

LENDING CLUB - CASE STUDY

-Tarun Tiwari

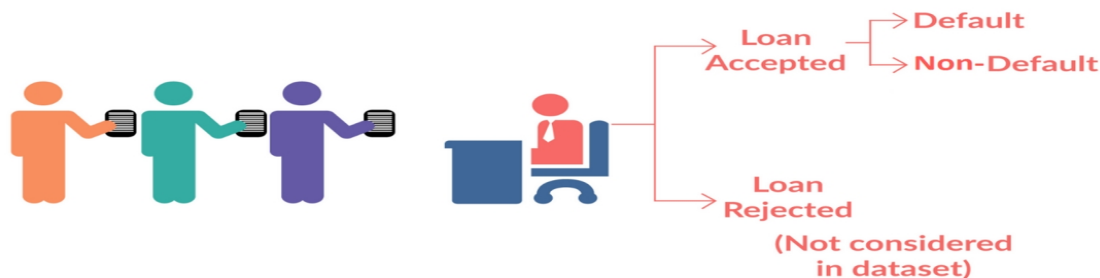
-Pranavi Chapala



Problem Statement

- ▶ You work for a consumer finance company which 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. Two types of risks are associated with the bank's decision:
- ▶ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- ▶ If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

LOAN DATASET



Data Dictionary

- ▶ **loan_amnt:** The listed amount of the loan applied for by the borrower.
- ▶ **Int_rate:** Interest Rate on the loan
- ▶ **Funded_amnt:** Amount recommended/approved by LC
- ▶ **Funded_amnt_inv** The amount funded by investors
- ▶ **Term:** The number of payments on the loan. Values are in months and can be either 36 or 60.
- ▶ **Purpose:** purpose of the loan like home, education, credit card, small business, vacation, medical etc...
- ▶ **Earliest_cr_line:** The month the borrower's earliest reported credit line was opened
- ▶ **Title :** The loan title provided by the borrower
- ▶ **Emp_length:** Employment service/length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- ▶ **Emp_title:** The job title supplied by the Borrower when applying for the loan
- ▶ **Installment:** The monthly payment owed by the borrower if the loan originates.
- ▶ **Grade:** LC assigned loan grade
- ▶ **Home_ownership:** The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are: RENT, OWN, MORTGAGE, OTHER



Problem Solving Methodology-EDA

▶ 1. Data Sourcing

- Importing libraries
- Reading the data into dataframe
- Printing the information and statistics

2. Data Cleaning

- ▶ Checking and dropping null values(missing values)
- Imputing all missing value columns according to there data type
- Checking outlier using box plot
- Removing outliers with Interquartile range

3. Data Derived Metrics

- ▶ Data Processing
- Data Filtering
- Dropping irrelevant columns
- Converting (object) columns into numeric

▶ 4. Univariate Analysis

- ▶ Continuous Variable
- Categorical Variable

▶ 5. Segmented Univariate Analysis

- ▶ Continuous Variable
- Categorical Variable

6. Bivariate Analysis

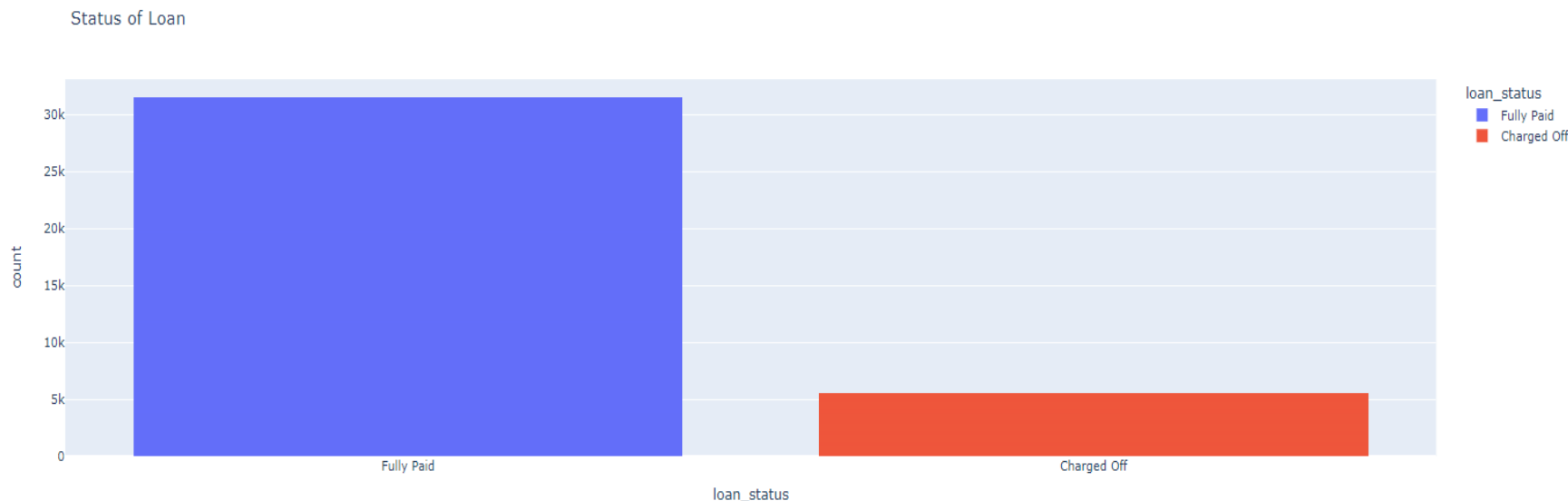
- ▶ Continuous Variable
- Categorical Variable
- Plotting word cloud and Live Frame plotting

▶ 7. Multivariate Analysis

- ▶ Correlation Matrics using heat map



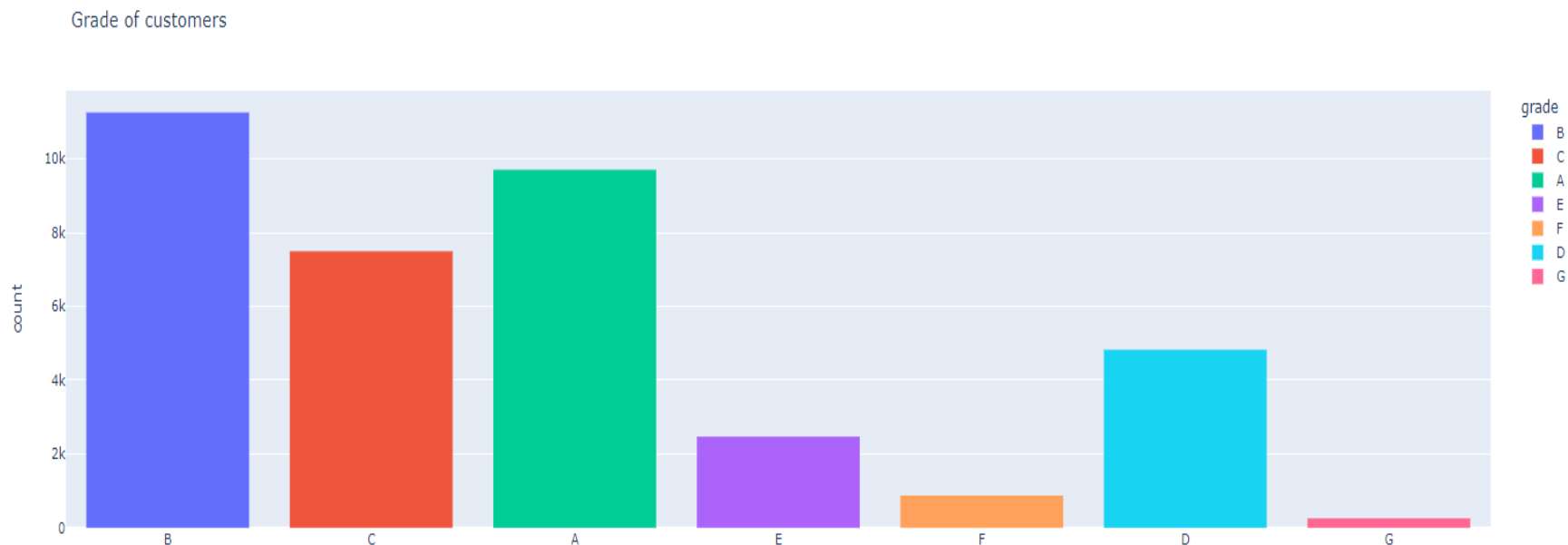
Univariate Analysis



Above graph loan_status shows class is biased by Fully Paid which is good for Lending Club. Fully Paid has 30k+ and Charged Off has 5K+ entries



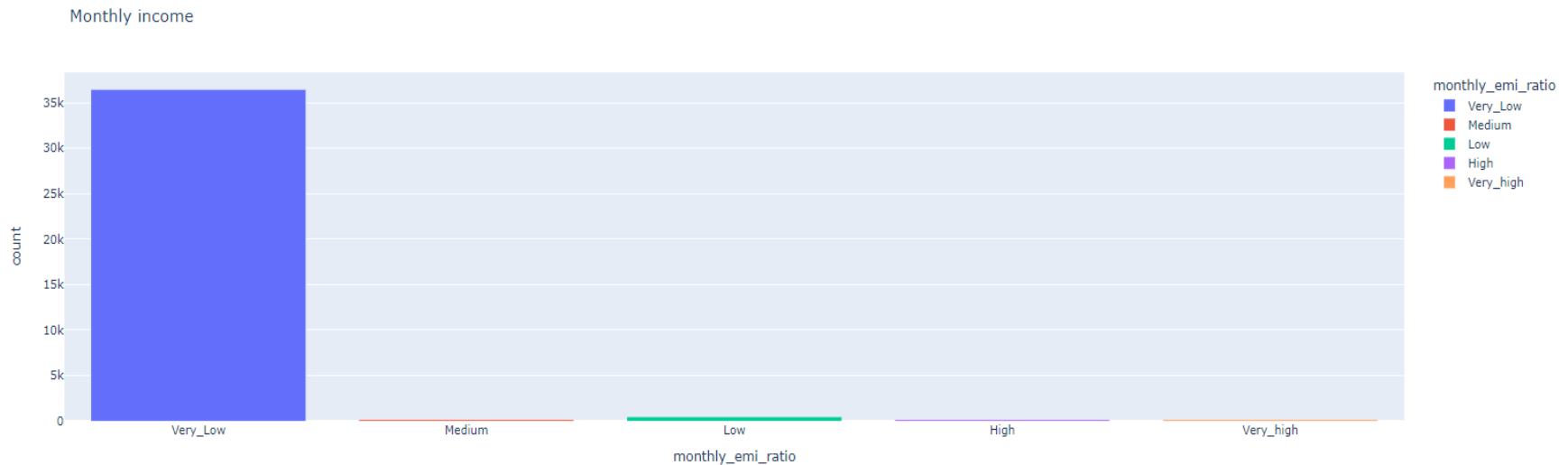
Univariate Analysis



Above graph shows 'B' grade customers have taken Loan Compare to 'A' Grade Customers. 'G' grade customers has low count.



Univariate Analysis

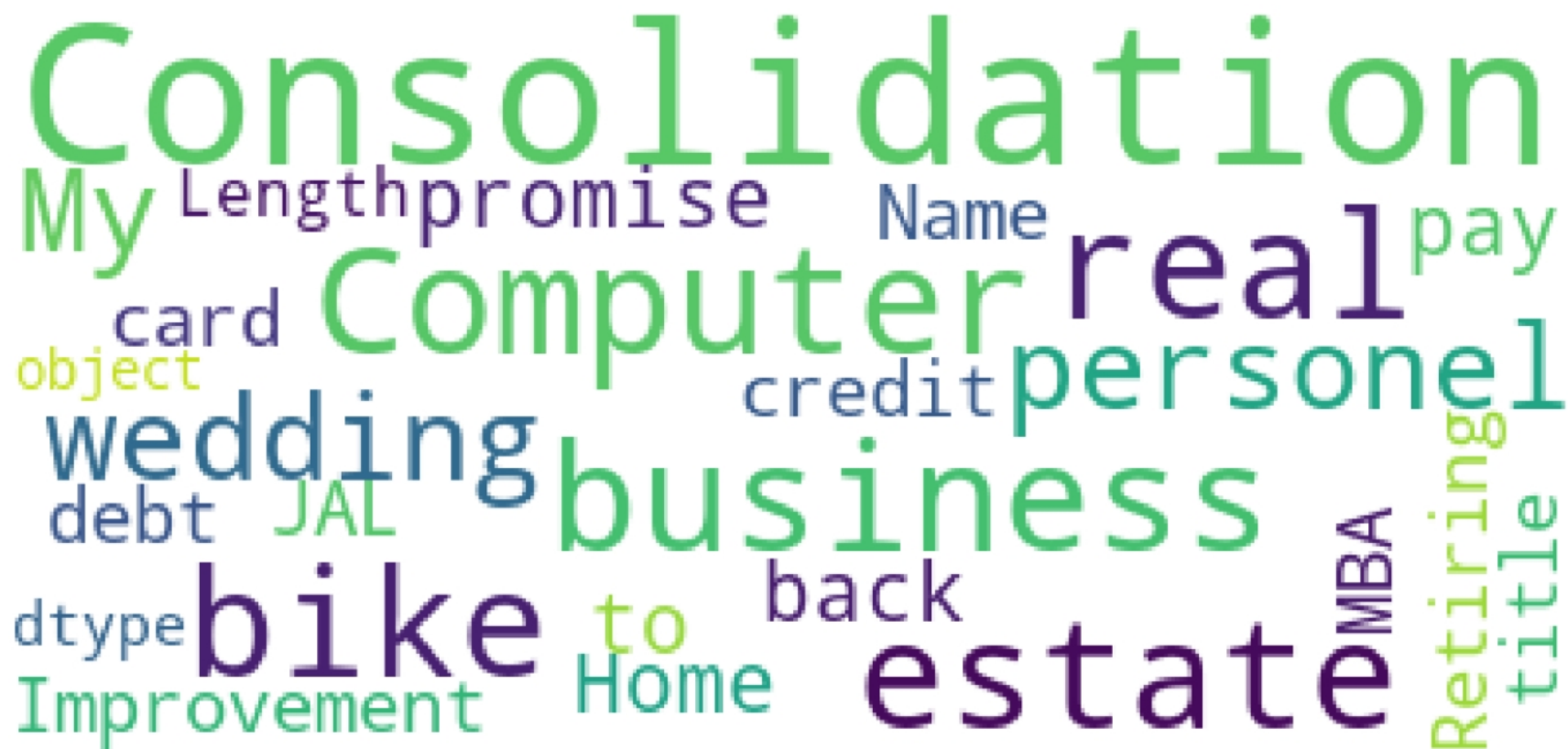


Above graphs shows the ratio of monthly income w.r.t monthly installments. According to Very_Low most of the customers paying monthly EMI less than 30% of monthly salary. About 60 customers paying EMI more than 50% of monthly salary.



Univariate Analysis

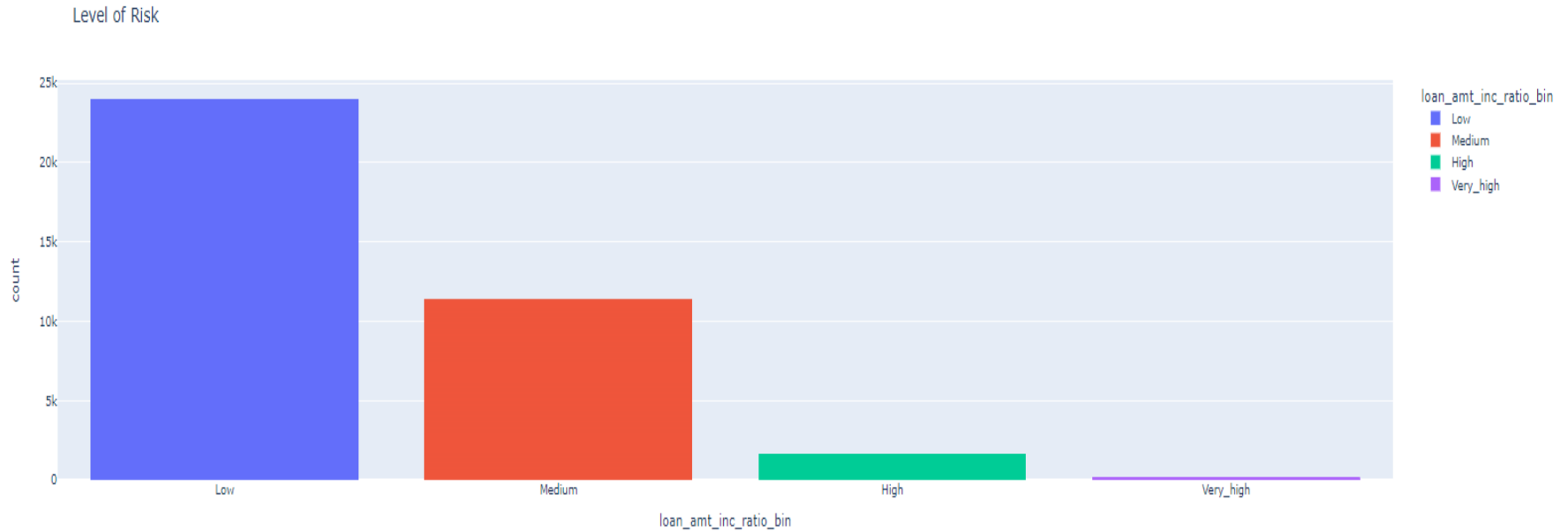
Most common words use Title



Fully Paid has 30k+ and Charged Off has 5K+ entries

Above image is plotting word cloud for title, which shows mostly used text from data. Like Consolidation is used many times so its Font size is bigger than other words.

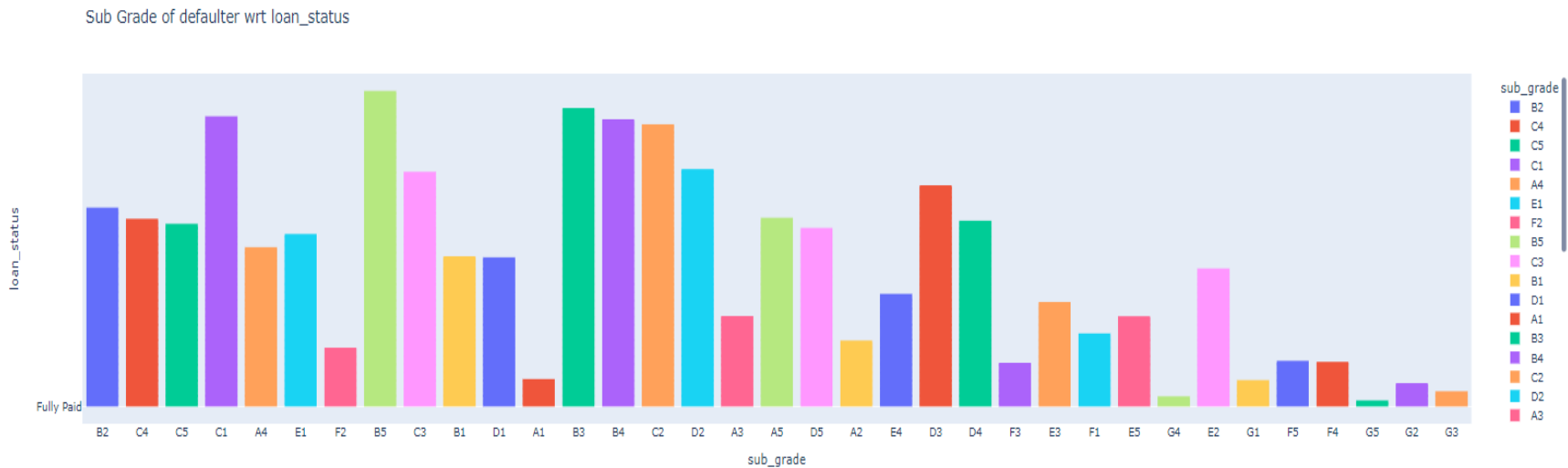
Univariate Analysis



Above graph shows the ratio of loan amount w.r.t annual income

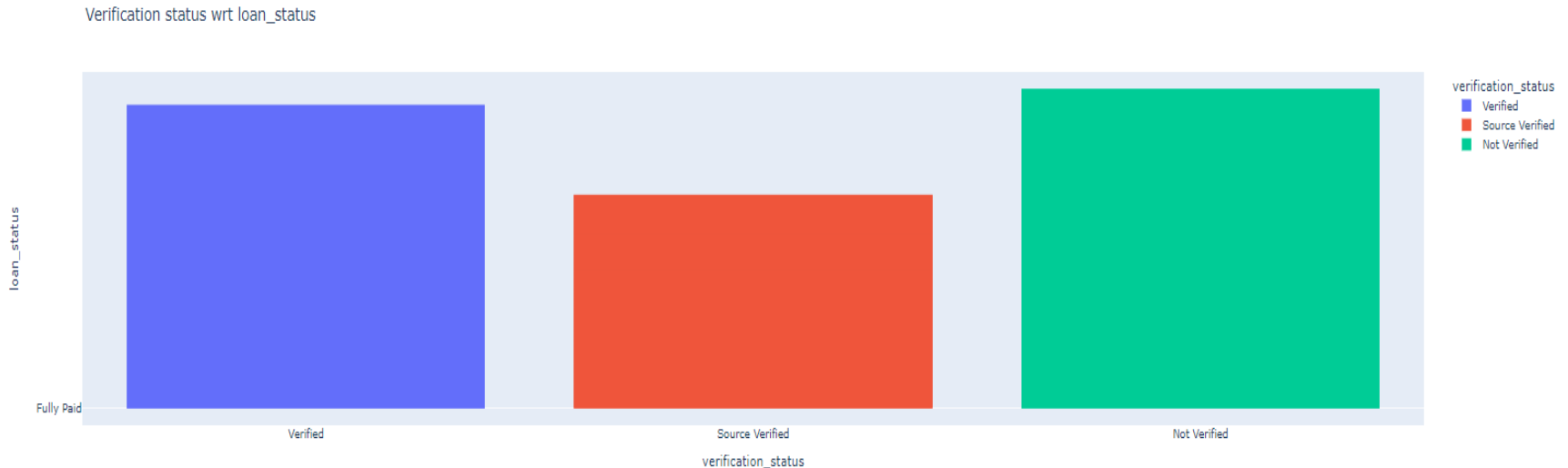


Segmented Univariate Analysis



From above plot for 'sub_grade' we can infer that the defaulter s rate is increasing wrt sub_grade, hence the chances of loan getting defaulted increases with the sub_grade from A1 moving towards G5.

Segmented Univariate Analysis

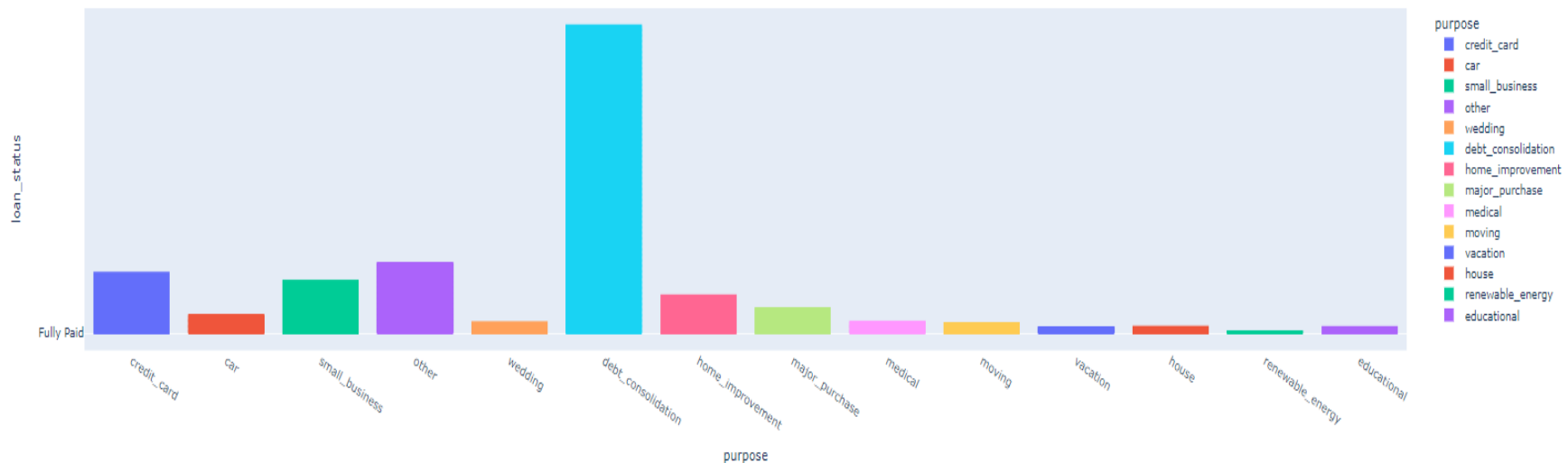


From above plot for 'verification_status' we can infer that the defaulters rate is increasing and is less for Not Verified users than Verified ones, but not useful for analysis.



Segmented Univariate Analysis

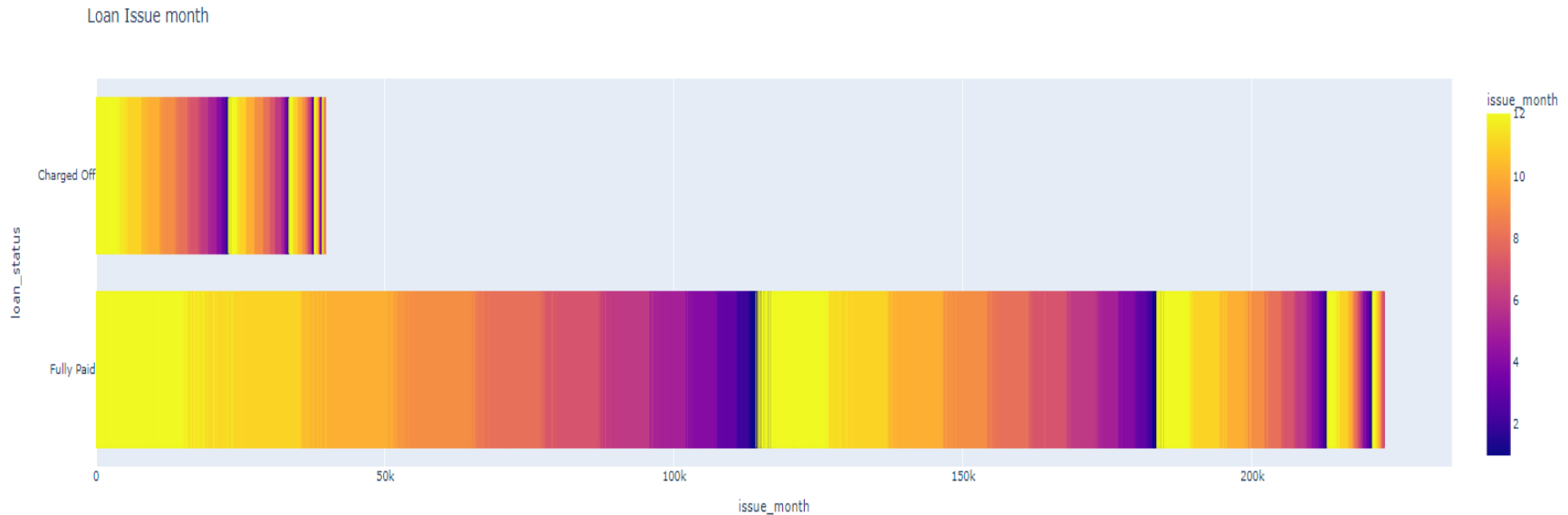
Purpose of taking loan wrt loan_status



From above plot for 'purpose' we can infer that the defaulters rate is nearly constant for all purpose type except 'small business', hence rate will depend on purpose of the loan



Segmented Univariate Analysis

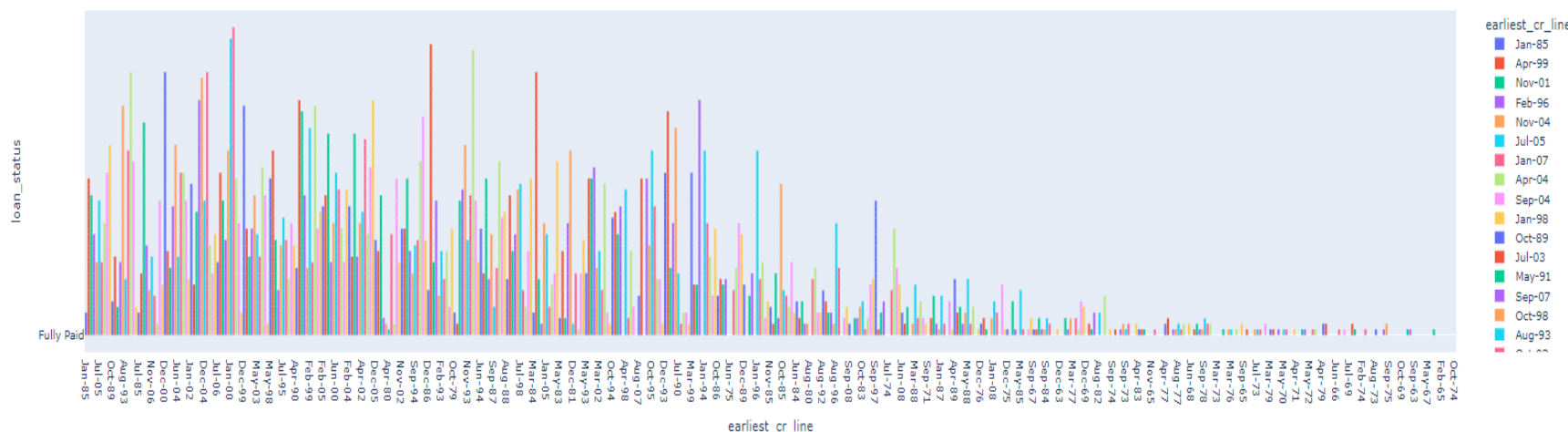


From above plot for 'month' we can infer that the defaulters rate is nearly constant here, not useful



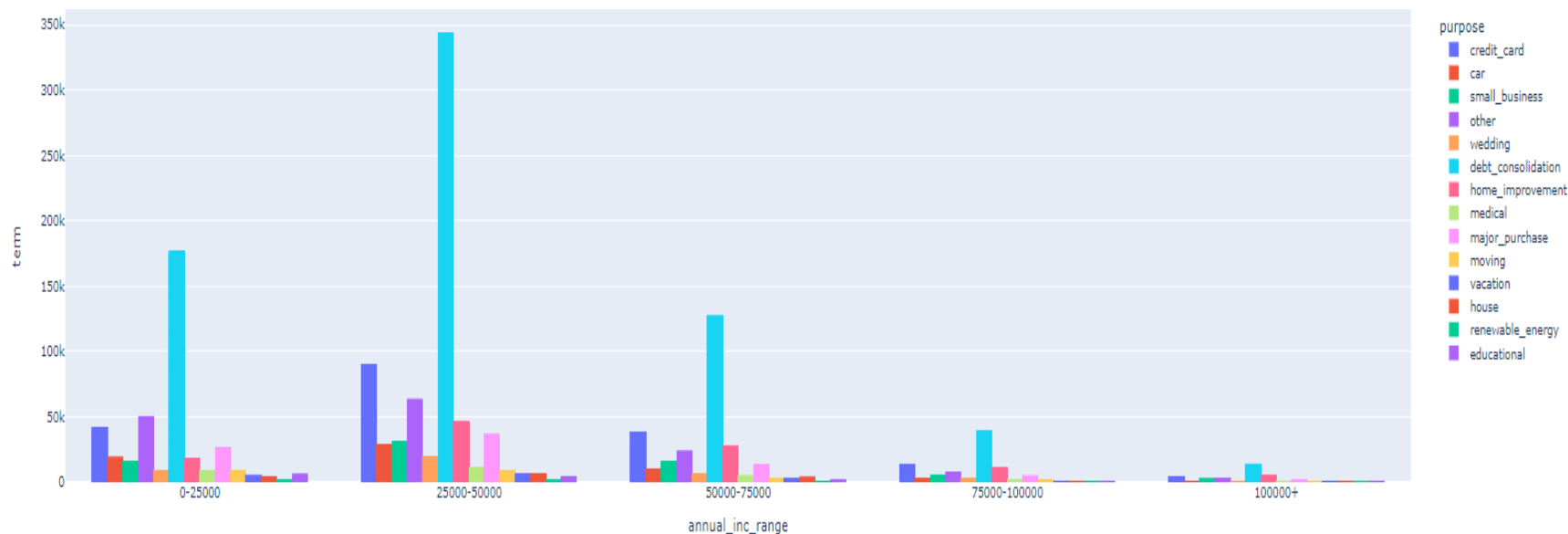
Segmented Univariate Analysis

Status of earliest crdit line



From above plot for 'earliest_cr_line' we can infer that the defaulters rate is nearly constant for all purpose type except year around 65, hence rate does not depend on earliest_cr_line of the person

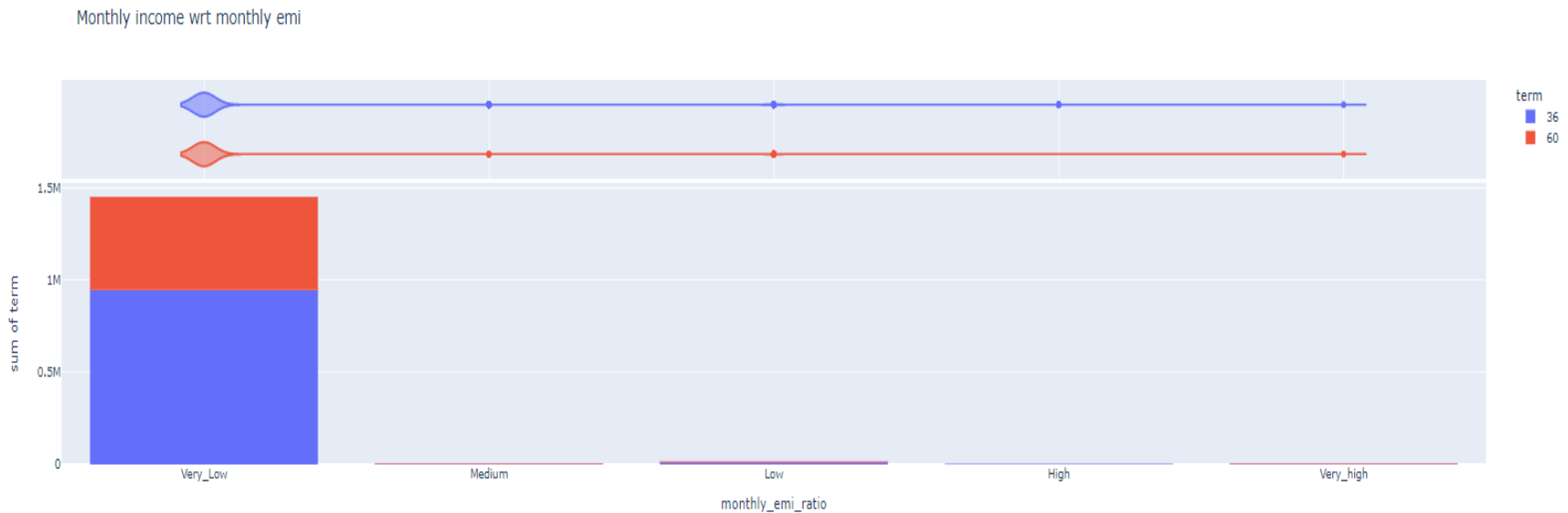
Bivariate Analysis



From above plot, we can infer it doesn't show any correlation



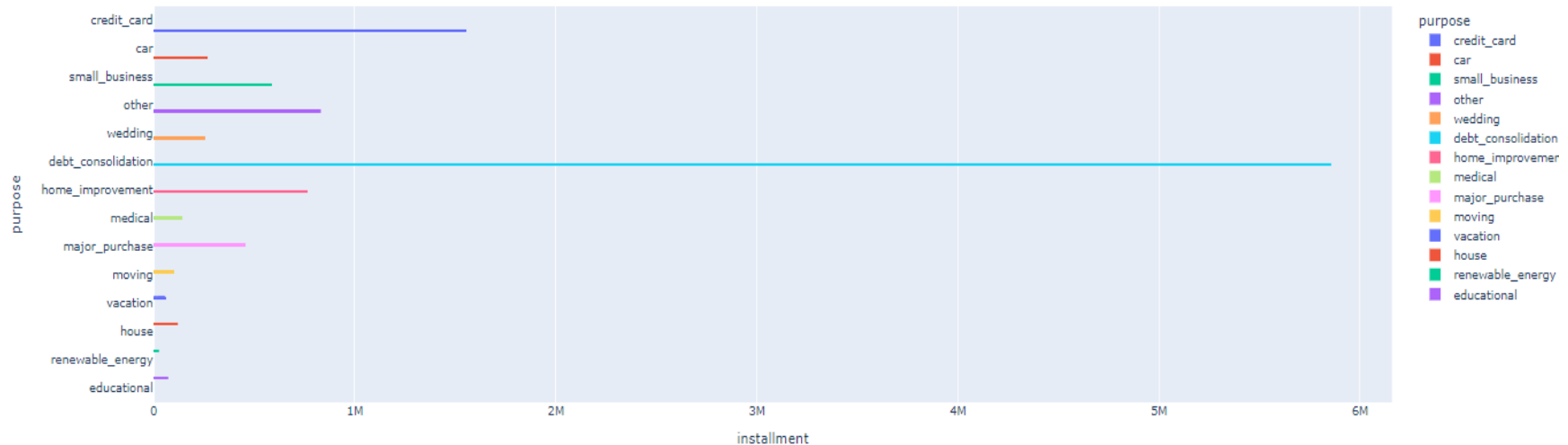
Bivariate Analysis



From above plot for `monthly_emi_ratio` we can infer that the defaulters rate is increasing w.r.t `term`.



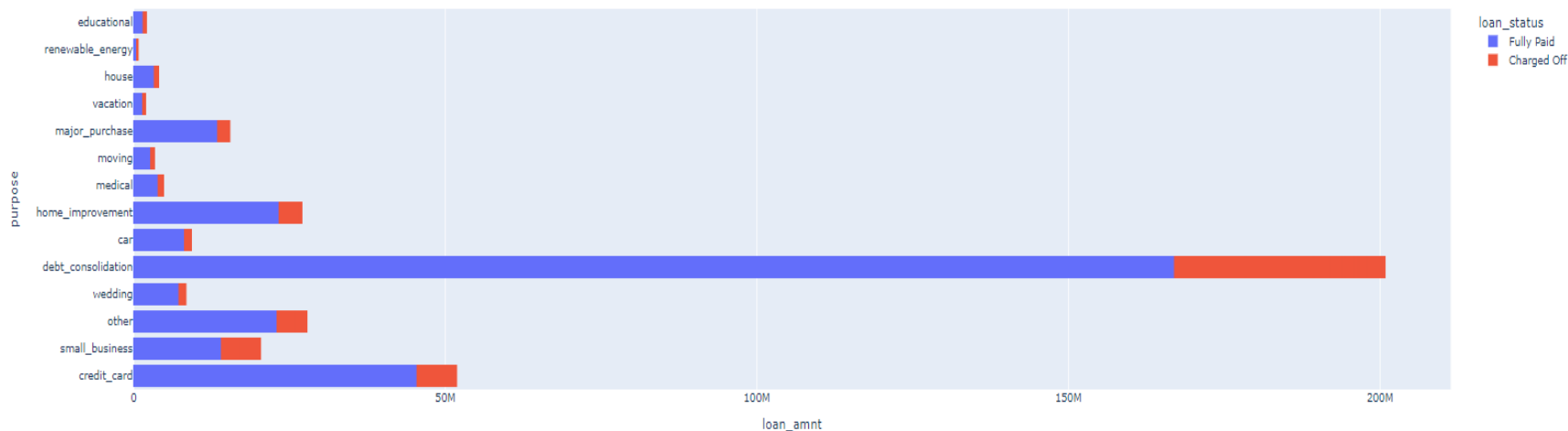
Bivariate Analysis



As we can see straight lines on the plot, default ratio increases for every purpose w.r.t installment except for small_business



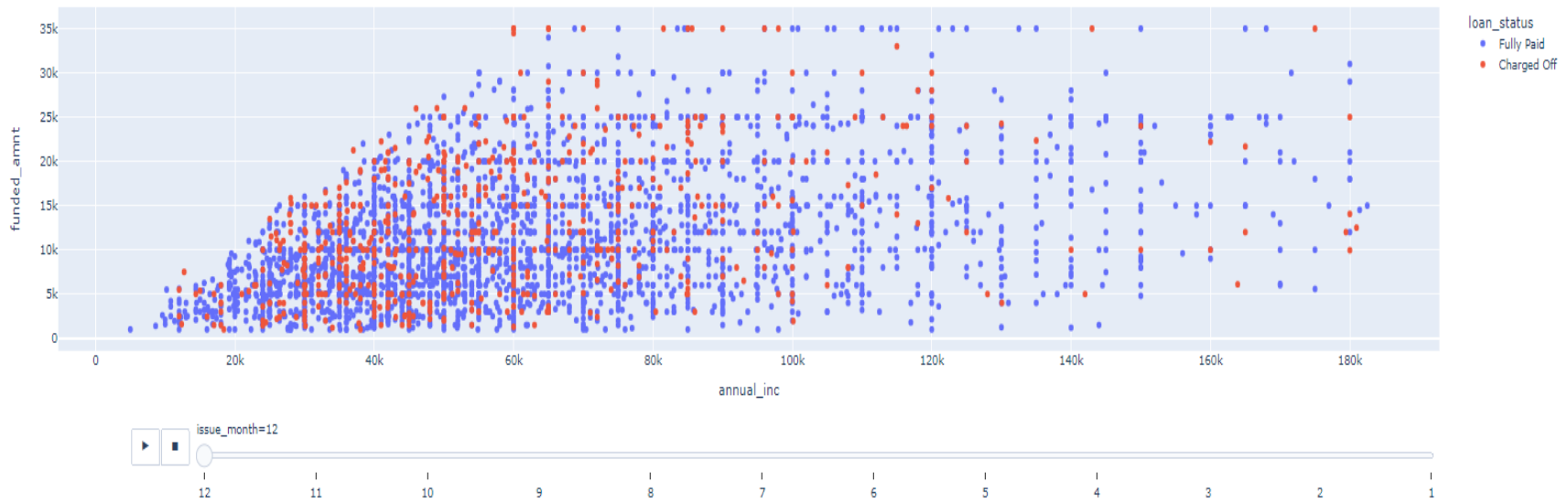
Bivariate Analysis



As we can see straight lines on the plot, default ratio increases for every purpose wrt loan_amnt



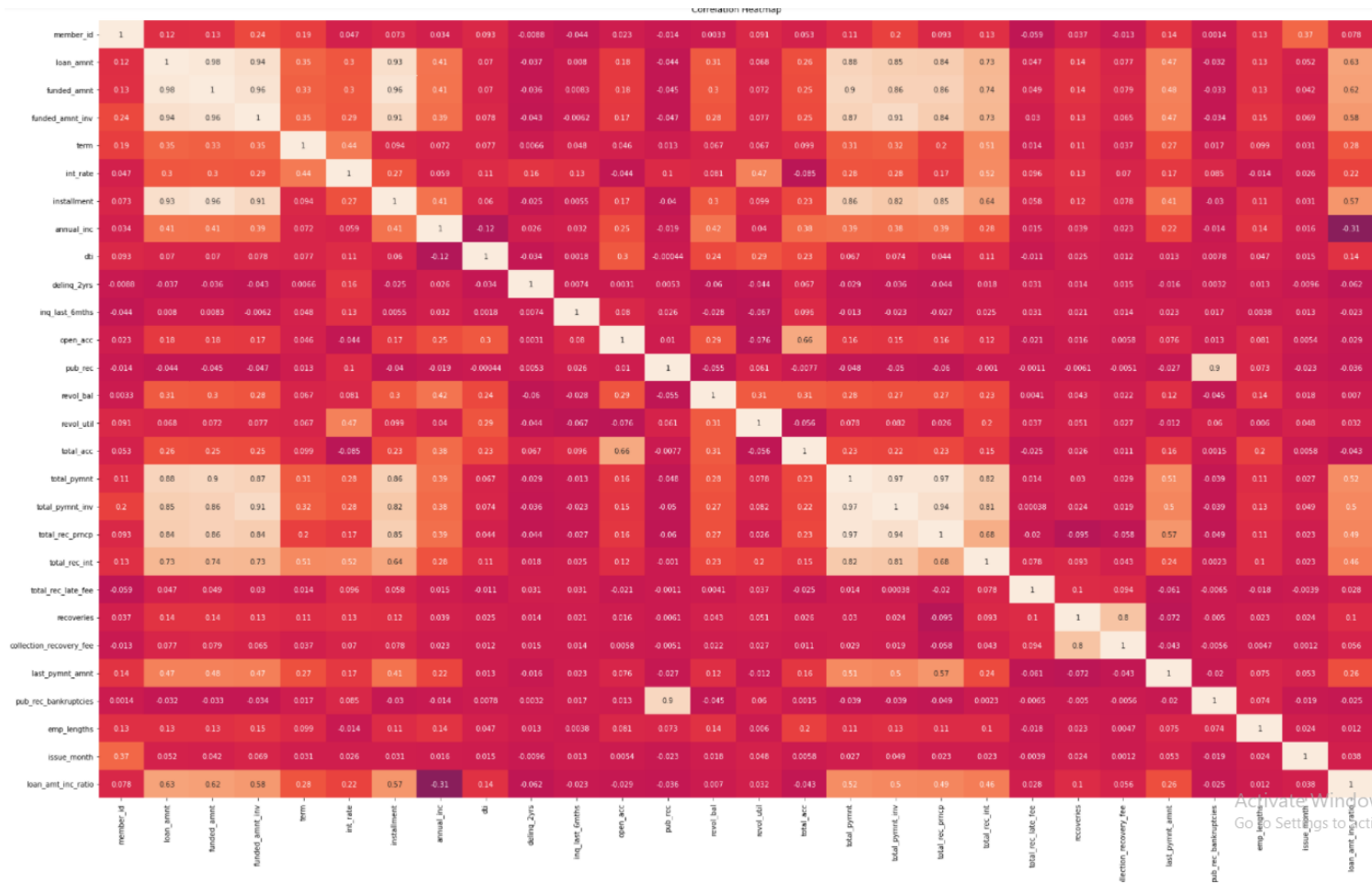
Bivariate Analysis



Above graph shows the scatter plot for ratio of Fully Paid and Charged off customers



Multivariate Analysis



Above graph shows Correlation Matrices

Conclusion

- ▶ As part of the Lending Club Case Study we have analyzed and would recommend that the following factors be considered by Consumer Finance Company to approve / reject the loan of an applicant:
 - ▶ Purpose
 - ▶ Term
 - ▶ Grade
 - ▶ Loan Amount
 - ▶ Interest Rate
 - ▶ Monthly Income
- ▶ Applicants who stay in rented house or mortgage are the maximum who take loans, so bank can take this as an advantage to grow customers.
- ▶ Customers with a loan purpose of Small Business are more likely to default as compared to other categories.
- ▶ The defaulters rate is increasing w.r.t term, hence the chances of loan getting defaulted is less for 36m than 60m.
- ▶ Applicant having purpose debt_consolidate are the maximum applicant for the loans.

