The business problem

The United Kingdom's government would like to launch a targeted advertising campaign to encourage people to get vaccinated against Covid-19. They want to understand:

1. The rate of vaccination uptake across different provinces in the United Kingdom.
2. Where the campaign should be targeted based on:
    a. The trends in recoveries overtime in a province or state.
    b. The trends in deaths overtime in a province or state.
    c. The number of people who have received a first dose but not a second dose of the vaccine.
3. Regions which have experienced a peak in hospitalizations.
4. The types of tweets that use the hashtags #vaccinated and #covid.

A note on the data

The data containing the number of daily covid cases, hospitalizations, deaths, and recoveries will be referred to as the covid data in this report. The data containing the daily number of vaccinations will be referred to as the vaccination data in this report.

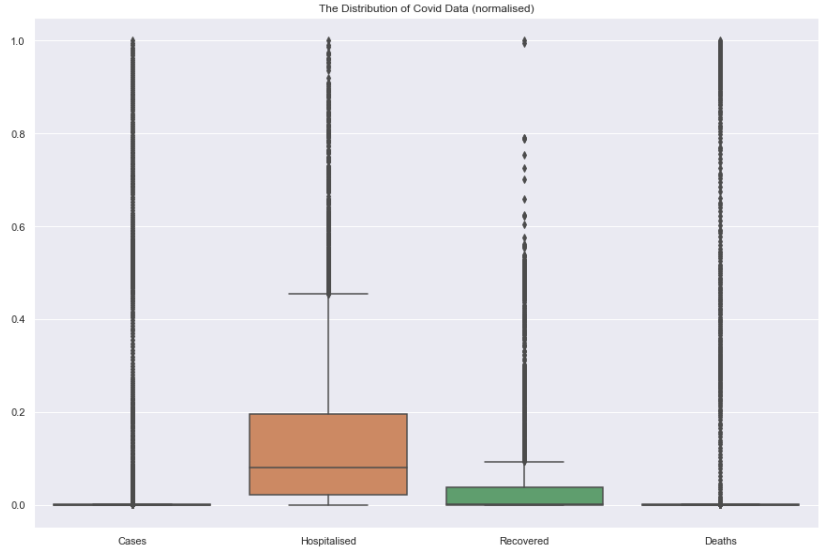
Both data sets are recordings of observations per day between the 1st of January 2020 and the 14th of October 2021. The data evolves with time and can be classified as time series data. This means that aggregations of values must be handled carefully – especially with the vaccination data. People who have received the first dose but not the second dose will likely become fully vaccinated six to eight weeks later. Aggregating the total number of people who have received the first dose over the period of study will not reflect the total number of partially vaccinated people currently in a Province/State accurately. It is better to study the trends overtime to make conclusions.

Initial exploration

To become familiar with the data, the column names and data types were identified. The date column was listed as an 'object' data type. The data shows a progression over time, so the date column is converted to a datetime object to make indexing and visualisation easier.

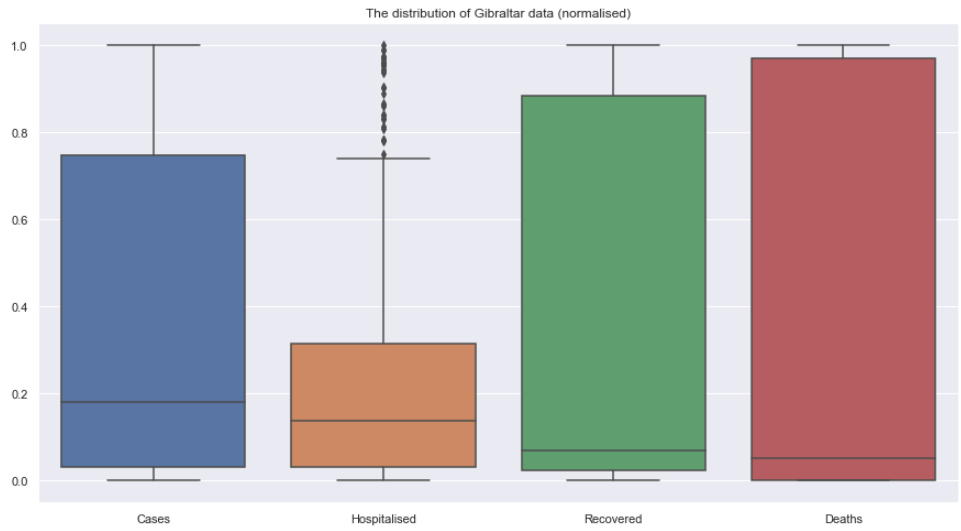
There were two observations with missing values in the covid data, and none in the vaccine data. The missing entries in the deaths, recovered, cases and hospitalized columns are imputed using the forward fill method. Mean imputation would not have been appropriate because the data is timeseries data, and it does not account for seasonality or trend components. Forward filling will keep the data moving in line with other data points around it.

The distribution of the data is shown on the boxplot below.



The box plots indicate that the data across all categories is skewed towards the right. This could indicate that there are provinces with extremely high case (or hospitalised, recovered etc) numbers. These may be erroneous.

The distribution of the province Gibraltar is shown on the two box plots below.



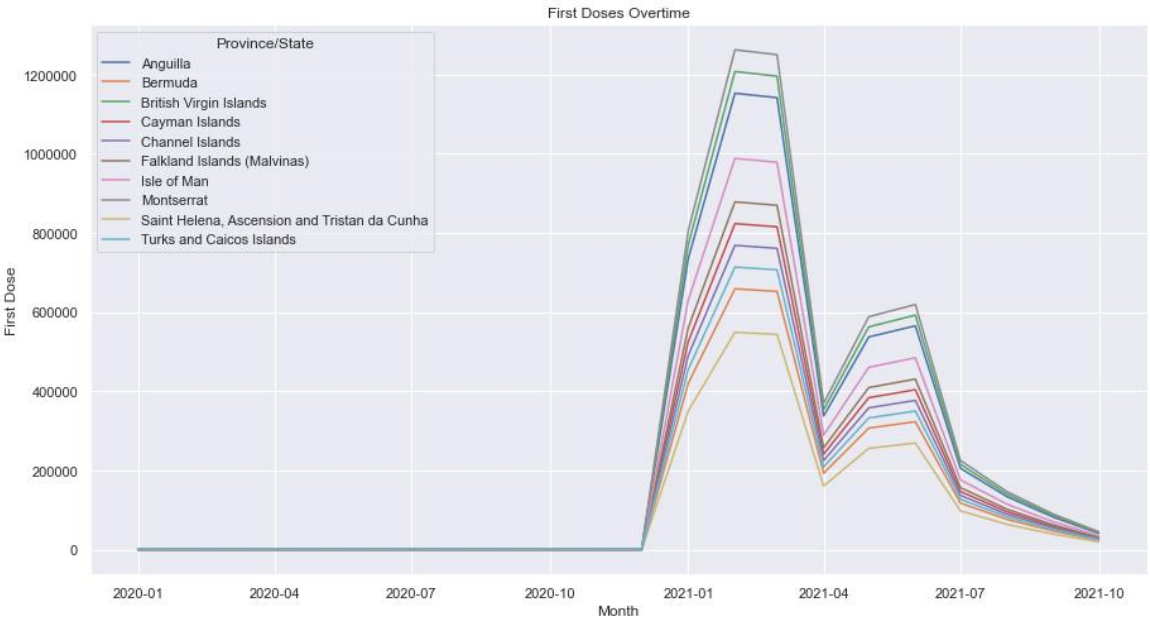
Gibraltar's maximum number of deaths exceeds its maximum number of cases. This would mean that more people died from Covid-19 on a single day in Gibraltar than the highest number of daily cases recorded. This is unreasonable and implies that something is wrong with the Gibraltar data. The data is removed from the covid data set. To keep the provinces under study consistent, Gibraltar is also removed from the vaccine data set.

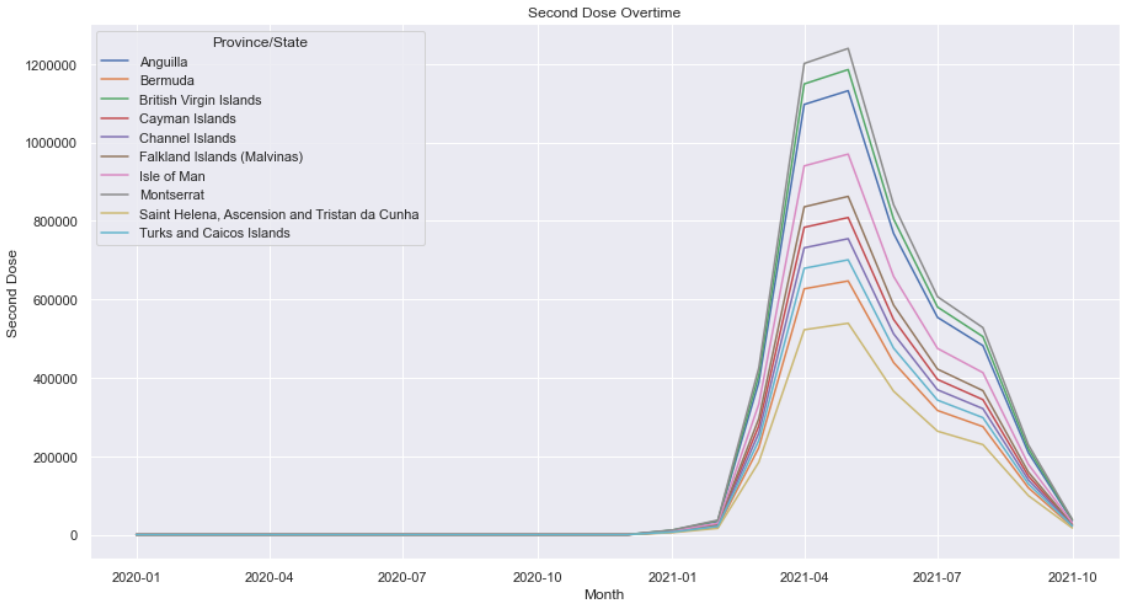
Vaccination insights across provinces

In the final full month of the study, the Isle of Man had the highest percentage of partially vaccinated people. Saint Helena, Ascension and Tristan da Cunha had the highest number of partially vaccinated people.

The ratio of partially vaccinated individuals to fully vaccinated individuals is close to one. This means that at the end of the period, for every partially vaccinated person there is a fully vaccinated person. No province stands out as having more partially vaccinated people than fully vaccinated people.

To pinpoint provinces to target, vaccinations over time are considered.



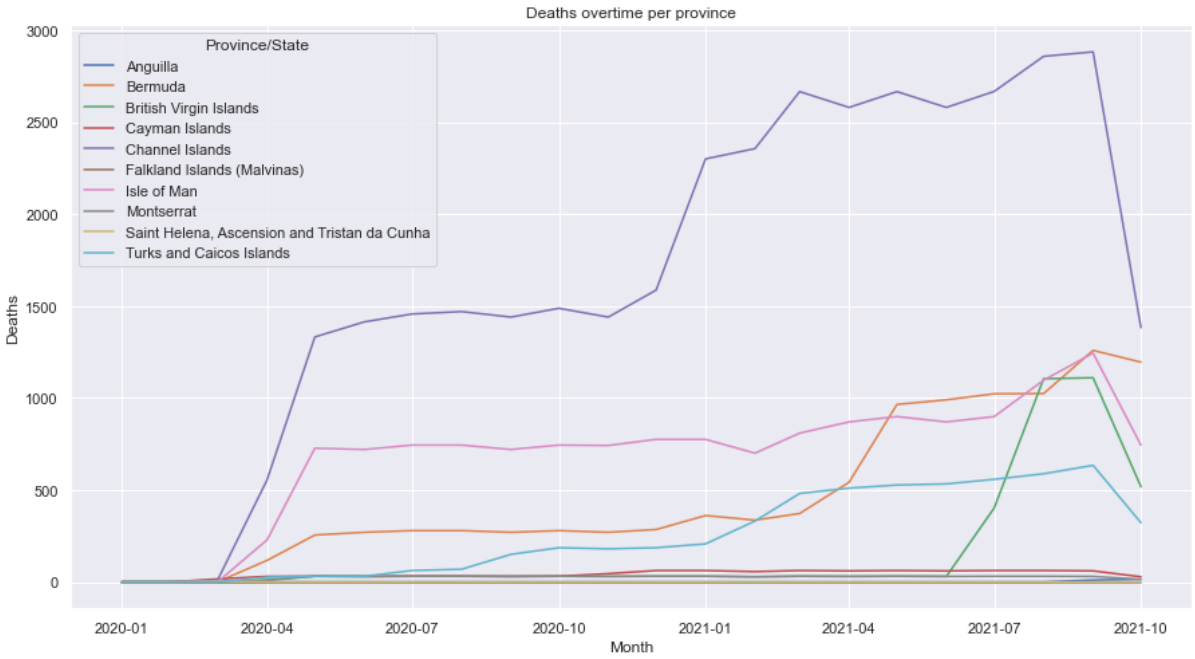


All provinces follow the same pattern. The number of daily second and first doses is declining. Between July 2021 and October 2021, the rate at which first doses are declining is slowing down. This could mean that first doses will begin to increase in a few months, or they will plateau.

In addition to this, the ratio of partially vaccinated individuals to fully vaccinated individuals is close to one. This means that at the end of the period, for every partially vaccinated person there is a fully vaccinated person. No province stands out as having more partially vaccinated people than fully vaccinated people.

The rate of vaccine uptake is uniform across provinces.

Analysis of deaths and recoveries across provinces

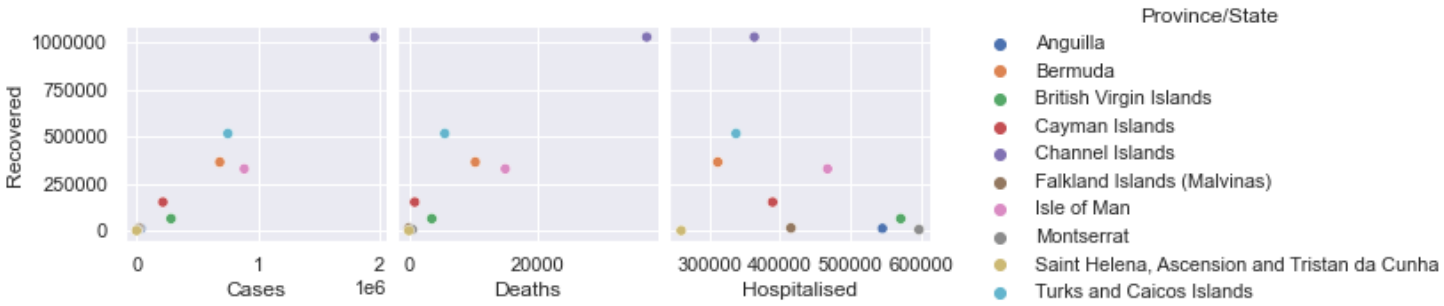


Turks and Caicos, The Isle of Man, The British Virgin Islands, The Channel Islands and Bermuda have had the largest number of deaths per month. From the above two graphs, it can be observed that the following provinces had a low number of monthly first and second dose vaccinations over the time period:

- The Channel Islands
- The Isle of Man
- Turks and Caicos
- Bermuda

These provinces may be suitable targets for the Government's marketing campaign.

The following pairplot below shows recoveries aggregated over the whole period plotted with the number of cases, deaths, and hospitalizations.



The Isle of Man has more deaths, cases and hospitalisations than recoveries. As previously mentioned, it also has the third lowest vaccination rate for both second and third doses. The Isle of Man remains a target for the government campaign.

Turks and Caicos also remains a target: for close to one million cases, they had just over 500 000 recoveries. The death numbers are low, but their vaccine uptake is also low. They had just over 300 000 hospitalizations. The British Virgin Islands also had many hospitalisations, despite a low number of deaths. This fact may be used in the campaign to encourage people to remain out of hospital by getting vaccinated.

The Channel Islands can be avoided as they had a relatively low number of deaths for close to two million cases.

Saint Helena, Ascension and Tristan da Cunha and Montserrat can be avoided. They have low case numbers and low death reports.

Have hospitalisations reached a peak?

Isle of Man, Turks and Caicos and the British Virgin Islands have been identified areas for the government to target.

Using the previous consultant's work, the data is plotted using a seven-day moving average.

The time series plot for the Isle of Man is shown below:

The peak can be identified between January 2021 and March 2021. On the upward slope, the actual values get close to the upper bound line. The data then plateaus and starts decreasing. The actual values get very close to the lower bound. This indicates that a peak has been reached.

The graphs below plot the time series for Turks and Caicos, and the British Virgin Islands respectively. One can see that they follow a similar patter to the Channel Islands, and that a peak in hospitalizations was reached in February of 2021.

Analysis of Twitter data

Common words associated with tweets about covid are Boris Johnson, 'protection', 'deaths' and 'signs'. While this does not say too much about the sentiment towards the Covid Vaccine, it can be noted that people are talking about the vaccine, Covid-19 and protecting themselves. The best hashtags to use are "Covid19" and "CovidIsNotOver". This corresponds with the indication from the covid data that a peak in Covid-19 infections has passed. People may be becoming more relaxed with restrictions and cases may begin to rise again. #CovidIsNotOver is a good existing hashtag to leverage in the campaign.