Senior Data Scientist – Home Assignment

For this assignment you can use <u>any</u> tool or programing language you want, it can be Python, R, Scala or any other developing tool or language.

Business problem:

- We have a tabular dataset - *home_assignment.feather*
- Dataset contains history of testing results on a product from production.
- Each sample represents a single unit in a production line.
- Product manager has notified the PE engineers that the test 'TLJYWBE' is very expensive, and we would like to try and predict its outcome:

$$\begin{cases} x < 10^{-5}, TestPass \\ x >= 10^{-5}, TestFaill \end{cases}$$

Please provide an E2E solution to the problem, show how you evaluate the solution's performance and explain its strengths and weaknesses.

We are expecting to get a power point presentation or a PDF document for the solution and the data resources used as a CSV file(s) after all the data manipulation practiced on (if there are any such files).

If you have used a programming language such as python or platforms like Jupyter Notebooks, please share the code or any other relevant appendix that you have used during the assignment via a VCS repository.

GOOD LUCK! 😊