LENDING CLUB CASE STUDY UPGRAD 2023

SUBMITTED BY:

POONAN SAHOO

RICKY MOHAPATRA

OBJECTIVE

- The objective of the case study is to implement EDA technique using a real world business problem and understand the insights and present the insights along with supporting charts and figures.

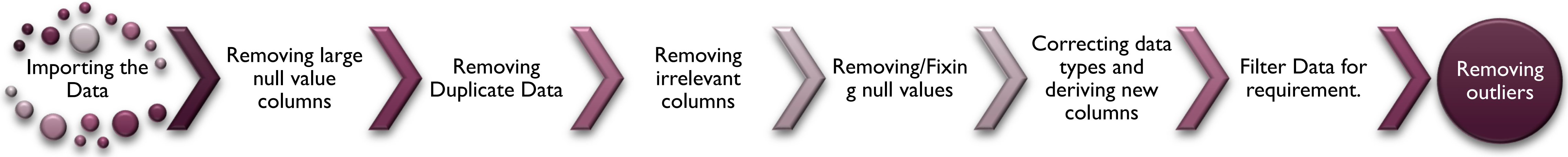
BUSINESS UNDERSTANDING

- The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables.

Dataset Details:

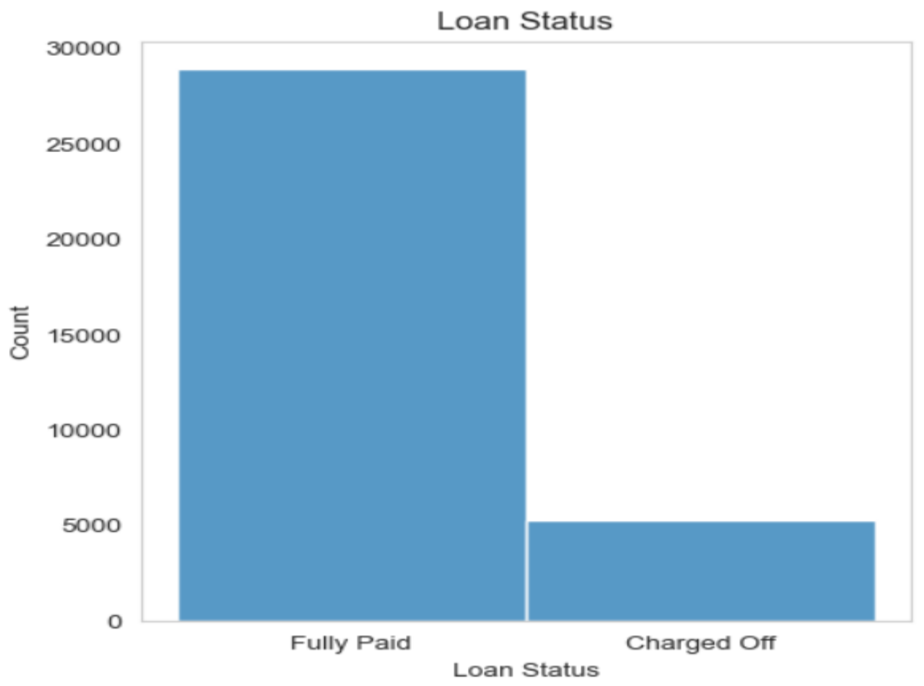
The data given below contains information about past loan applicants and whether they 'defaulted' or not. Data has details regarding approved loan not the rejected ones. It has 3 status of loan which is Fully Paid, Current and Charged-Off.

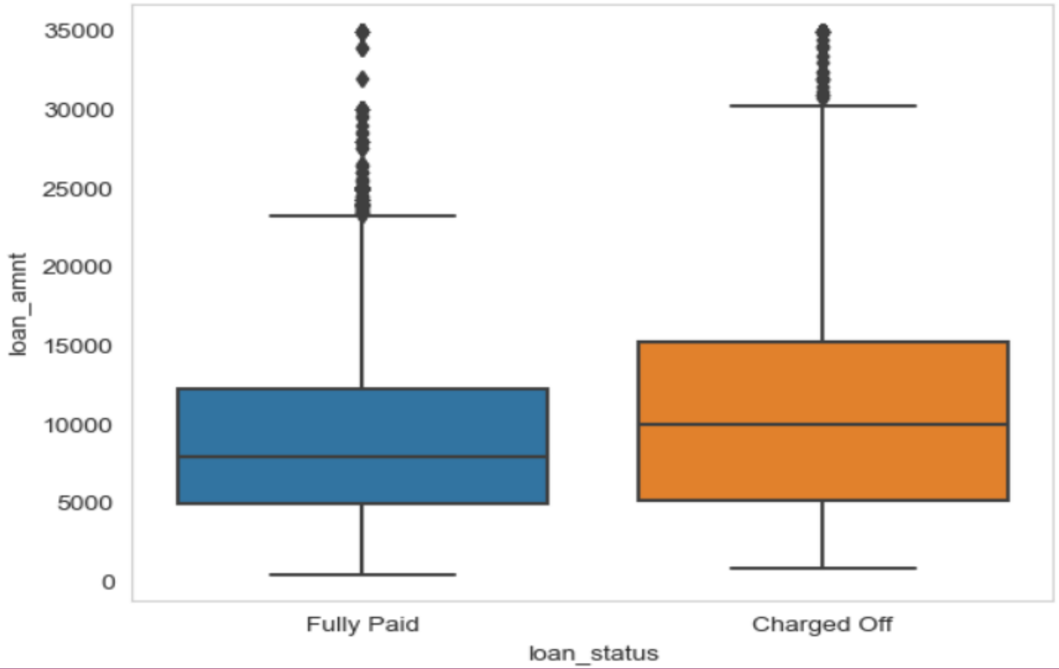
Data Clean-up and preparation process:



UNIVARIATE ANALYSIS

LOAN STATUS AND LOAN AMOUNT

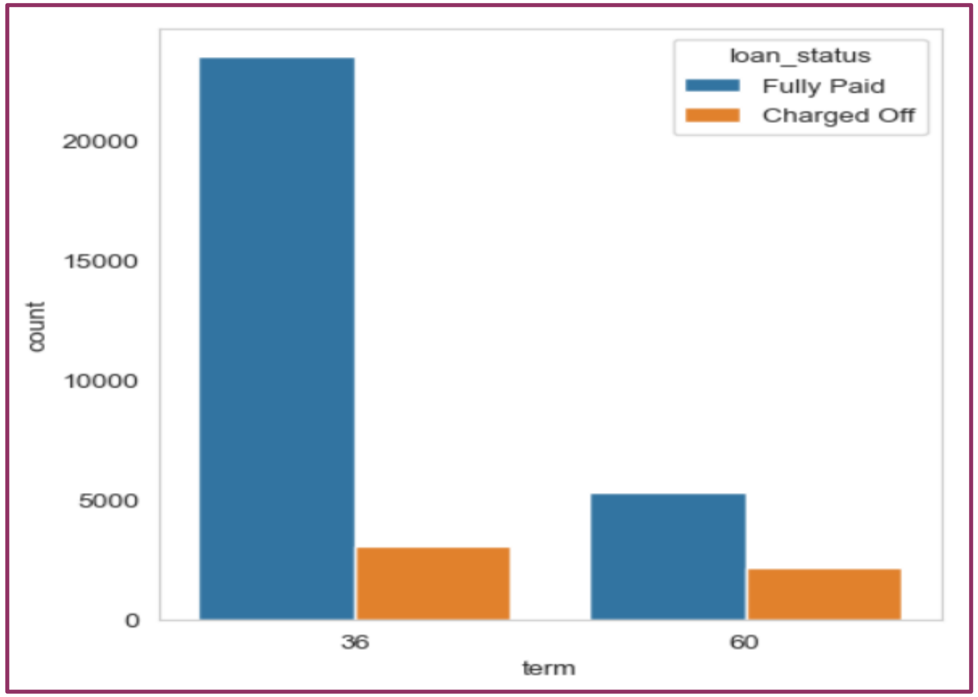


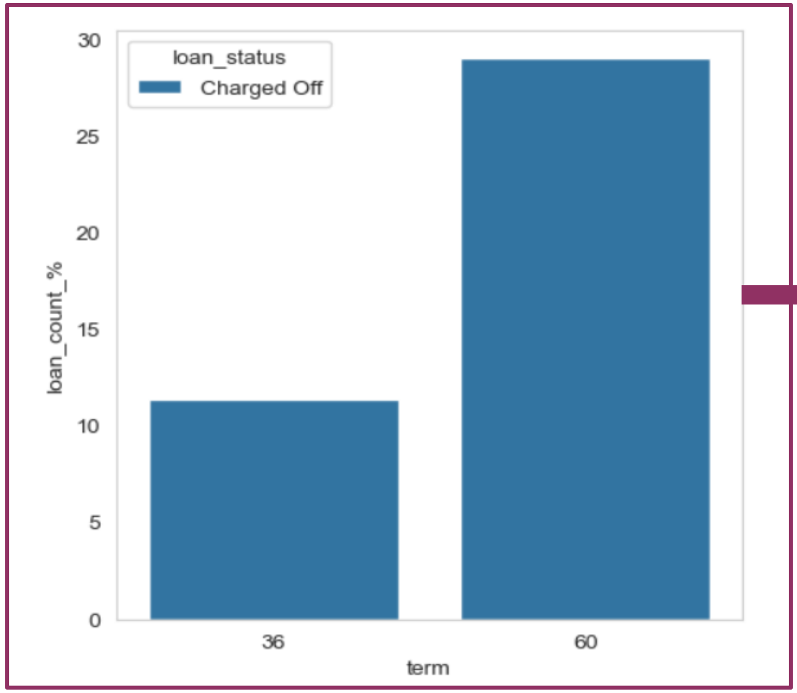


Loan Status: The number of charged off loan is much smaller(14.5%) compared to total count.

Loan Amount: It varies from 500 to 35000 with a median of 10000. Loan amount is majorly small and very few clients have taken large loans and higher is the loan amount higher is the chance of default

LOAN TERM





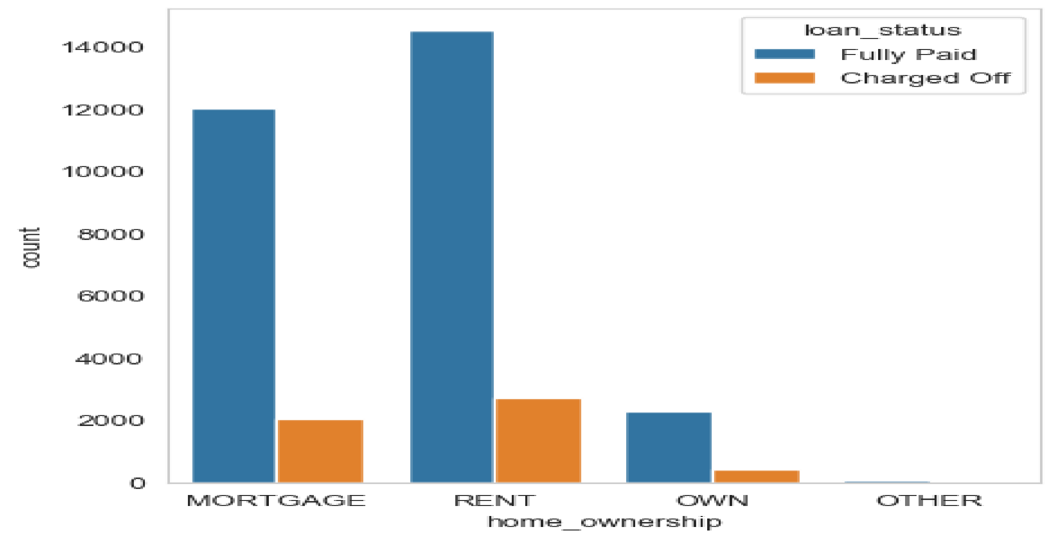
Loan Term: Loans with higher term(60 months) displayed higher charged-off loan count%. Indicates that loans with higher term are more likely to default

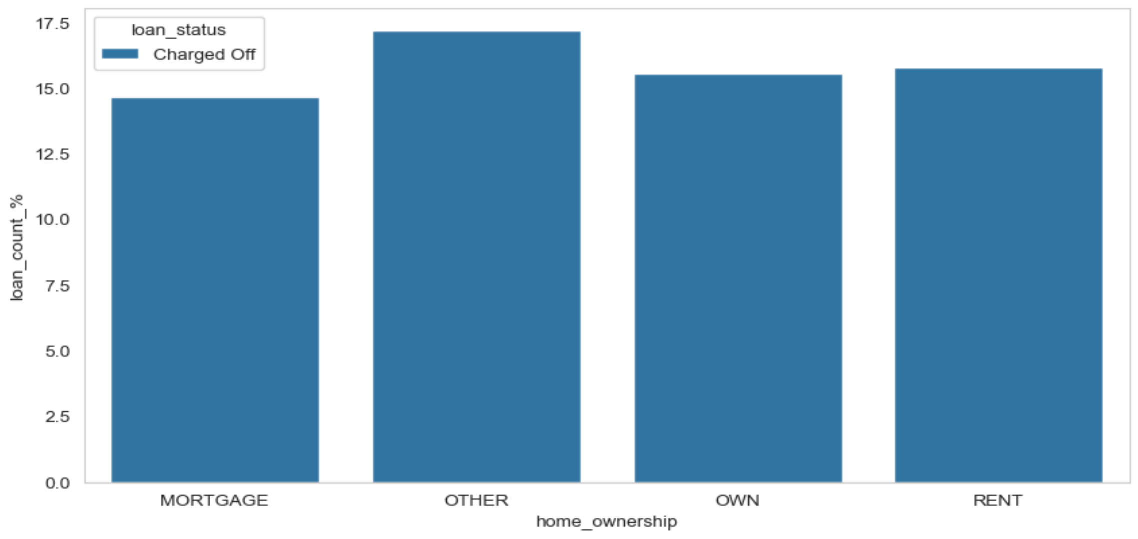
Loan Term: The Loans availed for 36 month term are much more than 60 months and have lower chance of defaulting.

term	loan_status	loan_count	total_loan_count	loan_count_%
36	Charged Off	3042	26614	11.430074
36	Fully Paid	23572	26614	88.569926
60	Charged Off	2185	7516	29.071315
60	Fully Paid	5331	7516	70.928685

Charged-off loan-count% = Count of charged-Off loans / (count of charged-off + count of fully-paid loans)

HOME OWNERSHIP TYPE

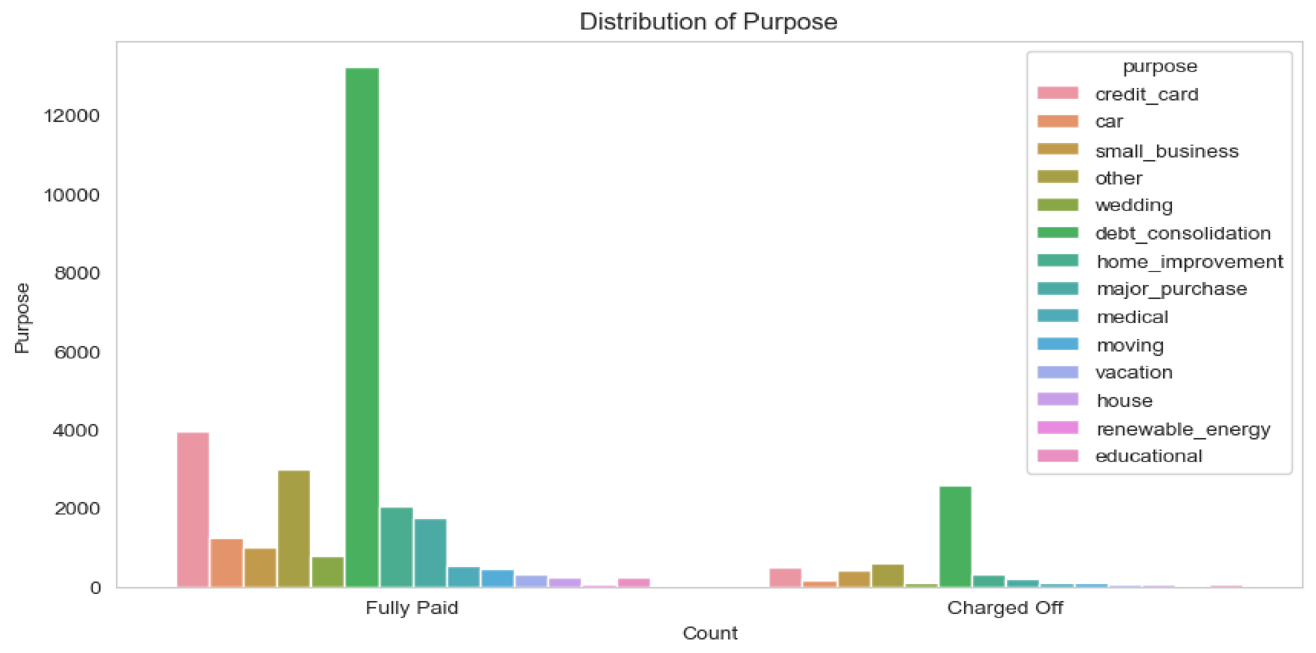




	home_ownership	loan_status	loan_count	total_loan_count	loan_count_%
0	MORTGAGE	Charged Off	2065	14087	14.658905
1	MORTGAGE	Fully Paid	12022	14087	85.341095
2	OTHER	Charged Off	16	93	17.204301
3	OTHER	Fully Paid	77	93	82.795699
4	OWN	Charged Off	420	2696	15.578635
5	OWN	Fully Paid	2276	2696	84.421365
6	RENT	Charged Off	2726	17254	15.799235
7	RENT	Fully Paid	14528	17254	84.200765

Home Ownership: While it is observed that for loans *under Rent and Mortgage ownership type*, the count of loans is higher however from the above table since there is not much difference between the charge-off loan-count% between categories, it is observed that the home ownership type does not have much relevance or impact on whether a borrower is likely to default and therefore we believe this variable is not useful to the analysis from a univariate standpoint.

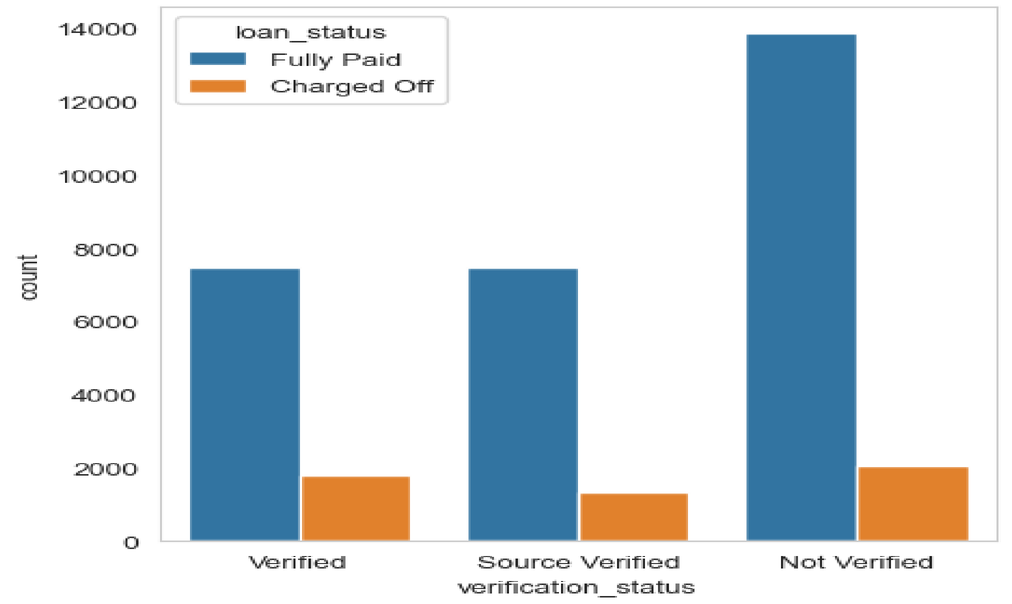
PURPOSE

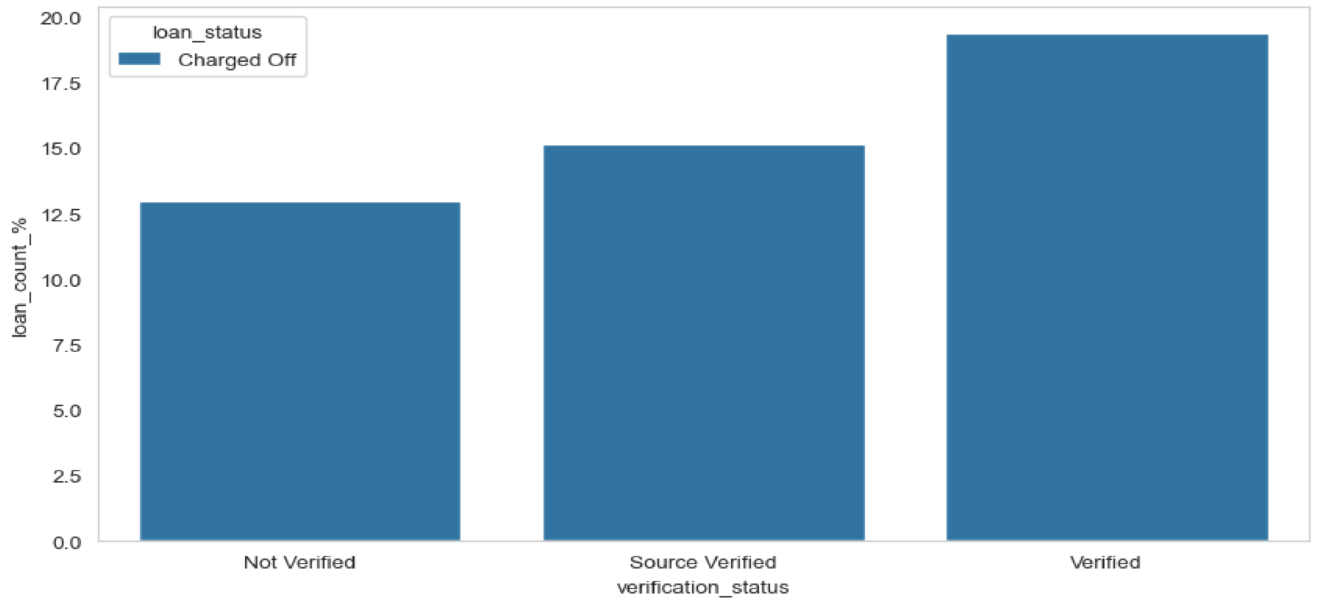


purpose	loan_status	loan_count	total_loan_count	loan_count_%
car	Charged Off	158	1426	11.079944
car	Fully Paid	1268	1426	88.920056
credit_card	Charged Off	499	4442	11.233679
credit_card	Fully Paid	3943	4442	88.766321
debt_consolidation	Charged Off	2570	15799	16.266852
debt_consolidation	Fully Paid	13229	15799	83.733148
educational	Charged Off	54	307	17.589577
educational	Fully Paid	253	307	82.410423
home_improvement	Charged Off	311	2360	13.177966
home_improvement	Fully Paid	2049	2360	86.822034
house	Charged Off	51	309	16.504854
house	Fully Paid	258	309	83.495146
major_purchase	Charged Off	215	1987	10.820332
major_purchase	Fully Paid	1772	1987	89.179668
medical	Charged Off	103	628	16.401274
medical	Fully Paid	525	628	83.598726
moving	Charged Off	90	543	16.574586
moving	Fully Paid	453	543	83.425414
other	Charged Off	590	3572	16.517357
other	Fully Paid	2982	3572	83.482643

Purpose: It is observed that around 49% of the charge off loans have originated under the purpose of '**debt consolidation**' along with that we also observe higher charge off percentage (16.3% of charge off loans) which indicates that 'debt consolidation' associated loans have a higher likeliness to default.

VERIFICATION STATUS



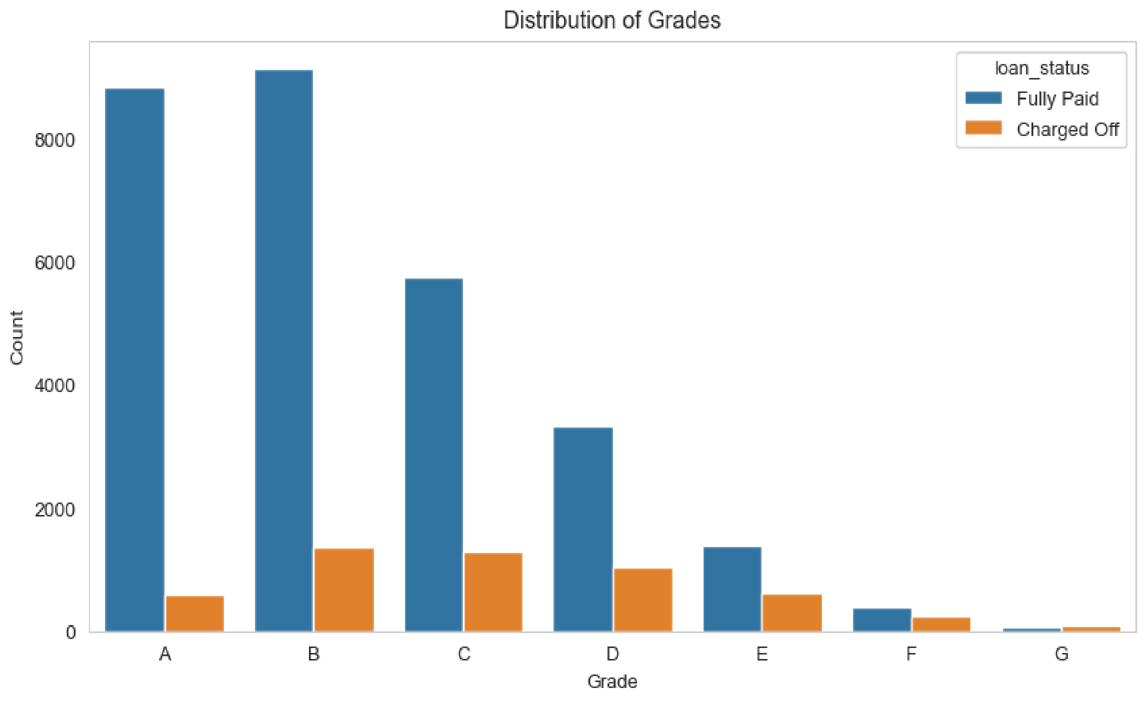


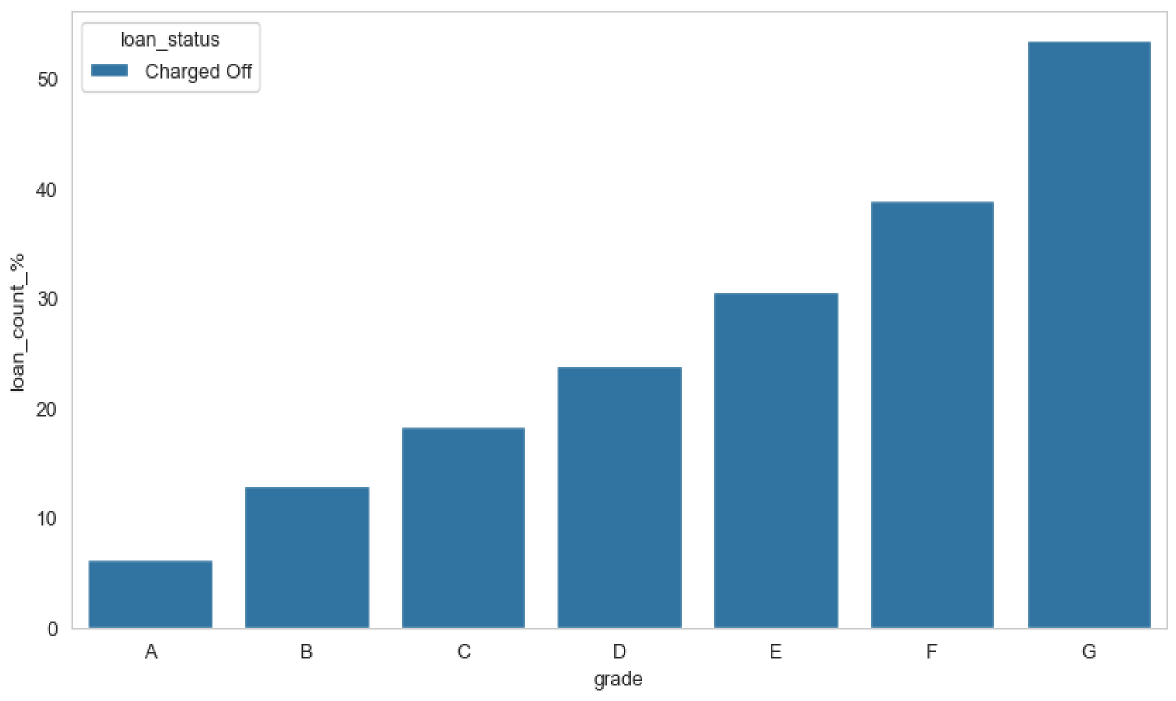
verification_status	loan_status	loan_count	total_loan_count	loan_count_%
Not Verified	Charged Off	2076	15987	12.985551
Not Verified	Fully Paid	13911	15987	87.014449
Source Verified	Charged Off	1343	8839	15.194026
Source Verified	Fully Paid	7496	8839	84.805974
Verified	Charged Off	1808	9304	19.432502
Verified	Fully Paid	7496	9304	80.567498

Verification Status: Verified borrowers are less than the unverified (9300 vs 16000 approx.) however the verified borrower category has displayed larger charge off percentage (19.5%) which is an indication that this category (verified by LC) is more likely to default based on the dataset provided.

This might be an indication that borrowers verified by LC do not come under much scrutiny which is why there is a chance that these borrowers might pass under the radar and thereby become more highly likely to default.

GRADES

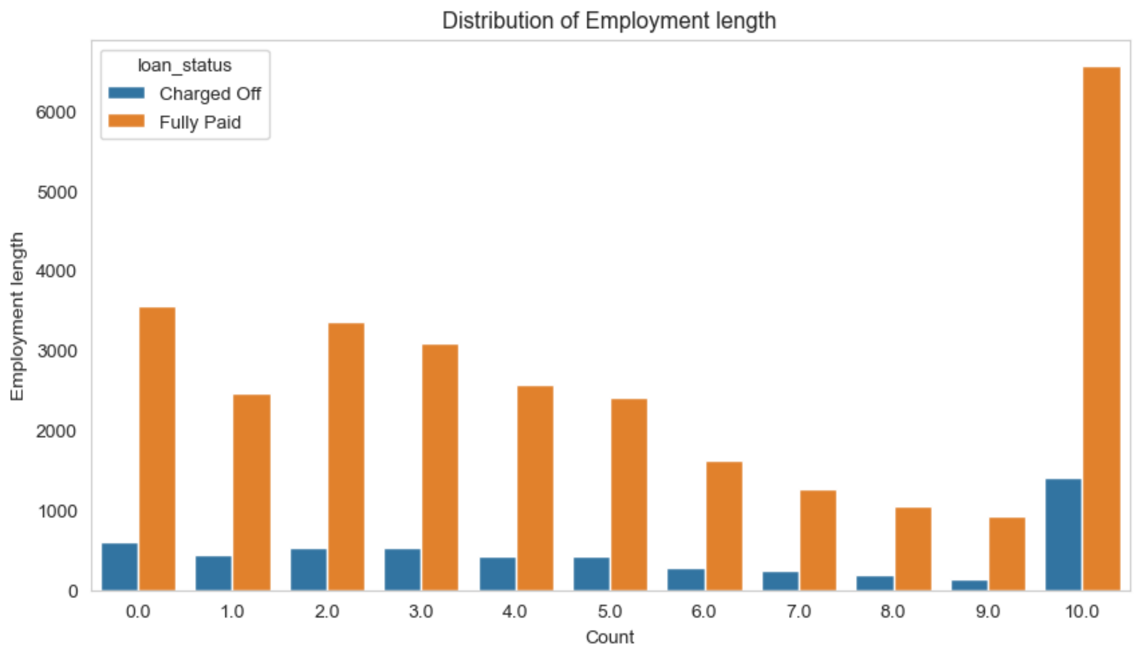


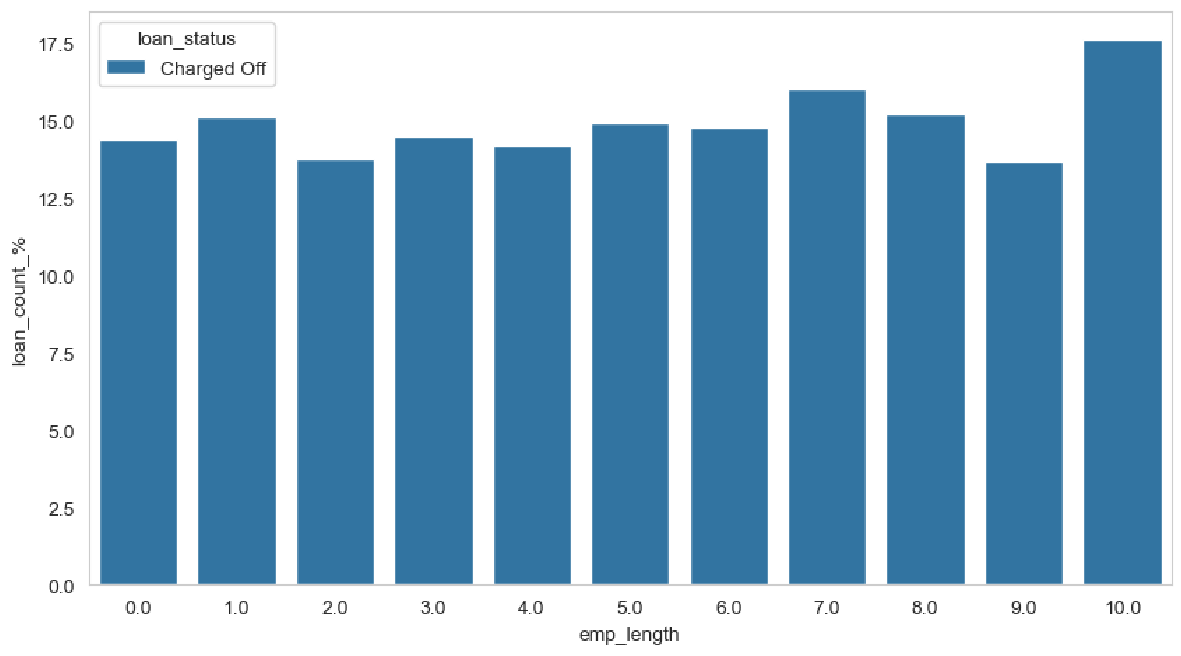


Grades: Most of the Charged-off are observed in 'B' and 'C' Grades, almost 50% of the Charge-off Population, but it is also observed that number of loans in lower sub-grades are more than its higher sub-grades, except for Grade C, where it is a opposite trend.

It can be inferred based on the observation above that loans belonging to sub prime grades i.e. as we move from A,B towards G we observe that the charge off percentage increases which indicates the loans with sub-prime grades are much more riskier and much more likely to default

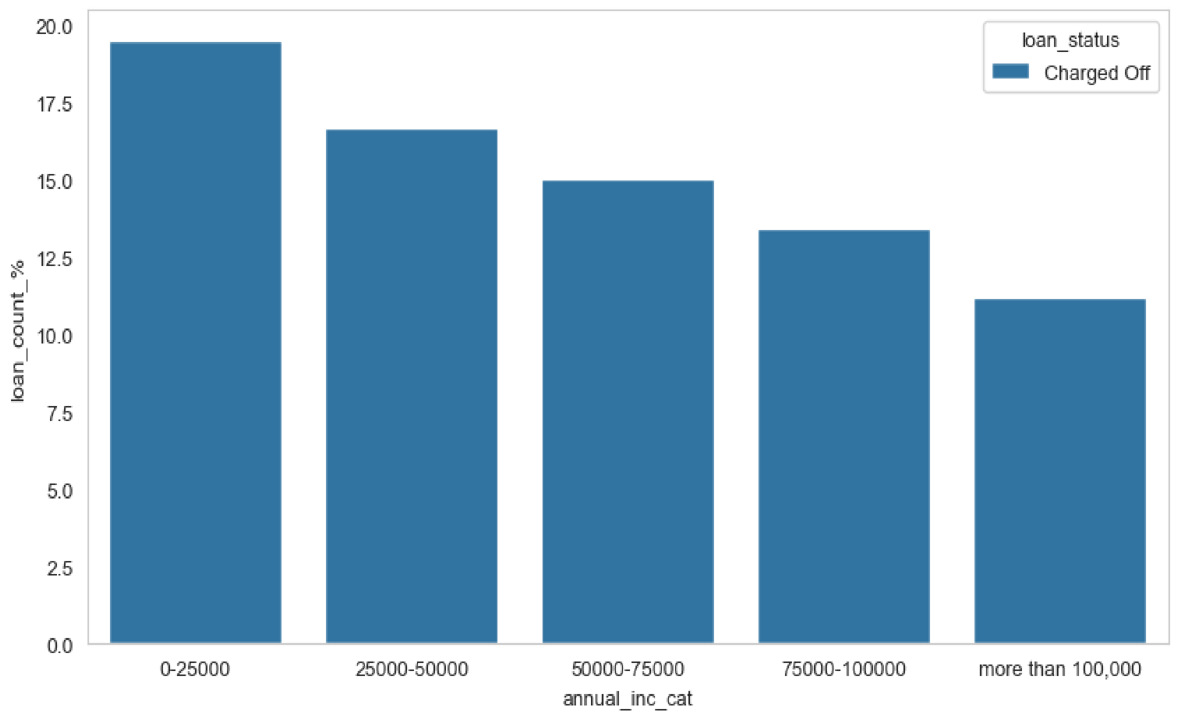
EMPLOYMENT LENGTH

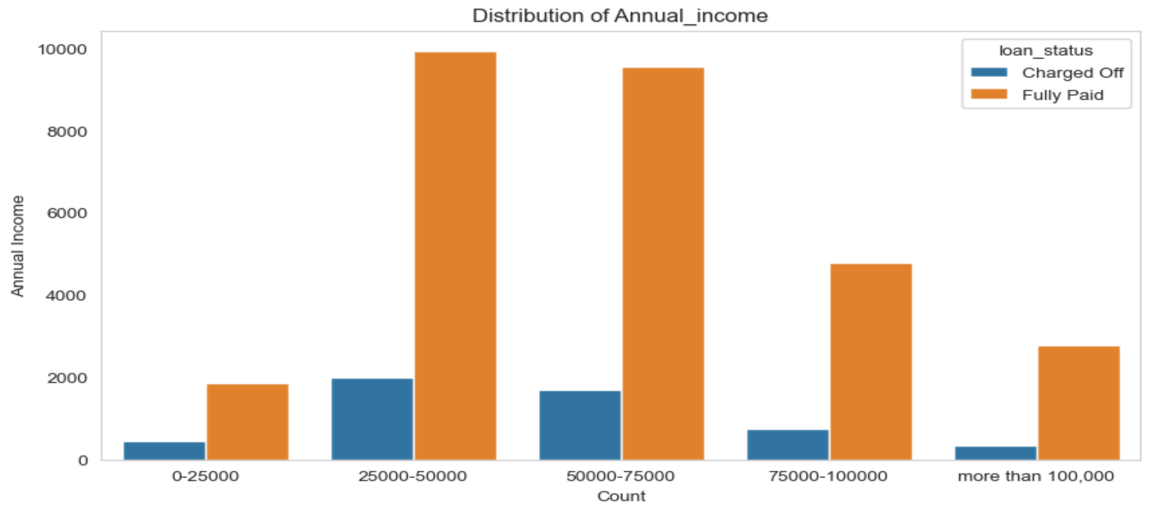


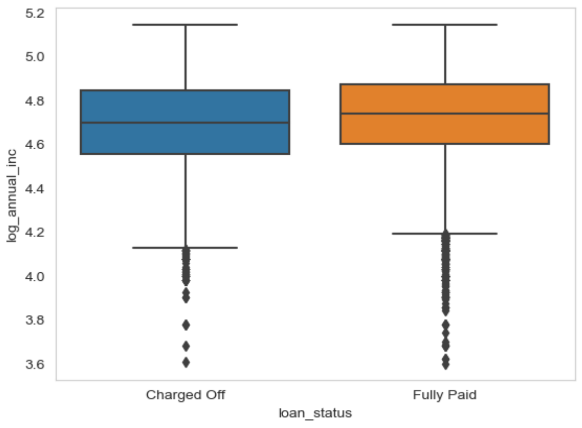


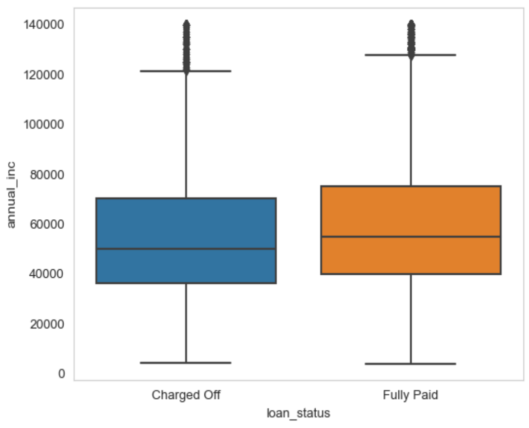
- **Employment Length**: It is observed that borrowers with employment length >10yrs have a higher no of loans(fully paid and charge off loans) as well as we notice the charge off% to be marginally higher for borrowers with employment length >10yrs.
- Based on the above graph we can conclude that employment length can not be considered as a primary indicator given it is unable to provide any concrete evidence to suggest the influence of this feature towards defaulting.

ANNUAL INCOME



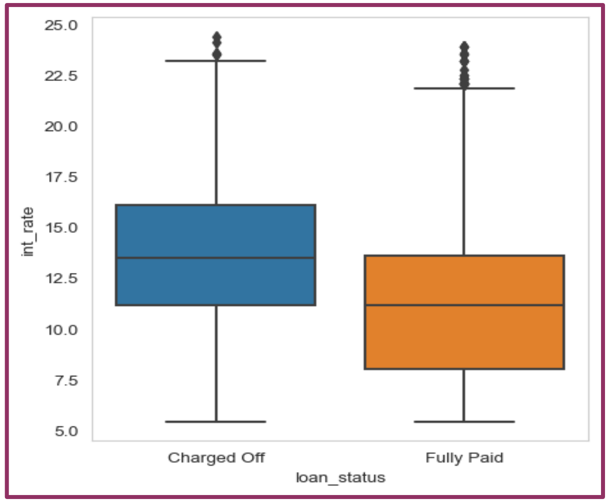


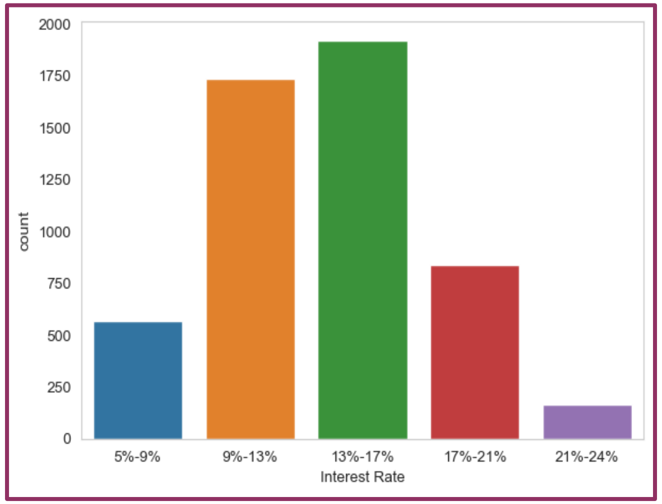


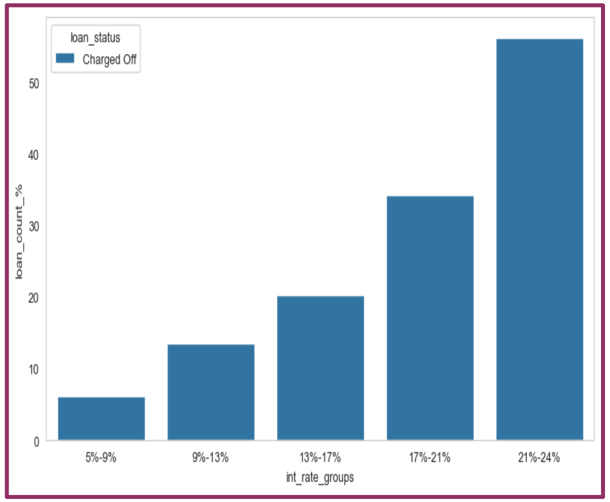


- **Annual Income:** The Charge-Off loan count% decreases with increasing annual income which indicates that with higher annual income accounts are less likely to default

INTEREST RATE AND DTI (DEBT TO INCOME)



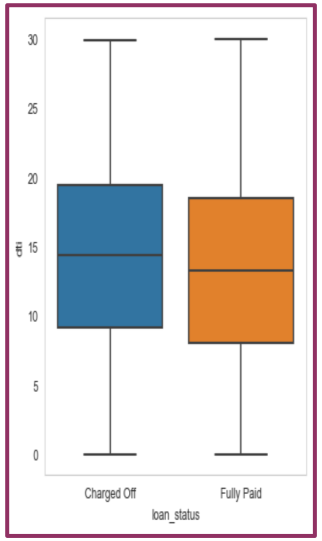


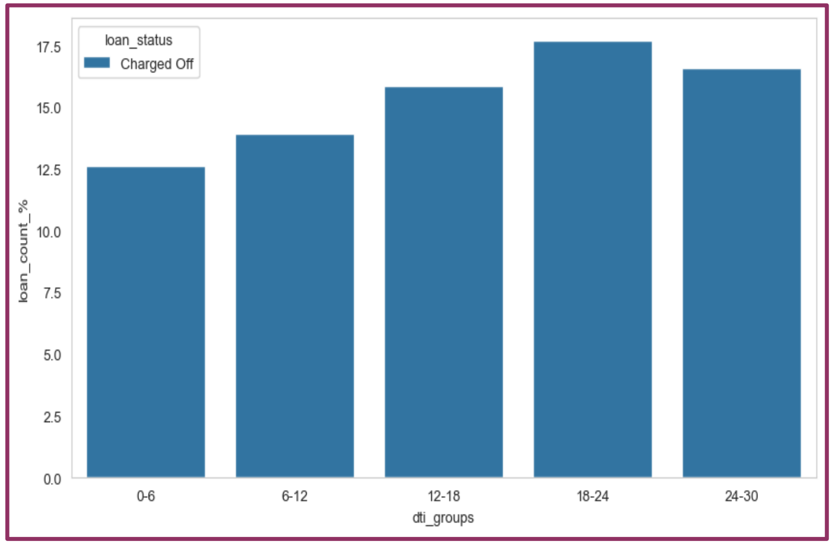


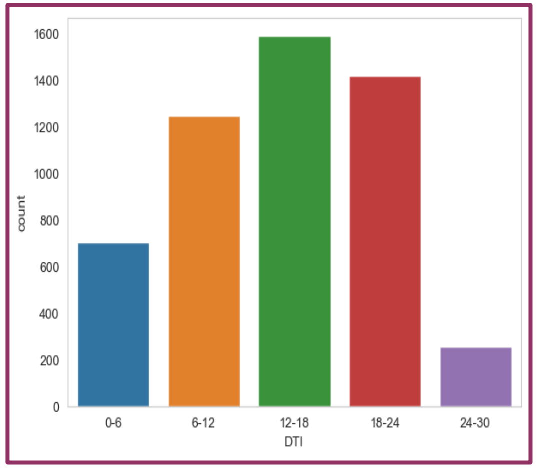
Interest Rate:

'Charged Off' accounts are charged higher Interest Rates when compared to 'Fully Paid' accounts, almost by 3%.

However since this information regarding interest rate is not available before a borrower is onboarded therefore this variable will not add any value to the overall analysis which is why we can ignore this feature in the context of univariate analysis.





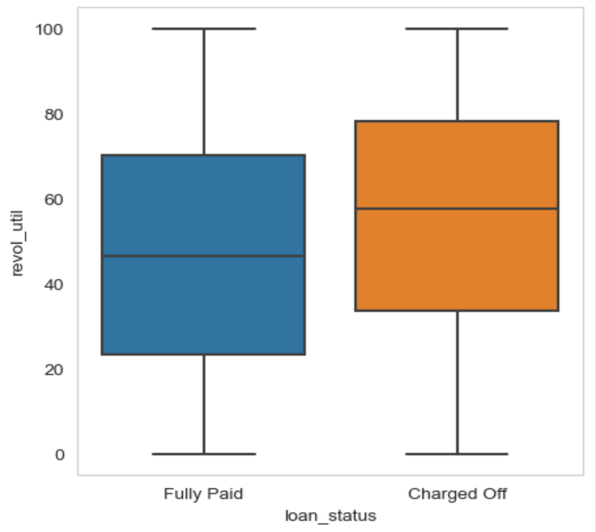


DTI : borrowers with higher debt to income ratios are likely to default more. These are based on the below noted observations:

1. Charged Off loans have marginally higher dti when compared to Fully Paid loans as per the box plot

2. The charged off percentages increased with increase in dti values since charge off% is higher for dti groups 12-18, 18-24, 24-30.

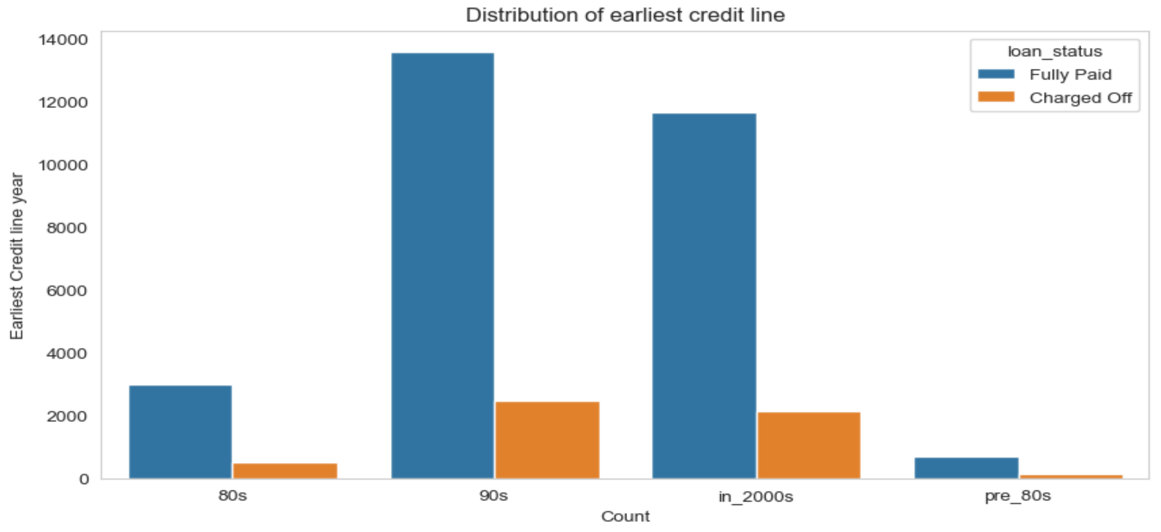
REVOLVING UTILISATION RATE AND EARLIEST CREDIT LINE

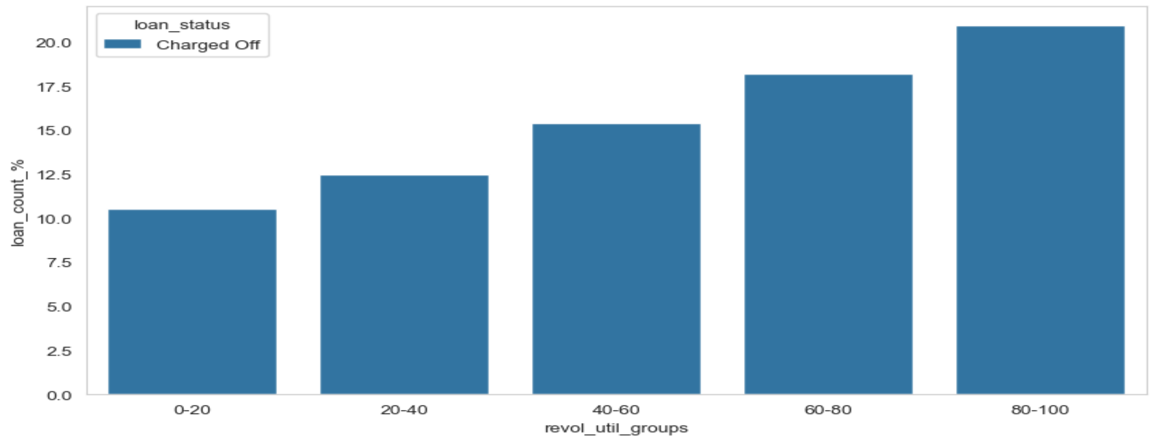


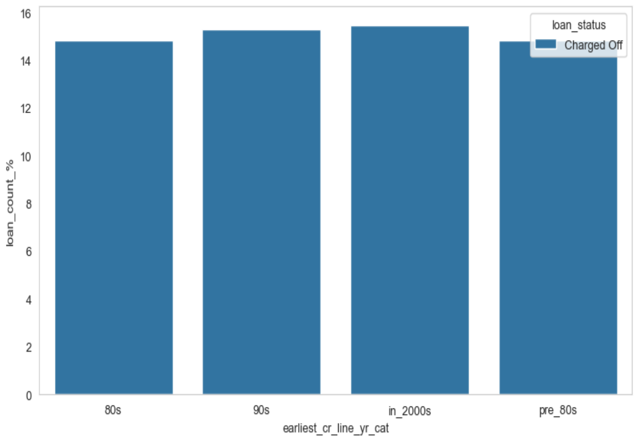
Revolving utilization rate:

The borrowers associated with higher revolving utilization rate are more likely to default and are riskier given that the charge-off loan-count% increased with the increase in revol-util rates.

This is a decent indicator for identifying loans that are likely to default.



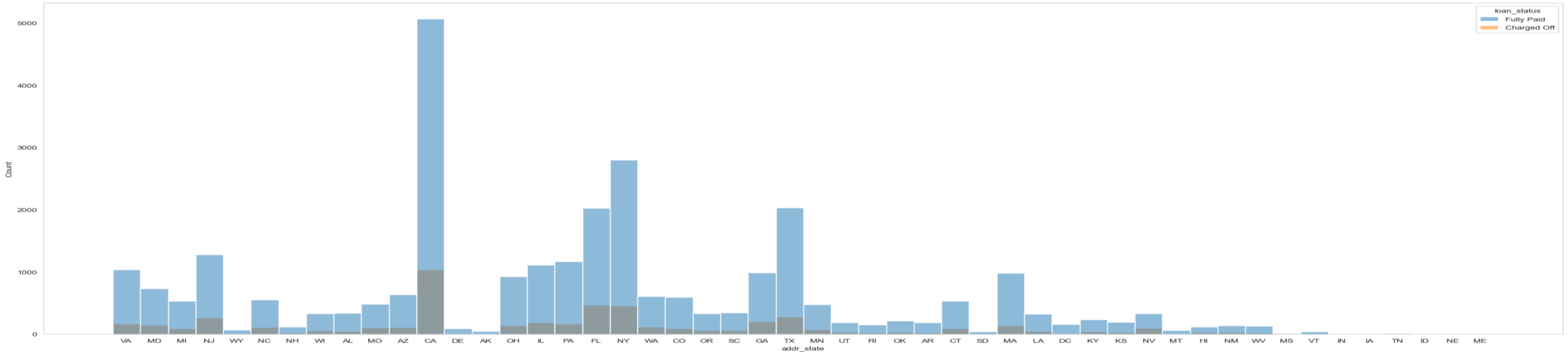




Earliest credit line:

The insights are inconclusive since the earliest credit line does not impact the no of charged off loans as not much difference is observed for the charged off loan count% in the above categories and therefore we do not consider this feature to be beneficial to the analysis in the context of univariate exercise

ADDRESS STATE



Top 15 states with highest charged-off loan count%:

	addr_state	loan_status	loan_count	total_loan_count	loan_count_%
53	NE	Charged Off	3	5	60.000000
61	NV	Charged Off	98	430	22.790698
77	SD	Charged Off	12	58	20.689655
25	ID	Charged Off	1	5	20.000000
22	HI	Charged Off	28	147	19.047619
18	FL	Charged Off	471	2495	18.877756
0	AK	Charged Off	12	65	18.461538
45	MO	Charged Off	107	594	18.013468
59	NM	Charged Off	30	169	17.751479
69	OR	Charged Off	68	400	17.000000
8	CA	Charged Off	1036	6104	16.972477
57	NJ	Charged Off	262	1546	16.946960
20	GA	Charged Off	200	1193	16.764459
38	MD	Charged Off	147	880	16.704545
51	NC	Charged Off	111	666	16.666667

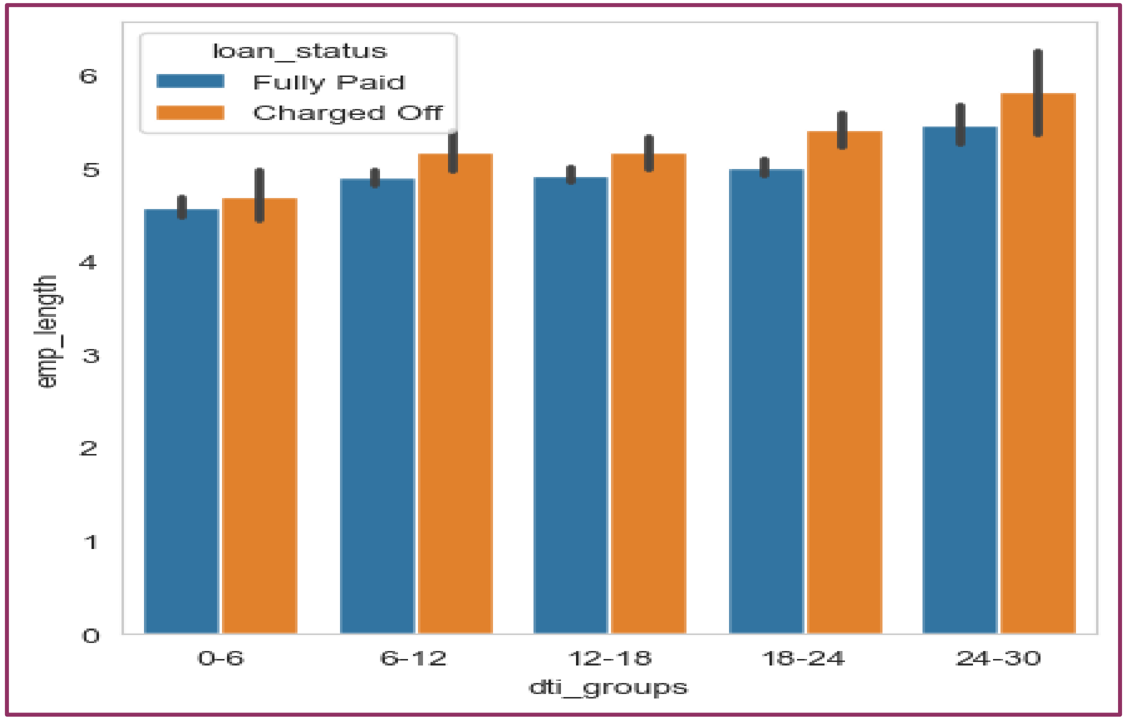
Inference: It is observed that majority of the borrowers are from the large urban cities like California, New york, Texas, Florida etc which shows propensity of the urban crowd to avail more loans and thereby the propensity to avail loans along with higher expense may lead to higher defaults.

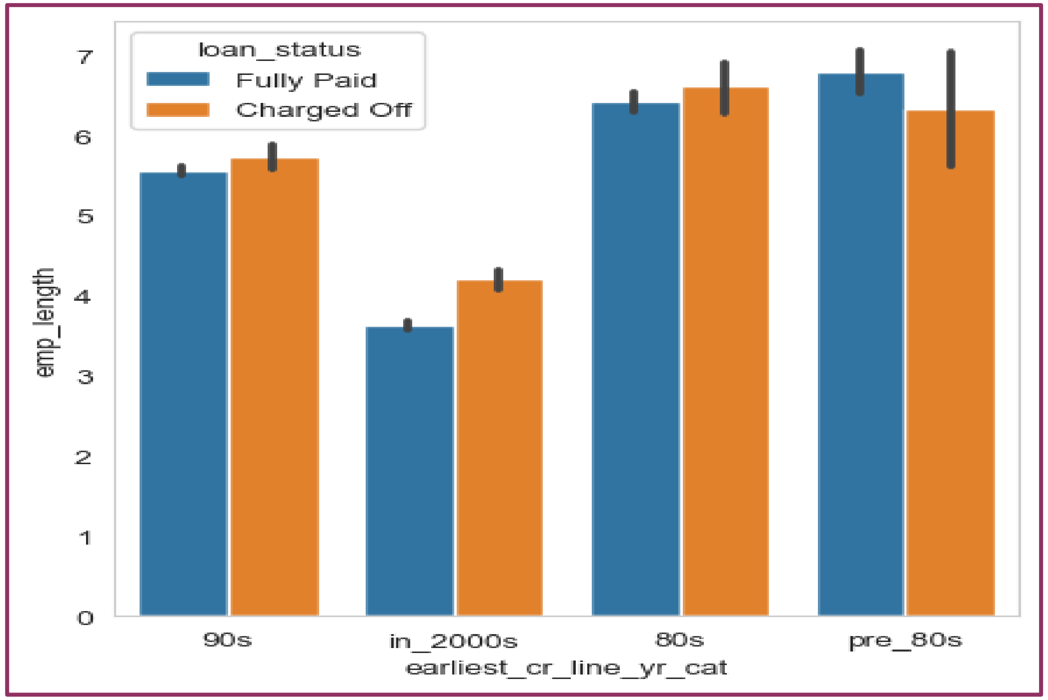
If we observe the charge off loan count% table it is noted that among the top 15 cities by charged-off loan-count% - Florida, California and New Jersey showcase higher charge off percentage along with higher no of loans(fully paid and charged off).

This to some extent indicates that loan approval of these major cities should go through more scrutiny given that the propensity to spend is more in such cities thereby more likely to default.

BIVARIATE ANALYSIS

1. EMPLOYMENT LENGTH VS DTI & EARLIEST CREDIT LINE



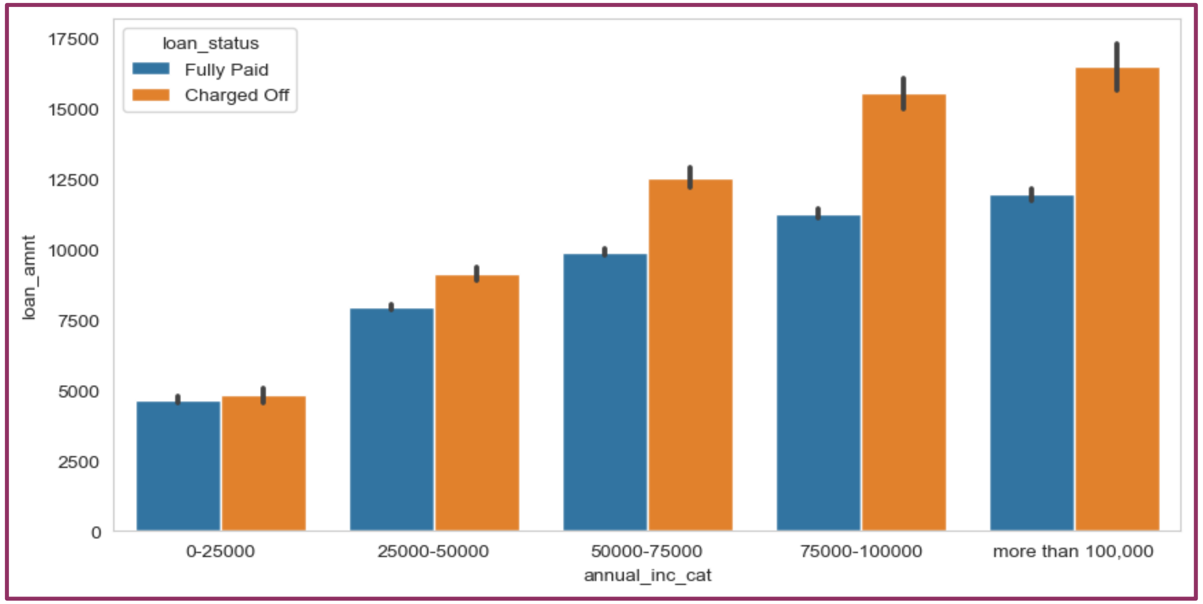


Employment Length vs DTI:

Both fully paid and charged off loans the employment length increases with the increase in dti which indicates that people tend to take up higher debt in comparison to their income with increase in their employment duration which ultimately leads to higher likeliness of defaults. Therefore DTI along with employment length becomes a crucial indicator for identifying charged off loans.

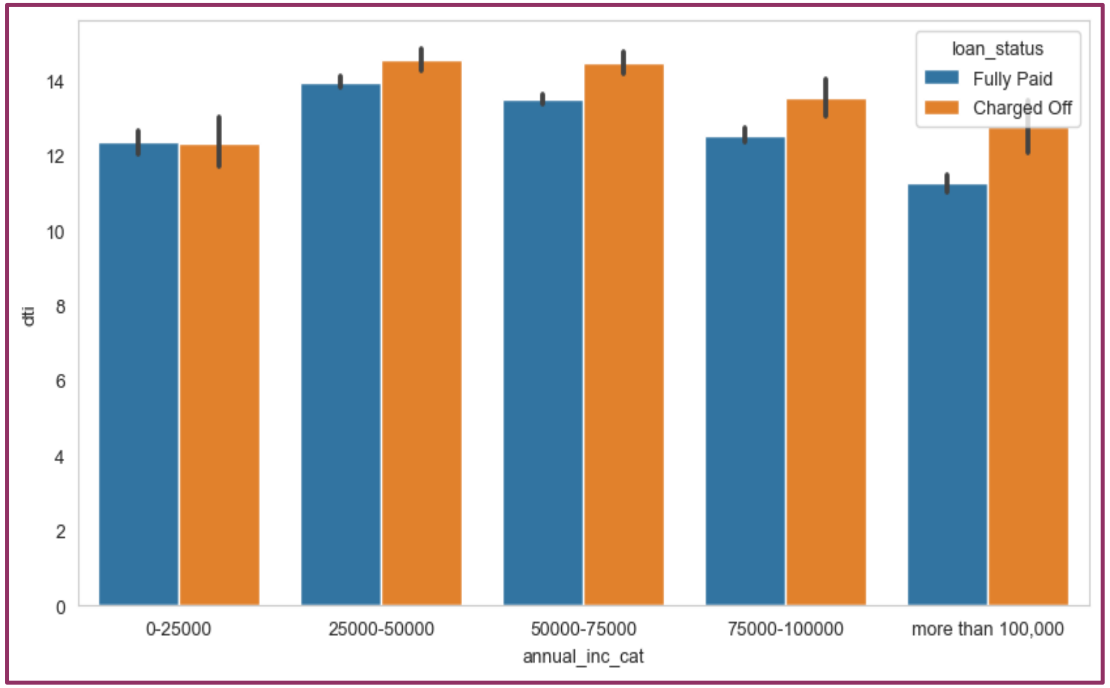
Employment Length vs Earliest credit line year: As we move from the pre80s(before year 1980) to the in2000s(year 2000 till present), the spread for employment length for borrowers defaulting has increased consistently through each category. This might be an indication that over time from before 1980 to the present, more no of borrowers with higher employment length have started to default and the gap has now reversed. While in pre80s the less experienced group used to default, this particular trend might have reversed in the recent times as the propensity to spend and acceptance of the credit culture has taken over.

ANNUAL INCOME VS LOAN AMOUNT & DTI



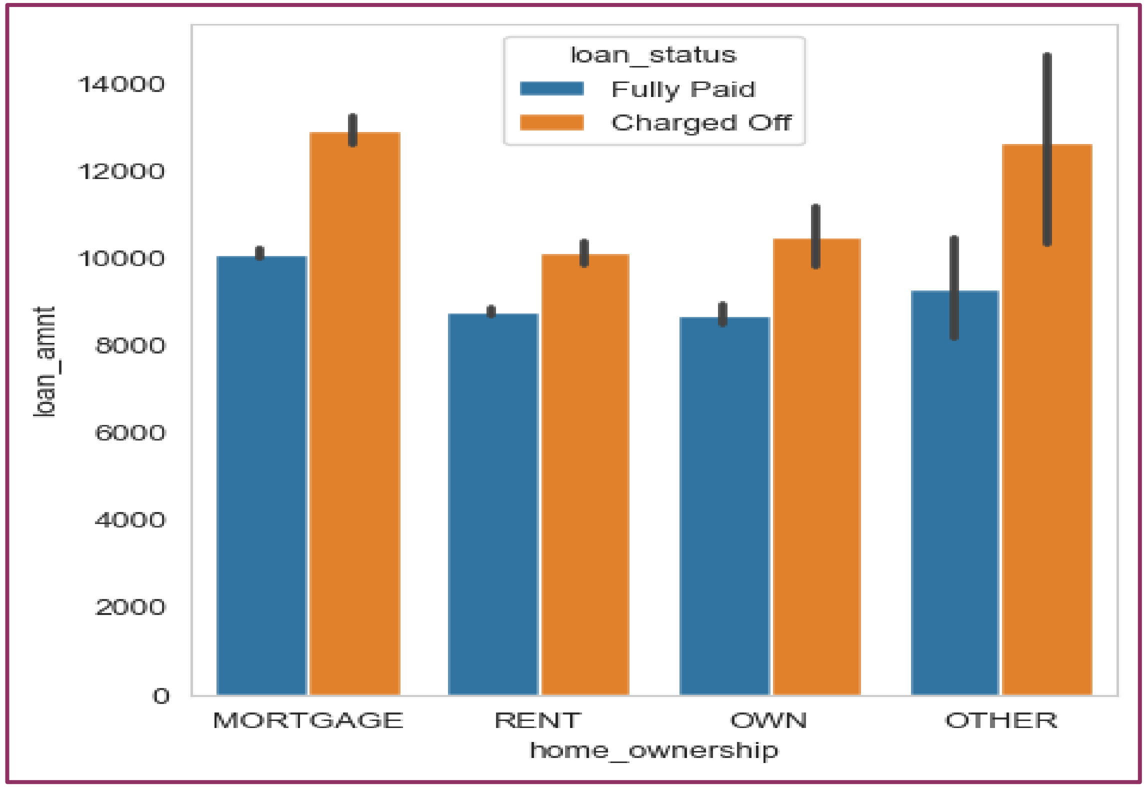
Annual income vs Loan amount: Employees with higher annual income got the loan approved for a amount higher than what was appropriate -1The loan amount is higher for the borrowers who defaulted

2.The above plot provides an interesting insight that with higher increasing annual income the disparity between loan amount for fully-paid and charged-off has increased. This would indicate that there are majority of cases where higher loan amounts have been sanctioned on the basis of higher annual income which has led to an increase in the no of charged-off loans. Approving higher Loan amount on the basis of high annual income might not be prudent.

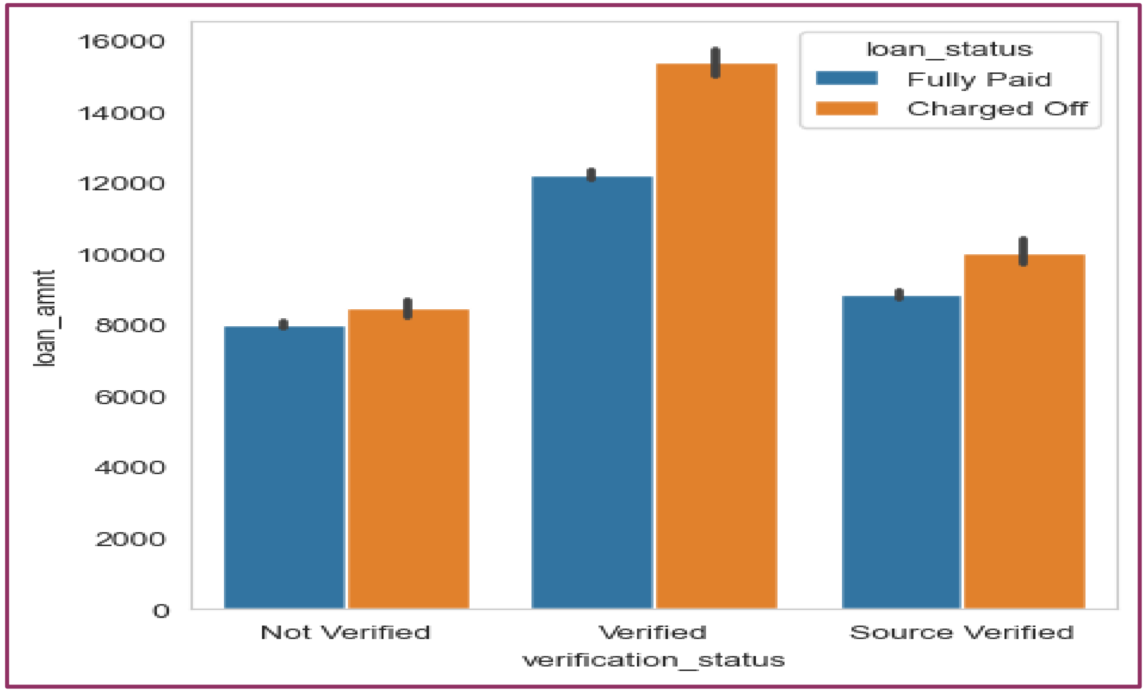


Annual income vs DTI: As per the univariate analysis, it has already been established that DTI and annual income and strong indicators for risk and probability of default. In combination we notice that the borrower groups with income (25000 -75000) showcase higher DTI compared to other groups. While this might not be a strong indicator however it can still be kept in consideration to identify the riskiness of borrowers

LOAN AMOUNT VS HOME OWNERSHIP & VERIFICATION STATUS



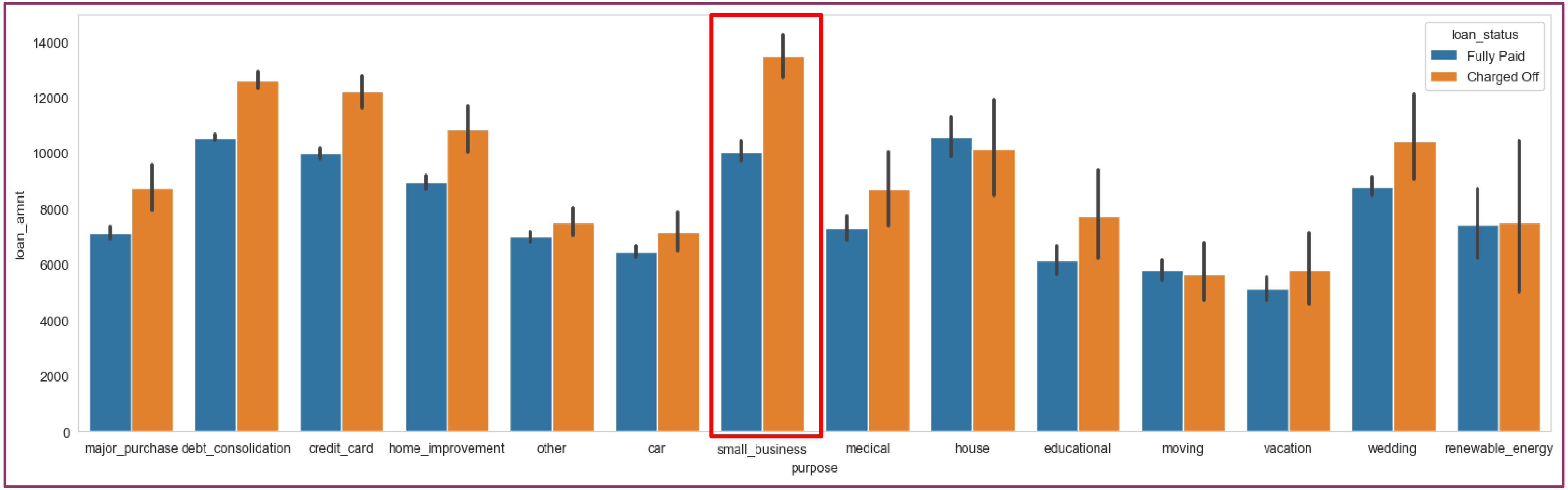
Loan Amount vs Home Ownership: We notice a larger spread and higher loan amount associated with defaulted loans compared to fully paid loans under the Mortgage and Other home ownership categories.



Loan Amount vs Verification Status: Borrowers verified by LC are have been provided higher loan amounts. Additionally it can be noted that the charged-off loans are associated with higher loan amount. This to some extent indicates that verified borrowers tend to get higher loan amounts and thereby the likeliness to default also increases if we refer to our univariate analysis with loan amount and verification status separately.

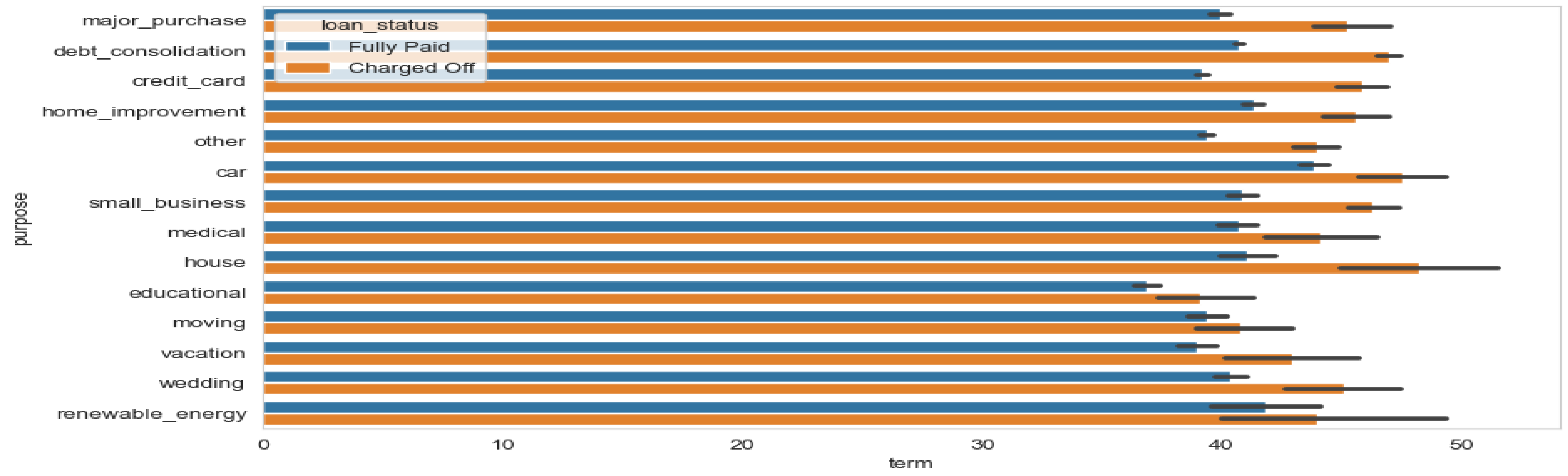
.

PURPOSE VS LOAN AMOUNT



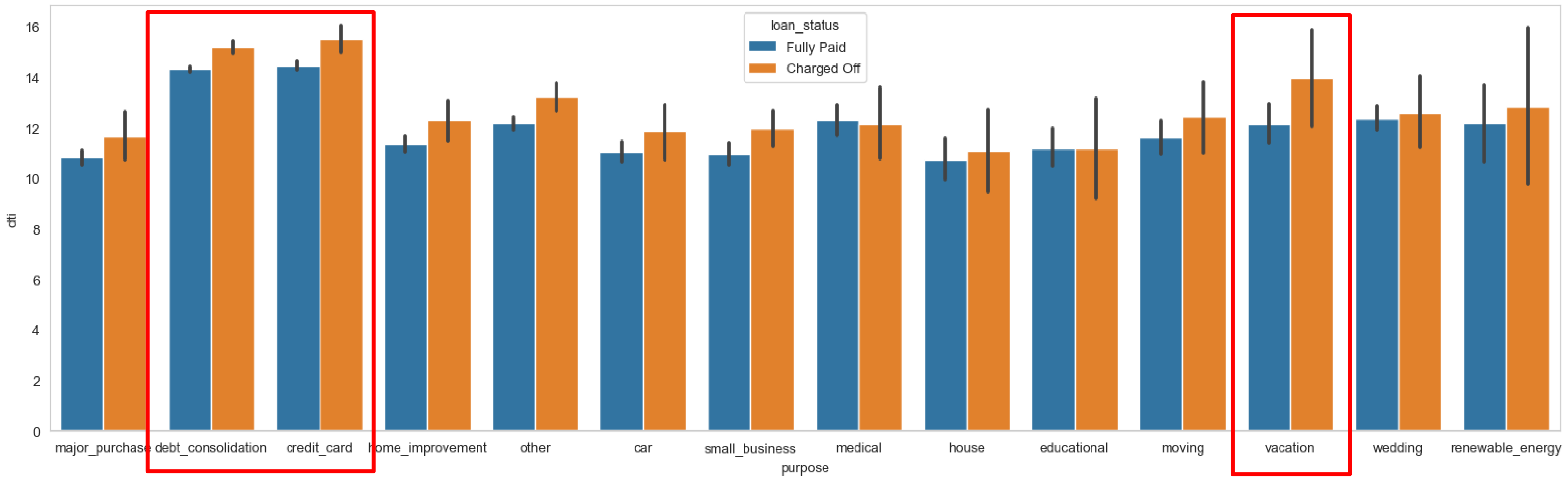
Loan Amount vs Purpose: For ***small business loans,*** the defaulted loans have a considerably higher loan amount associated compared to the fully paid loans. This can be an indication that there might be a need for limit setting for small business purpose loans which might be riskier given the purpose for which the loan is being used.

PURPOSE VS LOAN TERM



Loan term vs Purpose: For ***housing and debt consolidation*** purposes, the defaulted loans have a considerably higher loan term compared to fully paid loans. This to some extent indicates that for housing and debt consolidation purposed loans are more likely to default with increasing loan term.

PURPOSE VS DTI



Loan term vs Purpose: Loans for the purpose for ***debt consolidation, credit card and vacation*** are mostly availed by borrowers having higher debt to income ratio. Since higher DTI indicates more likeliness to default, it is important that based on this insight - such loans availed for purpose of debt consolidation, credit card and vacation should go through more approval processes to ensure the repayment capability of the borrower.

FINAL RECOMMENDATIONS

Major Driving factor which can be used to predict the chance of loans defaulting and reducing credit losses:
1. Annual Income, 2.DTI , 3.Verification Status, 4. Grades , 5. Loan Amount

There is a higher probability of loans defaulting in the scenario where:
1.Higher loan amounts (>10,000) are approved on the basis of borrowers having higher annual income or longer duration of employment
2.Higher loan amounts are approved for verified borrowers
3.Loans approved for borrowers with sub prime grades (Lower than C to G)
4.loans are approved for Borrowers with annual income in the range of 25000-75000 which showcase higher DTI (greater than 14)
5.Borrowers are from California, New Jersey and Florida
6.Higher loan amounts(>10,000) are sanctioned for loans towards small business purpose
7.Loans are approved for purpose of debt consolidation and housing with higher term (>4years)
8.Loans are approved for purpose of debt consolidation, credit card and vacation to borrowers with higher DTI (>14)
9.Higher term loans (60 months) are given out

THANK YOU