

# DATA SCIENCE 2

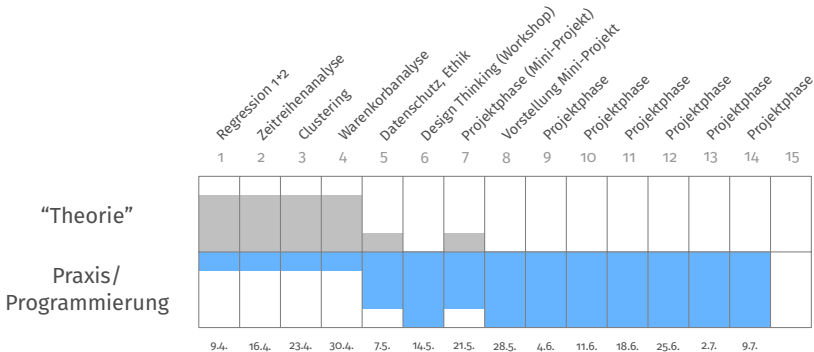
CANVAS WORKSHOP

PROF. DR. CHRISTIAN BOCKERMANN

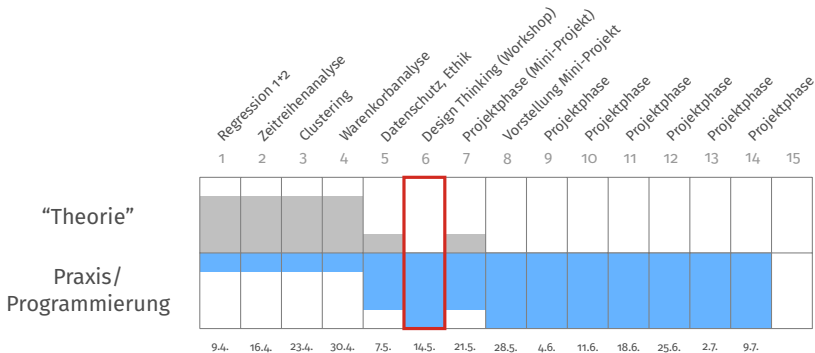
HOCHSCHULE BOCHUM

SOMMERSEMESTER 2024

## Themen der Vorlesung

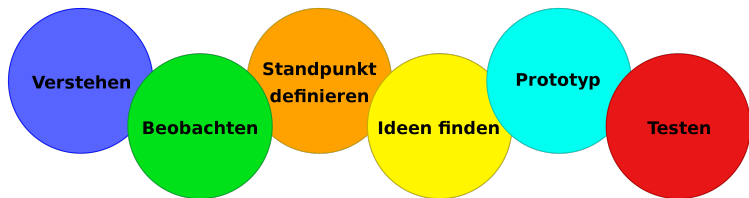


## Themen der Vorlesung



- 1 Design Thinking
- 2 Von der Idee zur Analyse
- 3 Projektphase

## Warum heute keine **Design Thinking** Vorlesung



- Design Thinking – Innovationsprozess
- Nach Terry Winograd, Larry Leifer, David Kelley (alle Stanford)

## Design Thinking – Phasen

1. **Frame a question**  
Finden Sie eine treibende Frage, die andere inspiriert nach kreativen Lösungen zu suchen
2. **Gather Inspiration**  
Inspirieren Sie zum Denken von Neuem, indem Sie herausfinden, was Menschen wirklich brauchen
3. **Generate Ideas**  
Lassen sie offensichtliche Lösungen hinter sich, um zu bahnbrechenden Ideen zu gelangen
4. **Make Ideas Tangable**  
Erstellen Sie grobe Prototypen um zu begreifen, wie Sie Ideen verbessern können
5. **Test to Learn**  
Verfeinern Sie Ideen, indem Sie Feedback sammeln und vorwärts experimentieren
6. **Share the Story**  
Erstellen Sie eine menschliche Geschichte, um andere zum Handeln anzuregen.

# Von der Idee zur Analyse

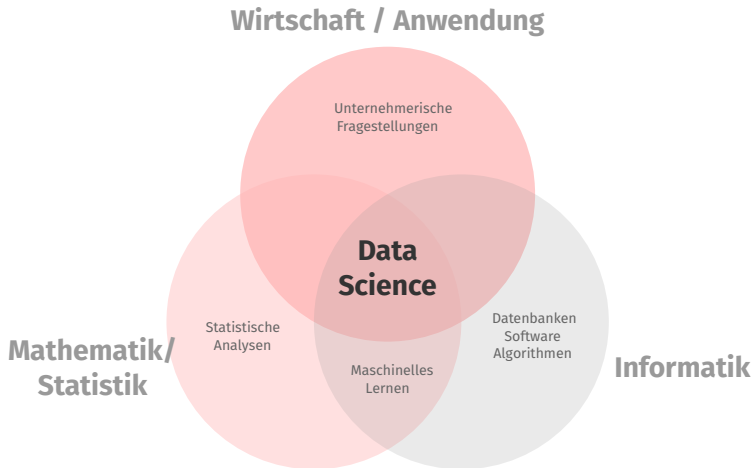
## Worum geht es im Kurs Data Science?

- Etablieren von datengetriebener Denk-/Arbeitsweise
- Kennenlernen von Methoden (ML)
- Wissenschaftliche Arbeitsweise
- Vorbereitung auf BA, Beruf

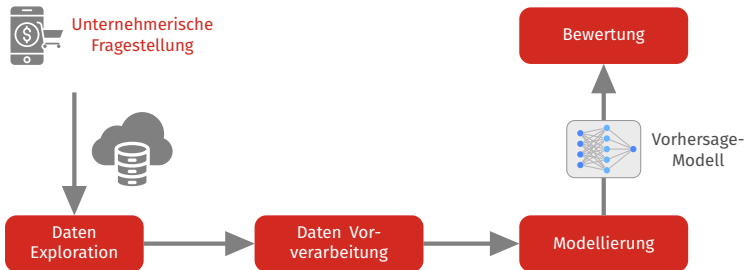


## Worum geht es im Kurs Data Science?

- Etablieren von datengetriebener Denk-/Arbeitsweise
- Kennenlernen von Methoden (ML)
- Wissenschaftliche Arbeitsweise
- Vorbereitung auf BA, Beruf
  
- Praktische Umsetzung erlernen
- Datenbasiertes *Story-Telling*



## Vorgehen bei der Datenanalyse



## Data Science 1

Beispiel für wirtschaftlichen Kontext, Anwendungsfall gegeben;  
vorgegebene Aufgaben:

1. Datenvorverarbeitung
2. Statistiken / Exploration
3. Modellierung / Evaluation

## Data Science 2

**Sie bekommen einen Datensatz.**

## Data Science 2

**Sie bekommen einen Datensatz.**

- Bewusst offene Aufgabenstellung

## Data Science 2

### **Sie bekommen einen Datensatz.**

- Bewusst offene Aufgabenstellung
- Eigenständig Datenanalyse “erarbeiten”
- Bericht/Dossier mit Beschreibung und Analyse erstellen
- Ergebnisse präsentieren
- Eigene Erfahrungen berichten

**Sie haben einen Datensatz bekommen.**

**Und nun?**

- Verstehen: Sammeln Sie Informationen über den Kontext (Fachwissen)



**Sie haben einen Datensatz bekommen.**

**Und nun?**

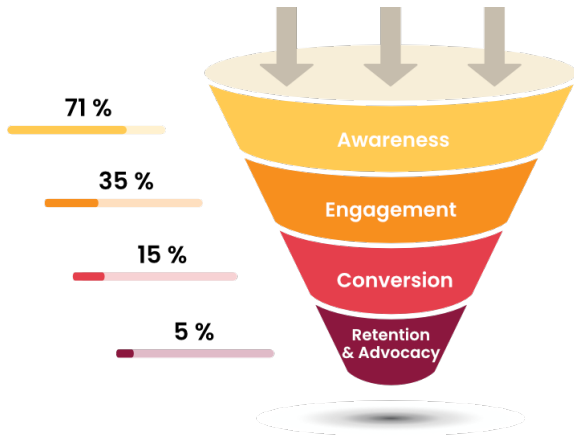
- Verstehen: Sammeln Sie Informationen über den Kontext (Fachwissen)
- Schauen Sie sich die Daten an
- Beschreiben, Fragestellung finden, Analysieren

## Datensatz:

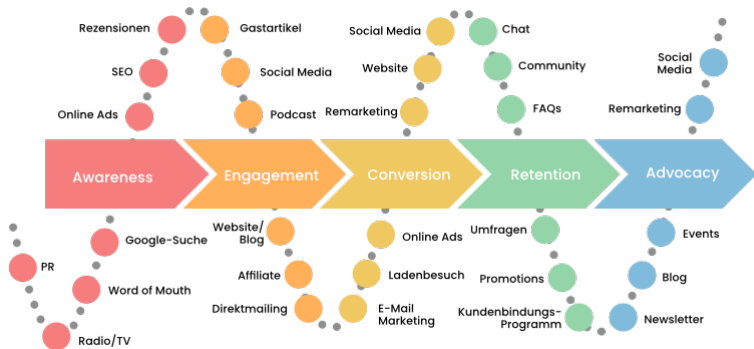
BestellNr	Datum	Kunde	AnzahlArtikel	Brand:A	Brand:B	Brand:C	Brand:D	Brand:E
239821	2019-12-02	3913	1	49.99	0	0	0	0
239822	2019-12-02	1024	1	0	0	21.99	0	0
239823	2019-12-03	0232	2	0	8.99	2.99	0	0
239824	2019-12-03	9218	1	0	0	0	0	13.49
239825	2019-12-04	4120	3	0	0	23.98	0	13.49

## Kontext verstehen

- Wo kommen die Daten her?
- Was beschreiben die Daten? Welche Details sind ggf. versteckt?
- Welche weiteren Daten gibt es in dem Kontext?

**Kontext: Sales Funnel**

## Kontext: Sales Funnel + Customer Journey



[<https://www.kaibader.marketing/sales-funnel/>]

**Datensatz:**

BestellNr	Datum	Kunde	AnzahlArtikel	Brand:A	Brand:B	Brand:C	Brand:D	Brand:E
239821	2019-12-02	3913	1	49.99	0	0	0	0
239822	2019-12-02	1024	1	0	0	21.99	0	0
239823	2019-12-03	0232	2	0	8.99	2.99	0	0
239824	2019-12-03	9218	1	0	0	0	0	13.49
239825	2019-12-04	4120	3	0	0	23.98	0	13.49

**Datensatz:**

BestellNr	Datum	Kunde	AnzahlArtikel	Brand:A	Brand:B	Brand:C	Brand:D	Brand:E
239821	2019-12-02	3913	1	49.99	0	0	0	0
239822	2019-12-02	1024	1	0	0	21.99	0	0
239823	2019-12-03	0232	2	0	8.99	2.99	0	0
239824	2019-12-03	9218	1	0	0	0	0	13.49
239825	2019-12-04	4120	3	0	0	23.98	0	13.49

**Exploration – nach Kunde**

- Durchschnittliche Kauf-Häufigkeit?
- Gibt es viele Wiederholungskäufer? In bestimmten Gruppen? (z.B. männlich/weiblich/Altersgruppe?)
- Durchschnittliche Zeit zwischen Käufen eines Kunden?

Angenommen, wir haben für jeden Kunden noch

- Alter, Geschlecht, ...
- Wohnort/Stadt

## **Modellierung – nach Kunde**

- Welche Kunden haben länger nicht gekauft?
- Ab wann gilt ein Kunde als “verloren”?
- Können wir Kunden erkennen, die nicht mehr kaufen?

Angenommen, wir haben für jeden Kunden noch

- Alter, Geschlecht, ...
- Wohnort/Stadt

## Modellierung – nach Kunde

- Welche Kunden haben länger nicht gekauft?
- Ab wann gilt ein Kunde als “verloren”?
- Können wir Kunden erkennen, die nicht mehr kaufen?

**Churn Prediction!**



## Formalisieren des Problems

- Sei  $T$  die durchschnittliche Zeit zwischen 2 Käufen
- Kunde ist “verloren”, wenn er  $2 \cdot T$  lang nicht kauft

## Formalisieren des Problems

- Sei  $T$  die durchschnittliche Zeit zwischen 2 Käufen
- Kunde ist “verloren”, wenn er  $2 \cdot T$  lang nicht kauft

## Modellierung / Lernaufgabe:

- Merkmal verloren berechnen (0 / 1)
- Aufgabe: Binäre Klassifikation, verloren vorhersagen

**Wie messen wir, ob wir gut vorgegangen sind?**

## Wie messen wir, ob wir gut vorgegangen sind?

### Average Order Value (AOV)

AOV misst die Höhe des durchschnittlichen Bestellwerts der Kunden.

$$AOV = \frac{\text{Umsatz}}{\#\text{Bestellungen}} \quad (1)$$

### Average Revenue per User (ARPU)

ARPU zeigt den Umsatz den ein Kunde in einer bestimmten Zeitperiode dem Unternehmen liefert.

$$ARPU = \frac{\text{Umsatz}}{\#\text{User}} \quad (2)$$

## Erfolgskontrolle

Wir müssen frühzeitig  
Güte-Kriterien festlegen!

## Canvas Strategien

- Strukturierung des Vorgehens
- Guideline für Fragestellungen

## Business Model Canvas

**The Business Model Canvas** Designed for: \_\_\_\_\_ Designed by: \_\_\_\_\_

Date: \_\_\_\_\_  
Revision: \_\_\_\_\_

<p><b>Key Partners</b> </p> <p>Who are my partners? Who are the best ones to do an activity together? What key resources do partners provide? What key activities do partners perform? What key channels do partners use? What key customer segments do partners serve? What key revenue streams do partners receive?</p>	<p><b>Key Activities</b> </p> <p>What key activities do my Value Proposition require? Do my Distribution Channels require certain activities? What key resources do my Key Partners require? What key revenue streams do my Key Partners receive?</p>	<p><b>Value Propositions</b> </p> <p>What value do we create for our customers? What value do our customers really value? What value do our customers get from us that they can't get elsewhere? What value do our customers get from us that they can't get elsewhere? What value do our customers get from us that they can't get elsewhere?</p>	<p><b>Customer Relationships</b> </p> <p>What type of relationship do we want with our Customers? How do we create, build and maintain a relationship with them? How do we create, build and maintain a relationship with them? How do we create, build and maintain a relationship with them? How do we create, build and maintain a relationship with them?</p>	<p><b>Customer Segments</b> </p> <p>For whom are we creating value? What segments are most attractive to us? What segments are most attractive to us? What segments are most attractive to us? What segments are most attractive to us?</p>
<p><b>Key Resources</b> </p> <p>What key resources do my Value Proposition require? Do my Distribution Channels require certain resources? What key resources do my Key Partners require? What key revenue streams do my Key Partners receive?</p>	<p><b>Channels</b> </p> <p>Through which Channels do our Customers Segment want to be reached? How are we reaching them now? How can we reach them better? Which are our most efficient Channels? Which are our most efficient Channels? Which are our most efficient Channels? Which are our most efficient Channels?</p>	<p><b>Cost Structure</b> </p> <p>What are the most important costs inherent in our business model? Which key resources are most expensive? Which key activities are most expensive? Which key channels are most expensive? Which key customer segments are most expensive? Which key revenue streams are most expensive?</p>		<p><b>Revenue Streams</b> </p> <p>For what value are our customers really willing to pay? For what do they really pay? How are they currently paying? How would they prefer to pay? How would they prefer to pay? How would they prefer to pay? How would they prefer to pay?</p>

www.businessmodelgeneration.com © 2010 Alexander Osterwalder & Yves Pigneur











## The Machine Learning Canvas (v0.4)

Designed for:

Designed by:

Date:

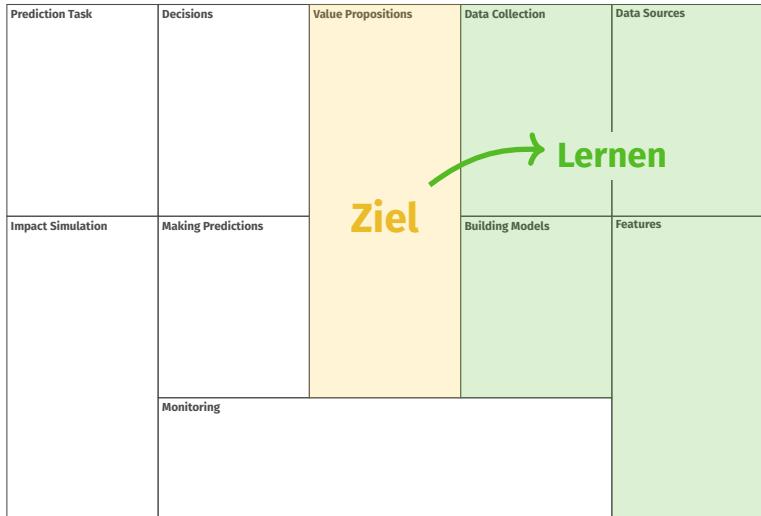
Iteration:

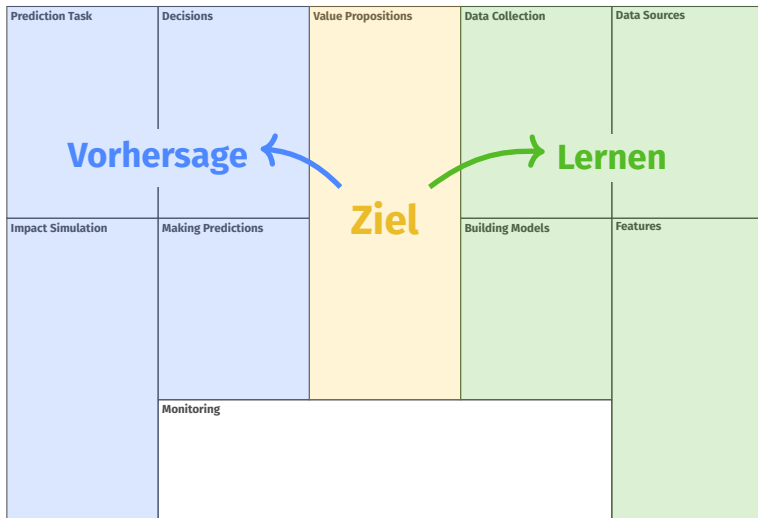
<b>Decisions</b> How are predictions used to make decisions that provide the proposed value to the end-user? 	<b>ML task</b> Input, output to predict, type of problem. 	<b>Value Propositions</b> What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving? 	<b>Data Sources</b> Which raw data sources can we use (internal and external)? 	<b>Collecting Data</b> How do we get new data to learn from (inputs and outputs)? 
<b>Making Predictions</b> When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction? 	<b>Offline Evaluation</b> Methods and metrics to evaluate the system before deployment. 		<b>Features</b> Input representations extracted from raw data sources. 	<b>Building Models</b> When do we create/update models with new training data? How long do we have to featurize training inputs and create a model? 
	<b>Live Evaluation and Monitoring</b> Methods and metrics to evaluate the system after deployment, and to quantify value creation. 			

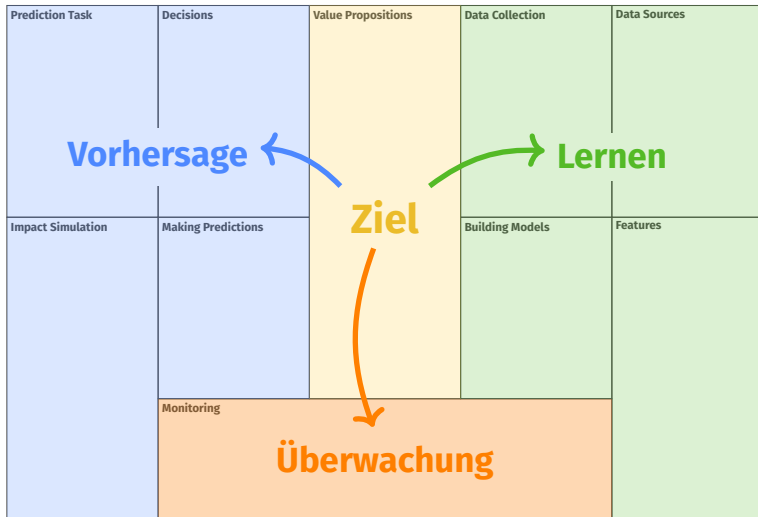


<b>Prediction Task</b>	<b>Decisions</b>	<b>Value Propositions</b>	<b>Data Collection</b>	<b>Data Sources</b>
<b>Impact Simulation</b>	<b>Making Predictions</b>		<b>Building Models</b>	<b>Features</b>
	<b>Monitoring</b>			









## Ziele

- Welche Ziele verfolgen wir mit der Analyse?
- Welchen Nutzen haben die Konsumenten der Analyse?

<b>Zu lösende Aufgabe</b>	<b>Skizze der Lösungsidee</b>	<b>Potentielle Kunden</b>	<b>Datenlage</b>
<b>Wie sieht bisherige Lösung aus?</b>	<b>Kennzahlen für Erfolg bzw. Misserfolg?</b>	<b>Potentielle Early Adopters?</b>	
<b>Potentielle Kosten?</b>		<b>Potentieller Nutzen?</b>	

[<https://www.eoda.de/wissen/blog/data-science-canvas/>]

# Projektphase



## Kaggle – DataScience als **Wettbewerb**



- Portal mit Data Science Challenges
- Jupyter Notebooks, Forum zum Austausch
- Für Challenges existiert Leaderboard
- Oft *accuracy* als Maß für Gewinner

## Kaggle Einstieg: Titanic Challenge

Getting Started Prediction Competition

### Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Kaggle · 43,384 teams · Ongoing

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#)

[My Submissions](#)

[Submit Predictions](#)

Overview

Description

Evaluation

Frequently Asked  
Questions

Ahoy, welcome to Kaggle! You're in the right place.

This is the legendary Titanic ML competition – the best, first challenge for you to dive into ML competitions and familiarize yourself with how the Kaggle platform works.

The competition is simple: use machine learning to create a model that predicts which passengers survived the Titanic shipwreck.

## Mini Projekt – 21.05. bis 28.05.

- Registrieren Sie sich bei [kaggle.com](https://www.kaggle.com)
- Melden Sie ein Team für ihre Gruppe an
- Probieren Sie die Titanic Challenge aus!
- Präsentieren Sie ihre Ergebnisse **am 28.05. um 9 Uhr** in der Vorlesung!

## Mini Projekt – 21.05. bis 28.05.

- Registrieren Sie sich bei [kaggle.com](https://www.kaggle.com)
- Melden Sie ein Team für ihre Gruppe an
- Probieren Sie die Titanic Challenge aus!
- Präsentieren Sie ihre Ergebnisse **am 28.05. um 9 Uhr** in der Vorlesung!

## Präsentation

- 2-3 Folien mit Ergebnis + Vorgehen

## Mini Projekt – 21.05. bis 28.05.

- Registrieren Sie sich bei [kaggle.com](https://www.kaggle.com)
- Melden Sie ein Team für ihre Gruppe an
- Probieren Sie die Titanic Challenge aus!
- Präsentieren Sie ihre Ergebnisse **am 28.05. um 9 Uhr** in der Vorlesung!

## Präsentation

- 2-3 Folien mit Ergebnis + Vorgehen

## Nächste Woche:

- Keine reguläre Vorlesung
- Gruppenarbeit von 9-12 Uhr im Raum möglich
- Möglichkeit für Fragen/Hilfestellung