LENDING CLUB CASE STUDY
SUBMISSION

Group Members: ARCHA S A & GOPIKA GS

Project Brief

- The main objective of the project is to analyze and understand the data about the loan applicants of a consumer finance company and identify the patterns which indicate whether a person is likely to default or not, so as to help the company in taking a suitable action regarding a loan application.

- The business objective is to identify the features that influence the tendency to default.

- Datasets used:
  - loan.csv – List of past loan applicants with their details.

Data Analysis Flow

Step 1 – Data Cleaning– The dataset is analyzed and the columns with more than 80% null values are dropped. Also the columns that are not useful for our analysis are also dropped.
The rows having missing values are also removed.

Step 2 – Derived Metrics – Additional columns that are useful for analysis is created like loan_year, loan_month and buckets for numerical values like int_rate, loan_amnt, etc are created.

Step 3 - Univariate Analysis – Univariate Analysis is carried out and the columns that are not useful are removed.

Step 4 – Segmented Univariate Analysis – The columns are analyzed with respect to the loan status.

Step 5 – Bivariate Analysis – The columns are compared with other columns with respective to the loan status.

Data Cleaning

- The data is checked for null values and the columns having more than 80% null values are dropped.

- Columns that are having a single unique value is dropped as it is not useful to our analysis.

- The columns that are not useful to our analysis are also dropped.

- The columns having redundant data are also dropped.

- Since we are interested in only the Fully Paid & Charged Off loans, we have removed the loan applications having a loan status of Current.
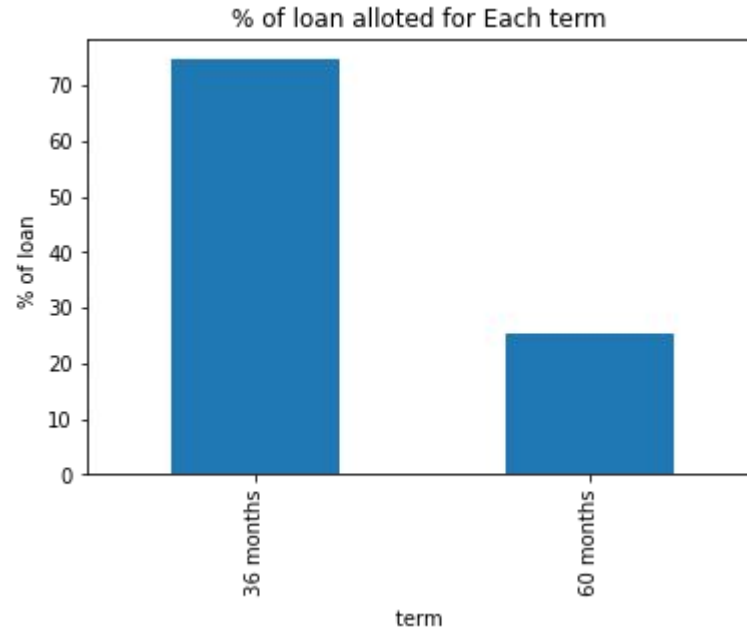
Univariate Analysis

Analyzing loan_amnt w.r.t loan_status
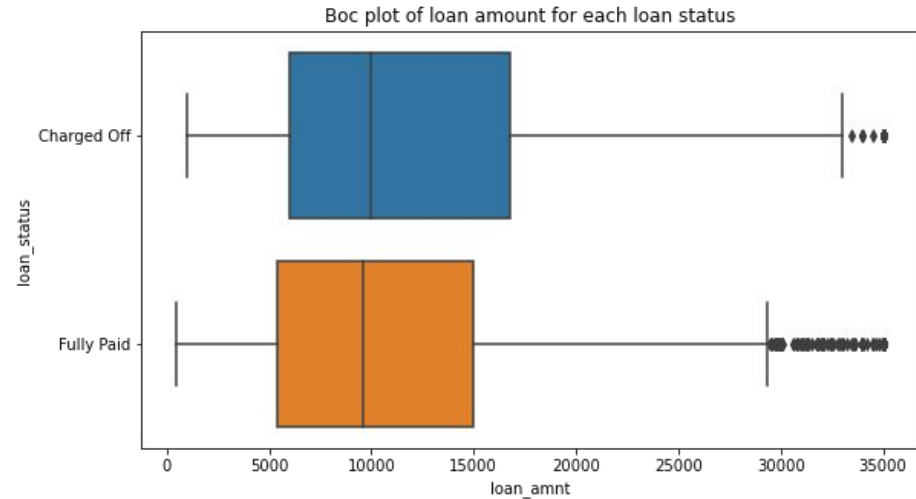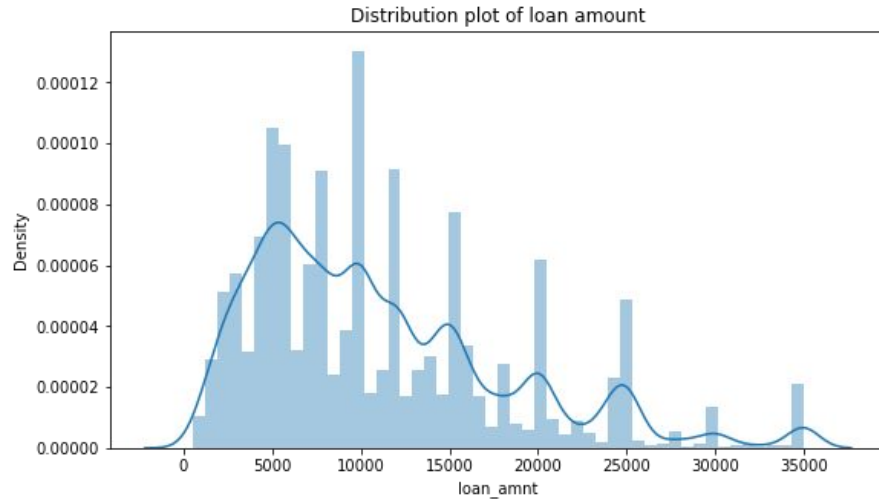
Observation :
The loan amount mostly lies within 5000-15000 and the distribution is clustered towards right with mean greater than median. Charged off loan has a higher mean than fully paid loan
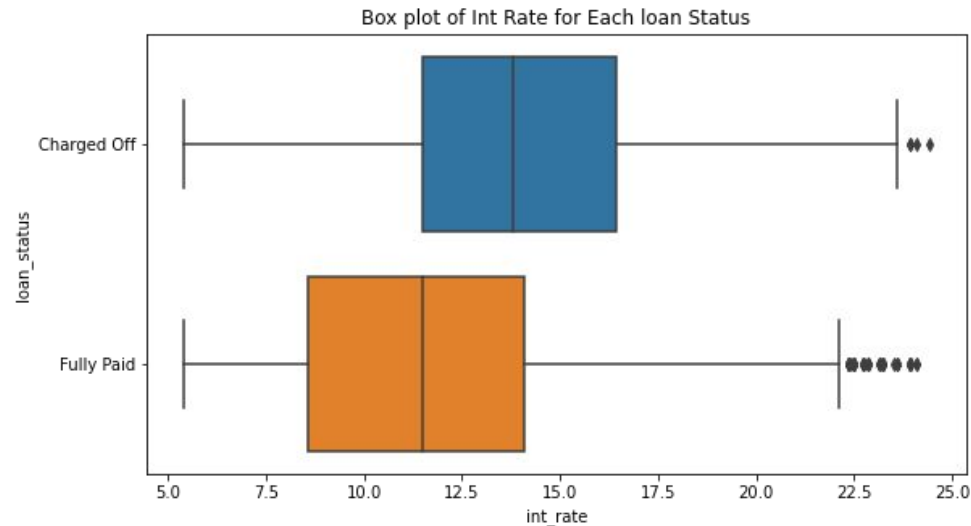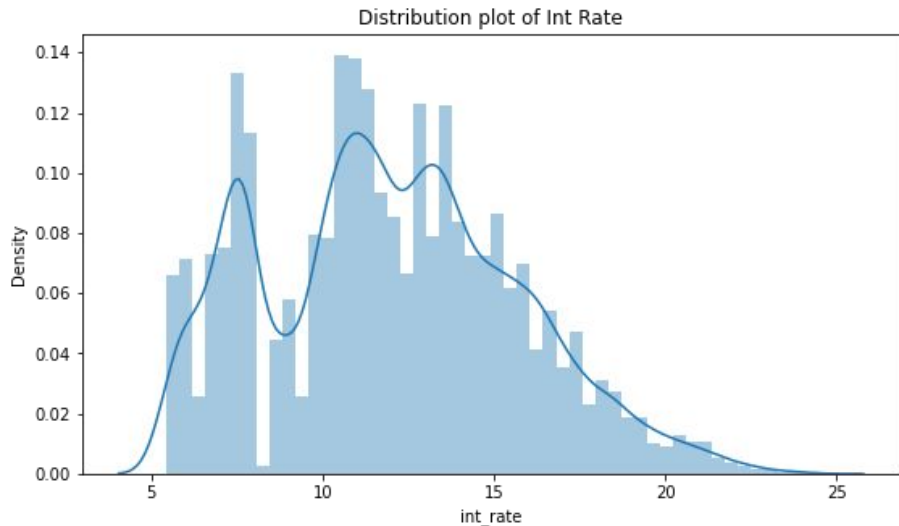
Analysing term

% of loan alloted for Each term

Observation :
More than 70% of loan are given for 36 months

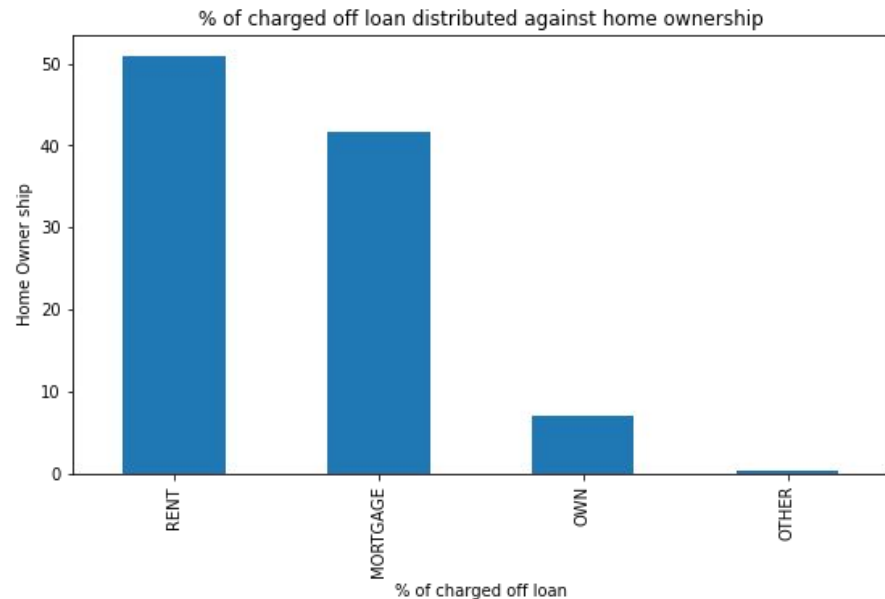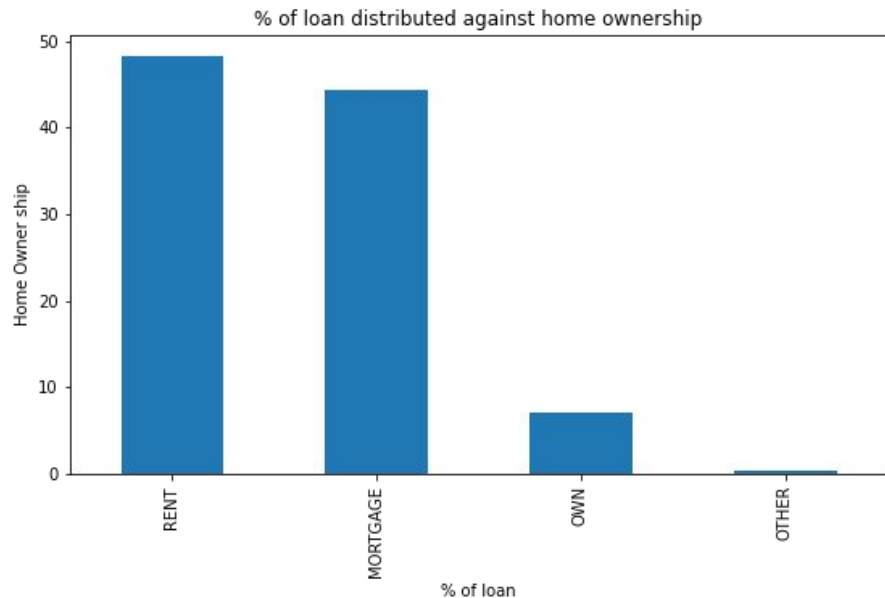## Analyzing interest rate w.r.t loan status

Observation :
Charged off loans of higher interest rate is higher compared to fully paid.
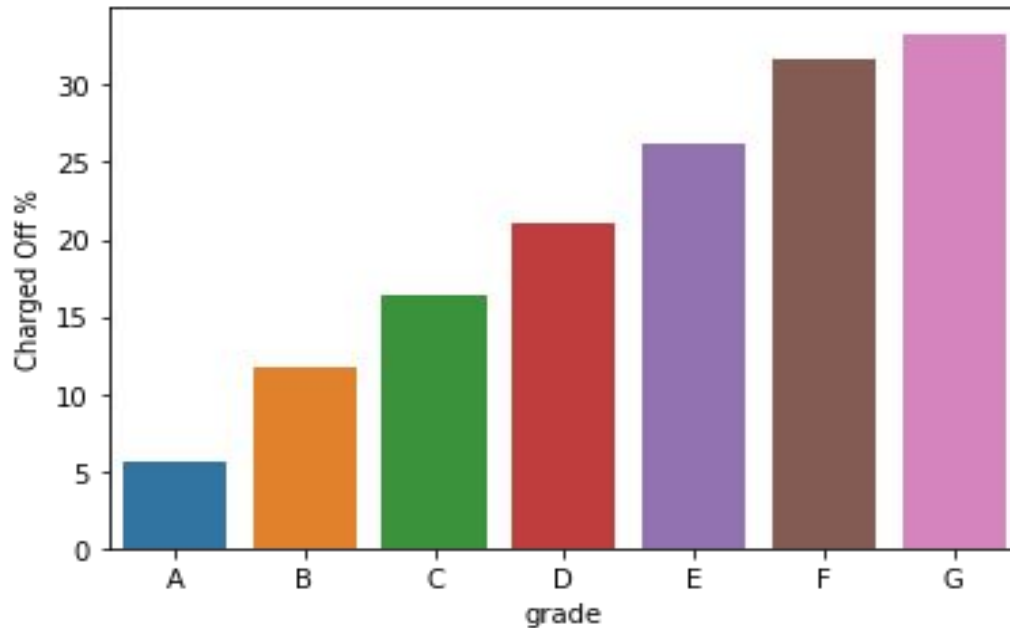
Analyzing home ownership w.r.t loan status

Observation :

Those having rented or mortgaged home are high chances for being a defaulter.
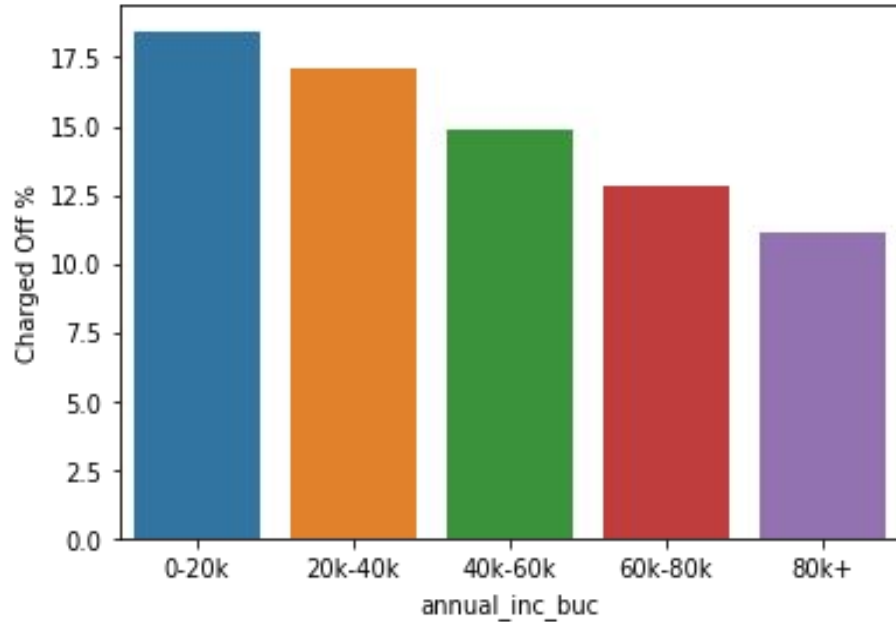
Bivariate Analysis

Analyzing the employee grade for Fully Paid and Charged Off loans

Observation:
Here it can be observed that the borrowers in the grade G & F are more likely to default
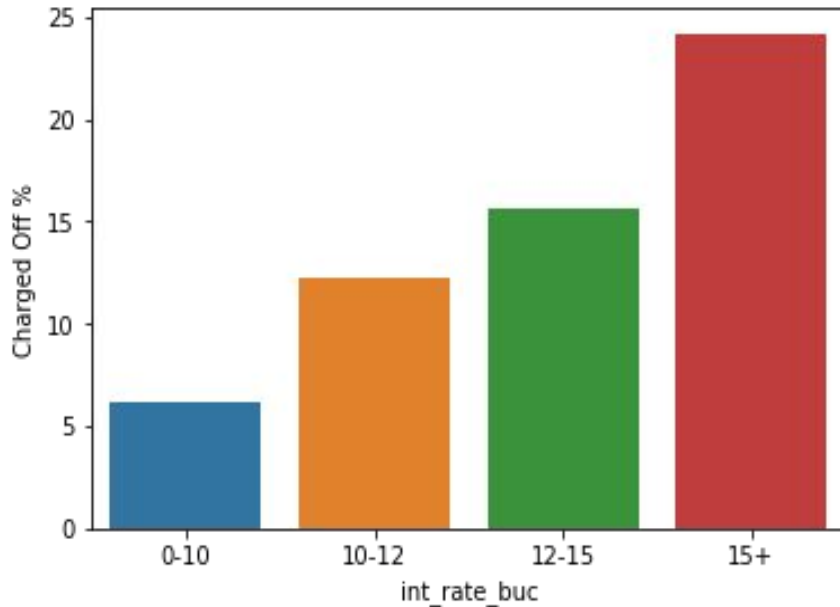The borrowers having Grade A has very less chances for defaulting

Analyzing the annual income for Fully Paid and Charged Off loans

Observation:
- Here it can be observed that the borrowers having income lower than 20k are more likely to default than those with higher incomes.
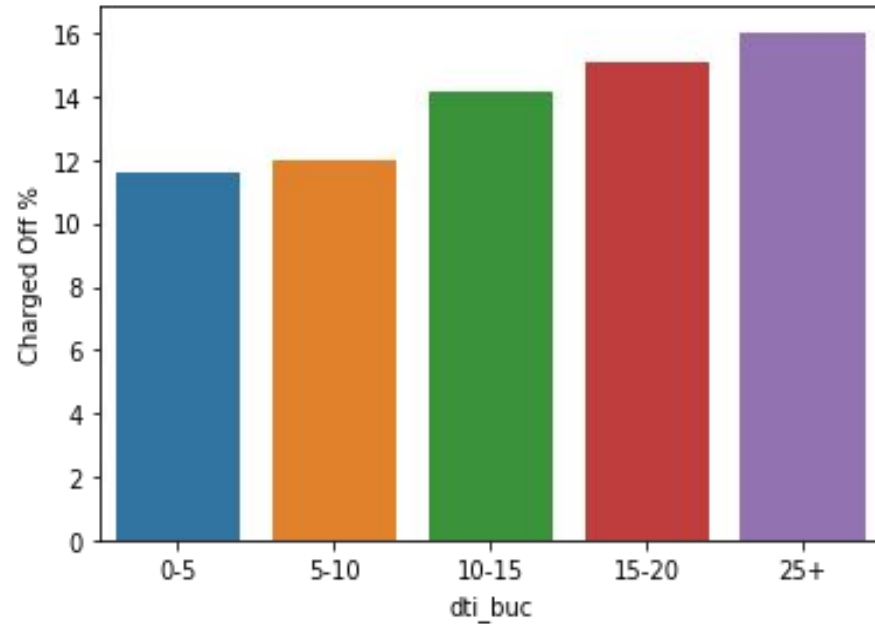- As the income increases, the charge off % decreases

Analyzing the interest rate for Fully Paid and Charged Off loans

Observation:
- Here it can be observed that the loans having interest rate greater than 15% are more likely to default.
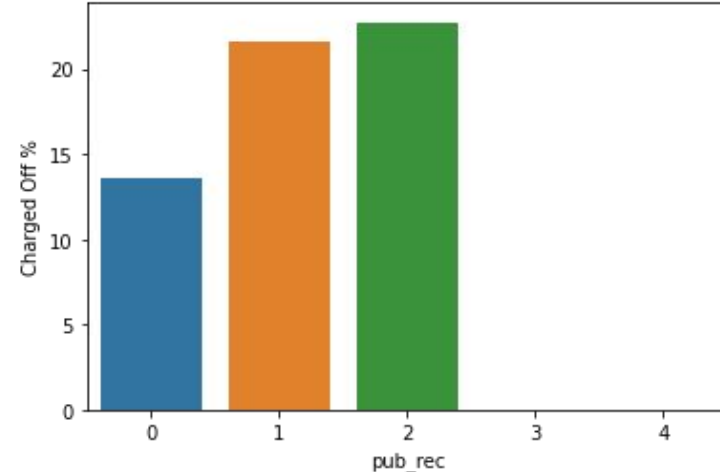- As the interest rate decreases, the charge off % decreases.
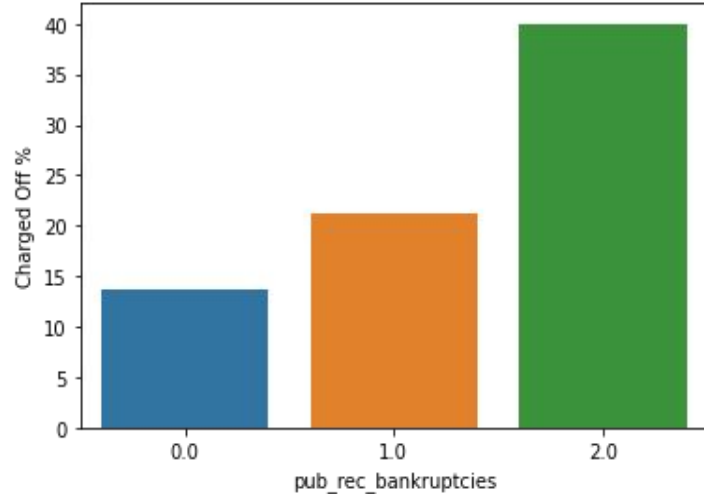
Analyzing DTI for Fully Paid and Charged Off loans

Observation:
- Here it can be observed that the borrowers having higher DTI more likely to default
- than those with lower DTI
- As the DTI decreases, the charge off % decreases

Analyzing publically recorded bankruptcies and derogatories for Fully Paid and Charged Off loans

Observation:
- Here it can be observed that the borrowers having publically recorded derogatories or previous bankruptcies are more likely to default than those who doesn't.

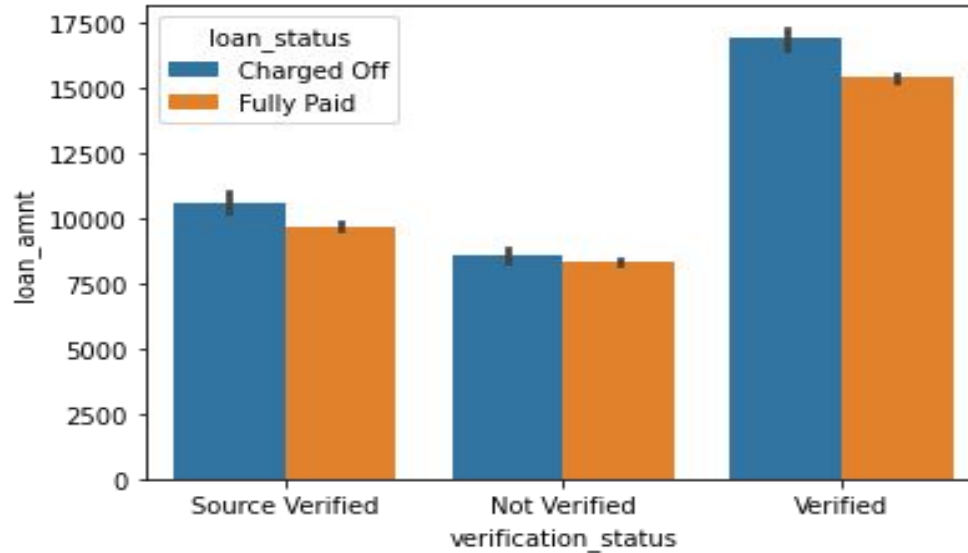# Analyzing purpose of loan for Fully Paid and Charged Off loans
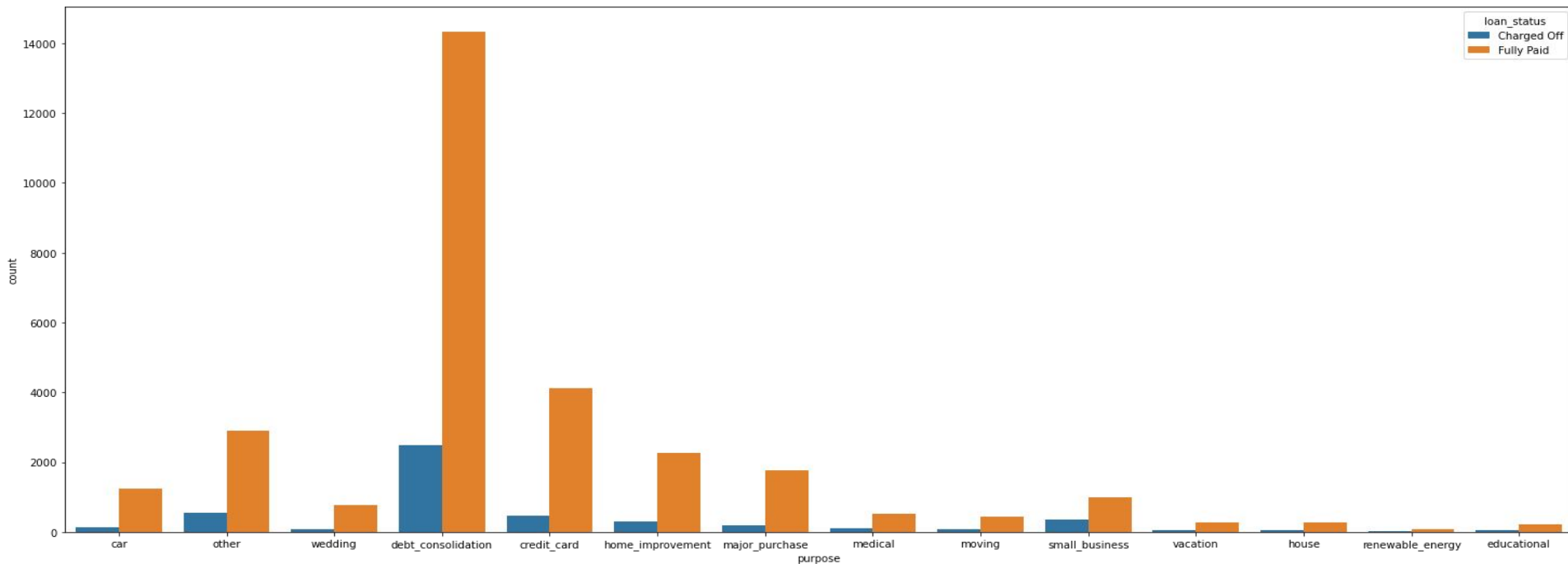
Observation:
- The plot shows that most of the loans where taken for the purpose of Debt Consolidation & Credit Card.
- The number of loans charged off is also high for these 2

Analyzing verification status for Fully Paid and Charged Off loans

Observation:
- As the loan amount increases LC make sure that the loan get verified properly

Conclusions

- On analysis, we can conclude that some of the features that influence the tendency to default are:

  1. The **Grade** of the loan applicant – The applicants in grades **G & F** are having a higher tendency to default.

  2. The **Interest Rate of the loan** – The loans with interest rate higher than 15% are more likely to default.

  3. The **Annual Income of the applicant** – The applicants having an annual income less than 20,000 are more likely to default.

  4. The **DTI** – The applicants having a higher DTI are more likely to default.

  5. The **Publically Recorded Bankruptcies and Derogatories** – The applicants who are having any publically recorded bankruptcies or derogatories are more likely to default.

  6. The **Home ownership** of the applicant - Those having rented or mortgaged home are high chances for being a defaulter.