

Chapter 4

towards the process mining of the computer science department

1.Introduction:

In the previous chapters we discussed Process mining in general, Now let us practice process mining with real data.

In this chapter we will cover, handling and exploration Event data using the process mining applications (Process discovery and conformance checking) . and by the end of this chapter we can improve the process, overcome the challenges in M'sila's university and we will be able to:

1. Explain how process mining can help in analyzing and improving computer science department processes
2. Identify opportunities for process mining in faculty of computer science department
3. Describe the data requirements in order to apply process mining
4. Interpret the results of various process mining techniques in the ProM lite tool
5. Apply ProM lite on real data to obtain process mining results .

2.study case :

process mining has been applied to a wide variety of organizations, which include municipalities, government agencies, banks, hospitals, multinationals, high-tech system manufacturers and their customers and media companies. In this case study, we will illustrate the use of process mining techniques for the analysis of the processes that daily take place in the computer science department as one of the departments of university Mohamed boudiaf in m'sila .

3.Mohamed BOUDIAF university :

The University of M'Sila - Mohamed Boudiaf is an Algerian public university located in M'Sila in Algeria. Its named after Mohamed Boudiaf began in 1985 with the opening of an institute of higher education in mechanics, followed in 1989 by those of civil engineering and management of urban techniques. In 1992, it became a university center and, in 2001, it became a university with four faculties and 23 departments. According to the statistics of the university for the year 2013, the number of faculties and institutes increased to 7 and 2 respectively. However, the number of research laboratories within the institution that is accredited by the Ministry of Higher Education and Scientific Research has increased from 7

to 23 laboratories. The University of M'Sila can accommodate 36,217 students and has 10,000 beds in the university residence.

4. Computer science department :



computer science department was created in September 1997. It proceeded its activities by the formation of the 1st promotion D.E.U.A in computer of management and system.

-In 1999 Registration of the first class of engineers, option information system.

-In 2003 Registration of the first postgraduate training promotion, industrial computer engineering option

- In 2007, the IT department adopted for the first time the new L.M.D pedagogical system by registering 69 students in the 2nd year of the Bachelor's degree program, an academic computer science specialty.

- The perspective of the IT department is unlimited including:
 - * Opening of other L.M.D formations
 - * Opening of the Master, organization of seminars, days, scientific colloquia, in order to promote the skills of all teachers.



5. Process mining applications :

process mining has been applied in numerous case studies ,In general there are four types of questions that a process mining project aims to answer:

- 1) What happened? Identifying the need to discover the executed process and its activities (e.g. what is the typical working day of a surgeon?).
- 2) Why did it happen? Understanding the activities and circumstances characterizing the situation or action (e.g. what caused the long waiting list?).
- 3) What will happen? Identifying the circumstances of when or how a specific activity will take place (e.g. is this patient likely to deviate from the normal treatment plan?).
- 4) What is the best that can happen? Identifying possible steps towards specific improvements (e.g. which check should be undertaken first to reduce flow time?).

Within process mining there are four perspectives that can be taken: control flow, performance, conformance, and organizational. The control flow perspective is most popular, with 60% of the papers addressing this perspective. This is not surprising since control flow discovery is usually one of the first steps taken in a process mining project (after event data exploration). Conformance checking (21%), and performance analysis (14%) are also applied regularly. The organizational perspective is addressed in 12% of the case studies.

Unsurprisingly, since most papers are authored by researchers, the most frequently used tool is ProM, which we also use throughout this study case . Other frequently used process mining tools are Disco and Celonis. Custom made tools and techniques are also often used.

It is however interesting to see which plug-ins are used in the case studies, as shown in the table below. The heuristics miner seems to be the most popular process discovery algorithm, followed by the fuzzy miner. The Alpha miner is also used a couple of times. However, recently a new process discovery technique has been developed, the inductive (visual) miner. This algorithm is better than the process discovery algorithms mentioned previously in several aspects.

Technique usage	%
Heuristic Miner	26
Fuzzy Miner	20
Trace Cclustering	11
Conformance Checker	8
Performance	8
Preformance Sequence Analyzer	7
Ontologis	5
LTL Chker	4
Markov Models	4
Trace Aligment with Tree Guide	4

table 4.1 : the most popular algorithms using in process mining

6.How to run a process mining project :

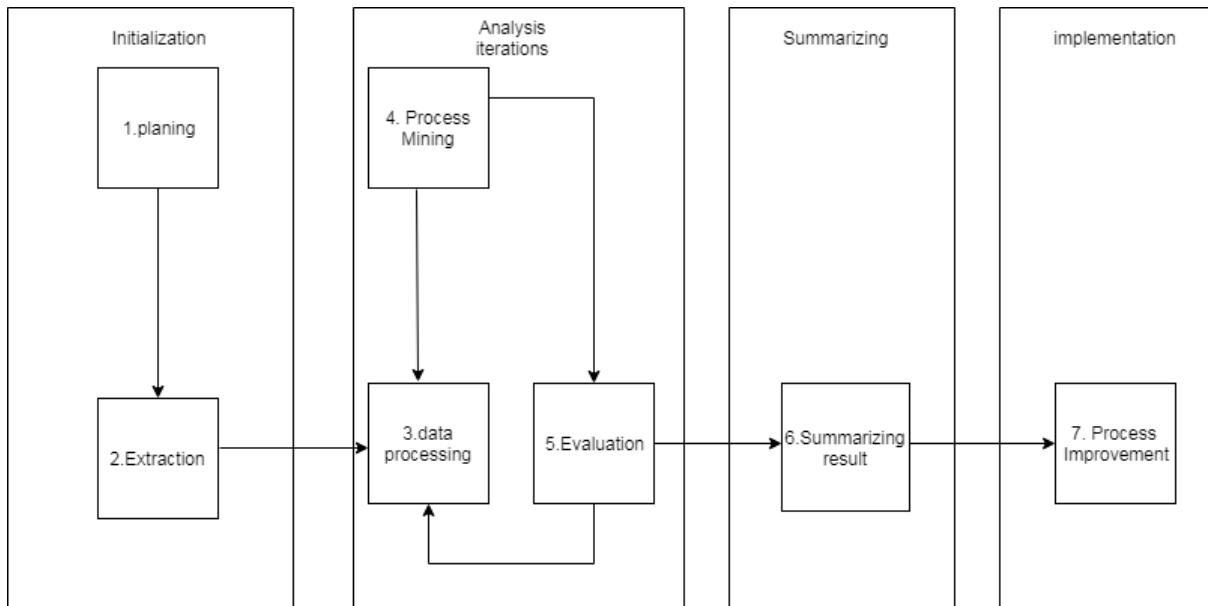


fig 4.1 the steps for run a process mining project

6.1 Planning:

In this phase the scope of the analysis is determined, as well as the questions to be answered. i follow the KISS (Keep It Simple Stupid) principle:

i start simple with a standard process without mixed diseases and with a clear case definition.

Also i select a process for which it is clear that the data is available and how it can be obtained , as you cannot perform process mining without the required data.

6.2Data extraction :

Starting with the questions and goal defined in step 1, the search for data starts. Whether a question can be answered depends on the available data, and its quality. Also consider which case notion to use, and which activities and timeframe to select (for instance, does extracting data from 5 years back make sense? What would you do tomorrow with the obtained insights?).

6.3Data processing :

The extracted data is unlikely to already be in a format ready for process mining analysis. Furthermore, depending on the data and questions defined, different views on the data might need to be taken. Furthermore, events might not be at the correct level of abstraction or detail, or other data sources and attributes could be added. Finally, filtering is important to remove those cases and events that are not of interest for this iteration.

6.4 process mining :

Once the data is converted and filtered, it is ready for process mining to be applied to answer each of the questions. In this phase the relevant plug-ins are selected and correctly configured. Different types of analysis can be performed: explorative (e.g. through visualizations), process discovery, conformance checking, or even process enhancement. Which specific plug-in to use depends on experience, which is hard to quantify.

6.5 Evaluate :

This is arguably the most important phase, as the results obtained in the previous step need to be interpreted and key findings and improvement suggestions need to be provided. Involvement of domain experts and/or stakeholders is therefore crucial. The results might also trigger new questions to be answered.

6.6 Summarizing result

The results obtained so far need to be summarized and reported (either formally in a written report, in a presentation, or in a brief discussion with the problem owner). It is however important that the report describes the steps taken (e.g. what are the questions, which data is used), such that the analysis is repeatable and verifiable. At the same time the reporting should not become too technical since the problem owner is mainly interested in the answers.

6.7 Process improvement

The insights obtained so far need to be implemented in practice, in order to improve the running processes. If this step is skipped, the analysis has no impact on the day-to-day business which is a waste of time, money, and effort. The actual implementation of process improvement however requires different expertises, and is often a separate project. Therefore make sure that the end products (i.e. reports) of the process mining project are clear and provide a solid base for the process improvement project.

7 . study case (registration service) :

The University of M'Sila belongs to the network of the major research universities of Algeria . It offers its students study programs (streams) in all major areas of knowledge and at all levels of diploma

The university that bears the name of its city m'sila, it contains a lot of services , in this project we chose the registration service as a case study to apply the process mining .

8 Event data extraction :

In this activity, we deal with the most common issues in collecting data from the real world. We will present some of the most common standards and show some pitfalls that could cause mistakes.

8.1 some problems to get the data set of computer science department :

this step takes a lot of the project time, it takes 60% of the project time and this represents a set of problems and the way I dealt with them was as follow :

8.1.1. access to data :

one of the most popular problems for a data scientist in general and especially for me as a process miner is accessing to the Ministry of Higher Education and Scientific Research's database to get the Registered students of computer science in University Academic Season (2017/2018) .solution:

we started to record the pathway of students from the day of receiving the inscription document but the question remains how to get exactly the time of receiving . Fortunately, the new smartphones and the computers allow recording the date and the time exactly when the student downloaded the inscription document and so on we decide to ask a small number of new student (100) to give us exactly the real-time of downloading the document which was registered on their phones/ computers.

8.1.1. Pen and copybook :

all of the offices in the department have their Work based on pen and copybook for registering the students without even using the pc so use the pen a lot of time to convert the data of students from papers to excel file.

8.1.3- tickets problems :

one of the steps that contain a lot of problems such as :

- some students buy the ticket and manage to lose it so they have to buy new ones. In the end, we found that some students with 3 tickets in 3 different times.

- the staff register the students without recording their specialties

solution:

- we were obliged to use excel to compare the whole list of all students who bought a ticket with the list that we got from computer science department.

- for students who have two tickets or more we chose the ticket with the most recent date.

finally we Applied the solutions, and we have retrieved the data, and built the CSV files that contain 100 student.10

9-Converting a CSV file to an event log

In this section we convert a CSV file or flat delimit file, to an event log, Since event logs are usually not ready made, we have to convert, for instance, CSV files, or other files, to an event log. And the easiest way to do this is through a CSV file that we got from the previous step (data extraction) , using the prom plugin to do that .

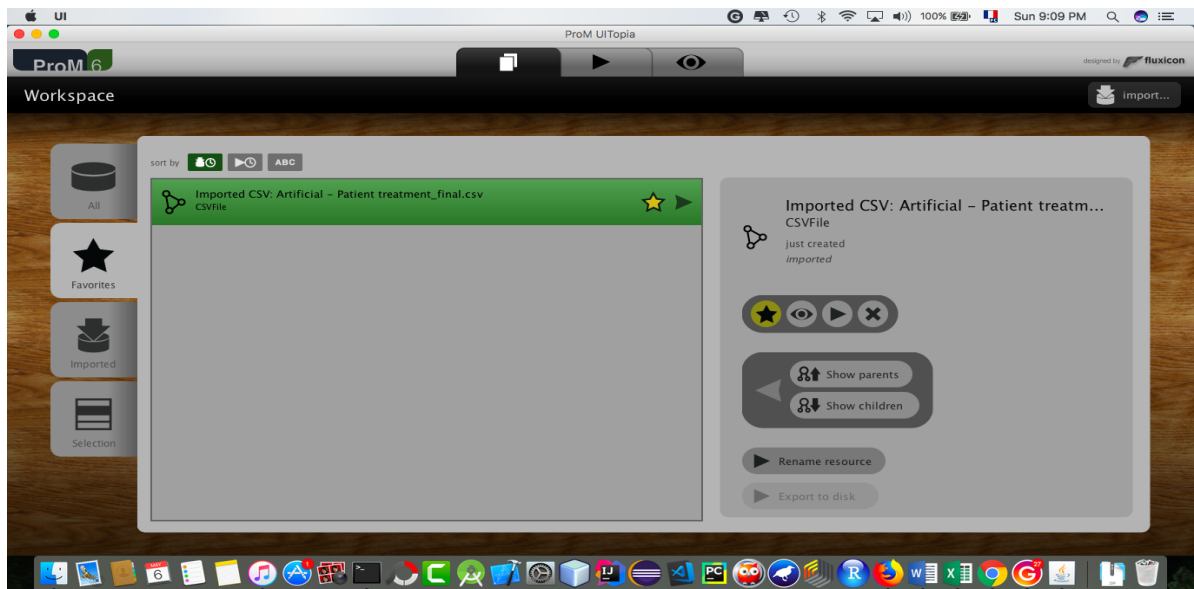


fig 4.2 how to import csv

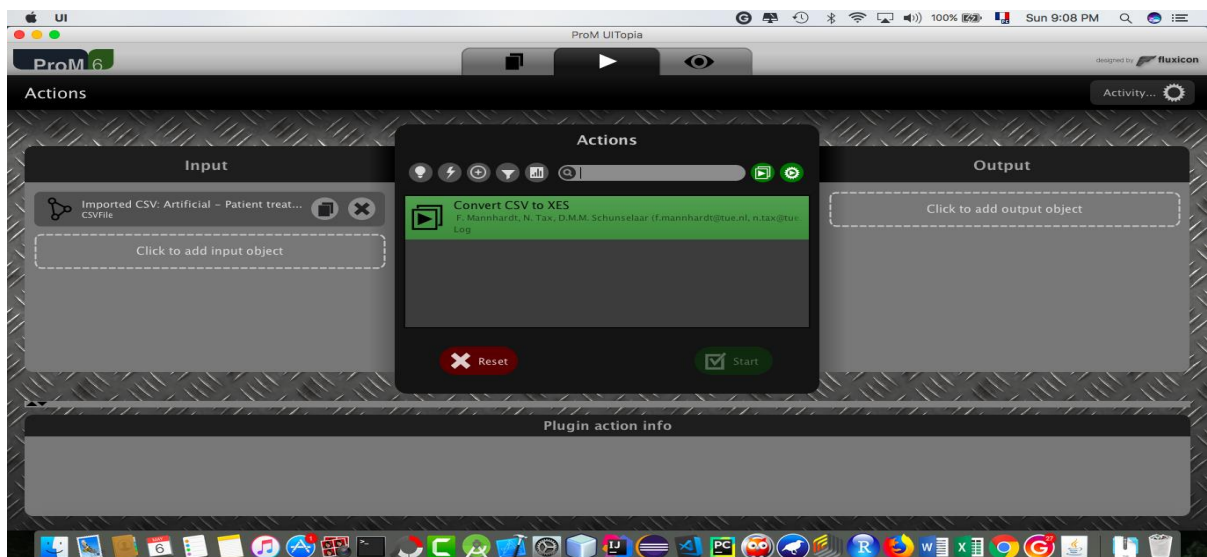


fig 4.3 how to convert csv

10 filter log :

Filtering is important to be able to discover high quality process models later on. It also allows us to focus on a particular part of the process, or only a particular type of case. and since every real life event log will contain some noise or erroneous data, so we wanted to remove that before doing the real analysis. And prom lite provides a plug in that we used to filter our event log

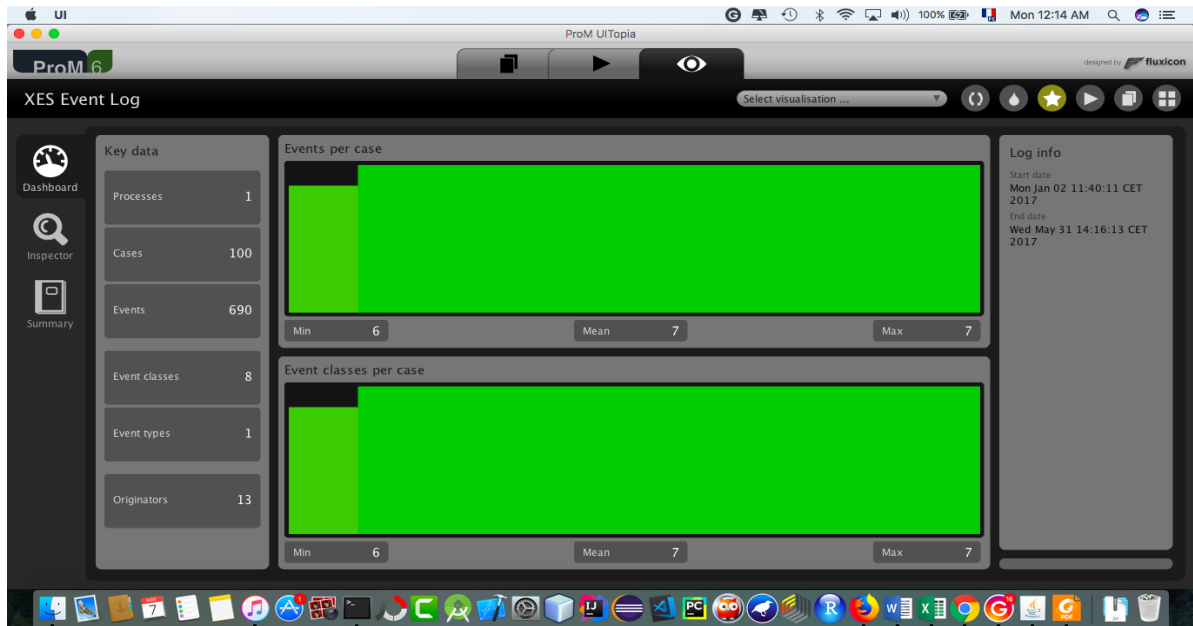


fig 4.4 event log before filter with simple heuristic

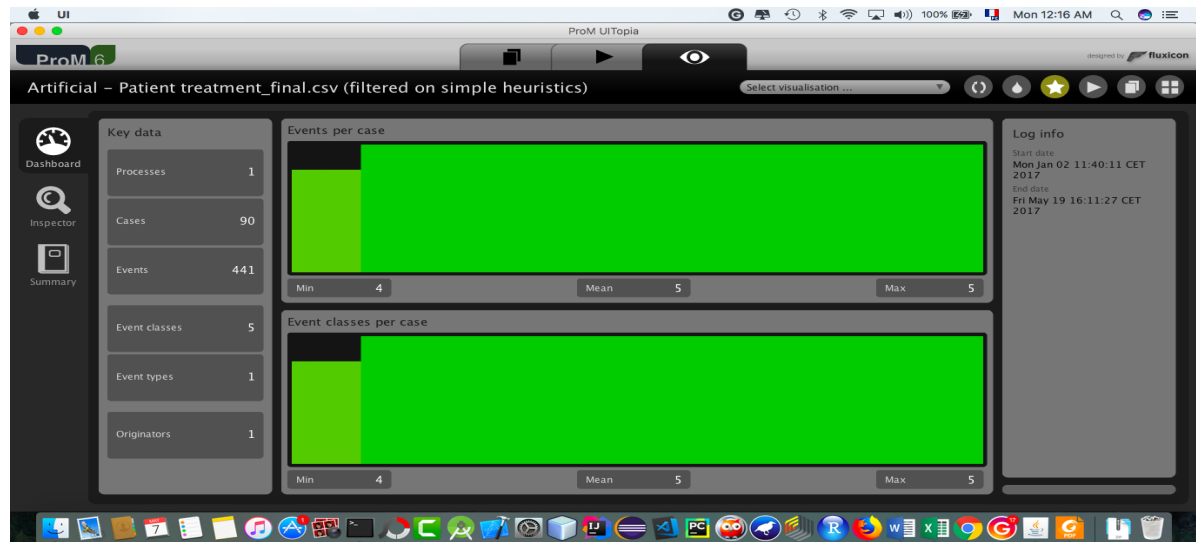


fig 4.5 event log after filter with simple heuristic

11-Questions Answered Based on an Event Log Only

Let us take a look at it, images are worth more than words:

11.1 with the dashboard (log dialog)

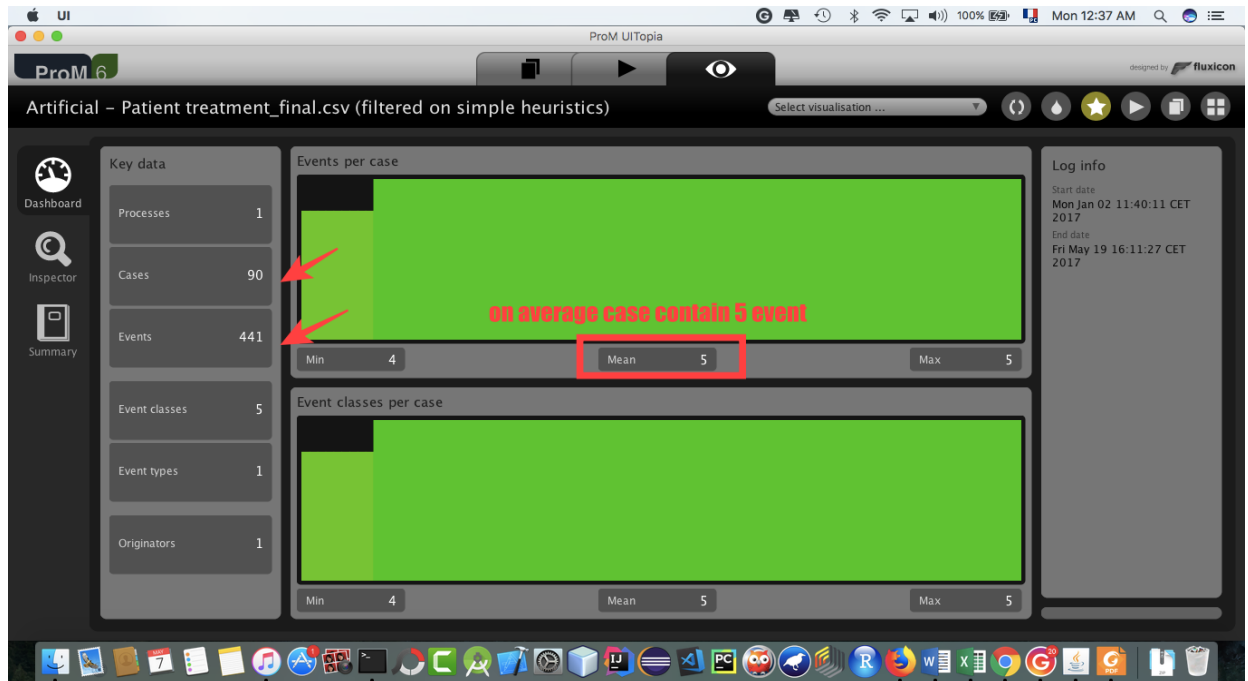


fig 4.6 the dashboard of PROMLite 1.1

11.2 with the dashboard (dotted chart)

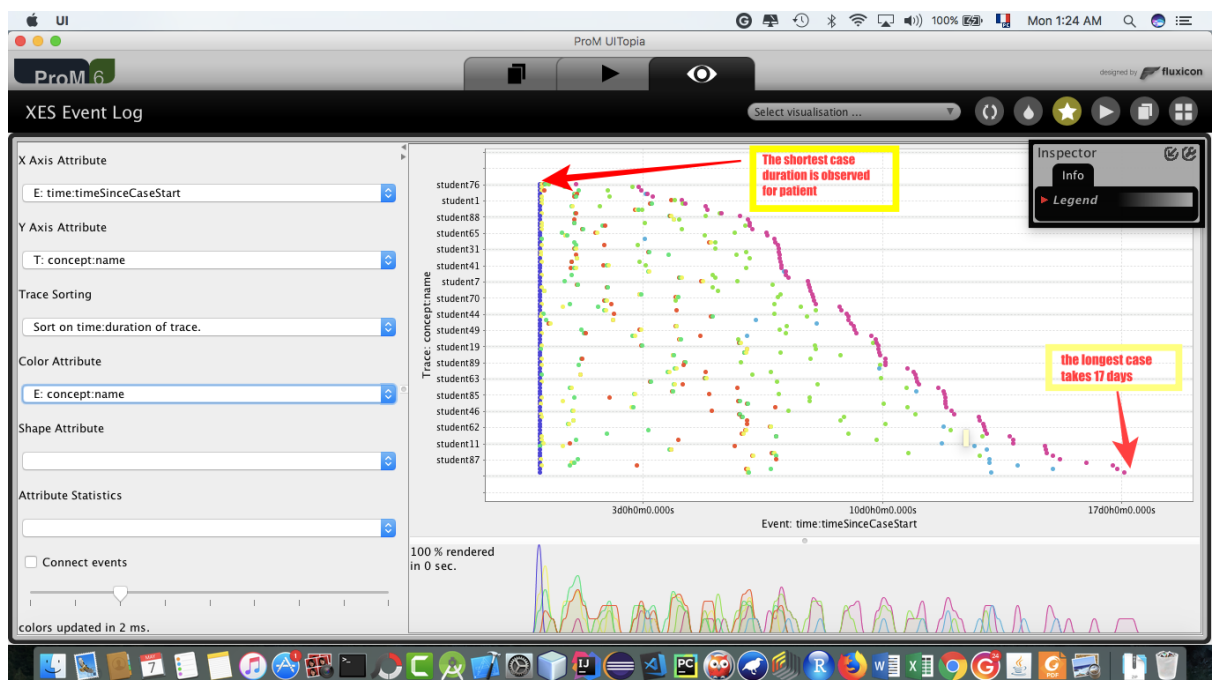


fig 4.7 the dotted chart in PROMLite 1.1

11.3 with the dashboard (export event log) :

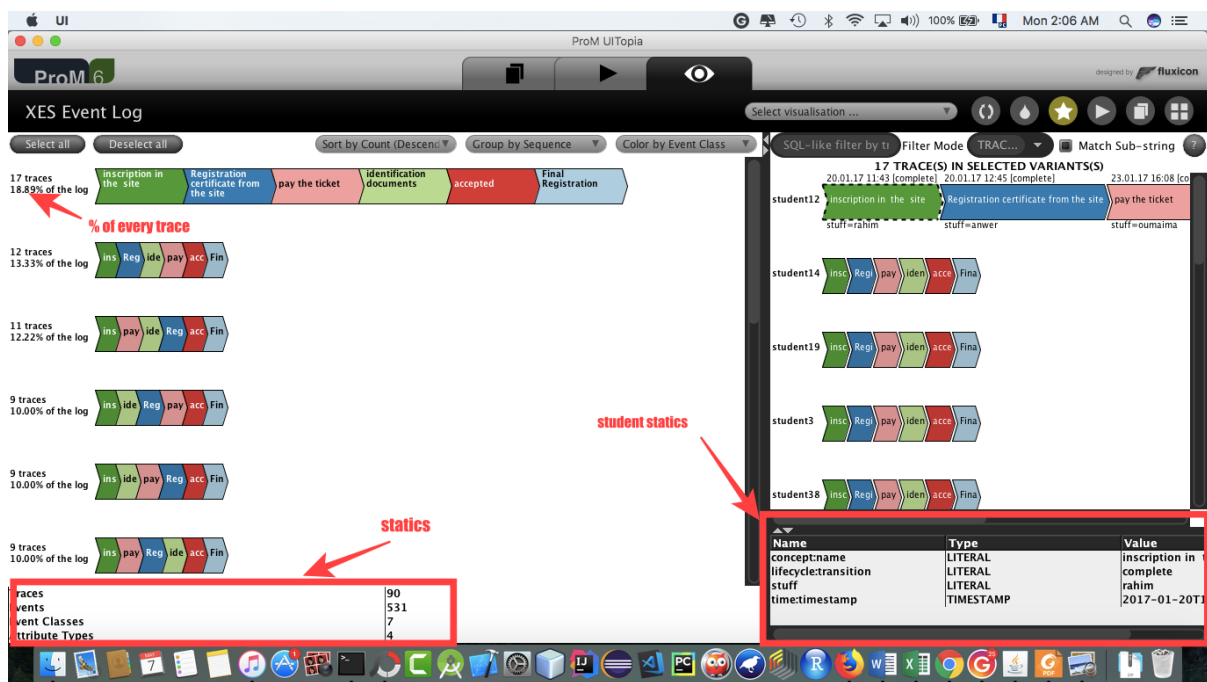


fig 4.8 the explorer of event log in PROMlite

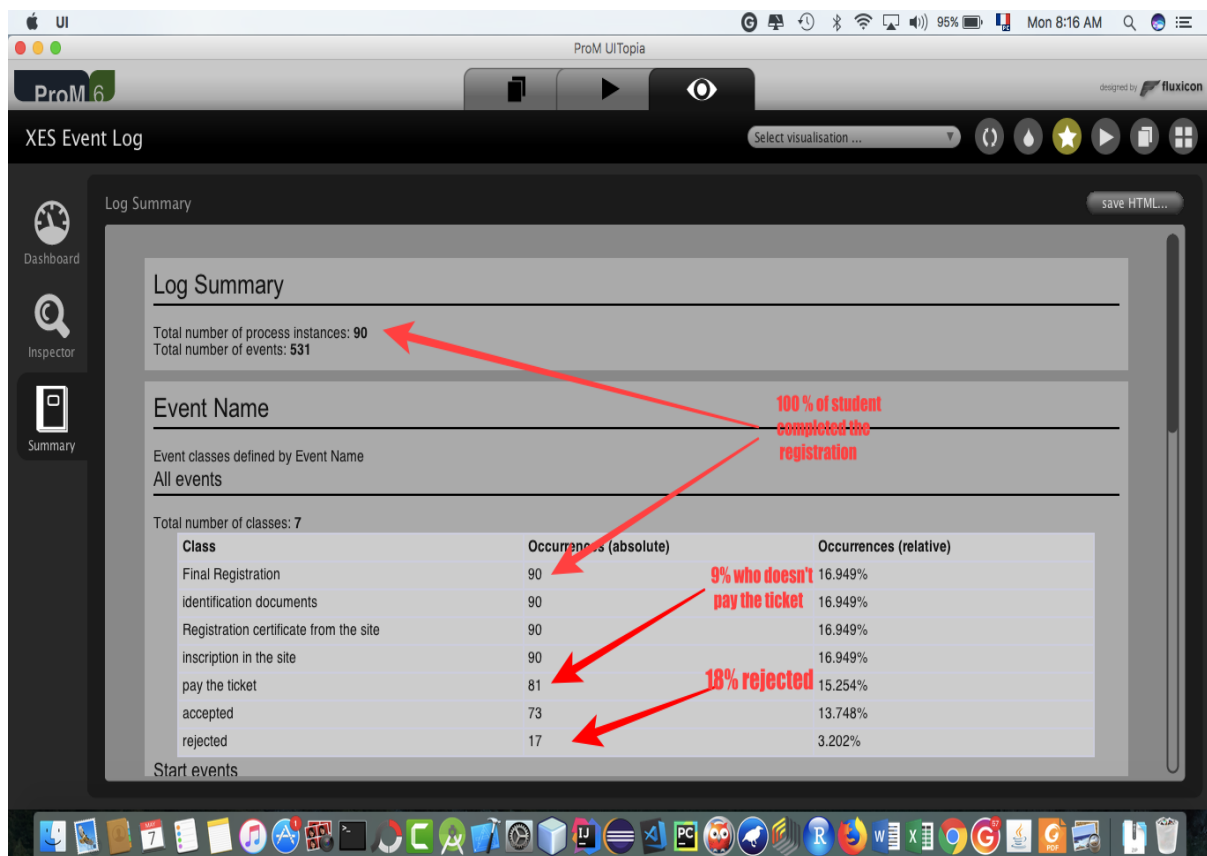


fig 4.9 the inspector of Prom lite

12-Questions Answered Based on process mining :

12.1 Discovery process model:

Based on inductive miner a get this process model:

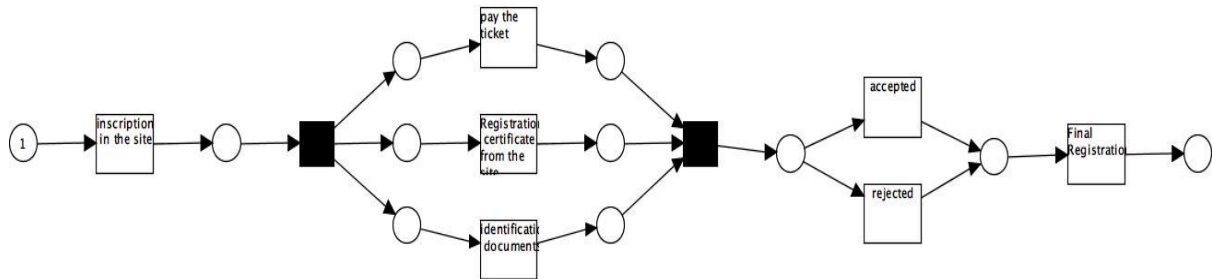


fig 4.10 inductive miner process model

Based on alpha miner a get this process model:

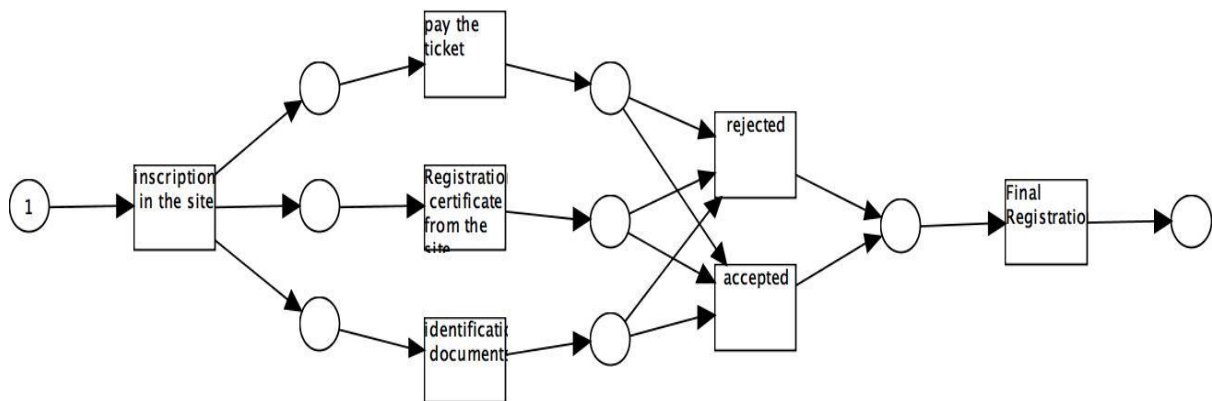


fig 4.11 alpha miner miner process model

12.2 Conformance checking with the inductive miner :

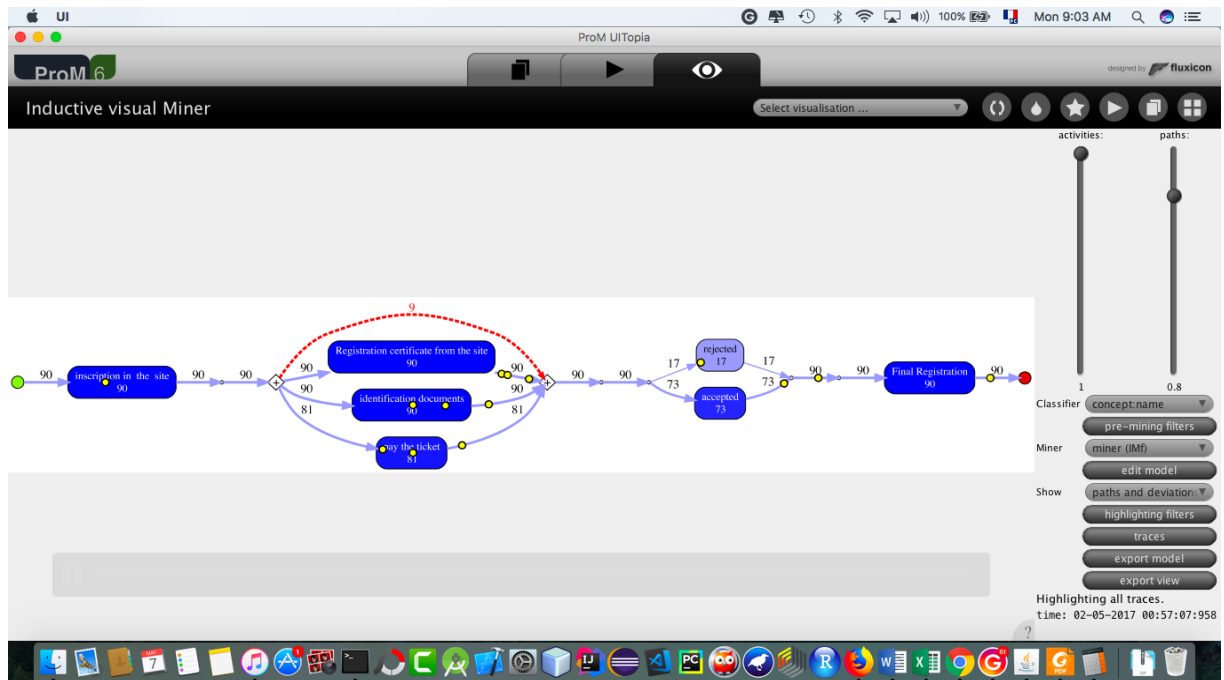


fig 4.12 Conformance checking with the inductive miner

as we notice the step of « buy the ticket » skipped 9 time .

12.3 Conformance checking using alignments :

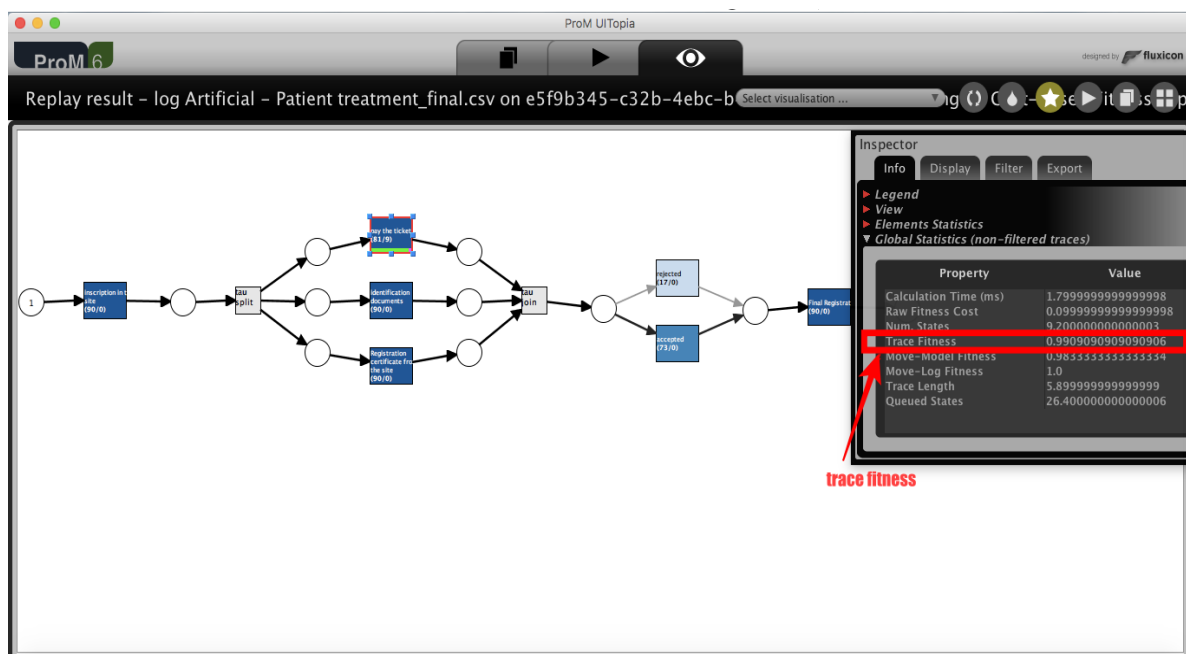


fig 4.13 Conformance checking using alignments

13 Performance analysis - Process Mining in computer sciences department:

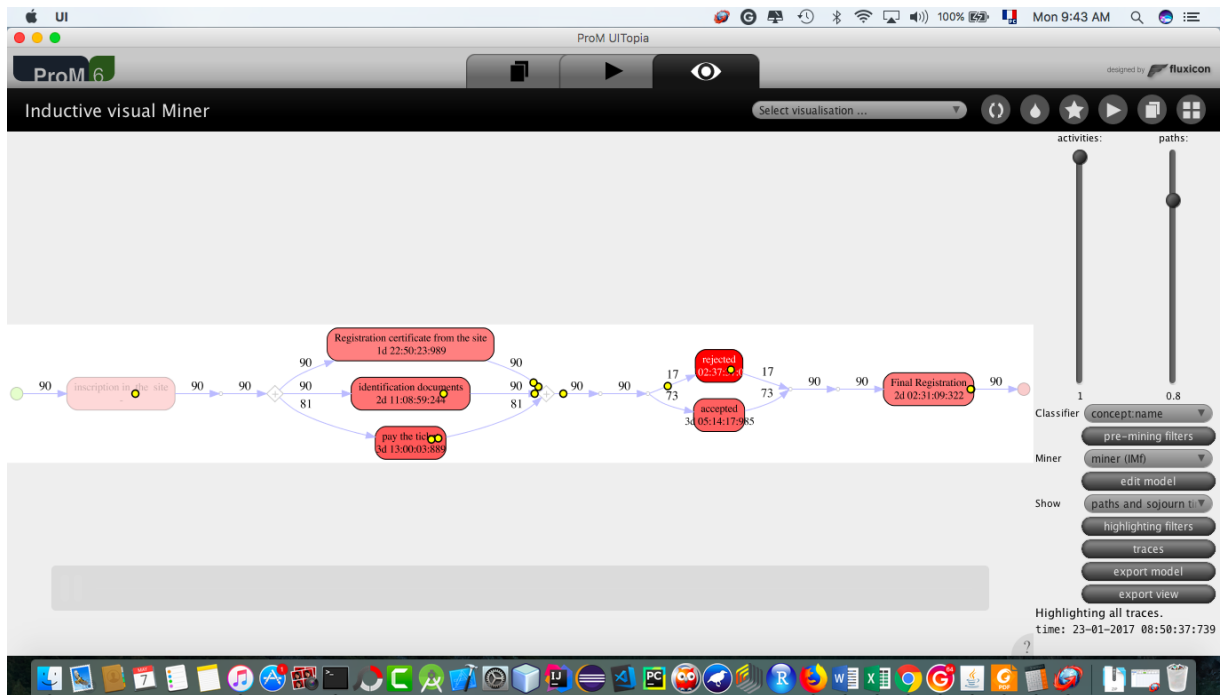


fig 4.14 Performance analysis - Process Mining in computer sciences department

14. other questions that be answers by process mining

1. What is the most frequent path in my process?

Answer :

The most frequent path in my process is :

<inscription in the site , ... , accepted , final registration >

2. How is the distribution of all cases over the different paths through the process?

Answer:

Firstly The pathway of any student always start from the inscription in the site as a point start then he have the three steps in parallel (pay the ticket , paid the identification document , paid the registration certificate from the site ,) then after few days Receive The response (rejected or accepted) .

4 How compliant are the cases (i.e. process instances) with the deployed process models :

Answer

the number of student who buy the ticket is minus

5. Where are the problems :

Answer

some student deviation from the real pathway because some stuff take some mistake and has bad controlled to the ticket

6.How frequent is the (non-)compliance?

Answer : 9 student .

7.What is the average/minimum/maximum throughput time of cases?

Answers :

Case throughput (average) : 9,45days

Case throughput (minimum) : 1 days

Case throughput (maximum) : 17 days

8.What is the average service time for each task?

Answer : x

-the inscription in the site : 0 days

-pay the ticket : 3,days

-paid identification document : 2 days

-paid registration : 1,96 days

-rejected = 7 days

-accepted = 3 days

-final registration

9.Which paths take too much time on average? How many cases follow these routings?

10.How much time was spent between any two tasks in the process model?

Answer :

The much time was spent between any two tasks in the process model is 3,48 days

10. How are the cases actually being executed?

Answer :

The number of case actually being executed is 7 case .

11. How many people are involved in a case?

Answer

The number of 150 student

10- What is the average/minimum/maximum service time for each task?

Answer

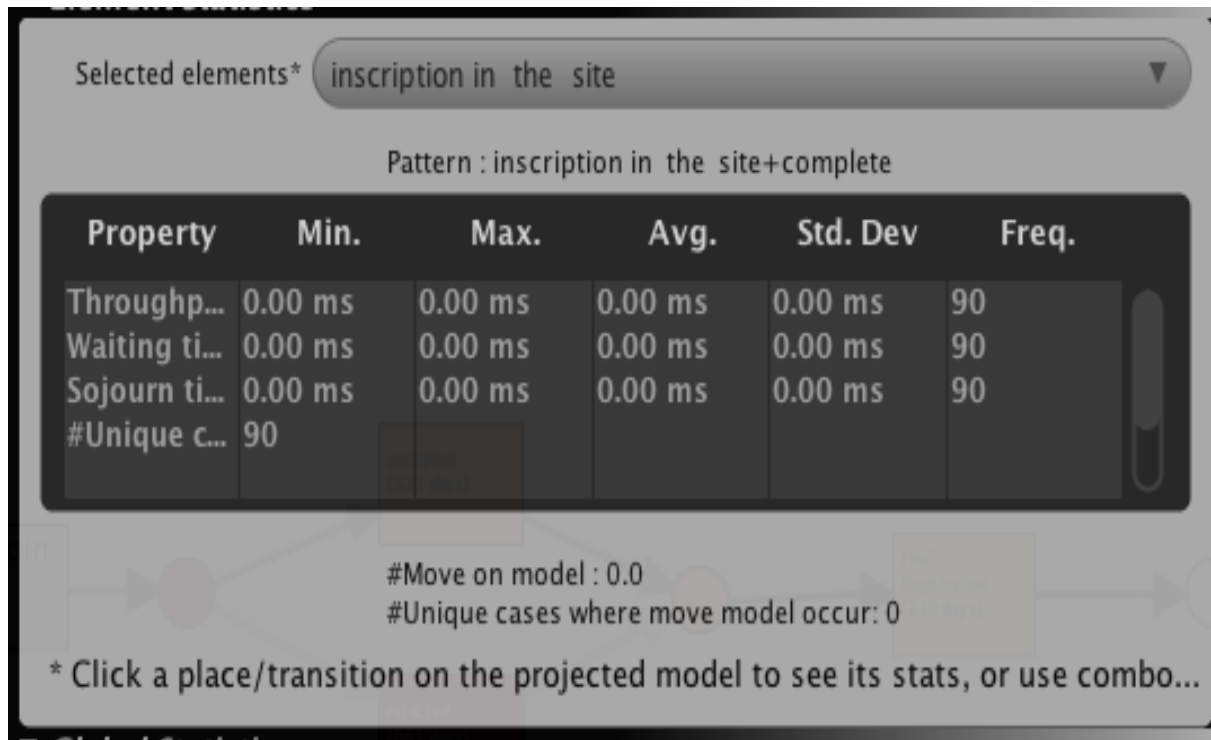


Fig 4.15 service time for inscription in the site

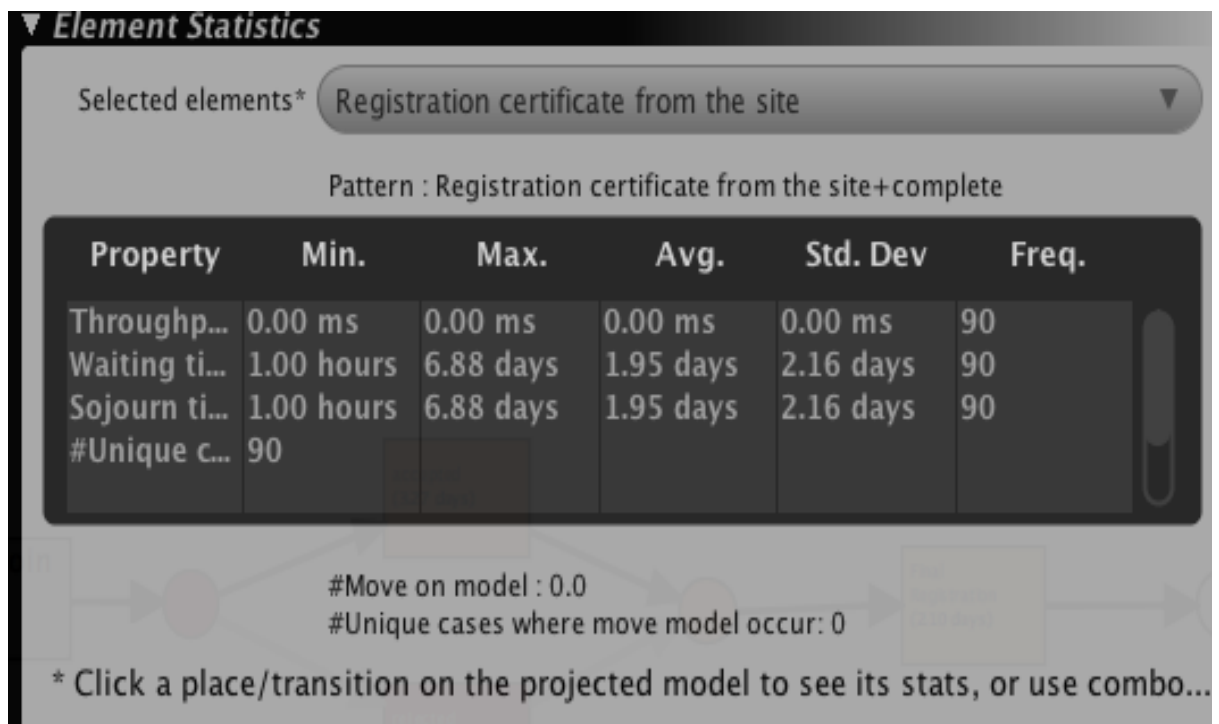


Fig 4.16 service time for registration certificate from the site

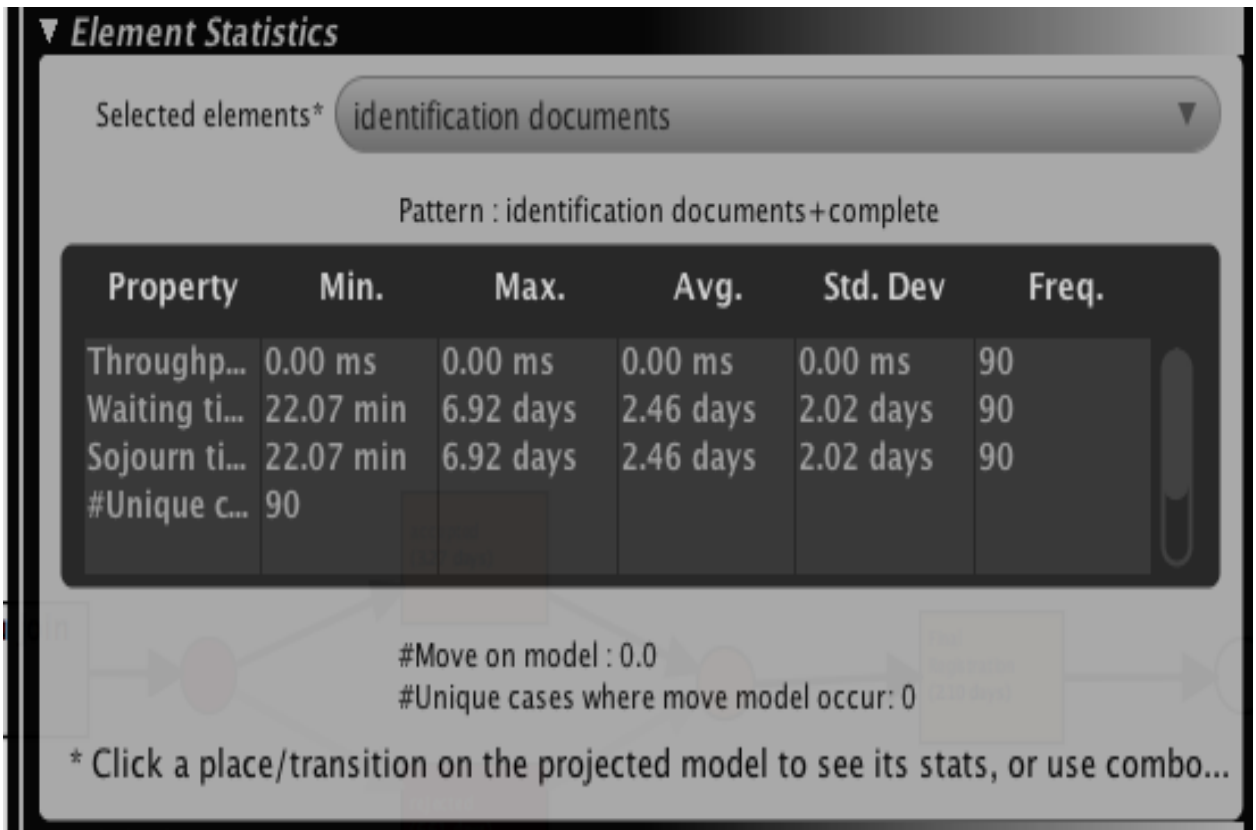


Fig 4.17 service time for identification documents pattern

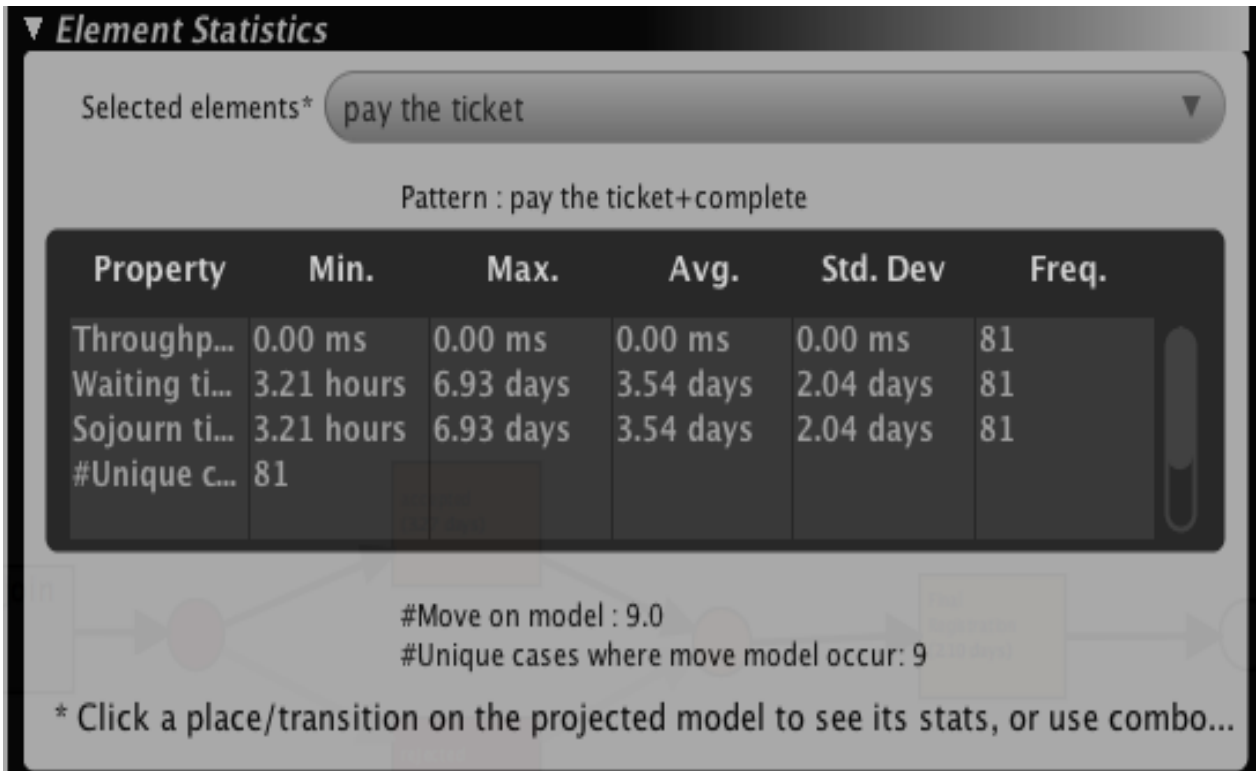


Fig 4.18 service time for pay the ticket

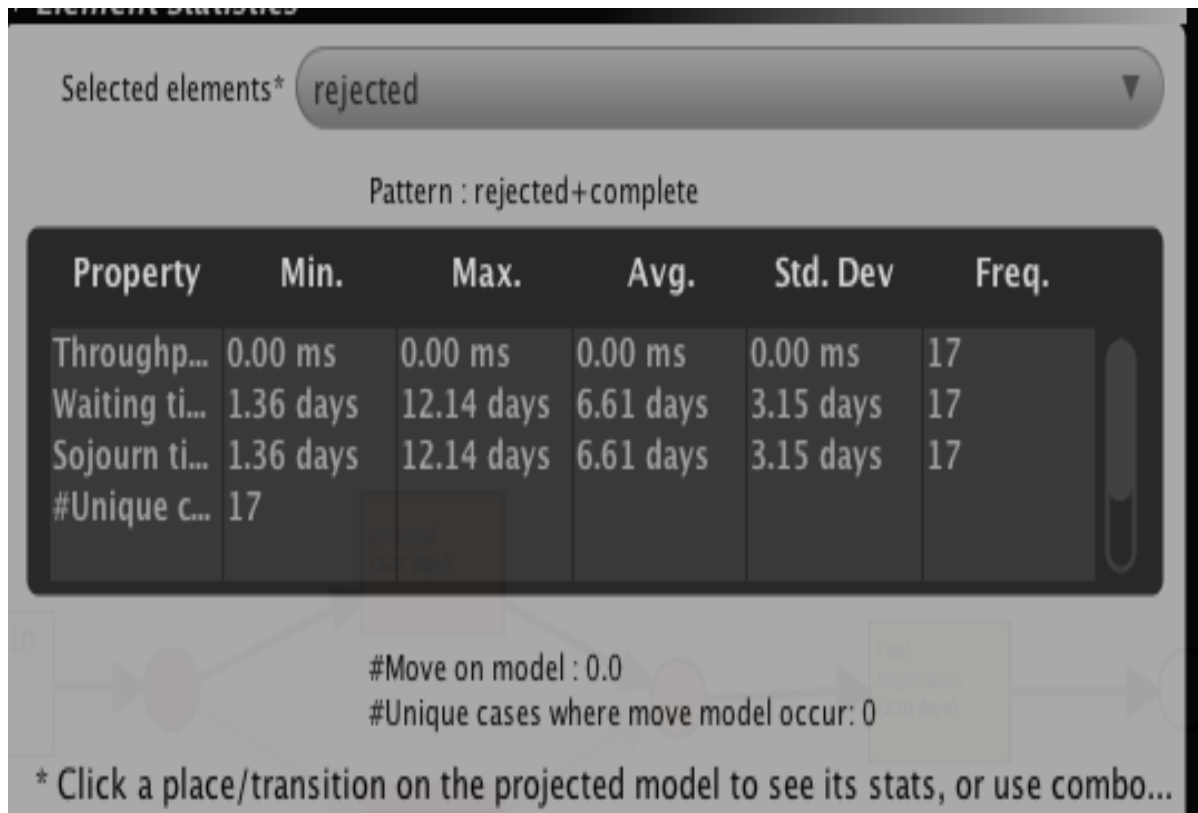


Fig 4.19 service time for rejected pattern

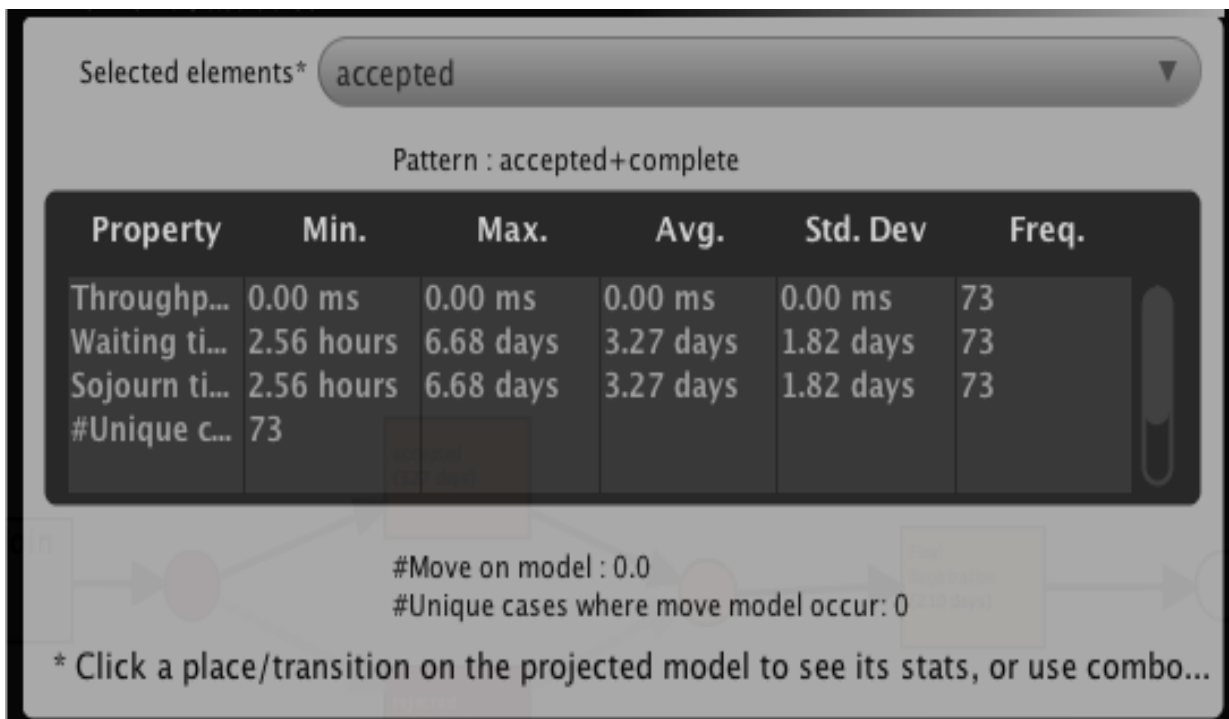


Fig 4.20 service time for accepted pattern

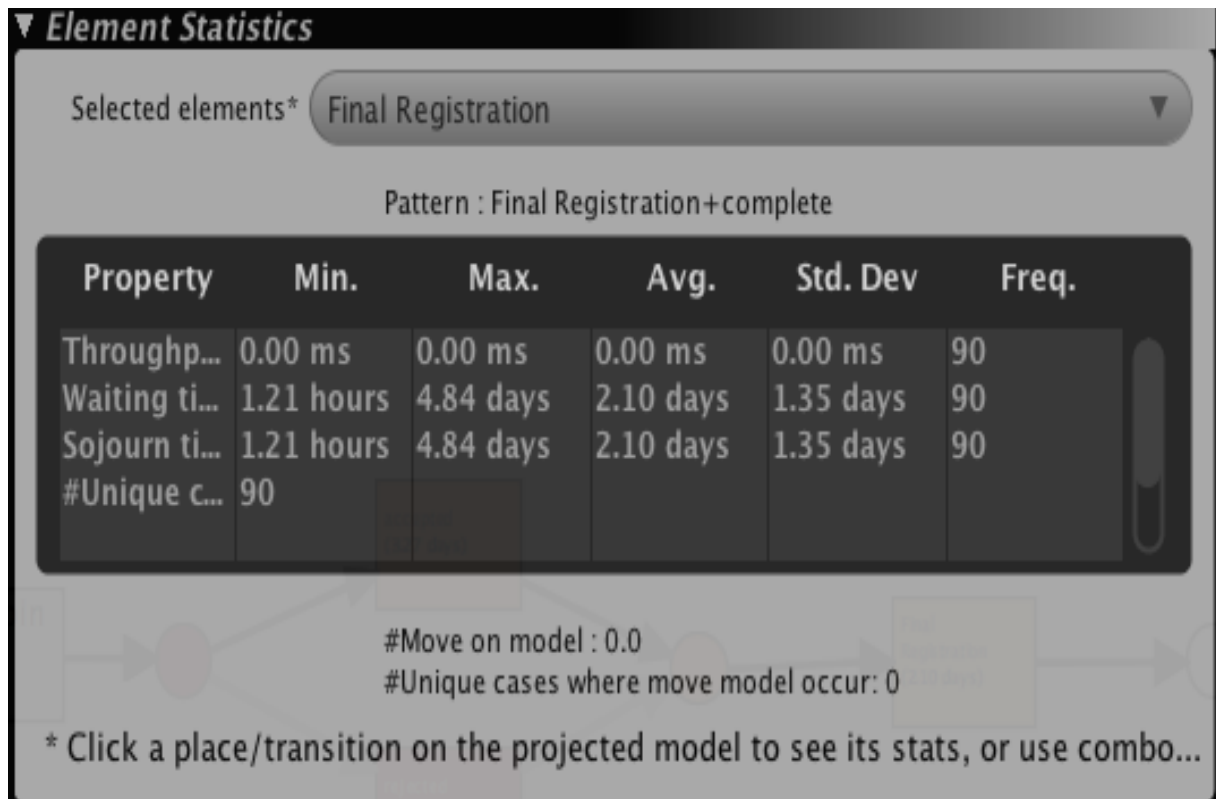


Fig 4.21 service time for final registration pattern

15.Conculsion :

In the end we practise the process mining with a real data and show how is powerful tools to deal with the process of any organisation

