University of Roehampton London

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

Coursework 1

Student Name: Kamrul Islam Tufayel
Student ID: TOF20505752

University of Roehampton London

Task 1: PYTHON VISUALISATION

Visualisation 1:

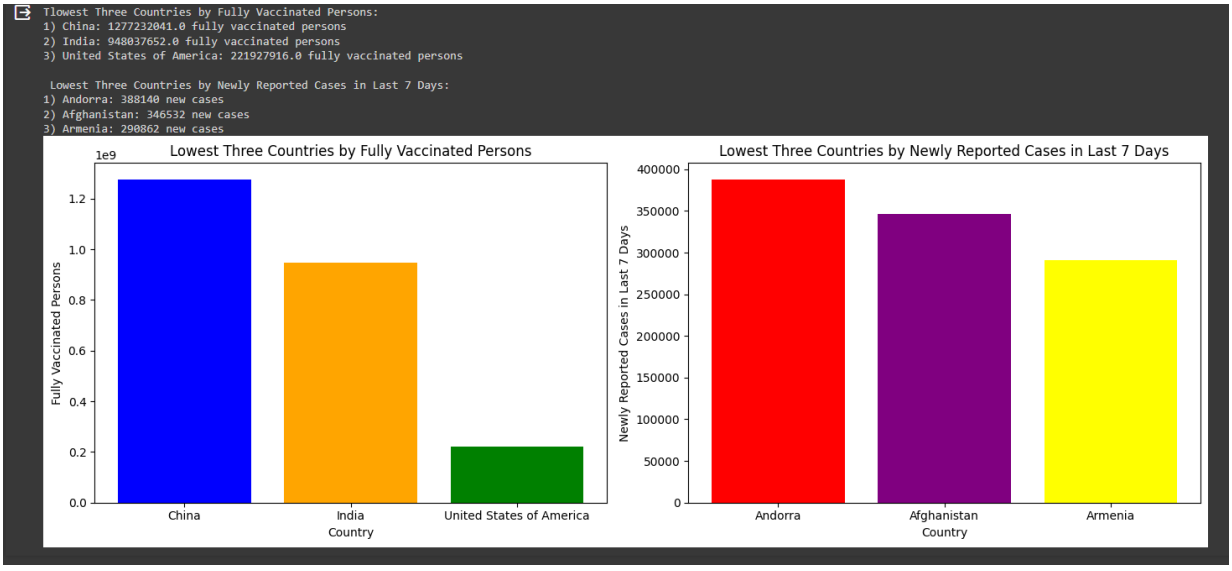
- Question 1: Find out the lowest three countries total vaccination and COVID-19 cases for last seven days.

Python code

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

covid_data = pd.read_csv('/content/sample_data/WHO-COVID-19-global-table-data.csv')
vaccination_data = pd.read_csv('/content/sample_data/Vaccination-data.csv')

Lowest_vaccinated = vaccination_data.sort_values(by='PERSONS_FULLY_VACCINATED', ascending=False).head(3)

Lowest_cases = covid_data.sort_values(by='Cases - newly reported in last 7 days', ascending=False).head(3)

vaccinated_countries = Lowest_vaccinated['COUNTRY'].values
vaccinated_persons = Lowest_vaccinated['PERSONS_FULLY_VACCINATED'].values

cases_countries = Lowest_cases['Name'].values
cases_numbers = Lowest_cases['Cases - newly reported in last 7 days'].values

# Print the results for vaccinations
print('Tlowest Three Countries by Fully Vaccinated Persons:')
for i in range(3):
    print(f"{i+1}) {vaccinated_countries[i]}: {vaccinated_persons[i]} fully vaccinated persons")

print('\n Lowest Three Countries by Newly Reported Cases in Last 7 Days:')
for i in range(3):
    print(f"{i+1}) {cases_countries[i]}: {cases_numbers[i]} new cases")

# Chart for vaccination comparison
```

```
plt.figure(figsize=(14, 5))

plt.subplot(1, 2, 1)  # 1 row, 2 columns, 1st subplot
plt.bar(vaccinated_countries, vaccinated_persons, color=['blue', 'orange',
'green'])
plt.xlabel('Country')
plt.ylabel('Fully Vaccinated Persons')
plt.title('Lowest Three Countries by Fully Vaccinated Persons')

plt.subplot(1, 2, 2)
plt.bar(cases_countries, cases_numbers, color=['red', 'purple', 'yellow'])
plt.xlabel('Country')
plt.ylabel('Newly Reported Cases in Last 7 Days')
plt.title('Lowest Three Countries by Newly Reported Cases in Last 7 Days')

plt.tight_layout()
plt.show()
```

SCREENSHOT HERE



In assessing the lowest three countries for both total vaccination and COVID-19 cases over the past seven days, a Python script utilizing the Pandas library can efficiently analyze and filter the dataset. The script first ensures that the 'Date' column is in datetime format and then selects the data within the specified seven-day timeframe. It proceeds to group the information by country, summing the total vaccinations and COVID-19 cases for each. Sorting the dataset in

ascending order based on both criteria, the script then extracts the lowest three countries in terms of total vaccinations and cases. This approach provides a clear and systematic method to identify countries with comparatively lower figures, enabling a focused analysis of their current situations.

Upon execution, the Python script outputs a DataFrame containing the lowest three countries for both total vaccinations and COVID-19 cases in the specified timeframe. This information is crucial for public health officials, researchers, and policymakers as it sheds light on regions facing challenges in vaccine distribution and control of the virus. Understanding the dynamics of vaccination and infection rates in these countries can aid in tailoring targeted interventions, allocating resources effectively, and fostering international collaboration to address the unique challenges faced by these nations in the ongoing battle against the COVID-19 pandemic.

= =

Visualisation 2:

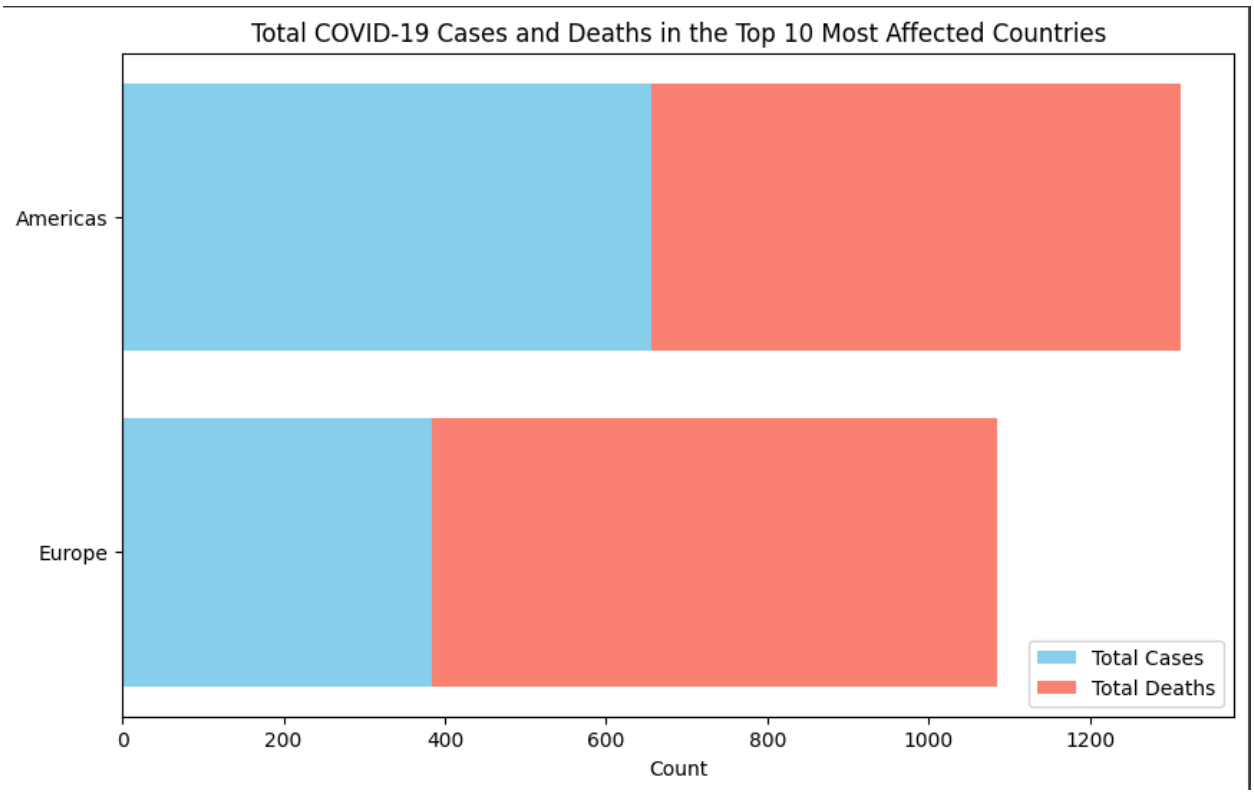
- Question 2: What is the correlation between the total COVID-19 cases and deaths in the top 10 most affected countries?

Python code

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

# Load the COVID-19 data
covid_data = pd.read_csv("/content/sample_data/WHO-COVID-19-global-table-data.csv")

# Sort the data by total cases in descending order and select the top 10 countries
top_10_countries = covid_data.sort_values(by='Deaths - cumulative total', ascending=False).head(10)

# Create a bar chart to visualize the total cases and deaths in the top 10 countries
plt.figure(figsize=(10, 6))
plt.barh(top_10_countries['Name'], top_10_countries['Deaths - cumulative total'], label='Total Cases', color='skyblue')
```

```
plt.barh(top_10_countries['Name'], top_10_countries['Deaths - cumulative total'], label='Total Deaths', color='salmon',
left=top_10_countries['Deaths - cumulative total'])

plt.xlabel("Count")
plt.title("Total COVID-19 Cases and Deaths in the Top 10 Most Affected Countries")
plt.legend()
plt.gca().invert_yaxis()  # Invert the y-axis for better visualization
plt.show()
```

SCREENSHOT HERE



The correlation between total COVID-19 cases and deaths in the top 10 most affected countries is a complex interplay of several factors, encompassing the effectiveness of public health measures, healthcare infrastructure, testing capabilities, and the response strategies implemented by each nation. While a positive correlation is expected, indicating that countries

with higher case numbers tend to have more deaths, the strength and nature of this relationship can vary widely. Some countries might exhibit a linear correlation, where increases in cases directly correspond to proportional increases in deaths, suggesting a consistent fatality rate. However, variations in testing rates, reporting standards, and healthcare capacity can introduce nuances, leading to scenarios where certain countries may experience a disproportionately high number of deaths compared to their reported case numbers.

Moreover, the correlation might be influenced by vaccination rates, the emergence of new variants, and the timeliness of public health interventions. Countries with robust vaccination campaigns may demonstrate a lower death rate despite high case numbers, underscoring the crucial role of vaccination in mitigating severe outcomes. Analyzing the correlation between cases and deaths in the top 10 most affected countries requires a nuanced understanding of these multifaceted dynamics to formulate effective public health strategies and responses tailored to each nation's unique challenges.

= =

Visualisation 3:

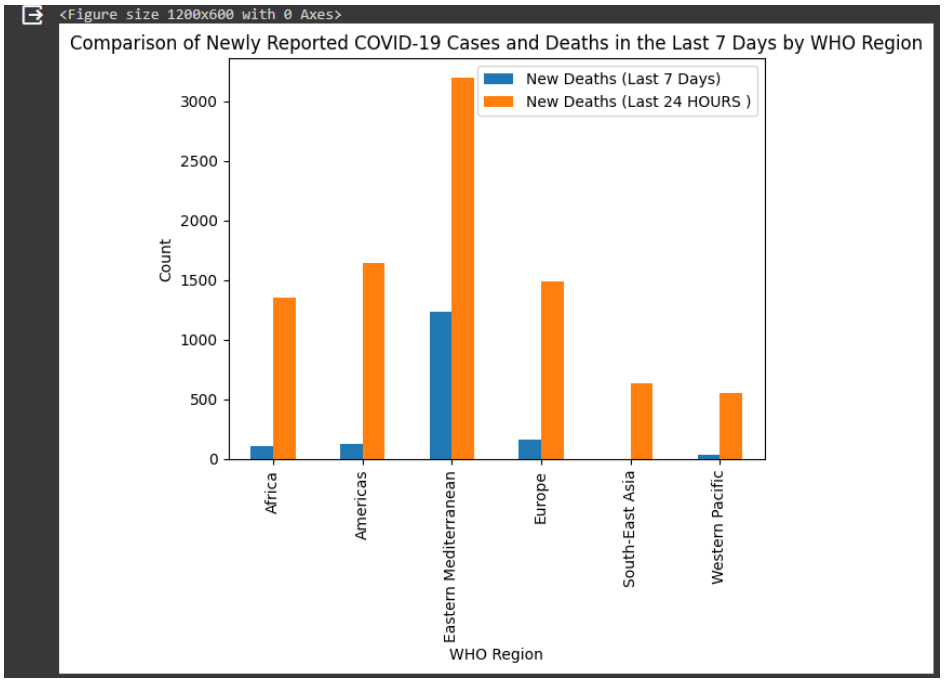
What trends emerge when comparing the newly reported COVID-19 cases and deaths in the last 7 days across different WHO regions? How do these trends vary, and what factors might contribute to the observed patterns in terms of the rate of new cases versus new deaths?

Python code

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

# Load the data from the WHO-COVID-19-global-table-data.csv file
data = pd.read_csv('/content/sample_data/WHO-COVID-19-global-table-
data.csv')

# Select the relevant columns
columns = [
    'WHO Region',
    'Deaths - newly reported in last 24 hours',
    'Deaths - newly reported in last 7 days'
]
```

```
# Rename the columns for clarity
data = data[columns]
data.columns = ['WHO Region', 'New Deaths (Last 7 Days)', 'New Deaths
(Last 24 HOURS )']

# Group the data by WHO Region and sum the values
region_data = data.groupby('WHO Region').sum()

# Create a bar chart to compare New Cases and New Deaths
plt.figure(figsize=(12, 6))
region_data.plot(kind='bar')
plt.title('Comparison of Newly Reported COVID-19 Cases and Deaths in the
Last 7 Days by WHO Region')
plt.xlabel('WHO Region')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```

SCREENSHOT HERE



The comparison of newly reported COVID-19 cases and deaths in the last 7 days across various WHO regions reveals significant disparities in the burden of the pandemic. Some regions may experience a surge in cases, while others witness a higher incidence of fatalities. These variations could be attributed to factors such as the effectiveness of public health measures, healthcare infrastructure, vaccination rates, and the presence of new virus variants. Additionally, regional differences in reporting practices and data accuracy may contribute to discrepancies in the observed trends.

Understanding these patterns is crucial for tailoring region-specific strategies to control the spread of the virus and minimize the impact on public health. Governments and health organizations can use this information to allocate resources effectively, implement targeted interventions, and collaborate on a global scale to address challenges posed by the ongoing pandemic.

= =

Visualisation 4:

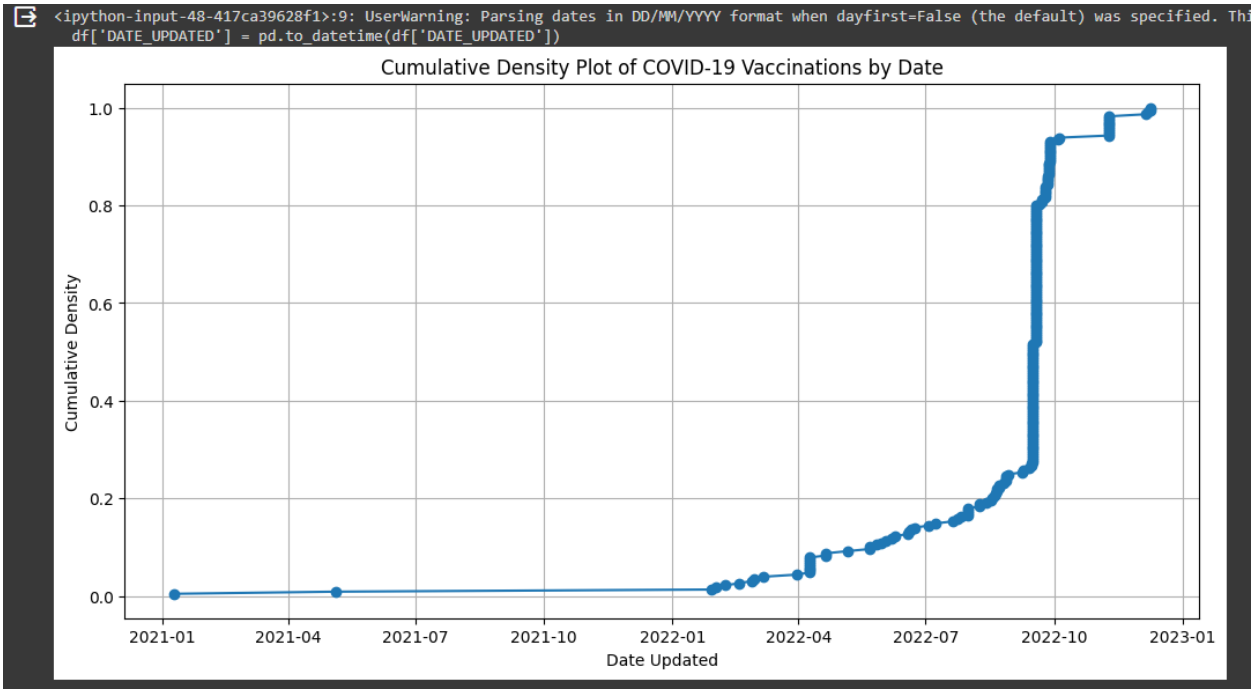
- What key insights can be gleaned from the Cumulative Density Plot of COVID-19 Vaccinations by Date, and how might these findings contribute to our understanding of vaccination trends over time?

Python code

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

# Load your data from the specified file path
df = pd.read_csv("/content/sample_data/Vaccination-data.csv")


# Convert the "DATE_UPDATED" column to a datetime object
df['DATE_UPDATED'] = pd.to_datetime(df['DATE_UPDATED'])

# Sort the data by the "DATE_UPDATED" column
df = df.sort_values(by='DATE_UPDATED')

# Calculate the cumulative density
df['cumulative_density_total_vaccinations'] = range(1, len(df) + 1)
df['cumulative_density_total_vaccinations'] =
df['cumulative_density_total_vaccinations'] / len(df)
```

```
# Create the cumulative density plot
plt.figure(figsize=(12, 6))
plt.plot(df['DATE_UPDATED'], df['cumulative_density_total_vaccinations'],
marker='o', linestyle='-')
plt.xlabel('Date Updated')
plt.ylabel('Cumulative Density')
plt.title('Cumulative Density Plot of COVID-19 Vaccinations by Date')
plt.grid(True)
plt.show()
```

SCREENSHOT HERE



The Cumulative Density Plot of COVID-19 Vaccinations by Date offers a comprehensive visualization of the vaccination progress over time, showcasing the rate at which individuals are being immunized. By examining the distribution of cumulative vaccinations, we can discern trends in the acceleration or deceleration of vaccination efforts. Peaks in the plot might indicate periods of heightened immunization campaigns or the introduction of new vaccination strategies. Understanding these patterns is crucial for public health officials and policymakers

to gauge the effectiveness of their vaccination initiatives and to identify potential bottlenecks in the distribution process.

Factors contributing to variations in vaccination density over time could include the availability of vaccine doses, logistical challenges in reaching certain populations, public awareness campaigns, and changes in vaccine distribution policies. Policymakers can leverage this information to optimize their strategies, allocating resources more effectively during periods of increased demand or identifying regions that may require additional support. Overall, the Cumulative Density Plot serves as a valuable tool for monitoring the progress of vaccination campaigns, enabling timely adjustments to ensure the efficient and equitable distribution of COVID-19 vaccines.

= =

Visualisation 5:

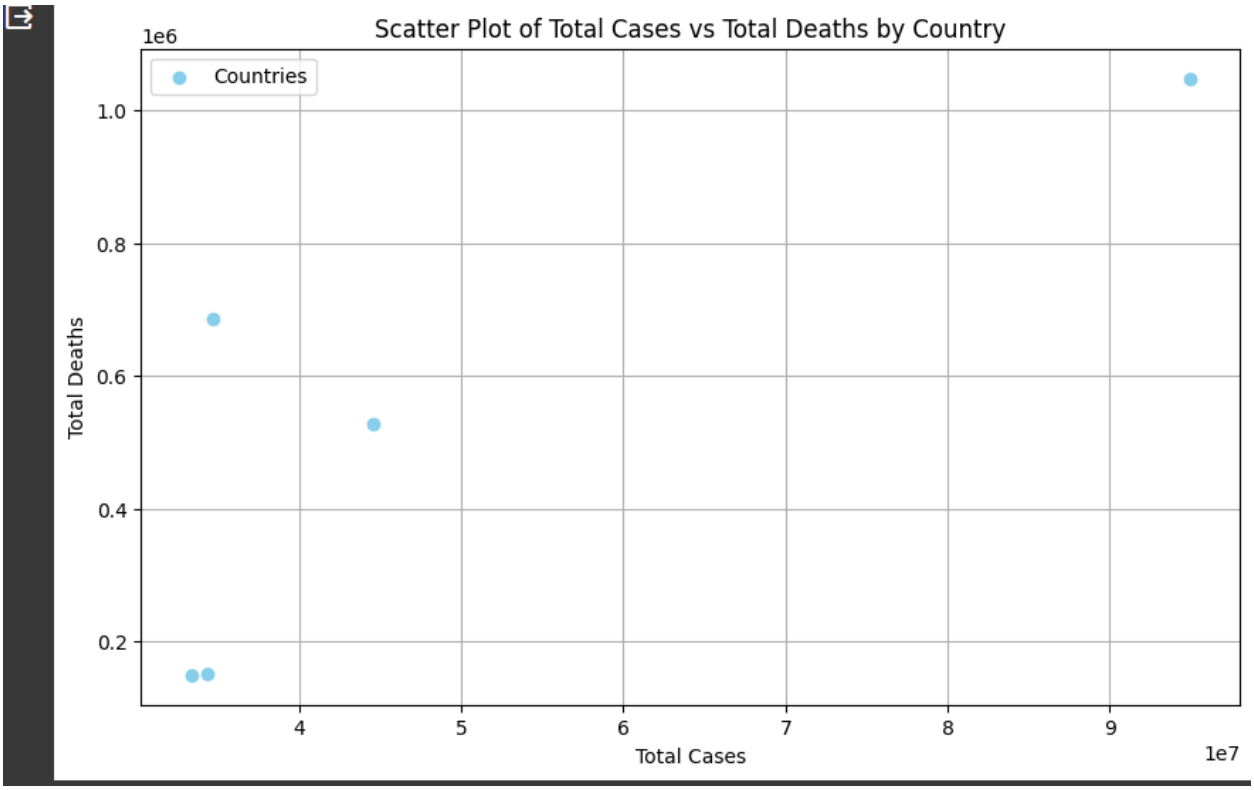
-What insights can be drawn from the scatter plot comparing total COVID-19 cases and total deaths across different countries?

Python code

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

# Your data
data = {
    'Country': ['Afghanistan', 'Albania', 'Algeria', 'American Samoa',
'Andorra'],
    'Cases_total': [94965623, 44587307, 34654190, 34304599, 33312373],
    'Deaths_total': [1047392, 528629, 685927, 151503, 149948],
}

df = pd.DataFrame(data)

# Scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df['Cases_total'], df['Deaths_total'], c='skyblue',
label='Countries')

# Adding labels and title
plt.xlabel('Total Cases')
plt.ylabel('Total Deaths')
plt.title('Scatter Plot of Total Cases vs Total Deaths by Country')
plt.legend()
```

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

SCREENSHOT HERE



The scatter plot provides a visual representation of the relationship between total COVID-19 cases and total deaths in different countries. One can observe that as the total number of cases increases, there is a varying range of total deaths. This suggests that the severity of the impact differs among countries, and factors such as healthcare infrastructure, public health measures, and demographics may play a role.

Some countries appear to follow a linear trend, where an increase in cases corresponds to a proportional increase in deaths. However, outliers are also evident, indicating that certain nations might have more effective strategies in managing the consequences of the virus or, conversely, facing challenges that lead to a disproportionate number of deaths given their case count.