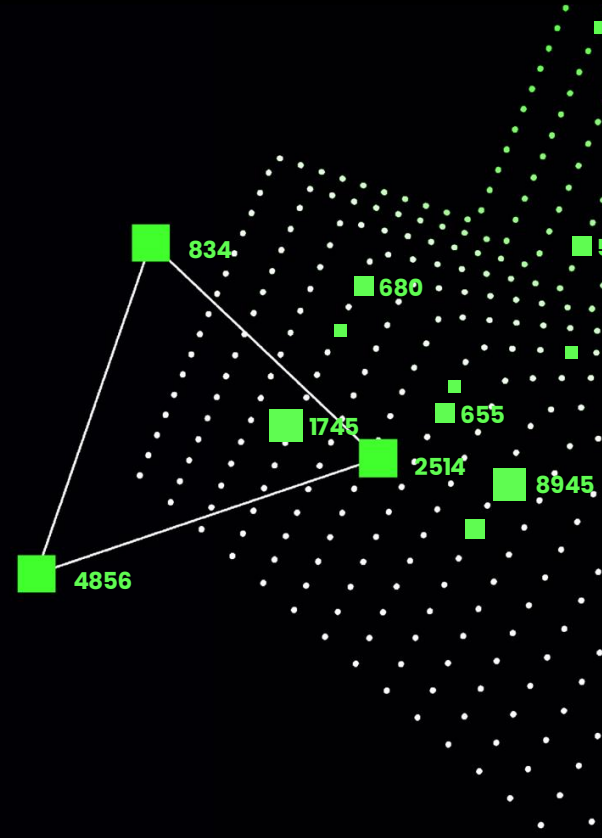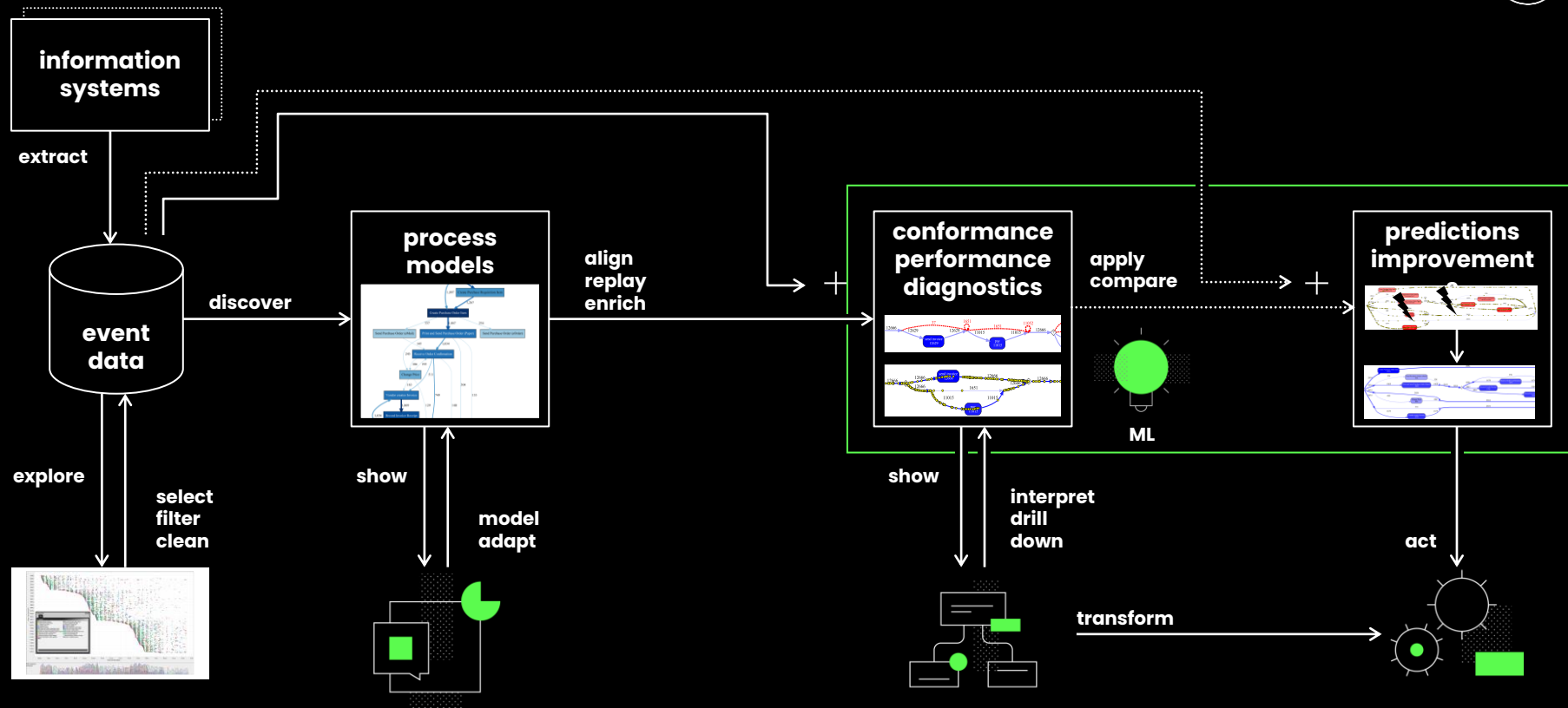# Comparative and Predictive Process Mining

## Process mining:
## From Theory to Execution

prof.dr.ir. Wil van der Aalst

www.vdaalst.com @wvdaalst | www.pads.rwth-aachen.de

information
systems

extract

discover

event
data

explore

select
filter
clean

process
models

align
replay
enrich

show

model
adapt

conformance
performance
diagnostics

apply
compare

ML

predictions
improvement

show

interpret
drill
down

act

transform

© Wil van der Aalst (use only with permission & acknowledgements)

# Basic **event data** enabling **process discovery** and **conformance checking**

| Case ID | Activity | Resource | Timestamp | Product | Prod-price | Quantity | Address |
|---|---|---|---|---|---|---|---|
| ... | ... | .... | ... | .... | ... | ... | ... |
| 6350 | place order | Aiden | 2018/02/13 14:29:45.000 | APPLE iPhone 6 16 GB | 639,00 € | 5 | NL-7751DG-21 |
| 6283 | pay | Lily | 2018/02/13 14:39:25.000 | SAMSUNG Galaxy S6 32 GB | 543.99 € | 3 | NL-7828AM-11a |
| 6253 | prepare delivery | Sophia | 2018/02/13 15:01:33.000 | APPLE iPhone 6 16 GB | .00 € | 3 | NL-7887AC-13 |
| 6257 | prepare delivery | Aiden | 2018/02/13 15:03:43.000 | SAMSUNG Galaxy S6 32 GB | 543.99 € | 1 | NL-9521KJ-34 |
| 6185 | confirm payment | Emily | 2018/02/13 15:05:36.000 | SAMSUNG Galaxy S4 | 329,00 € | 1 | NL-9521GC-32 |
| 6218 | confirm payment | Emily | 2018/02/13 15:08:11.000 | APPLE iPhone | 969,00 € | 2 | NL-7948BX-10 |
| 6245 | make delivery | Michael | 2018/02/13 15:14:04.000 | APPLE iPhone | 639,00 € | 3 | NL-7905AX-38 |
| 6272 | pay | Emily | 2018/02/13 15:20:36.000 | APPLE iPhone 6 16 GB | 639,00 € | 1 | NL-7821AC-3 |
| 6269 | pay | Charlotte | 2018/02/13 15:25:21.000 | SAMSUNG Galaxy S4 | 329,00 € | 1 | NL-7907EJ-42 |
| 6212 | prepare delivery | Sophia | 2018/02/13 15:43:39.000 | HUAWEI | .00 € | 1 | NL-7905AX-38 |
| 6323 | send invoice | Alexander | 2018/02/13 15:46:08.000 | APPLE iPhone 6 16 GB | 639,00 € | 1 | NL-7833HT-15 |
| 6246 | confirm payment | Jack | 2018/02/13 15:56:03.000 | SAMSUNG Galaxy S4 | 329,00 € | 3 | NL-7833HT-15 |
| 6347 | send invoice | Jack | 2018/02/13 15:57:42.000 | SAMSUNG Galaxy S4 | 329,00 € | 3 | NL-7905AX-38 |
| 6351 | place order | Zoe | 2018/02/13 16:17:37.000 | APPLE iPhone | | | 9521GC-32 |
| 6204 | prepare delivery | Sophia | 2018/02/13 16:31:28.000 | SAMSUNG Core Prime G361 | 135,00 € | 1 | NL-7828AM-11a |
| 6204 | make delivery | Kaylee | 2018/02/13 16:51:54.000 | SAMSUNG Core Prime G361 | 135,00 € | 1 | NL-7828AM-11a |
| 6265 | confirm payment | Lily | 2018/02/13 16:55:55.000 | SAMSUNG Galaxy S4 | 329,00 € | 4 | NL-9521GC-32 |
| 6250 | confirm payment | Jack | 2018/02/13 17:03:26.000 | MOTOR ●●● oto G | 199,00 € | 4 | NL-7942GT-2 |
| 6328 | send invoice | Lily | 2018/02/13 17:30:16.000 | APPLE iPhone 6s 64 GB | 858,00 € | 4 | NL-9514BV-16 |
| 6352 | place order | Aiden | 2018/02/13 17:53:22.000 | APPLE iPhone 6 16 GB | 639,00 € | 2 | NL-9514BV-16 |
| 6317 | send invoice | Jack | 2018/02/13 18:45:30.000 | APPLE iPhone 6s 64 GB | 858,00 € | 5 | NL-7907EJ-42 |
| 6353 | place order | Sophia | 2018/02/13 20:16:20.000 | APPLE iPhone 5s 16 GB | 449,00 € | 4 | NL-7751AR-19 |
| ... | ... | .... | ... | ... | ... | ... | ... |

**event =
case +
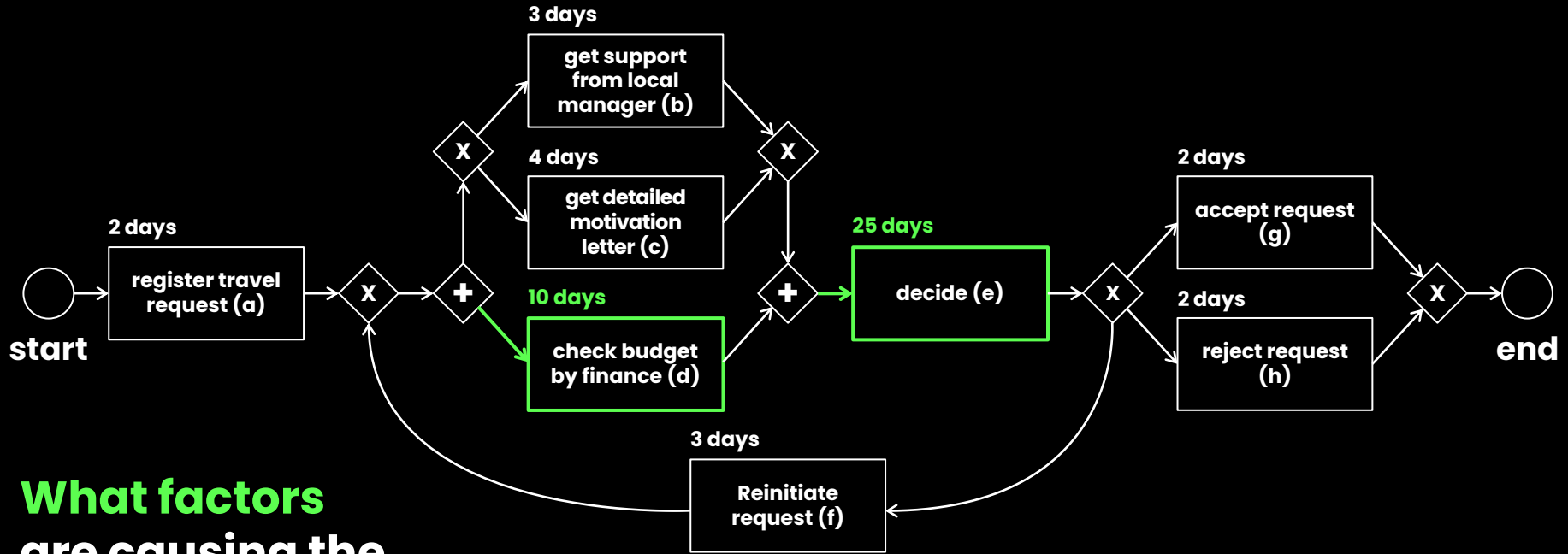activity +
timestamp +**

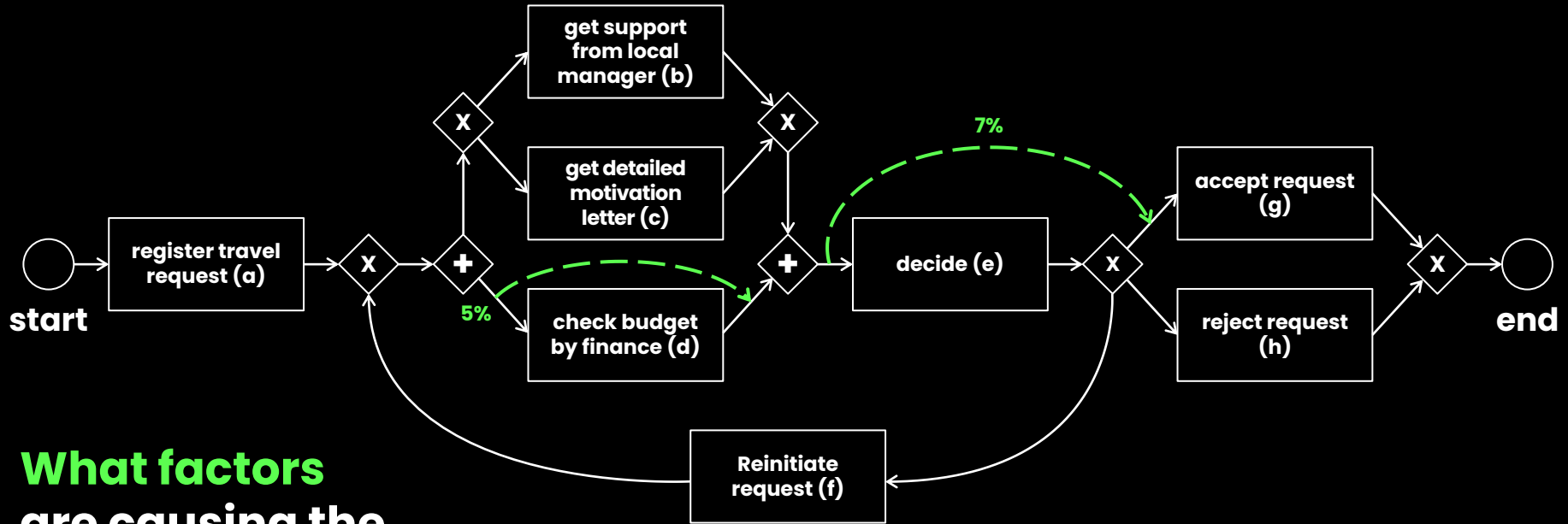# Decision point analysis
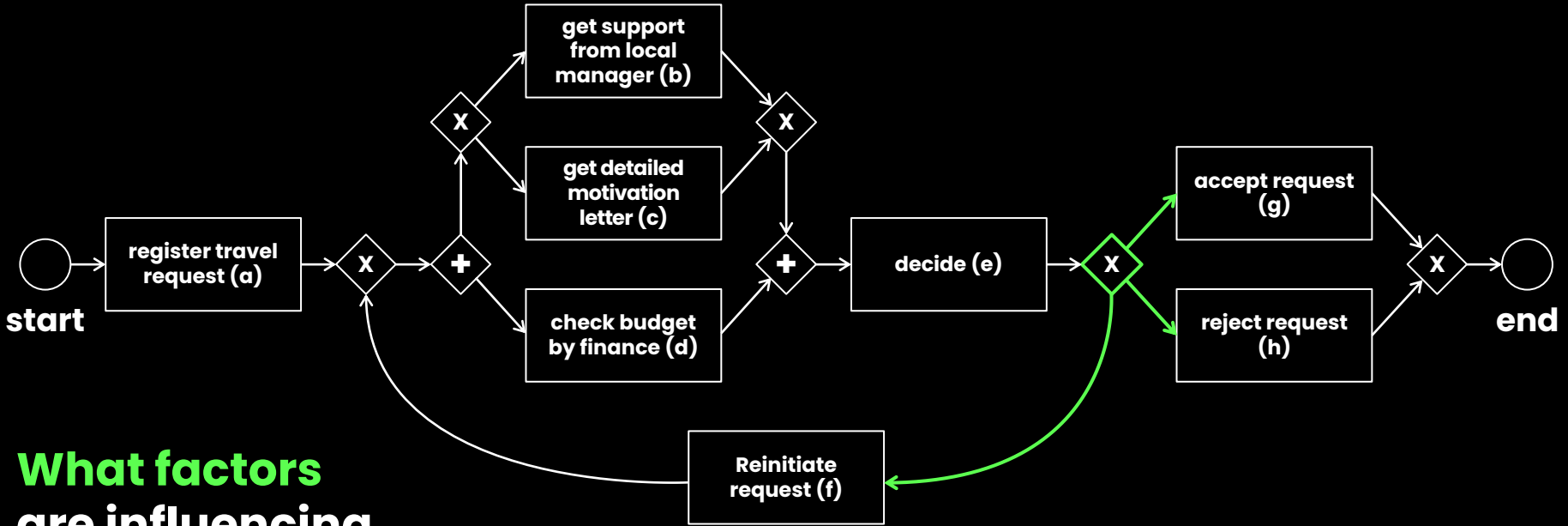


**What factors** are influencing the decision?

# Bottleneck analysis



**What factors** are causing the two bottlenecks?
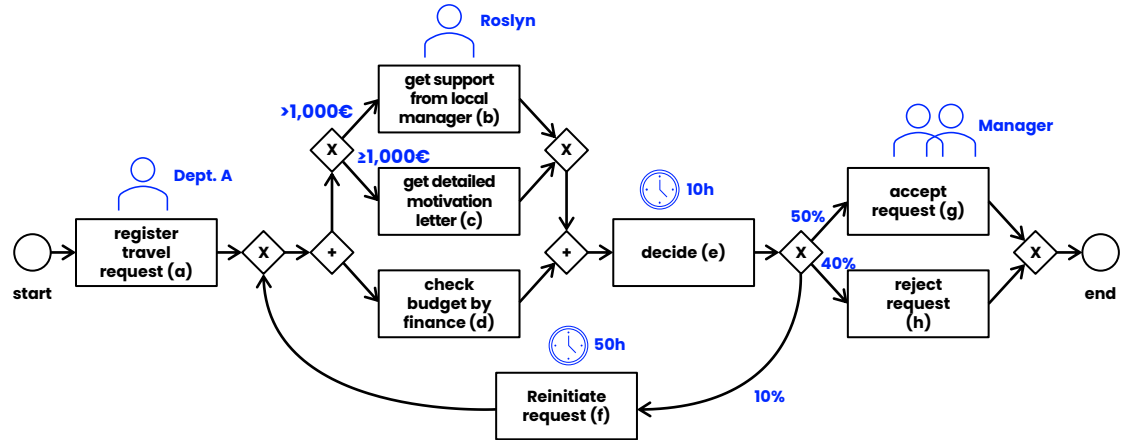
# Deviation analysis



**What factors** are causing the two deviations?

# Extending process models

## Such models

can be used to compare and predict!

- **durations**
- **probabilities**
- **roles**
- **decisions**
- **deviations**
- **resources**

# Compare: Location/Routing

## Aachen
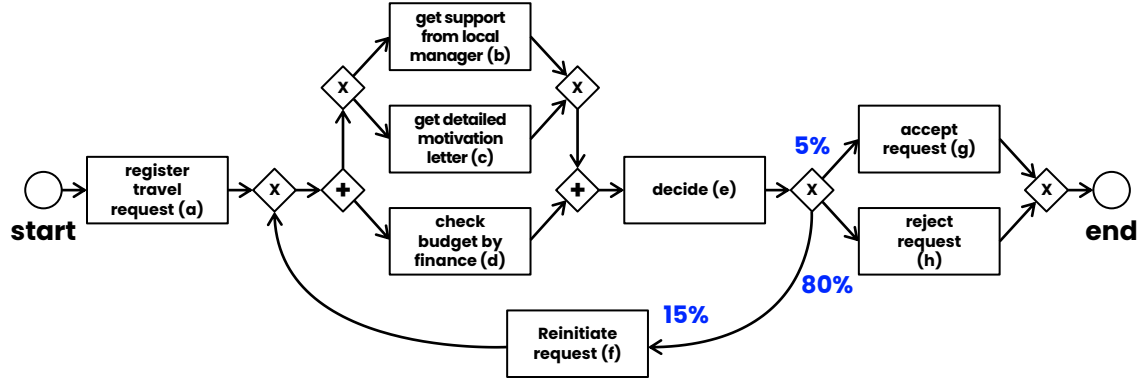


## Munich
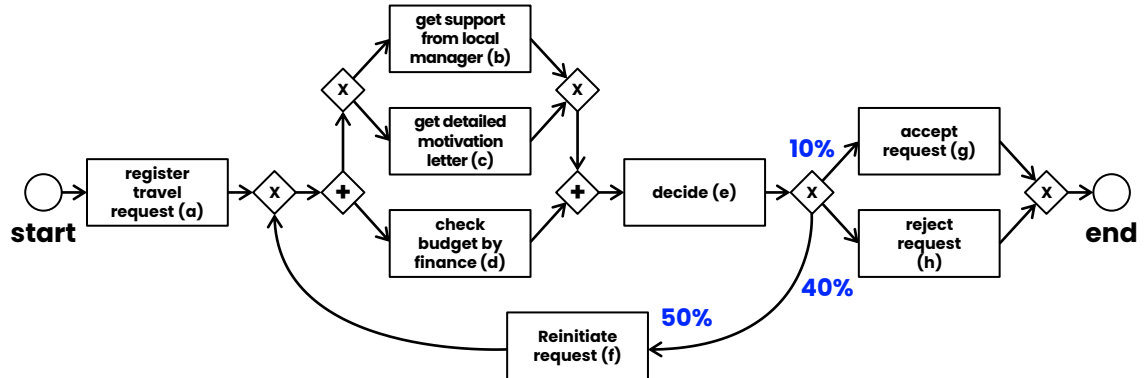
# Compare:

## February

## March

# Period/Delays
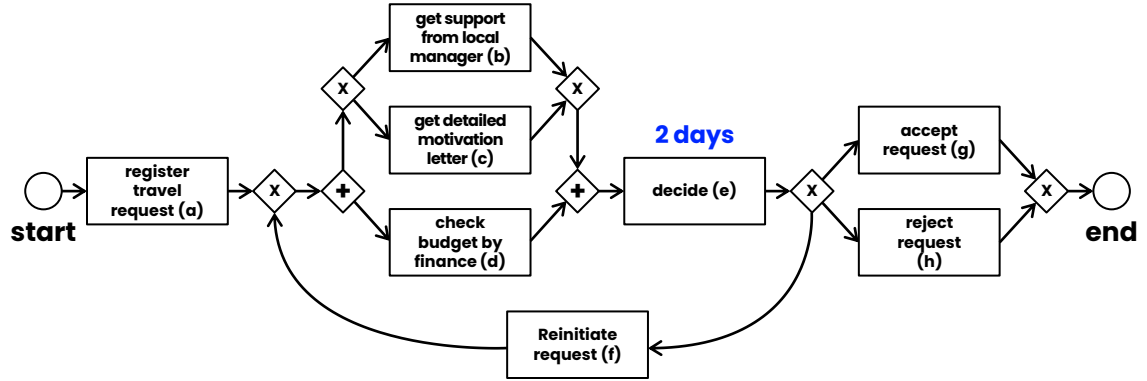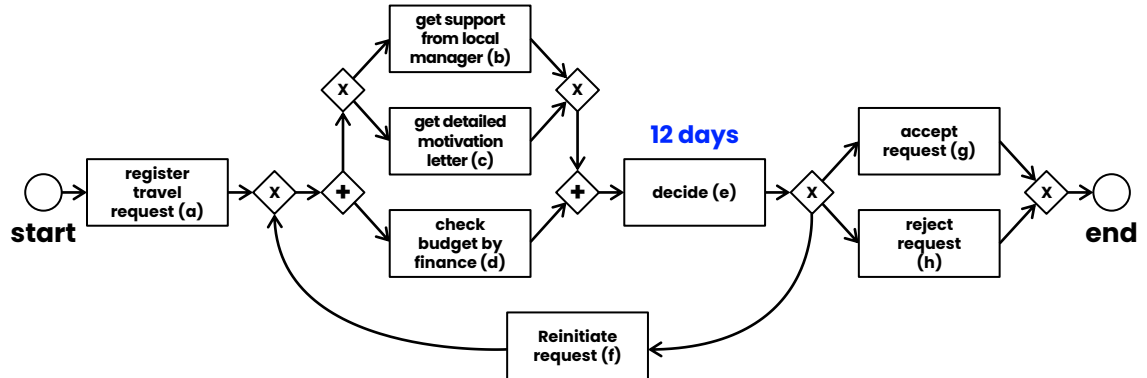
# Compare:

**Age/Deviations**
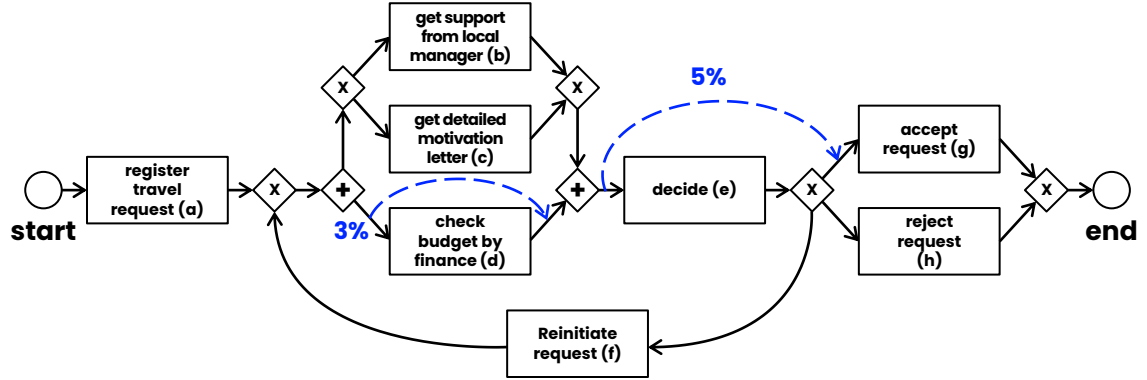
**> 50 year**

**≤ 50 year**
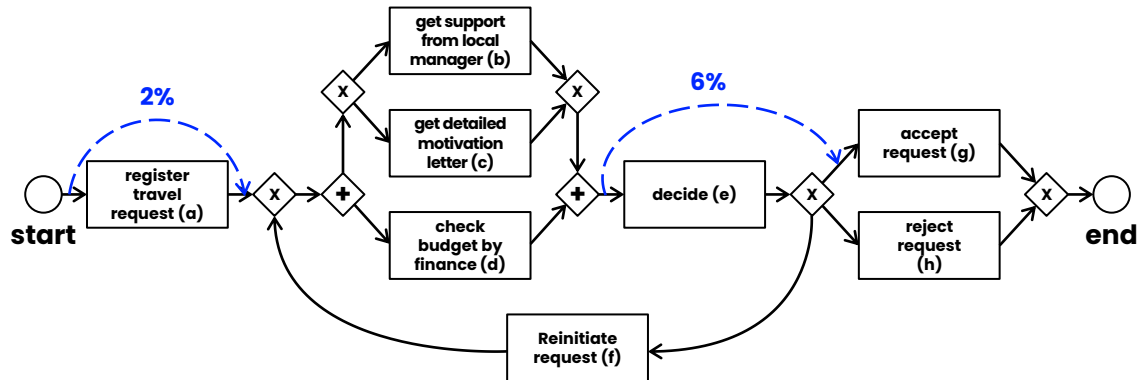
# Compare

**Find differences and commonalities considering multiple perspectives.**
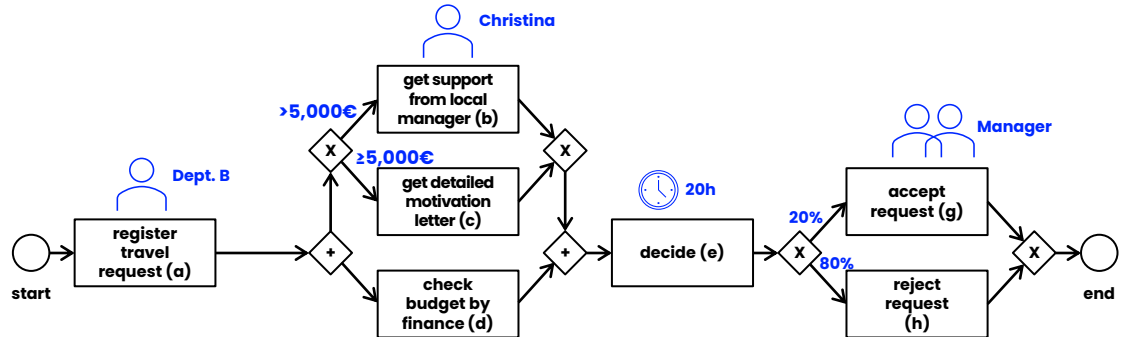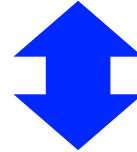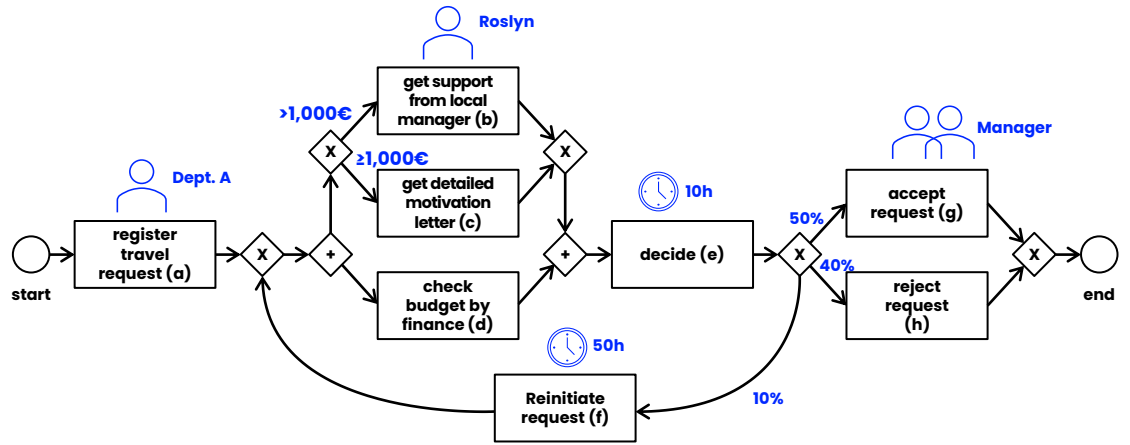
# Predict

We can use the enriched process model to **predict** the "path" of any case.

So we can **predict** whether the case will be **delayed**, **deviate**, **rejected**, etc.

**Comparison and prediction are closely related**



**Will be too late**

Roslyn

get support from local manager (b)

>1,000€

≥1,000€

get detailed motivation letter (c)

Dept. A

register travel request (a)

start

check budget by finance (d)

**Will deviate**

Reinitiate request (f)

50h

10h

decide (e)

50%

40%

10%

accept request (g)

Manager

reject request (h)

end
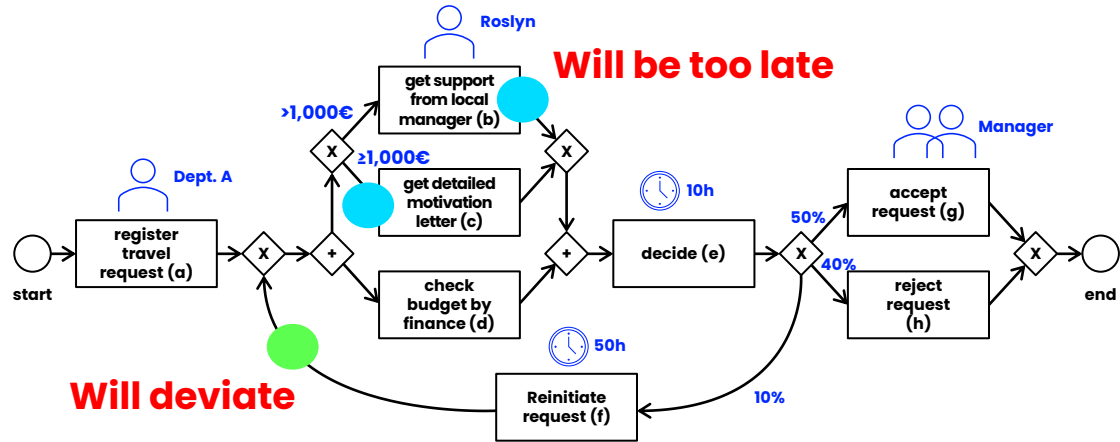
**Also for plain DFGs**

# <194 days

**63% pays directly**

Process Start
74,856

74,856

Create Fine
74,856

46,785

28,071    505    Payment
53,419

455    4,001

Send Fine
28,526

1,008    3,727

4,072    53,050

20,453

Add penalty
5,080

1,353

Process End
74,856

# ≥194 days

**0.2% pays directly**

Process Start
75,514

75,514

Create Fine
75,514

75,347

Send Fine
75,461

71,816    167

3,201

Add penalty
74,780

444    15,406    2,964

59,374    3,803    Payment
22,577

15,696

Process End
75,514

**Road traffic fines log**
**https://doi.org/10.1007/s00607-015-0441-1**

# <194 days

**on average
10 days until
payment**

Process Start
74,856

Create Fine
74,856

10 days

84 days  39 days   Payment
53,419

41 days    37 days

Send Fine
28,526

37 days   14 days

72 days

Add penalty
5,080

Process End
74,856

# ≥194 days

**on average
165 days until
payment**

Process Start
75,514

Create Fine
75,514

89 days

Send Fine
75,461

78 days        165 days

Add penalty
74,780        58 days

199 days   33 days

104 days    Payment
16,296

Process End
75,514

**Road traffic fines log
https://doi.org/10.1007/s00607-015-0441-1**

# Predictive process mining

## Identify situations
**(case, choice, stage, etc.)**

## Extract features

- **one target feature**
- **one or more input features**

## Build a model explaining
**the target feature in terms of the input features.**

| age | dev. | ... | time | outcome |
|-----|------|-----|------|---------|
| 41 | yes | | 4h | reject |
| 35 | no | | 8h | accept |
| 62 | yes | | 7h | reject |
| 24 | yes | | 8h | reject |
| ... | ... | ... | ... | ... |
| 71 | no | | 1h | accept |
| 71 | no | | 1h | reject |

# Machine Learning = Learning based on examples without being programmed

**input features**

**target features**

0.1895
0.2345
0.1895
0.7765
0.2345
0.6545
0.5653
0.765
0.1895
0.4455
0.765
0.765
0.7745
0.6545
0.2342
0.7765
0.765
0.2342
0.8845
0.2342
0.1895
0.2667
0.2345
0.9943
0.2345
0.2667
0.8845
0.7765
0.5653
0.7765
0.9943
0.4455
0.7745
0.765
0.2345
0.5653
0.9943
0.2342
0.2247
0.7745
0.5653
0.9845
0.2247
0.7765
0.8845
0.5653
0.9943
0.5653
0.9845
0.2247
0.7745
0.2667
0.7765
0.2345
0.5653
0.2247
0.7745

**Neural networks:**
**First a failure now the dominating paradigm**

**Focus on specific tasks (speech, images, etc.).**

| age | dev. | ... | time | outcome |
|-----|------|-----|------|---------|
| 41 | yes | | 4h | reject |
| 35 | no | | 8h | accept |
| 62 | yes | | 7h | reject |
| 24 | yes | | 8h | reject |
| ... | ... | ... | ... | ... |
| 71 | no | | 1h | accept |
| 71 | no | | 1h | reject |

# How about managing and improving operational processes?

**Fortunately, we can combine both!**

**Process mining can be used to create process-related machine learning problems.**

**We need process models** that are **understandable**!

**We are interested in improving end-to-end performance and compliance** (not a single task)!

**We do not have labeled data, we have SAP, Sales-force, Oracle, Microsoft, Infor, etc.** (holding thousands of tables)!

**What is the likelihood that we will loose this customer?**
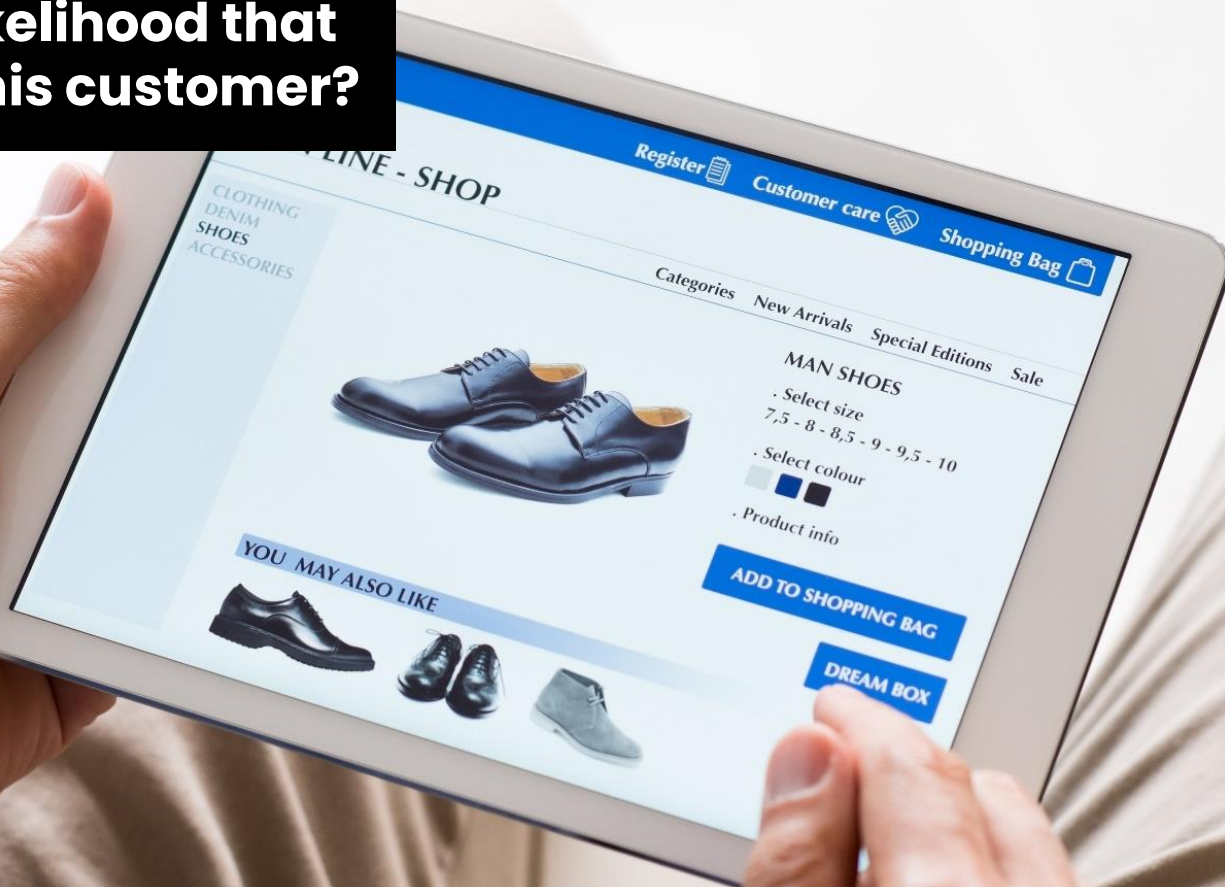
**Will this car be delivered on time?**

**Will we have enough free beds in the ICU tomorrow?**

Will this bag miss the plane?

**information systems**

extract

**event data**

discover

**process models**

align
replay enrich

**conformance performance diagnostics**

apply
compare

**predictions improvement**

ML

explore

select
filter
clean

show

show

act

**Identify situations (case, choice, stage, etc.)**

**Extract features:**
- **one target feature**
- **one or more input features**

**Build a model explaining the target feature in terms of the input features.**

information
systems

extract

event
data

discover

process
models

align
replay enrich

conformance
performance
diagnostics

apply
compare

predictions
improvement

Warning:
Do not start here!

ML

Start here!

show

interpret
drill
down

transform

act

© Wil van der Aalst (use only with permission & acknowledgements)