

# Analiza danych w Python

## Wstęp z użyciem języka Python i jego bibliotek

---

### Cele szkolenia

Celem szkolenia jest zapoznanie się z podstawami analizy danych opartych na dużych zbiorach danych oraz ich implementacją w języku Python (za pomocą bibliotek tego języka). Szkolenie jest przeglądem możliwości języka Python pozwalających na optymalną pracę z Big Data i Data Science.

### Umiejętności

Dzięki szkoleniu uczestnik dowie się:

- Co to jest Big Data.
- Jak zbierać i składować dane
- Jak używać języka Python do napisania programów analitycznych
- W jaki sposób przetwarzać duże dane: biblioteki: numpy, pandas, scipy
- W jaki sposób wizualizować dane: biblioteki:; matplotlib, seaborn

### Profil uczestników

Szkolenie przeznaczone jest dla analityków danych, architektów struktur danych, programistów oraz osób biorących udział w projektach związanych z Big Data.

### Przygotowanie uczestników

Od uczestników szkolenia wymagana są podstawowej wiedzy na temat danych oraz ich analizy a także podstawowej wiedzy na temat języka Python.

### Szczegółowy program szkolenia

1. Sesja wprowadzająca do big data
  - Definicja big data i jej znaczenie w dzisiejszym świecie.
  - Wyzwania i możliwości związane z analizą dużych zbiorów danych.
  - Przykłady zastosowań big data w różnych dziedzinach.
2. Podstawy statystyki
  - Podstawowe pojęcia statystyczne
  - Podstawowe miary rozkładu
3. Python - wybrane elementy
4. Analiza z NumPy i Pandas

- Tablice jedno i dwuwymiarowe w NumPy oraz podstawowe operacje
  - Series i DataFrame w Pandas
  - Wczytywanie i zapis danych w różnych formatach
  - Podstawowe atrybuty DataFrame (shape, indeksy, kolumny, typy danych)
  - Przydatne funkcje (describe, info, head, tail, sample)
  - Czyszczenie wartości zduplikowanych
  - Wartości brakujące - różne podejścia do radzenia sobie z nimi
  - Wykrywanie wartości odstających
  - Sortowanie danych
  - Filtrowanie (loc, iloc, query, where, isin, isnull, notnull)
  - Tabele przestawne
  - Grupowanie
  - Łączenie danych
  - Tworzenie nowych atrybutów
5. Analiza z użyciem Scipy
- Funkcje specjalne
  - Całkowanie
  - Statystyka
  - Analiza struktur wielowymiarowych
6. Wizualizacja danych z Matplotlib i Seaborn
- Rodzaje wykresów i wizualizacji - dobór wykresu, elementy storytelling
  - Podstawowe typy wykresów (histogram, wykres słupkowy, wykres typu scatter) w Pythonie

## Metoda realizacji szkolenia

Szkolenie realizowane jest w formie naprzemiennie następujących po sobie wykładów oraz ćwiczeń praktycznych. Szkolenie łączy w sobie wiedzę merytoryczną z praktycznymi przykładami jej wykorzystania w środowisku pracy.

## Liczba dni, liczba godzin szkoleniowych

3 dni, 24 godziny szkoleniowych