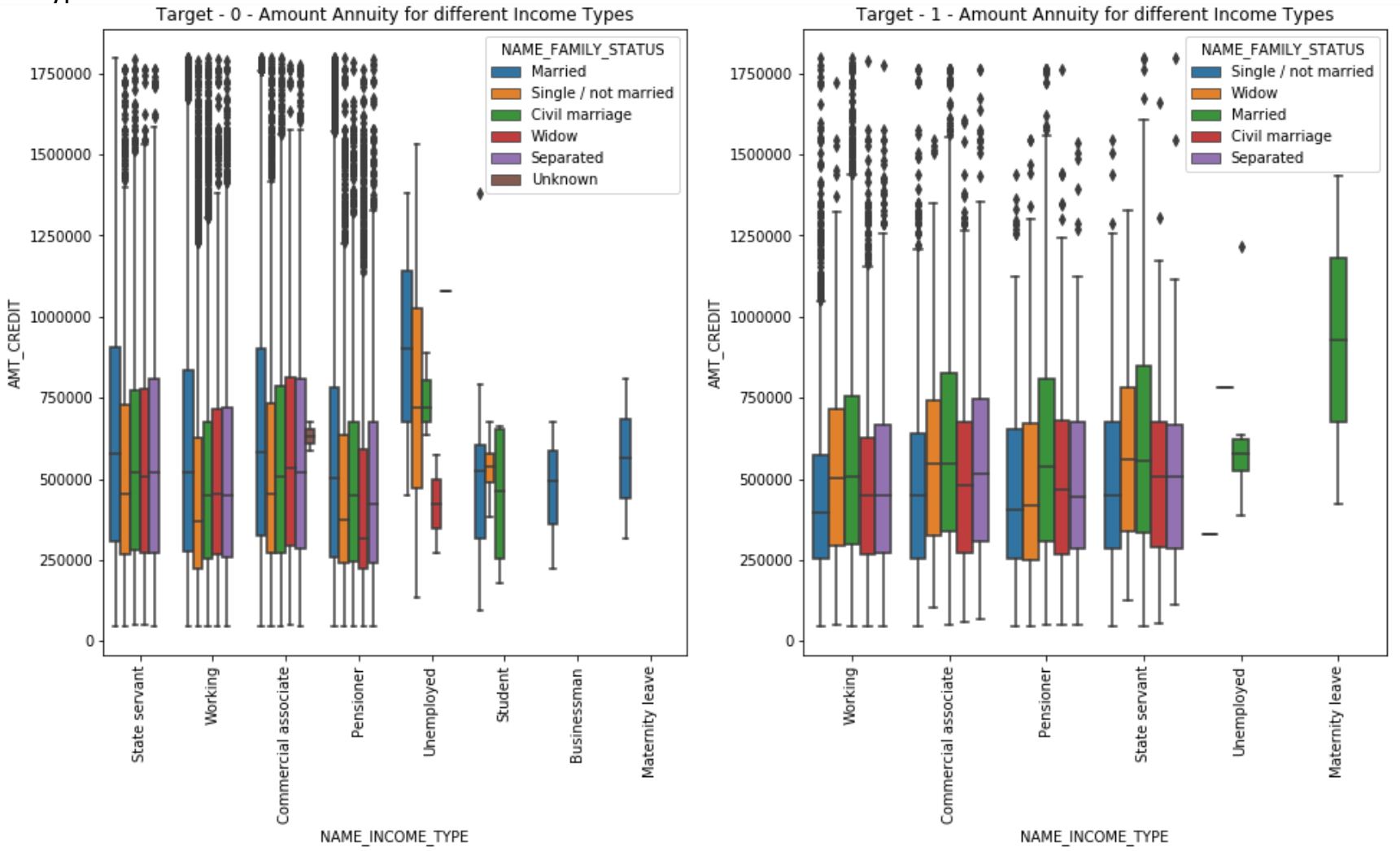
Observation: : that a significant number of ''Maternity Leave' applicants with 'Secondary/special' and 'State Servants' with 'Higher education' are more likely to be potential defaulters when compared with other education and income type. W.r.t loan credit



14. Multivariate/Bivariate analysis for Education Vs Loan credit Vs Income Type:

The below box plot represents trend of loan applications against 'Loan credit' vs 'Income Type' vs 'family status'. From an overall perspective it is evident that top 2 loan credit applicants are 'State Servants' and 'Unemployed' who are 'Married'. In contrast, 'Unemployed' 'Civil Marriage' and 'Student' who are 'Single/not married' are the least loan credit applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Maternity Leave' and 'Married' applicants and 'commercial' and ' Pensioner' who are 'Married'.

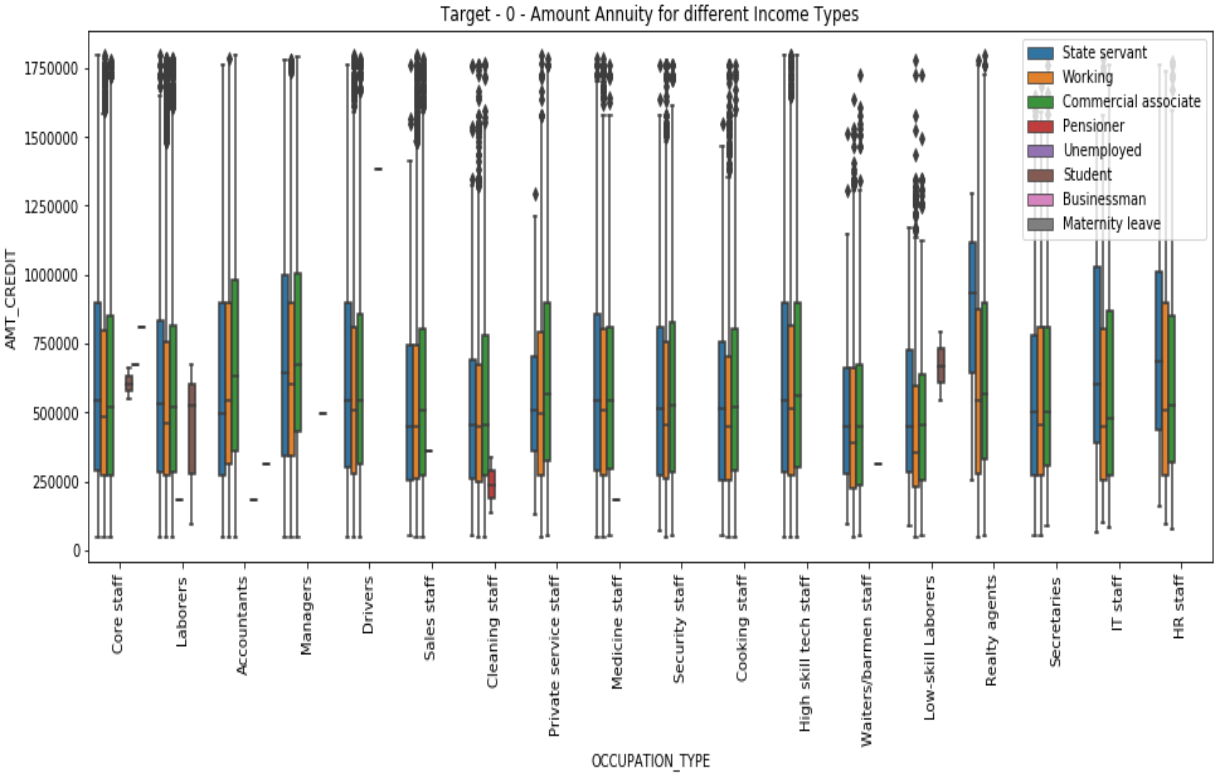
Observation: : that a significant number of ''Maternity Leave' applicants and 'commercial associates' who are 'Married' could also be potential defaulters when compared with other Income types w.r.t loan credit.

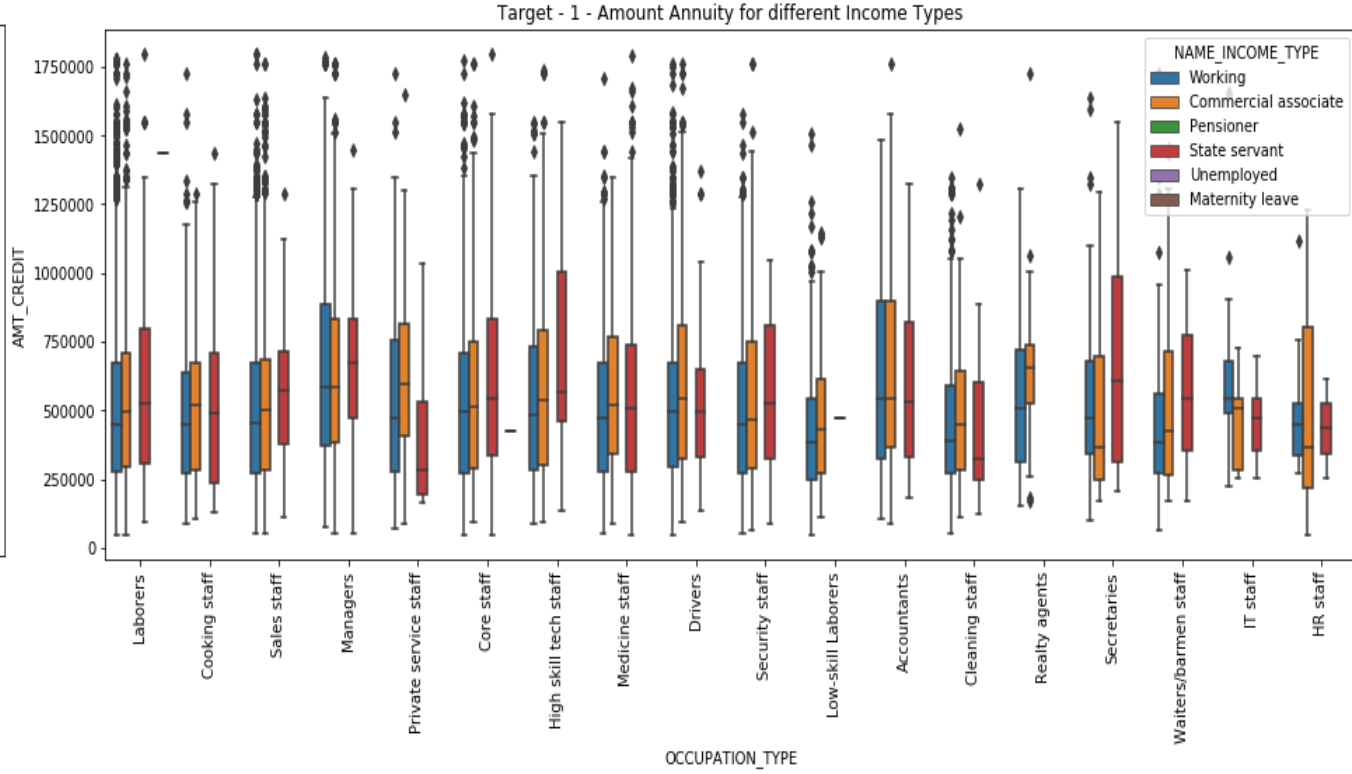


15. Multivariate/Bivariate analysis for Loan Annuity Vs Loan credit Vs Occupation Type:

The below box plot represents trend of loan applications against 'Loan Annuity' vs 'Loan credit' vs 'Occupation type'. From an overall perspective it is evident that top 2 loan credit applicants are 'Managers' ,'High skilled tech staff' who are 'State servants. In contrast, 'Low skilled Labors' ,'Waiters/Barmen staff' and 'Cleaning staff' who are either 'Commercial associates' or 'Working ' or 'State servants' are the least loan credit applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Secretaries', 'High skill tech staff' and 'Accountants' loan applicants who are either 'State servant', 'Working' or'Commercial associate'.

Observation: : Interesting insight with the below plots is that a significant number of ''Secretaries', 'High skill tech staff' applicants who are either 'State servant', 'Working' or 'Commercial associate' are likely to be potential defaulters when compared with other occupation type and loan annuity w.r.t loan credit.

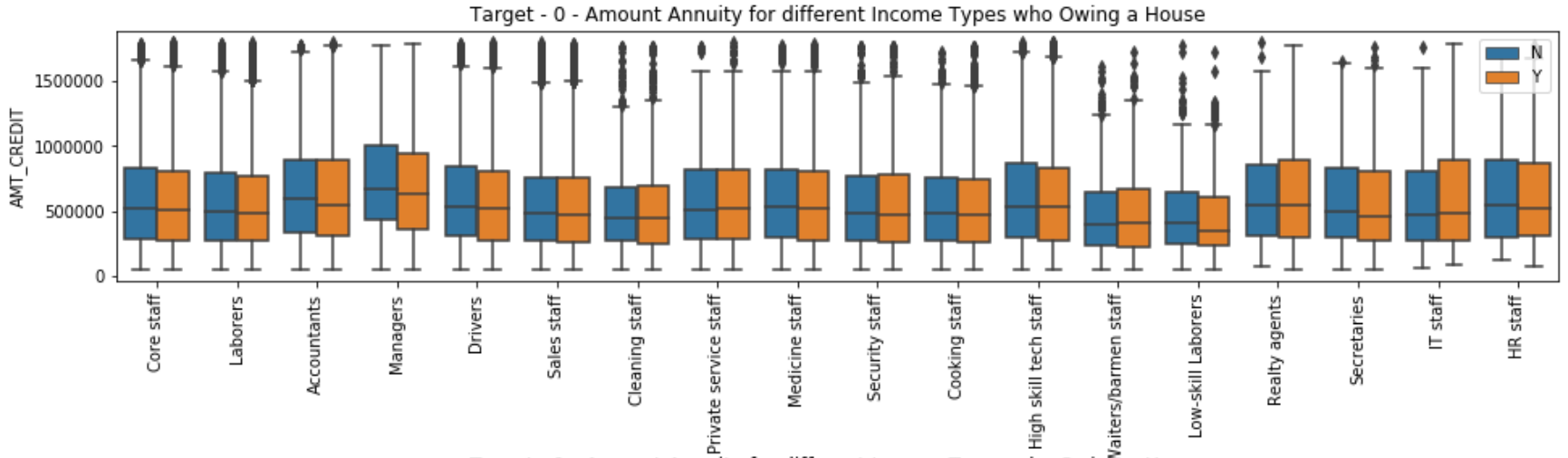


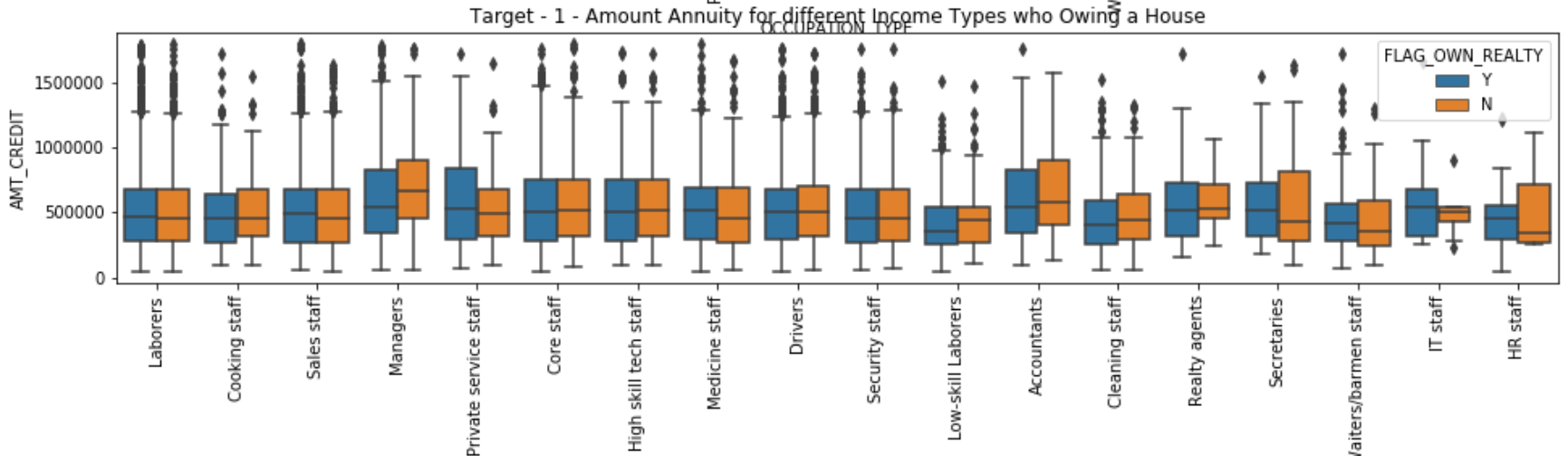


16. Multivariate/Bivariate analysis for Loan Annuity Vs Loan credit Vs Own House:

The below box plot represents trend of loan applications against 'Loan Annuity' vs 'Loan credit' vs 'Own House'. From an overall perspective it is evident that top 2 loan credit applicants are 'Managers' ,'High skilled tech staff' who have '**NO**' 'Own House' and 'Manages', 'IT Staff' who have 'YES' own house. In contrast, 'Cleaning staff', 'Low skilled Labors' ,**who have and do not have a Own house** are the least loan credit applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Managers', 'Secretaries' and 'Accountants' who Own a House.

Observation: : Interesting insight with the below plots is that a significant number of 'Managers', 'Secretaries' and 'Accountants' who Own a House are likely to be potential defaulters when compared with who do not own a house w.r.t loan credit.

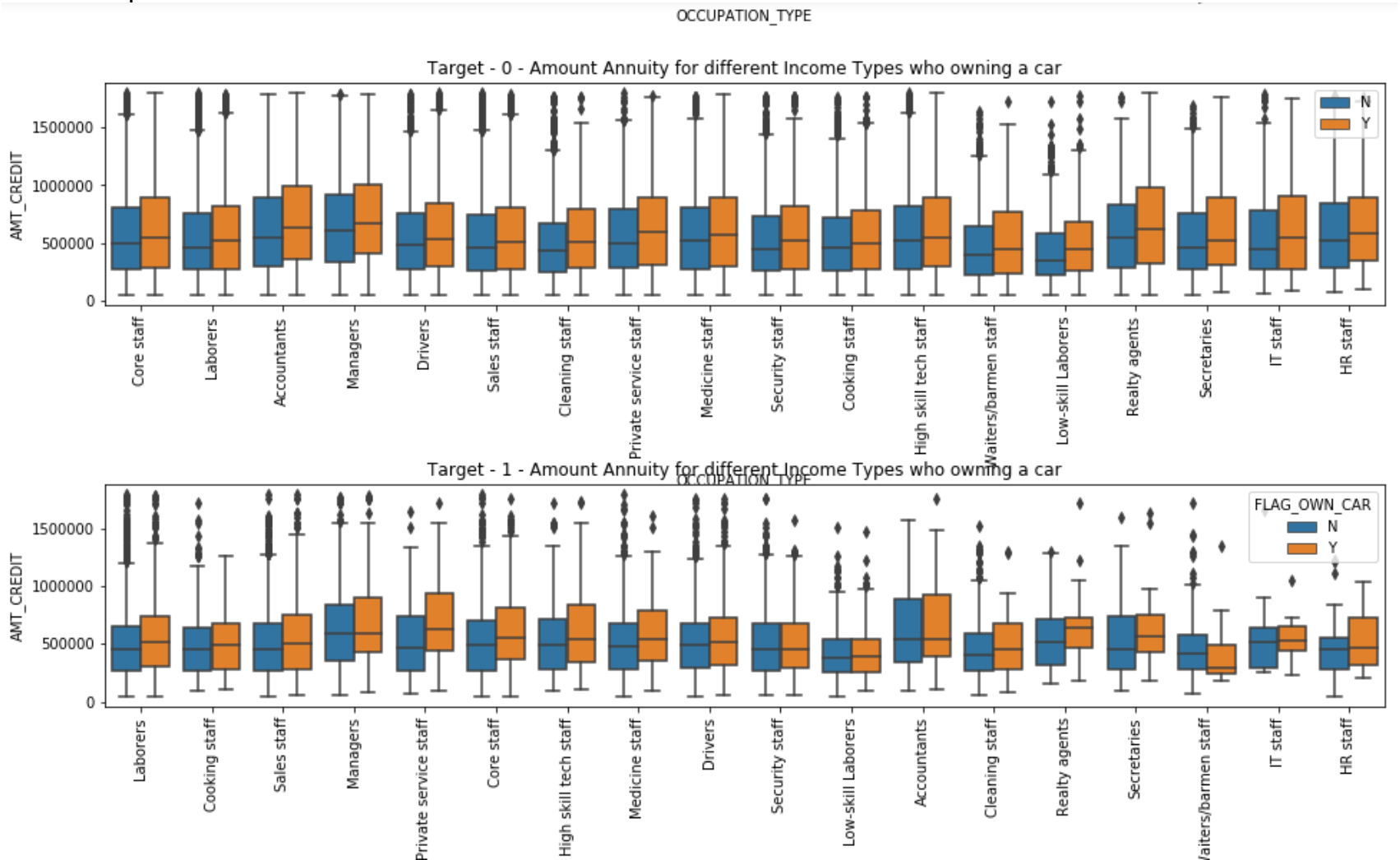




17. Multivariate/Bivariate analysis for Loan Annuity Vs Loan credit Vs Own Car:

The below box plot represents trend of loan applications against 'Loan Annuity' vs 'Loan credit' vs 'Own House'. From an overall perspective it is evident that top 2 loan credit applicants are 'Managers' ,'Accountants' who have '**NO**' 'Own Car' and 'Manages', 'Reality Agents' who have 'YES' own Car. In contrast, 'Cleaning staff', 'Low skilled Labors' ,**who have and do not have a Own Car** are the least loan credit applicants. Importantly, the loan applicants **with payment difficulties** are clearly depicted as 'Managers', 'Waiter/barmen staff' and 'Accountants' who Own a Car.

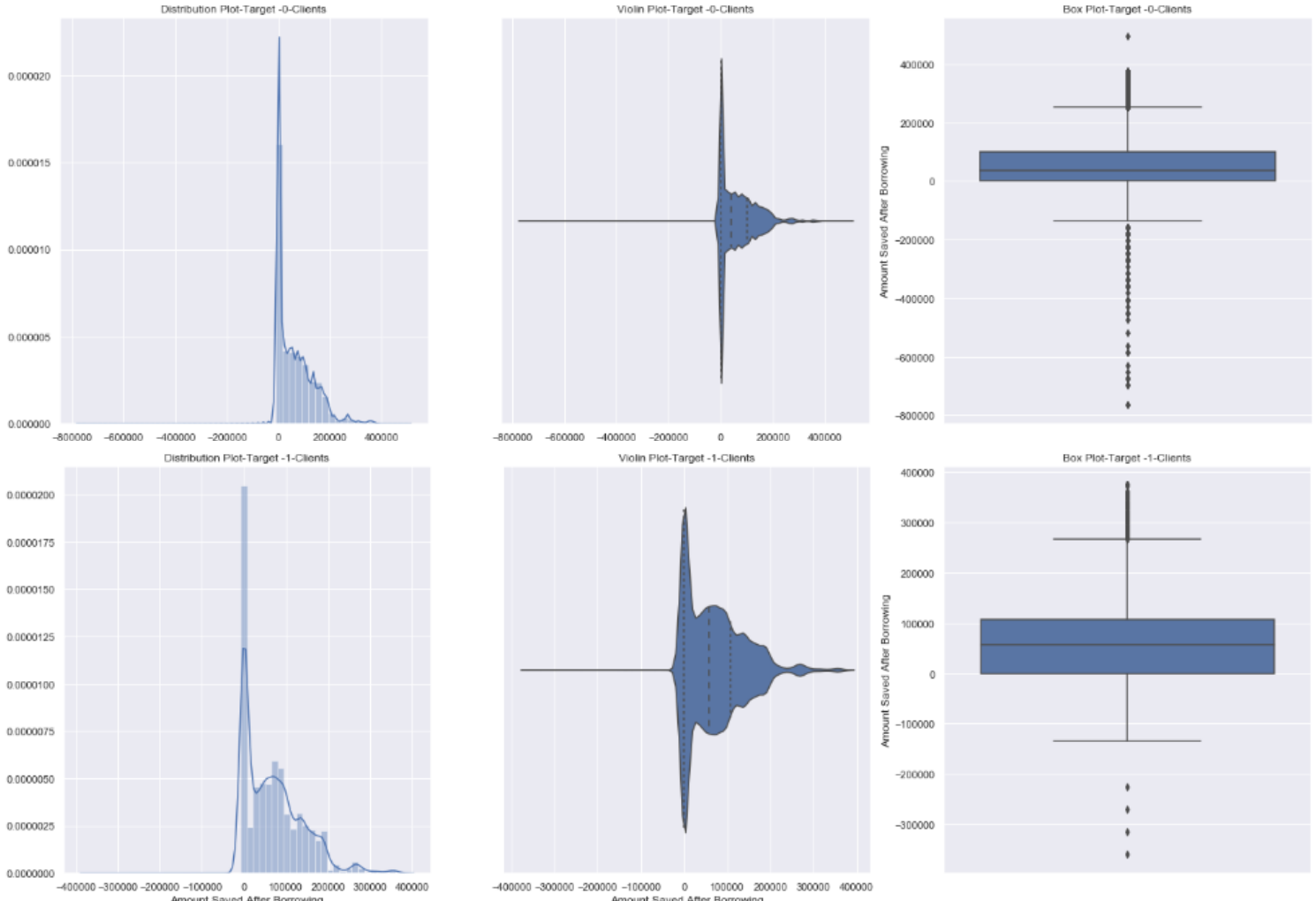
Observation: : Interesting insight with the below plots is that a significant number of 'Managers', 'Waiter/barmen staff' and 'Accountants' **who Own a Car** could also be potential defaulters when compared with who do not own a car w.r.t loan credit.



18. Multivariate/Bivariate analysis for Additional amount= (Amount credit - Goods Price):

The below box plot represents trend of loan applications against 'Amount credit' vs 'Goods Price'. From an overall perspective it is evident that high number of loan applicants in 'Target 0' group who receive additional credit more than the goods price (is close to ~1,50,000)and are less likely to default as the additional credit distribution is 0.000005 less below . In contrast, high number of loan applicants in 'Target 1' group who receive credit more than the goods price (is close to ~1,20,000)and are highly likely to default. Importantly, a significant number of loan applicants **with payment difficulties** are clearly in 'Target 1' group which is **0.0000050**.

Observation: : Interesting insight with the below plots is that a **significant number** of loan applicants in Target 1 group are more likely to default though they received loan credit higher than the goods price when compared with Target 0 loan applicants w.r.t additional credit(Box plot).



Conclusion: *Interesting insights from all the plots can be summarized*

1. *Loan applicants of Laborers', ' sales staff', 'Core Staff', who are not highly skilled labors could be potential loan defaulters when compared with other occupation types*
2. *Loan applicants of Working', 'Commercial staff', 'Pensioner', who receive low income are likely to be potential loan defaulters when compared with other income types.*
3. *Loan applicants of loan applicants who are employed less number of days are likely to be potential loan defaulters when compared with total number of applicants with more experience.*
4. *Loan applicants of though age group 20-25' are less number of applicants, they have more difficulties to pay the loan compared with age groups 25-30', '35-40' and are prone to risk of potential loan defaulters when compared with other age groups.*
5. *Loan applicants of credit buckets 245000—345000 and '445000--545000', are likely to be potential defaulters when compared with other credit amount range groups.*
6. *Loan applicants of a significant number of Married and single/Not married applicants are likely to be potential loan defaulters when compared with other Family status type w.r.t total applicants.*
7. *Loan applicants annuity groups of 11615 -21615', and ' 21615 - 31615 applicants are likely to be potential loan defaulters when compared with other loan annuity groups w.r.t total applicants.*
8. *Loan applicants of income groups 110000 - 120000', and '130000 - 140000' applicants are likely to be potential loan defaulters when compared with other income earned total for an year w.r.t total number of applicants*
9. *Loan applicants significant number of 'Higher education' and ' Secondary/special Education' applicants are likely to be potential loan defaulters when compared with other education types w.r.t total applicants.*
10. *Loan applicants of some loan applicants of type 'State Servant' and 'Commercial associate' are likely to be top potential loan defaulters when compared with other income types w.r.t loan annuity amount.*
11. *Loan applicants of a significant number of 'Academic Degree', 'Higher Education' applicants are likely to be potential loan defaulters when compared with other Education types w.r.t loan credited.*
12. *Loan applicants of a significant number of loan applicants 'Married' and 'Separated' are more likely to be potential defaulters when compared with other family status types w.r.t loan credit amount.*
13. *Loan applicants of a significant number of ''Maternity Leave' applicants with 'Secondary/special' and 'State Servants' with 'Higher education' are more likely to be potential defaulters when compared with other education and income type. W.r.t loan credit*
14. *Loan applicants of a significant number of ''Maternity Leave' applicants and 'commercial associates' who are 'Married' could also be potential defaulters when compared with other Income types w.r.t loan credit.*
15. *Loan applicants of a significant number of ''Secretaries', 'High skill tech staff' applicants who are either 'State servant', 'Working' or 'Commercial associate' are likely to be potential defaulters when compared with other occupation type and loan annuity w.r.t loan credit.*
16. *Loan applicants of a significant number of 'Managers', 'Secretaries' and 'Accountants' who Own a House are likely to be potential defaulters when compared with who do not own a house w.r.t loan credit.*
17. *Loan applicants of a significant number of 'Managers', 'Waiter/barmen staff' and 'Accountants' who Own a Car could also be potential defaulters when compared with who do not own a car w.r.t loan credit.*
18. *Loan applicants of a significant number of loan applicants in Target 1 group are more likely to default though they received loan credit higher than the goods price when compared with Target 0 loan applicants w.r.t additional credit(Box plot).*