1 Academic Integrity (0%, Mandatory)

Read all three of these pages


My general discussion: http://www.cs.ubc.ca/~tmm/courses/cheat.html

Department guidelines: http://www.cs.ubc.ca/about/policies/collaboration.shtml

Submitting this assignment is attesting that you fully understood that material.

2 Self-introduction (0%, Mandatory)

Introduce yourself on Piazza, in thread https://piazza.com/class/kikk5imzxi62xi?cid=7

3 Data Abstraction: Attributes (5%)

What type of attribute (categorical, ordinal, quantitative) are the following?

1. 50 meter race times
2. College major
3. Amazon rating for a product
4. Product name
5. Weight of your favorite cat
6. Personality of your favorite cat
7. Rank of your favorite cat in the local feline dominance hierarchy
8. Length of your favorite cat's longest whisker
9. Size of your favorite sweatpants
10. Color of your favorite sweatpants

Rubric: .5% each

Exercise Credit: adapted from Lex & Meyer, Utah

4 Data Abstraction: Central Park Squirrel Census (27%)

2018 Squirrel Census Data

https://data.cityofnewyork.us/Environment/2018-Central-Park-Squirrel-Census-Squirrel-Data/vfnx-vebw

For each field

- Write down the attribute type (categorical, ordinal, quantitative)
- Determine its cardinality (number of levels) for categorical or ordinal, or range for quantitative

Rubric: 18 fields. For each .5% type (9%), 1% cardinality/range (18%)
5 Data & Task Abstraction: Foreign Aid (68%)

Foreign aid dataset
For this dataset, read the field descriptions below, in domain language

3.1: Overall
- What is the dataset type(s)?
- How many fields/attributes does it have?
- How many items are there?

3.2: Analyze for each of the 7 fields
(AidData ID, Donor, Recipient, Year, Commitment Amount, Purpose Code, Purpose Name / Title Description)
- Determine its cardinality/range. For categorical attributes, write down number of unique levels. For quantitative attributes, specify range from min to max.

Does this characterization reveal any anomalies that you think might be dataset quality problems?

3.3: Write 2 questions you would like to answer with this data set, from the point of view of an aid worker reporting to the government of a country providing aid.

3.4: For each question, write the following information:
- Do you need a chart in order to answer this question? For any question that does not require a chart, create a new question that would benefit from one and use that instead.
- Which fields/attributes do you need to use to answer the question?
- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?
- Do you need to transform the data in order to answer the question? If yes, what transformations are needed?

3.5: Write 2 questions you would like to answer, from the point of view of a journalist reporting to the citizens of a country receiving aid.

3.6: For each new question, do step 3.4 above.
Rubric:
3.1: 1% each, 3%
3.2: 7 fields, 1% each type/cardinality, 2% last one = 17%
3.3 / 3.5: 2 questions, 6% each = 12, *2 = 24%
3.4 / 3.6: attributes 4%, last two 1% = 6*2=12%, *2 = 24%

*Exercise Credit: adapted from Bertini, NYU Tandon*