VIZUALIZACIJA PODATAKA – DZ2

**Matej Kovačević**

Zadatak: Napraviti jednostavnu vizualizaciju podataka po vlastitom izboru koju je potrebno opisati u formi dokumenta.

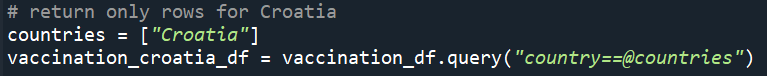
* This example includes the analysis of vaccination against COVID-19 in Croatia, and after processing and analyzing the data, it also enables a visual presentation of the processed data
* the data is extracted from the csv file and processed in the Spyder IDE using the Python programming language and its powerful libraries
* **Process:**

1. Read Csv

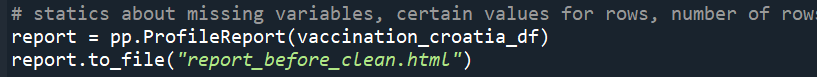


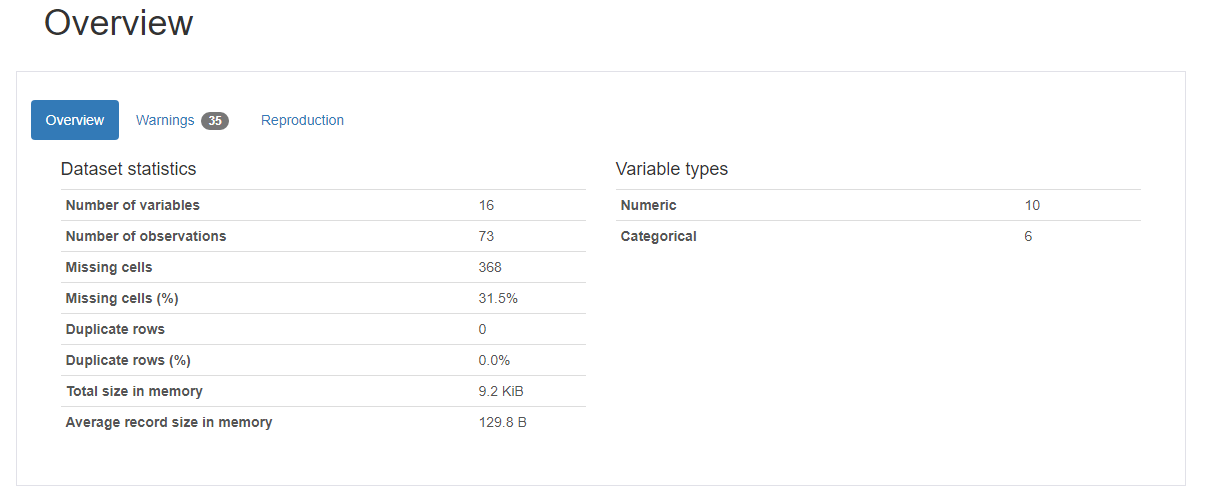
* Csv file contains data for all countries

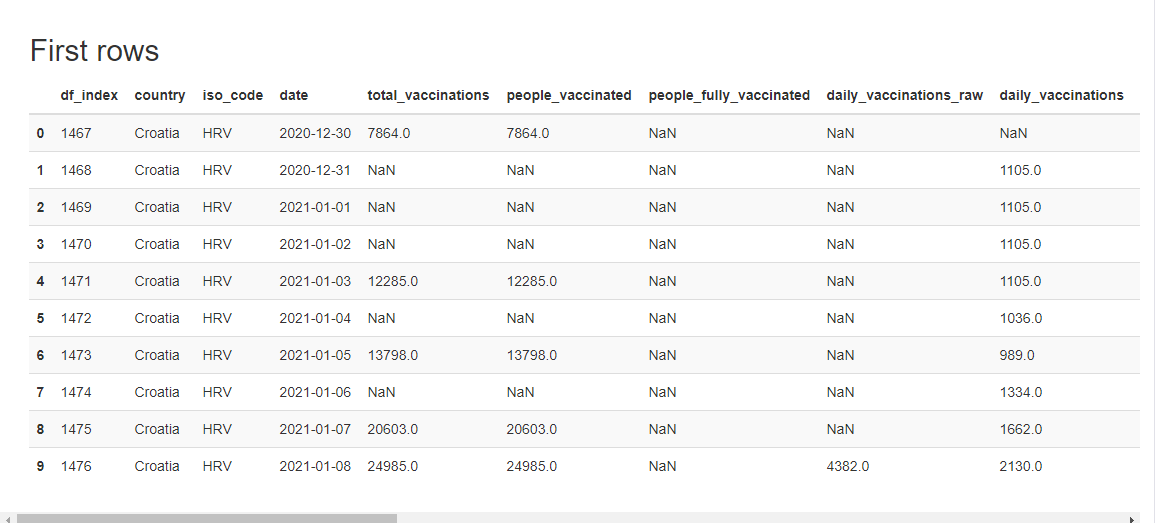
1. Get data for Croatia



1. Get statistics on empty cells, duplicate data, first rows and more



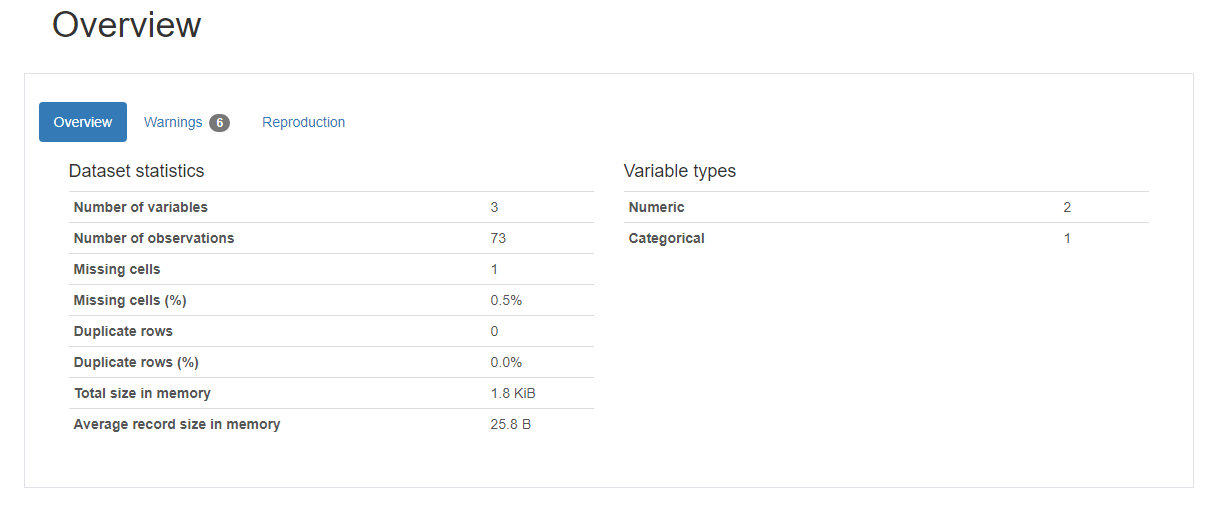




1. First analysis - the number of vaccinations per day in Croatia
2. retrieve only columns of interest



1. get report about data – only one empty cell

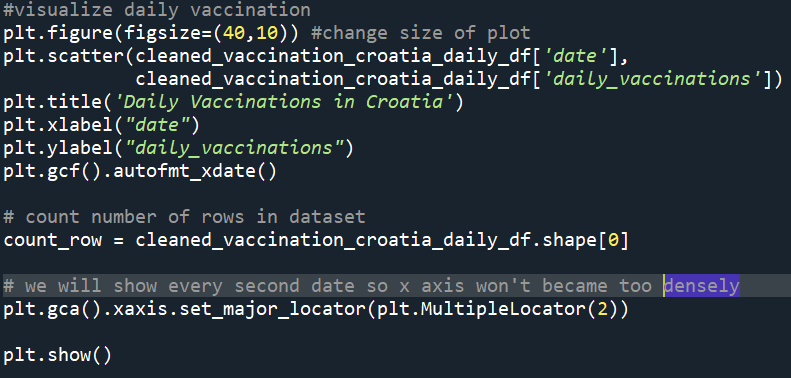


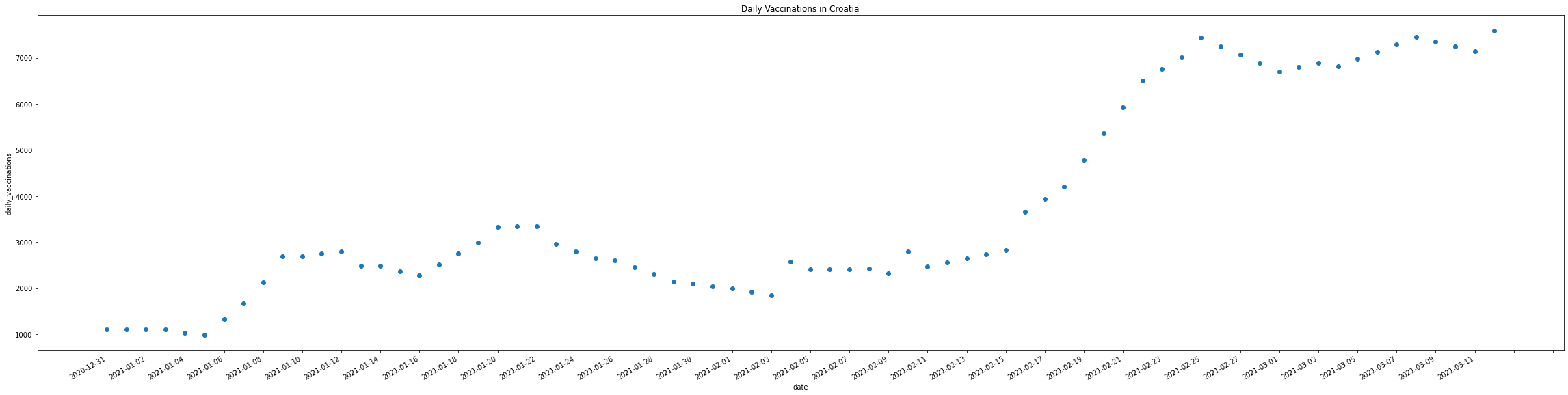


1. data cleaning – droping rows with missing cells



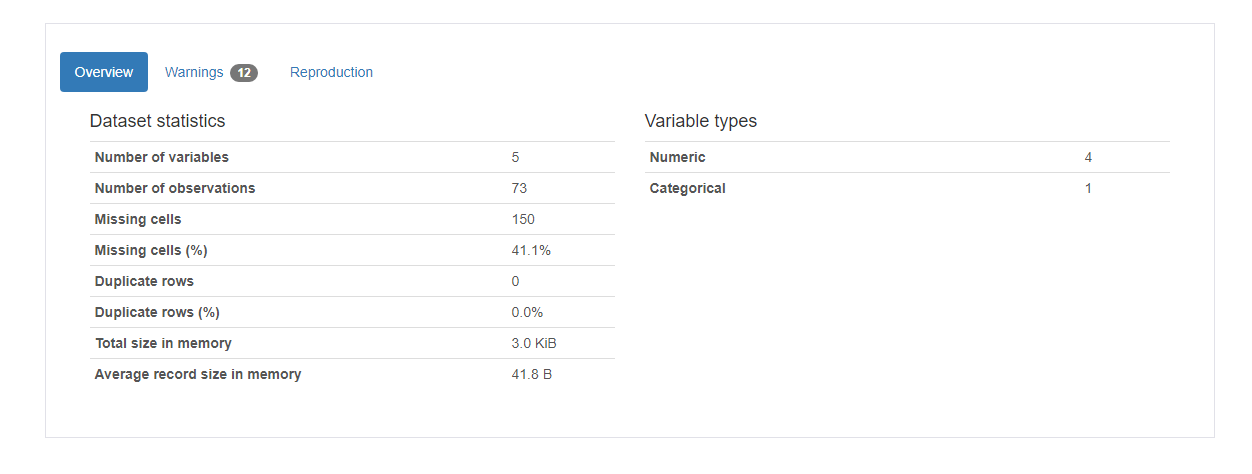
1. Visualize the data – using every second label for date so graph doesn’t look ugly



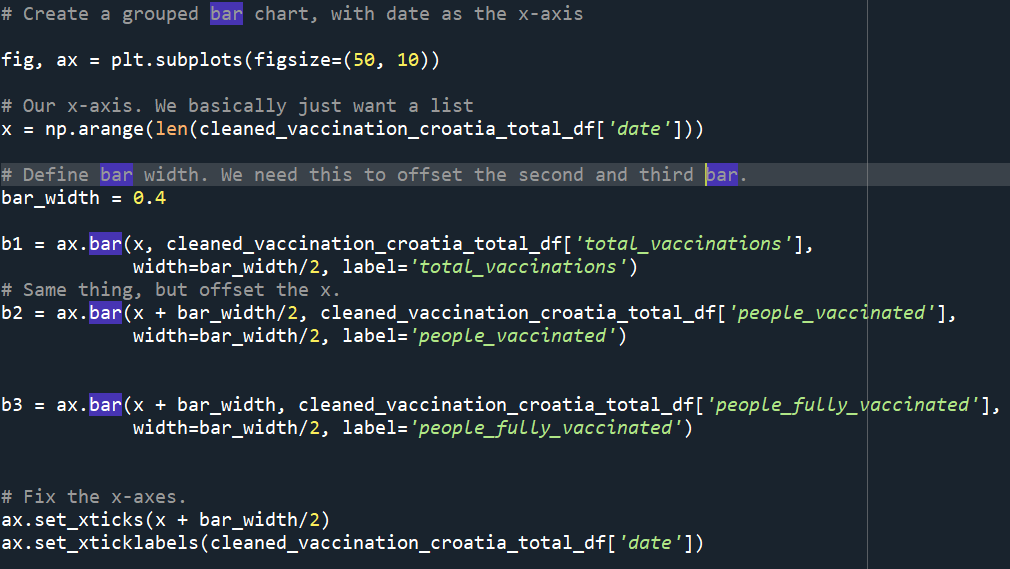


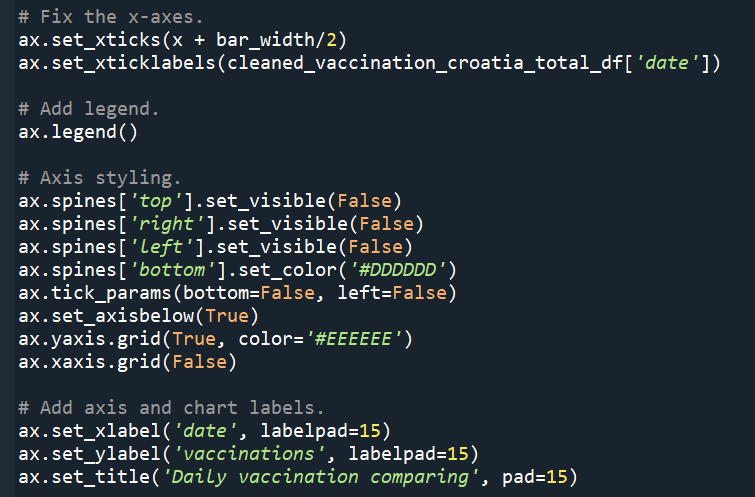
* In the graph, we note that the initial figures of approximately 1000 vaccinations per day starting on 31.12.2020. increased to more than 7000 vaccinations per day by 11.3.2021. The largest increase in the daily number of vaccinations was recorded around the middle of the February, just a few days after we received a new shipment of Pfizer vaccine.

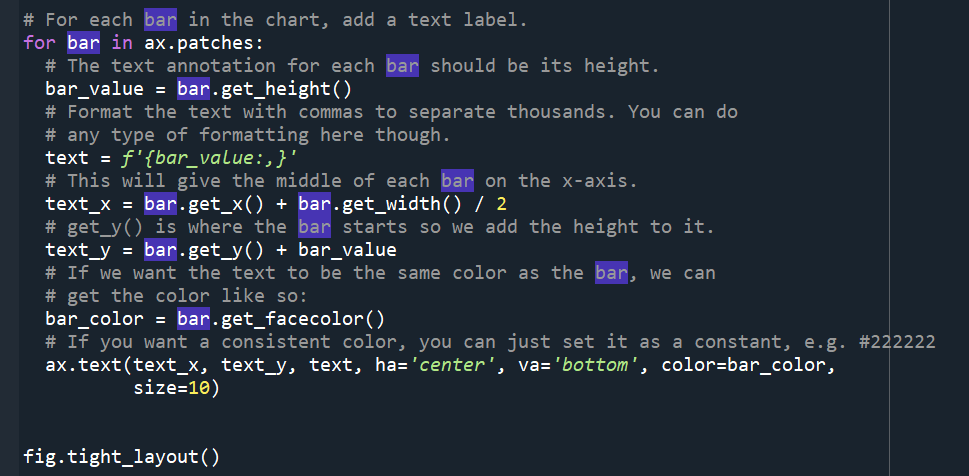
1. Second analysis - the number of vaccinations per day in Croatia
2. Repeat steps a) – c) from the first analysis – we will have to drop much more rows then in above example (Missing cells – 150)

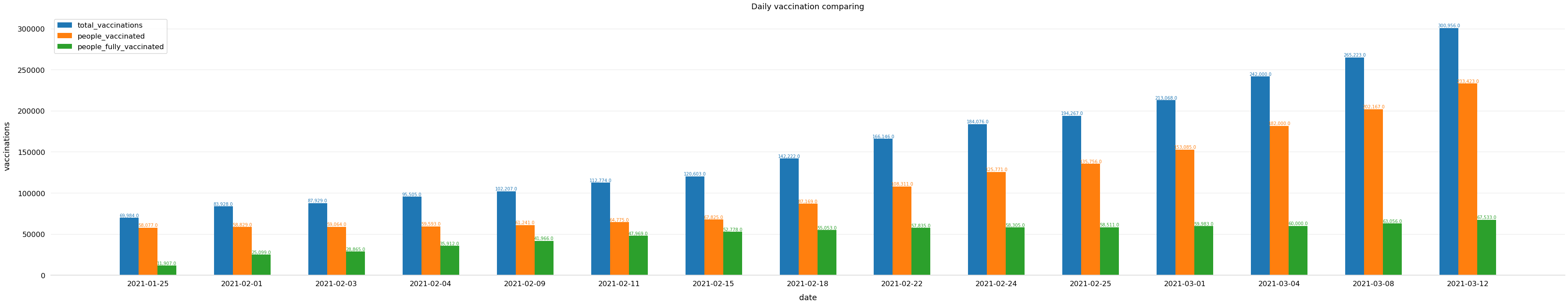


1. Visualize the data – creating a grouped bar chart









* **Total number of vaccinations** - this is the absolute number of total immunizations in the country
* **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people
* **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme
* What we can observe in these data and chart is that although the number of vaccinations drastically increases, the number of people who have full immunity, that is, the number of people who have received a double dose of the vaccine, that number stagnates.

***References:***

<https://www.kaggle.com/gpreda/covid-world-vaccination-progress> ***->*** dataset

<https://github.com/kovaccc/VaccinationAnalysis> -> my github repository