Assignment 1

Due Wed. September 17, 2025

Overview/Submission guide

1. [30 marks] Choose a real-world scenario where data would be collected. Pretend that you have access to real data from the scenario.
2. [3 marks] Devise a relational table with at least 12 tuples (rows) and at least 4 attributes (columns) for the scenario. Treat these values as the true values for the scenario. Be prepared to revise your scenario and relational table as you continue with the other parts of this question.
3. [3 marks] Prepare another relational table with generally similar values to the first relational table but with examples of all the following types of error:
   * 1. A measurement error
     2. A data collection error
     3. Noise
     4. An artifact
     5. Imprecision
     6. Bias
     7. Inaccuracy
     8. An outlier
     9. A missing value
     10. An empty value
     11. Inconsistent values
     12. Duplicate data
4. [24 marks] For each of the 12 types of error, explain how it is relevant to the relational table given for part (b) and why it is this type of error. You can look at section 2 of the notes for explanation and inspiration, but do not use exactly the same attributes or values as are used in the examples.
5. [70 marks] Ordered set representation of a data cube  
   Provide a file containing the ordered set representation of a data cube for a 3D cube. Read in the data and construct the totals around the outside of the cube. Display the cube.  
   Provide a file containing the ordered set representation of a data cube for a 4D cube. Read in the data and construct the totals around the outside of the cube. Display the cube.