Assignment 2 – Report

Simon Lindqvist, Abdalrahman Mohammed

Mail: siln22@student.bth.se, abmm22@student.bth.se

Date: 2024-05-26

Course: DV1597

Contents

Introduction

This report presents an analysis of COVID-19 data, focusing on deaths, cases, geographical distributions, various vaccination metrics, and a few additional by us curated questions related to the data. The analysis utilizes datasets on daily new cases and deaths, vaccination statistics, and hospital admission rates, sourced from The European Centre for Disease Prevention and Control (ECDC).

Data cleaning

Data cleaning is a crucial step in ensuring the accuracy and reliability of the analysis. The datasets used include data on daily deaths and cases, data on weekly vaccinations and their nature and data on hospital admission rates. The data cleaning performed was nothing but rudimentary, but for a reason. The data provided is sourced from the ECDC, which is the European centre for disease prevention. Their data is in turn collected from all the different countries in the EU/EEA. The could, and was chosen to be, trusted, and viewed as correct real-world data. ECDC should be trusted to provide already corrected data. Due to viewing the data as real-world data, it mustn't be tempered with to greatly during cleaning. It was chosen to view all data as correct, including missing values. If there are missing values, there is probably no data on it in this case. Interpolating missing values risks altering the analysis of the data we do have. When exploring the data one can also see that most of the missing values are contained to a few columns, for example refused doses in the second dataset (where almost all countries have no data, indicating most countries probably didn't keep record of this). For the same reason instances with missing values cannot be removed, since almost all would be deleted in such a filtering. There is also the fact that having a few missing values in a few places has less of an impact when most of our analysis would be conducted using summation of values into for example a weekly basis. To preserve the integrity of the original datasets, copies of each dataset were created. This allows any changes made during the cleaning process to be reversible and ensures that the original data remains untouched. Duplicates are removed from each dataset to prevent skewed results and ensure that the analysis is based on unique data points. For the vaccination dataset, an additional step is taken to remove rows where the reporting country does not match the region. This step is crucial as it eliminates a lot of duplicate data that arises from countries reporting for both them and their regions. Each dataset also undergoes normalization to convert columns into appropriate data types for analysis. This involves changing date columns to a date format and turning number columns into numeric types so we can do math and time series analysis accurately. We also change some columns to string types to work with text-based operations. [1]

Data analysis

Deaths, cases, and geography

The first thing examined was which countries reported the highest number of COVID-19 cases. We found that the percentages of cases relative to the population are notably high, with the Netherlands in Q1 2022 at 27.16% and France in Q1 2022 at 23.24% showing high numbers. Even the lower percentages, such as Germany in Q3 2022 at 5.25% and France in Q3 2022 at 4.08%, are still significant given their larger populations. All the countries listed are in Europe, indicating a significant impact of COVID-19 in Europe during the periods mentioned, especially in Western Europe. The high numbers can be attributed to several factors, including population density, travel connections, and possibly the efficiency of reporting and testing. The first quarter of 2022 shows the highest case numbers across all the countries mentioned. The case numbers decrease in the second and third quarters of 2022, reflecting either the impact of increased immunity from vaccination and previous infections, public health measures, or seasonal variations [2].

Given possible geographical patterns, a further exploration of the data was done using Geopandas. The total number of deaths and cases was visualized on a map of the countries in the EU. Examining this plot, we found that in 2020, France had the highest number of cases. Western European countries like France, Germany, Spain, and Italy reported high case numbers. Italy had the most deaths, showing strong signs of a harsh impact of covid in the early pandemic. While Italy had the highest death toll, other parts of western and central Europe was also hit hard. In 2021, France continued to have the most cases. Western and central Europe was still significantly affected when it came to cases. Germany and Poland reported the most deaths, which was in line with the general trend that central Europe remain with a lot of deaths with eastern Europe gaining a substantially higher death toll. In 2022, both France and Germany reported the highest number of cases. Western Europe was still heavily impacted, with significant case numbers in France, Germany, and Italy. Italy again reported the highest number of deaths. Most deaths were once again concentrated in western and central Europe, although showing a general decline since the previous year [3].

Possible reasons for these countries having such high total deaths and cases can perhaps be attributed to their larger populations, higher population density (Germany, France and Italy featuring a few high-density larger cities with high throughput of people). It is however worth noting that the data in the geographical plot is just total numbers, and not ratios, so larger countries are expected to get somewhat higher values. Though viruses spreading easier where more people reside is an important aspect too.

Vaccines

When exploring the vaccinations we first started by looking at which vaccines were the most popular, which age groups these were administered to and how many doses were administered. Our findings reveal that the most popular vaccine across the EU/EEA is Pfizer BioNTech's Comirnaty followed by Moderna's Spikevax and AstraZeneca´s - Vaxzervria. This is also reflected at the country level, where Pfizer BioNTech's Comirnaty stands out as the top vaccine in nearly all countries. Moderna's Spikevax also ranks consistently within the top three in most countries, showing it also had a great part in vaccination campaigns. AstraZeneca – Vaxzervria, while popular, varies in position, being the second or third most used in several countries.

There are, however, some exceptions. France is an outlier with 'Unknown' listed as the top vaccine, which might indicate different reporting standards when it comes to vaccine type. Hungary shows a unique pattern with BBIBP-CorV and Sputnik-V among the top three, showing its willingness to use non-western vaccines. Belgium and Denmark feature updated versions of the Pfizer-BioNTech vaccine, specifically Comirnaty Original/Omicron BA.1 and Comirnaty Original/Omicron BA.4/BA.5, indicating a possible preference for these newer variants. Liechtenstein stands out with Moderna as the top vaccine, which is less common compared to other EU/EEA countries where Comirnaty generally is the preferred one. We decided not to exclude unspecified vaccines from our analysis. Their presence could point to data aggregation issues or varying reporting standards. Keeping this category provides an interesting insight into potential variations in vaccine reporting practices [4].

Then the age groups to which the top 3 vaccines for each country was administered was also investigated. It shows that Pfizer BioNTech's Comirnaty, the most common vaccine across all countries, has the highest number of doses in the 25-49 age group, followed by the 50-59 and 60-69 age groups. Moderna's Spikevax is most common in the 25-49 age group, although generally less so than Comirnaty. In the 60-69 and 50-59 age groups also had a significant amount of Moderna doses. AstraZeneca's Vaxzervria is mostly administered to older age groups (60-69 and 70-79) [5].

The analysis shows a clear preference for Pfizer BioNTech's Comirnaty across the EU/EEA, with significant contributions from Moderna's Spikevax and AstraZeneca's Vaxzervria. Unique patterns in vaccine acceptance and usage are observed in certain countries, and the age group distribution shows that different countries have different vaccination strategies, or different availability of vaccines.

Vaccine Scepticism

To grasp different countries attitudes towards vaccines, a logical step is to measure scepticism against them. An easy way to do this would have been to use the refused vaccine column in the second dataset. However, the dataset reveals a large missing values for first dose refusals, with Iceland being the only country reporting instances of refusals. This might indicate either a degree of scepticism or a unique enthusiasm for reporting vaccine hesitancy. Due to incomplete data, a real assessment of vaccine hesitancy using this metric in other countries isn't possible. Thus, we focus on Iceland to draw any conclusions [6].

From February 8, 2021, to April 26, 2021, Iceland saw periods of vaccine refusals, which aligned with spikes in hospitalization rates during the same period. This suggests a potential correlation between first dose refusals and increased hospitalizations [7].

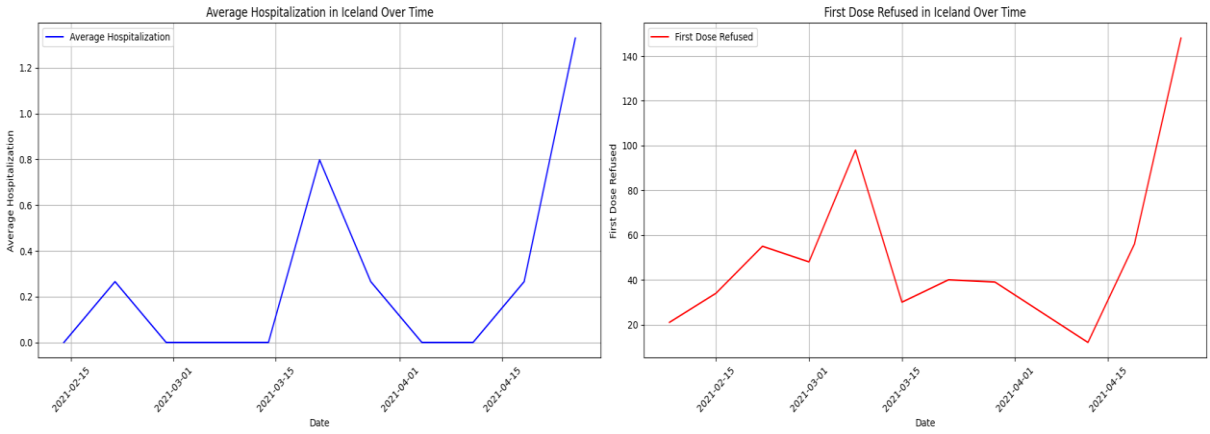

Figure 1. Plot of hospitalizations and vaccine refusals in Iceland during a period in 2021

A correlation plot between average hospitalization rates and first dose refusals in Iceland shows a moderate positive correlation coefficient of 0.6059, indicating that refusing the first dose may lead to higher hospitalization levels. To further investigate, we examined the percentage of people in each country who received the first dose, as a lower percentage might indicate higher refusal rates or inefficient vaccine administration. After eliminating duplicate age groups from our data, we calculated and sorted the percentage of people who received the first dose in ascending order. The plotted bar graph shows that Iceland has one of the highest vaccination percentages, suggesting it is not particularly sceptical about vaccines despite being the only country reporting refusal data. Romania and Bulgaria have the lowest percentages, indicating a possibility of higher scepticism or less efficient vaccine administration [8].

Vaccination of young people

The next step was to find investigate ratios of young and old people's vaccinations compared to the total populations. For young people age under 18 was chosen. We then started by creating an empty DataFrame and filling it with vaccination data for the age groups under 18 using the concat function. Next, we identified all unique countries in the dataset. For each country, we calculated the total number of individuals under 18 who have received their first dose of the vaccine. Then, we calculated the ratio of vaccinated individuals under 18 to the total population for each country and sort the results in descending order to highlight the countries with the highest proportion of vaccinated individuals under 18. However, we noticed that not all countries are present. This made us investigate what happens if we filter using the "Age<18" target group instead of separate age groups. Repeating the same process with the "Age<18" target group, we found that not the same countries are present. This divergence arises because some countries report using "ALL" and "Age<18" for target groups, while others use specific age groups or both methods. To address this, we combined the data into one plot using a union to ensure we include all relevant data, avoiding duplicates. By using a union of the two previous country groups, we created a new dataset that includes all the countries [9].

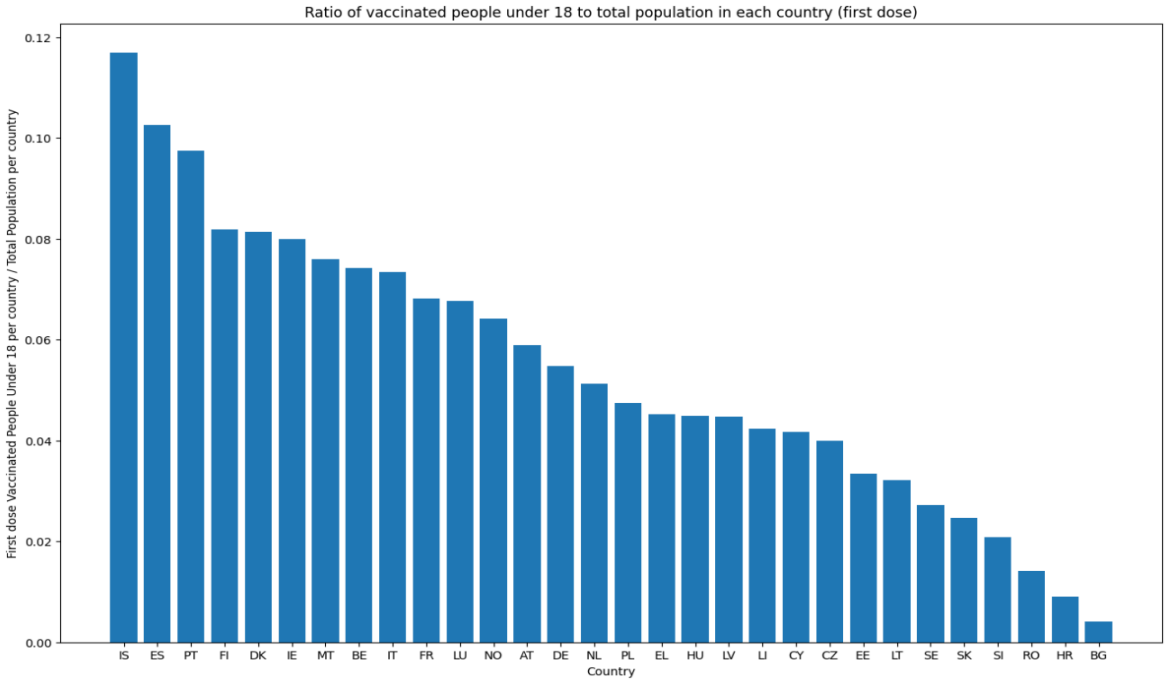

Figure 2. Plot of ratios of people<18 vaccinated by first dose compared to the total population.

The plot shows that Iceland, Spain, and Portugal have the highest proportion of vaccinated people under 18 compared to their total populations for the first dose of the vaccine. On the other hand, Romania, Hungary, and Bulgaria are the worst performers. This finding is unsurprising, as other plots in the Jupyter Notebook indicated that both Bulgaria and Romania had very low vaccination rates overall.

Vaccination of old people

After looking into the ratio of people under 18 getting the first dose, it was time to explore which countries had the oldest vaccinated population compared to their populations. We calculate the ratio of second doses to the total population for each country and plot this in descending order [10].

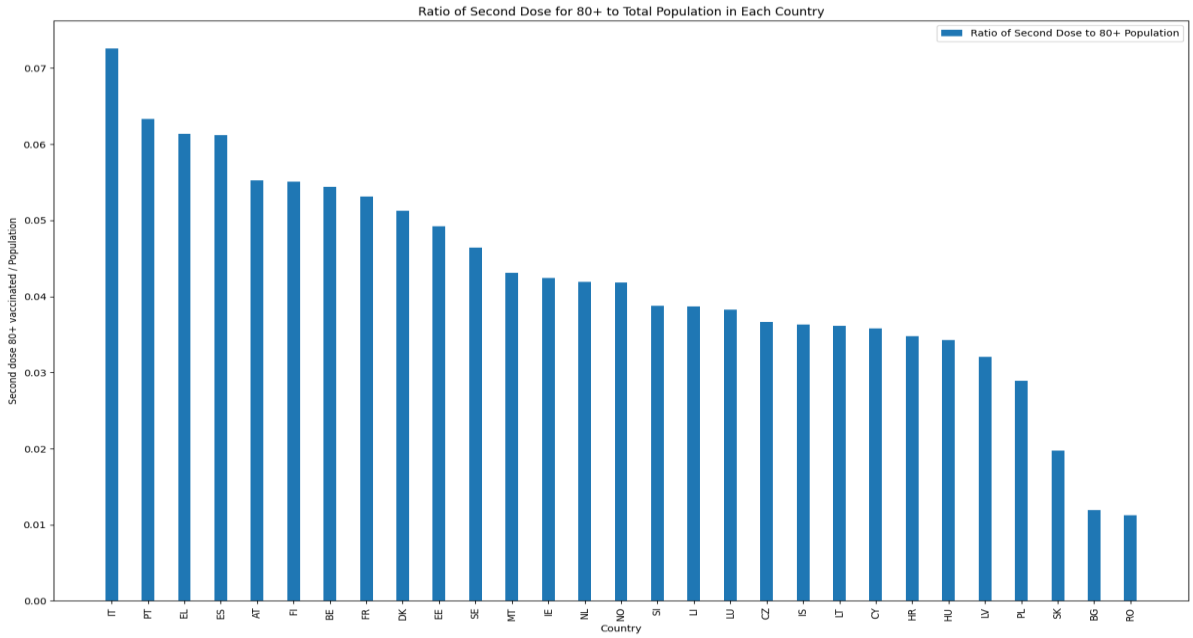

Figure 3. Plot of ratios of people 80+ vaccinated by second dose compared to the total population per country.

Upon examining the graph, four main outliers become apparent. Italy stands out with the highest ratio of vaccinated 80+ individuals compared to the total population. Conversely, Slovakia, Bulgaria, and Romania show exceptionally low numbers. Notably, Bulgaria and Romania were also among the worst performers for first dose vaccinations in the under-18 age group. This pattern suggests that these countries may struggle more than others in vaccinating their populations or that there may be higher levels of vaccine hesitancy among their populations. These countries also had the lowest percentages of first dose vaccinations, reinforcing the idea that they face significant challenges in their vaccination campaigns.

We chose to focus on the 80+ age group for two main reasons. Firstly, individuals aged 80 and above are among the most vulnerable to severe outcomes from COVID-19, including higher mortality rates and complications. Ensuring this age group is fully vaccinated is crucial to protecting them and reducing the strain on healthcare systems. Additionally, analysing the 80+ age group provides clear insights into how effectively countries are prioritizing and managing vaccinations for their most at-risk populations. While for example the 60+ age group is also important, the 80+ group represents a more immediate and urgent priority due to their higher risk factors. Measuring 80+ population is however liable to differences in expected lifespan of humans in different countries, which is something one should consider when analysing the data.

Hospitalization

We used the indicator column to filter for weekly new hospital admissions data and identify the five countries with the highest peaks, assuming they were hit the hardest. We then plot the data for 2020 and 2022 to compare the situations. In 2020, Belgium and Spain experienced early peaks in hospital admissions, with levels reaching ~30 per 100k in Belgium and ~65 per 100k in Spain by mid-March. Other countries peaked later in the year. By mid-October, admissions rose rapidly again, with Belgium peaking at around 40 and the Czech Republic reaching ~60 per 100k. In 2022, all five countries peaked in February at 20-40 admissions per 100k, except Latvia, which almost reached 60. Throughout the year, admissions remained steady without extreme surges, despite high case numbers early in the year. This suggests that vaccinations reduced COVID-19 lethality, resulting in fewer hospitalizations despite more infections. ICU admissions showed similar patterns, with the Czech Republic peaking at the end of 2020 and Latvia at the beginning of 2022. ICU rates dropped significantly later in 2022, indicating a possible correlation between hospital and ICU admissions. This proportionality suggests that vaccinations helped reduce severe cases, even if correlation does not imply causation. Proportionality was then tested for the countries in the ICU plot, showing a very significant correlation between normal and ICU hospital admissions, indicating these are directly proportional to each other [11].

Additional Questions

After the main and obligatory analysis, we still had some questions of our own. We wanted to know which country would provide us with the lowest risk of dying of we got infected with covid, if there was any correlation between vaccinations and deaths (i.e. if vaccines have any effect on lethality) and if there is any pattern for when people get sick. We argued all these were of vital interest to both us and the public since they allow both the individual and government institutions to plan accordingly. For example, a tourist might want to know the country with lowest covid lethality to minimise risk during vacation, or someone might want to know when in the year they should be extra cautious.

To answer the first question, we analysed the death rate per country during 2022, as this is the most recent data and should reflect the current situation best. Countries have likely optimized their COVID-19 responses by this time, because why would a country become worse at handling covid-19 over time, making 2022 data more relevant than data from earlier in the pandemic. First, we filter the deaths and cases data for 2022 and group it by country. Then the ratio of deaths to cases was calculated for each country, which would indicate how often a COVID-19 case results in death. From the graph, it is evident that the Netherlands has the lowest death-to-cases ratio in 2022. Thus, it can be concluded that the Netherlands was the safest place to contract COVID-19 in 2022. Given that this is the latest data available, it is the best guess to the current situation [12].

For the second question we started by plotting weekly vaccinations and deaths per 100k to look for patterns. The analysis period was set from the first week of vaccination reporting to the latest available full week of data on deaths. Our initial plots showed a potential decline in deaths following vaccination. To further examine this potential relationship, we opted to use a cross-correlation to account for the likely delay between vaccination and its effects on deaths, as vaccines do not provide immediate protection.

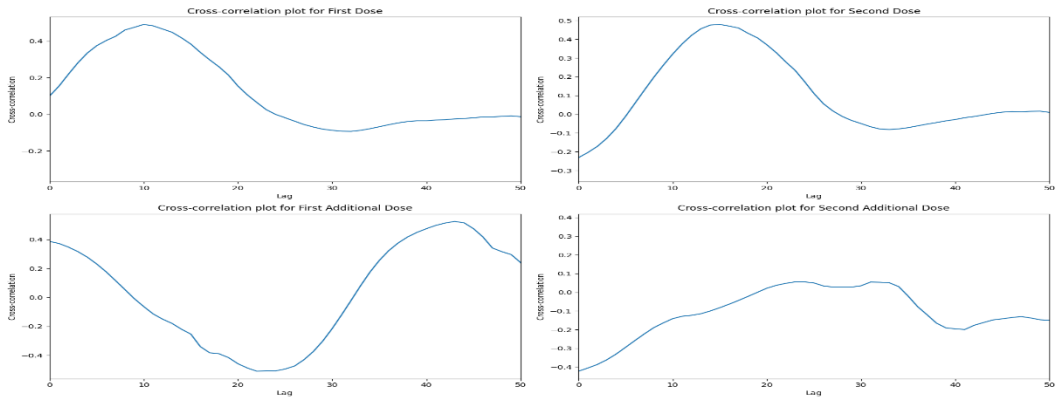

Figure 4. Cross correlation plot of vaccinations/deaths per 100k

For the first dose, we observed a peak correlation of nearly 0.5 at around 10 weeks post-vaccination. The second dose showed a similar peak correlation, occurring around 15 weeks after vaccination. For the first additional dose we saw an immediate correlation of about 0.4, and another peak at 40 weeks, though this longer delay might be due to factors unrelated to the first additional dose due to the extraordinary delay. The second additional dose showed no significant correlation, possibly due to already low death rates and a limited amount of data, making it difficult to detect patterns or correlations. Although correlation does not imply causation, the evidence suggests that vaccinations help reduce mortality, supported by both the observed trends in the graphs and the cross-correlation analysis. This is also supported by other findings in the data analysis where we found that even though there were more cases in the beginning of 2022, the rate of hospital admissions was the same, indicating less severe symptoms and less lethality. This coincides with a way larger share of people having gotten the vaccine in 2022, telling us vaccine might indeed reduce lethality [13]. For the third question we plotted the average number of cases for each month during 2020, 2021, and 2022. This approach helps identify any seasonal trends. We calculated the average number of cases per month by filtering the data for each year, grouping it by month, and calculating the mean number of cases [14]. The plot reveals a clear trend of increased COVID-19 cases during the colder months. For the most recent year, the trend doesn't quite appear due to the lack of data after October, the same period where case numbers typically rise in previous years. The overall pattern found excluding the lack of data is however consistent with other viruses like the flu, which also see higher transmission rates in winter in temperate areas like Europe [98]. Other research also supports this observation, indicating that coronavirus tends to spread less in warm, humid conditions and more in colder climates [99].

Conclusions

Western European countries, particularly France, Germany, and Italy, consistently reported high case numbers throughout 2020-2022, with Italy having the highest number of deaths early in the pandemic Possible explanations are some similarities between these countries such as population density and extensive travel networks. High population density in urban areas could have facilitated virus transmission, while robust healthcare and reporting systems may have resulted in more accurate and higher case counts, compared to other countries with less dense populations and a less efficient reporting system. In terms of vaccination, Pfizer's Comirnaty emerged as the most administered vaccine across the EU/EEA, followed by Moderna's Spikevax and AstraZeneca's Vaxzervria. This preference is likely due to these vaccines are at least partially manufactured in Europe, providing easier access and higher control of the vaccines. Some countries did however diverge from these trends, such as Hungary's significant use of non-western vaccines like the Chinese BBIBP-CorV and Russian Sputnik-V. This might be due to Hungary's geopolitical alliances, availability issues or a combination of both.

Despite incomplete data on vaccine refusals, Iceland's reporting indicated a possible correlation between first dose refusals and increased hospitalization rates. However, Iceland's high overall vaccination rate suggests a generally low level of vaccine scepticism, perhaps due to strong public health messaging and trust in government recommendations. The 25-49 age group received the most vaccinations, particularly with Comirnaty, reflecting targeted vaccination strategies aimed at protecting the working-age population. The prioritization of older age groups (60-69 and above) with AstraZeneca doses underscores efforts to protect the most vulnerable from severe outcomes. Iceland, Spain, and Portugal led in vaccinating individuals under 18, possibly due to wanting to act proactively. Conversely, Romania, Hungary, and Bulgaria lagged in vaccinating both young and elderly populations. This could be attributed to factors such as vaccine hesitancy, logistical challenges in vaccine distribution, lower trust in government health initiatives or economic reasons.

The analysis of hospitalization shows a strong correlation between hospitalization and ICU admissions and that the ratio of severe and normal cases didn't change over time, which suggests that vaccination efforts significantly reduced the severity of COVID-19 cases. When addressing additional questions, the Netherlands was identified as having the lowest death-to-case ratio in 2022, indicating effective management and treatment strategies. Cross-correlation analysis indicated that vaccinations likely contributed to the reduction in mortality rates, with significant correlations observed between vaccine administration and subsequent declines in death rates. Seasonal trends were also evident, with increased case numbers during colder months, consistent with patterns seen in other similar viruses.

References

[1] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 2

[2] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 3

[3] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 4

[4] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 5

[5] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 6

[6] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 7

[7] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 8

[8] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cells 9-10

[9] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cells 11-13

[10] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 14

[11] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cells 15-17

[12] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 18

[13] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 19-20

[14] S. Lindqvist, A.Mohammed, Submitted Jupyter Notebook, Code cell 21

[15] W. H. Organization, “World Health Organization”, 2023-10-03. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[16] Chen, S., Prettner, K., Kuhn, M. et al. Climate and the spread of COVID-19. Sci Rep 11, 9042 (2021)