Datasets Requirements

Overview

When receiving a task, you can face two different situations *regarding datasets:*

- **S1: YOU CAN GO WITH THE DATASETS PROVIDED IN THE TASK ITSELF**
  - First, check whether the task you received contains three datasets.
  - Second, check that the three datasets meet the requirements specified in this document.
    - If so, you should always use these datasets.
    - Otherwise, you must look for external datasets, as specified below.
- **S2: YOU NEED TO LOOK FOR DATASETS OUTSIDE OF THE TASK**
  - If your task doesn't contain datasets (for example, for the datasets section at the top of the task, you see "N/A"), *you must look for and upload at least three datasets that meet the requirements* specified in this document.
  - If your task contains three datasets, **BUT** they do not meet the requirements specified in this document, *you must look for and upload at least three datasets that meet the requirements* specified in this document.

When submitting your task, ensure that you **paste the links to all datasets you used**. Also, ensure the link is correct and that anyone clicking on it will be directed to your intended dataset.

✅ Minimum Requirements

- You must use **at least 3** datasets per task**,** and **the three datasets** must be used **in a combined way.**
- **Datasets must be:**
  - **Publicly accessible** (no paywalls).
  - **Open source** (free for use, no private license).
  - **In English**.
  - **Clean and structured** (headers on the first row, no corruption).
  - **Sufficiently large** → Try not to use very small datasets. For guidance as to what we consider sufficiently large, aim for at least **5 columns** and **100 rows**, but this is not a hard requirement.

Less than 1 GB total (counting the 3 Datasets together)

⚠️ **Reminder:** Double-check each dataset before submitting to ensure it opens correctly and meets all requirements.