Data Science 2

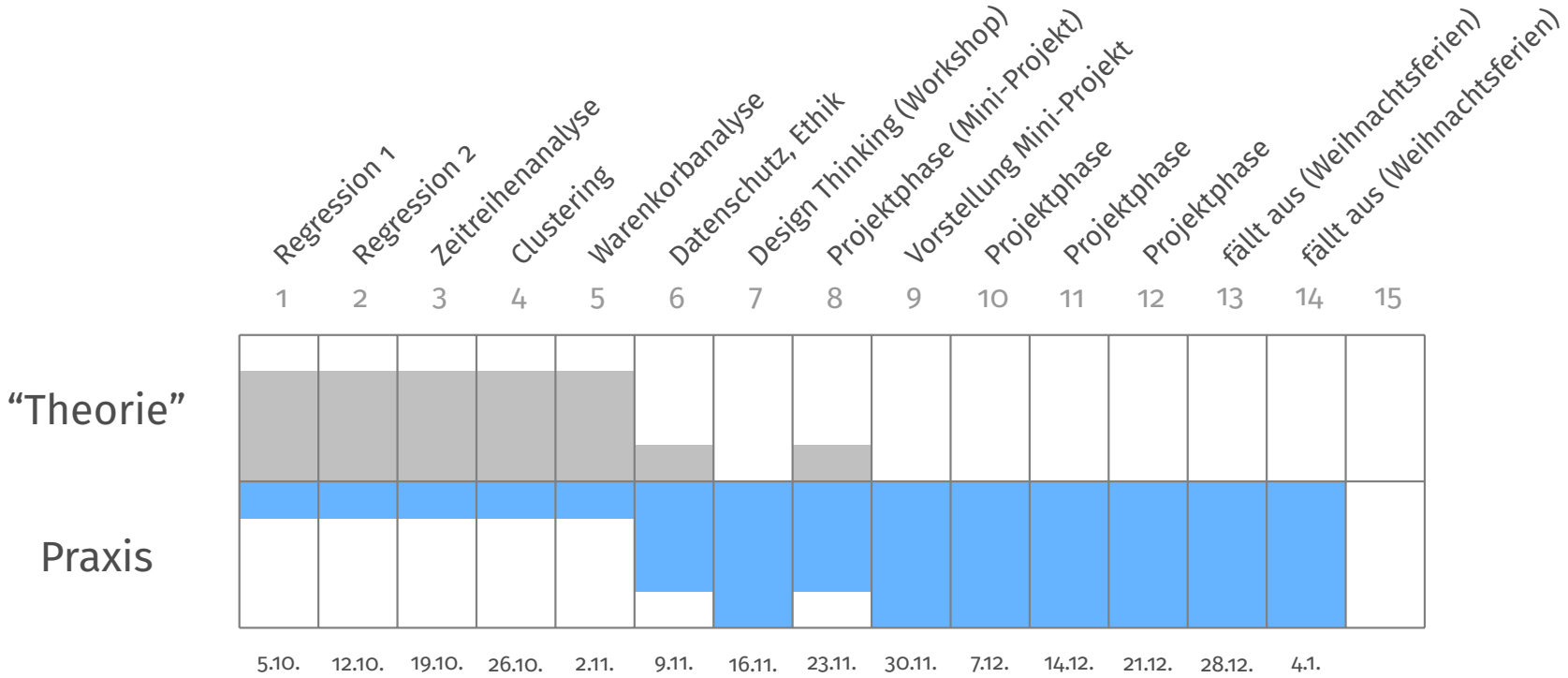
Vorlesung - Datenschutz, Ethik

Prof. Dr. Christian Bockermann

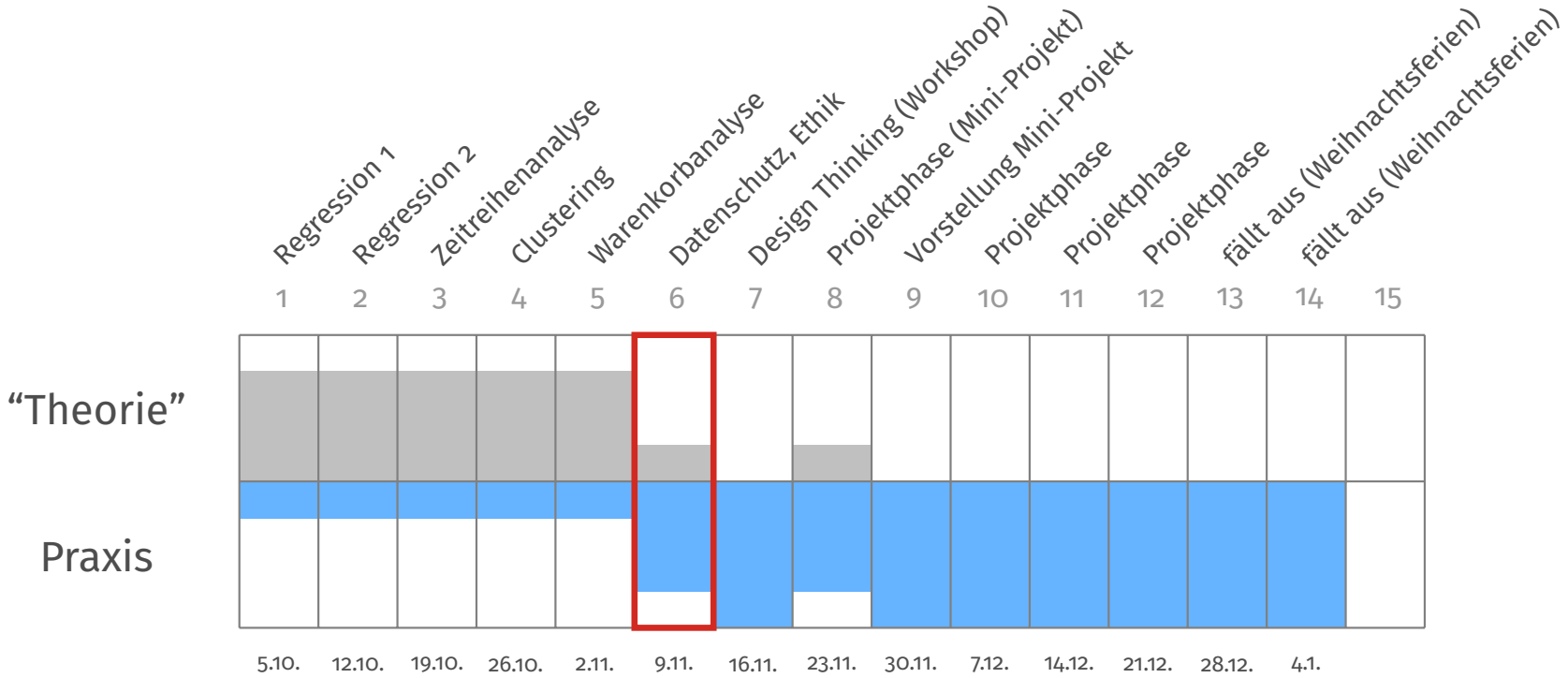
Hochschule Bochum

Wintersemester 2023 / 2024

Themen der Vorlesung



Themen der Vorlesung



Inhalt

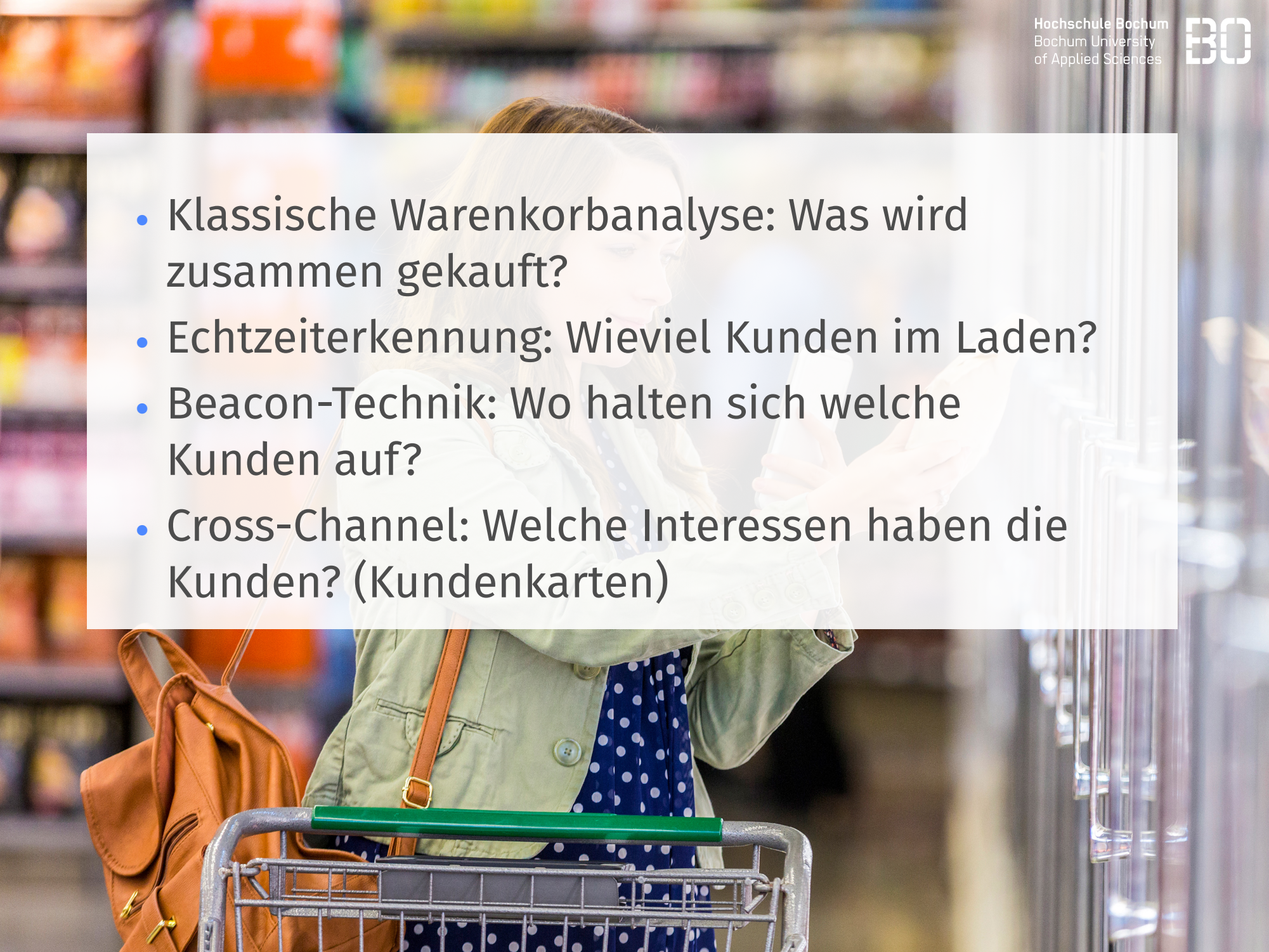
Vorwort

Diese Vorlesung erhebt keinen Anspruch auf

- Rechtliche Korrektheit und Verbindlichkeit
- Vollständigkeit und fachliche Korrektheit

Hochschule Bochum
Bochum University
of Applied Sciences

Datenanalyse im Handel



- Klassische Warenkorbanalyse: Was wird zusammen gekauft?
- Echtzeiterkennung: Wieviel Kunden im Laden?
- Beacon-Technik: Wo halten sich welche Kunden auf?
- Cross-Channel: Welche Interessen haben die Kunden? (Kundenkarten)

Warenkorb-Analyse

- Finden häufiger Mengen/Muster
- Vorlesung 6 in Data Science 2

Warenkorb-Analyse

- Finden häufiger Mengen/Muster
- Vorlesung 6 in Data Science 2

Kundenprofile

- Supermarkt kennt Produkte eines Kunden
- Über Kundenkarte zusätzlich Historie

Warenkorb-Analyse

- Finden häufiger Mengen/Muster
- Vorlesung 6 in Data Science 2

Kundenprofile

- Supermarkt kennt Produkte eines Kunden
- Über Kundenkarte zusätzlich Historie
- Kunden werden anhand ihrer Produkte gruppiert (Clustering, Vorlesung 5 in Data Science 2)
- Zielgerichtete Werbung, Promotions,...

Was weiss der Supermarkt?

- Einzugsbereich, Kundenaufkommen (Zeit)
- Einzelne Bons, Produktverkauf nach Tageszeit

Was weiss der Supermarkt?

- Einzugsbereich, Kundenaufkommen (Zeit)
- Einzelne Bons, Produktverkauf nach Tageszeit

# last 12 M	$ last 12 M	Class	Schoki
4	623,38	Low	X
6	1.492,23	High	X
7	914,98	Mid	
8	1.378,43	High	
3	416,18	Low	X

Was weiss der Supermarkt?

- Einzugsbereich, Kundenaufkommen (Zeit)
- Einzelne Bons, Produktverkauf nach Tageszeit

# last 12 M	$ last 12 M	Class	Schoki
4	623,38	Low	X
6	1.492,23	High	X
7	914,98	Mid	
8	1.378,43	High	
3	416,18	Low	X

Katzen	Vegan
X	
	X
X	X
	X
X	

Was weiss der Supermarkt?

- Einzugsbereich, Kundenaufkommen (Zeit)
- Einzelne Bons, Produktverkauf nach Tageszeit

# last 12 M	$ last 12 M	Class	Schoki
4	623,38	Low	X
6	1.492,23	High	X
7	914,98	Mid	
8	1.378,43	High	
3	416,18	Low	X

Katzen	Vegan
X	
	X
X	X
	X
X	

Sport	Mode
X	X
	X
X	X
	X

Was weiss der Supermarkt?

- Einzugsbereich, Kundenaufkommen (Zeit)
- Einzelne Bons, Produktverkauf nach Tageszeit

# last 12 M	$ last 12 M	Class	Schoki
4	623,38	Low	X
6	1.492,23	High	X
7	914,98	Mid	
8	1.378,43	High	
3	416,18	Low	X

Katzen	Vegan
X	
	X
X	X
	X
X	

Sport	Mode
X	X
	X
X	X
	X

3,522,031 views | Feb 16, 2012, 11:02am

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Hochschule Bochum
Bochum University
of Applied Sciences

Datenanalyse in der Medizin

Datenanalyse in der Medizin

- Erkennen von Krankheiten
  (vgl. “Husten” – Vorlesung 5, Data Science 1)
- Individualisierung von Therapien/Medikamenten

Datenanalyse in der Medizin

- Erkennen von Krankheiten
  (vgl. “Husten” – Vorlesung 5, Data Science 1)
- Individualisierung von Therapien/Medikamenten


Datenanalyse HZM, 06.02.2018

Typ-2-Diabetes: Das kosten die Folgeerkrankungen

© everythingpossible - Fotolia

Datenanalyse in der Medizin

- Erkennen von Krankheiten
  (vgl. “Husten” – Vorlesung 5, Data Science 1)
- Individualisierung von Therapien/Medikamenten


Datenanalyse HZM, 06.02.2018

Typ-2-Diabetes: Das kosten die Folgeerkrankungen

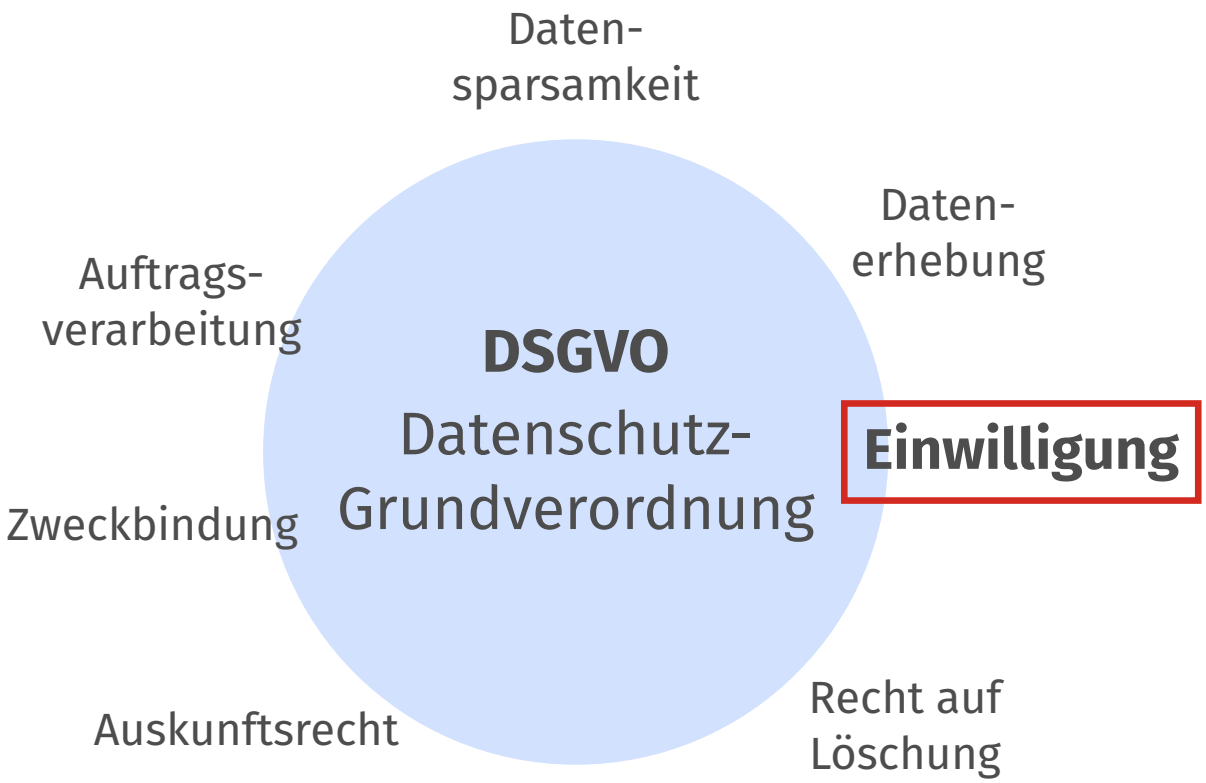

Mit einem schmutzigen Trick kann Facebook rausfinden, wann Sie Ihre Periode haben

Samstag, 23.02.2019, 10:15 · · von FOCUS-Online-Redakteur Florian Reiter

Doku: Das Dilemma mit den sozialen Medien (Netflix)

Datenschutz


Daten-
sparsamkeit
Daten-
erhebung
Auftrags-
verarbeitung
DSGVO
Datenschutz-
Grundverordnung
Einwilligung
Zweckbindung
Auskunftsrecht
Recht auf
Löschung


Daten-
sparsamkeit
Daten-
erhebung
Auftrags-
verarbeitung
DSGVO
Datenschutz-
Grundverordnung
Einwilligung
Zweckbindung
Recht auf
Löschung
Auskunftsrecht

Bevor Sie fortfahren

Google verwendet Cookies und Daten für Folgendes:

- Dienste anbieten und betreiben, z. B. Störungen prüfen und Maßnahmen gegen Spam, Betrug oder Missbrauch ergreifen
- Daten zu Zielgruppeninteraktionen und Websitestatistiken erheben, um zu verstehen, wie unsere Dienste verwendet werden

Wenn Sie zustimmen, verwenden wir Cookies und Daten auch für Folgendes:

- Qualität unserer Dienste verbessern und neue Dienste entwickeln
- Werbung ausliefern und die Effektivität von Werbung messen
- Personalisierte Inhalte anzeigen, abhängig von Ihren Einstellungen
- Personalisierte oder allgemeine Werbung bei Google und im Web anzeigen, abhängig von Ihren Einstellungen

Bevor Sie fortfahren

Google verwendet Cookies und Daten für Folgendes:

- Dienste anbieten und betreiben, z. B. Störungen prüfen und Maßnahmen gegen Spam, Betrug oder Missbrauch ergreifen
- Daten zu Zielgruppeninteraktionen und Websitestatistiken erheben, um zu verstehen, wie unsere Dienste verwendet werden

Wenn Sie zustimmen, verwenden wir Cookies und Daten auch für Folgendes:

- Qualität unserer Dienste verbessern und neue Dienste entwickeln
- Werbung ausliefern und die Effektivität von Werbung messen
- Personalisierte Inhalte anzeigen, abhängig von Ihren Einstellungen
- Personalisierte oder allgemeine Werbung bei Google und im Web anzeigen, abhängig von Ihren Einstellungen

Anonymisierung

Name	Vorname	PLZ	# last 12 M	$ last 12 M	Class
Fox	Peter	13124	4	623,38	
Lennon	John	12921	6	1.492,23	
Jackson	Peter	13174	7	914,98	
Tyson	Mike	12511	8	1.378,43	
Brown	Alice	13124	3	416,18	

Anonymisierung

Name	Vorname	PLZ	# last 12 M	$ last 12 M	Class
	Peter	13124	4	623,38	
	John	12921	6	1.492,23	
	Peter	13174	7	914,98	
	Mike	12511	8	1.378,43	
	Alice	13124	3	416,18	

Anonymisierung

Name	Vorname	PLZ	# last 12 M	$ last 12 M	Class
	Peter	13124	4	623,38	
	John	12921	6	1.492,23	
	Peter	13174	7	914,98	
	Mike	12511	8	1.378,43	
	Alice	13124	3	416,18	

Anonymisierung

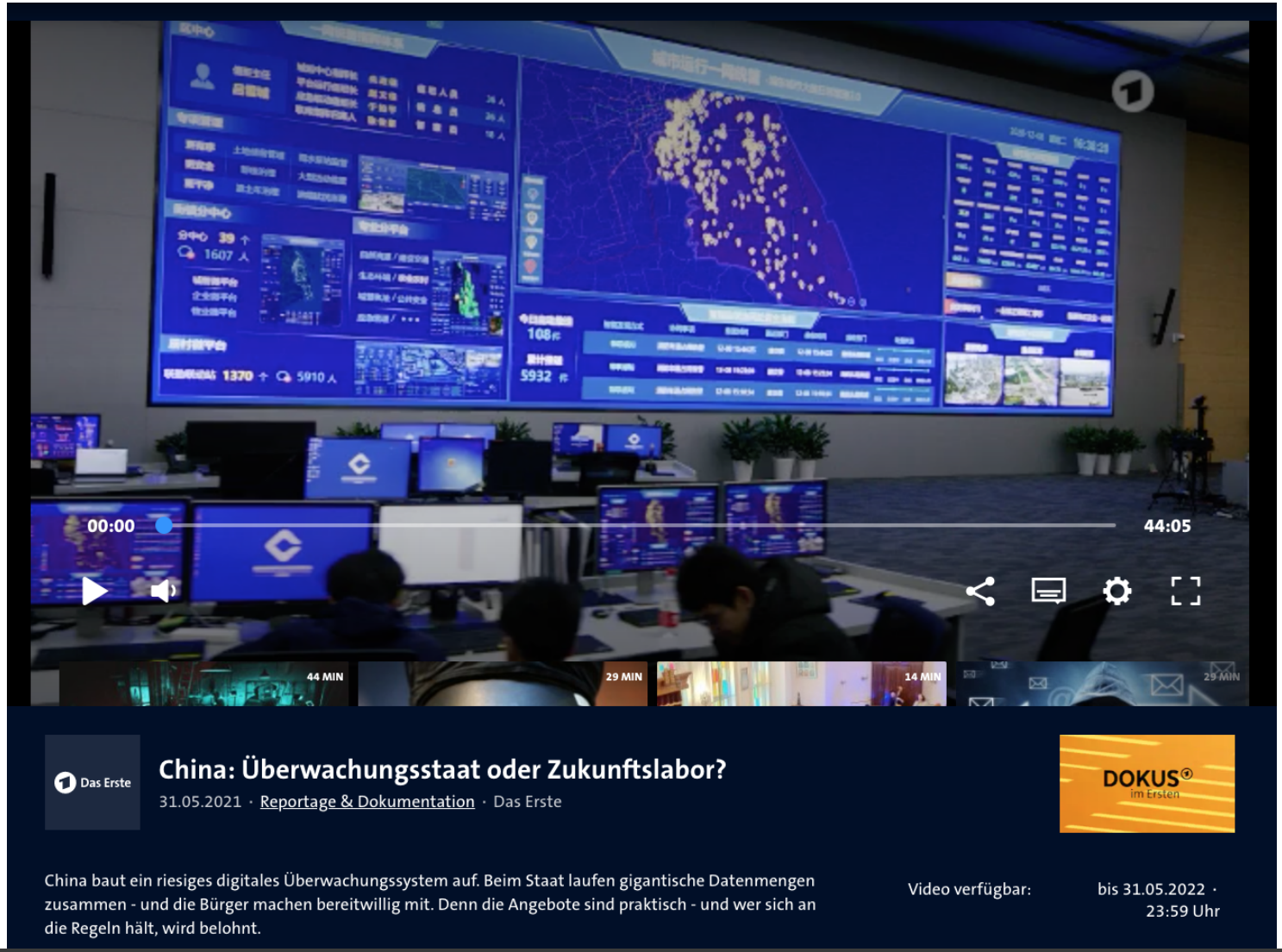
Name	Vorname	PLZ	# last 12 M	$ last 12 M	Class
	Peter	**13xxx**	4	623,38	
	John	**12xxx**	6	1.492,23	
	Peter	**13xxx**	7	914,98	
	Mike	**12xxx**	8	1.378,43	
	Alice	**13xxx**	3	416,18	

Pseudonymisierung

ID	PLZ	# last 12 M	$ last 12 M	Class
001	13xxx	4	623,38	
002	12xxx	6	1.492,23	
003	13xxx	7	914,98	
004	12xxx	8	1.378,43	
005	13xxx	3	416,18	

Personenbezogene Daten
durch Pseudonym ersetzen

China: Überwachungsstaat oder Zukunftslabor?



Learning Analytics

Learning Analytics – Wie gut ist (e-) Learning?

Learning Analytics – Wie gut ist (e-) Learning?

- Messen und Analyse von (e-)Learning
- Gezielte Förderung von Studierenden
- Planung/Gestaltung von Studienangeboten (vgl. Umfrage zu Master in DataScience)
- Privacy/Datenschutz hat höchste Priorität

Beispiel: Log Daten des Data-Science Servers

- Zugriffsprotokoll auf Web-Seite
- Zweck ist die technische Überwachung des Servers

Web-Server Access Logs

Typische Felder sind:

IP Adresse	Quelle IP Adresse der Verbindung
Datum	Datum+Uhrzeit des Zugriffs
Method+URI	Zugriffsmethode und URL
Status-Code	Status der Antwort (Ok, Fehler,..)
Referer	Vorangegangene URL
User-Agent	Bezeichnung des Browsers

Access Logs

Beispiel: `datascience.hs-bochum.de`

- Log Daten von 5.10.2020 bis 21.6.2021
- 1.405.942 Anfragen in ca. 260 Tagen
- von 1200 verschiedene IP-Addressen
- 24.419 verschiedene URLs
- 326 verschiedene Browser-Typen

Fragestellung

Führt die Nutzung des Jupyter-Servers zu besseren Noten in den Hausarbeiten/Klausuren des Kurses Data Science?

Fragestellung

Führt die Nutzung des Jupyter-Servers zu besseren Noten in den Hausarbeiten/Klausuren des Kurses Data Science?

- Nutzungsverhalten pro Teilnehmer:in
- Vorhersage: Nutzungsverhalten <-> Note?

Fragestellung

> Führt die Nutzung des Jupyter-Servers zu besseren Noten in den Hausarbeiten/Klausuren des Kurses Data Science?

- Nutzungsverhalten pro Teilnehmer:in
- Vorhersage: Nutzungsverhalten <-> Note?
- Aktivitätsprofile: Nachteule oder nicht?

Fragestellung

> Führt die Nutzung des Jupyter-Servers zu besseren Noten in den Hausarbeiten/Klausuren des Kurses Data Science?

- Nutzungsverhalten pro Teilnehmer:in
- Vorhersage: Nutzungsverhalten <-> Note?
- Aktivitätsprofile: Nachteule oder nicht?
- Endgeräte: Wer nutzt iPhone/Android/Windows/Linux/Mac?

Wieviel würden Sie preisgeben?

Learning Analytics in Data Science:

- Teilnahme an der Vorlesung/Übung
- Studiengang, Berufsziel
- Interessen, Hobbies
- Dauer der Nutzung des Servers
- Dauer der Arbeit an bestimmten Notebooks/Übungsblättern
- Bearbeitung/Nicht-Bearbeitung von einzelnen Ü-Aufgaben
- Exakte Zeitpunkte der Nutzung des Servers
- Endgerät, Browser,...
- Ungefährer Ort (Stadt) bei der Nutzung

Welche Plattformen/Apps nutzen Sie?

- TikTok, WhatsApp, Signal
- FaceBook, Twitter, Instagram
- ...