

Program szkolenia:

R dla analityków danych – od podstaw do interaktywnych wizualizacji

Informacje:

Nazwa:	R dla analityków danych – od podstaw do interaktywnych wizualizacji
Kod:	data-science-r-data-science
Kategoria:	Analiza danych
Grupa docelowa:	architekci developerzy analitycy
Czas trwania:	3 dni
Forma:	30% wykłady / 70% warsztaty

Pierwszy dzień szkolenia ma na celu wprowadzenie osób w świat programowania w języku R, poznać ciekawe metody usprawnienia codziennej pracy, oraz rozpocząć przygodę z wizualizacją przetworzonych danych.

Uczestnicy podczas tego dnia nauczą się zacząć pracę z R, dowiesz się jak skonfigurować środowisko by korzystanie z języka było jeszcze łatwiejsze, oraz zostanie dokonany przegląd składni oraz mechanizmów jakie funkcjonują w tym języku.

Wszystko będzie to poparte praktycznymi przykładami wraz z dobrymi praktykami które otworzą Ci furtkę do dalszej nauki języka R.

Kolejne dni mają na celu wprowadzenie programisty w świat praktycznego zastosowania R. W tej części nie skupimy się na samej składni języka ale głównie na tym jak i gdzie zastosować R, oraz jak może on nam pomóc w pracy codziennej.

Zostaną przedstawione możliwości wizualizacji danych w R, przygotowywanie dynamicznych dokumentów a nawet prezentacji multimedialnych.

Zalety szkolenia:

- Przykłady realnych problemów
- Świadomy dobór klasy modelu do klasy problemu
- Rzetelny fundament podstaw analizy danych

Szczegółowy program:

1. Wprowadzenie do środowiska R

- 1.1. Zastosowanie i użytkowni R - wprowadzenie
- 1.2. Rozwój zastosowań komercyjnych R
- 1.3. Instalacja środowiska R + Rstudio
- 1.4. Graficzny interfejs użytkownika w Rstudio lub R tools for Visual studio
- 1.5. Korzystanie z wbudowanych funkcji (R jako kalkulator)
- 1.6. Omówienie działania pakietów

2. Omówienie sposobów organizacji danych w R

- 2.1. Przegląd typów danych - co warto? czego unikać?
- 2.2. Struktury homogeniczne i heterogeniczne
- 2.3. Metody agregacji danych - gdzie wydajniej?

3. Praca ze źródłami zewnętrznymi

- 3.1. Import i eksport plików w środowisku R do różnych formatów
- 3.2. Konfiguracja połączeń do baz danych
- 3.3. Zagnieżdżanie R w procedurach na bazie SQL Server
- 3.4. Zmiana kontekstu przetwarzania danych - praktyczny przykład z SQL Server'em 2017

4. Mini case study (30 minut)

5. Funkcje oraz instrukcje sterujące

- 5.1. Tworzenie własnych funkcji
- 5.2. Wyrażenia logiczne z If - else oraz switch
- 5.3. Wykorzystanie przetwarzania iteracyjnego
- 5.4. Wybór optymalnych funkcji do postawionych problemów (Repeat,While,For,Apply)

6. Operacje z tekstem

6.1. Formatowanie tekstu

6.2. Przedstawienie użytecznych funkcji do pracy z danymi tekstowymi

6.3. Wyrażenia regularne oraz parsowanie tekstu

7. Od danych surowych do wartościowych

7.1. Czyszczenie danych

7.2. Przygotowanie danych do analizy

7.3. Wykrywanie wartości odstających

8. Zaczynamy od plot'a

8.1. Wykres słupkowy

8.2. Wykres liniowy

8.3. Edycja osi oraz możliwości parametryzacji.

9. Wizualizacje doładowane mocą pakietów

9.1. Wizualizacja danych z pakietem GGLOT2

9.2. Wizualizacja zależności (korelacji) z pakietem corrplot

9.3. Prezentacja danych na interaktywnych mapach

10. Wykorzystanie pakietu Rmarkdown

10.1. Dynamiczne dokumenty

10.2. Prezentacja multimedialna w R

10.3. Publikacja dynamicznych dokumentów

11. Interaktywne wizualizacje

11.1. Budowa dashboardu z pakietem Shiny

11.2. Przegląd możliwości pakietu Plotly