Lending Club Case study

Group members:

- Anurag Das
- Yoga Sree Durga K

Business Objective

- Lending club is a consumer finance company which specialises in lending various types of loans to urban customers.
- The dataset provided is the complete loan data for all loans issued through the time period 2007 t0 2011.
- The objective of the analysis is to use the information about past loan applicants and whether they ‘defaulted’ or not

Data Analysis Techniques

Data Understanding

- Getting to know the columns and the domain specific variables

Data Cleaning

- Removing all null valued columns
- Find missing %of each column and drop columns having missing values>75%
- Removing redundant and unnecessary columns/rows not required for the release
- Imputing numerical and categorical values
- Converting datatypes

Univariate Analysis

- Analysing each column and observing the frequency distribution of data of each column

Segmented Univariate Analysis

- Analysing continuous data variables with categorical columns

Bivariate Analysis

- Analysing two variable behaviour and finding the correlation between them

Recommendations

Analysing all plots and suggestions to the business for minimizing the loss and predicting the factors contributing to loan defaulters

High Level tasks

- The objective of the case study is to understand how customer attributes and loan attributes influence the tendency of defaulting

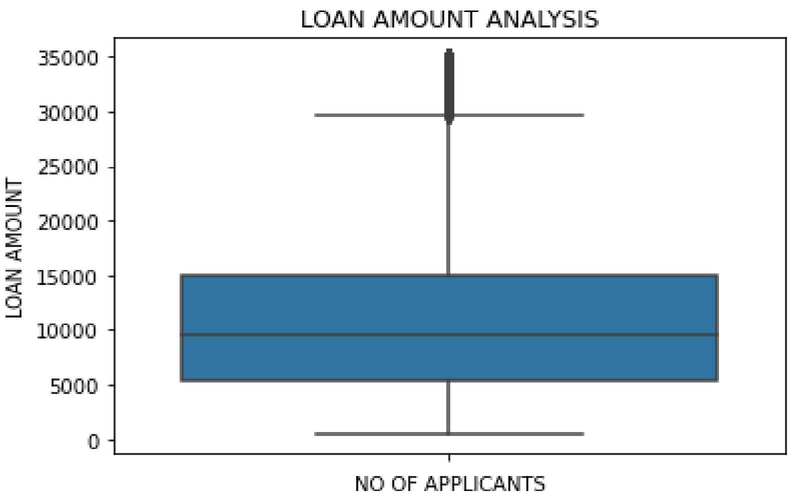
- Prepared the data by performing the following tasks:

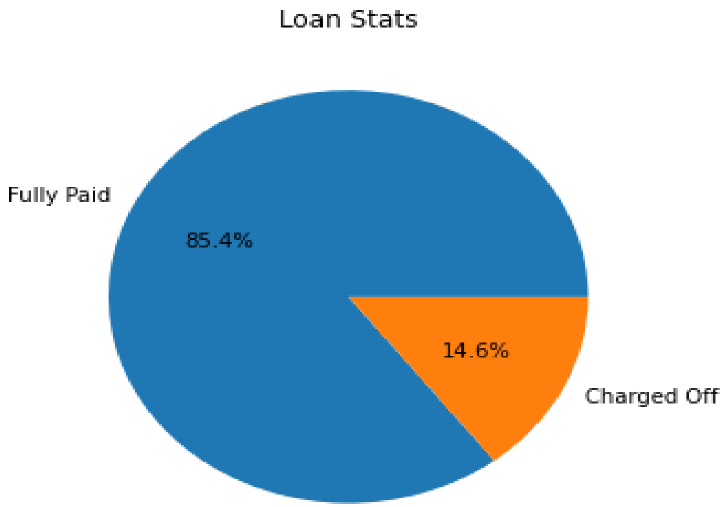
  - Imputed the NA values
  - Converted the datatypes
  - Derived new metrics for the analysis

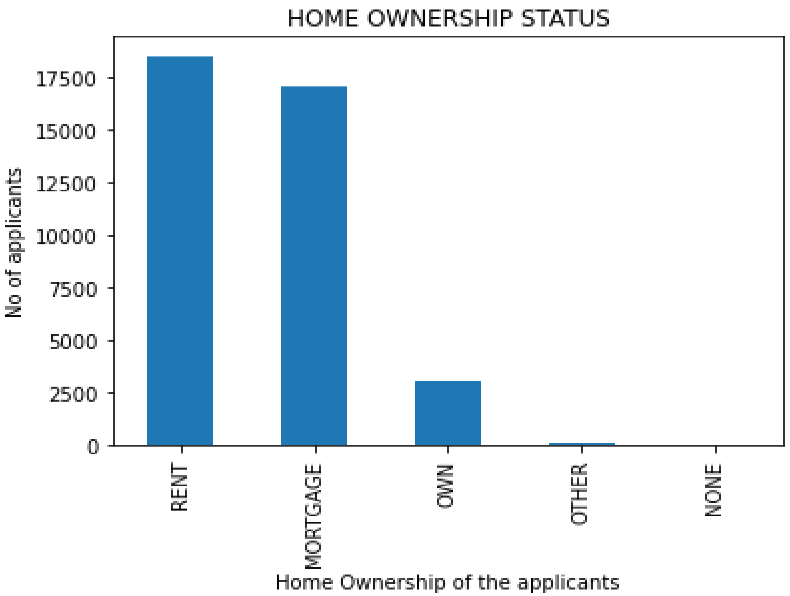
- For Univariate and Bivariate analysis, created the following plots:

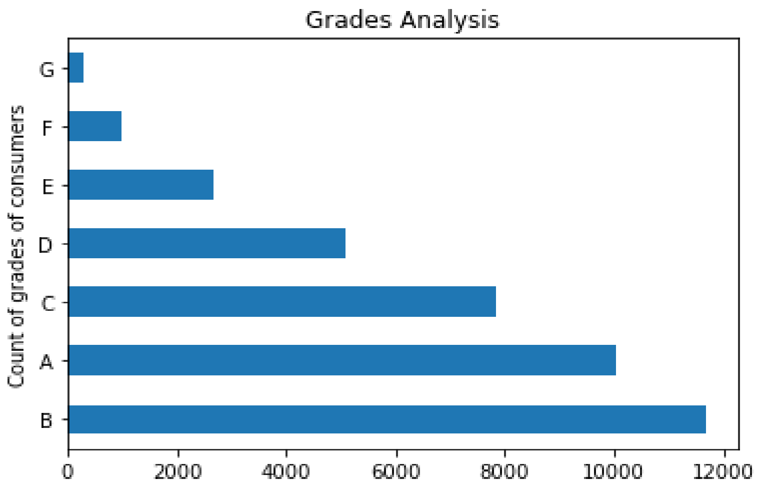
  - Histograms ,Bar charts, count plots to check the distribution of all driver variables
  - Boxplots to detect the outliers
  - Correlation and heatmap to understand the correlation between variables

Data Analysis - Univariate



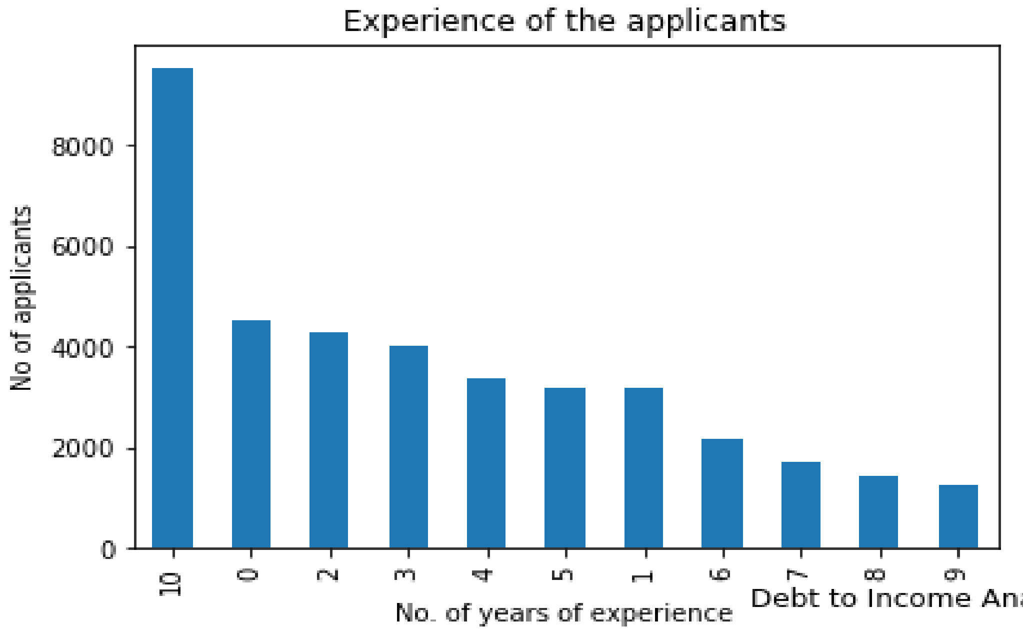


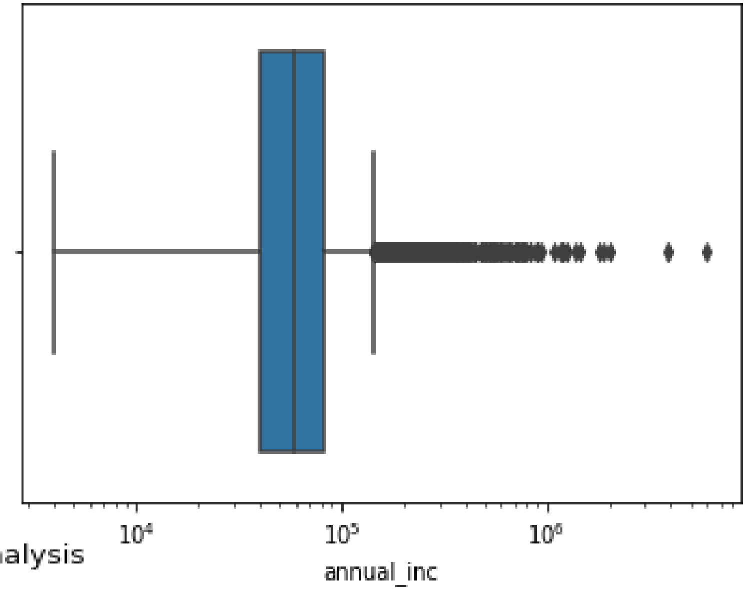




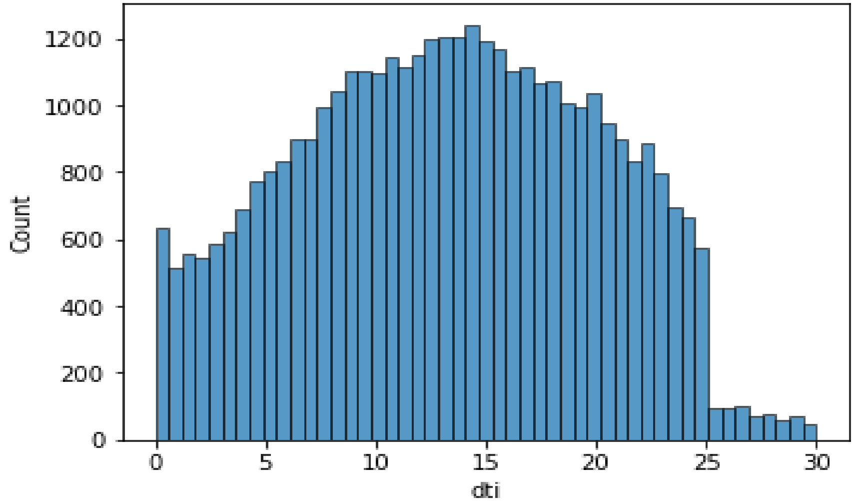
- **Inferences:**
  - Most of the loans applied are below 30000 with average number of consumers applying loans for 10000
  - More than 85% of the loans are fully paid
  - Majority of the applicants lived in Rented houses and houses in Mortgage
  - 56% of the loans are of Grade A and Grade B,hereby, most of the loans are top graded loans

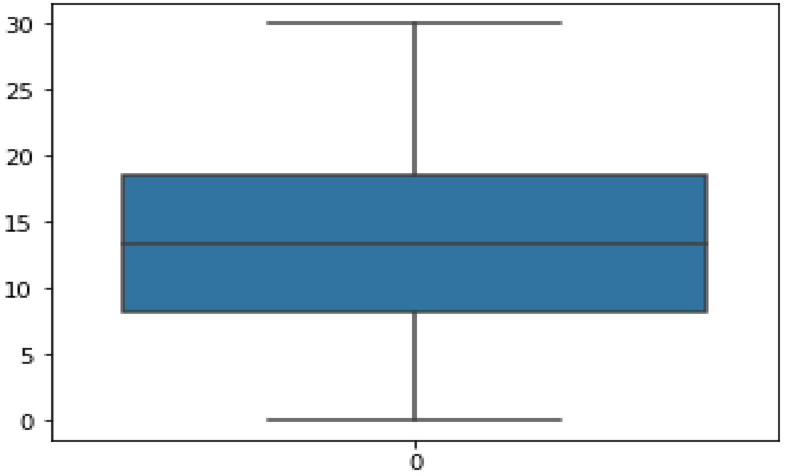
Data Analysis - Univariate





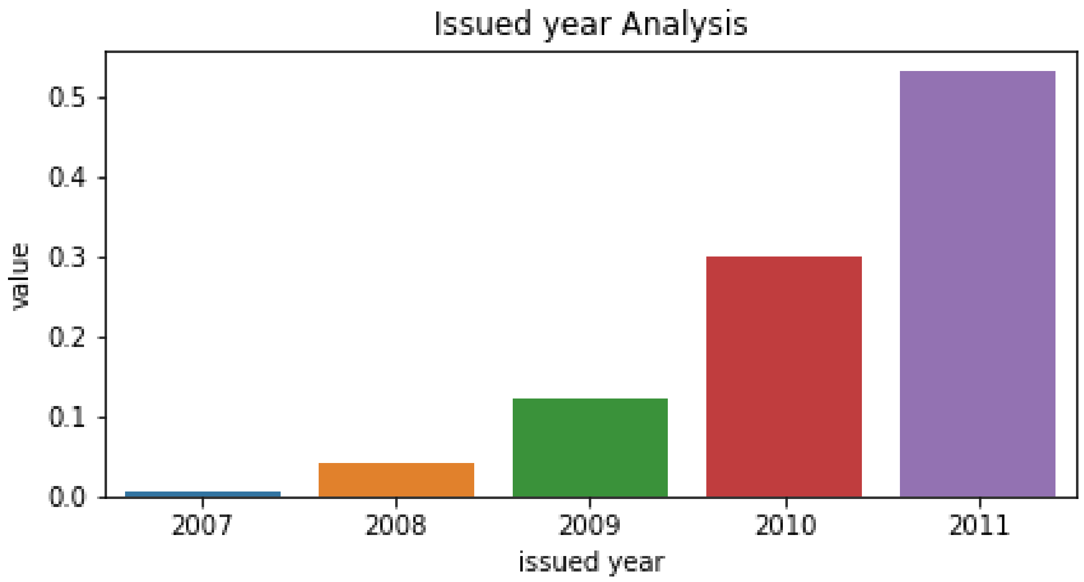
Debt to Income Analysis

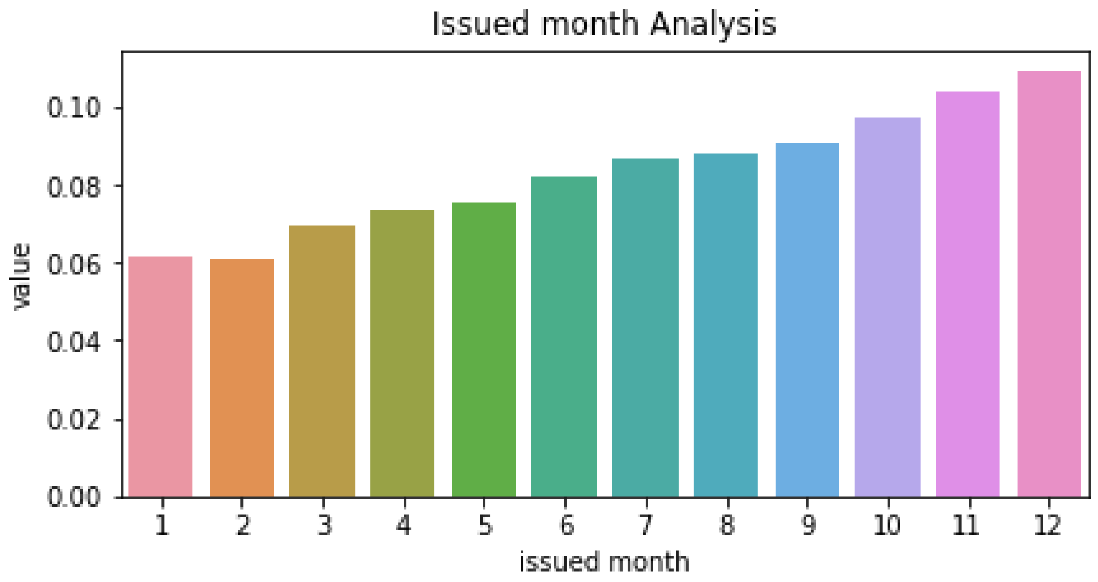




- **Inferences:**
  - ➢ Most of the loans applied are below 30000 with average number of consumers applying loans for 10000
  - ➢ More than 85% of the loans are fully paid
  - ➢ Majority of the applicants lived in Rented houses and houses in Mortgage
  - ➢ 56% of the loans are of Grade A and Grade B,hereby, most of the loans are top graded loans

Data Analysis - Univariate

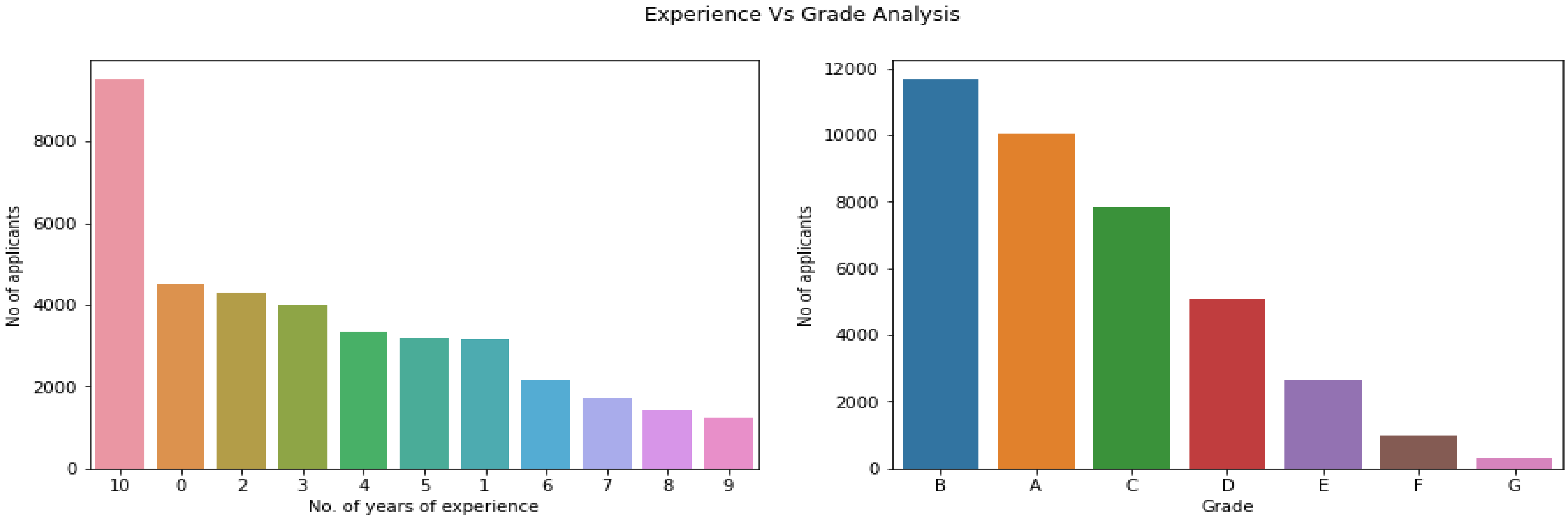




- **Inferences**

➢ The loans requested by the consumers increases every year and is the highest in 2011.Most of the loans were requested in October, November and December

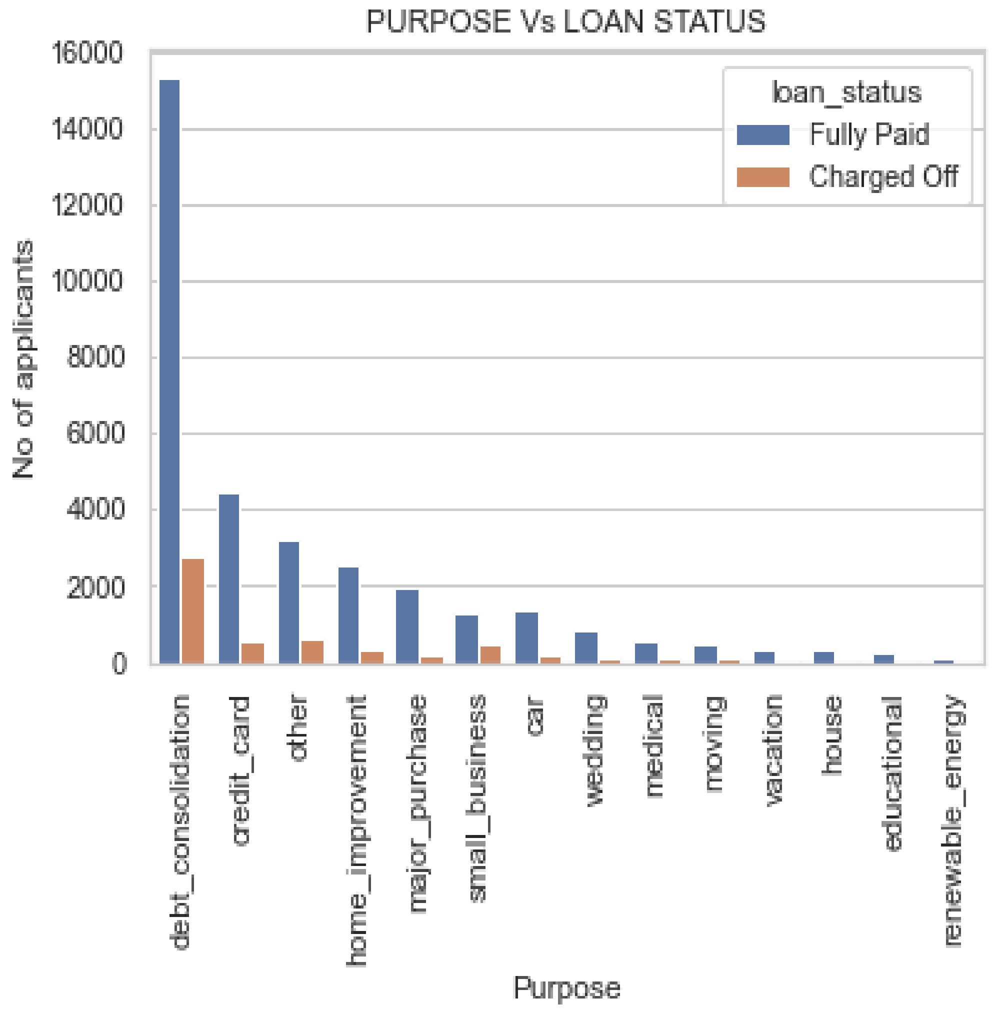
Data Analysis - Univariate

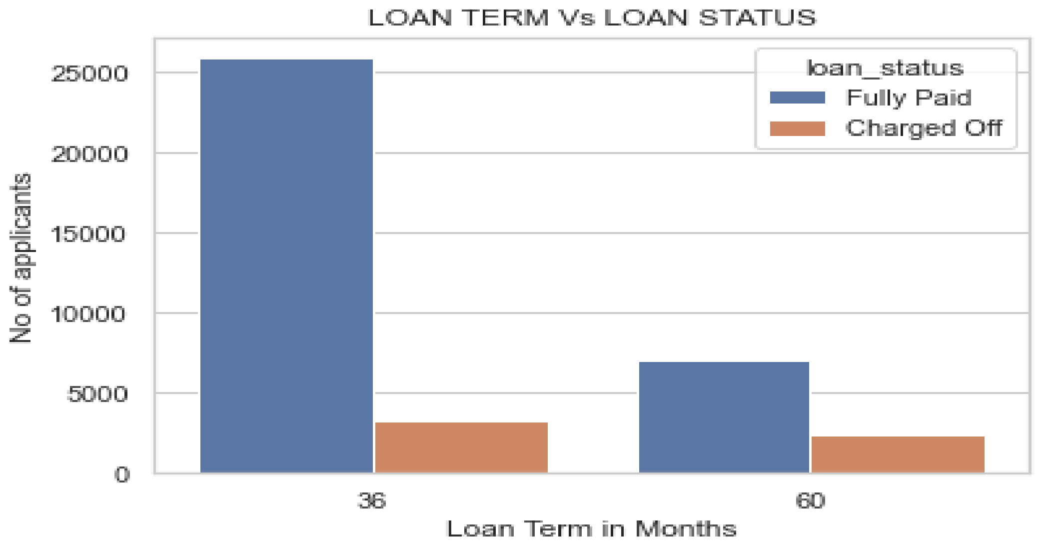


- **Inferences:**

People having more than 10 years of experience have applied for the highest grades loans such as B,A,C

Data Analysis-Segmented Univariate



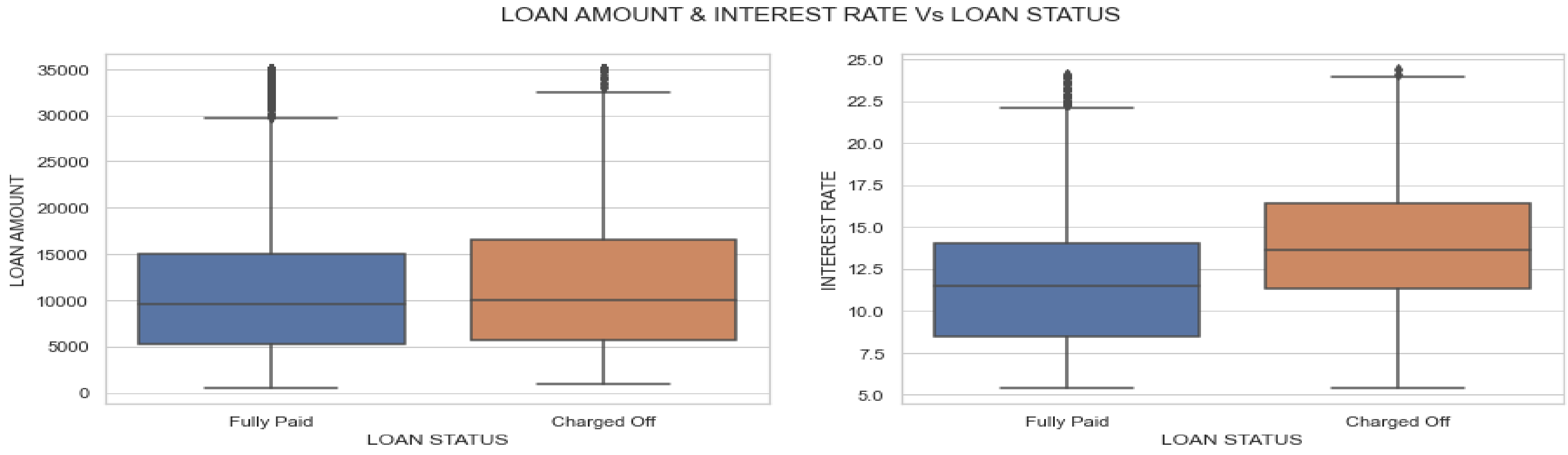


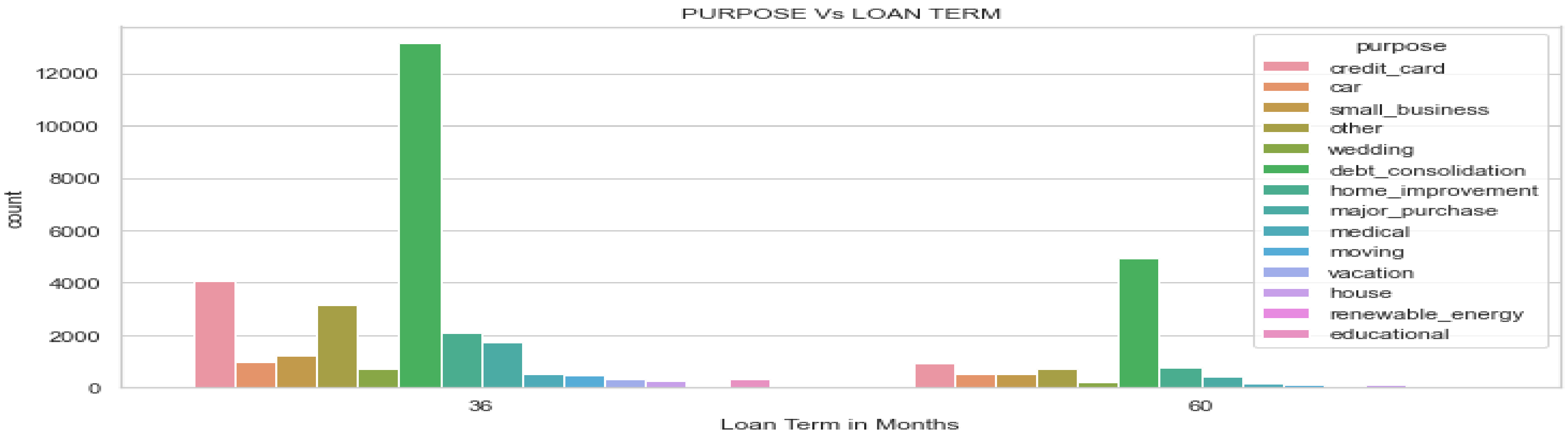
- **Inferences:**

➢ The loans taken for Debt consolidation is comparatively high than the loans taken for other purposes and have been fully paid

➢ The loans by consumers for a short duration is likely to be charged off

Data Analysis-Bivariate



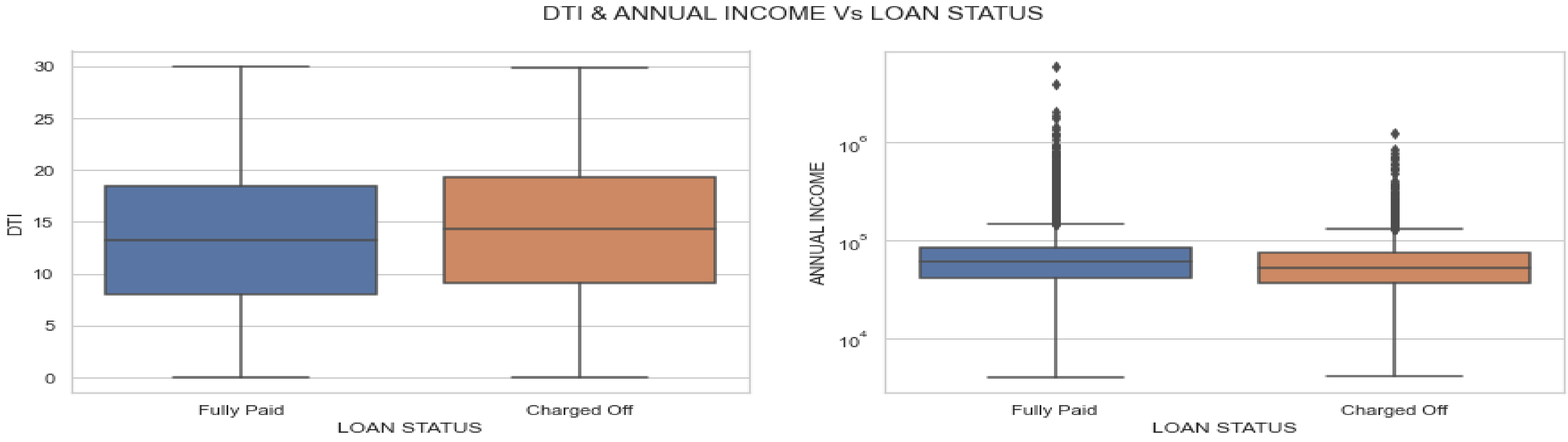


- **Inferences:**

➢ Loan amount when compared to the Loan status, the mean between Fully Paid and Charged Off is negligible, whereas, the mean value of the Charged off consumers is higher than that of the Fully Paid consumers

➢ Applicants taken Loan for shorter term are likely to be defaulters

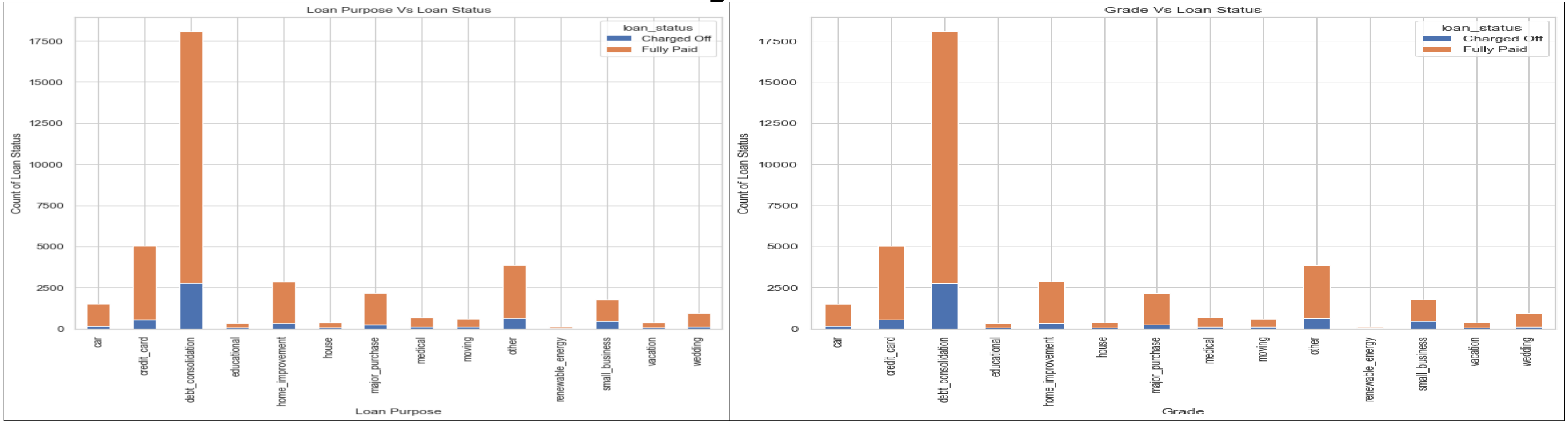
Data Analysis-Bivariate



- **Inferences:**

➢ DTI Vs Loan status shows that the mean of DTI of Charged off loans is slightly higher than the Fully paid loans. Also The annual income of Fully paid consumers are higher than the Charged off consumers

Data Analysis-Bivariate

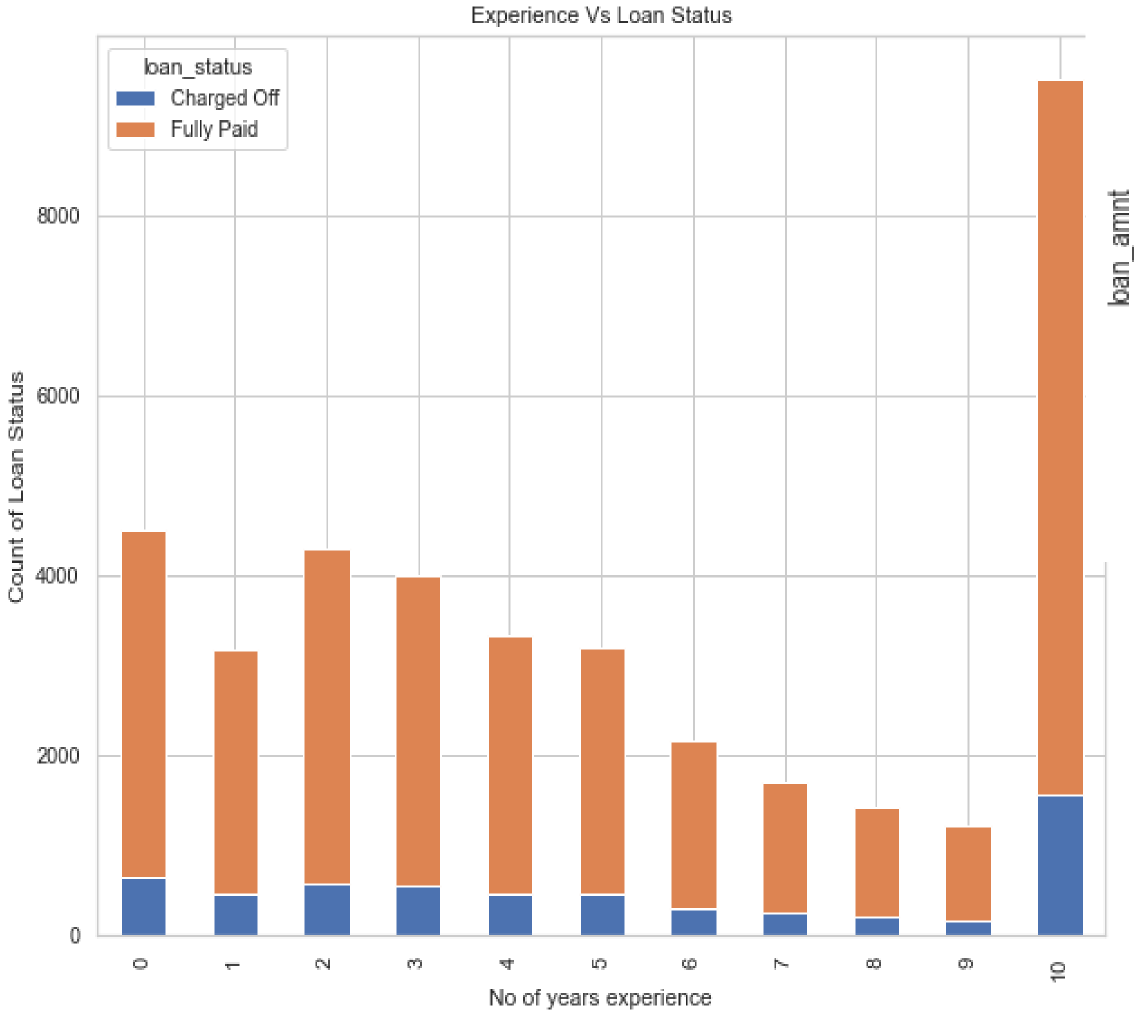


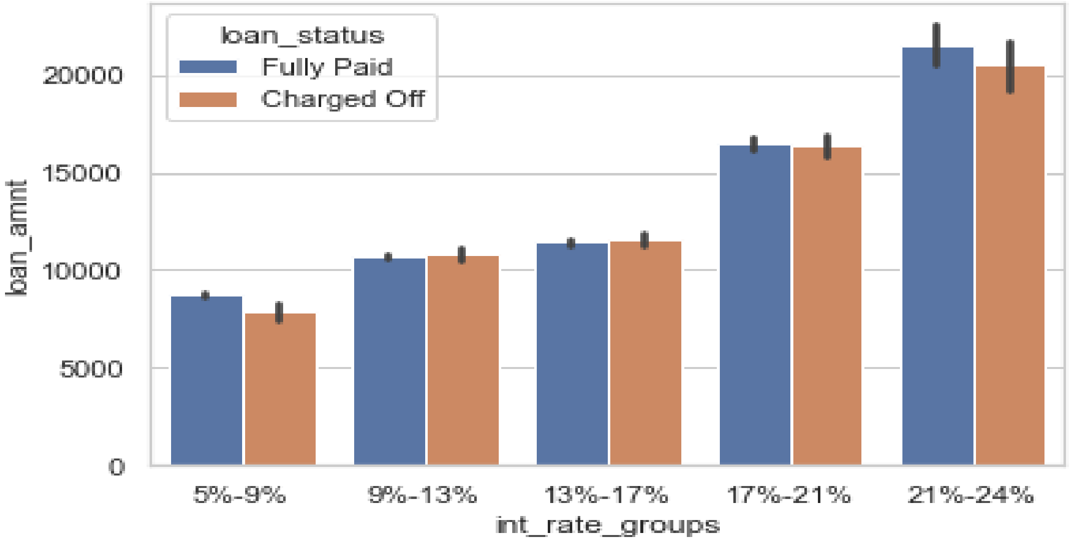
- **Inferences:**

➢ More applicants have defaulted for getting loan for debt consolation

➢ Majority of the applicants have defaulted for B grade loans

Data Analysis-Bivariate



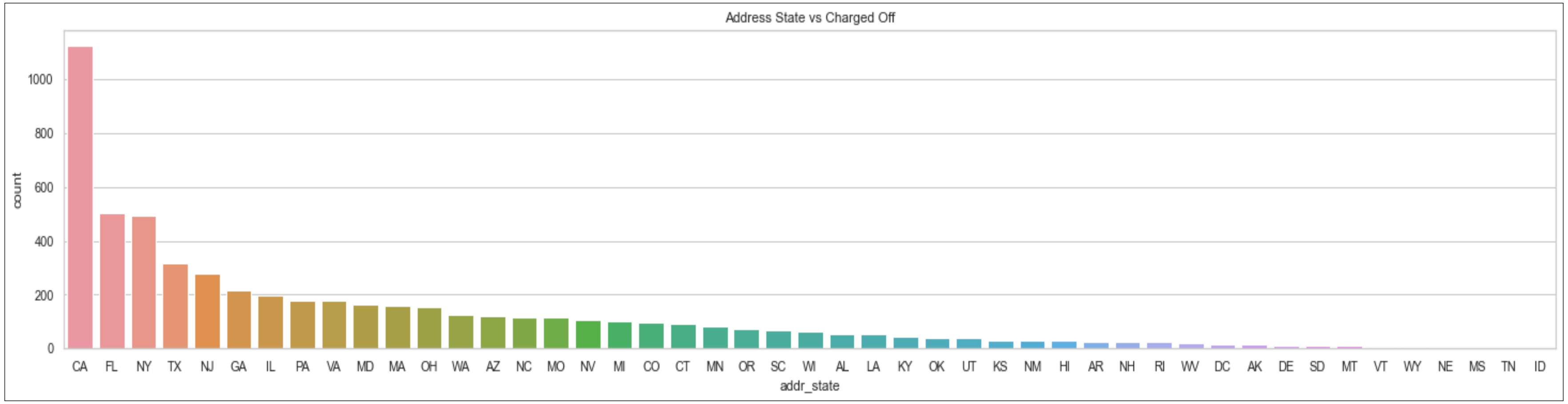


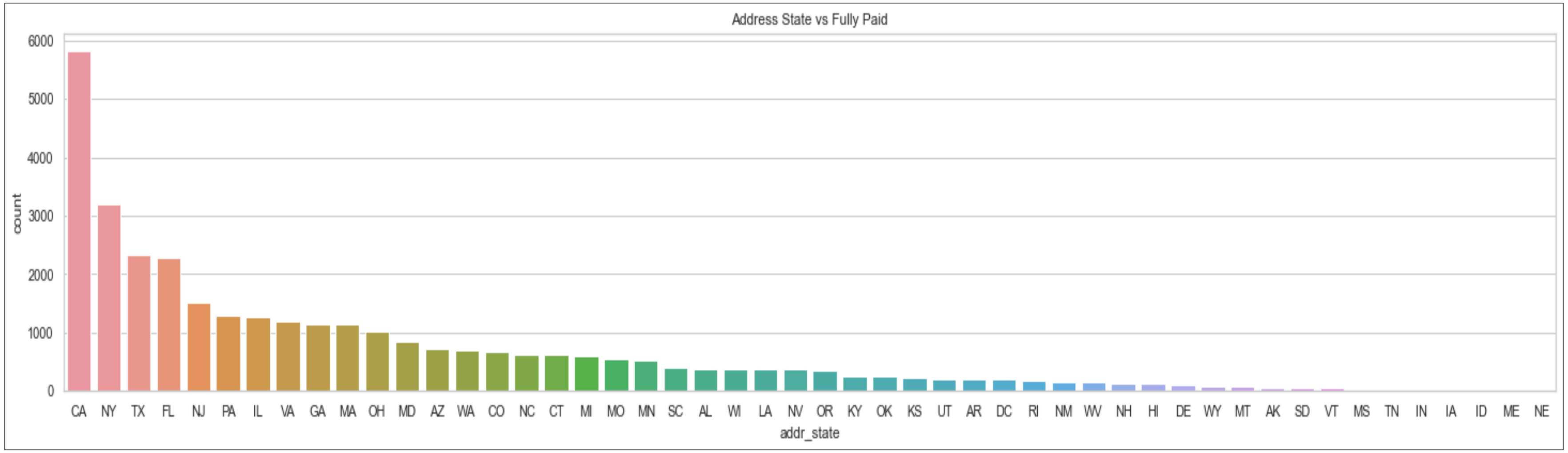
- **Inferences:**

➢ Applicants with work experience of 10+ years are likely to default

➢ The applicants who have taken loans with higher interest rates are likely to default. The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups. This can be a pretty strong driving factor for loan defaulting.

Data Analysis-Bivariate

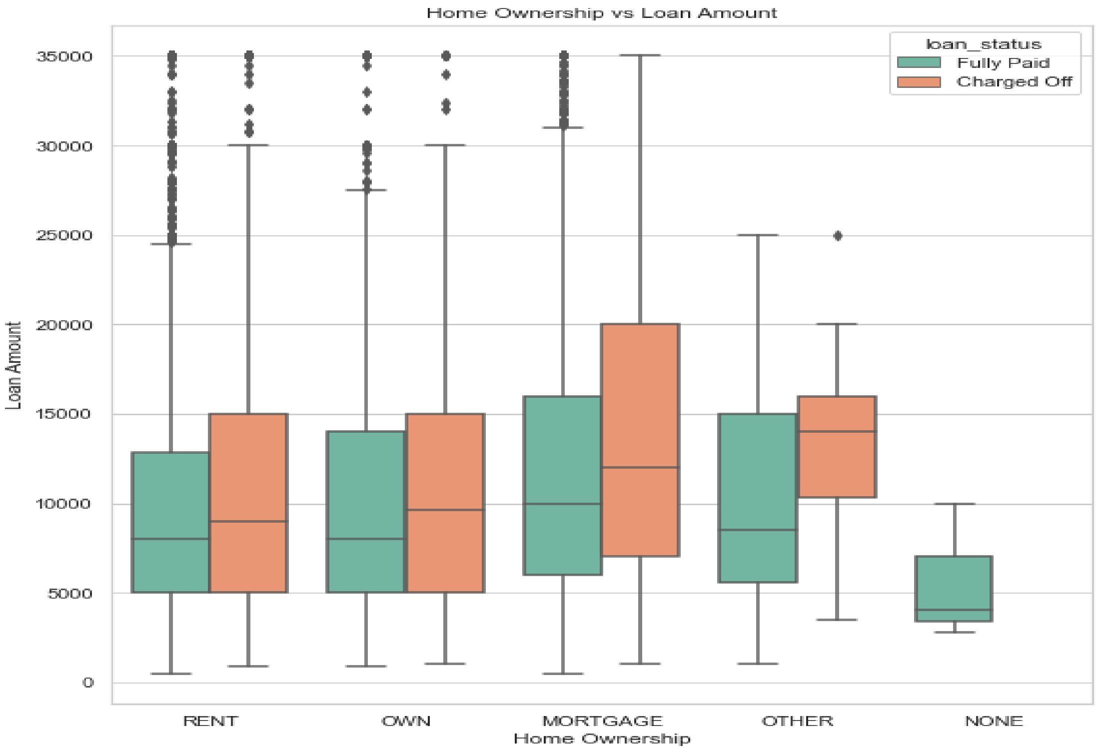


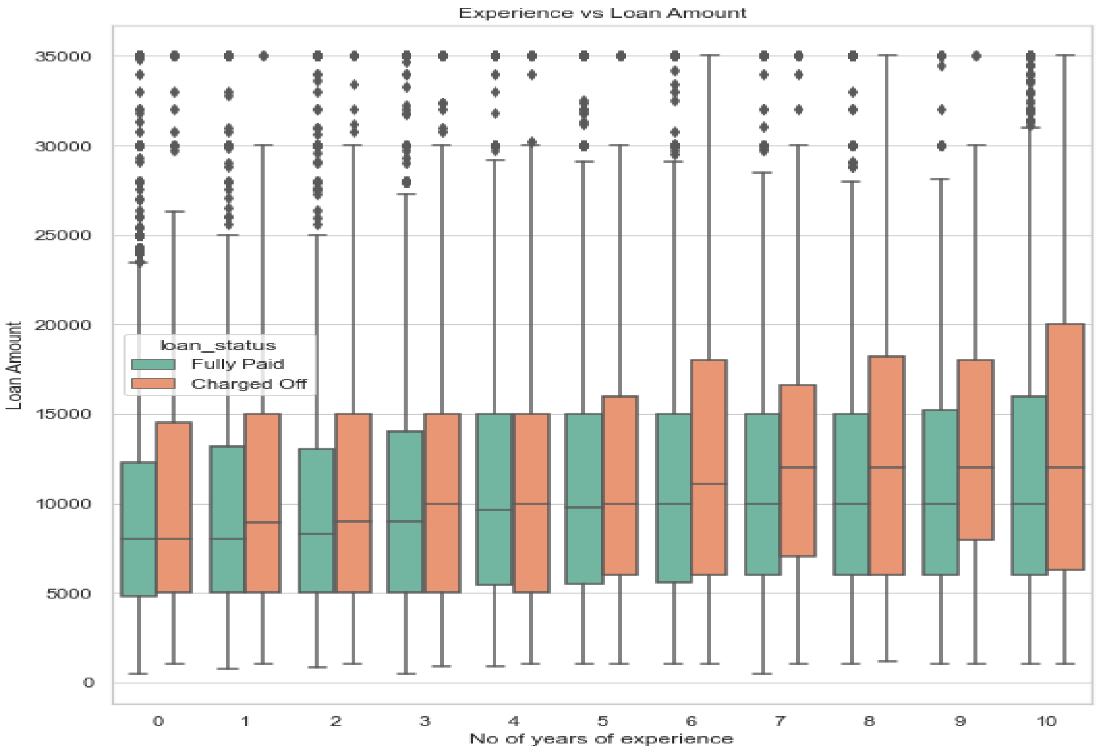


- **Inferences:**

The applicants from CA,FL and NY States are most likely to default

Data Analysis-Bivariate

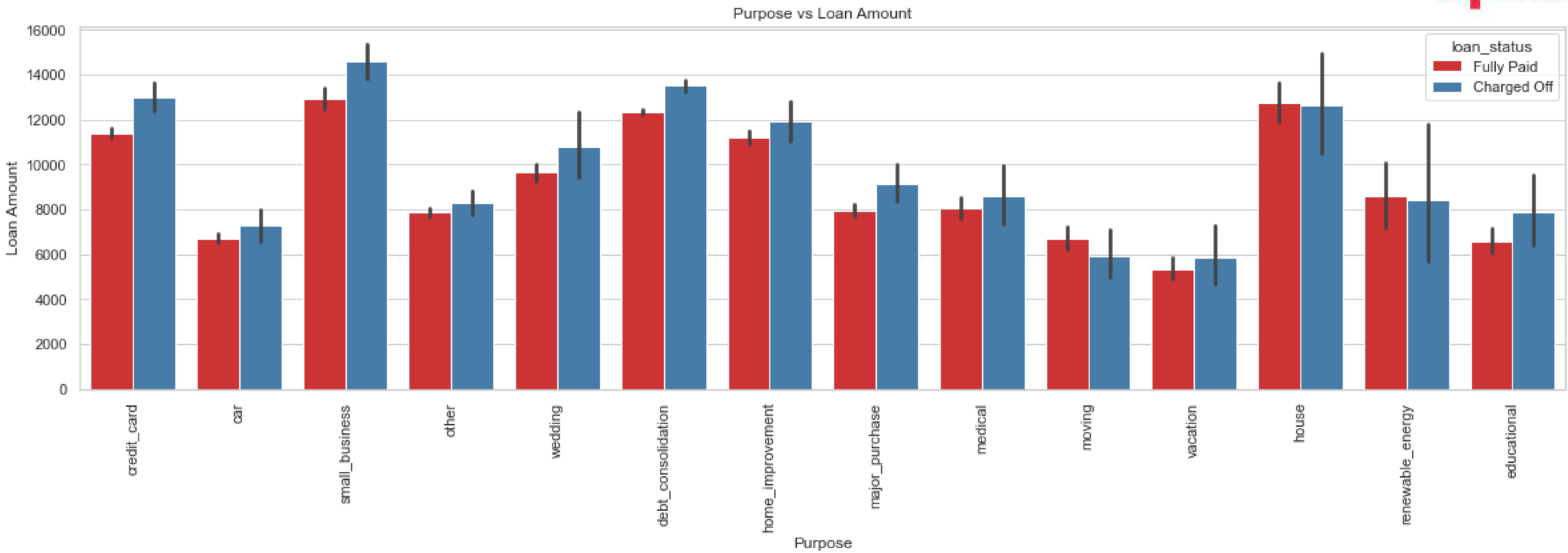




- **Inferences:**

➢ Applicants whose home is in mortgage and having work experience of more than 10 yers are likely to default

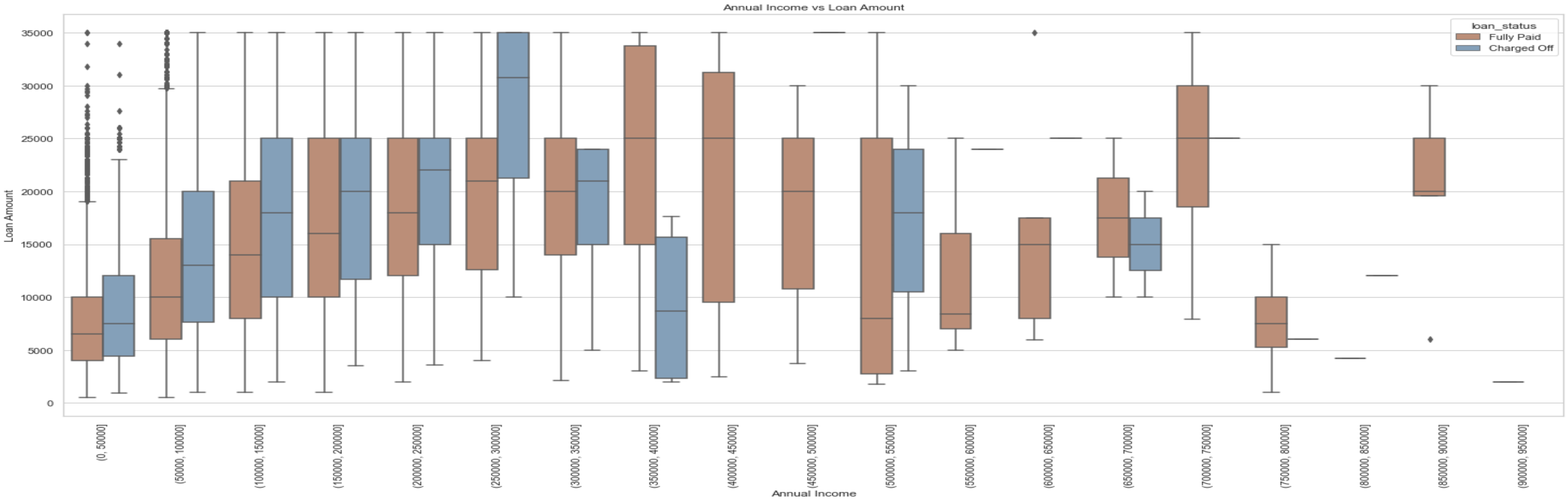
Data Analysis-Bivariate



- **Inferences:**

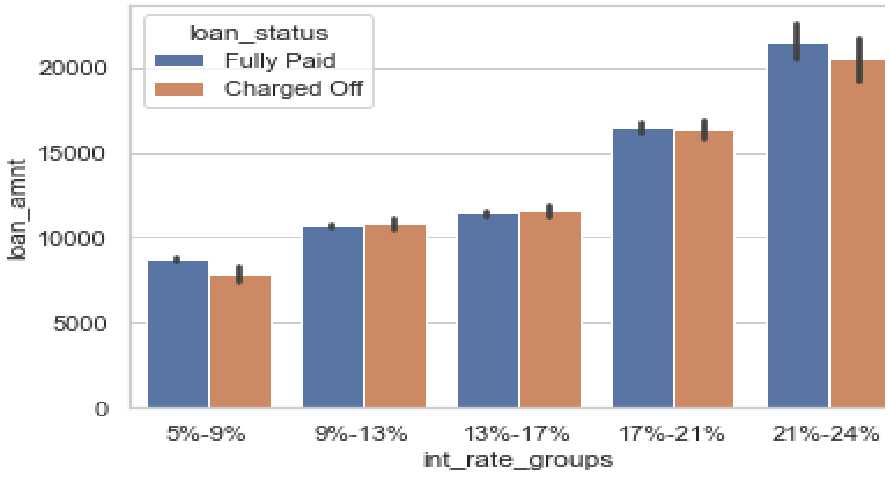
➢ Loans taken for small businesses, credit card, debt consolidation with loan amount greater than 12000 are likely to default

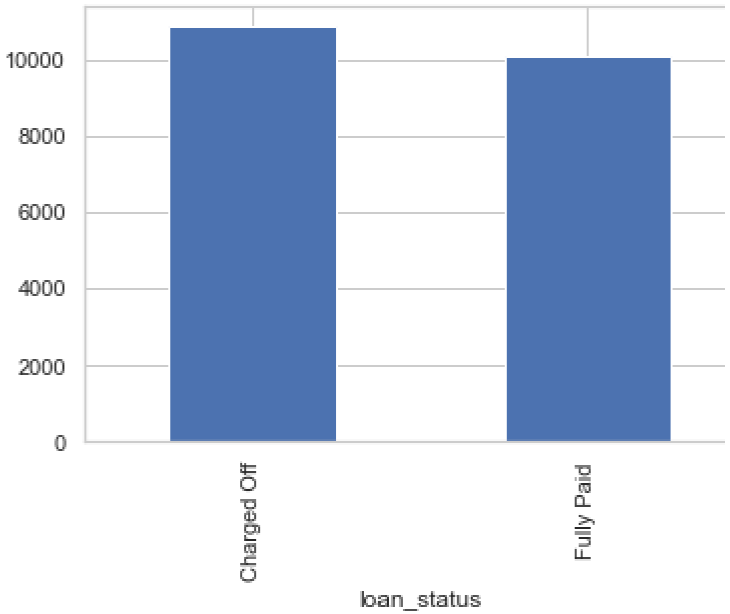
Data Analysis-Bivariate

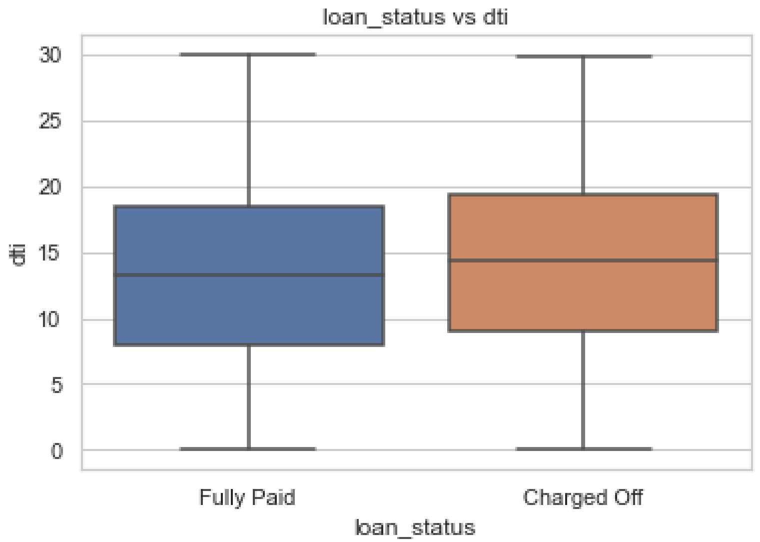


- **Inferences:**
- The Income group 250000-300000 who are applying for a loan of above 20000 are likely to default
- The Income group 350000-400000 applying for a loan below 15000 are less likely to default

Data Analysis-Bivariate

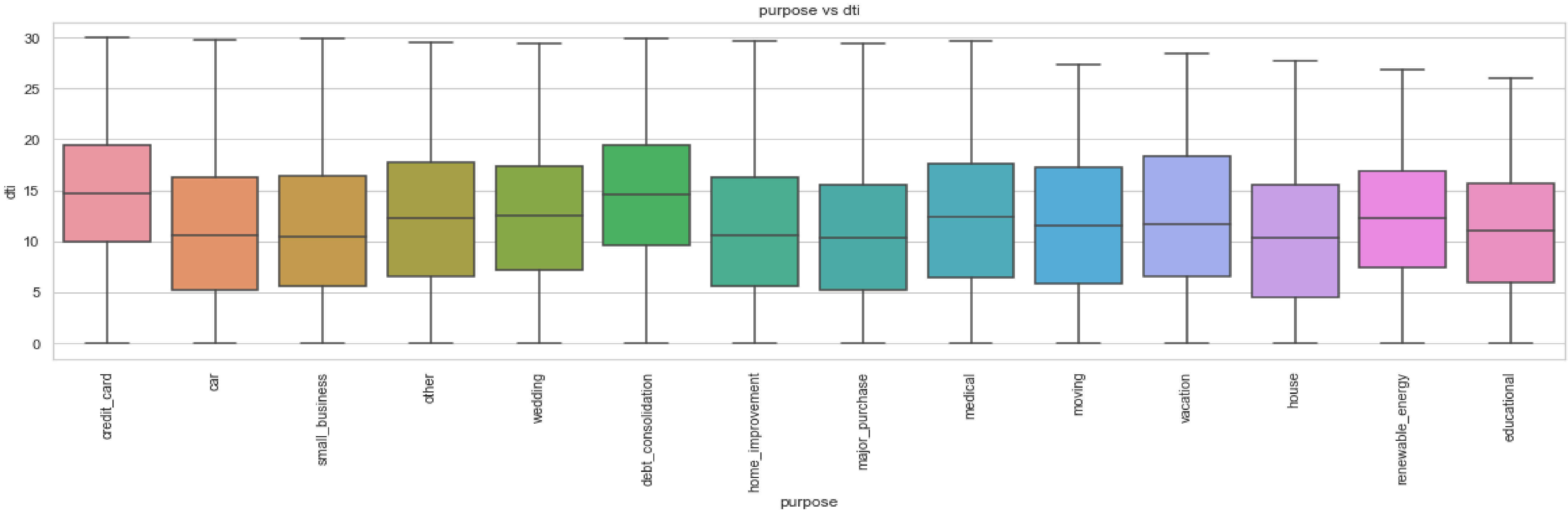






- **Inferences:**
- The applicants who have taken loans with higher interest rates are likely to default. The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups. This can be a pretty strong driving factor for loan defaulting.
- The applicants who have applied for higher loan amounts are likely to be defaulters
- Applicants with high DTI has bit more probability to default

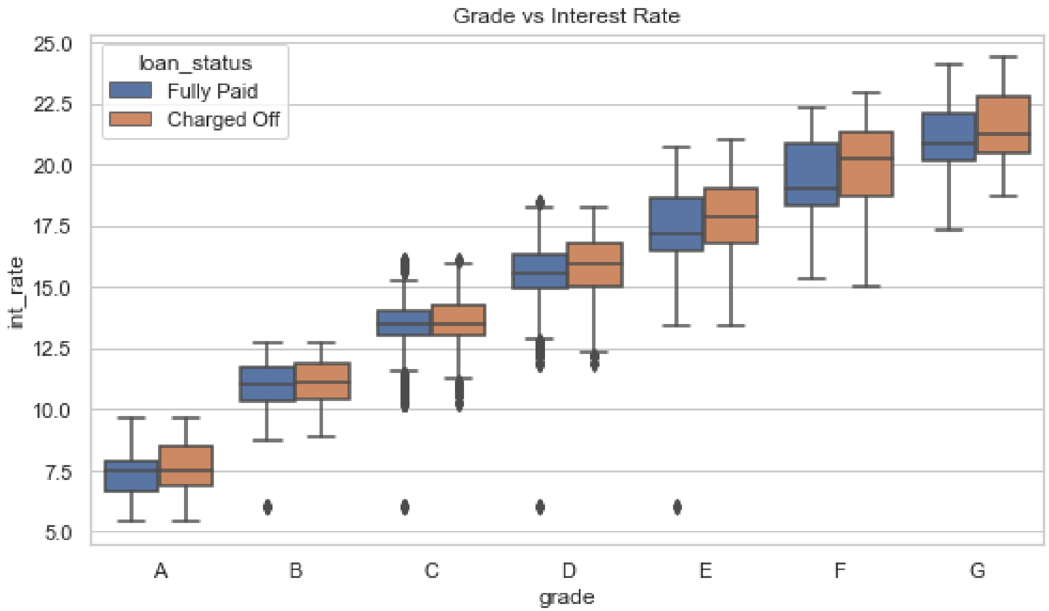
Data Analysis-Bivariate

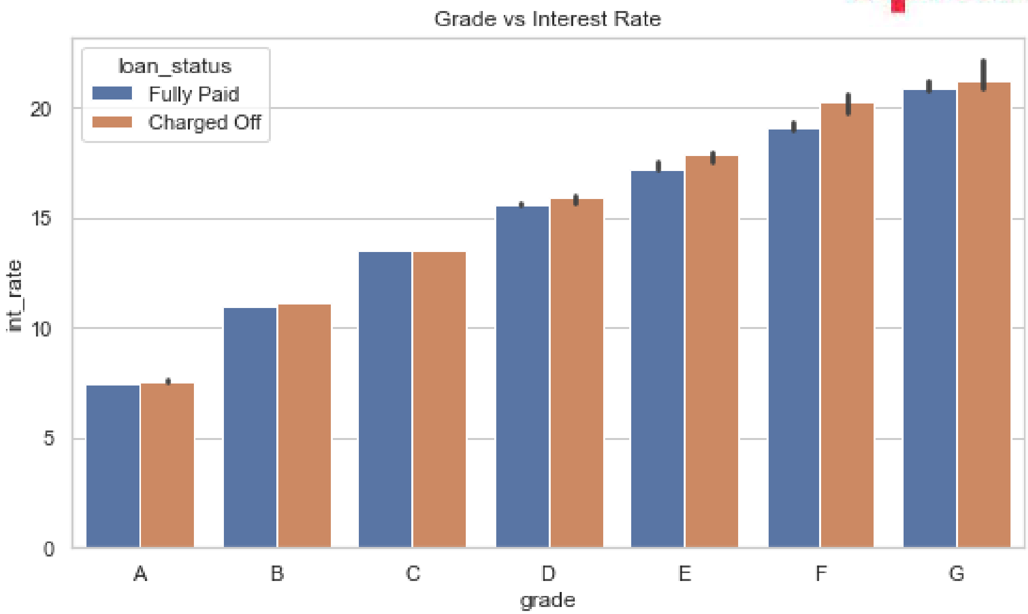


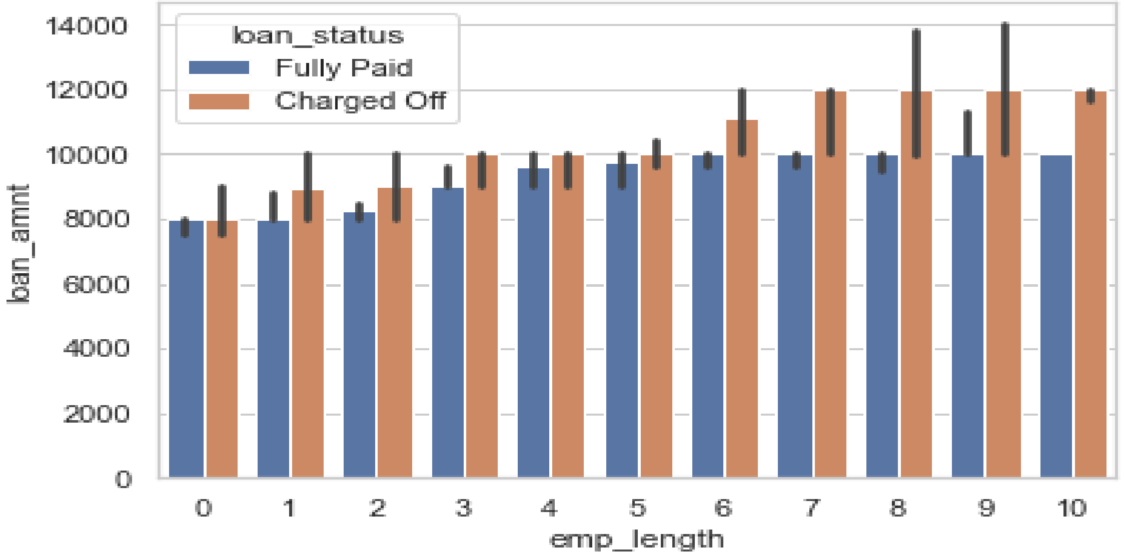
- **Inferences:**

People who took loan for credit card and debt consolidation purpose has more DTI than other purposes

Data Analysis-Bivariate







- **Inferences:**

➢ As grade decreases the interest rate gradually increases. and they are more and more prone to default the loan

➢ Applicants with higher work experience and took more loan amounts are likely to be defaulted

Recommendations:

In the following cases, the applicants are likely to default. So lending club should reduce/stop issuing loans taking into consideration the following factors:

- As grade decreases from A to G the interest rate gradually increases, and they are more prone to default the loan
- The applicants from CA,FL and NY are highly likely to default
- Applicants with higher work experience and took more loan amounts are likely to be defaulted
- The applicants who have taken loans with higher interest rates are likely to default. The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups.
- Applicants whose home is in mortgage and having work experience of more than 10 years are likely to default
- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default

Conclusion:

- The driving factors deciding loan default as per the data analysis, not limited to the below are:
- Grade of the loan provided to the applicant
- The state where the applicant is from according to the details provided during applying loan
- Higher interest rates of loans
- Shorter tenure of repayment
- Work experience of the applicant

THANK YOU