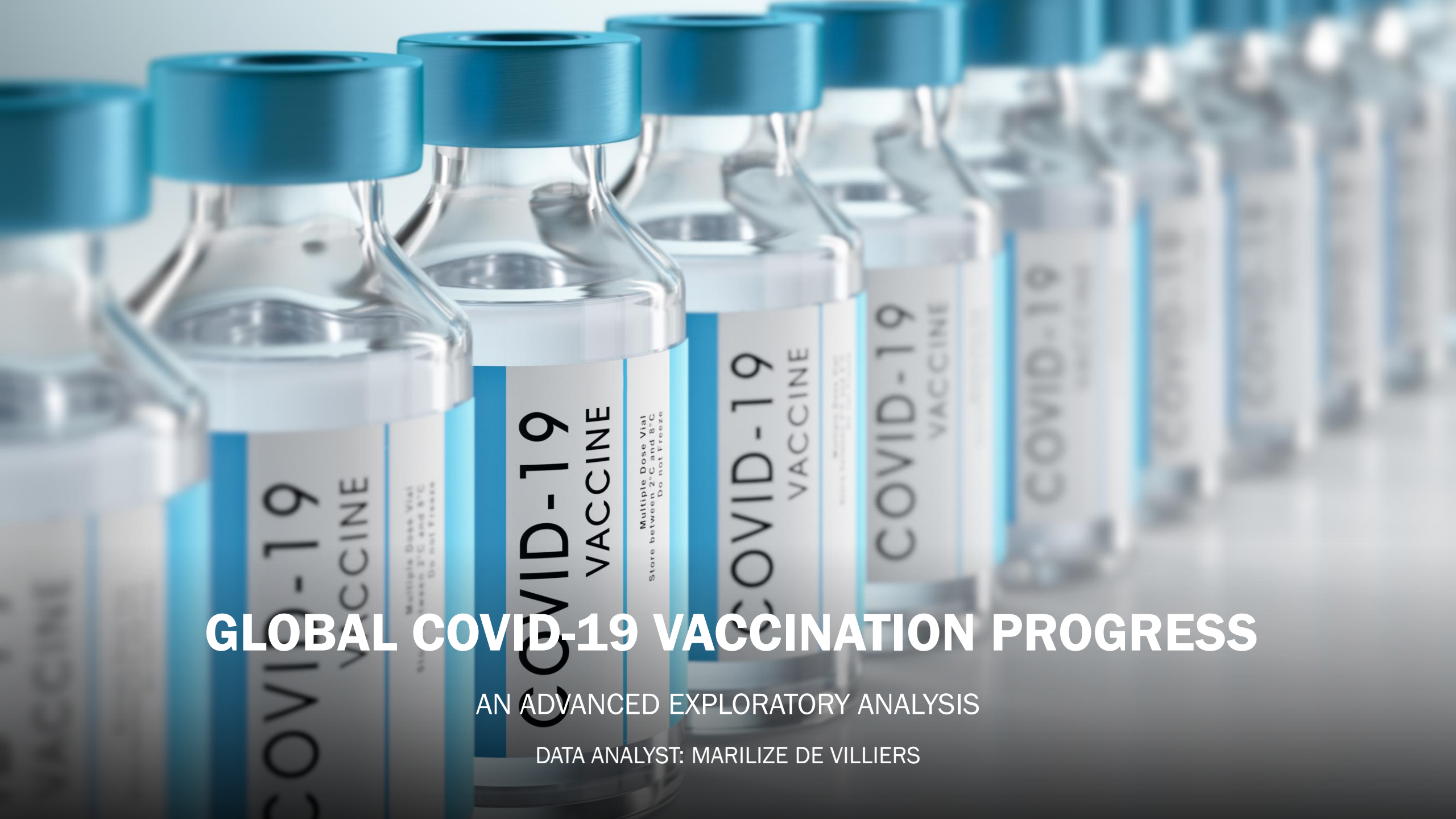


GLOBAL COVID-19 VACCINATION PROGRESS

AN ADVANCED EXPLORATORY ANALYSIS

DATA ANALYST: MARILIZE DE VILLIERS

OVERVIEW

CONTEXT

PURPOSE

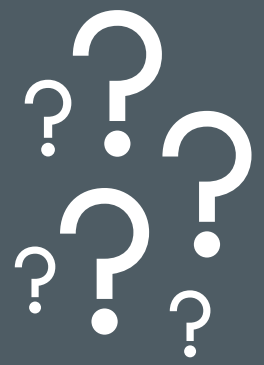
THE COVID-19 PANDEMIC HAS BEEN IN THE NEWS SINCE THE BEGINNING OF 2020, AND I HAVE BEEN INTERESTED IN HOW COUNTRIES ARE AFFECTED AND HOW THEY ARE PROGRESSING WITH THE COVID-19 VACCINATION ROLLOUTS. I AM ALSO INTERESTED IN WHICH COUNTRIES ARE USING WHICH VACCINES AND IF THE VACCINES HAVE HAD AN EFFECT ON CASES AND DEATHS.

THE DATA OF THIS PANDEMIC HAS BEEN AVAILABLE PUBLICLY SINCE IT IS A GLOBAL CONCERN.

THIS EXPLORATORY ANALYSIS IS A PERSONAL PROJECT THAT FORMS PART OF THE DATA ANALYTICS COURSE AT CAREERFOUNDRY, IN WHICH WE HAD TO FIND A TOPIC AND DATA THAT PIQUES OUR INTEREST.

OBJECTIVES & KEY QUESTIONS

Build an interactive **Tableau** dashboard that will visually showcase well-curated results of an **advanced exploratory analysis** conducted in **Python.**



1. Which countries were hit the hardest with covid-19 deaths/cases?
2. Which countries are using which vaccines?
3. Which countries are more advanced with the vaccination rollouts?
4. How has the vaccine rollouts been progressing?
5. Which countries had the first batch of vaccines?
6. Which, if any, factors have been a driving force for vaccine rollouts in each country?

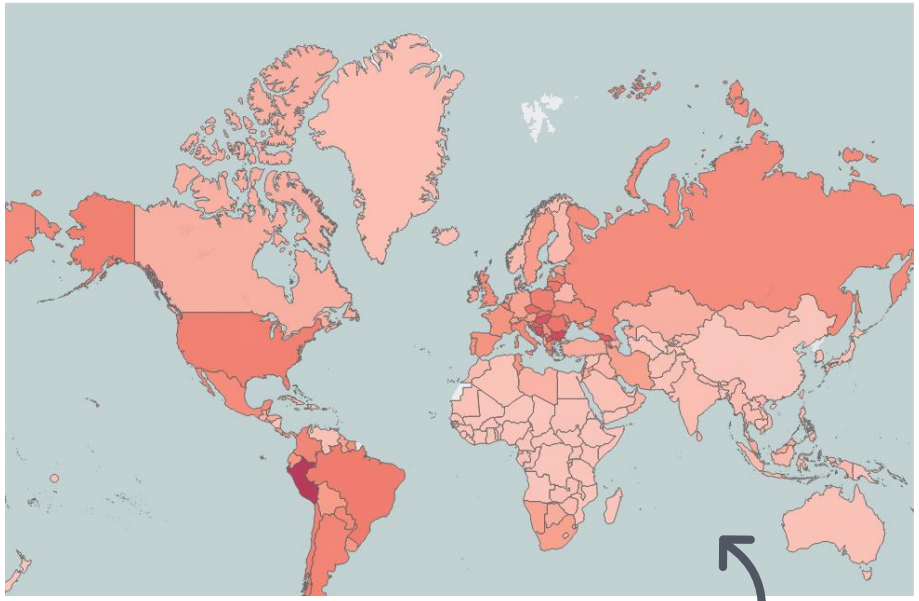
DATA SKILLS TOOLS

- Vaccination Dataset
- Population Dataset
- New Combined Dataset Time Series Data
- Geo Data (geojson)

- Sourcing open data
- Data wrangling & merging;
- Deriving variables; Grouping & Aggregating data;
- Exploratory analysis;
- Geospatial analysis using a shapefile;
- Regression analysis;
- Cluster analysis;
- Time-series analysis;
- Analysis narrative

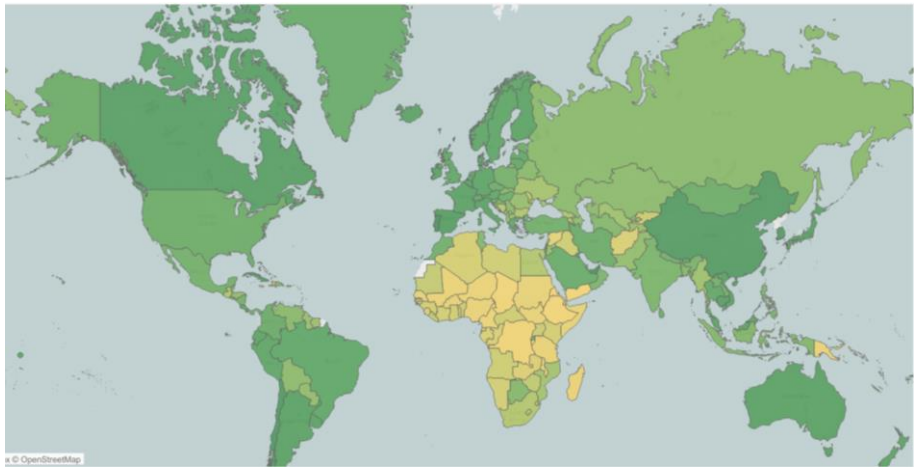
- Excel
- Python
- Tableau
- GitHub

Percentage of Deaths

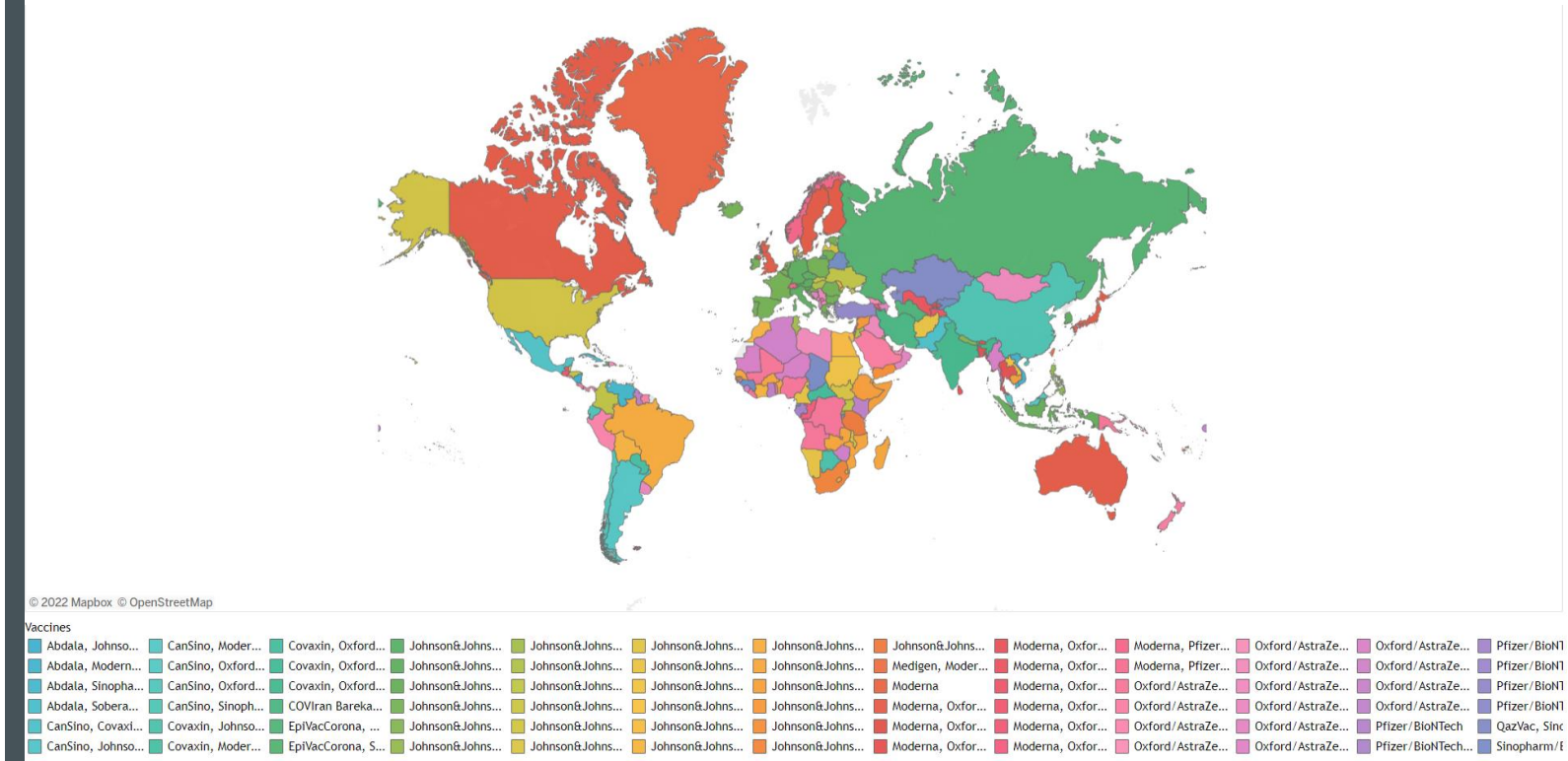


The darker the colour, the higher the number (more deaths and more vaccinations).

People Fully Vaccinated by 100



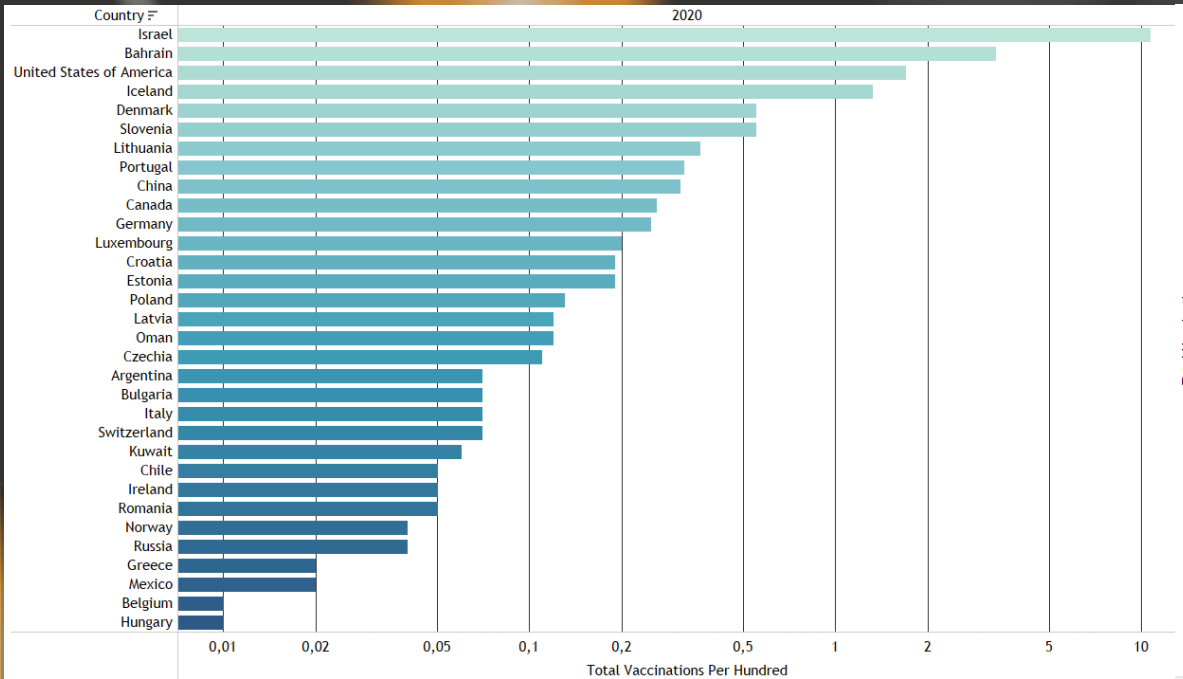
Vaccines used by Countries

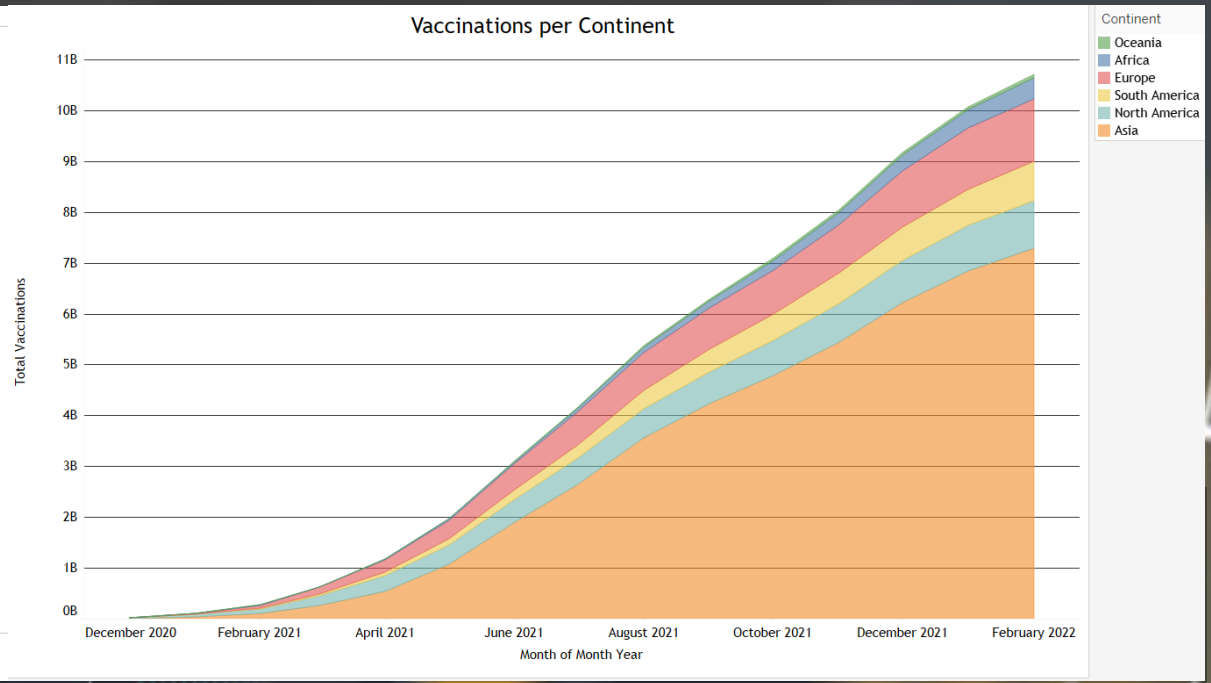


GEOSPATIAL ANALYSIS

ANALYSIS STEPS:

1. I sourced open data from Kaggle.com which led me to the GitHub profile of **Our World in Data** where I found the dataset containing all information on COVID-19 (deaths, cases, vaccinations, types of vaccines) from the beginning of the pandemic.
2. I wrangled, cleaned and merged datasets in Python.
3. Then, I did a Geospatial Analysis in Python to map which countries have the most deaths, most vaccinations, and are using which vaccines.





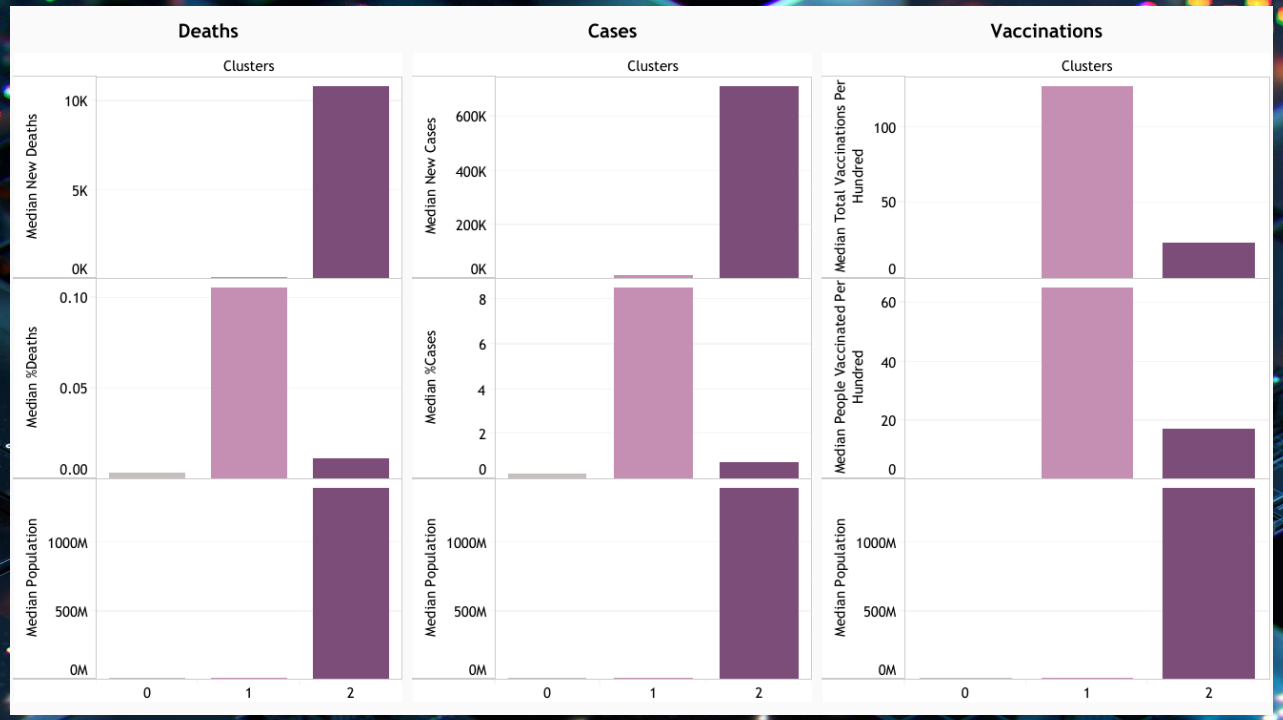
EXPLORATORY ANALYSIS

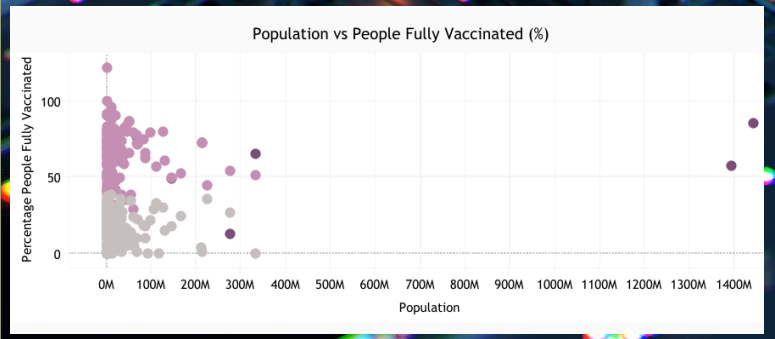
- The bar chart shows all the countries that had the first batches of vaccines (in December 2020 already).
- The line chart shows how the vaccination rollouts have progressed in each continent.

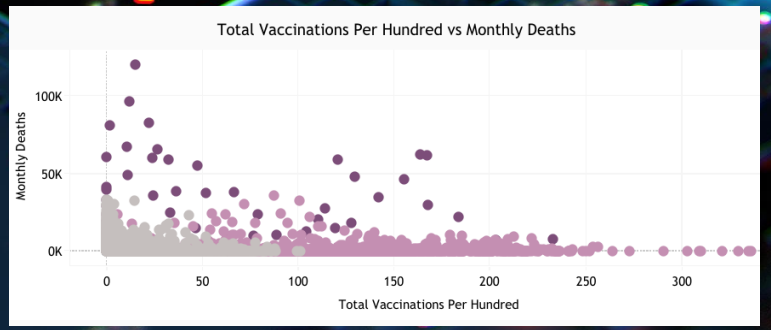
CLUSTER ANALYSIS

ANALYSIS STEPS:

1. I first used the elbow technique to determine how many clusters there are. The elbow techniques yielded three distinct groups.
2. The 3 clusters were plotted on scatterplots using k-means clustering.







INSIGHTS

PERU HAS THE HIGHEST PERCENTAGE OF DEATHS

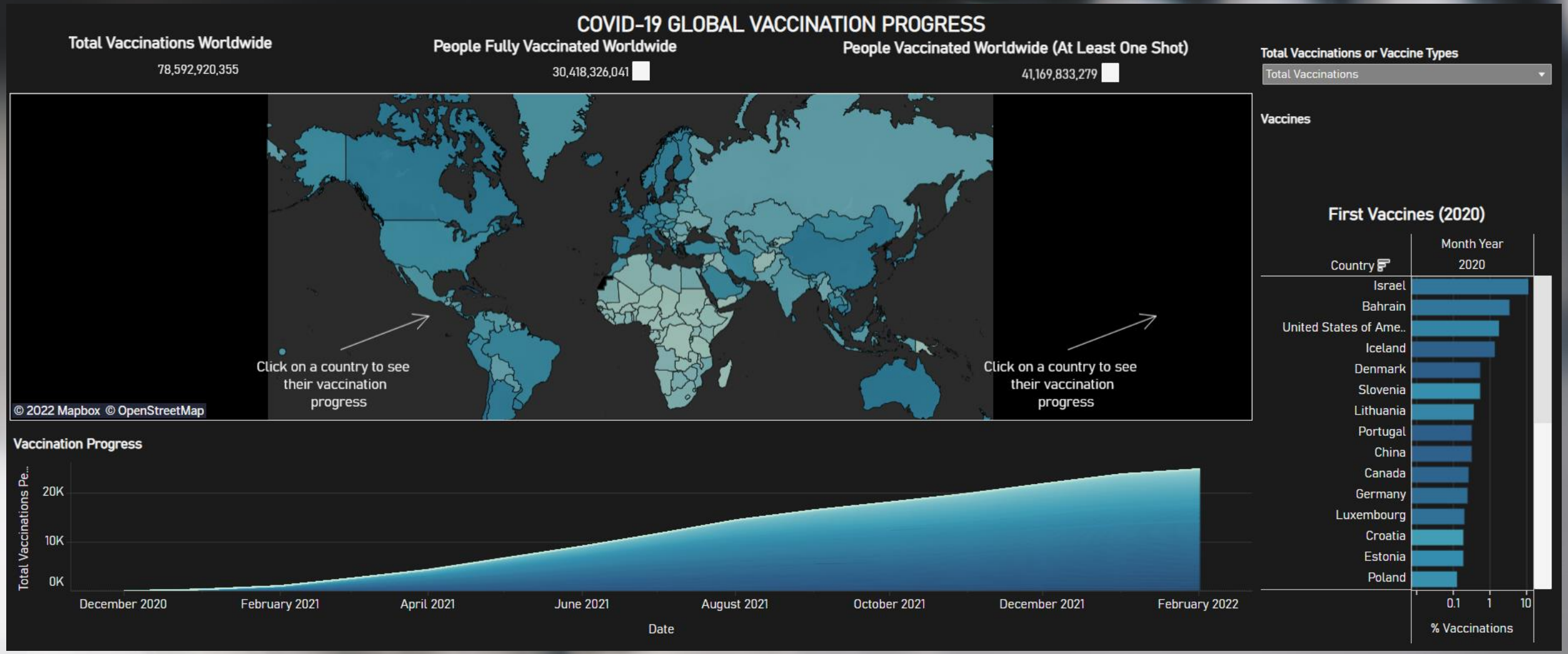
THE AFRICAN CONTINENT IS FAR BEHIND IN THE VACCINATION PROCESS

32

COUNTRIES STARTED WITH VACCINATION ROLLOUTS IN 2020

SEVERITY, POPULATION SIZE, DEVELOPMENT & EDUCATION ARE ALL DRIVING FACTORS FOR VACCINATION PROGRESS

TABLEAU DASHBOARD



CHALLENGES

SOLUTIONS

SUGGESTIONS

1. **CHALLENGE:** At some point my laptop could not handle the code I wrote, or it took extremely long to run the code.

   SOLUTION: The problem was the amount of RAM I had, so I installed an additional 8Gb RAM into my laptop.

2. **CHALLENGE:** As part of this advanced exploratory analysis, we were asked to prepare our time-series data for forecasting, and then as a bonus task run a forecasting code. My time-series data set was too small to run the forecast, since COVID-19 have only existed since 2019. I was only able to prepare the data set for a forecast.

   SOLUTION: I chose a different, unrelated data set to run the forecasting code.

WHAT WOULD I DO DIFFERENTLY?

1. I would wait a few more months or years and retry the forecasting code.
2. I would also research the countries within the 3 clusters to identify other factors that may have driven their vaccination progress, since the data set did not account for cultural, education or developmental factors.

MARILIZE DE VILLIERS
MARILIZEDV@GMAIL.COM

Feel free to click on the following icons to see my Tableau Dashboards and Storyboards, as well as my python codes on GitHub and my LinkedIn profile.

GitHub Repository

Vaccination Analysis Storyboard

Vaccination Operational Dashboard

LinkedIn Profile