

Berufsperspektiven Data Professional

Universität Zürich

28. Juni 2021

Klaus Glückert

Agenda

Overview

- > Problemstellung
- > Mega-Trends
- > Profile
- > Kompetenzen
- > Jobeinstieg
- > Bewerbung

Data-Scientist - Sexiest job of 21st century?

Problemstellung

3

Data

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

From the Magazine (October 2012)



Artwork: Tamar Cohen, Andrew J. Buboltz, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

**Artificial
Intelligence**

**Data
Science**

Drei Kräfte treiben die Datenrevolution an

Mega-Trends

5



1. data explosion



2. infrastructure



3. research

Data-Innovationen werden von Tech-Firmen gekauft

Mega-Trends

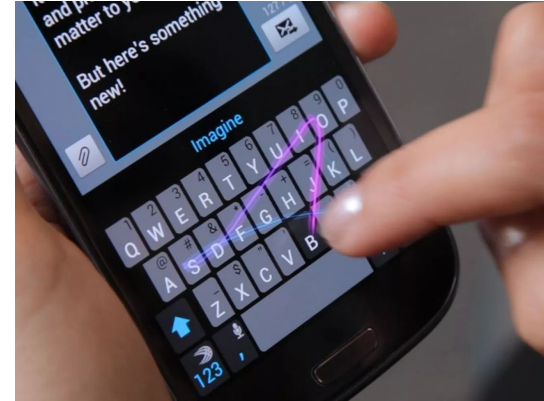
6



deep learning imagines computer game textures or invents new pixels to transform low resolution pictures to high quality



shadow-detection algorithms predict oil tank volumes for hedge funds



deep learning auto-corrects and predicts what you are going to write

TACHYUS

textio

ROSS

SIGOPT

Data-Innovationen werden von Tech-Firmen gekauft

Mega-Trends

7

Customer Service Manager

Job post for a Customer service role in Chicago

Our **passionate** team is hiring a **competitive** and **results-driven** customer service manager. We're a fun, **fast-paced** company, but we always **work hard**. As **our team** is rapidly **expanding**, we are looking for a **forward-thinking** leader. This is not just a job, it will be more than just a cog in the system. We want someone who is a **phenomenal** customer rep, a **proven** team player, and a **huge** plus.

You could attract more women to apply by changing your language.

Instead, you could try:

inspired by

Do you have a great **sense of humor**? Are you **driven by** the ability to set and exceed **results-oriented** goals? Do you do your best **under pressure** with tight **deadlines**? When you see something that's a problem, do you make a change or just complain? If so, this might be the place for you.

37

Textio Score
Below Average

Slightly masculine tone



Appeals to older people



20s 30s 40s 50s 60s

- Contains too many questions
- Sentences are too long
- Uses fixed mindset language

Textio is currently comparing your writing to 25,152 recent Customer service job posts in Illinois.

MACHINE INTELLIGENCE 3.0

ENTERPRISE INTELLIGENCE

VISUAL

- Orbital Insight
- planeta
- clarifai
- DEEPIVISION
- corfica
- ligoclan
- SPACE_KNOW
- Captricity
- netra
- deepomatic

AUDIO

- Gridspace
- TalkIQ
- nexidia
- twilio
- CAPIO
- Expect Labs
- Clover
- Mobvoi
- QuiriousAI
- popUP archive

SENSOR

- PREDIX
- IoT
- MAANA
- Sentena
- PLANET OS
- UPTAKE
- IMUBIT
- Radarsense
- thingworx
- KONUX
- Alluvium

INTERNAL DATA

- PRIMER
- IBM WATSON
- Gyrforg
- Palantir
- ARIMO
- Alation
- Saphro
- Outlier
- Digital Reasoning

MARKET

- mattermark
- Quid
- Datafox
- PREMISE
- Bottlenose
- MOTIVA
- enigma
- CB INSIGHTS
- Tracxn
- predata

ENTERPRISE FUNCTIONS

CUSTOMER SUPPORT

- DigitalGenius
- Kasisto
- ELOQUENT
- Wiseio
- ACTIONIQ
- zendesk
- Praet
- CLARABRIDGE

SALES

- collective
- sense
- fuse
- machines
- AVISO
- salesforce
- INSIDE SALES
- clari
- Zensight
- COM

MARKETING

- MINTIGO
- Lattice
- RADIUS
- Liftniter
- PERSADO
- brightfunnel
- retention
- COGNICOR
- AIRPR
- msgd

SECURITY

- CYCLANCE
- DARKTRACE
- ZIMPERIUM
- despinct
- Sentinel
- DEMISTO
- graphistry
- drawbridge
- SignalSense
- AppZen

RECRUITING

- textio
- entelo
- Wade & Wendy
- hi
- unitive
- SpringRole
- GIGSTER
- HireVue

AUTONOMOUS SYSTEMS

GROUND NAVIGATION

- drive.ai
- AdasWorks
- ZOOX
- Mobileye
- UBER
- Google
- TESLA
- Autonomy
- Auro Robotics

AERIAL

- SKYDIO
- SHIELD AI
- Airware
- dji
- LILY
- DroneDeploy
- pilotai
- SKYCATCH

INDUSTRIAL

- JAYBRIDGE
- OSARO
- CLEARPATH
- fetch
- KINDRED
- rethink
- robotics
- HARVEST
- robotics

PERSONAL

- amazon alexa
- Cortana
- Ailo
- facebook
- Siri
- Replika

AGENTS

- PROFESSIONAL
- butter.ai
- pogo
- SKIPFLAG
- clara
- x.ai
- slack
- talla
- Zoom
- sudo

INDUSTRIES

AGRICULTURE

- BLUE@RIVER
- mavrx
- tule
- TRACE
- GENOMICS
- Pivot Bio
- Terraviva
- AGRI-DATA
- Descartes Labs
- ub
- abundant

EDUCATION

- KNEWTON
- volley
- gradescope
- CTI
- coursera
- UDACITY
- alt school

INVESTMENT

- Bloomberg
- sentient
- SENTIUM
- KENSHO
- alphasense
- Dataminr
- CEREBELLUM CAPITAL
- Quandl

LEGAL

- blueJ
- BEAGLE
- Everlaw
- RAVEL
- Seal
- ROSS
- LEGAL ROBOT

LOGISTICS

- NAUTO
- Acer
- PRETECKT
- Routific
- clearmetal
- MARBLE
- PITSTOP

INDUSTRIES CONT'D

MATERIALS

- zymergen
- Citrine
- Eigen Innovations
- SIGHT MACHINE
- BINKGO
- BIOWORKS
- nanotronics
- CALCULARIO

RETAIL FINANCE

- TALA
- zest finance
- Lendo
- earnest
- affirm
- MIRADOR
- wealthfront
- Betterment

HEALTHCARE

PATIENT

- PULSE
- CareSkore
- ZEPHYR
- HEALTH
- IBM Watson Health
- OncoFlow
- BENTRIAN
- Atomwise
- Numerate

IMAGE

- BUTTERFLY
- 3SCAN
- ARTERYS
- enlitic
- BAYLABS
- Imaga
- Google DeepMind

BIOLOGICAL

- iCarbonX
- color
- GRAIL
- deep genomics
- RECURSION
- BAYMINIST
- Numerate
- Atomwise
- verily
- WHOLE BIOME

TECHNOLOGY STACK

AGENT ENABLERS

- OCTANE.AI
- howdy
- Maluuba
- KITT.AI
- OpenAI Gym
- Kasisto
- automat
- semantic machines

DATA SCIENCE

- DOMINO
- SPARKBEYOND
- rapidminer
- kaggle
- DataRobot
- yhat
- AYASDI
- dataiku
- seldon
- yseop
- bigml

MACHINE LEARNING

- CognitiveScale
- GoogleML
- context
- relevant
- Gyrforg
- HyperScience
- narrlogics
- minds.ai
- H2O.ai
- SCALED INFERENCE
- sparkcognition
- loop
- GEOMETRIC INTELLIGENCE
- deepsense.io
- reactive
- skymind
- bonsai

NATURAL LANGUAGE

- agolo
- APPLIEN
- LEXALYTICS
- Narrative Science
- loop
- spaCy
- LUMINOSO
- cortical.io
- MonkeyLearn

DEVELOPMENT

- SIGOPT
- HyperOpt
- fuzzyio
- okite
- rainforest
- lobe
- Anodot
- Signifai
- LAYER 6
- bonsai

DATA CAPTURE

- CrowdFlower
- diffbot
- CrowdAI
- import
- Paxata
- DATASIFT
- amazon
- mechanicalturk
- enigma
- WorkFusion
- DATALOGUE
- TRIFACTA
- parsehub

OPEN SOURCE LIBRARIES

- Keras
- Chainer
- CNTK
- TensorFlow
- Caffe
- H2O
- DEEPLARNING4J
- theano
- torch
- DSSTNE
- Scikit-learn
- AzureML
- neon
- MXNet
- DMTK
- Spark
- PaddlePaddle
- WEKA

HARDWARE

- KNUPATH
- TENSTORRENT
- Cirrascale
- NVIDIA
- intel
- nervana
- Movidius
- tensilica
- GoogleTPU
- 10 Labs
- Qualcomm
- Cerebras
- Isosemi

RESEARCH

- OpenAI
- maise
- ELEMENT
- vicarious
- KNOGGIN
- Numenta
- Kimera Systems
- Cogital

EXPLOSIONSARTIGE UNÜBERSICHTLICHKEIT VON STARTUPS UND TECH FIRMEN IM BEREICH DATA SCIENCE UND AI

...ABER IN WELCHEN BEREICHEN FINDET WERTSCHÖPFUNG STATT?

M2 Analytics: Customer experience enhancement and operation efficiency

Markets

9

Figure 7: Big Data Analytics Market Size by Business Category



Key Facts:

2015 : \$3 billion

2025 : \$30 billion

CAGR of 27%

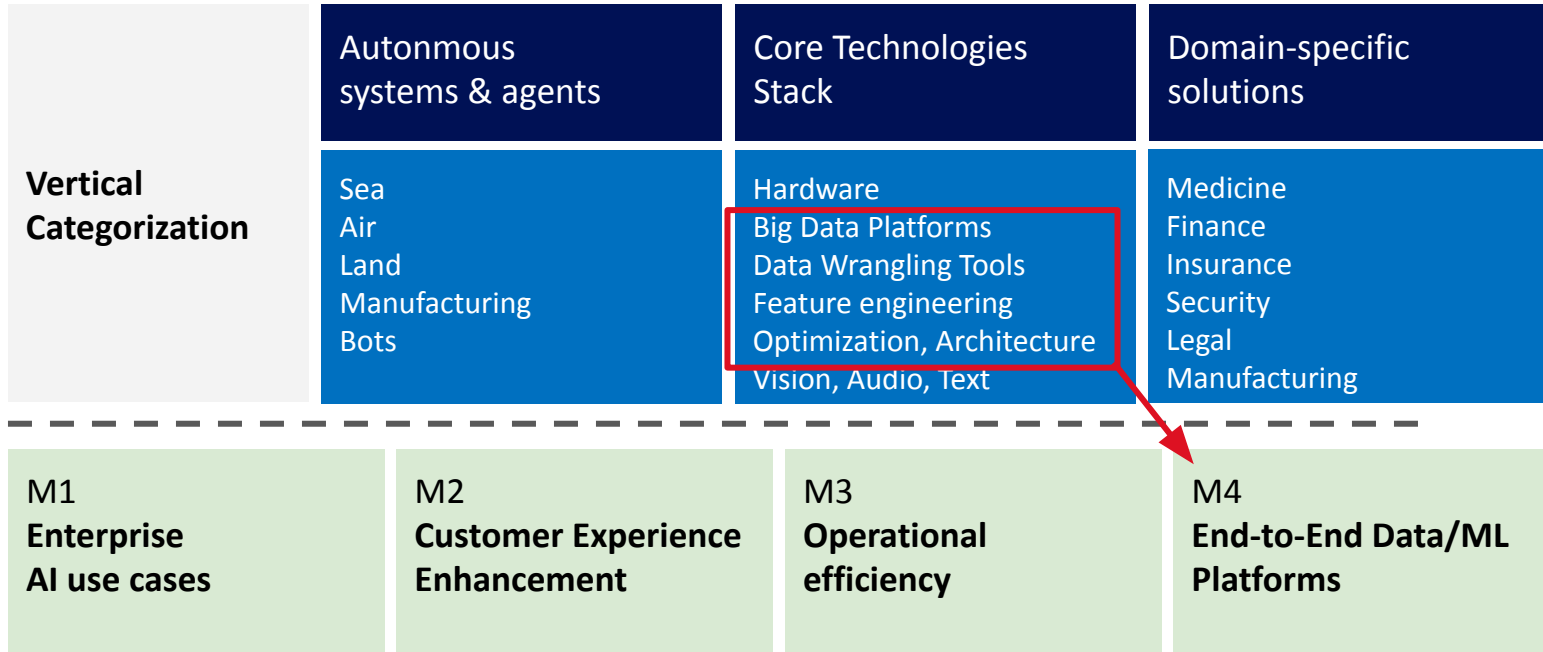
Klaus-Christian Gluckert

Source: Heavy Reading

Märkte und Wertschöpfung aufgrund von Data/Machine Learning Tech

Mega-Trends

10



End-to-End Data & ML Plattformen

Mega-Trends

11

Databricks for Data Science

An open and unified platform to collaboratively run all types of analytics workloads, from data preparation to exploratory analysis and predictive analytics, at scale.

The screenshot displays the Databricks workspace interface. At the top, there's a navigation bar with options like 'demo-cluster', 'File', 'View: Code', 'Permissions', 'Run All', and 'Clear'. Below this, a text box explains the purpose of the next cell: 'To demonstrate, in this next cell, we'll load the `silver_loan_stats` table from the `loans` database that we created in the previous notebook.'

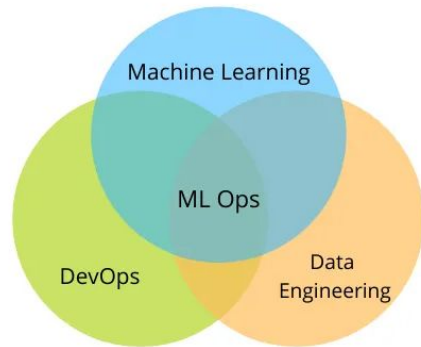
The main area shows a SQL query in Cell 01:

```
1 SELECT *
2 FROM silver_loan_stats
3 WHERE int_rate > 6.34
4 ORDER BY int_rate ASC
```

Below the query, a Spark job is executed, resulting in a table of loan data. The table has columns: loan_status, int_rate, revol_30d, issue_d, earliest_or_line, emp_length, verification_status, total_pymnt, loan_amnt, grade, annual_inc, dti, addr_state, and term. The data is filtered to show loans with an interest rate greater than 6.34, sorted by interest rate in ascending order.

At the bottom, a text box explains: 'Anytime the result of a SQL query or Spark command returns a DataFrame, the result will automatically be displayed in a visualization. The visualization engine can perform aggregations, counts, and group bys on the data to help you play with different visualizations quickly, without knowing exactly what they are looking for right away.'

Workspace: One central place to store and share notebooks, experiments, and projects backed with role-based access control.



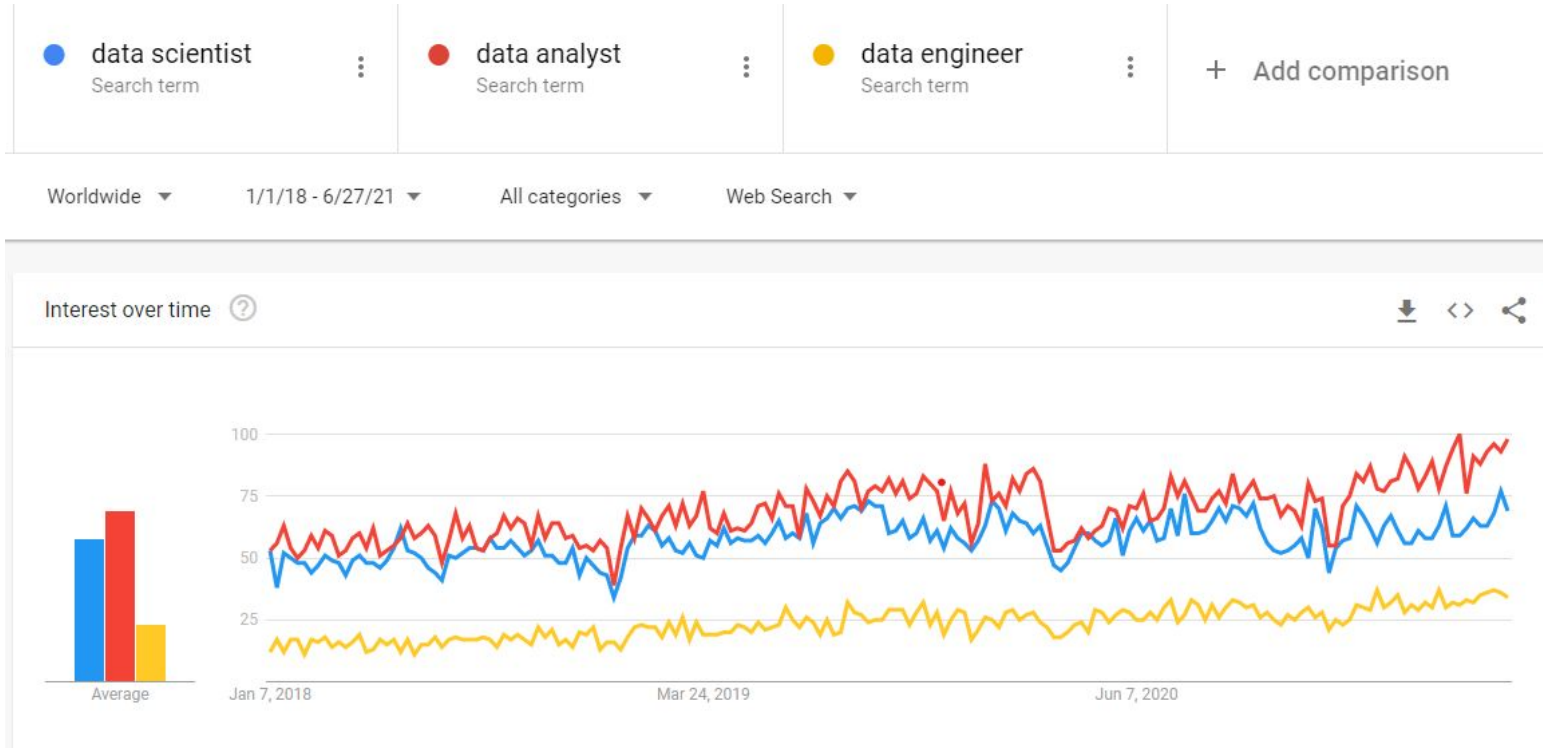
CLOUDERA
splunk >

Azure Data Platform

Automatisierung des “Data Scientists” (Analyst wird wichtiger...)

Mega-Trends

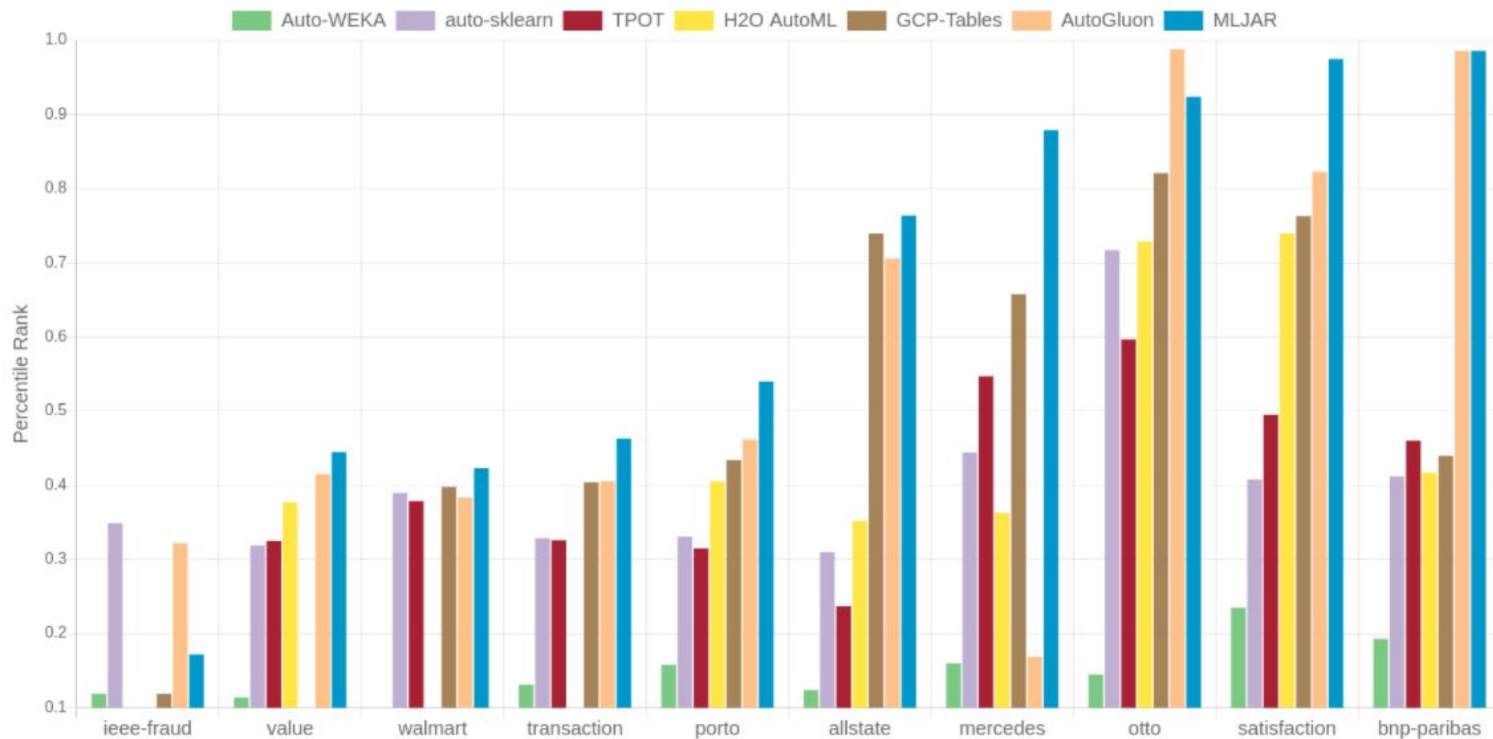
12



Automatisierung des “Data Scientists” (AutoML)

Mega-Trends

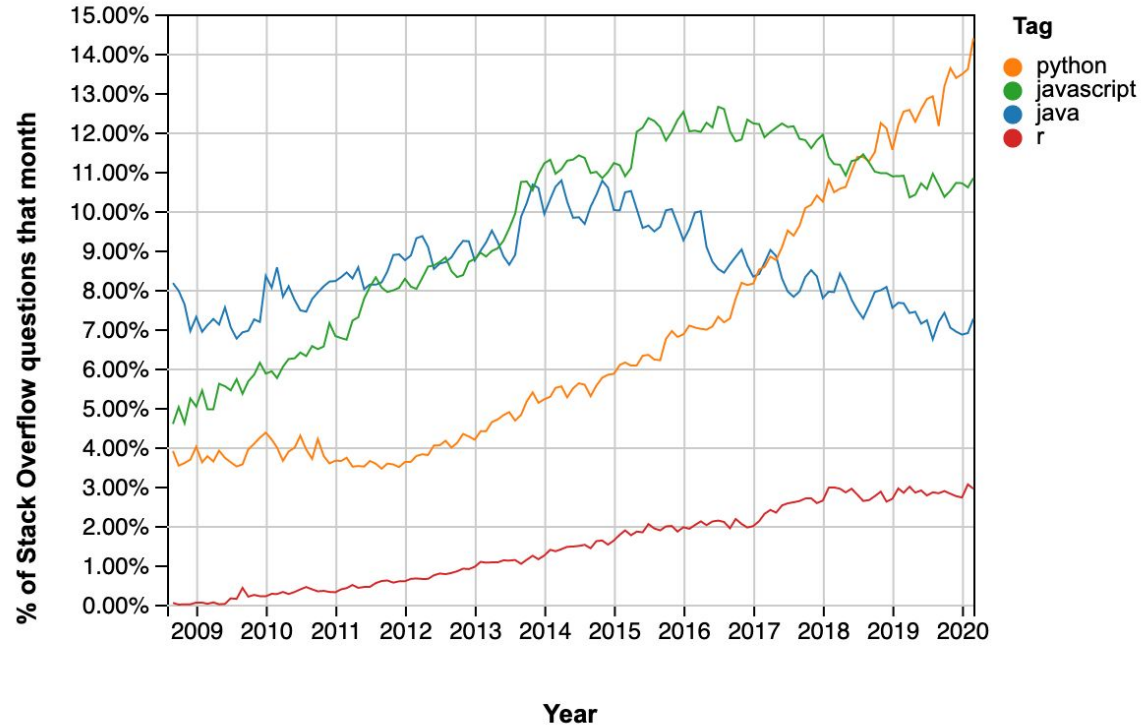
13



Python als Standard für Data Professionals

Mega-Trends

14



Federated Data Modelling, Privacy and Deep-Fakes

Mega-Trends

15

Klaus-Christian Glücker

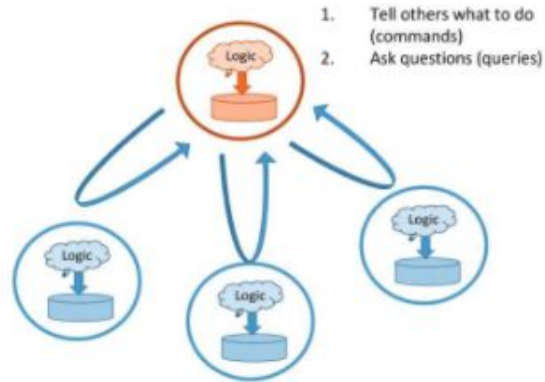


Vom Batch zu Stream-Processing (MLOps)

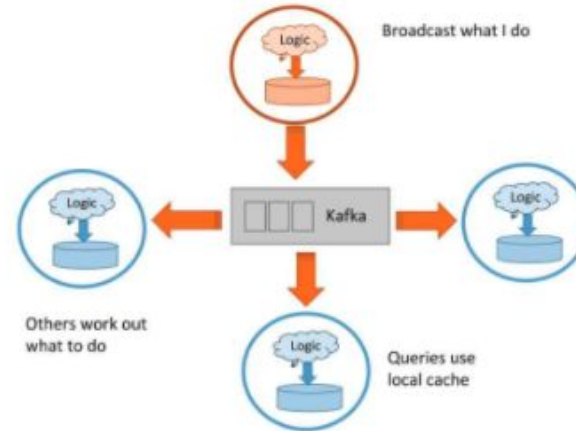
Mega-Trends

16

Request Driven



Event Driven



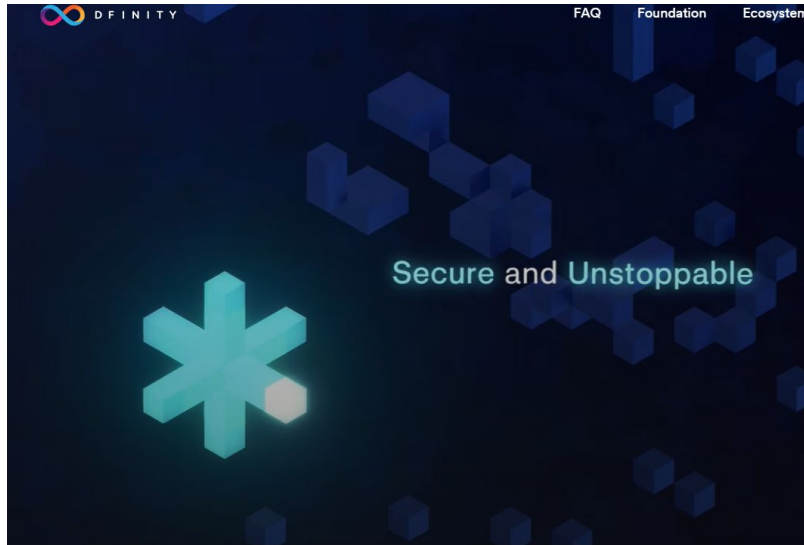
Beyond Microservices: Streams, State and Scalability (Gwen Shapira, QCon 2019)

<https://huyenchip.com/2020/12/27/real-time-machine-learning.html>

Dezentralisierung in Compute & Storage

Mega-Trends

17



DEFINITY

FAQ Foundation Ecosystem

Secure and Unstoppable

The Internet Computer

The world's first frictionless blockchain with web speed and internet scale

START CODING HOW IT WORKS

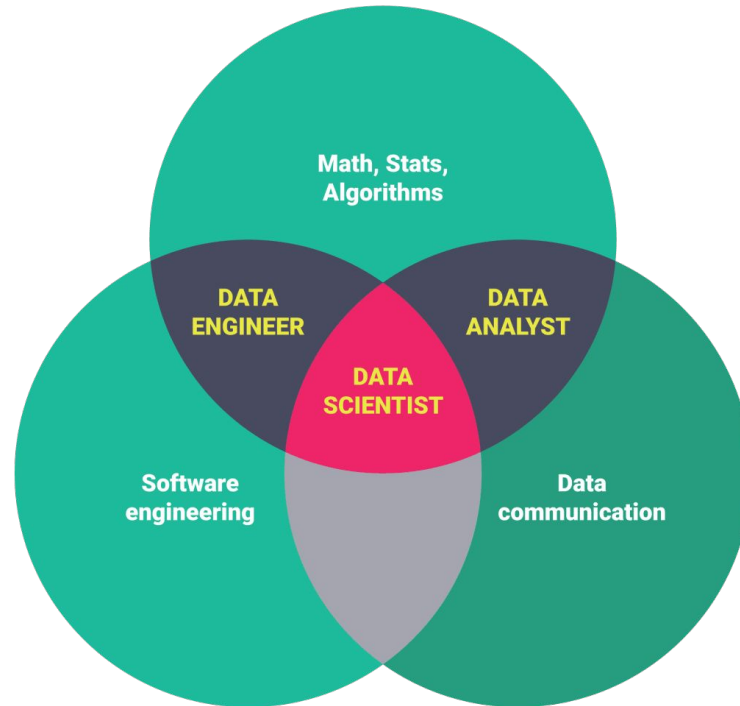
The decentralized
cloud is here.



Data Scientist als eierlegende Wollmilchsau

Profile

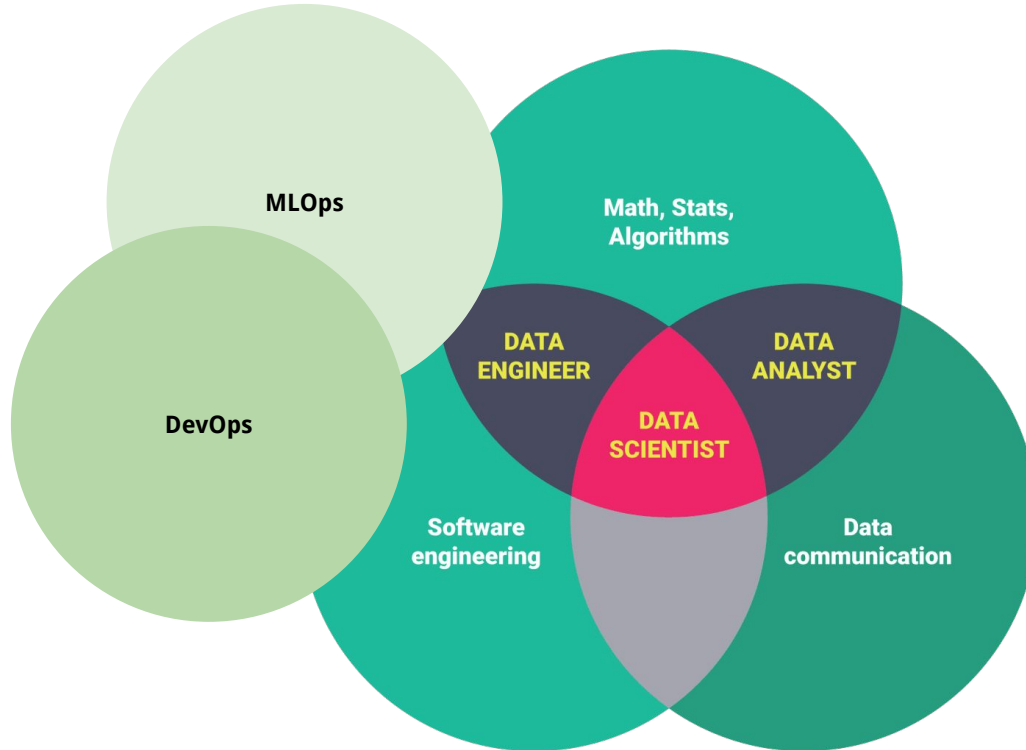
18



Data Scientist als eierlegende Wollmilchsau

Profile

19



Data Scientist als eierlegende Wollmilchsau

Profile

20



Pascal BORNET • 2nd
AI, Automation & Innovation Leader | Author...
4h • 🌐

+ Follow

Don't work hard, work smart!

About finding the right balance between [#ergonomic](#) and... see more

Working smart Vs Working hard



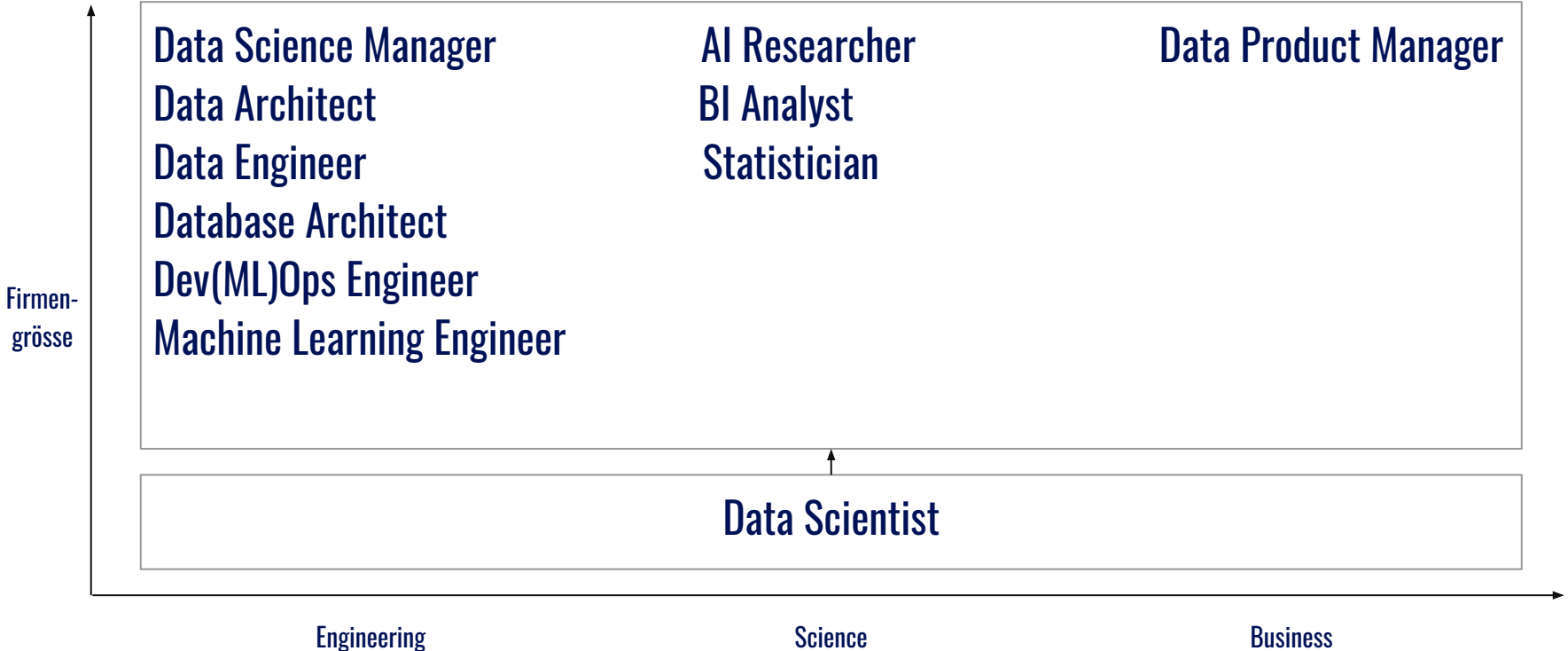
🔗 🏆 🌱 717

48 comments • 8,711 views

Rollensplit des Data Scientist

Profile

21



Gesuchte Profile Schweiz: “Engineering” und “Analytics” Profile

22

| A | B | C | D | E | F | G | H | I |
|--|--------------|-----------|----------|------------|---------|-----------|-----------|----------|
| 250 | 2.80% | 4.40% | 6.40% | 9.20% | 9.20% | 10.00% | 11.60% | 26.40% |
| | Intelligence | Architect | Big Data | Consultant | Analyst | Scientist | Analytics | Engineer |
| (Junior) Data Specialist (m/w/div) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| (Senior-) Consultant - Data Management & Strat | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| (Senior) Consultant - Forensic Technology, eDis | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| (Senior) Data Analyst (f/m) 80-100% | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| (Senior) Data Engineer (w/m/d) - Zürich | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| (Senior) Data Scientist 80-100 % | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Application Engineer in Data Governance (all ge | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Associate Data Analyst and Data Engineer | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| BetriebsselektrikerIn Data Center | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Big Data Application Developer | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

<https://www.jobagent.ch/search?terms=Data&provinces=ZH&page=1#results>

Beispiel Data Science Bootcamp Lernstoff

Kompetenzen



The Data Incubator

Data science training programm Silicon Valley, Data Science
2017 – 2017



Cornell-funded training program for data science and machine learning with < 2% acceptance rate. Modules covered:

Modules covered:

1. Python data manipulation
2. SQL
3. Web scrapping & Graph theory
4. Machine Learning with Scikit-Learn
5. Natural Language Processing
6. Time Series
7. Digital Signals
8. Map Reduce & MrJob library in Python
9. Apache Spark with PySpark and Scala
10. Interactive Data Visualization
11. Capstone Project

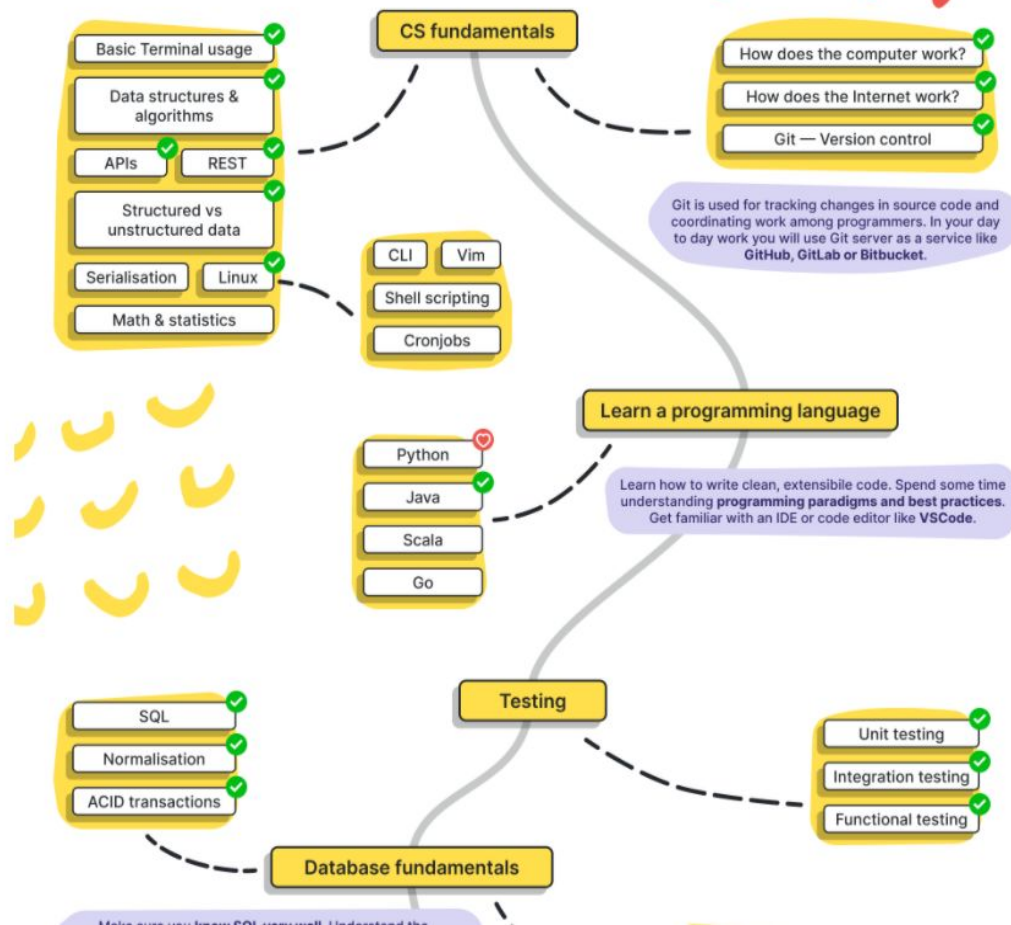
...und ein Data Engineer?

Kompetenzen

<https://github.com/datastacktv/data-engineer-roadmap>

- Personal recommendation
- General recommendation
- Cloud based

Data Engineer in 2021



Welche Möglichkeiten gibt es?

Jobeinstieg

- > Master in Computational Science zur Doppelqualifikation (Empfehlung: Statistik!)
- > UNI-Spinoff
- > Data Science Bootcamp
- > Unternehmensberatung
- > Startup
- > FAANG
 - Route als “Data Scientist” (M.A., Ph.D.)
 - Route als “Research Scientist” (Ph.D.)

However, this risk free approach is growing in popularity and has specifically permeated my field “Machine Learning”.

FAANG salary with an academic appointment is the

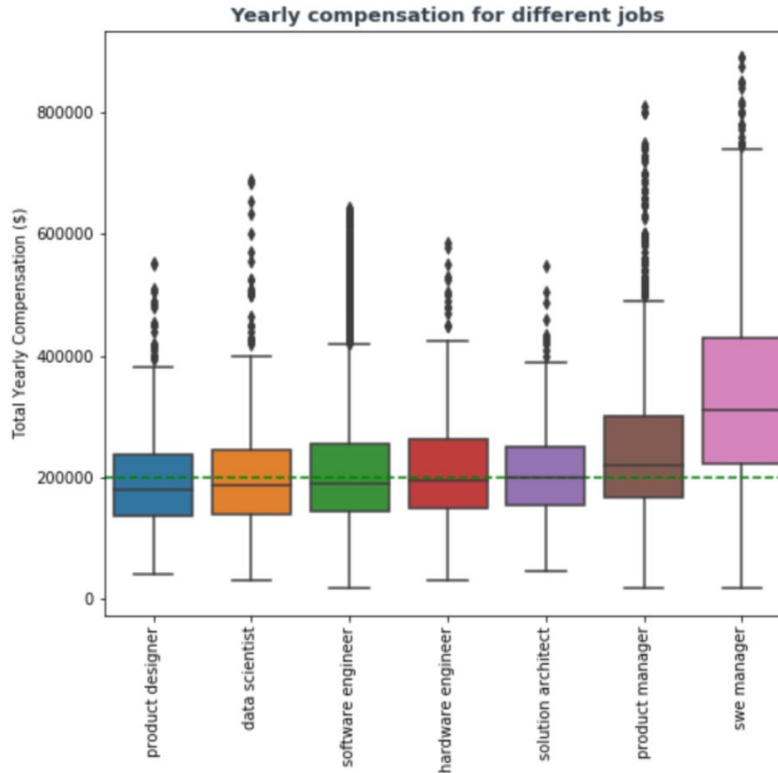
best job available in the world today.

<https://marksaroufim.substack.com/p/machine-learning-the-great-stagnation>

FAANG Gehälter

Jobeinstieg

26



<https://huyenchip.com/2020/01/18/tech-workers-19k-compensation-details.html>

- > Einstiegsgehalt (PHD): 120k
- > Stock-Options: 50k pro Jahr
- > Jahres-Bonus: 30k
- > Signing-Bonus: 50k

- > Total nach 4y: 1.25 Millionen
* inkl. Aktien-Wertsteigerung

Wo könnte ich z.B. als Psychologe/in dann überhaupt direkt einsteigen?

Jobeinstieg

27

- > Clinical Research
- > MedTech StartUp
- > Marketing Analytics
- > People Analytics (eher USA, nicht CH)



Ava is a digital health company with offices in Zurich, San Francisco, Belgrade and Makati that aims to advance women's reproductive health by bringing together artificial intelligence and clinical research. Our wearable device, smart app and proprietary predictive algorithms empower women by giving them unique clinically researched insights and personalized data about their menstrual cycle, fertile window, and pregnancy delivered in a way that's convenient and non-invasive. Ava was voted Best of Baby Tech at CES 2017, named a Women's Health "Editors' Choice" product and has been honoured as the best Swiss startup in 2017 and 2018. Our current key markets include USA, Germany, Switzerland and UK.

Would you like to join us on our challenging adventure? To strengthen our team, we are looking for a:

Clinical Research Scientist (60-100%)

Location: Zurich, Switzerland

Note: this joint job offer with the University Hospital of Zurich is limited to March 2023

Responsibilities:

- Lead and manage large clinical trials (e.g. trial design, statistical analysis plan, clinical documentation, CRO appraisal and selection, recruitment planning, monitoring, close out, etc.)
- Analysis and interpretation of complex real-world data
- Drive clinical development deliverables (e.g., CEP, CIP, CER etc.)
- Represent Ava globally in scientific events and conferences
- Synthesize scientific knowledge for internal and external use (e.g. marketing, regulatory filings, peer-reviewed journals)

Fragen der Teilnehmer/innen

Jobeinstieg & Bewerbung

- > Bewerbungsstrategien MA. vs PHD. ?
- > Wie gelte ich als PHD im Wechsel von der Uni in die Privatwirtschaft nicht als Berufseinsteiger/in?
- > Habe ich als Psychologe/in die richtigen Sach-Kenntnisse?
- > Welche Kompetenzen werden zwingend vorausgesetzt für Geisteswissenschaftler/innen?
- > Wo hätte man z.B. als Psychologe Vorteile gegenüber Computer-Wissenschaftlern?
- > Interview-Prozess (CH)