International Institute of Information Technology Bangalore

upGrad

LENDING CLUB CASE STUDY

Group Members:
Ajay Kumar Dabral
Abhishek Tiwary

PROBLEM STATEMENT

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:

1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

2. 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

The data given contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

The analysis has to be performed using EDA (Exploratory Data Analysis)

PROBLEM SOLVING APPROACH

Data Understanding and gathering Domain Knowledge

Data Cleaning

Analysis of Variables

Data Visualization

Conclusion

STEPS FOLLOWED

1. Data Understanding and gathering Domain Knowledge
2. Data Sourcing: Loading the Dataset
3. Data Cleaning:
    i) Fixing rows and columns
    ii) Fixing missing values
    iii) Standardize values i.e. converting to suitable format.
    iv) Outliar detection and removal
4. Univariate Analysis
5. Segmented Univariate Analysis
6. Bivariate Analysis
7. Derived metrices
8. Multivariate Analysis: Correlation and Heatmap of Variables

Data Understanding and Gathering Domain Knowledge

After cleaning the dataset we get 3 types of variables:

1. **Customer's Demographic Information:**

   These information is generally provided by the customer while taking loan from any financial institutions.

   e.g. emp_title, emp_length, home_ownership, annual_inc etc

2. **Loan Characterstics Information:**

   these are the attributes generally determined by the financial institutions and depends on the past history of the Borrower and purpose of loan.

   e.g. loan_amnt, funded_amnt, funded_amnt_inv, int_rate etc

3. **Credit information: Customer Behaviour variables:**

   These variables are not associated in identifying the default as they come in picture when the loan is approved.

   e.g. delinq_2yrs, earliest_cr_line, inq_last_6mths, open_acc etc.

Out of above three types Credit Information: Customer Behaviour Variables are least helpful in identifying loan defaults

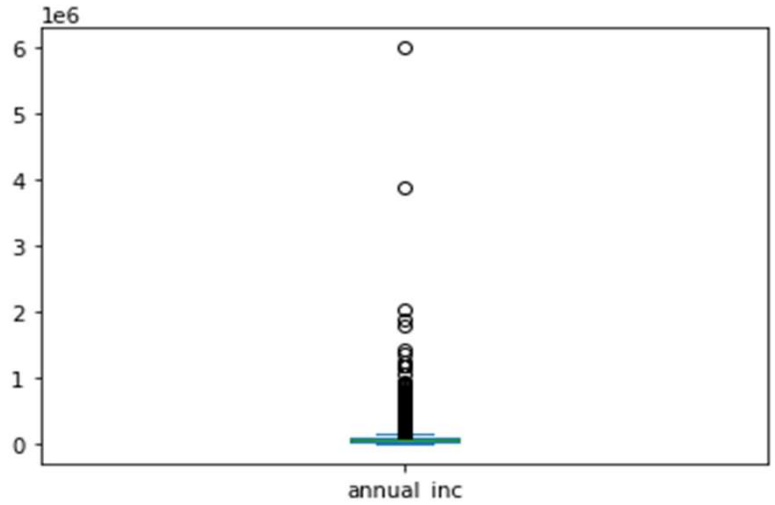
Data Loading and Cleaning

- Read loan.csv
- Dropping columns having all null values
- Dropping columns having same values for all Rows
- Dropping columns having too many missing values
- Dropping columns related to customer behaviour which do not help us in identifying defaulters
- Imputing missing values in "pub_rec_bankruptcies", "emp_lenth" and "emp_title"

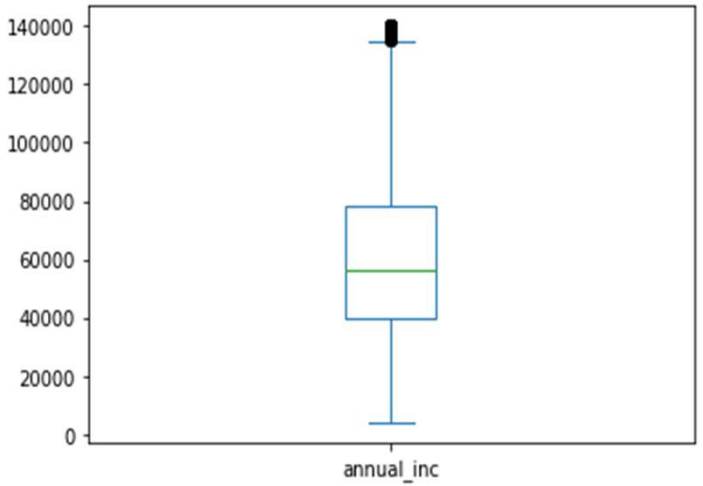
  replacing the missing values with mode.
- Dropping the rows which were having missing values in 'revol_util‘ field

Data Loading and Cleaning

- Removing outliers in annual_inc



Before removing outliers

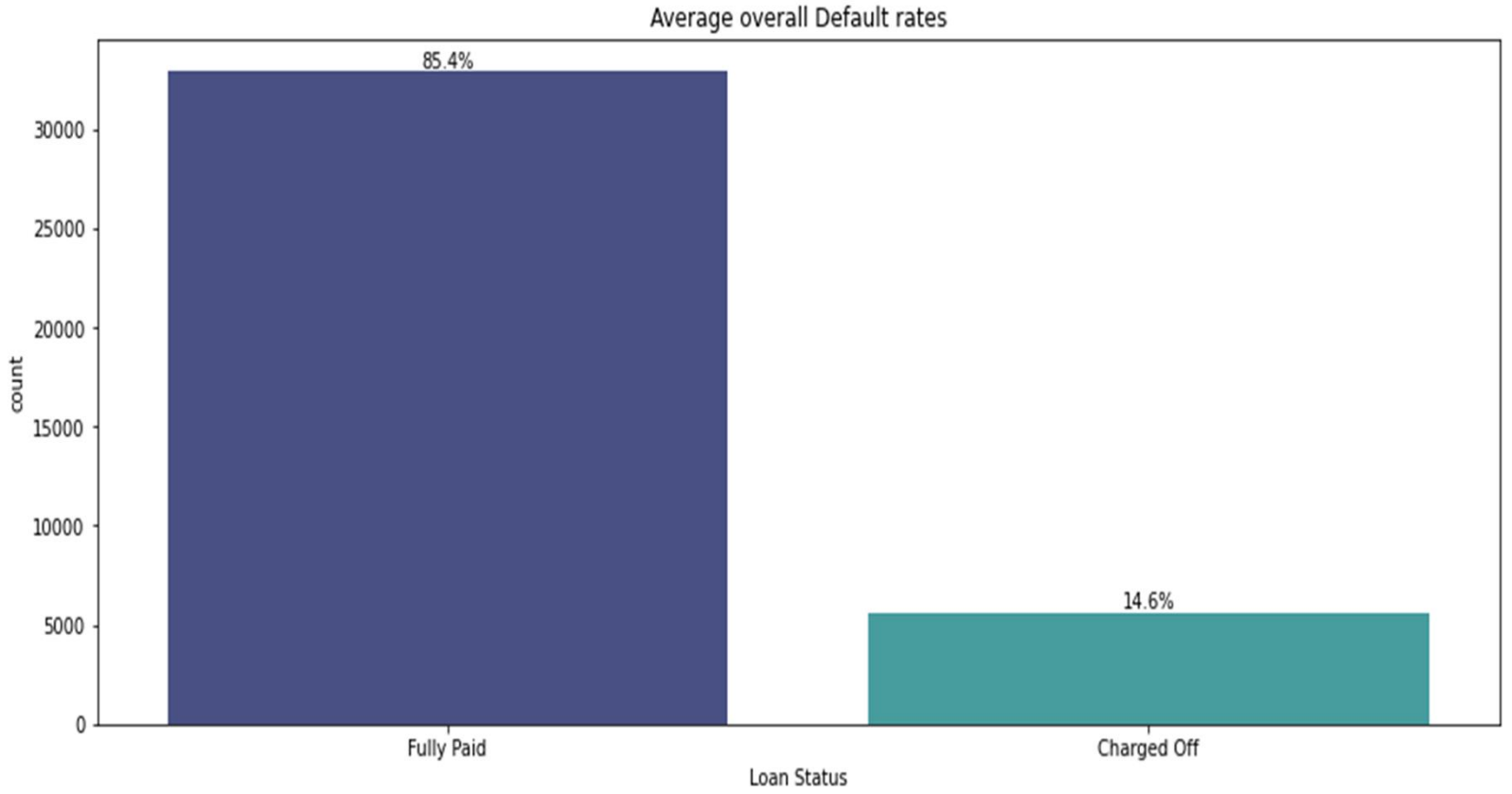


After removing outliers

- After removing outliers from annual income outliers from dti,loan_amnt, funded_amnt_inv are also removed.

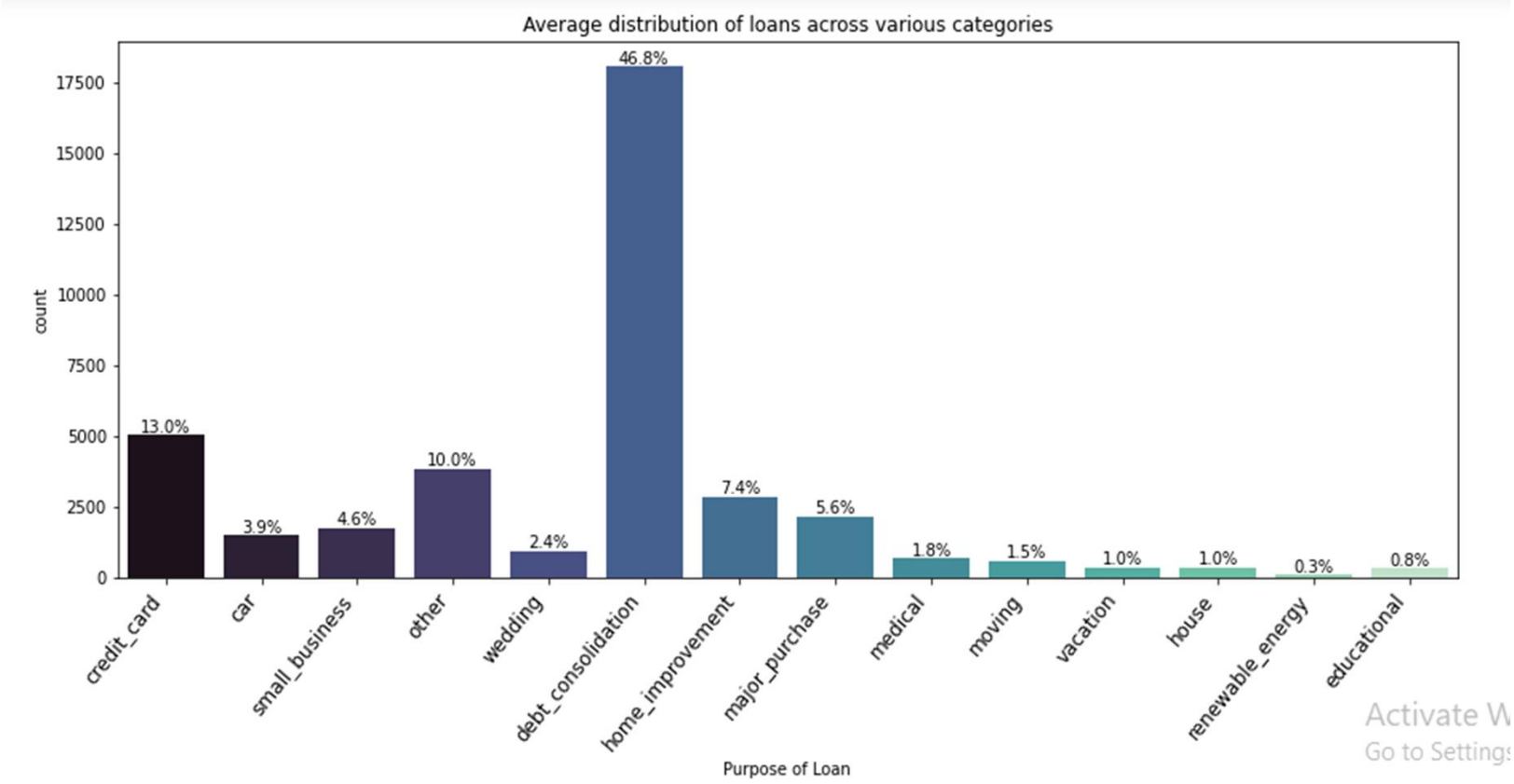
Conclusions from Univariate Analysis

- Number of Borrowers who have fully paid Loans are more than 5 times the customers who defaulted



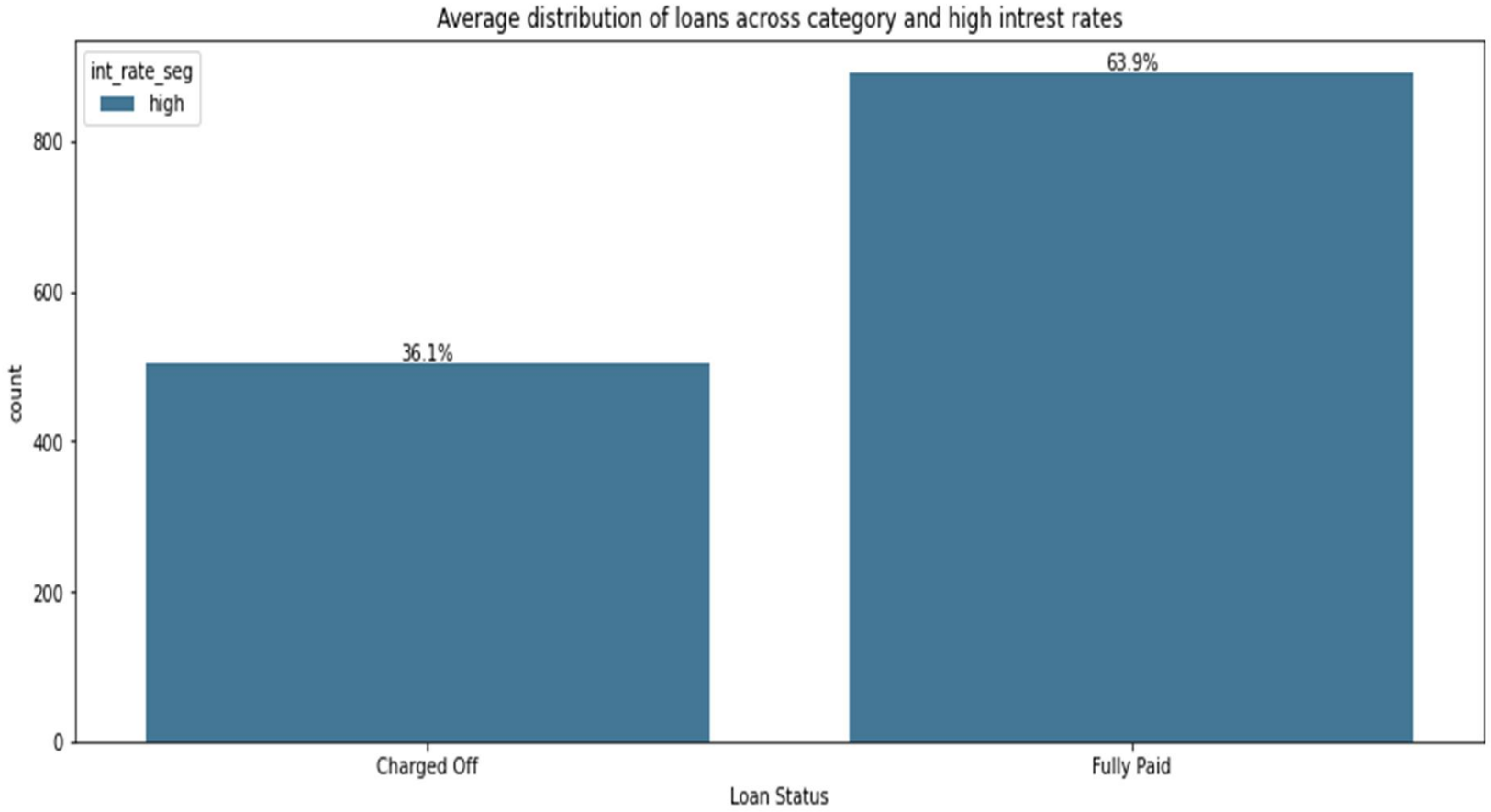
Conclusions from Univariate Analysis

- Maximum number of Borrowers have taken loans for Debt Consolidation.



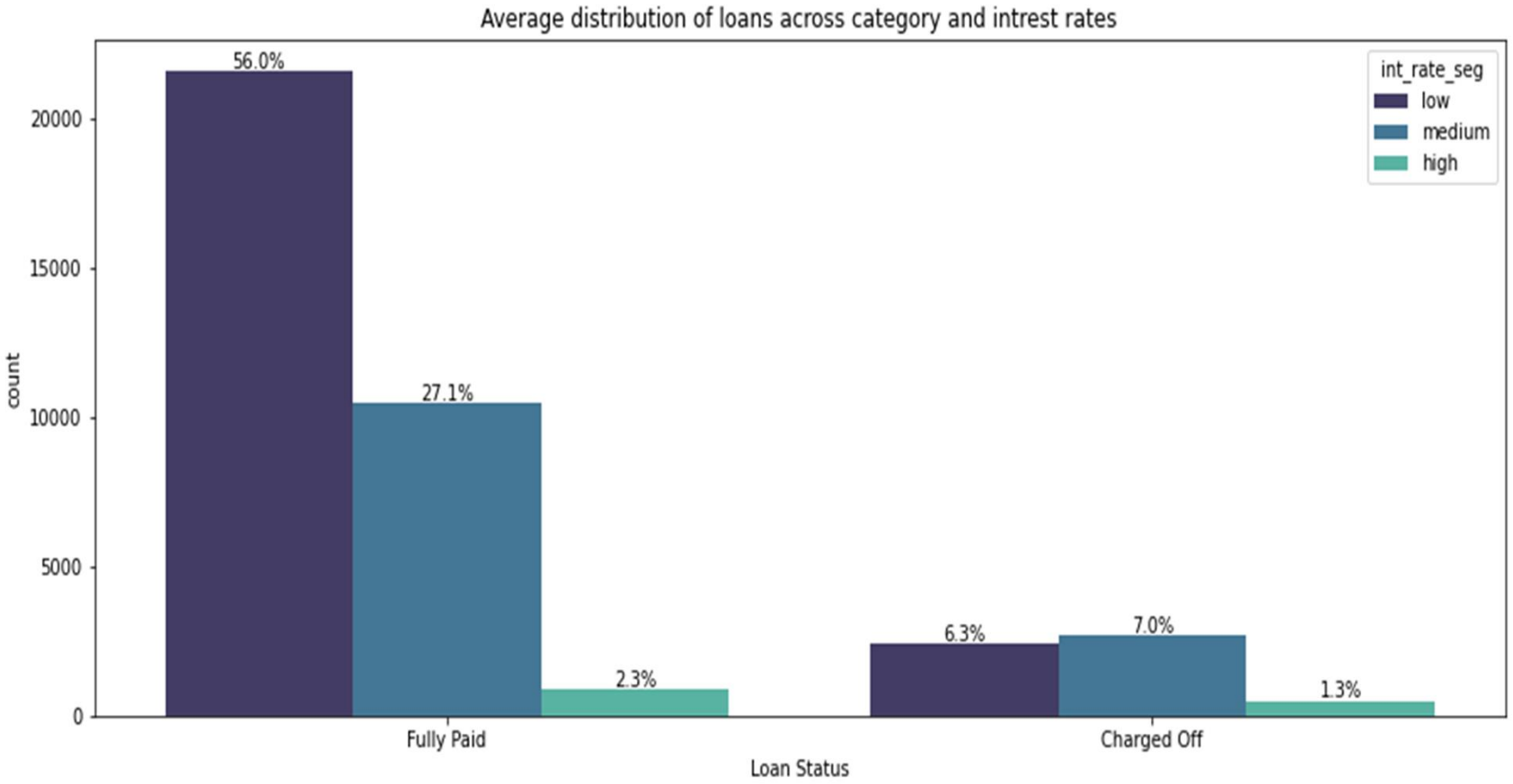
Conclusions from Univariate Analysis

- Borrowers who have taken loans at high interest rates have high chances of default



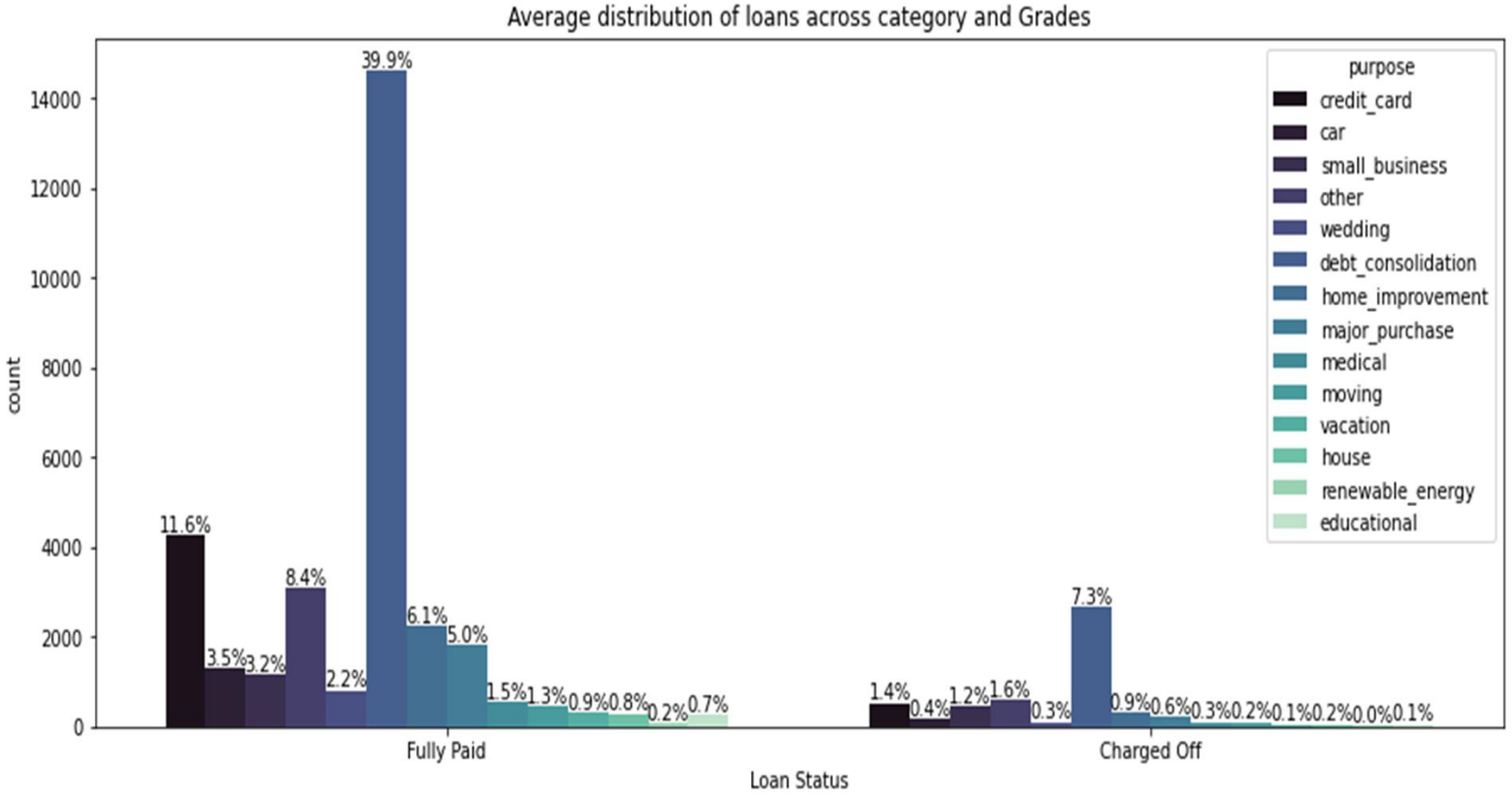
Conclusions from Bivariate Analysis

- More loans were given at low interest rates, and lesser loans were given at high interest rates



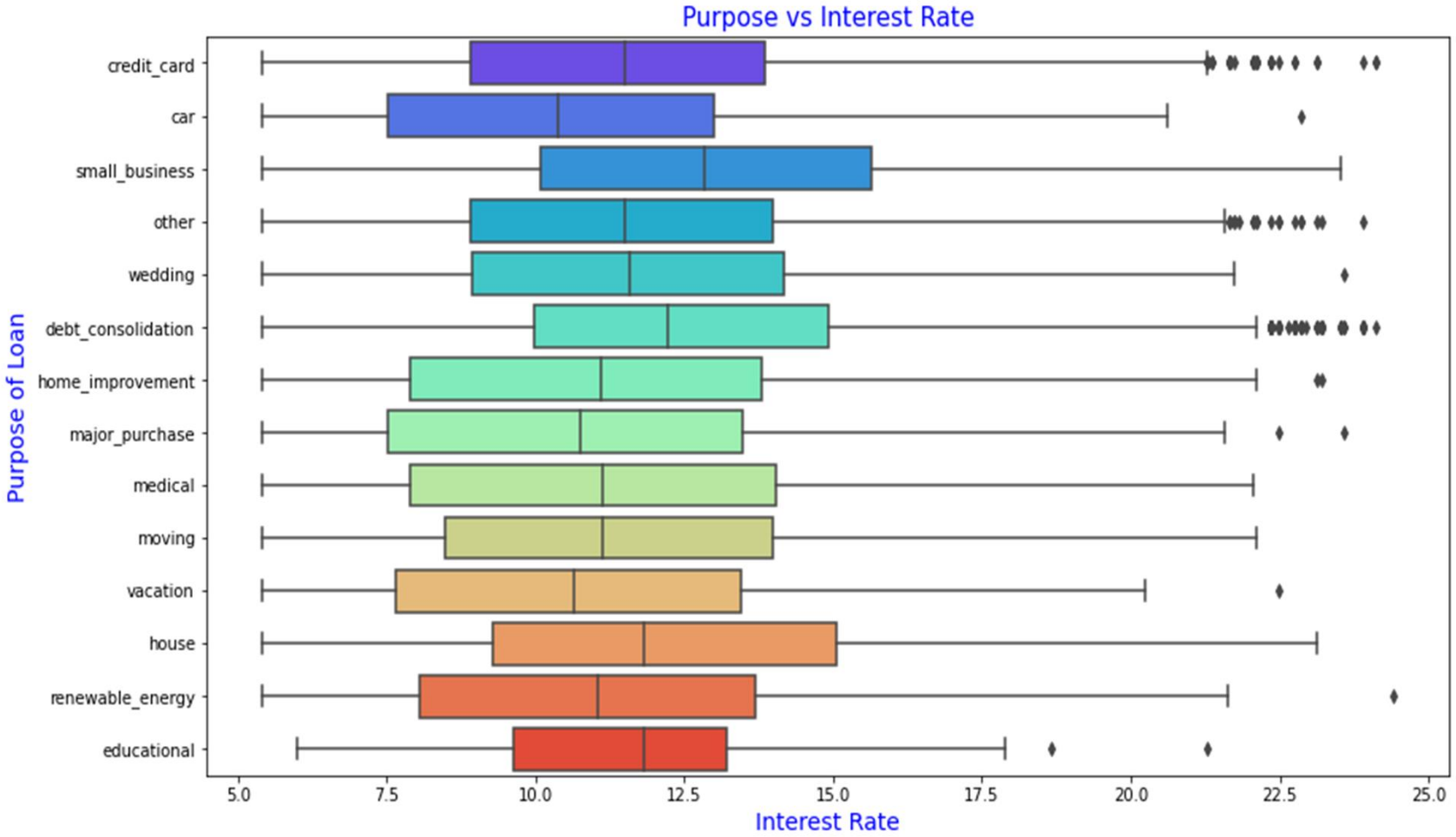
Conclusions from Bivariate Analysis

- Borrowers who have taken loans for Debt Consolidation and Credit card are more prone to loan defaults.



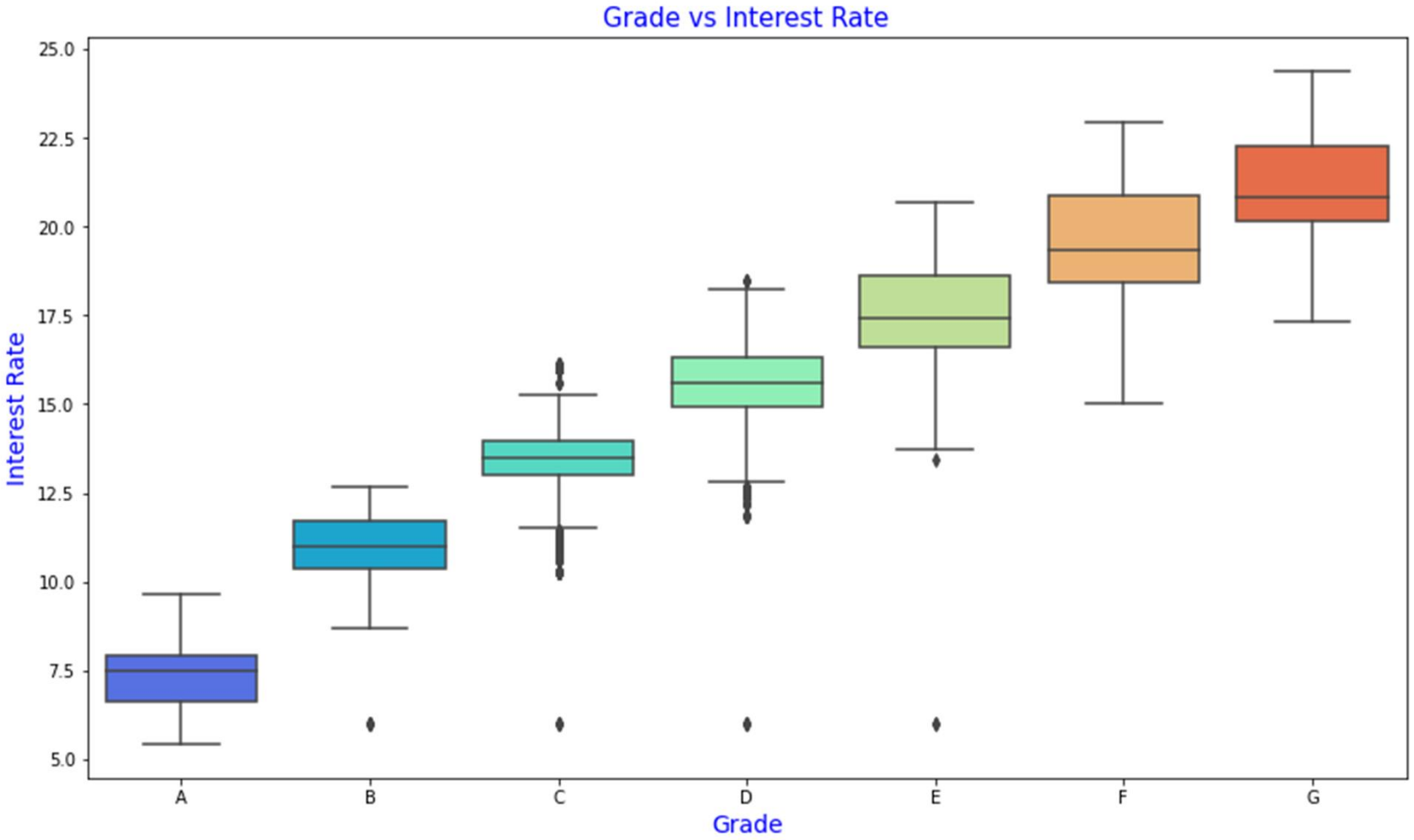
Conclusions from Bivariate Analysis

- High interest rates are charged for small business loan and house loan



Conclusions from Bivariate Analysis

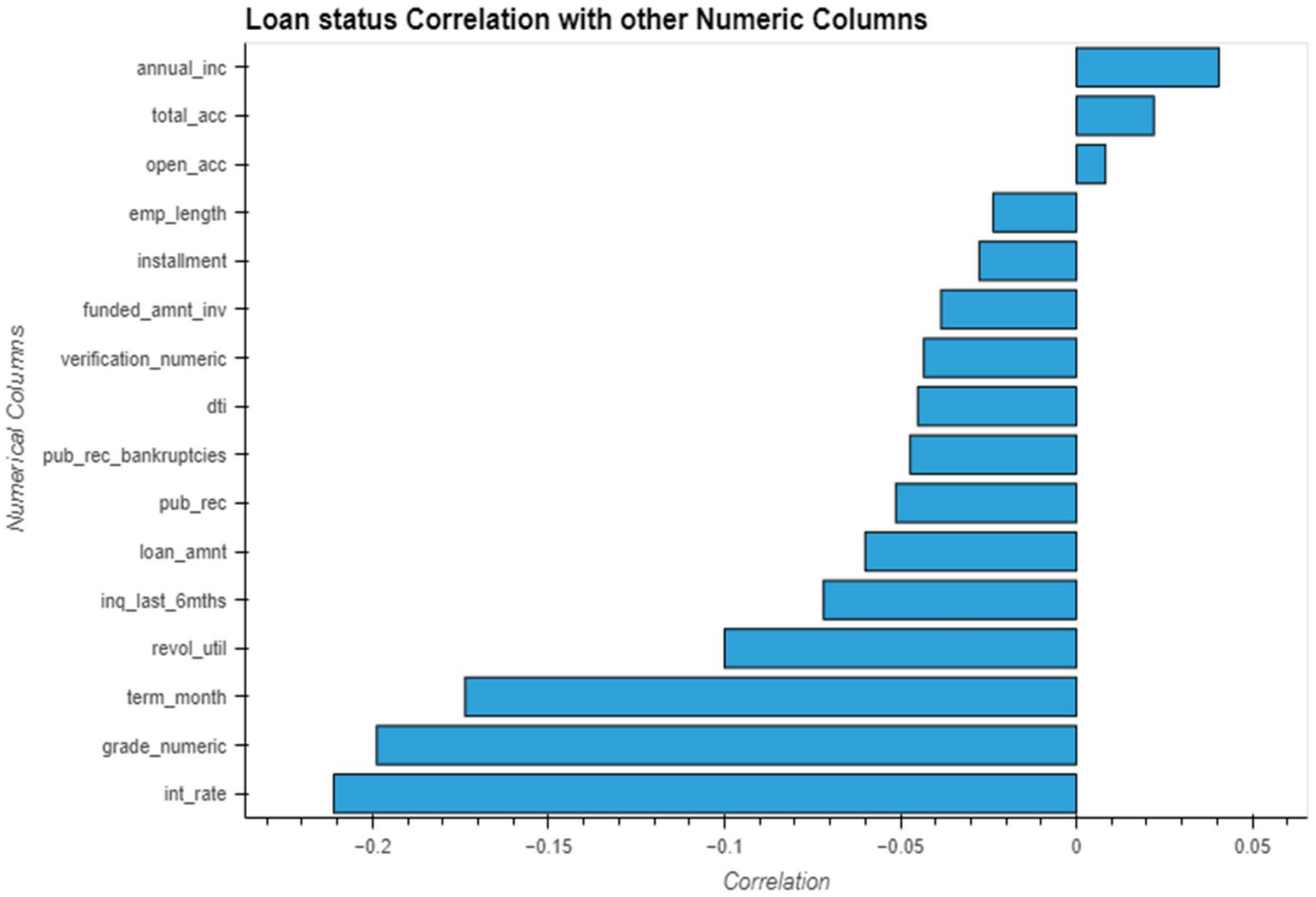
- Interest Rate increases as we move from Grade A to Grade G



Conclusions from correlations

- As annual income increases there are more chances of Fully paying loan
- Interest Rate increases as we move from Grade A to Grade G and so does the chances of default
- As the revolving credit utilization percentages of borrowers increases the loan default chances increases
- With increase in Term of loan, chances of default also increases.

Conclusions from correlations



Important Variable in Identifying Defaulters

1. Grades
2. Interest Rate
3. Revolving Credit Utilization
4. Annual Income
5. Purpose

Final Conclusions

- Approx. 15 % of Borrowers have defaulted on their loans, while 85 % have Fully paid their loans.
- As we move from Grade A to Grade G, interest rate increases and so does the loan defaults.

  This suggests that financial institution charges higher interest rates for borrowers having poor grades
- It is observed that the Borrowers having higher incomes are less likely to default on their loans
- Higher the revolving credit utilization percentage of a borrowers, higher number of loan default.
- Loans taken for the purpose of Debt consolidation and credit card have higher number of default.