**LOAN DEFAULT PREDICTION**

**Introduction**

In today’s financial landscape, understanding the factors influencing loan defaulting is crucial for lending institutions. This project aims to leverage data science and machine learning techniques to uncover the key determinants of loan default.

This insight can empower financial institutions to make more informed lending decisions, mitigate risks, and contribute to a more stable economic environment.

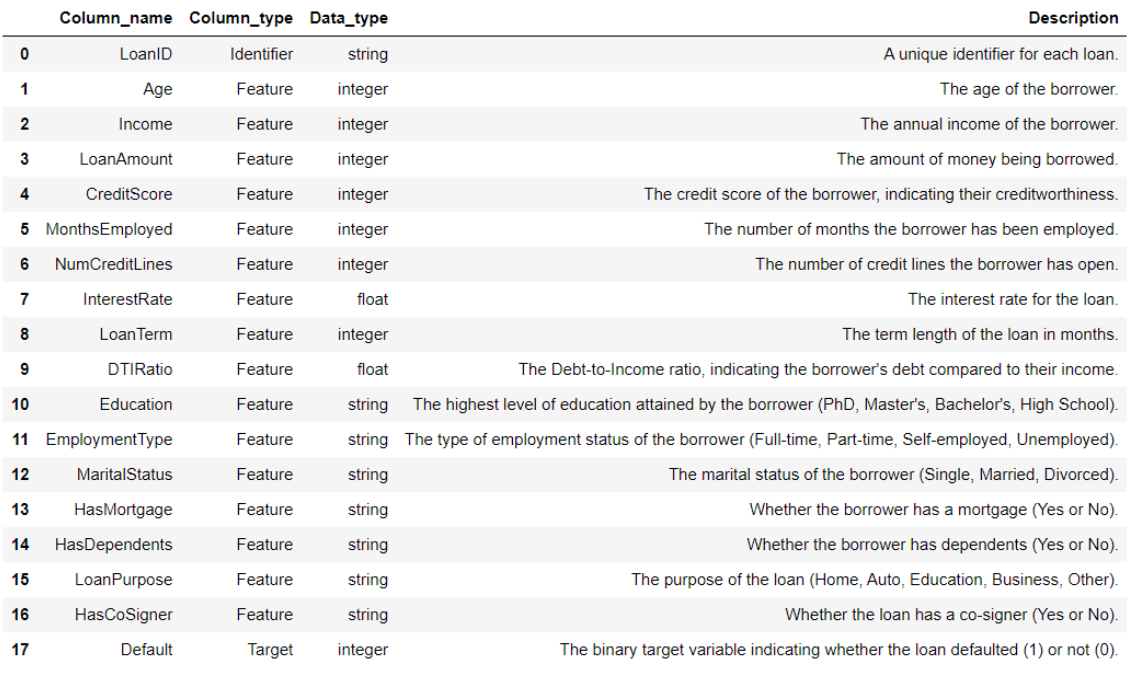
**Problem Statement**

Find out what factors tend to have an influence over loan defaulting and come up with a model that can be used by the lending institutions to predict people who can fail to repay their loan

**About the Data**

I used secondary data for this project; my dataset source is [Kaggle](https://www.kaggle.com/datasets/nikhil1e9/loan-default). I chose this project due to its rich and diverse features, which influence the chances of someone not repaying their loan.

Features of this dataset include 18 columns and approximately 255000 records.



**Research Questions**

1. **Demographics**
2. Does someone's age have a relationship with the loan default status?
3. Is there a relationship between income, loan amount, and default status?
4. Does employment type and months employed affect the default status?
5. Does someone’s education level relate to the default status?
6. Is there a relationship between marital status, dependents, and default status?
7. Is there a relationship between purpose and default?
8. **Economic Factors**
9. Is there a relationship between interest rates, loan terms, and loan amounts?
10. Is there a relationship between interest rate and default status?
11. Is there a relationship between DTI and loan default status?
12. **Stability**
13. Is there a relationship between credit score, number of credit cards, and loan status?
14. Do people who have mortgages tend to perform better than people who don’t pay loans?