**Podstawy analizy i reprezentacji danych projekt “Grupowanie państw na podstawie wyglądu ich flag”**

Artur Prasuła, Aleksandra Kowalczyk, Kacper Achramowicz

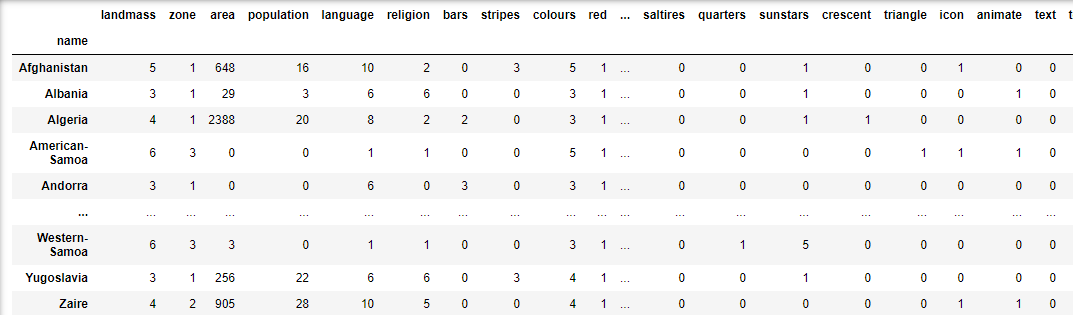
DISCLAIMER: EDYTUJEMY TEN PLIK CZCIONKA CONSTANTIA 14 DO PISANIA (bo lubię ją xD)

WRZUCAM SCREENY WYNIKÓW, A NIE KODU DO EWENTUALNEJ ZMIANY POTEM??

WNIOSKÓW NIE WRZUCAM NA RAZIE, BO W SUMIE NWM CZY MAM, CZY ZOSTAJĄ W JUPYTERZE. PLS ODPOWIEDŹ TUTAJ

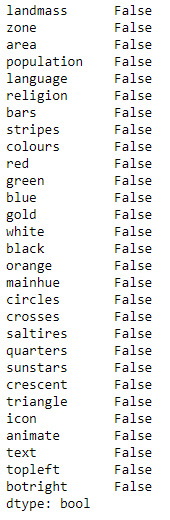
1. **Analiza eksploracyjna zbioru danych**

W celu przygotowania danych do analizy na początku pracy stworzyliśmy ramkę danych korzystając z pliku *flag.data.* Dodane zostały nazwy kolumn, ale też ujednoliciliśmy typ danych w całej ramce, tzn. zamieniliśmy kolory na liczby dziesiętne.



Zrzut Ramka danych zawierający zbiór z pliku

Następnym krokiem było sprawdzenie, czy dane posiadają jakieś braki. Wnioskiem z prostego testu było, że zbiór jest pełny. Wynik pokrył się z umieszczonym na stronie <https://archive.ics.uci.edu/ml/datasets/Flags>.



Zrzut : Test sprawdzający, czy występują puste dane. False oznacza, że dane są kompletne.

Dokonaliśmy również analizy pojedynczych atrybutów, ponieważ mogło okazać się to przydatne w dalszej analizie zbioru, głównie przy wyciąganiu wniosków z histogramu. Dzięki temu łatwiej było stwierdzić najczęściej występującą ilość obiektów poszczególnych atrybutów (np. najbardziej popularną ilość pionowych słupków we fladze).

