# Lending club case study

Data handling

- Importing the data and checking the data variables.
- Knowing about data using describe statement.
- Checking the missing value count and removing the missing values columns, if the missing value count is more than 40%.
- Dropping columns not necessary for analysis.
- Imputing the missing values using different techniques.
- Removing the signs from the values.
- Finding the outliers:loan_amount,funded_amount,annual_inc.
- Creating bins for easy imputation.

Univariate analysis

- Most of the employees having 10+ years of experiences

- Not verified accounts having the more portions, these is the one of the factors for defaulting.
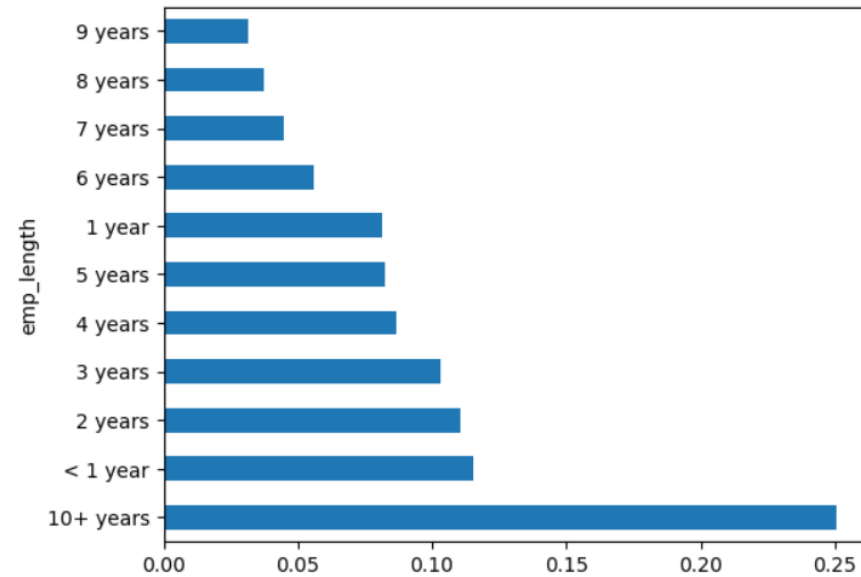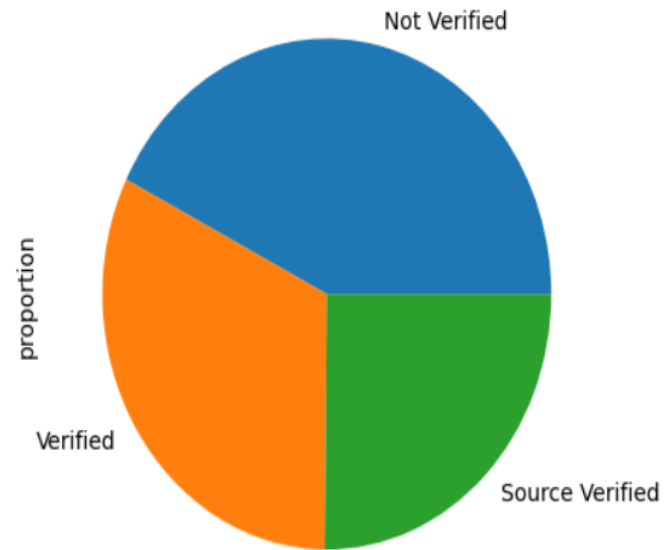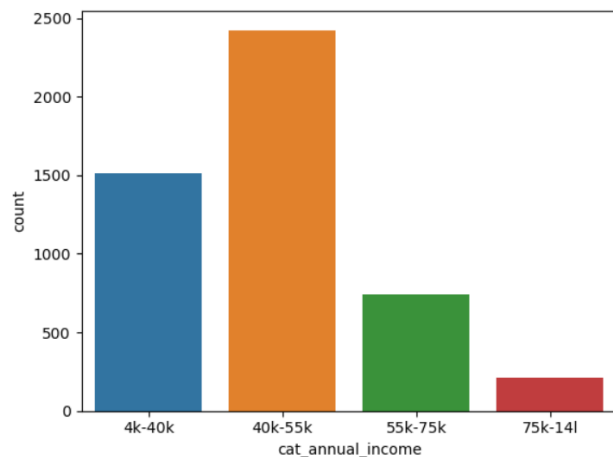
Univariate analysis
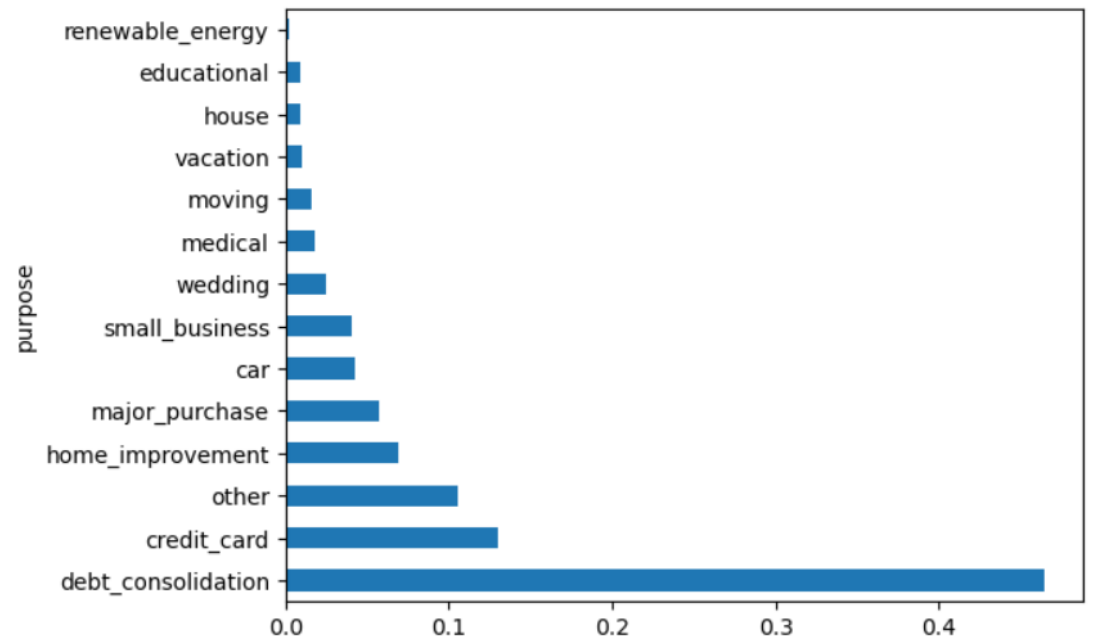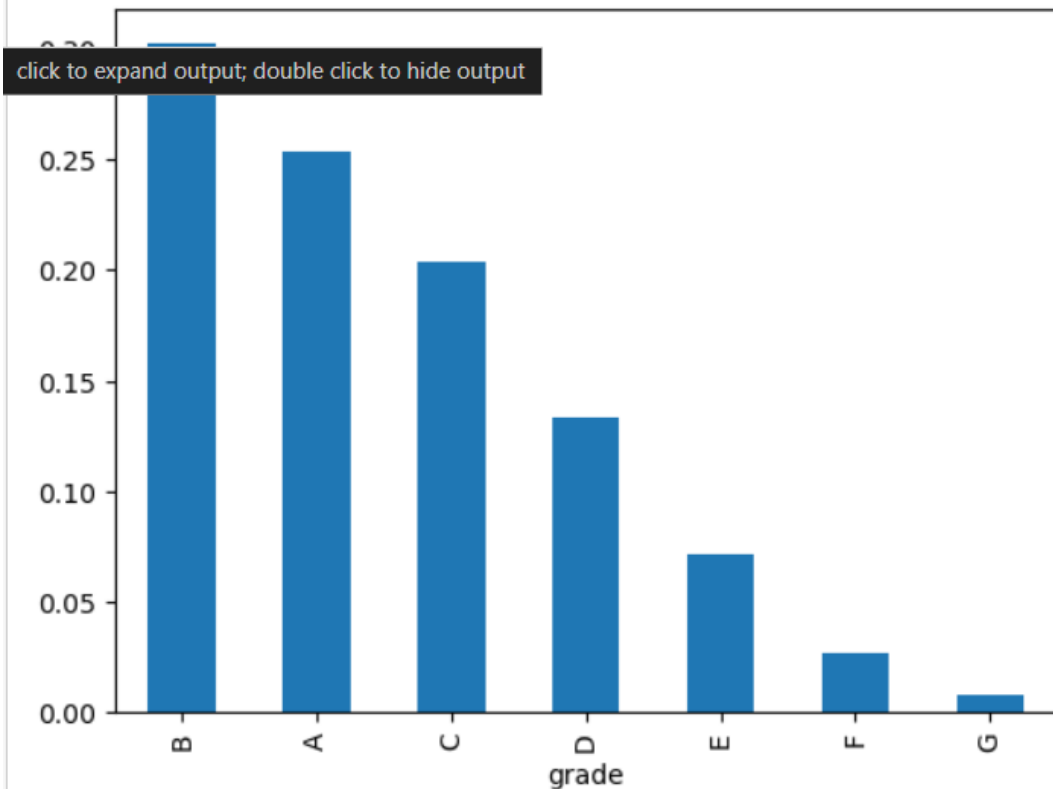
Bi-variate analysis

- Mostly charged off are having salary range of 40k-55k
- Charged off mostly having 9-15% of interest rate
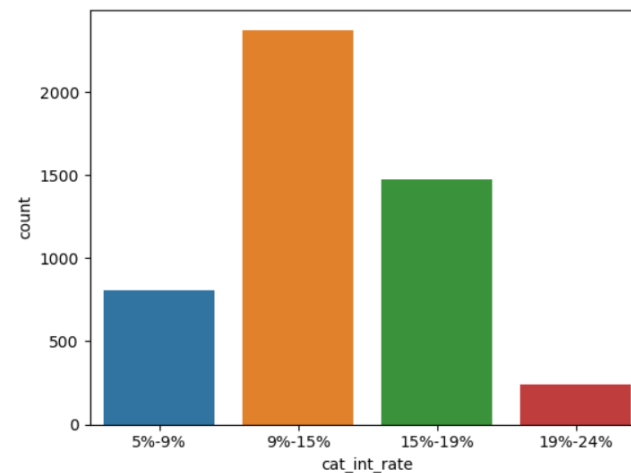


```
]: sns.countplot(x='cat_annual_income', data=df[df.loan_status == 'Charged Off'])
```

```
]: <Axes: xlabel='cat_annual_income', ylabel='count'>
```
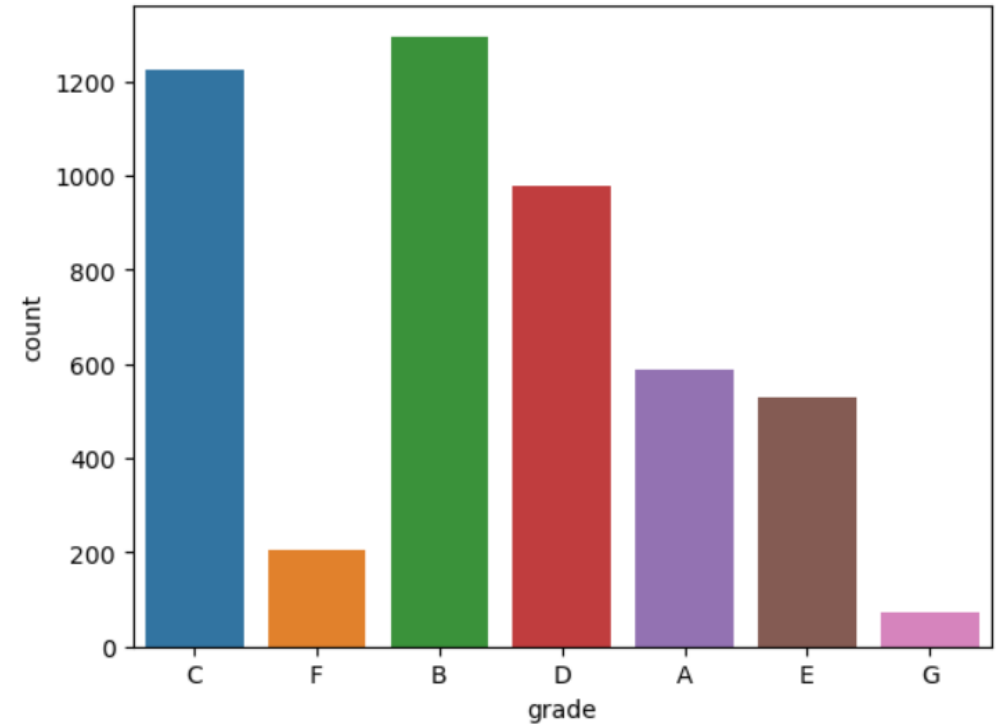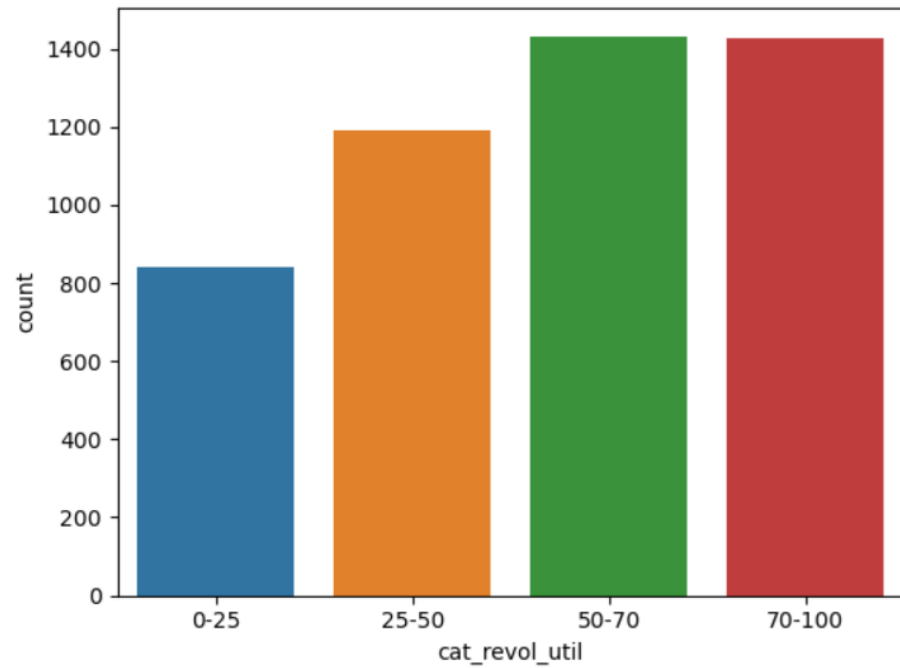


```
sns.countplot(x='cat_int_rate', data=df[df.loan_status == 'Charged Off'])
```
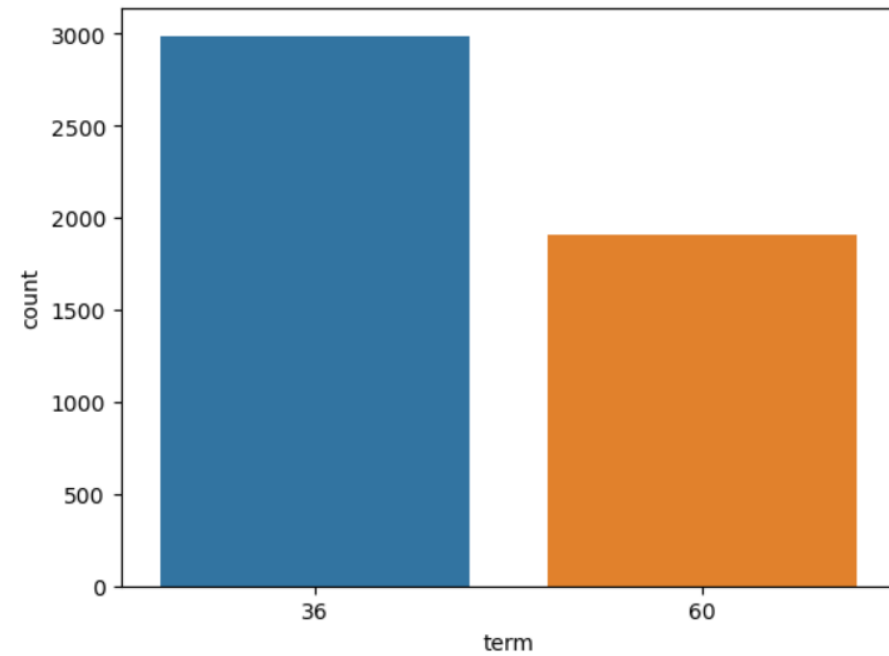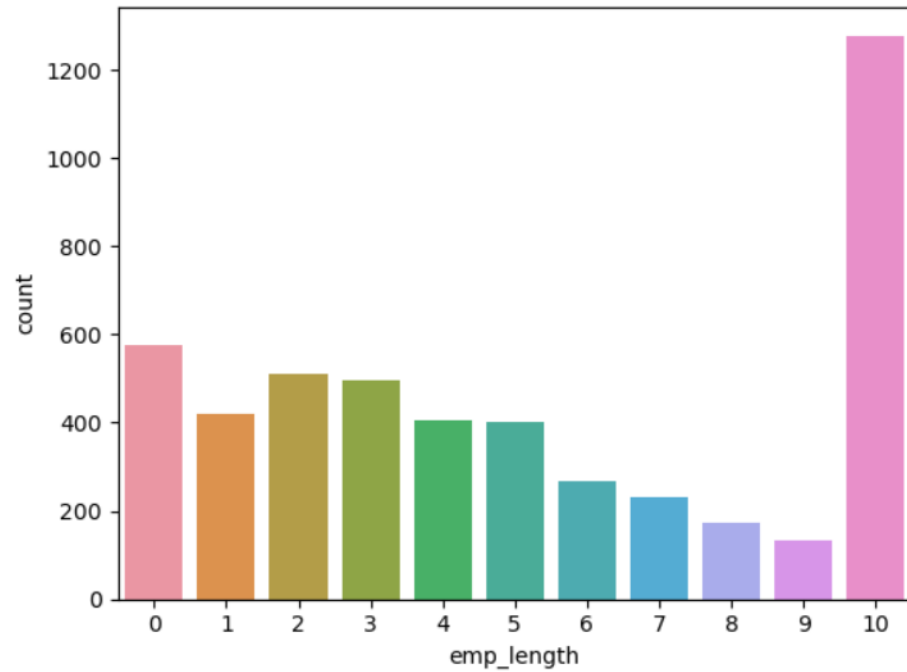
```
<Axes: xlabel='cat_int_rate', ylabel='count'>
```
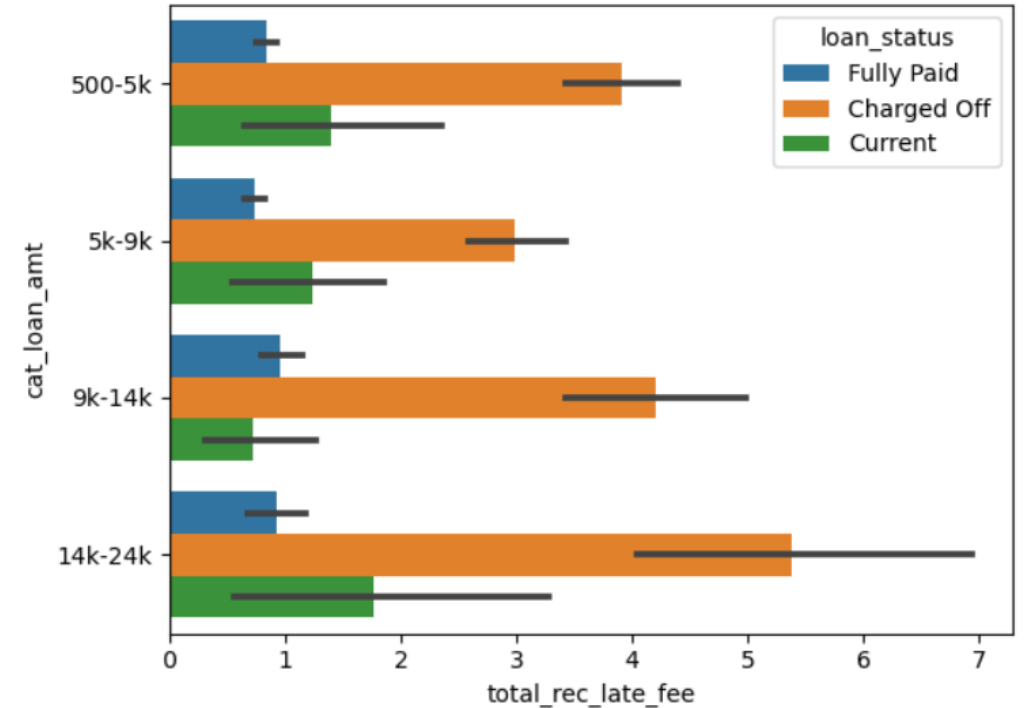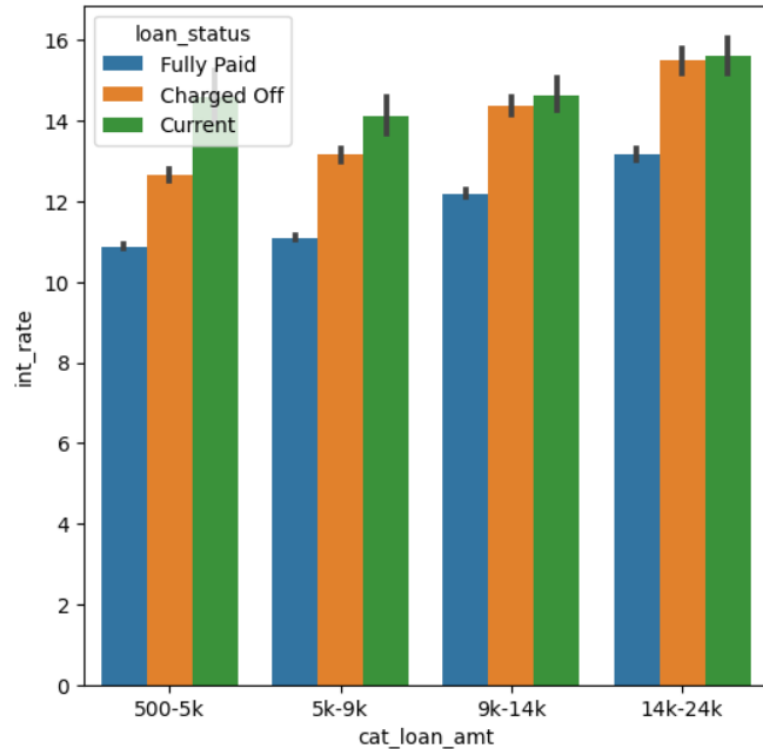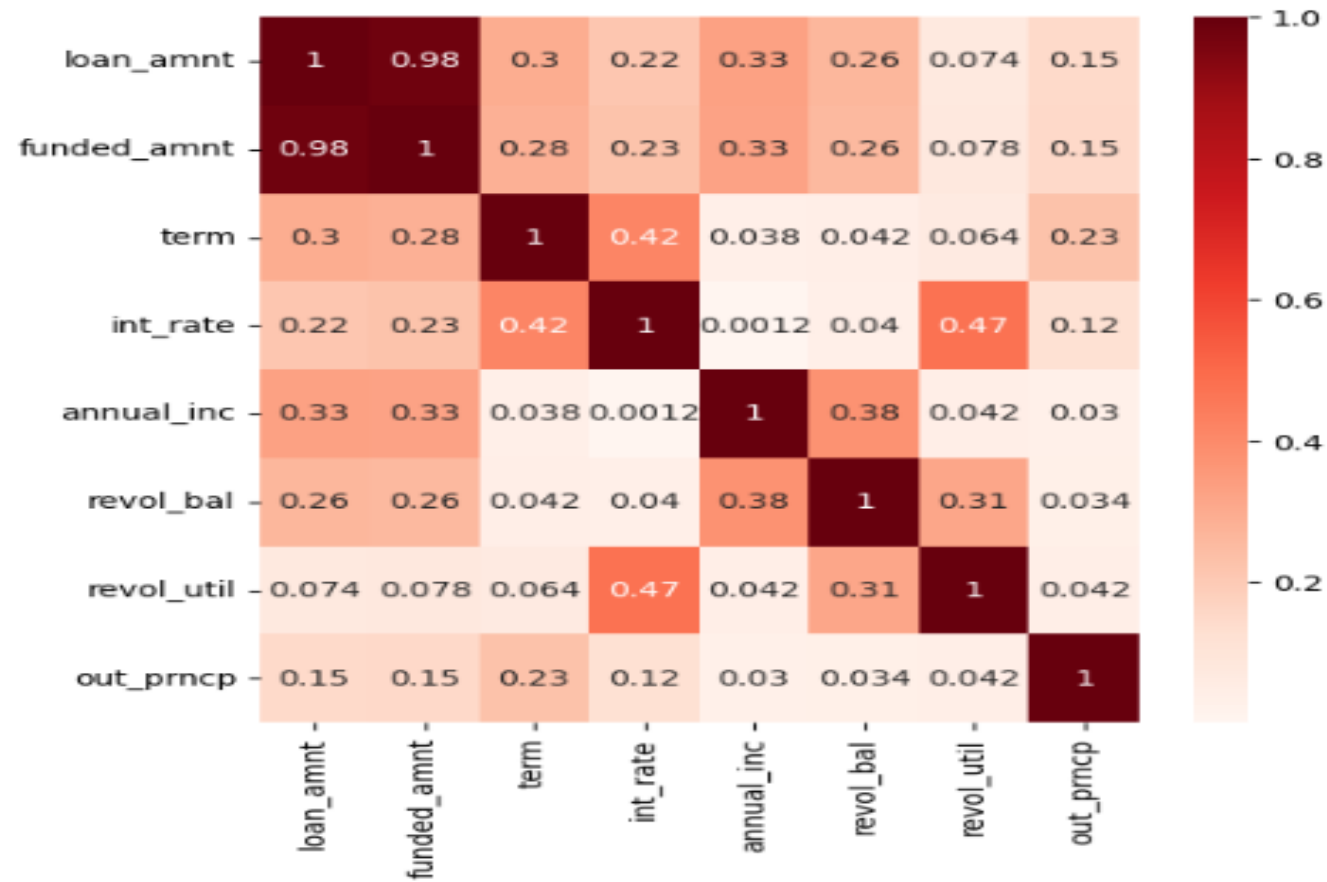
Bi-variate analysis

Bi-variate analysis
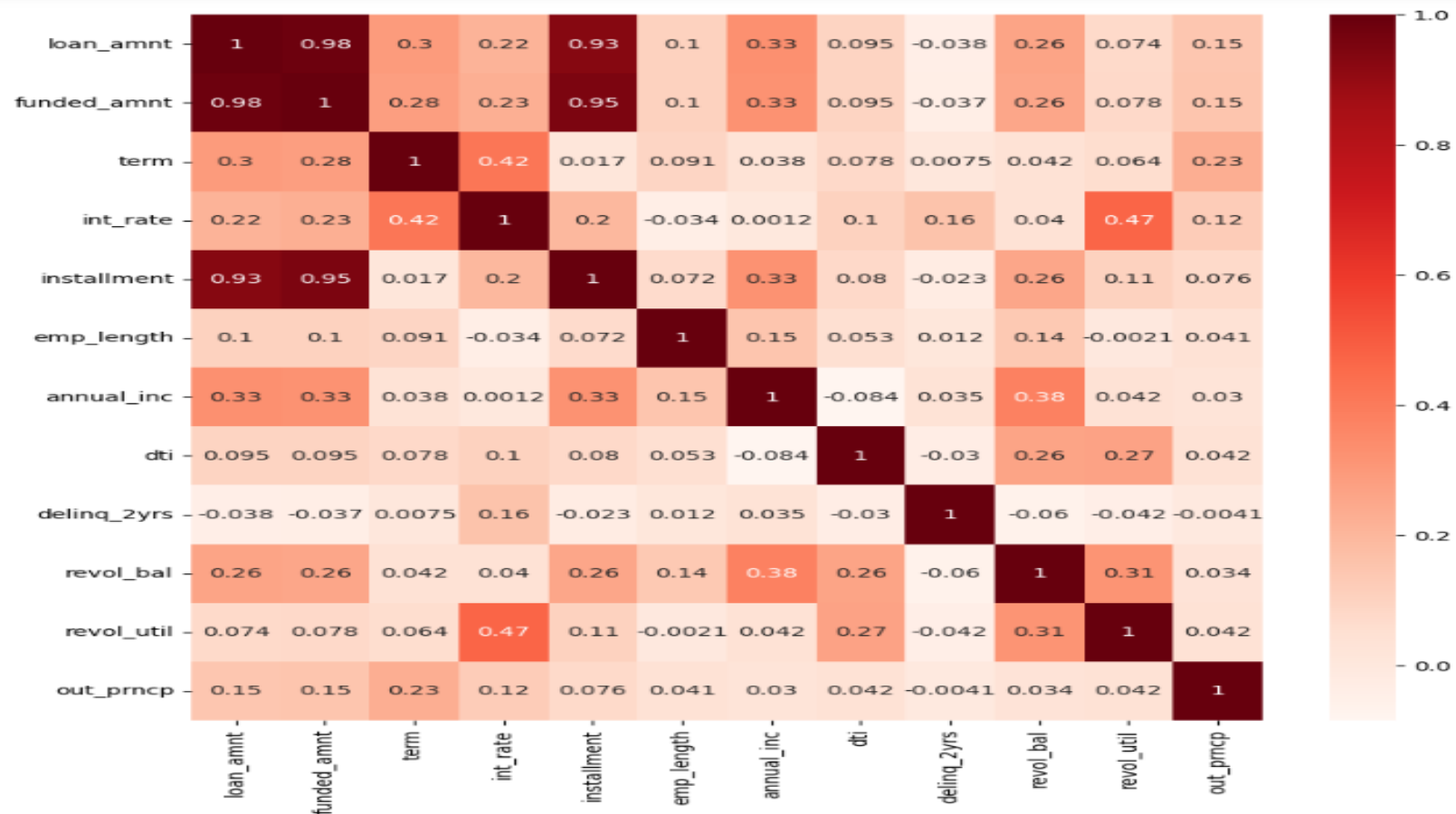
Multi variate analysis

Multi-variate analysis

Multi-variate analysis

Driver-Variables

Percentage of not verified customers is more than verified .	Emp_length is having more than 10+ years' experience
Salary between the 40k-55k	Revol_util is range between 70-100
B grade coustomers	Loan amount between 14k-24k