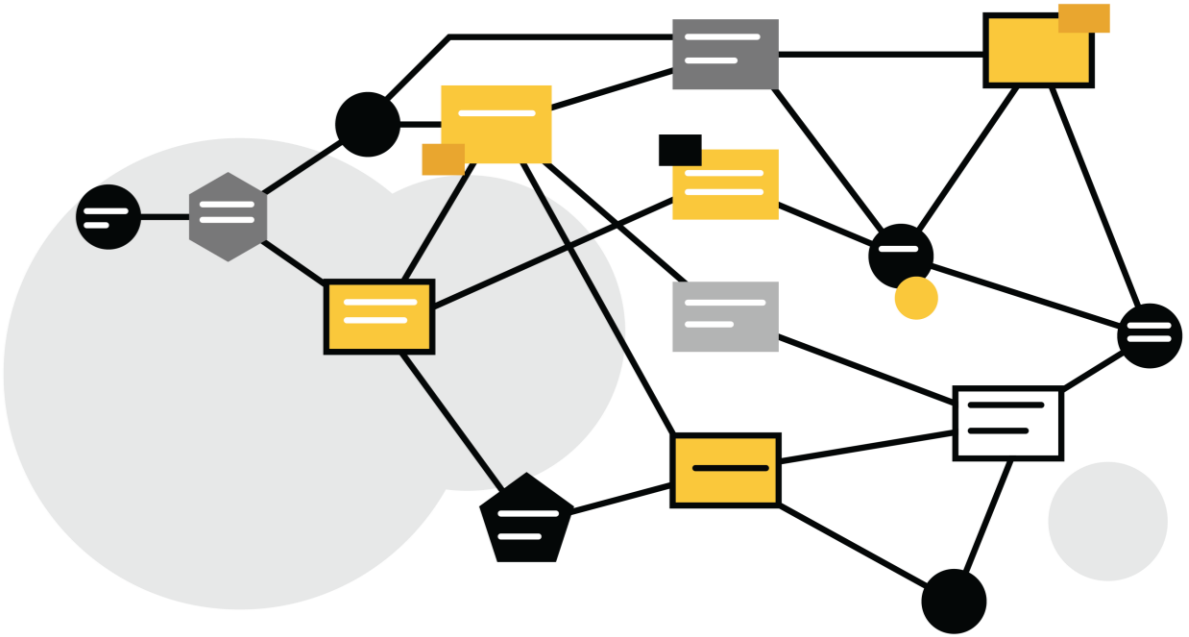
Project: Process Enhancement

Victoria Geisel & Natasha Randall



Process Mining – Prof. Dr. Dietlind Zühlke

17.01.2023

Table of Contents

Recap:
Process Enhancement

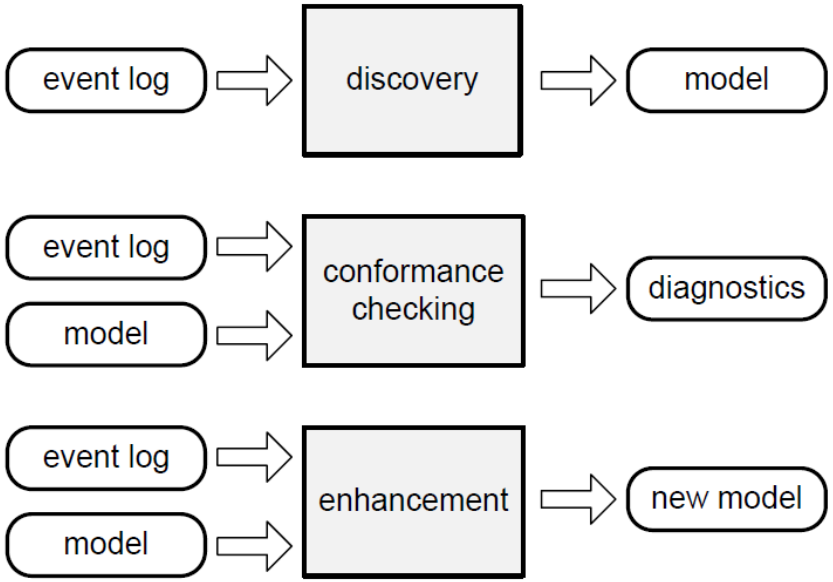
Definition of Process Enhancement

Enhancement

Increase or improve something in value, quality, desirability or attractiveness

Process Enhancement

A process model is extended or improved (repaired) using information extracted from some event log

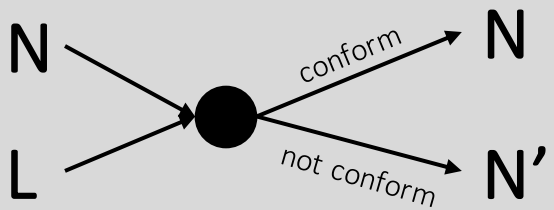


Recap: Model Repair & Extension

Model Repair

change the original model so that it reflects the reality better

- Missing activities in the model
- Missing activities in reality
- discrepancy in execution order



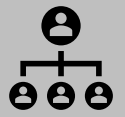
Extension

add a new aspect or perspective to an initial model in addition to the control flow perspective

- time perspective

- case perspective

- organization perspective



Research Hypotheses

Research Hypotheses

Hypothesis 1

Tools like Celonis and ProM can help us enhance a business process model with information on its processing times.

Hypothesis 2

Tools like Celonis and ProM can help us enhance a business process model with information on the company's organization (employees and resources).

Hypothesis 3

Tools like Celonis and ProM can help us adjust an existing process model to better reflect the reality of a business process.

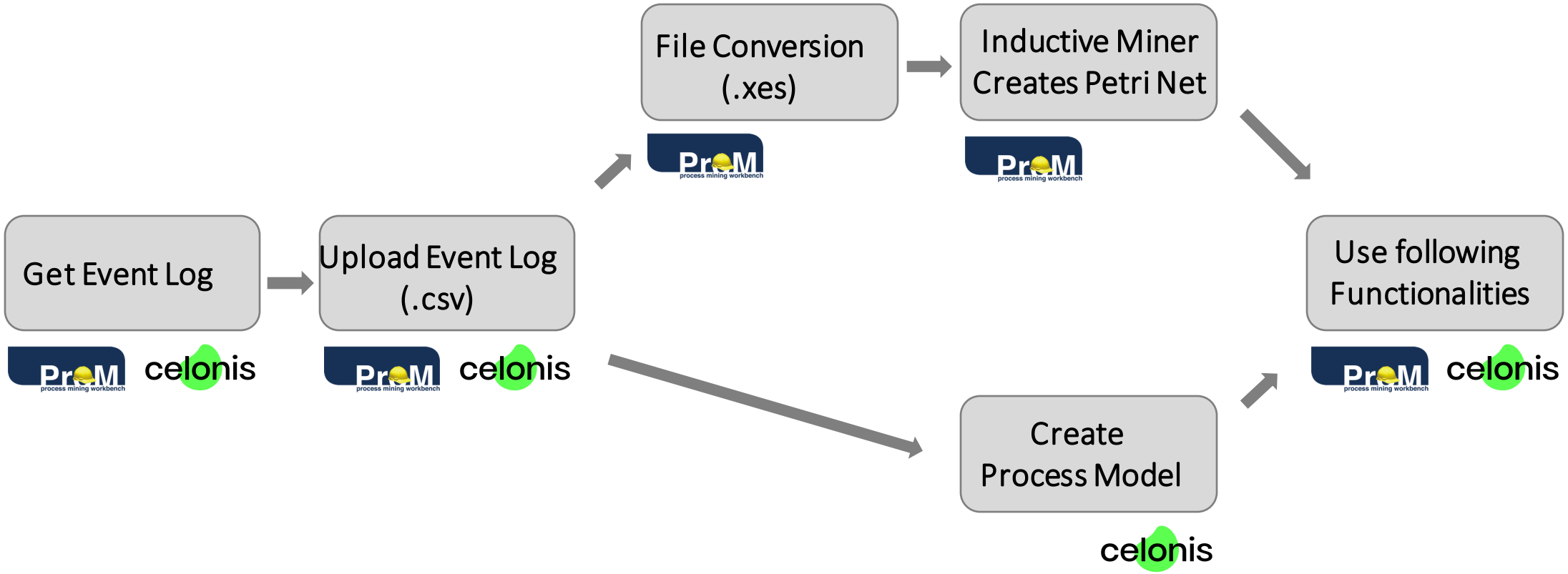
Project Concept

Victoria Geisel, Natasha Randall

PM – Prof. Dr. Dietlind Zühlke

Technology
Arts Sciences
TH Köln

Testing Concept



Enhancement with ProM

organization

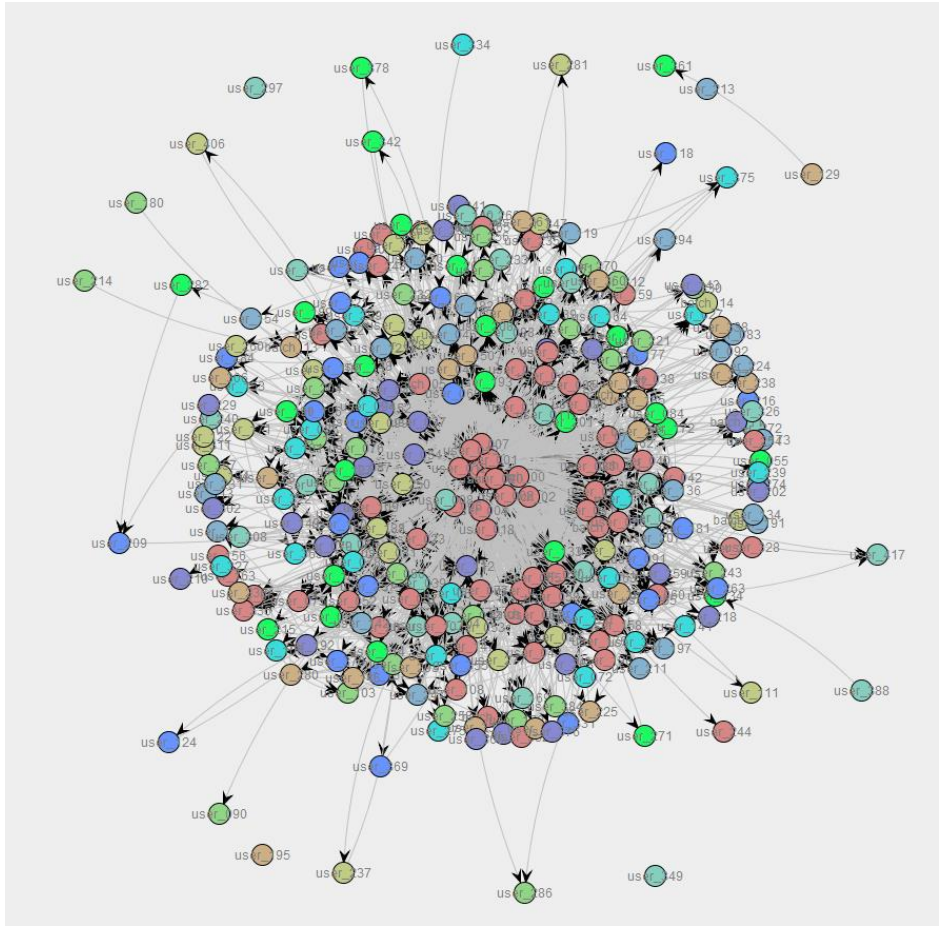
information concerning the **organizational structure**

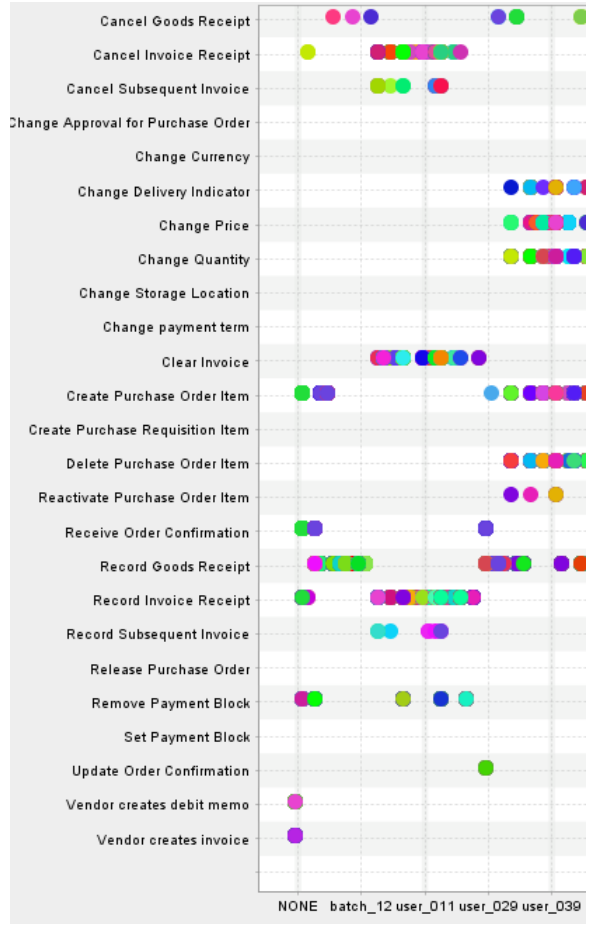
"Social Network Miner"
"Generate Log from Resource Perspective"

Which employees work together? Do employees work on the purchases of one or several vendors?

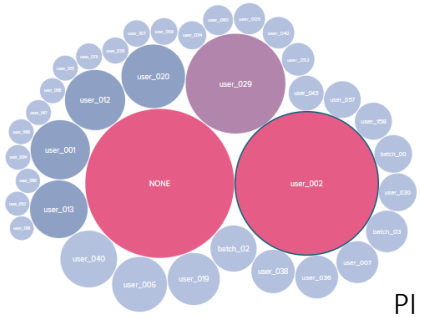
Which process activities are automated and which are handled manually?

Hand-over of work social network & dotted chart





Enhancement with Celonis



celonis

organization

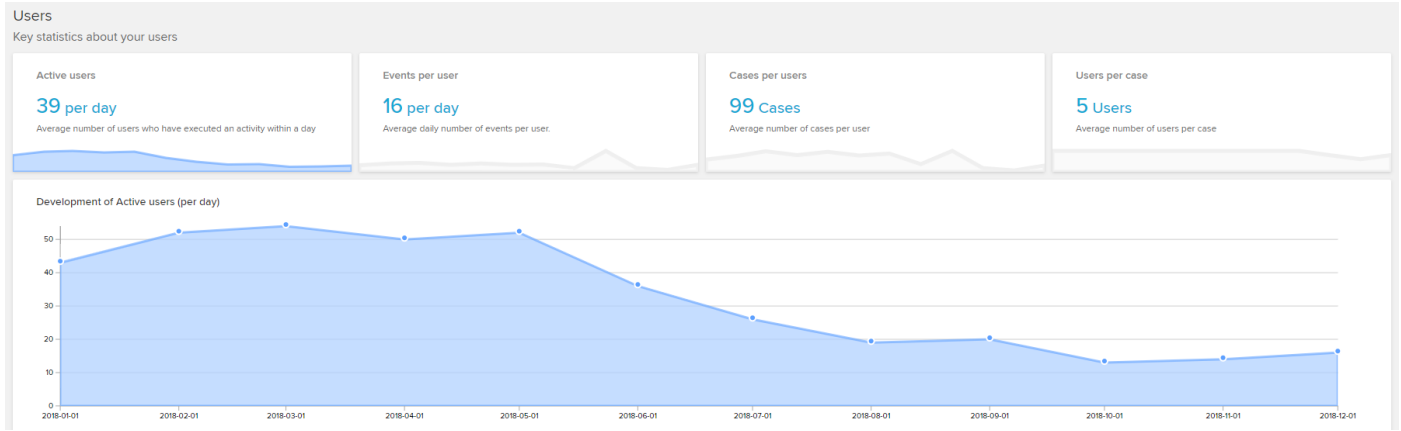
information concerning the organizational structure

"PI Social"

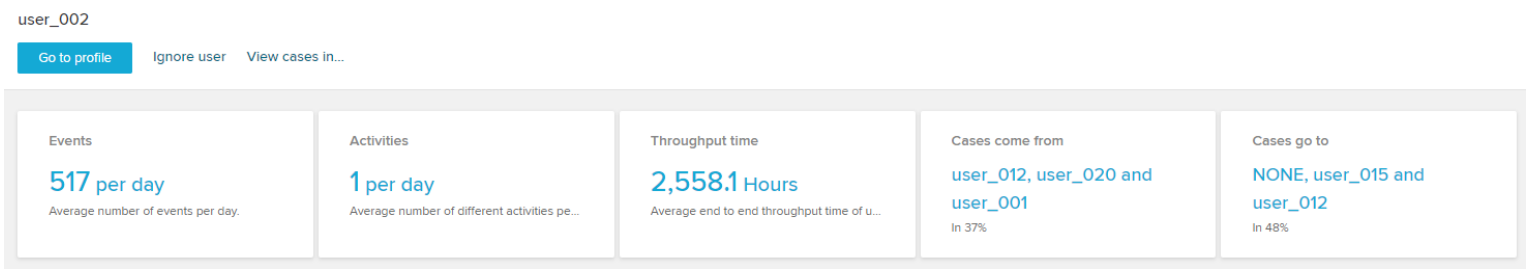
How many employees are working overall on the process each day?

How many activities is an employee working on each day?

PI Social: overview over most active users



PI Social: Analysis overview



PI Social: detailed working information for user_002

Enhancement with ProM

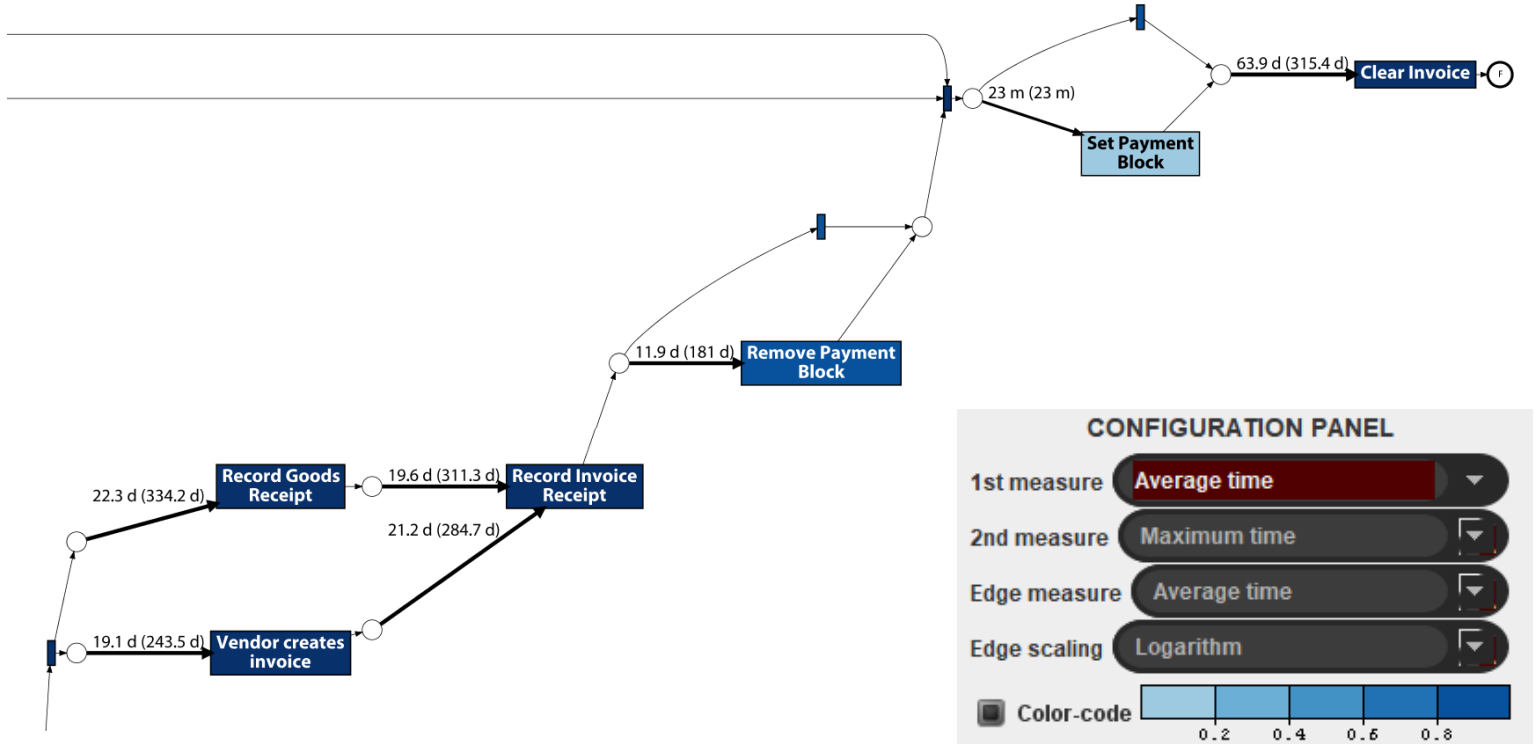
time

information concerning the **process performance**

"Multi-perspective PE"

How long does the process as a whole / a single activity take for execution?

Where are potentials for optimizing the process, because bottlenecks are occurring?



Multi-perspective process explorer – performance view

Enhancement with Celonis

time

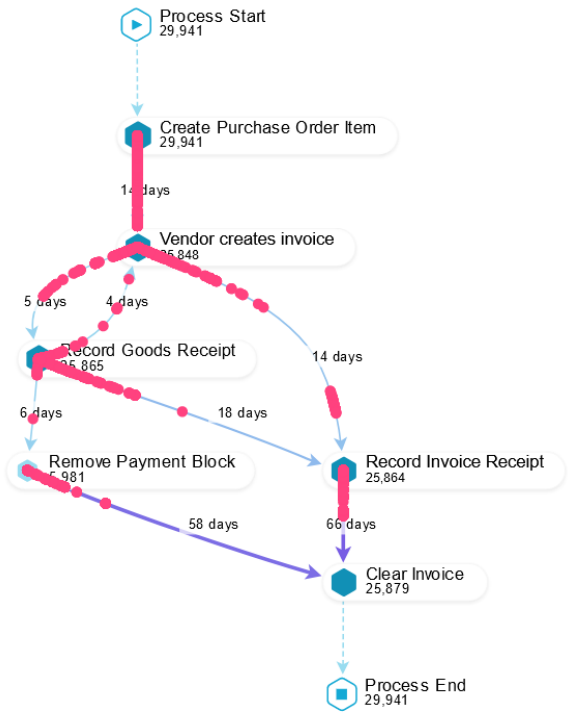
information concerning the process performance

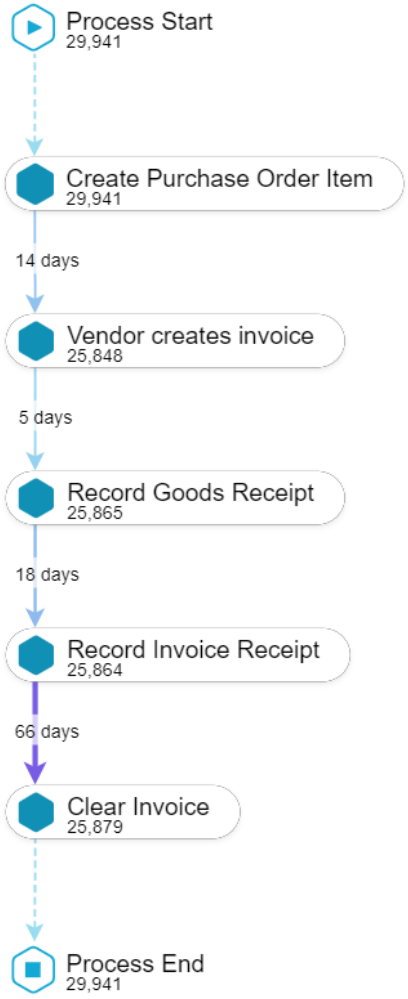
"Throughput time search"

"P. Explorer Animation"

How efficient are the employees throughout the day?

Where are potentials for optimizing the process, because bottlenecks are occurring?





Process Explorer Animation: Average Throughput-Time

Enhancement with ProM

case

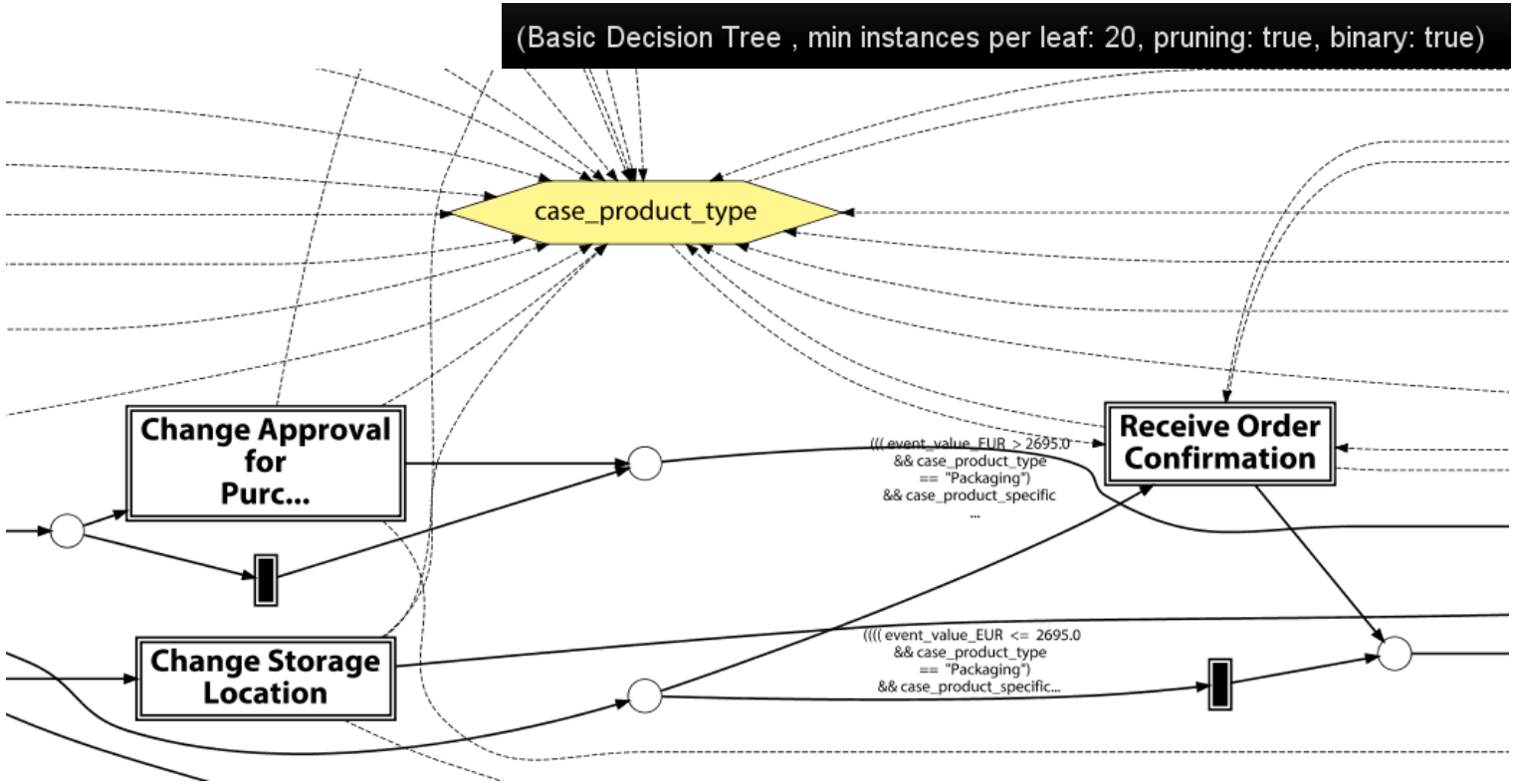
information concerning the **execution patterns**

"Pattern Abstractions"
"Decision Tree Mining"

What are the most frequent paths in the purchase order process?

How high is the rush?

How do case attributes relate to particular paths taken?



Enhancement with Celonis

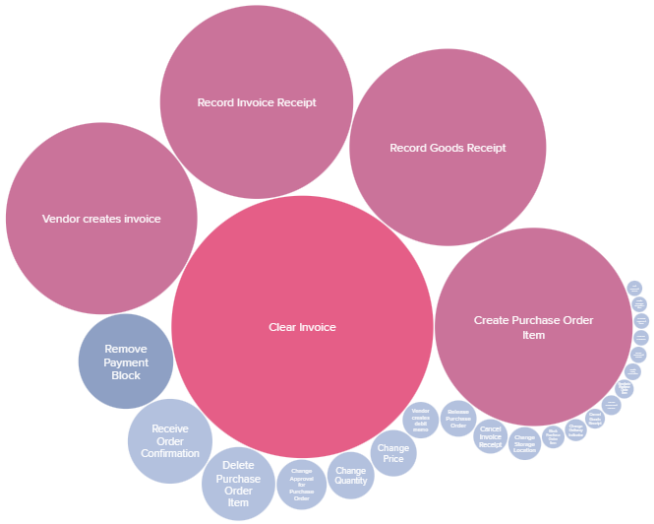
case

information concerning the execution patterns

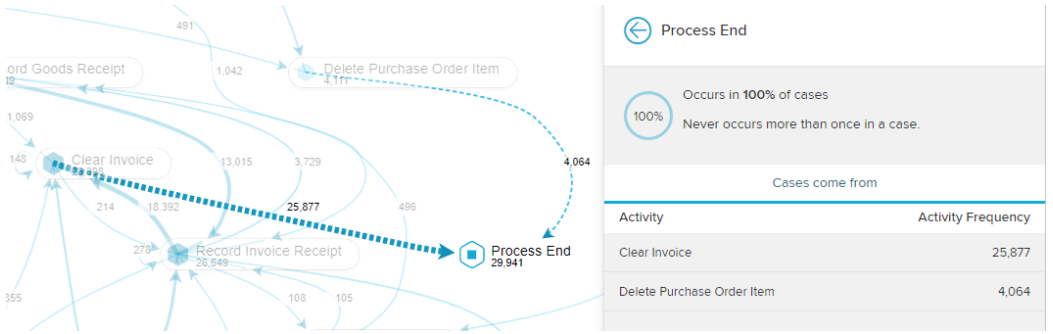
"Process Explorer"
"Variant Explorer"
"Activity Profile (PI Social)"

What is the Happy Path of the purchase order process?

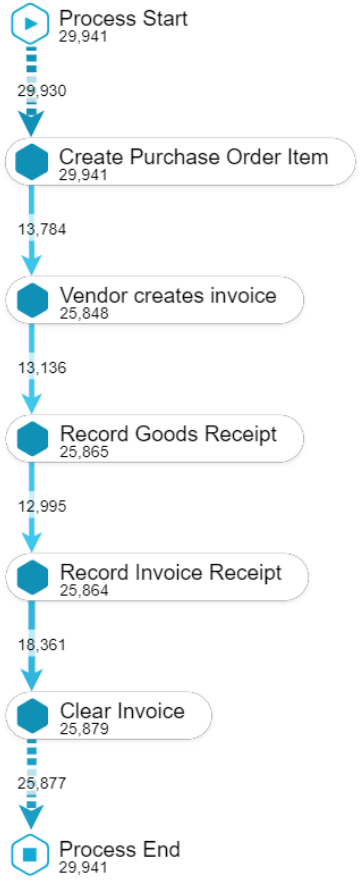
Which activities are executed most often?



PI Social: Activity Profile showing the most executed activities



Process Explorer: Execution analysis for end activity



Process Explorer: Happy Path

Research Questions

What is the distribution of throughput times? Why do throughput times vary?

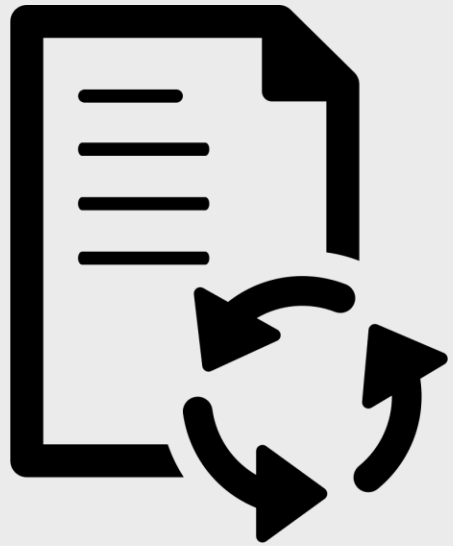
Why this is a useful research question in a business context:

- Time from purchase to pay is common KPI
- Important metric for customer satisfaction
- Can utilise case and organisational perspectives

Why look at throughput on this event log:

- Only one timestamp for each event
- Multiple goods receipt and invoice activities per case

Pre-processing the Event Log



Victoria Geisel, Natasha Randall

PM – Prof. Dr. Dietlind Zühlke

Technology
Arts Sciences
TH Köln

Filtering the Event Log

Filter only one process:

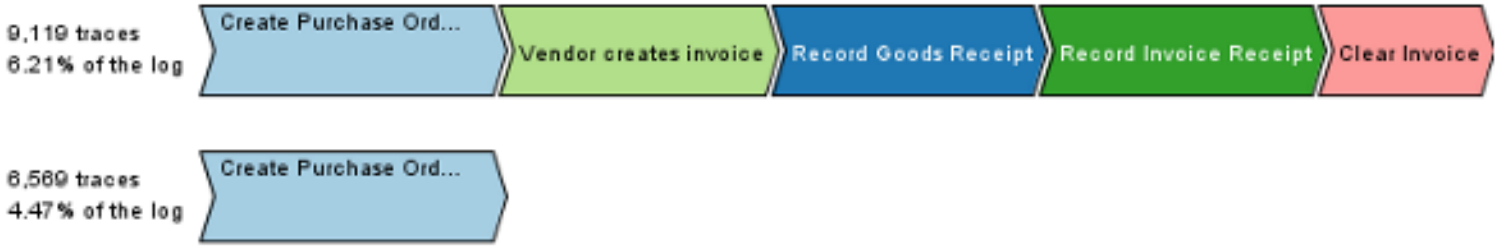
- 4 processes in event log

"an event log contains data related to a single process" (van der Aalst, 2016)

- Chose "3-way match, invoice before goods receipt" (contains 75% events)

Process prioritisation criteria: "Importance" and "Feasibility" (Dumas et al., 2013)

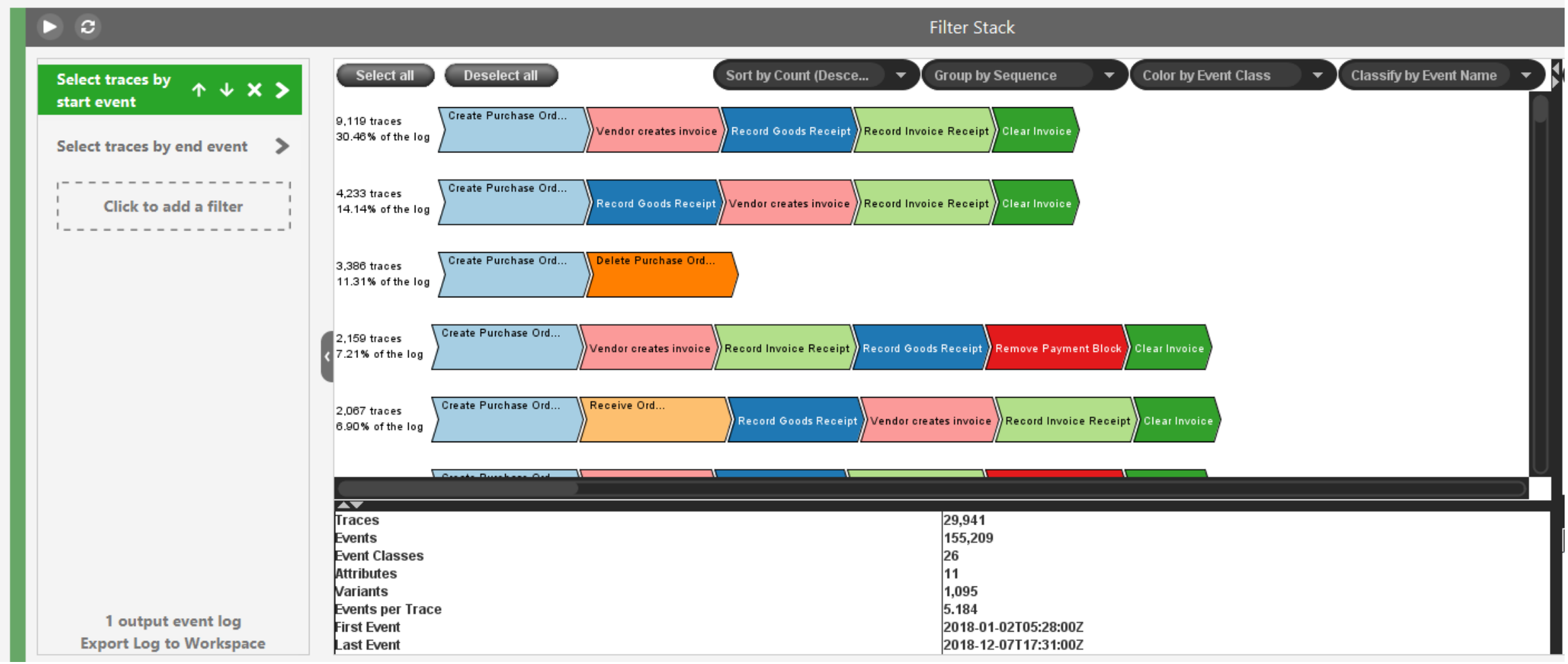
Filter only completed cases:



Top two most common cases in event log (ProM)

Filtering the Event Log: ProM

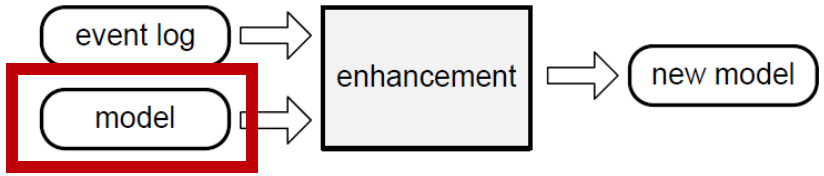
"Filter Event Log" plugin by D. Fahland



Discovering a Process Model

Modelled the data using inductive miner:

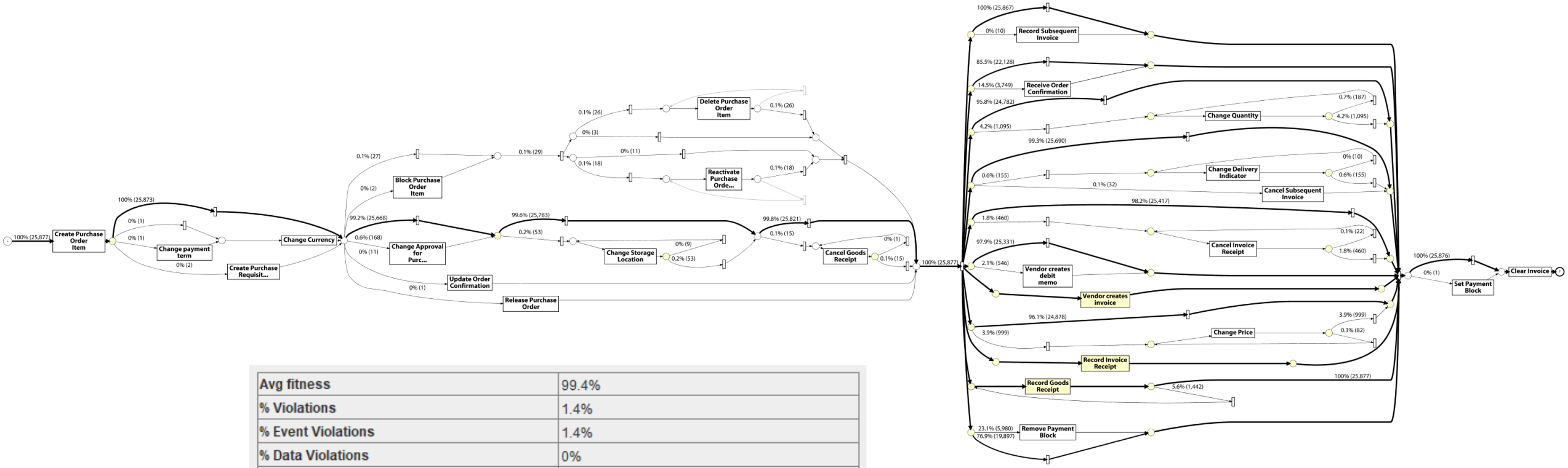
- Inductive miner "can handle infrequent behaviour and huge logs" (van der Aalst, 2016)
- Petri net needed for "Multi-Perspective Process Explorer" plugin



"Mine Petri Net with Inductive Miner"
ProM plugin by S. J. J. Leemans

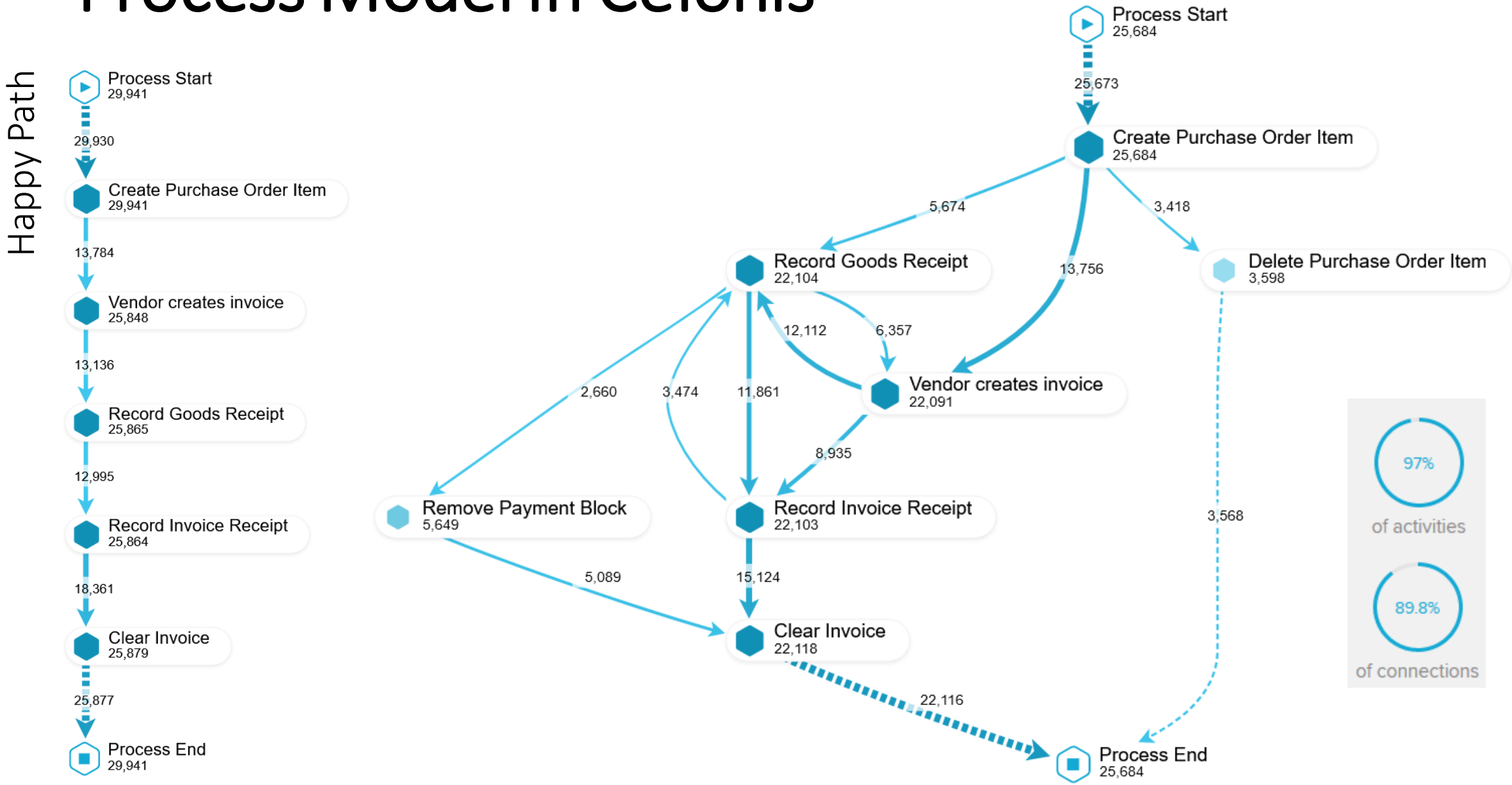
- Need >99% fitness to calculate timing information (Adriansyah & Buijs, 2013)
- Inductive miner run with low noise threshold – strongly overfitting, low precision

Process Model in ProM

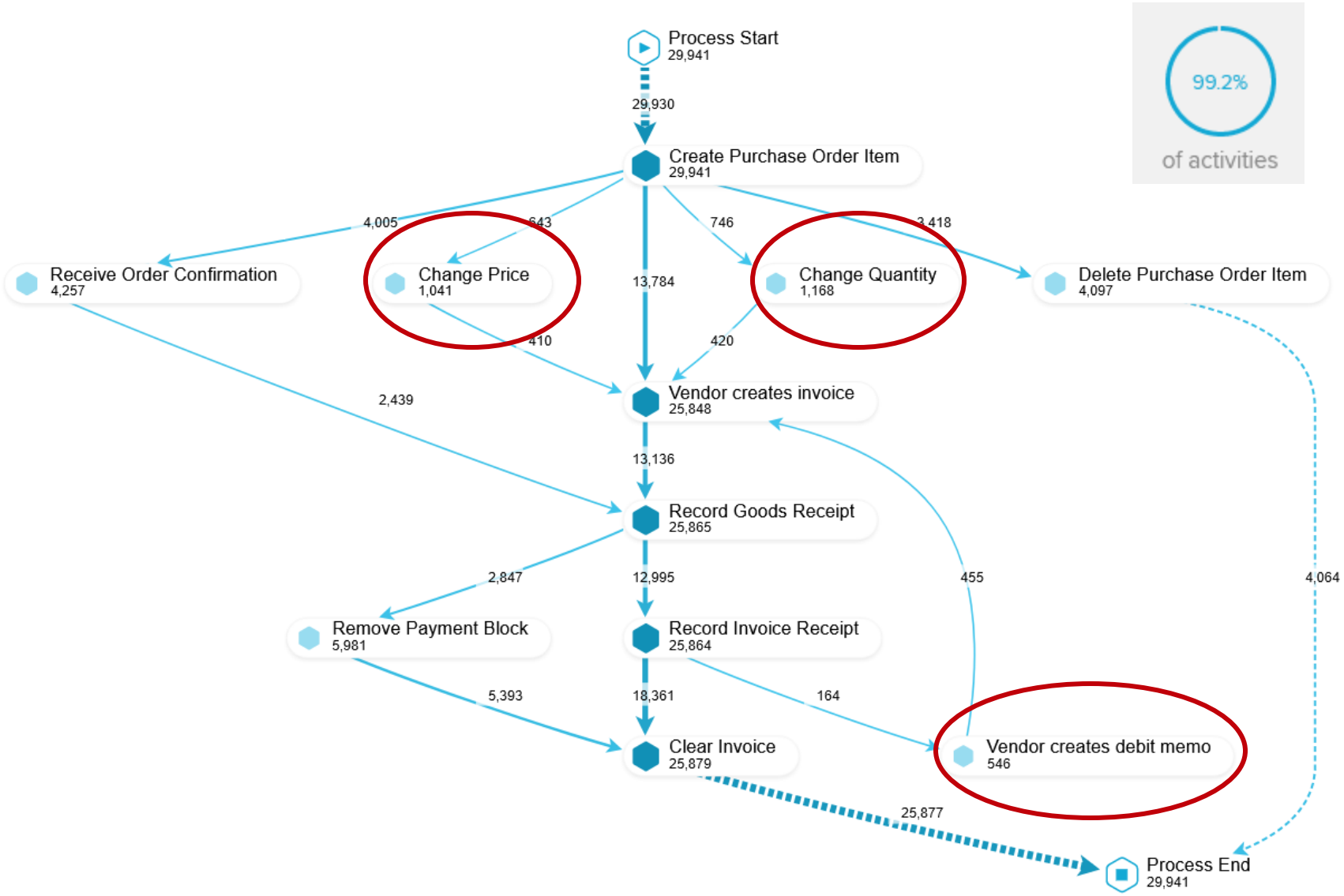


Avg fitness	99.4%
% Violations	1.4%
% Event Violations	1.4%
% Data Violations	0%
# Correct Events	144,375
# Wrong Events	1,898
# Missing Events	89

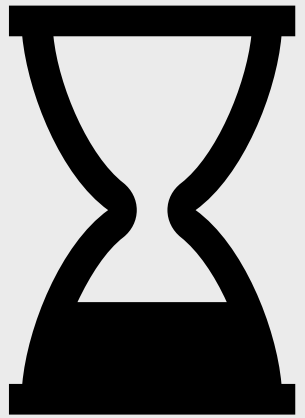
Process Model in Celonis



Process Model in Celonis

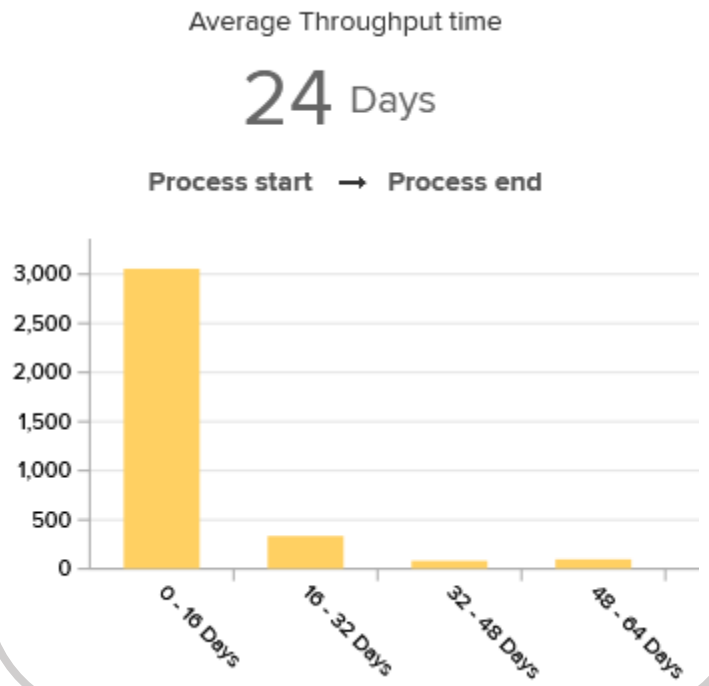


What is the distribution of throughput times?

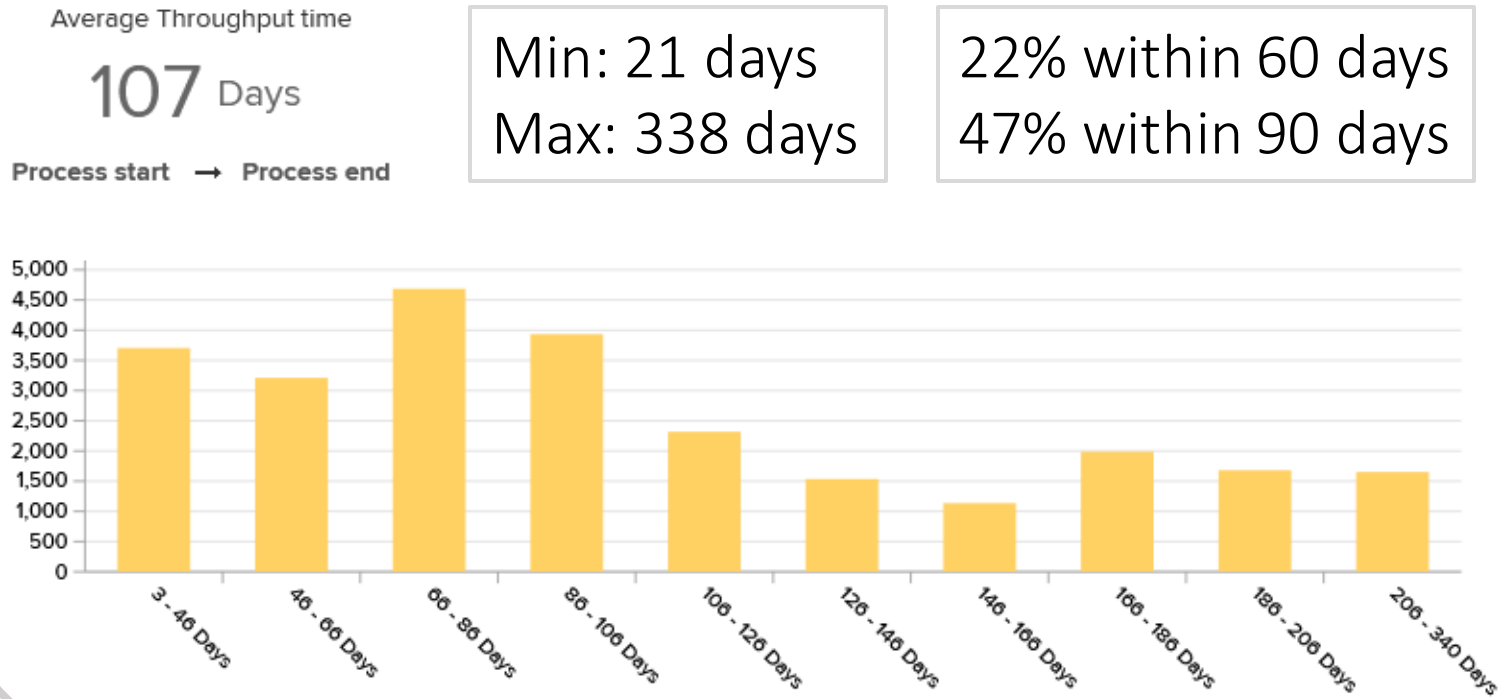


What is the distribution of throughput times?

Cases ending with "delete purchase order":



Cases ending with "clear invoice":



Min: 21 days
Max: 338 days

22% within 60 days
47% within 90 days

Why do throughput times vary?

Why do throughput times vary?

Event Log Attributes

Case attributes: *Example:*

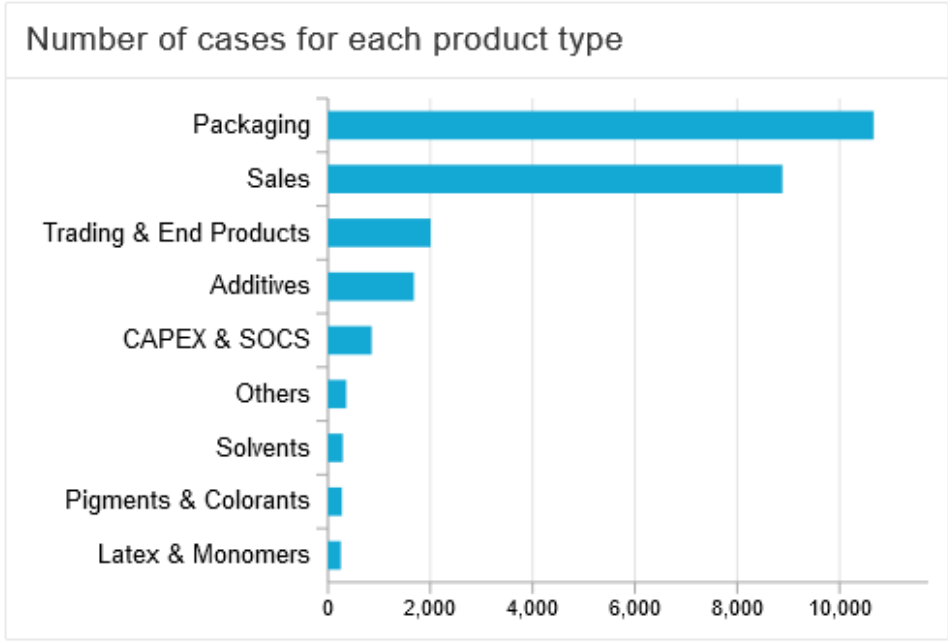
- Product type *(packaging)*
- Value in euros *(€1272)*

Resource attributes:

- Vendors *(vendorID_0136)*
- Employee on activity *(user_055)*

Control flow attributes:

- Payment block occurred *(TRUE)*



Generating Attributes

- Number of rework activities
- Automation rate
- Average vendor workload
- Number of line items

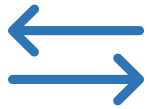
Creating an attribute dataset used to model throughput:

case_PO_id	avg_value_EUR	case_vendor	case_product_type	case_product_specific	rework_activities	payment_block	automation (%)	avg_vendor_workload	num_items	throughput
4507000266	134	vendorID_0103	Packaging	Labels	0	FALSE	0	49.272727	1	84 days 00:21:00
4507000268	300.5	vendorID_0105	Packaging	Labels	0	FALSE	0	13.402778	2	216 days 03:14:00
4507000272	769	vendorID_0137	Sales	Products for Resale	1	FALSE	25	3.647059	2	216 days 03:06:00
4507000273	598.5	vendorID_0138	Sales	Products for Resale	2	TRUE	0	9.121951	2	183 days 01:56:00
4507000287	581.5	vendorID_0148	Sales	Products for Resale	0	TRUE	50	3.714286	3	182 days 02:26:00

Modelling Throughput

```
throughput ~ event_value_EUR + rework_activities + payment_block + automation +
avg_vendor_workload + C(case_product_type) + C(case_product_specific) + C(case_vendor)
```

	coef	p-value
Intercept	110.417	0.000
Product Type: Additives	6.722	0.009
Product Type: Packaging	-6.1567	0.090
Product Type: Sales	15.2249	0.013
Specific Product Type: Containers	23.4936	0.000
Specific Product Type: Extenders	-18.3602	0.000
Specific Product Type: Labels	15.139	0.001
Specific Product Type: Products for Resale	-24.1504	0.000
vendorID_0104	-0.1407	0.982
vendorID_0106	-30.3705	0.000
vendorID_0120	34.2074	0.000
vendorID_0136	50.8701	0.000
vendorID_0171	9.0812	0.015
Value of Purchase Order (EUR)	0.0001	0.055
Number of rework Activities	11.4667	0.000
Payment block occurred	1.2151	0.460
% of automation	-0.2137	0.001
Vendor workload	-0.5985	0.000
Number of line items	0.4596	0.001



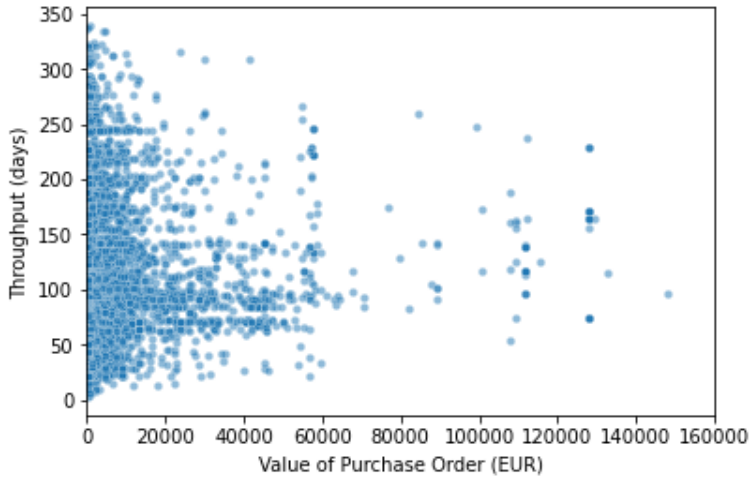
Different product types and different vendors have very different throughput times

Rework activities increase throughput

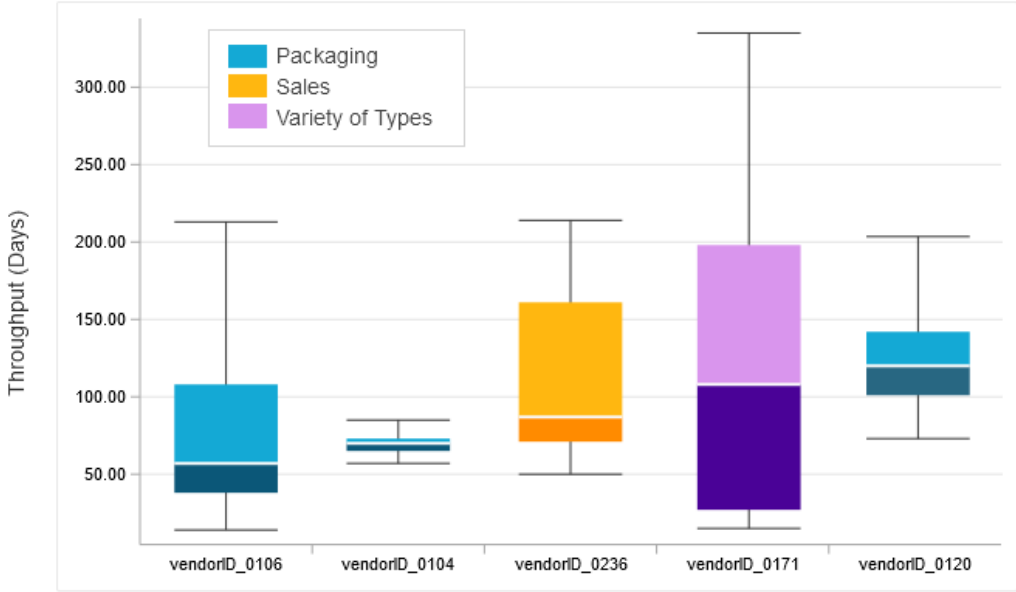
Higher levels of automation decrease throughput

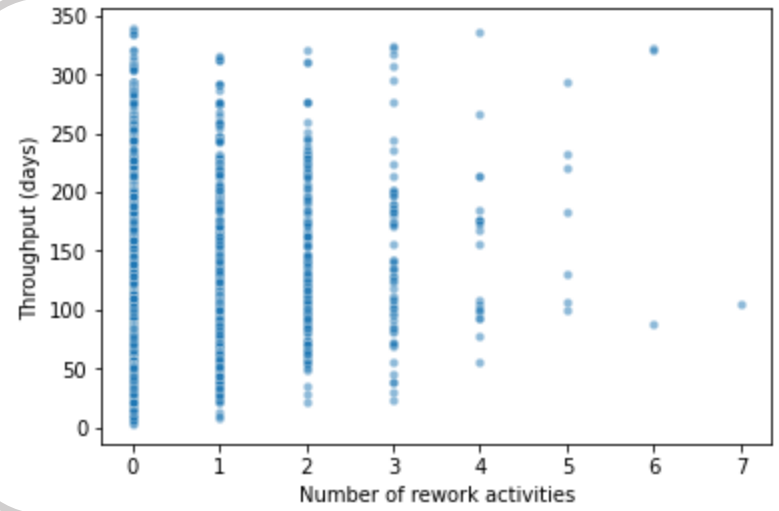
Purchase order values doesn’t affect throughput

No obvious correlation between value of purchase order and throughput time



Vendors working in the same product type area have different averages and variances in throughput times

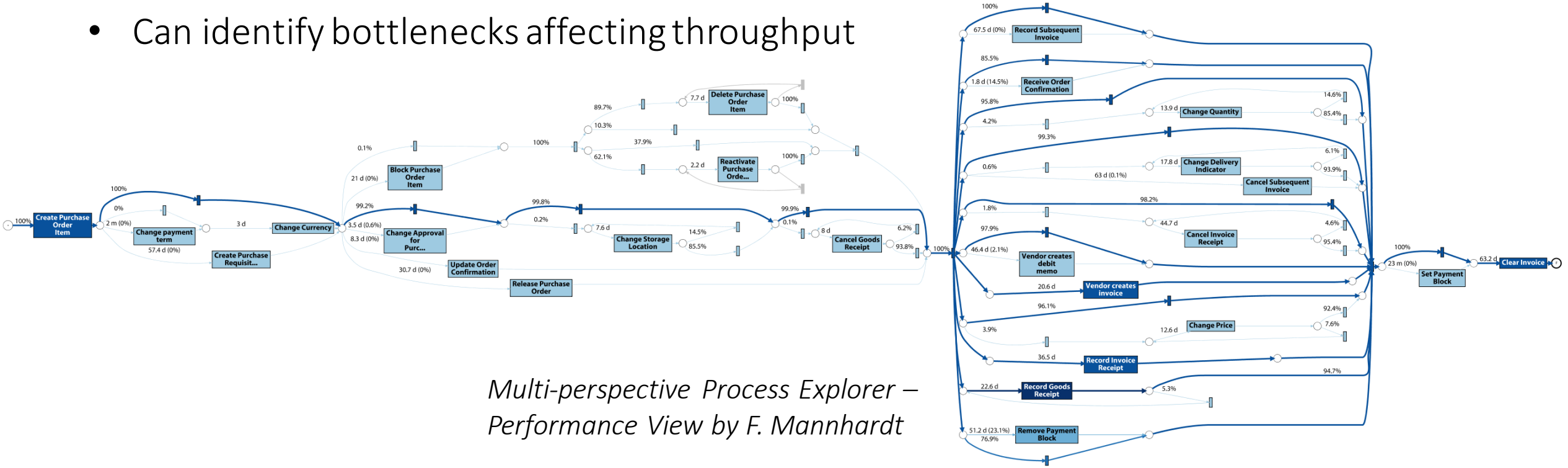




- High number of rework activities lead to higher throughput time
- But very large amount of variation in the plot
- Model does not fit well

Bottlenecks: ProM

- "Multi-perspective Process Explorer" calculates times between activities
- Can identify bottlenecks affecting throughput



Multi-perspective Process Explorer – Performance View by F. Mannhardt

Bottlenecks: Celonis

The largest bottleneck is at clear invoice

Bottlenecks

These connections increase process throughput time considerably

Record Invoice Receipt → Clear Invoice

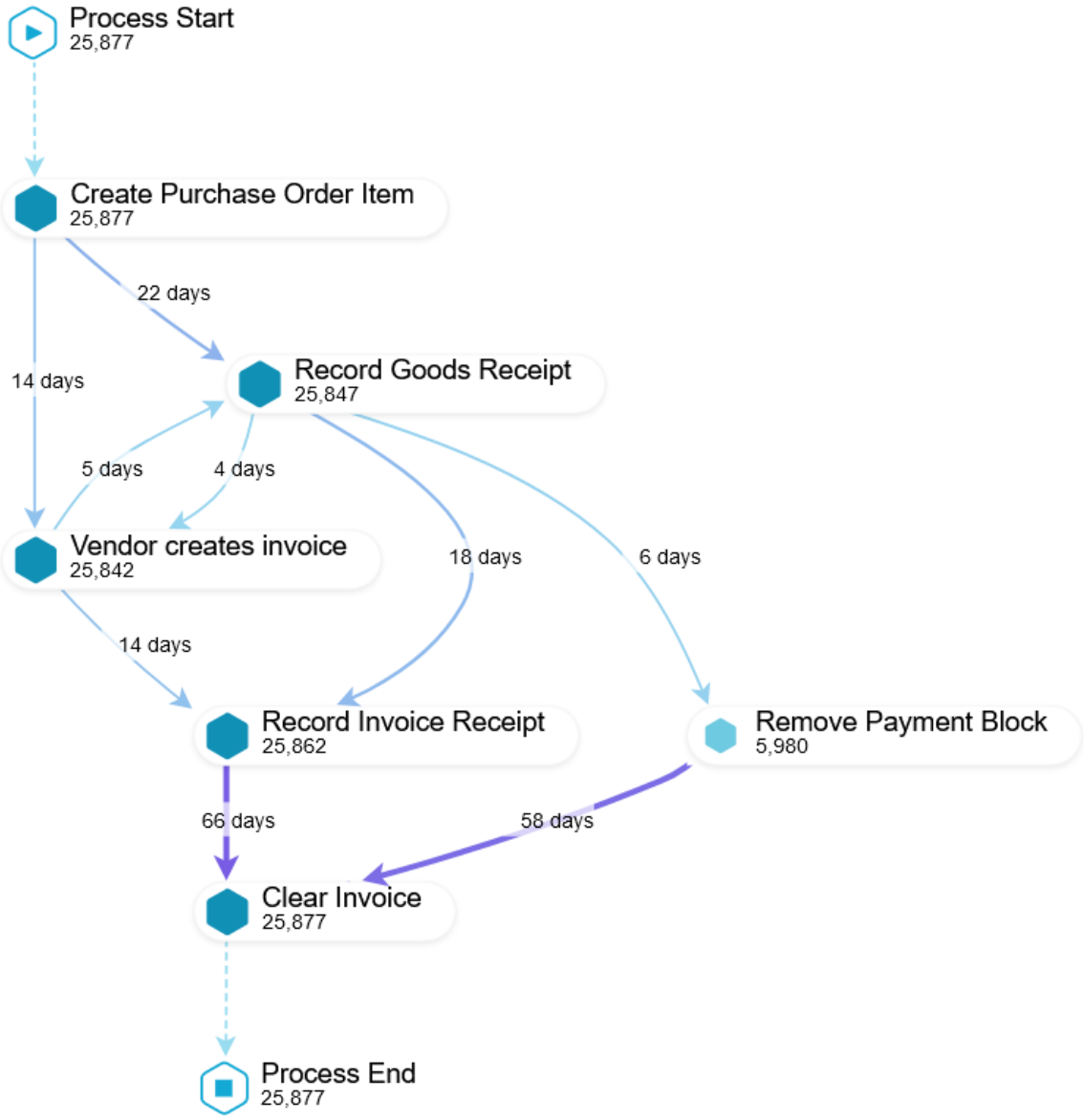
View cases in...

Throughput time	Cases affected
66 work day(s)	71%

Remove Payment Block → Clear Invoice

View cases in...

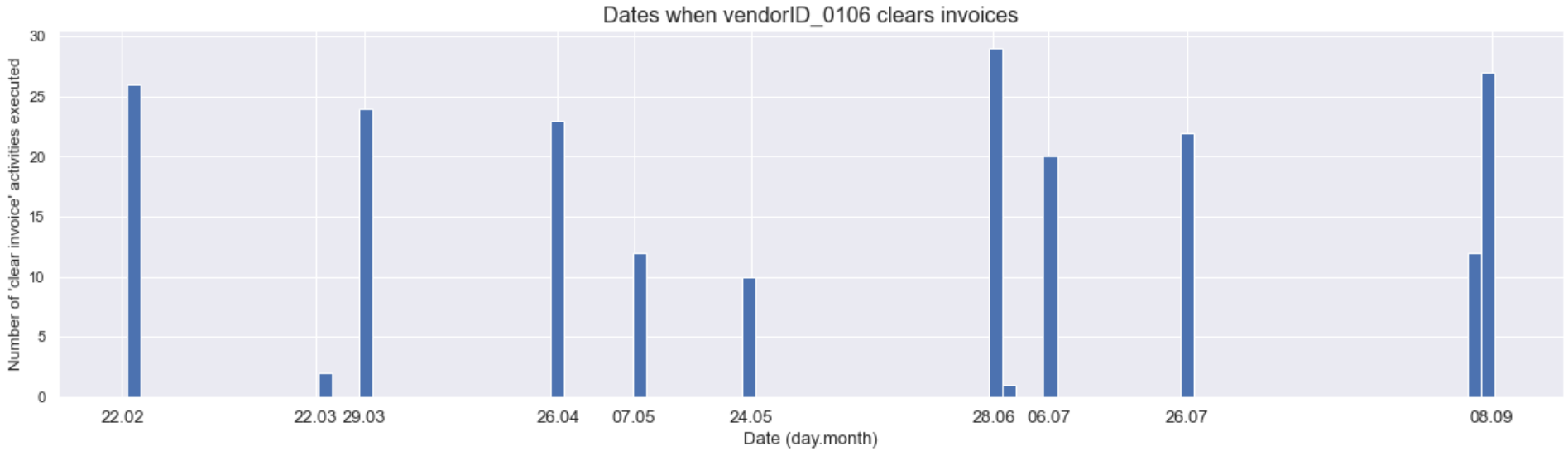
Throughput time	Cases affected
58 work day(s)	21%



Why is “Clear Invoice” a bottleneck?

- Companies delay clearing invoices to improve their own capital
- Vendors only clear invoices once a month in batches

→ This affects throughput calculations



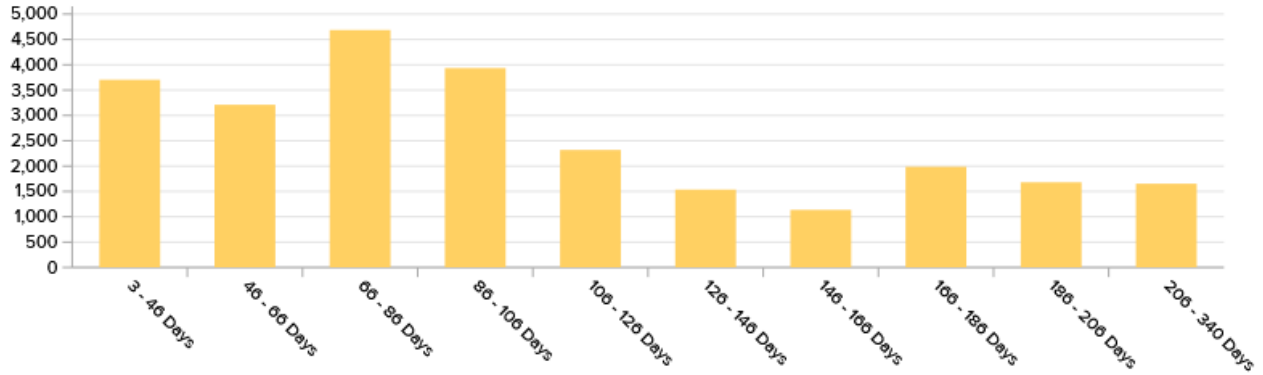
Research Questions Conclusion

What is the distribution of throughput times?

Average Throughput time

107 Days

Process start → Process end



Large variance in timings

Why do throughput times vary?

Can explain variance using event log attributes:

- More rework activities increase throughput
- Increasing automation rate decreases throughput
- Different vendor and product types have very different throughput

Bottleneck analysis revealed strongest influence is delay to 'clear invoice'

Research
Hypothesis
Conclusion

Conclusion

Hypothesis 1

Tools like Celonis and ProM can help us enhance a business process model with information on its processing times.

Research Questions

What is the distribution of throughput times?
Why do throughput times vary?

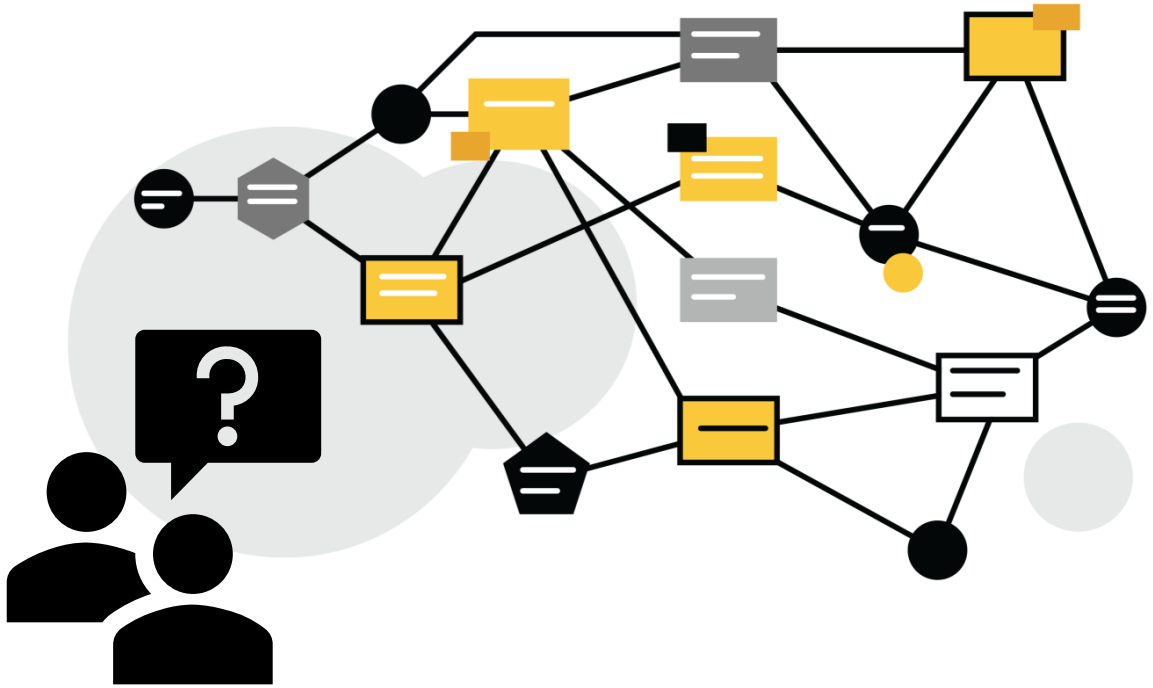
References

Van Dongen, B.F. *Dataset BPI Challenge 2019.* 4TU Centre for Research Data.
https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1
https://icpmconference.org/2019/icpm-2019/contests-challenges/bpi-challenge-2019/

- Van Der Aalst, W. (2012). *Process mining: Overview and opportunities*. ACM Transactions on Management Information Systems (TMIS), 3(2), 1-17.
- Van Der Aalst, W. (2016). *Process mining: Data science in action* (Vol. 2). Heidelberg: Springer.
- Leemans, S. J. (2022). *Robust Process Mining with Guarantees: Process Discovery, Conformance Checking and Enhancement* (Vol. 440). Springer Nature.
- Van Der Aalst, W., Josep Carmona (2022). *Process Mining Handbook.*. Springer
- Van Der Aalst, W., et al. (2012). *Process mining manifesto.* In International conference on business process management (pp. 169-194). Springer, Berlin, Heidelberg.
- Yasmin, F. A. (2019). *Enhancement in process mining: guidelines for process owners and process analysts* (Master's thesis, University of Twente).
- Verbeek, H. M. W., & Bose, R. J. C. (2010). *Prom 6 tutorial.* ProM.(nd). Extraído el, 27.
- Van der Aalst, W. M., Schonenberg, M. H., & Song, M. (2011). *Time prediction based on process mining.* Information systems, 36(2), 450-475.
- Mannhardt, F., De Leoni, M., & Reijers, H. A. (2015). *Extending process logs with events from supplementary sources.* In International conference on business process management (pp. 235-247). Springer, Cham.

Project: Process Enhancement

Victoria Geisel & Natasha Randall



Are there any questions?