DS & MLE internship recruitment task

You work as a data scientists at one of the top universities in the USA. One day a rector of the university comes to you with a task. She wants you to investigate the university's admittance criteria and to create an engine that would recommend candidates with the highest probability of graduating. The university wants to maximise the graduate rates by enrolling only such students, because it is beneficial for both students and the university. Students get their degree, which makes it easier for them to start their career, and the university earns a tuition for the whole length of a program, as dropouts are minimised.

You start your investigation by gathering two sets of data:

- **score_board.csv** - contains candidates data with information on whether they were admitted or not;
- **graduates.csv** - consists only of students data (so only those that started their education on the university are included) with information on whether they graduated or not.

Columns definitions:

- id - candidate id;
- year - recruitment year;
- gpa - Grade Point Average;
- maths_exam, physics_exam, cs_exam, art_exam, language_exam - various exam results;
- social_activity - score for the social activities (volunteering etc), values from 1-5, 1 means least active, 5 most active;
- essay_score, interview_score - scores for an entry essay and an interview;
- score - total score calculated to rank candidates;
- accepted - whether a given candidate was high enough in a ranking to be accepted and whether they decided to use this opportunity;
- graduated - whether a student graduated;

Objectives

1. Explore the datasets and present your findings.
2. Build a model that will predict if a person will graduate or not. Write up a summary of your modeling effort discussing the strong points and limitations of your approach.
    a. Be creative. Treat this task as an opportunity to present us your skills. The more you showcase the better. If you are aiming for a more engineering focused job you could, for example, create a docker image that would contain your model as REST service.
3. Prepare a report (e.g. a Jupyter Notebook or Rmarkdown document) that will describe:
    a. **your idea** (e.g. what algorithm you used and in what configuration, e.g. the architecture of a neural network, or how you engineered features you fed to the model)

    b. initial data wrangling
    c. how you **evaluated the model performance** using the available data
    d. your results
    e. highlights of everything else we should note about your solution, esp. **what you are the most proud of**, e.g. the novelty of the approach you used
4. You can use whatever tools you like. The ideal solution should be in the form of git repository containing both your **code** and the **report**.

You have time until **May 24th**. Please send us a link to your solution via email: **rekrutacja.ioki@pearson.com**, even if it will be unfinished, or you will achieve results that will not be satisfactory.