Lending Club Case Study

By:

Vineet Prakash

Ashita Ahuja

Approach

- 1. Data Cleaning
- 2. Univariate analysis
- 3. Segmented Univariate analysis
- 4. Bivariate Analysis
- 5. Conclusion

Data Understanding(Key variables)

- **id** - Unique id for loan listing
- **member_id** - Unique id for the borrower
- **loan_amnt** - Amount requested by the borrower
- **funded_amnt** - Amount approved by the lending club
- **funded_amnt_inv** - Amount lent by the investor or the final amount which the borrower received
- **term** - Time for which the amount is lent
- **int_rate** - Interest rate for the loan
- **installment** - Amount to be paid every month
- **grade** - Grade assigned by the lending club, parameter for determinig the risk , ranges from A-G
- **sub_grade** - Subgrade assigned from 1-5 for all the grades. Like A1-A5, based on those the interest keeps on increasing as the risk also increases
- **emp_title** - Job title for the borroweremp_length - No of employment years
- **home_ownership** - Indicates if the borrower has own house, rented, mortagaged or other
- **verification_status** - Indicates if income source is verified or not
- **annual_inc** - Annual income of the borrower

- **loan_status** - Indicates if the person has successfully paid a loan or is a defaulter
- **purpose** - Indicates the purpose of the loan
- **title** - Loan title provided by borrower , also describes the purpose
- **zip_code** - Zip code of his communication residence
- **addr_state** - Communication address
- **pub_rec_bankruptcies** - Number of public record bankruptcies
- **dti** - A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

Data cleaning

Following is the data cleaning process followed:

- Remove rows where the borrower is still in the process of paying out the loan

- Remove customer behaviour variables as those are not available at the time of loan application

- Drop columns where there are more than 50% NA values

- Drop columns where there is only one unique value as these will not contribute to the analysis

- Change interest rate column from string to numeric

- Binarize the target variable (Loan status) such that 1 represents default and 0 represents that the loan has been fully paid, for ease of analysis
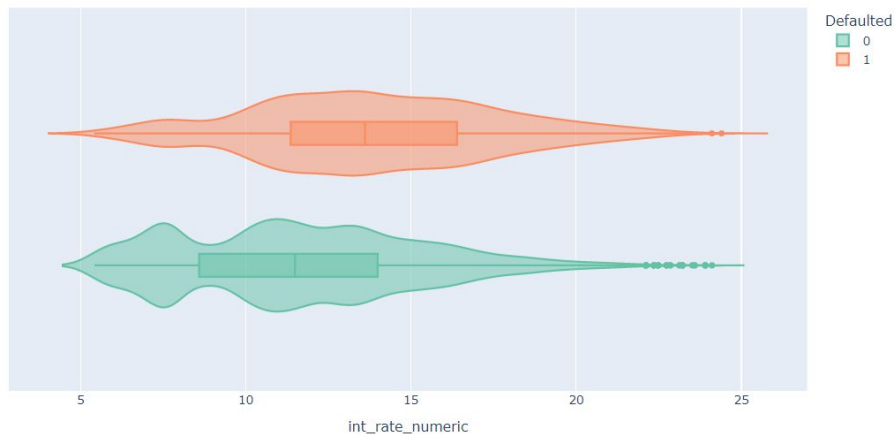
Data analysis

Loan Amount and Loan Status
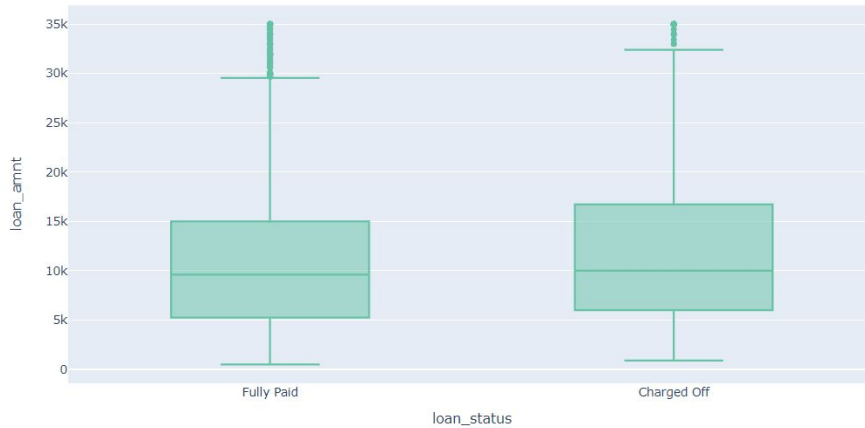
Charged off people have higher loan amounts than those who have fully paid their loans

Interest rate and loan status

loan_status	Charged Off	Fully Paid
int_rate_group		
5%-9%	0.060000	0.940000
9%-13%	0.130000	0.870000
13%-17%	0.190000	0.810000
17%-21%	0.290000	0.710000
21%-24%	0.410000	0.590000

People who have defaulted have higher interest rates on their loans as compared to those that haven't

Public record bankruptcies and loan status

loan_status	Charged Off	Fully Paid
pub_rec_bankruptcies		
0.000000	0.140000	0.860000
1.000000	0.220000	0.780000
2.000000	0.400000	0.600000

As record of pub_rec_bankruptcies
increase, no. of defaulters increases

Purpose and interest rate

|		
credit_card	0.100000	0.900000
debt_consolidation	0.150000	0.850000
educational	0.160000	0.840000
home_improvement	0.120000	0.880000
house	0.170000	0.830000
major_purchase	0.100000	0.900000
medical	0.160000	0.840000
moving	0.150000	0.850000
other	0.160000	0.840000
renewable_energy	0.190000	0.810000
small_business	0.270000	0.730000
vacation	0.140000	0.860000
wedding	0.100000	0.900000
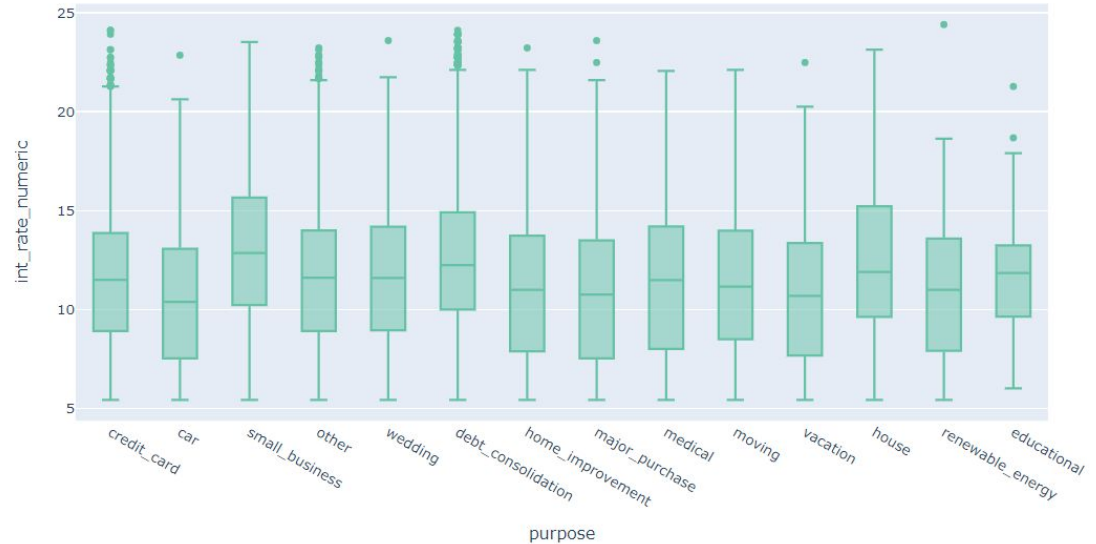
For small business ,debt_consolidation and house loans
are given at high interest rate and hence probability of
default increases

Verification status vs interest rate
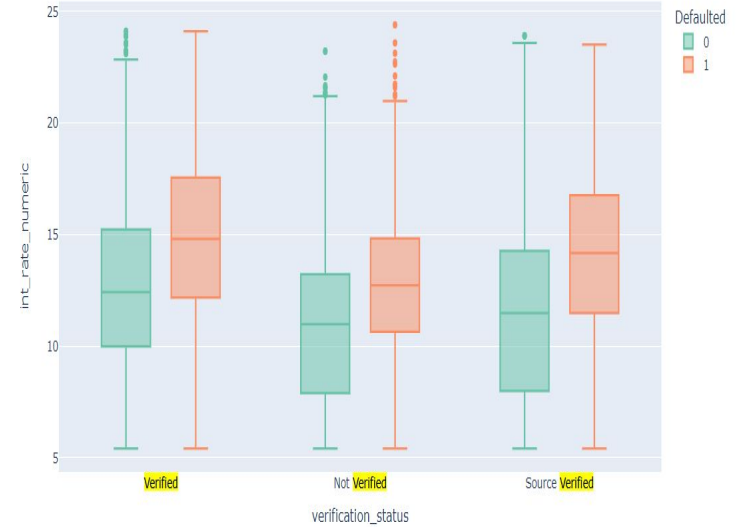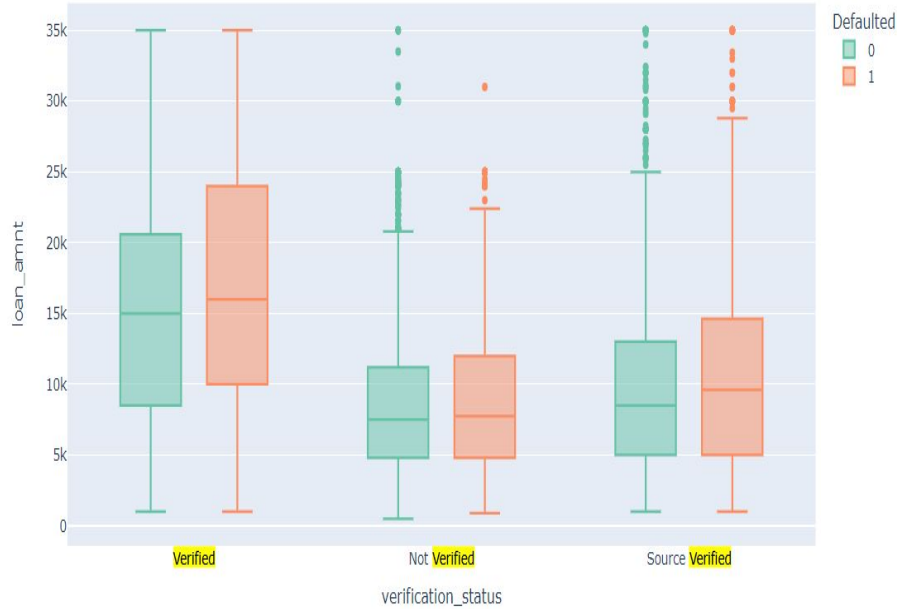
People who have their incomes verified by LC are the highest defaulters
and also their interest rates are higher

Interest rate vs grade
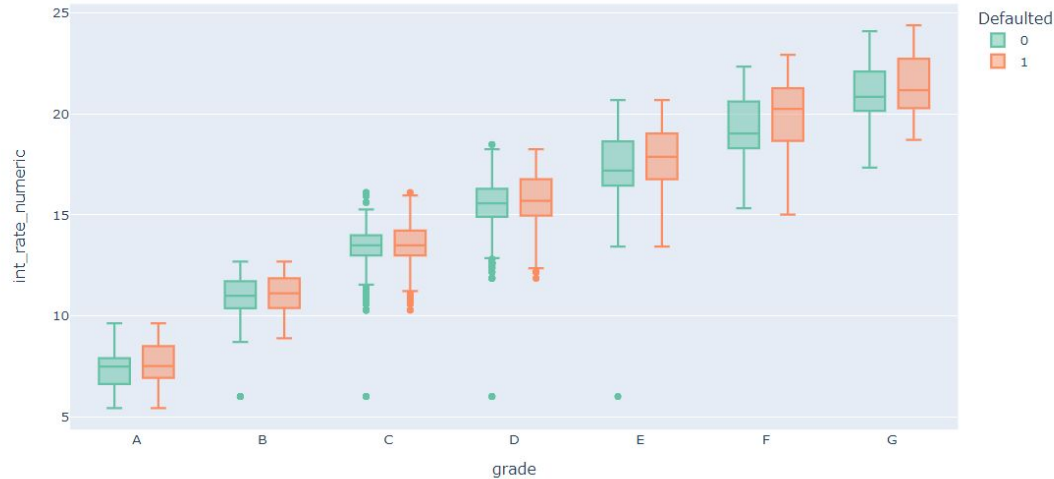
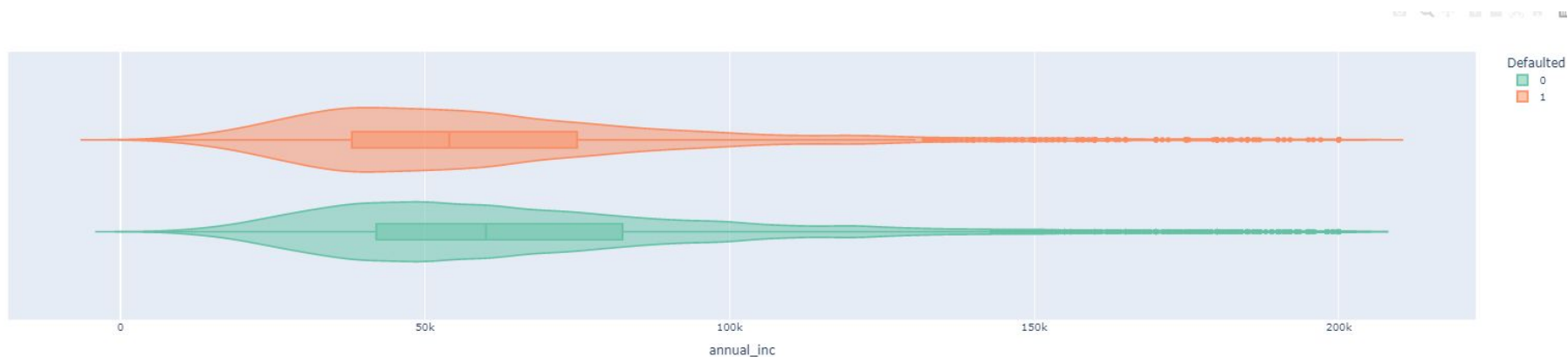| loan_status grade	Charged Off	Fully Paid
A	0.060000	0.940000
B	0.120000	0.880000
C	0.170000	0.830000
D	0.220000	0.780000
E	0.270000	0.730000
F	0.330000	0.670000
G	0.340000	0.660000

As loan grades and subgrades deteriorate, the probability of default increases as interest rate higher for lower grades
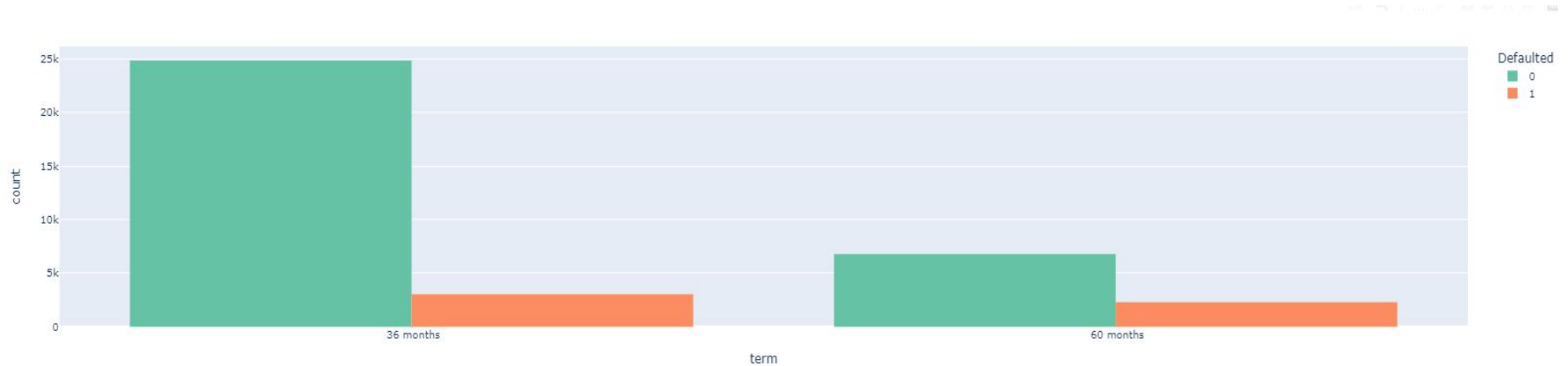
Debt to Income against Loan status
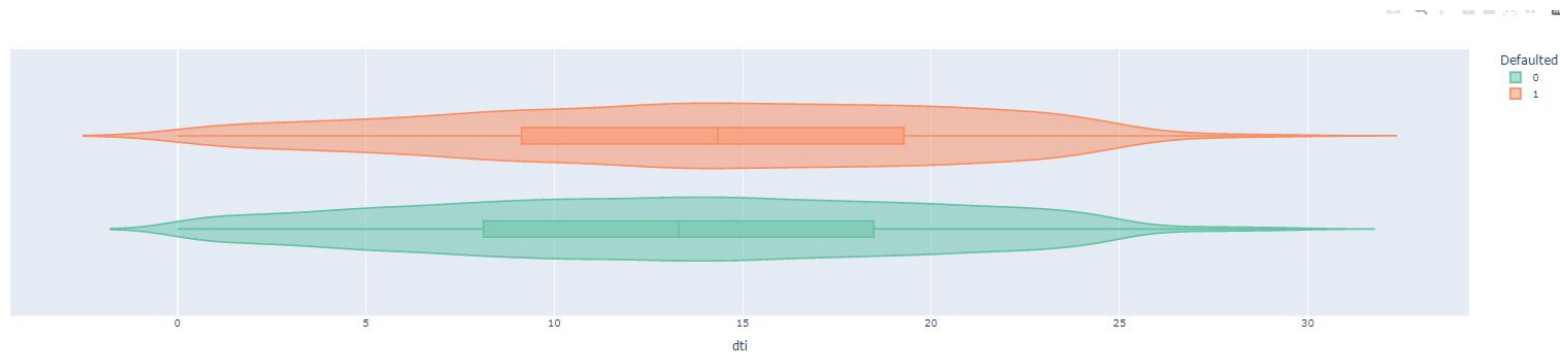
Defaulters, on average, have a higher
debt to income ratio than
non-defaulters

Annual income and loan status

Defaulters, on average, have a lower
annual income than non-defaulters

Loan term and loan status

loan_status term	Charged Off	Fully Paid
36 months	0.110000	0.890000
60 months	0.250000	0.750000

As loan tenure increases, probability of default increases

Recommendations

Based on the EDA performed, the following variables are strong drivers of default:

- Large loan amount
- Large interest rates for low grades
- Low loan grades and subgrades
- Low annual income
- High Debt-to-income ratio
- Loan taken for the purposes of small business
- High loan tenure

As such, lending club should:

- Avoid providing loans with higher interest rate to people with low grades
- Consider debt-to-income ratio before providing higher amount of loans
- Avoid providing large loan amount for very high interest rates
- Avoid providing loans to small businesses
- Restrict loans to smaller tenures