

Assignment 2: Data Science CSCI/ERTH/ITWS (15% written and 5% presentation of overall credit score)

Due: Presented during week 7 & 8 (10/14 and 10/21)

Write up due October 21st, 2026 by 11pm ET.

Presentation Slides Due: October 14th, 2026 by 11am EST.

Submission method: LMS (write up and presentation slides)

Please use the following file naming for electronic submission:

DataScience_Assignment2_team_number_or_name.ext

Late submission policy: first time – no penalty, otherwise 20% of score deducted each late day.

Assignment: Choose **ONE** plan (either plan A or B from Assignment 1, not both) and collect the data.

Document the data collection process and present the data that you collected in addition to submitting a written report. Conduct descriptive analysis of some features in the dataset and produce relevant measures, visualizations, etc. with some reasonable conclusions.

The weighting score for each question is included below. Please use the question numbering (1, 2, 3 and a, b, etc.) and prompts to structure your written report.

1. Report on your data management practical **experience** (not the plan but the application) from the data collection exercise using a minimum of 3-4 sentences (for each item) (5%)

a). The goal (what it was driven by), and the mode(s) of collection.

b). How data were acquired, any problems, lessons learned, etc.

c). Physical and logical organization of data/ metadata, success, failures.

d). Metadata and documentation collected/ stored, what provenance was identified.

e). A link to the data/ metadata (or a copy of it).

2. Describe how well (or not) the data management **plan** turned out in practice. 6% (6000-Level students)/ 8% (4000-Level, students)

a). Highlight what worked, what did not, and why (2-3 sentences for each of the 9 data management categories).

b). What would you do differently next time and why? (2-3 sentences total)

c). Any insights about the data or metadata collected: completeness, quality, etc.? (2-3 sentences total)

3. Conduct some preliminary/exploratory descriptive analysis of a feature (or more) in your dataset. 2%

These could include examining a variable's distribution, association between two variables, identifying outliers, trends, recurring patterns, etc.

Use methods such as:

- Summary statistics, boxplots, histograms, scatter, line, density, violin plots, network diagrams, etc.

NOTE: Keep the analytics descriptive, there is no requirement to do data mining, develop predictive models, etc. There is no minimum number of plots/measures, just make sure whatever analysis you present demonstrates the purpose for which the data were collected.

4. **6000-Level** student question: related to Q 1 and Q 2. 2%

a). What standards (data format, metadata, etc.) did you use (whether planned or not) and/or what ones couldn't you use and why? (2-3 sentences).

b). What standards/conventions would have been more helpful ('none' is not an acceptable answer) Minimum 1 sentence.

c). How do your choices of organization, formats, frameworks, and standards help capture and elucidate the provenance of your data product(s)? (3-4 sentences).

5. Oral presentation of the data collection and management in practice (5%).

Please submit your presentation (ppt, pdf or similar using the same naming scheme noted above).

Guidelines:

a). ~ 5 mins (MAX 😊)

b). 5-6 slides MAXIMUM

c). What you should present:

i. The goal (what it was driven by), and the mode of collection

ii. How data was acquired, any problems, lessons learned, etc.

iii. Physical and logical organization of data/ metadata, success, failures

iv. Other 'data management' aspects that you deem important

v. Metadata and documentation collected/ stored

vi. Present some of the data in some form (i.e. curation)

Include all references, citations and attributions of external information sources used in the writeup.

THE END