Data Processes – Second assignment

The COVID-19 pandemic significantly impacted society, particularly in 2020, with lasting effects in the following years. Two datasets containing synthetic data of patients from two hospitals have been generated. These datasets include different patient characteristics, such as sex, age, nationality, the date they visited the hospital's emergency department, their body temperature and oxygen saturation at the time of admission, as well as recorded symptoms and comorbidities. Additionally, the result of a PCR test is provided, with a positive result indicating COVID-19 infection and a negative result indicating the absence of infection

You are asked to define at least one business goal and one data mining goal, and for each, you must identify a Key Performance Indicator (KPI) to measure progress toward achieving these goals. Then, using appropriate data mining techniques, tools and models, you need to execute the technical part of the data project. Throughout this process, the KPIs and metrics generated from these techniques, tool and models must guide the approach to meeting the defined goals (if possible).

You need to prepare slides that explain all these elements:

- An introduction.
- The goals and KPIs.
- An exploratory data analysis.
- The preprocessing techniques applied (and the reasons for applying them).
- The techniques, tools, and algorithms used.
- The metrics obtained from the models and how they are used to guide the process to optimize the KPIs and achieve goals (if possible).
- The conclusions.

Notice that you do not need to write an entire report, only to prepare the slides.

A **ZIP file** that contains the slides in PDF format and the Python code used for the analysis must be uploaded to the Moodle assignment. The deadline for submission is **November 29th 2024, 23:59**.

The assignment must be completed in **groups of four students**, and don't forget to include all group members' names on the first slide.