

QTM-340: Data Science with Text
Homework 4
Due Tuesday, 10/20 4:20 PM via Canvas

Increasingly, a best practice in the field of data science is to perform an assessment of your data before you begin your project. This assessment has been described as a “data biography” (Krause), a “datasheet” (Gebru et al.), and a “data user guide” (Gradeck), among other formulations. Thus, the first part of your final project, which you will submit as Homework 4, is to **perform an assessment of your chosen or self-created dataset.**

Follow the guidelines from “[Datasheets for Datasets.](#)” **Your assessment should consist of seven short paragraphs, one on each of the following topics:**

1. Motivation
2. Composition
3. Collection Process
4. Preprocessing/cleaning/labeling
5. Uses
6. Distribution
7. Maintenance

Consult “[Datasheets for Datasets](#)” for the specific questions that each paragraph should address. You need not answer all of the questions in each section; just select the questions that are relevant to your dataset.

Submission details:

Your homework should consist of any sort of text document (Word, PDF, plain text) with the following sections:

1. The names of each of the people in your project group.
2. A short paragraph that describes your dataset, or your specific plans to create your dataset, including links if relevant.
3. The datasheet, consisting of the seven sections described above.

Upload your document to Canvas by the date specified above.