# Lending club case study.

# INTRODUCTION

Lending Club is the largest peer-peer marketplace connecting borrowers wth lenders. Borrowers apply through an online platform where they are assigned an internal score.
Lenders decide the following:

- whether to lend or not.
- the terms of loan such as interest rate, monthly installments, tenure etc.

# OBJECTIVE

1. To indentify variables which are strong indicators of defaulters.

2. Potential use of insights in approval or rejection of the application.

3. Identify Risky Loan Applicants

# DATA UNDERSTANDING

Types of variables:

- Customer related data
  (i.e., employment title, annual income etc.)
- Loan-related information
  (i.e., loan amount, interest rate, funded amount)
- Customer behavior
  (i.e fully paid or paid off)

# DATA CLEANING

```
# let's check the shape of the data set
raw_loan_data.shape
```

```
(39717, 111)
```

## DATA CLEANING

```
# identifying if there are any columns that is entirely null out of the given data set
# if true, we will not need those columns in the analysis ; we can drop them;

raw_loan_data = raw_loan_data[raw_loan_data.columns[(raw_loan_data.isnull().sum() != 39717)]]
```

```
# dropping of columns that are not relevant/helpful in the analysisdd
raw_loan_data.drop(columns=['chargeoff_within_12_mths','tax_liens','collections_12_mths_ex_med'],axis=1,inplace=True
```

```
raw_loan_data.shape
```

```
(39717, 51)
```

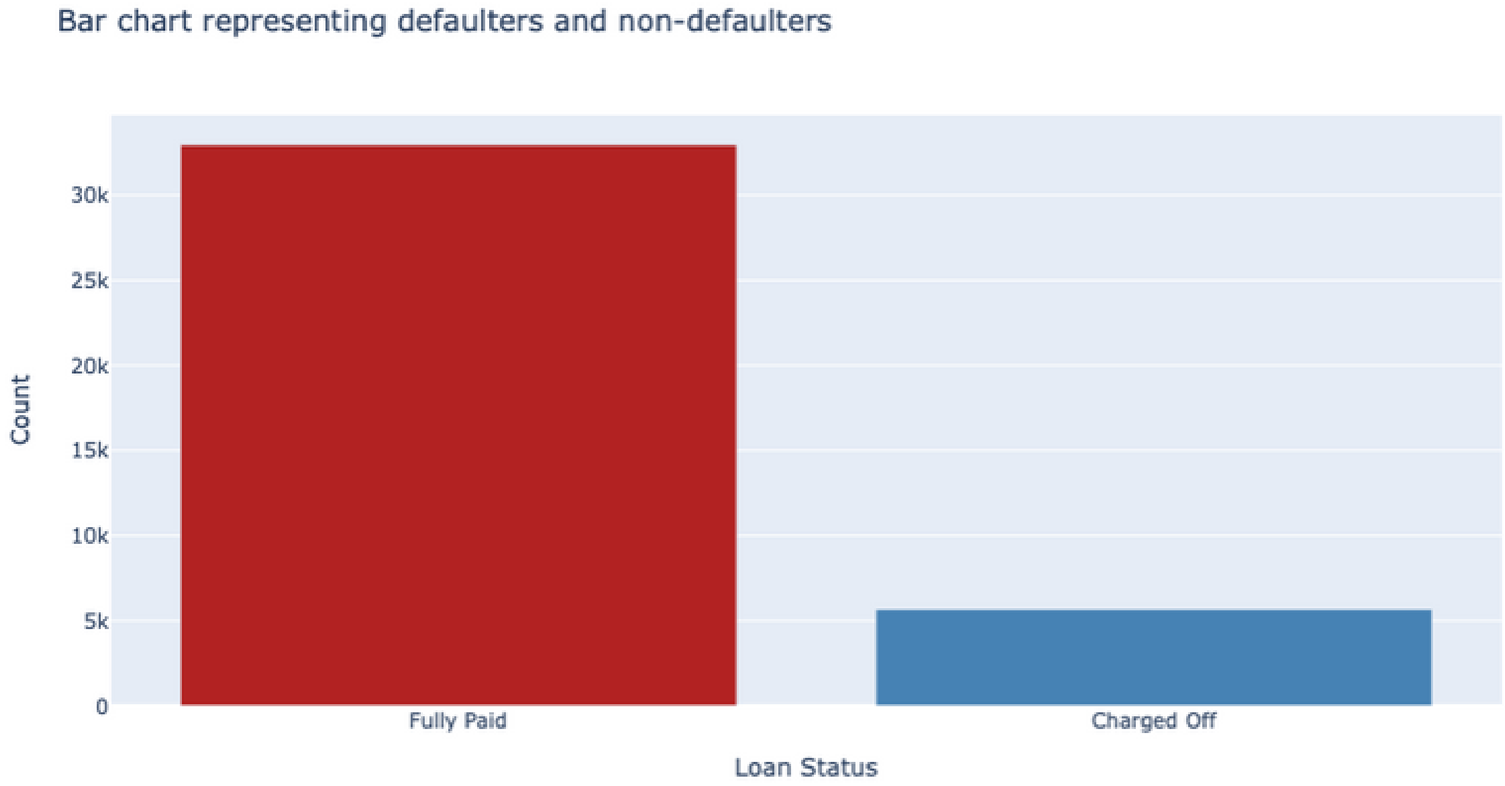
```
raw_loan_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39717 entries, 0 to 39716
Data columns (total 54 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   id                      39717 non-null  int64
 1   member_id               39717 non-null  int64
 2   loan_amnt               39717 non-null  int64
 3   funded_amnt             39717 non-null  int64
 4   funded_amnt_inv         39717 non-null  float64
 5   term                    39717 non-null  object
 6   int_rate                39717 non-null  object
 7   installment             39717 non-null  float64
 8   grade                   39717 non-null  object
 9   sub_grade               39717 non-null  object
 10  emp_title               37258 non-null  object
```

# DATA ANALYSIS

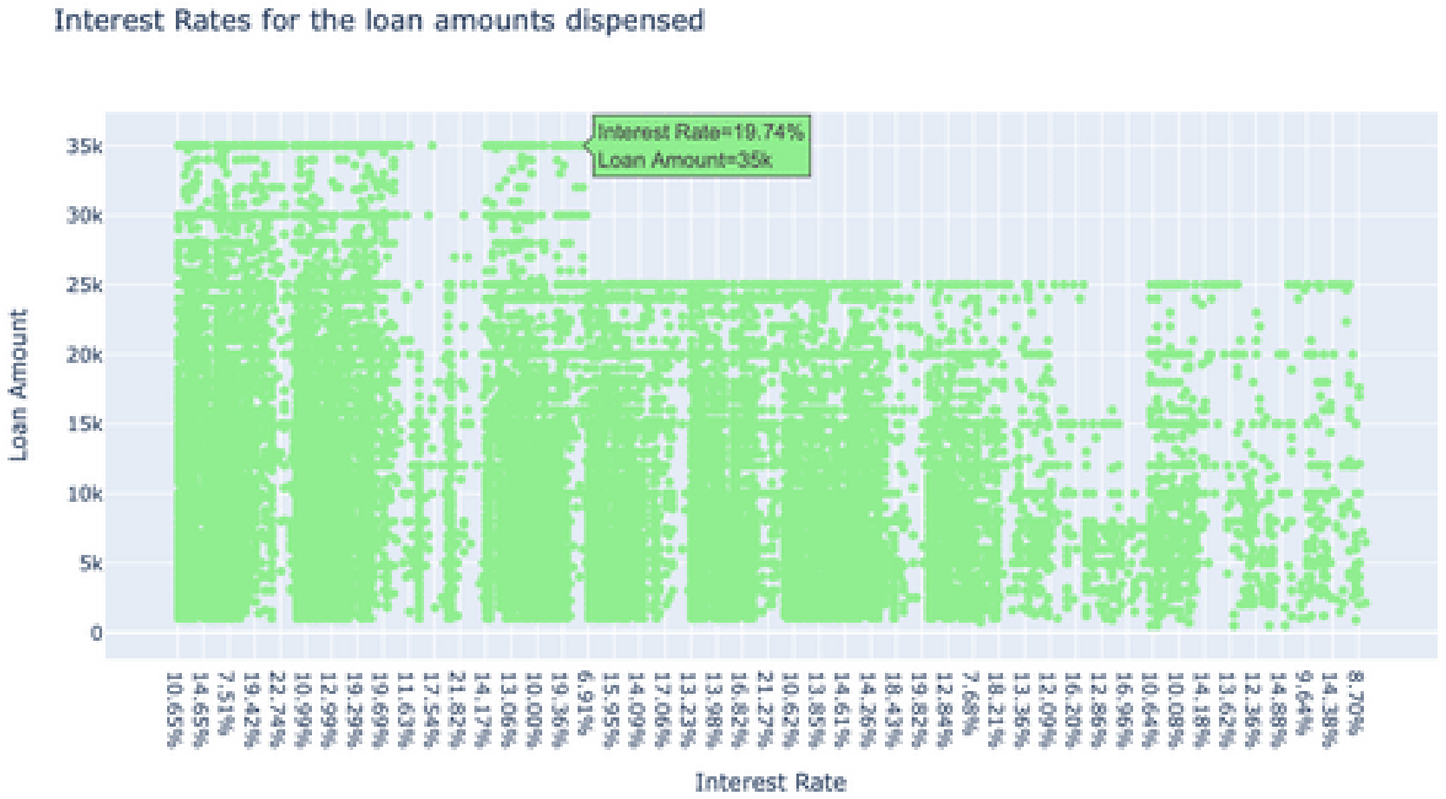
## A Univariate Analysis on Defaulters

- The plot clearly shows that around 14% of them are defaulters for a period of 2007-11
- Over the same period of time the amount of non-defaulters ave gradually been decreasing.



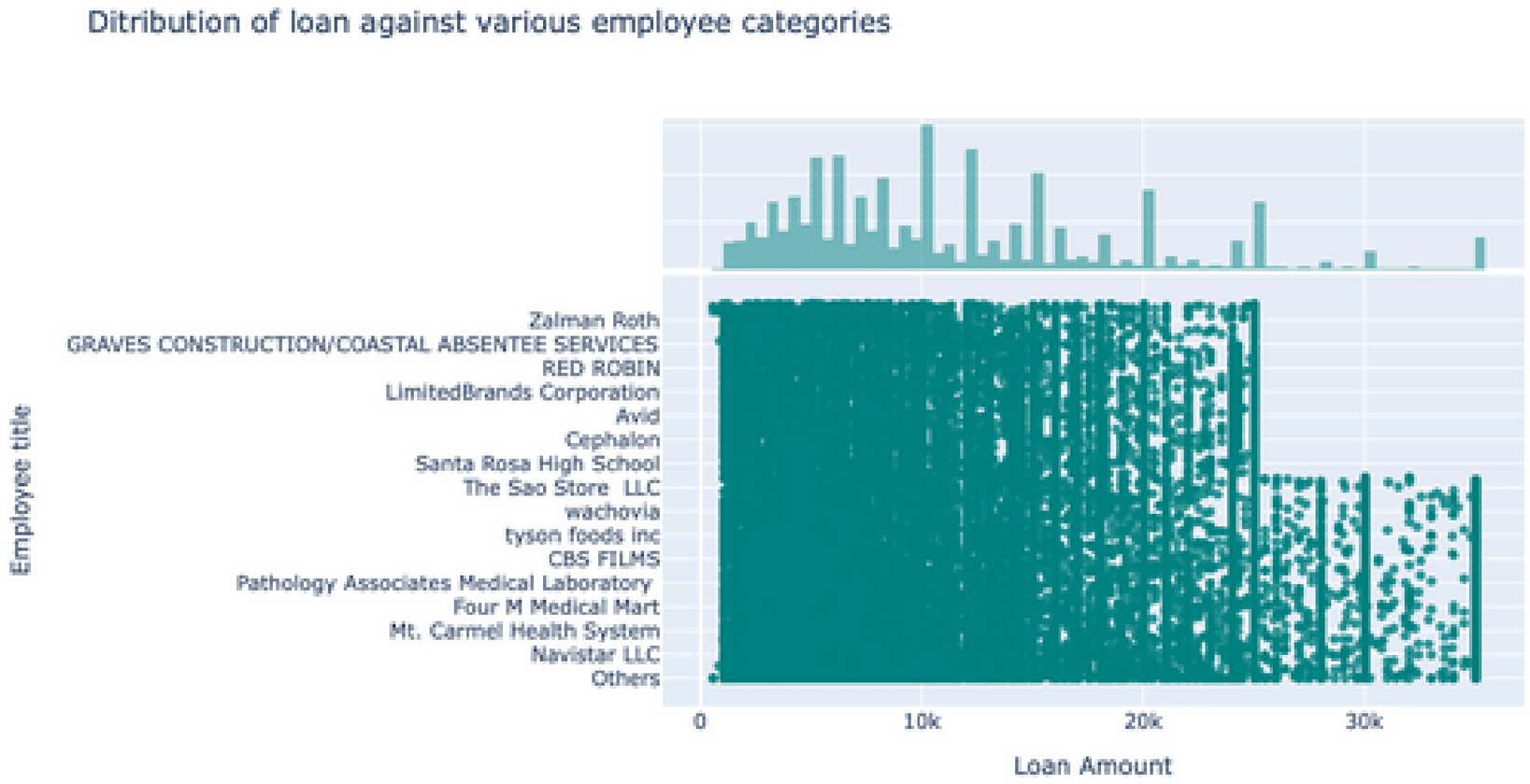
# A Univariate distribution of loan amount

# Bivariate Analysis:



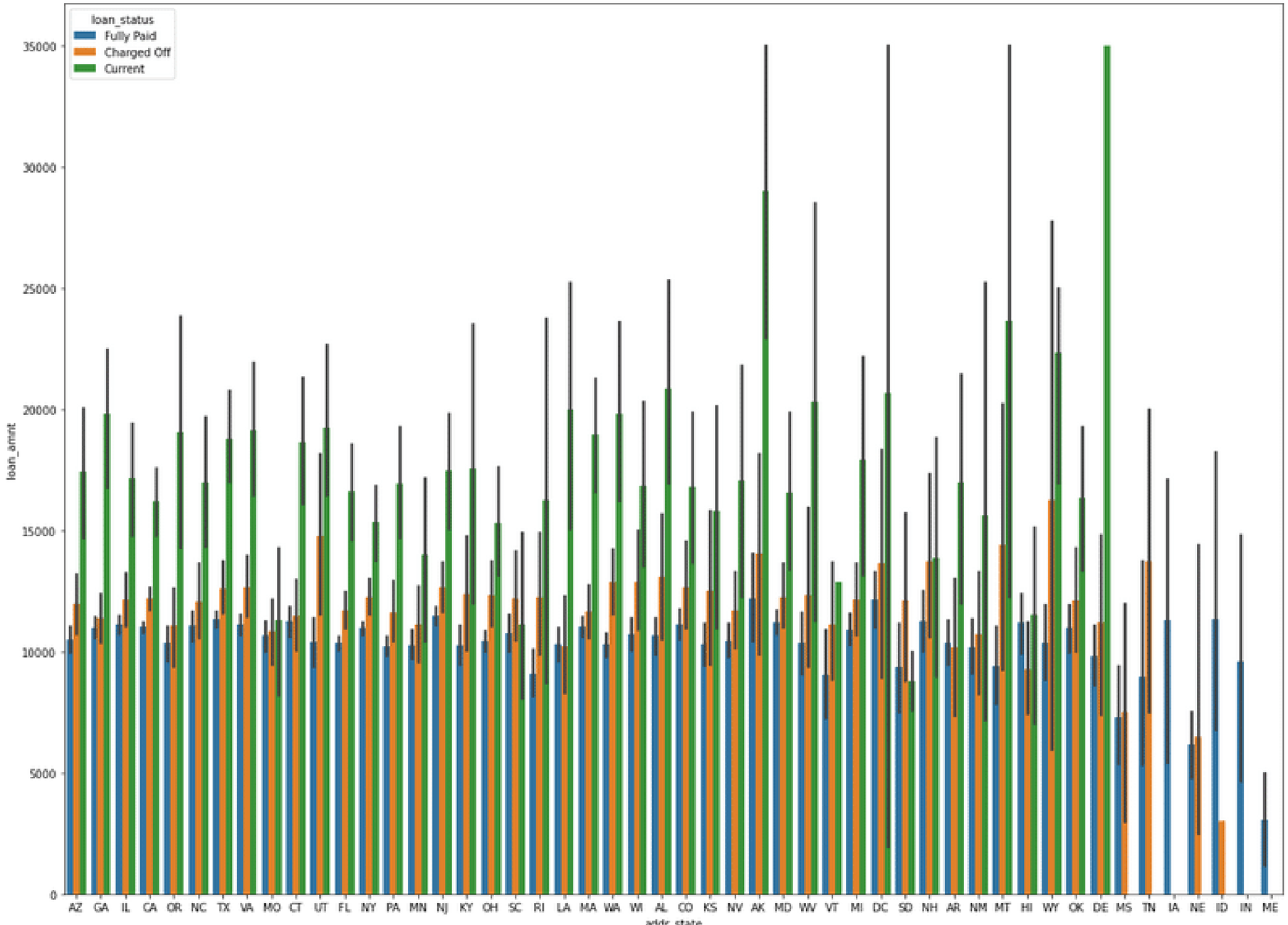

The above plot shows that the rate of interest increases as the loan amount increases.

# Bi-variate analysis on loan amount dispensed to employee titles




The analysis states the risk factor of issuing loan where it's analyzed based on the annual income of an Individual. There are people with average income lower than 50k taking loans of 25k or higher.

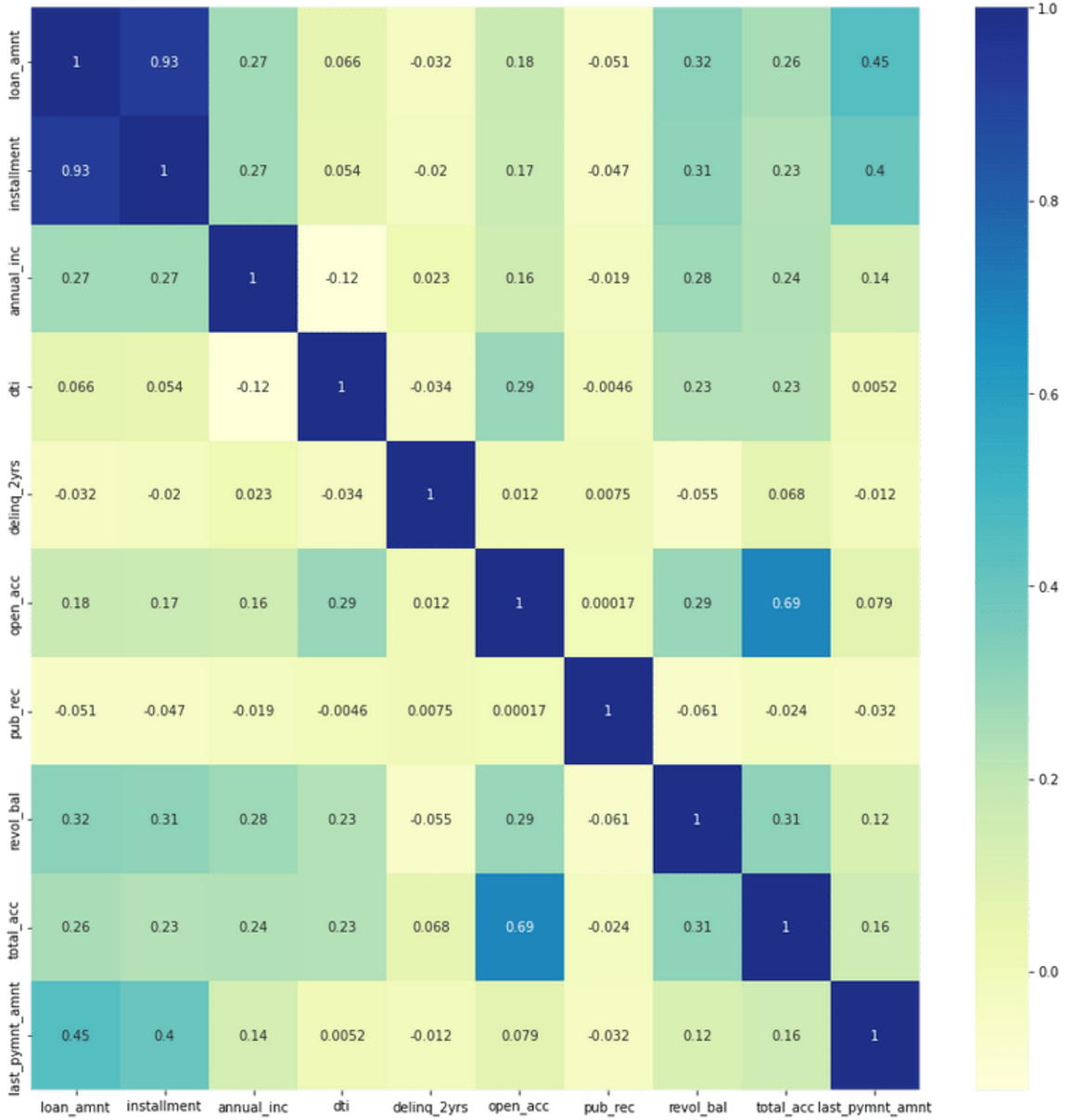
# Multivariate Analysis:




The state having the highest number of defaulters can be obtained from this plot.This information will benefit the company to a larger extent and also demands for more surevey to be conducted in such states before dispersal of any loan amount.

.

# Correlation Analysis:



# CONCLUSION

- From the previous slides and the analysis done one could easily determine whether the loan should be given or not based on the factors stated in the charts.
- Higher the loan amount borrowed,higher is the rate of interest. This benefits the company as it can minimize the chances of credit losses.
- People with an income of Rs.50000 on an average are taking loans of Rs.25000 or more which can be a potential threat to the company and can also cause credit loss.
- The state having highest number of defaulters was also obtained. This information will benefit the company to a larger extent and also demands for more survey to be conducted in such states before dispersal of any loan amount.
- Higher loan amount for longer term have lower grades.