**Dataset**

The dataset contains three files:

* network.csv
* inspections.csv
* sample\_submission.csv

The **network.csv** file contains a static picture of the gas pipeline network. Every row corresponds to a pipe and has a unique PipeId identifier. The table has 1.446.529 pipes.

The columns describe relevant features of each pipe. The complete list is:

PipeId - unique identifier for the pipe

Province - Spanish province where the pipe is located

Town - Town or city where the pipe is located

YearBuilt - Year in which the pipe was built and installed

Material - Material in which the pipe is built

GasType - Type of gas that runs through the pipe

Diameter - diameter of the pipe (in mm)

Length - Length of the pipe (in m)

Pressure - Pressure of the gas that runs through the pipe (bar)

NumConnections - Number of connections (external).

NumConnectionsUnder - Number of connections (internal and buried)

BoolBridle - Whether the pipe is bridled (True) or welded (False)

The **inspections.csv** file contains the results of the inspections carried out on the pipeline network from 2010 until 2020. On average, every pipe is usually inspected once every 2 years. The main result (target variable) is whether a leak or any other incident was found during the inspection. Severity of the incident is also reported on a scale from 1 (most severe) to 3 (least severe). The table contains 6.345.344 rows, each of which corresponds to a single inspection of a specific pipe. The columns are:

PipeId - unique identifier for the pipe

MaintenanceId - unique identifier for the inspection operation

InspectionYear - year in which the inspection took place

InspectionDate - date in which the inspection took place

MonthsLastRev - number of months elapsed since the last previous inspection.

Severity - Severity of the damage found (1: most severe, 3: least severe)

Incidence - Boolean whether an incident was found on the revision (1) or not (0).

Your task is to train a model with the information contained in these two files, whose purpose is to predict the inspections that were conducted in 2021 and 2022 on 909.733 pipes, whose PipeIds can be found in the **sample\_submission.csv** file. Note that these pipes should of course be described in the network.csv file and most of them will have undergone previous inspections in the inspections.csv file.

Your submission will need to have the same format as the **sample\_submission.csv** file, and it will be scored in the Kaggle platform according to the [AUC metric](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html).