For this project, we'll use the metro interstate traffic volume dataset available on the UCI machine learning repository[1]. The original dataset recorded the hourly westbound traffic on I-94 between Minneapolis and St. Paul in Minnesota, from 2012 to 2018. For the purpose of this practical work, the dataset has been shortened to July 2016 to 2018 and contains the following columns:

- holiday – US National holidays plus regional holiday, Minnesota State Fair
- temp – Average temperature recorded in the hour, expressed in Kelvin.
- rain_1h – Amount of rain that occurred in the hour, expressed in millimeters.
- snow_1h – Amount of snow that occurred in the hour, expressed in millimeters.
- clouds_all – Percentage of cloud cover during the hour.
- weather_main – Short textual description of the current weather
- weather_description – Longer textual description of the current weather
- date_time – Date and time recorded in the CST time zone. The format is YYYY-MM-DD HH:MM:SS.
- traffic_volume – Volume of traffic reported westbound on I-94 during the hour.

The objective is to develop models based on the data provided to predict **one-step** and **multi-step** forecasts of traffic volume. The project must follow the CRISP-DM methodology and include code for its phases in the Python language:

- data exploration and preparation.
- data pre-processing.
- creation of models to predict traffic_volume.

  statistical models:

  Select/derive a subset of data manageable by statistical models

  Model the series with adequate statistical models

  machine learning models:

  Using several regression algorithms

  deep learning models:

  Using LSTM and GRU networks
- evaluation of the models created. Compare them at least with one baseline model.
- selection of the best or better traffic_volume prediction models.

[1] https://archive.ics.uci.edu/dataset/492/metro+interstate+traffic+volume

The notebook must be annotated in Markdown, providing as much detail as possible about the steps you took to arrive at your solutions. You should explain how you interpreted the most relevant graphical figures, how you cleaned and pre-processed the data, how you assessed the models, and the commitments you made throughout their development.

Deadline and submission instructions

- The project should be submitted to Moodle in the discipline area by **24:00 on December 22**. From this date, the note will be penalised by 10%, and no projects will be accepted after December 24, 2024.
- The notebook should be placed in a ZIP file with the designation MINDD-StudentsNum.zip.
- The presentation and evaluation of the project will in group and individual from January 6th to 11th.