

Fundamentals of Data Science

2023/2024

First intermediate submission (20%): 3/11/2023 (20%)

Second intermediate submission (30%): 15/12/2023

Final report submission and presentation (50%): 15/01/2024

Assignment summary

The assignment focuses on the analysis of environmental and related data (climate, health, others). The general objective is to explore patterns and relations in data and report insights extracted from these patterns.

Objectives

First submission: data collection, reading and wrangling.

For this submission you need to collect and prepare the data, including:

- Structuring the data to a usable format
- Addressing noisy and inconsistent data, inc. outliers (correcting, removal)
- Addressing data completeness (imputation, removal)
- Merging datasets

Second submission: exploratory data analysis

For this submission you will focus on EAD, namely through visualizations.

Third submission: data modelling; reporting

For this submission you will apply statistical and/or machine learning models to identify patterns and predict outcomes.

Datasets

You need to integrate at least three datasets from the ones listed:

- OpenAQ
- EEA Industrial Reporting (pollutant emissions)
- European Climate Assessment & Dataset (or another climate dataset)
- UTD19 (road traffic)
- Others you identify