# Lending Club: EDA Case Study

By:

Debasish Mondal
Sharath Chandra

# Problem Statement

The primary objective of this case study is to identify the key driver variables that influence loan default, enabling the consumer finance company to make informed decisions on loan approvals and risk assessment. By conducting an Exploratory Data Analysis (EDA), we aim to discover patterns and relationships in the data that can help the company minimize credit loss, manage a healthy loan portfolio, and optimize their lending strategies.

Specifically, the case study will focus on:

1. Understanding how consumer attributes and loan attributes influence the likelihood of default.
2. Identifying the variables that are strong indicators of loan default.
3. Providing actionable insights and recommendations for the company's portfolio and risk assessment processes.

# Solution approach summary

- **Data Cleaning:**
  - Import the dataset and perform preliminary data cleaning, including handling missing values, dropping irrelevant columns, and converting data types as needed.
- **Univariate Analysis:**
  - Analyze the distribution of each variable in the dataset using appropriate visualization techniques.
- **Bivariate Analysis:**
  - Examine relationships between pairs of variables to understand their associations with loan default.
- **Feature Engineering:**
  - Create new variables or transform existing ones to better represent the information in the dataset.

# Solution approach summary

- **Multivariate Analysis:**
  - Investigate the relationships among multiple variables simultaneously using advanced techniques such as PCA, clustering, or regression analysis.
- **Identifying Key Driver Variables:**
  - Determine the variables that have the strongest relationship with loan default using correlation coefficients, statistical tests, or machine learning algorithms.
- **Interpretation and Recommendations:**
  - Summarize the findings from the analysis and provide actionable insights to help the company make informed decisions on loan approvals and risk assessment.

# Data summary

- Data contains 39717 rows and 111 columns where each row contains data related to user's each loan request
- Following are the important variables which we selected for our analysis and ignored others which may not provide info to the analysis
  Loan_amnt, term,