

Python: Pandas

1 Installieren in Anaconda

```
conda install pandas
```

2 Importieren

```
import pandas as pd
```

3 pandas.Series (Array, Spalte einer Tabelle)

3.1 Series-Objekt aus einer Liste erstellen

a) mit index: 0, 1, 2, ...

```
alter = [28, 19, 25, 35, 19]
series = pd.Series(alter)
```

b) mit Labels (index: "Emma", "Anna", "Berta", "Carla", "Dieter")

```
alter = [28, 19, 25, 35, 19]
labels = ["Emma", "Anna", "Berta", "Carla", "Dieter"]
series = pd.Series(alter, index=labels)
```

3.2 Statistik

Beschreibung der Tabelle

```
series.describe()
```

Anzahl unique values

```
series.value_counts()
```

3.3 Sortieren

a) nach Werten

```
series.sort_values()
```

b) nach Index

```
series.sort_index()
```

3.4 Balkendiagramm

```
conda install matplotlib
```

```
series.plot(kind='bar')
```

4 pandas.DataFrame (Tabelle)

4.1 DataFrame erstellen

a) DataFrame aus Dictionary erstellen

```
df = pd.DataFrame({
    'Name': ['Anna', 'Carla', 'Dieter', 'Berta'],
    'Alter': [25, 27, 22, 19],
    'Schuhgroesse': [35, 39, 43, 38],
    'Groesse': [1.59, 1.65, 1.87, 1.70]
})
```

b) DataFrame aus CSV-Datei erstellen

```
df = pd.read_csv('daten.csv')
```

4.2 DataFrame verändern

Eine Spalte als Index auswählen

```
df.set_index('Spalte')
```

Spalte hinzufügen

```
df['Neue-Spalte'] = df['Spalte'] * 2
```

4.3 DataFrame analysieren

Zeilen sortieren

```
df.sort_index()
df.sort_values(by='Spalte')
```

Aggregatfunktionen

```
df.sum()
df.mean()
```

Filter (Zeilen)

```
df.loc['Zeile']
df.loc[['Zeile1', 'Zeile2']]
```

```
df[df['Spalte'] > 21]
df.query(Spalte1 > 21 and Spalte2 == 35)
```

Projektion (Spalten)

```
df['Spalte']
df[['Spalte1', 'Spalte2']]
```

Filter und Projektion

```
df.loc[['Zeile1', 'Zeile2']][['Spalte1', 'Spalte2']]
```

Gruppieren

```
df.groupby('Spalte').sum()
df.groupby('Spalte').mean()
```