



Group Assignment:

Lending Club : EDA Case Study

Group facilitator : Sumit Rana

Team Member : Bhargava Srivathsan.G

ML\_AI\_C41 Batch

06-07-2022

- To Find the Various factors which could help in showcase the driving factors of the Loans default.
- To understand various methods of analysis like Univariate and Bivariate.
- To Gather various methods of visualizing the data and provide business recommendations.

The Input file contains the loan data in the period of 2007-2011

Total Rows in Original dataset: 39717

Total Columns in Original dataset: 111

Cleaned data :

Total Columns in Cleaned dataset: 46

Total Rows in Original dataset: 38577

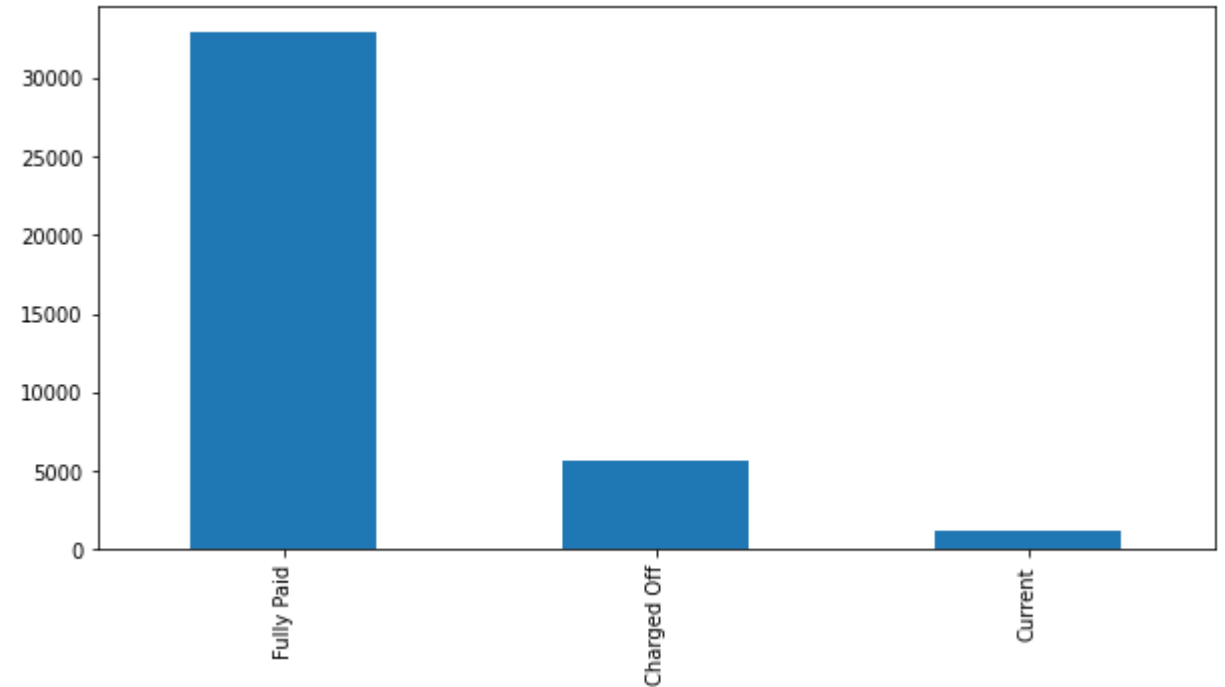
- There are columns in the dataset which contains all null values, so will have no impact on other variables or on the output of loan credit. So these columns can be removed.
- Some columns like 'policy code' have only one unique value. So they can be removed as well as constant value does not impact other variables.
- Loan status defines the current state of loan , so segmenting/filtering datasets on the basis of loan status will help us compare defaulters from non-defaulters.
- Cleaning removing columns which are unique identifiers for a row.
- These unique identifiers are used for giving identification to each dataset and does not impact or correlate with any variable.

The Analysis of the Univariate :

- Fully Paid 32950
- Charged Off 5627
- Current 1140

Recommendations:

We could see “fully paid” are 82 % of data following which “Charged off” are 14 % and “Currently” are 3 % of data.

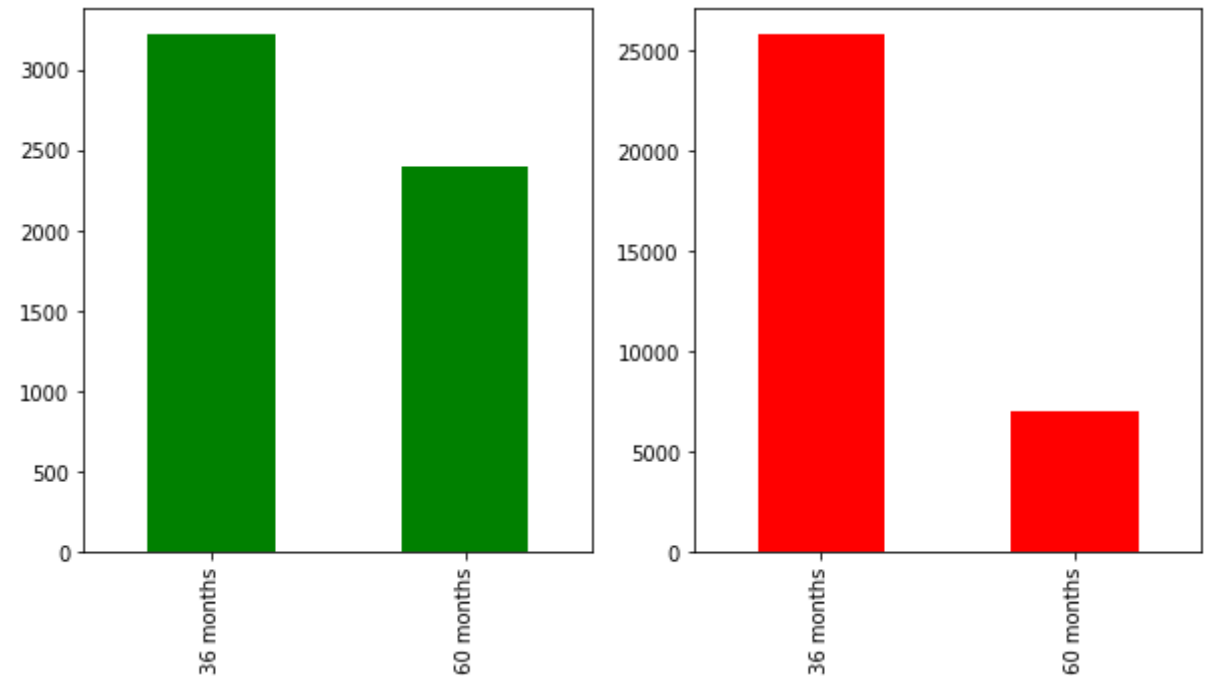


The Analysis of the Univariate : in “Term” Column

- loan\_dataset\_of\_defaulters
- loan\_dataset\_of\_non\_defaulters

Recommendations:

- “loan defaulters” are more than 3000 for 36 months and 2500 for 60 months.
- “loan non defaulters” are more than 25000 for 36 months and more than 5000 for 60 months.

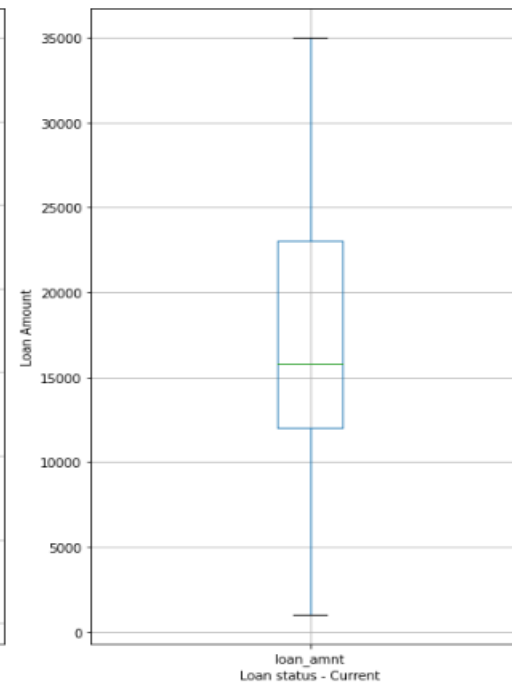
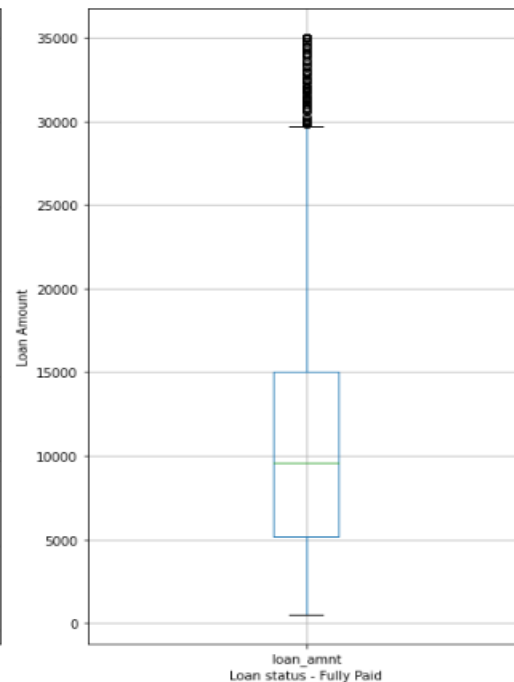
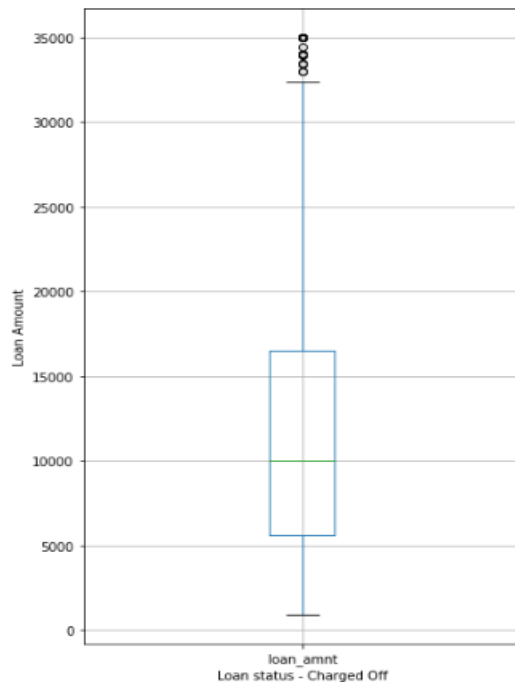


Insight:

To analysis “**Loan status**” Column in sub plot and dedication of outliers

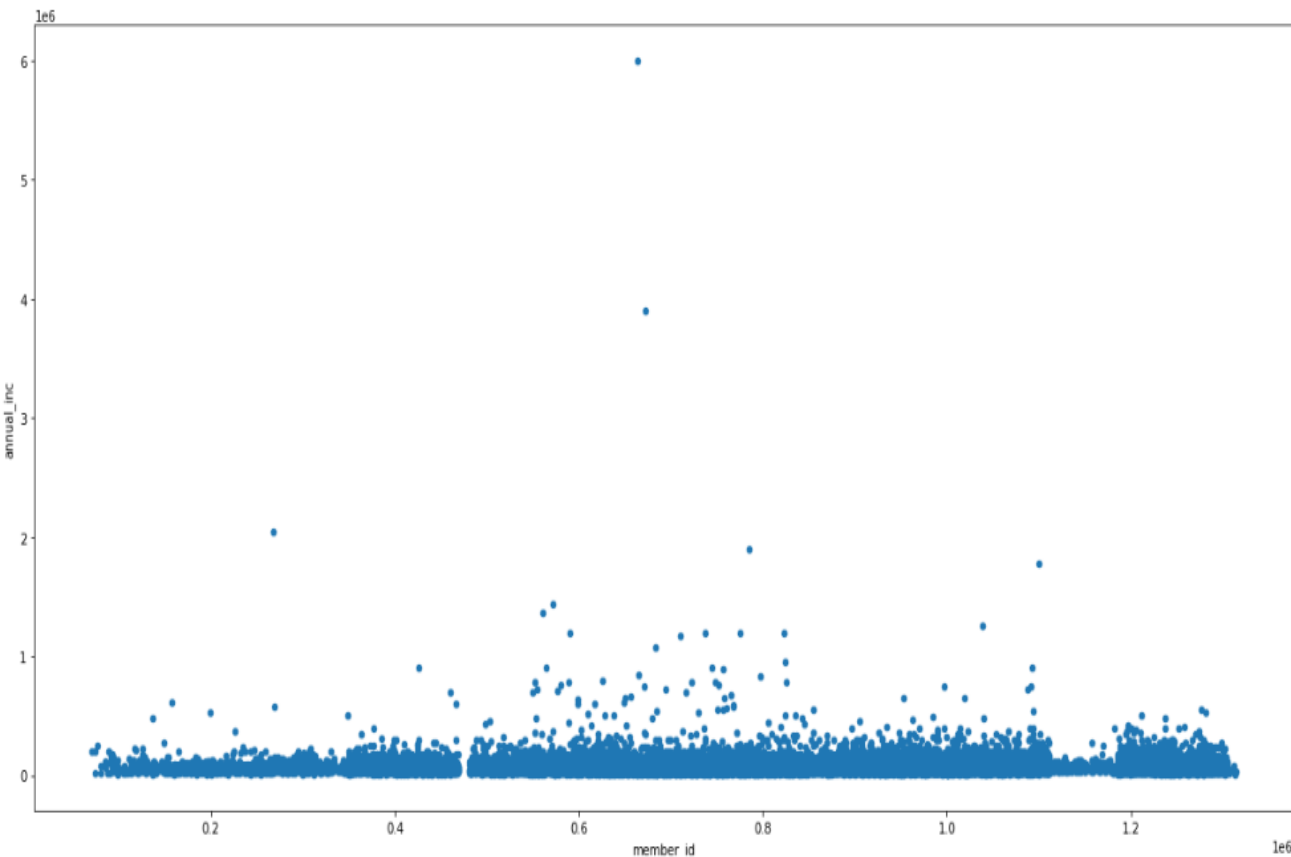
Recommendations:

- We can see the outlier in the “Charged off” more than 33000.
- We can see the outlier in the “Fully paid” more than 30000.
- We can don't see any outlier in the “Current.”



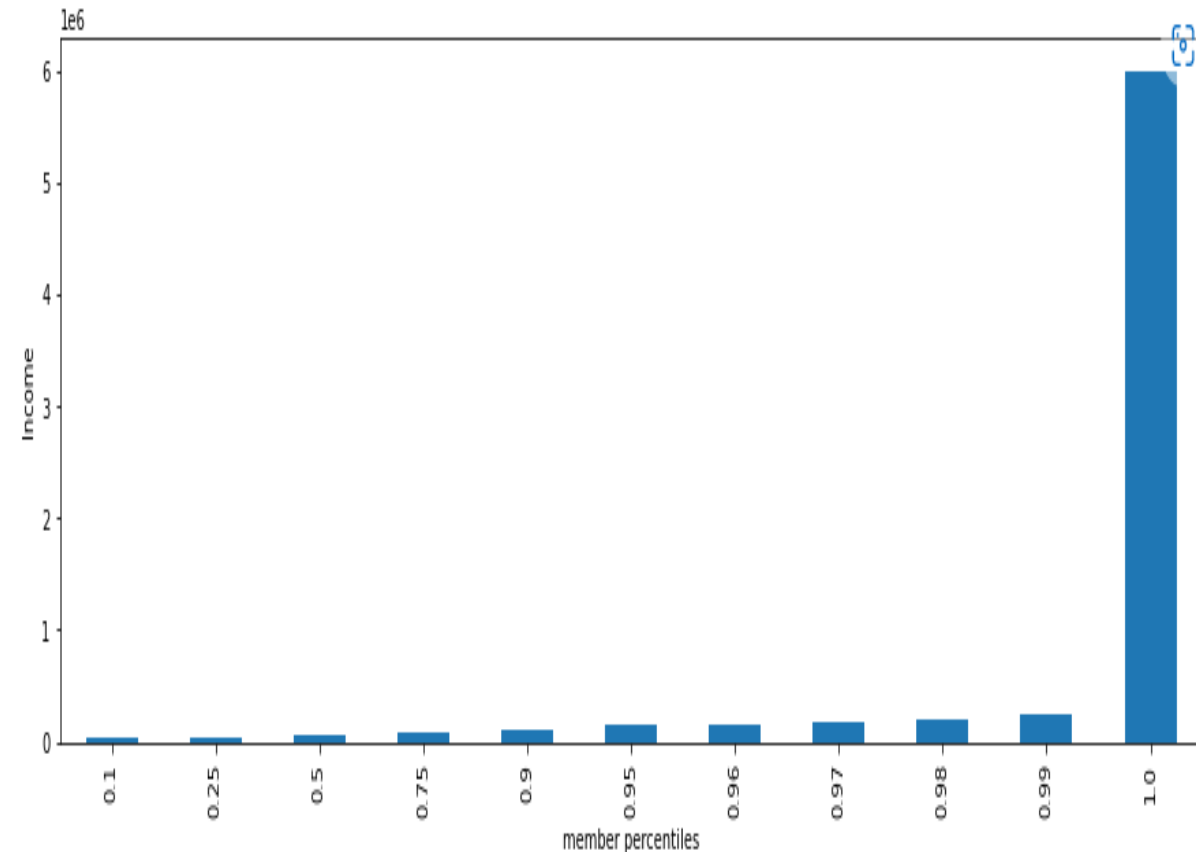
Insight:

To analysis “**Annual Income Analysis**” Column in finding the outliers



Recommendations:

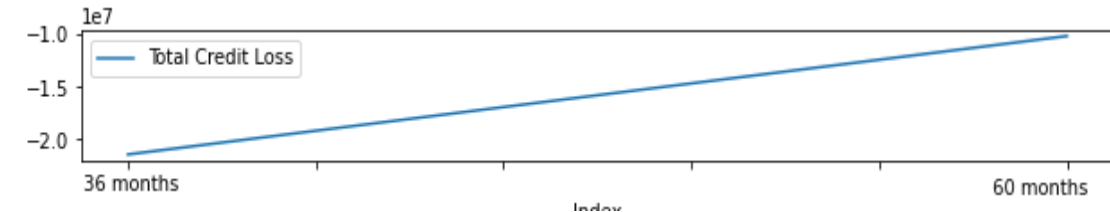
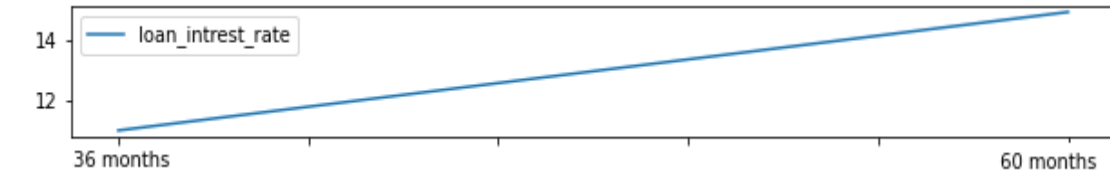
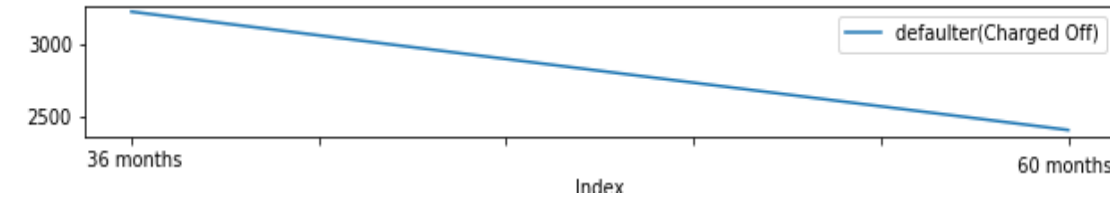
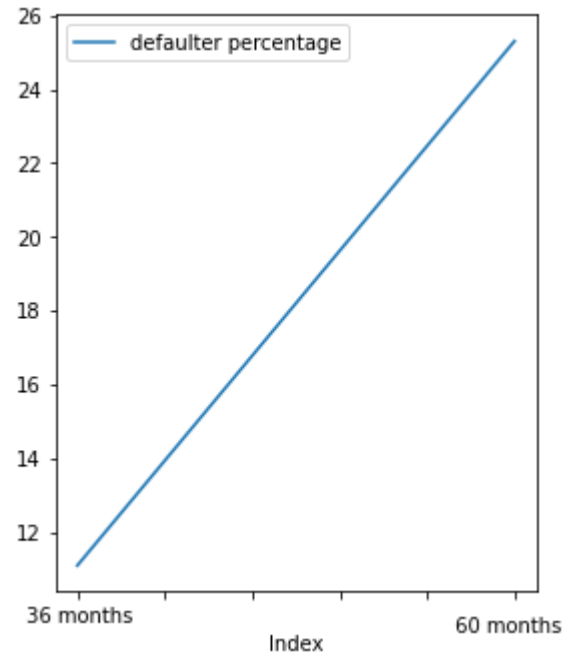
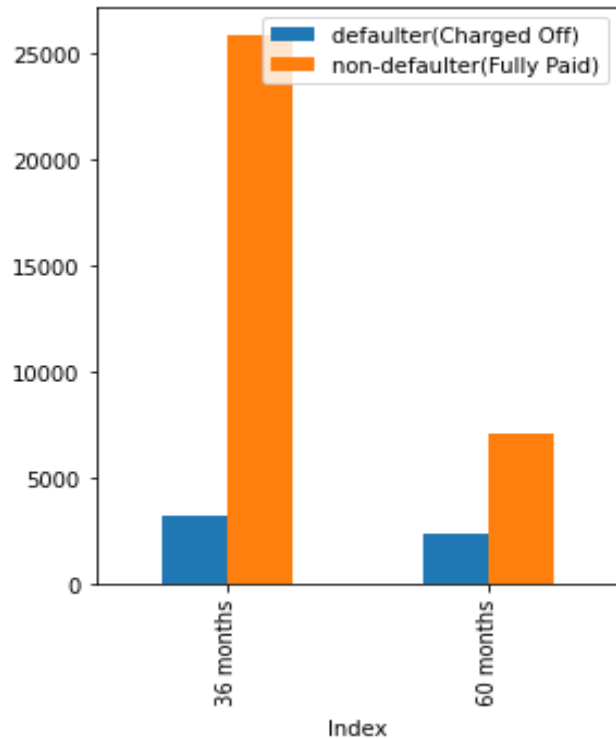
- We can see the Scatter plot for the outlier.
- In Particular we can see the last 1 quantile are the outliers comparing to the rest of the values.





Insight:

Here we are comparing 2 columns Loan Term & Loan Status and to find the relationship between them.



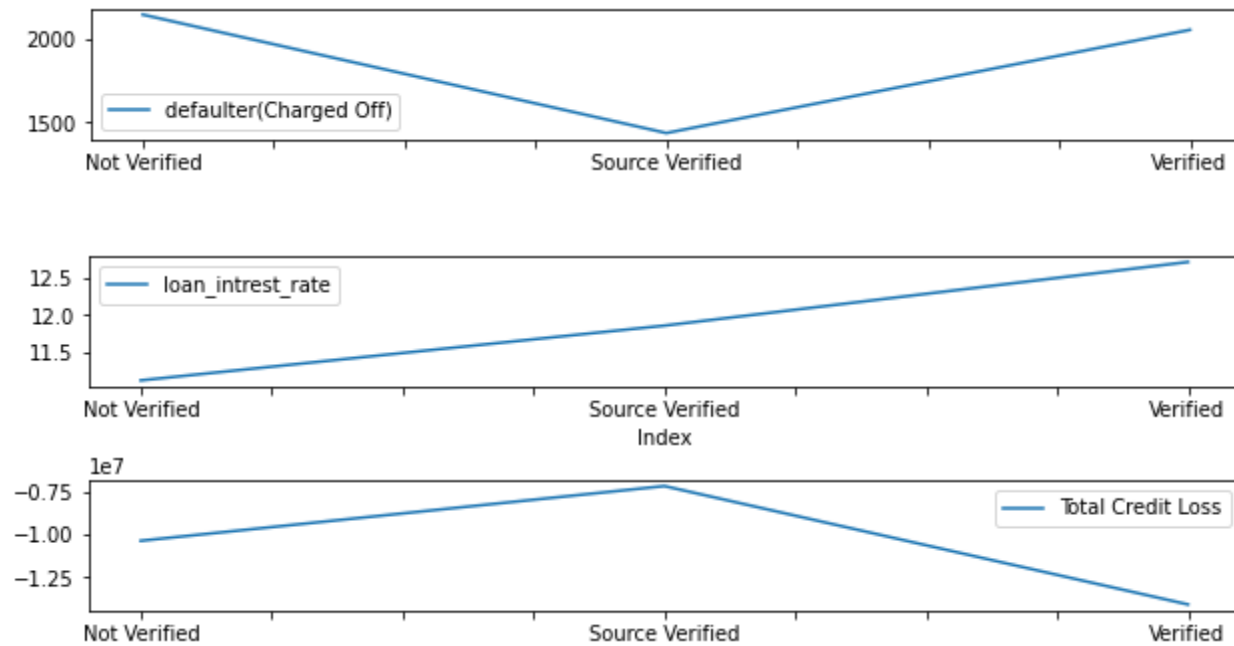
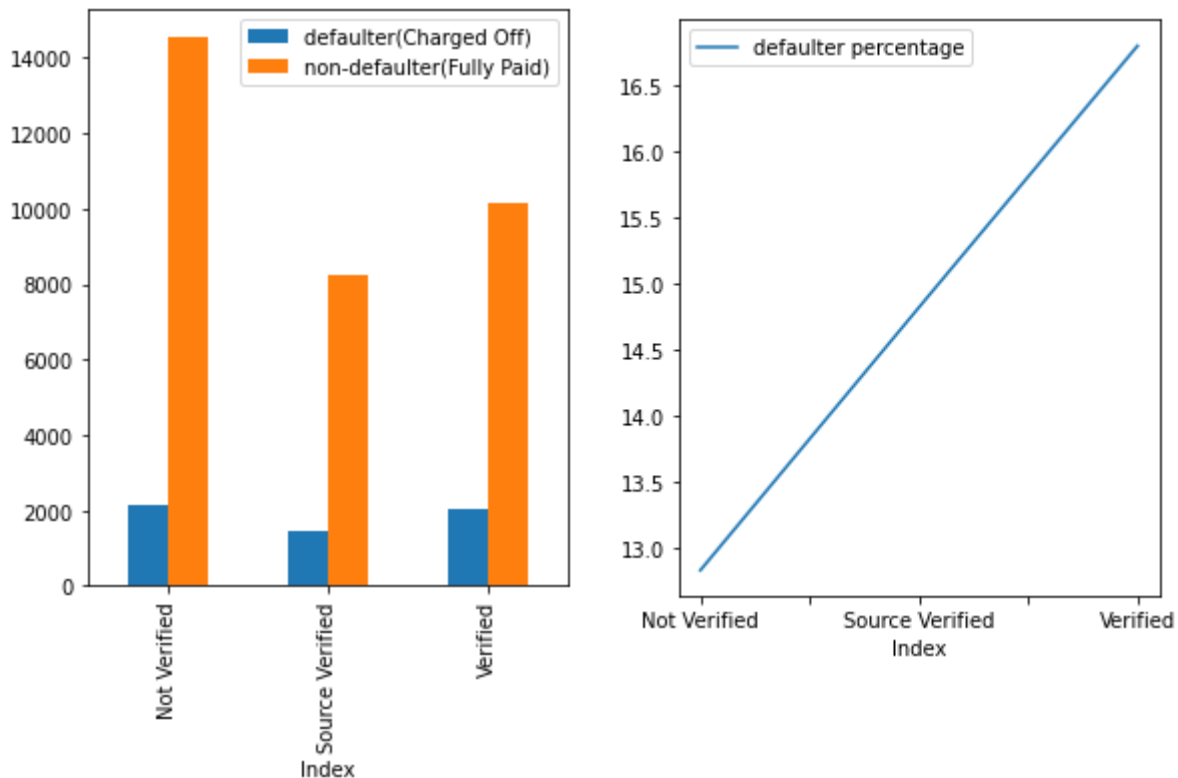
Recommendations:

- The Data shows the person whose having the 36 months are getting less default compared to the 60 months of data and also the line graph show that the % of defaulters are more in 60 months compared 36 months.
- The Line graph shows that the defaulter are getting reduced on Charge off for the 60 months period.
- The Loan interest and the Credit loss also getting increases for the 60 months.

# Data analysis - Bivariate Analysis – verification status & loan status upGrad

## Insight:

Here we are comparing 2 columns Verification status & Loan Status and to find the relationship between them.

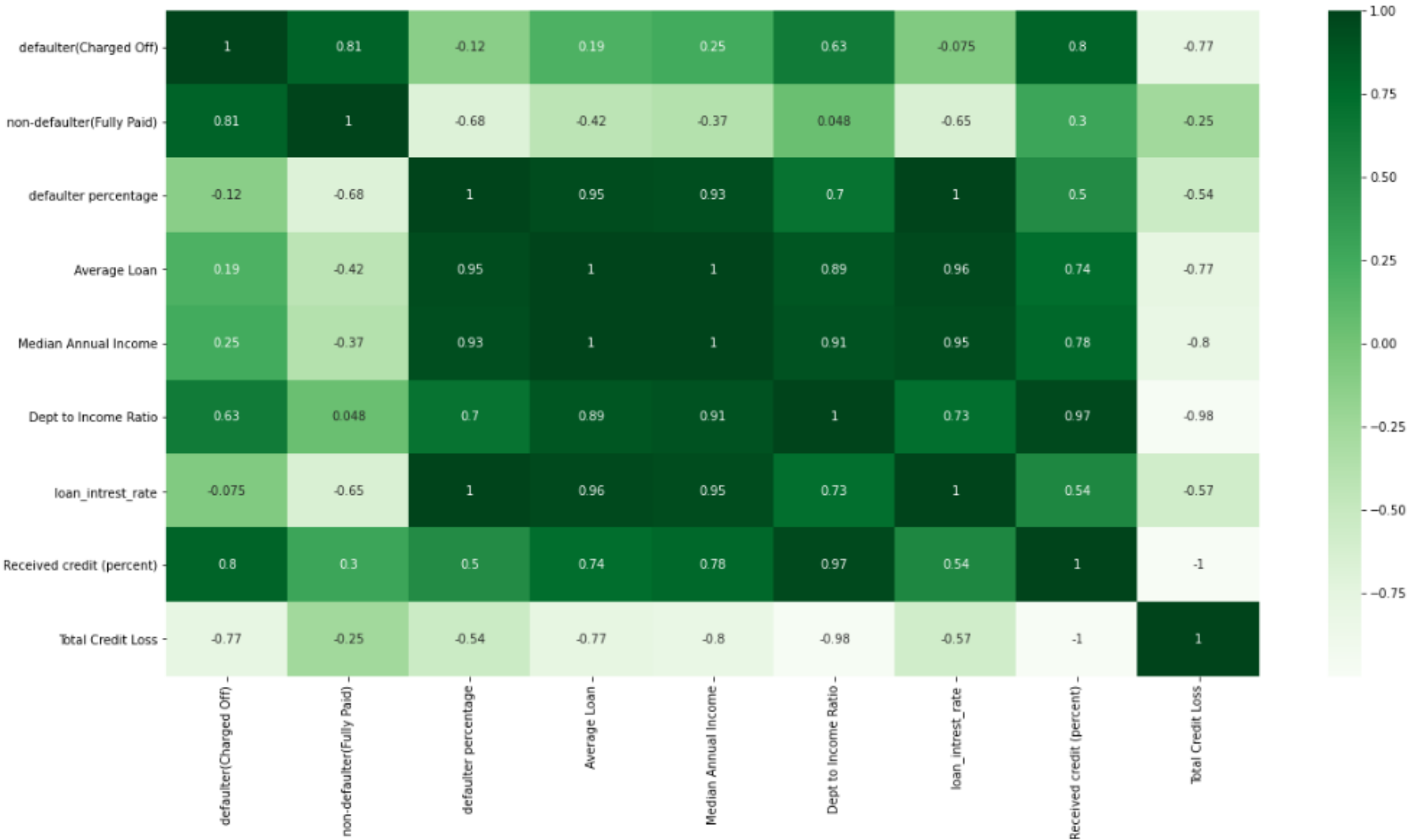


## Recommendations:

- The Data shows the person whose fully paid are more for the not verified and also the default rate is also high .
- The Line graph shows that the defaulter are getting reduced on Charge off for the Source Verified compared to the rest of 2 other values.
- The Loan interest increase for verified persons and the Credit loss also getting decreases for the verified .

Recommendations:

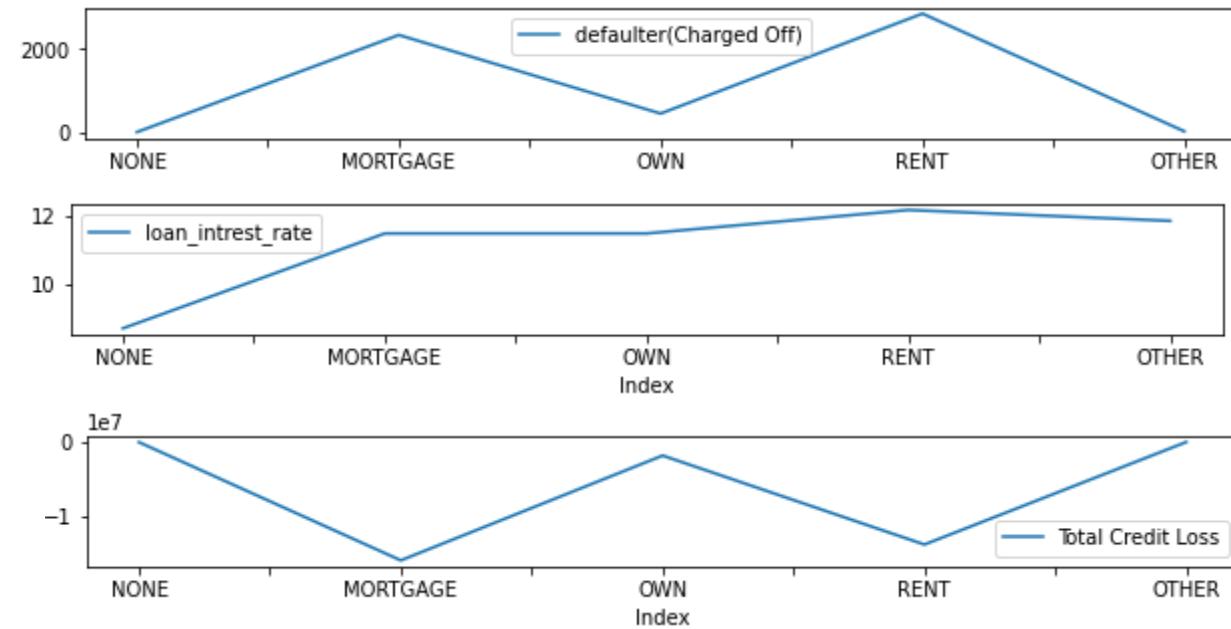
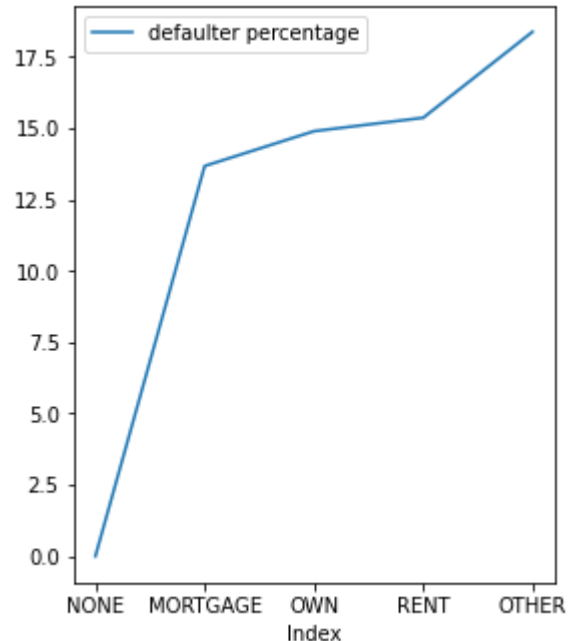
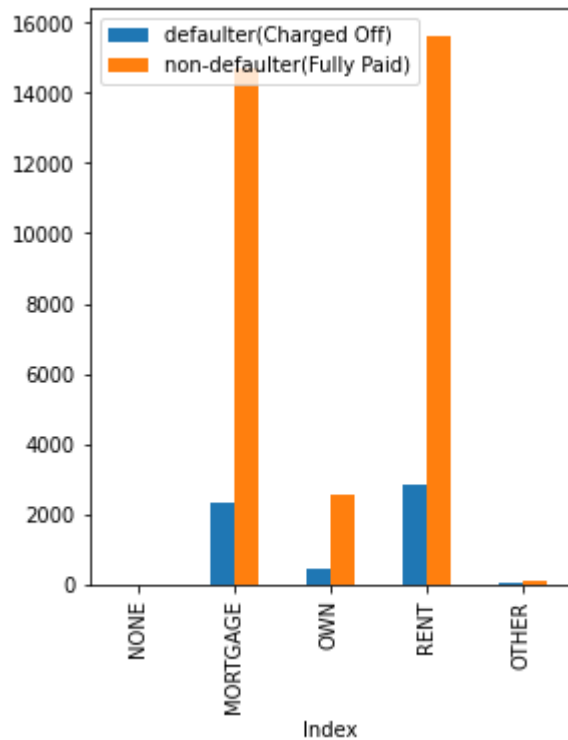
We could see the Positive correlation with Verification status and Loan status



# Data analysis - Bivariate Analysis – Home ownership to loan status upGrad

## Insight:

Here we are comparing 2 columns Home ownership to loan status and to find the relationship between them.



## Recommendations:

- The Data shows the person who are doing either Rent / Mortgage are able to pay fully .
- The Line graph shows that the defaulter are getting reduced on the person Owning the property.
- The Loan status and the Credit loss also getting decreases for the applicant who are letting the property to either Rent or Mortgage.

# Data analysis - Bivariate Analysis – Home ownership to loan status upGrad

Recommendations:

We could see the Positive correlation with Home ownership to loan status



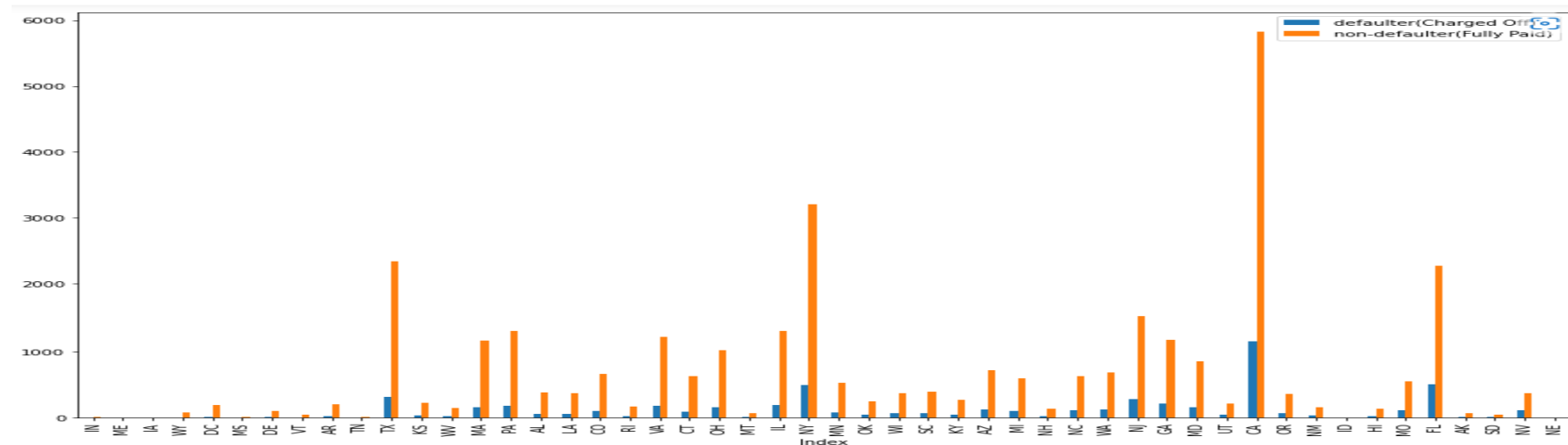
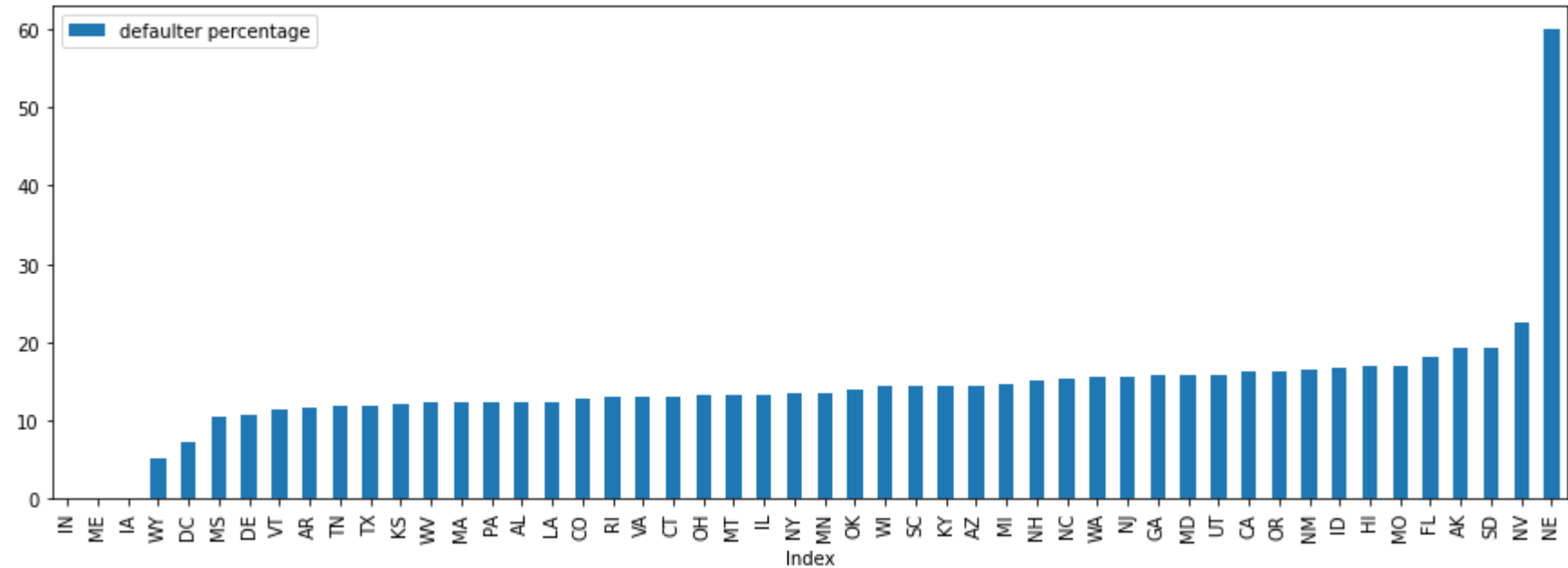
# Data analysis - Bivariate Analysis – Area & Loan Credit status upGrad

## Insight:

To find the relationship with the Area and Loan Credit status

## Recommendations:

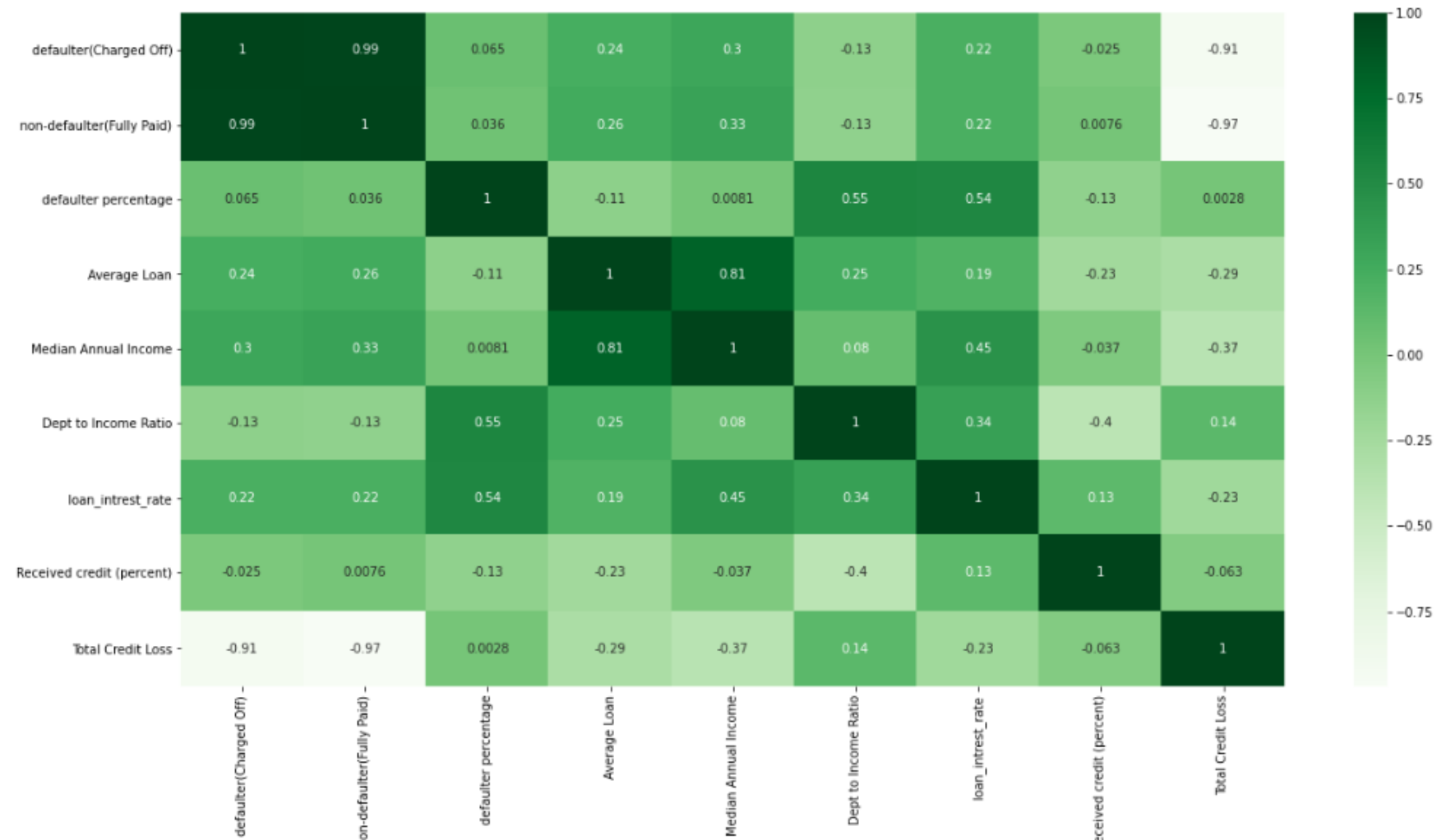
- The Individuals who belongs to CA and NY states with default rate equivalent to other state are tend to create less credit losses to business.
- The Individuals who belongs to NV state with higher default rates than other states are tend to create more credit losses to business.
- The Individuals who belongs to NE state with the highest default rate are also tend to create high credit losses.
- The Individuals who belongs to OH,VA, MA, PA States with average default rate are tend to create overall profitable to businesses.



# Data analysis - Bivariate Analysis – Area & Loan Credit status upGrad

Recommendations:

We could see the Positive correlation with Home ownership to loan status



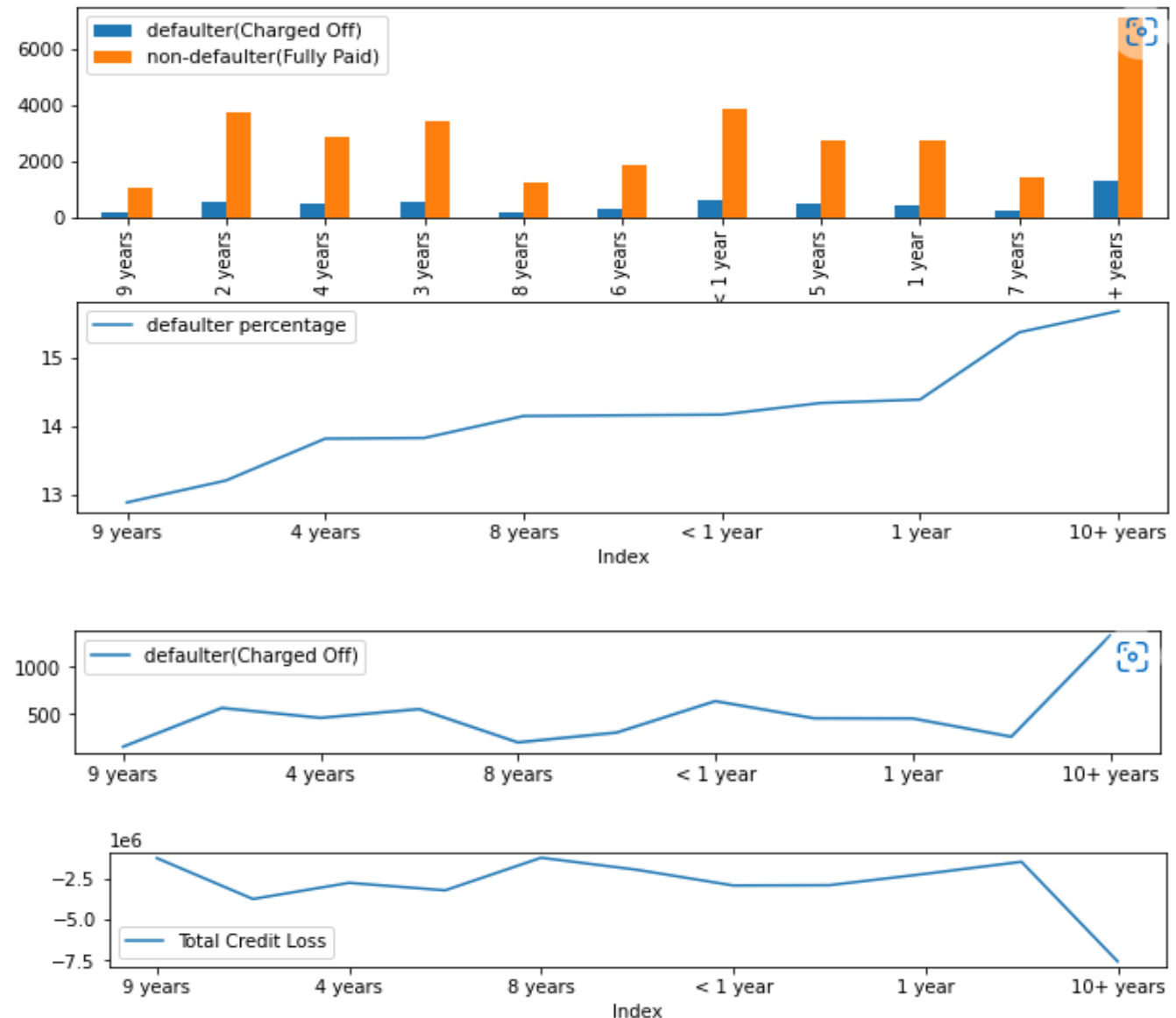
# Data analysis - Bivariate Analysis – Employment length of borrower & loan status

## Insight:

To find the relationship with the Area and Load Credit status

## Recommendations:

- The Individuals with 10+ years of experience are more likely to default, as they have higher loan eligibility with average interest rate .but there also a possibility to create the least credit losses i.e they are most profitable to business.
- The Individuals with 9 years of experience have the lowest interest rate with lowest default rates are considered to create the most credit loss to the business or they are least profitable. Also they have the highest loan recovery rate.



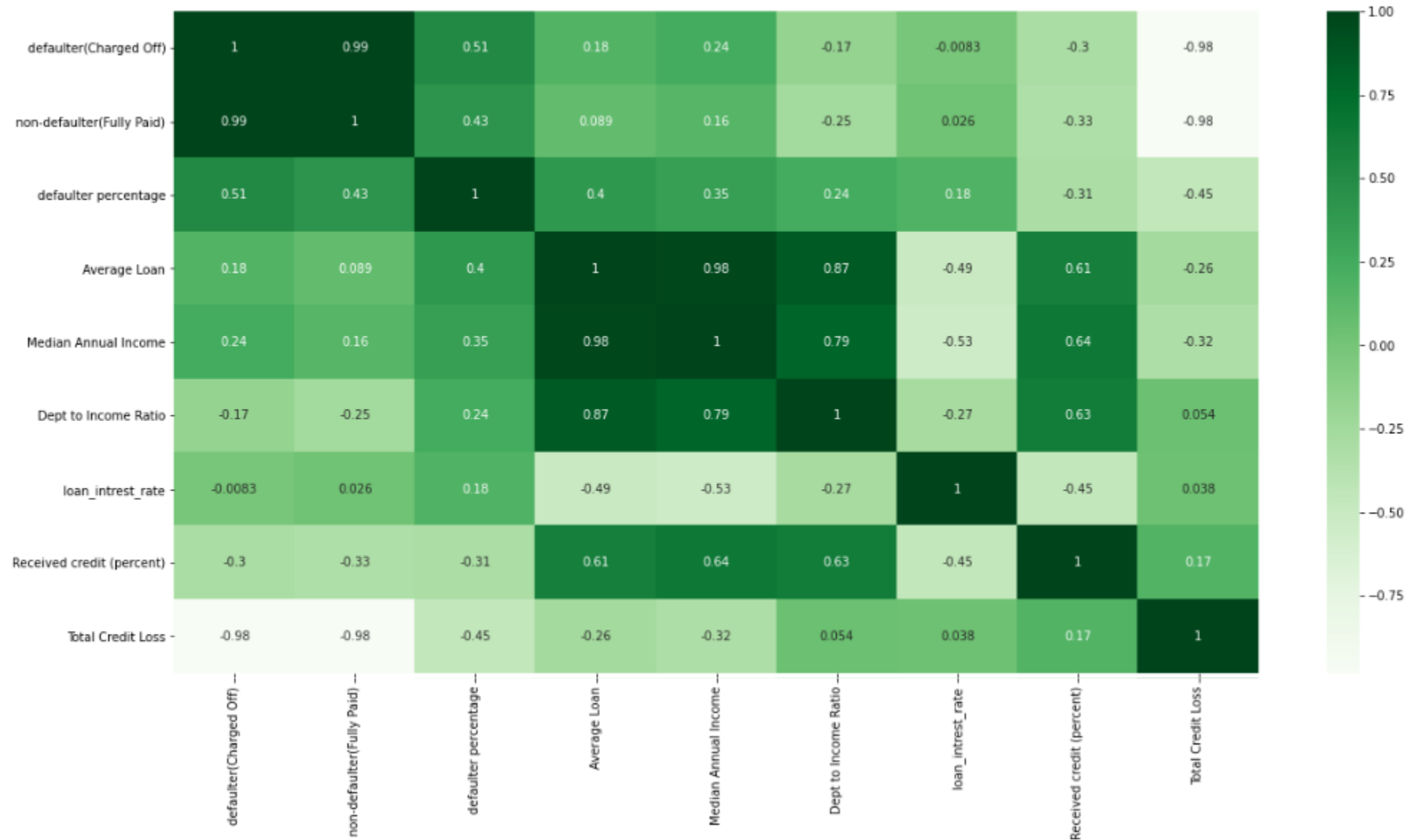


# Data analysis - Bivariate Analysis – Employment length of borrower & loan status

upGrad

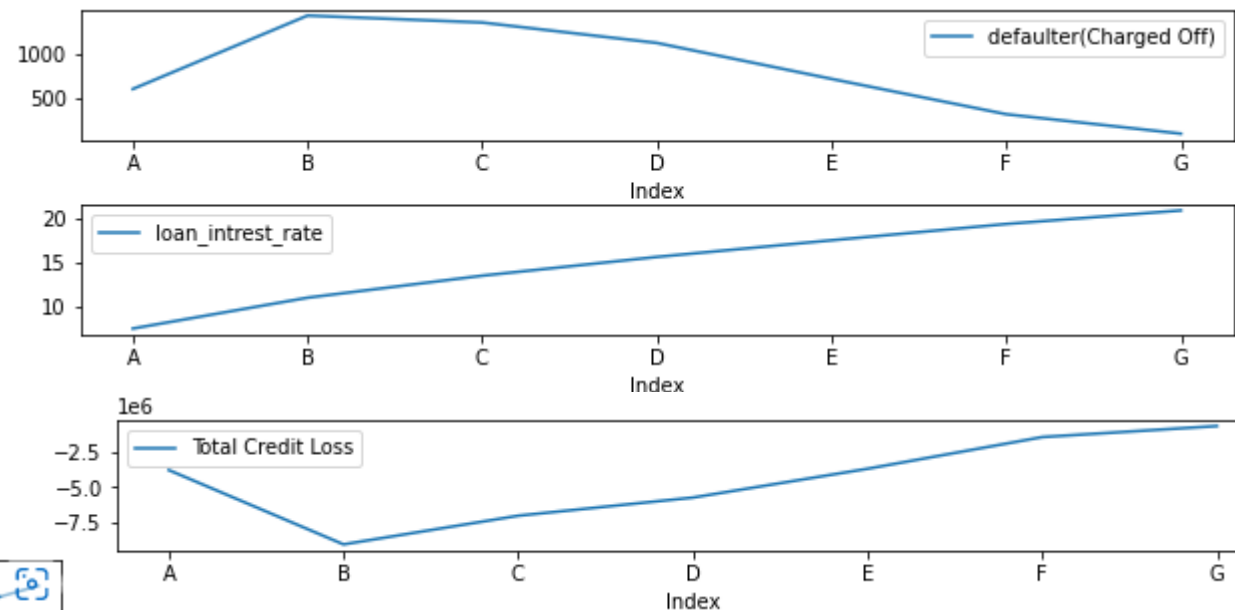
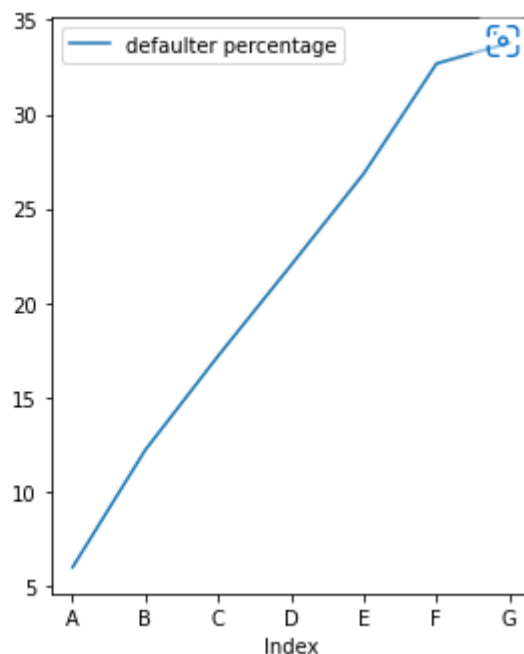
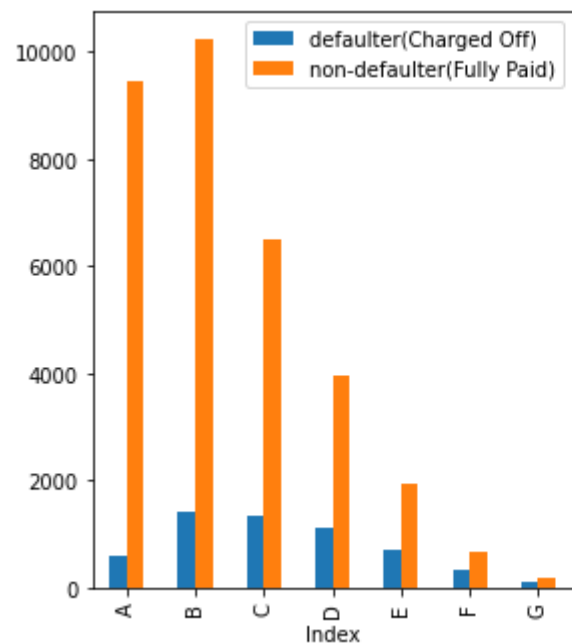
Recommendations:

We could see the Positive correlation with Home ownership to loan status



## Insight:

Here we are comparing 2 columns Grade & loan status and to find the relationship between them.

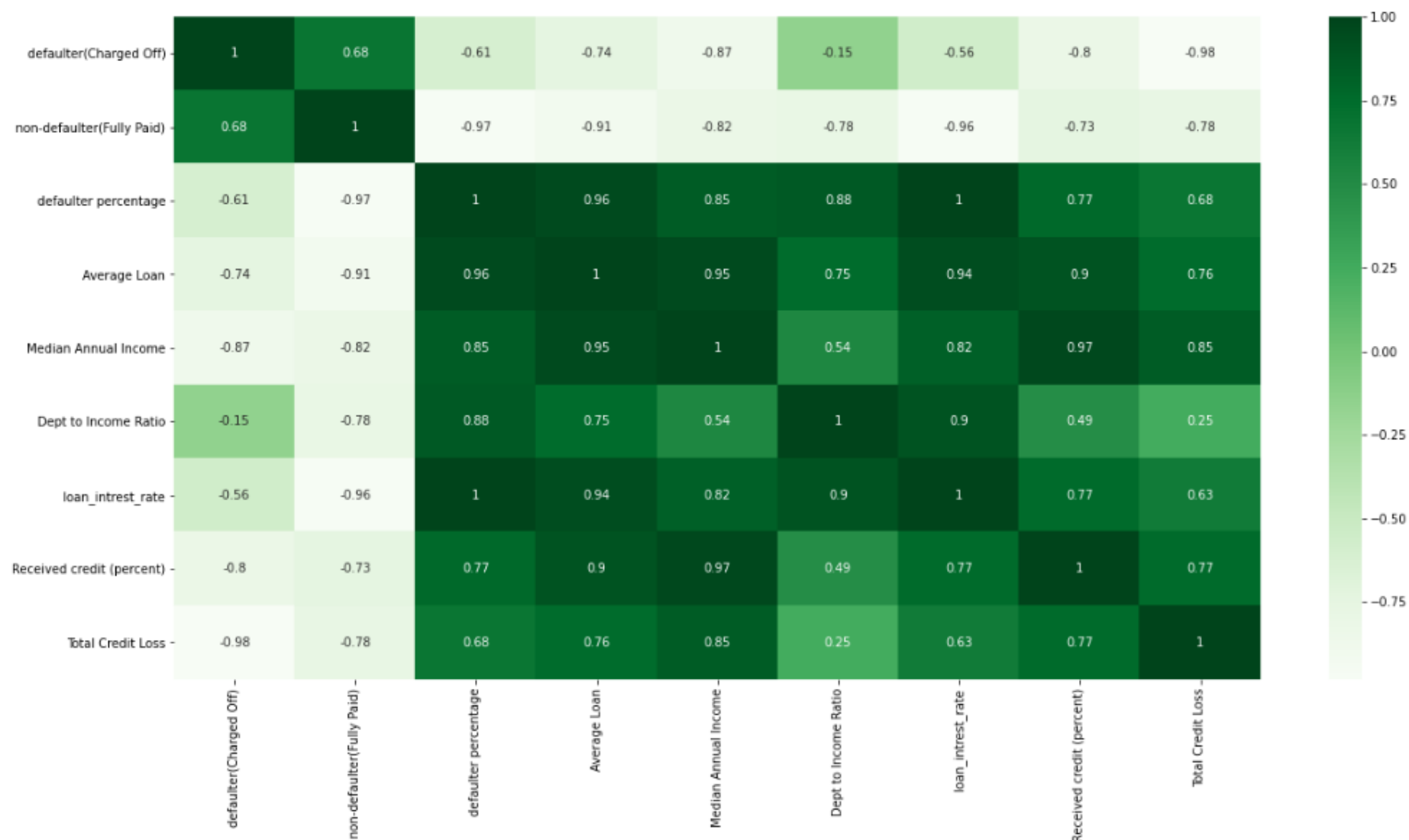


## Recommendations:

- The Individuals with 'A' LC-grade have the lowest debt and are less likely to default as they have lowest interest rates and lowest loan credit. But have the lowest recovery rates.
- The Individuals with 'B' LC-grade have the lower debt rate than average are the most profitable to business with low credit loss.

Recommendations:

We could see the Positive correlation with Home ownership to loan status



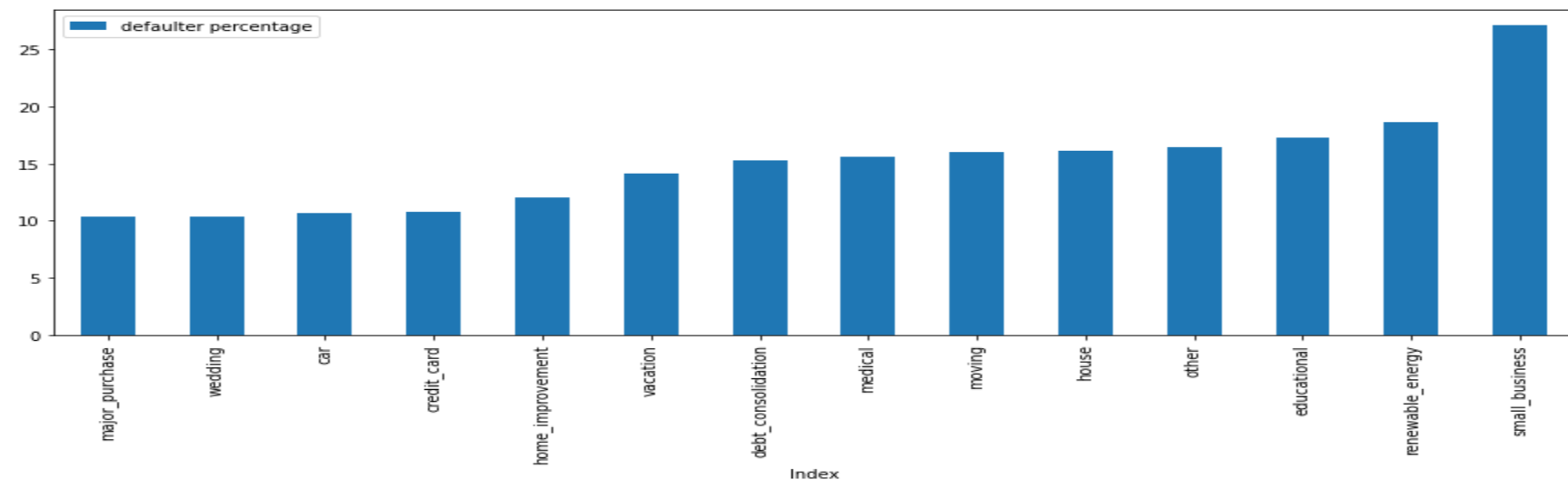
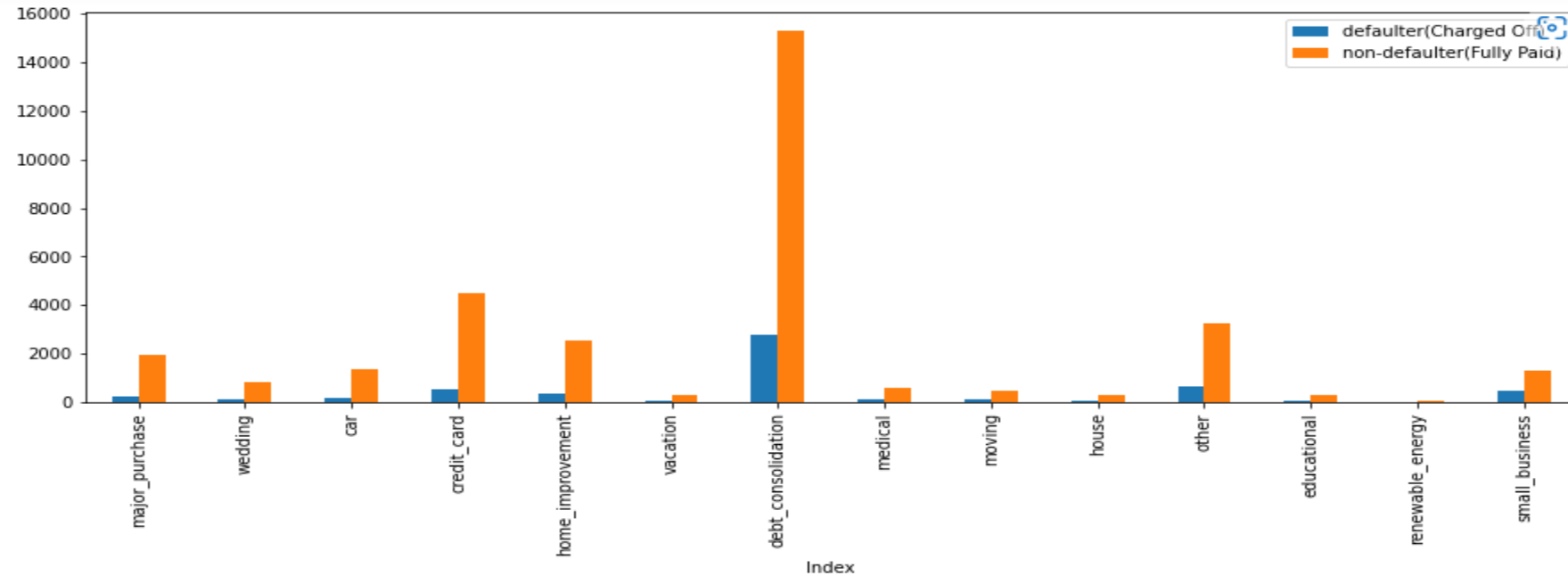
# Data analysis - Bivariate Analysis – Purpose of Loan and Loan status

## Insight:

Here we are comparing 2 columns Purpose of Loan and Loan status and to find the relationship between them.

## Recommendations:

- Loans taken for small business are more likely to create higher credit loss for the business with the highest default rates. They also have higher loan amounts with highest interest rates.
- Loans taken for 'debt consolidation' have an average default rate with loan eligibility of more than average are highly profitable to the business.
- Individuals with income verified and 10+ yrs experience are more profitable to business despite their highest default rate.



## Observation 1:

- Individuals whose income can be verified with income more than 70000, tend to default in the more on their loans , as they are eligible for higher loan amounts than non-verified individuals and have a higher dept. But individuals in this case tend to create lowest credit loss for a business as their interest rates are also higher.
- Individuals who own a house have a default rate of 14.89%, have the possibility to create higher bad credit or to create less profits with recovery. Also they can have low recovery rate out of all loans credited.
- Individuals who mention their home ownership as OTHER have the highest default rate of 18.37% , have the possibility to create losses in the overall recovery of the loan. Also they can have the lowest recovery rate.
- Individuals who RENT or MORTAGAGE the house have high recovery rate and have higher annual income are considered to be profitable in recovery for all loans.

## Observation 2

- Individuals who belongs to CA and NY states with default rate equivalent to other state are tend to create less credit losses to business.
- Individuals who belongs to NV state with higher default rates than other states are tend to create more credit losses to business.
- Individuals who belongs to NE state with the highest default rate are also tend to create high credit losses.
- Individuals who belongs to OH,VA, MA, PA States with average default rate are tend to create overall profitable to businesses.

## Observation 3:

- The Default Rate increases as the years of employment increases. Individuals with 10+ years of experience are more likely to default, as they have higher loan eligibility with average interest rate .but there also a possibility to create the least credit losses i.e they are most profitable to business.
- The Individuals with 9 years of experience have the lowest interest rate with lowest default rates are considered to create the most credit loss to the business or they are least profitable. Also they have the highest loan recovery rate.

## Observation 4:

- Individuals with 'A' LC-grade have the lowest debt and are less likely to default as they have lowest interest rates and lowest loan credit. But have the lowest recovery rates.
- Individuals with 'B' LC-grade have the lower debt rate than average are the most profitable to business with low credit loss.

## Observation 5:

- Loans taken for small business are more likely to create higher credit loss for the business with the highest default rates. They also have higher loan amounts with highest interest rates.
- Loans taken for 'debt consolidation' have an average default rate with loan eligibility of more than average are highly profitable to the business.
- Individuals with income verified and 10+ yrs experience are more profitable to business despite their highest default rate.