# Assignment : Lending Club Case Study

Team : Vivek Vikash and Vijay Mallepudi

Date : 26 Jun 2024

# Business Understanding

- We are working for Lending club a finance company which specializes in lending various types of loans to urban customers.

- When the 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

# Objective

The company wants to understand the driving factors (or driver variables) behind loan default (loan_status = 'Charged Off'), i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

# Problem Approach

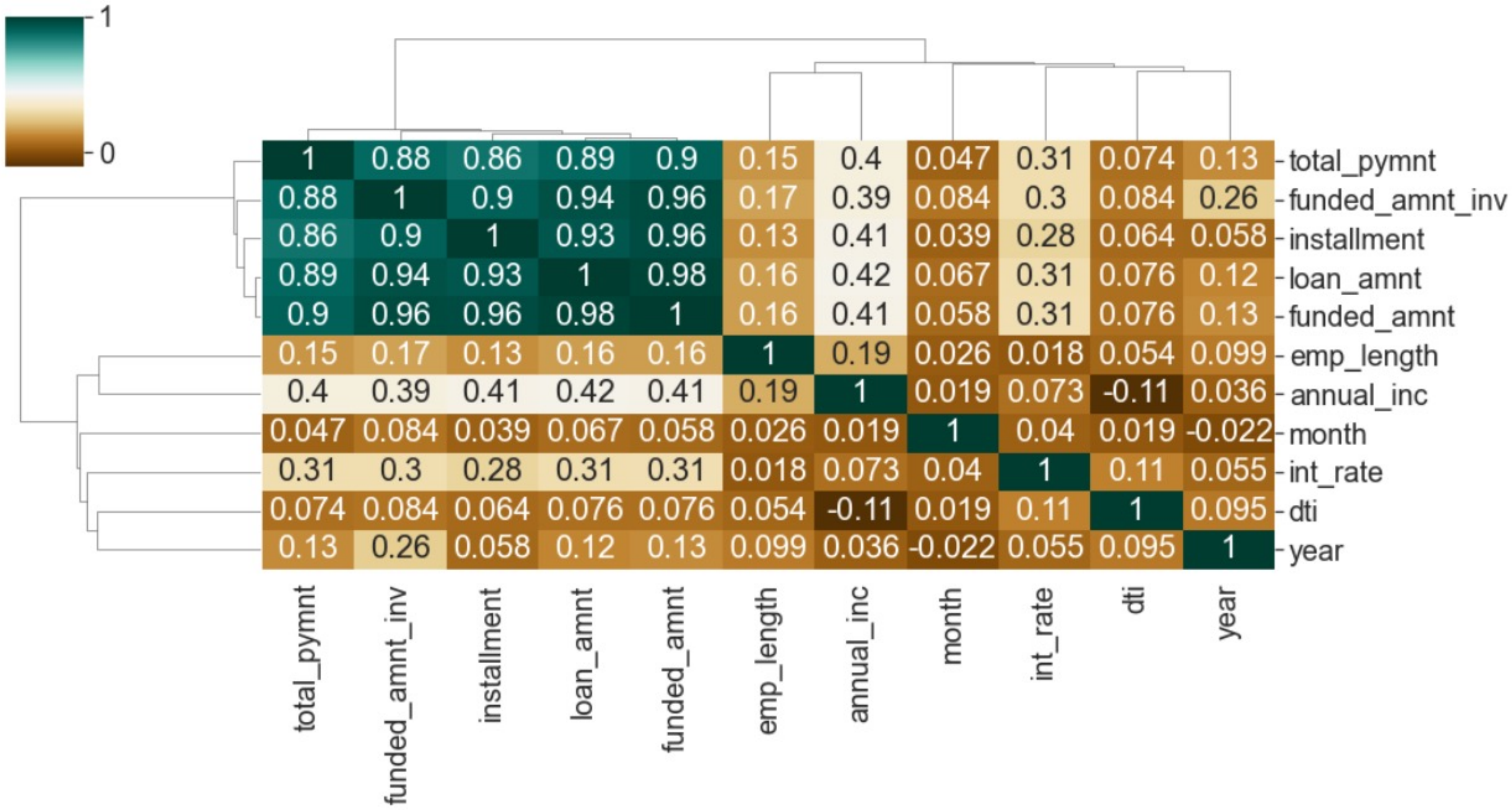
- Step 1: Data Cleansing
- Step 2: Univariate Analysis
- Step 3: Segmented Univariate Analysis
- Step 4 : Bivariate Analysis
- Step 5 : Conclusion

# Data Cleansing Process

- The data set is checked for any duplicate rows, No duplicates were present in the loan data set
- The data set is checked for any rows or columns having all missing values. It was observed that there were 54 columns present in the data set which were having all null values. Hence, all 54 columns were excluded from the data set.
- The EMP_LENGTH column has been refined to hold only the numeric part in order to use it in our analysis
- The missing values in the EMP_LENGTH column has been imputed with 0 value
- The ‘%’ symbol has been trimmed off from the  INT_RATE column
- Below columns have been derived for doing the Bivariate analysis:

  - **Year**
  - **Month**
  - **loan_amnt_cats**
  - **annual_inc_cats**
  - **int_rate_cats**

# Correlation Matrix




**<u>Observations:</u>**

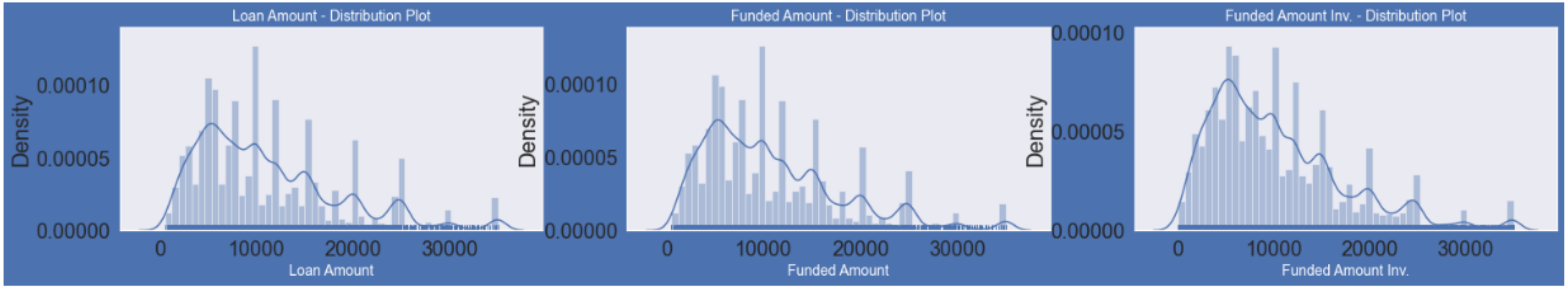
Loan amount, investor amount and funded amount are strongly correlated among each other.

Annual income with DTI (Debt-to-income ratio) is negatively correlated, that means whenever annual income is low DTI is high & vice versa.

There is a positive correlation between annual income and employment years which means income increases with work experience

# Univariate Analysis

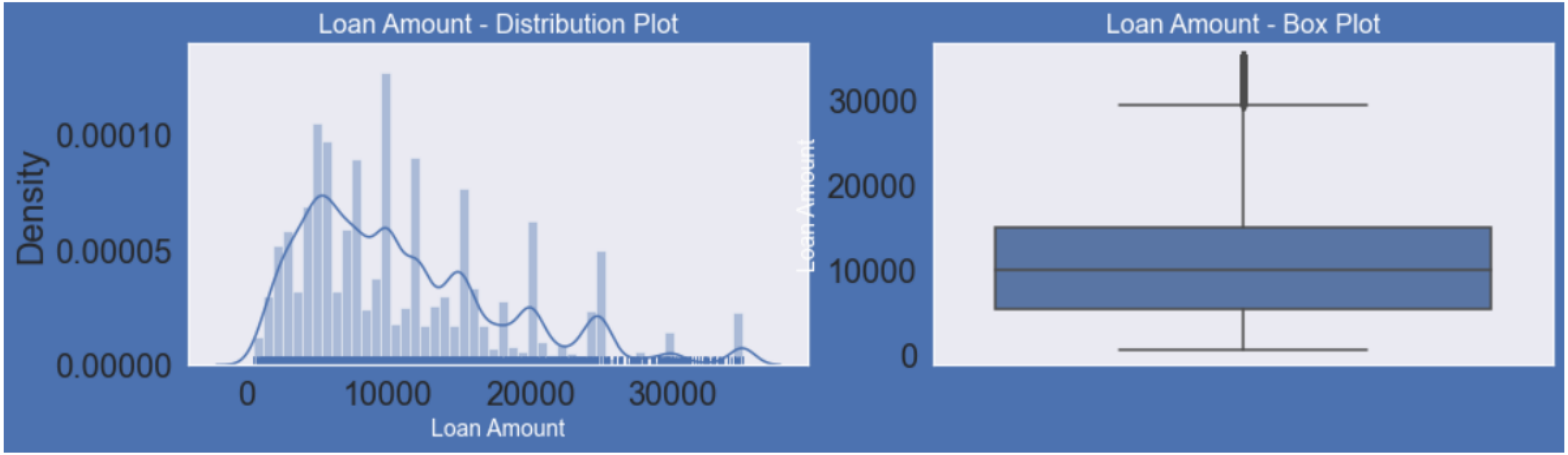
# Distribution Plot




**Observation:**

- Majority of the data lies with income range between 5000 to 15000
- The pattern seems to be similar for Loan amount, Funded Amount and Invested amount. Hence, we would be leveraging the Loan amount variable for further analysis

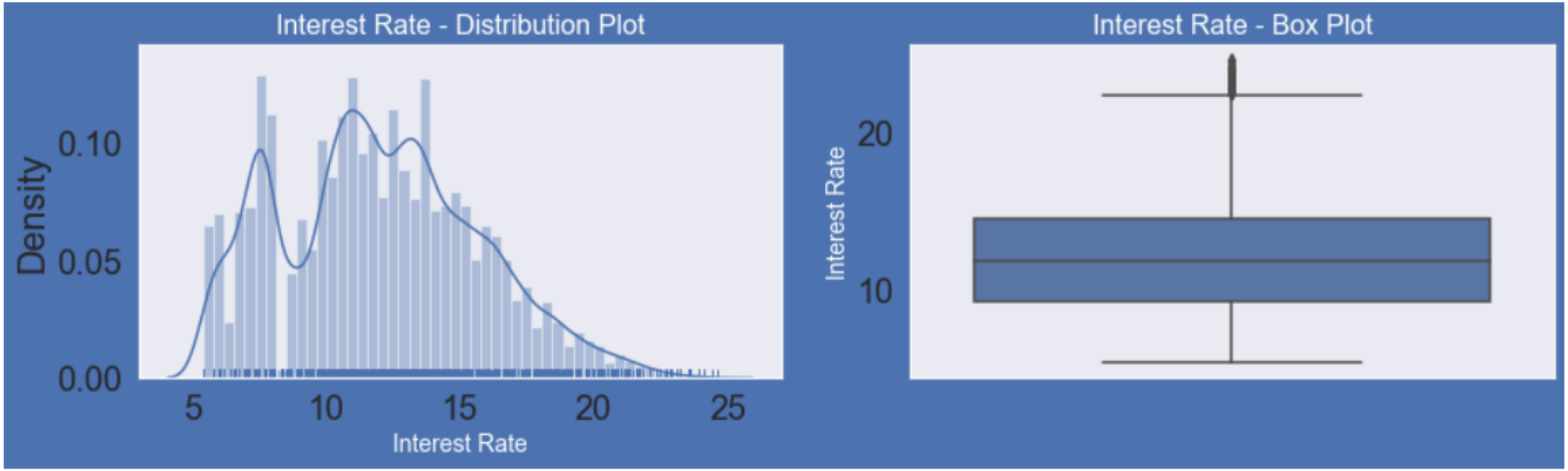
# Univariate Analysis on Loan amount-Quantitative Variables




**Observation:**

- Above plots show that most of the Loan amounts are in range of 5000 - 15000

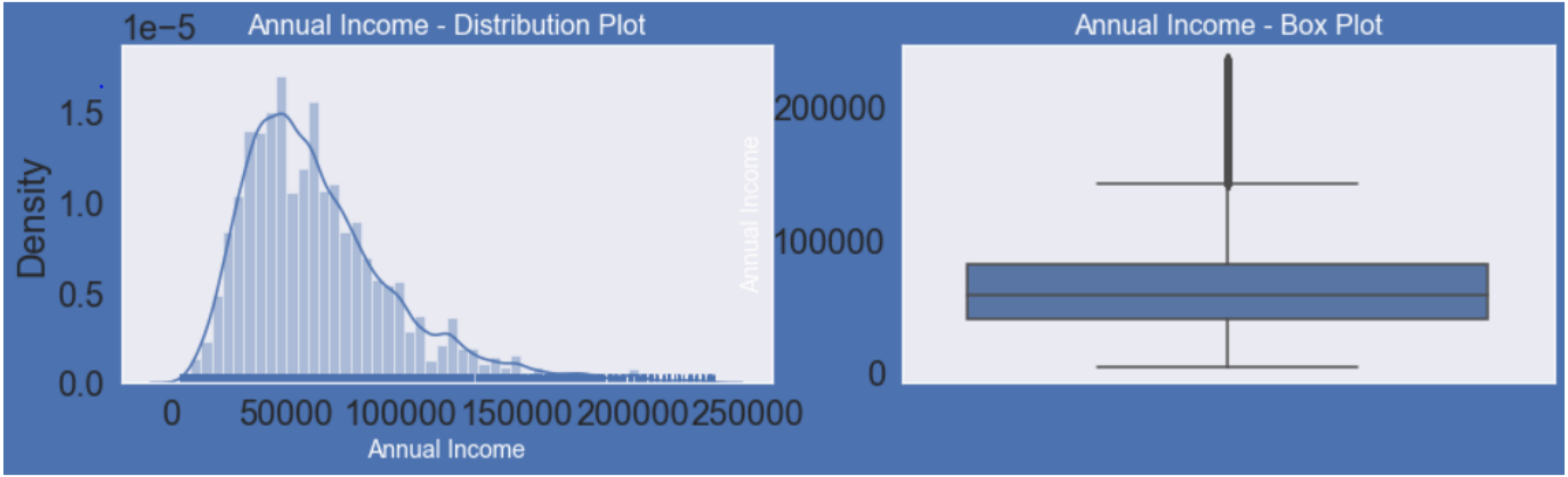
# Univariate Analysis on Interest Rate - Quantitative Variables




**Observation:**

- Below plots show that most of the Interest Rates on loans are in range of 10% - 15%

# Univariate Analysis on Annual Income - Quantitative Variables

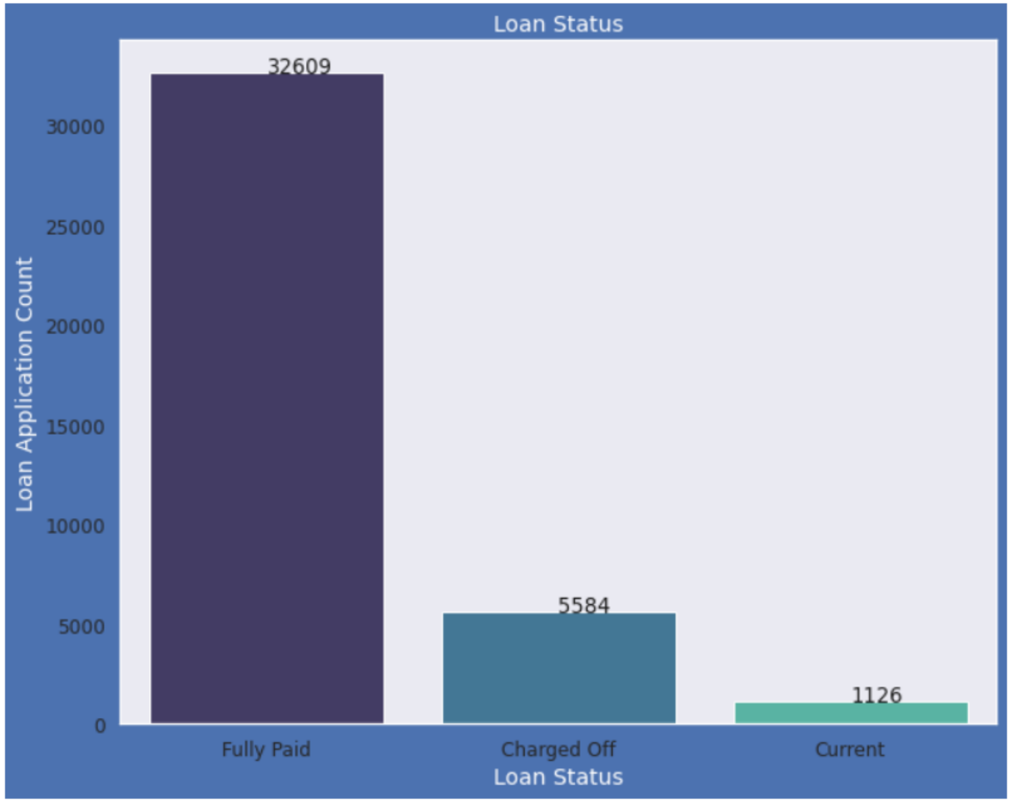



**Observation:**

- Below plots show that most of the borrower's Annual incomes are in range of 40k to 80k

# Segmented Univariate Analysis

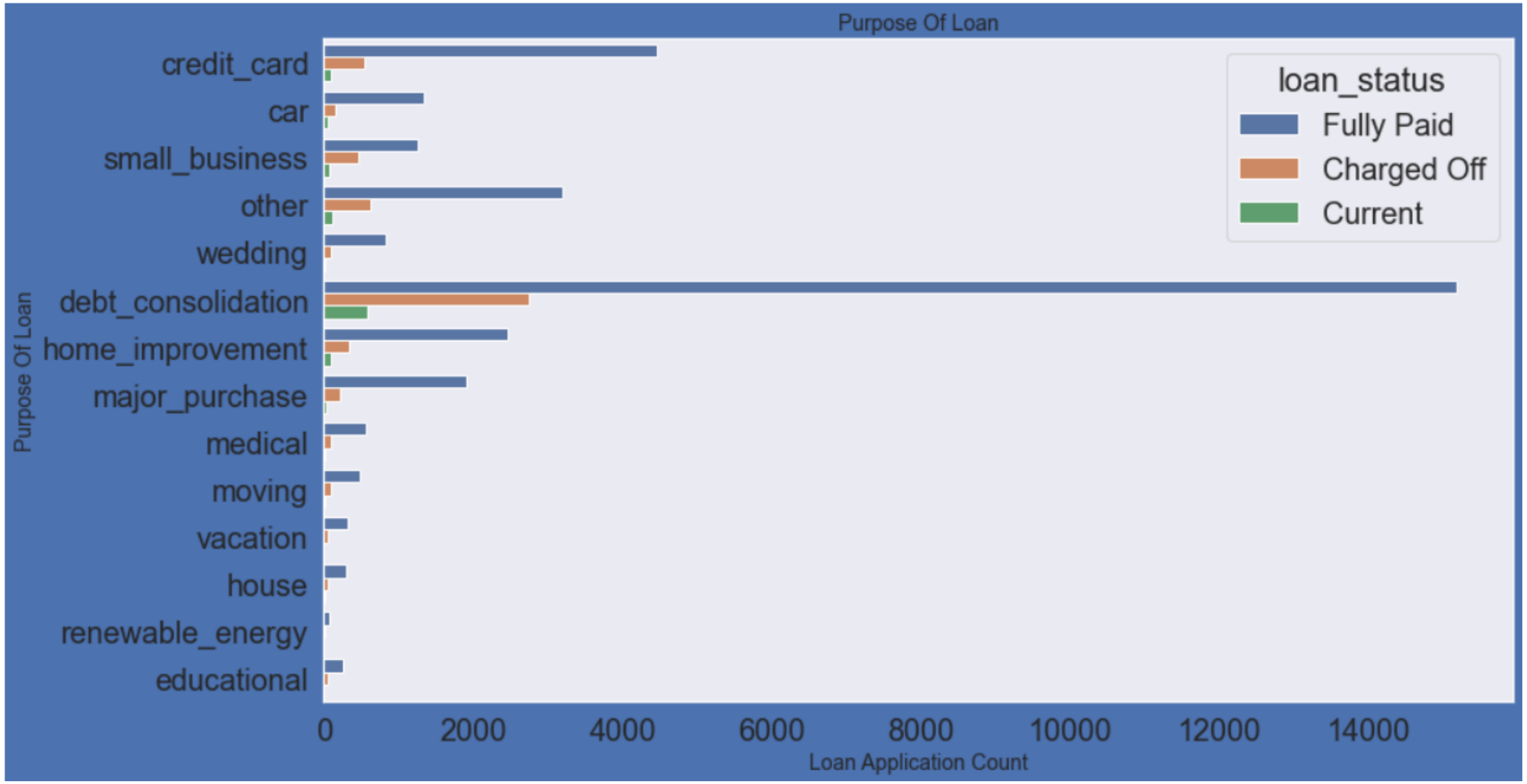
# Segmented Univariate Analysis – Loan Status




**<u>Observation:</u>**

- Close to 14% of loans were charged off out of the total loans issued

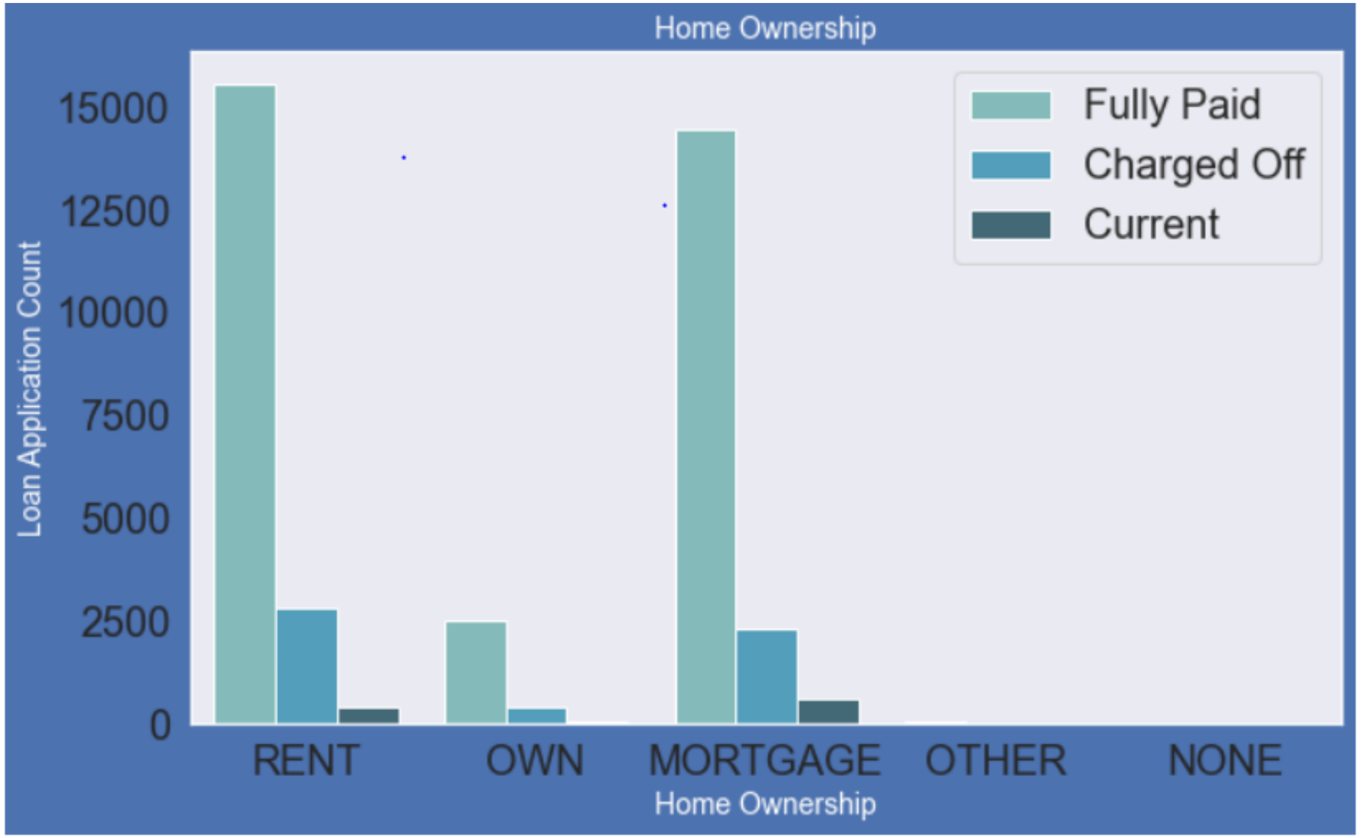
# Segmented Univariate Analysis – Loan Purpose




**Observations:**

- Most of the loans were taken for the purpose of debt consolidation & paying credit card bill
- Comparatively number of charged off count are also high for the loans taken for these 2 purpose

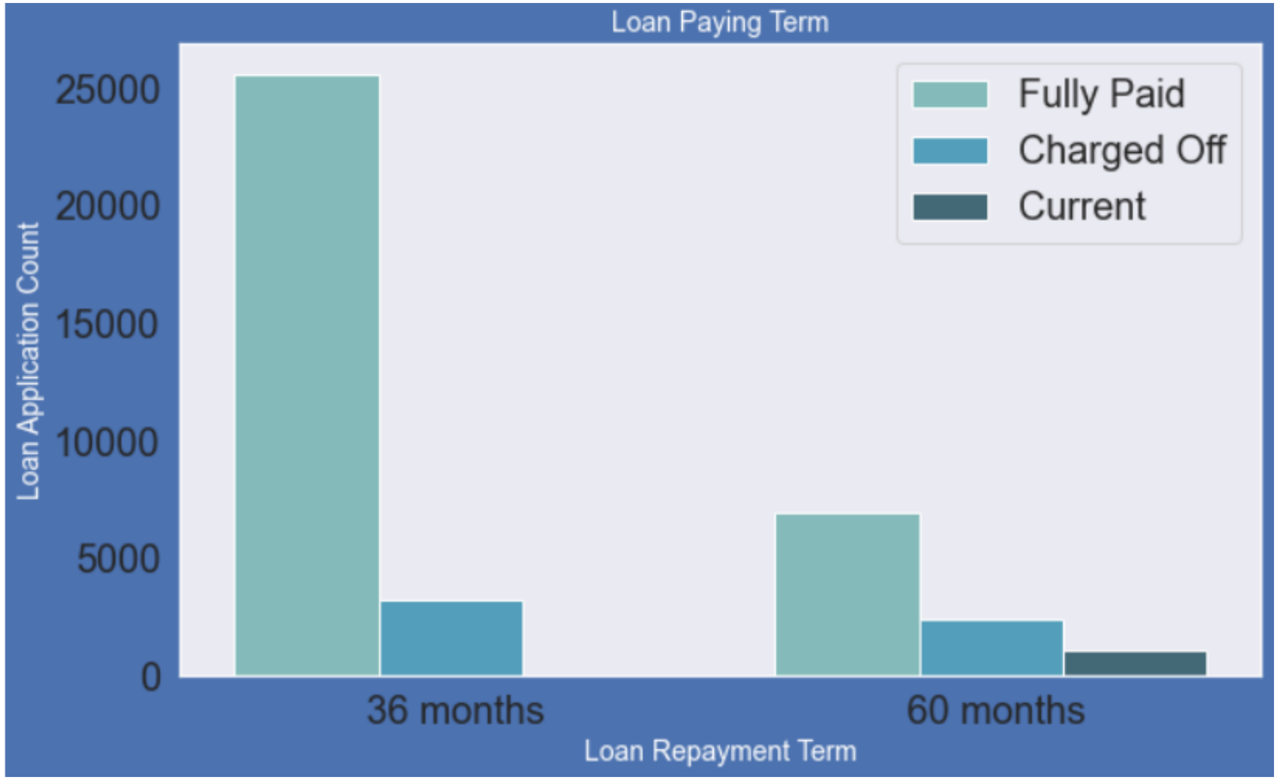
# Segmented Univariate Analysis – Home ownership




**<u>Observations:</u>**

- Most of people took loan who live in rented house or have mortgaged their property
- Charged off numbers are also high for the above 2 categories

# Segmented Univariate Analysis – Loan Repayment Term

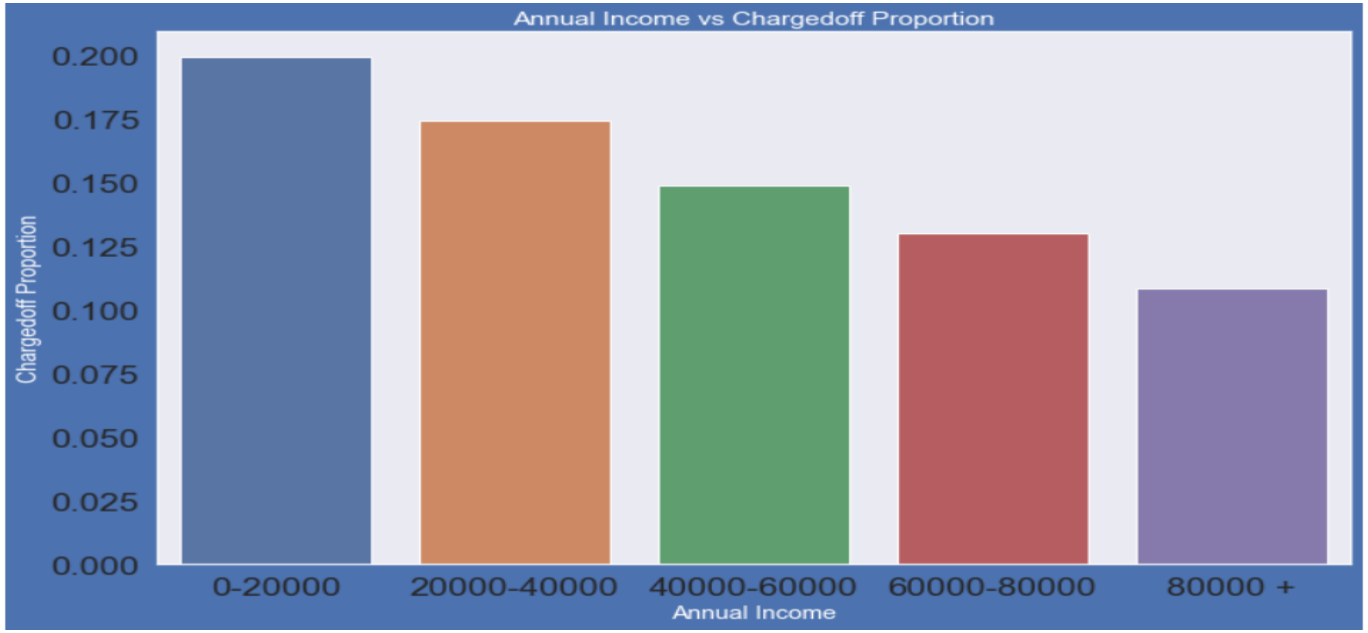



**<u>Observations:</u>**

- 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

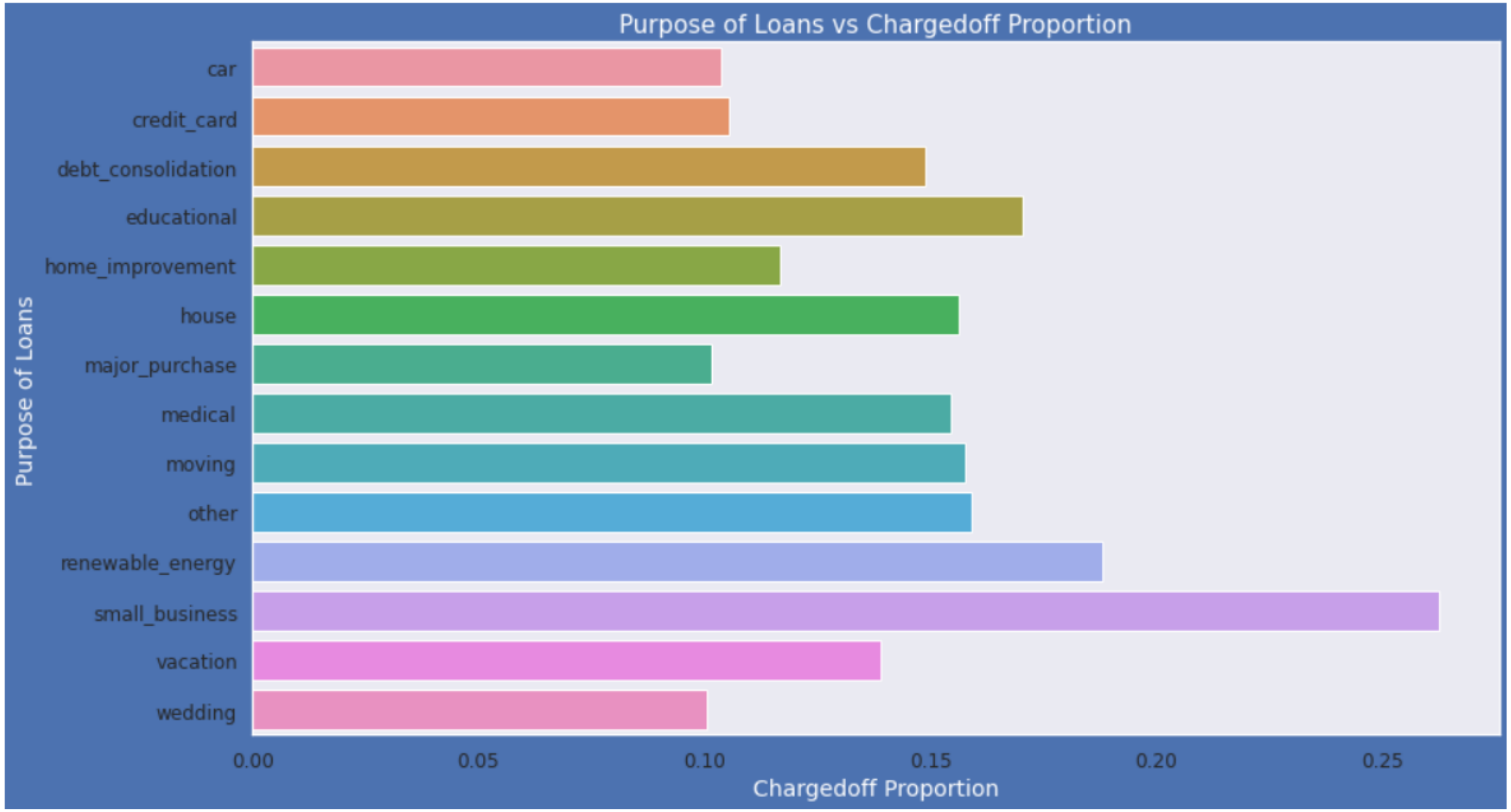
# Annual Income Vs. Charged Off Proportion




**Observations:**

- Income range of 0-20000 has high chances of charged off
- Income range of above 80000 has least chances of charged off
- Income range is inversely proportional to the Charged off

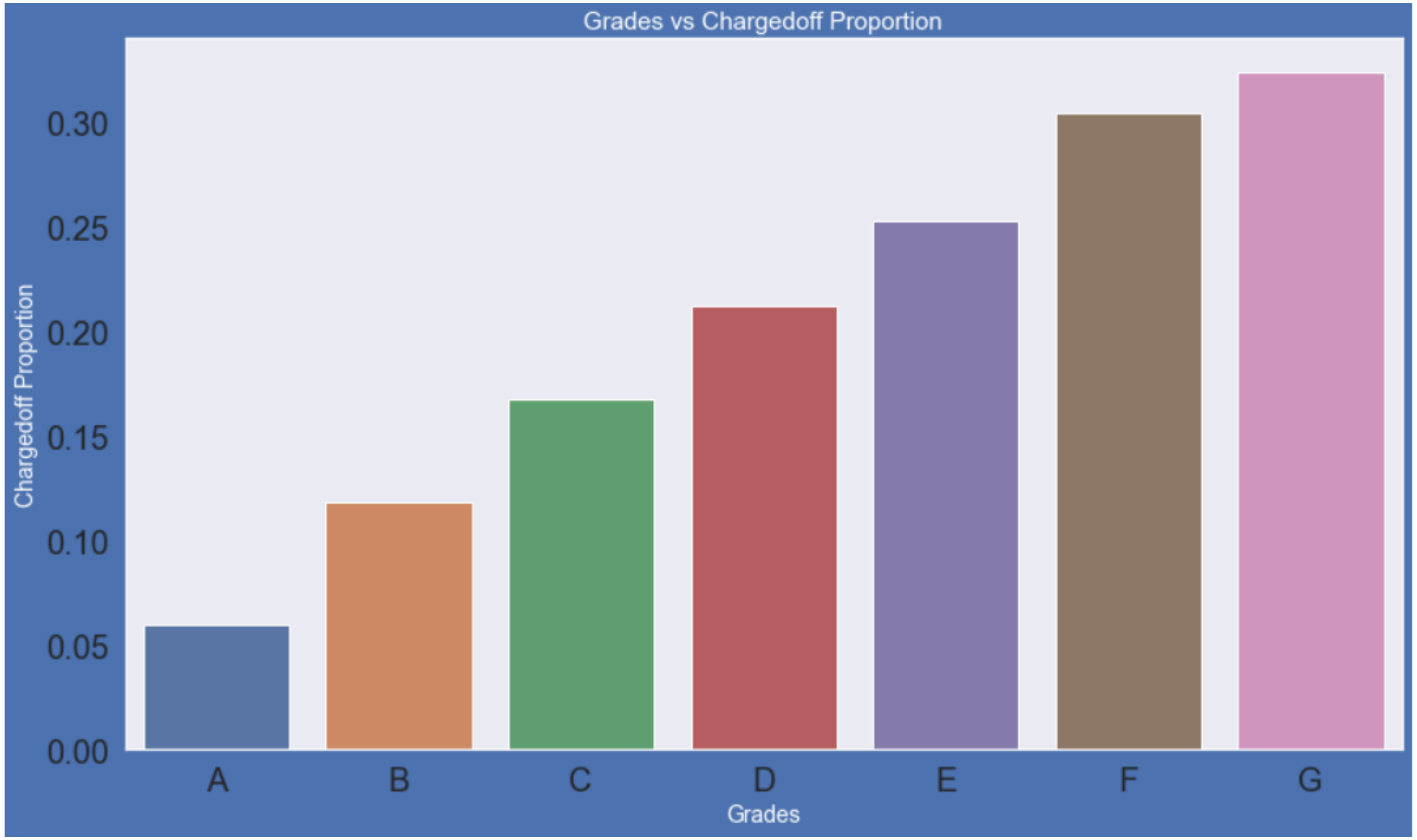
# Purpose of Loan vs. Charged Off Proportion




**Observations:**

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

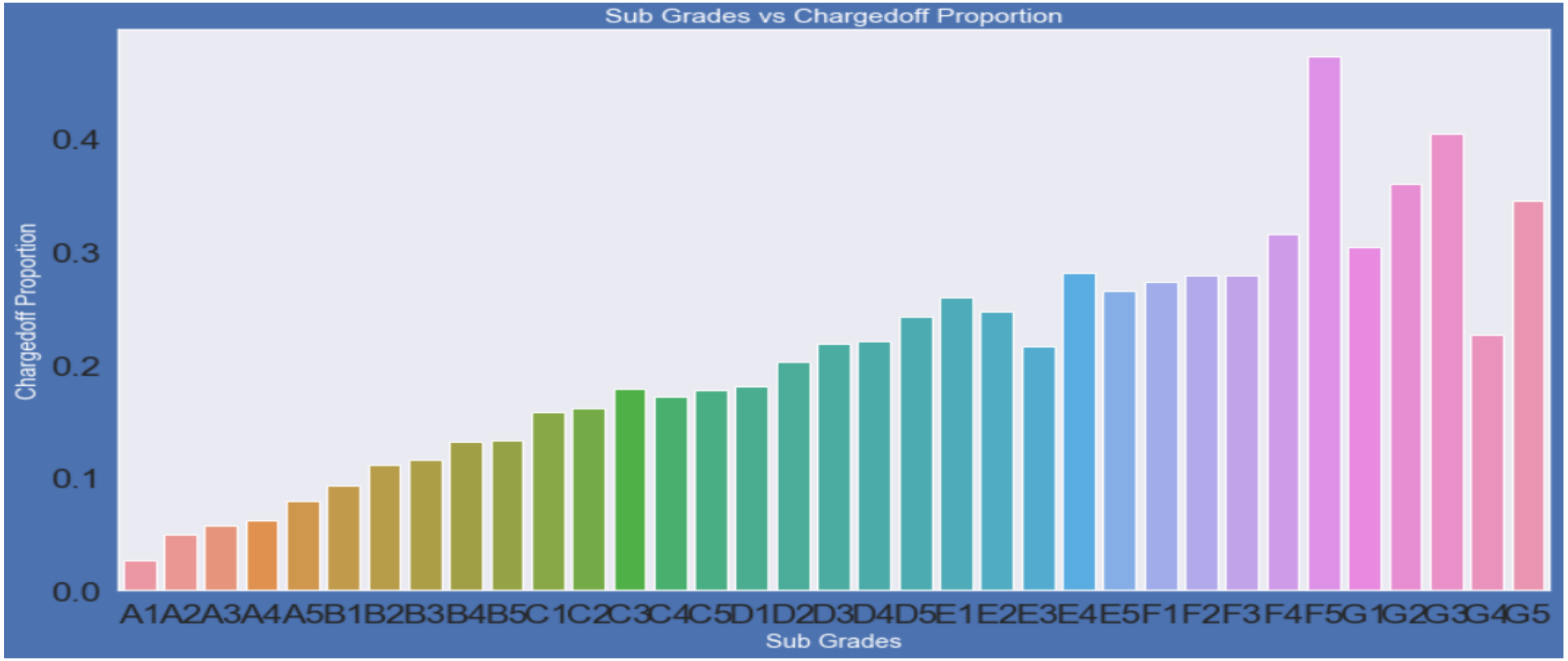
# Grade Vs. Charged Off Proportion




**Observations:**

- Grade "A" has very less chances of charged off
- Grade "F" and "G" have very high chances of charged off

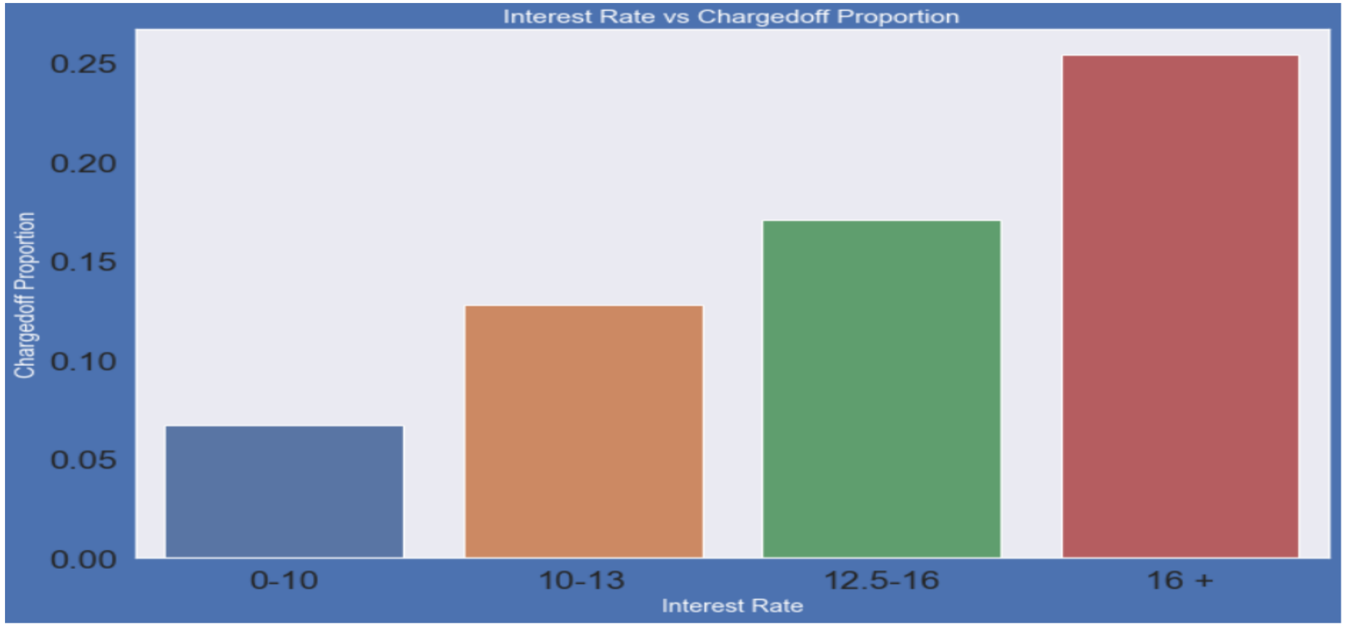
# Sub Grades Vs. Charged Off Proportion




**Observations:**

- Sub Grades of "A" has very less chances of charged off
- Sub Grades of "F" and "G" have very high chances of charged off

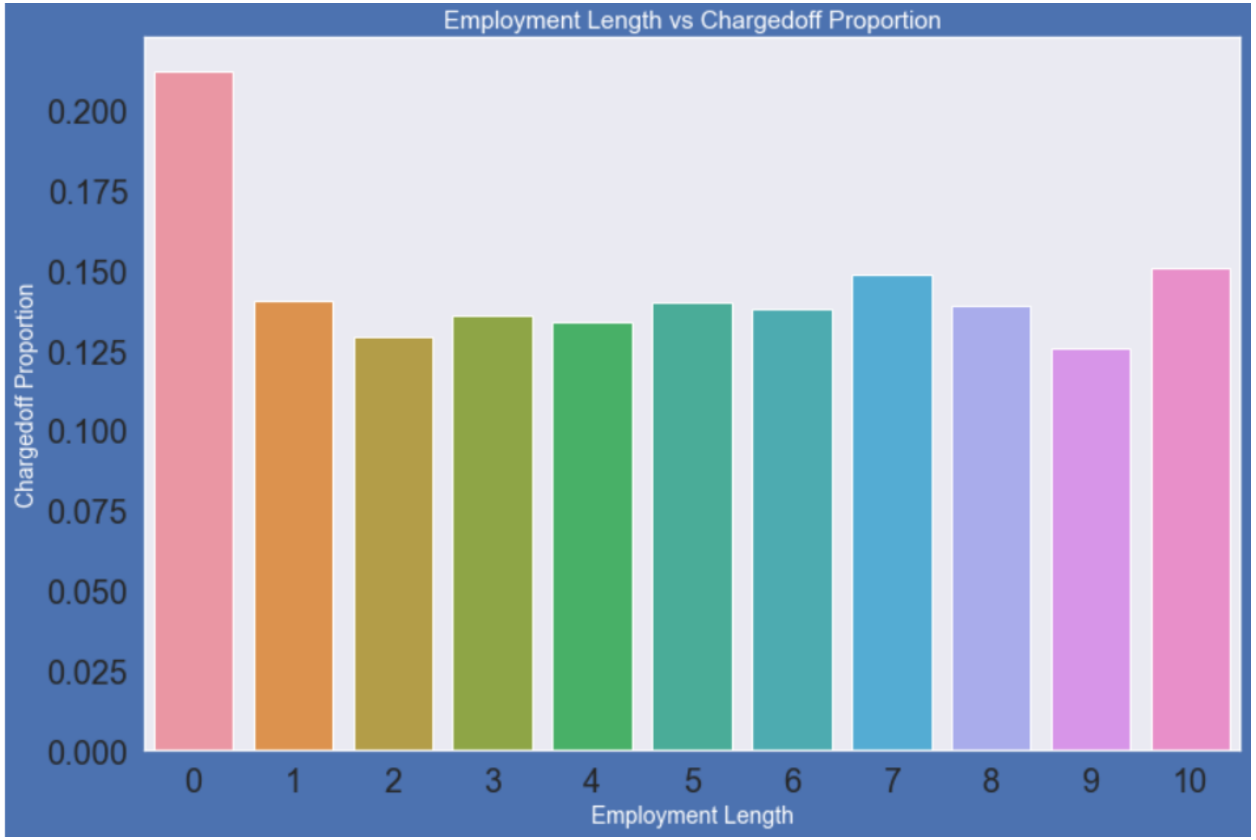
# Interest Rate Vs. Charged Off Proportion




**Observation:**

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

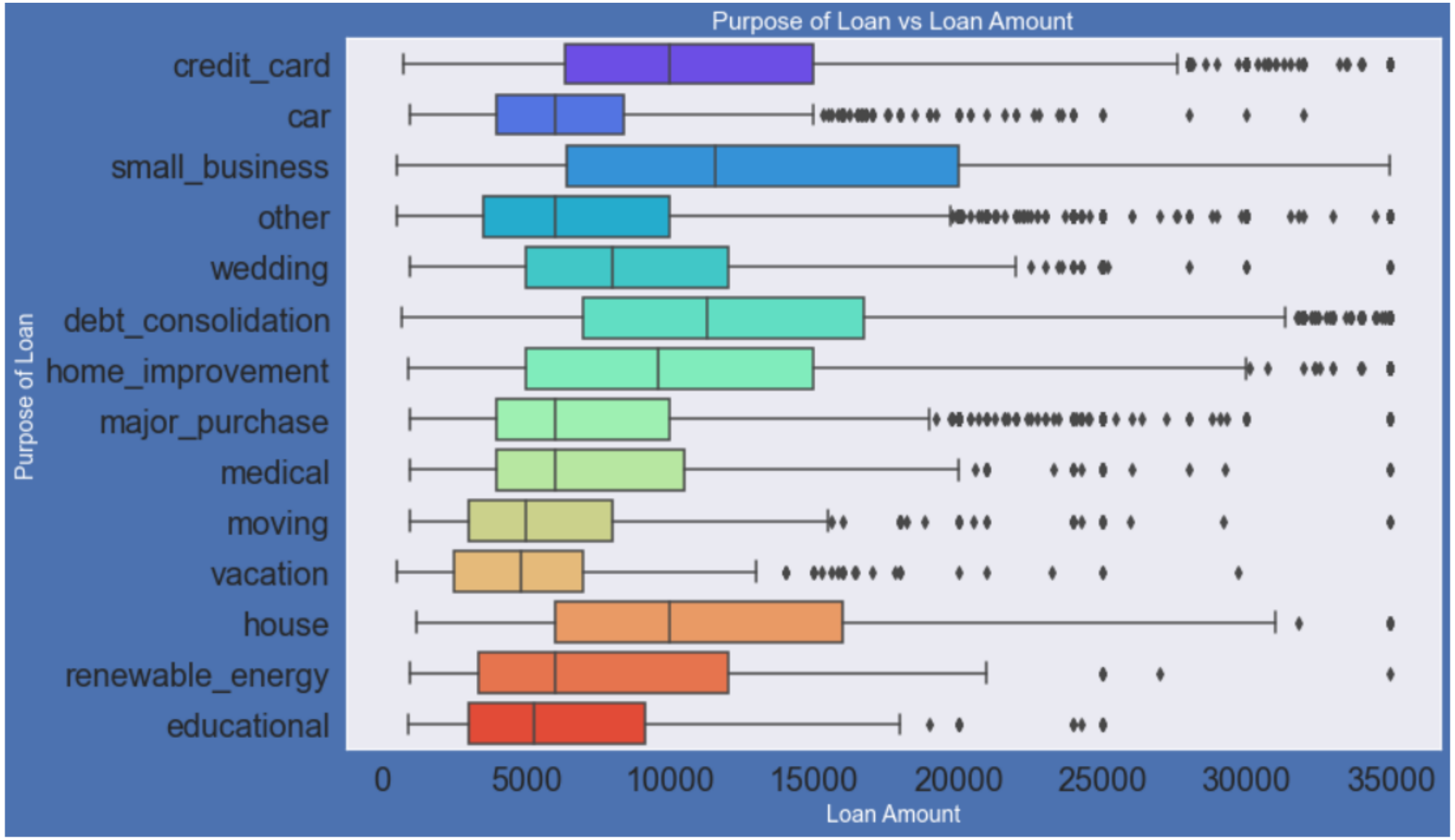
# Employee Length Vs. Charged Off Proportion




**Observation:**

- Those who are not working or have less than 1 year of work experience have high chances of getting charged off.

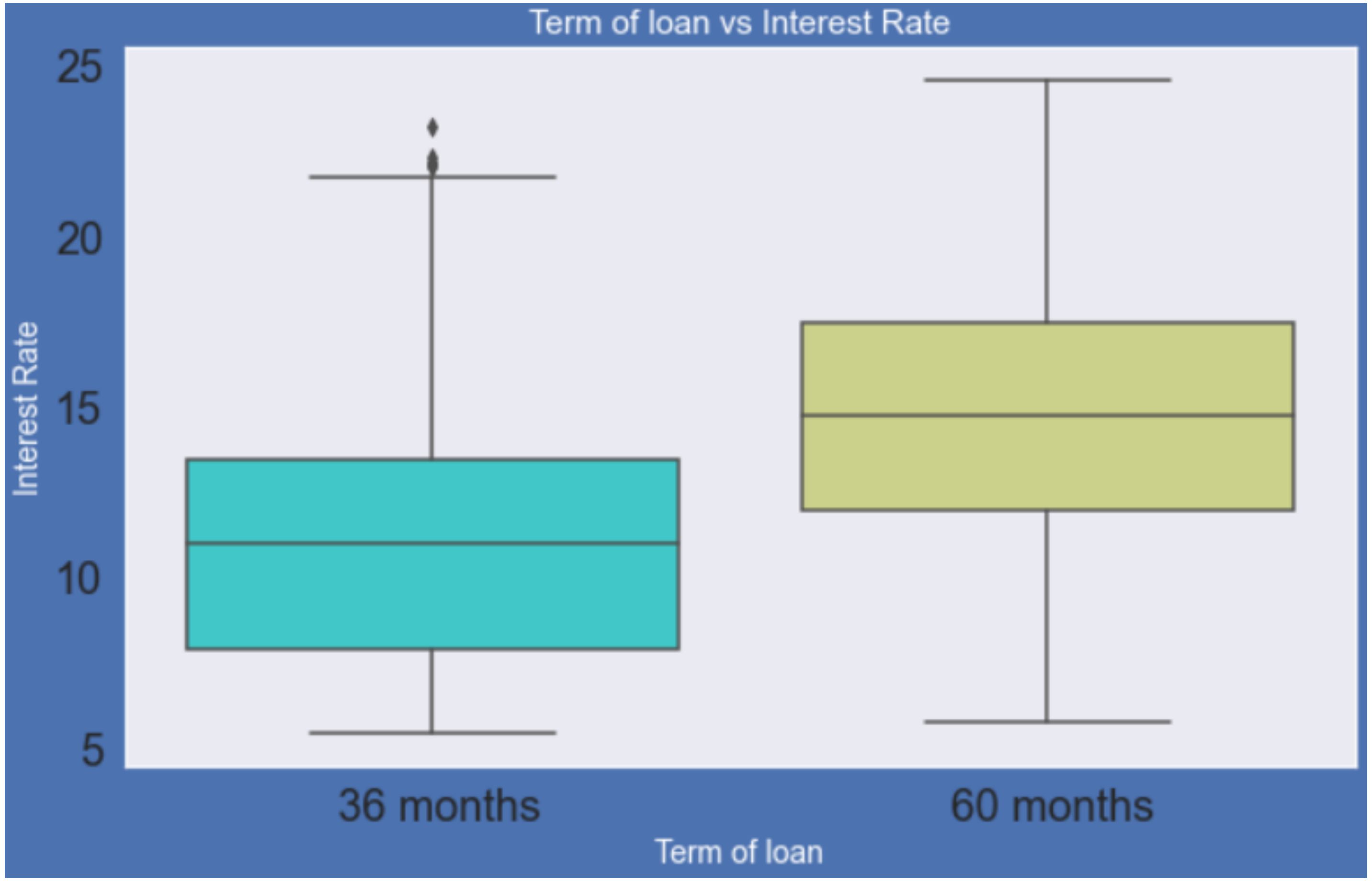
# Purpose of Loan Vs. Loan Amount




**Observation:**

- 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

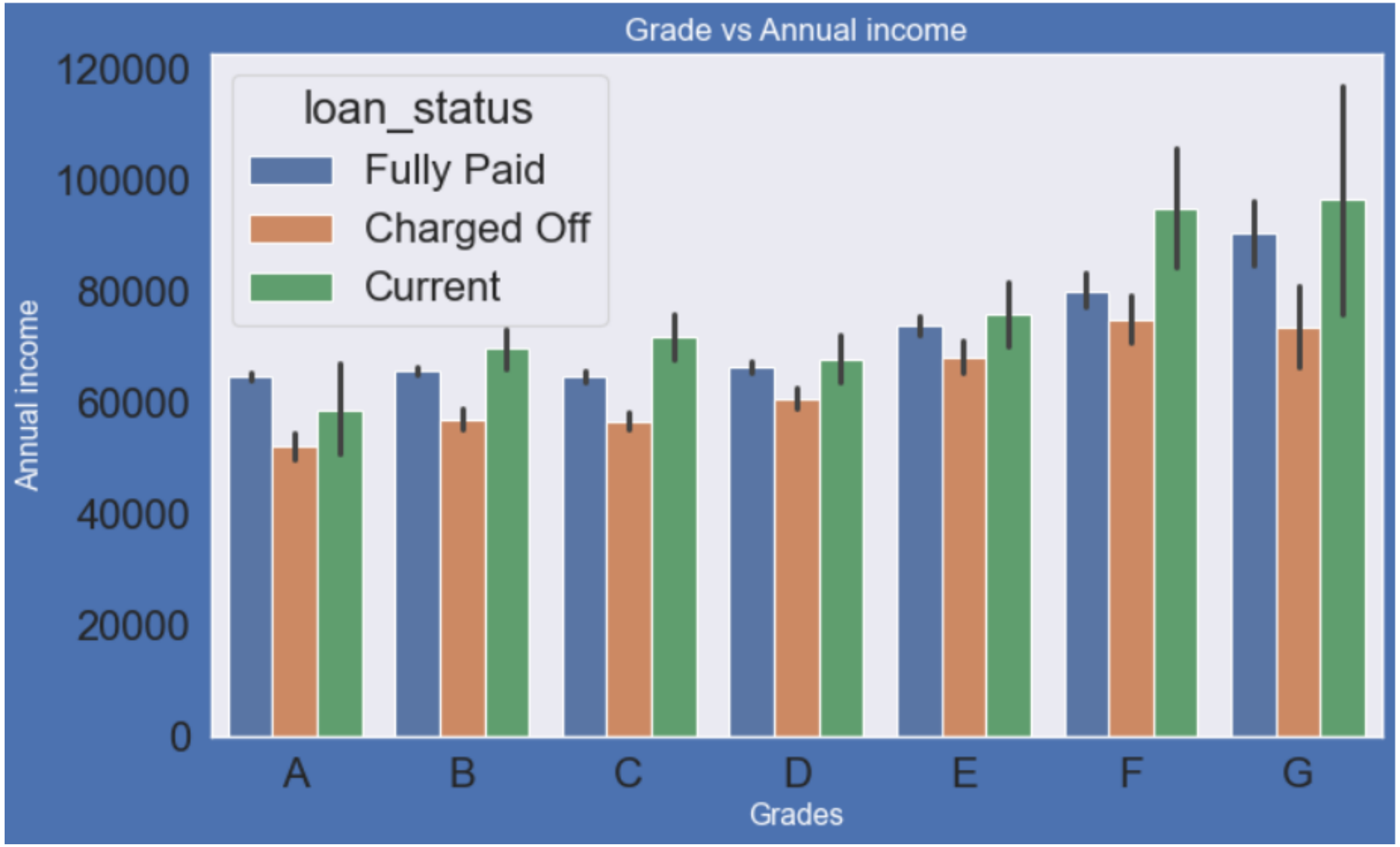
# Term Of Loan Vs. Interest Rate




**Observation:**

- 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

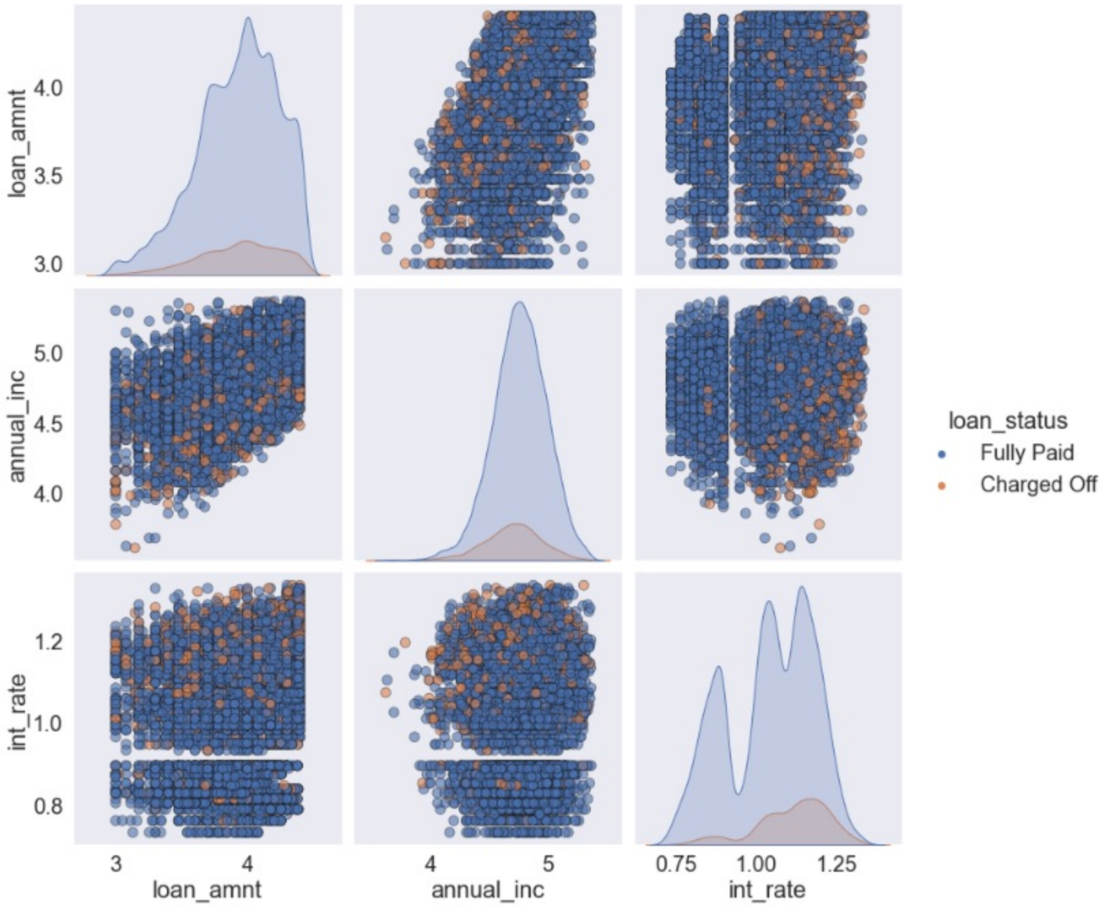
# Grade Vs. Annual Income




**<u>Observations:</u>**

- 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)

# Pair Plot




**Observation:**

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

# Conclusion

- Loans that were charged off mostly taken for the purpose of debt consolidation & paying credit card bill
- Customers whose loans were charged off mostly were staying in Rented home or their properties were mortgaged
- Loans which had a repayment schedule of 60 months were contributing more to the charged off numbers
- Loans which were taken for the purpose of debt consolidation and paying credit card bill had most off the charged off numbers as well
- Applicants with income range between 0-20000 have high chances of charged off
- Small business applicants have high chances of getting charged off
- Grades and sub grades F and G applicants have high chances of charged off
- Interest rate of more than 16% have high chances of charged off, charged off rate increases with higher interest rates
- Those who are not working or have less than 1 year of work experience have high chances of getting charged off
- The ones getting charged off have lower annual incomes than the ones who have paid fully for each and every grade