

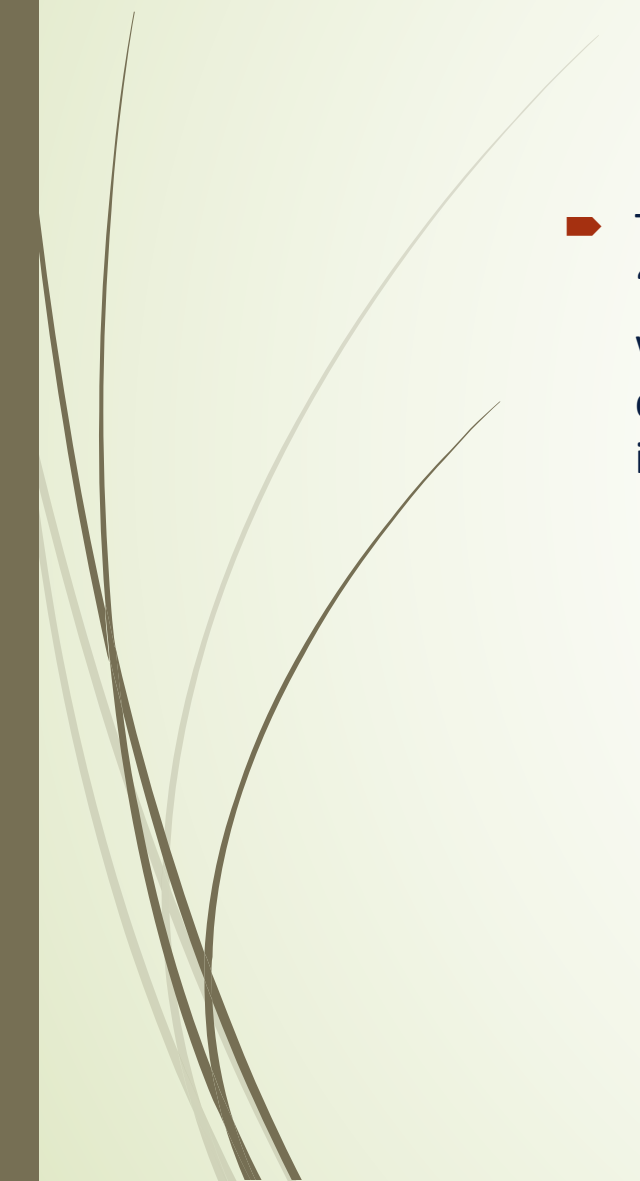


# LENDER CLUB CASE STUDY

- VIBHAV VERMA




# PROBLEM STATEMENT

- The loan dataset contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns using Exploratory Data Analysis(EDA) 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.
- 



# SOLUTION

- The below EDA steps were taken for the analysis:
    - Data Cleaning
    - Univariate and Segmented Analysis
    - Bivariate Analysis
    - Derived Metrics
- 



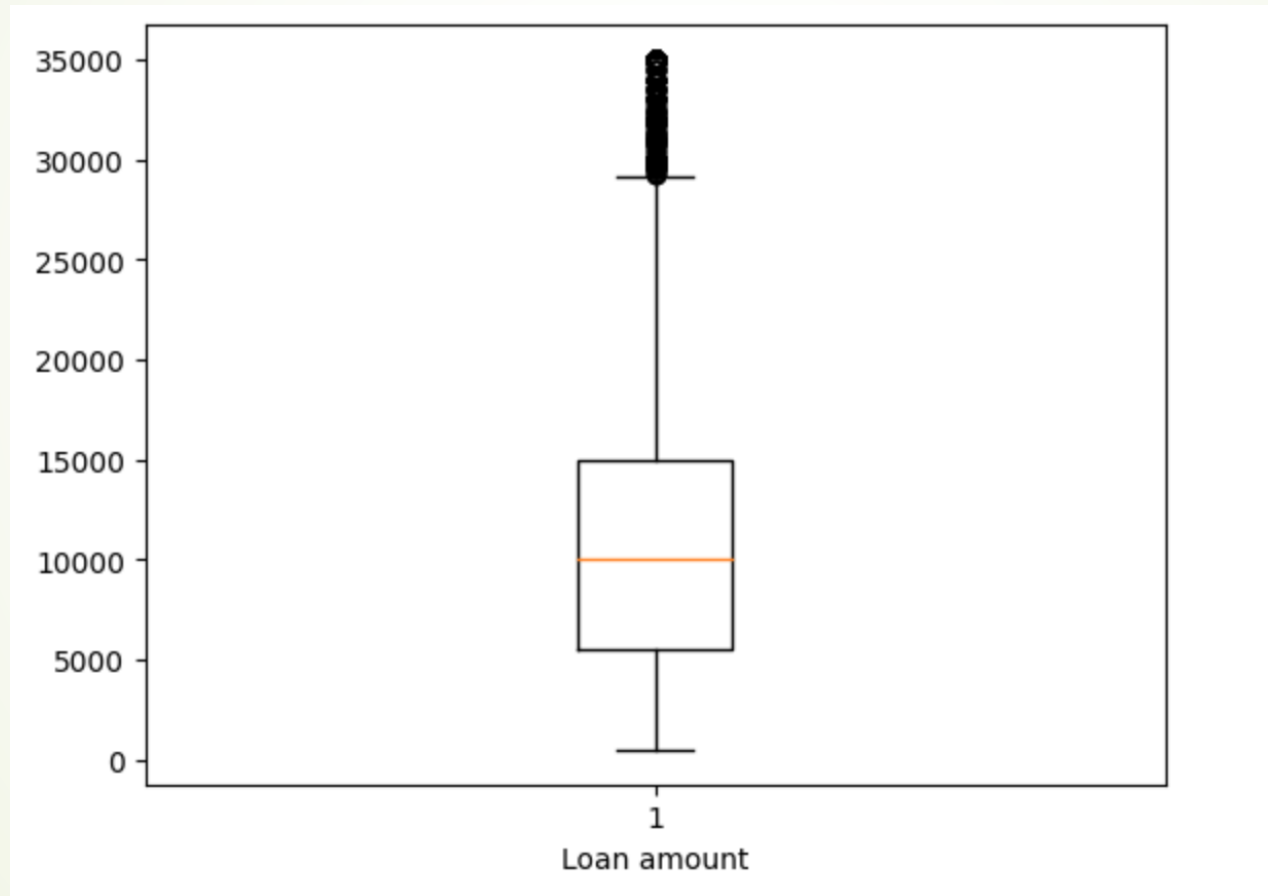
# DATA CLEANING



- Dropped columns with more than 75% missing values.
- Changed datatype of “next\_pymnt\_d” column to string as it had mixed types.
- Removed redundant string “month” from “term” column as from the data dictionary its understood that it is in months. Its data type was also changed to integer.
- Removed “%” string from “int\_rate” column as its understood that rate of interest is in percentage. Its data type was also changed from string to float.

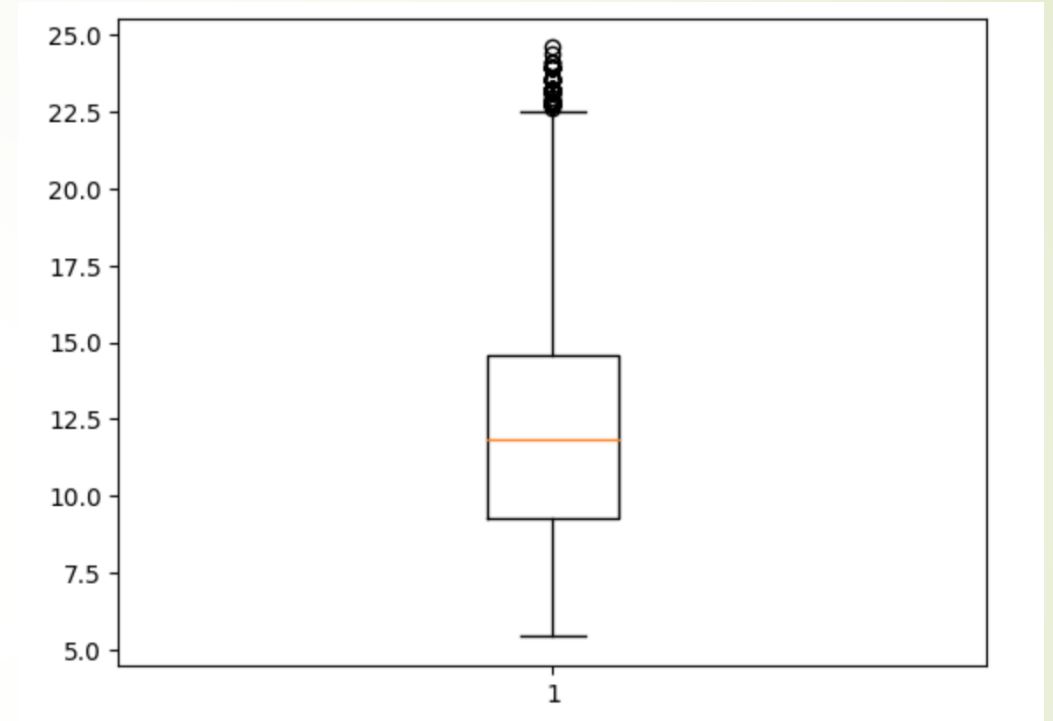
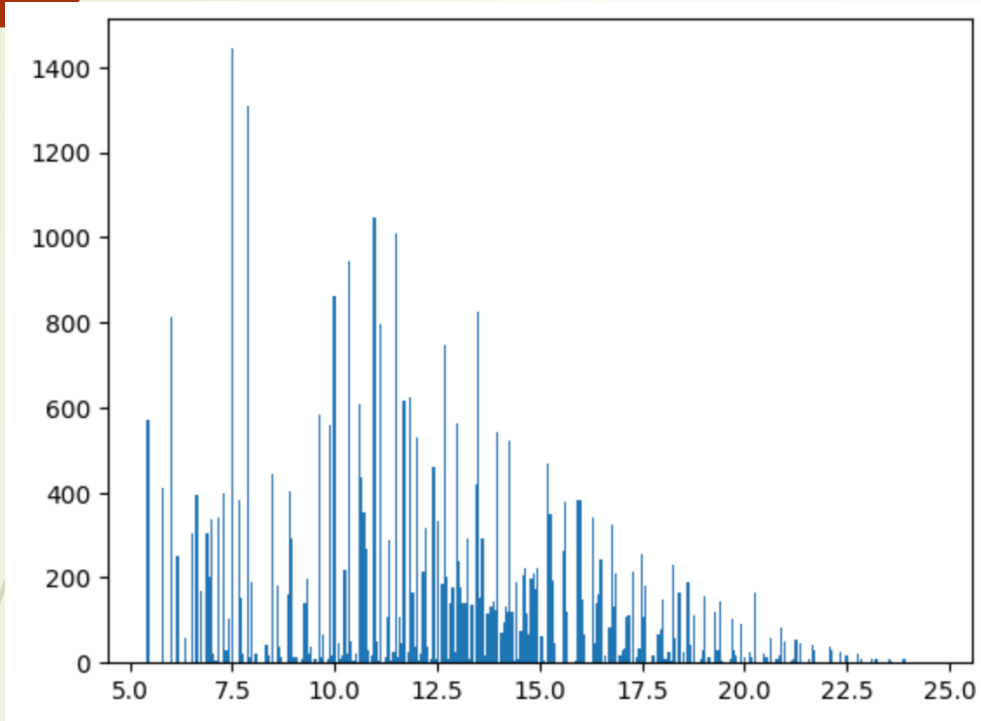
# UNIVARIATE ANALYSIS

## Loan amount



count	39717.000000
mean	11219.443815
std	7456.670694
min	500.000000
25%	5500.000000
50%	10000.000000
75%	15000.000000
max	35000.000000

## Rate of interest

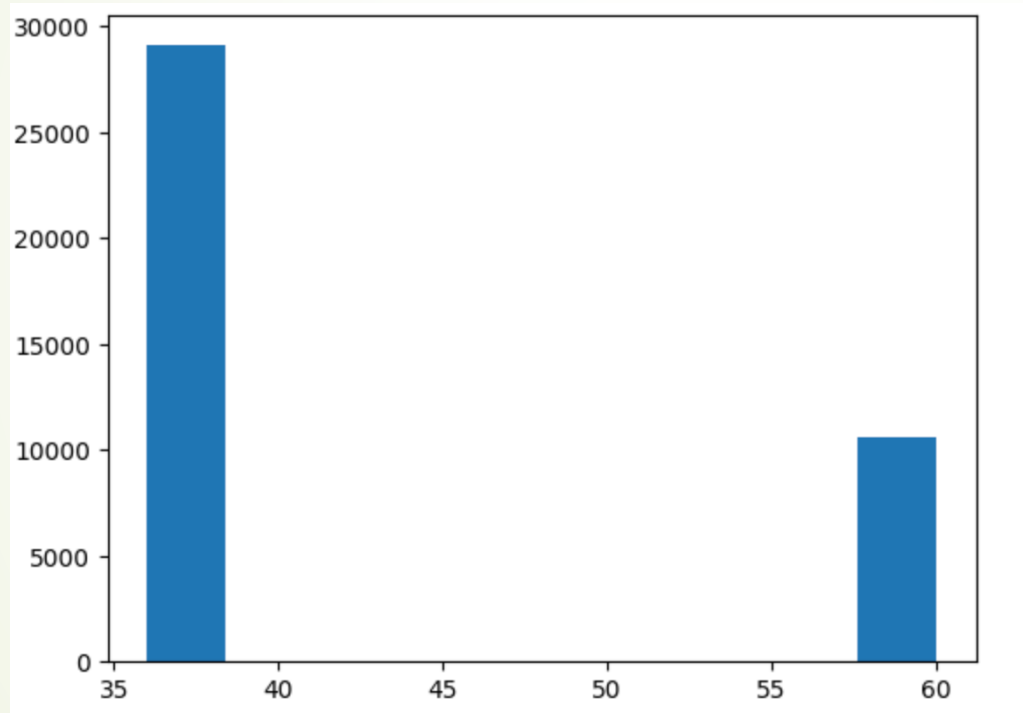


### Insights:

1. 50% of the loans lie in the range [5500, 15000].
2. Median loan amount is Rs. 10000.
3. 25% of loan amounts lie in the range [15000, 35000].
4. Around 1400 customers have an interest rate of 7.5%.
5. 50% of customers have interest rate between 9% to 14%.

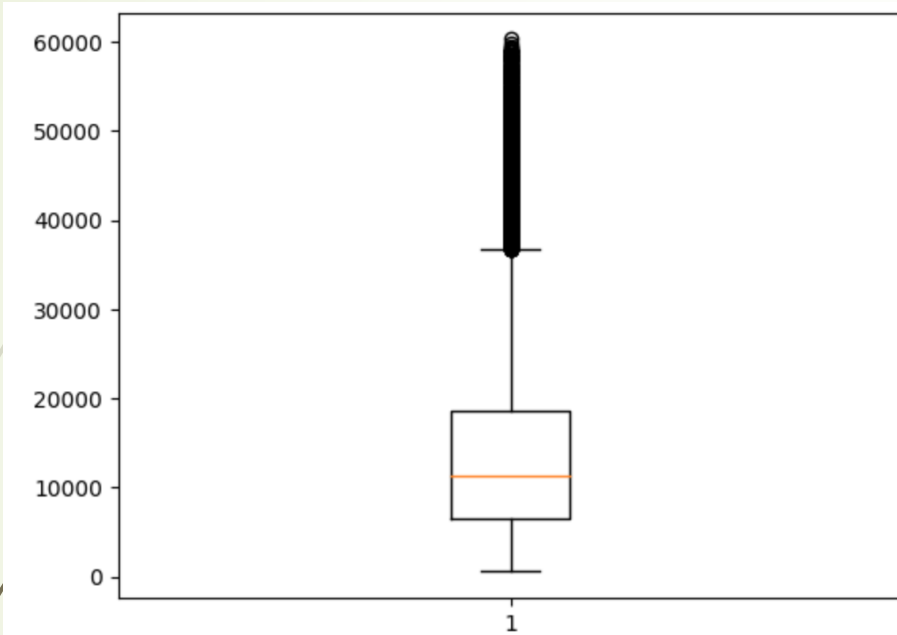
# SEGMENTED AND BIVARIATE ANALYSIS

- Data is grouped on loan time period



- Among customers having 60 months of interest term, 22% default their loans.
- Among customers having 36 months of interest term, 11% default their loans.
- So, people with higher loan time period have a higher tendency to default their loans.

A new metric "Principal amount" was derived from installment and loan time period.



loan_status	principal_amnt						
	mean	median	mode	max	min	q25	q75
Charged Off	15889.188868	12621.00	16563.60	60303.6	1049.76	6982.20	22121.4
Current	23728.563684	22091.40	16563.60	58887.0	1304.40	15508.20	31183.5
Fully Paid	13332.984528	11062.08	11199.96	59466.0	564.84	6315.03	17746.2

Principal amount aggregation on loan status

Among the 5627 customers that have defaulted:

1. 25% customers have principal amount less than ₹ 7000 approximately.
2. 50% customers have principal amount between ₹ 7000 and ₹ 22000.
3. 25% customers have principal amount above ₹ 22000 highest being close to ₹ 60000.

This proves the pattern that customers with high principal amount are more likely to default their loans.



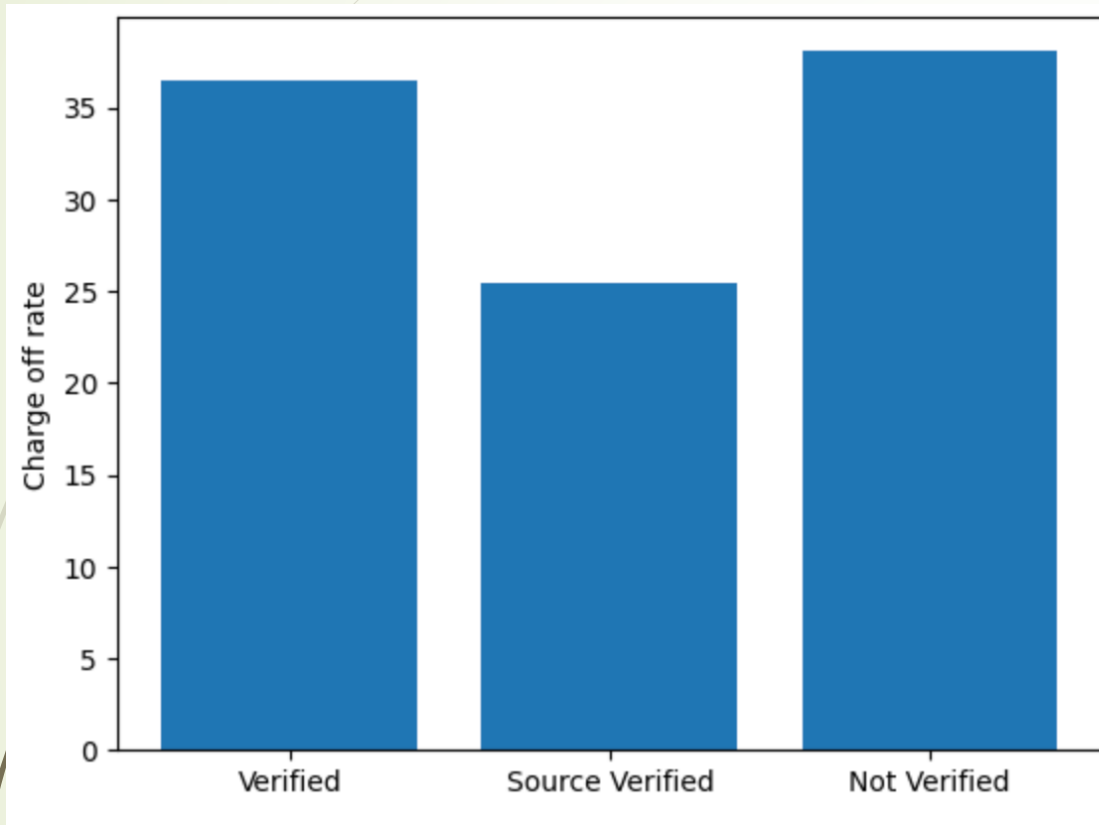
## Annual income aggregation on loan status

loan_status	annual_inc						
	mean	median	mode	max	min	q25	q75
<b>Charged Off</b>	62427.298034	53000.0	60000.0	1250000.0	4080.0	37000.00	75000.0
<b>Current</b>	75430.665105	65000.0	60000.0	650000.0	9960.0	47475.00	90000.0
<b>Fully Paid</b>	69862.503328	60000.0	60000.0	6000000.0	4000.0	41132.75	84000.0

1. The mean and median income of defaulted customers is the lowest.
2. The inter quartile range of annual income of defaulted customers is also the lowest among all.

So, customers having an annual income less than ₹ 75000 are more likely to default their loans.

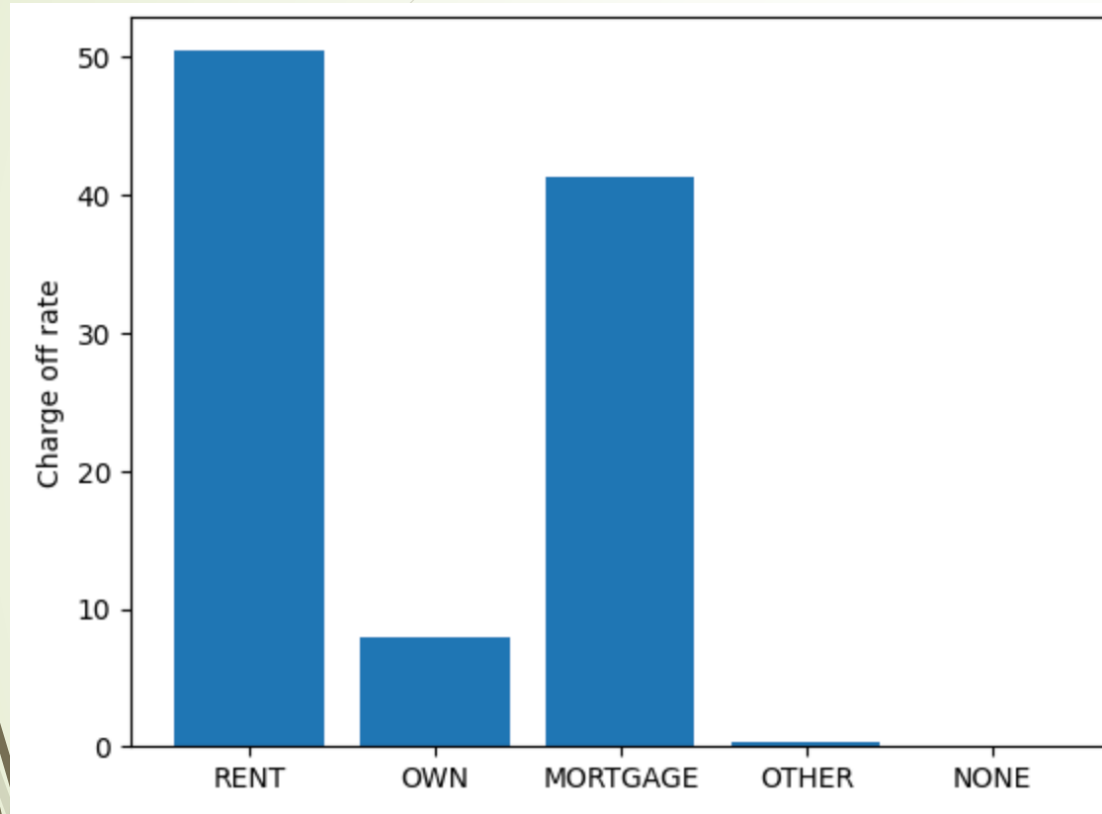
## Data grouped on income verification status and loan status



Default rate categorized on income verification status

- There isn't a direct correlation between the income verification status and loan default rate.
- This is since charge off rate is the roughly the same for loans that were verified and unverified loans.
- This raises the question on the bank's should income verification process. It probably needs an improvement.
- Charge off rate is lower for customers whose income source was verified. This might mean that these customers are working in a legitimate company and have correctly reported their income.
- This might also mean that not all verified income customers have reported their income honestly. They might have a business of their own due to which they are able to do so.

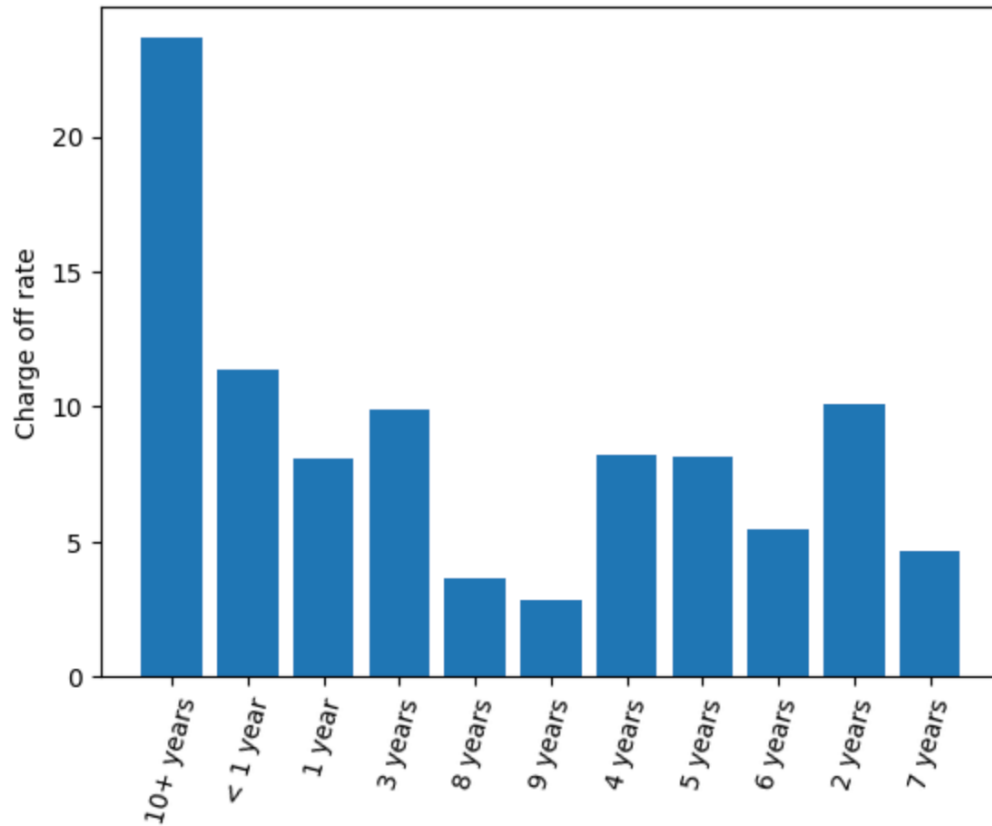
## Data grouped on home ownership and loan status



Default rate categorized on home ownership

- Customers who either live on rent or mortgage have around 50% and 40% charge off rate. Customers who own their property, have a charge off rate around 8%.
- This tells us that customers living on rent or mortgage are more likely to default their loans as they spend a part of their income on rent as well. This leaves them with little income to pay as salary.
- So, the bank should avoid giving loans to customers that don't own their place of living.

## Data grouped on employee length and loan status



- The charge off rate is highest in customers having more than 10 years of experience. This seems a bit counterintuitive as salary increases with employment length.
- One of the possible reasons can be that as employee grows in life, his dependable also increase like spouse, children, old parents etc. Due to this, he/she has more expenditures to take care of and hence is likely to default more often.

Default rate categorized on employee length



# ANALYSIS SUMMARY



1. Customers with higher loan time period have a higher tendency to default their loans.
2. Customers with high principal amount are more likely to default their loans.
3. Customers with high interest rate are more likely to default their loans.
4. Customers having an annual income less than ₹75000 are more likely to default their loans.
5. Customers living on rent or mortgage have a 50% rate of loan defaulting among the total defaulters.
6. Customers who have spent more than 10 years in a company are more likely to default than others.