

LENDING CLUB GROUP CASE STUDY

SUBMISSION

Name: Arun Prasad
Prashant Upadhyay

Abstract

- Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- The overall strategy is to do exploratory data analysis (EDA) on the data available with Lending Club and find out the defaulters risk to the company before the loan issue.
- Lending Club to issue Loan based on charged off % on basis of various parameters

Problem Solving Methodology

Business Understanding

- Define the problem

1

Get the data from source

2

cleaning data

3

univariate and bivariate
analysis and plotting

4

Derived metrics addition

5

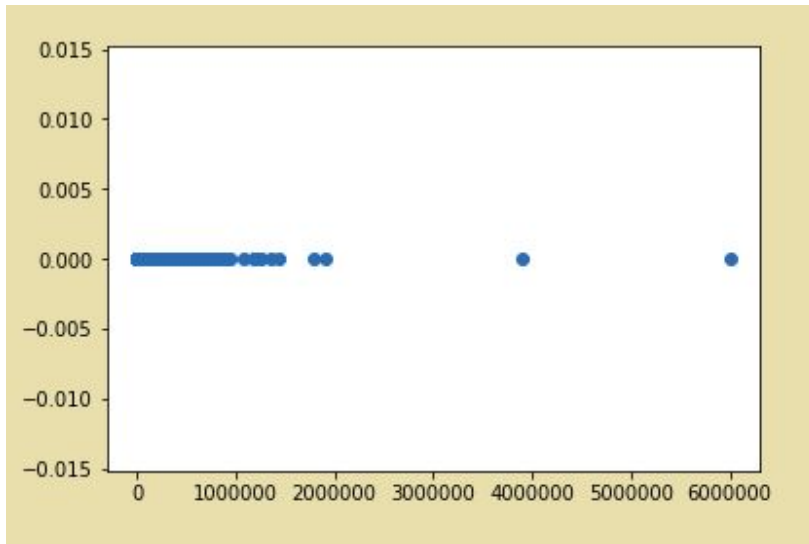
analysing for charged off %
and final conclusion

6

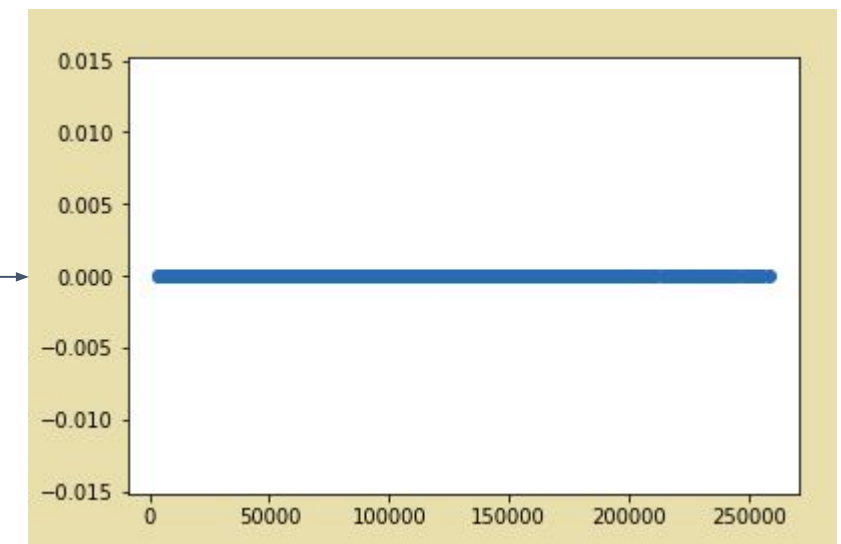
Data Cleaning

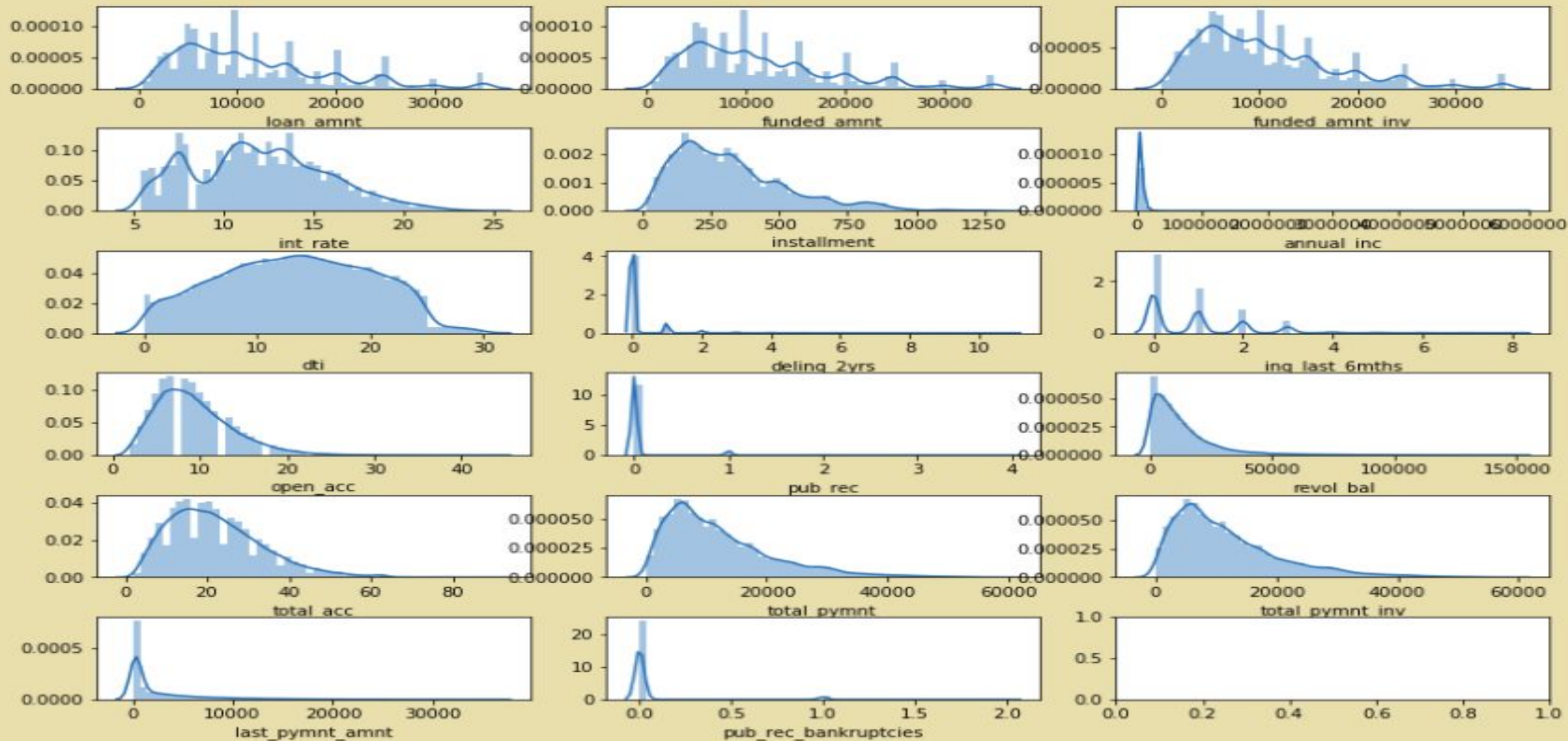
Data cleaning is done on the basis of below points

- Columns having all NaN value
- Columns having only single value
- Columns having 95% same values for continuous data
- Columns having outliers
- Rows having NaN values which are only 2% of total count



Outlier removal on the basis
on annual inc

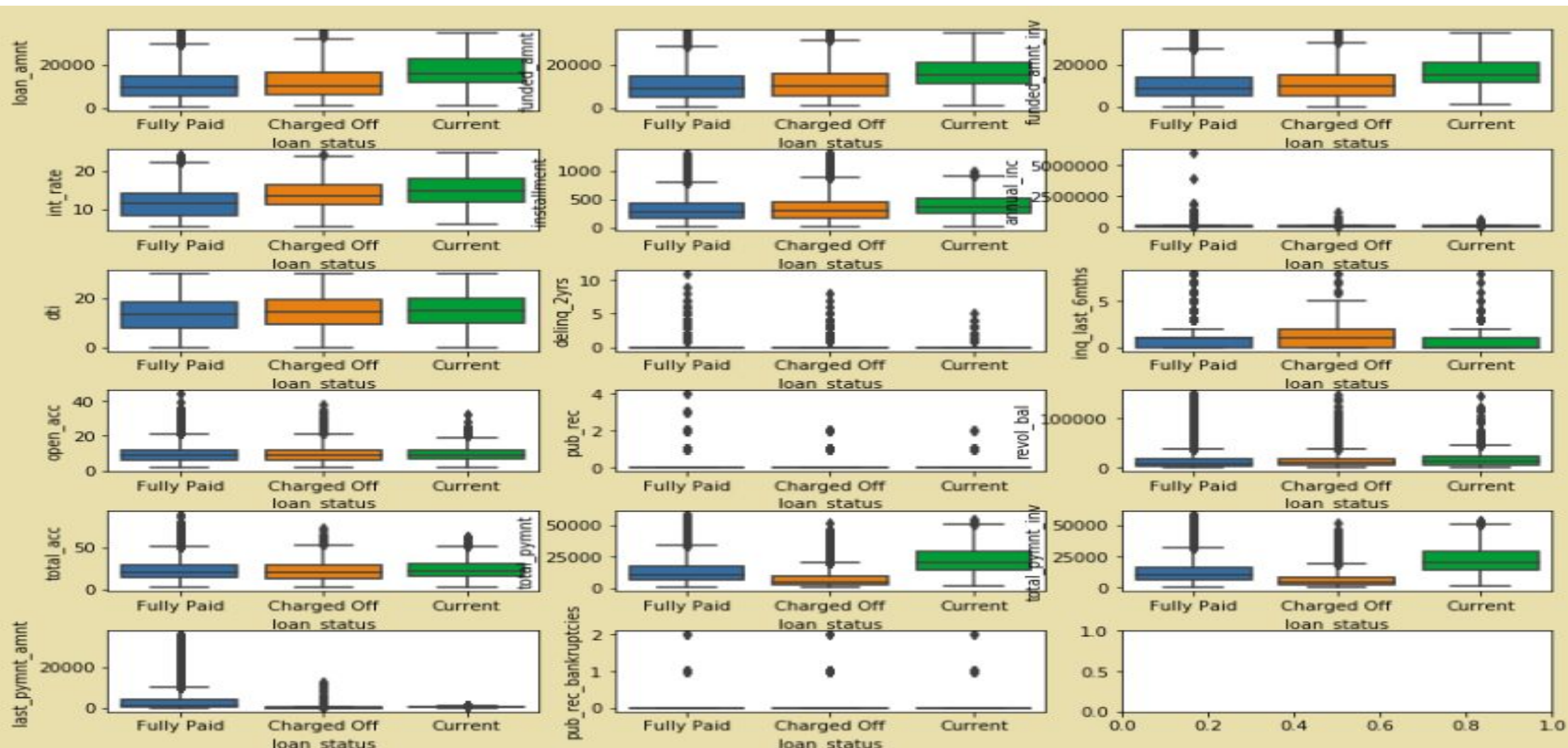






Univariate analysis

- In **int_rate_range** of 15 & above, more people are likely to default.
- Most of the loans were issued in the month of December.
- Borrowers from **states** like NV, TN, SD, AK, FL are more likely to default.
- Most of the people who borrowed the loan for *wedding* or *car* or *major purposes* fully paid back the loan whereas who borrowed the loan for *small_business* or *educational* or *renewable energy* are more likely to default.
- People who mortgaged their home are more likely to repay the loan than who own house or living on rent (**home_ownership**)
- Lower **Grade** people belonging to F or G are mostly defaulters and lenders should avoid them giving loans.

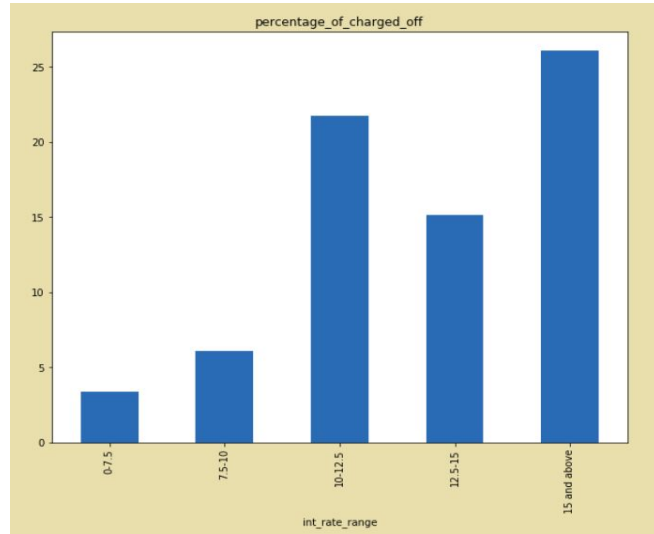




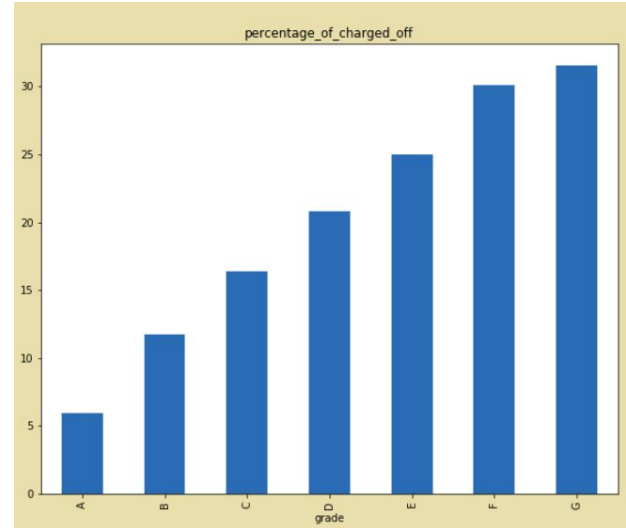
1. **loan_amnt, funded_amnt, funded_amnt_inv** are **highly correlated** to other and hence, for analysis any one of them can be used.
2. **total_acc and open_acc** are highly correlated as opening an account will also lead to increase in the count of total account.
3. **pub_rec and pub_rec_bankruptcies** are the variables that are highly correlated as entry of bankruptcies in public record will also lead to increase in public record.

Bivariate Analysis w.r.t Charged-Off %age

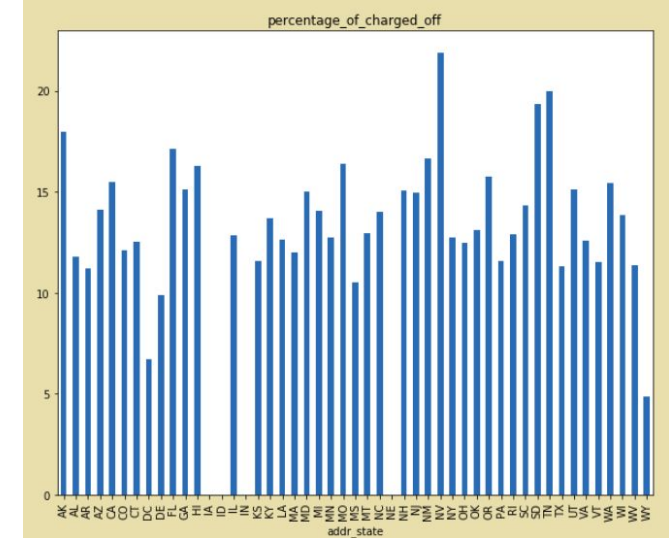
interest rate



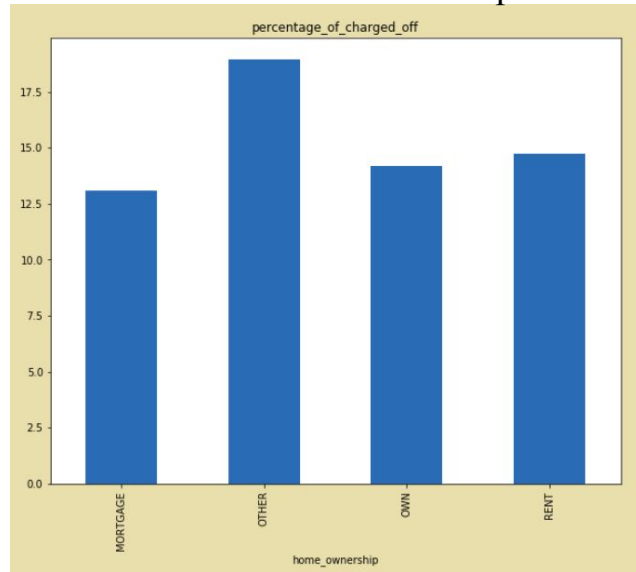
Grade



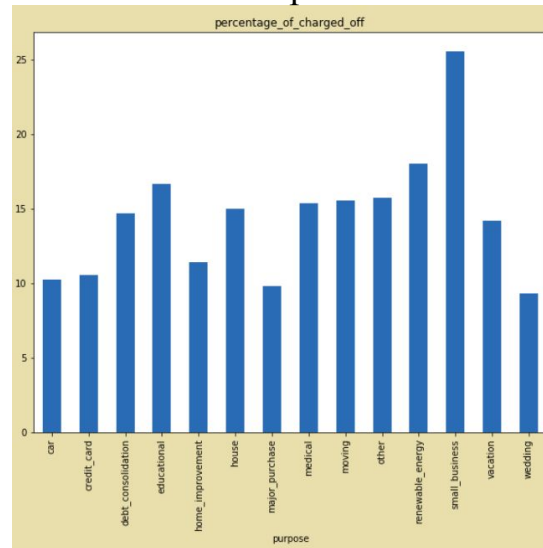
Addr State



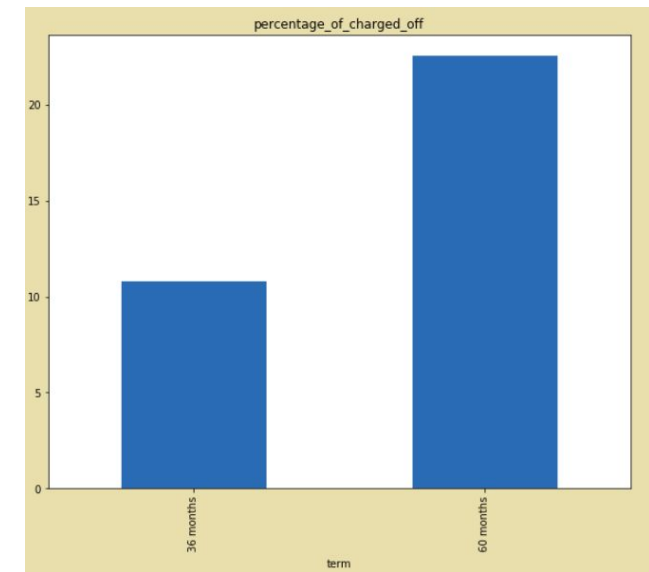
Home Ownership



Purpose of loan



Term



Conclusion

The features that can be useful in analyzing whether an loan applicant can be a defaulter or not are

1. grade
2. int_rate_range
3. addr_state
4. purpose
5. term
6. emp_length

- A, B & C grade borrowers are safe options for lending loan
- Higher Interest rates are more riskier for lending loan
- Lending loan to people from the following states NV, TN, SD, AK, FL can be riskier.
- Loan borrowed for the purpose of small business, renewable energy and educational loans are riskier.
- Loan taken for longer term are more likely to get defaulter than shorter term.
- Lending loans to borrowers whose employment length is not known is riskier.