

Wirtschaftsinformatik 2

Wochenaufgabe - Woche 5

In dieser Woche geht es um einen Einstieg in die Datenanalyse mit dem Modul Pandas. Dazu haben wir neben dem Kapitel 5.1 (bis Seite 132) des Buchs *Datenanalyse mit Python* noch einen Foliensatz `winf2-05-pandas-series` bereitgestellt, der einen etwas seichteren Einstieg ermöglichen sollte.

Für diese Wochenaufgabe werden allerdings Daten benötigt, die zunächst eingelesen werden müssen. Die Details dazu werden wir erst in Woche 6 behandeln. Es reicht daher, dass Sie den folgenden Code zunächst einfach ausführen und damit dann in der Variablen `solarStrom` das erforderliche Series Objekt erhalten.

```
import pandas as pd

url = 'https://datascience.hs-bochum.de/data/StromDE.csv'
data = pd.read_csv(url)
data.index = pd.to_datetime(data['Datum'])

# Extrahieren der Messwerte fuer produzierten Solarstrom
solarStrom = data['Solarstrom']

# Extrahieren der Messwerte fuer den Gesamtstrom Bedarf
stromBedarf = data['Verbrauch']
```

Damit Sie den Teil nicht abtippen müssen, haben wir auf der Materialseite der Vorlesung das Notebook `winf2-05-wochenaufgabe.ipynb` bereitgestellt, das Sie gerne als Vorlage in ihren Jupyter-Server hochladen können.

Aufgabe 1 (Statistiken mit Pandas Series)

Wenn der obige Code ausgeführt wurde, enthält die Variable `solarStrom` die Messwerte der mit Photovoltaik produzierten Strommengen in MWh in Deutschland pro Tag. Mit den Statistik-Funktionen von Series lassen sich nun eine Vielzahl von Fragen beantworten:

- Wie hoch ist die durchschnittliche Stromerzeugung aus Solarenergie?
- Was ist der Maximalwert?
- An welchem Datum wurde der Maximalwert erreicht?

Schauen Sie sich für die Beantwortung der Fragen gerne die Pandas-Dokumentation zu *Series* an (in den Folien verlinkt). Insbesondere die Method `idxmax()` ist hier hilfreich.

Aufgabe 2 (Rechnen und Statistiken mit Pandas Series)

Mit der produzierten Strommenge aus Solarstrom und dem Strombedarf lässt sich approximativ ableiten, wie hoch der Anteil von Solarstrom pro Tag ist.

- Berechnen Sie ein Series-Object **anteil**, das den prozentualen Anteil von Solarstrom am Gesamtstrombedarf enthält.
- Was war die maximale Abdeckung des Strombedarfs durch Solarstrom? An welchem Tag war das?
- Wie groß ist der durchschnittliche Anteil von produziertem Solarstrom am Strombedarf?