1. **Ask an important question you want answered (1 at least) : 2 Points**

Answer:   
1) How can we accurately predict whether a loan applicant will default on their loan based on their demographic information, financial history, and loan characteristics?

2) What are factors influence loan default risk among borrowers, and how can lenders use this information to make more informed lending decisions?

3) How can we improve loan approval processes?

1. **Answer why this question(s) is/are important to you: 2 Points**

Various factors influence loan default risk among borrowers, including personal financial stability, employment status, credit history, and loan terms.   
By analysing such factors I can provide valuable insights to the lenders to assess the creditworthiness of applicants and tailor loan terms accordingly.

Understanding the factors that contribute to loan default risk enables individuals to take proactive measures to mitigate these risks. For example, can work on improving the credit score, managing finances more effectively, and selecting loan options that align with their financial capabilities.

1. **Find and collect data (need one dependent variable and more than 3 dependent variables): 2 Points**

Data collection: <https://www.kaggle.com/datasets/laotse/credit-risk-dataset?resource=download>

Dependent variable: loan\_status

Independent variables:

1. person\_age
2. person\_income
3. person\_emp\_length
4. loan\_grade
5. loan\_amnt
6. loan\_int\_rate
7. loan\_percent\_income
8. cb\_person\_default\_on\_file
9. Describe your data (create a data dictionary)

**Independent variables:**

1. **person\_age:** **(INTEGER)** Age of the individual applying for the loan.
2. **person\_income:** **(INTEGER)** Annual income of the individual.
3. **person\_emp\_length: (FLOAT)** Employment length of the individual in years.
4. **loan\_grade: (CHARCTER)** The grade assigned to the loan based on the creditworthiness of the borrower.  
   A: The borrower has a high creditworthiness, indicating low risk.  
   B: The borrower is relatively low-risk, but not as creditworthy as Grade A.  
   C: The borrower's creditworthiness is moderate.  
   D: The borrower is considered to have higher risk compared to previous grades.  
   E: The borrower's creditworthiness is lower, indicating a higher risk.  
   F: The borrower poses a significant credit risk.  
   G: The borrower's creditworthiness is the lowest, signifying the highest risk.
5. **loan\_amnt:** **(INTEGER)** The loan amount requested by the individual.
6. **loan\_int\_rate: (FLOAT)** The interest rate associated with the loan.
7. **loan\_percent\_income: (FLOAT)** The percentage of income represented by the loan amount.
8. **cb\_person\_default\_on\_file: (INTEGER)** Historical default of the individual as per credit bureau records.
   1. Y: The individual has a history of defaults on their credit file.
   2. N: The individual does not have any history of defaults.

**Dependent variable:**

1. **loan\_status: (INTEGER)** Loan status, where 0 indicates non-default and 1 indicates default.

0: Non-default - The borrower successfully repaid the loan as agreed, and there was no default.  
1: Default - The borrower failed to repay the loan according to the agreed-upon terms and defaulted on the loan.