COVID-19,
Data analysis for UK Government

INDEX

1. CONTEXT

The UK government is boosting its efforts to increase vaccination rates through marketing campaigns.

This report summarises an analysis of the following data:

- **Quantitative data,** registered daily per region from 22/1/2020 to 14/10/2021, including:
  - Number of cases, deaths, recoveries and hospitalizations.
  - Number of people vaccinated with first and second doses.

- **Qualitative data**, relating to the #coronavirus hashtag in Twitter.

2. KEY INSIGHTS

2.1.   VACCINATION

- **The number of people with only the first dose has been decreasing, with the lowest values in the last two months.**
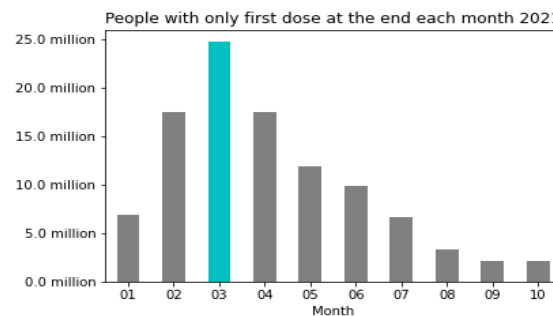

Figure 1
Inflection point in March as the second dose administration started

- **The percentage of people with only the first dose of the vaccine is below 5% in all regions, with residual variations of the percentage between regions.**

	Province/State	% Only First Dose
1	Turks and Caicos Islands	4.7231%
2	Isle of Man	4.7231%
3	Anguilla	4.7228%
4	British Virgin Islands	4.7227%
5	Cayman Islands	4.7226%
6	Channel Islands	4.7226%
7	Montserrat	4.7225%
8	Falkland Islands (Malvinas)	4.7225%
9	Gibraltar	4.7225%
10	Bermuda	4.7223%
11	Others	4.7222%
12	Saint Helena, Ascension and Tristan da Cunha	4.7221%

Figure 2

– **Gibraltar shows the highest number of people who have only had the first dose with 264,745**:
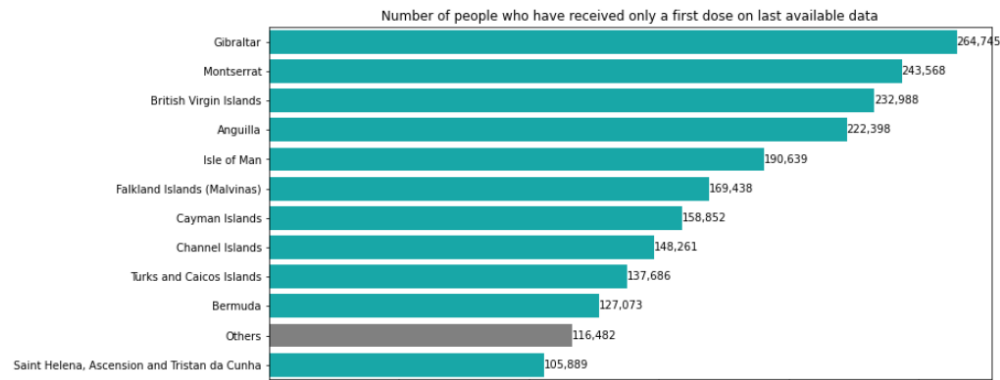

Figure 3

– **Gibraltar also presents the highest number of people fully vaccinated with almost 6 million.**
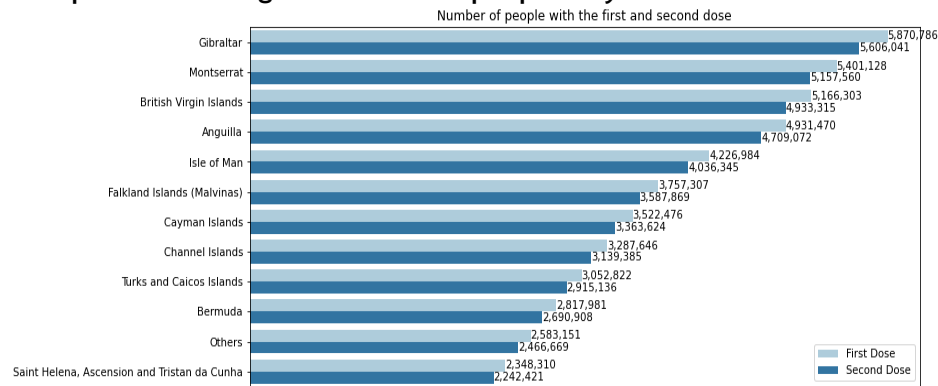

Figure 4

2.2. DEATHS, CASES AND RECOVERIES

– The table below shows the **total cases and deaths. "Others" province stands out** probably due to a higher population.

	Province/State	Cases	Deaths
1	Others	8,317,439	138,237
2	Channel Islands	12,135	100
3	Isle of Man	8,343	54
4	Gibraltar	5,727	97
5	Bermuda	5,548	95
6	Turks and Caicos Islands	2,910	23
7	British Virgin Islands	2,725	37
8	Cayman Islands	1,011	2
9	Anguilla	644	1
10	Falkland Islands (Malvinas)	69	0
11	Montserrat	41	1
12	Saint Helena, Ascension and Tristan da Cunha	4	0

Figure 5

- The table below shows the average death toll per day. **"Others" province stands out with almost 219 cases.** The remaining provinces have a residual value.

Province/State	
Others	218.729430
Channel Islands	0.158228
Gibraltar	0.153481
Bermuda	0.150316
Isle of Man	0.085443
British Virgin Islands	0.058544
Turks and Caicos Islands	0.036392
Cayman Islands	0.003165
Anguilla	0.001582
Montserrat	0.001582
Falkland Islands (Malvinas)	0.000000
Saint Helena, Ascension and Tristan da Cunha	0.000000

Figure 6

- **The death toll from March 2021 doesn't show a relevant increase, not even when there was a relevant increase in the number of cases, contrary to what happened before**. This can be explained by the vaccination campaign that started in 2021, which was especially effective from the second quarter of 2021.
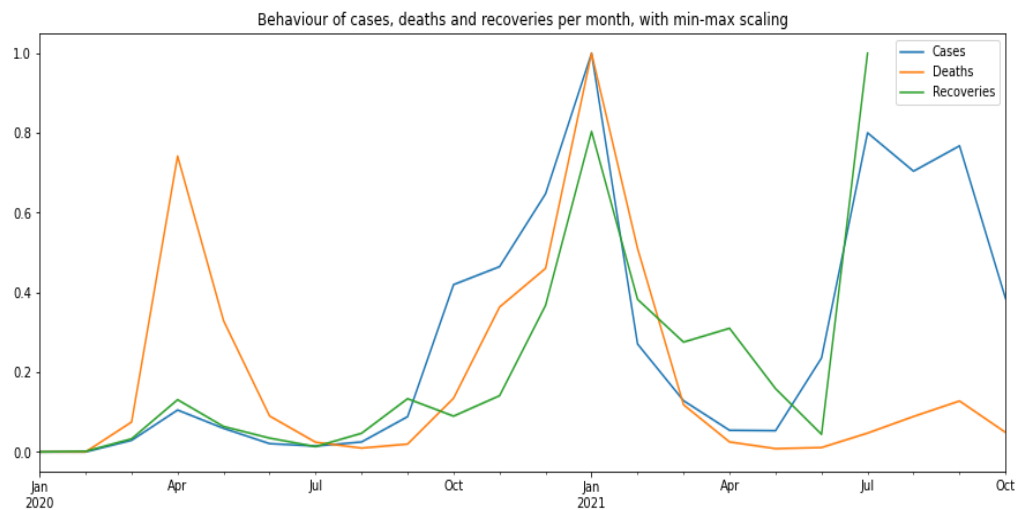

Figure 7

–   As expected, **"Others' region" stands out with a much higher death toll and a maximum value of 32,646 deaths.** Below shows the death toll monthly pattern, with **Gibraltar standing out at the end of January 2021, Bermuda and the Island of Man show an increase at the beginning of October.**
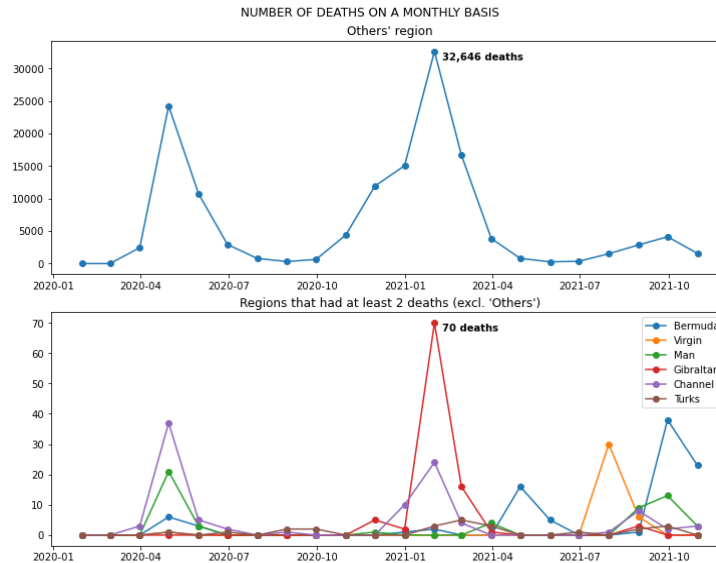

Figure 8

2.3.   RELATIONSHIP BETWEEN VACCINATION AND DEATHS/CASES/RECOVERIES

–   The heatmap below shows the relationship between variables from April to October 2021, highlighting:
   ▪   **Very high correlation between Cases/Deaths**
   ▪   **Negative correlation between Vaccines/Deaths and a positive correlation between Vaccines/Recoveries** (evidencing the positive effects of vaccination).
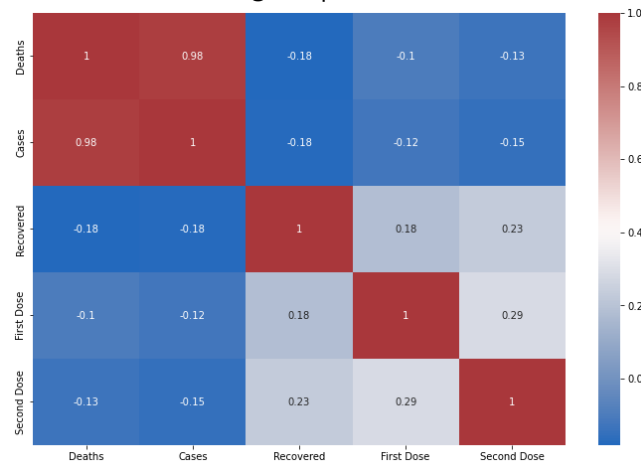

Figure 9

2.4.   HOSPITALISATION

The first figure, including boxplots per region based on **hospitalisation data up to April 2021** (previous second dose taking effect), evidenced a **right skewed distribution without outliers.** Applying the same scale on the boxplots using **data from April 2021**, shows a **decrease in hospitalisations** and a **less skewed distribution without outliers**.
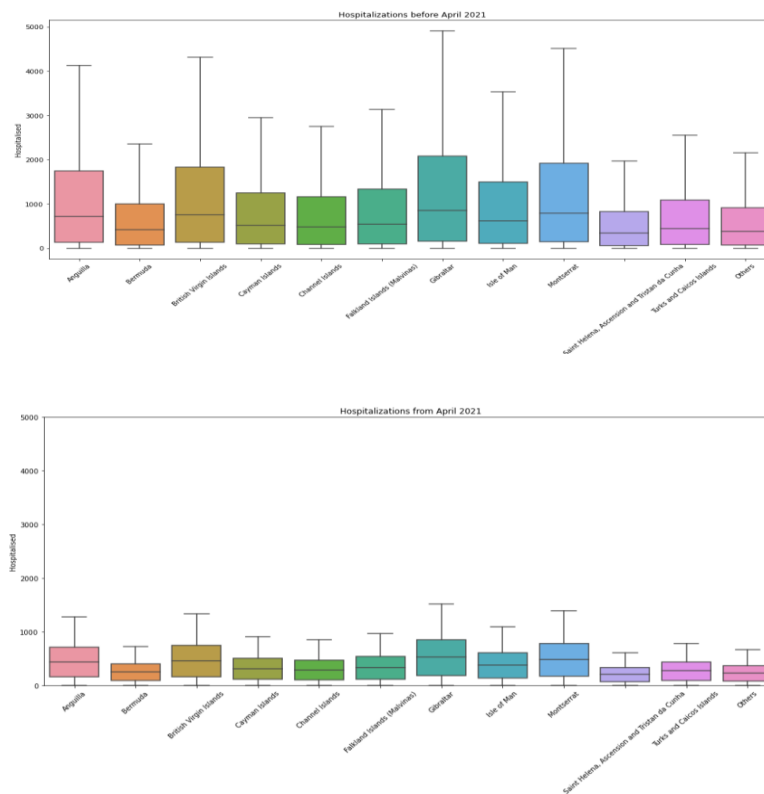

Figure 10

2.5.   TWITTER

– **Hashtags containing "covid" dominated the top5 hashtag usage**, with only one case that does not include this word.
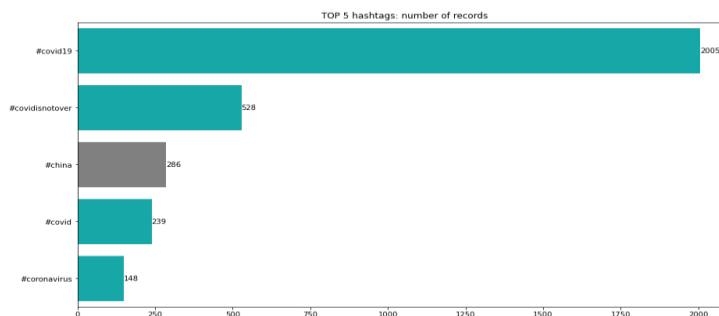

Figure 11

3. CONCLUSIONS AND RECOMMENDATIONS

Based on the data provided:

–   There is **no evidence to support prioritisation of vaccinations in regions with less fully vaccinated people** (residual differences in the percentage of people with only the first dose per region).

–   **Bermuda and British Virgin Islands** show an **increase of deaths at the end of September** that should be monitored.

However, the data provided presents the following inconsistencies:

–   "Others'" province stands out in the number of cases and deaths, but the number of people vaccinated does not follow this line where it is the second-to-last.

–   There is no information regarding recoveries from August 2021

Further analysis

–   Examine and **clarify data incoherencies to increase the assurance of the conclusion**.
–   Analysis of **population numbers per region** could improve the analysis.

4. DETAILS OF THE RATIONALE, TECHNIQUES AND ANALYSIS APPROACH

Using Python, through pandas, seaborn, matplotlib and datetime, the following procedures were done to structure, clean and analyse the data:

– **Import** the three datasets provided as DataFrames.

– **Merge the datasets** from covid_19_uk_cases and covid_19_uk_vaccinated.

– Use **if conditions** to print different answers based on the dataset.

– **First exploration of the data** through functions (dtypes, head, tail, shapes, describe) and graphs

– Check for **missing data**: changed "NaN" values to the previous values available using the forward fill function as these values are part of cumulative columns.

– **Create new columns** for further analysis, such as cumulative columns for vaccines and incremental columns for cases and deaths.

– Print datasets through tables to obtain overall values (figures 2, 5 and 6).

– As the percentage of people with only the first dose of the vaccine has residual variations between regions, the values were shown in a table revealed a repetition of similar percentages.

– **Lineplots** to understand the **trend over time** (figures 7 and 8).

– **Barplots** to compare distinct categories (figures 3, 4 and 11).

– **Smooth data** of cases/deaths/recoveries variables, **using min–max scaling** to show the three variables and observe their behaviour and correlation between all variables. Without smoothing, it wouldn't be possible as the absolute number of cases is much bigger.

– **Resample data by month** for a better understanding.

– **Create subsets by region** to compare each behaviour and plot it.

– **Annotate function** to highlight some aspects in graphs (figure 8).

– **Markers** in plots to highlight the values per month (figure 9).

– **Change the colour of a specific bar to highlight** (figure 1 and 11).

– **User-functions** and **import Nullformatter** to improve the y-axis visualisation (figure 1 and 11).

– Apply **groupby function** to analyse data monthly.

– **Heatmap** to evidence the **correlation** between variables (figure 9).

– **Filter or split the data** to evidence the impact of vaccination (figure 9 and 10).

– **Outliers checking** through **boxplots** (figure 10).

– Plot the second graph of boxplot, **changing boxplot scale** to use the same as the previous graphic to evidence the difference between the 2 periods analysed (figure 10).

– **Update strings, removing spaces and updating all words to lowercase using str.strip() and str.lower(),** to avoid repetition of a similar hashtag that has the same meaning and is only differentiated due to an upper case (example #Covid19 and #covid19).

– **Moving average** use case through hospitalisations Bermudas province data.