Data-driven Fraud Detection

The Data Mining Unit (Den Fælles Dataenhed)

Annika M. Jacobsen
09-10-2020

den fælles dataenhed

Data Mining

Acquiring Data from abroad

Cooperation across authorities

Monitoring

Data Mining - Overview

Den Fælles Dataenhed

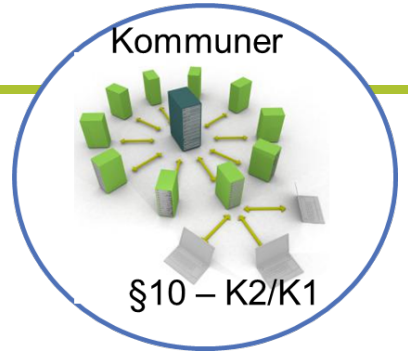
Data Sources

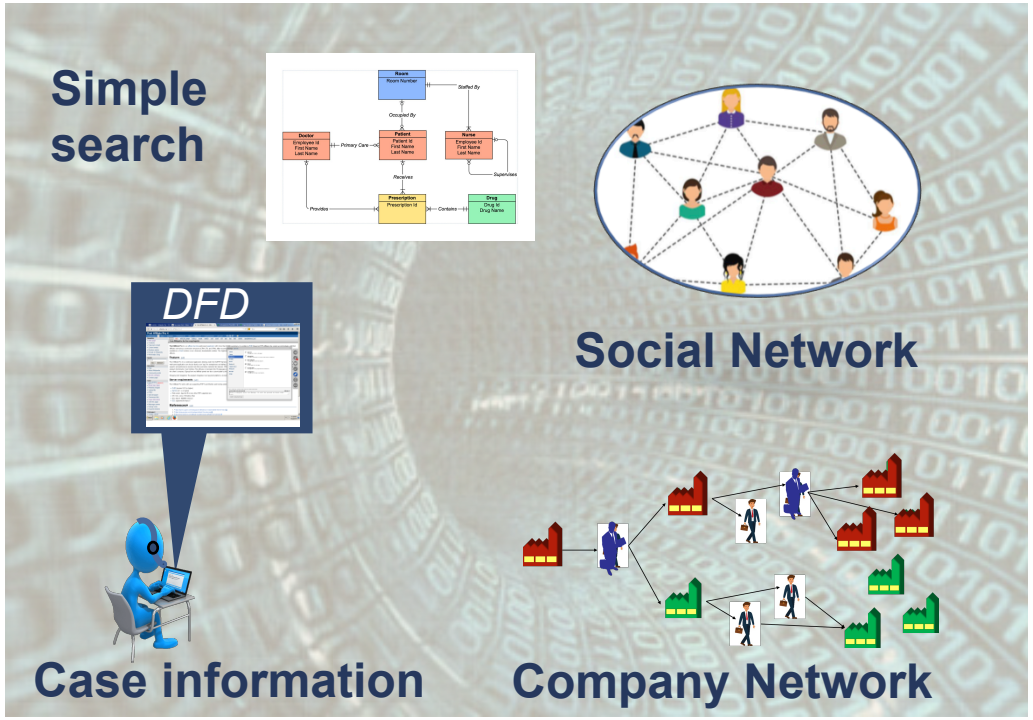
Social benefits receivers

- UDK (Housing benefit , Pension, family benefit, maternity/ paternity benefits
- KOM (Unemployment cash benefit, sick leave benefit and free childcare benefits)
- SIU (Student benefits)
- ATP (AUB, LG, ATP)

Data from public authorities

- CPR (Social network)
- BBR (Apartment size etc.)
- CVR (Company network)
- R75 (Car, bank accounts, assets, tax etc.)
- eIndkomst (employment, payments and work hour's)
- Regional Data (Check of stay in DK)
- STAR (Physical meeting In job-center)
- Etc.





Monitoring

Data Mining

Productions and Operations

Den Fælles Dataenhed

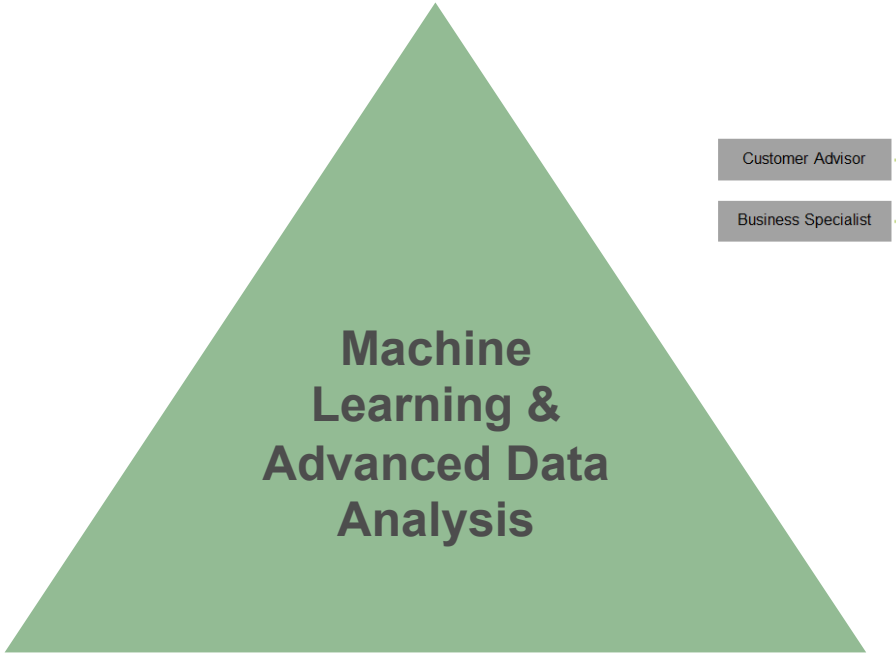
Person register
Housing register
Business register
Income register

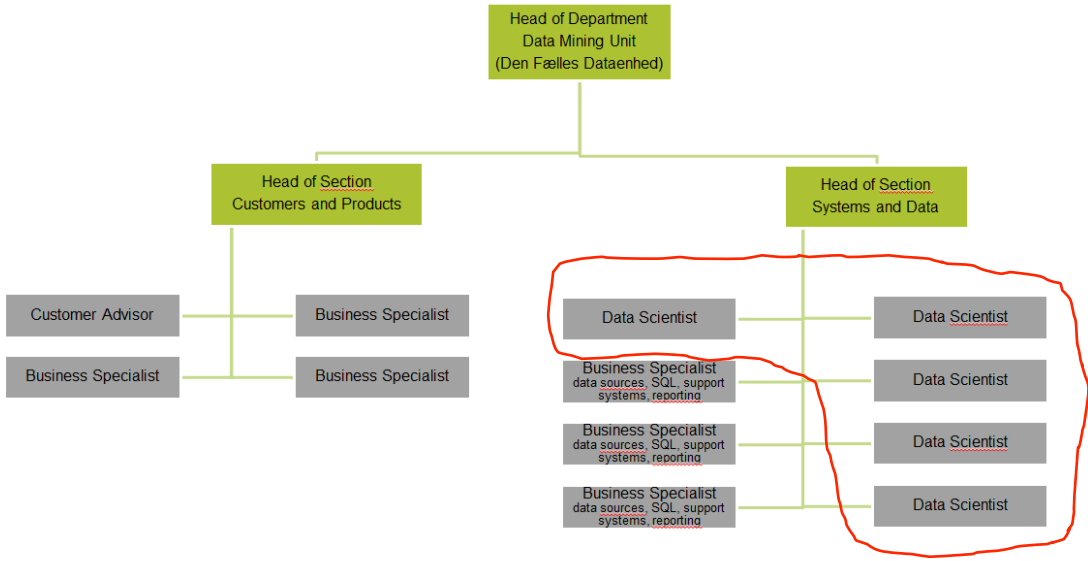
+

R75 (tax data)
Health services register
VAT register (companies)
...

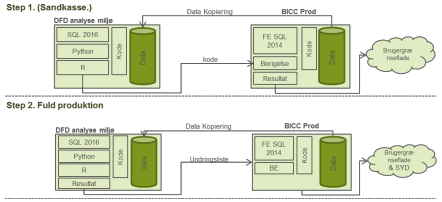
Data

Competences





Infrastructure



Overview From 2015 To 2019

Den Fælles Dataenhed

From 3 Data Sources in 2015 to 9 Data Sources in 2018

From simple experience-based criteria's to using complex machine learning techniques

From Fraud Detection to early error and Fraud Prevention and automatization

From 11 mill. EUR in 2016 to 62 mill. EUR in 2019

(1+1=2)

CPR
BBR
CVR

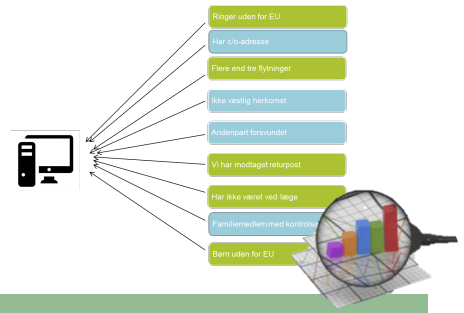
CPR
BBR
CVR
E-Ind.

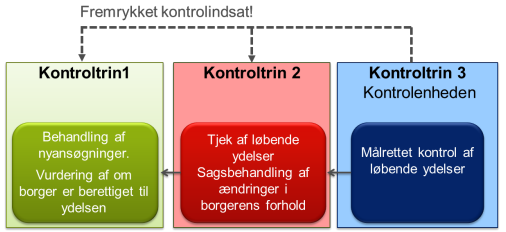
CPR
BBR
CVR
E-Ind.
R75

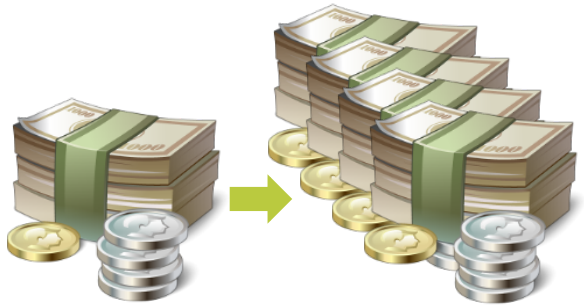
CPR
BBR
CVR
E-Ind.
R75
SU

CPR
BBR
CVR
E-Ind.
R75
SU
Region.

CPR
BBR
CVR
E-Ind.
R75
SU
Region.
Moms
STAR data







Data Sources

Criteria's

Control Stages

Revenue

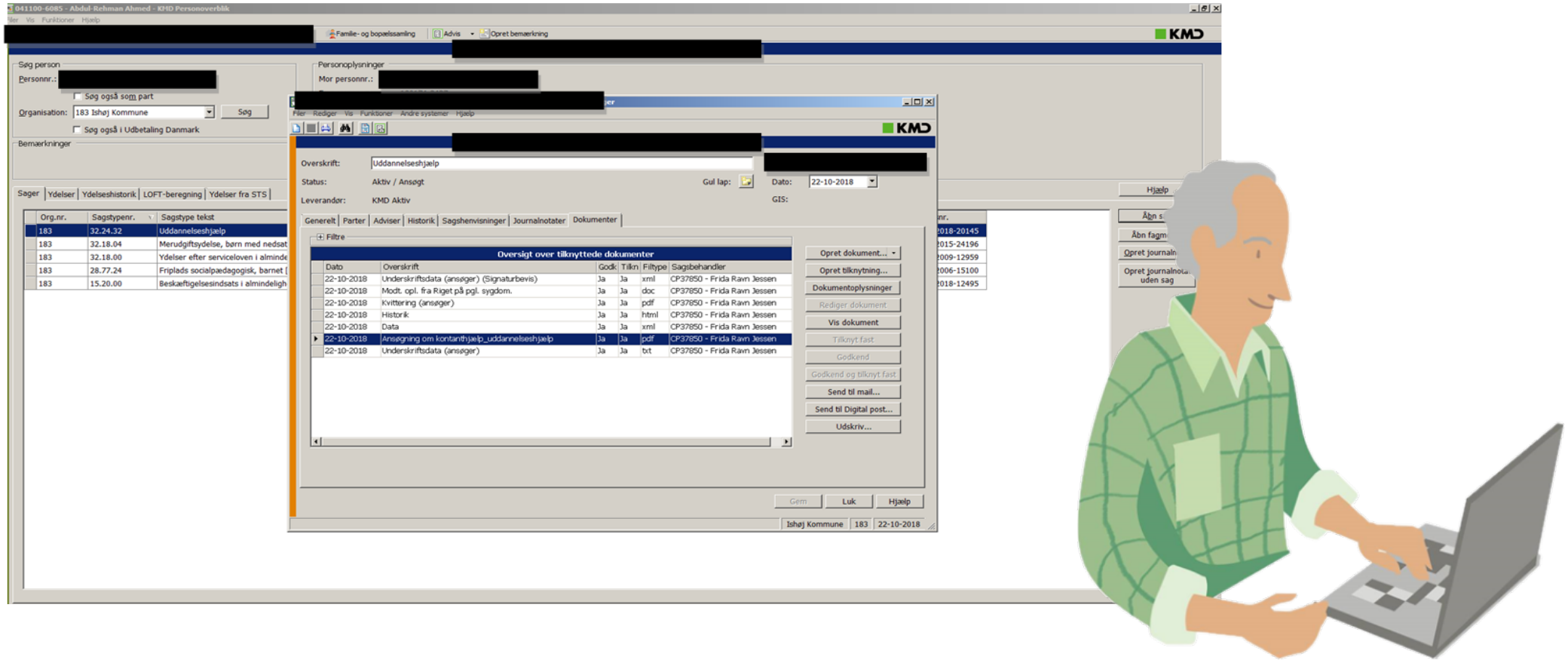
Multiple fathers on same address
Residence small with more residents than rooms
Incorrect address?

1

Case Informations

Receiving a Digital application of Unemployment cash benefit

Den Fælles Dataenhed

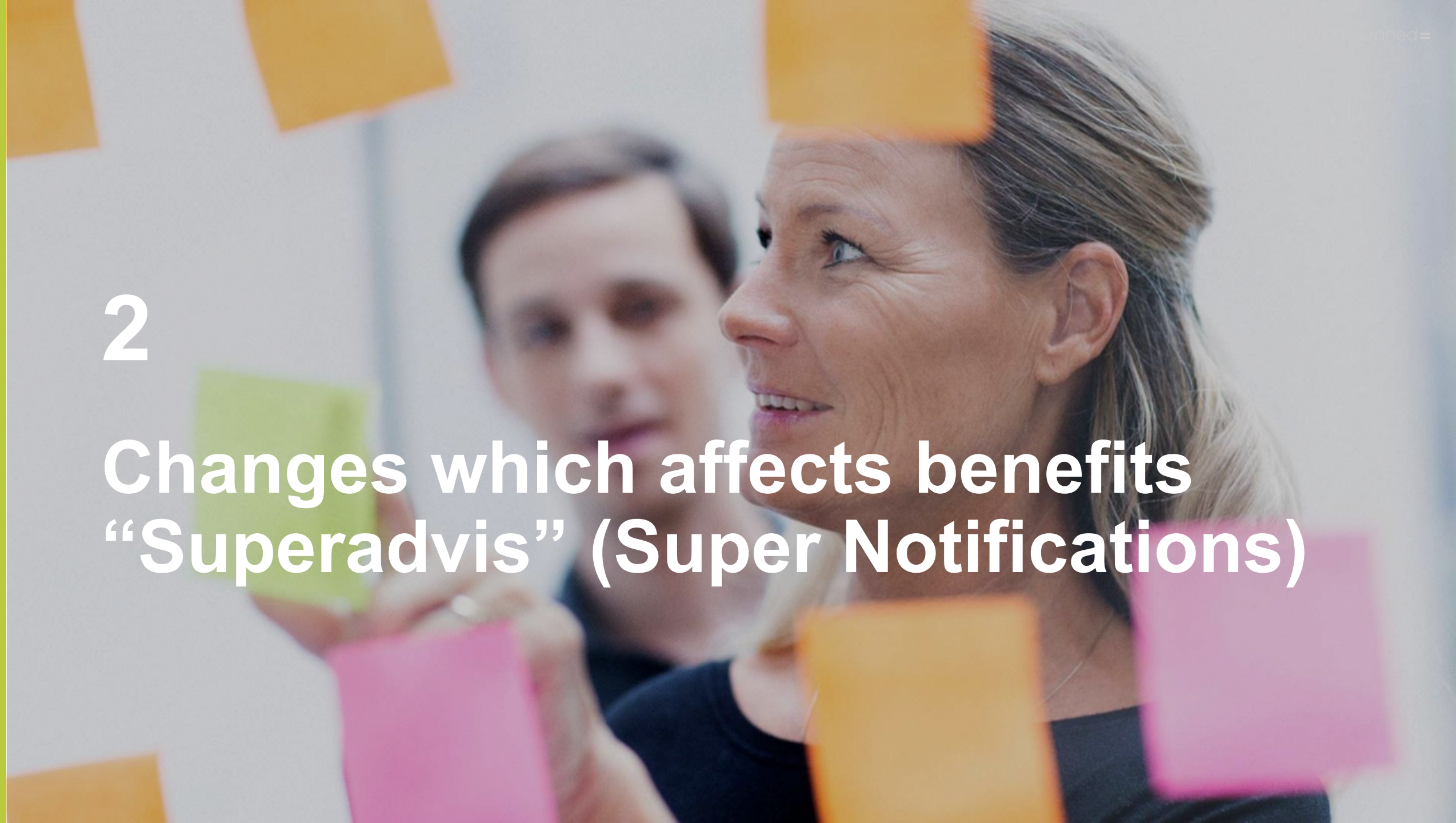


Now a ready information sheet ”betingelsesark” is used

Den Fælles Dataenhed

Ægtefælles børneopsparing	Ikke relevant
Bil	Nej
Ejendom	Nej
Aktier	Borger har ingen aktier
Overskydende skat	Den Fælles DataenheD.har ikke information om at borger har fået penge tilbage i skat
Ægtefælles bil	Ikke relevant
Ægtefælles ejendom	Ikke relevant
Ægtefælles aktier	Ikke relevant
Ægtefælles overskydende skat	Ikke relevant
A-kasse	Borger har ikke indbetalt til a-kasse
Offentlige ydelser	Borger har ikke faet udbetalt offentlige ydelser.
Kontanhjælps ansøgninger	Nej
Indkomst seneste tre måneder	Borger har fået tilbagetrukket penge fra eindkomst den sidste måned, se felt tilbagetrækninger
Indkomst seneste tre måneder	Borger har fået tilbagetrukket penge fra eindkomst den sidste måned, se felt tilbagetrækninger
Indkomst seneste tre måneder	Borgers bruttoløn er 15.205 kr. perioden018-07-01 til 2018-08-01. SE nr. [redacted]
Tilbagetrækning i eIndkomst	Der er ingen tilbagetrækning på borgeren i eindkomst
Klyngeindberetning i eIndkomst	Nej
SU	Nej
Regionskontakt	Nej
3 seneste regulære lønninger	Fra 2018-07-01 til 2018-07-31 15.205 kr (brutto) CV[redacted]
3 seneste regulære lønninger	Fra 2018-06-01 til 2018-06-30 8.756 kr (brutto) CVR[redacted]
3 seneste regulære lønninger	Fra 2018-05-01 til 2018-05-31 8.216 kr (brutto) CVR[redacted]

2



2

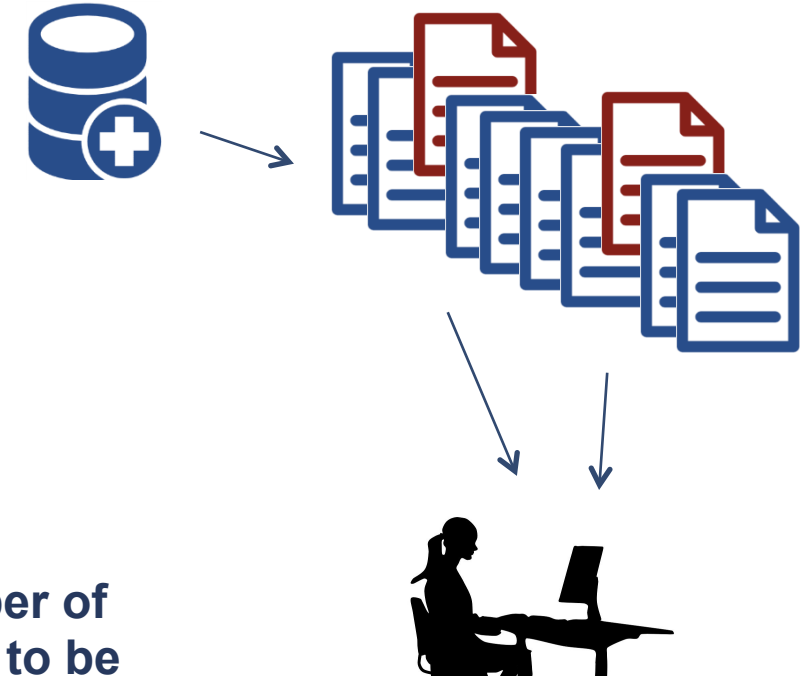
Changes which affects benefits “Superadvis” (Super Notifications)

§ 10 – Super-notification

The programs are finding cases which have received data about citizens that could have an impacted on the right to benefit outpayments

”Earlier the caseworker went trough every citizen100% receiving unemployment cash benefit every 3. month”

”Now the caseworker only looks at those cases selected by the datamining programs around 15%”



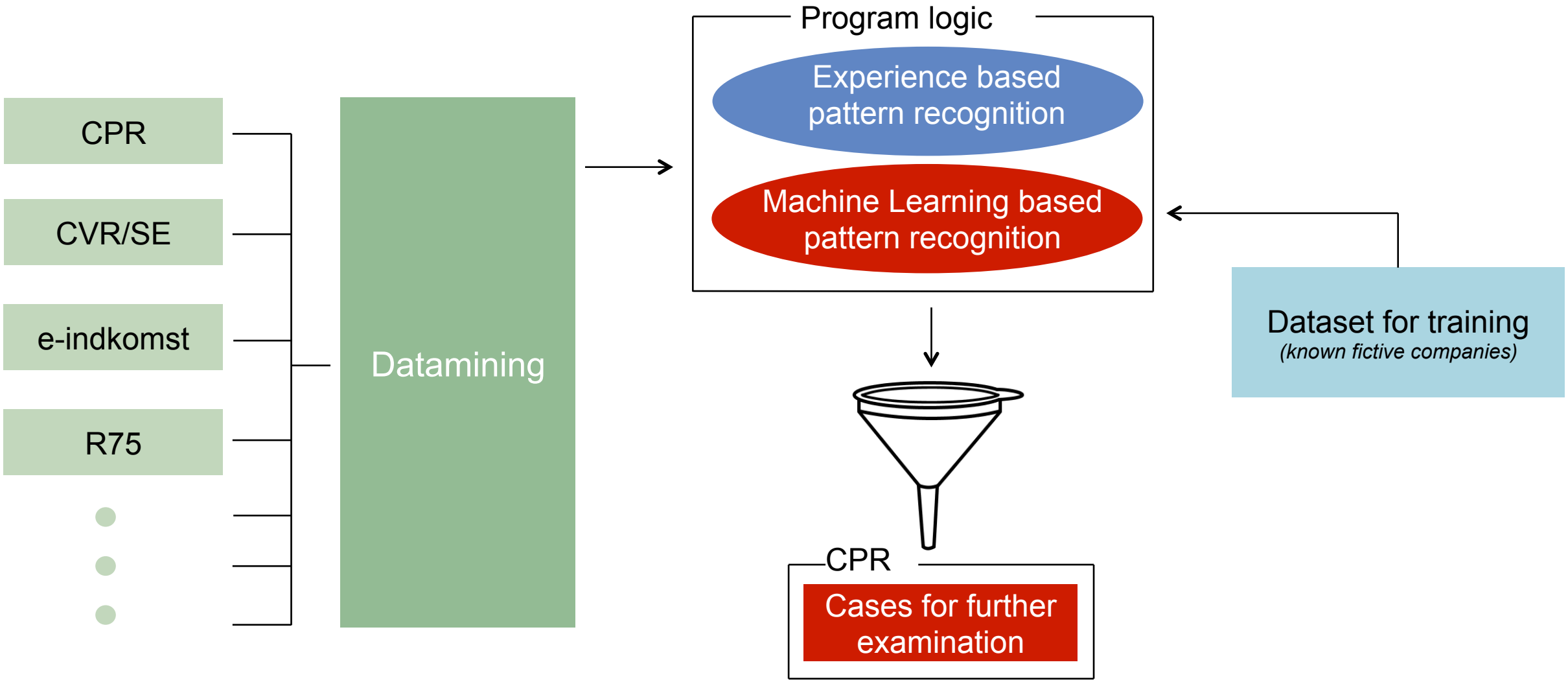
This way the number of cases which needs to be processed manual gets reduced by 85%

3

Fictive Employment, Machine Learning model - example

Company Analyses

Den Fælles Dataenhed



Keeps Going In Circles….

Den Fælles Dataenhed

- *Receive unemployment cash benefit*
- *Receive insurance unemployment benefit*
- *Receive three months salary from LG*
- *The company goes bankruptcy*
- *Receive sick leave benefit*

Earnings for keeping a residence permit

- *Establish a company*
- *Employ staff*
- *Transfer seniority from EU insurance for unemployment*
- *Receive family benefit*
- *Hire pregnant 4 months prior to child birth*
- Receive maternity/ paternity benefits

Suspected Fictive Employment

Identification of company specific information's

Statistic Around Companies

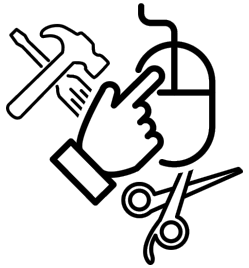
Den Fælles Dataenhed

Descriptive analyze

Owners resident

If we look at all Danish companies, then 1,5% of the owners are living outside of Denmark

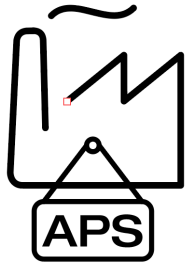
Are we looking at known companies with fictive hires, then 20,8% of the owners are living outside of Denmark



Branch

If we look at the branch, then 2% of all Danish companies are in real-estate- and clearing service

Are we looking at known companies with fictive hires, the 32% of the companies are in real-estate- and clearing service



Company type

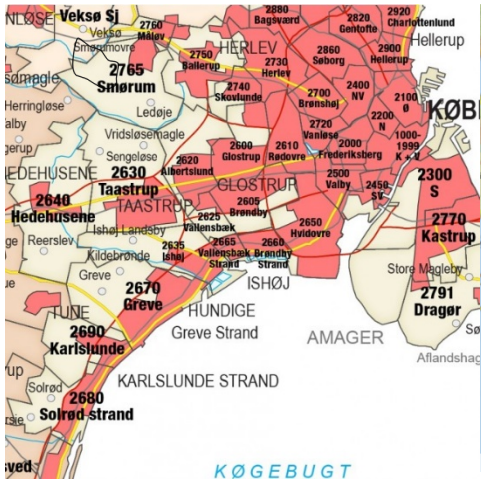
If we look at the company type, then 33% of all Danish companies are APS (Ltd with start value of 6.800 EUR)

80% of the companies with fictive hires are APS'es

Other Data Tracing

Retreated salary in the taxing system

Geographic placing

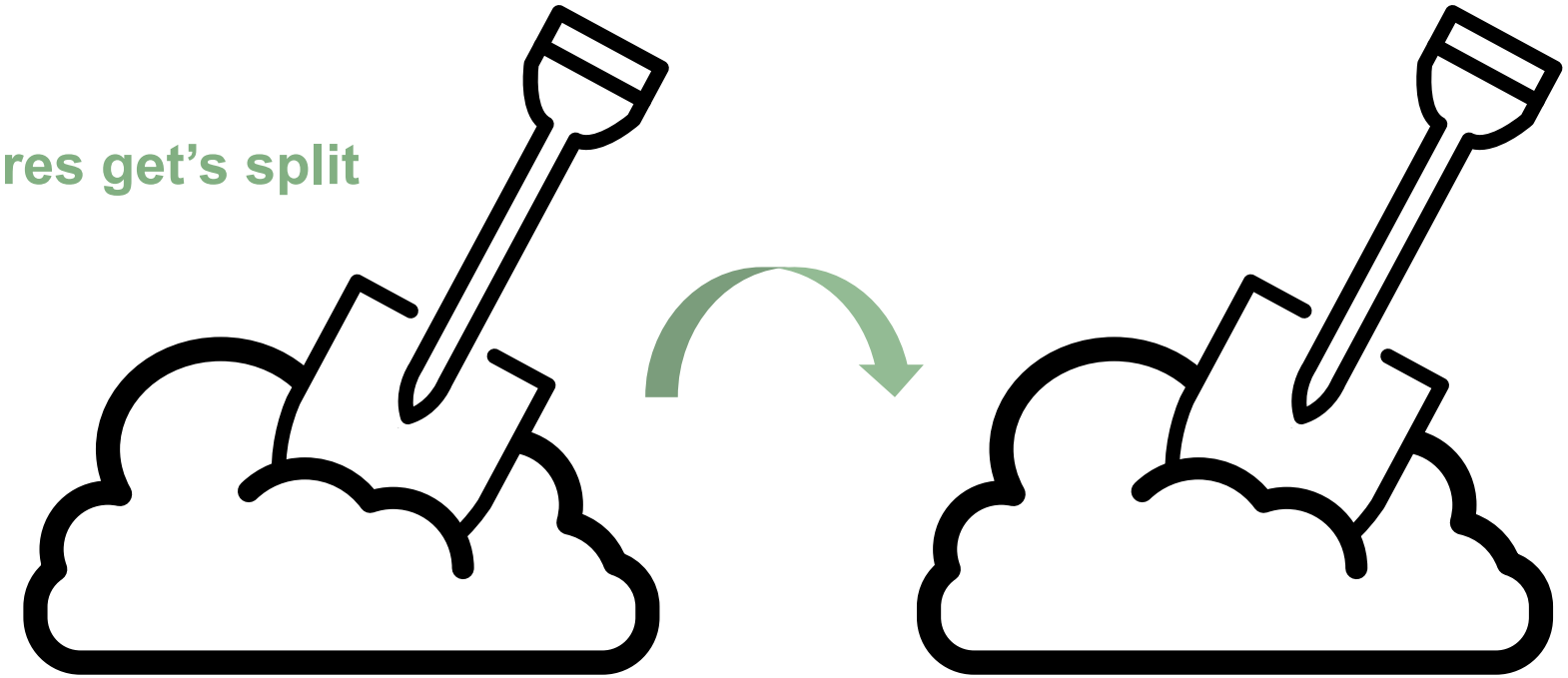


- ✓ 2600
- ✓ 2700

To many pregnant employed

Machine Learning In Praksis

Known fictive hires get's split into to piles



First pile gets used to train the algorithm

Second pile get used to se how god the algorithm is to identify fictive hires

$$P(y = snyd \mid x) = \beta_0 + \beta_{str}x_{str} + \beta_{branche}x_{branche} + \beta_{indrejsteEjer}x_{indrejsteEjer} + \ldots\ldots + \epsilon$$

Concrete case

Transport and cures company

Establish January 2014

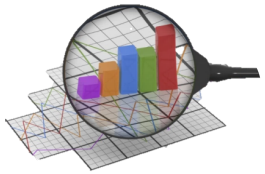
February 2014 owner hires his pregnant wife

May 2014 to April 2015 maternity benefits gets payed out

Maj 2015 til December 2015 Is not a work

January 2016 the owners hires his pregnant wife

April 2016 to March 2017 maternity benefits gets payed out again



April 2016 UDK's caseworker starts the case appointed from the Machine Learning model

UDK's caseworker receivs insufficient informations from the citizan

- ✓ No employment contract
- ✓ No pay check
- ✓ No payments on the wife's salary account

August 2016 the case gets sent to the tax authorities which examines the company

12. February 2018 - bankruptcy

March 2018 the tax authorities makes at decision that this hire was fictive

4

Fictive sickness, Machine Learning Model - example

Reduction of population

- Sick periods for the last 3 years
- Sort out citizens which are receiving insurance unemployment benefit
- Sort out citizens which are sick with a paragraph 56 or 58
- Sort out citizens which are on fleksløn – permanently sick and therefor on a part time job
- Sort out citizens which are only part time sick
- Sort out citizens which have very little illness

672.381

602.532

556.908

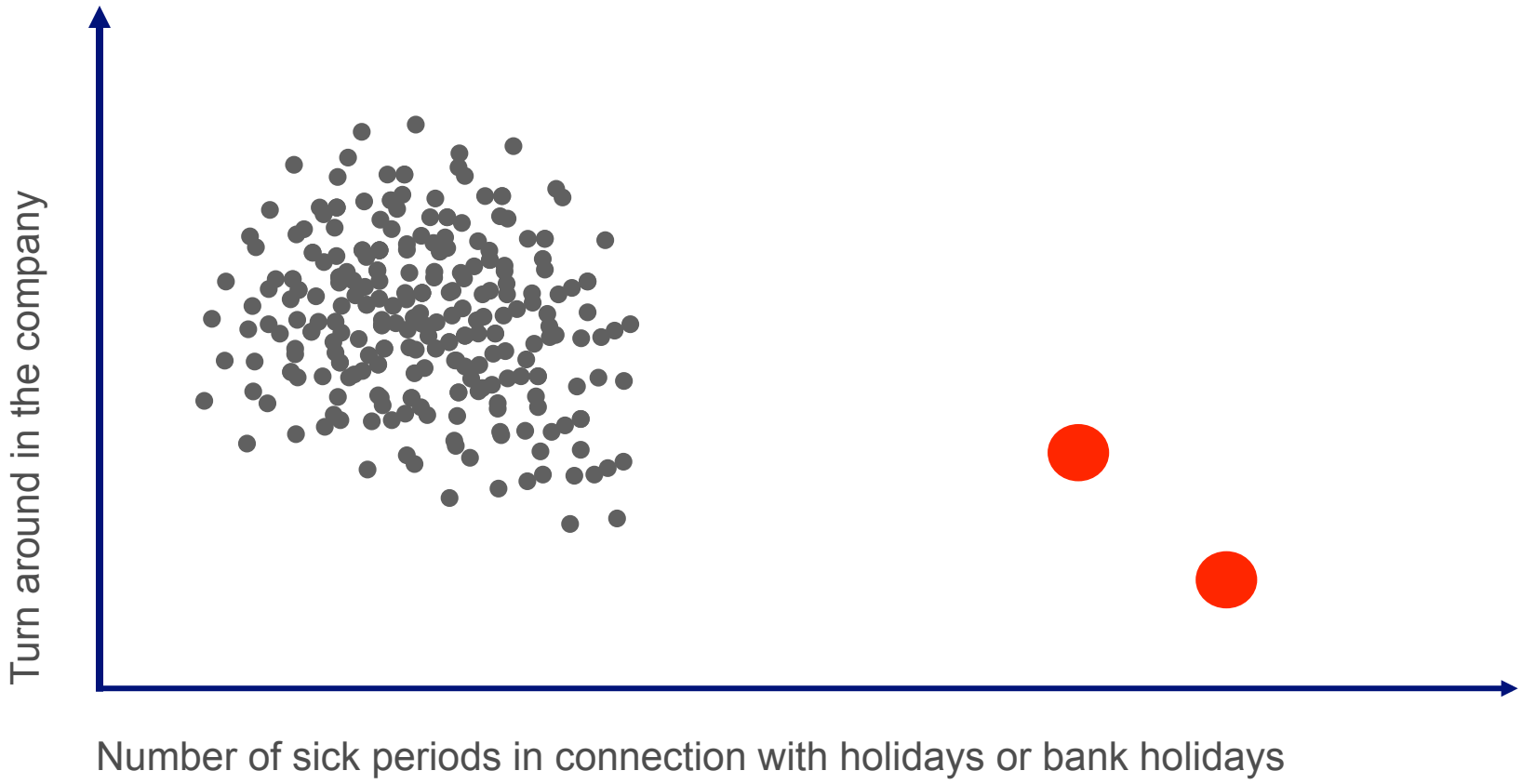
503.910

470.301

118.966

How does the model work?

The model is detecting “unnormal” behavior and situations using Machine Learning



Modules

- Holidays or bank holidays
- Relation to the owner and leader
- Brought cars in the companies
- Number of companies
- Number of branches
- Number of sick days
- Number of sick periods
- Number of bankruptcy's
- Sick from own company
- Periods under 3 and 8 weeks
- Creating a company while sick

All in all 22 modules

Concrete case: Hairdresser

Den Fælles Dataenhed

Frisørsalon etableret 23.11.2006

Borgers sygeperioder med sygedagpenge

Fra-dato	Til-dato
14-02-2019	19-02-2019
18-10-2018	22-10-2018
13-02-2018	19-02-2018
17-10-2017	23-10-2017
15-02-2017	20-02-2017
20-10-2016	24-10-2016

Kontrolgruppen vælger at undersøge sagen

- ✓ Næsten årlige enslydende sygeperioder i uge 7 + sommerferie + uge 42, tilbage fra 2014
- ✓ Anmodning om dagpengerefusion på mandage, som er en fast lukkedag ifølge hendes hjemmeside
- ✓ Opslag på Facebook på anmeldte sygedage/perioder
- ✓ Opslag på Facebook der annoncerer, at salonen er sommerferielukket – dage/uger som er sammenfaldende med, at hun efterfølgende har anmeldt at være syg i samme dage/uger

Registersamkøring

Udfaldstekst

Borger er sygemeldt fra egen virksomhed med branche frisør. Borger har været syg i uge 7 i 2016, 2017, 2018 og 2019, og i uge 42 i 2016, 2017 og 2018 (også syg i sommerferie perioder de sidste 3 år).

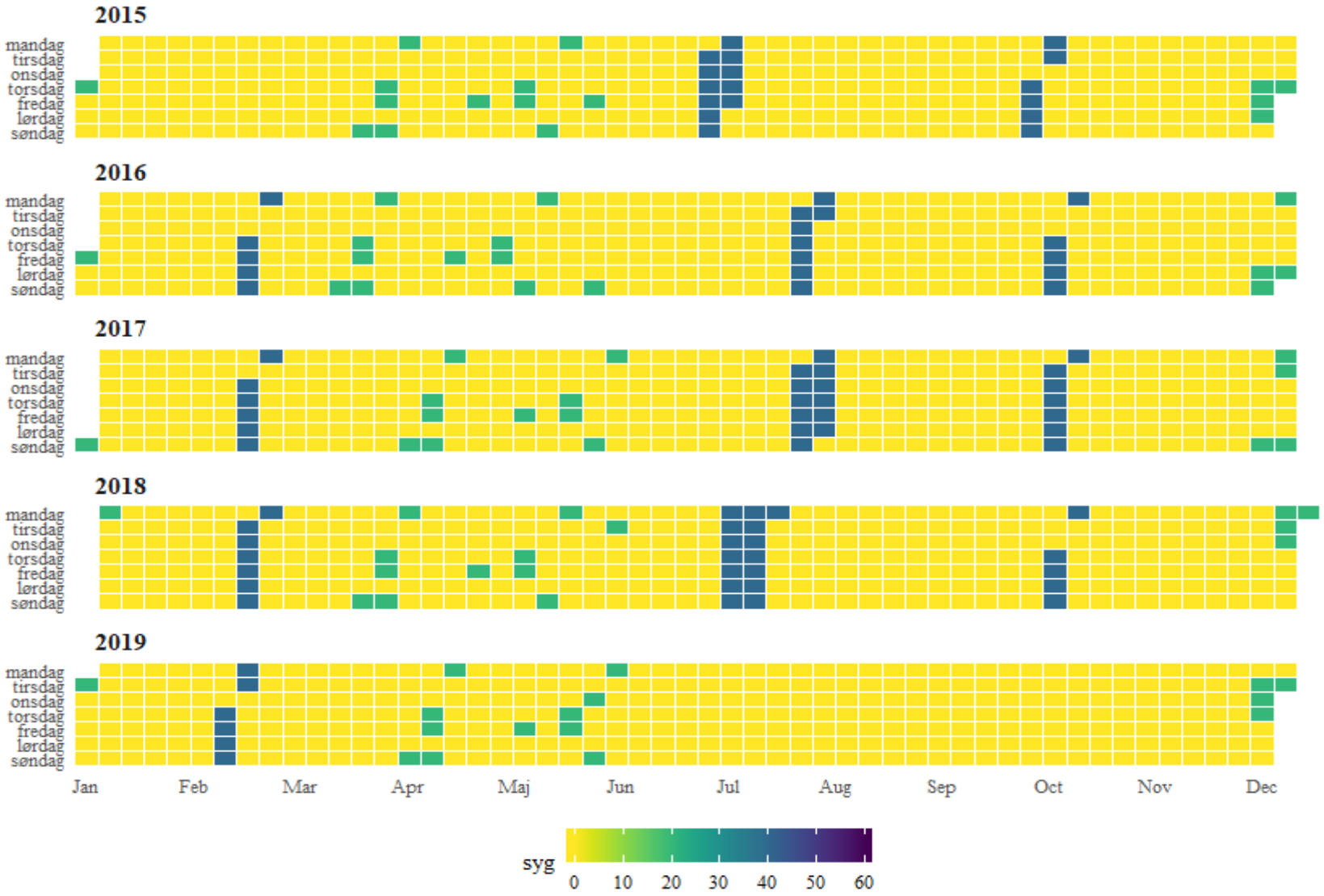
Udbetalt i alt: 55.889 kr. over en tre-årig periode

KR

Borger er partshørt og der skal træffes afgørelse

Graphical mapping of the sick periods

Den Fælles Dataenhed



Concrete case: Chef advisers

Den Fælles Dataenhed

Virksomhed etableret 17.03.2011

Registersamkøring

Ejer er syg 13 gange – ca. 1 måneds varighed ad gangen = 309.236 kr.

Datter og søn er syge 13 gange hver med ca. 1 måneds varighed = 302.041 og 332.288 kr.

Ansatte, nogle, med atypiske sygeforløb = 666.874 kr.

Udbetalt i alt: 1.610.439 kr. siden 2015

KR

Kontrolgrupper i to kommuner undersøger sagerne, og der er aftalt udgående virksomhedskontrol

Generelt undringsværdigt: Virksomheden har i 2018 modtaget 492.639 kr. i sygedagpenge-refusion
Bruttofortjeneste i 2018: 1.491.096 kr.
Personalomkostninger i 2018: 1.434.941 kr.

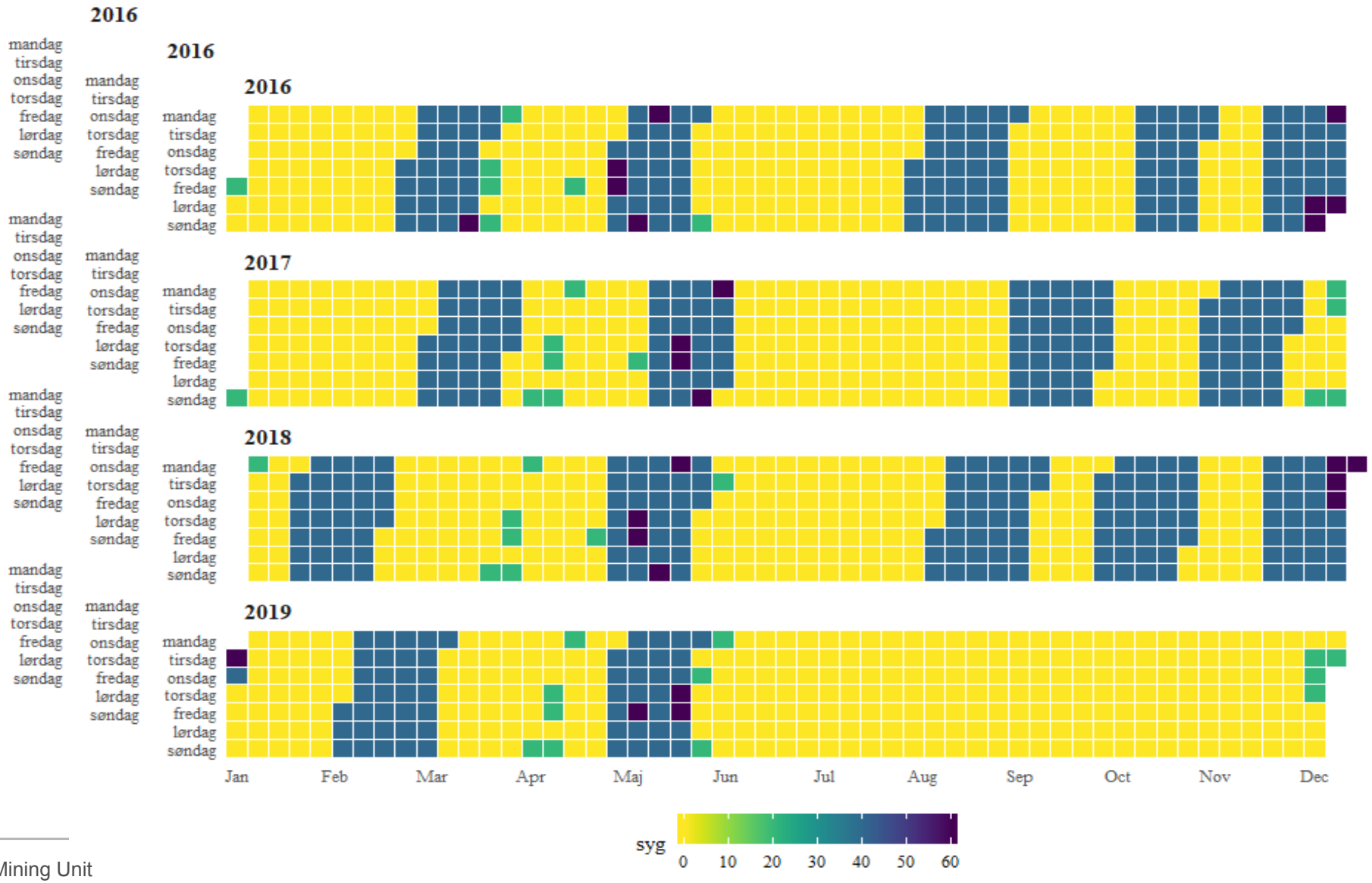
Graphical mapping of the sick periods

Den Fælles Dataenhed

Owner of the company/Mother

Daughter

Son



Questions?