

# Lending Club Case Study

Exploratory Data Analysis On Loan

**Author** – Ayan Mitra

**Date** – 24<sup>th</sup> Feb, 2024

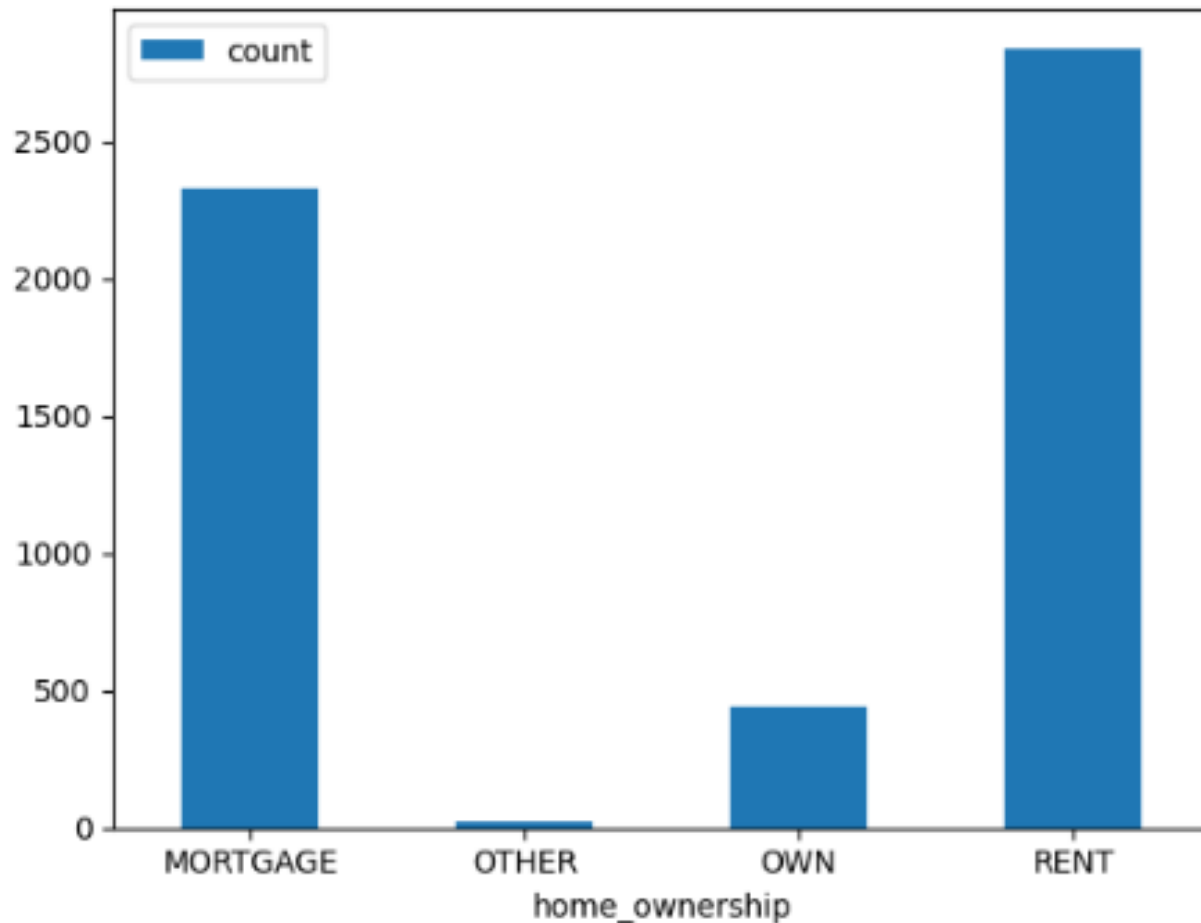
# Genesis of the Analysis

- 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

# Analyzing Data for Defaulters

# 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:

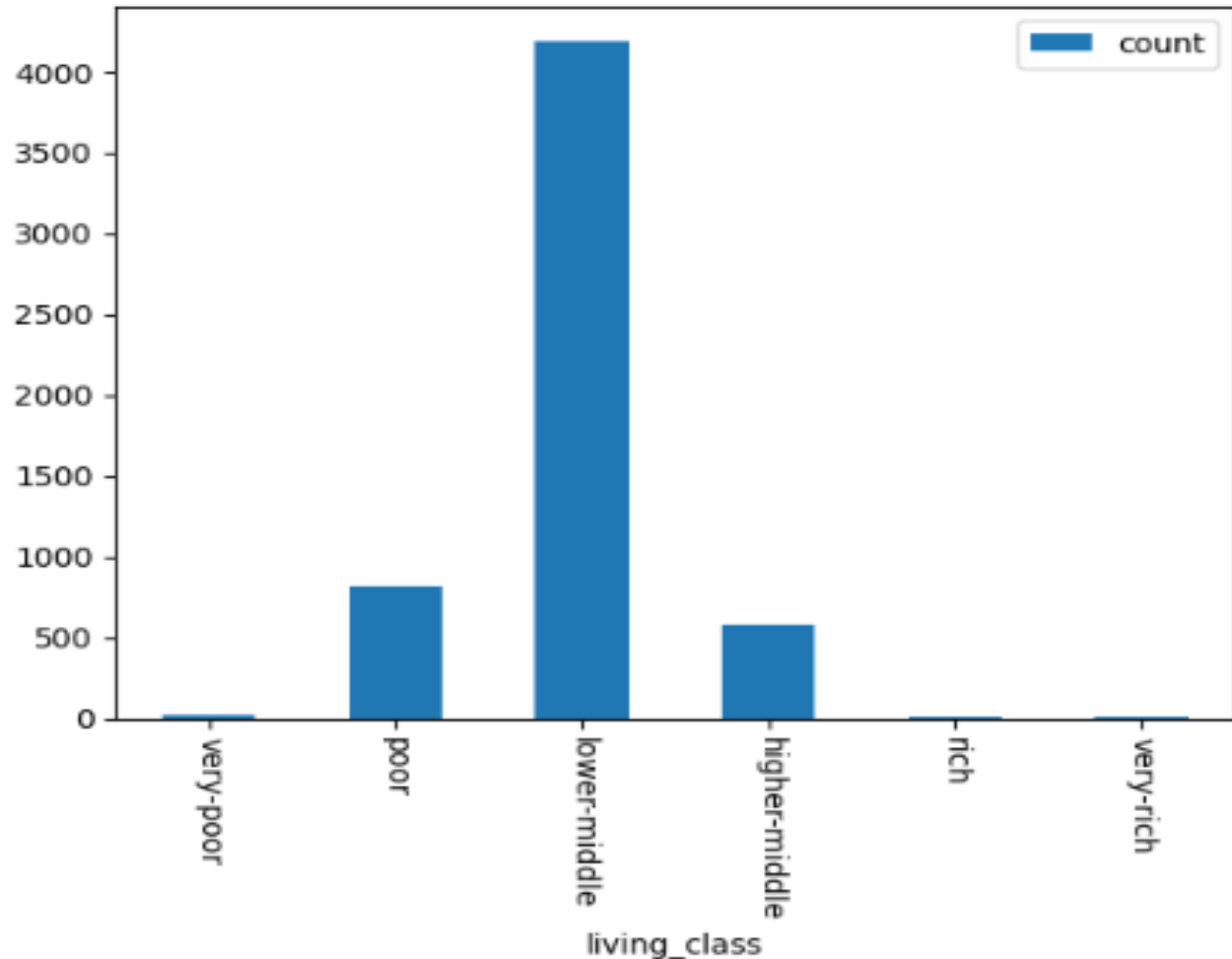
- People with rented and mortgaged house are defaulting their loan more than people with own house

## 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:

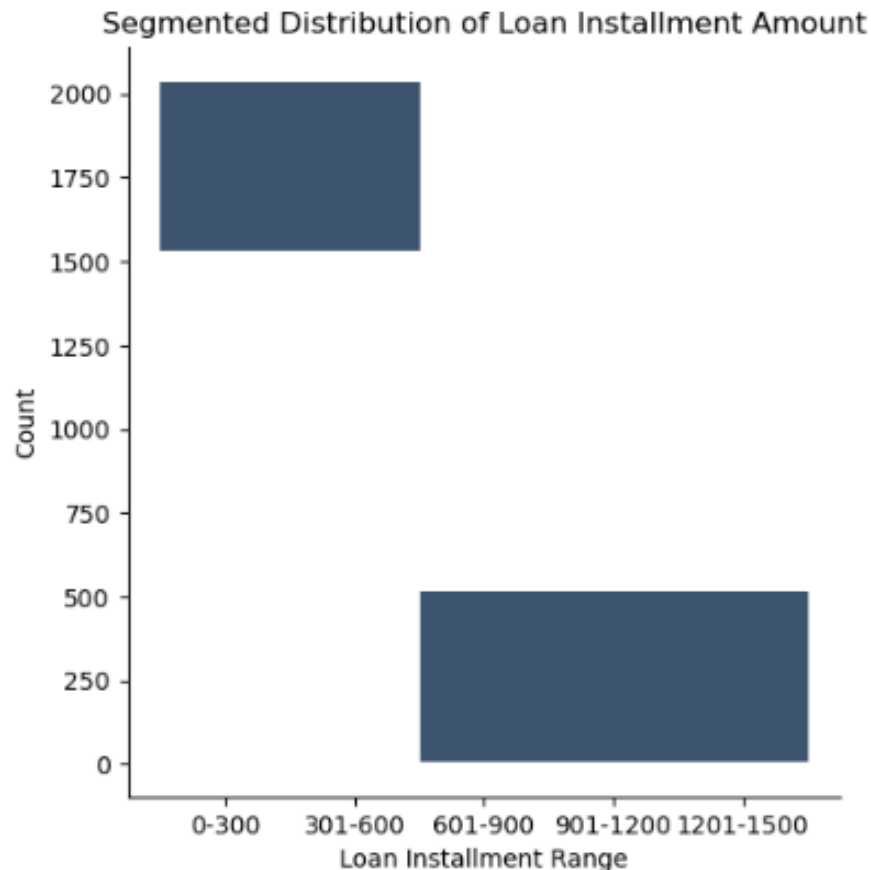
- People who have taken loan with a rate of interest between 11% to 15% are defaulting on their loan more

## 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:

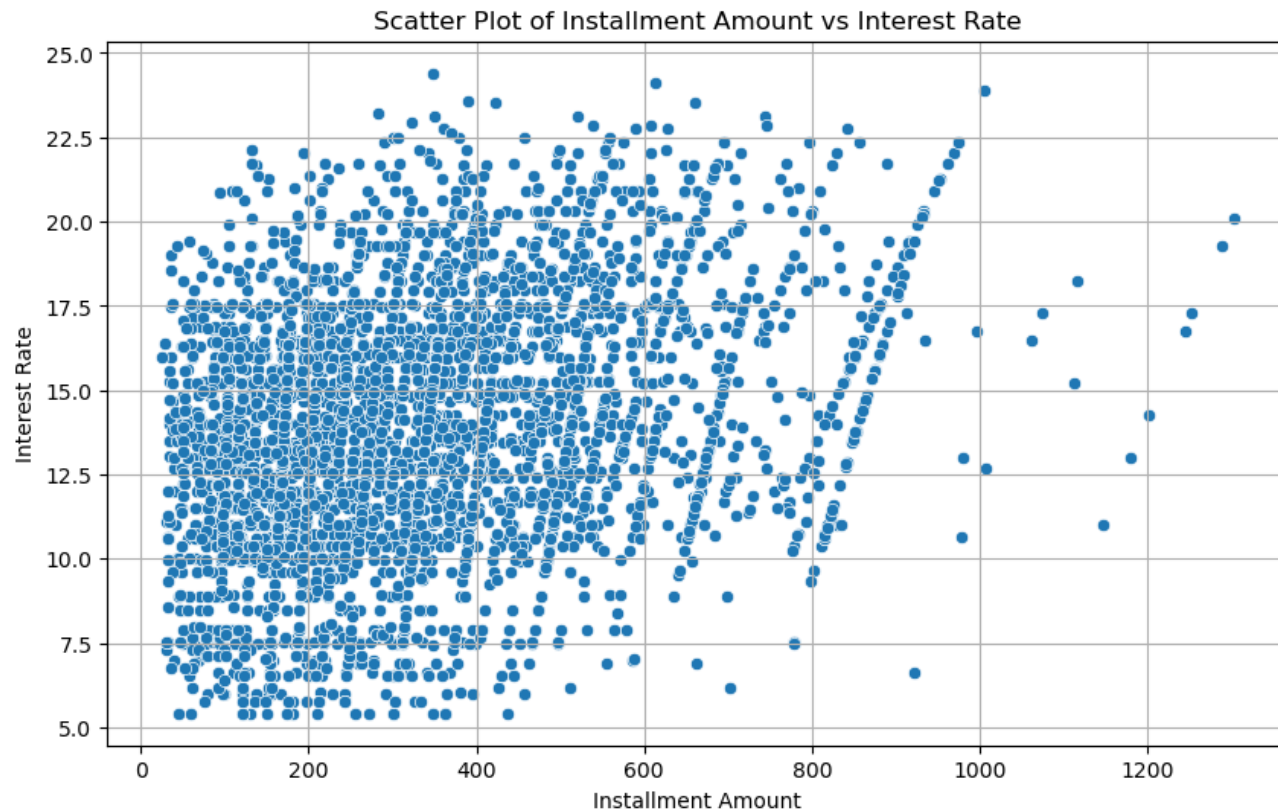
- Surprisingly, people with a smaller loan installment are defaulting the loan more

## 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 Defaulters in Lower-Middle Income Class for a correlation between monthly installment and interest rate

**Objective** – Since we know the largest population of loan defaulters 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%

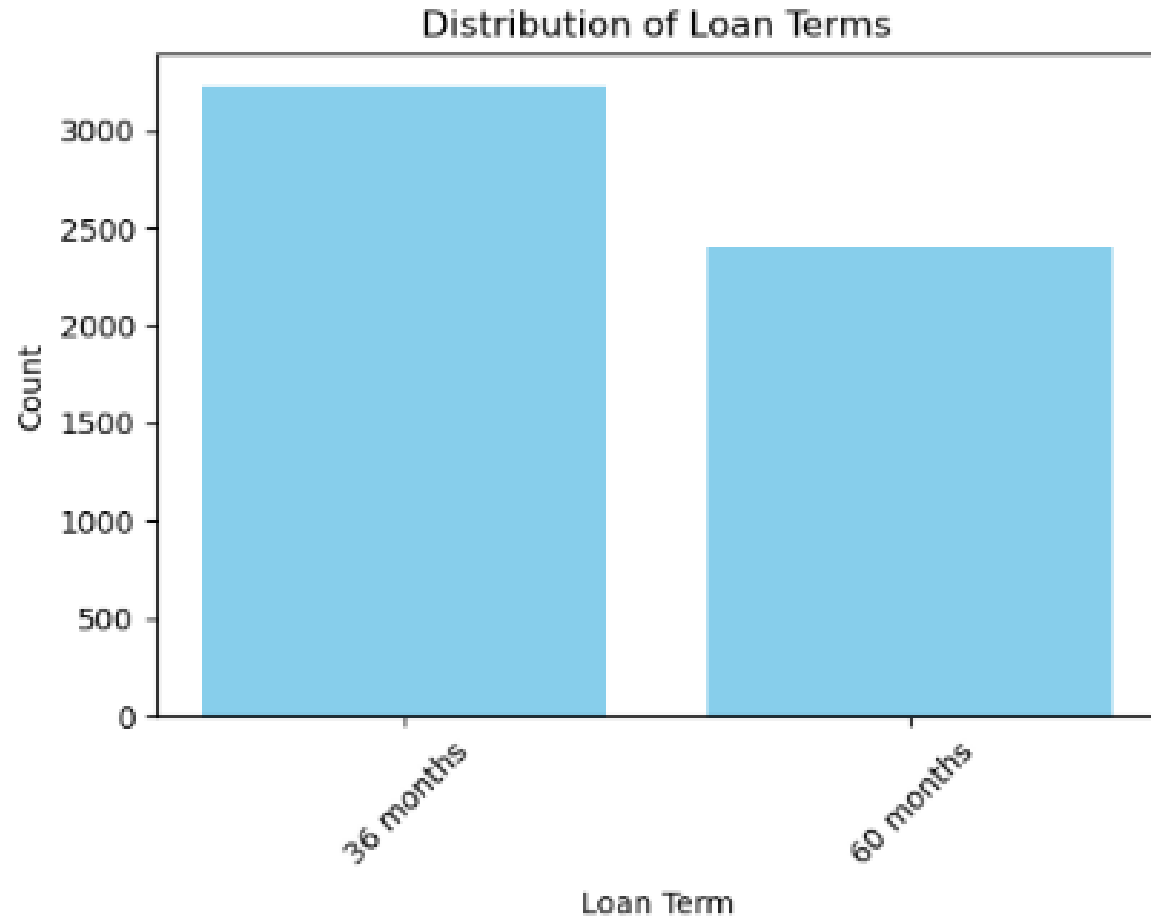
## Conclusion:

- 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:**

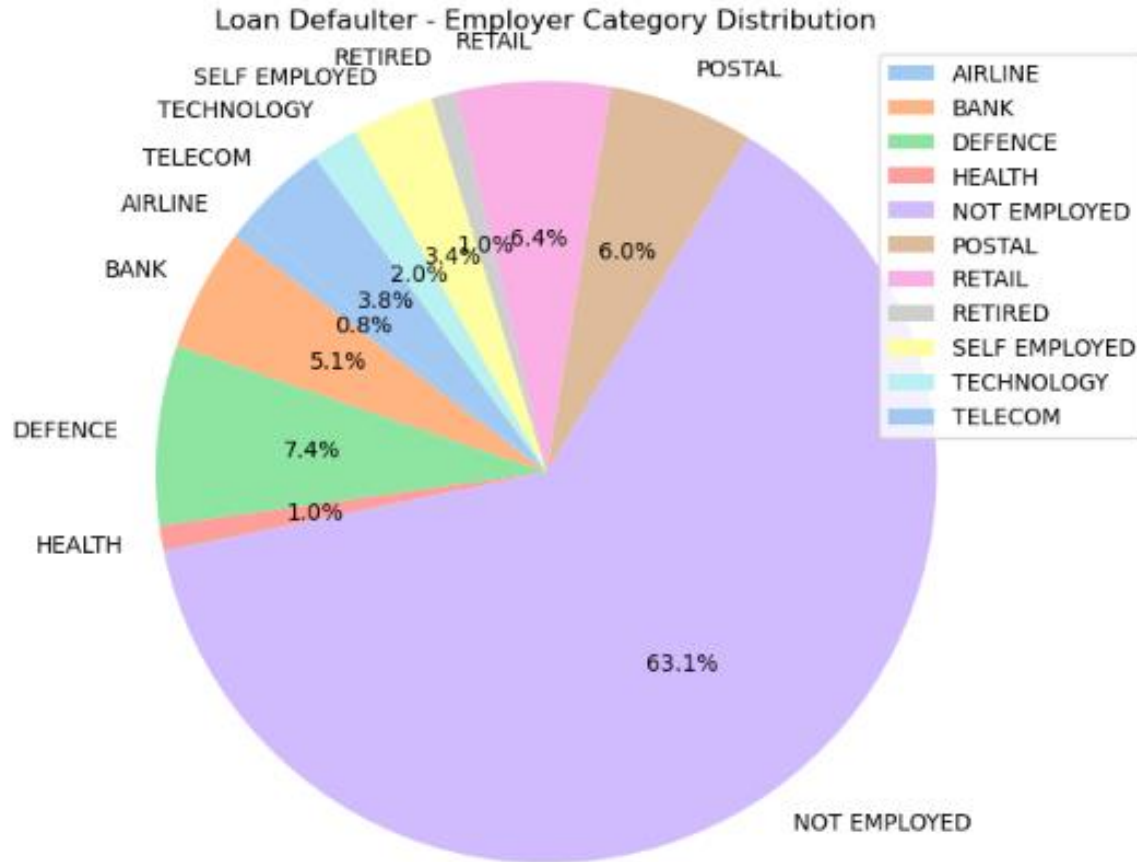
- The number of defaulters appeared to be slightly high in the 36 months loan term, but it is not very different from 60 months

**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:

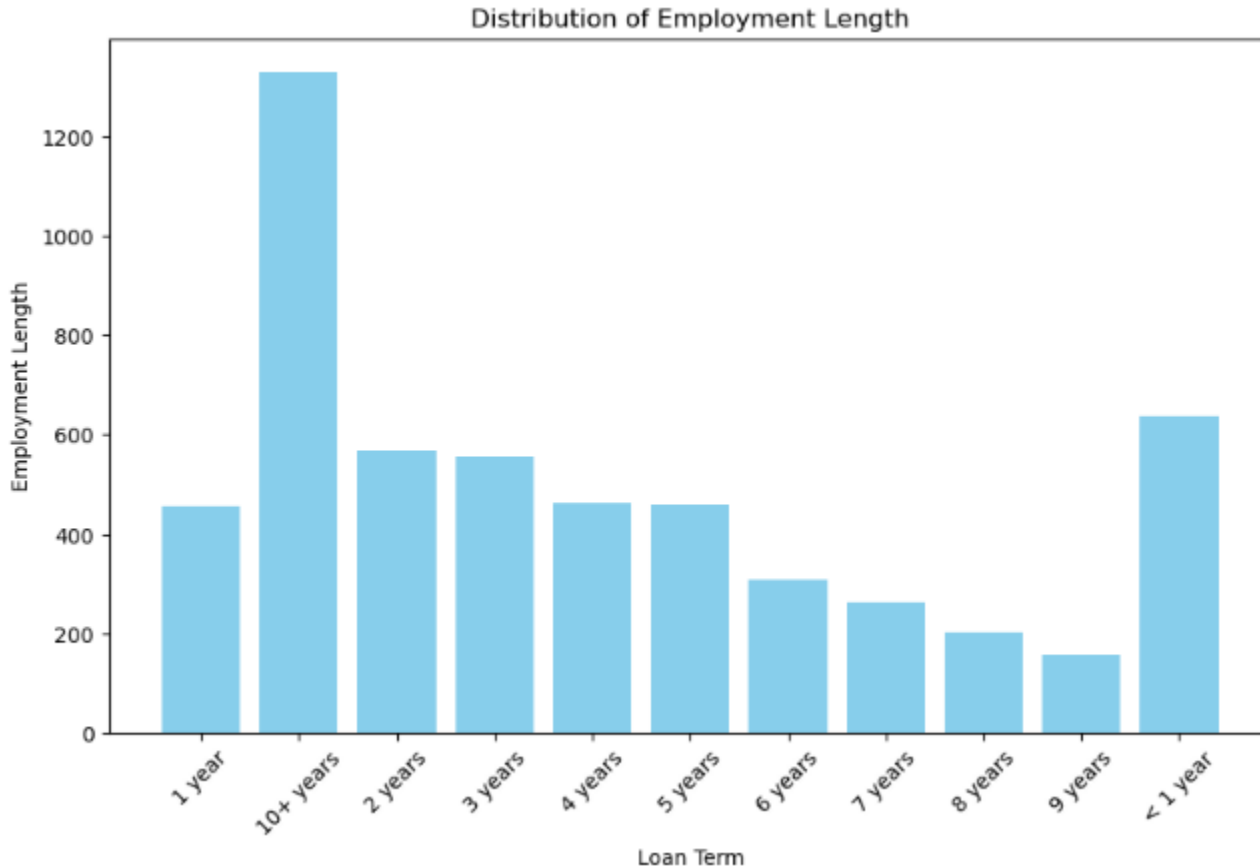
- People who are unemployed tends to be defaulting on their loans the most
- However, in the employed category, people whose employers are in Defense, Retail, and Postal service categories are defaulting more than the rest of the population

## 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:

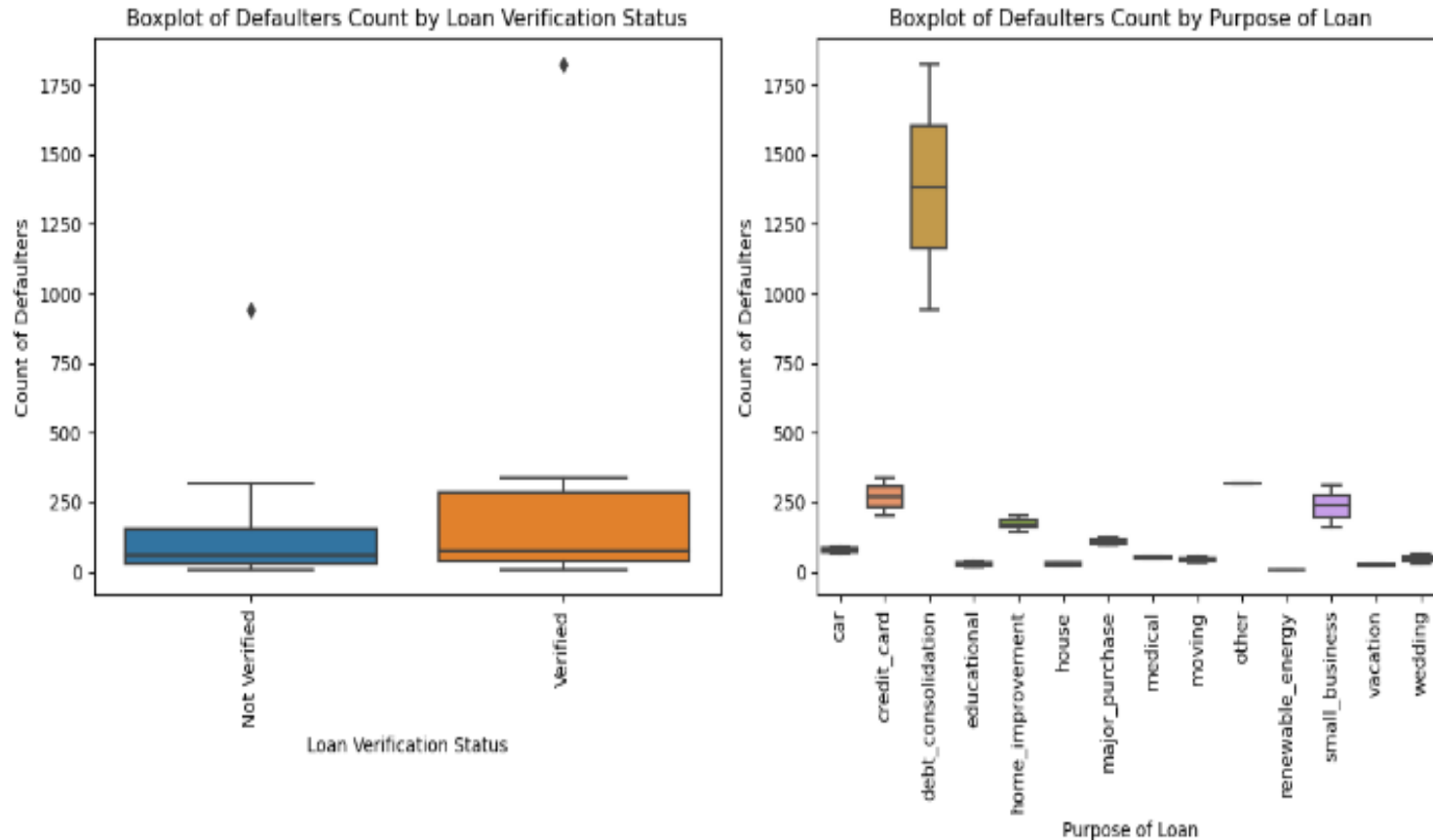
- People with 10+ years employment length are the highest loan defaulters
- In the group of people with less than 10 years of service, the highest loan defaulters falls into the <1 year of employment length category

## 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:

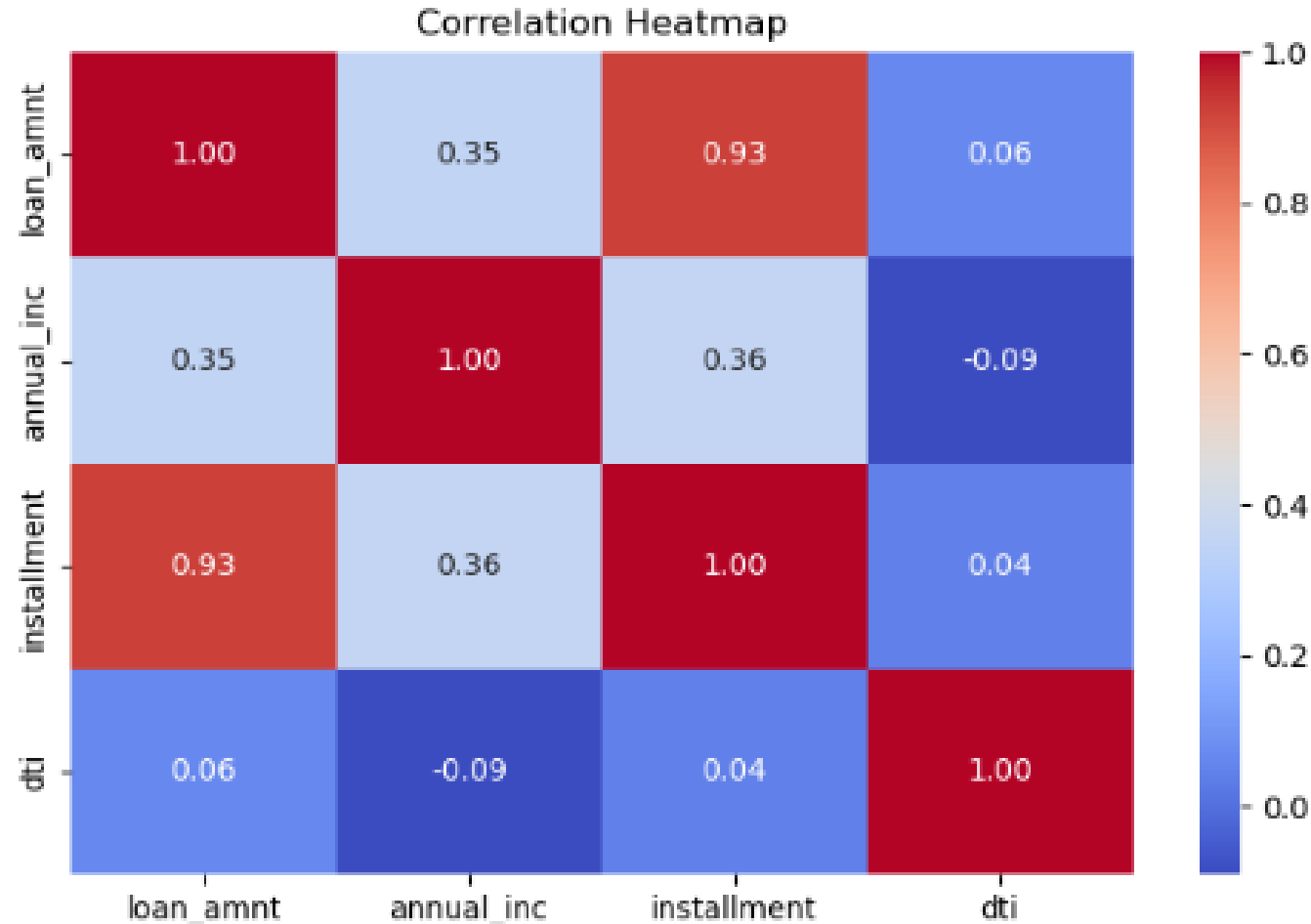
- Income verification seems to be not playing any major role on the loan being defaulted
- People taken loan for debt consolidation are the highest among defaulters
- The 25<sup>th</sup> and 75<sup>th</sup> quartiles of those defaulters who have taken loan for credit cards and small business are about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

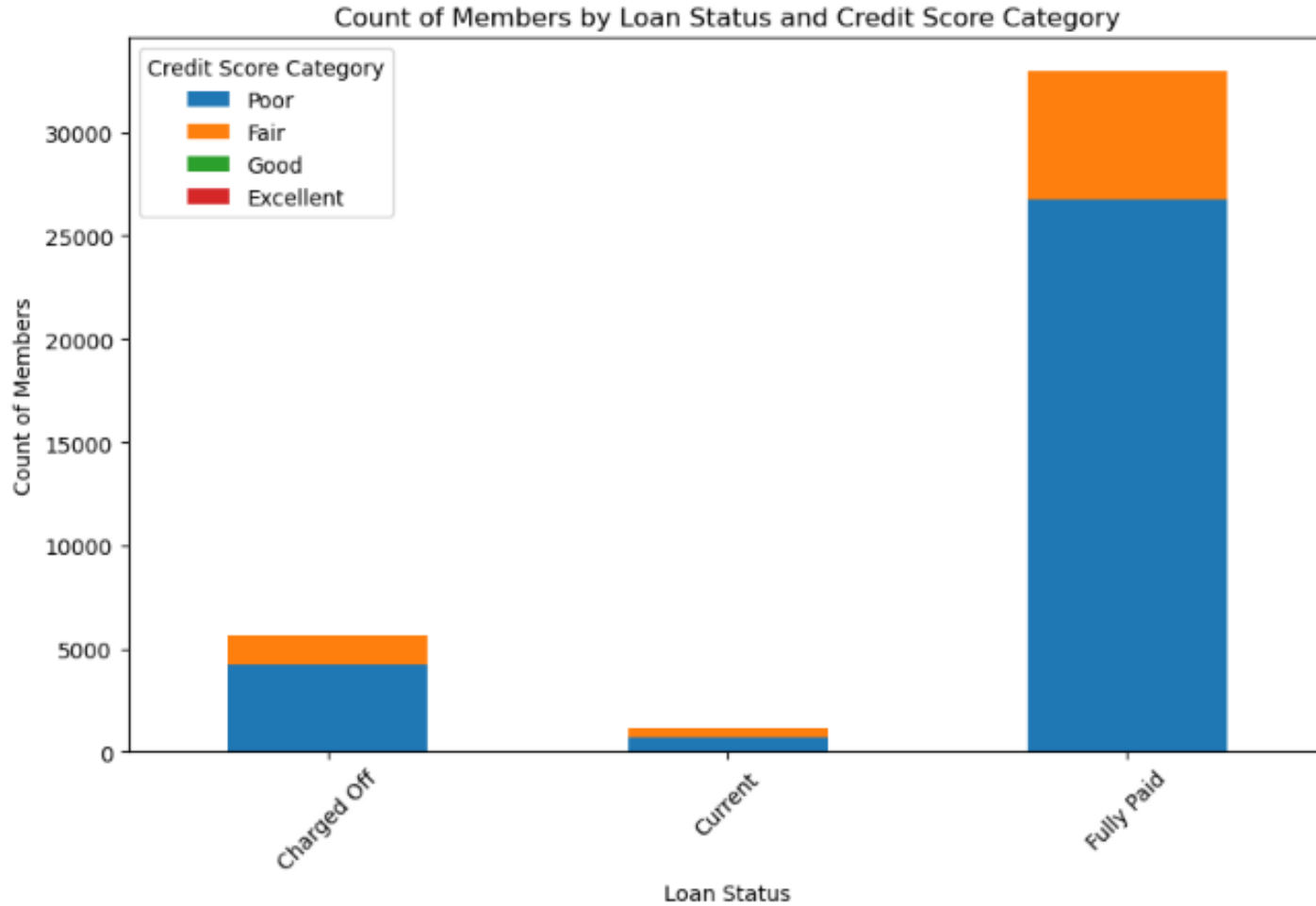
- Installment amount and loan amount are highly correlated
- Annual income and DTI are not correlated

## 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

# Analyzing Data for Current Loan Accounts

# Analysis on the loan collected

**Objective** – Determine how much of the current loan is already collected

	completed_term_percentage_group	group_count
0	0-25	0
1	26-50	0
2	51-75	0
3	76-99	1140

**Strategy:**

- Calculate the completed term percentage by taking total received principle divided by the funded amount.
- Then divide them in four buckets with ranges from 0-25%, 26%-50%, 51%-75%, and 76%-99%
- Take pivot on this buckets to find the counts

**Findings:**

- There are no loan where the term complete is less than 75%

**Conclusion:**

- 76% to 99% of all lend loan is already collected which is a good indication



# Analysis on the outstanding and late fee to funded ratios

**Objective** – Determine outstanding to funded, and late fee to funded ratio in percentages to find out the risk profile of all current loans

outstanding_per_bin mem_count		
0	0%-25%	1140
1	26%-50%	0
2	51%-75%	0
3	76%-100%	0

Outstanding Amount to Fund Ratio		
lat_fee_per_bin mem_count		
0	0%-25%	1140
1	26%-50%	0
2	51%-75%	0
3	76%-100%	0

Late Fee to Fund Ratio

**Strategy:**

- Calculate the outstanding amount, and late fee to fund ratio by taking those amounts divided by the funded amount.
- Then divide them in four buckets with ranges from 0-25%, 26%-50%, 51%-75%, and 76%-100%
- Take pivot on this buckets to find the counts

**Findings:**

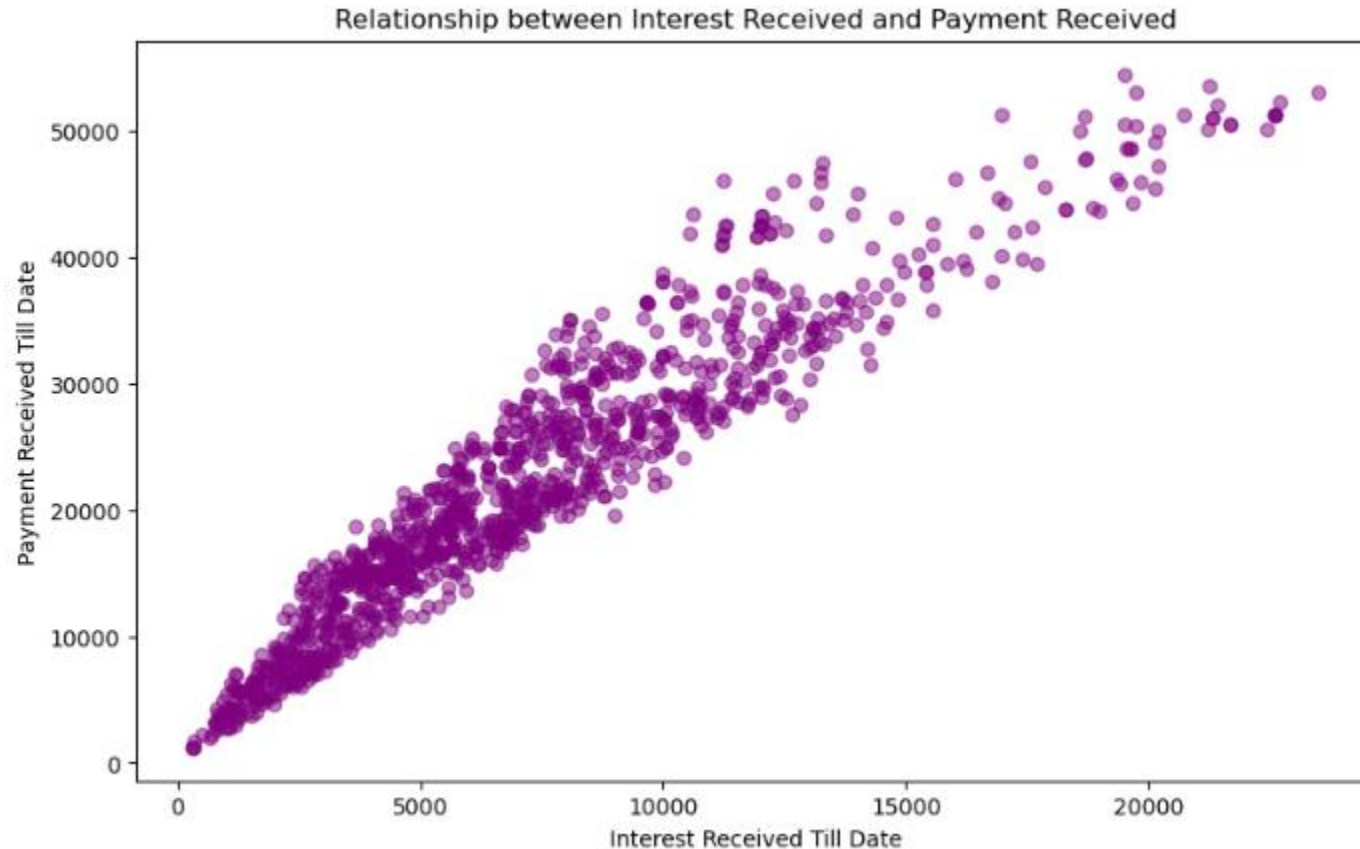
- There are no loan where the outstanding amount, and late fee ratio are greater than 25%

**Conclusion:**

- The outstanding amount to funded amount and late fee to funded amount ratios within 0%-25% indicates a low risk profile

# Analysis on the payment received vs. interest received

**Objective** – Determine if the interests are received along with payments and they are correlated



**Strategy:**

- Taken the total interest received and the total payment received columns
- Plotted a scatterplot to find the correlation

**Findings:**

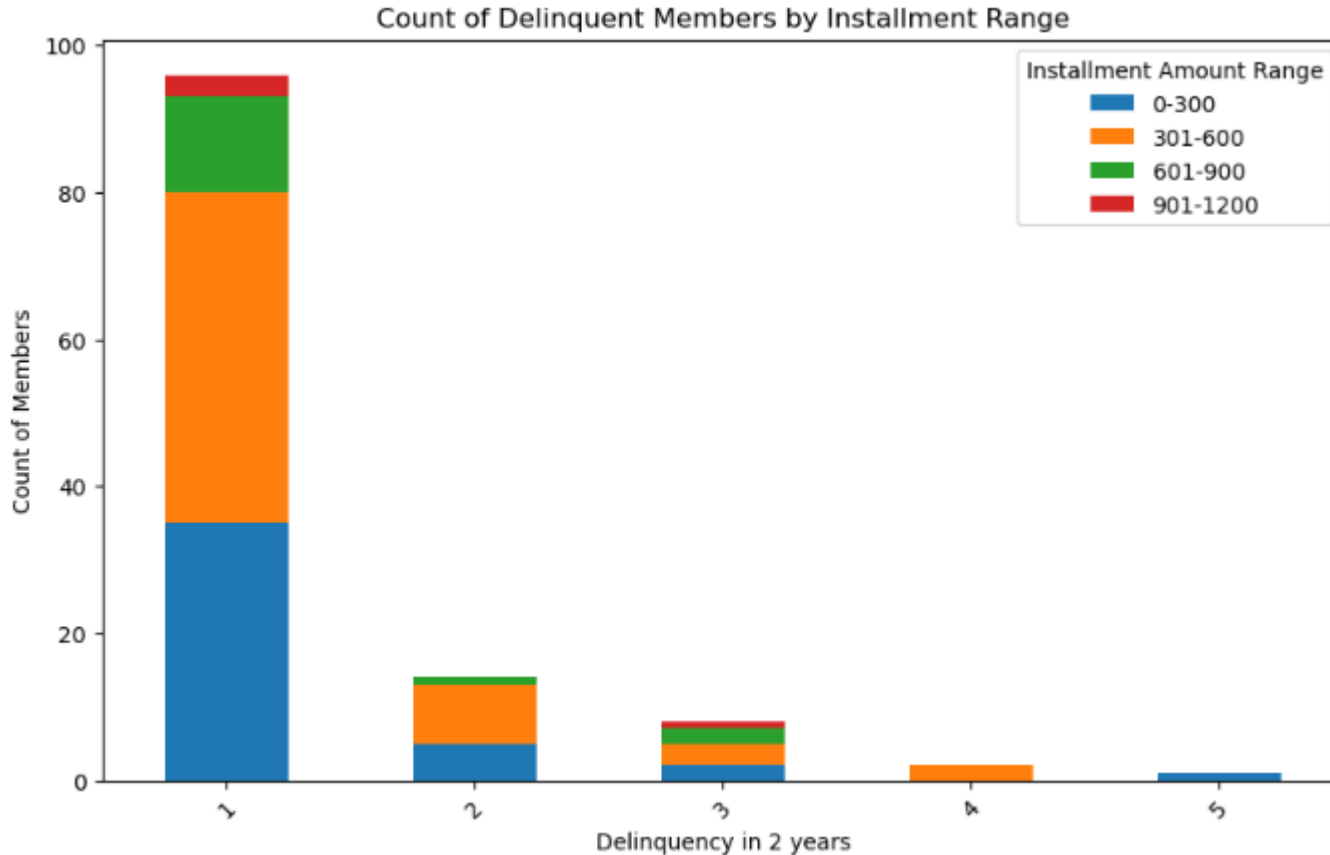
- We can see the interests are matching with payment received and the scatterplot denotes almost a straight line.

**Conclusion:**

- It seems borrower has paid off all accrued interest charges up to the current date.
- This suggests that the borrower is making timely and sufficient payments to cover both the principal amount and the interest charges, which is a positive indicator of loan performance

# Analysis Past 2 years Delinquency count w.r.t. Installment Payment

**Objective** – Determine if the number of times a borrower is being delinquent in the past 2 years has any bearing on their installment amount



## Strategy:

- Taken the number of times members being delinquent in the past 2 years
- Grouped their installments in 4 buckets
- Plotted a stack bar chart to see the correlation

## Findings:

- In the last 2 years, the highest number of times borrowers being delinquent is 1 time
- Borrowers who are paying an installment between 301 to 600 dollars are highest among the delinquents

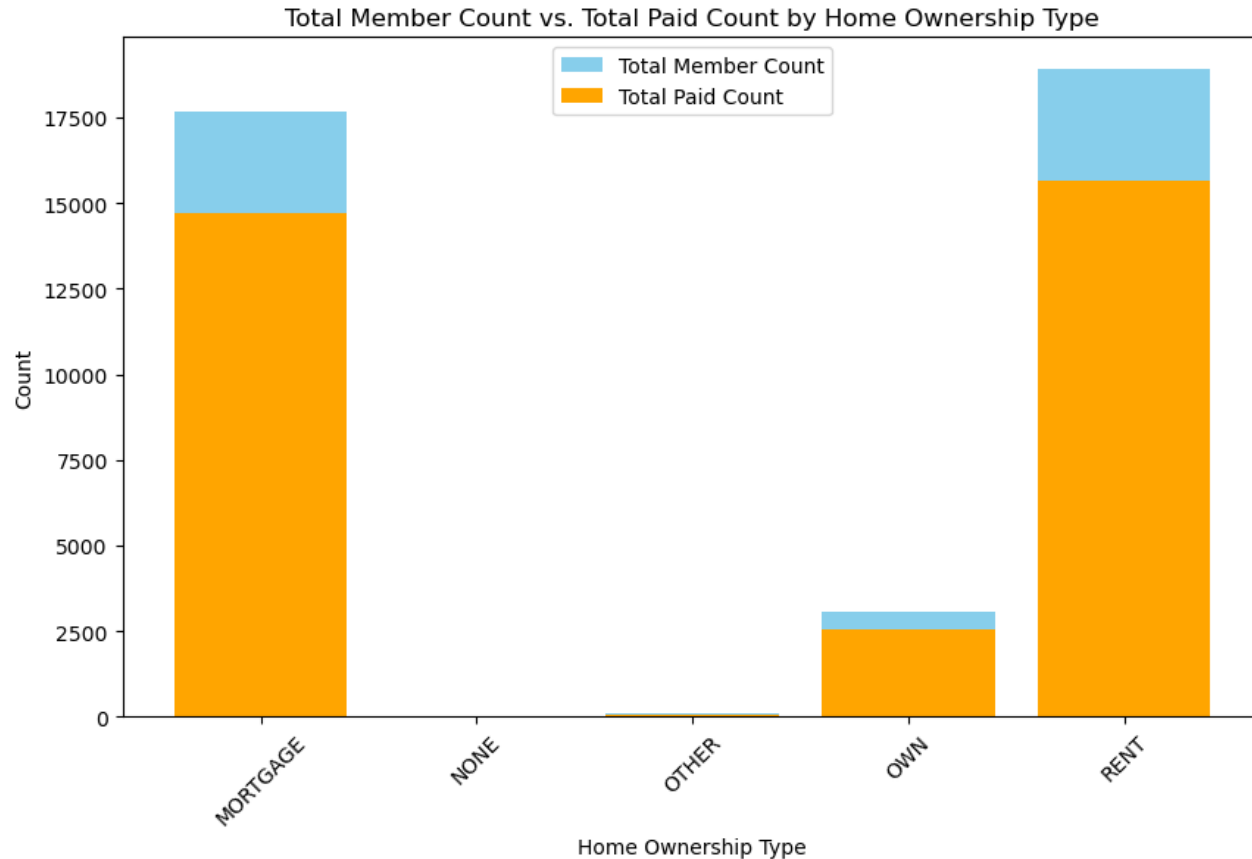
## Conclusion:

- It seems borrowers who are paying an installment amount  $\leq \$600$  are being delinquent at least once.
- The lending club may communicate the late /delinquency fee policies to such borrowers

# Analyzing Data for Fully Paid Accounts

# Analysis the home ownership of the fully paid loans

**Objective** – Determine how the home ownership impacts on the fully paid loans



## Strategy:

- Taken all the paid off loans
- Grouped the data by home ownership on the paid of loan dataset
- Grouped the data by home ownership on the entire dataset
- Merged the datasets and plotted a bar chart

## Findings:

- Borrowers are the largest where the home ownership is either Mortgage or Rent.

## Conclusion:

- ~82% of all borrowers in each home ownership category have paid off their loans.

# Analysis the employment length of the fully paid loans

**Objective** – Determine how the employment length impacts on the fully paid loans

	emp_length	mem_count_original	mem_count_paid	percentage
10	9 years	1258	1068	84.90
2	2 years	4388	3724	84.87
1	< 1 year	4583	3869	84.42
3	3 years	4095	3457	84.42
4	4 years	3436	2880	83.82
6	1 year	3240	2713	83.73
7	6 years	2229	1861	83.49
5	5 years	3282	2736	83.36
9	8 years	1479	1232	83.30
8	7 years	1773	1448	81.67
0	10+ years	8879	7157	80.61

## Strategy:

- Taken all the paid off loans
- Grouped the data by employment length on the paid of loan dataset
- Grouped the data by employment length on the entire dataset
- Displayed the result

## Findings:

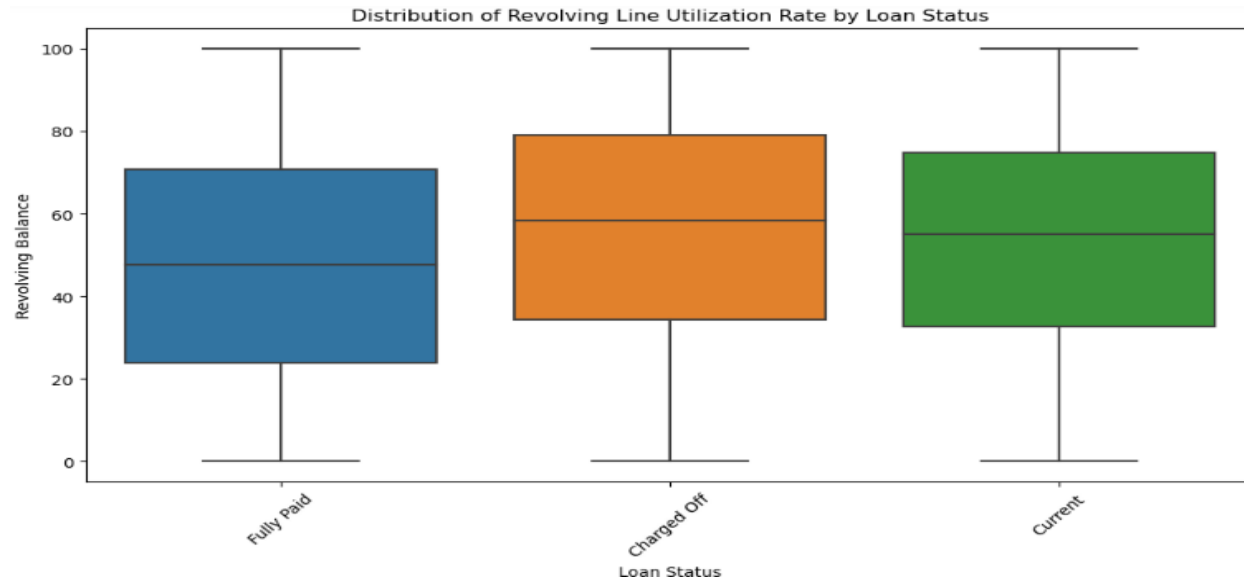
- Borrowers who have more than 10+ years of employment are the lowest to pay off their loans.

## Conclusion:

- >80% of all borrowers in each employment length category have paid off their loans.

# Analysis the revolving line utilization rate by loan status

**Objective** – Determine the distribution of the amount of credit the borrower is using relative to all available revolving credit under different loan status



## Strategy:

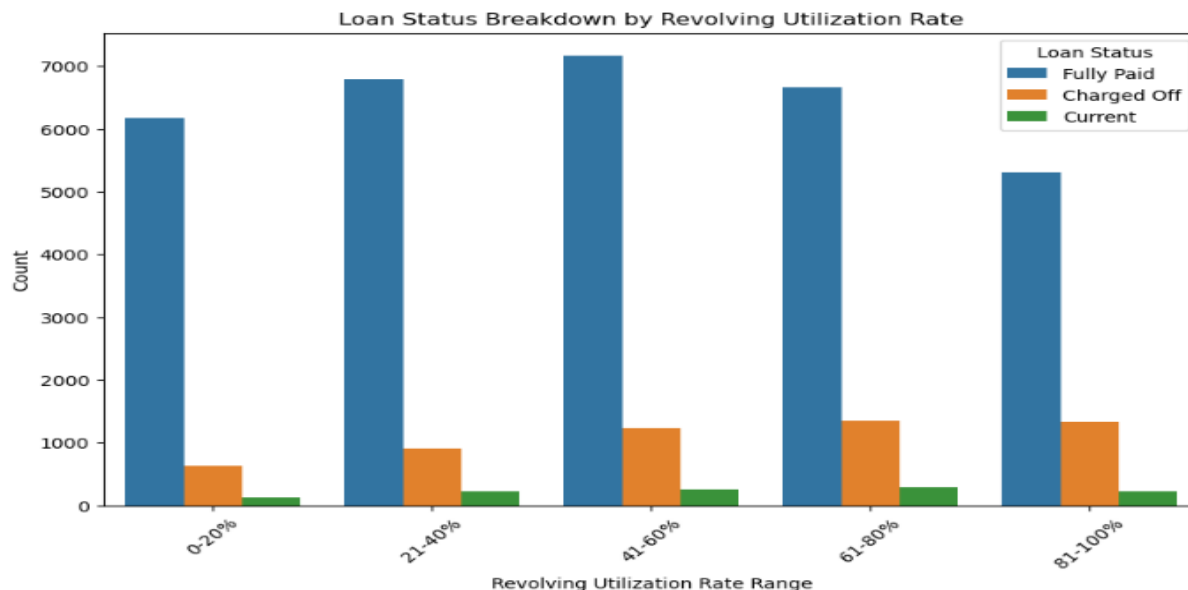
- Taken all the records in the entire dataset
- Converted the revolving utilization rate to a float value
- Plotted a box plot and a bar chart under each loan status

## Findings:

- Charged off loans are using the higher percentage of revolving line utilization rate

## Conclusion:

- A revolving line utilization rate typically denotes the ratio of a person's credit debt to their total credit limit.
- We can see, the median of the revolving line utilization rate for the fully paid and current loans are almost same, however it is much higher for the charged off loans.
- As a result the lending club may pay caution while granting loan to those borrowers whose revolving line utilization rate is higher than 50%, because they are most likely become defaulters



Thank you