

# EDA Case Study

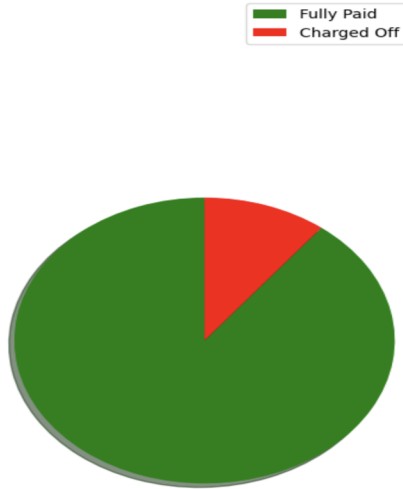
Sushanth Kamath & Ashutosh Aman

# Context on the assignment

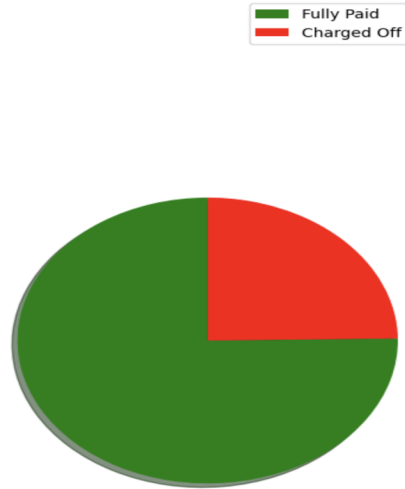
- Lending Club (LC) is the largest online marketplace facilitating a variety of retail loans.
- Lending Club (LC) wants to identify patterns from past loan applicants and whether an applicant is likely to default.
- The objectives of the case study are to :
  - Identify risky loan applicants to minimise credit loss
  - To understand how consumer attributes and loan attributes influence the tendency of default.
  - Identify variables that are possible causes for loan default
  - Provide recommendations on the attributes that could cause loan default

# Univariate Analysis

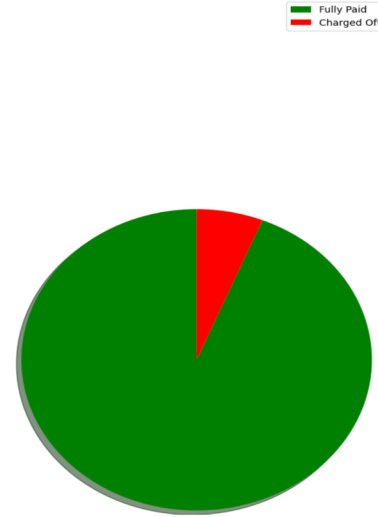
Distribution for 36 month term



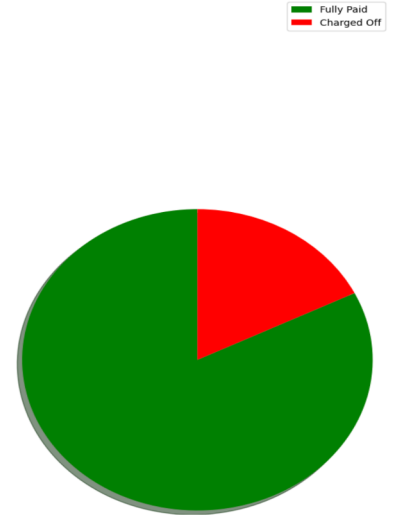
Distribution for 60 month term



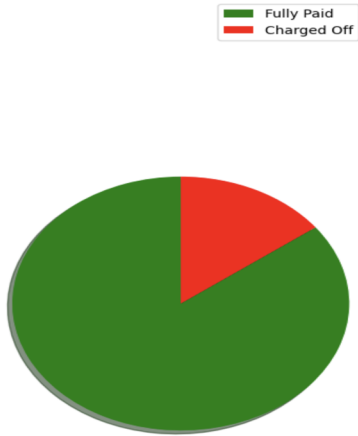
Distribution for Interest rate < 10 %



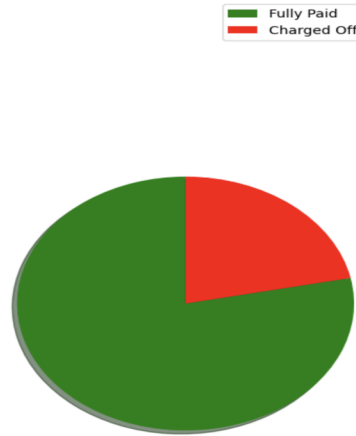
Distribution for Interest rate >= than 10 %



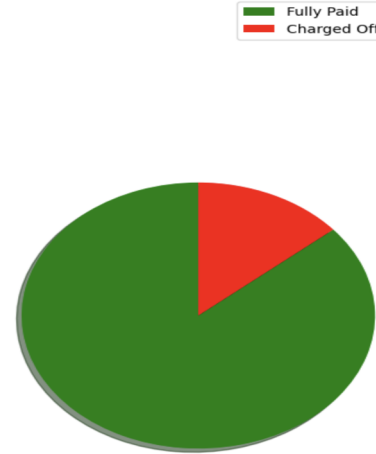
Distribution for pub\_rec 0



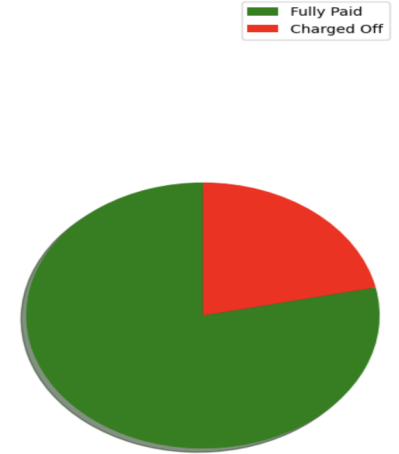
Distribution for pub\_rec >0

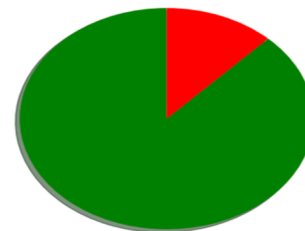
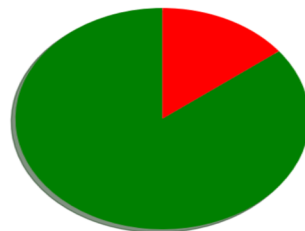
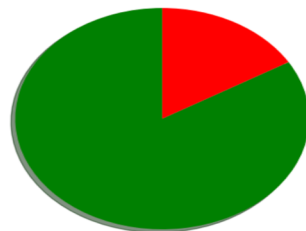
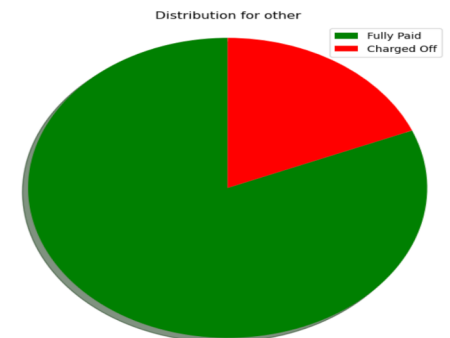
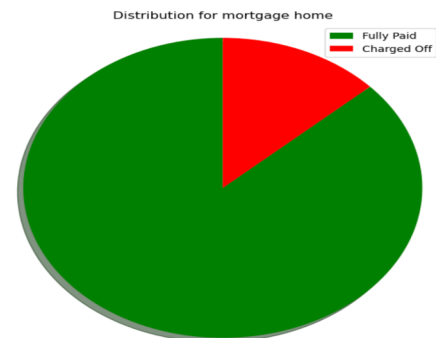
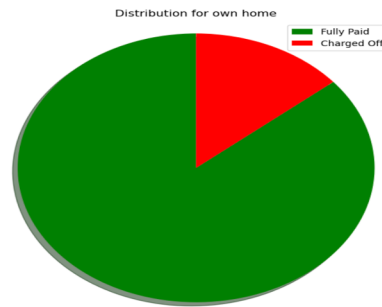
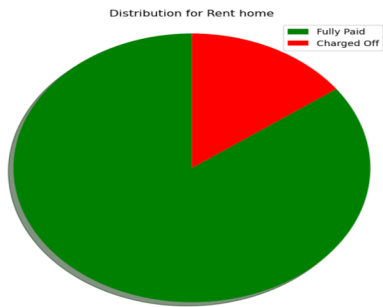


Distribution for 0 bankruptcy

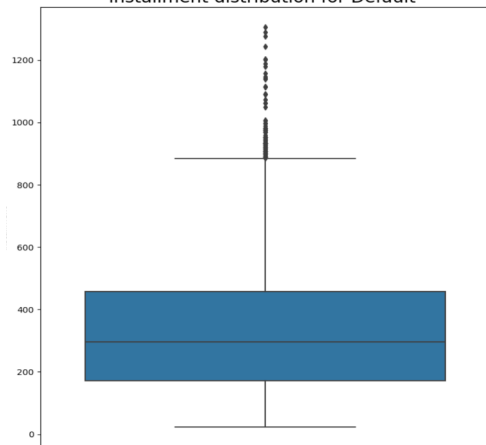


Distribution for bankruptcy >0

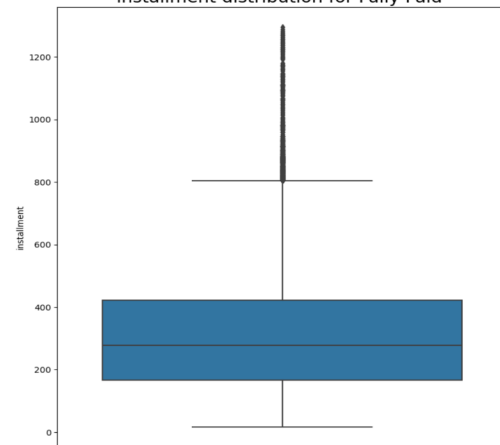




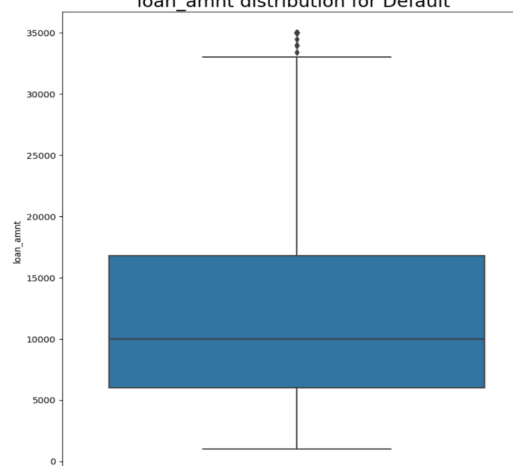
installment distribution for Default



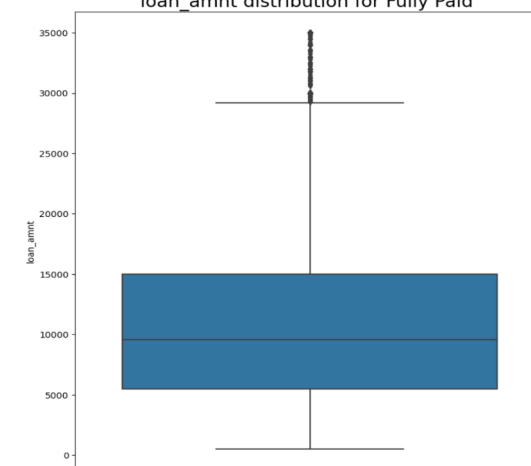
installment distribution for Fully Paid



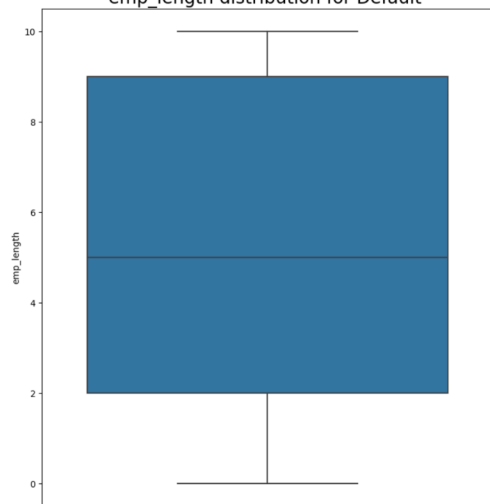
loan\_amnt distribution for Default



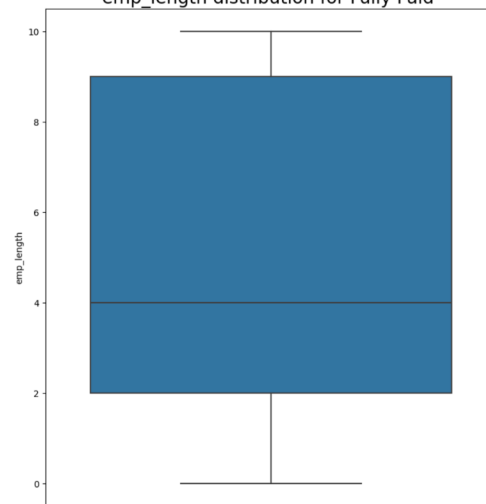
loan\_amnt distribution for Fully Paid



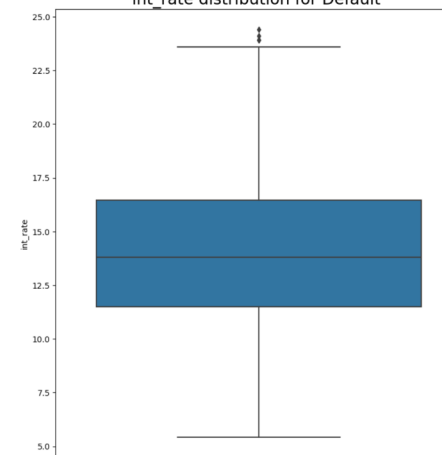
emp\_length distribution for Default



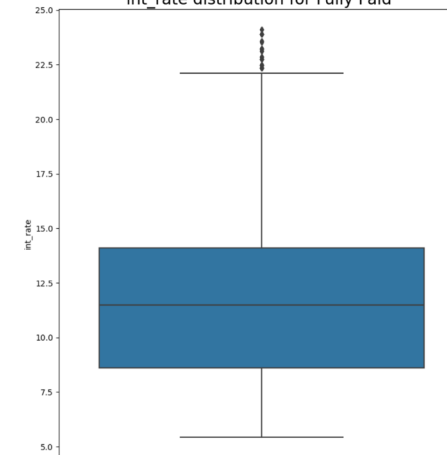
emp\_length distribution for Fully Paid

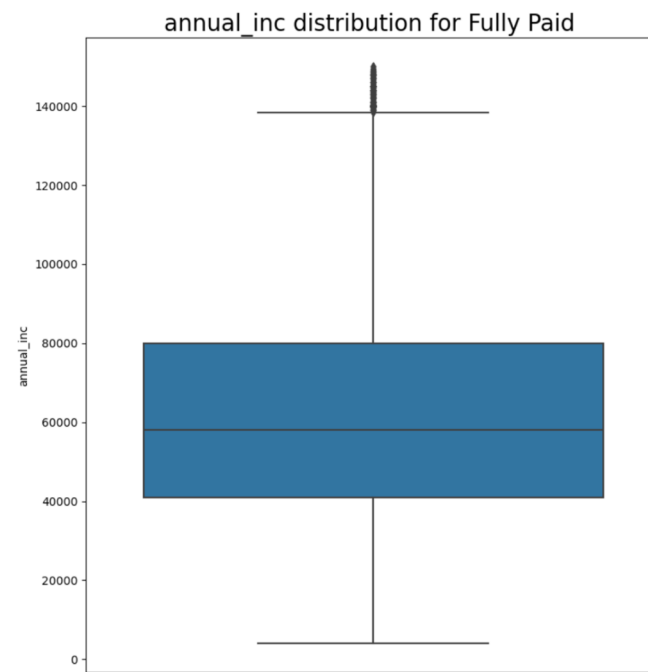
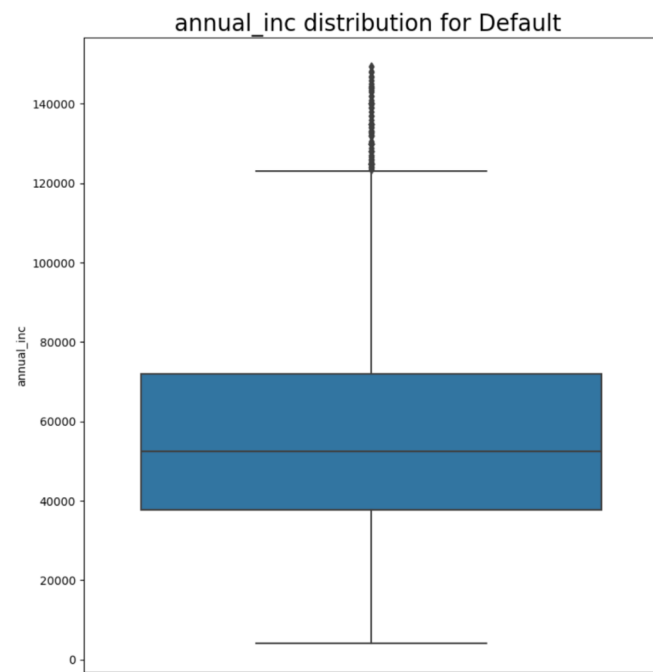


int\_rate distribution for Default



int\_rate distribution for Fully Paid



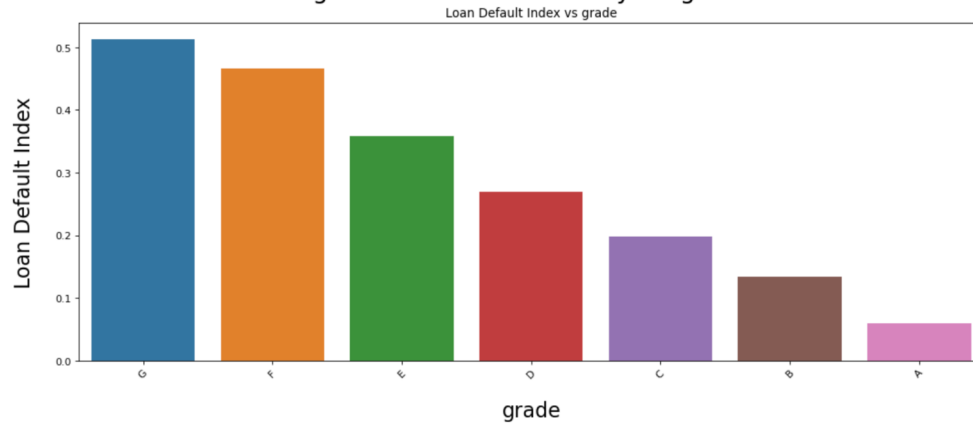


# Univariate analysis - Inference

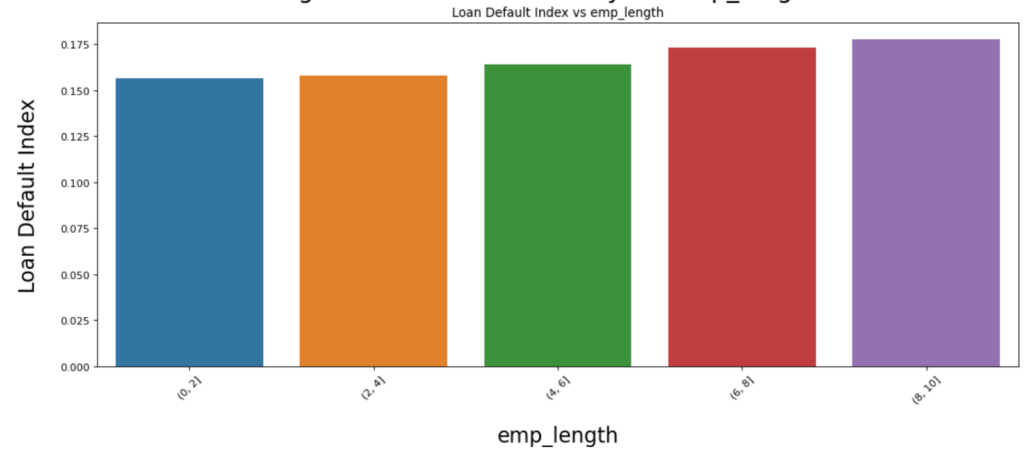
Field	Influences default ?	Inference
Loan tenure	Yes	Accounts with 60 month term have 15 percentage points higher probability of default
Annual Income	Yes	Accounts with income less than 100K have higher probability of default
Verification	No	No significant difference is observed between loan default and fully paid status based on whether the borrower's income source was verified or not
Employment length	No	No significant difference is observed between loan default and fully paid status based on the tenor of the employment of the loan borrowers
Interest Rate	Yes	Accounts with interest rate greater than 13% have higher probability of default
Bankruptcies record	Yes	Accounts with bankruptcy have 8 percentage points higher probability of default
Credit enquiries	Yes	Accounts with more than 1 occurrence have higher probability of default

# Segmented Univariate Analysis

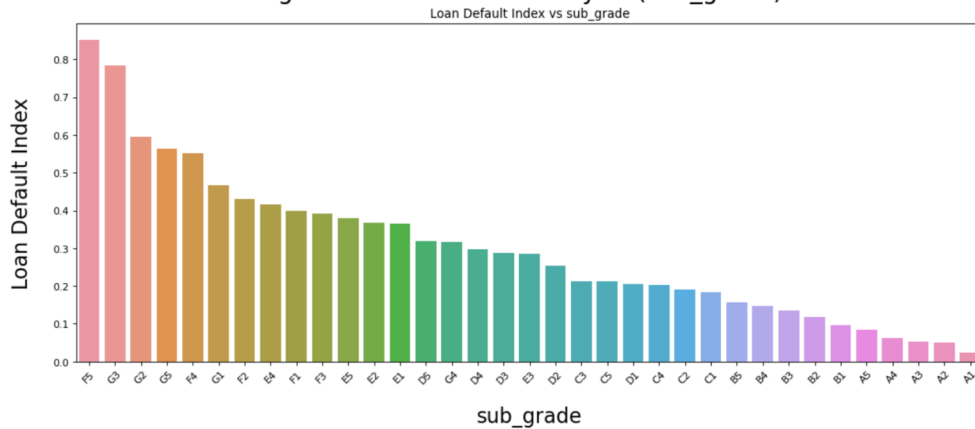
Segmented Univariate Analysis -grade



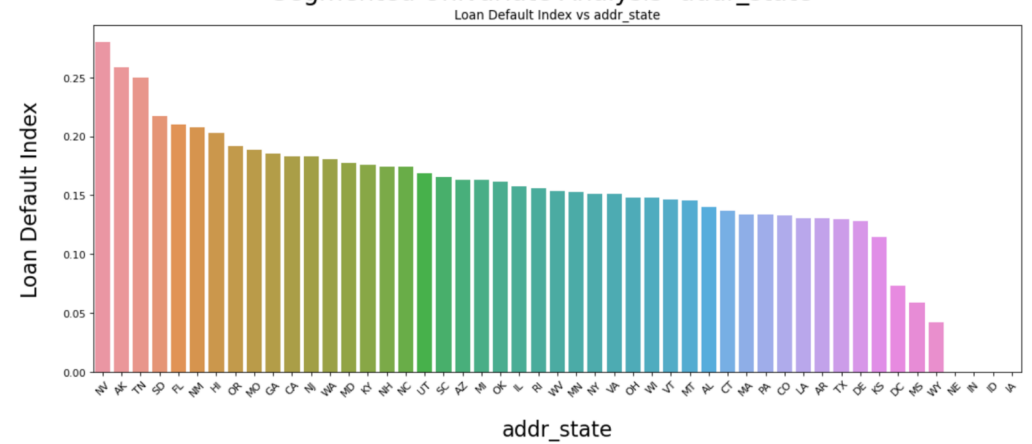
Segmented Univariate Analysis -emp\_length



Segmented Univariate Analysis (sub\_grade)



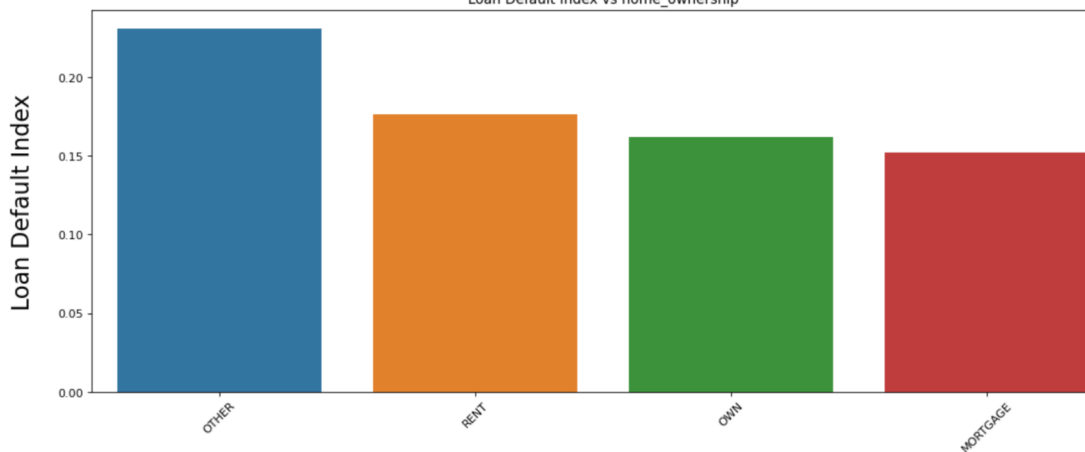
Segmented Univariate Analysis -addr\_state





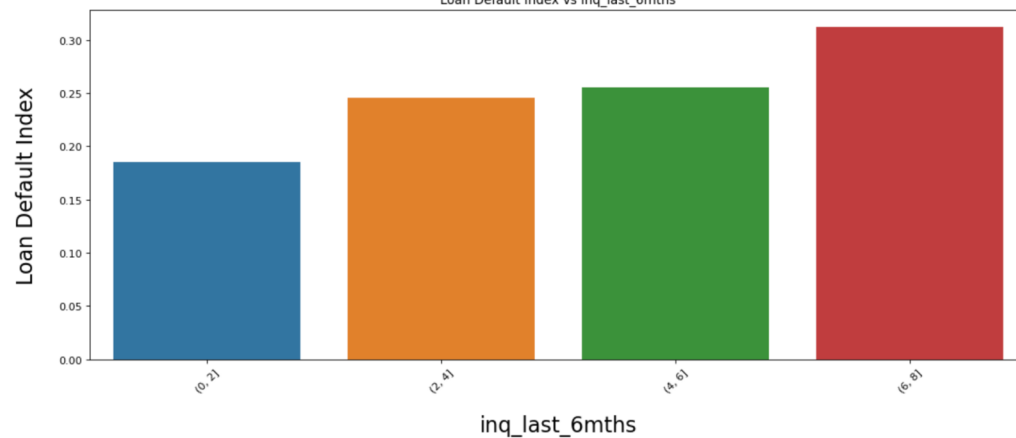
## Segmented Univariate Analysis (home\_ownership)

Loan Default Index vs home\_ownership



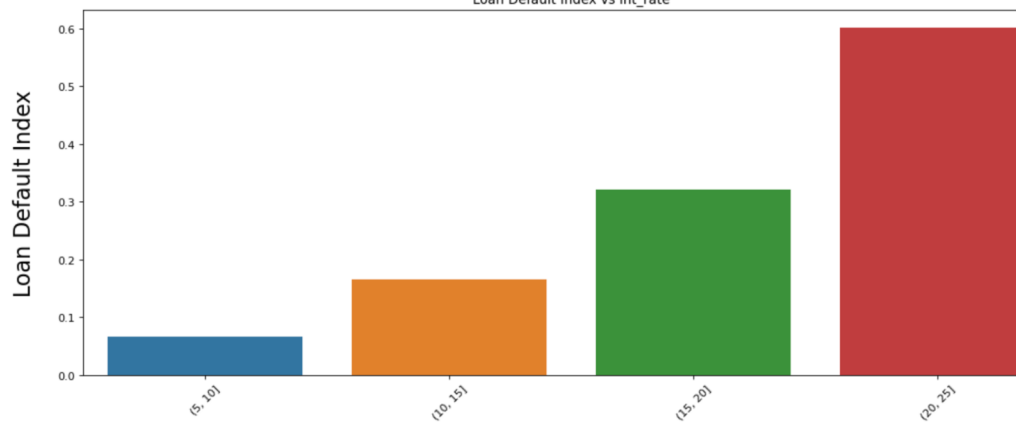
## Segmented Univariate Analysis -inq\_last\_6mths

Loan Default Index vs inq\_last\_6mths



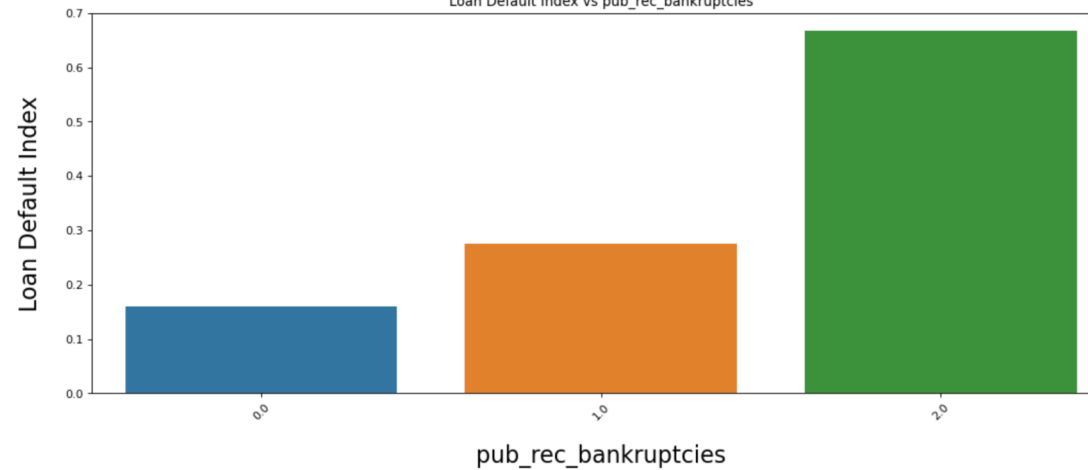
## Segmented Univariate Analysis -int\_rate

Loan Default Index vs int\_rate



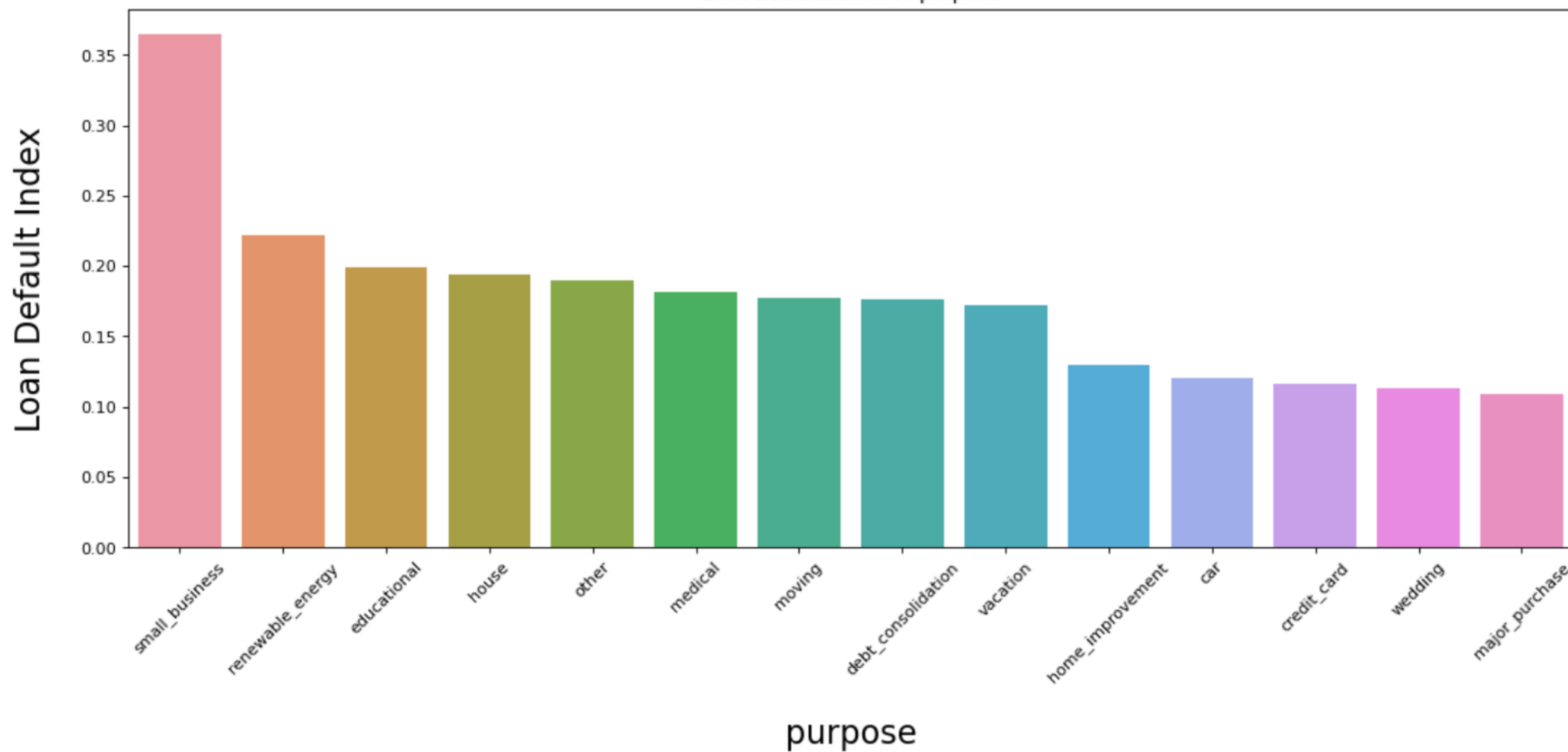
## Segmented Univariate Analysis -pub\_rec\_bankruptcies

Loan Default Index vs pub\_rec\_bankruptcies



# Segmented Univariate Analysis -purpose

Loan Default Index vs purpose



# Segmented Univariate Analysis- Inferences

Variable	Inference
Loan purpose	<p>The Loan default index is maximum at .35 for loan applicants borrowing to cater to the needs of small business indicating high risk lending by LC.</p> <p>Loans taken for personal purposes like wedding, purchase of car etc seem to have a lower index and hence less risky borrowers</p>
Interest Rate	<p>Loan Interest rates beyond 20% interest carries the highest default rate shown by the index value of 0.6. This is in line with the fact that higher the risk taken by LC, higher the returns.</p>
Credit inquiries	<p>From the data, it is observed that the default index is relatively higher at 0.3 for credit enquiries more than 6 times. Hence it is inferred that loan applicants who have a higher number of credit Inquiries are more likely to default compared to the ones that have lesser number of credit inquiries.</p>
State of applicant	<p>A relatively higher loan index value of 0.25 can be observed across loan applicants in the states of Nevada, Arkansas, Tennessee, South Dakota, Florida, Hawaii, New Mexico. Loan applicants from these states may have a higher probability of defaulting compared to the rest of the states in the USA</p>
Employment length	<p>The loan default index does not significantly vary across the length of the employment. Hence we assume that there may not be any influence on loan default by the number of years a loan applicant is employed</p>
Grade	<p>The risk of default seem to increase as the loan grade progresses from A to G as is evident from the default index values. Applicants assigned a loan grade of F &amp; G have a high default index of 0.5 as against low values for A&amp;B.</p>

# Bivariate Correlation

Attributes	Charged off trend	Fully Paid trend
Annual income	-ve	+ve
Pub_bankruptcies	+ve	-ve
Installment	+ve	-ve
Loan amount	+ve	-ve
Inq_last_6months	+ve	-ve
Interest rate	+ve	-ve

