

DATA SCIENCE 1

VORLESUNG 3 – VON DER KENNZAHL ZUR MESSUNG

PROF. DR. CHRISTIAN BOCKERMANN

HOCHSCHULE BOCHUM

WINTERSEMESTER 2026/2027

IHRE RETOUREN-KPIs

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

Woher kommen die Unterschiede?

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

Woher kommen die Unterschiede?

- **Anzahl oder Wert?** Eine Teilretoure ist nicht dasselbe wie eine Komplettretoure.

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

Woher kommen die Unterschiede?

- **Anzahl oder Wert?** Eine Teilretoure ist nicht dasselbe wie eine Komplettretoure.
- **Welche Retouren zählen?** Ist ein Transportschaden eine Retoure?

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

Woher kommen die Unterschiede?

- **Anzahl oder Wert?** Eine Teilretoure ist nicht dasselbe wie eine Komplettretoure.
- **Welche Retouren zählen?** Ist ein Transportschaden eine Retoure?
- **Welche Bestellungen zählen?** B2C und B2B zusammen oder getrennt?

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

Woher kommen die Unterschiede?

- **Anzahl oder Wert?** Eine Teilretoure ist nicht dasselbe wie eine Komplettretoure.
- **Welche Retouren zählen?** Ist ein Transportschaden eine Retoure?
- **Welche Bestellungen zählen?** B2C und B2B zusammen oder getrennt?
- **Welcher Zeitraum?** Bestelldatum oder Eingangsdatum? Und: Sind alle Retouren schon da?

$$\text{Retourenquote} = \frac{\text{Anzahl ...}}{\text{Anzahl ...}}$$

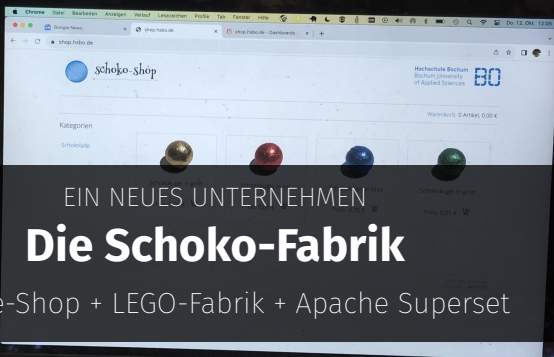
Woher kommen die Unterschiede?

- **Anzahl oder Wert?** Eine Teilretoure ist nicht dasselbe wie eine Komplettretoure.
- **Welche Retouren zählen?** Ist ein Transportschaden eine Retoure?
- **Welche Bestellungen zählen?** B2C und B2B zusammen oder getrennt?
- **Welcher Zeitraum?** Bestelldatum oder Eingangsdatum? Und: Sind alle Retouren schon da?

Eine KPI ohne Definition ist keine KPI.

Die Zahl hängt davon ab, was wir *festlegen*.

DIE SCHOKO-FABRIK



EIN NEUES UNTERNEHMEN

Die Schoko-Fabrik

Online-Shop + LEGO-Fabrik + Apache Superset



Warenkorb: 0 Artikel, 0,00 €

Kategorien

Schokolade



Schokokugel in gold

Preis: 0,05 € 



Schokokugel in rot

Preis: 0,05 € 



Schokokugel in blau

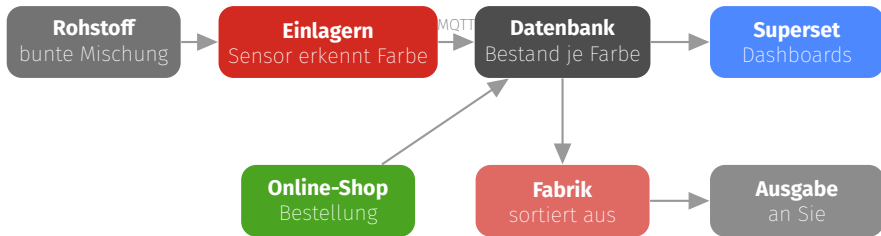
Preis: 0,05 € 



Schokokugel in grün

Preis: 0,05 € 

Vier Farben:  rot  blau  gold  grün



- Beim **Einlagern** erkennt ein Farbsensor jede Kugel – der Bestand wird aktualisiert.
- Jede **Bestellung** landet in der Datenbank, die Fabrik sortiert die Kugeln aus.
- Superset zeigt den **Bestand live**.

Die Fabrik kauft **gemischte** Schokokugeln ein – welche Farben drin sind, bestimmt der Zufall.

Die Fabrik kauft **gemischte** Schokokugeln ein – welche Farben drin sind, bestimmt der Zufall.

1. Wie viele Kugeln welcher Farbe stecken **im Schnitt** in einer Mischung?
2. Welche Farbe ist **am häufigsten**, welche am seltensten?
3. Welche Farben werden **am häufigsten bestellt**?

Die Fabrik kauft **gemischte** Schokokugeln ein – welche Farben drin sind, bestimmt der Zufall.

1. Wie viele Kugeln welcher Farbe stecken **im Schnitt** in einer Mischung?
2. Welche Farbe ist **am häufigsten**, welche am seltensten?
3. Welche Farben werden **am häufigsten bestellt**?

Und damit die eigentliche Entscheidung:

4. Sollten **seltene Farben teurer** sein?

Die Fabrik kauft **gemischte** Schokokugeln ein – welche Farben drin sind, bestimmt der Zufall.

1. Wie viele Kugeln welcher Farbe stecken **im Schnitt** in einer Mischung?
2. Welche Farbe ist **am häufigsten**, welche am seltensten?
3. Welche Farben werden **am häufigsten bestellt**?

Und damit die eigentliche Entscheidung:

4. Sollten **seltene Farben teurer** sein?

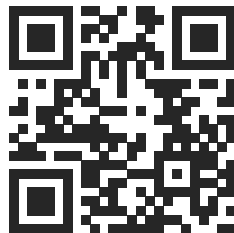
Angebot und Nachfrage – eine BWL-Frage, die wir mit Daten beantworten.
Mit **Metrik**, **Dimension** und **Filter** aus der letzten Vorlesung.

ÜBUNG: BESTELLEN UND AUSWERTEN

Der Schoko-Shop

shop.hsbo.de

- Bestellen Sie ein paar Kugeln.
- Geben Sie **nur einen Vornamen** oder ein **Pseudonym** an.
- Beobachten Sie in Superset, wie sich der **Bestand** verändert.



Alle anderen sehen Ihre Bestellung in Superset – überlegen Sie, was Sie angeben.

1 – Die Daten kennenlernen

- Welche Tabellen gibt es? Wie hängen sie zusammen?
- Was steht im Bestand, was in den Bestellungen?

2 – Angebot

- Wie verteilen sich die Farben im Bestand?
- Bauen Sie ein Diagramm, das sich live aktualisiert.

3 – Nachfrage

- Welche Farben werden wie oft bestellt?
- Passt die Nachfrage zum Angebot?

4 – Entscheiden

- Sollten seltene Farben teurer sein?
- Welche Kennzahl stützt Ihre Empfehlung?

WIE ZÄHLEN WIR DIE KUGELN?



**Woher kommt eigentlich die Zahl
„37 rote Kugeln“?**





- Erkennt der Sensor eine **goldene** Kugel als **rot**, ist der Bestand **falsch**.
- Dann ist auch die **Preisentscheidung** falsch – obwohl das Dashboard stimmt.



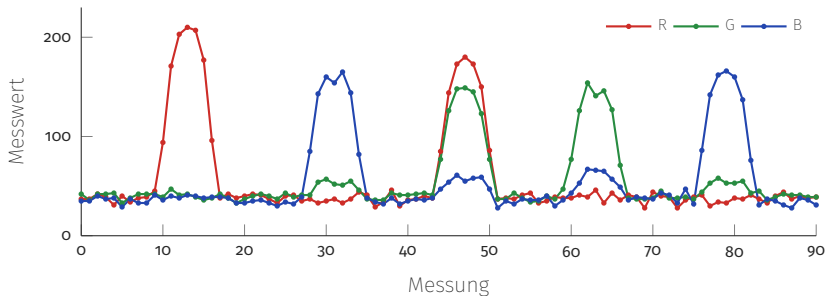
- Erkennt der Sensor eine **goldene** Kugel als **rot**, ist der Bestand **falsch**.
- Dann ist auch die **Preisentscheidung** falsch – obwohl das Dashboard stimmt.

Die Bestandszahl ist das Ergebnis einer Klassifikation.

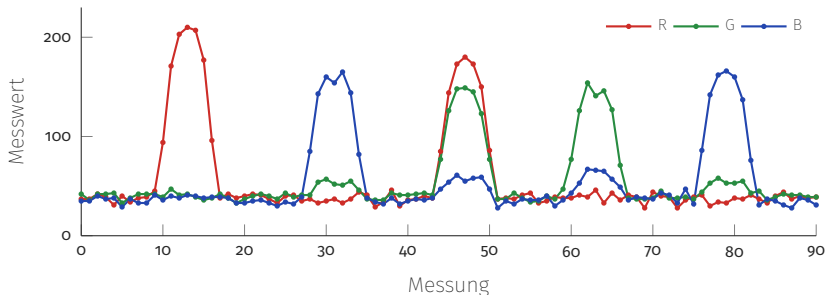
BI ist nur so gut wie die Daten, die vorne hereinfließen.

Der Farbsensor

Er liefert einen Strom von Messwerten: Rot, Grün, Blau (RGB) – per MQTT



schematisch



schematisch

Wo sind hier die Kugeln? Welche Farbe haben sie?

Chat: Wie viele Kugeln sehen Sie?

1. **Was misst der Sensor?**

Drei Zahlen pro Messung – und dazwischen das leere Band.

1. **Was misst der Sensor?**

Drei Zahlen pro Messung – und dazwischen das leere Band.

2. **Wann ist eine Kugel da?**

Eine Kugel erzeugt 5–10 Messungen. Wo fängt sie an, wo hört sie auf?

1. **Was misst der Sensor?**

Drei Zahlen pro Messung – und dazwischen das leere Band.

2. **Wann ist eine Kugel da?**

Eine Kugel erzeugt 5–10 Messungen. Wo fängt sie an, wo hört sie auf?

3. **Welche der Messungen nehmen wir?**

Die erste und letzte sind mit dem Hintergrund vermischt.

1. **Was misst der Sensor?**

Drei Zahlen pro Messung – und dazwischen das leere Band.

2. **Wann ist eine Kugel da?**

Eine Kugel erzeugt 5–10 Messungen. Wo fängt sie an, wo hört sie auf?

3. **Welche der Messungen nehmen wir?**

Die erste und letzte sind mit dem Hintergrund vermischt.

4. **Wie wird aus drei Zahlen eine Farbe?**

Und wer sagt der Maschine eigentlich, was „rot“ ist?

1. **Was misst der Sensor?**

Drei Zahlen pro Messung – und dazwischen das leere Band.

2. **Wann ist eine Kugel da?**

Eine Kugel erzeugt 5–10 Messungen. Wo fängt sie an, wo hört sie auf?

3. **Welche der Messungen nehmen wir?**

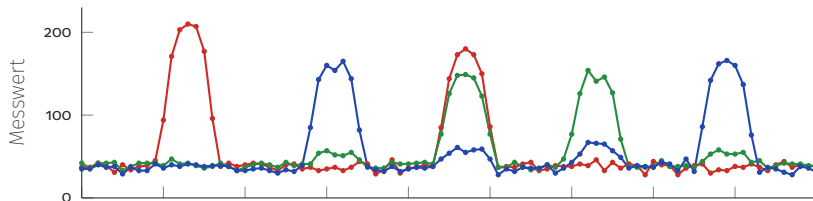
Die erste und letzte sind mit dem Hintergrund vermischt.

4. **Wie wird aus drei Zahlen eine Farbe?**

Und wer sagt der Maschine eigentlich, was „rot“ ist?

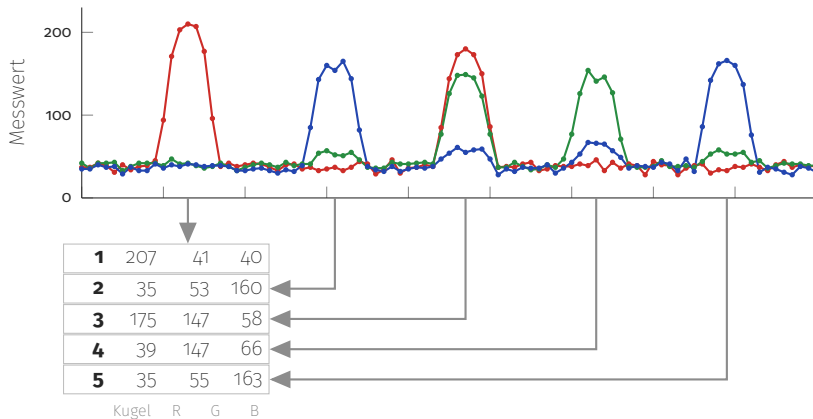
5. **Was kann schiefgehen?**

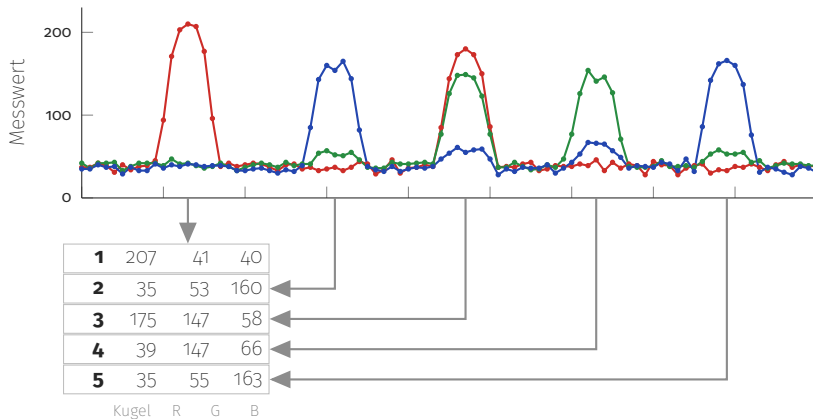
Licht, Abstand, Lage der Kugel, ähnliche Farben ...



1	207	41	40
2	35	53	160
3	175	147	58
4	39	147	66
5	35	55	163

Kugel R G B





Aus dem **Messstrom** wird ein **Datensatz**: eine Zeile pro Kugel.

Trainingsdaten, die von Hand erzeugt wurden.

R	G	B	I	label
58	13	20	75	 rot
11	30	21	44	 gruen
57	46	34	110	 gold
14	33	57	69	 blau
73	79	85	176	 fussball

...

208 Kugeln in fünf Klassen – jede Zeile ist ein **Beispiel**.

Trainingsdaten, die von Hand erzeugt wurden.

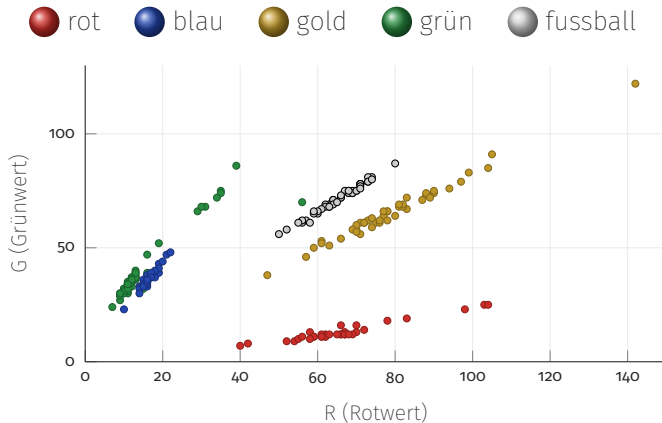
R	G	B	I	label
58	13	20	75	 rot
11	30	21	44	 gruen
57	46	34	110	 gold
14	33	57	69	 blau
73	79	85	176	 fussball

...

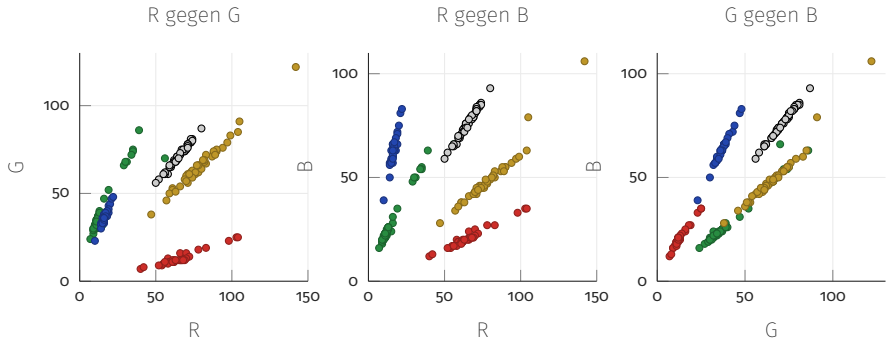
208 Kugeln in fünf Klassen – jede Zeile ist ein **Beispiel**.

Eine neue Kugel auf dem Band:

R	G	B	I	label
52	41	30	98	



Diskussion: Wo würden Sie die Grenzen ziehen?



- Welche Farben lassen sich in welcher Darstellung gut trennen – und welche nicht?
- Reichen zwei Werte, oder brauchen wir alle drei?

Aus Vorlesung 1 – jetzt am Beispiel der Kugeln:

Lernen

R	G	B	label
58	13	20	rot
11	30	21	gruen
14	33	57	blau
...			

Trainingsdaten

Lernverfahren

Modell
?

Aus Vorlesung 1 – jetzt am Beispiel der Kugeln:

Lernen

R	G	B	label
58	13	20	rot
11	30	21	gruen
14	33	57	blau
...			

Trainingsdaten

Lernverfahren

Modell
?

Anwenden

R	G	B	label
62	15	22	?

neue Kugel auf dem Band

Aus Vorlesung 1 – jetzt am Beispiel der Kugeln:

Lernen

R	G	B	label
58	13	20	rot
11	30	21	gruen
14	33	57	blau
...			

Trainingsdaten

Lernverfahren

Modell
?

Anwenden

R	G	B	label
62	15	22	?

neue Kugel auf dem Band

Vorhersage
„rot“

Wie würden **Sie** solche Trainingsdaten erzeugen?

- Woher wissen wir, welche Farbe eine Kugel **wirklich** hat?
- Wie viele Beispiele brauchen wir – und von welchen Farben?
- Was kann dabei schiefgehen?

Messwerte + bekannte Farbe = Beispiele zum Lernen.

Am 05.11. lernt daraus ein Entscheidungsbaum, Farben zu erkennen.



- Bei den Retouren entscheidet die **Definition** über die Zahl.
- In der Fabrik entscheidet die **Messung** über die Zahl.
- Hinter einer einfachen Kennzahl kann ein **Modell** stecken.

Der Datensatz: `kugeln_hausaufgabe.csv` – 205 Kugeln mit R, G, B, I und `label`

Ihre Aufgabe: Welche Farbe haben diese drei Kugeln?

Kugel	R	G	B	I	label
A	67	13	21	84	?
B	29	66	48	104	?
C	50	56	59	122	?

- **Wie** Sie vorgehen, entscheiden Sie selbst. Auch das **Werkzeug** ist frei: Excel, Pandas, Diagramme, Papier und Bleistift ...
- Halten Sie fest, **wie** Sie zu Ihrer Entscheidung gekommen sind.
- Wie sicher sind Sie sich – bei jeder der drei Kugeln?

Einstieg mit Pandas: `df = pd.read_csv("kugeln_hausaufgabe.csv")`