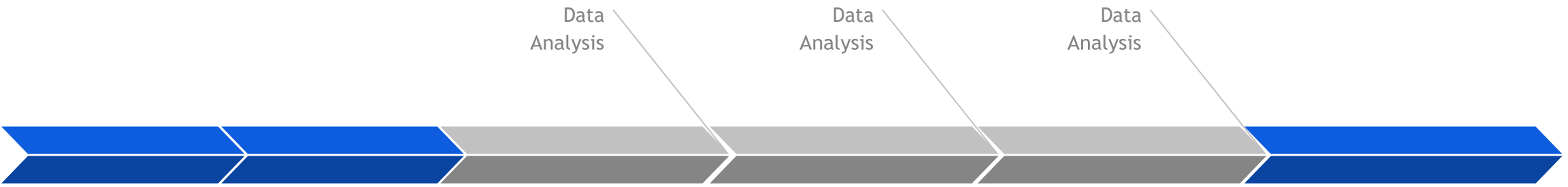
Lending Club Case Study

Group Members:
Sunilkumar Munolli

ABSTRACT

- Lending club is the largest online loan marketplace, facilitating personal loans,business loans, and financing of medical procedures.
- Borrowers can easily access lower interest rate loans through a fast online interface.
- The objective of analysis is to use the information about past loan applicants and findwhether they ‘defaulted’ or not.



Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analyzing each column, plotting the distributions of each column.

SegmentedUnivariate Analysis

Analyzing the continuous data columns with respectto the categorical column

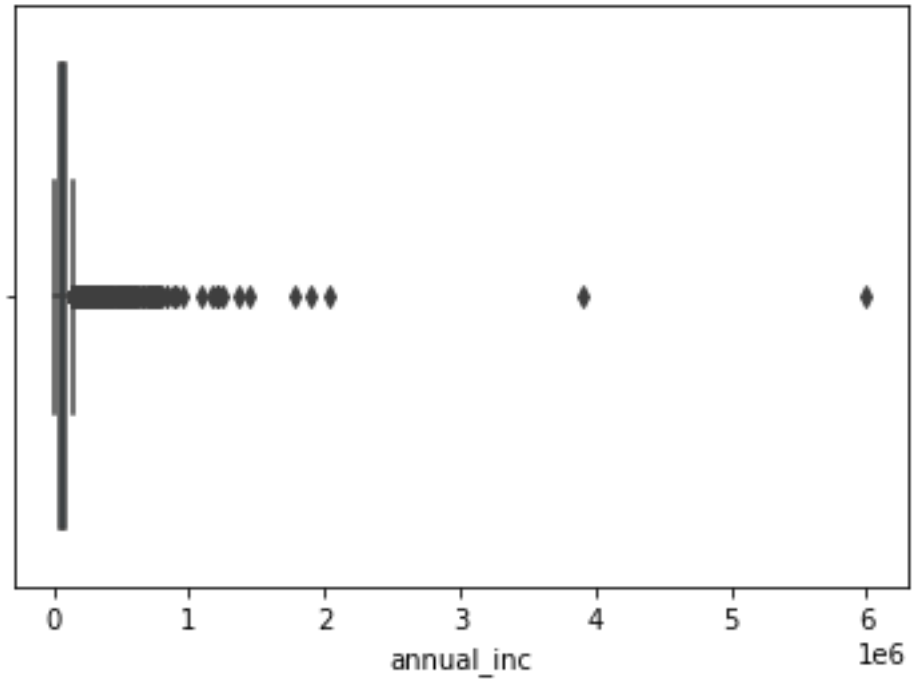
Bivariate Analysis

Analyzing the two variable behavior like term and loan status with respect to loan amount.

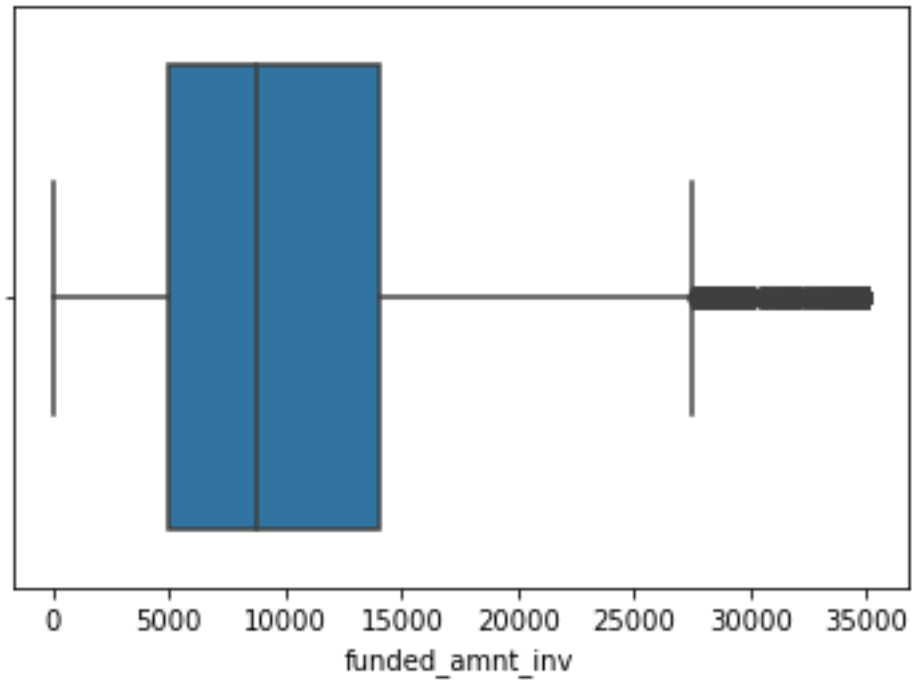
Recommendations

Analyzing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.

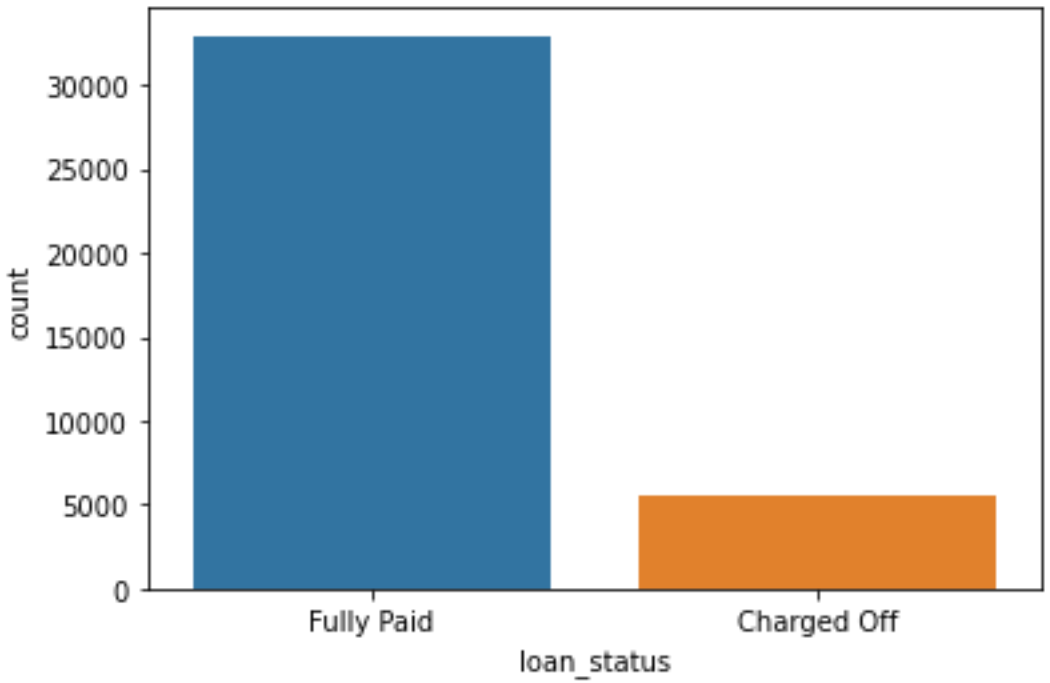
Univariate Analysis



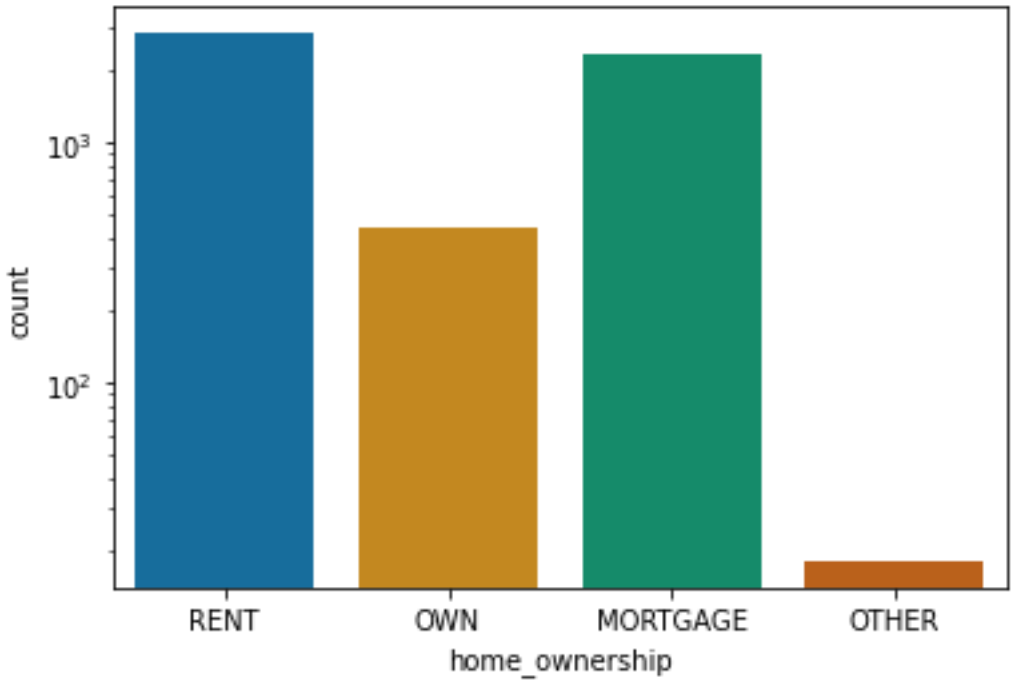
- Removing outliers from annual Income up to 95% to make it easy for plot visualization.



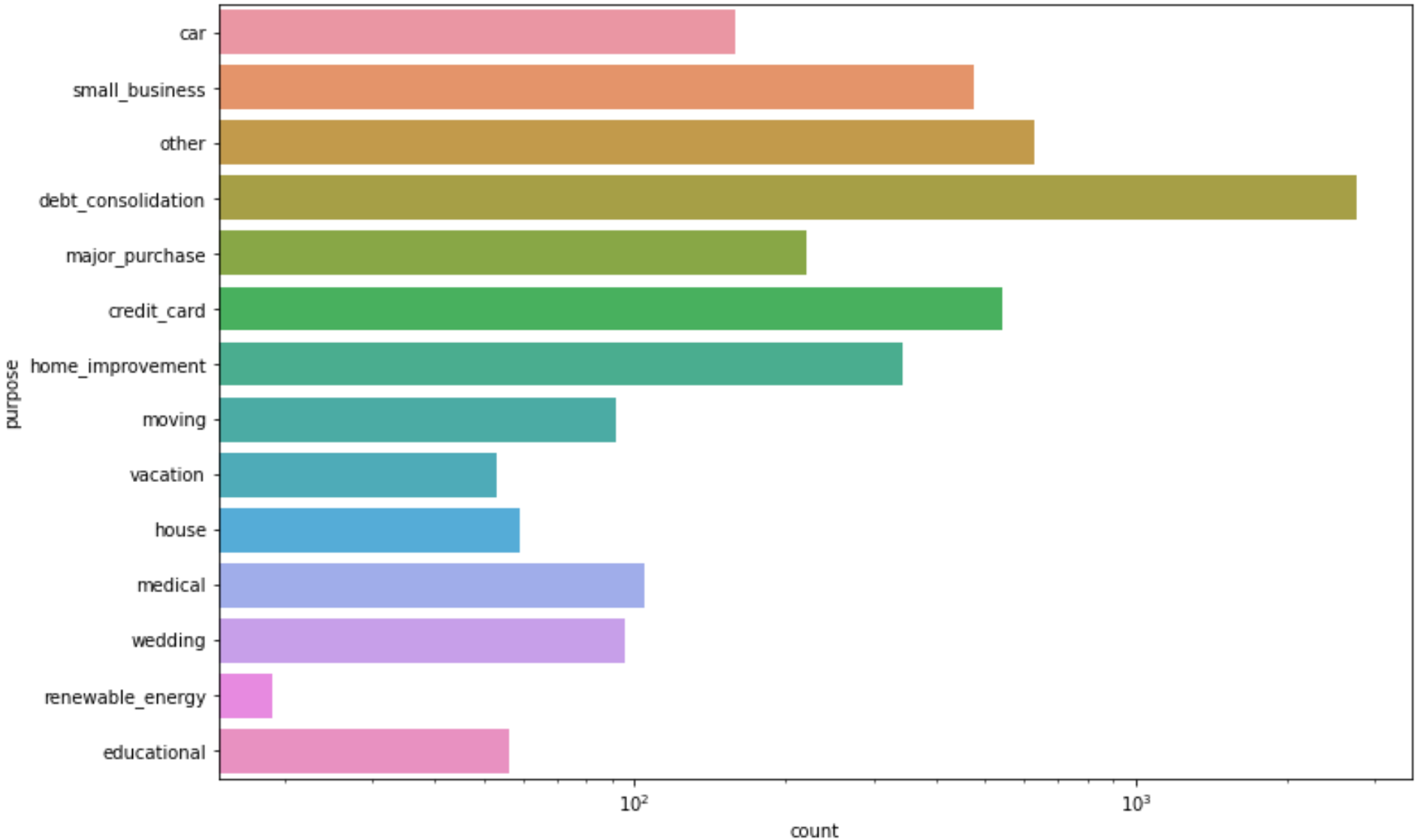
- Checking for outliers for funded amount from the investor for any loan rejection or approval



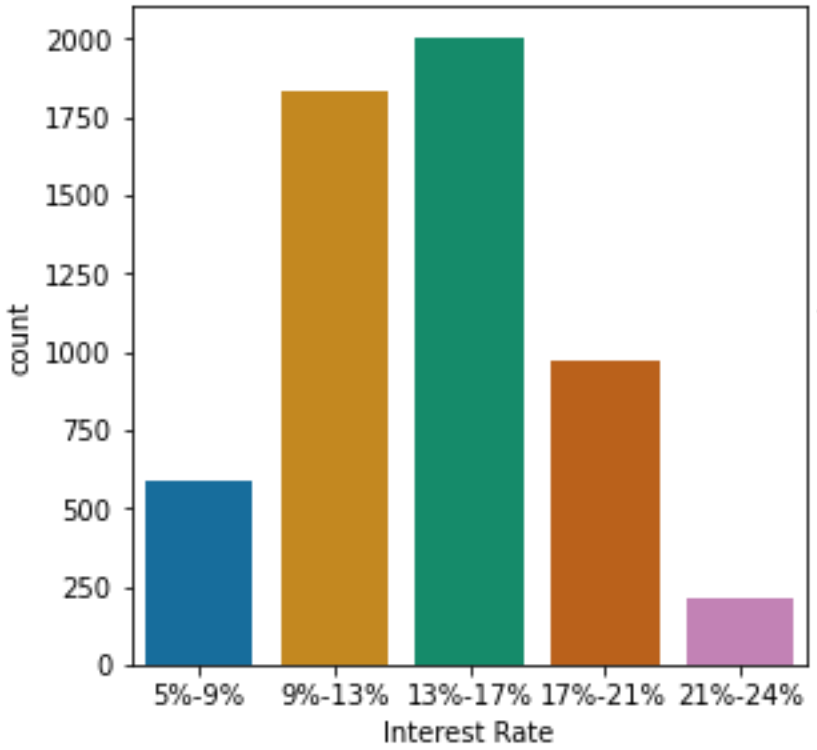
We are Visualizing the Categorical data for the loan status for customers who are fully paid and for the customer who is charged off.

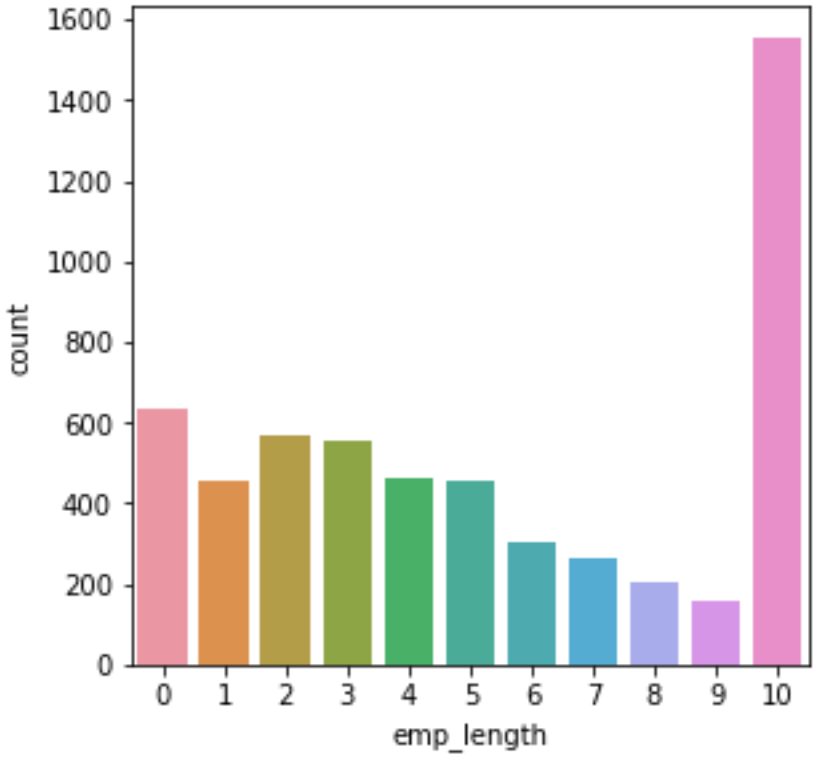


From this plot, we can observe that customers who are staying in rental homes have high chances of being charged off.

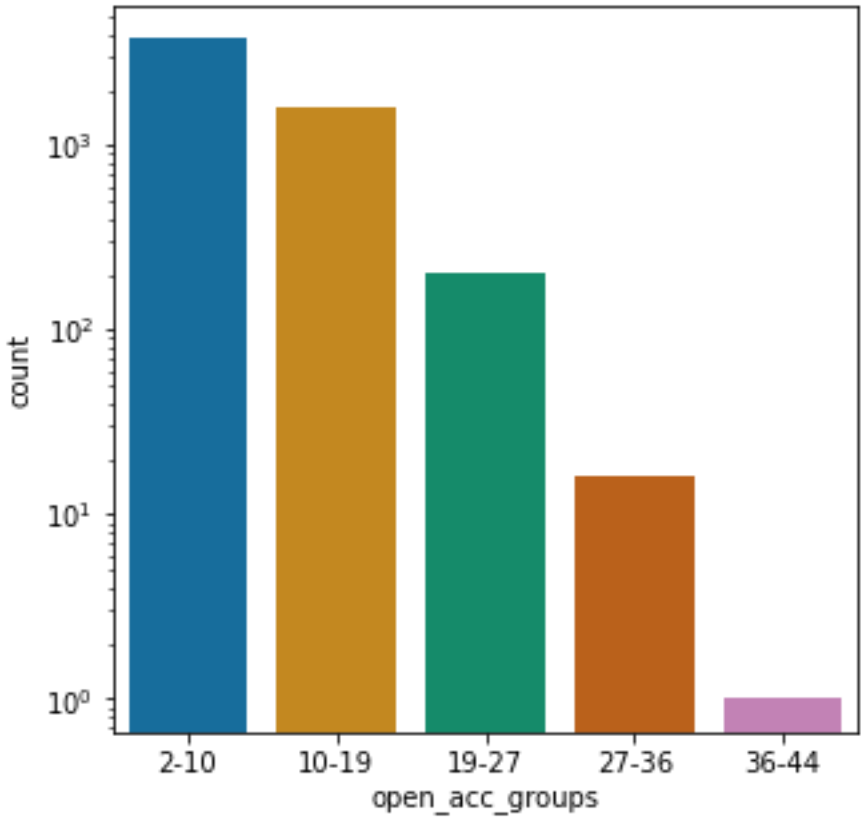

car
small_business
other
debt_consolidation
major_purchase
credit_card
home_improvement
moving
vacation
house
medical
wedding
renewable_energy
educational
purpose
10^2
10^3
count

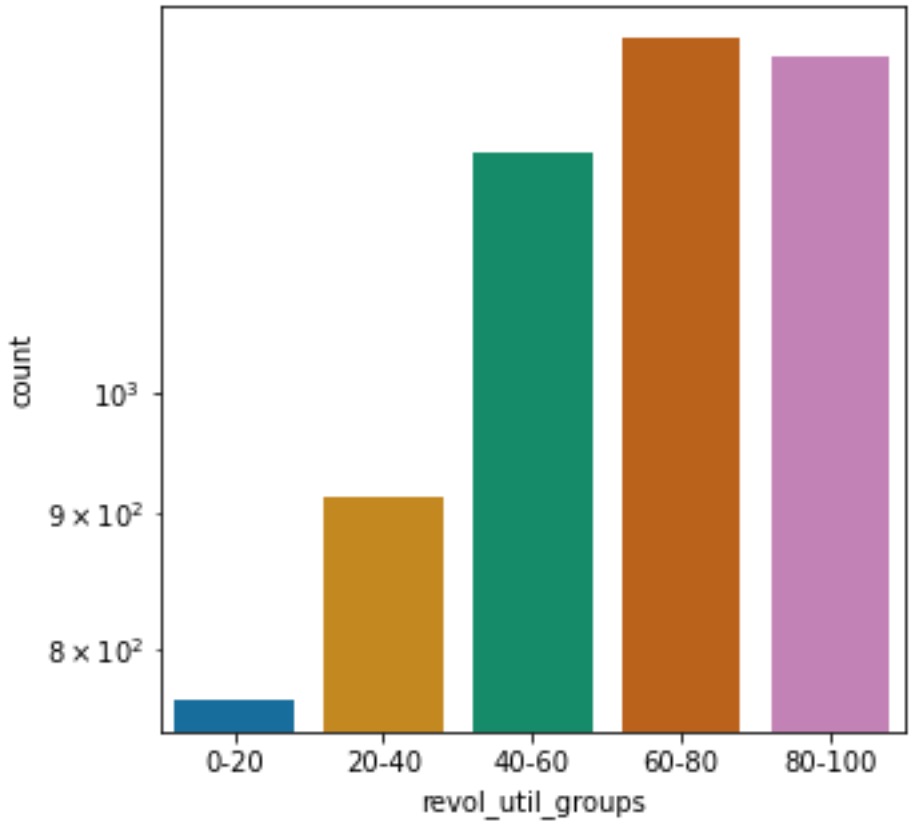
- From the above plot, we can infer the purpose of the loan from the customers is due debt_consolidation which means to repay the existing the loans.

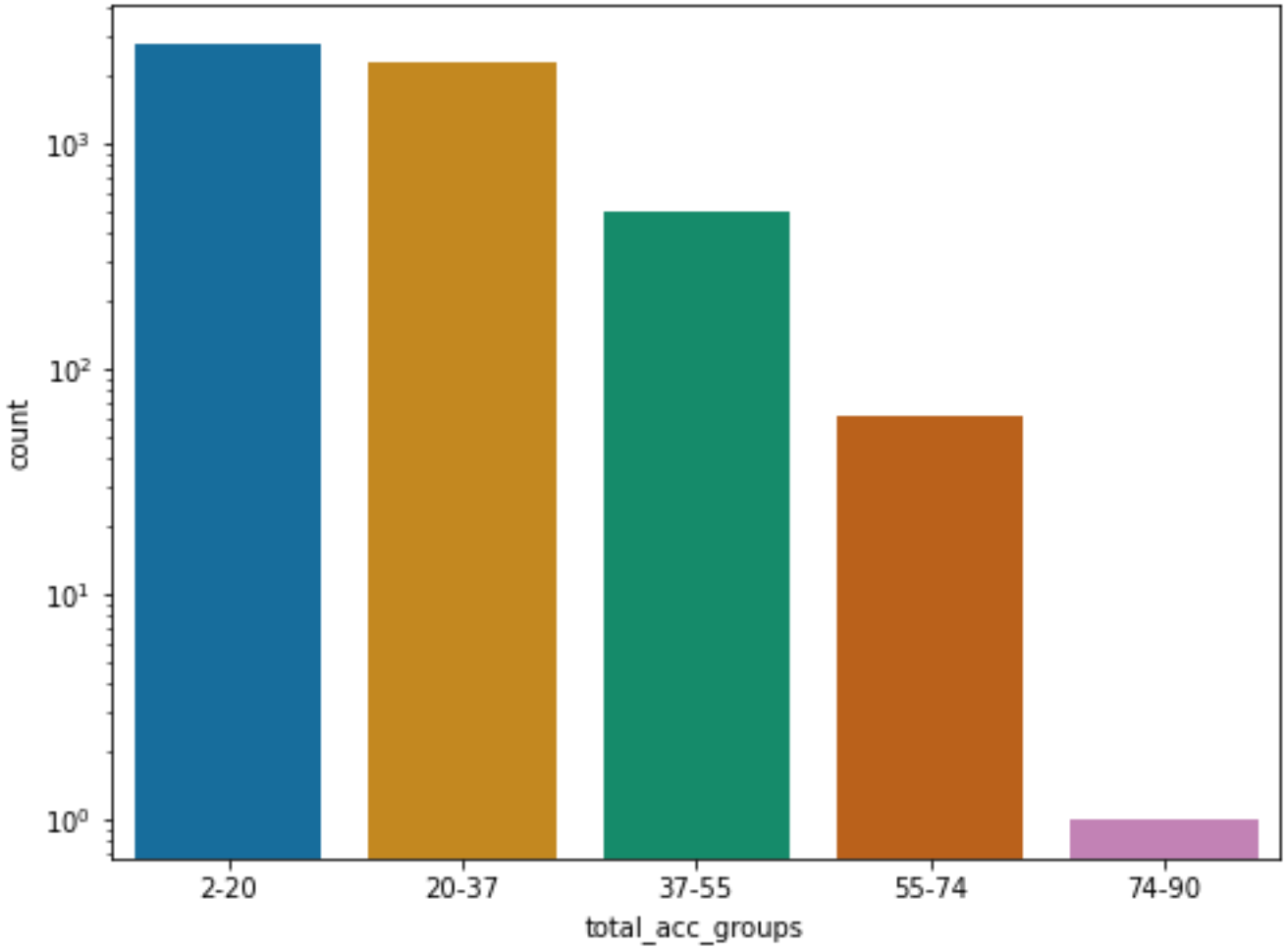


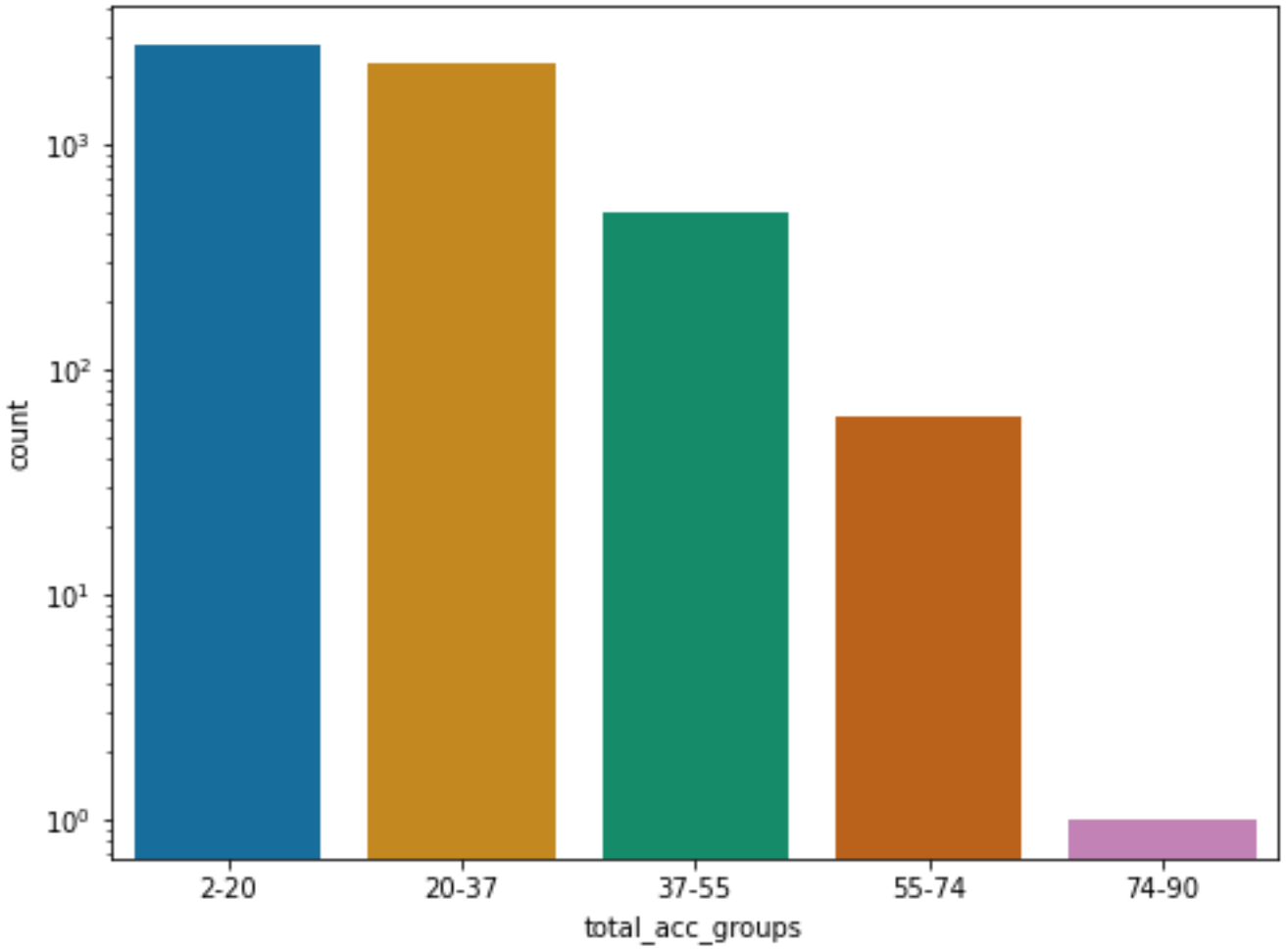


Maximum number of customers who are charged off is having around 10 years of Employment length and when interest rate falls under 13 – 17 %

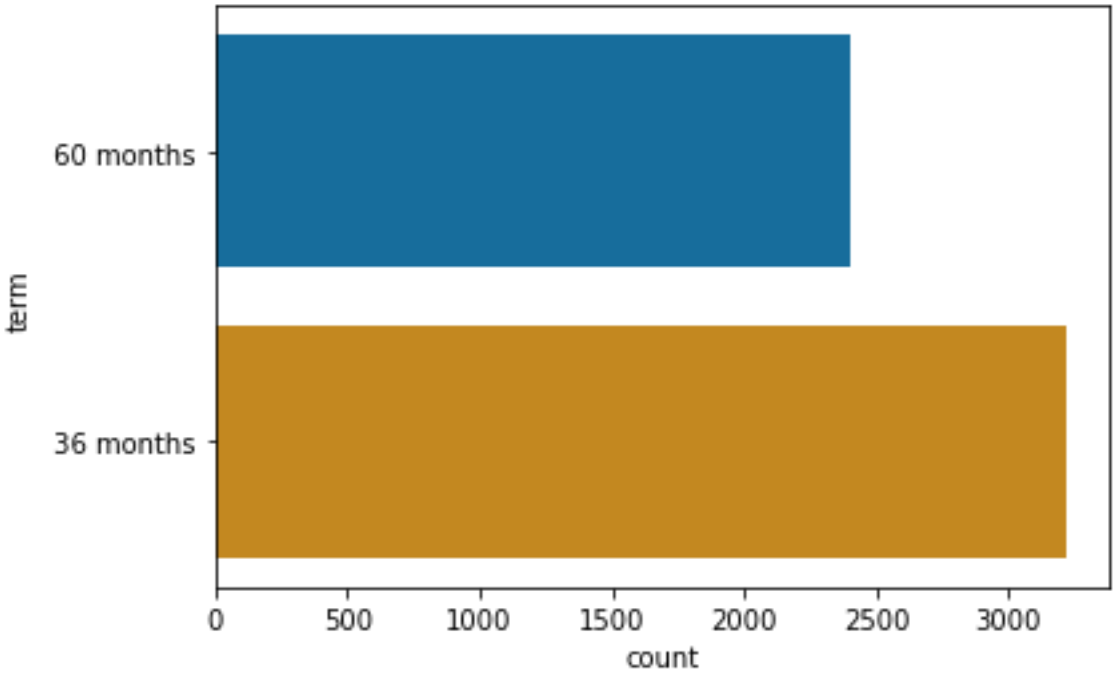

count
10^3
10^2
10^1
10^0
2-10
10-19
19-27
27-36
36-44
open_acc_groups

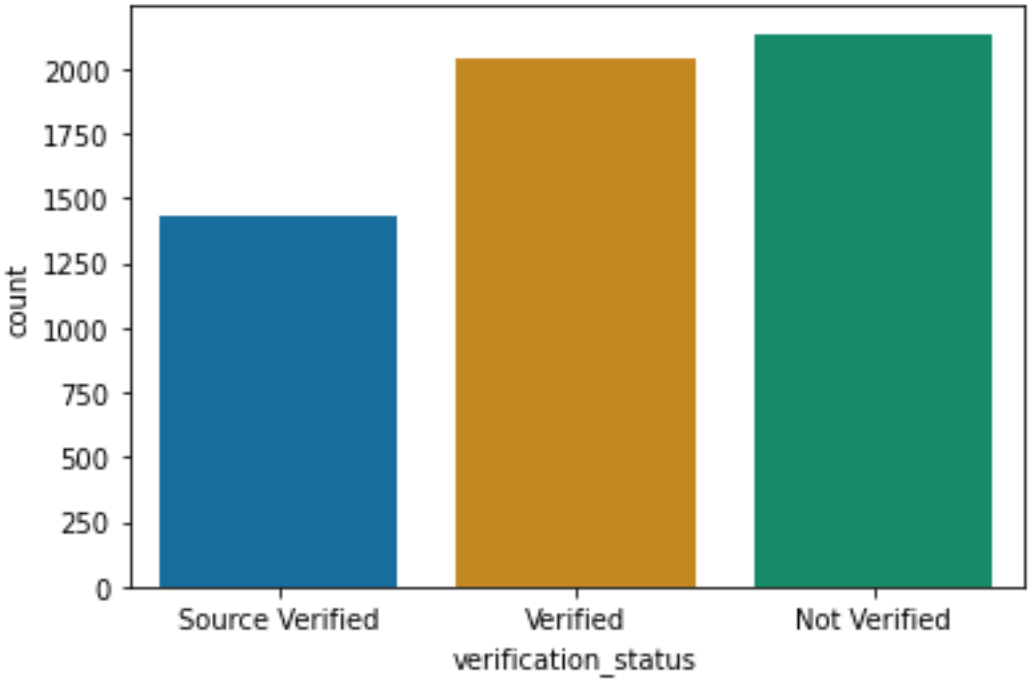

count
10^3
9×10^2
8×10^2
0-20
20-40
40-60
60-80
80-100
revol_util_groups


count
10^3
10^2
10^1
10^0
2-20
20-37
37-55
55-74
74-90
total_acc_groups

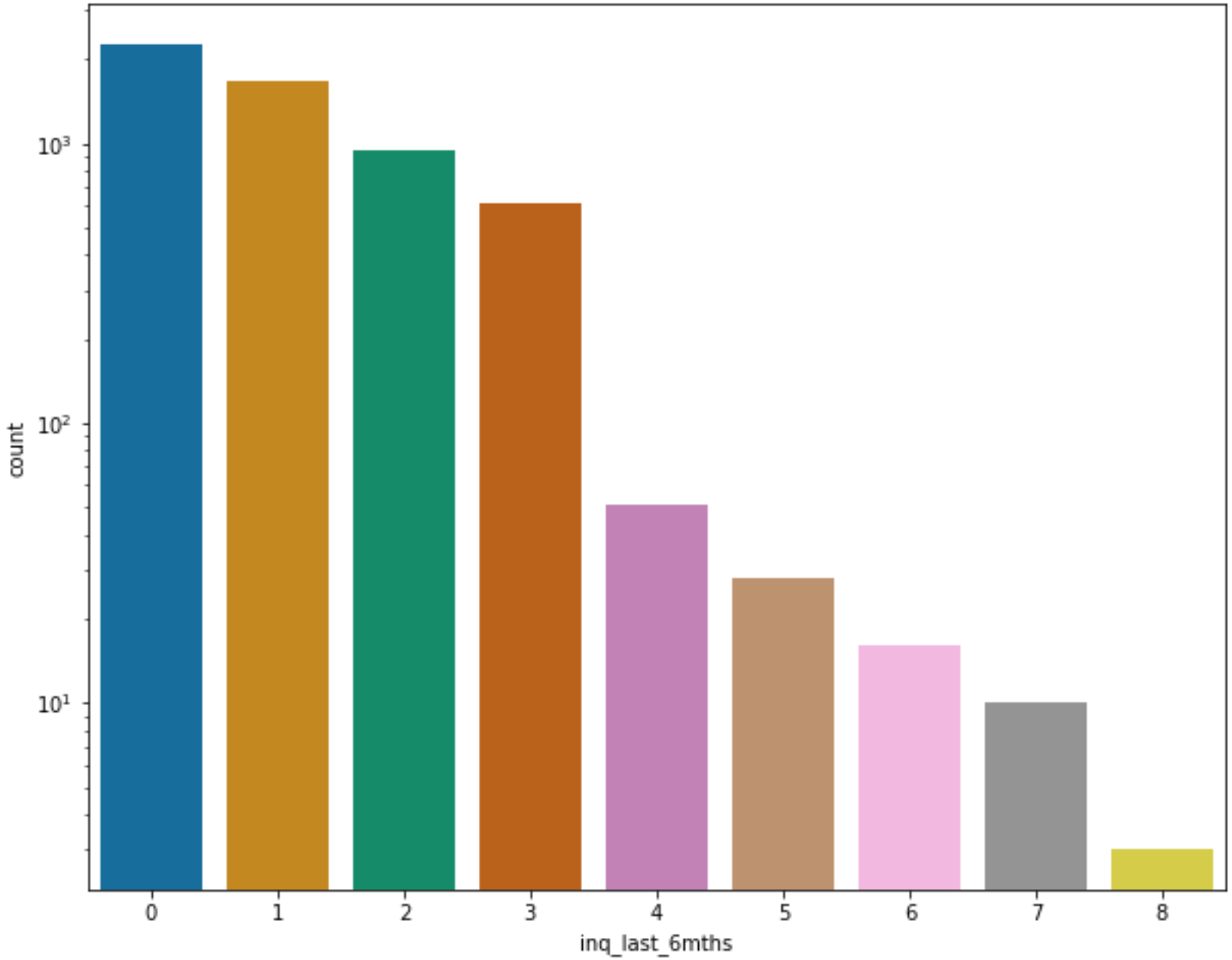


- From the plot1 and plot 3borrowers who is having open_acc and total_acc 2 -10 has been charged off
- From the plot 2 borrowers who is having rev_util 60 – 80 has been charged off
- From the plot 4 borrowers who is having annual_inc between 3 – 31k has been charged off

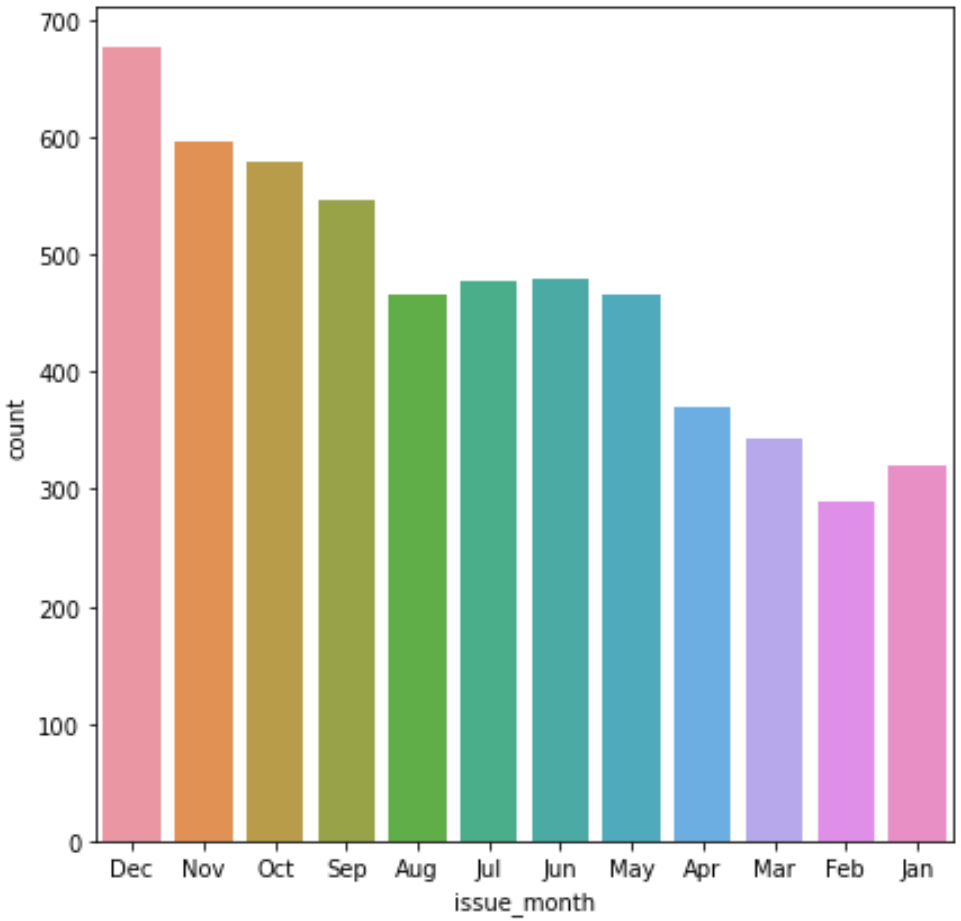


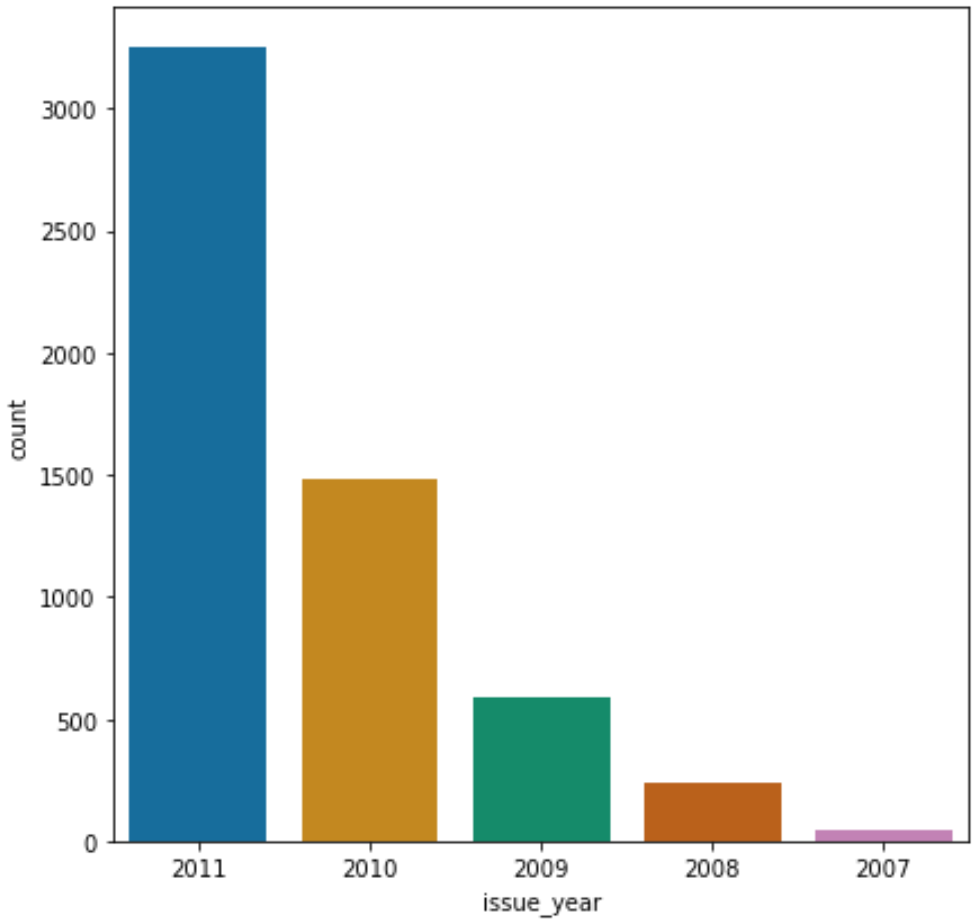


- From the plot1 borrowers who is having loan term for 36 months is having high chances to be charged off
- From the plot 2 borrowers whose source is not verified is having high chances to be charged off


count
10^3
10^2
10^1
0
1
2
3
4
5
6
7
8
inq_last_6mths

- The number of inquiries in past 6 months (excluding auto and mortgage inquiries). The maximum cases of a customer to be defaulter when the number of inquires in the past 6 months is 0.





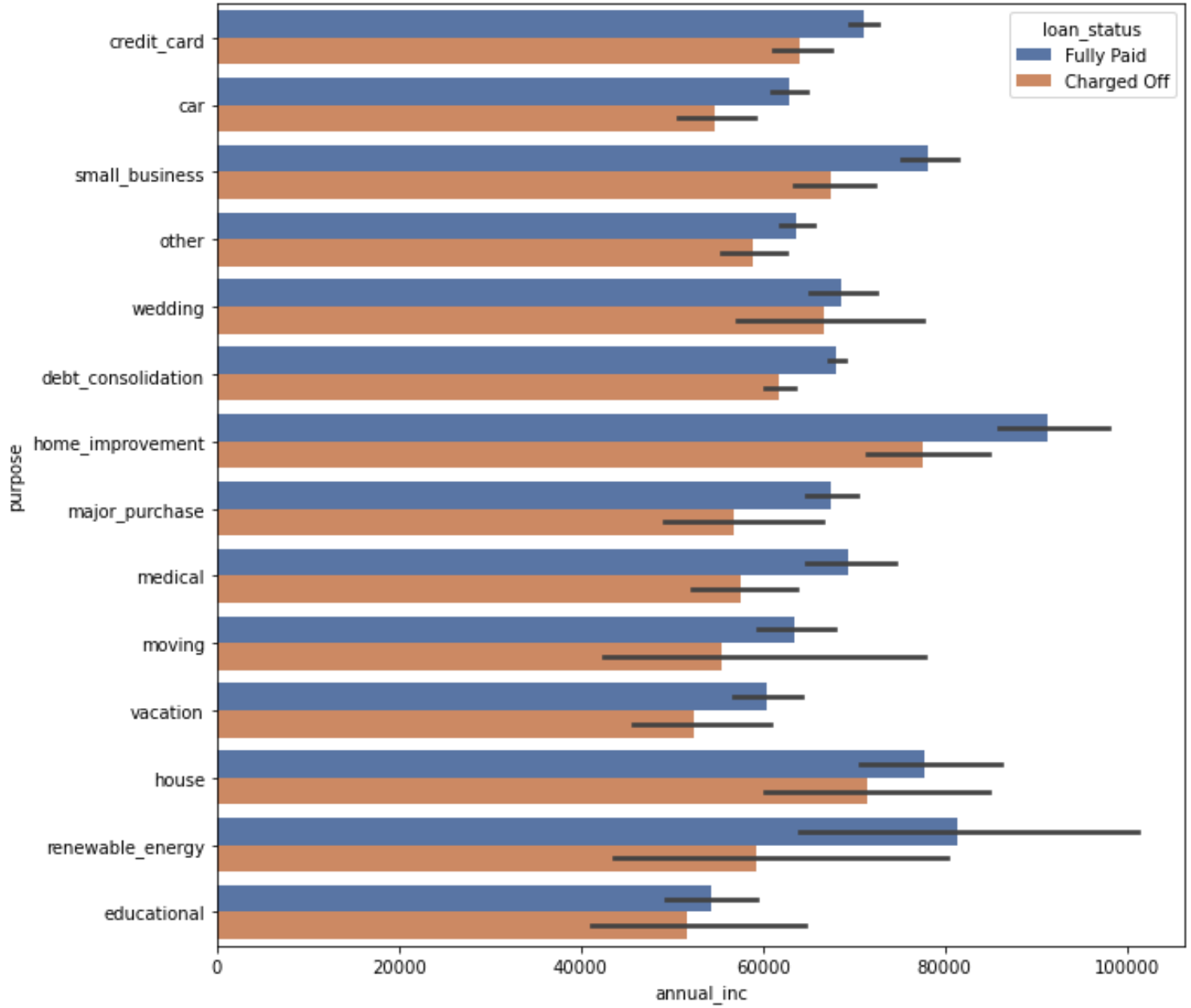
- The maximum number of defaults occurred when the loan was issued in Dec. Loan issued in the year 2011 were also charged off high as compared to other years

Consolidate Univariate Analysis Observation:

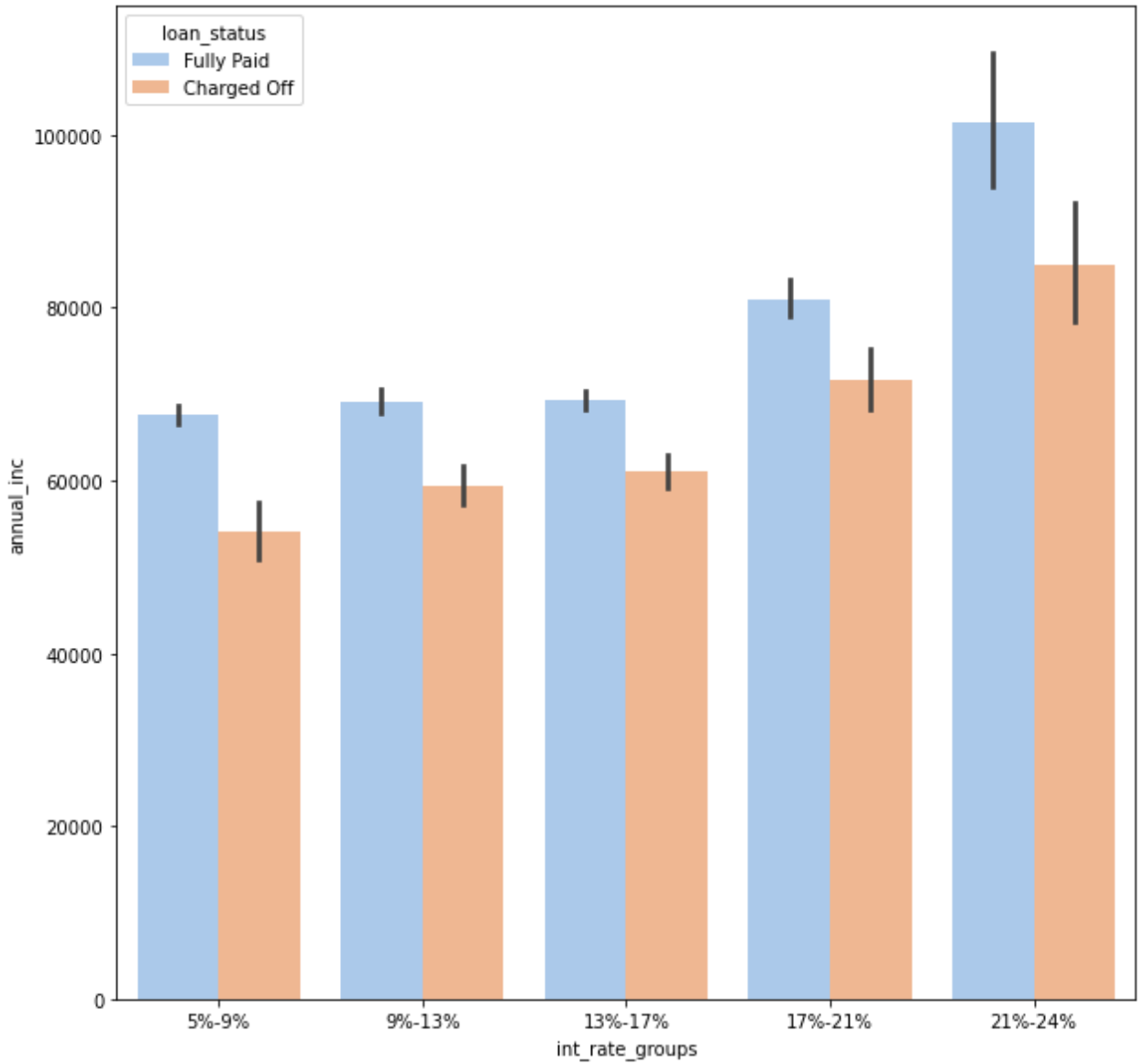
1. The applicant's homeownership is 'RENT'
2. The purpose of the loan is Debt clearance
3. The applicants who receive interest at the rate of 13-17%
4. Applicants who have an income of range 31201 - 58402
5. Applicants with employment length of 10
6. When funded amount by investor is between 5000-10000
7. Loan amount is between 5429 - 10357
8. Data is between 12-18
9. When monthly installments are between 145-274
10. Term of 36 months
11. When the number of derogatory public records is 0
12. When the purpose is 'debt_consolidation'
13. Grade is 'B'
14. And a total grade of 'B5' level.
15. Also there is a very interesting observation from the date issued. The late months of a year indicated the high possibility of defaulting.

The high number of loan defaults in 2011 could be due to the financial crisis in the USA (Assuming the data is of US origin)

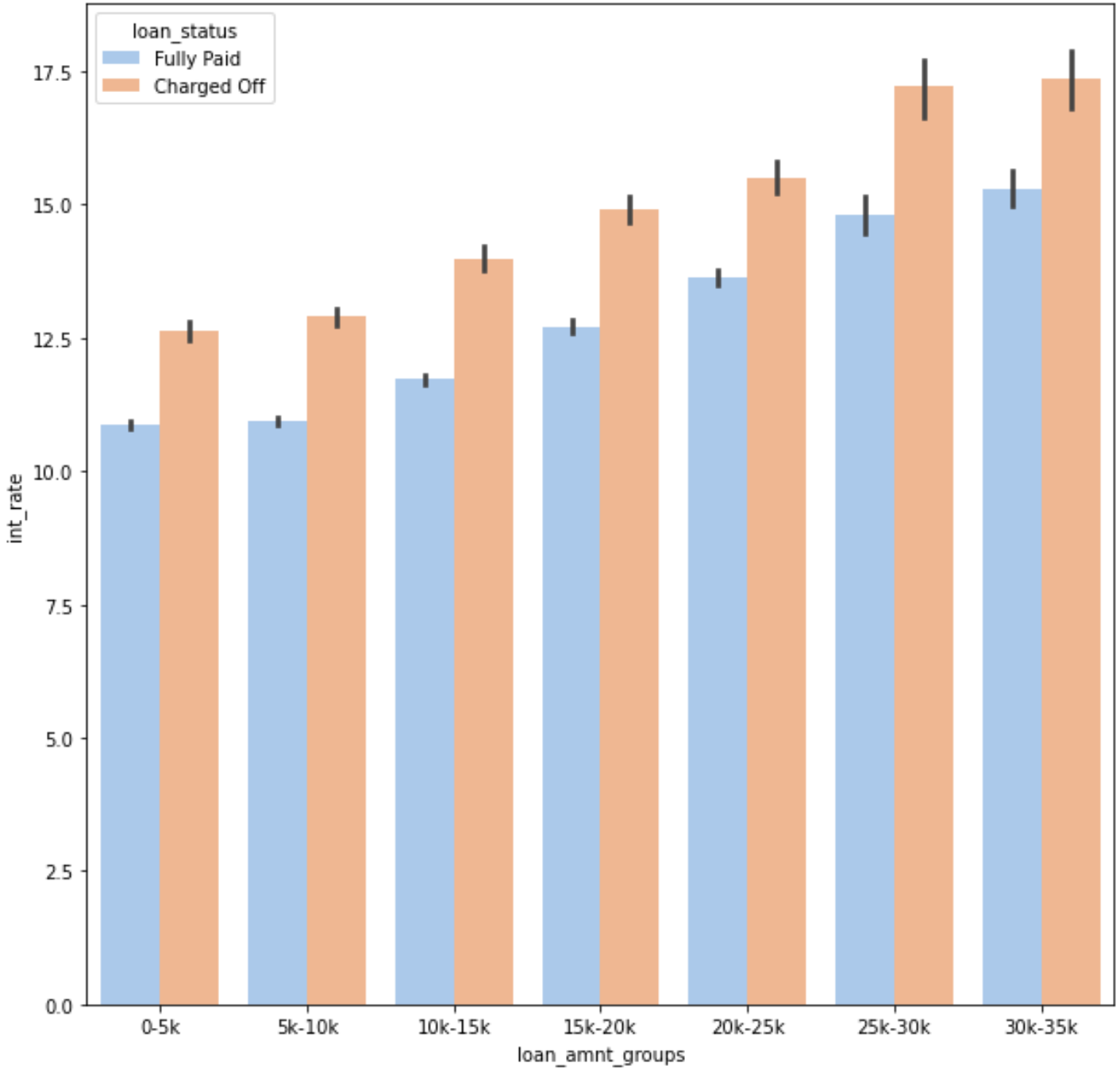
BI-VARIATE ANALYSIS

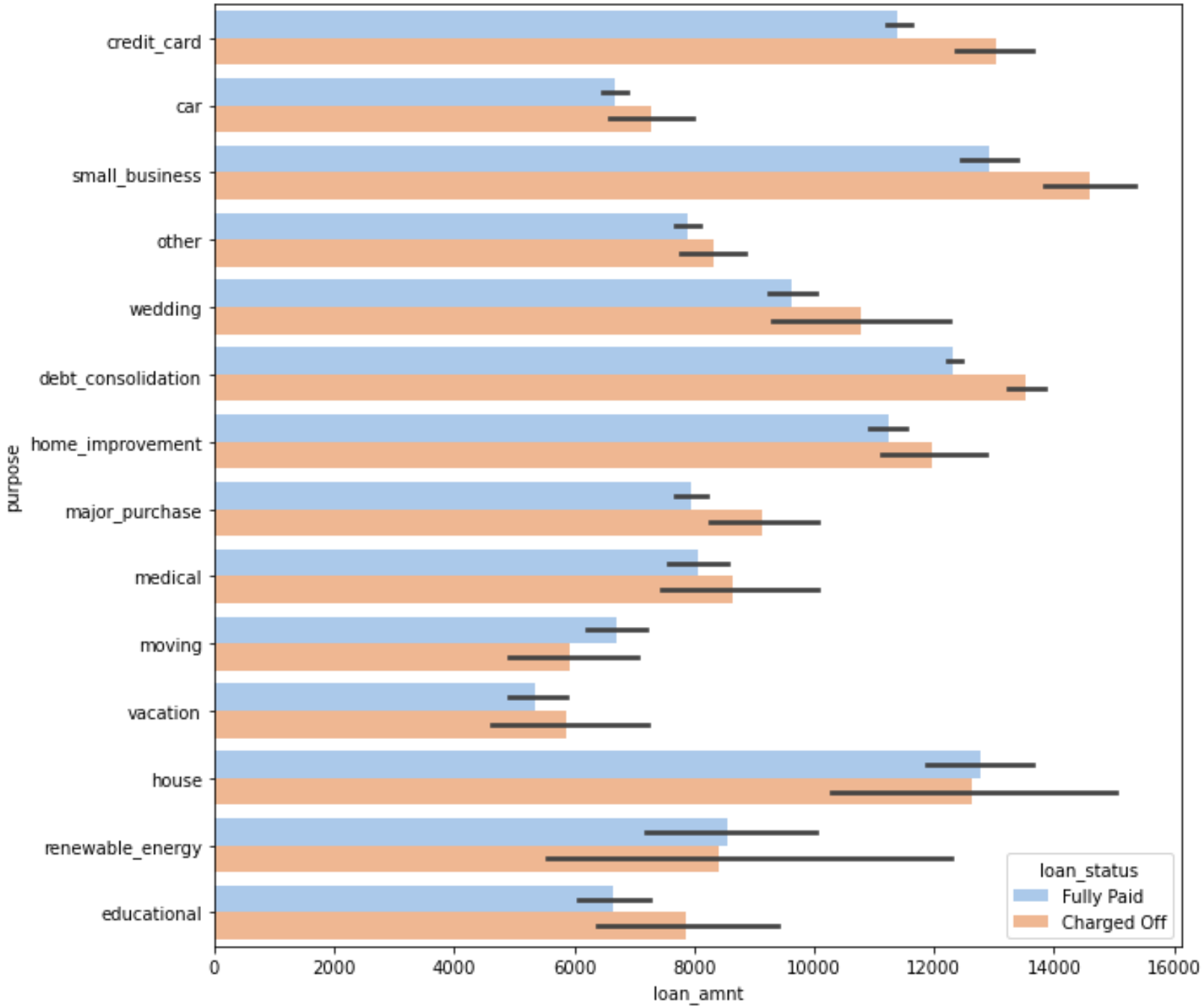

loan_status
Fully Paid
Charged Off
credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose
0
20000
40000
60000
80000
100000
annual_inc

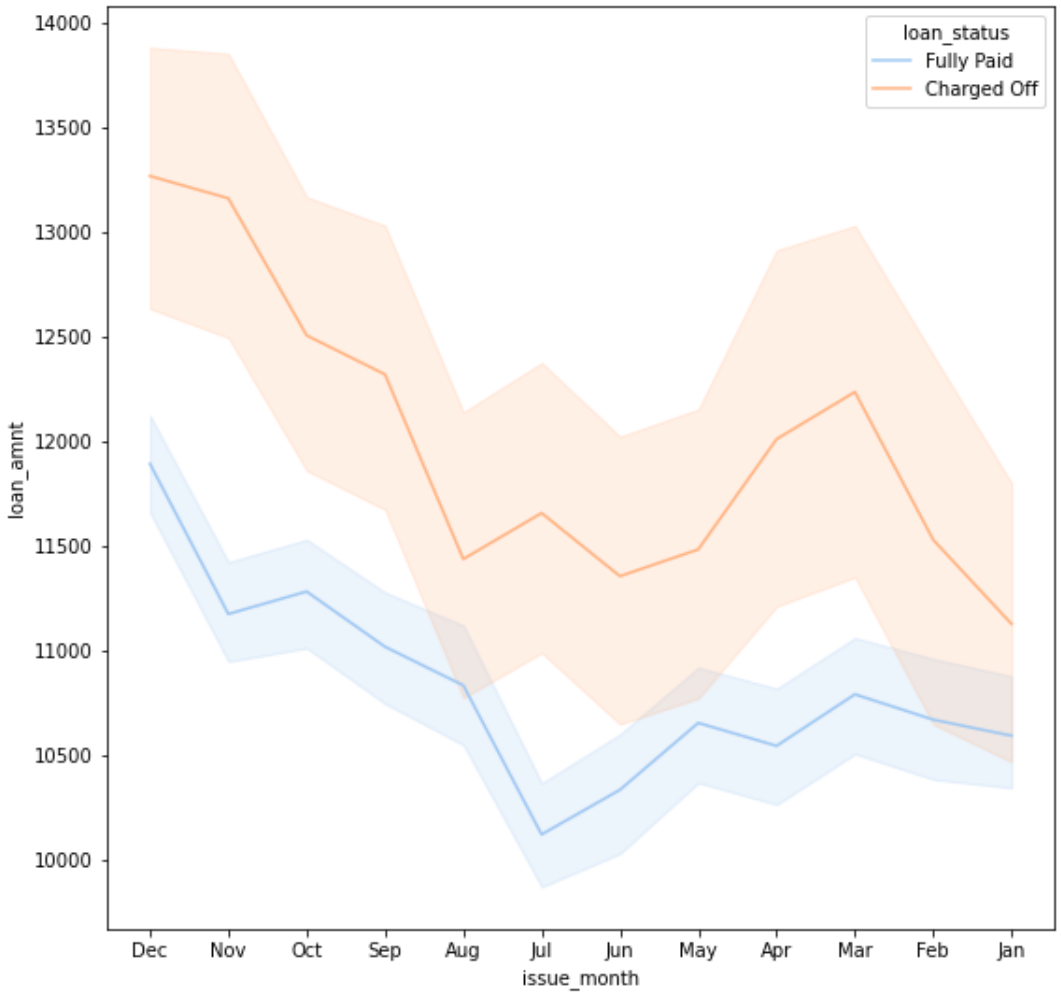
- Applicants with higher salaries applied for home improvement, small business, house, and renewable energy.

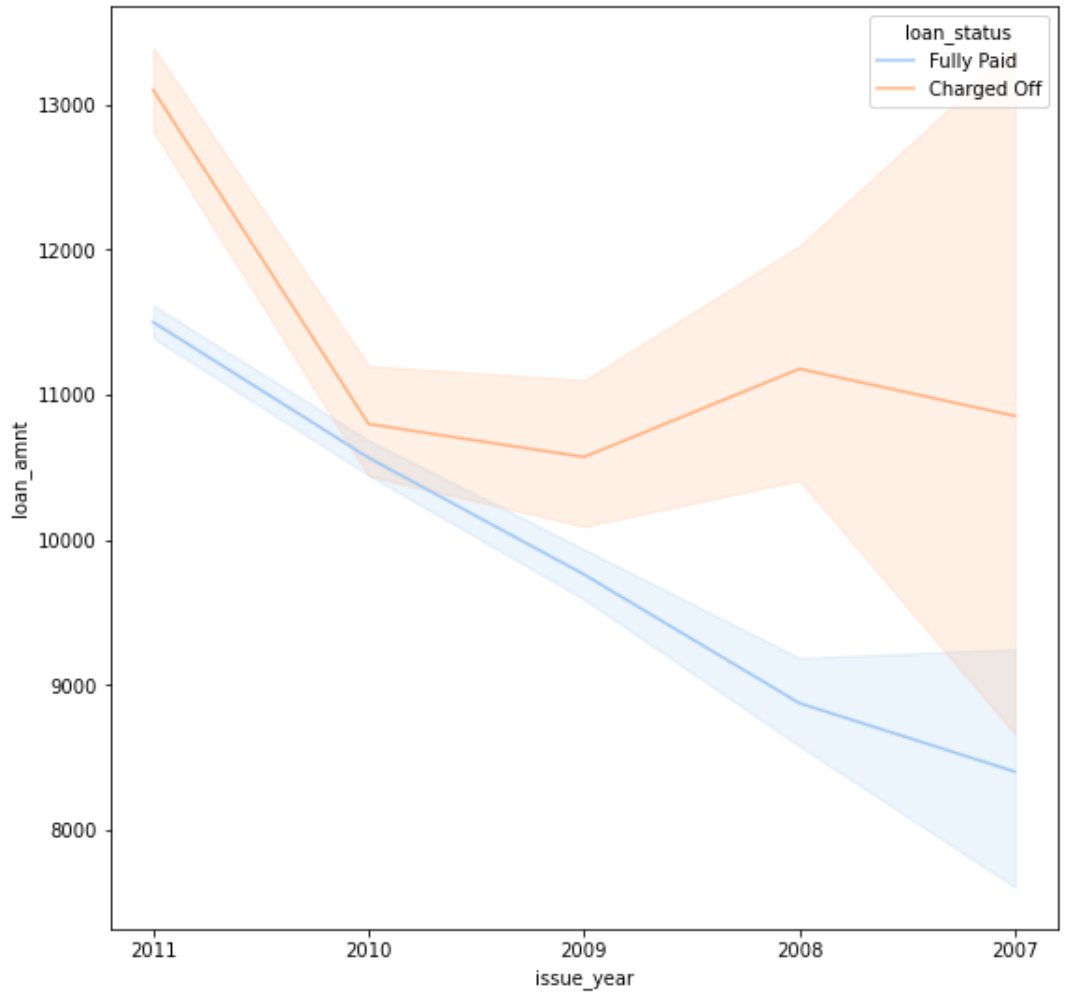


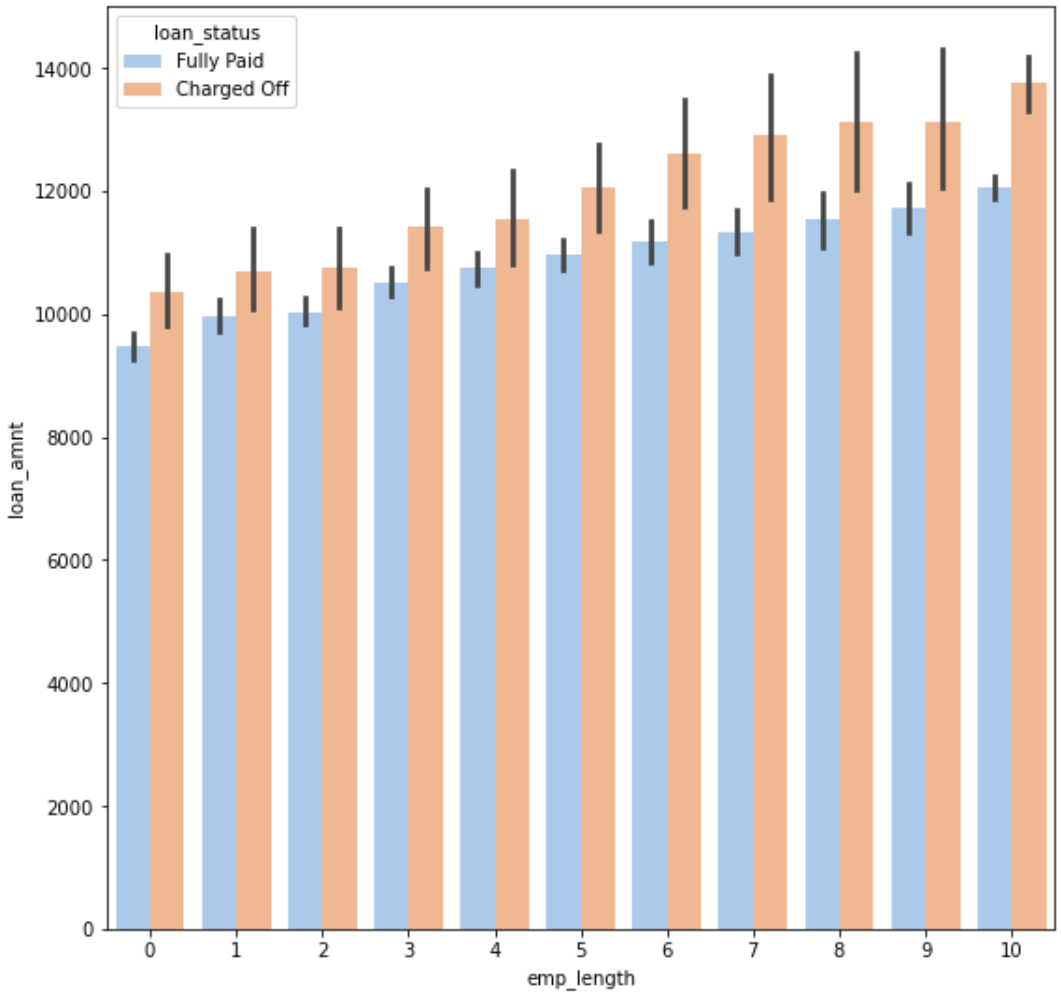
The probability of defaulter is when Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k

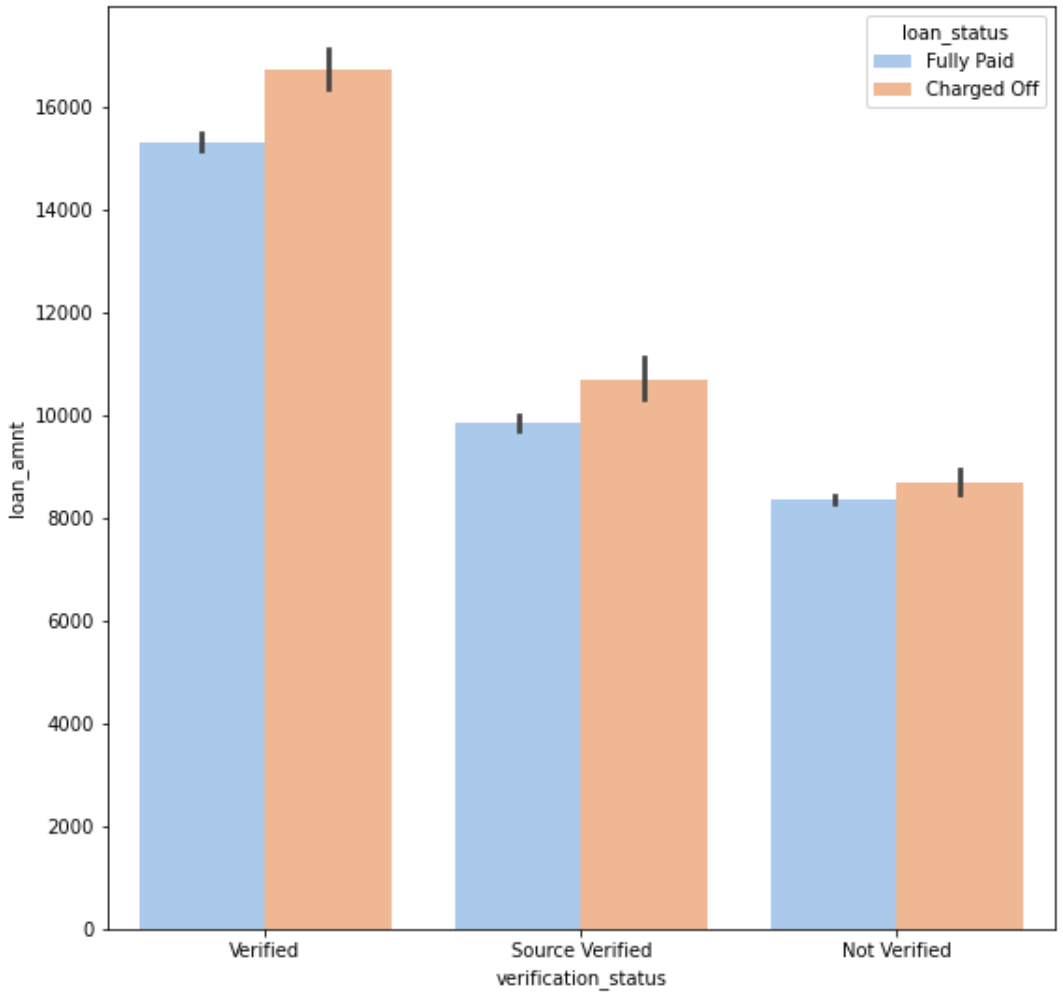

loan_status
Fully Paid
Charged Off
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
int_rate
0-5k
5k-10k
10k-15k
15k-20k
20k-25k
25k-30k
30k-35k
loan_amnt_groups

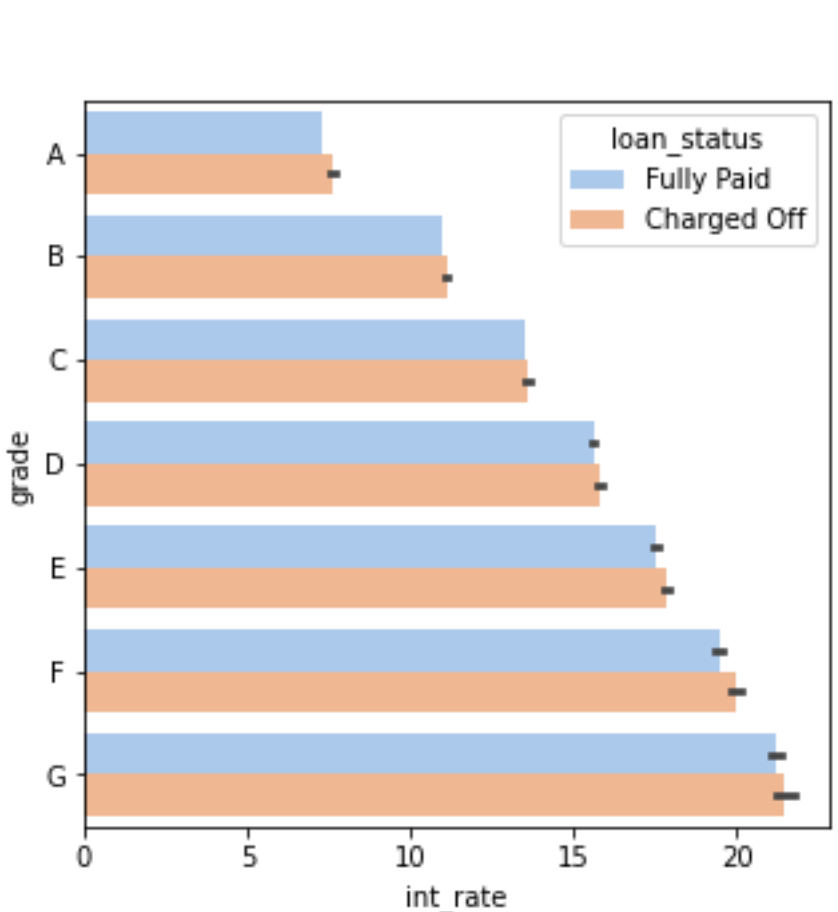

credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose
0
2000
4000
6000
8000
10000
12000
14000
16000
loan_amnt
loan_status
Fully Paid
Charged Off

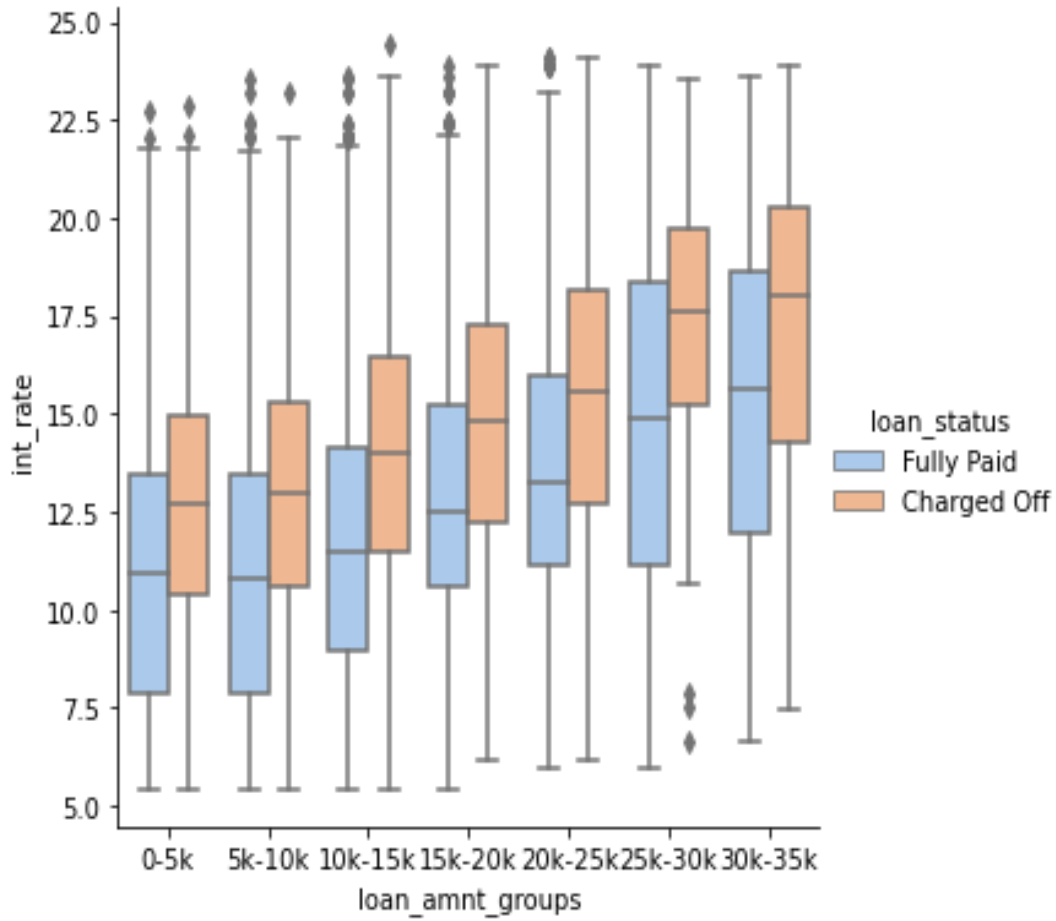

loan_status
Fully Paid
Charged Off
loan_amnt
14000
13500
13000
12500
12000
11500
11000
10500
10000
Dec
Nov
Oct
Sep
Aug
Jul
Jun
May
Apr
Mar
Feb
Jan
issue_month

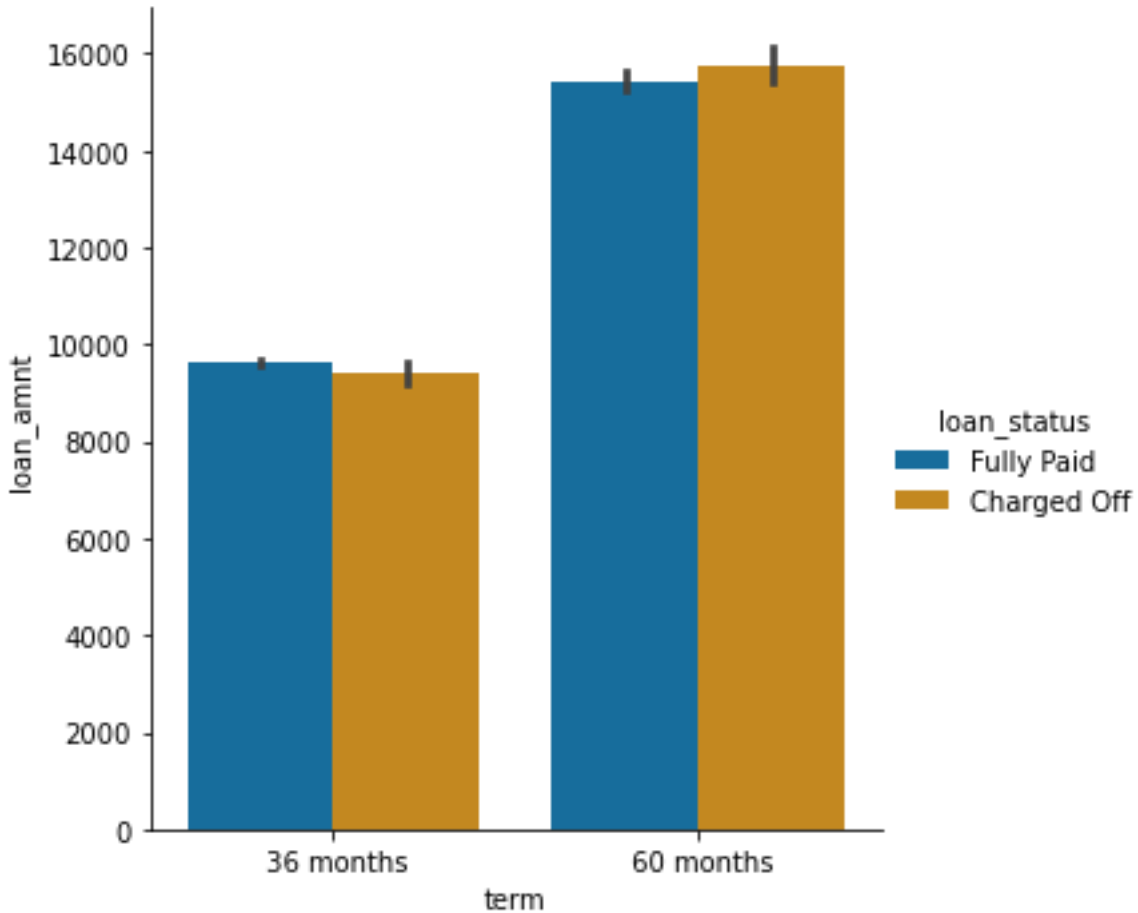

loan_status
Fully Paid
Charged Off
loan_amnt
13000
12000
11000
10000
9000
8000
2011
2010
2009
2008
2007
issue_year


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


loan_status
Fully Paid
Charged Off
A
B
C
D
E
F
G
grade
0
5
10
15
20
int_rate


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


16000
14000
12000
10000
8000
6000
4000
2000
0
loan_amnt
36 months
60 months
term
loan_status
Fully Paid
Charged Off

Consolidated Bi- Variate observations:

The above analysis with respect to the charged-off loans. There is the probability of defaulting when :

1. Applicants taking a loan for 'home improvement' and have an income of 60k -70k
2. Applicants whose homeownership is 'MORTGAGE and have an income of 60-70k
3. Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
4. Applicants who have taken a loan in the range 30k - 35k and are charged an interest rate of 15-17.5 %
5. Applicants who have taken a loan for a small business and the loan amount is greater than 14k
6. Applicants whose homeownership is 'MORTGAGE and have a loan of 14-16k
7. When employment length is 10yrs and loan amount is 12k-14k
8. When the loan is verified, and the loan amount is above 16k

Conclusions

- Lending club should reduce the high-interest loans for 36 months tenure, they are prone to loandefault.
- Grades are a good metric for detecting defaulters. The lending club should examine more informationfrom borrowers before issuing loans to Low grades (Grade B and Sub Grade 5).
- Small business loans default more. Lending clubs should stop/reduce issuing the loans tothem.
- Borrowers with mortgage homeownership are taking higher loans and defaulting to the approved loans. The lending club should stop giving loans to this category when the loan amount requested is morethan 14K – 16K.
- Borrowers who receive interest at the rate of 21-24% and have an income of 70k-80k are prone to loan default