

# DATA AND INFORMATION QUALITY PROJECT REPORT

PROJECT ID: ...

PROJECT NUMBER (1/2/3)

ASSIGNED DATASETS:

STUDENTS: (NAME SURNAME ID)

ASSIGNED TASK: (Classification/Clustering) (only for projects 1 and 2)

FOR PROJECT 1 AND 2 YOU MUST INCLUDE:

1. SETUP CHOICES
  - a. Chosen ML algorithms
  - b. Chosen ML performance evaluation metrics
  - c. Imputation/outlier detection techniques selected
2. PIPELINE IMPLEMENTATION
  - a. Description of the steps you performed
3. RESULTS
  - a. Description of the main results obtained
  - b. ML performance comparison between the imputation/outlier detection techniques you have implemented

FOR PROJECT 3 YOU MUST INCLUDE:

1. DATASET PROBLEMS
  - a. Describe the quality problems you found in the dataset
2. PIPELINE IMPLEMENTATION: description of the steps you performed
  - a. Explain the techniques used in the phases of normalization/standardization, data cleaning, data deduplication and the order in which they were used
  - b. Justify your choices