

COURSEWORK

The aim of this project is to assess your ability to analyze data samples using statistical learning techniques. Your tasks are as follows:

Identify an interesting dataset and **clean** it.

- 1) provide a brief **background** of your study, concisely describing its significance and getting the reader engaged in your problem.
- 2) **describe** the data, discussing the main information you will be using.
- 3) explicitly formulate your **research question(s)**.
- 4) describe and implement your **approaches and methodologies** to answer your research question(s).
- 5) **Report** the results of your implementations and drawing useful overarching, as well as specific and properly contextualized, **conclusions** and **interpretations** from them. Make use of relevant tables and visuals to underpin your findings and conclusions.
- 6) Discuss the **limitations** of your analyses.

The work must be presented in a short report. The report is maximum **5 pages, double-spaced, with font size 12 (Times New Roman), and with 1-inch margins**.

Breakdown of marks: Tasks 4) + 5) = 50%, Tasks 1) + 2) + 3) + 6) = 35%, others (language, typography, presentation, etc) = 15%.

Important note: I will be evaluating your knowledge of **this course** here. **Please focus your main methodologies on the techniques taught in the course!**

Submission deadline for report: 14:00 on Thursday 27th February 2024

A post-scriptum

Short reports like this are challenging to prepare and present. From past experience, your main challenges will be:

- 1) to **choose** the information/analyses to present – you will definitely run more analyses than those in your report and you will need to choose which ones are more pertinent and interesting.
- 2) to maximize your **signal-to-noise ratio** – everything in the report must convey important information.
- 3) to present your results and interpretations **concisely** and **precisely** – find a way to make convincing visuals and precise arguments on your results and descriptions. That means no beating around the bush, no diversions, and direct conveyance of information in an engaging way.