

Datenanalyse mit Python und Pandas

Daten systematisch analysieren und mit Pandas effizient auswerten

01 INTRO

Python und Pandas gehören zu den wichtigsten Werkzeugen für die praktische Datenanalyse. Daten lassen sich aus unterschiedlichen Quellen übernehmen, strukturieren, filtern, transformieren, zusammenführen und mit vergleichsweise wenig Code auch bei größeren Datenbeständen systematisch untersuchen.

Der Kurs vermittelt den praktischen Umgang mit Pandas entlang eines vollständigen Analyseprozesses. Im Mittelpunkt stehen DataFrames, Auswahl und Transformation von Daten, Aggregationen, Zeitreihen und die Verbindung unterschiedlicher Datenbestände. Die Teilnehmer arbeiten durchgehend mit realistischen Datensätzen und entwickeln reproduzierbare Analysen in Python.

DAUER

3 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Grundkenntnisse in Python; Pandas-Kenntnisse sind nicht erforderlich

02 TRAININGSPROGRAMM

Inhalte

Python als Werkzeug für die Datenanalyse

- typischer Workflow einer Datenanalyse
- Python-Ökosystem für Data Analytics
- Jupyter Notebook und Entwicklungsumgebungen
- NumPy, Pandas und ergänzende Bibliotheken
- reproduzierbare Analyseabläufe

Pandas und das DataFrame-Modell

- Series und DataFrame
- Index und Spalten
- Datentypen
- DataFrames erzeugen
- grundlegende Operationen
- Datenbestände inspizieren

Daten importieren und exportieren

- CSV und Textdateien
- Excel
- JSON
- SQL-Datenquellen
- Importparameter
- Zeichensätze und Trennzeichen
- Export von Analyseergebnissen

Daten auswählen und filtern

- Spalten und Zeilen auswählen
- loc und iloc
- boolesche Filter
- kombinierte Bedingungen
- Sortierung
- Sampling
- Abfragen mit query

Daten transformieren

- neue Spalten berechnen
- Werte ersetzen
- Mapping
- String-Operationen
- Datentypen konvertieren
- kategorische Daten
- vektorisierte Operationen

Fehlende und problematische Werte

- fehlende Werte erkennen
- NaN, None und NA
- fehlende Werte behandeln
- Dubletten erkennen
- Wertebereiche prüfen
- Plausibilitätskontrollen

Gruppieren und aggregieren

- groupby
- Aggregationsfunktionen
- mehrere Aggregationen
- benannte Aggregationen
- Transformationen innerhalb von Gruppen
- Pivot-Tabellen
- Kreuztabellen

Datenbestände zusammenführen

- merge
- join
- concat
- Join-Typen
- Schlüssel und Indizes
- Kardinalitäten
- typische Fehler beim Zusammenführen

Datums- und Zeitdaten

- Datumswerte konvertieren
- Pandas-Datetime-Typen
- Komponenten extrahieren
- Zeitdifferenzen
- Zeitreihen
- Resampling
- rollierende Auswertungen

Statistische Auswertungen mit Pandas

- deskriptive Kennzahlen
- Verteilungen
- Quantile
- Korrelationen
- Häufigkeiten
- gruppierte Statistiken
- Ausreißer erkennen

Effiziente Analysen und reproduzierbare Workflows

- vektorisierte Berechnungen
- apply sinnvoll einsetzen
- typische Performance-Fallen
- Method Chaining
- Analysecode strukturieren
- Zwischenergebnisse kontrollieren
- Analysen reproduzierbar gestalten

Praktisches Analyseprojekt

- Datensatz einlesen und untersuchen
- Datenqualität beurteilen
- Daten transformieren
- Fragestellungen formulieren
- Kennzahlen entwickeln
- Ergebnisse auswerten und dokumentieren

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	3 Tage
Durchführung	Präsenz oder Live Online, vorzugsweise als Inhouse-Schulung
Schwerpunkt	Inhouse-Schulungen für Unternehmen und Teams
Sprache	Deutsch, englische Fachbegriffe und Dokumentation
Unterlagen	Kursunterlagen, Beispiele und Übungsaufgaben
Technik	Eigener Rechner mit Python, Jupyter und den gängigen Analysebibliotheken

Inhouse-Schulungen

Eigene Datenbestände und typische Fragestellungen können in die Übungen einbezogen werden.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Entwickler, technische Mitarbeiter, Analysten und Fachanwender, die Daten mit Python analysieren und Pandas systematisch einsetzen möchten.

Voraussetzungen

Grundkenntnisse in Python, insbesondere Variablen, Kontrollstrukturen, Funktionen und grundlegende Datenstrukturen.

05 LERNZIELE

Was Sie nach dem Kurs können

Nach dem Kurs können die Teilnehmer:

- Daten mit Pandas einlesen und strukturieren
- DataFrames sicher auswählen, filtern und transformieren
- Datenbestände aggregieren und miteinander verbinden
- Datums- und Zeitdaten auswerten
- typische Datenqualitätsprobleme erkennen
- statistische Grundausswertungen durchführen
- reproduzierbare Datenanalysen mit Python erstellen

06 FAQ

Häufige Fragen

Ist der Kurs eine Einführung in Python?

Nein. Die für Pandas benötigten Python-Konzepte werden bei Bedarf erläutert, grundlegende Python-Kenntnisse werden jedoch vorausgesetzt.

Was unterscheidet den Kurs von Datenaufbereitung und Data Cleaning?

Dieser Kurs behandelt Pandas als umfassendes Werkzeug für die Datenanalyse. Der Cleaning-Kurs konzentriert sich speziell auf Datenqualität, Bereinigung, Transformation und die systematische Aufbereitung problematischer Datenbestände.

Werden reale Daten verwendet?

Ja. Die Übungen orientieren sich an typischen Datenbeständen und Fragestellungen aus der Praxis.

Werden auch Datenbanken behandelt?

SQL-Daten können als Datenquelle verwendet werden. Datenbankentwicklung und SQL selbst sind jedoch nicht Schwerpunkt des Kurses.

UC IT Service · Ulrich Cuber · kontakt@uc-it.de

<https://www.uc-it.de/coaching/datenanalyse-pandas/>