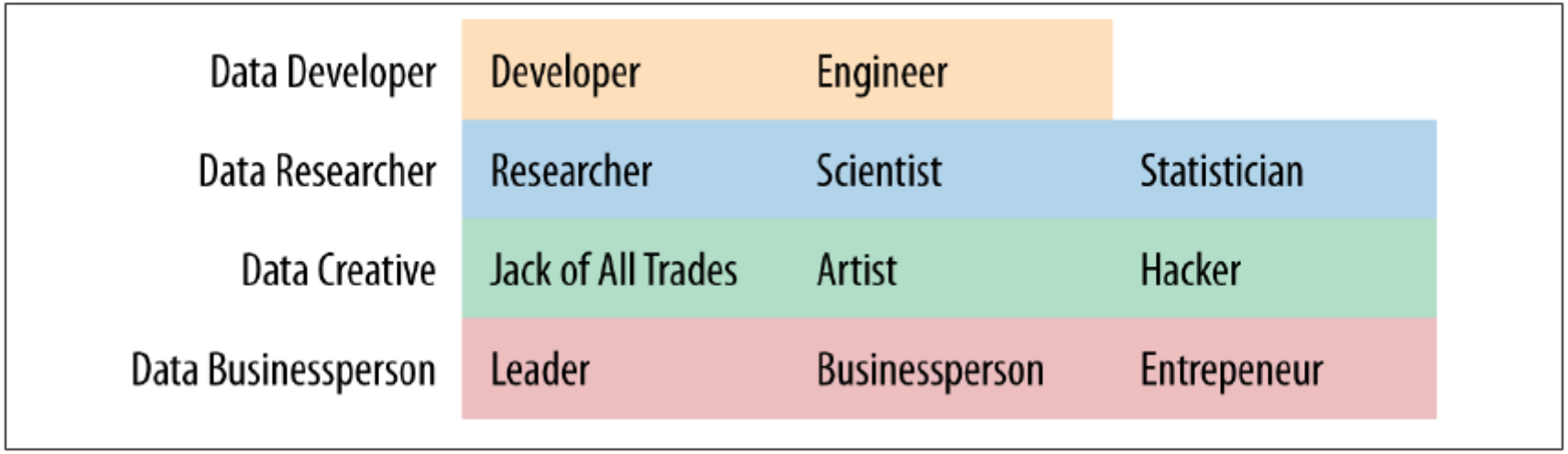
Data Scientist – Role and Use Cases

Analyses of Elevators & Escalators – a Pilot Project for a comprehensive Availability Management System
Energy Data Management – pillar of intelligent Virtual Power Plants

The Role of a Data Scientist
Self-Identification

I think of myself as an …

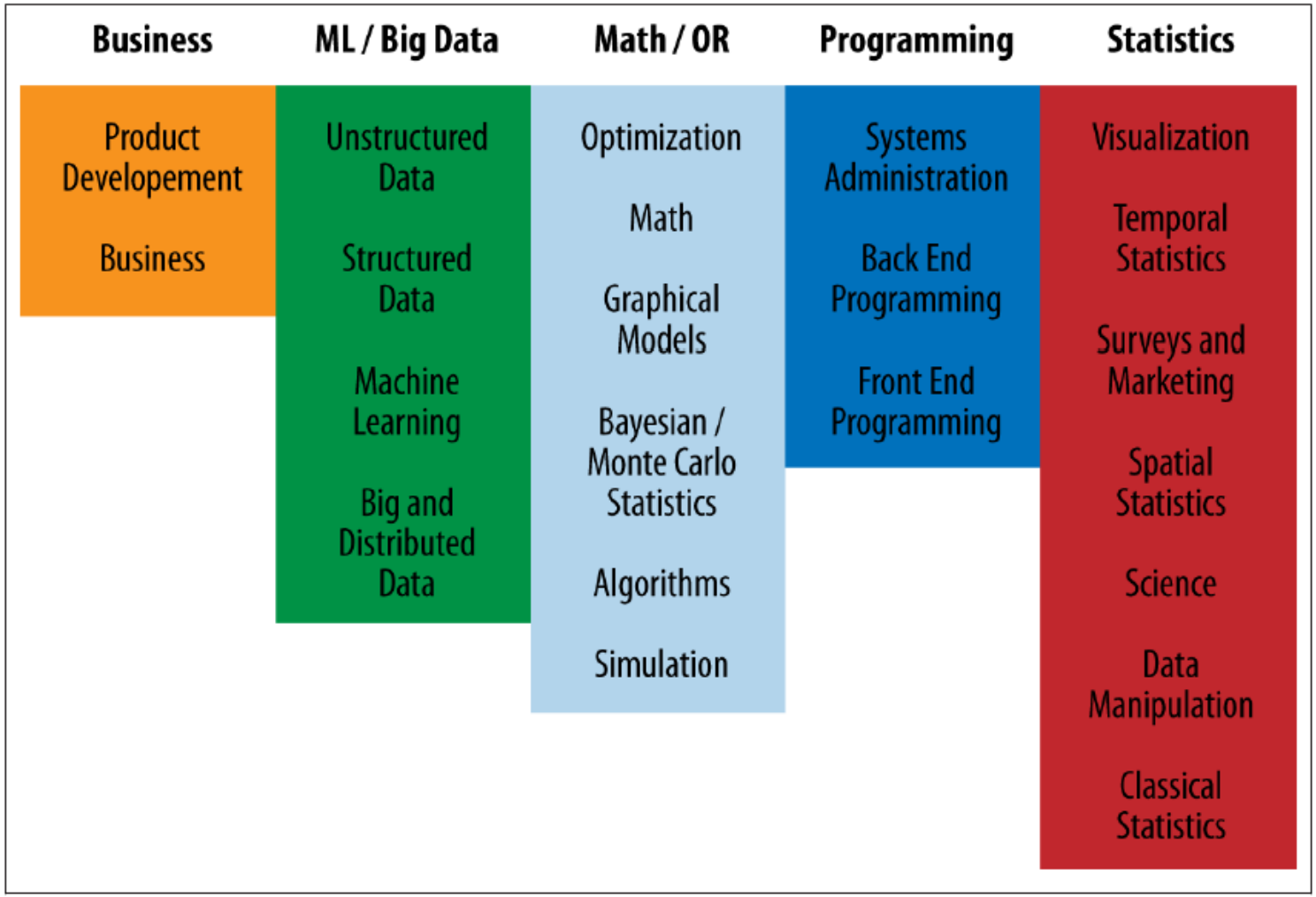


Source: Harris/Murphy/Vaisman (2013): Analyzing the Analyzers

The Role of a Data Scientist
Skills

DB Mobility Networks Logistics

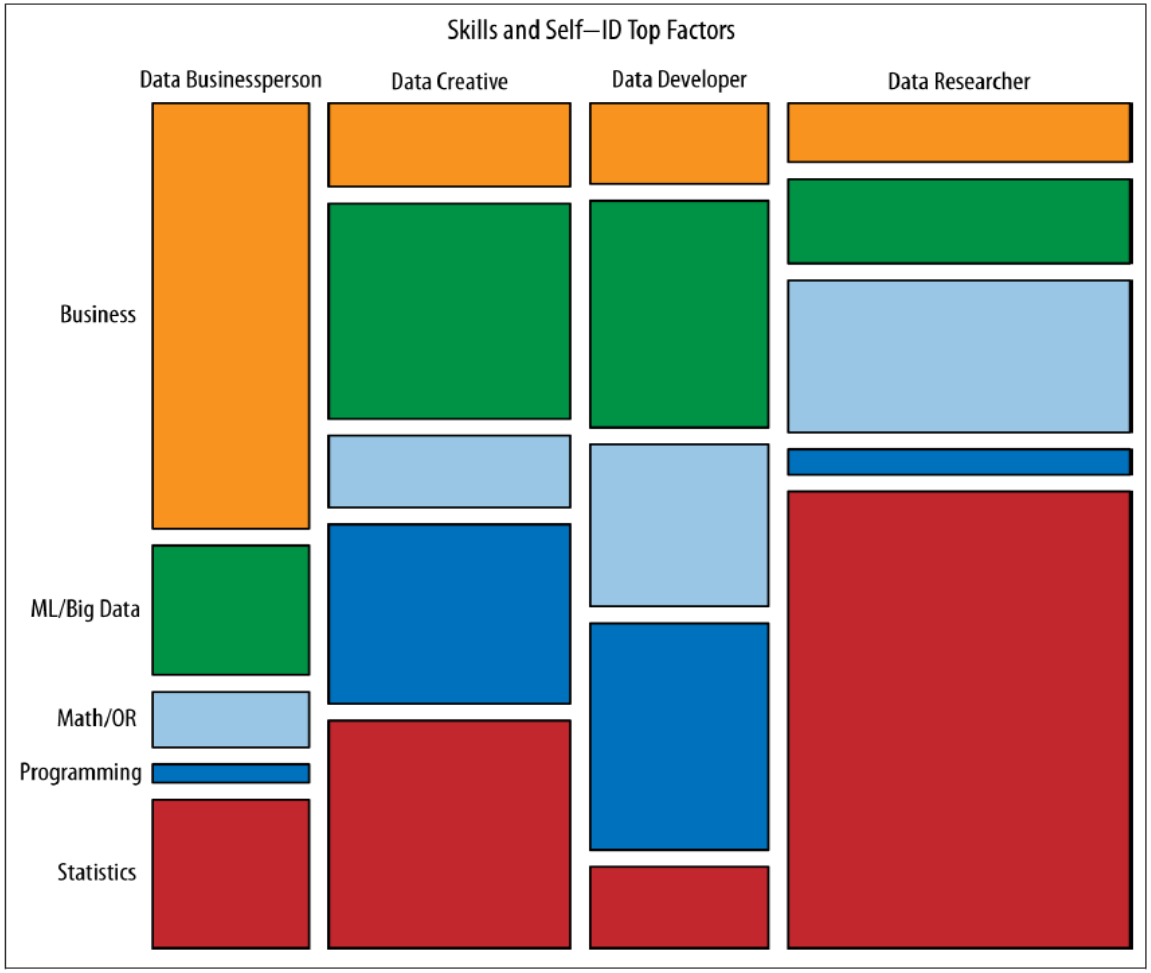
What skills do you bring to your work?



Source: Harris/Murphy/Vaisman (2013): Analyzing the Analyzers

The Role of a Data Scientist
Skill sets differ: collaboration is essential!

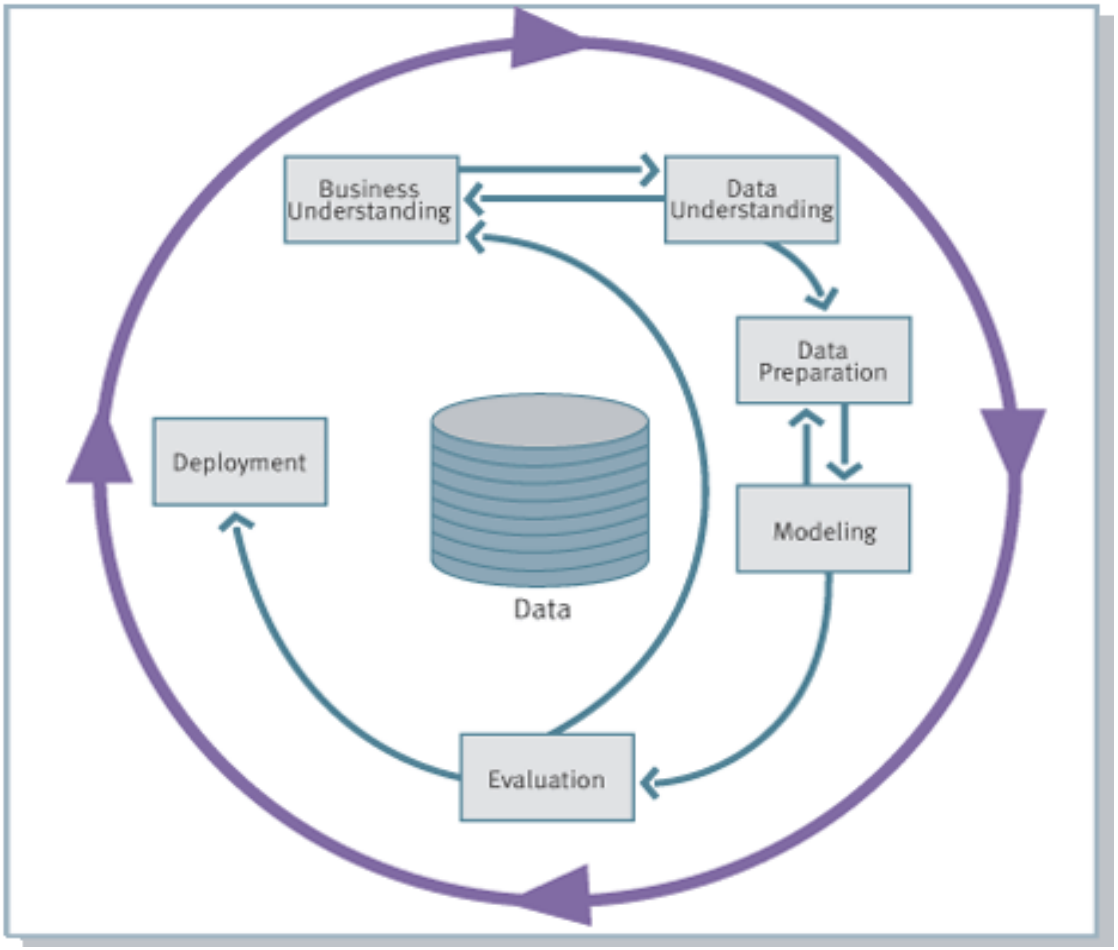
Data Scientists assigning themselves to different
“Self-IDs” reveal different and correlated skill ratings



Source: Harris/Murphy/Vaisman (2013): Analyzing the Analyzers

Data Science:
Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Analyses of Elevators and Escalators: Summary

Internal and external Data

- Sensor Data
- Malfunction Notifications
- Master Data
- Service Contracts
- Spare Parts
- Traveler Traffic Data
- Event data (e.g. exhibitions)
- Geo- and Weather Data
- Social Media Data

Infrastructure

Data Quality

Patterns

Clusters

Causality

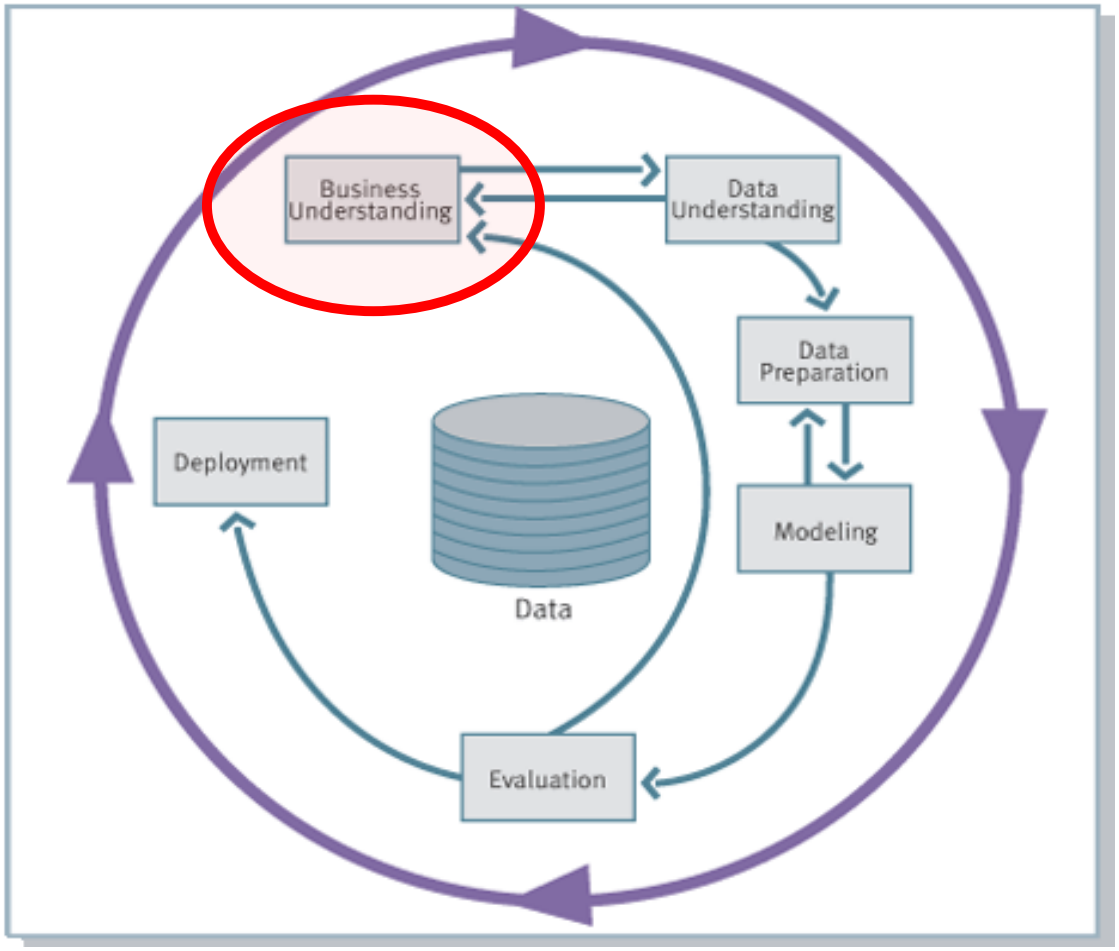
Models

Goals

- **Measurement, Visualization and Forecasting of Availability**
- **Assessment and Optimization of Service Level Agreements**
- **Customer Information and Customer Feedback**

Analyses of Elevators and Escalators: Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Business Understanding

Business questions:

1

Utilization of current and future sensor data:

- How to keep results of already ongoing analysis?
- Does it make sense to integrate external data e.g. weather?
- Can infrastructure data deliver more insights when looking also on geo-data?
- How to interprete the current status delivery of different data sources?

2

Security of client information:

- Which information can be detected from the data analysis?
- How fast insights can be gained?
- Which new applications can be created? How to integrate a client feedback?
- How to derive quick client information based on value borders?

3

Availability and visualization of malfunctions:

- How to improve the availability of elevators and escalators by data analytics?
- How to predict the outage behaviour?
- How to visualize malfunctions on dashboards?
- How to transfer the pilot project into an industrialized product?

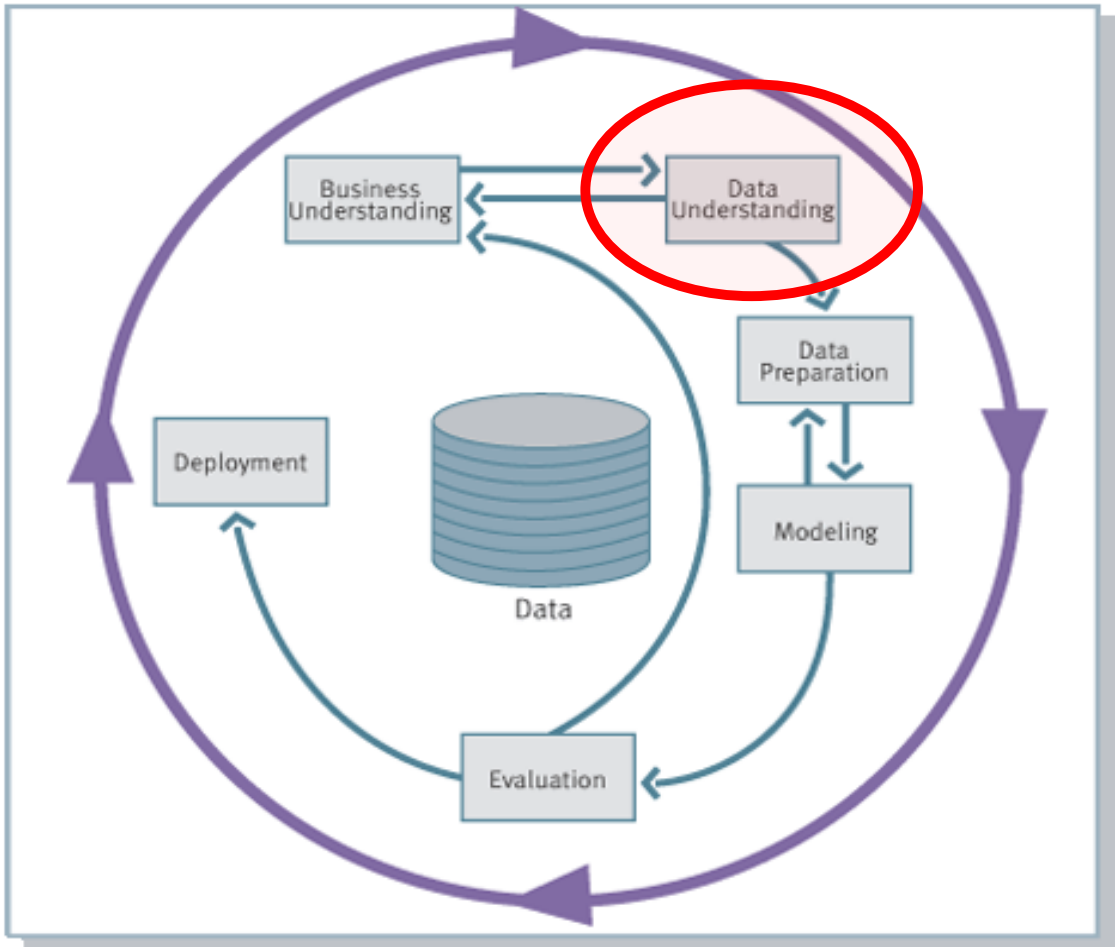
4

Infrastructure building:

- What are the system requirements of an appropriate solution?
- How to define the target architecture?
- Which components (e.g. sandbox) have to be part of a solution?

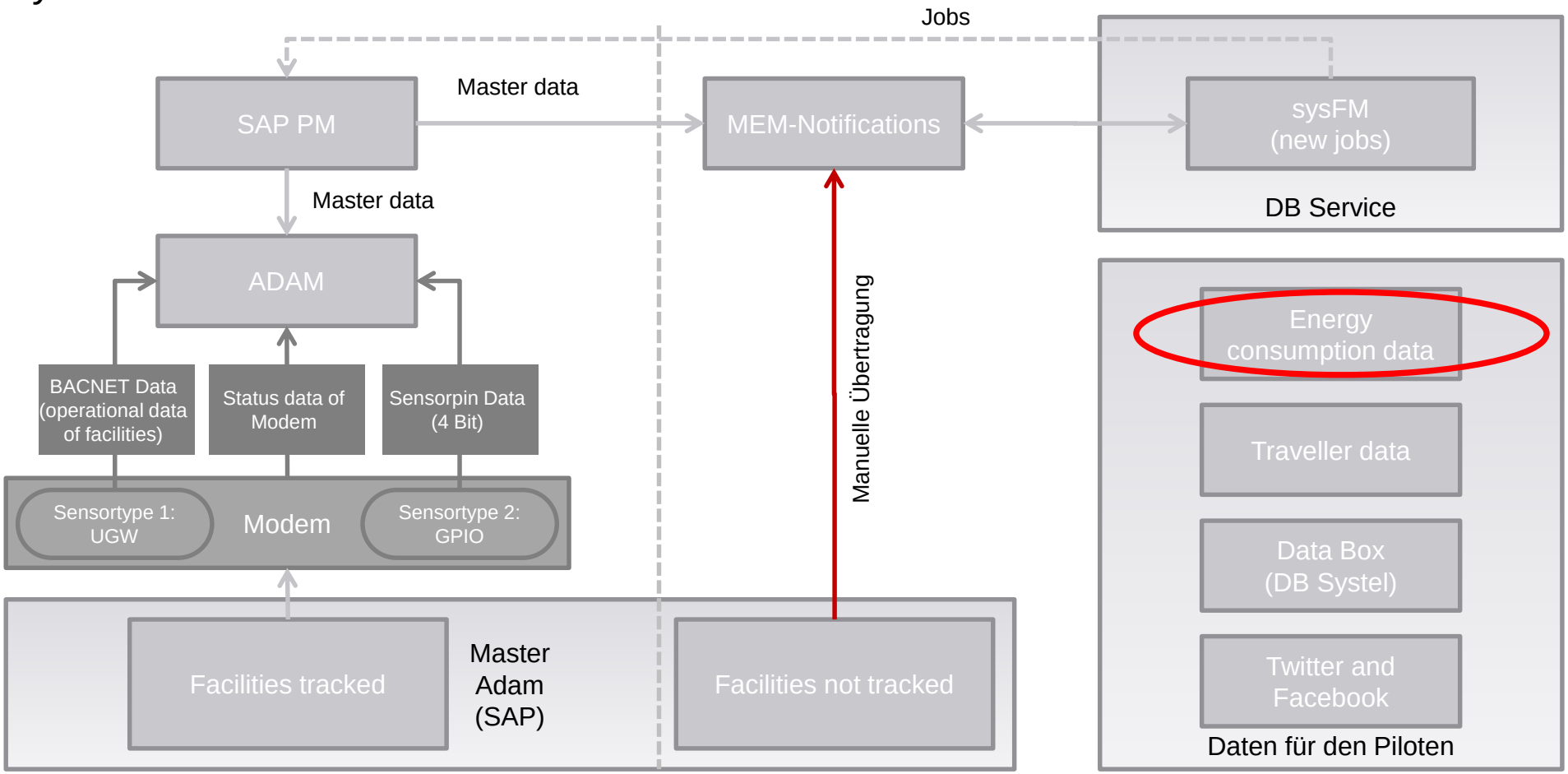
Analyses of Elevators and Escalators: Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Data Understanding

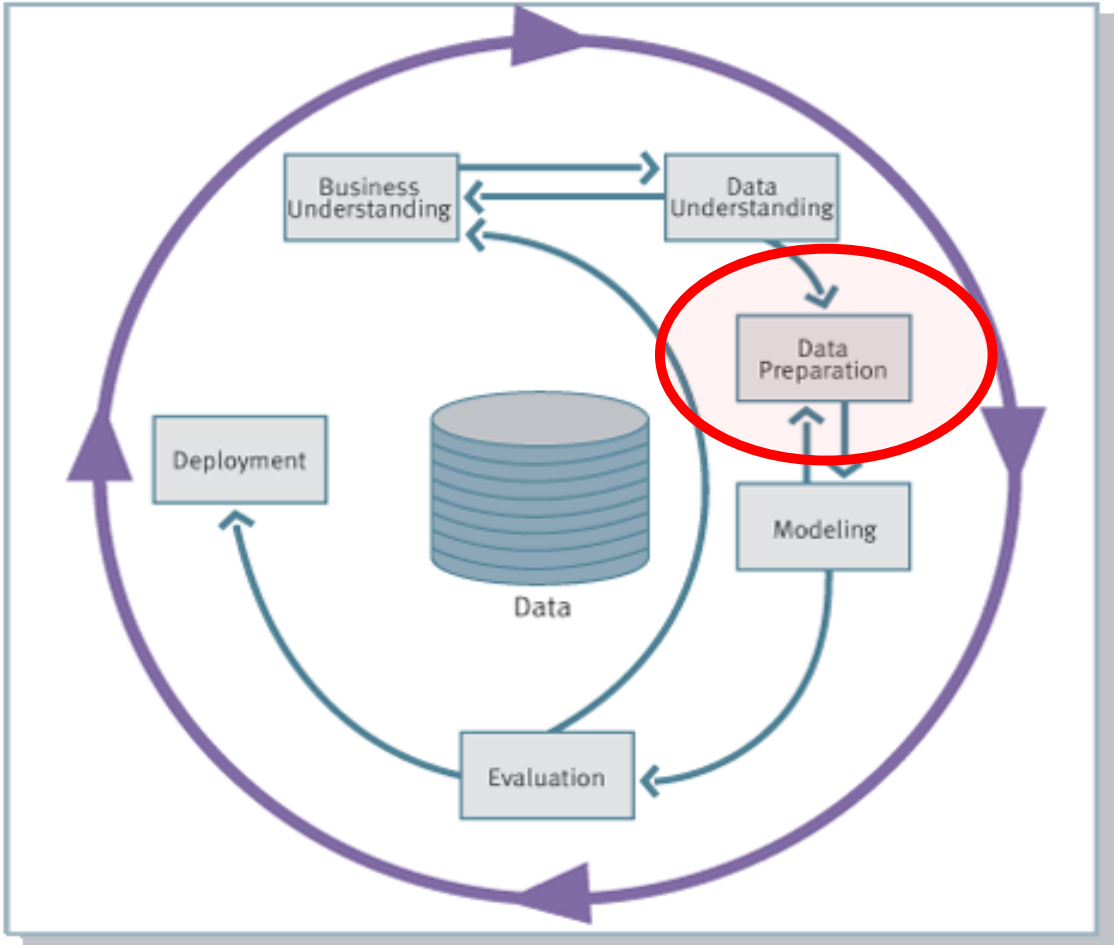
System and Architecture



► **Current solution's architecture: parallel working systems with different perspectives without (common) analytical „Back-End".**

Analyses of Elevators and Escalators: Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Data Preparation
Text analysis: MEM
Step 1: Breakdown of texts in rows according to time stamp

Example: Content of free text field „BESCHREIBUNG“

MELDUNGS NUMMER	EQUIPMENT	ERSTELLT	Zeitstempel	BESCHREIBUNG
M-00447614	10132576	24.04.2015 08:13:54		[Fr 24.04.2015 08:13:54 - Peters, Timo - Erfassung]: HF 62 stillgelegt aufgrund unerträglichen Schlaggeräuschen [Fr 24.04.2015 14:55:07 - Demus, Silvio - Gemeldet]: Handlauf Antriebskette Rückmeldung Herr Bauer [Do 07.05.2015 10:18:22 - Hermann, Georg - Gemeldet]: Meldung DB Services Bauer. Anlage wieder i.O.
M-00458242	10132576	12.05.2015 07:33:25		[Mi 13.05.2015 09:40:21 - Fürst, Antal - Gemeldet]: Rückmeldung DB Services Hr. Bauer
M-00463711	10132576	21.05.2015 09:33:28		[Do 21.05.2015 09:33:28 - Knodel, Andreas - Erfassung]: Bei Handlaufumlenkung oben links starke Schlaggeräusche. Es befindet sich auch Abrieb dran!
M-00491492	10132576	01.07.2015 15:01:55		[Mi 01.07.2015 15:01:55 - Knodel, Andreas - Erfassung]: Bitte danach schauen [Mi 15.07.2015 14:39:59 - Peters, Timo - Gemeldet]: HF steht [Mo 20.07.2015 15:21:16 - Zill, Silvio - Gemeldet]: Rückruf Techniker Stanzu wieder i.O.

1 row

Example break down of „BESCHREIBUNG“

M-00447614	10132576	24.04.2015 08:13:54	24.04.2015 08:13:54	HF 62 stillgelegt aufgrund unerträglichen Schlaggeräuschen
M-00447614	10132576	24.04.2015 08:13:54	24.04.2015 14:55:07	Handlauf Antriebskette Rückmeldung Herr Bauer
M-00447614	10132576	24.04.2015 08:13:54	07.05.2015 10:18:22	Meldung DB Services Bauer. Anlage wieder i.O.
M-00458242	10132576	12.05.2015 07:33:25	13.05.2015 09:40:21	[Mi 13.05.2015 09:40:21 - Fürst, Antal - Gemeldet]: Rückmeldung DB Services Hr. Bauer
M-00463711	10132576	21.05.2015 09:33:28	21.05.2015 09:33:28	[Do 21.05.2015 09:33:28 - Knodel, Andreas - Erfassung]: Bei Handlaufumlenkung oben links starke Schlaggeräusche. Es befindet sich auch Abrieb dran!
M-00491492	10132576	01.07.2015 15:01:55	01.07.2015 15:01:55	Bitte danach schauen
M-00491492	10132576	01.07.2015 15:01:55	15.07.2015 14:39:59	HF steht
M-00491492	10132576	01.07.2015 15:01:55	20.07.2015 15:21:16	Rückruf Techniker Stanzu wieder i.O.

3 rows

► **Breakdown of MEM-notifications successful**

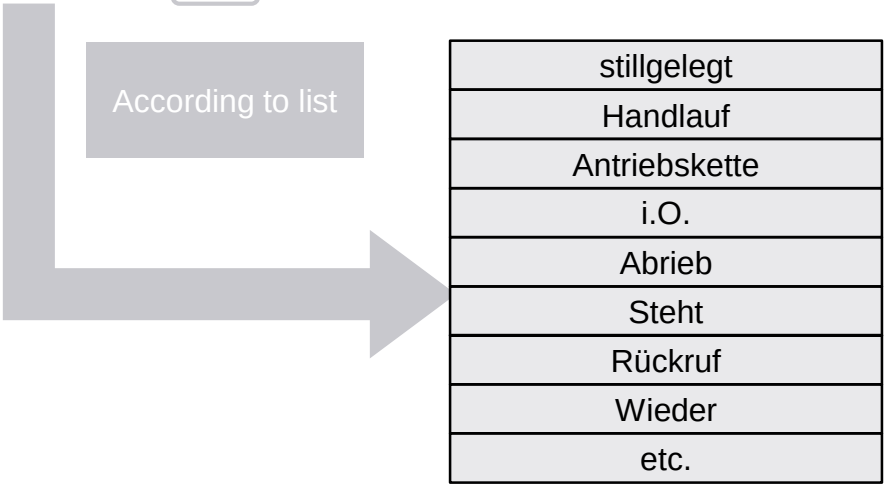
Data Preparation
Text analysis: MEM
Step 2: Detection of mostly used terms

Example: Content of free text field „BESCHREIBUNG“

M-00447614	10132576	24.04.2015 08:13:54	24.04.2015 08:13:54	HF 62 stillgelegt aufgrund unerträglichen Schlaggeräuschen
M-00447614	10132576	24.04.2015 08:13:54	24.04.2015 14:55:07	Handlauf Antriebskette Rückmeldung Herr Bauer
M-00447614	10132576	24.04.2015 08:13:54	07.05.2015 10:18:22	Meldung DB Services Bauer. Anlage wieder i.O.
M-00458242	10132576	12.05.2015 07:33:25	13.05.2015 09:40:21	[Mi 13.05.2015 09:40:21 - Fürst, Antal - Gemeldet] Rückmeldung DB Services Hr. Bauer
M-00463711	10132576	21.05.2015 09:33:28	21.05.2015 09:33:28	[Do 21.05.2015 09:33:28 - Knodel, Andreas - Erfassung]: Bei Handlauf umlenkung oben links starke Schlaggeräusche. Es befindet sich auch Abrieb dran!
M-00491492	10132576	01.07.2015 15:01:55	01.07.2015 15:01:55	Bitte danach schauen
M-00491492	10132576	01.07.2015 15:01:55	15.07.2015 14:39:59	HF steht
M-00491492	10132576	01.07.2015 15:01:55	20.07.2015 15:21:16	Rückruf Techniker Stanzu wieder i.O.

Solution:

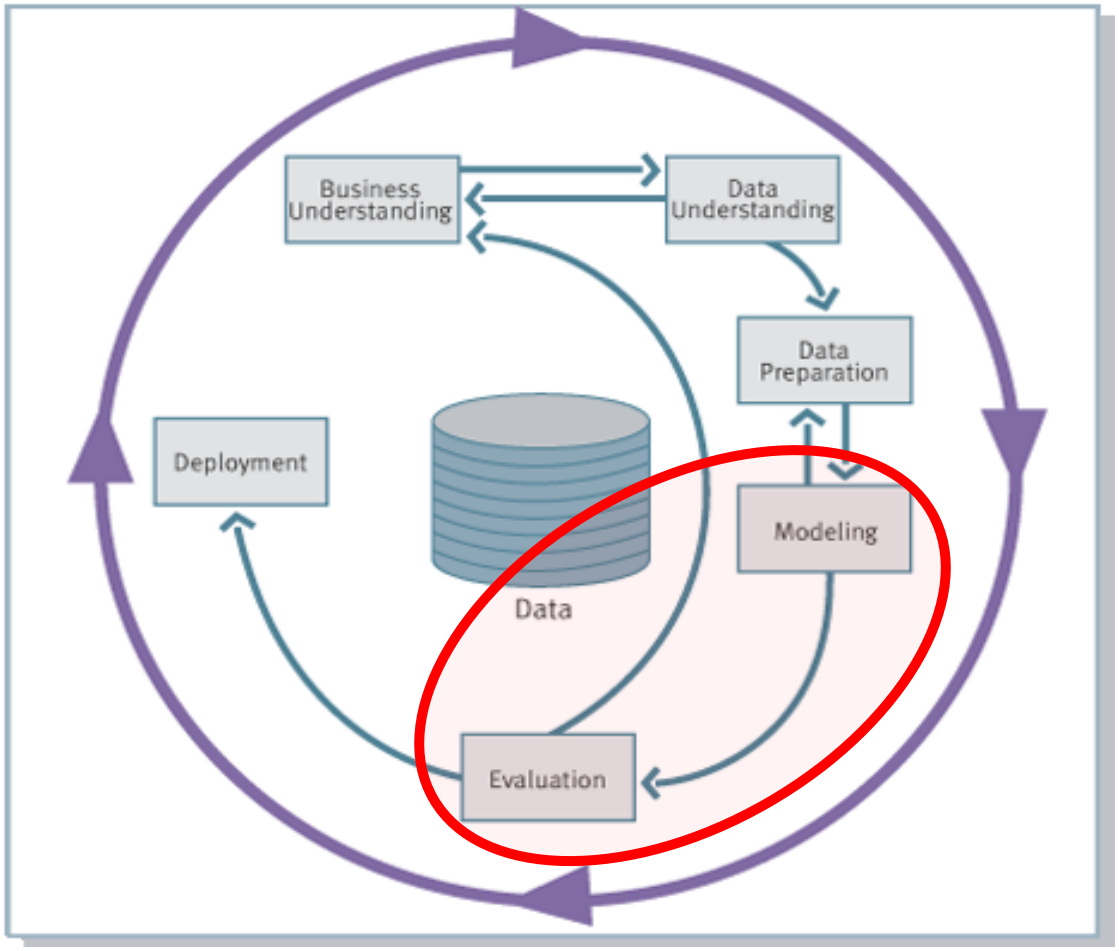
- List of all used words
- Creation of Stop-List: Eliminations of words without content (e.g. „ist“). Considerations of abbreviations (e.g. „i.O.“) and related words („außer Betrieb“)
- Creation of synonyms (e.g.. „Glas“ for Glasscherben, Verglasung, Glastür, etc.)



► **Mapping of free text successful**

Analyses of Elevators and Escalators: Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Data Preparation / Modeling
Text analysis: MEM
Step 3: Assessment of created list of words

MEM-notifications – Most frequently used words in BESCHREIBUNG, which allow detection of reasons

reasons

Elevators

Notifications with reasons:

3.267 (18%)

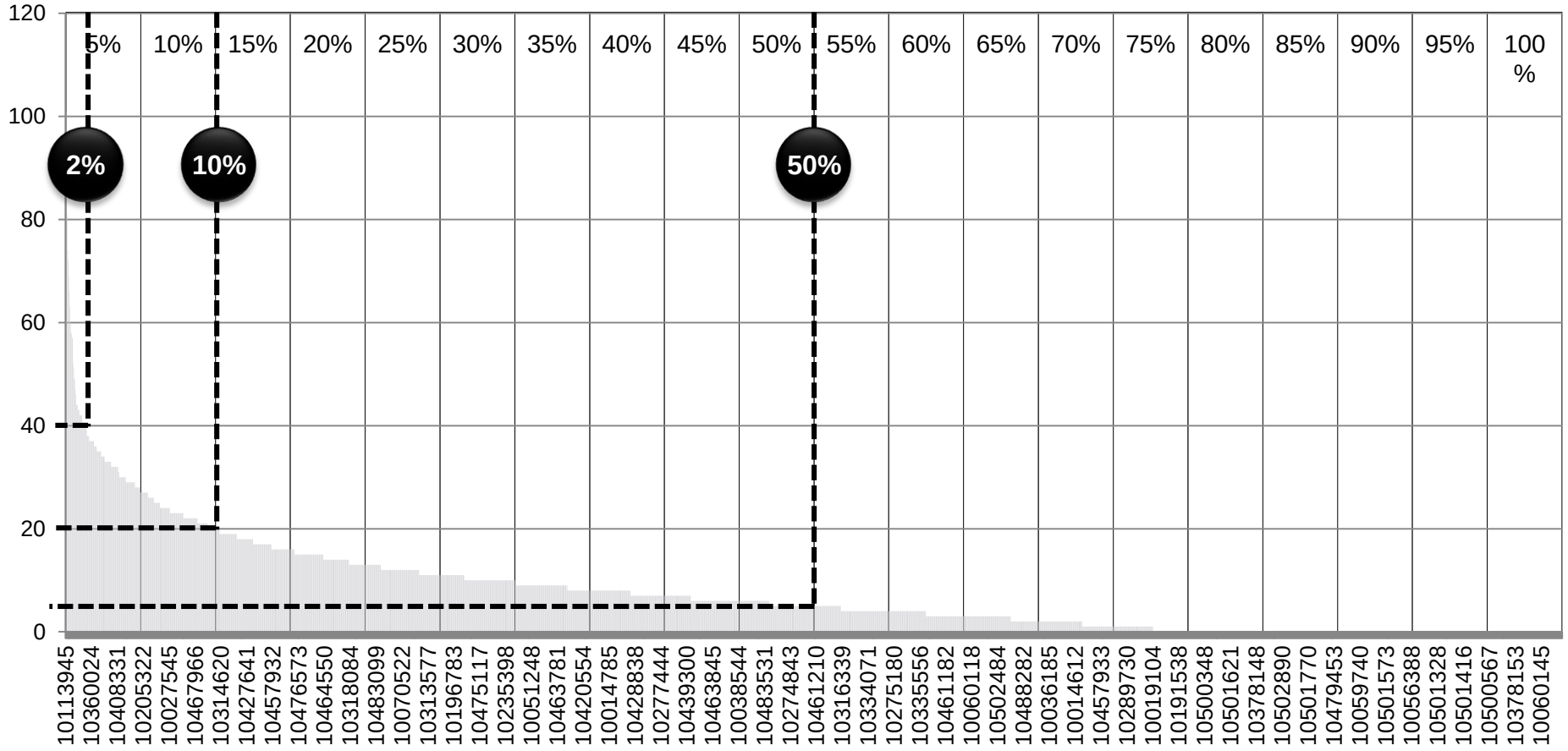
Notifications without reasons

17.719 (82%)

Zeile 1		Zeile 2		Zeile 3		Zeile 4	
Term	Anzahl	Term	Anzahl	Term	Anzahl	Term	Anzahl
tür	2275	tür	201	tür	110	tür	48
kabine	415	ersatzteil	104	ersatzteil	54	ersatzteil	31
taste	389	taste	56	taste	29	taste	14
beleuchtung	220	feuerwehr	38	feuerwehr	23	feuerwehr	7
ständig	191	sprechverbindung	28	evu	19	wasser	6
geräusch	180	evu	28	ständig	11	schlüssel	5
sprechverbindung	152	kamera	23	schacht	10	kabine	5
sammelstörung	140	ständig	21	lichtschranke	10	evu	5
leuchten	129	beleuchtung	21	kabine	10	lichtschranke	5
wasser	127	wasser	20	schlüssel	10	hydraulik	4
feuerwehr	95	lichtschranke	20	bedienung	9	system	3
schacht	91	kabine	18	bündi-unbündig	9	schacht	3
rot	83	nagel	16	wasser	8	ständig	3
verbindung	79	schlüssel	15	platine	7	steuerung	3
bedienung	77	verbindung	14	fremdkörper	6	handlauf	3
bündig-unbündig	76	geräusch	14	geräusch	6	nagel	2

- Elevators: 20% of all MEM-notifcations led to reasons. Top reason: „Tür“
- Spare parts are significant for all other reparations.

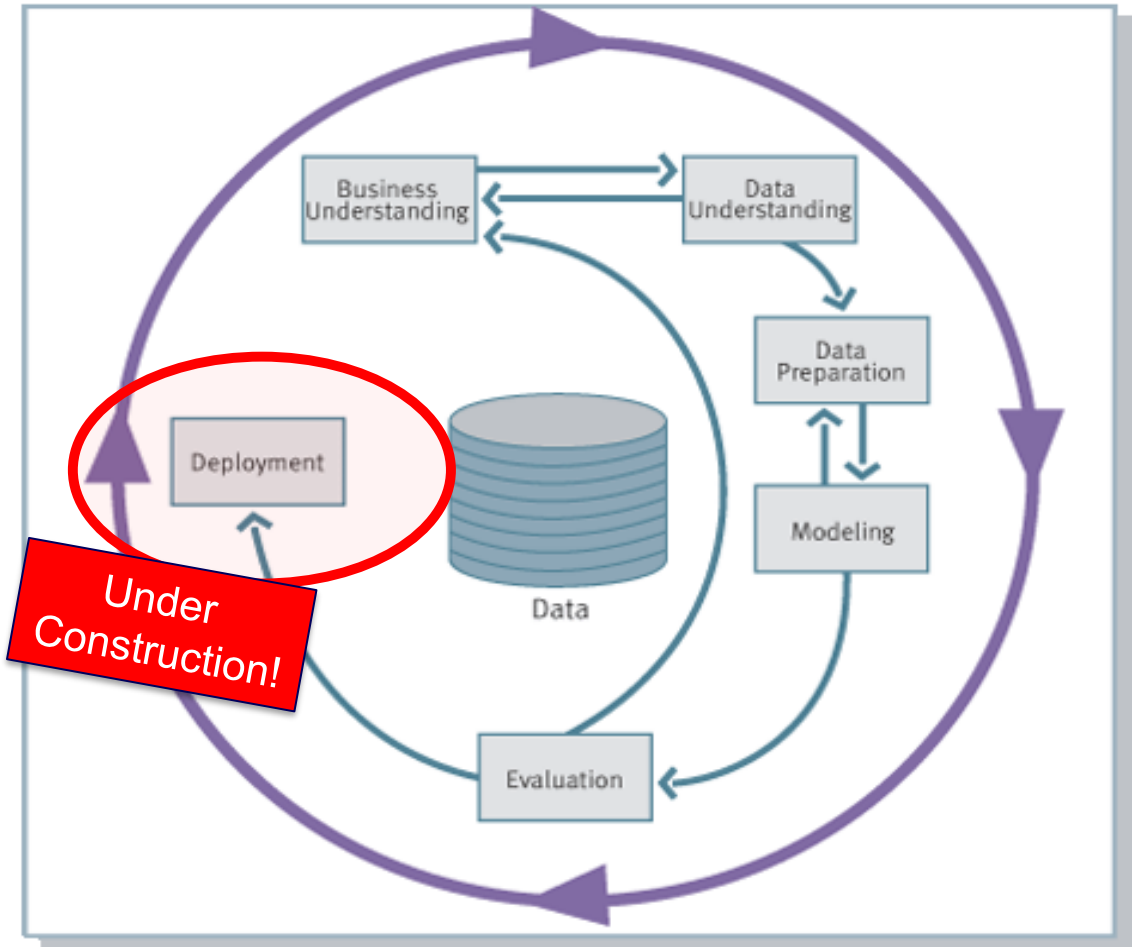
Modeling
Overview repeating notificiations



► 2% of all elevators cause more than 40 notifications per year and elevator.
50% with less than 5 notifications per year.

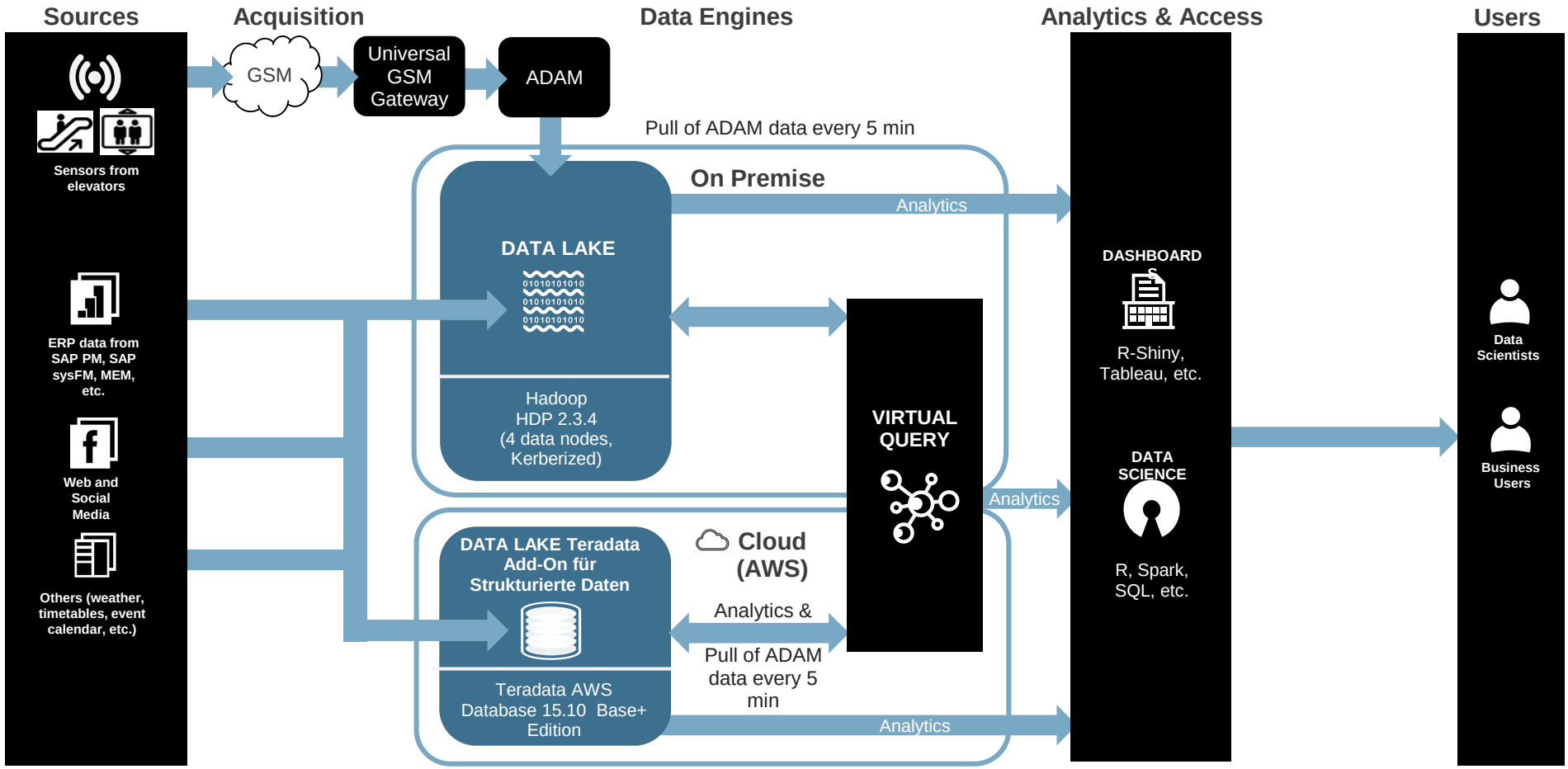
Analyses of Elevators and Escalators: Overview of the Process

Reference: Cross Industry Standard Process for Data Mining (CRISP-DM)



Current Infrastructure for the Pilot

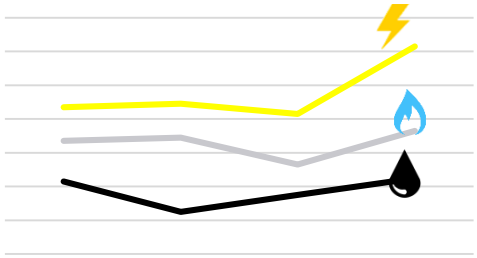
Pilot Verfügbarkeitsmanagement



► **Basement for architecture are existing explorations systems plus added components to allow industrialization of solution.**

Increase of energy efficiency:
High prices & consumption as external & internal drivers

Rising prices



- Electricity + 165%
- Gas + 82%
- Diesel + 69%

Total primary energy

91,026 TWh
=
22.7 Millions HH
4 Persons HH with 4.000 kWh/a

Water

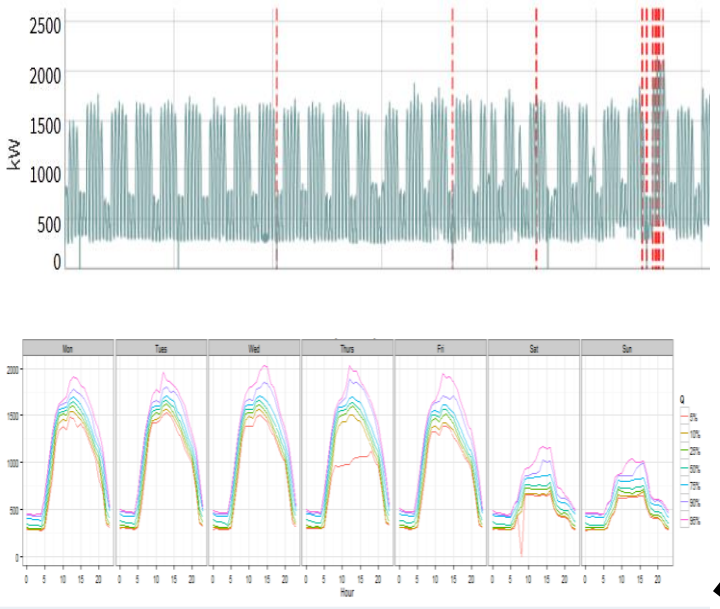
9,31 Mio. m³
=
212.000 HH
4 Persons HH with 120l/day

Diesel

415 Mio. liter / year via DB Energie

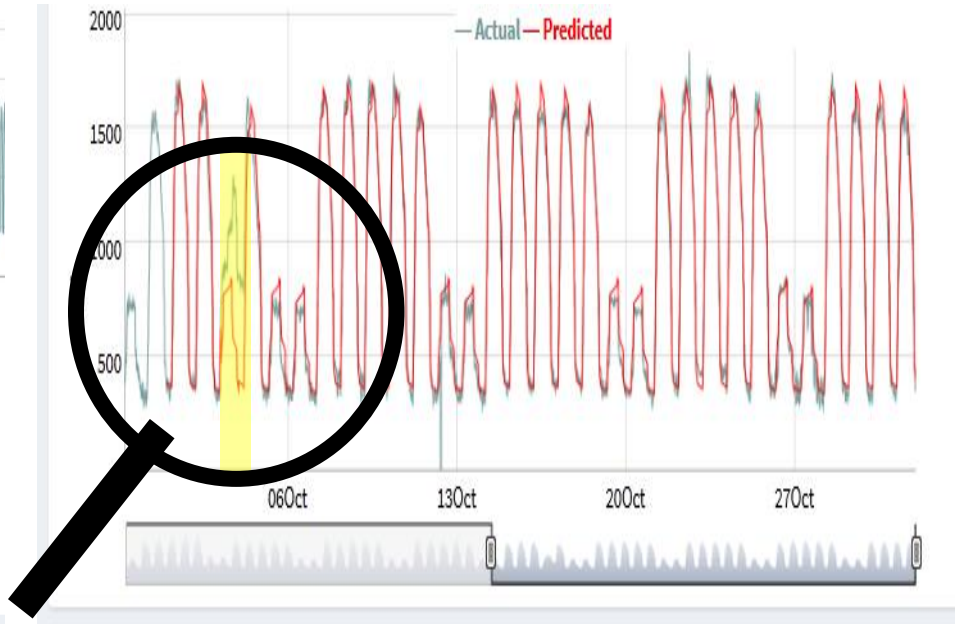
Increase of energy efficiency: Beginning with basic analysis of stationary consumption

Necessities:
Data availability & quality



- Transparency is key for intelligence

Customer based analysis lead to higher efficiency potential and predictions of higher quality

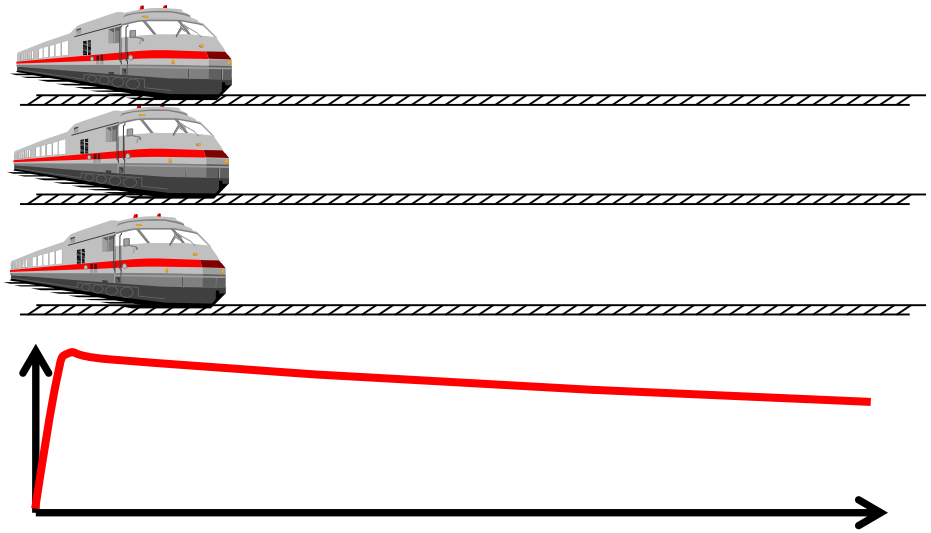


- Supervision of predictors and model training lead to higher reliability

- Manuel notifications
- External predictors
- Data Mining
- Event Pattern Management
- Sub-Customer

Increase of energy efficiency: Efficiency potential detection in various systems

Efficiency potential 1: Parallel departure of ICE

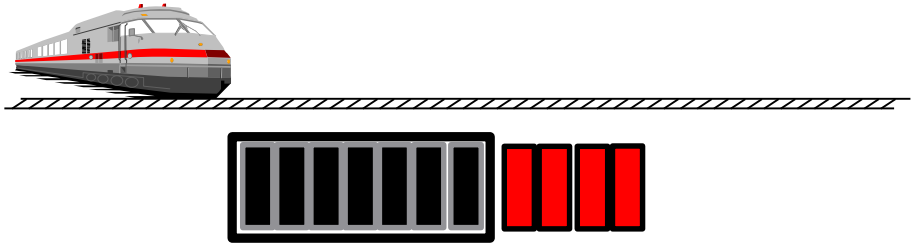


Constant need of 8 MW to reach top speed

6 minutes of accelerating

= consumption of ~18.000 households

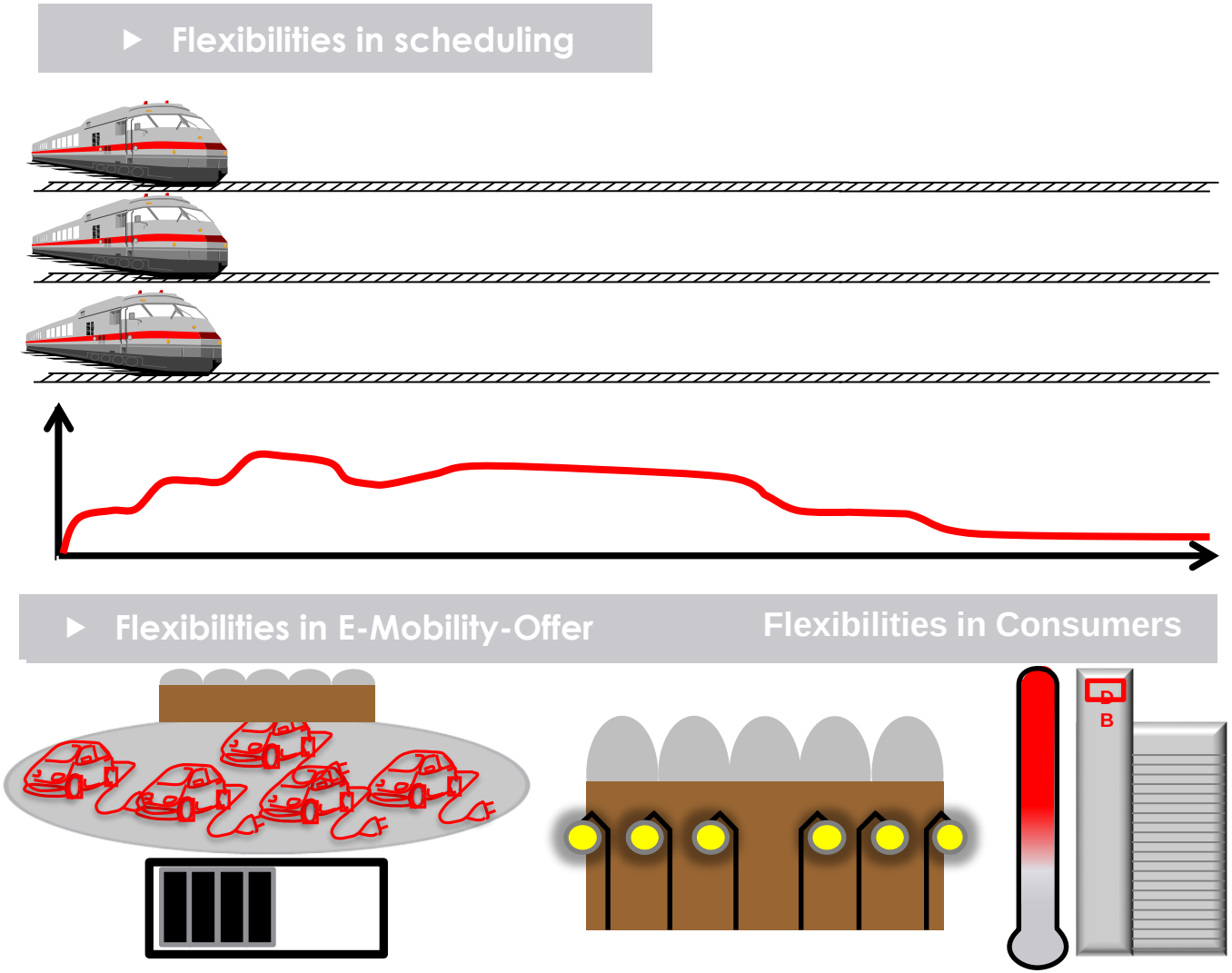
Efficiency potential 2: Win back of braking energy



Track München-Stuttgart:

20% win back of utilized energy

Increase of energy efficiency:
Down the road: Interactive systems based on real-time analysis



Thank you for your attention!

Questions or suggestions ...?

i5.startup
@deutschebahn.com