

DATA SCIENCE 1

TUTORIAL DAY 2

PROF. DR. CHRISTIAN BOCKERMANN

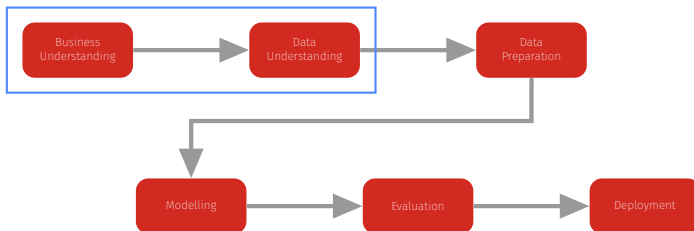
HOCHSCHULE BOCHUM

WINTERSEMESTER 2025/2026

Ziel des Tutorial Day 2

In der digitalen Kompetenzwoche ging es um das Thema Python und Pandas. Pandas ist das recht wichtige Python Modul, mit dem sich tabellarische Daten einlesen, verarbeiten und analysieren lassen.

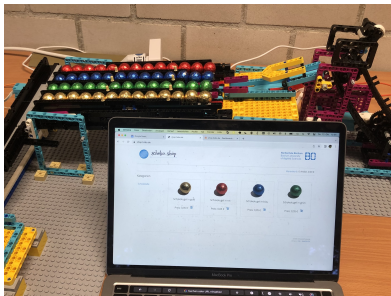
Zwei wichtige Schritte in der Datenanalyse sind das *Business Understanding* und das *Data Understanding*:



In diesem Tutorial geht es darum, sich mit einem Anwendungsfall zu befassen, für den Daten analysiert, visualisiert und verstanden werden müssen.

Anwendungsfall: Online Shop

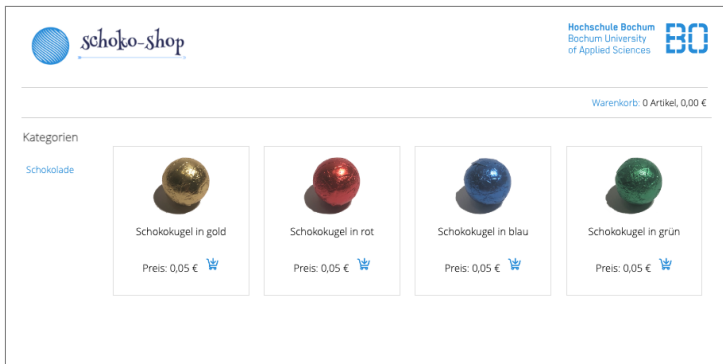
Es geht in diesem Anwendungsfall um den Schoko-Kugel Shop, den wir als LEGO-Modell am Fachbereich Wirtschaft als Demonstrator betreiben. Über diesen Online-Shop können Schokokugeln in verschiedenen Farben bestellt werden. Die Ware (Kugeln) wird in gemischten Beuteln angeliefert und über eine automatisierte Sortieranlage nach Farbe sortiert. Ein Kommissionsroboter arbeitet dann die Bestellungen und gibt die entsprechenden Kugeln aus.



Links: Die Sortieranlage mit den Silos/Rutschen für die jeweiligen Farben der Schokokugeln.

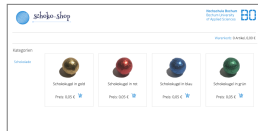
Teil 1: Der Online Shop

Der Online Shop → <https://shop.hsbo.de>



Der Online Shop ist einem realen Shop nachempfunden und nimmt Bestellungen entgegen, die dann an die anderen Systeme (Auslieferung) weitergereicht werden.

Datenfluss im Online-Shop

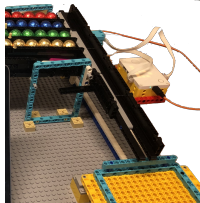


Datenbank

Über einen Broker (Vermittler) werden die Daten einer Bestellung an die Auslieferung und ein Dashboard weitergeleitet.

Kanal: `shop/orders`

Message Broker



Dashboard



Welche Daten sind wichtig?

Für diesen Anwendungsfall kann man sich natürlich zunächst mal fragen, welche Daten genau erforderlich sind. Das wären z.B.:

für **Bestellabwicklung, Controlling**

- Bestelldaten (Name, Adresse, Produkte)
- Umsatzdaten (Preis, Anzahl)

für **Marketing, Prognosen**

- Mögen KiTA-Kinder lieber rote Kugeln?
- Wann kauft welche Altersgruppe am häufigsten ein?
- Welche Produkte sind unbeliebt? (Preis-Anpassung?)

Tutorial Daten und Aufgaben

Die Daten der Bestellungen aus dem Online-Shop sind für Sie verfügbar. Die Daten befinden sich in der Datei

`https://data.hsbo.de/bestellungen.csv`

Bei ihrer Bestellung können die Kunden auch zusätzliche Informationen (Schulkind/Studierende/Mitarbeiter/Gast) und die Altersgruppe angeben. Auch diese Daten sind verfügbar:

`https://data.hsbo.de/kunden.csv`

Auf dem Data Science Server finden Sie zudem auch eine Jupyter-Notebook Datei mit den Tutorial Aufgaben:

`https://datascience.hs-bochum.de`