



Agenda - Beispiel

Dies ist unsere empfohlene Agenda für das Thema Data Science. Wir verfügen darüber hinaus über weiteres Kursmaterial, um Themen einfach auszutauschen.

Definition Data Science

- Ziele
- Methoden
- Abgrenzung zu anderen Disziplinen

Einrichten und Erweitern von Python

- Unterschied Python 2.x und 3.x
- Anaconda / Conda environments
- Installing Packages
- Jupyter Notebooks

Grundlagen der Sprache und Basisfunktionen

- Datentypen
- Comprehension
- Bedingungen
- Kontrollfluss
- Definition von Funktionen
- Dataframes und Manipulation von Daten mit dem Modul pandas
- Data-In – von Datenbank zu pandas
- Slicing und Filterung von dataframes
- Sortierung
- Umgang und Erkennung fehlender Werte
- Umgang und Erkennung Duplikate
- Zusammenführen mehrerer Datensätze
- Einbinden externer Datenquellen

Erweiterte Funktionen

- Datenbanken
- APIs
- Basic Webscraping
- Daten Visualisieren
- Grundlagen von matplotlib
- Grafiken mit einer Variablen
- Grafiken mit mehreren Variablen mithilfe von Farben und Formen
- Deskriptive Statistik mit pandas
- Einfache eindimensionale Statistiken
- Bedingte statistiken
- Statistische Verteilungen
- Mehrdimensionale Statistiken
- Prädiktive Modellierung und Machine Learning mit dem Modul scikit-learn
- Einführung
- Varianten von Machine Learning Problemen
- Varianten von Machine Learning Algorithmen
- Holdout Strategien
- Feature Engineering
- Feature Importance
- Ensembles
- KPIs und Metriken

Use-case: Eigener oder bereitgestellter Datensatz