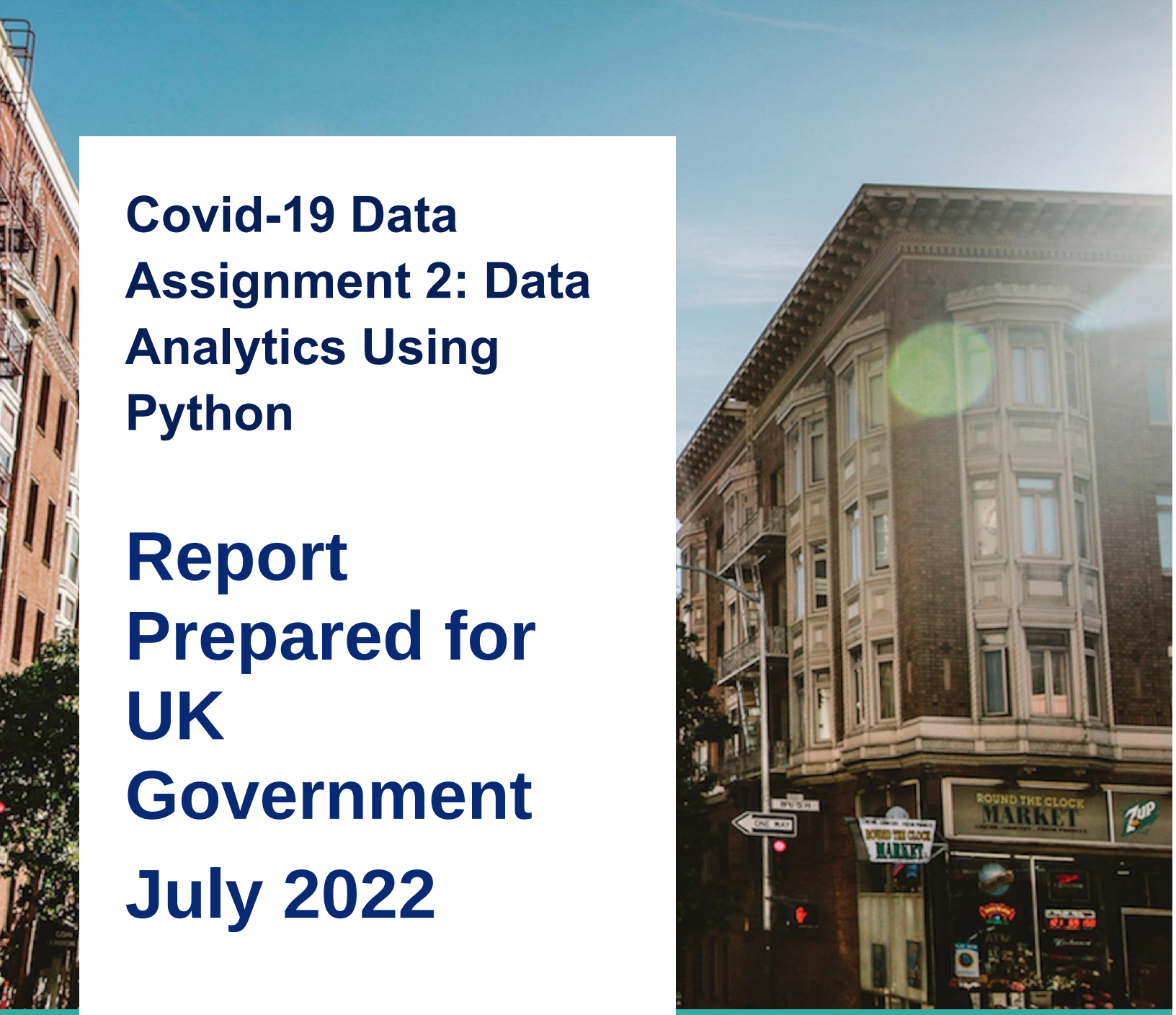


Covid-19 Data Assignment 2: Data Analytics Using Python

Report Prepared for UK Government July 2022

JULY 111 **2022**

UK Government / LSE Career Accelerator
Authored by: Zikomo Smith

Executive Summary

Background/context of the brief

The government wanted to identify trends and patterns that can be used to inform its marketing approach to increase the number of fully vaccinated people.

Row Labels	Total Vaccinations	Sum of Eligible for Second Dose	Cases	Deaths
Anguilla	4,709,072	222,398	644	1
Bermuda	2,690,908	127,073	5548	95
British Virgin Islands	4,933,315	232,988	2725	37
Cayman Islands	3,363,624	158,852	1011	2
Channel Islands	3,139,385	148,261	12135	100
Falkland Islands (Malvinas)	3,587,869	169,438	69	0
Gibraltar	5,606,041	264,745	5727	97
Isle of Man	4,036,345	190,639	8343	54
Montserrat	5,157,560	243,568	41	1
Others	2,466,669	116,482	8317439	138237
Saint Helena, Ascension and Tristan da Cunha	2,242,421	105,889	4	0
Turks and Caicos Islands	2,915,136	137,686	2910	23

Topline results

- **Hospitalisations reached peak across all regions in March of 2021** and have subsided since.
- **The Isle of Man** and **Bermuda** saw significant increases in deaths towards the end of 2021.
- **Deaths have plateaued** across the remainder of the regions.
- **Data quality remains an issue** – the total vaccinations number is not reflective of reality (Anguilla has a population of 15,090.
- **COVID tweets tend to have negative sentiment**

Topline recommendations

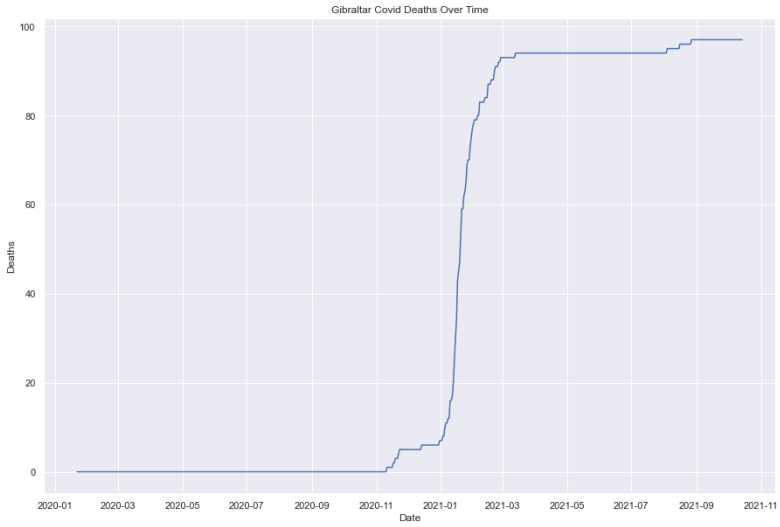
1. Target campaigns at areas with highest number of people eligible for second doses **(Gibraltar, Montserrat, British Virgin Islands.)**
2. Target campaigns at areas with increasing death rates to get them to work to keep themselves as safe as possible **(Isle of Man & Bermuda)**
3. **Include positive Twitter sentiment as a new campaign KPI**
4. **Implement data quality assurance process to ensure clean data** and to prevent repeating past mistakes.

Table of Contents

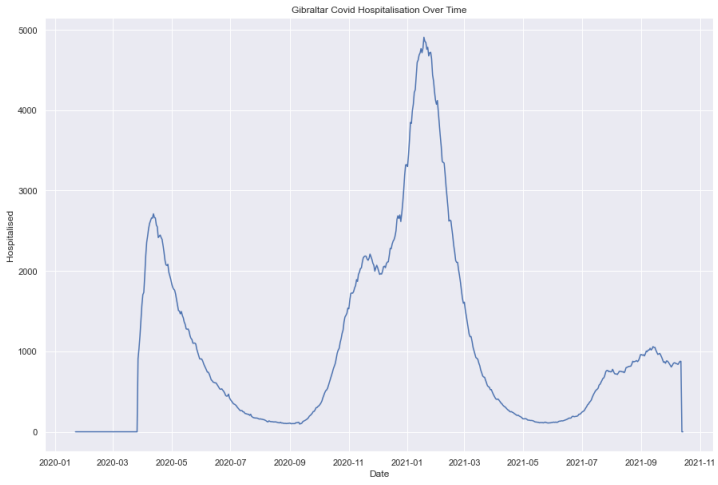
1.Analytical approach

The analytical approach was as follows:

1. *Determine the shape and type of data present and how they might inform insights*
    a. The analysis used line plot graphs to assess whether the data was cumulative or not. Based on the analysis of Gibraltar, we found that death data was cumulative as shown below:



Vaccination data, however, showed the total numbers per day as shown below:



As a result, the analysis aimed to compare spikes in the non-cumulative date, with the plateaus and rises in the cumulative data to generate insights.

The analysis also focused on trends over time data as this suited the nature of the dataset.

b. The analysis uncovered rows without data and backfilled them to ensure the completeness of the entire dataset.

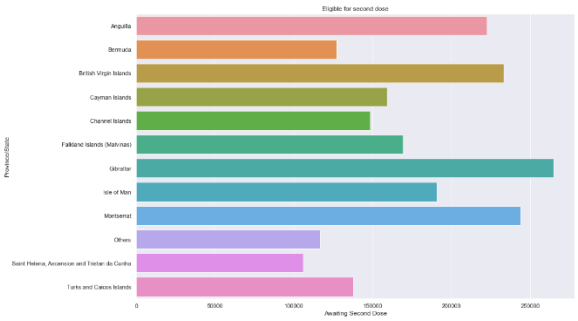
The analysis discovered that Bermuda was missing data for Deaths, Cases, Recovered, and Hospitalised fields on 21/09/2020 and 22/09/2020.

	Province/State	Country/Region	Lat	Long	ISO 3166-1 Alpha 3-Codes	Sub-region Name	Intermediate Region Code	Date	Deaths	Cases	Recovered	Hospitalised
875	Bermuda	United Kingdom	32.3078	-64.7505	BMU	Northern America	0	2020-09-21	NaN	NaN	NaN	NaN
876	Bermuda	United Kingdom	32.3078	-64.7505	BMU	Northern America	0	2020-09-22	NaN	NaN	NaN	NaN

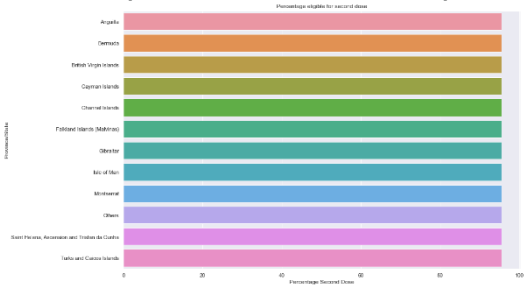
For the cumulative fields, the analysis used a backfill to ensure that the fields were filled with numbers smaller than those found on future dates. This was so the new numbers did not disturb the cumulative nature of the dataset.

2. *Determine whether absolute values or ratios provide more compelling data*
   a. The analysis aimed to provide multiple views of the dataset to allow us to find the most compelling insights. The analysis employed sum & grouping functions to ensure that we could calculate both ratios and absolute values.

   An example of an absolute value barplot:



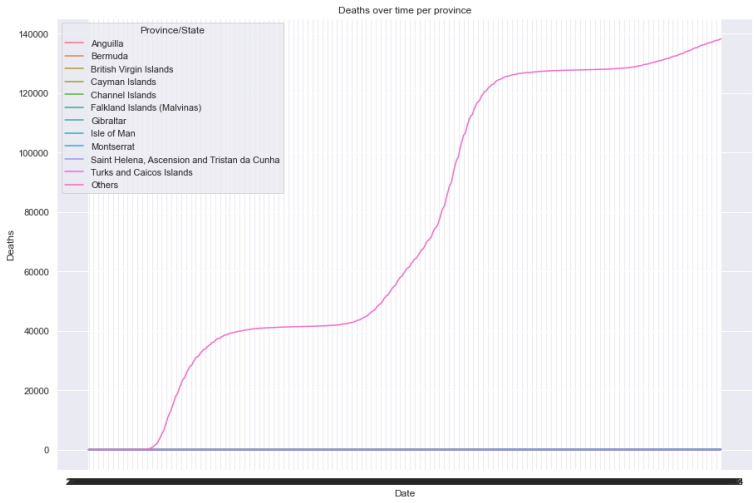
   An example of a ratio value barplot:



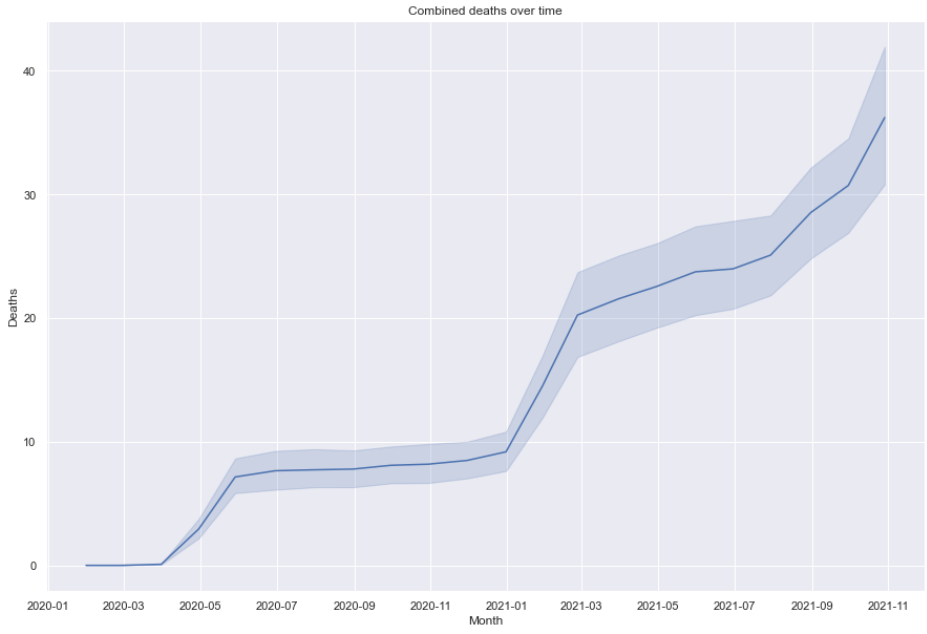
   The analysis clearly shows that the absolute value chart shows more meaningful differences. This helped inform the insights present later on in the report.

3. *Find meaningful datasets and correlations*
    a. The analysis found that the ‘Others’ Province/State dataset was skewing the overall data so that it was removed from the final analysis.

Aggregate deaths over time including ‘Others’:



Aggregate deaths over time excluding ‘Others’:



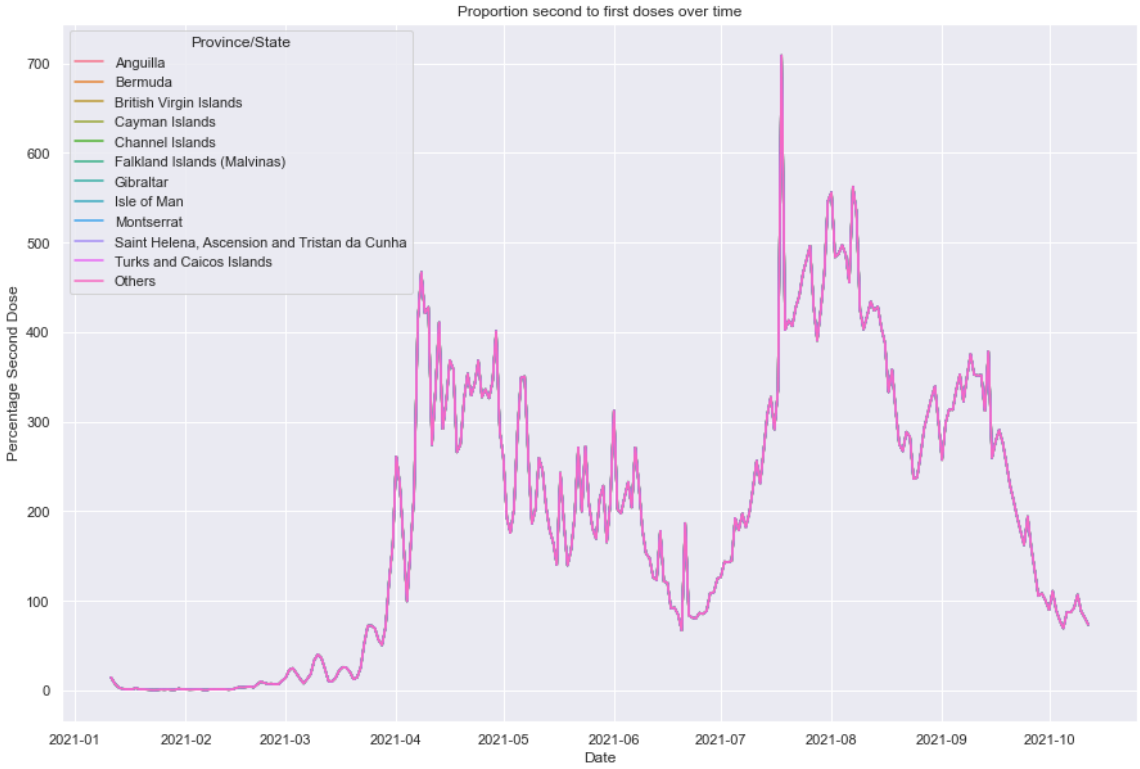
Aggregate deaths excluding ‘Others’ shows less of a plateau and captures the rise in deaths towards the end of the time.
‘Others’ is also a vague title for the data – it is not clear which UK region it refers to.

Thus, the analysis focused on the smaller Provinces/States where the data and trends were clearer and more actionable for the Government.

b. The analysis aimed to assess the context of COVID tweets and determining whether they focused on positive or negative stories by looking at clusters of hashtags.

2.Province/ State Visualizations, Patters, Trends, and Insights

1. The below chart visualises the proportion of second doses compared to the first dose as a percentage.

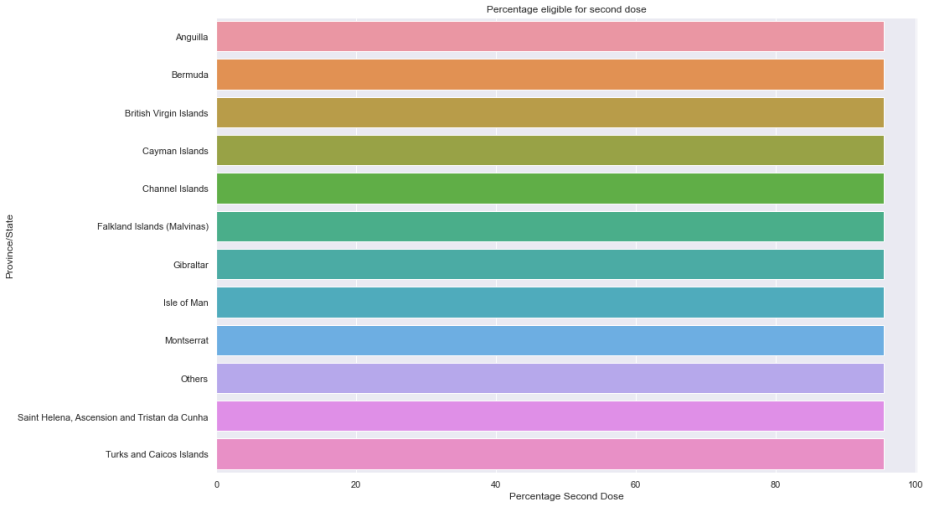


Patterns & trend

- We can see that second doses started to exceed first does in April 2021.

- This trend mostly continued throughout the year, though second doses fell below even in late June and in October 2021.

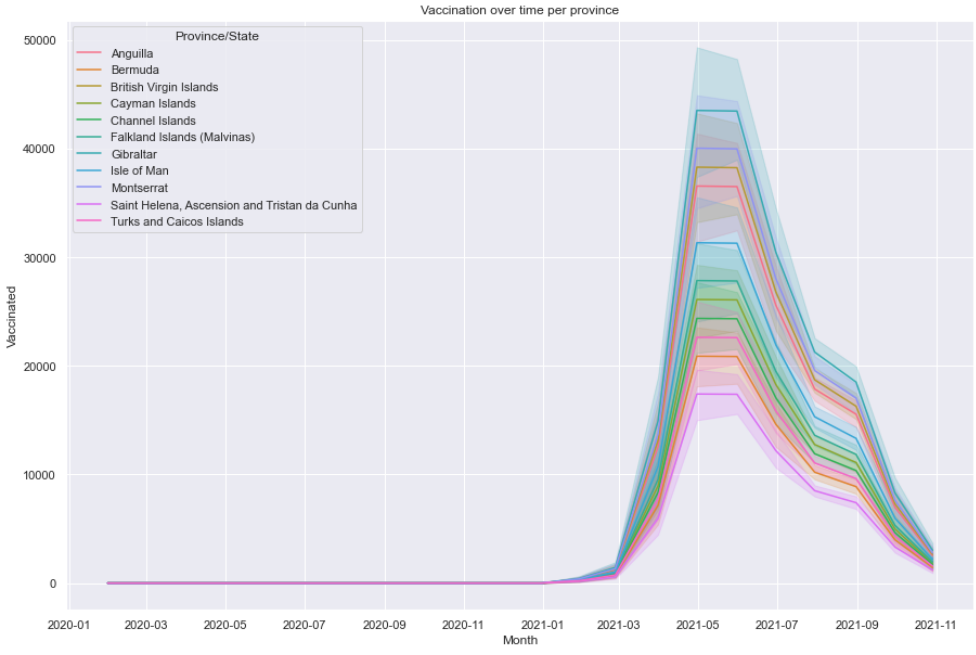
2. **The visual below the percentage of people that have had the second dose who have also had the first dose.**



Insight

- The percentage eligible chart suggests that the government is rolling out first and second vaccines in a similar manner across provinces

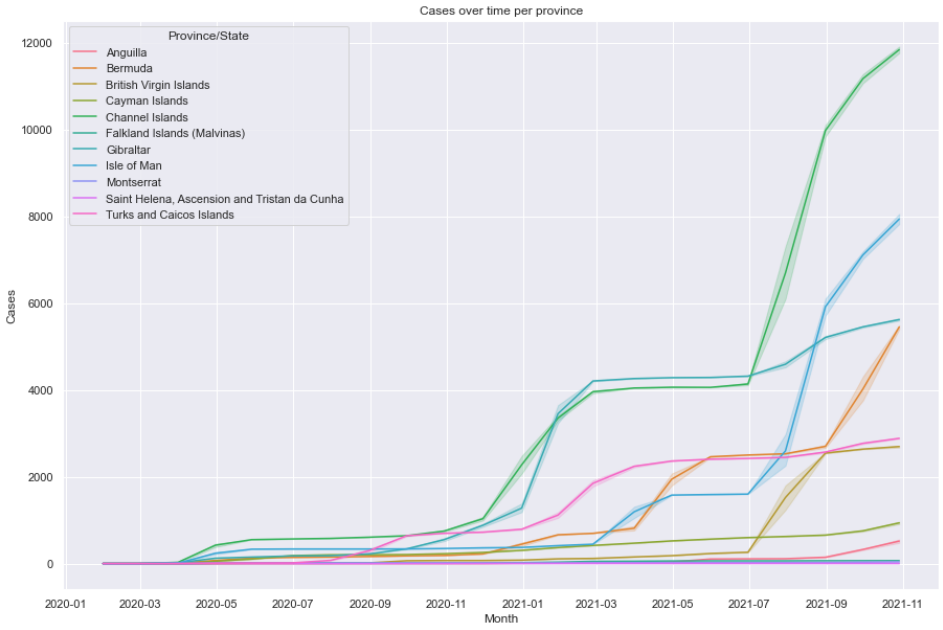
3. **The visual below shows the number of vaccinations per province over time**

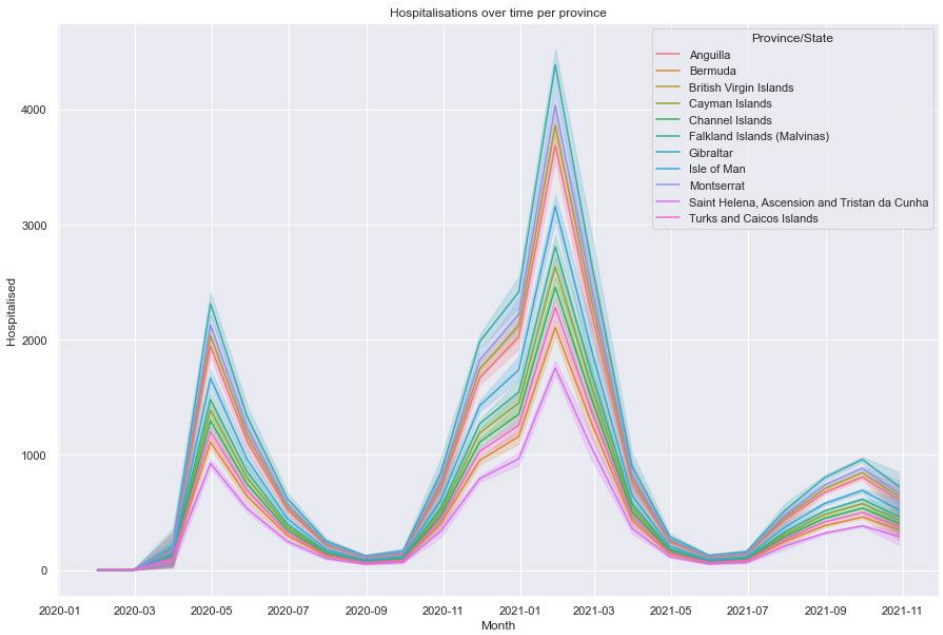


Insight

- The vaccination chart further emphasizes that the government is rolling out first and second vaccines in a similar manner across provinces given that the shape of the graph is similar across Provinces.

4. **The visual below shows the number of cases per province over time. The visual underneath that shows the hospitalisations over time per province.**

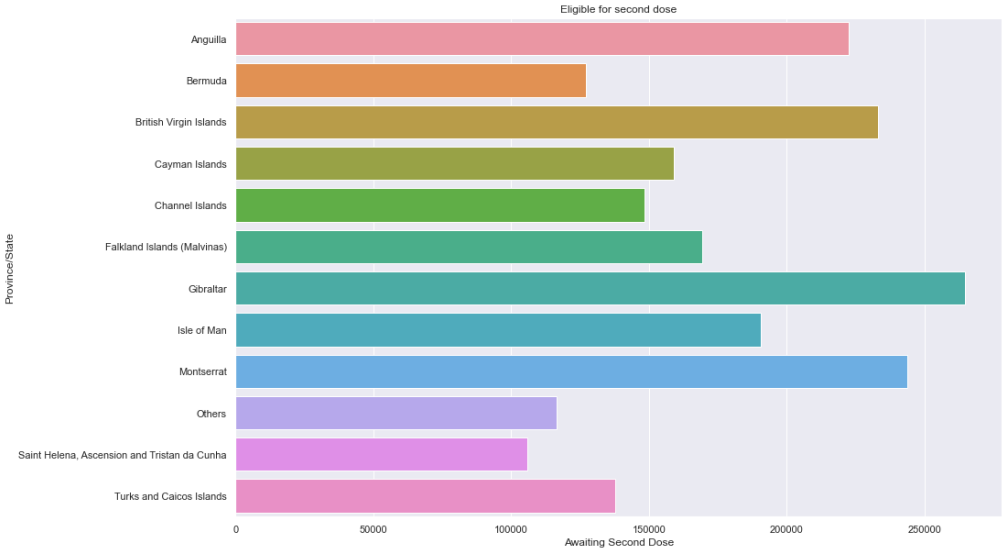


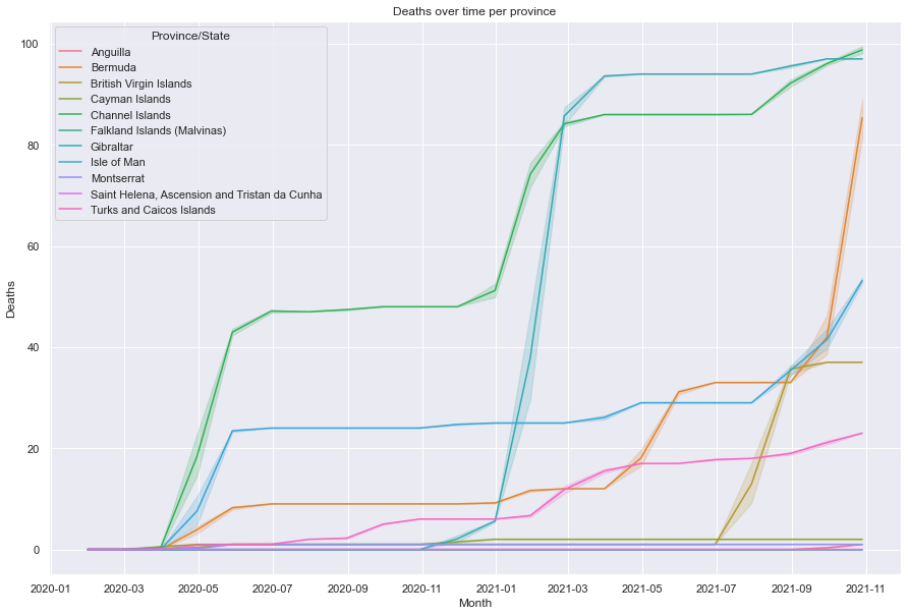


Insight

- The above charts demonstrated that cases no longer track hospitalisations, which demonstrates the effectiveness of the vaccine.
- Indeed, hospitalisations decreased after the spike in vaccinated individuals around July 2021.

5. **The visual below shows the total number of people eligible for a second dose. The second visual shows the deaths over time per province.**



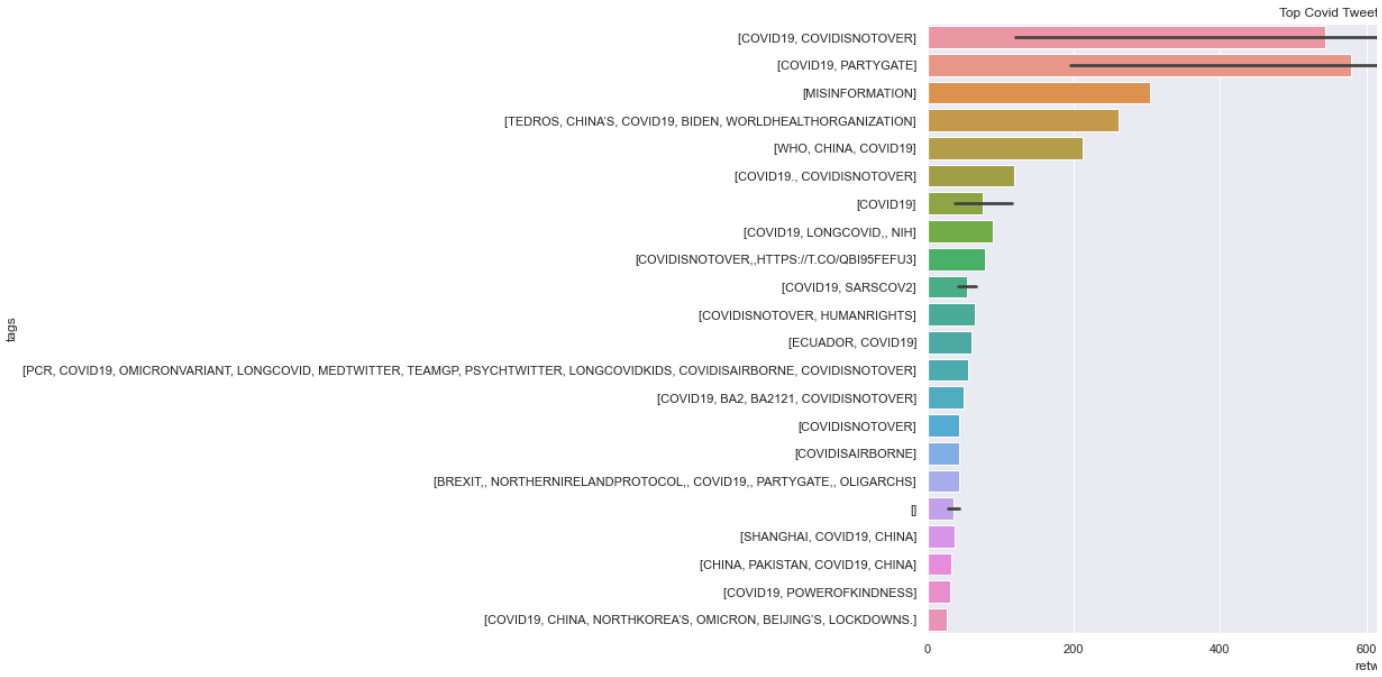


Insight

- The above charts demonstrate that each country is having a different covid experience.
- Gibraltar, Montserrat, British Virgin Islands have the highest number of people eligible for second doses.
- The Isle of Man & Bermuda have seen death rates increase towards the end of the year.
- This suggests the government may have to develop a bespoke marketing strategy per Province.

3.Twitter Sentiment Analysis

1. **The visual below shows the most retweeted covid tweets and the other hashtags they are associated with.**



Insight

- The above chart demonstrates that the top covid tweets focus on political issues (party gate), and topics likely to incur negative sentiment (misinformation, COVIDISNOTOVER, Human Rights, Oligarchs).
- The Government will want encouraging messaging that cuts through, so it would be good to see tweet sentiment change in line with marketing efforts.

Thank You