Data Scientist - Case

You are provided with the training_data.csv. This file belongs to a previous Kaggle competition, which had the aim to create a credit scoring model. You will be elaborating a model for a pre approved credit operation. This means that only clients that are selected by you will be able to take loans.

Thinking outside the box is encouraged. You are free to use the approach you find best, but are required to work with Python and complete the following tasks:

- EDA (exploratory data analysis)
- Elaborate a credit scoring model
- Policy Questions:
    - How would you structure the operation with regards to the use of the credit scoring model?
    - Would you consider maximizing profit, revenue or the number of loans taken? Do you think these are good ideas or propose other considerations?
    - How could you further improve your model?

Consider that:

- The case is open-ended on purpose, show us what you can do.
- While we care about clean code and polished graphs, we are more interested in your ideas while trying to solve this **business problem**.
- Be sure to add explanations regarding all content, decisions and assumptions included in the notebook.
- Presenting the notebook should take 25 minutes maximum, and that consists of the next stage of the selection process.

Deliverables

- Jupyter Notebook

Deadline

- 3rd of August at 03:00 UTC

Data dictionary

‘SeriousDlqin2yrs’ (target variable)

Person experienced 90 days past due delinquency or worse

‘RevolvingUtilizationOfUnsecuredLines’

Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits

‘age’

Age of borrower in years

‘NumberOfTime30-59DaysPastDueNotWorse‘

Number of times borrower has been 30-59 days past due but no worse in the last 2 years

‘DebtRatio’

Monthly debt payments, alimony, living costs divided by monthly gross income percentage

‘MonthlyIncome‘

Monthly income

‘NumberOfOpenCreditLinesAndLoans‘

Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)

‘NumberOfTimes90DaysLate’

Number of times borrower has been 90 days or more past due

‘NumberRealEstateLoansOrLines’

Number of mortgage and real estate loans including home equity lines of credit

‘NumberOfTime60-89DaysPastDueNotWorse’

Number of times borrower has been 60-89 days past due but no worse in the last 2 years

‘NumberOfDependents’

Number of dependents in a family excluding themselves (spouse, children etc.)