



VIKASKUMAR GUPTA
RAGHAVA PARISA

LENDING CLUB CASE STUDY

PROBLEM STATEMENT



Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).



Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.



In other words, borrowers who default cause the largest amount of loss to the lenders.

BUSINESS OBJECTIVE

- **As a Data Scientist working on Lending Club analysis which contains the customers historical data about the loan applicants, We need to perform EDA so that how the data is showing the trends.**
- **Basically, whenever the lender refuses to pay back the amount that owned causes the credit loss which leads to financial loss to the respective bank. These type of customers comes under 'Charged-Off' or the 'Defaulters'.**
- **Dataset contains different type of loan parameters for a particular period (2007-2011).**
- **The loan company wants to understand the driving factors (or driver variables) behind loan default for approving the loan to the customers.**
- **The variables which are strong indicators of 'Defaulters'.**
- **The company can utilize this knowledge for its portfolio and risk assessment.**

STEPS INVOLVED



Data Exploring

- 1) No. Of columns
- 2) Understanding the column data types
- 3) Using Data Dictionary understanding the meaning of each column



Data Cleaning

- 1) Identify and removing the columns & rows with Null, Missing & Unique values.
- 2) Deriving new metric/columns for better understanding/identification.



Exploratory Data Analysis

- 1) Univariate
- 2) Segmented Bivariate
- 3) Bivariate



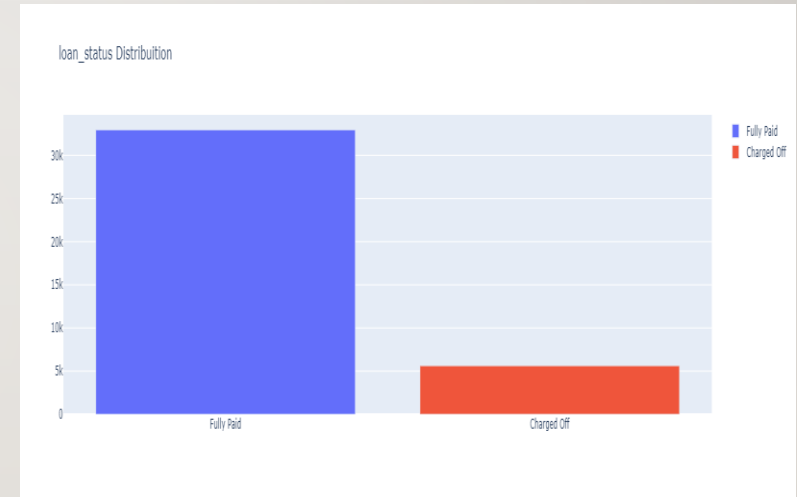
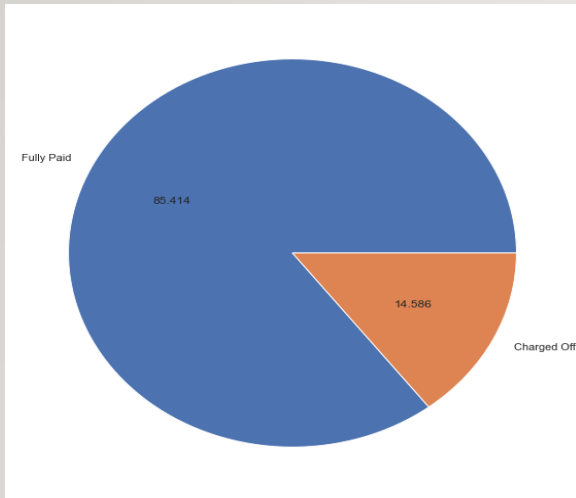
Data Visualization

- 1) Using Matplotlib, Plotly and Seaborn visualizing the graphs through Dist, Bar, Hist, Pie charts.

VARIABLES INVOLVED IN ANALYSIS

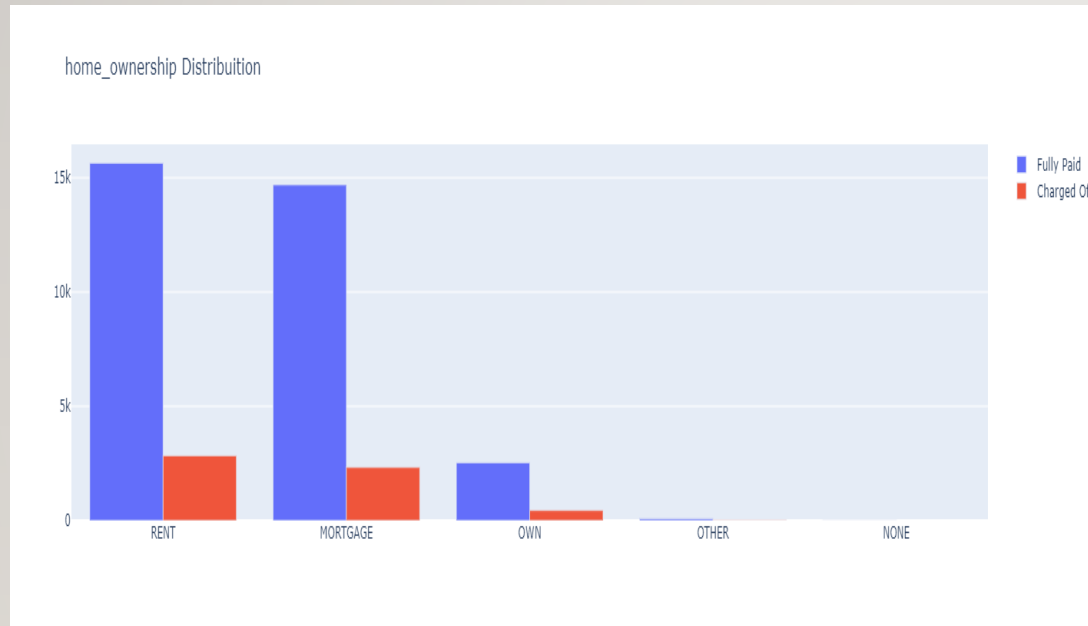
- **Loan_status** - Current status of the loan
- **home_ownership** - The home ownership status provided by the borrower during registration. Our values are RENT, OWN, MORTGAGE, OTHER.
- **Purpose** - A category provided by the borrower for the loan request.
- **Int_rate** - Interest Rate on the loan
- **Verification_status** - Indicates if income was verified by LC, not verified, or if the income source was verified
- **emp_length** - Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- **Term** - The number of payments on the loan. Values are in months and can be either 36 or 60.
- **Grade** - LC assigned loan grade
- **Sub_grade** - LC assigned loan subgrade
- **Addr_state** - The state provided by the borrower in the loan application
- **Year** - The year which the loan was funded

LOAN STATUS ANALYSIS



- The average default rate is 14.58%
- There are more Fully Paid Customers compared to Charged Off

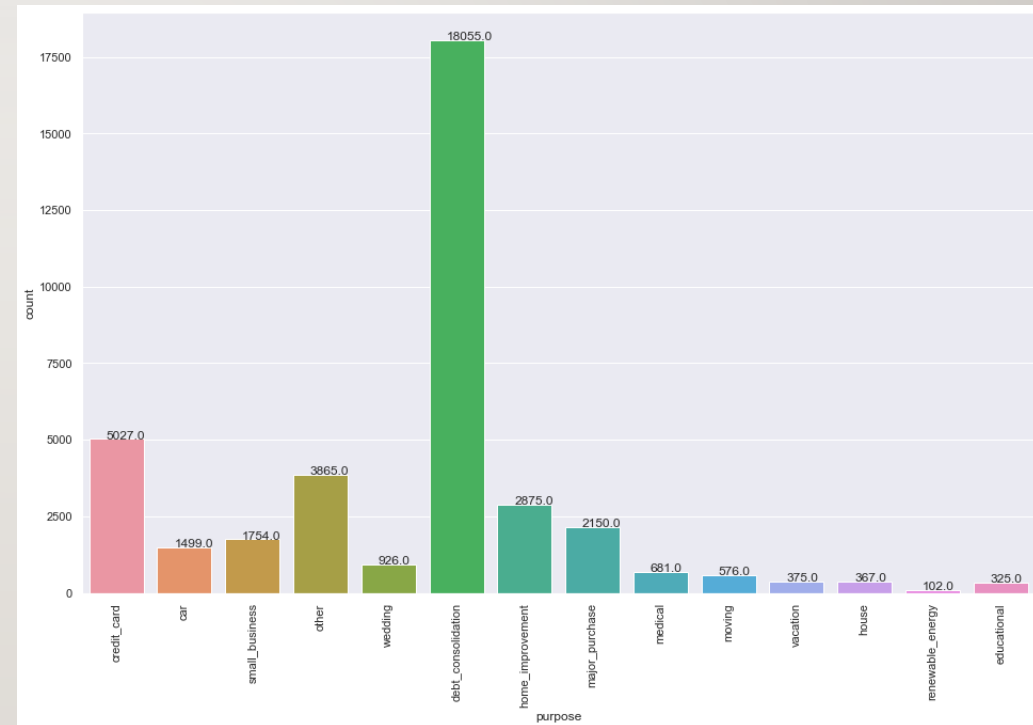
HOME OWNERSHIP ANALYSIS



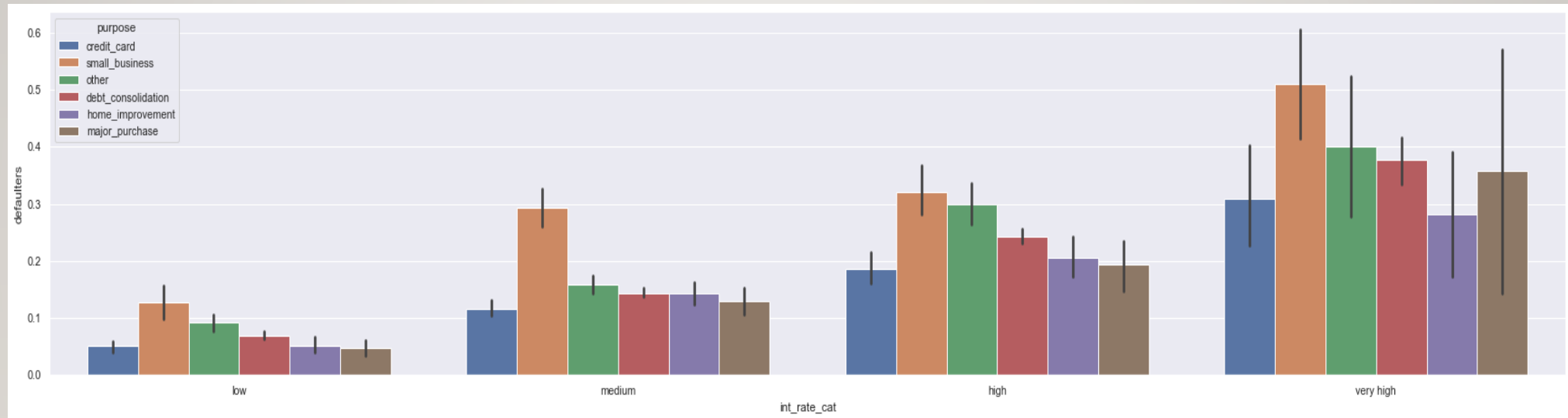
- Customers who live in Rent applied for loans comparatively, this could be because they have taken loans for constructing a new house or for other purposes.

PURPOSE ANALYSIS

- Most of the customers who has taken loan for the Purpose of Debt_consolidation

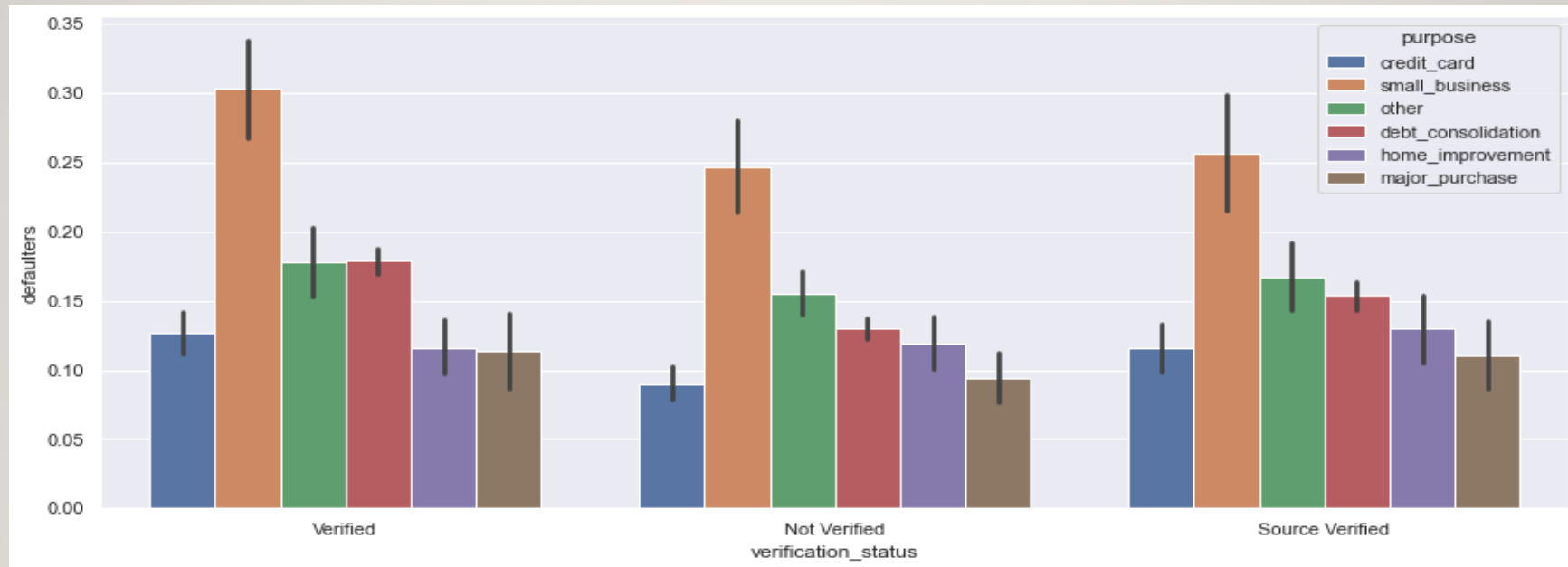


INTEREST RATE ANALYSIS



- Most of the applicant have large defaulter rate when they have the higher interest rate.
- By using the analysis bank can process the loan approval based on the defaulter's rate.
- To the above graph loan applicants in the small business category is the larger defaulter because as their interest rate is high, they may not be able to pay the loan back on time which may cause a loss to the bank.

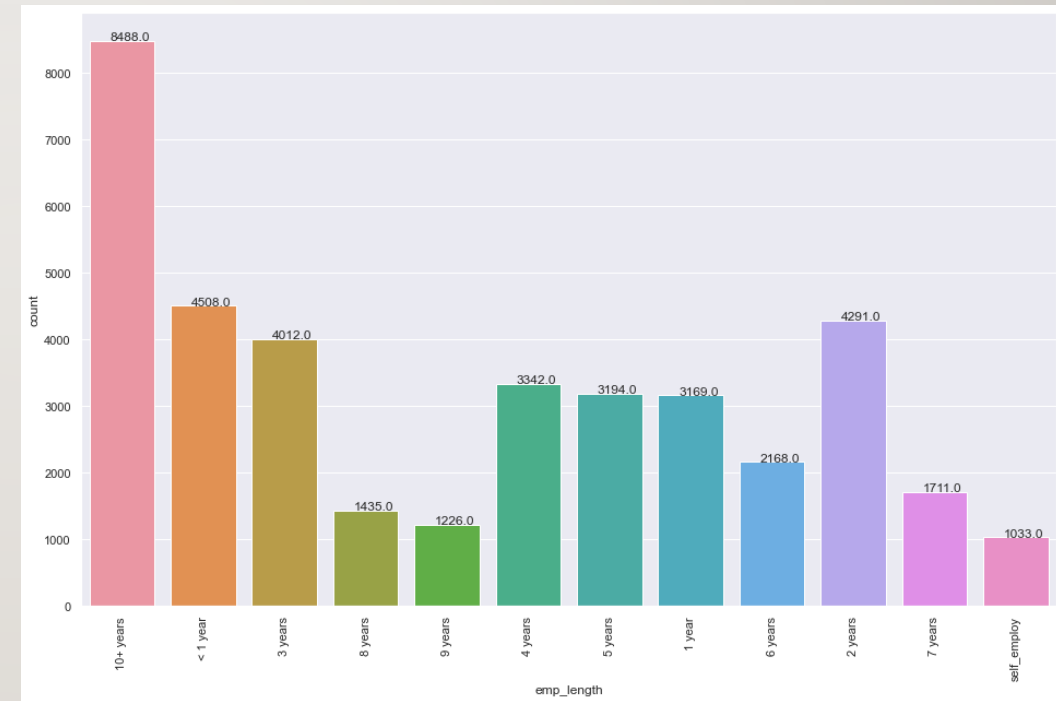
VERIFICATION STATUS ANALYSIS



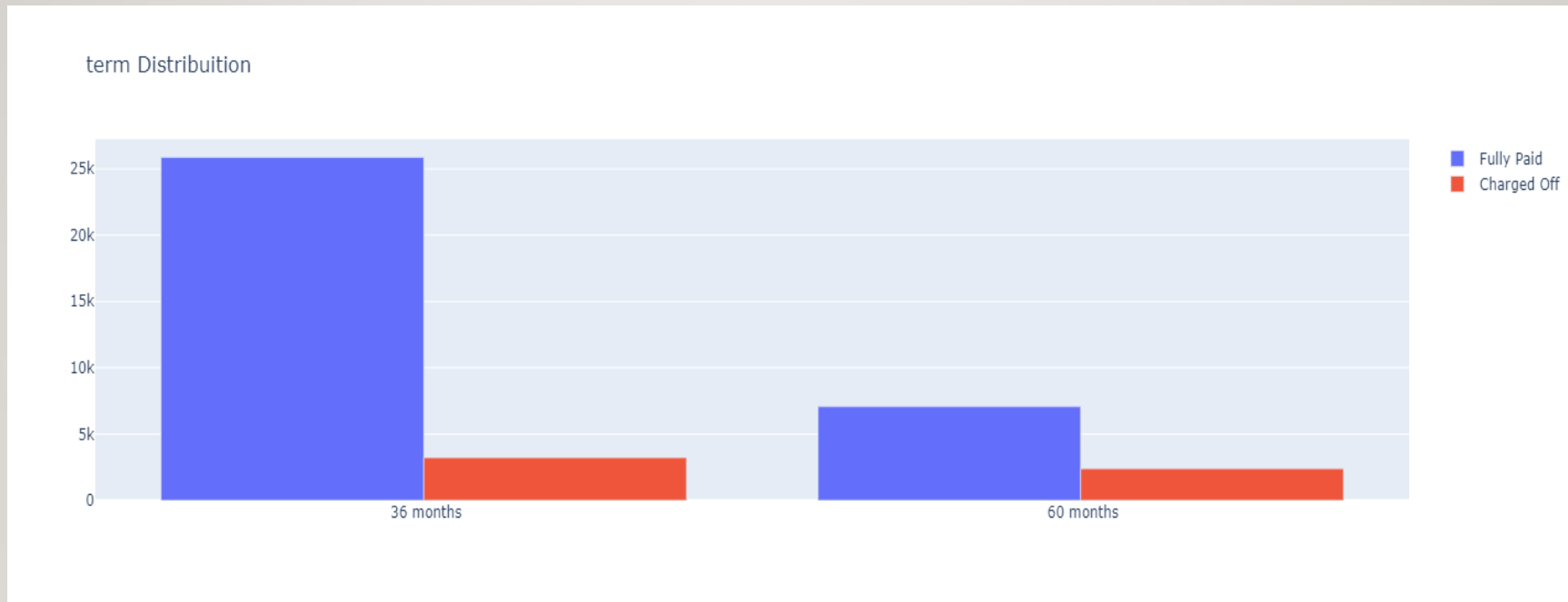
- Most of the applicants whose loan got verified on purpose of Small business.
- As the highest defaulter value is showing for the small business category bank should be more careful before approving the loan.

EMPLOYEE LENGTH ANALYSIS

- Customer with higher experience has a greater number of loans comparatively.

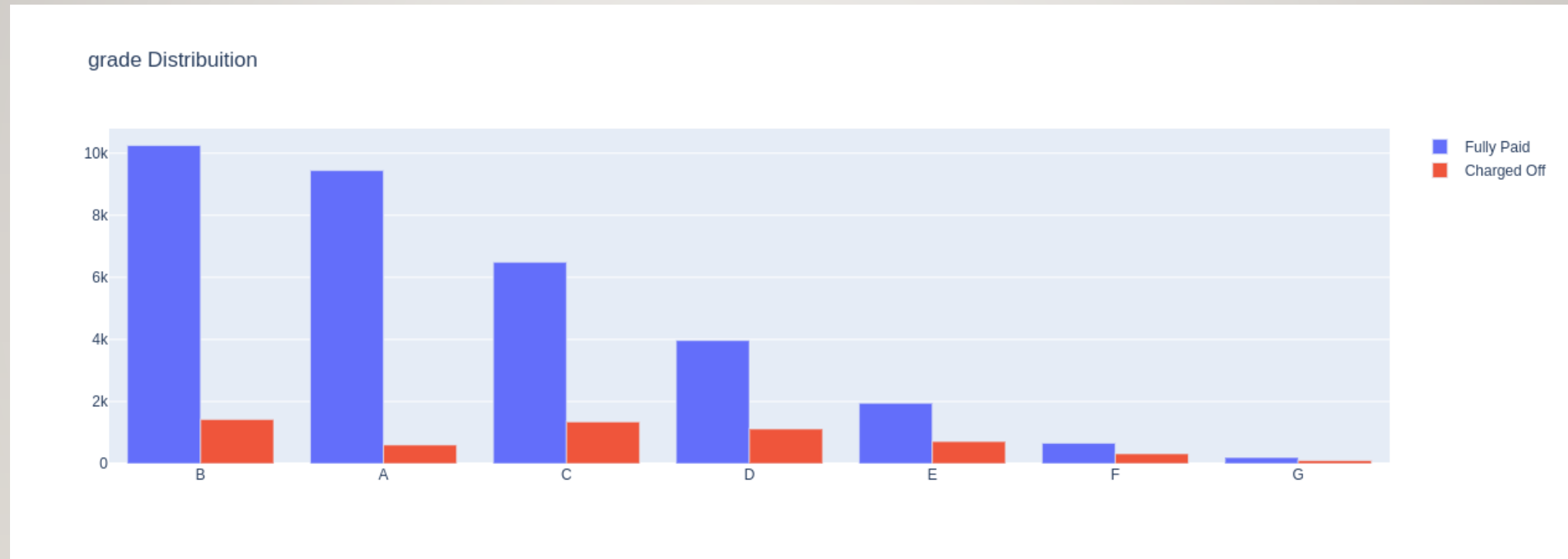


TERM ANALYSIS



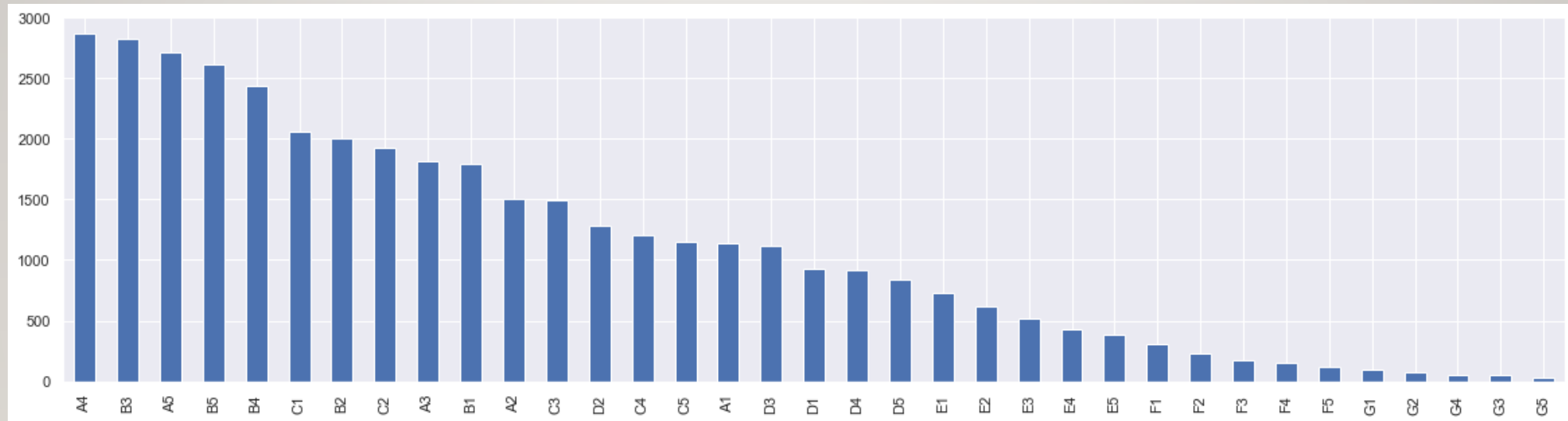
- Most of the customers have taken for 36 months loan tenure over the 60 months tenure

GRADE ANALYSIS



- Most of the customers who belongs to grade B has Fully paid their loans than compared to D, E, F and G grade customers

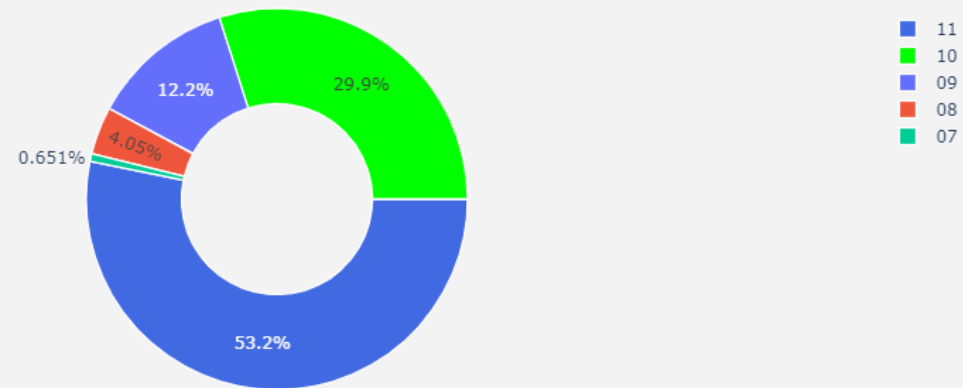
SUBGRADE ANALYSIS



- Top 5 Applicants are falling under A4, B3, A5, B5, B4 section.
- Most of the applicants are from B grade in loan sanctioned category.

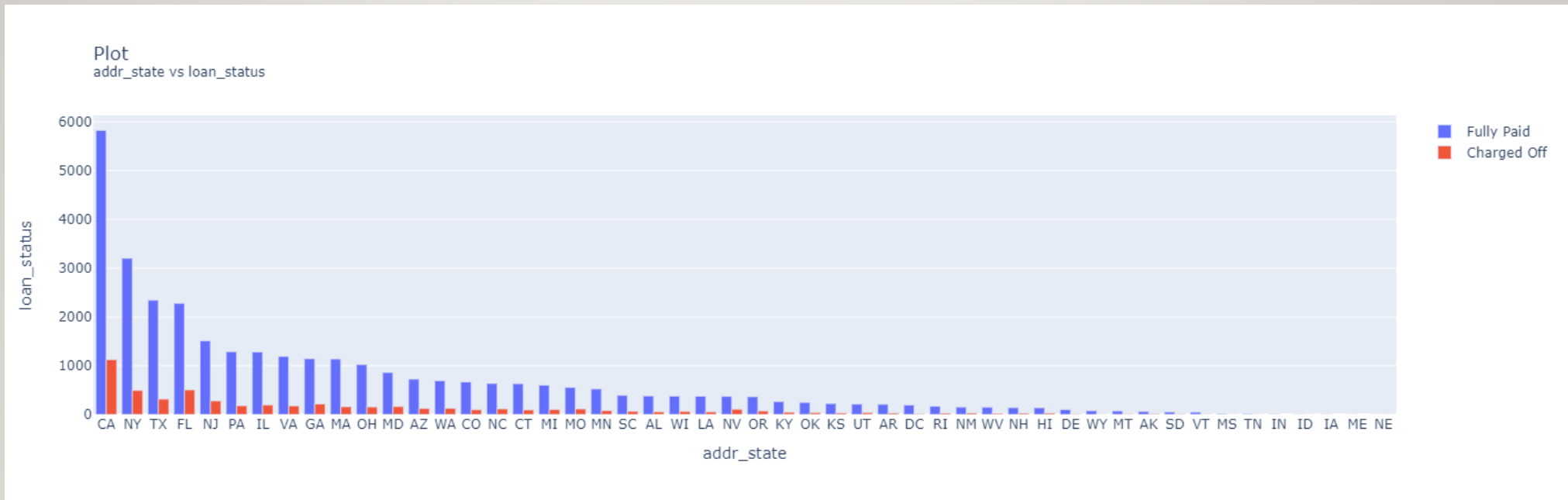
ISSUE DATE ANALYSIS

Year Wise Status Report



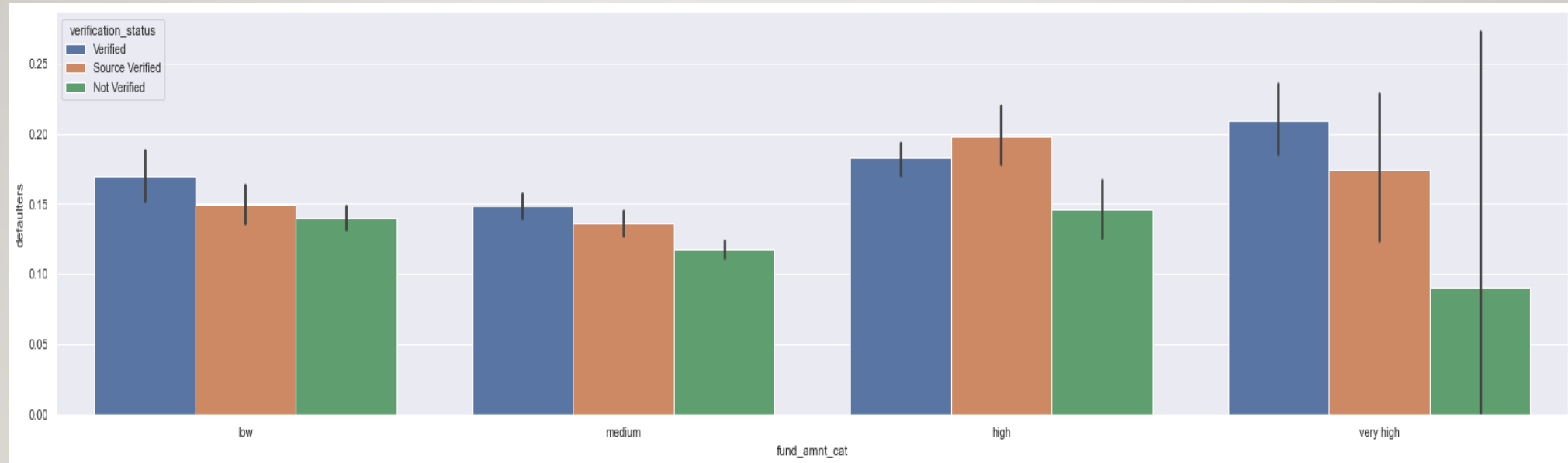
- Most of the applicants has applied loan in the 2011 which is around 53.2%.
- As the year goes on from 2007-2011 the number of loan applicant is increasing.

ADDRESS STATE ANALYSIS



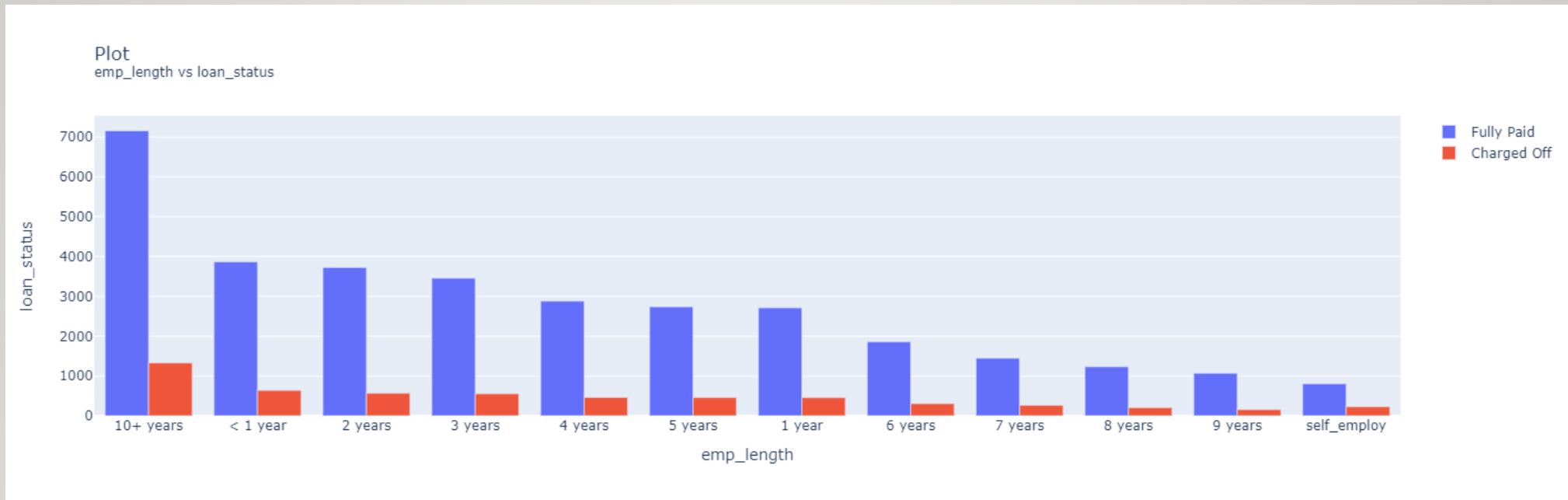
- As a state wise analysis California (CA) has the more applicants comparatively with the other states.
- Most of the Defaulters are coming from Nebraska (NE) because the bank hasn't received any of the loan amounts from that state.

FUNDED AMOUNT W.R.T VERIFICATION STATUS



- In Funded amount segment, when the applicants background verification is done tend to be defaulters.

EMPLOYEE LENGTH W.R.T LOAN STATUS



- Applicants with higher experience has taken more loans when compared with applicants having less experience.
- As the Charged Off representing in the graph says Defaulters among all the experienced applicants are consistent.

RECOMMENDATIONS

- ❖ LC needs to be more careful by adding few more additional check before approving the loan to small business category as they have higher Defaulter value.
- ❖ LC need to focus more on the small states like Nebraska(NE) as they have higher Defaulter rates which lead to loss of money.
- ❖ LC can approve the loan for the applicants who's applying for the purpose of Debt consolidation with high probability as they have less Defaulters.
- ❖ We Recommend to LC that the following variables are important to process the loan application based on the mentioned deriving features:

Predictor
Purpose of Loan
Term
Employment Length
Grade
Interest Rate
Home Ownership

THANK YOU