Case Study

As Data Scientist you have been asked to create a solution which would allow the decision makers to decide whether the person next month will be a defaulter or not. Yes=1, No=0. Please use the dataset in the csv file attached.

- As part of your work you should demonstrate how you would develop a machine learning model to accomplish this task. Please note that your final submission should be either a jupyter notebook or a python file.

- Machine Learning Tasks
    - Perform an Exploratory Data Analysis on the dataset including but not limited to
        - Understand correlation in dataset
        - Identify potential biases that my exist in the dataset
    - Create required Features for modelling, including but not limited to
        - Feature selection.: Different ways of selecting the top features from the dataset. Which one to prefer considering O(n).
        - Feature reduction
    - Develop an optimal model for the given dataset, including but not limited to
        - Optimal model selection
        - Handling of imbalanced dataset
        - Hyperparameter tuning- Strategy to save the experiments done using various parameter tuning Regularization techniques for better generalization
        - Model Explainability. Its OK to use any available open source package. Prefer the explanation on test data.
    - Deploy model for scoring on test data,
        - Model scoring. Which metric is chosen and why.
        - Drift identification
    - In the end create present your solution to the Interview panel using either of these:
        - Dashboards by using any Dashboarding tool to present the exploration done on the dataset.
        - ML Model prediction App using any open source (Streamlit/Flasgger) etc.
        - Presentation slides if the above does not work.

- Engineering Tasks. It would be an added advantage if you can showcase the same problem solving on any cloud platform by performing the engineering tasks:

    - Develop machine learning pipeline using any one cloud platform (AWS/Azure/GoogleCloud)
    - Set CI & CD for Machine learning models.
    - Implement Model Monitoring strategy.
    - Expose the machine learning model via API Gateway.