**TASK ONE**

Our approach maintains a series of initial variables and calculates the answers to all questions within a single for-loop when reading from the covid\_data.csv file.

**The highest and second-highest number of infections.**

Four variables to hold the highest number of infections and its corresponding country, the second-highest number of infections and its corresponding country, are declared and initialized with the first country values from the file. The algorithm loops through the remaining countries in the file, and when the sum of a country's new confirmed cases is larger than the highest number of infections, it updates the highest number of infections with the new confirmed cases. The previous highest number of infections becomes the second-highest number of infections. The countries are also updated accordingly.

**The countries with the highest infection and death rates.**

Assign the value 0, and the first country read from the file to the initial variables that will hold the highest infection (and death) rate and their countries accordingly. Next, the algorithm loops through the countries, calculate the infection (and death) rates for each country. When a larger infection (and death) rates are found, update the highest infection (and death) rates accordingly as well as their corresponding countries.

**The Overall death rate.**

We find the total confirmed cases by summing up all the confirmed cases for each country. A similar thing is done to find the total deaths. The overall death rate is calculated by dividing the total number of deaths by the total number of confirmed infections.

**The Positive and Negative Trends**

The correlation coefficient is calculated over the recent 7 (1-week data) data points for each country. The countries with a positive correlation coefficient are retrieved as the countries with a positive trend while those with negative correlation are retrieved as countries with a negative trend. Among all the countries with the positive trends, the country with the highest positive correlation has the steepest increase while the country with the lowest negative correlation has the steepest decrease. We generated integers from 1 - 7 as the x-values and used the recent 1-week data as they-values in order to calculate the correlation.

**The country that peaks earliest.**

To find the country whose number of infections per day peak the earliest, we first found the most initial peak point for each country and compare it with a global earliest peak point. If a country peaked earlier than the global one, we update the global earliest peak and keep that country until another peak is found.

TASK TWO

Our approach to solving this task uses pattern matching. We use the KMP string matching algorithm to find the pattern, partial\_time\_series.csv data from the covid\_data.csv file. We first read the partial time series data into a list. Given the list, we use the KMP algorithm to search for it in the covid\_data.csv file and return the date and the country where the match occurred in the file. The main idea of the KMP algorithm is that whenever we have a mismatch, we don't discard what has been matched already. Instead, we leverage the fact that we already know some of the characters in the text of the next window. The algorithm skips matching characters in the next window of the pattern with the next window of the confirmed list that we know will match and match the other characters to decide whether the current window matches or not.