- ***Data source:*** https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv/data

Project plan:

- *Explain what you intend to study with your project.*
  - We intend to study how different features of loans lead to their default (or) the loan's grade. Via EDA, we'd like to better assess the data-set and glean important understanding that would better inform the models that we develop.

- *What is the ultimate objective?*
  - To predict loan defaults (or) loan grades via appropriate machine learning models.

- *What types of models are you considering?*
  - In the case of loan defaults, this would require classification models in which case we would consider logistic regression, trees, and random forests with different parameters. Since there are 145 columns in this dataset, we would also consider dimension reduction methods like PCA.
  - In the case of loan grade, we'd use other regression methods to predict the loan grade.

- ***Address the following questions:***

  - *Why is this project interesting?*
    - This project is interesting because it has real world implications and is likely reflective of the type of research question(s) that a financial institution would give to it's data scientists. Thus, this is a useful project to work on for students who are interested in financial services or FinTech. Also, this dataset has a robust set of debt performance metrics and other personal data like location and income. Thus, there are a lot of potential models to explore and it will be fun to see which models produce the most accurate predictions.

  - *What challenges and obstacles might you anticipate with this project?*
    - We anticipate multicollinearity within the dataset along with potential issues related to appropriate feature selection given the number of different independent variables that one can choose from.