

Praca domowa numer 2

11 grudnia 2021

Firma, dla której pracujesz jako analityk wygrała przetarg na przeprowadzenie badania na rzecz firmy ubezpieczeniowej dotyczącego ryzyka wystąpienia trzęsienia ziemi i jego szkodowości. Twoim celem jest wykorzystanie analizy skupień (taką technikę badawczą zaproponowała Twoja firma) do wyznaczenia obszarów w którym występuje nasilenie zjawiska trzęsienia ziemi, oraz wskazanie takich rejonów w których potencjalne szkody materialne i ludzkie będą największe.

Zleceniodawca dostarczył bazę danych *Global Significant Earthquake Database* opracowaną przez *National Centers for Environmental Information* obejmującą znaczące trzęsienia ziemi od roku 2250 przed naszą erą do dnia dzisiejszego. Znajdziesz ją w pliku *earthquakes.csv*.

Po załadowaniu zbioru danych ustaw ziarno generatora liczb losowych (ang. *seed*) na wartość odpowiadającą numerowi Twojego albumu, oraz wylosuj 95% podpróbę danych.

Zbiór danych jest gotowy do analizy. Rozpocznij pracę od zapoznania się z danymi oraz uporządkowania bazy danych. Zwróć uwagę, iż niektóre obserwacje nie są kompletne.

Jako metodę analizy wykorzystaj jeden z poznanych wariantów analizy skupień.

Opracowanie powinno mieć formę raportu. Jesteś zobowiązana/zobowiązany dostarczyć raport w formie drukowanej (wydruk dwustronny) oraz elektronicznej (akceptowane formaty .pdf, ewentualnie .doc). Nieprzekraczalny termin dostarczenia raportu to 14 stycznia 2022 godzina 18.00. Jeśli go nie dotrzymasz

zostaniesz zwolniona/-ny z pracy¹. Prace należy przesłać na adres elektroniczny pstrawinski@wne.uw.edu.pl, a wersję drukowaną pozostawić na portierni budynku 00-241 Warszawa, Długa 44/50 wejście od ulicy Długiej. Opracowanie należy opatrzyć imieniem, nazwiskiem i numerem indeksu autora. Opracowania anonimowe nie będą brane pod uwagę².

Wskazówka: Zamawiający oceniając wartość merytoryczną opracowania będzie przede wszystkim zwracał uwagę na argumentację uzasadniającą podjęte w trakcie analizy decyzje. Ważne decyzje powinny zostać zilustrowane odpowiednimi wykresami i/lub wartościami stosownych statystyk.

Nie ma błędnych odpowiedzi są mniej lub bardziej trafne.

Limit długości tekstu: 15000 znaków

¹Praca zostanie uznana niewykonaną

²Praca zostanie uznana niewykonaną