Lending Club Case Study

Submitted by –

Rajesh Kumar Singh

Swarnali Sen

Exploratory Data Analysis

Data Sourcing

- Understand the data source
- Import the data for analysis

Data Cleaning

- Fix rows and columns
- Fix missing values
- Standardise values
- Fix invalid values
- Filter data

Univariate Analysis

- Separate plots for each variable
- Bar plots, Pie plots or Histograms for categorical variables
- Box plots or Violin plots for numerical variables

Segmented Univariate Analysis

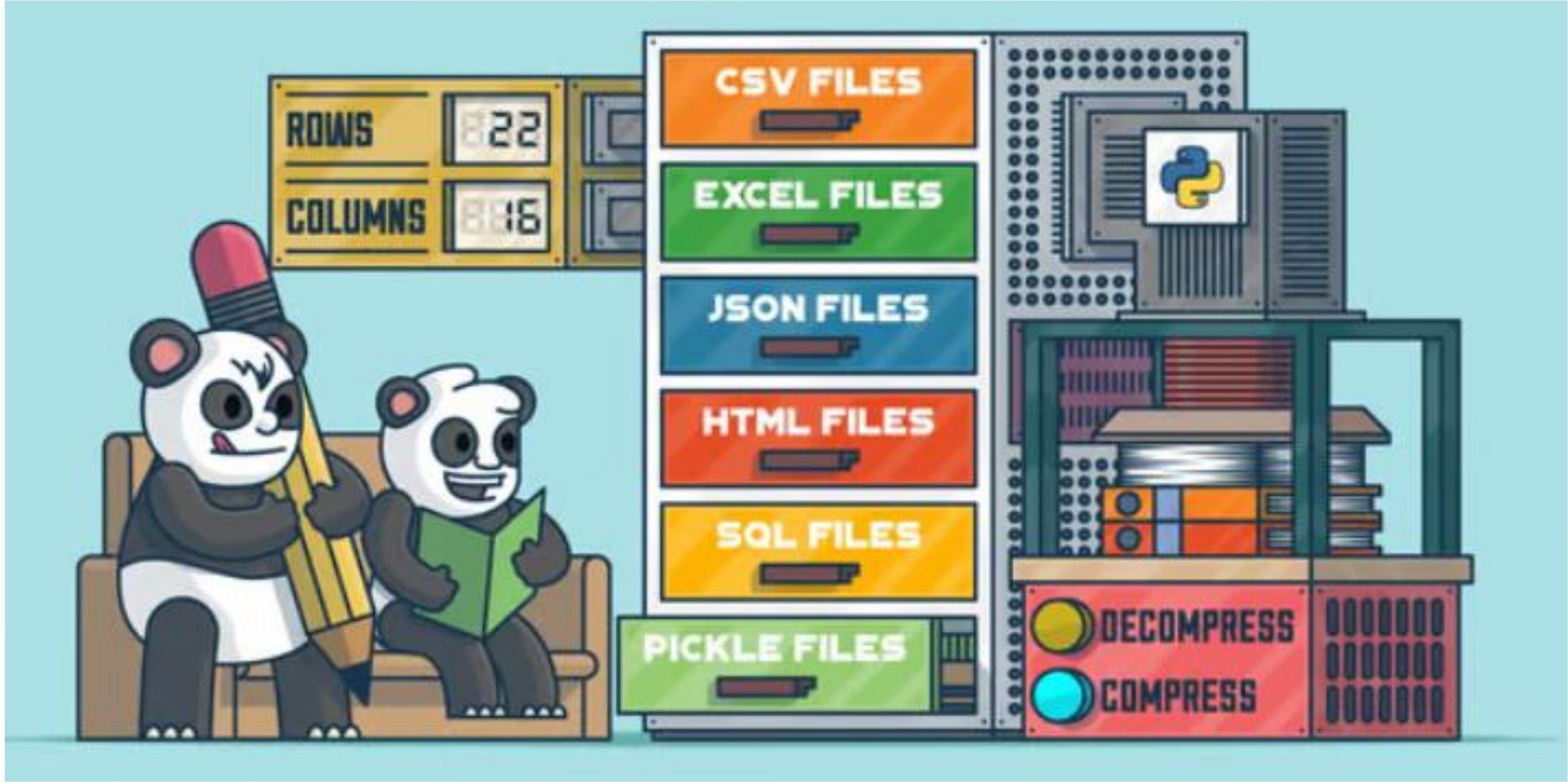
- Divides the data variable into subsets (segments)
- Calculate statistical measures(mean/mode/median etc) for each segment
- Visualization by plotting

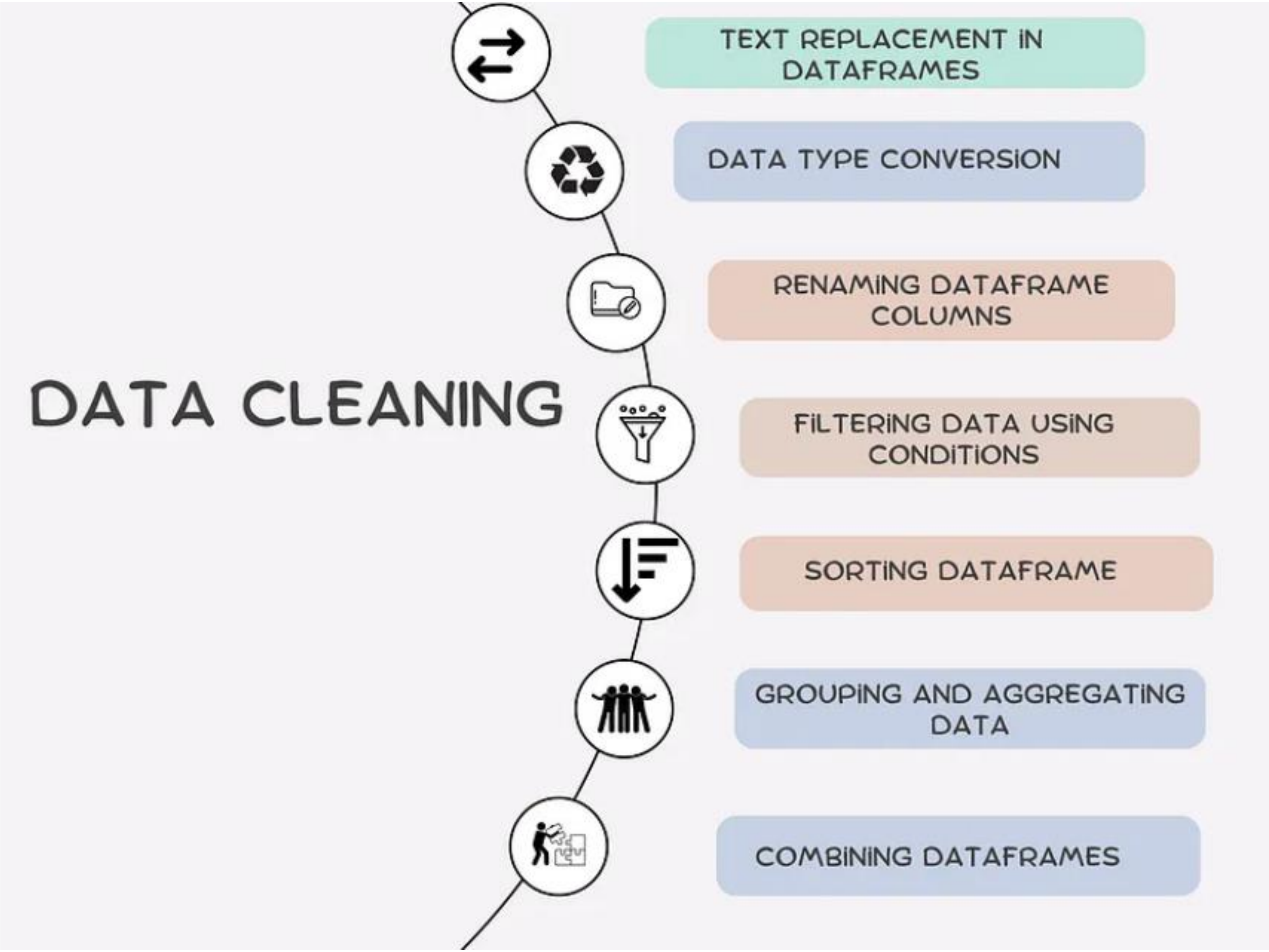
Bivariate Analysis

- Relationships between pairs of variables
- Scatter plot, Strip plot, Correlation Matrix or Heatmap
- Observe the pattern or trends

Data Sourcing

- Sourcing the data from different types of files
- Importing the data in python for analysis using pandas library



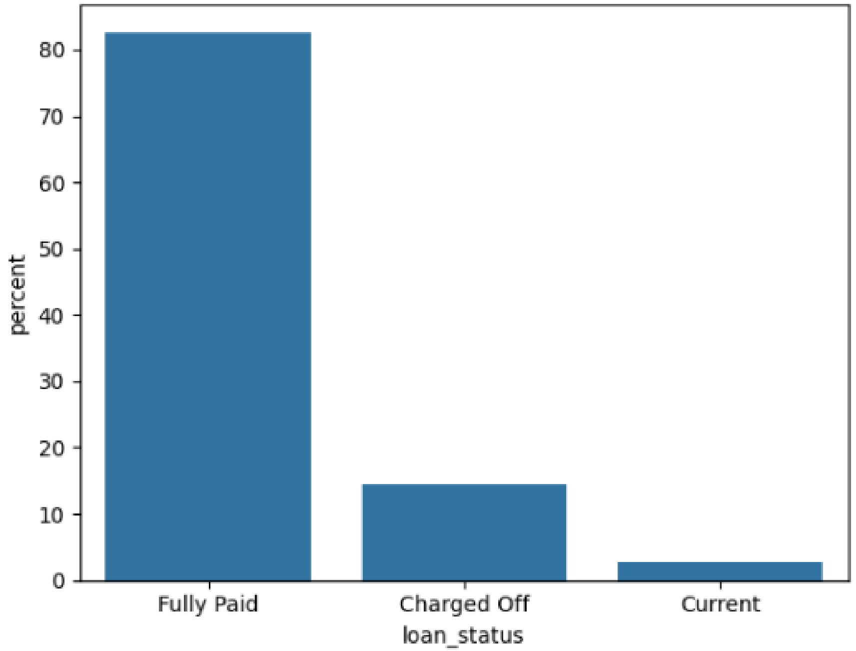

DATA CLEANING
TEXT REPLACEMENT IN DATAFRAMES
DATA TYPE CONVERSION
RENAMING DATAFRAME COLUMNS
FILTERING DATA USING CONDITIONS
SORTING DATAFRAME
GROUPING AND AGGREGATING DATA
COMBINING DATAFRAMES

Overview of Case Study

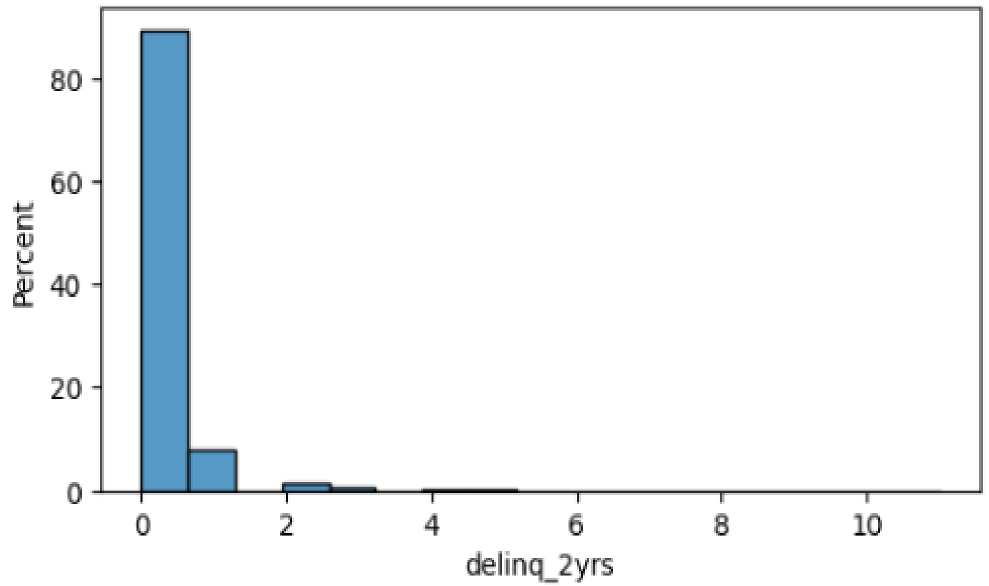
- Performing EDA on the shared loan data will provide the driving factors that can be utilized for the risk assessment for the new applicants.
- Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. If one can identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

Univariate & Segmented Univariate Analysis

Loan Status



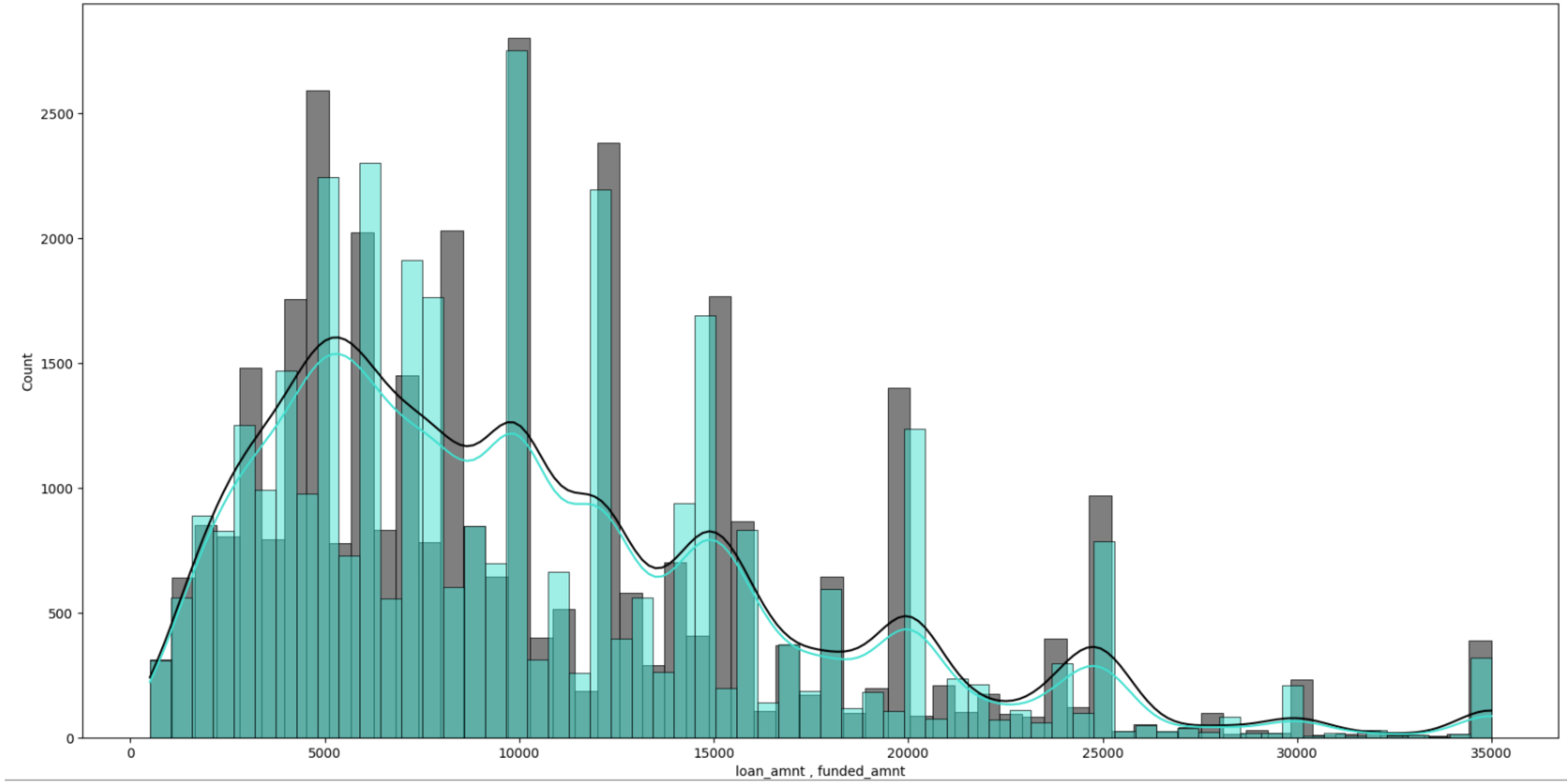
Delinquency of 30+ days for past 2 Years



Observation

- Around 80% of the issued loans are Fully Paid & doesn't have any record of monthly late payments
- Around 15% of the issues loans are Charged Off

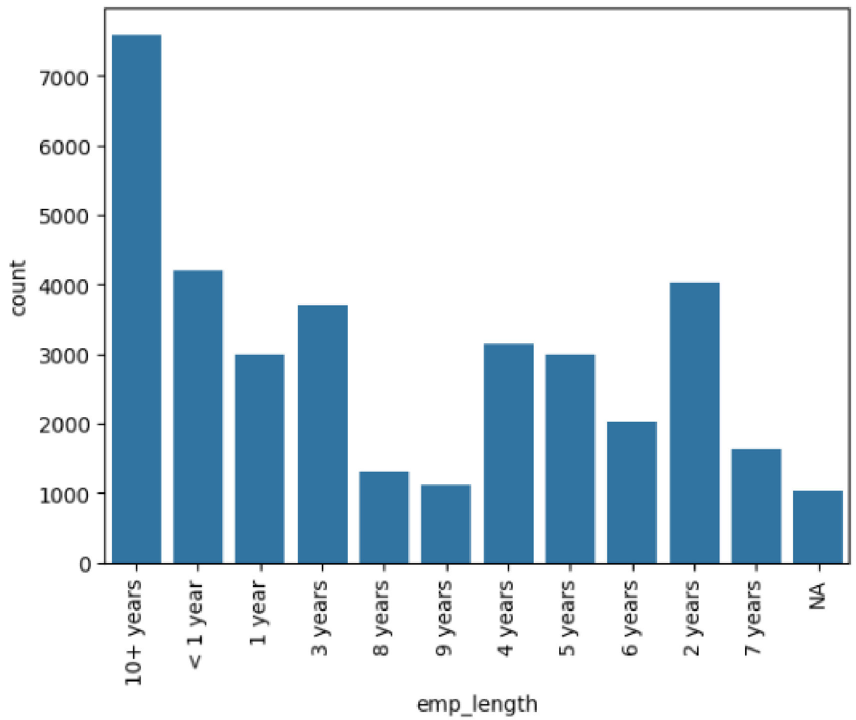
Loan Amount & Funded Amount

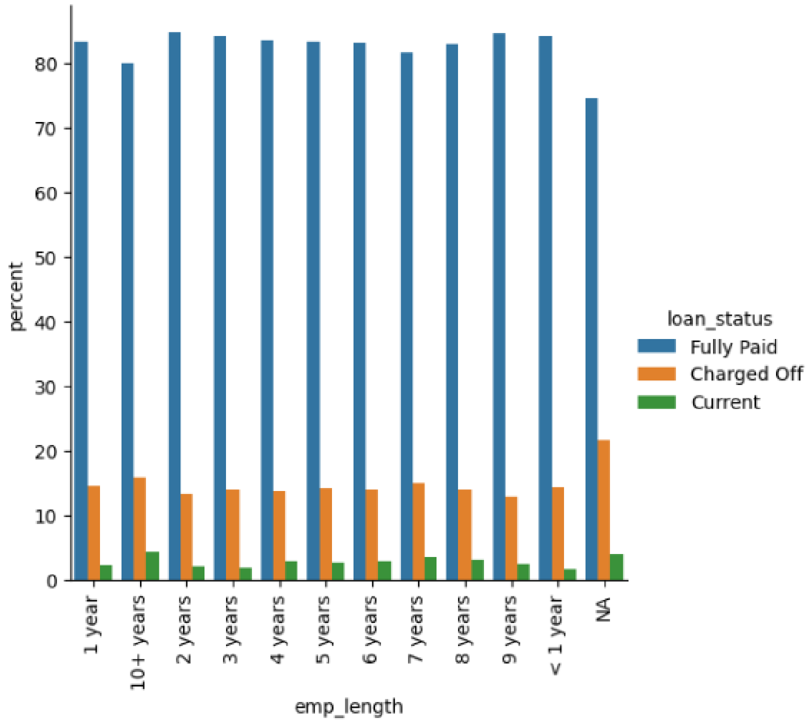


Observation

- The requested & funded amount for most of the loans lies in the range of 5000-15000
- The pattern is similar for both of them
- Funded ampunt is slightly lower than requested amount

Employment Length

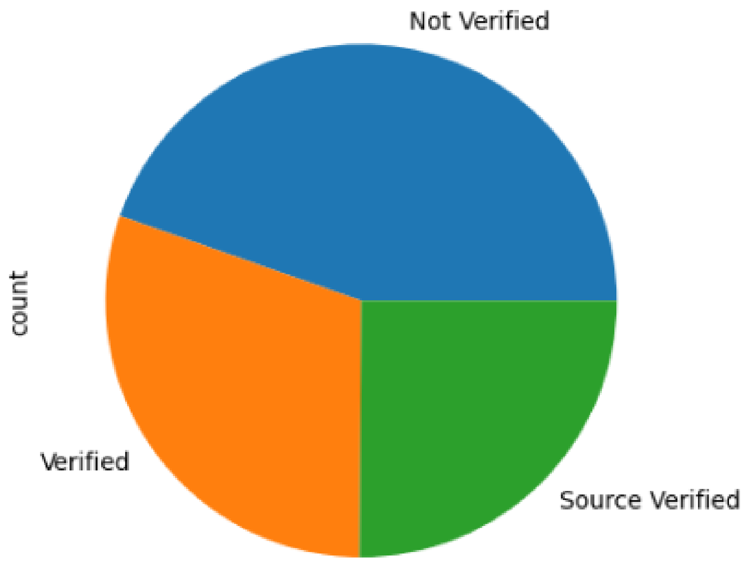


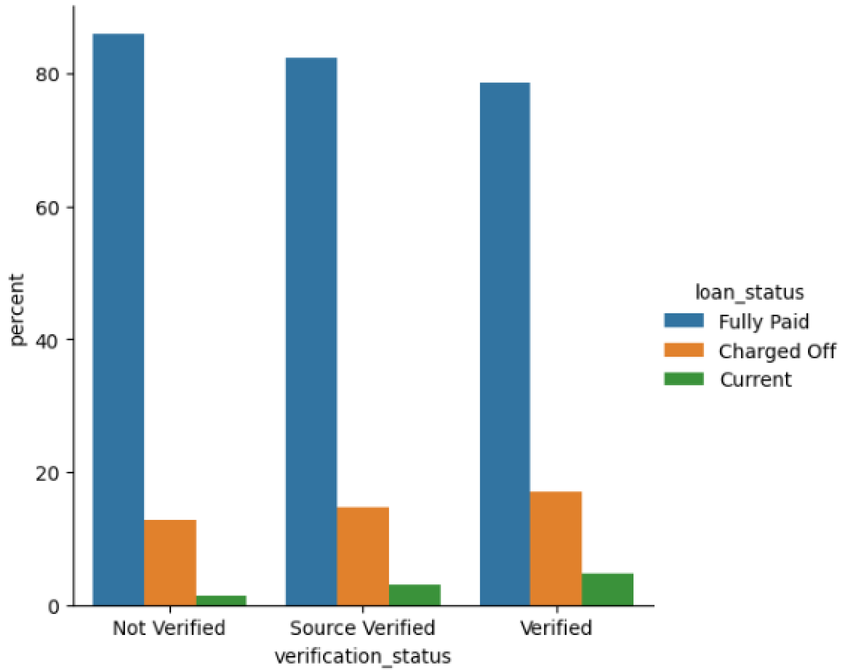


Observation

- Employees working for more than 10 years have mostly applied for loan
- Borrowers who haven’t provided employment details have high percentage of Charged Off loans

Verification Status

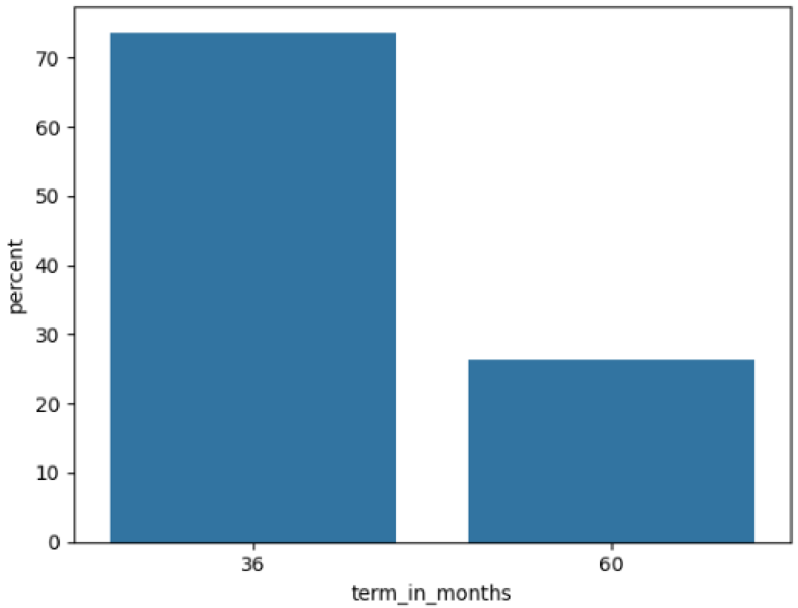


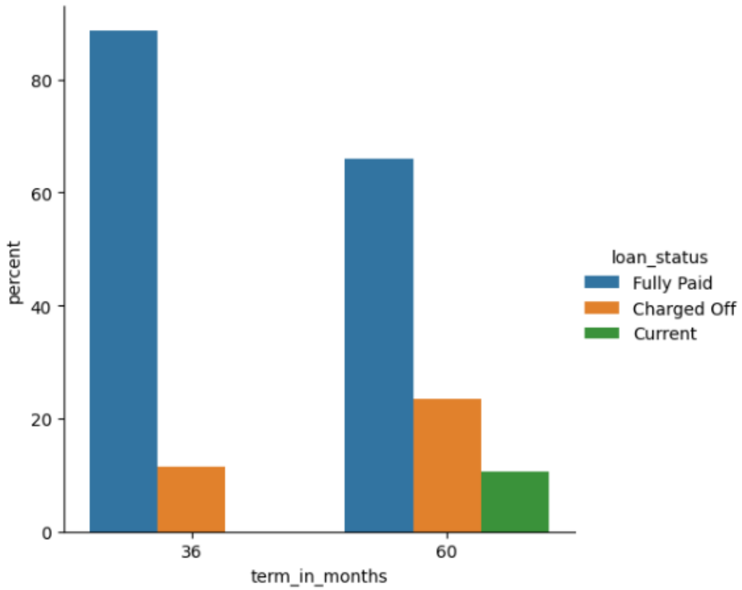


Observation

- Most of the cases, annual income is not verified
- No useful pattern of loans that are charged off w.r.t verification status

Term in Months

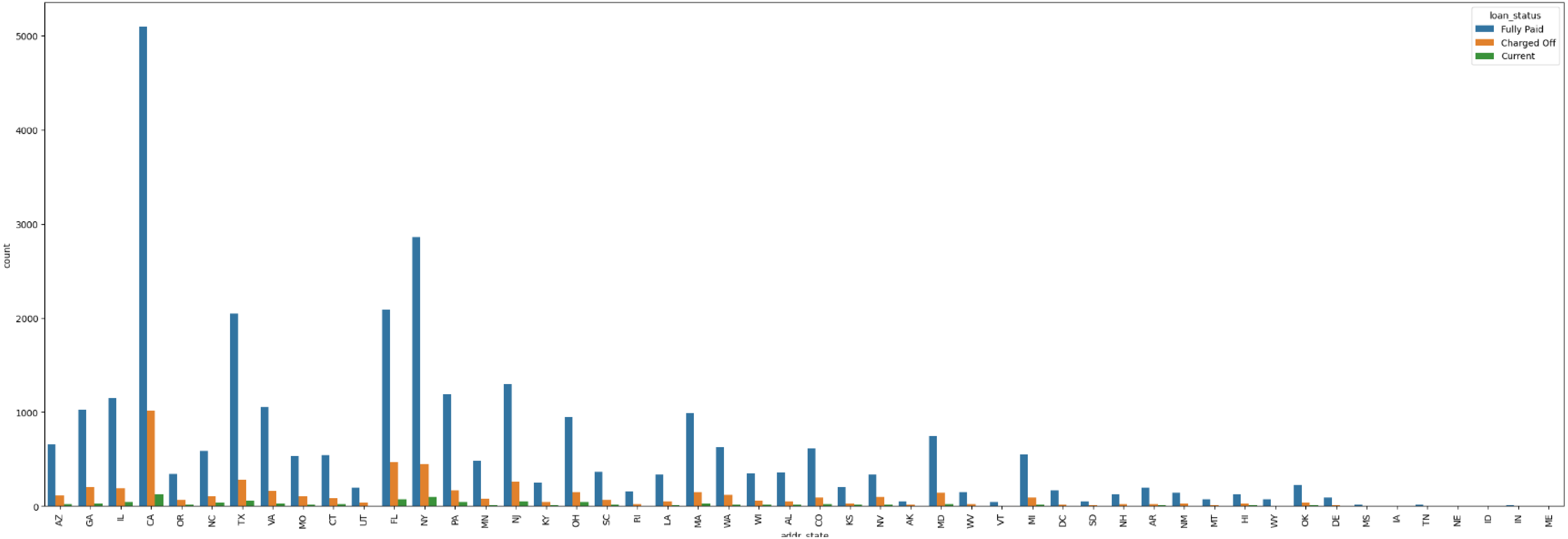




Observation

- Most no of loans are issued for the term of 36 months
- Most no of Charged Off loans are there in 60 months term

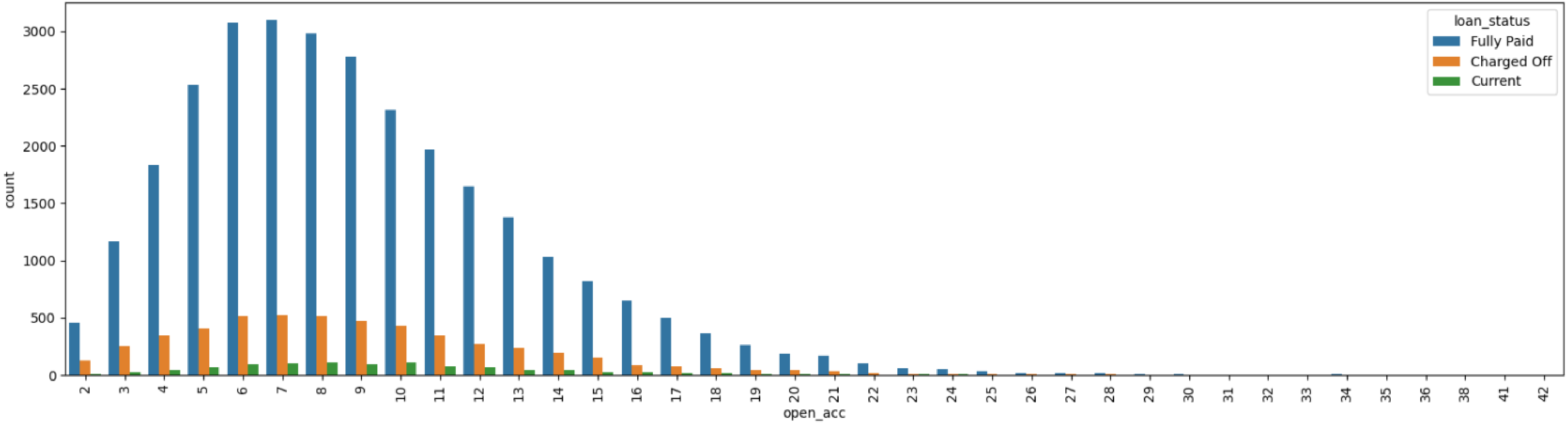
Address State



Observation

- Most number of loans are applied in the states of California (CA), Florida (FL) , New York (NY), Texas (TX)
- No specific patterns in terms of Charged Off loans

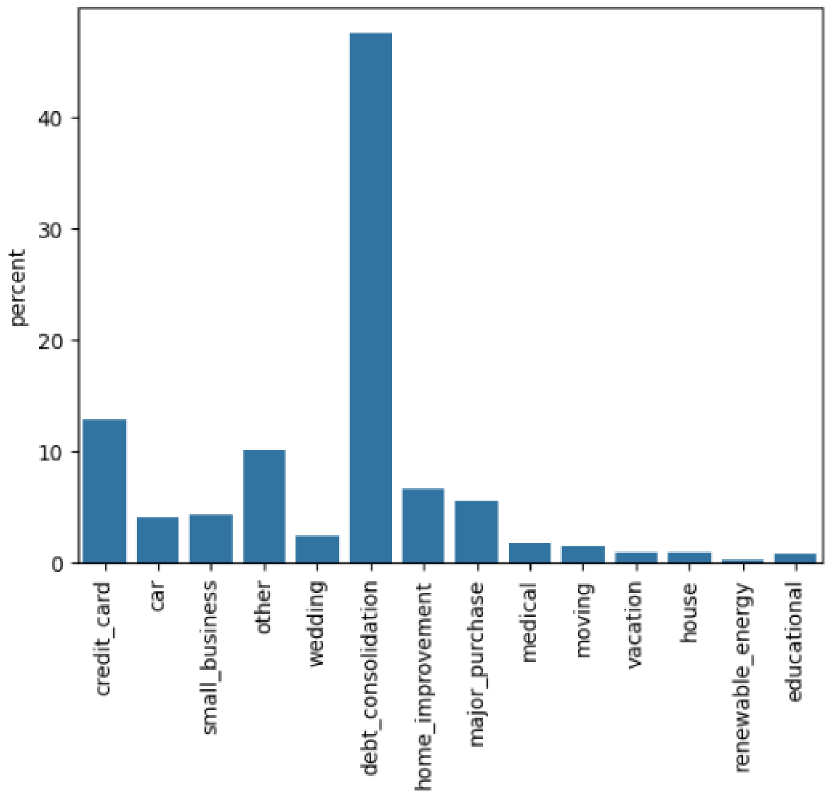
Open Credit Accounts

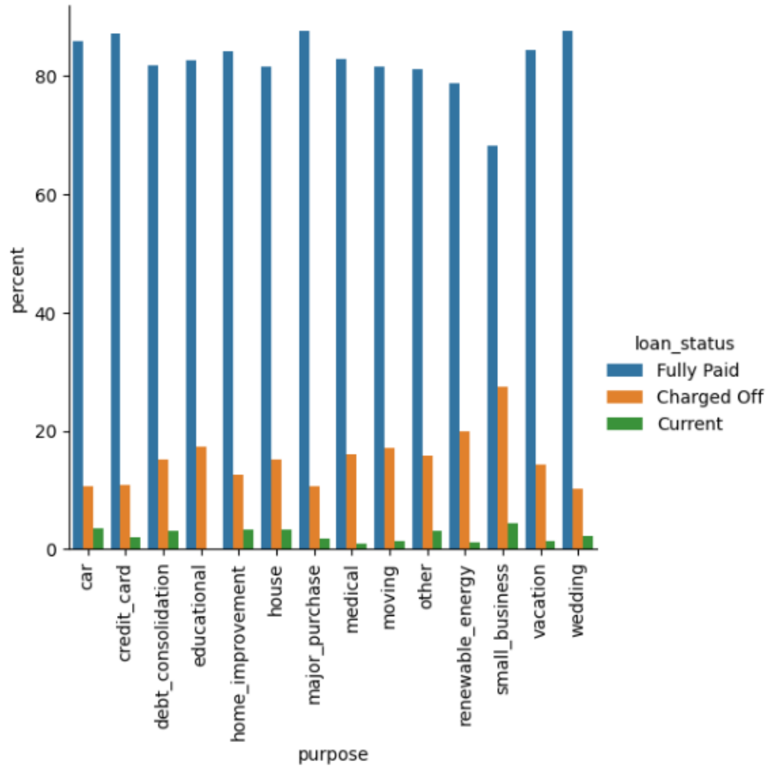


Observation

- Borrowers who are issued loan have mostly 3-15 Open Credit Accounts
- No specific patterns in terms of Charged Off loans

Purpose of Loan

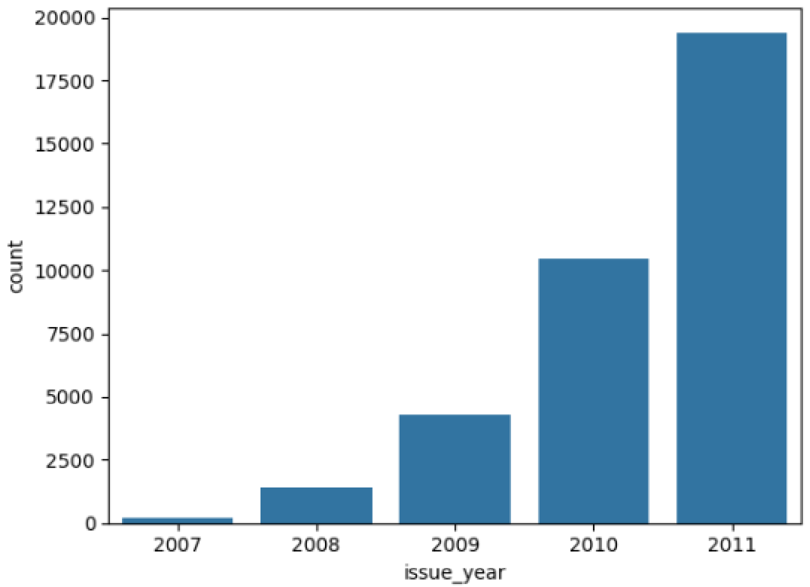




Observation

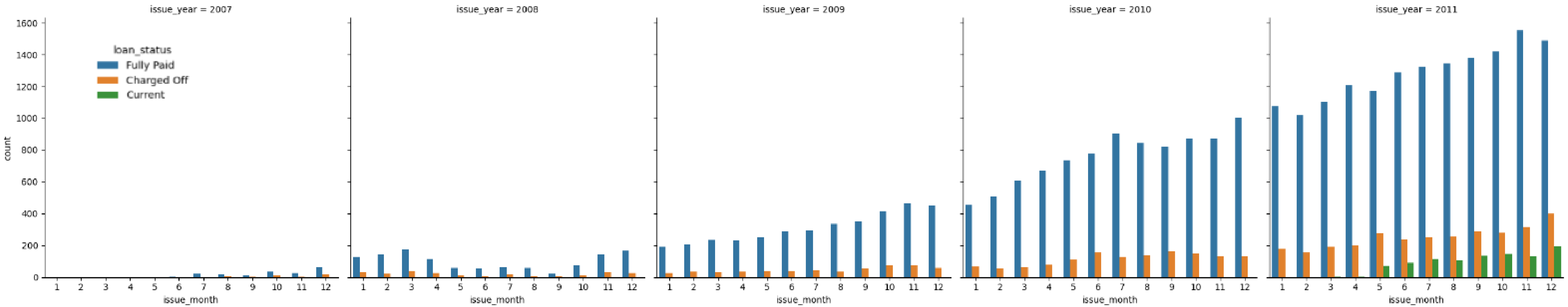
- Almost half the loans are for the purpose of debt consolidation
- Higher percentage of Charge Offs are for the loans with purpose of Small Business

Issue Month & Year

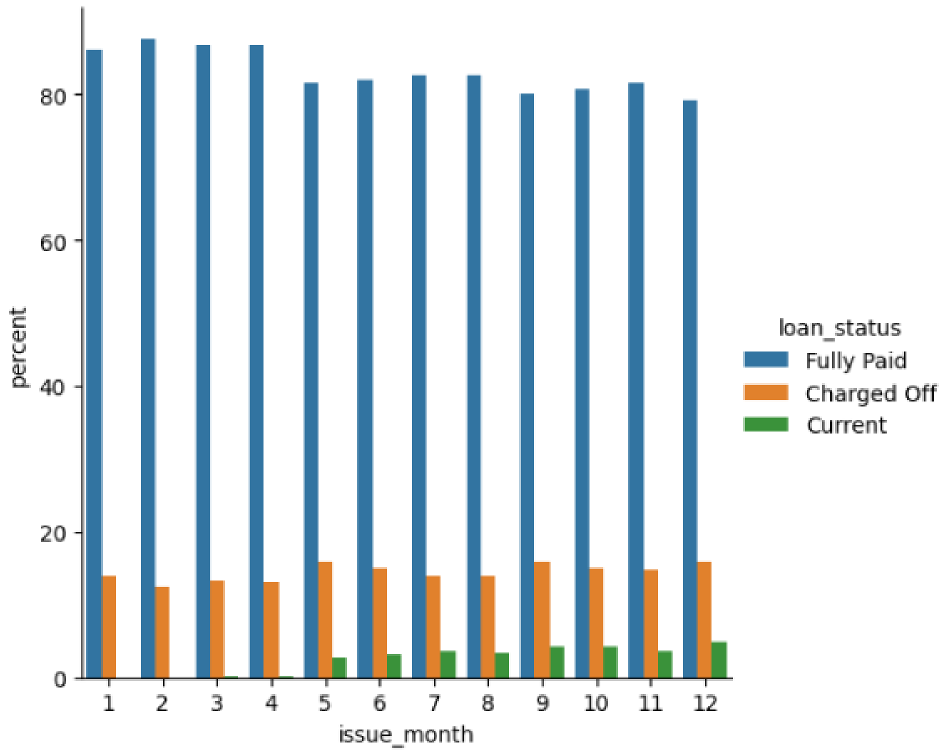


Observation

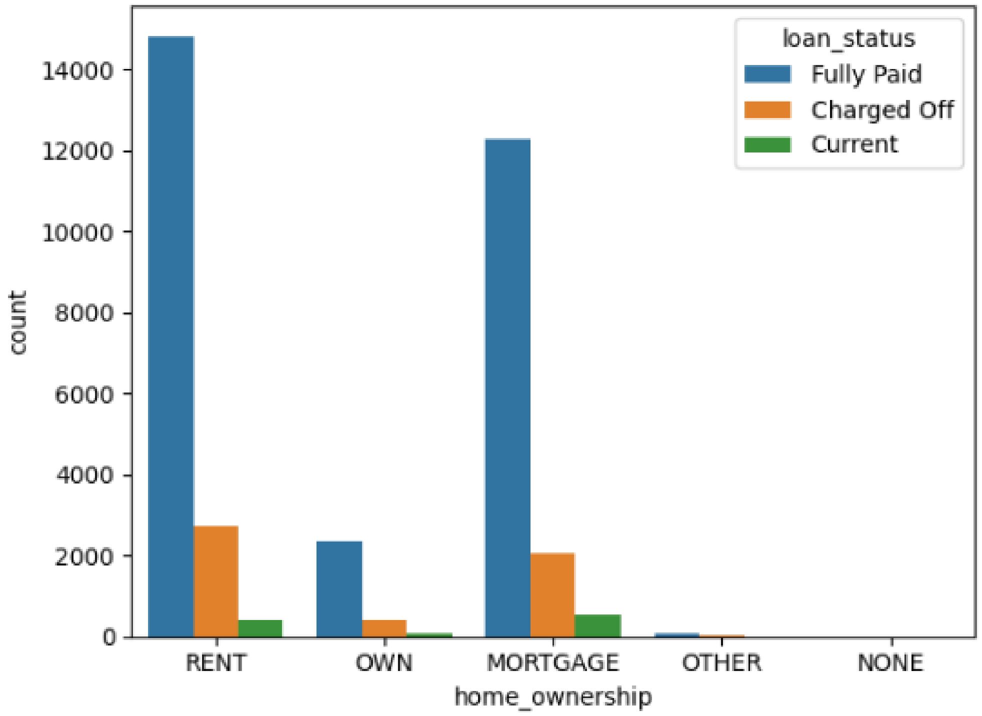
- Most loans are issued in the year 2011
- Total number of loans & Charge Off loans are growing w.r.t year & month



Issue Month



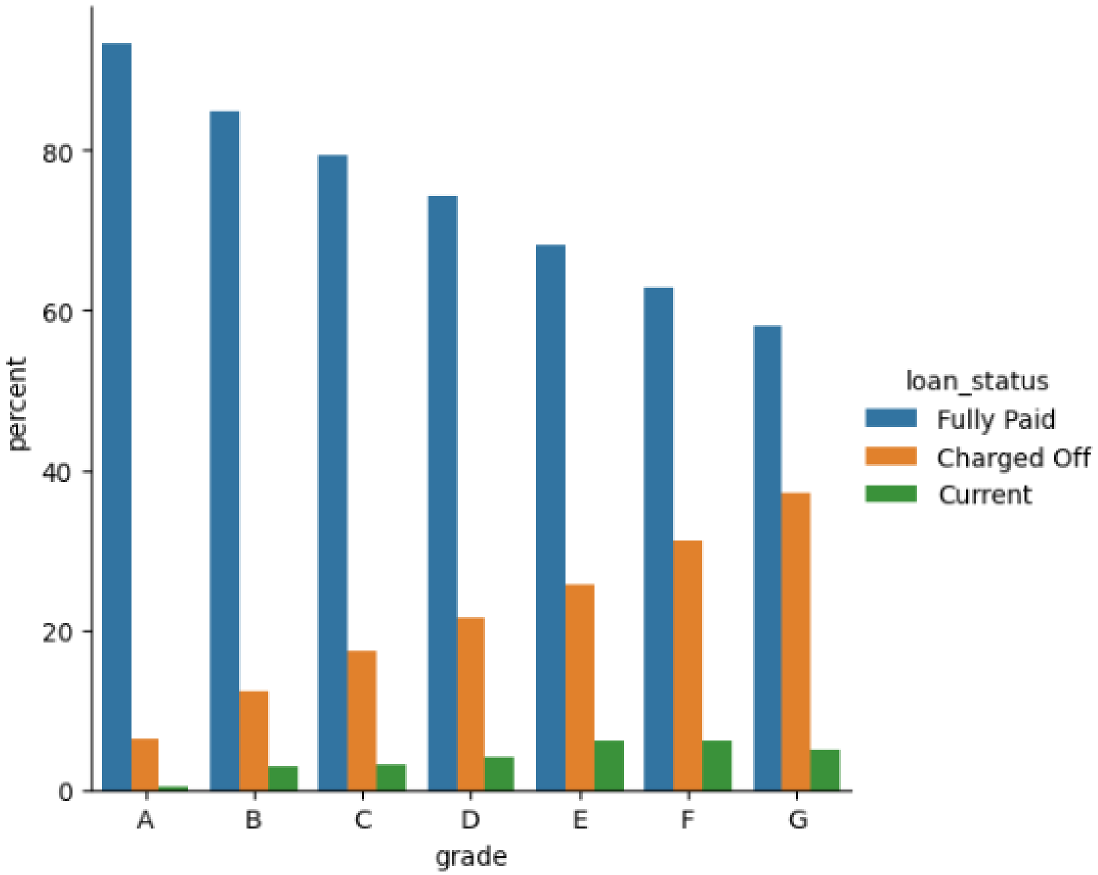
Home Ownership



Observation

- Across the years, loans issued in the month of February has least percentage of Charge Offs
- Loans are applied mostly by the borrowers who doesn't have own property
- No useful relation of Charged Off loans with Home Ownership

Grade

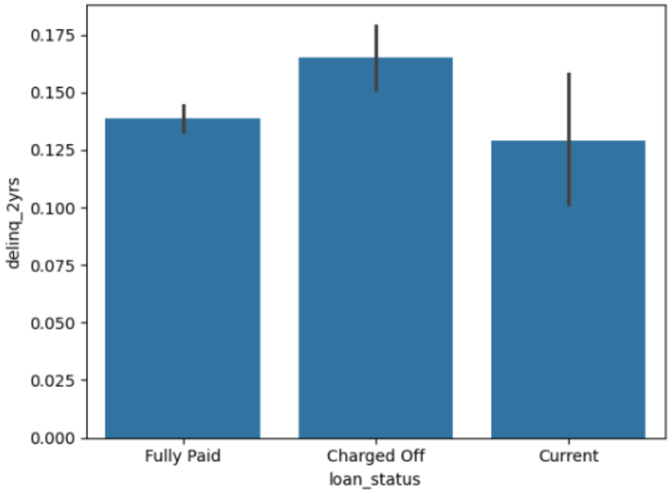


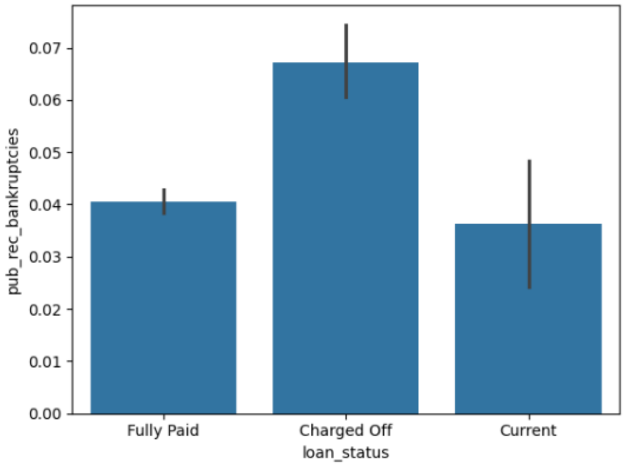
Observation

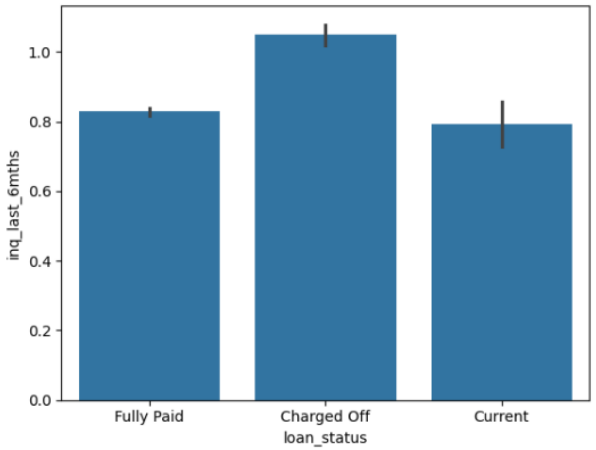
- With increasing grade, the percentage of Charged Off loans are increasing

Bivariate Analysis

Loan Status



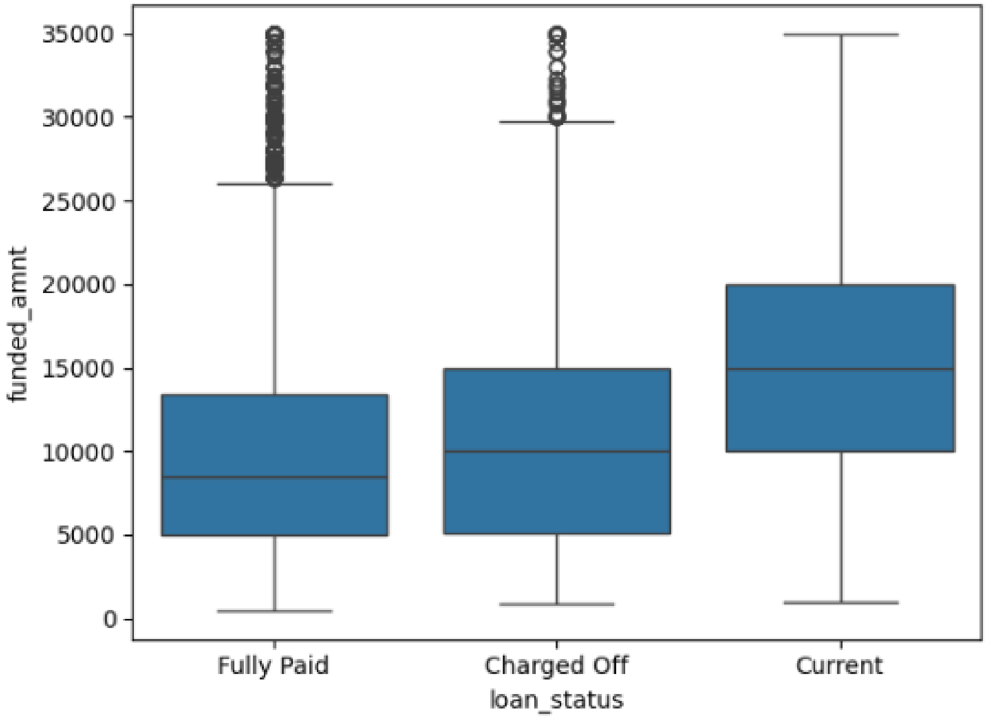


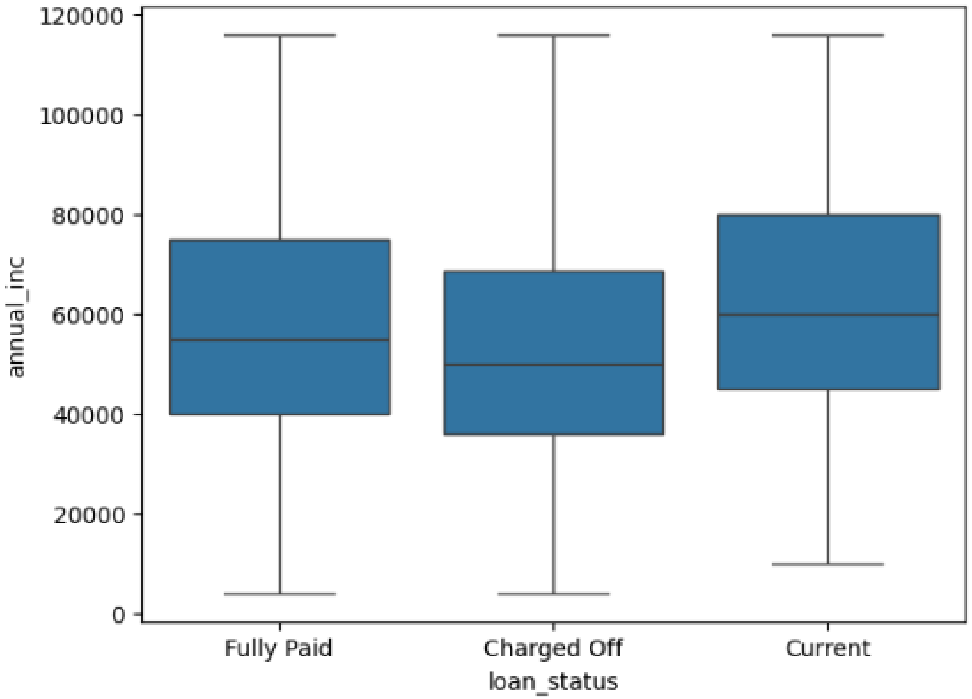


Observation

- Chances of loans being Charged off are higher if there are high number of monthly late payments or bankruptcies or enquiries

Loan Status

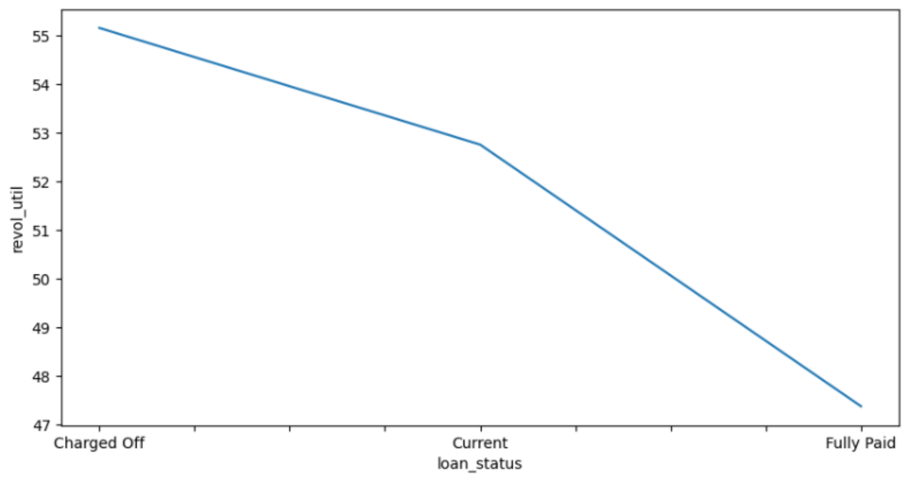


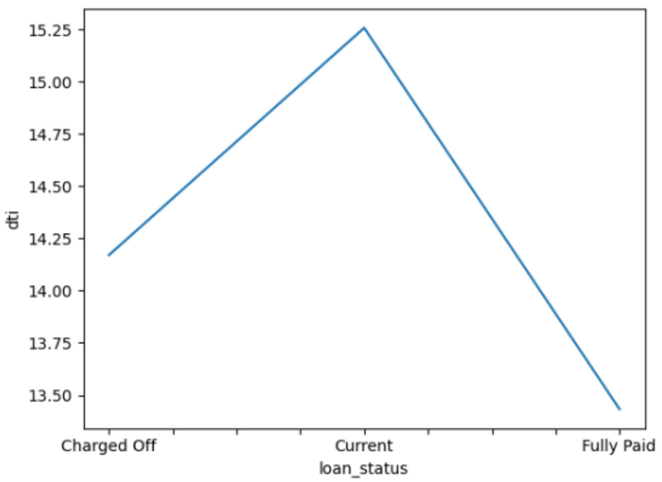


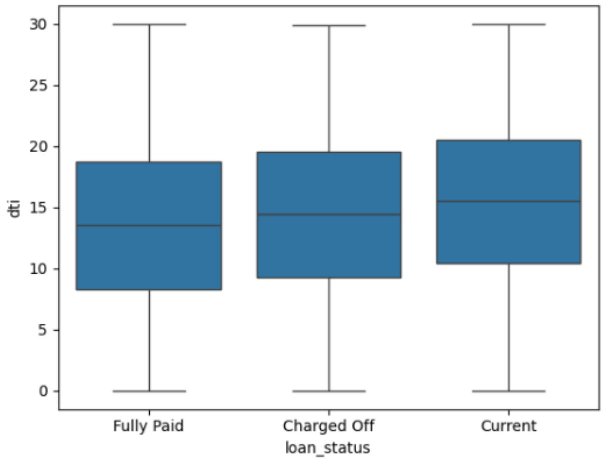
Observation

- Median of funded amount lies higher for Charged Off loans w.r.t Fully Paid loans
- Median of annual income lies lower for Charge Off loans w.r.t Fully Paid loans

Loan Status



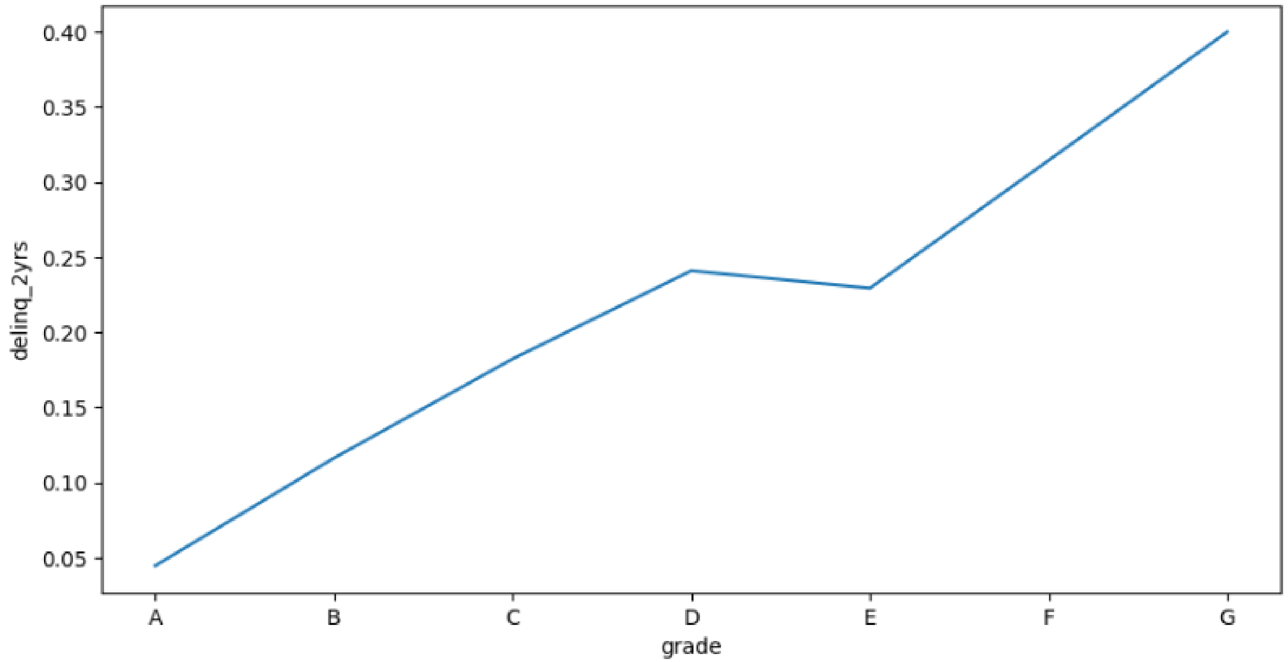


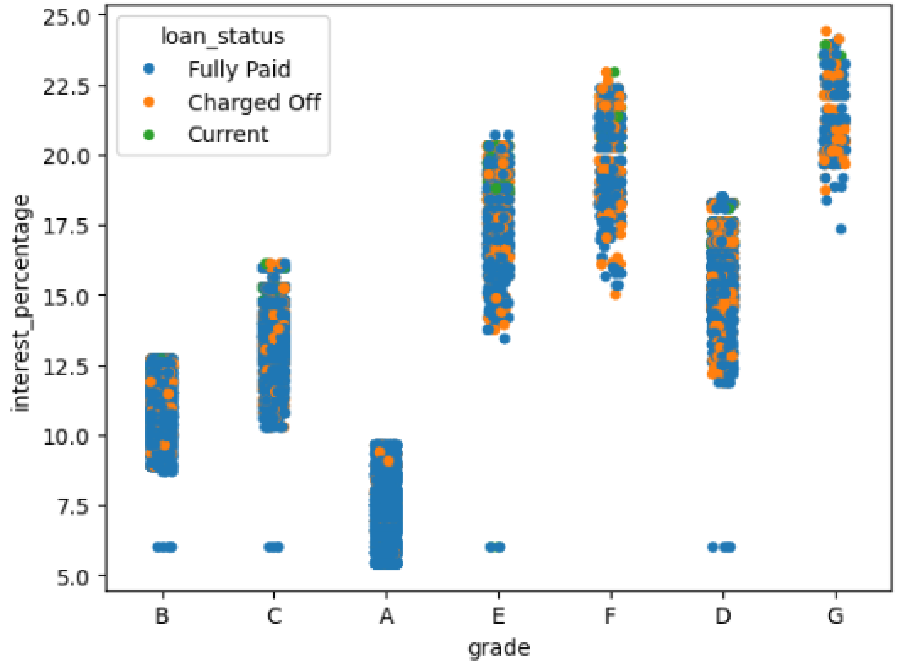


Observation

- Revolving line utilization rate & dti are in average higher for the Charged Off loans w.r.t Fully Paid loans

Grade

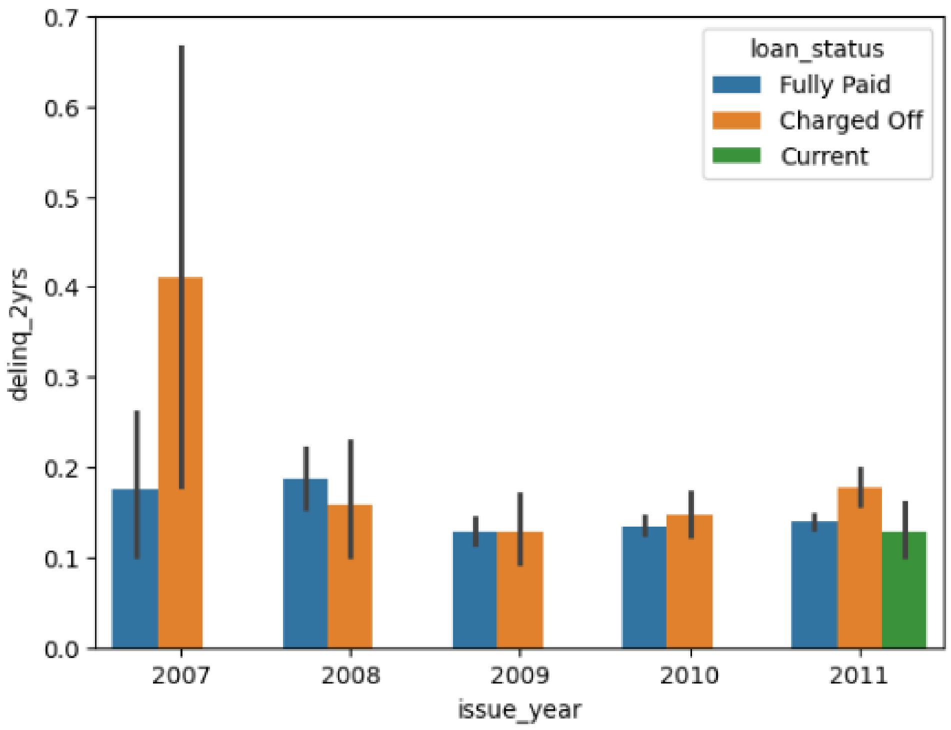


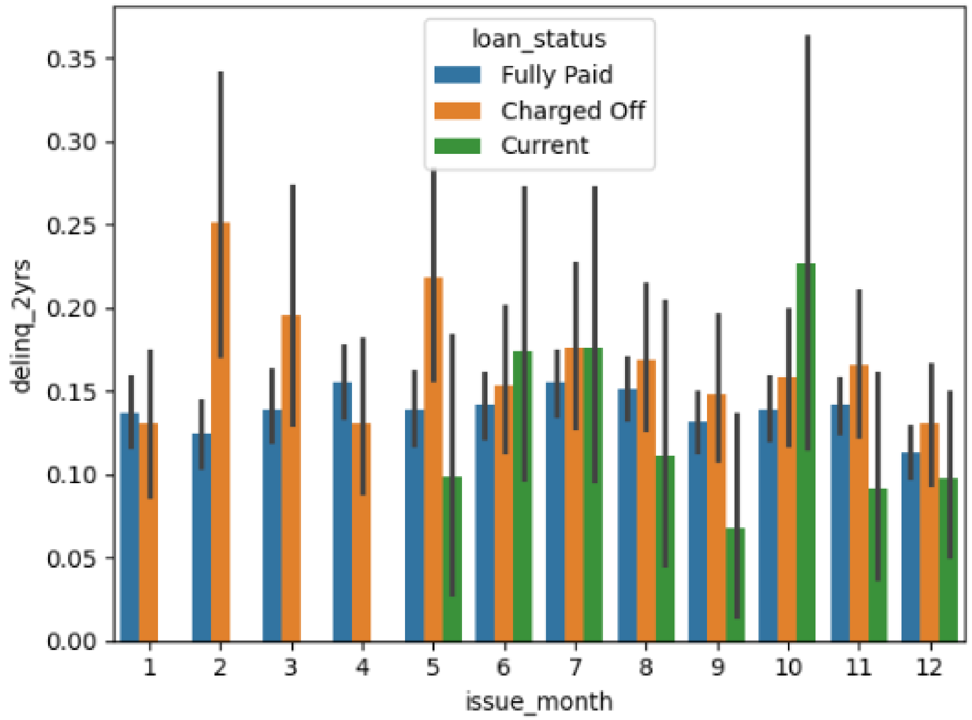


Observation

- Monthly late payments are more for the loans of higher grades
- Interest rate is increasing with grade
- Higher grade, higher interest rate loans are having more Charge Offs

Issue month and year

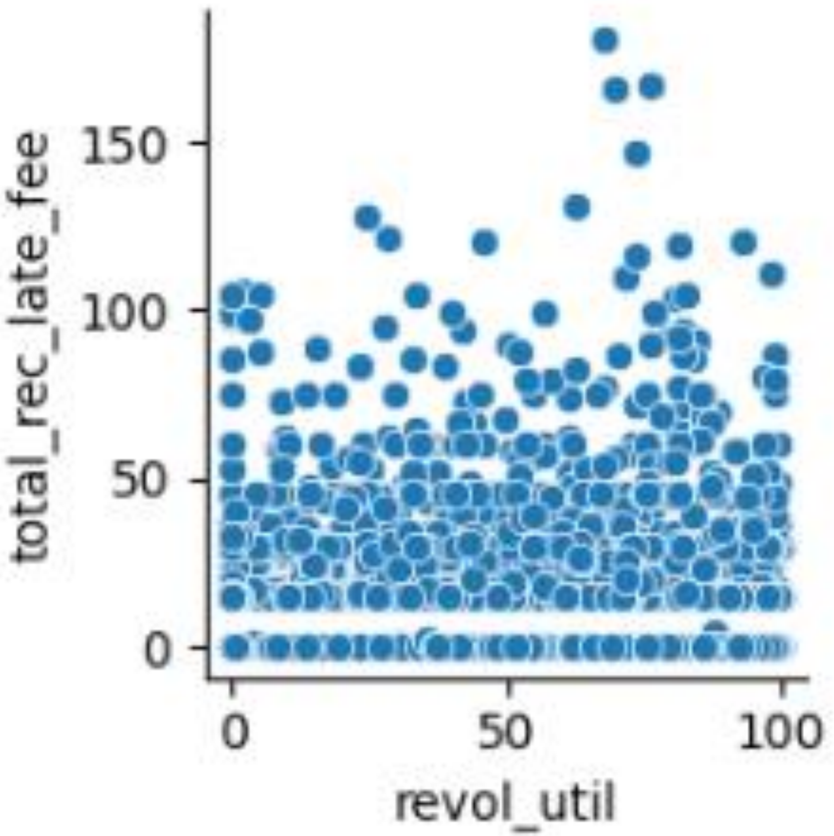




Observation

- Most number of monthly payment delays as well as Charged Off loans happened in the year 2007
- Charged Off loans are high for the loans issued in the month of February, March & May where monthly payments are delayed

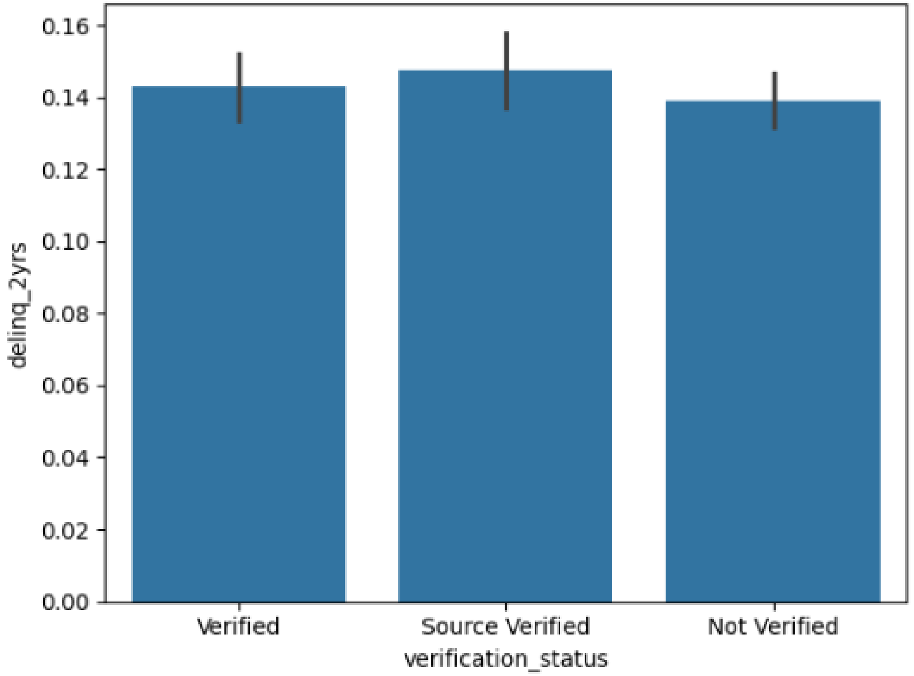
Total amount of Late Fee

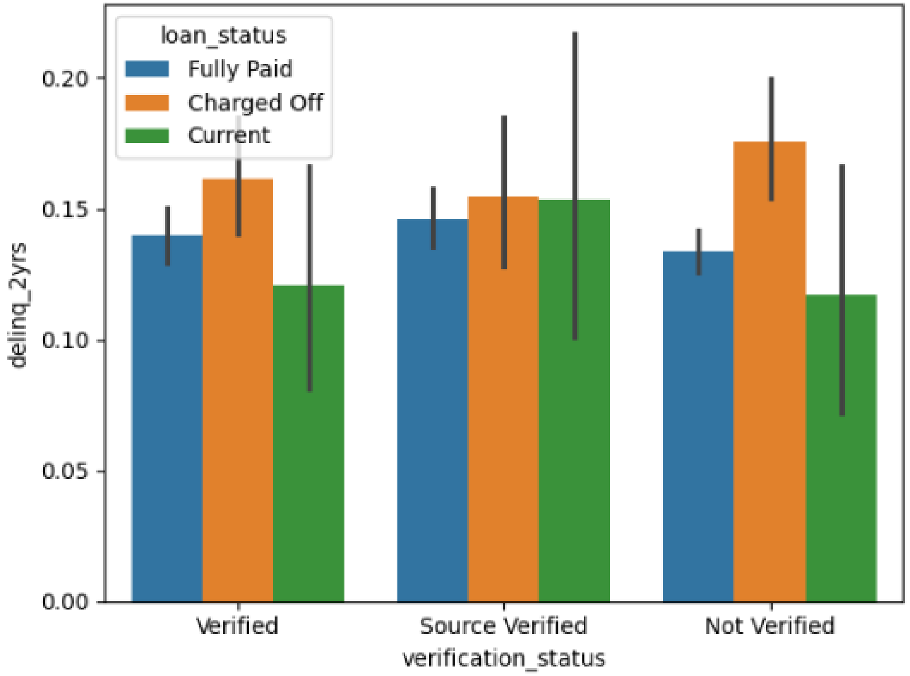


Observation

- For higher Revolving line utilization rate, the total amount of late fees is higher

Verification Status

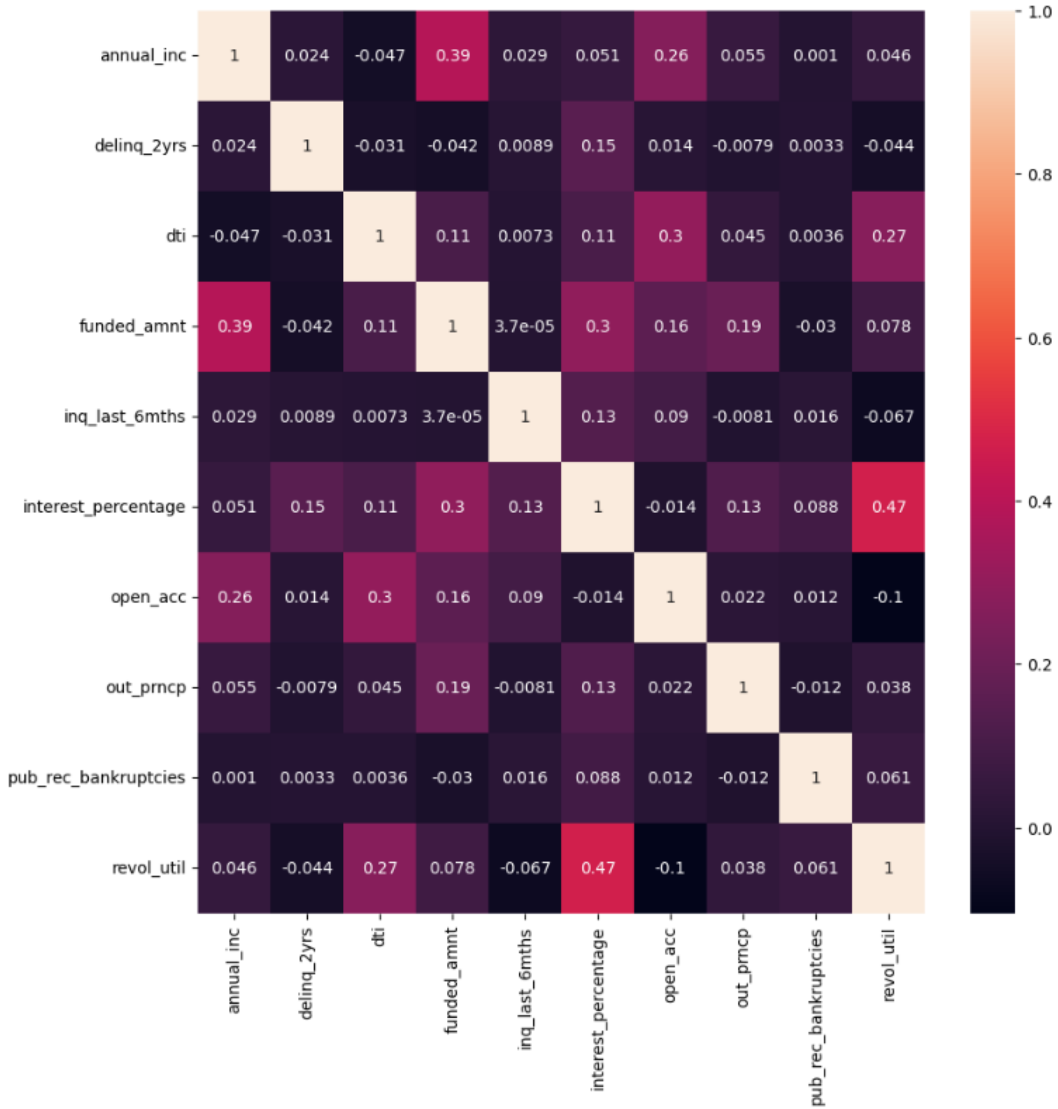




Observation

- No significant pattern between income verification status & monthly late payments
- Charged Off loans are slightly higher for Not Verified incomes when monthly payments are considered

Correlation Matrix



Observation

- Relatively high positive correlation between the following pairs
  - Revol Util Rate & Interest Percentage
  - Funded Amount & Interest Percentage

Conclusion

- **Univariate Analysis**
  - 80% of the loans are in fully paid status and almost 15% are charged off.
- **Segmented Univariate Analysis**
  - Borrowers with higher Grades have a high chance of default.
  - The borrower with a higher term of the loan has a high chance of defaulting.
  - The borrower with the purpose of "Small Business" has a high chance of defaulting.
- **Bivariate Analysis**
  - The borrower who has a high revolving utilization rate has a high chance of defaulting.
  - The borrower making late monthly credit installment payments has a high chance of defaulting.
  - The borrower with a higher DTI-Debt to Income ratio has a high chance of defaulting.
  - The borrower with a higher number of public record bankruptcies has a high chance of defaulting.
  - The borrower who has made a higher number of loan inquiries has a high chance of defaulting.

Thank You