

PB2

Pandas & Matplotlib



<https://pandas.pydata.org/>

Inhaltsverzeichnis

- / Was ist Pandas
- / Installation
- / Datenstrukturen in Pandas
- / Grundlegende Operationen
- / Kaggel
- / Pandas Library

Was ist Pandas

- / Python Library für Datenmanipulationen
- / Einlesen von Daten
- / Manipulieren
- / Transformieren
- / Speichern

Installation von Pandas

- / Öffne Console
- / `pip install pandas`

Datenstrukturen in Pandas

/ **Series:**

- / Eindimensionale Datenstruktur in Pandas (ähnelt einem Array)
- / Jede Series hat einen Index, der die Position der Elemente eindeutig identifiziert
- / Beispiel: `pandas_series.py`

/ **DataFrame:**

- / Zweidimensionale Datenstruktur in Pandas (ähnelt Tabelle mit Zeilen und Spalten)
- / Jedes DataFrame hat einen Index (Zeilenbeschriftung) und Spaltenüberschriften
- / Beispiel: `pandas_dataframe.py`

Pandas Befehle

/ df.head()	-> Ausgabe der ersten Zeilen
/ df.tail()	-> Ausgabe der letzten Zeilen
/ df.info()	-> Ausgabe von einer Zusammenfassung
/ df['Date Time']	-> Filtern der Spalte 'Date Time'.
/ df.describe()	-> Ausgabe des Schemas
/ df['T (degC)'].max()	-> max wert der spalte 'T (degC)'.
/ df['T (degC)'].min()	-> minwert der spalte 'T (degC)'.
/ df.mean()	-> Mittelwert

/ <https://pandas.pydata.org/docs/reference/index.html#api>

Daten lesen aus CSV

/ Pandas hat eine read Funktion für viele Datentypen

```
import pandas as pd

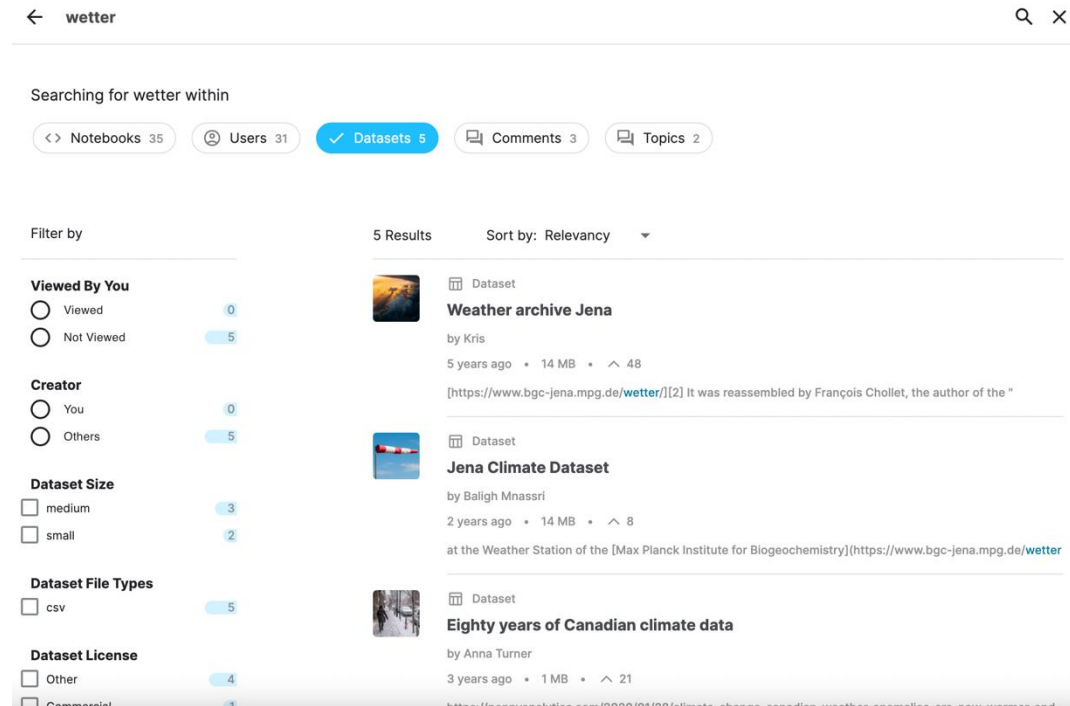
# Laden des CSV-Datensatzes in ein DataFrame
df = pd.read_csv('beispieldaten.csv')

# Anzeigen der ersten 5 Zeilen des DataFrames
print(df.head())
```


Kaggle

- / Plattform für Datenset's und Wettbewerbe.
- / Registrierung nötig.
- / Diverse gratis Kurse.

- / <https://www.kaggle.com/>
- / Suche nach beliebigen Begriff
- / Filter nach Datensets
- / Öffne ein beliebiges Datenset



The screenshot shows the Kaggle search results for the term 'wetter'. The search bar at the top left contains the text 'wetter'. Below the search bar, there are filters for 'Notebooks' (35), 'Users' (31), 'Datasets' (5), 'Comments' (3), and 'Topics' (2). The 'Datasets' filter is selected. The results are sorted by 'Relevancy'. There are 5 results displayed. The first result is 'Weather archive Jena' by Kris, with 14 MB and 48 views. The second result is 'Jena Climate Dataset' by Baligh Mnassri, with 14 MB and 8 views. The third result is 'Eighty years of Canadian climate data' by Anna Turner, with 1 MB and 21 views.

← wetter

Searching for wetter within

<> Notebooks 35 Users 31 **Datasets 5** Comments 3 Topics 2

Filter by

Viewed By You

☐ Viewed 0

☐ Not Viewed 5

Creator

☐ You 0

☐ Others 5

Dataset Size

☐ medium 3

☐ small 2

Dataset File Types

☐ csv 5

Dataset License

☐ Other 4

☐ Commercial 1

5 Results Sort by: Relevancy

Weather archive Jena

by Kris

5 years ago • 14 MB • 48

[https://www.bgc-jena.mpg.de/wetter/][2] It was reassembled by François Chollet, the author of the "

Jena Climate Dataset

by Baligh Mnassri

2 years ago • 14 MB • 8

at the Weather Station of the [Max Planck Institute for Biogeochemistry](https://www.bgc-jena.mpg.de/wetter

Eighty years of Canadian climate data


by Anna Turner


3 years ago • 1 MB • 21


https://www.scribd.com/document/32000000/Climate-change-canadian-weather-station-1948-2008-weather-and-

Kaggle



/ Download das Datenset (Anmeldung nötig)

 Search

 KRIS · UPDATED 5 YEARS AGO

 48

New Notebook

 Download (14 MB) 

Einzelaufgabe (10 Min)

- / Series:
 - / Gebe in der Series Python Applikation den Wert bei Index 'b' aus
 - / Berechne statistische Operationen für die Series:
 - / Mittelwert
 - / Summe aller Werte
- / DataFrame:
 - / Gebe in der DataFrame Python Applikation die Werte für Spalte 'Name' aus
 - / Greife auf Zeile 1 zu
 - / Gebe eine Statistikübersicht zum gesamten DataFrame aus

Einzelaufgabe (20 Min)

- / Vertrautheit mit den grundlegenden Funktionen von Pandas
- / Lade ein Beispieldatensatz aus Kaggle in ein DataFrame
- / Sehe dir die Daten an und deren Beschreibung
- / Manipuliere die Daten, indem du sie filterst, Gruppierst und Aggregierst

Weiterführende Befehle

- / Merge:
 - / Zwei Dataframes basierend auf gemeinsamen Spalten oder Indizes zusammenführen
 - / Ähnlichkeiten zu SQL JOIN-Operationen
 - / Beispiel: «pandas_merge.py»

matplotlib

<https://www.jumpingrivers.com/blog/customising-matplotlib/>

Inhaltsverzeichnis

- / Was ist Matplotlib
- / Installation
- / Grundlegende Plot Typen
- / Grundlegende Anpassungen

Was ist Matplotlib

- / Weit verbreitete Python Bibliothek zur Erstellung von statischen, animierten und interaktiven Visualisierungen.
- / Nützlich für Erstellung von Diagrammen und Grafiken zur Datenanalyse
- / Anpassbar und erweiterbar

Installation von Matplotlib

- / Öffne Console
- / `pip install matplotlib`

Grundlegende Plot Typen

/ **Line Plot:**

- / Nützlich, um Datenpunkte über eine kontinuierliche Skala hinweg zu verbinden und Trends oder Veränderungen darzustellen. (matplotlib_line.py)

/ **Scatter Plot:**

- / Nützlich, um die Beziehung zwischen zwei Variablen zu visualisieren. Datenpunkte werden als Punkte dargestellt. (matplotlib_scatter.py)

/ **Bar Plot:**

- / Nützlich, um kategoriale Daten zu vergleichen. Die Höhe der Balken repräsentiert den Wert. (matplotlib_bar.py)

/ **Histogramm:**

- / Nützlich, um die Verteilung einer kontinuierlichen Variablen zu visualisieren. Es zeigt, wie viele Datenpunkte in jedem Intervall liegen. (matplotlib_histogramm.py)

Grundlegende Anpassungen

- / Die Diagramme können jeweils angepasst werden, um diese besser Darzustellen.
Dazu gehören:
 - / Titel
 - / Achsen
 - / Legenden
 - / Farben

- / Siehe „matplotlib_anpassungen.py“

Einzelaufgabe (20 Min)

/ Erstelle ein Line Plot mit folgenden DataFrame

```
# Beispiel-Daten erstellen
data = {
    'Year': [2015, 2016, 2017, 2018, 2019, 2020],
    'Sales': [250, 300, 350, 400, 450, 500]
}
```

/ Füge Titel, Achsenbeschriftungen und eine Legende hinzu

/ Erstelle nun ein Scatter Plot und passe die Punktgrösse und –farbe an

Gruppenaufgabe Datenanalyse mit Pandas (30 Min / 2 Personen)

- / Gegeben sind «sales.csv» und «products.csv»
- / Beantworten sie folgende Fragestellungen:
 - / 1. Welches Produkt hat den höchsten Gesamtumsatz erzielt?
 - / 2. Wie viele Bestellungen wurden pro Kategorie getätigt?
 - / 3. Wie hoch ist der durchschnittliche Preis pro Kategorie?
- / Tipp:
 - / Verwenden sie «groupby» und «merge»

Gruppenaufgabe Datenanalyse mit Pandas (40 Min / 2 Personen)

- / Gegeben sind die vorherigen Daten
- / Stellen sie folgende Werte grafisch dar:
 - / Bar Plot: Gesamtumsatz pro Produkt
 - / Histogramm: Verteilung der Verkaufsmenge
 - / Interpretieren sie die Ergebnisse im Plenum