

BDEA Aufgabe 1: Interaktive Datenanalyse

Installation

Docker Setup (mysql:8.0)

- `docker compose up`
- Datenbank `movie_tv` wird automatisch durch die `setup.sql` Datei erstellt

Jupyter Notebook

- Beinhaltet alle benötigten python packages und installiert sie beim Ausführen mithilfe von magic z. B.: `%pip install pandas`

Daten Laden

- Daten entweder über Kaggle (siehe Daten) laden
- Oder über mein GitHub Repository: [BDEA](#)
- Die Daten müssen in einem Ordner namens „kaggle“ in dem Ordner mit dem Notebook liegen

Quellen

Daten

- [Amazon Prime Movies and TV Shows](#)
- [Apple TV+ Movies and TV Shows](#)
- [Disney+ Movies and TV Shows](#)
- [Netflix Movies and TV Shows](#)
- [Paramount+ Movies and TV Shows](#)

Daten säubern

- [Hilfe zur datensäuberung](#)

MySQL

- [Window Functions](#): `CUME_DIST()`, `DENSE_RANK()`, `ROW_NUMBER()`
- [Rand\(\)](#)
- [WITH \(Common Table Expressions\)](#)
- [JOIN Clause](#)

Pandas

- [DataFrame](#): Arbeiten mit den Daten in python
- [read_csv](#): CSV dateien einlesen
- [to_sql](#): Dataframe automatisch in die Datenbank speichern
- [read_sql_query](#): SQL Abfrage in python schreiben

SQLAlchemy

- [create_engine](#): Datenbankverbindung aufbauen

Seaborn

- [Correlation matrix](#)
- [Pairplot](#)
- [Boxplot](#)
- [jointplot](#)
- [lineplot](#)