University of St.Gallen
Institute of Computer Science

Introduction

and First Steps

Process Mining

Session 1

From insight to impact.

Outline

1. Introduction
   - Process Mining
   - This Course

2. First Steps

Process Mining Success Stories (self-reported)

- ***Deutsche Telekom*** — *Procurement / purchase-to-pay*: >€10M saved annually; cash discount capture rate reached 96%; paid-on-time rate >90%
- ***Accenture*** — *Request-to-order / invoice approval*: 50% improvement in request-to-order cycle time; 30% reduction in invoice approval time; $35M in annualized working capital benefits.
- ***Vodafone*** — *Invoice processing*: 553% improvement in "touchless" (no manual intervention) invoice rate.
- ***Siemens*** — *Order-to-cash / manufacturing / inventory*: automation rate up 24%, global rework down 11%, ~10 million fewer manual touches per year, 1.5 million additional items delivered on-time annually, and $8.7M in economic value added from digital inventory management optimization.
- …

Source: Celonis, cf. also Forrester: The Total Economic Impact™ of Celonis

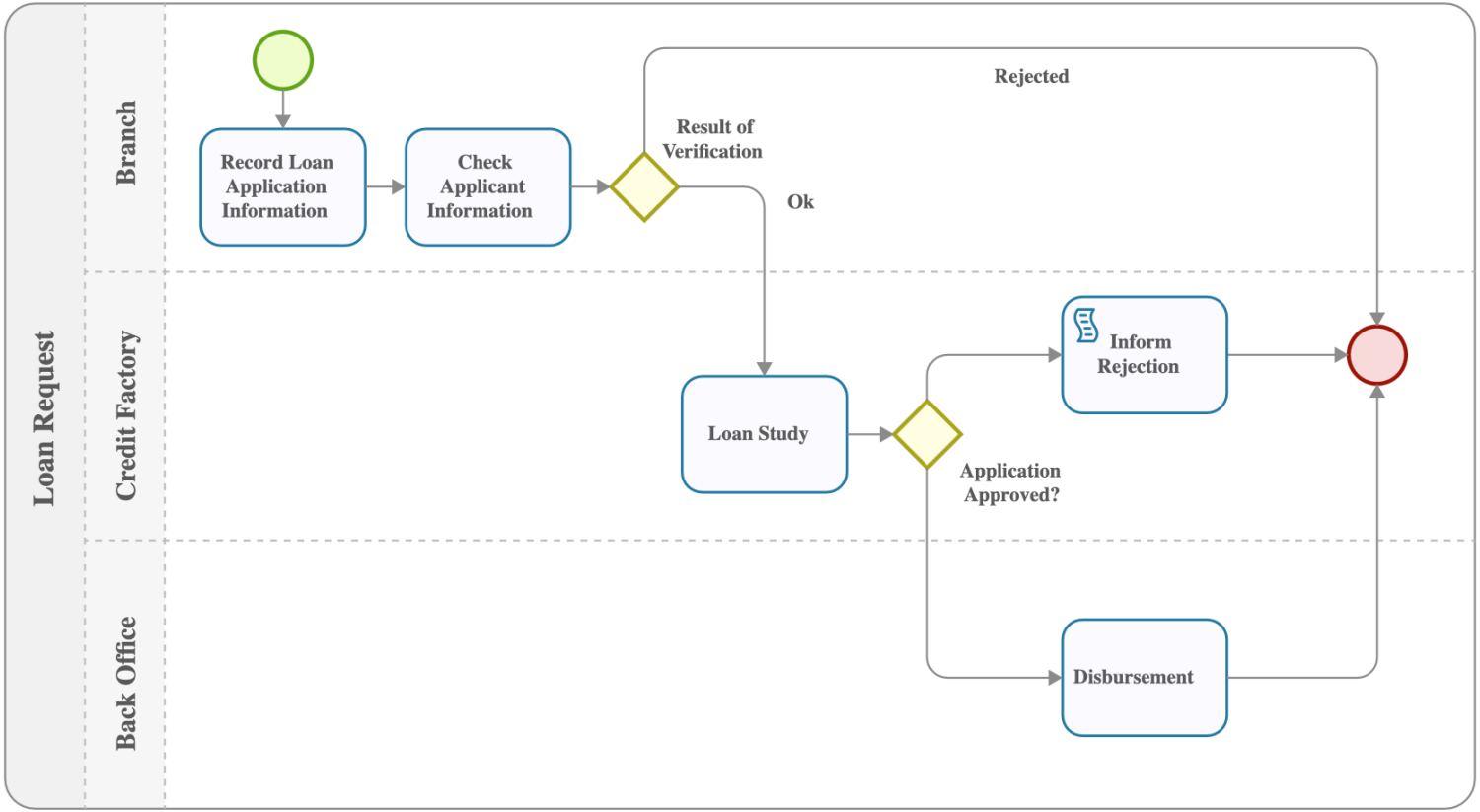
Process Mining

= *Data mining applied to event logs. An event log records the execution of a process.*

Outline

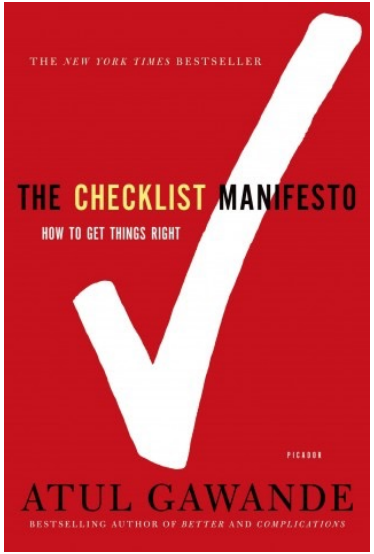
1. Process Mining
   - What is a process?
   - Why bother?
   - What is process mining?

What is a process? – A first simple example (in BPMN)



Source: https://www.bizagi.com/en/blog/guide-to-bpmn

What is a process? – A checklist



Surgical Safety Checklist

World Health Organization | Patient Safety
A World Alliance for Safer Health Care

Before induction of anaesthesia
(with at least nurse and anaesthetist)

Has the patient confirmed his/her identity, site, procedure, and consent?
- ☐ Yes

Is the site marked?
- ☐ Yes
- ☐ Not applicable

Is the anaesthesia machine and medication check complete?
- ☐ Yes

Is the pulse oximeter on the patient and functioning?
- ☐ Yes

Does the patient have a:

Known allergy?
- ☐ No
- ☐ Yes

Difficult airway or aspiration risk?
- ☐ No
- ☐ Yes, and equipment/assistance available

Risk of >500ml blood loss (7ml/kg in children)?
- ☐ No
- ☐ Yes, and two IVs/central access and fluids planned

Before skin incision
(with nurse, anaesthetist and surgeon)

- ☐ **Confirm all team members have introduced themselves by name and role.**
- ☐ **Confirm the patient's name, procedure, and where the incision will be made.**

Has antibiotic prophylaxis been given within the last 60 minutes?
- ☐ Yes
- ☐ Not applicable

Anticipated Critical Events

To Surgeon:
- ☐ What are the critical or non-routine steps?
- ☐ How long will the case take?
- ☐ What is the anticipated blood loss?

To Anaesthetist:
- ☐ Are there any patient-specific concerns?

To Nursing Team:
- ☐ Has sterility (including indicator results) been confirmed?
- ☐ Are there equipment issues or any concerns?

Is essential imaging displayed?
- ☐ Yes
- ☐ Not applicable

Before patient leaves operating room
(with nurse, anaesthetist and surgeon)

Nurse Verbally Confirms:
- ☐ The name of the procedure
- ☐ Completion of instrument, sponge and needle counts
- ☐ Specimen labelling (read specimen labels aloud, including patient name)
- ☐ Whether there are any equipment problems to be addressed

To Surgeon, Anaesthetist and Nurse:
- ☐ What are the key concerns for recovery and management of this patient?

This checklist is not intended to be comprehensive. Additions and modifications to fit local practice are encouraged.

Revised 1 / 2009

© WHO, 2009

What is a process? – A textual description

The loan approval process involves several steps designed to evaluate a borrower's financial stability and ability to repay the requested loan.

The process begins when the borrower submits a loan application to a lender. This application includes details about the borrower's identity, employment, income, credit history, and the purpose of the loan.

The lender conducts an initial review of the application to ensure it is complete and meets basic eligibility criteria. This step includes verifying that the borrower meets minimum credit score and income requirements.

The lender pulls the borrower's credit report to assess their credit history, score, and debt-to-income ratio. This step helps the lender evaluate the borrower's financial habits and ability to manage debt.

The borrower must provide proof of income, such as pay stubs, tax returns, or bank statements. The lender may also contact the borrower's employer to verify employment details.

For secured loans, the lender evaluates the collateral provided by the borrower (e.g., a property for a mortgage or a car for an auto loan) to determine its value and ensure it is sufficient to secure the loan.

The lender calculates the borrower's debt-to-income (DTI) ratio to assess whether they have sufficient income to cover existing debts and the new loan payment.

During underwriting, the lender conducts a detailed analysis of the borrower's financial profile, creditworthiness, and the loan's associated risks. This step may involve automated or manual reviews.

Based on the underwriting results, the lender decides whether to approve or deny the loan application. If approved, the terms of the loan (interest rate, repayment period, and monthly payment) are finalized and communicated to the borrower.

The borrower reviews the loan agreement, including terms and conditions, and signs the necessary documentation if they agree to the terms.

Once the loan agreement is signed, the lender disburses the loan amount. The funds are transferred to the borrower's account or directly to the entity receiving the payment (e.g., a seller in a mortgage).

After disbursement, the borrower begins repaying the loan as per the agreed schedule. The lender monitors repayments and communicates with the borrower regarding any issues or missed payments.

What is a process? – A process as an *event log*.

Case ID	Activity Name	Timestamp	Role Name	Resource	Loan Amount	Applicant Income	Loan Term	Credit Score
1	Submit Application	1/14/25 9:00	Applicant	Mike Johnson	14591	96013	12	680
1	Initial Review	1/14/25 9:42	Loan Officer	Emily Clark	14591	96013	12	680
1	Credit Check	1/14/25 10:52	Credit Analyst	David Brown	14591	96013	12	680
1	Final Approval	1/14/25 11:25	Loan Manager	Helen Carter	14591	96013	12	680
2	Submit Application	1/14/25 9:00	Applicant	Sarah Lee	49201	41124	12	720
2	Initial Review	1/14/25 10:16	Loan Officer	Emily Clark	49201	41124	12	720
2	Credit Check	1/14/25 10:55	Credit Analyst	Susan White	49201	41124	12	720
2	Final Approval	1/14/25 11:26	Loan Manager	Michael Green	49201	41124	12	720
3	Submit Application	1/14/25 9:00	Applicant	Alex Field	6573	95917	48	750
3	Initial Review	1/14/25 9:34	Loan Officer	Michael Green	6573	95917	48	750
3	Credit Check	1/14/25 10:06	Credit Analyst	Susan White	6573	95917	48	750
3	Final Approval	1/14/25 11:10	Loan Manager	Michael Green	6573	95917	48	750
4	Submit Application	1/14/25 9:00	Applicant	James Black	38687	75755	24	750
4	Initial Review	1/14/25 10:25	Loan Officer	Michael Green	38687	75755	24	750
4	Credit Check	1/14/25 11:35	Credit Analyst	Laura Grey	38687	75755	24	750
4	Final Approval	1/14/25 12:21	Loan Manager	Michael Green	38687	75755	24	750
5	Submit Application	1/14/25 9:00	Applicant	Philip Marks	45948	37981	12	680
5	Initial Review	1/14/25 9:54	Loan Officer	Emily Clark	45948	37981	12	680
5	Credit Check	1/14/25 11:17	Credit Analyst	Susan White	45948	37981	12	680
5	Final Approval	1/14/25 11:51	Loan Manager	Michael Green	45948	37981	12	680
6	Submit Application	1/14/25 9:00	Applicant	John Doe	16781	108431	36	680

What is a process? – More examples

Business

- Procure-to-Pay (P2P)
- Quote-to-Cash (Q2C)
- Hire-to-Retire (H2R)
- Record-to-Report (R2R)
- Lead-to-Order (L2O)
- Idea-to-Offering (I2O)
- Request-to-Service (R2S)
- Issue-to-Resolution (I2R)
- Campaign-to-Lead (C2L)
- Demand-to-Supply (D2S)
- …

Engineering

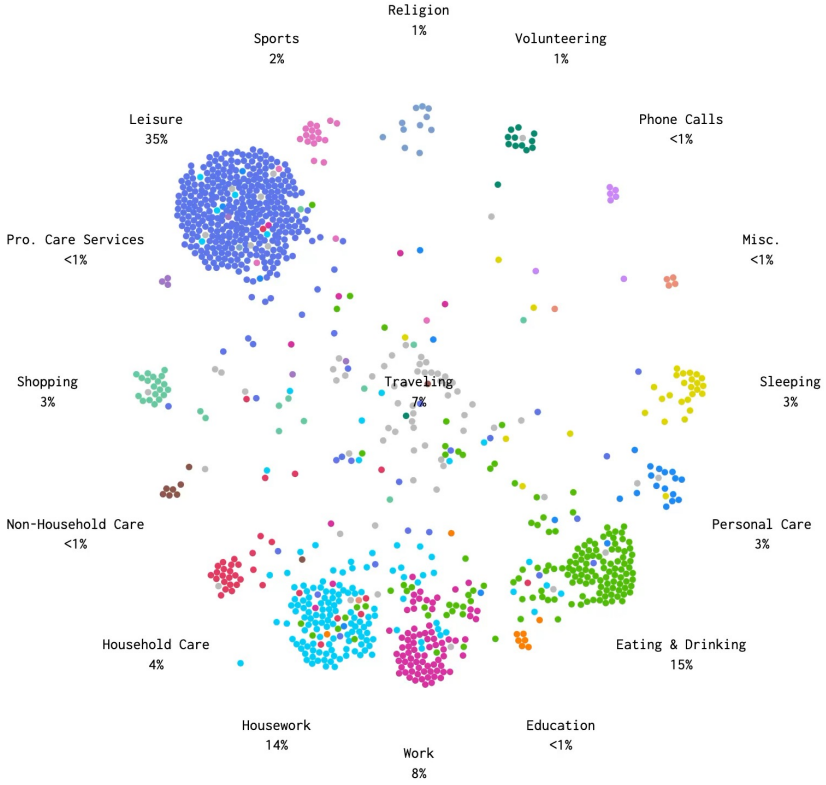
- Develop a software feature (Plan to Release)
- Testing a system under realistic usage conditions
- …

Transport/Logistics, Administration, Manufacturing, Science, Health Care, …

- Shipping a consignment from location A to B
- Swiss federal referendum
- …

Other Event Logs

- ***Technical logs***: web server access logs, auth/security logs, application logs, clickstreams, e-commerce transaction logs,
- ***Collaboration platforms***: GitHub/GitLab issues & PRs, bug trackers (Jira, Bugzilla), Wikipedia edit history, stack overflow, customer support chat/ticket transcripts
- ***Physical devices***: hardware device usage data, public transit / GPS trip logs, smart home/sensor logs



A day in the life of americans (Nathan Yau)

Process

= *a model of how a set of actors collaborate (to reach a common goal).*

Actors

Humans ⟷ AI Agents ⟷ Machines

higher-level interfaces

more predictable

synonyms for "process"

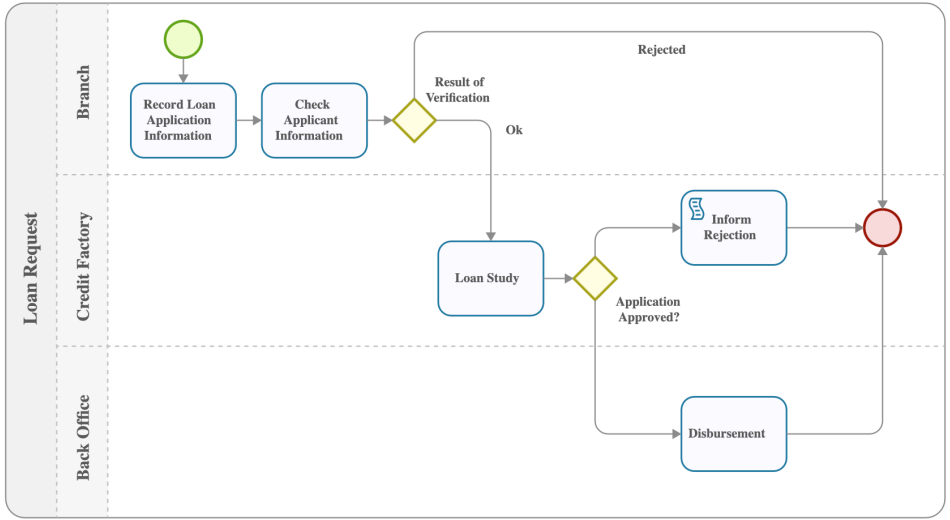
process,
workflow,
(legal, diplomatic) procedure/protocol

(communication, network) protocol
(coordination, network) algorithm
service orchestration/choreography

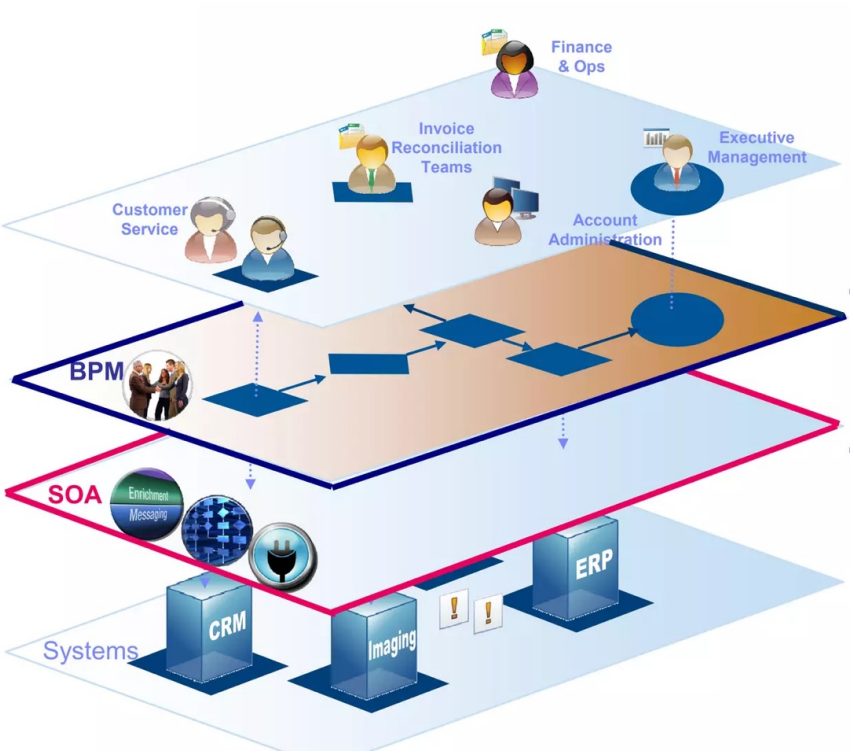
A process can have any mix of actors.

What is common across most processes?

- multiple interacting steps
- multiple actors
  - different actors may belong to different organizational units
  - we can also abstract from which actor is executing each step
- multiple information objects maybe used across multiple steps
  - e.g. folders, documents
- multiple actors act concurrently and are distributed in space



What is a (business) process? – Software engineering view



© 2013 IBM, Source: https://www.slideshare.net/SuraGonzalez/bpm-overview

Process Participants

Business Process
(Application Integration Layer)

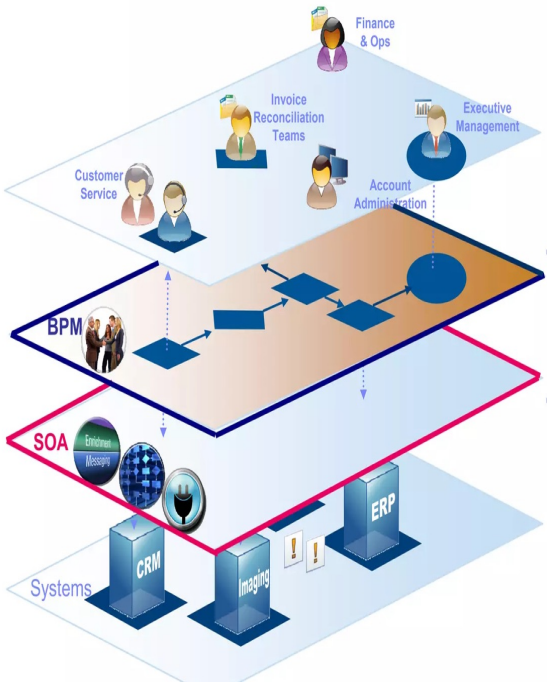
Integration Infrastructure Layer

Enterprise Applications

What is a (business) process? – Software engineering view

A piece of software

- high up in the stack, on top of applications
- typically integrates multiple applications
- often semi-automated, i.e., contains *human tasks*
- often long-running
- embodied as
  - orchestration of services/events
  - explicitly as a (BPMN) process model in a BPMS
    - e.g., Camunda, SAP, IBM, Oracle, Appian, …



© 2013 IBM, Source: https://www.slideshare.net/SuraGonzalez/bpm-overview

Processes – Why bother?

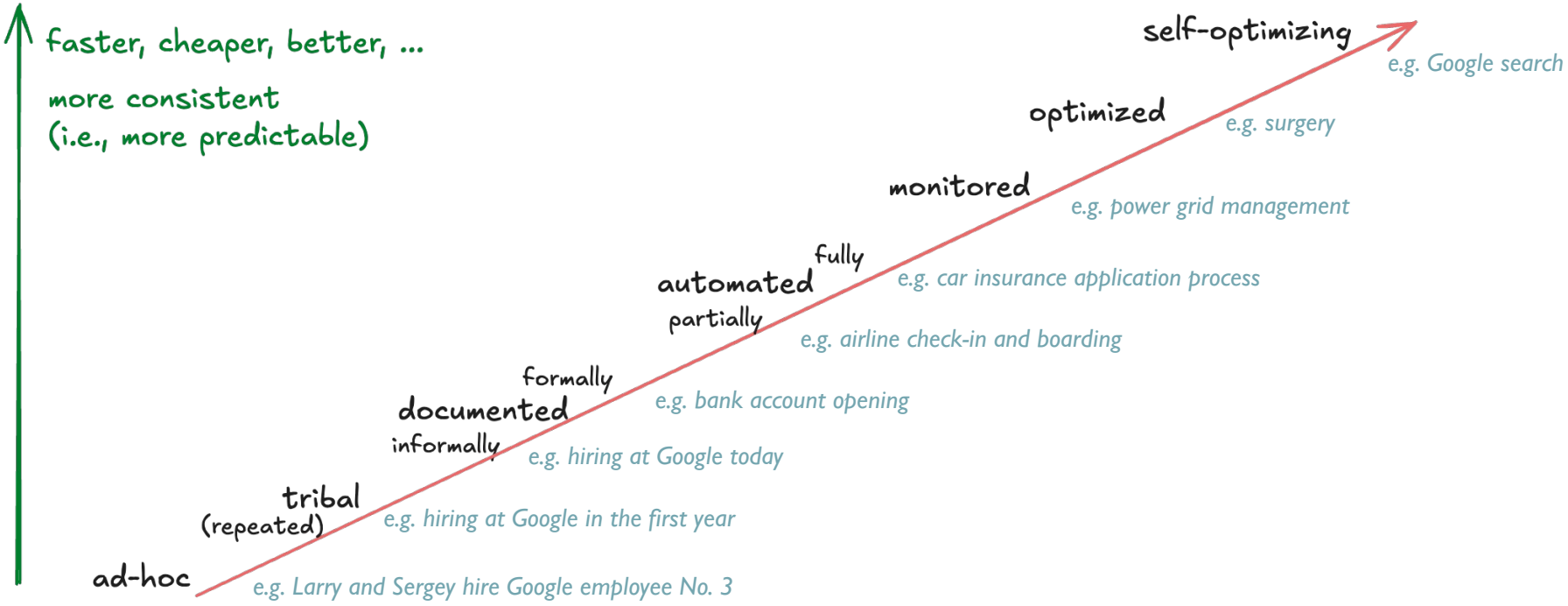
Processes – why bother?

Your processes determine your

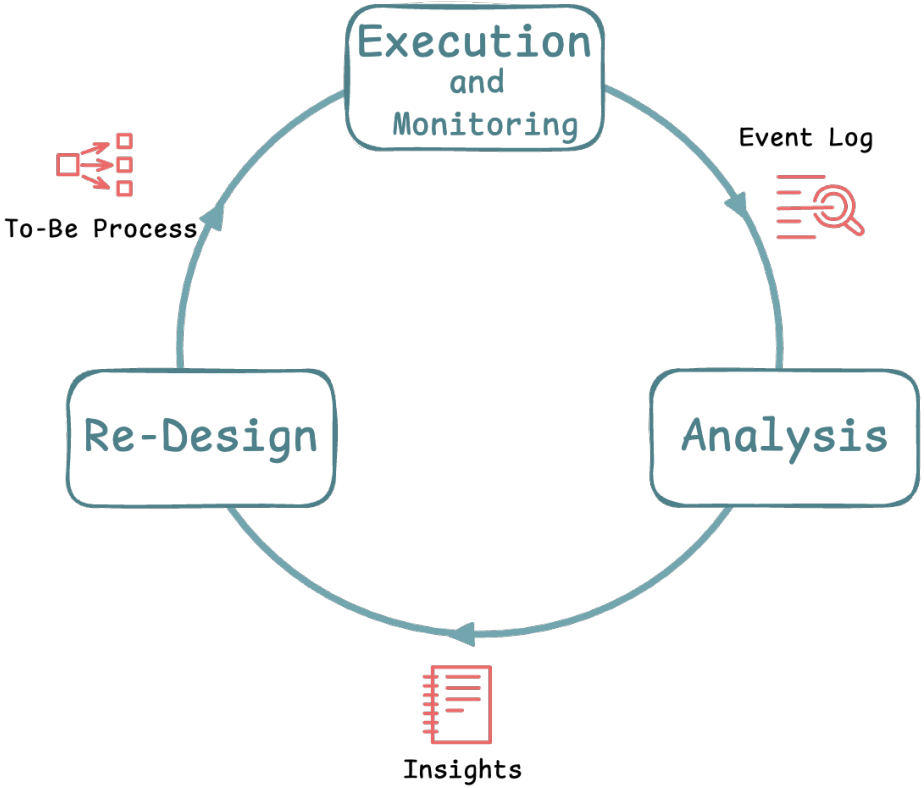
- efficiency, effectiveness, scalability, robustness, agility etc. and therefore
- your competitiveness

processes = "software" assets of your organization *(also non-automated ones)*

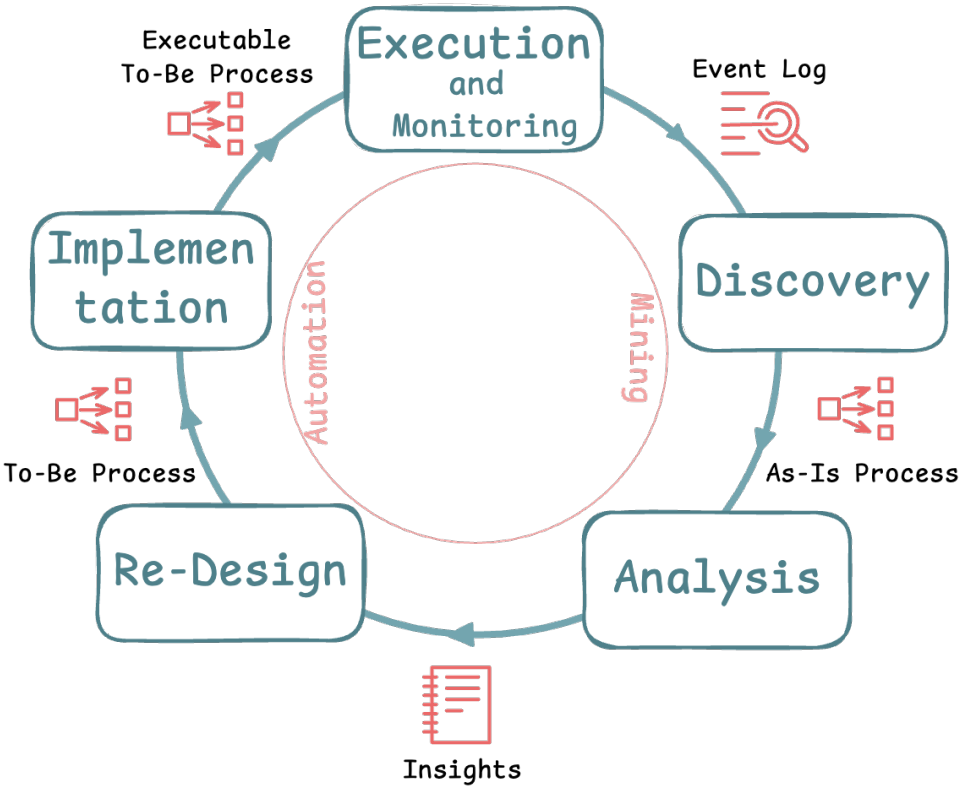
A difference across processes: Process Maturity, cf. also CMM



How do processes improve?

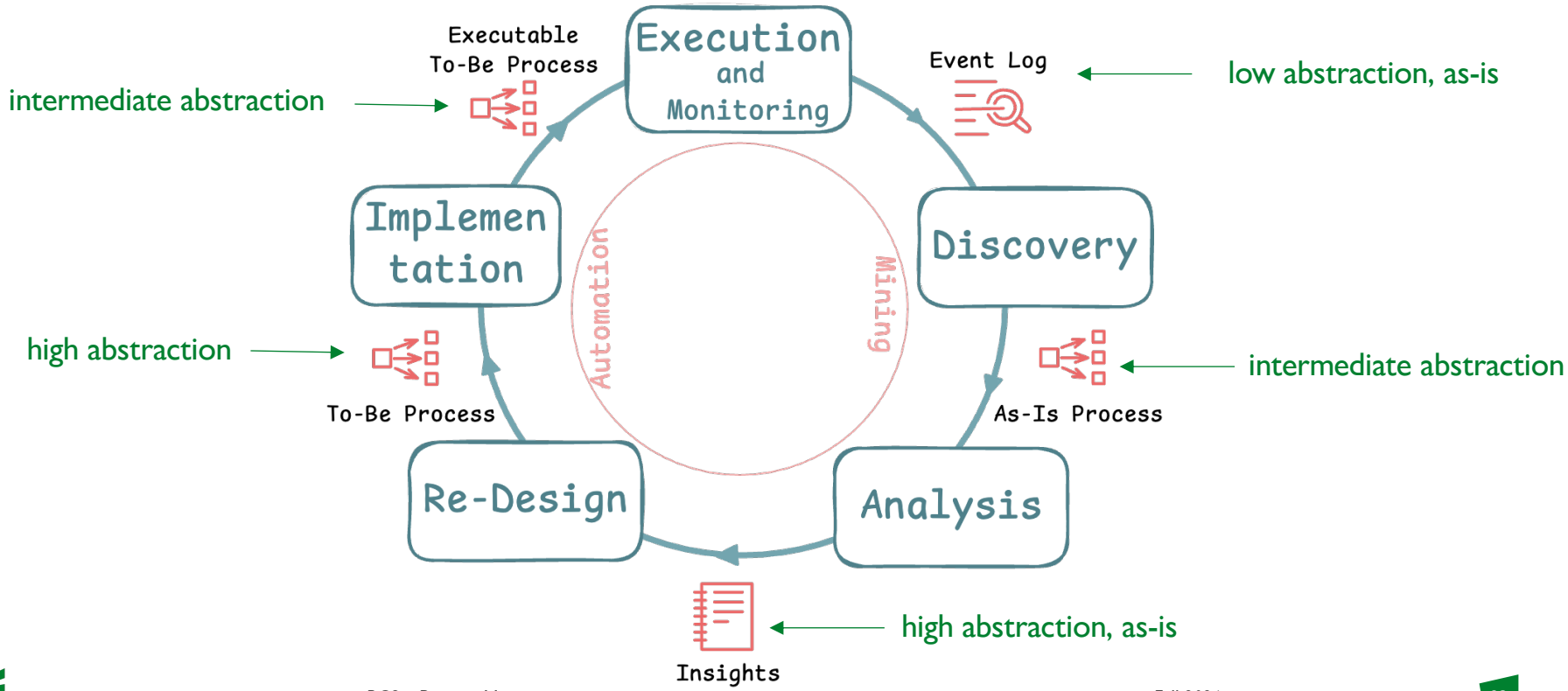


How do processes improve?

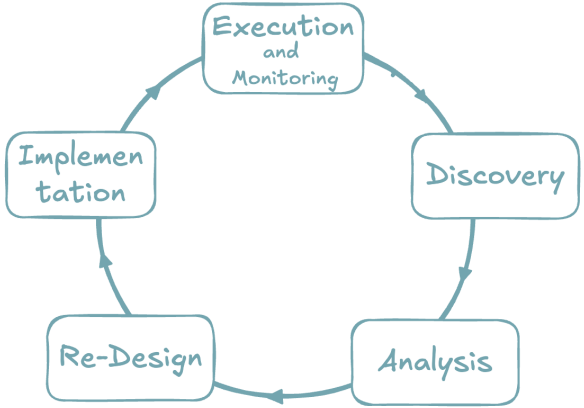


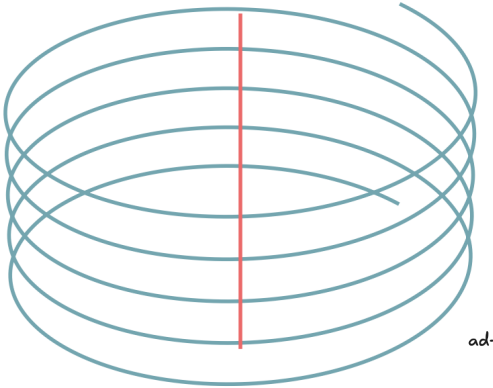
Process Engineering Lifecycle
(cf. BPM Lifecycle)

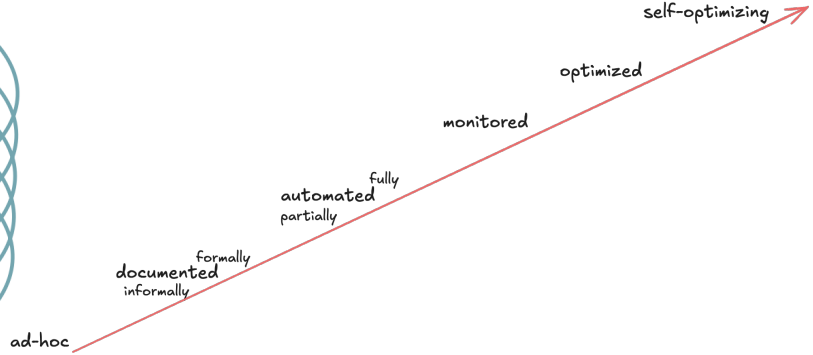
Models on different abstraction levels



Continuous Process Improvement







Process Mining

= *Data mining applied to event logs. An event log records the execution of a process.*

What is process mining? – First examples

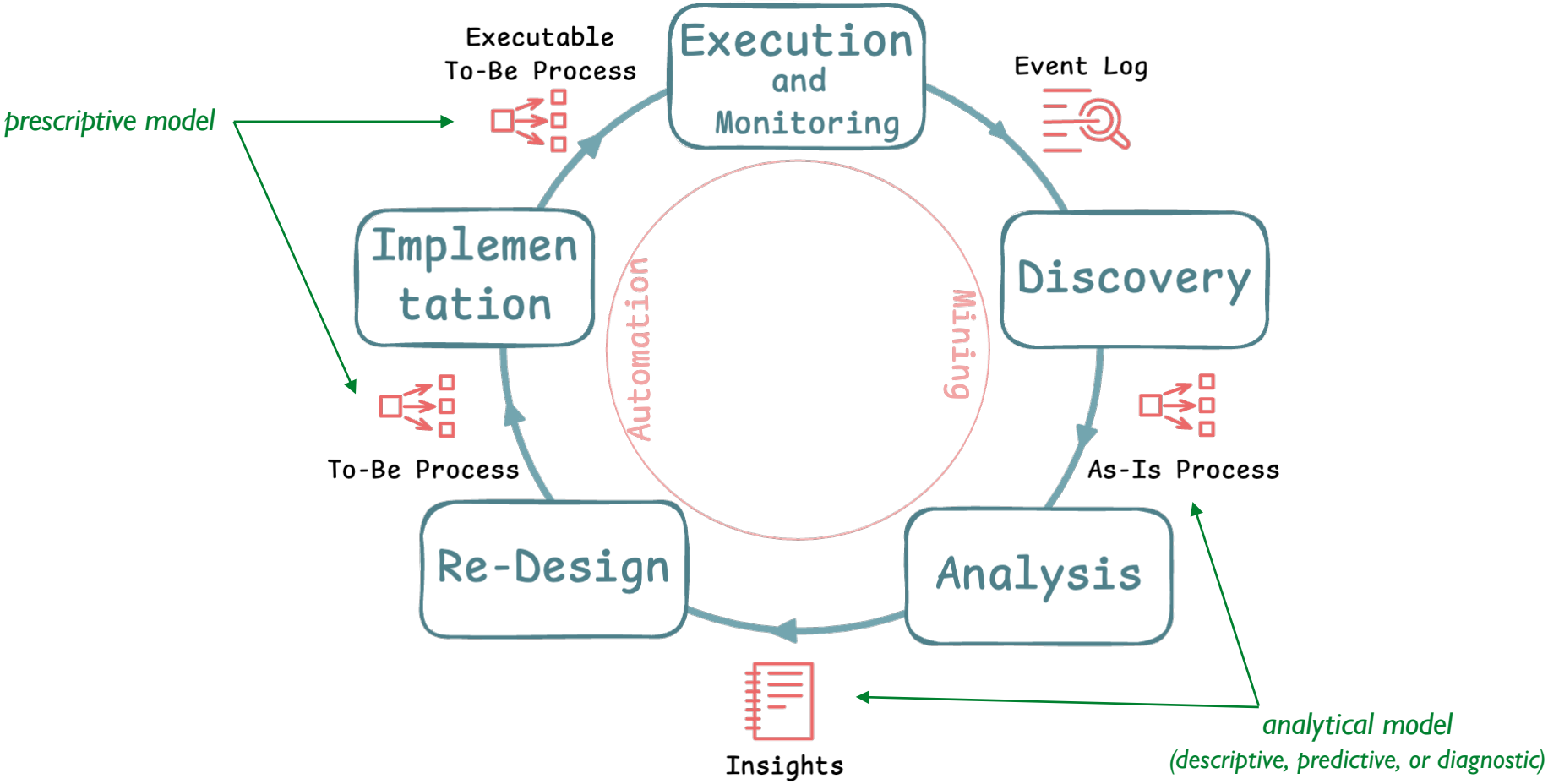
Event log = recording of the process execution

https://fluxicon.com/book/read/dataext/

CaseID	Timestamp	Medium	Activity	Service Line	Urgency
case9700	20.8.09 11:46	Phone	Registered	1st line	0
case9700	20.8.09 11:50	Phone	Completed	1st line	0
case9701	23.9.09 12:23	Phone	Registered	1st line	0
case9701	23.9.09 12:27	Phone	Completed	1st line	0
case9705	20.10.09 14:21	Phone	Registered	Specialist	2
case9705	20.10.09 16:48	Phone	At specialist	Specialist	2
case9705	19.11.09 10:31	Phone	In progress	Specialist	2
case9705	19.11.09 10:32	Phone	Completed	Specialist	2
case3939	15.10.09 11:48	Mail	Registered	Specialist	2
case3939	15.10.09 11:48	Mail	Offered	Specialist	2
case3939	20.10.09 17:18	Mail	In progress	Specialist	2
case3939	20.10.09 17:19	Mail	At specialist	Specialist	2
case3939	21.10.09 14:49	Mail	In progress	Specialist	2
case3939	21.10.09 14:49	Mail	In progress	Specialist	2
case3939	28.10.09 10:17	Mail	In progress	Specialist	2
case3939	28.10.09 10:18	Mail	Completed	Specialist	2
case9704	20.10.09 14:19	Mail	Registered	1st line	0
case9704	20.10.09 14:24	Mail	Completed	1st line	0
case9703	20.10.09 14:40	Phone	Registered	1st line	0
case9703	20.10.09 14:58	Phone	Completed	1st line	0
case9702	24.8.09 12:24	Mail	Registered	2nd line	2
case9702	24.8.09 12:30	Mail	Offered	2nd line	2

Process Mining answers questions such as:

- What is the actual behaviour of the process in practice? (**Discovery**)
  - *e.g. what sequence of steps (coding, testing, build, …) was actually used to develop the feature? How consistently?*
- To what extent does the actual behavior deviate from the normative or expected behavior (**Conformance**).
  - *e.g. was every pull request reviewed?*
- Where are perfomance bottlenecks that limit throughput or other KPIs? (**Performance**)
  - *e.g. was mean time to release high because of excessive reviews or too many bugs?*


Execution and Monitoring
Event Log
Discovery
As-Is Process
Analysis
Insights
Re-Design
To-Be Process
Implementation
Executable To-Be Process
Automation
Mining
prescriptive model
analytical model
(descriptive, predictive, or diagnostic)

Why do we need to discover a process?

Prescriptive model does not exist:

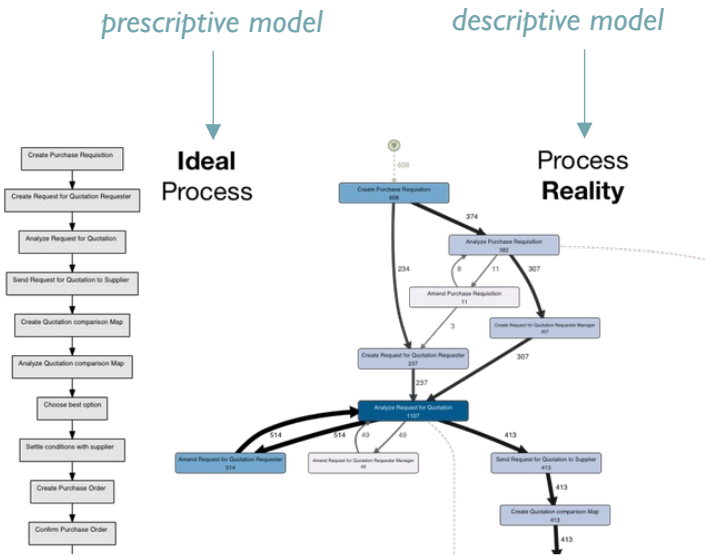
- low process maturity
- prescriptive model is implicit in the minds of different process stakeholders, where each stakeholder
  - sees only part of the process
  - has different expectations based on different goals

Prescriptive model does not specify everything:

- it allows for flexibility (easy adaptation to context change)
- there are too many exceptional behaviors to specify

Prescriptive model is not strictly followed:

- humans: misunderstanding or misaligned with prescription
- machines: incorrect implementation of the prescription



Source: Fluxicon

What is process mining? – More examples

Answer questions like:

- What are differences in travel approval throughput time between national and international trips?
- Can we predict whether an invoice will be paid in time?
- Can we predict ahead of time that an incident resolution will transgress the deadline of its SLA?
- Why do so many parking tickets appear neither paid, officially dismissed, nor credit collected?
- Can the cost of loan approval process be reduced by detecting some ”bad” loan applications earlier in the process?
- Which parts of the process are unnecessarily repeated? Why?
- Where are further automation opportunities?
- …

Why process mining?

Process mining enables data-driven improvement of business processes. Potential improvements include

- increased operational efficiency (e.g., process re-engineering or -automation),
- better management of compliance and risk, and
- enhanced customer experience

Success Stories

- Process Mining in Action , Springer, 2020
  - reports from pioneering case studies including Siemens, Uber, BMW, ABB, Bosch and others
- Celonis Promotional Success Story Videos
  - Example Postfinance

The Process Mining Market is on the Rise

Process Mining Software Market Size to Surpass USD 27.72 Billion by 2030, exhibiting a CAGR of 49.4%

As per the report by Fortune Business Insights, the global process mining software market size is projected to reach USD 27.72 billion in 2030, at a CAGR of 49.4% during the forecast period, 2023-2030

June 08, 2023 04:52 ET| Source: **Fortune Business Insights** Follow

INSIGHTS

SAP to acquire Business Process automation startup Signavio for a reported US$1.2 billion

Staff Writers / Thu 28 Jan 2021

Insights > SAP to acquire Business Process automation startup Signavio for a reported US$1.2 billion

Salesforce to acquire process intelligence firm Apromore

Apromore deal marks a significant step in accelerating Salesforce's innovation in agentic process automation

IBM acquires Italy's myInvenio to integrate process mining directly into its suite of automation tools

Ingrid Lunden @ingridlunden / 2:03 PM GMT+2 • April 15, 2021 Comment

Appian acquires process mining company Lana Labs

Kyle Wiggers @Kyle_L_Wiggers August 5, 2021 2:05 PM

EDITORS' PICK | Jun 2, 2021, 06:00am EDT | 63.454 views

Celonis Raises $1 Billion At $11 Billion Valuation, Making It New York's —And Germany's — Most Valuable Startup

Slide adapted from F. Zerbato

The Celonis Story (part 1)

1998 Academic research into process mining starts

2007 First Commercial Process Mining Tool (Futura Reflect)

2009 Fluxicon founded (by two Computer Scientists) as spin-off from academic research

- *sleek product Disco, but revenue remains mostly through service engagements*

2011 Celonis founded (two BA students, one computer scientist) in Munich

- *bootstrapping through applied projects with big companies (Bavarian Radio, Siemens)*
- *developing a product through some public funding (TU Munich) and service engagements*

2015-2017 Scaling up

- *Standard connectors (SAP, ServiceNow, ..)*
- *Strategic Partnerships (SAP, IBM, Deloitte)*
- *Expansion into US and wider Europe*

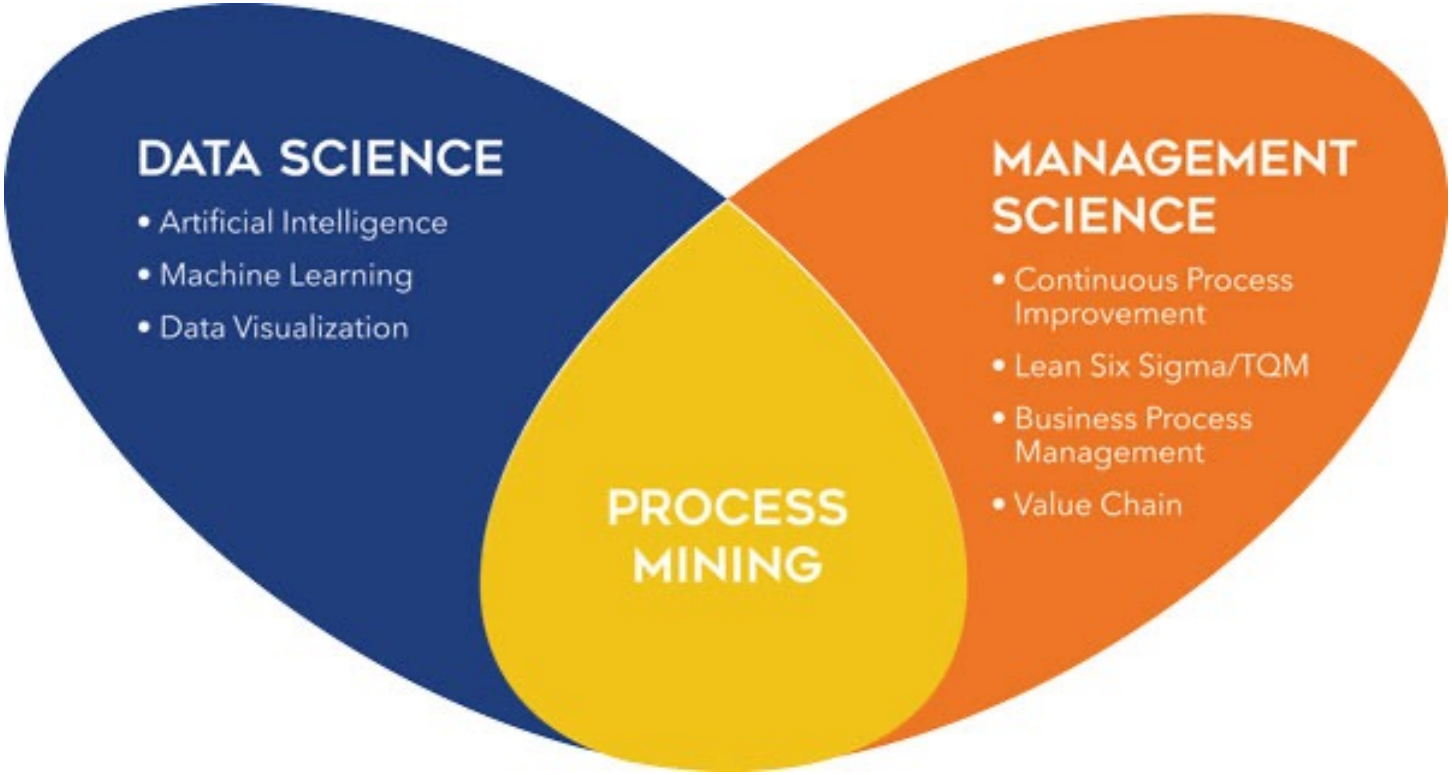
2018 Unicorn status

Gartner Market Report

April 2024

First report for Process Mining: 2018

Figure 1: Magic Quadrant for Process Mining



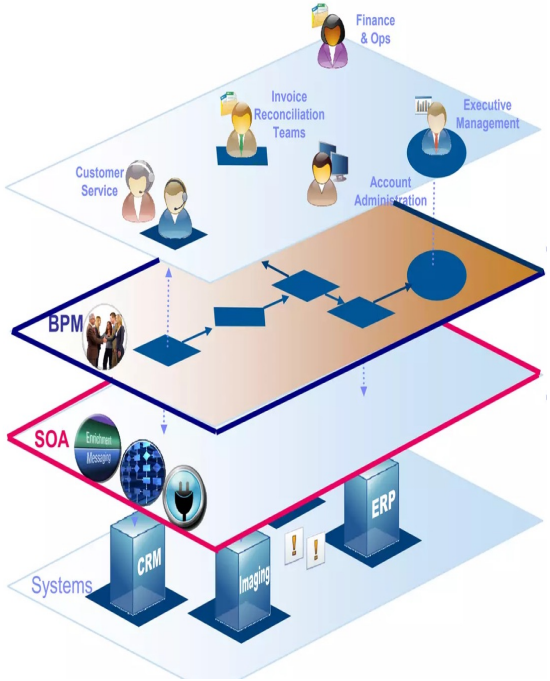
Source: Informs

Business Driver: Process Automation

Process automation is often non-existent or rudimentary, but there is a large automation potential

- process participants work with applications, and send around data in ad-hoc fashion, often using copy/paste between applications

- global market for process automation
  - USD 9.78 billion in 2022
  - USD 21.15 billion by 2030 (10.10% CAGR)



© 2013 IBM, Source: https://www.slideshare.net/SuraGonzalez/bpm-overview

The Celonis Story (part 2)

2019 – today several acquisition in integration, automation and analytics technology

2020 Launched Execution Management System (EMS) moving into automation and monitoring

2022 valued at $13.2 billion

2023 Acquisition of Symbio, a provider of classical BPM software and some AI capabilities

This Course

Course Design Rationale

Process mining is a "bleeding edge" topic:

- discipline is young
  - limited *canonical* content, i.e., established, widely adopted standardized content
  - frequently emerging new content
  - emerging content can suffer from *academic* or *marketing* bias (hype)

Course Design:

- 7 sessions (6 weeks) on established practical technical content
- acquire emerging content together in a *Seminar*
  - reflect critically together in discussion – cut through the hype

Course – Part 1 (weeks 1-6): *Lectures with some integrated practical work*

Session 1: Introduction
- *Event Logs, Event Log Inspection, First Exploration*

Session 2: Exploratory Initial Log Analysis
- *Visualizations, Log Enrichments*

Session 3: Process KPIs
- *Process Outcomes, Dashboards*

Session 4: Behavioral Analysis and –Modeling
- *Variants, Process Maps, BPMN, Petri Nets*

Assignment 1, part 1 due

Session 5: Performance Mining
- *Rework, Load and Demand, Bottlenecks*

Session 6: Conformance Checking
- *Business Rules and Normative Process Models*

Session 7: Deviance Diagnostics
- *Explanations and Rule Mining*

Assignment 1, part 2 due

Practical Work:

- Python Notebook Process Mining analysis leveraging pandas, pm4py, Marimo, plotly
- Using Process Mining Tools such as *Fluxicon Disco, Celonis, Apromore*

Assignment 1

- analyze a real-world event log *(details later in this session)*

- assignment tasks are progressively released with each lecture
  - but due only two times

- contains many exercises with definite answers, reflection and interpretation of analytic results, and some open questions that require some creativity

- coding in Python notebook and some use of process mining tools (Disco, Celonis)
  - hand-in the notebook including result interpretations/reflections

Part 2: Seminar

= A method to *collectively* acquire an emerging topic – often used at university

Skills

- acquire knowledge that is not available in curated form (textbooks)
- communicate deep technical knowledge *(your responsibility to go deep)*
- critically reflect what you have read and tried with others to assess value, maturity etc.

These skills are useful in practice, e.g., in *technology spikes*

Your task: Acquire knowledge (read/hands-on technology exploration) – Reflect – Present

- Two scheduled preparation consultations:
  1. discuss your understanding of your acquired knowledge
  2. discuss presentation
- hand-in your slides and collateral material before your talk

Course – Part 2 (weeks 7-12) : *Seminar*

Choose a flavor based on your preference

- Discovery and Analysis Algorithms
  - *e.g. inductiver miner*
- Data Set 360
  - applied analysis around a known or *new* data set
- Process Mining Tool 360
  - e.g. *Apromore, Celonis*
- Advanced and Emerging Techniques

Advanced and Emerging Process Mining Techniques

- Predictive Monitoring
- Process Simulation
- Declarative mining and querying
- Object-centric process mining
- Event Knowledge Graphs
- Performance Spectrum Visualization
- *Celonis* PQL
- Queries in natural language
- Data curation
- …

Example Seminar Topic: Inductive Miner (Algorithms)

Topic: How does this algorithm learn a process model from an event log?

Practical part: Run the algorithm on the running examples of Part 1, vary parameters and critically assess the results.

Theoretical part: Understand and present the algorithm.

Literature: Original conference paper on the inductive miner. *(Optional secondary literature exists on the topic and can help, but is not required reading.)*

Data: Data sets from lectures in part 1.

Appeal: How induction works under the hood. Existing public videos and presentations help with the understanding.

Practical and theoretical part exist in every topic, but may have different weight in each topic.

Learning Objectives

- Understand basic concepts of processes, the process improvement lifecycle, and the value of process mining
- Elevate and broaden data mining skills using popular Python libraries as well as being able to use state-of-the-art process mining tools to
  - to discover process models, behavioral patterns, and multi-perspective process models
  - to obtain and interpret performance mining results
  - to obtain and interpret conformance checking results
  - understand the benefits and limitations of these state-of-the-art process mining tools
- Being able to understand and to create imperative and declarative process models

- Elevate skills to assimilate, reflect, and clearly present state-of-the-art technical/research papers and/or state-of-the art technology, i.e., give a technical talk
- Being able to critically reflect and discuss ideas presented in a technical talk

Examination Summary

- Assignment 1: Data analysis project (50%) – complete in groups of at most two students, group grade
  - *quality of your data analysis*
- Technical seminar talk (40%) – (individual) talk with individual grade
  - *depth of detail based on your technical preparation and clarity*
- Participation in Seminar discussions (10%)

Questions?

My Background *(for this course)*

- Research in Process Mining since 2019, Process Automation since 2007
- Contributed to technical assessment of the Process Mining startup *MyInvenio* acquisition by IBM
- BPMN co-author
- 15+yrs Research in Concurrency / Process Modeling

Process Mining – First Steps

and some tools

Outline

1. Introduction

2. First Steps
   - Basic Concepts
   - First Log Exploration
   - Assignment 1 Introduction

Basic Concepts

and some tools

Event Log

Case ID | Timestamp | Activity | Attributes

Instances | Events

	CaseID	Timestamp	Medium	Activity	Service Line	Urgency
1	CaseID	Timestamp	Medium	Activity	Service Line	Urgency
2	case9700	20.8.09 11:46	Phone	Registered	1st line	0
3	case9700	20.8.09 11:50	Phone	Completed	1st line	0
4	case9701	23.9.09 12:23	Phone	Registered	1st line	0
5	case9701	23.9.09 12:27	Phone	Completed	1st line	0
6	case9705	20.10.09 14:21	Phone	Registered	Specialist	2
7	case9705	20.10.09 16:48	Phone	At specialist	Specialist	2
8	case9705	19.11.09 10:31	Phone	In progress	Specialist	2
9	case9705	19.11.09 10:32	Phone	Completed	Specialist	2
10	case3939	15.10.09 11:48	Mail	Registered	Specialist	2
11	case3939	15.10.09 11:48	Mail	Offered	Specialist	2
12	case3939	20.10.09 17:18	Mail	In progress	Specialist	2
13	case3939	20.10.09 17:19	Mail	At specialist	Specialist	2
14	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
15	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
16	case3939	28.10.09 10:17	Mail	In progress	Specialist	2
17	case3939	28.10.09 10:18	Mail	Completed	Specialist	2
18	case9704	20.10.09 14:19	Mail	Registered	1st line	0
19	case9704	20.10.09 14:24	Mail	Completed	1st line	0
20	case9703	20.10.09 14:40	Phone	Registered	1st line	0
21	case9703	20.10.09 14:58	Phone	Completed	1st line	0
22	case9702	24.8.09 12:24	Mail	Registered	2nd line	2
23	case9702	24.8.09 12:30	Mail	Offered	2nd line	2

https://fluxicon.com/book/read/dataext/

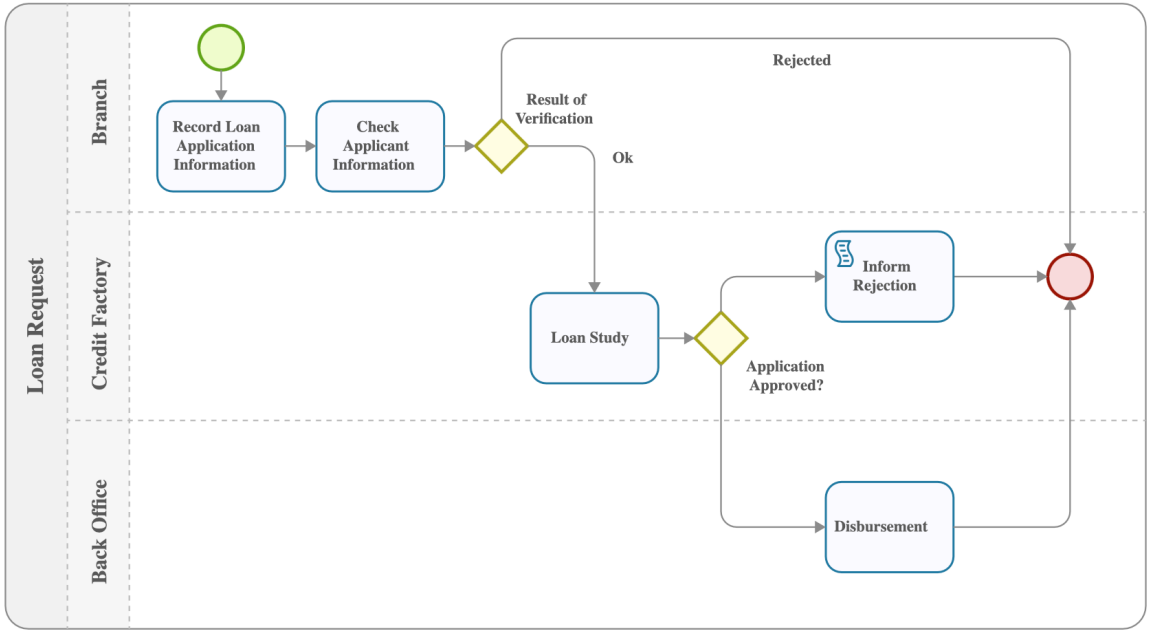
An *event log* is a (multi-)set of events.

Each event has a *case id*, at least one *timestamp* and an *activity* name plus zero or more attributes.

A *case* (a.k.a. *process instance*) is the group of events that have the same case id.

Example:

Event Log from Process Model



Case ID	Applicant	Timestamp	Activity	Department
case4200	Walter White	16.09.2024	Record Loan Application Information	Branch
case4200	Walter White	17.09.2024	Check Applicant Information	Branch
case4711	Don Draper	18.09.2024	Record Loan Application Information	Branch
case4200	Walter White	18.09.2024	Loan Study	Credit Factory
case4711	Don Draper	19.09.2024	Check Applicant Information	Branch
case4711	Don Draper	20.09.2024	Loan Study	Credit Factory
case4200	Walter White	21.09.2024	Inform Rejection	Credit Factory
case4711	Don Draper	21.09.2024	Disbursement	Back Office

First Event Log Exploration (Microsoft Excel)

Open *purchasing-Example.csv* (download from Canvas) in Microsoft Excel (or similar).

- How many cases does the log have?
- How many activities does case 42 have?
- What is the last date of an activity in case 42?
- What is the start date and end date of the entire log?
- How many different values for *Role* occur in the entire log?

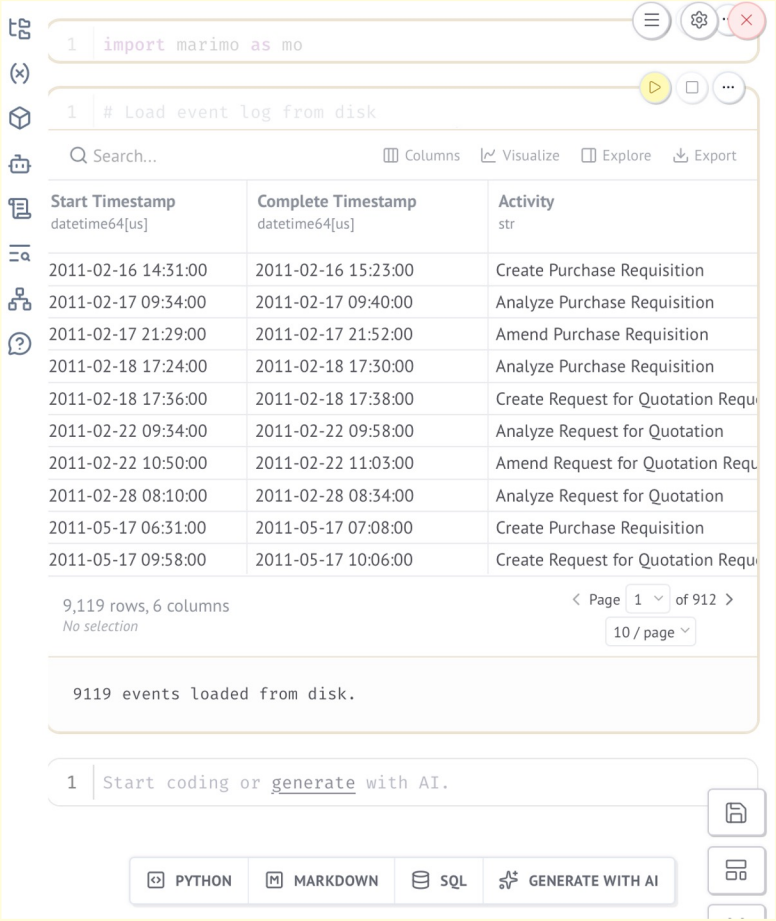
Python Notebook Setup

Recommended setup:

- Marimo
  - *built-in interactivity*
  - *reproducible*
  - *git friendly*

Alternatives:

- Install *(or update)* Visual Studio Code with Jupyter, Data Wrangler, and Copilot plugins.
- JupyterLab
- PyCharm Professional (free for students)
  - *the Community edition does not have Jupyter support*
- Google *Colab*



Marimo

= A *reactive* Python notebook environment

- when a cell changes then all *dependent* cells are automatically re-computed
- cell dependency must be acyclic (DAG of variables/cells)

- rich set of interactive elements out of the box
- annotated Python rather than json → git-friendly
- usage modes: notebook, app, slide

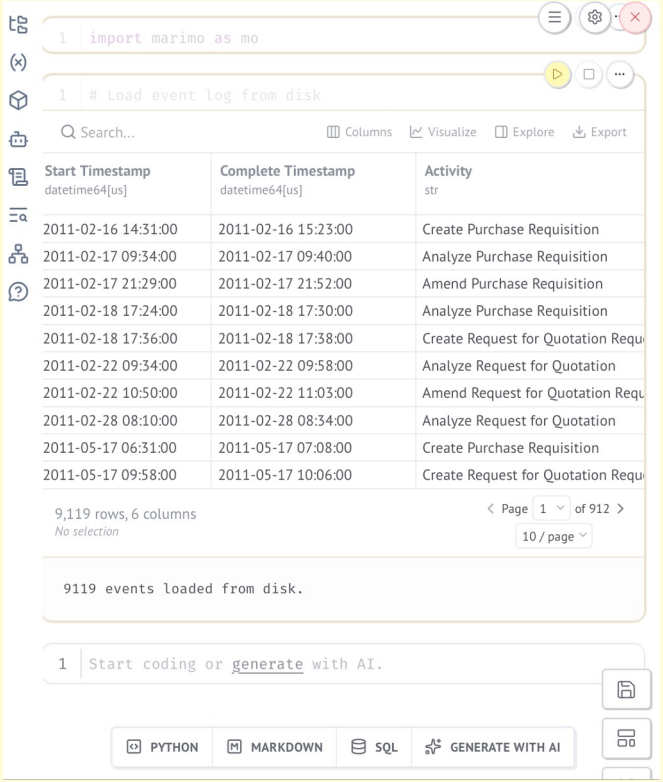
Get started:

```
pip install marimo
marimo edit my_marimo_notebook.py
```

First Event Log Exploration (Python)

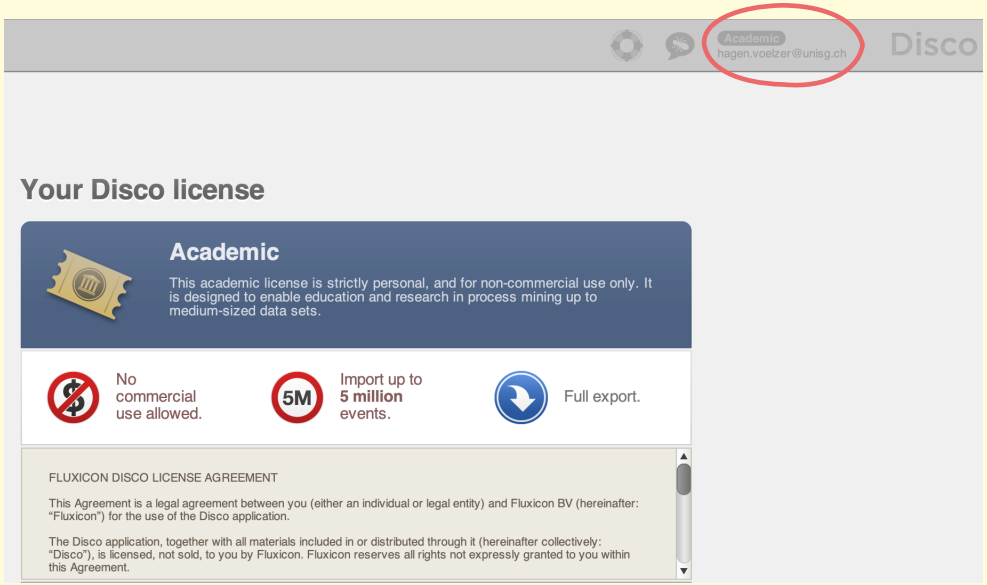
- Open *purchasing-Example.csv* in *pandas*. You can start from the notebook *purchase_example_start.py* (download from Canvas).
  - What is the start date and end date of the entire log?
  - How many activities does case 41 have?
  - How many different values for *Role* occur in the entire log?
  - How many different values for *Resource* occur in the log?
  - How many cases* were touched by *Esmeralda Clay*? In which role(s) did she operate?

* *slightly more difficult*

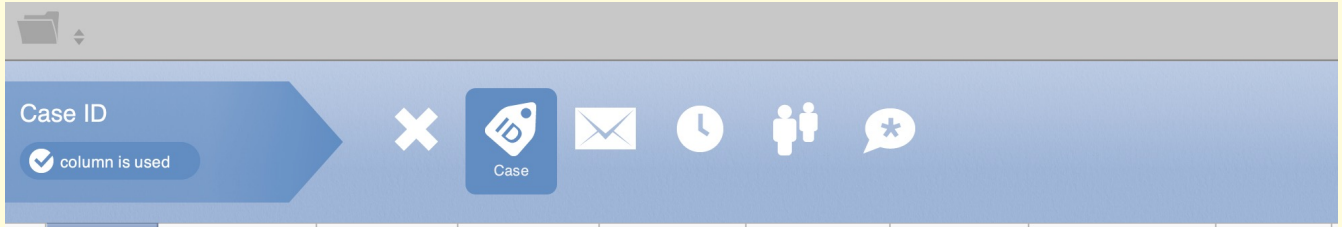


Fluxicon *Disco*

- Download Fluxicon Disco
- Apply for an academic license using your *unisg.ch* e-mail address
  - *This is an automated e-mail roundtrip and should not take long, but you can proceed for now also with the trial license.*
- Open *purchasing-Example.csv* (download from Canvas) in Disco



Disco csv Loader



	Case ID	Activity	Start Date	End Date	Agent Position	Customer ID	Product	Service Type	Resource
1	Case 1	Inbound Call	9.3.10 8:05	9.3.10 8:10	FL	Customer 1	MacBook Pro	Referred to Servicer	Helen
2	Case 1	Handle Case	11.3.10 10:30	11.3.10 10:32	FL	Customer 1	MacBook Pro	Referred to Servicer	Helen
3	Case 1	Call Outbound	11.3.10 11:45	11.3.10 11:52	FL	Customer 1	MacBook Pro	Referred to Servicer	Henk
4	Case 2	Inbound Call	4.3.10 11:43	4.3.10 11:46	FL	Customer 2	MacBook Pro	Referred to Servicer	Susi
5	Case 3	Inbound Call	25.3.10 9:32	25.3.10 9:33	FL	Customer 3	MacBook Pro	Referred to Servicer	Mary
6	Case 4	Inbound Call	6.3.10 11:41	6.3.10 11:51	FL	Customer 4	iPhone	Referred to Servicer	Fred
7	Case 5	Inbound Call	18.3.10 10:54	18.3.10 11:01	FL	Customer 5	MacBook Pro	Product Assistance	Kenny
8	Case 6	Inbound Call	25.3.10 17:09	25.3.10 17:13	FL	Customer 6	MacBook Pro	Referred to Servicer	Harold
9	Case 6	Inbound Call	25.3.10 17:16	25.3.10 17:18	FL	Customer 6	MacBook Pro	Referred to Servicer	Nancy
10	Case 6	Inbound Call	26.3.10 8:36	26.3.10 8:40	FL	Customer 6	MacBook Pro	Referred to Servicer	Elena
11	Case 7	Inbound Call	18.3.10 11:49	18.3.10 11:50	FL	Customer 7	MacBook Pro	Product Assistance	Karen
12	Case 8	Inbound Call	11.3.10 9:20	11.3.10 9:23	FL	Customer 8	MacBook Pro	Referred to Servicer	Karen
13	Case 9	Inbound Email	19.3.10 19:47	21.3.10 8:17	FL	Customer 9	MacBook Pro	Product Assistance	Samuil
14	Case 9	Call Outbound	21.3.10 8:32	21.3.10 8:33	FL	Customer 9	MacBook Pro	Product Assistance	Samuil
15	Case 9	Handle Email	21.3.10 8:33	21.3.10 8:33	FL	Customer 9	MacBook Pro	Product Assistance	Samuil
16	Case 10	Handle Email	27.3.10 11:29	27.3.10 11:30	FL	Customer 10	iPhone	Product Assistance	Jochem
17	Case 11	Inbound Call	27.3.10 8:09	27.3.10 8:11	FL	Customer 11	iPhone	Product Assistance	Irena
18	Case 12	Inbound Call	29.3.10 9:28	29.3.10 9:29	FL	Customer 3	MacBook Pro	Product Assistance	Wil
19	Case 13	Inbound Call	5.3.10 10:13	5.3.10 10:15	FL	Customer 12	MacBook Pro	Product Assistance	Ton
20	Case 14	Inbound Call	4.3.10 7:49	4.3.10 7:50	FL	Customer 3	iPhone	Product Assistance	Wil

load all data attribute columns unless you have a good reason not to

First Event Log Exploration (Fluxicon Disco)

Open *purchasing-Example.csv* in Disco

- What is the start date and end date of the entire log?
- How many activities does case 41 have?
- How many (sequential) variants does the log have?
- How many different activities occur in the log?
- How many different values for Role occur in the entire log?

Assignment 1 Introduction

You are asked to analyze a real-life event log. The event log can be downloaded here.

The event log records the process of managing road traffic fines by a local police in Italy. This process is regulated by Italian law.

A brief description of the process activities and log data attributes follows. This is some, but limited domain knowledge. Feel free to share your own experience or domain knowledge about how traffic fines are handled in Italy.

Otherwise, it is up to our process mining skills to contribute to the **two main goals**:

- understand what precisely has happened according to the log and
- what improvements to the process and to the data collection we can suggest. This includes improvements to what data is collected, in what form it is represented and fixing any data quality issues that you might find.

Process Activity Description

Activity	Description
Create Fine	The initial creation of the fine in the information system. It initializes the event log attributes *amount*, *dismissal*, *points* and *totalPaymentAmount*.
Send Fine	A notification about the fine is sent to the offender.
Insert Fine Notification	The notification is received by the offender.
Add penalty	An additional penalty is applied.
Payment	A payment made by the offender is registered.
Send for Credit Collection	Unpaid fines are sent for credit collection. A separate process is started by a collection agency to collect the money of the unpaid fines.
Insert Date Appeal to Prefecture	The offender appeals against the fine to the prefecture*.
Send Appeal to Prefecture	The appeal is sent to the prefecture by the local police.
Receive Result Appeal from Prefecture	The local police receives the result of the appeal. **
Notify Result Appeal to Offender	The local police informs the offender of the appeal result. **
Appeal to Judge	The offender appeals against the fine to a judge.

** A prefecture in Italy is an administrative body representing the national government in each province.*

*** If the prefecture dismisses the fine, the appeal is deemed accepted, and the obligation to pay the fine is cancelled.*

Event Log Attributes Description

Attribute	Description
amount	The amount due to be paid for the fine (in case a penalty is added, the value includes the penalty amount as well as the initial fine amount).
article	The number of the article of the Italian road-traffic law that is violated by the offender (e.g., article 157 refers to stopping and parking vehicles).
dismissal	A flag indicating whether and by whom the fine is dismissed. **It is initialized to NIL on creation of the fine**. We know the meaning of: - G: dismissed by the judge - #: dismissed by the prefecture
expense	The additional amount due to be paid for postal expenses.
lastSent	N/A
matricola	N/A
notificationType	A flag indicating to whom the fine refers. P: car owner, C: car driver.
resource	A numeric code indicating the employee who handled the case in the corresponding activity.
paymentAmount	The amount paid by the offender in one transaction.
points	Penalty points deducted from the driving license. In Italy, each driver starts with 20 points on their license and may loose points for each offence, based on the gravity.
totalPaymentAmount	The cumulative amount paid by the offender.
vehicleClass	A flag indicating the kind of vehicle driven or owned by the offender.

More General Instructions for Assignment 1

Use a Python notebook (template on Canvas) for the assignment as your main report (which you submit after weeks 3 and 6). Besides the code, your textual answers, the generated output such as charts, you can also include screenshots from external tools. Structure the notebook for readability, cf. screenshot.

By default, create Python code to answer any question unless indicated otherwise in the task explicitly.

Before executing any code, form a *hypothesis*, i.e., think about what the answer could be or in what range you expect the answer to be and compare with the answer you are getting from the code. If you find the answer surprising, say so in your report. You might also want to double check the correctness of your code. For some tasks, we ask you to write down a hypothesis explicitly.

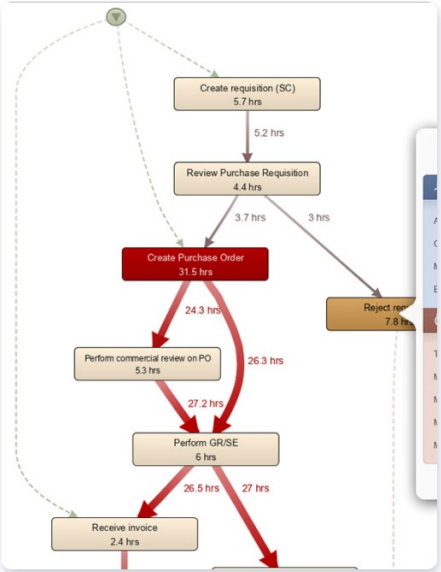
If the answer raises a new question, also write the question down, even if you might not have time to investigate it. If the answer implies some suggestion for improving the process or the data collection (or the data representation), write it down as a finding. Sometimes, we ask explicitly for such a reflection, but even if we don't, it will help to solve some tasks later if you do.

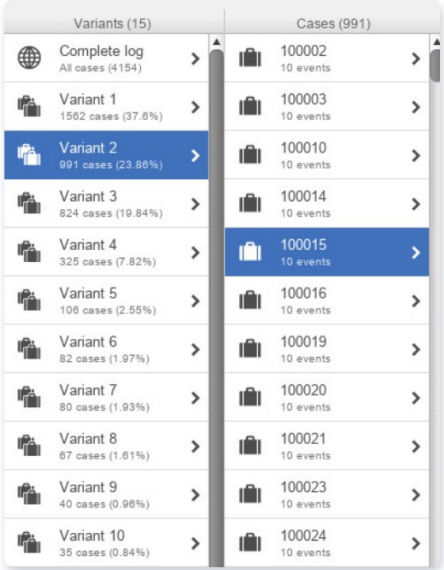
Assignment 1: Tasks from Session 1 (First Event Log Exploration)

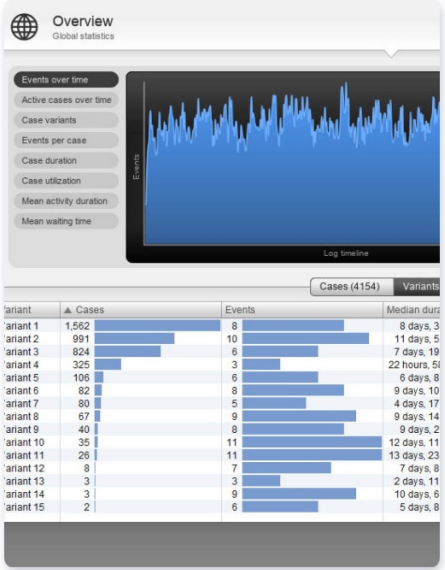
1. Load the event log using pm4py (see assignment template on canvas). Answer the following questions *[Code optional]*:
   a) What time interval does the log cover?
   b) How many distinct values for the attribute *vehicleClass* does the log have?
   c) What are *min, median, max* of the *initial value* of the attribute *amount* (associated with a *Create Fine* event) ? Reflect the result: what do you recommend to be investigated further?
   d) How many events have a value > 0 for the attribute *points*? What are the associated activities? How many case are affected?

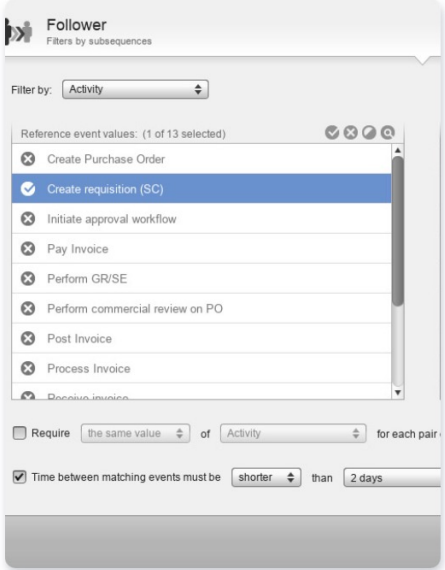
Gallery

Disco



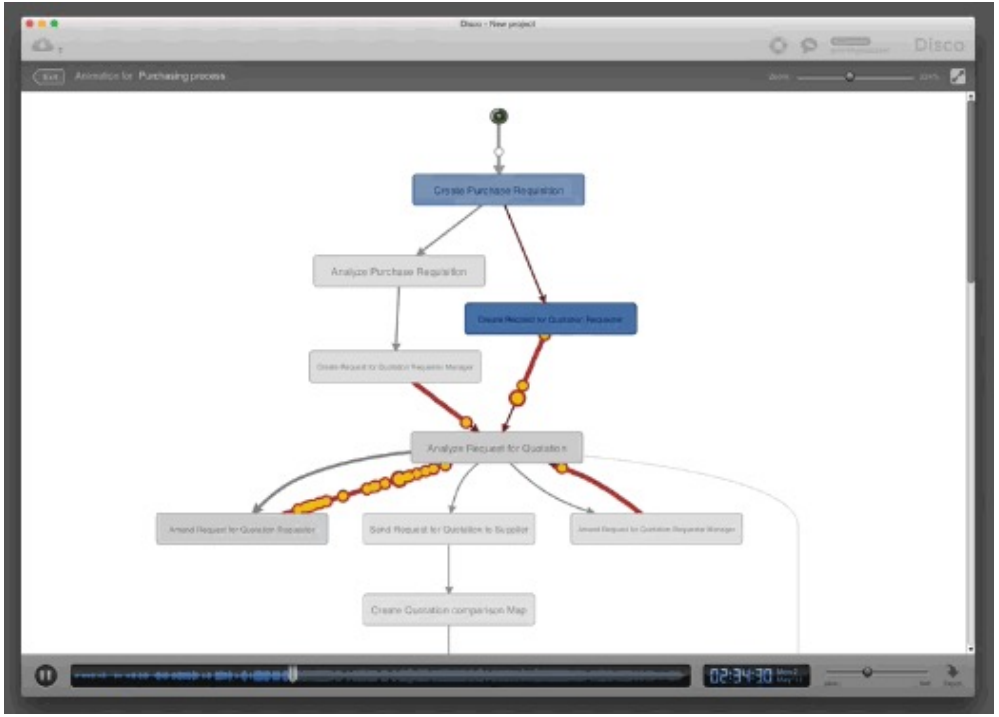




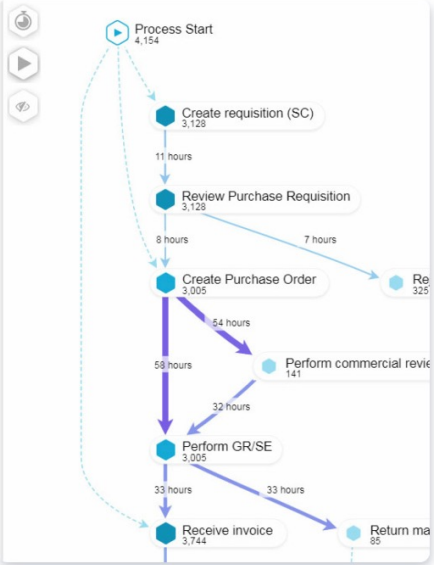


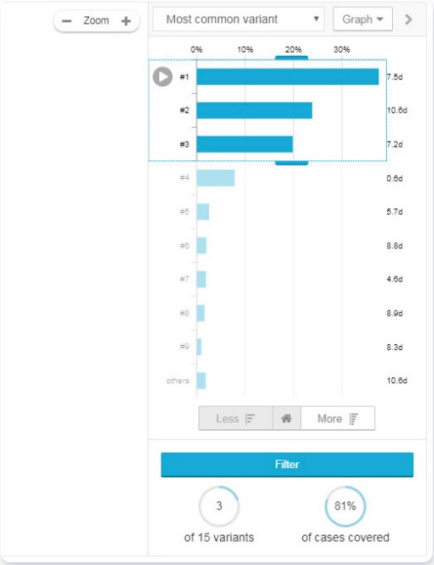
Source and more info: PM Software

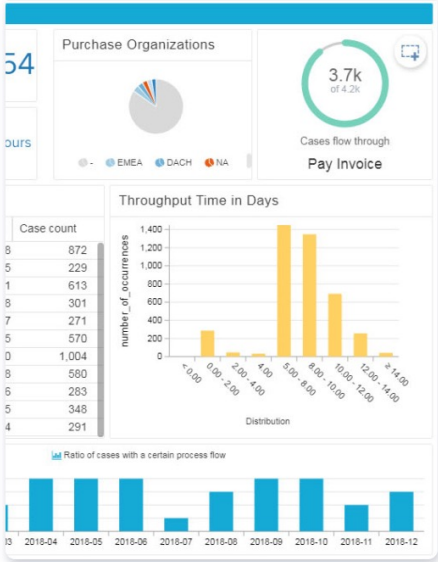
Disco Animation Example

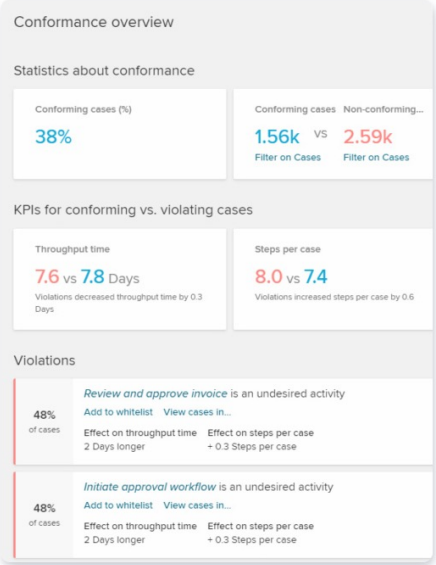


Celonis



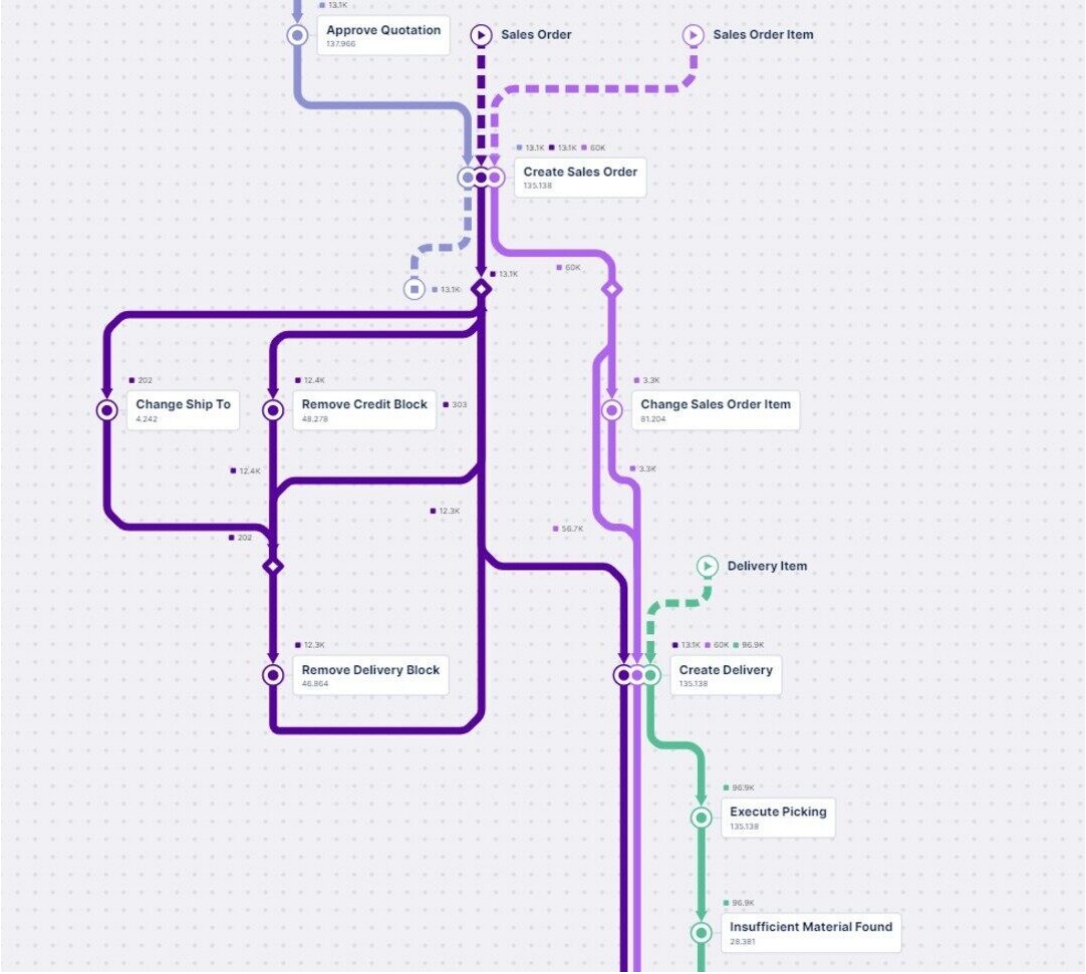






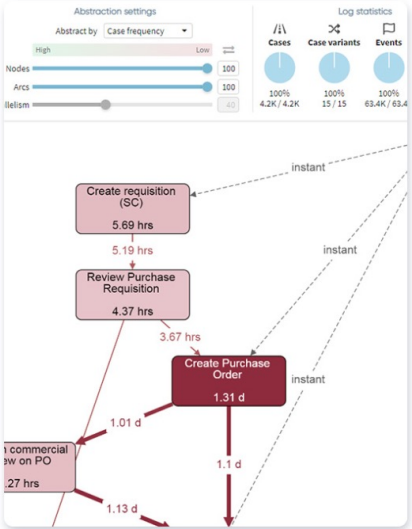
Source and more info: PM Software

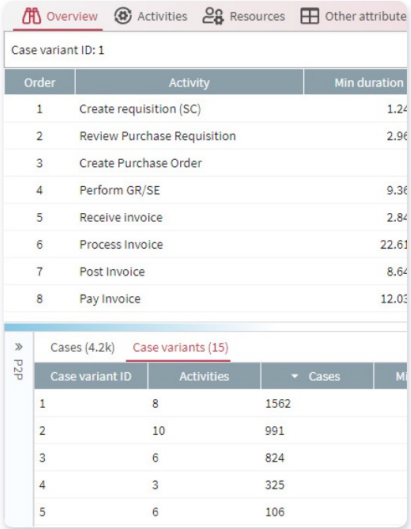
Celonis Object Centric Process Mining

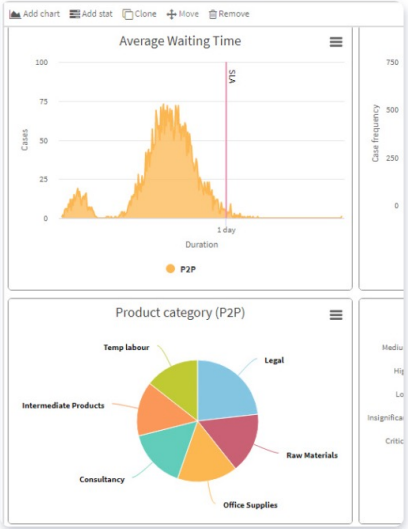


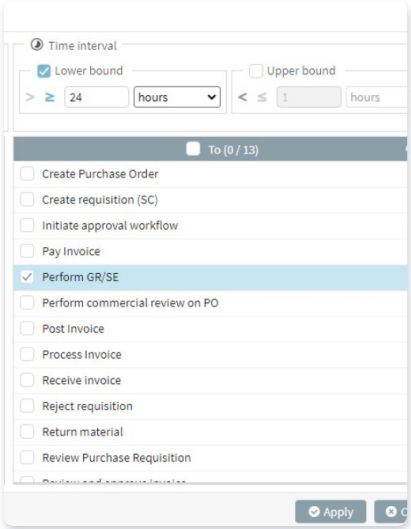
Source: https://www.pads.rwth-aachen.de/cms/pads/der-lehrstuhl/aktuelle-meldungen/~yevvy/celonis-process-sphere-provides-easy-acc

Apromore









Source and more info: PM Software