

# DATA SCIENCE 2

CANVAS WORKSHOP

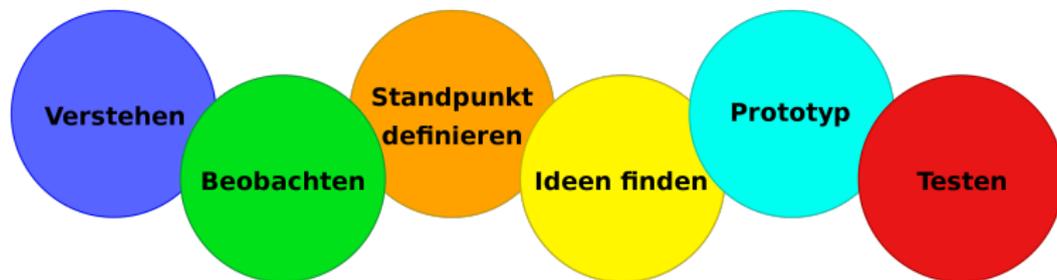
PROF. DR. CHRISTIAN BOCKERMANN

HOCHSCHULE BOCHUM

WINTERSEMESTER 2024/2025

- 1 Design Thinking
- 2 Von der Idee zur Analyse
- 3 Beispiel: Spritpreise

## 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.

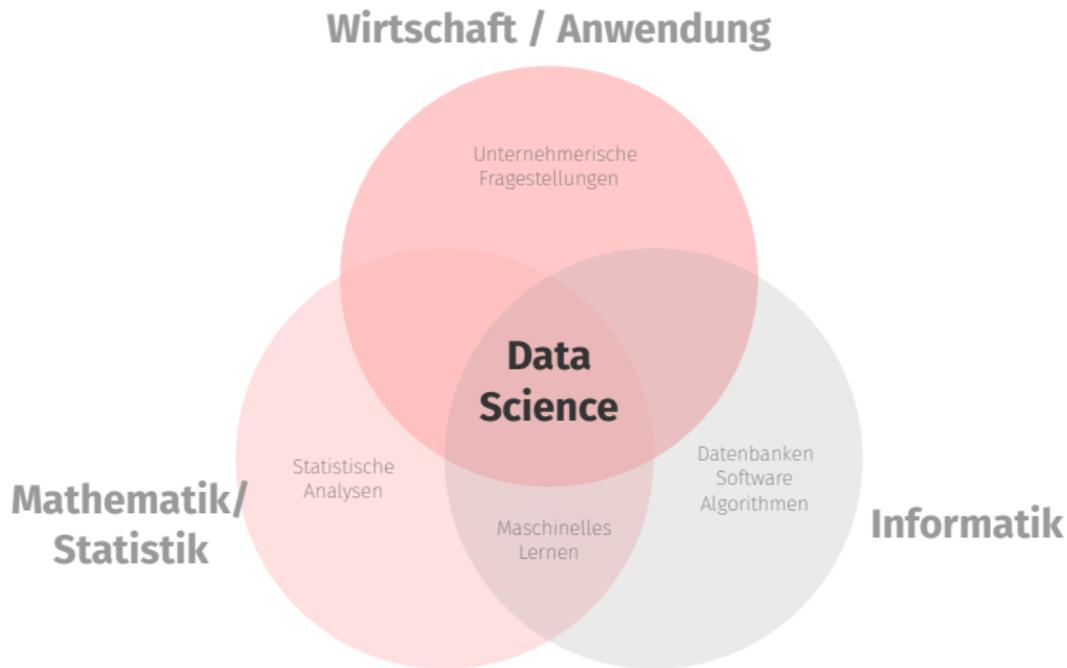
# Von der Idee zur Analyse

## Worum geht es im Kurs Data Science?

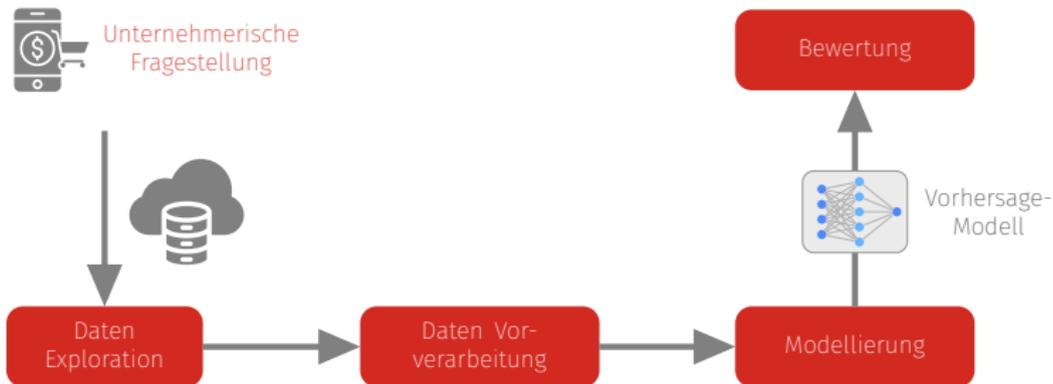
- Etablieren von datengetriebener Denk-/Arbeitsweise
- Kennenlernen von Methoden (ML)
- Wissenschaftliche Arbeitsweise
- Vorbereitung auf BA, Beruf

## 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/ein Thema.**

## Data Science 2

**Sie bekommen einen Datensatz/ein Thema.**

- Bewusst offene Aufgabenstellung

## Data Science 2

### **Sie bekommen einen Datensatz/ein Thema.**

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

**Sie haben einen Datensatz/ein Thema bekommen.**

**Und nun?**

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

## **Sie haben einen Datensatz/ein Thema bekommen.**

### **Und nun?**

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

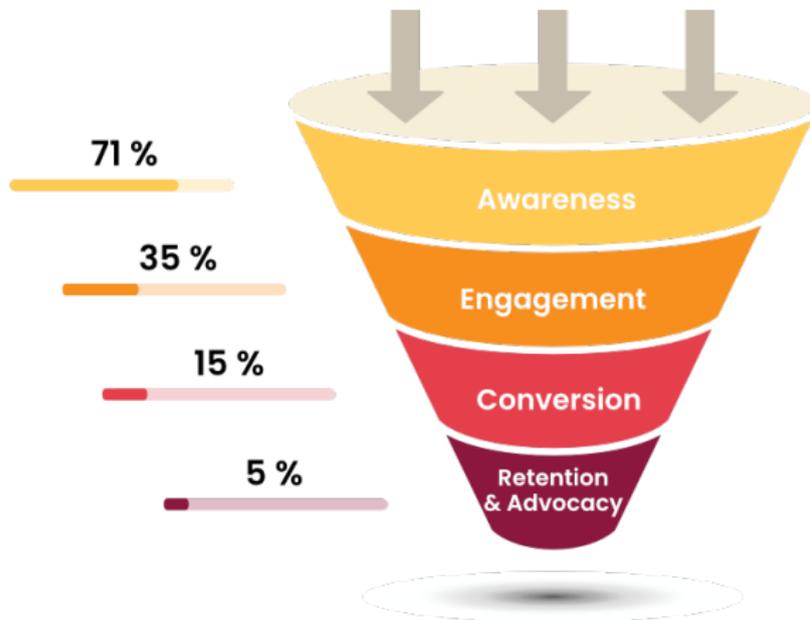
## Datensatz:

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

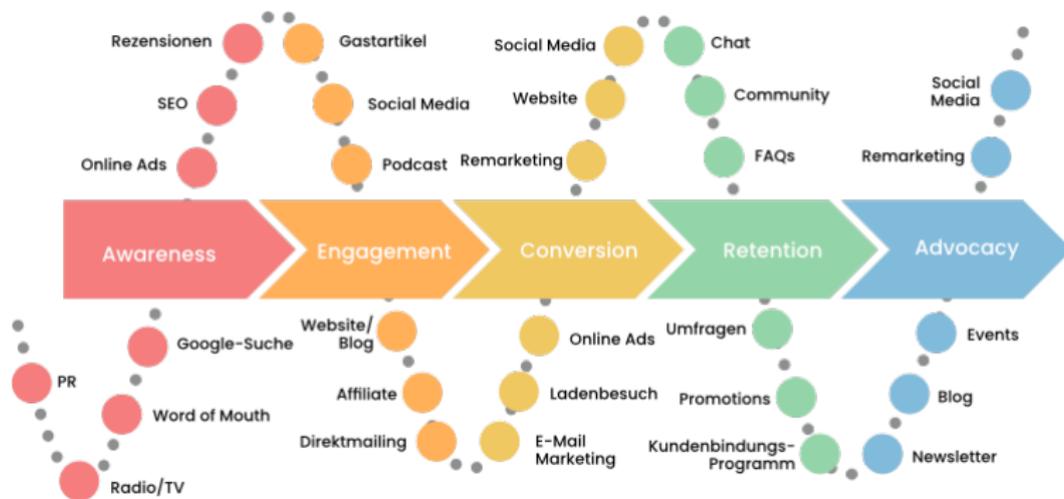
## Kontext verstehen

- 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: \_\_\_\_\_
Version: \_\_\_\_\_

<p><b>Key Partners</b></p> <p>Who are our partners? Who are the most important? What do they do for us? What do they expect from us? What do they expect from our partners? What do they expect from our customers? What do they expect from our suppliers? What do they expect from our competitors?</p>	<p><b>Key Activities</b></p> <p>What key activities do our Value Propositions require? What key activities do our Channels require? What key activities do our Customer Relationships require? What key activities do our Customer Segments require? What key activities do our Key Resources require? What key activities do our Revenue Streams require?</p>	<p><b>Value Propositions</b></p> <p>What value do we create for our customers? What value do our customers expect? What value do our customers want? What value do our customers need? What value do our customers want to avoid? What value do our customers want to be different from? What value do our customers want to be better than? What value do our customers want to be unique?</p>	<p><b>Customer Relationships</b></p> <p>What type of relationship do we want with our Customers? What type of relationship do our Customers want? What type of relationship do our Customers need? What type of relationship do our Customers want to avoid? What type of relationship do our Customers want to be different from? What type of relationship do our Customers want to be better than? What type of relationship do our Customers want to be unique?</p>	<p><b>Customer Segments</b></p> <p>For whom are we creating value? Who are our most important customers? Who are our most profitable customers? Who are our most loyal customers? Who are our most innovative customers? Who are our most influential customers? Who are our most vocal customers?</p>
<p><b>Key Resources</b></p> <p>What key resources do our Value Propositions require? What key resources do our Channels require? What key resources do our Customer Relationships require? What key resources do our Customer Segments require? What key resources do our Key Activities require? What key resources do our Revenue Streams require?</p>	<p><b>Channels</b></p> <p>Through which Channels do our Customers Segments want to be reached? How are we reaching them now? How can we reach them better? How can we reach them more efficiently? How can we reach them more effectively? How can we reach them more sustainably? How can we reach them more uniquely?</p>	<p><b>Cost Structure</b></p> <p>What are the most important costs incurred in our business model? What are the most important costs incurred in our business model? What are the most important costs incurred in our business model? What are the most important costs incurred in our business model? What are the most important costs incurred in our business model? What are the most important costs incurred in our business model? What are the most important costs incurred in our business model?</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 can they pay better? How can they pay more efficiently? How can they pay more effectively? How can they pay more sustainably? How can they pay more uniquely?</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 can they pay better? How can they pay more efficiently? How can they pay more effectively? How can they pay more sustainably? How can they pay more uniquely?</p>		

www.businessmodelgeneration.com

## The Machine Learning Canvas (v0.4)

Designed for:

Designed by:

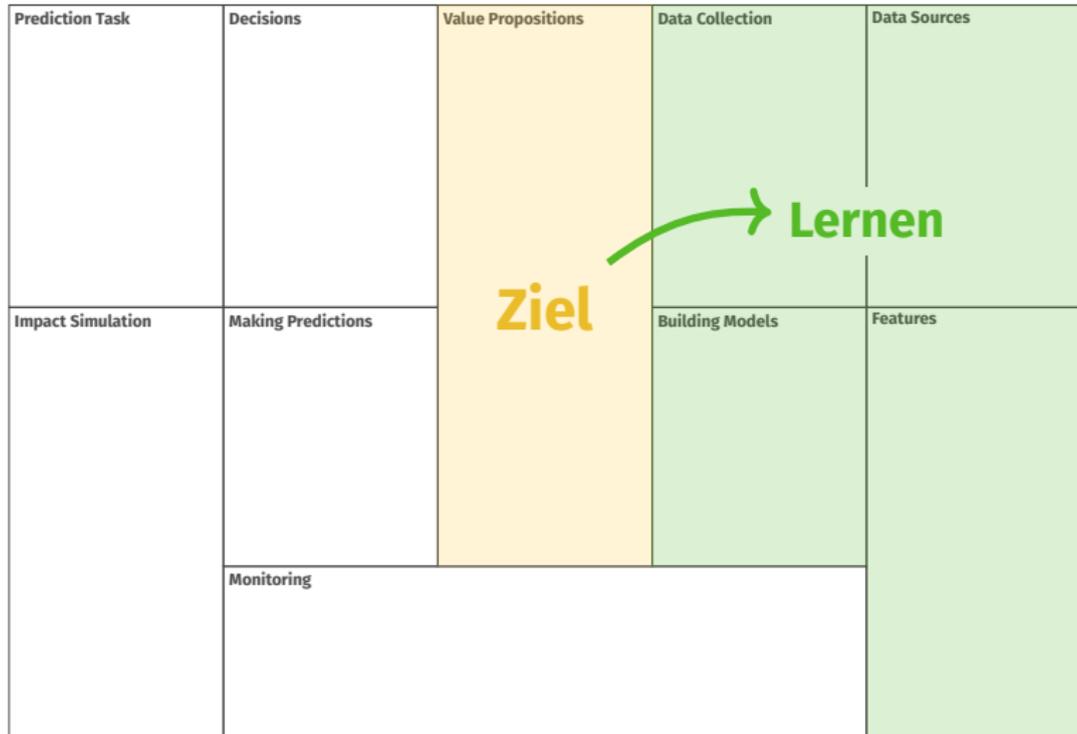
Date:

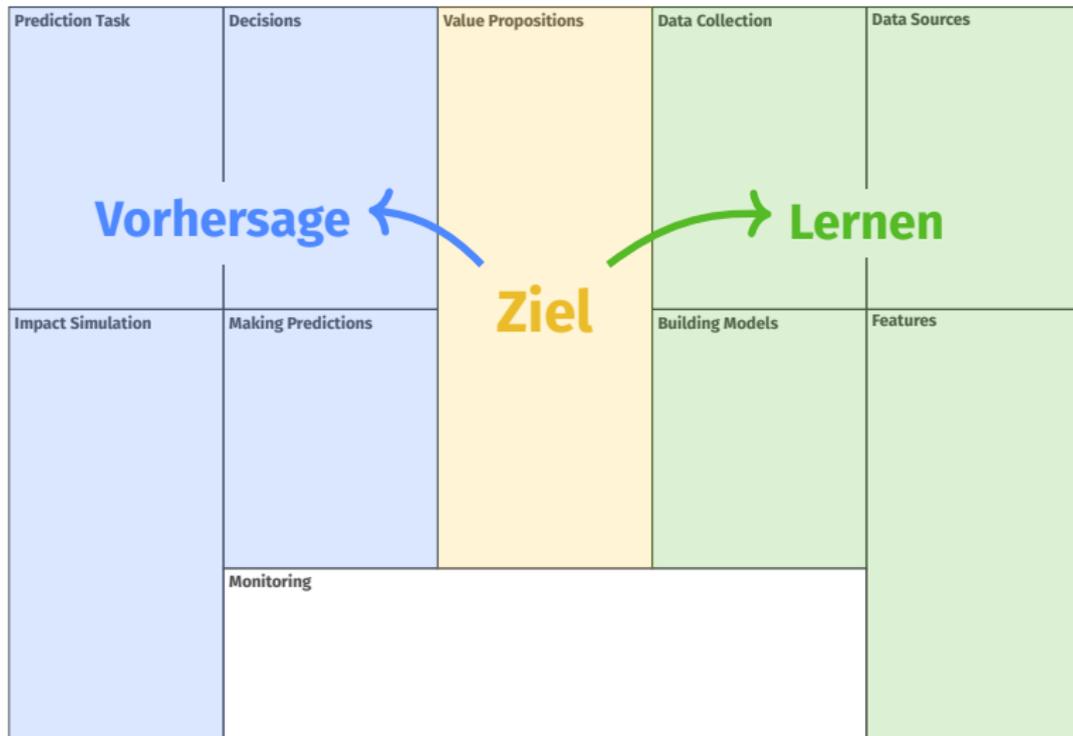
Iteration:

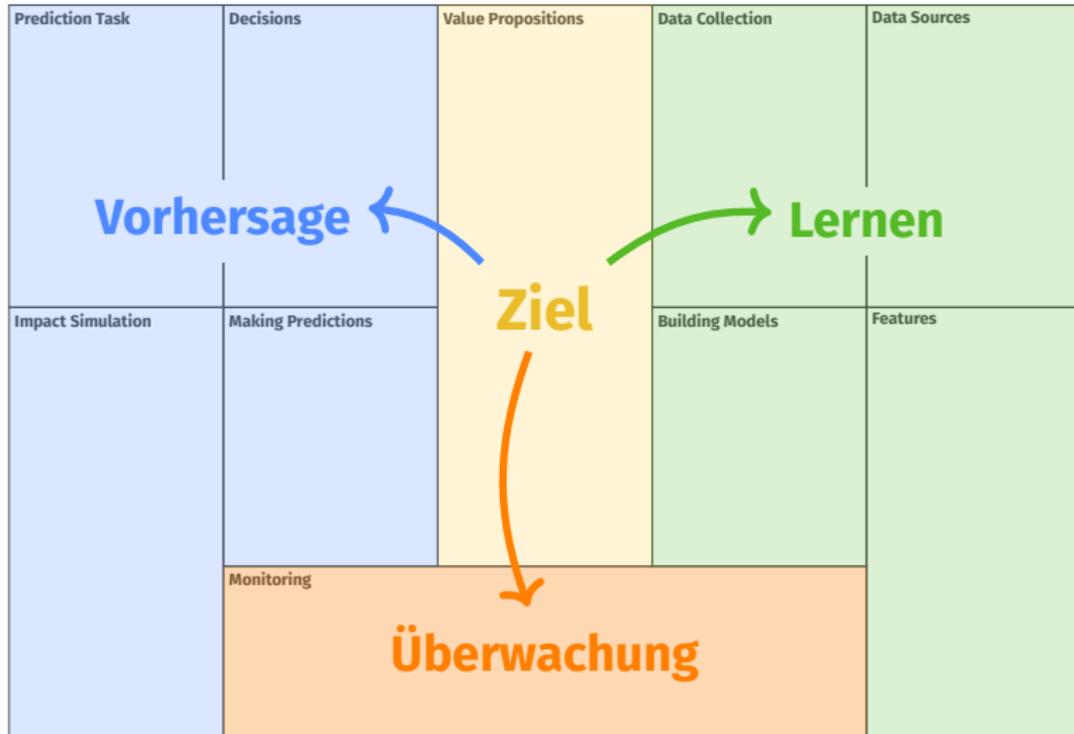
<p><b>Decisions</b> </p> <p>How are predictions used to make decisions that provide the proposed value to the end-user?</p>	<p><b>ML task</b> </p> <p>Input, output to predict, type of problem.</p>	<p><b>Value Propositions</b> </p> <p>What are we trying to do for the end-user(s) of the predictive system?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

<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?

# Beispiel: Spritpreise

## Analyse der Spritpreise

- Veröffentlichungspflicht der Tankstellen
- Spritpreise seit 2014



## Spritpreise für alle!