

Lending club case study

Manimaran October 11, 2023

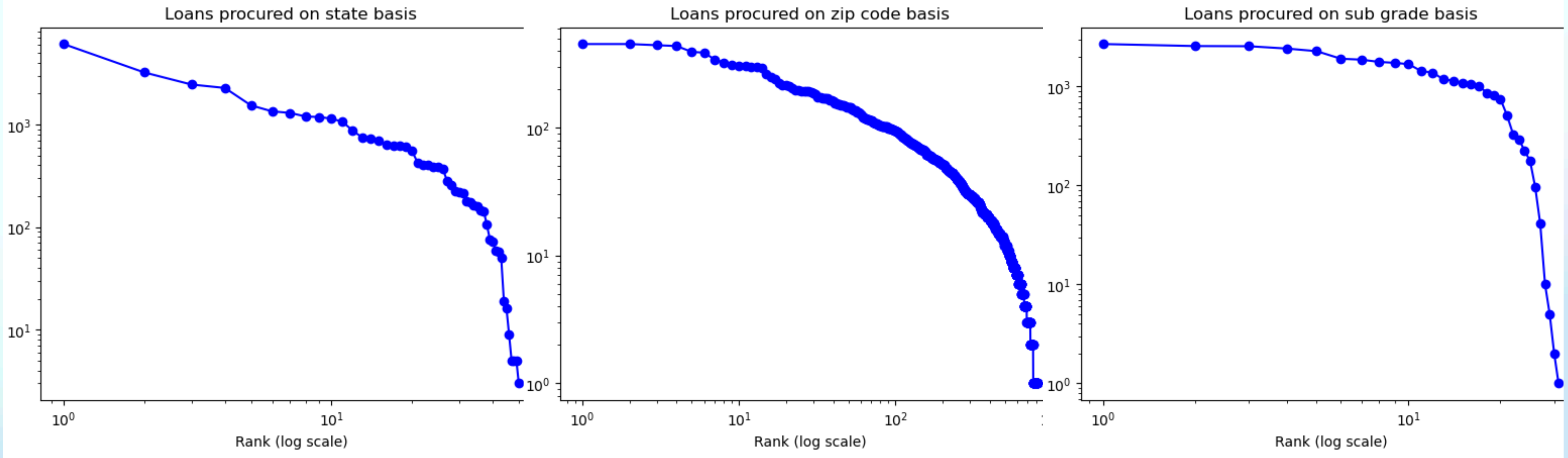
Given the data set of a company called 'Lending Club', a consumer finance company, EDA (Exploratory Data Analysis) needs to be performed on the given data set to help the company make decisions for loan approvals based on the applicants' profile.

- The approach is fairly simple.
- Perform data cleaning and manipulation to make the data suitable for analysis, remove the outliers.
- Perform data analysis on the data set and arrive at insights.
- The analysis range from adding derived metrics, performing univariate, segmented univariate and bivariate analysis.
- Visual presentation of the results.

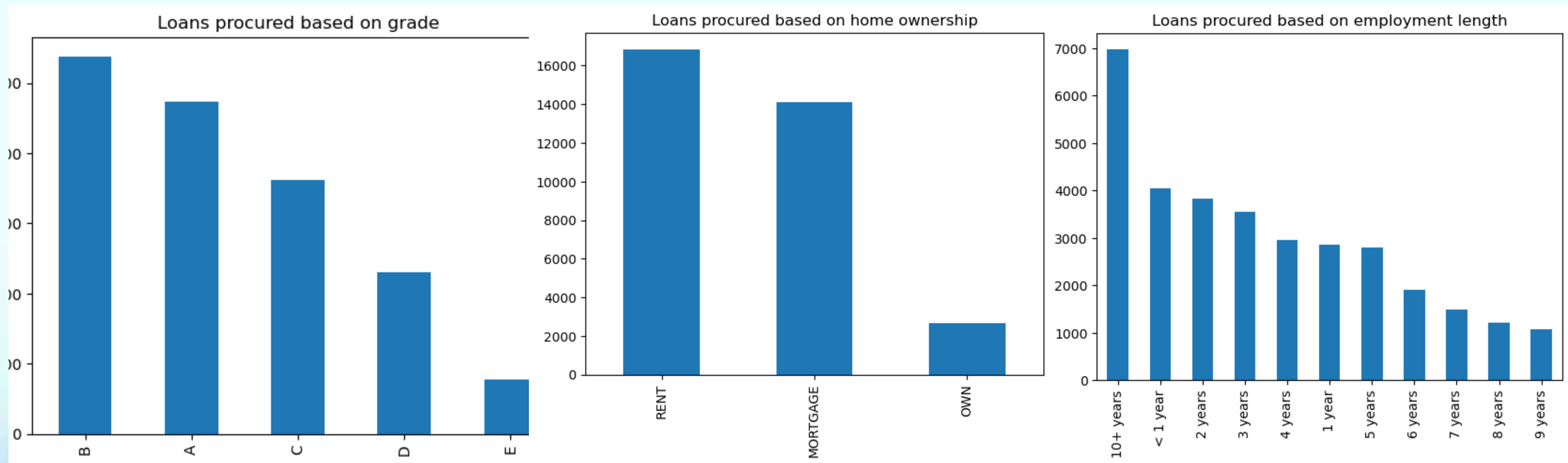
Univariate Analysis

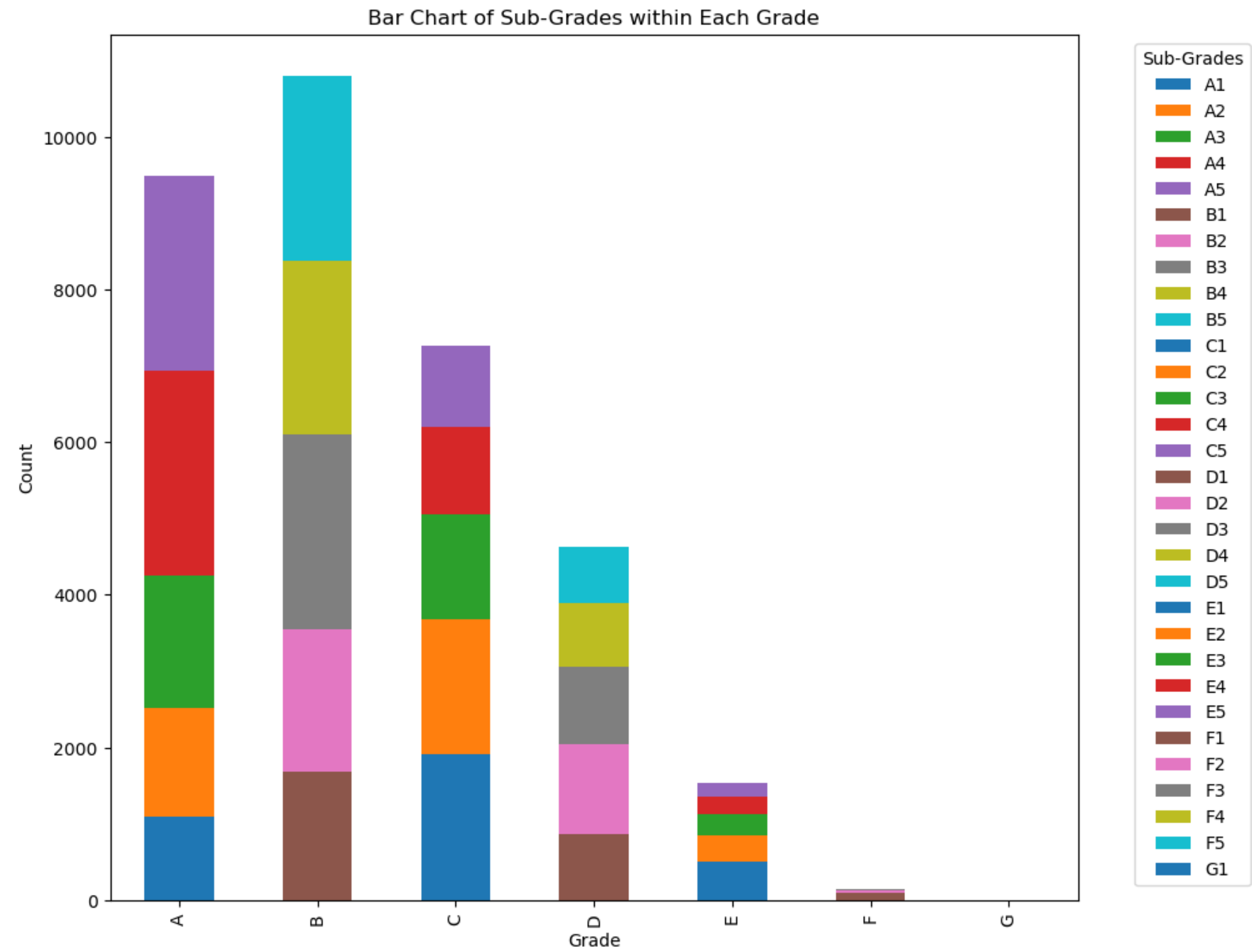
Univariate analysis - Unordered categorical variables

- Univariate analysis is performed on the loans procured based on the state, zip code and sub-grade and then plotted on a log-log scale
- Further univariate analysis is performed on loans procured based on the grade, home ownership, employment length, income verification status and plotted using bar graphs.
- Stacked bar charts are plotted on the number of loans procured based on the zip codes in each state and sub-grades within each grade
- The log-log scale can be seen in the next slide.



Univariate analysis on grade, home ownership and employment length can be seen in the following slide

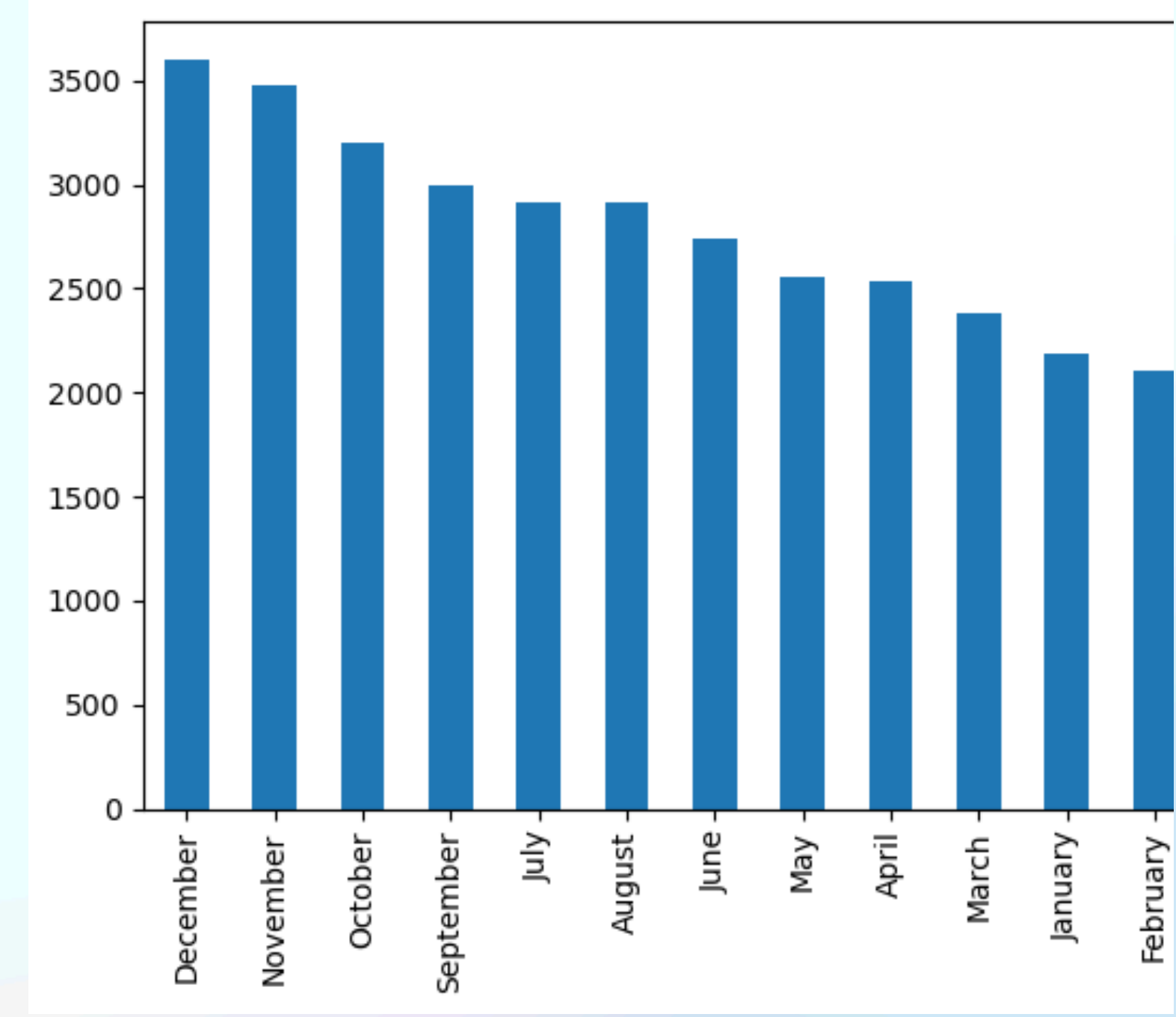
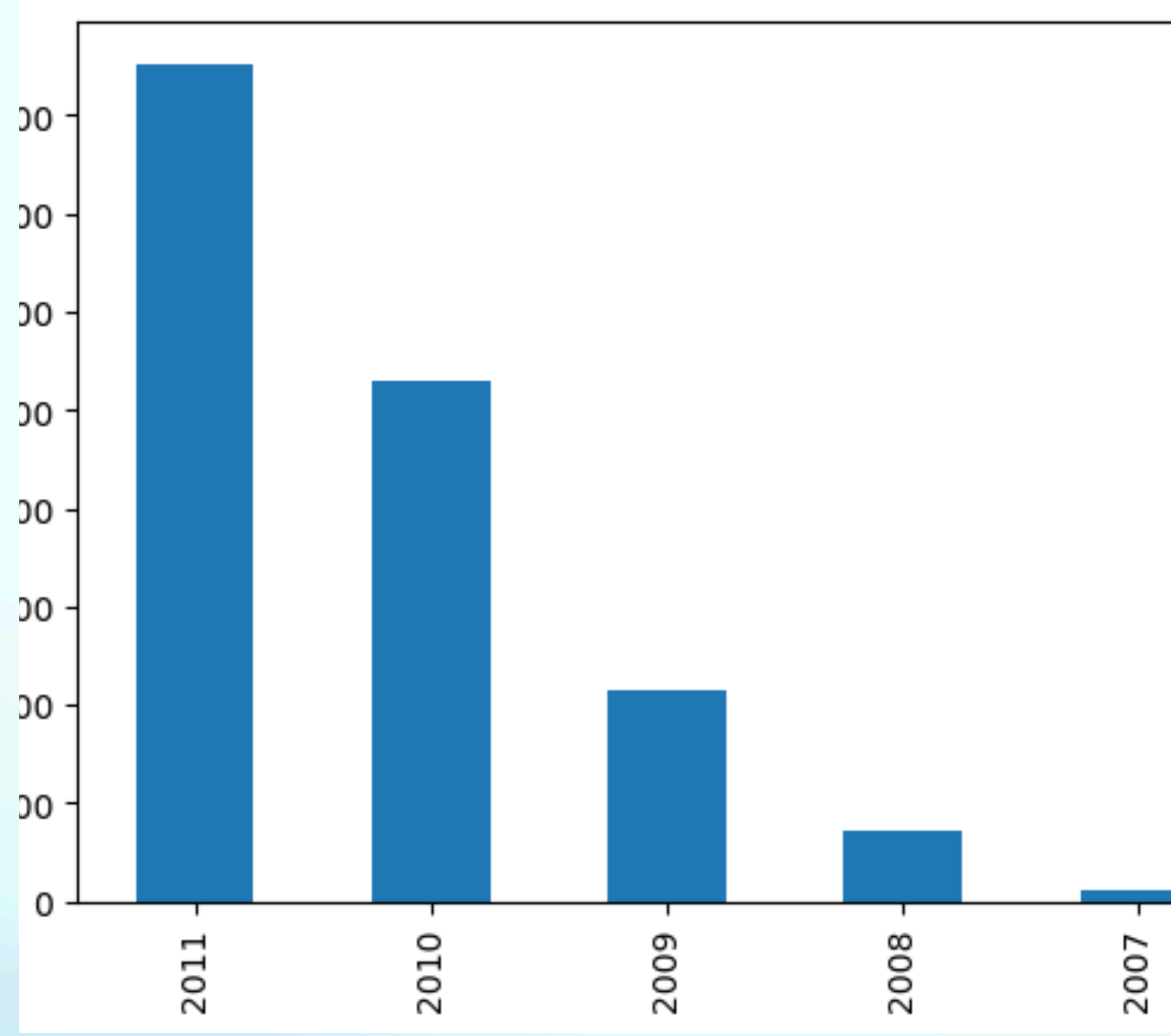


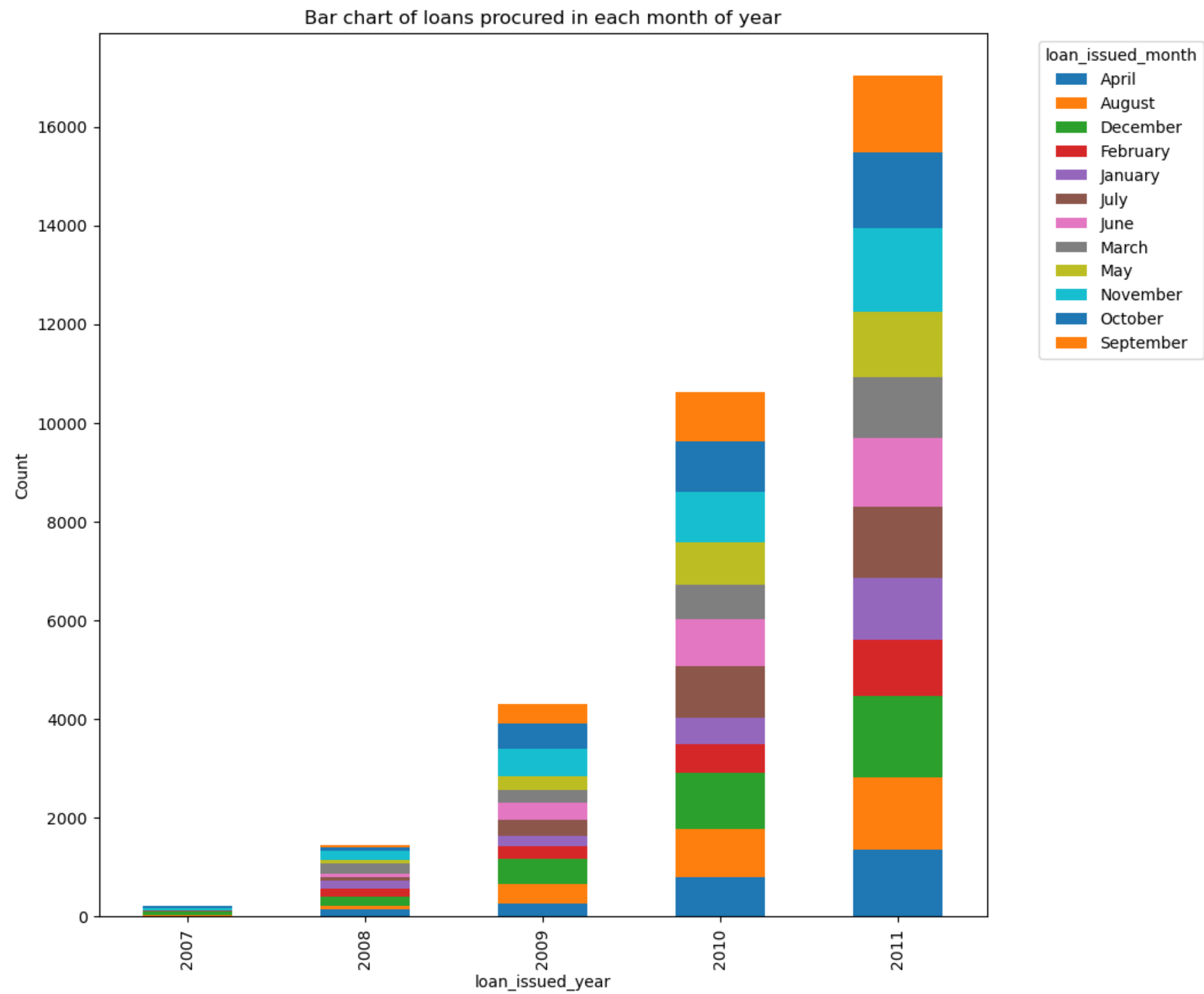


Univariate Analysis

Univariate analysis - ordered variables

- Univariate analysis is performed on the loans procured based on the loan issued year, loan issued month,
- Graphs in the following slide

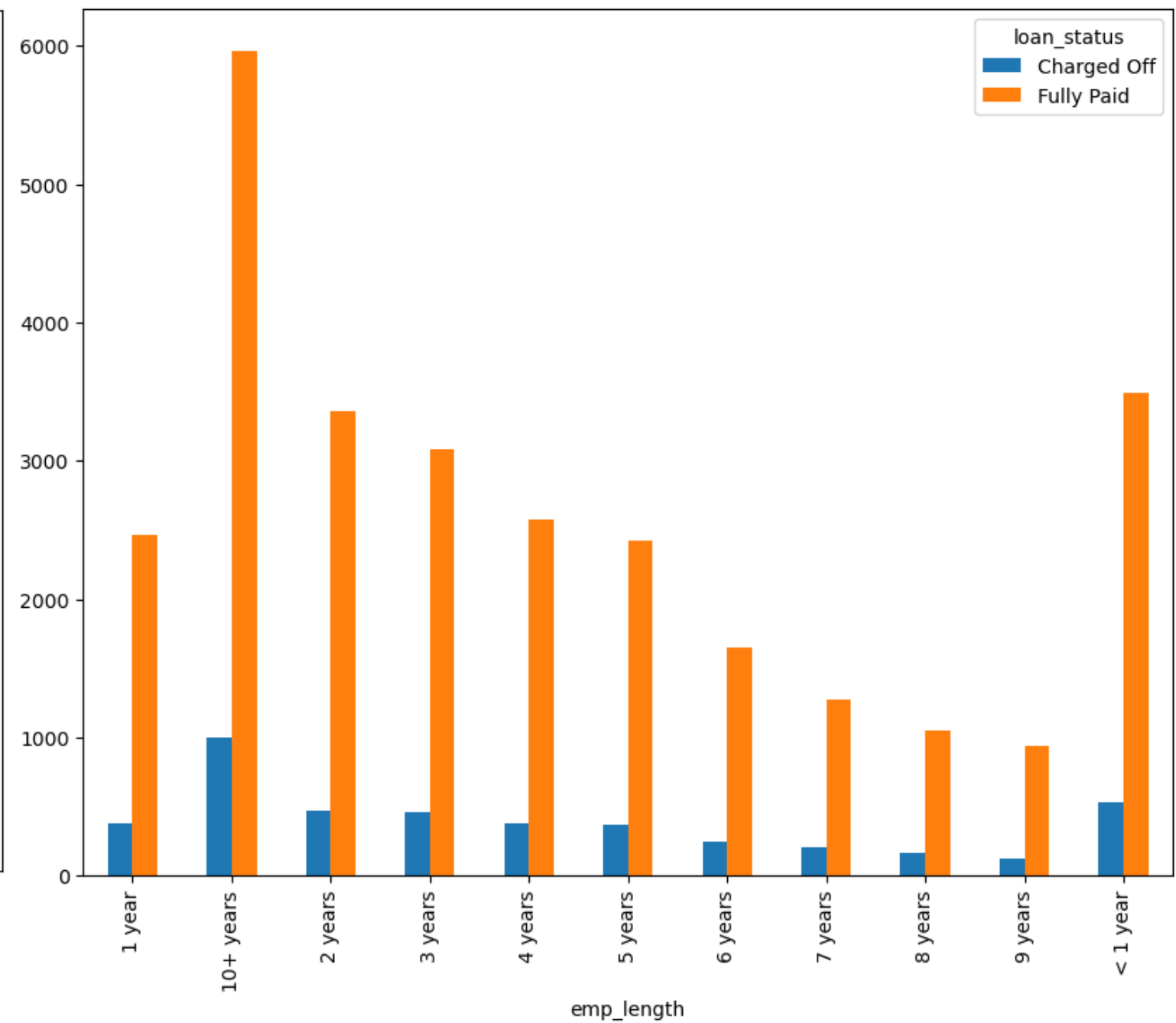
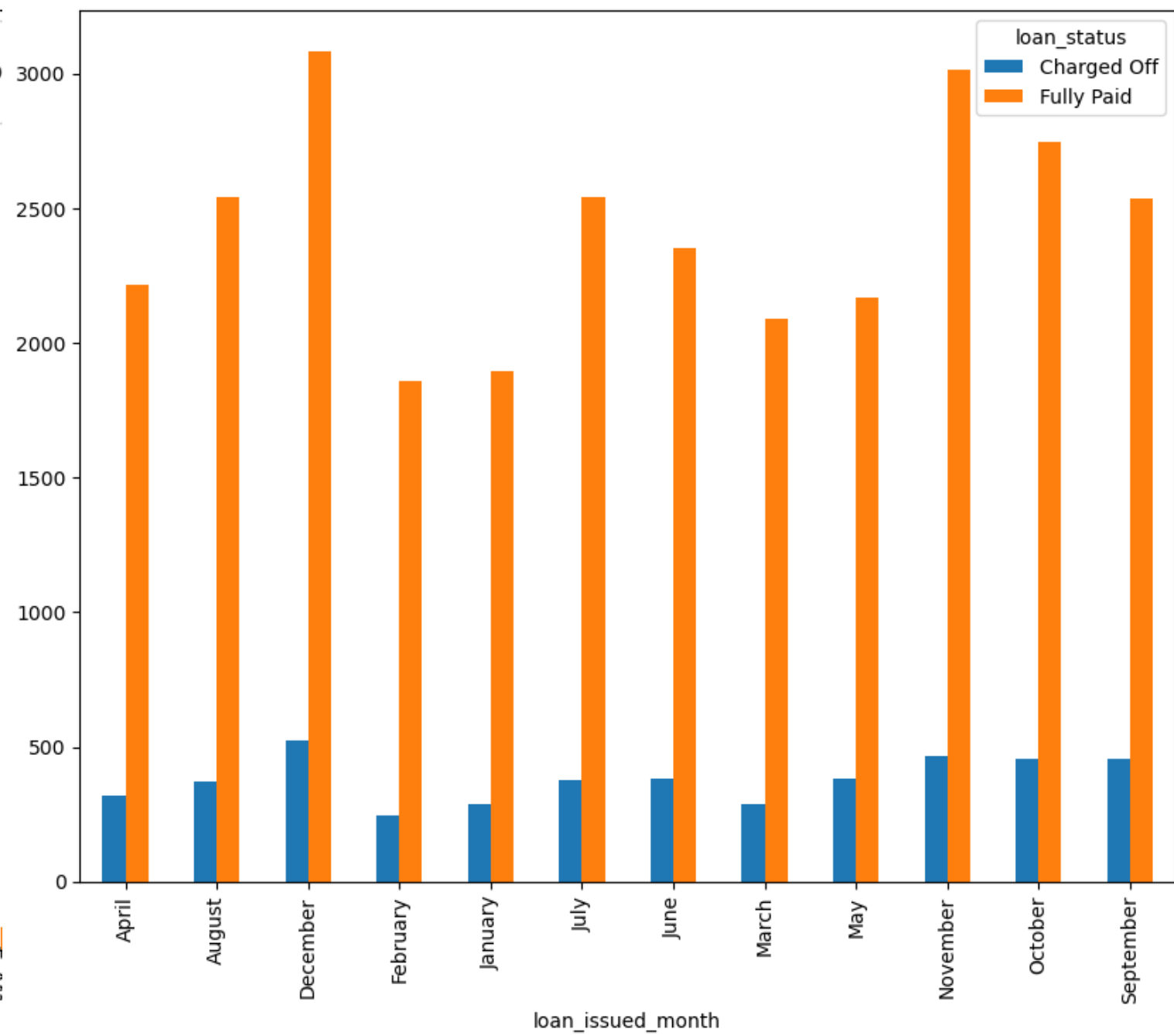
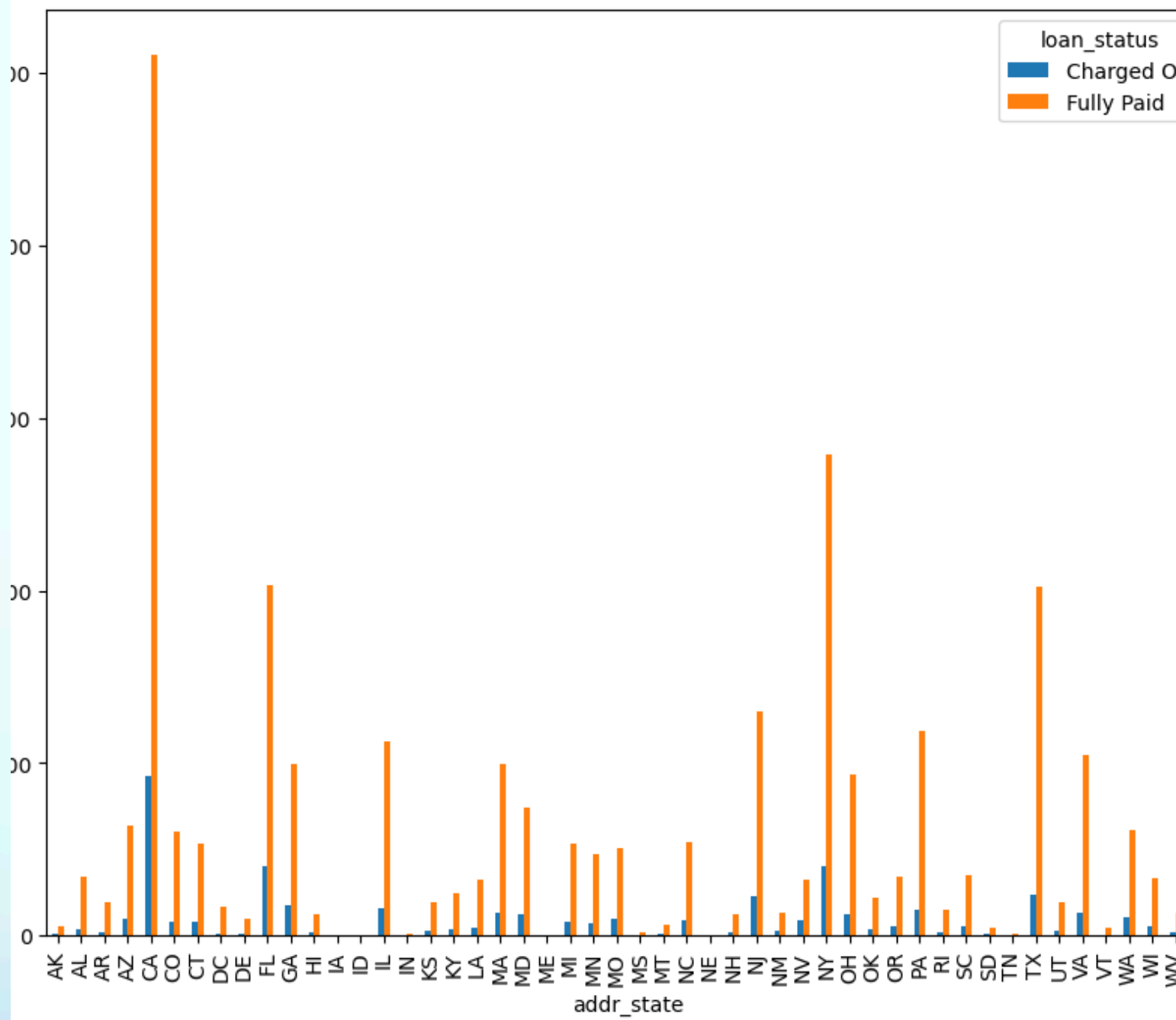




Segmented univariate analysis

Segmenting against the status of the loan

- Here multiple variables are segmented and analysed against the status of the loan.
- The variables analysed are state, loan issued month, sub-grade, term, employment length, home ownership, income verification status and purpose.
- Following slide contains the visual representation of some of the analysis mentioned above.
- More is shown in the Jupyter notebook attached



Bivariate Analysis

- Bivariate analysis is performed on the following and scatter plots are plotted. Interesting insights have emerged.
 - Annual income vs Instalment
 - Annual income vs Interest rate
 - Instalment vs Interest rate
 - Funded amount vs DTI
 - Annual income vs Revolving line utilisation rate
 - Total accounts vs Revolving line utilisation rate
 - Funded amount vs Interest rate
 - Funded amount vs Annual income
- Slides for some of these are included in the following slides

