

Homework #1

Instructor: Dr. Zafeirakis Zafeirakopoulos *Name:* Muhammet Fikret ATAR *Student Id:* 1801042693
Assistant: Gizem Süngü *1801042693*

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

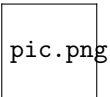


Figure 1: The complete dataset links

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.

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 |

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?
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.)

Hypothesis 1 : In countries with high population density,the total number of cases is higher

How did i obtain the hypothesis 1: I used the answers to questions 3 and 17(population) to prove this hypothesis. When we look at the regions with high population density(USA,Brazil,Russia), we see that the total number of cases(2044341698,737354507,4741333902) is proportionally high.

Hypothesis 2 : Countries with higher economic performance have lower death rates

How did i obtain the hypothesis 2:There are two countries Iraq and France which have the same situations in most of the variables. Country iraq has higher corona death rates than country france. Country iraq has also lower economic rate than country france.

Hypothesis 3 : As the economic performance increases, the number of vaccinations for at least one dose increases

How did i obtain the hypothesis 3:Iraq and Turkey;Country turkey has higher the number of vaccinations than country iraq. Country iraq has also lower economic rate than country turkey.

Hypothesis 4 : According to that the total number of new tests per day increases, the rate of positive cases increases

How did i obtain the hypothesis 4: When we look at America and Slovakia, America has done more tests and found more positive cases.

Hypothesis 5 : The number of deaths from heart disease is higher in countries with high median age.

How did i obtain the hypothesis 5:Brazil and Bulgaria,Country Bulgaria has heart disease rates than country brazil. Country brazil has also median age rate than country bulgaria.[177.96]-[424.68]

Hypothesis 6 :The higher the life expectancy in the countries, have lower death rates.

How did i obtain the hypothesis 6:Australia-Cameroon I have observed that the mortality rate is lower in Austria, where life extacy is high.

Hypothesis 6 :In countries with higher rates of male smokers, more people die from heart disease. How did i obtain the hypothesis 6:Bosnia and Herzegovina - Brazil ;There are two countries Bosnia and Herzegovina and Brazil which have the same situations in most of the variables. Country Bosnia and Herzegovina has higher corona death rates than country Brazil. Country Bosnia and Herzegovina has also higher male smokers rate than country brazil.

As a result of my observations, it can be said that changes in covid data also affect the effect of many factors.Such as economic performance, median age, diabetes prevalence, number of female smokers and male smokers and handwashing facilities affect the data of countries resulting from the epidemic.