**IFN650** Business Process Analytics

Assignment 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Key Information  |  |  | | --- | --- | | Type | Group Assignment  (3-4 students) | | Period | 2023/SEM-1 | | Due Date | **Wednesday 31 May 11:59** | |  | Deliverables Required  |  | | --- | |  | | **Professional Written Report: 50%** (incl.15% individual marks) | |  | |

## Background/Overview

In this assignment, you are required to demonstrate your understanding of process mining by using different tools and techniques to analyse execution data. In answering specific questions about the data, you will apply various process mining techniques in order to interpret event log(s) and draw meaningful evidence-based insights from the data.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Part A** | |  | **Part B** | |  | **Part C** | |
| **Disco** | **15%** |  | **ProM** | **20%** |  | **Exploratory Analysis**  **(Multiple Process Mining Tools)** | **15%** |
| Process Discovery | 5% |  | Process Discovery 1&2 | 10% |  | Improvement Recommendations | 15% |
| Process Comparison | 5% |  |  |
| Filtering | 5% |  | Process Conformance & Performance | 10% |  |  |  |

## Deliverables

### Written Report

You are required to submit a written report that answers the assignment questions together with screenshots and explanations to illustrate the process mining results. Key findings/insights should be discussed in depth.

* Your report should be a **maximum of 15 pages** of content for Parts A and B together with a **maximum of 5 pages** of content for Part C for each member.

### Part A: Disco Analysis (15%)

You are to use an event log ‘BPI\_Challege\_2017.xes’. It is a real-life event log publicly available for process mining analysis. Unless otherwise specified, the complete, unfiltered, original log should be used to answer each of the three questions using the Disco software.

1. Analyse and interpret the following **process models** generated by Disco.
   1. Compare and contrast two process models (maps) – one generated using the setting: 100% activity and 100% paths and the other generated using the setting 50% activity and 50% paths. Explain why we should not use a model with 0% paths to understand process behaviours.
   2. Investigate the case variants detected from the log. Overall, how many case variants are present in the log? Report on the top five (5) most frequent case variants and their respective frequencies. How much of the log do these five (5) case variants cover? Investigate the variants with low frequencies. How many variants only have less than 10 cases? Explain the implications of a high number of case variants when you try to generate a representative process model.
2. **Compare** their process behaviours (i.e., Are the two process models quite similar or very different? Observe the activities/paths/rework loops etc) and performance (i.e., throughput times, # cases, bottlenecks) of two groups of cases. Describe your observations.
   1. Group A: Cases for Business loans (i.e., attribute LoanGoal – value Business goal).
   2. Group B: Cases for Home Improvements.
3. **Sequentially** apply the following **filters** to the original log.

1) Only keep cases whose behaviour is shared by at least 50 cases;

2) Filter out (discard) all cases that were not started and completed between 1 April 2016 and 31 Dec 2016;

3) Only keep the cases with “W\_call after offers” being the last event.

Show the overview screen with the statistics of the filtered log.

Using the **filtered** log, answer the questions:

1. How many cases are there in the log?
2. What is their mean duration?
3. How many variants are there in the log?
4. How significant is the problem of rework for this process?
5. Are there any bottlenecks detected? If so, which activities/paths are involved?

### Part B: ProM (20%)

Use the event log ‘RequestForPayment.xes’ for Part B. This log is provided as part of BPI Challenge 2020[[1]](#footnote-1). The data comes from the travel reimbursement process at TU/e. This log contains Requests for Payment (should not be travel related): 6,886 cases, 36,796 events.

Please undertake an exploratory analysis of the log first. You can filter the original log if you wish when **answering each of the four questions using the ProM framework** (all required plug-ins are available in ProM Lite).

1. Discover the process models using the **Alpha** **Miner** and the **Inductive Miner** (Petri Net) algorithms (90% paths for inductive miner). Compare the similarities and differences between the two Petri-nets discovered.
2. Discuss the different visualisations (process tree, BPMN) available from the **Inductive Miner** algorithm (process tree). Write down a brief process description based on the discovered BPMN model.
3. Using the Petri nets models discovered by the Inductive Miner algorithm (at 90% paths), replay the log (‘**Replay a Log on Petri Net for Conformance Analysis’** plug-in). Explain how well the models describe the process behaviour seen in the log, including a discussion of the following trace fitness metrics.
   1. Does the model completely fit the log?
   2. If not, how many cases fit the models and how many do not?
   3. Where are the problems for the non-fitting process cases?
4. Analyse the process using the **Inductive visual Miner** and identify potential bottlenecks and deviations. Note: Inductive visual Miner is also made available as part of QuickVisualiser: <http://leemans.ch/leemansCH/quickvisualiser/>

### Part C: BPI Challenge 2011 Log Exploratory Analysis (15%) [Individual Task]

Use ‘Hospital\_log.xes’ for Part C. This is a real-life log, taken from a Dutch Academic Hospital. This log contains 46560 events in 824 cases. Apart from some anonymisation, the log contains all data as it came from the Hospital's systems. Each case is a patient of a Gynaecology department. The log contains information about when certain activities took place, which group performed the activity and so on. Many attributes have been recorded that are relevant to the process. Some attributes are repeated more than once for a patient, indicating that this patient went through different (maybe overlapping) phases, where a phase consists of the combination of Diagnosis & Treatment.

van Dongen, Boudewijn (2011): Real-life event logs - Hospital log. Version 1. 4TU.ResearchData. dataset. <https://doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54>

**[Individual Task]** Imagine that you are asked to present **one data-informed process improvement recommendation** to stakeholders. Your task is to extensively analyse this process (guided by the stakeholder questions) and **provide one improvement recommendation based on these analysis insights**. Please note that this is quite a complex process. You can filter the log as you see fit. You must explain your assumptions, and filtering rules and **justify your recommendation** using the process mining results (including screenshots).

You are encouraged to investigate multiple tools (e.g., **Disco, Celonis, ProM Lite, Quick Visualiser, Apromore**, etc) to derive your insights**.**

## Group Work

The assessment tasks are to be completed in a group of **three or four students**. Each student must individually register in a group on Canvas. All group members are expected to contribute equally to the assignment deliverables (Parts A and B) and conduct Part C individually.

## Submission

Your report must be submitted online, using the link on Canvas.

* Your team should submit one assignment, uploaded by any one member of your team.

### Format of Submission

You must submit a single .pdf file (named “**IFN650\_Assigment-2\_Group-XX.pdf**”), containing part A, B and the individual parts C of each member. The formatting of the whole report must be consistent.

### Late Assessment Policy

## Assignments submitted without an approved extension will not be marked and will receive a grade of 1 or 0%. You can [apply for an automatically approved 48-hour extension](https://qut.to/late-assessment), or if special circumstances prevent you from meeting the assignment due date, you can apply for a longer extension. If you don't have an approved extension you should submit the work you have done by the due date and it will be marked against the assessment criteria.

1. <https://doi.org/10.4121/uuid:52fb97d4-4588-43c9-9d04-3604d4613b51> [↑](#footnote-ref-1)