Data Processes assignment

Assignment details

Covid19 pandemics has hit society in the last two years.

We have simulated a dataset of patients suffering covid that have been admitted to hospitals in the last 2 years.

The dataset contains data from patients admitted in different hospitals diagnosed with COVID-19 (age, sex, days in hospital, days in ICU, exitus, destination after being admitted in ER, and some medical parameters collected when they were firstly admitted in ER: temperature, heart rate, blood glucose, O_2 saturation, systolic blood pressure, and diastolic blood pressure).

The goal of the assignment is double:

1) to design a project plan in which a business goal must be detailed and consequently a plan designed. In this sense we must remember that important questions that we have been analysed all this period: factors that influence the survival, survival curves, anomaly detection, ….
2) to develop the project with all the tasks that have been explained in the project plan.

The assignment consists of two parts which must be clearly differentiated in the report that you have to elaborate:

- Project plan.
- Technical report, which must focus on the analysis of the data. In this regard, you must consider univariate and bivariate analysis, survival curves (e.g., Kaplan–Meier), and any other analysis that may help to understand the survival of a patient. You must also train and test different models to predict the survival (the most important part of the technical execution is the previous analysis though).

You must also include any code that you might have prepared and a README.txt document where you explain what files you delivered and the description of each one.

Do not forget to include the names of all the members in the group in the documents.

A ZIP file that contains all these files must be uploaded to the Moodle task for the assignment. The deadline to upload this file is **January 9th 2022, 23:55**.

The assignment must be done in **groups of 4 people**.