

Homework #1

Instructor: Dr. Zafeirakis Zafeirakopoulos
Assistant: Gizem Süngü

Name: Akif Kartal

Student Id: 171044098

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- It is not a group homework. Do not share your answers to anyone in any circumstance. Any cheating means at least -100 for both sides.
- Do not take any information from the Internet.
- No late homework will be accepted.
- For any questions about the homework, come to my office hour.
- After the office hour, no questions about the homework by email will be responded.
- Submit your homework (both your latex and pdf files in a zip file) into the course page of Moodle.
- Save your latex, pdf and zip files as "Name_Surname_StudentId".{tex, pdf, zip}.
- The deadline of the homework is 22/04/21 23:55.

Problem 1

(100 points)

Homework 1 considers a Covid-19 dataset which is published on [Github](#). Please download any document type that you prefer of the dataset from the links which are shown in Figure 1. The dataset is updated daily and

 Download our complete COVID-19 dataset : [CSV](#) | [XLSX](#) | [JSON](#)

We will continue to publish up-to-date data on confirmed cases, deaths, hospitalizations, testing, and vaccinations, throughout the duration of the COVID-19 pandemic.

Figure 1: The complete dataset links

includes data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. The data set has the following basic columns:

- iso_code: Short name of the country
- continent: The continent where the country exists
- location: The country name
- date: The date when the data about various variables are taken.

You are responsible to implement a program which reads the given dataset from the file and computes the data for the following questions. Any programming language that you prefer will be accepted. Putting comments on your functions that you implement is must. Each question must be appended to a file which is called "output{.csv, .txt}". The file contains the first 18 questions listed below. The 18th question will be written in this document.

1. How many countries the dataset has?
2. When is the earliest date data are taken for a country? Which country is it?

3. How many cases are confirmed for each country so far? Print pairwise results of country and total cases.
4. How many deaths are confirmed for each country so far? Print pairwise results of country and total deaths.
5. What are the average, minimum, maximum and variation values of the reproduction rates for each country?

Table 1: The format of the output for the questions 5, 6, 7, 8, 9, 10, 12, 13.

Country	minimum	maximum	average	variation
value	value	value	value	value

6. What are the average, minimum, maximum and variation values of the icu_patients (intensive care unit patients) for each country?
7. What are the average, minimum, maximum and variation values of the hosp_patients (hospital patients) for each country?
8. What are the average, minimum, maximum and variation values of the weekly icu (intensive care unit) admissions for each country?
9. What are the average, minimum, maximum and variation values of the weekly hospital admissions for each country?
10. What are the average, minimum, maximum and variation values of new tests per day for each country?
11. How many tests are conducted in total for each country so far?
12. What are the average, minimum, maximum and variation values of the positive rates of the tests for each country?
13. What are the average, minimum, maximum and variation values of the tests per case for each country?
14. How many people are vaccinated by at least one dose in each country?
15. How many people are vaccinated fully in each country?
16. How many vaccinations are administered in each country so far?
17. List information about population, median age, # of people aged 65 older, # of people aged 70 older, economic performance, death rates due to heart disease, diabetes prevalence, # of female smokers, # of male smokers, handwashing facilities, hospital beds per thousand people, life expectancy and human development index.

Table 2: The format of the output for the question 17

Country	population	median age	# of people aged 65 older
value	value	value	value

18. Summarize all the results that you obtain by the first 17 questions (except question 2).

Table 3: The format of the output for the question 18

Country	q#3	q#4	q#5_min	q#5_max	q#5_avg	q#5_var
value	value	value	value	value	value	value

19. Comment the results based on your observations. Write your opinions about the reasons of increasing infection rates by giving examples from the results. Feel free to explain any situation that you observe. More observations more opportunities will bring you for the second homework.
(**Solution**) (Write your observations here.)

About this question my observations are following;

Note: These observations was made with April 16 data.

1. My First observation is number of cases will affect number of death directly.
 - For example **Azerbaijan** has 294.211 cases and 4.045 death, on the other hand **Bhutan** has 934 cases and 1 death.
2. If the median age is low in a country, then deaths is low and vice versa.
 - For example **Japan** has 521.796 cases and 9.506 death, on the other hand even tough **Jordan** has more cases than Japan which is 679.138 cases, it has 8.057 death. Because Japan has 48,2 median age and Jordan has 23,2 median age.
3. Economic performance will affect number of vaccinations are administered in each country.
 - For example **Austria** has 45436,686 economic performance and it has 2.197.468 administered vaccinations, on the other hand, **Afghanistan** has 1803,987 economic performance and it has 120.000 administered vaccinations.
4. Death rates due to heart disease will **not** affect total deaths.
 - For example **Central African Republic** has 75 death and 435,727 death rates due to heart disease, on the other hand even tough **Chad** has less case number than central african republic. It has 168 death and 280,995 death rates due to heart disease.
5. Handwashing facilities will **not** affect total cases.
 - For example **Serbia** has 97,719 handwashing facilities and **Senegal** has 20,859 but even tough serbia has two times more population than senegal it has 16 times more cases than senegal.
6. Hospital beds per thousand people will **not** affect total deaths.
 - For example **Japan** has 13,05 hospital beds per thousand people and **Banglades** has 0,8 but total deaths number are almost same. Japan has 9506 total death and banglades has 10081. Also, as we have seen in 2nd observation **Jordan** has 1,4 hospital beds per thousand people but even if it has more cases than japan, it's death number is less than japan.
7. Number of smokers will affect both total cases and deaths.
 - For example **Serbia** has 37,7 female smokers and 40,2 male smoker and **Oman** has 0,5 female smokers and 15,6 male smoker even though their population is close serbia has 3 times more death and cases than oman.
8. Economic performance will affect life expectancy.
 - For example **Turkey** has 25129,341 economic performance and 77,69 life expectancy. Beside this, **Hungary** has almost same economic performance and life expectancy. It has 26777,561 economic performance and 76,88 life expectancy.

System Requirements

In order to run program you need to install followings;

- Python
- openpyxl library for python.
- any other library if needed.
- **owid-covid-data.xlsx** file in same directory.

After running wait a few minute program to finish, then **output.csv** file will be created.