Lending Club Case Study

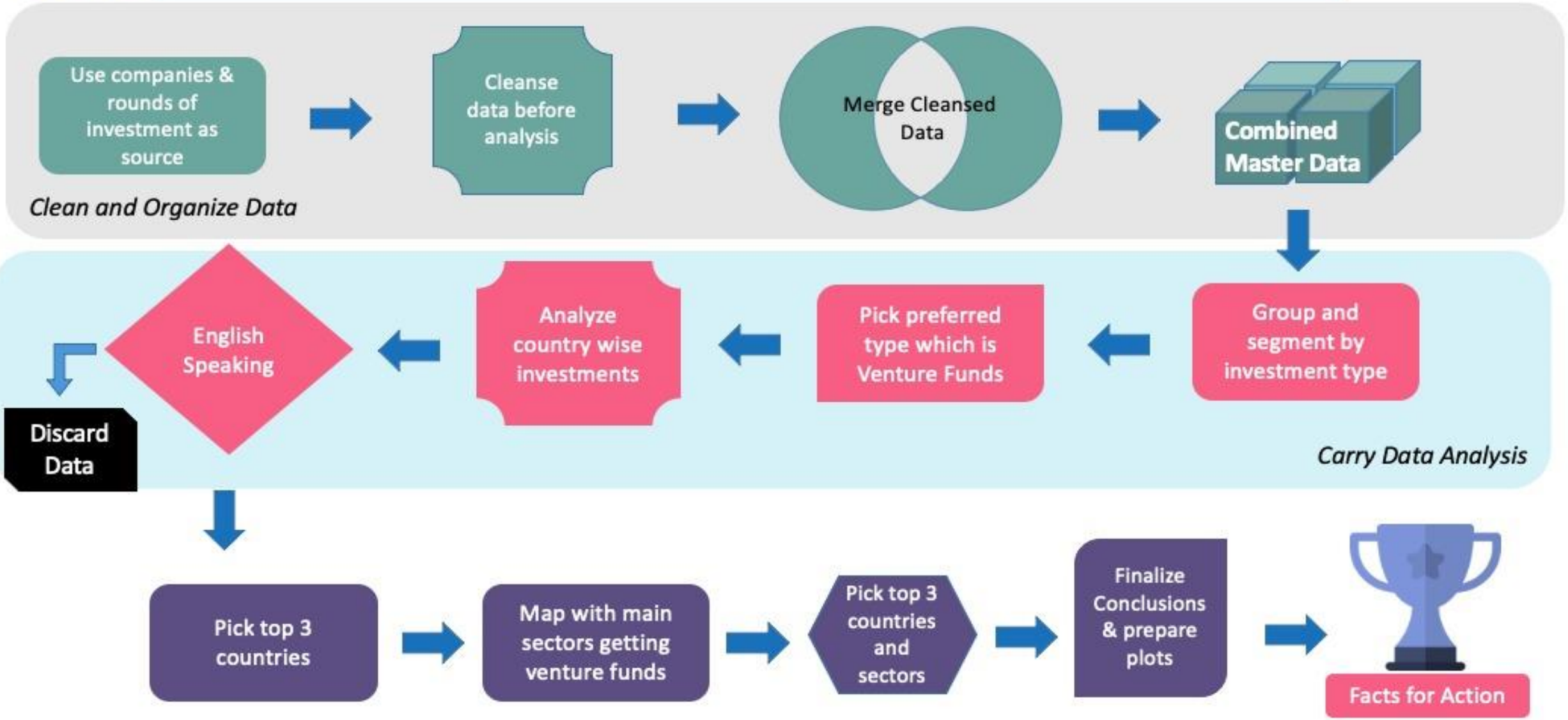
Name: RAJU SHANKAR VITKAR

Lending club Case Study Strategy

Company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Approach to solving the problem



Problem Analysis

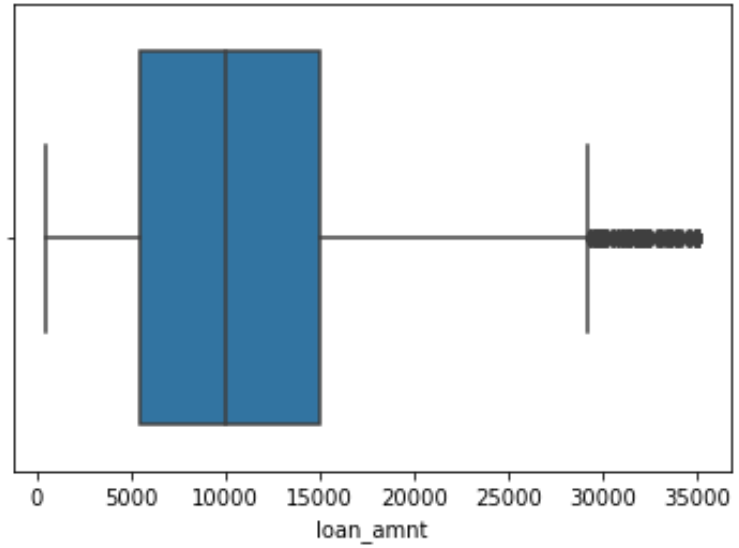
- Person applies for a loan, there are **two types of decisions** that could be taken by the company:

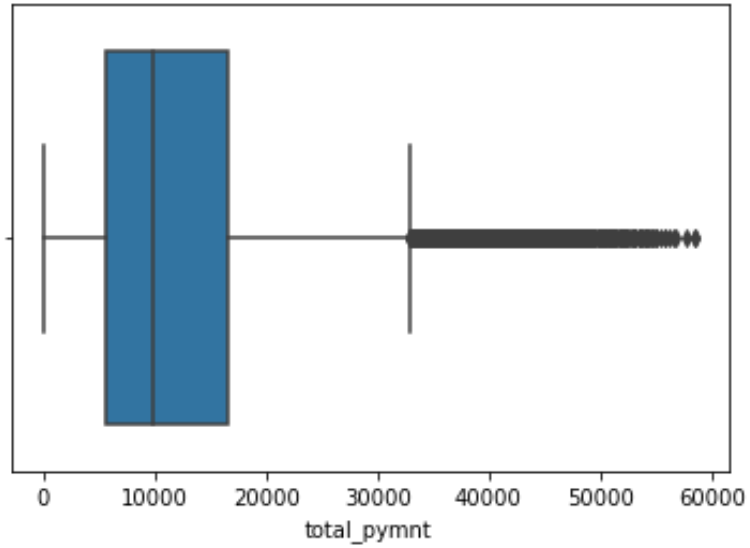
1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
    - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
    - **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
    - **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan
2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

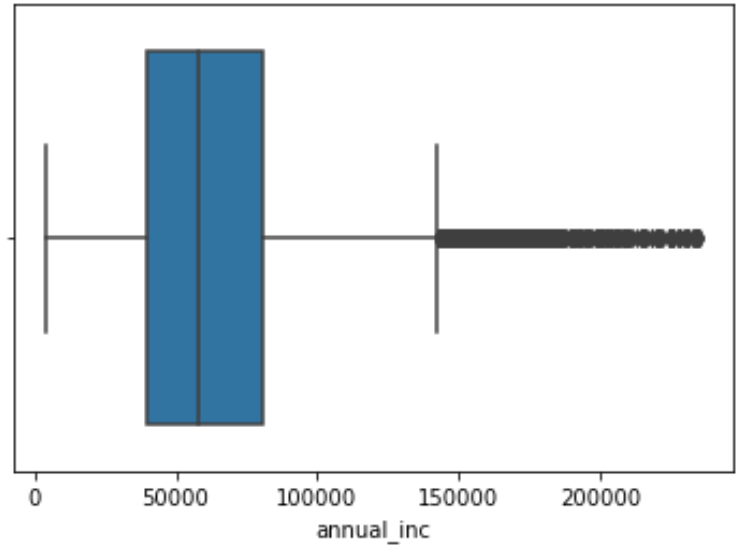
STEPS

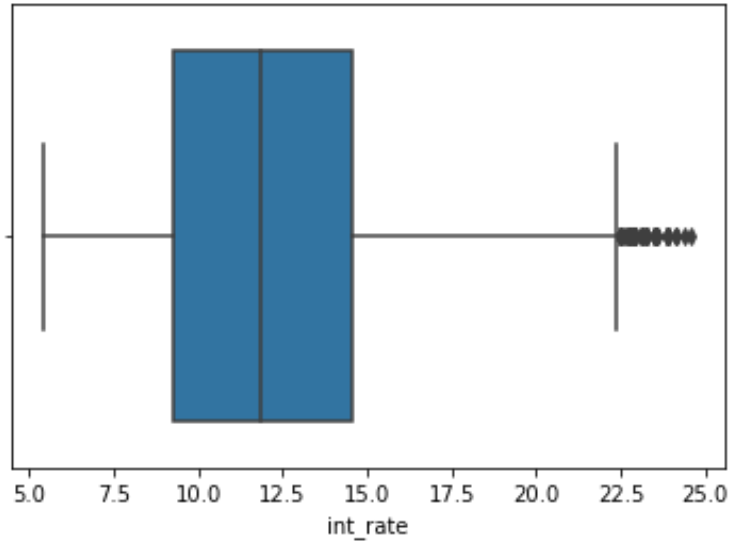
1. Data understanding and Sourcing.
2. Check for Data quality issue and Binning
3. Check for Data imbalance and univariate, segmented univariate & Bivariate analysis, correlation
4. Merging of application data with previous application data.
5. Data analysis by univariate, segmented univariate, Bivariate analysis and correlation
6. Recommendation and Risks

Univariate Analysis



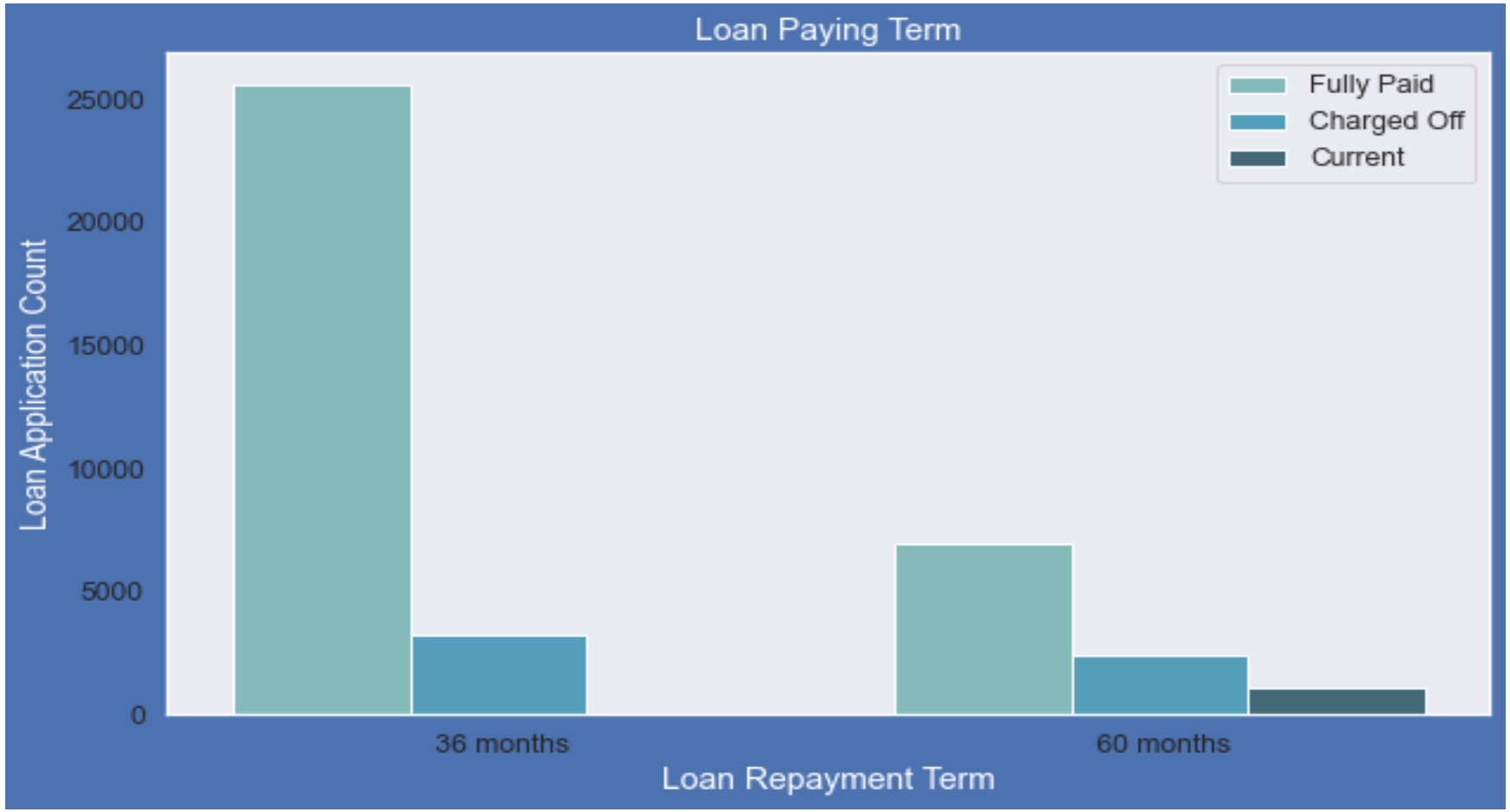






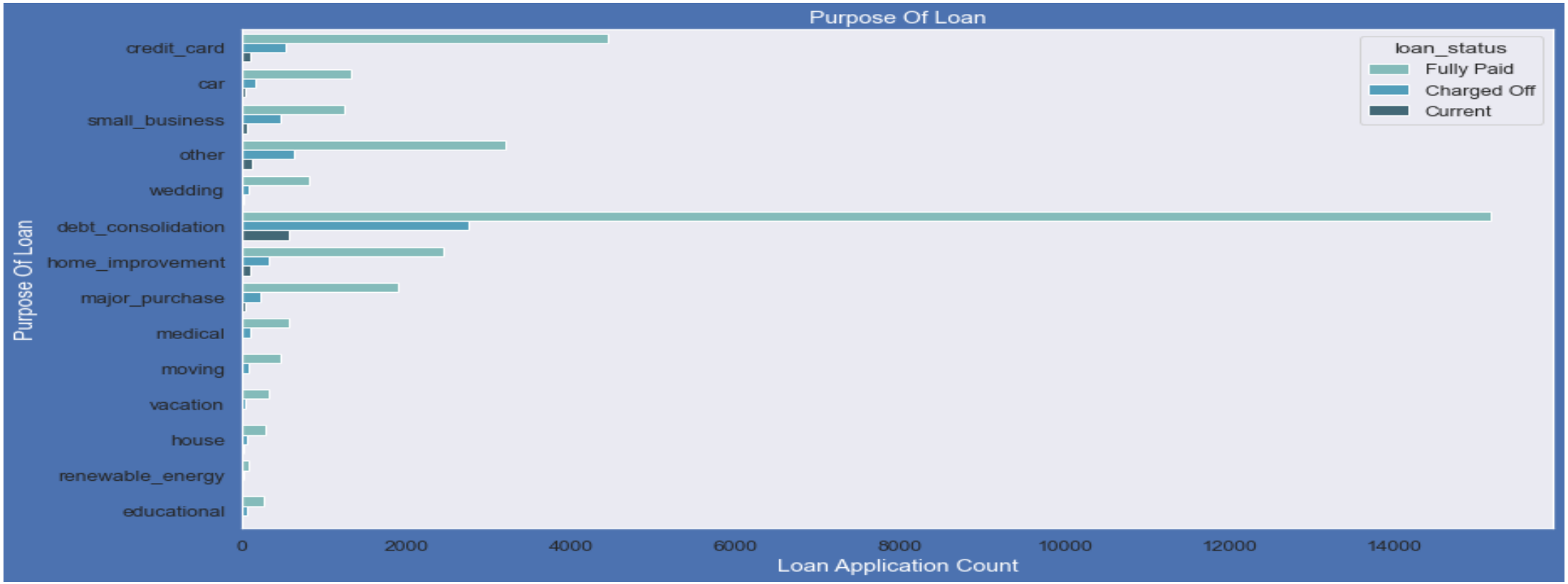
Loan Paying Term

- Below plot shows that those who had taken loan to repay in 60 months had more % of number of applicants getting
- charged off as compared to applicants who had taken loan for 36 months.

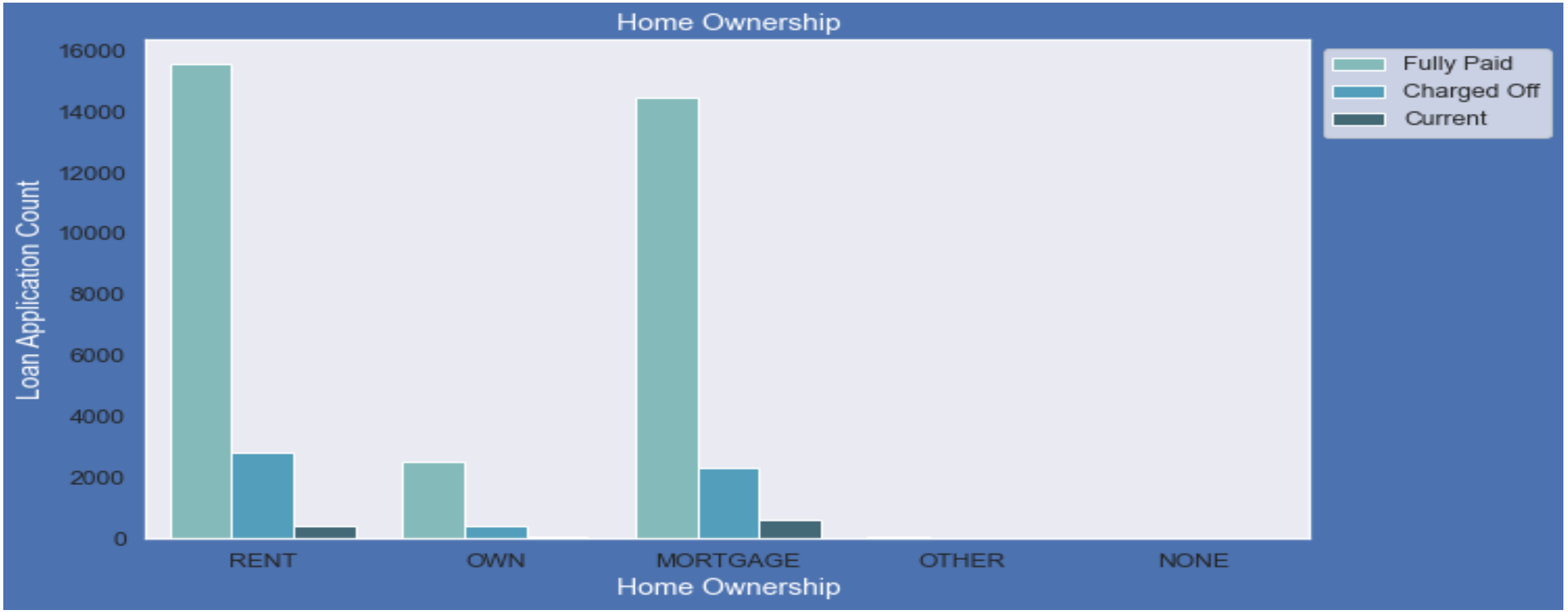


Bivariate Analysis

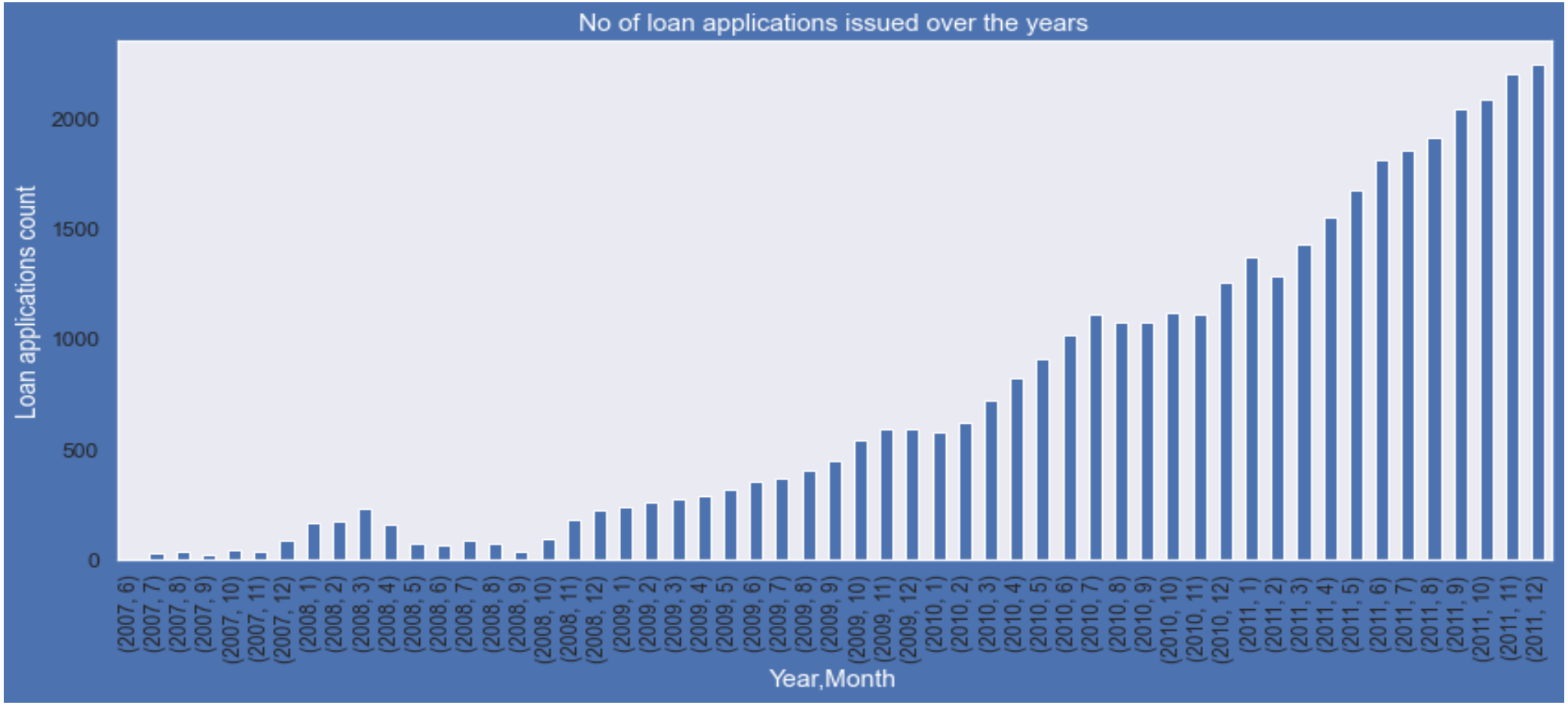
- Below plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.
- Number of chraged off count also high too for these loans.



Below plot shows that most of them living in rented home or mortgaged their home.
Applicant numbers are high from these categories so charged off is high too.

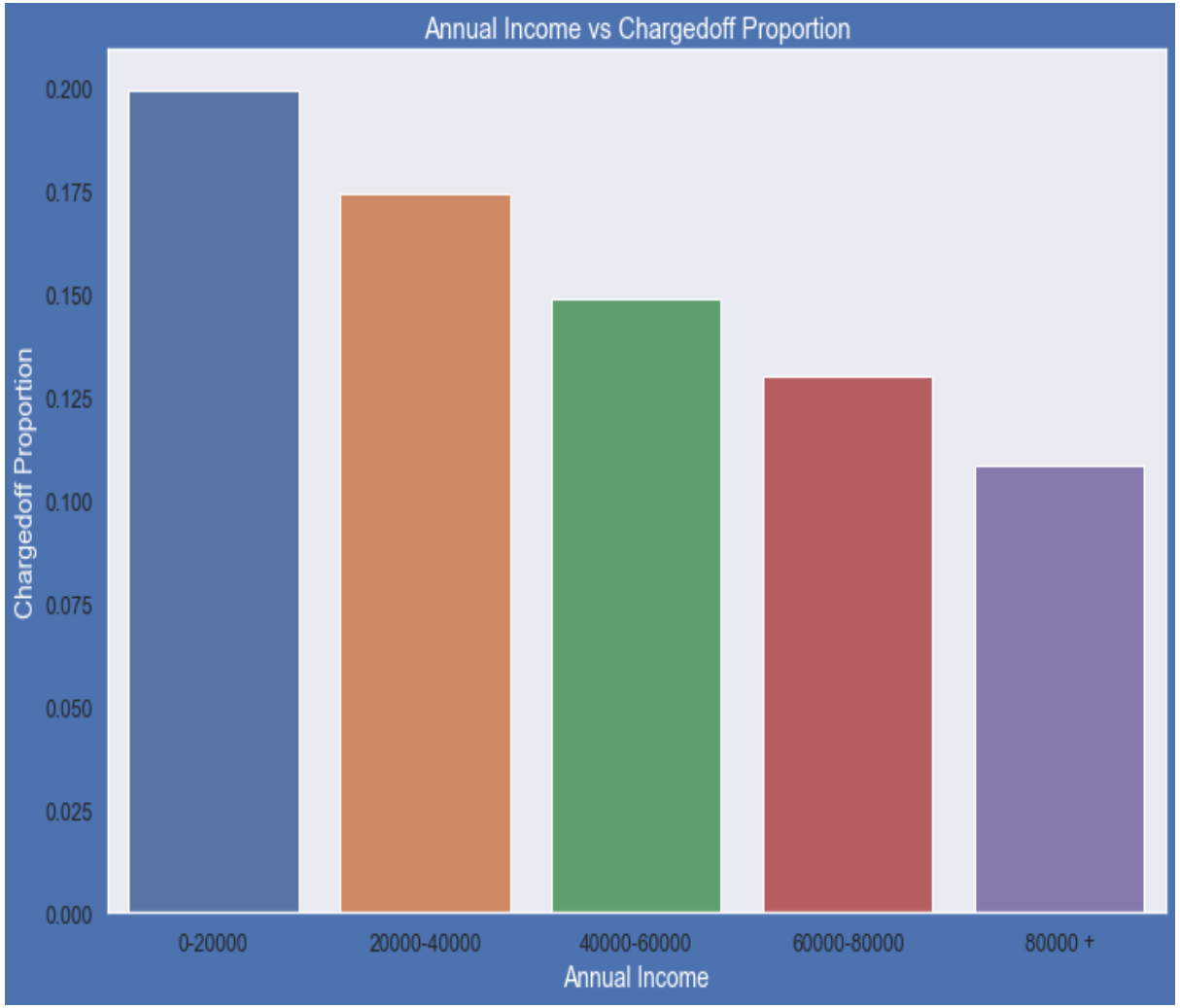


So the increase in number of loan applications are adding more to number of charged off applications.
Number of loans issued in 2008(May-October) got dipped, may be due to Recession.

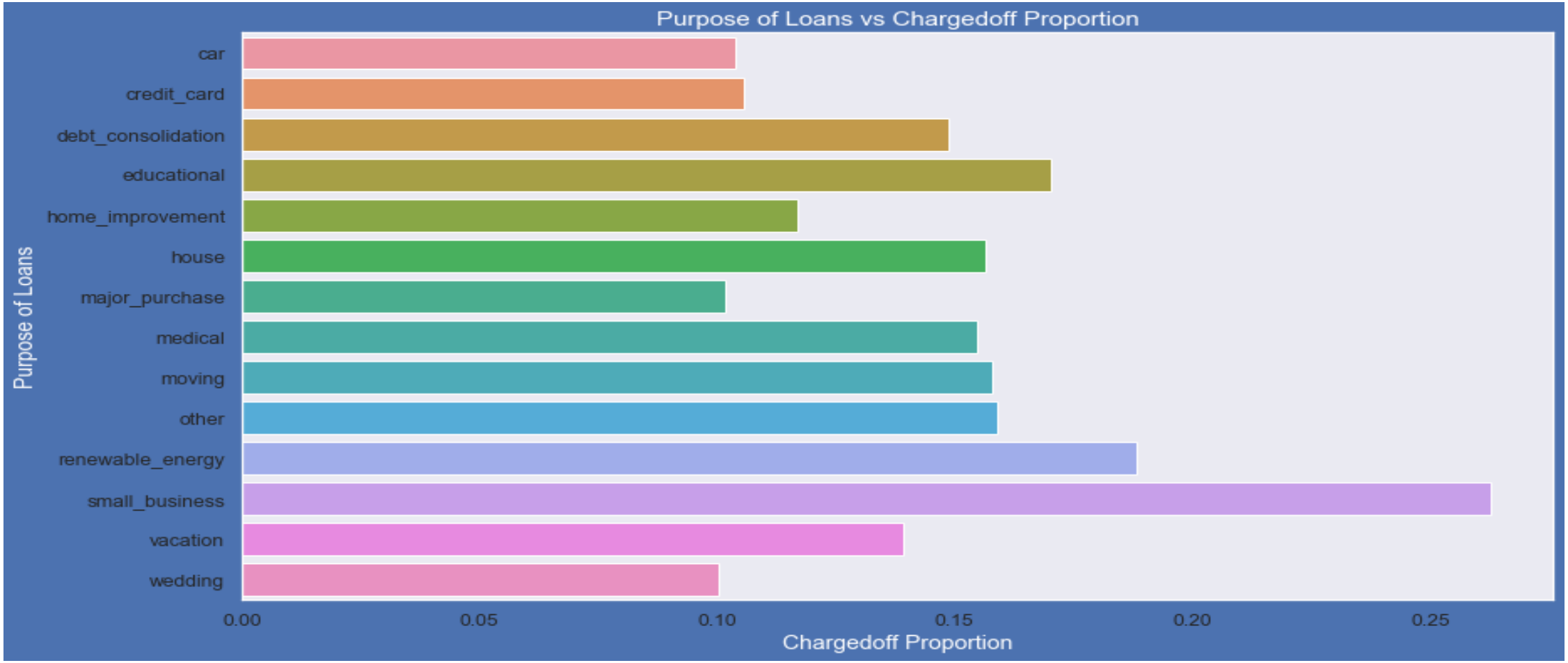


Below are the observations from this plot:
1.Income range 80000+ has less chances of charged off.
2.Income range 0-20000 has high chances of charged off.
3.Notice that with increase in annual income charged-off proportion got decreased.

loan_status	annual_inc_cats	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
0	0-20000	237	9	943	1189	0.20
1	20000-40000	1514	170	7004	8688	0.17
2	40000-60000	1729	345	9534	11608	0.15
3	60000-80000	1024	240	6597	7861	0.13
4	80000 +	1080	362	8531	9973	0.11



Observations from this plot:
1.small Business applicants have high chances of getting charged off.
2.renewable_energy where charged-off proportion is better as compare to other categories.



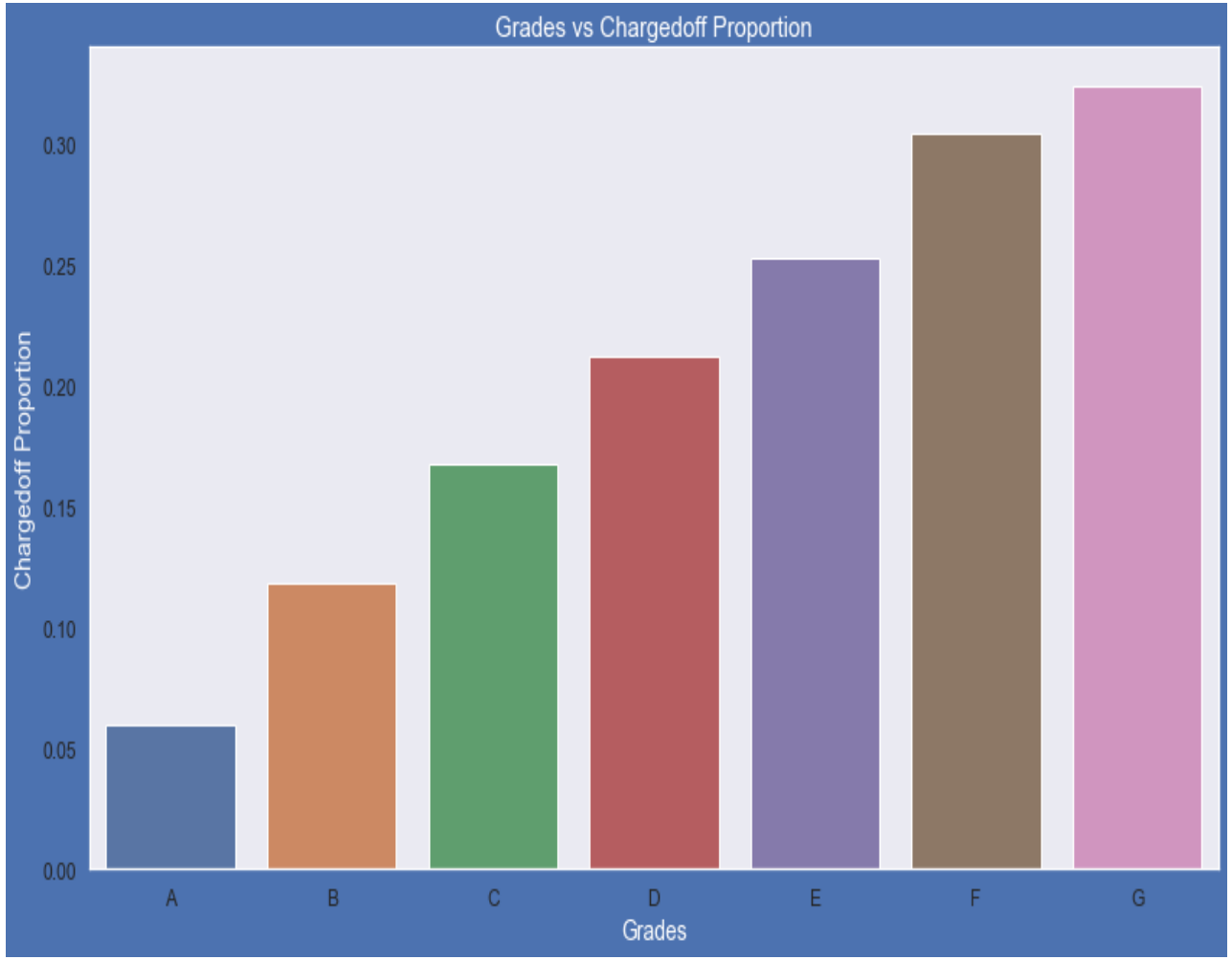
Key Observations:

1.Grade "A" has very less chances of charged off.

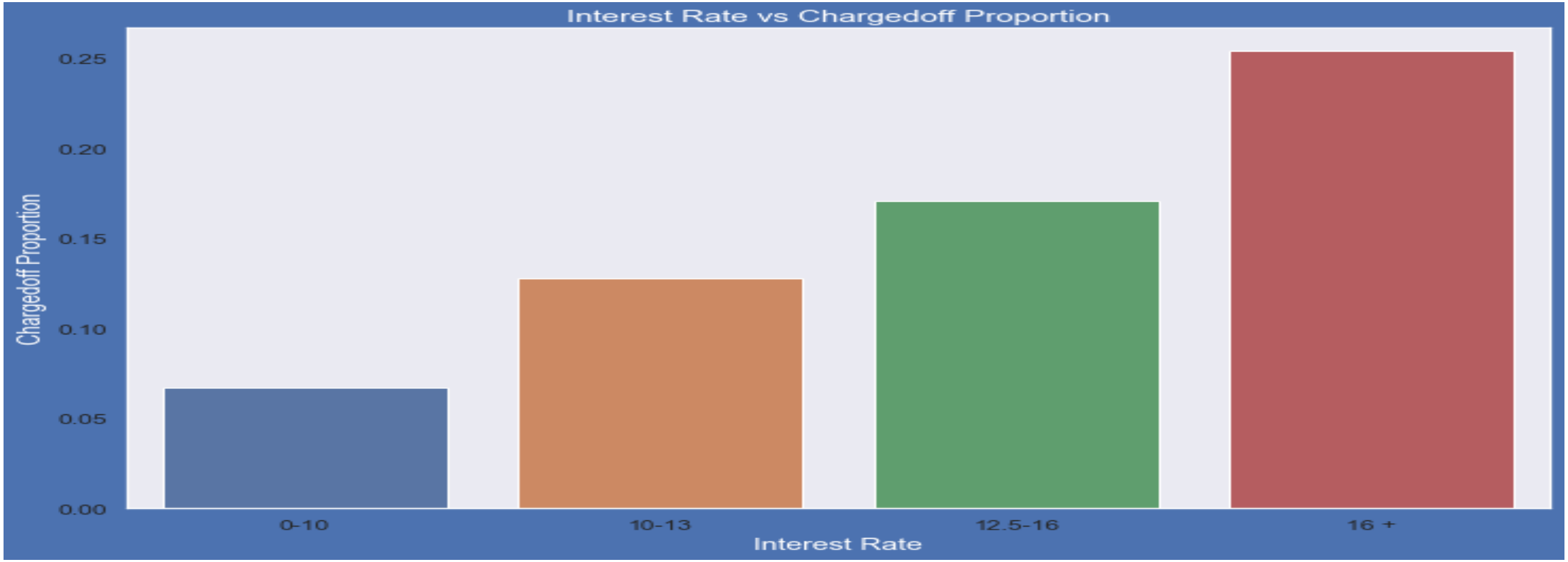
2.Grade "F" and "G" have very high chances of charged off.

3.Chances of charged of is increasing with grade moving from "A" towards "G"

loan_status	grade	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
6	G	99	17	190	306	0.32
5	F	312	71	642	1025	0.30
4	E	704	176	1906	2786	0.25
3	D	1112	216	3916	5244	0.21
2	C	1342	262	6411	8015	0.17
1	B	1414	344	10167	11925	0.12
0	A	601	40	9377	10018	0.06



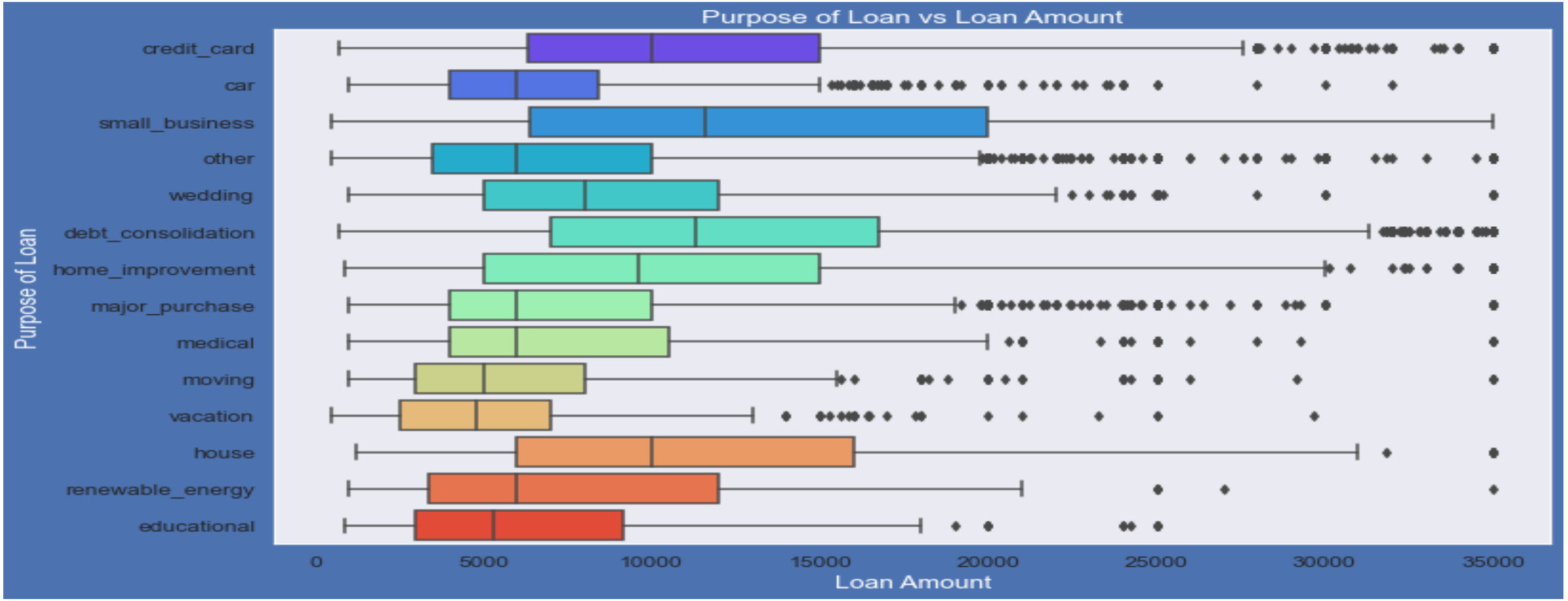
Key points:
1. interest rate less than 10% has very less chances of charged off. Interest rates are starting from minimum 5 %.
2. interest rate more than 16% has good chances of charged off as compared to other category interest rates.
3. Charged-off proportion is increasing with higher interest rates.



Key Points:

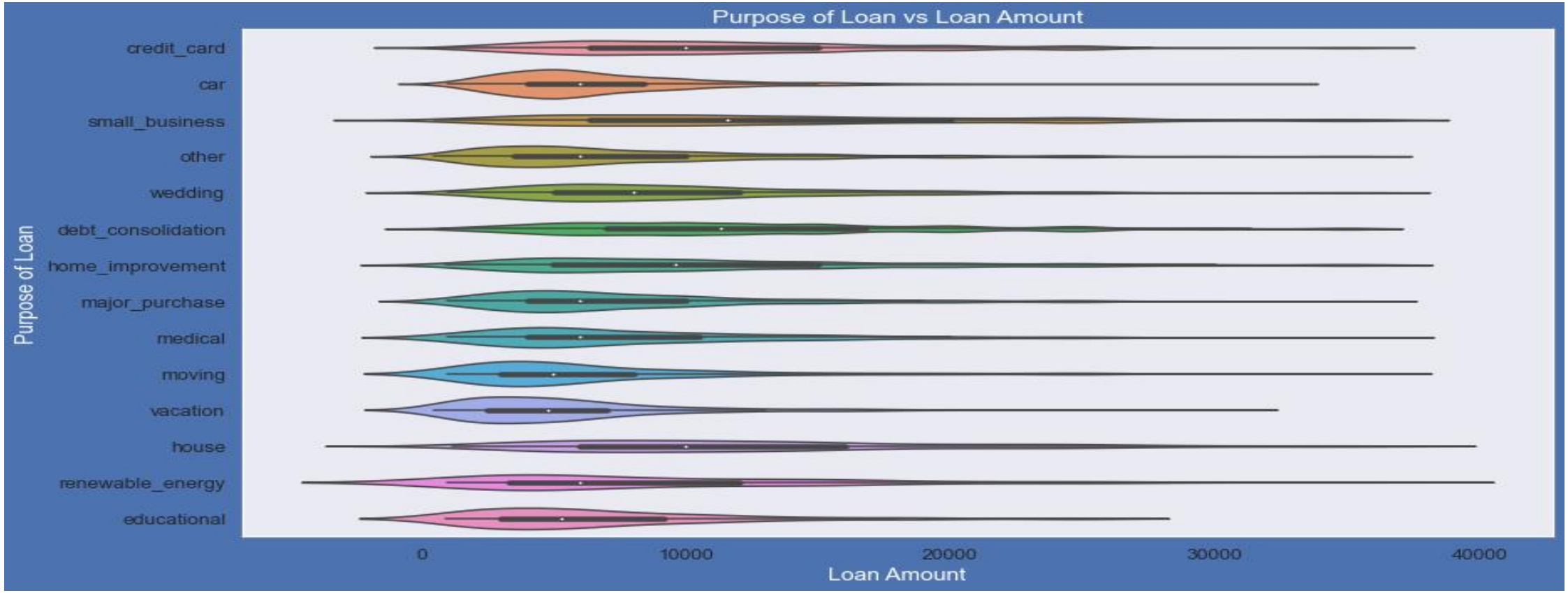
1.Median,95th percentile,75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.

2.Debt consolidation is second and Credit card comes 3rd.



Violin Plot - It shows the distribution of quantitative data across several levels of one (or more) categorical variables such that those distributions can be compared. Here is the observation :-

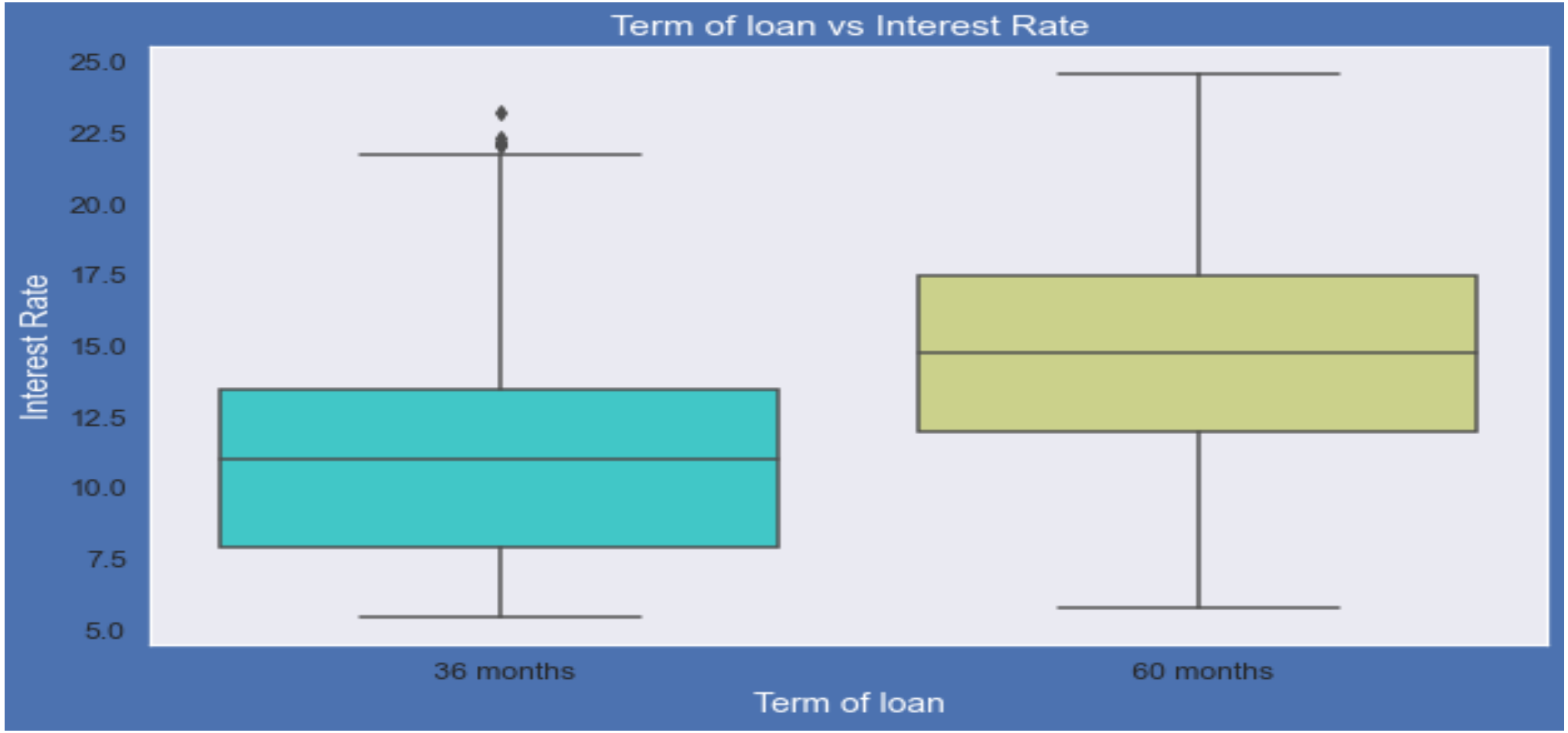
Loan taken for small business purpose, Debt consolidation and Credit card are somewhat evenly distributed as compare to loan taken for other purposes.



Term of loan vs Interest Rate observations:

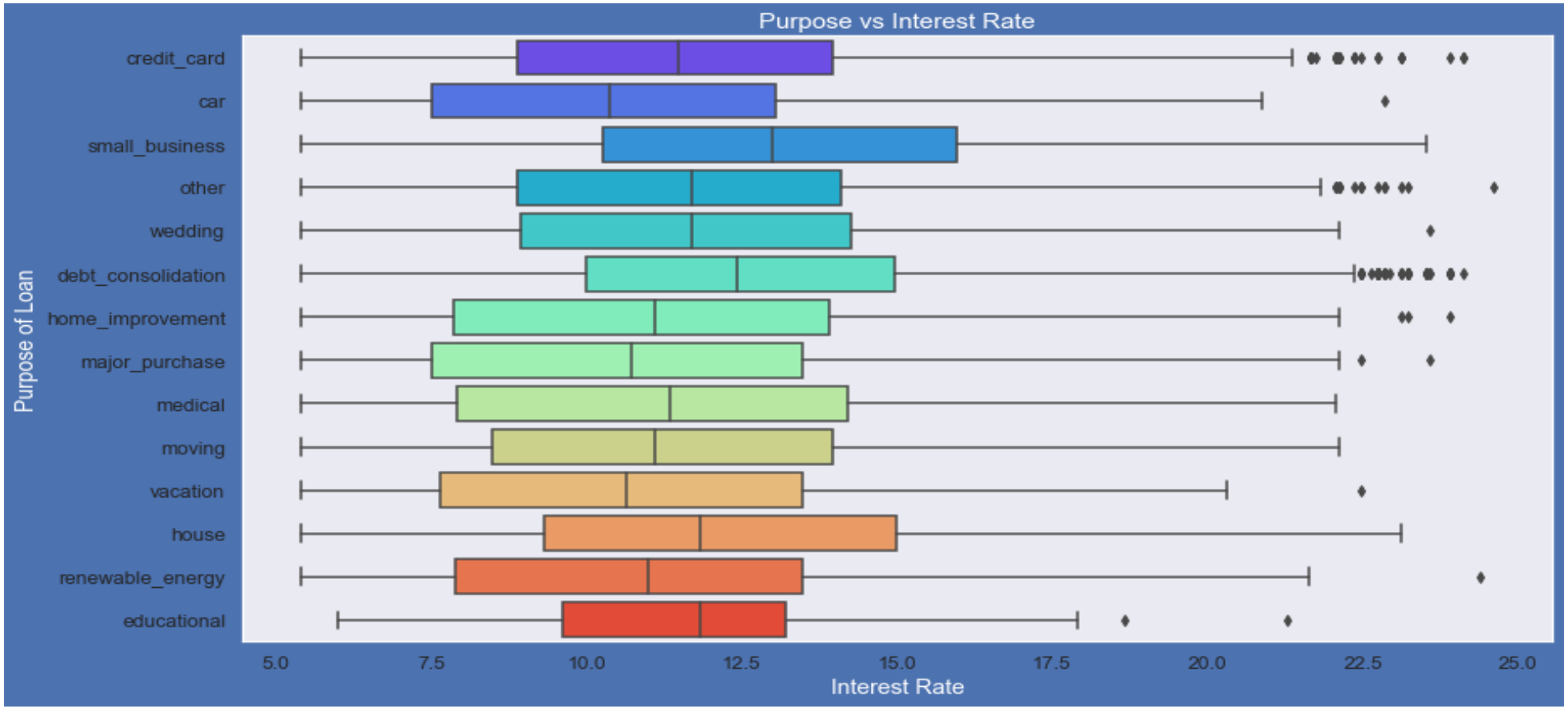
1.It is clear that average interest rate is higher for 60 months loan term.

2.Most of the loans issued for longer term had higher interest rates for repayment.



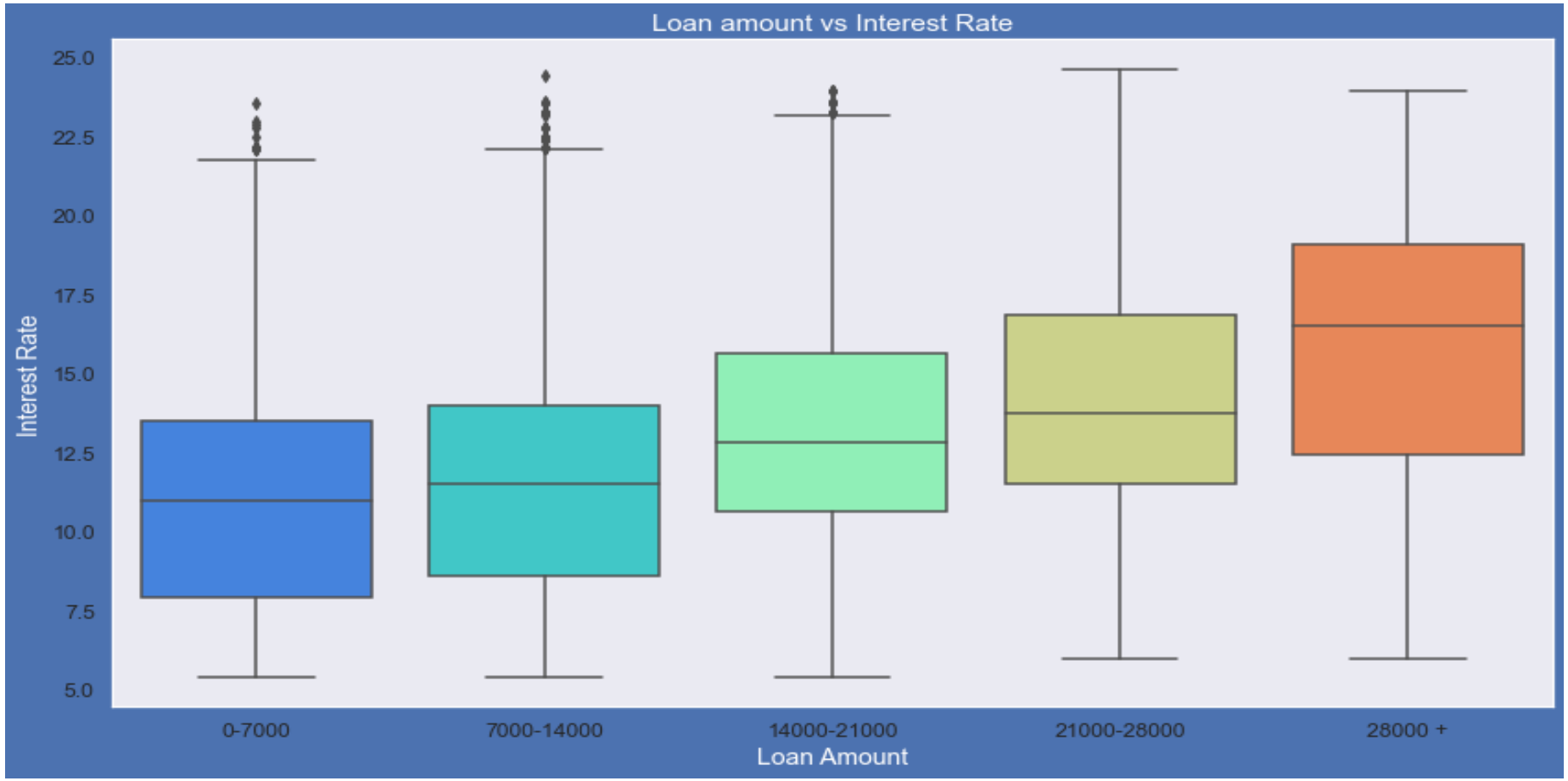
Purpose vs Interest Rate observations:

1. It is clear that average interest rate is highest for small business purpose.
2.Loans taken for small business purposes had to repay the loan with more intrest rate as compared to other.
3.Debt consolidation is 2nd where borrowers had to pay more interest rate.



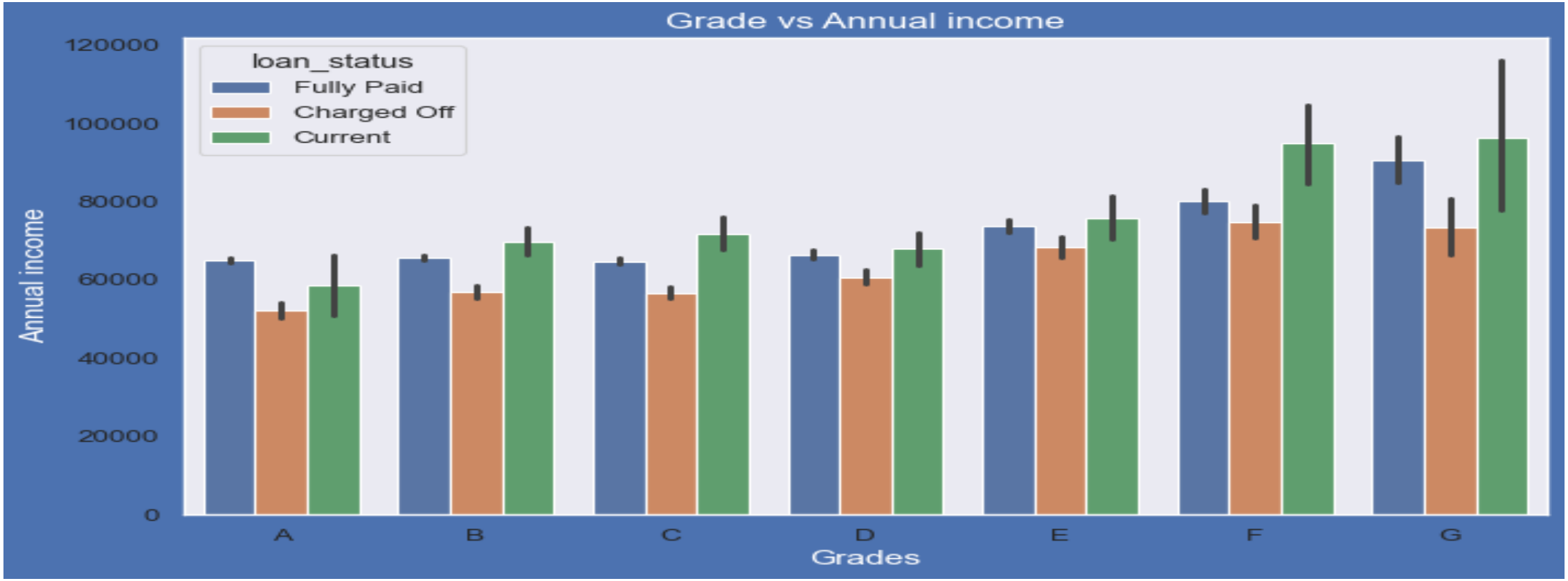
Loan Amount vs Interest Rate observations:

1. It is clear that interest rate is increasing with loan amount increase.
2. probably when loan amount is more it is taken for longer loan term, we saw earlier that longer the loan term more the interest rate.



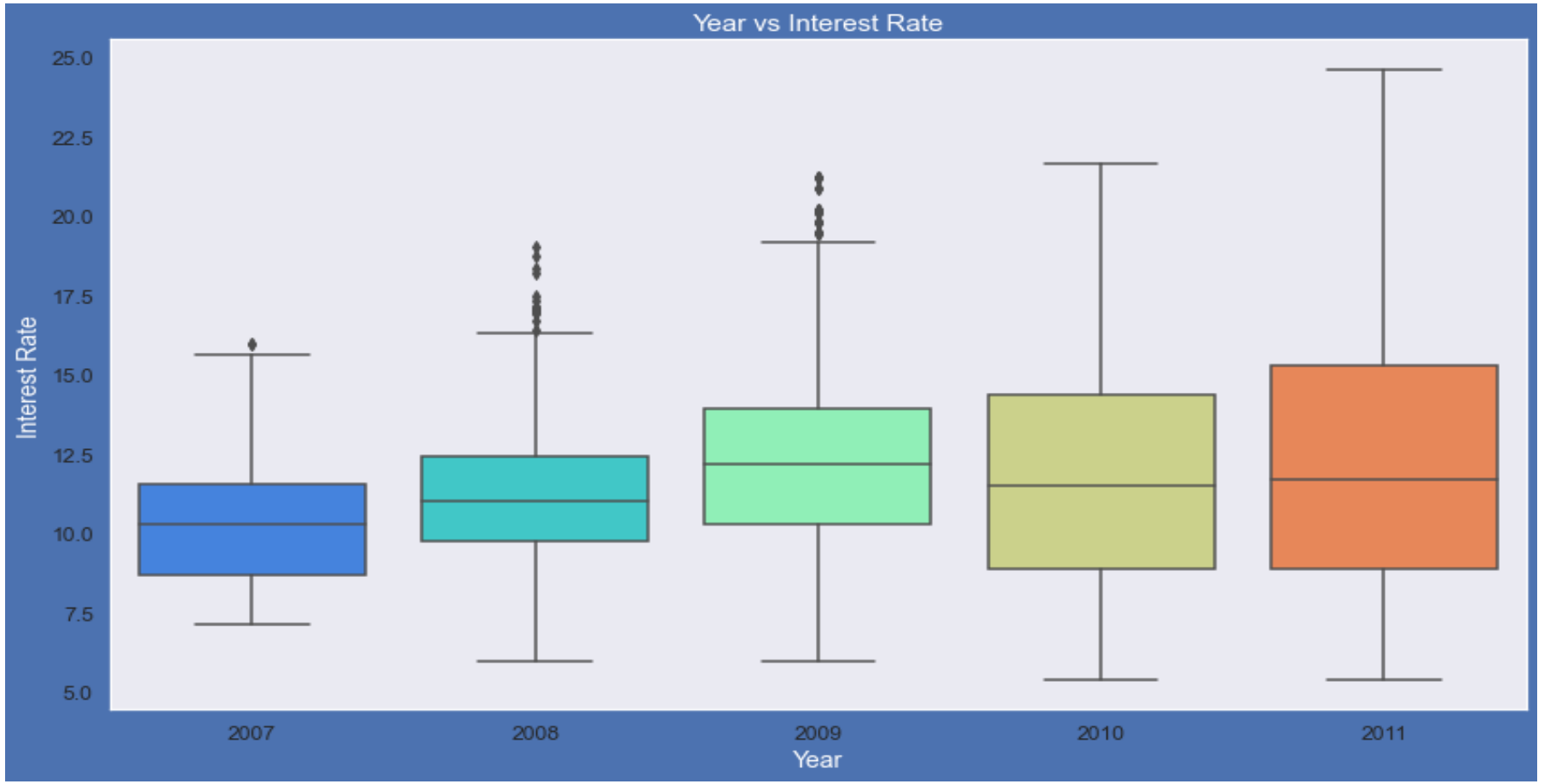
BarPlot to show variation of annual income across grade for every loan status

From this we can conclude that the ones getting 'charged off' have lower annual incomes than the ones who 'paid fully' for each and every grade (i.e. at same interest range)

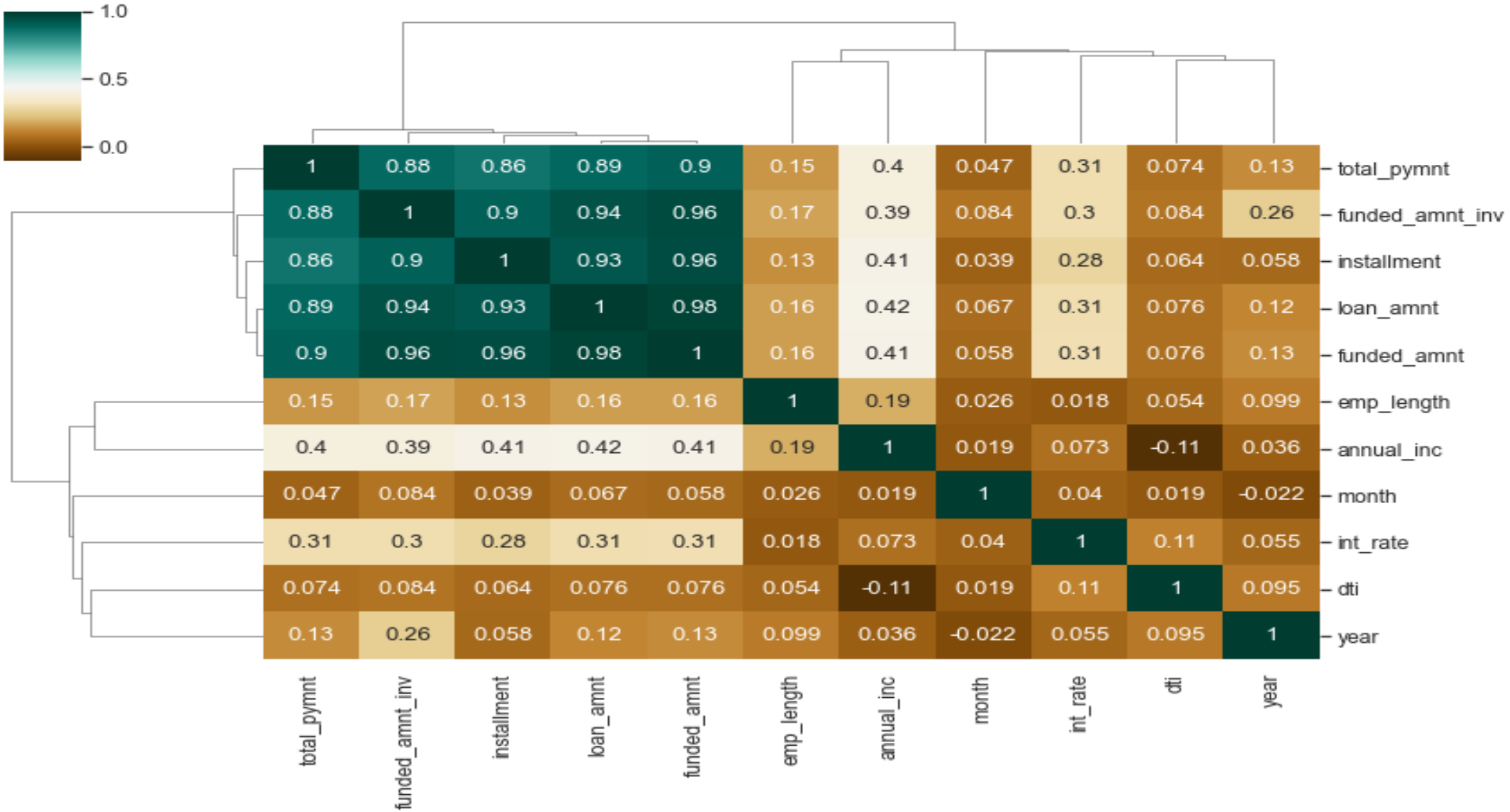


Box plot for year vs interest rate

Plot shows interest rate is increasing slowly with increase in year.



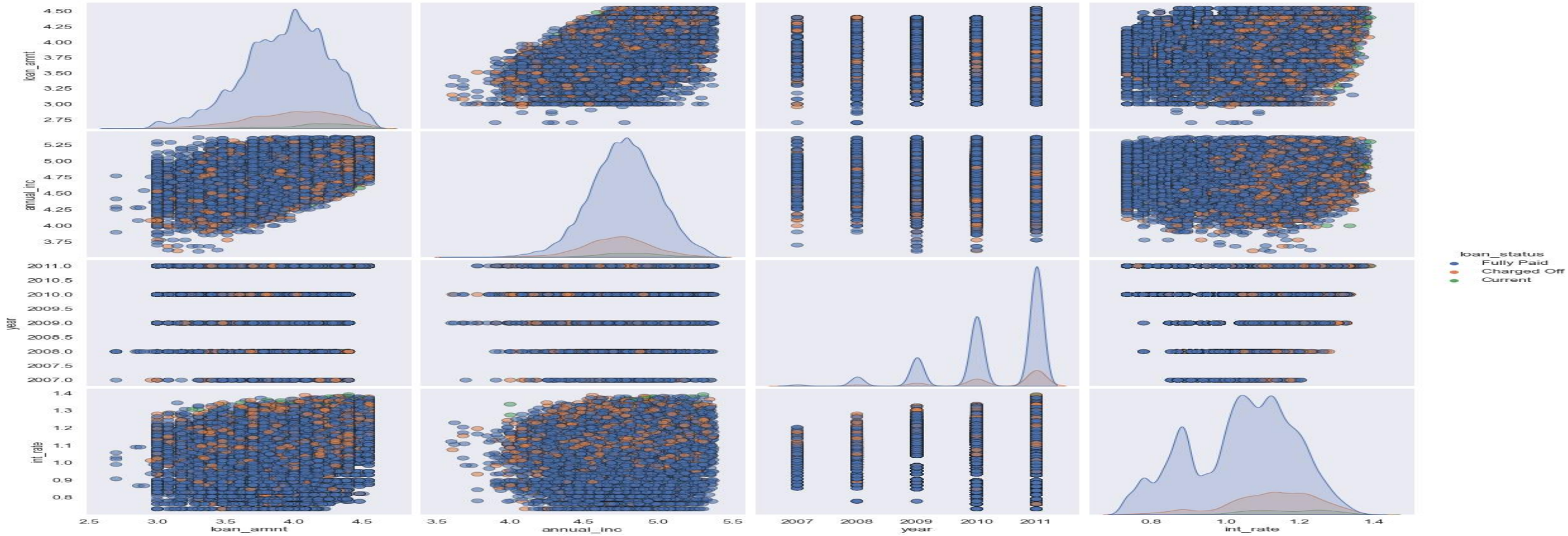
Correlation Matrix - Quantitative Variables :



	total_pymnt	funded_amnt_inv	installment	loan_amnt	funded_amnt	emp_length	annual_inc	month	int_rate	dti	year
total_pymnt	1	0.88	0.86	0.89	0.9	0.15	0.4	0.047	0.31	0.074	0.13
funded_amnt_inv	0.88	1	0.9	0.94	0.96	0.17	0.39	0.084	0.3	0.084	0.26
installment	0.86	0.9	1	0.93	0.96	0.13	0.41	0.039	0.28	0.064	0.058
loan_amnt	0.89	0.94	0.93	1	0.98	0.16	0.42	0.067	0.31	0.076	0.12
funded_amnt	0.9	0.96	0.96	0.98	1	0.16	0.41	0.058	0.31	0.076	0.13
emp_length	0.15	0.17	0.13	0.16	0.16	1	0.19	0.026	0.018	0.054	0.099
annual_inc	0.4	0.39	0.41	0.42	0.41	0.19	1	0.019	0.073	-0.11	0.036
month	0.047	0.084	0.039	0.067	0.058	0.026	0.019	1	0.04	0.019	-0.022
int_rate	0.31	0.3	0.28	0.31	0.31	0.018	0.073	0.04	1	0.11	0.055
dti	0.074	0.084	0.064	0.076	0.076	0.054	-0.11	0.019	0.11	1	0.095
year	0.13	0.26	0.058	0.12	0.13	0.099	0.036	-0.022	0.055	0.095	1

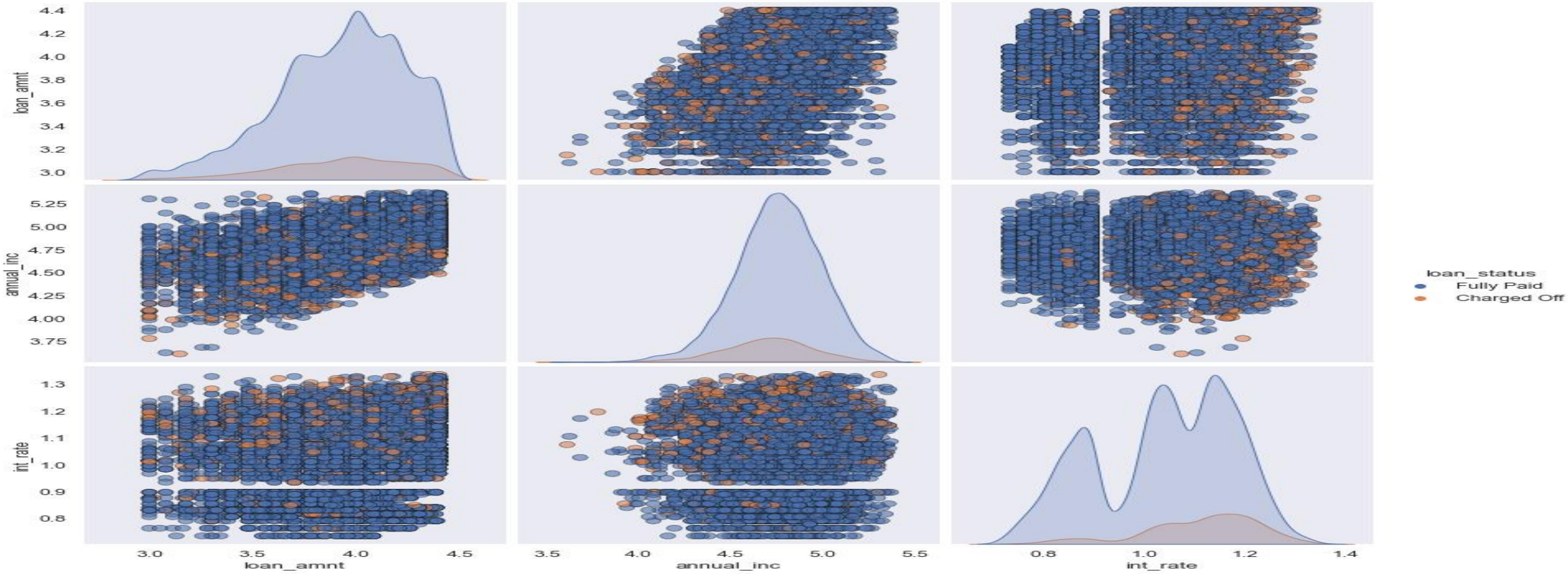
Multivariate Analysis - Pair Plots - 1

- Higher interest rate higher charged off ration
- Higher the annual income higher the loan amount slightly
- Increase in number of charged off with increase in year
- Interest rate is increase with loan amount increase



Plot-2

- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly
- Interest rate is increase with loan amount increase this result in high charged off



Total Loan Issued

debt_consolidation	46.93
credit_card	12.92
other	10.05
home_improvement	7.49
major_purchase	5.51
small_business	4.6
car	3.9
wedding	2.38
medical	1.74
moving	1.47
house	0.96
vacation	0.96
educational	0.82
renewable_energy	0.26

Conclusions

- Observation is that loan amount, investor amount, funding amount are strongly correlated.
- Annual income with DTI(Debt to income ratio) is negatively correlated.
- Debt income ration is the percentage of a consumer monthly gross income that goes toward paying debts.
- That means when annual income is low DTI is high & Vice versa
- Positive Correlation between annual income and employment years
- That means income increase with work experience.