LENDING CLUB CASE STUDY

Problem Statement

A **consumer finance company** 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. lending loans to 'risky' applicants is the largest source of financial loss for the company. If these risky loan applicants can be identified, then such loans can be reduced thereby cutting down the amount of credit loss.

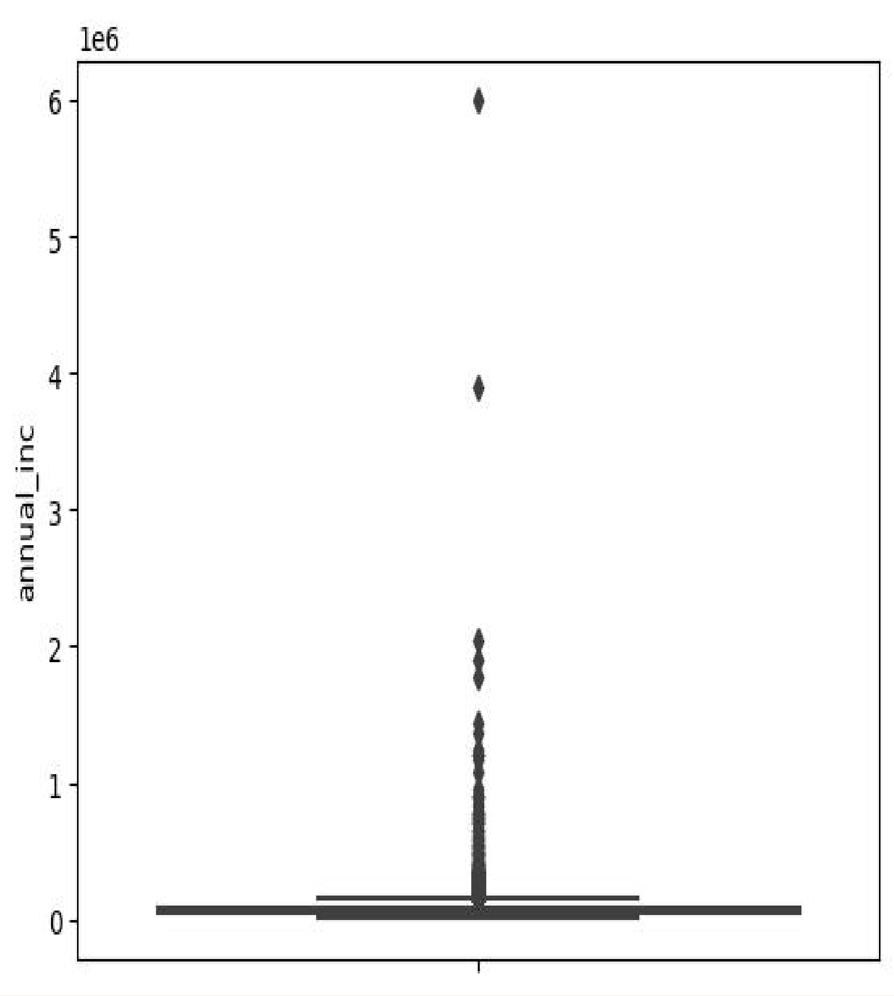
Given information about past loan applicants and whether they 'defaulted' or not, the aim is to identify patterns and drivers which indicate if a person is likely to default.

Data understanding and data cleaning

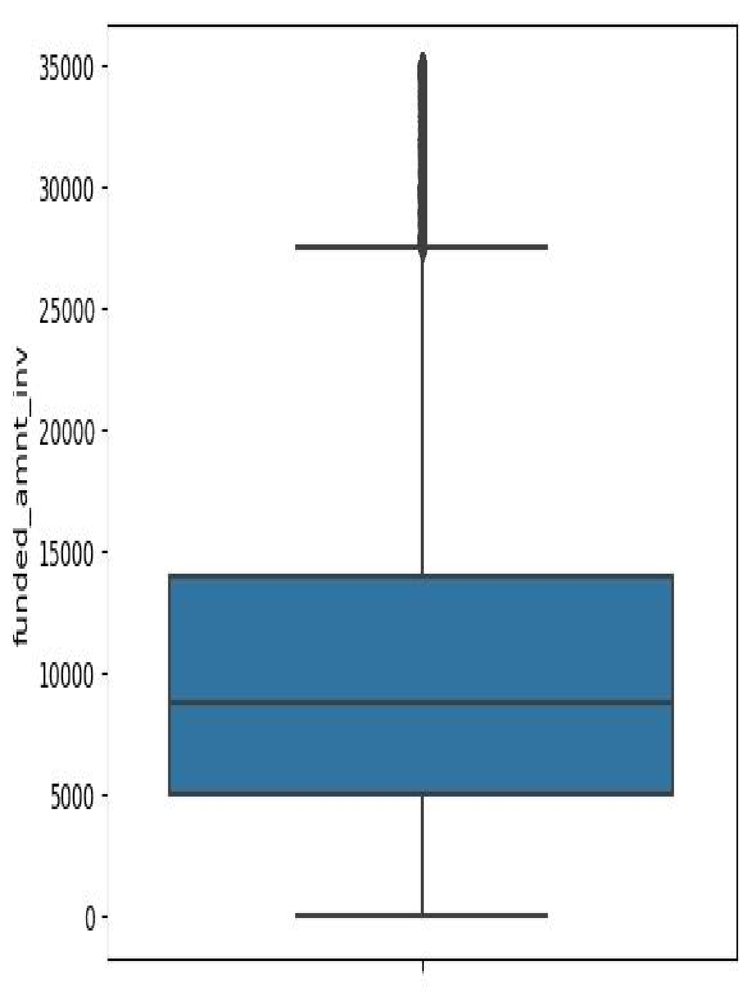
- The dataset provided with the case study contains information about past loan applicants and whether they ‘defaulted’ or not.
- Filter the given dataset to contain just the scenarios where the loan is approved and is either fully paid or defaulted. Our entire analysis will be done on this dataset
- Filter out the customer behavior variables and focus only on the applicant and the loan characteristics

Univariate analysis

- In the given dataset, 32847 loans were fully paid and 5580 loans were defaulted
- Annual income – outliers beyond 99%
- Interest Rate – steady spread, no outliers, mean of 11.9%
- Funded_amnt_inv – steady spread. No outliers. 150 records where funded_amnt_inv < 1

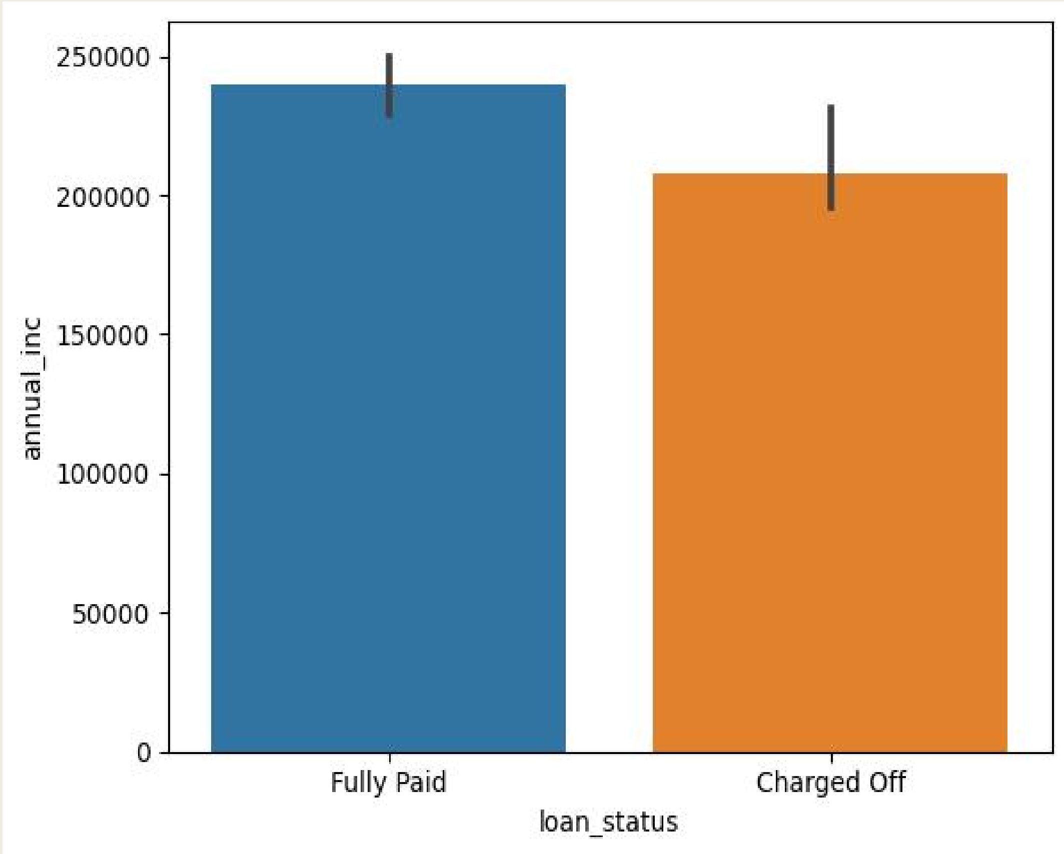


Univariate – annual_inc

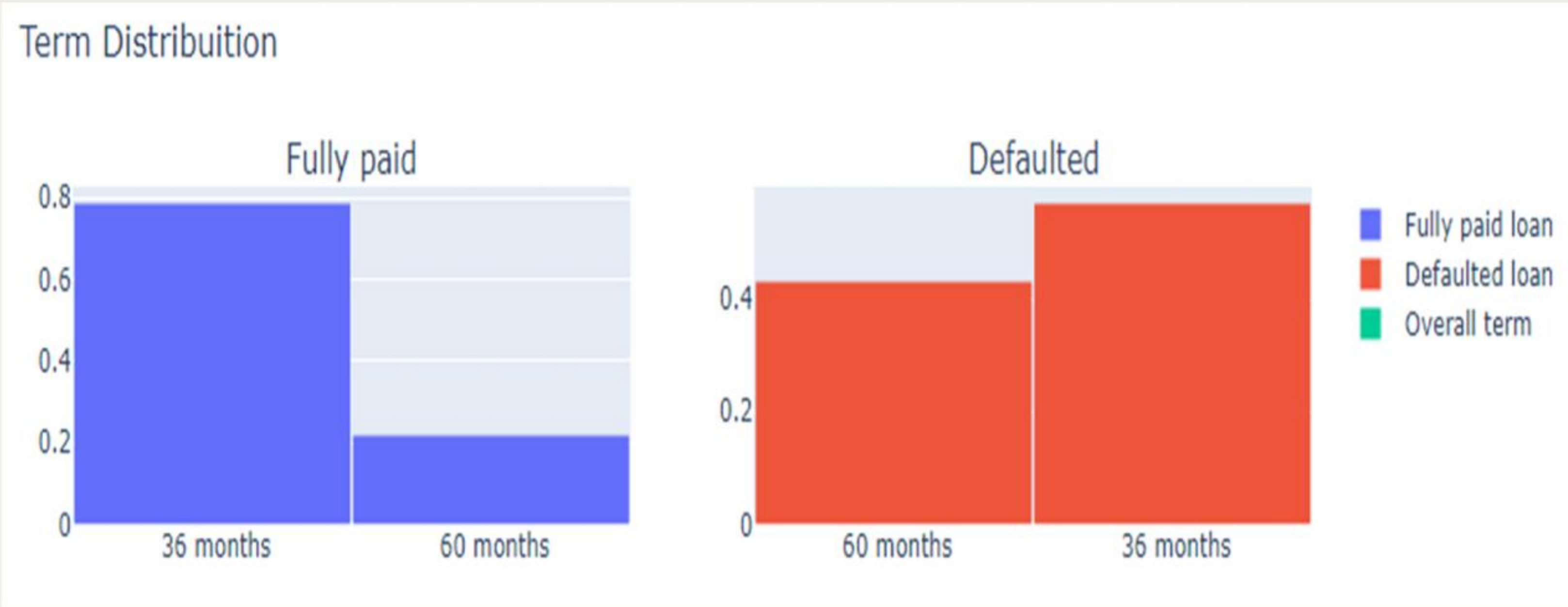


Univariate – funded_amnt_inv

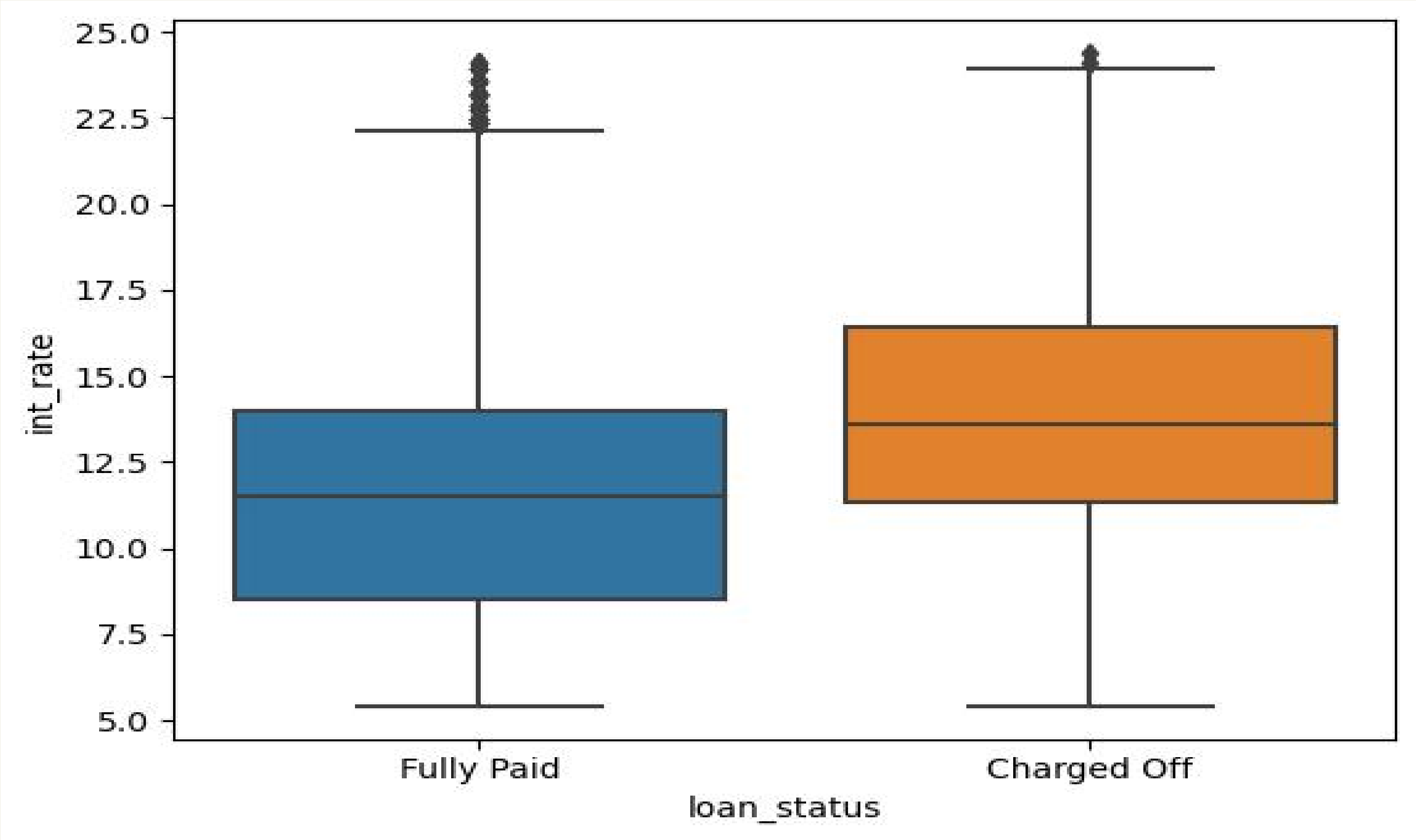
Bivariate analysis – Annual income vs Loan Status



Bivariate analysis – Loan term vs loan status



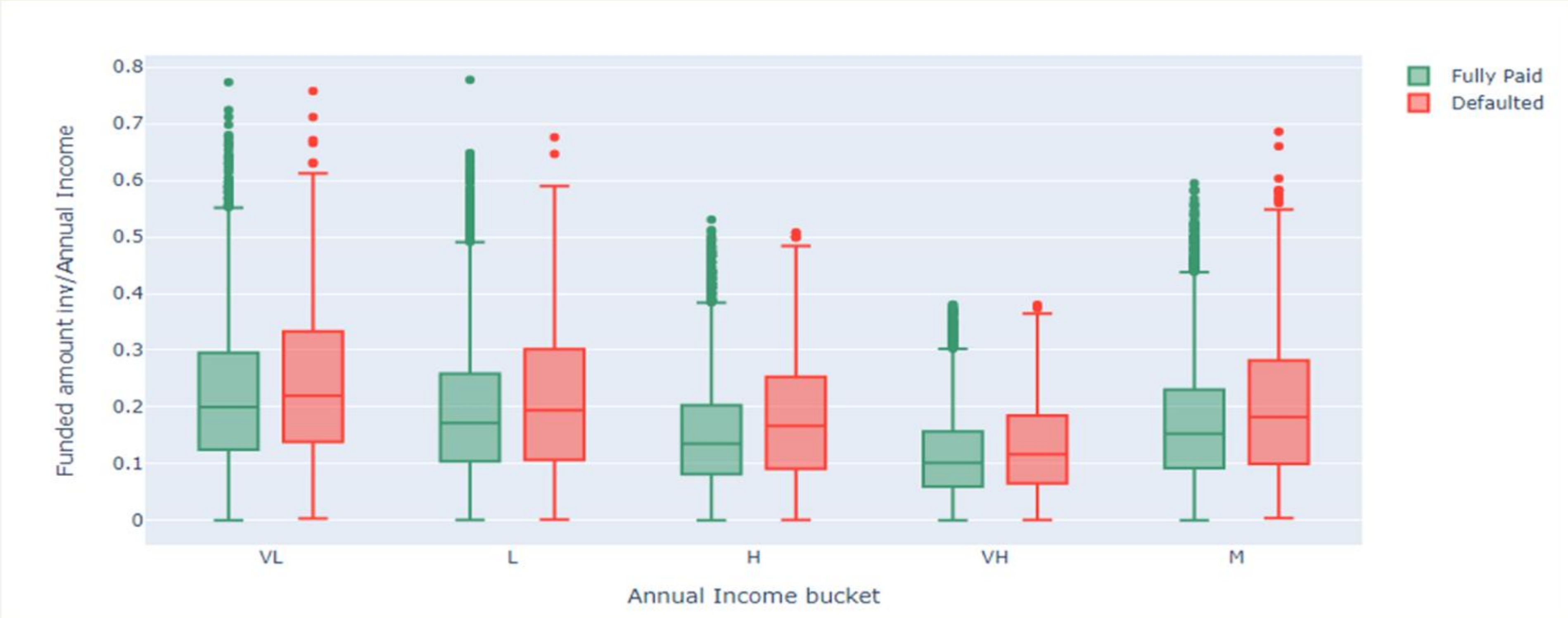
Bivariate analysis – Interest rate vs Loan Status



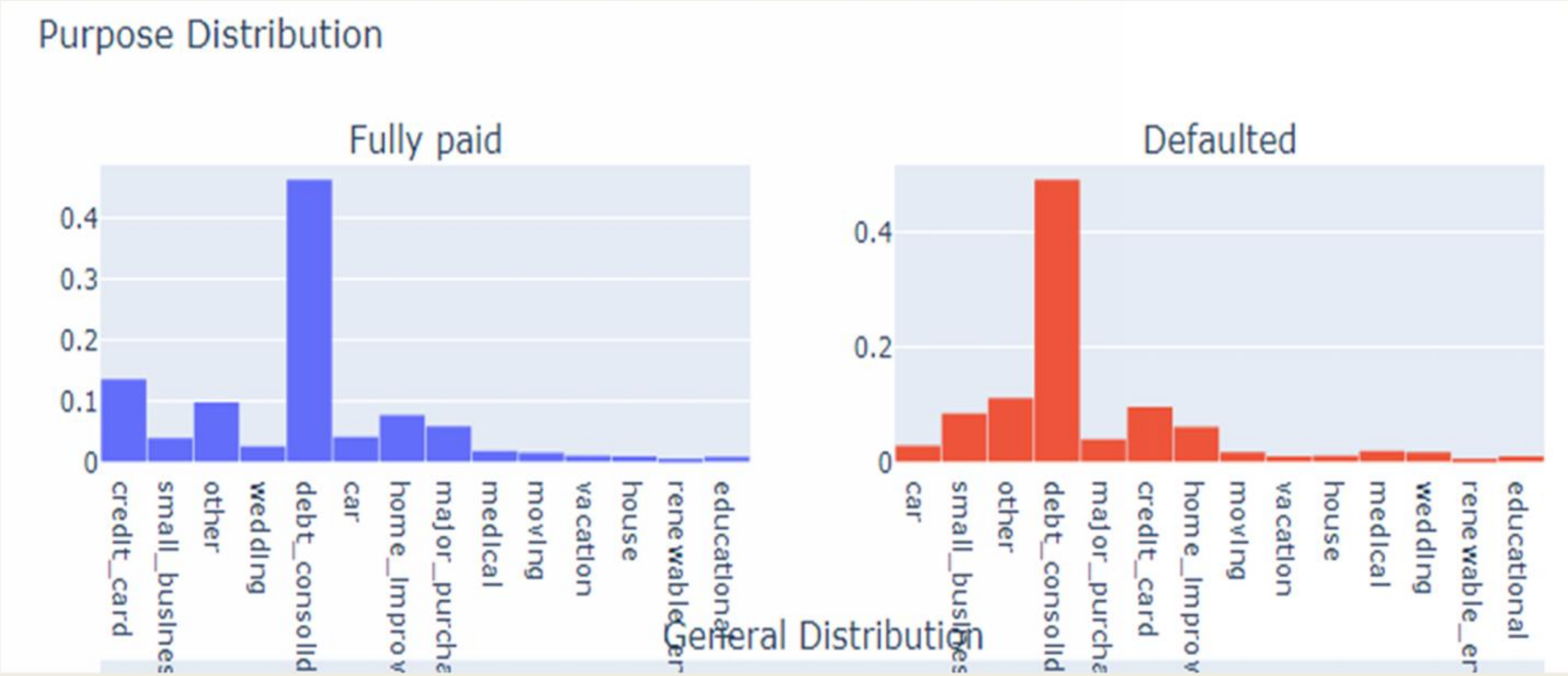
Bivariate analysis – Grade vs Loan Status

Bivariate analysis – Sub-grade vs Loan Status

Bivariate analysis – funded_amnt_inv/annual income vs annual inc



Bivariate analysis – Purpose vs loan status



Recommendations

Based on the EDA, here are the driving factors that indicate if a person is likely to default

- Loan term – Likelihood of default is more with a loan term of 60 months
- Interest rates – Higher interest rates likely lead to more defaults. Fully paid loans tend to have interest rates in the range of 8.5% to 14% .
- Grades and Subgrades –Customers belonging to grade A , (subgrade A1 to A5) are more likely to fully pay the loans. Grade D (Subgrade D2, D3), Grades E(subgrade E1, E2, E4), F (Subgrade F1) and G are more likely to default.
- For any given annual income bucket, if the funded amount is beyond a certain percentage of annual income,, there is more likelihood of defaulted loans. For example, in the medium annual income bracket, if funded amount is greater than 44% of annual income, likelihood of default increases
- Purpose - Customers with small businesses are more likely to default. Loans that were taken for wedding or credit card are more likely to be fully paid