

Lending Club EDA

By,
Tritam Mohanty



Business Problem

- Overview:
 - 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.
- Expectations
 - Derive driving factors for loan default and risk assessment
- This Data Analysis was to explore the business driven and data driven conclusion to discover the driving factors for loan default.



Approach to Problem

- Data Understanding.
- Data Cleanup
- Deriving New Columns.
- Performing Univariate Analysis.
- Performing Segmented Univariate Analysis.
- Performing Bivariate Analysis.



Understanding Data

- 39,717 records and 111 features
- Removal of Data
 - Duplicate Records
 - 50% of columns with no values
 - Columns with 60% or more missing values
 - Columns with less relevance for the analysis
 - Outliers
- Formatting Data
 - Fixing Columns
 - Standardize Precision
 - Correcting Data Types and invalid values

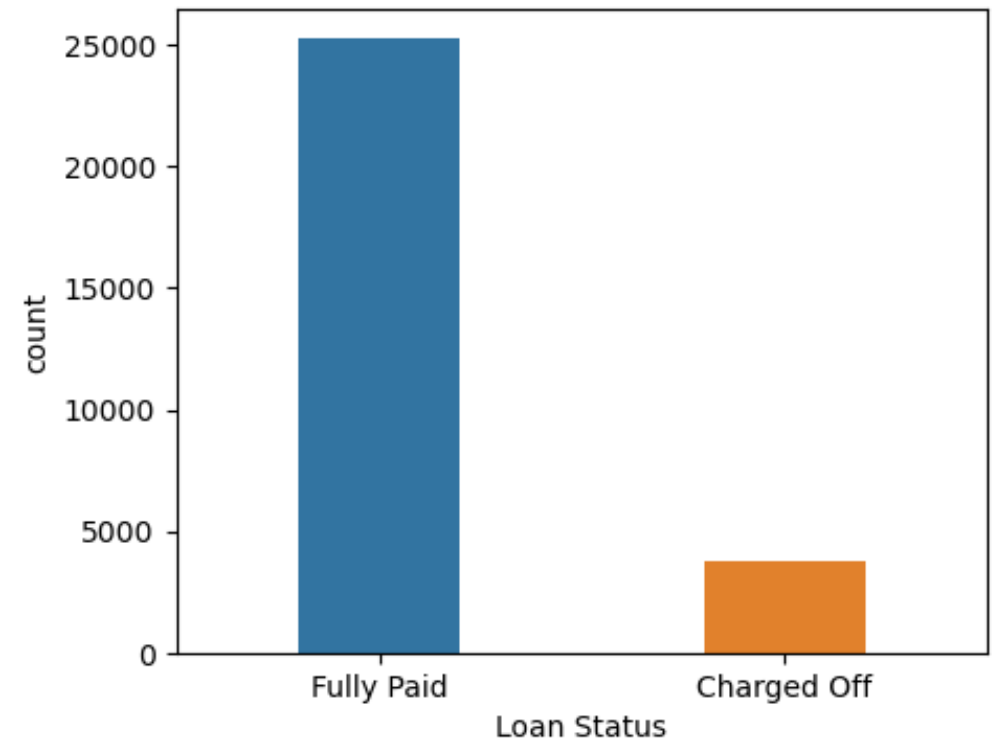


Key Business Metrics

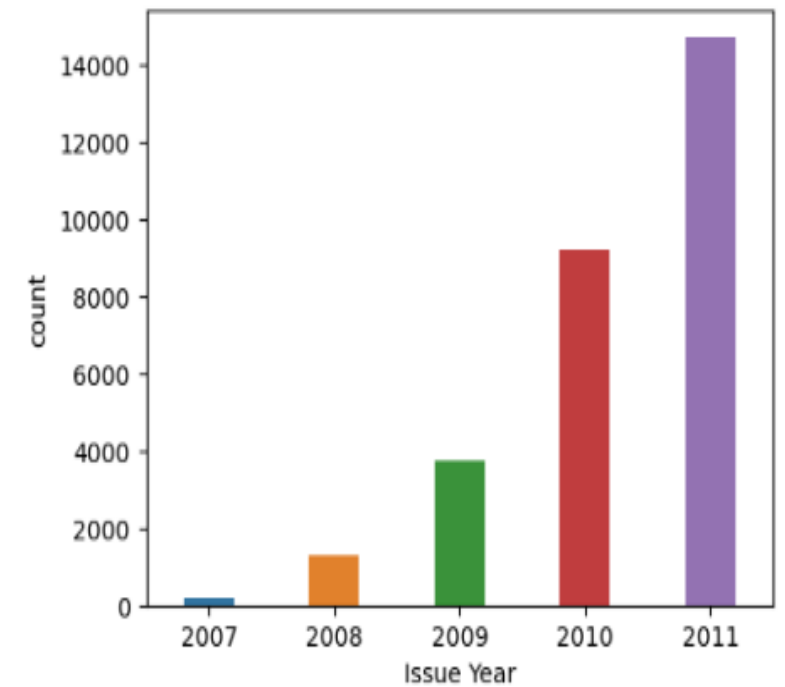
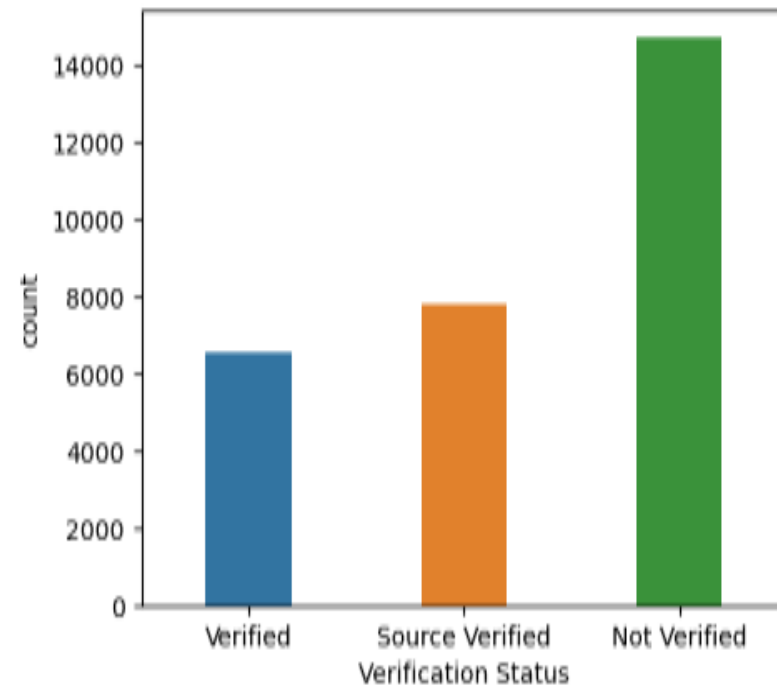
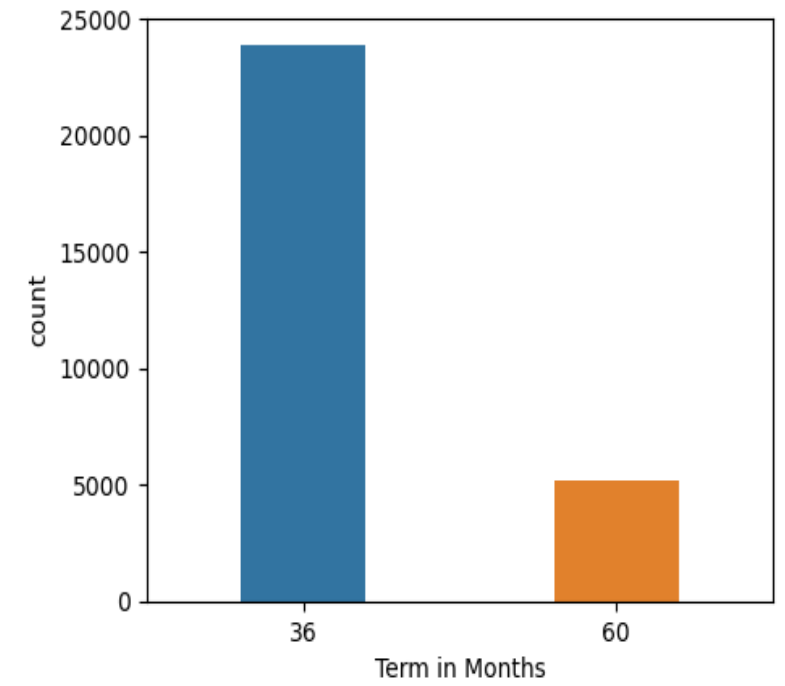
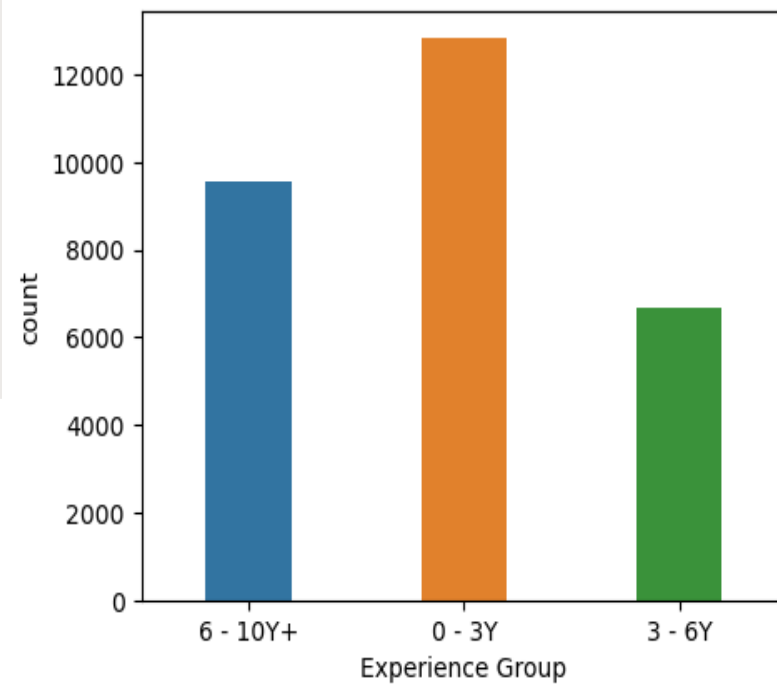
- Positive Metrics
 - 87% fully paid
 - Increase in business year on year
 - 57.49% of the funded amount could be recovered for Charged Off loans
 - Profit of 17.31% from Fully Paid loans
 - 96.14% match of funded amount to loan amount
- Negative Metrics
 - 13% Charged off loans
 - 50% borrowers not verified
 - Avg Charged Off from 2007 to 2011 is 13.6%
 - 42.51% of the Charged Off loan were not recovered
 - 3.85% when loan amount did not match funded amount

Univariate Analysis

- 87% Full paid and 13% Charged off loans
- 36.1% (Maximum) borrowers in interest rate < 10%
- Maximum borrowers in 0-50K income group (49.05%) income group and 0-3Y(44.13%) employment length
- 82.1% borrowers fall in 36 months term and 17.8% in 60 months term
- 50% borrowers are not verified
- Grades A and B have minimum defaults
- Loans issued show huge growth year on year
- More loans taken towards Year End may be due to Festive season



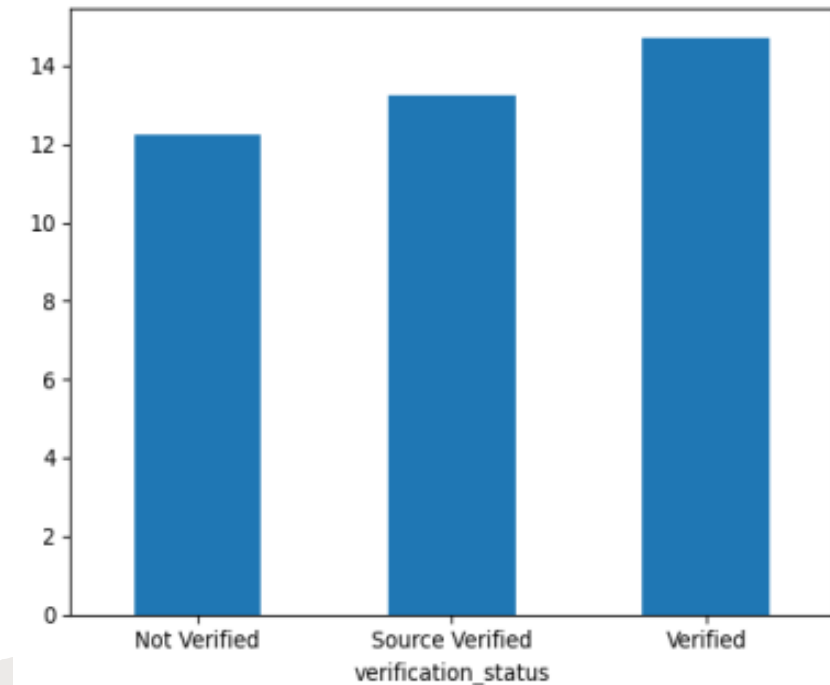
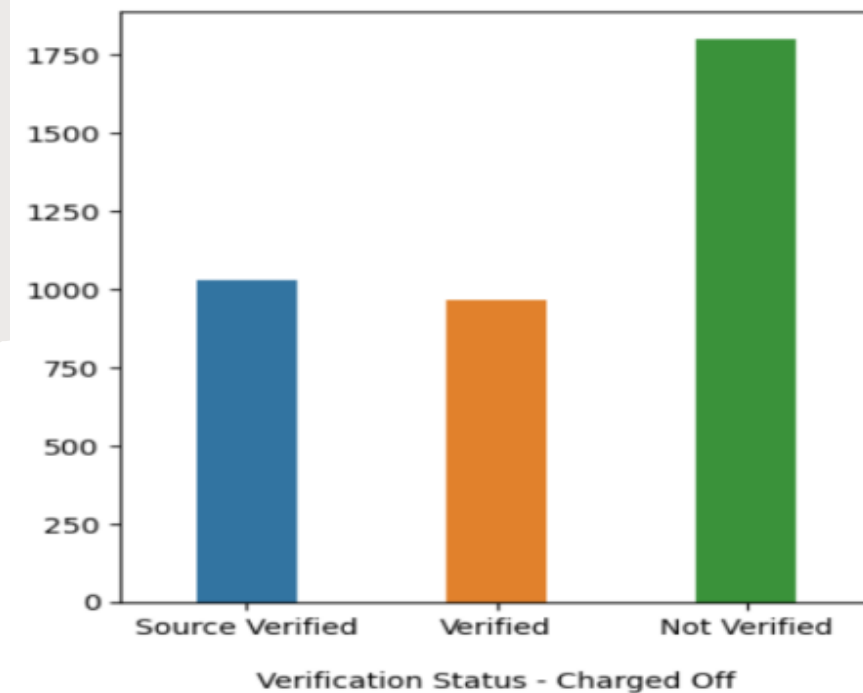
Univariate Analysis



Bivariate Analysis

Verification status / Loan Status

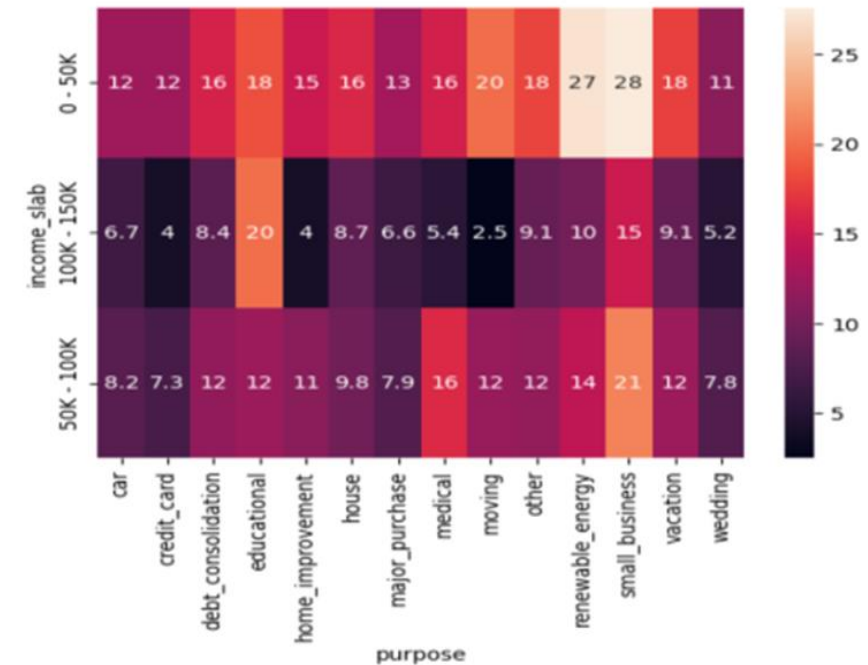
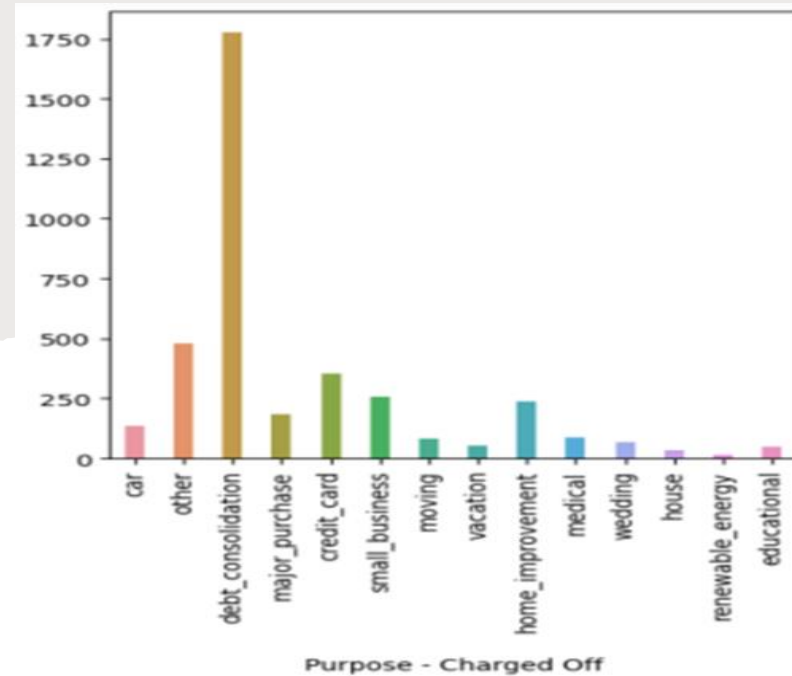
- More number of charged off loans from Not Verified Applicants
- Verified Applicants default 2.7% more when compared against total loans in Verified category



Bivariate Analysis

Purpose

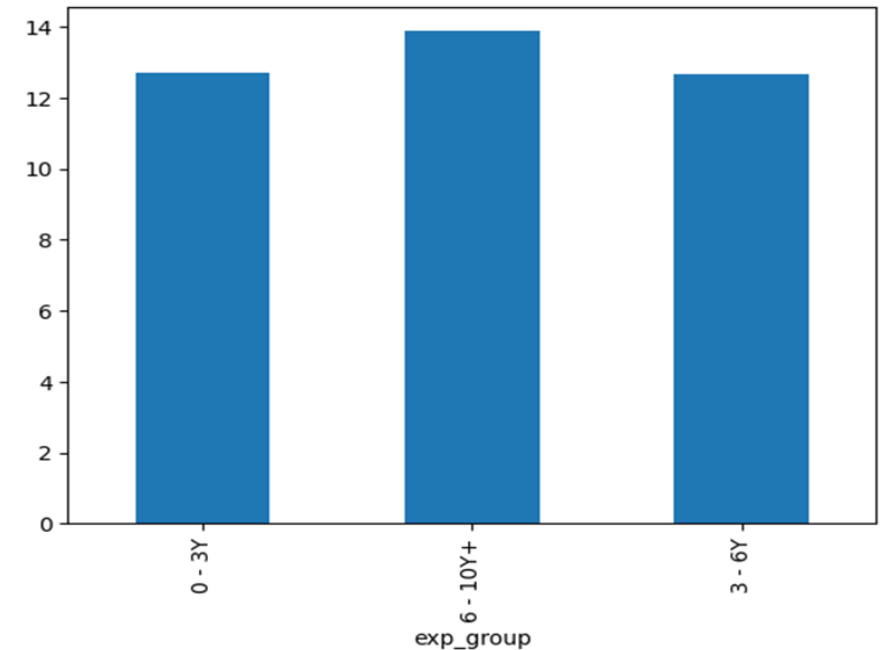
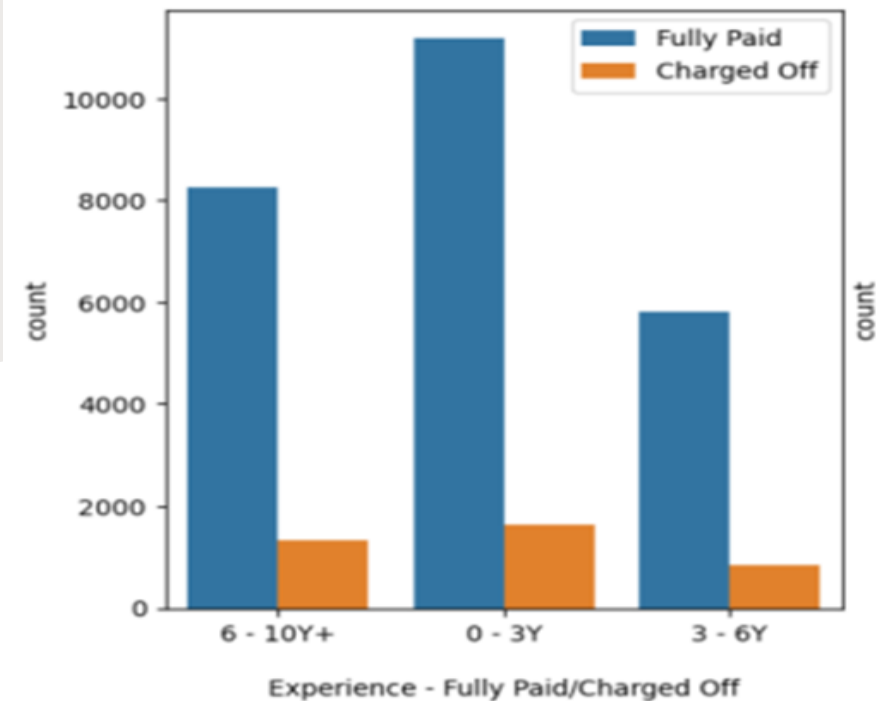
- Small businesses, education and renewable energy are the top 3 defaults in loan categories
- Small businesses most likely default in income slabs below 50K and higher than 100K
- For income range 50-100K, education is the top contributor followed by small businesses



Bivariate Analysis

Employment Length

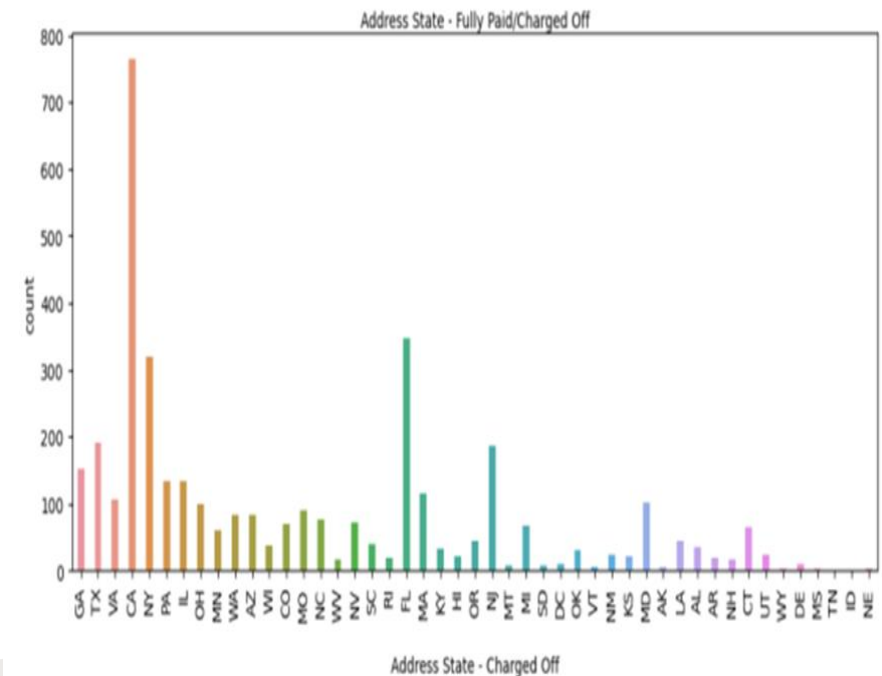
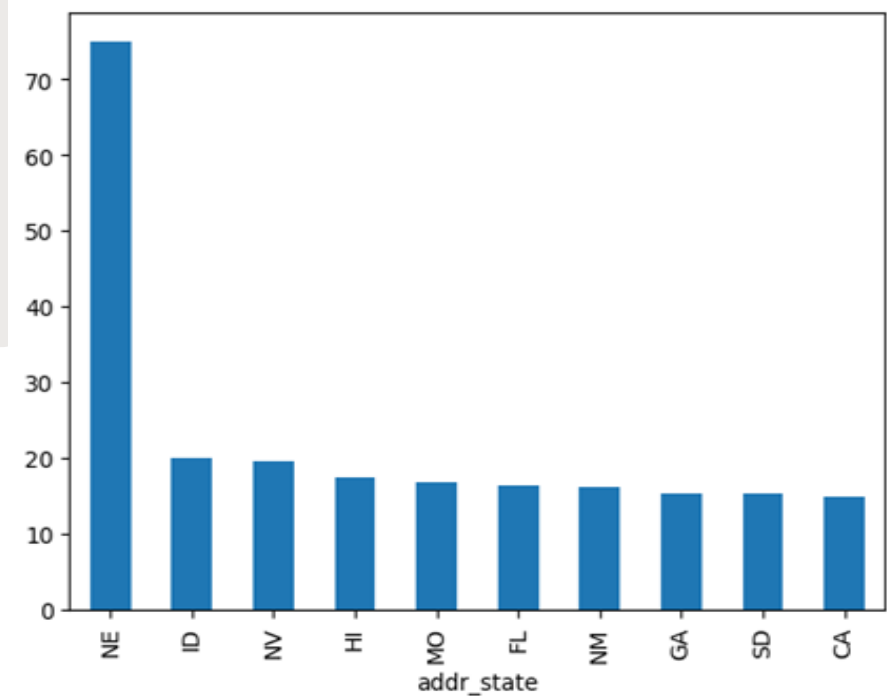
- We have more borrowers who have fully paid and charged off in **0-3 Years** employment length
- Borrowers in **6-10 Years** employment length default more when compared with total borrowers in the same group
- More default comes from large fund amounts (12K or more) and the surprisingly amounts below 3K also that they borrow and are not able to repay



Bivariate Analysis

State

- We have more borrowers who have fully paid and charged off from **CA and FL**
- Borrowers in **NE, ID and NV** default more when compared with total borrowers in the same area
- More verification preferred for top 10 default regions marked in the graph below



Recommendations



Verify applicants before providing loans



Consider revisiting or improving the verification process as verified applicants also default more



Recommend applicants from Grades A and B. Applicants from Grade C and above have higher default rates



More caution and verification is recommended for funding higher amounts with a tenure of 60 months as they tend to default more



Recommend more controls while funding applicants with multiple open accounts as they default more



Recommend lower interest rates with thorough evaluation of other factors like annual income etc. before choosing higher interest rates

Recommendations



Recommend caution while funding small businesses in general and specially in income range below 50K



Loans borrowed for education and renewable energy also default more. More verification/controls to be exercised in these categories



Loans funded with amounts 12K and above default more in all income groups specially in 6-10Y+. A better verification process needs to be exercised here



Recommend Not to fund applicants with Inquiries in last 6 months



Recommend Not to consider applicants with public records and public record bankruptcies



Recommend more verification for applicants coming from Nebraska, Nevada Idaho, California and Florida

The background of the image is a complex network graph. It consists of numerous small, semi-transparent nodes (dots) in shades of grey and light pink, interconnected by a dense web of thin, light grey lines. The nodes are distributed across the entire frame, creating a textured, interconnected appearance.

Thank You