

Instruction 5 – Assignment 5

IMPORTANT!!!

The answers for this assignment should be:

1. Printed and delivered at the room Pav D3 (see box "Process Mining – Assignment 5")
2. Posted at the Study web in the folder "1BM45_Process mining -> assignment 5 -> Submitted"

Deadline: 14/03/2008 - 06:00 p.m. (**strict** deadline for both steps above!)

Table of Contents

1. Assignment	2
2. Input Material	4
Process Description	4
Data Location	4
3. Structure of Report	5

1. Assignment

This assignment aims at assessing your problem-solving skills when using data mining and process mining techniques to analyze process executions. To do so, we have created a simulated log which mimics a real-life process. Your analysis should cover the following points:

- What are the five most frequent paths for this process? How much of the log do they account for? What are their average throughput times?
- How does the process model that describes the behavior in the log look like? Does this model completely fit the log? If not, how many instances fit this model and how many do not? Where are the problems for the non-fitting process instances?
- What are the roles in the organization? Which employees can perform both immediate and internal repairs? Who is handing over work to whom? Who are the central employees for this process?
- Is the rule "*immediate repairs that could not be solved should be handled by an internal team again before being sent to an external team*" being indeed obeyed? Which percentage of process instances complies with this rule? Which percentage does not comply with this rule?
- What kind of repair is more common in the organization? Which percentage of repairs can be fixed in a first attempt?
- Where are the bottlenecks in the system? Which bottlenecks are due to long waiting times and which are due to long execution times? Which three tasks have the longest average Sojourn times?
- What are the business rules that usually apply for the moments of choice in the model? (Note: At least one rule should be reported!)
- The technical employees responsible for repairing *heating* systems (i.e. the plumbers) have been experiencing many problems with these systems. More specifically, they would like to avoid that more problems happen during winter time because it is very unpleasant for the tenants and the repair becomes very urgent (and, therefore, expensive). Thus, it would be nice if one could perform a preventive check for the most relevant heating systems (i.e., the ones that more likely will have problems soon). By combining data from different databases of the rental housing organization, it was possible to generate data about the heating system and boilers in the houses (about 5400) of the organization in combination with information about performed repairs (i.e. Building (B), Plumber (P) or

Electronic (E)). This data has been combined in the Excel file *RentalHousing.xls*. Note that, in this file, a value 1 under B means that one or more building (B) repairs are performed for that building.

Given this scenario, your task here is to come up with recommendations on which group of houses should undergo a preventive inspection of the heating system. You should clearly motivate your recommendations.

The results of your analysis should be described in a *report* that has managers as the target audience. Since managers are usually very busy, you should use a concise and clear language when reporting and discussing your results. Section 3 contains more details on how this report should look like. Additionally, to give you a bit more context about the process, Section 2 includes a general process description. This section also contains pointers to the data to be used as input during your analysis.

2. Input Material

Process Description

The running example is inspired on processes of the Dutch rental housing organizations. These organizations rent houses at cheaper prices than in the private sector. They have many processes, like registration, personal information update, complaints handling, etc. The data to be analyzed in this assignment focuses on the *process to handle requests (or complaints) for house repairs*. The process starts when a tenant contacts the company to file a complaint. If the complaint indeed involves a repair in the house, a ticket is created and an appointment is made to inspect the house such that the actual problem can be detected/confirmed. Additionally, the inspector estimates how much time will be needed to fix the problem. Easy fixes are usually performed together with the inspection. More complicated fixes require a new appointment and can be performed by an internal or external team. When the repair is ready, the client is informed and the ticket number is communicated to the financial administration so that they can take care of the payment to the appropriate institutions. Once the payment is in place, the complaint is archived.

Data Location

The event log and the Excel sheet with the data to be used by data mining or process mining techniques are located at StudyWeb, folder "Handouts" (See e-mail with subject Assignment 5). The material contains two files:

1. EventLogAssignment5.mxml
2. RentalHousing2007.xls

3. Structure of Report

The answers for this assignment should be given in a report. The reason for doing so is that these documents are always produced when showing managers the results of process mining analysis in real-life situations. This way they can communicate the results for other levels in the hierarchy. Therefore, to guide you in producing this report, we have defined an outline with the points your report should include. This outline is as follows:

Outline

1. **Cover page**

Size: 1 page

<<Include the title of your report, your name and student number, place and date>>

2. **Introduction**

Size: At most 2 pages

<<The introduction should clearly state the aim of this report. As the target audience are managers of rental housing companies, the introduction should indicate what these managers can find in the report in an appealing way (so that the managers feel like reading the rest of the document).>>

3. **Analysis Results**

Size: At most 1 page of text per question answered (Note that this page limit does not include figures!)

<<This section should present the results of your analysis. The idea is that you create subsections that address one or more of the points outlined in Section 1.

For every point, you should include:

- a) The questions addressed;*
- b) Your answer for these questions;*
- c) Screenshots of mined models/results that support your answer >>*

4. **Conclusion**

Size: At most 2 pages

<<This section should summarize your main findings and suggestions on how to improve the process that you have analyzed>>