

Explorative Datenanalyse mit Python

Daten untersuchen, Zusammenhänge erkennen und Fragestellungen entwickeln

01 INTRO

Explorative Datenanalyse beginnt häufig dort, wo noch nicht genau feststeht, welche Aussagen ein Datensatz überhaupt ermöglicht. Ziel ist es, Strukturen, Verteilungen, Zusammenhänge und Auffälligkeiten systematisch zu untersuchen und daraus belastbare Fragestellungen oder Hypothesen zu entwickeln.

Der Kurs verbindet deskriptive Statistik, Visualisierung und praktische Analyse mit Python. Im Vordergrund steht nicht die Produktion möglichst vieler Diagramme oder Kennzahlen, sondern deren sinnvolle Auswahl und Interpretation.

DAUER

2 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Python-Grundkenntnisse und grundlegender Umgang mit Pandas

02 TRAININGSPROGRAMM

Inhalte

Aufgabe der explorativen Datenanalyse

- EDA im Analyseprozess
- Exploration versus Bestätigung
- Fragestellungen entwickeln
- Hypothesen
- Daten und Kontext
- typische Denkfehler

Datensatz verstehen

- Beobachtungen und Merkmale
- numerische und kategoriale Variablen
- Skalenniveaus
- Datentypen
- Datenherkunft
- Stichprobe und Grundgesamtheit

Datenqualität vor der Analyse

- fehlende Werte
- Dubletten
- Wertebereiche
- Plausibilität
- Ausreißer
- mögliche Verzerrungen
- Grenzen des Datensatzes

Univariate Analyse

- Häufigkeiten
- Lageparameter
- Streuung
- Quantile
- Verteilungen
- Schiefe
- robuste Kennzahlen

Verteilungen visualisieren

- Histogramme
- Boxplots
- Dichte
- Balkendiagramme
- empirische Verteilungen
- sinnvolle Skalierung

Beziehungen zwischen Variablen

- numerisch versus numerisch
- kategorial versus numerisch
- kategorial versus kategorial
- Scatterplots
- gruppierte Auswertungen
- Kreuztabellen

Korrelation und Zusammenhang

- Kovarianz
- Korrelation
- Pearson und Rangkorrelation
- Korrelationsmatrizen
- nichtlineare Zusammenhänge
- Korrelation versus Kausalität

Gruppen und Segmente untersuchen

- Gruppierung
- Vergleich von Teilpopulationen
- Aggregationen
- Verteilungen innerhalb von Gruppen
- auffällige Segmente
- Simpson-Paradoxon als Warnsignal

Ausreißer und ungewöhnliche Beobachtungen

- statistische Verfahren
- visuelle Erkennung
- multivariate Auffälligkeiten
- Datenfehler versus interessante Beobachtung
- Einfluss auf Kennzahlen

Zeitliche Strukturen

- Trends
- saisonale Muster
- Veränderungen
- rollierende Kennzahlen
- Vergleich von Zeiträumen
- Ereignisse und Strukturbrüche

Von Beobachtungen zu Hypothesen

- Muster formulieren
- alternative Erklärungen
- Confounder
- Scheinkorrelation
- Auswahlverzerrung
- Grenzen explorativer Aussagen
- nächste Analyseschritte ableiten

Exploratives Analyseprojekt

- unbekannten Datensatz erschließen
- Analysefragen entwickeln
- geeignete Kennzahlen auswählen
- Visualisierungen einsetzen
- Auffälligkeiten untersuchen
- Ergebnisse und Unsicherheiten kommunizieren

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	2 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 Fragestellungen können als Grundlage für die Exploration verwendet werden.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Data Analysts, Entwickler, technische Fachkräfte und Anwender, die Daten nicht nur auswerten, sondern systematisch untersuchen und interpretieren möchten.

Voraussetzungen

Python-Grundkenntnisse sowie grundlegende Erfahrung mit Pandas. Tiefergehende Statistikkenntnisse sind nicht erforderlich.

05 LERNZIELE

Was Sie nach dem Kurs können

Die Teilnehmer können Datensätze strukturiert explorieren, geeignete statistische Kennzahlen und Visualisierungen auswählen, Zusammenhänge und Auffälligkeiten erkennen sowie Ergebnisse kritisch interpretieren.

06 FAQ

Häufige Fragen

Ist das ein Statistikkurs?

Statistische Grundlagen werden dort behandelt, wo sie für die Exploration benötigt werden. Der Schwerpunkt liegt auf praktischer Datenanalyse und Interpretation.

Was unterscheidet den Kurs von der Datenvisualisierung?

Hier ist Visualisierung ein Werkzeug zur Untersuchung von Daten. Der Visualisierungskurs behandelt die Darstellung selbst: Auswahl, Gestaltung und Kommunikation geeigneter visueller Darstellungen.

Wird Machine Learning behandelt?

Nein. Explorative Datenanalyse kann die Grundlage für spätere ML-Verfahren bilden, ist aber ein eigenständiger Analyseschritt.