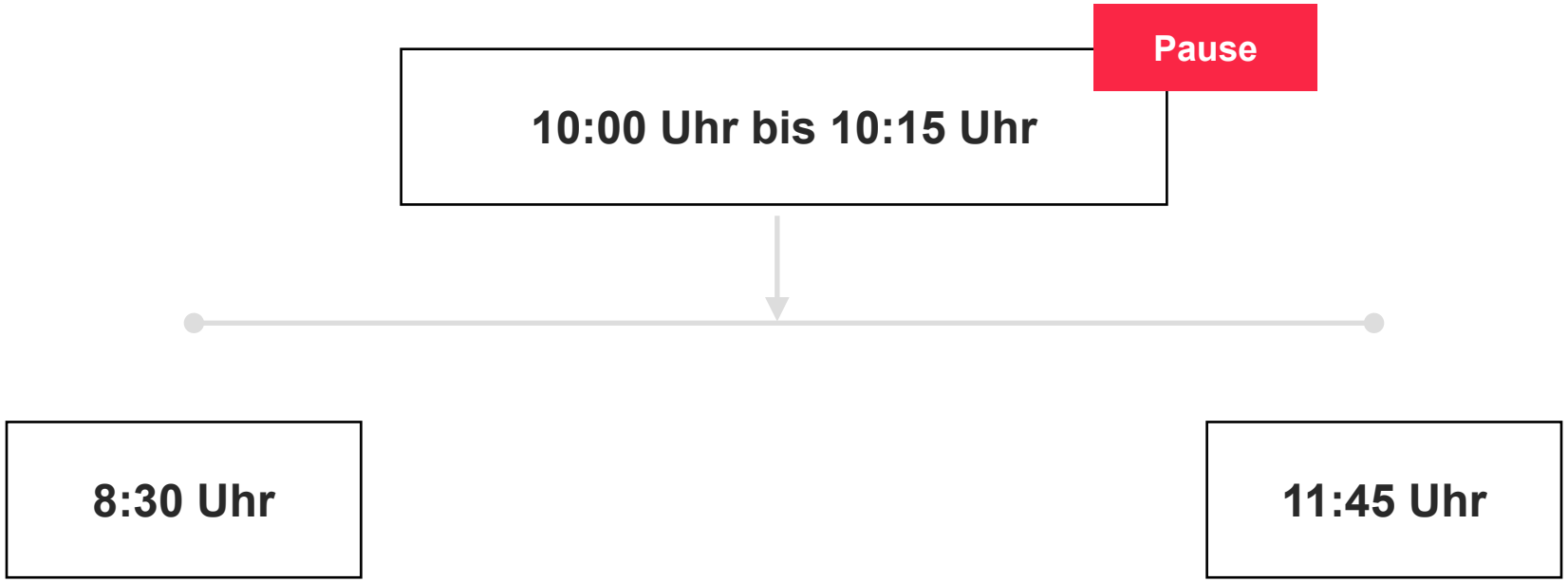
DATA MINING

21. April 2020

Zeitplan



Wer bin ich?

Tin Votan

- Wirtschaftsingenieurwesen / Master of Science
  Karlsruher Institut für Technologie

- Wirtschaftsingenieurwesen / Bachelor
  Technische Universität Dresden

- Start-up-Gründer
- Machine Learning Engineer

votan@lehre.dhbw-stuttgart.de

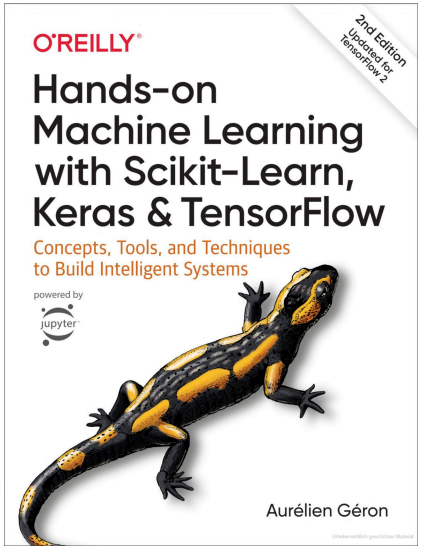
linkedin.com/in/tinvotan/

Quellen

Géron, Aurélien (2019): *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2. Edition, Sebastopol/CA (United States of America): O'Reilly Media Inc.

Bizer, Christian (2020): *Data Mining - Introduction and Organization*. (Vortrag, 12.02.2020). Mannheim: Universität Mannheim - Data and Web Science Group.

https://github.com/ageron/handson-ml2 (Abruf: 19.04.2020)



Was ist Data Mining?

riesige Datenmengen

4 Petabytes pro Tag

Muster in den Daten

Interesse und Verhalten von 1 Mrd. Nutzer

Entscheidungsgrundlage für Strategien

Vorhersagen treffen

https://www.brandwatch.com/blog/facebook-statistics/

Definition

Erkundung und Analyse von sehr großen Datenmengen um bedeutungsvolle Muster zu erkennen

Nicht-unbedeutende Extraktion von:
- **impliziten,**
- **vorher unbekannten und**
- **potenziell nützlichen**

Informationen aus Daten

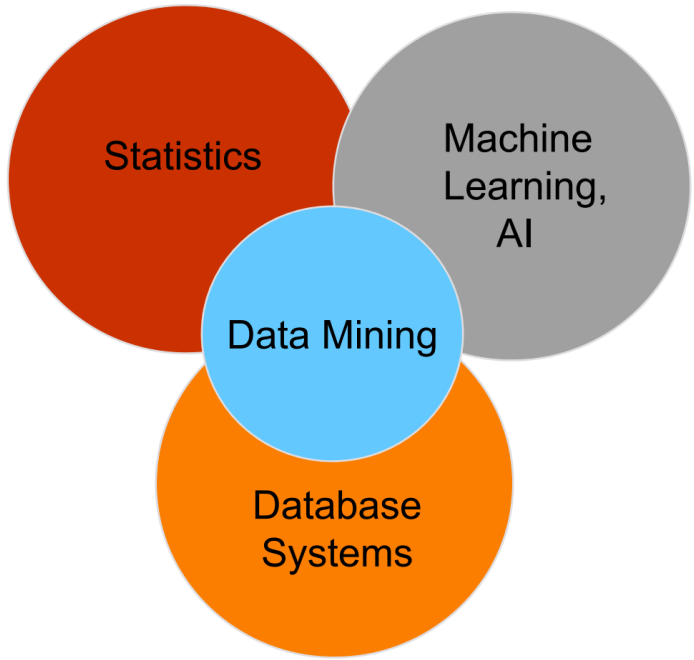
Data-Mining Methodik

1. Interessante Muster in großen Datenmengen erkennen
2. Entscheidungen / Strategien aufgrund der abgebildeten Entscheidungen treffen
3. Vorhersagen über mögliche Ereignisse basierend auf den erkannten Mustern treffen

Ursprünge

Unzulänglichkeiten der traditionellen Verfahren überwinden

- große Datensätze
- hohe Dimensionalität der Daten
- Heterogenität und komplexe Struktur von Daten
- explorative (erkenntnisreiche) Analyse fern vom "*Hypothese stellen und testen*" - Paradigma



Vorgehensweise

1. Daten sammeln

2. Daten erkunden

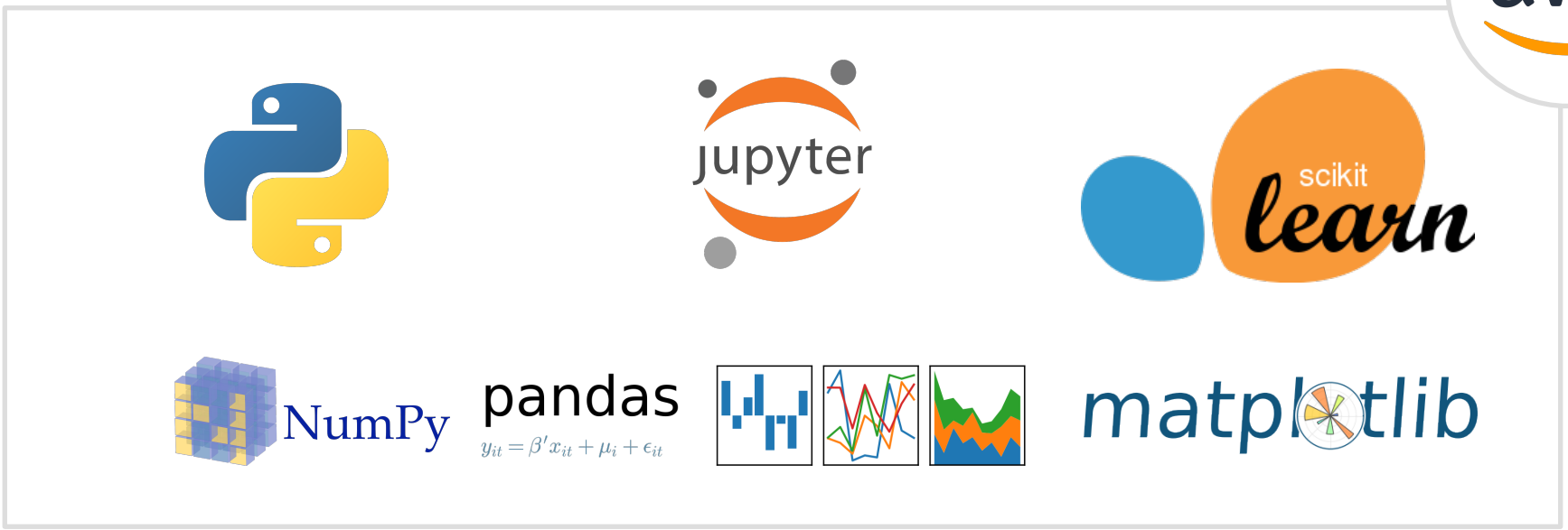
3. Daten aufbereiten

4. Algorithmus auf Daten trainieren

5. Algorithmus evaluieren

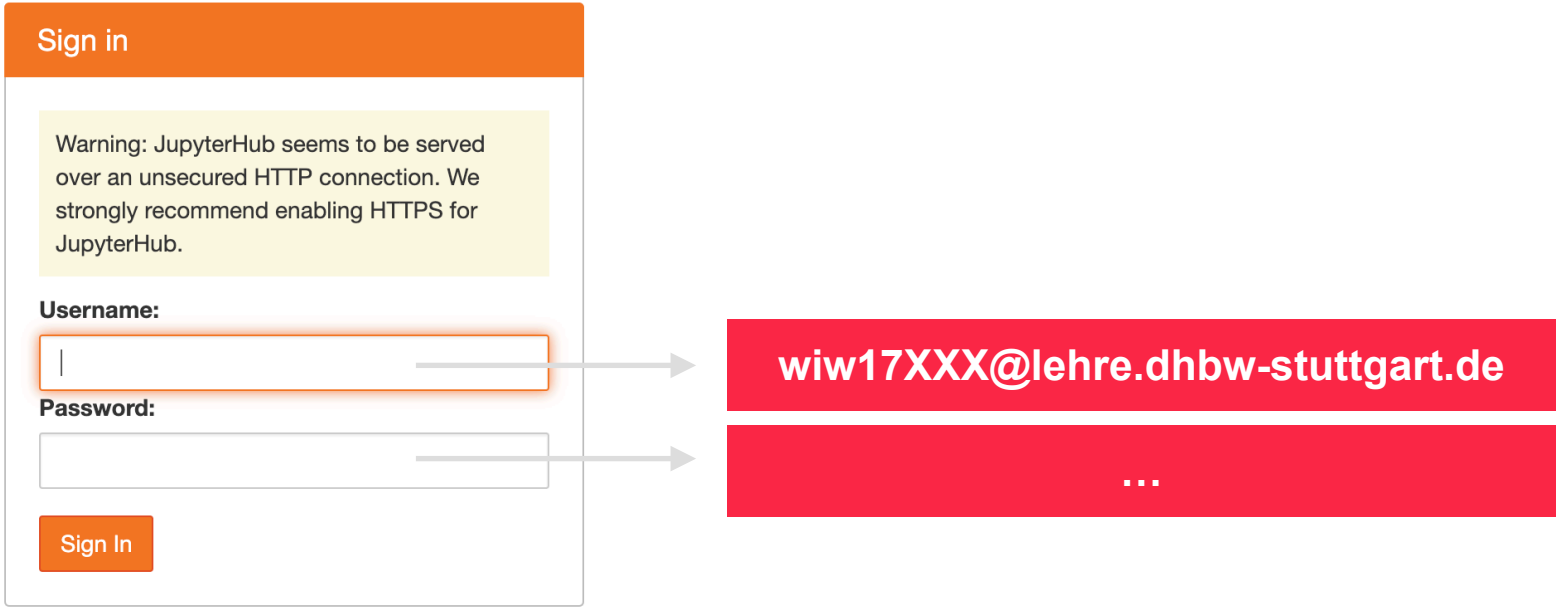
Die heutige Veranstaltung

Vorhersage von der Immobilienpreise in den einzelnen Distrikten von Kalifornien 1990

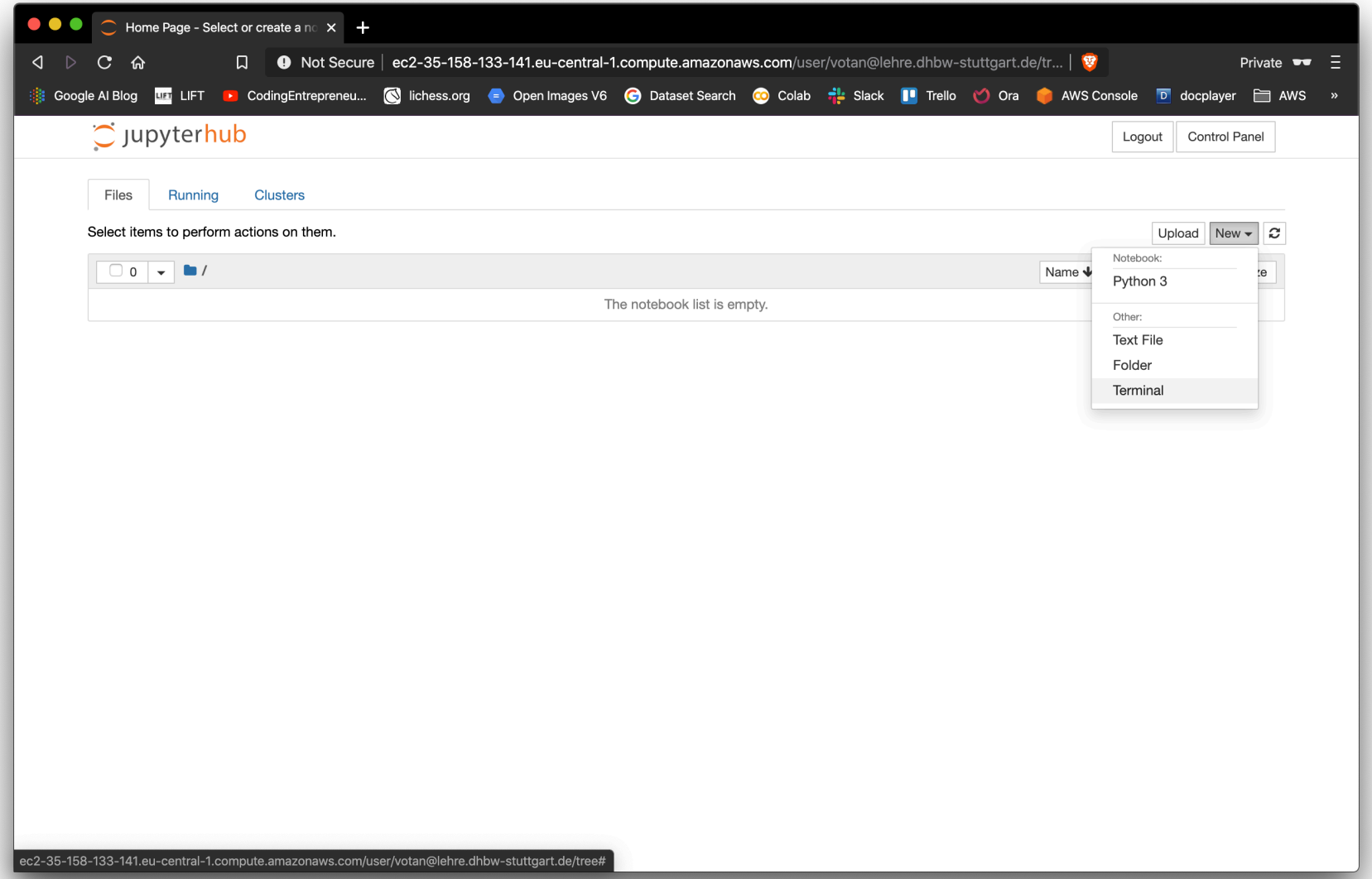


JupyterHub: Anmeldung auf AWS Client

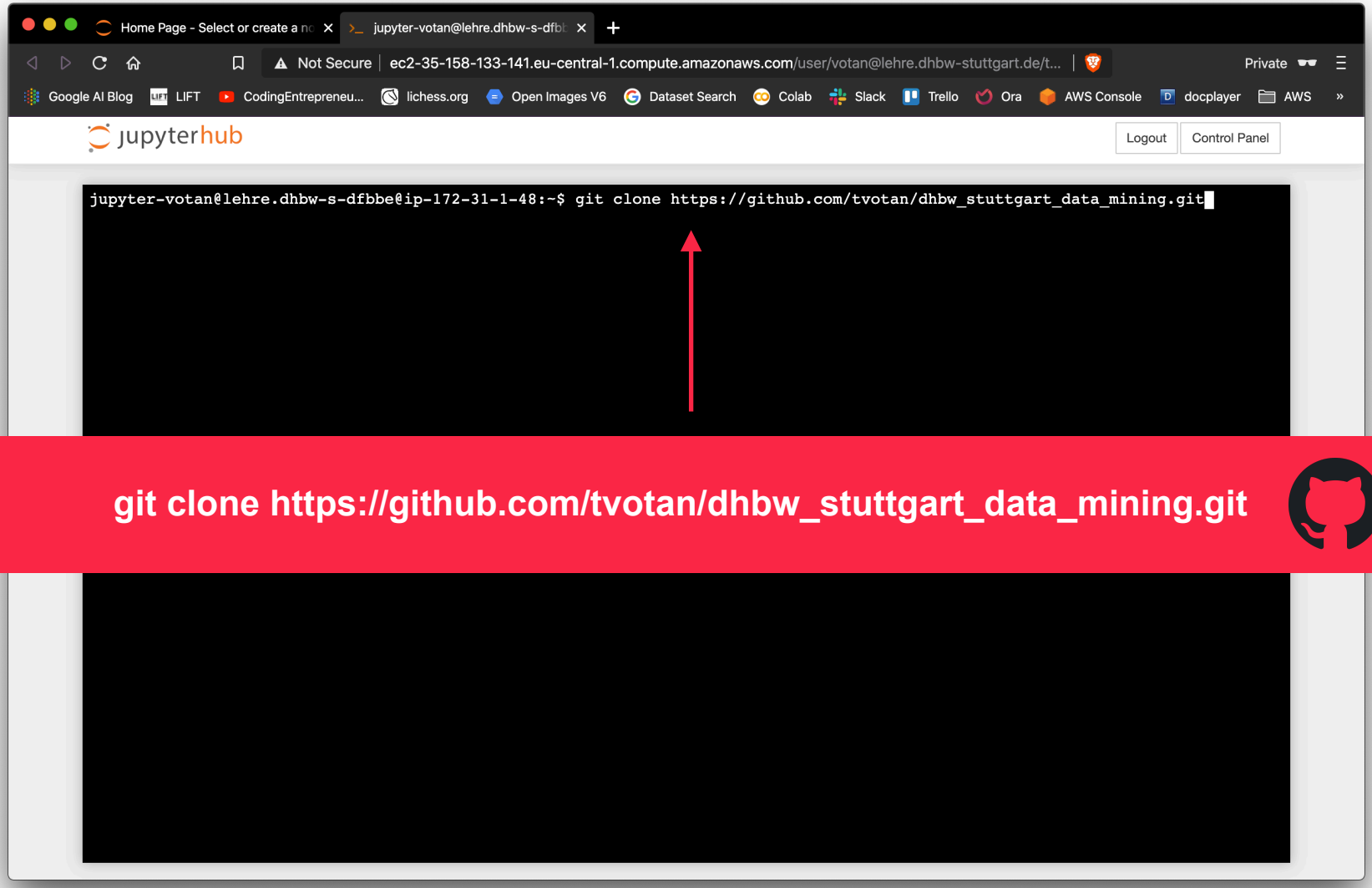
http://ec2-35-158-133-141.eu-central-1.compute.amazonaws.com/hub/login



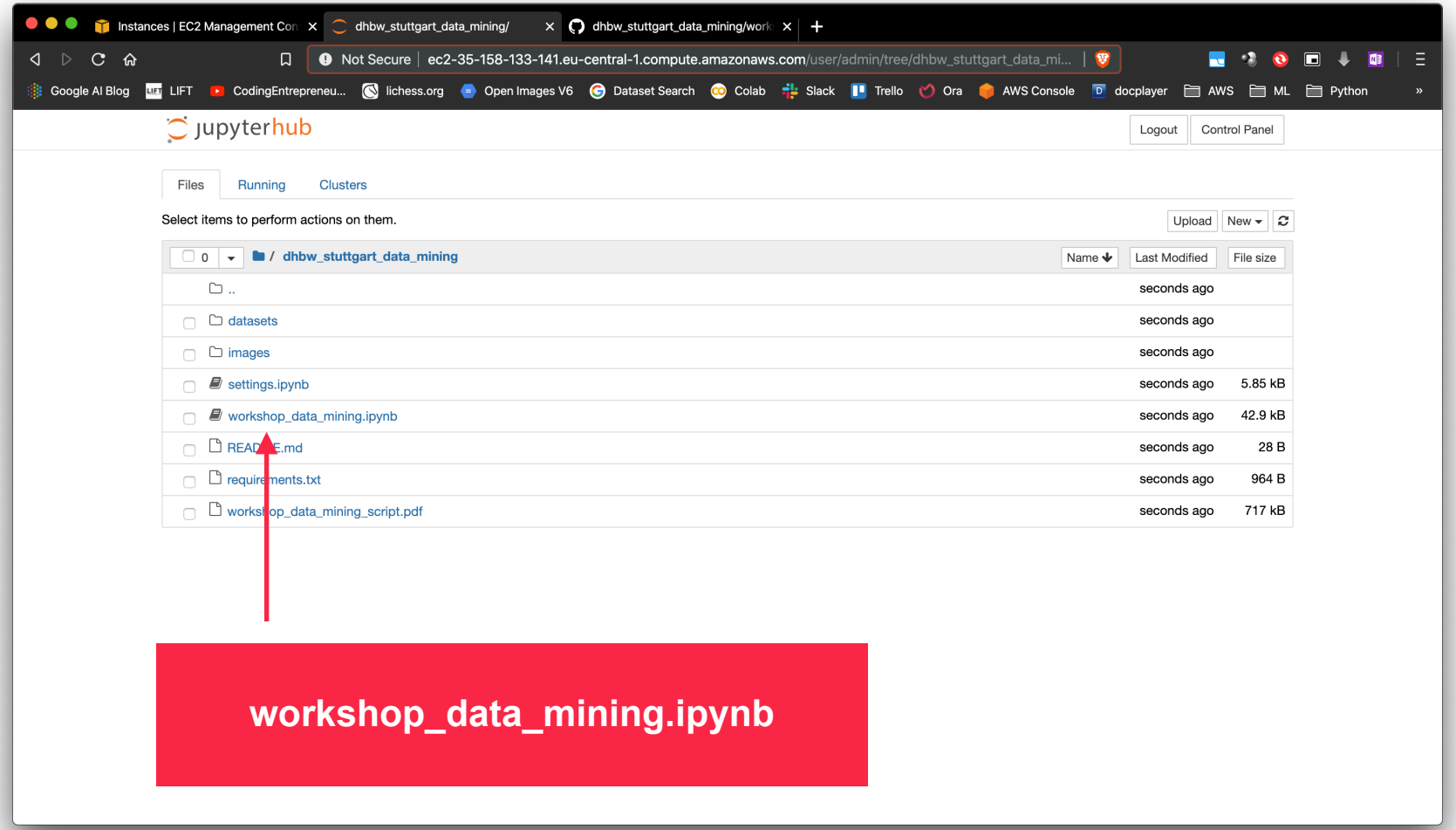
JupyterHub: Terminal öffnen



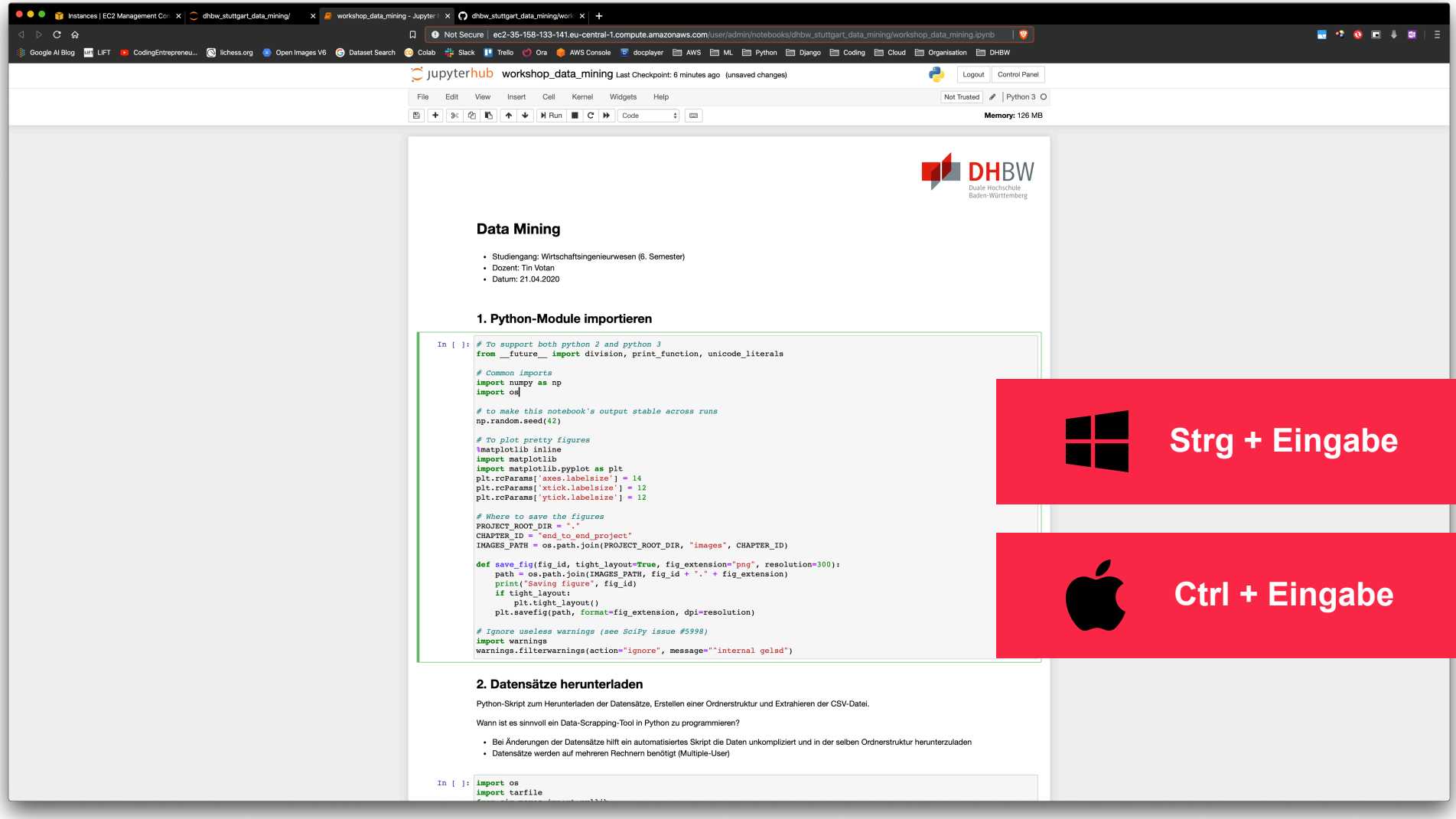
JupyterHub: Repository herunterladen



JupyterHub: Notebook öffnen



JupyterHub: Ausführen von Python-Code



Strg + Eingabe

Ctrl + Eingabe