

Beschaffung

Was machen Datenwissenschaftler ohne Daten?

In den vergangenen Python Übungen haben wir uns damit begnügt, Daten in einem Programm kodiert abzulegen, mit diesen zu arbeiten und sie innerhalb des Jupyter Notebooks auszugeben. Stellen Sie sich nun aber einen Datensatz vor, der tausende oder gar Millionen von Einträgen enthält. Vermutlich sind Sie nicht gewillt, ein Programm zu bearbeiten, bei dem Sie über die erste Million Codezeilen hinweg scrollen müssen, bevor Sie die eigentliche Funktion finden.

Offensichtlich gibt es also Möglichkeiten, Daten von verschiedenen Orten zu beschaffen. Das trennt nicht nur unseren Programmcode von den Daten, sondern ermöglicht auch die Anbindung unterschiedlicher Quellen, Formate und Protokolle. Dieses Kapitel beschäftigt sich daher mit verschiedenen Arten, Daten in Python zu laden, in andere Formate umzuwandeln und auch wieder abzuspeichern.