**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.

We answered (a) and (b) by initializing four variables to hold the highest number of infections and its corresponding country, the second-highest number of infections and its corresponding country with the first country values from the file. When we encounter a sum of a country's new confirmed cases which larger than the highest number of infections, we update the highest number of infections with the new confirmed cases. The previous highest number of infections becomes the second-highest number of infections. We update the countries at the same time.

We answered (c) and (e) by first assigning 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, we calculate the infection (and death) rates for each country. When a larger infection (and death) rates are found, we update the highest infection (and death) rates accordingly as well as their corresponding countries.

We answered (c) by finding the total confirmed infection cases. 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.

We answered (f), (g), (h), and (I) by calculating the correlation coefficient over the recent 7 (1-week data) data points for each country. The countries with a positive correlation coefficient are considered 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.

(j) 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.