

Data Science 2

Übungsblatt 3

Aufgabe 1 (Stromverbrauch)

Der Datensatz zum Stromverbrauch 2019 für Deutschland befindet sich in der Datei

https://data.hsbo.de/strom/DE_Stromverbrauch_2019.csv

In dieser Aufgabe geht es darum, aus dem Datensatz eine Zeitreihe mit den Verbrauchswerten zu extrahieren.

1. Lesen Sie den Datensatz in einen DataFrame ein und schauen Sie sich die verfügbaren Spalten an. Fügen Sie eine Spalte **date_and_time** hinzu, die aus der Spalten **Datum**, einem Leerzeichen und der Spalte **Uhrzeit** besteht.
2. Benutzen Sie die Funktion **pd.to_datetime(...)** um eine Spalte **datetime** mit Zeitstempeln zu erzeugen.
3. Wählen Sie aus dem DataFrame die Spalte **Gesamt [MWh]** als einzelnes **Series** Object aus und setzen Sie den **index** des Series Objektes auf die **datetime** Spalte.
4. Benutzen Sie die **resample** Funktion und **mean()** um Durchschnittswerte pro Tag zu bekommen. Plotten Sie den Verlauf der Durchschnittswerte.
5. Erstellen Sie einen Plot für die Durchschnittswerte für die ersten 14 Tage.
6. Benutzen Sie **rolling** und **mean** um einen gleitenden Durchschnitt über die letzten 14 Tage zu berechnen. Erstellen Sie einen Plot, der ihre zuvor berechneten Tageswerte und den gleitenden Durchschnitt enthält.