



Credit Risk Analysis for Lending Club

BY: SAHIL AVASTHI & SAKET INGALE

Contents

- Business Context
- Dataset
- EDA Objective
- EDA Approach
- EDA Key Insights
- EDA Conclusion
- Recommendations

Business Context

Lending Club

- Lending Club is a financial technology company that operates a peer-to-peer (P2P) lending platform.
- Founded in 2006, it initially allowed individuals to lend money directly to other individuals seeking personal loans, bypassing traditional financial institutions. Over time, LendingClub expanded its offerings and services.

Business Context

- In this case study we'll be analyzing past loan applicants of a company which is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures for approving loan applications.
- Borrowers, can easily access lower interest rate loans through a fast online interface exposed by the company

Problem Statement

- 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.

Dataset

The Dataset contains the information for various loan listings for a lending company in USA distributed across various business attributes.

No of records in dataset: 39717 rows

No of attributes in dataset: 111 columns

Target Variable Name: loan_status

Loan Status Values

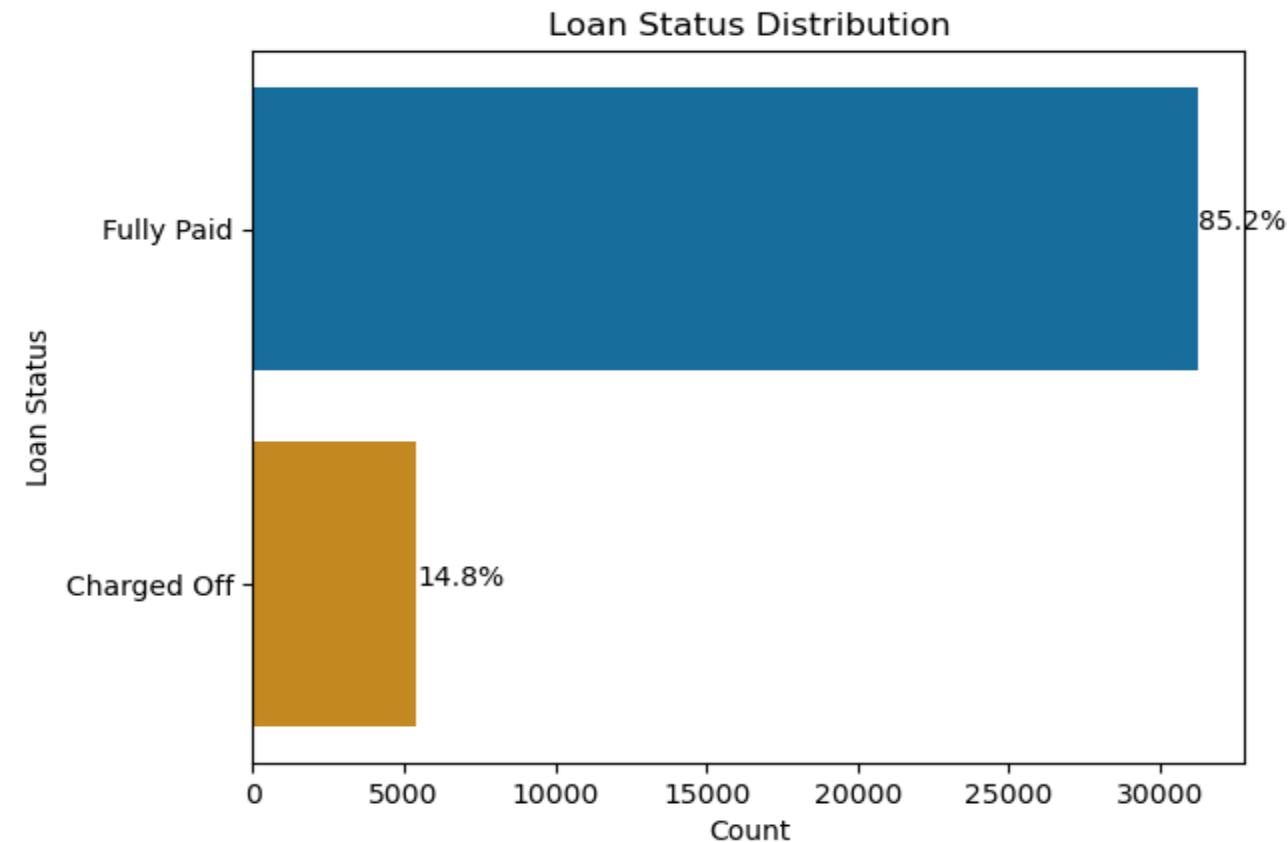
Fully paid

Applicant has fully paid the loan (the principal and the interest rate)

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

*Note: We have not considered records with 'Current' loan status since it will not be used for charge_off loans driving factors analysis



EDA Objective and Approach

EDA Objective

Understand how **consumer attributes** and **loan attributes** influence the tendency of default

Understand the **driving factors (or driver variables)** behind **loan default**, i.e. the variables which are strong indicators of default

EDA Approach Followed

Data Cleaning

- Fixing Rows and Columns
- Missing Values Treatment
- Filtering Useful Data

Univariate Analysis

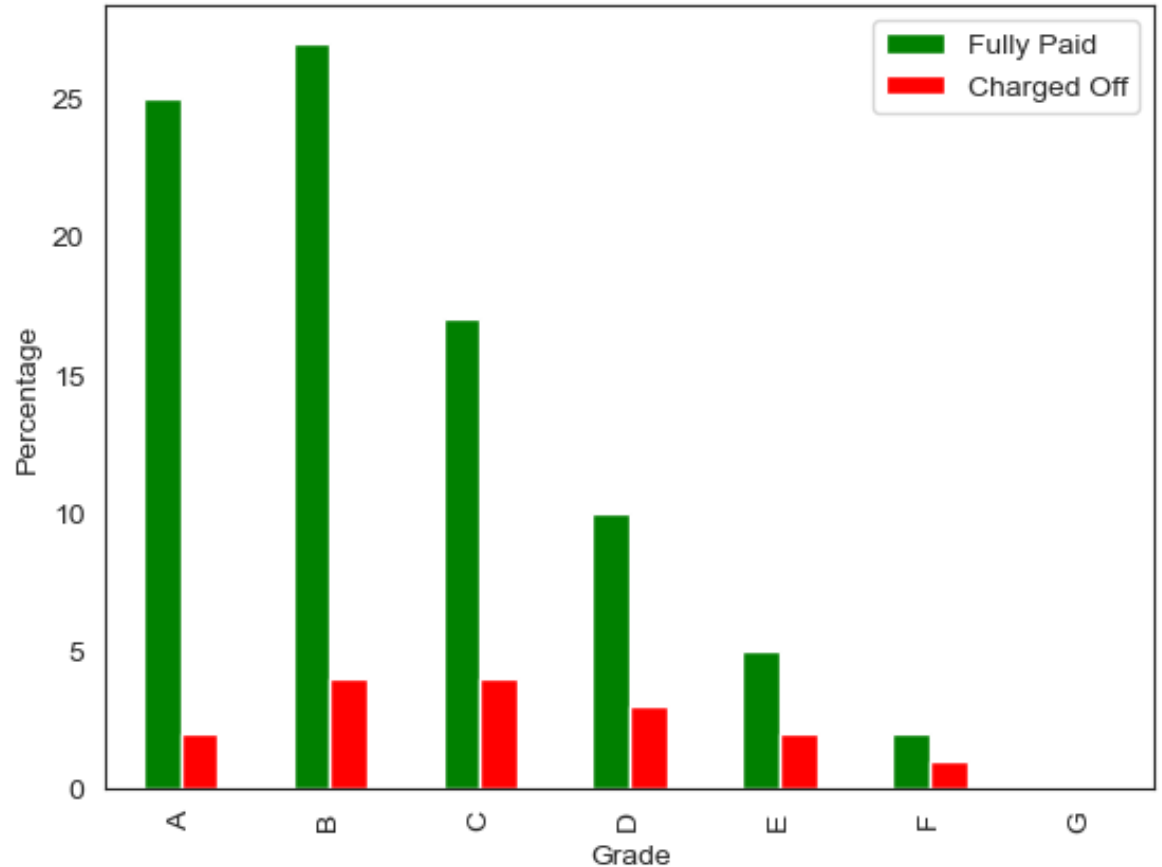
- Univariate Analysis on Important Variables
- Segmented Univariate Analysis against Target Column

Bivariate Analysis

- Correlation Analysis on Numerical Features
- Segmented Bivariate Analysis against Target Column

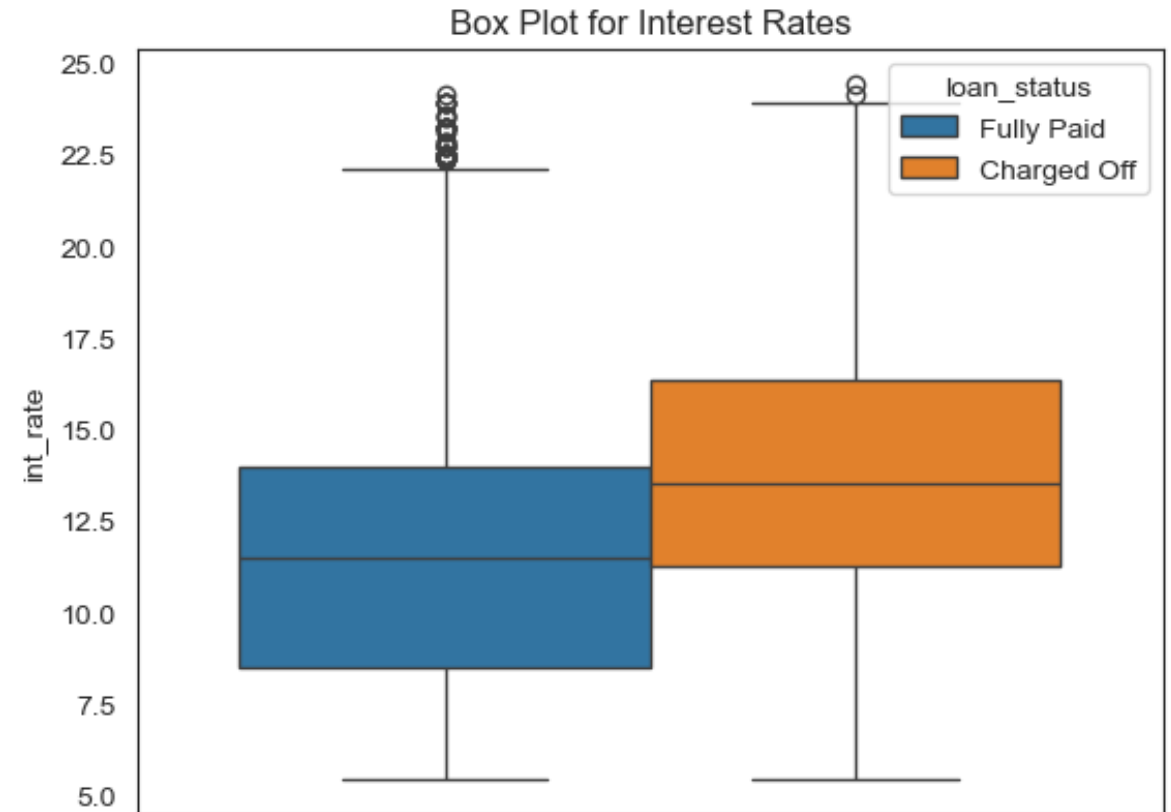
Key Insights about Grades

- Grades A to G Indicate various Loan Grades assigned by Lending Club to each loan record
- As we can observe in the percentage bar plot, It's clearly visible that Customers with Grades A E or F default less
- However, **Customers with Grade B or C default more**
- Customers with Grade B or C loans can be assessed further to avoid loan defaults in future
- Thus, Loan Grade Is one of the driving variable behind default loans



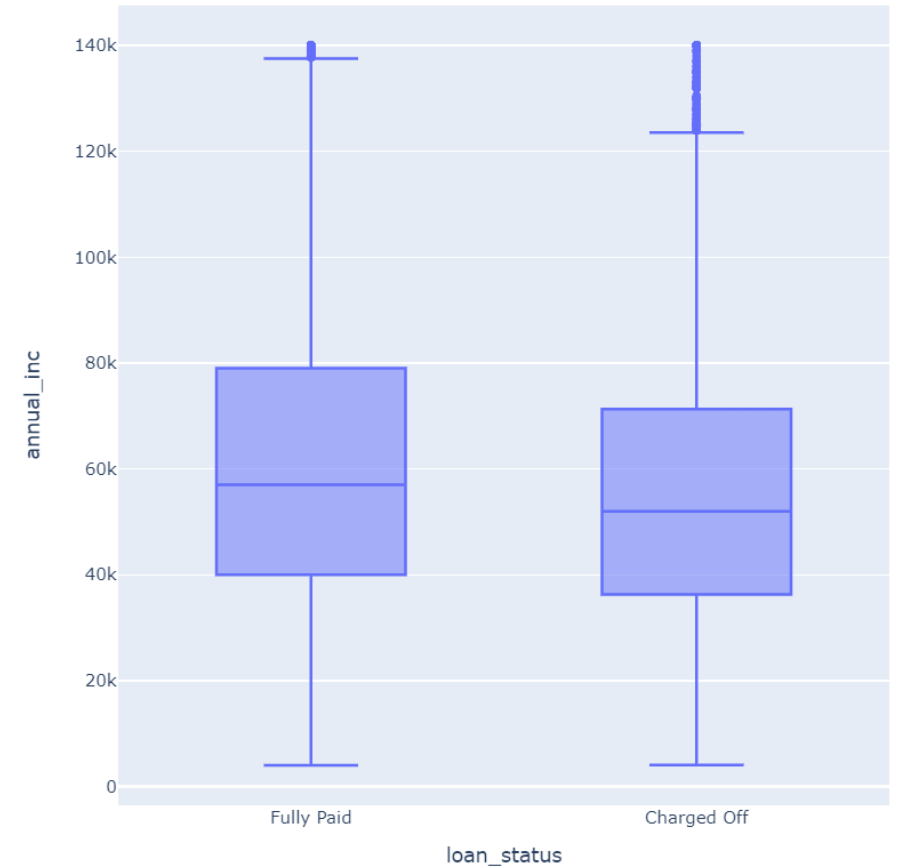
Key Insights about Interest Rate

- Int_rate column indicates the interest rate for each loan record in percentage
- As we can observe in the plot, It's clearly visible that Customers with higher interest tend to default
- However, **Customers with interest rate more than 13.75% appear to default more likely**
- Customers with high interest rate loans can be assessed further to avoid loan defaults in future
- Thus, Interest Rate Is one of the driving variable behind default loans



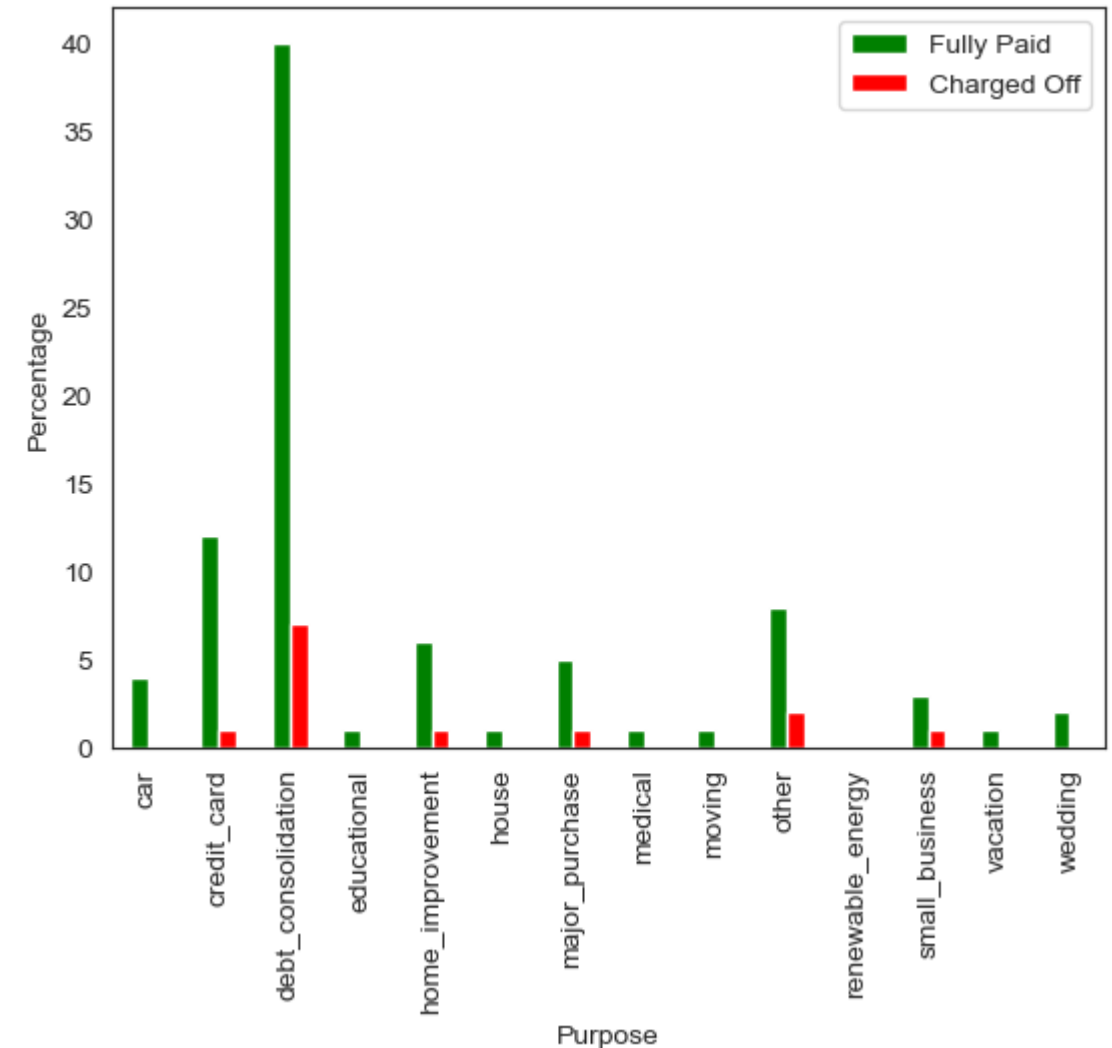
Key Insights about Annual Income

- `annual_inc` column indicates self-reported annual income provided by the borrower during registration.
- As we can observe in the plot, It's clearly visible that Income distribution (25 - 75 percentile) for Charged off is 36 - 72 ; Income distribution (25 - 75 percentile) for Fully Paid is 40 - 79.
- This means **customers income between 36 and 40 has higher chance of default**
- Customers with this income range can be assessed further to avoid loan defaults in future
- Thus, Annual Income Is one of the driving variable behind default loans



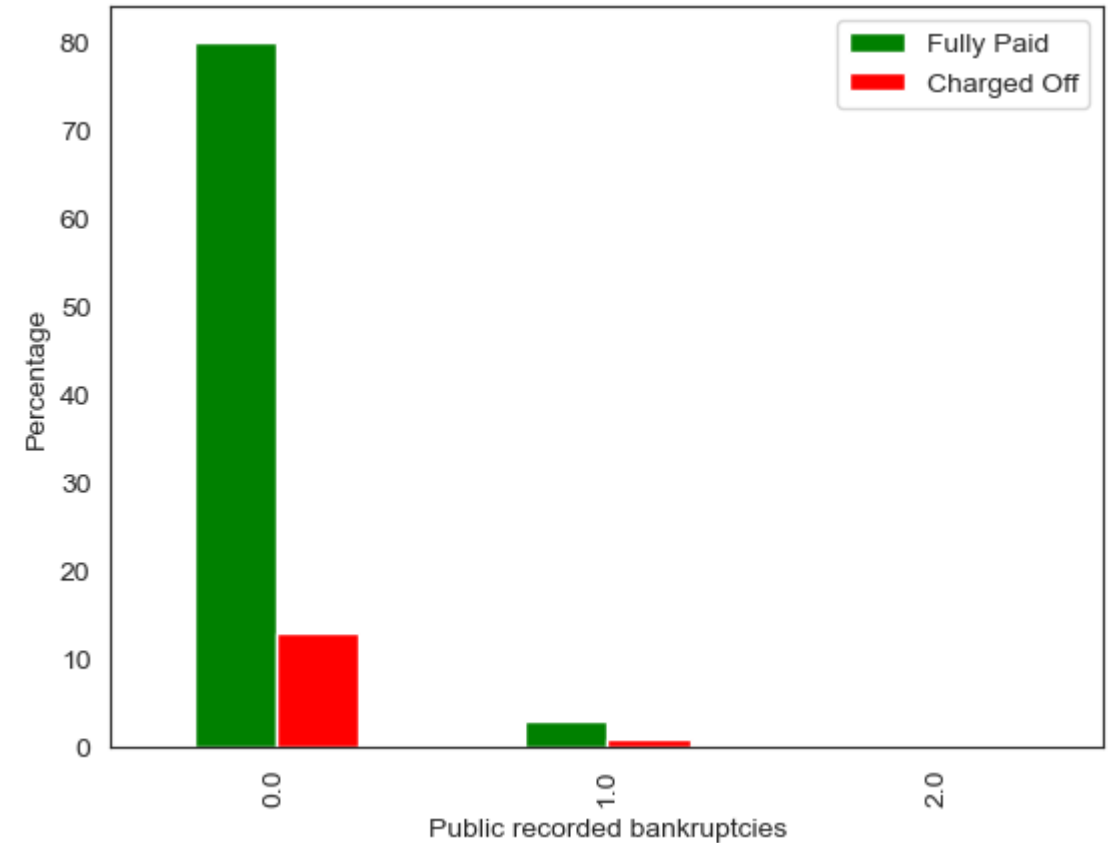
Key Insights about Loan Purpose

- Purpose column Indicate A category provided by the borrower for the loan request.
- Customers who took loan for debt consolidation are more prone to default as well as they can repay the full loan
- Hence loans taken for debt consolidation can be approved with further check only
- Thus, Loan Purpose is one of the driving variable behind default loans



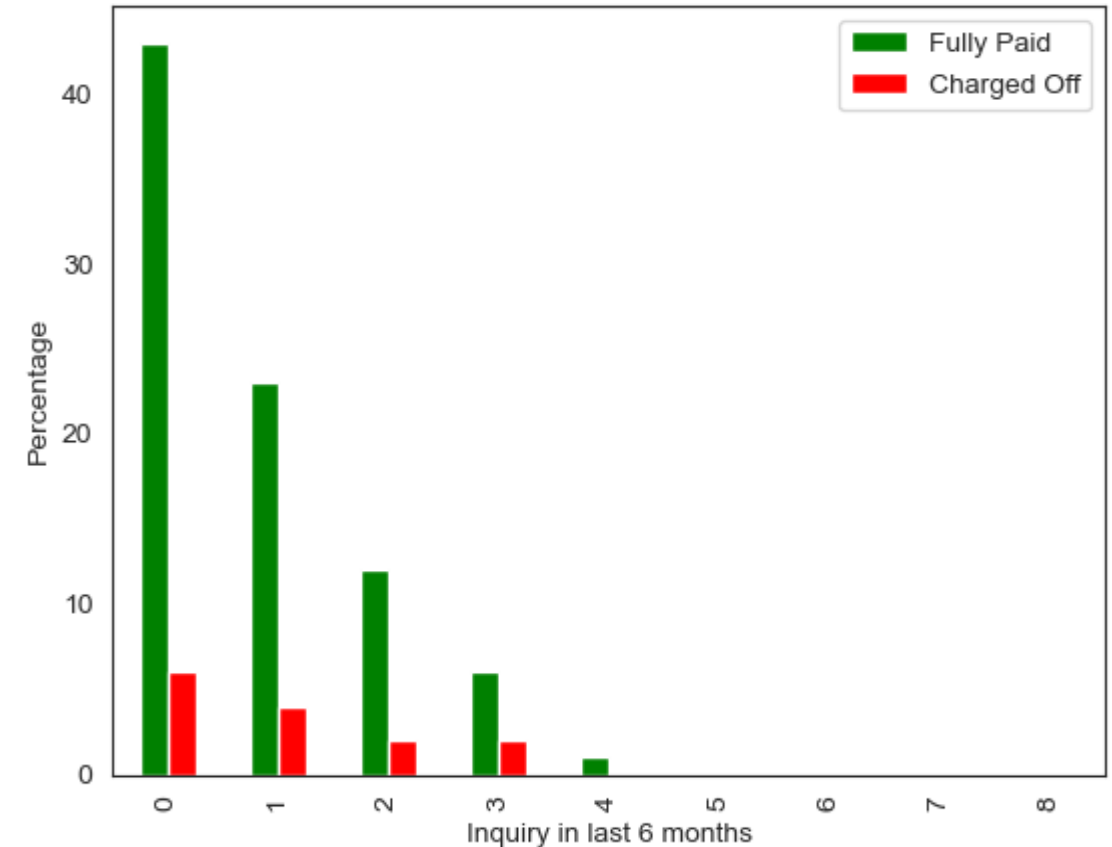
Key Insights about Public Bankruptcies

- Column `pub_rec_bankruptcies` Indicates Number of public record bankruptcies
- As we can observe in the plot, customers with both zero and more bankruptcies tend to default
- Chances of default is higher (around 50%) for people with more than zero public bankruptcies
- Thus, number of public bankruptcies for customers can be analyzed before approving loans



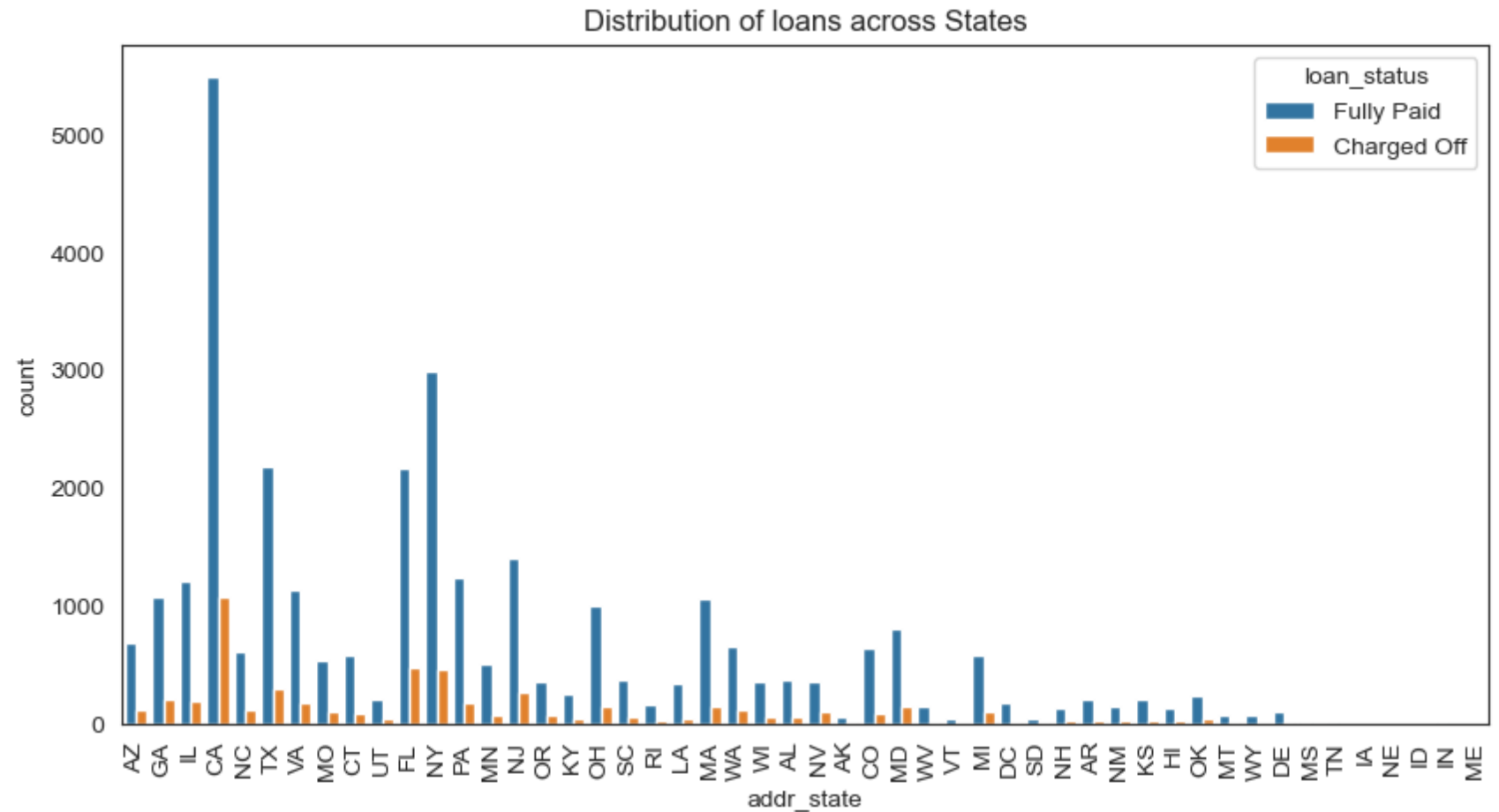
Key Insights about No of Inquiries

- Column `inq_last_6mths` Indicates The number of inquiries in past 6 months
- As we can observe in the plot, ratio of heights of percentage appear to be higher for more than 1 inquiry in last 6 months
- Thus, number of inquiries in last 6 months can be a good indicator towards default loans



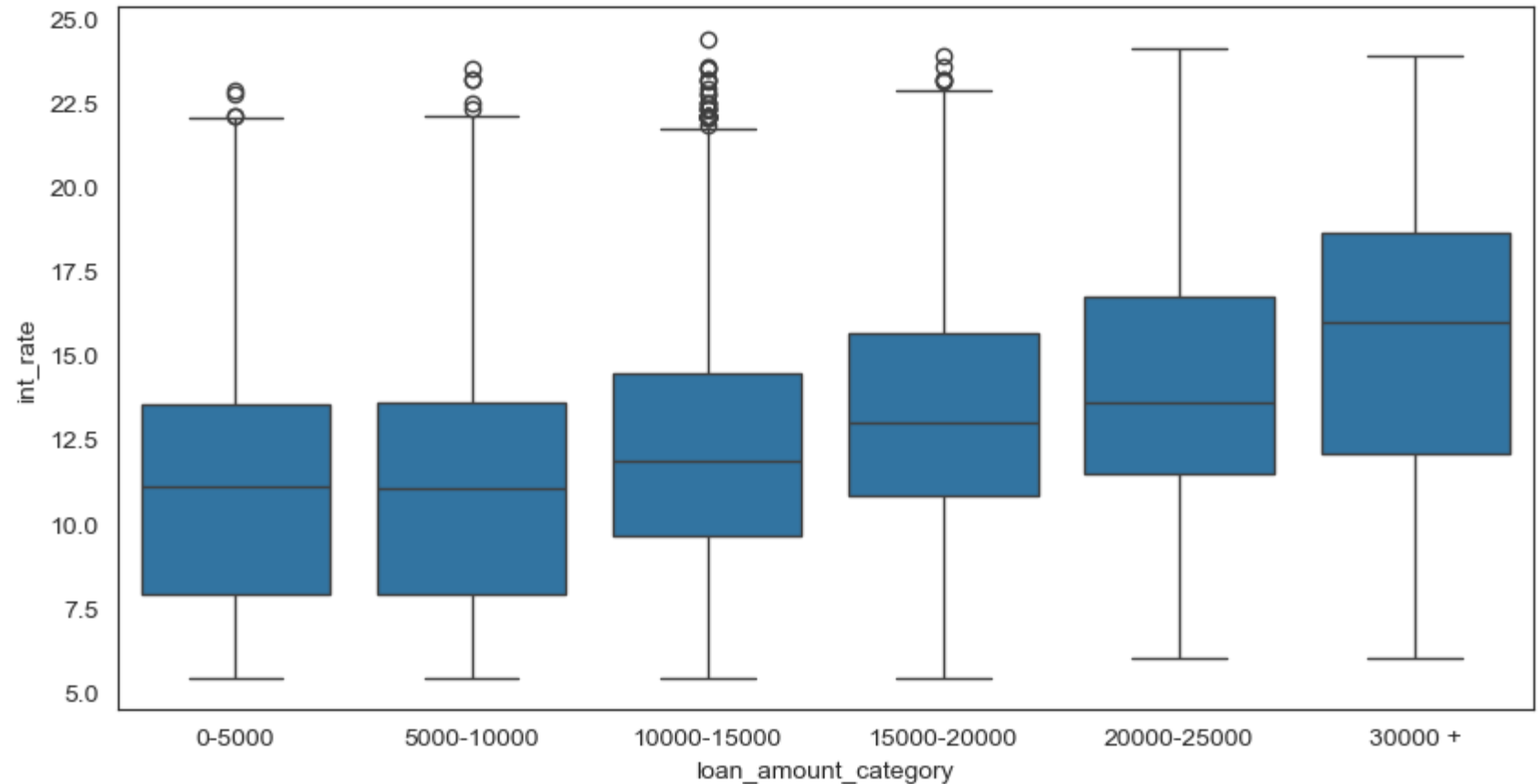
Key Insights about Customer State

- Column `addr_state`
Indicates the state provided by the borrower in the loan application
- As we can observe in the plot, people in California state tend to take loans more
- States CA, FL and NY show comparatively higher chances of default loans



Key Insights Interest rate Vs Loan Amount

- As we can observe in the bivariate boxplot, people in there is increasing trend between loan amount and interest rates
- Higher interest rates significantly increase the total loan amount
- This might motivate customers to charge off during adverse financial conditions



EDA Conclusion

Driving factors behind charging off loans identified using Exploratory Data Analysis are

1. Grades
2. Interest Rates
3. Annual income
4. Purpose of the loan
5. Public recorded bankruptcies
6. No of inquiries in last 6 months

Other key insights that are identified are –

1. Customers with 10+ years & less than 1 year have higher probability to default
2. Customers who own property default less; however, Customers on Rent are slightly more prone to default
3. Customers who took loan in CA are more prone to default as well as they can repay the full loan
4. Customers with 1 inquiries in last 6 months are more likely to default
5. Larger amount of loan contributes to higher chance of defaulting as it increases the interest rate

Recommendations

1. DTI impacts the repayment of the loans. This we have verified through Grades, interest rates and annual income therefore **before approving loans DTI should be verified correctly and thoroughly**
2. Most of the analysis confirm that loans which are approved with higher Grades has high probability of repayment therefore this concludes that **banker should approve more loans with higher grades**
3. Customers who are taking loans for debt consolidation or credit card should be thoroughly verified against **other factors such as loan Grades and DTI**
4. We have observed that there is negative correlation between Public recorded bankruptcies & Installments therefore **customers with more Public recorded bankruptcies should be verified correctly and thoroughly**
5. We have observed that there is negative correlation between Public records & loan amounts therefore **customers with more Public records should be verified correctly and thoroughly before approving loans**