Programming Task

Create a python package that takes CSV data as an input and outputs high-level statistics of the data such as mean, median, outliers percentage, unique value count or percentage, and null value count or percentage for each column.

Things to consider:

1. Code should be production-ready. Candidates should assume the submitted code will get deployed in production immediately. Try to cover all the best practices possible you will include in the production codebase.
2. Create a flask API as an entry point and dockerize the code.
3. The high-level statistics should be returned as JSON Response.
4. We will highly admire modular and clean code structures.
5. Candidates should structure code in such a way that in the future if any other file formats will be provided the codebase should have minimal change in code structure.
6. Candidates should consider both numeric and non-numeric columns.

Please submit the code zipped in a single zip file and a document that includes the thought process of why you decided on those approaches and the limitations.

Analytics Task

Attached now the data for the coding / modelling challenge that we would like you to work on.

The task at hand is a **binary classification problem**, for which both a **training** and a **validation** data set are provided as csv files.

Notes:

- We would like you to explore both data sets, note down your key observations along with a short summary / visualization on this classification task.
- You are encouraged to ask questions if there's any clarification needed. We will try to clarify them as soon as possible.
    - Hint: some numbers could be in an unexpected format.
- Here you should build a classifier on the training data, with the goal of achieving the best performance possible on the validation data.
- You are allowed to manipulate / remove any independent variables in either data set in any way you see fit to achieve this goal.

- We would like you to use Python or R, and you are free to choose any packages and / or to create user defined functions to complete this task.
- We would like to get any code you write for this project, and would appreciate it if it is well commented on so that it's easier for us to read.

Submission details:

You can submit your assignment to this email thread or send it at:
akashshrestha@lftechonology.com

If you have any questions regarding the task, please contact at:
shishirbhattarai@lftechnology.com