Lending Club Case Study

Group Members

Sanjay M. Belgaonkar

Praneeth Reddy

Problem Statement

- We 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
- The aim of this case study 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.

Analysis approach

- In this case study, we have used EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- Following steps are followed in the analysis:
  - Overview of the data set
  - Data Cleaning
  - Handling of missing values
  - Addition of derived metrics
  - Univariate analysis
  - Bivariate analysis
  - Conclusion

Data Cleaning

Steps followed in the data cleaning

1. Find out how many empty rows and columns are there in the data set.
2. Find out which columns have all null values and remove them.
3. Drop the additional columns which are not considered for the analysis.

- After the data cleaning the number of columns reduced from 111 to 26.

Handling Missing Values

- We have dropped the columns having more than 90% of missing values.

- Then the number of columns reduced to 21.

Derived Metrics

- Year and Month columns are created using issue_d column.
- Categorical division of the following columns is done.
  - Loan Amount
  - Annual Income
  - Interest Rate
  - Debt to Interest Ratio

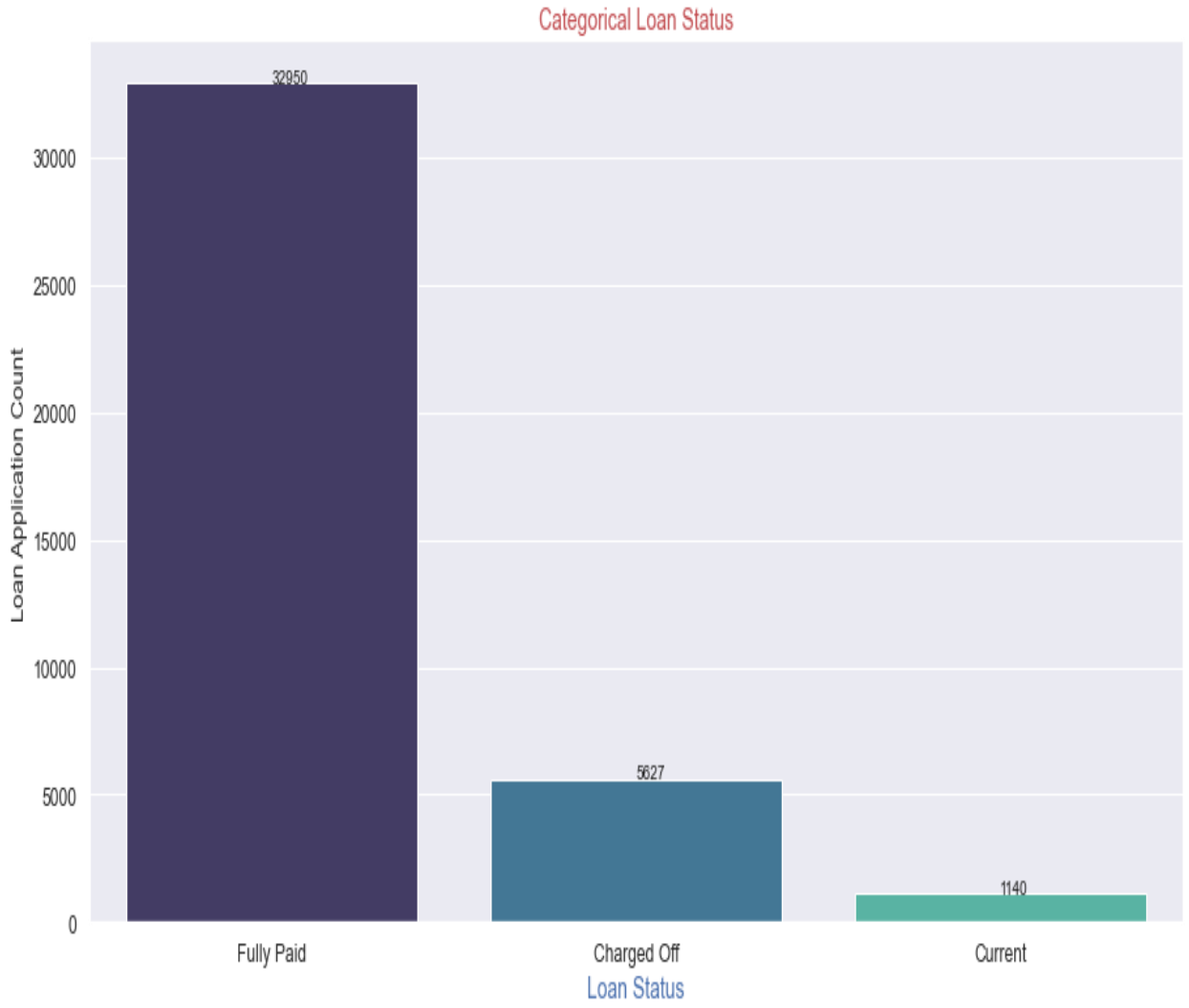
Univariate Analysis

- Following variables are considered for the Univariate analysis.
  - Loan Status
  - Loan Amount
  - Funded Amount
  - Investors' Funded Amount
  - Annual Income
  - Term of loan
  - Employment Length

Observations from Univariate Analysis

- Loan amount is in the range 5000 to 15000.
- Around 14.16% loans were charged off out of total loan issued.
- Loan status is

Fully Paid	83 %
Charged off	14 %
Current	3 %



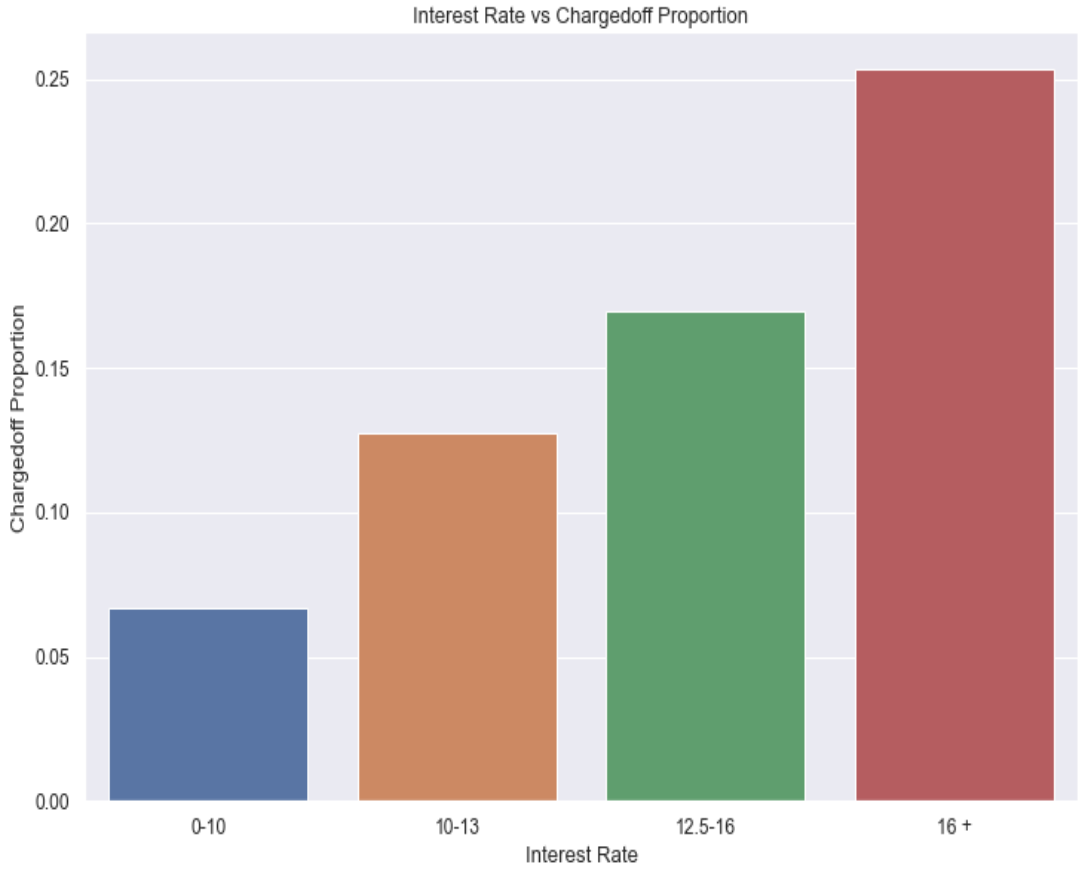
Bivariate Analysis

- Following are the variables considered for bivariate analysis.
  - Purpose of loan
  - Loan amount
  - Interest rate
  - Annual income
  - Grades
  - Verification Status

Results of Bivariate Analyses

Observations from Bivariate Analysis

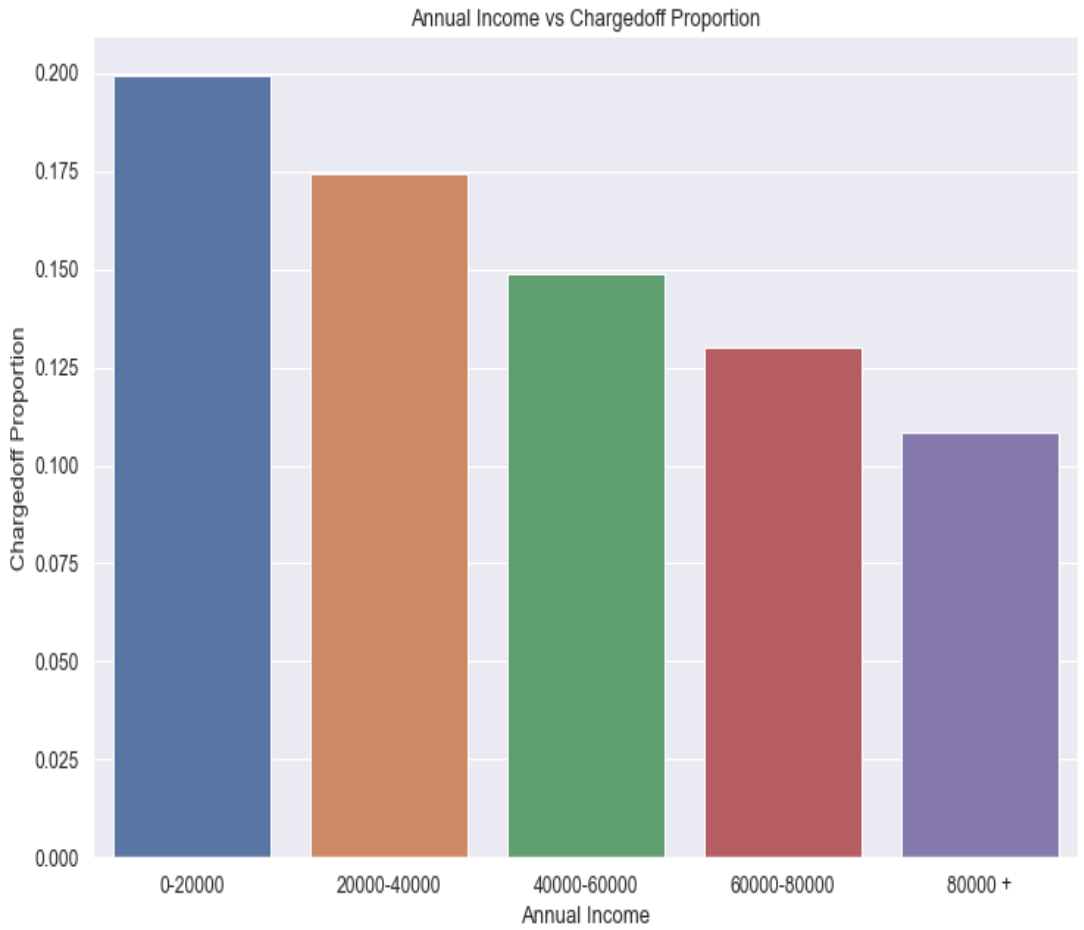
Interest rate vs. Charged off proportion



- Charged off proportion is increasing with higher interest rates.

Observations from Bivariate Analysis

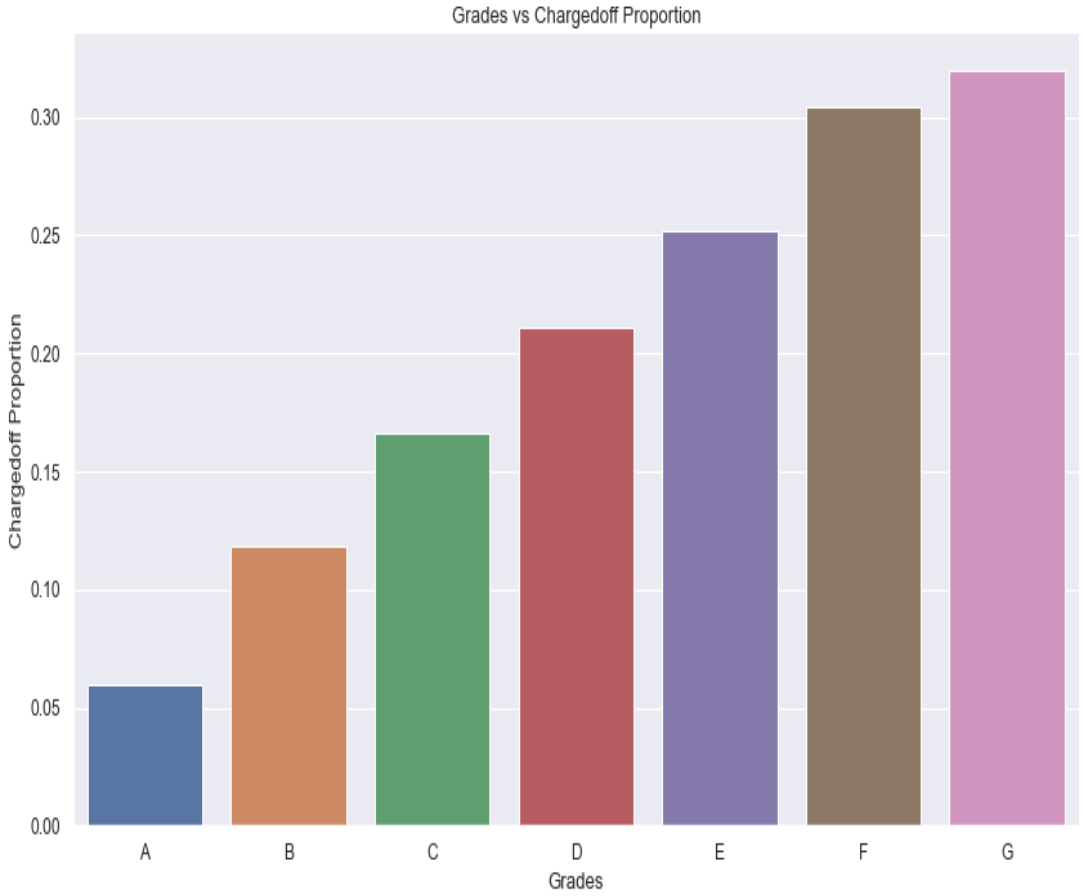
Annual income vs. Charged off proportion



- Income range 0-20,000 has high chances of charged off.

Observations from Bivariate Analysis

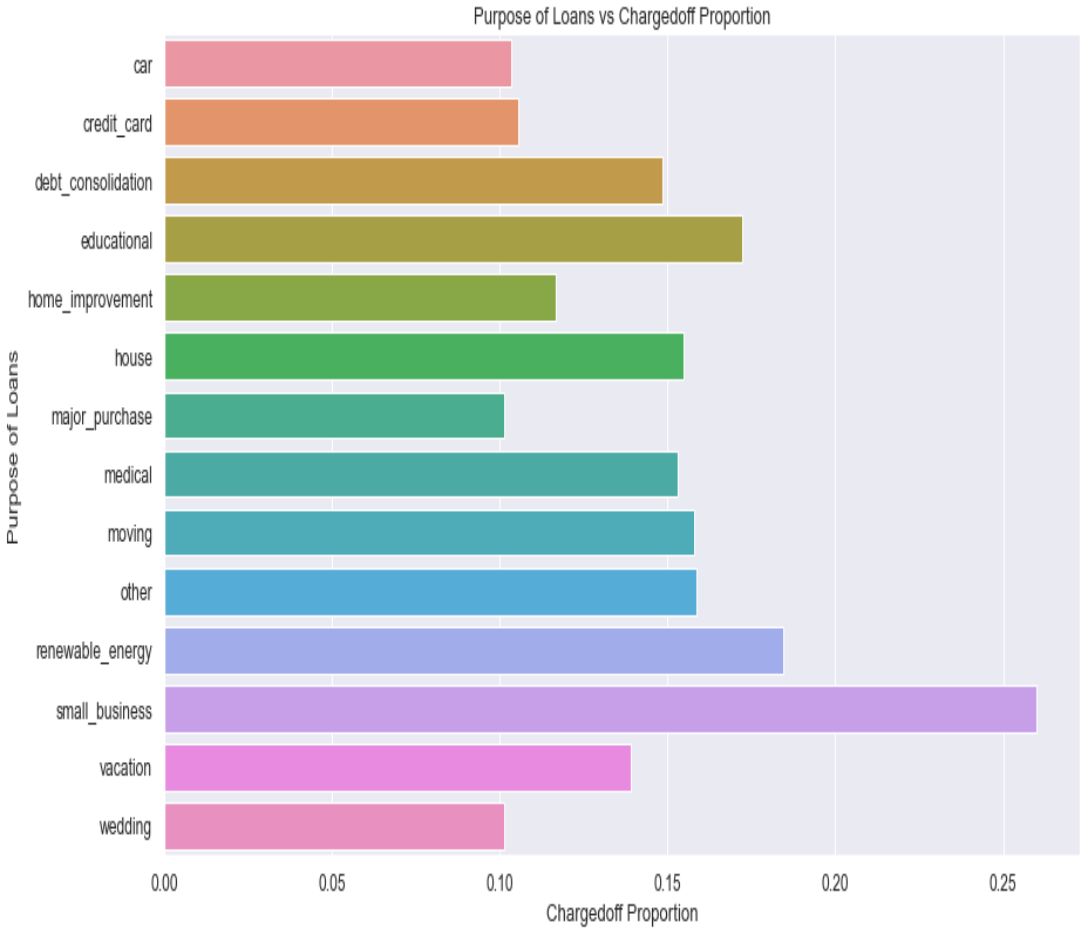
Grades vs. Charged off proportion



- Chances of charged off is increasing with grades moving from "A" towards "G“.

Observations from Bivariate Analysis

Purpose of loan vs. Charged off proportion



- Small Business applicants have high chances of getting charged off.

Conclusions

- From our EDA Analysis we conclude that a loan applicant is likely to be charged off under the following circumstances:
  - If the applicant's income falls in the range 0 – 20,000.
  - If the applicant goes for longer term loan then applied interest rates will be higher.
  - If the loan application is for small businesses.
  - If the applicant falls in the F or G grades.

Thank You.