Credit EDA case study

Presented By-
Shivam Gupta and Simran Jain

Data _Set

- This dataset has 3 files as explained below:
- *1. 'application_data.csv'* contains all the information of the client at the time of application.
  The data is about whether a **client has payment difficulties.**
- 2. '*previous_application.csv*' contains information about the client's previous loan data.
- 3. '*columns_description.csv*' is data dictionary which describes the meaning of the variables.

ANALYSIS OF THE DATA

1. Upon analysis of the data we found out that the data has 307511 rows and 122 columns.

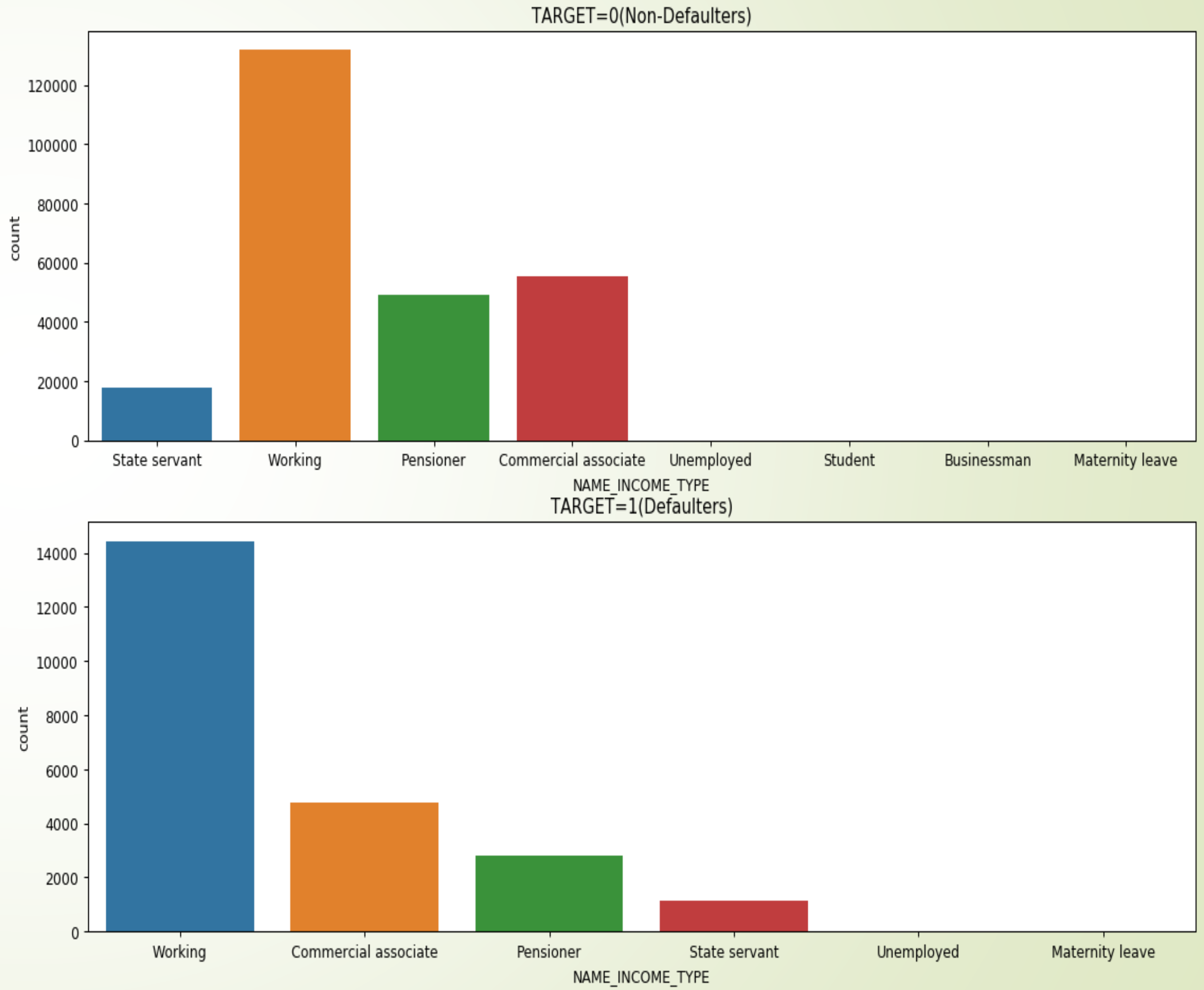
2. Further analysis revealed that data has some missing values which needs to be treated.

3. Idea behind the case study was to minimize the risk of losing money while lending it to lenders.

4. This case study aims to identify patterns which indicate if a client has difficulty paying their instalments 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.

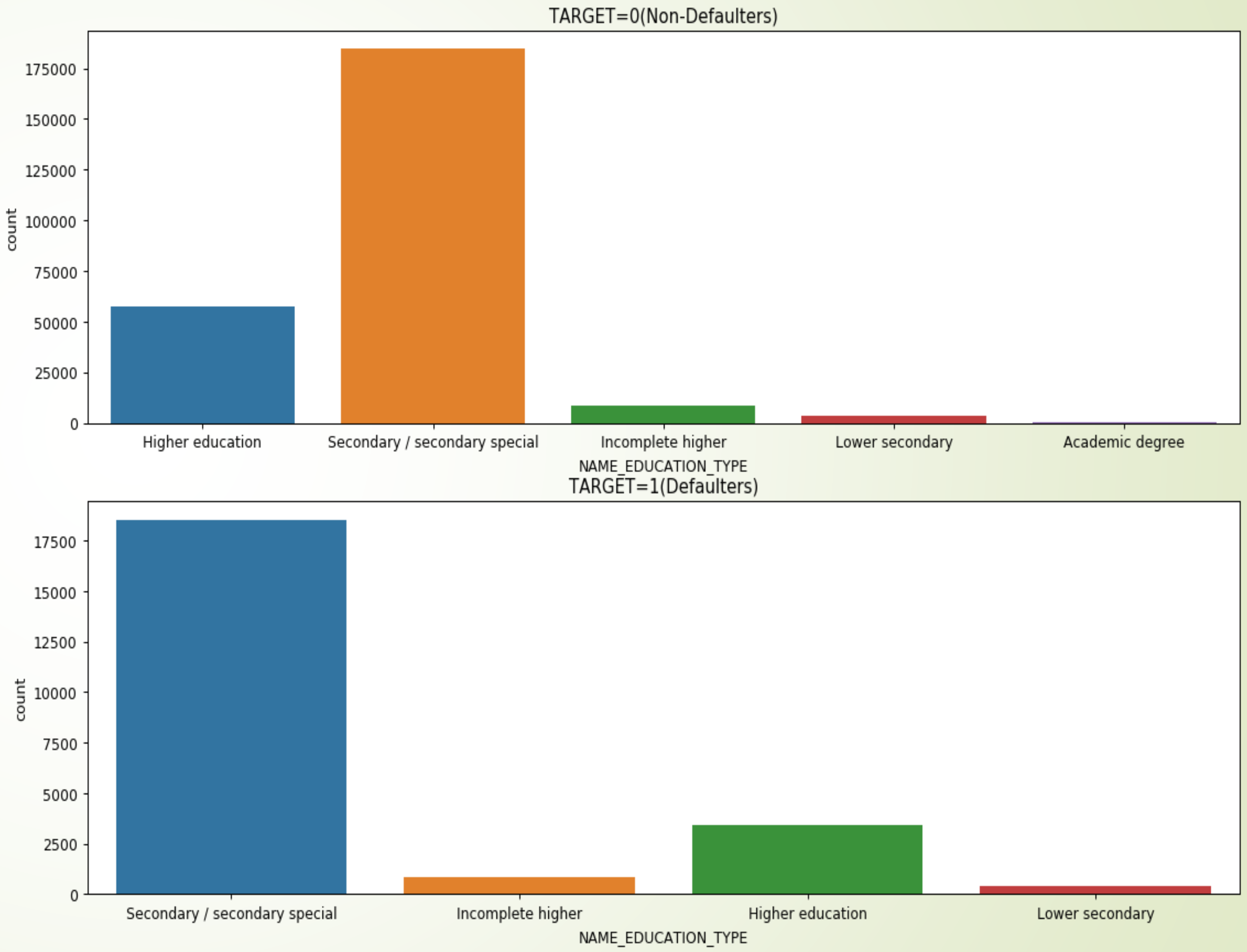
Analysis of the data Graph 1:

- On analyzing the data we could see that for name_income_type those who are working have the highest number in both the list for defaulters as well as non defaulters.
- While the state servant contribute to the lowest and is more likely to be a non defaulter.



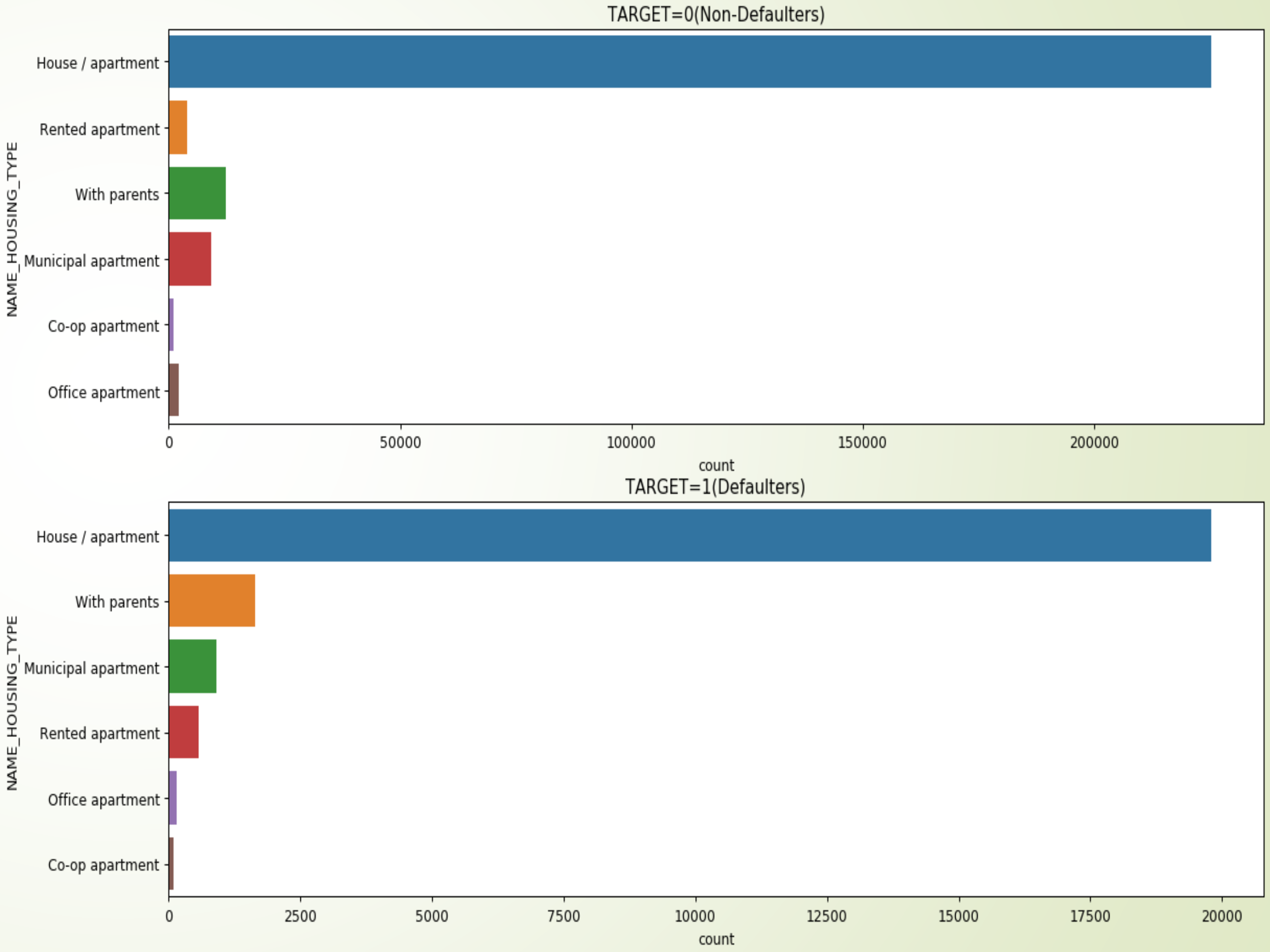
GRAPH 2:

- On analyzing the data we could see that for name_education_type those who have completed secondary education have the highest number in both the list for defaulters as well as non defaulters.
- While those who have completed higher education tend to be the non defaulters.



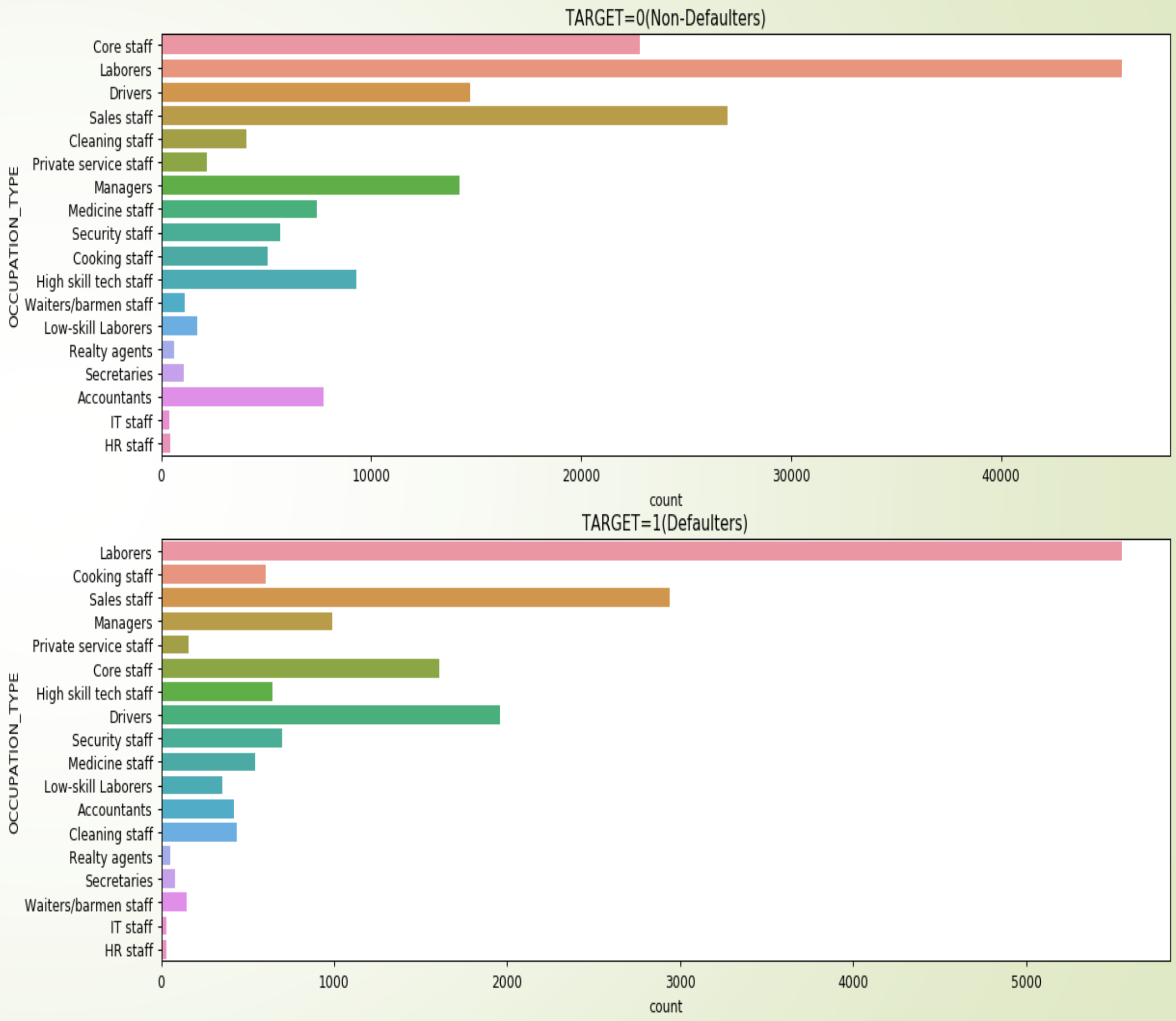
GRAPH 3:

- On analyzing the data we could see that for name_housing_type those who have house/apartments or live with parents tend to be non defaulters.
- While those live in office apartment are more likely to be a defaulter.



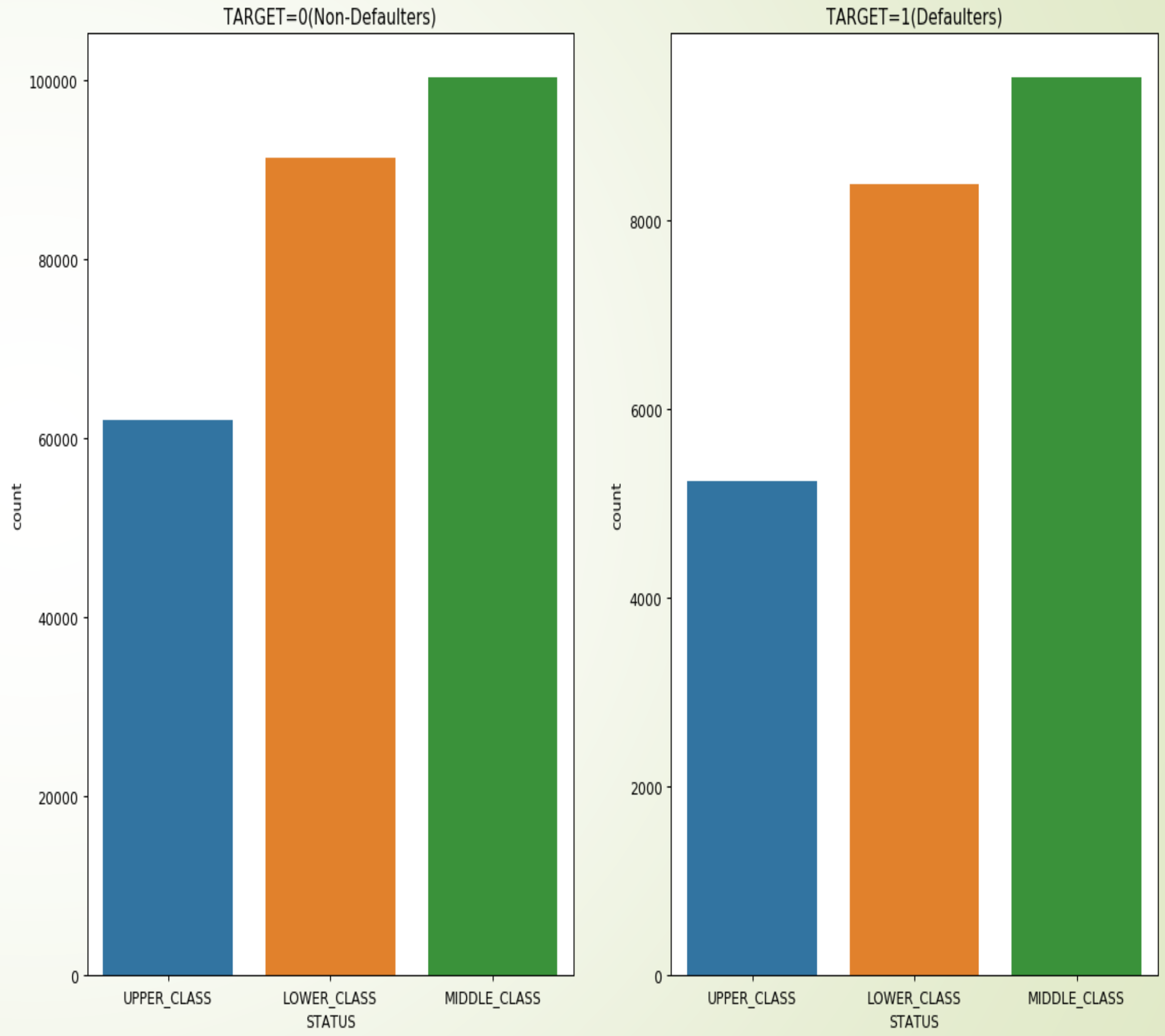
GRAPH 4:

- On analyzing the data we could see that for occupation_type those who laborers tend to be non defaulters.
- As well as drivers , sales staff , managers , high skill tech staff seems to be non defaulter.
- While it staff and HR staff seem to be a defaulter.



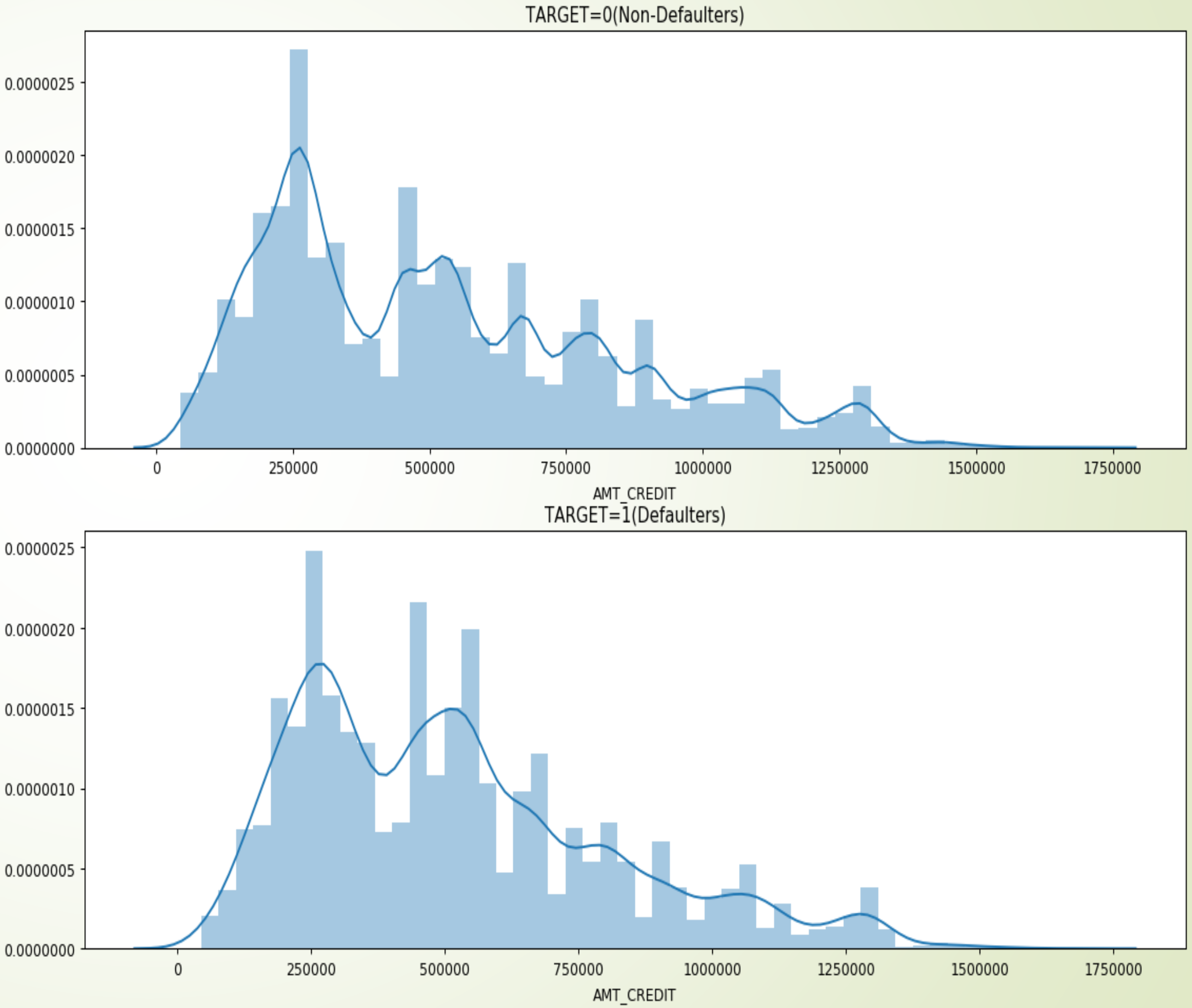
GRAPH 5:

- On analyzing the data we could see that based on status those who belong to middle are more likely to be a defaulter.
- While the upper class tends to be the non defaulter.



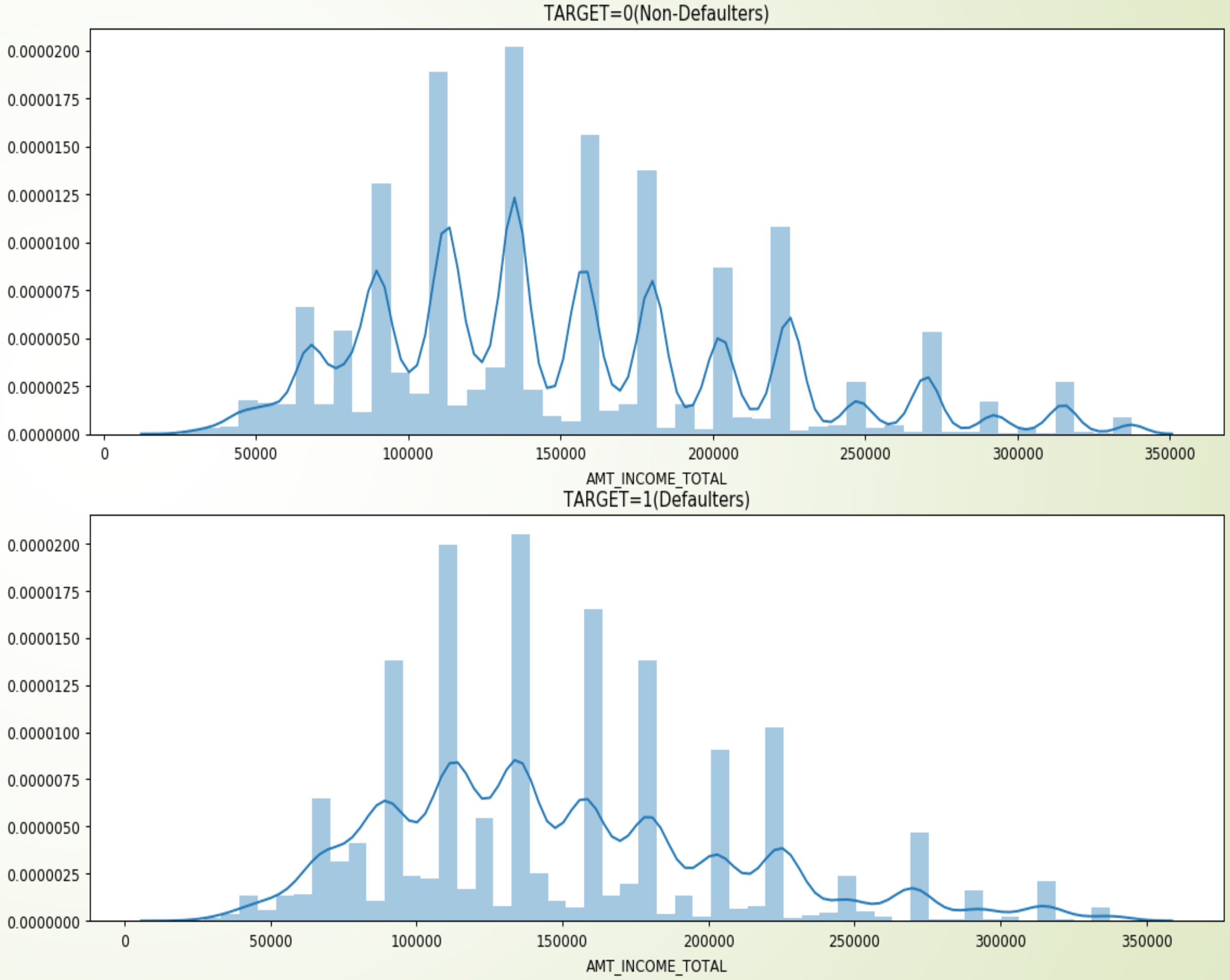
Graph 6:

- From the graph we can incur that when the amount credit is in between 250000 and 500000 then the probability of the person being defaulter is more.
- While when the amount credit is around 250000 the probability of not being a defaulter is more.



Graph 7:

- We can make an observation that when amount income total is in between 100000 and 150000 then probability of being non defaulter is more.
- While in case of non defaulter we can see fluctuation in the probability.



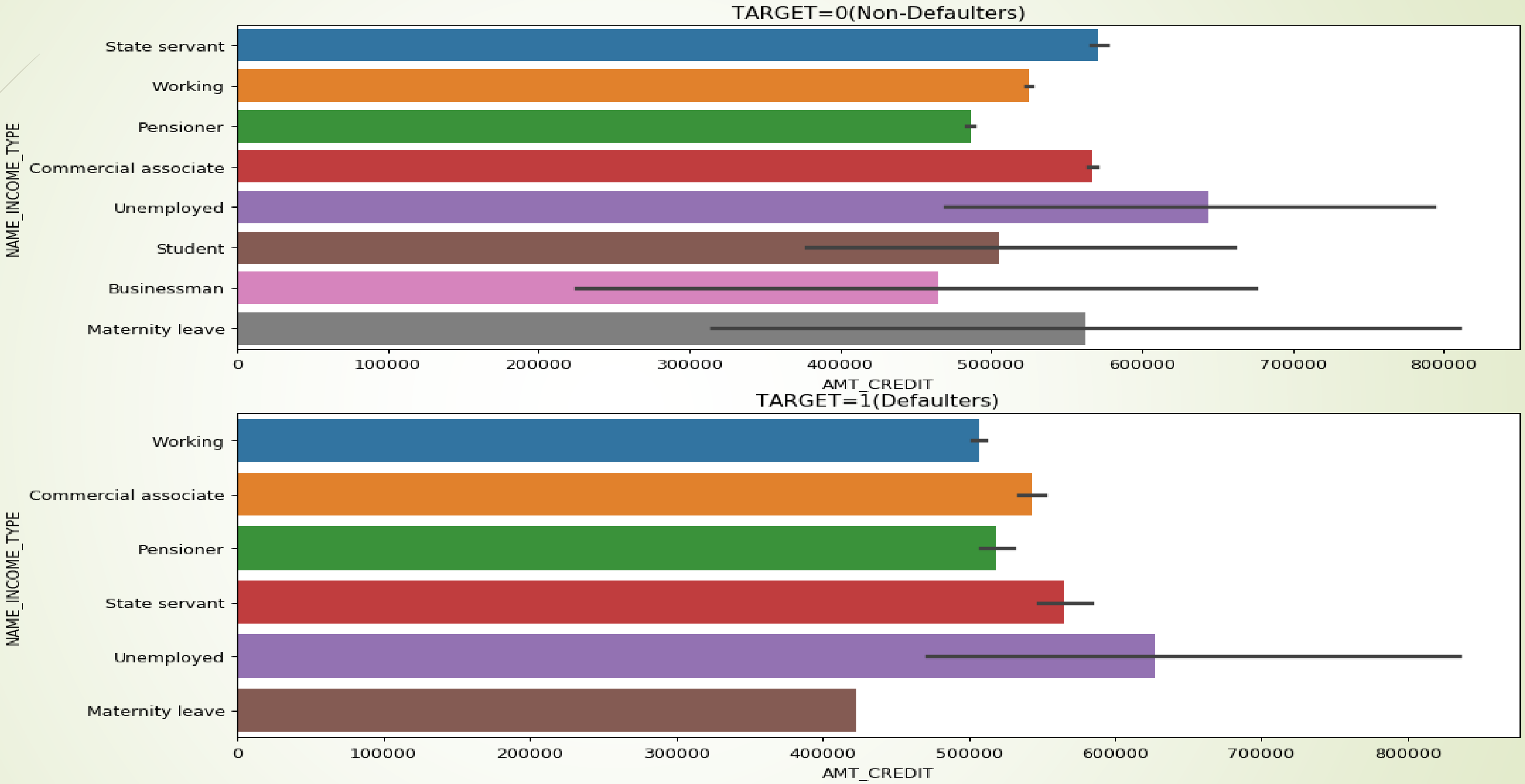
For defaulters and non defaulters some more analysis:

- Graph 8(based on contract_type) : when we analyzed the data we found out that non defaulters count is more for the revolving loans.
- Graph 9(based on gender) : when we analyzed the data we found out that females tend to be defaulters as compared to males.
- Graph 10(based on family status) :on analysis we found out that married people turn out to be non defaulters.
- Graph 11(based on organization type) : on analysis we found out that those whose organization type is government turn out to be non defaulter.
- Graph 12(based on credit score) : we couldn’t compute anything based on credit score.
- Graph 13(count children): based on no of children those who have 0 children have high chances of being defaulter.
- Graph 14(amt annuity) : those with annuity between 0-10000 have high chance of being non defaulter.
- Graph 15(amt good price) : for those whose amt good price is in between 60000 and 80000 have high probability of being defaulter.

Variables correlation that are highly correlated in both defaulters and non defaulters

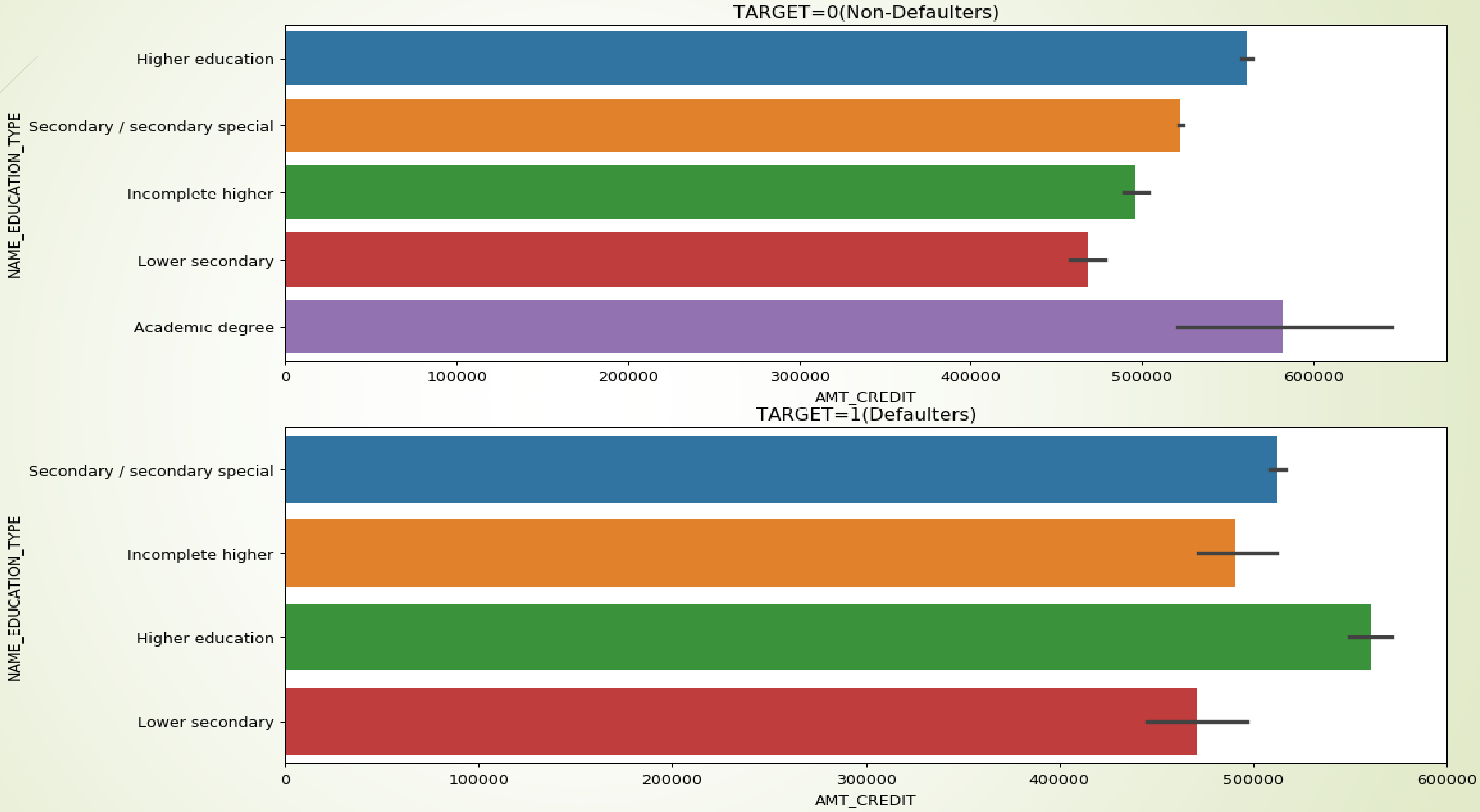
VARIABLE	DEFAULTER	NON-DEFAULTER
1.Amount credit v/s amt good price	0.98	0.98
2.cnt_family V/s count children	0.88	0.88
3.amt annuity to amt credit	0.74	0.76
4.amt good price to amt annuity	0.74	0.76
5.amt credit v/s amt income total	0.3	0.32
6.amt annuity v/s amt income total	0.38	0.4
7. amt good price v/s amt income total	0.3	0.33
8.days birth v/s days registration	0.29	0.34
9.count children v/s days birth	0.26	0.34
10. days birth v/s days id publish	0.26	0.29

Bivariate analysis(name_income_type v/s amt_credit):



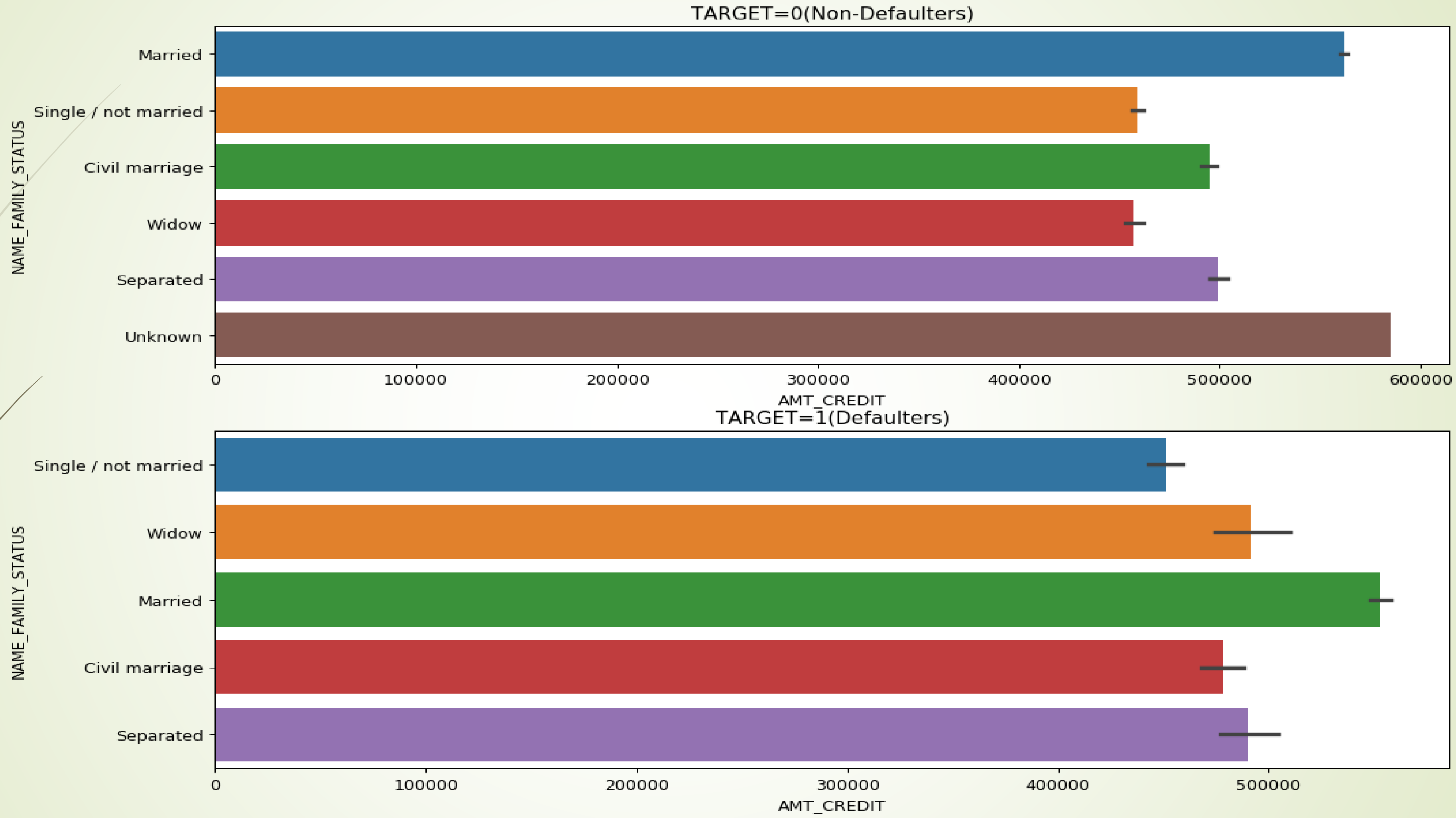
- On analysis we can conclude that Pensioner high credit likely to default.

Graph 2(name_education_type v/s amt_credit):



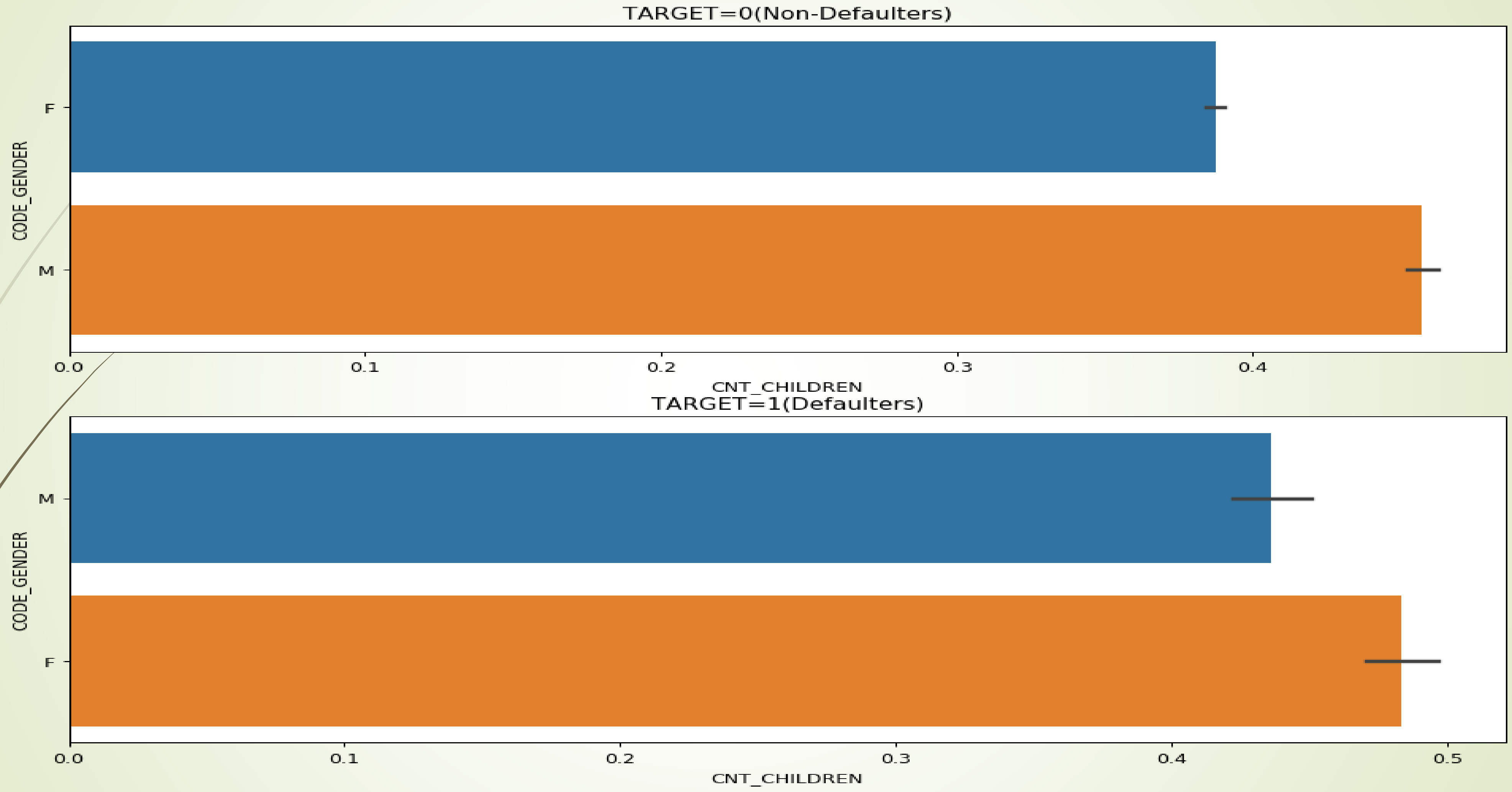
- On analysis we can conclude that academic degree having less to default

Graph 3(name_family_status v/s amt_credit):



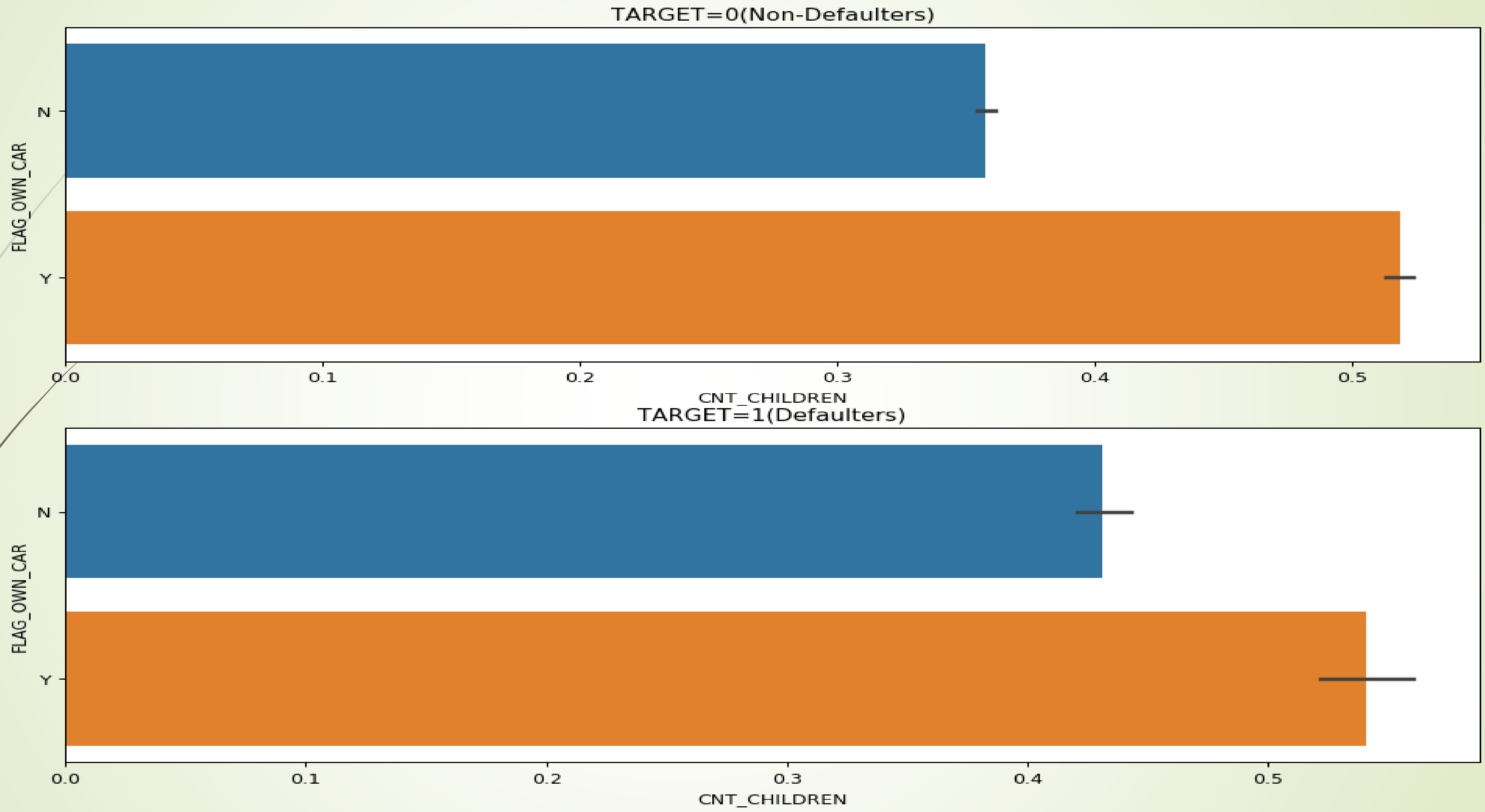
- On analysis we can conclude that widow with high credit likely to default

Graph 4(code_gender v/s cnt_children):



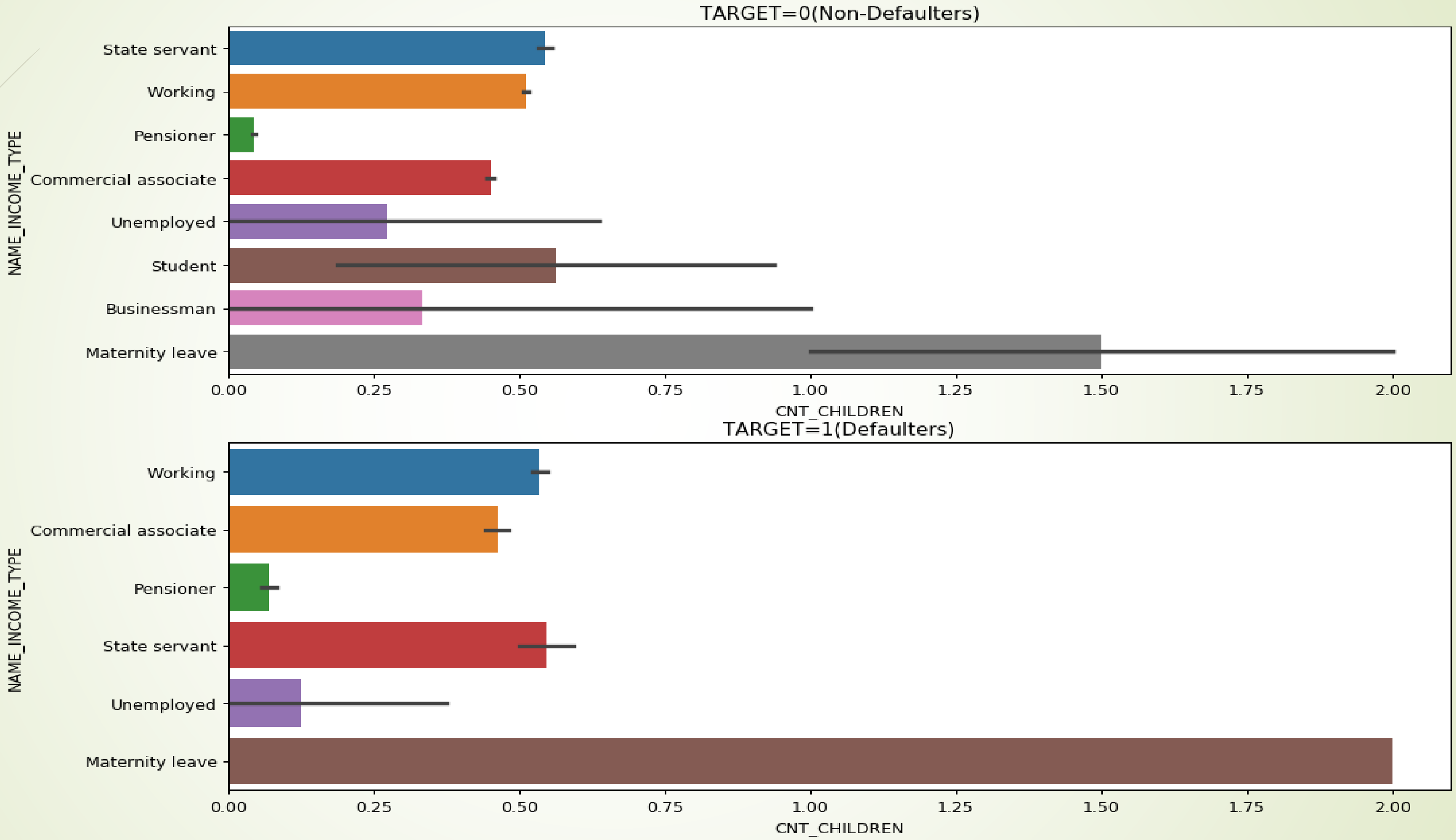
- On analysis we can conclude females having high children likely to default

Graph 5(flag_own_car v/s cnt_children):



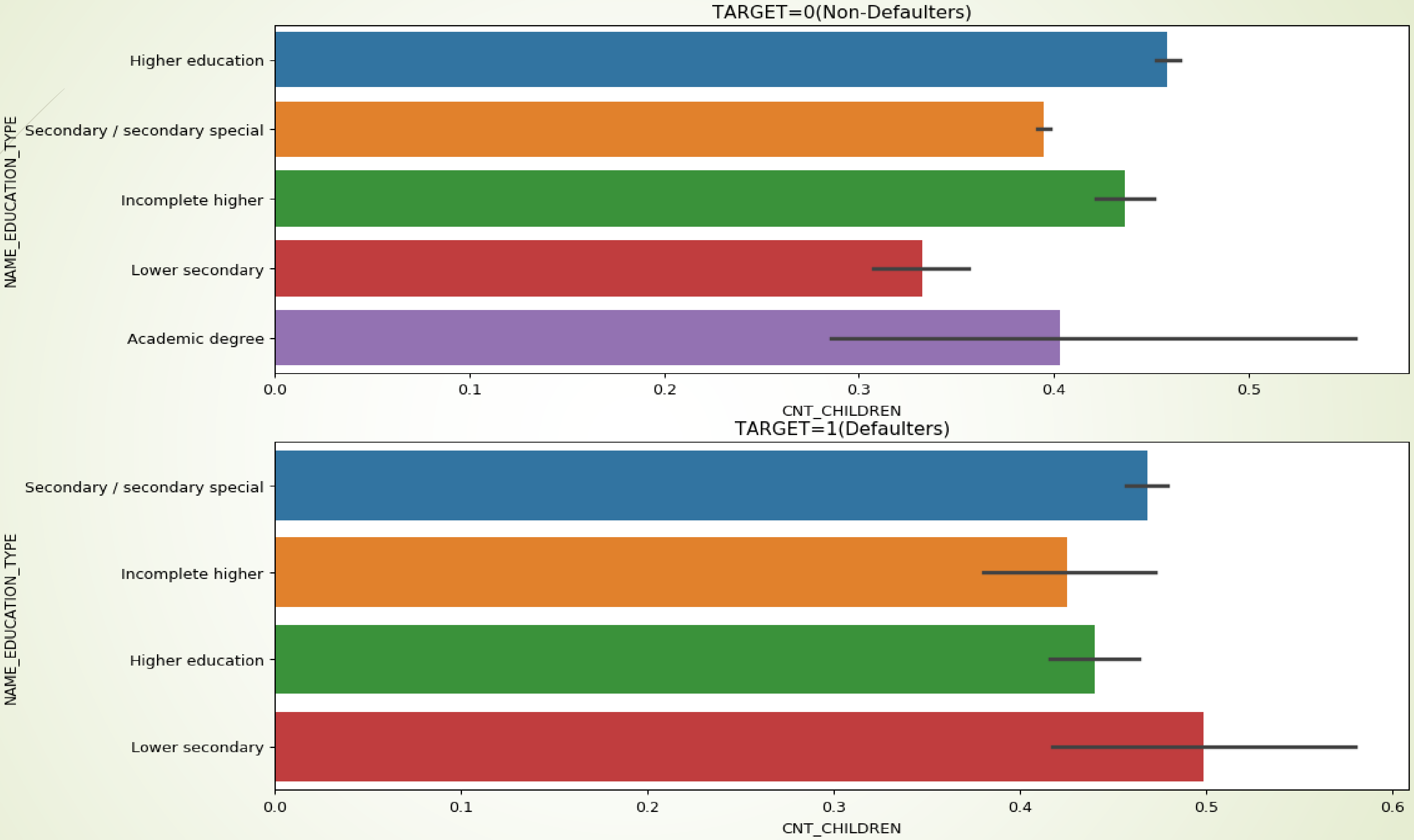
- On analysis we can conclude that if person doesn't owns a car and has more children then likely to default

Graph 6(name_income_type v/s cnt_children):



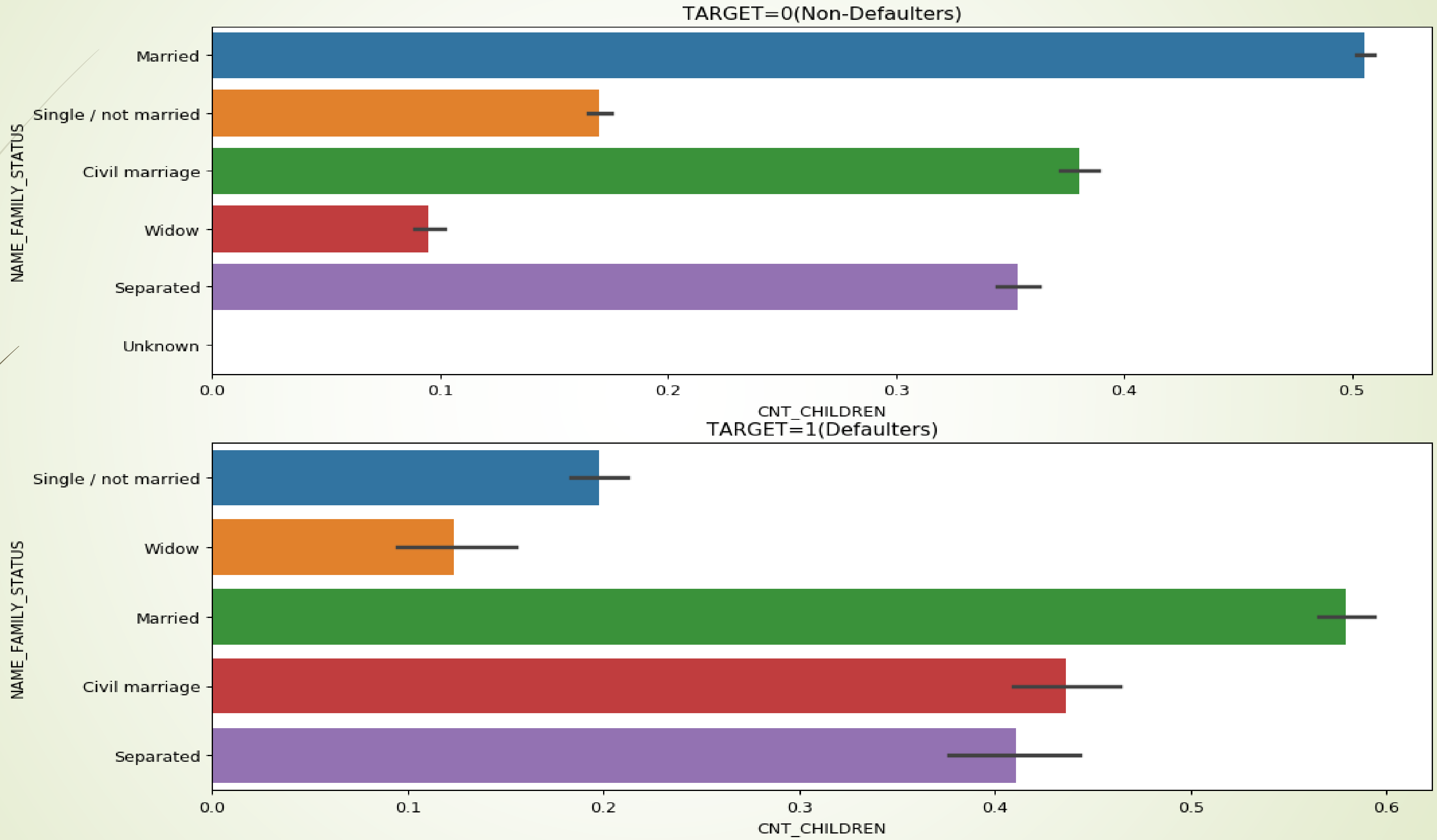
- On analysis we can conclude that maternity leave and having more children likely to default

Graph 7(name_education_type v/s cnt_children):



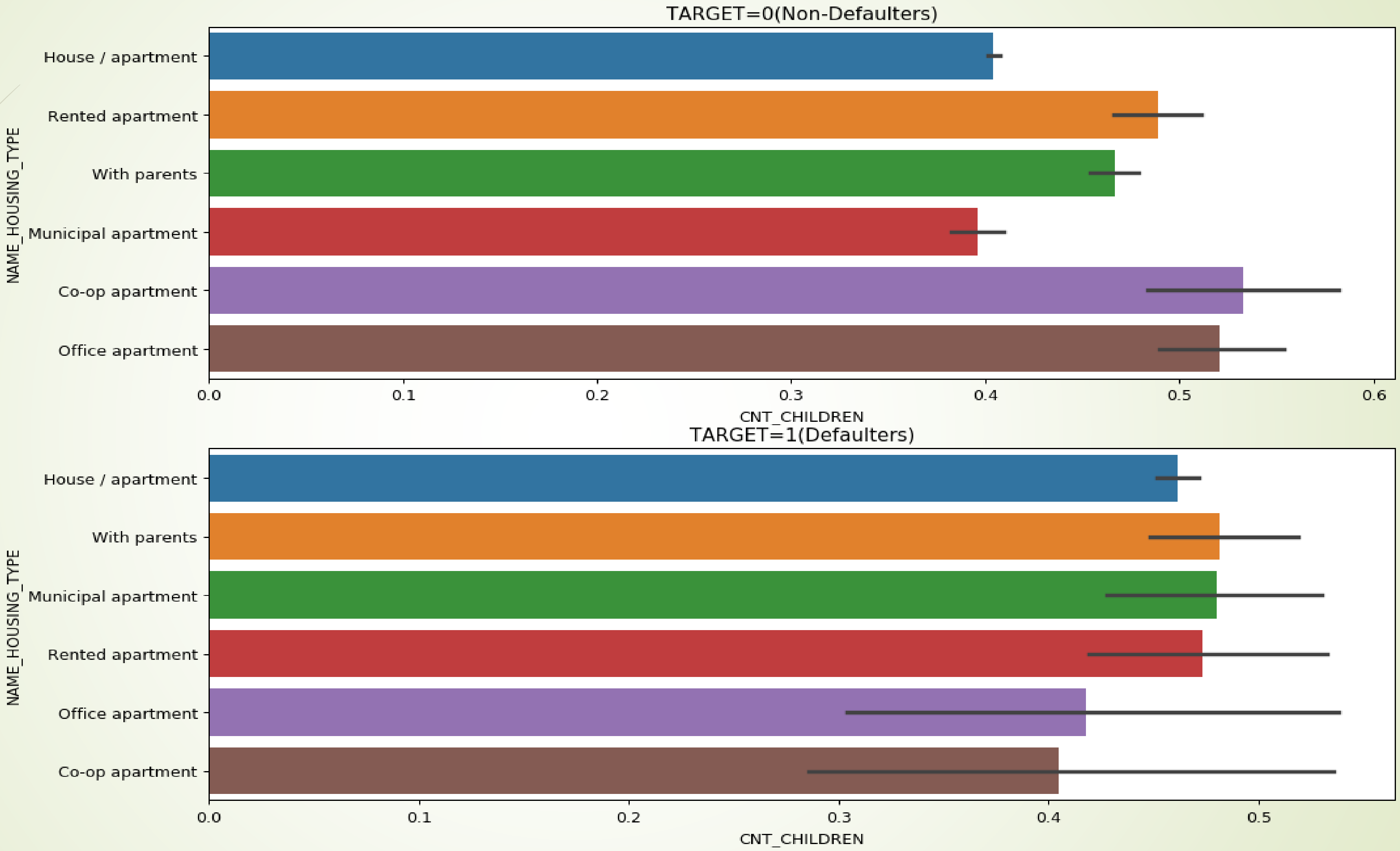
- On analysis we can conclude that lower secondary education and more number of children likely to default

Graph 8(name_family_status v/s cnt_children):



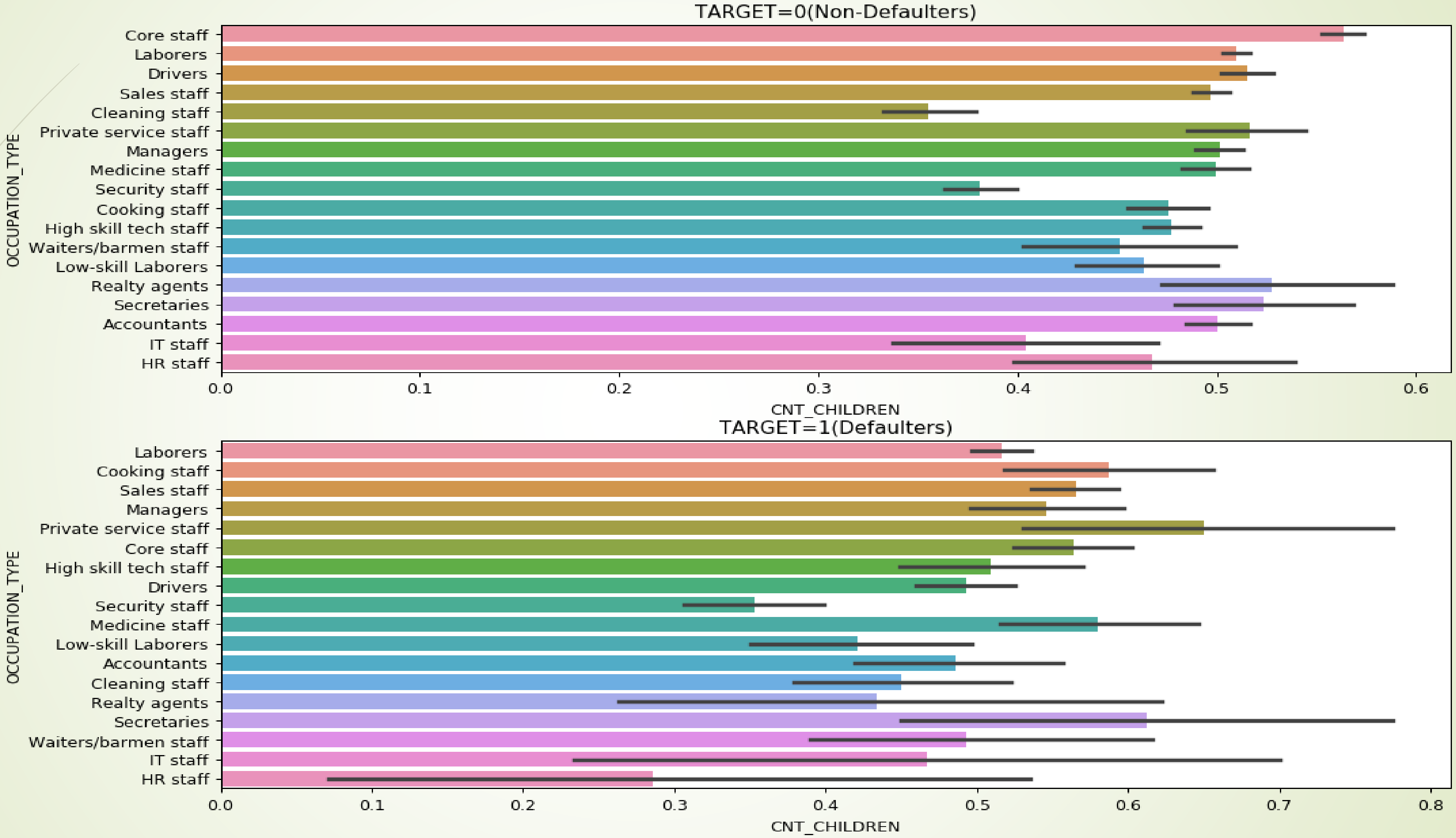
- On analysis we can conclude that married with more number of children likely to default

Graph 9 (name_family_status v/s amt_credit):



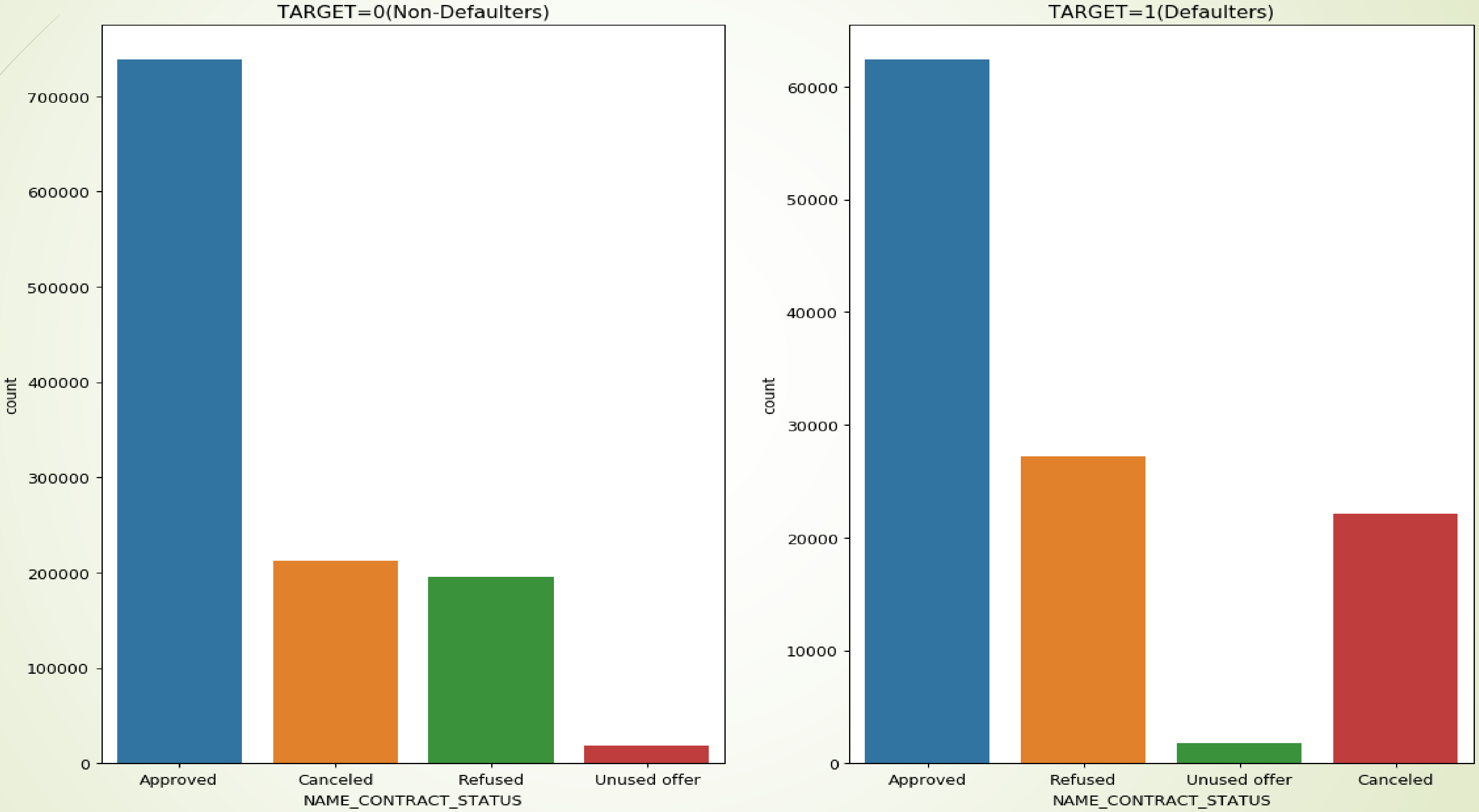
- On analysis we can conclude that people with municipal apartments having more children are likely to default and house/apartment having more children are likely to default.

Graph 10(cnt_children v/s occupation_type):



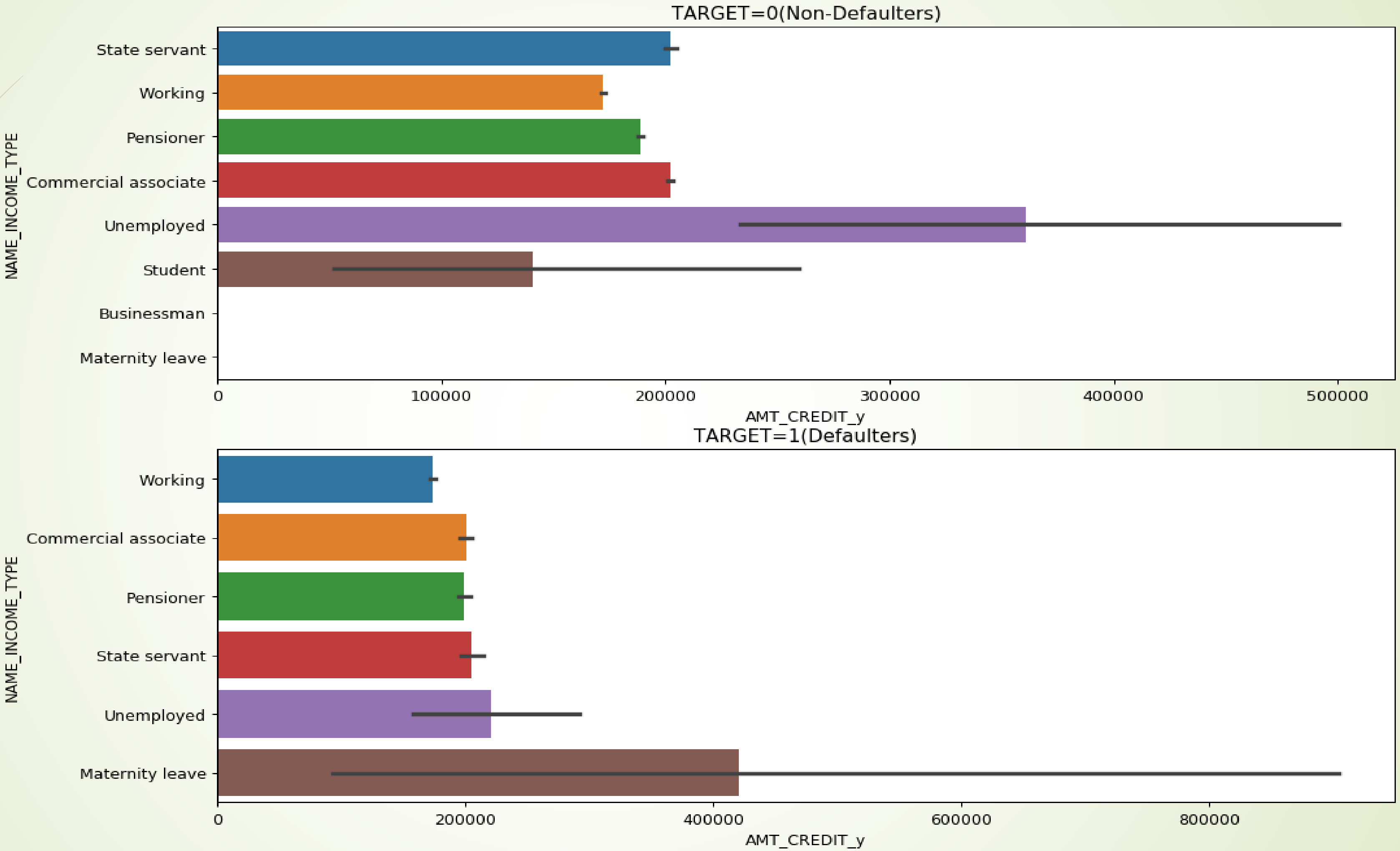
- On analysis we can conclude that private service staff , cooking staff , cleaning staff and secretaries with more children likely to default

Graph 1:name_contract_status(after merging the dataset previous application and *application_data*):



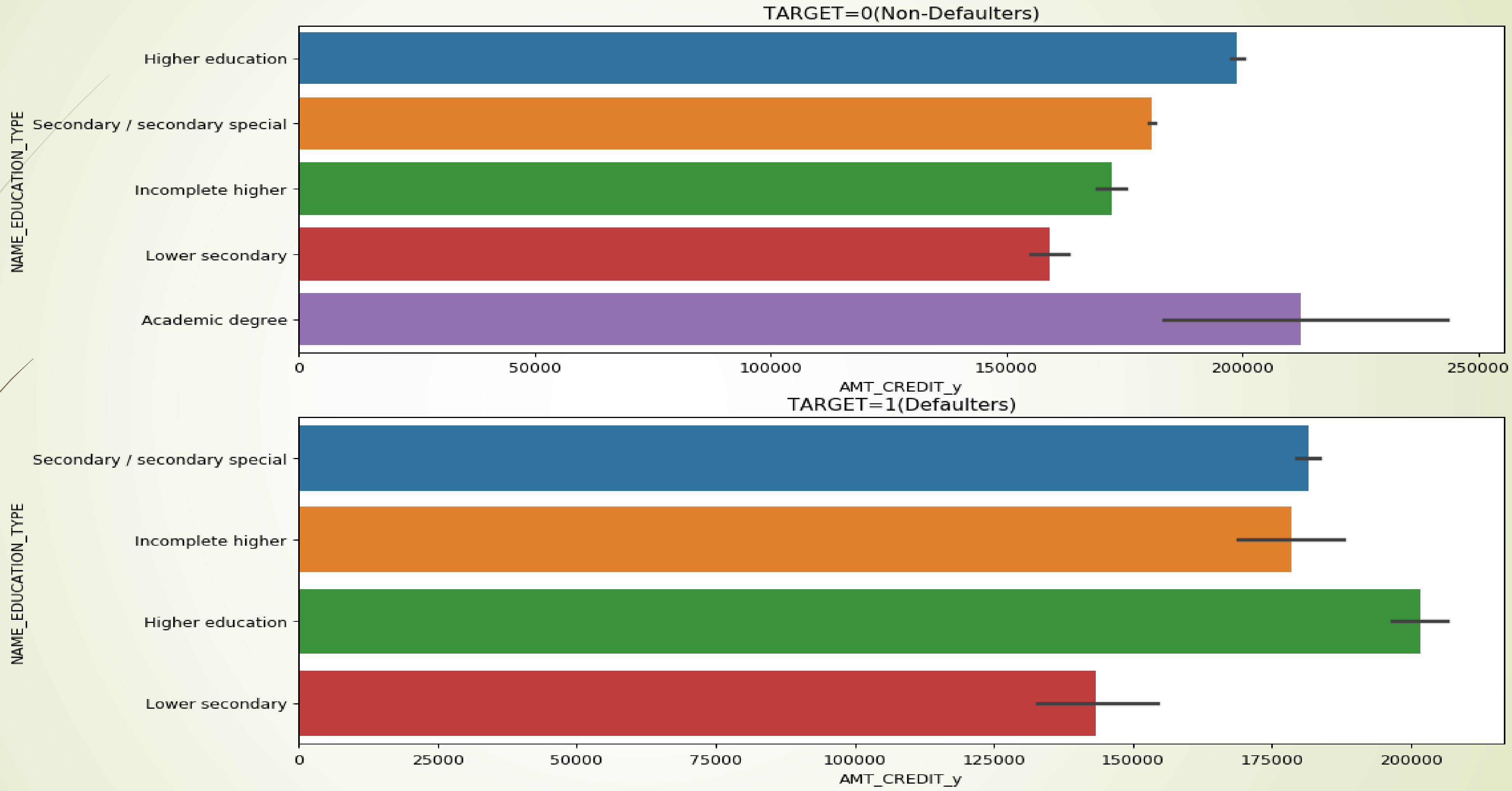
- On analysis we can conclude the defaulters are moslty those whose previous application was refused.

Graph 2:(name_income_type v/s amt_credit):



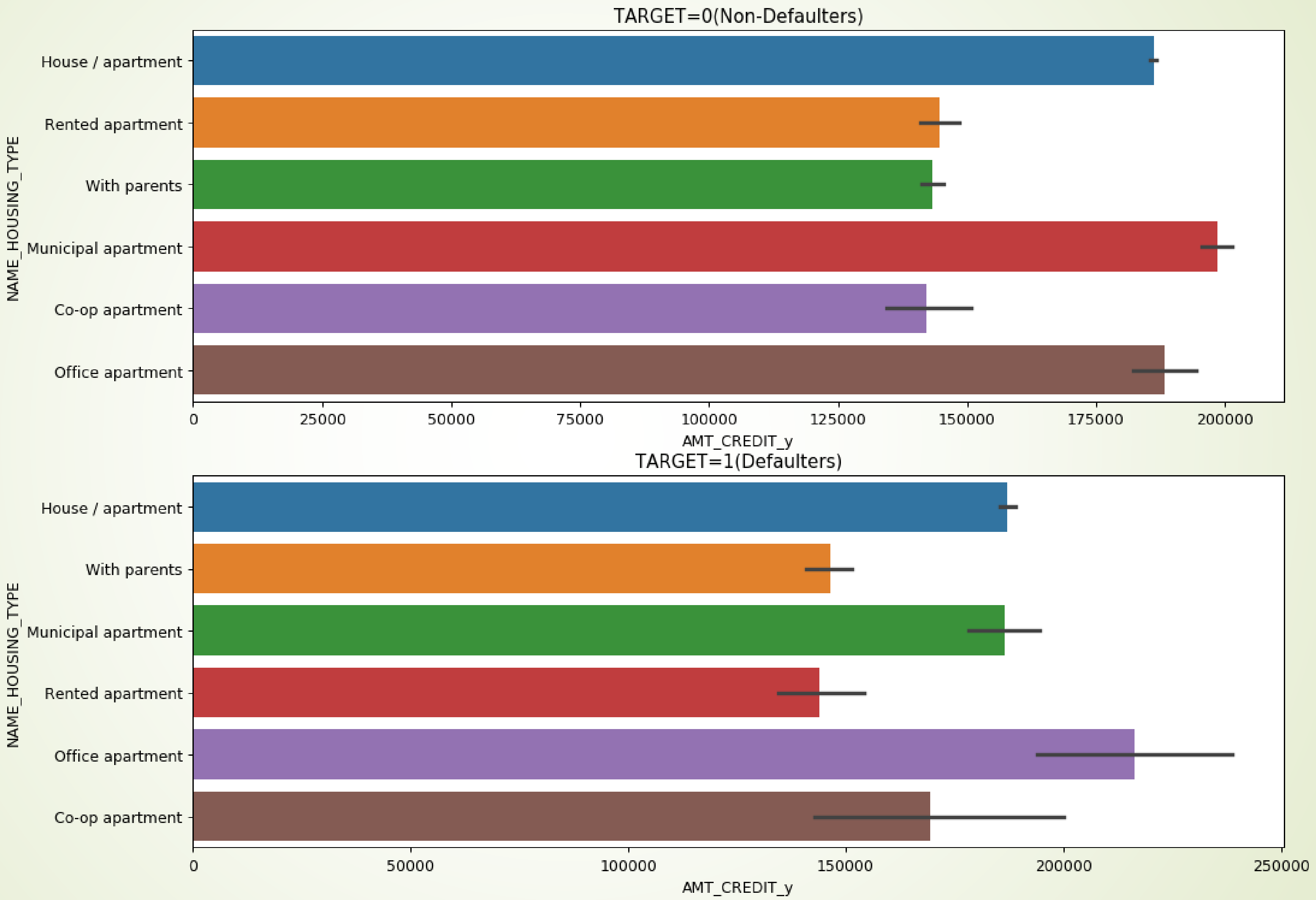
- On analysis we can conclude that if name income type is maternity leave and applied for large amount of credit in previous application then he likely to go default.

Graph 3 (name_education_type n v/s amt_credit):



- On analysis we can conclude that if the name education type is academic degree and in the previous application if the credit is high he likely to non –default if the name education type is Higher degree and in the previous application if the credit is high he likely to Default

Graph 4(name_housing_type v/s amt_credit):



- On analysis we can conclude that if the name housing type is office apartment and previously applied for high credit amount more than 2lakhs he likely to go default

THANKYOU