**STAGE ONE TASK DATA ANALYTICS BOOTCAMP**

**CARRIED OUT BY: MOSOBALAJE ABDULMALIK AYOBAMI**

**TITLE: COVID 19 CASE STUDY**

**INTRODUCTION**

The report gives an insight to the data trends of the Covid19 virus pandemic of 2020; analysing the spread, infection and recovery rates and aim to provide statistical explanation to the pandemic. The data used in this study includes infection rates, recovery rates, and mortality rates across various regions, as well as responses during the pandemic.

**METHODOLOGY**

**Data Collection:**

Dataset was imported from the kaggle data repository. The specific web link being <https://www.kaggle.com/datasets/imdevskp/corona-virus-report/versions/166?resource=download>. The data set was annotated to be sourced from publicly available records such as that from the World Health Organisation (WHO) and the John Hopkins University, USA.

**Data Variables (Features) and Details:**

Data for 208 countries, containing the parameters of “Country/Region”, “Confirmed” “Deaths”,“Recovered”,“Active”,“New cases”, “New deaths”,“New recovered”,”Deaths / 100 Cases” , ”Recovered / 100 Cases”, “Deaths / 100 Recovered”, “Confirmed last week”, “1 week change”, “1 week % increase”, “WHO Region”.

**Limitation for data collection:**

* Data informing on lockdown rates and responses not publicly available.
* Dataset informing on vaccination and infection rates is much too large (10000+) columns to be successfully condensed and merged with the base data for regions and countries.

**Analysis Timeline and Flow of Operation**

Modules and Libraries Used: Python Pandas, Matplotlib, Seaborn, Numpy and Plotly.

Cleaning of the data: No null or blank value was found and all data columns have appropriate data types~thus the dataframe requires no extra cleaning.



***Flow of analysis***

**Key Insights and Trends:**

Top 10 countries with the highest number of cases confirmed: US, Brazil, India, UK, Russia, S. Africa, Colombia, France, Canada, Peru.

Lowest five countries in terms of number of cases confirmed: Western Sahara, Holy See, Greenland, St. Kitts and Nevis, Dominica

Countries with the highest active cases: US, Brazil, India, UK, Russia

Countries with the largest percentage one week increase: Papua new guinea, Gambia, Bahamas, Zimbabwe, Libya

Countries with the smallest percentage one week increase (lowest spread rate): Jordan (negative. Thus, decreases), Brunei, Dominica, E. Guinea, Fiji

Countries with the smallest percentage one week increase: Yemen, UK, Belgium, Italy, France, Hungary, Netherlands, Mexico, Spain, Western Sahara

Ranking of cases by WHO region (Highest to lowest): Americas, Europe, S. East Asia, Eastern Mediterranean, Africa, Western Pacific.

Countries with highest case fatality rate: Yemen, UK, Belgium, Italy, France, Hungary, Netherlands, Mexico, Spain, Western Sahara.

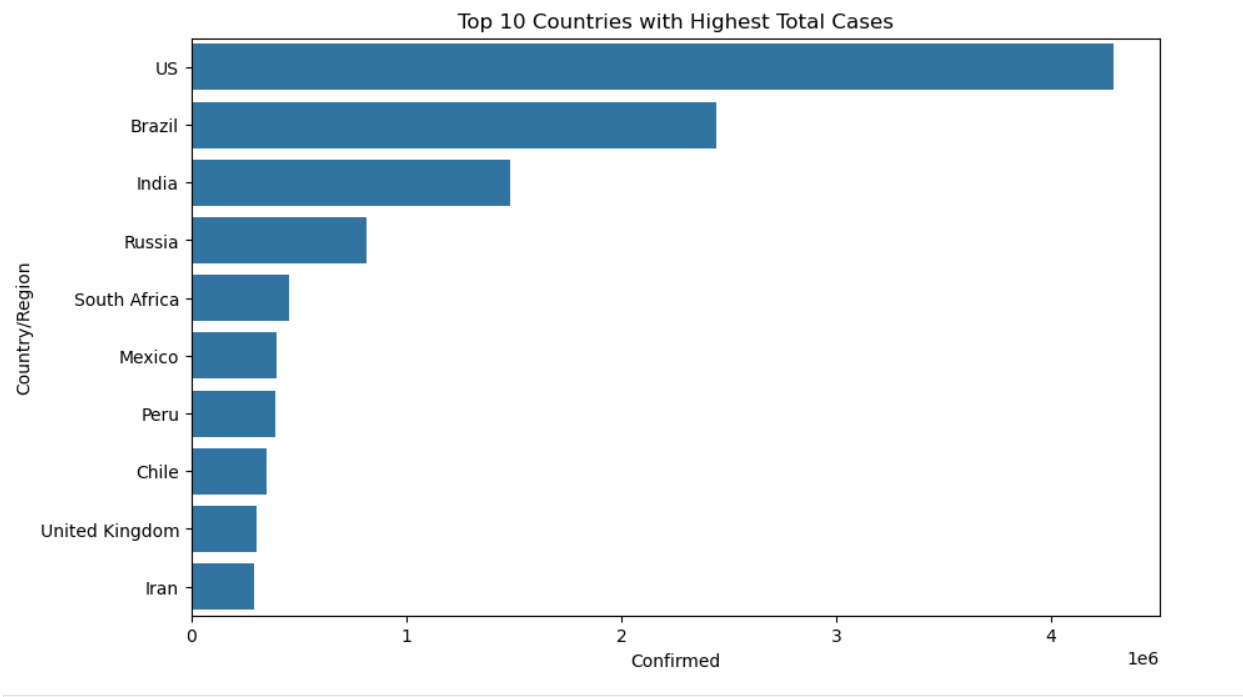
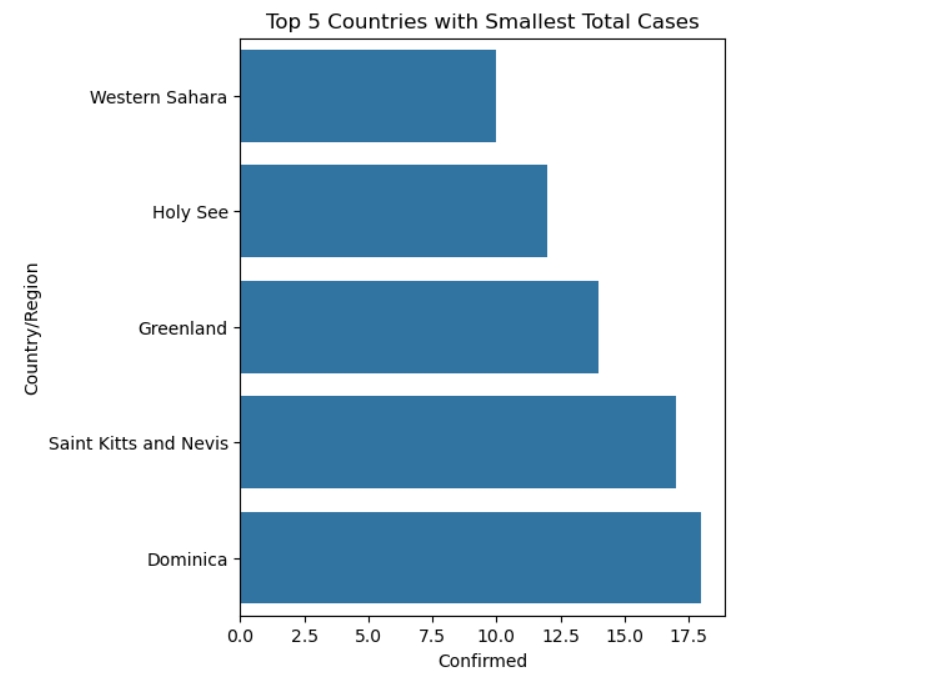
### **Conclusion**

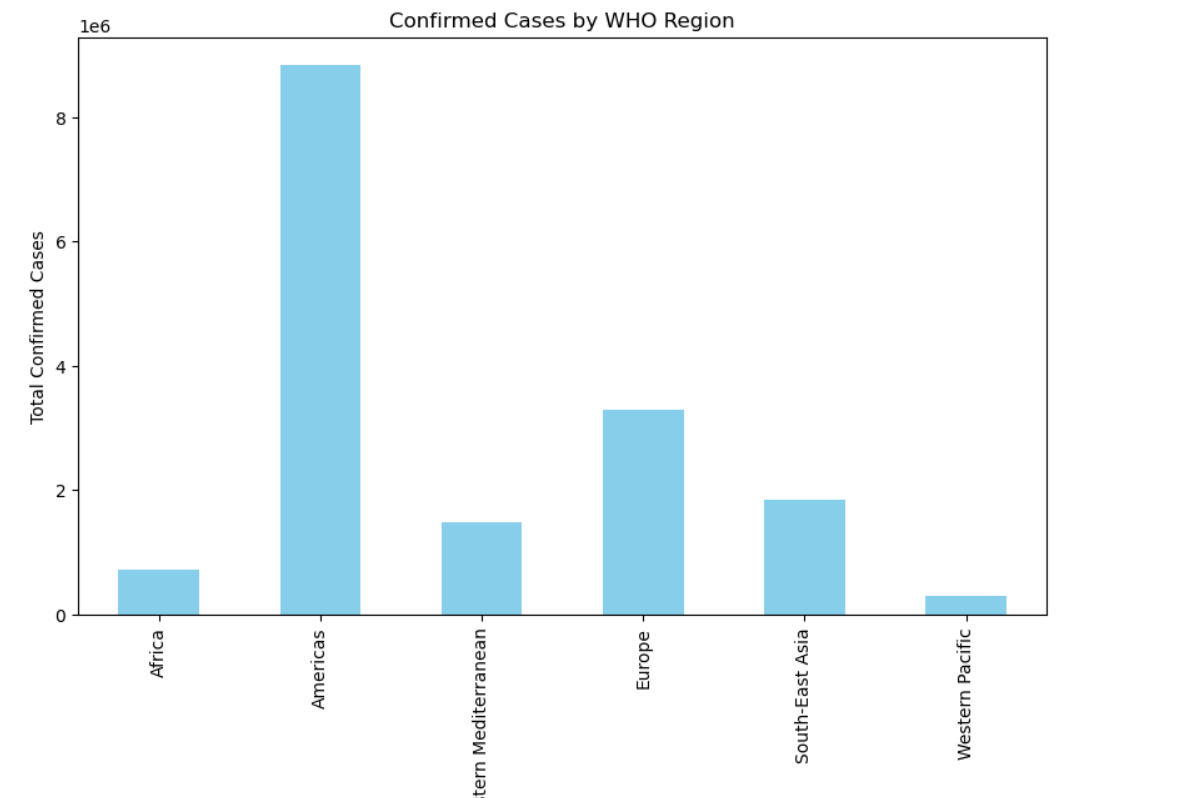
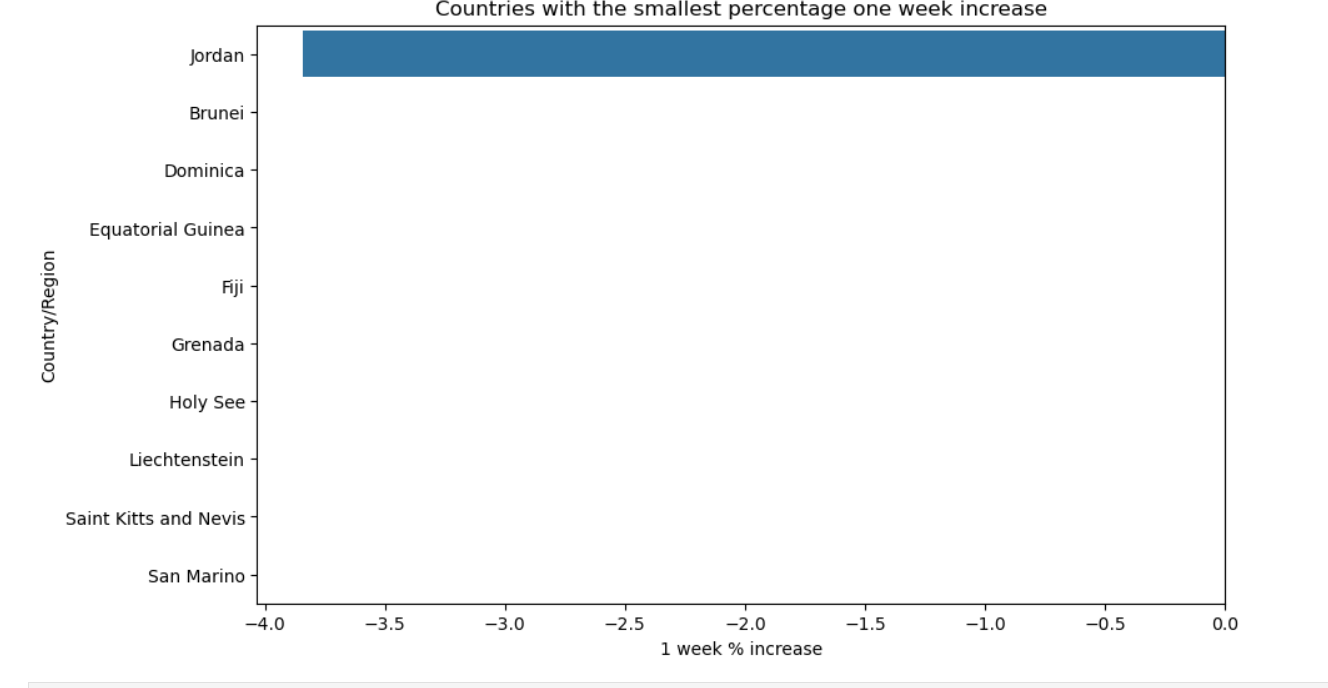
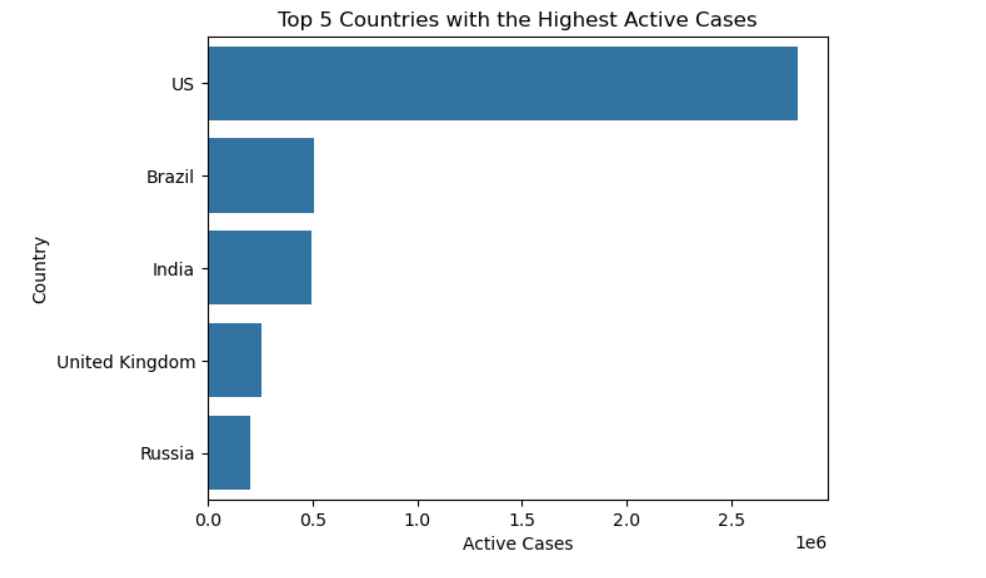
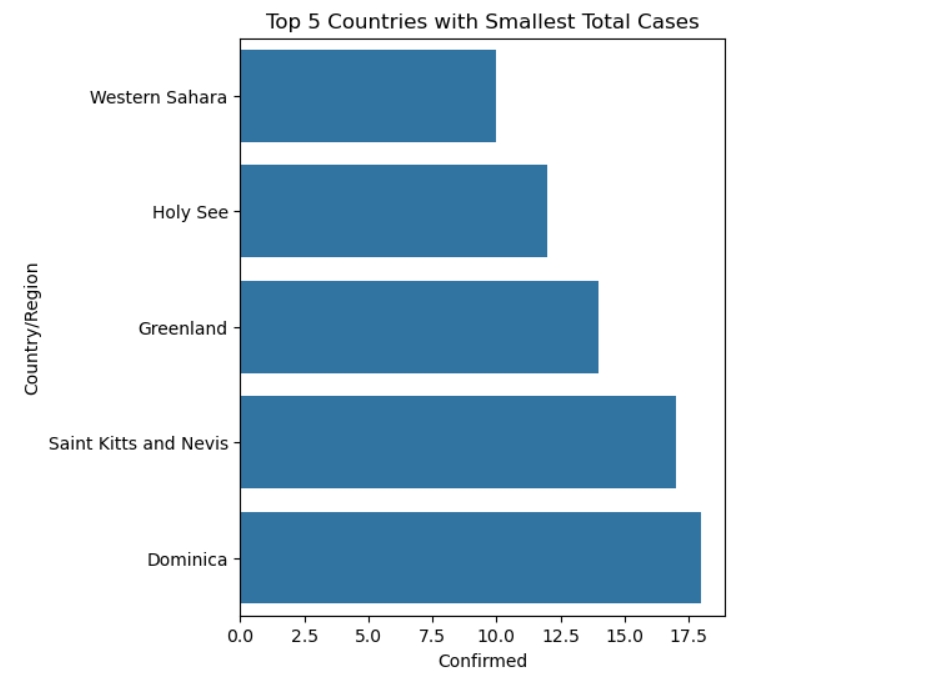
The analysis of COVID-19 data has revealed significant global disparities in the spread, severity, and control of the pandemic. Key findings include:

1. Countries with the Highest Cases: The United States, Brazil, and India have consistently reported the highest confirmed cases due to their large populations, high transmission rates, and initial delays in implementing effective containment measures. These countries also reported the highest active cases, highlighting the persistent challenge of controlling the pandemic.
2. Countries with the Lowest Cases: Small and isolated countries like Western Sahara, Holy See, and Greenland reported the lowest number of confirmed cases. This is largely attributable to geographic isolation, small populations, and strict border controls.
3. Regions with Rapid One-Week Case Increases: Papua New Guinea, Gambia, and Zimbabwe exhibited the fastest-growing case rates over one week, suggesting a potential surge in transmission due to insufficient healthcare infrastructure, delayed interventions, or limited public health measures.
4. Regions with Declining Spread Rates: Countries like Jordan and Brunei demonstrated negative or minimal case increases, indicating effective control measures, vaccination campaigns, and adherence to public health guidelines.
5. WHO Region Analysis: The Americas and Europe recorded the highest numbers of cases, reflecting their population sizes, economic activity, and initial pandemic spread. Conversely, Africa and the Western Pacific reported significantly fewer cases, likely due to differences in population density, movement patterns, and timely responses.
6. Case Fatality Rates: Yemen, UK, Belgium, and Italy exhibited the highest case fatality rates, pointing to overwhelmed healthcare systems, an aging population, and possibly delays in treatment or resource allocation during the peak of the pandemic.

### Recommendations

1. Strengthen Global Surveillance and Response:
2. Targeted Interventions for High-Risk Countries:
3. Support for Low-Income Regions:
4. Regional Collaboration:
5. Public Awareness Campaigns:
6. Focus on Long-Term Preparedness:

**DATA VISUALISATION AND GRAPHS**

****