

Introduction

About Dataset:

- ❑ The given data contains information about past loan application and whether the applicant 'defaulted' or not
- ❑ The database has details of around 40,000 applicants with 100+ features

Objective:

- ❑ Understanding the driving factors behind loan default

Data Cleaning and Manipulation

- ❑ Total no. of features present in dataset = 111
- ❑ Features with all NaN values = 54
- ❑ Following columns contains most number of NaN values after removing columns with all NaN values:

```
In [14]: #Check for missing data in remaining columns
loan.isna().sum().sort_values(ascending=False).head(10)
```

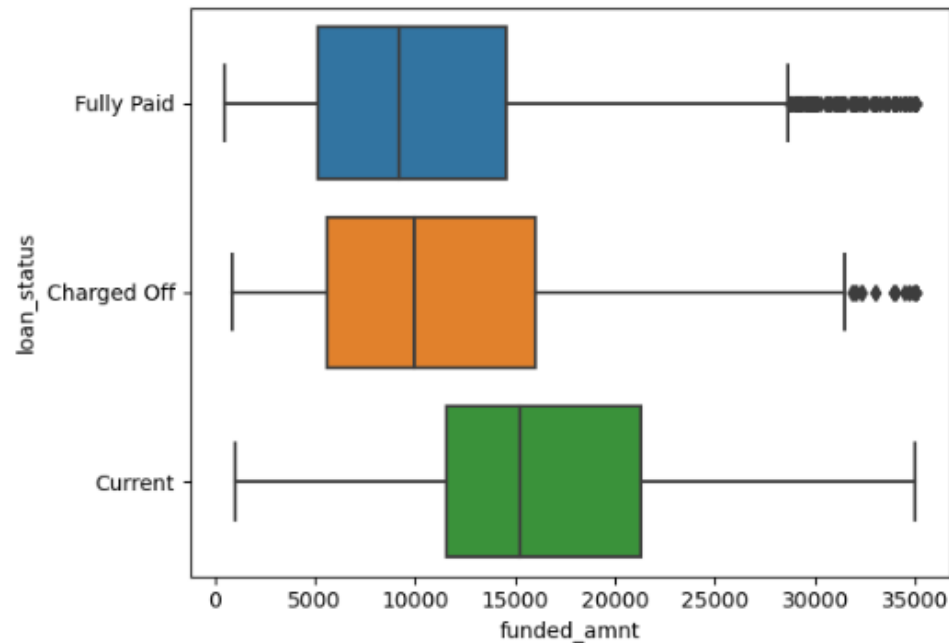
```
Out[14]: next_pymnt_d          38577
mths_since_last_record      36931
mths_since_last_delinq      25682
desc                        12940
emp_title                   2459
emp_length                  1075
pub_rec_bankruptcies        697
last_pymnt_d                71
chargeoff_within_12_mths    56
collections_12_mths_ex_med  56
dtype: int64
```

- ❑ Further analysis results in removal of top 4 columns mentioned above
- ❑ 'emp_title' column is converted into categorical variable to determine whether employee title is present or not
- ❑ For remaining columns with NaN values, applicants with NaN values are dropped to make the dataset free of NaN values
- ❑ Duplicates are removed using 'id' and 'member_id' columns

Data Analysis - Univariate & Bivariate Analysis

1. funded_amnt v/s loan_status

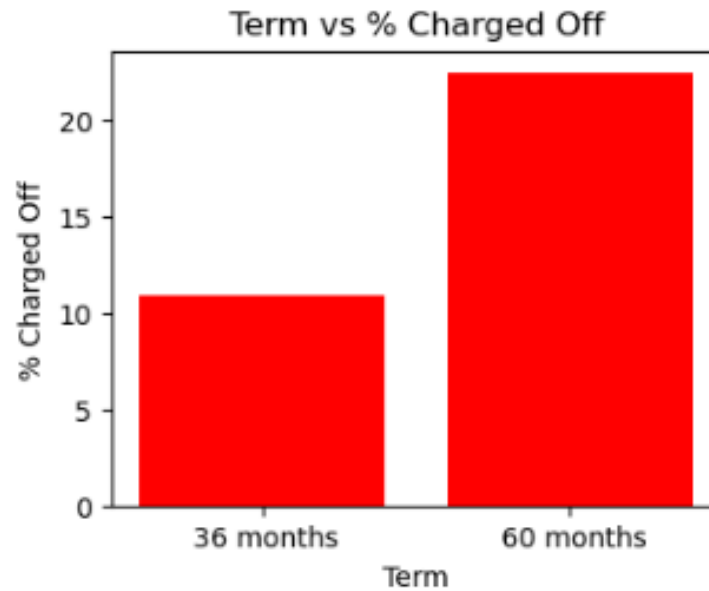
- *funded_amnt*: The total amount committed to that loan at that point in time
- Analysis:



- Conclusion: The 25th percentile, mean and 75th percentile of funded amount is similar for both "Fully Paid" and "Charged Off" loan. This indicates that the funded amount doesnot affect loan status

2. term v/s loan_status

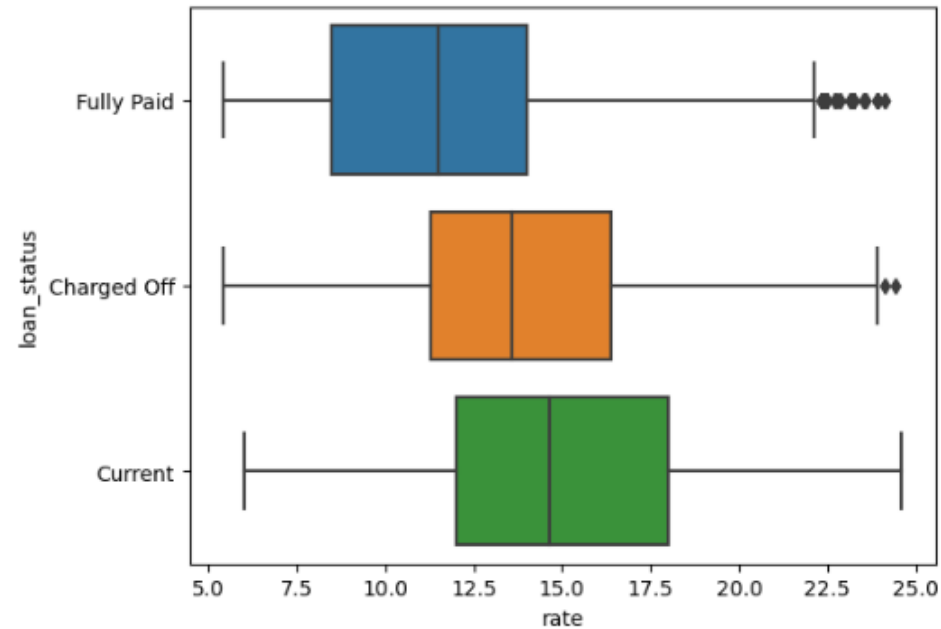
- *term*: Tenure of the loan either 36 months or 60 months
- Analysis:



- Conclusion: Loans with high duration has higher chance of default

3. int_rate v/s loan_status

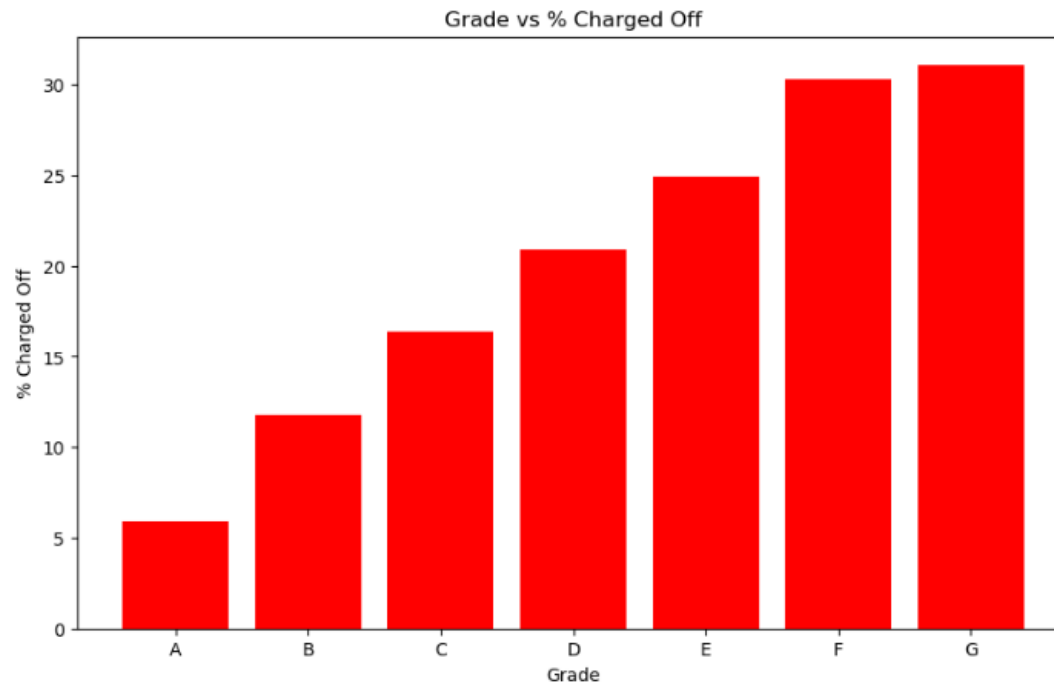
- *int_rate*: Interest rate on the loan
- Analysis:



- Conclusion: Default loans have higher mean interest rate as compared to loans fully paid

4. grade v/s loan_status

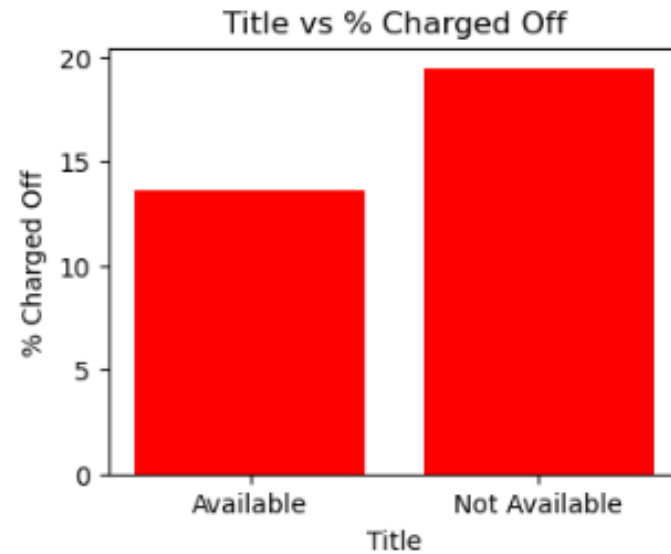
- *grade*: Grade of loan assigned by LC
- Analysis:



- Conclusion: Lower grade loans have higher chances of default as compared to better grade loans

5. emp_title v/s loan_status

- *emp_title*: Job title of the borrower mentioned at the time of loan application
- Analysis:

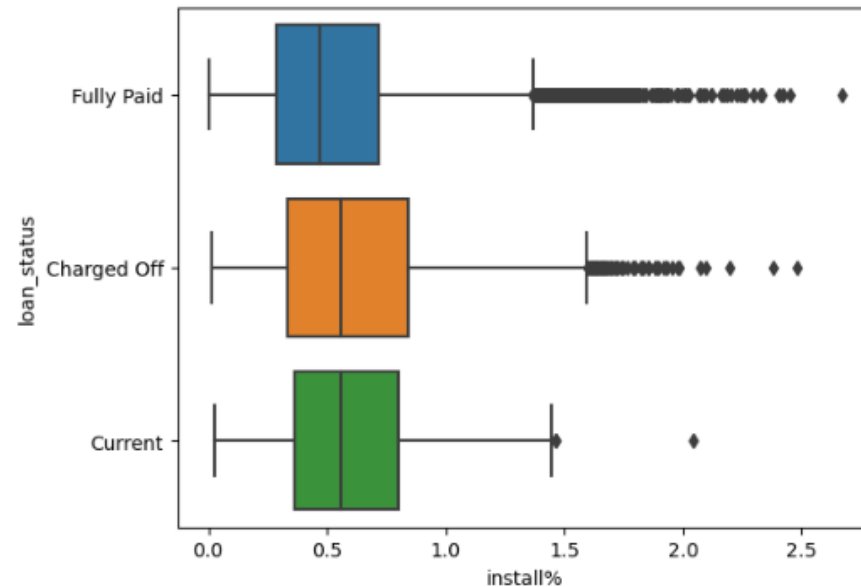


- Conclusion: Loan has higher chances of default when employee title is not furnished during loan application

6. installment & annual_inc v/s loan_status

- *installment*: Monthly payment to be paid
- *annual_inc*: Annual income of the borrower

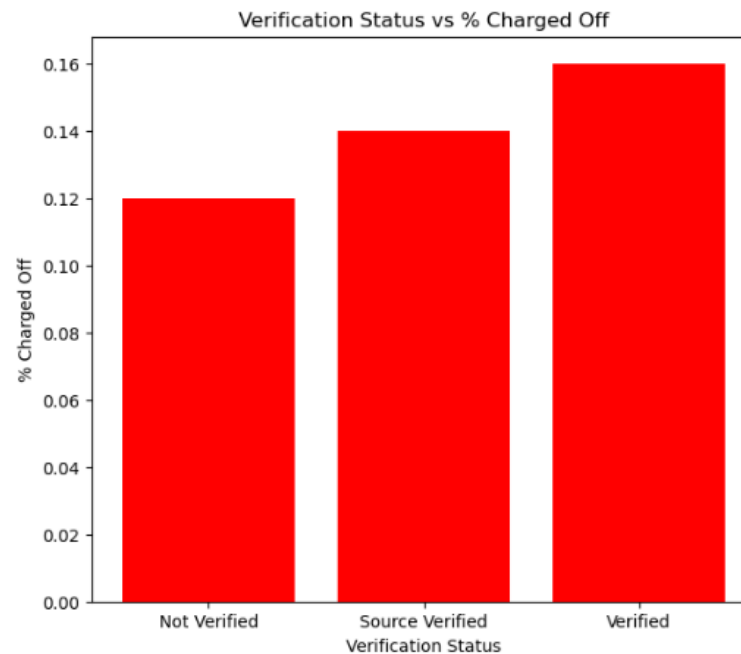
- Analysis:



- Conclusion: Borrowers who defaulted had a higher mean % of total income to be paid as installment

7. verification_status v/s loan_status

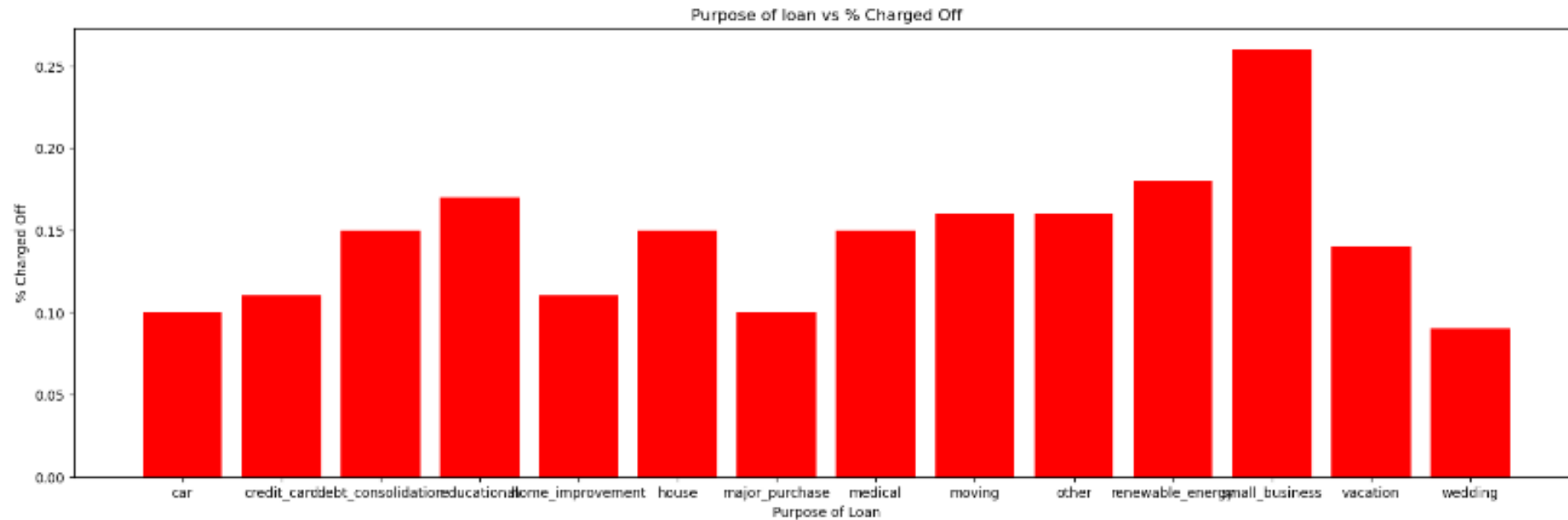
- *verification_status*: Indicates if income of borrower was verified
- Analysis:



- Conclusion: Counterintuitively, individuals with verified income has higher chances of default

8. purpose v/s loan_status

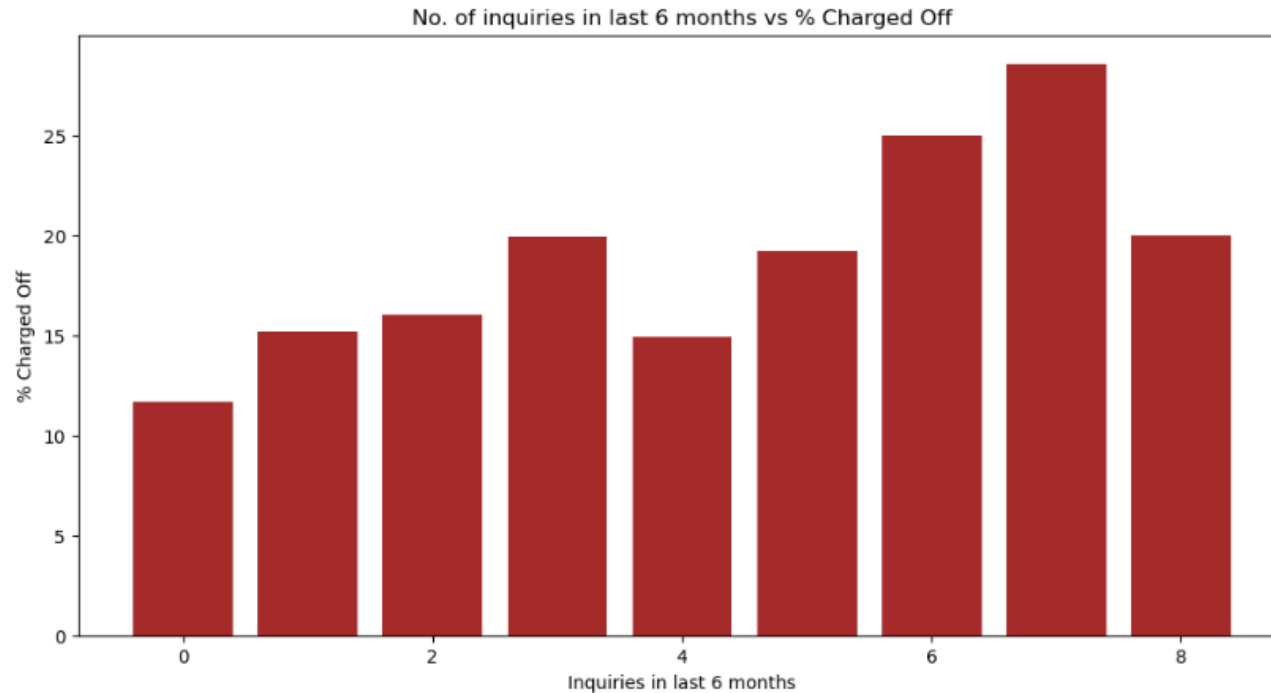
- *purpose: Purpose of requesting for loan*
- Analysis:



- Conclusion: Individuals borrowing for small business tends to default more

9. inq_last_6mths v/s loan_status

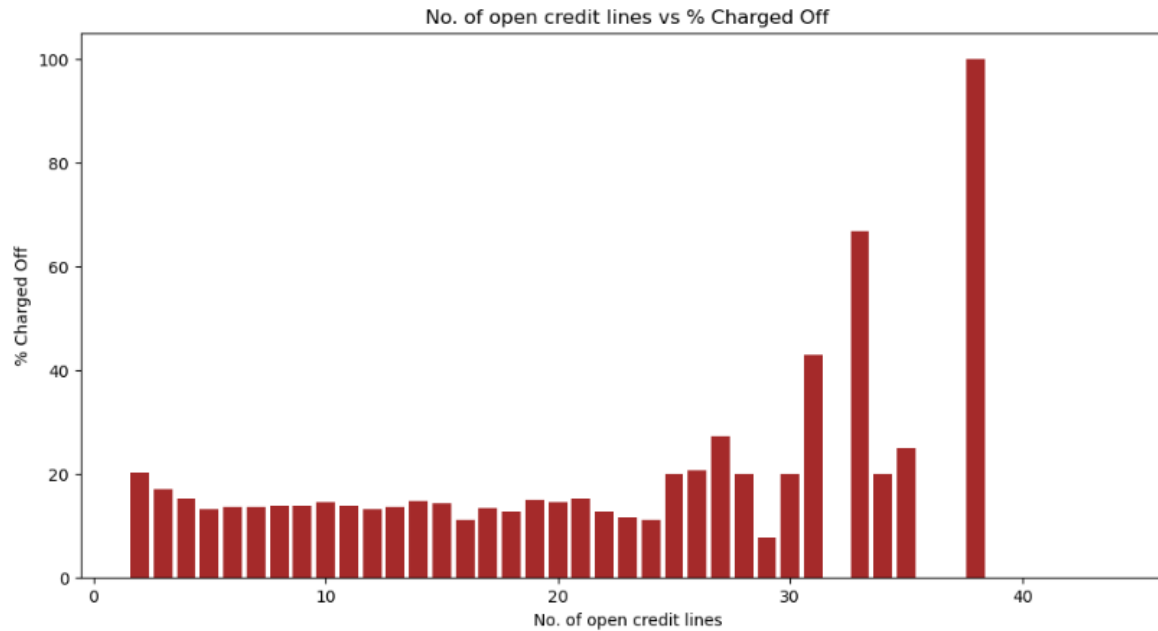
- *inq_last_6mths*: No. of enquiries in the last 6 months
- Analysis:



- Conclusion: In general, the chances of default increases with increasing no. of credit enquiries in last six months

10. open_acc v/s loan_status

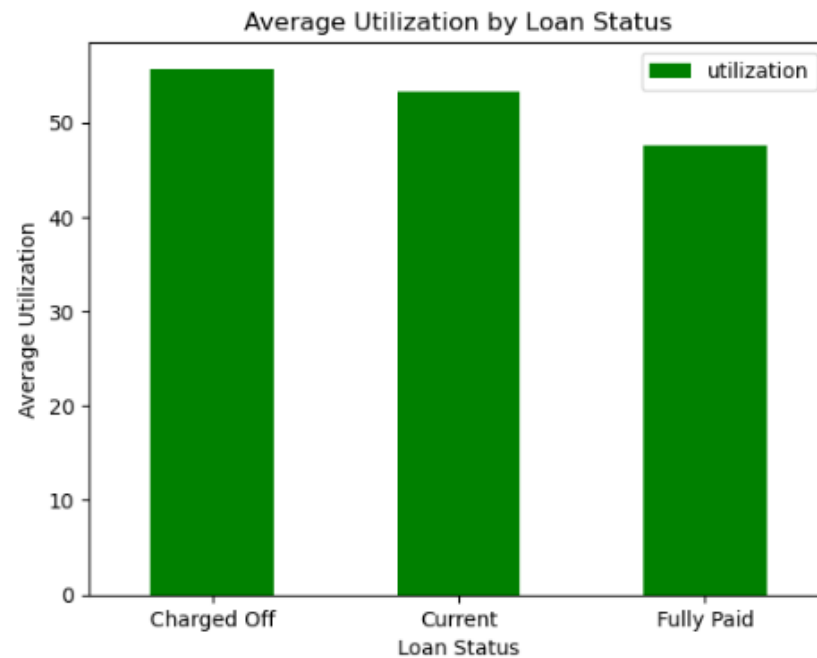
- *open_acc*: No. of open credit line for the borrower
- Analysis:



- Conclusion: The chances of default reduces as no. of open credit lines increases from 1 to 24. The trend reverses beyond 24 credit lines

11. revol_util v/s loan_status

- *revol_util*: Utilization of revolving line of credit
- Analysis:



- Conclusion: Borrowers with higher average utilization tends to default more

Conclusion

- Important attributes that influences tendencies of default are:
 - 'term': Tenure of the loan
 - 'int_rate': Interest rate of the loan
 - 'grade': Grade of loan assigned by LC
 - 'emp_title': Borrower with job title on loan application tends to default less
 - 'installment/annual_inc ': Borrower with high installment/annual income ratio are more likely to default
 - 'purpose': Individuals with loan for business requirement tends to default more
 - 'inq_last_6mths': In general, the chances of default increases with increasing no. of credit enquiries in last six months
 - 'open_acc': The chances of default reduces as no. of open credit lines increases from 1 to 24. The trend reverses beyond 24 credit lines
 - 'revol_util': Borrowers with higher average utilization tends to default more