Lending Club Case Study

Problem Statement:

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. The company provides loans to individuals based on various factors such as income and loan purpose. However, some borrowers fail to repay their loans, leading to financial losses to the company. That is the reason the company has to make a decision for loan approval based on applicant's profile when the company receives a loan application.

The company wants to understand the **driving factors (or driver variables)** behind loan default (charged Off). Here the ultimate goal is to identify patterns and derive insights from historical data to:

1. Predict loan defaults accurately.
2. Identify high-risk applicants early.

Dataset Overview and Insights:

1. The data about past loan applicants and whether they 'defaulted' or not. The complete loan data for all loans issued through the time period 2007 to 2011.
2. The Loan status is categorized into three types 1. Fully Paid, 2. Current, 3. Charged-off
3. The dictionary to understand the variables.

Approach:

1. Understand the data set
2. Data preparation (cleaning, standardizing and treating outliers)
3. Data Analysis
4. Conclusion

1. Understanding the data set

1. There are no headers and footers in the given data set
2. There are 39717 rows and 111 columns in total
3. Out of 111 columns 54 columns are all null values
4. Out of 111 columns 9 columns which have only one value for all rows
   - 4.1. ➔ **pymnt_plan, initial_list_status, collections_12_mths_ex_med, policy_code, application_type, acc_now_delinq, chargeoff_within_12_mths, delinq_amnt, tax_liens**
5. Out of 111 Columns there are 3 columns which has all values unique
   - 5.1. ➔ **Id, member_id, url**
6. From observing the Dictionary, it's clear that there are few variables which will not contribute to our analysis
   - 6.1. ➔ **title, emp_title, zip_code, last_credit_pull_d, addr_state, desc, out_prncp_inv, total_pymnt_inv, delinq_2yrs, revol_bal, out_prncp, total_pymnt, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, next_pymnt_d, mths_since_last_delinq, mths_since_last_record, pub_rec, pub_rec_bankruptcies**
7. We have grade and sub_grade. Sub_grade is sub category of the variable grade

2. Data Preparation

2.1. Dropped **all columns** where all values are **null**

2.2. Dropped **all columns** where values are **identical across all rows**

2.3. Dropped **all column** where all values are **unique**

2.4. Dropped **all columns** that were not **relevant** to our analysis

2.5. Dropped the records where the **loan_status ='Current'**, as this will not be useful for our analysis

2.6. After cleaning the data, we have the relevant variables in our data set

➔ **loan_amn,funded_amnt,funded_amnt_inv,term,int_rate,installment, grade, sub_grade, emp_length, home_ownership, annual_inc, verification_status, issue_d, loan_status, purpose, dti, earliest_cr_line, inq_last_6mths, open_acc, revol_util, total_acc, issue_d_month, issue_d_year**

2.7. Assigning the **mode** value to null values in **“emp_length”** as the mode 10+years value is almost twice the count of the next three most frequent values.

2.8. Dropped the missing values for **revol_util** as the missing percentage is very low, .1%

2.9. Removed “%” symbol from **“revol_util”** and **“int_rate”** and converted their values to float

2.10. Derived two variables **“issue_d_month”, “issue_d_year”** from **“issue_d”**

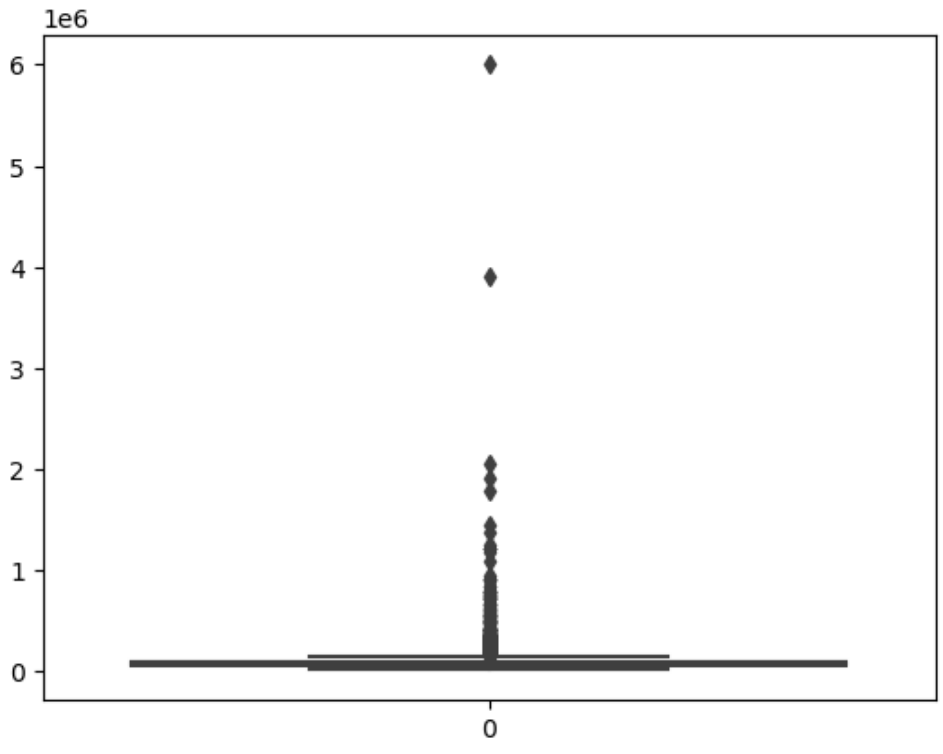
2.11. Formatted the **emp_length column** to improve readability and visualization

2.12. Imputing the value for **“home_ownership”** variable where value is **“NONE”** and assume that to **“OTHER”**

2.13. As variable **“term”** has only two values 30 months and 60 months, we will keep them as it

2.14. Handling Outliers of numeric variables in our current dataset

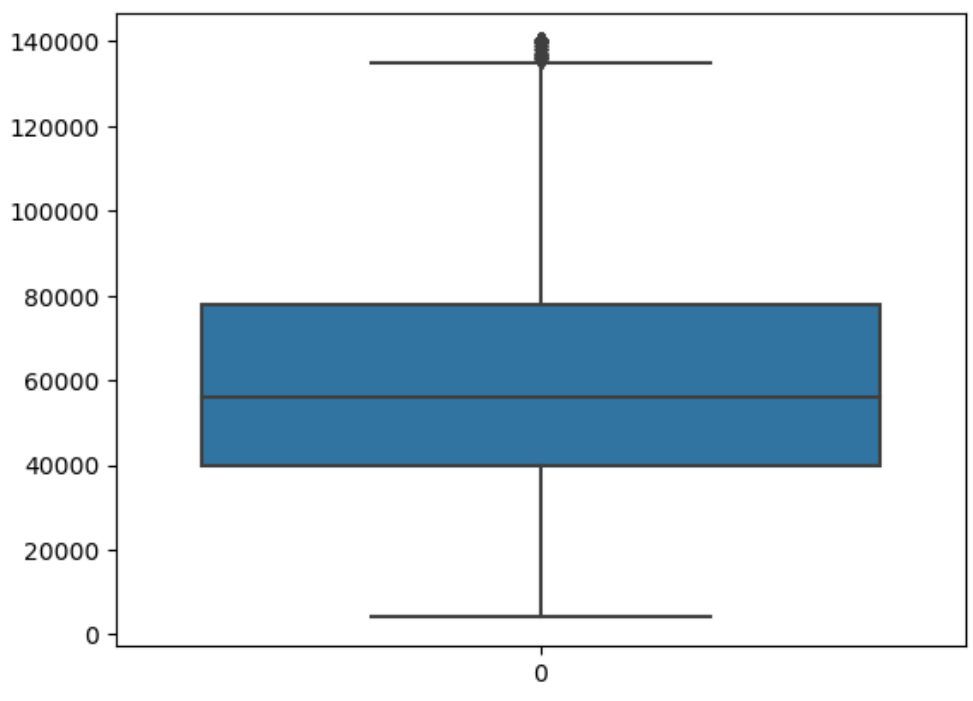
A. **Handling Outlier for annual_inc**



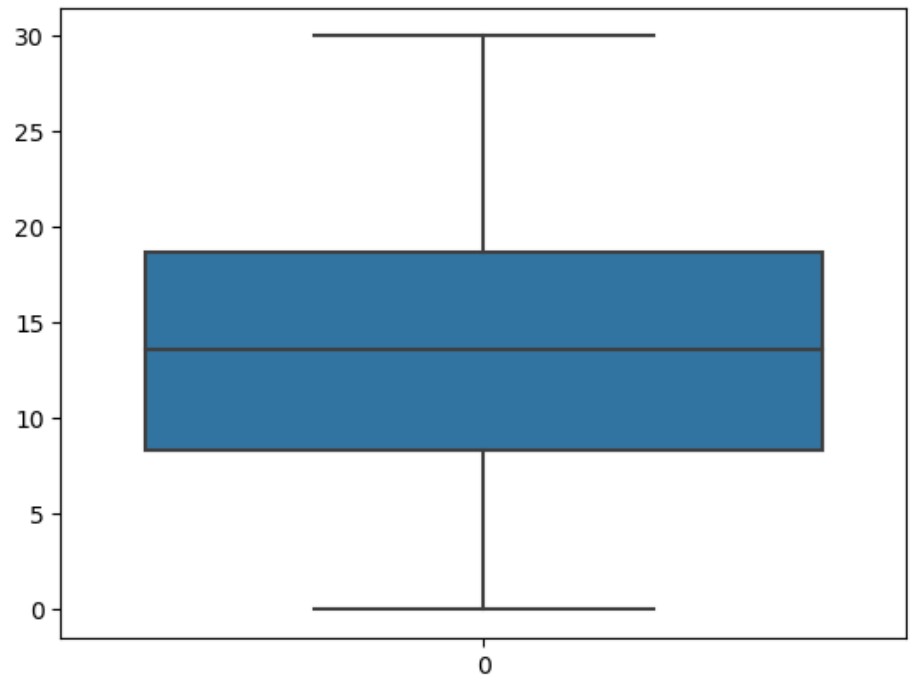
The above representation clearly indicating the presence of outliers in annual_inc

The value after 0.95 quantile seems to be disconnected with the distribution

Hence, we have restricted our data set to 0.95 quantile to avoid the irrelevant outlier data

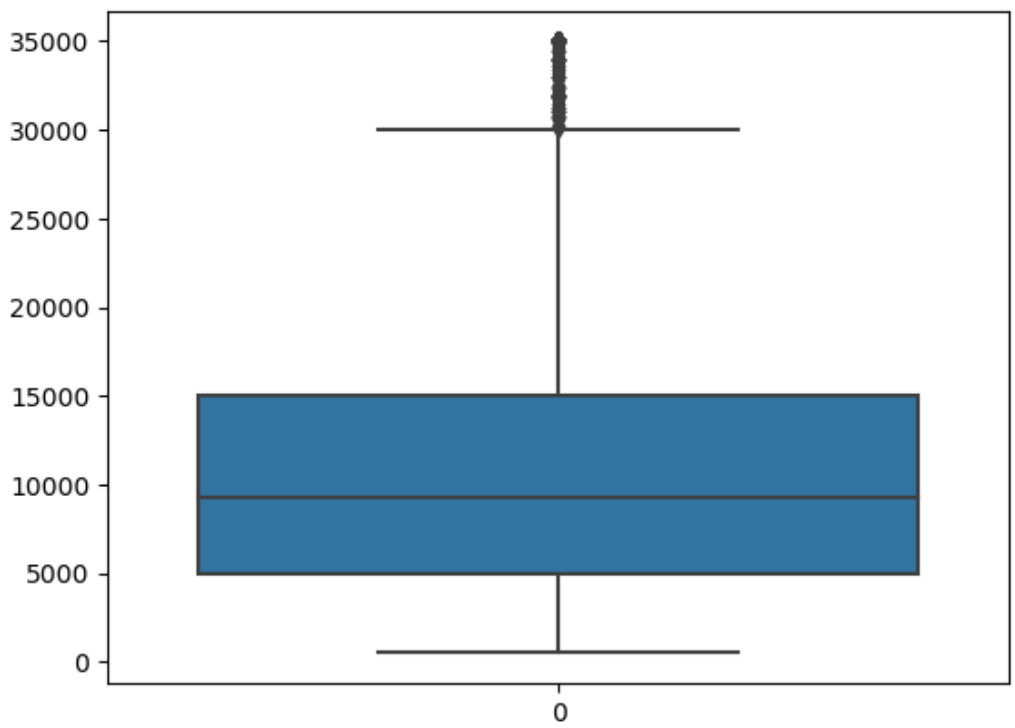


B. **Handling Outlier for dti**



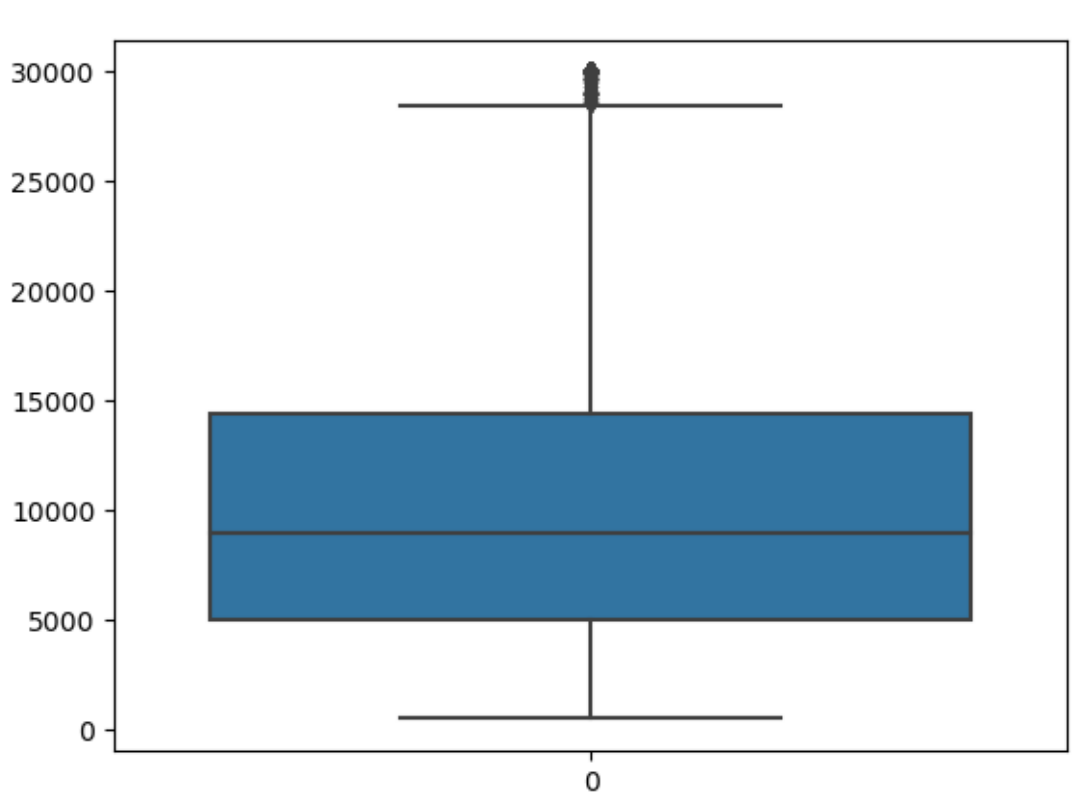
There are no outliers identified with dti, hence keeping the dti variable data as it is

C. **Handling Outlier for loan_amt**

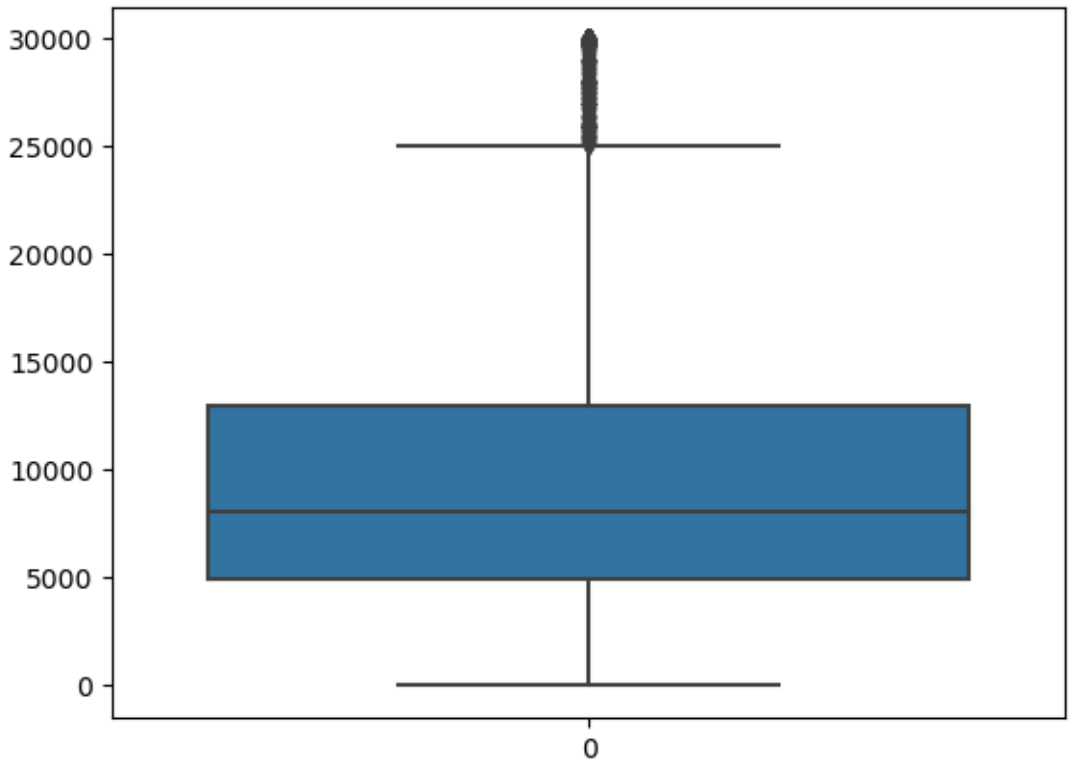


From the above representation, it is evident that there are outliers for loan_amnt

From the quantile information, we can observe the values are continuous until 0.98 and there is sudden spike, hence considering the value only inside the quantile 0.98



D. **Handling Outlier for funded_amnt_inv**

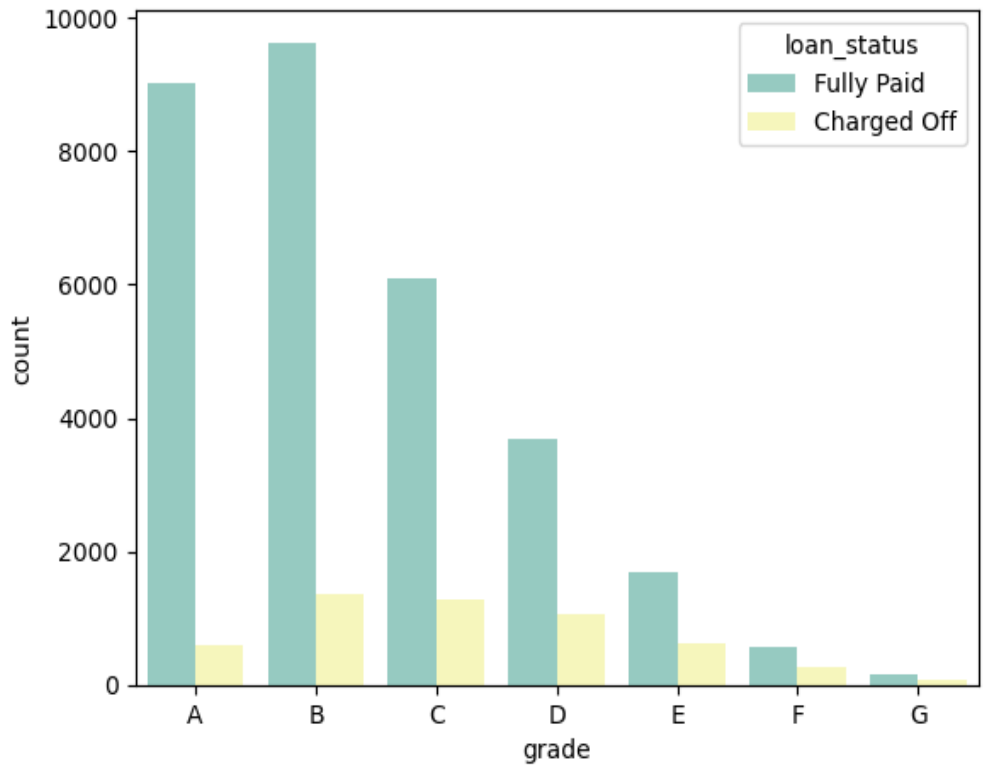


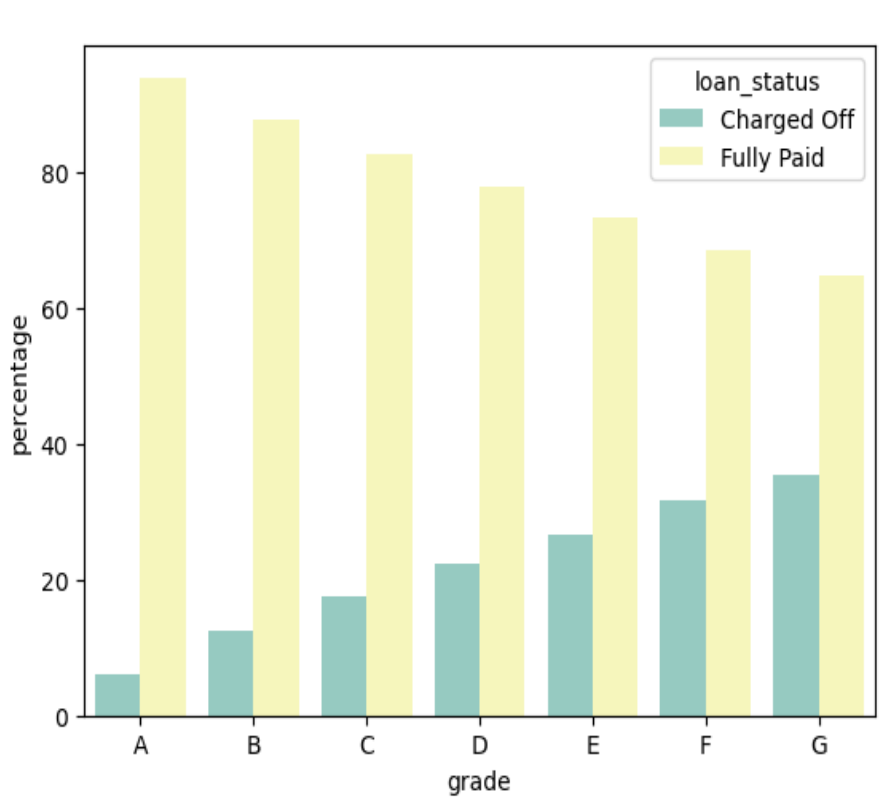
Though there are some values far from distribution, the distribution is pretty continuous and there is no need to remove outliers / extreme values for these above columns

3. Data Analysis

3.1.Data Analysis Part-1 (Using loan_status with one more column)

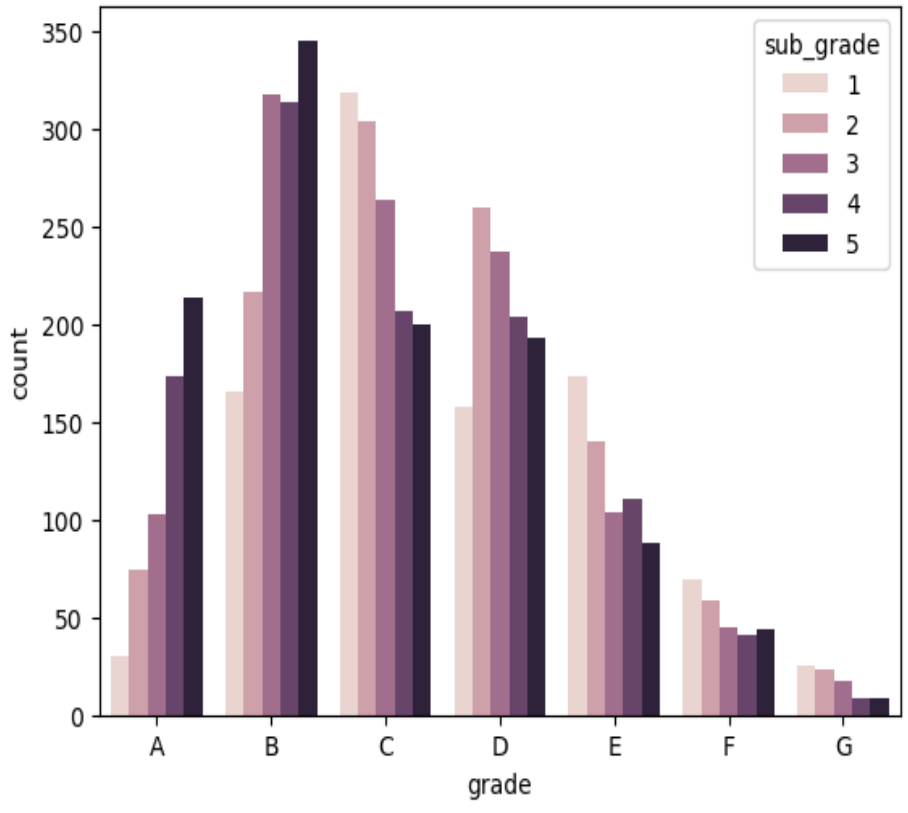
3.1.1) Analyzing Grade:

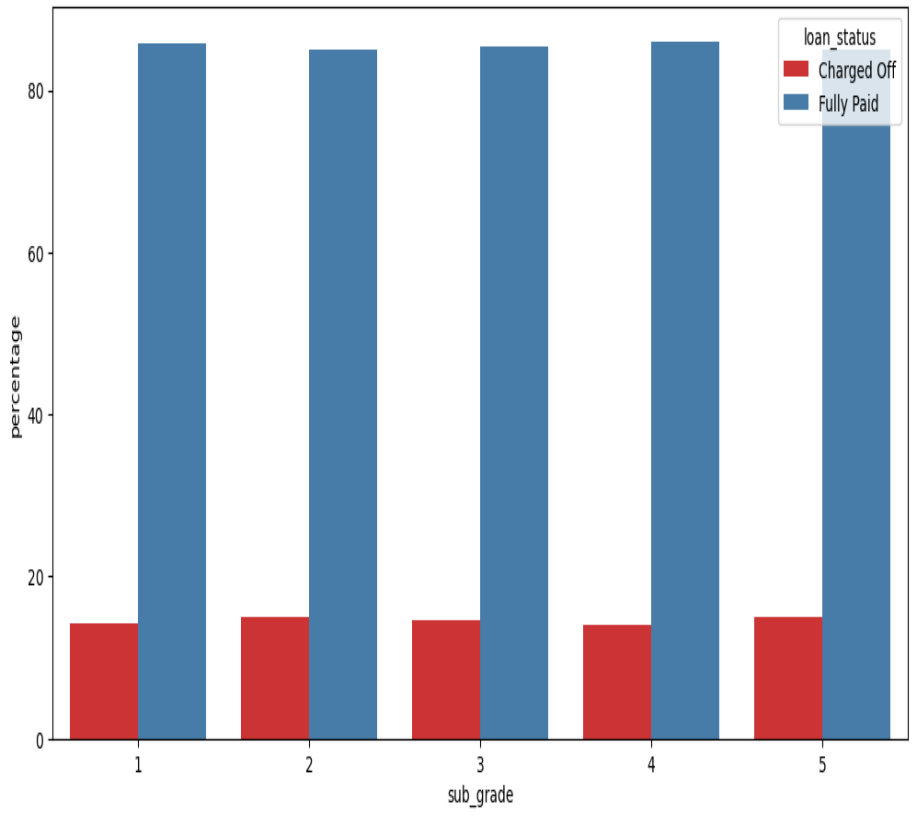




As can be seen above the highest percentage of defaulters (Charged Off) are in grade G (next highest in F and then in E).

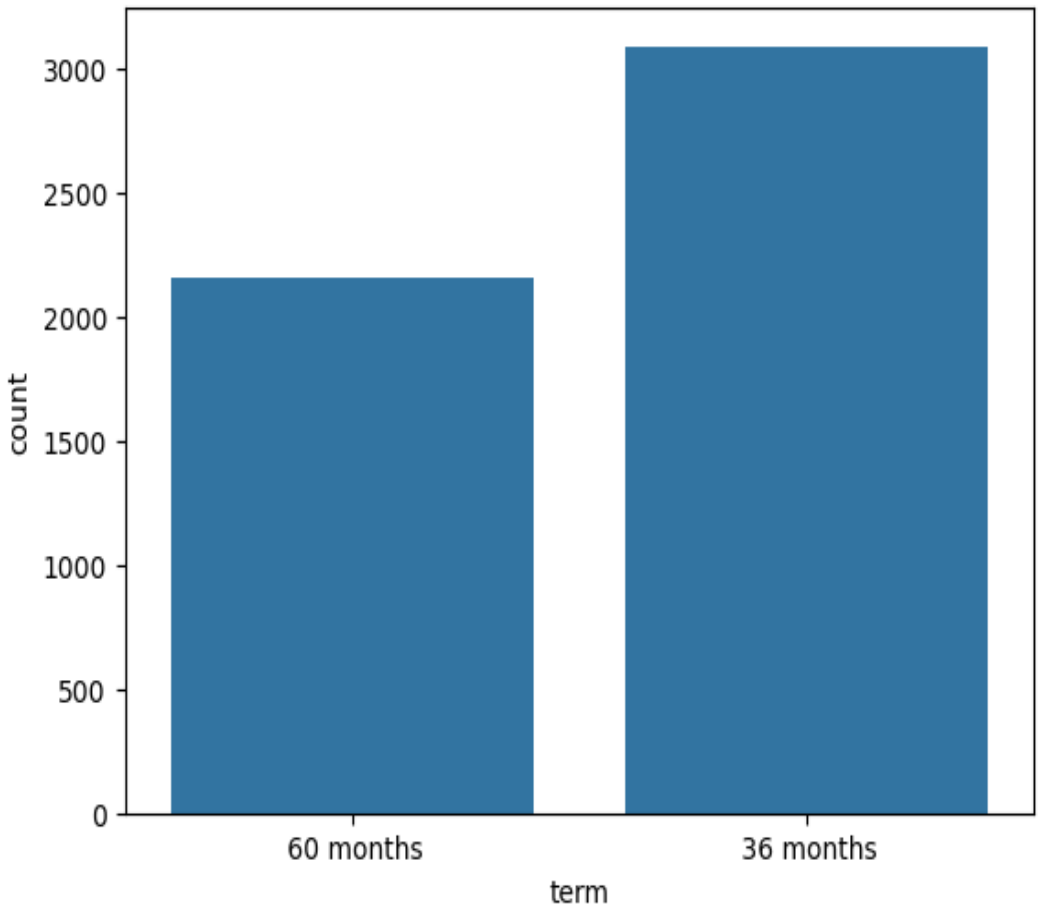
3.1.2) Analyzing sub_grade:

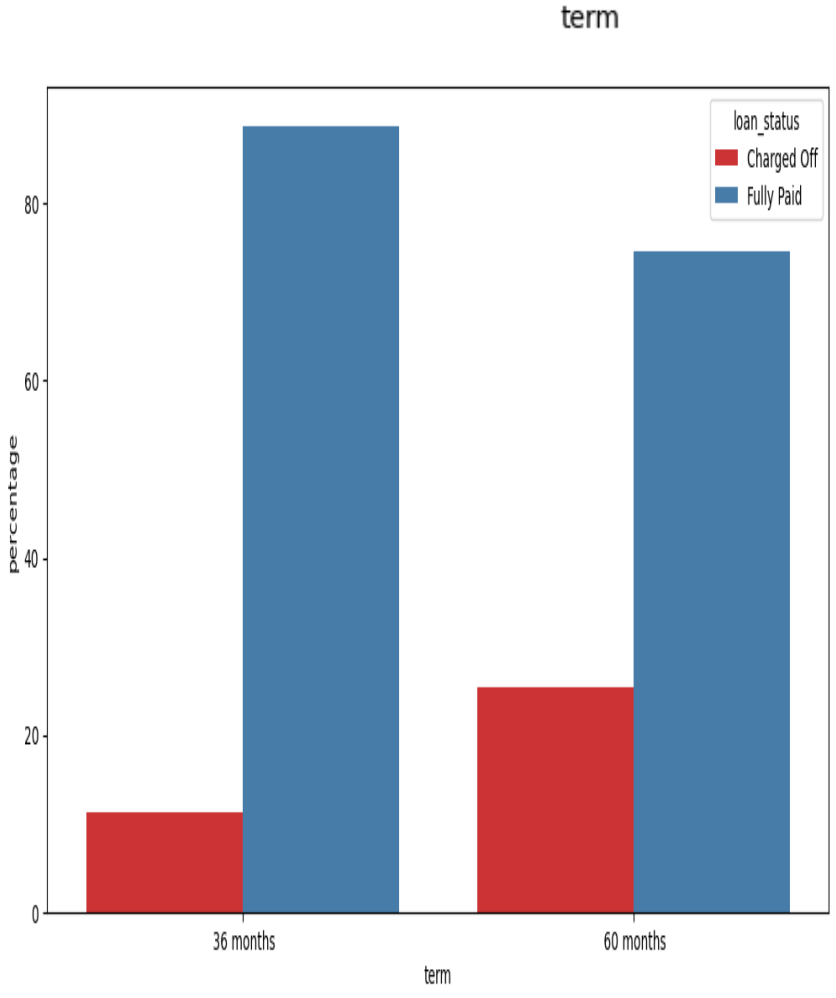




No impact. All sub grades show equal percentages of Charged Off and Fully Paid loans.

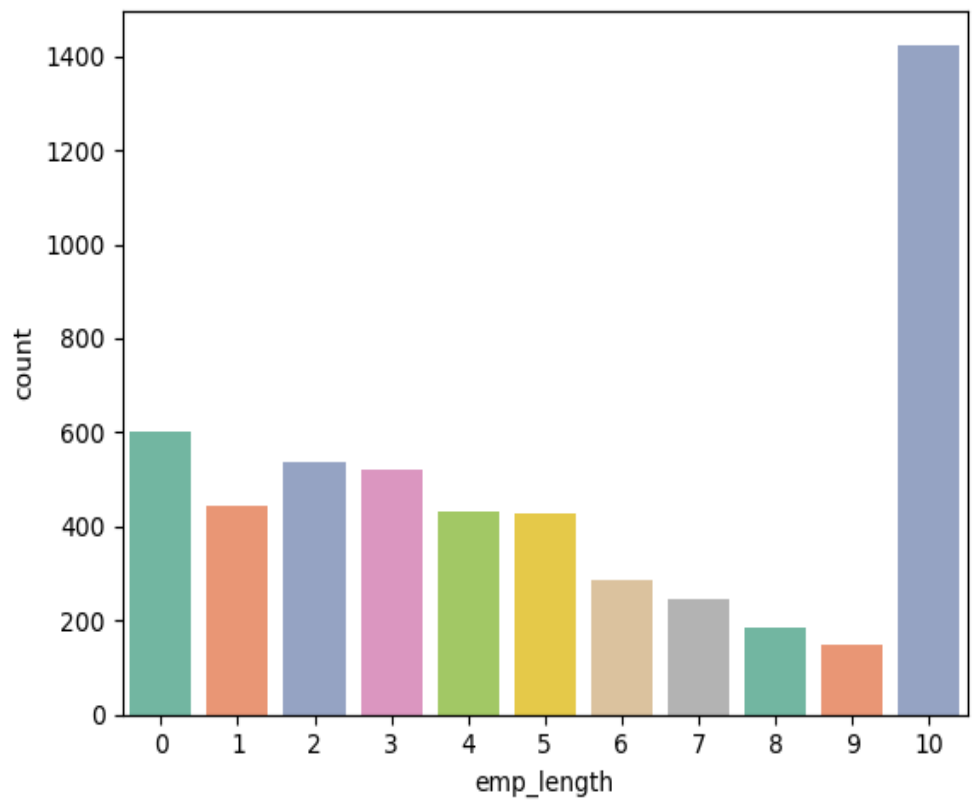
3.1.3) Analyzing term

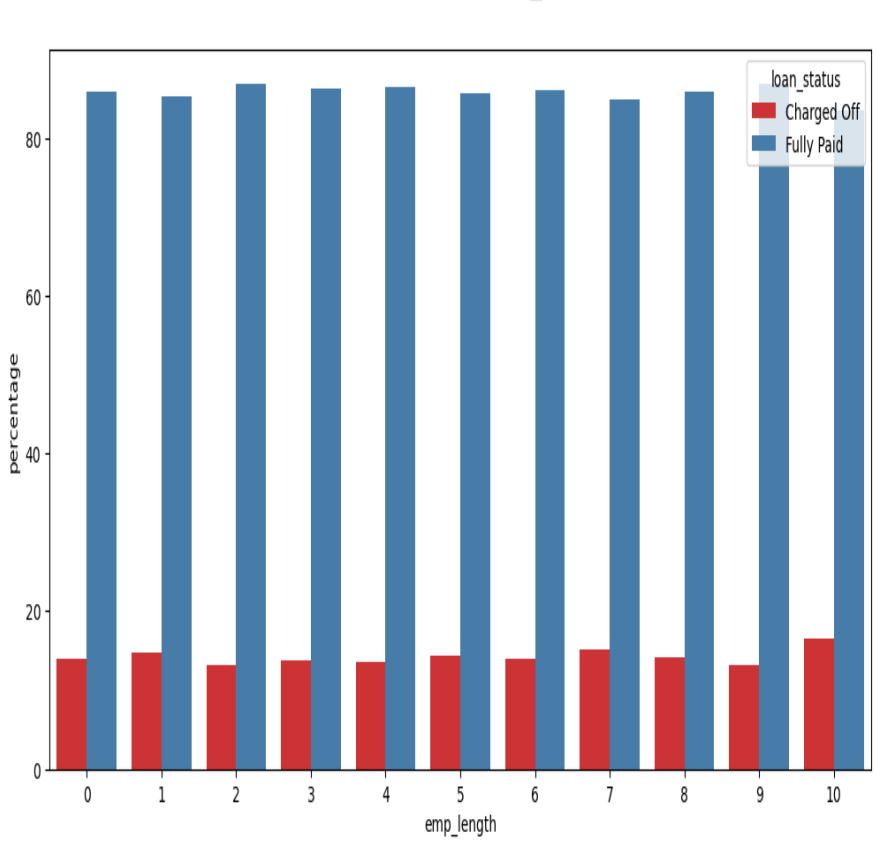




60 Months term shows higher percentage of Charged Off Loans.

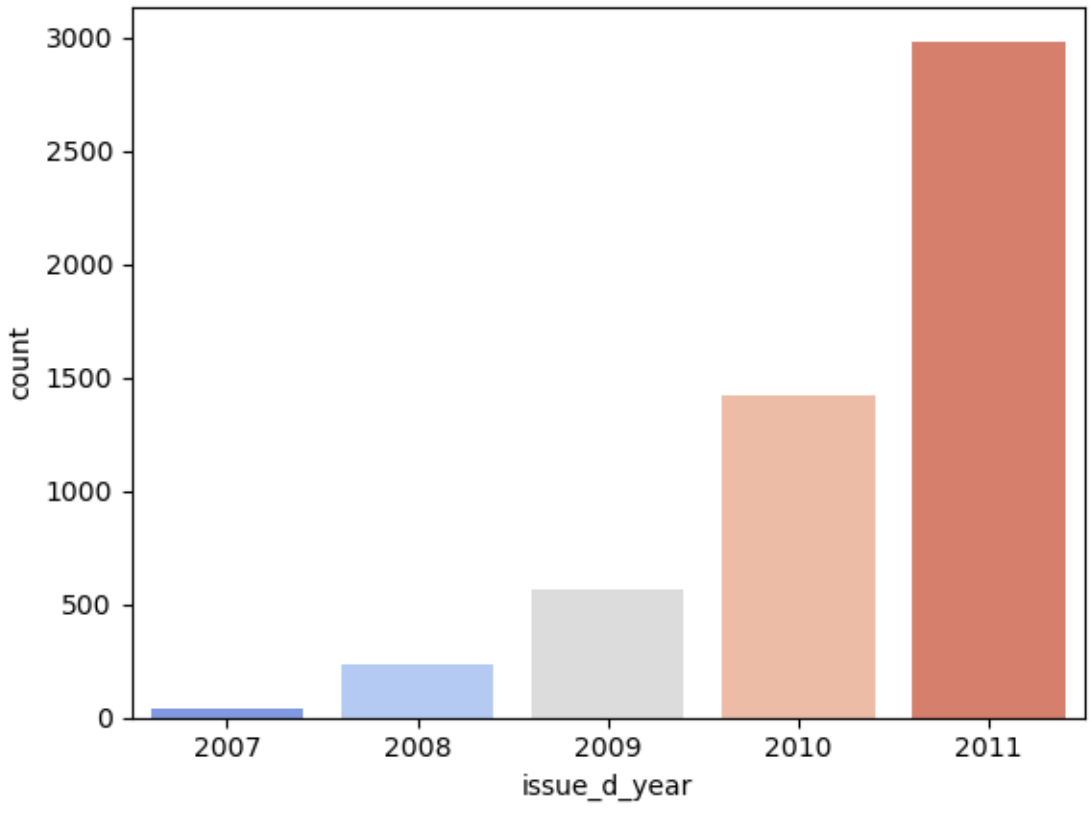
3.1.4) Analyzing emp_length:



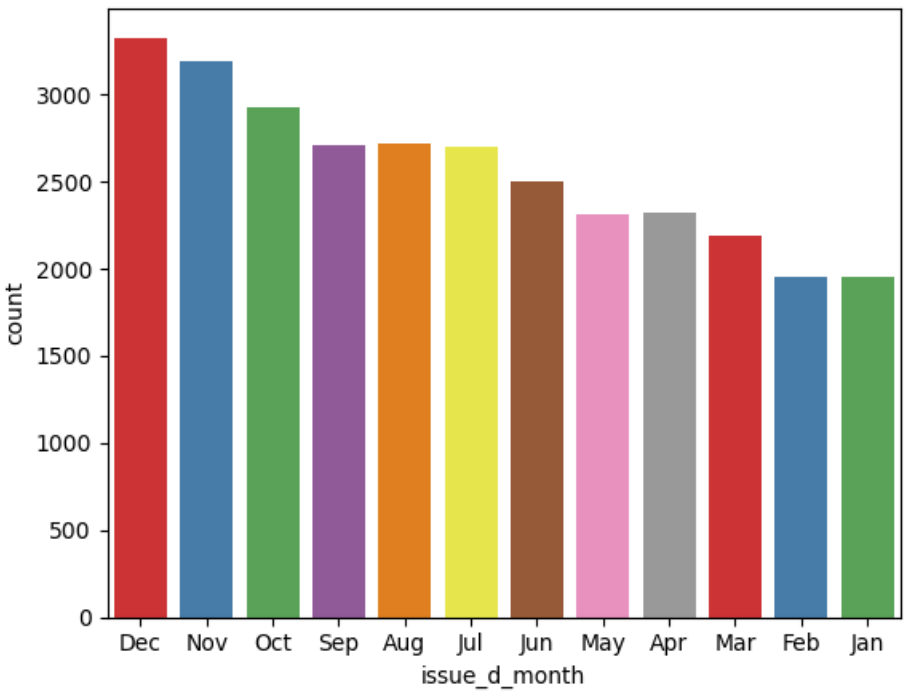


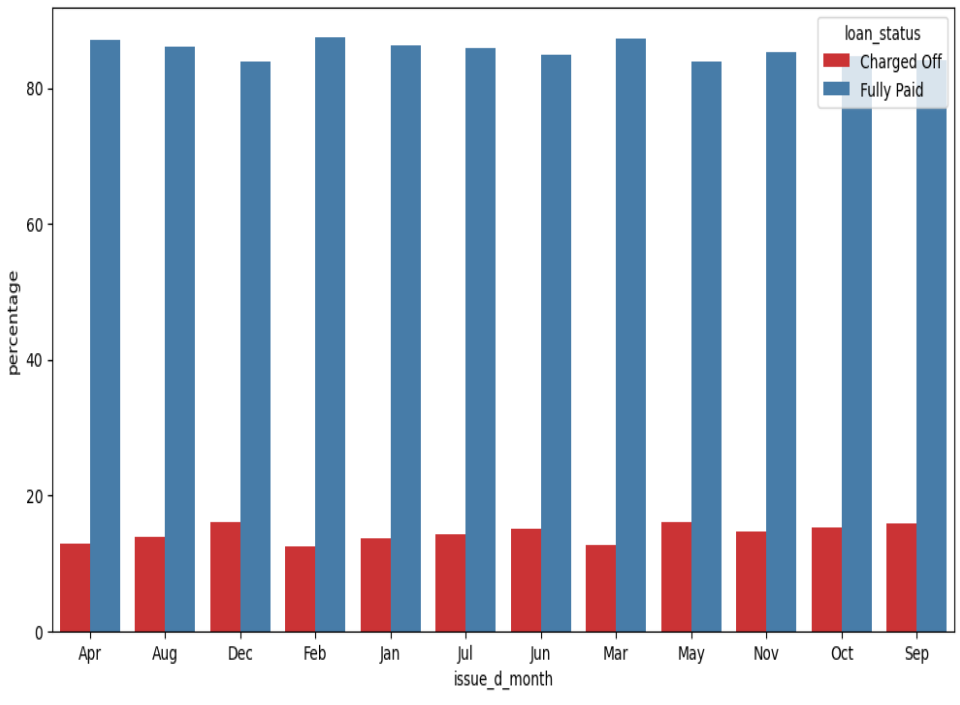
No Impact. All emp_length values have almost equal percentage of defaulters though emp_length >=10 years have little higher percentage of defaulters.

3.1.5) Analyzing issued_d_year (derived variable from issue_d)



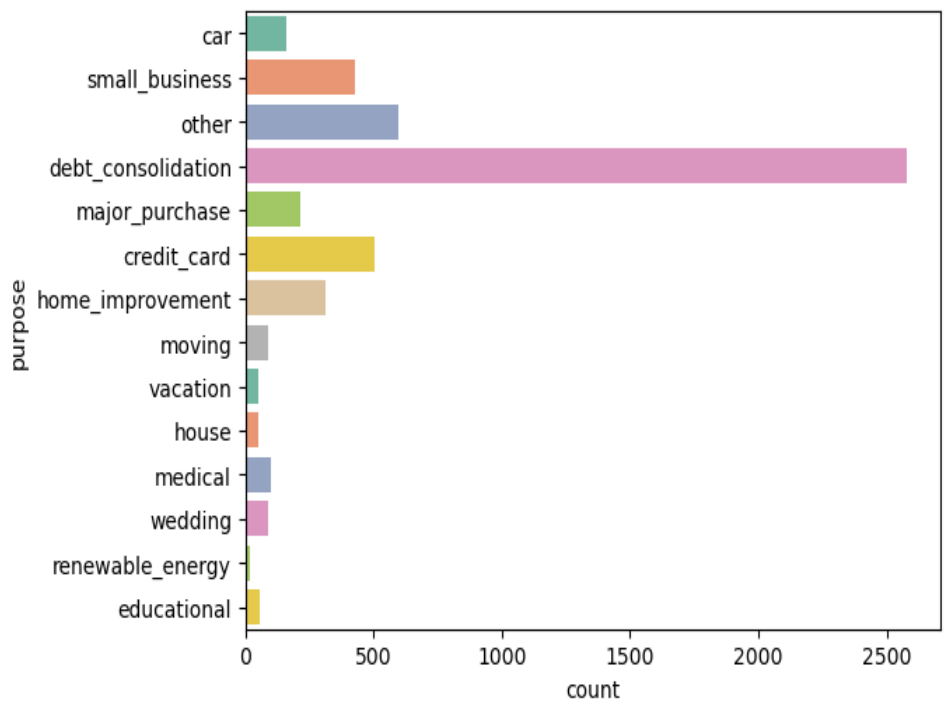
3.1.6) Analyzing issued_d_month (derived variable from issue_d)¶

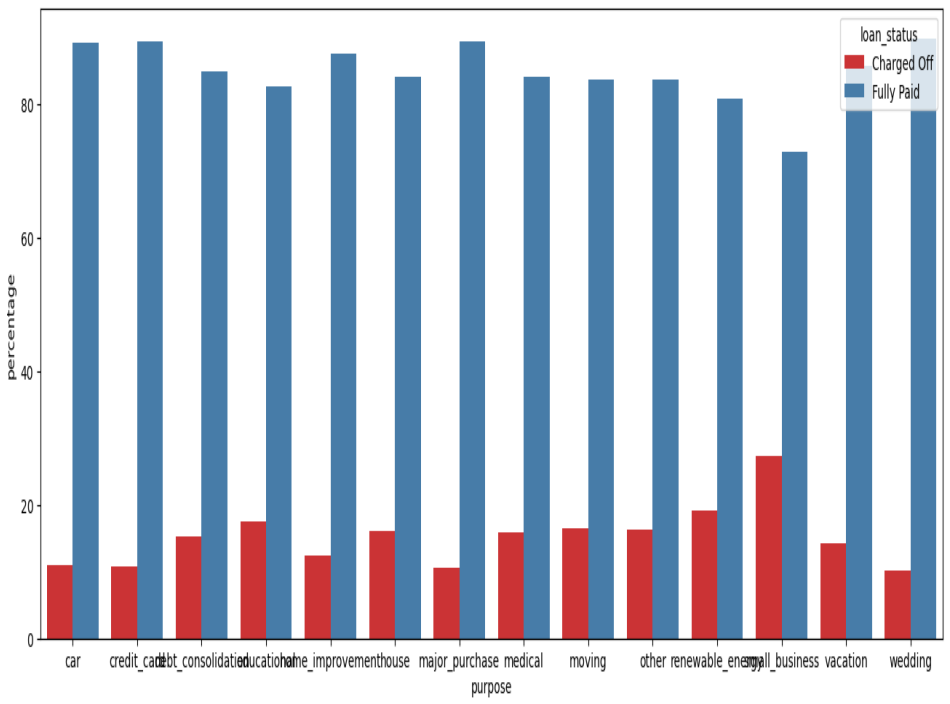




No Impact

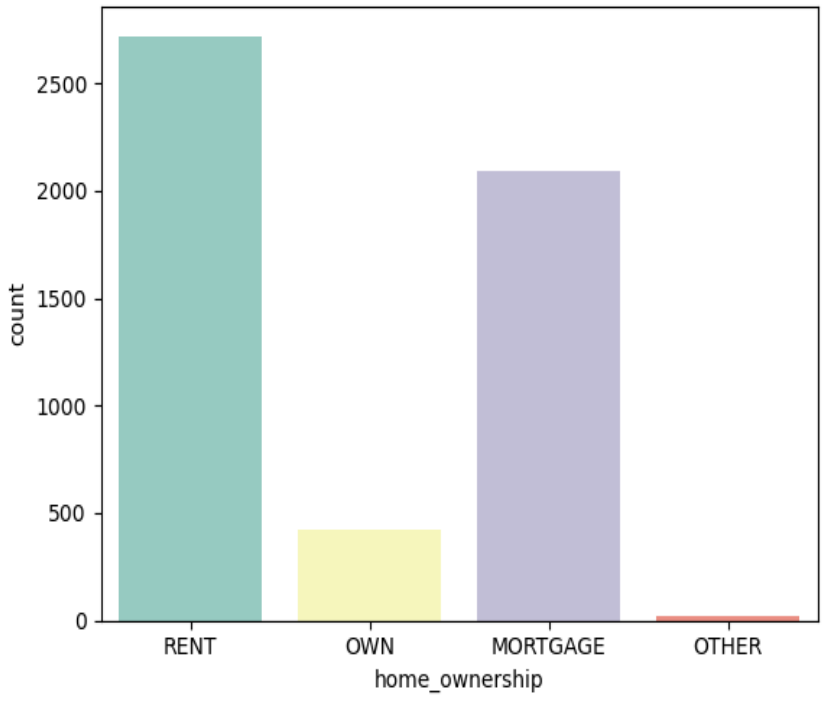
3.1.7) Analyzing purpose:

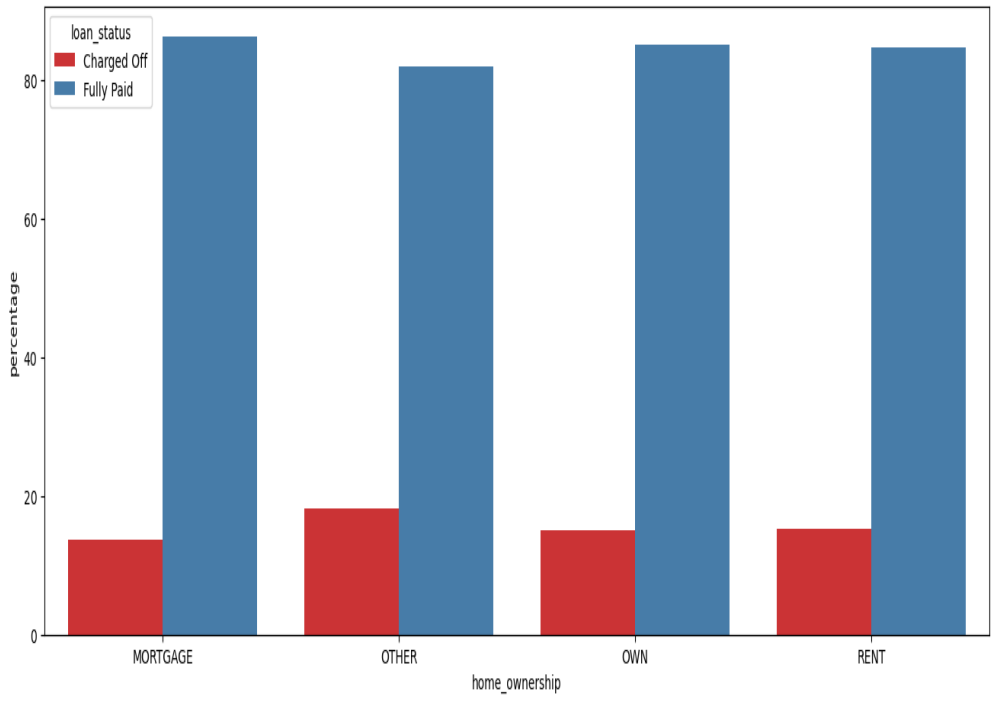




Highest percentage of defaulters (Charged Off) are there for loans taken for 'small_business'.

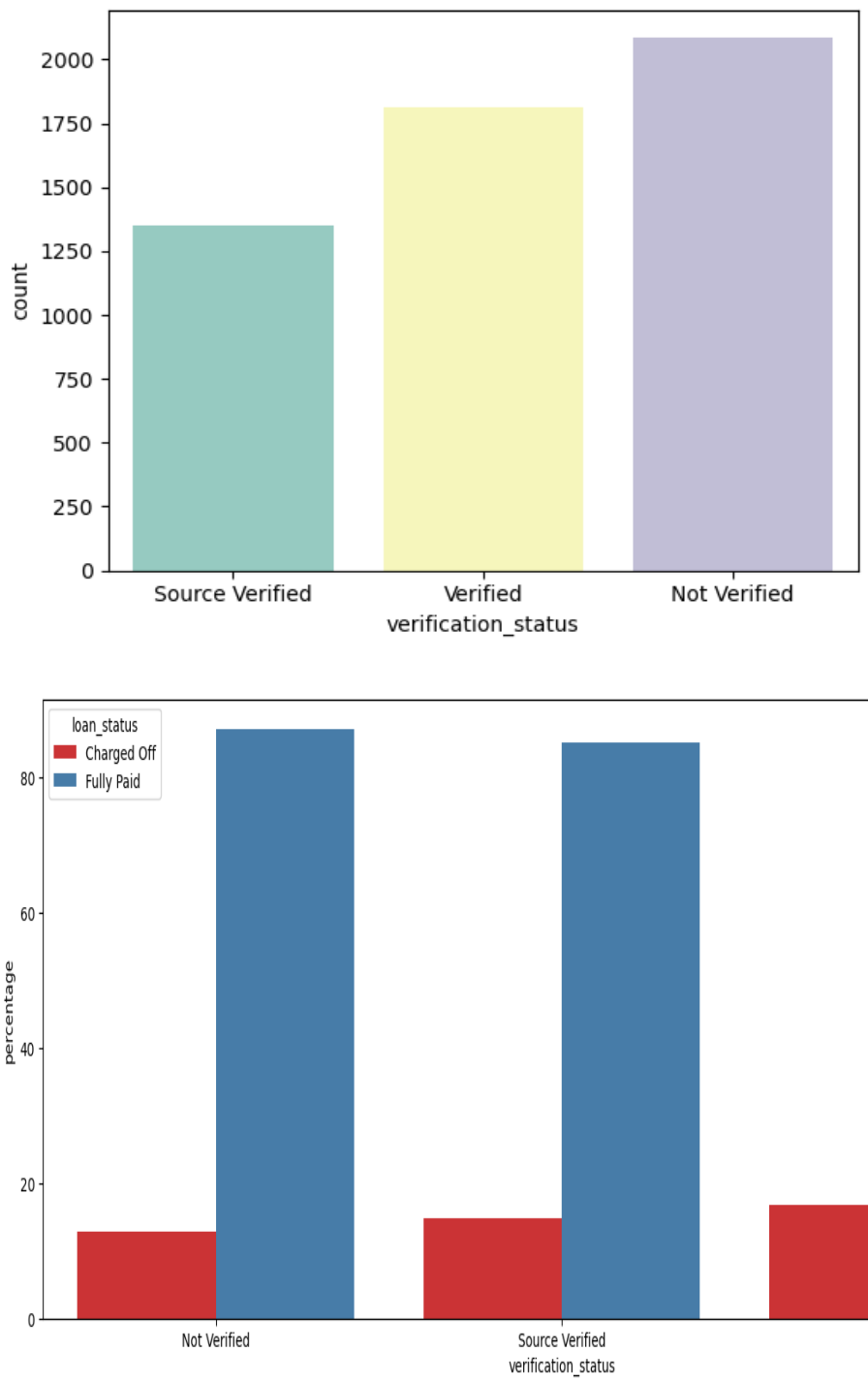
3.1.8) Analyzing home_ownership:



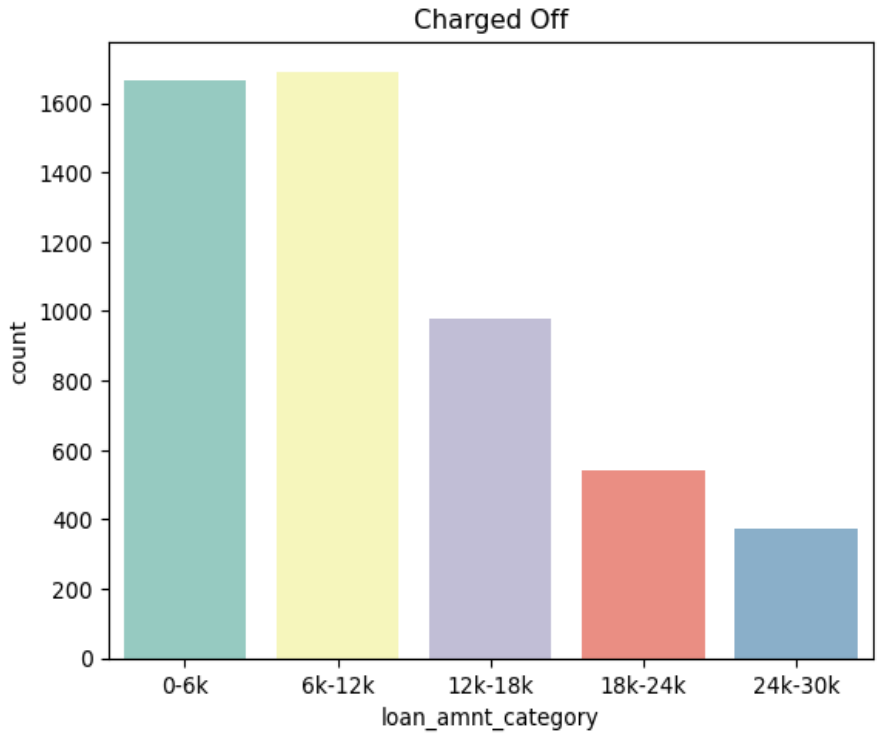


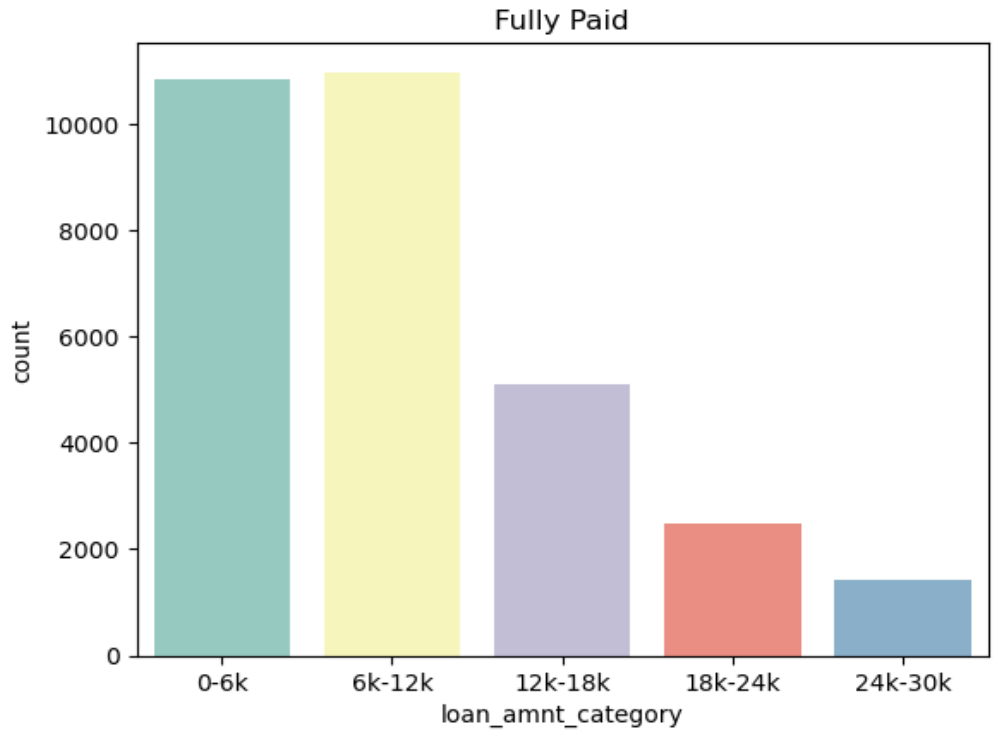
Home_Ownership has almost no significant impact. Though Home_ownership = OTHER has slightly higher percentage of defaulters(Charged Off loans) as compared to other values of home_ownership.

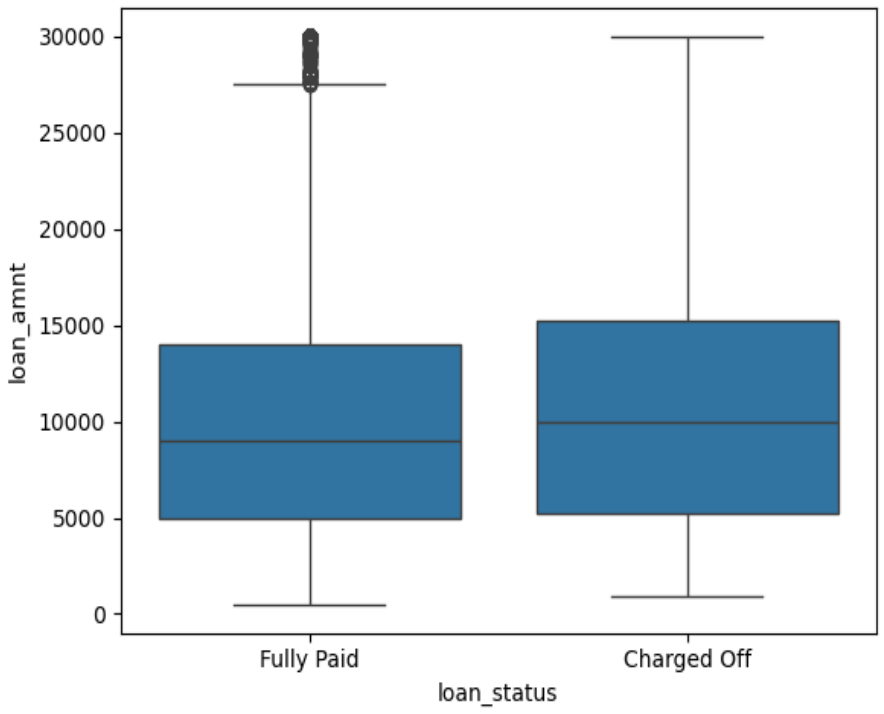
3.1.9) Analyzing verification_status:

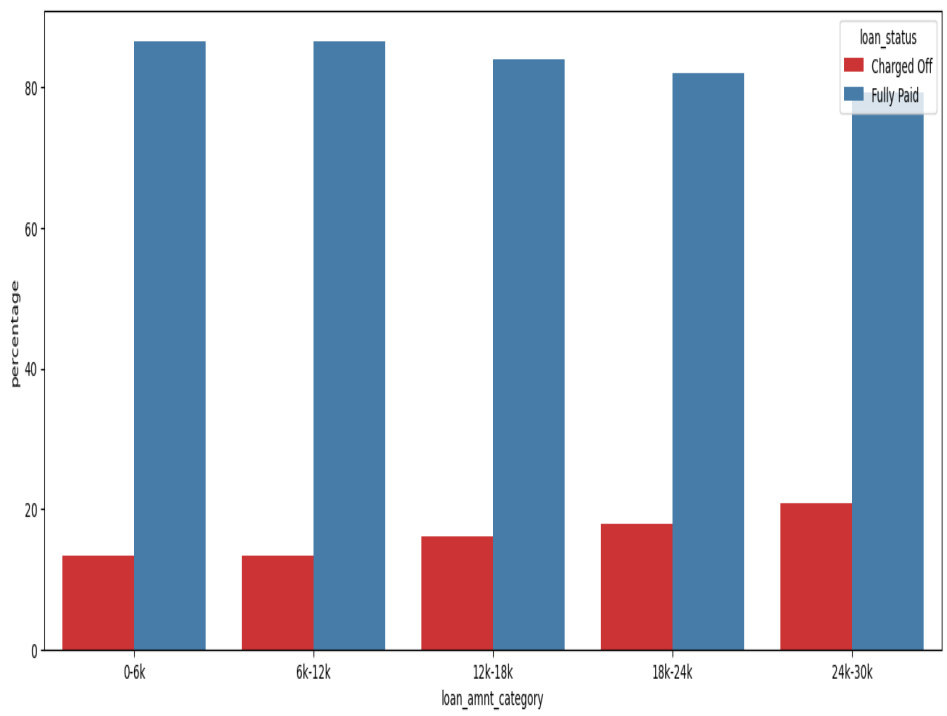


3.1.10) Analyzing loan_amnt:



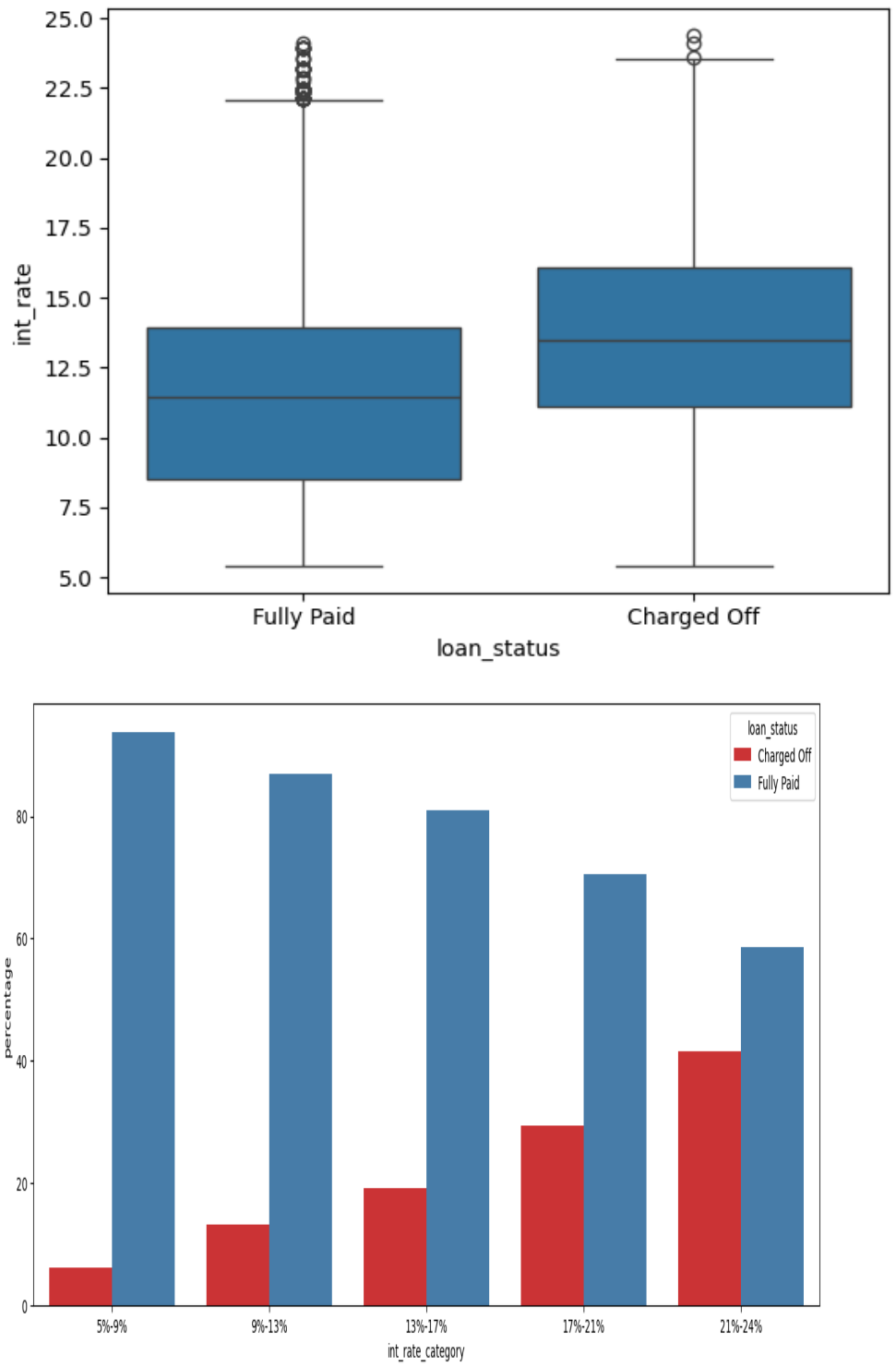






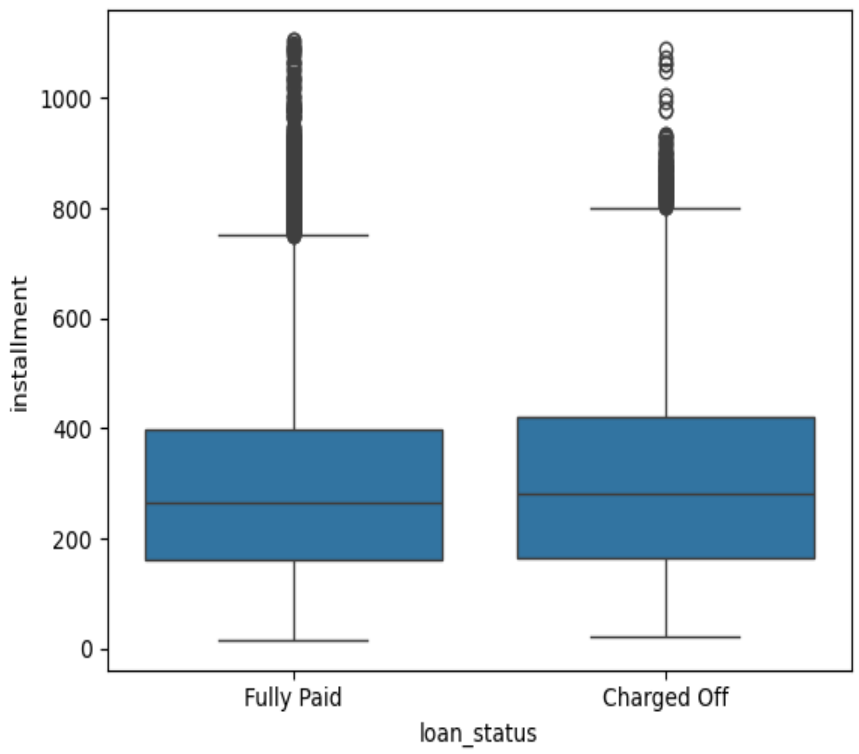
It seems that as the loan amount increases the possibility of default(Charged Off) increases. The loan_amnt_category with highest loan seems to have highest percentage of loan default(Charged Off).

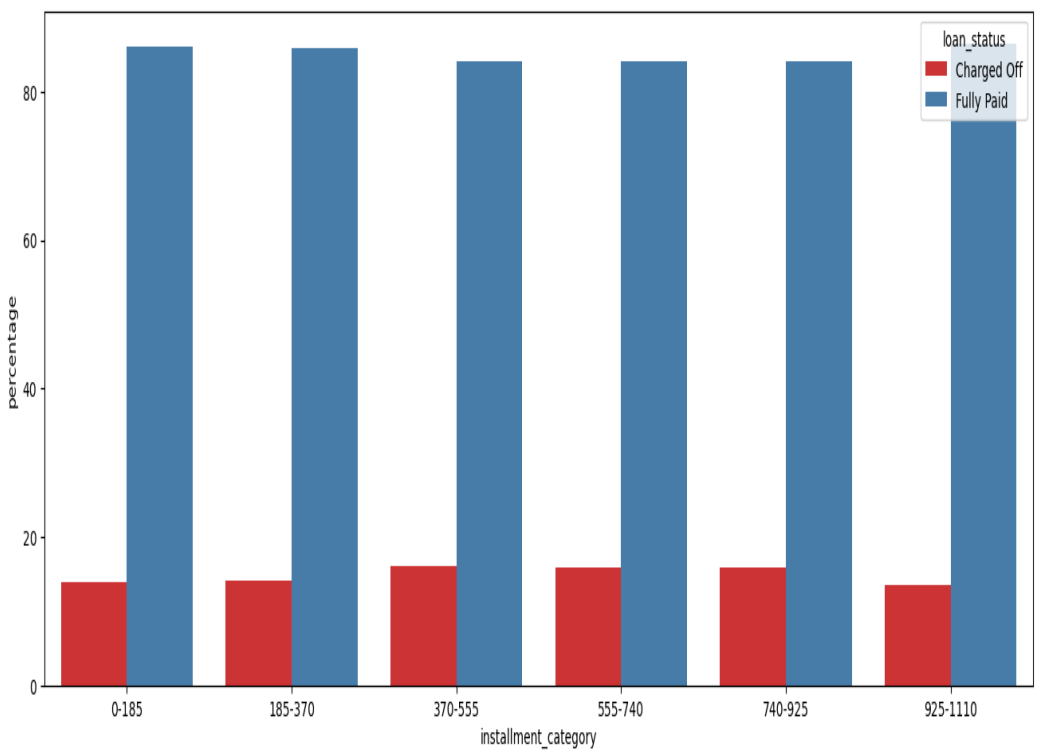
3.1.11) Analyzing int_rate:



Above graph shows that the highest percentage of loan defaults(Charged Off) are for loans where interest rate is high around 17%-24%

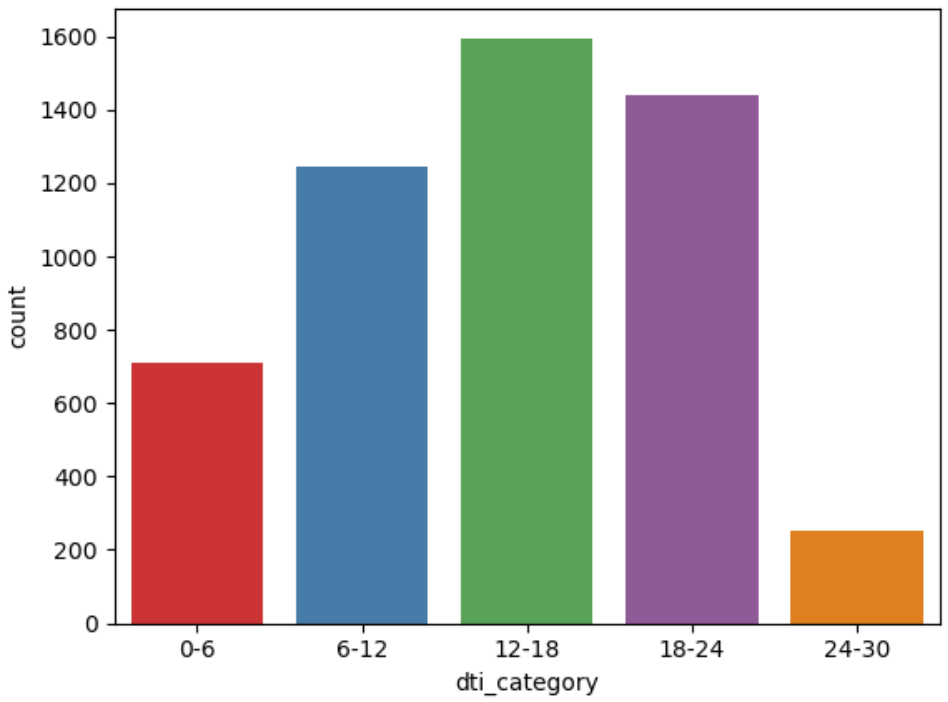
3.1.12) Analyzing installment:

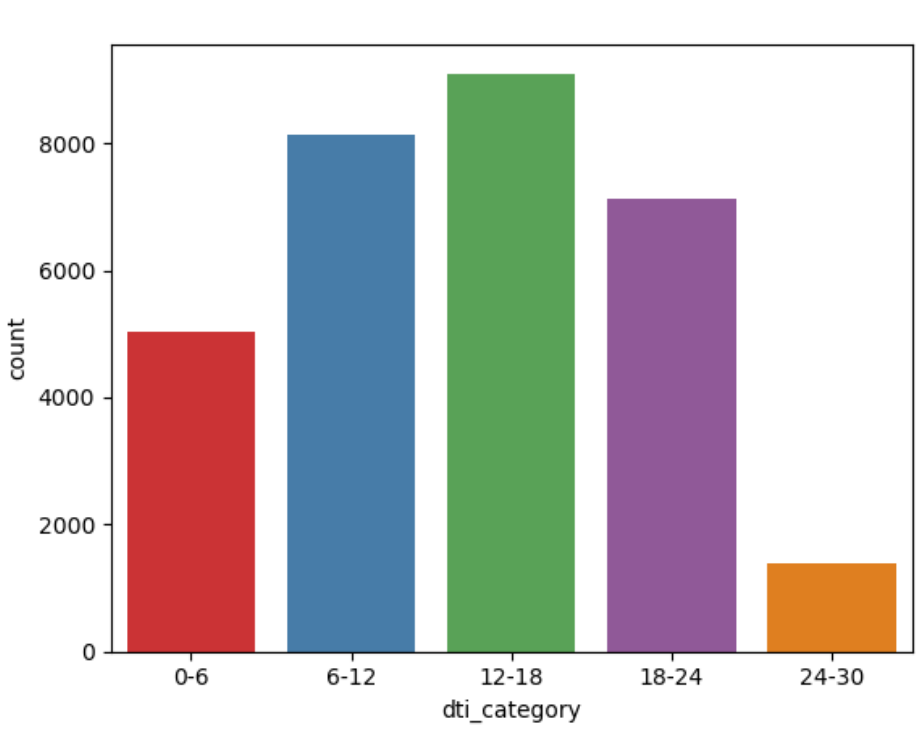


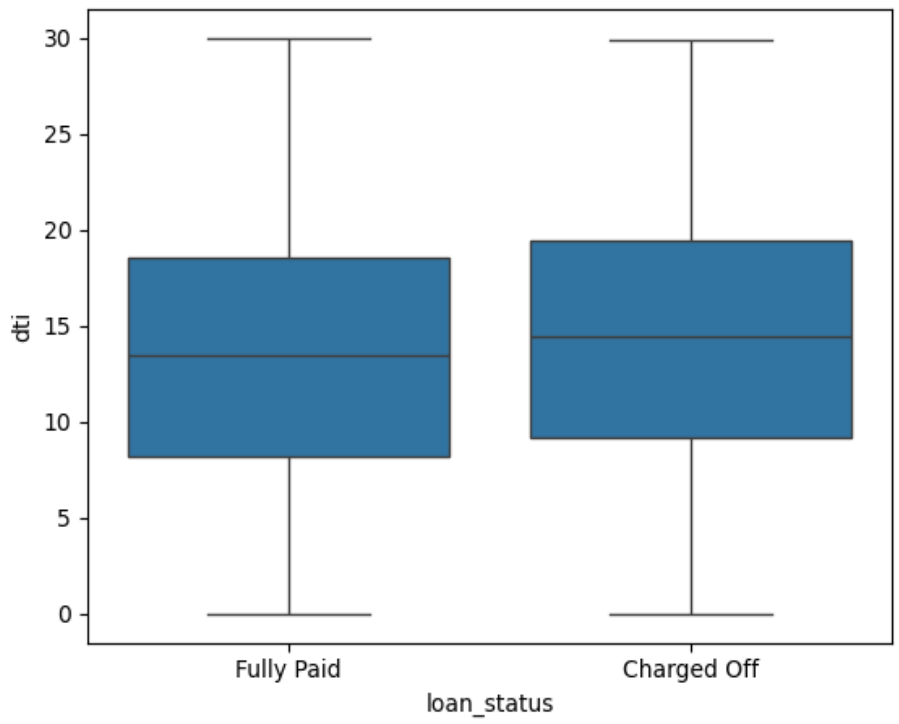


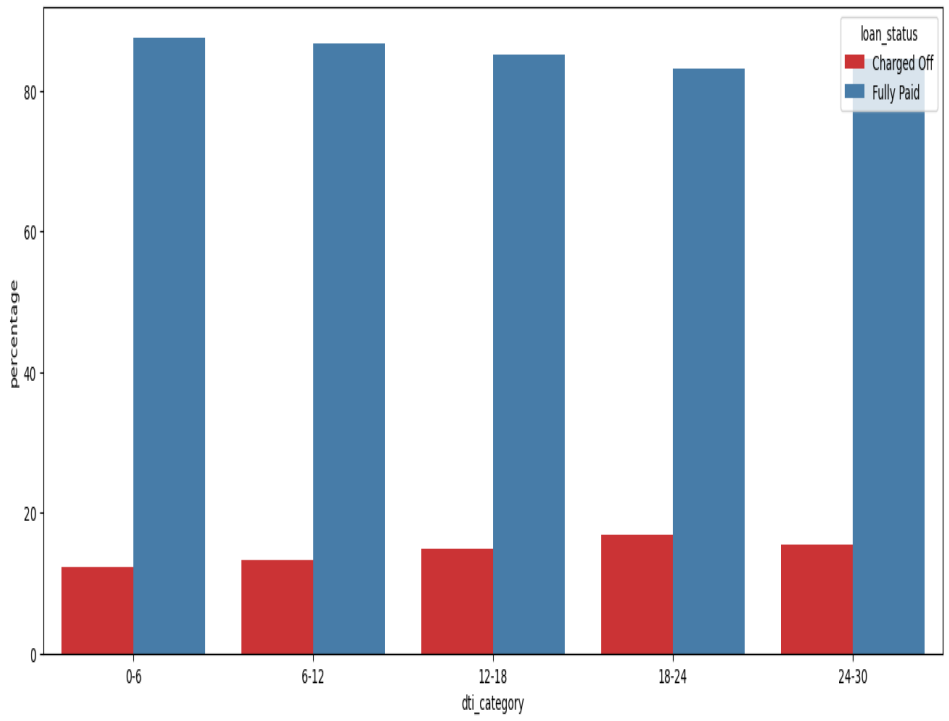
No Impact. Negligible differences

3.1.13) Analyzing dti:


1600
1400
1200
1000
800
600
400
200
0
count
0-6
6-12
12-18
18-24
24-30
dti_category

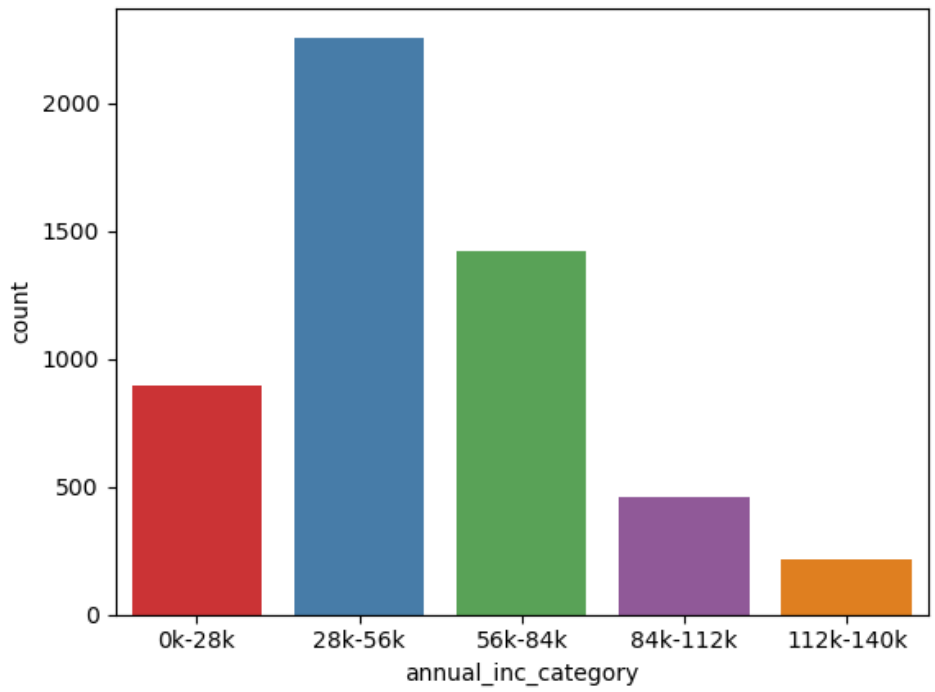

8000
6000
4000
2000
0
count
0-6
6-12
12-18
18-24
24-30
dti_category

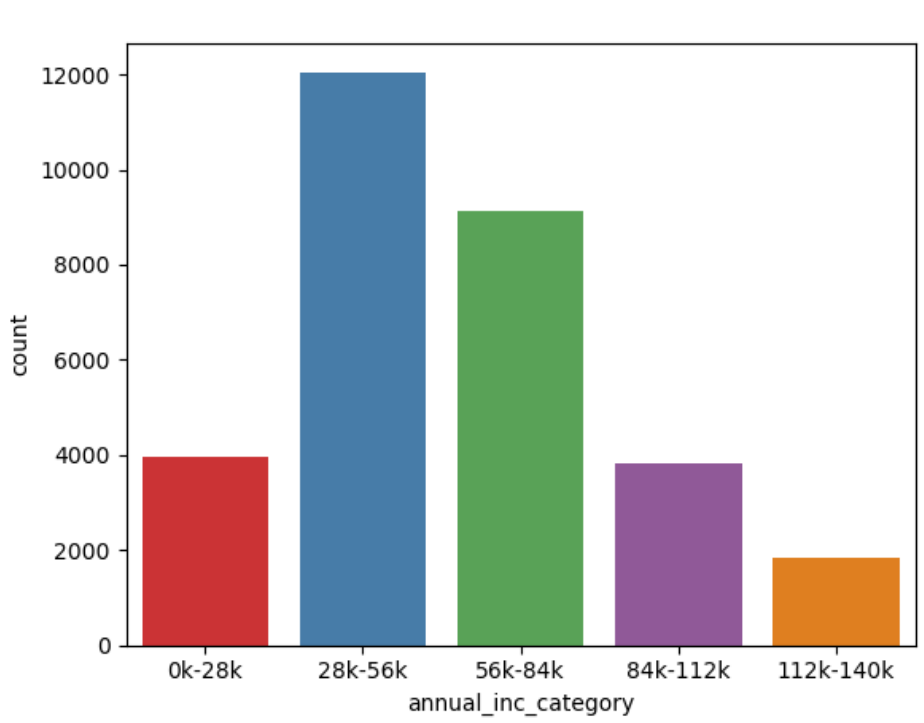


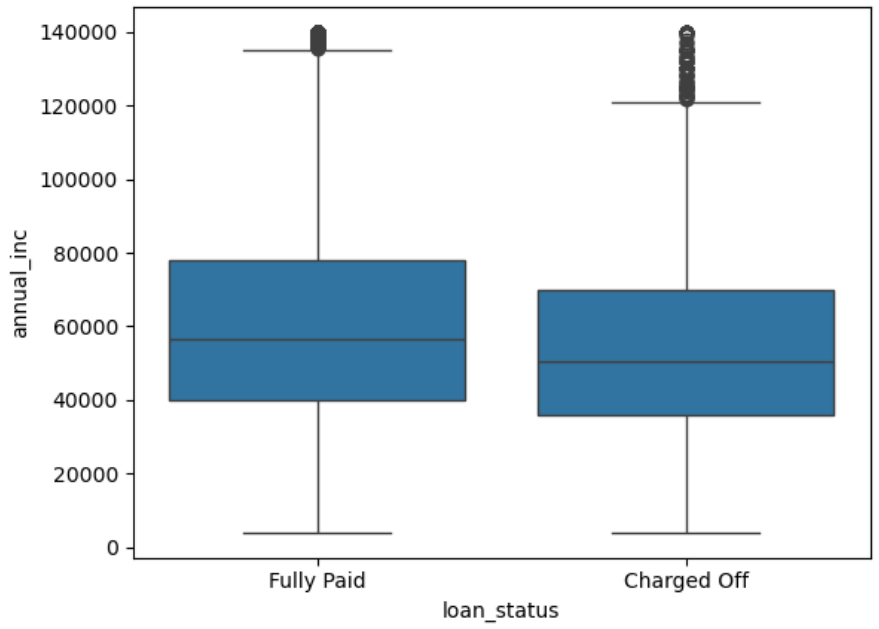


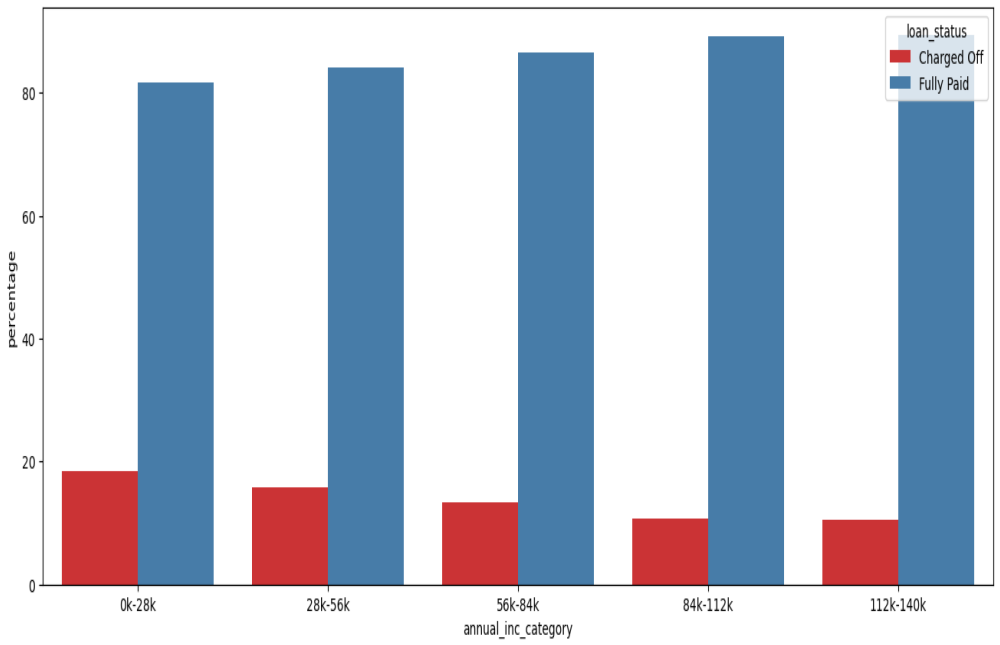
Not very significant differences but higher percentage of defaults(charged off) can be seen for dti ranging between 18-30. It can be seen that in general when dti increases the risk of loan being Charged Off increases.

3.1.14) Analyzing annual Income:


2000
1500
1000
500
0
count
0k-28k
28k-56k
56k-84k
84k-112k
112k-140k
annual_inc_category


12000
10000
8000
6000
4000
2000
0
count
0k-28k
28k-56k
56k-84k
84k-112k
112k-140k
annual_inc_category



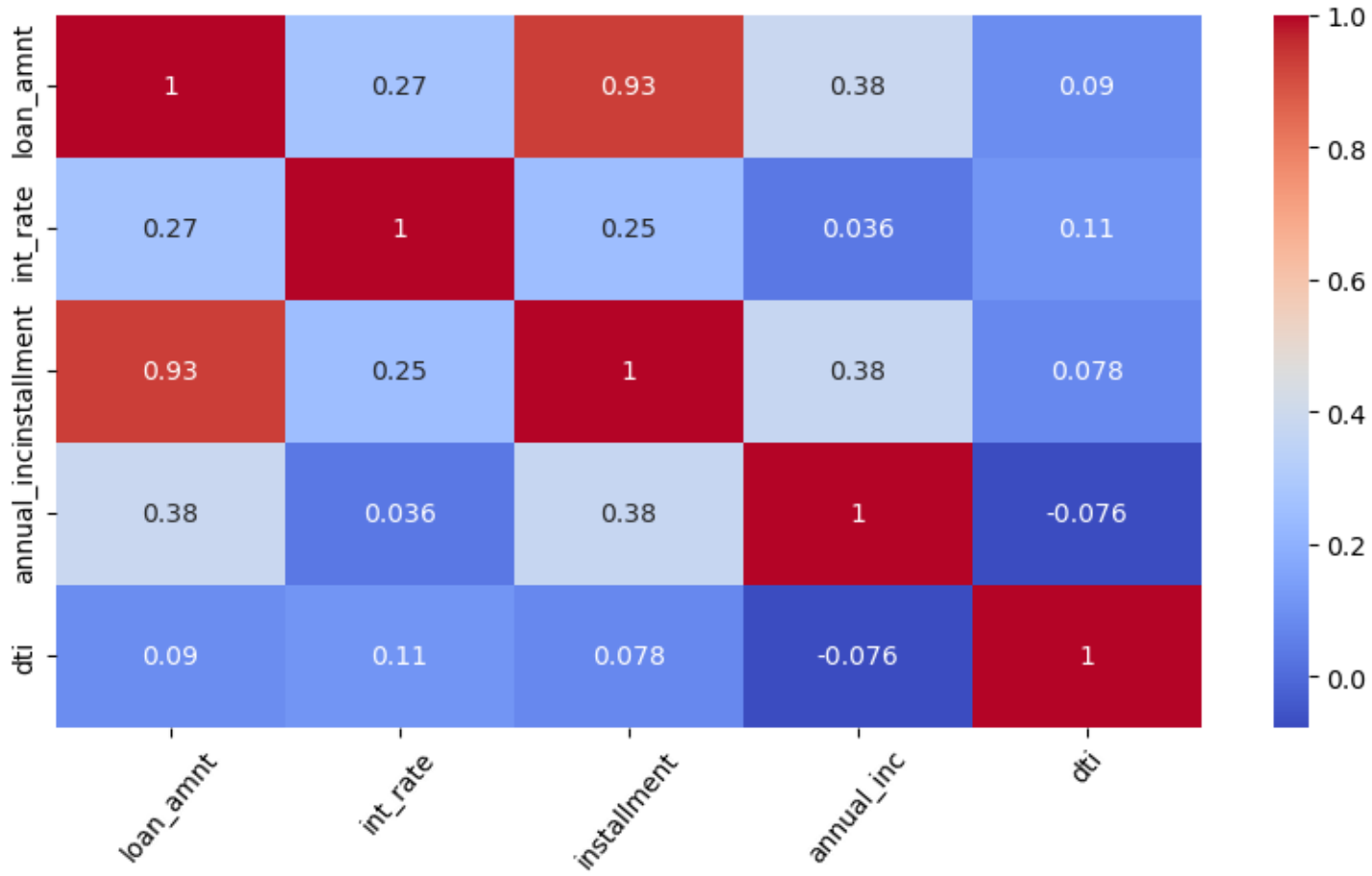


The people with low annual income have more default rate. The highest percentage of defaulters can be seen in annual income range from 0 to 28K and second highest default rate is in annual income range from 28k to 56k.

Observations from above Analysis Part 1: - using Loan_Status and one more variable

1. Among all grades, The highest percentage of defaulters (Charged Off) are in grade G (next highest in F and then in E).
2. 60 Months term shows higher percentage of Charged Off Loans, as compared to 36 months.
3. All emp_length values have almost equal percentage of defaulters though emp_length >=10 years have little higher percentage of defaulters. So emp_length does not seem to have much impact on risk of a loan being Charged Off.
4. Among all values of 'purpose', highest percentage of defaulters(Charged Off) are there for loans taken for 'small_business'.
5. Home_Ownership has almost no significant impact. Though Home_ownership = OTHER has slightly higher percentage of defaulters (Charged Off loans) as compared to other values of home_ownership.
6. Different verification_status values shows almost negligible differences in percentages of defaulters, though verification_status = 'verified' shows (slightly higher than others) highest percentage of defaulted loans as compared to other values of verification_status. So this will have little impact on the risk calculation of a loan.
7. As the loan amount increases the possibility of default(Charged Off loan) increases. The loan_amnt_category with highest loan seems to have highest percentage of loan default(Charged Off). 18k-24k, 24k-30k are two ranges of loan_amnt with highest possibility of a loan being Charged Off as compared to other lower ranges of loan_amnt.
8. Among various interest Rate values, the highest percentage of loan defaults(Charged Off) are for loans where interest rate is high around 17%-24%.
9. Installments do not have much impact on the risk of a loan being charged off, though installments ranging from 370-975 have slightly higher possibility of being charged off as compared to other installment categories.
10. Among all the dti values, the higher percentage of defaults(charged off) can be seen for dti ranging between 18-30. It can be seen that in general when dti increases the risk of loan being Charged Off increases.
11. The people with low annual income have more default rate. The highest percentage of defaulters can be seen in annual income range from 0 to 28K and second highest default rate is in annual income range from 28k to 56k.
12. All sub grades show equal percentages of Charged Off and Fully Paid loans. Sub grades individually does not have much impact on possibility of default

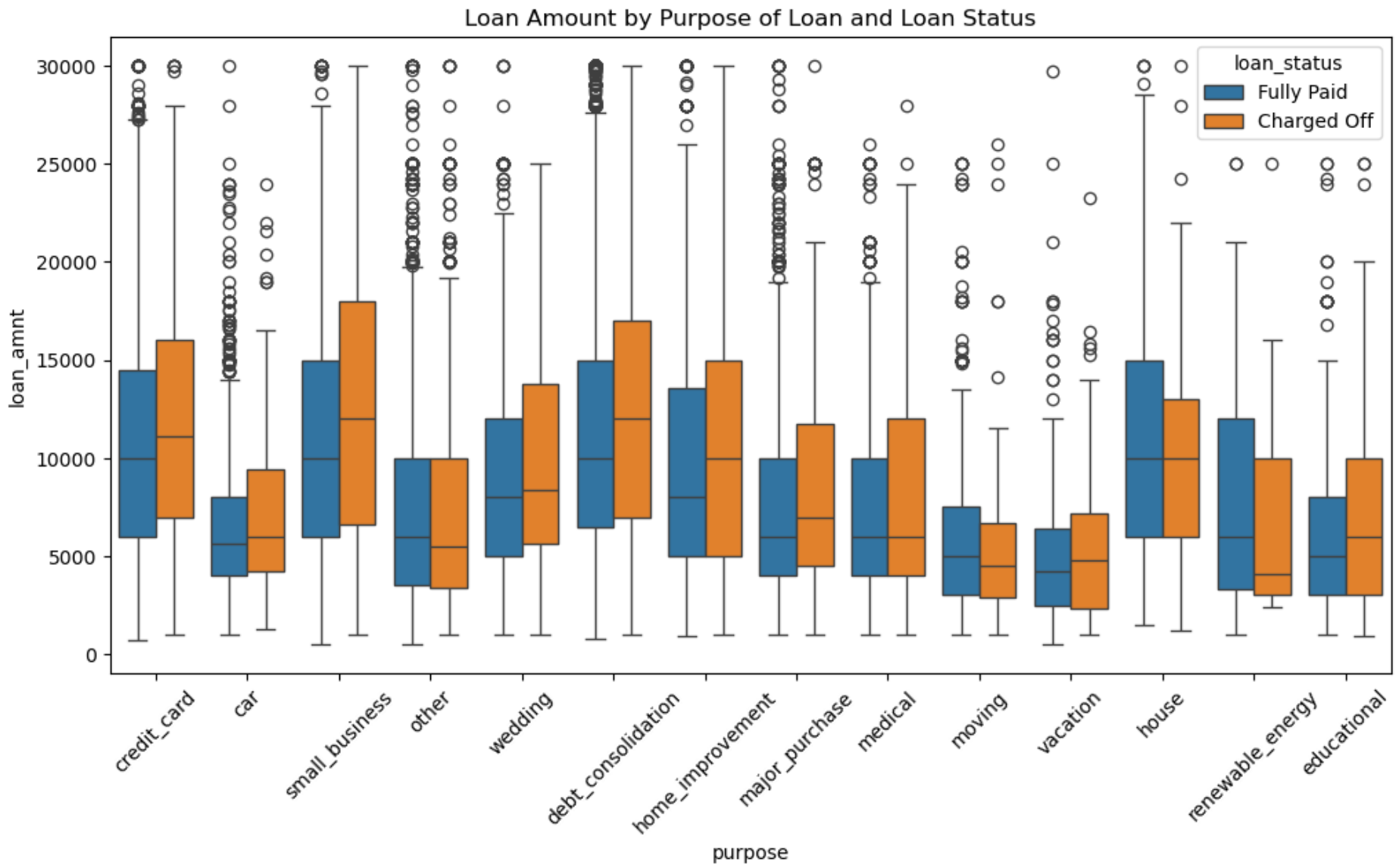
Understanding the correlation between numerical variables

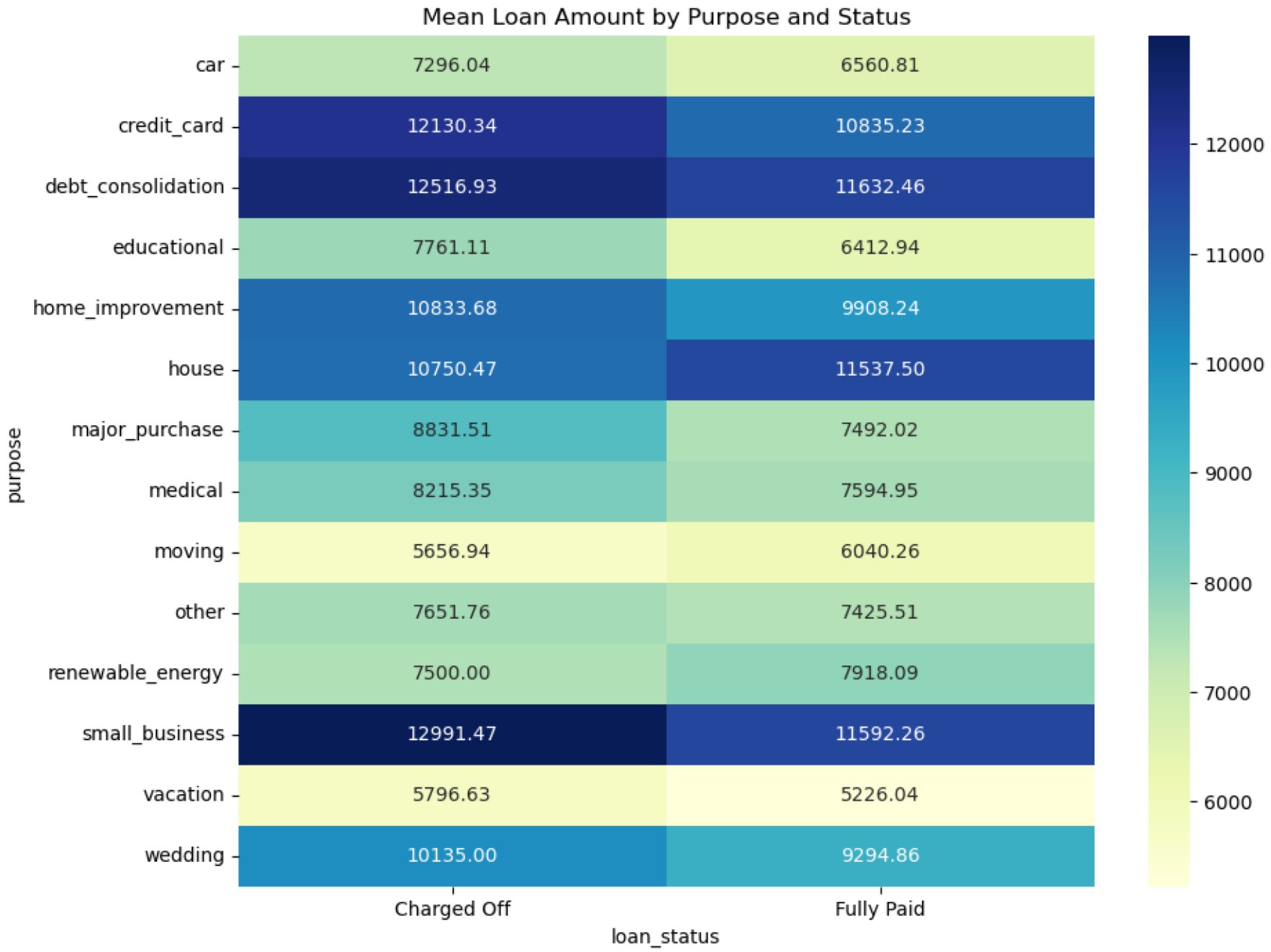


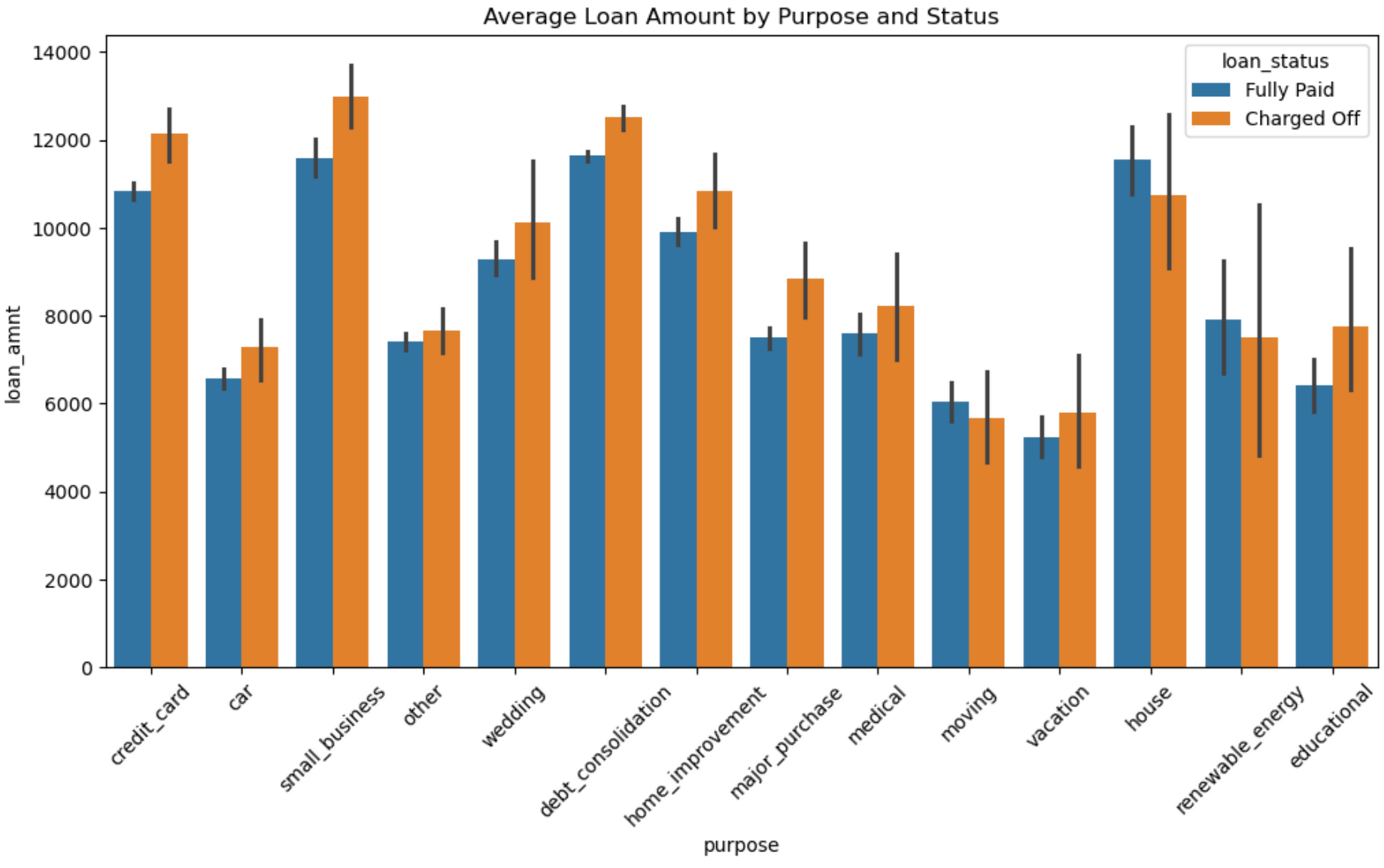
3.2. Data Analysis Part 2 - using loan_status and two more column

Analyzing loan_amnt with other columns for more insights

3.2.1) Loan amount vs Purpose over loan_status

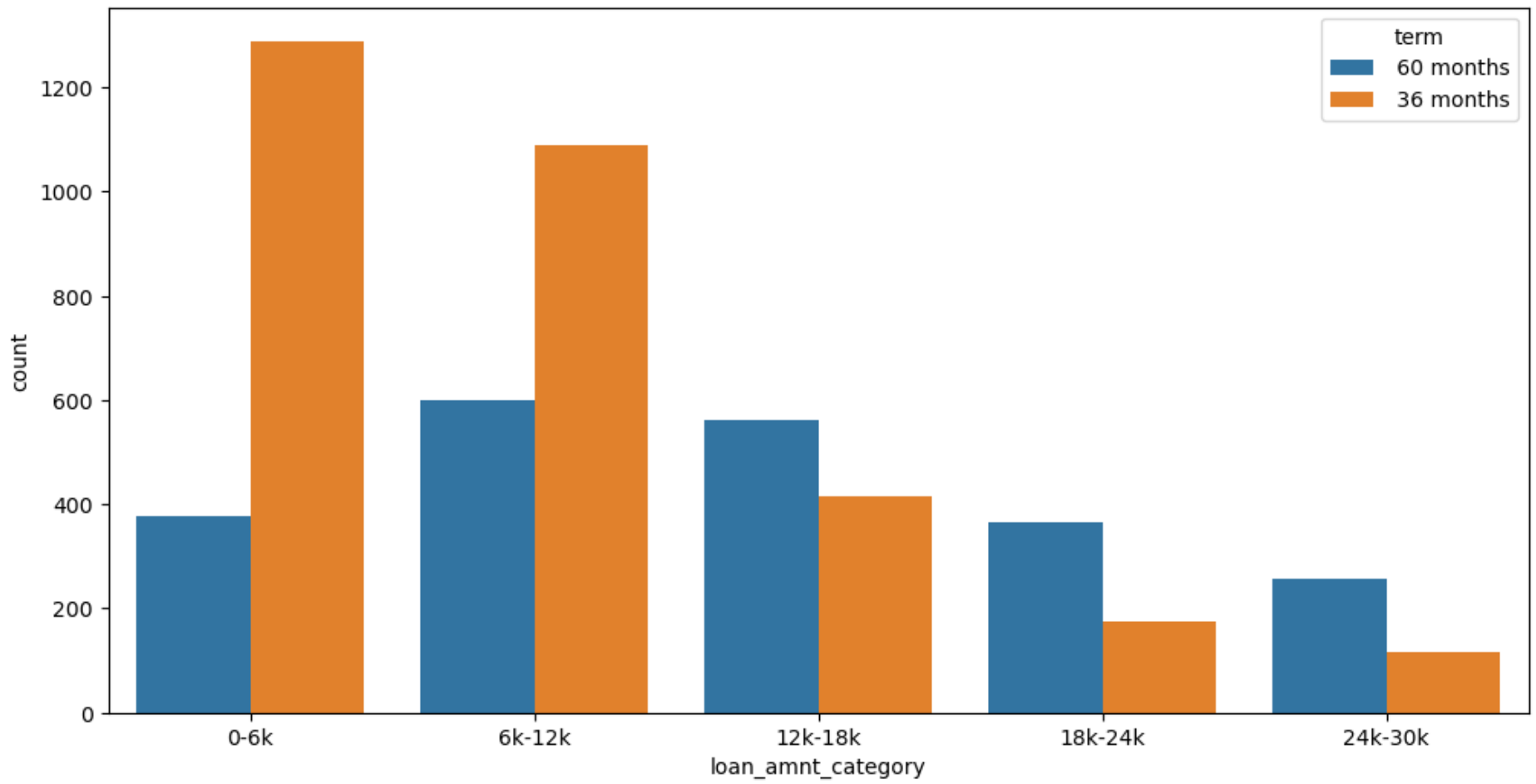


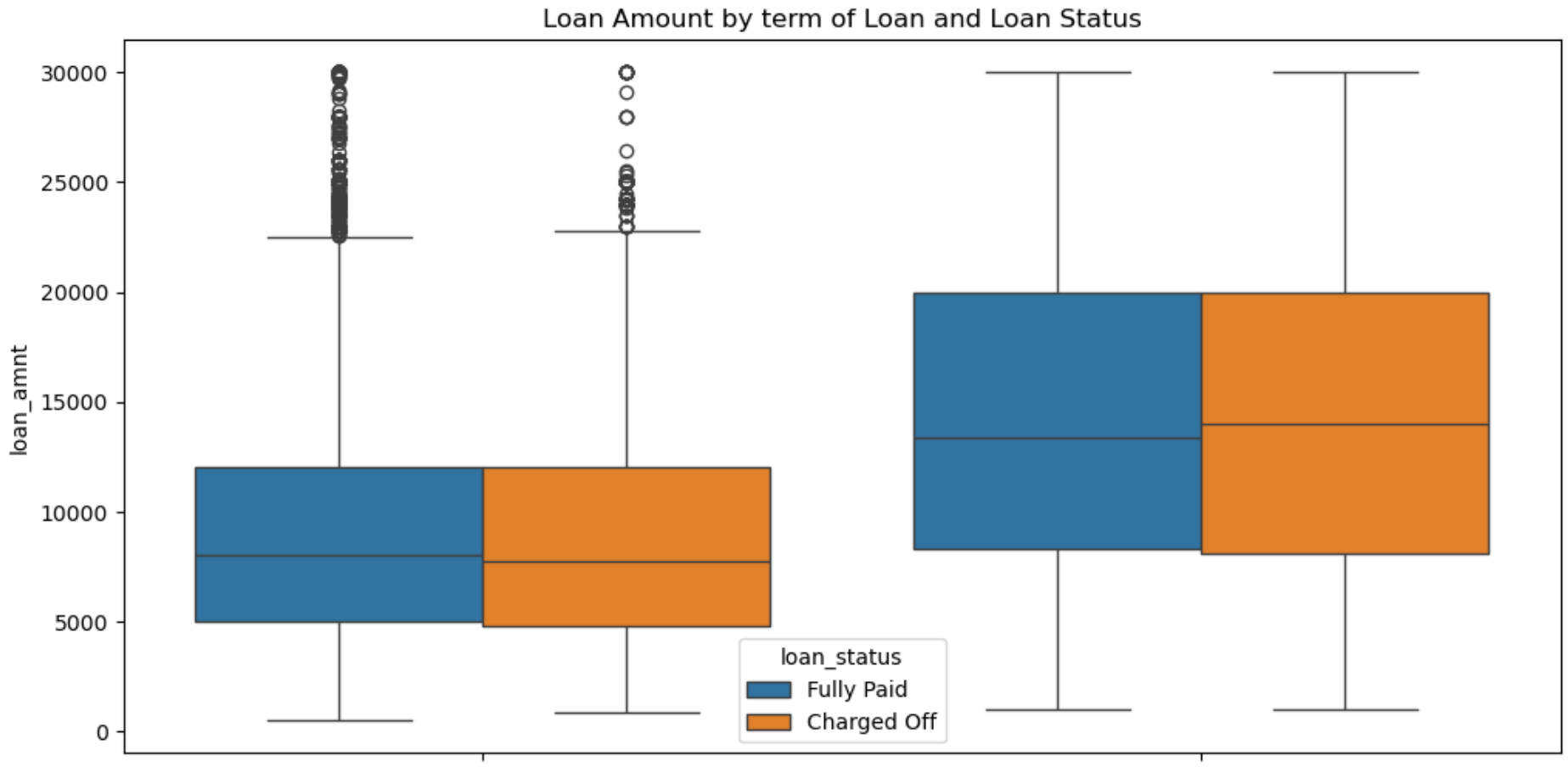


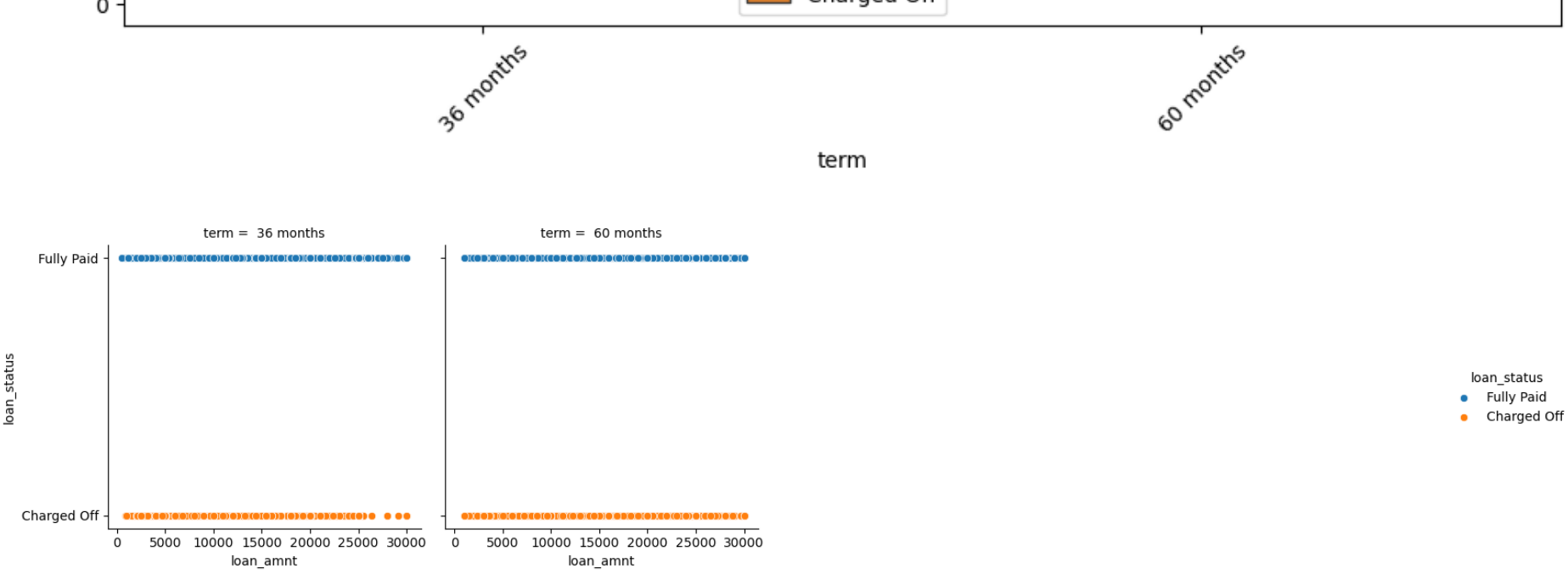


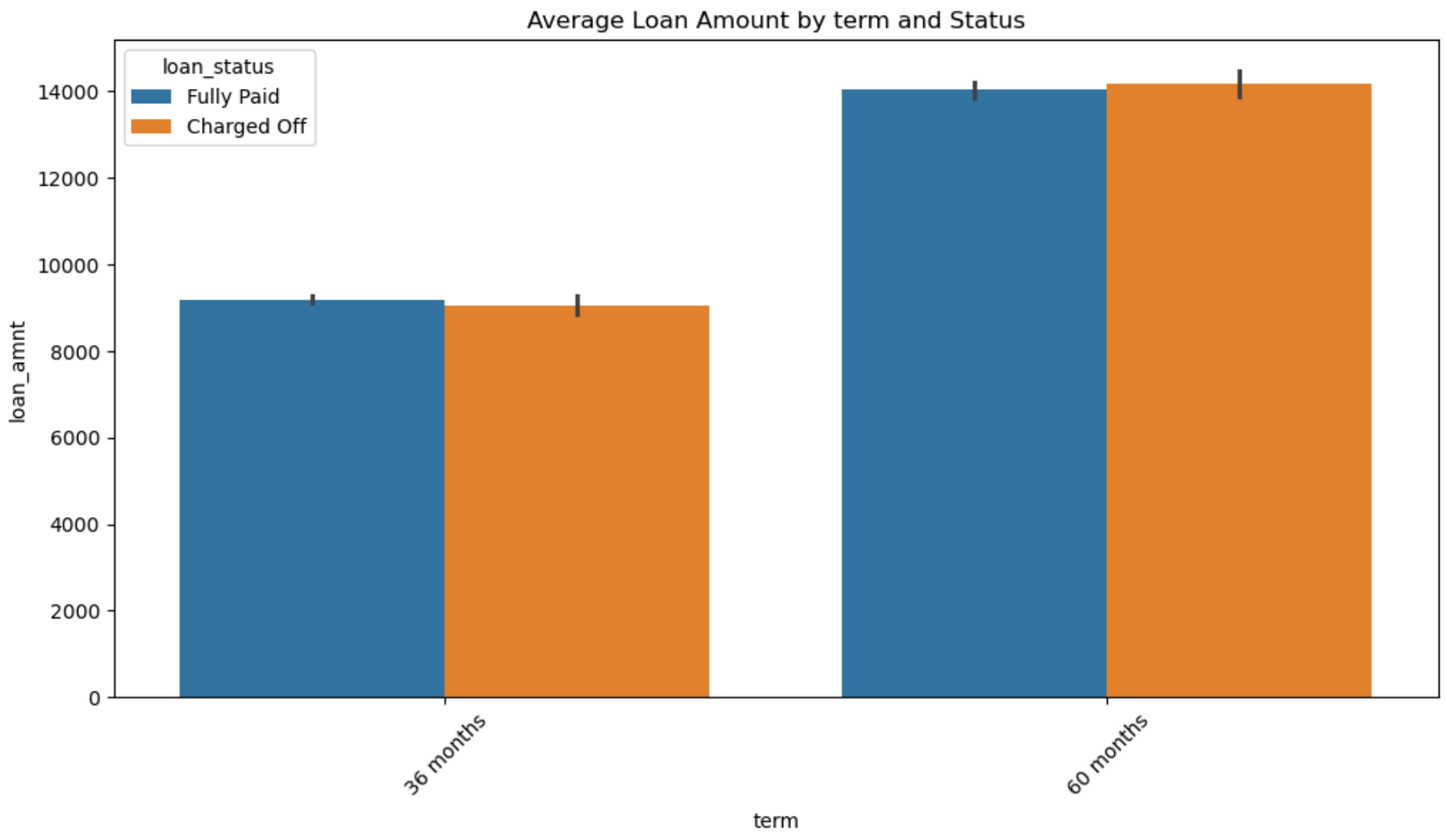
This again confirms that 'Small_Business' have the highest average loan amount and are most risky type of loan.

3.2.2) Loan amount vs term over loan_status



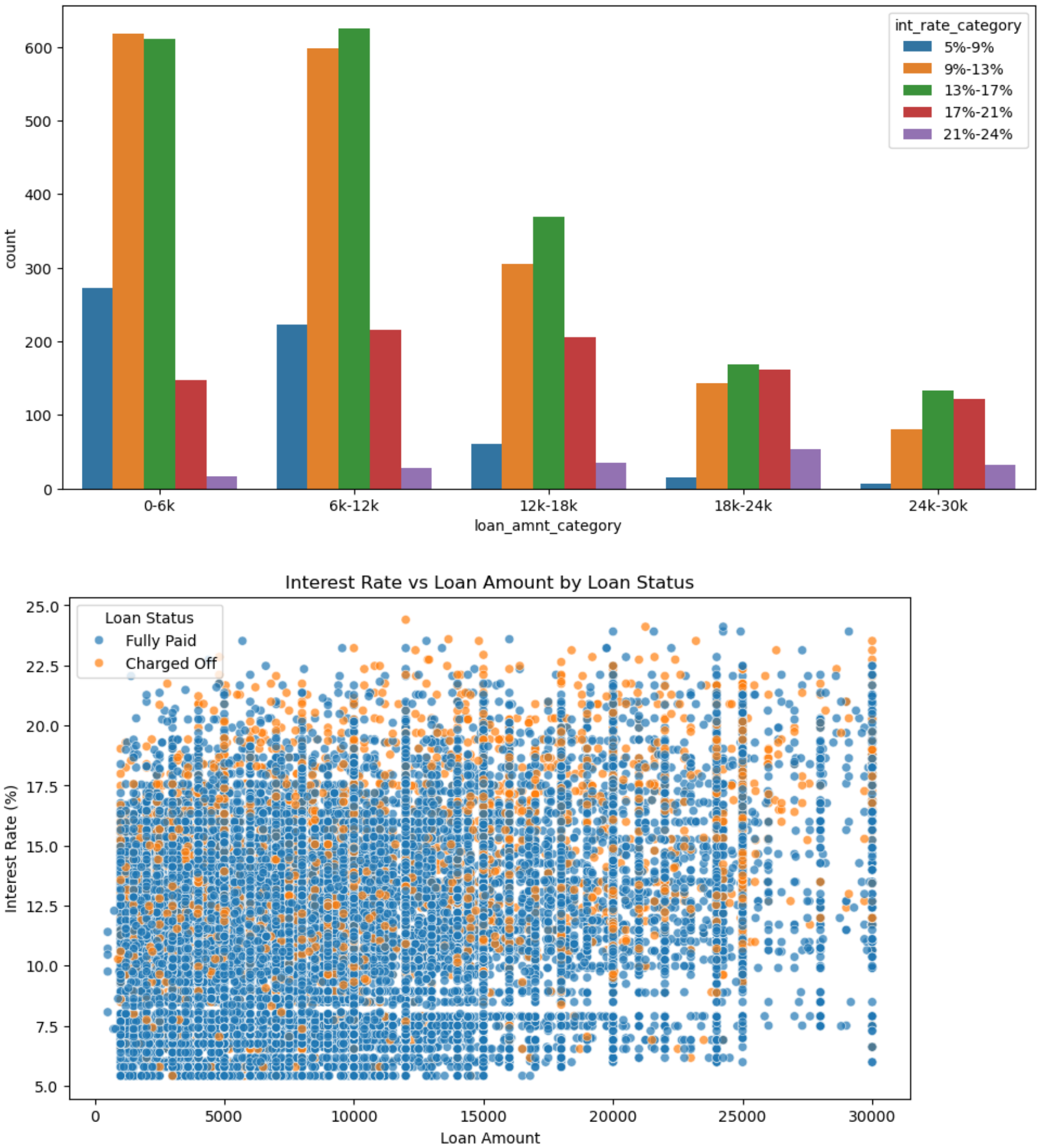


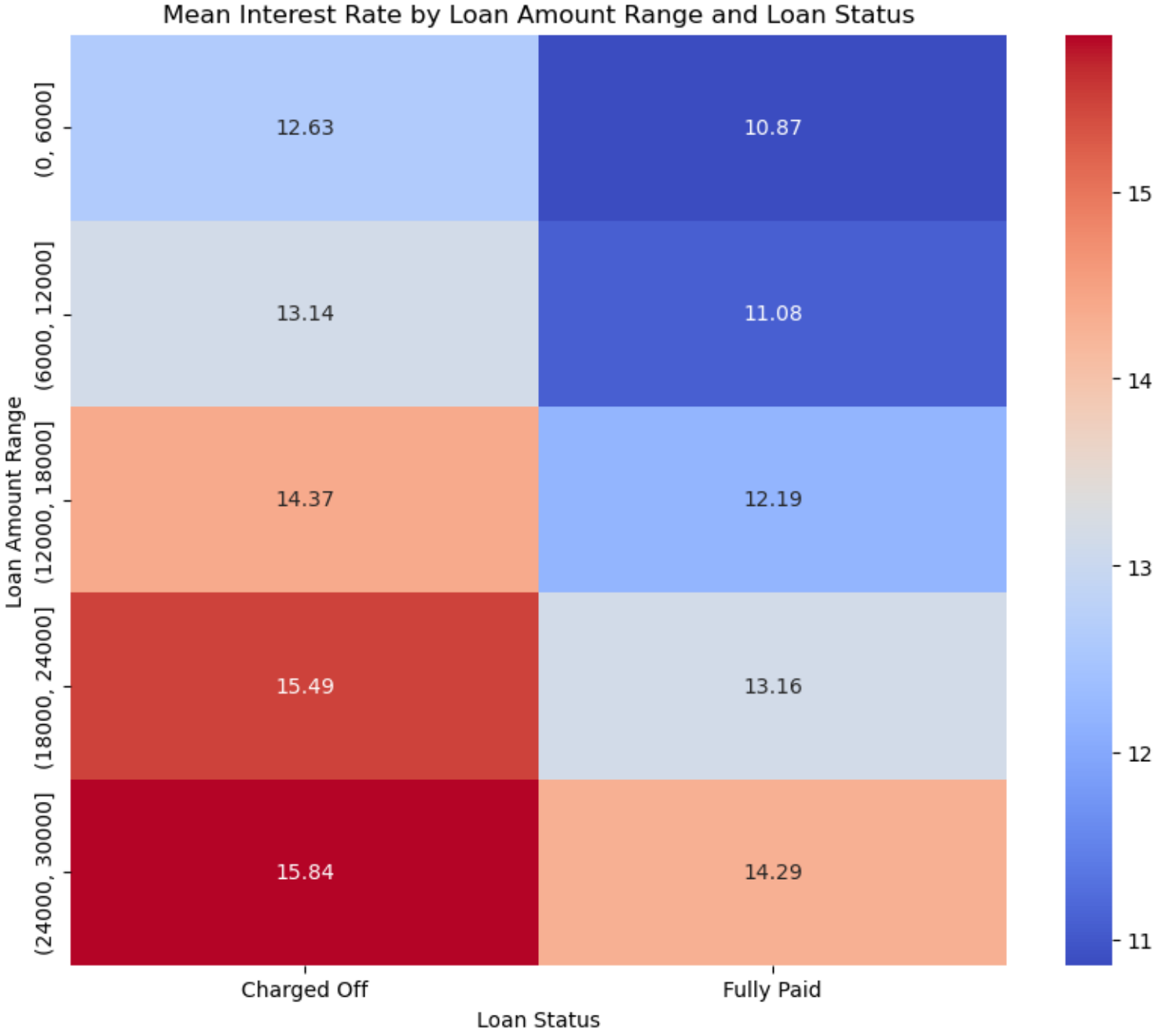




60 Months term have higher mean loan amount. This confirms our two observations that higher loan amount have higher risk of being Charged Off (as we know by previous analysis that 60 months term have higher percentage of Charged Off loans as compared to 36 months term) and 60 Months term is associated with higher risk of being Charged Off.

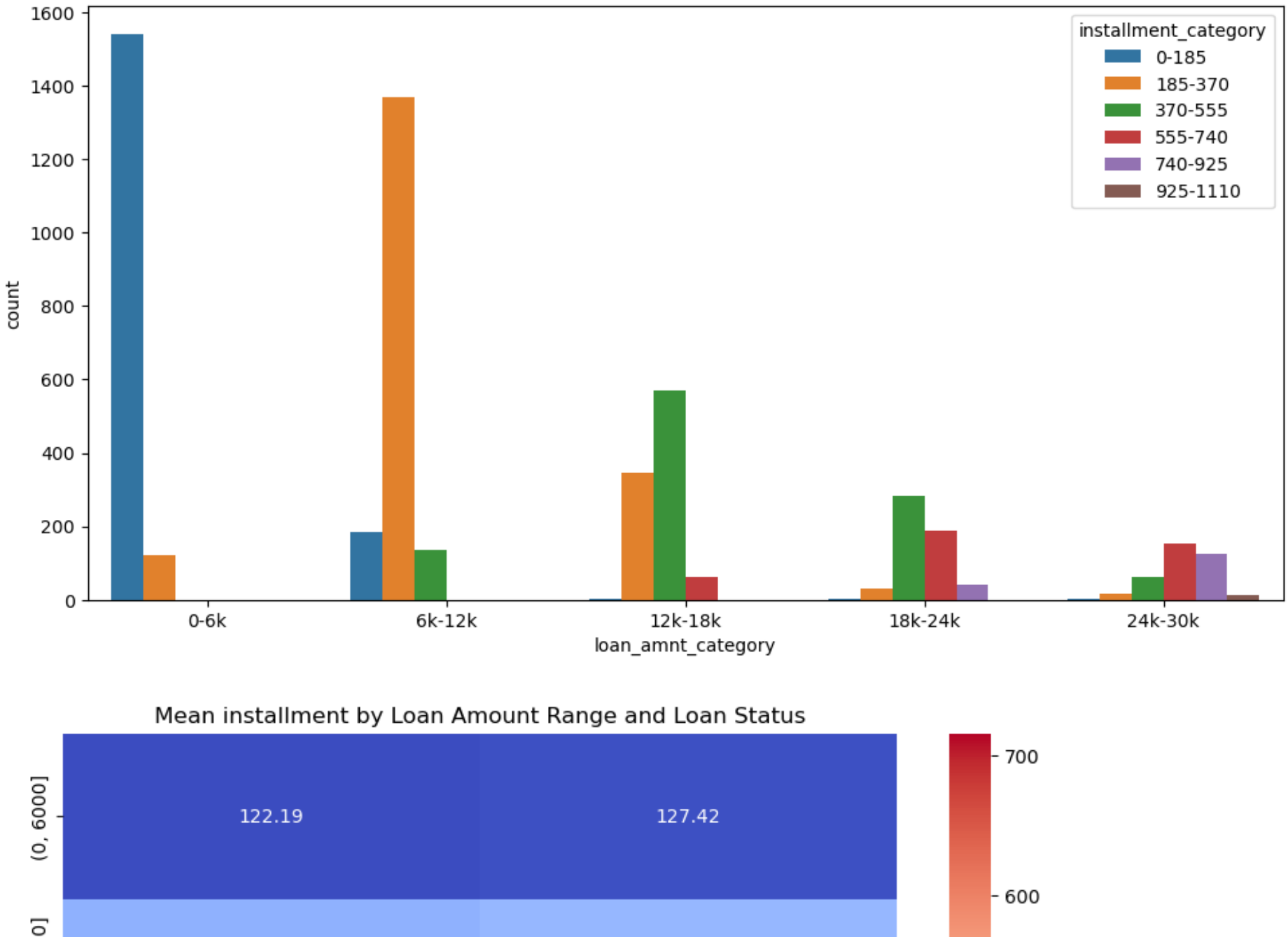
3.2.3) Loan amount vs interest rate over loan_status

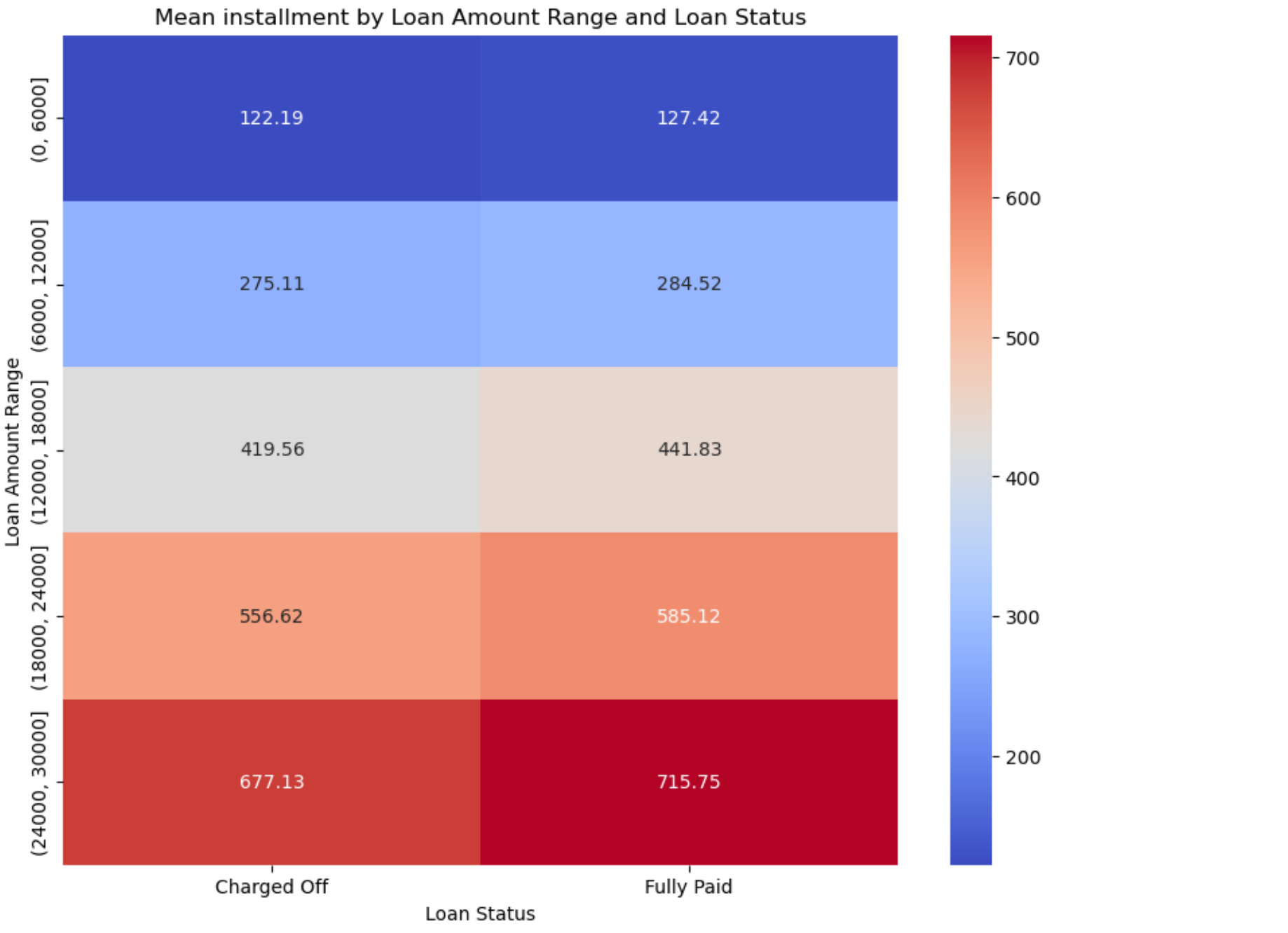




This shows that higher loan amounts have higher interest rates. And from our previous analysis it was seen that both of these factors are associated with high risk of loan being Charged Off.

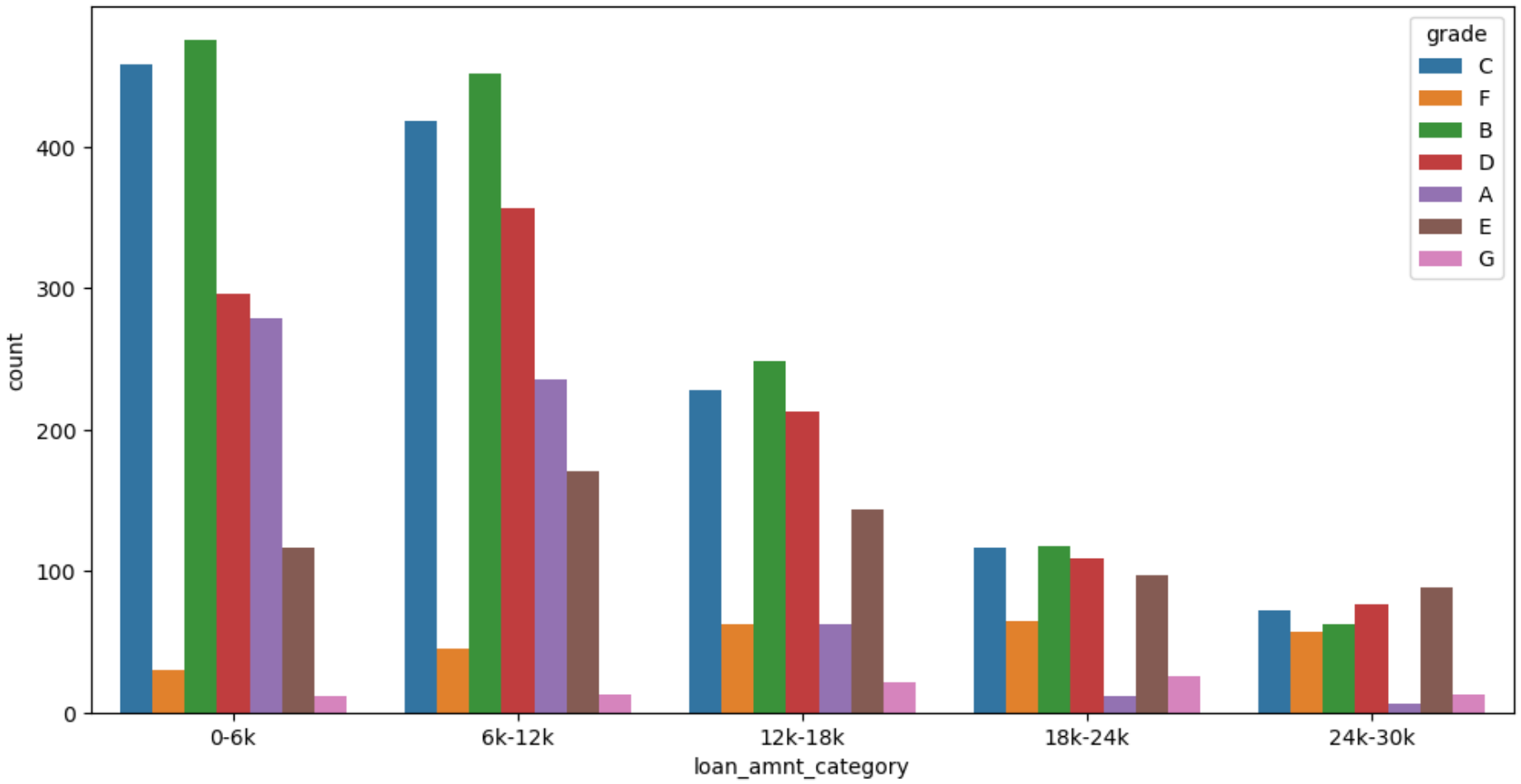
3.2.4) Loan amount vs installments over loan_status

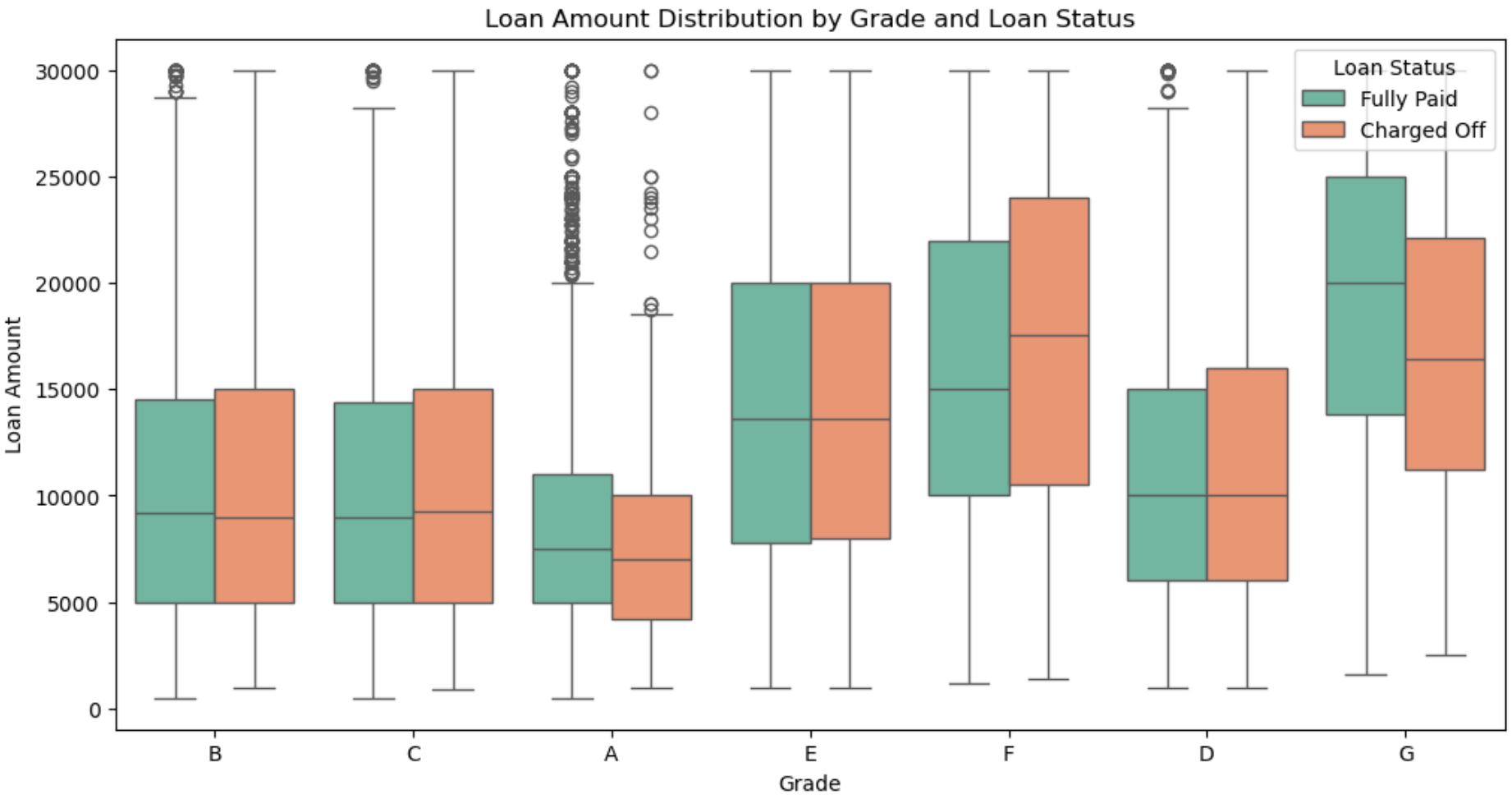


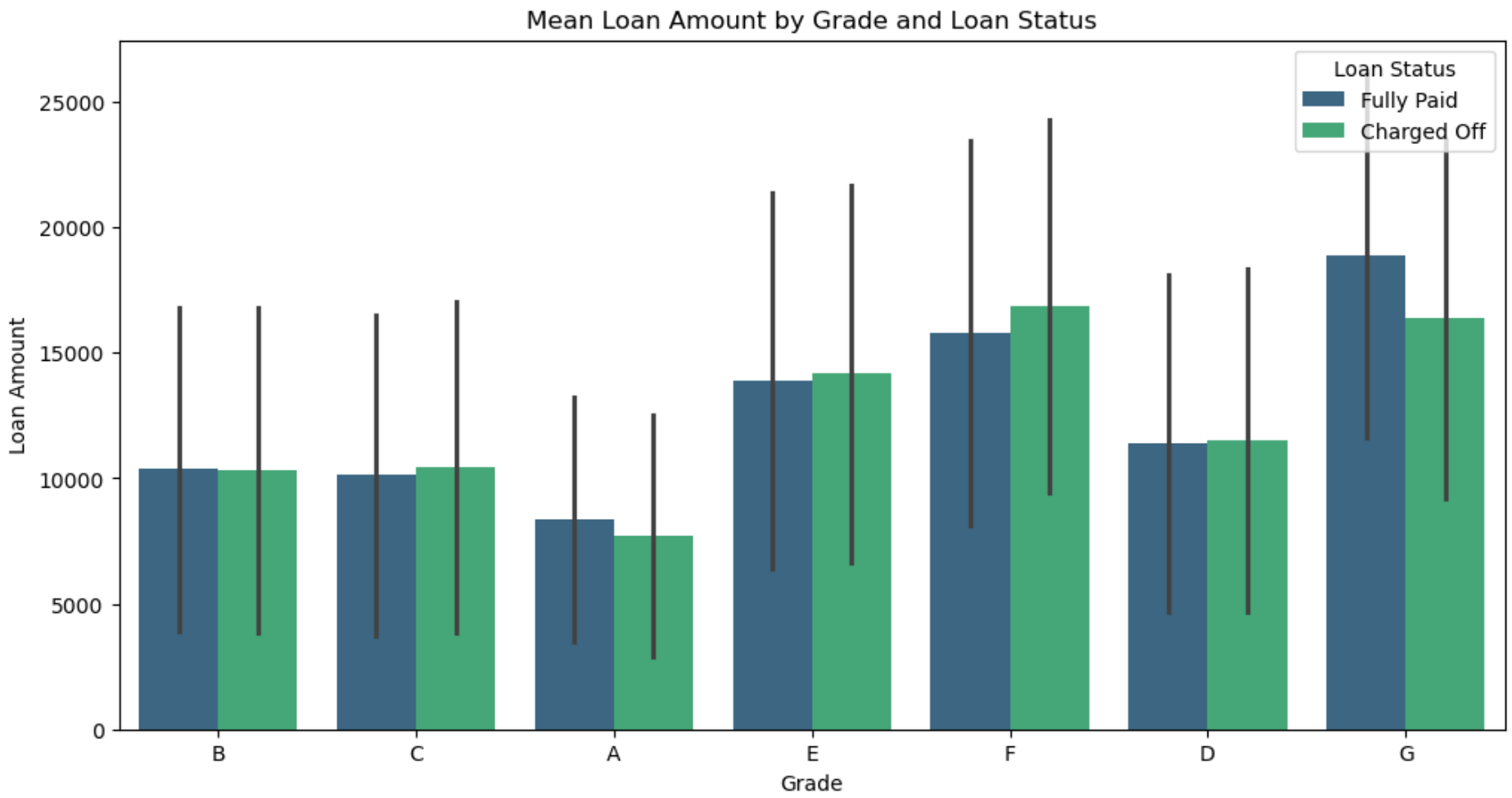


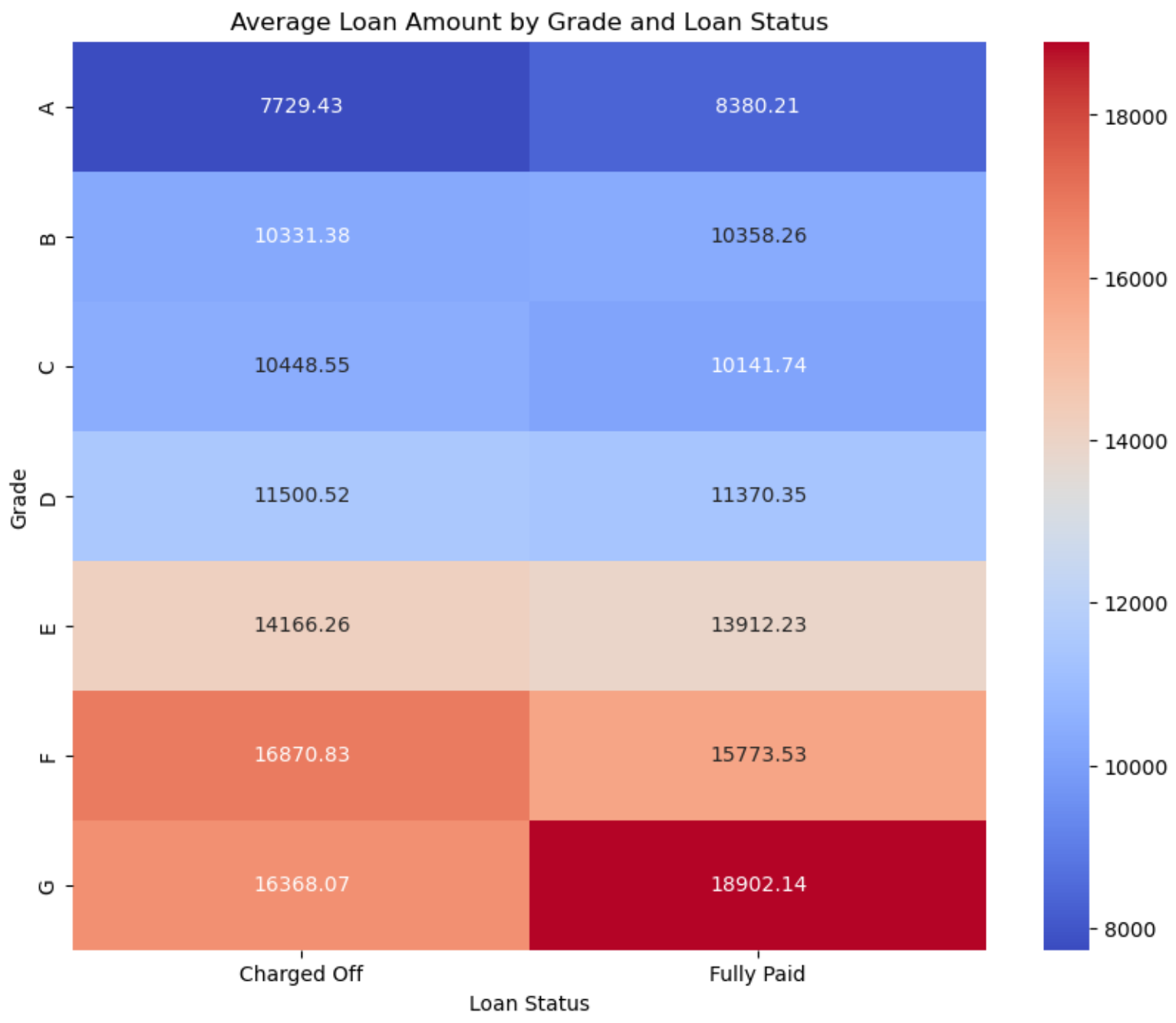
This shows that the mean installment increases as the loan amount increases.

3.2.5) Loan amount vs Grade over loan_status


grade
C
F
B
D
A
E
G
count
0
100
200
300
400
0-6k
6k-12k
12k-18k
18k-24k
24k-30k
loan_amnt_category


Loan Amount Distribution by Grade and Loan Status
Loan Status
Fully Paid
Charged Off
Loan Amount
0
5000
10000
15000
20000
25000
30000
B
C
A
E
F
D
G
Grade



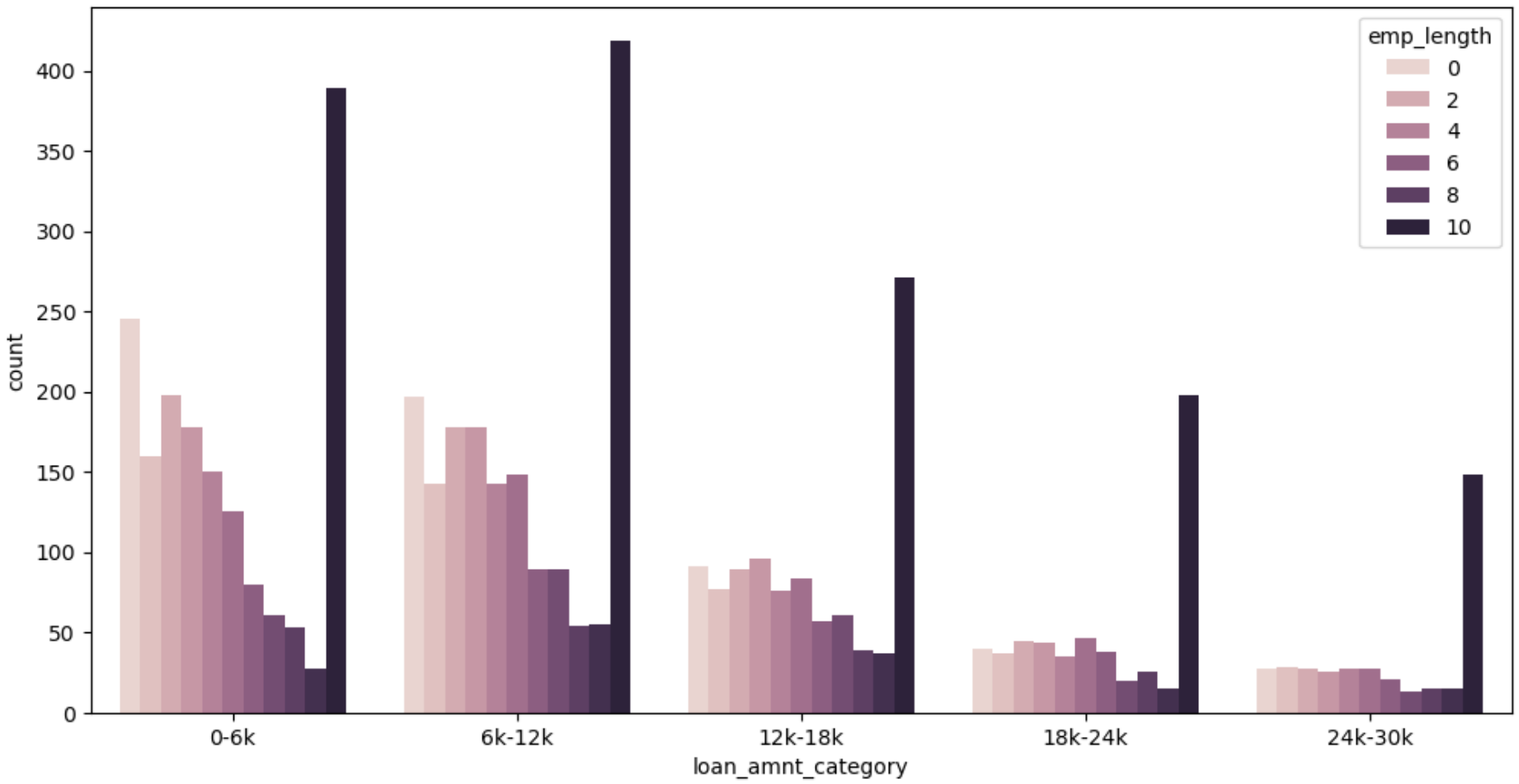


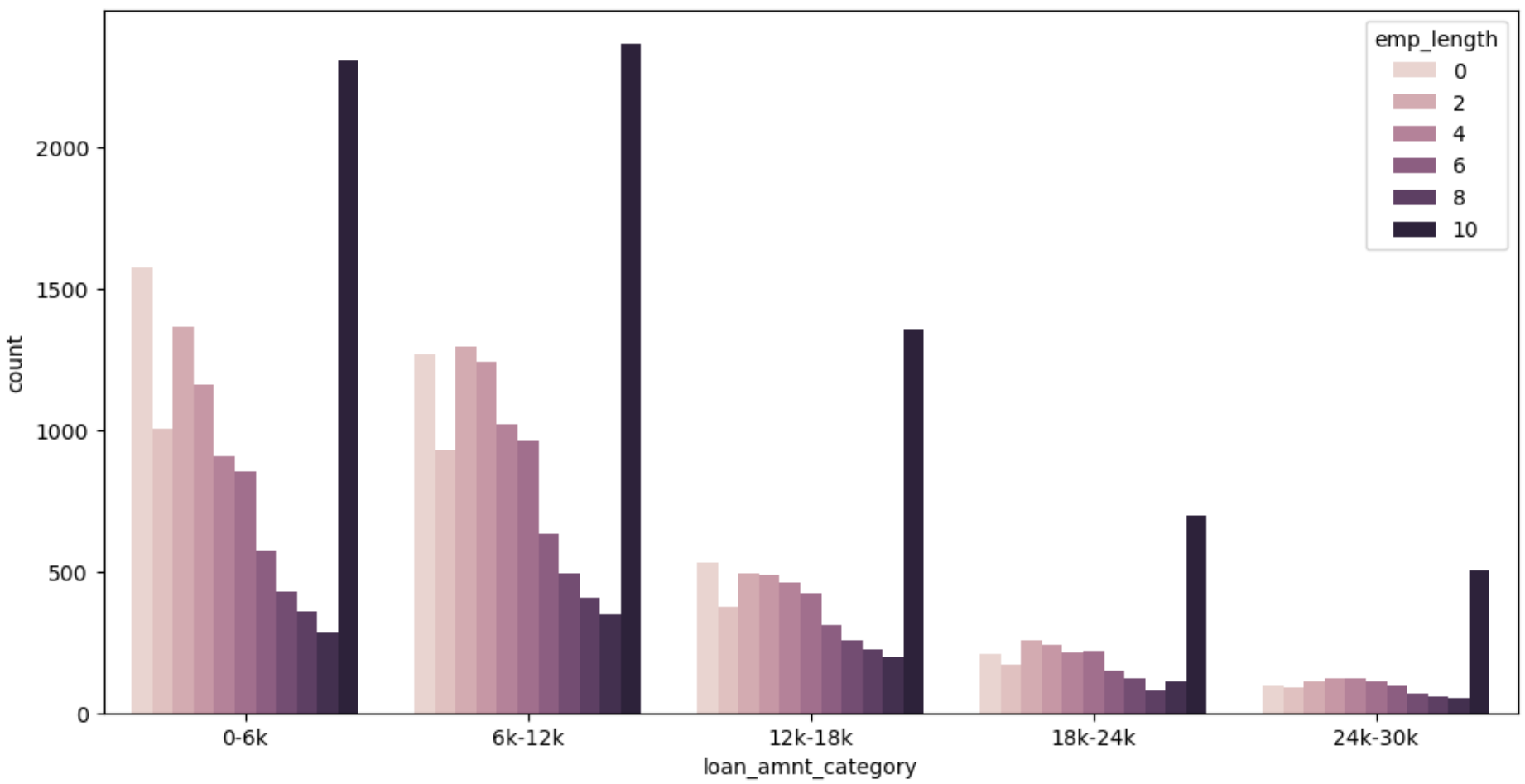
Grades E,F,G has higher average loan amounts in the Charged Off segment, as we have seen in previous analysis that these three also have highest percentage of

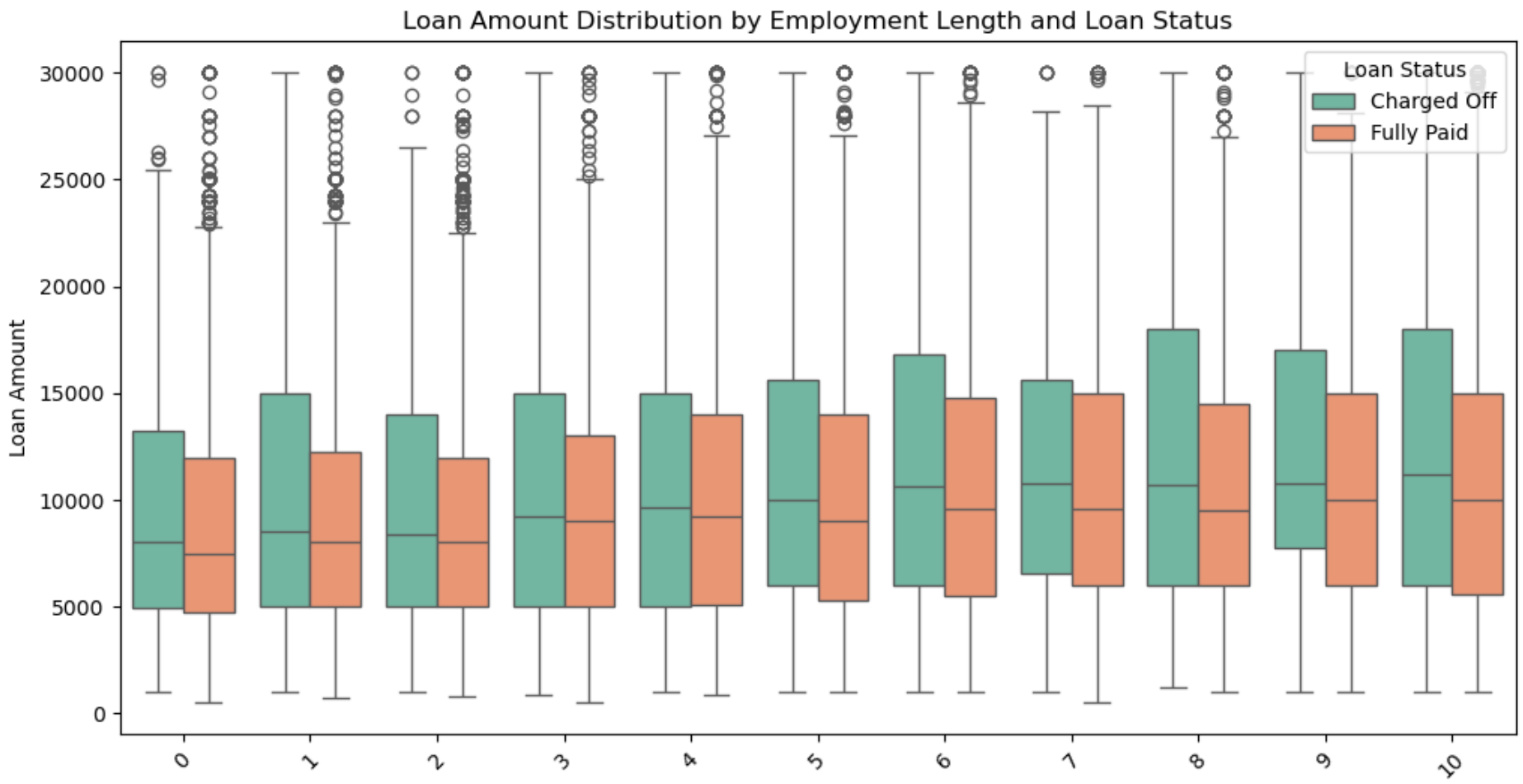
Charged Off loans as compared to other grades. Thus this confirms our findings. It also confirms a strong link between higher

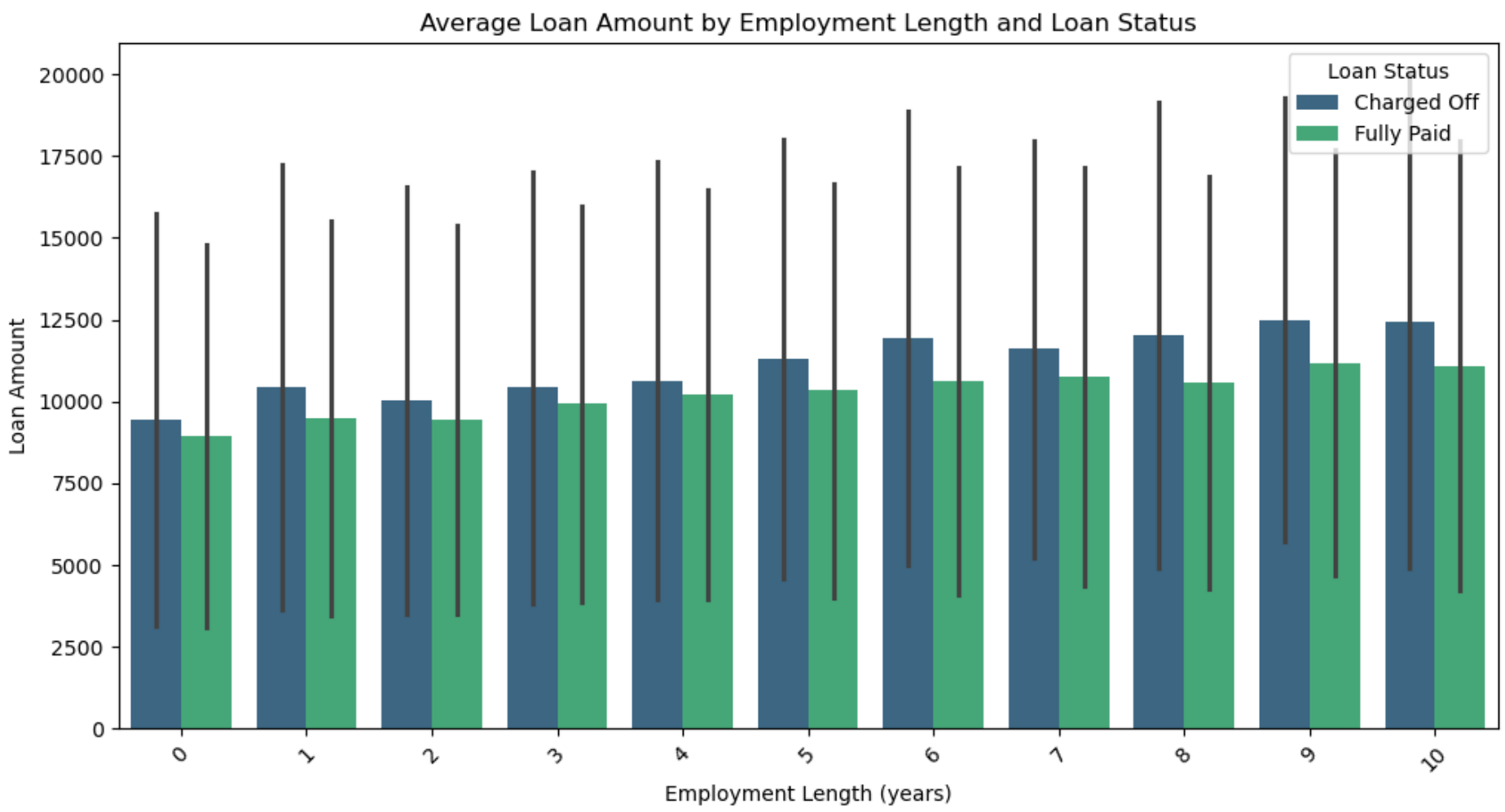
loan amounts being more risky and prone to being Charged Off.

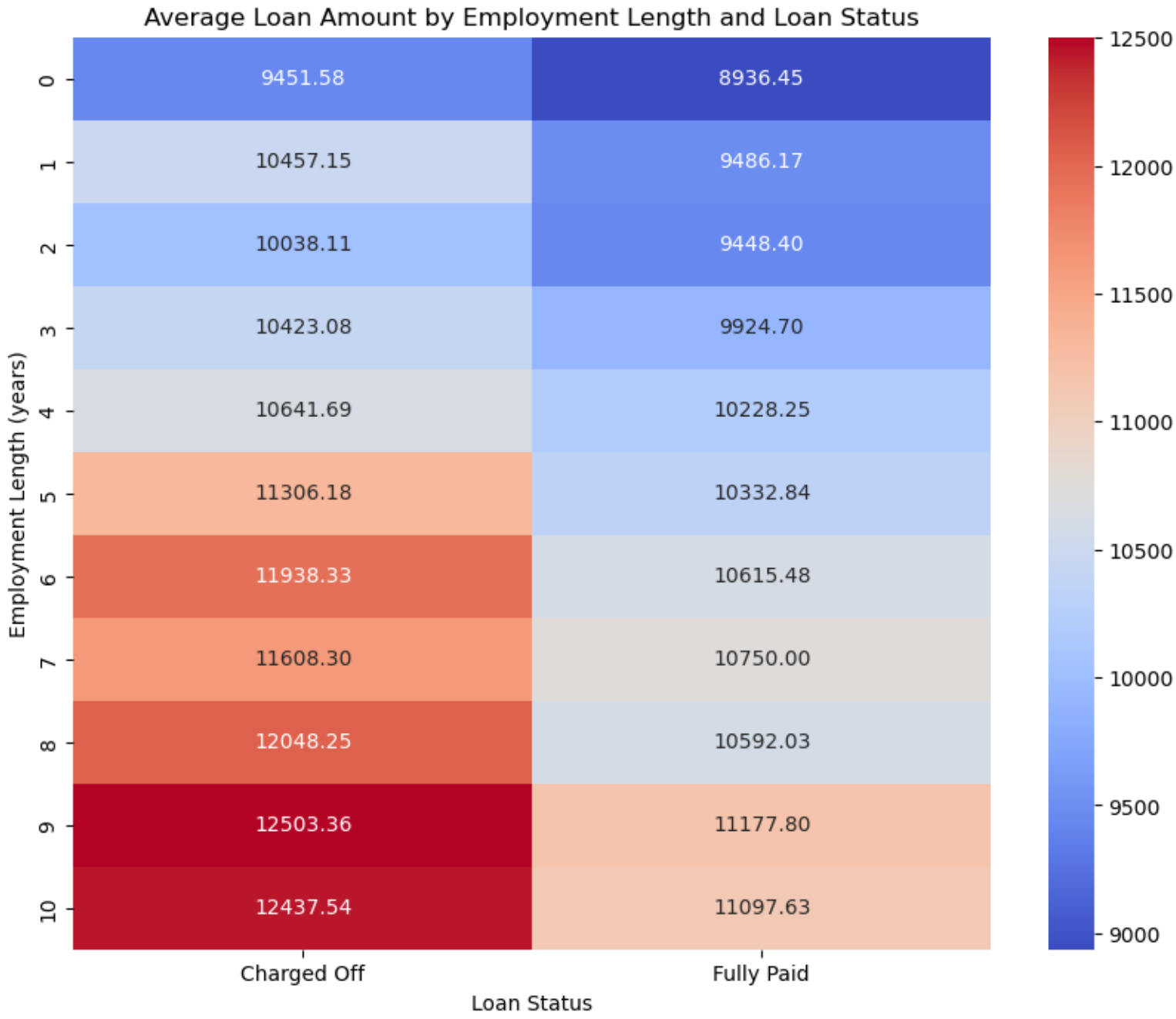
3.2.6) Loan amount vs Employee Length over loan_status





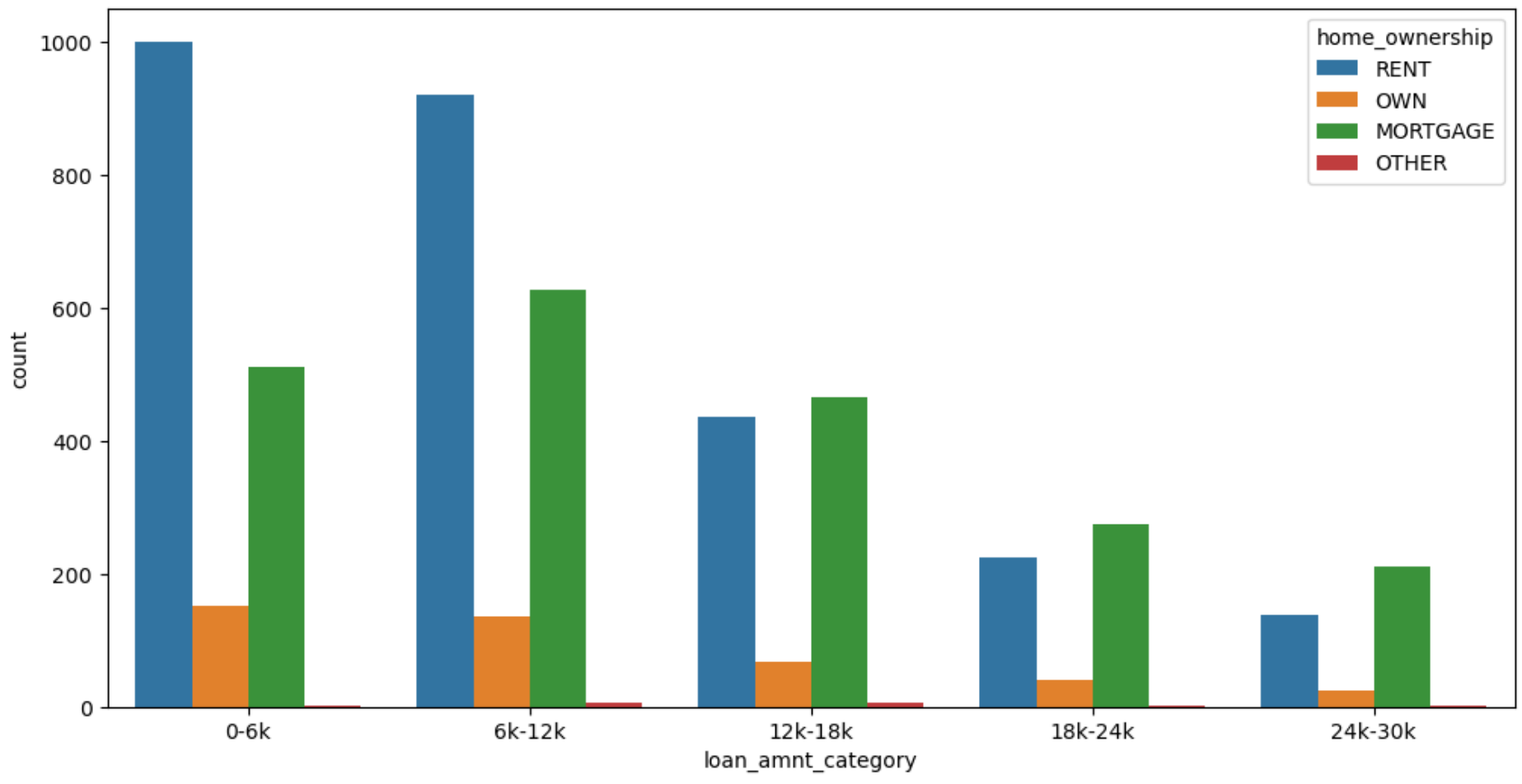

Loan Amount Distribution by Employment Length and Loan Status
Loan Status
Charged Off
Fully Paid
Loan Amount
30000
25000
20000
15000
10000
5000
0
0
1
2
3
4
5
6
7
8
9
10
Employment Length (years)

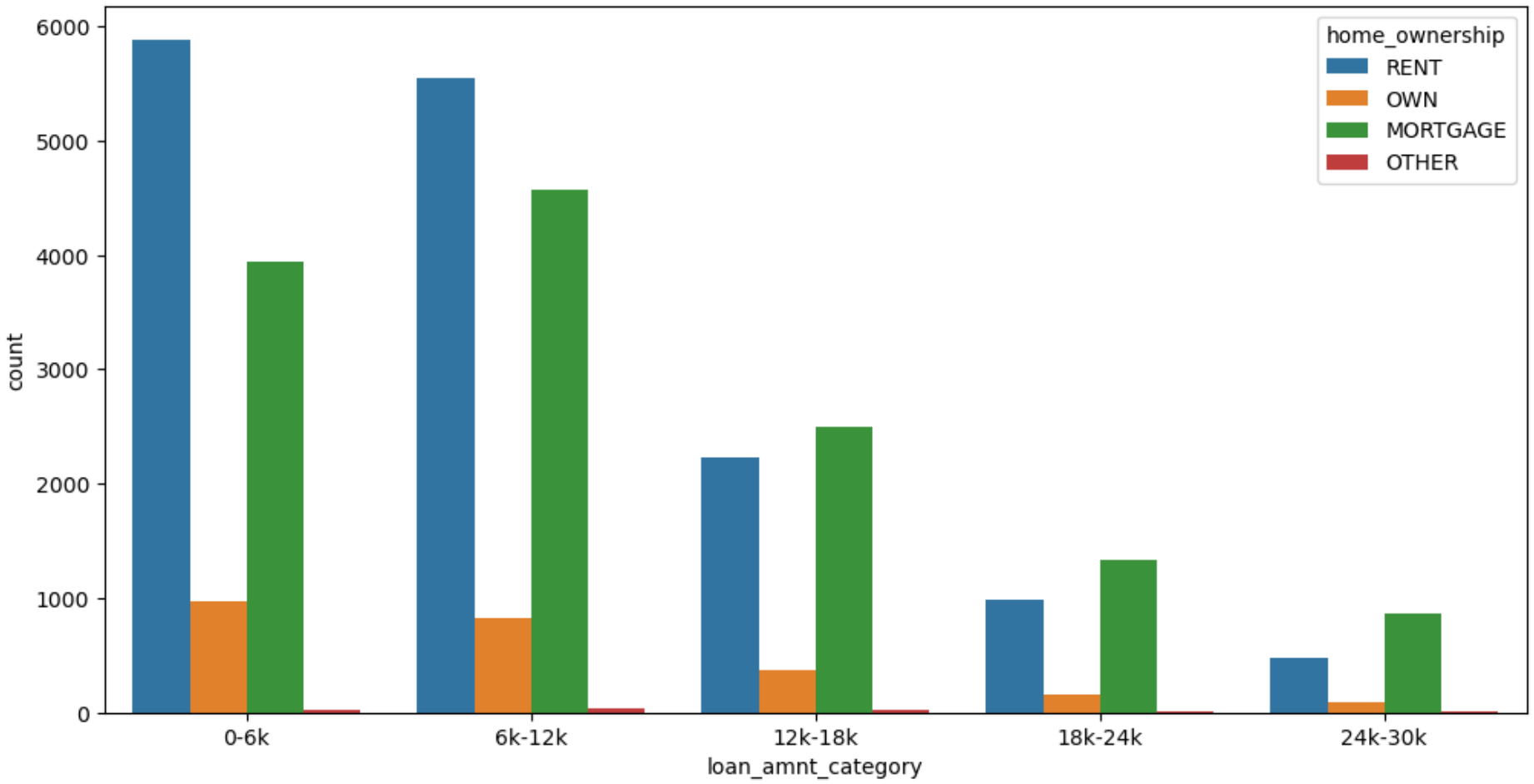

Average Loan Amount by Employment Length and Loan Status
Loan Status
Charged Off
Fully Paid
Loan Amount
20000
17500
15000
12500
10000
7500
5000
2500
0
0
1
2
3
4
5
6
7
8
9
10
Employment Length (years)

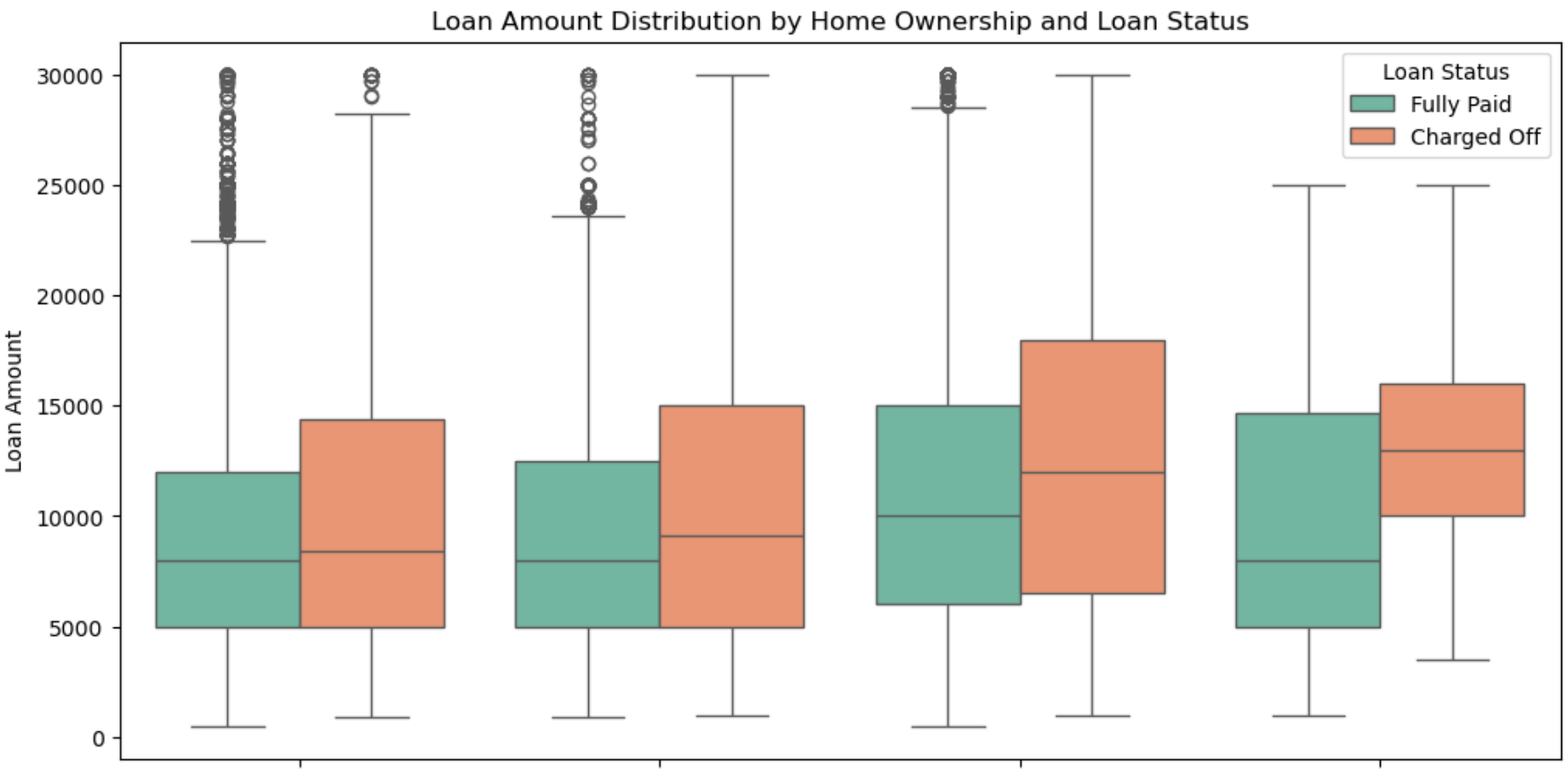


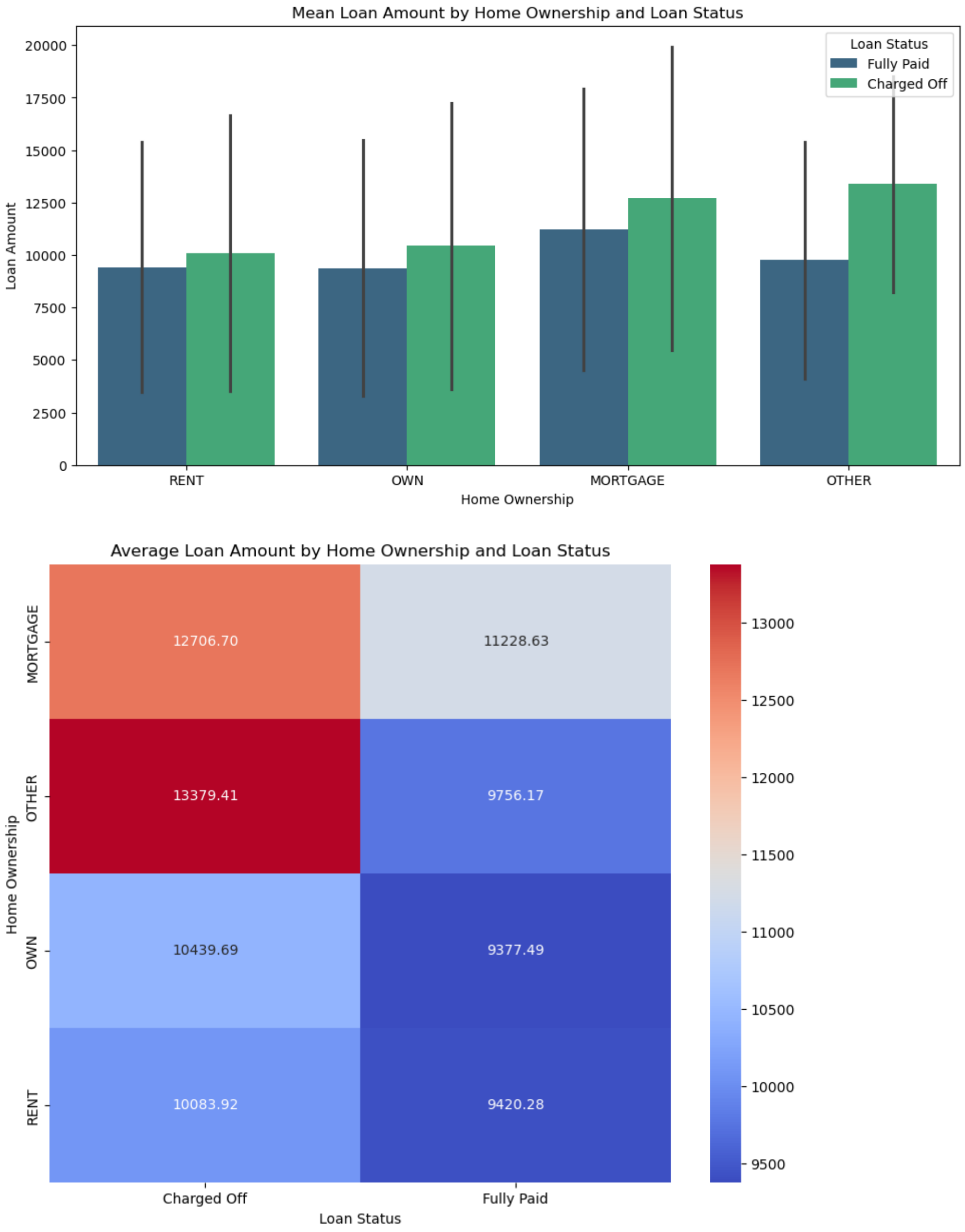
A slight trend can be seen that the average loan amount increases as the emp_length increases. Also, for same emp_length the charged Off loans have higher average loan amount.

3.2.7) Loan amount vs Home Ownership over loan_status



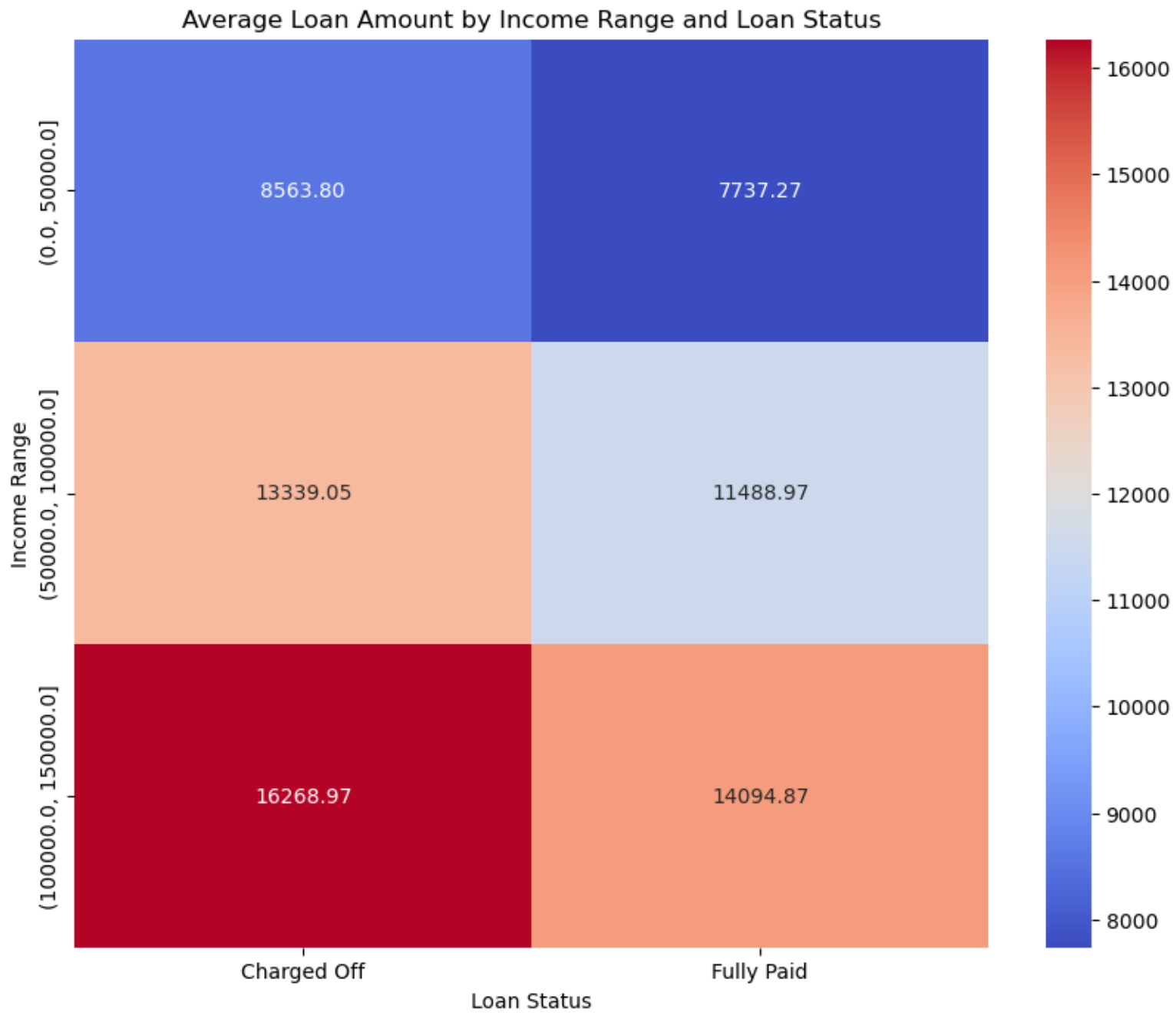

home_ownership
RENT
OWN
MORTGAGE
OTHER
count
6000
5000
4000
3000
2000
1000
0
0-6k
6k-12k
12k-18k
18k-24k
24k-30k
loan_amnt_category

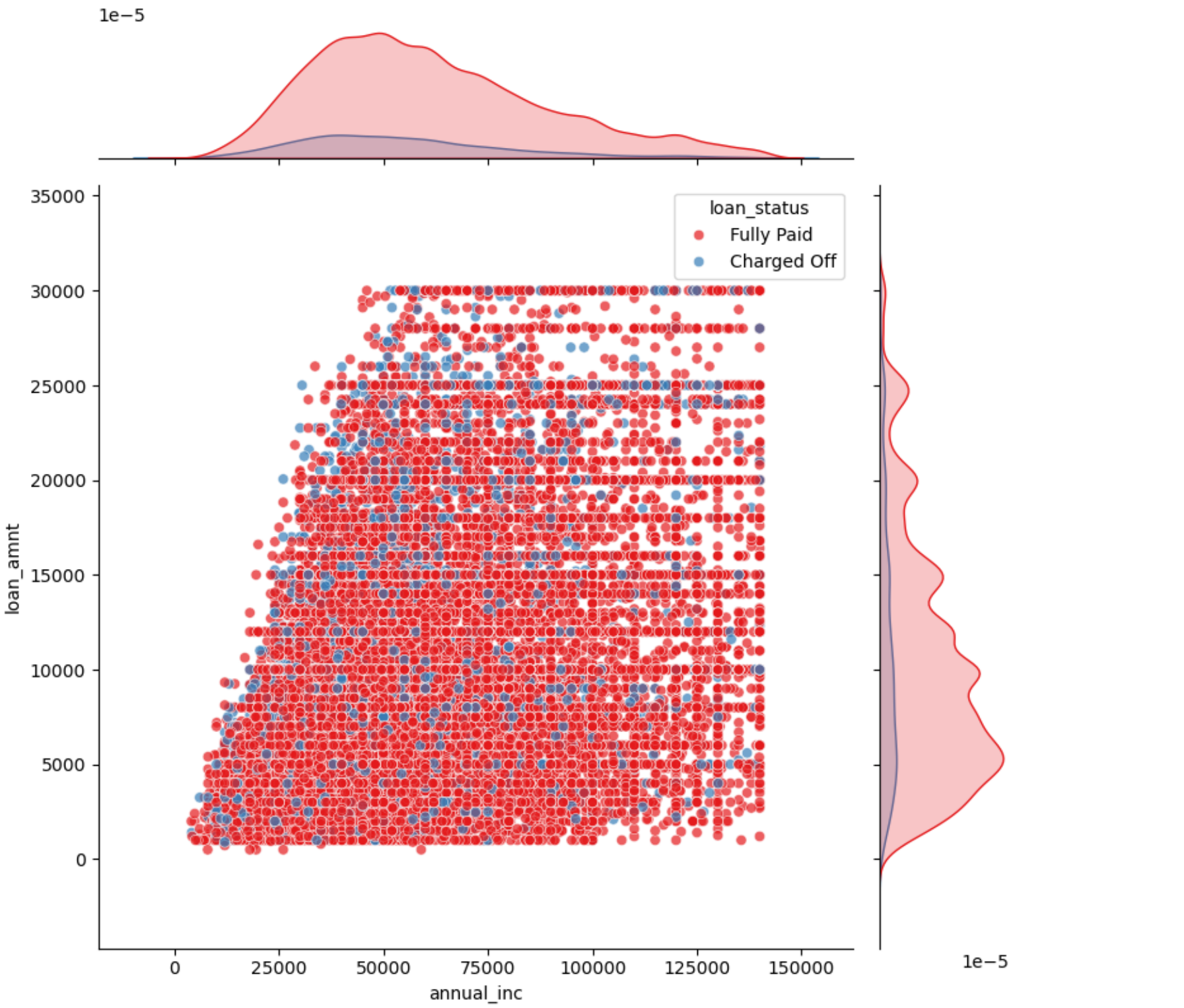

Loan Amount Distribution by Home Ownership and Loan Status
Loan Status
Fully Paid
Charged Off
Loan Amount
30000
25000
20000
15000
10000
5000
0
RENT
OWN
MORTGAGE
OTHER
Home Ownership

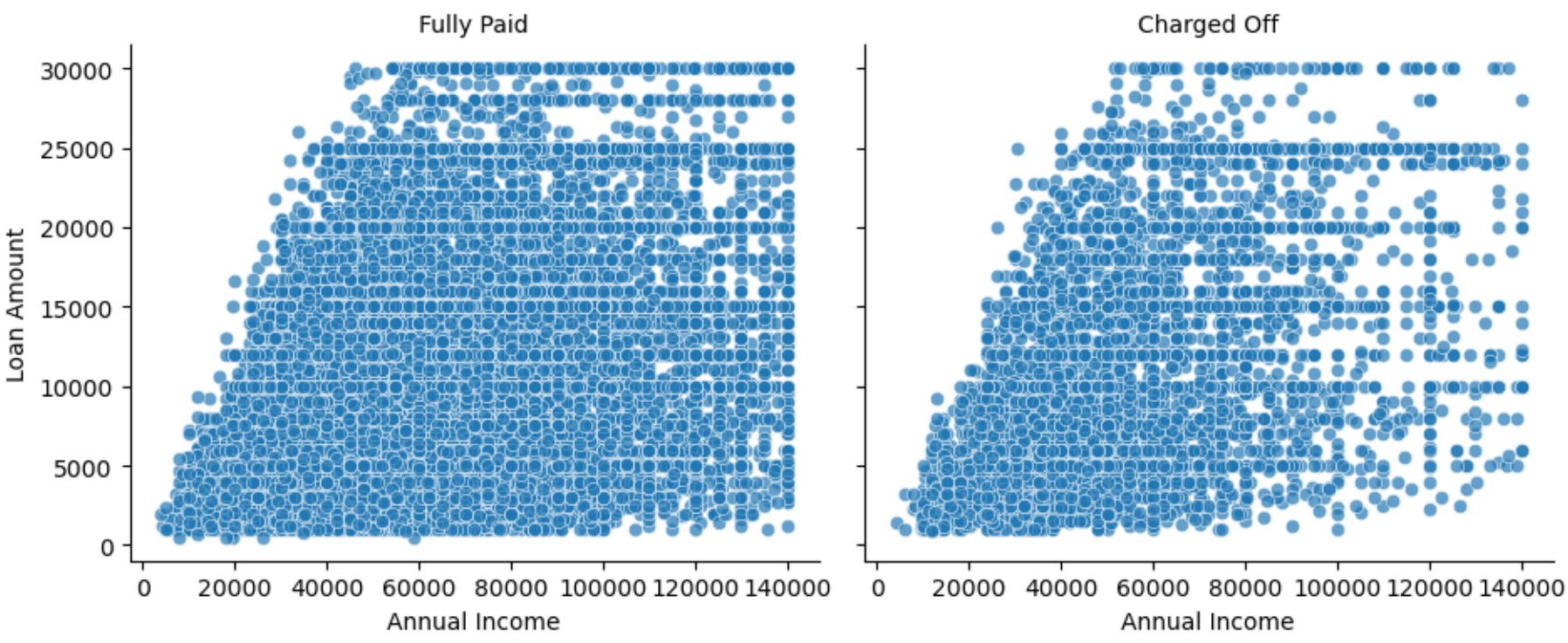


Home_Ownership value 'OTHER' has the highest average loan amount. And from previous analysis it was shown that the home_ownership value 'OTHER' has the highest percentage of Defaulters (Charged Off loans) among various home_ownership values. Also, it can be seen for each category of home_ownership has higher average loan amount for Charged Off as compared to Fully Paid.

3.2.8) Loan amount vs annual_income over loan_status¶



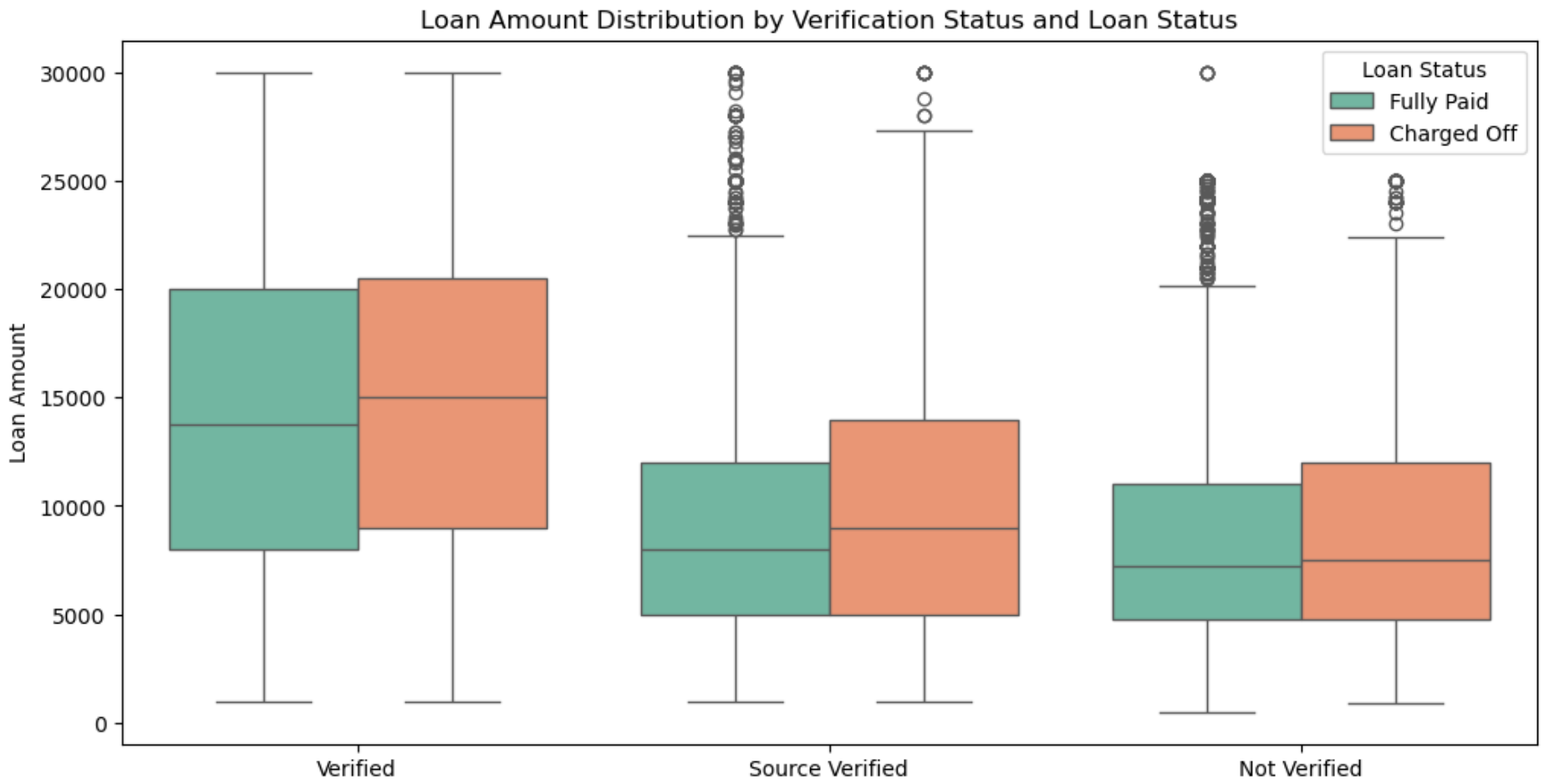


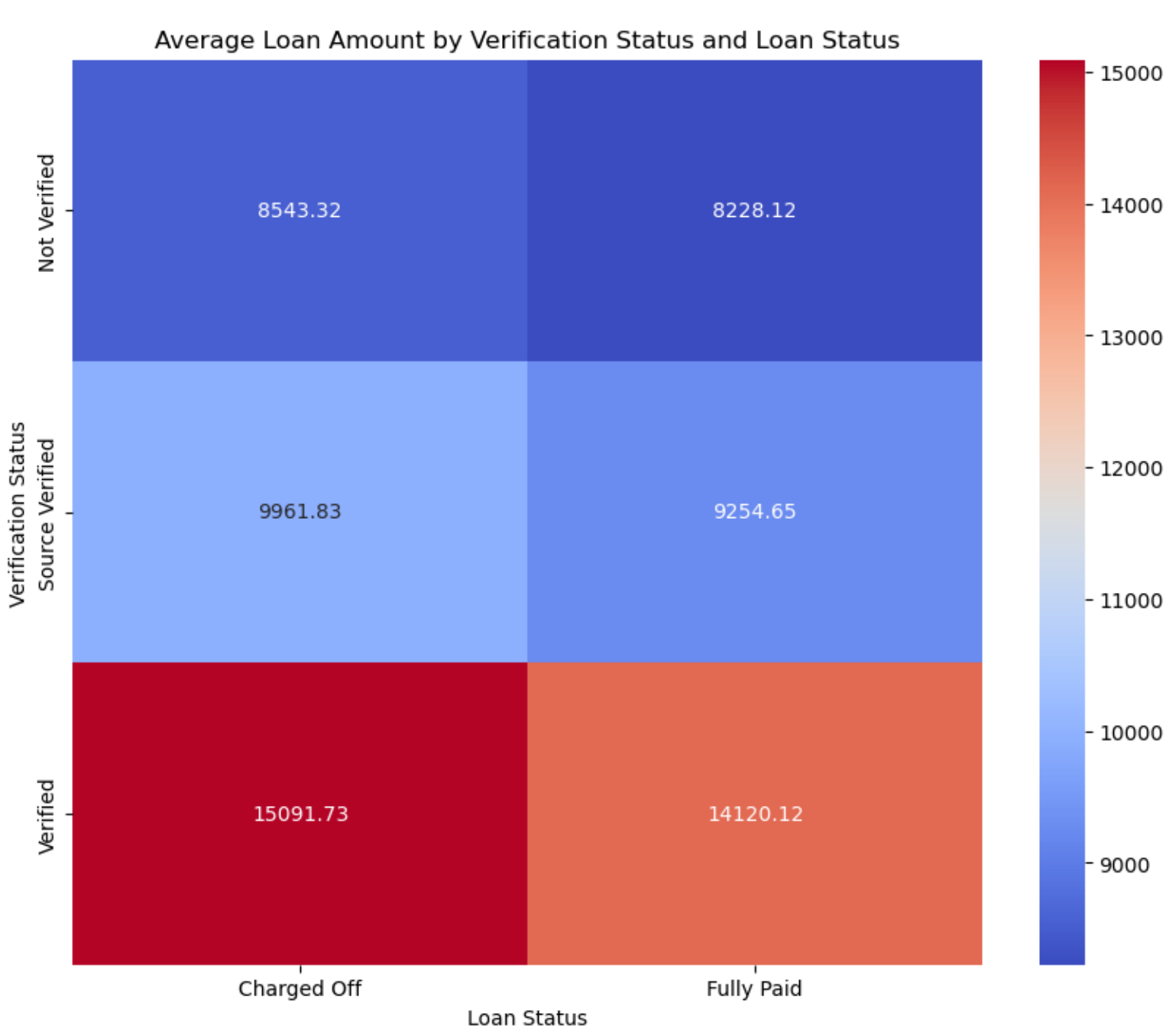


From scatter plot, it can be seen that Charged Off loans are mostly concentrated towards the lower income segment (0 to 60k).

Also, it can be seen from heatmap that in each income segment the average loan amount is higher for Charged Off than for Fully Paid

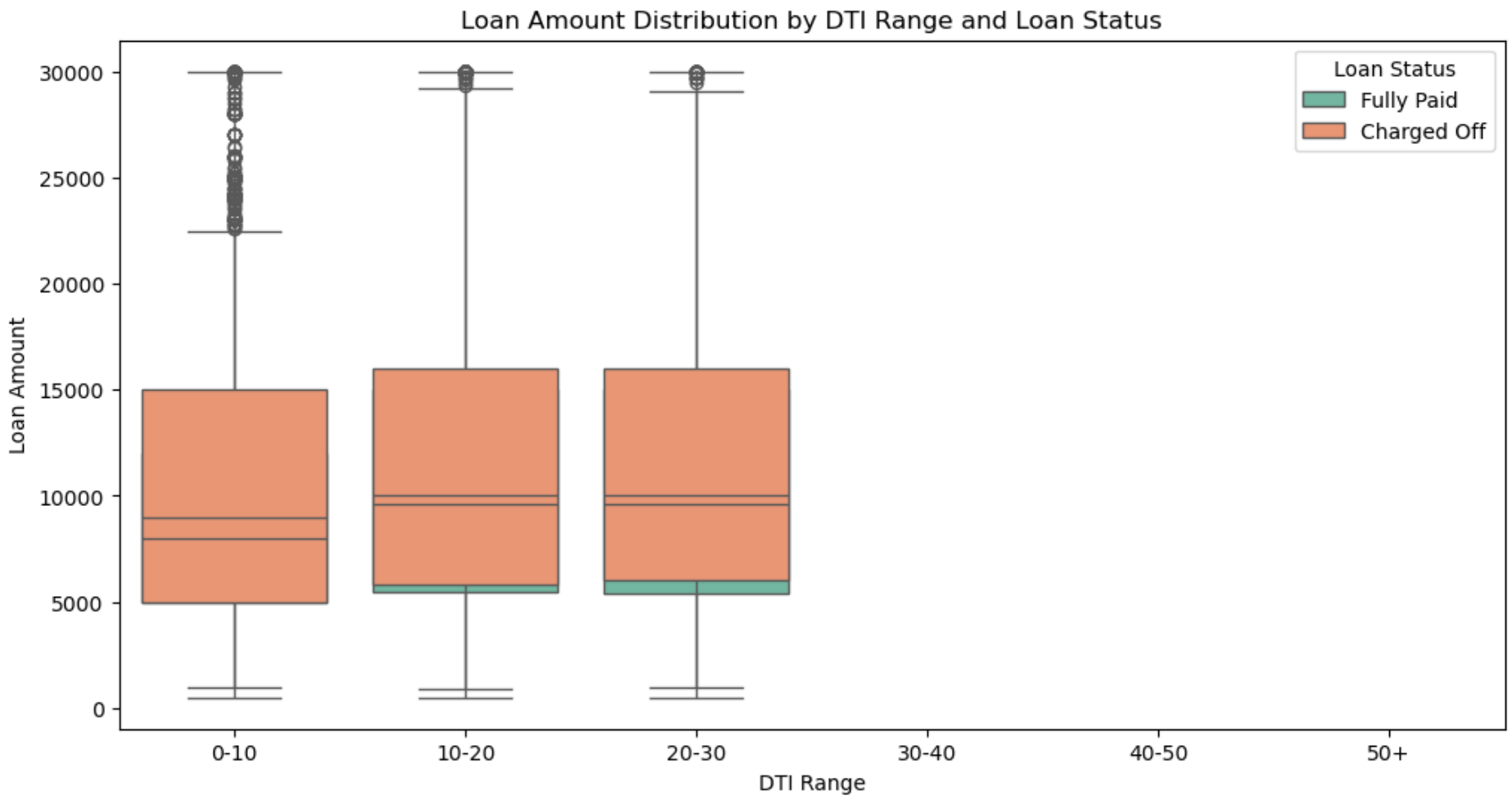
3.2.9) Loan amount vs Verification Status over loan_status

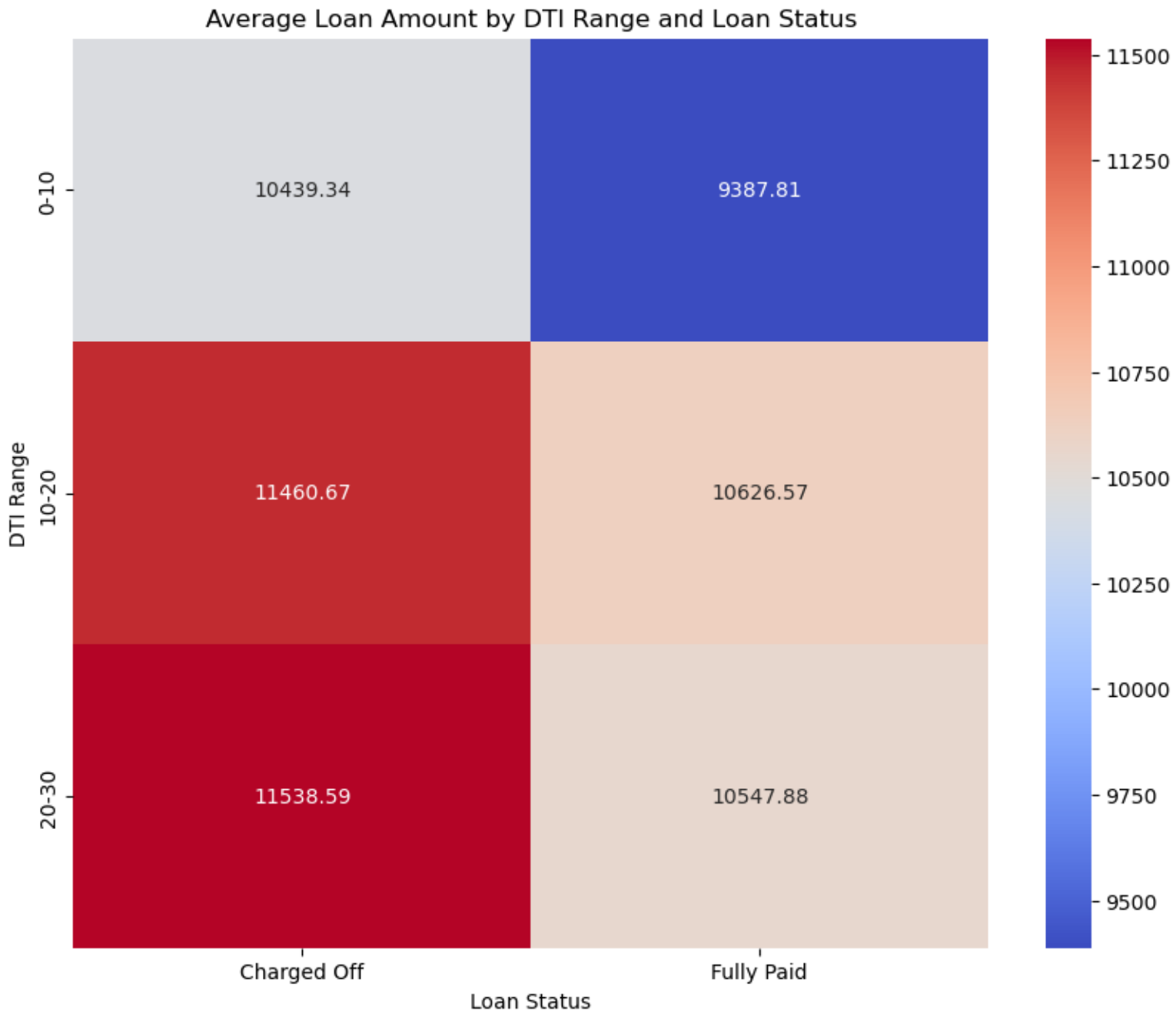

Loan Amount Distribution by Verification Status and Loan Status
Loan Status
Fully Paid
Charged Off
Loan Amount
30000
25000
20000
15000
10000
5000
0
Verified
Source Verified
Not Verified
Verification Status


Average Loan Amount by Verification Status and Loan Status
Not Verified
Source Verified
Verified
Verification Status
8543.32
8228.12
9961.83
9254.65
15091.73
14120.12
Charged Off
Fully Paid
Loan Status
15000
14000
13000
12000
11000
10000
9000

From Part 1 analysis it was shown that verification_status 'Verified' has slightly higher percentage of default loans as compared to other values of verification_status. From heatmap it is seen that the highest values for average loan amounts are for verification_status 'Verified'. Also, for each category of verification_status the Charged Off loans have higher average loan amount than for Fully Paid.

3.2.10) Loan amount vs Debt to Income over loan_status

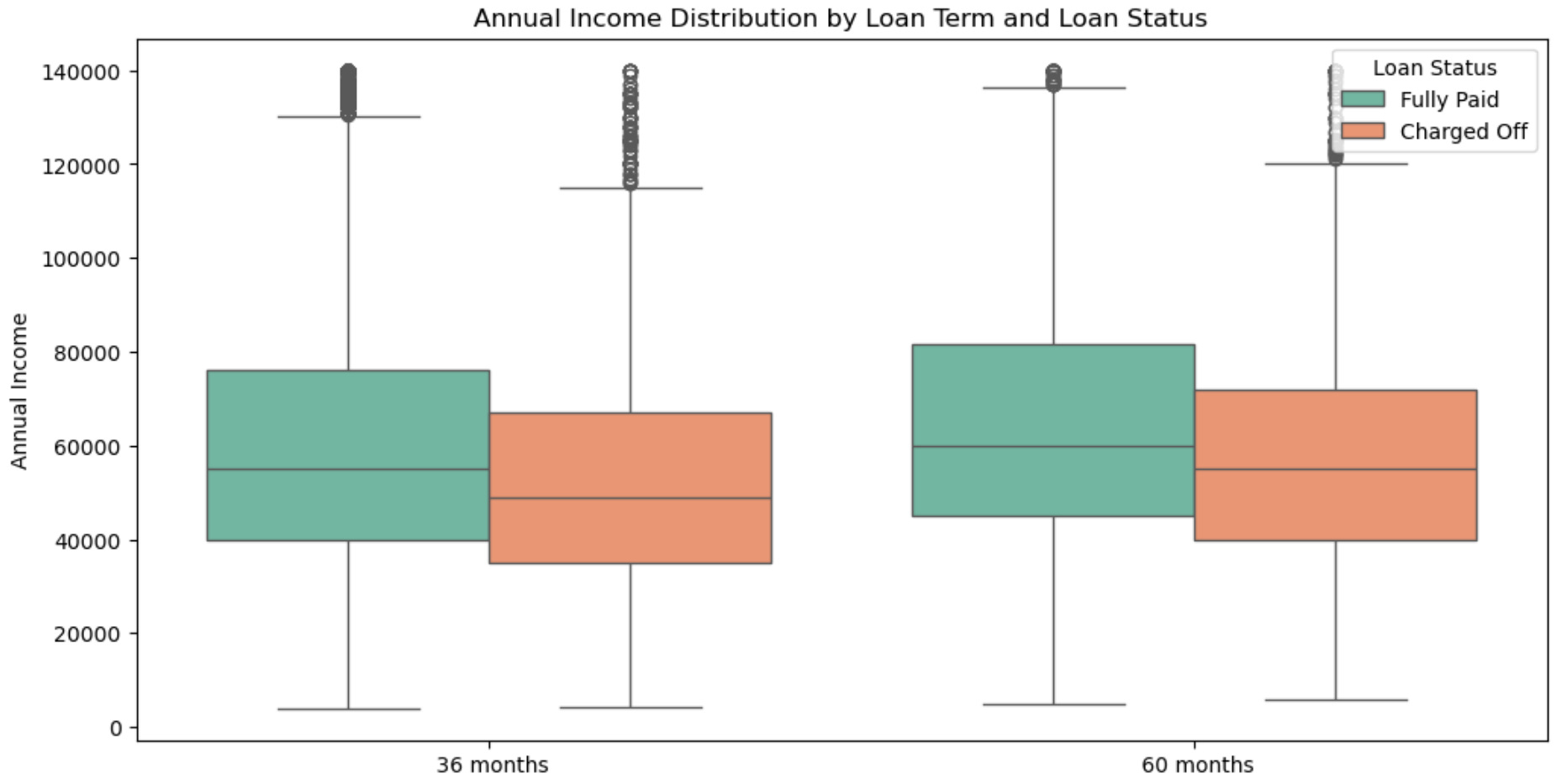


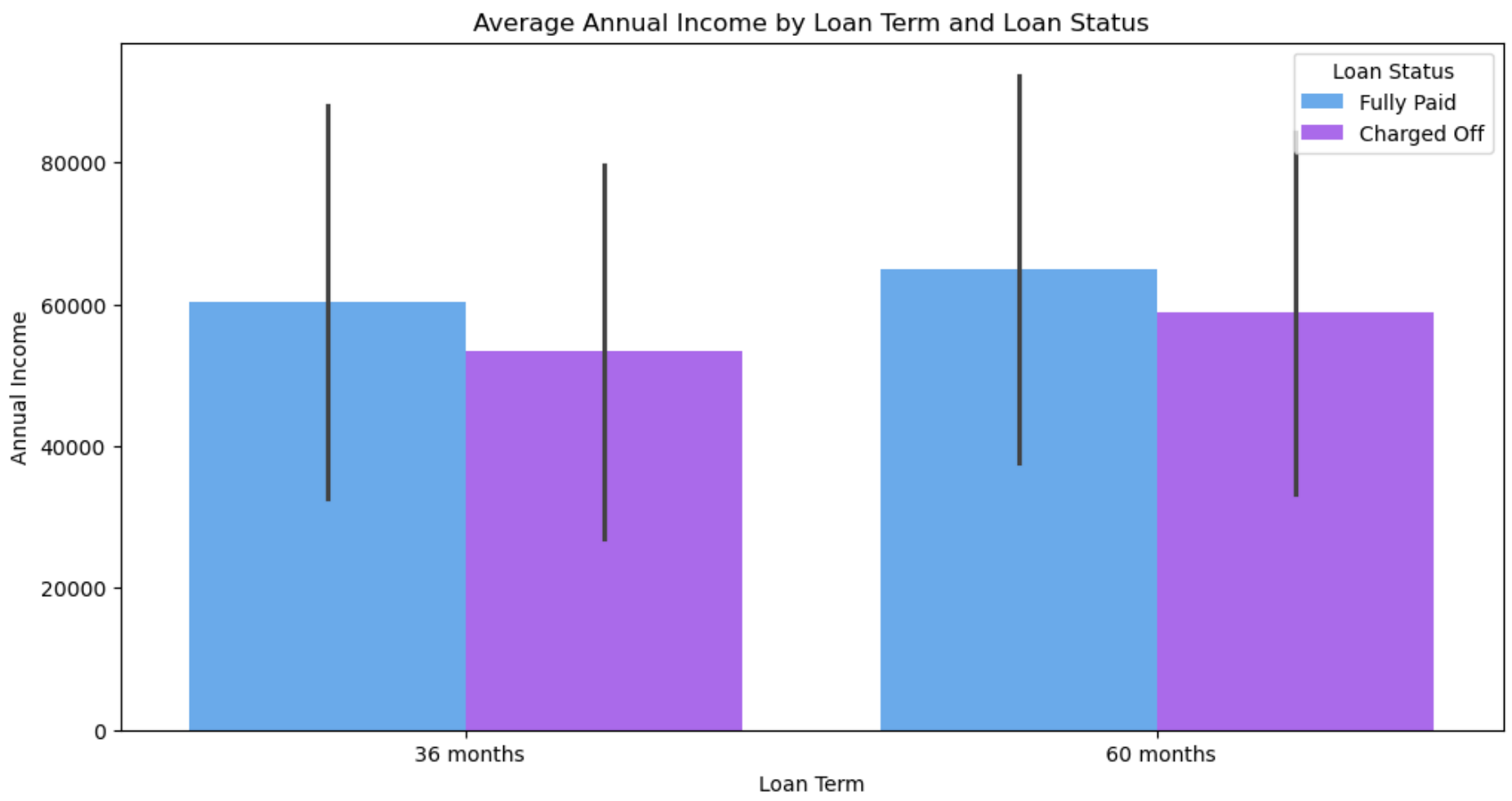


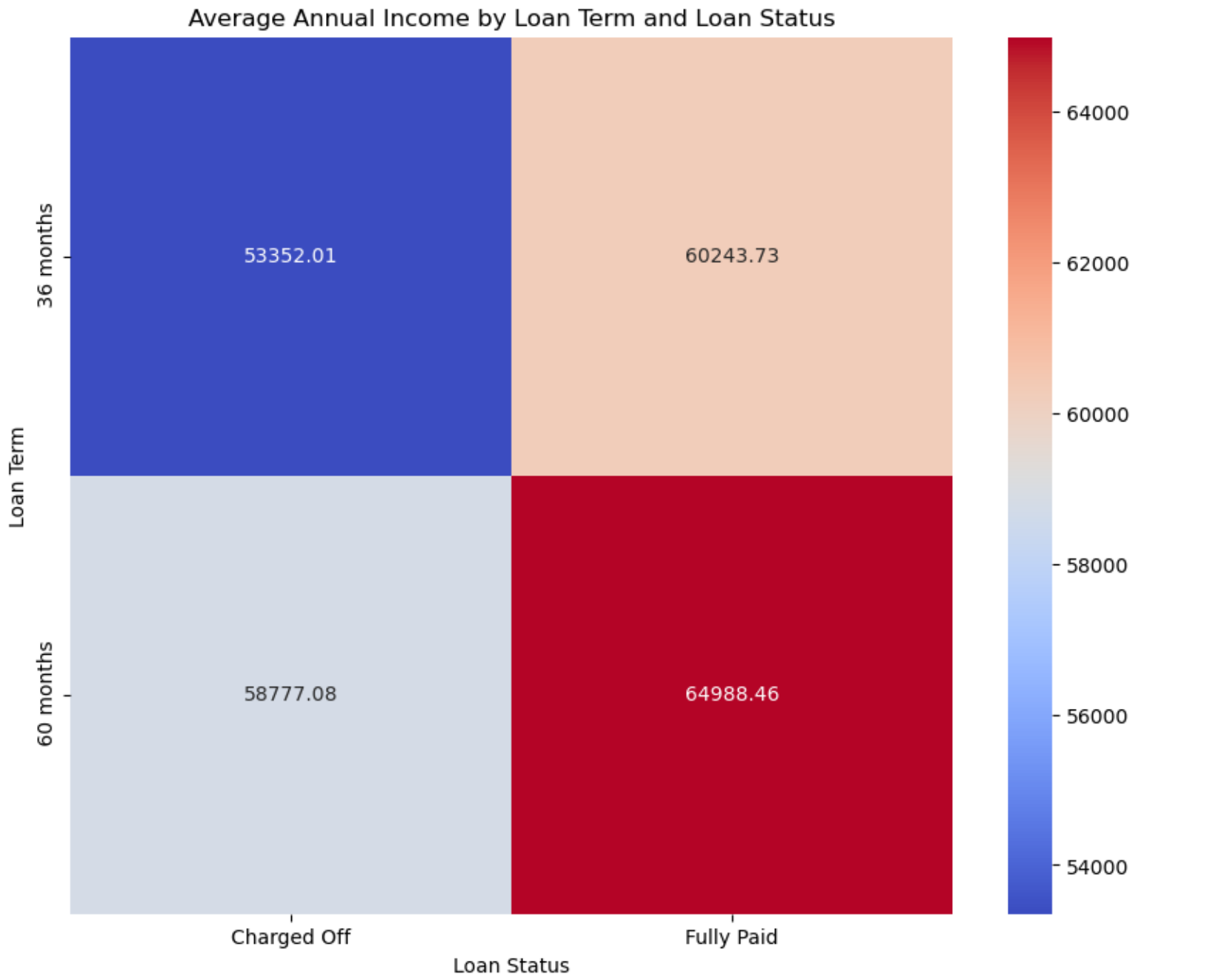
From heatmap it can be seen that for each segment of DTI the average loan amount for Charged Off is higher than for Fully Paid. Also, from scatter plot it can be seen that most Charged Off loans are located in DTI range of 15-25. From heatmap it can be seen that higher DTI range has higher average loan amount.

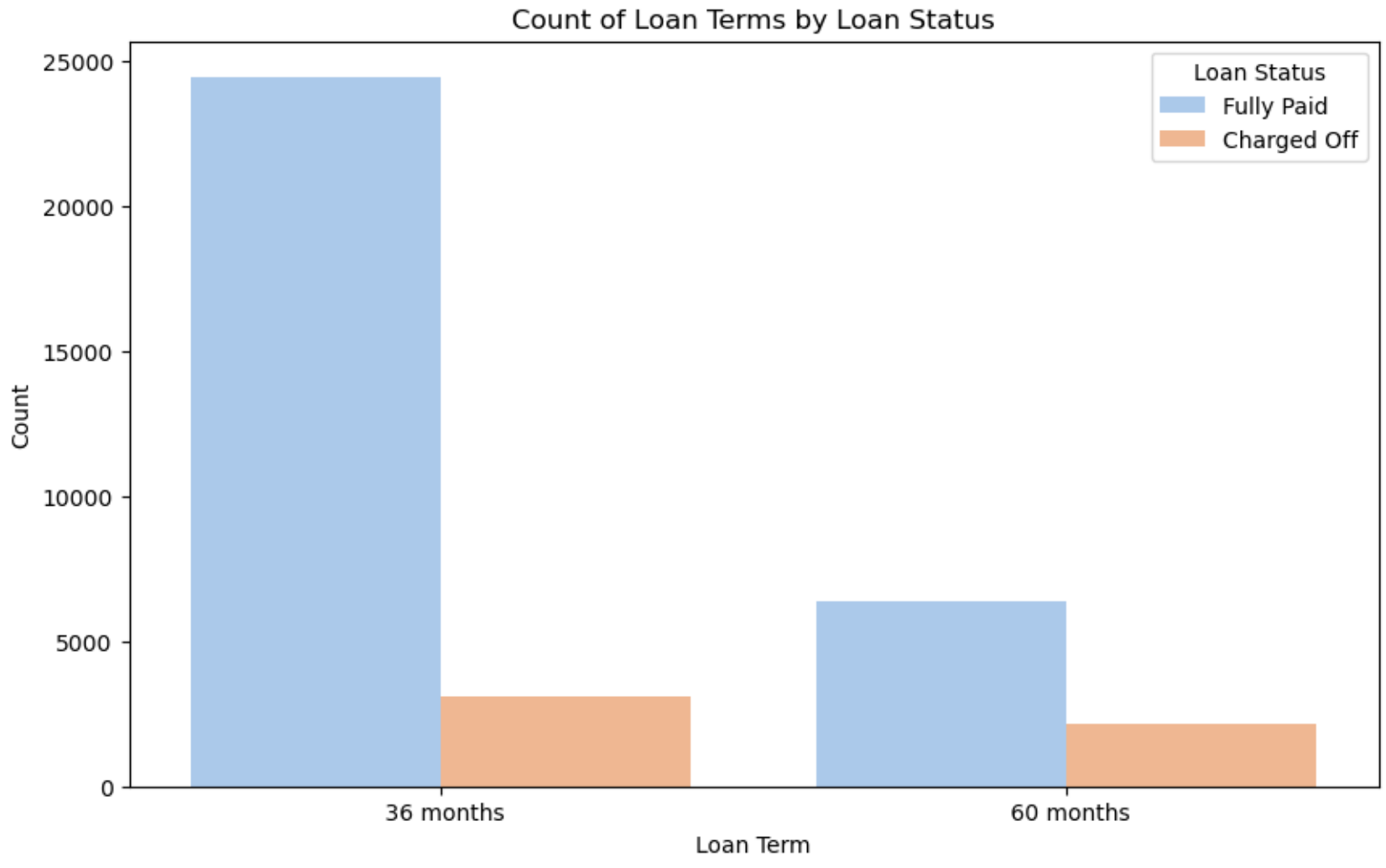
Analyzing annual income with other columns for more insights

3.2.11) Annual income vs Term over loan_status


Annual Income Distribution by Loan Term and Loan Status
Loan Status
Fully Paid
Charged Off
Annual Income
140000
120000
100000
80000
60000
40000
20000
0
36 months
60 months
Loan Term

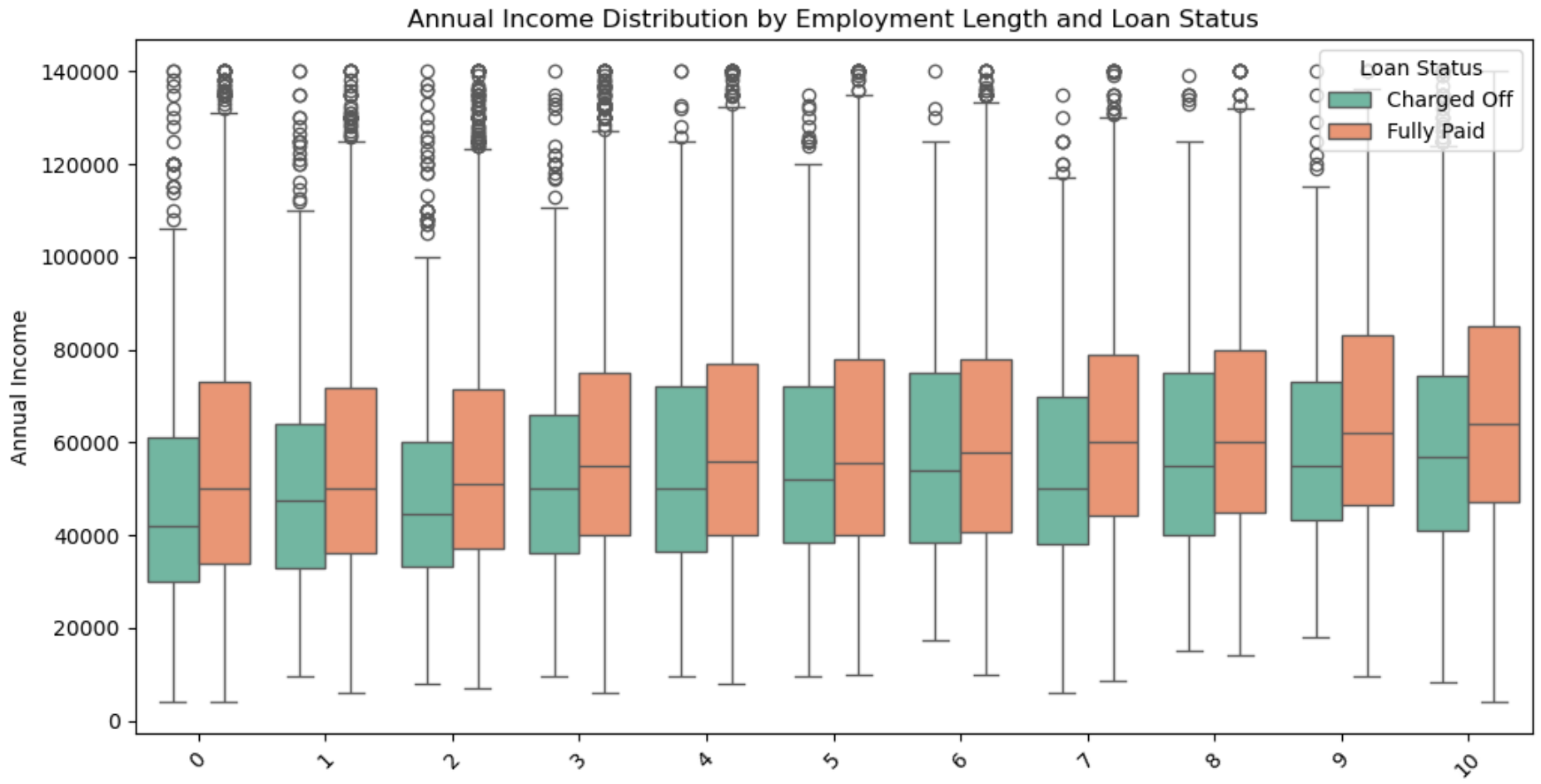

Average Annual Income by Loan Term and Loan Status
Loan Status
Fully Paid
Charged Off
Annual Income
80000
60000
40000
20000
0
36 months
60 months
Loan Term

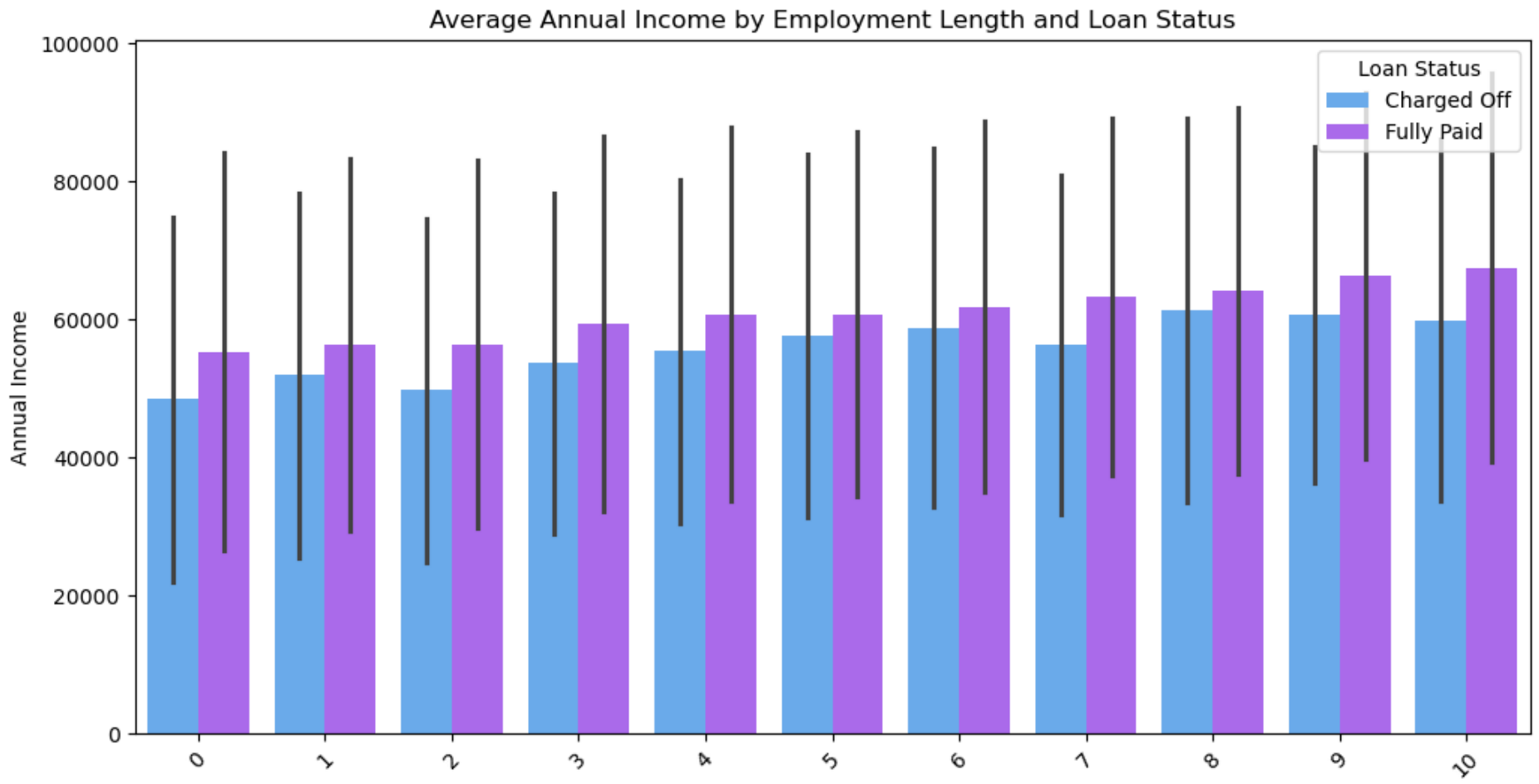


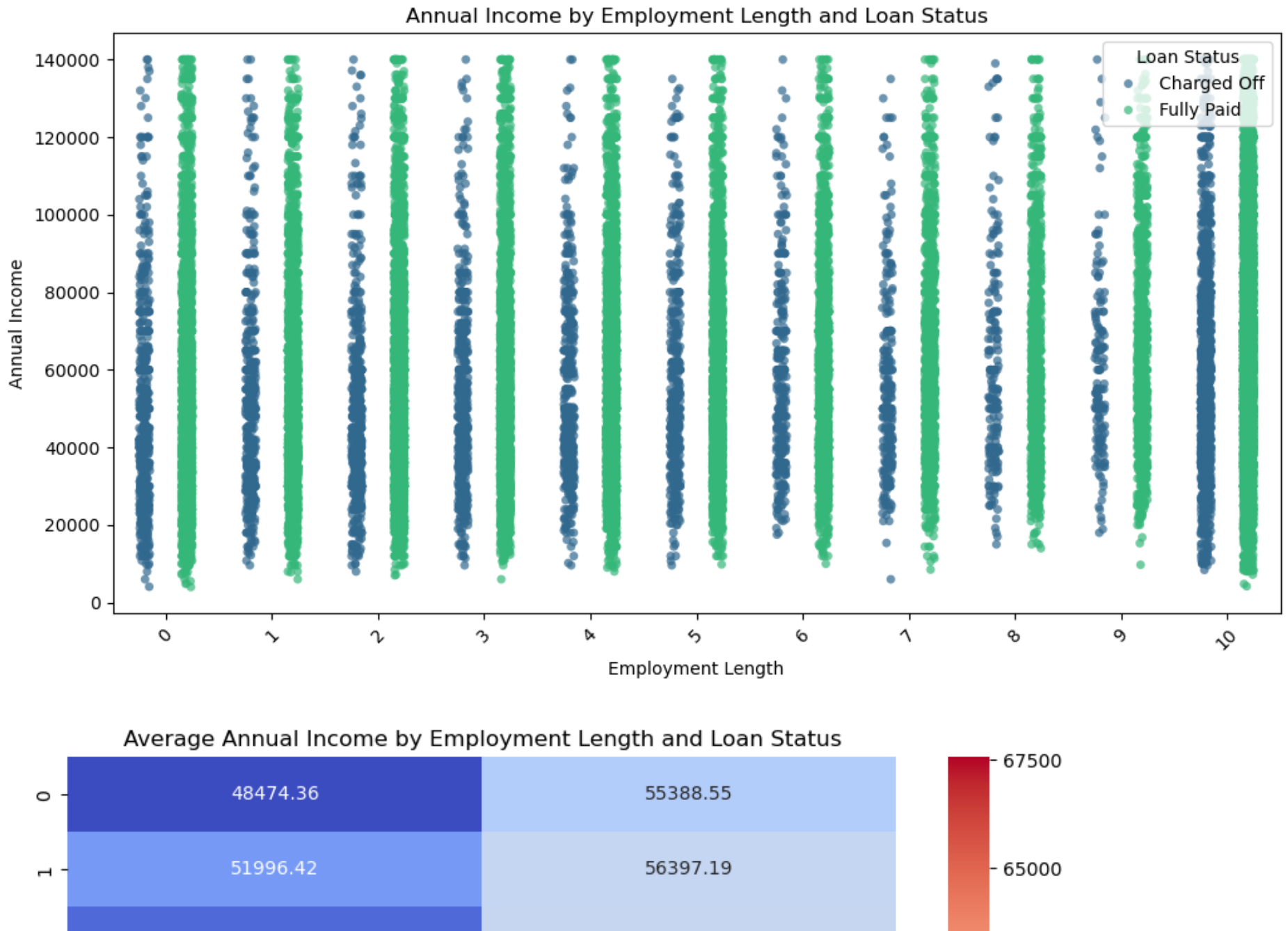


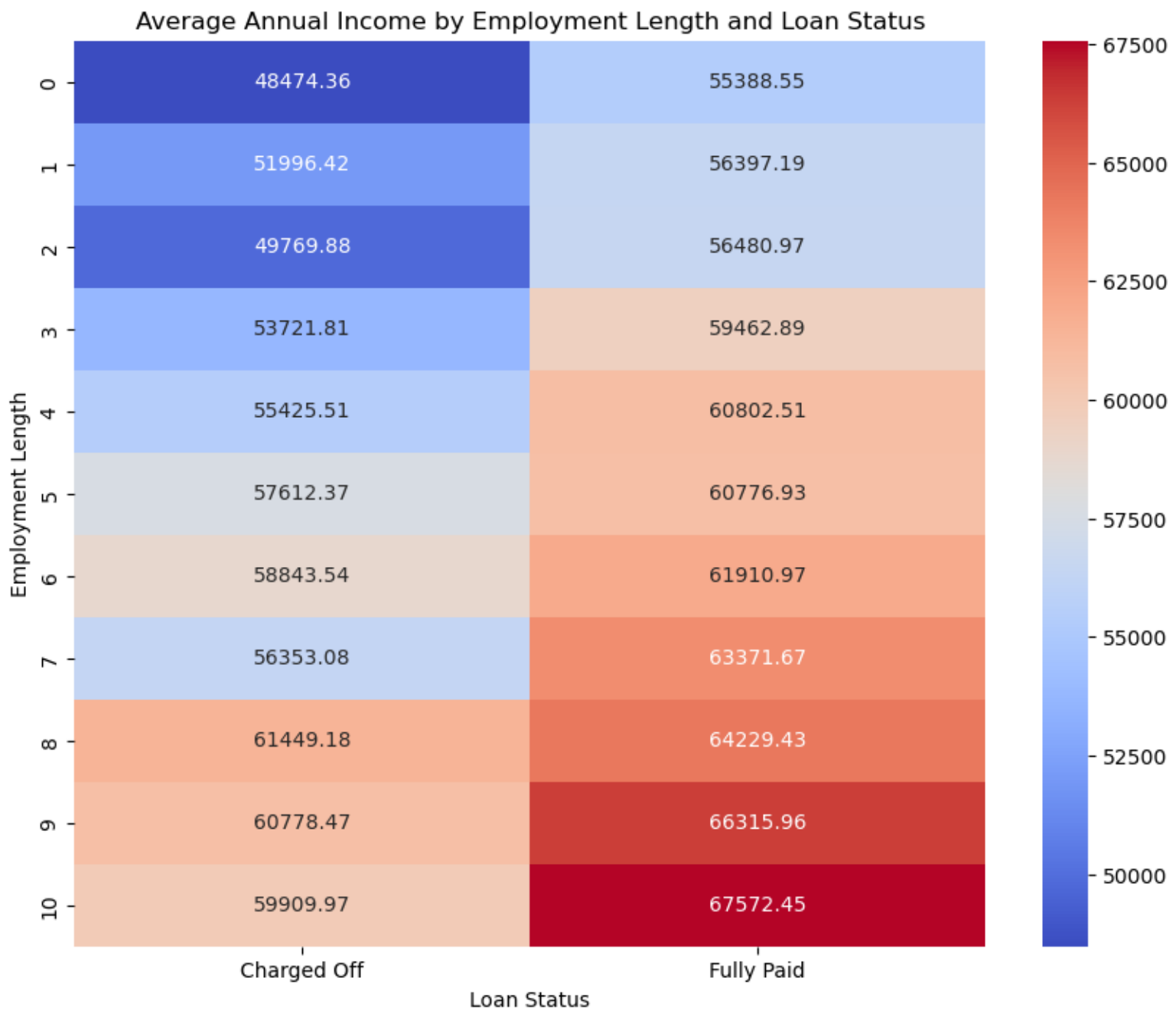
Average annual income is higher for 60 months terms as compared to 36 months. Also, for same term the average annual income is higher for Charged Off as compared to Fully paid.

3.2.12) Annual income vs Employee Length over loan_status


Annual Income Distribution by Employment Length and Loan Status
Loan Status
Charged Off
Fully Paid
Annual Income
0
20000
40000
60000
80000
100000
120000
140000
0
1
2
3
4
5
6
7
8
9
10
Employment Length

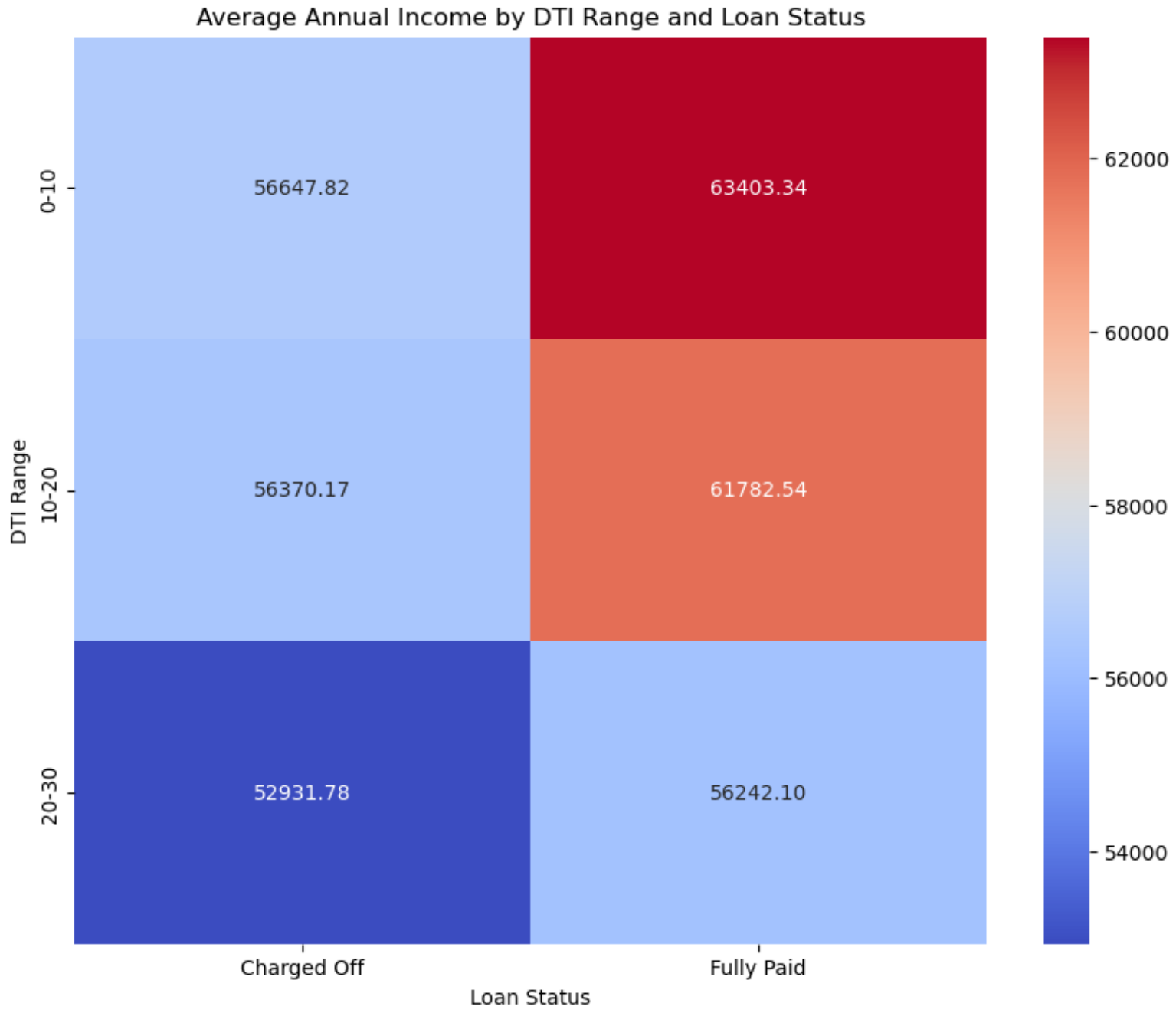

Average Annual Income by Employment Length and Loan Status
Loan Status
Charged Off
Fully Paid
Annual Income
0
20000
40000
60000
80000
100000
0
1
2
3
4
5
6
7
8
9
10
Employment Length





It can be seen that as emp_length increases the average annual income increases. Also, it can be seen from heat map that in each emp_length segment the Fully Paid loans have higher average annual income than the Charged Off

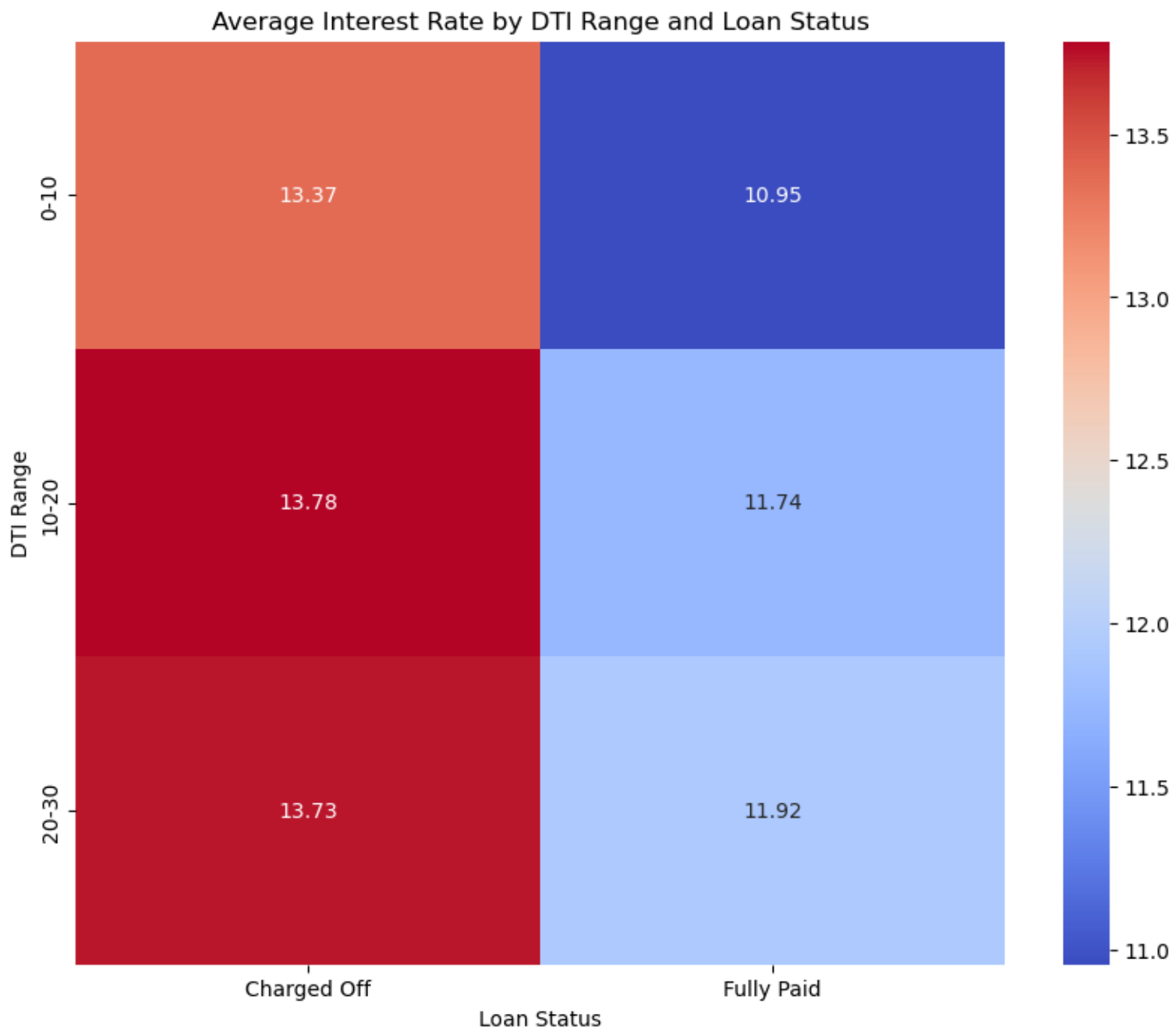
3.2.13) Annual income vs Debt to Income over loan_status



As the DTI increases the average annual income decreases. For the same category of DTI the Fully Paid loans have higher income than the Charged Off loans.

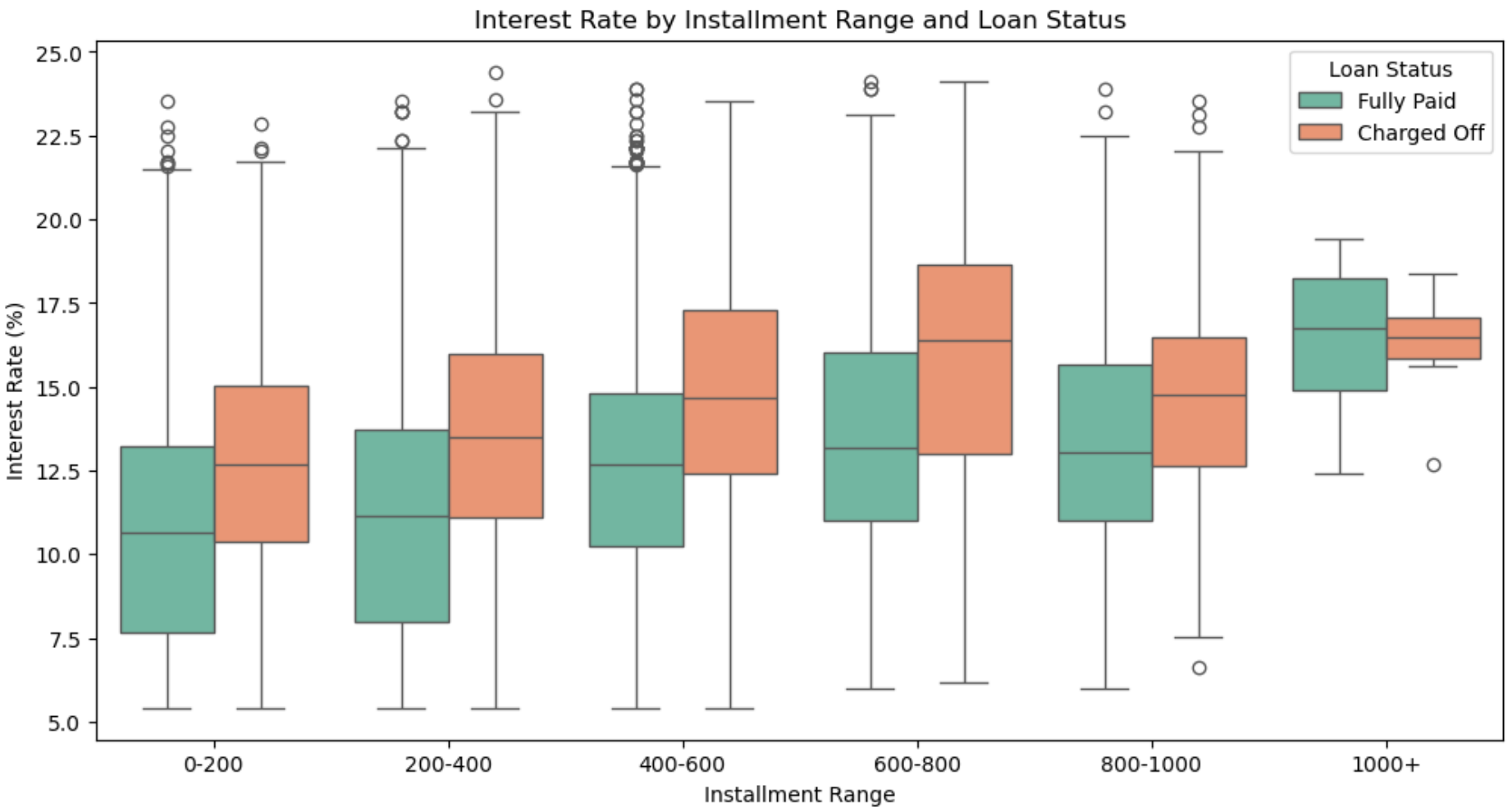
Analyzing interest rate with other columns for more insights

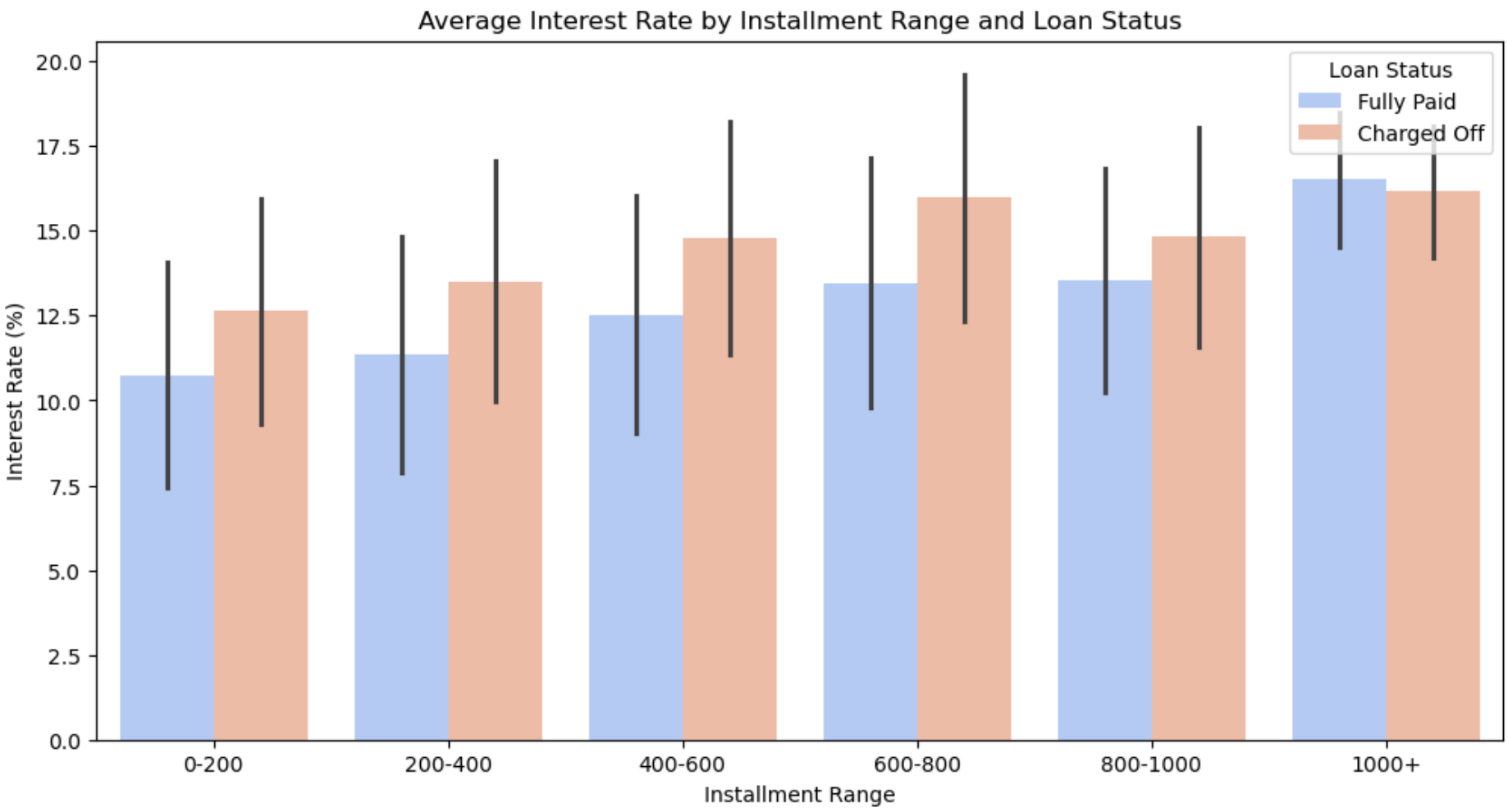
3.2.14) Interest Rate and Debt To Income

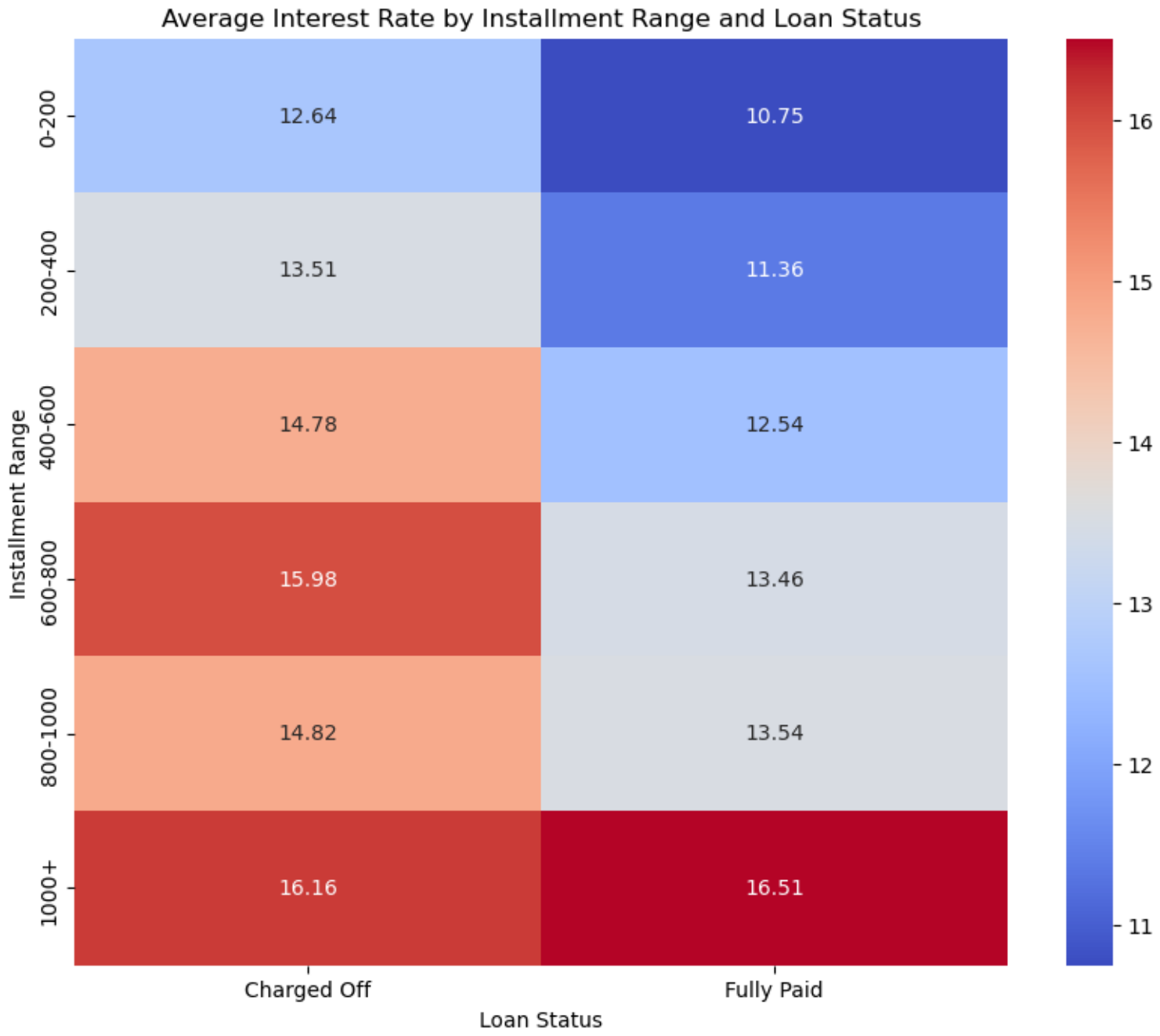


Average Interest rates are low for Fully paid Loans as compared to Charged Off loans.

3.2.15) Interest Rate and Installment

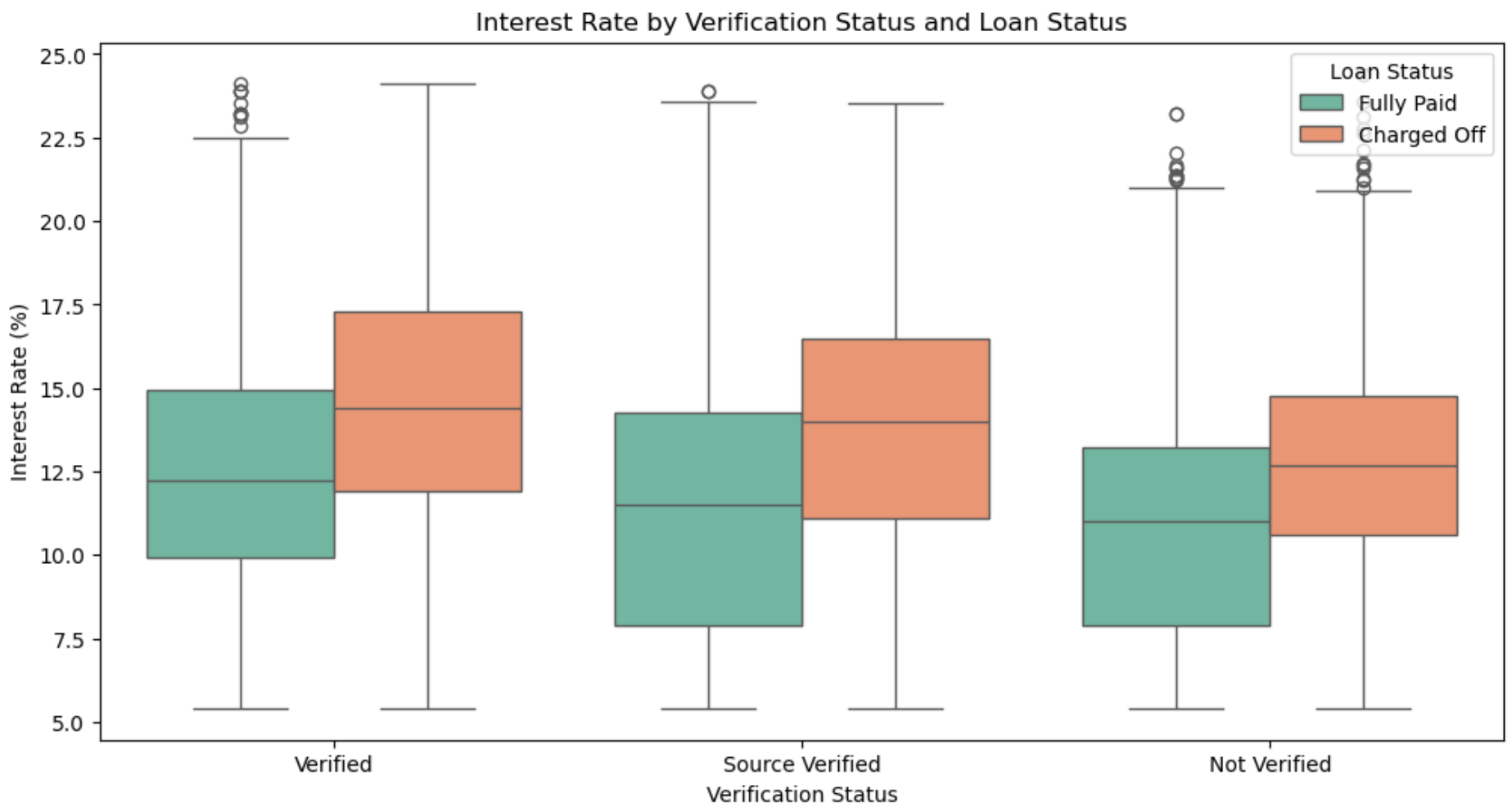

Interest Rate by Installment Range and Loan Status
Loan Status
Fully Paid
Charged Off
Interest Rate (%)
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
0-200
200-400
400-600
600-800
800-1000
1000+
Installment Range

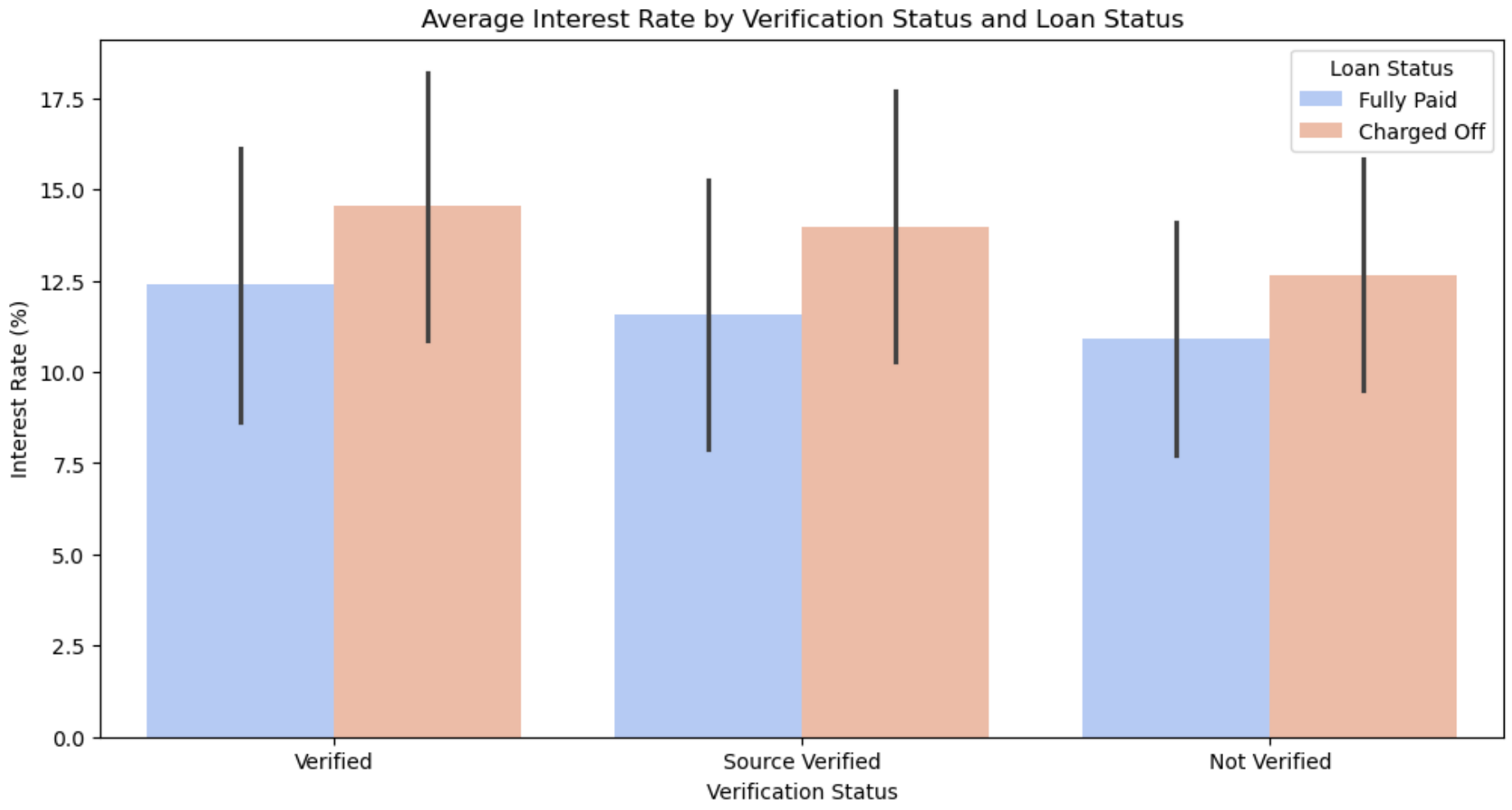

Average Interest Rate by Installment Range and Loan Status
Loan Status
Fully Paid
Charged Off
Interest Rate (%)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0-200
200-400
400-600
600-800
800-1000
1000+
Installment Range

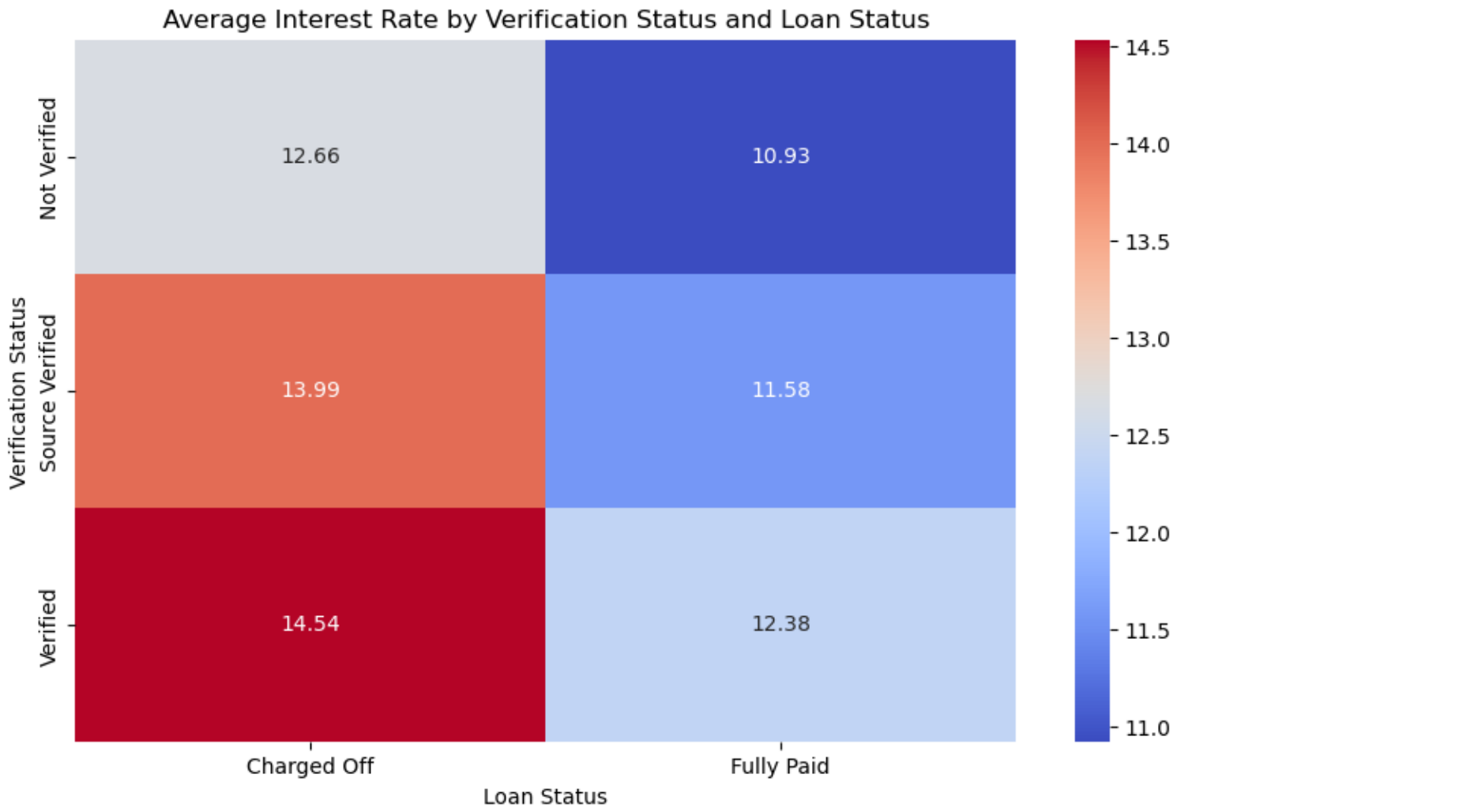


Average interest rate increases as number of installments increases. From previous analysis higher interest rates are linked with higher Charged off loans.

3.2.16) Interest Rate and Verification Status







It seems that verification status as 'Verified' are linked with highest interest rates. And as can be seen in heat plot, that the each category of verification status has higher interest rates for Charged Off loans as compared to Fully paid.

Observations from Data Analysis Part 2 - using loan_status and two more columns

1. Confirms that 'Small_Business' have the highest average loan amount and are most risky type of loan.
2. 60 Months term have higher mean loan amount. This confirms our two observations that higher loan amount have higher risk of being Charged Off (as we know by previous analysis that 60 months term have higher percentage of Charged Off loans as compared to 36 months term) and 60 Months term is associated with higher risk of being Charged Off.
3. Higher loan amounts have higher interest rates. And from our previous analysis it was seen that both of these factors are associated with high risk of loan being Charged Off.
4. Mean number of installments increases as the loan amount increases.

5. Grades E,F,G has higher average loan amounts in the Charged Off segment, as we have seen in previous analysis that these three also have highest percentage of Charged Off loans as compared to other grades. Thus this confirms our findings. It also confirms a strong link between higher loan amounts being more risky and prone to being Charged Off.
6. A slight trend can be seen that the average loan amount increases as the emp_length increases. Also, for same emp_length the charged Off loans have higher average loan amount.
7. Home_Ownership value 'OTHER' has the highest average loan amount. And from previous analysis it was shown that the home_ownership value 'OTHER' has the highest percentage of Defaulters (Charged Off loans) among various home_ownership values. Also, it can be seen for each category of home_ownership has higher average loan amount for Charged Off as compared to Fully Paid.
8. From scatter plot, it can be seen that Charged Off loans are mostly concentrated towards the lower income segment (0 to 60k). Also, it can be seen from heatmap that in each income segment the average loan amount is higher for Charged Off than for Fully Paid.
9. From Part 1 analysis it was shown that verification_status 'Verified' has slightly higher percentage of default loans as compared to other values of verification_status. From heatmap it is seen that the highest values for average loan amounts are for verification_status 'Verified'. Also, for each category of verification_status the Charged Off loans have higher average loan amount than for Fully Paid.
10. From heatmap it can be seen that for each segment of DTI the average loan amount for Charged Off is higher than for Fully Paid. Also, from scatter plot it can be seen that most Charged Off loans are located in DTI range of 15-25. From heatmap it can be seen that higher DTI range has higher average loan amount.
11. Average annual income is higher for 60 months terms as compared to 36 months. Also, for same term the average annual income is higher for Charged Off as compared to Fully paid.
12. It can be seen that as emp_length increases the average annual income increases. Also, it can be seen from heat map that in each emp_length segment the Fully Paid loans have higher average annual income than the Charged Off
13. As the DTI increases the average annual income decreases. For the same category of DTI the Fully Paid loans have higher income than the Charged Off loans.
14. Average Interest rates are low for Fully paid Loans as compared to Charged Off loans.
15. Average interest rate increases as number of installments increases. From previous analysis higher interest rates are linked with higher Charged off loans.
16. It seems that verification status as 'Verified' are linked with highest interest rates. And as can be seen in heat plot, that the each category of verification status has higher interest rates for Charged Off loans as compared to Fully paid.

For clarity, repeating observations from analysis part one.

Observations from above Analysis Part 1: - using Loan_Status and one more variable.

1. Among all grades, The highest percentage of defaulters (Charged Off) are in grade G (next highest in F and then in E).
2. 60 Months term shows higher percentage of Charged Off Loans, as compared to 36 months.
3. All emp_length values have almost equal percentage of defaulters though emp_length >=10 years have little higher percentage of defaulters. So emp_length does not seem to have much impact on risk of a loan being Charged Off.
4. Among all values of 'purpose', highest percentage of defaulters(Charged Off) are there for loans taken for 'small_business'.
5. Home_Ownership has almost no significant impact. Though Home_ownership = OTHER has slightly higher percentage of defaulters (Charged Off loans) as compared to other values of home_ownership.
6. Different verification_status values shows almost negligible differences in percentages of defaulters, though verification_status = 'verified' shows (slightly higher than others) highest percentage of defaulted loans as compared to other values of verification_status. So this will have little impact on the risk calculation of a loan.
7. As the loan amount increases the possibility of default(Charged Off loan) increases. The loan_amnt_category with highest loan seems to have highest percentage of loan default(Charged Off). 18k-24k, 24k-30k are two ranges of loan_amnt with highest possibility of a loan being Charged Off as compared to other lower ranges of loan_amnt.
8. Among various interest Rate values, the highest percentage of loan defaults(Charged Off) are for loans where interest rate is high around 17%-24%.
9. Installments do not have much impact on the risk of a loan being charged off, though installments ranging from 370-975 have slightly higher possibility of being charged off as compared to other installment categories.
10. Among all the dti values, the higher percentage of defaults(charged off) can be seen for dti ranging between 18-30. It can be seen that in general when dti increases the risk of loan being Charged Off increases.
11. The people with low annual income have more default rate. The highest percentage of defaulters can be seen in annual income range from 0 to 28K and second highest default rate is in annual income range from 28k to 56k.
12. All sub grades show equal percentages of Charged Off and Fully Paid loans. Sub grades individually does not have much impact on possibility of default.

Recommendations:

1. Loans with higher loan amounts are more prone to default (Charged Off), so all such loans should be scrutinized more than other loans.
   Verification process for such loans should be made more stringent.
   And if possible break that loan into multiple parts which would be passed after retreiving principal of previous part.
   Or it should be passed with higher collateral.
2. All of the following are indicators of possible higher loan amounts - 60 months term, Grades G(E & F), 'Small_Business' as Purpose, Home_Ownership = 'OTHER'.
3. Verification process should be improved because for verified loans the average loan amounts is high, the average interest rates are high and are more prone to being Charged Off as compared to other types of verification_status.
4. Higher interest rates are more prone to being Charged off, So increasing interest rates should be done only when it is absolutely necessary for business and that loan should be very thoroughly verified.
5. Interest rates increases as the number of installments increases, so loans with higher number installments should be more scruitinized before allowing them and customer should be tried to be convinced to avail a loan with low number of installments to avail low interest rates.

4. Conclusion:

The analysis reveals that loans with higher amounts, longer terms, lower grades, and higher interest rates are more likely to default. To mitigate this risk, it is recommended to tighten the verification process, particularly for high-value loans and loans with longer terms or higher interest rates. Additionally, loans with a larger number of installments should be scrutinized more closely, and borrowers should be encouraged to opt for loans with fewer installments to reduce risk. By implementing these measures, Lending Club can reduce defaults and improve the overall sustainability of its loan portfolio.

Authors:

1. Abhishek Raghuvanshi

2. Subhrabindu Khuntia