Autorzy: Tomasz Ancukiewicz, Bartosz Jackowiak, Czcibor Babula, Mateusz Kudła

Opiekun: Fabian Wiktorowski (przedstawiciel firmy Roche)

PROFILOWANIE DANYCH W FORMACIE JSON

1. **Założenia:**
   Głównym zadnie było przygotowanie narzędzia bądź rozwinięcie biblioteki *pandas-profiling* przeznaczonej do przetwarzania plików w formacie *JSON* i wyznaczania występujących w nich charakterystyk.

2. **Wymagania funkcjonalne :**
   - Stworzenie narzędzia od podstaw bądź rozwinięcie istniejącej biblioteki *pandas-profiling*
   - Profilowanie zagnieżdżonych plików w formacie *JSON*
   - Określanie charakterystyk
   - Wykonanie wizualizacji podobnej do generowanej przez bibliotekę *attaccama*

3. **Wykorzystane biblioteki:**
   - *pandas*
   - *pandas-profiling*
   - *statistics*
   - *jinja2*

4. **Realizacja:**
   Opracowane przez nas rozwiązanie zostało zrealizowane zgodnie z wymaganiami, i jest to rozwinięcie biblioteki *pandas-profiling*. Dodatkowo charakterystyki generowane przez nasze rozwiązanie eksportowane są do plików **.html* z podziałem na poszczgólne tabele.

5. **Uruchomienie:**
   W celu uruchomienia skryptu użyta powinna zostać poniższa komenda:

   ```
   ./main.py -i <plik_wejściowy> [-o <folder_wyściowy>] [-s] [-f] [-t]
   ```

   Gdzie, dodatkowe parametry oznaczają odpowiednio:
   - `-i` – plik wejściowy w formacie *JSON*
   - `-o` – katalog wyjściowy zawierający wyniki profilowania (domyśla wartość: Output)
   - `-s` – przeglądarka z wynikami nie otwiera się automatycznie
   - `-f` – uruchomienie normalizacji pliku *JSON* przed utworzeniem encji
   - `-t` – zapisuje utworzone obiekty do katalogu `tables` w katalogu wyjściowym

6. **Ograniczenia:**
   Narzędzie zostało przetestowane na pliku o rozmiarze 50 MB. Z uwagi na fakt, że wygenerowane tabele są przechowywane w pamięci RAM, maksymalny możliwy rozmiar pliku może zmienić w zależności od komputera.