

# Lending Club Case Study by Anandhakumar Dharmalingam

# Problem Statement:

- Using the concepts of EDA to decipher which types of customers default on a loan
- Need to understand how consumer attributes and loan attributes influence the tendency of default.

# Meta Data Description:

- This loan.csv data set is all about the loan approval process and the borrowers repayment follow up is observed.

# Finding the consumer attributes & loan attributes in the LOAN data set:

- **consumer attributes :**
- id, member id, grade, sub\_grade, emp\_title, emp\_length, home\_ownership, annul\_inc, desc, purpose, title, zip\_code, addr\_state,
- **loan attributes**
- loan\_amnt, funded\_amnt, funded\_amnt\_inv, term, int\_rate, verification\_status, Issue\_d, loan\_status, pym, dti, delinq\_2ys, earliest\_cr\_line, inq\_last\_6mths, mths\_since\_last\_delinq, mths\_since\_last\_record, open\_acc, pub\_rec, revol\_bal, revol\_util, total\_acc, out\_prncp, out\_prncp\_inv, total\_pymnt, total\_pymnt\_inv, total\_rec\_prncp, total\_rec\_int, total\_rec\_late\_fee, recoveries, collection\_recovery\_fee, last\_pymnt\_d, last\_pymnt\_amnt, next\_pymnt\_d, last\_credit\_pull\_d

# Classifying the required variable of Loan Dataset for our analysis:

- ***Categorical Variable:***
  - **Ordered**
    - Term, int\_rate, grade, sub\_grade, emp\_length, verification\_status, issue\_d, loan\_status, delinq\_2yrs, earliest\_cr\_line, inq\_last\_6mths, mths\_since\_last\_delinq, mths\_since\_last\_record, revol\_util, pub\_rec, last\_pymnt\_d, next\_pymnt\_d, last\_credit\_pull\_d
  - **Unordered**
    - Emp\_title, home\_ownership, desc, purpose, title, zip\_code, addr\_state,
- ***Quantitative/Numeric variable:***
  - loan\_amnt, funded\_amnt, funded\_amnt\_inv, instalment, annual\_inc, dti, open\_acc, revol\_bal, total\_acc, out\_prncp, out\_prncp\_inv, total\_pymnt, total\_pymnt\_inv, total\_rec\_prncp, total\_rec\_int, total\_rec\_late\_fee, recoveries, collection\_recovery\_fee, last\_pymnt\_amnt

# Final Findings

- After analysing all the related features available in the dataset, we have come to an end, deducing the main *driving features* for the **Lending Club Loan Default** analysis:
- *The best driving features for the Loan default analysis are:* **term, grade, purpose, revol\_util, int\_rate, installment, annual\_inc, funded\_amnt\_inv**



