

Lending Club Case Study

Group Members:

Shruti Pandit

Allena Venkata Sai Abhishek

Index:

- Identifying Business Problem
- Data Description
- Data Preparation and Processing
- Data Cleaning and Manipulation
- Models:
 - Univariate Analysis
 - Bivariate Analysis
- Summary of Findings
- Conclusion



Identifying Business Problem

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 find whether they 'defaulted' or not.



Data Description

Lending Club provides us with 4 years of historical data. This dataset contained information pertaining to the borrower's past credit history and Lending Club Loan information. The total dataset consisted of over 39K+ records which was sufficient for our team to conduct the analysis models.

Variables present within the dataset provided an ample amount of information which we could use to identify relationships & gauge their effect upon the success or failure of a borrower fulfilling the terms of their loan agreement.

We required only variables that had a direct or indirect response to a borrower's potential to default. To achieve this, we prepared the data by choosing select variables that would best fit the criteria.

Data Preparation and Processing

Prior to Data Modelling, the data was reviewed, cleaned and prepared as follows:

- Removed columns that obviously had no relation to the analysis in question
- Removed columns that had bad quality data(i.e. missing values in observations, unintelligible values etc.)
- Established derived columns from existing columns to facilitate model analysis(Eg-Derived Categories from loan amount etc.)
- Filled Null values with 0s for better modeling

Data Modelling

We selected following data models to determine which return the best predictive capability. We utilized multiple assessment parameters to determine the most successful model-

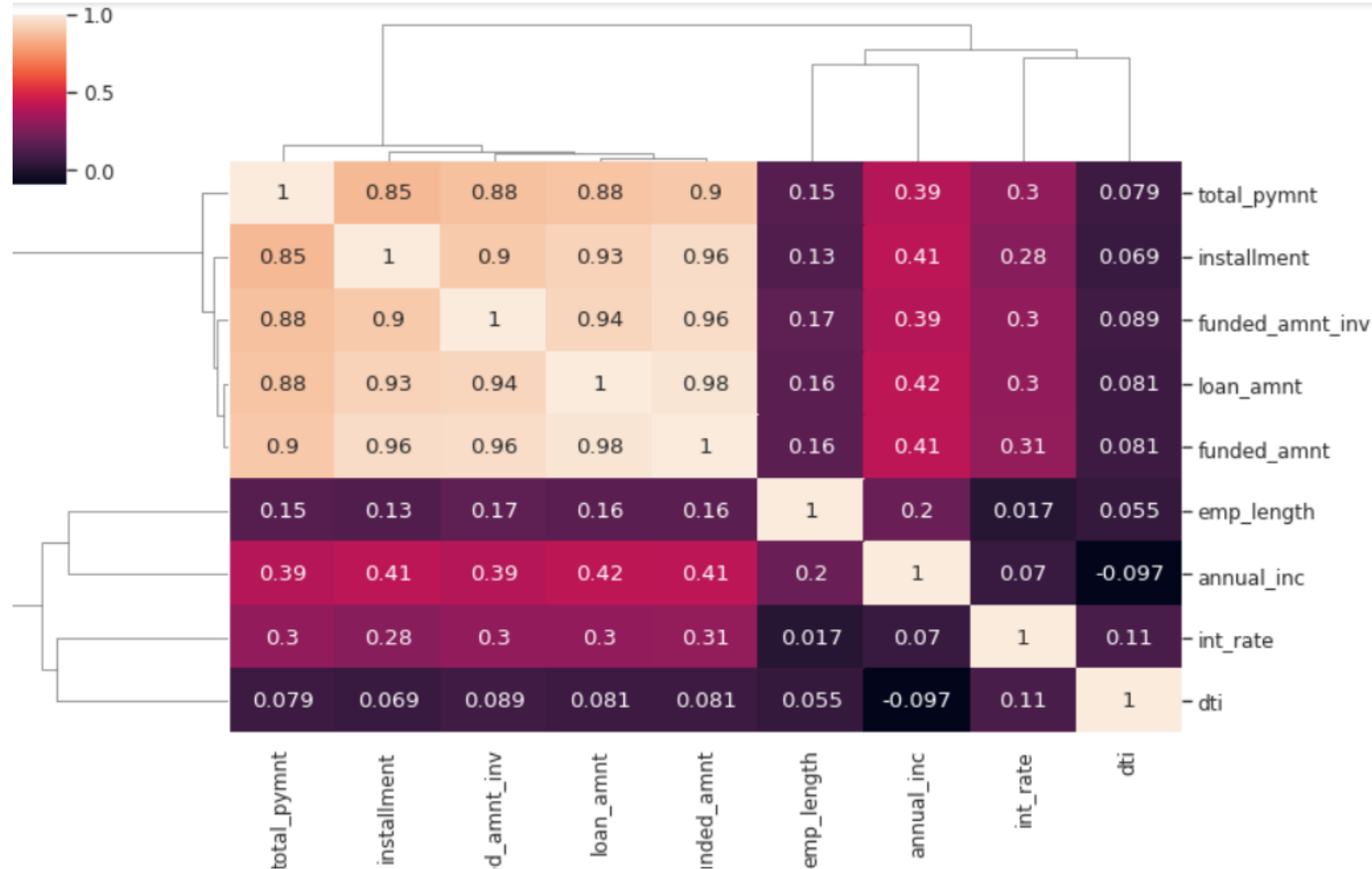
- Univariate analysis-Box Plots
- Bivariate Analysis-Heat Maps
- Univariate Analysis-Distribution Plots
- Univariate & Bivariate Analysis-Bar Plots
- Bivariate analysis-Violin Plots

Finding 1:Heatmap

Findings from Heatmap are as Follows-

1. Annual Income to Debt To Income Ratio i.e. dti are negatively correlated
2. Loan Amount, Investor Amount and Funding Amount are strongly correlated
3. Positive correlation between Annual Income and employment years
4. Positive correlation between annual income and funded amount that means people with high income gets high funded amount
5. Positive correlation between annual income and total payment

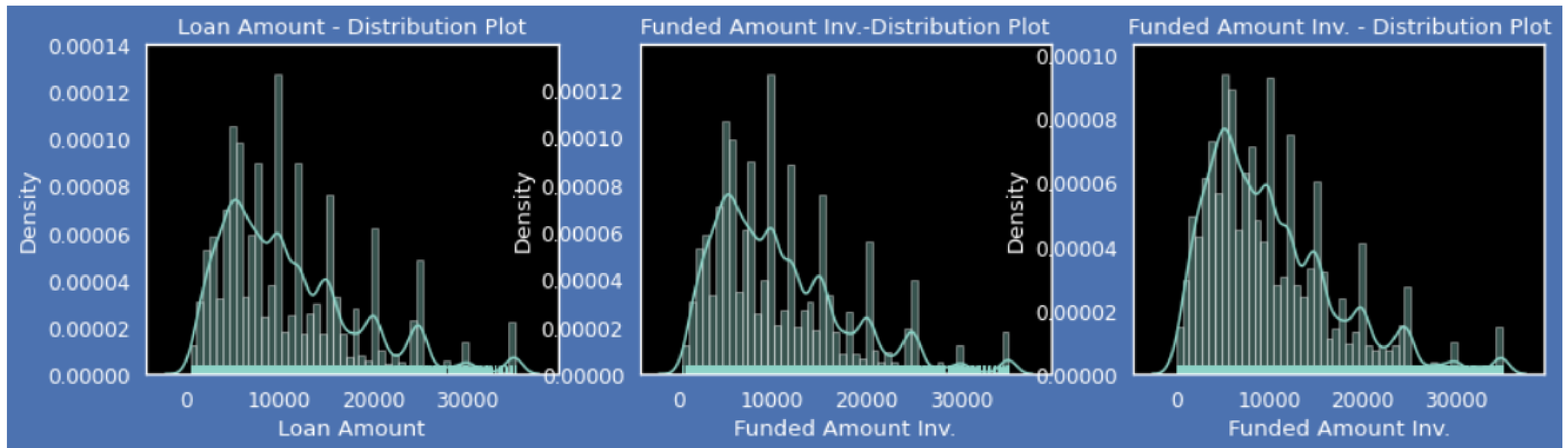
Finding 1: Heatmap



Finding 2:Univariate analysis

Findings are as follows-

1.Distribution of amounts for all three looks very much similar.

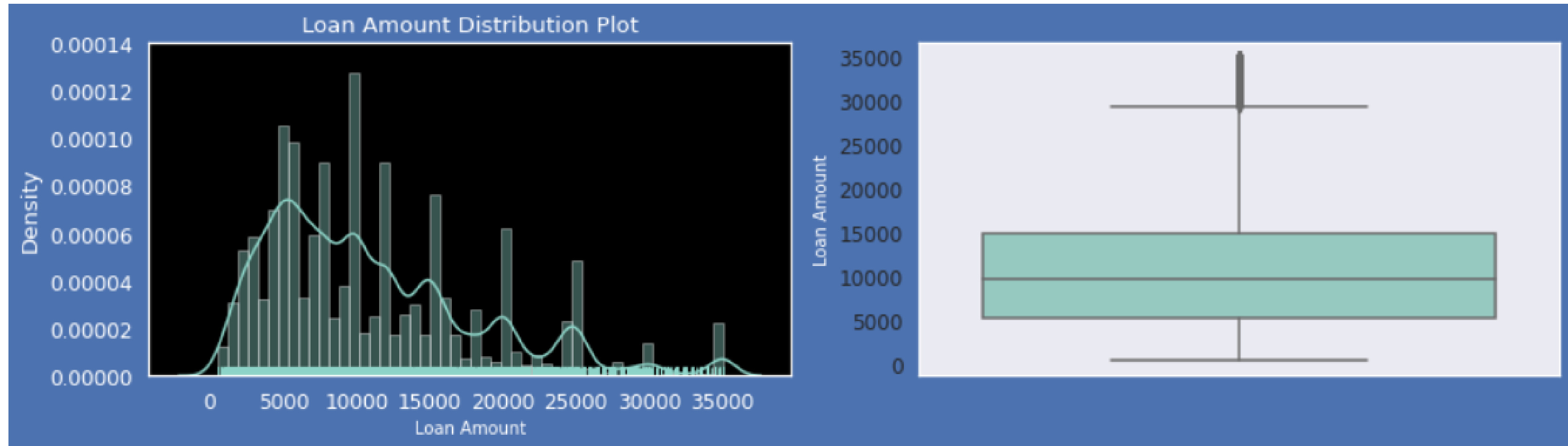


Finding 3: Univariate analysis-Loan Amount Variable



Findings-

1. Most of the loan amounts are in the range 5000-15000



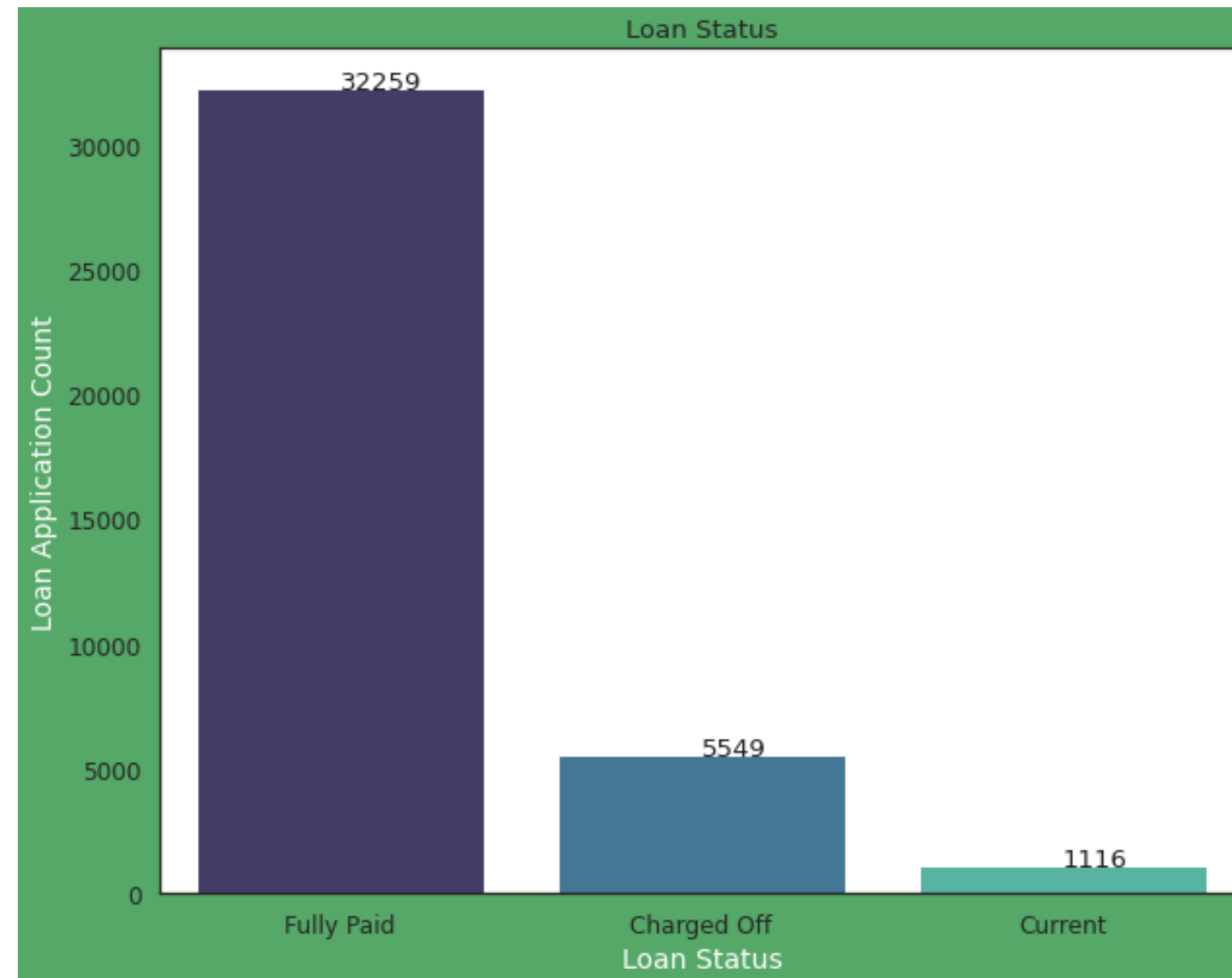


Finding 4:Univariate analysis-Loan Status

UpGrad

Findings-

1.The following plot shows that around 14% loans were charged off out of the total loans issued

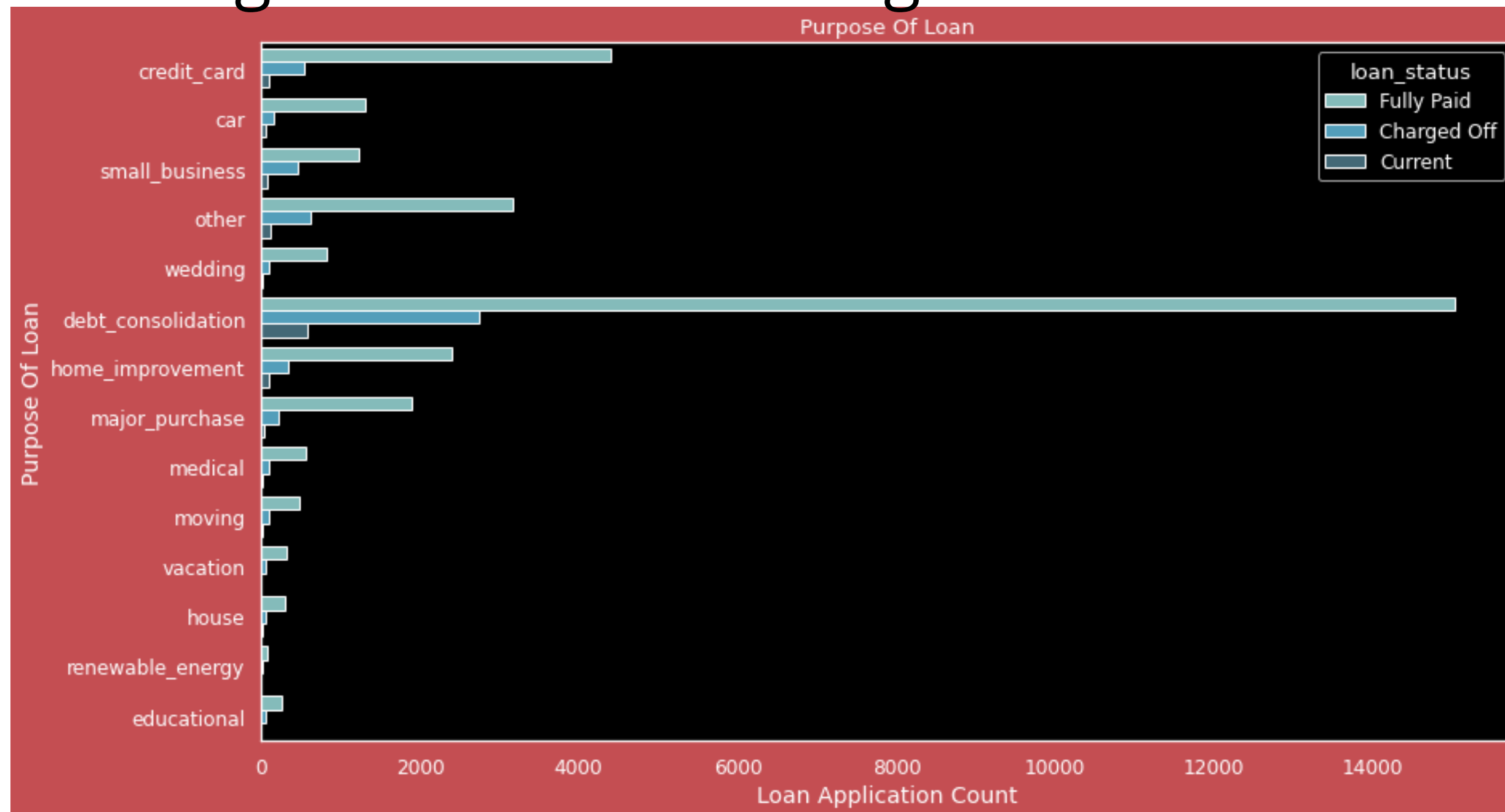




Finding 5:Univariate analysis-Loan Purpose



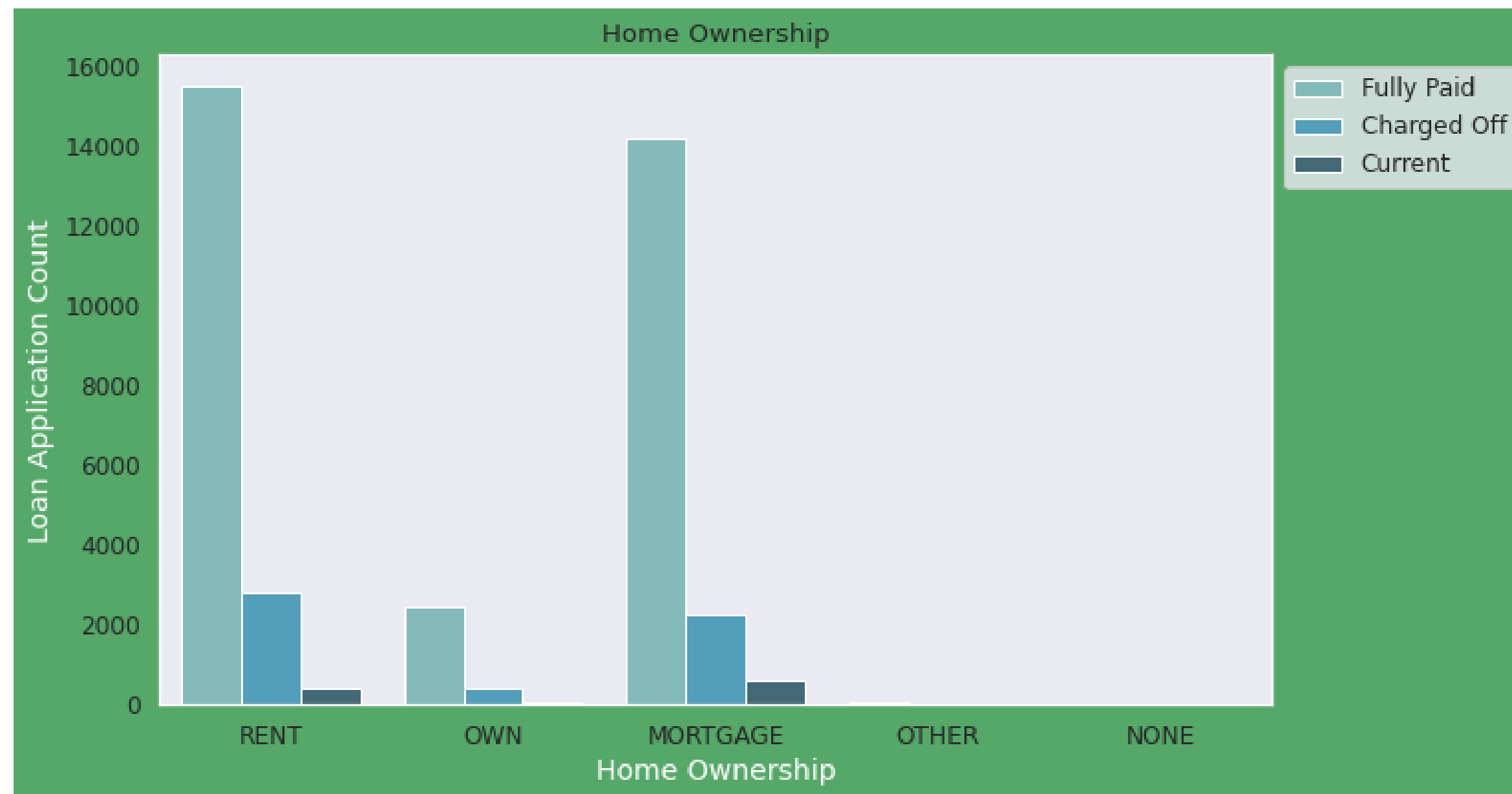
- 1.The Count Plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.
- 2.Number of charged off count also high too for these loans.



Finding 6: Univariate analysis-Home Ownership



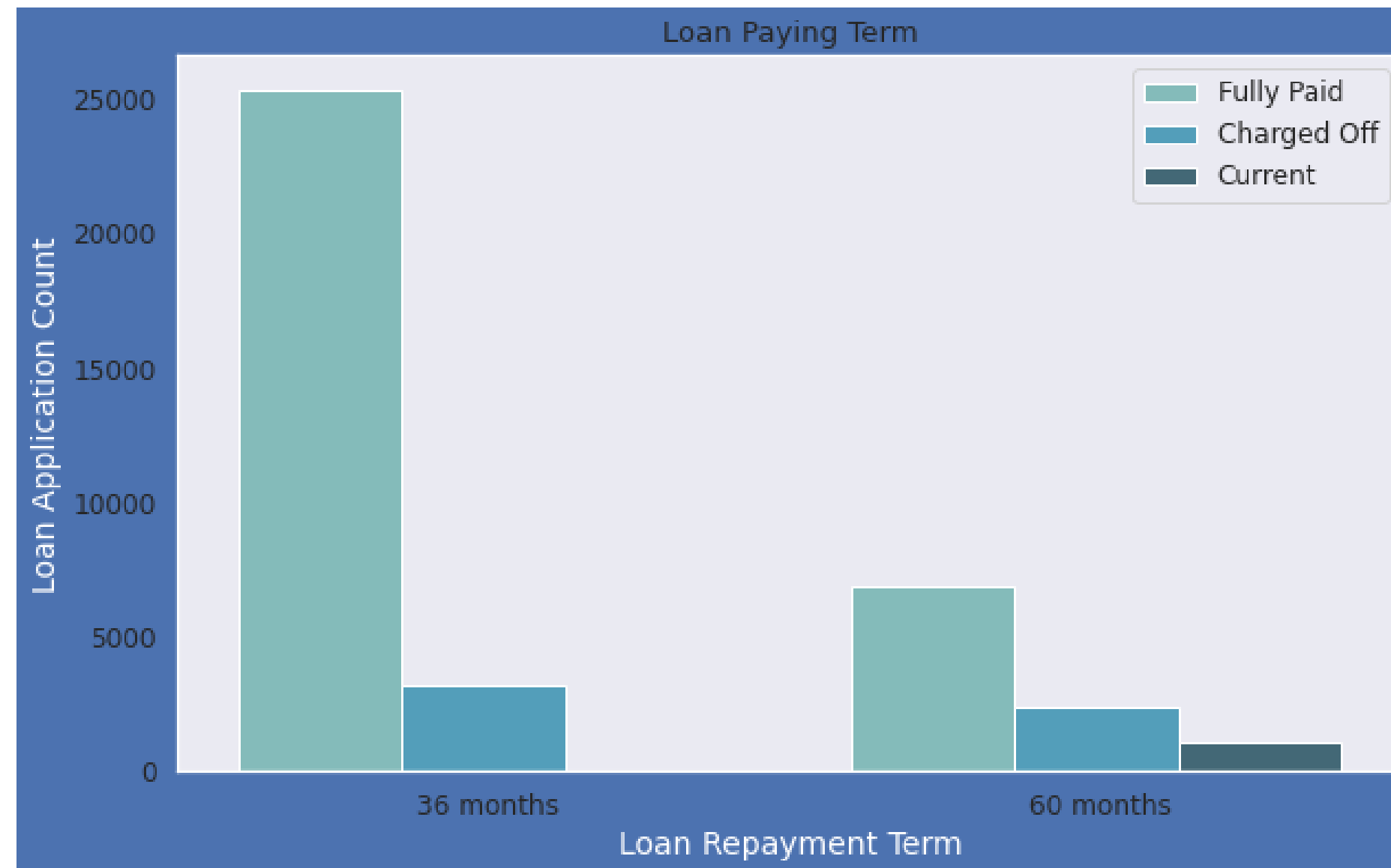
1. The Count plot shows that most of them living in rented home or mortgaged their home.
2. Applicant numbers are high from these categories so charged off is high too.



Finding 7: Univariate analysis-Loan Paying Term



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

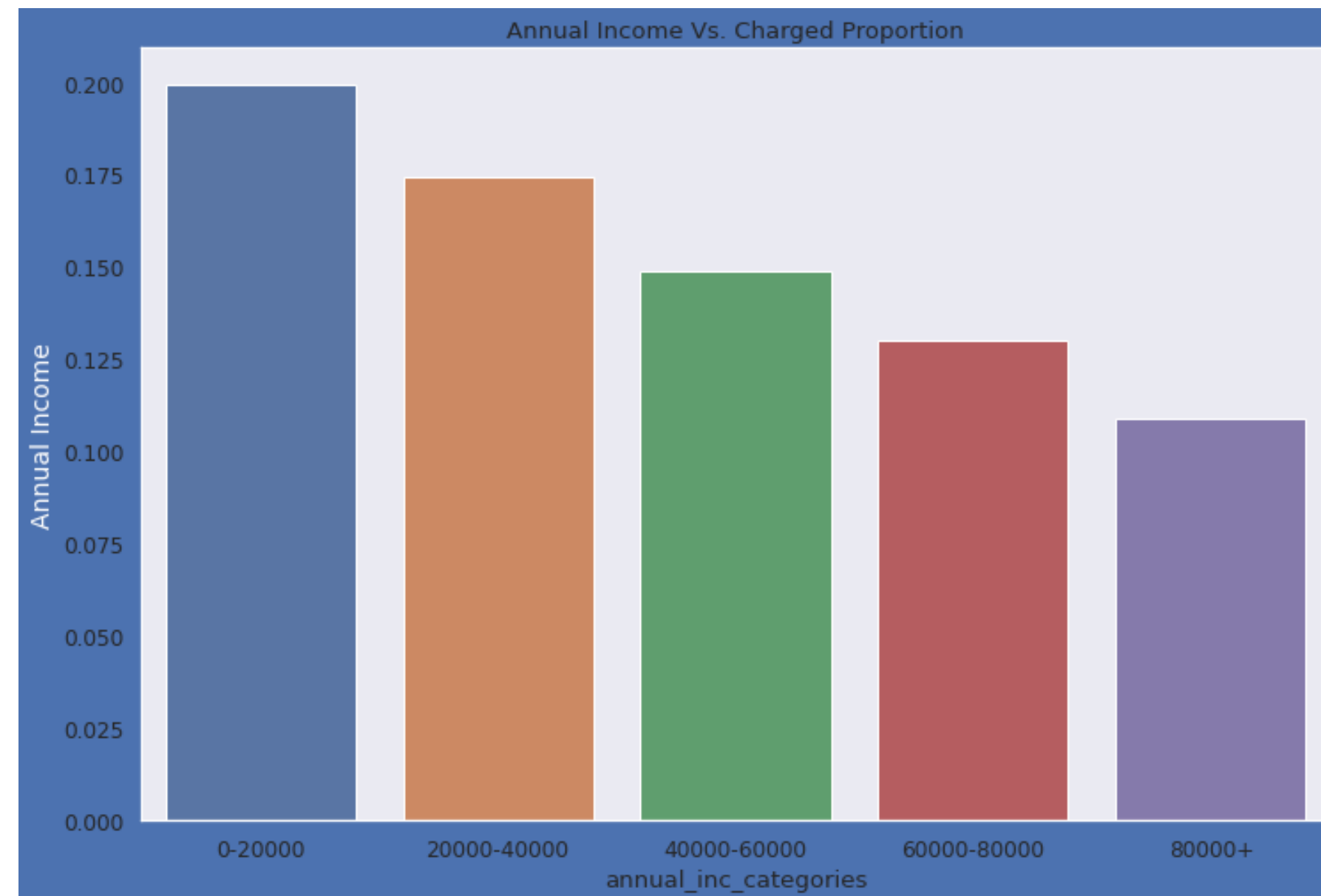




Finding 8: Bivariate analysis-Annual Income Vs. Charged Off Proportion



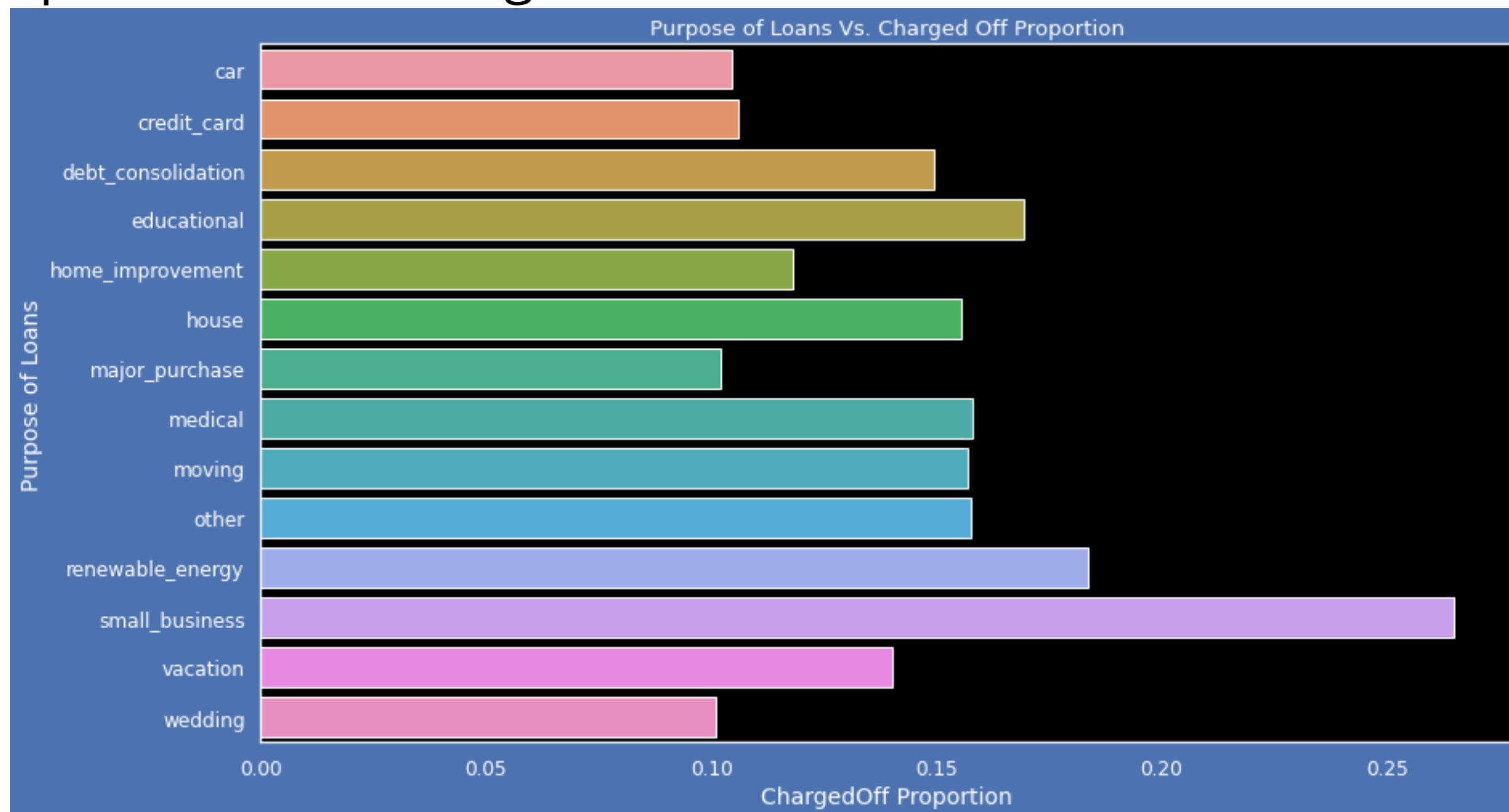
1. Income range 80000+ has less chances of charged off.
2. Income range 0-20000 has high chances of charged off.
3. Notice that with increase in annual income charged off proportion got decreased. So, they are inversely proportional.



Finding 9: Bivariate Analysis Purpose of Loan Vs Charged Off Proportion



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.

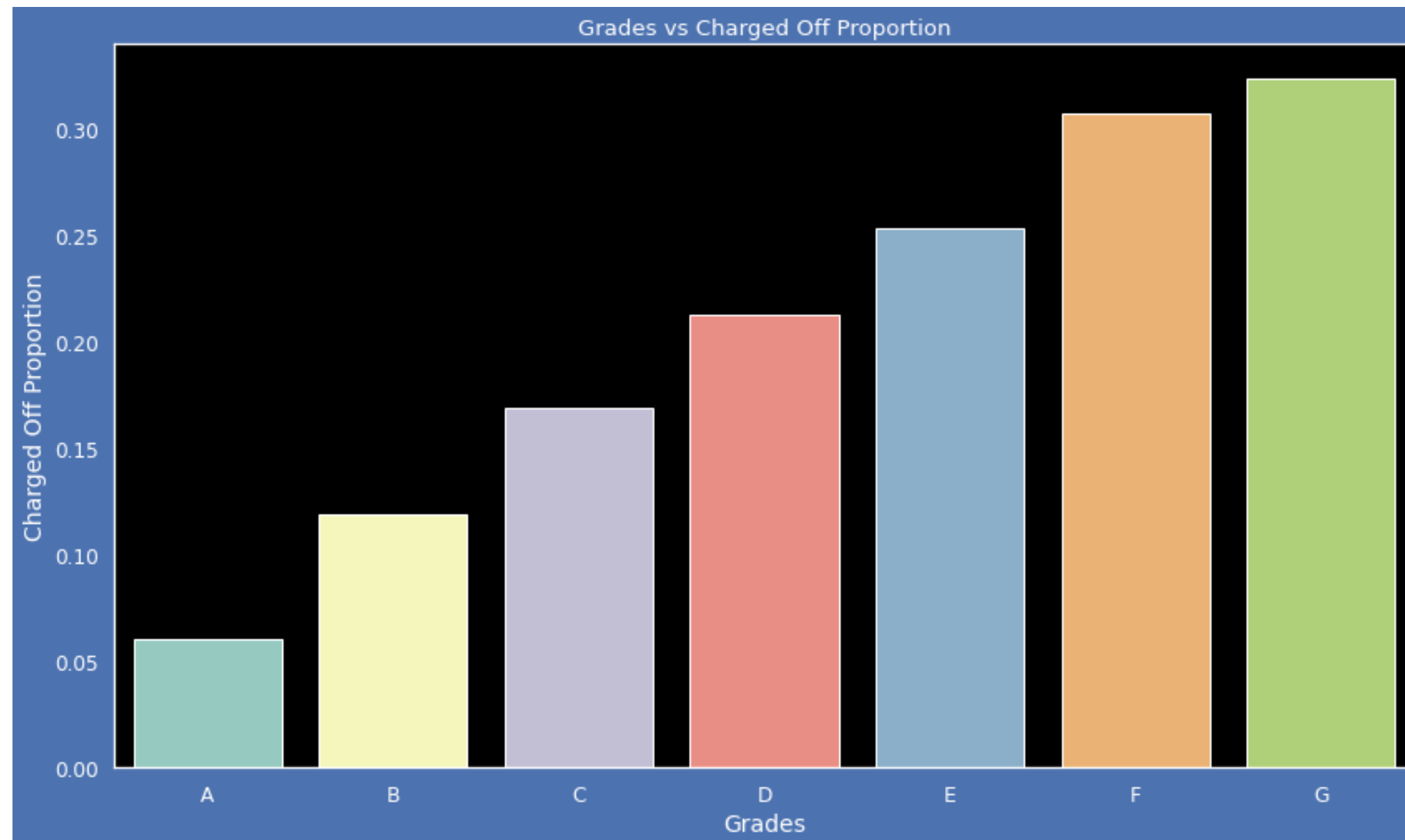




Finding 10: Bivariate analysis of grade against Charged Off Proportion



1. Grade A has least chances of getting charged off
2. Grade F and Grade G have high chances of getting charged off
3. Chances of getting charged off is increasing with grades moving from A to G



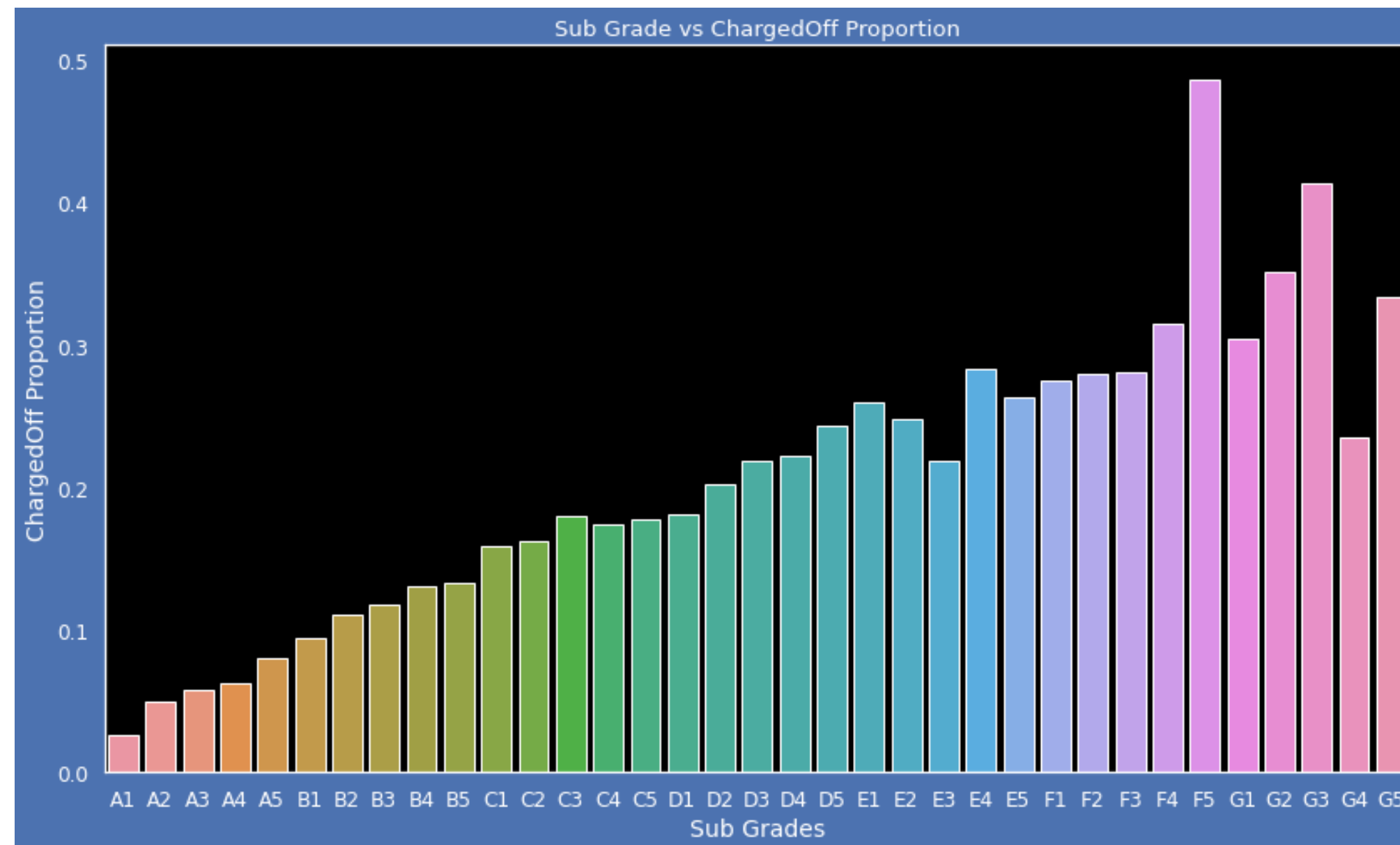


Finding 11: Bivariate Analysis on sub grade against Charged Off Proportion



Findings:

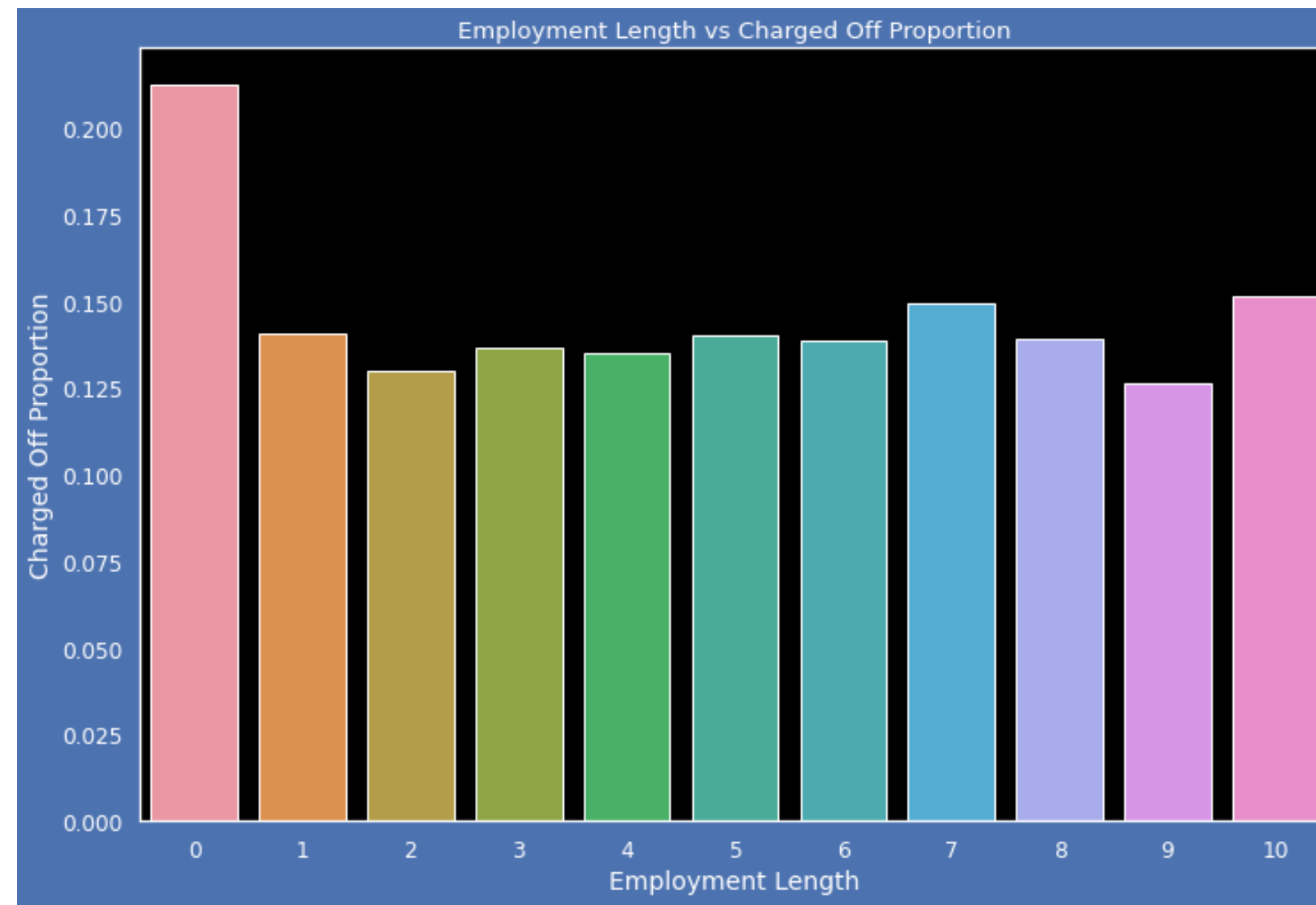
1. Sub Grade A has least chances of getting charged off
2. Sub Grade F and Sub Grade G have high chances of getting charged off
3. Chances of getting charged off is increasing with Subgrades moving from A to G



Finding 12: Bivariate Analysis on Employment Length Vs. Charged Off Proportion



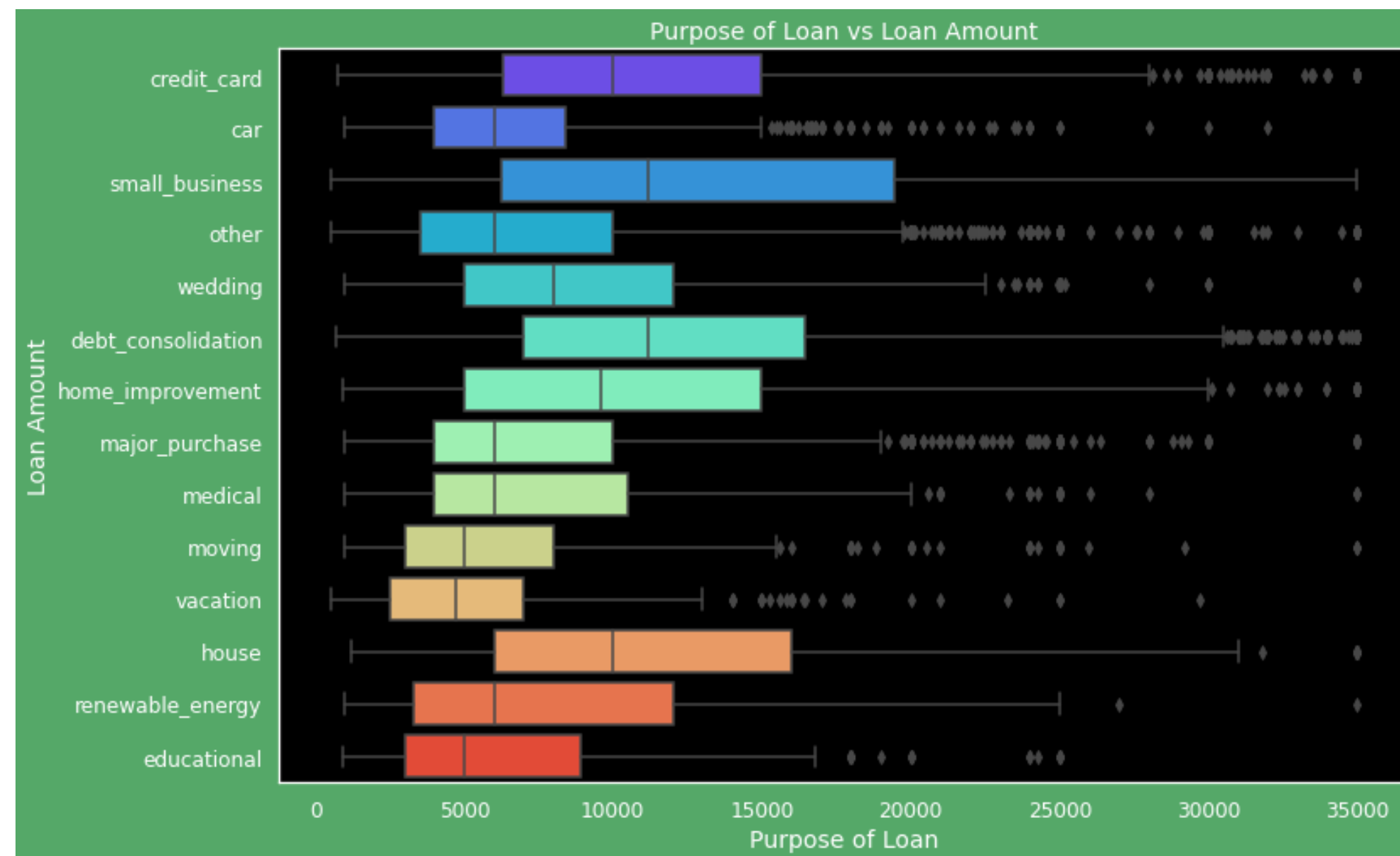
1. Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
2. It makes sense as with less or no experience they don't have source of income to repay loan.
3. Rest of the applicants have more or less same chances of getting charged off.



Finding 13: Bivariate Analysis Purpose of Loan Vs Loan amount



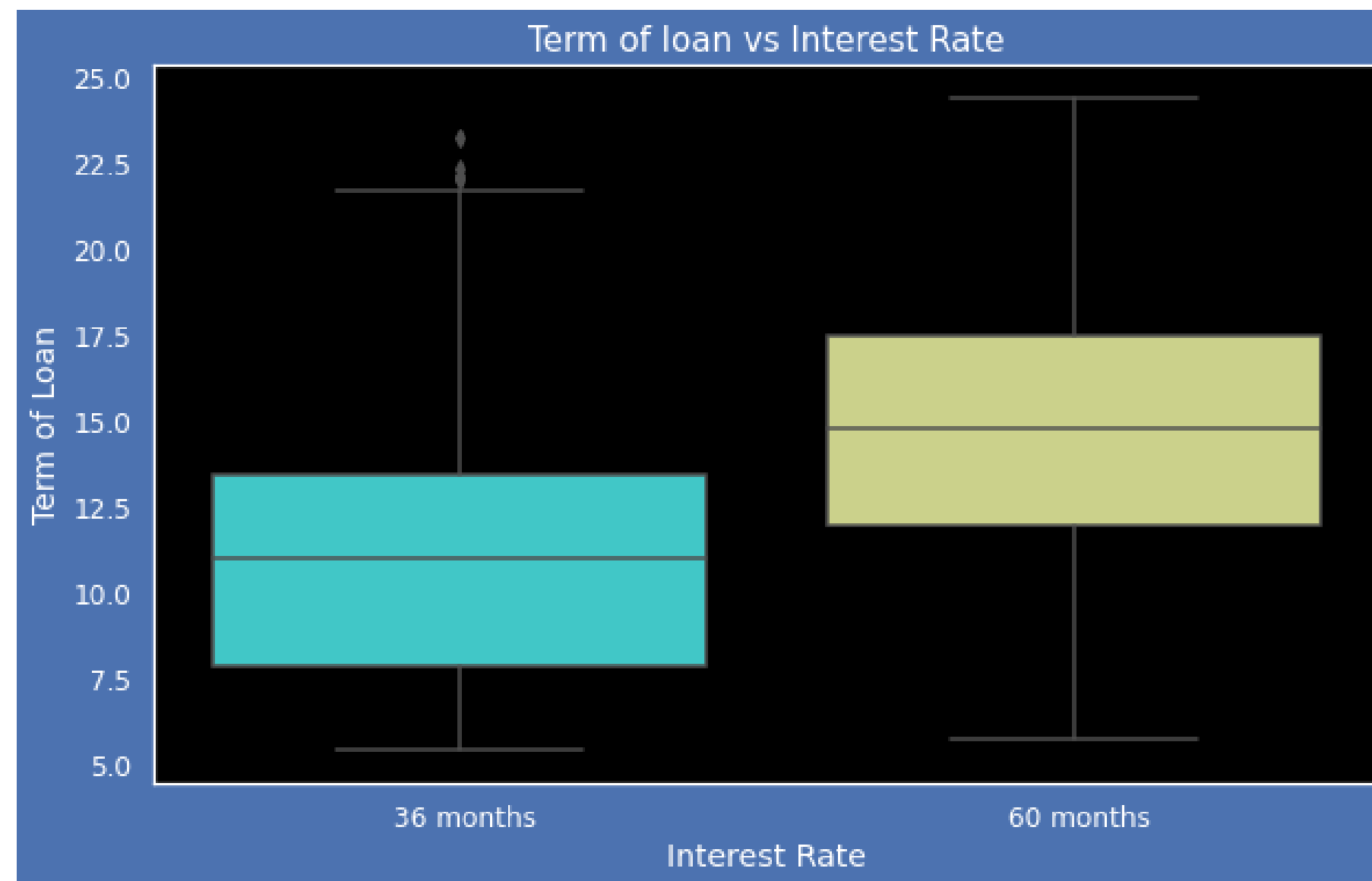
1. Median, 95th percentile, 75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.
2. Debt consolidation is second and Credit card comes 3rd.



Finding 14: Bivariate Analysis-Term of Loan vs. Interest Rate Box Plot



It is clear that average interest rate is higher for 60 months loan term. Most of the loans issued for longer term had higher interest rates for repayment.



Finding 15: Bivariate Analysis - Purpose vs Interest Rate

Box Plot

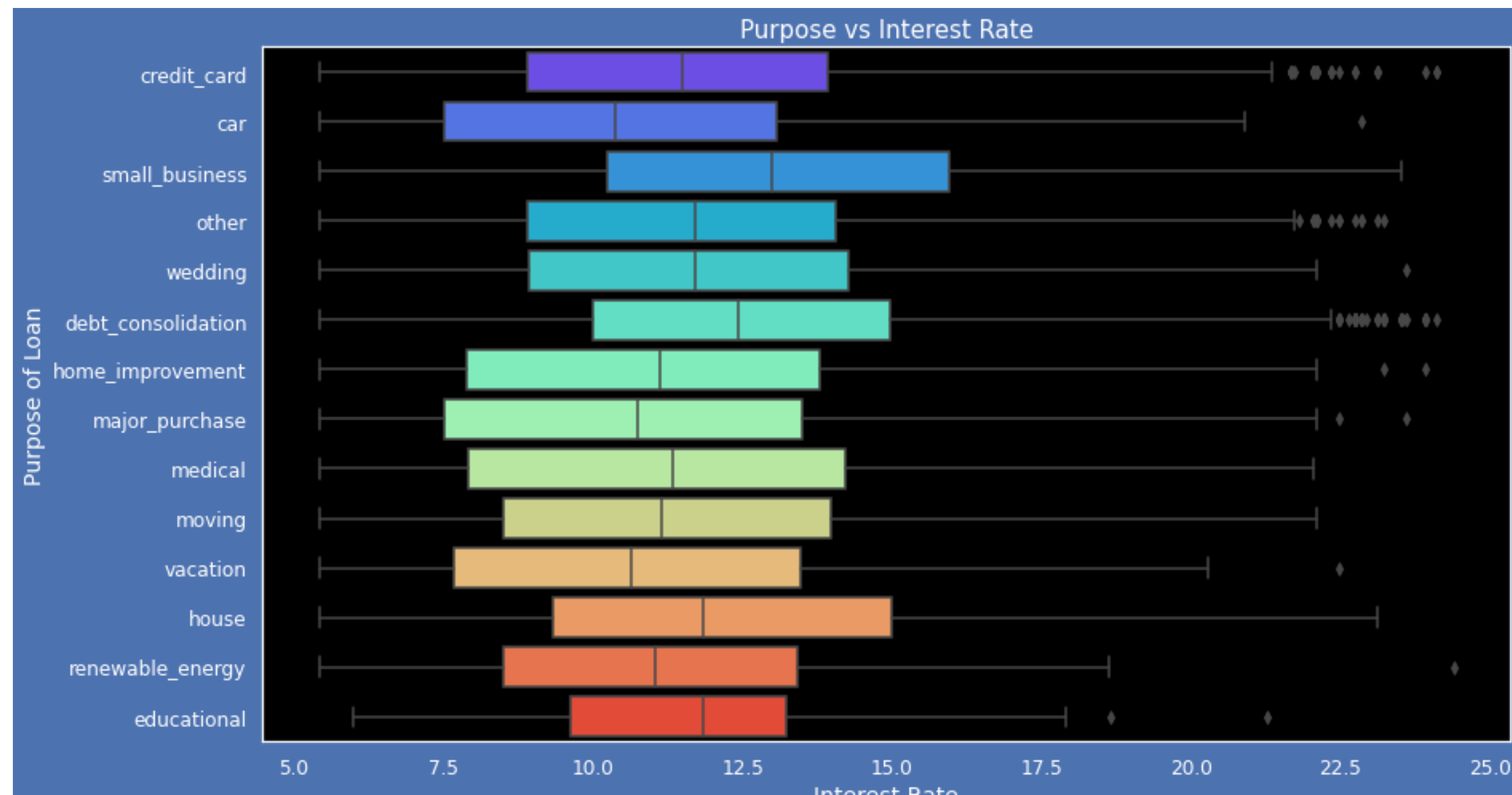


Observations:

It is clear that average interest rate is highest for small business purpose.

Loans taken for small business purposes had to repay the loan with more interest rate as compared to other.

Debt consolidation is 2nd where borrowers had to pay more interest rate.



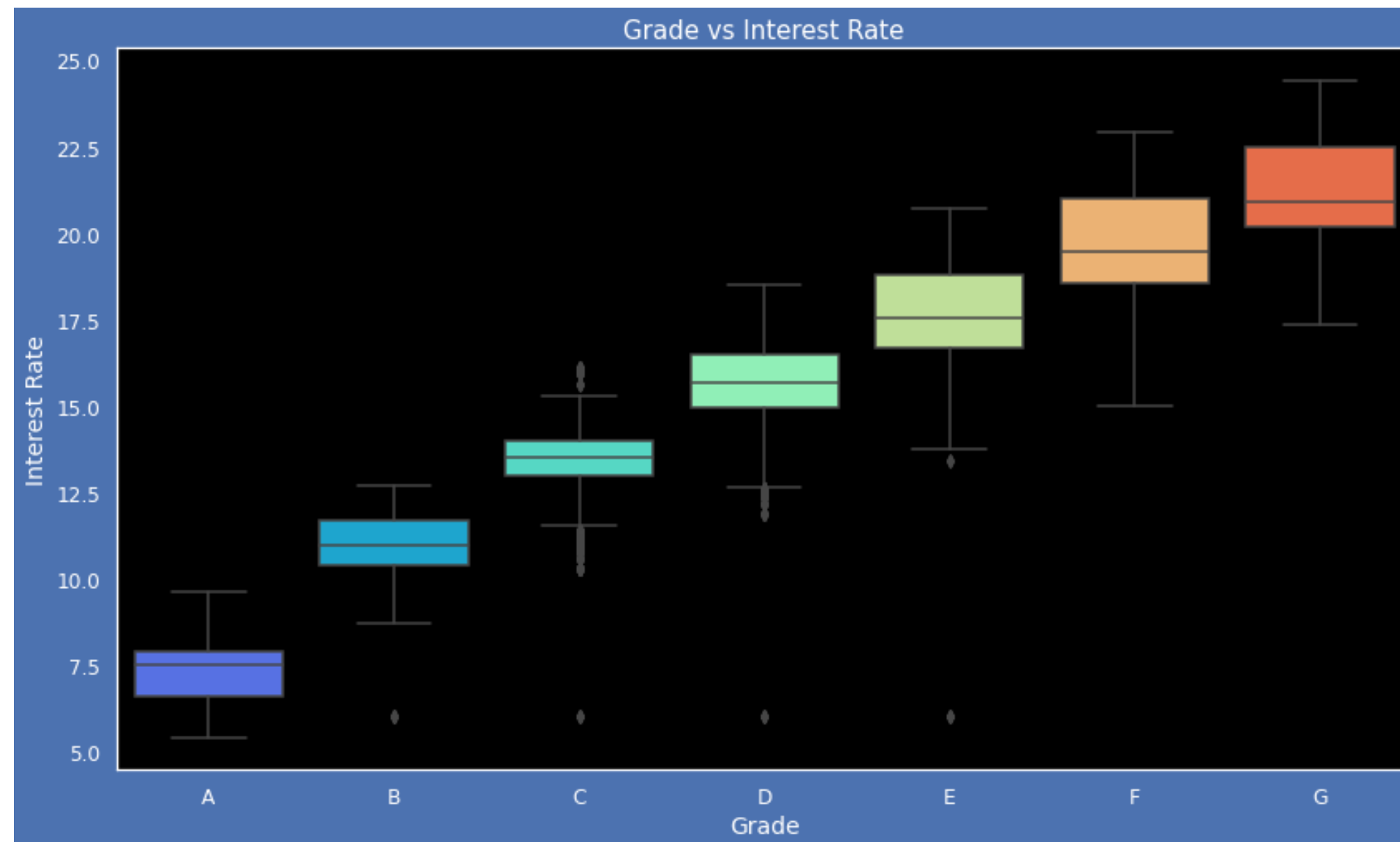
Finding 16: Bivariate Analysis - Grade vs Interest Rate



A-grade is a top letter grade for a lender to assign to a borrower.

The higher the borrower's credit grade, the lower the interest rate offered to that borrower on a loan.

It is clear that interest rate is increasing with grades moving from A to F.



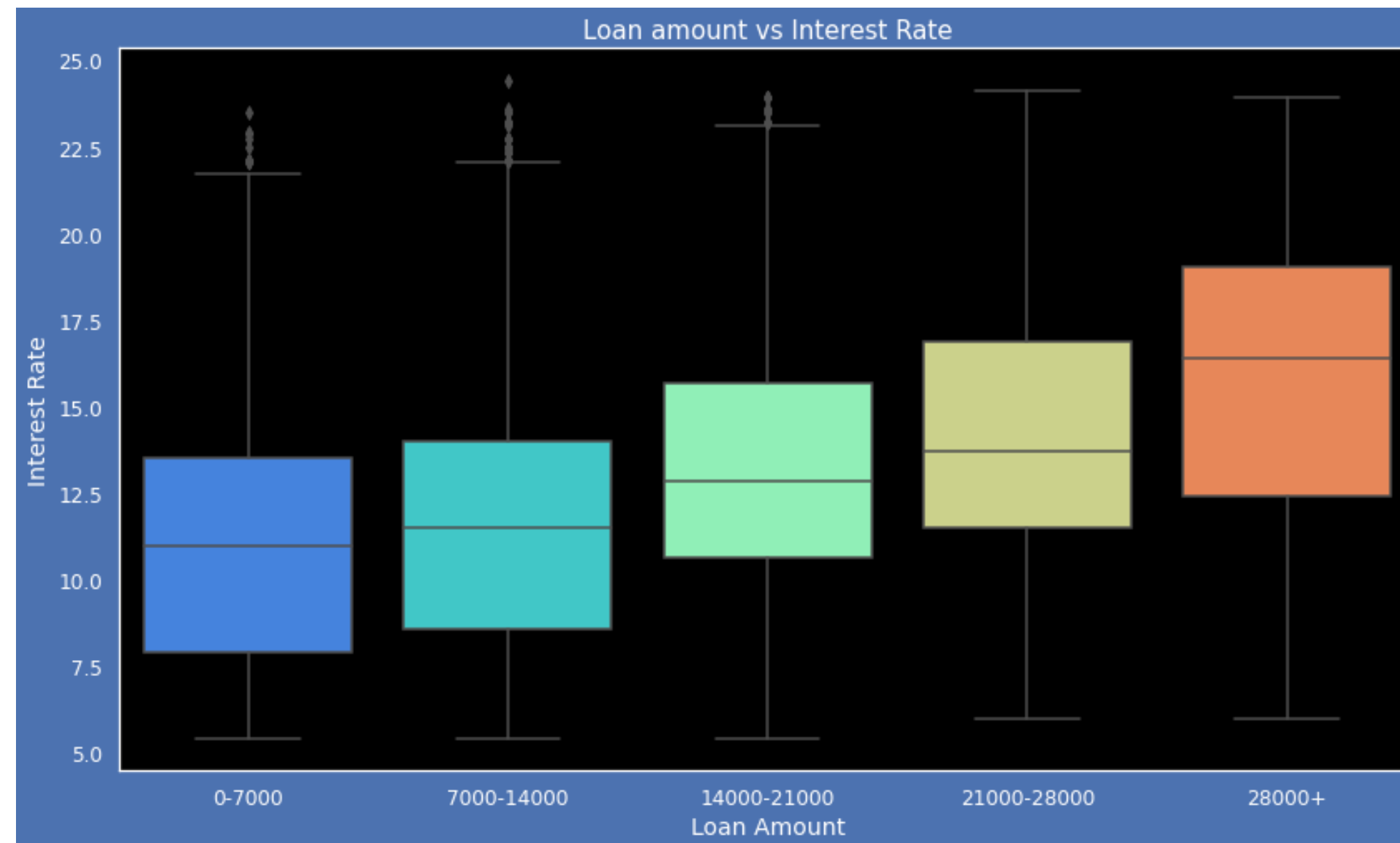
Finding 17: Bivariate Analysis - Loan Amount vs Interest Rate



Observations:

It is clear that interest rate is increasing with loan amount increase.

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.



Finding 18: Bivariate Analysis - DTI vs Interest Rate

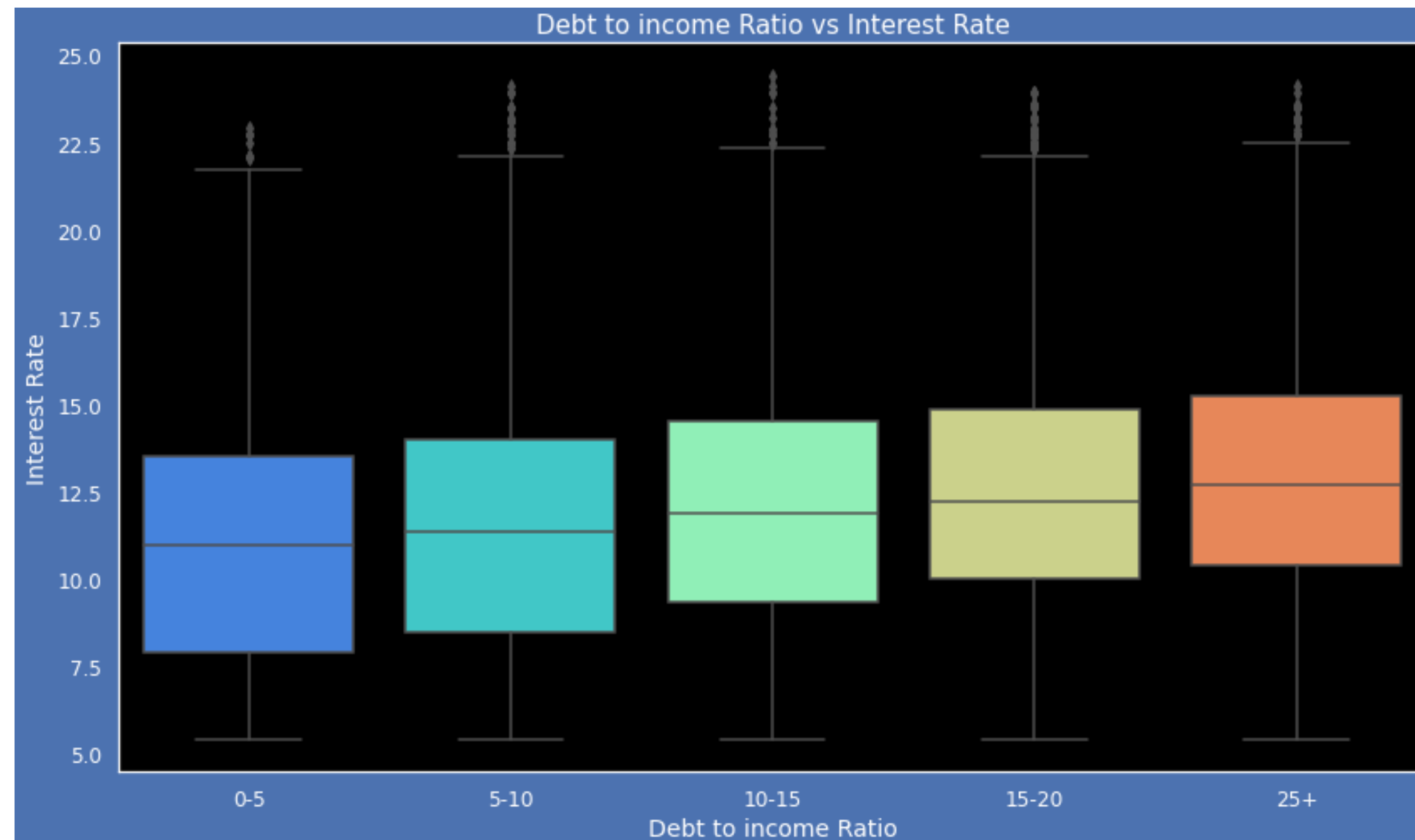
Box Plot



Observations:

If your DTI is low enough you may get a lower interest rate

Plot shows no significant variation but there is slight increase in interest rate with increase in DTI.



Conclusion



1. Lending Club will be in profit if they give more loan amount as interest rate increases with higher loan amount
2. A-grade is a top letter grade for a lender to assign to a borrower. Grades should be evaluated thoroughly before lending money
3. A derogatory item is an entry that may be considered negative by lenders because it indicates risk and hurts your ability to qualify for credit or other services. Public records and collections are derogatory items because they reflect financial obligations that were not paid as agreed. Those who already have pub_rec value 1 or 2 have charged off chances higher than who have no Derogatory Public Record. pub_rec count 3-4 has less numbers so cannot reach on any conclusions.
4. States NV, CA and FL states shows good number of charged offs in good number of applications. So, Lending Club should issue less loans in these states.
5. Those who are not working or have less than 1 year of work experience have high chances of getting charged off. So, Lending Club should issue less loans to these kind of applicants.
6. Small Business Owners default most. So, Lending Club should issue less loans to these kind of applicants.



UpGrad