Lending Club Case Study

Prepared by

Bhanu Sameera Velpuri

Sudhakar Vuriti

Objective

The objective is to identify potential default applicants while reviewing their available credentials to ensure that Lending companies will not incur loss by giving loan to an undeserved applicant using Exploratory Data Analysis techniques.

There are Two types of risks 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.

Two scenarios associated with this decision:

1. **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.
2. **Loan rejected:** If company not approved the loan because of not meeting the requirement's of the bank rules.

Executive Summary

Univariate Analysis

- Juniors who have 1-4 yrs of experience are the highest loan seekers.
- A large proportion of applicants are applying loan for debt consolidation.
- A large portion of applicants have applied for loan grade B.
- A large portion of applicants have applied for B3-C1 loan subgrade followed by A1-B2
- Most of the loans are of 36 months tenure.
- Applicants who rent or mortgage home are the highest loan seekers.

Segmented Univariate Analysis

- Most of the charged off applicants have loan amount less than 15K.
- Applicants who are professionals (10+ years) have taken highest average loan amount.
- Charged off applicants report higher percentage of income going to debts(dti) compared.
- Most of the applicants who took the loan for debt consolidation have paid fully
- Loan was paid by all applicants whose average annual income is high compared to charged off applicants
- Interest rate is higher, the chance for 'charged off' is more.
- Most of the applicants with higher income have their home mortgaged.
- Applicants with highest average income for Loan grade F are being charged off for highest average loan amount.

Executive Summary

<u>Bivariate Analysis</u>

- There is high correlation between annual income to loan amount (0.42) , there 40 % chance that higher income individuals apply to higher loan amounts.
- Very low correlation between annual income to interest rate (0.068), there is positive but negligible influence on higher income individuals to apply for loan based on interest rate.
- Annual income to DTI has very minor but negative correlation (-0.11).
- Loan amount to DTI has positive but little positive correlation (0.071).

Conclusion

- ✓ It is safe to provide loans to employees with higher income, low dti value for 36 months tenure.
- ✓ It is also safe to provide loans to applicants whose home is mortgaged and purpose of the loan is debt consolidation.

Steps followed for the case study :

Step 1: Data Understanding : Tried to understand the complete data set.

Step 2: Data Cleaning : Removed all unnecessary columns. All columns and rows that contains 'NaN' values as these does contribute to the analysis.

Step 3: Univariate Analysis : Tried to analyze the data in individual variables to infer its value towards identifying potential defaults.

Step 4: Segmented Univariate Analysis : This was done by grouping categorical data with quantitative variables to make analysis more insightful compared to Univariate analysis.

Step 5: Bivariate/Multivariate Analysis : Here we used two or more variables to draw a better conclusion to understand correlation between them.

Step 6: Recommendations/Results : Here, we made an effort to provide insightful recommendations to ensure that Lenders will not incur loss by offering bad loans to the applicants who has no intent to payback.

Data Cleaning Steps :

- Delete columns: Deleted unwanted columns.
- Remove outliers: Removed high and low values that would disproportionately affect the results of our analysis.
- Missing values: Treat missing values with appropriate approach.
- Duplicate data: Remove identical rows, remove rows where some columns are identical.
- Filter rows: Filter by segment to get only the rows relevant to the analysis.

Strategy Followed for conducting analysis :

- ✓ The whole project is to analyze and understand how consumer and loan attributes are influencing the tendency of Charge off.

- ✓ During analysis used the different plots.

- Histograms / Bar charts to check out the distribution of all the variables
- Scatter plots to detect the outliers
- Box plots to see the quantile percentages distribution across variables.
- Performed the Bivariate analysis to understand how different variables interact with each other and how its affecting the Charge off% and fully paid%.
- Heat map for correlation between quantitative variables.

Univariate Analysis
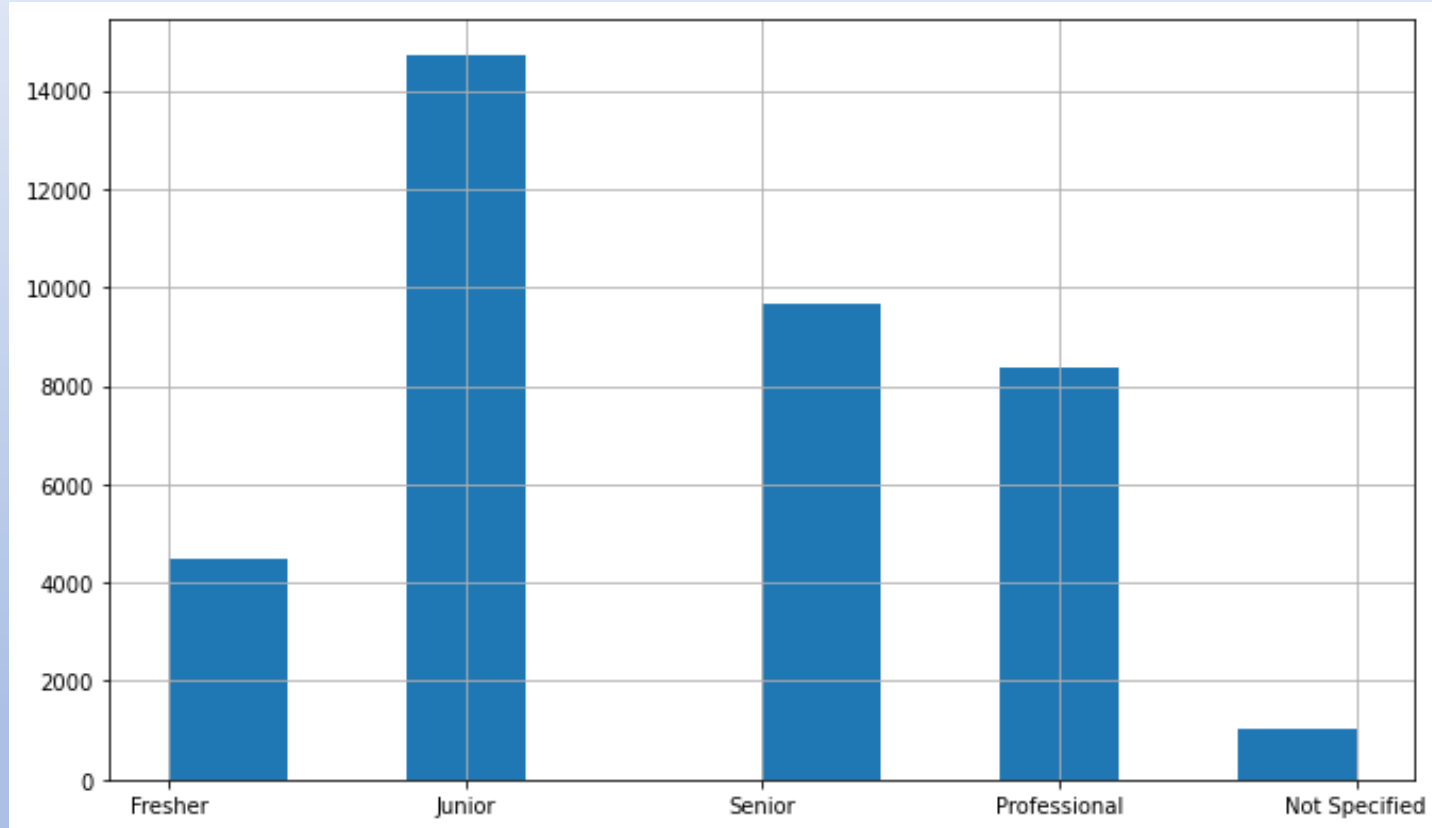
Juniors who have 1-4 yrs of experience are the highest loan seekers.
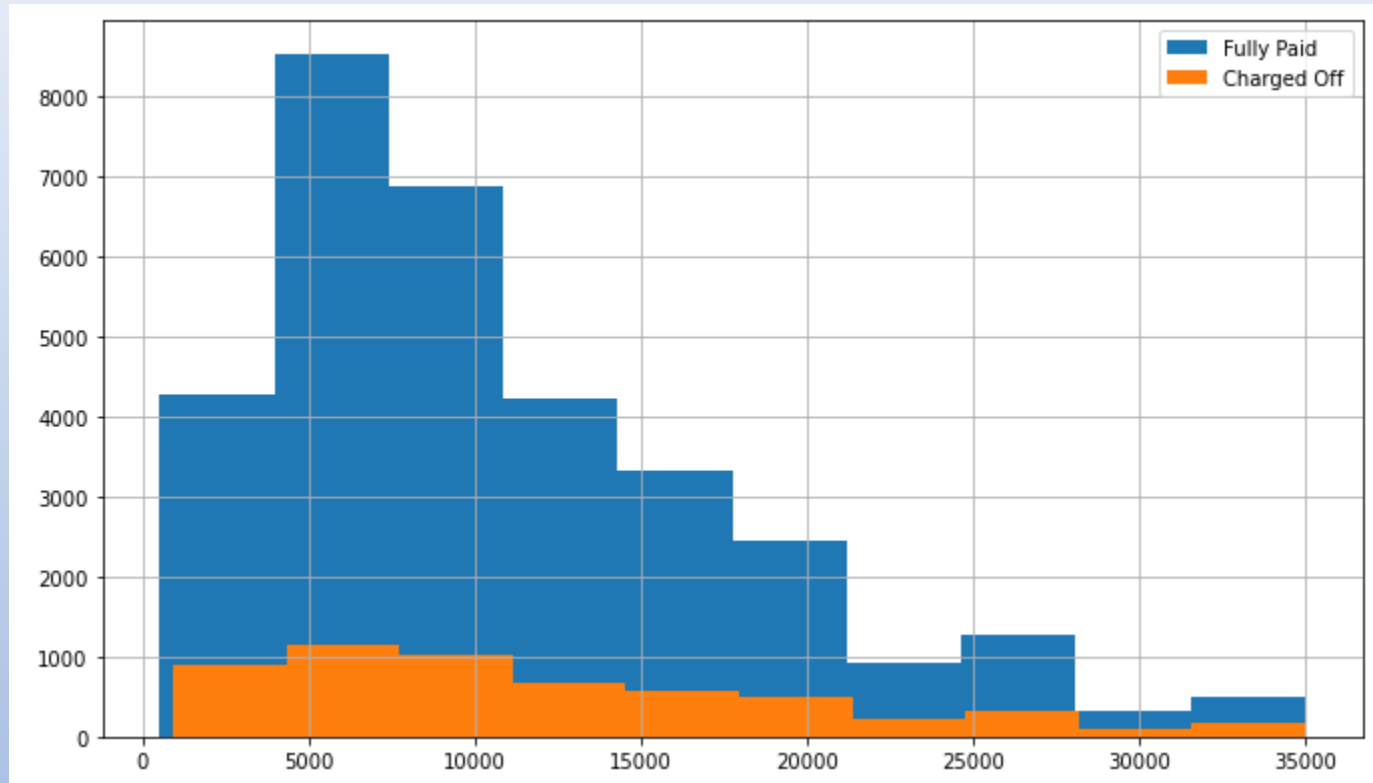
Charged off applicants have loan amount below 15k

loan_status	Charged Off	Fully Paid	All
emp_exp			
Fresher	637	3847	4484
Junior	2027	12673	14700
Senior	1383	8275	9658
Professional	1318	7052	8370
Not Specified	227	802	1029
All	5592	32649	38241

Juniors who have 1-4 yrs of exp are highest applicants for loan which has highest applicants being charged off and fully paid.

Boxplot grouped by loan_status

dti
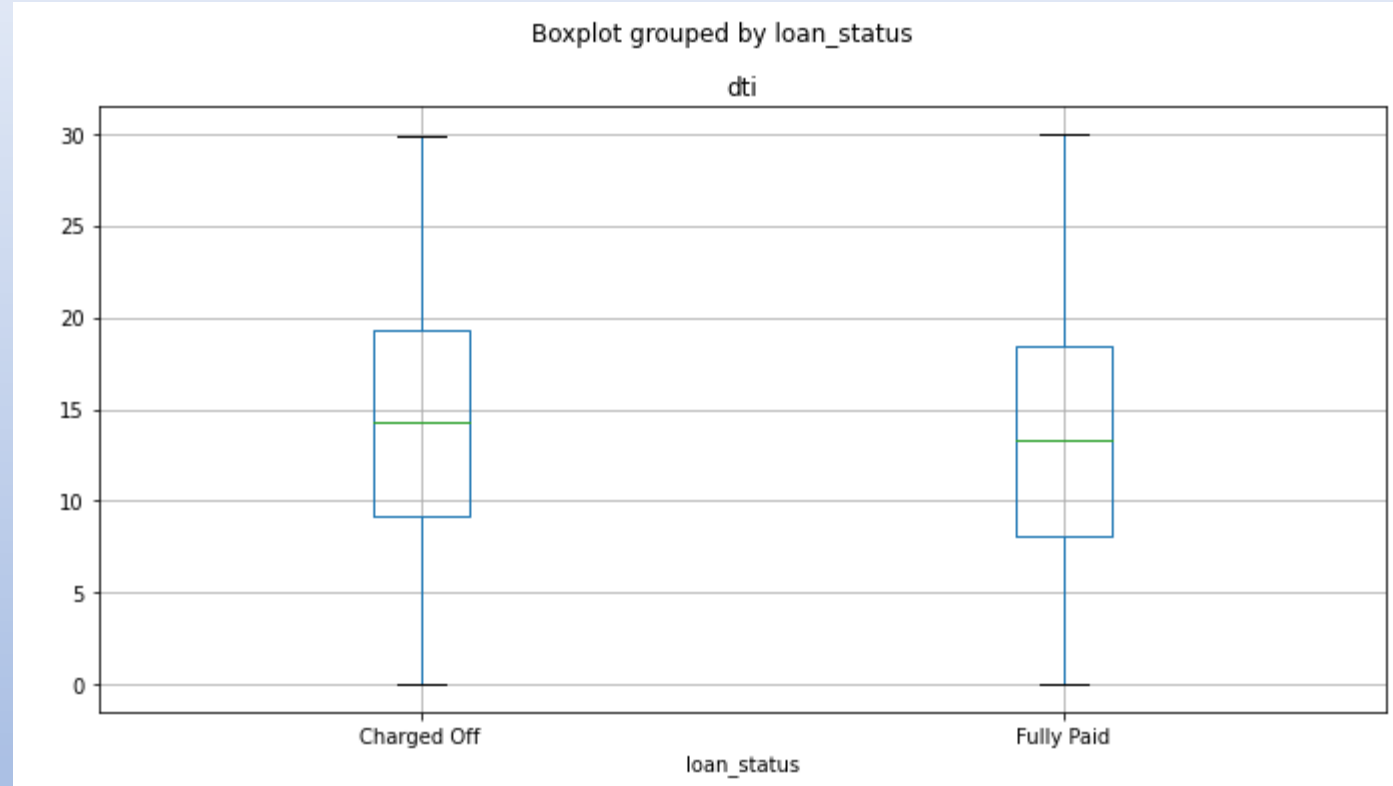
Charged off applicants report higher percentage of income going to debts compared to fully paid

loan_status	annual_inc		loan_amnt	
grade	Charged Off	Fully Paid	Charged Off	Fully Paid
A	52182.0	65073.0	7801.0	8633.0
B	57224.0	65827.0	10869.0	10880.0
C	56670.0	64962.0	11006.0	10700.0
D	61055.0	66711.0	12196.0	12021.0
E	68936.0	74190.0	15813.0	15452.0
F	75371.0	80857.0	19078.0	17303.0
G	73392.0	90505.0	18558.0	20827.0

Loan grade F have highest average income being charged off for highest average loan amount.

Bivariate & Multi Variate Analysis
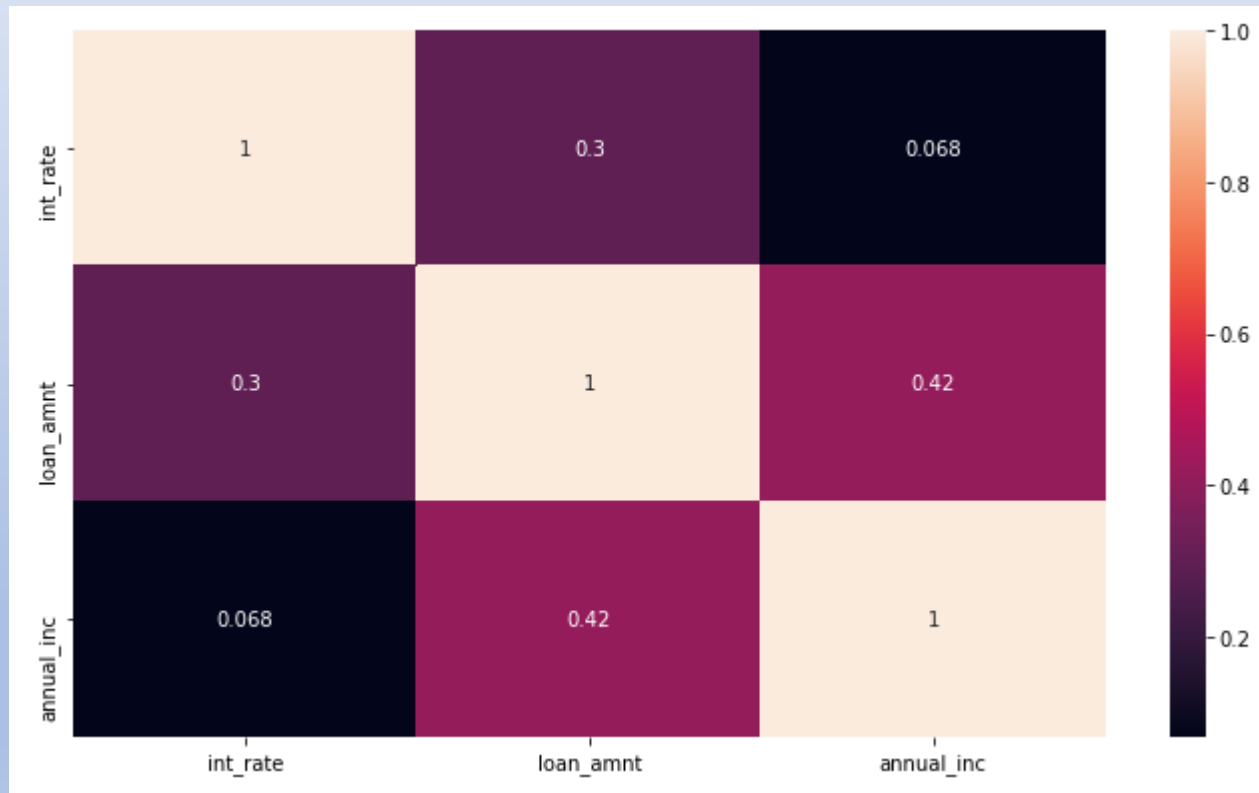
- There is high correlation between annual income to loan amount (0.42) , there 40 % chance that higher income individuals apply to higher loan amounts.

- There is high correlation between annual income to loan amount (0.42) , there 40 % chance that higher income individuals apply to higher loan amounts.
- Very low correlation between annual income to interest rate (0.068), there is positive but negligible influence on higher income individuals to apply for loan based on interest rate.
- Annual income to DTI has very minor but negative correlation (-0.11).
- Loan amount to DTI has positive but little positive correlation (0.071).

Thank you