

LENDING CLUB CASE STUDY

EXPLORATORY DATA ANALYSIS

Group Members:
Pragyan Choudhury
Eswar Janjanam

Problem Statement

Company Profile:

The Lending Club is one of the largest online marketplace for the loans. Consumer can easily apply for the loans in various categories. The loans are funded by Investors based on the Credit ratings of the Customers.

Problem:

Lots of loans are Defaulted and Charged Off. The company wants to understand the driving factors behind loan default. The company can utilize this knowledge for its portfolio and risk assessment to make better decisions.

Objective:

As a Data Analyst, perform EDA to understand how consumer attributes and loan attributes influence the tendency of loans default.

Data Availability

The data available contains the information about past loan applicants and whether they 'defaulted' or not. The dataset does not contain any information for the loans which were 'NOT Approved'.

The provided dataset can be analyzed using Exploratory Data Analysis (EDA) technique.

Datasets provided:

- loan.csv

The historical data for the approved loans which have either been fully-paid or Charged Off. It is in comma separated value (csv) file format.

- Data_Dictionary.xlsx

Contains information about the various columns. It is required for Data Understanding. It is in excel format.

Analysis Approach



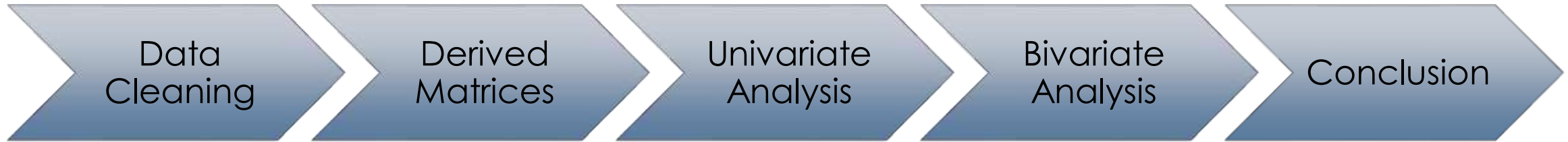
Data Cleaning:

- Drop columns with all null values/ single categorical value / random values
- Impute null values for important columns
- Convert values into proper data.

Derived Matrices:

- Create derived Metrics for year and month for date.
- Convert Quantitative values into categories.
- Calculate ratios and percent and other derived metrics.

Analysis Approach Contd.



Univariate Analysis:

- Check distribution and frequencies for numerical and categorical data
- Analyze the variable against segments of other variables.

Bivariate Analysis:

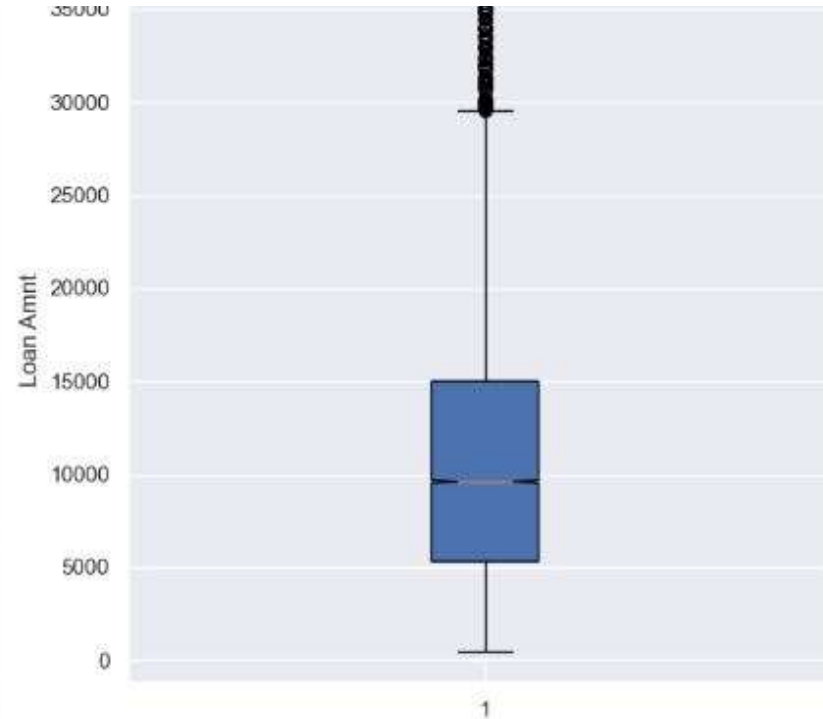
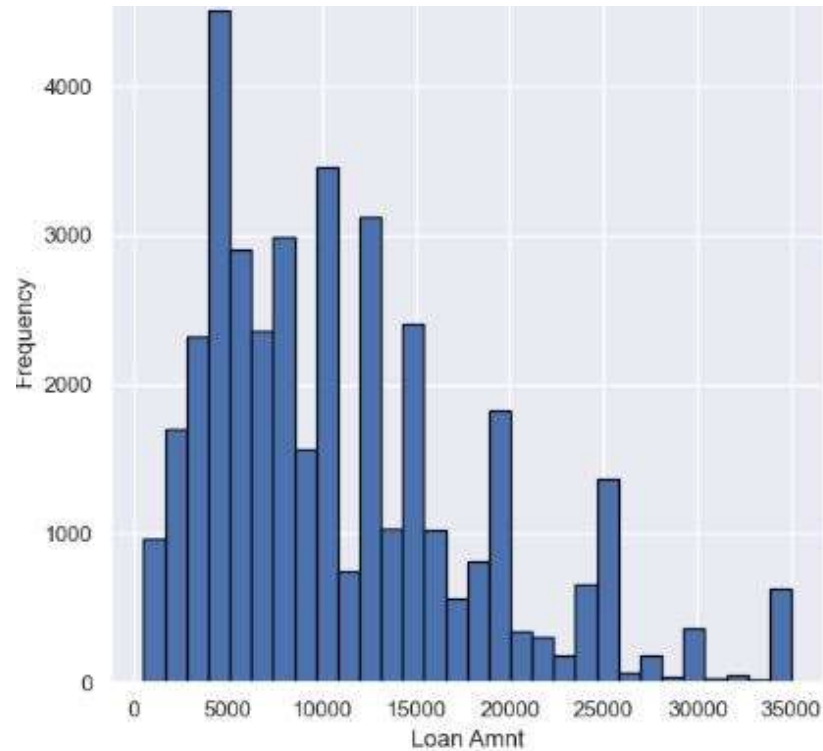
- Analyze how 2 variables influence each other.
- Check for the correlations between the variables.

Conclusion:

- Publish insights and observations from the analysis.

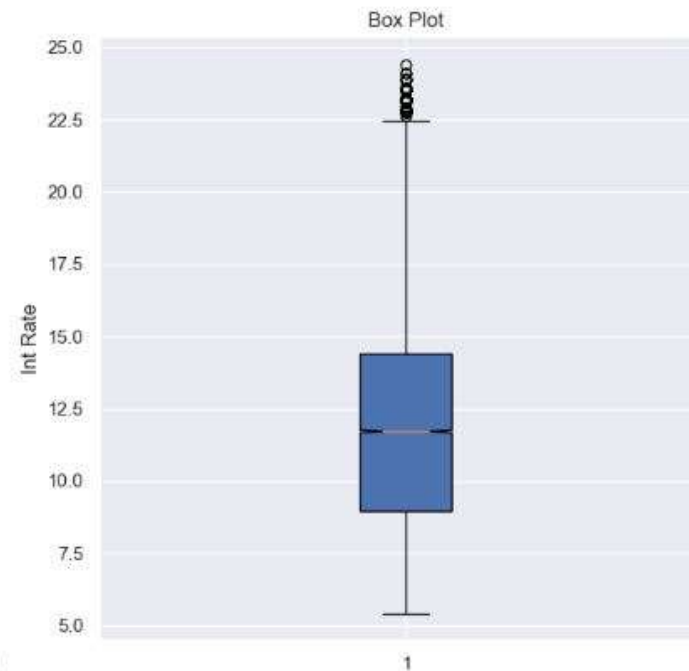
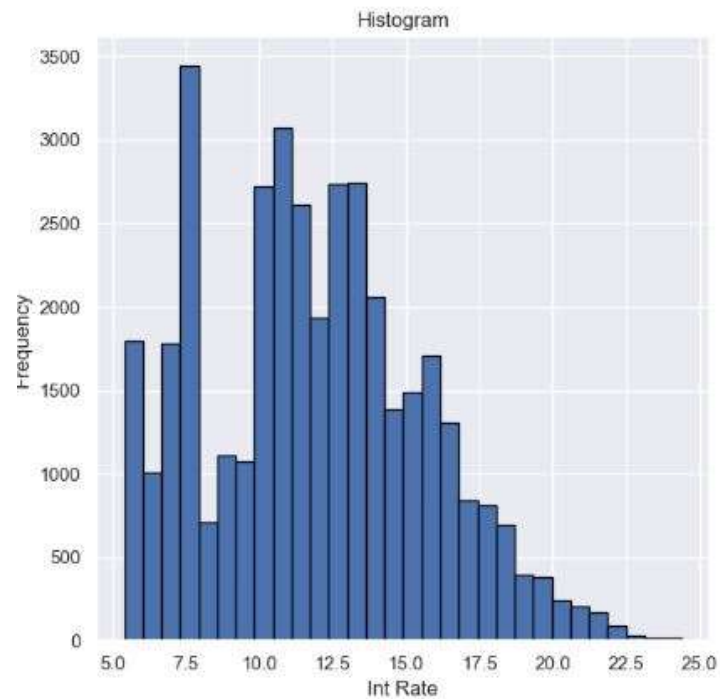
Analysis

Loan Amount: Most of the people opted for loans took the loan amount in the range \$5000 to \$15000 where maximum people took loan amount of around \$5000. The median amount is around \$10000.



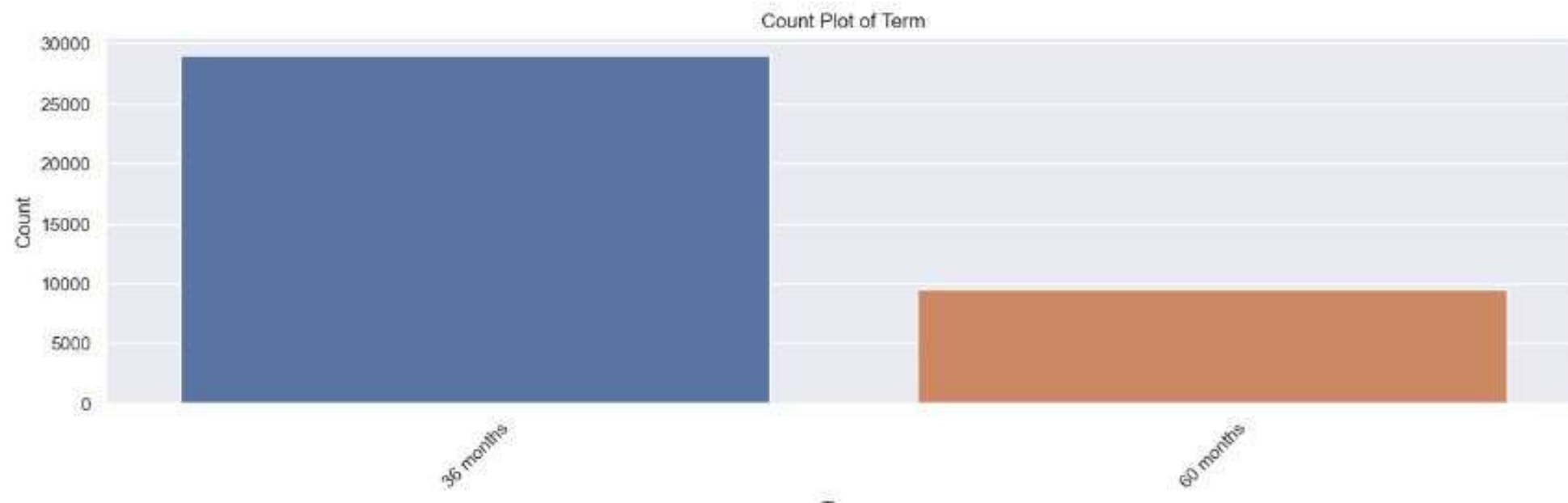
Analysis Contd.

Int_rate: The interest for most of the loan is in the range of around 8.5% to 14.5% with a median interest of around 12%.



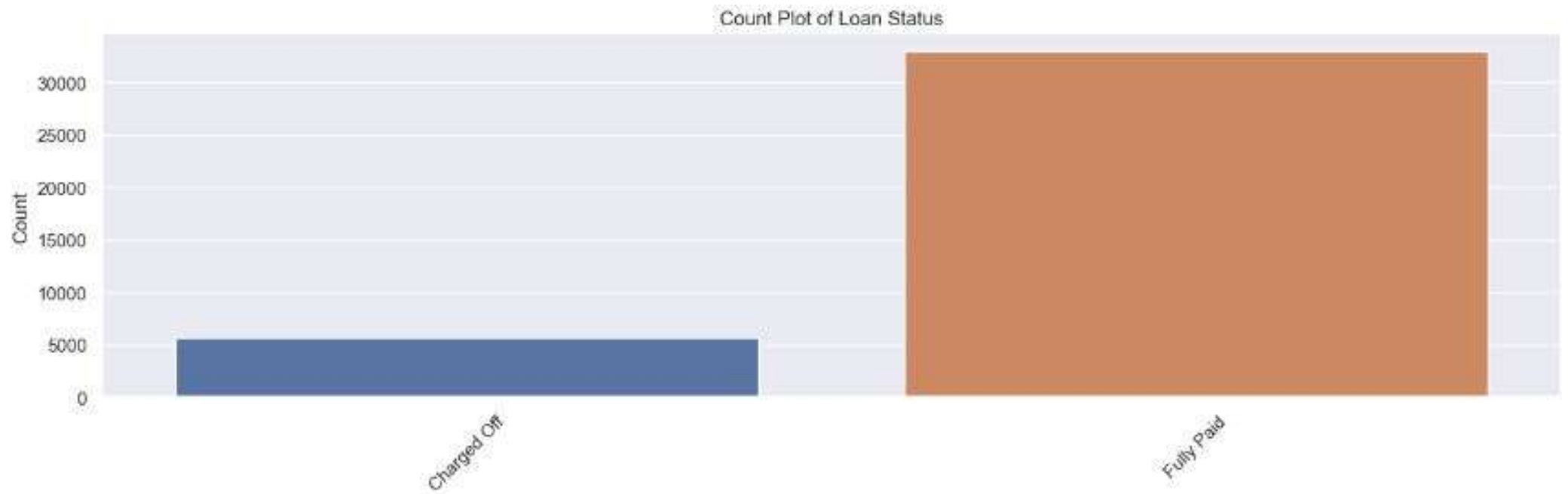
Analysis Contd.

term: There are two types of loans. One is of 3 years and the other is 5 years where most of the loans are 3 years only.



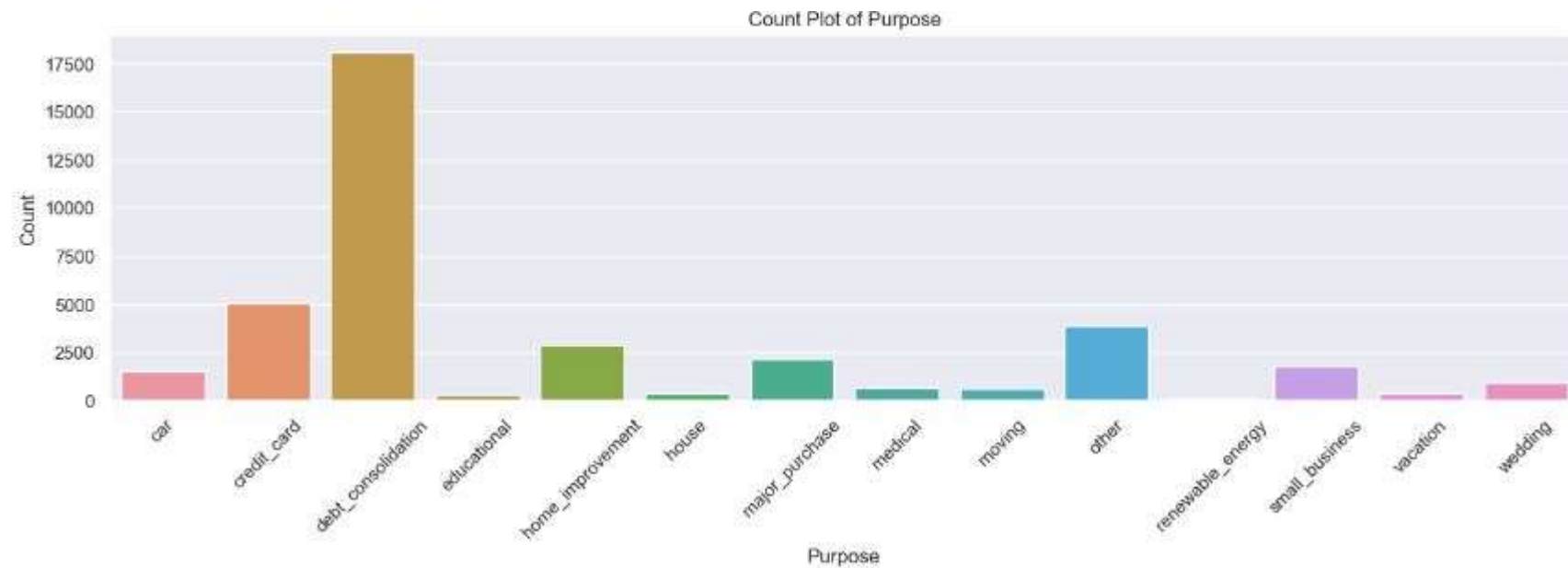
Analysis Contd.

loan_status: Almost 15% of the issued loans are Charged Off.



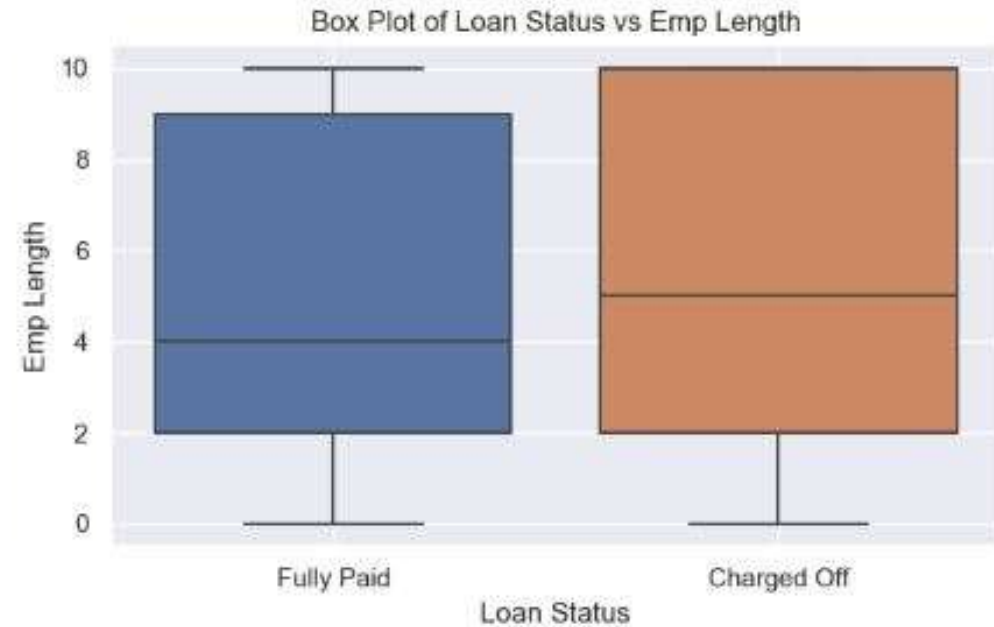
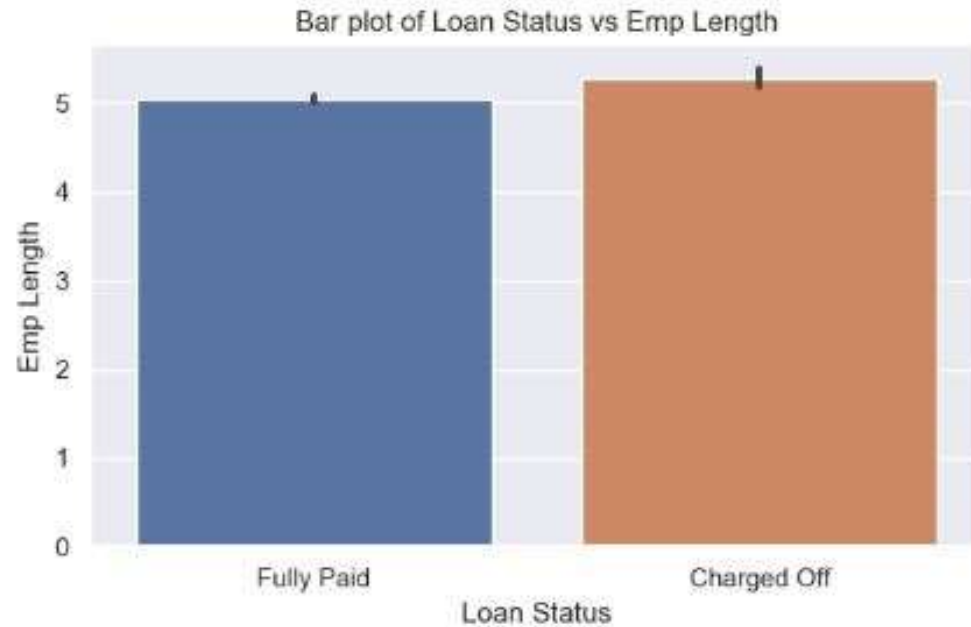
Analysis Contd.

purpose: Most of the people who took loans are for Debt Consolidation followed by credit card bill repayment..



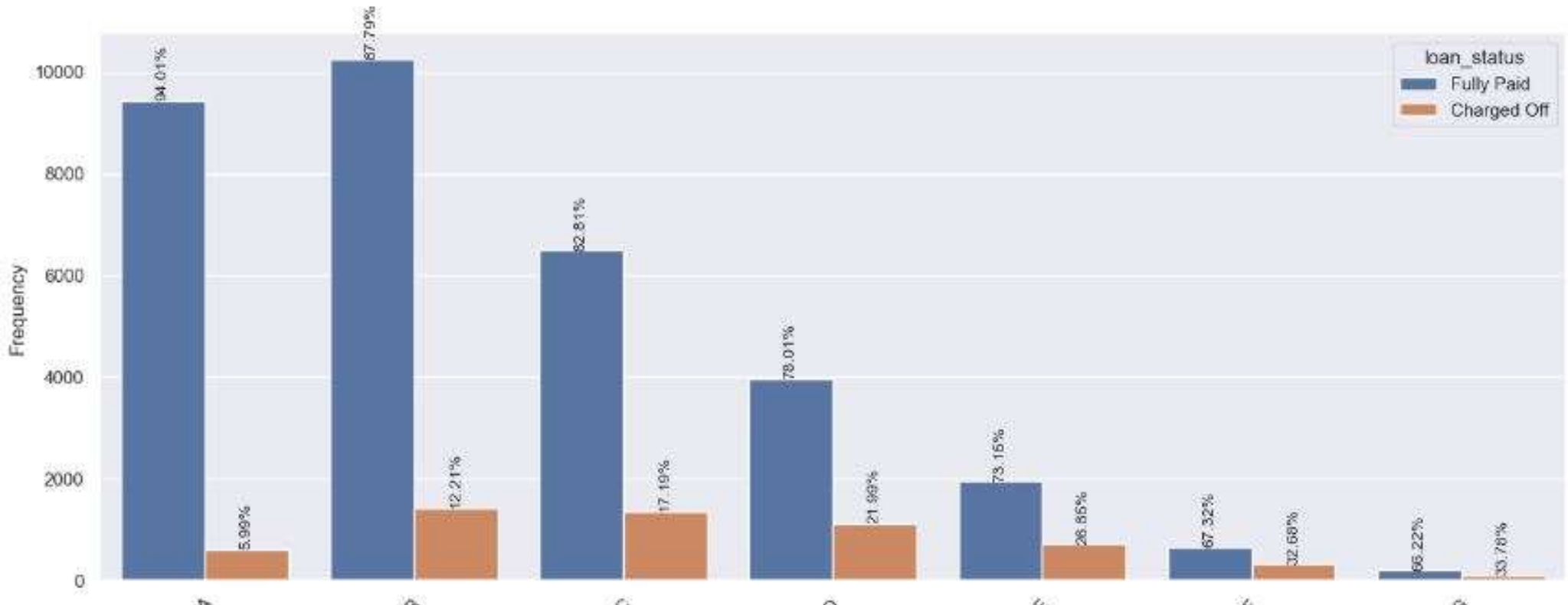
Analysis Contd.

emp_length Vs Loan Status: The median employment length for fully paid is 4 years and for charged off is 5 years. Most of the loans are taken by people having employment length greater than 10 years and charged off loans is also higher in 10+ years employment length.



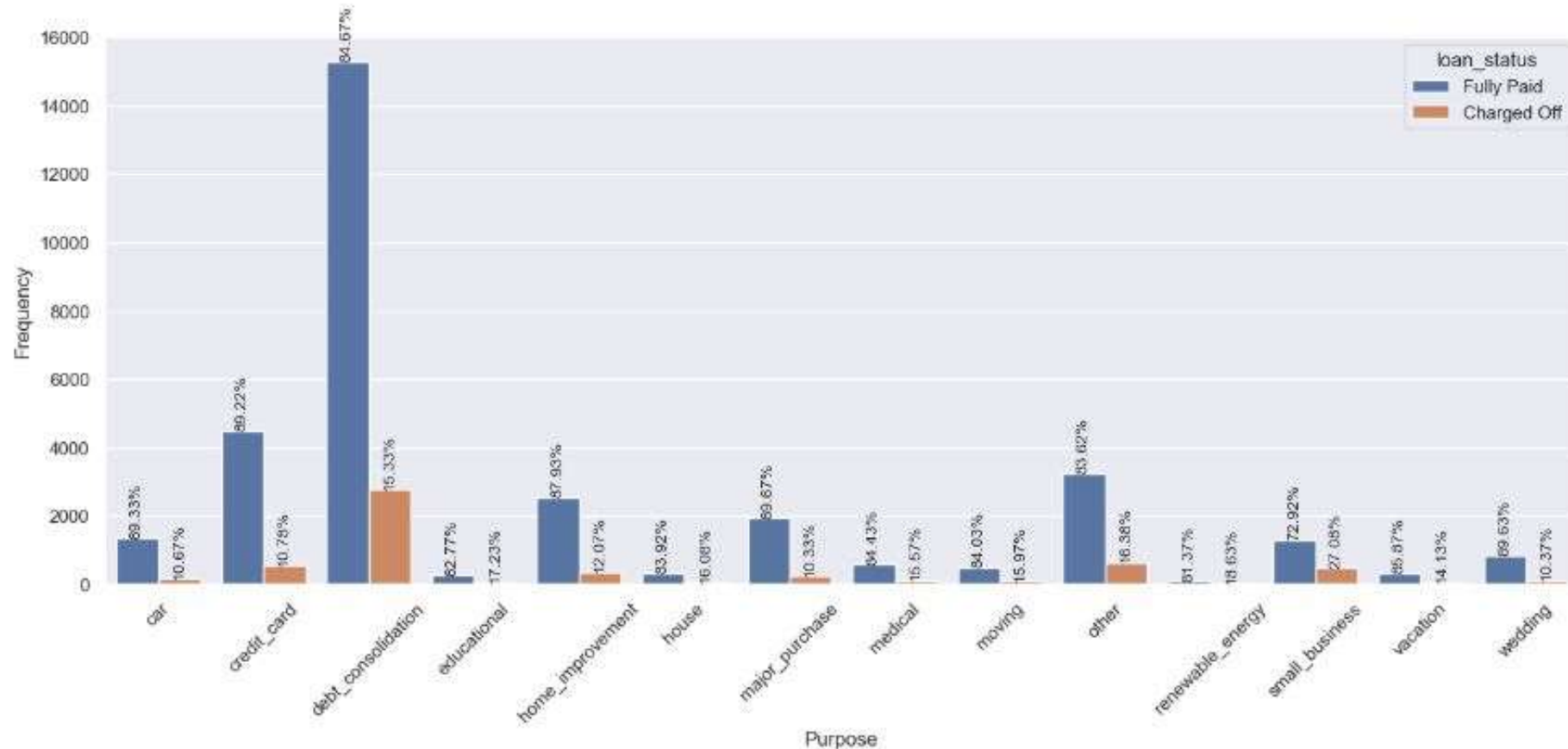
Analysis Contd.

Grade Vs Loan Status : Charged Off loans are comparatively less in A and B grade customers and more in G grade which is almost 34%.



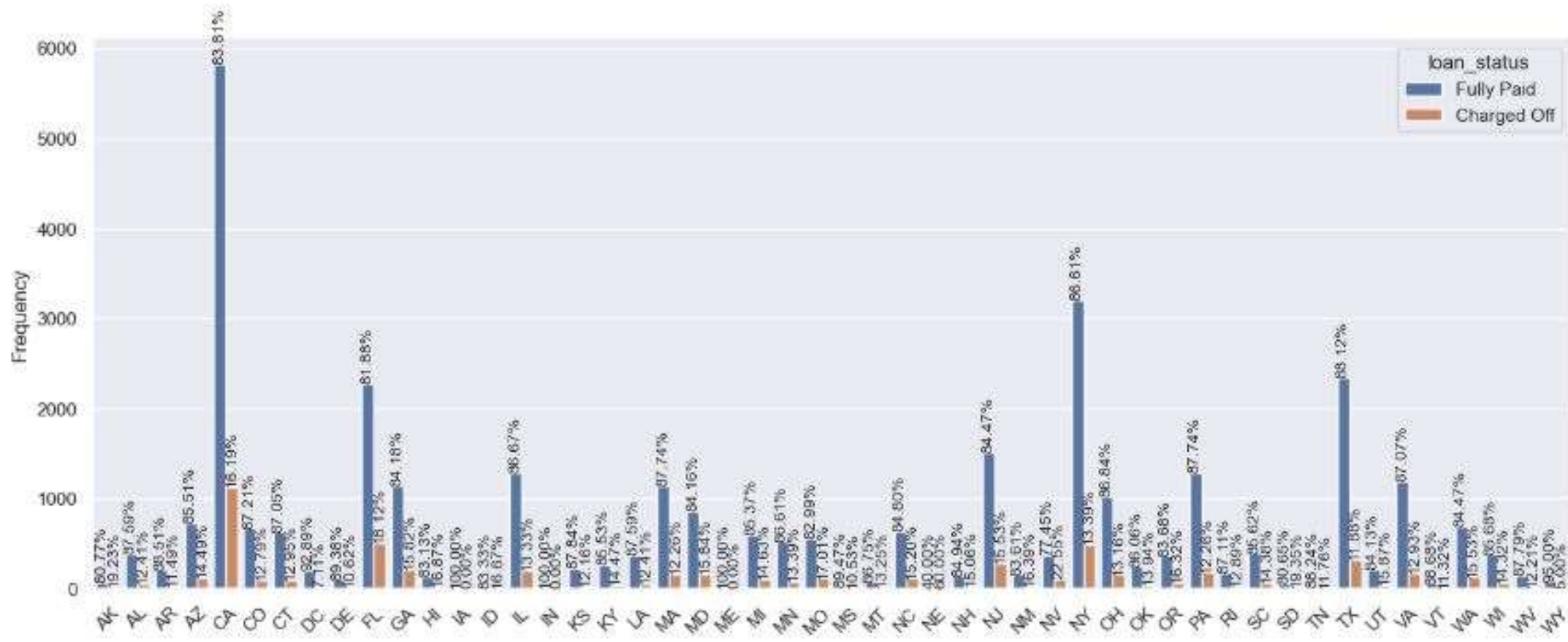
Analysis Contd.

Purpose Vs Loan Status : Most of the loans are taken for debt consolidation. But the percentage of charged off is highest in small business which is 27% followed by renewable energy and educational loans.



Analysis Contd.

Addr_state Vs Loan Status : Though California(CA) people took more loans, 60% of loans given to Nebraska(NE) people are charged off. 22.5% of loans given to Nevada(NV) people are charged off..



Recommendations

- For higher loan amount, there is higher chance of charged off.
- For higher interest rates is directly related to charged off.
- The loans given to people of low annual income have high risk of getting charged off.
- If dti is more for a person, his loan have high chance of getting charged off.
- If the term of the loan is more, probability to charged off is more
- Loan issued to G Grade are most likely to be charged off
- Loans given to small business, renewable energy and education are more likely to be charged off
- Loans given to people from Nebraska are 60% charged off. One loan in every five loans given to Nevada people is likely to be charged off.