

# Python für Datenanlysten

## Projekt – Methoden der Datenanalyse

In diesem Projekt findest du die Aufgabenstellung für das Abschlussprojekt zum Thema Python für Datenanalysten. Beginn der Projektarbeit ist am 15.07.2024 um 9:00 Uhr.

Die Projekte sind bis zum 19.07.2024 um 20:00 Uhr in Form einer komprimierten Datei mit dem Titel „vorname\_nachname\_datananalysis.zip“ bei den Teaching Assistants abzugeben. \_\_\_\_\_



Der Datensatz enthält 3 Dateien:

**animes.csv:** Enthält eine Liste von Anime mit Titel, Titel-Synonymen, Genre, Dauer, Rang, Popularität, Bewertung, Ausstrahlungsdatum, Episoden und vielen anderen wichtigen Daten über einzelne Anime. Diese Informationen bieten ausreichende Einblicke in Trends im Laufe der Zeit in Bezug auf wichtige Aspekte von Anime. Der Rang ist im CSV-Format als Gleitkommazahl angegeben, enthält jedoch nur ganzzahlige Werte. Dies liegt an NaN-Werten und ihrer Darstellung in pandas.

**profiles.csv:** Enthält Informationen über Benutzer, die Anime schauen, nämlich Benutzername, Geburtsdatum, Geschlecht und eine Liste ihrer Lieblings-Anime.

**reviews.csv:** Enthält Informationen über Bewertungen von Benutzern zu Animes, einschließlich Textrezensionen und Bewertungen.

## STAKEHOLDER ANFORDERUNGEN

---

- **ANALYSIERE DEN DATENSATZ BEANTWORTE FOLGENDE FRAGEN:**

Wie haben sich die Genres auf den Beliebtheitsscore ausgewirkt?

Wie hat sich die Länge der Episoden auf den Beliebtheitsscore ausgewirkt?

Gibt es demographische Unterschiede in der Beliebtheit?

Welche anderen Parameter hatten Einfluss auf den Beliebtheitsscore?

- **TRIFF VORHERSAGEN, WELCHE GENRES IN ZUKUNFT EINEN HOHEN BELIEBTHEITSSCORE ERZIELEN WERDEN**
- **ENTSCHEIDE, AUF WELCHE METRIKEN PRODUZENTEN ACHT GEBEN MÜSSEN**
- **DIE ERKENNTNISSE MÜSSEN IN EINEM DASHBOARD ZUSAMMENGEFASST WERDEN**
- **ERKENNTNISSE ÜBER DIE DATEN MÜSSEN NACHVOLLZIEHBAR DOKUMENTIERT WERDEN**

## VORGEHEN

---

- **FÜHRE EINE DESKRIPTIVE DATENANALYSE DURCH, UM DIR EINEN ÜBERBLICK ÜBER DEN DATENSATZ ZU VERSCHAFFEN.**

Joine die Datensätze und ergänze fehlende Werte, falls für die Analyse benötigt.

Führe Visualisierungen durch, um erste Zusammenhänge der Daten zu erkennen.

Welche Auffälligkeiten weist der Datensatz auf? Gibt es Daten, die zu wenig repräsentiert werden?

Welche Korrelationen gibt es?

- **FÜHRE EINE EXPLORATIVE ANALYSE DURCH UND BEANTWORTE DIE OBEN GESTELLTEN FRAGEN**

- **FÜHRE EINE PREDIKTIVE ANALYSE DURCH, UM DAS ZUKUNFTSVERHALTEN DER DATEN ZU ERAHNEN.**

Beachte dabei, die Daten mit Feature Engineering oder Normalisierung zu präparieren.

Probiere dich an mehreren Regressionsalgorithmen und dokumentiere deren Performance. Entscheide dich für einen Algorithmus und begründe deine Entscheidung

- **ÜBERFÜHRE DEINE ANALYSIS UND MODELLIERUNG IN EINE STREAMLIT APP.**

Deine App sollte dem Stakeholder die Möglichkeit geben, deine Analysis nachvollziehen zu können.

Gib dem Appnutzer die Chance, sich mit Klicks interaktiv durchzuklicken

Das Model sollte nicht bei jedem Event neu kompilieren. Lass die Vorhersage nur dann laufen, wenn ein Nutzer dies explizit in der App mit einem Knopfdruck ansteuert.