

IT-Summit 2017

Thomas Pflügl

Data Scientist - ENGEL Austria GmbH

Hallo, mein Daten-Schatz!

Big Data, Little Data, No Data in der Industrie

Hallo, mein Daten-Schatz!

Am 08. März 2017 ist Internationaler Frauentag

"Dating Scientist"



Source: © boxwoodavenue.com, morgenpost.de

Hallo, mein Daten-Schatz!

Am 08. März 2017 ist Internationaler Frauentag

"Dating Scientist"

Frauen in Berlin

Zahlen, Daten und Fakten zum Internationalen Frauentag

1.779.366

Frauen in Berlin

Familienstand

ledig

46%

verheiratet

34%

11%

verwitwet

9%

geschieden

44 Jahre

Durchschnittsalter der Berliner Frauen

Branchen, in denen die meisten Frauen arbeiten

Gesundheits- und Sozialwesen

129.378

Erziehung und Unterricht

48.520

Handel
73.977

Dienstleistungen
(Freiberuflerinnen,
Technik,
Wissenschaft)

55.407

Mobil in Berlin

Wegeanteile nach Verkehrsmitteln

28%
ÖPNV

32%
zu Fuß

28%
Auto

12%
Fahrrad

Erwerbstätigkeit

866.100 erwerbstätige Frauen

12 %
selbständig

88 %
angestellt

Einkommen

mittleres Nettoeinkommen
(nur Erwerbstätige)

1425 € **1600 €**

33,5 Stunden

durchschnittliche
Arbeitszeit pro Woche

Berlinweit in der Mehrheit

Frauen
51,1%

Männer
48,9%

Aufteilung in den Bezirken

Frauenanteil

- 48 bis 49%
- 49 bis 50%
- 50 bis 51%
- 51 bis 52%
- 52 bis 53%
- 53 bis 54%

IT-Summit 2017



Zur Person: Thomas Pflügl - Data Scientist

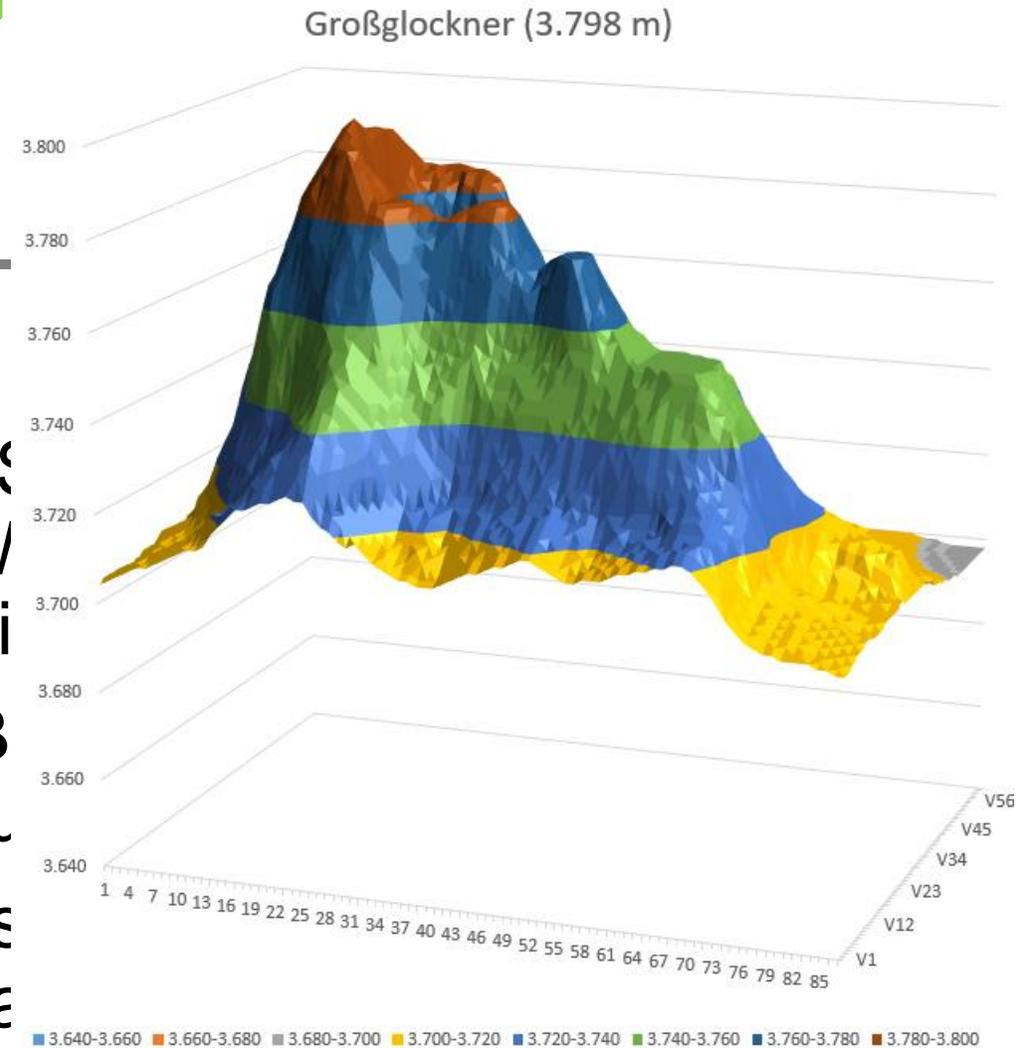
- 51 Jahre
- Landesalpinreferent OÖ, Bergsteigen, Skitouren, Klettern & Klettersteige, Sprach- und Wortschätze, Reisen, Alte Musik, Psychologie und Philosophie
- Komme aus dem Asset Management Business (Advanced Risk Management Tools to quantify and reduce risk)
- Deutsch, Englisch, Spanisch, Französisch, (Italienisch), ein paar Script-/Programmiersprachen wie Visual Basic, SQL, Python, (R), HTML, PHP, JavaScript
- seit 08-2015 bei ENGEL AUSTRIA GmbH (*Bekanntnis zu Data Science*)

Source: © Eigene Grafik

IT-Summit 2017

Zur Person: Thomas Pflügl -

- 51 Jahre
- Landesalpinreferent OÖ, Bergsteigen, Klettern & Klettersteige, Sprach- und W
Alte Musik, Psychologie und Philosophie
- Komme aus dem Asset Management B
Management Tools to quantify and redu
- Deutsch, Englisch, Spanisch, Französis
Script-/Programmiersprachen wie Visual
HTML, PHP, JavaScript
- seit 08-2015 bei ENGEL AUSTRIA GmbH (*Bekanntnis zu Data Science*)



Source: © Eigene Grafik

ENGEL

www.engelglobal.com

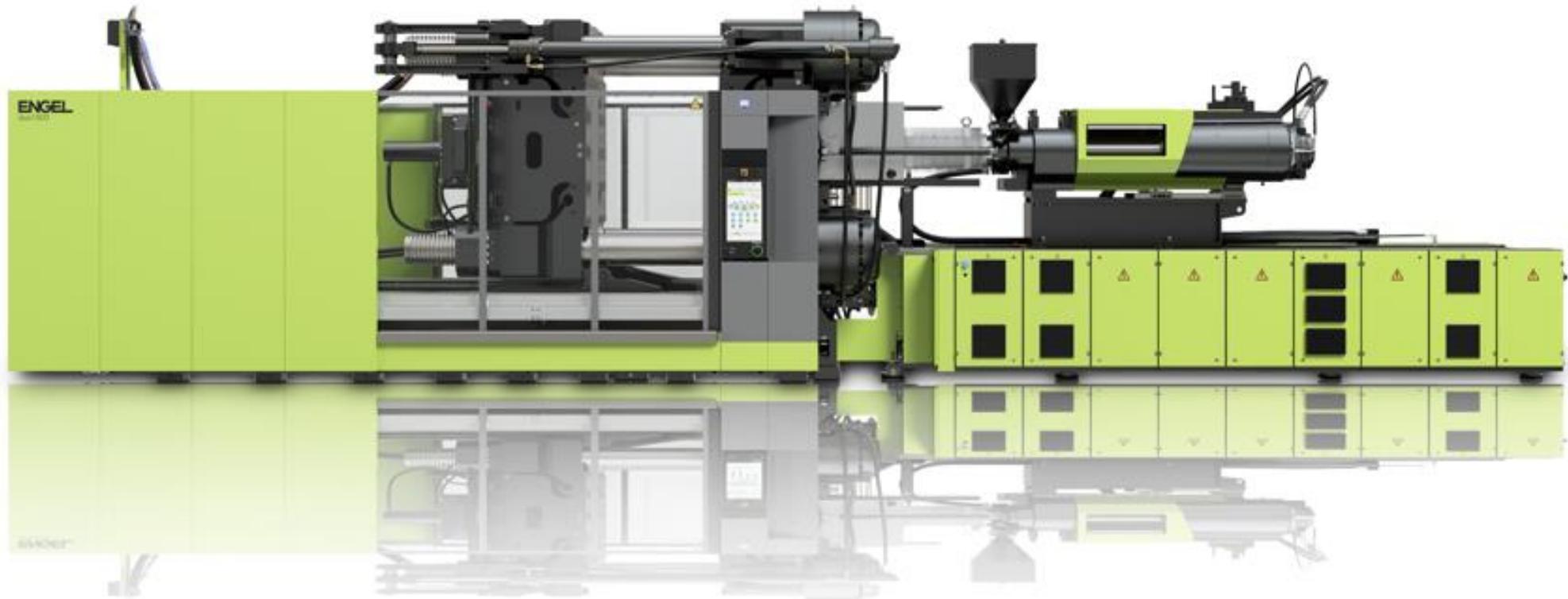
Die ENGEL Success Story? Kurz erklärt:



Source: © Salzburger Nachrichten/APA/Herbert Pfarrhofer

Die ENGEL Success Story

ENGEL: Weltmarktführer bei Spritzgießmaschinen



Produktionswerke

9 Produktionsstandorte

Europa, Nordamerika und Asien
(China, Korea)

Schwertberg, St. Valentin, Dietach, Hagen (D), Kaplice (CZ),
York (US), Shanghai + Changzhou (CN), Pyungtaek-City (KR)

Die ENGEL Success Story

Niederlassungen/Vertretungen und Mitarbeiter

In über 85 Ländern



weltweit 5.400



Österreich 3.000

Source: © dreamstime.com, fotolia.com

Die ENGEL Success Story

ENGEL - Umsatz 2015/2016

1,25 Milliarden Euro

ENGEL - Investitionen im Geschäftsjahr 2016/17

100 Millionen Euro

ENGEL Business Unit - Automotive

Innenraum

- Instrumententafel
- Sitzschale
- Armlehne
- Lautsprecherabdeckung
- Türverkleidung

Klimatisierung

- Filter
- Gehäuse
- Lüfter
- Ventilatoren
- Belüftungsklappen

Beleuchtung

- Frontscheinwerfer
- Heckleuchten
- Reflektoren
- LED Technologie

Was stellen
ENGEL
Spritzgieß-
maschinen
her?



ENGEL Business Unit - **Medical**

Medizintechnik

- Einwegspritzen
- Infusionssysteme
- Dialysesysteme

Healthcare

- Zahnreinigungsprodukte
- Babycare



ENGEL Business Unit - Packaging

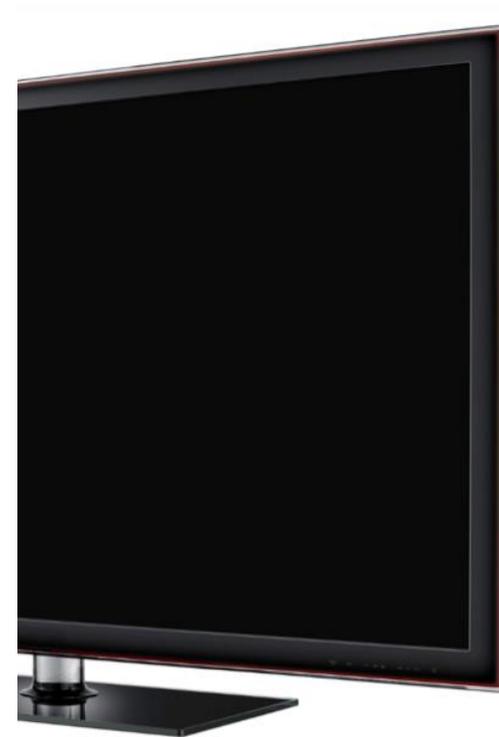
Dünnwandverpackungen

- Becher und Behälter
- Rund- und Rechteck-Eimer

ENGEL Business Unit - Teletronics

Displays

- Flachbildschirme
- Monitore
- Laptops, Tablets, E-Book Reader



ENGEL Business Unit - Technical Moulding

Möbel und Büroartikel

- Bürostühle
- Griffe
- Möbelrollen
- Stempel

Spielwaren

- Spielbausteine
- Spielfiguren
- Strategiespiele
- Spielzeugfahrzeuge

Haushalt

Sport und Freizeit

- Sonnenbrillen
- Skibrillen
- Ski & Snowboards



Big Data - Kulturwandel - Mentaler Wandel

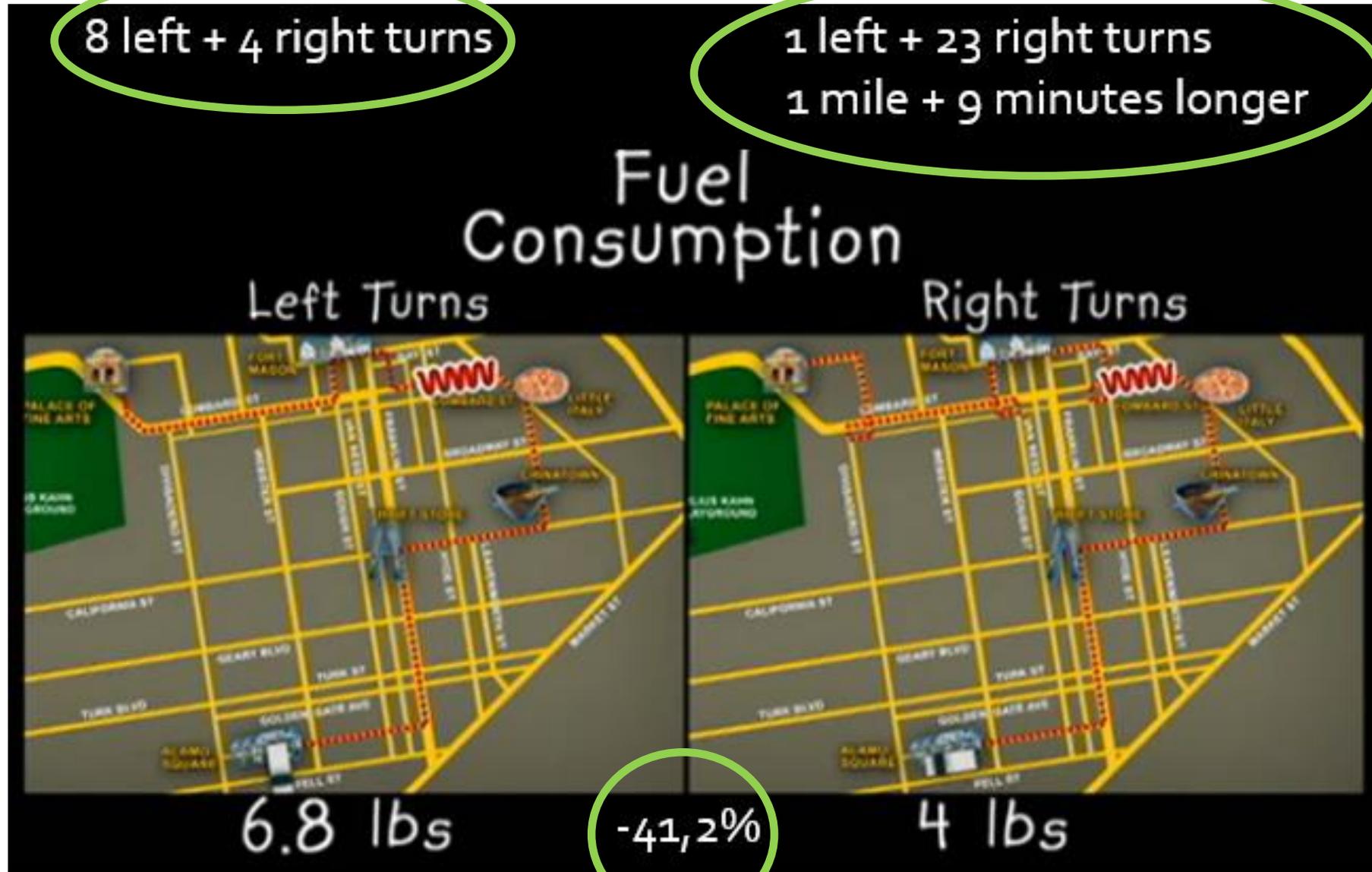
Traditioneller Ansatz (modell-getrieben, top down, question first)

- Analyse von **kleinen Teilmengen** der Informationen/Daten
- von der Frage (Hypothese) zur Antwort \Rightarrow hypothesentestende Methoden
- wiederholbar und strukturiert
- Struktur und Schema für Daten (einschränkendes Modell der Realität)

Big Data Ansatz (daten-getrieben, bottom up, data first)

- Analyse von **allen** Informationen/Daten
- **von den Daten zur Erkenntnis** \Rightarrow **keine Theorie notwendig**
- rechenintensiv, iterativ und explorativ (Histogramm, Scatterplot, Ausreißer, statist. Lagemaße wie arithm. Mittel/Median, bzw. Streumaße wie Varianz/IQR)
- Daten sind die Struktur, Analyse beschreibt und verallgemeinert Muster einer Datengesamtheit

UPS - Do the right thing! - Von den Daten zur Erkenntnis



Source: © auto-news.de

UPS - Do the right thing! - Von den Daten zur Erkenntnis

On-Road Integrated Optimization and Navigation

- UPS wendet Big Data seit über 20 Jahren an
- >700 Mitarbeiter an ORION Entwicklung beteiligt
- 250 Millionen Datenpunkte zur Routenoptimierung
- 1000 Seiten ORION Algorithmus - eine Übung in der Anwendung von **Heuristiken** (Suche nach der besten Antwort durch "**gezielte Ignoranz**")
- "Traveling Salesman Algorithm" (Problem des Handlungsreisenden)
 - **Ersparnis \$300 - \$400 Mio. pro Jahr!**
 - 350.000 **mehr** Pakete/Jahr mit 1.100 **weniger** Trucks
 - 29 Mio. **weniger** Meilen
 - 10 Mio. Gallonen = 38 Mio. Liter **weniger** Treibstoff
 - 22.000 Tonnen **weniger** Kohlendioxid



Source (Grafik): © geek.com, Source (Text): © qz.com

UPS - Do the right thing! - Von den Daten zur Erkenntnis

On-Road Integrated Optimization and Navigation



- ⇒ **Einfache Regel beim Abbiegen: ca. 90% nach rechts**
- ⇒ **Aber hohe Komplexität:**
Anzahl Möglichkeiten für 120 Stopps?



Source (Grafik): © geek.com, Source (Text): © qz.com

Die Kombinatorische Explosion

There are 720 ways to do 6 things: $1 \times 2 \times 3 \times 4 \times 5 \times 6$

Explosion

Live!

1	1
2	2
3	6
4	24
5	120
6	720
7	5.040
8	40.320
9	362.880
10	3.628.800
11	39.916.800
12	479.001.600
13	6.227.020.800
14	87.178.291.200
15	1.307.674.368.000



Source: © pixabay.com

Was macht ein Data Scientist?

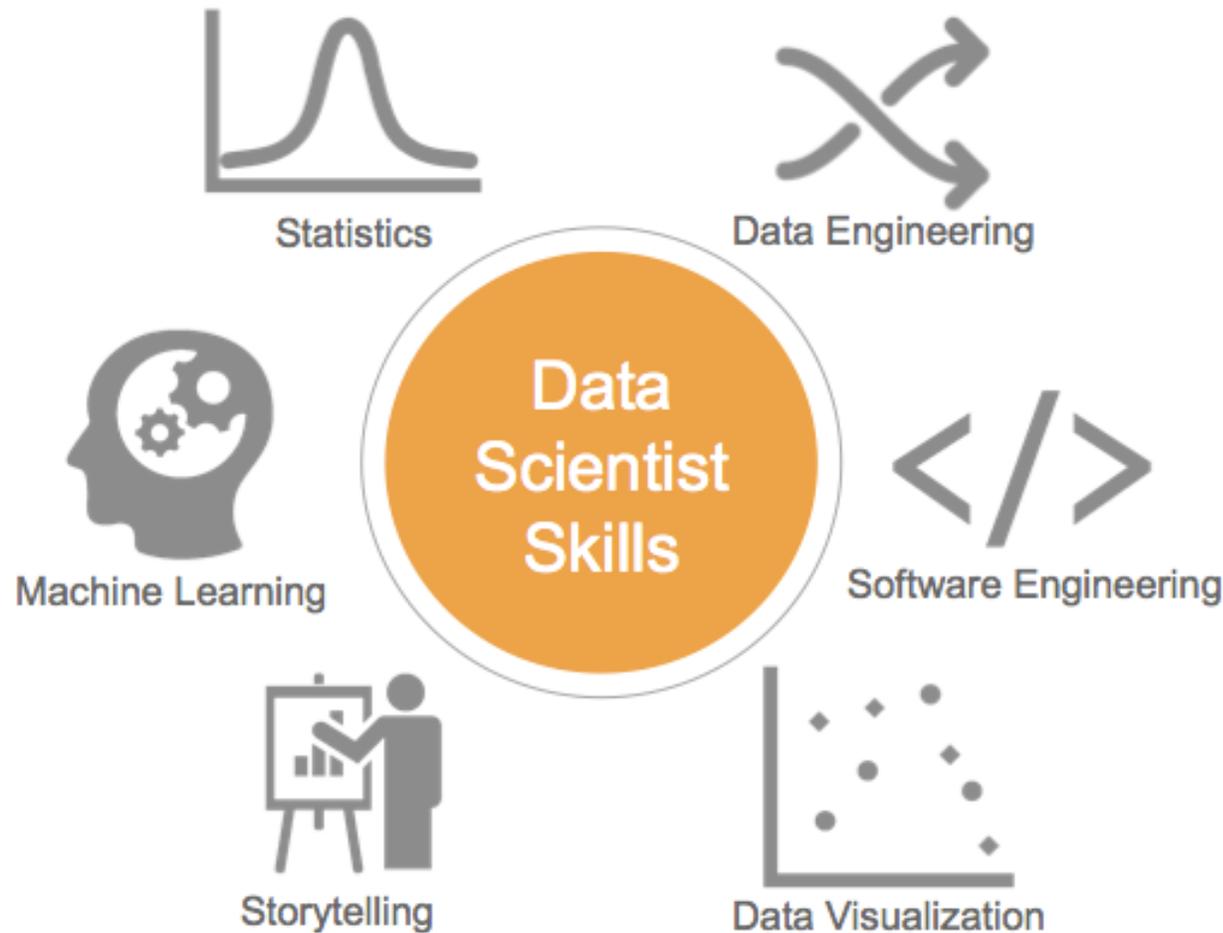
Data → **Information**
→ **Knowledge**



<https://www.google.at/search?&tbm=isch&q=Sexiest+Job>

Source: © 7wdata.be

Was macht ein Data Scientist? - Typen, Talente & Trends



"The educated person is not the person who can *answer the questions*, but the person who can *question the answers*."

Source (Grafik): © vijaybhat.com, Source (Text): © intelligence.de

Was macht ein Data Scientist? - Decoding Data Science

Eine neuer Beruf im Zeitalter von Big Data (was immer das diese Woche heißt...)

- Begleiter des digitalen **Kulturwandels über sämtliche Datenwertschöpfungsketten**
- Blick für das Ganze | Neugier, was in den Daten steckt | Daten sichtbar machen
- Technisches Know-how und **kommunikative Stärke/Sprachvielfalt**
⇔ vs. *Inkompetenzkompensationskompetenz*
- Das *richtige* Problem identifizieren
- Zeitaufwand: 70% Datengewinnung, Transformation, Aufbereitung ⇔ vs. 30% Analysis
- ***The process can be more valuable than a model***
- Data Science ist ein Teamsport:
enge Zusammenarbeit mit Führungskräften (Forschungsrichtung) und Fachabteilungen
- **Datenspezialisten & Data Scientists integrieren!** ⇒ aber nicht zu stark in die Unternehmen und in das Tagesgeschäft einbinden
- Experten empfehlen: **Data Science als separates Kompetenzzentrum aufbauen (nicht etwa in die IT integrieren)** ⇒ gewisses Maß an Unabhängigkeit bewahren
- Unterschied zwischen Statistiker und Data Scientist:
Data Scientist = skeptischer Statistiker, Kreativität ist Top

Was macht ein Data Scientist bei ENGEL?

Aufgaben

- Abteilung **Business Development**
- CSD-Customer Service Division
- Garantie und Kulanz
- Qualitätsmanagement
- Komplexe Auswertungen
- Entwicklung von Tools
- Research:
 - Betreuung von Forschungsprojekten mit mehreren Universitäten und Think Tanks
 - Betreuung Master-Arbeiten (extern und intern)

Comparing Apples to Apples



@cesifoti

Source: © cesifoti, Cesar A. Hidalgo

Was macht ein Data Scientist bei ENGEL?

Aufgaben

Apple I.

First in the world, then in our heads.



Apple II.

First in someone's head, then in the world.



@cesifoti

Source: © cesifoti, Cesar A. Hidalgo

Was macht ein Data Scientist? - Die ENGEL-Datenwelt

ZDF - Zahlen - Daten - Fakten

- ENGEL baut rund 45-50 Maschinen/Woche
- 90.000 Maschinen/Serialnummern
- 170.000 Materialien
- 5.000 - 5.500 Parameter Maschinendaten
(12.000 bei Mehrfarbenmaschinen)
- ca. 2.000 Maschinen online

- Maschinendiagnose
- Symptomkatalog
- Benutzerverhalten/Ergonomie CC 300-Steuerung
- FRED (**F**ailure **R**ate **E**stimating **D**evice):
 - Ausfallswahrscheinlichkeit und Automatisierte Ersatzteil-Pakete

Daten - die Wirbelsäule der Unternehmen 2020

Komplexitätsreduktion? Komplexitätsbewältigung?

- Jeder liebt die Vielfalt - kaum jemand mag die Komplexität
- Wenn wir es uns zu einfach machen, wird alles kompliziert
- "komplex" (qualitativ) vs. "kompliziert" (quantitativ)
- Aus der Komplexität lässt sich (fast) immer etwas machen
- Alle "echten" Big-Data-Szenarien sind komplex
- **Erst die Komplexität schafft Wettbewerbsvorteile**



Source: © osteopathie-wiegleb.de, Eigene Grafik

Daten - die Wirbelsäule der Unternehmen 2020

Komplexitätsreduktion? Komplexitätsbewältigung?



- **360° Kundenblick:** gelingt nur, wenn im "richtigen Leben" auch
- Medienbrüche abbauen: gelingt nur, wenn wir auch eigene mentale Medienbrüche erkennen und abbauen
- Methodenarsenal: technisch + digital + mental
- **"Endlich beginnen die Schwierigkeiten"** (Niccolò Machiavelli, 1469-1527)
- **Fehlerkultur:** **Kultur des Ausprobierens ("THANK YOU, ERRORS!")**
 Fehler sind eine wichtige Informationsquelle!

Source: © osteopathie-wiegleb.de, Eigene Grafik

Datenqualität - die Bandscheiben der Unternehmen 2020

Datenqualitätsdimensionen für Stammdaten



Source (Text): © "Entwicklung einer Methode zur Stammdatenintegration", Alexander Schmidt, Logos Verlag Berlin GmbH, 2010
Source (Foto): © panthermedia.net / Le Moal Olivier

Datenqualität - die Bandscheiben der Unternehmen 2020

Datenqualitätsdimensionen für Stammdaten

Datenqualitätsdimension	Definition
Genauigkeit (Accuracy, Correctness)	Grad, in dem Daten korrekt, glaubwürdig und frei von Fehlern sind bzw. in dem sie den korrekten Zustand eines Realweltobjektes widerspiegeln
Vollständigkeit (Completeness)	Grad, in dem Daten über die ausreichende Breite, Tiefe und Umfang zur Erfüllung einer Aufgabe verfügen bzw. in dem notwendige Werte oder Attribute eines Datensatzes nicht fehlen
Aktualität (Timeliness)	Grad, in welchem das Alter der Daten für deren Nutzung angemessen ist
Konsistenz (Consistency)	Grad, in dem die Daten mit ihrer Definition übereinstimmen
Eindeutigkeit (Uniqueness)	Grad, in dem es für jedes Realweltobjekt nur eine einzige, eindeutige Repräsentation in den Applikationen gibt
Zugänglichkeit (Accessibility)	Grad, in dem Datennutzer Zugriff auf von ihnen benötigte Daten haben

Source (Text): © "Entwicklung einer Methode zur Stammdatenintegration", Alexander Schmidt, Logos Verlag Berlin GmbH, 2010

Source (Foto): © panthermedia.net / Le Moal Olivier

Big Data, Little Data, No Data in der Industrie

No Data:

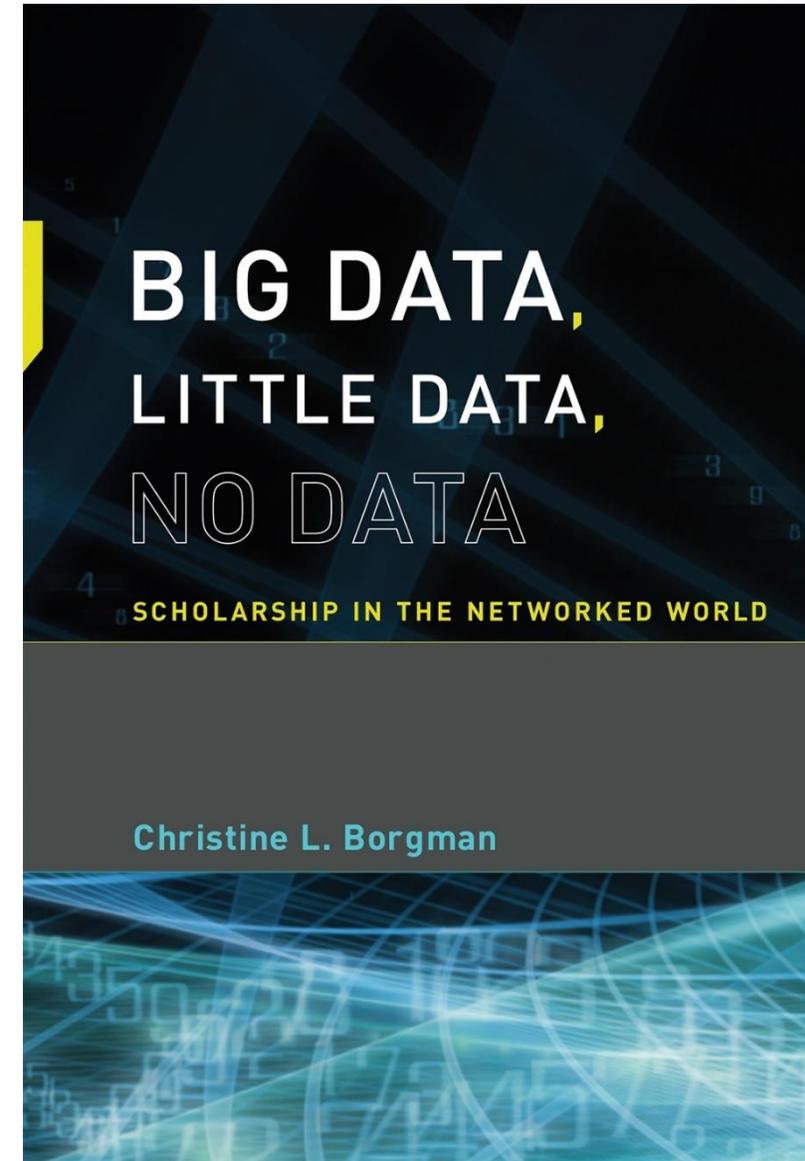
- Prozessketten vervollständigen oder umsetzen (Disziplin bei einem "Pflichtfeld")
- Vor 200x gibt es keine Daten, da der Prozess per 01.04.200y eingeführt wurde.
- Unerreichbarkeit der eigenen Daten

Little Data:

- Zu wenige Daten, um statistische Aussagekraft zu haben
- Datenschutz und Datensicherheit

Big Data:

- Nicht "viele" sondern vielfältige Daten
- Information wird immer stärker zur **strategischen Ressource**, deren sinnvolle Erschließung und Nutzung wettbewerbskritisch ist.
- **Gartner: Unternehmenswert richtet sich bald nach dem Datenreichtum**
- **Verknüpfungskreativität:** verschiedene Abteilungen und quer durch alle Disziplinen (Bsp.: Servicetickets/Maschine, Temperatur in Brasilien)
- **Viele neue Daten** ⇒ aus vielen neuen Quellen ⇒ viele neue Sichten



Source: © mitpress.mit.edu

Effizient | Plastifiziertertechnologie von ENGEL

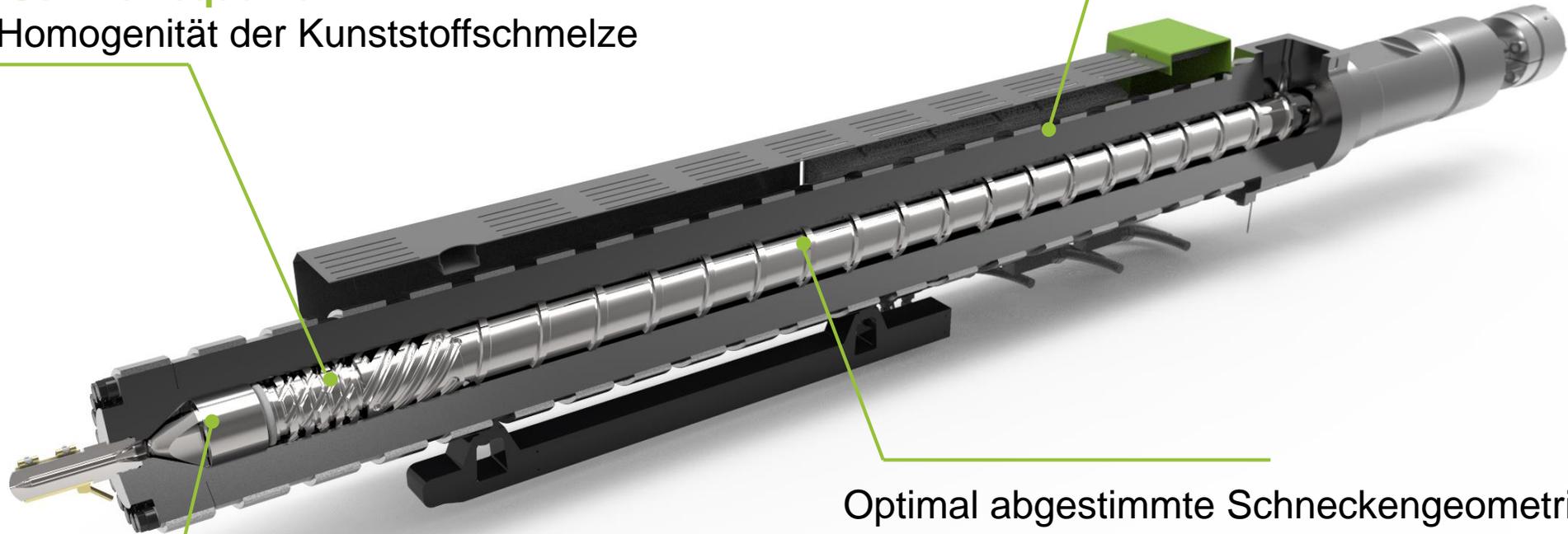
Qualität und Leistungsfähigkeit beim **Dosiervorgang**

Hohe Lebensdauer und Wirtschaftlichkeit

Sehr gute Beständigkeit gegenüber Verschleiß und Korrosion

Allerbeste Schmelzequalität

Konstante Homogenität der Kunststoffschmelze



Optimal abgestimmte Schneckengeometrie

Breites Einsatzspektrum bei höchsten Leistungen

Stabiles Förderverhalten

Präzise Wiederholgenauigkeit

e-connect.monitor

Schneckenüberwachung **ohne Maschinenstillstand**

Mit Ultraschall in den Massezylinder blicken

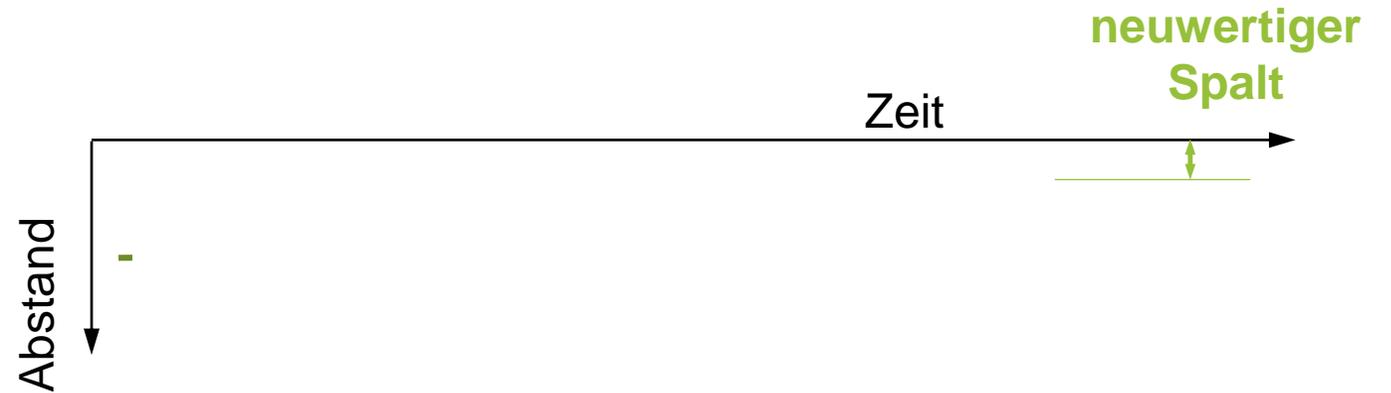
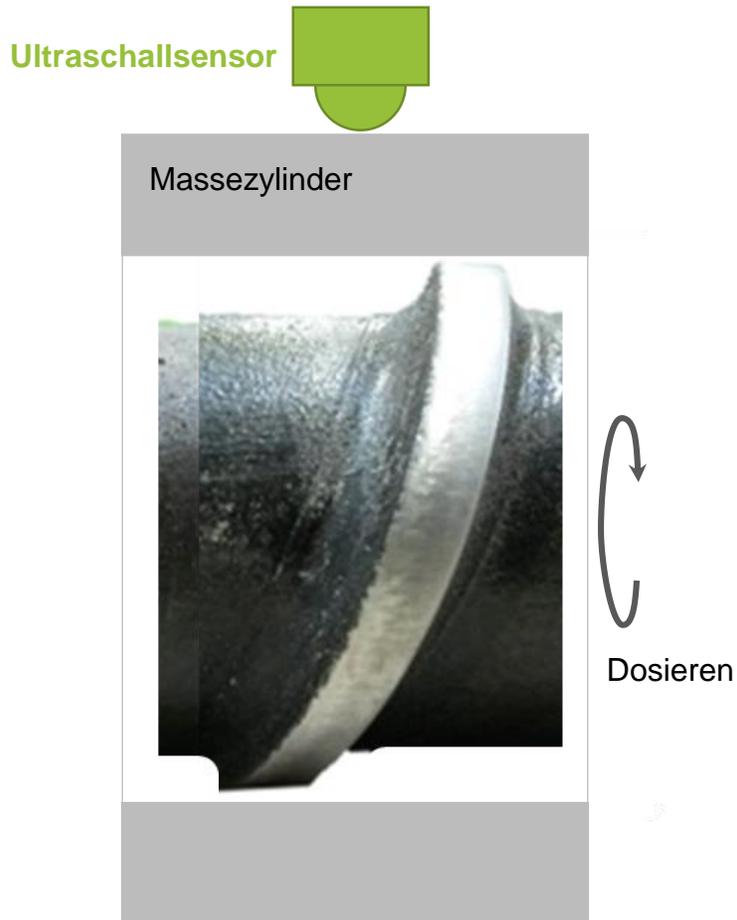
- Ermittlung des **Verschleißes des Schneckenstegs**
- Messung der **Ultraschall Reflexionen** im Massezylinder
- Messung **während des Betriebs** möglich
- **Keine Demontage** der Schnecke und des Flanschs notwendig (**die Wartezeit zieht sich in ihr Schneckenhaus zurück...**)
- **Aufzeichnung** der Messdaten durch einen **ENGEL Service Techniker**
- **Automatisierte Auswertung** der Messdaten **bei ENGEL**



e-connect.monitor

Schneckenüberwachung **ohne Maschinenstillstand**

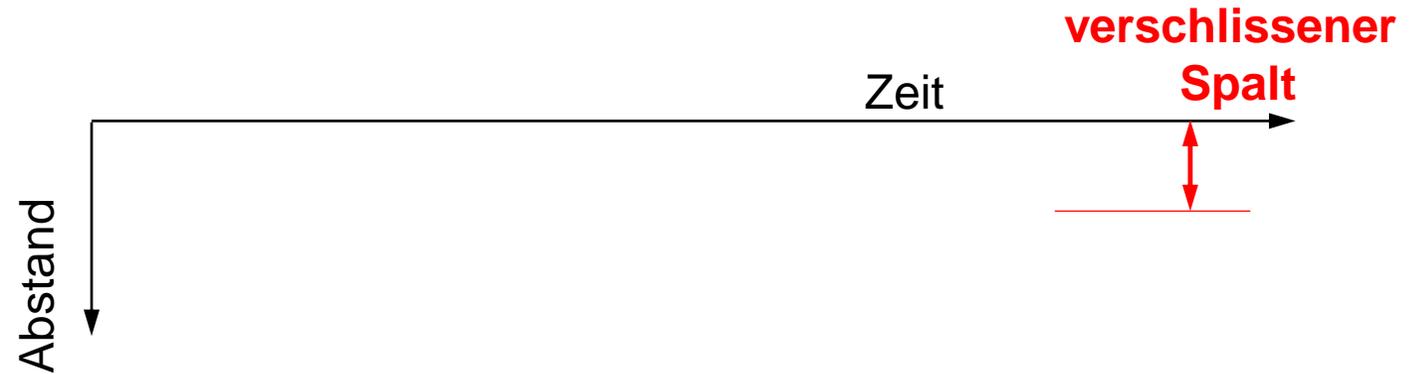
Wie funktioniert die Messung?



e-connect.monitor

Schneckenüberwachung **ohne Maschinenstillstand**

Wie funktioniert die Messung?



Visualisierung OSZI Daten - vitale vs. sterbende Pumpe

Programmiersprache: Python 3.5.2 | Anaconda 4.2.0 (64-bit)

Variable 1: ungefilterter Pumpendruck-Istwert | Variable 2: LogN (Pumpendruck)

Python Live!

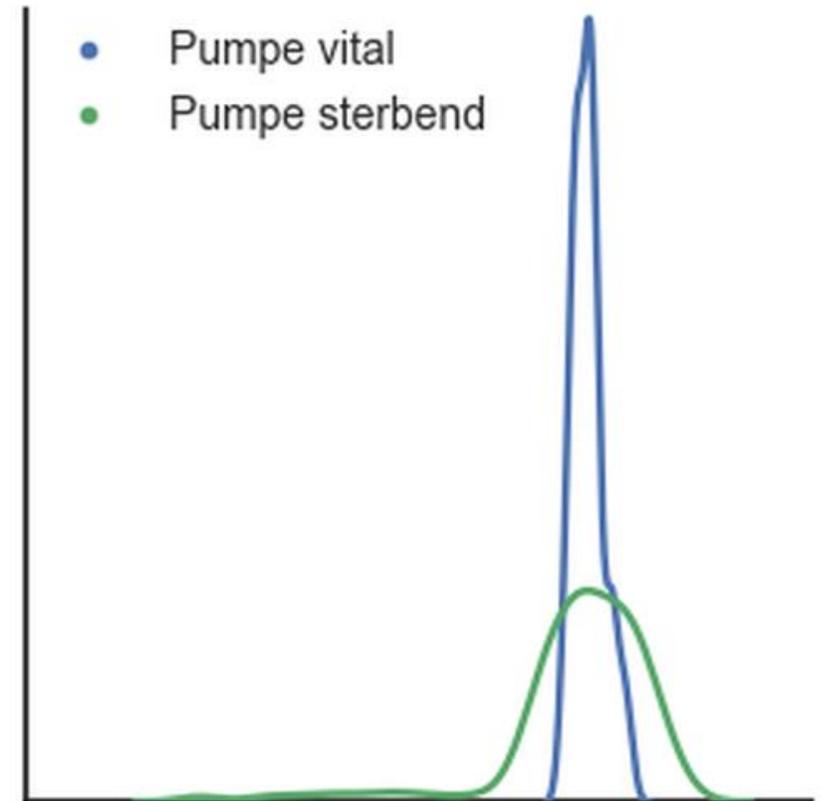
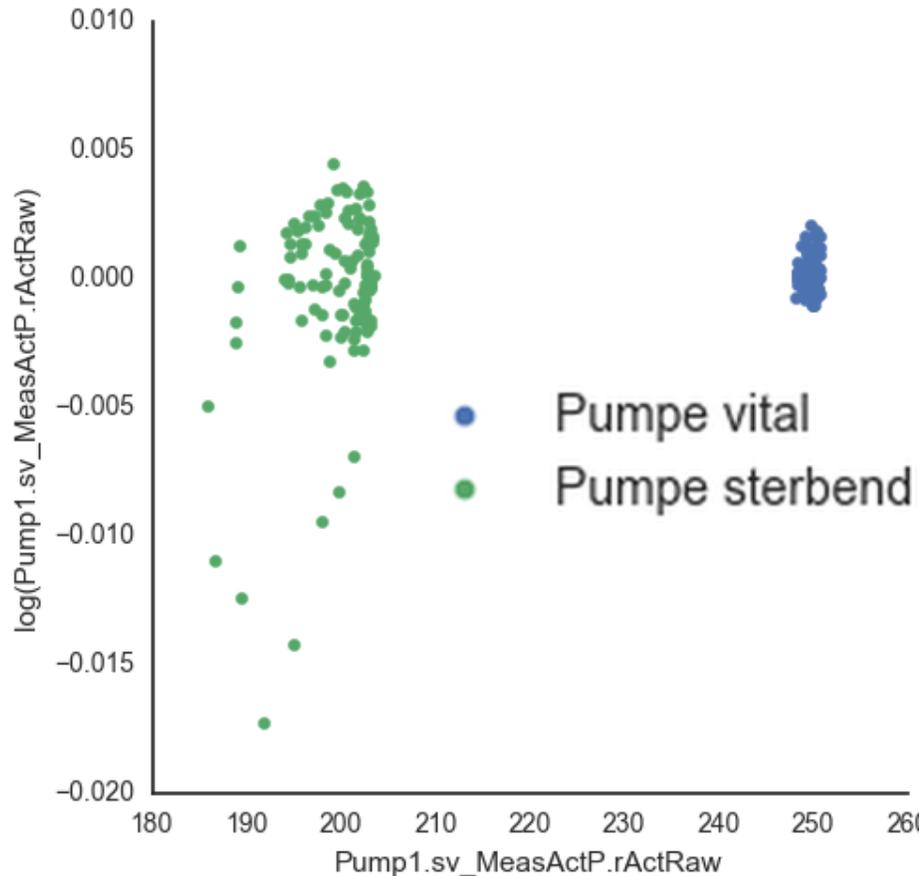


Source: © pixabay.com

Visualisierung OSZI Daten - vitale vs. sterbende Pumpe

Programmiersprache: Python 3.5.2 | Anaconda 4.2.0 (64-bit)

Variable 1: ungefilterter Pumpendruck-Istwert | Variable 2: LogN (Pumpendruck)



Visualisierung OSZI Daten - vitale vs. sterbende Pumpe

Programmiersprache: Python 3.5.2 | Anaconda 4.2.0 (64-bit)

Variable 1: ungefilterter Pumpendruck-Istwert | Variable 2: LogN (Pumpendruck)

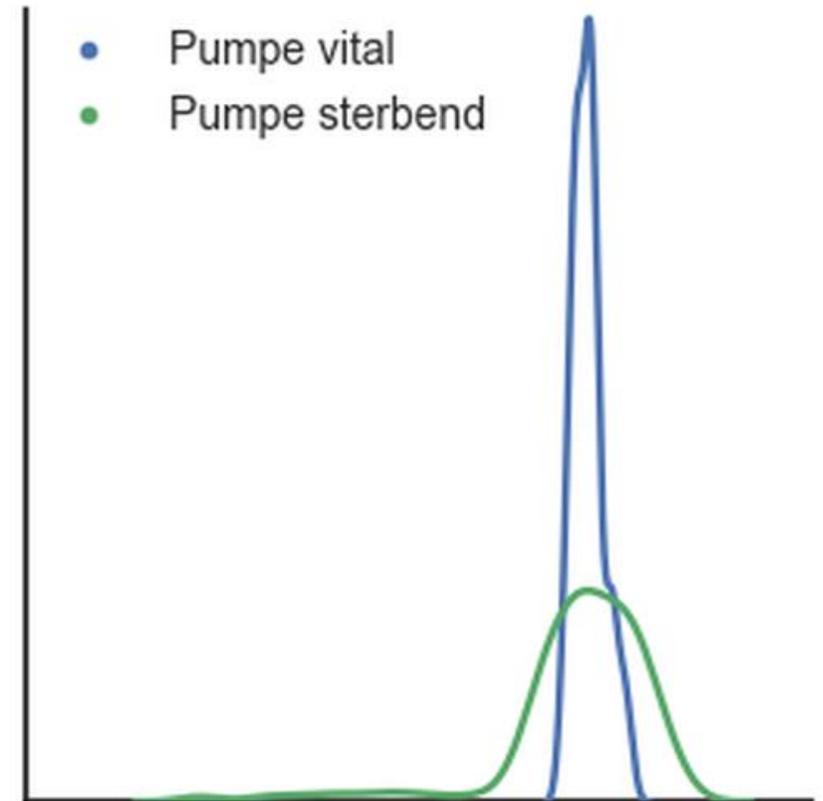
Kurtosis

Klassische Interpretation:

- Wölbung = Maßzahl für die Steilheit bzw. "Spitzigkeit" einer (eingipfeligen) Wahrscheinlichkeitsfunktion, statistischen Dichtefunktion oder Häufigkeitsverteilung.

Moderne Interpretation:

- sagt mehr über "tail extremity/tail risk/fat tails" (Extremwertverteilung) aus
- das **Gesamtgewicht der tails relativ zum Rest der Verteilung**



5 Forderungen an Politik, moderne Unternehmen & CEOs

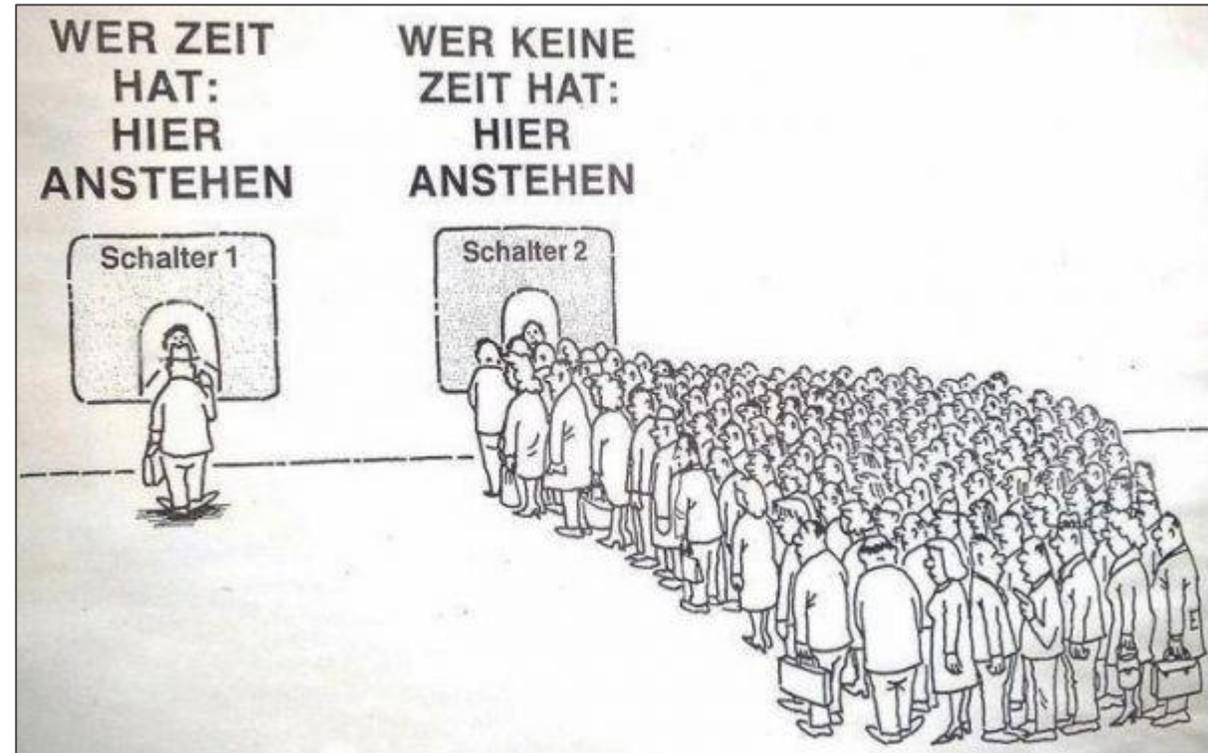
Schritte zum Digitalen Klimawandel

- Förderung **aller Mitarbeiter** in der **digitalen Kompetenz** (digital literacy)
- Förderung der **Datenqualität** in der gesamten Datenwertschöpfungskette
- **Abteilungsübergreifendes Denken fördern** und Wissensilos brechen
(Informationsmonopole sind oft sehr teure Machtspiele)
- **Geburtshilfe** beim Digitalen Zwilling leisten
- Einführung eines **Daten-Ministers**

Zuletzt ein Tipp aus der Statistik:

Die tägliche Entscheidung im Supermarkt

- Die meisten Kunden sind Rechtshänder und tendieren unterbewusst nach rechts.
- Je weiter rechts eine Schlange liegt, desto länger dauert es im Mittel.
- Wählen Sie eher die **linke** Schlange!

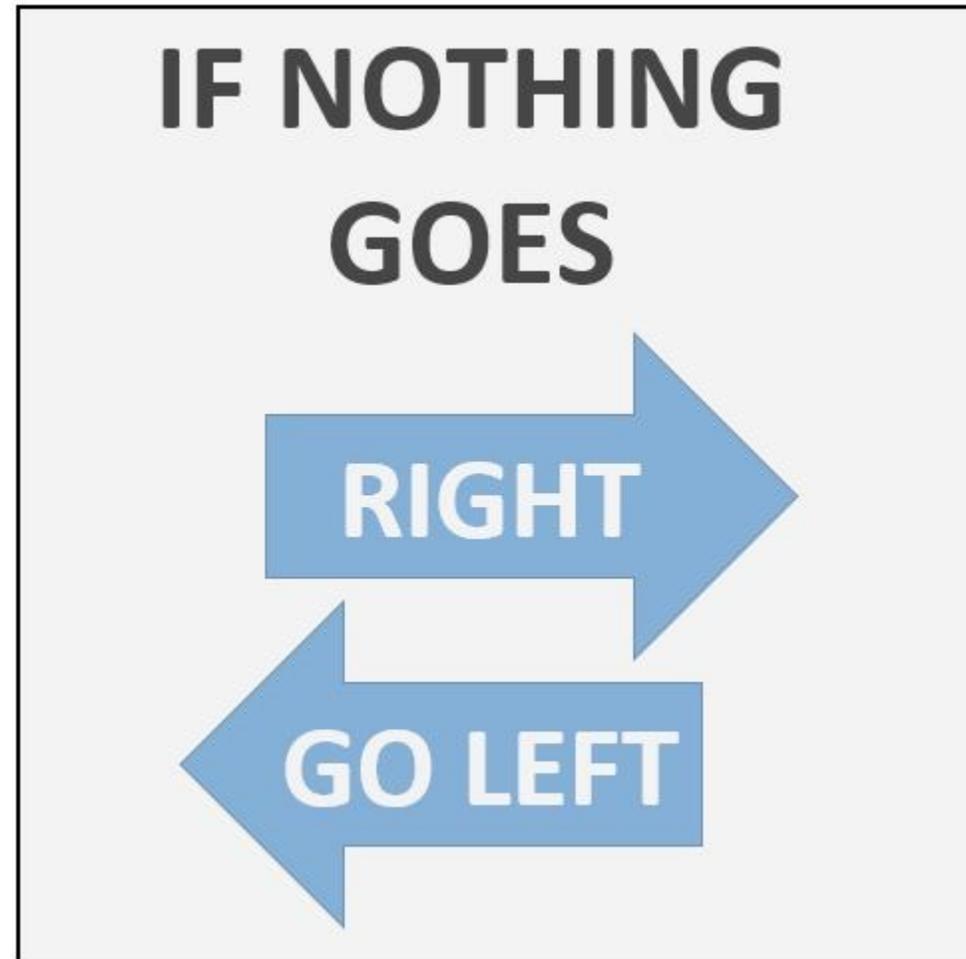


Big Data and Decision Making

Data is easy.

Deciding is hard.

Big Data and Decision Making



Source: © Eigene Grafik

ENGEL

www.engelglobal.com

IT-Summit 2017

Thomas Pflügl - Data Scientist



ENGEL Austria GmbH

Source: © shutterstock.com

IT-Summit 2017

Thomas Pflügl - Data Scientist



Danke!

Kontakt Daten: thomas.pfluegl@engel.at



ENGEL Austria GmbH