Appendices

1 Experiment Instrument

Prior questionnaire

Name:

Gender: M / F

Group:

1. It was easy for me to understand the process that was described in the assignment

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

2. I know what is expected from me

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

3. Please rate how certain you are that *you can* do the things discussed below (on a scale from 0 ('Cannot do at all') to 100 ('Highly certain to do'):

   (a) How certain are you that *you can* properly *build an event log*, using the process information that was provided in this case?

   /100

   (b) How certain are you that *you can* properly select the *most suited process instance* to serve the purpose, using the process information that was provided in this case?

   /100

   (c) How certain are you that *you can* properly identify the *most suited process activities* to serve the purpose, using the process information that was provided in this case?

/100

(d) How certain are you that *you can* properly take into account the *analysis consequences* of your decisions during the event log building, using the process information that was provided in this case?

/100

State the current time: __ : __

The assignment

Build the architecture of an event log for the case process in such a way that it facilitates the process analysis phase.

The architecture of an event log gives an unambiguous answer to the questions:

- which process instance will be followed throughout the process (and which unique identifier will be used for that purpose)?
- which activities will be stored (and which table fields will serve to this end)?
- which attributes will be stored (and which table fields will serve to this end)?

Please state for your selection of the process instance:
- what benefits this selection has
- what possible downsides this selection has

Please state for your selection of activities, in combination with the selection of the process instance:
- what possible downsides could arise

Context: Process Mining project at e-tailer

A large, international e-tailer decides to analyse its on-line ordering process. Annually, they ship around 53 000 products for a total value of 1.4 million euro. The company's product catalogue contains more than 20 000 products and it has an active customer base of 30 000 customers. The process owner would like to analyse the on-line ordering process through a process mining analysis. Like every project sponsor, he is interested in as many as possible insights in the process. In the kick-off meeting, a description of the process and information on the stored data is provided, along with some more explicit expectations of the analysis.

Description of the process, as provided by the process owner

"The process goes through two phases. First there is a web phase in which the customer browses the website and orders a product. Then there is the physical phase in which the order is picked, packed and delivered. For the whole process, we would for instance like to know what percentage of the cases is following our process model, like we designed it.

Everything starts with the webshop. The marketing team is responsible for filling this store with interesting items. They also have to pay attention to the lay-out, using the available space to guide the customers to items they might buy. If you want, you can contact Mrs. Lane of Marketing for more information on that part. She has a lot of valuable information for you, I am sure.

When customers visit the webshop, they can browse through the catalogue and add an item of interest to the shopping cart in case they decide to purchase that item. This is done by clicking the "add to cart" button. Placing items in the shopping cart creates an order, with each item (during that browsing session) being a part of that order. Each item can be seen as an order line. After placing an item in the shopping cart, the customer can decide to continue shopping, or to place the order. If the customer leaves however the browsing session, the order stays unfinished and there is no possibility to return to this order. Our IT guys still need to create a cookie-based system so that the customer can continue on that order, next time he visits our shop. Before making this investment, it would be interesting to know how many times a visitor creates an order (and for how much money), and then abandons the browsing session without payment. However, coming back to my story, if the customer places the order, the payment registration is launched.

Once the payment is launched, an invoice is created that matches that order. The system creates this invoice automatically. Most of the times, it is the accounting department's respon-

sibility to assure this process step. Sometimes another department. It would be nice to have some insights in which departments do a better job in collecting the cash in a quickly fashion. During the phase of the invoice creation, an automatic inventory check already takes place. In case all ordered items are available, one invoice is created for these items. Items that are out of stock, are grouped in a different invoice. The created invoice(s) are displayed to the customer who then can choose a payment method. The first option is using a credit card and then payment is registered instantly. If they prefer to wire transfer the money instead, they get an e-mail with the necessary details. In this case, payment is only registered a few days later, when the bank transfer is completed.

Once the payment is registered, we start picking and packing the order, as registered in a pick-pack document. If the full order is available in stock, there is only one set of pick-pack instructions. However, in case of partial non-availability, several pick-pack documents may be created on one order. The pick-pack document, we call it the PP, is available for our internal warehouse employees. This document contains information on what exactly they need to pick, where it is located, how they need to pack it, and where to store it for the shipping phase. Next, the pick-pack document triggers the shipping, which results in a delivery at the client.This delivery is normally handled by our delivery partner for the customer's country. The customer signs for receipt, which is stored in our information system, through a mobile hand-held device that the drivers use to collect a signature for receipt. We wonder what the time distribution is between picking, packing, and shipment. Actually, everything from picking on, is of particular interest. It's also important to know what the time distribution is between payment and delivery since this will help improve customer satisfaction.

Once a user has placed a successful order, we use that information to send them a monthly update of the new, interesting items in our webshop, based on their browsing history and previous purchases.

To aid understanding of the process, a visual illustration of a typical order-delivery process can be seen in Figure 1."

Description of the data

"Attached, you can find an overview of the data tables, as they appear in our database. On the left we give the names of the columns and on the right a description of the contents of each column. Underneath each table we note the primary and foreign keys."

Figure 1: A visual representation of the online order process

Browsing session

Browsing session	
session_ID	unique session id
IP_address	IP address of visitor
start_date	time the session started
end_date	time the session ended

```
Primary key: session_ID
Foreign key: session_ID REFERENCES Browsed_products(session_ID)
```

Browsed products

Browsed_products	
session_ID	session ID
product_ID	product ID
timespan	time the product was browsed (in seconds)

```
Primary key: session_ID, product_ID, timespan
Foreign key: product_ID REFERENCES Product(product_ID)
Foreign key: session_ID REFERENCES Browsing_session(session_ID)
```

Product

Product	
product_ID	product ID
product_name	name of the product
unit_price	price per unit
currency	currency of the product (USD or GBP or EUR)
supplier	name of the supplier
margin	margin made on the product

`Primary key: product_ID`

Customer

Customer	
customer_ID	customer ID
address	address of the customer
phone_nr	phone nr of the customer
name	first and last name of the customer

`Primary key: customer_ID`

Order

Order	
order_ID	order ID
customer_ID	customer ID
session_ID	session ID
creation_time	time (DD/MM/YYYY hh:mm:ss) the order was created (first item in cart)

`Primary key: order_ID`

`Foreign key: customer_ID REFERENCES Customer(customer_ID)`

`Foreign key: session_ID REFERENCES Browsing_session(session_ID)`

Order line

Order_line	
order_ID	order ID
line_nr	line number
creation_time	moment (DD/MM/YYYY hh:mm:ss) that order line was created
product_ID	product ID
quantity	quantity of the product ordered
invoice_nr	invoice ID of the matching invoice
invoice_line_nr	line number of the matching invoice

```
Primary key: order_ID, line_nr
Foreign key: order_ID REFERENCES Order(order_ID)
Foreign key: product_ID REFERENCES Product(product_ID)
Foreign key: invoice_nr, invoice_line_nr
             REFERENCES Invoice_line(invoice_nr,line_nr))
```

Invoice

Invoice	
invoice_nr	invoice ID
invoice_type	type of invoice: 'In-stock', 'Ordered'
department	department that is responsible for the invoice: 'ACC', 'CCB', 'DDF'
VAT_type	the type of VAT that applies to the invoice: '6', '12', '21'
creation_time	date (DD/MM/YYYY) the invoice was created
posting_time	time (DD/MM/YYYY hh:mm:ss) invoice was posted in the general ledger
clearing_time	time (DD/MM/YYYY hh:mm:ss) invoice was paid
accounting_period	accounting period invoice was posted (MM/YYYY)

```
Primary key: invoice_nr
```

Invoice line

Invoice_line	
invoice_nr	invoice ID
line_nr	the number of the line
order_ID	order ID
order_line_nr	the line of the order to which this invoice line applies
value_VAT_incl	amount incl. VAT
value_VAT_excl	amount excl. VAT

```
Primary key: invoice_nr, line_nr
Foreign key: invoice_nr REFERENCES Invoice(invoice_nr)
Foreign key: order_ID, order_line_nr
             REFERENCES Order_line(order_ID, line_nr)
```

Payment

Payment	
payment_nr	payment ID
invoice_nr	invoice ID
amount	amount that was paid
payment_method	method that was used to make the payment: 'CreditCard', 'Transfer'
payment_date	date (DD/MM/YYYY) the payment was registered

```
Primary key: payment_nr, invoice_nr
Foreign key: invoice_nr REFERENCES Invoice(invoice_nr)
```

Pick-Pack

Pick-Pack	
pp_nr	pick-pack ID
order_ID	order ID
order_line_nrs	the set of order lines to which this PP-doc applies
if_complete	order completely available or not ('0', '1')
activity_start	start moment of the activity (DD/MM/YYYY hh:mm:ss)
activity_end	end moment of the activity (DD/MM/YYYY hh:mm:ss)
activity_type	type of the activity: 'pick', 'pack'
package_type	type of the package that is used: 'box', 'bag', 'other'
location	warehouse location
gate_nr	gate where the package will be picked up for shipping
max_weight	weight restriction for the country that the shipment goes to

`Primary key: pp_nr,order_ID`

`Foreign key: order_ID REFERENCES Order(order_ID)`

Shipment

Shipment	
shipment_ID	shipment ID
pp_nr	pick-pack ID
start_date	start date of the shipment (DD/MM/YYYY)
end_date	end date of the shipment (DD/MM/YYYY)
delivery_address	delivery address of the shipment

`Primary key: shipment_ID`

`Foreign key: pp_nr REFERENCES Pick-Pack(pp_nr)`

Delivery

Delivery	
delivery_ID	delivery ID
shipment_ID	shipment ID
time_stamp	moment of delivery acceptance by customer (DD/MM/YYYY hh:mm:ss)
X-coordinates	GPS coordinates of the delivery
Y-coordinates	GPS coordinates of the delivery

Primary key: delivery_ID

Foreign key: shipment_ID REFERENCES Shipment(shipment_ID)

Follow-up questionnaire

Name:

Group:

Team:

State the current time: __ : __

1. I had sufficient time to read the case and complete the assignment

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

2. This case assignment was very difficult to execute

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

3. Overall, I believe that applying the procedure to build the event log from the data was easy

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

4. Building the event log, given the procedural approach, was difficult for me

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

5. I am confident that I am now competent to build an event log structure for a process mining analysis in practice

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

6. I found the concepts of the procedure to build an event log from a relational database difficult to understand

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

7. I found the procedure to build an event log from a relational database easy to understand

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

8. Understanding the link between the database tables and the event log consequences was frustrating

completely disagree	1	2	3	4	5	6	7	completely agree
	☐	☐	☐	☐	☐	☐	☐	

2 Grading instrument

2.1 Goal of the project and key questions

From the project description (see Section 1) the subjects need to derive that the goal of the process mining project relates to efficiency. The efficiency goal is derived from the following key questions that are formulated in the process description:

- How many times does a visitor create an order (and for how much money) and then abandons the browsing session without payment?
- What percentage of the cases is following our process model like we designed it? Everything from picking on, is of particular interest.
- What is the time distribution between picking, packing and shipment?
- What is the time distribution between payment and delivery?
- Which departments do a better job in collecting the cash in a quickly fashion?

2.2 Grading

Table 1 explains grading instructions, tying the level of understanding shown by the subjects to a certain grade. Table 2 contains examples of good and bad justifications of the event log structure that need to be made by the subjects to show an understanding (or its absence).

Table 1: Explanation of grades

Grade	Explanation
A	The subjects understand the majority of the underlying concepts, both theoretically and technically (like the cardinality of relationships for example) to a level that can be expected from graduate students. Some (minor) mistakes are permissable. They furthermore understand the relations between the concepts they discuss and the potential impact on the quality of the event log.
B	The subjects understand to a certain degree the underlying concepts. However, they are not fully knowledgeable about the true technical functionalities (like for example copying or multiplying information from header to child). They furthermore understand the relations between the concepts and work towards a goal.
C	The subjects lack a clear understanding of some of the underlying concepts. They are not always able to translate concepts to the event log structure. The subjects have no view on the big picture of 'building a qualitative event log'. They make their own assumptions instead of thinking from a stakeholder's perspective.
D	The subjects lack a clear understanding of the event log structure, where process instance, activity (and hence timestamps) and attributes are related to each other. The subjects make their own assumptions instead of thinking from a stakeholder's perspective.

Table 2: Example expressions of good and bad justifications

Good justifications	Bad justifications
The selection of activities is based on the availability of a timestamp: activity "create order" has "creation_time" as timestamp in the order table.	The selection of activities is not based on whether there is a timestamp available in the data tables. For instance selecting the activity "customer places an order" while there is not timestamp available.
The selection of the process instance is based on the project goal: the order is chosen as instance as it allows to study the efficiency of the process.	The selection of the process instance is based on the first interaction with the client: the browsing session is taken as instance although it does not help to answer the key questions (see Section 2.1).
The selection of the process instance level is based on the key questions that need to be solved: the instance at the level of an order will suffice to answer the key questions.	The selection of the process instance level is based on the granularity of the available information: the instance is at the level of order line as there is an "Order_line" table.

3 Statistical tests affecting variables

With respect to the affecting factors that were measured via scale variables, the mean and standard deviation for each variable for the procedure group and the control group can be found in Table 3. To use a t-test to compare the means of these variables, the normality assumption was tested, using the Shapiro-Wilk test. The variables in bold are the ones that comply with the normality assumption in both groups and with the assumption of homogeneity of variances across the two groups (tested by the Brown-Forsythe test). However, the t-tests did not show any statistical differences between the two groups. This possibility was anticipated, given the small sample size. The statistical power of the t-tests varied between 0.051 and 0.167, which clearly indicates the low probability of the tests to capture real effects.

Table 4: Results Shapiro-Wilk test

Variable	Statistic procedure group	P-value procedure group	Statistic control group	P-value control group
Academic achievement	0.915	0.392	0.913	0.378
PEOU Ass. 1	0.833	0.036	0.752	0.004
PEOU Ass. 2	0.82	0.025	0.744	0.003
PEOU Proc. 1	0.802	0.015	0.731	0.002
PEOU Proc. 2	0.868	0.095	0.858	0.073
PEOU Proc. 3	0.647	<0.001	0.866	0.089
PEOU Proc. 4	0.908	0.268	0.907	0.263
PEOU Proc. 5	0.85	0.058	0.819	0.025
PEOU Proc. 6	0.935	0.494	0.904	0.245
S EFF 1	0.935	0.499	0.809	0.019
S EFF 2	0.929	0.439	0.883	0.141
S EFF 3	0.817	0.023	0.902	0.232
S EFF 4	0.939	0.542	0.893	0.183
ENV Ass. 1	0.886	0.153	0.647	<0.001
ENV Ass. 2	0.92	0.359	0.909	0.276

Academic achievement based on a max score of 100
PEOU: perceived ease of understanding of assignment (Ass.) or of procedure (Proc.)
S EFF: self-efficacy
ENV: environment when executing the assignment (Ass.)
Duration Ass: duration of event log building (in mins)

Table 3: Descriptive statistics of affecting variables

Variable	Procedure group		C ontrol group	
	Mean	St.d.	Mean	St.d.
PEOU Ass. 1	6.1	(0.738)	6.1	(0.568)
PEOU Ass. 2	5.2	(0.789)	5.1	(1.287)
PEOU Proc. 1	3.3	(0.675)	3.1	(1.37)
PEOU Proc. 2	4.5	(0.972)	5.2	(1.317)
PEOU Proc. 3	3.6	(1.075)	3.8	(1.398)
PEOU Proc. 4	3.6	(1.43)	4.2	(1.398)
PEOU Proc. 5	4.5	(1.179)	3.5	(1.581)
PEOU Proc. 6	4	(1.633)	4.6	(0.966)
S EFF 1	69.7	(13.817)	62.8	(7.642)
S EFF 2	63.9	(10.847)	63.5	(11.797)
S EFF 3	65	(14.337)	65.5	(11.414)
S EFF 4	61	(15.776)	56.5	(13.344)
ENV Ass. 1	4.2	(1.989)	6.4	(1.075)
ENV Ass. 2	4.9	(1.101)	5.5	(1.354)

Table 5: Results Brown-Forsythe test

Variable	Statistic	P-value
Academic achievement	1.135	0.305
S EFF 2	0.024	0.88
S EFF 4	0.09	0.768
ENV Ass. 2	1.371	0.257
PEOU Proc. 2	0.643	0.433
PEOU Proc. 4	0	1
PEOU Proc. 6	1.231	0.282

Academic achievement based on a max score of 100
PEOU: perceived ease of understanding of assignment (Ass.) or of procedure (Proc.)
S EFF: self-efficacy
ENV: environment when executing the assignment (Ass.)
Duration Ass: duration of event log building (in mins)

Table 6: T-test of affecting factors

Variable	Statistic	P-value	Confidence interval	Power
Academic achievement	0.735	0.475	[-6.476 ; 13.226]	0.089
S EFF 2	0.079	0.938	[-10.247 ; 11.047]	0.051
S EFF 4	0.689	0.5	[-9.228 ; 18.228]	0.091
ENV Ass 2	-1.087	0.291	[-1.759 ; 0.559]	0.125
PEOU Proc. 2	-1.353	0.193	[-1.787 ; 0.387]	0.167
PEOU Proc. 4	-0.949	0.355	[-1.929 ; 0.729]	0.105
PEOU Proc. 6	-1	0.331	[-1.861 ; 0.661]	0.102

Academic achievement based on a max score of 100
PEOU: perceived ease of understanding of assignment (Ass.) or of procedure (Proc.)
S EFF: self-efficacy
ENV: environment when executing the assignment (Ass.)
Duration Ass: duration of event log building (in mins)

Table 7: Mann-Whitney-Wilcoxon test of control variables

Variable	Statistic	P-value
PEOU Ass. 1	49.5	1
PEOU Ass. 2	53	0.839

PEOU Ass.: perceived ease of understanding of assignment

4 Details of discussed concepts per procedure step

Table 8: Frequency of concepts discussed during 'Step 1: State goal'

Concept	Frequency	
	Procedure	Control
Project Goal	3	4
Cornerstone	0	1
Nice to have project goal	4	1
Must have project goal	4	1
Efficiency goal	4	6
Compliance goal	4	5
Key questions (to articulate goal)	5	4
Process instance	1	0

Table 9: Frequency of concepts discussed during 'Step 2 - Identify process cornerstones'

Concept	Frequency	
	Procedure	Control
Activity	1	0
Cornerstone	5	1
Efficiency goal	0	1
Key questions (to articulate goal)	0	1
Process instance	1	0

Table 10: Frequency of concepts discussed during 'Step 3 - Identify key tables'

Concept	Frequency	
	Procedure	Control
Cardinality of tables	1	0
Timestamp	3	0
Cornerstone	3	0
Document	1	0
Key questions (to articulate goal)	1	0

Table 11: Frequency of concepts discussed during 'Step 4 - Identify relationships (cardinality) between tables'

Concept	Frequency	
	Procedure	Control
Cardinality of tables	5	5
Timestamp	2	0
Attribute	1	0
Artificial multiplication	0	1
Key questions (to articulate goal)	2	0

Table 12: Frequency of concepts discussed during 'Step 5- Select process instance document'

Concept	Frequency	
	Procedure	Control
Project Goal	2	0
Cardinality of tables	5	3
Self-loops	1	0
Artificial multiplication	2	0
Document	1	1
Must have project goal	2	0
Efficiency goal	5	1
Compliance goal	3	0
Key questions (to articulate goal)	5	0
Process instance	5	3
Start document	1	4
End document	1	1

Table 13: Frequency of concepts discussed during 'Step 6 - Select process instance level'

Concept	Frequency	
	Procedure	Control
Project Goal	0	3
Cardinality of tables	3	1
Activity	2	0
Cornerstone	0	1
Attribute	1	0
Self-loops	1	0
Artificial multiplication	3	2
Key questions (to articulate goal)	3	0
Process instance	2	8

Table 14: Frequency of concepts discussed during 'Step 7 - Select activities'

Concept	Frequency	
	Procedure	Control
Project Goal	1	0
Cardinality of tables	2	4
Timestamp	5	11
Activity	6	12
Cornerstone	4	1
Attribute	2	3
Self-loops	0	1
Artificial multiplication	3	0
Nice to have project goal	1	0
Document	0	1
Attribute-dependent activity	3	1
Compliance goal	0	1
Key questions (to articulate goal)	1	7
Process instance	3	3

Table 15: Frequency of concepts discussed during 'Step 8 - Select attributes'

Concept	Frequency	
	Procedure	Control
Project Goal	0	2
Cardinality of tables	0	1
Timestamp	0	3
Attribute	4	11
Case attribute	5	0
Event attribute	5	0
Attribute aggregation	2	1
Key questions (to articulate goal)	2	1
Process instance	0	1

Table 16: Frequency of concepts discussed during 'Step 9 - Consider attributes to incorporate in activities'

Concept	Frequency	
	Procedure	Control
Activity	0	1
Attribute	0	2