Problem Statement:

Title: Loan Default Prediction and Evaluation Criteria

Background: For financial institutions, the ability to predict loan defaults is a critical factor in managing financial risk, maintaining profitability, and ensuring sustainable growth. Defaulted loans can significantly disrupt the financial health of an institution, as they lead to substantial losses and may potentially affect the institution's credibility. Traditional risk assessment methods, which largely rely on a borrower's credit score, income, and collateral, have not been entirely efficient in predicting loan defaults. This inefficiency could be attributed to their inability to capture complex patterns and relationships in the data, thereby necessitating a more advanced approach.

Problem: The candidate needs to predict the likelihood of a borrower defaulting on a loan using a machine learning model. This model should consider a variety of factors that potentially influence a borrower's ability to repay a loan, such as income, employment status, credit score, loan amount, interest rate, and other relevant variables. The prediction model should also handle data imbalances, missing values, and outliers effectively. Second, the ability to address the problem, including the understanding of the problem, the proposed solution, the ability to build a predictive model, and the understanding of machine learning algorithms, data analysis, data pre-processing, feature engineering, and model evaluation methods.

In terms of the data science interview, the evaluation criteria should encompass the following:
- **Problem Understanding**: The candidate's understanding of the loan default prediction problem.
- **Proposed Solution**: Their proposed approach to solve the problem, includes data preprocessing, feature engineering, model selection, and model training.
- **Model Development**: The candidate's ability to develop a predictive model, including handling imbalanced data, missing values, and outliers.
- **Model Evaluation**: Their understanding of various model evaluation metrics relevant to the problem, such as precision, recall, F1 score, AUC-ROC, etc.
- **Knowledge and Experience**: Their understanding of machine learning algorithms, data analysis, and other relevant areas.
-

Evaluation Criteria:
- EDA and pre-processing
- Feature importance
- Modelling and results
- Business solution/ Interpretation of results obtained
- How to handle an imbalance dataset

System Design Tasks:

- Design system architecture to deploy this ML Model in production
- How do you perform canary build?
- What should be the strategy for ML Model Monitoring?
- How do you perform load and stress testing?
- How do you track, monitor and audit ML training?
- Design framework for continuous delivery and automation of machine learning tasks.