**Time Series Analysis of Covid Impact on the UK, SWEDEN and AUSTRALIA**

**Qr code

Description automatically generated**

**Authors**

**MoureenCaroline Ochieng| Sophie Burroughs | Chiratidzo Matowe | Pelumi Oluboye | Chang Xu | Anna Lisowska**

**1.0 Introduction**

At the start of the year 2020, social and physical isolation practices were introduced to curb the spread of a global outbreak. In order to analyse the effects of those practices, we chose 3 countries namely: United Kingdom, Sweden and Australia. Our choice to select these 3 countries was due the varying levels of restrictions. The aim of our research was to further perform a time series analysis using our datasets collected from different sources and then produce a forecast of confirmed cases for the upcoming months. Secondly, we sought to further investigate and gain insight on the impact that covid-19 restrictions had a lasting effect on Covid case numbers.  The report is a comprehensive description of the key findings and the questions we answered using detailed diagrams. We further discuss the individual efforts done by each individual within the team in the pursuit to answer the project’s key objectives.

**1.1 Aim**

Our work aimed to evaluate the impact of covid-19 in 3 distinctive countries by using past and recent data in order to learn future trends that could be applied to further covid outbreaks.

**1.2 Research objectives**

1. Investigate and comparatively assess the effect of covid-19 new cases, total cases and deaths across the 3 different countries.
2. Understand the correlation of data variables within the dataset in order to predict covid-19 outbreak.
3. Analyse the variations in restriction levels across the 3 countries. Investigating whether the restrictions were effective in stopping the spread of the virus.
4. Look into the growth of Covid over time using time series analysis to calculate moving average in new cases across the 3 different countries and the correlation between new cases and deaths. Looking into the seasonality of Covid and investigating the correlation differences between different countries and waves.
5. Investigate the relationship between new daily cases in relation to daily death cases across the countries.

**2.0 Background**

The first cases of Covid-19 were detected in each of these 3 countries in January 2020 but a serious outbreak was confirmed in March 2020. We have chosen these 3 countries based on their differing government responses to the pandemic. We aim to analyse the effect of these restrictions on the number of Covid cases and deaths and also on the forecasted cases and deaths in the future.

**2.1 Restrictions in UK**

On 23 March 2020, the UK implemented a strict lockdown closing schools, offices and public leisure facilities. By July 2020, these restrictions were reduced slightly, schools had a phased reopening, non-essential shops reopened and restaurants, bars and leisure facilities were allowed to reopen with social distancing regulations. As autumn arrived, the regulations were increased again, a curfew was brought in and a limit of 6 people could gather in groups. People were again encouraged to work from home. By November 2020, a second strict national lockdown was imposed on the country, this lasted for 4 weeks with a few weeks of slightly less stringent regulations before a third lockdown came into force on 3 January 2021.

By February 2021, a roadmap to lift regulations was published and this began with schools reopening on 8 March 2021 and outdoor gatherings were allowed. By May 2021, outdoor gatherings of up to 30 people were allowed, small groups could meet indoors and sporting events could take place outside with spectators. By July 2021, most legal limits on social contact were removed and all elements of the economy were allowed to reopen. This signalled the end of strict lockdown in the UK.

**2.2 Restrictions in Sweden**

Unlike most of the international community, Sweden took a more lenient approach to restrictions in response to the Coronavirus pandemic. In March 2020, Sweden allowed things to continue as normal with only a few minor restrictions in place. The Public Health Agency did initiate contact tracing and a strategy to protect the most vulnerable citizens. It was advised that those with symptoms avoid social contact, work from home and minimise travel. Eventually, large gatherings were banned and there was some advice that higher education institutions switch to remote learning.

In December 2020, the Swedish Prime Minister announced new tougher restrictions including advising face masks be worn on public transport and closure of non-essential public services. In January 2021, a law was passed to allow for the use of lockdown measures and to legally limit social gatherings. Some further restrictions were implemented in June and July 2021, with some international travel bans and some limits on the amount of people that could gather in a social setting as well as working from home being encouraged. These rules were lifted in autumn 2021 before being tightened again in December 2021 as a response to the new Omicron variant. By February 2022, all bans on gatherings and leftover restrictions were removed.

**2.3 Restrictions in Australia**

In March 2020, the Australian government declared a human biosecurity emergency in response to the outbreak. The crucial difference between their response and the response of the UK and Sweden is that they closed their borders to all non-residents on 20 March 2020. In addition to this extreme measure, many individual states closed their borders and social distancing rules were introduced. This included closing non-essential services such as pubs, clubs and other social gathering venues. They did however, keep most business operations open such as construction, manufacturing and some retail.

Australia was one of the few countries to pursue a ‘zero-covid’ strategy until late 2021, aiming to minimise domestic transmission. This policy involved strict regulations on international arrivals and swift responses to any local Covid outbreaks with lockdowns. Some local outbreaks were responded to with very strict lockdowns that received global attention, they aimed to reach a point of zero new cases. There were no deaths from Covid in Australia from 28 December 2020 to April 2021.

**3.0 Steps Specification**

In terms of our analysis, we began by deciding on a topic. We all had some interest in health-based data and the pandemic has provided a lot of interesting data that we can analyse. Firstly, we looked into available datasets and the most comprehensive of these was sourced from ‘Our World in Data’. As this dataset was so large, we wanted to narrow this down and so chose to focus on 3 countries that we could compare based on their differing strategies to Covid restrictions. To add to this research we brought in additional data that could provide figures for levels of restrictions across the country.

Having chosen the UK, Sweden and Australia, we narrowed down the dataset accordingly and then set about cleaning the data. Figure 3.0 below summarises the steps we took in order to achieve our outcome for the project.

**3.1 Data collection**

We collected our data from Our World in Data and Kaggle. Our choice to use different sources because the data sources held useful variables that answered the questions we had. The data collected from these sources was structured meaning it was retrieved in a fixed format.

**3.2 Dataset Cleaning**

This step involved correcting inaccurate, insufficient, duplicate, or other wrong data in the dataset. This entailed locating data mistakes, then correcting them by modifying, updating, or eliminating data. The data cleaning process enhanced our data quality and contributed to the provision of more precise, dependable, and consistent information for decision-making for our project. Particularly we discovered that we initially had 67 columns in our dataset, however, among the 67 columns, 23 columns had 2377 records of null values. We had to further delete the column entries to avoid poor forecasting results. The remaining columns had fewer null values, for example, one of the columns with the highest value of missing values was 716. Mean imputation was our chosen method for computing the missing values. The missing values were substituted with a computed mean value. We filtered the data so that there was fair comparative analysis among the 3 countries. Our chosen dates for analysis were from  01 February 2021 to 31 July 2022. Secondly, we removed any occurrences of duplicate values

**3.3 Exploratory Data Analysis**

We used this phase to do initial investigations on the data and check for patterns. We also used this phase to check for summary statistics.

Diagram

Description automatically generated

**4.0 Implementation and Execution**

The successful implementation of our project was done using a rigorous data analysis process. To carry out a successful implementation of the project, we generated useful questions that we could use to achieve the research objectives stated in section 1.2. Data preprocessing and data cleaning and data exploration were fundamental steps for us to achieve a clean dataset. The team followed an Iterative development. We broke down the project work into parts such as those discussed in Figure 3.0. We further did iterative cycles in order to improve the quality of our work.

**4.2 Team Execution and Tasks**

|  |  |
| --- | --- |
| **Tasks** | **Team Member Name** |
| **T**opic definition | Unanimous Decision |
| **D**ata Collection | All |
| **D**ata Exploration phase 1 (understanding the data) | All |
| **D**ata Exploration phase 2(data cleaning and understanding the variables) | Oluboye Pelumi and Anna |
| **D**ata Analysis Phase 1(Time series analysis) | Chang Xu |
| **D**ata Analysis Phase2(Understanding the relationship between restrictions, deaths and new cases) | Moureen and Sophie |
| **R**eport Writing | Chiratidzo and Sophie |
| **P**reparation of powerpoint | All |

**4.2.1 Tools and Libraries**

The following tools and libraries were used to execute our project:

* Python
* Pandas
* Numpy
* MatplotLib
* Seaborn
* Altair

**4.3 Communicating, Visualising and Reporting**

For this section, we have used diagrams or plots to communicate our findings and explain the information we found.The following sections give detailed explanations of how we addressed the research objectives in Section 1.2.

**4.3.1 Investigating and assessing the effect of covid-19 deaths impact among the 3 countries.**

The diagrams below answer Objective I. It shows the comparative analysis of the total Covid cases and the new cases recorded among the 3 countries. The UK is the country that has the greatest number of cases in comparison to the other two, followed by Australia and Sweden. Among the 3 countries Sweden had lesser restrictions in comparison to the other two countries. Surprisingly, countries like Australia that had stricter restrictions had more total cases in comparison to Sweden which had a lenient approach to implementing Covid restrictions. The total cases in the UK show a  steep growth regardless of the strict restrictions that were put in place at the start of covid outbreak. A potential reason for this could be to do with population density as the UK is more densely populated than the other two countries, this is not something that we could control for.

Chart, histogram

Description automatically generatedChart, line chart

Description automatically generated

**Figure 4.3.1.0:** *New Cases across the 3 Countries* **Figure 4.3.1.1:** *Total Cases across the 3 Countries*

**4.3.2 Understanding correlation of data variables within the dataset in order to predict covid-19 outbreak.**

We used a correlation matrix so that we depicted the correlation between all the possible pairs of values in a table. We chose to use this as it  is a powerful tool to summarise our dataset and identify and visualise patterns in the given data. The following are the summaries of the findings :

* New cases and total cases are highly correlated with a value of 0.8.
* New cases and new cases per million is highly correlated with a value of 0.8.
* New deaths and new deaths per million is highly correlated with a value of 0.8.

A picture containing calendar

Description automatically generated

**4.3.3 Analysing the effect Restrictions had on the Total Deaths.**

We addressed Research Objective iii, by analysing the variations in restriction levels across the 3 countries, through the use of line plots. We brought in a new dataset to provide the data on levels of restrictions, restriction level is measured by reviewing all containment policies implemented in the country at a single time and then giving stricter policies a higher value. Figures 4.3.3.0, 4.3.3.1, 4.3.3.2 and 4.3.3.3 below show the results obtained from the analysis. These figures generally show a trend that validates our assumptions of Covid restrictions and the fact that they played a major role in lowering the total number of deaths. It can be noted in figure 4.3.3.0 that Sweden saw a period of increased restrictions where the deaths continued to rise.

This is an interesting insight from a policymaking perspective as we can clearly see that increased levels of restrictions align with lower total death figures. However, we would need to be careful that this result was not affected by other factors such as public obedience to the restrictions at earlier stages in the pandemic and the results of the vaccination regime in the later part of the pandemic.

Chart, line chart

Description automatically generated

**Figure 4.3.3.0:** *Restriction Levels against Total Deaths in Sweden.*

Chart, line chart

Description automatically generated

**Figure 4.3.3.1:** *Restriction Levels against Total Deaths in Australia.*

Chart, line chart

Description automatically generated

**Figure 4.3.3.2:** *Restriction Levels against Total Deaths in the UK.*

Chart, scatter chart

Description automatically generated

**Figure 4.3.3:** *Comparison the effects of restrictions in the UK, Australia and Sweden.*

Figure 4.3.3 shows a scatterplot which displays the relationship between restriction levels and total deaths across the 3 countries. We can see from the graph that Sweden (green) is showing a slight negative correlation between restriction levels and total deaths meaning that there were fewer deaths when restrictions were made stronger. Australia (yellow) also shows this slight negative correlation with stronger restrictions resulting in fewer deaths. The UK, on the other hand, shows a more mixed picture, with high death rates even with strong restriction policies. This could be explained by a lag in the UK implementing policies or other influential factors such as increased population density or increased civil disobedience to the restrictions. Overall, based on this graph and the slight negative correlation shown in the data, there is an argument for restrictions being effective in limiting Covid deaths.

**4.3.4 Time Series Analysis of Covid Growth over time using Average Cases and Death**

We addressed Research Objective IV, by investigating the trend of average new cases along with the average deaths growth over time.The key question we intended to solve was to check if the  three countries adapt to different lockdown policies. Particularly, was there any difference in the case growth pattern and correlation between death growth and case number growth. In order to do this we checked for a correlation between weather seasons and total deaths by doing an analysis using a boxplot at 4.3.4.0. We discovered that all of the countries display distinct seasonality in case growth and death growth. However, for the UK and Sweden, this aligns with their actual seasons, yet Australia follows a similar pattern of case growth and death growth but they have their summer when Europe has its winter. This suggests that the seasonality displayed in the data is not necessarily the result of the actual season, it can also be influenced by lockdown restrictions and other variables.

**United Kingdom                   Sweden                             Australia**

**Chart, box and whisker chart

Description automatically generated Chart, histogram

Description automatically generatedChart, box and whisker chart

Description automatically generated**

**Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated   Chart, box and whisker chart

Description automatically generated**

**Figure 4.3.4.0:** *Boxplot showing Seasonality*

The below further time series analysis showed the trends more clearly across the 3 different countries. We used the moving average because covid identification was influenced by a window period of about 14 days, this also helps to flatten the curve. This moving average analysis was used to identify the growth situation of covid. Hence, we built the average moving of the new cases based on the window period information. The data used in this analysis was from February 2020 to August 2021, hence the diagram is showing a collection of 30 months in total. Each graph has three peaks, therefore we assumed that the outbreak cycle is 10 months.

The trend is such that new average cases in all the 3 countries are the same. The diagram shows a periodic growth. New cases are still reported to date in the UK and Australia. Sweden has the lowest number of new average cases. The new average deaths also show a periodic growth, with Australia showing the greatest growth in the last few months to date. Overall, the trend is showing the opposite trend between the two variables(new cases and deaths), when there is an increased number of deaths, there are fewer cases. However, from the period between July 2020 to October 2020 and April 2021 and July 2021 in the UK, there was no marginal difference in the new cases deaths recorded. In Sweden the period between August and November 2020, and April 2021 to January 2022, there is also no marginal difference in the new cases deaths recorded. Australia's case is unique as it shows no recorded new cases and average deaths from November 2020 to July 2021.  In addition, it can be noted that  there is lag between the new case growth and new deaths. In Stage 1, when the vaccine when the virus first identified: there is no correlation between the death and cases being identified in the UK and Sweden. Based on the observation more death was observed compared to the people get infected.

Graphical user interface, application

Description automatically generated

**Figure 4.3.4.1:** *Time Series Analysis across the 3 different Countries.*

**4.3.5 Investigate the relationship between new cases in relation to daily new cases and daily new deaths across the 3 countries.**

Objective V, was addressed using the below histograms showing the daily new cases and daily new deaths and daily deaths. It can be observed that the general trend when there are new cases recorded it also contributed to the spike in deaths. The UK among all the countries has the highest number of deaths and highest number of new cases, followed by Sweden and lastly, Australia.

Chart, histogram

Description automatically generated

**Figure 4.3.5.0:** *Daily New Cases and Daily New Deaths for the UK*

Chart, bar chart, histogram

Description automatically generated

**Figure 4.3.5.1**:*Daily New Cases and Daily New Deaths for Sweden*

Chart, bar chart, histogram

Description automatically generated

**Figure 4.3.5.2:** *Daily and New Cases for Australia*

We further analysed the above data using violin plots. Figure 4.3.5.2.0 and 4.3.5.2.1 are a representation of the daily new cases and daily deaths. The size of the dots,is a representation of the new occurrence of cases. A bigger dot represents more occurrences of daily new cases and the same applies vice versa. The UK had the greatest number of daily new cases, followed by Sweden and Australia. The same trend applies for the daily new deaths.

Chart, line chart

Description automatically generated

**Figure 4.3.5.2.0:** *Violin Plot showing the Daily new Cases*

Graphical user interface, application

Description automatically generated

**Figure 4.3.5.2.1:** *Violin Plot showing the Daily new deaths.*

**5.0 Conclusion**

The aim of our research was to show the relationship that restrictions had on new covid cases and deaths. It can be concluded that restrictions played a major role in the spread of covid and reducing the number of deaths. Particularly,  implementing tight restrictions was a great contributing factor to reducing the spread of covid and reducing the deaths of citizens, for instance Australia experienced fewer cases of deaths and infections. On the other hand, the UK experienced the greatest number of cases and deaths; the country is densely populated and could have played a major role regardless of the strict measures they initially had. Lastly, Sweden could have quickly reached herd immunity early due to the country’s relaxed approach to implementing restrictions which could have contributed to the fewer cases in deaths and new infections.

**6.0 Challenges and Future Work**

In some instances we couldn’t plot detailed plots because we did not have enough data. For example, we did not have enough data for the month of January 2020 in order to start the seasonality information for the BoxPlots.

In addition, implementing prediction algorithms could be useful to check Covid prediction for the upcoming months and years in order to reduce the spread of Covid-19.

**References**

[1] <https://ourworldindata.org>

[2]<https://www.instituteforgovernment.org.uk/charts/uk-government-coronavirus-lockdowns>

[3] <https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Australia>

[4]<https://en.wikipedia.org/wiki/Swedish_government_response_to_the_COVID-19_pandemic>

[5]<https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/928729/S0803_CO-CIN_-_Time_from_symptom_onset_until_death.pdf>

[6] <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>