

Take Test: Assignment 2 Answer Sheet

Test Information

Description Deadline: 10/10/2017

- Instructions
- Feel free to talk to other members of the class in doing the homework. We are more concerned that you learn how to solve the problem than that you demonstrate that you solved it entirely on your own. You should, however, write down your solution yourself. Please try to keep the solution brief and clear.
 - Please use Piazza first if you have questions about the homework. Also feel free to send us e-mails and come to office hours.
 - The homework is due at 11:59 PM on the due date. We will be using Compass for collecting the homework assignments. **Please submit your both answers (fill in blank in compass) and details (via pdf) in Compass (<http://compass2g.illinois.edu>).** Please do NOT hand in a hard copy of your write-up. Contact the TAs if you are having technical difficulties in submitting the assignment. **We do NOT accept late homework!**
 - The homework should be submitted in pdf format. You are required to submit the source code, and use the file names to identify the corresponding questions. For instance, 'Question1.netid.py' refer to the python source code for Question 1, replace netid with your netid. Compress all the files (pdf and source code files) into one file. Submit the compressed file ONLY.
 - For each question, you will NOT get full credit if you only give out a final result. Necessary calculation steps are required.

Multiple Attempts Not allowed. This test can only be taken once.

Force Completion This test can be saved and resumed later.

❖ Question Completion Status:

Save All Answers

Close Window

Save and Submit

QUESTION 1

24 points

Save Answer

Assume that a base cuboid of 10 dimensions contains only 3 base cells:

$(a_1, a_2, a_3, c_4, \dots, c_9, c_{10}), (b_1, b_2, b_3, c_4, \dots, c_9, c_{10}), (c_1, c_2, c_3, c_4, \dots, c_9, c_{10})$

where $a_i \neq b_i, b_i \neq c_i$ and $a_i \neq c_i$, for any $i=1, 2, 3$. There is no dimension with concept hierarchy. The measure of the cube is count. The count of each base cell is 1.

a. cuboids are there in the full data cube.

b. The complete cube will contain distinct aggregated (i.e., non-base) cells.

c. An iceberg cube will contain distinct aggregated cells, if the condition of the iceberg cube is count ≥ 2 .

iceberg cube is $count > 2$.

d. The closed cell with $count = 3$ has non-star dimensions.

QUESTION 2

36 points[Save Answer](#)

This question aims to provide you a better understanding of measures as well as cuboid structures. You can find an artificially generated dataset called Q2data in the attachment of Assignment 2 (also on compass). It contains 50 lines, where each line representing a business. For each business, it has (*Business id, State, City, Category, Price, Rating*). We now want to construct a cube over four dimensions (*Location, Category, Rating, Price*) with count as the measure. Note that in the Location dimension, there is a concept hierarchy, i.e. City and State.

You should write scripts to manipulate the data. There is no restrictions on programming languages and you are allowed to use any built-in functions. In addition, you also need to include necessary steps in your .pdf file, to show how you get the results and submit your code. Please process the data, and answer the following questions.

- a. There are cuboids in this cube.
- b. There are cells in the cuboid (*Location(city), Category, Rating, Price*).
- c. Now let's drill up by climbing up in the Location dimension, from City to State. There are cells in the cuboid (*Location(State), Category, Rating, Price*).
- d. Further, there are cells in the cuboid (**, Category, Rating, Price*).
- e. The count for the cell (*Location(state) = 'Illinois', *, rating = 3, Price = 'Moderate'*) is .
- f. The count for the cell (*Location(city) = 'Chicago', Category='food', *, **) is .

QUESTION 3

40 points[Save Answer](#)

Given the dataset Q3data which contains 100 transactions, each line is a transaction and each transaction contains item(s) separated by spaces. Please do frequent pattern mining on this dataset and answer the following questions. If the answer is not an integer, please round the result to 3 decimal places.

- a. Suppose the minimum support is 20.
 - 1. The number of frequent patterns is .
 - 2. The number of frequent patterns with length 3 is .
 - 3. The number of max patterns is .
- b. Suppose we decrease the minimum support to 10.
 - 1. The number of frequent patterns is .
 - 2. The number of frequent patterns with length 3 is .
 - 3. The number of max patterns is .

4. The confidence measure of the association rule $(C, E) \rightarrow A$ is .

5. The confidence measure of the association rule $(A, B, C) \rightarrow E$ is .

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

Save All Answers

Close Window

Save and Submit