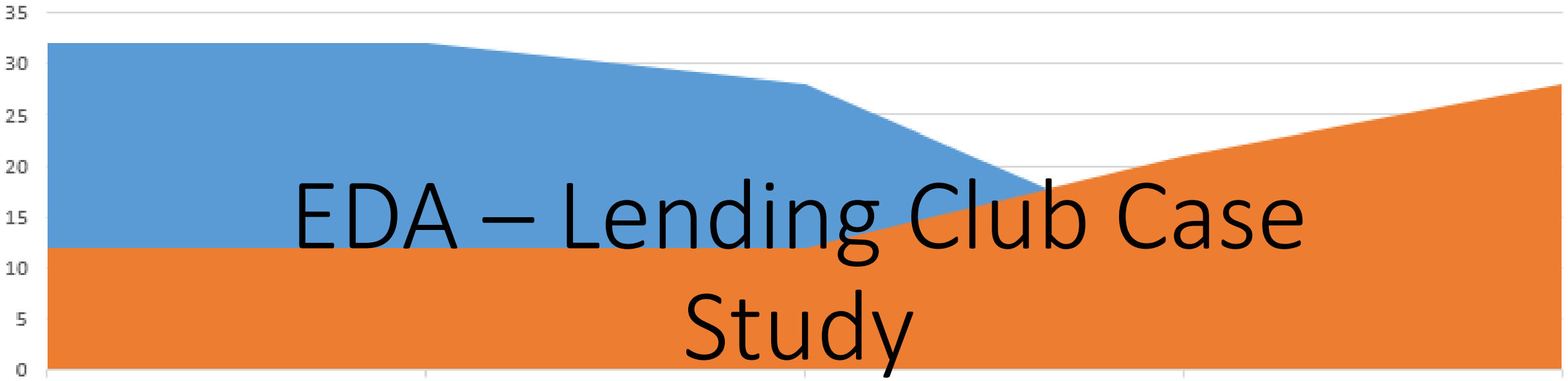


EDA – Lending Club Case Study

By,
Srikanth Navuduri
T.S. Prabakaran

Data Cleaning

- ✓We have removed the column which has only null values
- ✓We have removed the column which has only one value (ex:- only zeros)
- ✓We populate median value for the missing values for some of the columns

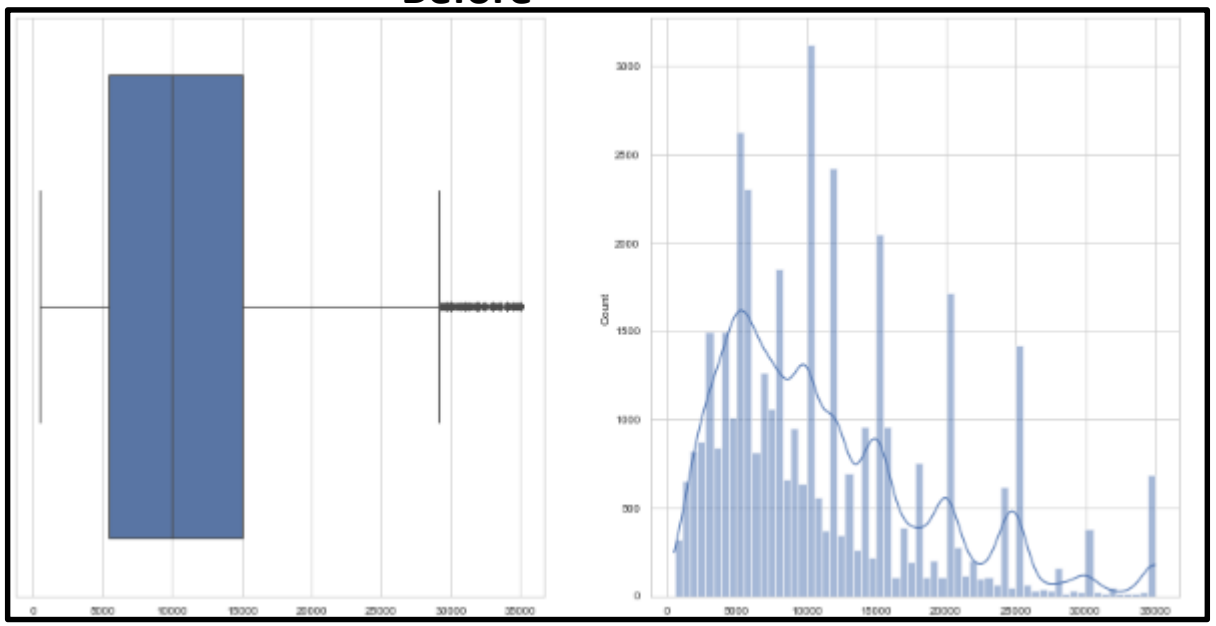
Univariant Analysis

By using univariant analysis, we analyze the amount field and remove the outlier in the data.

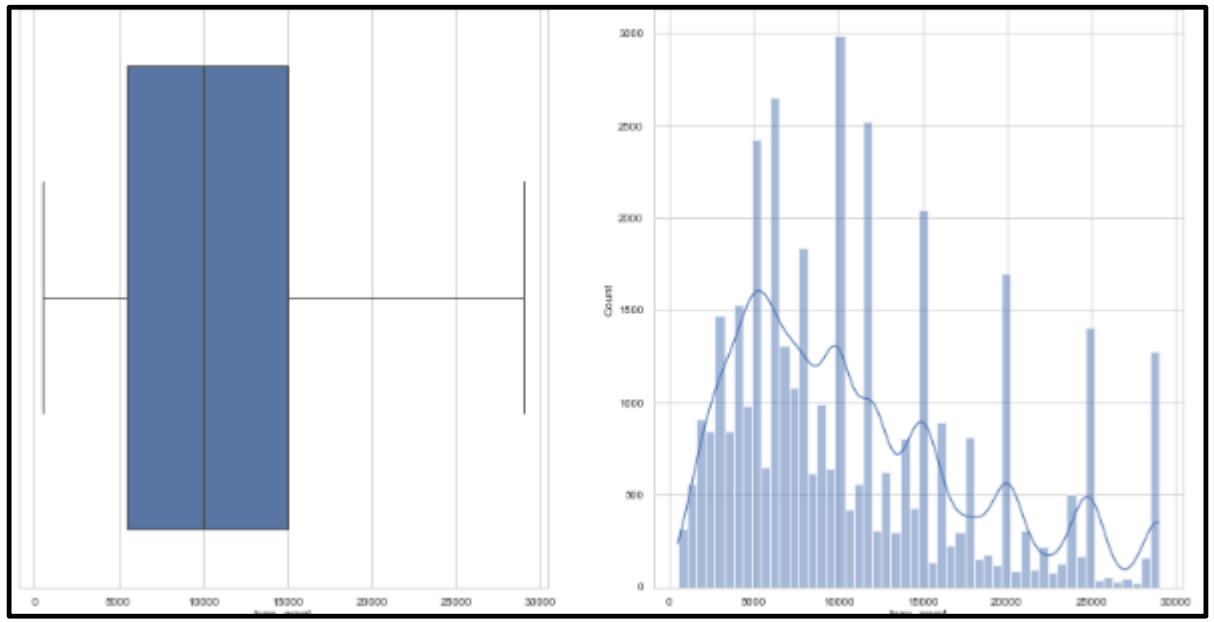
Example:-
If we take, loan_amnt field by setting upper limit we treated outlier and normalize the data.
Like wise we have done outlier correction for all the amount fields

```
upper_limit = 29000
lower_limit = 0
loan_data['loan_amnt'] = np.where(loan_data['loan_amnt'] >= upper_limit, upper_limit, np.where(loan_data['loan_amnt'] <= lower_l
```

Before



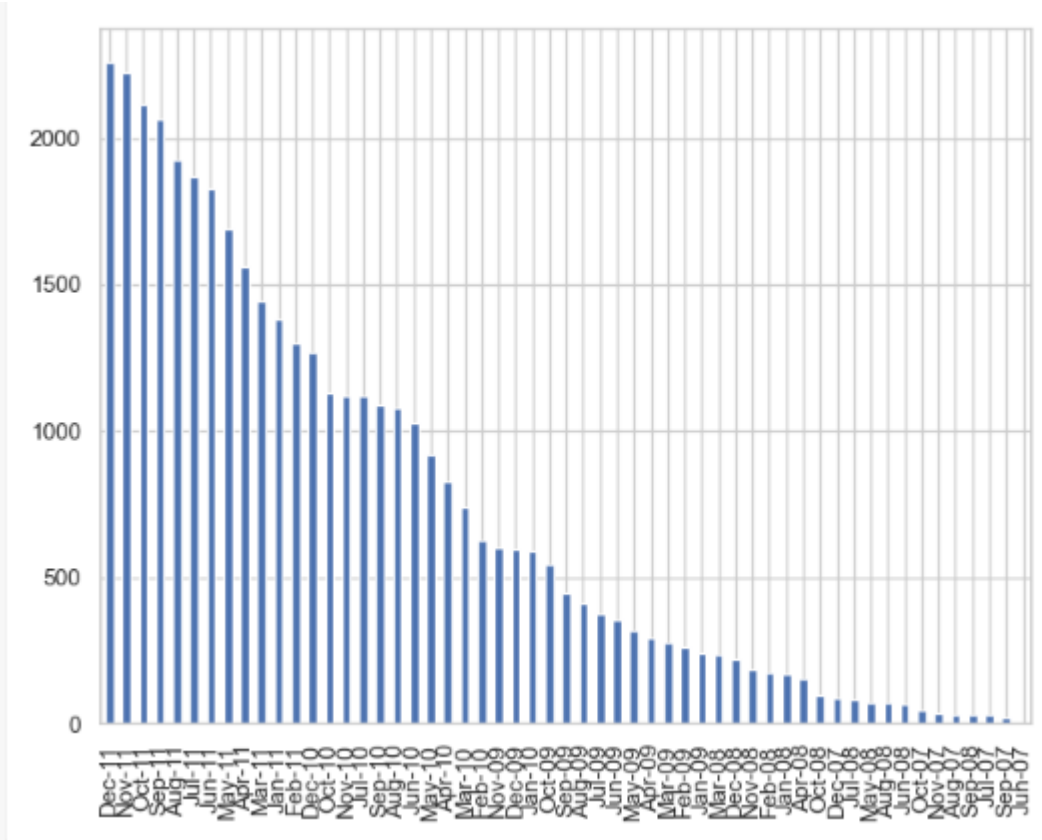
After



Univariant Analysis

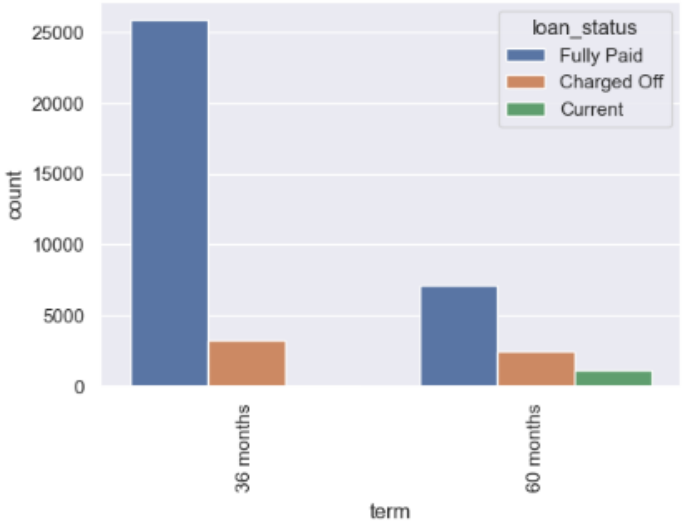
By using univariant analysis, we analyze the issue_d column and the below inferences

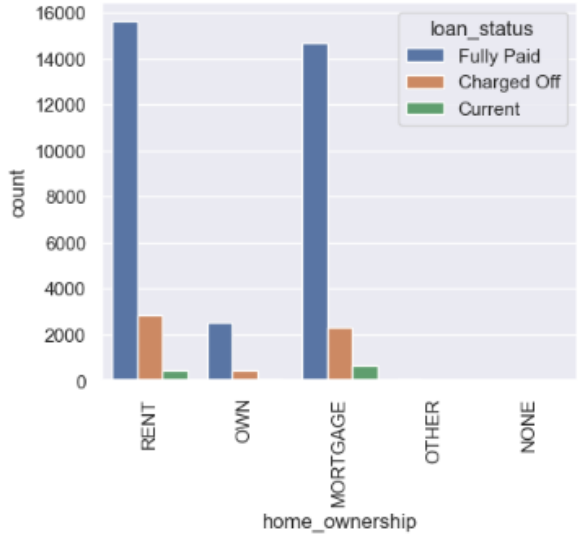
- Most number loan sanctioned in the period from May 2011 to Dec 2011.
- Most number of “charged of loan from in this period May 2011 to Dec 2011.

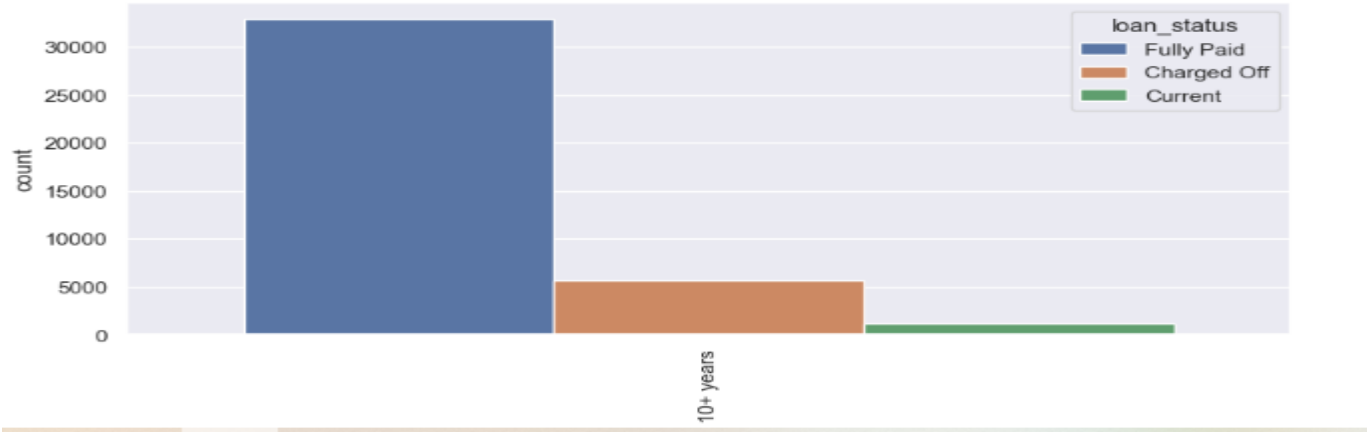


Bivariate Analysis

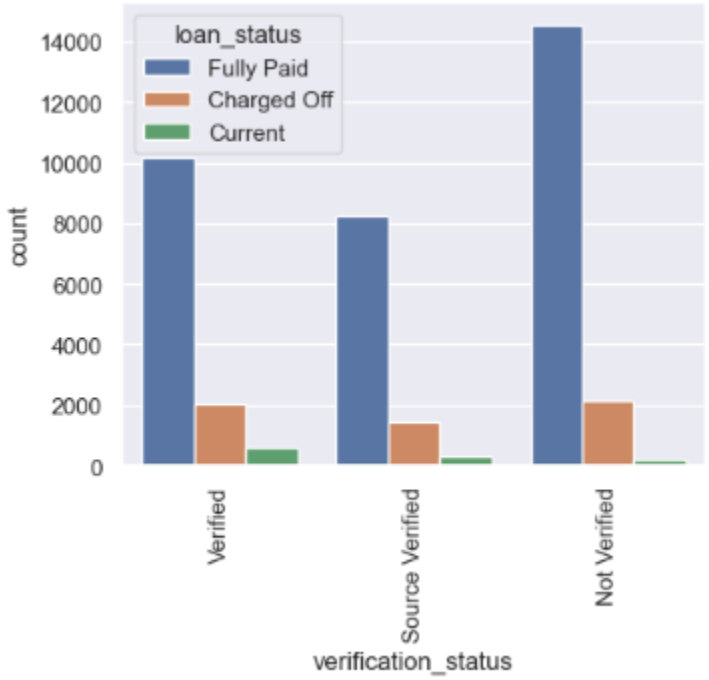
In Bivariate analysis, we working with categorical columns - with loan_status
There are columns with one value across all the rows, ignoring such columns
analyzing them using bar-charts

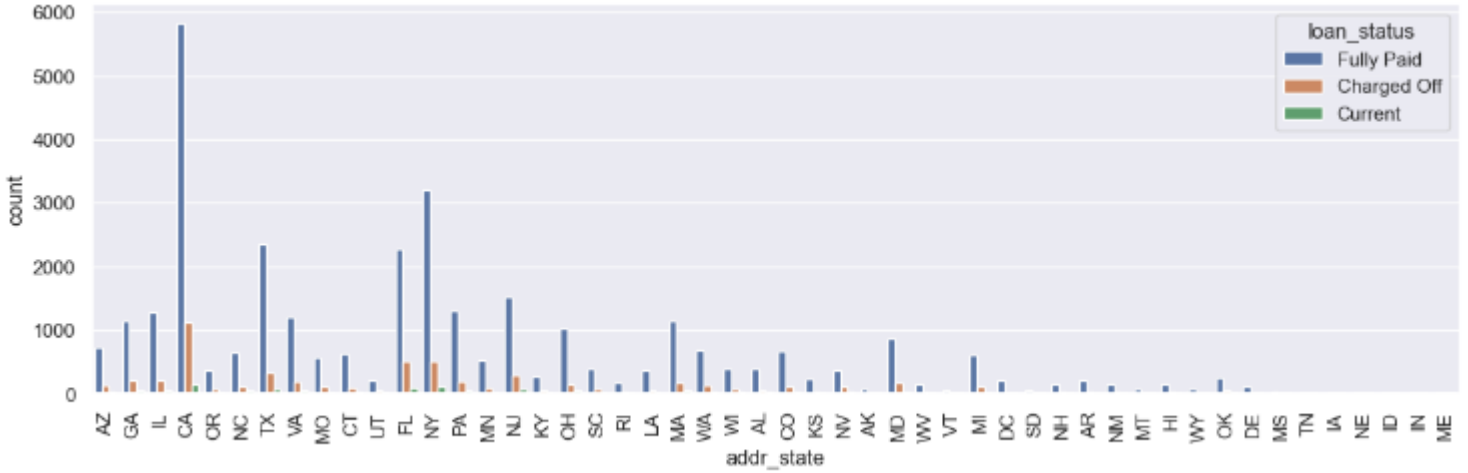


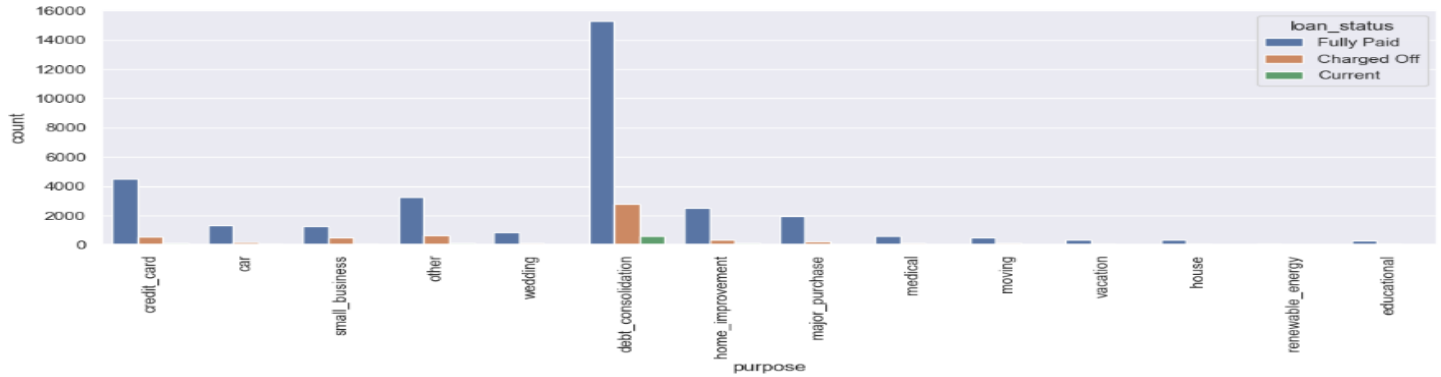




Bivariate Analysis







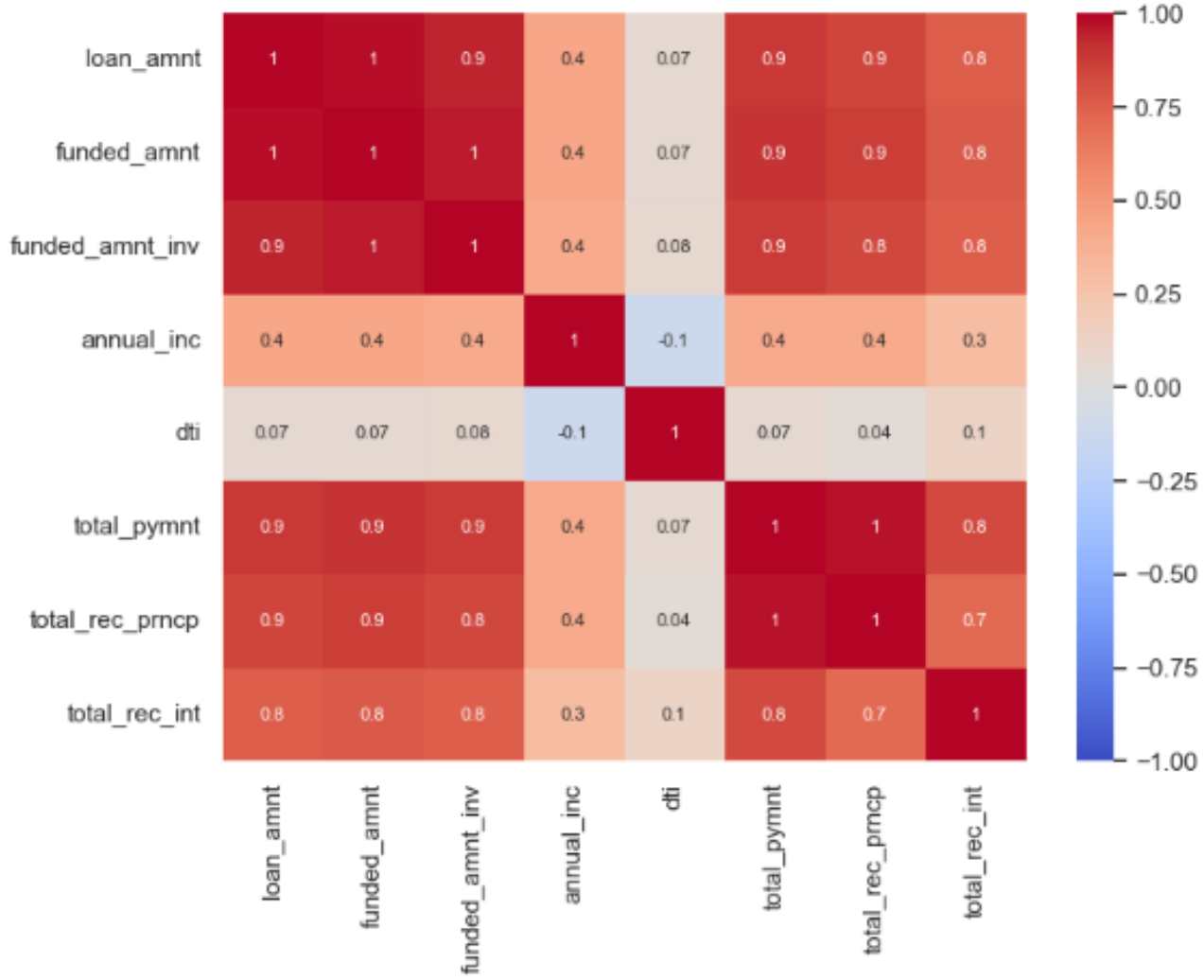
By using this analysis we have inferred the below findings.

- Employee has 10 years of experience, taking more loan.
- Most of loan charged off which term is 60month compared to 36 months
- CA,NY TX, FL are the states where loans given are more
- Most of the loan sanctioned for “Debt-consolidation” purpose

Multivariate Analysis

In Multivariate analysis, we plot heat maps using amount fields,.

- loan_amnt, funded_amnt, funded_amnt_inv are Highly correlated
- Total_pymnt, total_rec_prncp, total_rec_int are Highly correlated



Inferences

- ✓ Most of loan has sanctioned without proper verifying the income
- ✓ Most loans which sanctioned from May 2011 to Dec 2011 is "Charged Off"
- ✓ Employee has 10 years of experience, taking more loan.
- ✓ Most of loan charged off which term is 60month compared to 36 months
- ✓ CA,NY TX, FL are the states where loans given are more
- ✓ Most of the loan sanctioned for “Debt-consolidation” purpose

Conclusion

- ❖ Since most of loan has sanctioned without proper verifying the income, process needs to tighten there, to avoid financial loss