

LENDING CLUB CASE STUDY SUBMISSION

Group Members:

1.Hanumant Jadhav

2.Vaibhav Patil.

Case Study Objective

Identification of risky applicants using EDA is the aim of this case study.

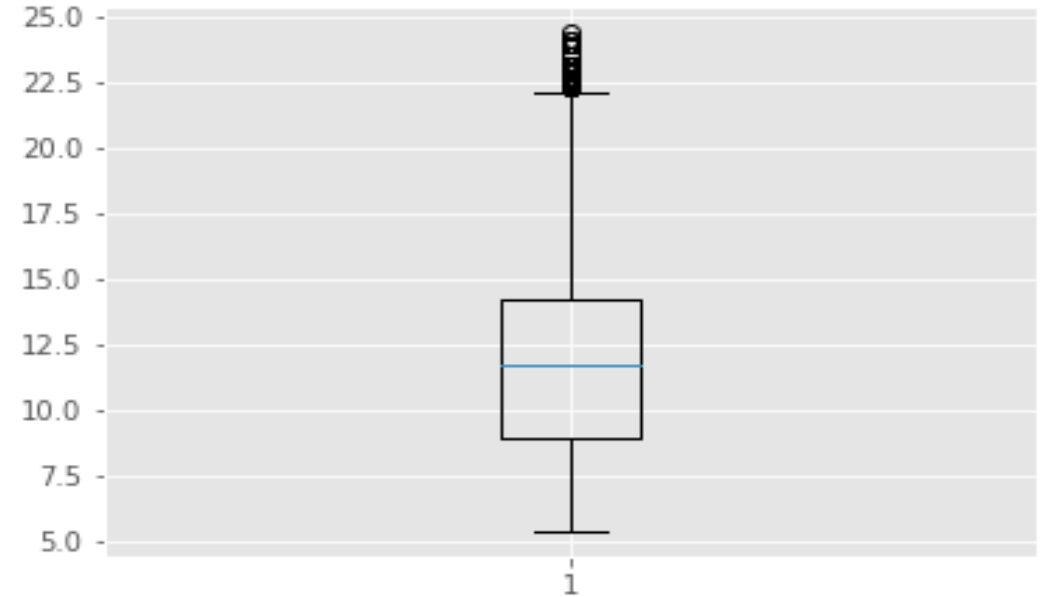
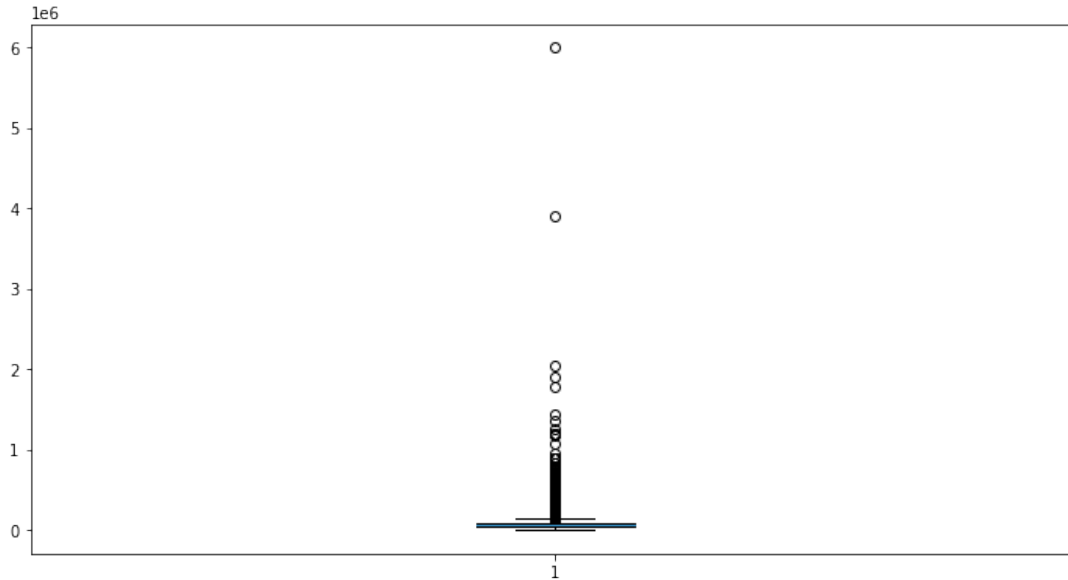
Problem Solving Using EDA

- Understanding Of Dataset.
- Data Cleaning(Imputed the NA values for all the driver variables)
- Data Preparation
- Exploratory Data Analysis .(Performed Univariate, Bivariate, Multivariate analysis)

Steps to be followed in EDA

- Import the required libraries for Exploratory Data Analysis.
- Load the data into the data frame.
- Display the top 5 rows from dataset to find any extra rows like Header row.
- Delete unnecessary columns as we are not able to see complete dataset.
- Check the shape of complete dataset.
- Check whether any columns values are missing in dataset.

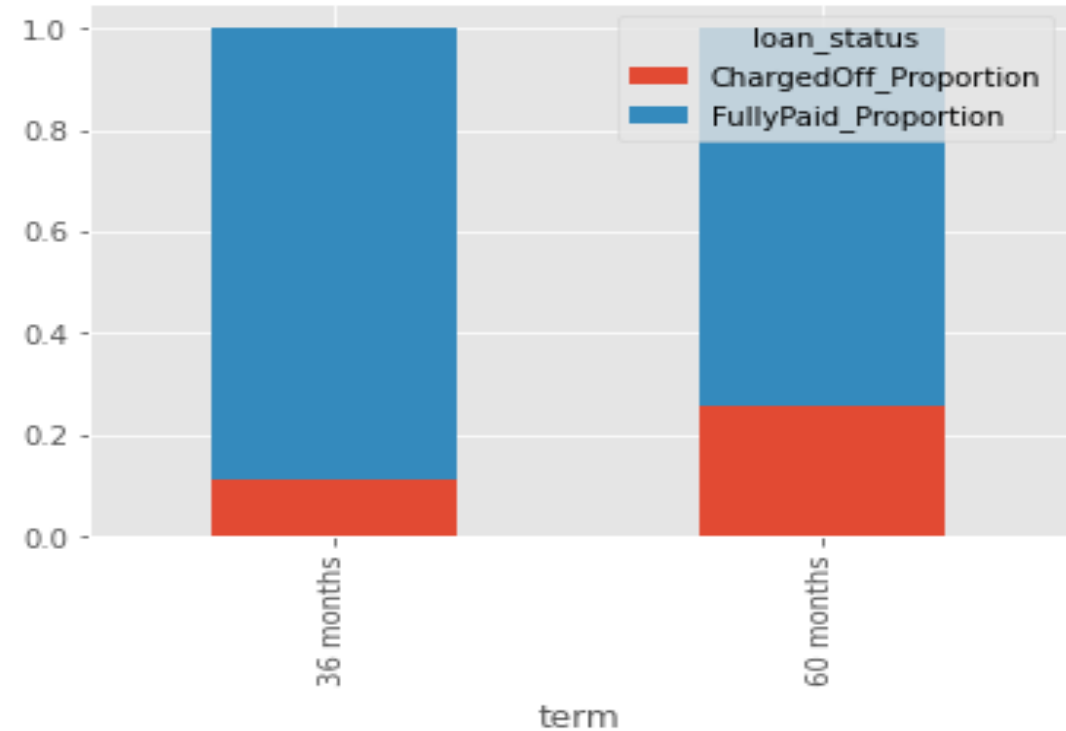
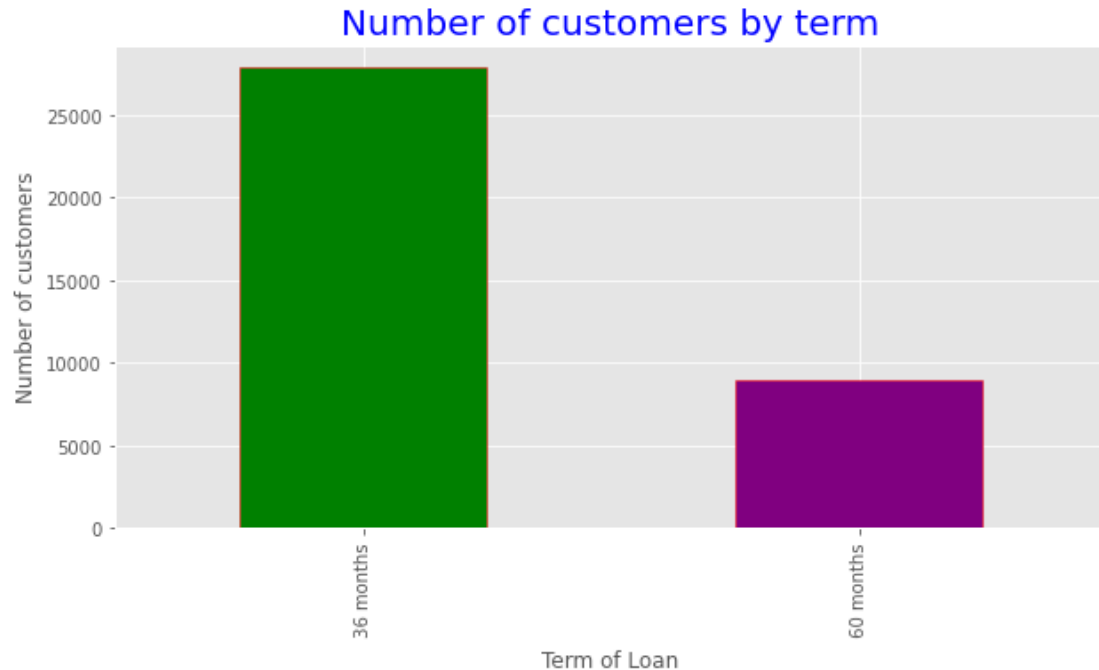
Univariate Analysis : Draw Boxplot for Annual Income



Observation:

- There are outliers for annual income, So we need to remove it.
- It seems the interest rate is having % sign in it. Let's remove it from interest rate column.

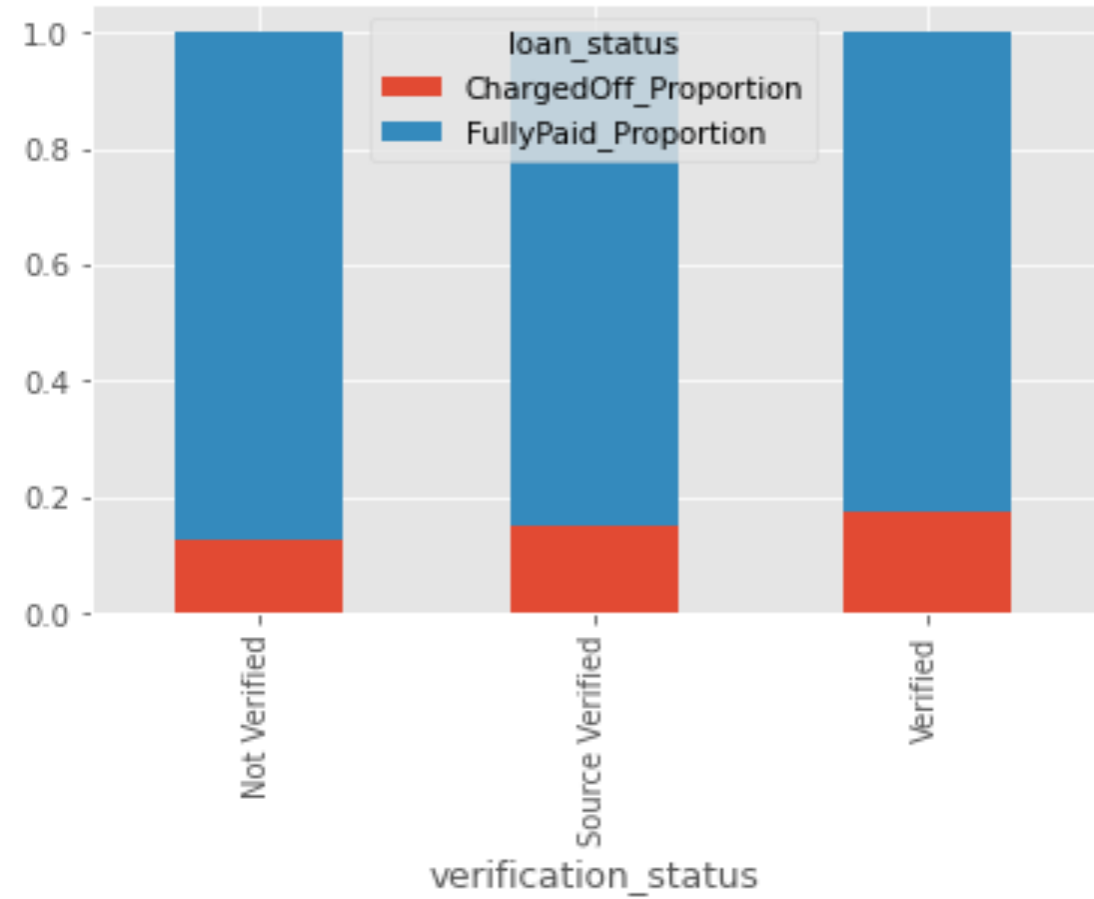
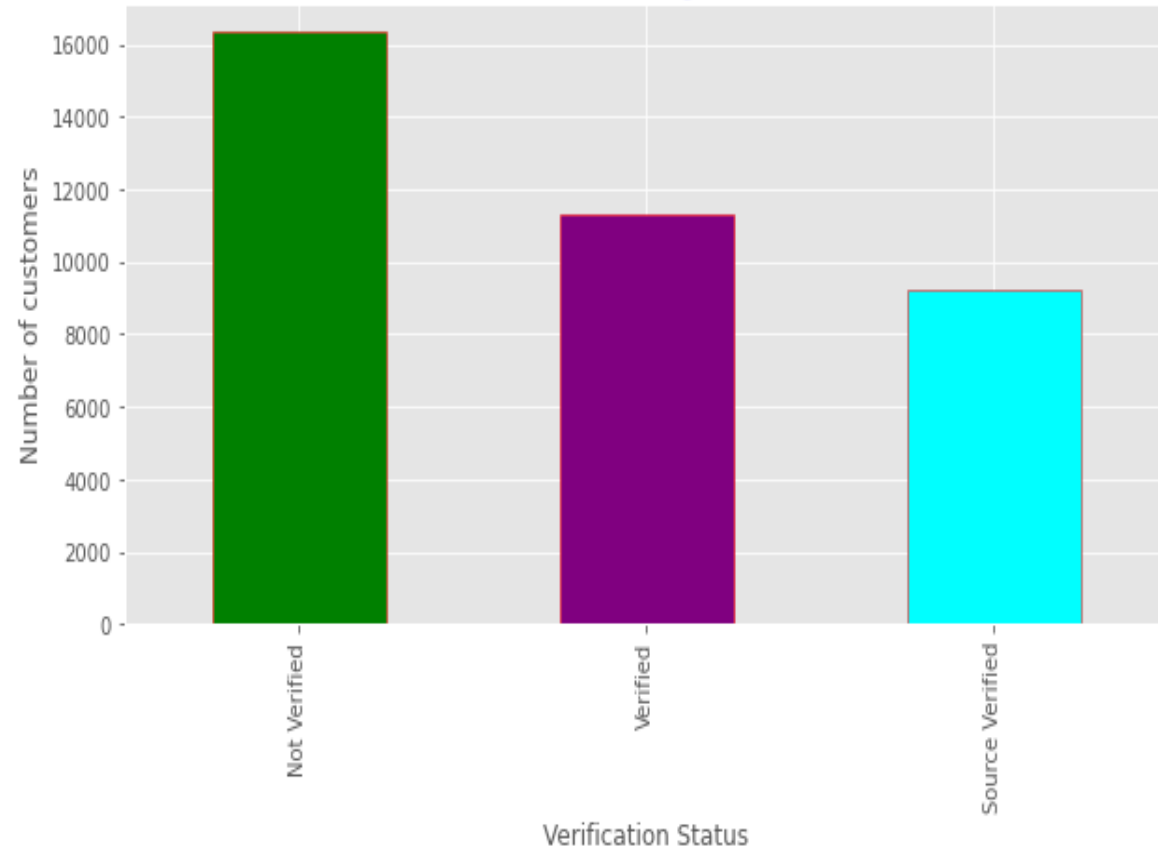
Number of customers by term



Observation:

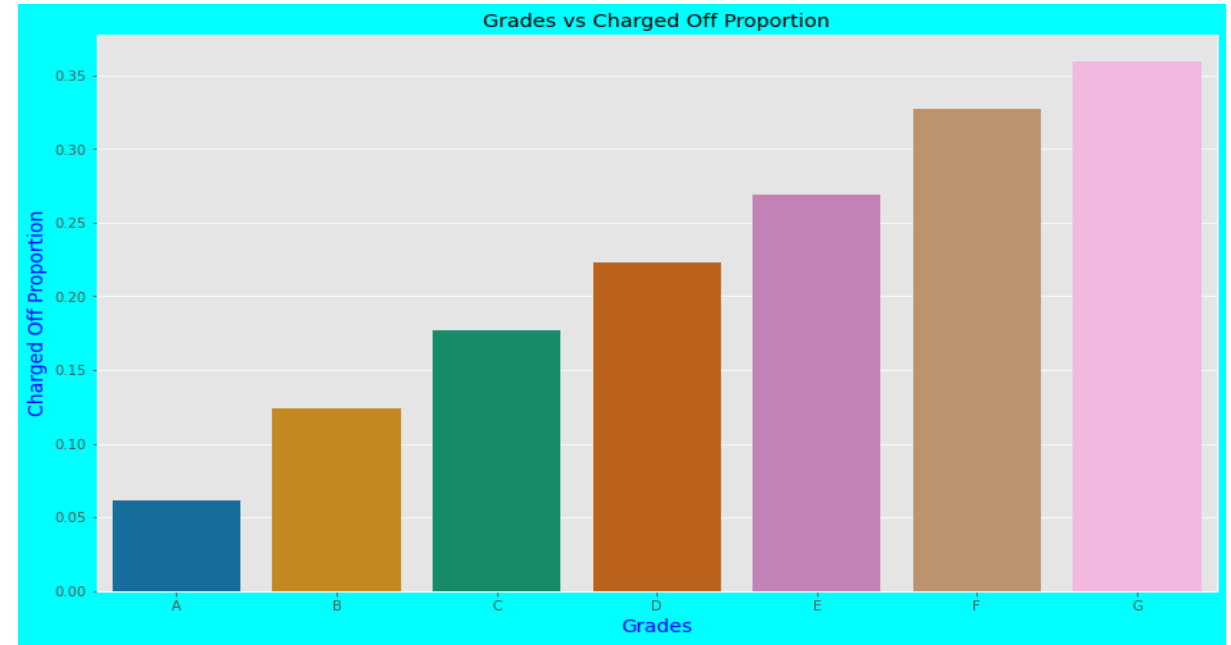
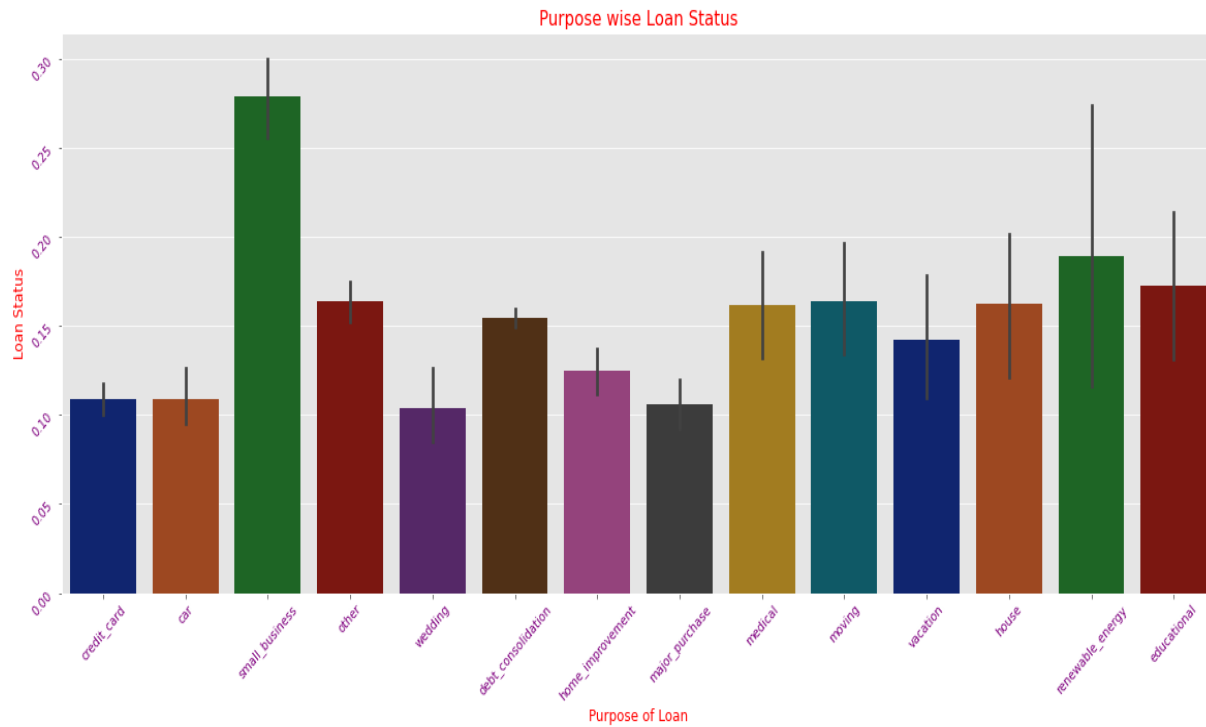
- Most of the customers are preferring 36 months over 60 months.
- Customers who is applying for loan period of 60 months has greater defaulters than customers who is applying for 36 months.

Number of customers by verification status



Observation:

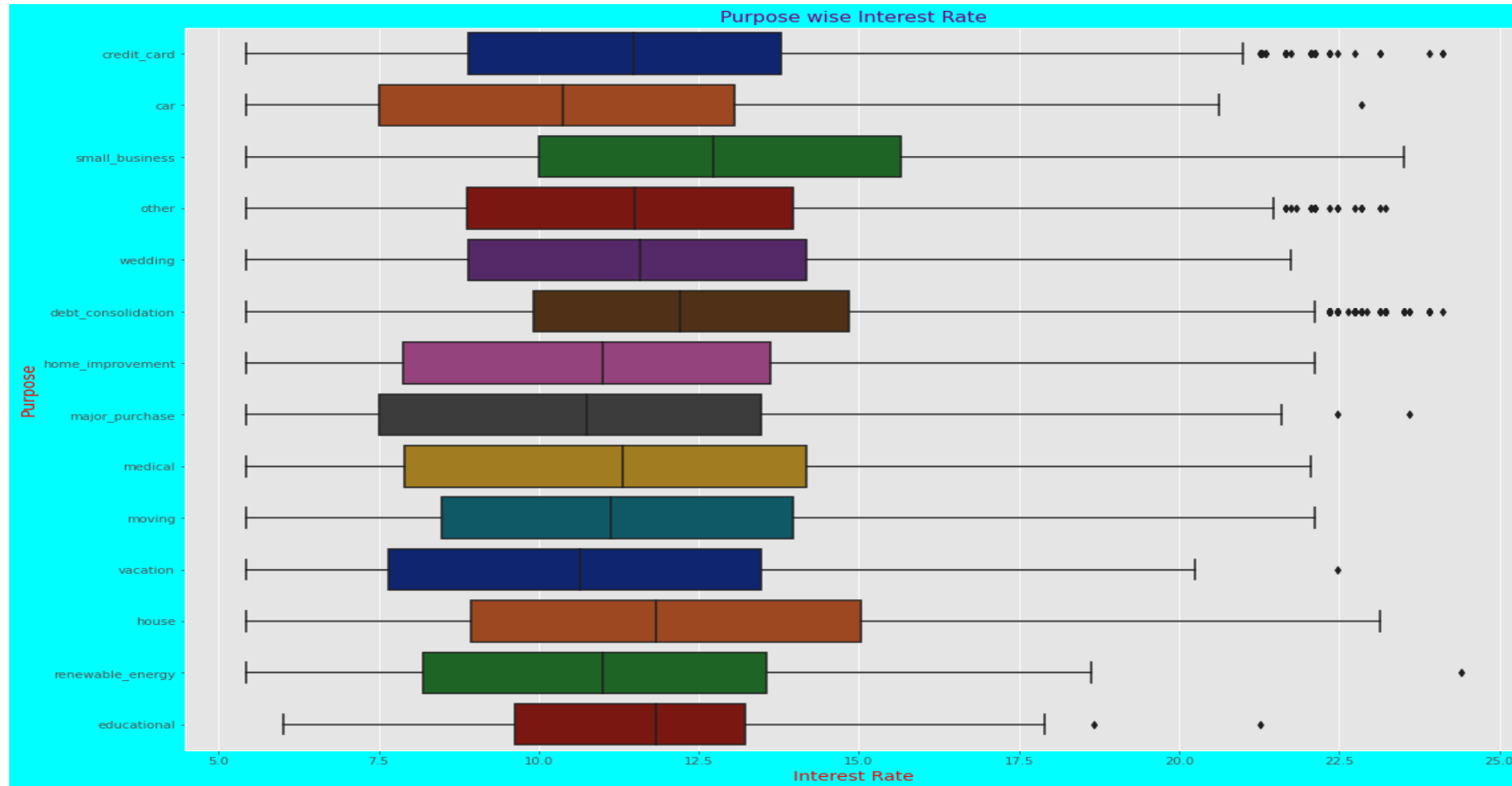
- Most of the customers verification status is Not Verified followed by Verified.
- Verified status is having more defaulters as compared to others. Similarly for Not Verified verification status less defaulters.



Observation:

- From above we can identify that small business purpose is having large number of Charged Off.
- Grade G has highest number of defaulters.
- Defaulters are increasing if we move from grade A to grade G.

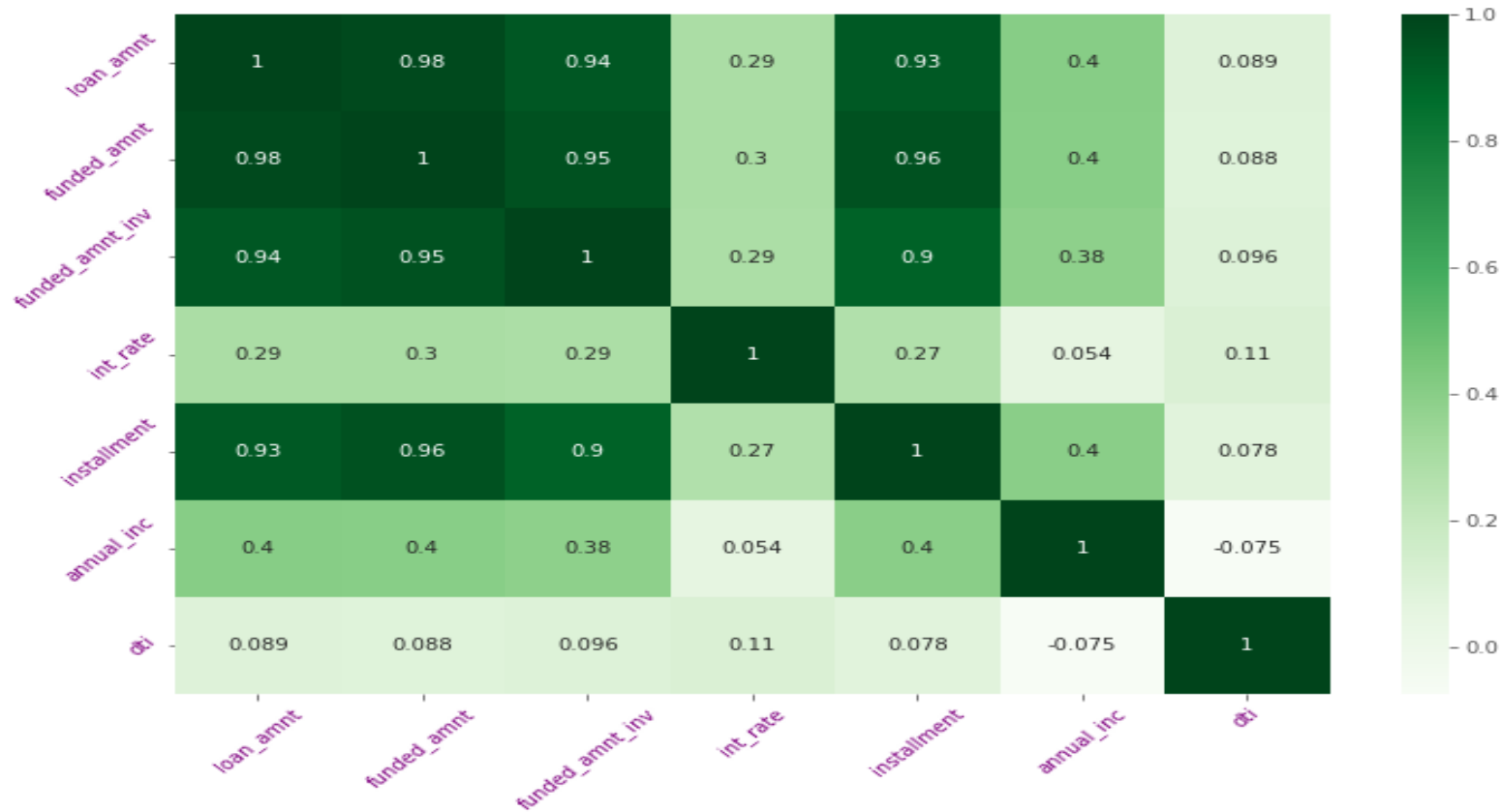
Bivariate Analysis : Purpose wise Interest Rate



Observation:

- Average interest rate for small business is greater than other loan purposes.

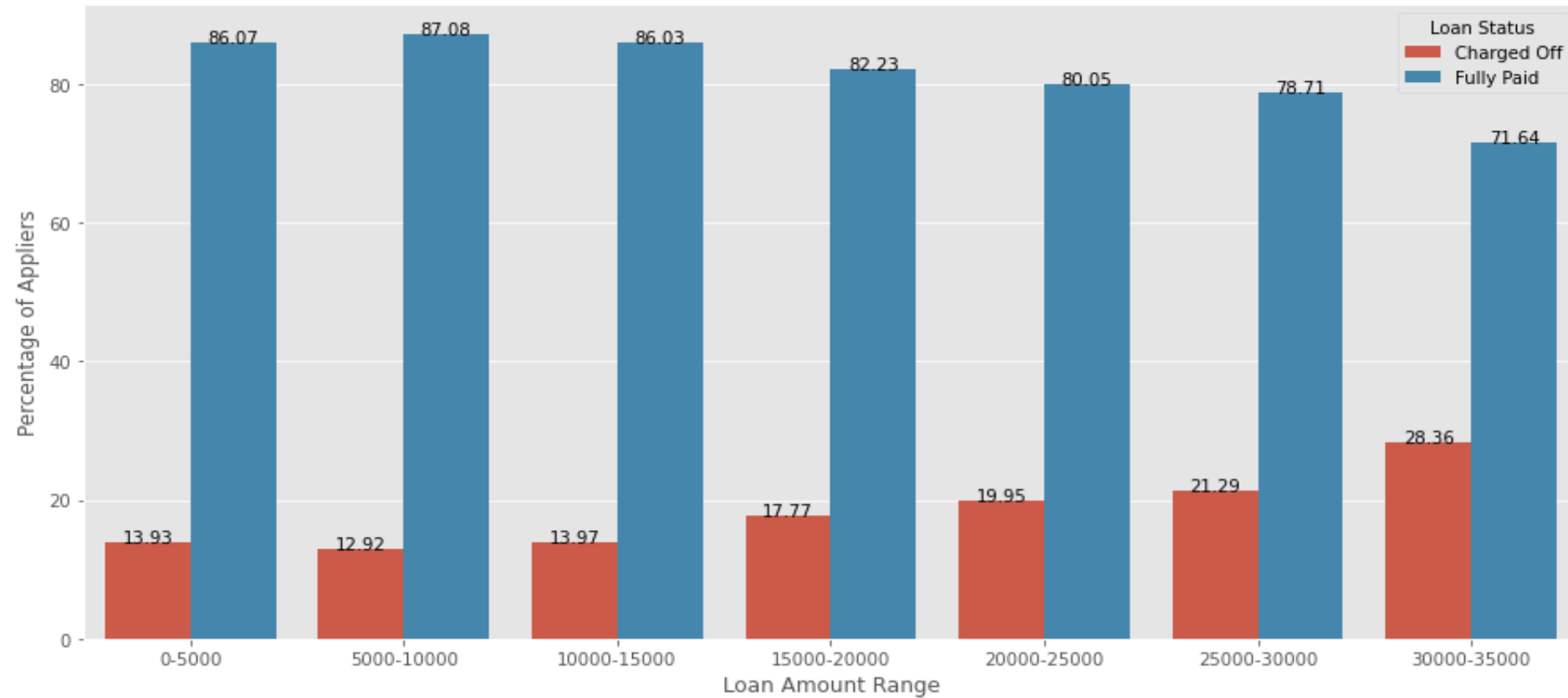
Bivariate Analysis



Observation:

- Annual Income of customer is negatively correlated with Debt to Income Ratio.
- Loan Amount, Investor Amount and funding amount are strongly correlated.

Multivariate Analysis:



Observation:

Applicants with loan amount greater than 15,000 tend to default and with less than 15,000 tend to repay