University of Roehampton London

Data Visualisation Module (CMP020X302A) - (UG) Autumn Term 2023-2024 Coursework 1

Name:Md Atikujjaman

Student ID: ATI21508718

Task 1: PYTHON VISUALISATION

University of Roehampton London

Visualisation 1:

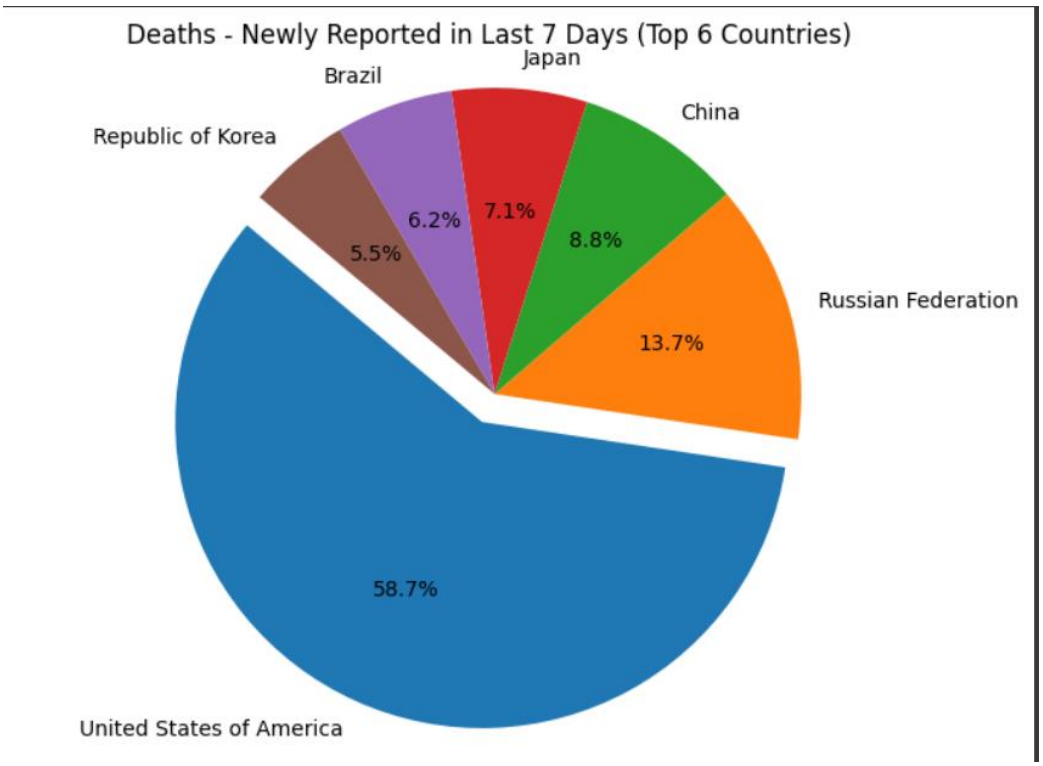
- Question 1 : What is the percentage of top six countries Deaths - newly reported in last 7 days, Visualize this question with pie chart?

Python code

```
 import pandas as pd
import matplotlib.pyplot as plt

# Read data from a CSV file
df = pd.read_csv('/content/sample_data/1. WHO-COVID-19-global-table-data.csv')

# Filter out rows with continent names and handle missing values
df = df[~df['COUNTRY'].str.contains('continent', case=False, na=False)]

# Sort the DataFrame by 'Deaths - newly reported in last 7 days' in descending order
df = df.sort_values(by='Deaths - newly reported in last 7 days', ascending=False)
custom_colors = ['#FF9999', '#66B2FF', '#99FF99', '#FFCC99', '#c2c2f0', '#ffb3e6']


# Keep the top 6 countries
df = df.head(6)

# Create a pie chart
```

```
plt.figure(figsize=(8, 8))
plt.pie(df['Deaths - newly reported in last 7 days']
, labels=df['COUNTRY'],
autopct='%1.1f%%',explode=(0.1, 0, 0, 0, 0, 0)
,startangle=140)
plt.title('Deaths - Newly Reported in Last 7 Days
(Top 6 Countries)')
plt.subplots_adjust(bottom=0.3)

plt.axis('equal')  # Equal aspect ratio ensures that
the pie chart is circular.

# Show the pie chart
plt.show()
```

SCREENSHOT



Explanation of visualization: The most recent COVID-19 death numbers for the last seven days in the top six countries are summarised in this pie chart in an easy-to-read manner. Its main objective is to give visitors a quick visual grasp of how fatalities have been distributed during this particular time period.

The graph clearly shows that the United States of America has had the greatest number of deaths reported, indicating the pandemic's major effects on the nation during this time. On the other hand, although the Republic of Korea has the smallest portion of the pie, it is evident that it has managed the COVID-19 situation relatively well over the past seven days and has the lowest recorded fatality rate.

Complex data is made easily accessible to a wide range of stakeholders, including the public and policymakers, through this visualisation. It is a useful tool that provides rapid insights into current patterns in COVID-19 mortality in these countries, supporting resource allocation decisions and directing focused actions to address the disparate effects of the pandemic in various regions.

= =

Visualisation 2:

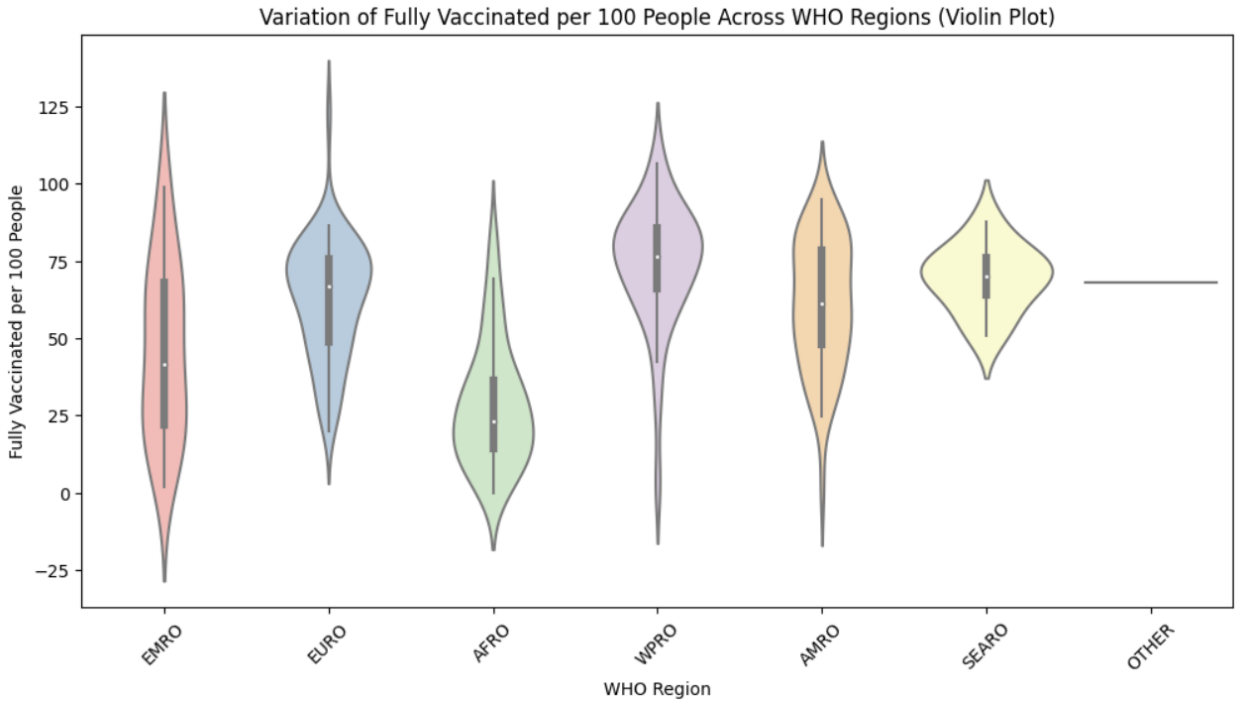
- - Question 2. How does the number of fully vaccinated individuals per 100 people vary across different WHO regions, visualized using a violin plot?

Python code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Read the data from a CSV file into a DataFrame
df = pd.read_csv('/content/sample_data/2.
Vaccination-data.csv')

# Group the data by WHO regions and calculate the
average fully vaccinated per 100 people
region_avg_fully_vaccinated =
df.groupby('WHO_REGION')['PERSONS_FULLY_VACCINATED_PE
R100'].mean().sort_values()
# Create a violin plot
plt.figure(figsize=(12, 6))
sns.violinplot(x='WHO_REGION',
y='PERSONS_FULLY_VACCINATED_PER100', data=df,
palette='Pastel1')
plt.xlabel('WHO Region')
```

```
plt.ylabel('Fully Vaccinated per 100 People')
plt.title('Variation of Fully Vaccinated per 100
People Across WHO Regions (Violin Plot)')
plt.xticks(rotation=45)
plt.show()
```

SCREENSHOT



Explation of the visualisation: This violin plot, we can see how different WHO areas differ in terms of the proportion of completely vaccinated individuals per 100 persons. The breadth of each violin indicates how frequent various vaccination rates are, and each violin represents a unique WHO region. Immunisation rates are correlated with violin width and violin narrowness. The number of individuals that are fully vaccinated out of 100 is indicated on the y-axis. In order to determine where some places are performing better or worse, this allows us to compare vaccination coverage across regions. This map, in short, provides us with an overview of the status of vaccination campaigns around the globe.

Visualisation 3:

-Question 3:In what ways does the three-dimensional scatter plot depict the correlation among the overall number of COVID-19 cases, the total number of

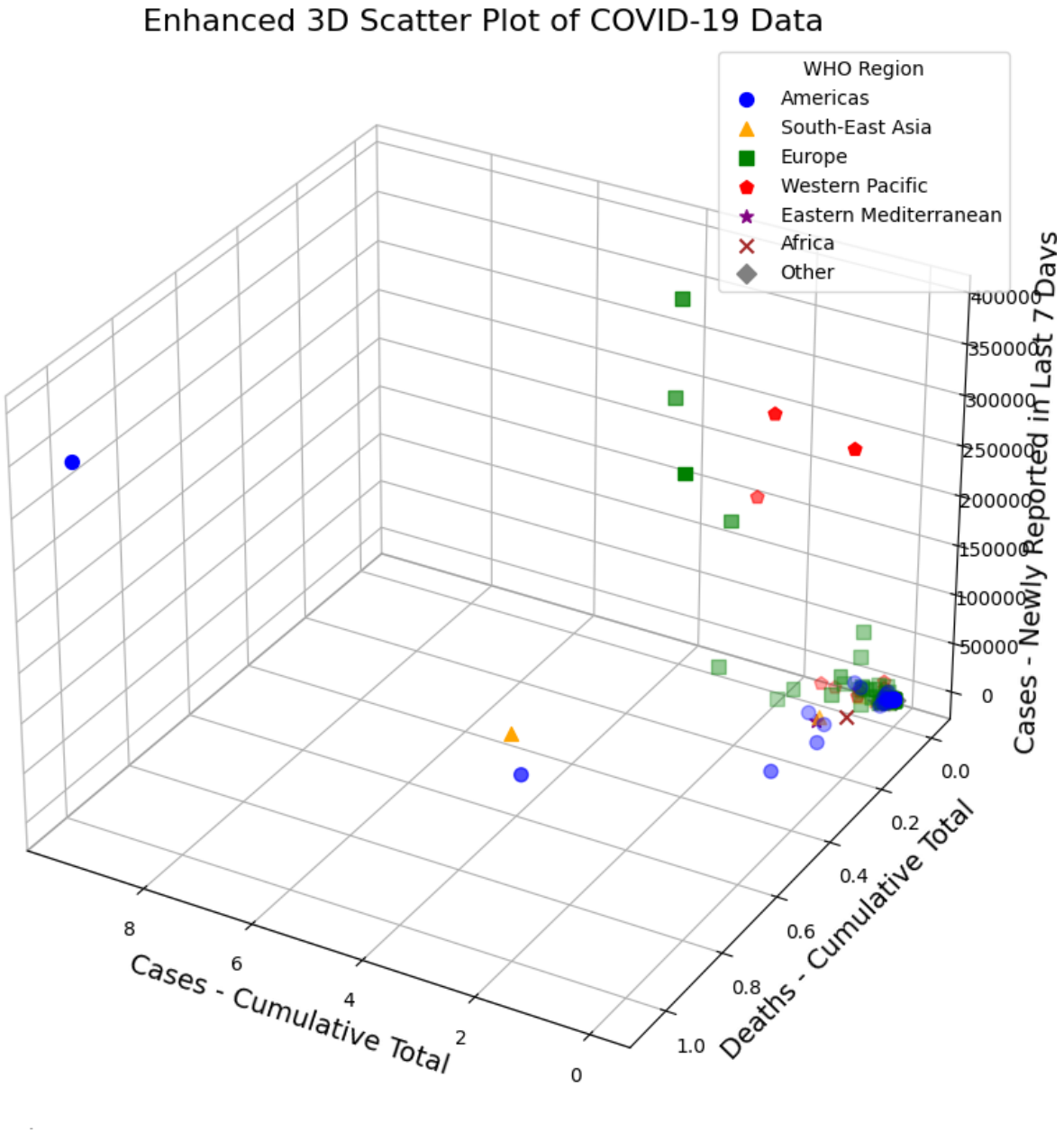
fatalities, and the quantity of newly reported cases within the last 24 hours in various WHO regions?

Python code

```
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np

# Load  dataset
df = pd.read_csv('/content/sample_data/1. WHO-COVID-19-global-table-data.csv')

cases_cumulative = df['Cases - cumulative total']
deaths_cumulative = df['Deaths - cumulative total']
cases_newly_reported = df['Cases - newly reported in last 7 days']
WHO_region = df['WHO Region']

# Define colors and markers for each WHO region
region_styles = {
    'Americas': ('blue', 'o'),
    'South-East Asia': ('orange', '^'),
    'Europe': ('green', 's'),
    'Western Pacific': ('red', 'p'),
    'Eastern Mediterranean': ('purple', '*'),
    'Africa': ('brown', 'x'),
    'Other': ('gray', 'D')
}

# Create a new figure for 3D plotting
fig = plt.figure(figsize=(12, 10))
ax = fig.add_subplot(111, projection='3d')

# Scatter plot with varying color and marker
for region, (color, marker) in region_styles.items():
    region_data = df[WHO_region == region]
    ax.scatter(region_data['Cases - cumulative total'],
               region_data['Deaths - cumulative total'],
```

```
                region_data['Cases - newly reported in last 7
days'],
                color=color, marker=marker, s=50, label=region)

# Create a legend
ax.legend(loc='upper right', title="WHO Region")

# Labels and title with increased font size
ax.set_xlabel('Cases - Cumulative Total', fontsize=14)
ax.set_ylabel('Deaths - Cumulative Total', fontsize=14)
ax.set_zlabel('Cases - Newly Reported in Last 7 Days',
fontsize=14)
ax.set_title('Enhanced 3D Scatter Plot of COVID-19 Data',
fontsize=16)

# Customize grid and background
ax.grid(True)
ax.xaxis.pane.fill = ax.yaxis.pane.fill = ax.zaxis.pane.fill =
False
ax.xaxis.pane.set_edgecolor('gray')
ax.yaxis.pane.set_edgecolor('gray')
ax.zaxis.pane.set_edgecolor('gray')

# Adjust initial view angle
ax.view_init(elev=30, azim=120)

# Show the plot
plt.show()
```

SCREENSHOT



Explation of the visualisation: A 3D scatter plot of COVID-19 data broken down by WHO area is seen in the image. Three axes make up the plot:

The total number of COVID-19 instances that have accumulated over time in each location appears to be shown by the x-axis with the caption "Cases - Cumulative Total".

The total number of fatalities in each region that have been linked to COVID-19 is probably shown on the y-axis with the label "Deaths - Cumulative Total".

The number of new COVID-19 cases reported in the previous week is displayed on the z-axis, which reads "Cases - Newly Reported in Last 7 Days."

Various WHO areas are represented by different markers and colours:

Americas are represented as blue circles, Triangles of orange colour for Southeast Asia, Europe's green squares, The Western Pacific represents red pentagons ,For the Eastern Mediterranean, purple stars, Africa's brown crosses, Grey diamonds in other, unspecified areas..

In order to facilitate data comprehension, the plot is improved with a clear backdrop and grid. It also has a legend in the top right corner that links the colours and markers to the appropriate WHO areas. Based on data from the WHO, the general goal of this visualisation is to present a comparative picture of the impact of COVID-19 in various regions of the world.

Visualisation 4:

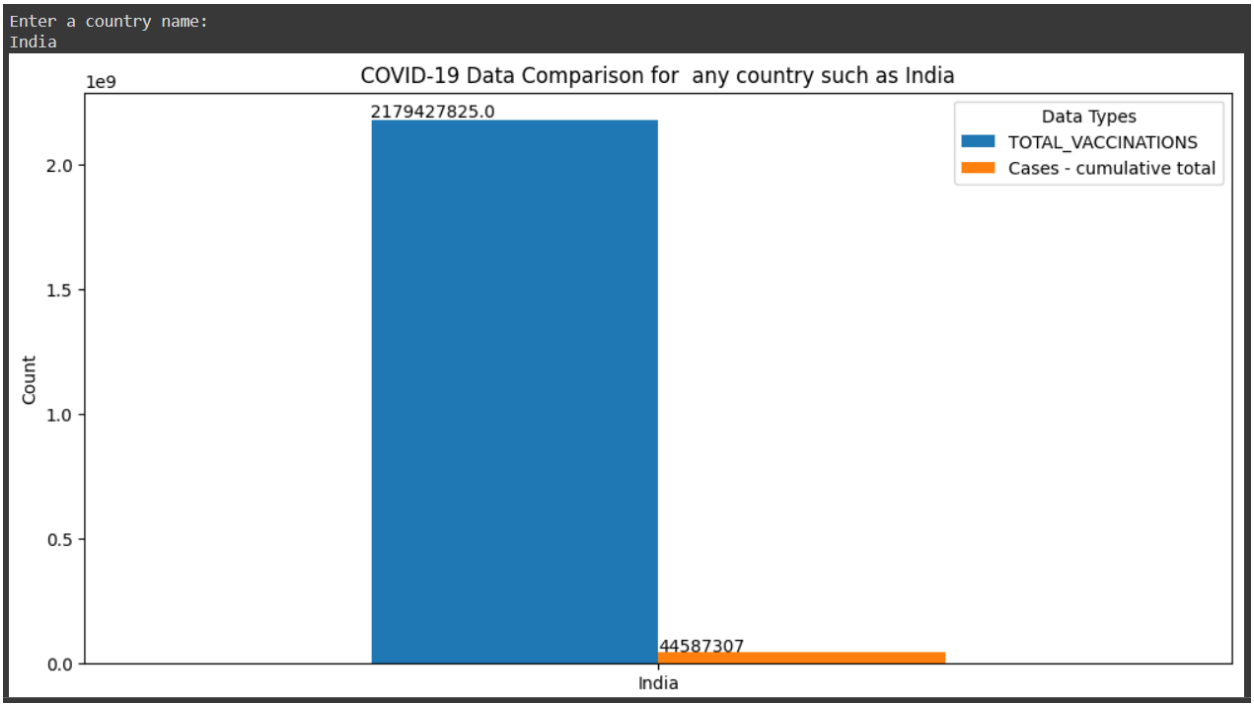
- Question 4: How does comparing a country's total COVID-19 vaccinations with its total cumulative COVID-19 cases reveal insights, visualized using a bar chart and user need to input country name?

Python code

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the COVID-19 data into a dataframe
covid_data = pd.read_csv('/content/sample_data/1.
WHO-COVID-19-global-table-data.csv')
vaccination_data =
pd.read_csv('/content/sample_data/2. Vaccination-
data.csv')
merged_data = pd.merge(covid_data, vaccination_data)

# Take user input for the country name
# Display a message without a box and get user input
print("Enter a country name:")
country_name = input().strip()

# Filter data for the specified country and the
specific columns
```

```
filtered_data = merged_data[merged_data['COUNTRY'] ==
country_name][['TOTAL_VACCINATIONS',
                                    'Cases - cumulative
total']]

if not filtered_data.empty:
    # Create a bar chart to compare the two
categories for the specified country
    plt.figure(figsize=(12, 6))
    ax =
filtered_data.set_index(merged_data['COUNTRY'][filter
ed_data.index]).plot(kind='bar', ax=plt.gca())
    plt.xlabel('Category')
    plt.ylabel('Count')
    plt.title(f'COVID-19 Data Comparison for
{country_name}')
    plt.xticks(rotation=0)
    plt.legend(title='Data Types')

    # Adding the text on the bars
    for p in ax.patches:
        ax.annotate(str(p.get_height()), (p.get_x() *
1.005, p.get_height() * 1.005))

    plt.show()
else:
    print(f"No data found for {country_name}. Please
check the country name or it may not have data for
the selected categories.")
```

SCREENSHOT



Explation of the visualisation: This visualization allows me to compare a country's COVID-19 cases with its vaccination rates.i add there dynamic fixturesfor user input. User can input a country's name, and it shows a bar chart with two bars. One bar represents the total COVID-19 cases, while the other shows the total number of vaccinations administered. This visual comparison helps me to understand how vaccination efforts relate to a country's COVID-19 situation.

Visualisation 5:

- Question 5 : Which are the top three countries with the highest COVID-19 cases and the highest number of fully vaccinated individuals, visualized using a swarm plot?

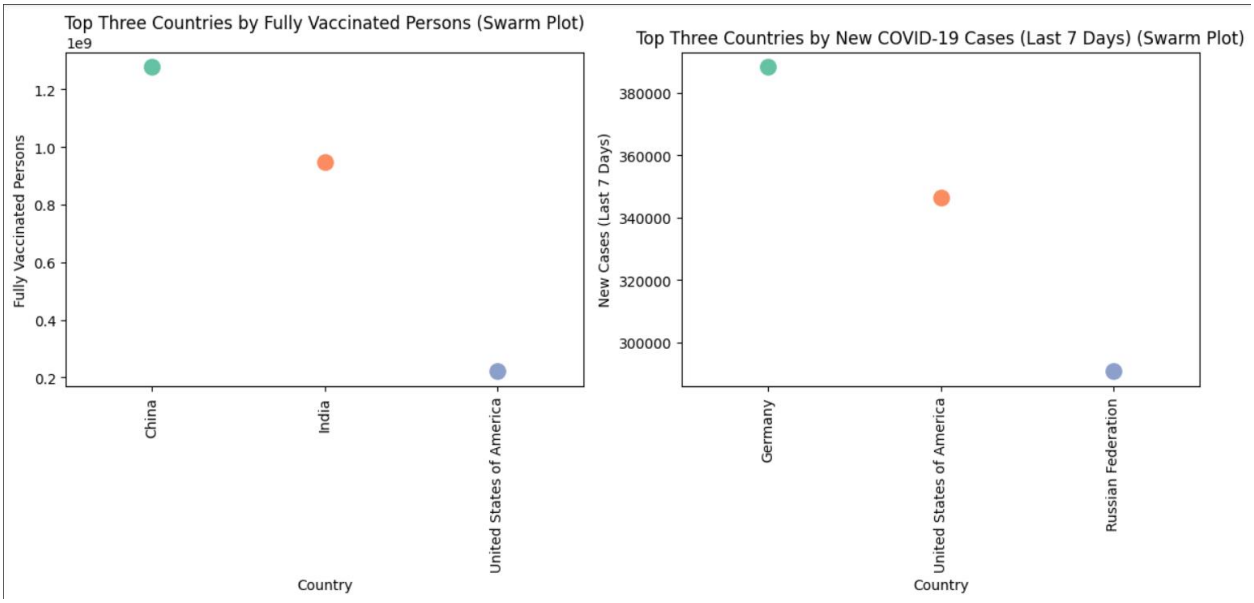
Python code

```
# find out top three countries COVID-19 cases and
vaccinated countries ? visualize this one with
swarmplot
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Load the CSV data into Pandas DataFrames
covid_data = pd.read_csv('/content/sample_data/1.
WHO-COVID-19-global-table-data.csv')
vaccination_data =
pd.read_csv('/content/sample_data/2. Vaccination-
data.csv')

# Sort and find the top three vaccinated countries
top_vaccinated =
vaccination_data.sort_values(by='PERSONS_FULLY_VACCIN
ATED', ascending=False).head(3)

# Sort and find the top three countries with the
highest COVID-19 cases reported
top_cases = covid_data.sort_values(by='Cases - newly
reported in last 7 days', ascending=False).head(3)

# Create a figure with subplots
plt.figure(figsize=(12, 6))
dot_colors = ['#990000', '#FF6600', '#000099']
# Swarm plot for top vaccinated countries
plt.subplot(1, 2, 1)
sns.swarmplot(x='COUNTRY',
y='PERSONS_FULLY_VACCINATED', data=top_vaccinated,
palette='Set2', size=12)
plt.xlabel('Country')
plt.ylabel('Fully Vaccinated Persons')
plt.title('Top Three Countries by Fully Vaccinated
Persons (Swarm Plot)')
plt.xticks(rotation=90)

# Swarm plot for top countries with new COVID-19
cases
plt.subplot(1, 2, 2)
```

```
sns.swarmplot(x='COUNTRY', y='Cases - newly reported
in last 7 days', data=top_cases,
palette='Set2',size=12)
plt.xlabel('Country')
plt.ylabel('New Cases (Last 7 Days)')
plt.title('Top Three Countries by New COVID-19 Cases
(Last 7 Days) (Swarm Plot)')
plt.xticks(rotation=90)
```

SCREENSHOT



Explation of the visualisation: This graphic uses a swarm plot to show the top three nations in terms of the total number of people who have received all recommended vaccinations as well as the greatest number of COVID-19 cases. The top three nations with the highest percentage of completely vaccinated citizens are depicted on the left side of the plot as a sequence of dots. Comparably, the figure on the right side uses a similar swarm of dots to show the top three nations with the greatest number of newly reported COVID-19 cases during the previous seven days.

An effective method of comparing these nations is through the use of swarm plots, which highlight any possible relationships between immunization campaigns and the state of COVID-19. The size of the dots has been changed to improve visibility, which makes it simpler to see patterns and variances in the numbers of cases and vaccinations throughout these high-achieving nations.

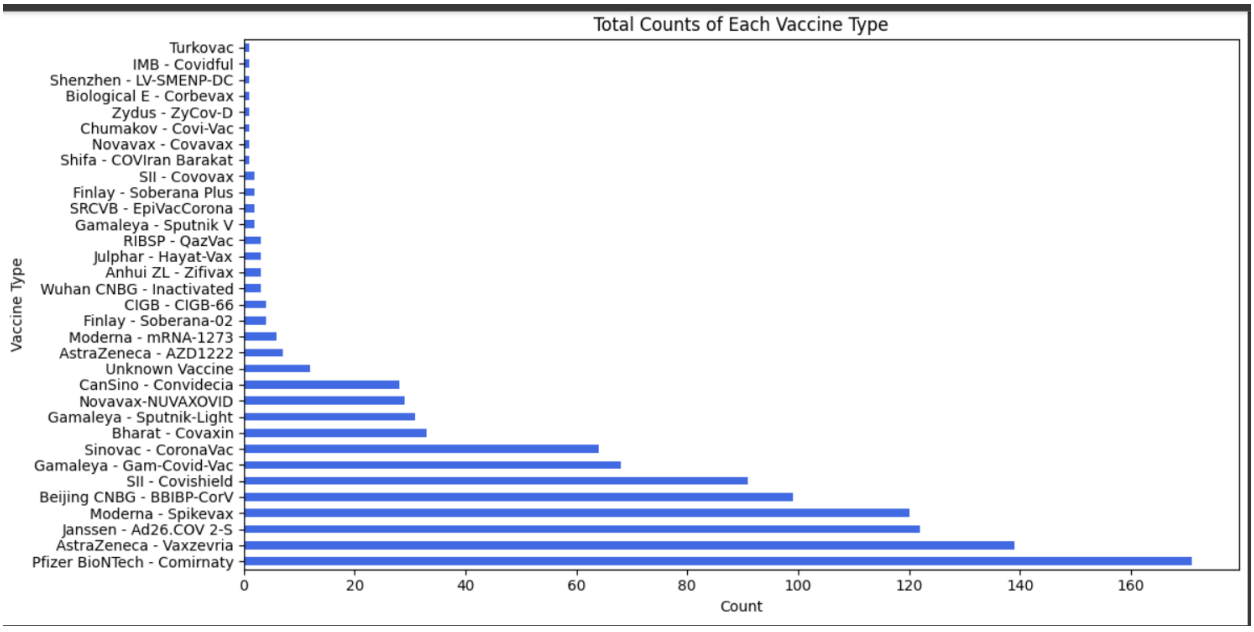
Visualisation 6:

- Question 6 : What are the total counts of each vaccine type used in the dataset?

Python code

```
import pandas as pd
import matplotlib.pyplot as plt

# Read the data from a CSV file into a DataFrame
df = pd.read_csv('/content/sample_data/2.
Vaccination-data.csv')

# Split and count the occurrences of each vaccine
type
vaccine_counts =
df['VACCINES_USED'].str.split(',').explode().str.stri
p().value_counts()

# Create a bar chart for visualization
vaccine_counts.plot(kind='barh', color='royalblue',
figsize=(12, 6))
plt.xlabel('Count')
plt.ylabel('Vaccine Type')
plt.title('Total Counts of Each Vaccine Type')
plt.tight_layout()

# Show the bar chart
plt.show()
```

SCREENSHOT



Explation of the visualisation: The total counts of each vaccine type utilised in COVID-19 vaccination efforts are visually represented by this bar chart. The information was taken out of the dataset that was supplied, which tracks the use of different vaccine kinds. The length of each horizontal bar on the chart represents the frequency or count of that vaccine being delivered, and each bar refers to a different type of vaccine. The vaccination type with the highest count is positioned at the top of the descending order of bars. Viewers will be able to easily determine which vaccine kinds have been utilized most frequently in COVID-19 vaccination campaigns in this way.

Gaining knowledge about the distribution and prevalence of various vaccination types over different regions or nations might be facilitated by examining this visualization. It facilitates comprehension of the decisions taken in immunization programmers and can enlighten conversations on vaccine accessibility, effectiveness, and distribution tactics. Additionally, it offers a clear and understandable means of informing a wide range of people—including legislators, medical experts, and the general public—about the importance of the COVID-19 immunization campaignn.

Visualisation 7:

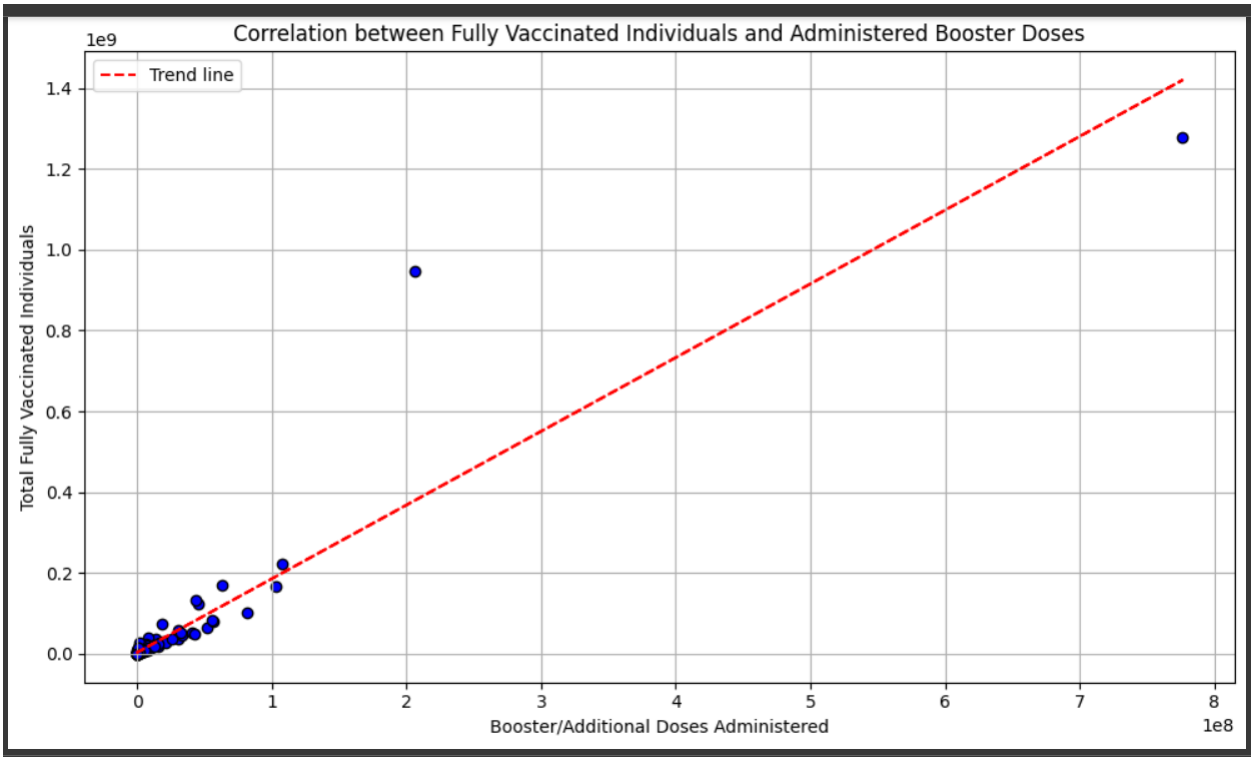
- Question 7: What does the scatter plot and trend line in this visualization reveal about the correlation between the number of administered booster/additional doses and the total number of fully vaccinated individuals?

Python code

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# Read the data from the CSV file into a DataFrame
df = pd.read_csv('/content/sample_data/2.
Vaccination-data.csv')

# Filter out rows with missing data in
'PERSONS_FULLY_VACCINATED' and
'PERSONS_BOOSTER_ADD_DOSE'
df = df.dropna(subset=['PERSONS_FULLY_VACCINATED',
'PERSONS_BOOSTER_ADD_DOSE'])

# Extract the relevant columns
fully_vaccinated = df['PERSONS_FULLY_VACCINATED']
booster_doses = df['PERSONS_BOOSTER_ADD_DOSE']

# Create a scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(booster_doses, fully_vaccinated, alpha=1,
color='blue', edgecolor='k')

# Add a trend line
z = np.polyfit(booster_doses, fully_vaccinated, 1)
p = np.poly1d(z)
plt.plot(booster_doses, p(booster_doses), "r--",
label='Trend line')

# Enhance the plot with a grid, labels and a legend
plt.xlabel('Booster/Additional Doses Administered')
plt.ylabel('Total Fully Vaccinated Individuals')
```

```
plt.title('Correlation between Fully Vaccinated
Individuals and Administered Booster Doses')
plt.legend()
plt.grid(True)

# Show the plot with tight layout
plt.tight_layout()
plt.show()
```

SCREENSHOT



Explation of the visualisation: The association between the number of individuals who have gotten all recommended vaccines and the booster or additional doses administered in various regions or countries is depicted in this scatter plot. The y-axis shows the total number of individuals who have gotten all advised immunisations, while the x-axis shows the number of boosters or extra doses administered. Every point on the diagram denotes a different country or region.

Individual data points are represented by the blue dots on the scatter plot, and the trend line fitted to the data by linear regression is shown by the red dotted line. The trend line aids in determining the general correlation between booster dosages and people who have received all recommended

vaccinations. If the trend line slopes upward, it indicates a positive connection and a propensity for more people to receive all recommended vaccinations as booster doses are given. On the other hand, a trend line that slopes downward would indicate a negative association. In order to evaluate this relationship visually and ascertain whether booster doses actually have a significant effect on attaining complete vaccination status, a scatter plot is utilised.

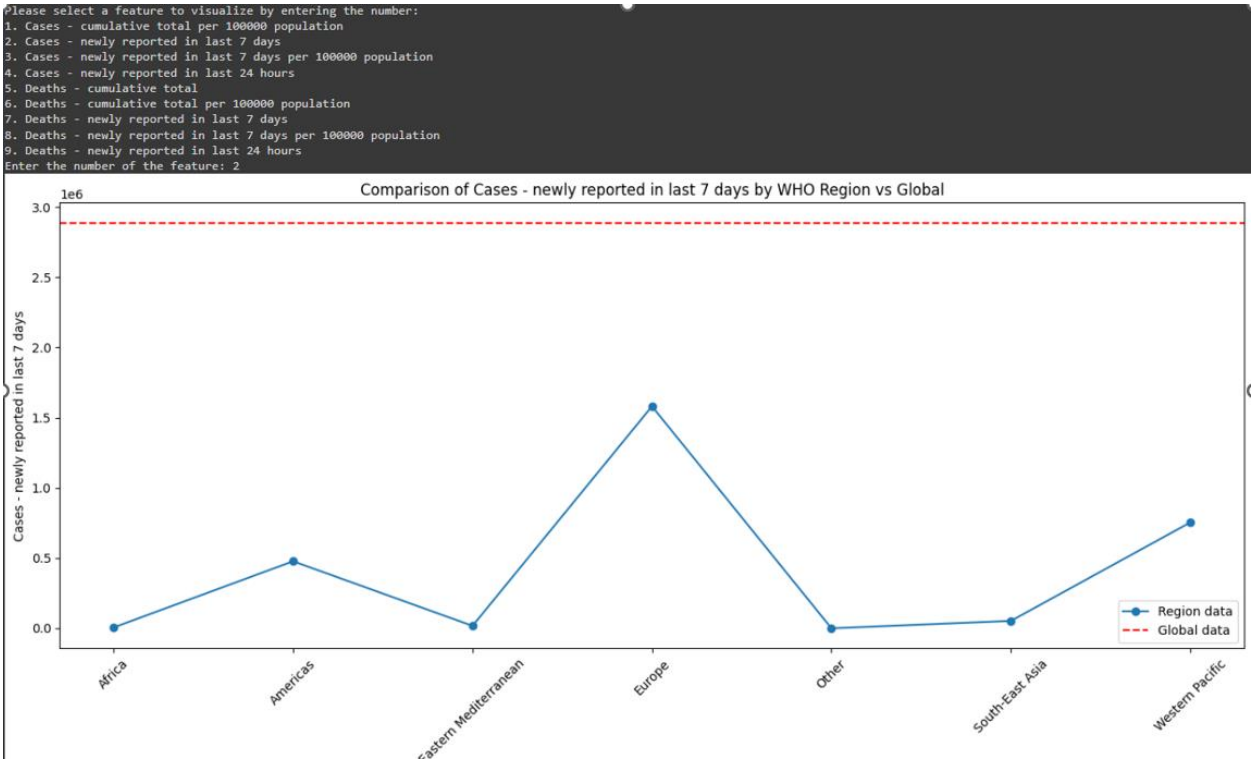
Visualisation 8:

- Question 8 : What does the line graph comparing the selected feature between global and WHO regions reveal about the differences and patterns in COVID-19 data?

Python code

```
# Extract the global data for the selected feature
global_data = global_df[feature].iloc[0]

# Create the plot
plt.figure(figsize=(14, 7))
plt.plot(region_grouped_data.index,
region_grouped_data.values, marker='o', linestyle='-
', label='Region data')
plt.axhline(y=global_data, color='r', linestyle='--',
label='Global data')
plt.title(f'Comparison of {feature} by WHO Region vs
Global')
plt.xlabel('WHO Region')
plt.ylabel(feature)
plt.legend()
plt.xticks(rotation=45)  # Rotate x-axis labels for
better readability
plt.tight_layout()  # Adjust plot to fit into the
figure area
plt.show()
```

SCREENSHOT



Explation of the visualisation: The differences and patterns in COVID-19 data between WHO regions and the global average are clearly displayed in this line graph. Every line denotes a different WHO area, illustrating the variations across these regions for the specified data feature (COVID-19 instances, vaccines, etc.). Important information can be gleaned from the slopes and locations of the lines in relation to the red dashed global average line. A region may have greater values for the selected attribute than the world average, which could indicate a higher disease prevalence or better vaccination rates, if its line continuously falls above the global line. On the other hand, regions with lower values are indicated by lines below the global standard; these regions may suggest lower vaccination rates or effective pandemic control.

Health officials, decision-makers, and academics can use this graphic as a quick reference to pinpoint areas that differ from worldwide trends in COVID-19 data and need particular interventions or assistance. In order to effectively combat the ongoing epidemic, it also highlights regions with remarkable performance or issues, which serves as a guide for decision-making and resource allocation. The graph also highlights significant patterns and trends that deepen our understanding of the various ways that different regions are handling the COVID-19 crisis.

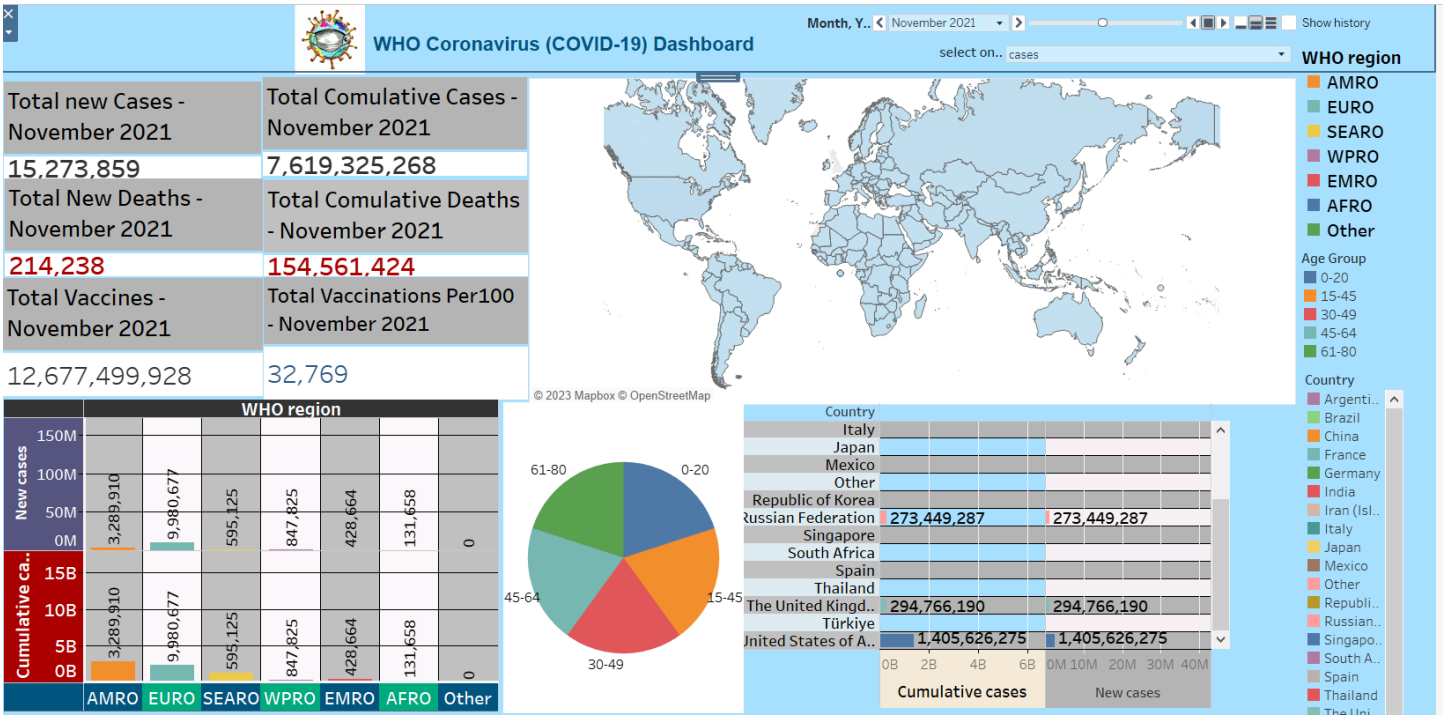
Targeted initiatives and policies to improve global public health outcomes can be informed by this information.

University of Roehampton London

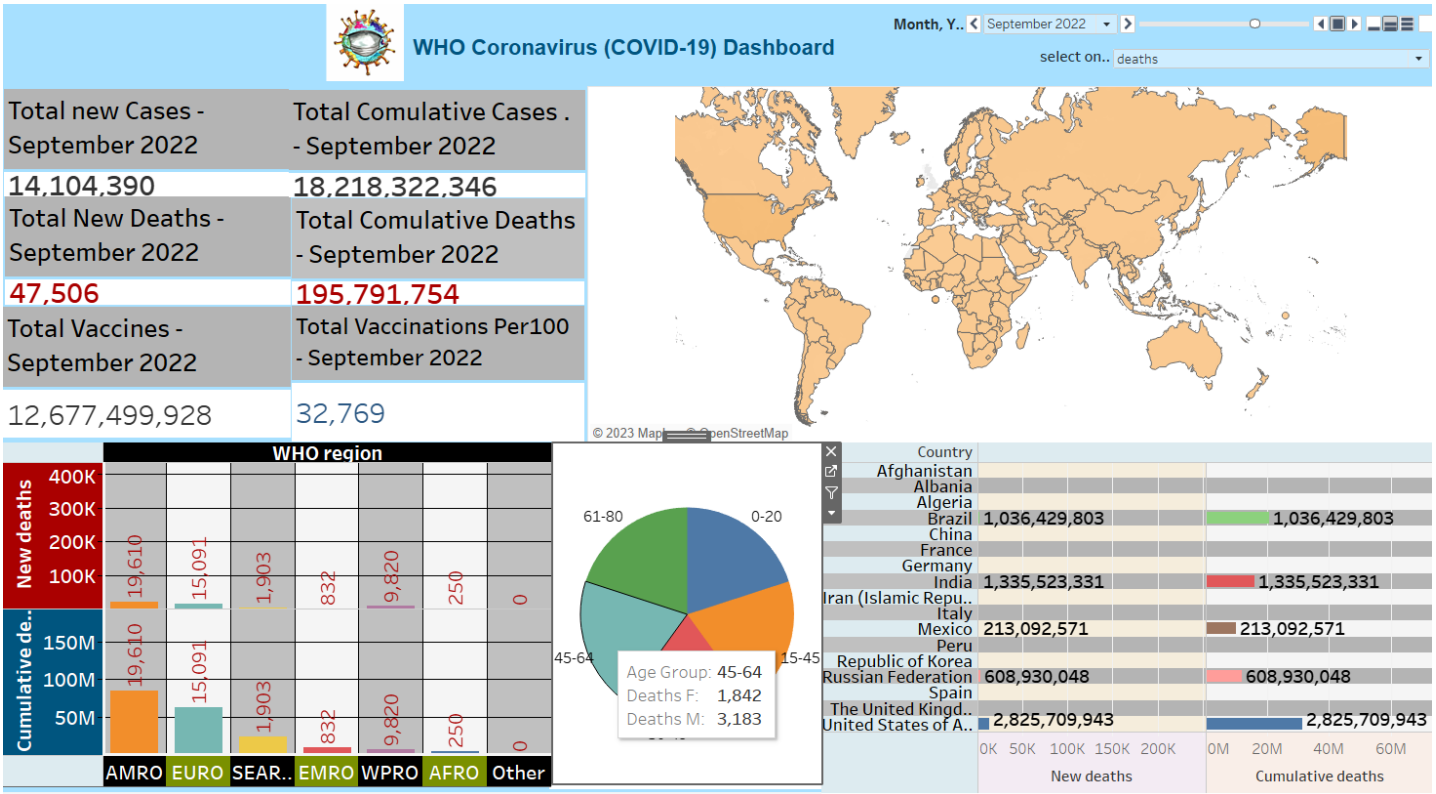
Task 2: TABLEAU VISUALISATION

Dashboard:

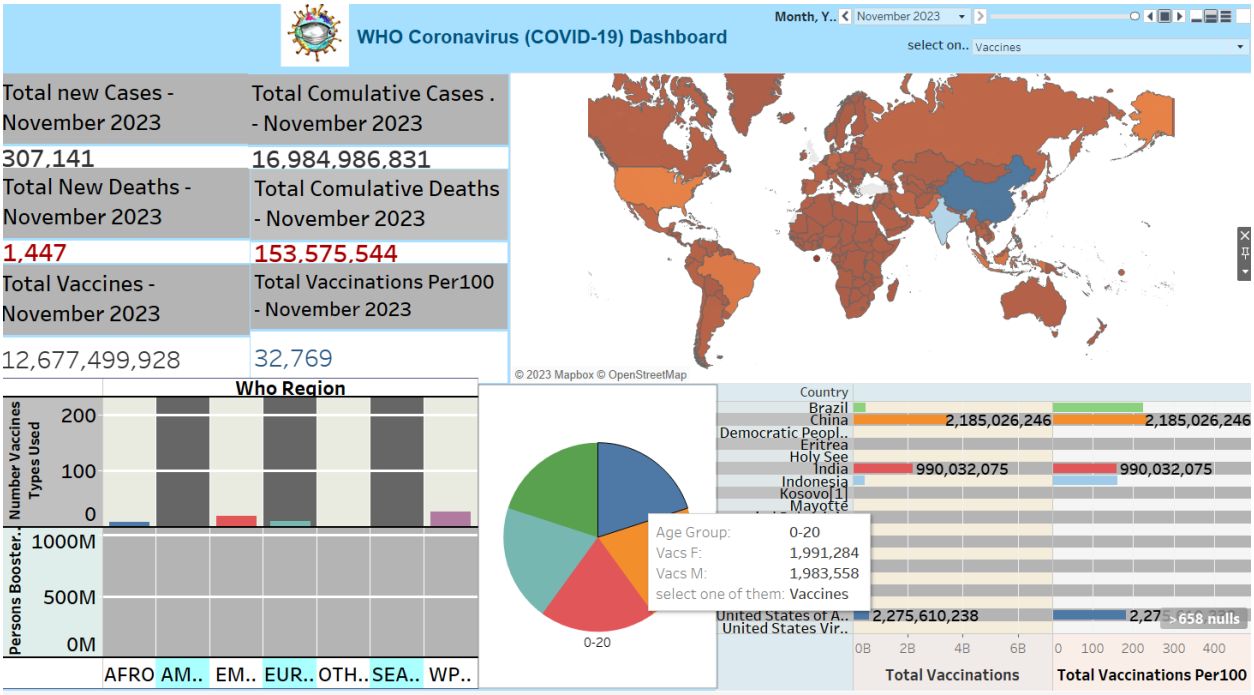
SCREENSHOT HERE



Dashboard: Covid-19 cases update



Dashboard: Covid-19 Deaths update



Dashboard: Covid-19 Vaccines update

Explation of the Deshboard: The WHO COVID-19 Dashboard is an interactive platform that displays the most recent worldwide pandemic statistics based on user-selected parameters and refreshes dynamically. Users can select cases, fatalities, or vaccination data from a selection feature at the top of the dashboard, and the display will update with the appropriate metrics. In the centre is a visualisation of a world map, with colour gradients that change to show the density of the user-selected statistic (such as the distribution of cases, death rates, or vaccine coverage). The most important numbers are shown clearly, giving an overview of the most recent and cumulative totals as well as the vaccination rates per 100 persons for the selected period.

These results are accompanied by a bar chart and a pie chart, which provide a comparative and demographic perspective by breaking down the data distribution between WHO regions and age groups. Granular analysis is possible due to the complete table that lists data particular to each country. The dashboard, which uses a time-series methodology, tracks the pandemic's path and modifies the information shown to the date chosen, offering both a historical and a present viewpoint. This dashboard is an essential tool for tracking and analysing the COVID-19 pandemic's emerging patterns, the effects of interventions, and the progress of current worldwide vaccination campaigns for public health authorities, academics, and the general public.

Question1: How can one analyze the trend of COVID-19 cases, deaths and vaccines in a specific country, using the interactive map with date?

Answer:

Explation: Users have access to an extensive tool for analysing trends and consequences over a range of time periods thanks to the interactive visualisation maps for COVID-19 related data. For example, by choosing Kazakhstan on the February 2021 instances map, readers may see that there were 27,524 new cases and 7,003,125 cumulative cases; the map's colour intensities show how far the pandemic has travelled. The severity of the health crisis in the Russian Federation is demonstrated by the mortality map for April 2021, which shows 11,278 new fatalities and 3,139,590 total deaths, underscoring the serious toll that the virus has taken. With Russia having received 175,974,391 vaccines in total and a vaccination rate of 120.6 per 100 persons, the vaccination map for 2022 makes it possible to analyse response efforts and get insight into the reach and efficacy of vaccination campaigns. These interactive maps are a priceless tool that let users explore intricate health information and see how the COVID-19 epidemic is evolving over time.

Question 2: How can the data visualisations that show the total number of doses of the vaccine and booster doses given, coupled with new and cumulative fatalities from COVID-19, provide a thorough picture of the pandemic's trends and the immunisation campaigns conducted in the various WHO regions?

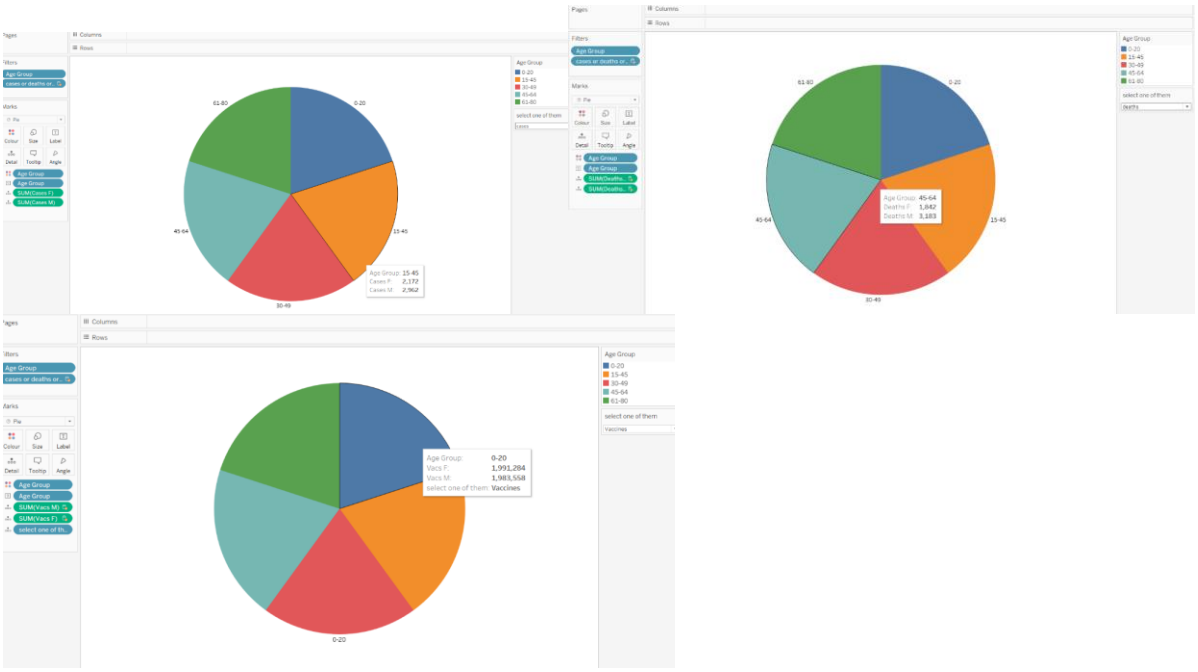
Answer:

Explation: The first chart displays cumulative and new cases, emphasising areas with a high overall case burden and areas where the virus is spreading quickly. This is essential for pinpointing hotspots and comprehending the temporal and spatial distribution of the illness.

The dark part of the pandemic—the new and cumulative deaths—is highlighted in the second chart. This information is critical for evaluating the virus's impact on mortality, the efficacy of medical interventions, and potentially the occurrence of more deadly variants in various geographic areas.
Finally, the vaccine dose chart illustrates the variety of vaccines used and the amount of booster doses given, providing insight into the breadth and depth of immunisation programmes. A greater variety of vaccines may be

the result of improved resource accessibility as well as a more varied vaccine strategy. On the other hand, the quantity of booster shots may indicate the degree to which a region values boosting immunity within its people.

Question 3: How do pie chart visualisations clarify the distinct effects of the pandemic on different demographic groups?and show the distribution of COVID-19 cases, deaths, and vaccinations by age group and gender.

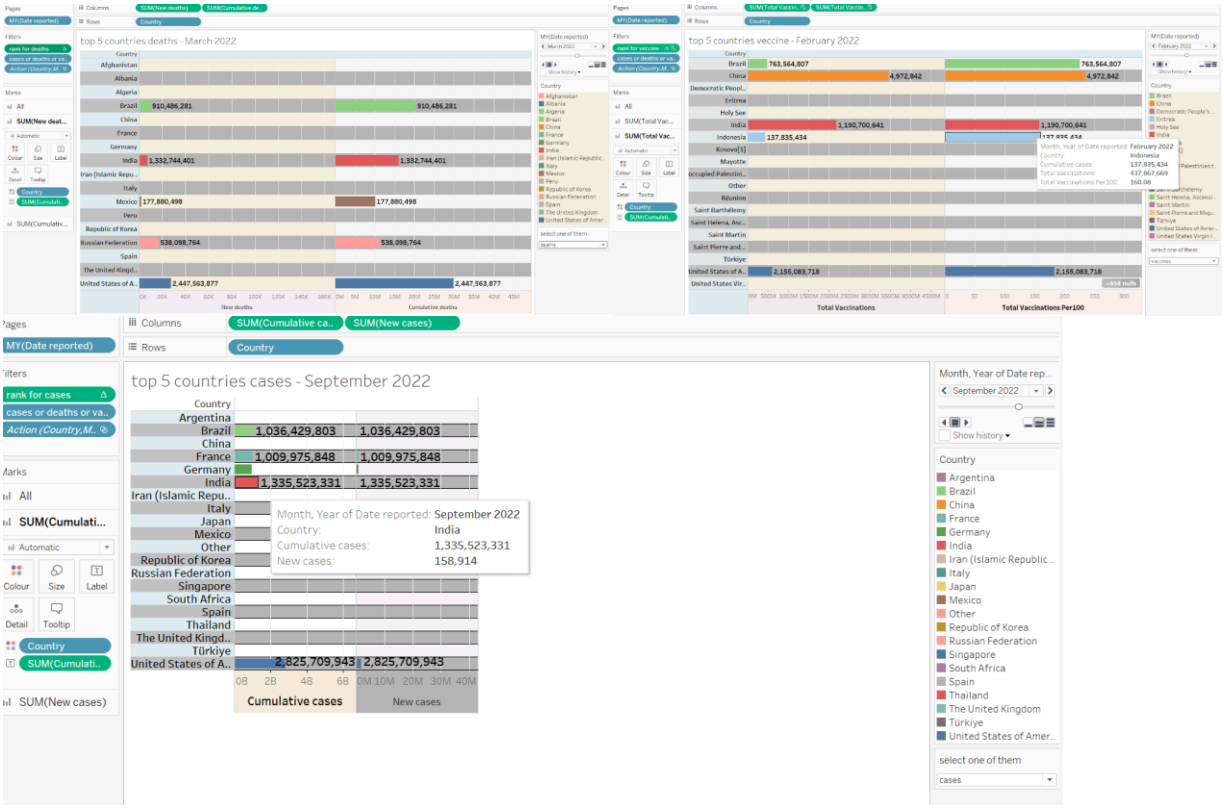


Explation: COVID-19 Cases by Age Group: The first pie chart illustrates the distribution of COVID-19 cases by age group. Larger slices imply a greater number of reported cases within specific age groupings, assisting in determining the most afflicted age groups.

Deaths in COVID-19 by Age Group:
The second graphic is about mortality, and it shows which age groups have the highest death rates. This can be critical for prioritising healthcare resources and safeguards for the most vulnerable populations.

COVID-19 Vaccinations by Age Group: The third graphic summarises vaccination efforts by age group, demonstrating success in immunising prospective high-risk populations and offering insights into the reach of vaccination programmes.

Question 4: How can we learn more about the extent of the pandemic and the efficacy of the response in these areas from the bar chart visualisations for the top five countries based on COVID-19 cases, fatalities, and vaccinations?



Explation: The visuals provide a comparative viewpoint on the pandemic's effects and responses in several countries. The bar charts show the overall cumulative numbers and fresh counts for cases and fatalities within a certain timeframe, indicating the seriousness of the outbreak and the continuing hazards. The charts illustrate the overall number of vaccines administered, providing insight into the scope of the immunisation programme. Stakeholders can see trends by comparing the top nations, such as which areas have greater rates of illness or vaccine coverage and maybe why. These understandings are essential for planning public health initiatives and allocating resources to lessen the pandemic's consequences.

REFERENCES

1. Age and gender dataset collected from https://globalhealth5050.org/the-sex-gender-and-covid-19-project/dataset/

2.who covid -19 data collected from https://covid19.who.int/data