

Lending Club Case Study

Exploratory Data Analysis On Loan

Author – Ayan Mitra

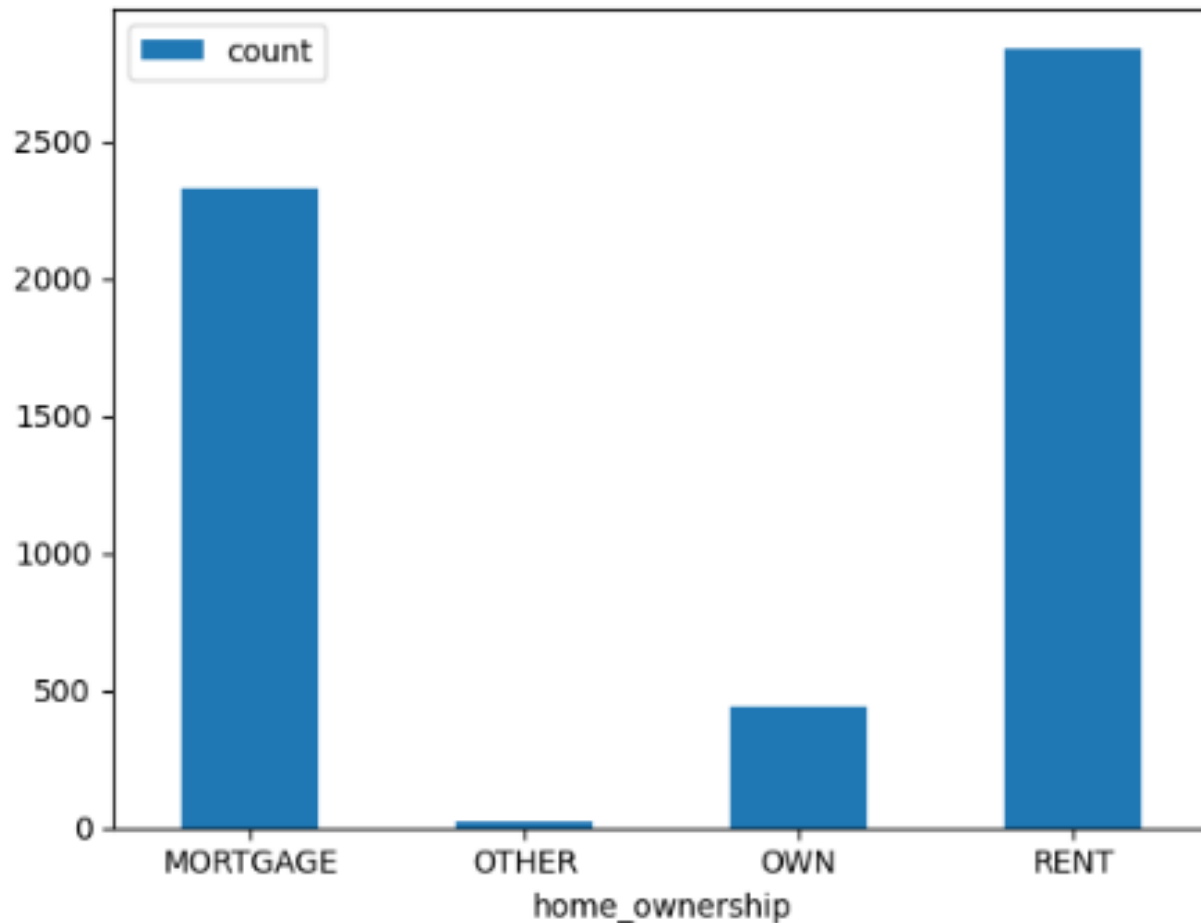
Date – 24th Feb, 2024

Genesis of the Analysis

- We have received ~40k loan records for analysis
- The data is cleaned by taking out the unnecessary columns, and by formatting important numeric and date columns
- Derived values are created before analysis
- A subset of data is taken by income groups for the defaulters to perform various analysis
- Various charts and graphs are produced to describe the data
- A conclusion is given at the end of each analysis

Analysis of Defaulters by Home Ownership

Objective – Find if people are failing to payoff their loan because of their home ownership status



Strategy:

- Grouped the defaulters by their home ownership and stored them into a different dataset
- Used the dataset to generate a bar chart

Findings:

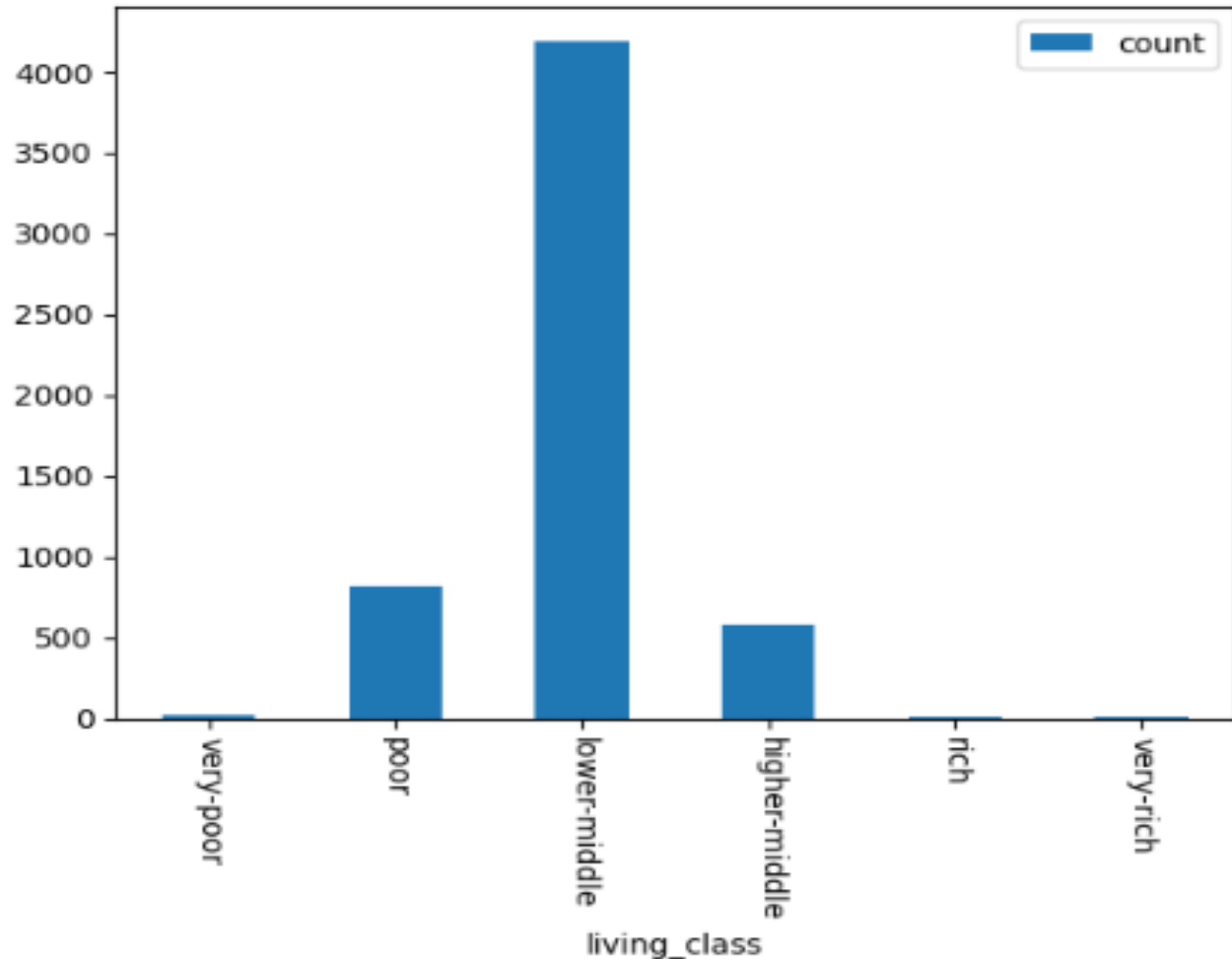
- People with rented and mortgaged house are defaulting their loan more than people with own house

Conclusion:

- The lending club should prefer giving loans to those who have their own house than people renting houses or having a mortgage on their home.

Analysis of Defaulters by Income Class

Objective – Find if people are failing to payoff their because of their income class



Strategy:

- Grouped the defaulters by their annual income category in 6 classes
 - Very poor (up to \$10,000)
 - Poor (\$10,001 to \$30,000)
 - Lower-Middle (\$30,001 to \$100,000)
 - Higher-Middle (\$100,001 to \$500,000)
 - Rich (\$500,001 to \$1,000,000)
 - Very-Rich (Greater than \$1,000,000)

Findings:

- People in the Lower-Middle Income class tends to default on their loan more than other income classes

Conclusion:

- Although, the number of loan-takers are more in this category, the lending club should pay caution while disbursing loan to people in this income range

Analysis of Defaulter in Lower-Middle Income Class w.r.t. Interest Rates

Objective – Since we know the largest population of loan defaulters are in lower-middle income category, does the interest rates play any role for defaulting the loan

int_range

11-15 2023

16-20 1355

6-10 598

21-25 220

0-5 0



Strategy:

- Grouped the interest rates under which the loan is taken for the people in the lower-middle class income category in the following buckets
 - Upto 5%
 - 6% to 10%
 - 11% to 15%
 - 16% to 20%
 - 21% to 25%

Findings:

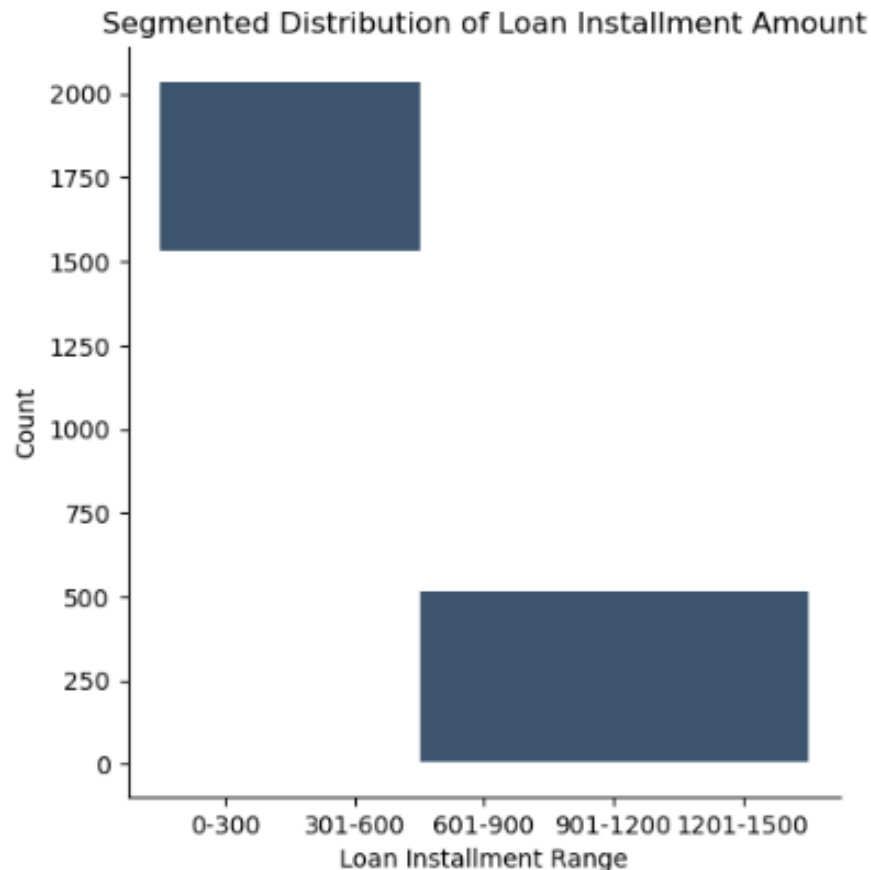
- People who have taken loan with a rate of interest between 11% to 15% are defaulting on their loan more

Conclusion:

- Although, the number of loan-takers are more with the rate of interest between 11% to 15%, but we can see the count from other categories is still higher.
- The lending club can be cautious deciding the rate of interest for loan-takers in the income category

Analysis of Defaulter in Lower-Middle Income Class w.r.t. Installment Amount

Objective – Since we know the largest population of loan defaulter are in lower-middle income category, does the installment amount play any role for defaulting the loan



Strategy:

- Grouped the installment payment for the lower-middle class income category in the following buckets
 - Upto \$300
 - \$301 to \$600
 - \$601 to \$900
 - \$901 to \$1200
 - > \$1200

Findings:

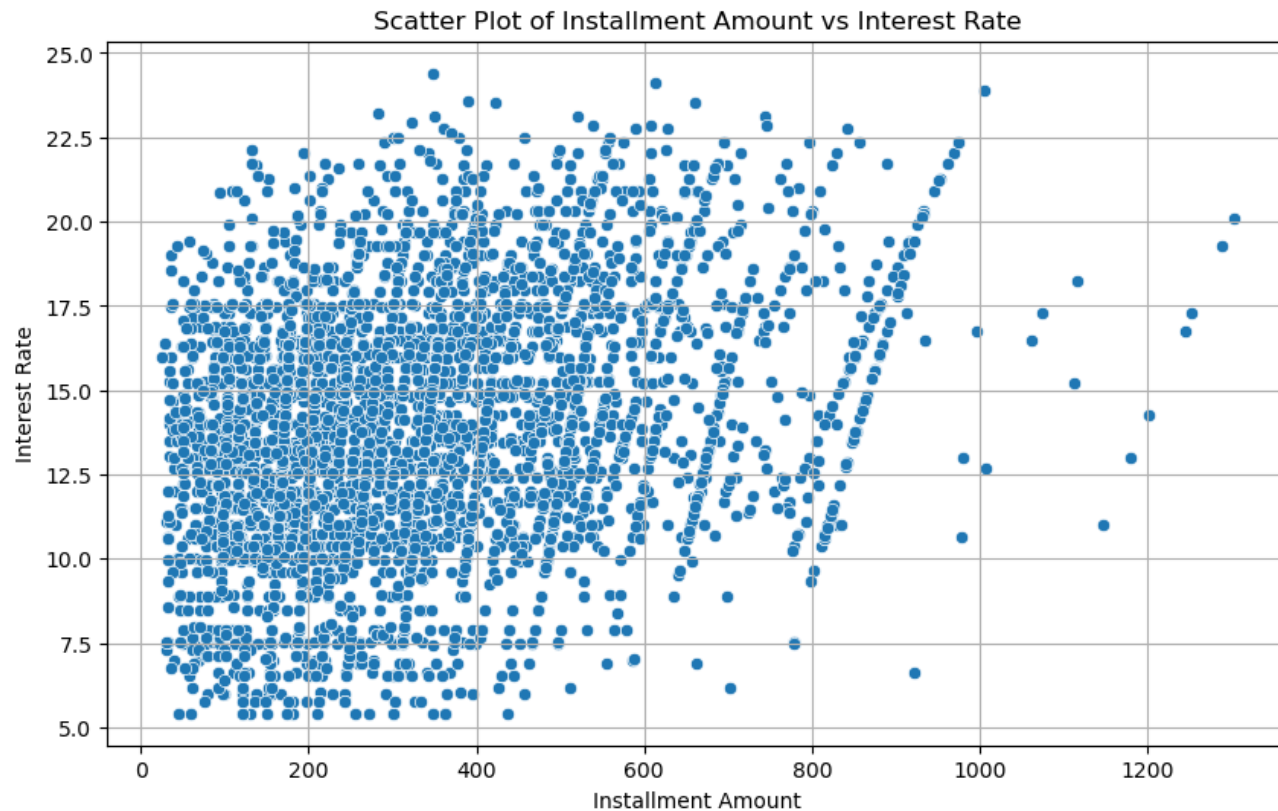
- Surprisingly, people with a smaller loan installment are defaulting the loan more

Conclusion:

- The results are a bit contradictory. It seems people with a smaller installment to pay are tending to default on their loan more than people with bigger installment to pay
- The lending club should recalibrate their installment payment formulas so that it stands >\$600 per month

Analysis of Defaulter in Lower-Middle Income Class for a correlation between monthly installment and interest rate

Objective – Since we know the largest population of loan defaulter are in lower-middle income category, is there a correlation between the monthly installment amount and interest rate.



Strategy:

- Taken the installment amount and interest rates to plot a scatterplot to find the correlation

Findings:

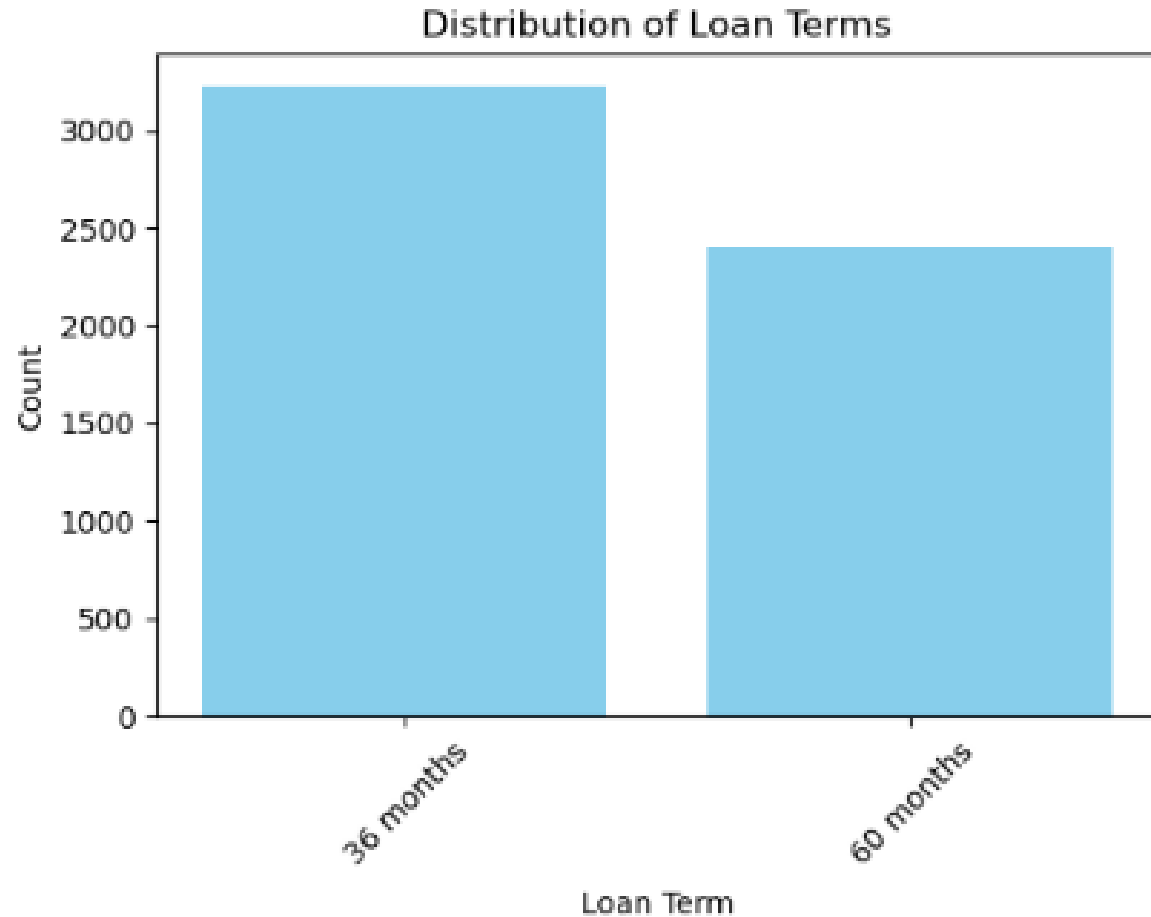
- The highest density was found \$100 to \$400 range with interests between 7.5% to 17.5%

Conclusion:

- Loan takers who typically pay a lower monthly installment with an interest rate between 10% to 17% are defaulting on their loan more than the rest of the group

Analysis of Defaulter on Loan-Terms

Objective – Find if the loan term makes any difference for defaulting the loan



Strategy:

- Take the loan-term field and plot a bar chart to see the distribution

Findings:

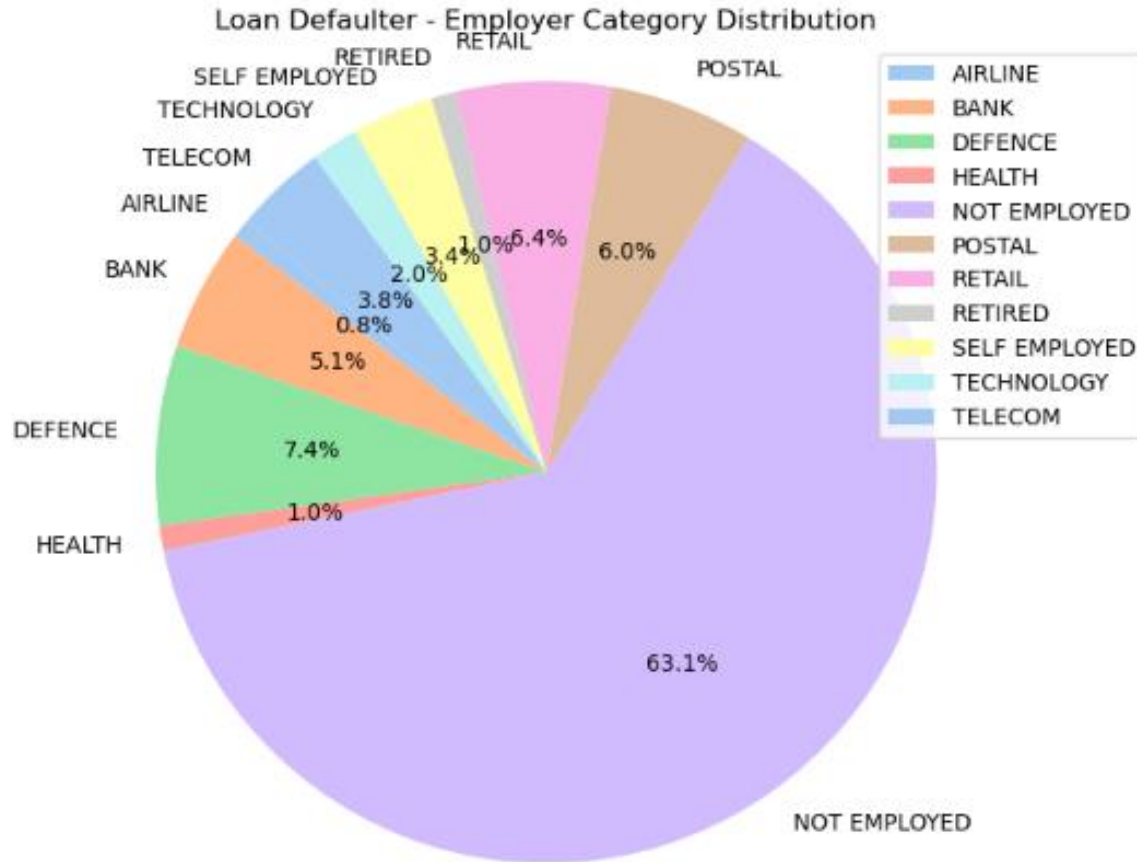
- The number of defaulters appeared to be slightly high in the 36 months loan term, but it is not very different from 60 months

Conclusion:

- The loan term doesn't seem to have an impact on defaulting a loan

Analysis on Employer Category

Objective – Find if the employer's category has any role on defaulting on loan



Strategy:

- For this analysis only those employers are taken where there are more than 5 loan defaulters
- Cleansed the employer title data for better analysis (e.g. employers like Walmart, US Postal Services etc. are named differently in various records)
- Next grouped the data by the employers and categorized them in 22 different buckets
- Finally plotted a pie chart to see the distribution

Findings:

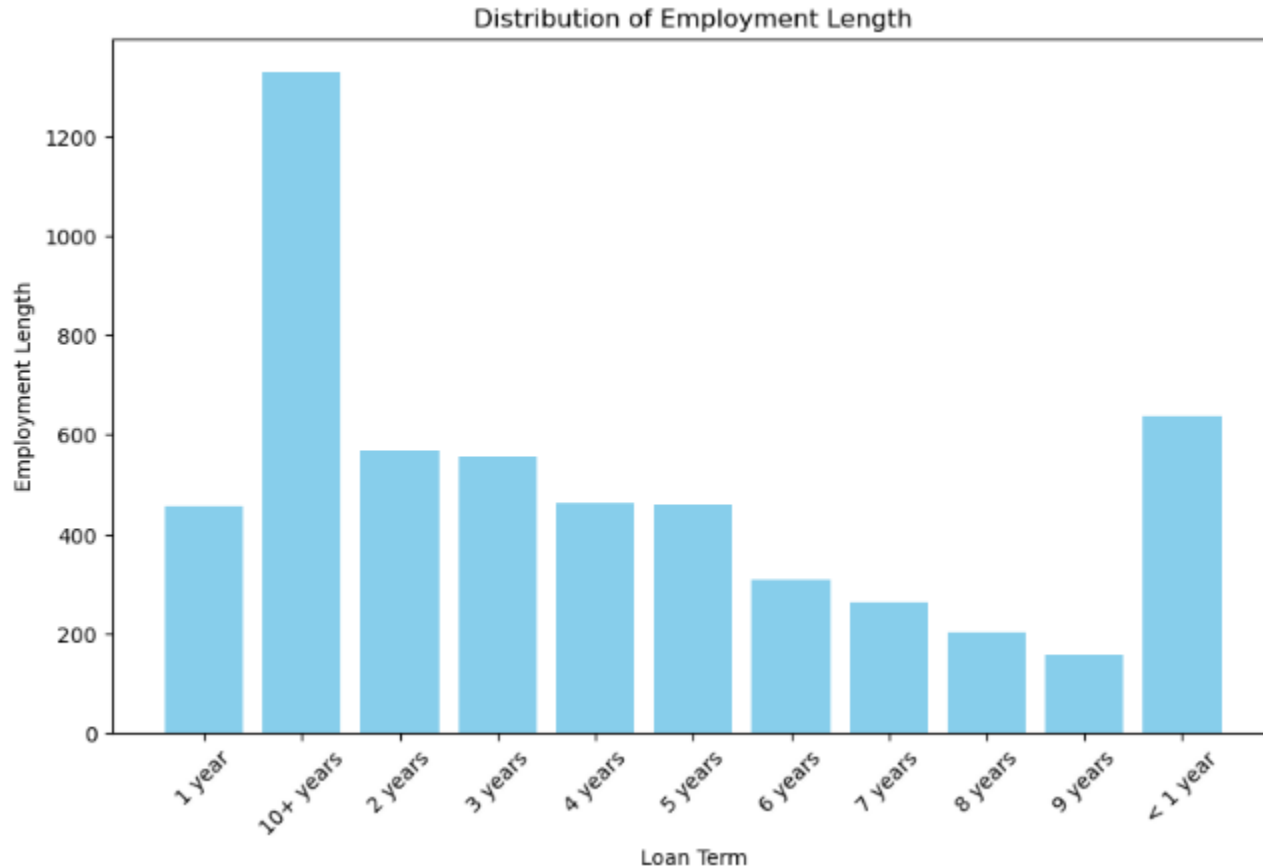
- People who are unemployed tends to be defaulting on their loans the most
- However, in the employed category, people whose employers are in Defense, Retail, and Postal service categories are defaulting more than the rest of the population

Conclusion:

- The lending club must be very careful granting loans to unemployed
- The lending club must pay caution while lending loans to those people who employers are in Defense, Retail, or Postal service

Analysis of Defaulters against their employment length

Objective – Find how the employment length impacts on loan defaulters



Strategy:

- Taken the employment length and plotted a bar chart to find the distribution

Findings:

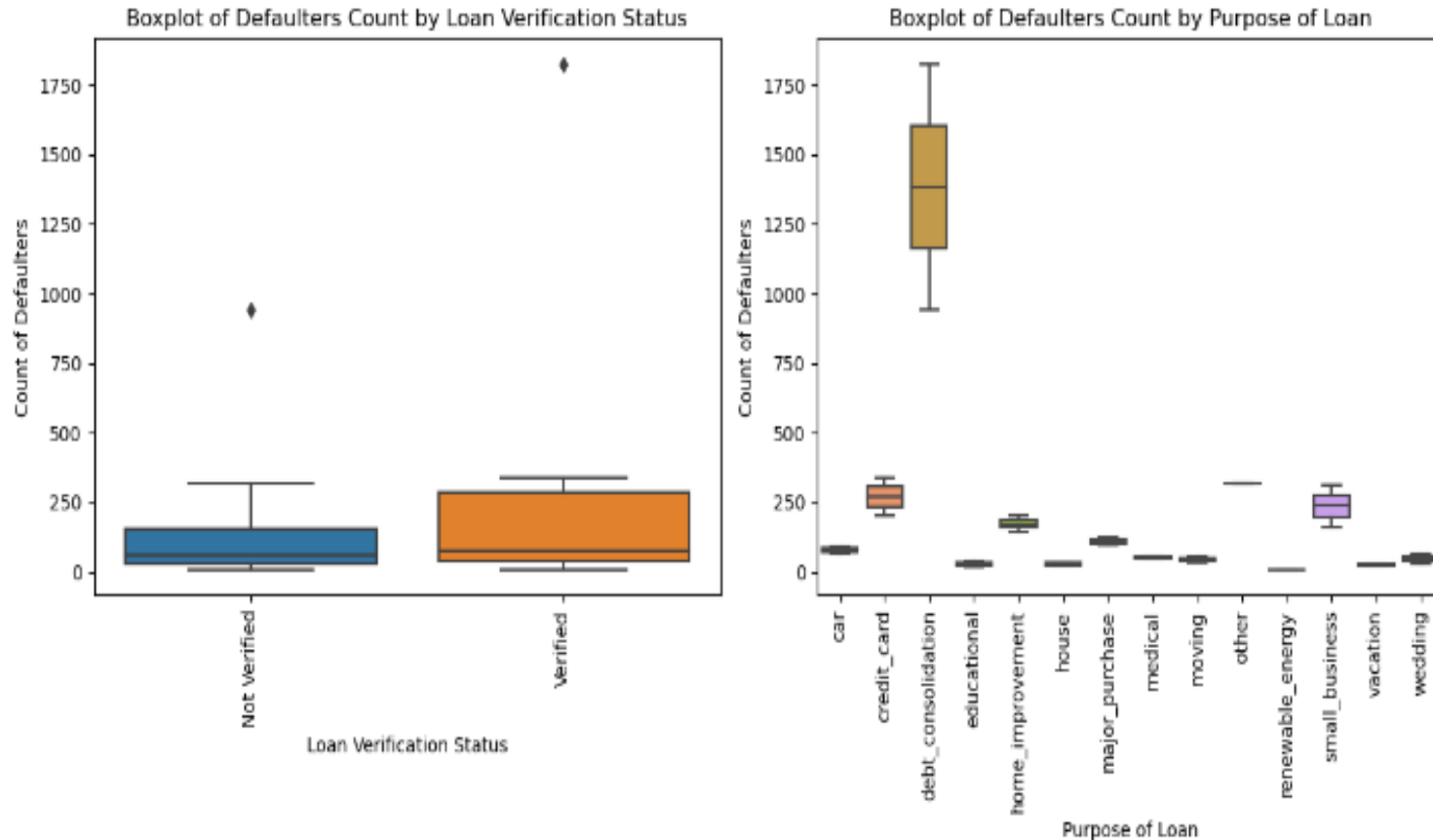
- People with 10+ years employment length are the highest loan defaulters
- In the group of people with less than 10 years of service, the highest loan defaulters falls into the <1 year of employment length category

Conclusion:

- While the population of defaulters is largest in 10+ years employment length category, it is obvious that people are taking and defaulting on loan if they have <1 year of service.
- The lending club need to pay caution while lending loans to people with <1 year of employment
- People with 6 to 9 years of employment length tend to be defaulting less on their loan

Analysis of Loan Verification Status and Purpose

Objective – Find out the defaulters are distributed among the income verification status and the purpose for which the loan is being taken.



Strategy:

- Categorized the data by loan verification status and purpose of the loan
- Plotted a boxplot graph to see the distribution of the data

Findings:

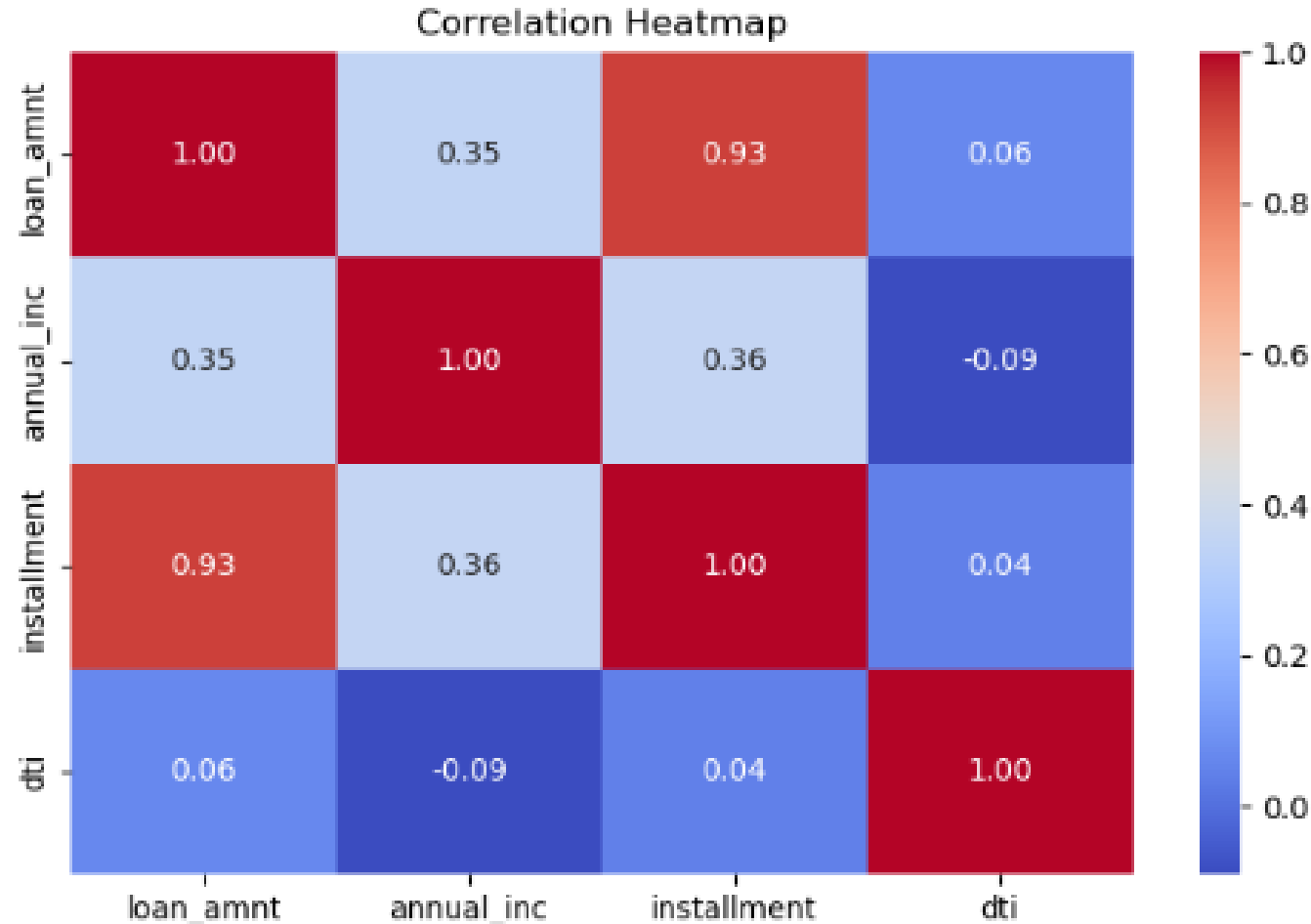
- Income verification seems to be not playing any major role on the loan being defaulted
- People taken loan for debt consolidation are the highest among defaulters
- The 25th and 75th quartiles of those defaulters who have taken loan for credit cards and small business are about the same

Conclusion:

- The lending club need to very careful while lending loan to those who are taking it for debt consolidation

Analysis between Loan Amount, Annual Income, Installment & DTI

Objective – Is there a correlation between Loan Amount, Annual Income, Installment & DTI among the defaulters group



Strategy:

- Taken the Loan Amount, Annual Income, Installment & DTI fields
- Derived the correlation between these fields
- Plotted a heatmap to find the correlation

Findings:

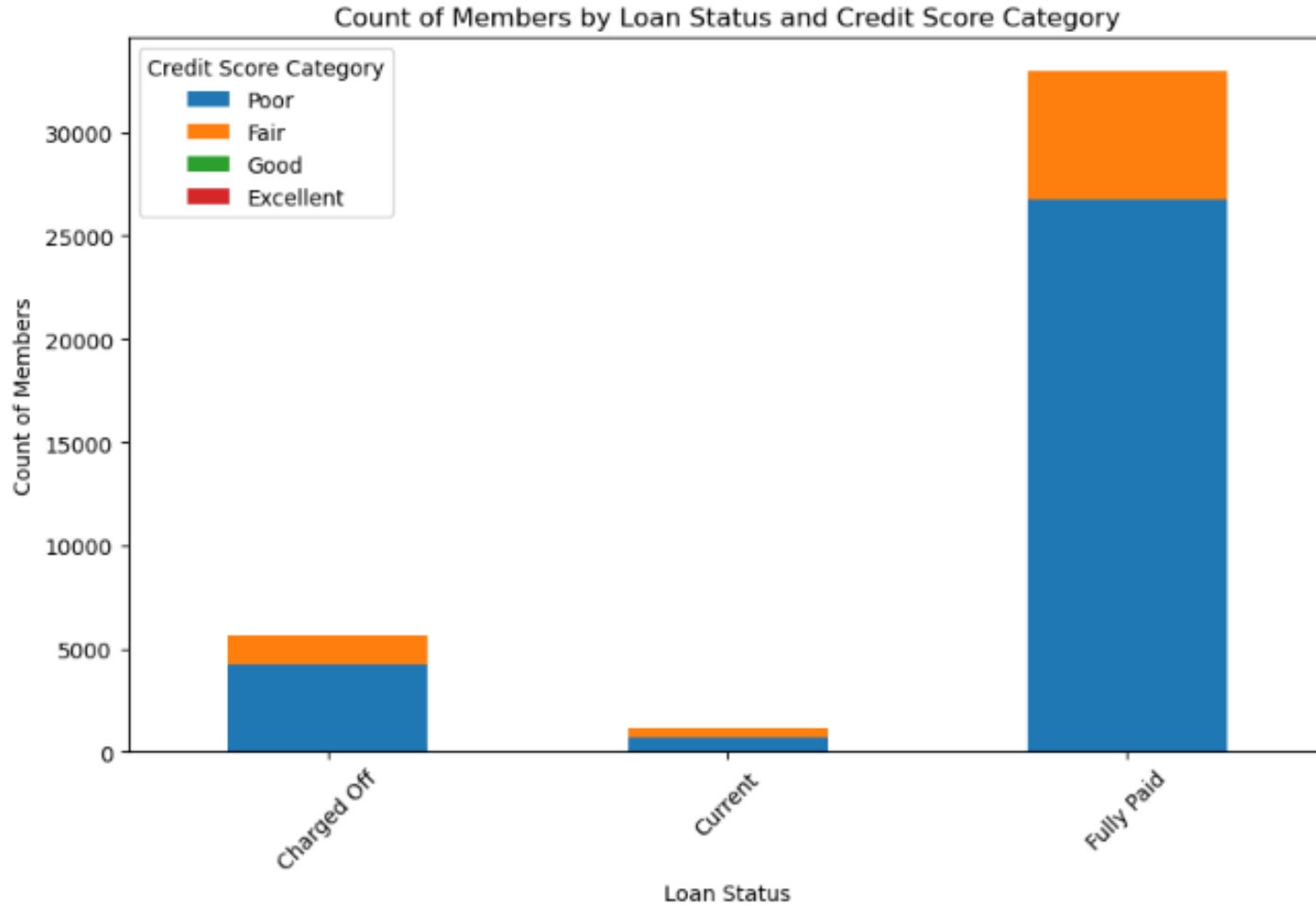
- Installment amount and loan amount are highly correlated
- Annual income and DTI are not correlated

Conclusion:

- The lending club need to very careful while determining the loan amount and the installment amount
- The installment amount is also correlated with annual income of the person

Analysis the data by determining a credit score

Objective – Determine a credit score and see the distribution among the complete dataset



Strategy:

- Taken the Loan Amount, Annual Income, Installment & DTI as individual dimensions
- Computed a credit score by normalizing the data in the above dimensions
- Determined a credit score category between Poor, Fair, Good and Excellent
- Plotted a stacked chart to find the distribution of credit rating

Findings:

- There is a large number of members who are poorly rated on their credit but still granted a loan

Conclusion:

- The lending club must use a scoring mechanism (like credit rating) before granting a loan to its members

Thank you