

## 1. Problem statement

My aim is to build a machine learning model that can evaluate and learn from previous loans to help recommend the best loans for individuals to invest in.

## 2. Context

The finance sector is focused around one essential mathematical problem – how can we assess and quantify risk? While this is usually calculated by large firms, in recent years more and more opportunities have arisen for individuals to not only buy but also sell financial products. LendingClub, a financial services company headquartered in San Francisco, California, enables borrowers to create unsecured personal loans between \$1,000 and \$40,000 and investors to search and browse the loan listing on LendingClub website. This puts normal people in the same position as banks, allowing them to select loans that they want to invest in based on the information supplied about the borrower, amount of loan, loan grade, and loan purpose.

But with individuals acting as banks, they then have the same problem as banks do – how can they accurately assess the risk of giving a loan to maximize their return? With Machine Learning, I aim to help answer this question by building a model that can evaluate and learn from previous loans to help recommend the best loans for individuals to invest in.

## 3. Stakeholders

- CEO/CFO
- Data science team
- Me

## 4. Success criteria

- Understand the target variable default and its relationship with independent variables
- What loan grade and loan purpose are safe for investors?
- Find a machine learning model that can correctly predict all good loans as good loan. Looking for the best precision score.

## 5. Constraints

- New to the finance industry
- Depend on historical loan data
- Macroeconomic environment plays a major role in loan default

## 6. Scope + risks

- Predicting loan default by using Python and machine learning is outside of my expertise as I am new to data science

## 7. Data source

<https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv>