

Lending Club case study

Analysis & Report by:

Chandana Kalyan
Sri Krishna Chaitanya Hari

Nov 2024

Problem Statement

Identify risky loan patterns from Lending Club dataset and gather some insights which can help reduce future losses.

Objective

The goal of this case study is to apply Exploratory Data Analysis (EDA) techniques on a given dataset and effectively present findings in a business-focused manner

Process flow



1. Understanding of Data



2. Cleaning of Data



3. EDA

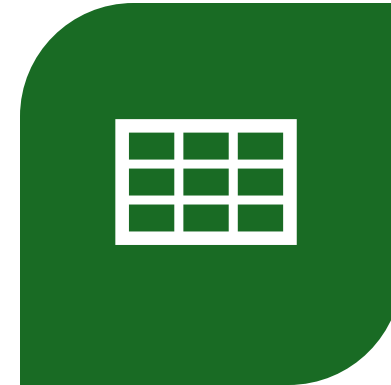


4. Observations/
recommendations

Understanding of Data



IT'S A LOAN DATA SET WITH NEARLY 40 THOUSAND RECORDS CONTAINING LOAN STATUS, FINANCIAL ASPECTS, CUSTOMER PROFILE ALONG WITH STAGE-WISE APPLICATION STATUS



WE ALSO RECEIVED DATA DICTIONARY, DESCRIBING EACH VARIABLE IN THE DATA SET

Cleaning Data

The initial data has approximately 39 thousand rows and 111 columns.

All the columns with Null data (both 100% and less than 5%) have been dropped.

All the records with “Current” status have been dropped.

All the columns with Unique value as 1, have been dropped.

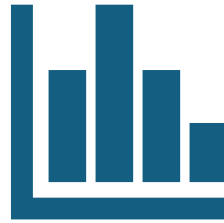
All the columns with irrelevant data have been dropped.

Fixing special character and incorrect data types.

EDA



Univariate Analysis

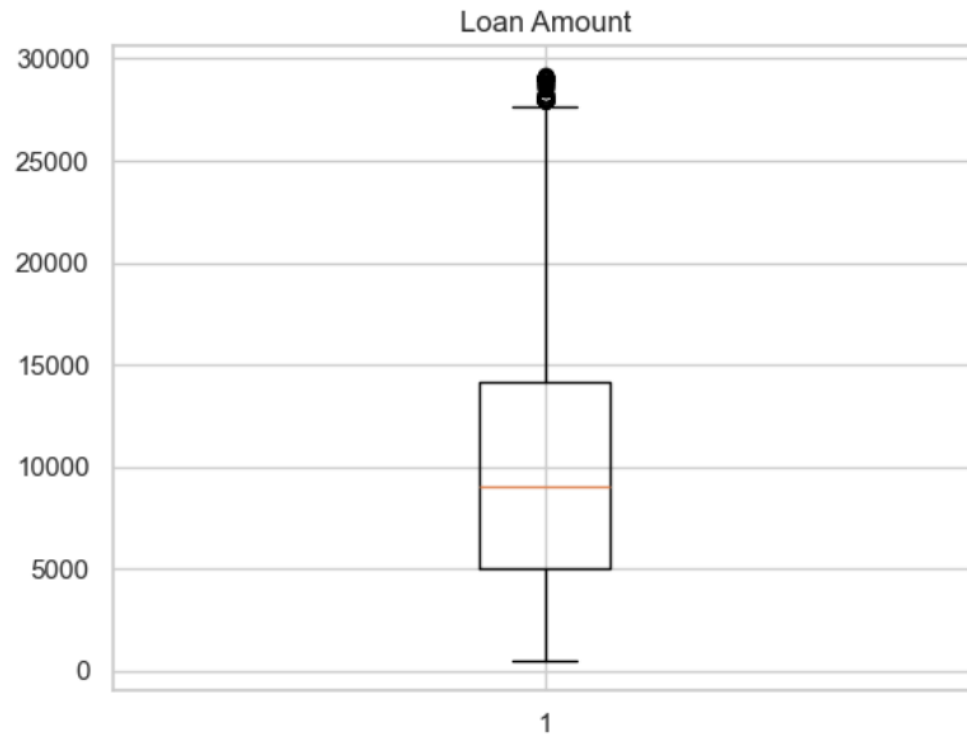


Bi- Variate Analysis



Segmented Analysis

Univariate Analysis – Loan Amount distribution, Annual Income distribution



Notes:

- Max. loan amounts fall in between 5000 and 15000
- Median is at 9500



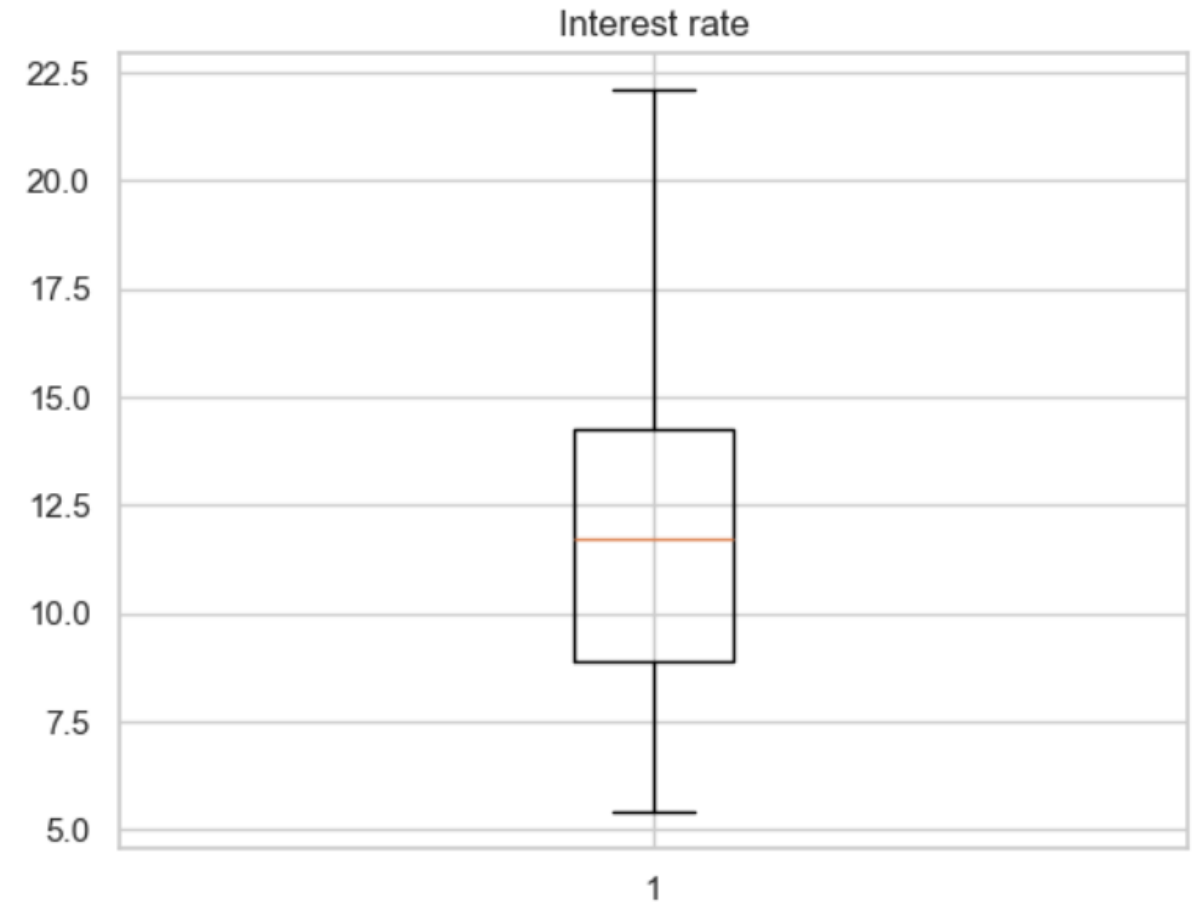
Notes:

- 75% of annual income is near 80000
- Range of annual income varies from 4000 to 139500

Univariate Analysis – Rate of Interest distribution

Notes:

- The max. rate of interest provided is 22.1% and lowest at 5.4%
- Most of the ROI fall under 14% with median at 11.7%

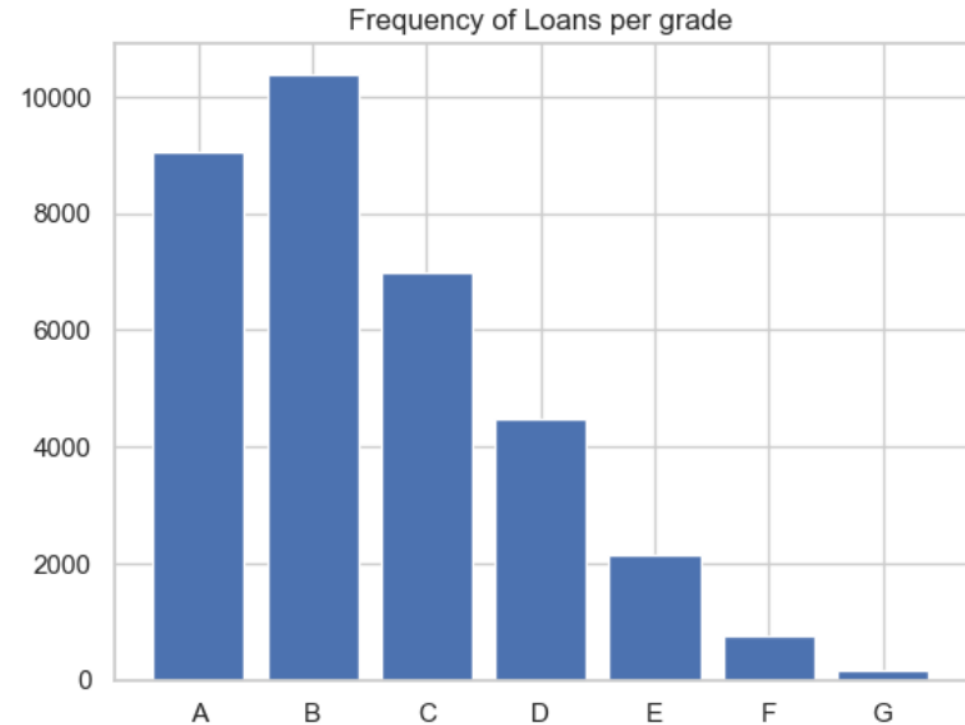


Univariate Analysis – Loan term & Grade



Notes:

- Most loans are granted for 36 term limit.



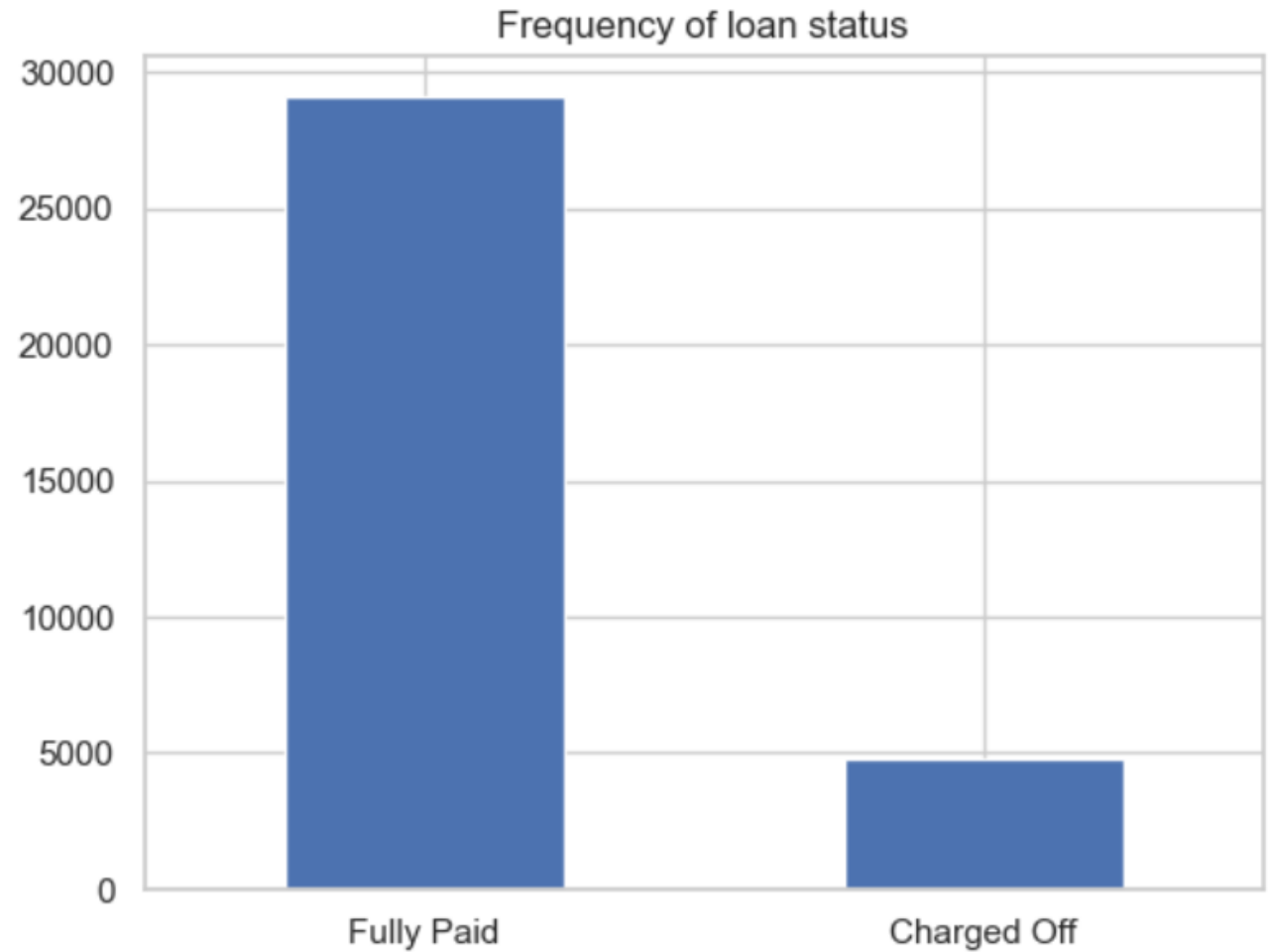
Notes:

- Highest loans are granted for category: Grade B
- The number of loans approved reduced with grade, signifying the risk with increase in grade

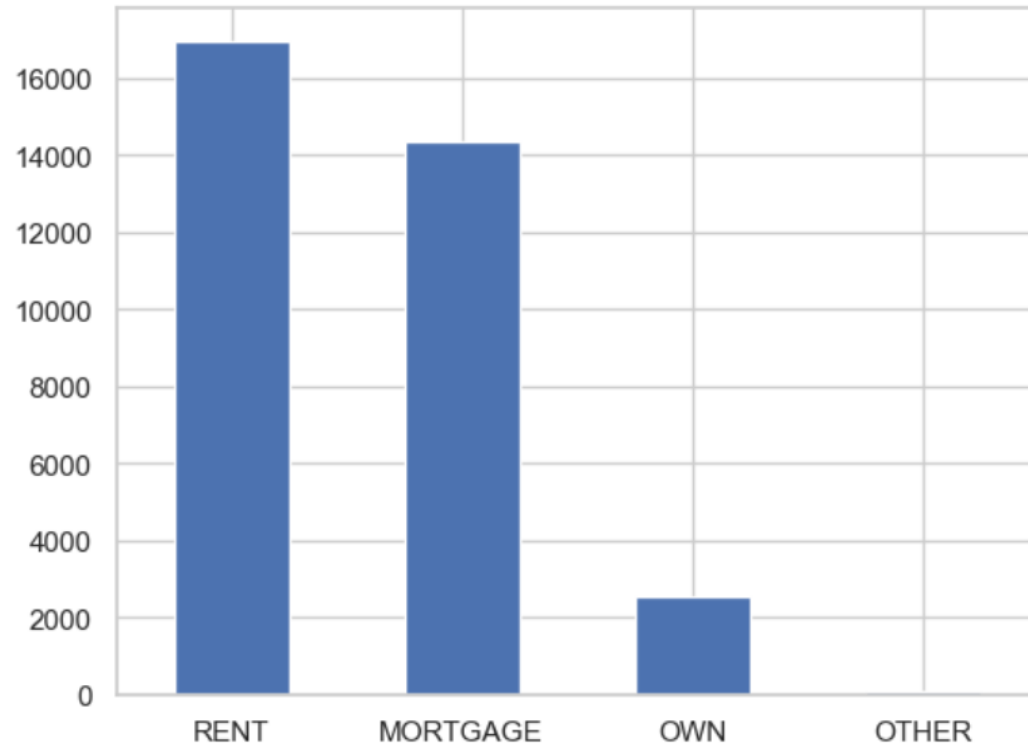
Univariate Analysis – Loan status

Notes:

- Charged Off loans are comparatively very less (< 20%)

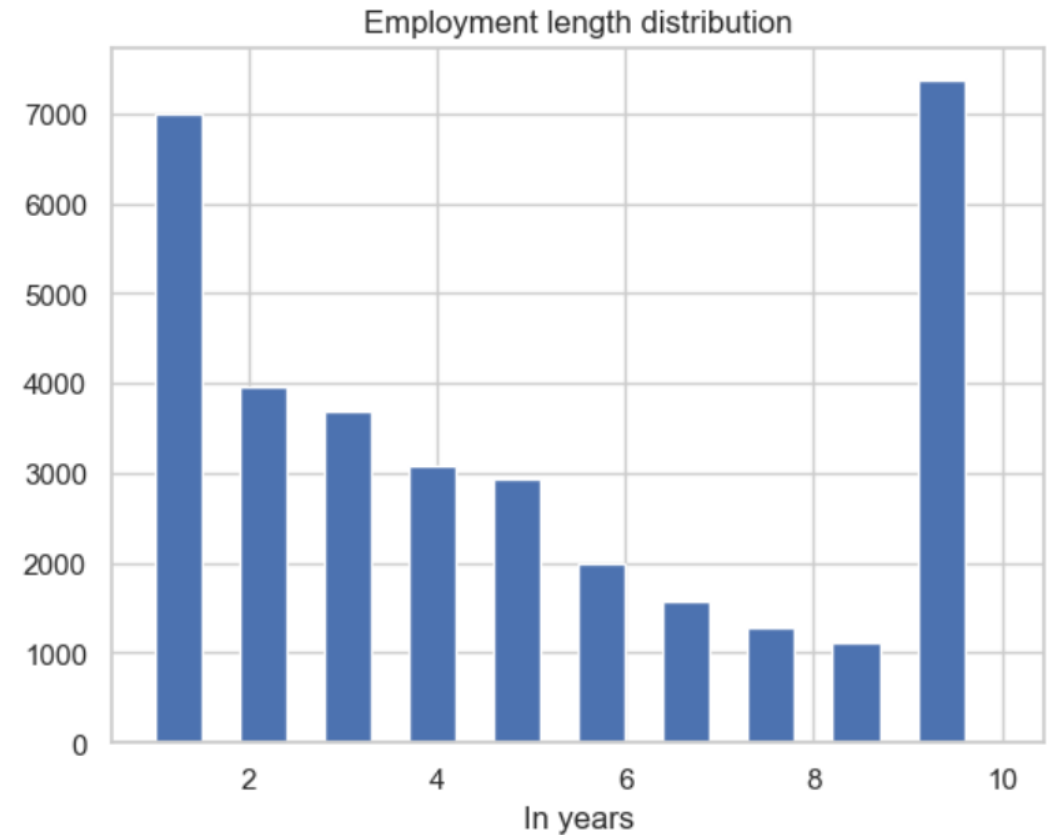


Univariate for home ownership & employment length



Notes:

- Customers who are living on Rent have taken more loans
- Customers having own homes opted for less loans



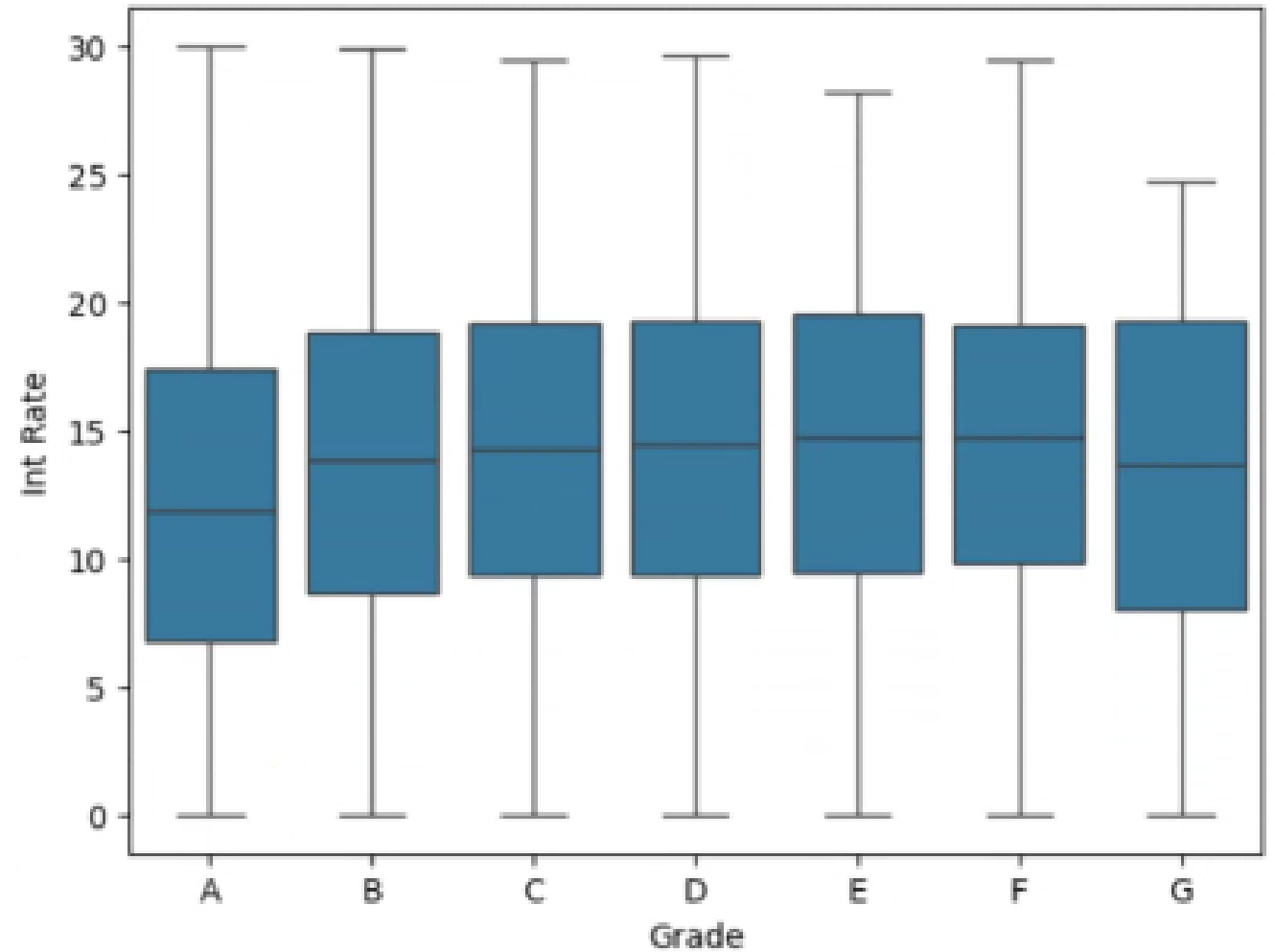
Notes:

- People employed for more than 9 years tend to take most number of loans

Segmented Univariate for Interest rate by Grade

Notes:

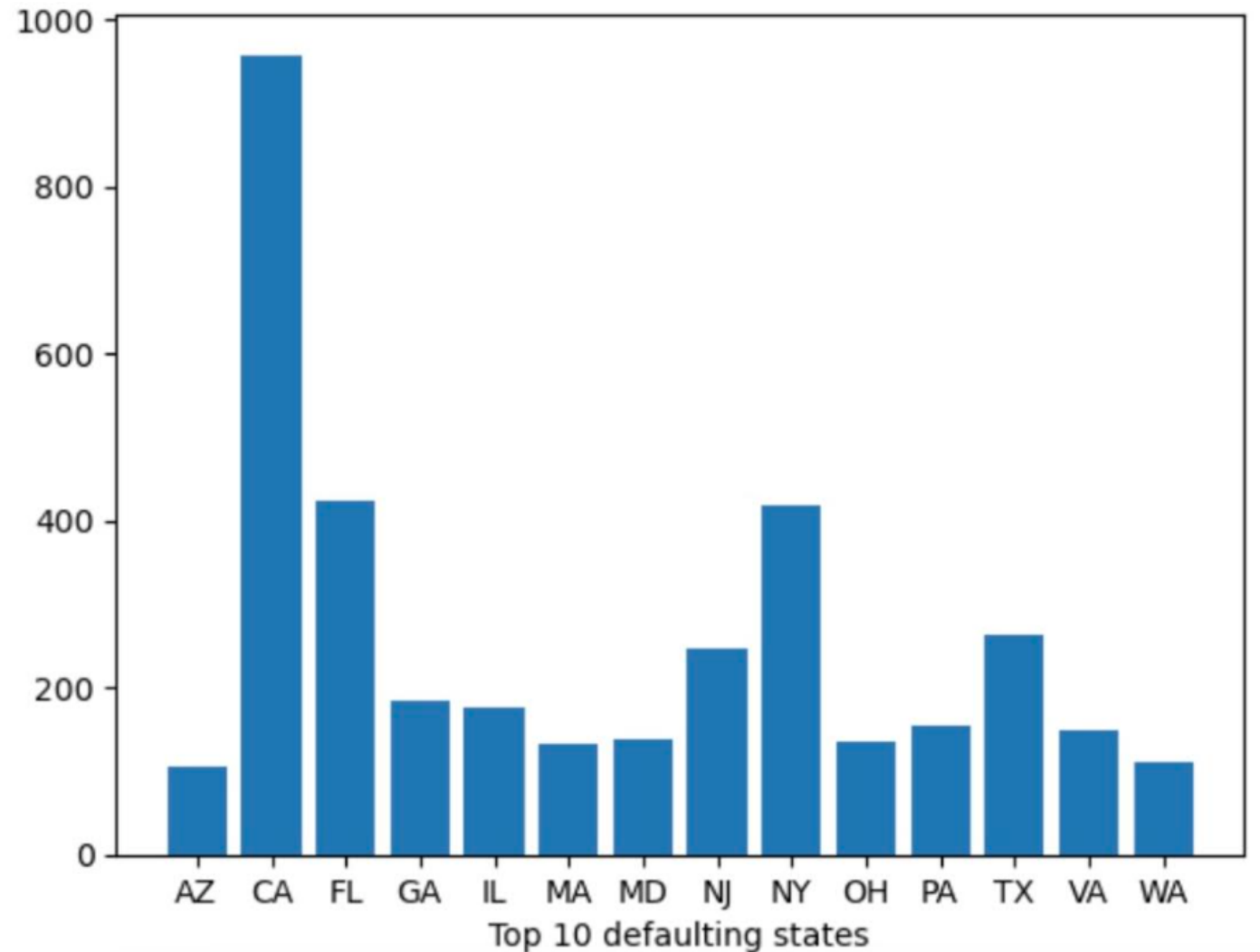
- A grade pays lower interest compared to others.



Segmented Univariate - Top 10 defaulting states

Notes:

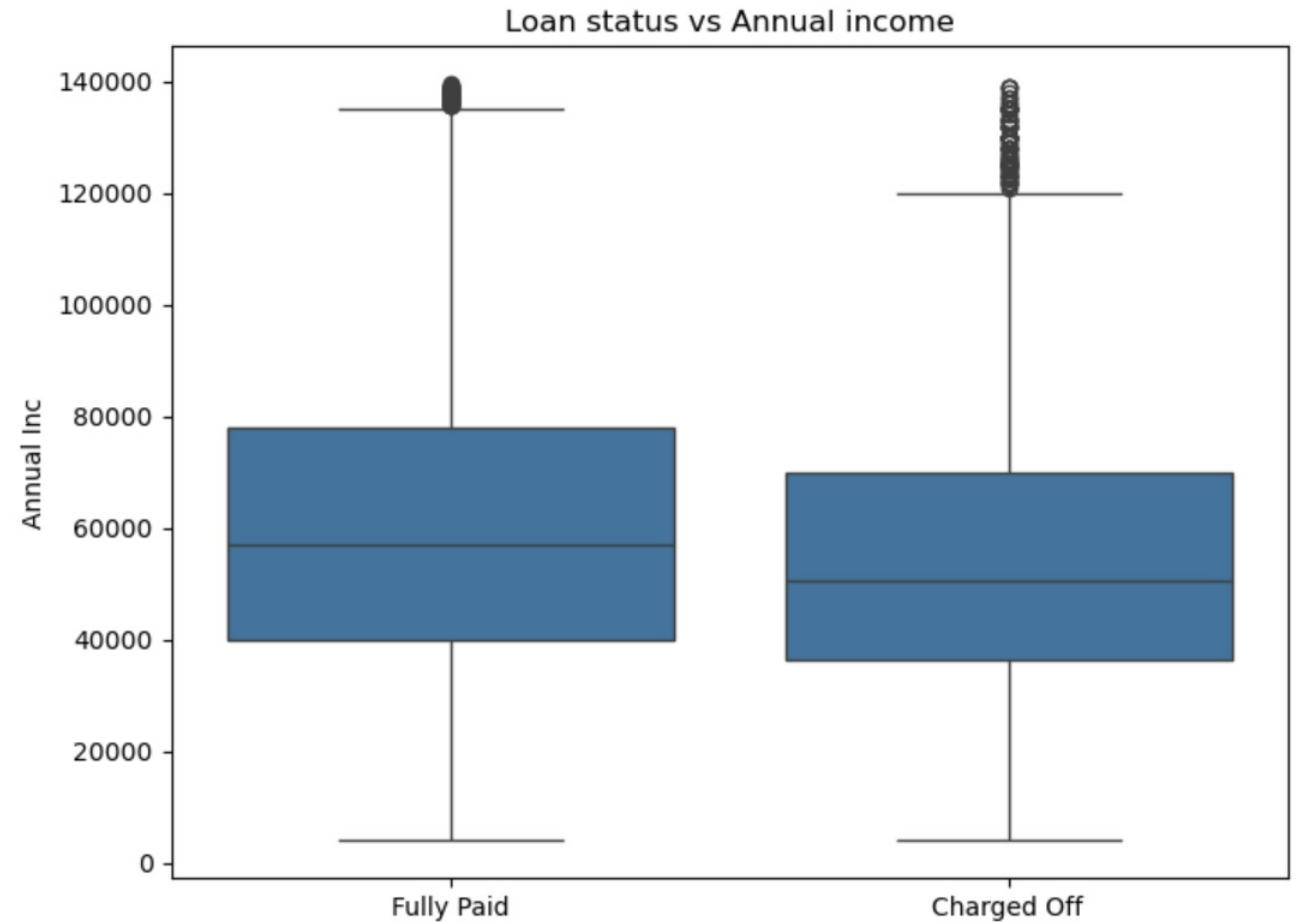
- Top 3 States like California, Florida and New York have highest default rates



Segmented Univariate Analysis – Annual income

Notes:

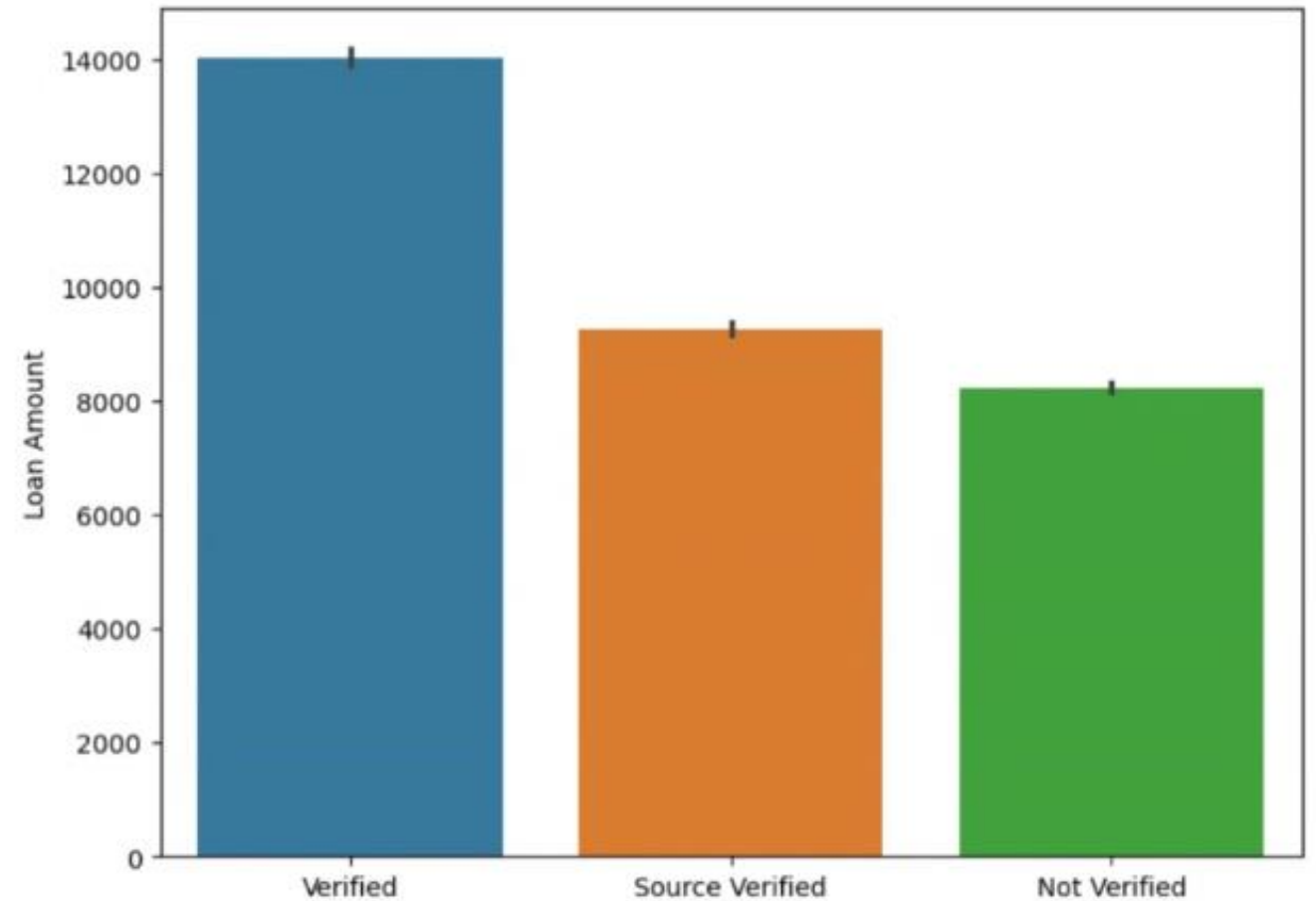
- The median of Annual Income for both Fully Paid and Charged Off loans



Loan Amount vs. Verification status

Notes:

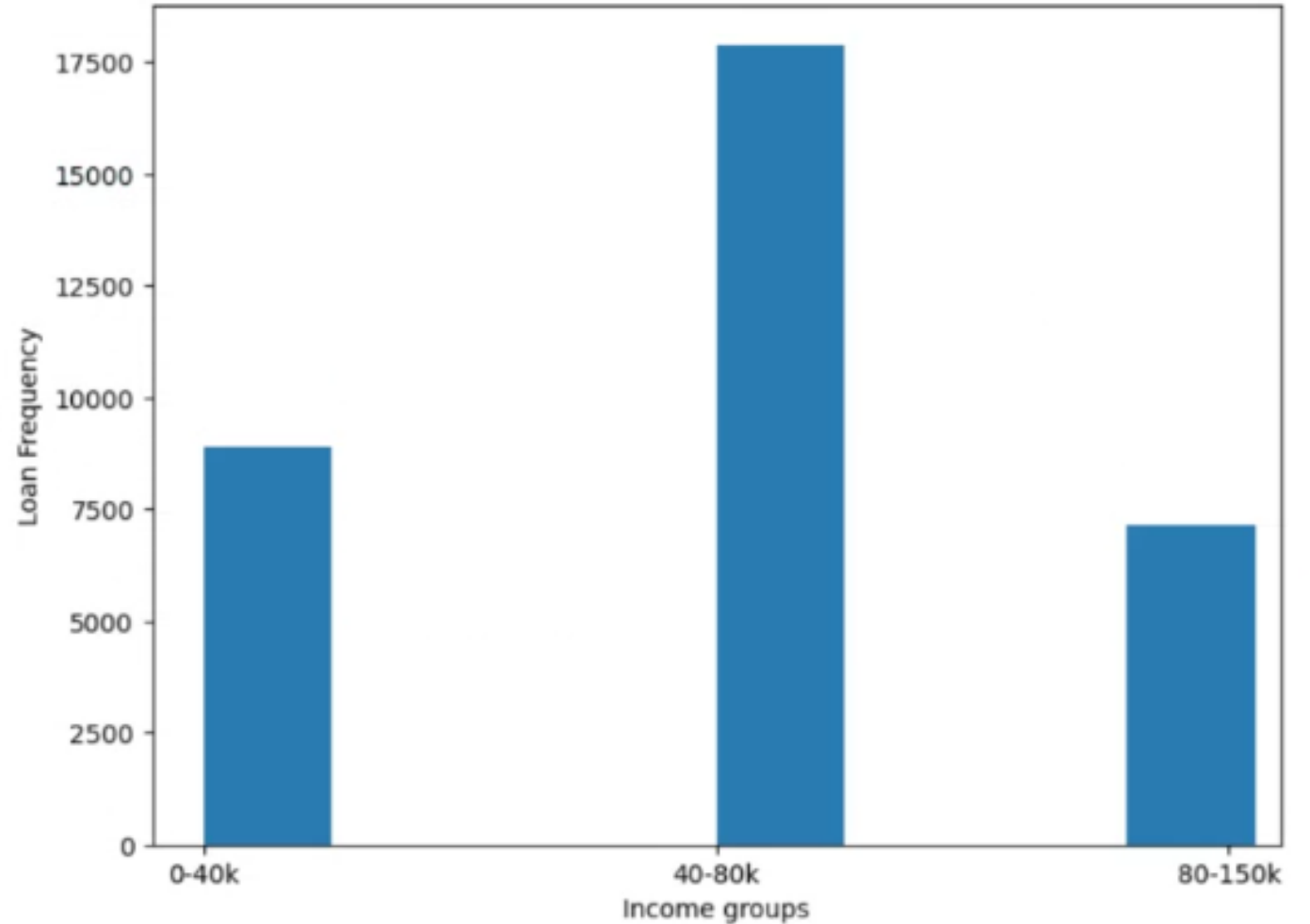
- Verified applications tend to get more loan amount approved compared to source verified and non-verified



Loan frequency per Income group

Notes:

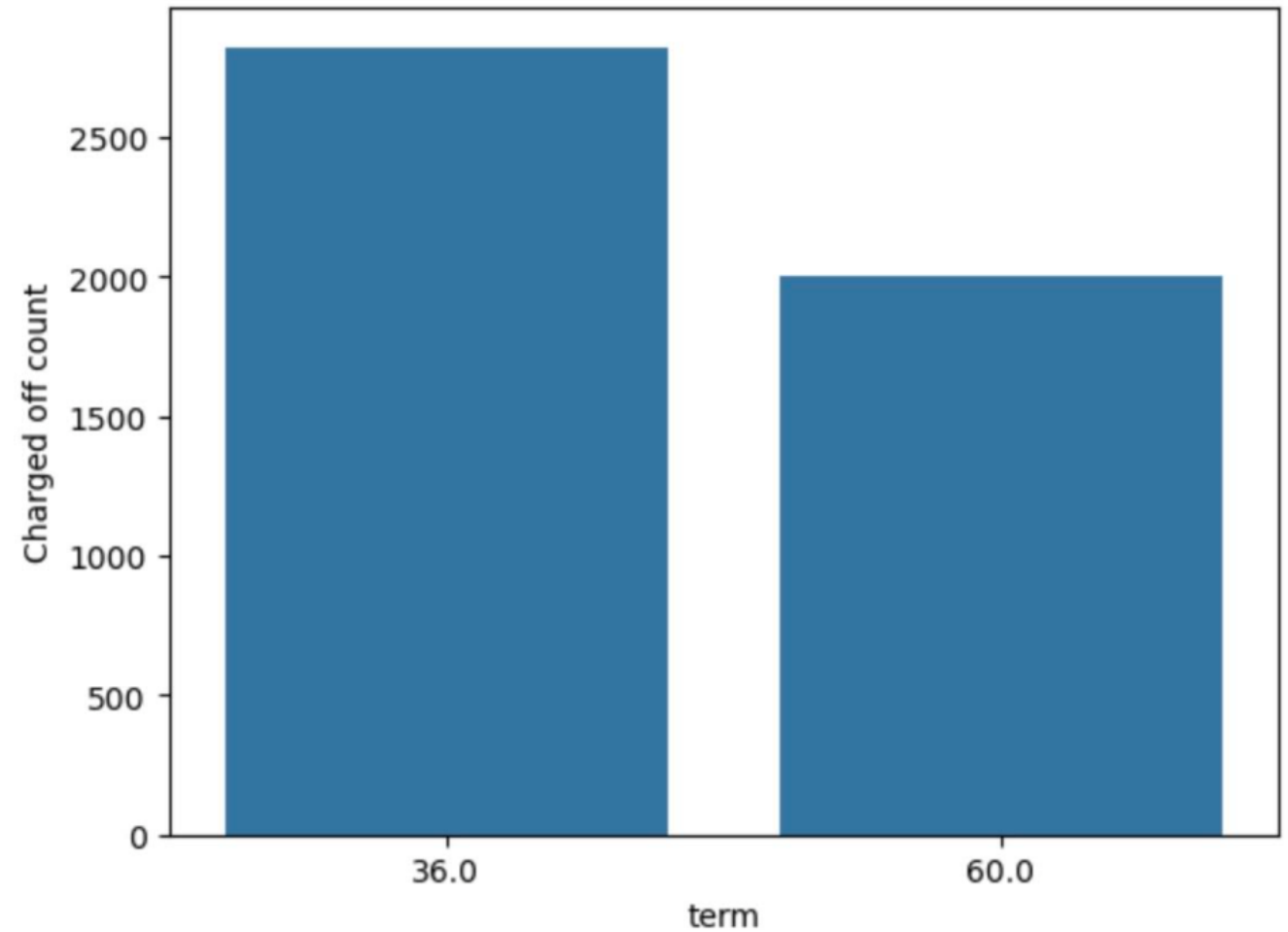
- People with income falling in the bucket range of 40k-80k opted for more loans



Loan Term for Charged off loans

Note:

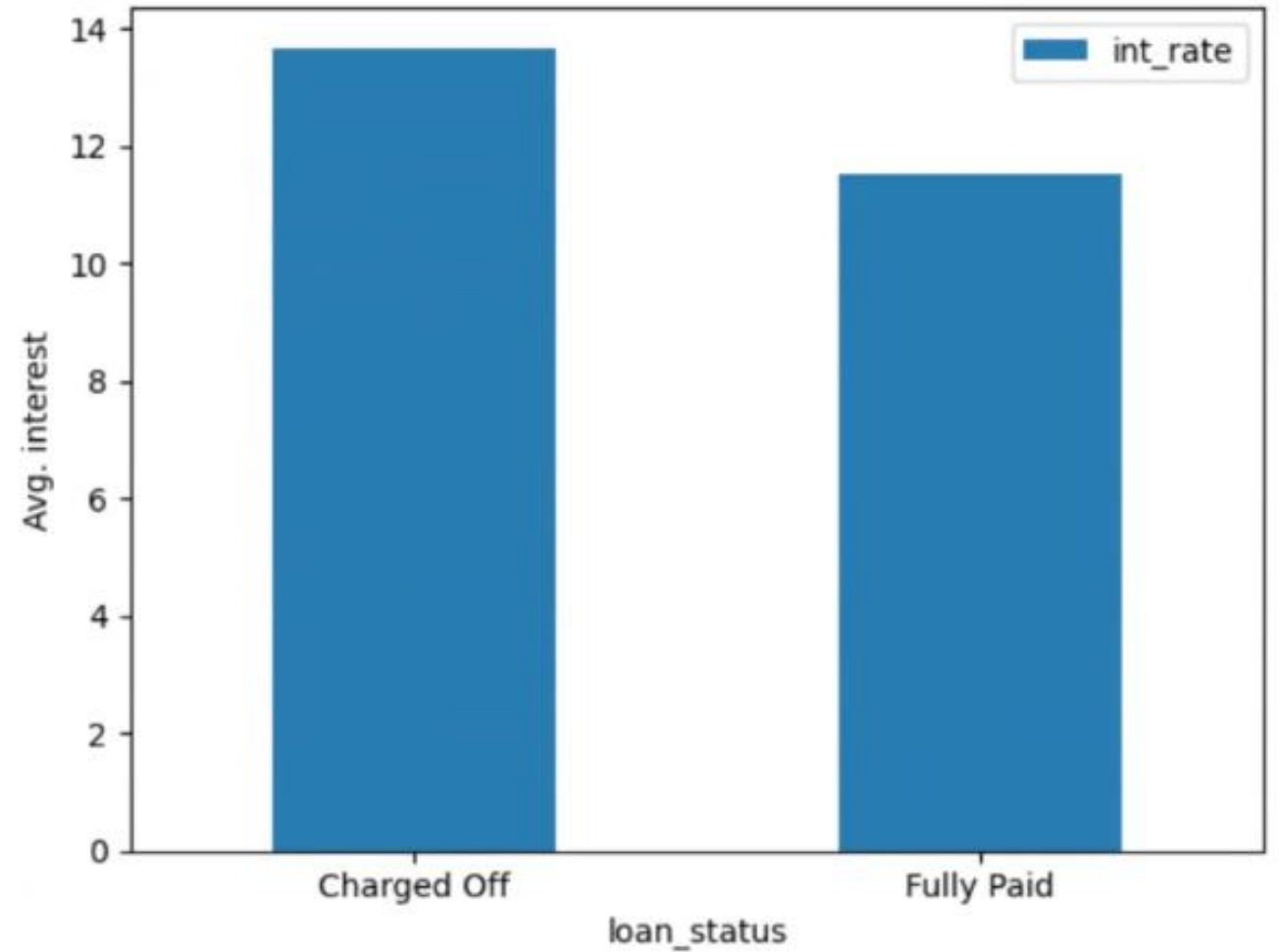
- People who opted for 36 months defaulted more than loans with 60 months.



Charged Off vs Fully Paid based on Interest rate

Notes:

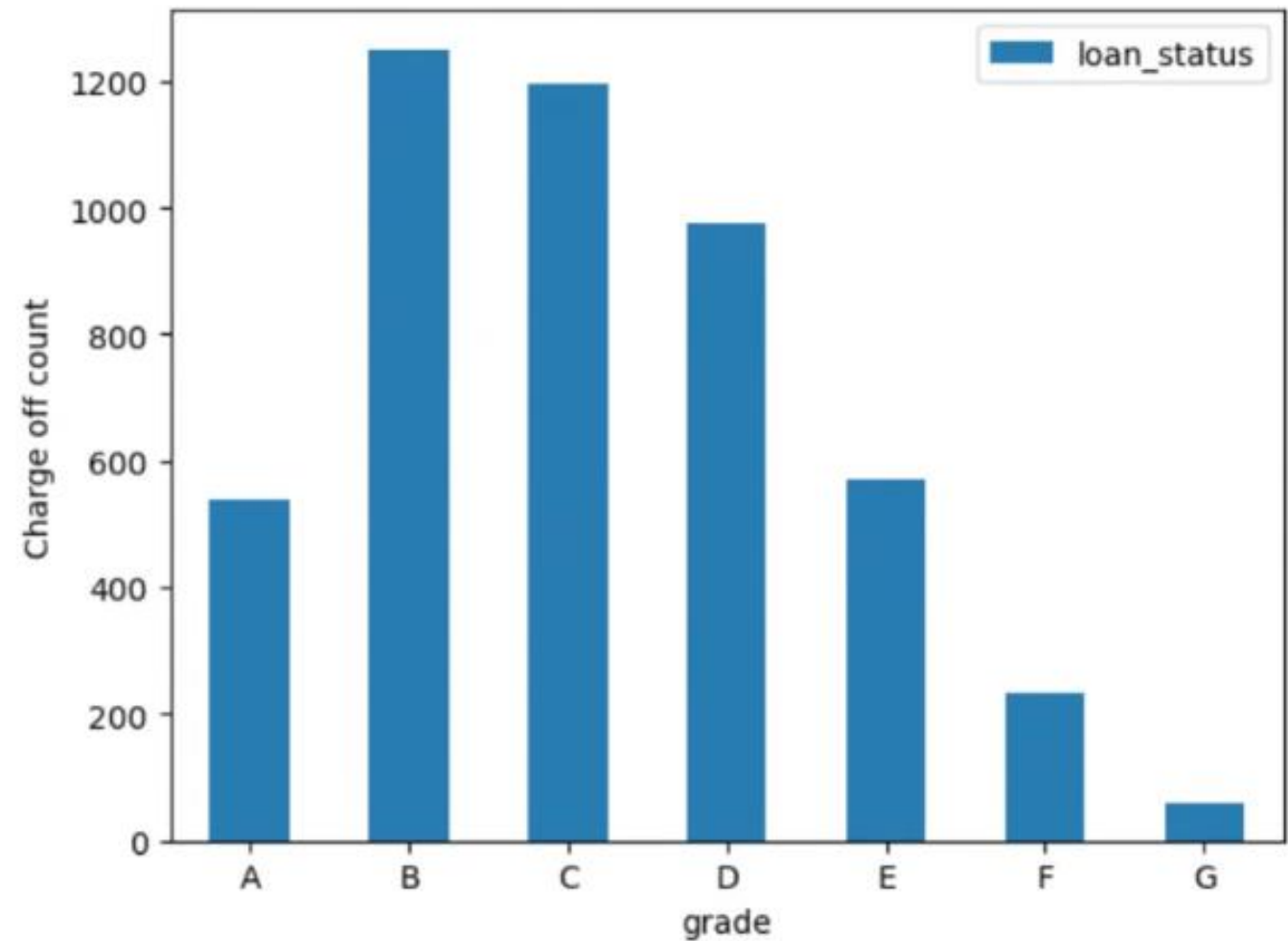
People who took loans on higher rate of interest defaulted the most.



Charged Off loans based on Grade

Notes:

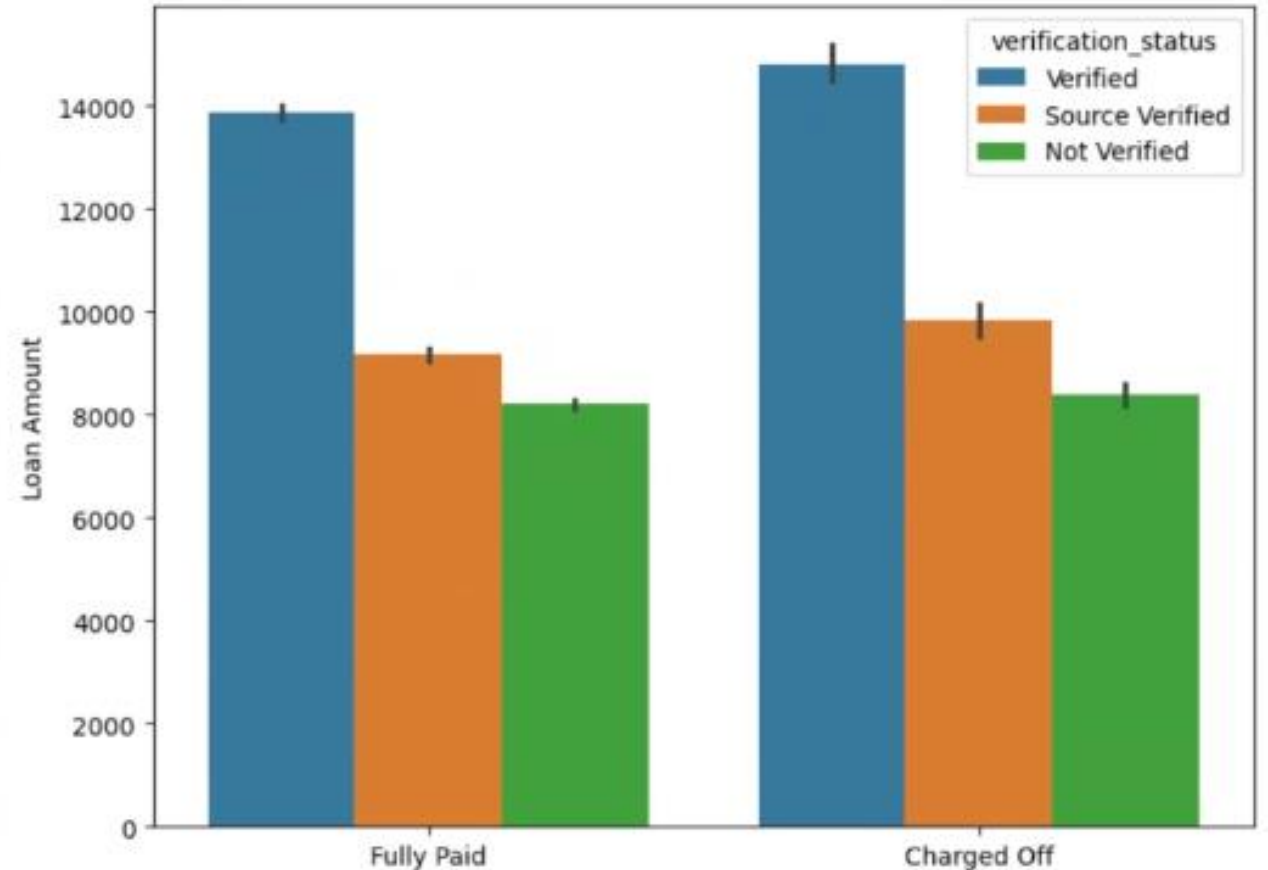
- LC assigned loan grades B & C have higher defaulters than others



Segmented Analysis – Loan amount based on verification status

Notes:

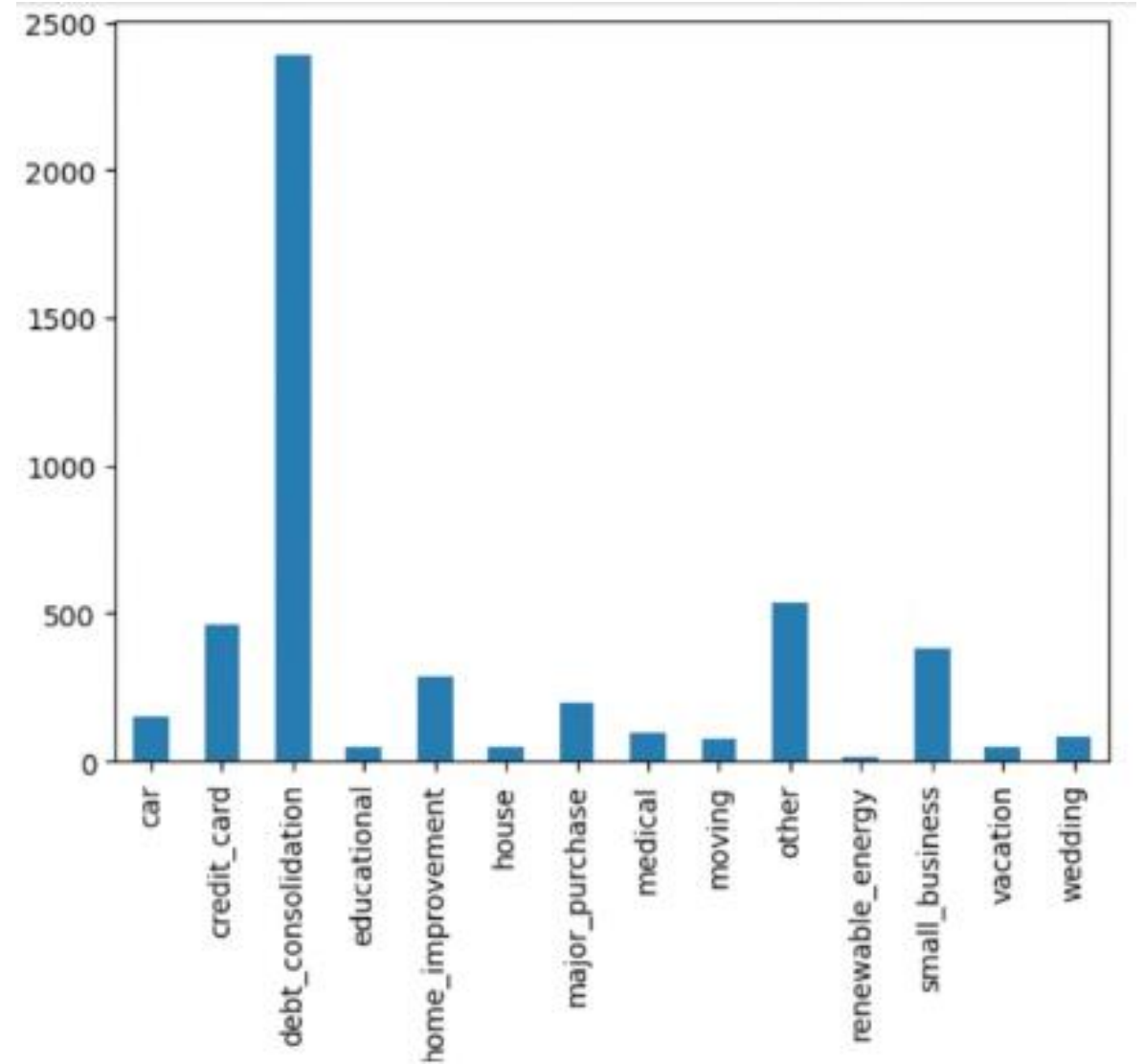
- Verification status does not have a significance on defaulters because the Charged Off loans were equally or more verified than Fully Paid loans



Purpose of loan for Charged –off loans

Notes:

- Debt consolidation, credit cards and small business have higher chances of defaulting



Observations:

- People with more than 80K annual income have lesser chance of defaulting the loan
- Max loans are in the range of 5000 and 15000. Larger loan amounts are relatively rare but could represent a higher risk for default
- Most of the loans are given in the Grade-B category as well as highest defaulters are in B. Comparatively A category has better ratio of paid to charged-off
- Higher the rate of interest , more are the chances of defaulting
- People in rental accommodation tend to apply for more loans and they also tend to be the defaulters
- Consider issuing loans on longer terms for complete repayment of loans.
- Interesting aspect in Verification status is that it does not have a major significance on defaulters because the Charged Off loans were equally or more verified than Fully Paid loans.
- Beware of states like CA, NY, FL, TX, NJ, while approving loans.
- Beware of loan applications with purpose of Debt consolidation, credit cards, small business as they have higher chances of defaulting.