Lending Club Case Study

Analysis on Loan Status Charged Off Status

Basic Details of Loan Dataset

- **Shape of Dataset - (39717,111)**

- **Info of Dataset - {dtypes - int64(13), float64(74), object(24)}**

- **Null Values on (Axis = 1) = 0**

- **Null Values on (Axis = 0) = 54**

- **Duplicated on id = 0**

- **Dropping columns which are not important for the analysis**

- **Dropping Column which contains more than 30% blank rows**

- **Filtering out data for loan status is current as they are still paying EMIS**

- **After Dropping Blank Columns shape - (38577,22)**

- **Columns Need to be taken care –**

  - Pub_rec_bankruptcies

  - Emp_length

- **Columns for Analysis**

  - Loan Amount

  - Funded Amount

  - Interest Rate

  - Loan Status

  - Purpose

  - Total Payment

  - Annual Income

Data Cleaning

► **Data Cleaning on Columns –**

  ► Employment Length – Only Numeric should be available

  ► pub_rec_bankruptcies – Fill with Not Known at Blank Places

  ► Revol Util – Remove % Symbol

  ► Interest Rate - Remove % Symbol

  ► Columns to Numeric data types –

    ► loan_amnt

    ► funded_amnt

    ► int_rate

    ► funded_amnt_inv

    ► Installment

    ► annual_inc

    ► Dti

    ► emp_length

    ► total_pymnt

► **Data Cleaning on Columns –**

  ► Create Two Separate columns – Year and Month on Loan issued date

  ► Remove Few more Columns -

    ► Application Type

    ► Policy Code

    ► Instalment

    ► Payment Plan

    ► Initial Status

  ► Create more columns for Derive analysis

    ► Loan Amount Category

    ► Annual Income Category
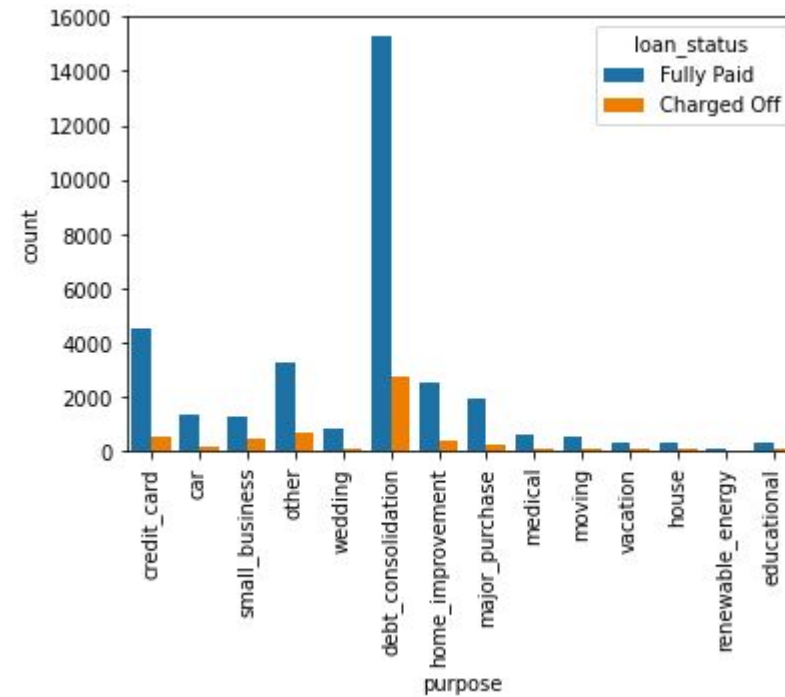
    ► Interest Rate Category

Data Analysis

Loan Status – Fully Paid, Current and Charged Off

```
Fully Paid      32950
Charged Off      5627
Name: loan_status, dtype: int64
```
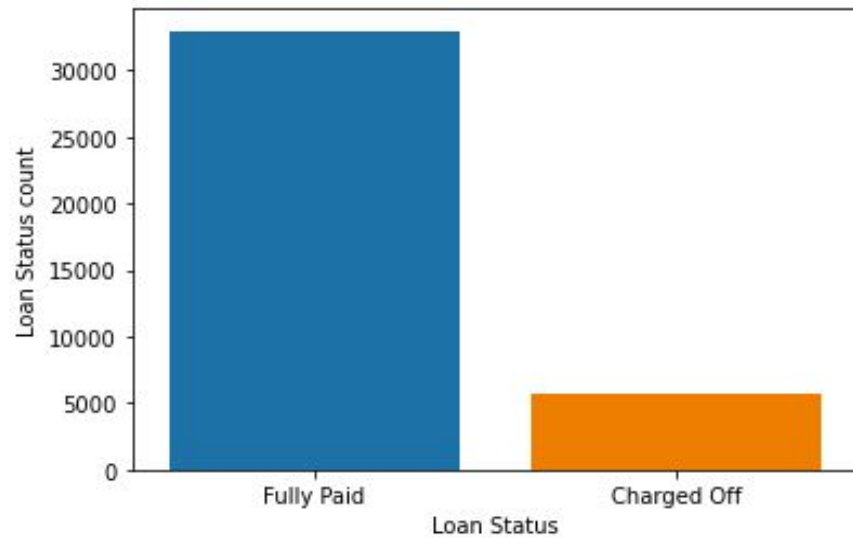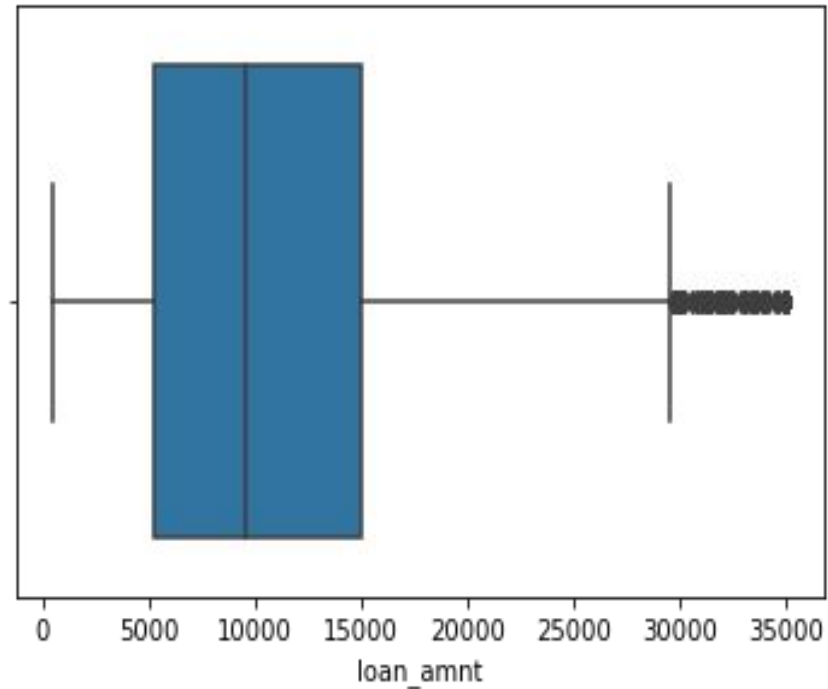

Plot for Purpose column – debt consolidation is the purpose in which charged off and current status loans count are low
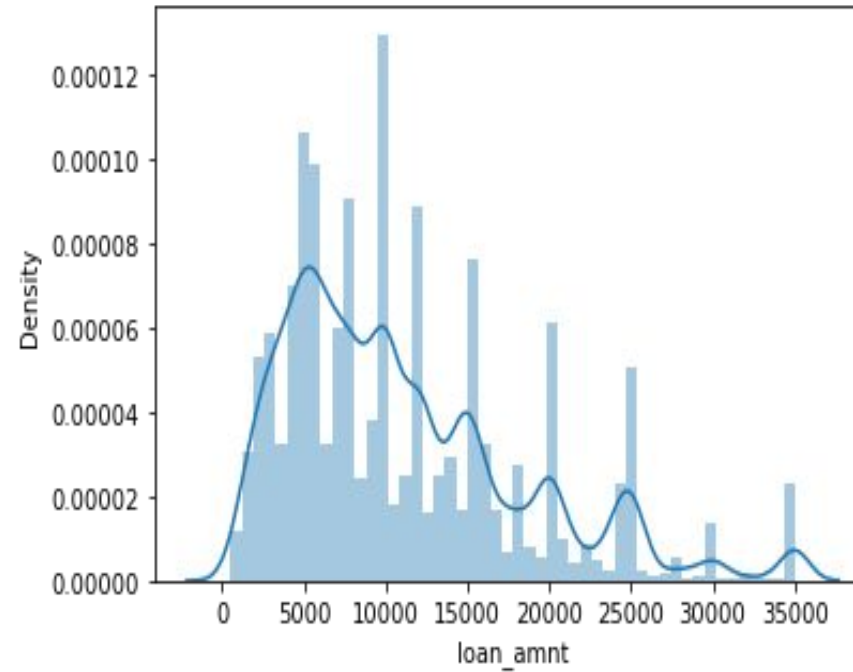
Univariate Analysis

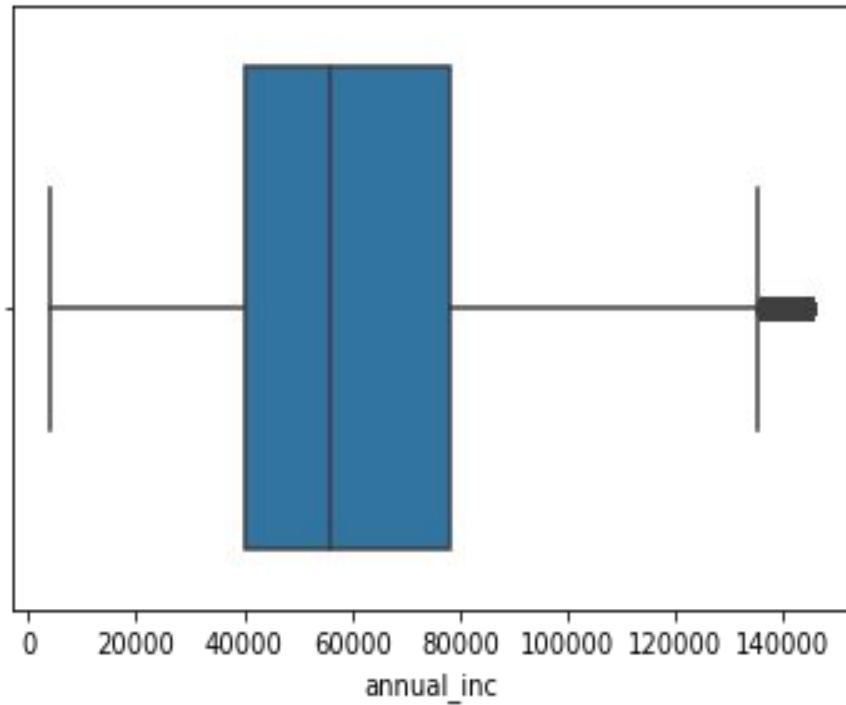**Box Plot for Loan Amount – Ranges between 5000-15000**
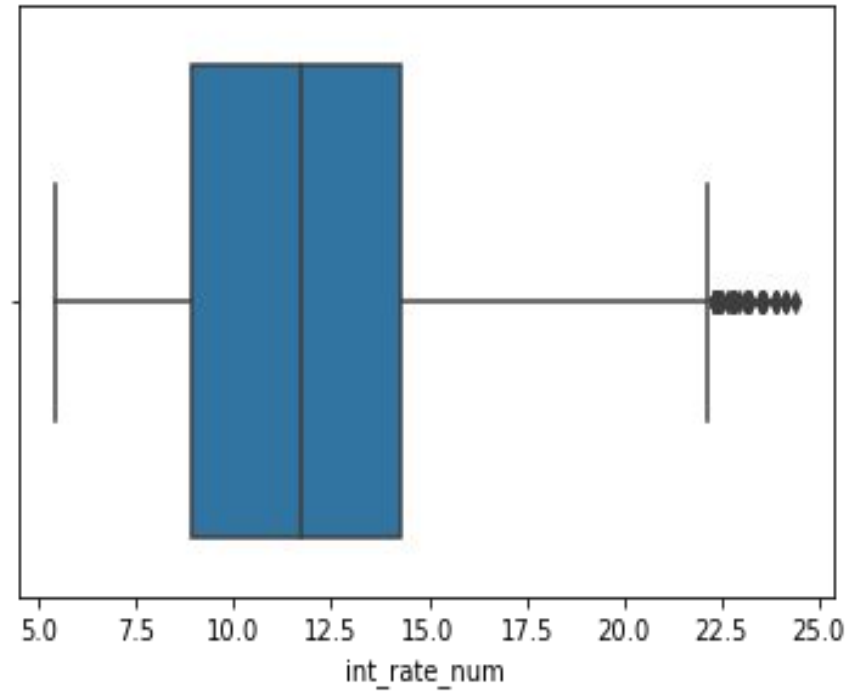
Dist Plot for Loan Amount

Univariate Analysis

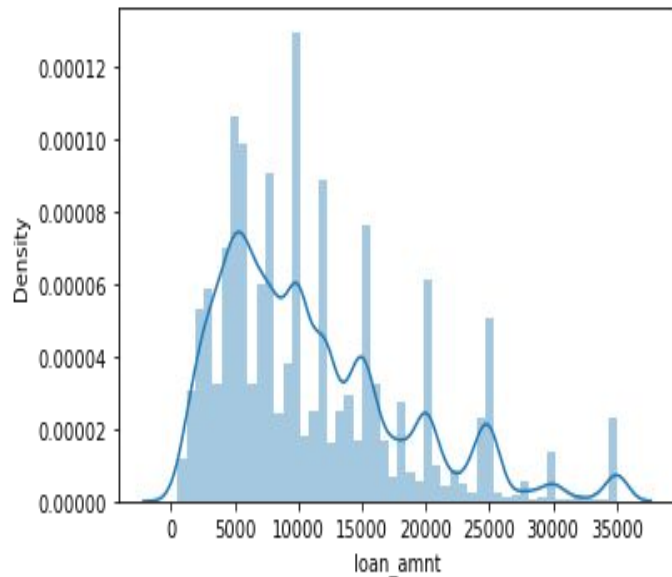**Box Plot for Annual Income – Ranges between 40000-80000**
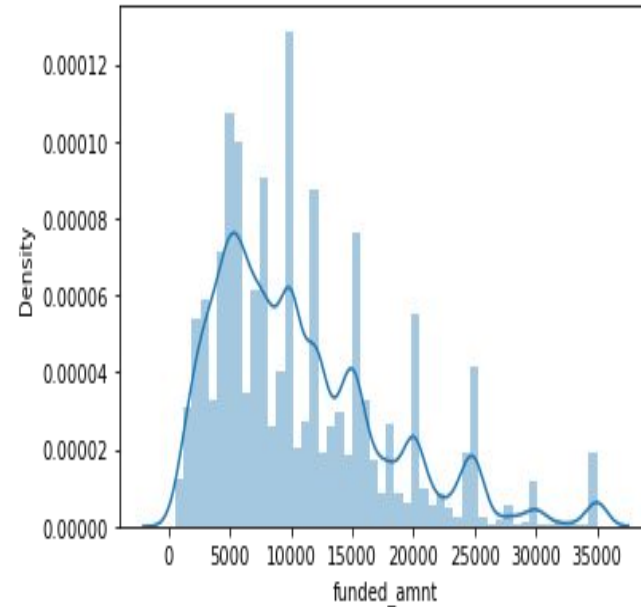
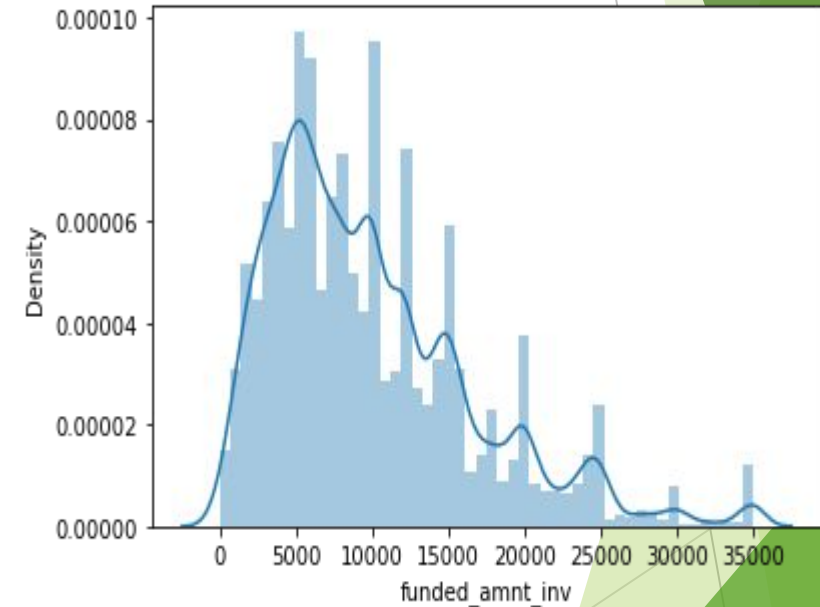**Box Plot for interest rate – ranges between 8%-14%**

Univariate Analysis

Distribution Plot for Loan Amount
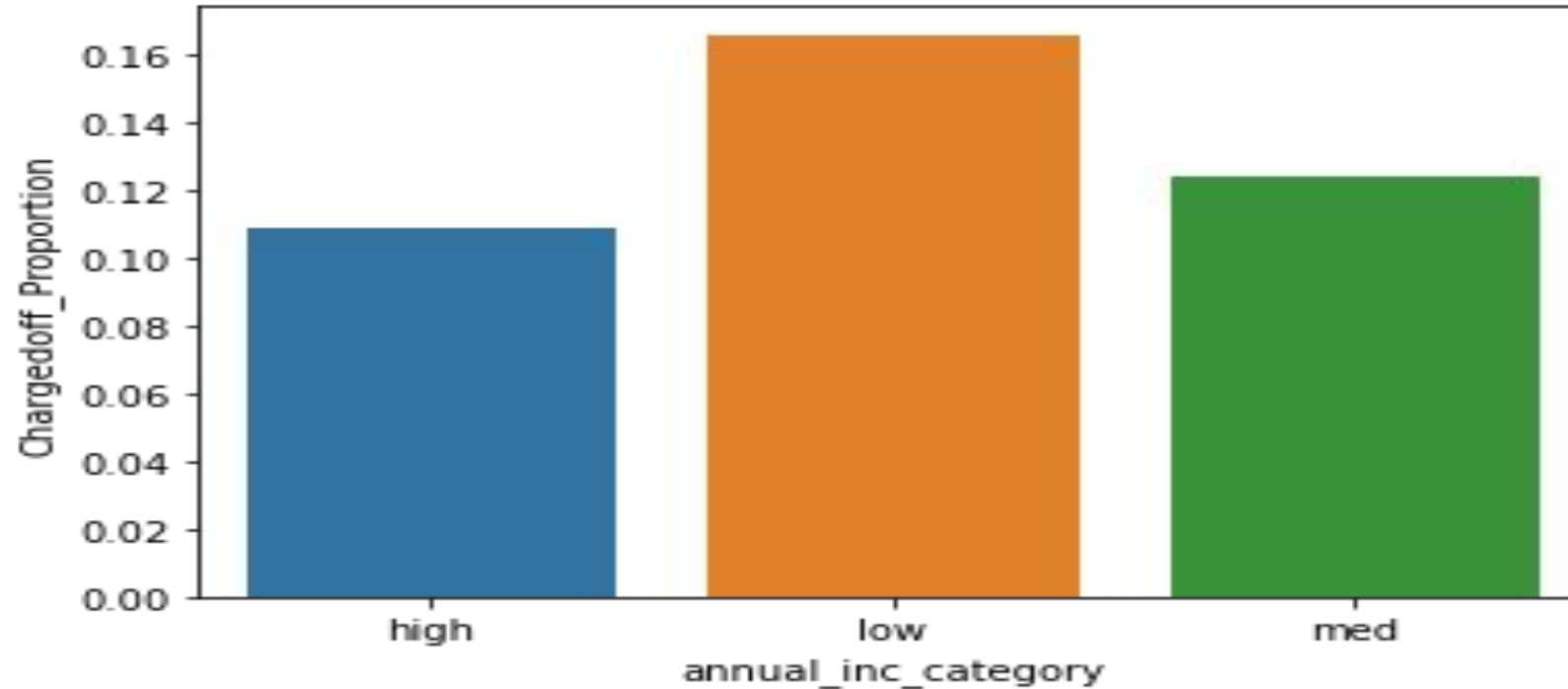
Distribution Plot for Funded Amount

Distribution Plot for Funded Amount invert

Loan Amount Funded Amount and Funded Aount_invert are more or less same pattern so considering Loan Amount for Analysis

Bivariate Analysis

Bar Plot for the Charged OFF data with respect to Annual Income Category
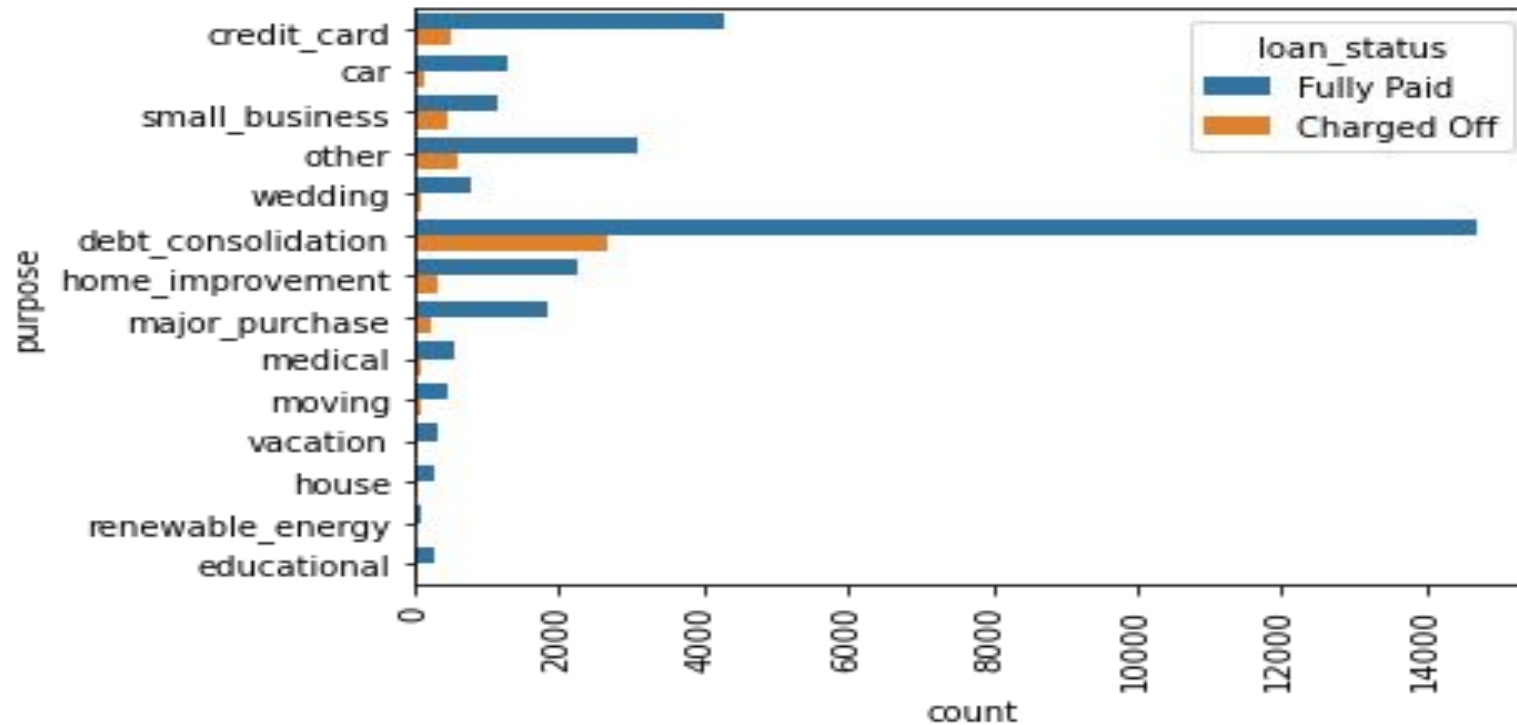

- ► Highlights –
    - ► Income Ranges 80000+ has less chances of charged off which come under high category
    - ► Income Ranges 0-2000 has high chances of charged off which come under low category
    - ► With increase of Annual Income Charged off possibility got decreased
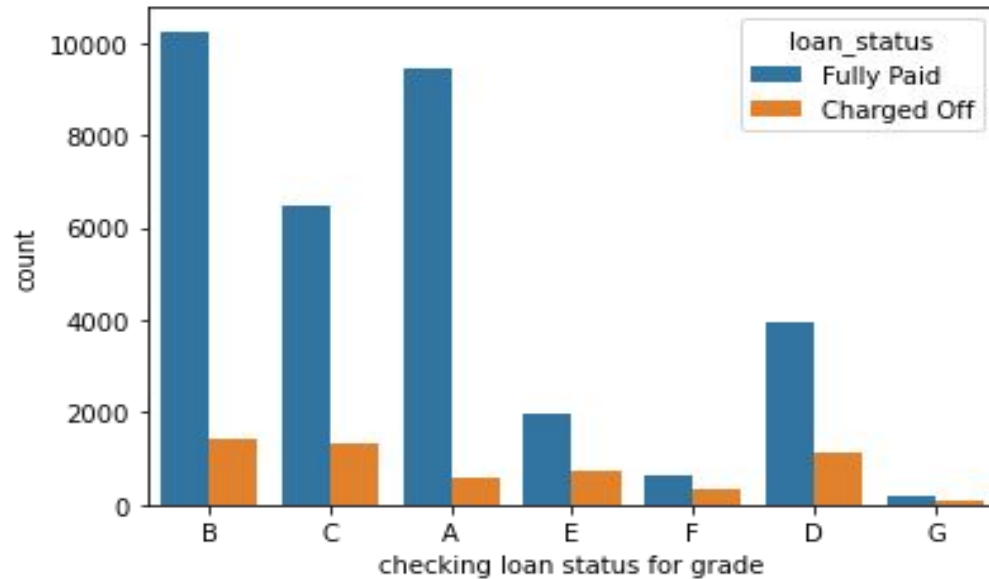
Bivariate Analysis

Bar Plot for the Loan status with respect to Purpose of Loan

- ► Highlights –
    - ► Renewal Energy charged off possibility is better than other category
    - ► Small Business has high chances of charged off

Bivariate Analysis
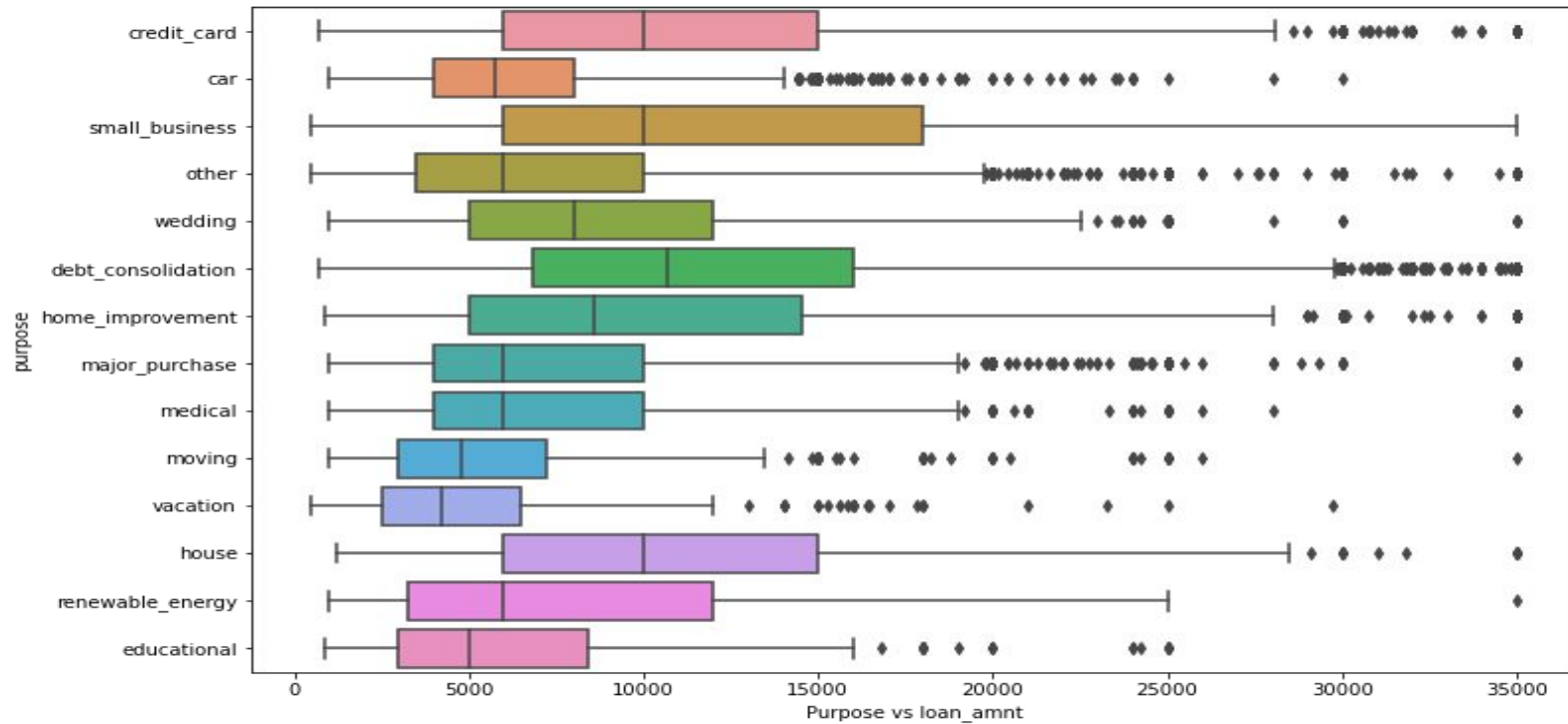
Bar Plot for the Loan status with respect to Grade System

- ► Highlights –
  - ► Grade A has high chances of charged off
  - ► Grade F and G has high chances of charged off
  - ► Pattern Shows that chances of charged off increases moving towards from A to G

Bivariate Analysis
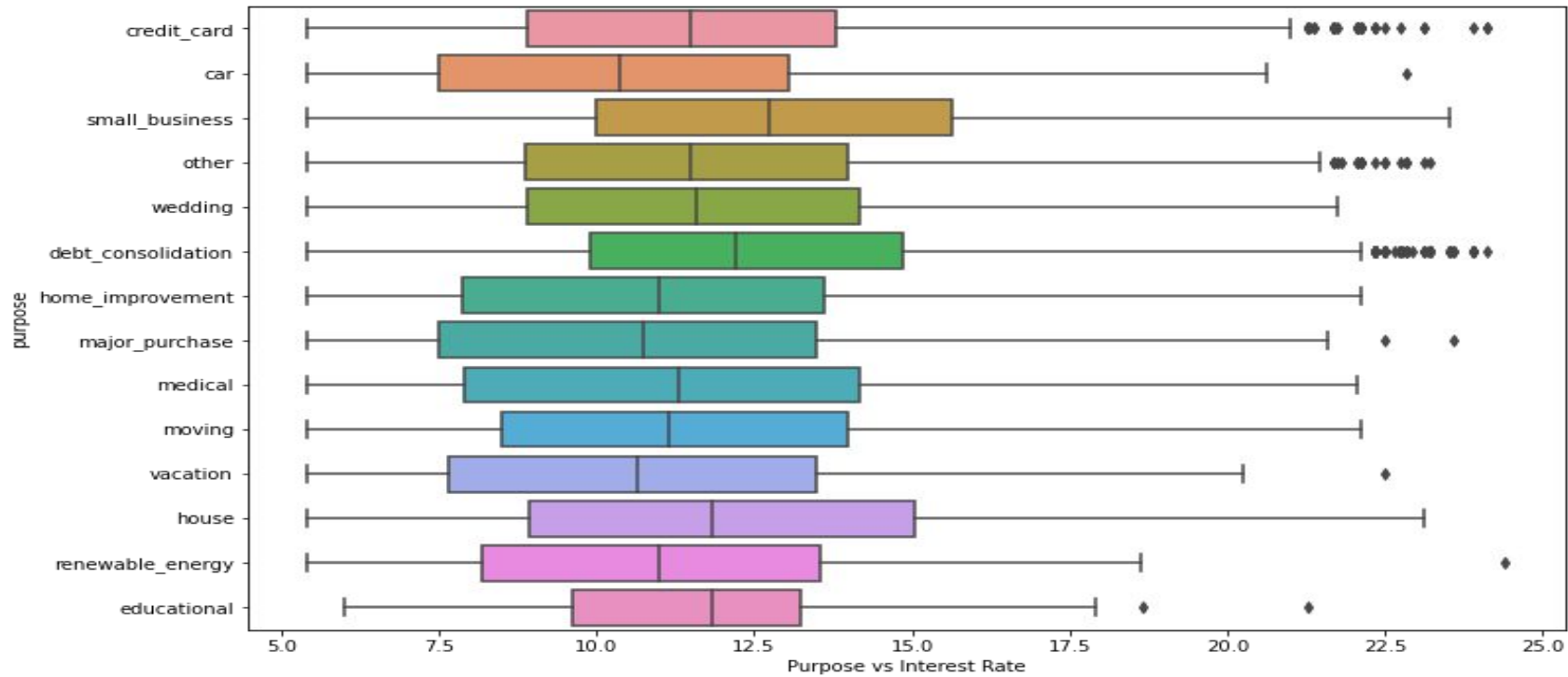
Box Plot for the Loan Amount vs Purpose of Loan

- ► Highlights –

  - ► Debt Consolidation is second and Credit Card is the 3rd reason to get Purpose of Loan

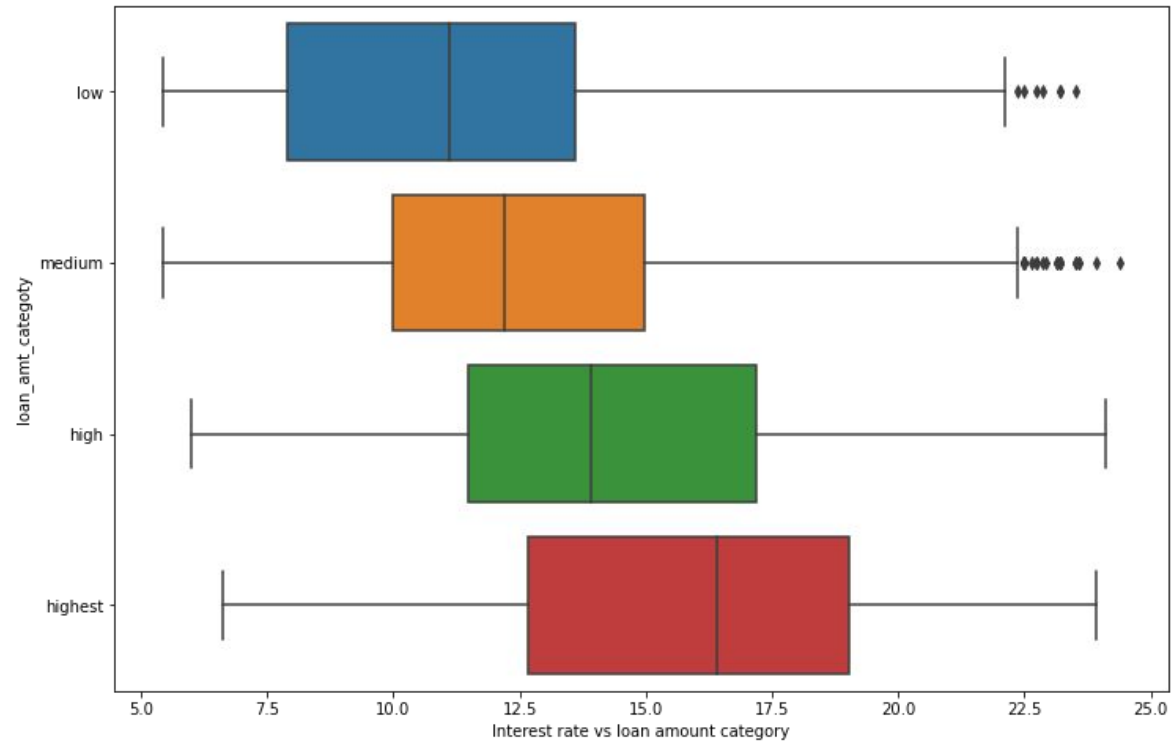 - ► Small Business is the 1st reason to get the Loan

Bivariate Analysis

Box Plot for the interest rate vs Purpose of Loan

- ► Highlights –
  - ► Average Interest Rate is higher for Small Business
  - ► Loans taken for small business had to repay with more interest rate

Bivariate Analysis

Box Plot for the interest rate vs Loan Amount

► Highlights –

   ► Interest Rate is increases with Loan Amount