Exploratory Data Analysis (EDA) for bank loan defaulters

Business Scenario:

1. Why this data was collected?
2. To understand the background of bank members applying for loan.
3. To know the maximum and minimum ask of loan and funded amounts.
4. How much interest is getting collected? Revenue generated.
5. To understand if there are any defaulters and if there is any pattern in that.
6. Use Case:
7. Identify the pattern of defaulter’s behaviour.
8. Deciding whether to fund a particular bank member or not.
9. Business Problem:

The bank or investing or money lending organisation provides loans to many people by doing certain background checks. Still there are so many fraud and money laundering cases where the banks go bankrupt. When a bank provides loan for a period, it charges interest over the amount. If the people are paying their loans back in the given time frame, the banks will be in some profit. But if the people default, the losses for the bank will be far bigger than the profits with no defaulters. Hence it is very much important to map the behaviour of defaulters, predict who will be most likely to default and avoid giving loan to them.

1. Data

Data Dictionary:



1. What kind of analysis you perform?
2. Univariate, bivariate
3. EDA
4. Use the tools of your trade
5. Python
6. R
7. Analysis points:
8. Distribution of Numerical attributes (distplot and histogram)
9. Loan amount,
10. Interest rate,
11. instalment,
12. annual income,
13. dti,
14. revo\_bal,
15. Unique and counts of Categorical attributes ( countplot, pie-chart)
16. Term:
17. Employement length
18. Home Ownership
19. Verification status
20. Purpose/title
21. Relation between two numerical values
22. 6C2
23. Correlation Matrix is between numerical values
24. Relation between Numerical and categorical values
25. All Numerical attributes wrt loan status
26. All Numerical attributes wrt