

Gramener Case Study



Submitted By:
Mahesh Tyagi , Gaurav Jain, Navneet Singh & Sanjay Kakkar
PGD-Data Science
Cohort -Mar'2018

Problem Statement & Analysis Approach

Problem Statement:

A consumer finance company which specializes in lending various types of loans to urban customers faces two types of risks while deciding on approving or denying a loan.

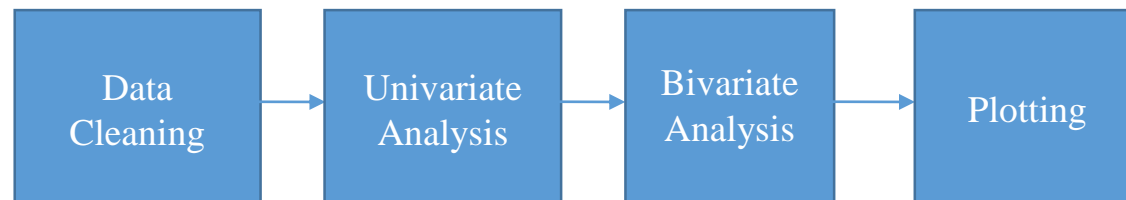
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Analysis Approach:

Identify customers who are likely to default by using.

- Univariate Analysis and Bivariate Analysis.
- Plots to summarize the results.

Analysis flowchart:



Note: This presentation only includes findings that impact default percentage,.

Details - Sample Data

- ❑ Data of period Jun'2007 – Dec'2011
- ❑ Total loan records of 39,717 applicants and their loan status
- ❑ Information of loan is spread among 111 attributes in dataset

Data Cleaning

- ❑ No duplicate records found
- ❑ 54 attributes have been removed from dataset on account of containing NA only
- ❑ 9 attributes have been removed containing observation as Zero value/only Zero+NA / same value in all the records
- ❑ Date attributes are formatted with date types
- ❑ Standard precision formatted for all numeric attributes
- ❑ Derived attributes like loan burden % has been added($\text{installment}/\text{income}$)

Univariate Analysis findings

Count of Loan Status

Charged Off	Current	Fully Paid
5627	1140	32950

Summary of funded amount

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
500	5400	9600	10950	15000	35000

Summary of funded amount committed by Investor

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	5000	8975	10400	14400	35000

Summary of Interest rates

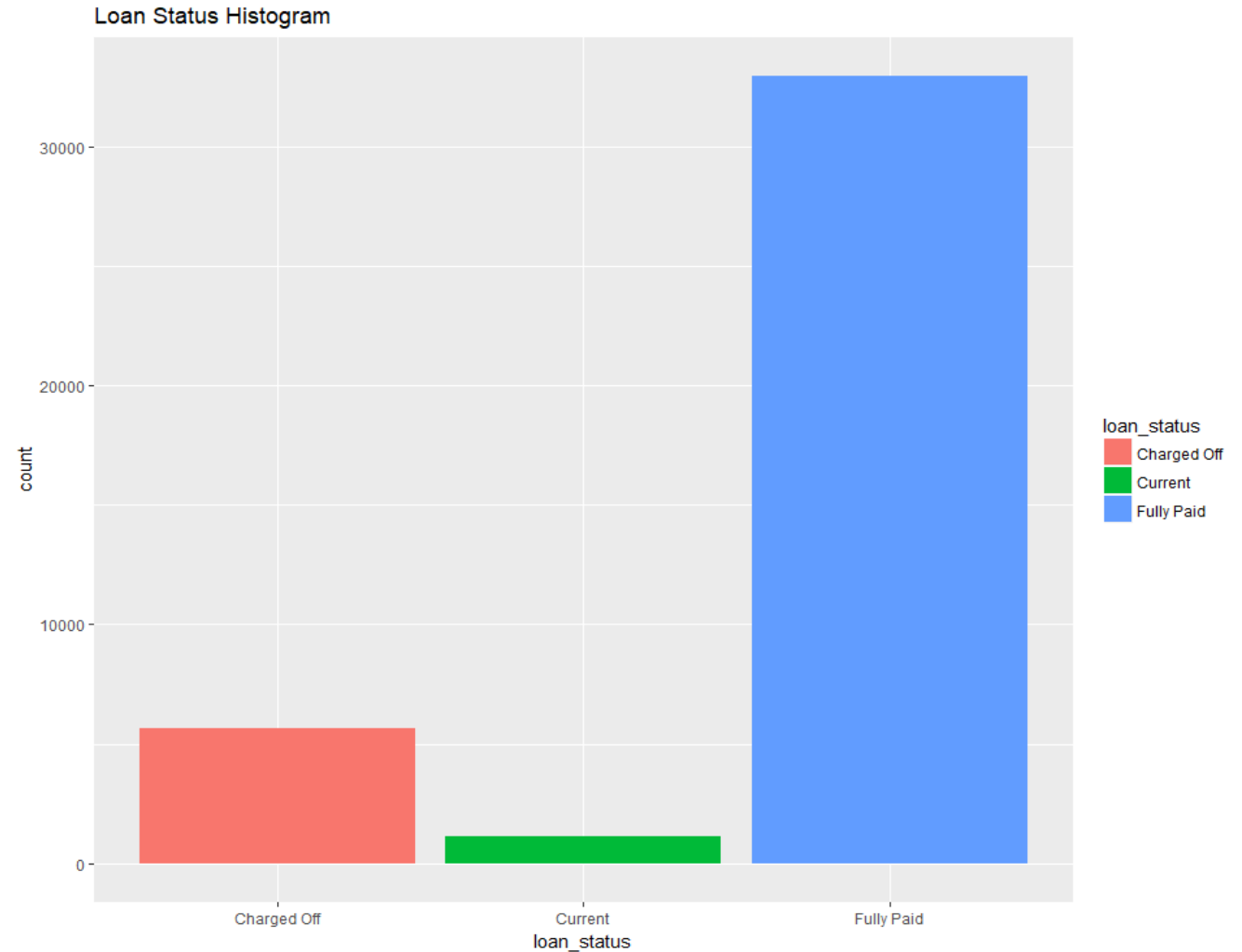
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.42	9.25	11.86	12.02	14.59	24.59

Count of verification status

Not Verified	Source Verified	Verified
16921	9987	12809

Summary of assigned loan grades

A	B	C	D	E	F	G
10085	12020	8098	5307	2842	1049	316



Bivariate Analysis– Verification Status

Verification status against Loan Status:

Analysis

Verification status was analyzed against loan status to understand which type of verification status is more likely to default

The Graph indicates

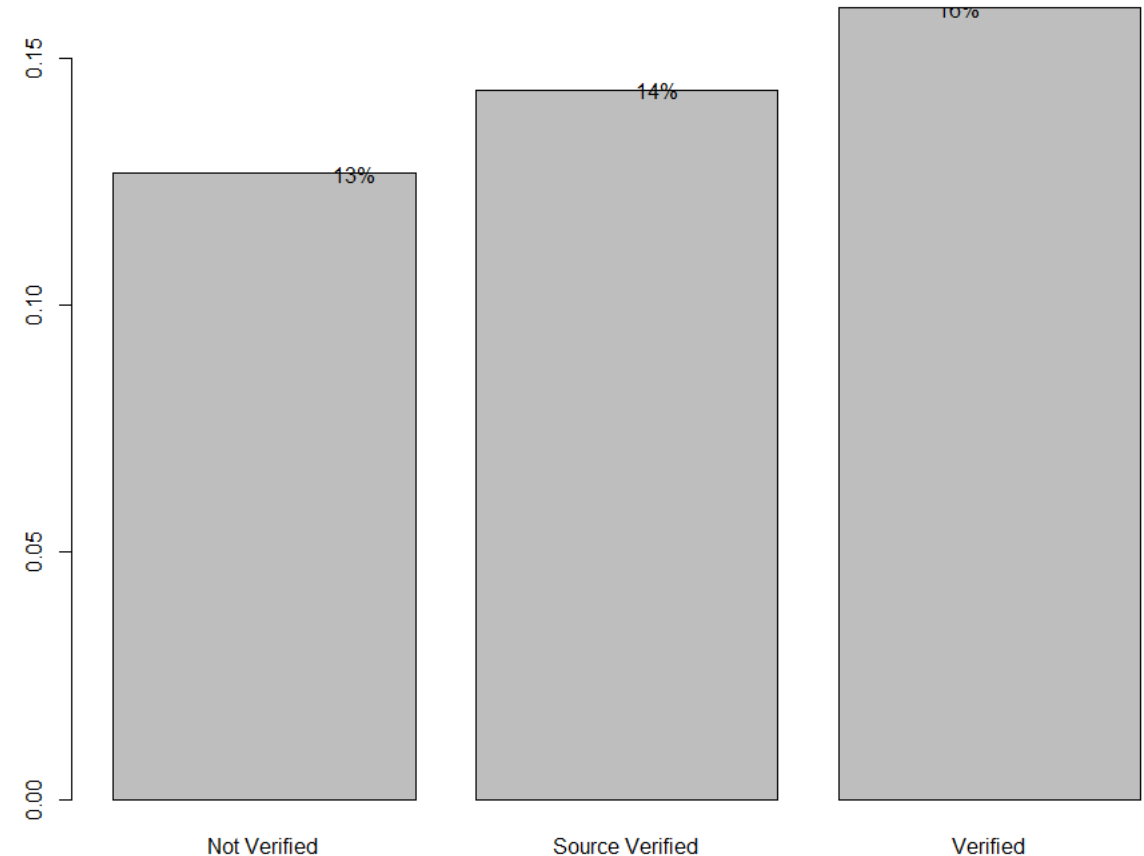
- 13% of not verified customers defaulted.
- 14% of source verified customer defaulted.
- 16% of verified customer defaulted.

Conclusion

The analysis concludes that Verified customers are more likely to default as compared to “Not Verified” customers.

Business Sense

Verified customers should have been less likely to default as compared to “Not Verified” customers.



Bivariate Analysis– Bankruptcy record

Bankruptcy record against Loan Status:

Analysis

Bankruptcy record was analyzed against loan status to understand its impact on “default” percentage.

The Graph indicates

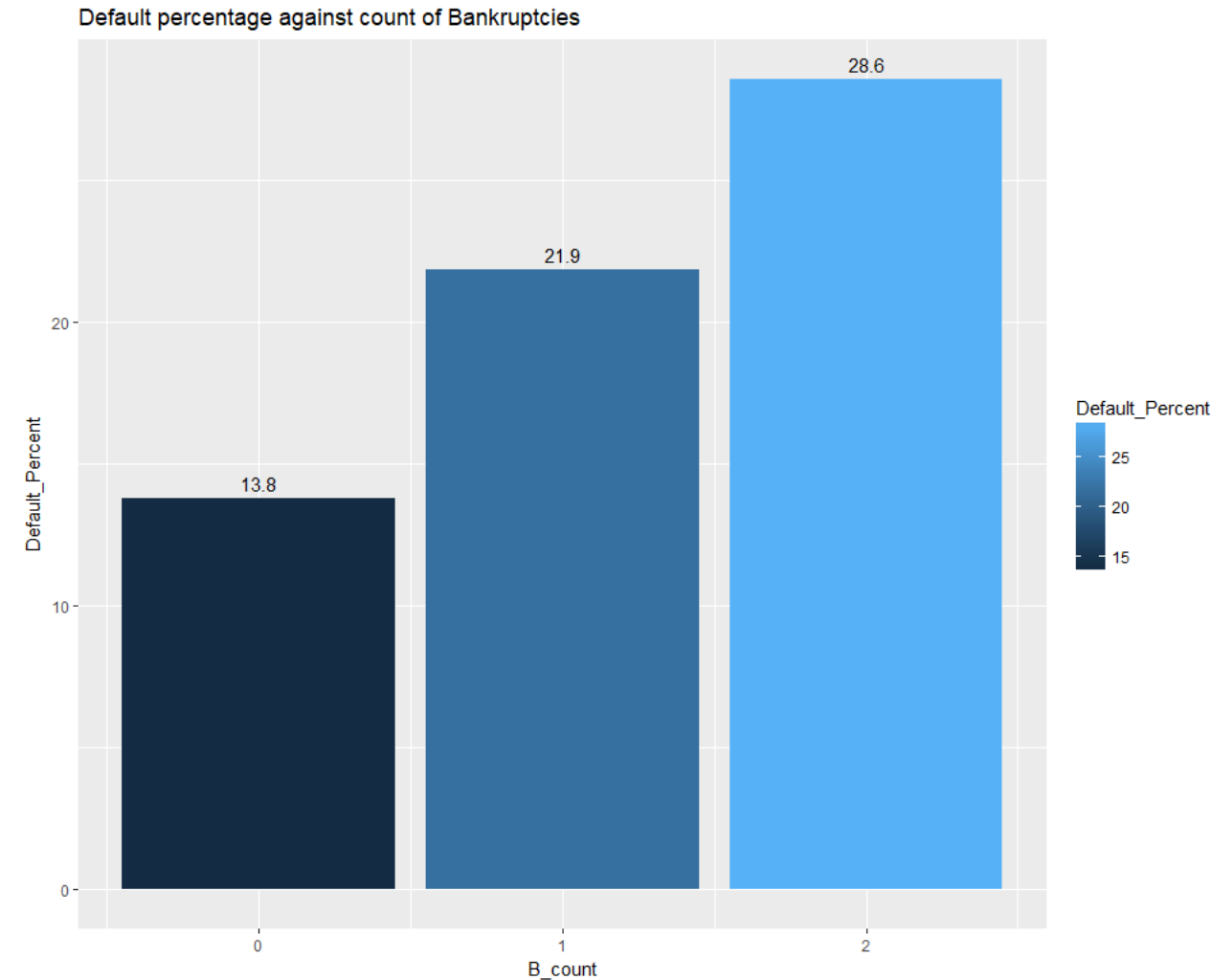
- 13.8% of customers having no bankruptcy record defaulted.
- 21.9% customers having one bankruptcy record defaulted.
- 28.6% customers having two bankruptcy records defaulted.

Conclusion

- Customers having bankruptcy record are more likely to default
- Higher the count of bankruptcy, higher the chances of default.

Recommendation

Customer’s with bankruptcy record should be avoided while lending or should be given loan against security or with higher interest rates along with additional recovery measures.



Bivariate Analysis– Loan Burden

Loan Burden against Loan Status:

Analysis

Loan Burden calculated as: $\text{EMI}/(\text{Annual income}/12)$ was analyzed against loan status to understand its impact on “default” percentage.

The Graph indicates

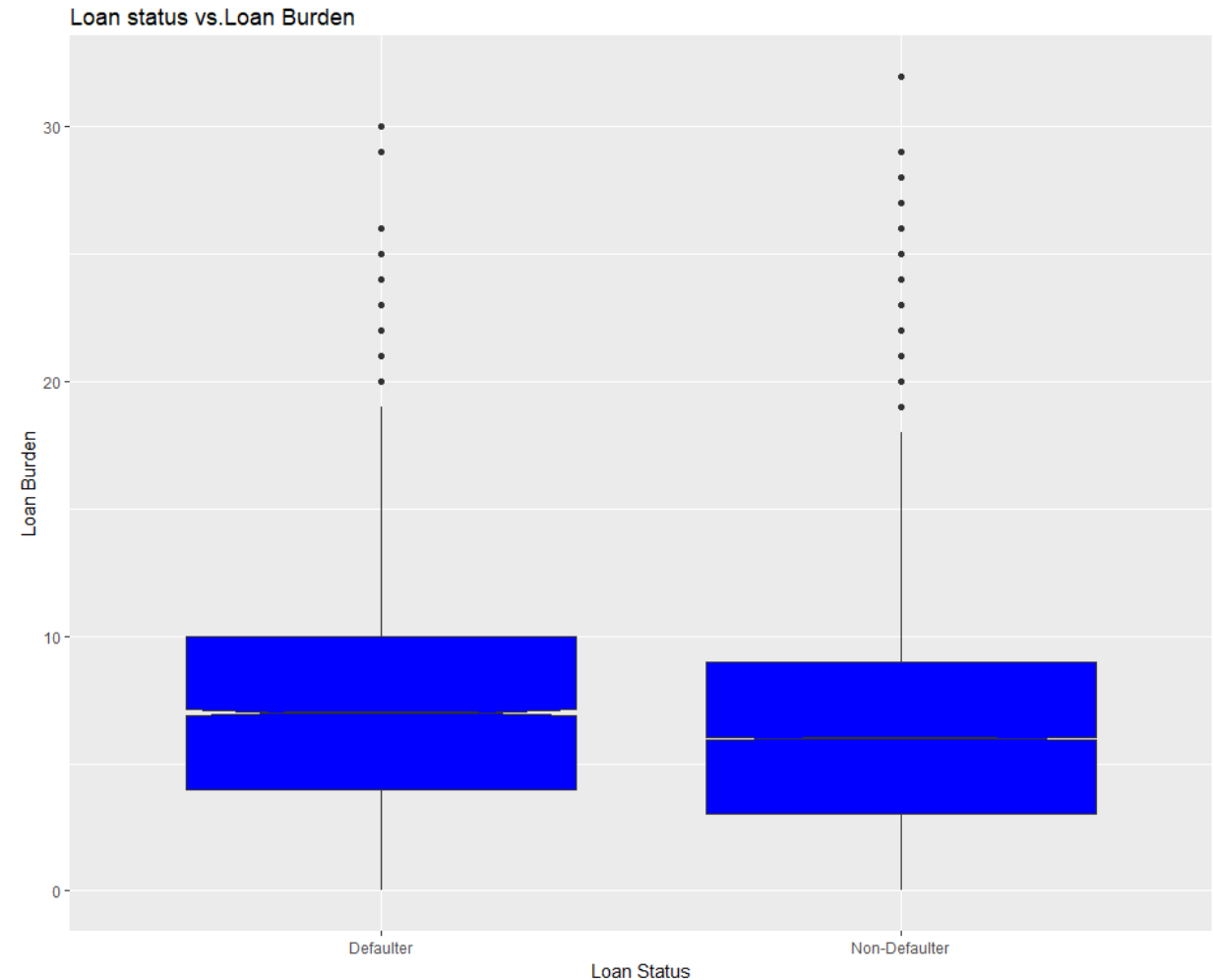
- Charged of customer have a higher median of loan burden as compared to fully paid.

Conclusion

- Customers are likely to default if their EMI takes out a huge chunk from their income.

Recommendation

EMI should be carefully finalized keeping into consideration the customers income



Bivariate Analysis– Purpose

Purpose of loan against Loan Status:

Analysis

Purpose of loan was analyzed against loan status to understand which category is vulnerable to “default”.

The Graph indicates

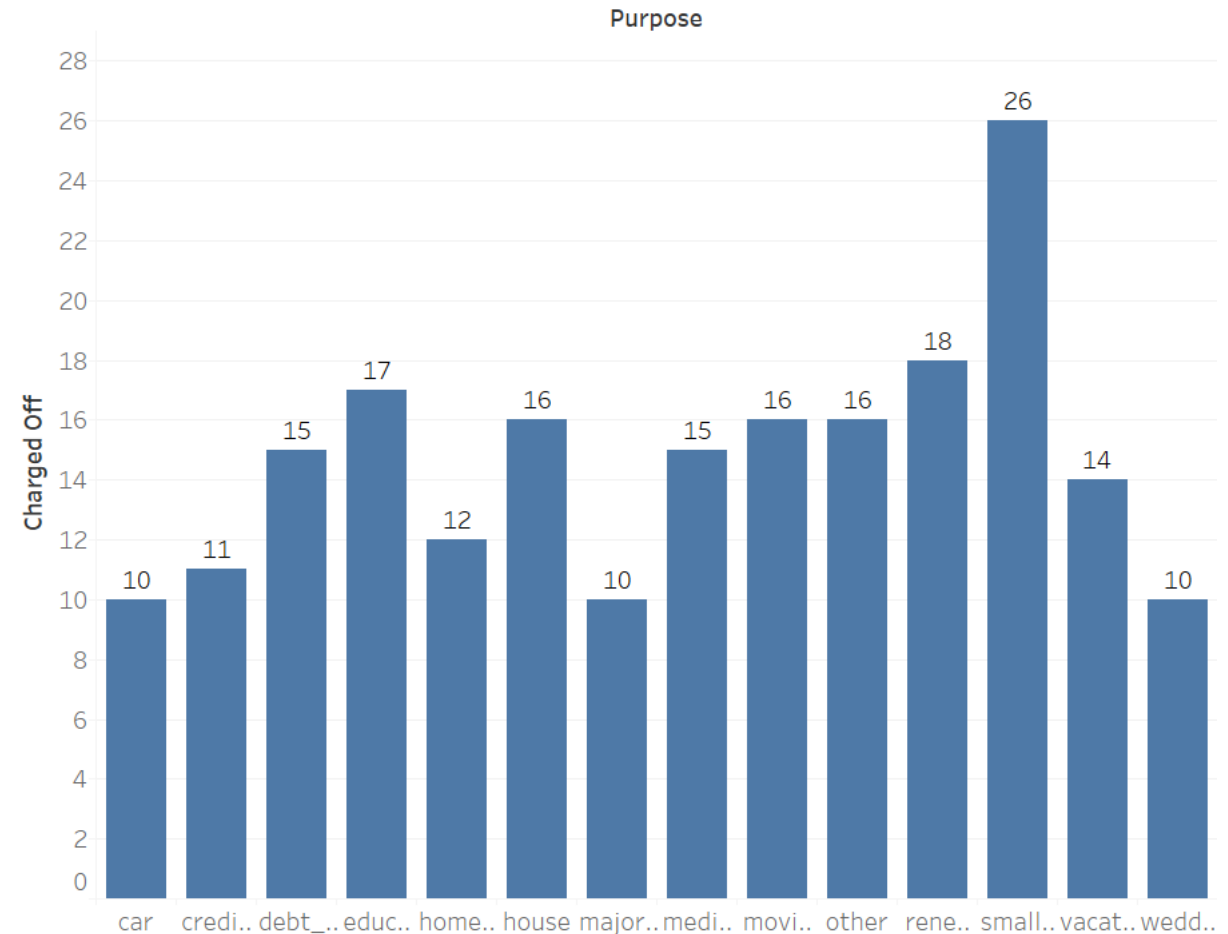
- Small_Business category had the highest percentage of defaults (26%).
- Categories, “Car”, “major_purchase” and “wedding” are the lowest contributors to default percentage.

Conclusion

- Loan’s with “Small_Business” are more likely to defaults as compared to other categories.

Recommendation

Loan for Small business can perhaps be treated as commercial loan and attach a collateral against the loan.



Bivariate Analysis– Address State

Purpose of loan against Loan Status:

Analysis

Default count and default percentage was compared state wise to get an understanding of which states are contributing the most in both aspects.

The Graph indicates

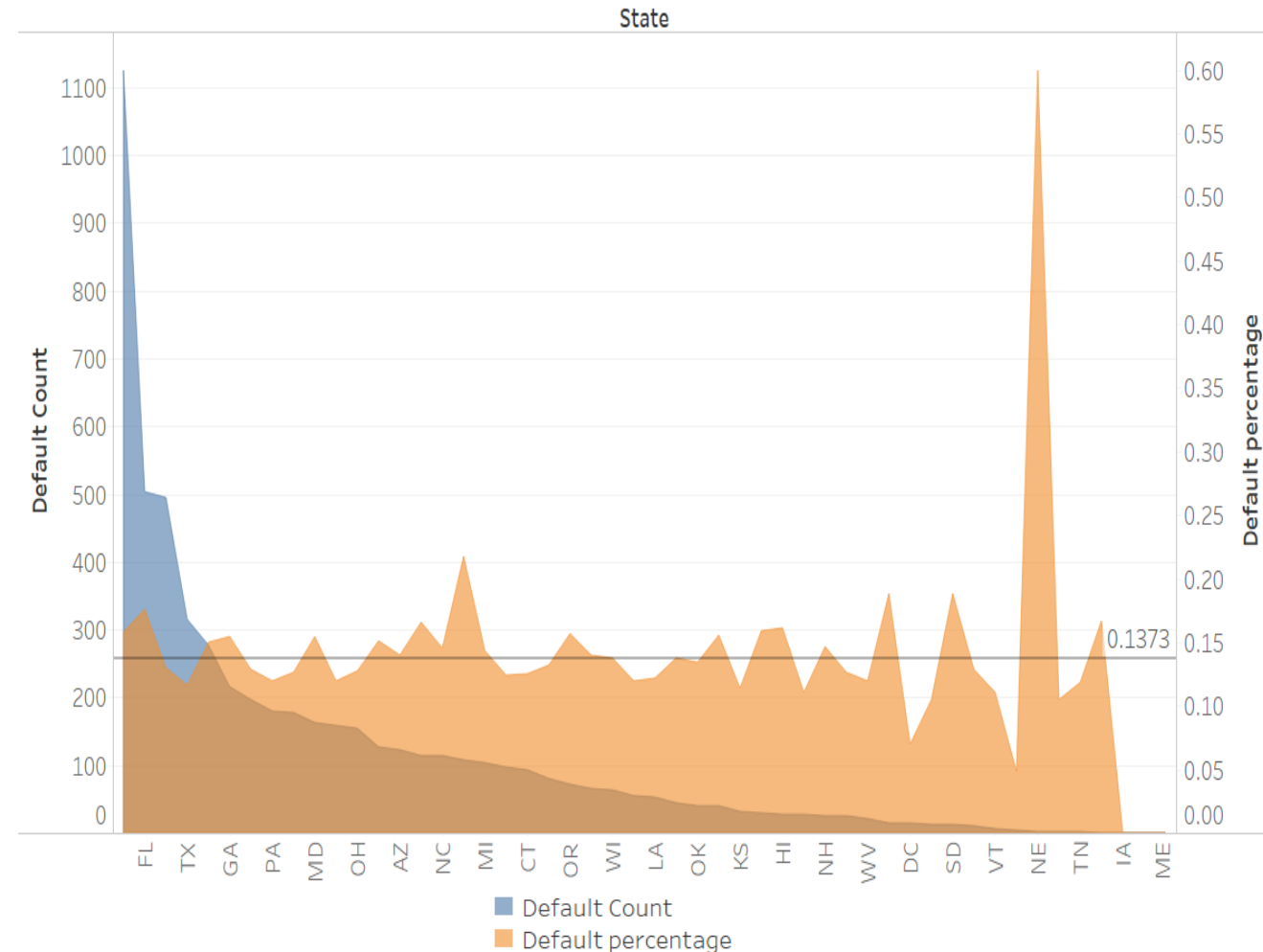
- While there are states with very default percentage but almost negligible count.
- A few states are relatively high on default count and above average default percentage with CA, FL and NJ being top three.

Conclusion

- Loan's approved in states with average default percentage greater than 13.7% are more likely to default compared to the rest.

Recommendation

While States like CA and FL are high on default percentage, they are also high business states hence instead of taking strict measure, exercising some caution to bring down default percentage should be the key takeaway.



Thanks