Introduction and context

To act as a data analyst working with the UK government to analyse COVID-19 data (from January 2020 to October 2021).

As part of its goal to increase the number of fully vaccinated individuals (people who have received a first and second dose of the vaccine), the UK government is planning to launch a series of marketing campaigns to promote the vaccine. The UK government wants to identify trends and patterns that can be used to inform its marketing approach to increase the number of fully vaccinated people.

Data Sources provided by UK Government

- Covid-19 Cases CSV file containing case, death, hospitalisation and recovery data for UK provinces/states.
- Covid19 Vaccinations CSV file containing first and second dose vaccinations for UK provinces/states
- Tweet data scraped via Twitter API

Analysis Methodology

Use Python to analyse data and identify patterns and trends that could help inform the UK government marketing strategy to promote the vaccine and ultimately increase the vaccination rate.

Report Contents

1 Vaccinations

1.1 Total vaccinations by province/state

Table 1 lists the total number of vaccinations by UK Province/State. Findings are organised into Vaccinated (those receiving both vaccine doses) and those who are partially vaccinated (only receiving first vaccine dose).

As the UK government is interested in identifying regions with a significant population of people who are not fully vaccinated findings are order partial vaccinations in descending order.

The province/states with the largest number of people who have are partially vaccinated are Gibraltar, Montserrat, British Virgin Islands and Anguilla. All four province/states have in excess of 200,000 people who have not received a second vaccine dose.

The province/state with the least number of people partially vaccinated is Saint Helena, Ascension and Tristan da Cunha (HATDC).

In Gibraltar although almost 96% of people (5,606,041) are fully vaccinated a significant number of people remain partially vaccinated (254,745).

Table 1: *Partial and fully vaccination totals by province/state*

Province/State	Vaccinated	Partially Vaccinated	% Ratio of Interest
Gibraltar	5606041	264745	4.509532
Montserrat	5157560	243568	4.509577
British Virgin Islands	4933315	232988	4.509763
Anguilla	4709072	222398	4.509771
Isle of Man	4036345	190639	4.510048
Falkland Islands (Malvinas)	3587869	169438	4.509560
Cayman Islands	3363624	158852	4.509669
Channel Islands	3139385	148261	4.509640
Turks and Caicos Islands	2915136	137686	4.510122
Bermuda	2690908	127073	4.509363
Others	2466669	116482	4.509299
Saint Helena, Ascension and Tristan da Cunha	2242421	105889	4.509158

1.2 Vaccinations ratio of interest by province/state

In terms of the ratio of populations who are partially / fully vaccinated, Turks and Caicos Islands have the highest percentage of individuals who have received a first dose but not a second dose. Saint Helena, Ascension and Tristan da Cunha (HATC) have the smallest percentage of people who have only received their first dose.

Figure 1: *Vaccinated and partially vaccinated by province/state*

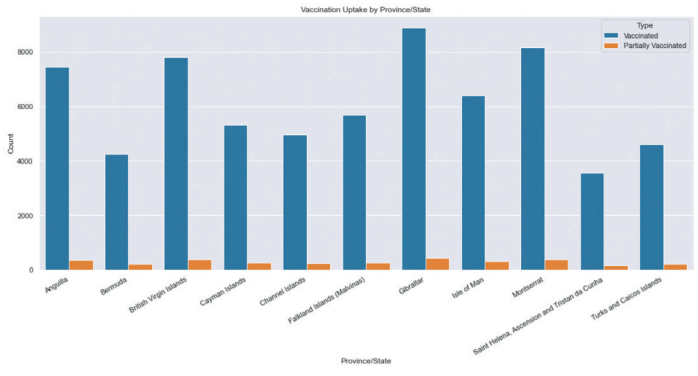


Figure 1 illustrates this ratio and shows:

- in terms of actual numbers, there is little difference across province/state in the volume of people who have not yet received a second dose.
- there does not appear to be a particular province/state that has a significantly higher or lower full vaccination engagement rate

1.3 Vaccinations over time (Gibraltar sample)

As Gibraltar offer the largest population that has not been fully vaccinated, further analysis used Gibraltar as a sample population to examine how their vaccination engagement changed over time. Table 2 details vaccination engagement findings by month.

Table 2: *Gibraltar Partial vaccination engagement by month*

Month	Vaccinated	Partially Vaccinated	% Ratio of Interest	First Dose	Second Dose
2020-02-01	0	0	0.000000	0	0
2020-03-01	0	0	0.000000	0	0
2020-04-01	0	0	0.000000	0	0
2020-05-01	0	0	0.000000	0	0
2020-06-01	0	0	0.000000	0	0
2020-07-01	0	0	0.000000	0	0
2020-08-01	0	0	0.000000	0	0
2020-09-01	0	0	0.000000	0	0
2020-10-01	0	0	0.000000	0	0
2020-11-01	0	0	0.000000	0	0
2020-12-01	0	0	0.000000	0	0
2021-01-01	0	0	0.000000	0	0
2021-02-01	13174	906844	98.568071	920018	13174
2021-03-01	43413	1310574	96.793692	1353987	43413
2021-04-01	514345	840610	62.039699	1354955	514345
2021-05-01	1296344	-897952	-225.394084	398392	1296344
2021-06-01	1342955	-707923	-111.478319	635032	1342955
2021-07-01	895996	-217516	-32.059309	678480	895996
2021-08-01	649870	-421858	-185.015701	228012	649870
2021-09-01	572892	-411096	-254.082919	161796	572892
2021-10-01	239096	-142890	-148.525040	96206	239096
2021-11-01	37956	5952	13.555616	43908	37956

Some monthly findings show negative values for those partially vaccinated. These negative values are a result of the months when more people are recorded as receiving a second vaccination than those receiving their first vaccination.

Figure 2: *Gibraltar first and second dose vaccination engagement by month*

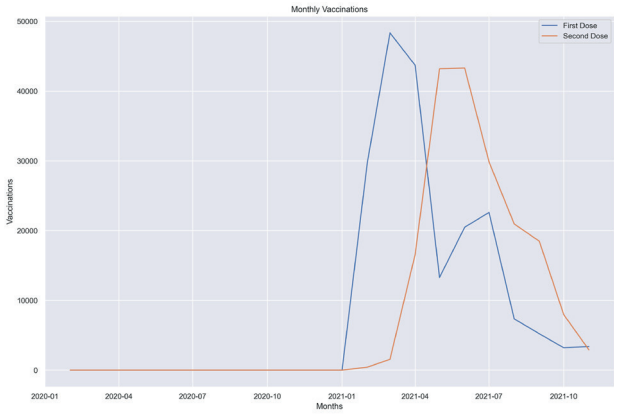


Figure 2 shows that first dose vaccination uptake begins in January 2021 and uptake increases rapidly over the first 3 months. As expected and as a result of the vaccination programme 2 month gap between vaccine first and second dose provision, uptake of the second dose begins 2 to 3 months after initial engagement with the first dose.

For Gibraltar vaccinations appear to have already peaked; first dose peak in April 2021 and second dose peak in June 2021.

Vaccine engagement decreases rapidly after peaks have been reached. First dose engagement does experience a second wave of engagement whereas second dose engagement does not.

Figure 3 shows vaccination engagement by months for all province/states. It can be observed that this follows a similar rise, peak and decline as that of Gibraltar.

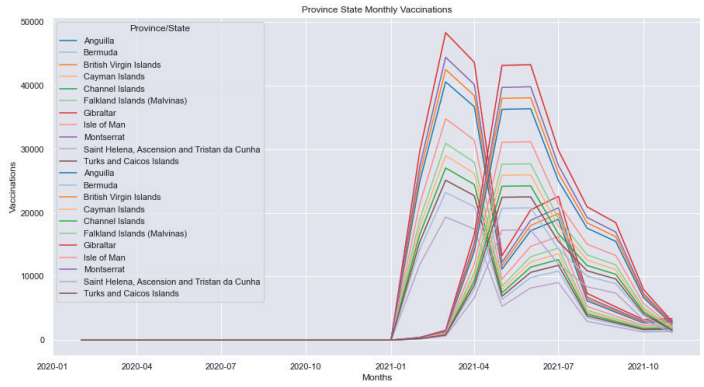


Figure 3: *First and second dose vaccination engagement by month for all province/states*

2 Deaths

2.1 Province/state deaths

Figure 4 shows the different levels of death recoded in the province/states. The Channel Islands and Gibraltar both record the highest levels of covid deaths, although high levels of deaths are recorded approximately 9 months earlier within the Channel Islands than they are for Gibraltar.

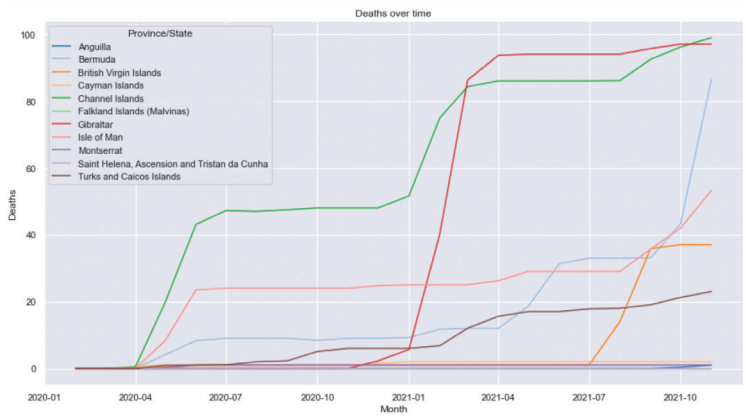
The Isle of Man, Bermuda and to a lesser degree, Turks and Caicos Islands, all show a lower and less rapid rate of death. However, this trend changes quite dramatically in the last months of recorded data for Bermuda where death rates can be seen to rise rapidly.

A similar pattern can be observed for the British Virgin Islands. Here deaths remain very low throughout until July 2021 when a rapid rise in deaths begins to occur.

All other province/states experience consistent and relative low levels of death: Anguilla, Cayman Islands, Falkland Islands, Montserrat, Saint Helena, Ascension and Tristan da Cunha.

Death rates for all province/states apart from Anguilla, Cayman Islands, Falkland Islands, Montserrat, Saint Helena, Ascension and Tristan da Cunha, do not appear to have reached their peak.

Figure 4: *Monthly deaths by province/state*



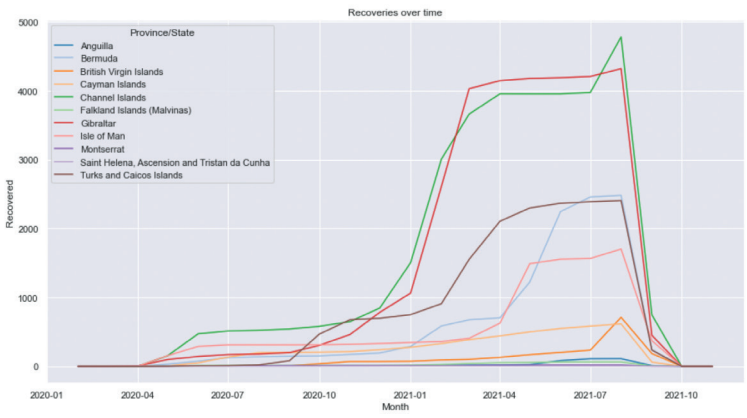
3 Recoveries

3.1 Province/state recoveries

All provinces/states appear to follow a similar trend in terms of the growth, peak and decline of recoveries.

Figure 5 shows that higher rates of recovery are generally not observed until January 2021 and that most provinces/states experience a sustained peak of recovery levels. Province/states then show a rapid decrease in recovery rates at the end of summer 2021. The greatest drop in recovery rates can be observed within the Channel Islands and Gibraltar.

Figure 5: *Monthly recoveries by province/state*



3.2 Province/state recovery and death

It is interesting that when viewed together, monthly death and recovery findings appear to correlate. At the end of the data collection period, deaths are rising for the same province/states that show a sharp decrease in recoveries during this time period. In addition, vaccination engagement is also very low.

4 Hospitalisations

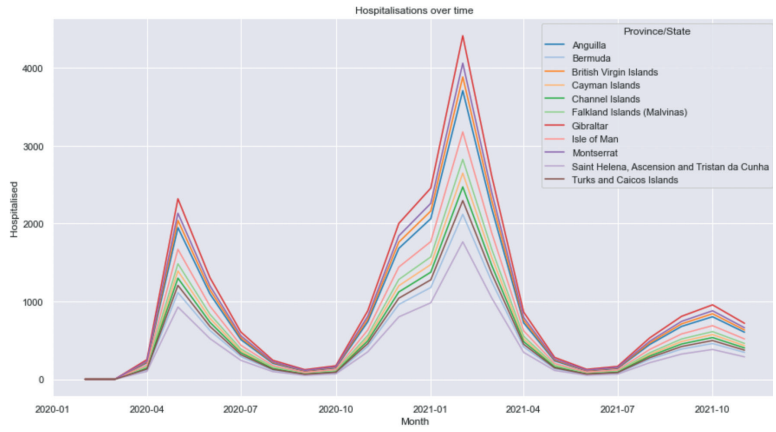
4.1 Province/state hospitalisations

Figure 6 shows three distinct hospitalisation peaks occurring in all provinces/states. The second peak is the highest and occurs in February / March 2021. The first wave is not as high and occurs at the beginning of the covid pandemic during April / May 2020.

A smaller third peak can be observed in October 2021.

Gibraltar appears to have the highest levels of hospitalisations and Saint Helena the lowest.

Figure 6: *Monthly hospitalisations by province/ state*



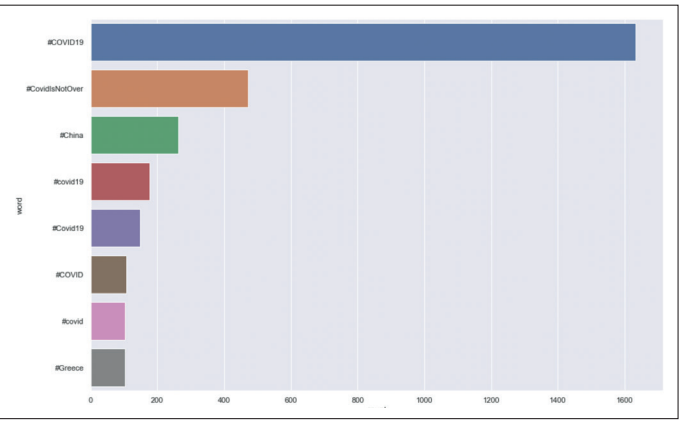
5 Qualitative Twitter data

5.1 Twitter data and tweets

In order to further understand and add context to the quantitative data-led findings, twitter data was scraped and supplied for analysis. Tweet content and data were analysed within the intention of determining the most popular hashtags and tweet words.

Analysis shows that hashtags / words containing 'covid' dominate the content of tweets posted during the date range of supplied Twitter data. This report highlights that the twitter data was sampled at a different time range to the other quantitative data sources. The UK government may wish to consider the relevance in terms of any correlations arising from this analysis.

Figure 7: *Most popular hashtags by count*



Most popular words and hashtags also contained references to China and Greece. Python analysis was able to scan the wording of these tweets to enable deeper understanding of their context and potential relevance.

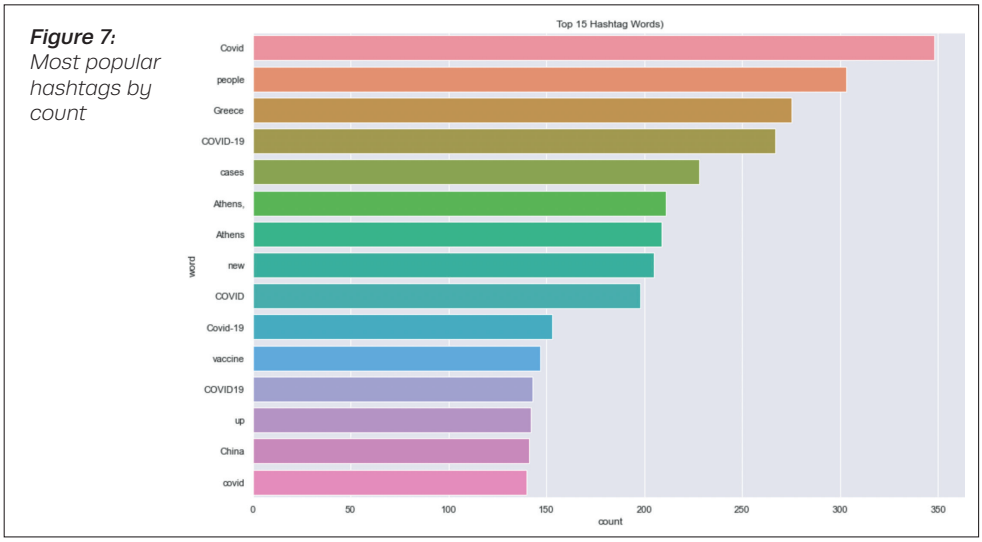


Figure 7: *Most popular hashtags by count*

Examination of tweet messages suggest a resurgence in covid cases in China is occurring. Tweets relating to Greece did not appear to be linked to the covid pandemic.

6 Channel Island sample population

6.1 Channel Island findings

The Channel Island population was requested to be used as a smaller sample to investigate moving average hospitalisations using a seven day window. Appendix 3 shows the findings and Python created visualisations that for this sample study.

6.2 Further UK government questions

The UK government has requested further information to assist their data-driven decision making process. This has taken the form of three questions. Responses are provided as follows:

Question 1: Quantitative and qualitative data

Quantitative data is numerical in format whereas qualitative data is object or text based. Qualitative data can also be referred to as categorical in type as it can contain data counts the amount of items within a category.

Quantitative data is used for statistical analysis and as such can effectively provide high level over view aggregated findings. Although numerical data can offer powerful insights, qualitative information can enrich and contextualise findings further. For example, analysis of customer social media content can reveal insights into customer engagement and buying behaviours. Such information is a very powerful tool to inform business development and sales direction.

Question 2: What is the value of continuous improvement?

The process of continuous improvement is underpinned by a supportive business culture which is built around the following 4 principles:

- customer centricity
- product quality
- process efficiency
- business profitability

Essentially continuous improvement aims to make consistent and incremental enhancements to business processes. It places the customer at the heart of all it's activities and offers products and services that consistently reflect changing customer preference and behaviour. As a result of sustained customer engagement, business profitability is also maximised.

Question 3: Are there any other data ethic requirements to be consider?

The analysis client, the UK government, adheres to all data protection requirements and has good governance in place. They also wish to know if there is anything else they need to implement from a data ethics perspective.

Data ethics frameworks are concerned with the morals and values that can guide appropriate behaviour in terms of the impact that all areas of data use can have on people and society. Ethical and regulatory practice do not always align. In order to maintain the sustainability of data sharing activities, the

In order to maintain the sustainability of data sharing activities, the UK Government Centre for Data Ethics and Innovation (CDEI) recently identified the need for personal data sharing to be 'conducted in a way that is trustworthy, aligned with society's values and people's expectations. Public consent is crucial to the long term sustainability of data sharing activity.' ('Addressing trust in public sector data use', CDEI, May 2020)

The key findings of this report also identify opportunities for the public sector to maximise the beneficial use of personal information. Among these opportunities are the introduction of accessible digital records for members of the public as well as the creation and sharing of clear guidelines that communicate expectations in terms of personal information.

This report advises that these two opportunities be explored further by the client. To enable further discussion and required opportunities, the full report can be accessed online at GOV.UK

7 Conclusion and recommendations

Overall all province/states have similar proportions of their populations requiring a second vaccination. In addition, all province/states currently have low levels of vaccine engagement. Also, recoveries and hospitalisations look to have already peaked in all province/state.

Therefore to aid recommendations as to which province/state the UK government should target in the first phase of their marketing campaign, the following can be considered.

The largest populations requiring a second vaccine are Gibraltar, Montserrat, the British Virgin Islands and Anguilla. If a decision was made purely based of volume of people, targeting these populations could leave to least volume of death as a result of vaccine uptake by the most amount of people.

However, choosing to exclude populations based on their smaller size presents significant ethical considerations. As such this report recommends that the first phase of the marketing campaign includes Bermuda, Isle of man, Cayman Islands and Turks and Caicos Islands as their populations are experiencing rising levels of death. This report also recommends the inclusion of Gibraltar based not only on populations size but also on their current high levels of death. Consideration also should be given to the Channel Islands as their levels of death are also currently high.

Arguably where funding constraints will result in the inability to target all province/states, this report advises that the Falkland Islands and St Helena along with other province/states are considered for inclusion in subsequent campaigns. Their moderate and smaller populations also appear to be experiencing lower levels of death / rates that are stable and not rising.