Lending Club Case Study

Anil Tiwari and Anirudh Naithani

Agenda

- Problem statement
- Analysis approach
- Data Understanding
- Data Cleaning & Pre-processing
- Univariate analysis with visualizations
- Bivariate analysis with visualizations
- Multivariate Analysis
- Suggestions
- References & Useful Links

Problem Statement

- Consumer finance company which specializes in lending various types of loans to urban customers needs to assess risk of:
  - Not approving the loan which are likely to be paid resulting in a loss of business to the company
  - Approving the loan which are not likely to be paid leading to a financial loss for the company
- Based on the data given containing information about past loan applicants and whether they ‘defaulted’ or not. Analyzing data to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Analysis Approach

01

Data quality issues are addressed by removing columns which has not a number, null or higher percentages.

02

Manipulation dates is done to perform year analysis.

03

Univariate analysis is done for Loan amount, Employment length in years, Applicant's grade, Loan Purpose, number of application growth with bar plots.

04

Bivariate analysis is performed on applicant grade and interest rate, employment length and loan amount, and employment length and Debt to Income ration with box plots.

05

Insights are explained comments, summary slide.

Data Understanding

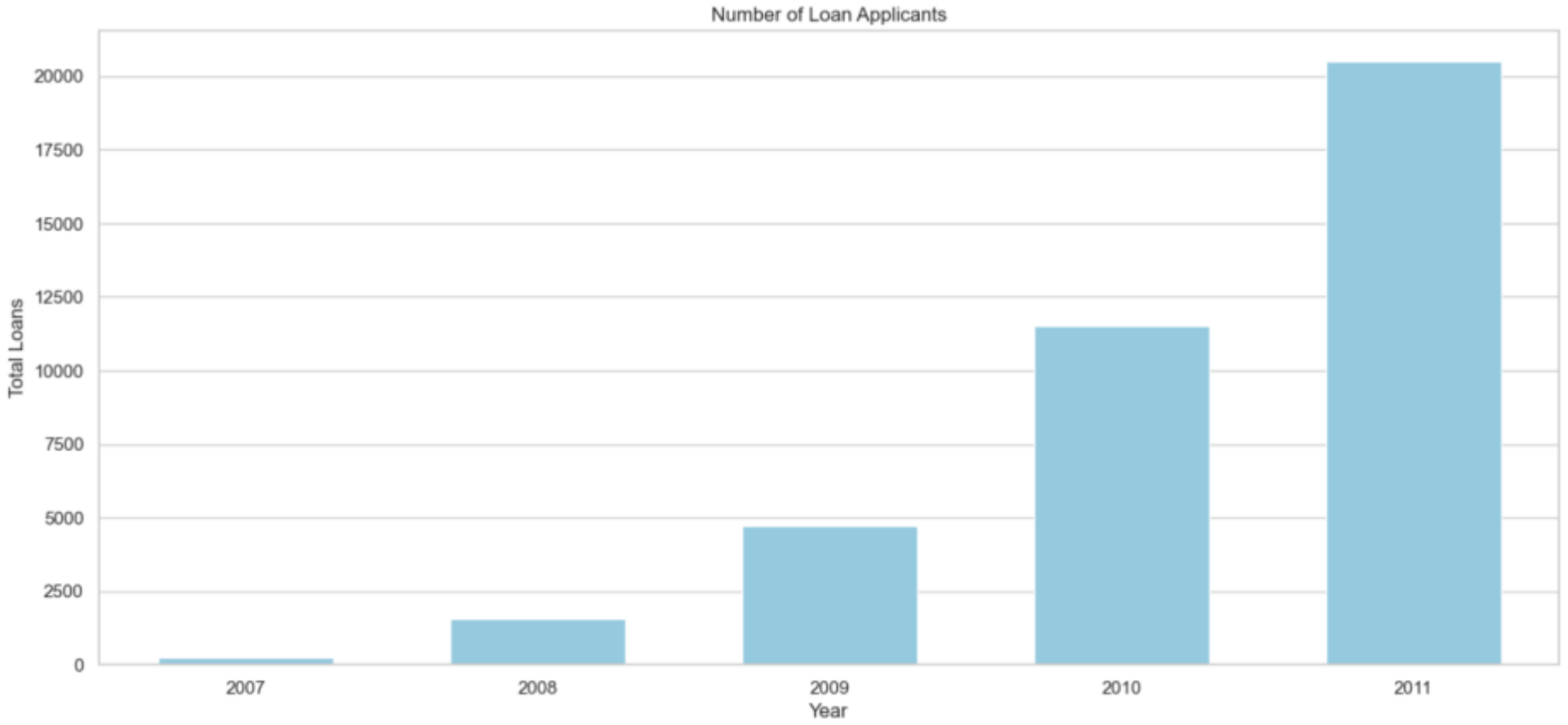
- **Dataset Attributes:**
  - **Primary Attribute**
  - **Loan Status: The Principal Attribute of Interest (loan_status)**
  - **Decision Matrix:**
    - **Loan Acceptance Outcome**- There are three potential scenarios:
    - **Fully Paid**
    - **Current**
    - **Charged-off**
  - **Loan Rejection**
- **Key Columns of Significance:**
  - **Customer Demographics**
  - **Loan Characteristics**
  - **Excluded Columns**

Data Cleaning and Pre-processing

- 1.Loading data from loan CSV
- 2.Checking for null values in the dataset
- 3.Checking for unique values
- 4.Checking for duplicated rows in data
- 5.Dropping Records & Columns
- 6.Common Functions
- 7.Data Conversion
- 8.Outlier Treatment
- 9.Imputing values in Columns

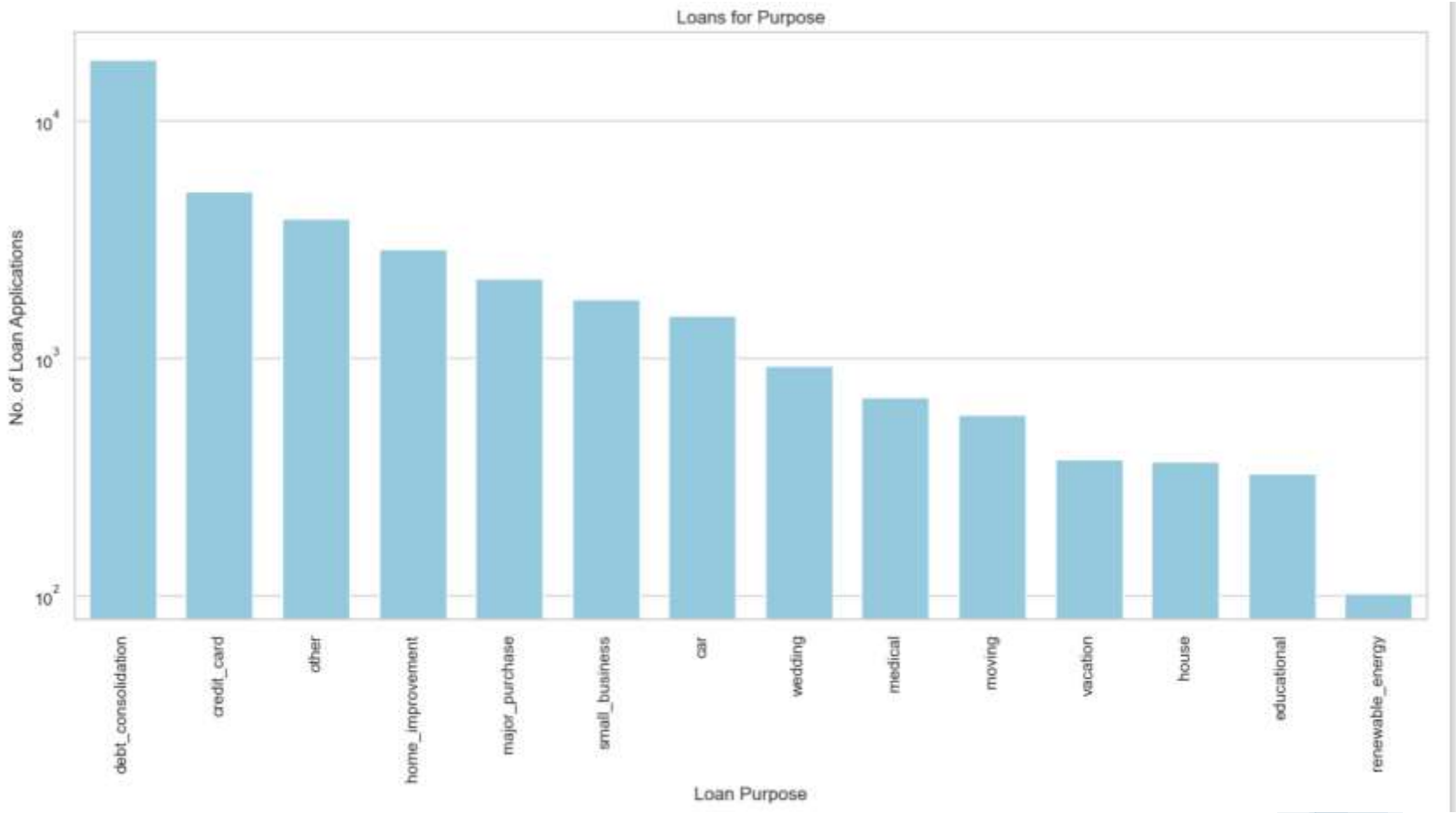
Univariate Analysis 1

- In univariate analysis of number of loan applications vs year, we found sequential growth in number of loan applications from 2007 to 2011



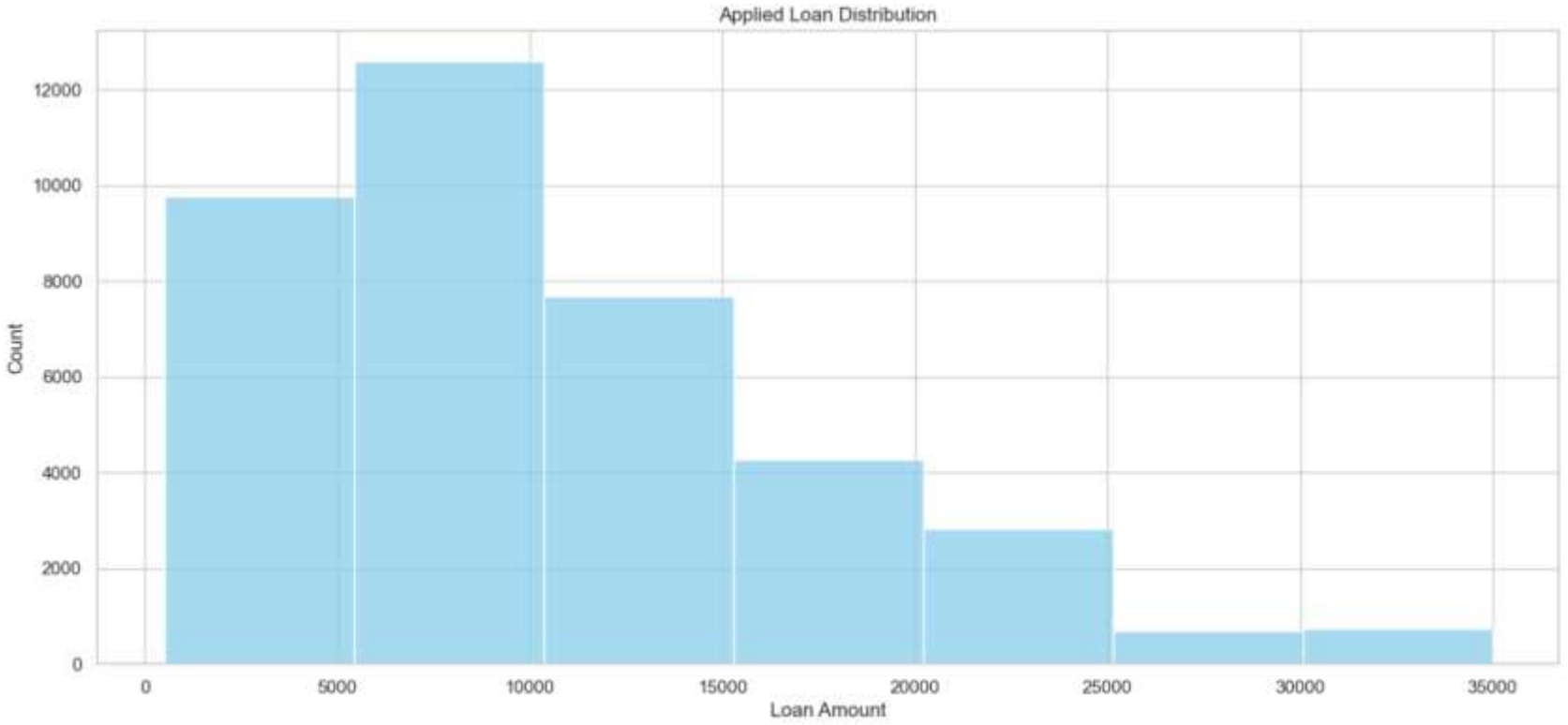
Univariate Analysis 2

- In univariate analysis of purpose for loan, it was observed that highest number of loans are for depreciating assets or consumption (debt consolidation, credit card, major purchase, wedding) compared to appreciating assets (small business, home, education)



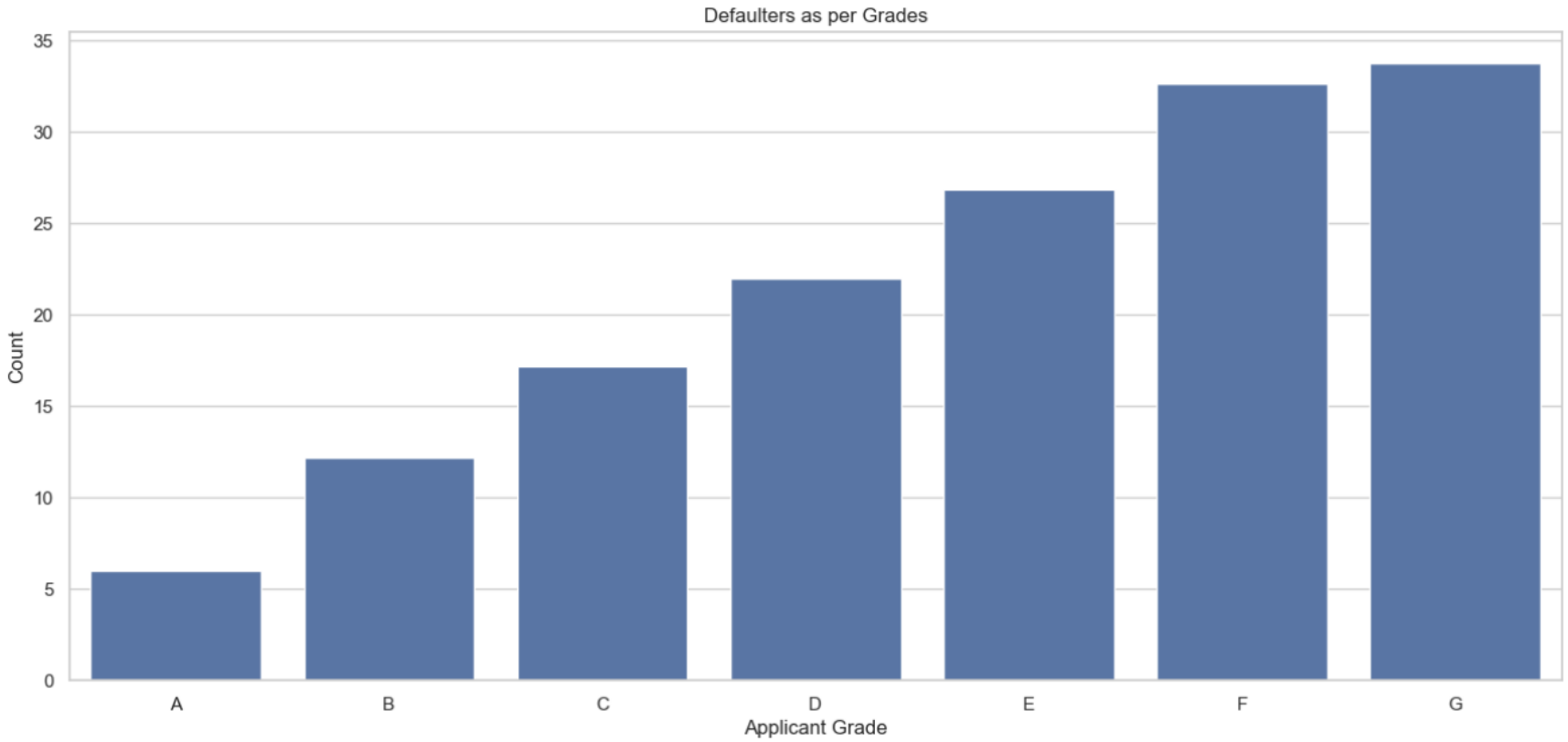
Univariate Analysis 3

- In univariate analysis of loan amount, it was observed that the maximum number of loans applied for the amount 5000 or 10000



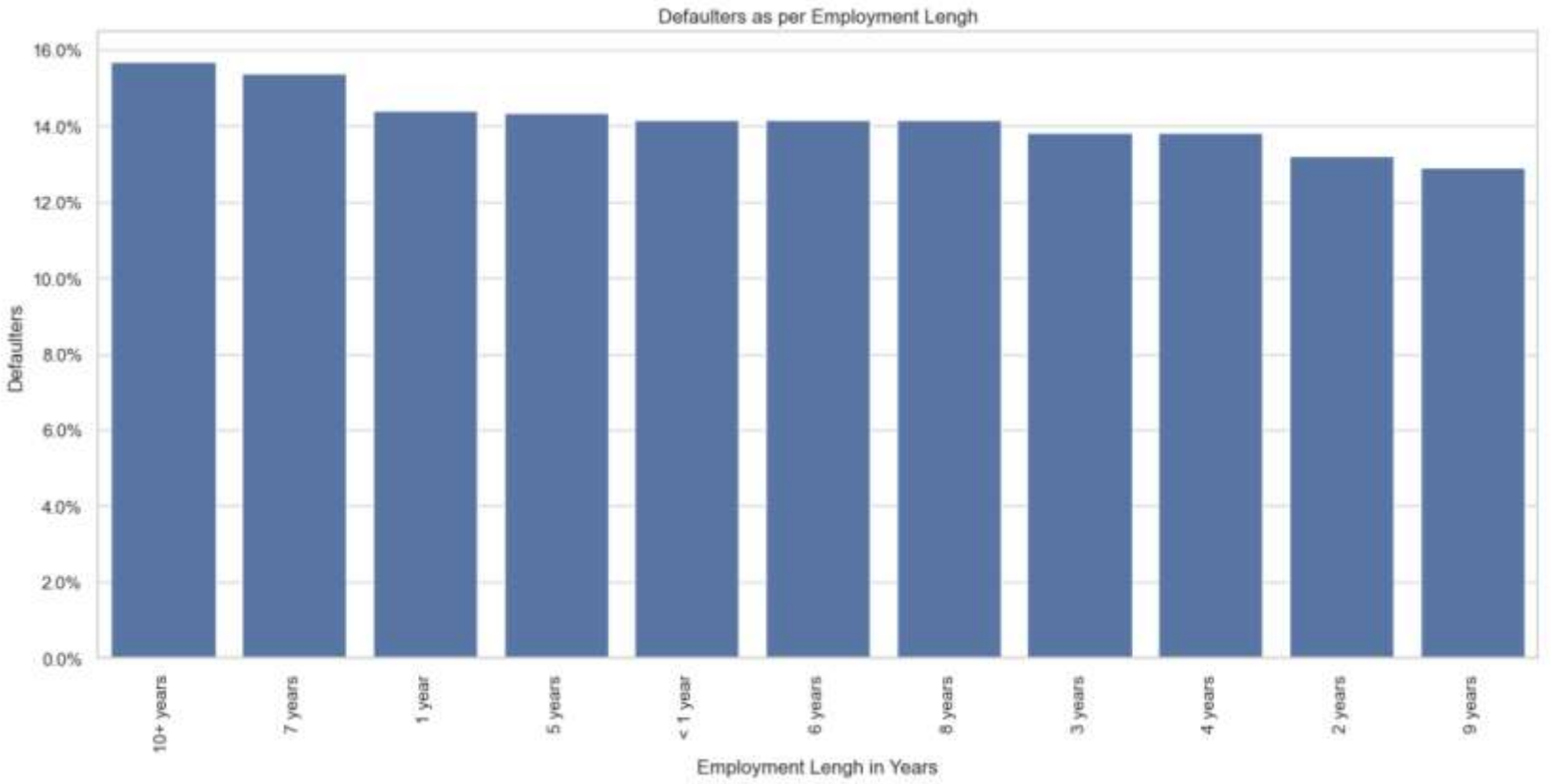
Univariate Analysis 4

- In univariate analysis of applicant grade, it was overserved that as Grade goes down from A to G, defaulters percentage increase



Univariate Analysis 5

- In univariate analysis of applicant employment length in years, it was overserved that as Risk of defaults is maximum for customer with employment length of 10+ years. Otherwise, there is no strong correlation.



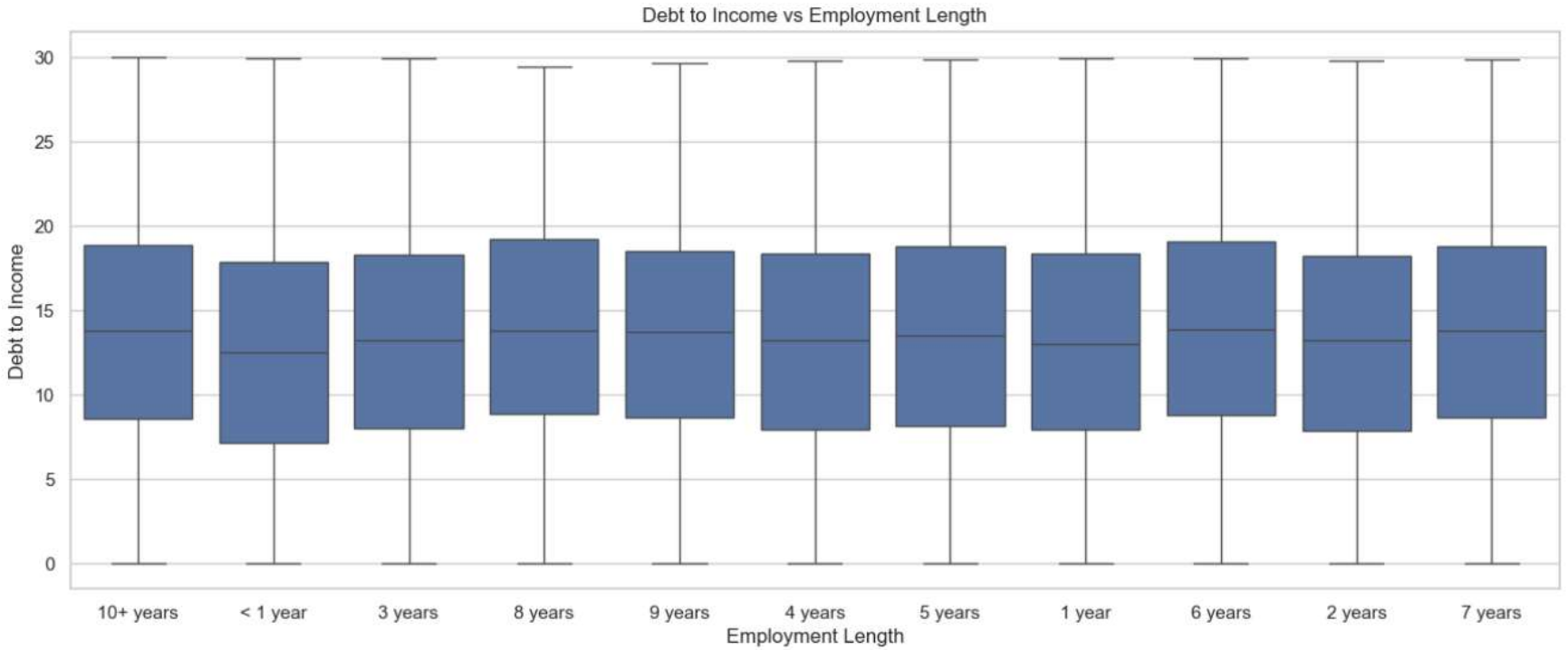
Univariate Analysis 6

- In univariate analysis of loan purpose, it was overserved that Highest number of loans defaults are for productive purposes (business, renewal energy, education, home etc.) while defaults are less for consumption purpose (major purchase, wedding, car etc.)

purpose	count
small_business	27.08
renewable_energy	18.63
educational	17.23
other	16.38
house	16.08
moving	15.97
medical	15.57
debt_consolidation	15.33
vacation	14.13
home_improvement	12.07
credit_card	10.78
car	10.67
wedding	10.37
major_purchase	10.33

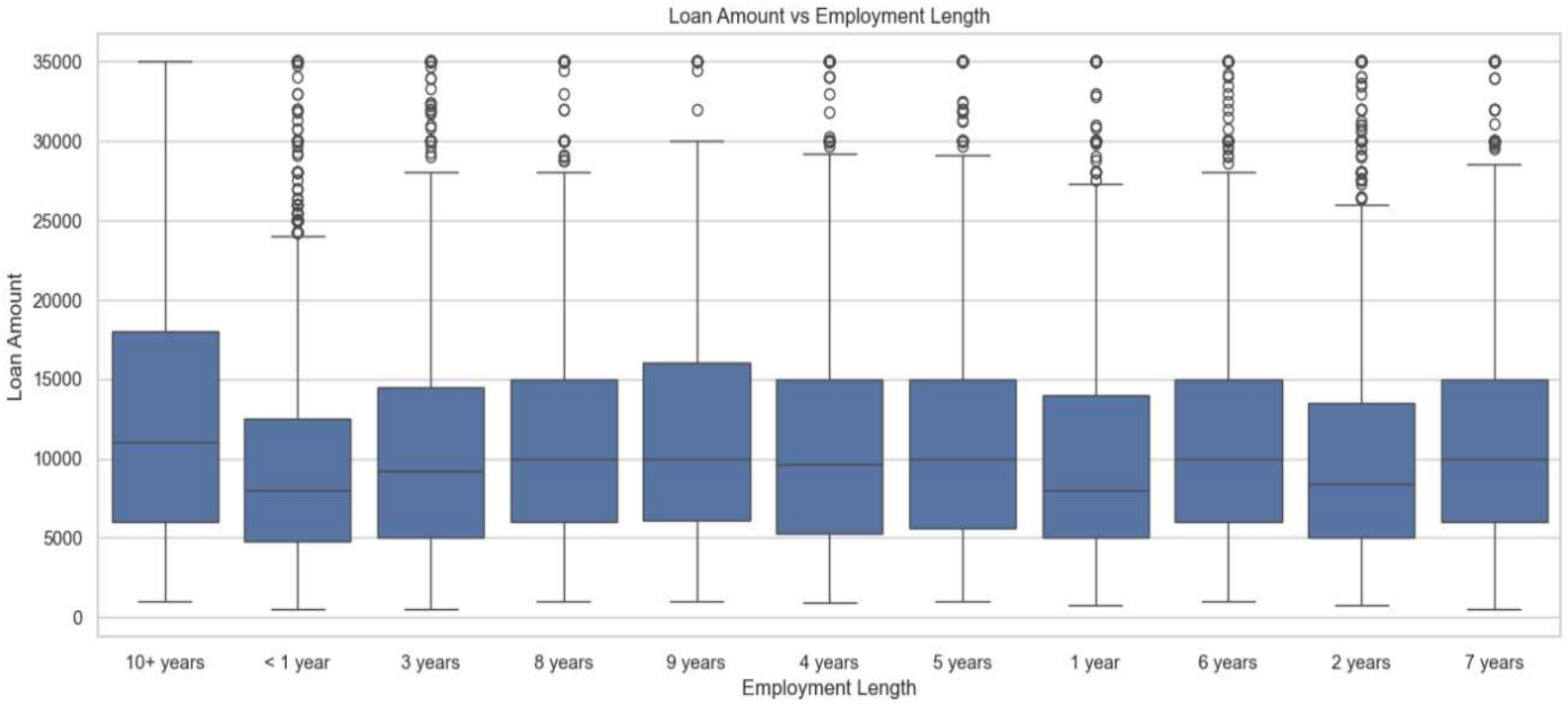
Bivariate Analysis 1

- In Bivariate analysis of employment debt to Income vs Employment length in years, it was observed that applicants with higher employment length (10+, 9, 8, 7 ,6 years) has higher Debt to Income as compared to lower employment length (<1, 1, 2, 3 years).



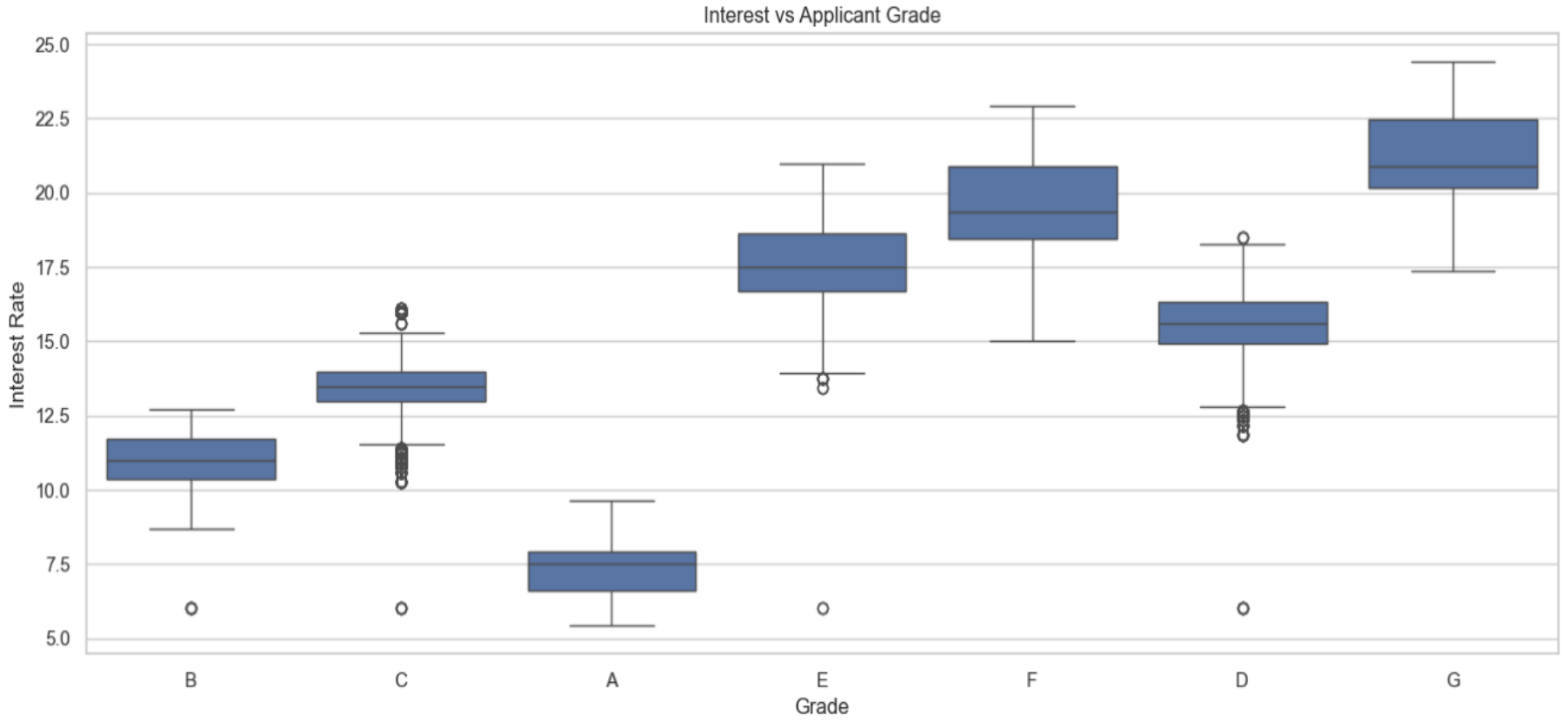
Bivariate Analysis 2

- In bivariate analysis of Employment length and loan amount, it was observed that the applicants with 10+ years was applying for more loan amount compared to 1 year or less.



Bivariate Analysis 3

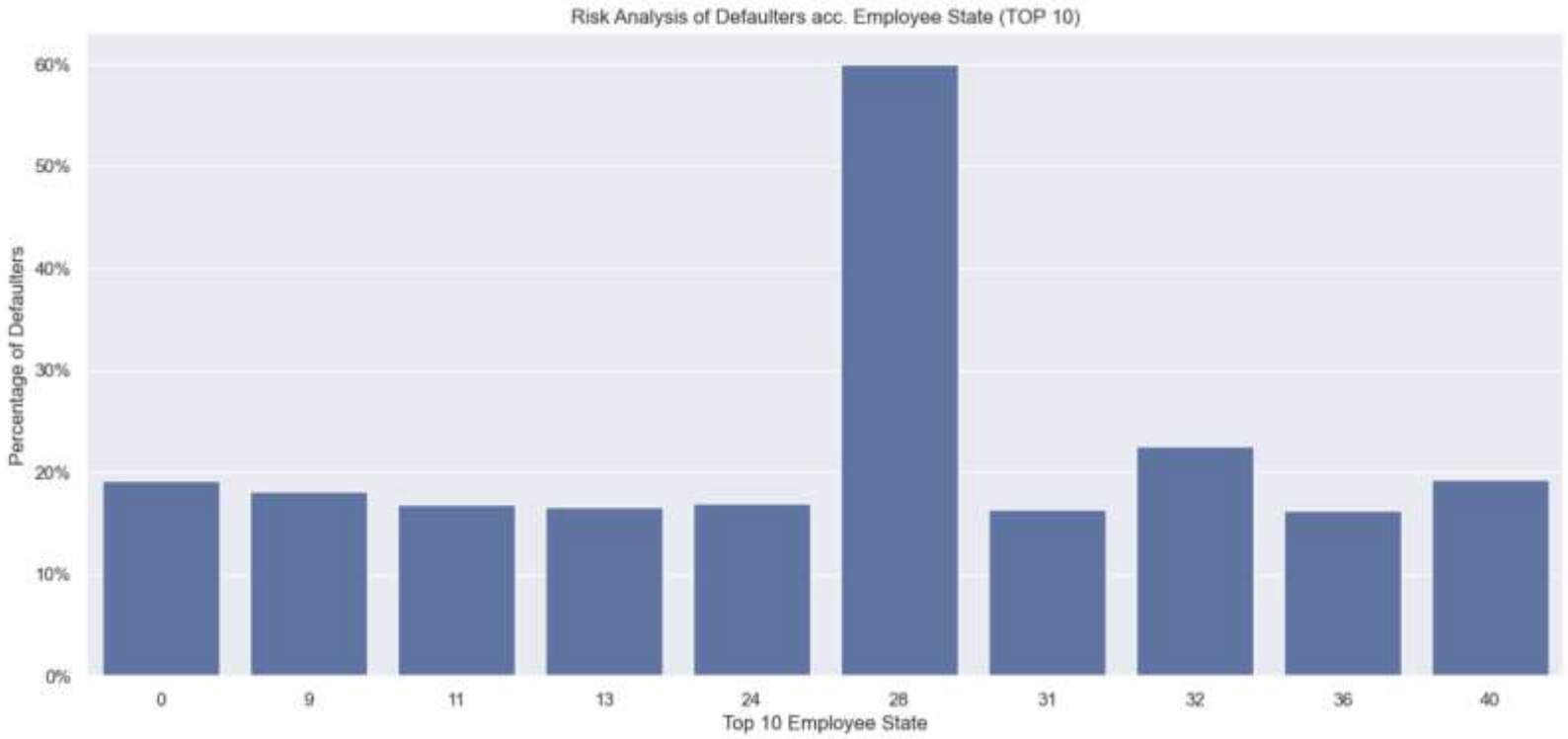
- In bivariate analysis between interest rate and applicant grade showed that higher grade applicants (A, B, C) were offered loans at lower interest rate (within 15%) where as lower grade (D, E, F, and G) applicants are being offered loan at higher rates (above 15%)



Multivariate Analysis

- The percentage of Defaulters is maximum from NE (address) which is 60%

	addr_state	defaulter_perc
28	NE	60.00
32	NV	22.55
40	SD	19.35
0	AK	19.23
9	FL	18.12
24	MO	17.01
11	HI	16.87
13	ID	16.67
31	NM	16.39
36	OR	16.32



Summary

- Lower grade (D, E, F, and G) offered higher loan amount as they are more likely to default (findings from univariate analysis 4, and bivariate analysis 3)
- Employment length of 10+ years applicants shall be approved less loan amount as they as they are more likely to default, and have higher debt to income ratio (findings from univariate analysis 6
- Loans for productive purposes (business, renewal energy, education, home etc.) shall be offered less amount as they are more likely to default (findings from univariate analysis 5, and bivariate analysis 2).

Suggestion

- **Careful Evaluation for Debt Consolidation Loans**: Carefully evaluate applicants seeking debt consolidation loans, considering potential interest rate adjustments or offering financial counselling services to manage the associated risks.
- **Consider Housing Stability:** Take housing status into account during the underwriting process to assess housing stability and its impact on the applicant's ability to repay the loan.
- **Review Verification Process:** Review the verification process to ensure effective assessment of applicant creditworthiness. Consider improvements or adjustments based on the review findings.
- **Monitor & Adjust for Regional Risk Trends:** Monitor regional risk trends, especially in states like **California, Florida, and New York.** Adjust lending strategies or rates accordingly in high-risk regions.
- **Thorough Assessment for High Loan Amounts:** Conduct more thorough assessments for loan amounts of $15,000 or higher. Consider capping loan amounts for higher-risk applicants to mitigate potential defaults.
- **Adjust Interest Rates Based on DTI Ratios:** Review the interest rate determination process and consider adjusting rates based on Debt-to-Income (DTI) ratios to align with the borrower's ability to repay.
- **Consider Income Levels for Affordability:** Consider offering financial education resources and set maximum loan amounts based on annual incomes below $40,000 to ensure loan affordability for borrowers.

Reference & URLs

- **Technologies: Package Version Documentation**
  - Python https://www.python.org/
  - Matplotlib https://matplotlib.org/
  - Numpy https://numpy.org/
  - Pandas https://pandas.pydata.org/
  - Seaborn https://seaborn.pydata.org/

- **GitHub Repository Link:**
  https://github.com/aniltiwari-tech/lending-club-case-study