

WIRTSCHAFTSINFORMATIK 2

DATENVERARBEITUNG MIT PYTHON

PROF. DR. BERND BLÜMEL, PROF. DR. CHRISTIAN
BOCKERMANN, PROF. DR. VOLKER KLINGSPOR

HOCHSCHULE BOCHUM

SOMMERSEMESTER 2025

Was Sie letzte Woche gelernt haben sollten:

- Wie werden Tabellen in Pandas dargestellt?
- Was ist das CSV-Datenformat?
- Wie werden Daten in Pandas eingelesen?

Pandas DataFrame

```
import pandas as pd  
order = pd.read_csv('https://data.hsbo.de/order.csv');
```

```
-----  
HTTPError                                Traceback (most recent call last)  
Cell In[17], line 1  
----> 1 pd.read_csv('https://data.hsbo.de/order.csv')
```

Pandas DataFrame

```
import pandas as pd  
order = pd.read_csv('https://data.hsbo.de/order.csv');
```

```
-----  
HTTPError                                Traceback (most recent call last)  
Cell In[17], line 1  
----> 1 pd.read_csv('https://data.hsbo.de/order.csv')
```

```
File /opt/conda/lib/python3.11/urllib/request.py:643, in HTTPDefaultErrorHandler.h  
p, code, msg, hdrs)  
    642 def http_error_default(self, req, fp, code, msg, hdrs):  
--> 643     raise HTTPError(req.full_url, code, msg, hdrs, fp)
```

```
HTTPError: HTTP Error 404: Not Found
```

Pandas DataFrame

Es wird die Datei **orders.csv** mit dem nachfolgenden Code eingelesen und ergibt die unten dargestellte Tabelle:

```
import pandas as pd
orders = pd.read_csv('https://data.hsbo.de/orders.csv')
```

	Datum;BestellungNr;KundeNr;Artikel_Gold;Artikel_Rot;Artikel_Blau;Artikel_Gruen;Betrag
0	2023-10-12;77;78;0;0;0;0;0.05
1	2023-10-12;78;79;0;0;0;0;0.15
2	2023-10-12;79;80;1;0;0;0;0.15
3	2023-10-12;80;81;0;0;0;0;0.15
4	2023-10-12;81;82;0;0;0;0;0.15

Pandas DataFrame

Es wird die Datei **orders.csv** mit dem nachfolgenden Code eingelesen und ergibt die unten dargestellte Tabelle:

```
import pandas as pd
orders = pd.read_csv('https://data.hsbo.de/orders.csv')
```

	Datum;BestellungNr;KundeNr;Artikel_Gold;Artikel_Rot;Artikel_Blau;Artikel_Gruen;Betrag
0	2023-10-12;77;78;0;0;0;0;0.05
1	2023-10-12;78;79;0;0;0;0;0.15
2	2023-10-12;79;80;1;0;0;0;0.15
3	2023-10-12;80;81;0;0;0;0;0.15
4	2023-10-12;81;82;0;0;0;0;0.15

Was fällt Ihnen auf?



Warenkorb: 0 Artikel, 0,00 €

Kategorien

Schokolade

Rübezahl | Schokolade



Schokokugel in gold

Preis: 0,05 € 🛒

Rübezahl | Schokolade



Schokokugel in rot

Preis: 0,05 € 🛒

Rübezahl | Schokolade



Schokokugel in blau

Artikel aktuell leider nicht
verfügbar.

Rübezahl | Schokolade

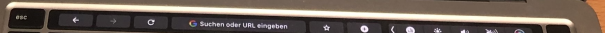
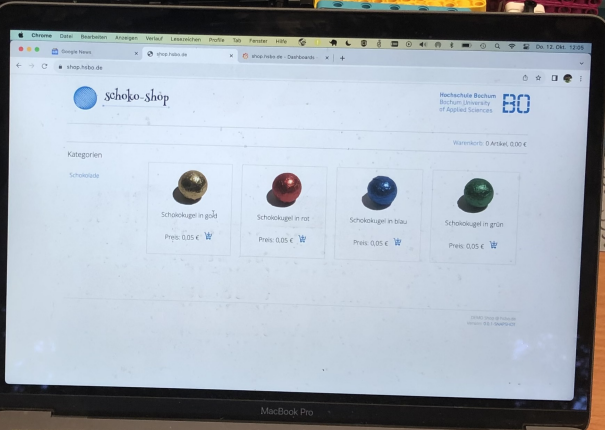
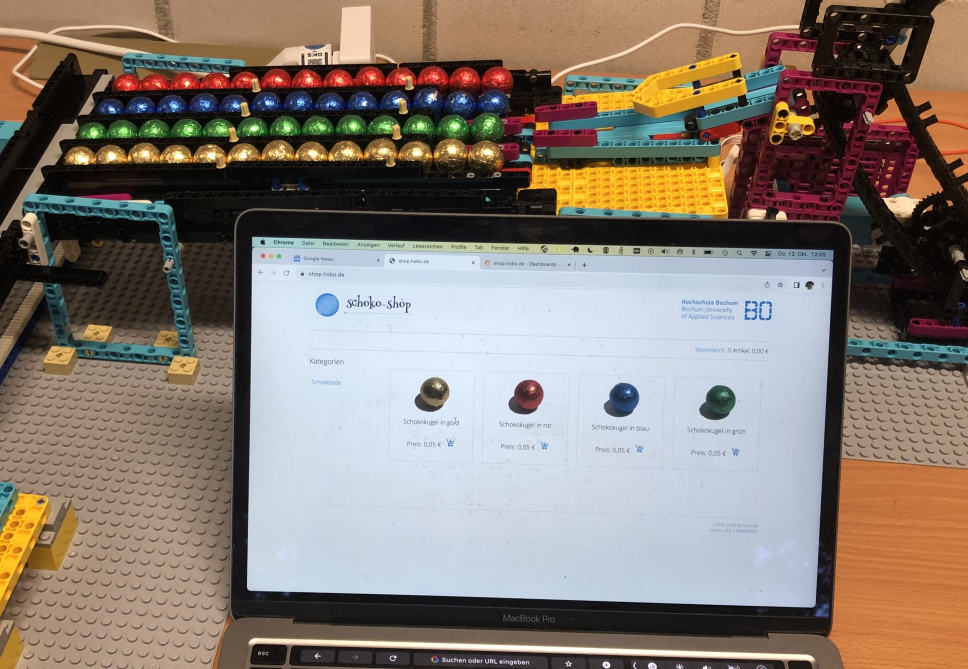


Schokokugel in grün

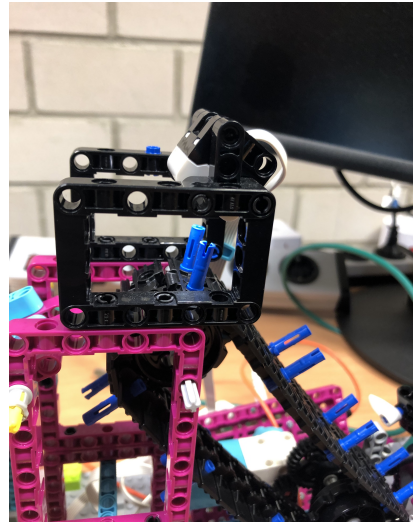
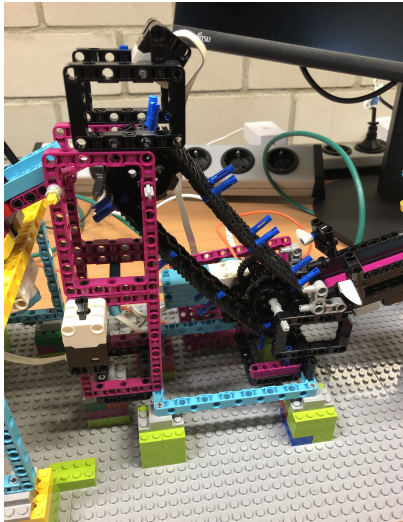
Preis: 0,05 € 🛒

Shop-Daten

shop.hsbo.de
Version: 0.0.1

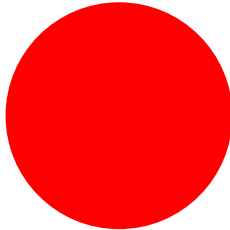


Klassifikation von Schoko-Kugeln



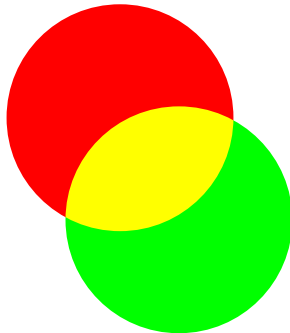
Rückblick (Winf1, Vorlesung 1): Farben im Computer

Gängiges Farb-Schema: RGB Farben (red/green/blue)



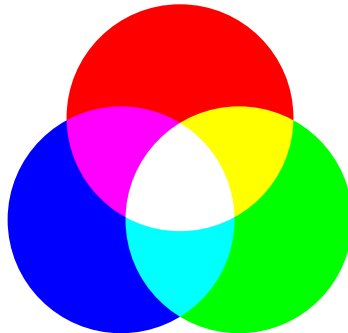
Rückblick (Winf1, Vorlesung 1): Farben im Computer

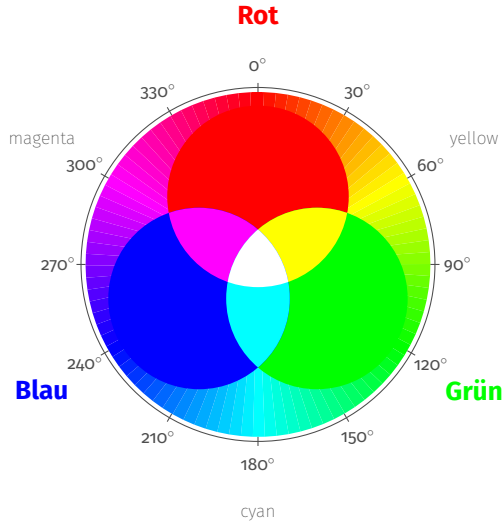
Gängiges Farb-Schema: RGB Farben (red/green/blue)



Rückblick (Winf1, Vorlesung 1): Farben im Computer

Gängiges Farb-Schema: RGB Farben (red/green/blue)

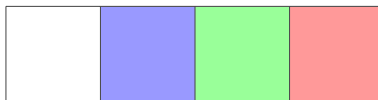




Farben im Computer

Datentyp Long belegt 4 Byte Speicherplatz im PC:

Grün-Anteil



$$b \cdot 256^2 + g \cdot 256^1 + r \cdot 256^0 = \text{Farbwert}$$

Blau-Anteil

Rot-Anteil

LEGO Farbsensor

- Vordefinierte Farben erkennen
- RGB Farbwerte auslesen



```
senor = ColorSensor('B')  
  
# Farbe als String auslesen  
(red,green,blue,intense) = sensor.get_rgb_intensity()  
  
# intense ist der gemessene Helligkeitswert
```

Klassifikation von Schoko-Kugeln

Sensorische Messung x



RGB Farbwerte

38,43 % Rot
9,01 % Grün
12,94% Blau

Sensor Daten als Beispiel

```
import pandas as pd  
df = pd.read_csv('https://data.hsbo.de/kugeln.csv')
```

	ID	R	G	B	I	label
0	564	16	39	28	59	gruen
1	708	22	48	83	100	blau
2	516	40	7	12	50	rot
3	527	62	11	19	77	rot
4	715	17	37	63	79	blau

Wie sind die durchschnittlichen R/G/B Werte für blaue Kugeln?

Heutige Übung

- Übungsaufgaben mit Hilfe von IBIX
- Bonuspunkt-Aufgaben **ab Donnerstag** verfügbar

Heutige Übung

- Übungsaufgaben mit Hilfe von IBIX
- Bonuspunkt-Aufgaben **ab Donnerstag** verfügbar

Selbststudium nächste Woche

- Thema: *Tabellen mit Pandas (DataFrame) Teil 2*
- Foliensatz *Pandas DataFrame 2* (datascience.hs-bochum.de)
- Fragen gerne bis Dienstag per Mail an den Dozenten

Wochenaufgabe

- Verfügbar auf

<https://datascience.hs-bochum.de>
(Vorlesungen -> Semester -> Wirtschaftsinformatik 2)