



Data Analytics Dataset Guidelines

OVERVIEW

When sourcing a dataset for your final project, it is important that you conduct this process with the utmost integrity and fairness to the dataset's owner, as well as the subjects represented within the dataset.

CRITERIA

The dataset should have at least two tables or two files that you will JOIN together. It can be stored in any format (.csv, json, sql database...).

PROHIBITED AND PROBLEMATIC DATASETS

Prohibited or problematic datasets could include any of the following:

- Personal information or any information that could be used to identify a person (e.g., personal health records that are linked to names).
 - Confidential or sensitive information (e.g., private sales information from a company used w/o authorization).
 - Information that could be considered obscene, defamatory, or harassing of others (use your best judgement).
 - Information that would infringe on someone else's intellectual property and/or information obtained illegally or in an unauthorized manner (e.g., data extracted from running an unauthorized/illegal script on a site you don't own).
-

DATASET REVIEW

If you are unsure if your dataset includes any of the above, please consult with your instructional team.