Lending Club Case Study

Team Members :-
Arjun Shome (Group Facilitator)
Neelam Tom Sunder Rajeev

Problem Statement :-

I work for a **consumer finance company** which specialises 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

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

•**Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)

•**Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

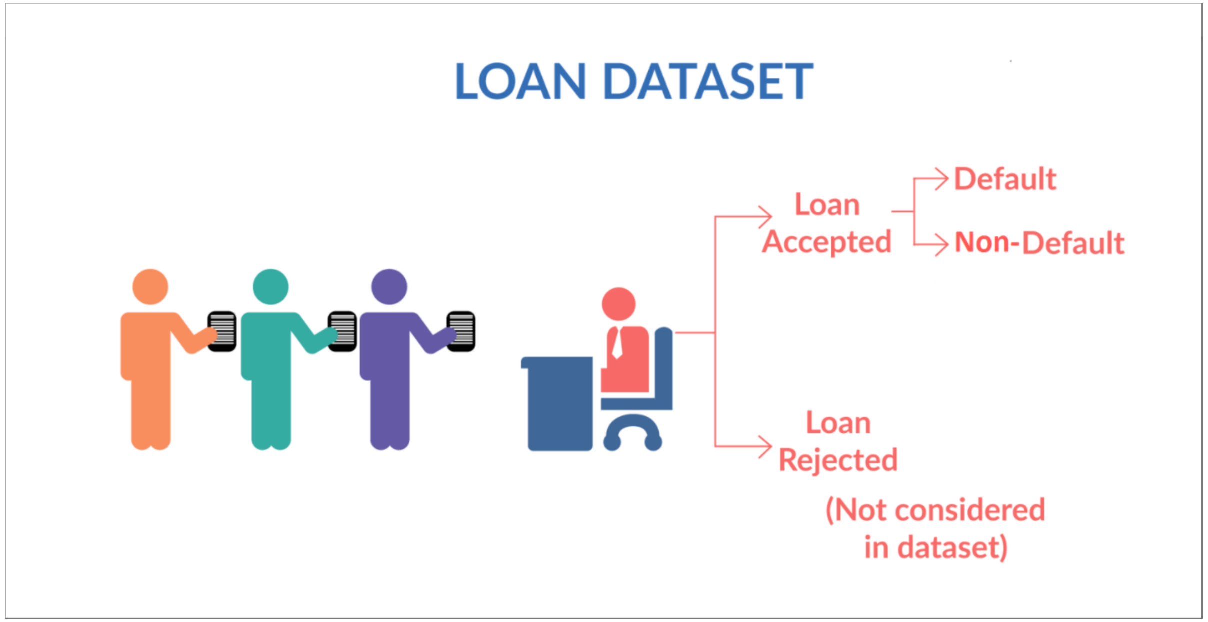
•**Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan

•**Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Objective : -

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, 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 utilize this knowledge for its portfolio and risk assessment.



Problem solving methodology

Data Analysis

Data Analysis

Data Analysis

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analysing each column, plotting the distributions of each column.

Segmented Univariate Analysis

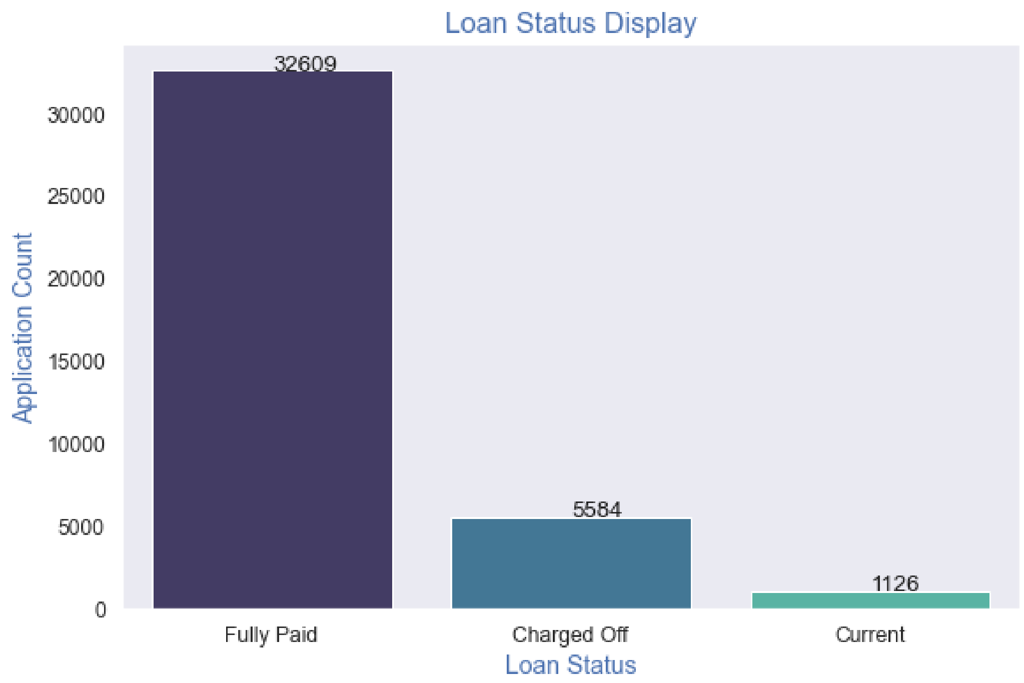
Analysing the continuous data columns with respect to the categorical column

Bivariate Analysis

Analysing the two variable behaviour like term and loan status with respect to loan amount.

Recommendations

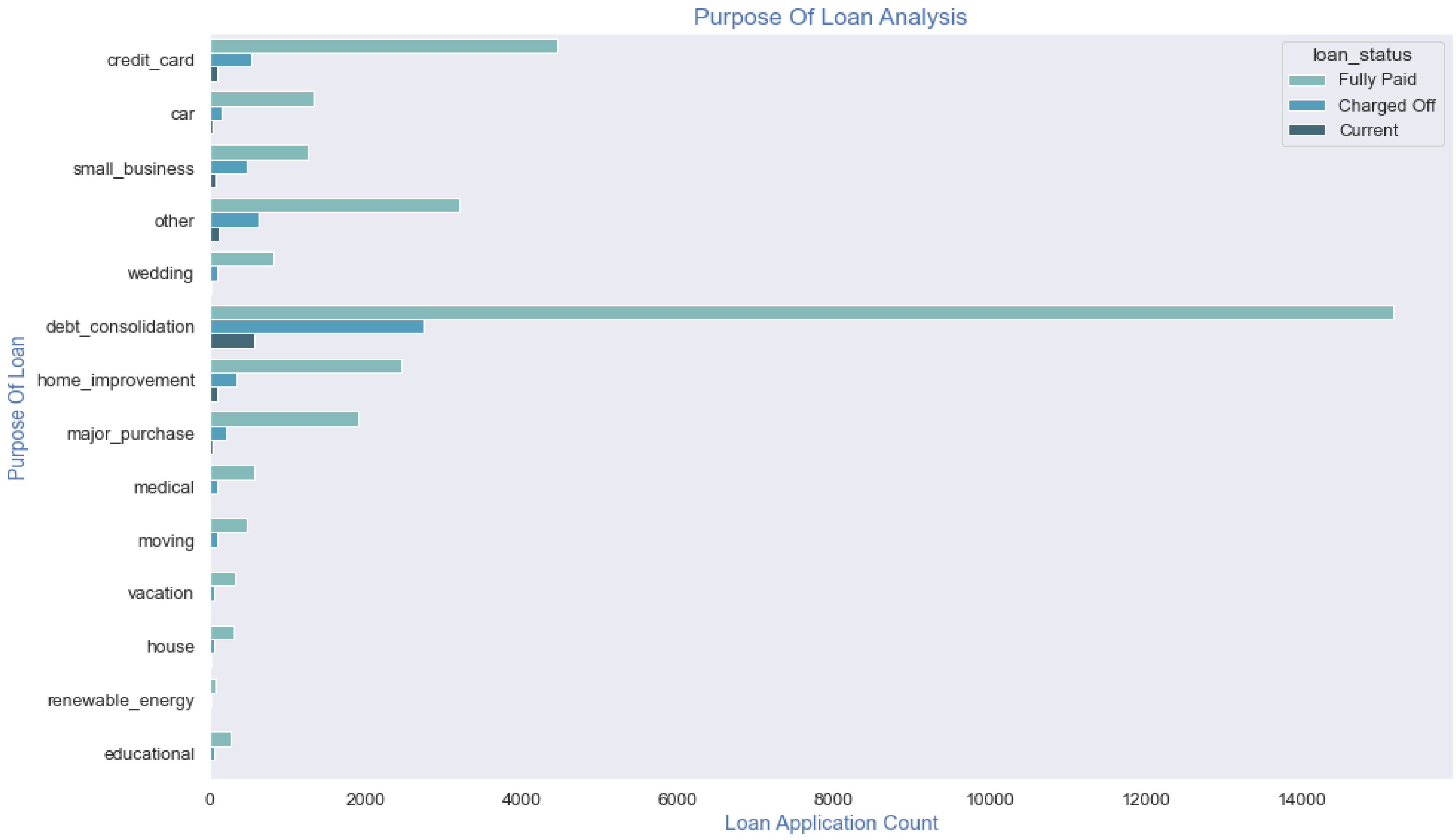
Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.



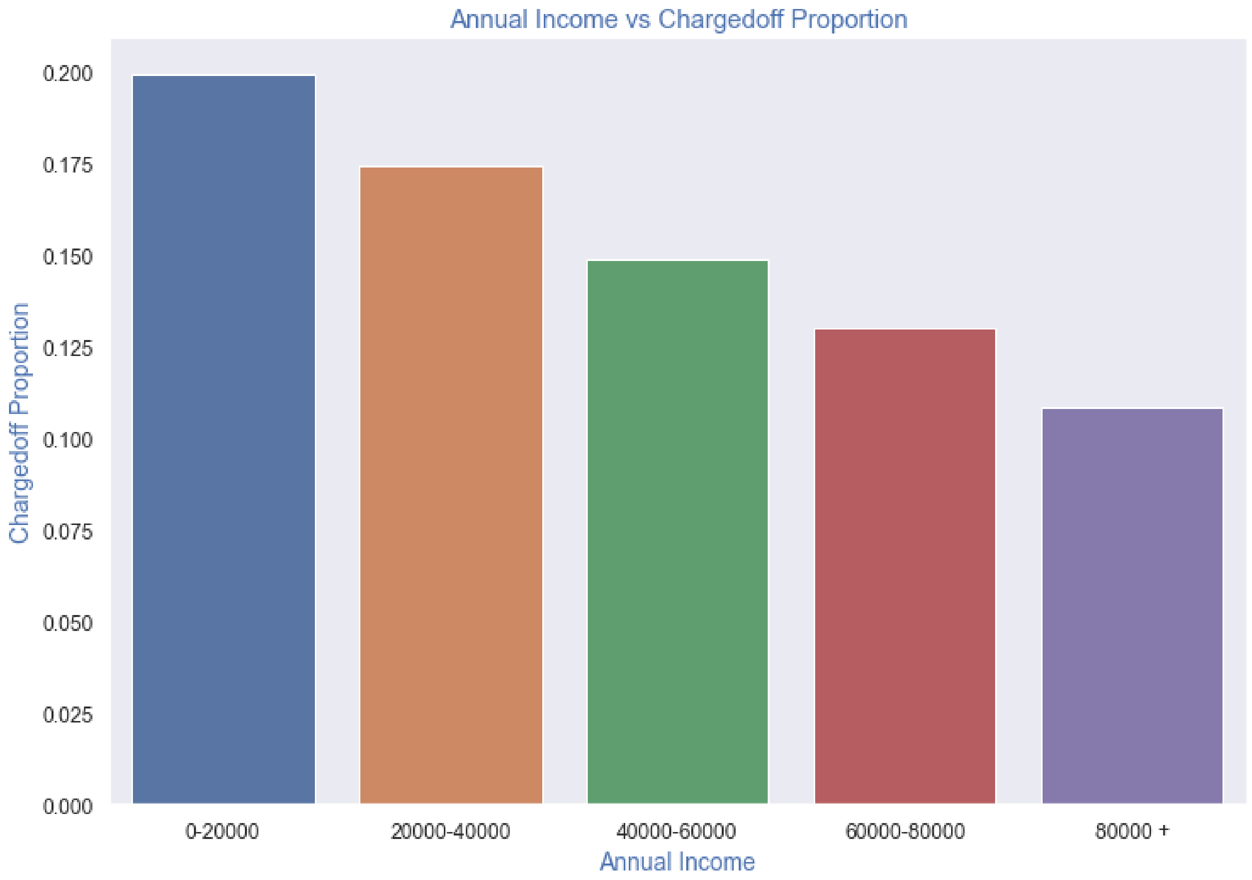
Currently from this figure we observe that most of the loans currently are already Fully paid.

There are few loans which are charged off.

Currently there are 1126 loans, installment of which is still being given by the borrowers.



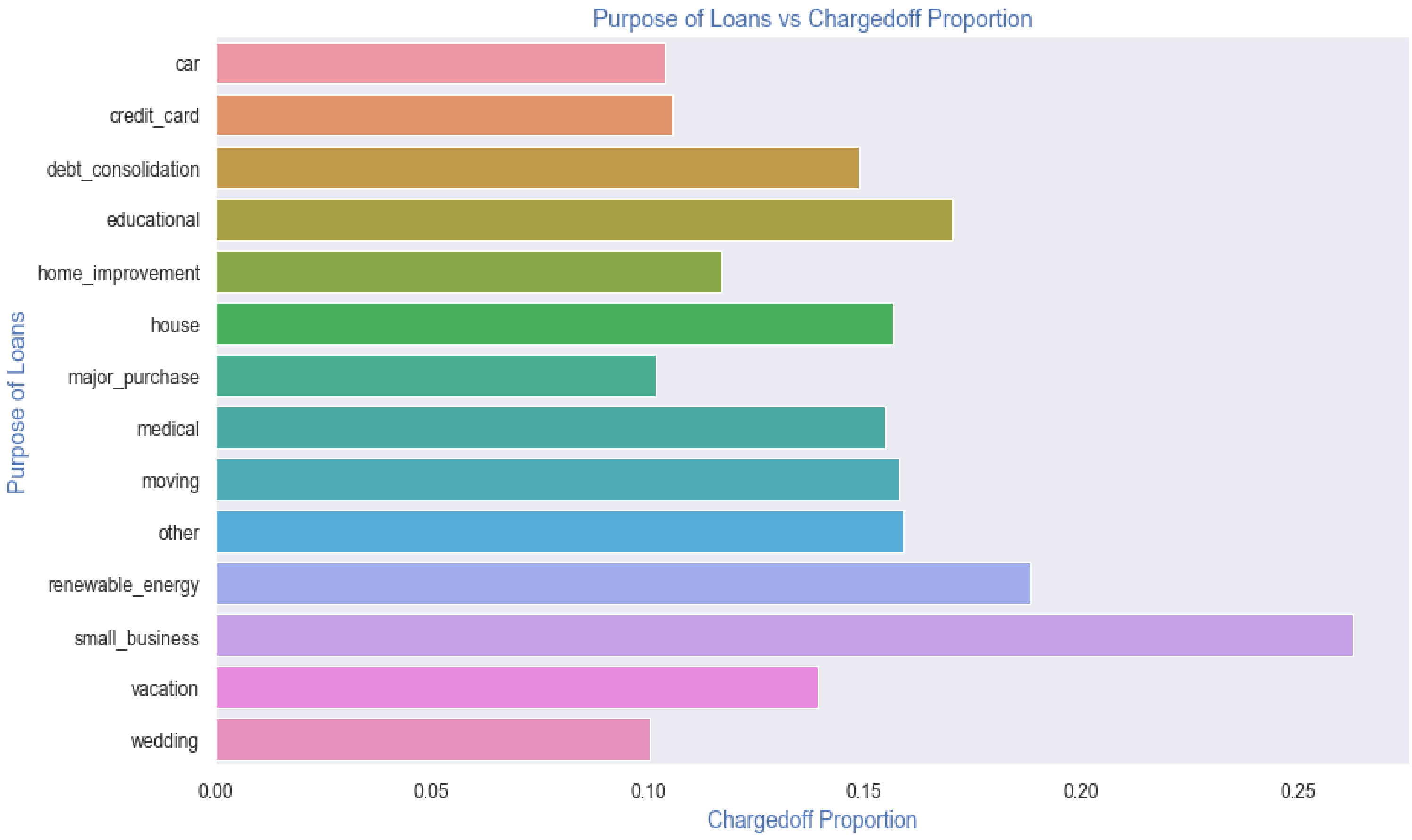
We observe that maximum loans were taken for debt consolidation
Followed by, payments of credit card bill, and most loans which were charged off belongs to the loans borrowed for debt consolidation, holding the maximum number of loans which are active currently.

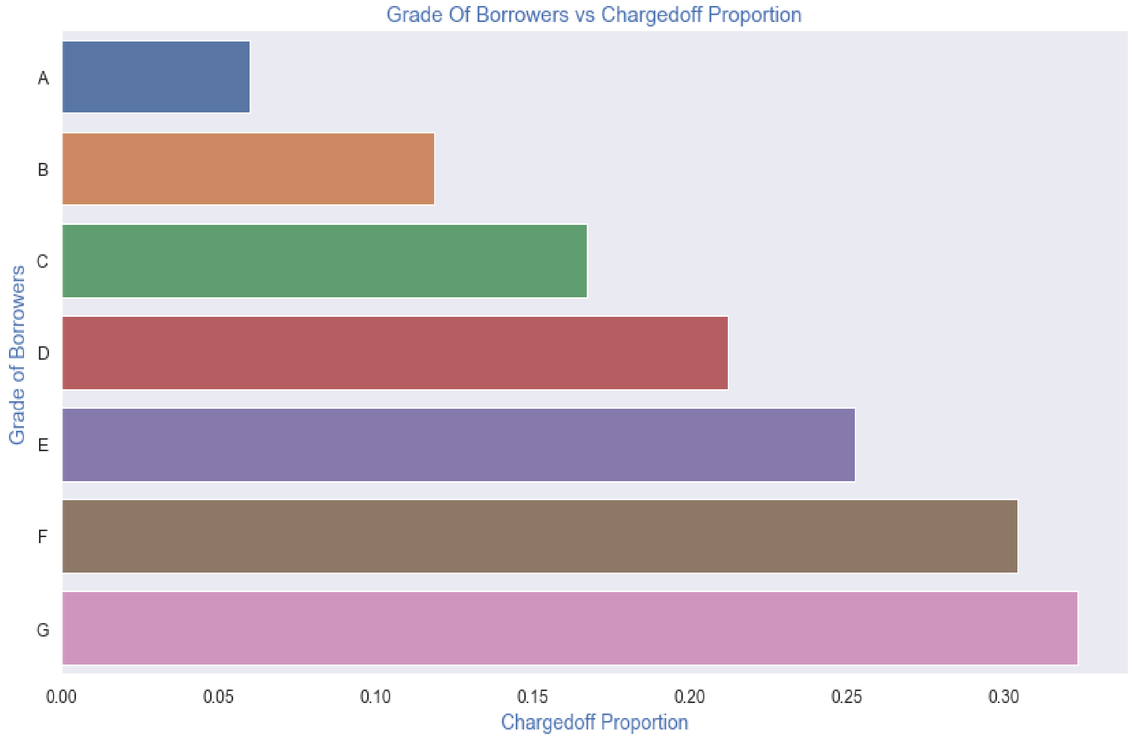


We see that with less annual income of the borrowers the loans are most likely to get charged off.

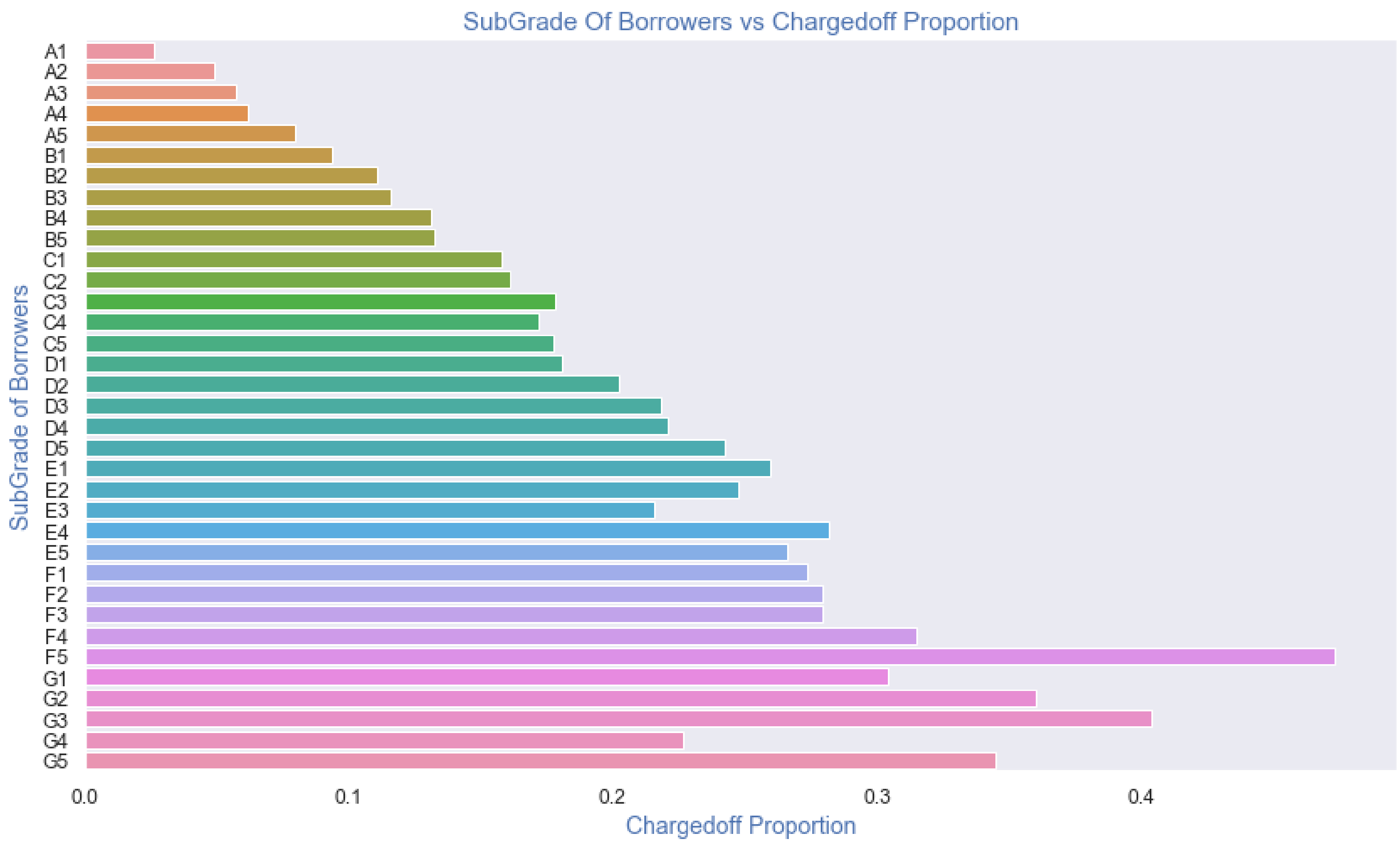
This is a valid and an important point for the lender to consider while giving out loan, to avoid risk.

Here, we see that loans that were taken with the purpose to start a small business has the highest risk of getting charged off.

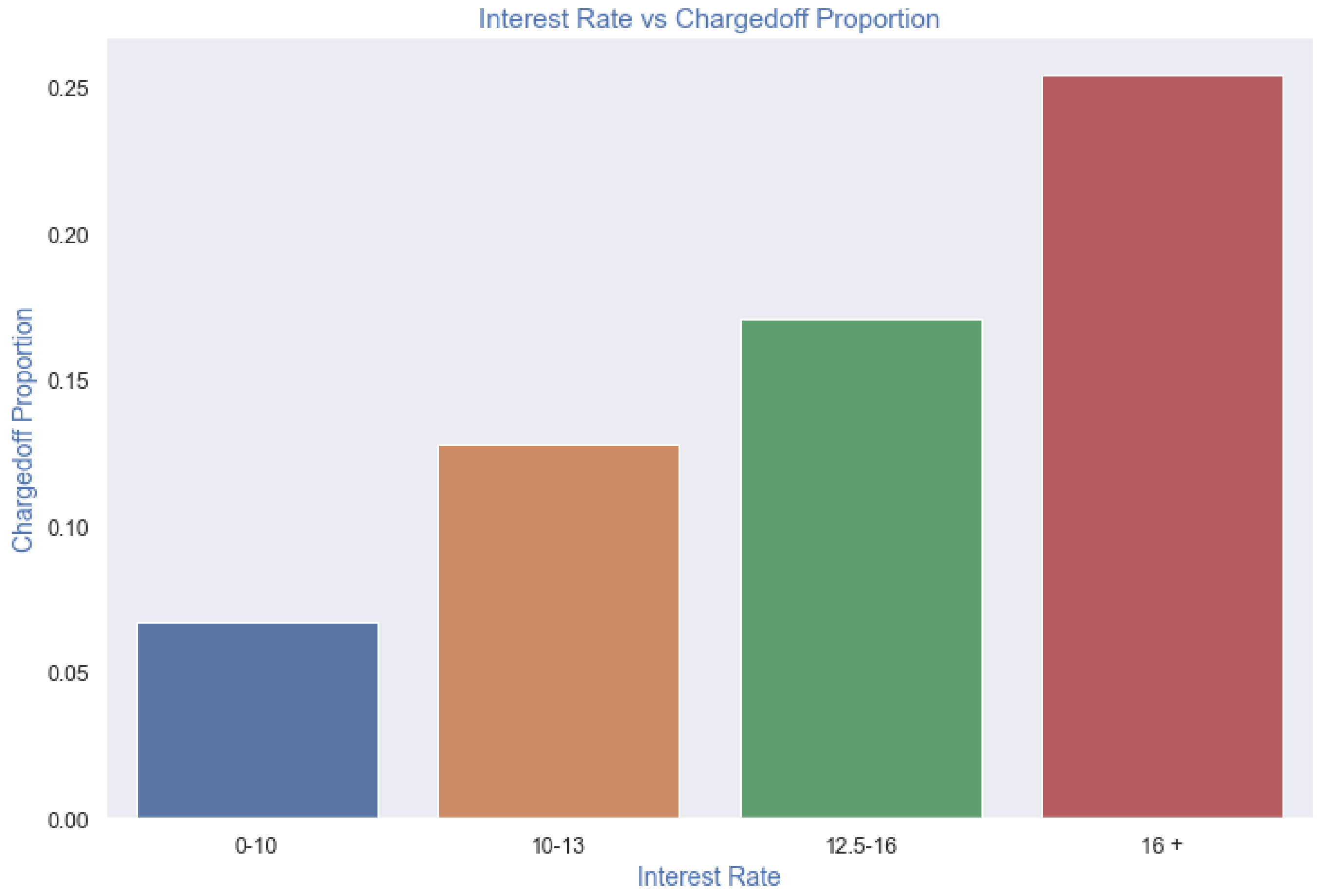




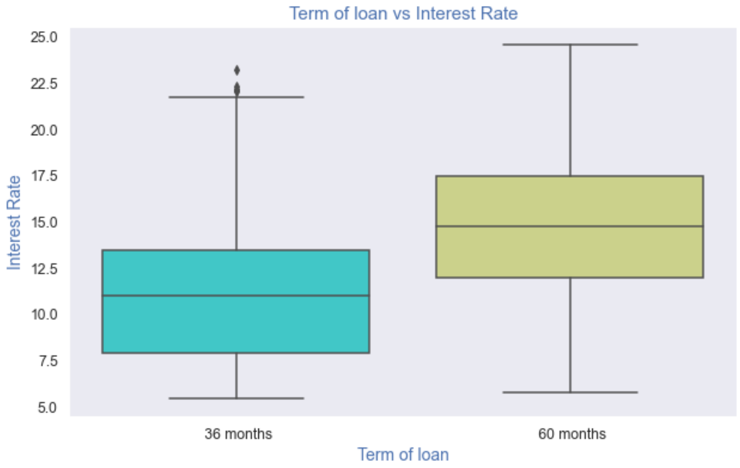
- Borrowers having 'G' and 'F' grade has the highest chance of defaulting the loan or their loan getting charged off.
  Another important variable to be considered while lending loan.
- Grade 'A' Borrowers are most likely to fully repay the loan amount with least chance of their loan getting charged off.

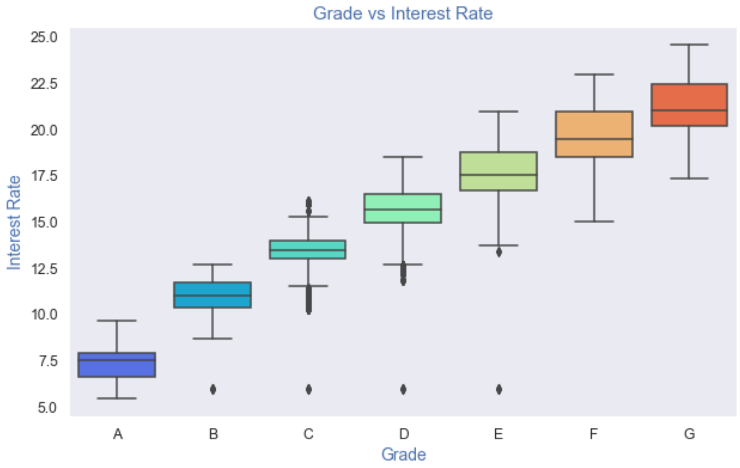


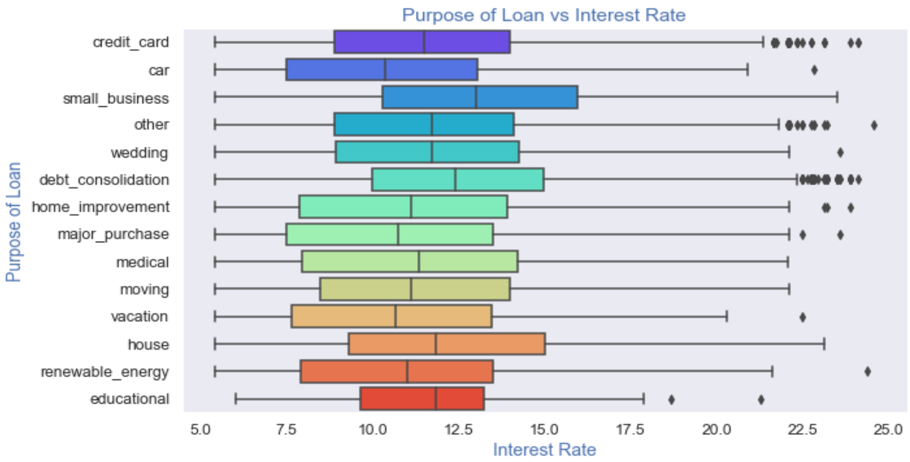
Among the sub-groups loans should be avoided for ('F4', 'F5', 'G1', 'G2', 'G3' and 'G5') borrowers to avoid risk of the loans getting charged off.



The Interest rates should be placed with caution to avoid the loans getting charged off, usually loans with higher interest rate gets charged off.



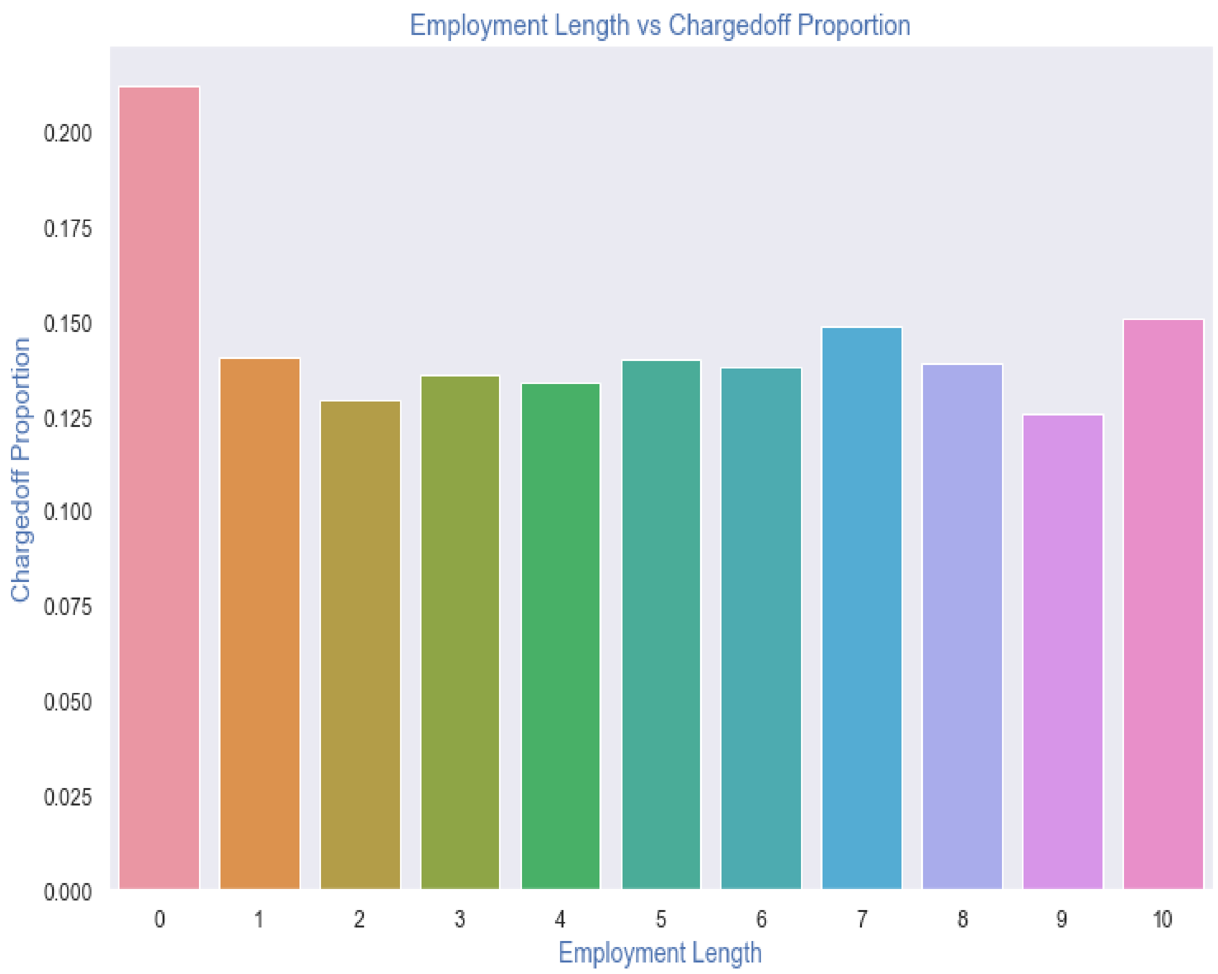




In previous slide we saw that loans with **higher interest rates** has the **higher chance of getting charged off**, so

- With the longer tenure of the loan the interest is growing higher.
- Providing loans to Grade 'F' and 'G' borrowers on higher interest rates.
- Providing loans to borrowers who are taking loans with the purpose to start a small business.

All these three points increases the chance to the loans getting charged off.



Loans are possibly going to get charged off if provided to a borrower whose work experience is less than a year,

which typically makes sense as the guy will have comparatively less money and experience.

Conclusion :-

Lending Club should reduce higher interest loans for 60 months tenure, which are prone to loan default.

Lending club should consider checking the grades of the borrowers before lending them loans, preferably to Grade 'A' customers, who has the best record to fully repay the loans.

Loans for the purpose of starting a small business should be reduced by lending club as they will tend to default more.

Lending club should check the work experience of the customers and try to provide loans to the borrowers who has more than a year of work experience at least.

Lending club should prefer giving loans to borrowers having higher Annual income at least above 60000.