

SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 3 Data 18.10.2024 Temat: "Wykorzystanie pakietu Pandas do manipulacji i przetwarzania danych w Pythonie" Wariant 1	Imię Nawisko: Mariusz Jagosz Informatyka II stopień, niestacjonarne, 1 semestr, gr.A
---	---

1. Polecenie: wariant 1 zadania

Dane znajdują się pod linkiem: Development Assistance for Health Database 1990-2020 <http://ghdx.healthdata.org/record/ihme-data/development-assistance-health-database-1990-2020>

2. Opis programu opracowanego (kody źródłowe, rzuty ekranu)

Kod znajduje się na repozytorium pod linkiem: [unibb/Nauka o Danych I/Task3 at main · mariuszjagosz/unibb](https://github.com/mariuszjagosz/unibb-Nauka-o-Danych-I-Task3-at-main)

Kod został również przesłany na platformę elearningową

W celu wykonania kodu należy pobrać dane z powyższego linku i umieścić plik CSV w tym samym folderze, w którym znajduje się notebook Jupyter.

3. Wnioski

Wnioski z analizy datasetu DAH (1990-2020):

1. **Średnia i rozrzut DAH:** Średnie wsparcie wynosi 2519,36 k\$, z dużym rozrzutem danych i niską medianą (44 k\$), co sugeruje asymetrię – większość krajów otrzymuje niewielkie wsparcie, podczas gdy nieliczne mają wartości znacznie wyższe.

2. **Brakujące dane:** W około 14% rekordów brakuje danych, co może wpływać na wyniki.
3. **Wartości odstające:** Występują znaczne wartości odstające, co wymaga dodatkowej uwagi przy analizie.
4. **Korelacja z HIV DAH:** Silna dodatnia korelacja (0,83) między ogólnym DAH a wsparciem na HIV wskazuje, że większe wsparcie ogólne idzie w parze z pomocą na HIV.

Ogólnie, dane wymagają dalszego oczyszczenia, a analiza potwierdza duże rozbieżności w poziomie wsparcia dla różnych krajów i obszarów.