Schritt-für-Schritt-Anleitung: Kaggle-Datensatz mit Büchern herunterladen und in Python nutzen

# 1. Registrierung und API-Token

Gehe auf kaggle.com, melde dich an oder erstelle einen Account.

Klicke oben rechts auf dein Profilbild → „Account“.

Scrolle runter zu API → „Create New API Token“.

Es wird eine Datei kaggle.json mit deinen Zugangsdaten heruntergeladen.

# 2. Kaggle API lokal einrichten

Lege die Datei kaggle.json im Ordner ~/.kaggle/ ab (Linux/Mac) oder %USERPROFILE%\.kaggle\ (Windows).

Installiere das Kaggle-Python-Paket (falls nicht schon):

bash

Code kopieren

pip install kaggle

# 3. Datensatz herunterladen

Öffne ein Terminal oder Jupyter Notebook und führe aus:

bash

Code kopieren

kaggle datasets download zygmunt/goodbooks-10k

unzip goodbooks-10k.zip -d goodbooks

Der Datensatz enthält 10.000 Bücher mit vielen Infos (Titel, Autoren, Genres, Bewertungen).

# 4. Daten in Python laden und anschauen

python

Code kopieren

import pandas as pd

Schritt-für-Schritt-Anleitung: Kaggle-Datensatz mit Büchern herunterladen und in Python nutzen

1. Registrierung und API-Token

Gehe auf kaggle.com, melde dich an oder erstelle einen Account.

Klicke oben rechts auf dein Profilbild → „Account“.

Scrolle runter zu API → „Create New API Token“.

Es wird eine Datei kaggle.json mit deinen Zugangsdaten heruntergeladen.

2. Kaggle API lokal einrichten

Lege die Datei kaggle.json im Ordner ~/.kaggle/ ab (Linux/Mac) oder %USERPROFILE%\.kaggle\ (Windows).

Installiere das Kaggle-Python-Paket (falls nicht schon):

bash

Code kopieren

pip install kaggle

3. Datensatz herunterladen

Öffne ein Terminal oder Jupyter Notebook und führe aus:

bash

Code kopieren

kaggle datasets download zygmunt/goodbooks-10k

unzip goodbooks-10k.zip -d goodbooks

Der Datensatz enthält 10.000 Bücher mit vielen Infos (Titel, Autoren, Genres, Bewertungen).

4. Daten in Python laden und anschauen

python

Code kopieren

import pandas as pd

[goodbooks-10k/samples/books.csv at master · zygmuntz/goodbooks-10k · GitHub](https://github.com/zygmuntz/goodbooks-10k/blob/master/samples/books.csv)

[GitHub - zygmuntz/goodbooks-10k: Ten thousand books, six million ratings](https://github.com/zygmuntz/goodbooks-10k)

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Beispiel: Scrape Spiegel-Archiv für 2024 Hardcover Belletristik

url = "https://www.bestsellerliste.de/" # ...

resp = requests.get(url)

soup = BeautifulSoup(resp.text, "html.parser")

# dann Tabellen extrahieren, in DataFrame speichern...

import requests

from bs4 import BeautifulSoup

import pandas as pd

# Ziel-URL

url = "https://de.wikipedia.org/wiki/Liste\_der\_meistverkauften\_Belletristikb%C3%BCcher\_in\_Deutschland"

# Wikipedia-Seite abrufen

response = requests.get(url)

soup = BeautifulSoup(response.text, "html.parser")

# Alle Tabellen auf der Seite finden

tables = soup.find\_all("table", {"class": "wikitable"})

books = []

for table in tables:

rows = table.find\_all("tr")

headers = [th.text.strip() for th in rows[0].find\_all("th")]

for row in rows[1:]:

cells = row.find\_all(["td", "th"])

if len(cells) >= 3:

titel = cells[0].text.strip()

autor = cells[1].text.strip()

jahr = cells[2].text.strip()

verlag = cells[3].text.strip() if len(cells) > 3 else ""

books.append({

"Titel": titel,

"Autor": autor,

"Jahr": jahr,

"Verlag": verlag

})

# In DataFrame umwandeln und speichern

df = pd.DataFrame(books)

df.to\_csv("bestseller\_buecher.csv", index=False, encoding="utf-8")

print("✅ CSV-Datei 'bestseller\_buecher.csv' erstellt mit", len(df), "Einträgen.")

Tools

ChatGPT kann Fehler machen. Überprüfe wichtige Informationen. Siehe Cookie-Voreinstellungen.

Tools

ChatGPT kann Fehler machen. Überprüfe wichtige Informationen. Siehe Cookie-Voreinstellungen.