

SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 1 Data 28.09.2024 Temat: "Ustalenia platformy Jupyter. Użycie biblioteki pandas" Wariant 1	Imię i Nazwisko: Mariusz Jagosz Informatyka II stopień, niestacjonarne, 1 semestr, gr.A
---	--

1. Polecenie: wariant 1 zadania

Zadanie dotyczy pobrania danych z pliku, tworzenia ramki danych, wykonania poszczególnych zadań poniżej na podstawie odpowiedniego zbioru danych:

Mój wariant zadania: Wariant 1

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:

Kod znajduje się na repozytorium pod linkiem: [unibb/Nauka o Danych I/Task1 at main · mariuszjagosz/unibb \(github.com\)](https://unibb.github.io/Task1-at-main-mariuszjagosz-unibb-github-com/)

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

Dzięki bibliotece pandas możemy przetwarzać duże zbiory danych, dzięki takiej strukturze jak DataFrame. Używając tej biblioteki można ładować i

zapisywać dane w różnej formie, takiej jak pliki csv czy Excel. Pandas posiada wiele metod, które pozwalają w relatywnie prosty sposób manipulować danymi, czyścić dane i je transformować. Pozwala również na szybki podgląd danych statystycznych za pomocą metod takich jak „describe()”. Dodatkowo używając biblioteki Matplotlib wraz z biblioteką pandas możemy wizualizować dane tworząc wykresy różnego typu. Działając na bardzo dużych plikach możemy użyć argumentu chunksize, który dzieli operacje na mniejsze kawałki.