Analiza i Wizualizacja Danych w Pythonie – Laboratorium

Ćwiczenie 1: Ładowanie i prezentacja danych

1. Napisać program, który umożliwi:
    a. Ładowanie do zmiennych danych z plików **csv**, **xlsx**, **txt**,
    b. Wyświetlanie załadowanych zbiorów danych
    c. Zapisywanie przetworzonych zbiorów do formatów **csv**, **xlsx**, **txt**

    Ścieżki do plików lub ich nazwy powinny być podawane jako parametry wywołania programu lub w wprowadzane przez użytkownika w wyniku interakcji z programem. Podobnie, powinna być możliwość podania, czy zbiory są etykietowane, a jeśli tak, która kolumna zawiera etykiety. Sprawdzić poprawność załadowanych danych(np. wartości `NaN`)
2. Napisać kod, który umożliwi uzyskanie następujących informacji na temat pobranego zbioru danych:
    a. Liczba wierszy i kolumn
    b. W przypadku danych etykietowanych określenie, ile jest różnych kategorii w zbiorze (wyświetlenie tych kategorii)
    c. Sprawdzenie rozłożenia kategorii, tzn. czy są równomiernie reprezentowane w zbiorze (na podstawie liczby przykładów należących do poszczególnych kategorii).
    d. Dla każdego atrybutu obliczenie mediany i wartości średnich.
3. Narysować uporządkowany (rosnąco lub malejąco) ciąg wartości średnich w postaci wykresu słupkowego.
4. Wybrać ze zbioru 20% atrybutów o największych wartościach średnich i zapisać je w postaci nowego zbioru (z etykietami, jeśli występują) do nowego pliku (do plików **csv**, **xlsx**, **txt**).
5. Napisać program, który umożliwi przetestowanie napisanego kodu.
6. Zrealizować zadania z punktów 1-4 w wersji obiektowej (np. klasy `DataLoader` i `DataDescriber`).

Uwaga: w celu realizacji zadania można posłużyć się biblioteką numpy oraz matplotlib.

Źródła wiedzy:

https://matplotlib.org/stable/tutorials/pyplot.html

https://numpy.org/doc/stable/