CPSC 436V, Foundations 1 Answer Template

Out: Thu Jan 13 2022. Due: Fri Jan 21 2022, 6pm.
Submit to Gradescope, as a file PDF format following this gdoc answer template.
Make a copy of this document by navigating on the toolbar: “File” > “Make a copy”

0 Name & Student Number:

Kevin Cai
41146127
kevincgc
p4u9a

1 Academic Integrity (0%, Mandatory)

I attest that I have read the three pages on academic integrity and fully understand this material.

2 Self-introduction (0%, Mandatory)

I have introduced myself on Piazza

3 Data Abstraction: Attributes (5%)

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

1. 100 meter race times: quantitative
2. College major: categorical
3. Amazon rating for a product: ordinal
4. Product name: categorical
5. Weight of your favorite cat: quantitative
6. Personality of your favorite cat: categorical
7. Rank of your favorite cat in the local feline dominance hierarchy: ordinal
8. Length of your favorite cat's longest whisker: quantitative
9. Size of your favorite sweatpants: ordinal
10. Color of your favorite sweatpants: categorical

Rubric: .5% each

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

Field name	Attribute type	Cardinality	Range
X	Quantitative		-73.98115858 to -73.94972177
Y	Quantitative		40.76491067 to 40.80011852
Unique Squirrel ID	Categorical	3018	
Hectare	Categorical	339	
Shift	Categorical	2	
Date	Quantitative		10062018 to 10202018
Hectare Squirrel Number	Ordinal	23	
Age, Location	Categorical	3	
Primary Fur Color	Categorical	4	
Highlight Fur Color	Categorical	11	
Combination of Primary and Highlight Color	Categorical	22	
Color notes	Categorical	133	
Above Ground Sighter Measurement	Ordinal	40	
Specific Location	Categorical	288	
Running, Chasing, Climbing, Eating, Foraging, Kuks, Quaas, Moans, Tail flags, Tail twitches, Approaches, Indifferent, Runs from	Categorical	2	
Other Activities	Categorical	308	
Other Interactions	Categorical	198	
Lat/Long	Categorical	3023	

Rationales (if needed):

- If the cell was left blank for a categorical attribute, it was counted as its own category of “Unknown,” so an attribute like Primary Fur Color had potential values of Black, Cinnamon, Gray and Unknown.

- Lat and long are both quantitative attributes by themselves, but when combined as a single attribute, I think it makes more sense to consider the position as a categorical attribute. Adding positions doesn’t make sense and subtracting gives the Manhattan distance between two points, which generally isn’t useful.

- I interpreted Above Ground Sighter Measurement as the relative height above ground and not as a specific height measurement in metric/imperial units.

Rubric: 18 fields. For each .5% type (9%), 1% cardinality/range (18%)

5 Data & Task Abstraction: Foreign Aid (68%)

5.1: Overall
-What is the dataset type(s): Table
-How many fields/attributes does it have: 7
-How many items are there: 499

5.2: Analyze for each of the 7 fields

Field name	Attribute type	Cardinality	Range
AidDataID	Categorical	499	
Donor	Categorical	25	
Recipient	Categorical	129	
Year	Quantitative		1991 to 2010
Commitment Amount	Quantitative		0.152847 to 101905000
Purpose Code	Categorical	127	
Purpose Description	Categorical	134	

Does this characterization reveal any anomalies that you think might be dataset quality problems?
Yes, I highly doubt that Norway sent Madagascar $0.152847 USD in 2010 for Teacher training. Additionally, there are 127 codes but 134 descriptions, even though identical codes should have the same description.

5.3:
Q1: How is the aid distributed proportionally for the countries that we have sent aid?

Q2: How has the annual aid given changed over time?

5.4:
5.4.1:
- Do you need a chart in order to answer this question and why?
Yes, the data can be presented much more clearly as a pie chart rather than a list of percentages.

- Which fields/attributes do you need to use to answer the question?
Donor, Recipient, Commitment Amount

- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?
All needed data is here.

- Do you need to transform the data in order to answer the question? If yes, what transformations are needed?
Yes, I would need to calculate the total amount sent for each country by summing up the amounts for each recipient when the donor is my country. I would also need the total amount sent by summing up all amounts when the donor is my country. I can then calculate the proportional amount sent to each country.

5.4.2:

- Do you need a chart in order to answer this question and why?
Yes, the data can be presented as a scatterplot showing by how many percentage points each year deviates from the mean over the timespan.

- Which fields/attributes do you need to use to answer the question?
Donor, Year, Commitment Amount

- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?
All needed data is here.

- Do you need to transform the data in order to answer the question? If yes, what transformations are needed?
Yes, you have to calculate the yearly mean aid provided over the period of 1991-2010 and then calculate by how many percentage points each year deviates from the mean.

5.5:
Q1: How is the purpose of aid distributed?

Q2: How much aid was actually received versus how much was committed by donors?

5.6:
5.6.1:
- Do you need a chart in order to answer this question and why?
Yes, a pie chart would be much better at showing distribution between categories than a list of values.

- Which fields/attributes do you need to use to answer the question?
Recipient, Purpose Code

- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?
Yes, all needed data is given.

- Do you need to transform the data in order to answer the question? If yes, what transformations are needed?
I would have to sum up the aid for each purpose.

5.6.2:
- Do you need a chart in order to answer this question and why?
Yes, a bar chart with an amount given bar and an amount committed bar for each donor country would greatly enhance the readability of the difference between how much a donor commits and how much they actually send compared to a list of values.

- Which fields/attributes do you need to use to answer the question?
Donor, Recipient, Commitment Amount

- Do you have all the data you need to answer this question, or would you need additional data fields that are not provided here?
I would need Amount Received in addition to Commitment Amount to answer this question.

- Do you need to transform the data in order to answer the question? If yes, what transformations are needed?

No transformations are necessary.

Rubric:
5.1: 1% each, 3%
5.2: 7 fields, 1% each type/cardinality, 2% last one = 17%
*5.3 / 5.5: 2 questions, 6% each = 12, *2 = 24%*
*5.4 / 5.6: attributes 4%, last two 1% = 6*2=12%, *2 = 24%*