



Lending Club Case Study

By.

Allena Venkata Sai Abhishek



Index

- Identifying Business Problem
- Data Description
- Data Preparation and Processing
- Data Cleaning and Manipulation
- Models:
 - Univariate Analysis
 - Bivariate Analysis
 - Segmented Univariate Analysis
 - Multivariate Analysis
- Conclusion
- Recommendation

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 the analysis is to use the information about past loan applicants and make a data-driven decision about 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 with 111 columns/features 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
- Deleted rows containing either 50% or more than 50% NaN Values
- Outliers were Identified and treated



Data Modelling

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

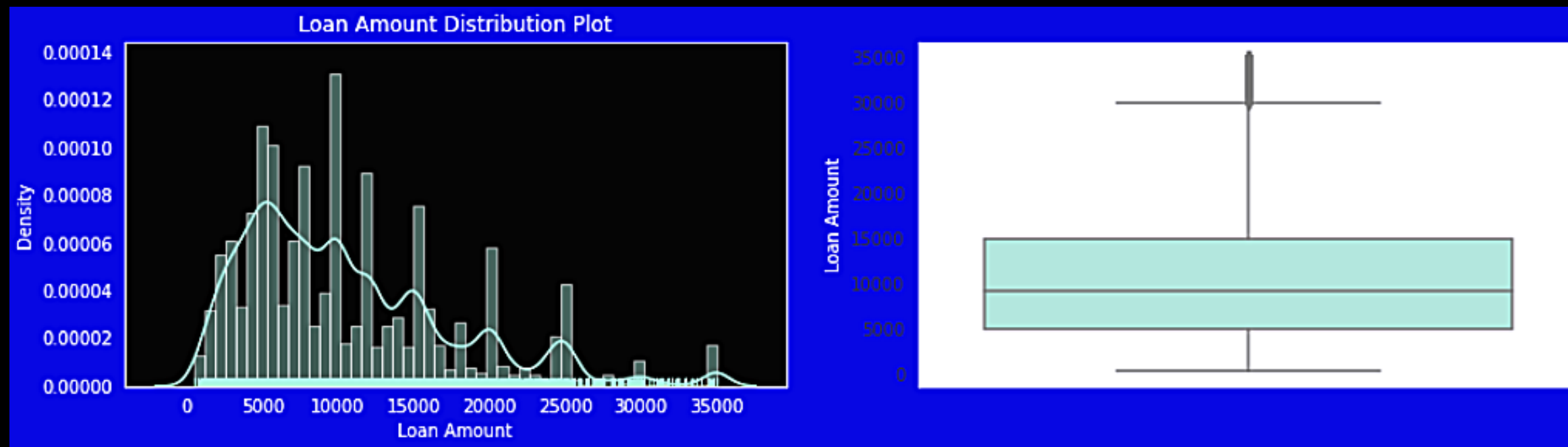
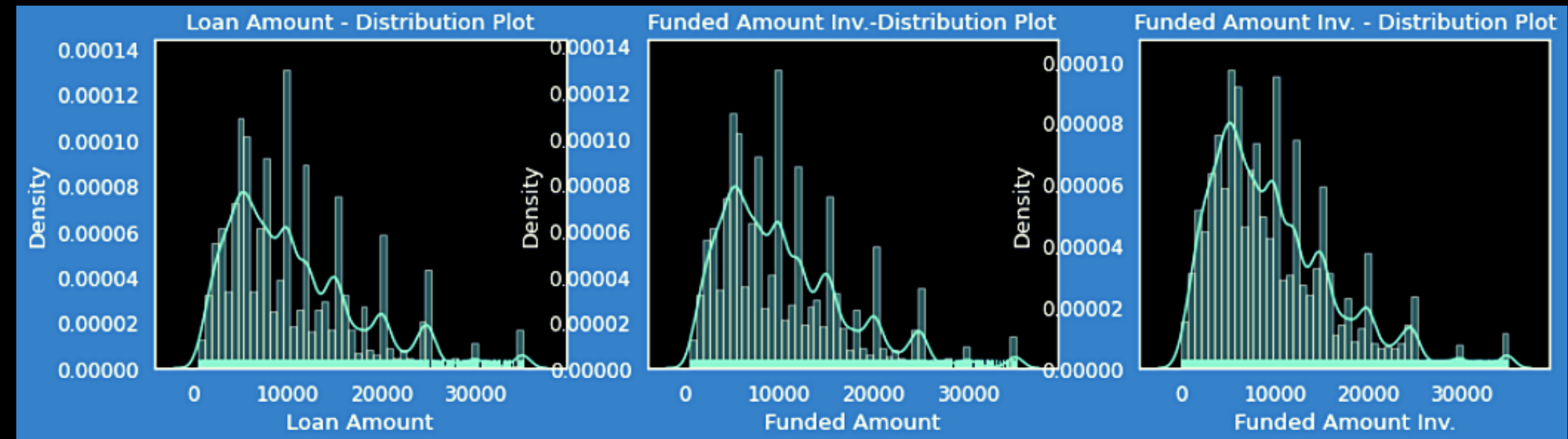
- **Univariate analysis-Box Plots**
- **Univariate Analysis-Distribution Plots**
- **Univariate & Bivariate Analysis-Bar Plots**
- **Bivariate analysis-Violin Plots**
- **Bivariate Analysis-Heat Maps**
- **Bivariate Analysis-Correlation plot**
- **Bivariate Analysis-Box Plots**
- **Segmented Analysis-Count Plot**
- **Segmented Analysis-Bar Plots**
- **Multivariate analysis-Density Plots**
- **Multivariate analysis-PairPlots**

1. Univariate Analysis

Univariate analysis-Loan, Funded, Funded Inv Amount Variable Loan Amount Variable

Findings

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

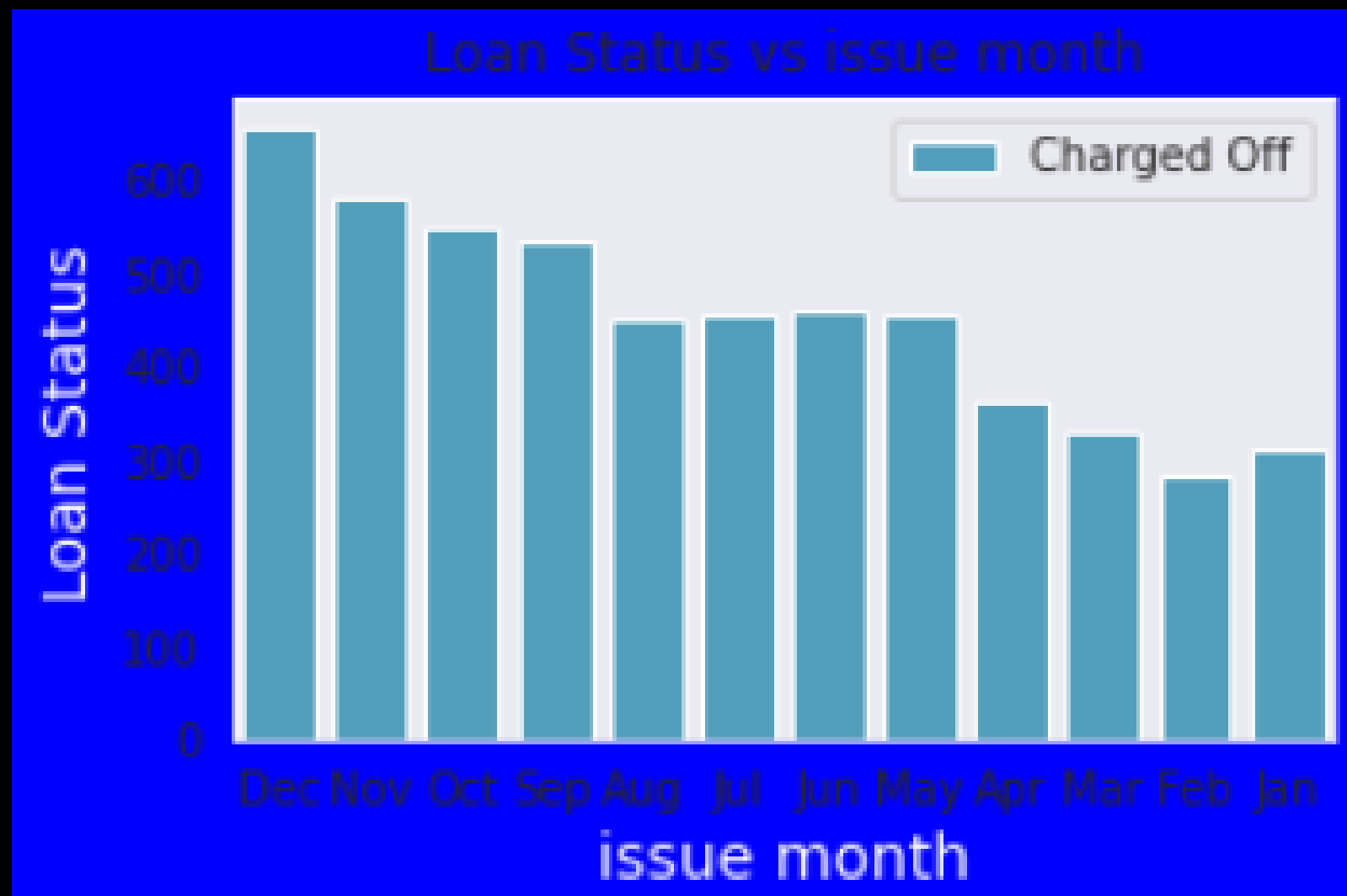


- Most of the loan amounts are in the range 5000-15000

Univariate analysis-Issue Month & Year

Findings

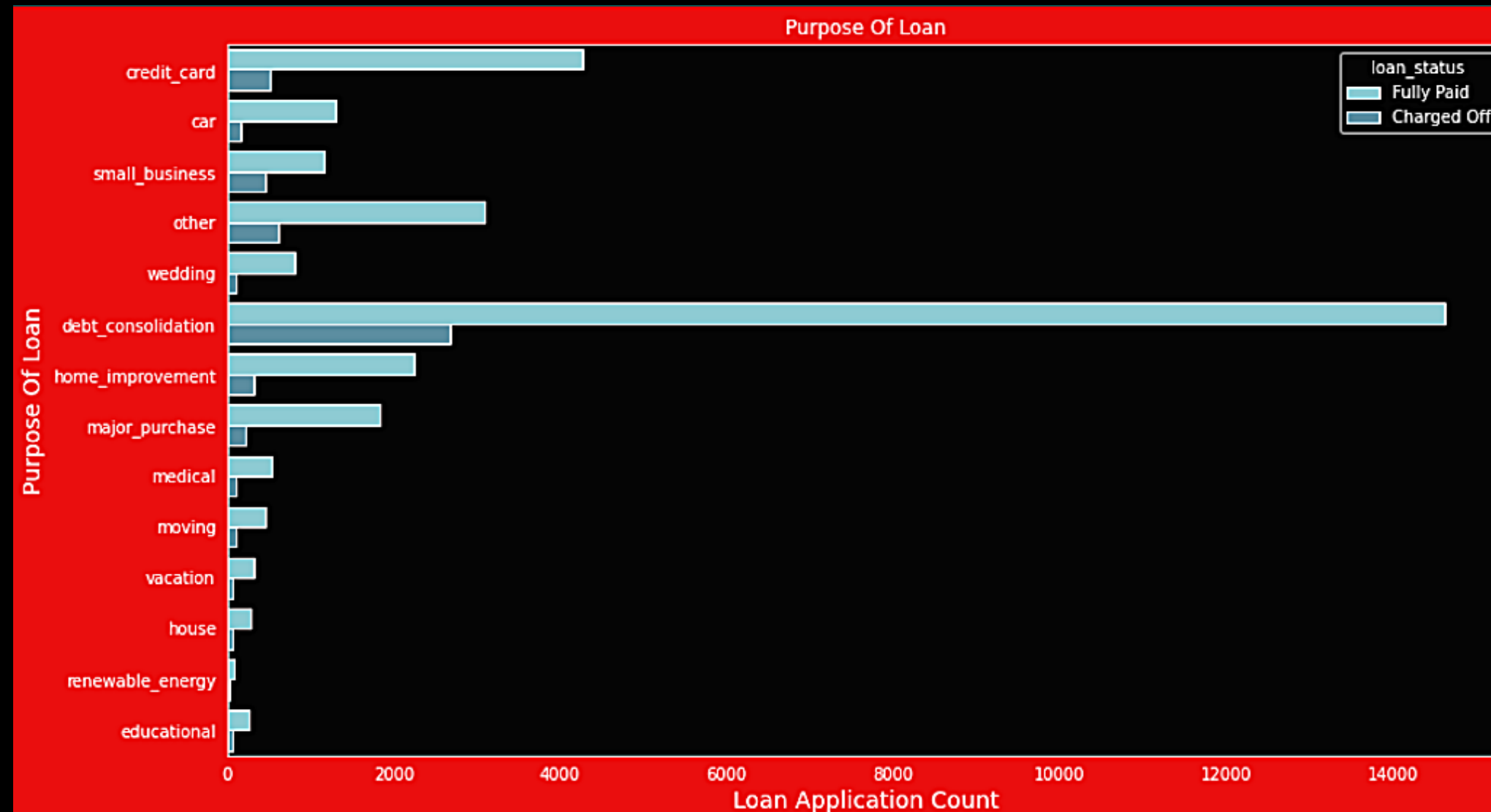
Most of the charged off are in December & in 2011



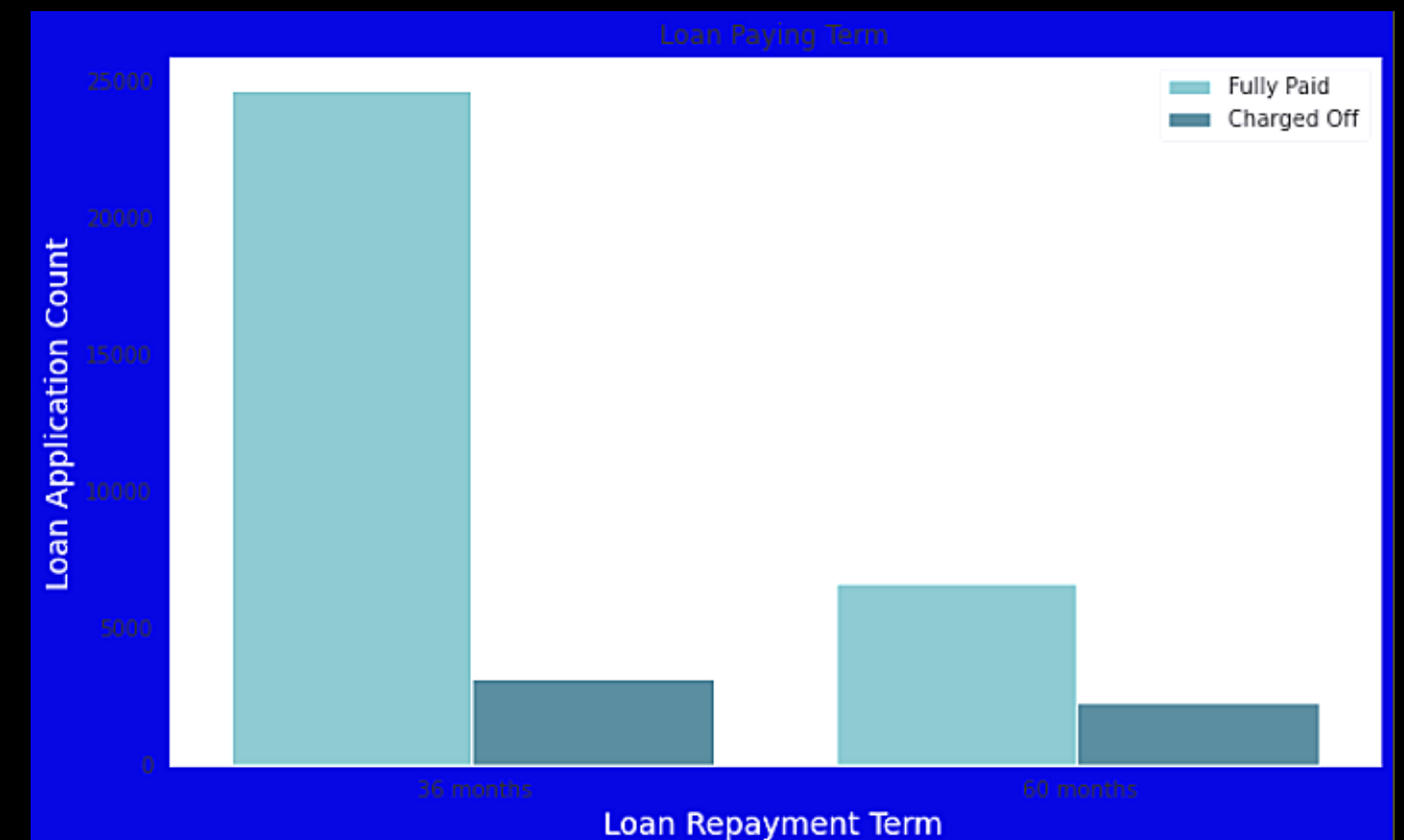
Univariate analysis - Loan Purpose & Loan Paying Term

Findings

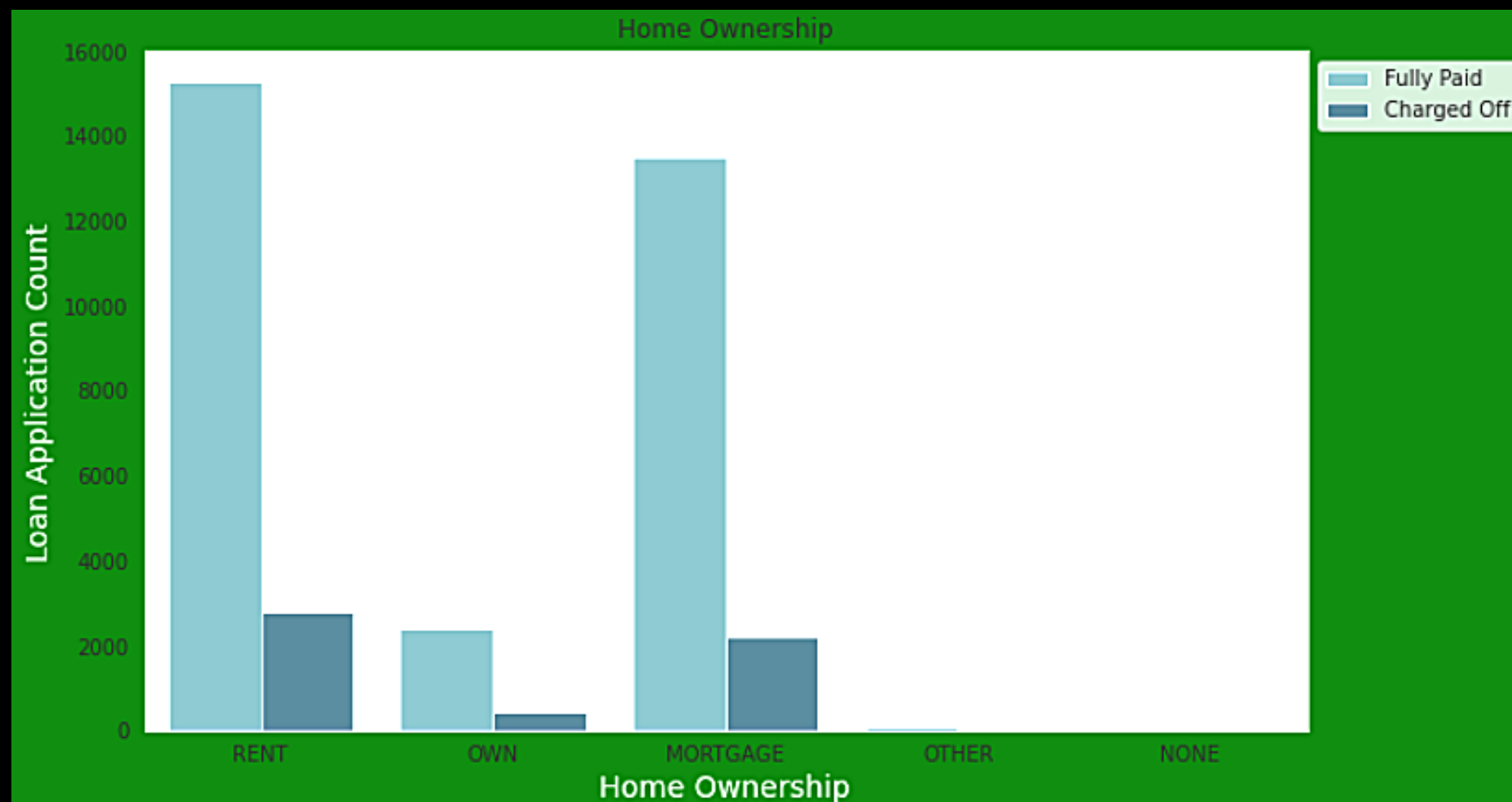
- The Count Plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill.
- Number of charged off count also high too for these loans.



- The 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.

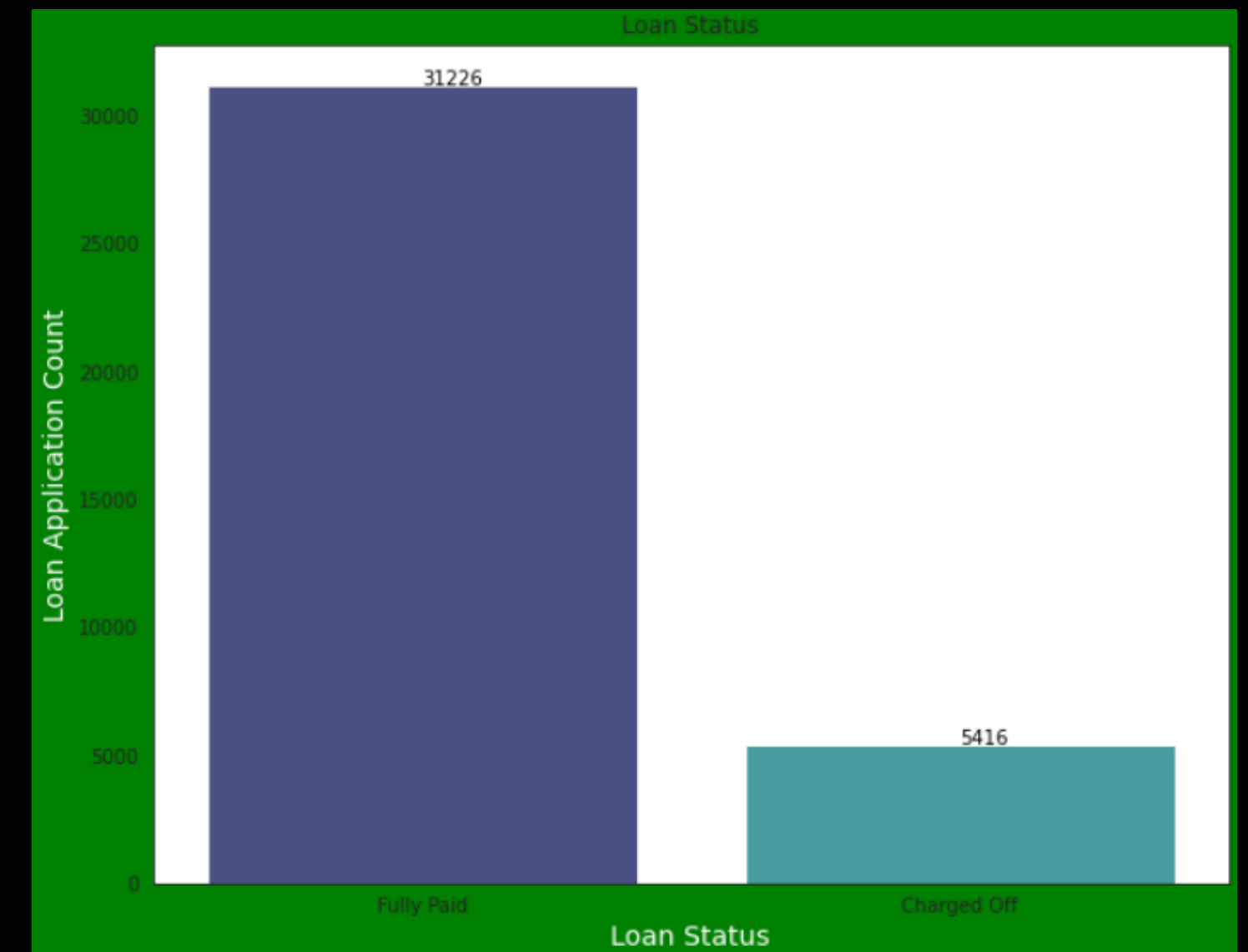


Univariate analysis - Loan Status & Home Ownership



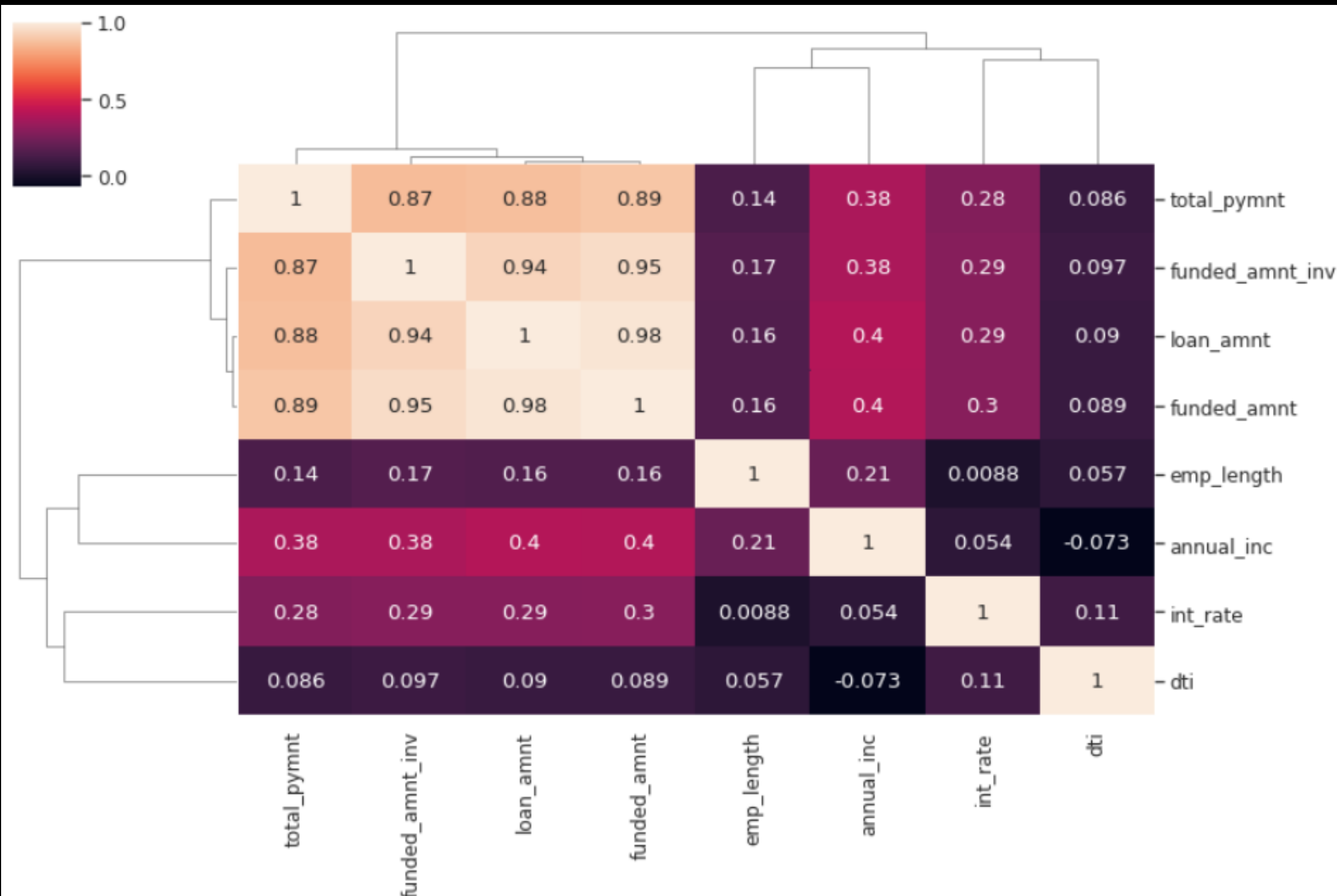
Findings

- The following plot shows that around 14.7% loans were charged off out of the total loans issued



- The Count 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.

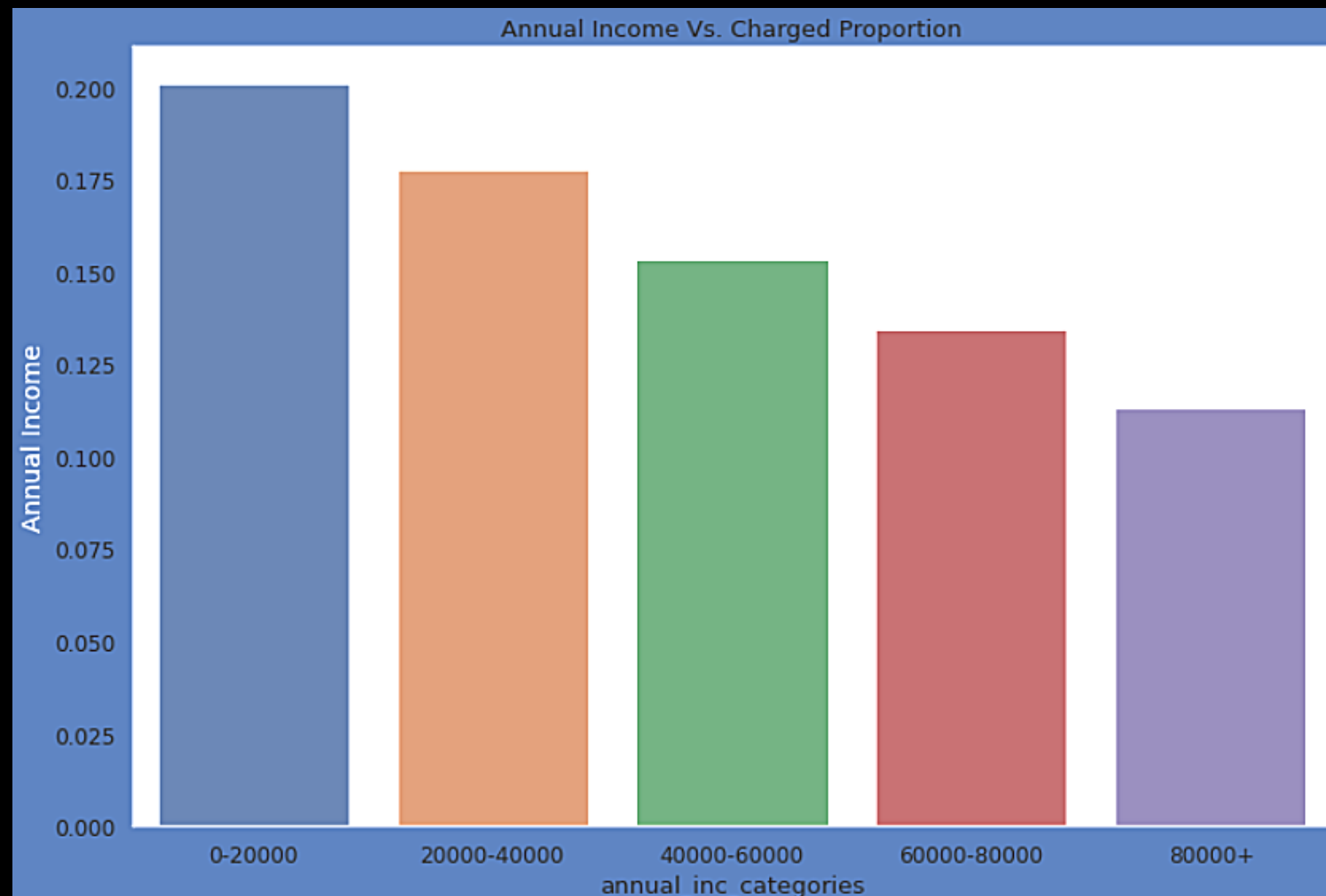
2. Bivariate Analysis



Bivariate analysis - Correlation

- Annual Income to Debt To Income Ratio i.e. dti are negatively correlated
- Loan Amount, Investor Amount, and Funding Amount are strongly correlated
- Positive correlation between Annual Income and employment years
- Positive correlation between annual income and the funded amount that means people with high income get a high funded amount
- Positive correlation between annual income and total payment

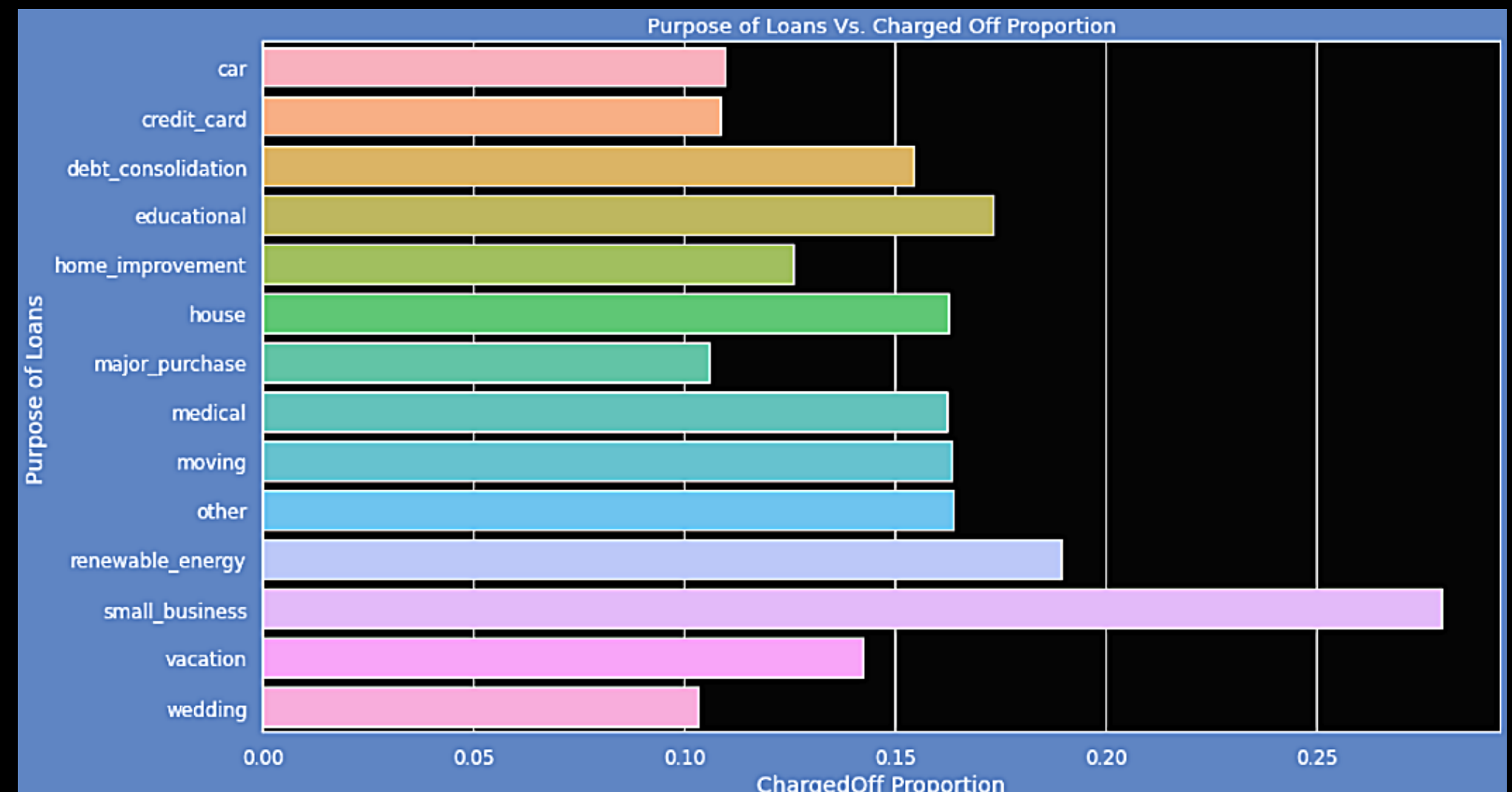
Bivariate analysis - Annual Income Vs. Charged Off & Purpose of Loan Vs Charged Off



- Small Business applicants have high chances of getting charged off.
- Renewable_energy where charged off proportion is better as compare to other categories.

Findings

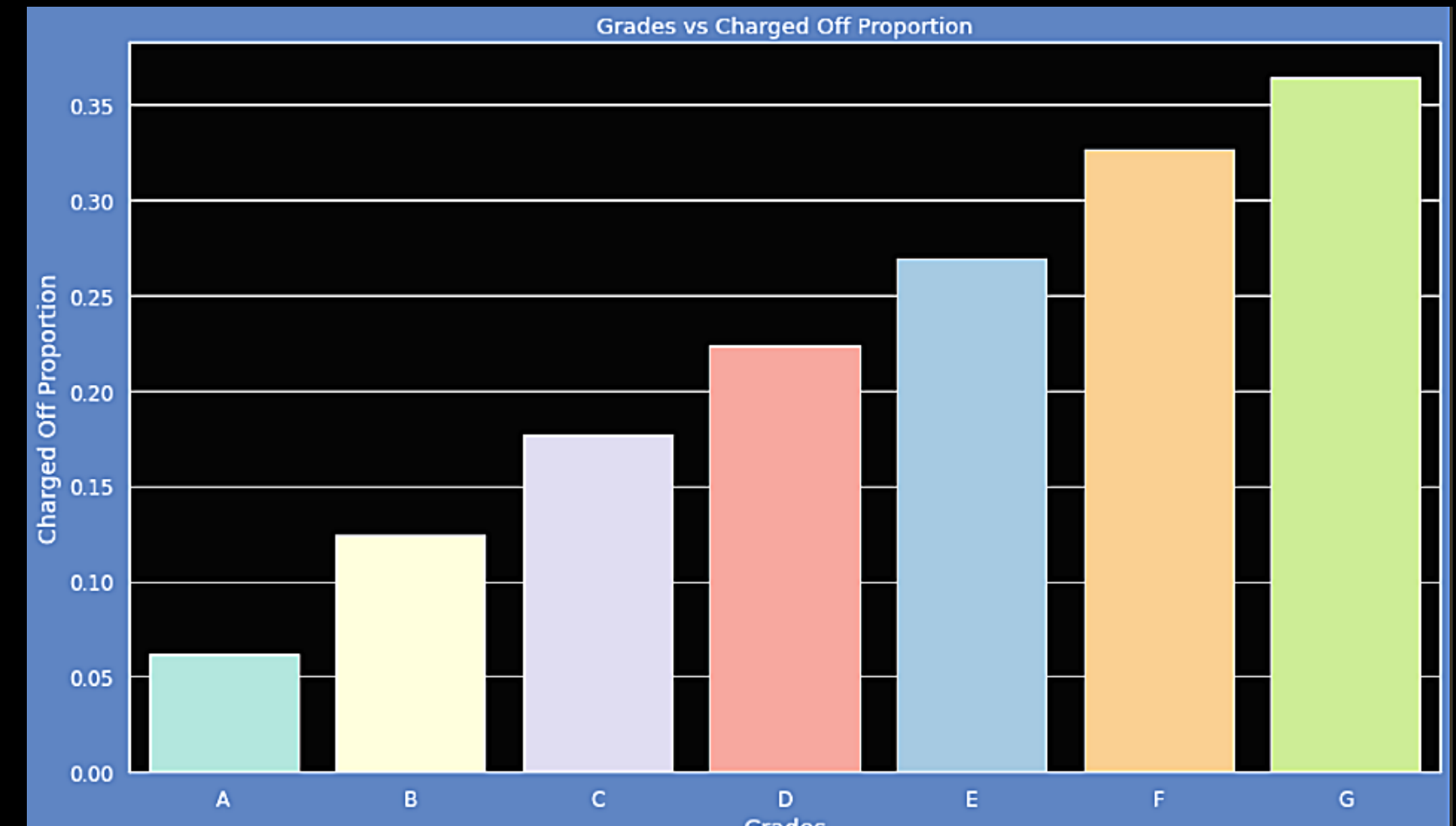
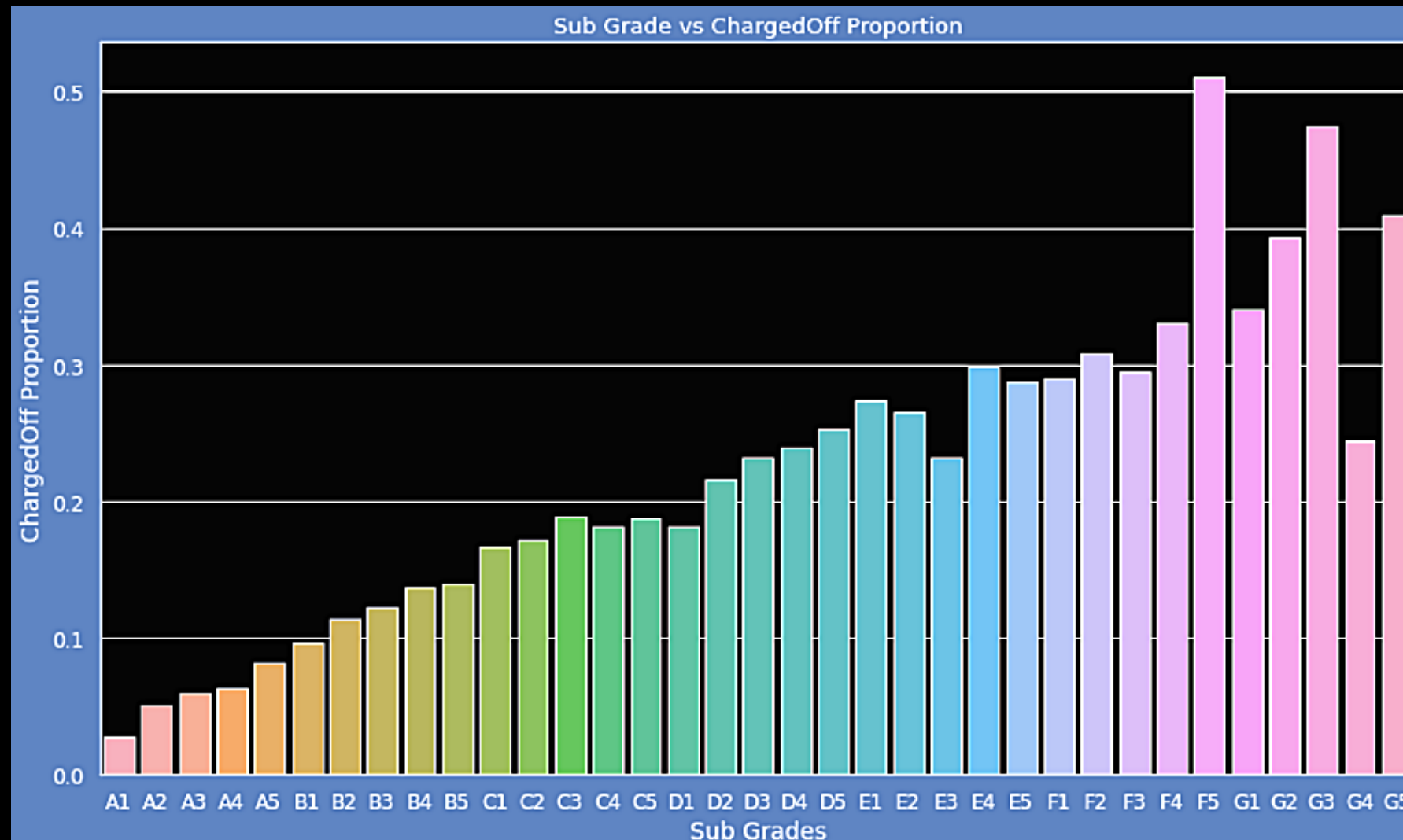
- Income range 80000+ has fewer chances of charged off.
- Income range 0-20000 has high chances of charged off.
- Notice that with increase in annual income charged off proportion got decreased. So, they are inversely proportional.



Bivariate analysis - Grade against Charged Off

Findings

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

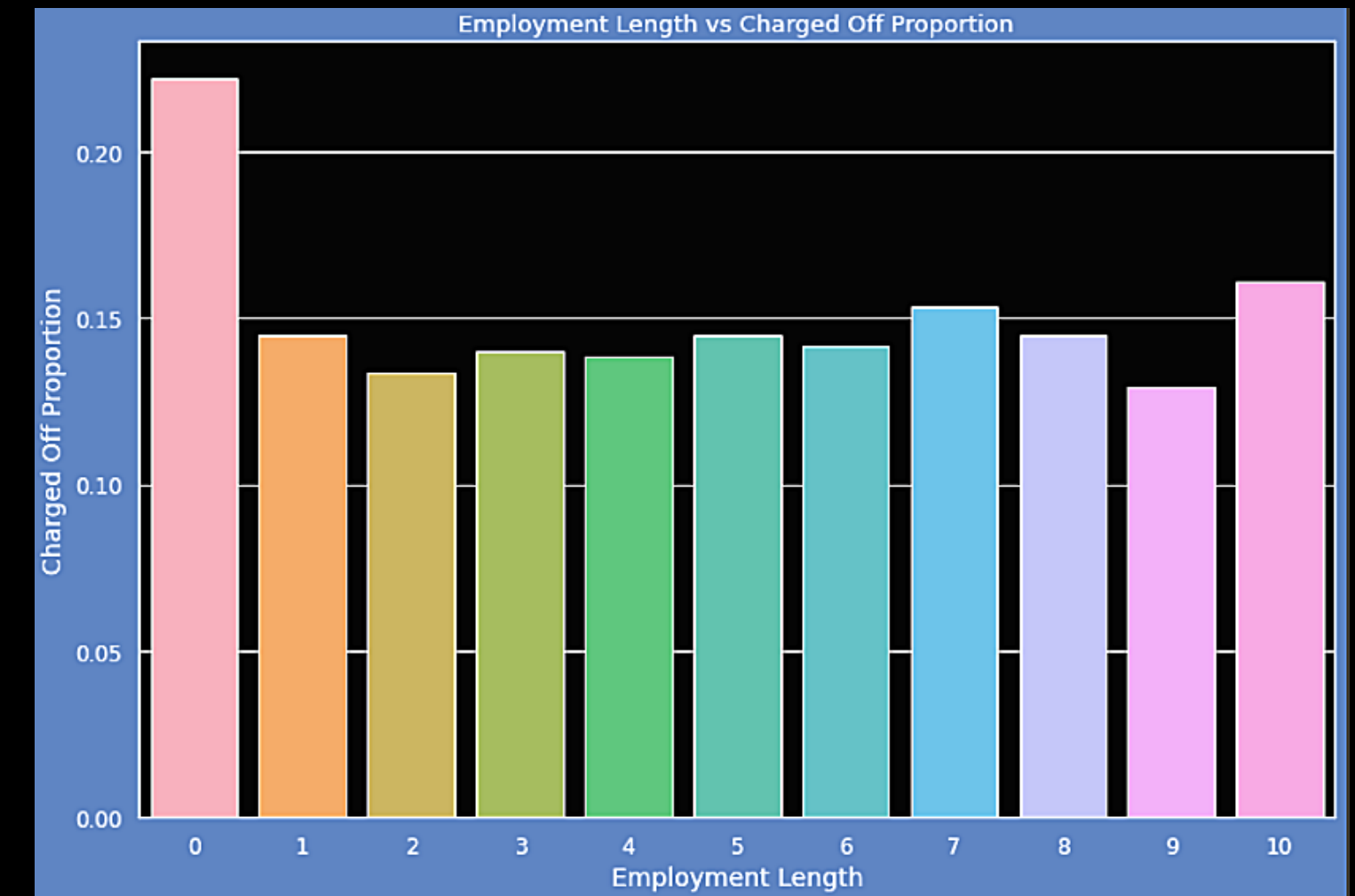


- Sub Grade A has least chances of getting charged off
- Sub Grade F and Sub Grade G have high chances of getting charged off
- Chances of getting charged off is increasing with Subgrades moving from A to G

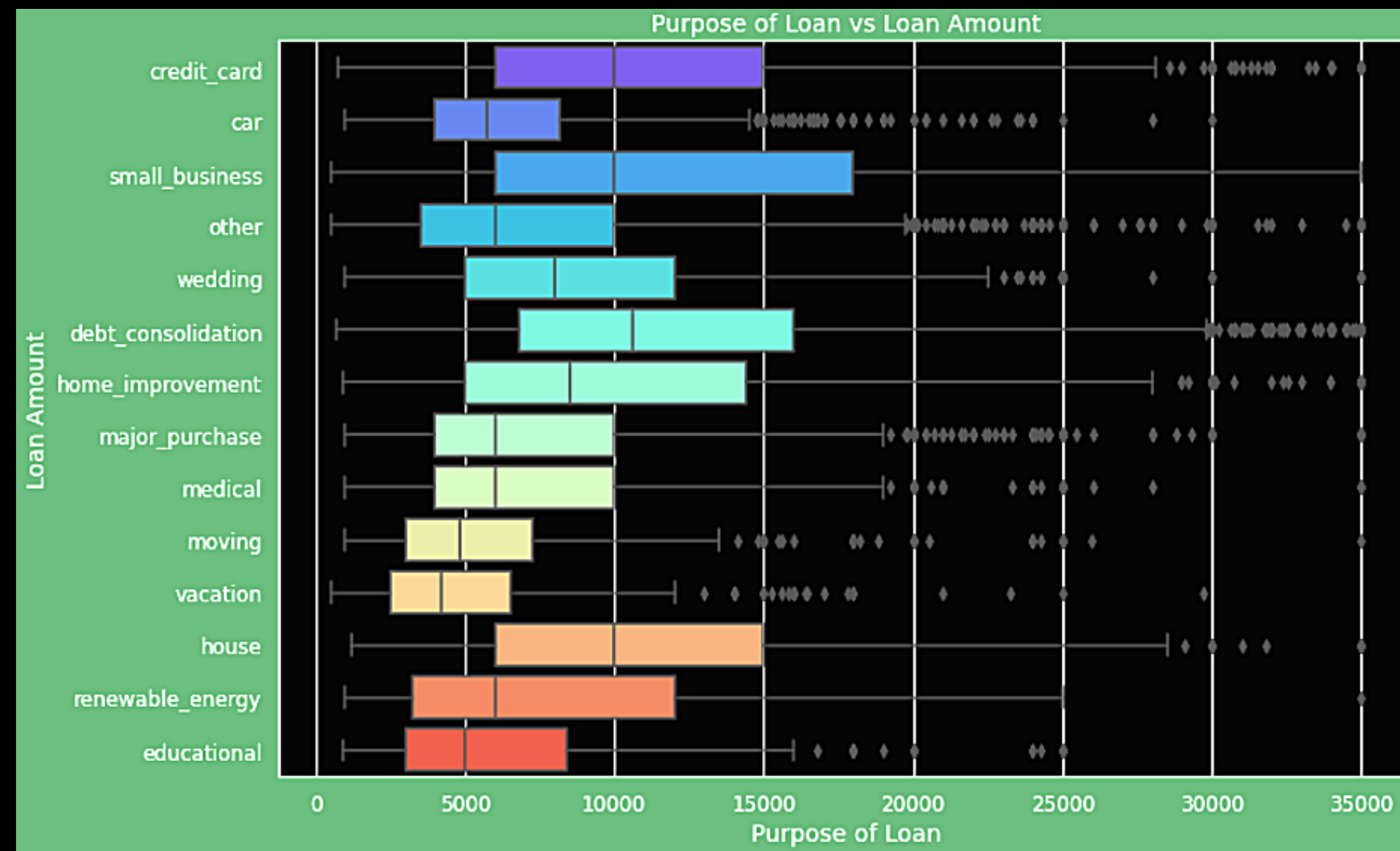
Bivariate analysis - Employment Length vs Charged Off

Findings

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



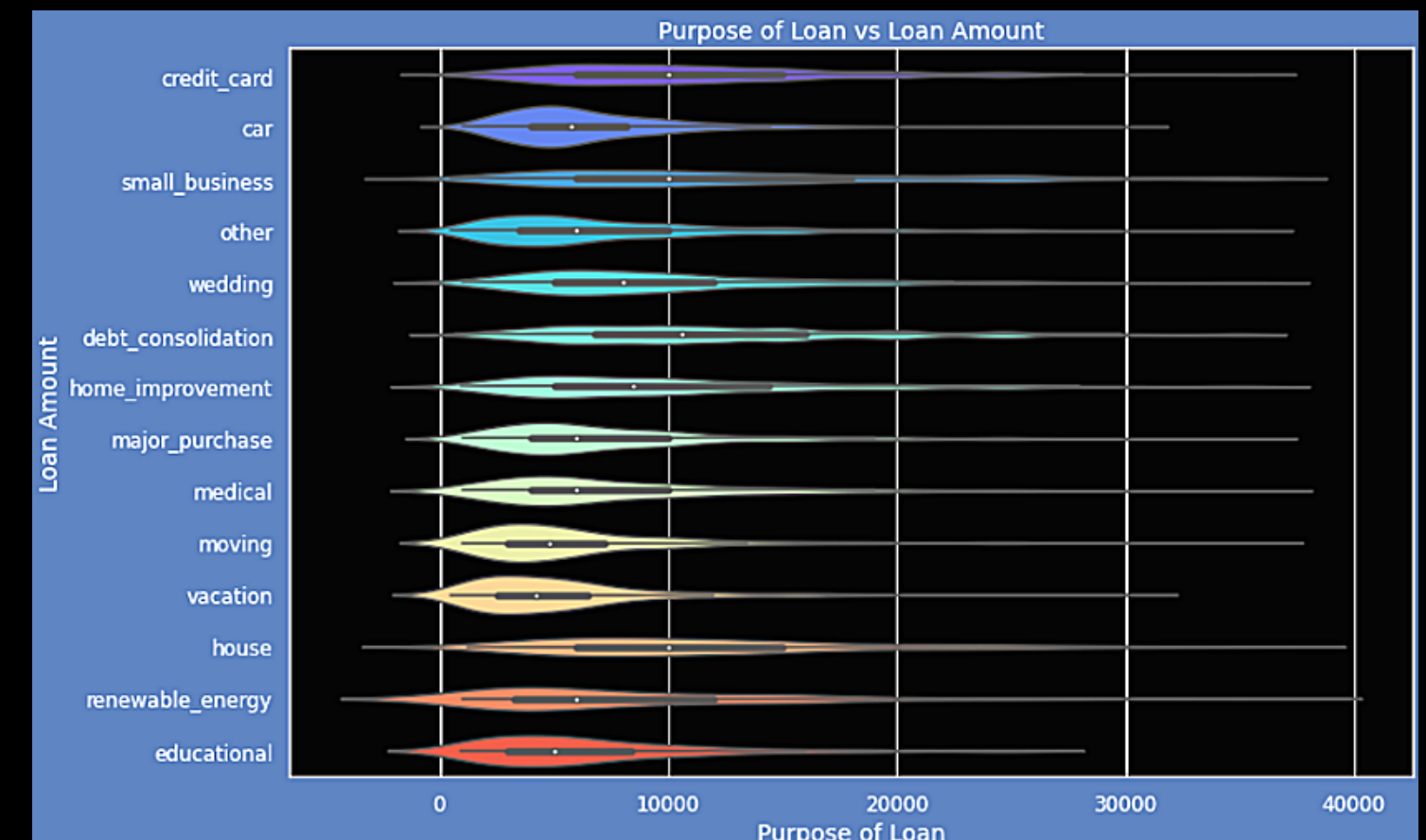
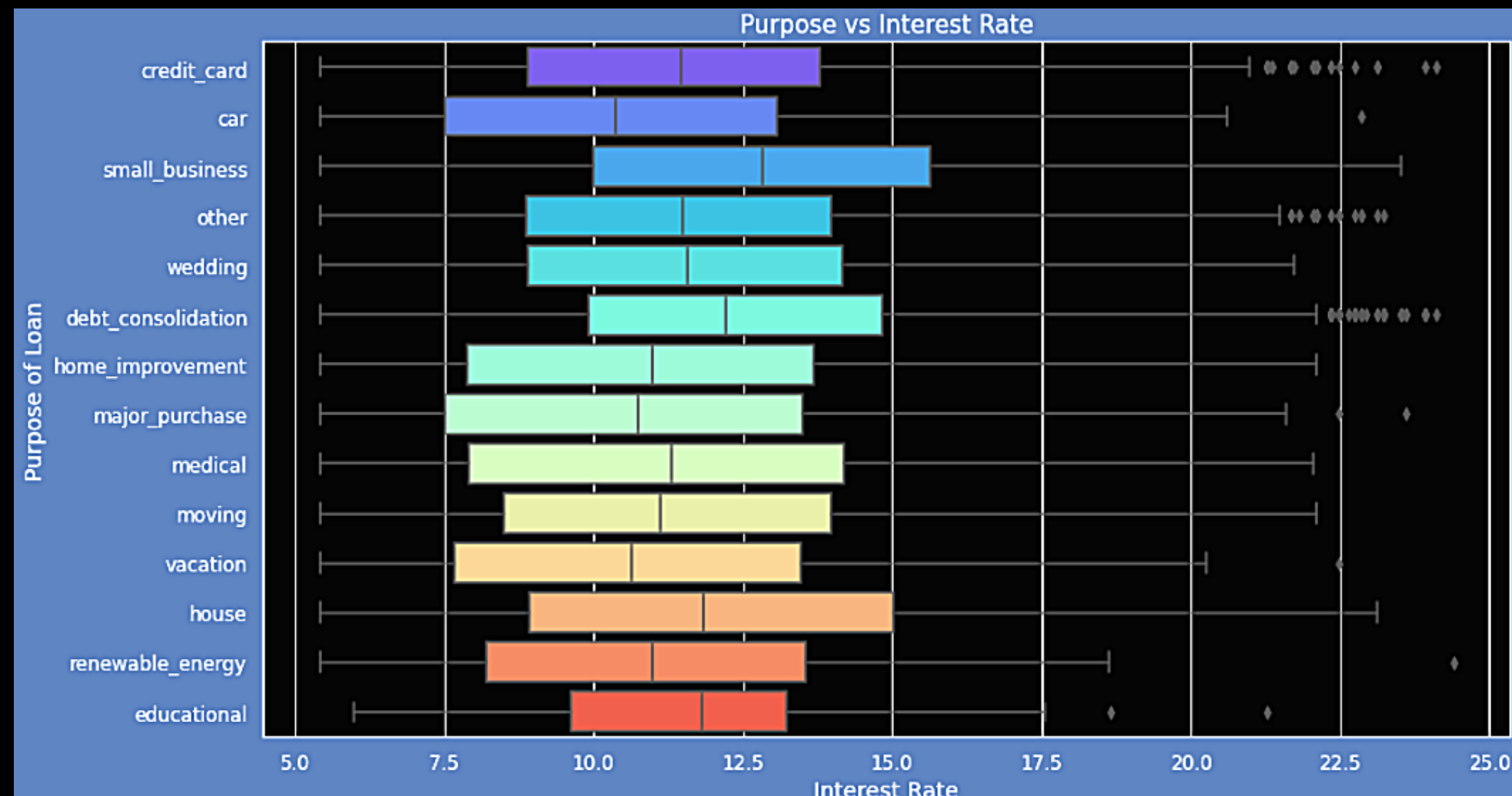
- Median, 95th percentile, 75th percentile of loan amount is highest for loan taken for small business purpose among all purposes.
- Debt consolidation is second and Credit card comes 3rd.



Bivariate analysis - Purpose of Loan vs loan Amount & Purpose vs Interest Rate

Findings

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

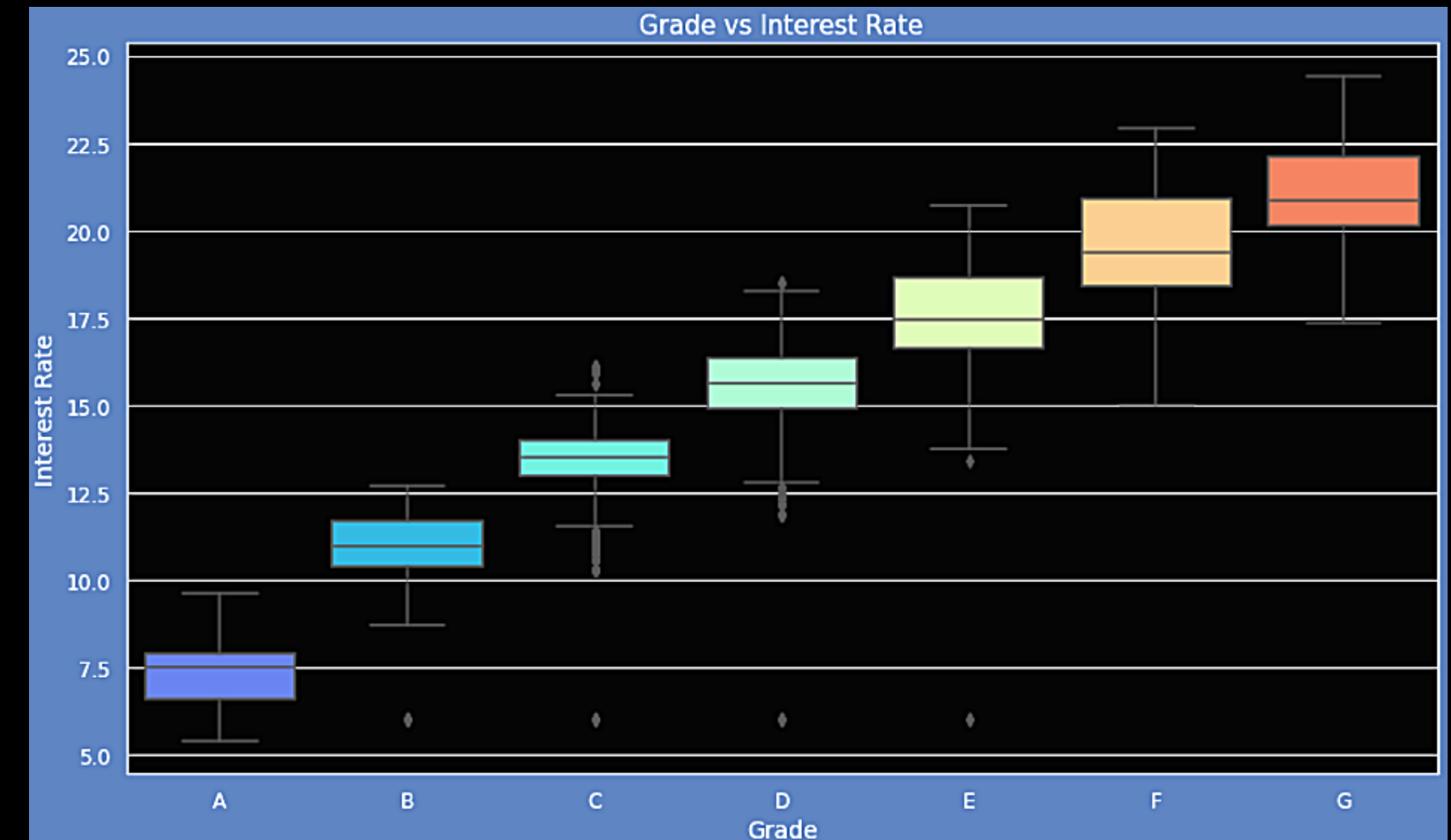
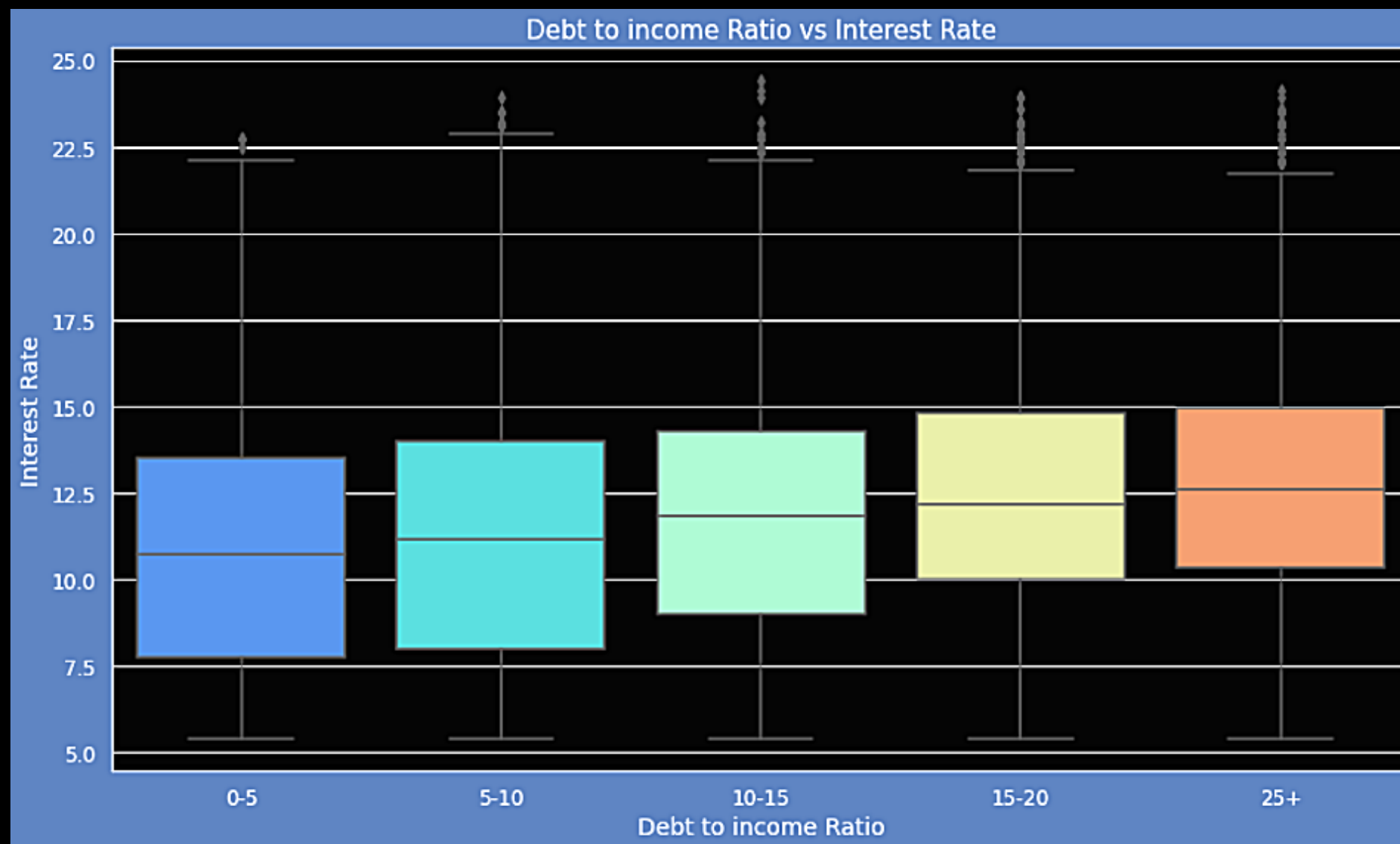


- It is clear that average interest rate is highest for small business purpose.
- Loans taken for small business purposes had to repay the loan with more interest rate as compared to other.
- Debt consolidation is 2nd where borrowers had to pay more interest rate.

Bivariate analysis - Grade vs Interest Rate & DTI vs Interest Rate

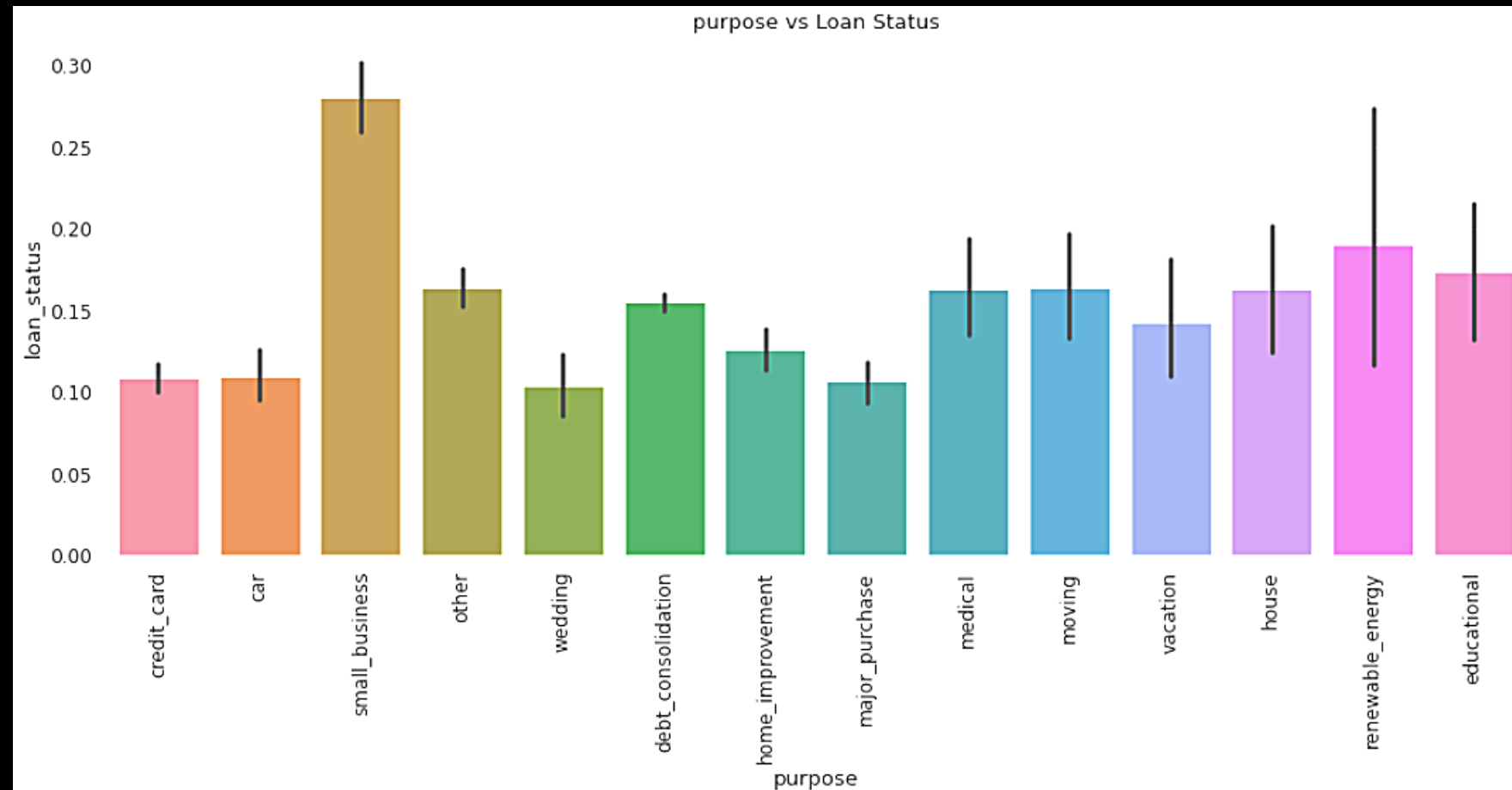
Findings

- 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.



- 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.

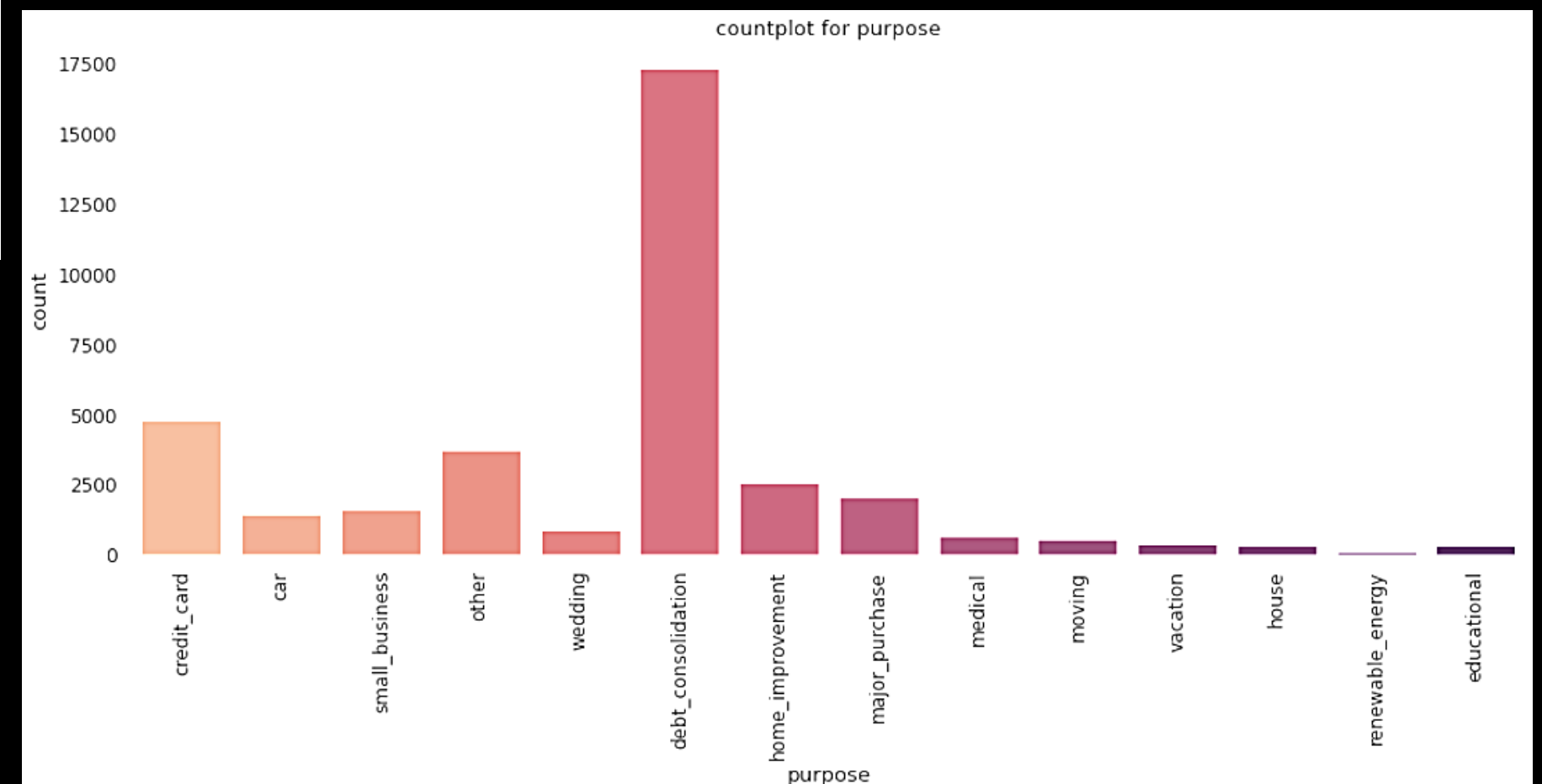
3. Segmented Analysis



The purpose of loan also plays a major driving factor, that will be analyzed below

- Small business has the most loans defaults, followed by renewable energy and education.
- Small business has the most loans defaults, followed by renewable energy and education but the count plot depicts that there is more samples of debt_consolidation.

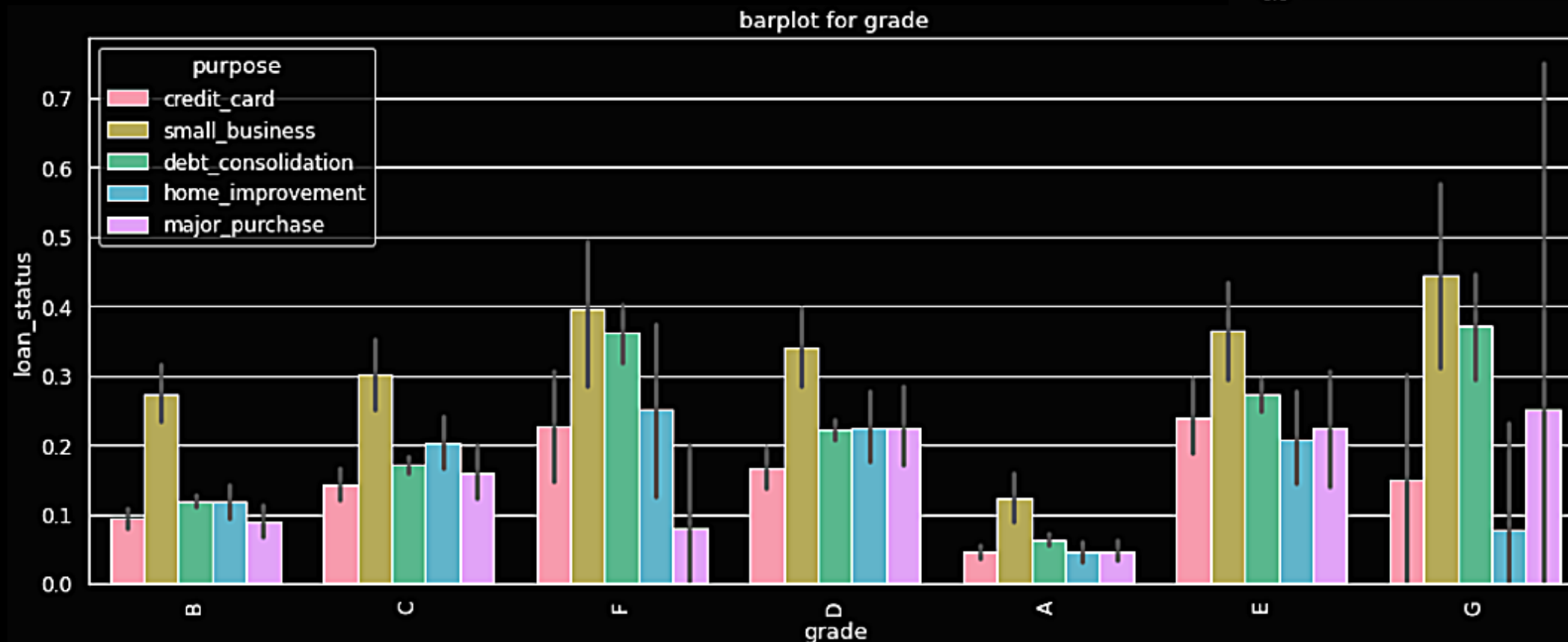
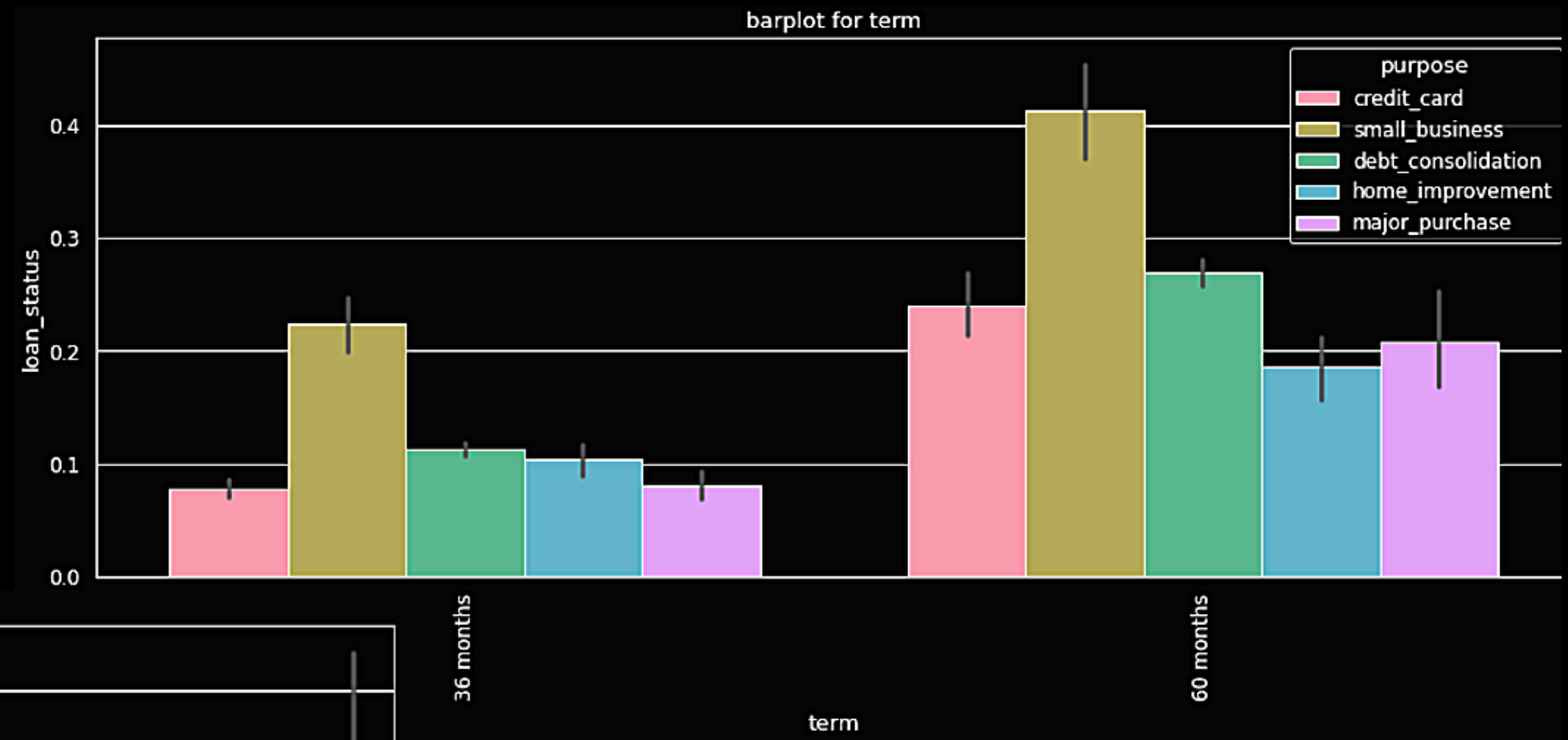
- The small business has more sample compared to education followed by renewable energy
- Now we analyze the major 5 purposes of loan according to the countplot above
- We are keeping the purpose of the loan as constant & analysing loan status against the term for top 5 purposes with the term



Segmented analysis - Term & Grade for 5 major purposes of Loan

Findings

- small business has the most loans defaults, followed by debt_consolidation for both terms.



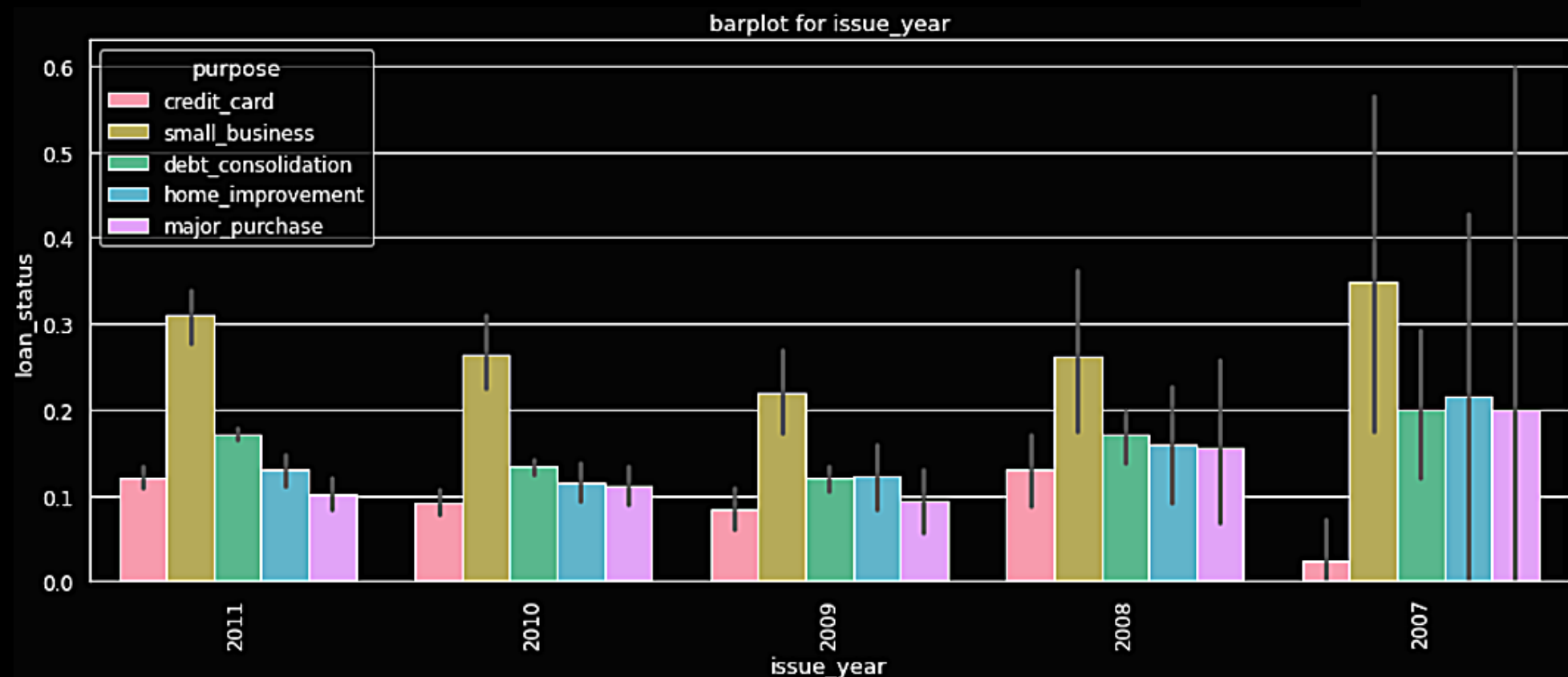
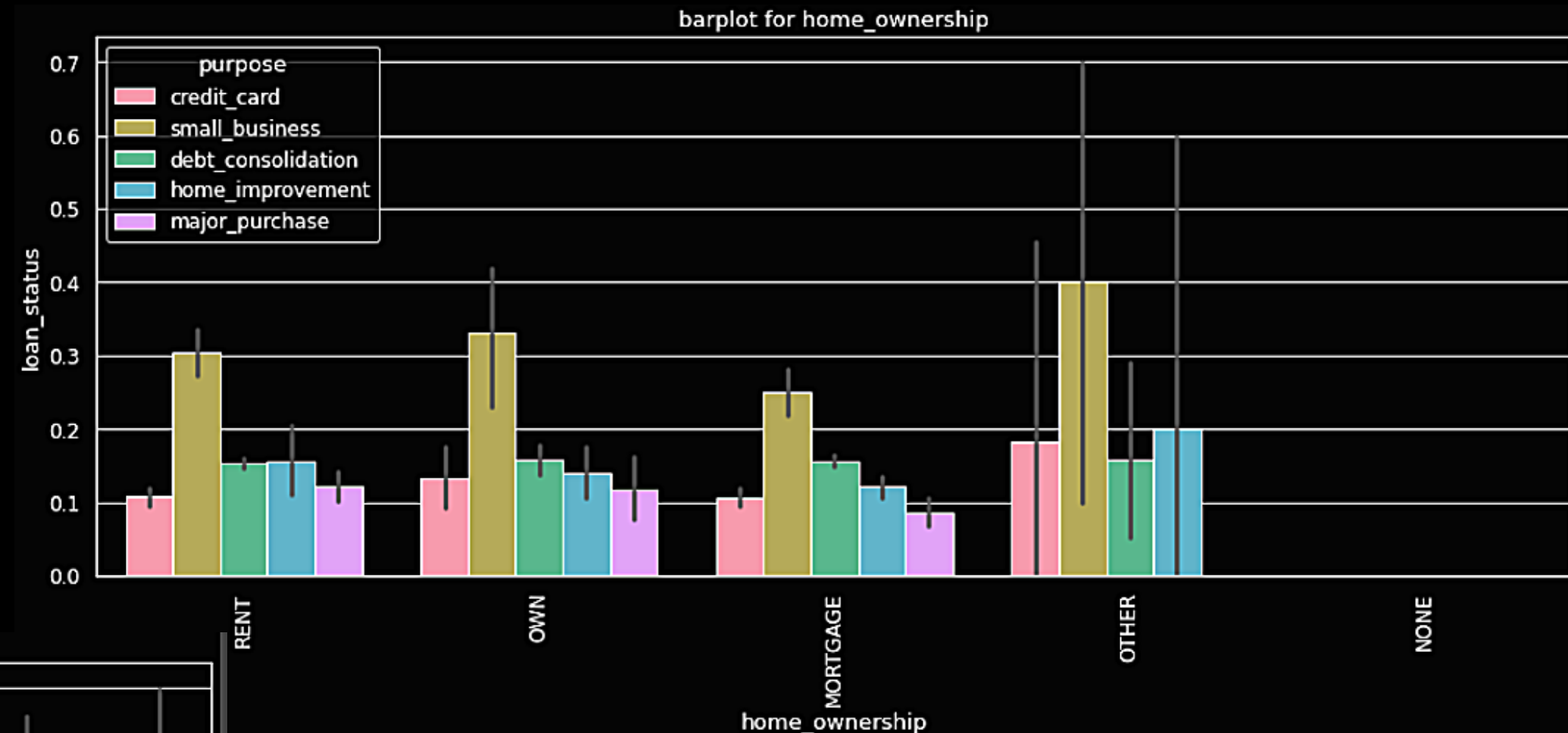
- In case of all Grades, the small business has the most loans defaults.

Segmented analysis - Home Ownership & Issue

Year for 5 major purposes of Loan

Findings

- In case of all Home ownership, the small business has the most loans defaults.

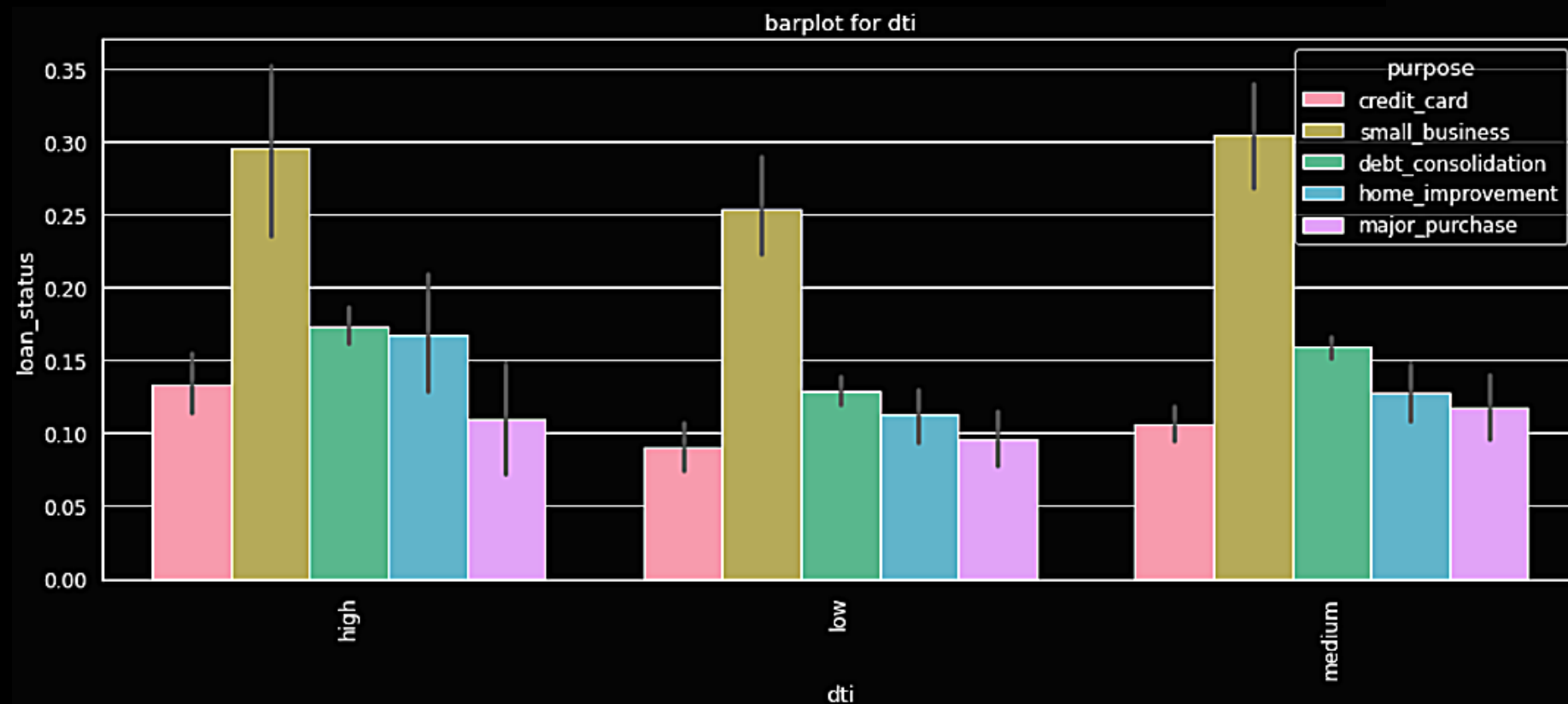
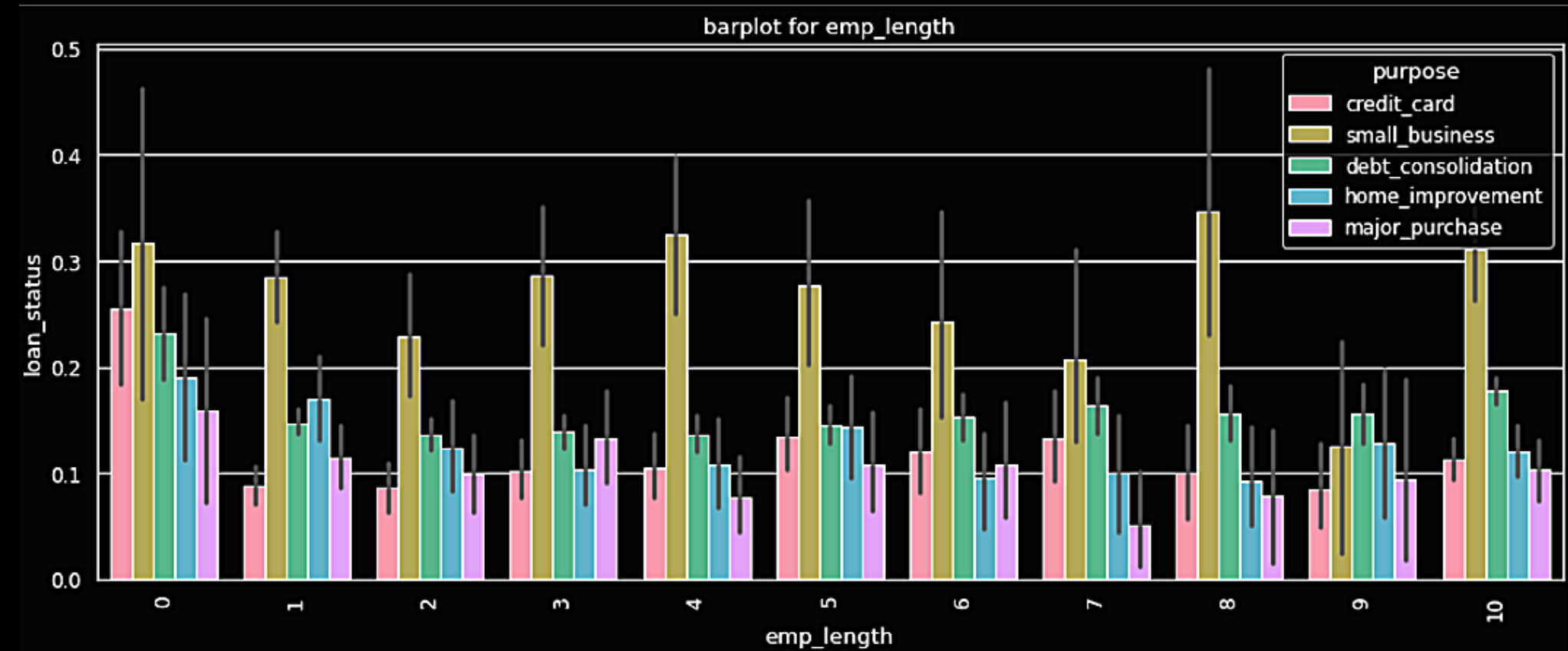


- For all issue year, the small business has the most loans defaults.

Segmented analysis - emp_length & for categorical variable - dti for 5 major purposes of Loan

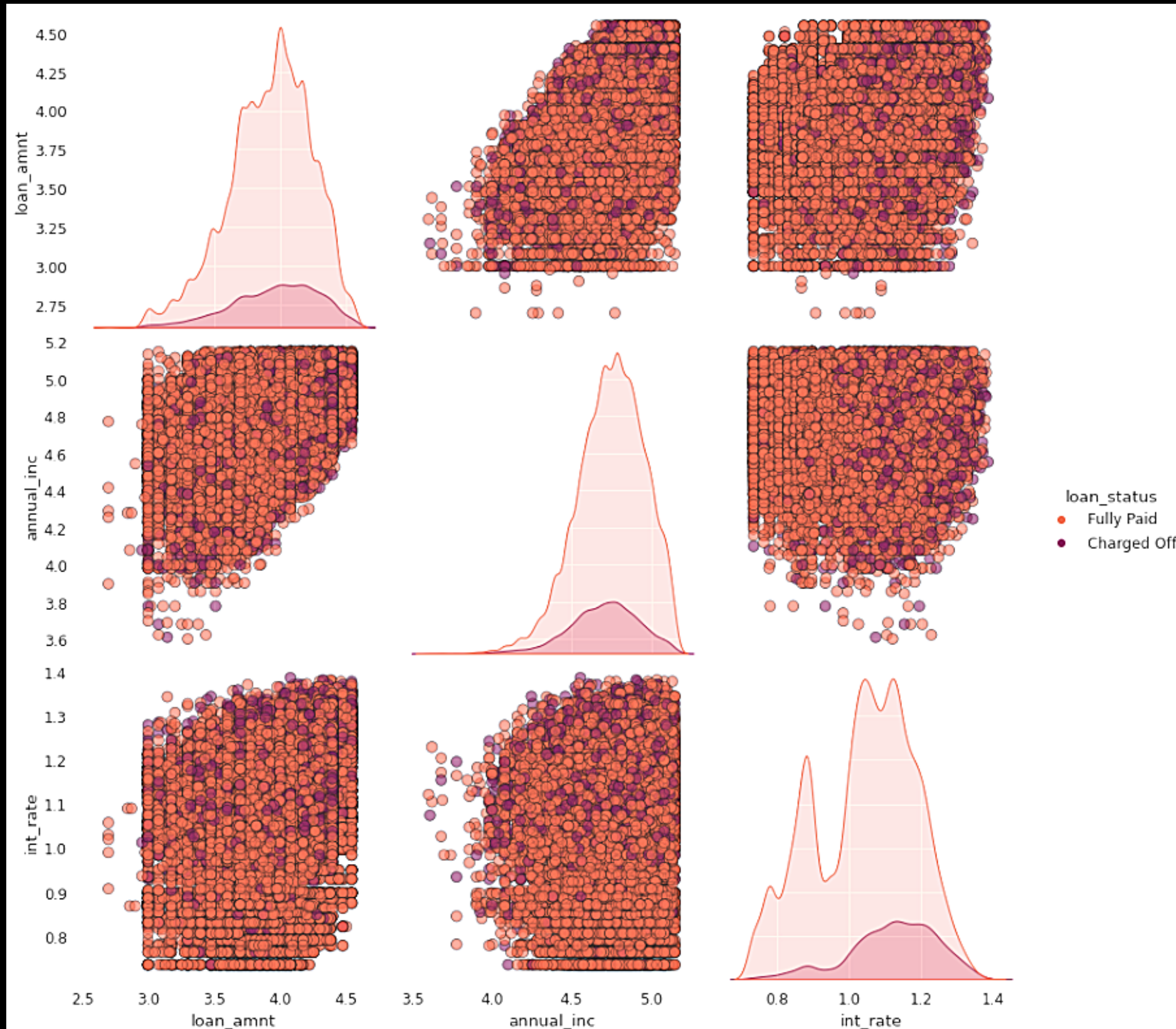
Findings

- For all emp_length, the small business has the most loans defaults.



- Binning has been done to use the categorical variable
- same trend across loan purposes

4. Multivariate Analysis



We created a pair plot colored by loan status with a density plot of the diagonal and format the scatter plots.

Findings

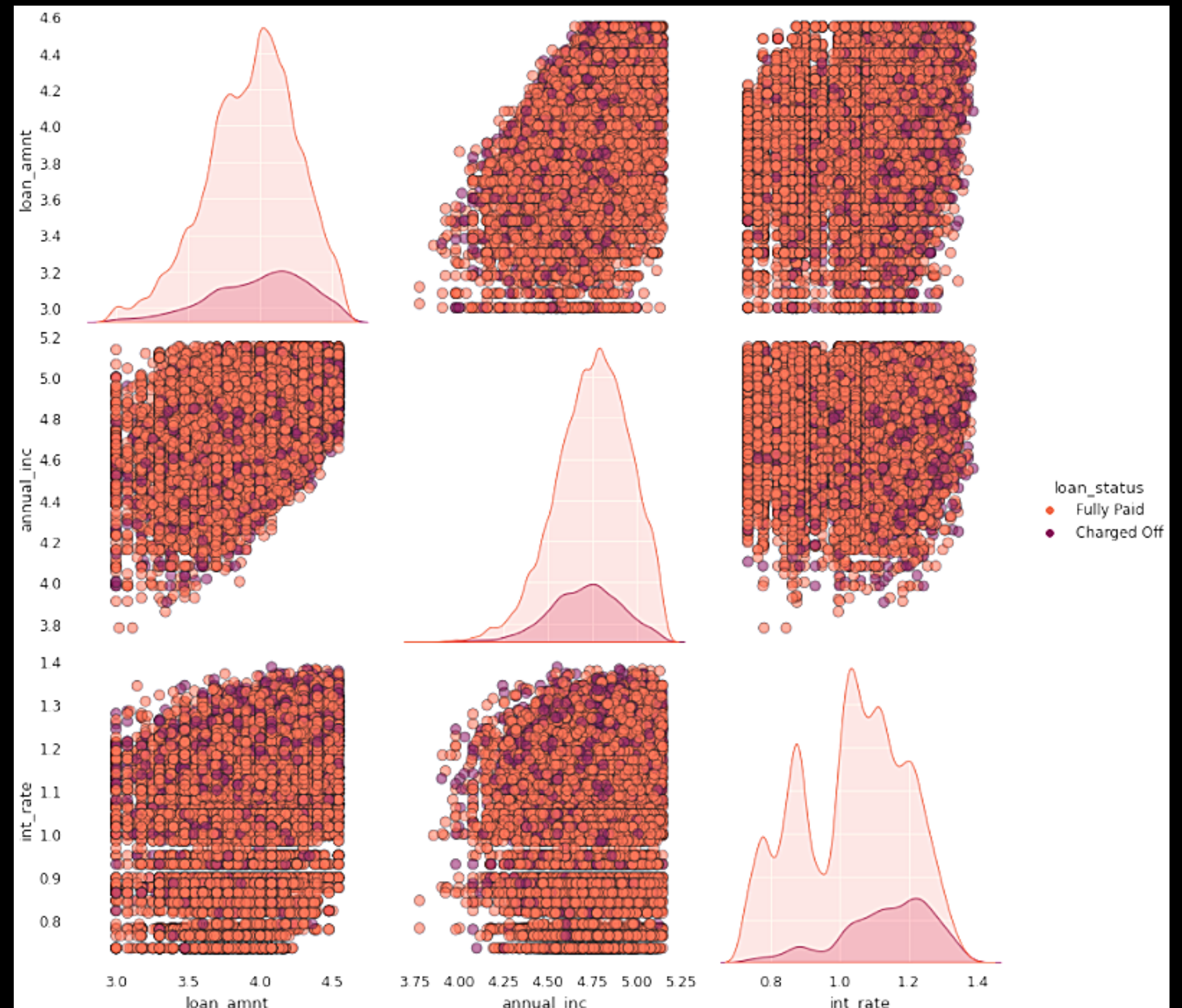
- Higher is the charged off ratio for higher interest rate.
- Slightly Higher the loan amount for a higher the annual income .
- Yearly, there is an increase in number of charged off.
- With increase in loan amount there is an increase in interest rate

4. Multivariate Analysis

A pair plot for multivariate analysis by choosing few important columns from dataset. Above plot is too dense so just pick year 2011 for getting some insights.

Findings

- Higher is the charged off ratio for higher interest rate.
- Higher the annual income higher the loan amount slightly
- Interest rate is increasing with loan amount & in increases is there in high charged off.





Conclusion

Minor Impact

- Higher loan amount (above 15.5K)
- Higher installment amount (above 325)
- Lower annual income (below 36.7K)
- Higher debt to income ratio (above 14.7%)
- Applicant's address state (NV, SD, AK, FL, etc.)
- Loan issue month (May, Sep, Dec)

Combined impact

- High interest rate & loan amount for lower income applicants
- Longer repayment term & High installment
- Home ownership (other) and loan purpose (car, moving or small business)
- Loan purpose & Residential state
- Income group & loan purpose

Heavy impact

- Higher revolving line utilization rate (above 57.5%)
- Higher interest rate (above 13%)
- Repayment term (5 years)
- Loan grade & sub-grade (D to G)
- Missing employment record
- Derogatory public records (1 or 2)
- Public bankruptcy records (1 or 2)
- Loan purpose (small business, renewable energy, educational)



Conclusion

- 1.If more loan amount is given Lending Club will be in profit because the interest rate increases with higher loan amount.
- 2.Grades should be analyzed thoroughly before lending money as A-grade is a top grade for lending.
- 3.A derogatory item can be considered as a negative entry by the lenders as it indicates public records & collections reflect financial obligations that were not paid as agreed. Inturn, risk and hurts your ability to qualify for credit or other services. Those who have pub_rec value 1 or 2 have higher chances of charged-off than who have no derogatory Public Record. pub_rec count with 3 or 4 cannot reach on any conclusions as there are less numbers.
- 4.States NV,CA and FL states & Dec, May, Sep are the Loan issue month shows high numbers of charged offs in good number of applications. So, the Lending Club should issue less loans in these states & during these issue months.
- 5.Less loans to these kind of applicants who are not working or have less than 1 year of work experience have high chances of getting charged off.
- 6.Lending Club should issue less loans to Small Business Owners as they defaults the most.



Conclusion

Loan Acceptance

1. For Weddings, Major Purchases, Car and Credit Card purposes.
2. Loan amount ranging between 5,000 USD and 10,000 USD.
3. Annual Income of applicants more than 1,00,000 USD.
4. Calculated interest rate for loan less than 7.5%.
5. Grade A&B.
6. Loan applicant is owning a house.
7. Loans for 36 months term.

Loan Risk Factors

1. Loans for purpose of Small Business.
2. Loans for loan amount 30,000 USD and above.
3. Loan applicants annual income whose annual income is less than 25000 USD.
4. Loans having calculated interest rate more than 15%.
5. Loans in grade E, F&G.
6. Loan applicant is either living on rent or mortgage.
7. Loans for 60 months term.