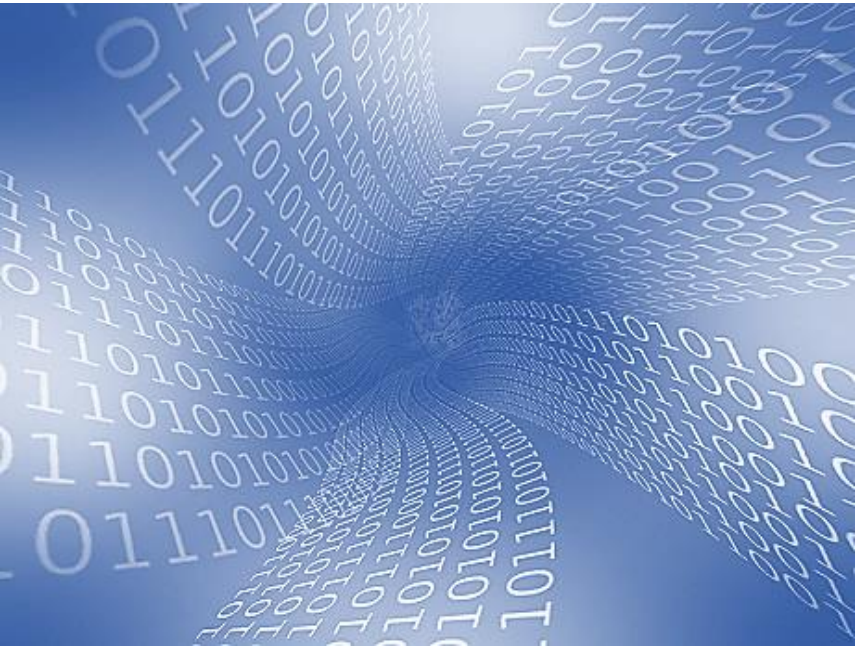
Prof. Dr. Bastian Amberg
School of Business & Economics
Department of Information Systems

Freie Universität Berlin



Business Intelligence

01 (Meta-)Introduction

Prof. Dr. Bastian Amberg
(summer term 2024)
17.4.2024

Who am I?

Department
Wirtschaftsinformatik

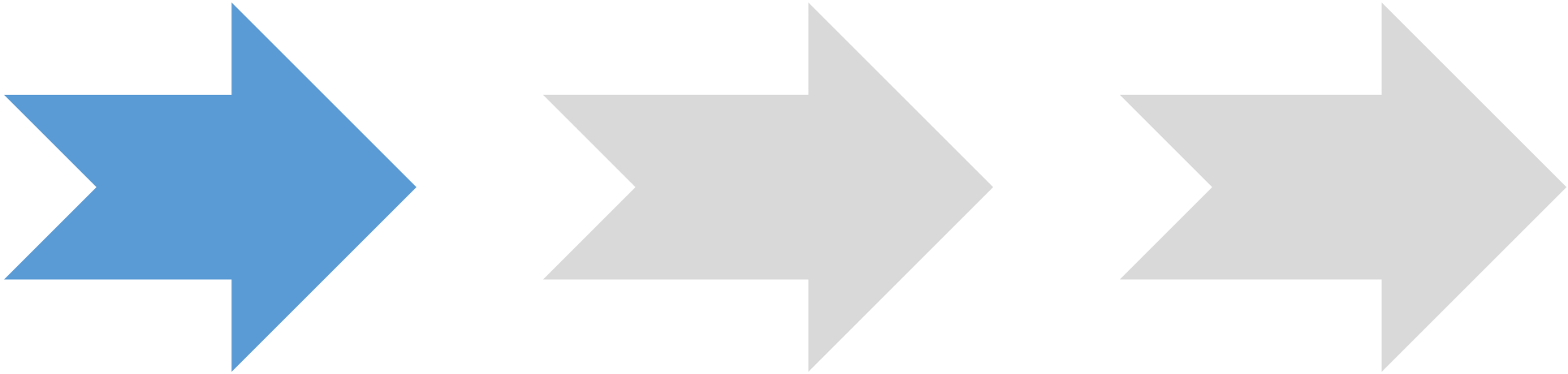
Freie Universität Berlin

Professur für Wirtschaftsinformatik	Professur für BWL, Information und Organisation
Prof. Dr. Natalia Kliewer Operations Research & Analytics Planungssysteme in Transport und Verkehr Robuste Effizienz Revenue Management	**Prof. Dr. Martin Gersch** E-Business Informationsmanagement Service Engineering Entrepreneurship Education

Juniorprofessur für Advanced Decision Analytics	Assoziierte ECDF-Professur	Assoziierte Professur	Assoziierte Professur	Juniorprofessur für BWL, insb. Digital Entrepreneurship und Diversity
Prof. Dr. Bastian Amberg Entscheidungs-unterstützungs-systeme Robuste Effizienz In Dienstleistungs-industrien	**Prof. Dr. Daniel Fürstenau** Demnächst wieder Digitale Infrastrukturen IT-Governance & IT-basierte Risiken Digital Transitions in Healthcare	**Prof. Dr. Christian Meske** Digital Workplace Collaboration Technologies Management of DT Processes	**Prof. Dr. Hannes Rothe** Digital Entrepreneurship Digital Infrastructures & Ecosystems Organizing Data & Knowledge	**Prof. Dr. Janina Sundermeier** Digital Entrepreneur-ship Unternehmerische Diversität Gründungsbezogene Persönlichkeits-merkmale

Nähere Informationen unter: http://www.wiwiss.fu-berlin.de/fachbereich/bwl/pwo/

Agenda



(1)
Metaintroduction

(2)
Decision Support and Business Intelligence

(3)
Data-analytic thinking

Ref.

Herzlich Willkommen…

…zur Pflichtvorlesung im 2. Semester des Studiengangs **Master-Wirtschaftsinformatik**

Interaktive Vorlesung/Seminar mit Übungseinheiten/Miniprojekten am Rechner

Seminaristischer Unterricht mit **regelmäßiger** und **aktiver Teilnahme**: auch Sie produzieren Inhalte

Unterrichtssprache: Deutsch;
Materialien in der Regel auf Englisch

Modul: Business Intelligence

Hochschule/Fachbereich/Institut: Freie Universität Berlin/FB Wirtschaftswissenschaft/Institut für Wirtschaftsinformatik

Modulverantwortliche/r: Dozentinnen und Dozenten des Moduls

Zugangsvoraussetzungen: Keine

Qualifikationsziele:

Die Studentinnen und Studenten können mithilfe von Methoden der intelligenten Datenanalyse Erkenntnisse aus der Analyse großer und komplexer Datenmengen gewinnen. Sie besitzen die Fähigkeit, Simulationssysteme für die Entscheidungsunterstützung insbesondere unter Unsicherheit im betriebswirtschaftlichen Umfeld zu entwerfen, zu implementieren und einzusetzen. Sie sind in der Lage, die vermittelten Modelle, Methoden und Algorithmen in der den Fragestellungen angemessenen Weise auszuwählen und anzuwenden sowie die Handlungsempfehlungen aus der Methodenanwendung abzuleiten. Das Seminar am PC adressiert insbesondere auch überfachliche Qualifikationsziele, insb. eigenständiges Arbeiten, analytisches Denken, Präsentationsfähigkeiten sowie Fähigkeiten zur technikassistierten Aufgabenlösung in Teams.

Inhalte:

Spezielle Modelle und Algorithmen des Datamining, Modelle, Methoden und Grundlagen der Simulation sowie Nutzung einschlägiger Softwarewerkzeuge zur Datenanalyse und Simulation, eine Auswahl aus speziellen Techniken, wie z. B. Clustering, Assoziationsanalyse, Klassifikation, diskrete und ereignisgesteuerte, stochastische, agentenbasierte Simulation etc.

Lehr- und Lernformen	**Präsenzstudium** (Semesterwochenstunden = SWS)	**Formen aktiver Teilnahme**	**Arbeitsaufwand** (Stunden)	
Seminarsitischer Unterricht	1	Unterrichtsgespräch, Beantwortung von Diskussionsfragen, Diskussion von Anwendungsproblemen	Präsenzzeit Seminaristischer Unterricht	15
			Vor- und Nachbereitung Seminaristischer Unterricht	30
Seminar am PC	2	Kurzvorträge mit Diskussion, Diskussion von Literatur und Anwendungsbeispielen	Präsenzzeit Seminar am PC	30
			Vor- und Nachbereitung Seminar am PC	30
			Prüfungsvorbereitung und Prüfung	75
Veranstaltungssprache:		Deutsch		
Pflicht zur regelmäßigen Teilnahme:		Ja		
Arbeitszeitaufwand insgesamt:		180 Stunden		6 LP

Ref.

Rahmenbedingungen

Umfang der Veranstaltung: 6 LP/ **3 SWS**

Mittwoch, 10-12 Uhr
Freitag, ~~12-14 Uhr~~ 14-16 Uhr

Dies beinhaltet:
→ interaktive Vorlesungen
→ digitale Übungen, Zeit für Vor- und Nachbereitung der Vorlesungen

Terminplan und Ankündigungen im **Blackboard** beachten!

Kommunikation und Materialienbereitstellung über Blackboard
Kurs Business Intelligence (SoSe 2024) (WIWISS_S_10180206_24S)

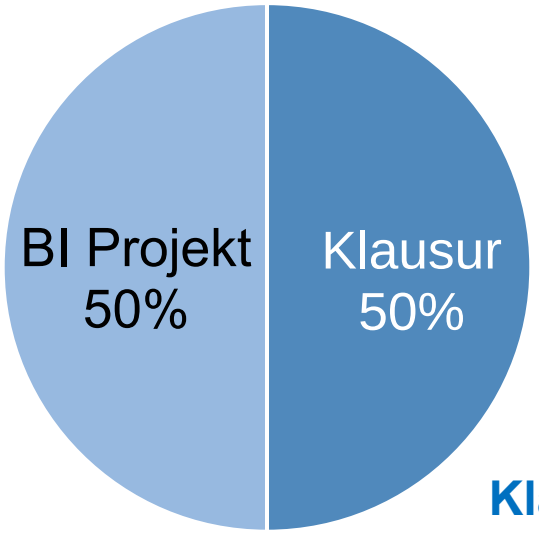
Unklarheiten direkt im Kontext, am Ende der Veranstaltung oder zu Beginn der nächsten Veranstaltung klären, ansonsten: **bastian.amberg@fu-berlin.de**
Oder Blackboard-Forum „Organisatorische Fragen“ / „Inhaltliche Fragen“

Prüfungsleistung besteht aus **Klausur** und **Mini-Projekt**

Ref.

Prüfungsleistung

Zusammensetzung



BI-Projekt

macht 50% der Note aus.

Bearbeitungsdauer ca. 6 Wochen, semesterbegleitend

Sie finden sich in 4er-Gruppen zusammen und bearbeiten einen Datensatz auf relevante Fragestellungen.

Sie wählen sich die dafür geeigneten Methoden und Werkzeuge selbstständig aus.

Ca. 20-minütige Präsentation der Fragestellung, des Vorgehens und Diskussion der Ergebnisse und Erstellen einen kurzen Projektberichts. (genaue Vorgaben hierzu später)

Ref.

Klausur

macht 50% der Note aus.

Dauer 60 Minuten, im Klausurenzeitraum

(1.Termin im Zeitraum 22.7. bis 3.8.)

(2.Termin im Zeitraum 23.9. bis

In der Veranstaltung lernen Sie **Methoden und Werkzeuge** der Business Intelligence kennen.

Einzelleistung

Verständnisfragen und Anwendung

(Beispielaufgaben im späteren Veranstaltungsverlauf)

Literatur

Provost, F.; Fawcett, T.: Data Science for Business. Fundamental Principles of Data Mining and Data-Analytic Thinking. O'Reilly, 2013

Berthold, M. R.; Borgelt, C.; Höppner, F.; Klawonn, F.: Guide to Intelligent Data Analysis. Springer, 2011

Müller, R. M.; Lenz, H.-J.: Business Intelligence. Springer, 2014

Lusti, M.: Data Warehousing und Data Mining, Springer, 2002

Additional readings: **see Bibliography** at the end of each presentation

Ref.

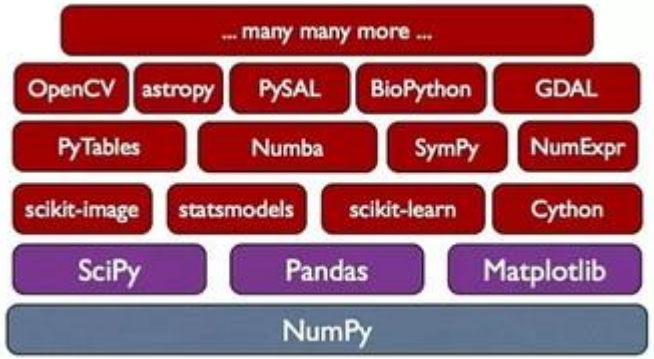
Software

im Verlauf des Kurses

Python Packages

- Let's agree on commonly using Python version ≥ 3.7 https://www.python.org/downloads/ (latest version is 3.12.3)
- Install Anaconda (https://www.anaconda.com/distribution/) (or the according packages sciPy, Pandas, NumPy, scikit-learn, Matplotlib in your own environment)

Deep Learning:
Tensorflow (https://www.tensorflow.org/) with conda (install)



Konstanz Information Miner (**KNIME**)*

- http://www.knime.org/

Weka Data Mining Software*

- http://www.cs.waikato.ac.nz/ml/weka/

*For individual needs

Ref.

Overall goals of this class

Freie Universität Berlin

We have three goals. After this course:

- You know how to solve business problems by data-analytic thinking
- You have an overview about principles of how to model and solve upcoming business problems.
- You know several tools and ways of how to practically implement solution methods

Ref.

Main focus areas

Freie Universität Berlin

Data Warehousing / Data Engineering

- How to **store and access** huge amounts of data?

Data Mining / Data Science

- How to **derive knowledge and profitable business action** out of large databases?

Ref.

Schedule

Freie Universität Berlin

		Wed., 10:00-12:00			Fr., 14:00-16:00 (Start at 14:30)	Self-study	
Basics	W1	17.4.	(Meta-)Introduction	19.4.		Python-Basics	Chap. 1
	W2	24.4.	Data Warehouse – Overview & OLAP	26.4.	*[Blockveranstaltung SE Prof. Gersch]*		Chap. 2
	W3	1.5.		3.5.	Data Warehouse Modeling I		Chap. 3
	W4	8.5.	Data Warehouse Modeling II	10.5.	Data Mining Introduction		
Main Part	W5	15.5.	CRISP-DM, Project understanding	17.5.	Python-Basics-Online Exercise	Python-Analytics	Chap. 1
	W6	22.5.	Data Understanding, Data Visualization	24.5.	*No lectures, but bonus tasks* *1.) Co-Create your exam* *2.) Earn bonus points for the exam*		Chap. 2
	W7	29.5.	Data Preparation	31.5.			
	W8	5.6.	Predictive Modeling I	7.6.	Predictive Modeling II (10:00 -12:00)	BI-Project	Start
	W9	12.6.	Fitting a Model I	14.6.	Python-Analytics-Online Exercise		|
	W10	19.6.	*Guest Lecture*	21.6.	Fitting a Model II		|
	W11	26.6.	How to avoid overfitting	28.6.	What is a good Model?		|
Deep-ening	W12	3.7.	Project status update Evidence and Probabilities	5.7.	Similarity (and Clusters) From Machine to Deep Learning I		|
	W13	10.7.		12.7.	From Machine to Deep Learning II		|
	W14	17.7.	Project presentation	19.7.	Project presentation		End
Ref.					*Klausur 1.Termin ~ 22.7. bis 3.8.* *Klausur 2.Termin ~ 23.9. bis 5.10.*	Projektbericht	

Case Study

Informations on our digital Python exercises

Freie Universität Berlin

A) Exercises for self-study - based on Jupyter notebooks:

Notebooks can be found in Blackboard:
"Kursmaterialien > Readings & Übungen > Python-Übungen > Jupyter Notebooks"

Chapters are unlocked at

Python-Basics:

17.04.: Chapter 1 – Erste Schritte	*solutions: 24.04.*
24.04.: Chapter 2 – Strings & String-Funktionen	*solutions: 02.05.*
02.05.: Chapter 3 – Bedingungen	*solutions: 08.05.*

Python-Analytics:

15.05.: Chapter 1 – Spezielle Datentypen	*solutions: 22.05.*
22.05.: Chapter 2 – Datenanalyse	*solutions: 29.05.*

B) Exercises via Webex:

Python-Basics	~ 17.5.
Python-Analytics	~ 14.6.

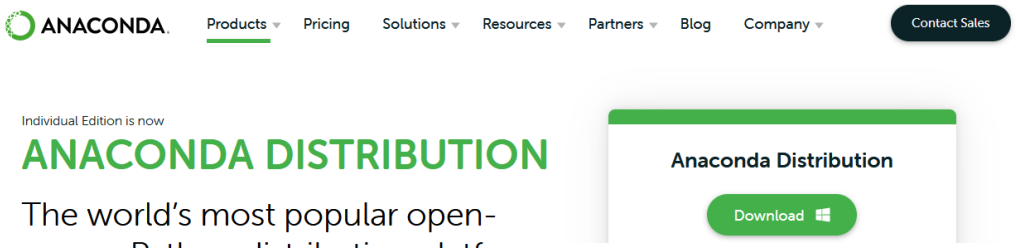
The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

jupyter
https://jupyter.org/

Ref. **Questions during/after exercises:** bastian.amberg@fu-berlin.de

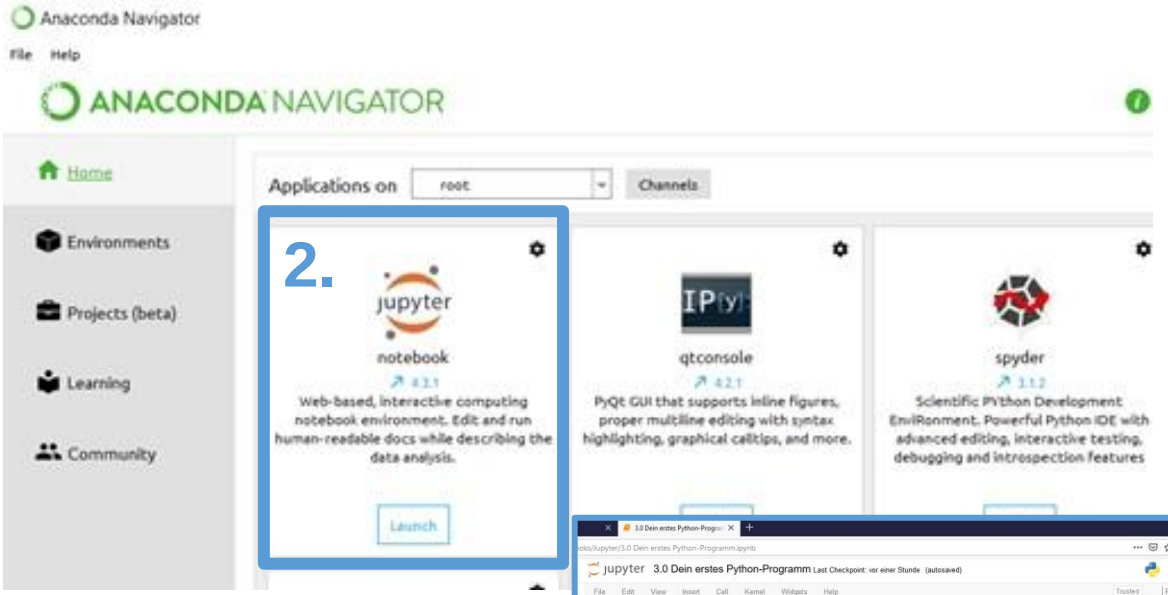
Getting Started with Jupyter-notebooks

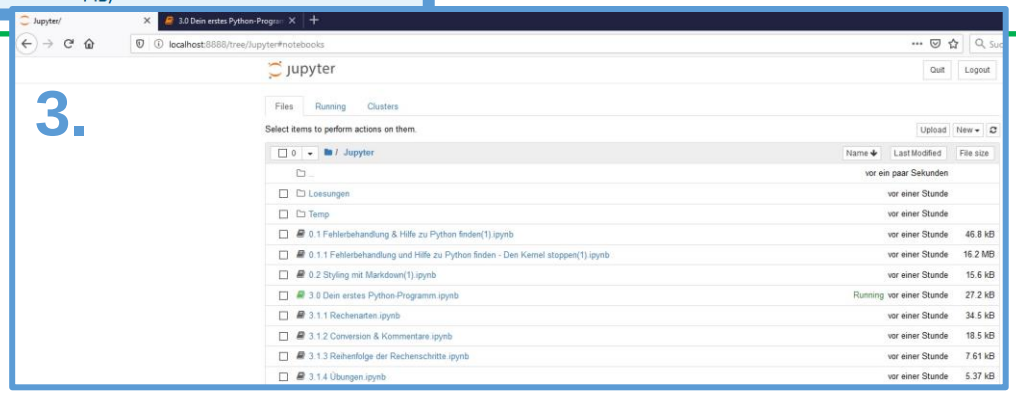
Freie Universität Berlin

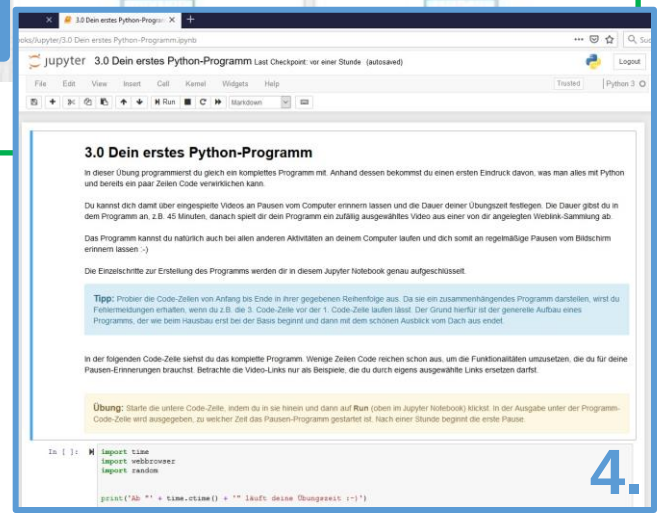


https://www.anaconda.com/distribution/

1.







Alternative zum schnellen Testen ohne Installation: Sind die Notebooks einmal lokal gespeichert, können sie z.B. über *CoCalc (Collaborative Calculation and Data Science Service)* geöffnet, kopiert, umbenannt, ausgeführt und bearbeitet werden. Es ist – Stand April 2024 – allerdings die Erstellung eines (kostenlosen) Accounts notwendig. Link zu CoCalc: https://cocalc.com/ bzw. https://cocalc.com/features/jupyter-notebook . Nach Login kann ein Notebook oder mehrere hochgeladen werden (unter „File > Upload… > Upload" oder unter „File > Open… > Upload) und es kann zum Editieren kopiert, umbenannt und bearbeitet werden. Das bearbeitete Notebook kann gespeichert und anschließend heruntergeladen werden, um es lokal zu sichern.

What are your expectations?

?

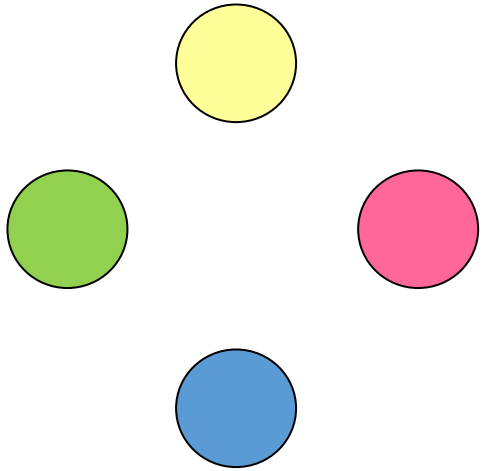
?

Kahoot
www.kahoot.it
(über Smartphone oder Laptop)
PIN folgt

What is your background?

!

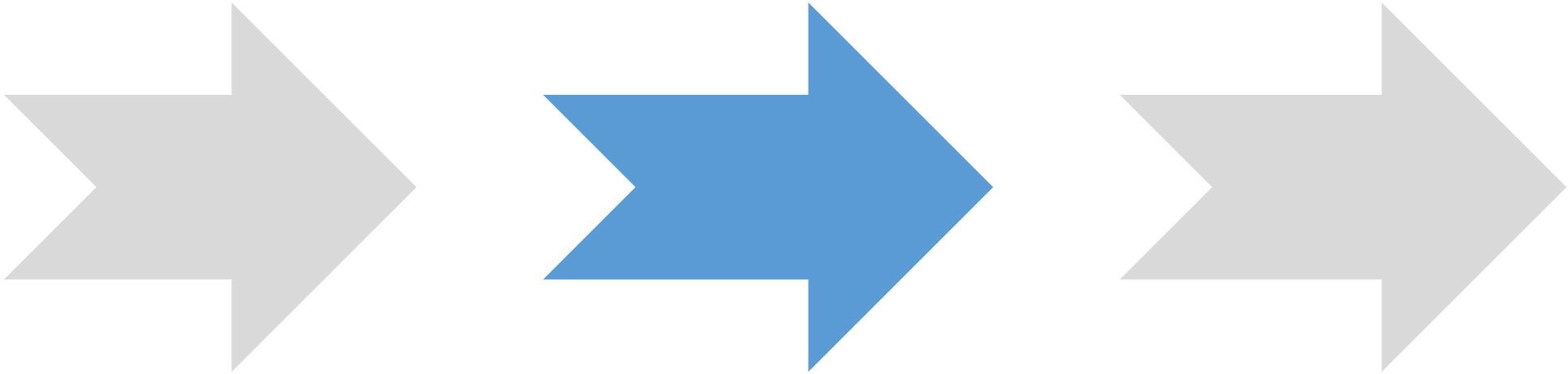
!



Ref.

Agenda

Let’s put BI into perspective



(1)
Metaintroduction

(2)
Decision Support and Business Intelligence

(3)
Data-analytic thinking

Ref.

Decision Support Systems

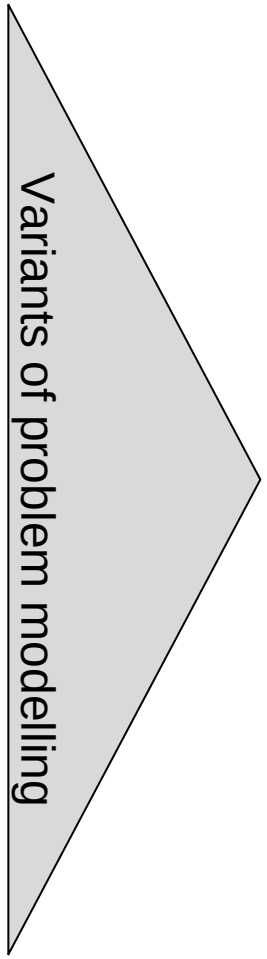
And variants of problem modelling

Decision Support Systems in the broadest sense can be defined as

Computer technology solutions that can be used to support complex decision making and problem solving.
(Shim et al. 2002)

Broad definition that encompasses many areas

- Application systems
- Mathematical modeling
- Data driven modeling



Common **mathematical modelling**

- All the relevant variables and relations can be identified
- The nature of the problem can adequately be caught by mathematical models
- Optimal solution for the underlying decision problem can be derived from the model (e.g., linear programming)

Data driven (empirical) **modelling**

- Problem too complex to identify all the relevant variables and relationships
- Gain insight into problem structure by analysis of historical (transactional) data (e.g., data mining)
- Entails trial-and-error experiments and oftentimes black-boxing

Ref. Power (2007)

Business Intelligence: Definition

There is no unique or mathematical definition of Business Intelligence.

The Data Warehousing Institute defines Business Intelligence as…

… the process, technologies and tools needed to ***turn data into information, information into knowledge and knowledge into plans*** *that drive profitable business action. Business intelligence encompasses data warehousing, business analytics tools, and content/knowledge management.*

(http://www.tdwi.org/)

Another (similar) definition:

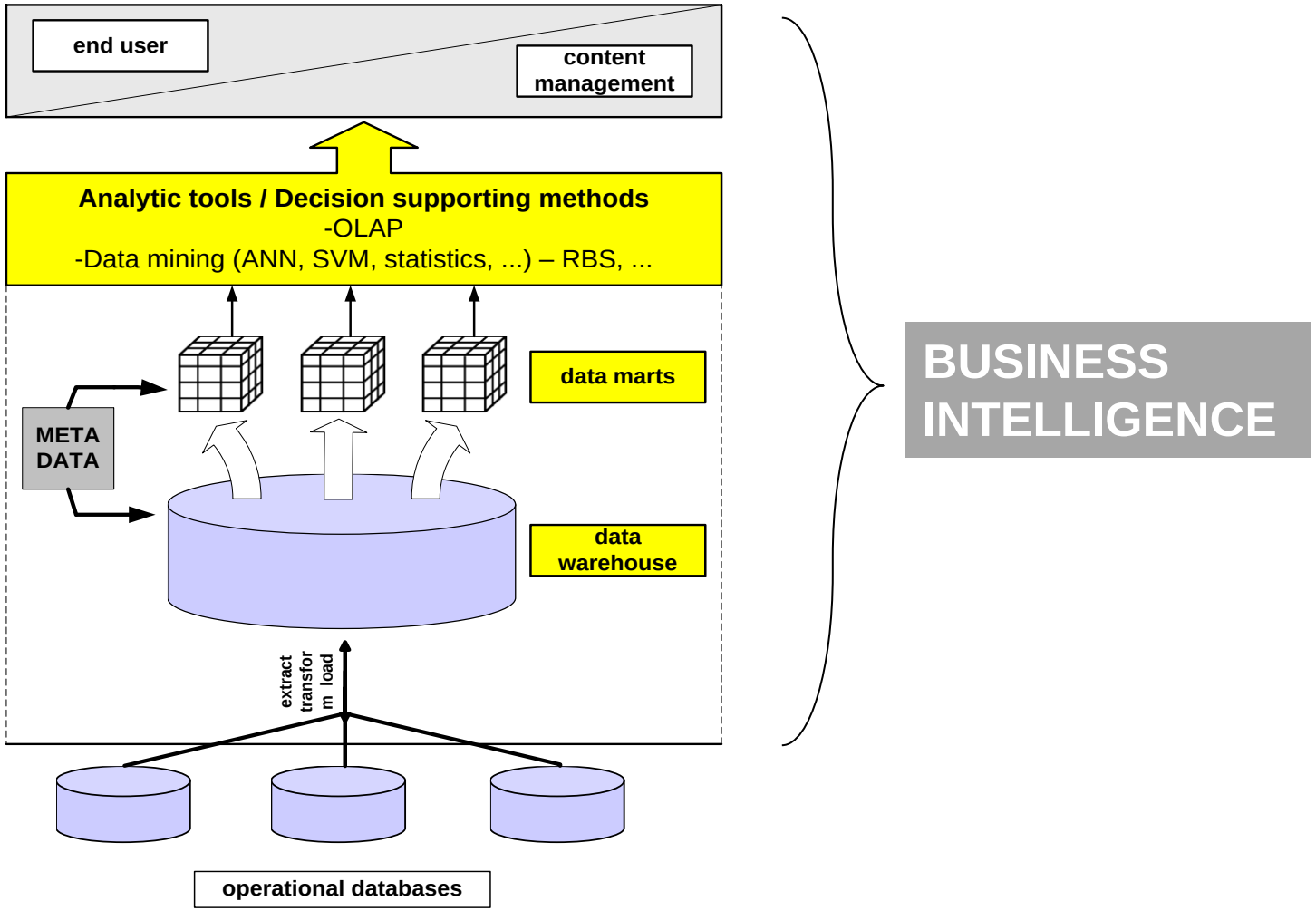
Business intelligence is the conscious, methodical transformation of data from any and all sources into new forms to provide information that is business driven and results oriented. It will often encompass a mixture of tools, databases and vendors.

(Mike Biere, specialist for business intelligence analytic tools, IBM)

Ref.

Other view on Business Intelligence

A holistic view on BI



Ref.

Benefits of Business Intelligence

Freie Universität Berlin

Increased profitability

Distinguish between profitable and non-profitable customers

Improved Customer-Relationship-Management

Analysis of aggregated customer information to provide better customer service, increase customer loyalty

Decreased costs

Lower operational costs, improve logistics management

Decreased risk

Apply Business Intelligence methods to credit data can improve credit risk estimation

Business Intelligence can help improve businesses in a variety of fields:

Customer analysis
→ customer profiling

Sales channel analysis
Behavior analysis
→ fraud detection, shopping trends, web activity, social network analysis

Business productivity analysis
→ defect analysis, capacity planning and optimization, risk management, increase sustainability

Supply chain analysis
→ supply and vendor management, shipping, distribution analysis, sustainable supplier management

Business Intelligence supports decision makers with the required information at the right time and location, and with sufficient quality (format, visualization, validity etc.)

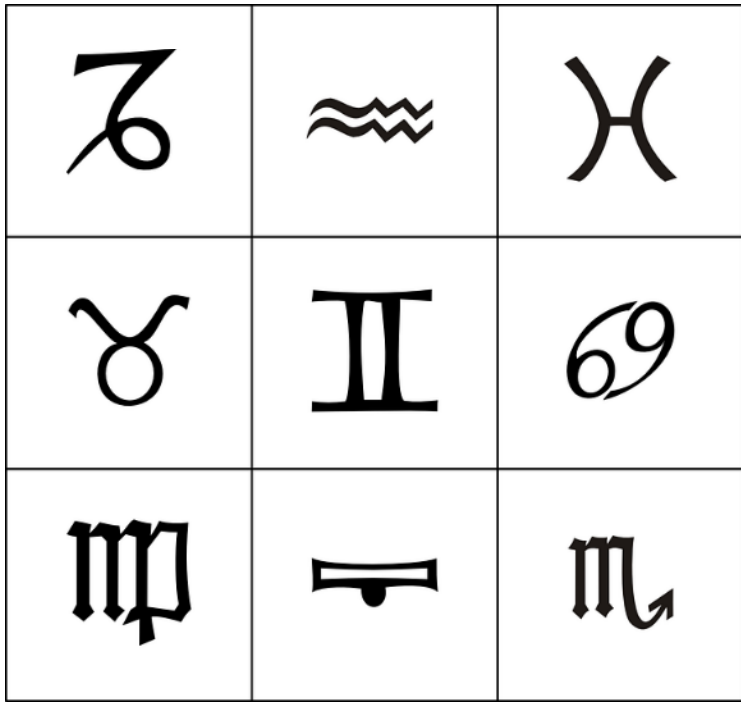
Ref.

Discerning „Data“ from „Information“

Excursus: Semiotics

Semiotics is a field of research in Epystemology („meaning making“) and studies signs (icons, characters, symbols) and sign systems.

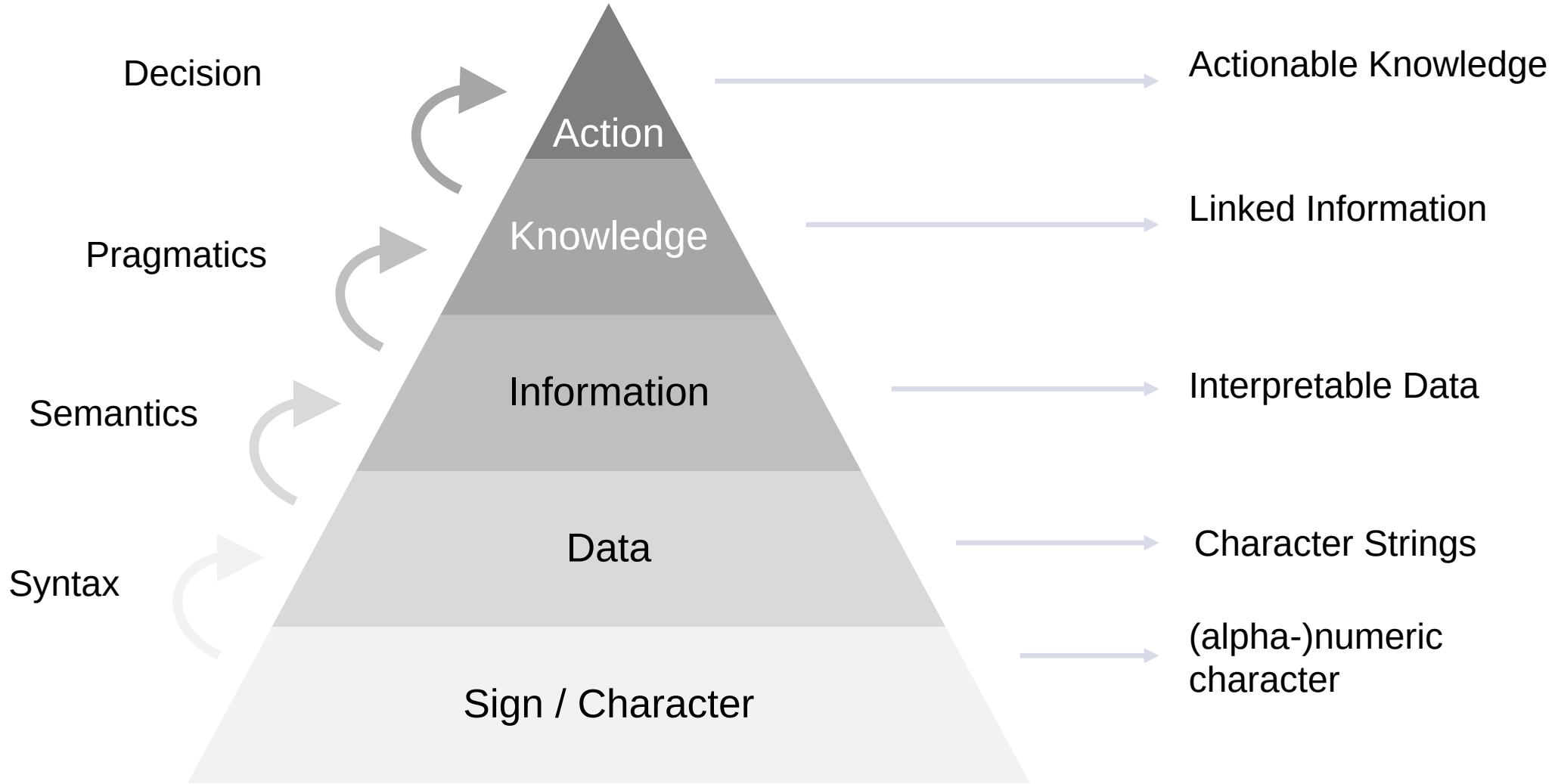
Layer of observation	Object of observation	Example
Syntax	*“Relations among signs in formal structures”*	Is the communicated sign (orthographically/ gramatically) correct?
Semantics	*“Relation between signs and the things to which they refer; their […] meaning”*	What is the meaning of the sign? What is the meaning of „display“ or „windows“?
Pragmatics	*“Relation between signs and the effects they have on the people who use them”*	What is the use of this sign?



Ref. Krcmar (2015), Chandler (2007)

Data transformation (1/2)

Knowledge Pyramid



Ref. Bodendorf (2006)

Data transformation (2/2)

Transforming Data into information; into knowledge; into action

From data to information

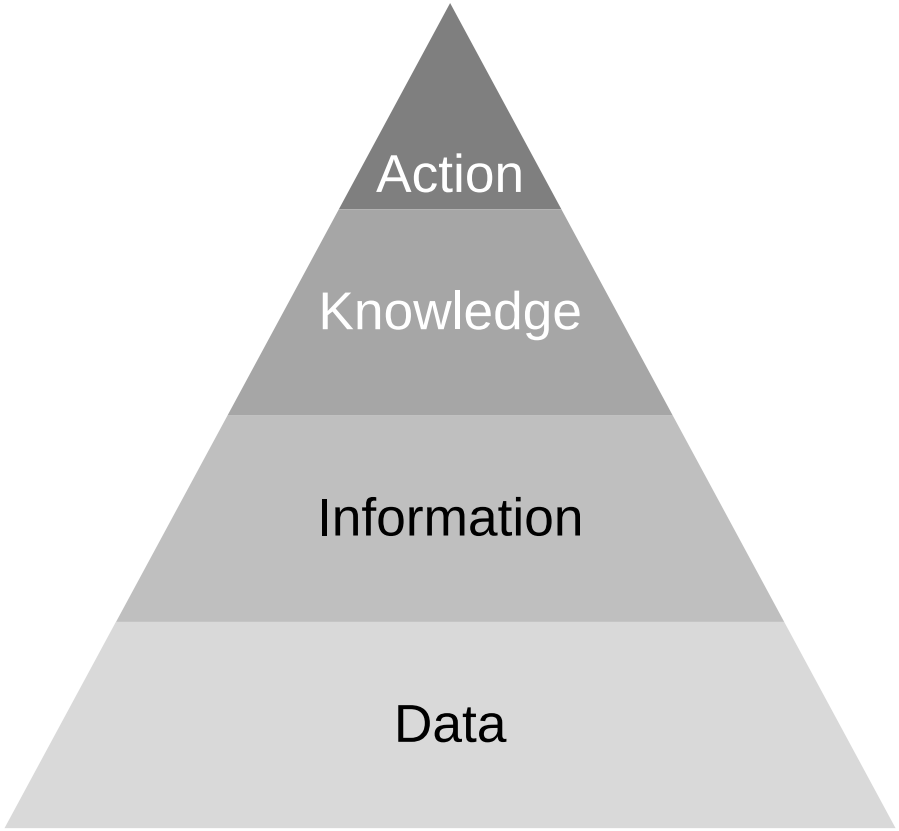
- **Select relevant data** and figure out which configuration makes the data more significant

From information to knowledge

- Piles of information are accumulated and analyzed in different ways
- In this process **analytical tools** are involved

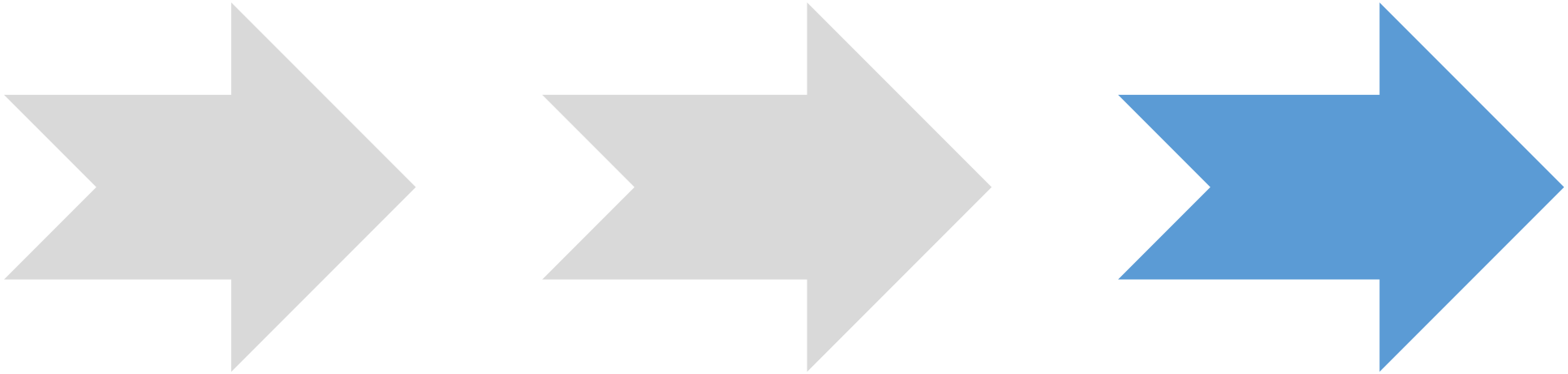
From knowledge to actionable (critical) knowledge

- Knowledge is considered to be critical if it can be used to form a plan of action for **solution of a business problem**



Ref.

Agenda



(1)
Metaintroduction

(2)
Decision Support and
Business Intelligence

(3)
Data-analytic thinking

Ref.

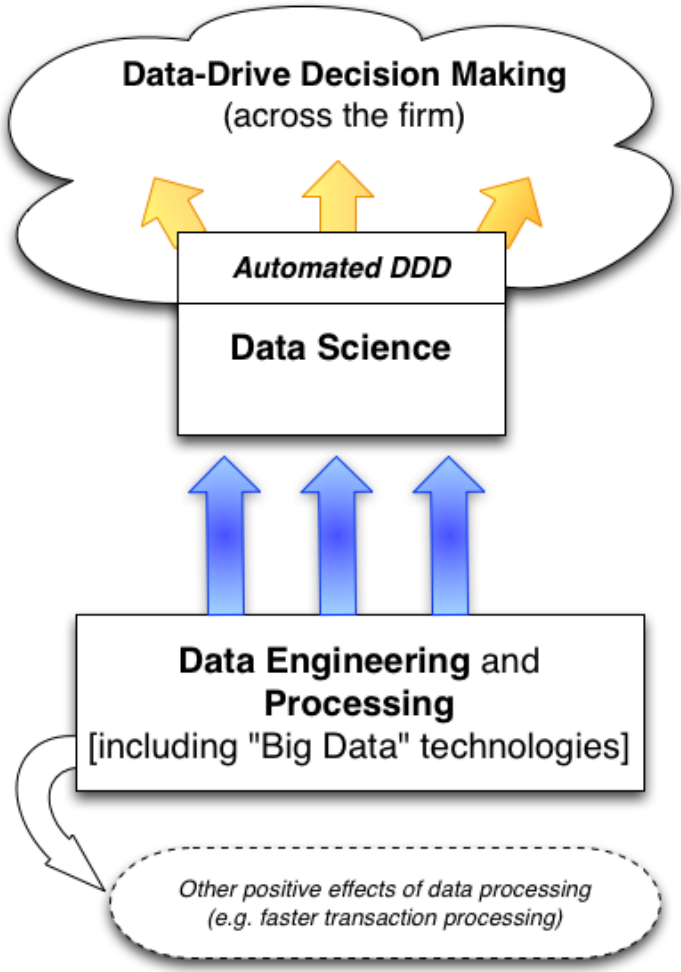
Data Driven Decision-making (DDD)

Legitimizing decisions on the basis of data

Data-driven Decision Making (DDD) is a *"practice of basing decisions on the analysis of data rather than purely on intuition."*
Provost & Fawcett (2013)

Type-1 decision:
"discover" something new within your data

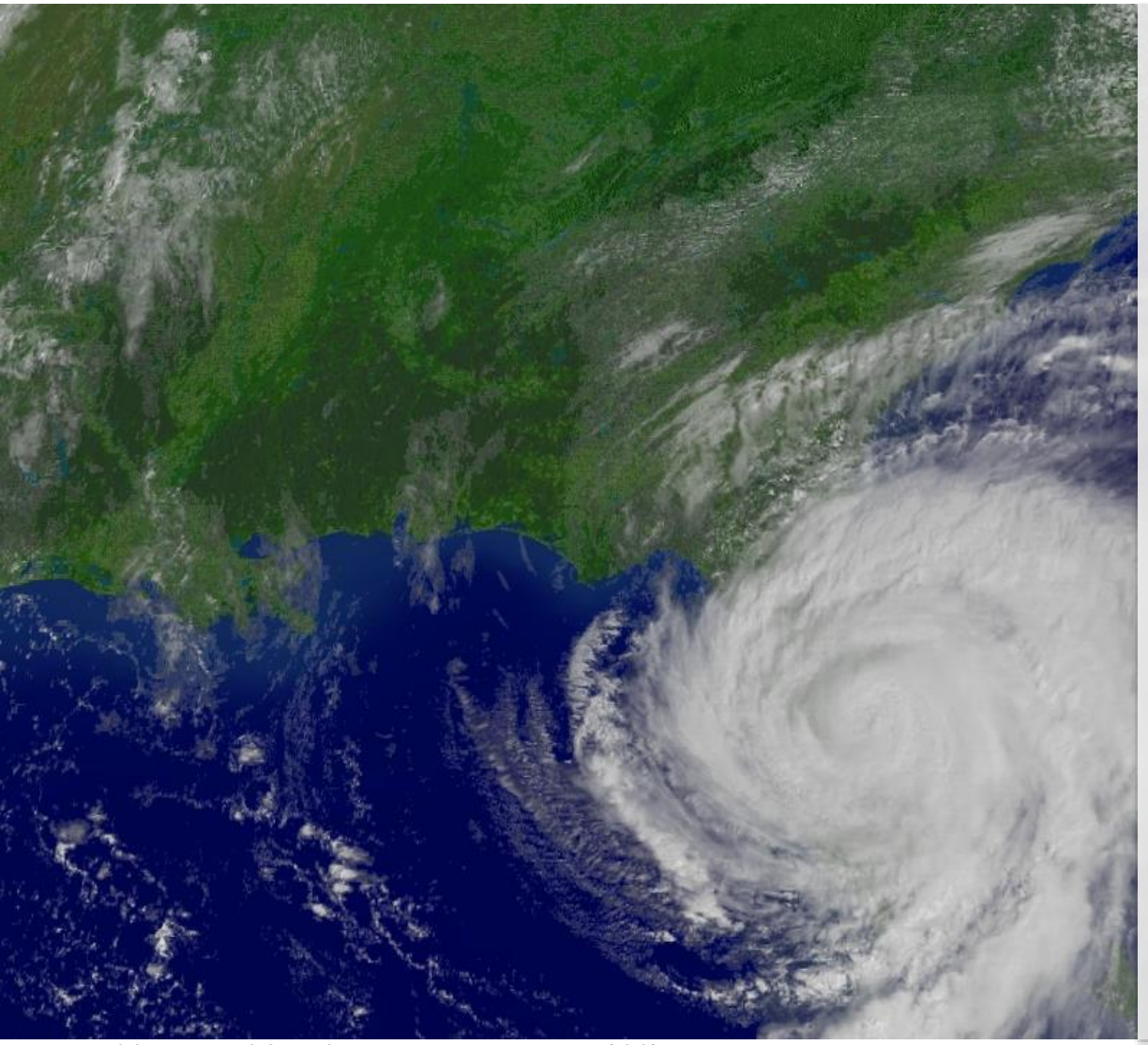
Type-2 decision:
repeat decisions at massive scale
(automatic decision making)



Ref. Provost & Fawcett, 2013 (Link)

Type 1 decision

Discover something new: Hurricane Frances (1/2)



Hurricane Frances was on its way, barreling across the Caribbean, threatening a direct hit on Florida's Atlantic coast. Residents made for higher ground, but far away, in Bentonville, Ark., executives at Wal-Mart Stores decided that the situation offered a great opportunity for one of their newest data-driven weapons ... ***predictive technology****.*

A week ahead of the storm's landfall, Linda M. Dillman, Wal-Mart's chief information officer, pressed her staff to come up with ***forecasts*** *based on what had happened when Hurricane Charley struck several weeks earlier. Backed by the trillions of bytes' worth of shopper history that is stored in Wal-Mart's data warehouse, she felt that the company could 'start predicting what's going to happen, instead of waiting for it to happen,' as she put it.*

(Hays, New York Times, 2004)

Ref. Link to context information, Provost & Fawcett, 2013 (Link)

Img. Frances | Wikipedia

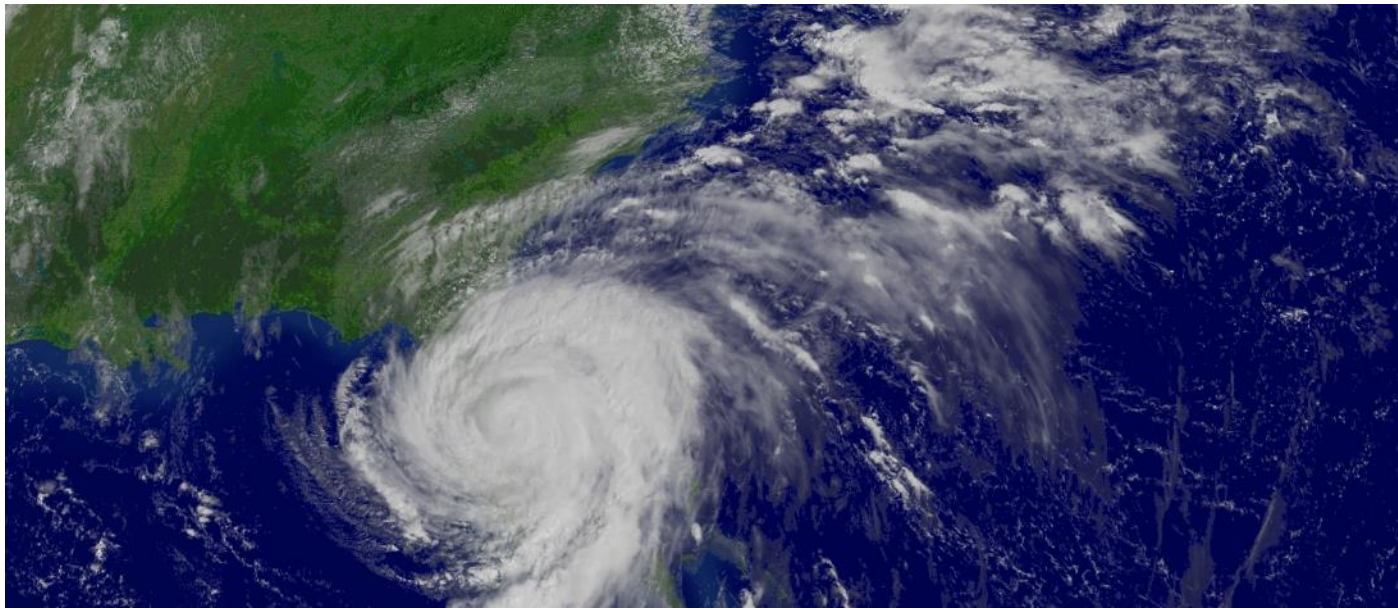
Type 1 decision

Discover something new: Hurricane Frances (2/2)

Why might data-driven prediction be useful?

Analytic thinking: discover sales patterns due to the hurricane that are not obvious!

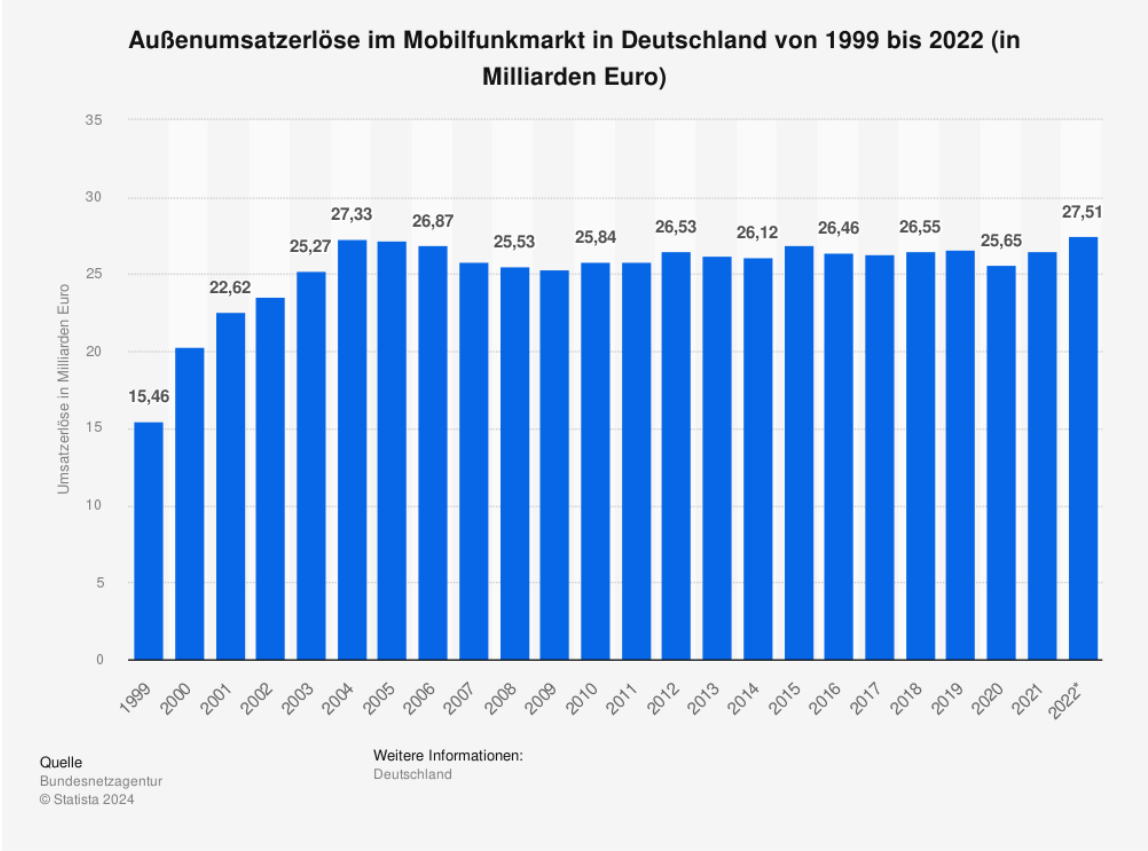
Identify unusual local demand for products (not through hypothesis testing, but data exploration)



Ref. Provost & Fawcett, 2013 (Link)

Type 2 decision

Automatic decision making: Predicting Customer Churn



Mobile TeleCom market is highly saturated.

Many TeleCom companies have great issues with customer retention.
(current churn rates between 1 and 3 % for Deutsche Telekom mobile)

Customer churn is expensive.

Who should get a retention offer?

Raising accuracy of a prediction has huge effects on profitability

Ref. Churn rates from Statista

Knowing about DDD, how do we proceed?

We follow the path, desribed by semiotics

Data processing and “Big Data”

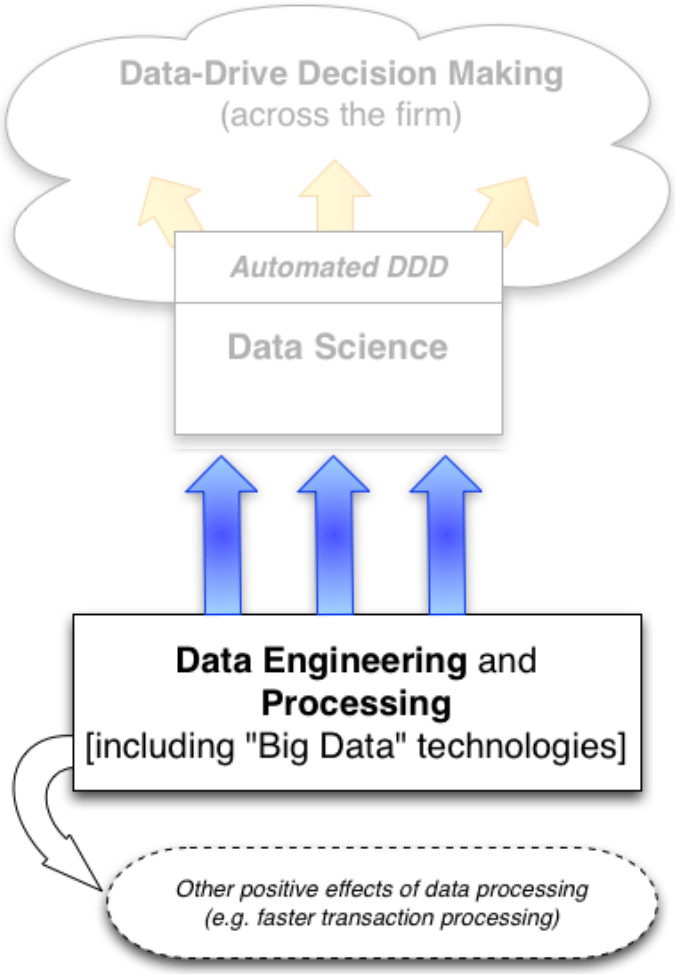
Where‘s the challenge?

Data engineering and data processing are critical to support data science

But: data engineering is not data science

Data engineering handles for instance Data Warehouses, which

- collect data from operational databases
- accumulate **historical data**
- provide the basis for Business Intelligence applications



Where is the catch?

“Big data” means data sets that are too large for traditional data processing systems
Using big data technologies is associated with additional **productivity growth**

Ref.

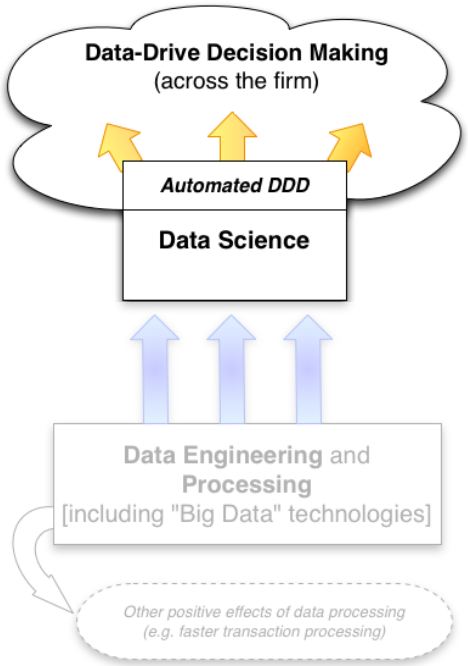
Data Science vs. Data Mining

Freie Universität Berlin

Data science:
A set of *fundamental principles* that guide the extraction of knowledge from data

Data mining:
Extraction of knowledge from data via tools/technologies that *incorporate* the principles

"Data science is an interdisciplinary field aiming to turn data into real value […] Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and predictions, and the exploitation of results taking into account ethical, social, legal, and business aspects"
(Van der Aalst 2016)



In this class, we do both!

Ref.

Data Science

Fundamental principles/concepts which we will get to know

… use well-defined stages for analysis

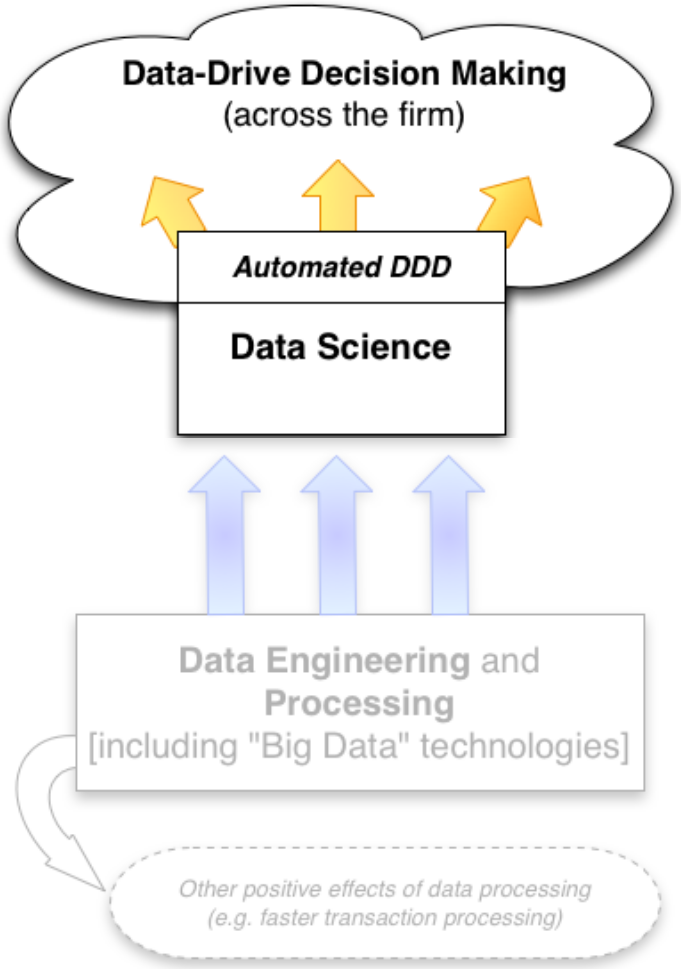
… find informative descriptive attributes

… be careful with overfitting

… think about the context when evaluating results

… data & data science capability as strategic asset

… and many more



Ref.

Example: Data & data science capability as strategic asset

Freie Universität Berlin

Reading

- Data and the capability to make decisions from data are complementary assets.

1990s: Signet Bank aims at modeling of profitability of credit card customers.

However, there was only data for the terms they had offered in the past.

They conducted experiments in order to build predictive models from the data (charge-offs!).

Today one of the largest credit-card issuers.

Ref.

Some examples

In which fields do we use data for gaining competitive advantages?

Marketing (Advertising)

Online advertising

Recommendations for cross-selling

Customer relationship management

Production

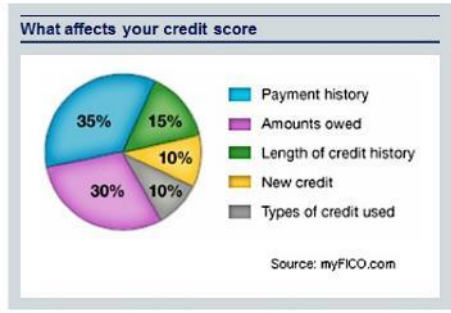
Robotics and automation

Finance

Credit scoring and (high-frequency) trading

Fraud detection

Workforce management



Retail

Supply Chain Management (Wal-Mart, Amazon etc.)

Services

Voice-activated services

Human-machine communication (Siri, Alexa, Google Assistant)

Augmented Reality Services

Ref.

Fragen?

- ✓ Metaintroduction
- ✓ Decision Support and Business Intelligence
- ✓ Data-analytic thinking

Todos for next Week

1. Read short example about data & data science capability as strategic asset (case of Signet Bank)
   Kursmaterial > Readings/Übungen

2. Read Goes (2014): How does the current acknowledgement of data as a valuable resource influence our research domain?
   Kursmaterial > Readings/Übungen


EDITOR'S COMMENTS

Big Data and IS Research

By: Paulo B. Goes
Editor-In-Chief, *MIS Quarterly*
Salter Professor of Technology and Management
Head, Management Information Systems
Eller College of Management
University of Arizona
pgoes@eller.arizona.edu

MIS Quarterly

Publication history	1977 - present
Frequency	Quarterly
Ranking (Jourqual)	A+
Impact factor (2015)	5.384
Cited half-life	>10.0 years

3. Start your Python journey
   Kursmaterial > Readings/Übungen > Python Übungen > Jupyter Notebooks

4. Remember your basics on data base design (including normalization)

Ref. Goes (2014), VHB Jourqual (Gesamtliste)

Image: Cover MISQ

Bibliography

- Bodendorf, Freimut. *Daten-und Wissensmanagement*. Springer-Verlag, 2006.
- Chen, Hsinchun, Roger HL Chiang, and Veda C. Storey. "Business intelligence and analytics: From big data to big impact." *MIS quarterly* 36.4 (2012): 1165-1188.
- Chandler, Daniel. *Semiotics: the basics*. Routledge, 2007.
- Gluchowski, Peter, and Peter Chamoni, eds.: *Analytische Informationssysteme: Business Intelligence-Technologien und-Anwendungen*. Springer-Verlag, 2015.
- George, Gerard, Martine R. Haas, and Alex Pentland. "Big data and management." *Academy of Management Journal* 57.2 (2014): 321-326.
- Krcmar, Helmut. "Informationsmanagement." *Informationsmanagement*. Springer Berlin Heidelberg, 2015. 85-111
- McAfee, Andrew, et al. "Big data." *The management revolution. Harvard Bus Rev* 90.10 (2012): 61-67..
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51-59.
- Power, Daniel J. "A brief history of decision support systems." *DSSResources. COM, World Wide Web, http://DSSResources. COM/history/dsshistory. html, version* 4 (2007).
- Russom, Philip. "Big data analytics." *TDWI best practices report, fourth quarter* (2011): 1-35.
- Schieder, C., Dinter, B., Gluchowski, P.: *Metadatenmanagement in der Business Intelligence – eine empirische Untersuchung unter Berücksichtigung der Stakeholder-Perspektiven*. In: 12th International Conference on Wirtschaftsinformatik, 2015.

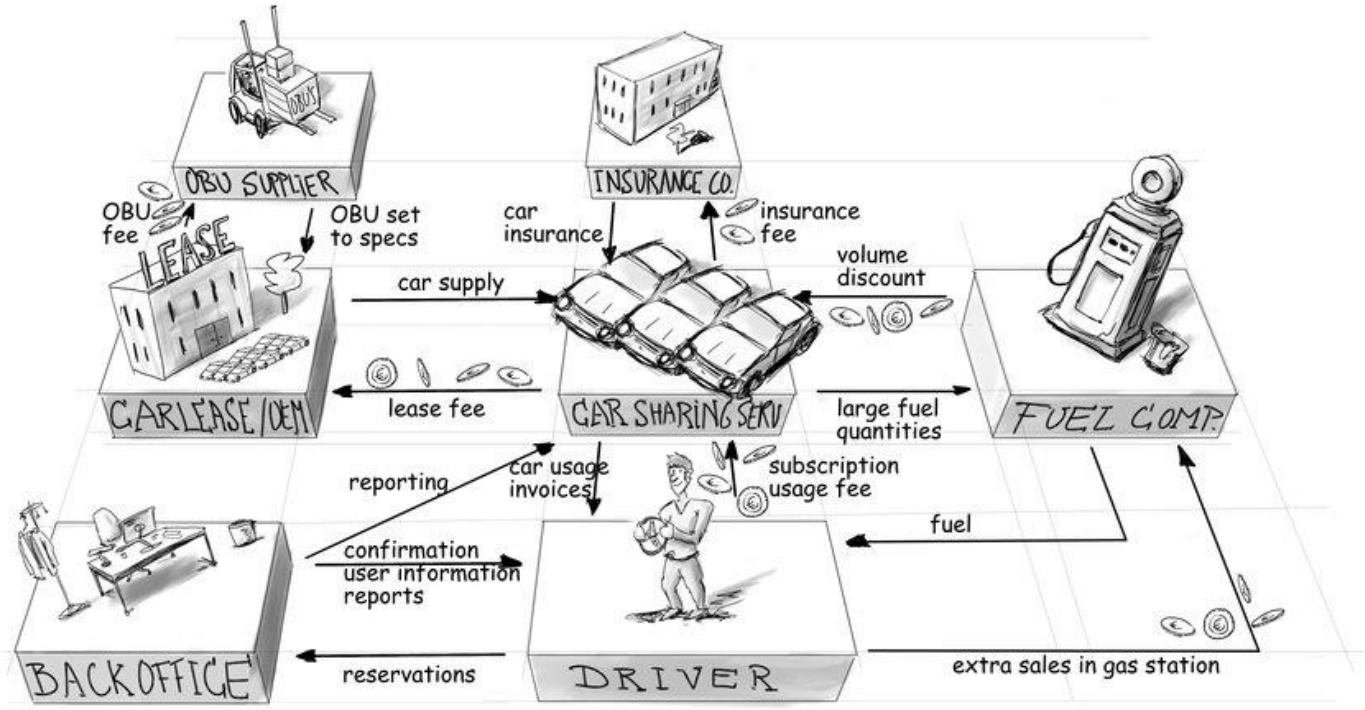
Ref.

Data - Why should companies care?

Data changes business models

“[C]ompanies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable”

(McAfee und Brynjolfsson, 2012, Result of 330 Interviews with Top-Managers and analysis of annual reports in Northern America)



Ref. McAfee Brynjolfsson (2012) , Osterwalder Pigneur (2010)