

Dziennik projektowy

Programowanie w R

Inżynieria i analiza danych

Zespół:

Martyna Bielenda

Jakub Chorzepa

Damian Kobyliński

1. Pobranie danych z WHO

- Niestety zbyt dużo dawno nieuzupełnianych danych lub ich brak. Mało zaburzeń/chorób do wyboru, osobne bazy danych dla każdego państwa.

2. Pobranie danych z IHME (Institute for Health Metrics and Evaluation-baza danych prowadzona przez Uniwersytet w Waszyngtonie) z wybranymi latami, chorobami, regionami oraz przedziałami wiekowymi.

3. Wstępna obróbka-rozdzielenie i przefiltrowanie, oraz analiza danych w Excelu

- Sprawdzenie wiarygodności i ilości danych, zaplanowanie podziałów.

4. Załadowanie pliku z gotową bazą danych do RStudio.

5. Podział ramki na występowanie według regionu oraz przedstawienie tych danych na osobnych wykresach słupowych dla każdego regionu

6. Pogrupowanie ramki z występowaniami zaburzeń według przedziałów wiekowych, utworzenie czterech badanych kategorii wiekowych: 10 – 19, 20 – 29, 30 – 39, 40 – 49, wyświetlenie danych na wykresie radarowym

7. Utworzenie prognozy występowania danych chorób w 2020 roku

- Ze względu na predykcje wykorzystując regresję liniową, w wyniku wyszły niepoprawne dane (z danych wynika, że stwierdzonych zaburzeń jest więcej niż ludzi na ziemi)

8. Podział ramki na śmierci według regionu oraz przedstawienie tych danych na wykresach

9. Podział ramki występowania według lat

- Zliczenie ich występowania oraz przedstawienie ich na wykresie liniowym

10. Pogrupowanie ramki występowania zaburzeń ze względu na płeć oraz przedstawienie tych danych na dwóch wykresach słupowych

11. Analiza wynikowych wykresów i danych

12. Stworzenie plakatu z rysunkami oraz najważniejszymi wykresami z opisami