

# Datenmanipulation & Datenanalyse mit Python

Dr. Simon Schölzel

(updated: 25. März 2024)

---

1

Einführung und Motivation

2

Grundlagen der Programmierung mit Python

3

Datenmanipulation und Datenanalyse mit `pandas`

4

Ausblick

1	Einführung und Motivation
2	Grundlagen der Programmierung mit Python
3	Datenmanipulation und Datenanalyse mit <code>pandas</code>
4	Ausblick

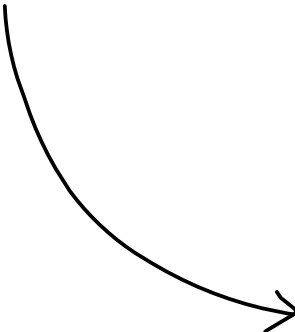
# 1 Einführung und Motivation

## 1.1 Einordnung der Vorlesung in den Kontext der Data Literacy

Ridsdatel et al. (2015), S. 2

„Data Literacy is the ability to collect, **manage, evaluate**, and apply data, in a critical manner.“

### Kompetenzmatrix „Data Literacy“



Daten- Management	Datenorganisation
	Datenmanipulation
	Datenkonvertierung
	Metadatenerzeugung und -verwendung
	Datenheilung, -sicherheit und -wiederverwendung
Datenevaluation	Datenaufbewahrung
	Datenwerkzeuge
	Grundlegende Datenanalyse
	Dateninterpretation (Datenverständnis)
	Nutzen von Daten zur Identifizierung von Problemen
	Datenvisualisierung
	Datenpräsentation (verbal)
	Datengetriebene Entscheidungsfindung

# 1 Einführung und Motivation

## 1.1 „Data Scientist: The Sexiest Job of the 21st Century“

# Data Scientist:

## *The Sexiest Job of the 21st Century*

**Meet the people who  
can coax treasure out of  
messy, unstructured data.**

by Thomas H. Davenport  
and D.J. Patil

**W**hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review October 2012

Quelle: [Harvard Business Review](#)

“It’s a **high-ranking professional** with the training and **curiosity** to make discoveries in the world of **big data**.”

“What abilities make a data scientist successful? Think of him or her as a **hybrid of data hacker, analyst, communicator, and trusted adviser**. The combination is extremely powerful—and **rare**.”

“But we would say the dominant trait among data scientists is an **intense curiosity**—a desire to go beneath the surface of a problem, **find the questions at its heart, and distill them into a very clear set of hypotheses that can be tested**.”

“On this front, demand has raced ahead of supply. Indeed, the shortage of data scientists is becoming a serious constraint in some sectors.”

# 1 Einführung und Motivation

## 1.1 „Data Scientist: The Sexiest Job of the 21st Century“

### 1. Data Gathering, Preparation, and Exploration

Sammlung von strukturierten und unstrukturierten Daten  
Datenaufbereitung (DATA CLEANING oder DATA WRANGLING)  
Explorative Datenanalyse (EDA)

### 2. Data Representation and Transformation

Restrukturierung von Daten  
Vertrautheit mit Datenbanken und Dateiformaten  
Verarbeitung verschiedener Datenmodalitäten

### 3. Computing with Data

Anwendung von Programmiersprachen (z.B. R, Python)  
Effiziente Programmierung „makes or breaks a product“  
Cloud, Edge, and Distributed Computing

### 4. Data Visualization and Presentation

Data Storytelling und Plot Engineering  
Implementierung von Dashboards  
Erstellung eines Analysereports für die Stakeholder

### 5. Data Modeling

Explanatory Modeling (traditionelle Statistik/Inferenz)  
Predictive Modeling (Machine Learning/Prediction)

### 6. Science about Data Science

Akademische Forschung im Bereich Data Science  
Weiterentwicklung von Open Source Tools  
Etablierung von branchenweiten Best Practices

Quelle: [Donoho \(2017\), The Full Scope of Data Science](#)

# 1 Einführung und Motivation

## 1.1 „Data Scientist: The Sexiest Job of the 21st Century“

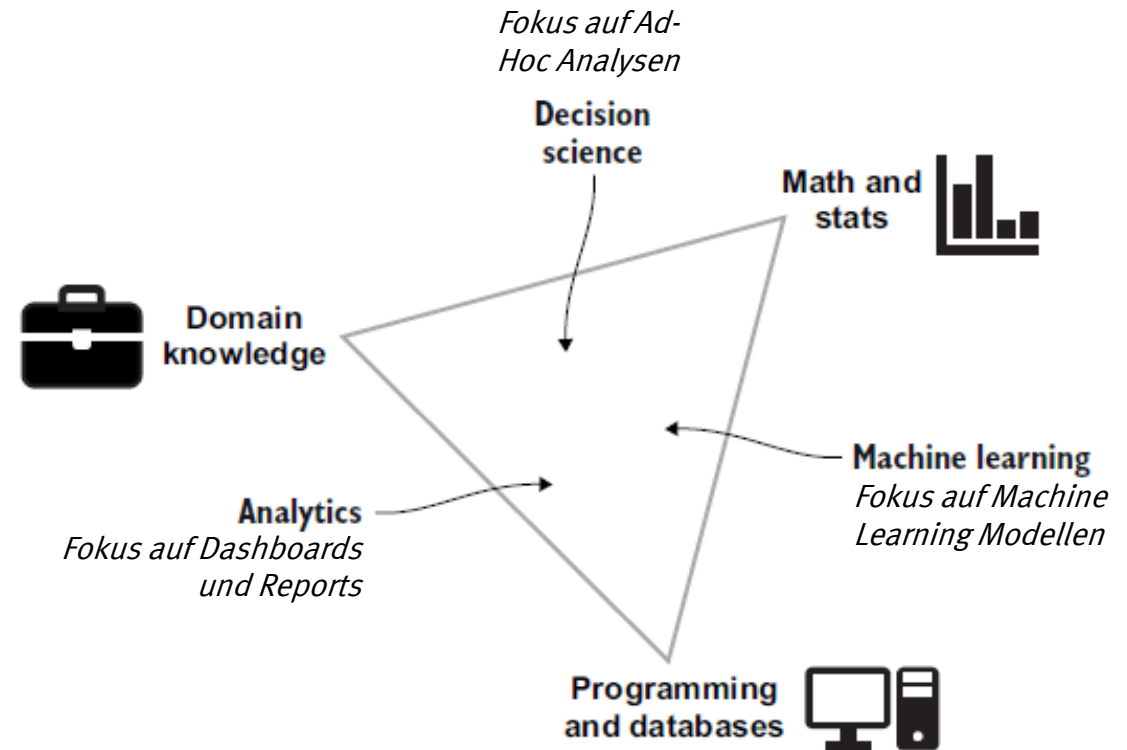
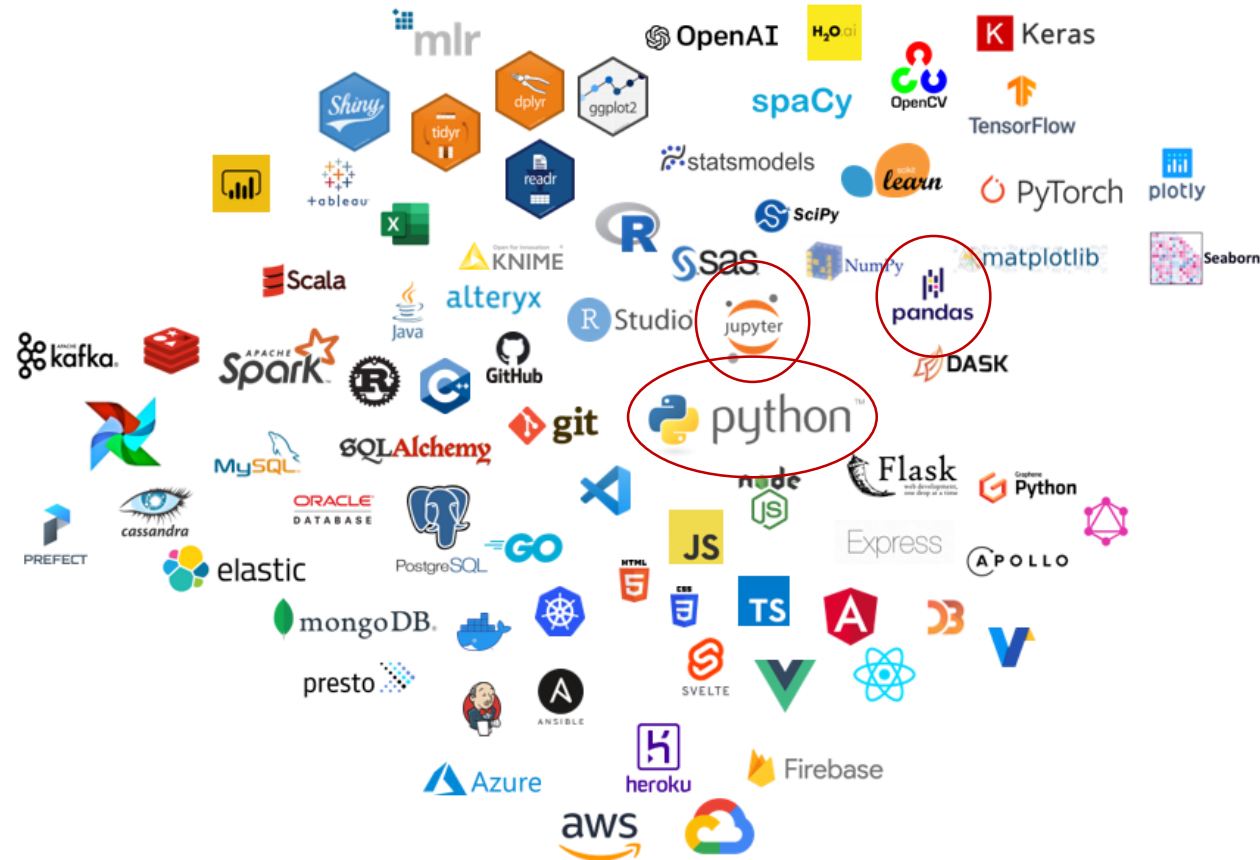


Figure 1.1 The skills that combine to make data science and how they combine to make different roles

Quelle: [Robinson/Nolis \(2020\), Build a Career in Data Science](#)

# 1 Einführung und Motivation

## 1.2 Data Science Tech Stack

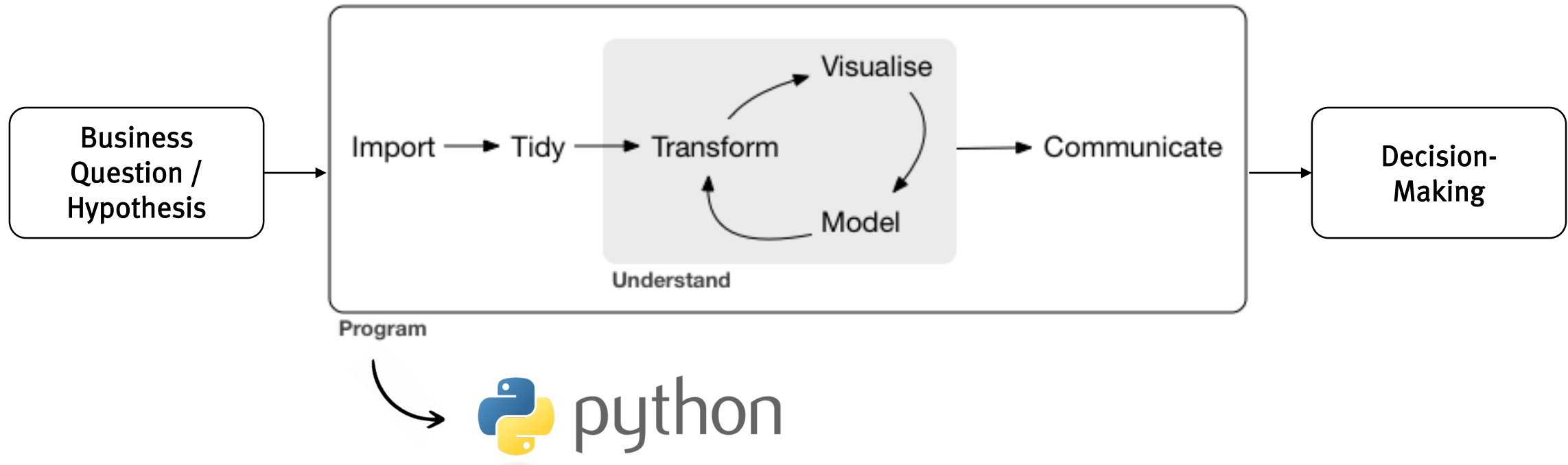




# 1 Einführung und Motivation

## 1.3 Der Typischer Data Science Workflow

Quelle: [Wickham/Grolemund \(2021\)](#)



- » **Reflexion über Daten:** Viele praxisrelevanten Fragestellungen erfordern Daten, um eine Evidenz-basierte Antwort liefern zu können. Kenntnisse einer Programmiersprache erleichtern es, frühzeitig und strukturiert über die benötigten Daten sowie Datenmodelle reflektieren.
- » **Tooling:** Zur Durchführung der Datenanalyse braucht es das richtige Rüstzeug. Python ist eine flexible Programmiersprache, die Funktionen für die Datenbeschaffung, -bereinigung, -manipulation, -analyse, und vieles mehr mitbringt.
- » **Praxisrelevanz:** Python ist Open Source, sehr beliebt in der IT-Community und weit verbreitet in der Praxis. Kompetenzen im Umgang mit Python stellen einen klaren „Wettbewerbsvorteil“ für BWL/VWL-Absolvent:innen im 21. Jahrhundert dar.
- » **Datenqualität in der VUCA-Welt:** VUCA führt dazu, dass reale Daten häufig „noisy“/„dirty“ sind. Kompetenzen im Bereich Datenmanipulation und -analyse erlauben den autonomen Umgang mit Daten.
- » **Vorbereitung auf eine empirische Abschlussarbeit:** Durch die Kenntnis einer Programmiersprache sind die Weichen gestellt, um erfolgreich und eigenständig eine empirische Abschlussarbeit schreiben zu können.
- » Spaß an der Arbeit mit Daten!