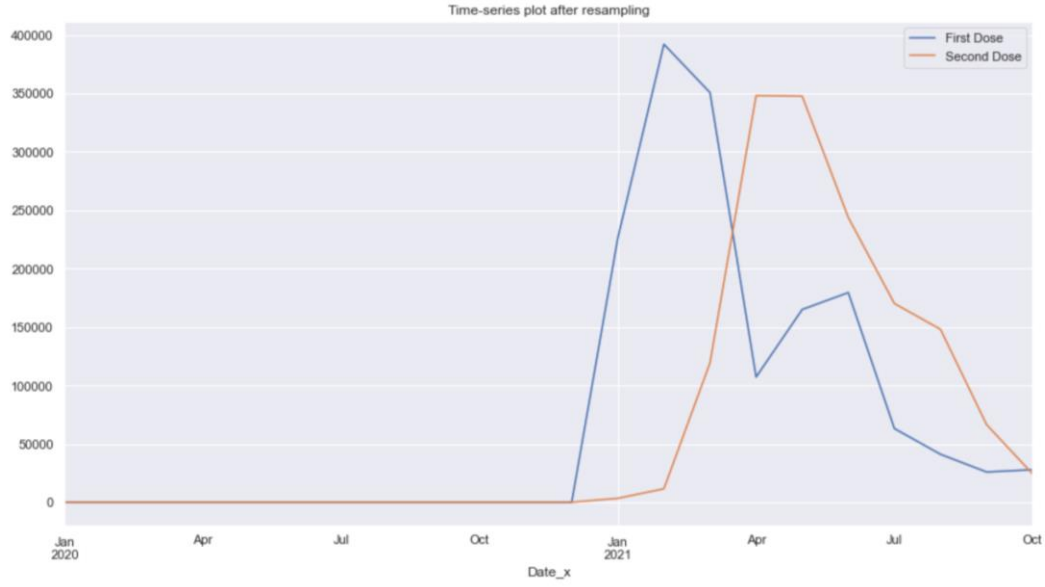
The government's objective is to promote the vaccination by setting up strong marketing strategies and respectively increase the rate of it. The main questions needed to be looked in are:

- What are the total vaccinations for a particular region? Look into the first, second, total and overtime.
- Where should they target their marketing campaigns first?
    - Areas with the biggest number of people who received the first dose only.
    - Analyse the deaths trend across all regions and confirm if a peak has been reached.
    - Explore the areas having the greatest number of recoveries so it can be avoided in the initial campaigns.
- Find the regions experiencing a peak in hospitalisation. Are there any regions that have not reached that peak yet?
- Explore the types of Twitter data points and the tweets that have #coronavirus and #vaccinated hashtags.

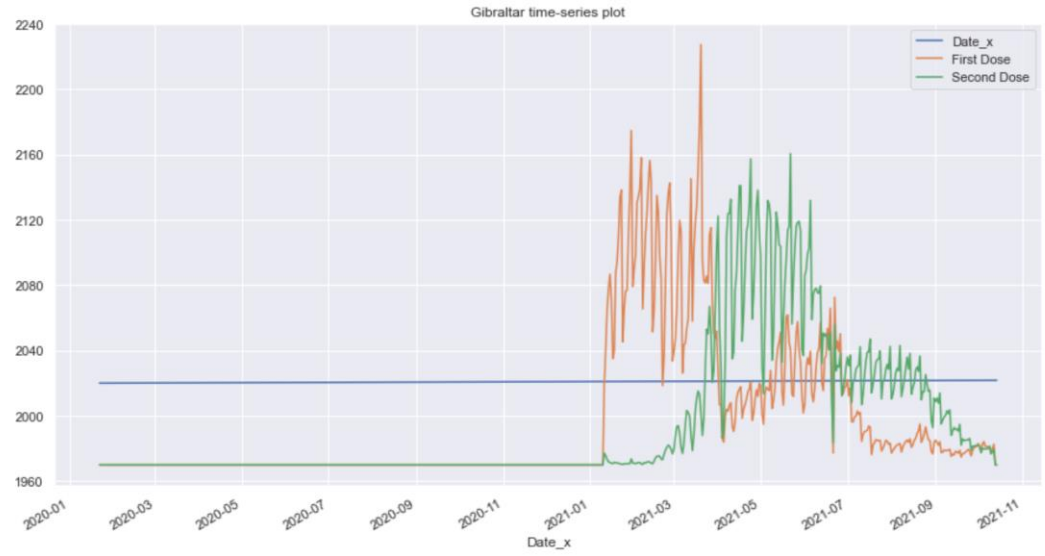
The data provided had to be analysed first so that accurate insights could be drawn. Both csv files have the same number of rows, being 7584. With regards to the number of missing values, cov_19_uk_vaccinated csv file does not have any, compare to covid_19_uk_cases, which has two rows of NaNs on dates 2020-09-21, 2020-09-22 for Bermuda state. The columns having NaN are on Deaths, Cases, Recovered, Hospitalised.

When looking at the trend for all regions, the number of first dose vaccinations starts increasing from January 2021, hitting its peak by March 2021 to 94,000 vaccines and then massively decreasing around April 2021. It then continues to gradually increase. However, the highest number reaches only 34,000 in June 2021, which, is down by triple the number of vaccines registered in March. Furthermore, they continue to steadily decrease until October 2021.

With regards to the second dose vaccinations, the numbers start increasing from March 2021 and having its peaks from May until end of June 2021. The reason the second dose are that late to first is due the wait time procedure required in between the two vaccinations in order to get the full efficiency of it. The percentage difference between the total first and second dose vaccinations is not high, being only 2% down.



Similar trend can be seen when looking at a particular state/province. On Gibraltar, second dose is surpassing the first dose numbers starting from March 30th, 2021, however, an unusual sharp drop in vaccines was noticed in April 2021, along with another decline detected in July 2021.



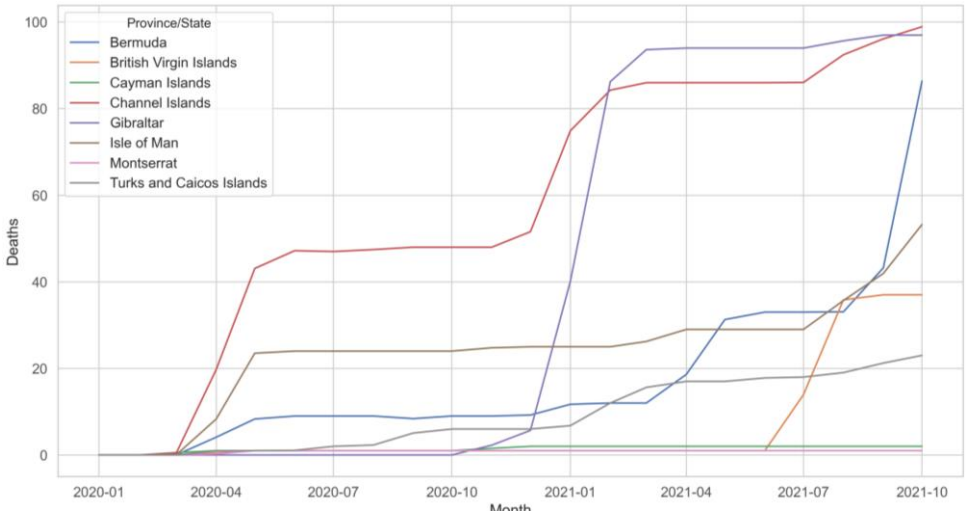
By calculating the percentage of first dose to fully vaccinated, Turks and Caicos Islands where the highest in not fully vaccinated, followed by Isle of Man and Anguilla. The ones having the highest percentage of fully vaccinated were Saint Helena, Ascension and Tristan da Cunha, Others, Bermuda, and Gibraltar.

Province/State_x	First Dose	Second Dose	num.missing.2nd_dose	%num.missing.2nd_dose
Turks and Caicos Islands	3052822	2915136	137686	0.045101
Isle of Man	4226984	4036345	190639	0.045100
Anguilla	4931470	4709072	222398	0.045098
British Virgin Islands	5166303	4933315	232988	0.045098
Cayman Islands	3522476	3363624	158852	0.045097
Channel Islands	3287646	3139385	148261	0.045096
Montserrat	5401128	5157560	243568	0.045096
Falkland Islands (Malvinas)	3757307	3587869	169438	0.045096
Gibraltar	5870786	5606041	264745	0.045095
Bermuda	2817981	2690908	127073	0.045094
Others	2583151	2466669	116482	0.045093
Saint Helena, Ascension and Tristan da Cunha	2348310	2242421	105889	0.045092

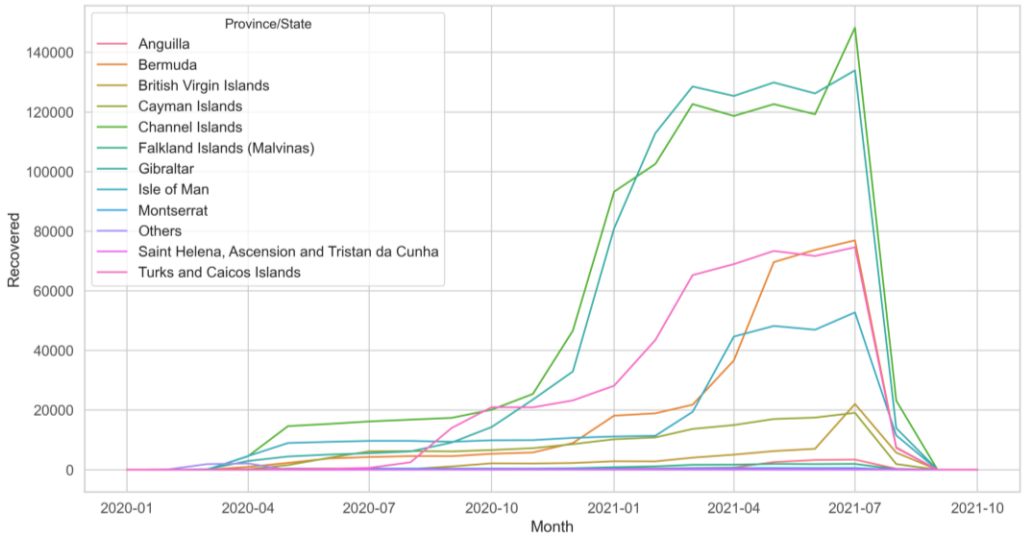
When looking at the trend of the death cases below, it can be clearly seen a rise starting from March 2020. Channel Islands has a consistent increase throughout the period hitting its highest peak in October 2021.It is important to notice that Gibraltar shows sharp elevated levels of deaths starting from December 2020 and continues so throughout.

Overall, the states and provinces have a trend of deaths increase starting from March 2021 until October 2021, except for Montserrat and Cayman Islands with a steady small number of cases during all the period provided in the data. British Virgin Islands are also experiencing small numbers of cases up until a sudden peak in July 2021.

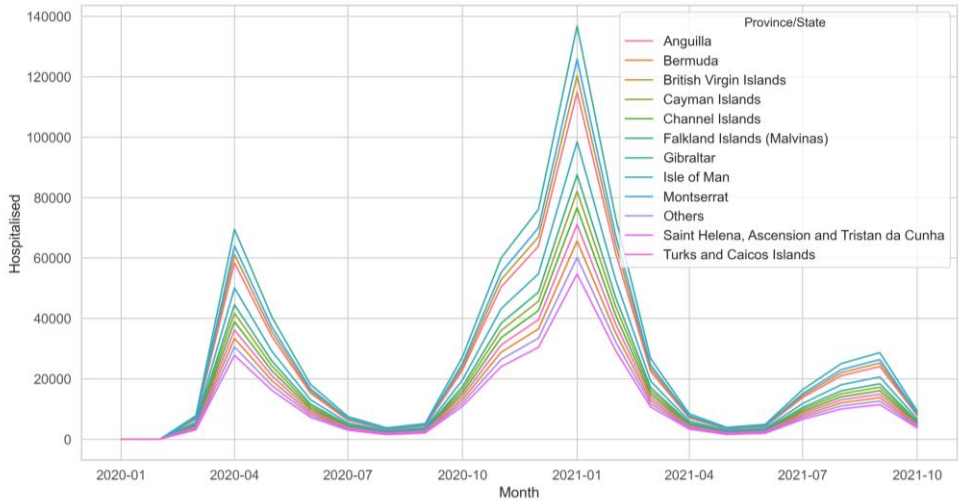
There were also groups of regions that were skewing the data having no cases registered to extremely high numbers e.g “Others.” It was crucial in stripping them out and see a clear picture of deaths across the territories.

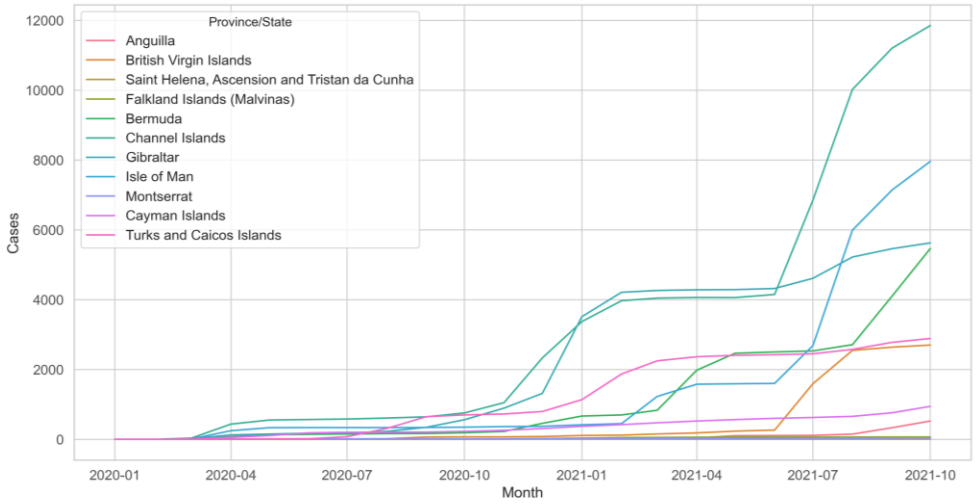


Furthermore, the region having the most recoveries are Channels Islands and Gibraltar. It has not been consistent overtime, on the contrary, it can be noticed a sharp increase starting from November 2020, having a steadiness from April 2021 until June 2021, and then followed by a sudden drop in August 2021. It is also important to note that there is a trend for all regions dropping in recoveries around August 2021 and reaching its lowest numbers by September 2021.



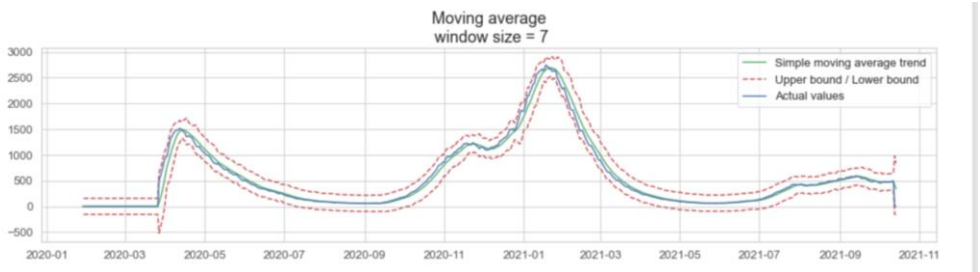
When assessing the trend between hospitalisations and cases, an insight could be brought on the discrepancy of that. There is a peak in hospitalisation across regions in January and when comparing these with the cases during the same period, the numbers are substantially low, which is wrong. It can be argued that a peak should be reached for regions as Channel Islands, Isle of Man, Gibraltar, and Bermuda according to the “Cases” visual below.



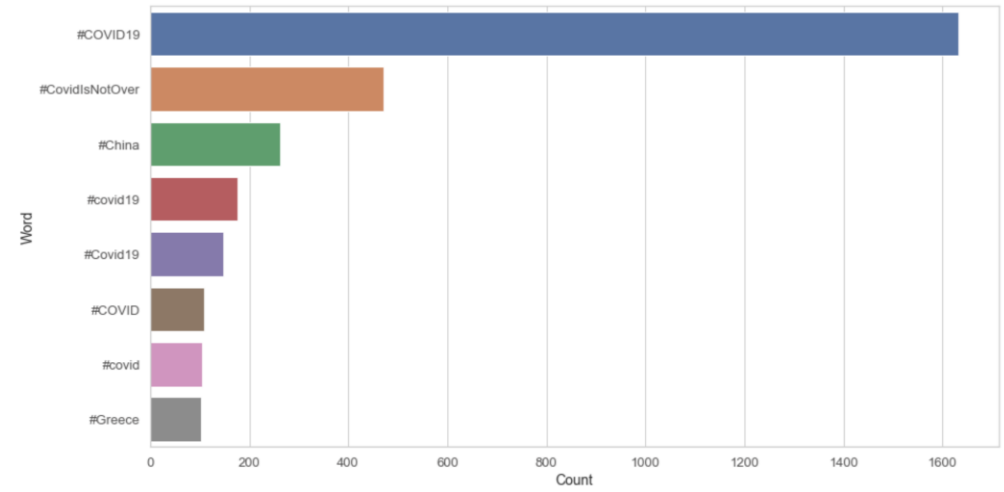


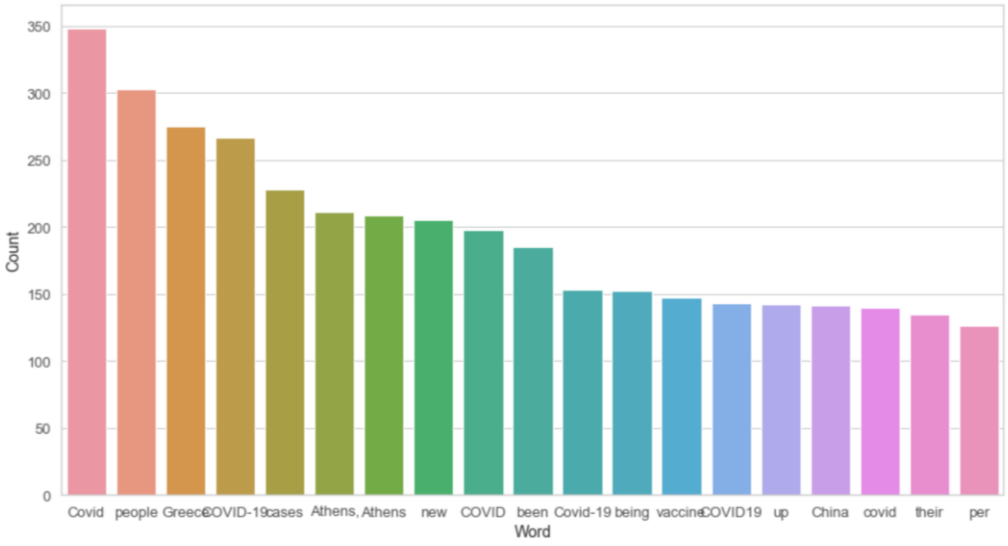
Moreover, when completing the unfinished work of the external consultant who resigned, a reason on the functions proposed by him was to:

- Show trends by using rolling average for a time series in order to smoothen the short-term variations.
- Returning the top three days with the biggest difference between daily and rolling seven days mean for the purpose of detecting any errors in values that would affect the trend cycle.



Having looked at the tweets data set, it could be investigated into additional hashtags related to Covid-19 as well as stripping out a list of stop words that would help us filter through the important content of the tweets text and bring the most used words. As a result, the top hashtags and keywords were based on Covid discussion and majority of the content was about encouraging the vaccination and side effects this virus can bring if not taking the jab.





In conclusion, the data provided was missing the number of actual population and discrepancies mentioned prior, which made the analysis more of an assumption rather than a fact.

It can be argued that for the initial marketing campaigns, based on the rate calculation of recoveries and cases below, top three regions to be excluded are Saint Helena, Ascension and Tristan da Cunha, Falkland Islands, Cayman Islands.

Province/State	Cases	Deaths	Recovered	Hospitalised	highest_recov
Saint Helena, Ascension and Tristan da Cunh	1438	4	1135	259773	79%
Falkland Islands (Malvinas)	20482	0	14754	415650	72%
Cayman Islands	217756	911	152052	389669	70%
Turks and Caicos Islands	752618	5612	515923	337710	69%
Gibraltar	1413853	25412	956103	649459	68%
Montserrat	9556	539	6376	597486	67%
Bermuda	685442	10353	363999	311547	53%
Channel Islands	1957978	37130	1027626	363690	52%
Isle of Man	887133	15051	328319	467605	37%
Anguilla	35315	24	12708	545540	36%
British Virgin Islands	284961	3573	64359	571506	23%
Others	1621650674	46987145	4115	285768	0%

A big focus should be made on promoting the vaccination continuously as the death cases show a high increase across all regions.

Also, the percentage of second dose missing values across regions is around 4,5% and without knowing the actual number of populations, a suggestion on which regions to priorities for marketing campaigns cannot be built based on that.

Moreover, the additional questions the government is interested in perfectly relates to this project.

- Due to insufficient historic data in that scenario both qualitative and quantitative forecasting methods are important. Focusing on quantitative only would have been enough if we had accurate historic data of population, cases, and hospitalisation. However, with the help of the qualitative, more relevant predictions can be made with the assistance of experts and patients in the case of Covid-19 scenario. Delphi method, Market research or Panel consensus methods would help and compile opinions to make accurate forecasting on marketing campaigns across regions.
- Another factor to consider especially when dealing with qualitative data is to have a data ethics framework. Even though the government adheres to the data protection rules, is there a formalised documentation on that? Having a strong code of ethics in place will only bring in transparency and positive social influence.
- Hence, another important strategy that will bring long term success is to always seek "Continuous Improvement." This will not only affect positively the current Covid-19 project but will impact the quality of future processes. Prioritising, sharing issues/ideas, review progress and execute the change across different departments is key.