



Lending Club Case Study

Jayashree Jayabalan

Apoorva Srivastava



Problem Statement

The consumer finance company must decide whether to approve or reject loan application based on applicant's profile. The company is looking for factors that can possibly lead to an applicant defaulting loan, be identified as part of the analysis from their past data. This can help or guide them make the decision to approve given loan application or not.



Approach of Analysis - EDA

- The approach identified to solve the problem Exploratory Data Analysis.
- EDA has following stages.
 - Data Sourcing
 - Data Cleaning
 - Univariate Analysis & Segmented Univariate Analysis
 - Bivariate Analysis
 - Derived Metrics
- For this Analysis, we are going to focus on Univariate and Bivariate Analysis
- We have used Python – pandas data frame and matplotlib.pyplot library for our analysis



EDA – Data Sourcing

Data has been sourced from UpGrad learning platform. The data includes:

- Loan.csv – Contains the historical data of loans approved by the finance company.
- Data Dictionary – Provides description for the columns in loan.csv
- The loan.csv was loaded into pandas data frame.



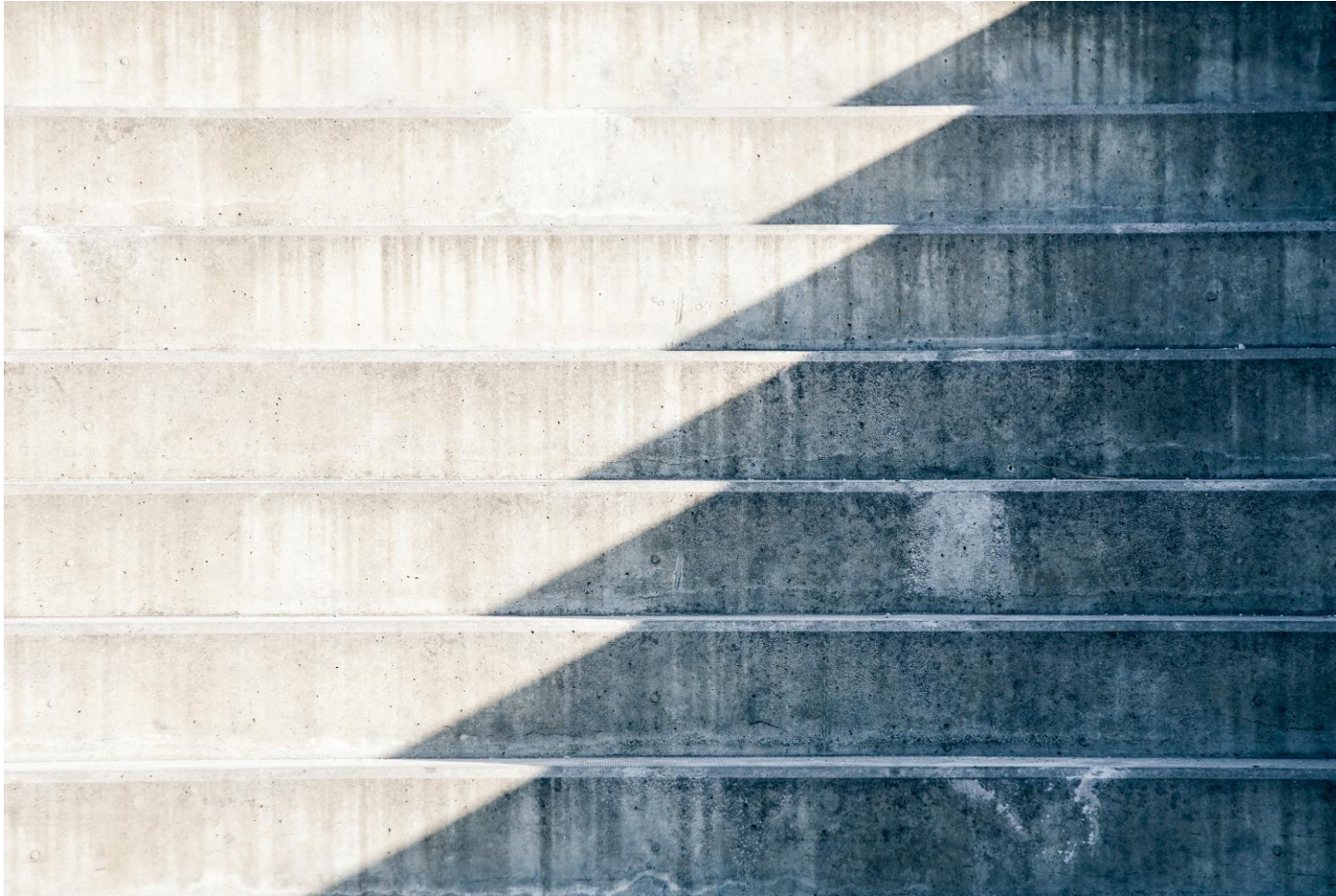
EDA – Data Cleaning - Columns

- ✓ Following steps were carried out to clean the data.
 - ✓ All columns having all null values were dropped
 - ✓ All columns having more than 50% null values have been dropped from the data frame
 - ✓ All columns having variance outside the desirable range have been dropped
 - ✓ Finally, we ended up with 14 columns



EDA – Data Cleaning - Rows

- ✓ Following steps were carried out to clean the data.
 - ✓ All rows having no Job description, but salary have been dropped.
 - ✓ Irrelevant data for analysis like ongoing loans have been removed.
 - ✓ Removed rows having outliers in annual income.
 - ✓ Removed rows having loan amount outliers.
 - ✓ No of Rows are 33690.



EDA – Data Cleaning – Data type conversion

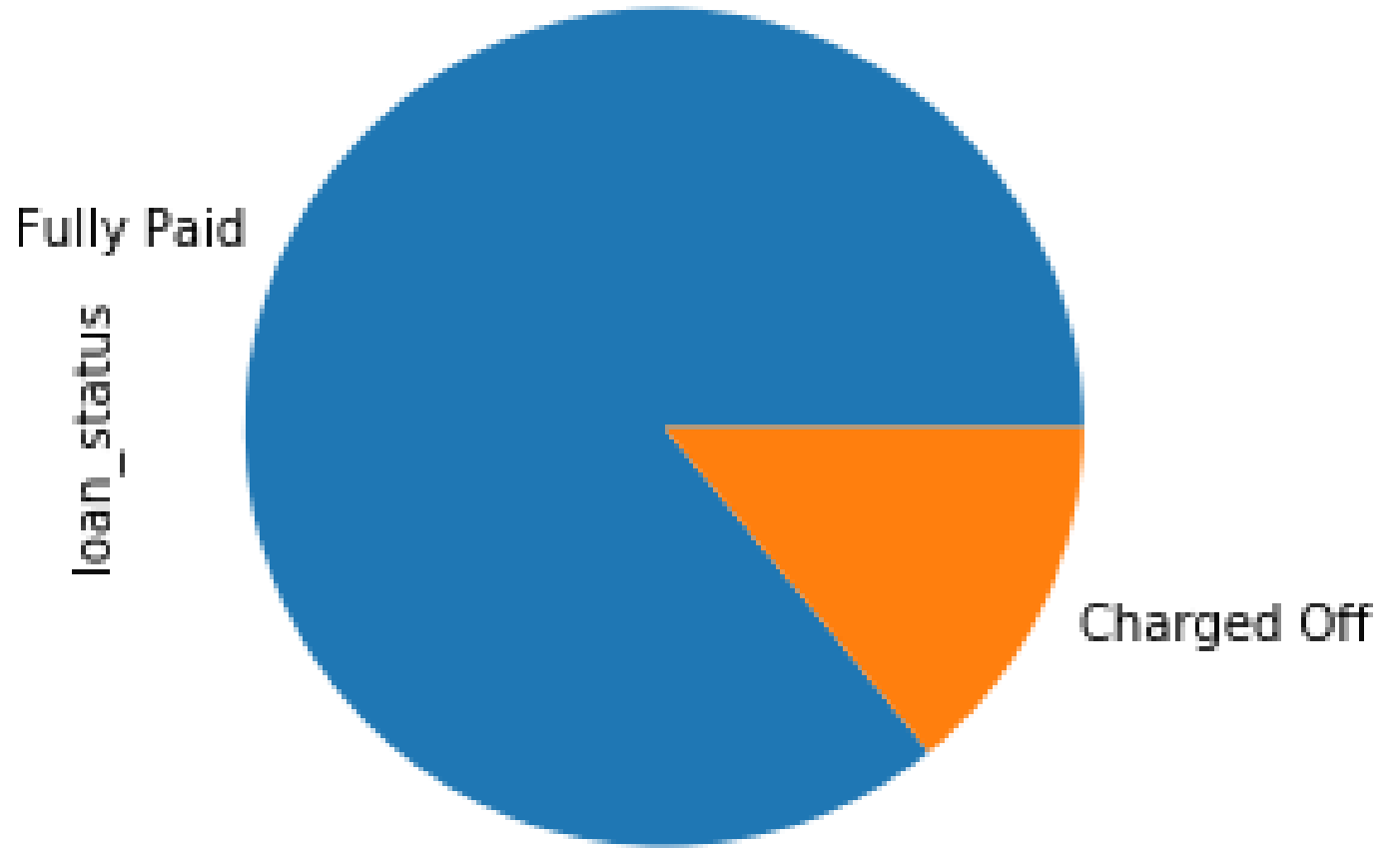
Interest rate has been
converted into float.

Type of Variables identified for Analysis

- Grade and Sub Grade – Ordered Categorical Variables
- Bankruptcy – Numerical variable
- Loan amount – Numerical variable
- Purpose – Unordered categorical variable
- Interest Rate – Numerical variable
- Monthly income and Monthly installment – Continuous variables (numerical)

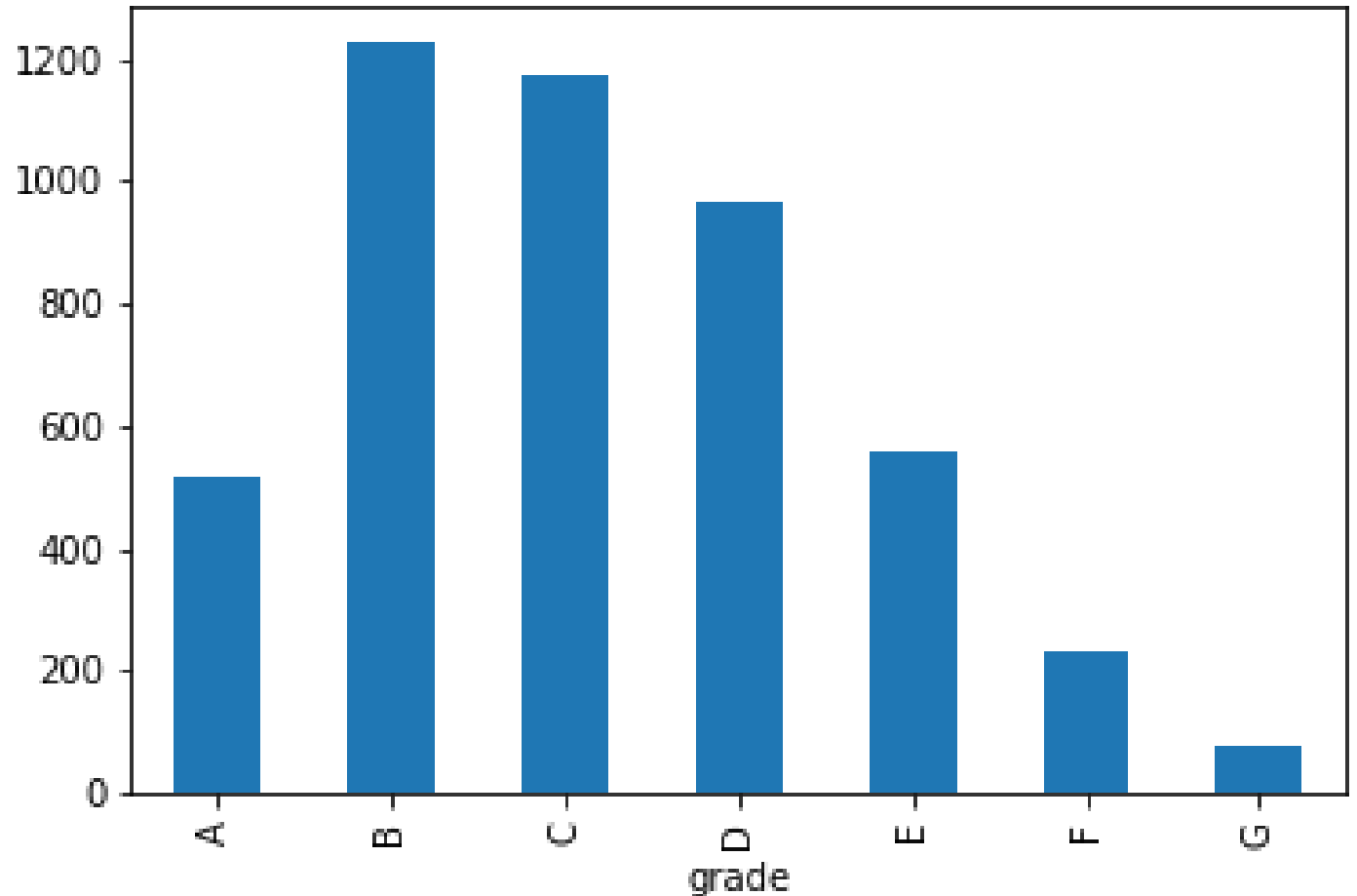
EDA – Data Analysis

- ❖ We have plotted various charts for EDA with values derived from the data frame.
- ❖ To start with, we have plotted a pie chart to display the proportion of Charged_Off (defaulted) loan Vs Fully paid loans.



EDA – Univariate Analysis

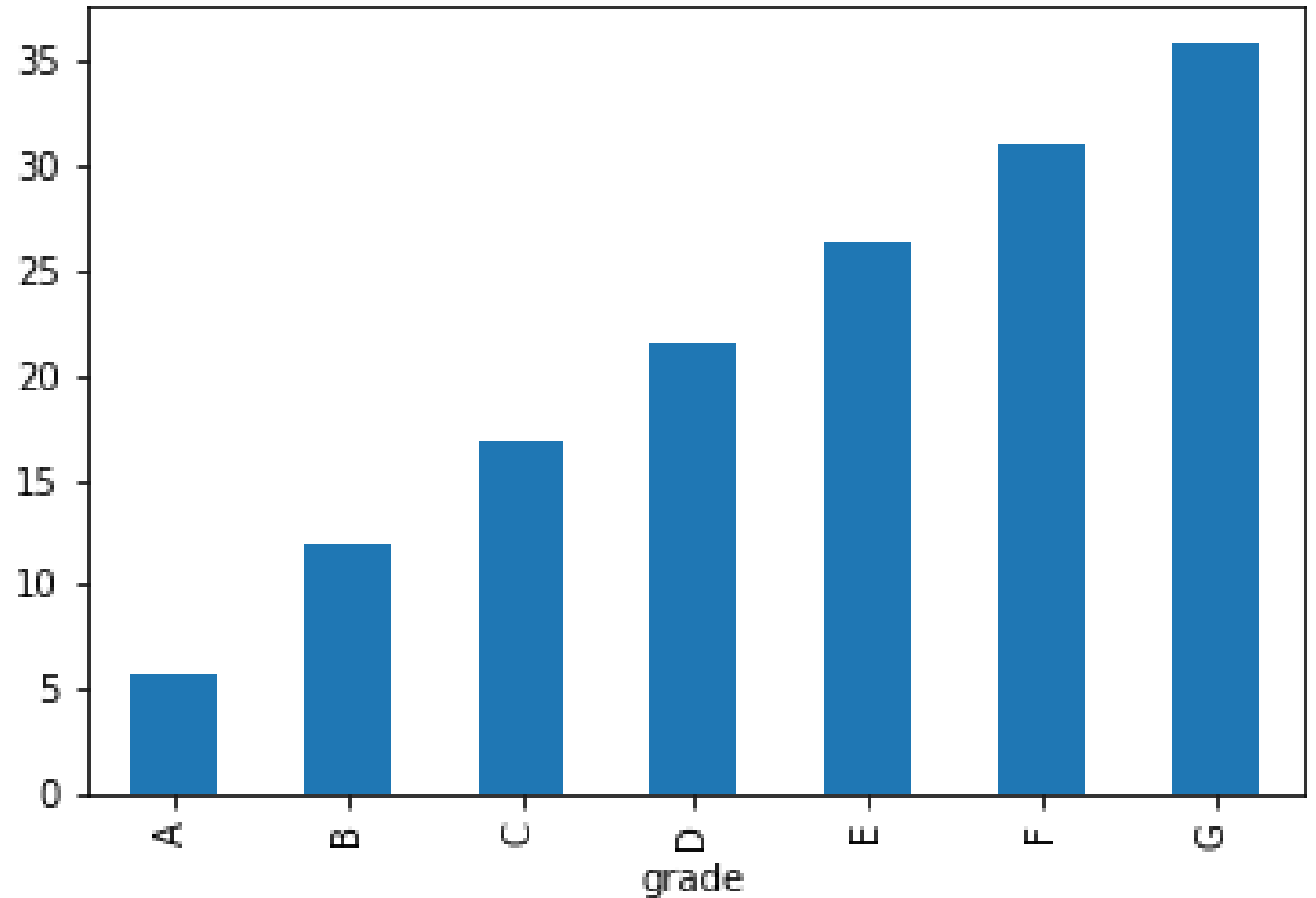
- We have plotted a bar chart to display the distribution of loans against Grades of loan applicants..
- From the chart we can see that the most charged off loan is from Grade 'B'
- But this is not significant enough to say that Grade 'B' applicants are going to be defaulters as maximum number of loan applicants belong to 'B' Grade.

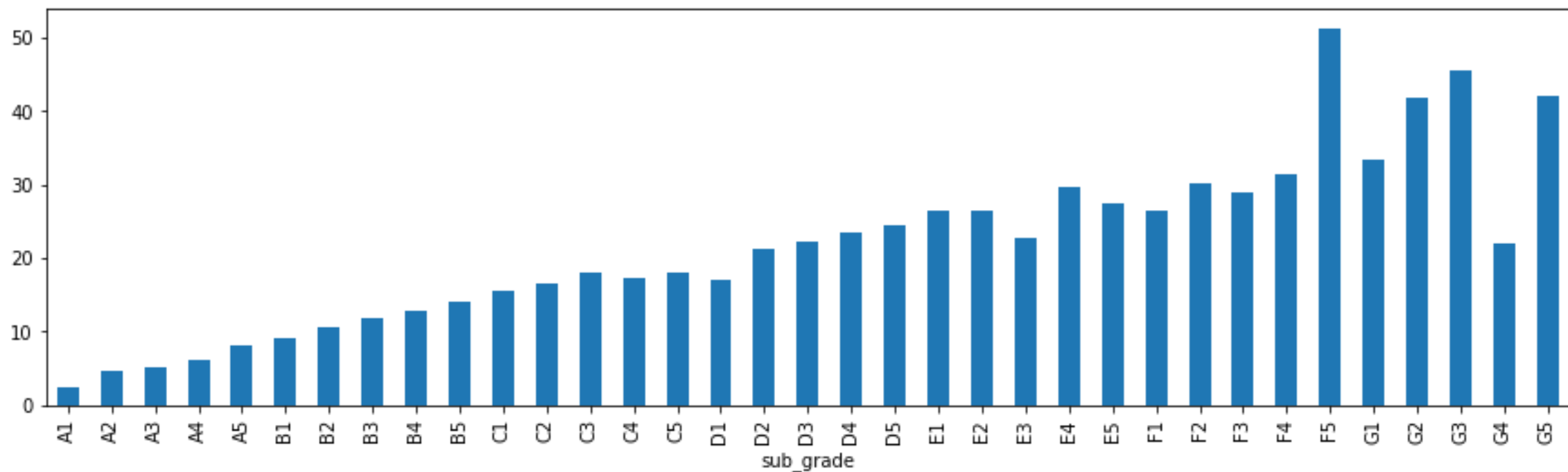


EDA – Univariate Analysis

As the factor Grade couldn't derive any conclusion based on number of loan applicants, we thought of taking the percentage or ratio of Charged off number vs total number and plot it.

This bar chart shows that lower the Grade it is, higher is the chance of defaulting.





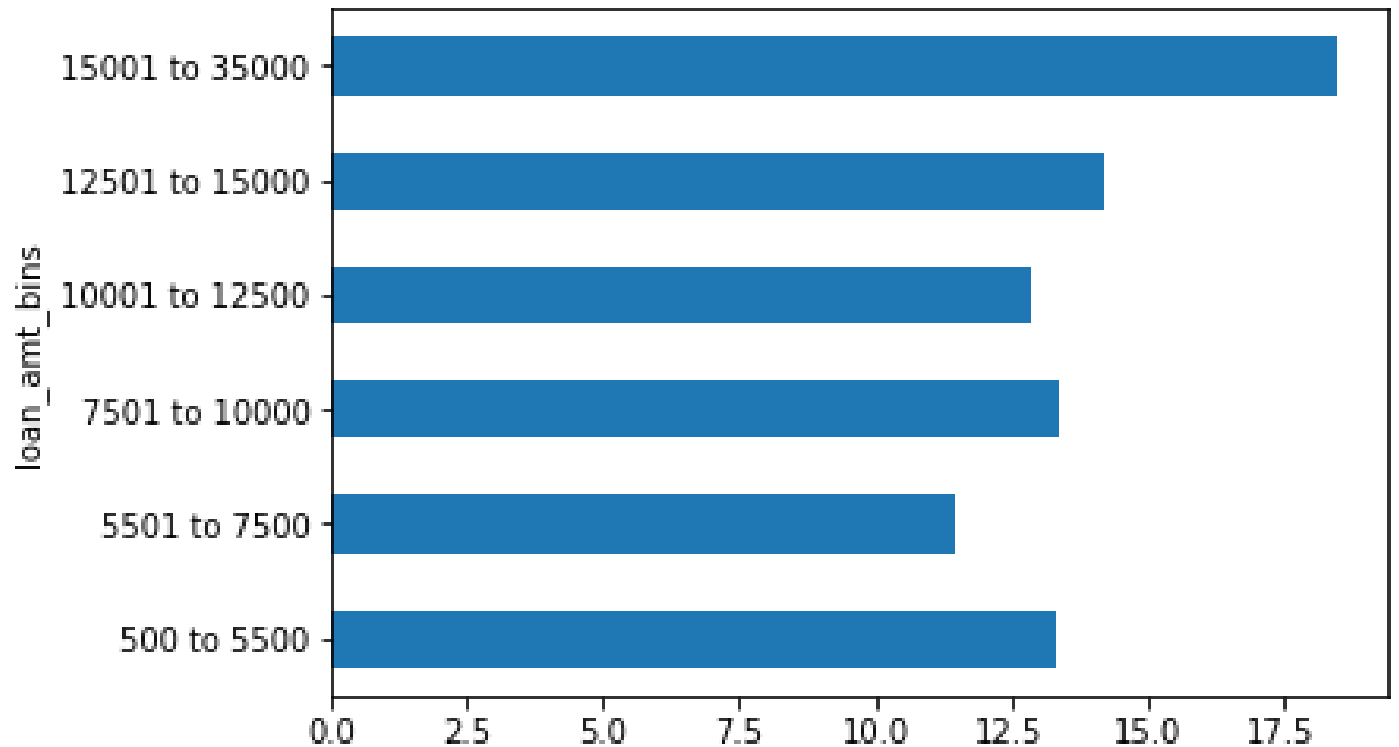
EDA – Univariate Analysis

To support our conclusion on the factor “Grade”, we have plotted the ratio of “sub grade charged off applicants Vs sub grade total applicants”.

This has strengthened our conclusion as we see majority of charged off cases have come from the lower grades “F & G”

EDA – Univariate Analysis

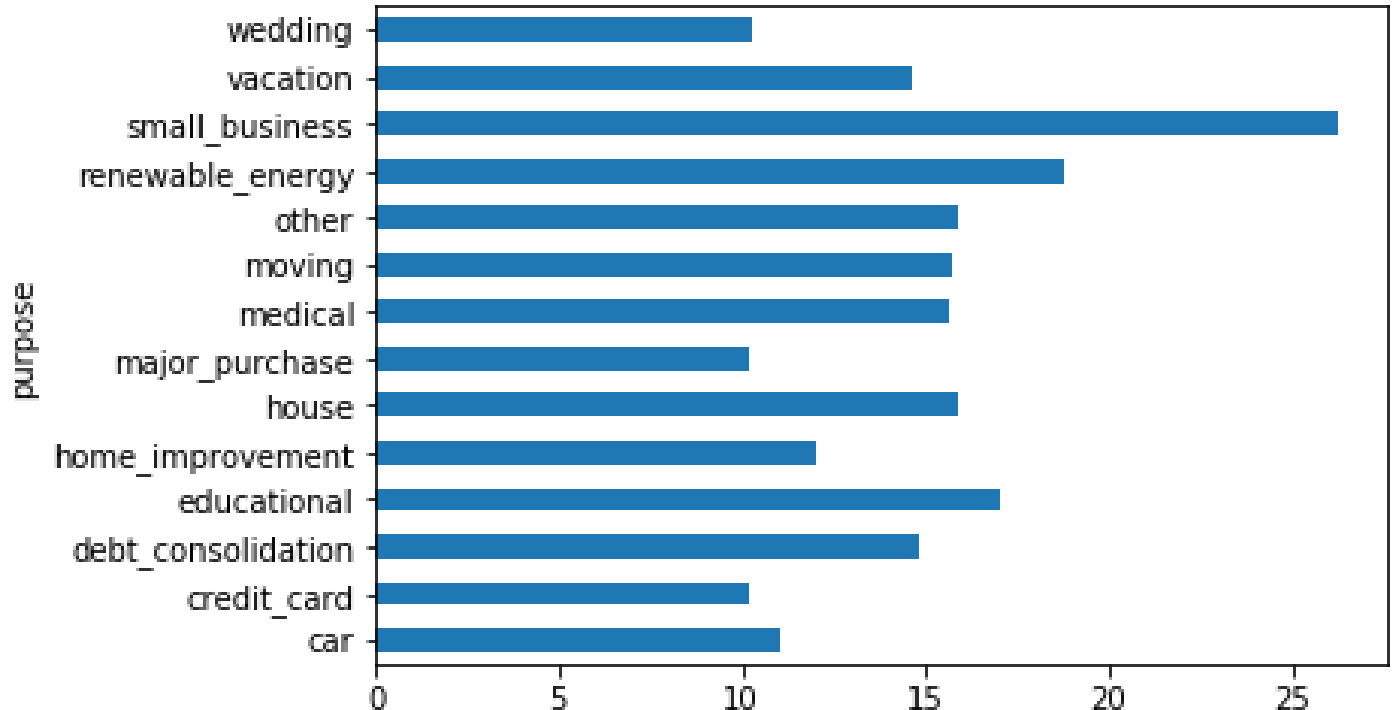
Another plot based on loan amount. This just shows that the loan amount is very high then the risk of applicants defaulting is also more.

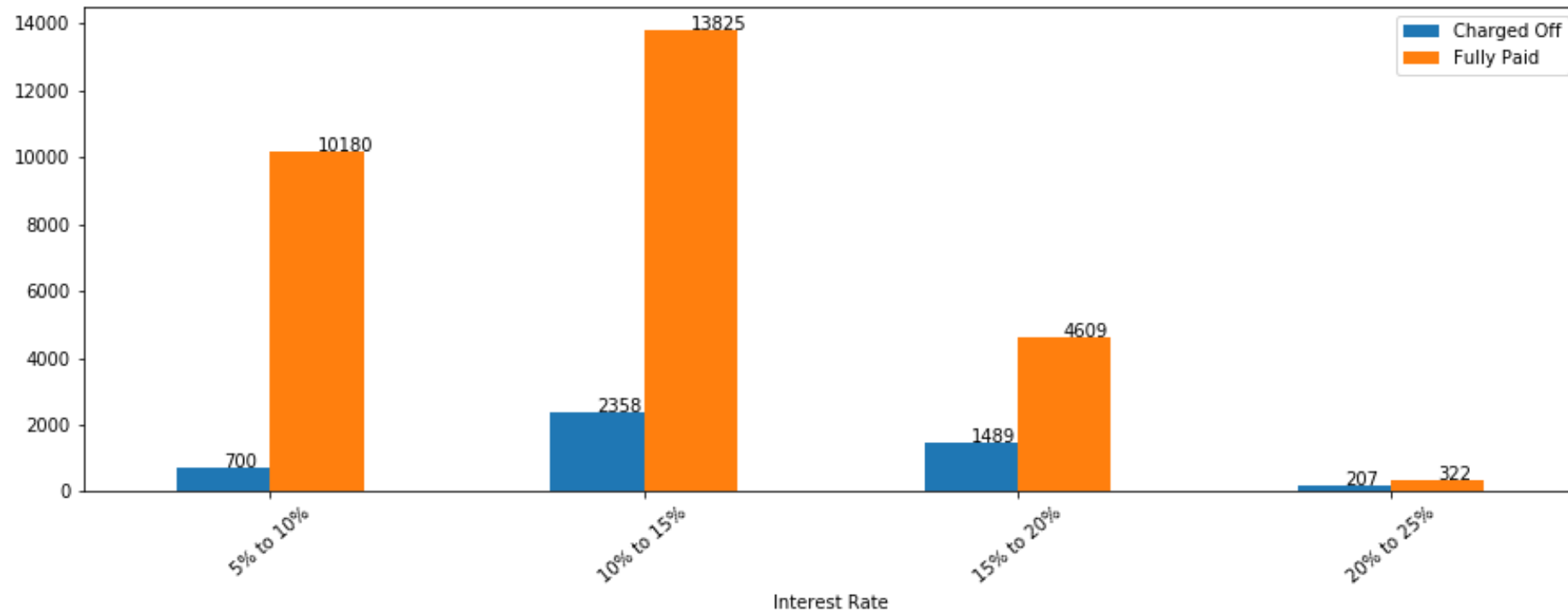


EDA – Univariate Analysis

Another plot based on purpose of loan. This just shows that applicants availing loan for small businesses are more like to be defaulted.

The next threat comes from categories renewable energy or students availing educational loans.





EDA –Univariate Analysis

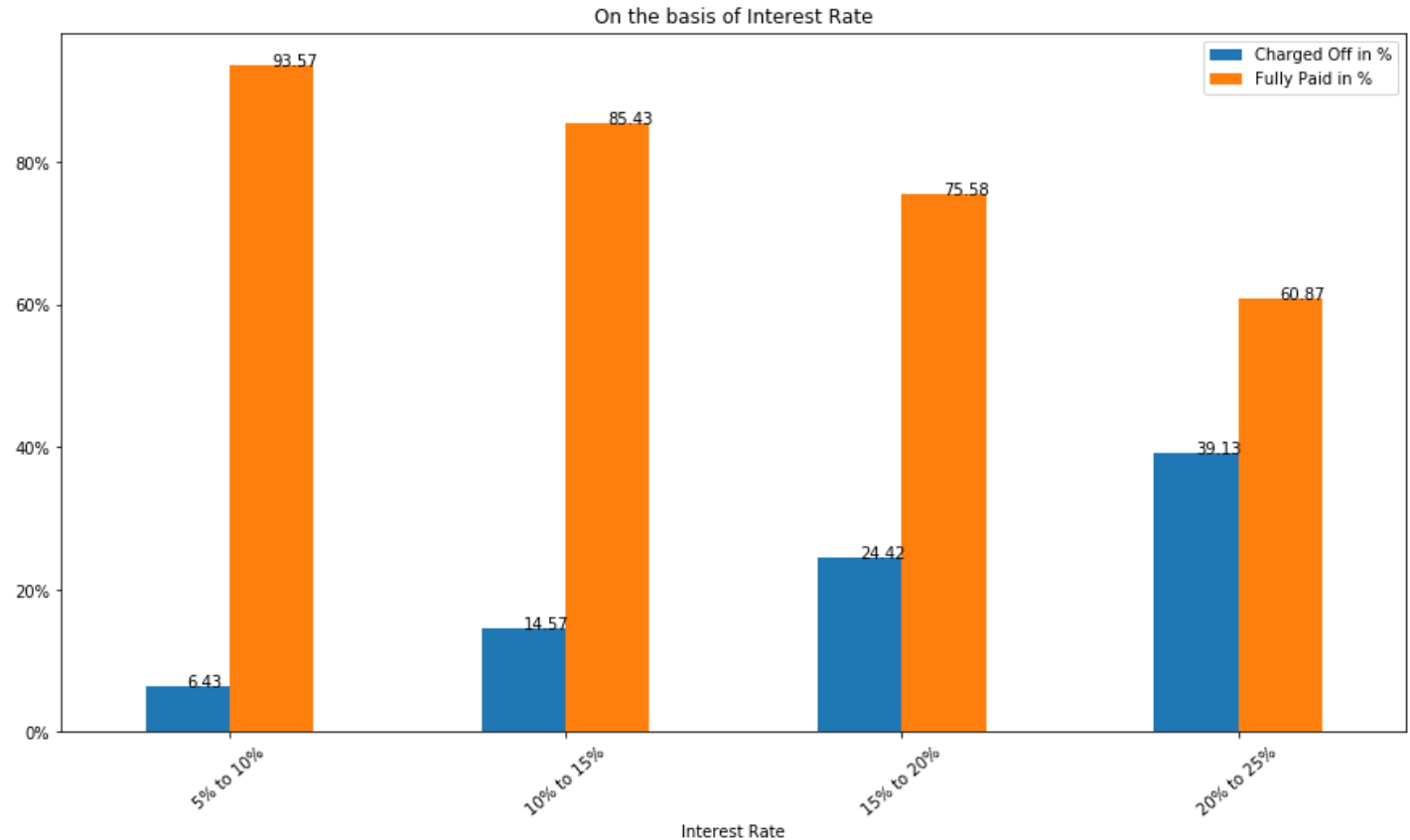
Presenting the bar graph based on Interest Rate. The data is grouped into range of buckets and plotted.

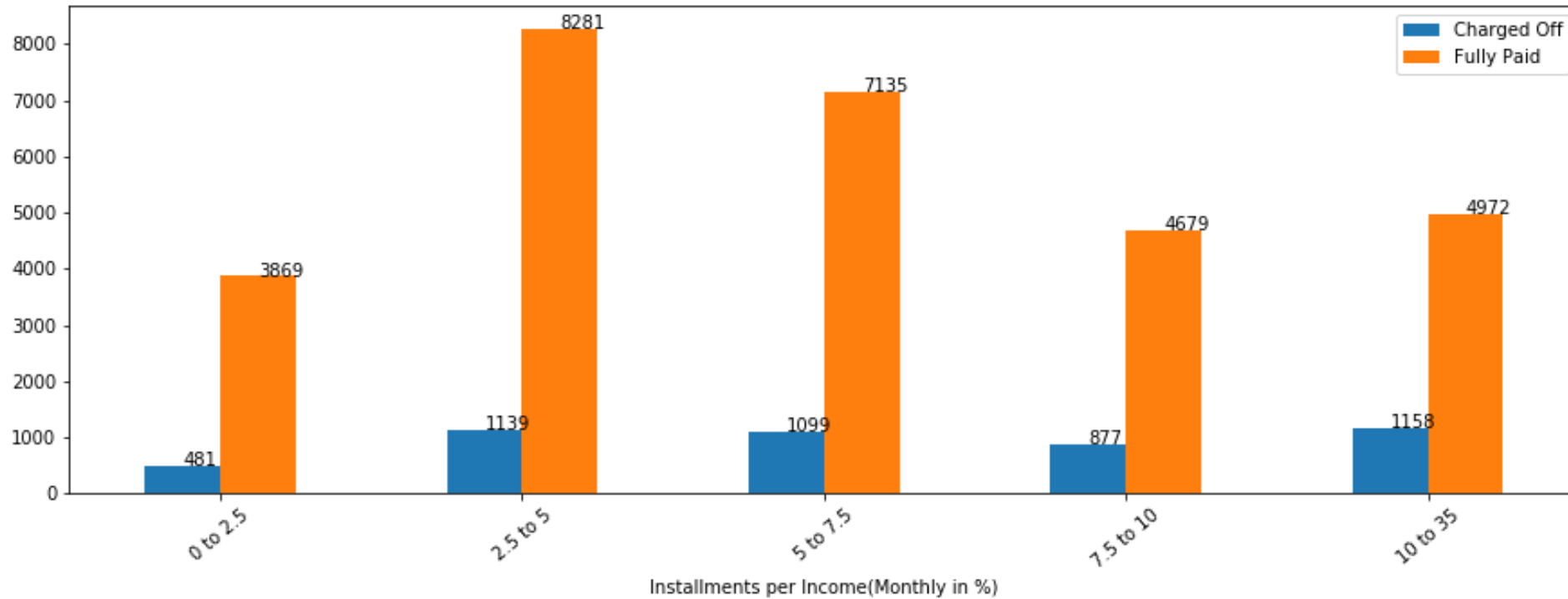
We couldn't produce any conclusion from the graph

EDA – Univariate Analysis

We have then plotted the % of charged off applicants Vs Fully Paid applicants %.

This plot clearly shows the interest rate being a factor for applicants defaulting the loan.





EDA –Bivariate Analysis

The Bivariate analysis has been conducted based on monthly installment of loan and monthly income of the applicant. Monthly income has been derived from annual income.

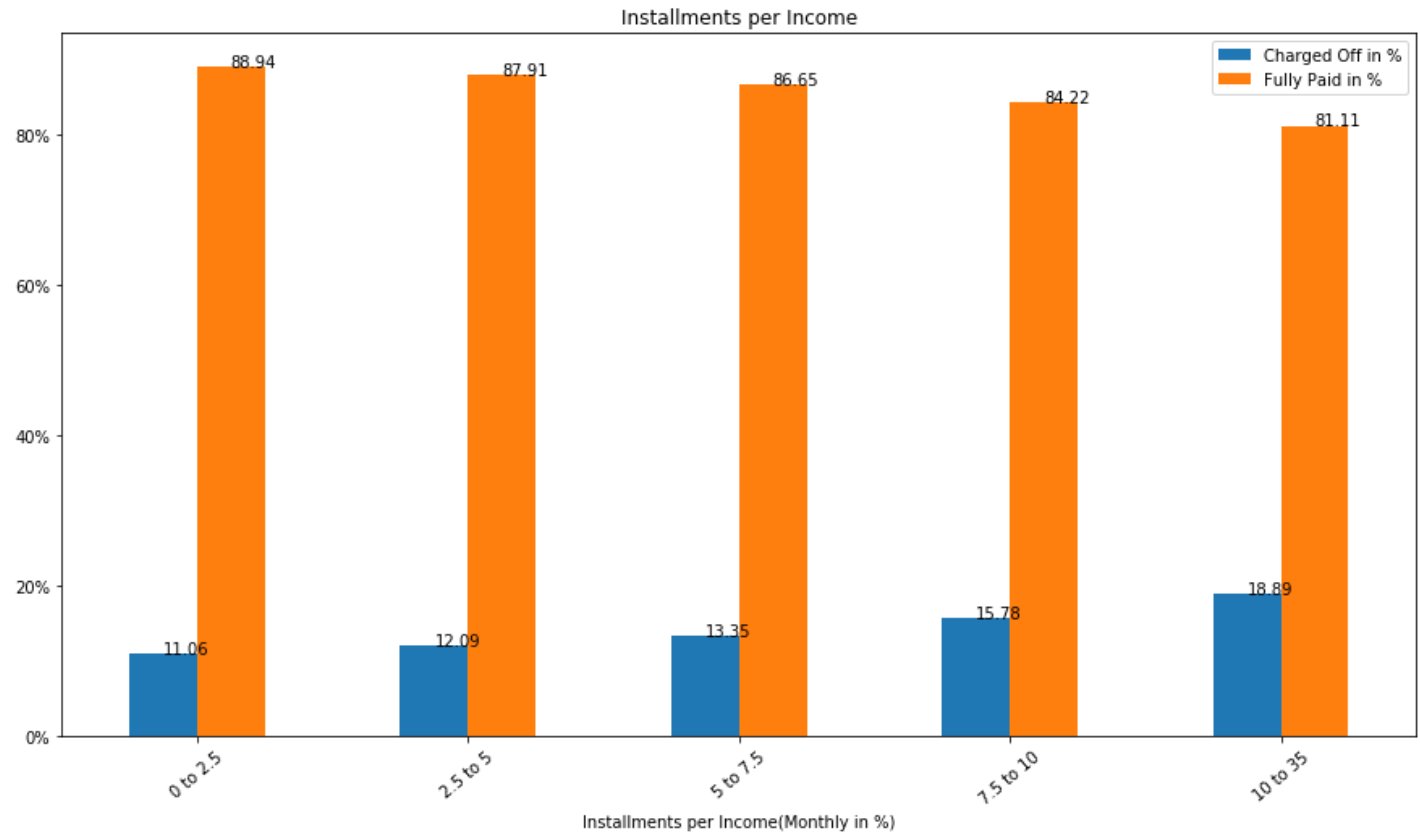
When the data is plotted based on number of applicants falling the range of ratio of monthly installment to income, it seems that the ratio above 2.5% almost has the same effect on applicants charging off.

EDA – Bivariate Analysis

However, when the number of applicants in 'Y' axis is replaced by % of applicants, we can see significant results.

The monthly installment increases when monthly income increases. Hence, they are positively correlated.

Our conclusion from the analysis is that the installment increases with monthly income. This further has an impact on the loan as the installment increases, the defaulting ratio also increases.



That's the end of our analysis.

Thank you for providing the
opportunity to perform the analysis
on Lending club case study.

