Started on	Tuesday, July 30, 2019, 2:17 PM
State	Finished
Completed on	Tuesday, July 30, 2019, 2:17 PM
Time taken	38 secs
Points	10.00/10.00
Grade	**100.00** out of 100.00

Information

Now, as usual, we'd like you to go through Lantz's example, coding along with him and answering the following questions as you go. This is pp 266 - 283, with a lot of great examples of how to visualize the data, stirred in with learning the algorithm.

And look for the request to save your output for the Homework.

Question **1**

Correct

1.00 points out of 1.00

You would like to have the results be as specific as possible, like what particular brand of milk is associated with what other purchases.

Select one:

- True
- False ✔

Not so -- as Lantz notes on p 267, we might want to delete any too-specific info, so that we can see the meaningful associations, like what else people generally buy when they buy milk.

The correct answer is 'False'.

Question **2**

Correct

1.00 points out of 1.00

Transaction data lends itself to matrix representation.

Select one:

- True
- False ✔

Not necessarily. As Lantz says on p 267, a transaction can be shown as a comma-separated list of any number of items, with features scattered throughout. Not very readily clarified by feature columns. Lantz describes how easy remedies don't really fix this problem, on p 268.

The correct answer is 'False'.

Question **3**

Correct

1.00 points out of 1.00

A "sparse matrix" is a good solution to having a huge variety of data features.

Select one:

- True ✔
- False

Yes, indeed. You should try to get comfortable with the one for the groceries, that Lantz introduces on p 269. These are a major tool to solve the "big data" problem of having really scattered features.

The correct answer is 'True'.

Question **4**

Correct

1.00 points out of 1.00

The R "arules" package lets you answer useful questions about a sparse matrix of data, like which ones have a certain level of support.

--> Homework: Save an image of your "histogram" like the one Lantz has at the bottom of p 272, to turn in.

Select one:

- True ✔
- False

Yes -- see p 272.

The correct answer is 'True'.

Question **5**

Correct

1.00 points out of 1.00

Using the R "apriori()" function, you may have to guess a few times, at things like the right levels of confidence and support.

Select one:

- True ✔
- False

Yes, you can see Lantz guessing at these, at the bottom of p 275.

The correct answer is 'True'.

Question **6**

Correct

1.00 points out of 1.00

The "apriori" library is made to let you use the same rules multiple times, in a production environment.

Select one:

- True ✔
- False

It is. This is why there are capabilities for saving detailed rules by an easy name, and for writing them to save in permanent storage. When you have a big, elaborate mess of output, you may want to come back to it.

The correct answer is 'True'.

Question **7**

Correct

1.00 points out of 1.00

In Apriori terms, applied to sales, "lift" means how much more likely an item is to be purchased, given a particular association.

Select one:

- True ✔
- False

This is how Lantz describes "lift" on p 278.

The correct answer is 'True'.

Question **8**

Correct

1.00 points out of 1.00

Lantz concludes that the "lift", of buying milk in association with buying potted plants, probably is not "actionable."

Select one:

- True ✔
- False

Right -- the "lift" data is there, described on p 279, but it's probably an inexplicable one.

The correct answer is 'True'.

Question **9**

Correct

1.00 points out of 1.00

Lantz felt it might be an important discovered "gem", that people who buy herbs are nearly four times more likely to buy root vegetables.

--> Homework: Save an image of your sorted output of the grocery rules, like Lantz shows at the top of p 281.

Select one:

- True ✔
- False

He surely did, on p 281. This is the kind of new information that "big data" analysts get paid to discover.

The correct answer is 'True'.

Question **10**

Correct

1.00 points out of 1.00

Lantz concludes that the apriori "subset()" function isn't especially useful.

Select one:

- True
- False ✔

No, he really likes this function, and shows why. See p 282.

The correct answer is 'False'.

Question **11**

Complete

Not graded

We want to base online and remote face-to-face discussions on the topics of most value to you.

Please think carefully about all the material you read, then write a prompt for discussion you would like to hear - either:

a. Something that you aren't sure about, which you'd like to have explained in class, or

b. A topic you liked a lot, that you'd like to discuss in class.

na

Thanks!