A wooden pencil with a dark eraser tip is positioned diagonally across the frame, resting on a piece of paper. The paper features a line graph with a grid. The graph has a vertical axis with labels '100' and '50', and a horizontal axis with labels '93' and '98'. The pencil is pointing towards the bottom right of the frame, with its tip near the '93' label on the horizontal axis. The background is a blurred, light-colored surface.

LENDING CLUB CASE STUDY: EDA ASSIGNMENT

BY –

GAURAV SINGH GAUR

TABLE OF CONTENTS

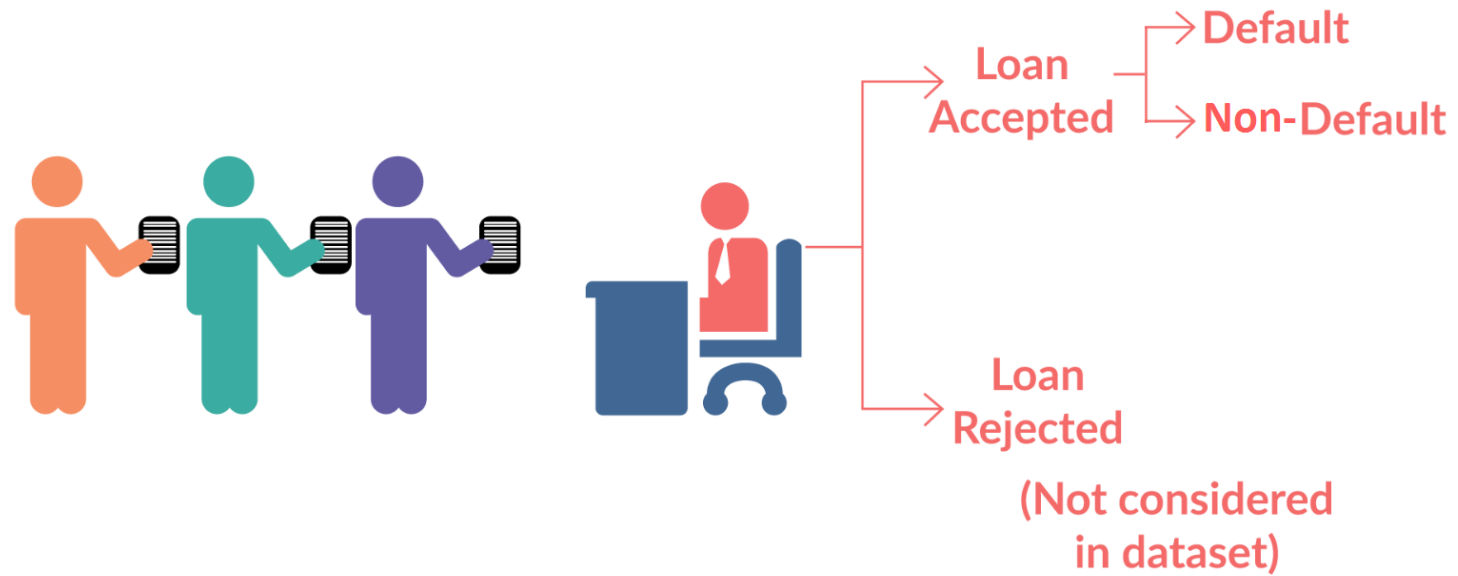
1. Objective
2. What is EDA?
3. EDA Steps & Analysis
4. Conclusion

OBJECTIVE: LENDING CLUB CASE STUDY

- Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.
- The main objective is to be able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- Identification of such applicants using EDA is the aim of this case study. Perform an analysis 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.

OBJECTIVE: LENDING CLUB CASE STUDY

LOAN DATASET



CASE STUDY OBJECTIVE

Identification of Loan Applicant traits that tend to 'default' paying back

Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

Loan Lending Organizations may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants.

WHAT IS EDA?

- EDA(Exploratory Data Analysis) is a data exploration technique to understand the various aspects of the data.

EDA STEPS

- Data Sourcing & Understand the data
- Data Cleaning : Filtering data from null values and redundancies
- Univariate Analysis
- Bivariate & Multivariate Analysis
- Conclusion

DATA SOURCING

- 1. **'loan.csv'** : - It contains the complete loan data for all loans issued through the time period 2007 to 2011.
- 2. **'Data_Dictionary.csv'** :- describes the meaning of these variables.

UNDERSTANDING THE DATA

- To understand the data I carried out below activities: -
 1. Went through Column descriptor
 2. Identified the shape of the data set
 3. Checked data head & tail
 4. Looked for unique values per column
 5. Checked Dataset description

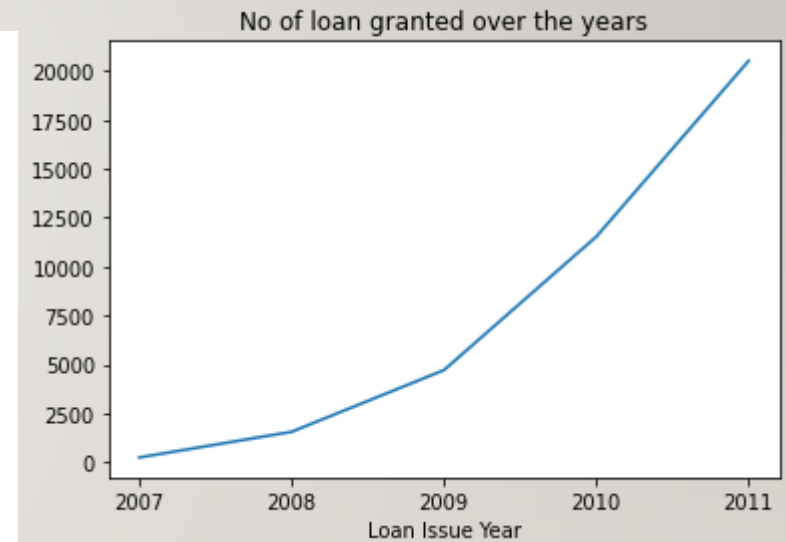
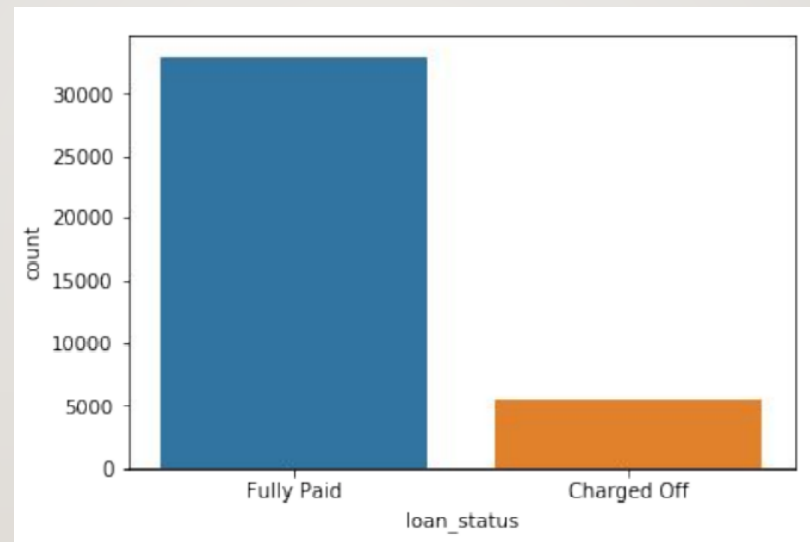
DATA CLEANING & FILTERING

- To Clean & filter the data I carried out below steps: -
 1. Find the columns having more than 30% null values and remove such columns.
 2. Removed Rows having null values equals to or greater than 30%.
 3. Identified & Removed Columns where were either Redundant or not required in the conclusive analysis.
 4. Removed outliers.
 5. Treat missing values with appropriate approach

UNIVARIATE, BIVARIATE AND MULTIVARIATE ANALYSIS

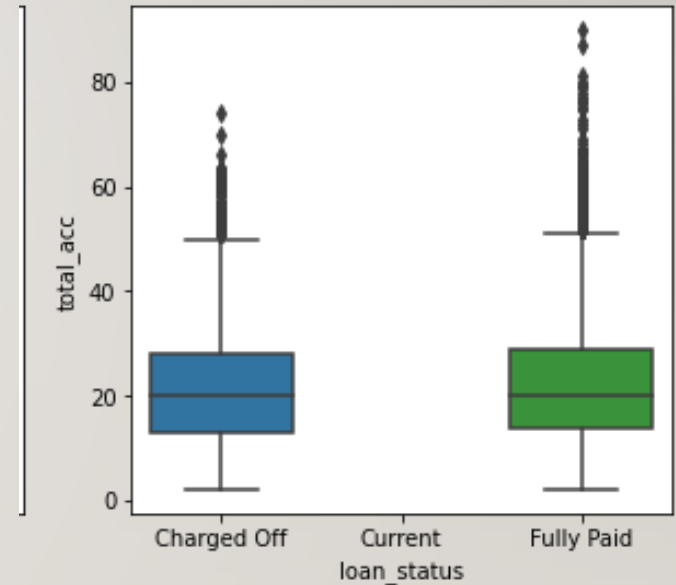
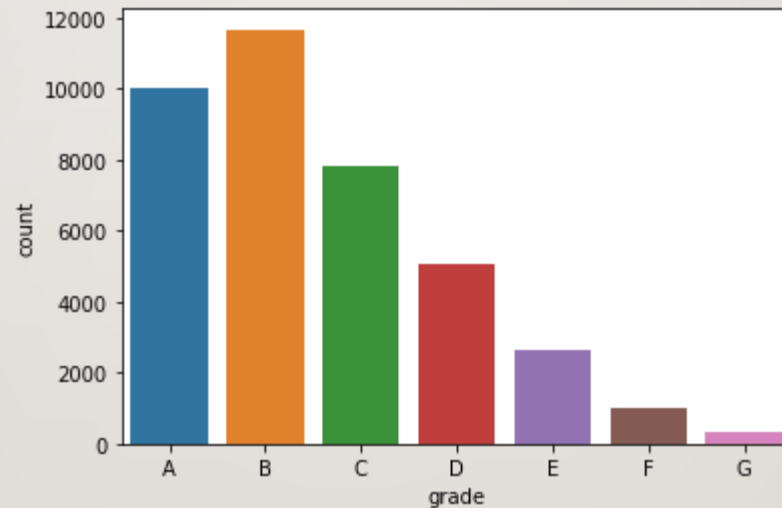
LOAN STATUS

- **Analysis : -**
- 1- Very less defaulters, only about 15%
- 2- Demand for loan has been increasing

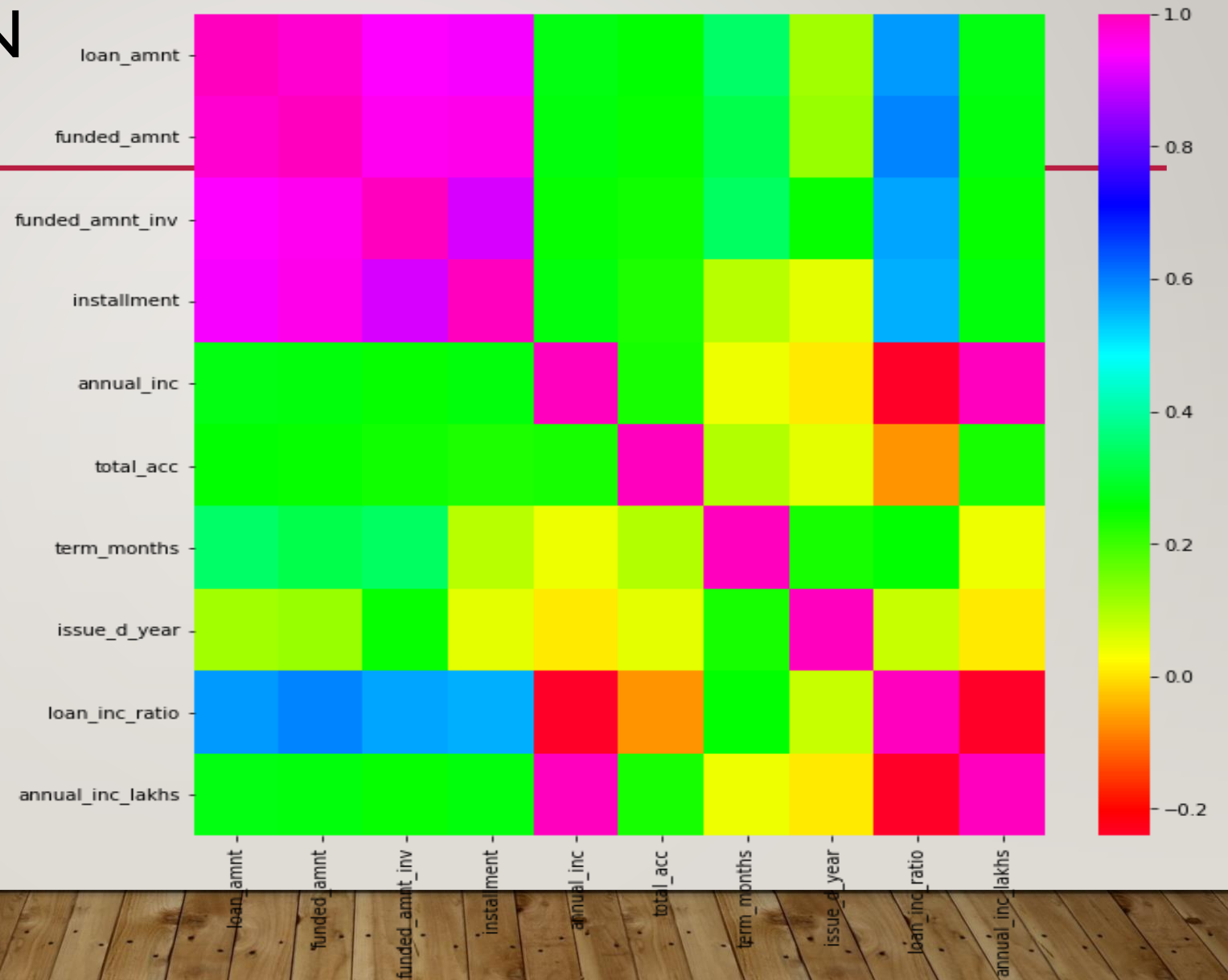


LOAN GRADE

- **Analysis :** -
- 1- Majority Loans in A & B Category
- 2- Majority loan are in between 5000-15000.
- 3. Higher interest rate has higher tendency to charge-off

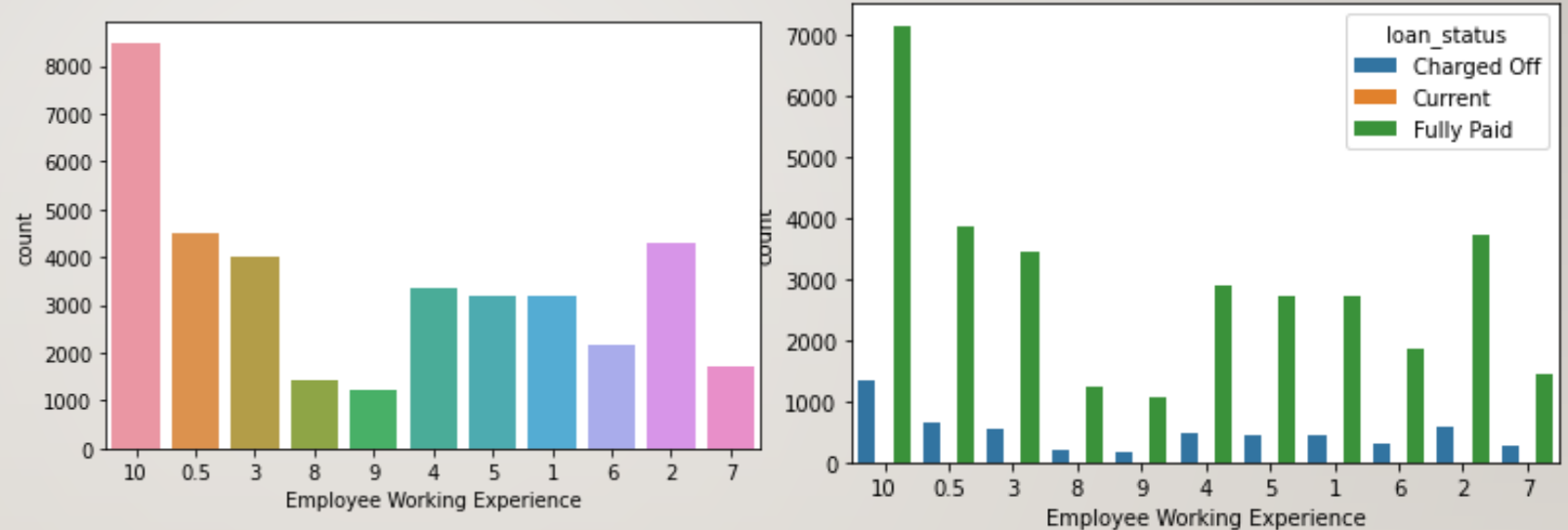


CORRELATION



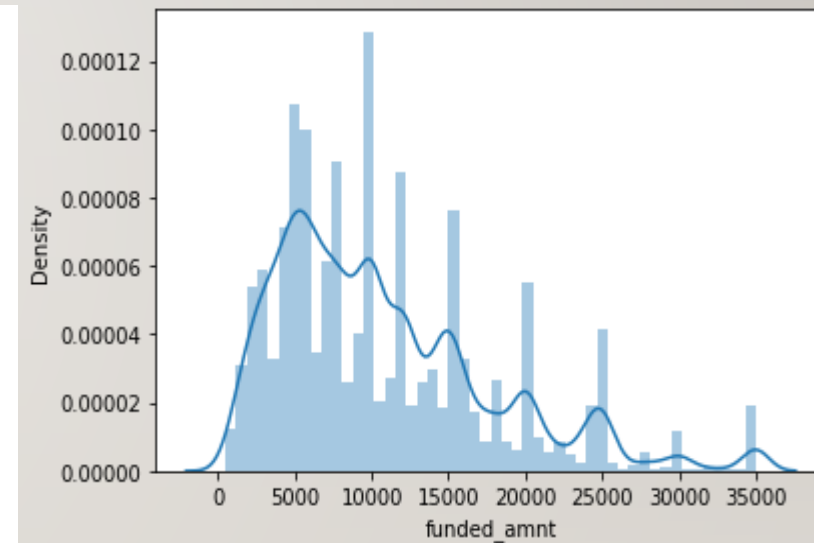
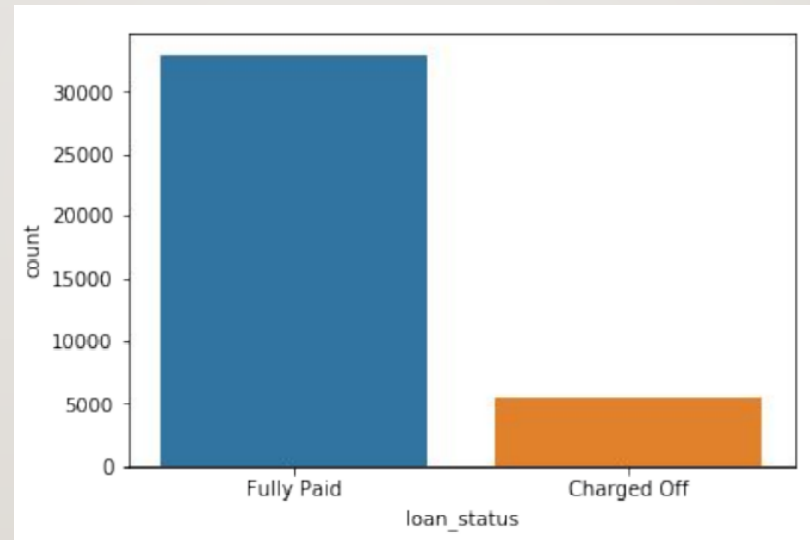
LOAN APPLICANT V/S WORK EXPERIENCE.

- **Analysis : -**
- 1- Majority has more than 10 yrs of exp.
- 2- Loan default rate is high in 2-3yrs of exp.



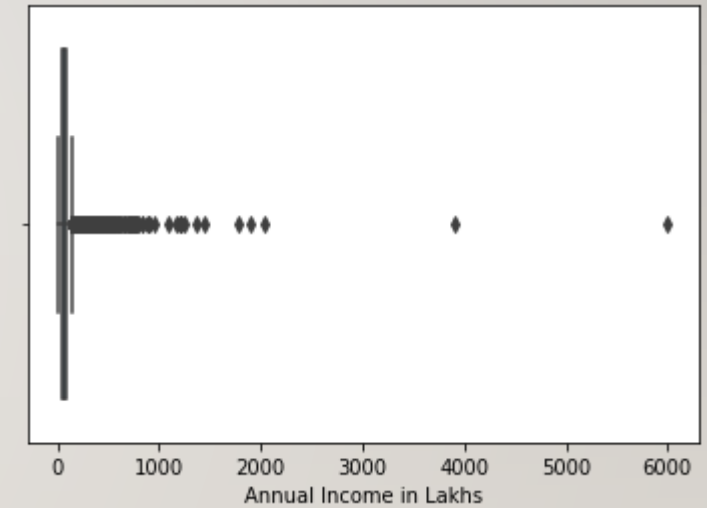
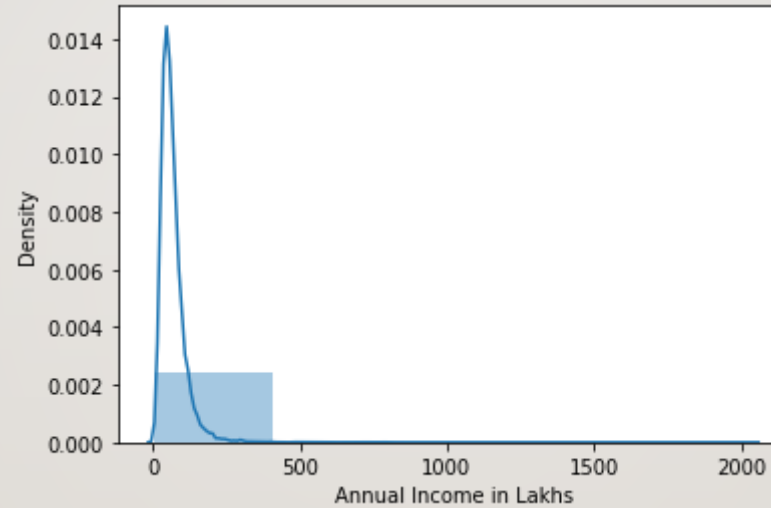
LOAN AMOUNT

- **Analysis : -**
- I - Most of the loan is in-between 5000-15000



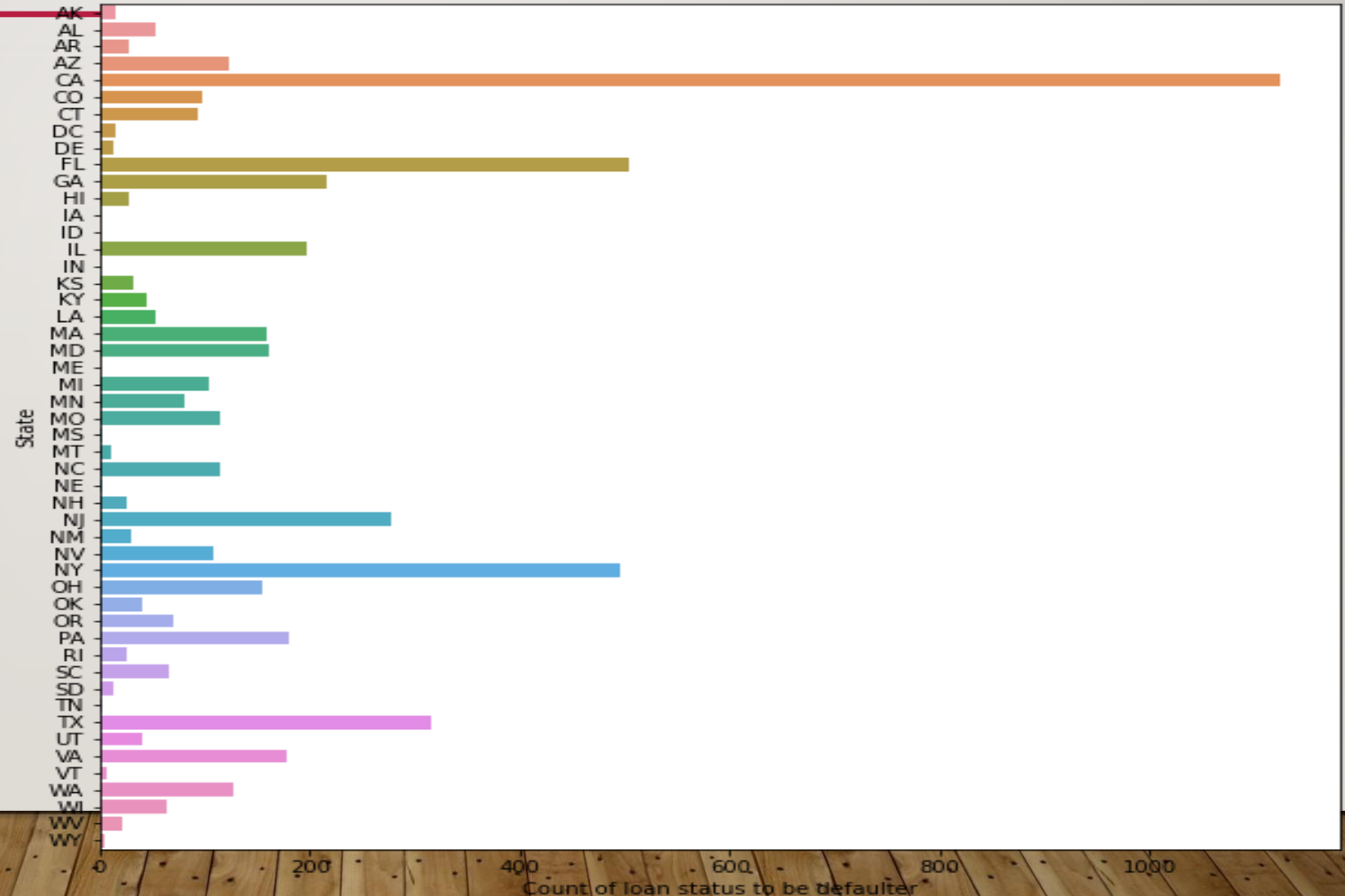
INCOME OF LOAN APPLICANTS

- **Analysis : -**
- 1- Majority income is between 0-4 lakhs
- 2- Only two outliers here.



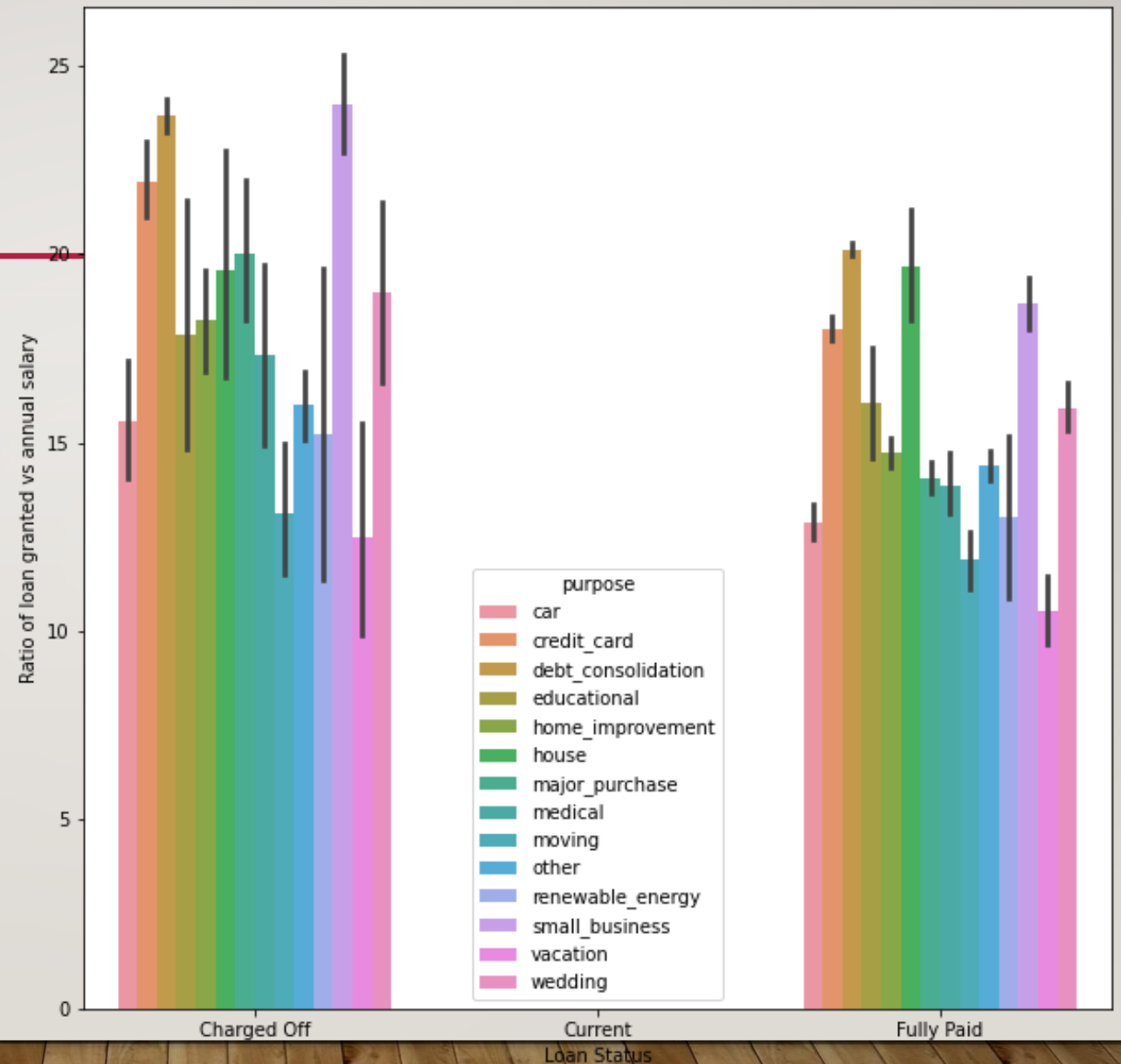
LOAN DEFAULTER BY STATE

- **Analysis : -**
- I- CA state has most defaulters

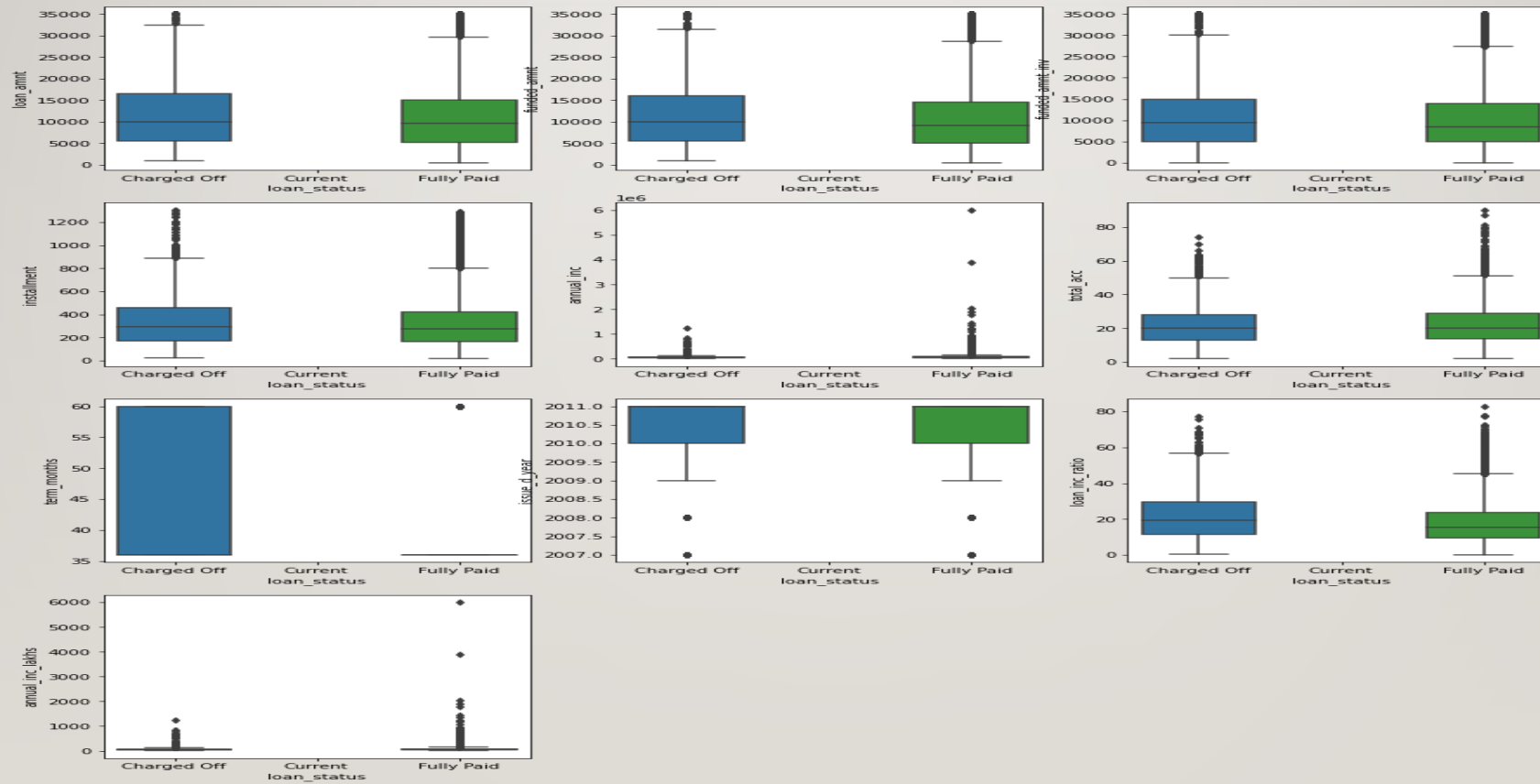


LOAN PURPOSE

- **Analysis:-**
- Most loans were given for car and small_business



MORE ANALYSIS



CONCLUSION

1. Loans having higher interest rate have more defaulters.
2. Most defaulter applicants belonging to CA state, as tendency to default is high.
3. MT,IN states have very less or say negligible defaulters.
4. Low grade loans have high tendency to default
5. When the purpose is debt consolidation check applicant thoroughly as it has high tendency to default.

THANK YOU