SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

|  |  |
| --- | --- |
| Laboratorium Nr 3  Data 19.10.2024  Temat: "Wykorzystanie pakietu Pandas do manipulacji i  przetwarzania danych w Pythonie"  Wariant 10 | Anna Więzik  Informatyka  II stopień, niestacjonarne,  1 semestr, gr.1b |

1. Polecenie:

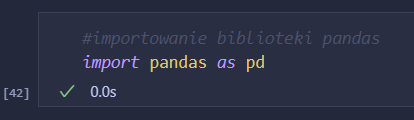
Premise Child Health COVID-19 Health Services Disruption Survey 2020 [http://ghdx.healthdata.org/record/ihme-data/premise-c hild-health-covid-19-health-services-disruption-survey-2 020](http://ghdx.healthdata.org/record/ihme-data/premise-c%20hild-health-covid-19-health-services-disruption-survey-2%20020)

1. Link do repozytorium:

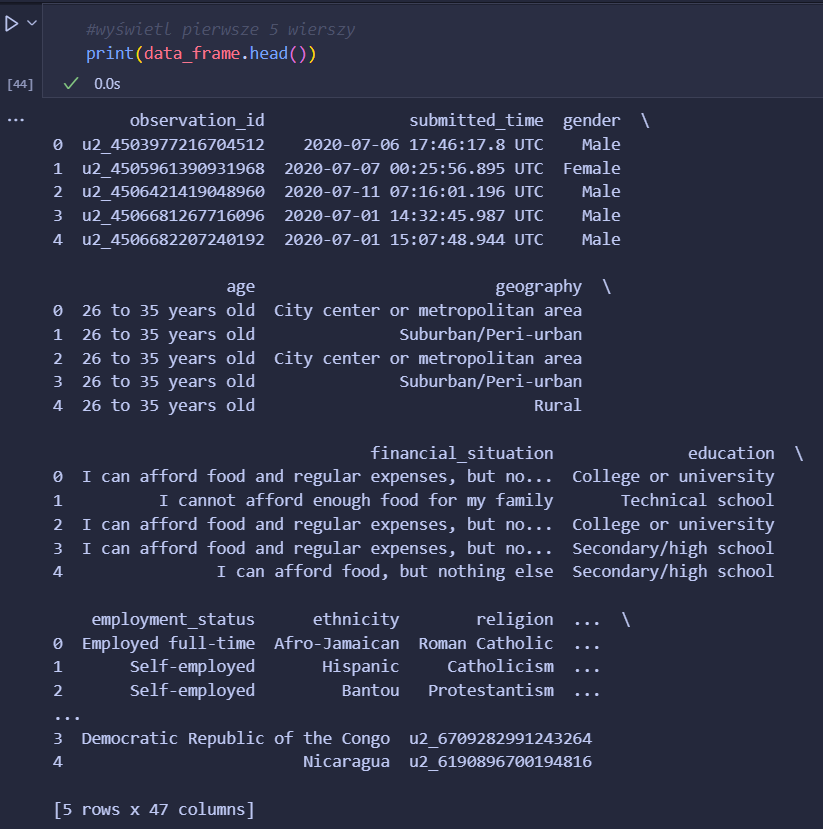
Link: <https://github.com/AnaShiro/NoD1_2024>

1. Opis programu opracowanego

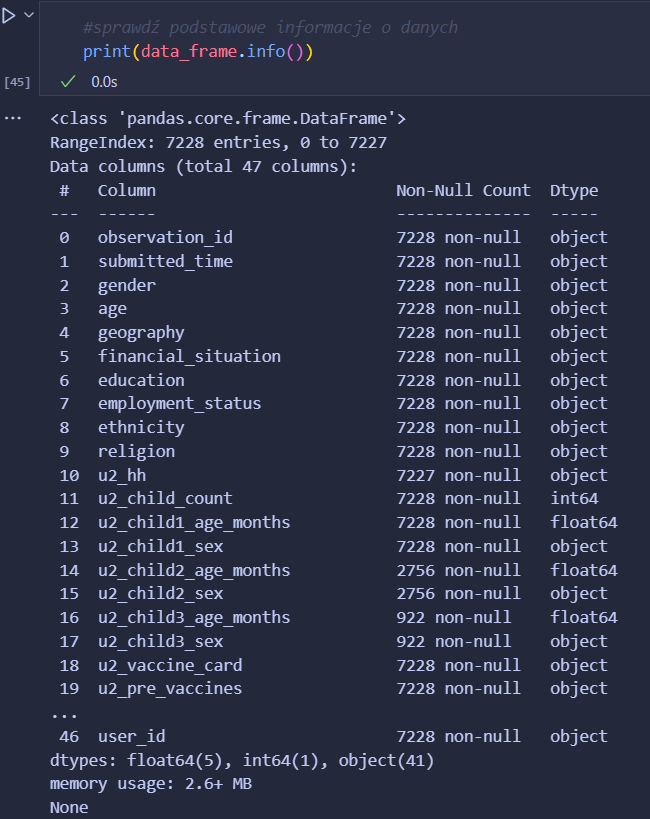
* Wczytywanie danych i wyświetlanie podstawowych informacji
  + Wczytaj dane z pliku



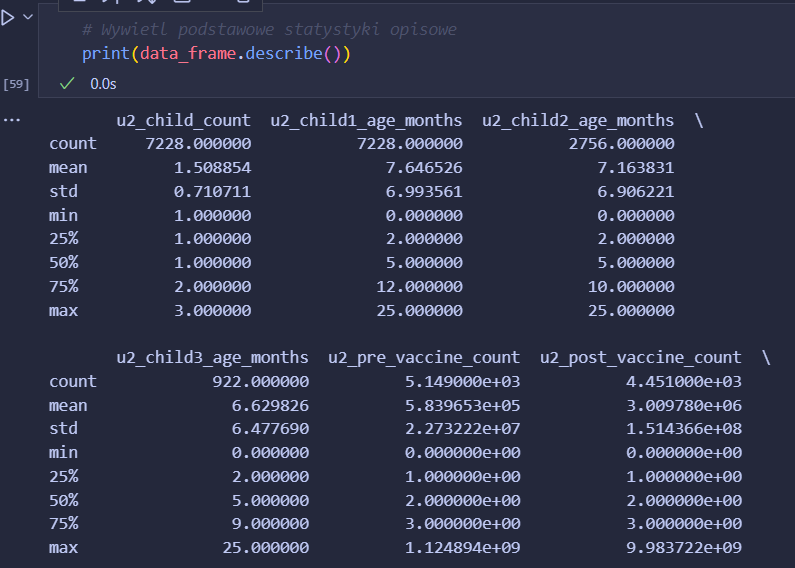
* + Wyświetl pierwsze 5 wierszy



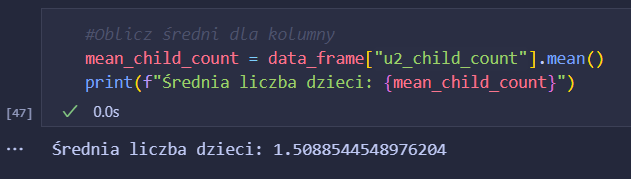
* + Sprawdź podstawowe informacje o danych



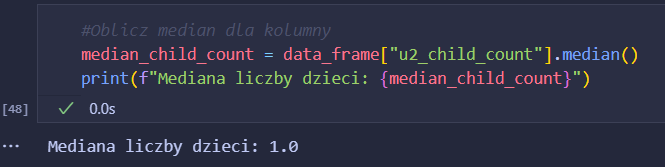
* + Wyświetl podstawowe statystyki opisowe



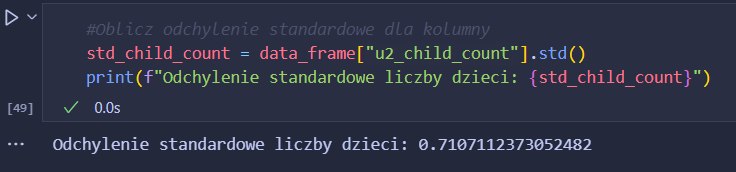
* Obliczanie podstawowych statystyk
  + Oblicz średnią



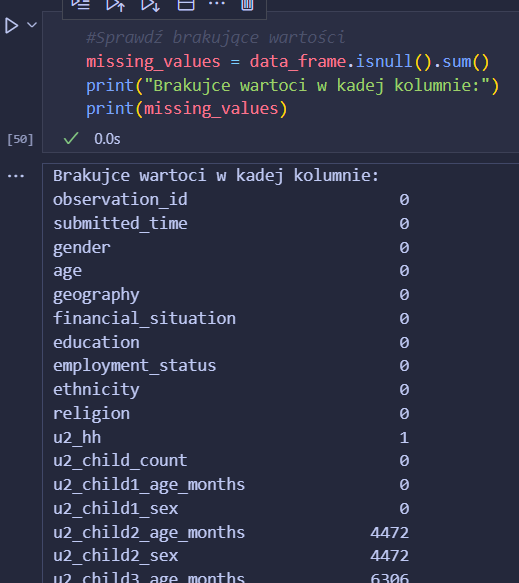
* + Oblicz medianę



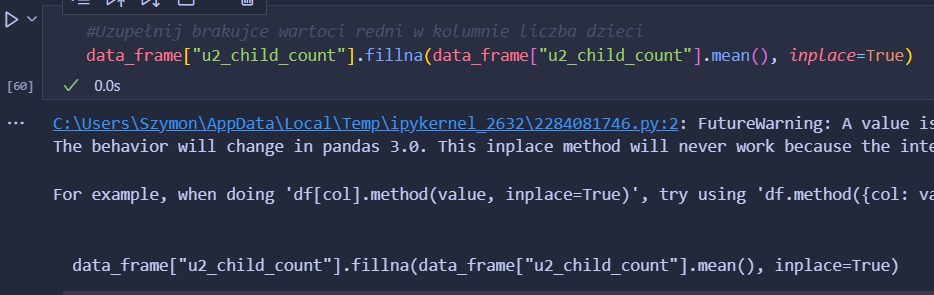
* + Oblicz odchylenie standardowe



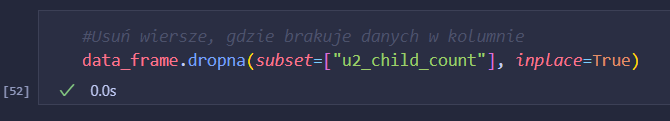
* Identyfikacja i obsługa brakujących danych
  + Sprawdź brakujące wartości



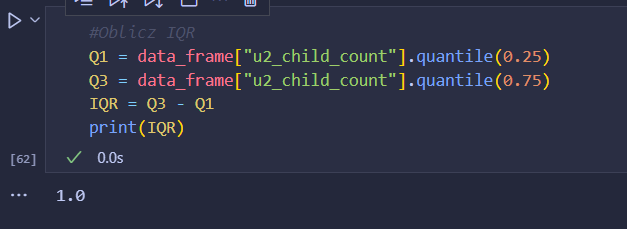
* + Uzupełnij brakujące wartości średnią



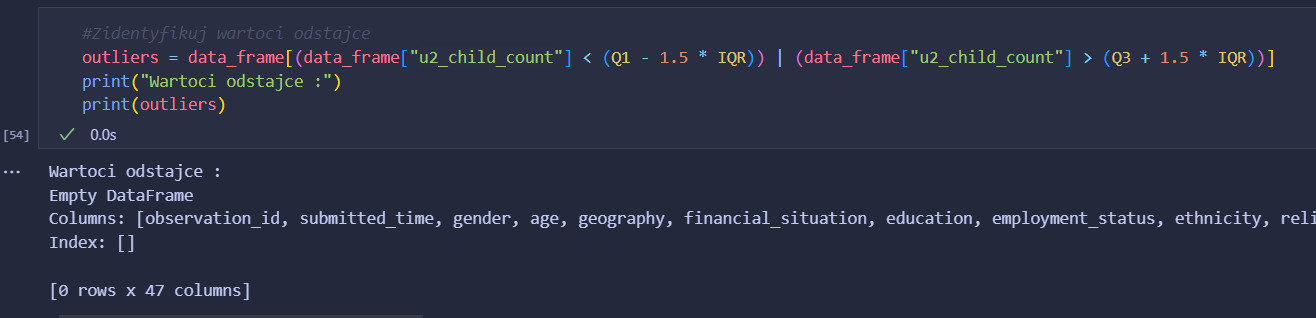
* + Usuń wiersze w których brakuje danych



* Wykrywanie wartości odstających
  + Oblicz IQR



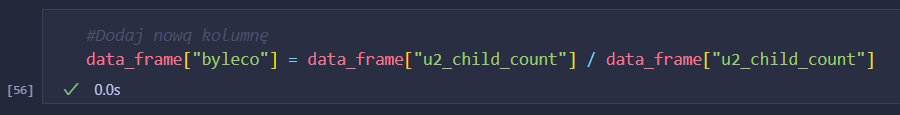
* + Zidentyfikuj wartości odstające



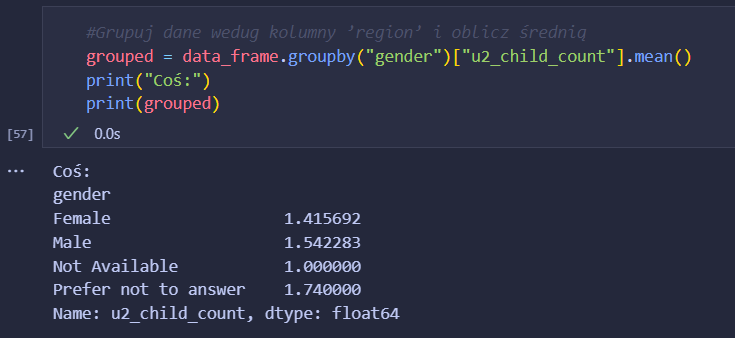
* Analiza zależności między kolumnami
  + Wykonaj wykres rozrzutu



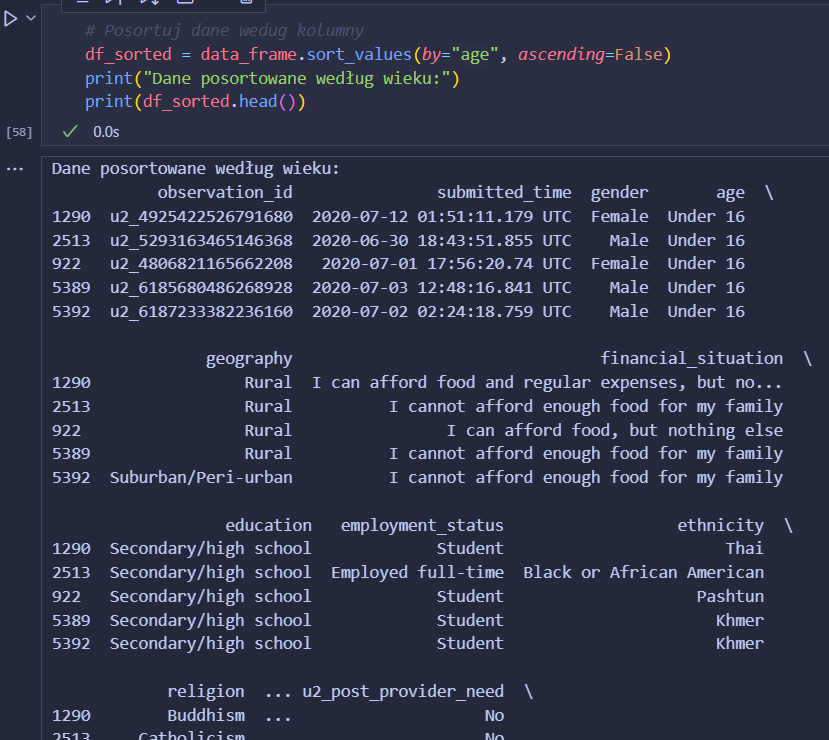
* Przekształcenie danych
  + Dodaj nową kolumnę



* + Grupuj dane według kolumny



* + Posortuj dane według kolumny



1. Wnioski

Zaczęliśmy od wczytania danych z pliku CSV i wyświetlenia podstawowych informacji o zbiorze danych. Następnie obliczyliśmy podstawowe statystyki opisowe dla wybranych kolumn, aby zrozumieć rozkład danych. Brakujące dane mogły wpłynąć na jakość analizy, więc musieliśmy je zidentyfikować i odpowiednio obsłużyć. Wartości odstające mogły zaburzać wyniki analizy, dlatego musieliśmy je zidentyfikować. Zbadaliśmy zależności między różnymi kolumnami poprzez obliczenie współczynników korelacji. Na koniec przekształciliśmy dane, tworząc nowe kolumny, grupując dane i sortując je.

Wykonując te zadania, nauczyliśmy się podstawowych i zaawansowanych technik manipulacji danymi w Pandas. Zrozumieliśmy, jak efektywnie wczytywać, analizować i przekształcać dane, co jest niezbędną umiejętnością w analizie danych i nauce o danych.