Course 2 Assignment – Data Analytics Using Python – UK covid-19

1. Introduction:

To better understand the current Covid-19 pandemic. The United Kingdom's government would like to gain some insights, using a data driven approach, on the number of vaccinated and hospitalised individuals across its overseas territories from January 2020 to October 2021.

The results of this data analysis project will be used to support the government's strategy in launching a series of marketing campaign to promote vaccination. The questions listed below will help in finding trends and patterns in the existing covid-19 situation.

Questions:

- Q1. What is the total vaccination per region?
- Q2. Where should the first marketing campaign be targeted?
- Q3. What other types of Twitter data points and tweets have both #coronavirus and #vaccinated hashtags?
- Q4. Which regions have experienced a peak in hospitalisation?

2. Materials – methods – analysis - results

2.1Description of each data sets:

a. Covid cases:

Contains the number of covid cases, deaths, hospitalised and recovered. The data set is made of 7584 rows and 12 columns which are shown in the following table along with their data types and missing values if any:

Column name	Data type	Missing values (if any)
Province/State	Object	0
Country/Region	Object	0
Lat	Float	0
Lon	float	0
ISO 3166-1 Alpha 3-Codes	Object	0
Sub-region name	Object	0
Intermediate region code	Integer	0
Date	Object	0
Deaths	Float	2

Cases	Float	2
Recovered	Float	2
Hospitalised	Float	2

The IDs with missing values are 875 and 876, located in Bermuda.

b. Vaccinated:

Contains the number of covid cases, deaths, hospitalised and recovered. The data set is made of 7584 rows and 11 columns, which are shown in the table below, along with their data types and missing values if any:

Column name	**Data type**	**Missing values (if any)**
Province/State	Object	0
Country/Region	Object	0
Lat	Float	0
Lon	float	0
ISO 3166-1 Alpha 3-Codes	Object	0
Sub-region name	Object	0
Intermediate region code	Integer	0
Date	Object	0
Vaccinated	Integer	0
First dose	Integer	0
Second dose	Integer	0

2.2 Methods and analysis:

Q1. Total vaccination for a particular region?

The table below shows the total number of people who have received their first dose and second dose. Where second dose is the equivalent as being fully vaccinated. From January 2020 to October 2021, Gibraltar has received the highest number of first dose (5870786) and the highest number of people being fully vaccinated (5606041). In contrast Saint Helena et al have the lowest number of first dose (2348310) and fully vaccinated (2242421).

Province/State	First Dose	Second Dose
Gibraltar	5870786	5606041
Montserrat	5401128	5157560
British Virgin Islands	5166303	4933315
Anguilla	4931470	4709072
Isle of Man	4226984	4036345
Falkland Islands (Malvinas)	3757307	3587869
Cayman Islands	3522476	3363624
Channel Islands	3287646	3139385
Turks and Caicos Islands	3052822	2915136
Bermuda	2817981	2690908
Others	2583151	2466669
Saint Helena, Ascension and Tristan da Cunha	2348310	2242421

Figure 1 below, displays a line plot for first dose and second dose administered. For the first dose, all states follow a similar pattern. They start receiving the first dose in January 2021 and reach a first peak around March 2021. It reaches a second and lower peak in summer 2021.

Following the wave of first dose vaccination is the wave of second dose vaccination, which kick-in by the end of March 2021. And sees a steady rise between April 2021 and June 2021. And starts to decline by July 2021.

The usual time delay between taking first and second dose is 8-12 weeks (NHS, 2022). This could partially explain the time delay between first dose and second dose vaccination.

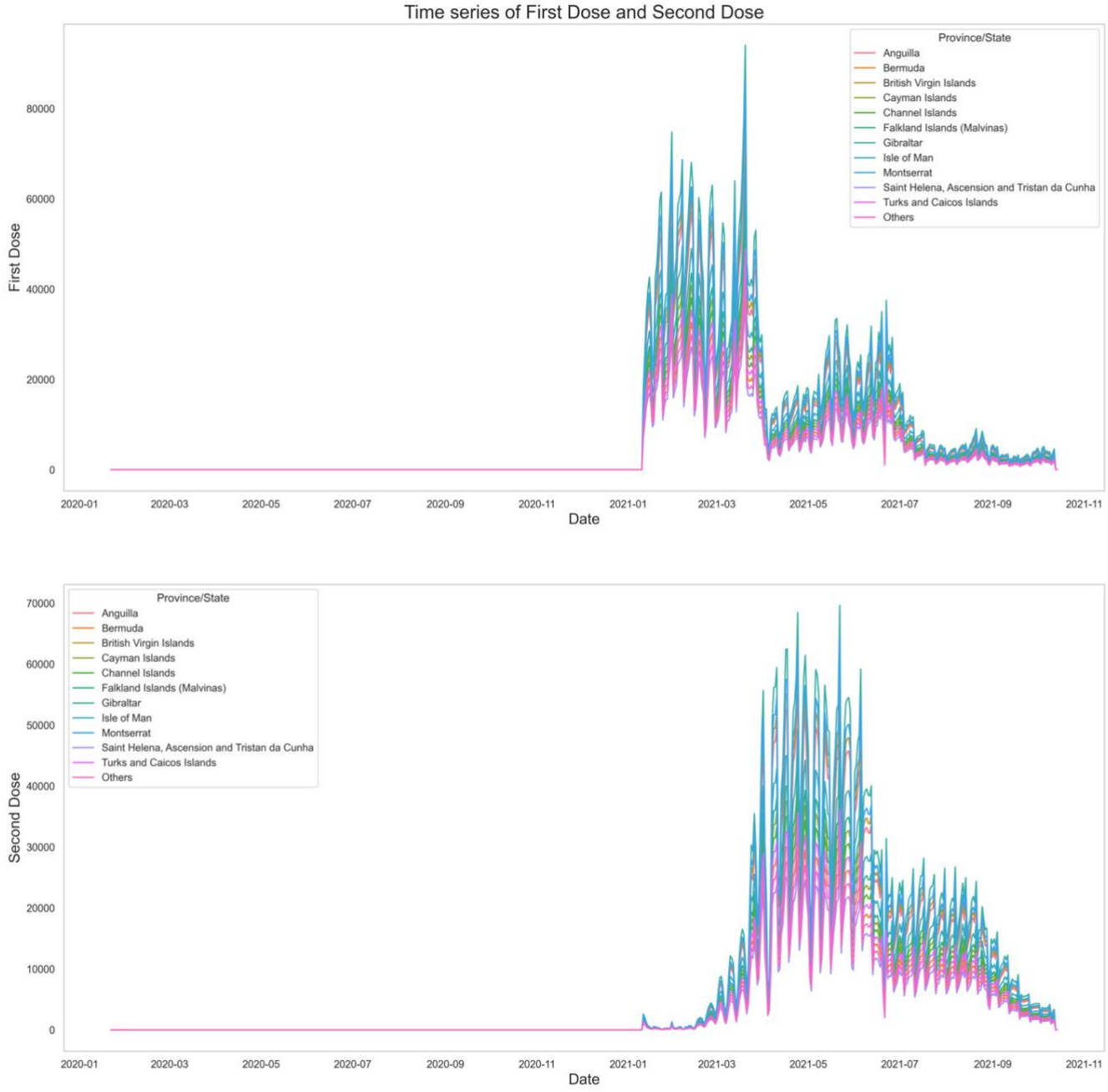


Figure 1. Shows the time series of first dose and second dose. Province/Sates are being colour coded.

Q2. Where should the first marketing campaign be targeted?

The marketing campaign should focus on regions who have the highest percentage of partially vaccinated people. Where partially vaccinated people, are people who have only received their first dose. This is illustrated in the table below.

Table 2 shows the number of people who are partially vaccinated and the percentage of people who have only received their first dose. Gibraltar has the highest number of partially vaccinated people (264745); however, it does not have the highest percentage of partially first dosed vaccinated people (23.068%). It is Turks and Caicos Islands that have the highest percentage of first dosed vaccinated people (23.071%).

Province/State	Vaccinated	First Dose	Second Dose	Partially vaccinated	percentage of first dose only
Turks and Caicos Islands	2915136	3052822	2915136	137686	0.023071
Isle of Man	4036345	4226984	4036345	190639	0.023070
Anguilla	4709072	4931470	4709072	222398	0.023069
British Virgin Islands	4933315	5166303	4933315	232988	0.023069
Cayman Islands	3363624	3522476	3363624	158852	0.023069
Channel Islands	3139385	3287646	3139385	148261	0.023068
Montserrat	5157560	5401128	5157560	243568	0.023068
Falkland Islands (Malvinas)	3587869	3757307	3587869	169438	0.023068
Gibraltar	5606041	5870786	5606041	264745	0.023068
Bermuda	2690908	2817981	2690908	127073	0.023067
Others	2466669	2583151	2466669	116482	0.023067
Saint Helena, Ascension and Tristan da Cunha	2242421	2348310	2242421	105889	0.023066

From the data above, it is difficult to determine in which region the marketing campaign should start solely based on the percentage of people who have received their first dose only.

Figure 2 shows the number of covid-19 cases, and number of recovered cases over time. The Channel Islands have the highest number of recovered cases compared to other regions. Whereas the British Virgin Islands and the Cayman Islands have the lowest number of recovered cases. This is because, the latter have a much lower number of Covid-19 cases compared to the two other regions.

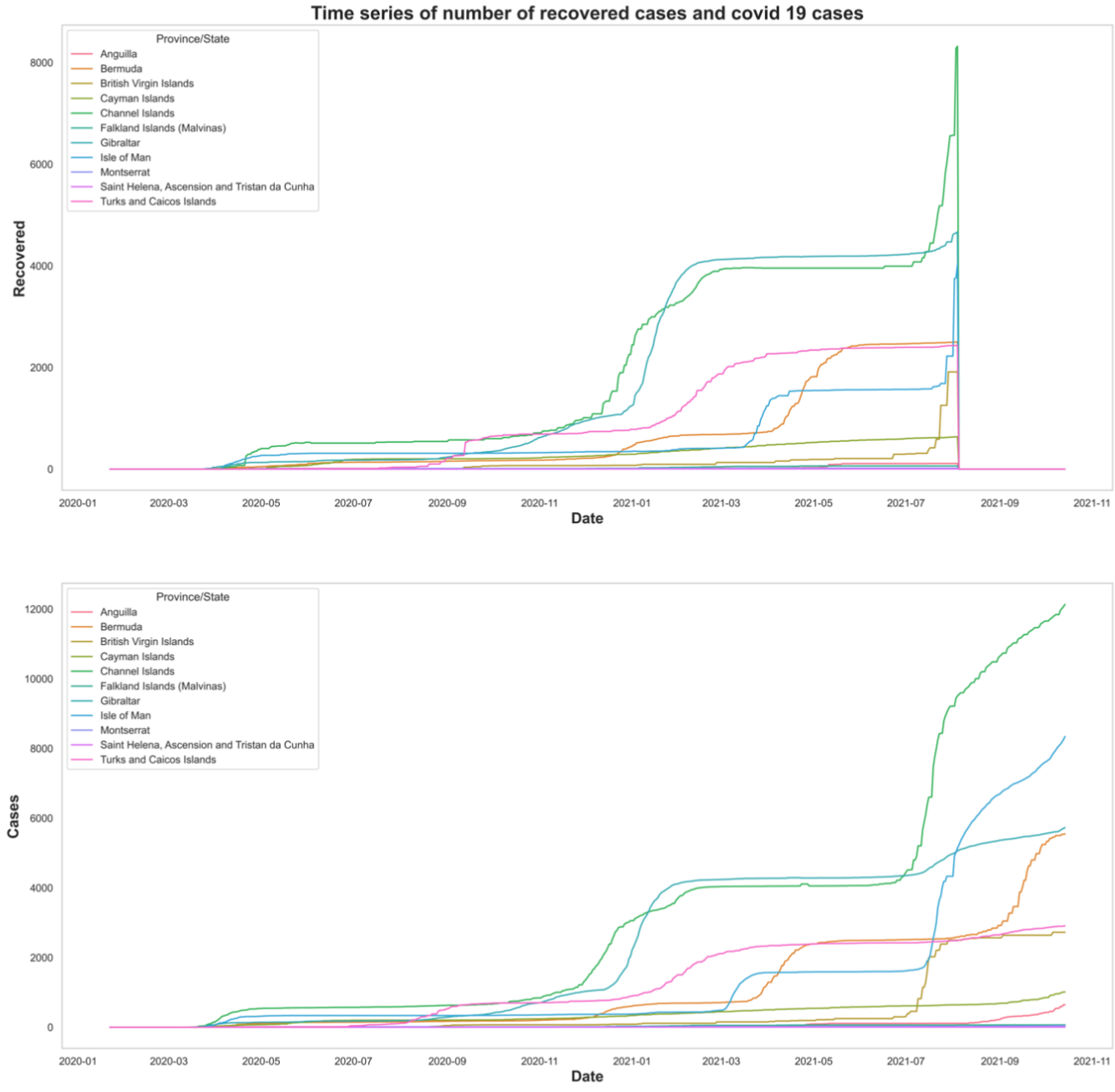


Figure 2. Top charts show the number of recovered cases, whereas bottom charts show the number of covid-19 cases over time. The "Others" state/province has been removed from this data sets as it displayed very high and unusual numbers. This shown in figure 1 of the appendix.

Figure 3 shows the number of deaths per region. In each regions the number of deaths has not reached a peak. Its pattern is a steady increase, followed by a plateau and then rises again. This is particularly the cases for Channel Islands, Bermuda and Gibraltar. Montserrat and British Virgin Islands have a very low number of covid 19 cases, which has been stagnant for a year.

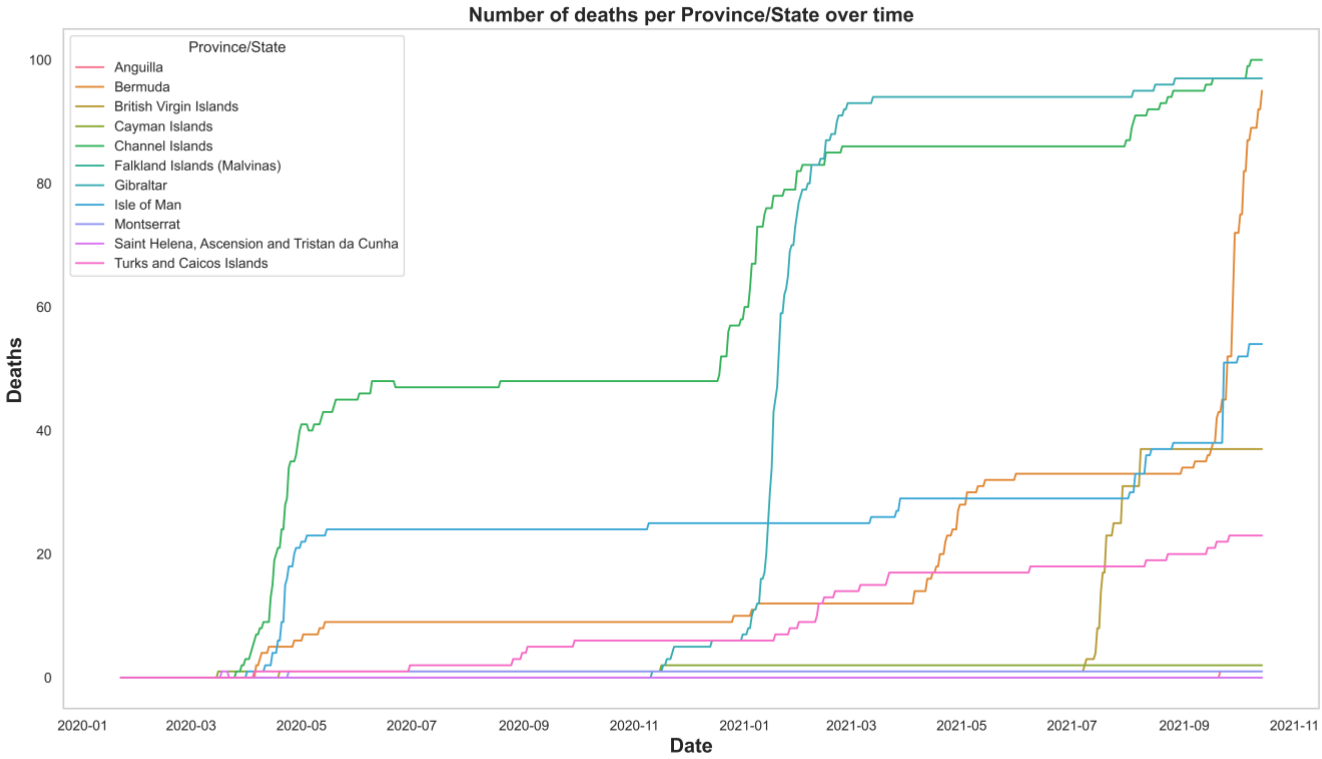


Figure 3. Deaths per region.

Q3. What other types of Twitter data points and tweets have both #coronavirus and #vaccinated hashtags?

Out of the 4776 tags, 2120 contained the hashtag related to covid or covid-19. This is shown in figure 4 below. Qualitative information from social media could help bring a better understanding of the existing situation in different provinces/state.

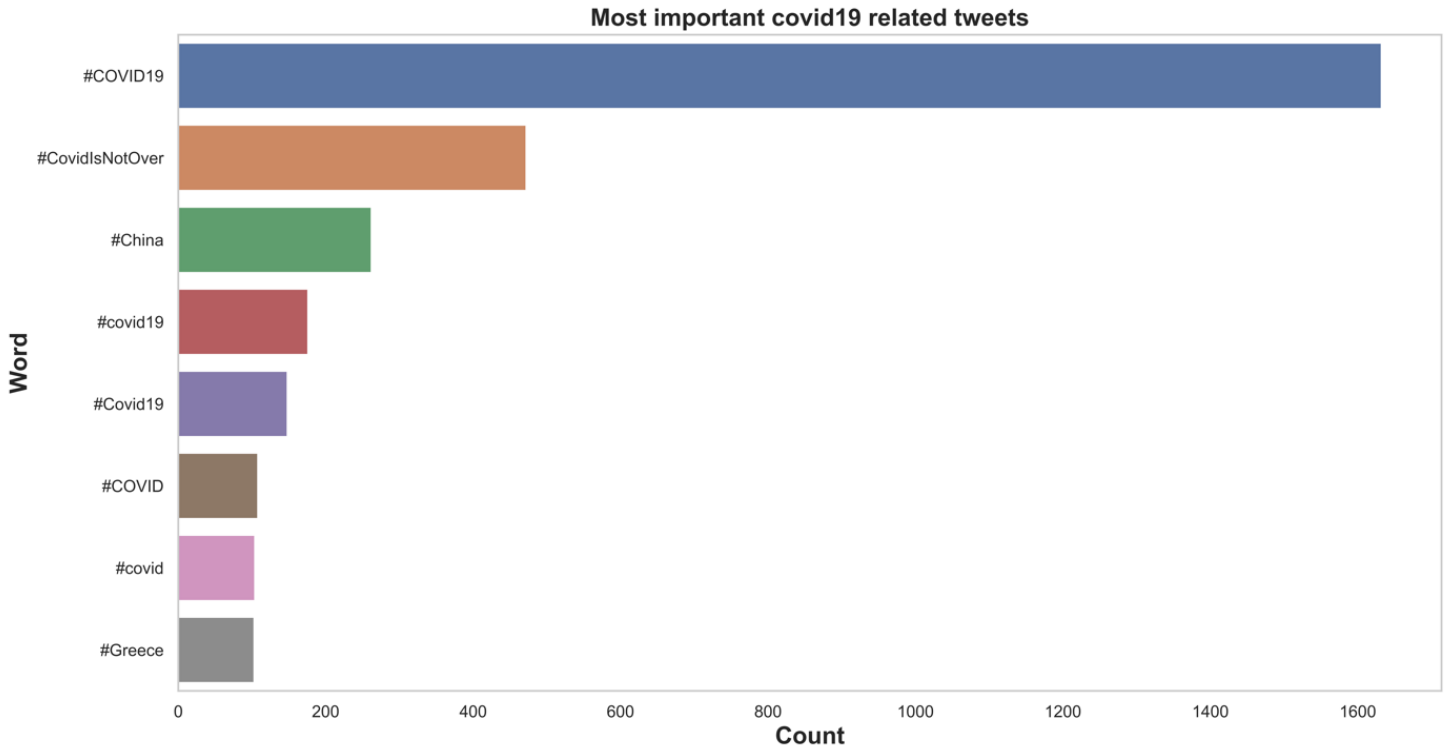
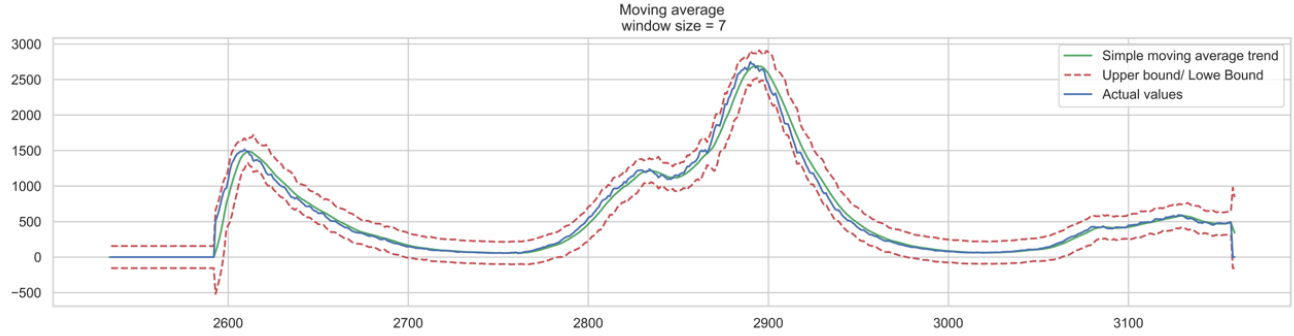


Figure 4. Shows the most important tweets related to covid-19 from the 'tweets data set'.

Q4. Which regions have experienced a peak in hospitalisation?

Figure 5 shows a time series forecasting using a 7, 15, 30 days moving average. This is used to predict the number of hospitalisations. In this scenario, the figure shows the hospitalisation number for the Channel Islands. The time series has a higher accuracy with a lower moving average. The Channel Islands have reached a peak in terms of hospitalisation.

Other states and provinces have also reached their hospitalisation peak, and this is show in figure 6.



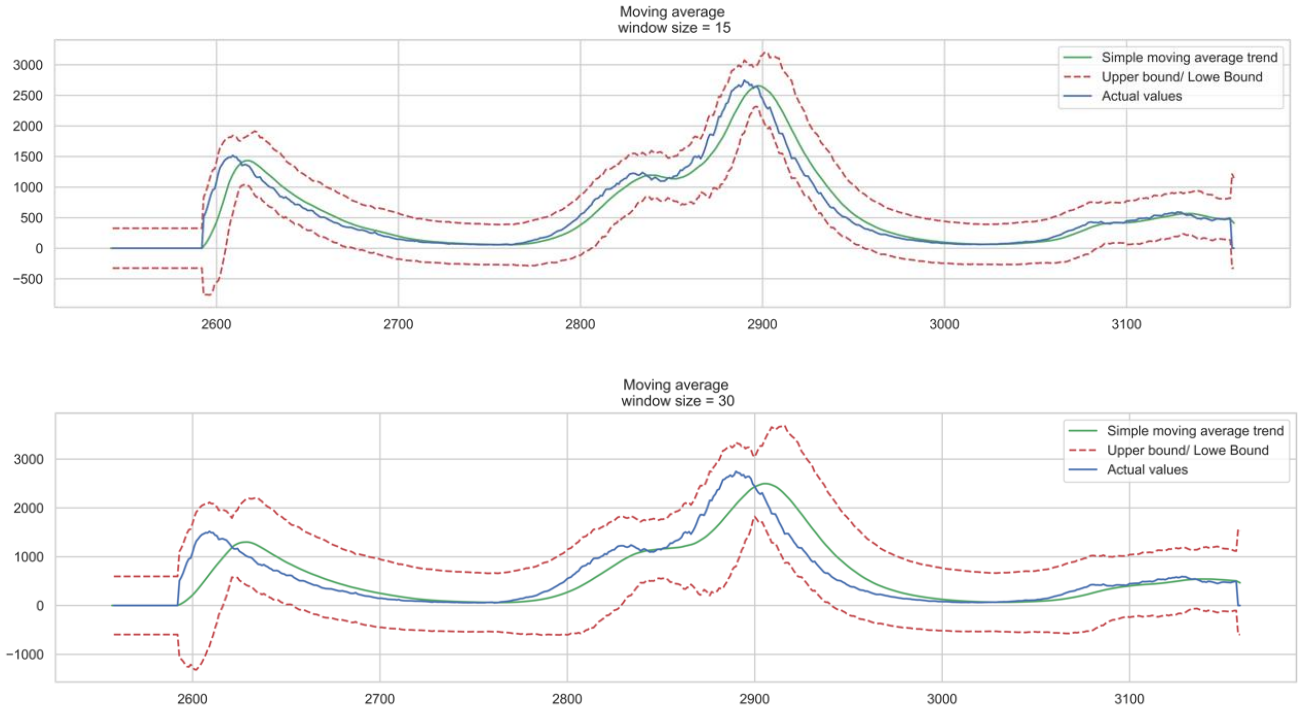


Figure 5. Shows a timer series moving average for a window period of 7, 15 and 30 days.

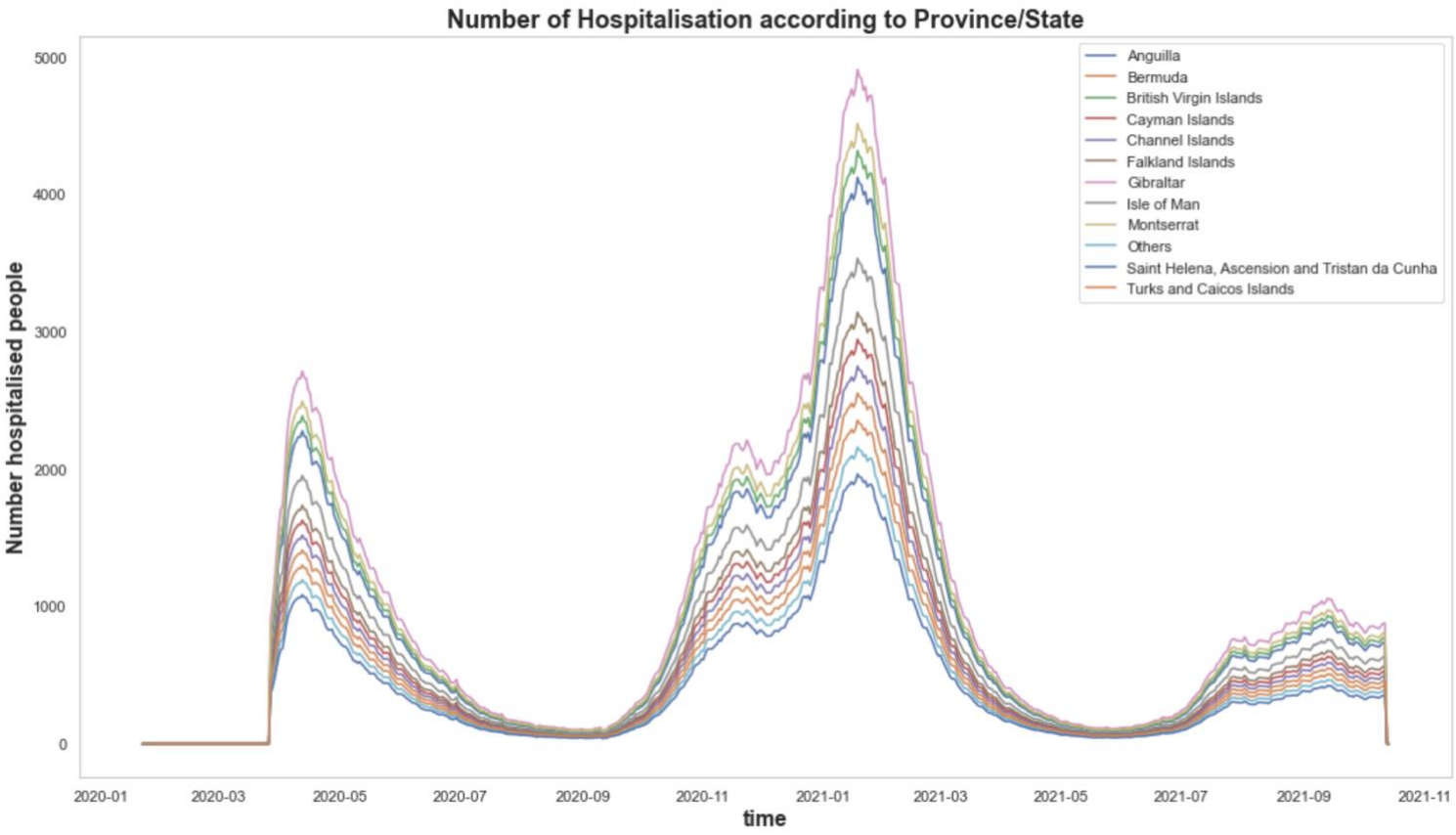


Figure 6. Shows the hospitalisation number for different provinces/states.

3. Conclusion

To summarise, all regions have accomplished a full vaccination rate of nearly 80%. The number of deaths seems to increase so is the number of covid cases and recovered cases. From the given data sets it is difficult to single out one region where the marketing campaign should first start.

Both the vaccination and hospitalisation rate follow a pattern that is too similar for each region. Thus, the data should be further investigated and verified. For instance, the “Others” region has an unusual number of covid-19 cases.

Some recommendation for future analysis includes:

- Diversity of the data: Try to collect demographics data, understanding which part of the population is at risks (Children, elderly). Who dies, and who has been vaccinated. This could further enhance the marketing campaign.
- Data collection: verify how the data are collected, to ensure accuracy and quality. Because good data provides good analysis. And hence, supports better decision making.

4. Appendix

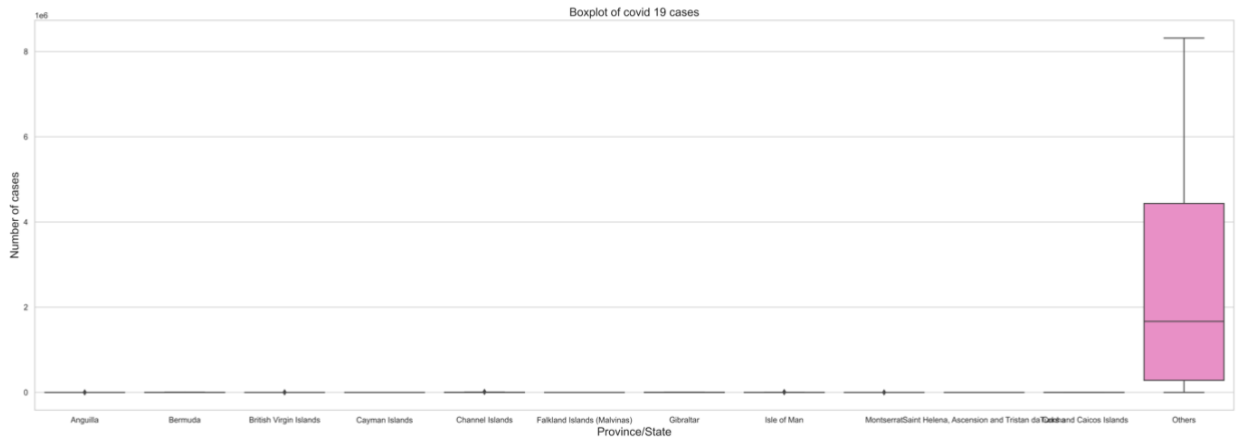


Figure 1 above, shows a boxplot of covid 19 cases. It can be clearly seen that the “Others” state is an outlier as its number of covid cases range between 0 and 8 million, whereas other state/provinces ha.

5. Reference

NHS (2022) *How to get 1st and 2nd dose of the coronavirus (COVID-19) vaccine*. Available at: https://www.nhs.uk/conditions/coronavirus-covid-19/coronavirus-vaccination/how-to-get-a-coronavirus-vaccine/how-to-get-a-1st-and-2nd-dose/