Feedback — Quiz Week 6

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Sat 21 Nov 2015 5:13 PM CET**. You got a score of **4.00** out of **5.00**. You can attempt again, if you'd like.

Question 1

In the first lecture of week 6 the three types of operational support are discussed: detect, predict and recommend. To which of the following four generic data science questions does prediction relate?

Your Answer		Score	Explanation
⚪ What happened?			
⚪ Why did it happen?			
⦿ What will happen?	✔	0.50	Indeed, predictions look into the future and predict what will happen, or what is most likely to happen.
⚪ What is the best that can happen?			
Total		0.50 / 0.50	

Question Explanation

Operational support is discussed in more detail in lecture 6.1: 'Operational support: detect, predict and recommend'. The

Question 2

Consider the example tabular data shown below:

Customer				Order				Order-Delivery			Delivery		
Cust_ID	**Name**	**Gender**		**Ord_ID**	**Cust_ID**	**NrOfItems**		**Ord_ID**	**Del_ID**		**Del_ID**	**Time**	**Truck_ID**
1	Wil	Male		11	1	3		11	21		21	01-Jan	1

| 2 | Joos | Male |
| 3 | Anne | Female |

| 12 | 1 | 4 |
| 13 | 2 | 2 |

11	23
12	21
13	22
13	24

22	01-Jan	2
23	02-Jan	1
24	03-Jan	2

Of course more data can be recorded in each table, such as the address of the customer, the time the order was confirmed, etc.

Based on the data shown above, customer Wil for example has two orders (orders 11 and 12), and order 11 has two deliveries (deliveries 21 and 23).

We believe that the interaction of a customer with our company varies by gender, i.e. placing orders and receiving deliveries is different. If we want to discover this difference, what would we need to select as the case or process instance?

Your Answer	Score	Explanation
○ The order-delivery relation is our process instance.		
◉ The order is our process instance.	✖ 0.00	If one would select the order as the process instance it is not possible to see how one customer places different orders. Therefore this is not the correct process instance.
○ The customer is our process instance.		
○ The delivery is our process instance.		
Total	0.00 / 0.50	

Question Explanation

The difficulties encountered when converting data to an event log are discussed in lecture 6.2: 'Getting the right event data'.

Question 3

Consider the data as shown in the previous question.

How many deliveries are recorded for order number 13?

Note: enter a number only (and no sentences etc.), f.i.: '1' or '11' is correct input (but maybe not the

correct answer!) while 'One' is not correct input.

You entered:

2

Your Answer		Score	Explanation
2	✔	0.50	
Total		0.50 / 0.50	

Question Explanation

For order number 13 the table order-delivery contains two records: deliveries 22 and 24 are related to this order. Therefore the correct answer is 2.
The difficulties encountered when converting data to an event log are discussed in lecture 6.2: 'Getting the right event data'.

Question 4

Again consider the data of the previous two questions.

If we select the order as our process instance, what issue do we encounter?

Your Answer	Score	Explanation
⦿ A single delivery occurs for more than one order, hence this is event is duplicated in the event log.	✔ 0.50	This is indeed the main issue that is encountered. This would create issues when we want to investigate the utilization of trucks, we should not count this delivery twice for the same truck. Including exactly the same event multiple times in event logs is therefore a tricky thing to do with sometimes unexpected results.
○ An order cannot be selected as a process instance since there is no case identifier and related events.		
○ There are no relations from the order table to any other table, hence we get an event log that does not contain much detail.		

Total	0.50 / 0.50

Question Explanation

The difficulties encountered when converting data to an event log are discussed in lecture 6.2: 'Getting the right event data'.

Question 5

Guideline for Logging number 6 reads:

Events should be at least partially ordered. The ordering of events may be stored explicitly (e.g., using a list) or implicitly through an attribute denoting the event's timestamp. If the recording of timestamps is unreliable or imprecise, there may still be ways to order events based on observed causalities (e.g., usage of data).

Why is it so important that events are (partially) ordered?

Your Answer		Score	Explanation
☐ It is crucial to know the order of events to have correct basic statistics such as the number of times an activity was executed or by whom.	✔	0.12	For this example statistics the order of events is irrelevant since it can be calculated by simply counting frequencies, the order of events is irrelevant.
☑ It is crucial to know the order of events in order to correctly discover a process model.	✔	0.12	Without knowing the (partial) order of events, it is impossible to discover the control flow which a case follows.
☑ It is crucial for replay on a process model.	✔	0.12	When replaying a trace on a process model the order of events matter since it has to be determined which events are executed in what order for the case/trace to see whether deviations from the process model occur.
☐ It is crucial to know the order of events to correctly discover the working together social network, which shows	✔	0.12	For this example the order of events is irrelevant, since only whether two users work together on the same case is required information.

which users work
together in the same
case.

Total	0.50 / 0.50

Question Explanation

The guidelines for logging are explained in more detail in lecture 6.3: 'Guidelines for logging' and guideline 6 as of 4:30 min.

Question 6

Guideline for Logging number 7 reads:

If possible, also store transactional information about the event (start, complete, abort, schedule, assign, suspend, resume, withdraw, etc.). Having start and complete events allows for the computation of activity durations.

Which type of analysis is not possible when only the complete/end time of an activity is recorded, and not when it started etc.?

Your Answer	Score	Explanation
○ Discovery of a process model.		
○ Creation of a dotted chart.		
● Calculation of resource utilization, e.g. when is a resource working and when not.	✔ 0.50	Without knowing when work began on an activity, and ideally also when work was suspended and resumed, it is impossible to provided detailed estimations on resource utilization. Of course, one could assume that a resource starts working on an activity as soon as it becomes available, but usually this is not the case in practice.
○ Creation of basic statistics such as the number of times an activity was executed or by whom.		

Total	0.50 / 0.50

Question Explanation

The guidelines for logging are explained in more detail in lecture 6.3: 'Guidelines for logging' and guideline 6 as of 4:30 min.

Question 7

What statement regarding the difference between process mining analysis and general BI analysis is true?

Your Answer	Score	Explanation
◉ Process mining explicitly incorporates the process in the analysis whereas BI analysis does not (or statically).	✔ 0.50	This is indeed the main advantage since the process greatly influences which conclusions can be made from the data, and adds crucial information during analysis.
○ Process mining analysis is supported by more commercial tools than BI analysis.		
○ Process mining exists longer than BI analysis.		
Total	0.50 / 0.50	

Question Explanation

The difference between process mining tools and 'general' BI tools is explained in lecture 6.4: 'Process mining software'.

Question 8

Why is the second stage in the L* lifecycle model, creation of a control-flow model, an important phase?

Your Answer	Score	Explanation

○ The control-flow model is required for social network analysis.

○ The control-flow model is required for visual analysis such as the dotted chart and the log dashboard.

◉ The control-flow model should be well-connected to the event data to perform further analysis such as conformance checking and bottleneck analysis.	✔ 0.50	This is indeed the main reason, the discovered control-flow model is used for further analysis and therefore should have a good connection with the event data.
Total	0.50 / 0.50	

Question Explanation

The L* lifecycle model for process mining is discussed in lecture 6.5: 'How to conduct a process mining project'.

Question 9

Why is application of process mining on a spaghetti-type process potentially more beneficial?

Your Answer	Score	Explanation
◉ The process is more complex and may be harder to understand. Therefore, the potential benefits of using process mining may be more substantial.	✔ 0.50	Spaghetti processes are indeed less-structured, and therefore the process is less understood. Process mining can provide significant insights in how the process runs, and where it could be optimized.
○ In general, spaghetti-type processes are more expensive because they are more labor intensive, require more steps and are less supported by IT systems.		
○ Within spaghetti-typed processes more data is		

available regarding case
characteristics, changed
data, etc.

Total	0.50 / 0.50	

Question Explanation

Analyzing spaghetti typed processes is explained in more detail in lecture 6.7: 'Mining spaghetti processes'.

Question 10

Why does the single perfect process model not exist?

Your Answer	Score	Explanation
○ All of the three reasons mentioned here.		
◉ There simply does not exist a process model that scores very good on all four quality dimensions (replay fitness, precision, generalization and simplicity).	✖ 0.00	This is indeed a correct statement, especially for real-life processes there are trade-offs between the quality dimensions to be made as the perfect model does not exist. However, this is not the only valid reason and therefore the answer is not correct.
○ A different process model (based on the same event data) is required depending on the usage of the process model.		
○ Every user has a different taste of what is a nice looking process model based on the same event data.		
Total	0.00 / 0.50	

Question Explanation

The comparison between process models and maps are made in lecture 6.8: 'Process models as maps'.