EDA Case Study

Ravi Kiran Edla
&
Vaibhav Sharma

Case Study Info

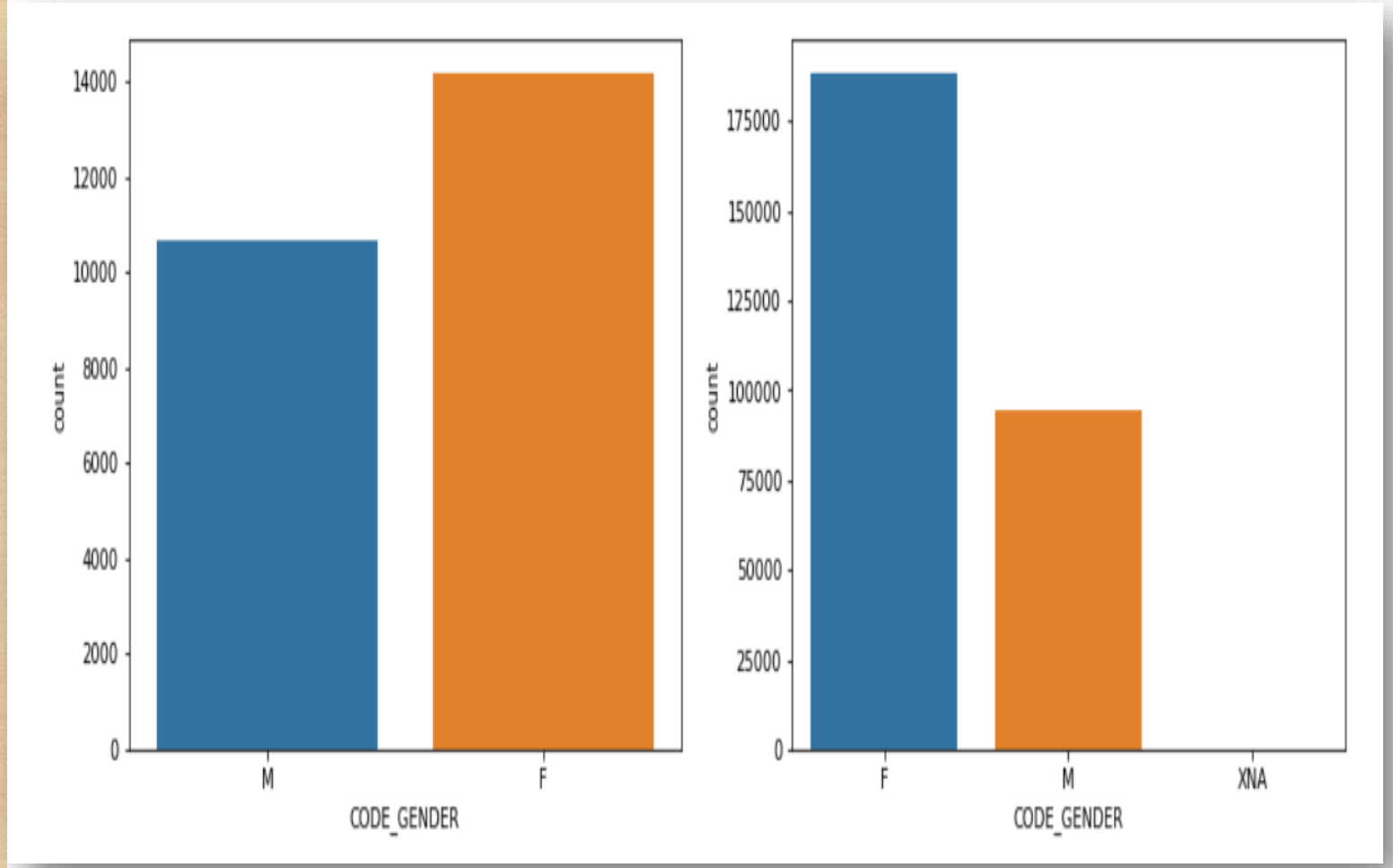
Applying Exploratory Analysis techniques for a business scenario.

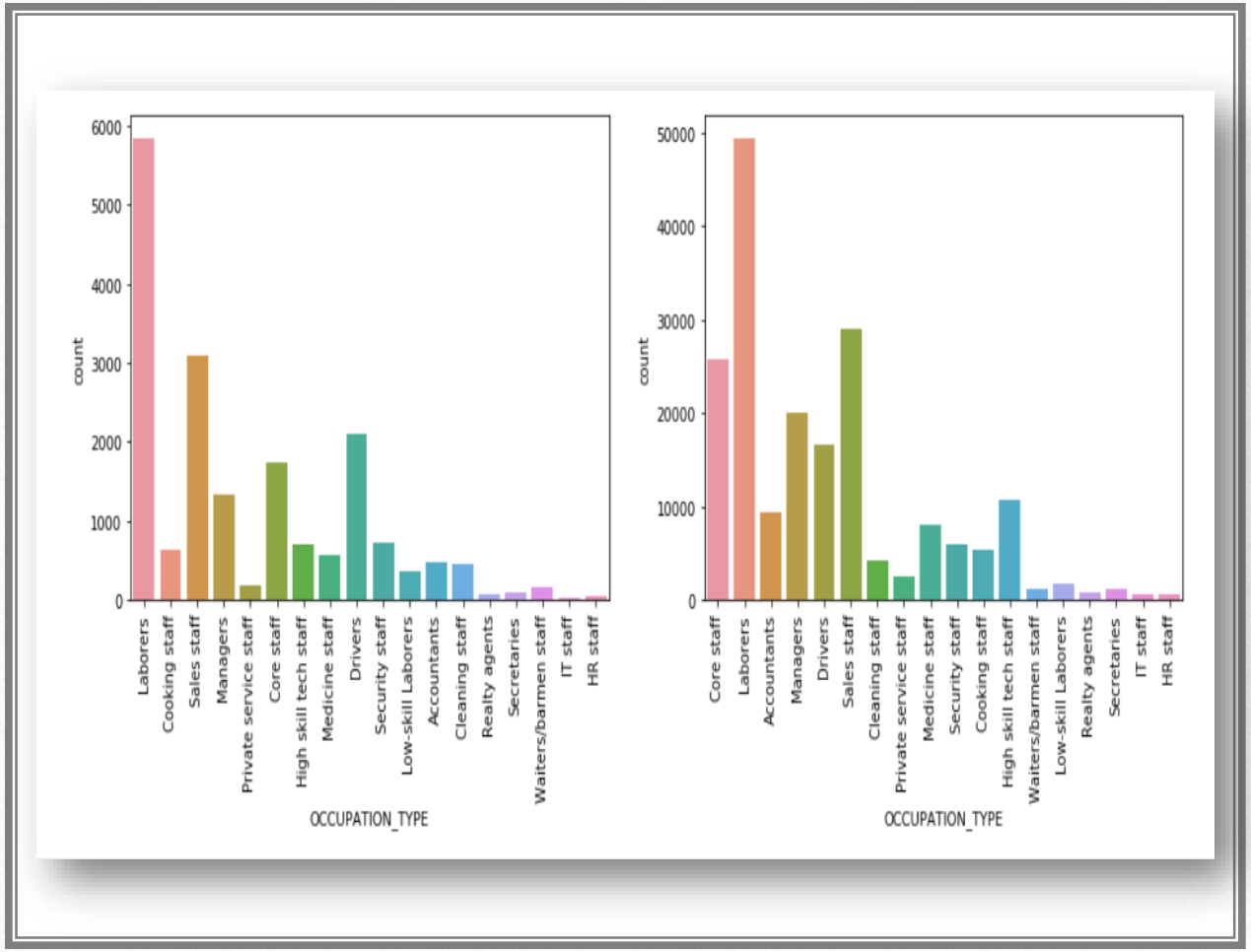
Deriving insights of risk analysis in banking and financial services from the given application data, by analyzing and plotting the data in graphs prepared in the Jupyter notebook.

Understanding how data is used to minimize the risk of losing money while lending to customers.

Male Vs Female of Defaulter Vs Non-Defaulter

- Graph displaying the data analysis of defaulters.
- This shows that Males have a higher chance of not returning their loans compared with Females.

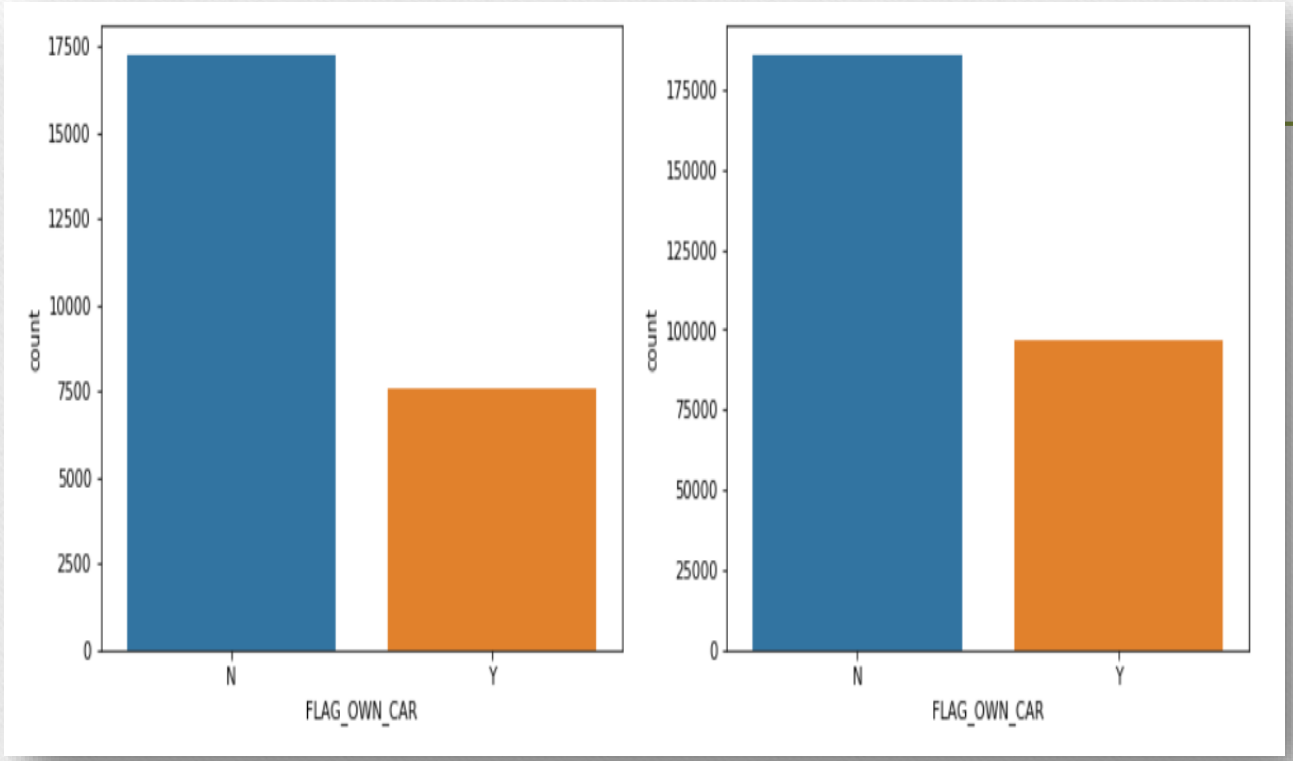




Occupation types Defaulter/Non-Defaulter.

- Graph displaying the data analysis of various defaulters.
- This shows that Laborers have a higher chance of not returning their loans. Whereas, HR Staff and IT Staff have very minimal chance of defaulting their loans. Hence, it is advised to prioritize the sanctioning of loans to HR Staff and IT Staff applicants.

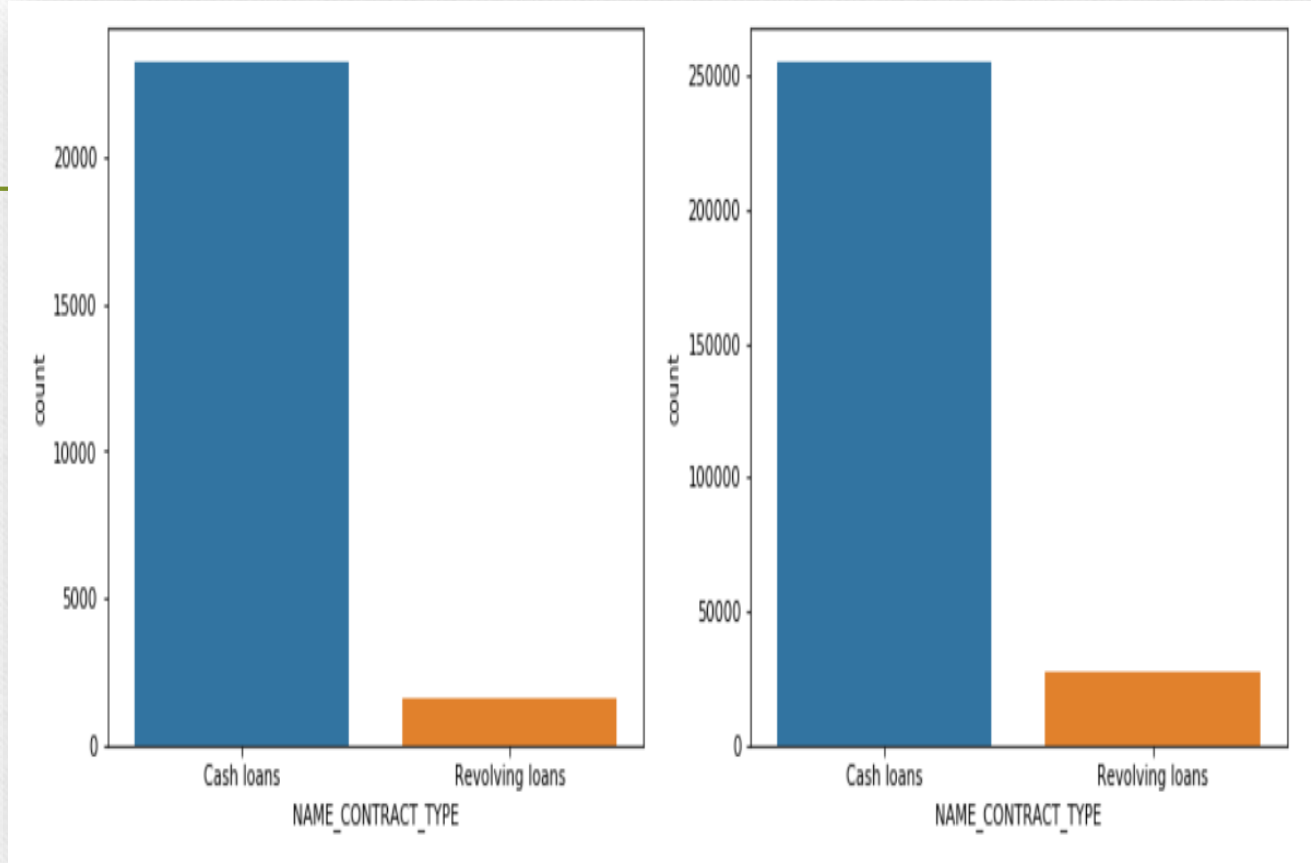
Owned car applicants Defaulter/Non-Defaulter

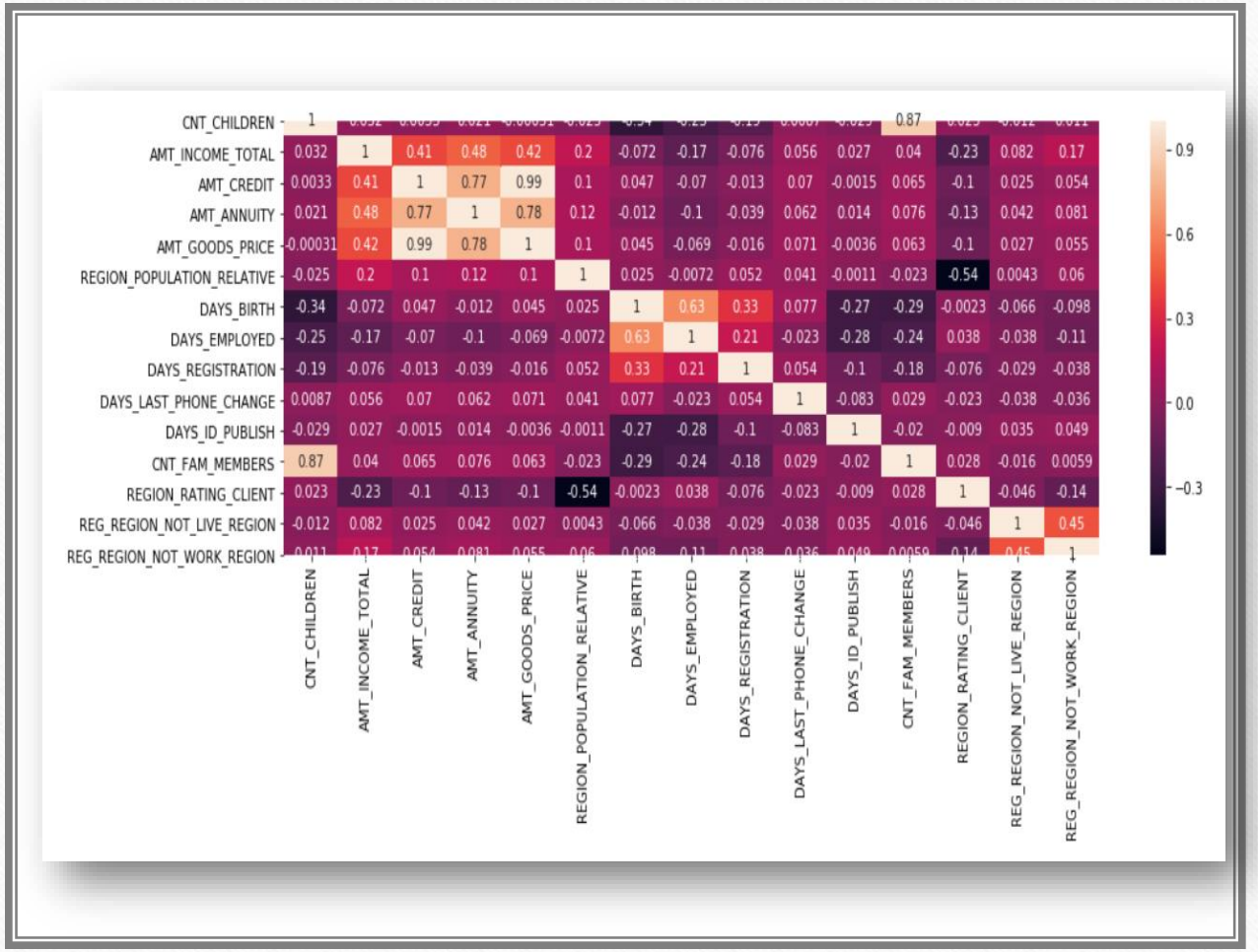


- Graph displaying the data analysis of various defaulters.
- The clients that owns a car are almost a half of the ones that doesn't own one.
  The clients that owns a car are less likely to not repay a car that the ones that own.
  Both categories have not-repayment rates.
  Hence, we can consider for loans towards car loan applicants.

Cash loans Vs Revolving loans in Defaulter/Non-Defaulter

- Graphs displaying the data analysis of various defaulters against Cash loans Vs Revolving loans.
- From the graphs, we can identify that the defaulters in Revolving loans are minimal when compared to defaulters to Cash loans.
- Thus, it is prescribed to prefer only Revolving loans.



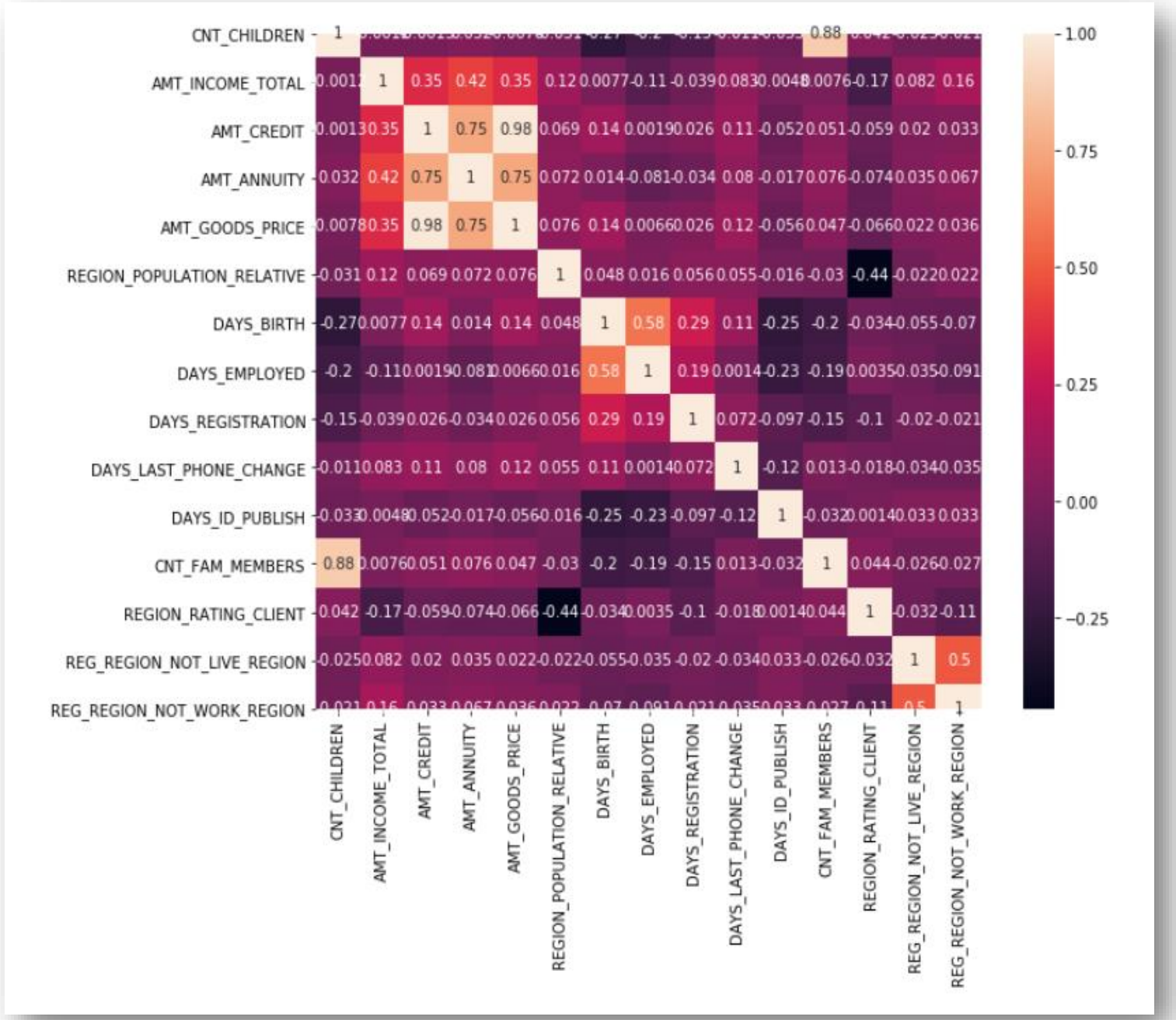


Correlations Analysis for Non-defaulters

- From the Correlation matrix, we notice that the AMT GOODS PRICE is highly correlated to AMT CREDIT, i.e., Credit amount of the loan against the consumer's loans it is the price of the goods for which the loan is given.
- REGION POPULATION RELATIVE is highly non-correlated to REGION RATING CLIENT, i.e., the amount of loan credit effects the Normalized population of region where client lives.

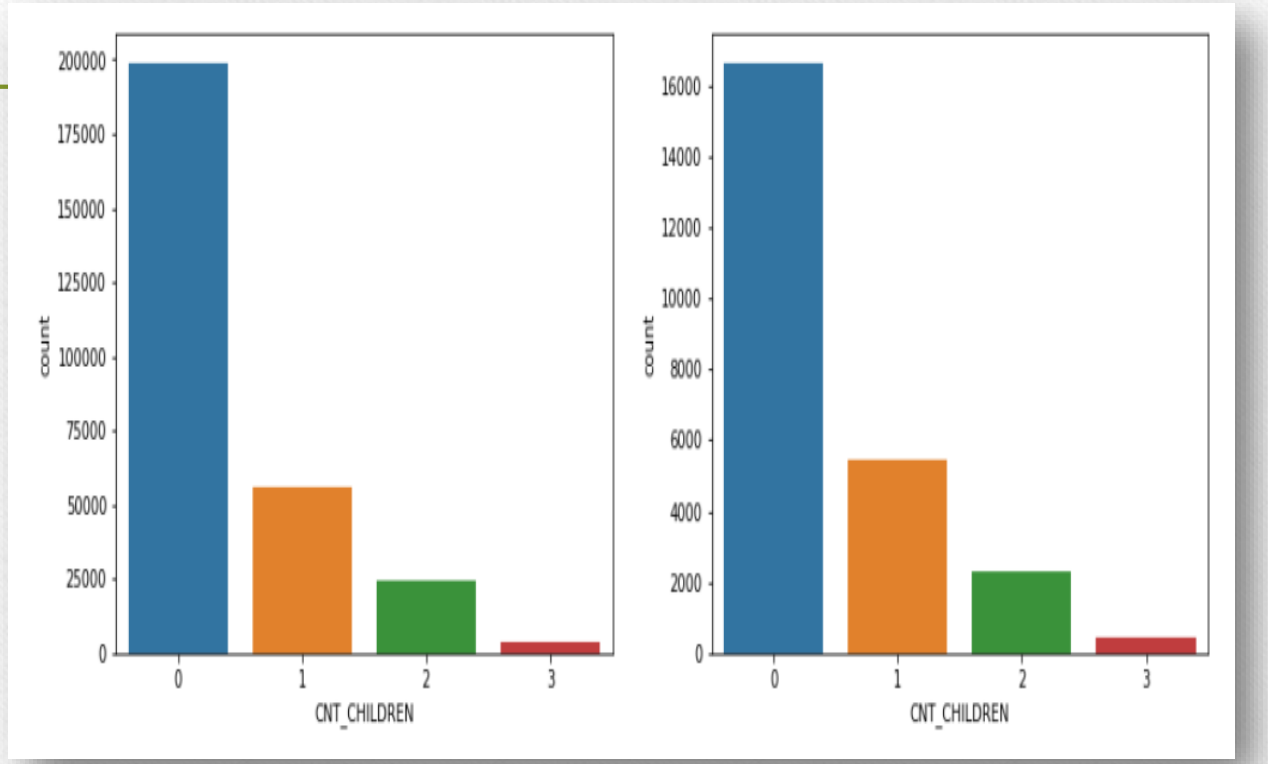
Correlations Analysis for Defaulters

- From the Correlation matrix, we notice that the we notice that the AMT GOODS PRICE is highly correlated to AMT CREDIT. i.e., Credit amount of the loan against the consumer's loans it is the price of the goods for which the loan is given.
- REGION POPULATION RELATIVE is highly non-correlated to REGION RATING CLIENT. i.e., the amount of loan credit effects the Normalized population of region where client lives.



Count of Children

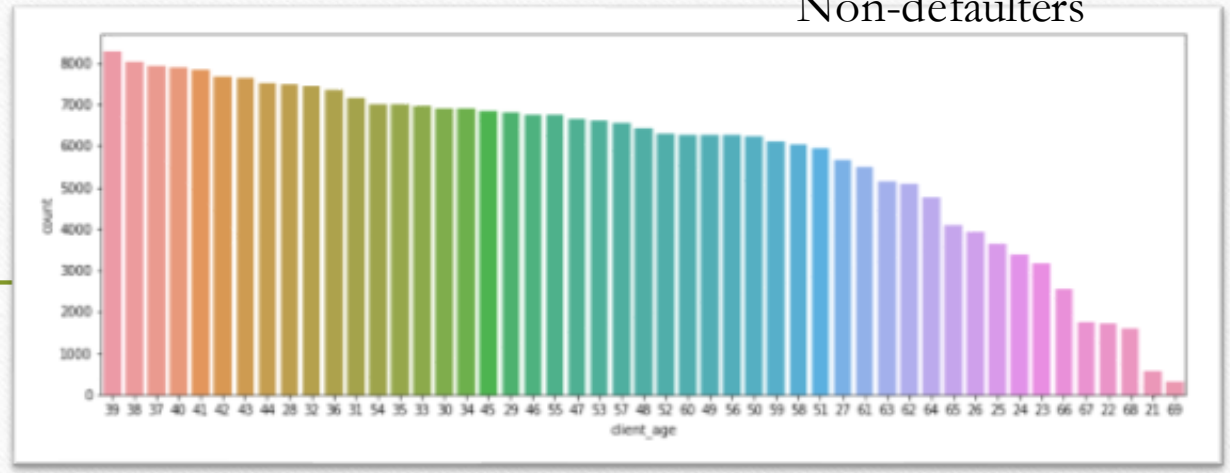
- From the graphs we can notice that most of the clients taking a loan have no children.
- The number of loans associated with the clients with one children are 4 times smaller than the number of loans associated with the clients with two children are 8 times smaller.
- Hence, it is always recommended to prefer loans to applicants with No-children.



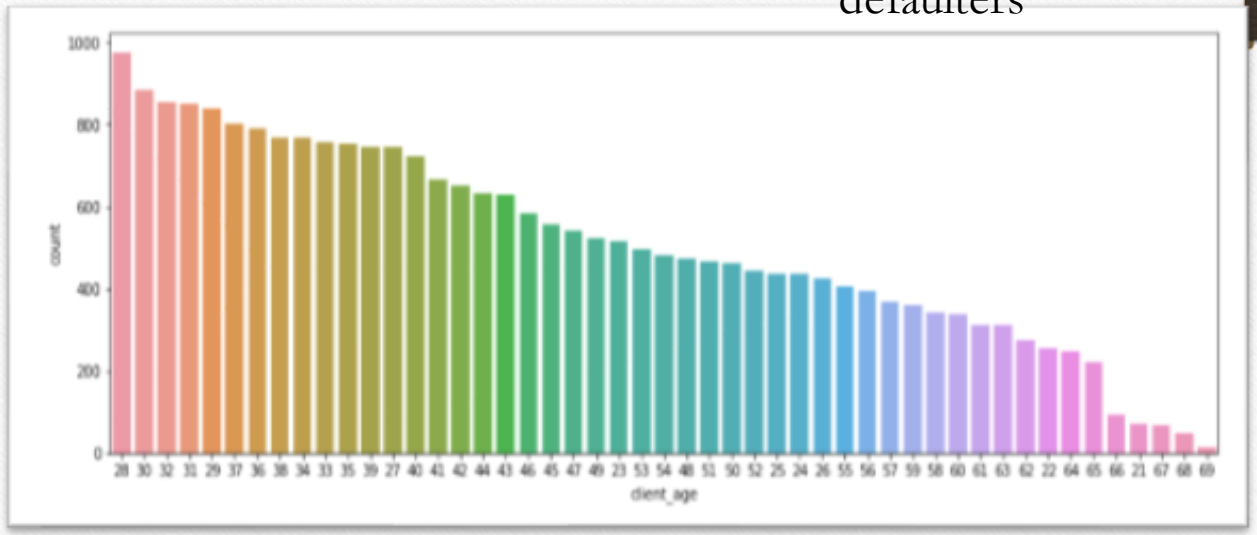
Client Age factor analysis between defaulters and non-defaulters.

- From graphs we notice that the defaulters are more in the age group is between 28 and 30. Whereas, the non-defaulters age group is more in between 37 and 40.
- Hence, it is less risky in giving loans to the applicants of age between 37 to 40.

Non-defaulters

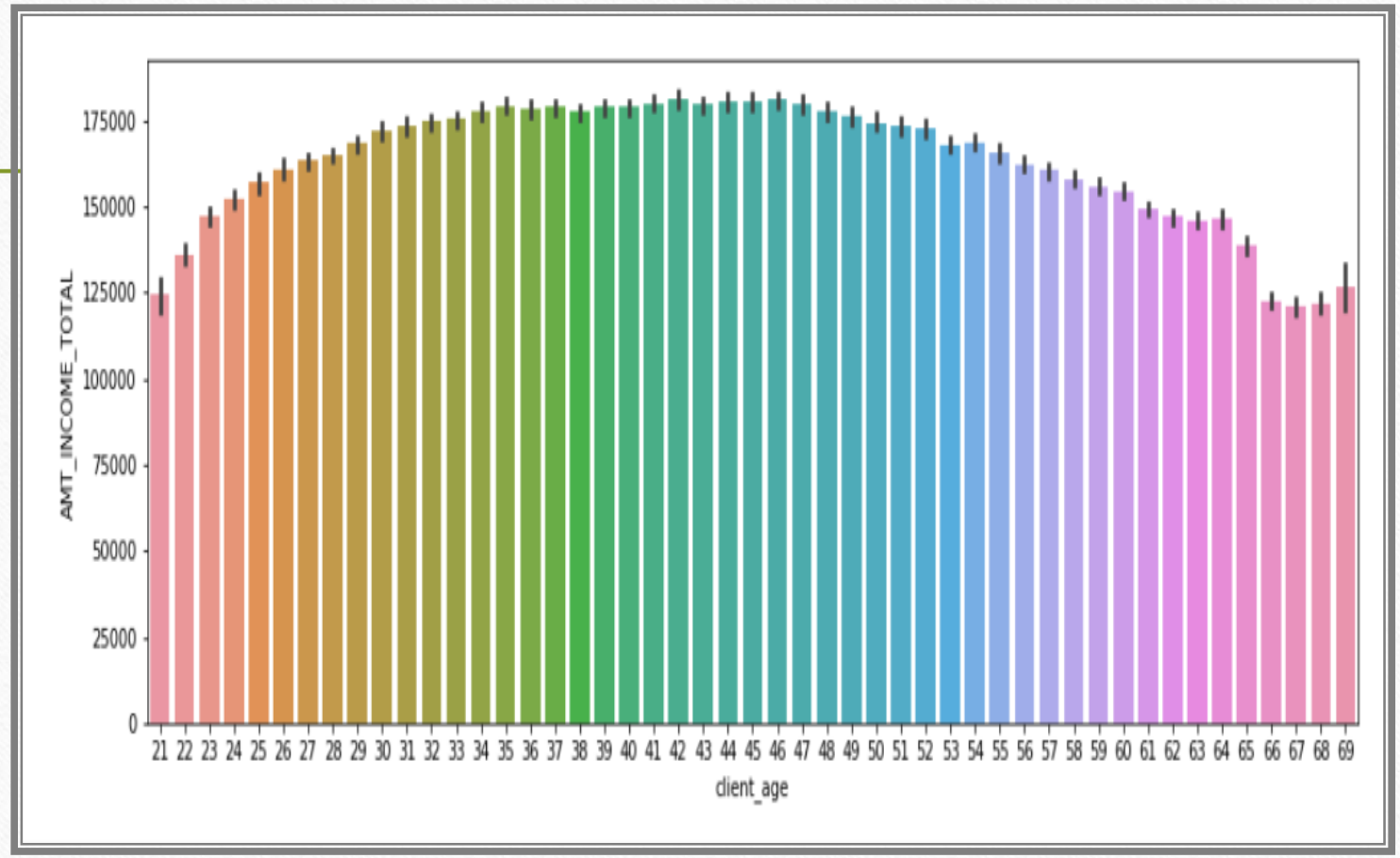


defaulters



Age of the applicant vs Amount of income

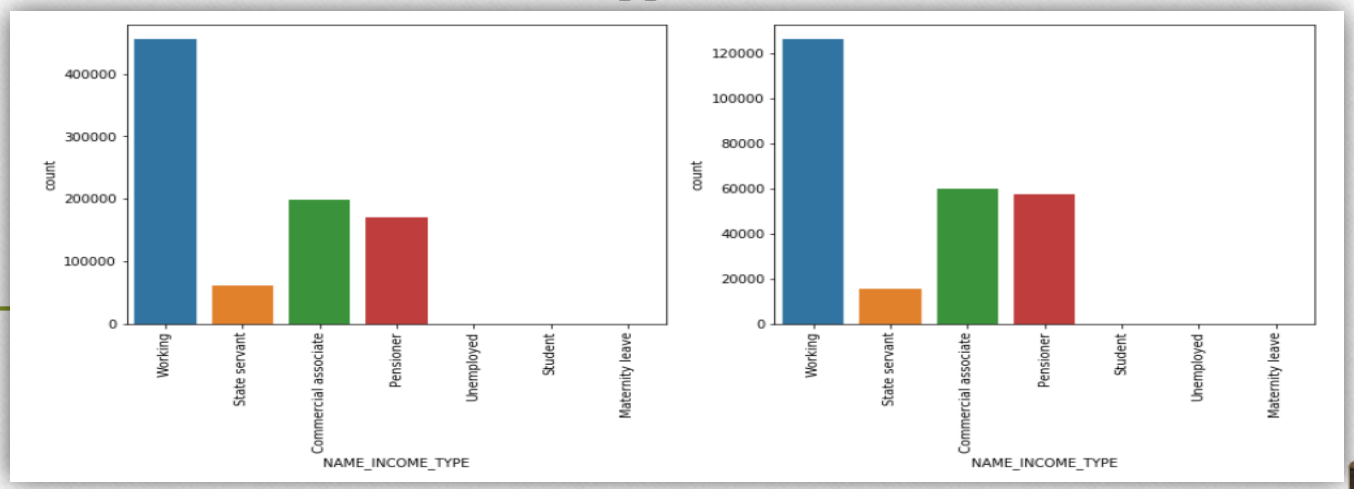
- From the graph, it is observed that the amount of income is high in the age group of 35 to 49.
- Hence, it is recommended to target the applicants of the age group of 35 to 49 for granting the loans.



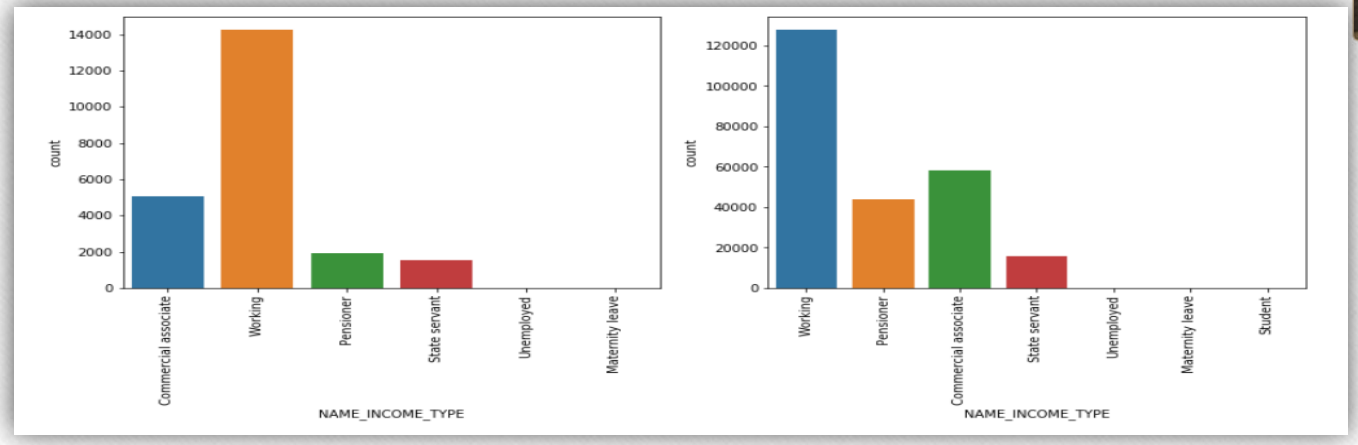
Loan status for various Income types

- From the graph plots, we understand that there is high tendency of approving loans to the applicants with high incomes whereas, the applicants with low income rates are getting cancelled or refused or else are unused too.
- Hence, we can target more towards to High income applicants.

Approved loan vs Cancelled loan



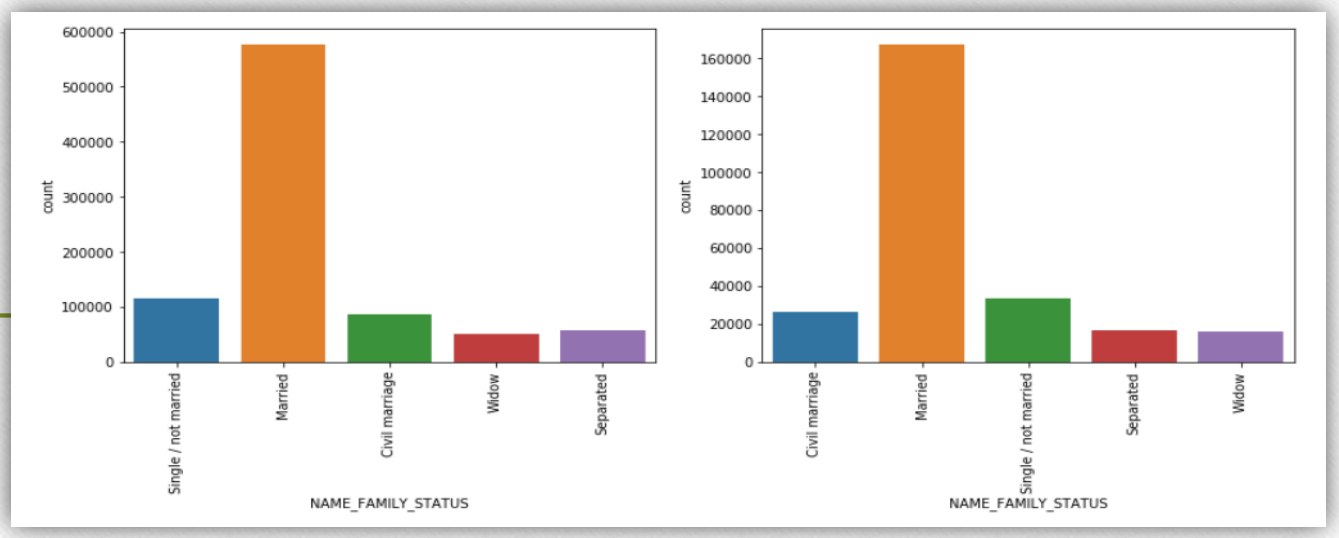
Unused Offer loan vs refused loan



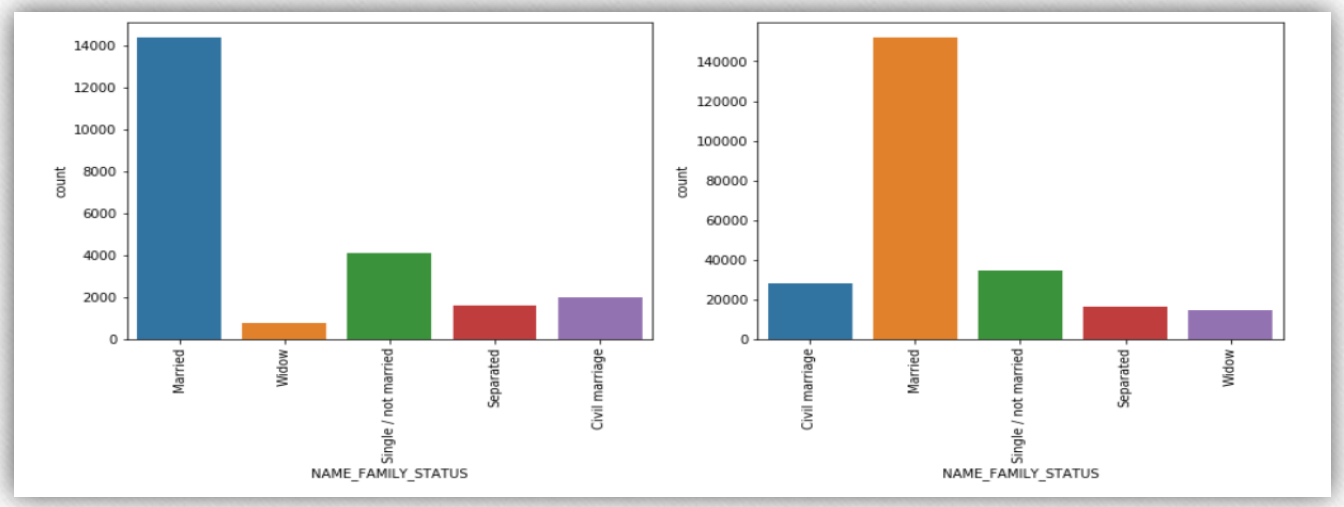
Loan status based on Family status

From the graph plots, we understand that there is always a flickering tendency for grant of loans against Married applicants who are the majority to get approval of loans and also they are the one's who have major tendencies towards loan cancellations or refusal or even not using the offers.

Approved loan vs Cancelled loan



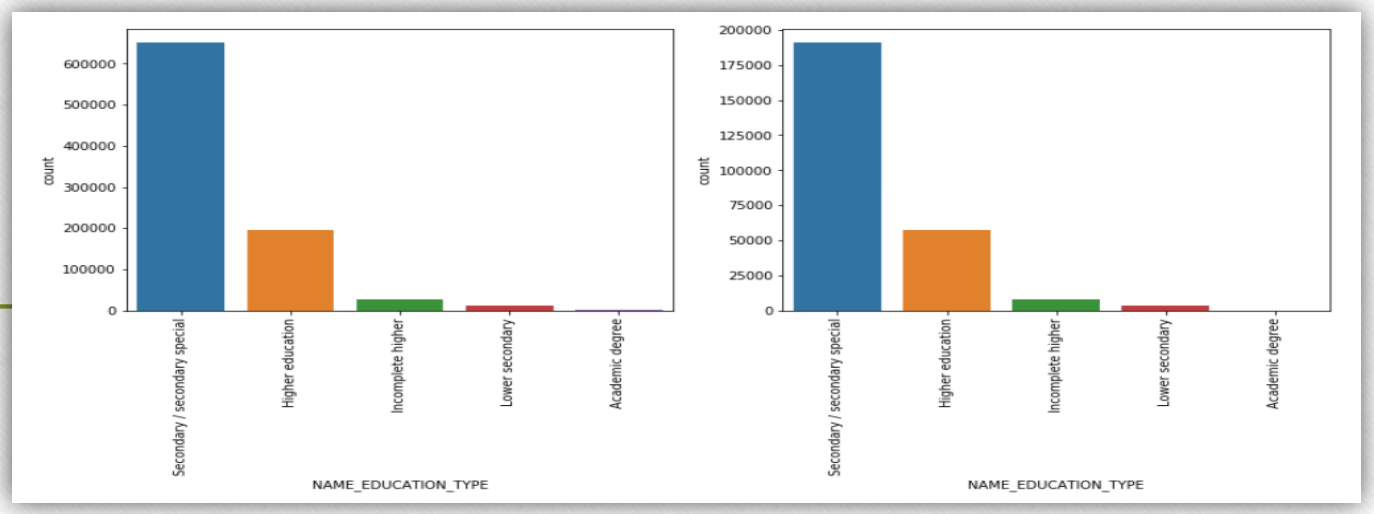
Unused Offer loan vs refused loan



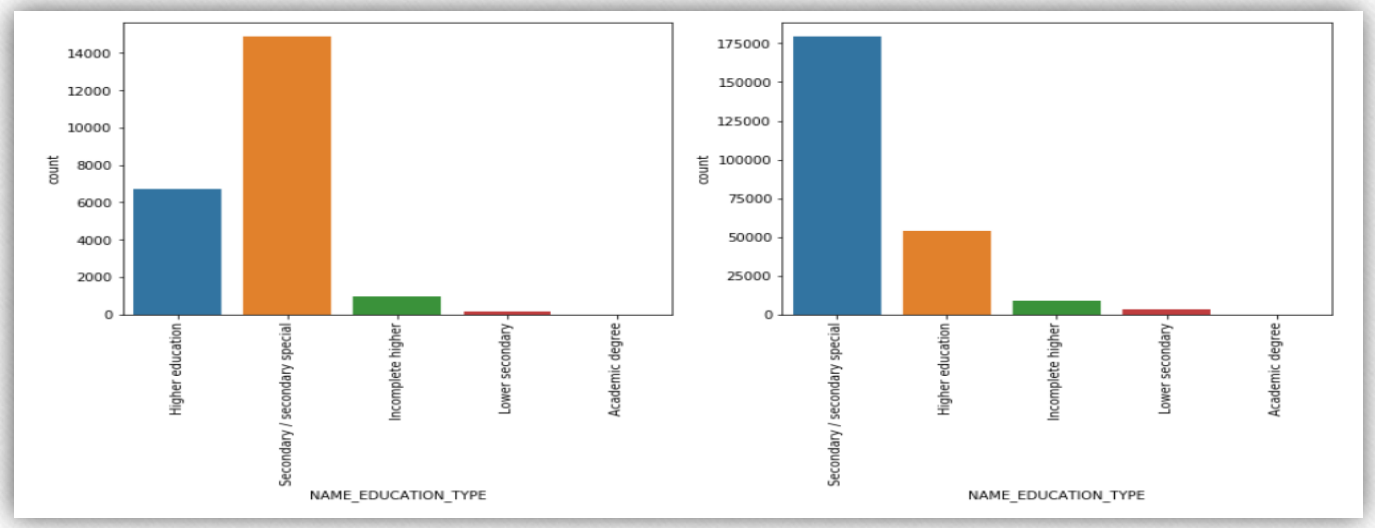
Loan status based on Education

From the graph plots, we understand that there is always a flickering tendency for grant of loans against applicants having education 'Secondary/Secondary special' who are the majority to get approval of loans and also they are the one's who have major tendencies towards loan cancellations or refusal or even not using the offers.

Approved loan vs Cancelled loan



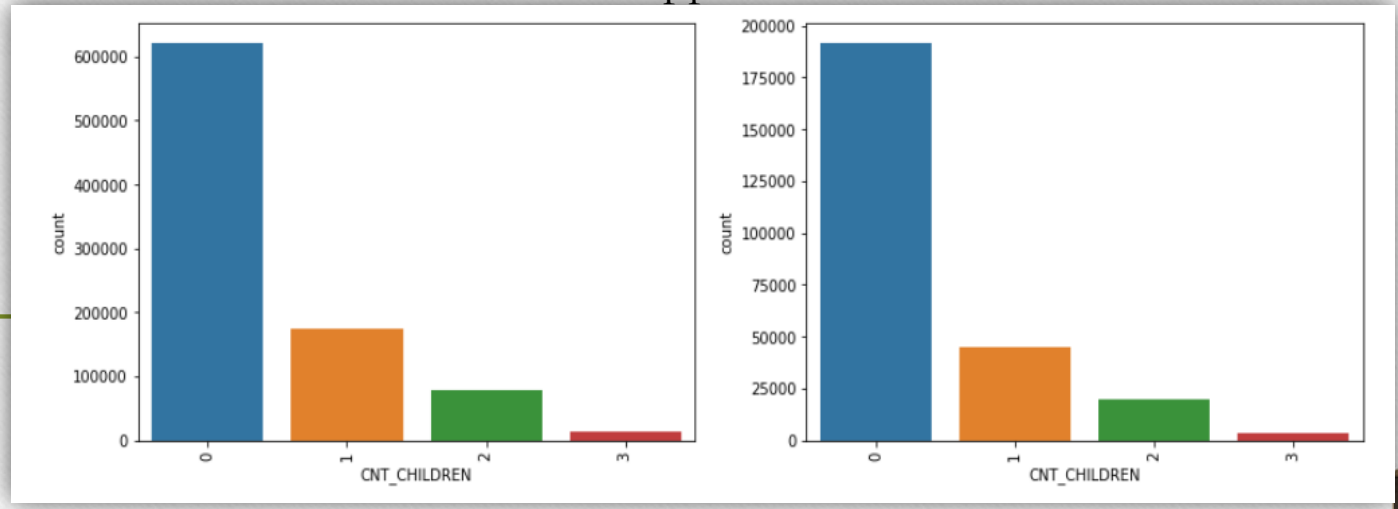
Unused Offer loan vs refused loan



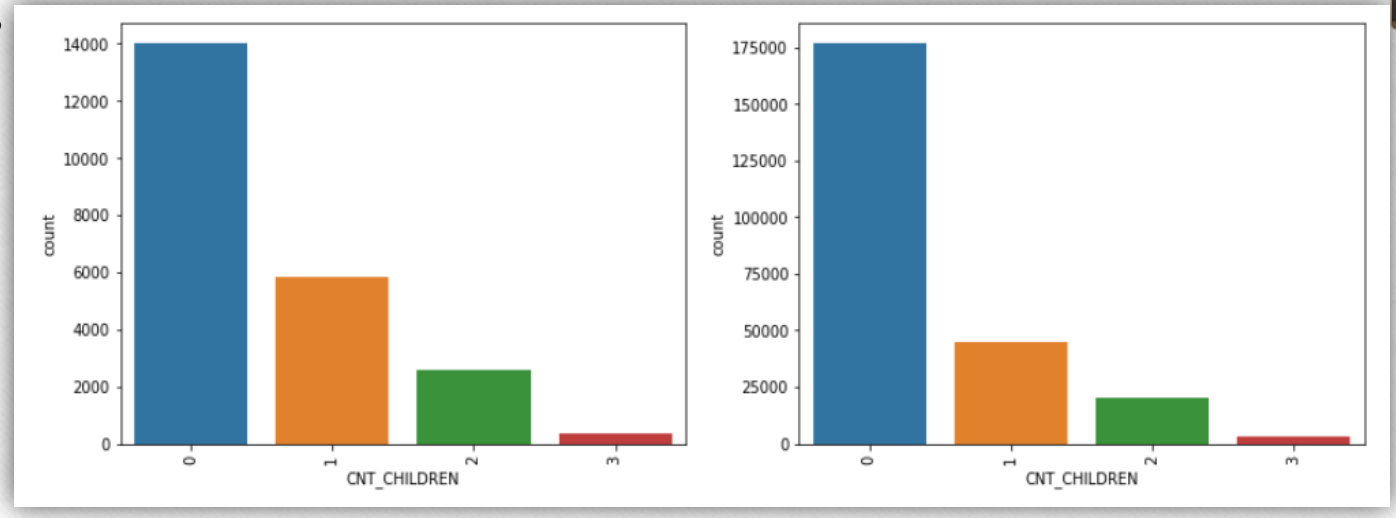
Loan status based on the number of Children

From the graph plots, we understand that there is high tendency of approving loans to the applicants with no children whereas, the applicants with more children are getting cancelled or refused or else are unused too.
Hence, we can target more towards applicants with no-children.

Approved loan vs Cancelled loan



Unused Offer loan vs refused loan



Reference links

- Quick commands help:
  - www.hackerrank.com
  - www.geeksforgeeks.org
  - https:/pandas.pydata.org/pandas-docs
  - https:/seaborn.pydata.org/
  - https:/stackoverflow.com/