

# DATA SCIENTIST

## MAIN ASSESSMENT

---

### Part I – Predictive part (8-10 hours)

#### *Requirements*

- Download any necessary files from that page: [Speech Corpus](#) and advice that page for any further information regarding the dataset structure.
- Follow the flow of the provided assignment.ipynb file and answer at least queries 1 and 2 or 1 and 3.

### Part II – Reproducibility (1-2 hours)

#### *Requirements*

- Create a public GitHub repo and upload any required files for reproducing your notebook.
- Create and upload also a Docker folder and provide the necessary instructions so as to conduct any of your experiments through that image.

### Part III – Presentation (2-4 hours)

#### *Requirements*

- Create a brief presentation following the next structure:
  - 1-2 slides: introduce the problem that you faced
  - 1-2 slides: data structure
  - 3-4 slides: your solutions and results
  - 1-2 slides: insights and future improvements

Keep in mind that your audience will be both Data Scientists and part of C-level (e.g., CFO, CEO), and you will have 15 minutes to present your draft.

**Good luck**