Lending Club Case Study

Shivashankar Kankanmeli

Problem Statement

The Loan dataset given by a company which is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Company wants to avoid lending loans to risky applicants by not lending the loan. From the given Loan dataset, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Data understanding

Following is the summary of Data from Loan Dataset:

- Loan Dataset has 39717 entries (Rows) and 111 columns
- Out of 111 columns most of the column data has either NA or Zero values and those columns are of not significance to perform analysis and find the common patterns of defaulters.
- The columns/variables considered for the analysis are loan_amnt,funded_amnt_inv,term,int_rate,installment,grade,sub_grade,emp_title,emp_length,home_ownership,annual_inc,verification_status,issue_d,loan_status,purpose,title,revol_bal,revol_util,application_type
- Main column variables used for evaluation are Loan amount, Term, Interest rate, Installment, Annual Income, Verification status, Loan status, Purpose
- Data Types are float64(3), int64(2), object(14)
- Boxing or Binning is done on following assumptions
- Interest Rates
    - Low Interest: 7
    - Mid Interest : 7.1-11
    - High Interest : 11+
- Income:
    - low Income: 0-50000
    - mid Income: 50001-100000
    - high Income: greater than 100000

Data Analysis

- From the given Dataset, first Identified the count of Loan status and found out that

  Loan status with Fully Paid count is 32145
  Loan status with Charged off count is 5399
  Loan status with Current count is 1098

- We can ignore the entries with Loan status with **Current** as they are still paying the Interest and might not be a possibility of going into a Defaulter status

- Univariate Analysis are being done on Loan status, Interest Rate, Purpose, which has given clarity on count of loan entries based on category

- Bivariate Analysis has given few insights on the patterns of Defaulters and let us provide few recommendations. Few samples are given below

	loan_status	int_rate	Count
0	Charged Off	high interst rate	4205
1	Charged Off	low interst rate	155
2	Charged Off	medium interst rate	1039
3	Fully Paid	high interst rate	17608
4	Fully Paid	low interst rate	3454
5	Fully Paid	medium interst rate	11083

```
loan_status    purpose
Charged Off    car                     155
               credit_card             510
               debt_consolidation     2679
               educational              52
               home_improvement        330
               house                    59
               major_purchase          211
               medical                 101
               moving                   84
               other                   596
               renewable_energy         18
               small_business          462
               vacation                 50
               wedding                  92
Fully Paid     car                    1293
               credit_card            4389
               debt_consolidation    14996
               educational             265
               home_improvement       2455
               house                   295
               major_purchase         1869
               medical                 555
```

Data Analysis

- Multivariate Analysis on loan status, Annual Income and Interest Rate has given the following input

```
loan_status    annual_inc     int_rate
Charged Off    high income    high interst rate       490
                              low interst rate          6
                              medium interst rate      93
               low income     high interst rate      1823
                              low interst rate         90
                              medium interst rate     525
               medium income  high interst rate      1892
                              low interst rate         59
                              medium interst rate     421
Fully Paid     high income    high interst rate      2857
                              low interst rate        465
                              medium interst rate    1522
               low income     high interst rate      6564
                              low interst rate       1225
                              medium interst rate    4451
               medium income  high interst rate      8187
                              low interst rate       1764
                              medium interst rate    5110
dtype: int64
```

Recommendation

- After following the steps of Data Understanding, Data Analysis and Data visualization (samples are shown in next few slides). Following are the recommendation given

- There are around 5000 Charged Off entries, who became the defaulters.

- Charged off people with Low and Medium Income people became more defaulters than High Income people.

- Charged off loan status with High and Medium interest rate are the defaulters

```
loan_status   annual_inc      int_rate
Charged Off   high income     high interst rate     490
                              low interst rate        6
                              medium interst rate    93
              low income      high interst rate     1823
                              low interst rate       90
                              medium interst rate   525
              medium income   high interst rate     1892
                              low interst rate       59
                              medium interst rate   421
Fully Paid    high income     high interst rate     2857
                              low interst rate      465
                              medium interst rate   1522
              low income      high interst rate     6564
                              low interst rate      1225
                              medium interst rate   4451
              medium income   high interst rate     8187
                              low interst rate      1764
                              medium interst rate   5110
```

Recommendation

- Recommendation to the Loan lending company is to, do more scrutiny or Investigation for the applicants who are ready to take loan with high interest rates and with Low or Medium income

- Don't consider the defaulters to give loan again.

- Lend less loan with more EMI terms to the applicants with low or Medium income after Investigation.
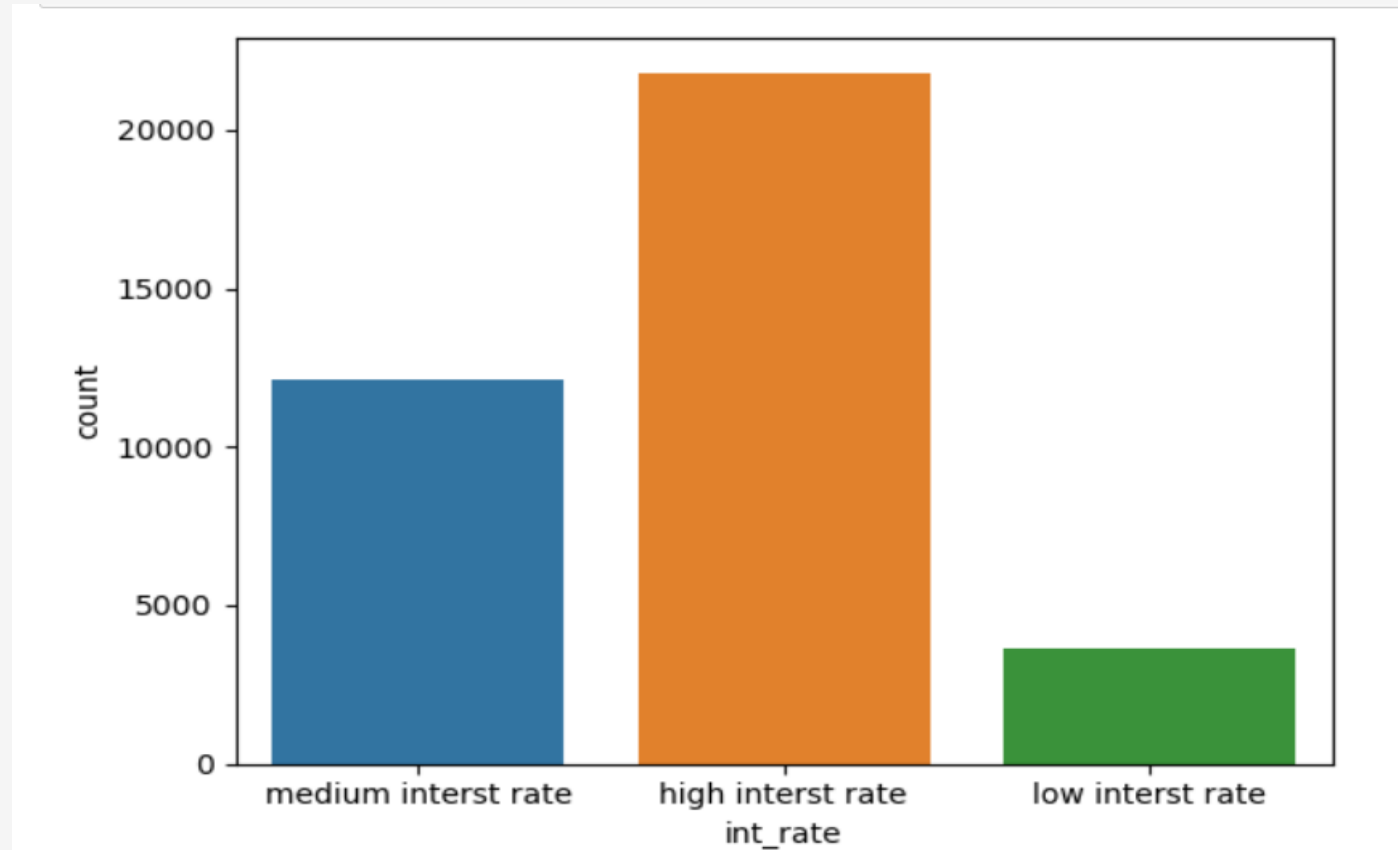
Data Visualization

- Univariate Analysis:

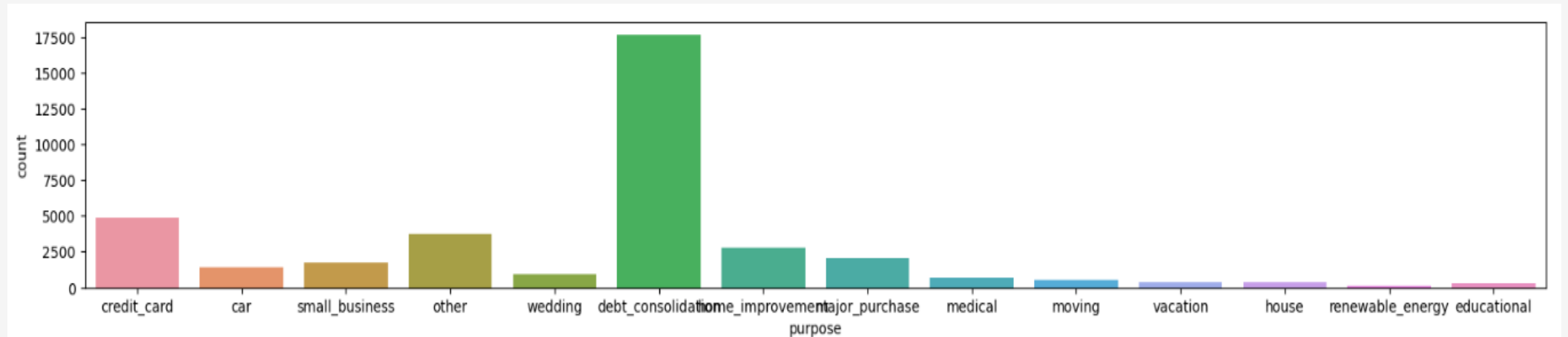- Count of Loan status with Charged Off status is around 5000
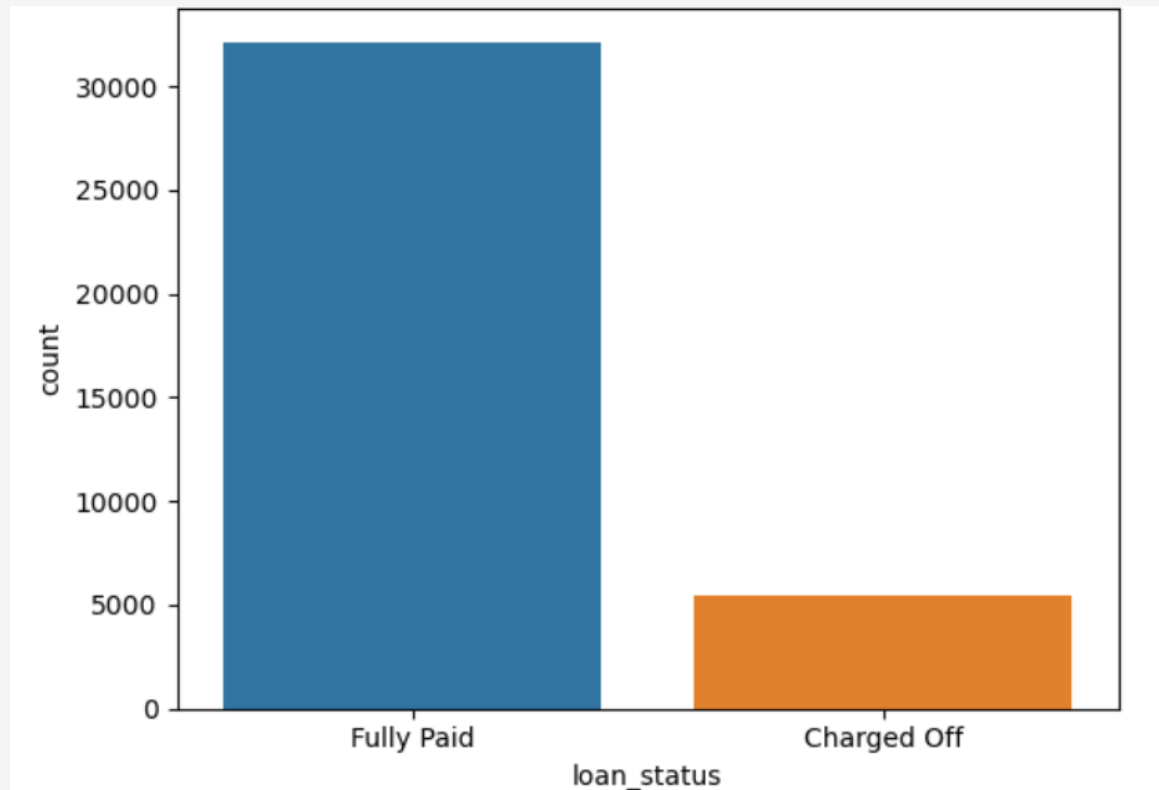
Data Visualization

- Univariate Analysis:

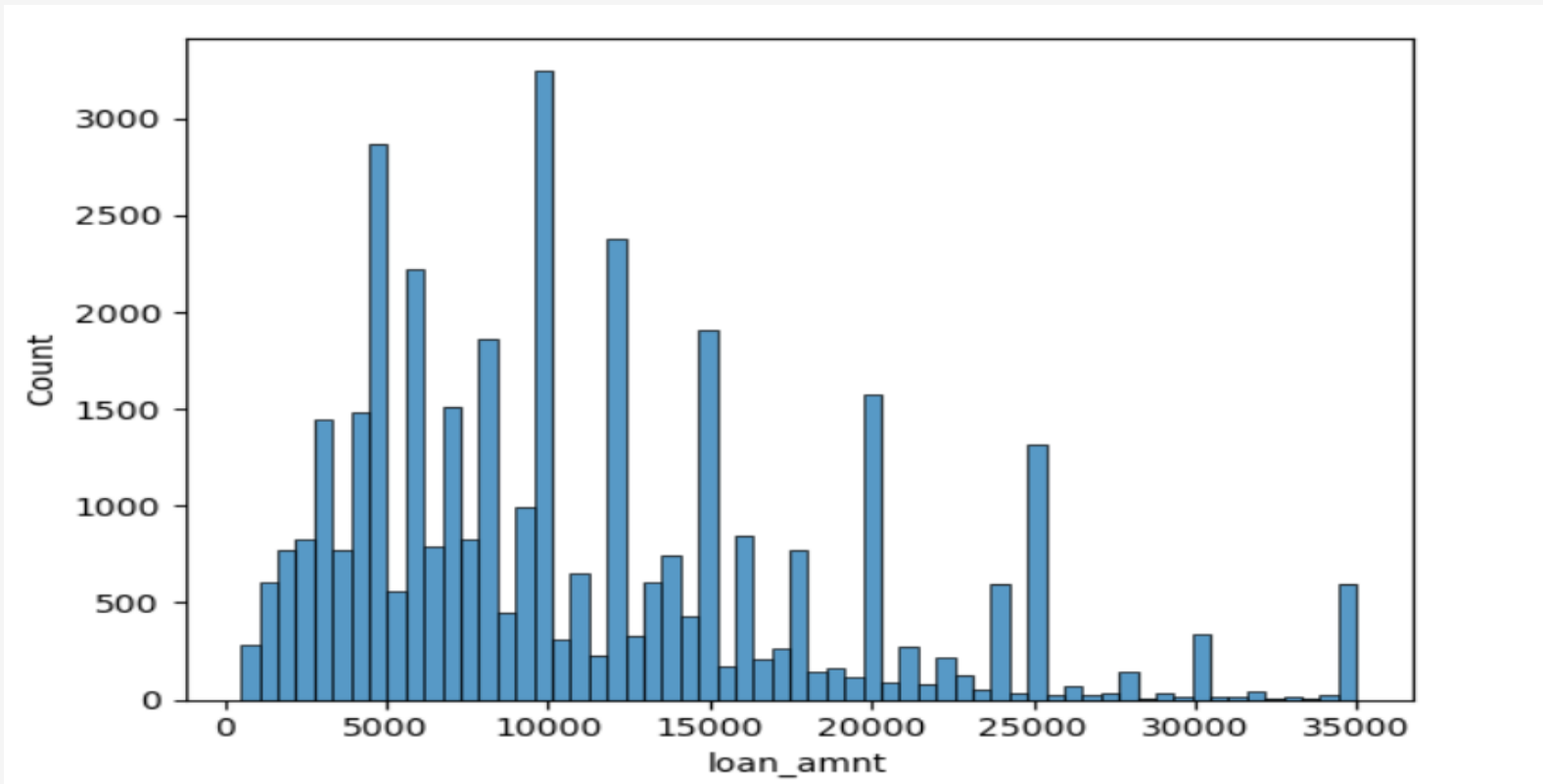- Count of applicants with High interest rates are more than Medium interest and Low interest rates

Data Visualization

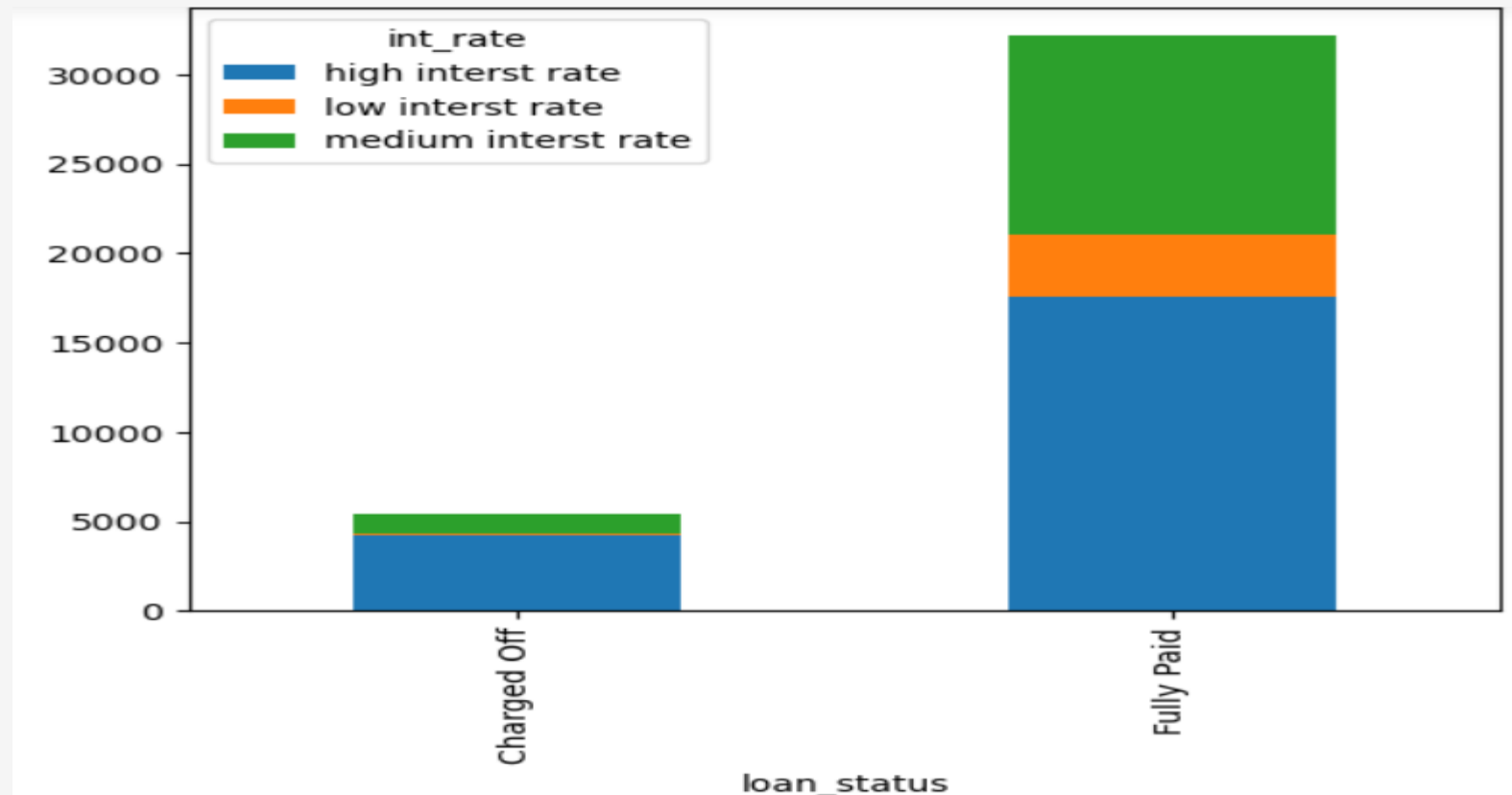- Univariate Analysis:

- Data visualization on Purpose variable

Data Visualization

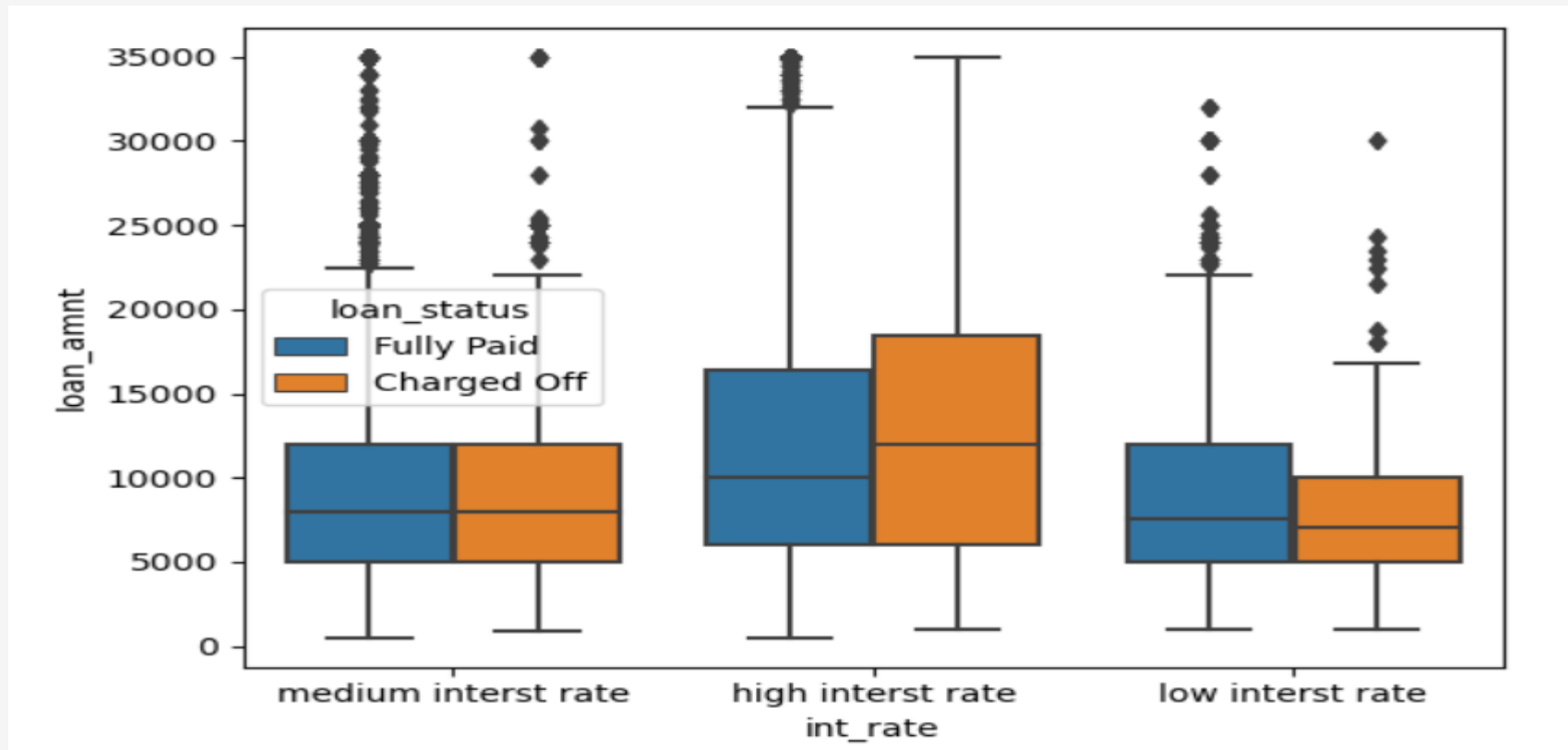- Univariate Analysis:

- Histogram plot on Loan Amount

Data Visualization

- Bivariate Analysis:

- Stacked Bar plot on loan status and Interest rate. Which shows applications with High Interest rate are defaulters

Data Visualization

- Multivariate Analysis:

- Box plot on Interest Rate, loan amount and loan status, which shows High and low interest rates are defaulters

Thank You