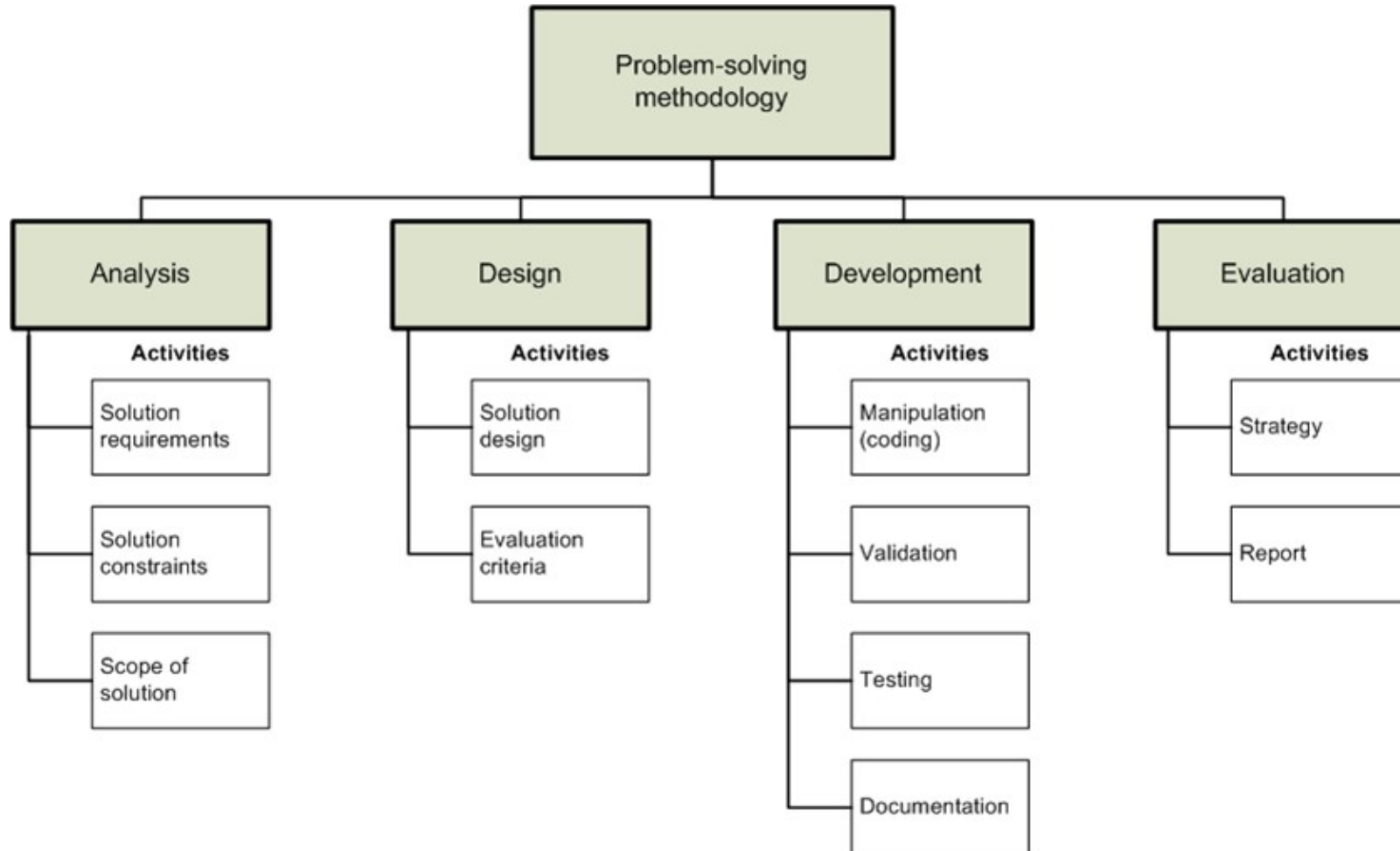


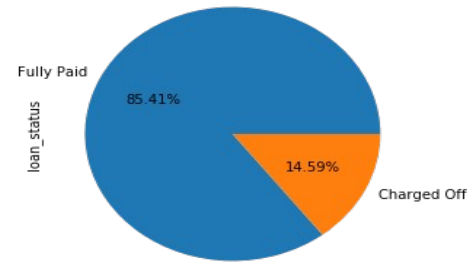
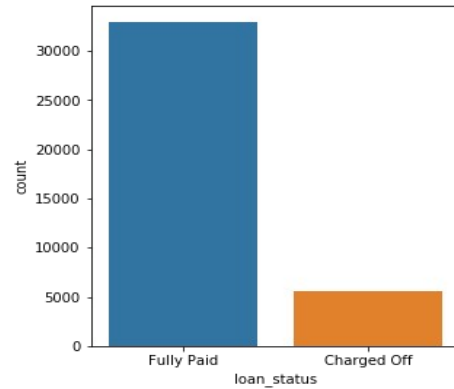
LENDING CLUB CASE STUDY ASSIGNMENT SUBMISSION

Name: Akshay Jindal, Bighnesh Mishra

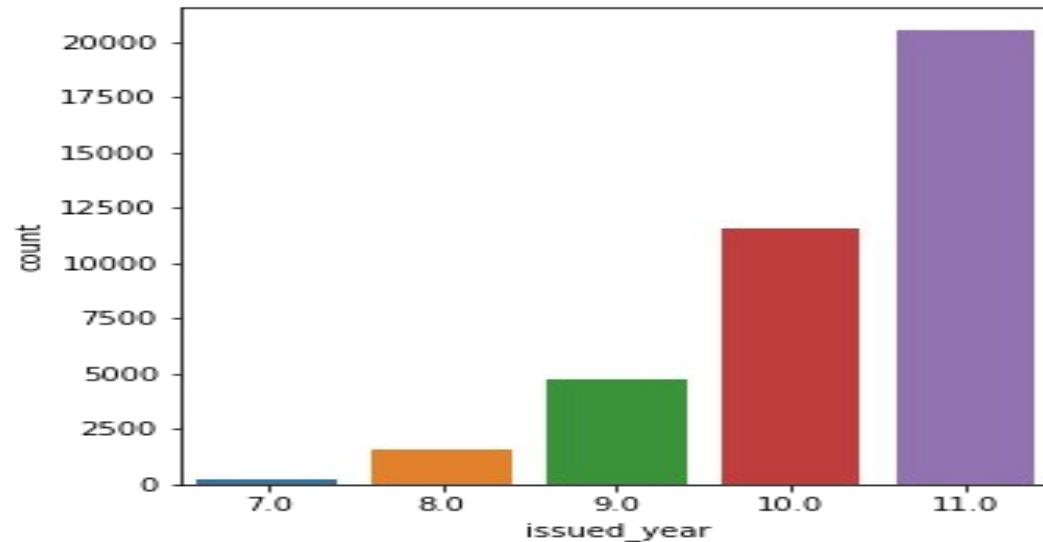
Problem Solving Methodology



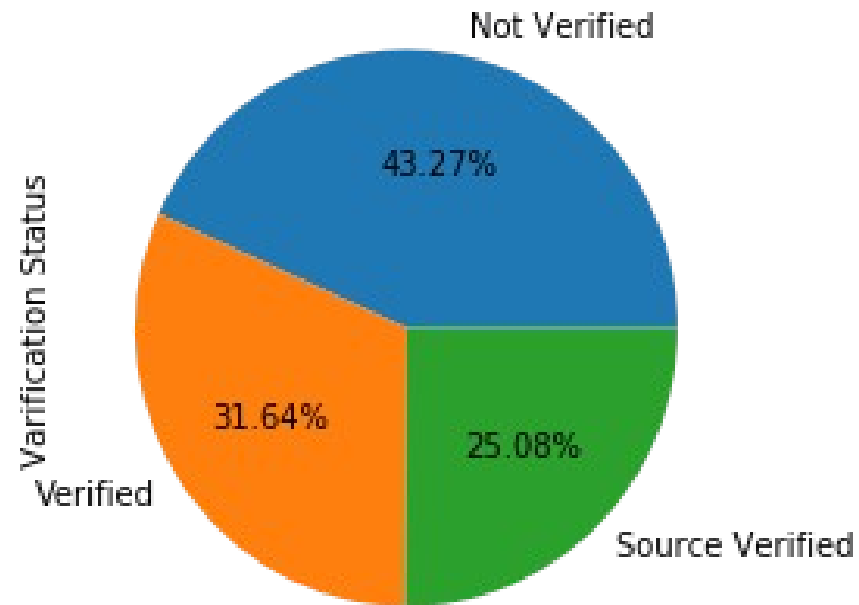
Increase of loan distribution over the years



- The study has been performed on data-set (2007-2011) in which 14.59 % of Charged Off users exist
- Over the years the loan distribution also increased. A huge boost can be seen in year 2011.



Percentage of verified and non-verified borrowers



Most of the users are not verified

- Most of the loan users are not verified. It could be a factor of increasing Charged Off users.
- 5.55 % Charged Off users are not verified.

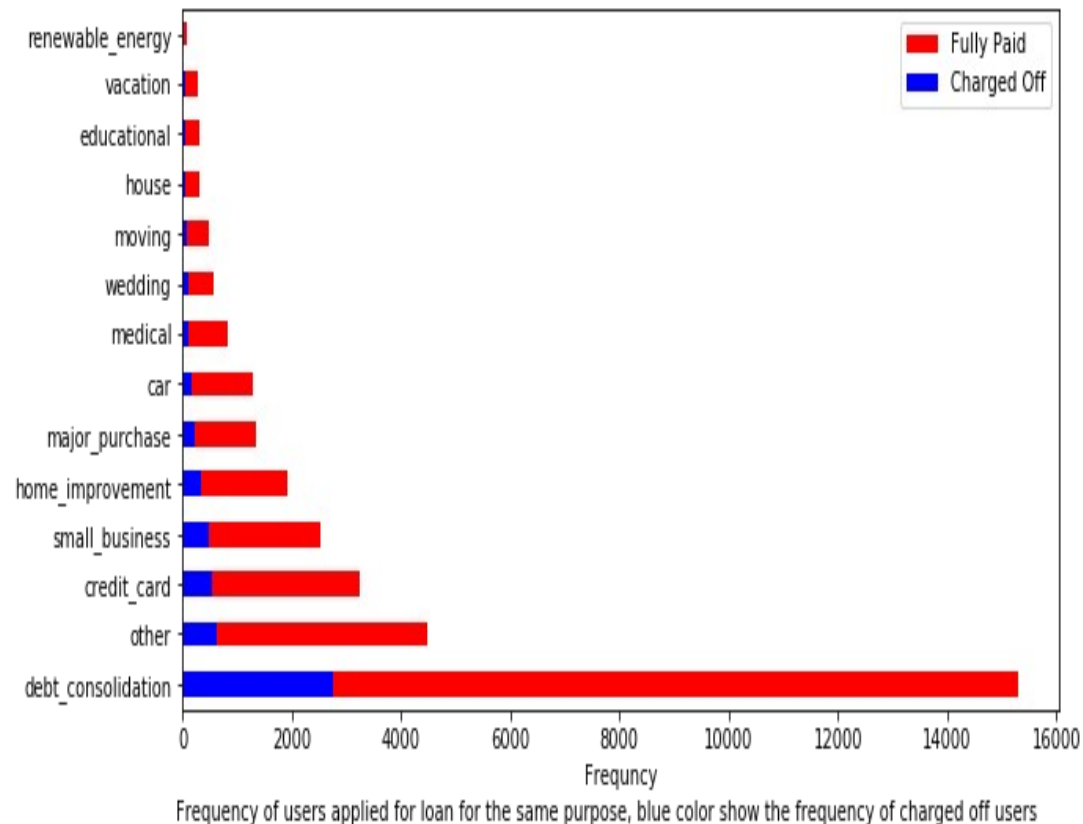
Verification status	Not Verified	Source Verified	Verified
Loan status			
Charged Off	5.55	3.72	5.32
Fully Paid	37.72	21.37	26.32

Correlation between loan attributes

	dti	funded_amnt	loan_amnt	funded_amnt_inv	annual_inc
dti	1	0.0621941	0.0624358	0.070663	-0.12153
funded_amnt	0.0621941	1	0.98179	0.956172	0.264798
loan_amnt	0.0624358	0.98179	1	0.937922	0.268999
funded_amnt_inv	0.070663	0.956172	0.937922	1	0.251981
annual_inc	-0.12153	0.264798	0.268999	0.251981	1

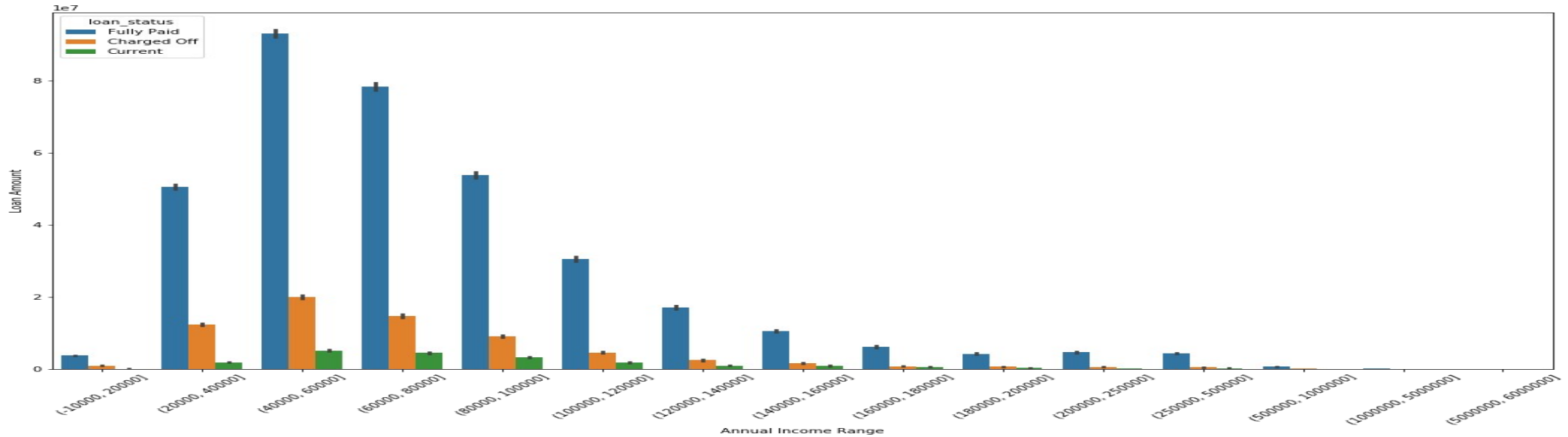
1. The above table a very positive correlation can be seen between funded amount and loan amount.
2. The positive relation between loan amount and amount funded by the investors.
i.e The amount whatever recommended by the lending club most of the time investors also issue the same or a little less amount to the borrower.
3. A negative relation between DTI and annual income. i.e increment in the DTI decreases the annual income.

Users' purpose of taking loan



- Here the blue line indicates the Charged Off users.
- Red line indicates the count of Fully paid users.
- Most of the users had taken loan for the purpose of debt consolidation.
- The top three purposes are as follow:
 1. Debt Consolidation
 2. Other
 3. Credit Card.

Annual income range which holds the maximum number of defaulters



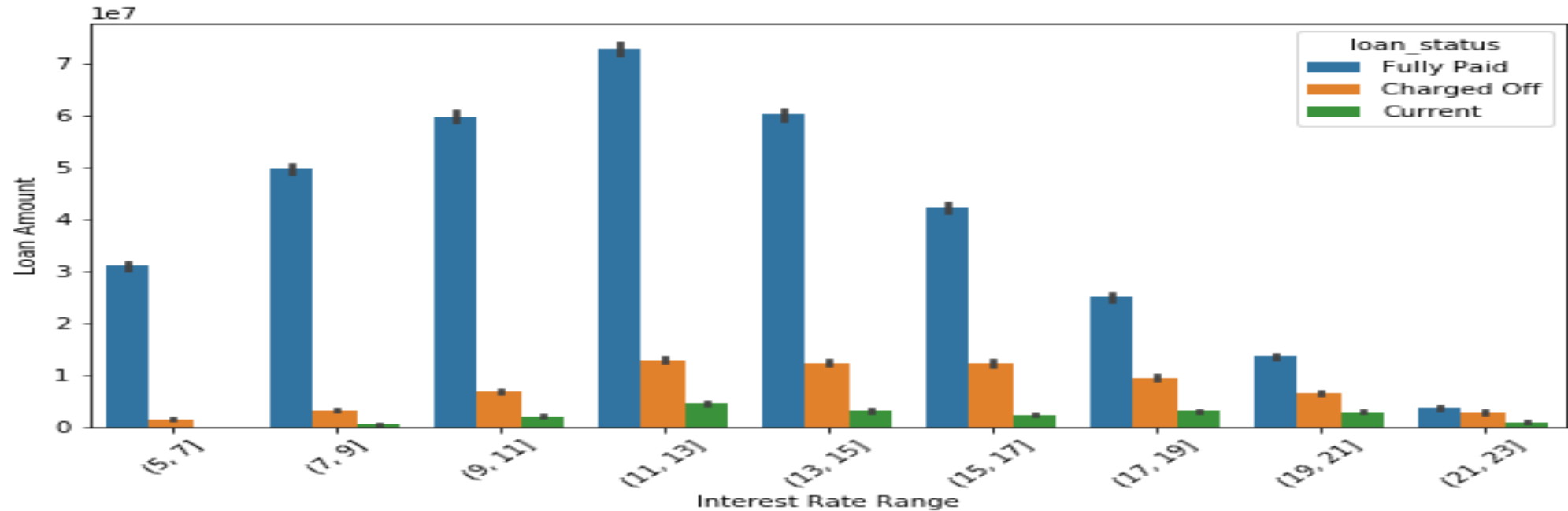
Most of the defaulters lie between the income range of 40K to 60K.

Top three defaulter income categories:

1. 20k-40k
2. 40k-60k
3. 60k-80k

Interest rate in which the highest number of defaulters lie

Plot 3



The above figure shows that most of the defaulters are lying in the interest range of 11 to 21.

Loan amount issued to the sub grade and employment length

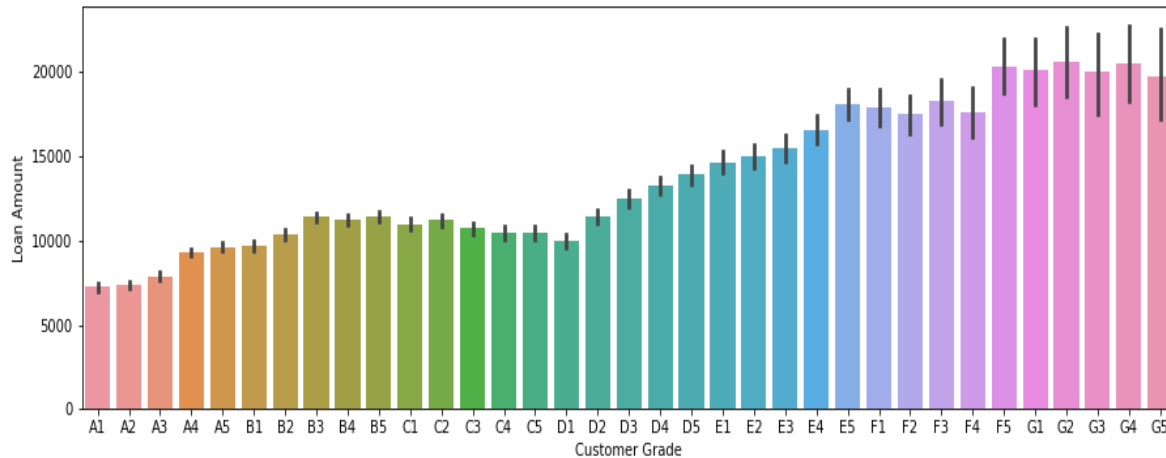


Figure1 .

- 1) Display how the loan amount distributed between differnt different sub grades.
- 2) Most of the about is issued to sub grade G1,G2 .. and so on.

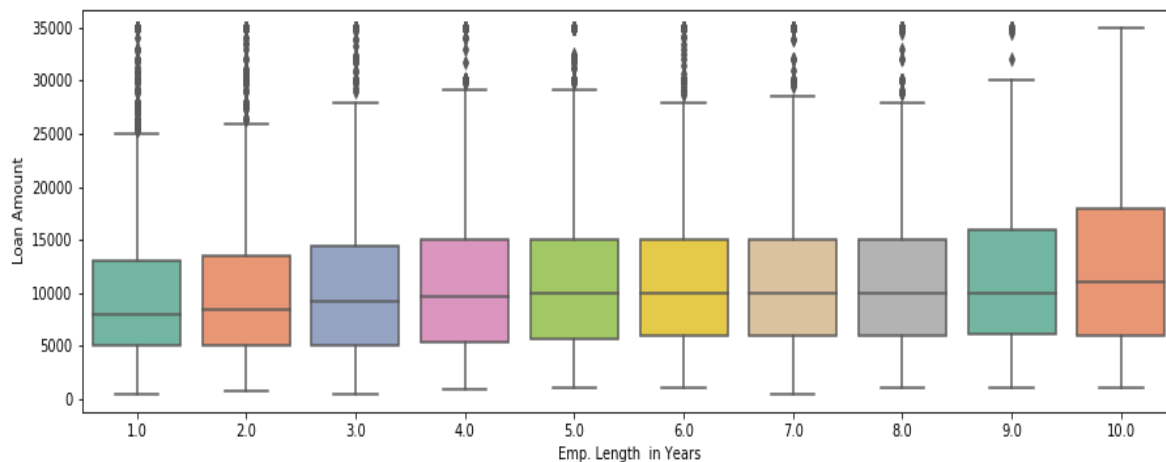


Figure 2.

- 3) Shows that the 10 & 10+ years employment length received the maximum loan amount.
- 4) Loan amount is almost equally distributed between 25 and 75 percentile. Except the 10 years employment length



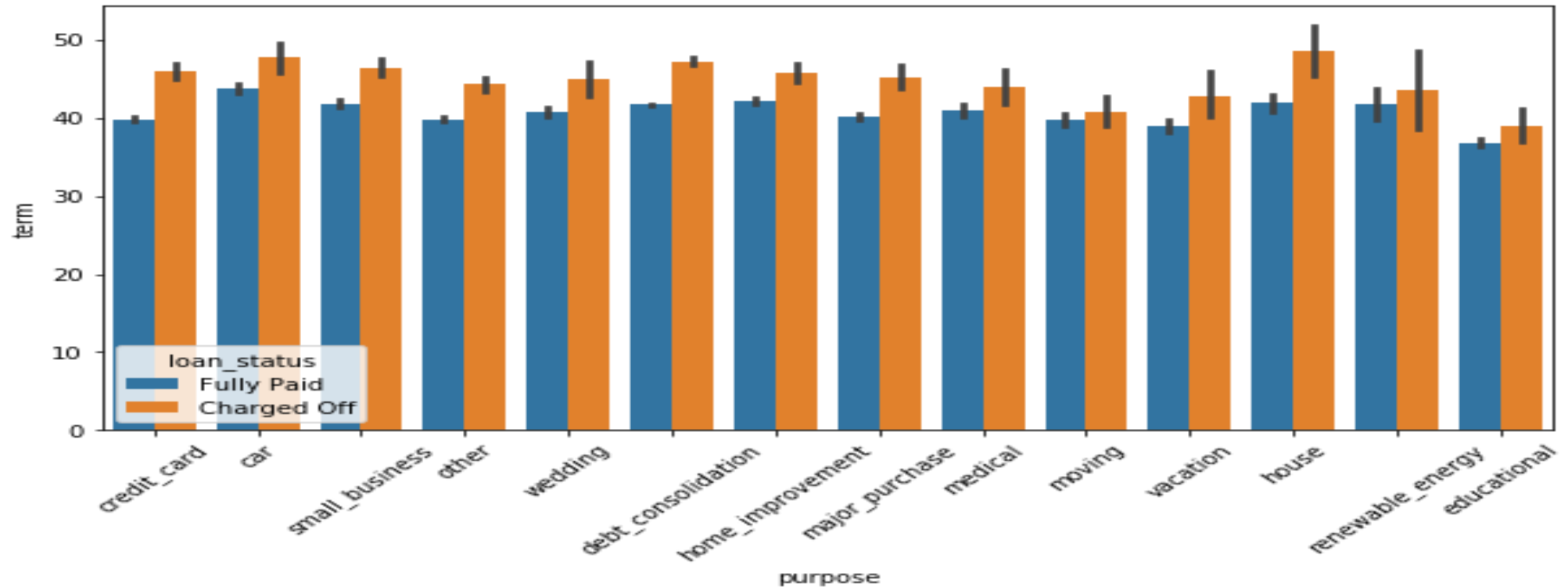
Top 9 US states in which most of the defaulters lie



Addr State	CA	FL	GA	AZ	CO	CT	AL	HI	AR
Charge d Off	7.89%	3.53%	1.51%	0.86%	0.69%	0.66%	0.38%	0.20%	0.19%
Fully Paid	40.82%	15.96%	8.02%	5.09%	4.68%	4.43%	2.67%	0.97%	1.46%

Most of the defaulters are based from the CA and FL states.

Plot between loan term and purpose



All the defaulters had applied for long term loan period.

1. The yellow bar denotes the Charged off users loan term
2. The blue bar refers to the Fully Paid user loan term.

From the analysis, it is found that the following univariate and bivariate variable can play a vital role to identify the defaulters. Variables are as follow :

1) Verification Status

- As seen in the graph ~42% people were not verified. The borrower should come from a verified source.

2) Employment Length

- The employment length is higher for the defaulters.

3) Interest Rate

- Interest is higher for the defaulters.

4) DTI

- DTI is higher for the defaulters

5) Annual Income

- Annual income is lower for the defaulters

6) Term

- Most of them has applied for a higher term i.e 60 months

7) Grade

- Sub grade G has the highest number of defaulters.

8) Address State

- Most of the defaulter are living in California, Florida