LENDING CLUB GROUP CASE STUDY

Submitted by:
Vasanthi Gavini
Manisha Shedge

Problem statement

- When a company gets a loan request, it must decide whether or not to approve the loan based on the applicant's profile. The bank's decision is related with two sorts of risks:
- If the candidate is likely to pay back the loan, the company will lose business if the loan is not approved.
- Approving the loan may result in a financial loss for the company if the applicant is unlikely to pay back the loan, i.e. is likely to default.

About Data

- The provided dataset is the “loan CSV”
- While talking about the data the total length of data :
- No of rows:39717
- No of columns: 111
- Also the data dictionary is provided
- Which has variables and its description

About 115 rows and 2 columns.

Problem understanding

- The business goal is to identify high-risk loan applicants who are likely to default, and to reduce loan approval in order to reduce credit loss. These driving elements can be used by the organization to assess its business risk.

Data understanding

- There is many more columns which are not included or not needed in the analysis so we removed them.
- And that columns are(pymnt_plan, initial_list_status,collections_12_mths_ex_med,policy_code, acc_now_delinq, application_type,pub_rec_bankruptcies, tax_liens, delinq_amnt)
- some of the columns which is realated to the post approval loan.
- some of the column like id,member_id,url,title,emp_title,zip_code,last_credit_pull_d ,addr_state which is not relevant to loan defaulting so we are removed them.
- desc colmun which is description of loan description provided by borrower is not needed so we removed them

- We are not dealing with the current laon so we exclude that column too.

Data analysis

- After the understanding the data we removed the NAN values,missing values.

(emp_length ,revol_util)

- After that checking for the data types

Analysis

- Univariate Analysis:

Analyze the impact of each categorical and continuous variable on the Default count in relation to the overall loan status. Choose the influential variable that has a positive or negative relationship with the Default status. Choose a variable to infer loan characteristics and aid in the identification of defaults.

Bivariate & Segmented Analysis

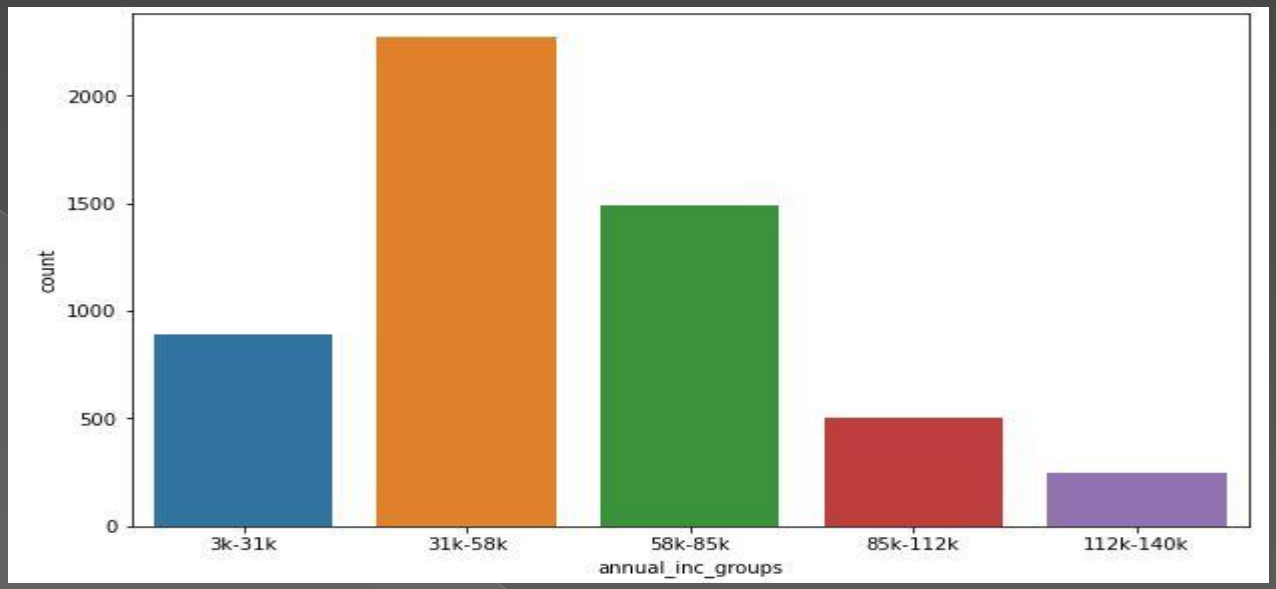
- Impact Univariate Variables are mixed into two feature sets and the Default count is examined.2. Segmented analysis using the Binning Discrete Categorical Variable and the real Categorical Variable to see the relationship and pattern towards Loan Default.3. Analyzed the data spread by comparing the actual continuous variable to the actual categorical variable.4. Variable combinations are identified using analytical reasoning and commercial relationships.

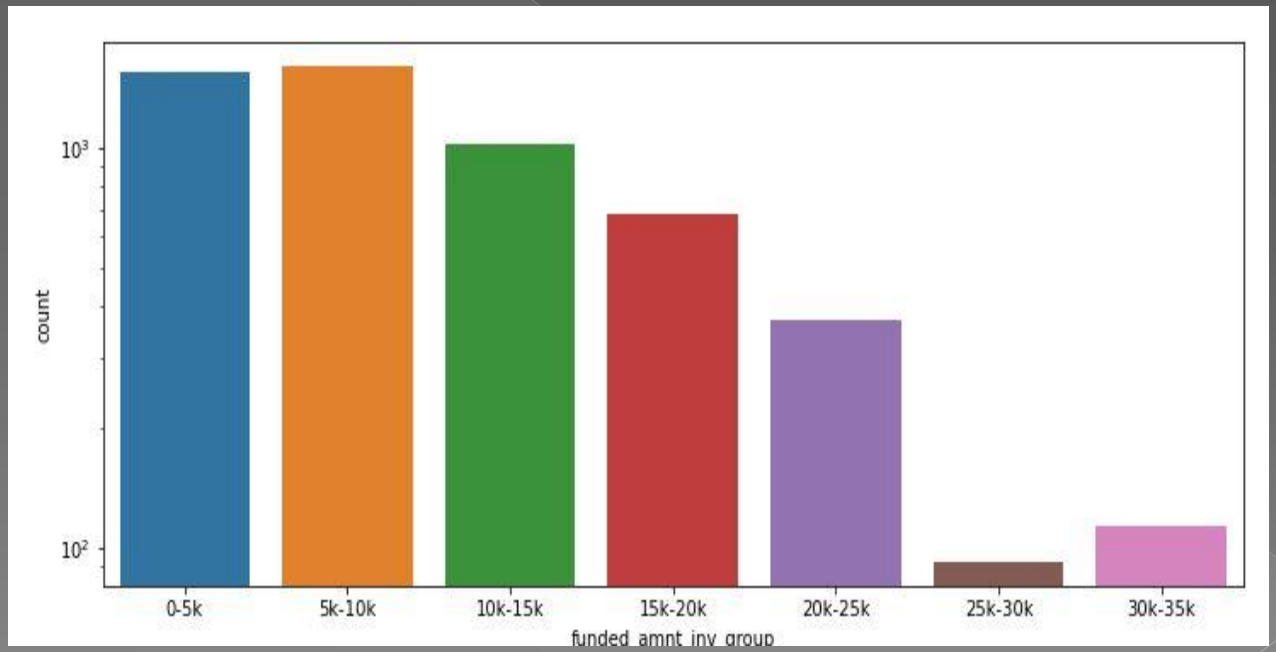
Result

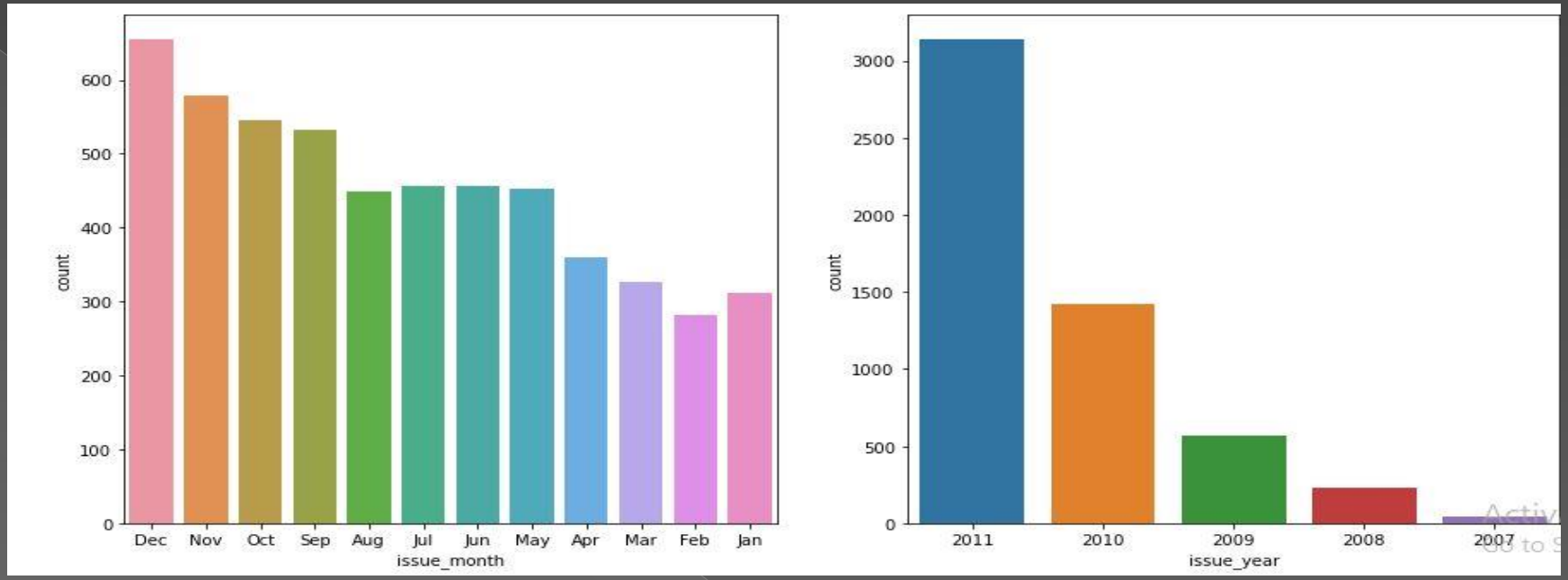
The results of the preceding study for each variable in terms of charged-off loans imply the following. When there is a higher chance of defaulting

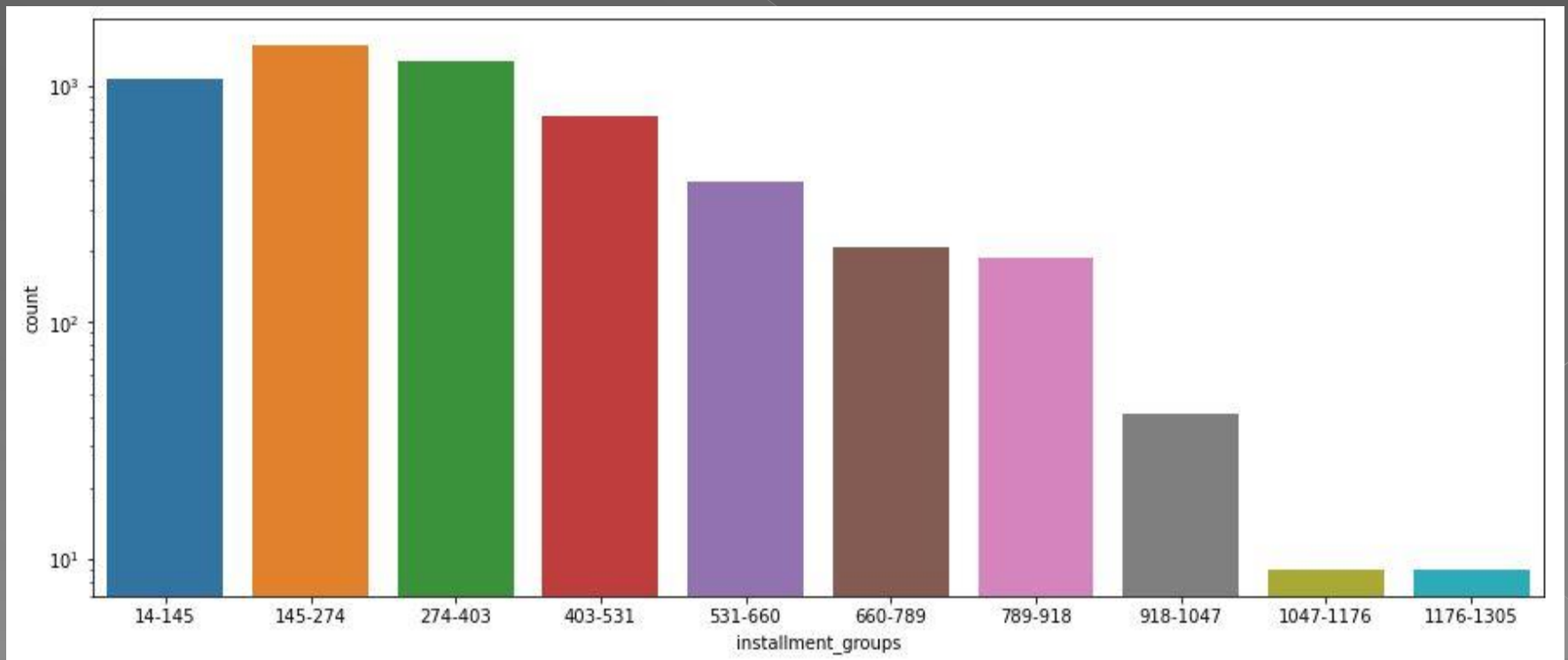
- Applicants whose home ownership status is 'RENT'
- Those that utilise the loan to pay off other debts
- Applicants who obtain a 13-17 percent interest rate
- Applicants with incomes between 31201 and 58402
- Applicants with 20-37 open accounts
- Applicants having a ten-year employment history
- When the amount invested by the investor is between 5,000-10,000

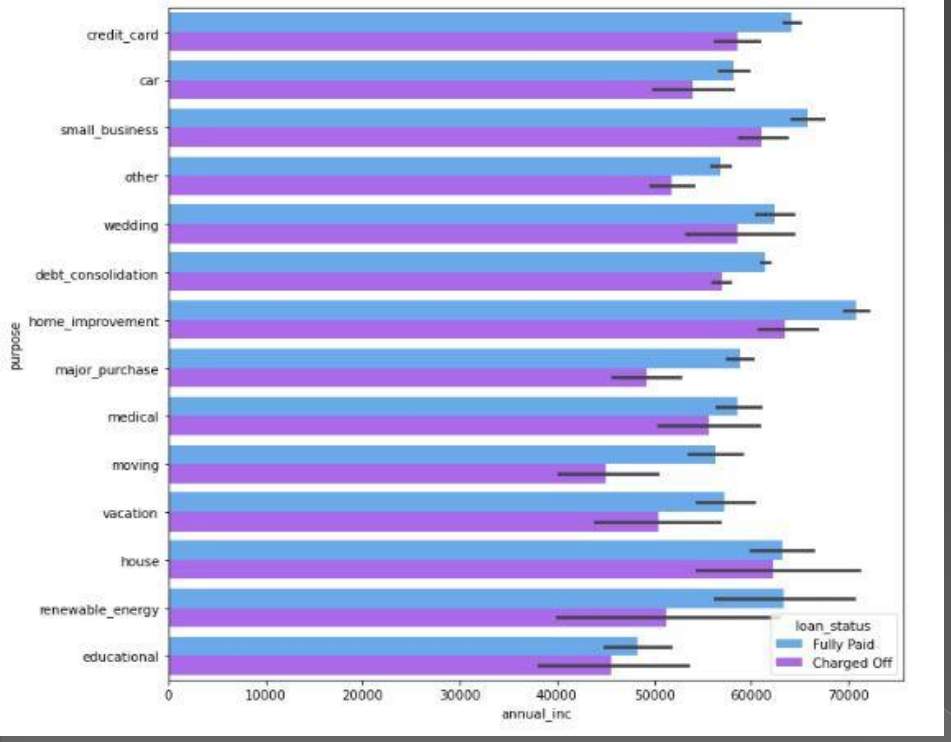
- The loan amount ranges from 5429 to 10357.
- Dti ranges from 12 to 18.
- When the monthly instalments range from 145-274
- 36-month contract
- When the loan status is unknown
- When there have been no inquiries in the last six months and there are no unfavourable public documents
- When the goal is to consolidate debt
- The grade is 'B,' with a total grade of 'B5'.

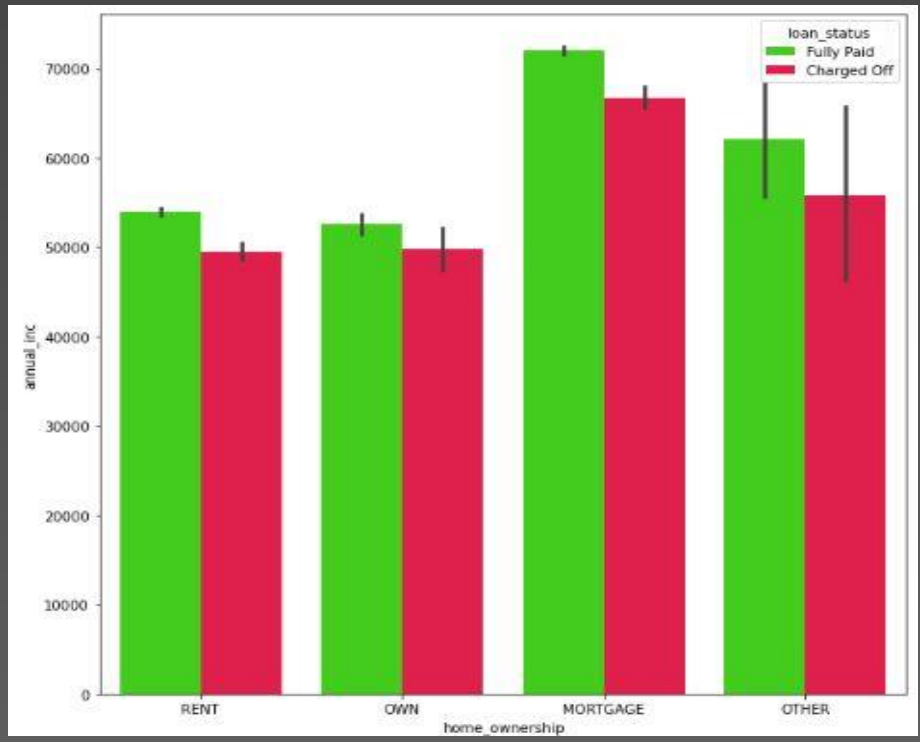

2000
1500
1000
500
0
count
3k-31k
31k-58k
58k-85k
85k-112k
112k-140k
annual_inc_groups

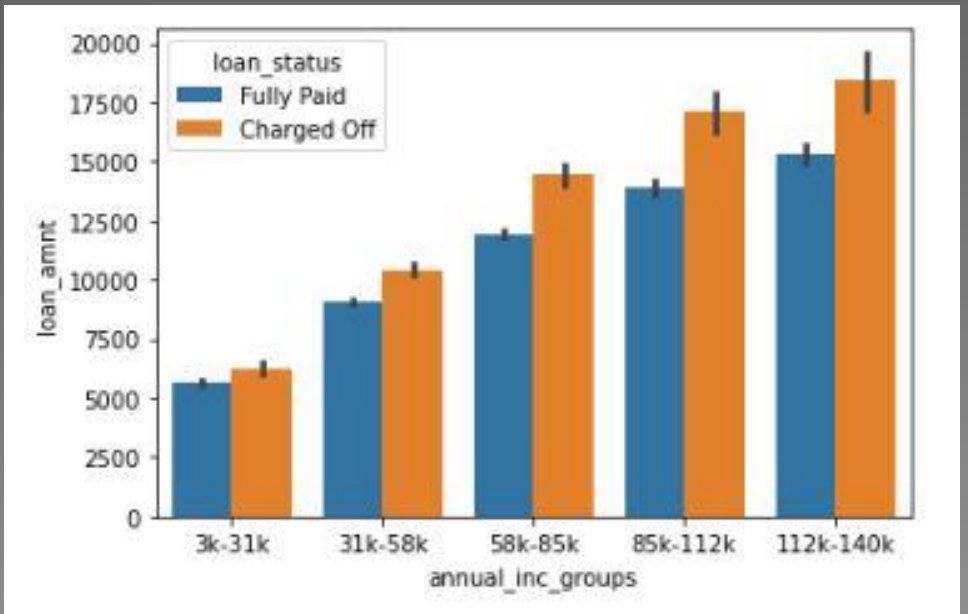

10^3
10^2
count
0-5k
5k-10k
10k-15k
15k-20k
20k-25k
25k-30k
30k-35k
funded_amnt_inv_group

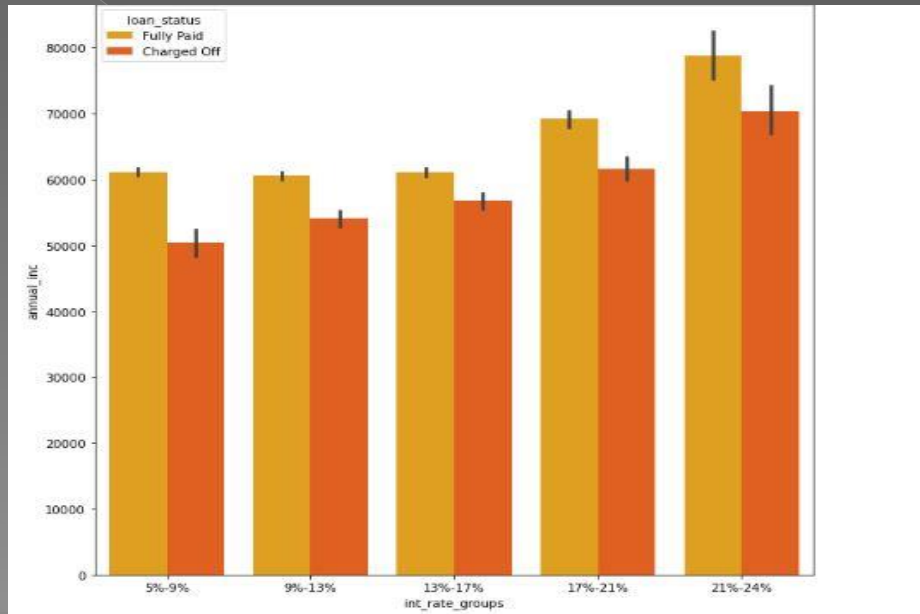

count
600
500
400
300
200
100
0
Dec
Nov
Oct
Sep
Aug
Jul
Jun
May
Apr
Mar
Feb
Jan
issue_month
count
3000
2500
2000
1500
1000
500
0
2011
2010
2009
2008
2007
issue_year

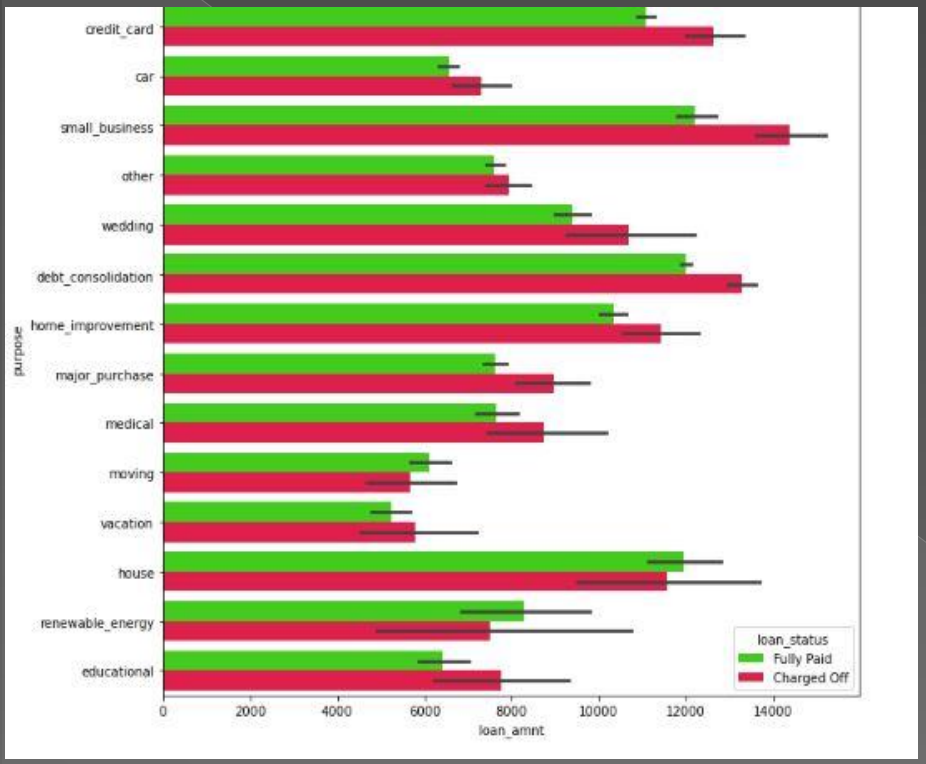

count
10^3
10^2
10^1
14-145
145-274
274-403
403-531
531-660
660-789
789-918
918-1047
1047-1176
1176-1305
installment_groups

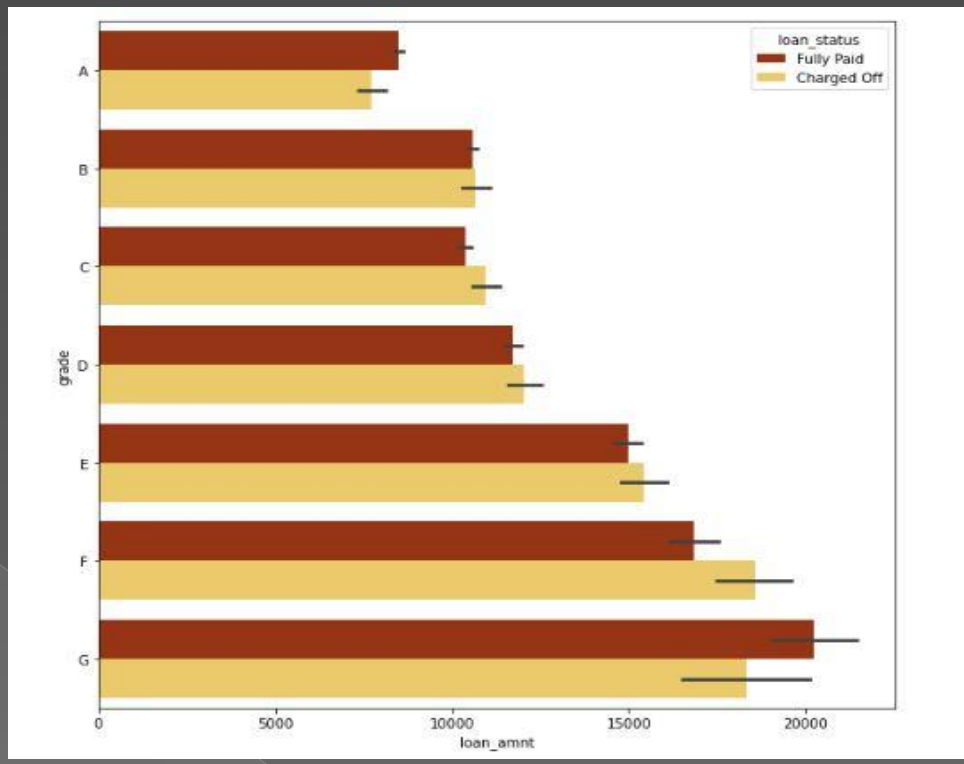

credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose
loan_status
Fully Paid
Charged Off
0
10000
20000
30000
40000
50000
60000
70000
annual_inc

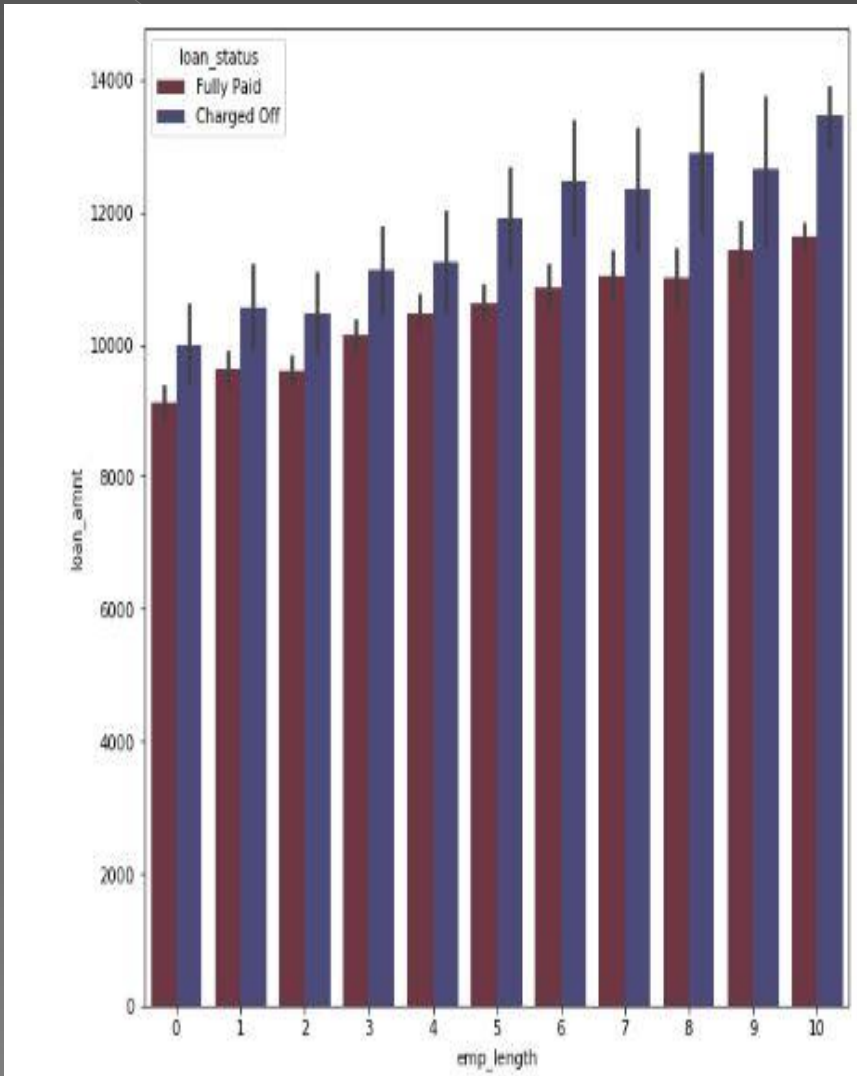

loan_status
Fully Paid
Charged Off
annual_inc
0
10000
20000
30000
40000
50000
60000
70000
RENT
OWN
MORTGAGE
OTHER
home_ownership

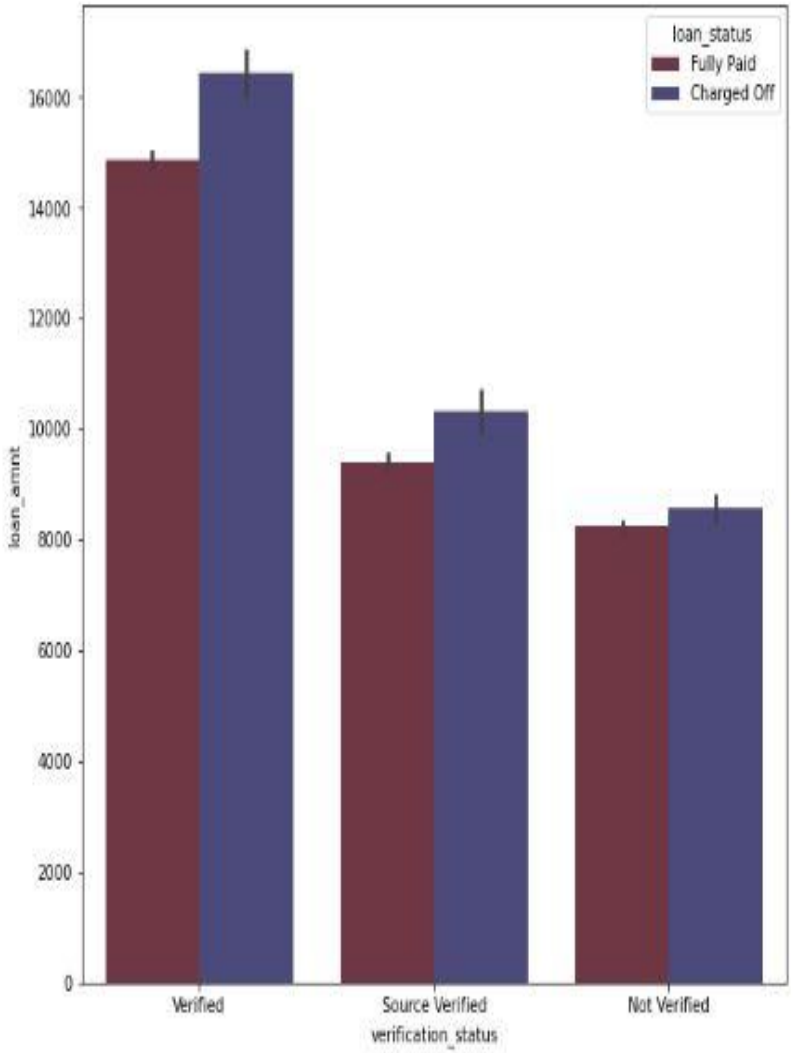

loan_status
Fully Paid
Charged Off
loan_amnt
0
2500
5000
7500
10000
12500
15000
17500
20000
3k-31k
31k-58k
58k-85k
85k-112k
112k-140k
annual_inc_groups

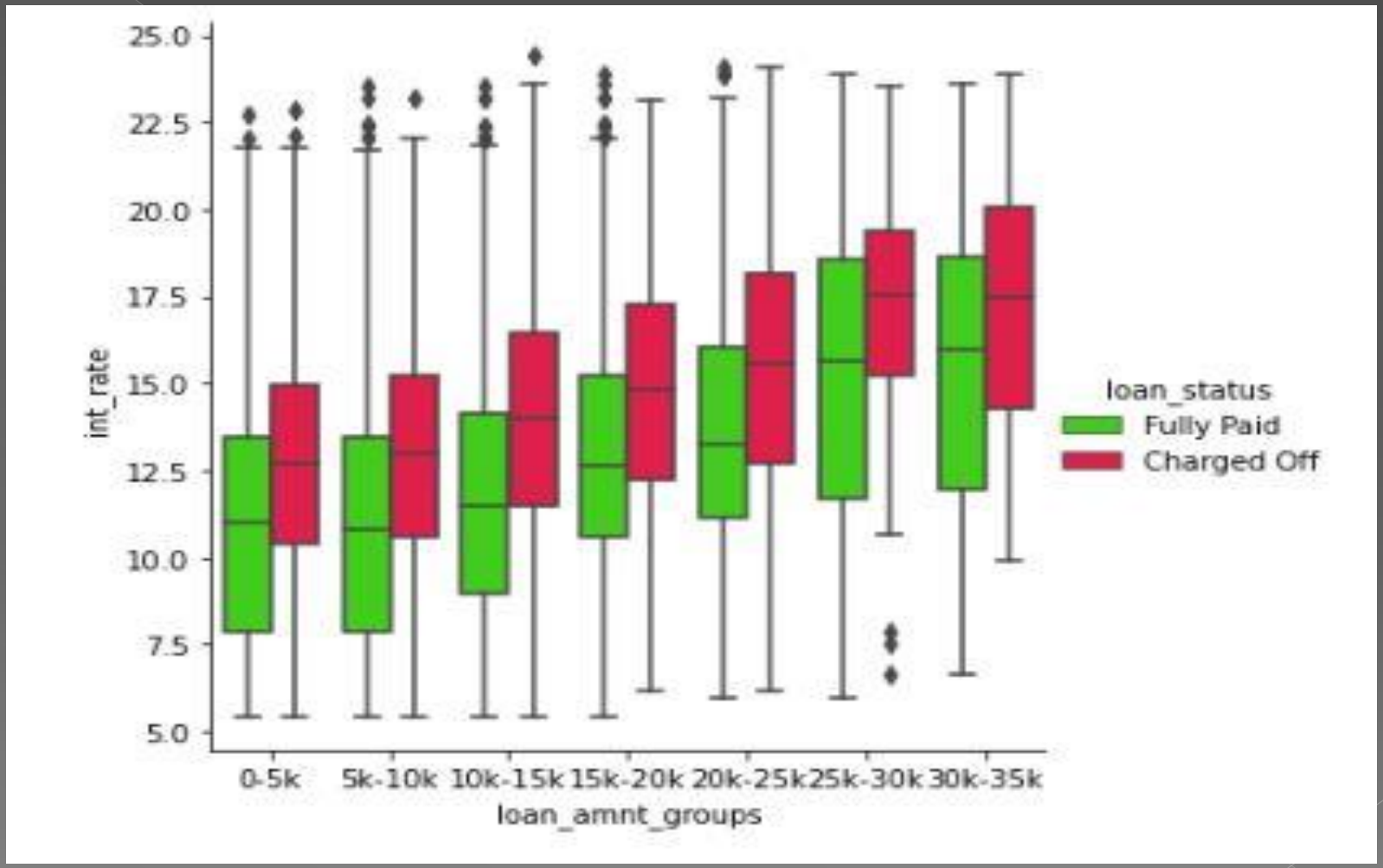

loan_status
Fully Paid
Charged Off
annual_inc
0
10000
20000
30000
40000
50000
60000
70000
80000
5%-9%
9%-13%
13%-17%
17%-21%
21%-24%
int_rate_groups


credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose
0
2000
4000
6000
8000
10000
12000
14000
loan_amnt
loan_status
Fully Paid
Charged Off


A
B
C
D
E
F
G
grade
0
5000
10000
15000
20000
loan_amnt
loan_status
Fully Paid
Charged Off


loan_status
Fully Paid
Charged Off
loan_amnt
0
2000
4000
6000
8000
10000
12000
14000
0
1
2
3
4
5
6
7
8
9
10
emp_length


loan_status
Fully Paid
Charged Off
loan_amnt
0
2000
4000
6000
8000
10000
12000
14000
16000
Verified
Source Verified
Not Verified
verification_status


25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
int_rate
0-5k
5k-10k
10k-15k
15k-20k
20k-25k
25k-30k
30k-35k
loan_amnt_groups
loan_status
Fully Paid
Charged Off

conclusion

- Candidates with a household income of 60,000 to 70,000 who are taking out a loan for "home repair“
- Candidates with a mortgage and a household income of 60-70,000
- Candidates with a 70000-80000 annual income and a rate of interest of 21-24%
- Candidates who have accepted a loan in the range 30000 - 35000 and are charged interest rate of 15-17.5 %
- Candidates who have obtained a small business loan for a sum larger than 14,000.
- Candidates with a 'MORTGAGE' on their property and a 14-16000 loan
- Once the grade is F and the loan balance is between 15,000 and 20,000
- When the period of employment is ten years and the loan amount is twelve to fourteen thousand dollars
- Whenever the loan is verified and the amount is greater than $16,000For grades G and higher than 20% interest rate