Analysis of COVID-19 cases and lockdowns in Australia

Summary

COVID-19 is an infectious respiratory disease, caused by SARS-CoV 2 virus which has spread worldwide; causing a pandemic. This disease is highly contagious and can spread from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe. It was first discovered in Wuhan, China in December 2019. It then spread across the globe hitting Australia in late January 2019.

Description/goals

Our group would like to observe COVID-19 cases within Australia and how they affect lockdowns within each state. We will also be examining how the number of COVID -19 patients affect the length of lockdowns, when it is necessary for a lockdown to occur.

Our goals include:

- A heatmap of Australia showing the number of COVID-19 cases in each state during each month of the year.
- Prediction of when it hits a certain number of cases a lockdown will occur within the state with the length of the lockdown.
- Prediction of deaths by COVID-19 depending on the percentage of people getting the vaccinated and how that contributes to lockdowns.

Datasets

The dataset that we will be using was obtained from M3IT's COVID-19_Data repository on GitHub. Because the dataset is updated daily, we have decided to use the data only up to 1/10/2021 as we downloaded the dataset 1/10.

The dataset is in a tabular form. The dataset details the confirmed COVID-19 cases, deaths caused by COVID-19, tests, recovered patients, hospitalized patients, patients in ICU and ventilation and the vaccines for each Australian state and the date the data was recorded. For our case, we will remove the hospitalized, recovered, ICU, and ventilation patients columns as we are more concerned with the number of cases rather than the severity.

We will also be adding more data by going through news articles/reports to find out when certain states of Australia were declared in lockdown as this dataset does not have this information.

Dataset source:
https://github.com/M3IT/COVID-19_Data/blob/master/Data/COVID_AU_state.csv

Techniques

The techniques we plan to use to analyse our dataset include:

- K-means clustering - To create heatmaps and easier visualise the map of Australia of COVID-19 cases
- K-nearest neighbour - Predicting if the state would be in lockdown or not depending on the number of cases.
- Linear Regression - Predicting the length of the lockdown and death rates.

Project plan

Our project plan will follow in these milestones

Milestone #1 - Preparing and Cleaning the data

We will retrieve the two datasets and join them together. Once joined we will clean the dataset of any outliers and data not needed for the analysis. Now we could use the dataset to display obvious trends in case numbers, vaccinations, and deaths through heat maps and graphs.

Due Date: Week 9

Milestone #2 - Use data and techniques to make predictions about lockdowns.

Now we will use the trends observed to then make predictions by using techniques like linear regression and K-nearest neighbour of when Australia will be in lockdown depending on the amount of cases and the length of the lockdown.

Due Date: Week 12

Milestone #3 - Analyse the predictions and confirm the outcomes

With the new data obtained by the predictions, we can now make an analysis and answer the questions we have set for ourselves.

Due Date: Week 13