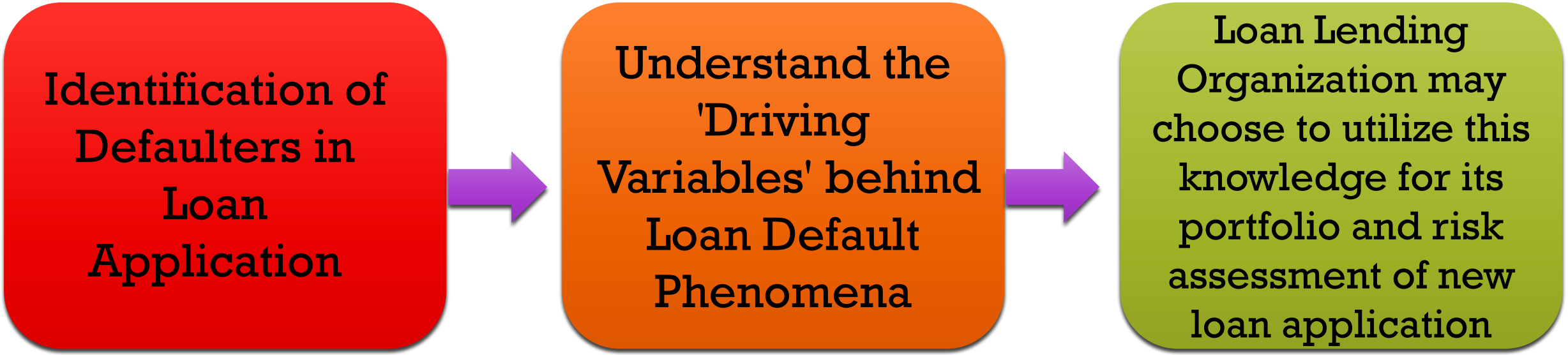
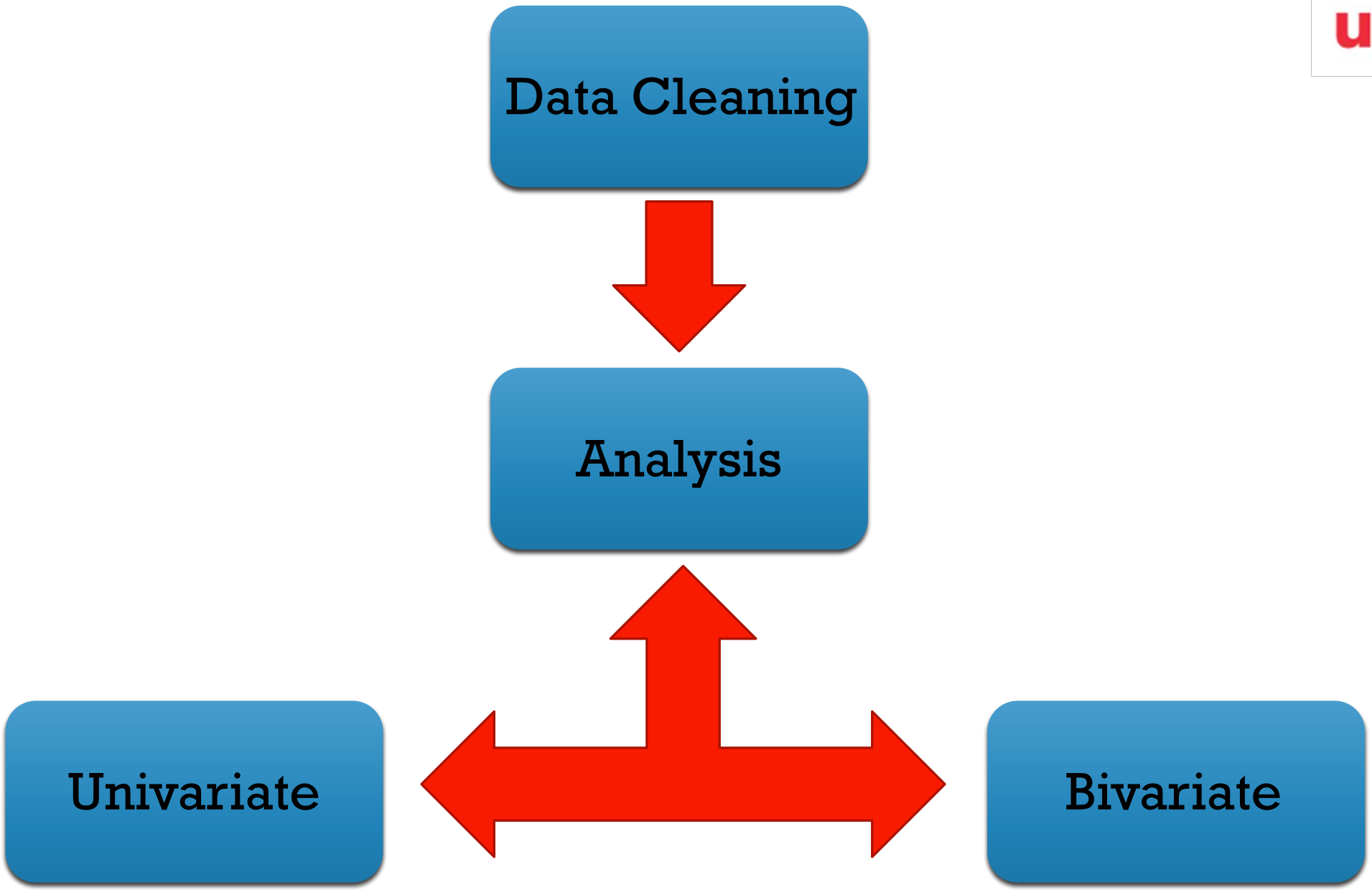
LENDING CLUB CASE STUDY

Group Members:

1. Vikas Upadhyaya
2. Harish Kumar C

Case Study Objectives





* Flow Diagram of EDA

Data Cleaning Steps

- **Delete columns:** Delete unnecessary columns.

- **Remove outliers:** Remove high and low values that would disproportionately affect the results of your analysis.

- **Missing values:** Treat missing values with appropriate approach.

- **Duplicate data:** Remove identical rows, remove rows where some columns are identical.

- **Filter rows:** Filter by segment, filter by date period to get only the rows relevant to the analysis.

Analysis

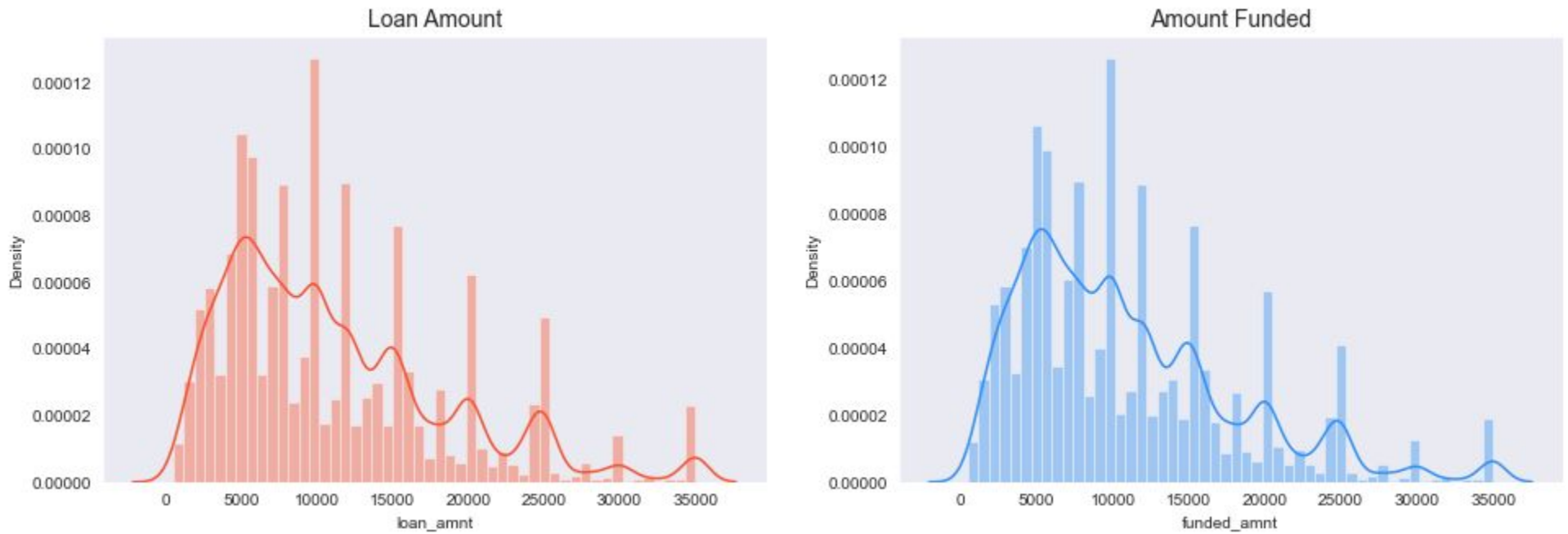
- The essence of the whole project is to analyse and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.

- We performed data cleaning and preparation on the Loan dataset:
  - Imputed the NA values for all the variables
  - Created two new columns:
    - Profit and Loss column
    - Ratio of funded amount and annual income

- During univariate analysis we have created:
  - Histograms and Bar charts to check out the distribution of all the driver variables
  - Box plots to detect the Outliers
  - Performed the Multivariate analysis to understand how different variables interact with each other.

Univariant Analysis

Quantitative variables



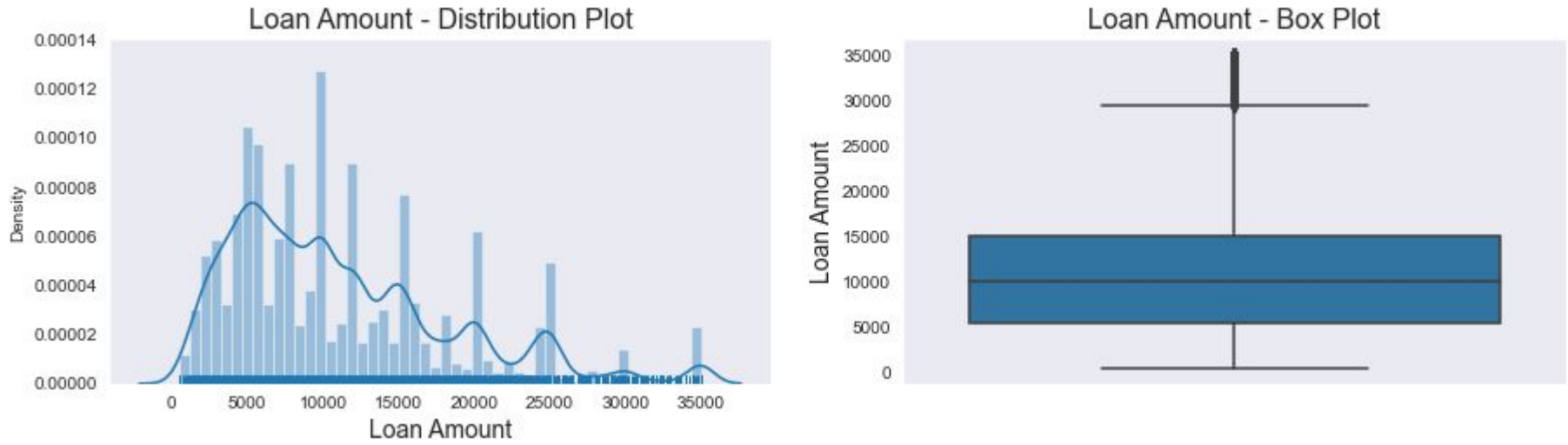
Lets see distribution of two loan amount fields using distribution plot.

Quantitative Variables

Observation:

- Distribution of amounts for all two looks very much similar.
- Hence will work with only loan amount column for rest of our analysis.

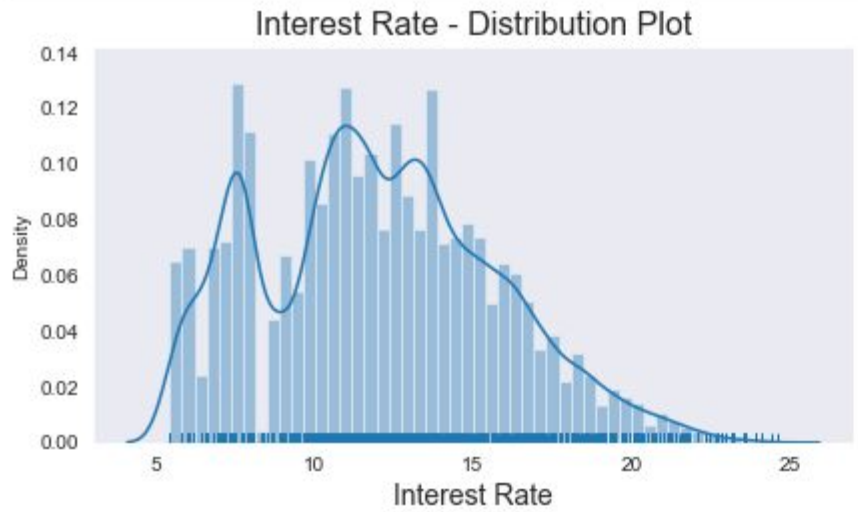
Loan amount - Quantitative variable

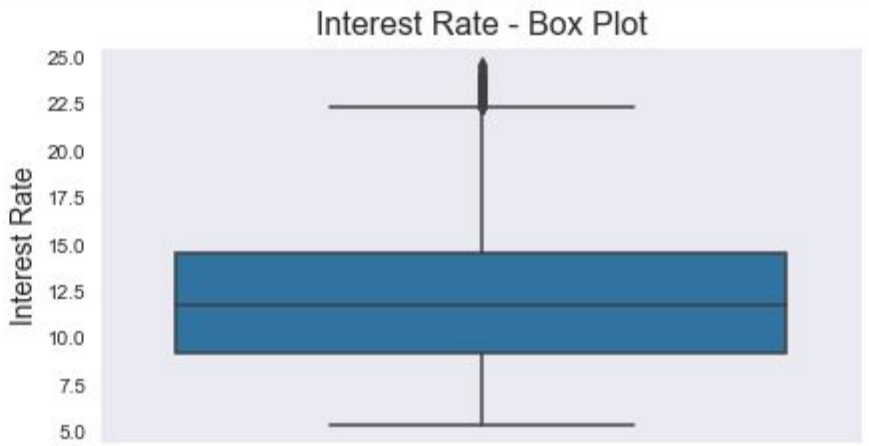


Observations:

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

Interest Rate - Quantitative one

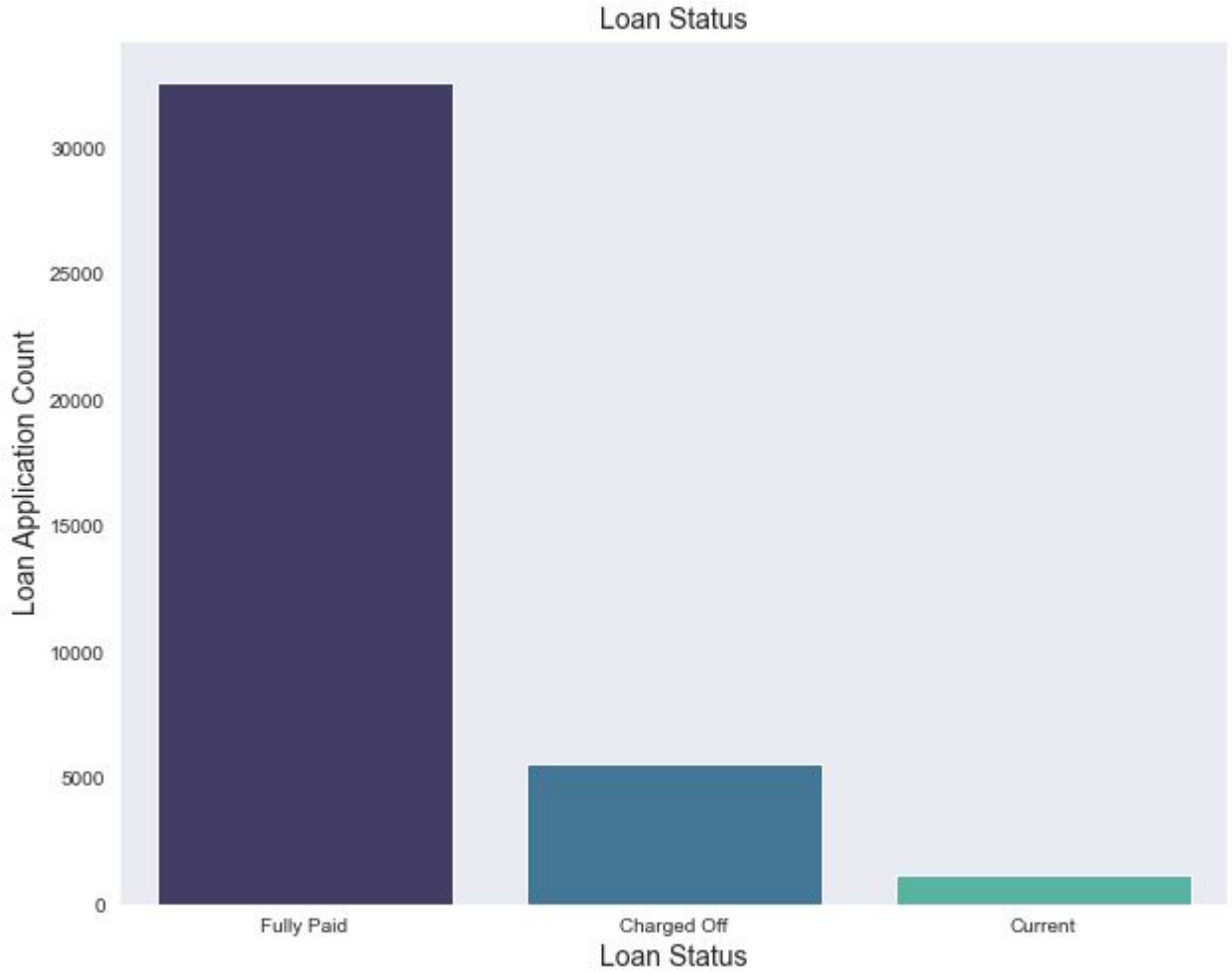




Observations:

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

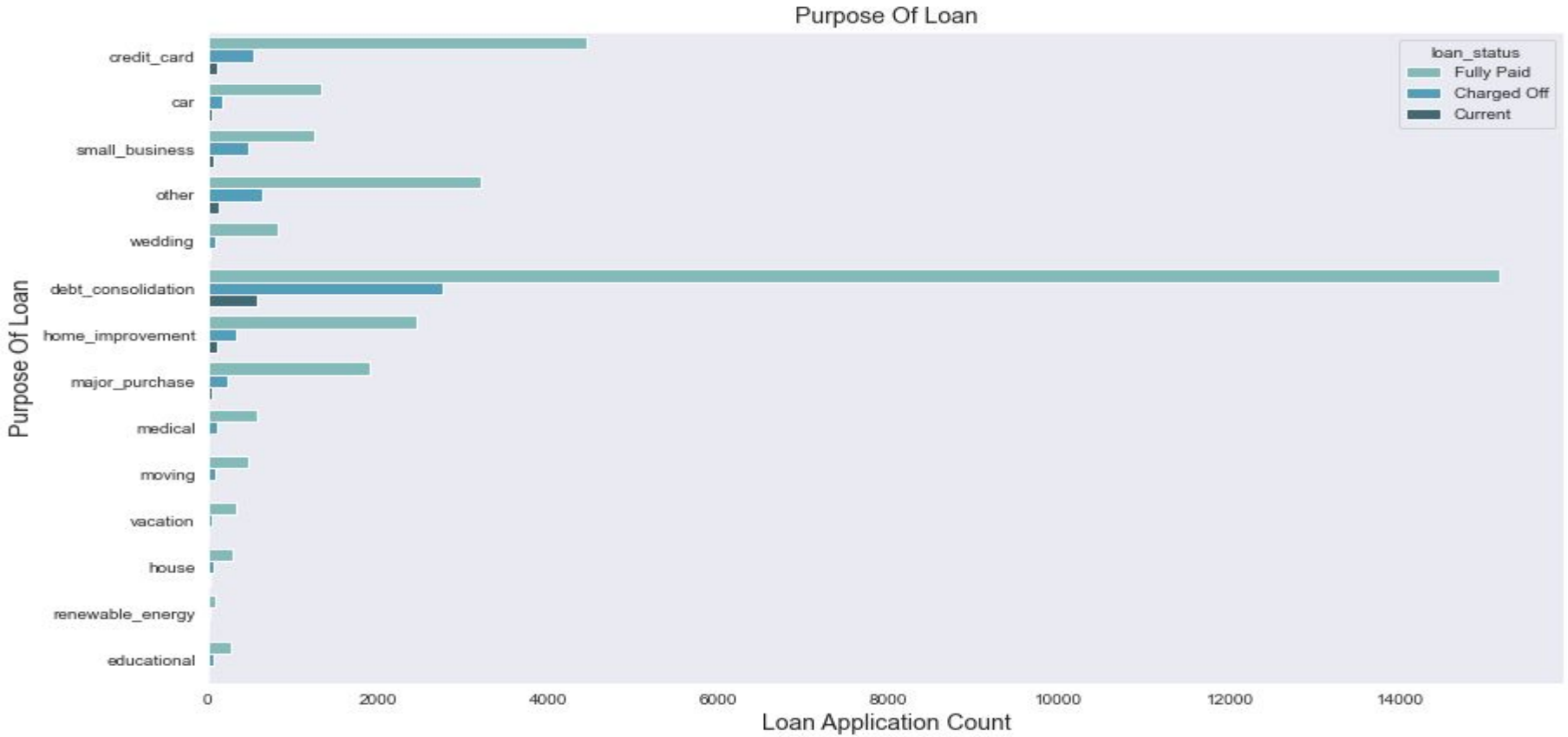
Univariant Analysis on Loan status
Unordered Categorical



Observations:

- Above plot shows that close to 14% loans were charged off out of total loan issued.

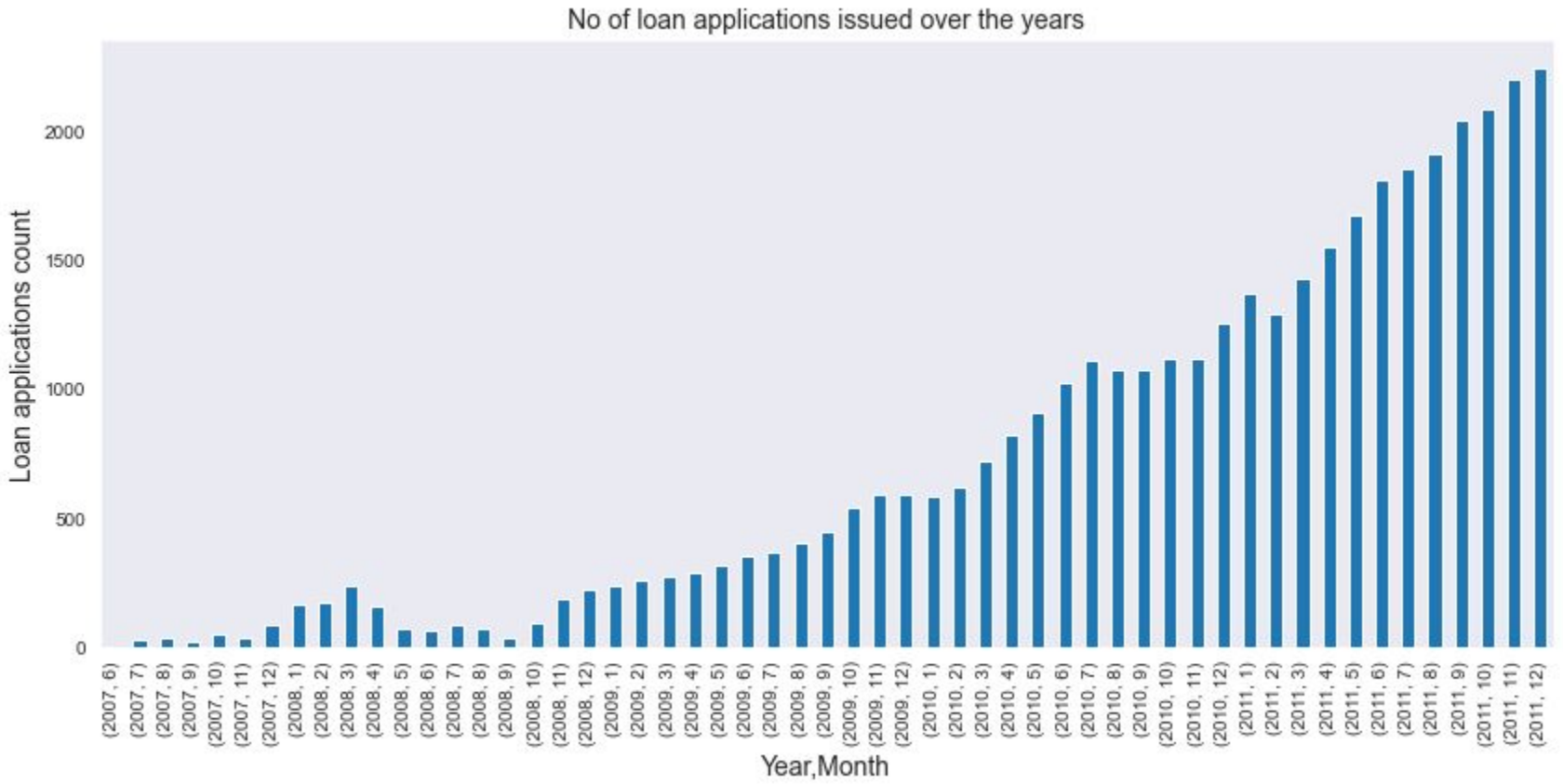
Univariant Analysis on Loan Purpose Unordered Categorical



***Observations*:**

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

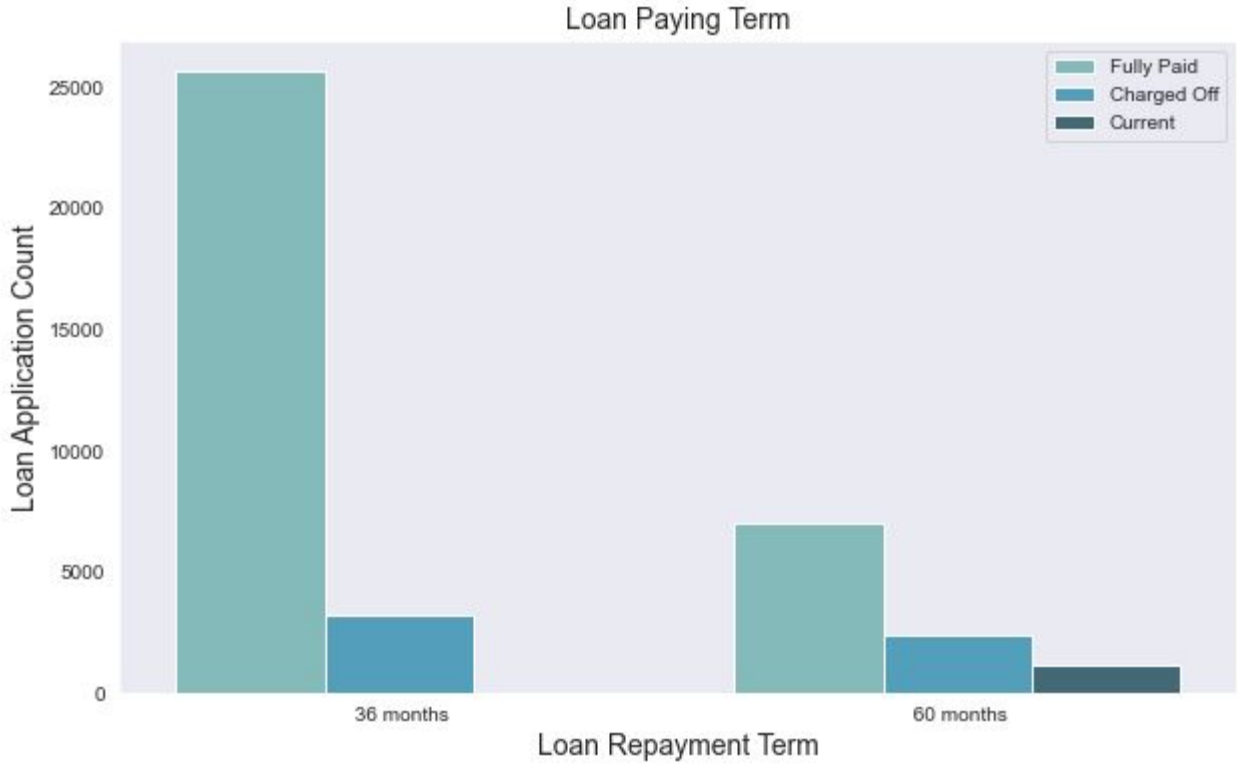
Univariant Analysis on Derive Columns Ordered Categorical



Observation:

- Is that count of loan application is increasing every passing year.
- so increase in number of loan applications are adding more to number of charged off applications

Univariant Analysis on loan paying term Ordered Categorical

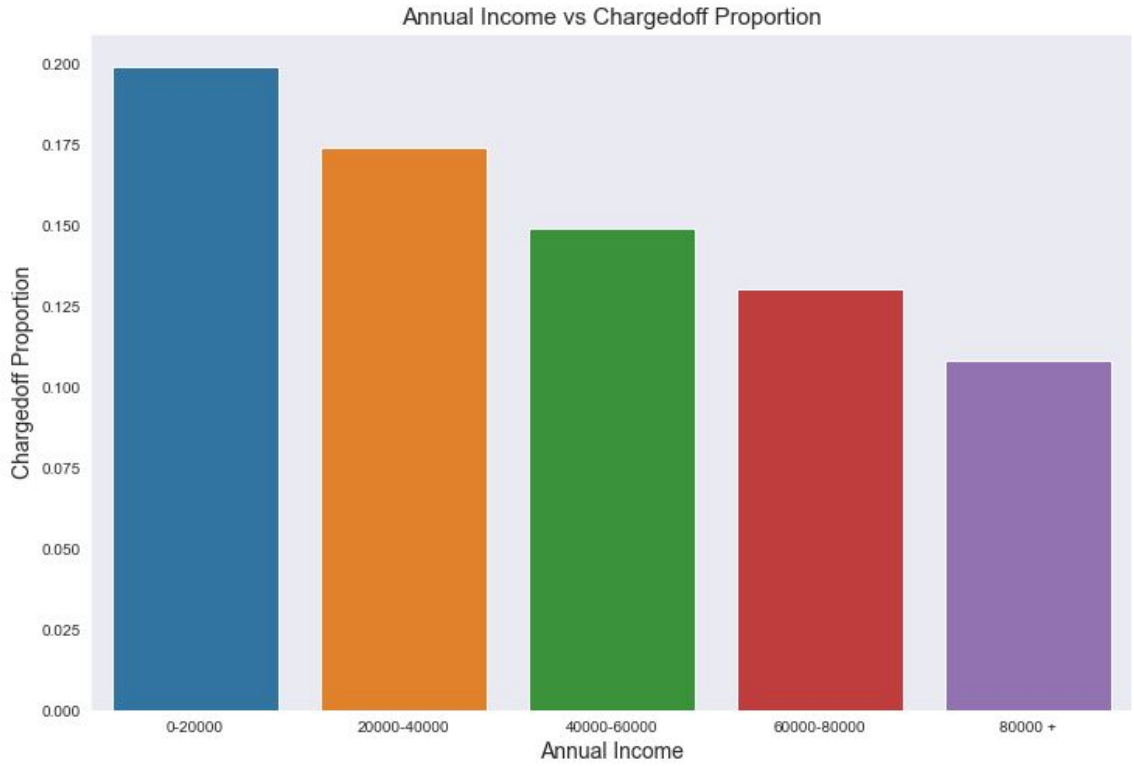


Observations:

- 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.

Bivariant Analysis

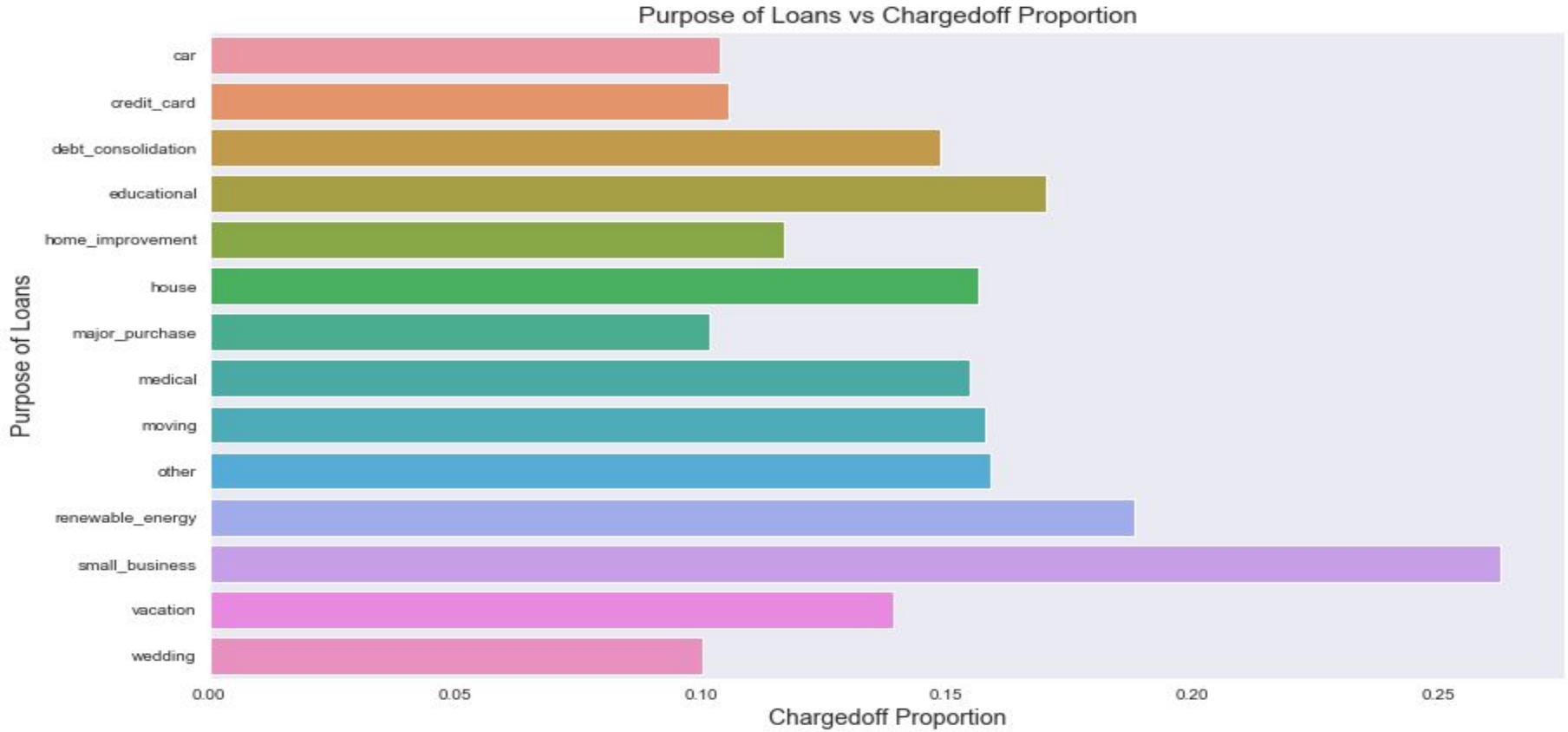
Bivariate Analysis on annual income against Charged-Off_Proportion



Observations:

- Income range 80000+ has less 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.

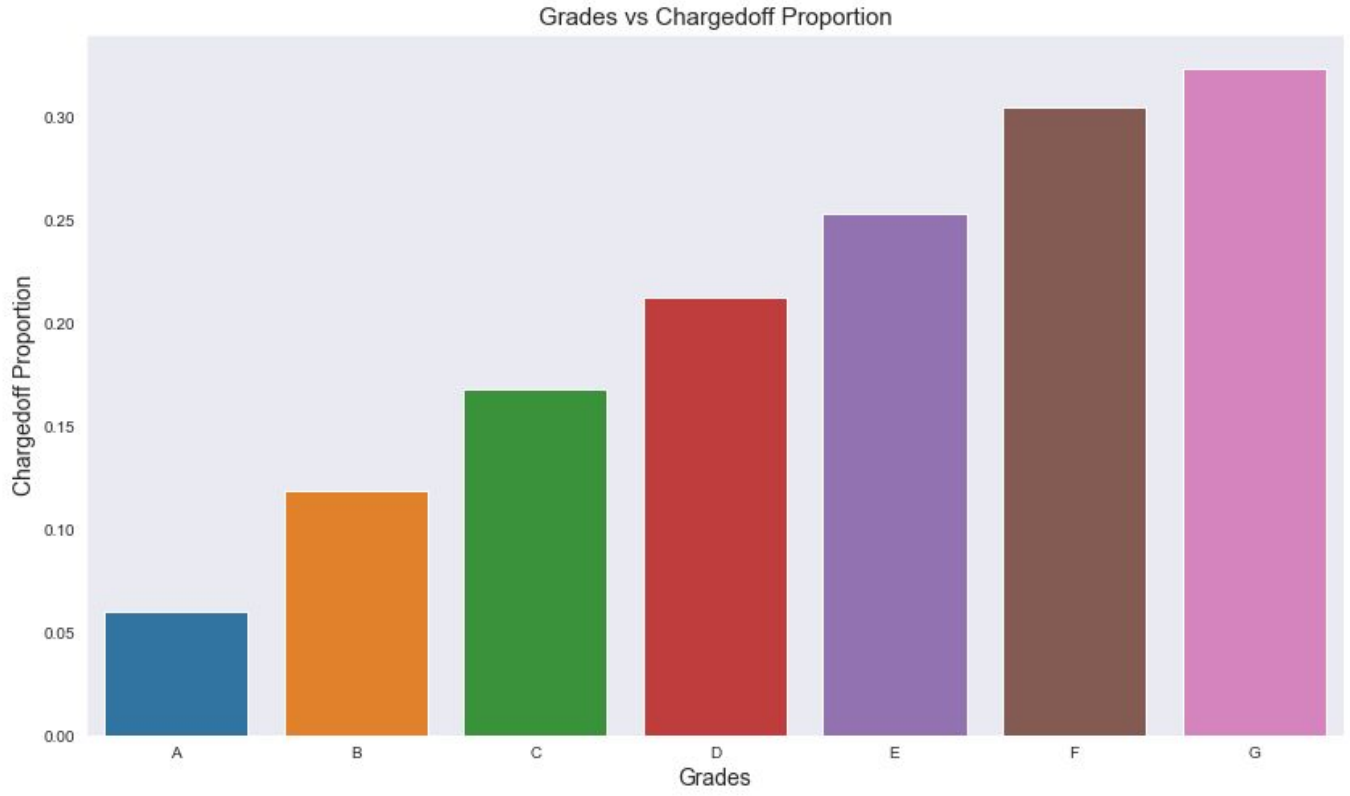
Bivariate Analysis on purpose of Loan against Charged-off Proportion



Observations:

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

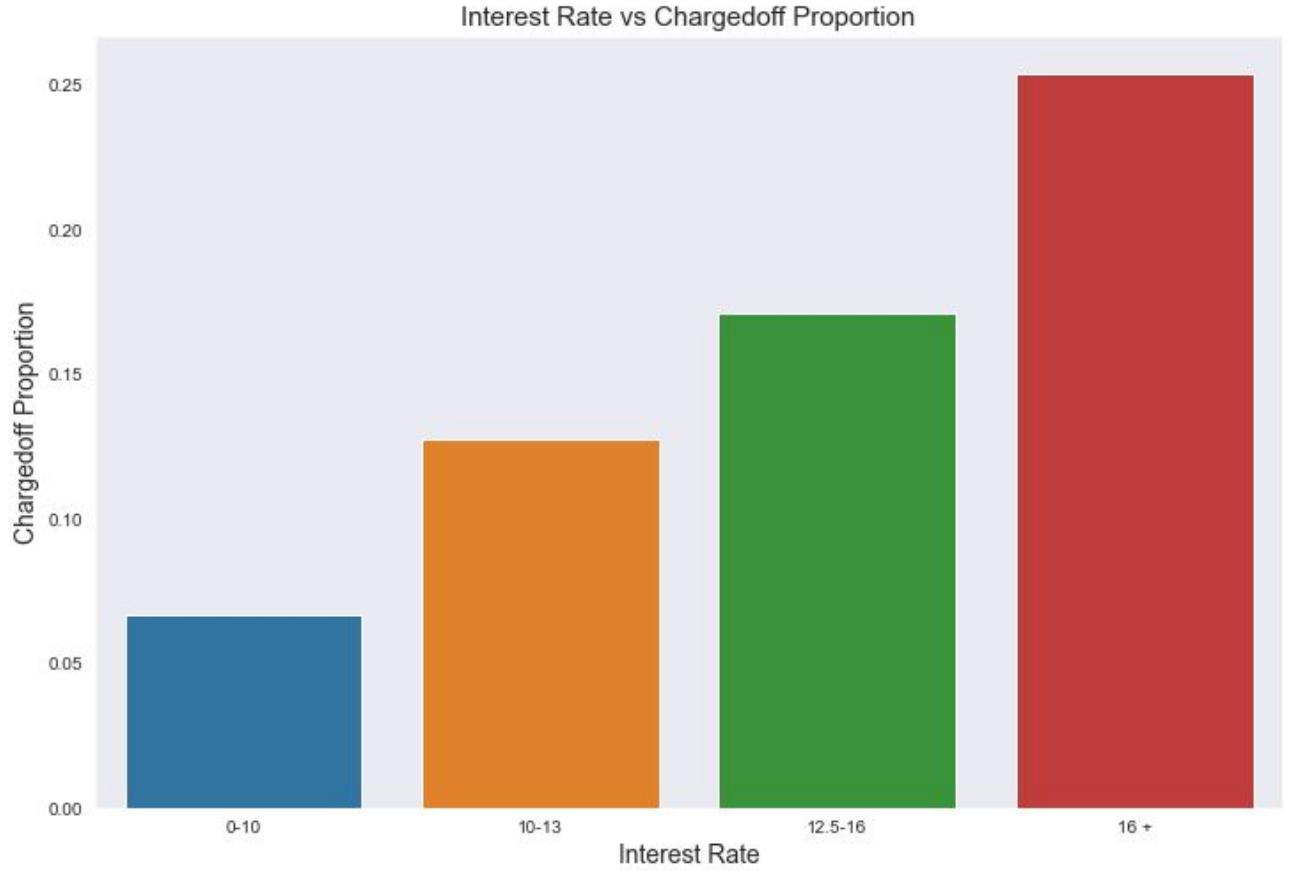
Bivariate Analysis on grade against Charged-off Proportion



Observations:

- Grade "A" has very less chances of charged off.
- Grade "F" and "G" have very high chances of charged off.
- Chances of charged off is increasing with grade moving from "A" towards "G"

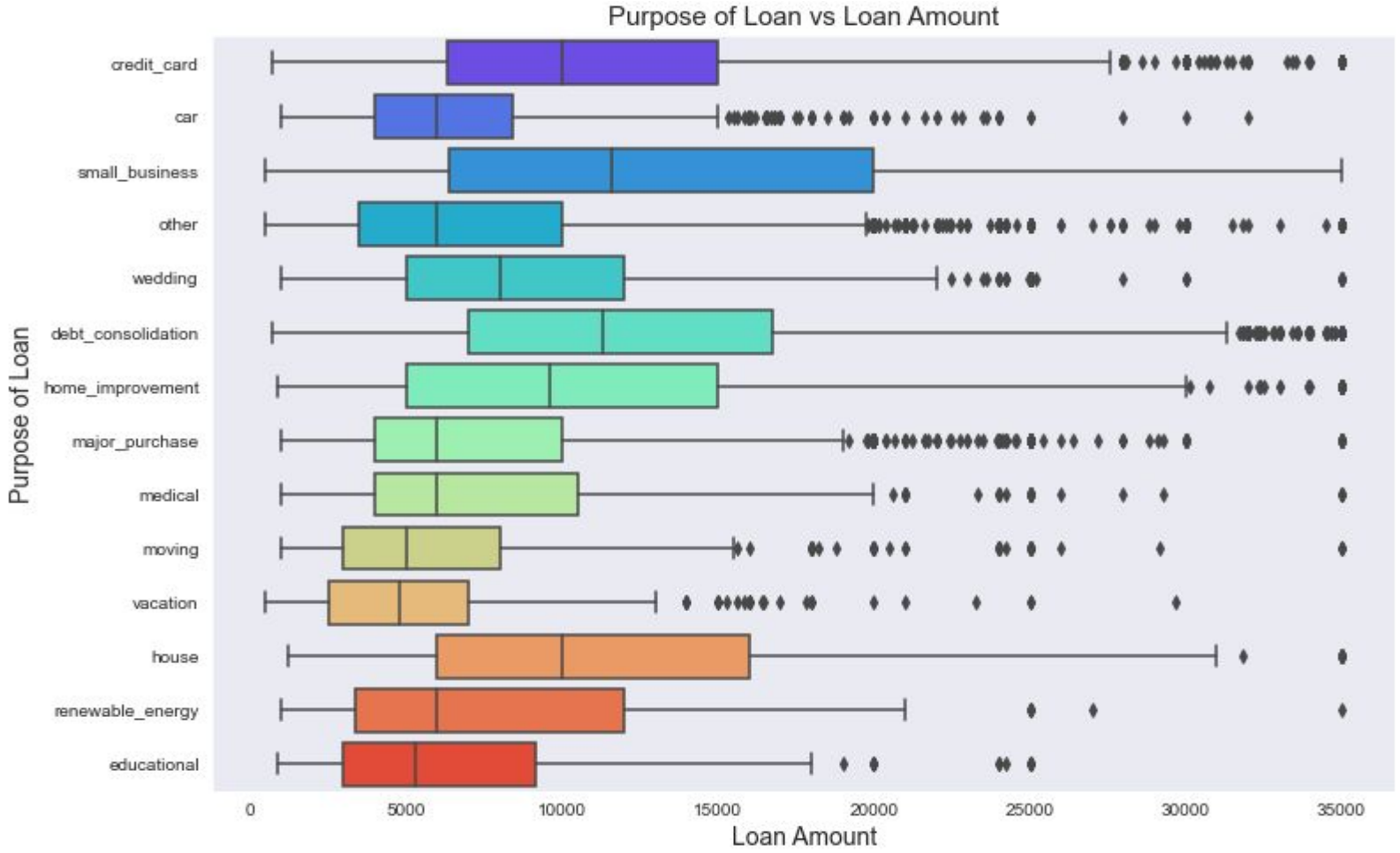
Bivariate Analysis on interest rate against Charged-off Proportion



Observations:

- 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.

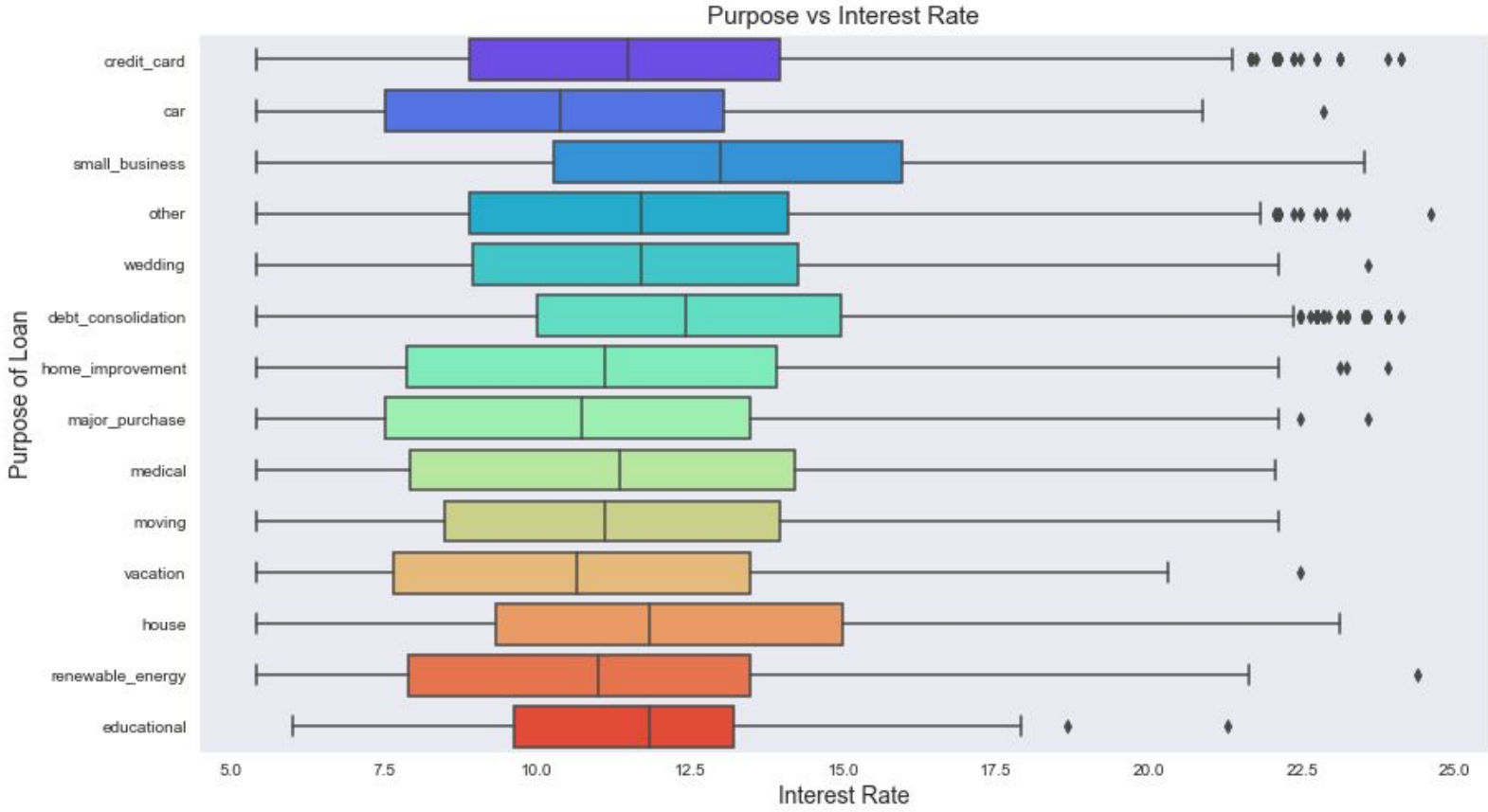
Bivariant Analysis using Boxplot
Purpose of loan vs Loan amount



Observations:

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

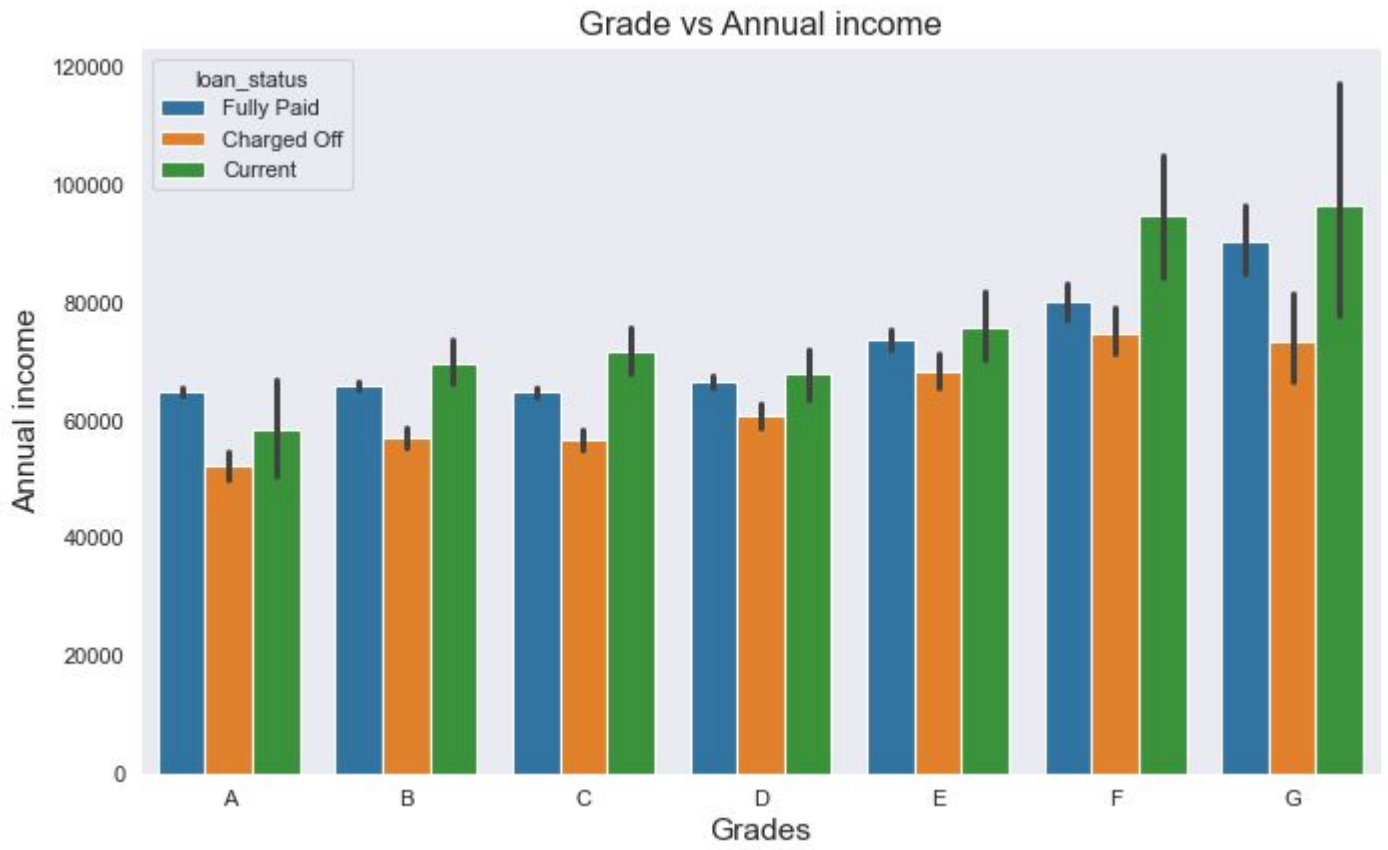
Bivariate Analysis - Purpose vs Interest Rate



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.

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



Observations:

- 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)

Conclusion

1. Lower annual incomes are more likely to be defaulter
2. Higher the interest rate higher charged off ratio
3. Those who are not working or have less than 1 year of work experience have 4. high chances of getting charged off
5. Grade "F" and "G" have very high chances of charged off.
6. Small Business applicants have high chances of getting charged off.

Thank You