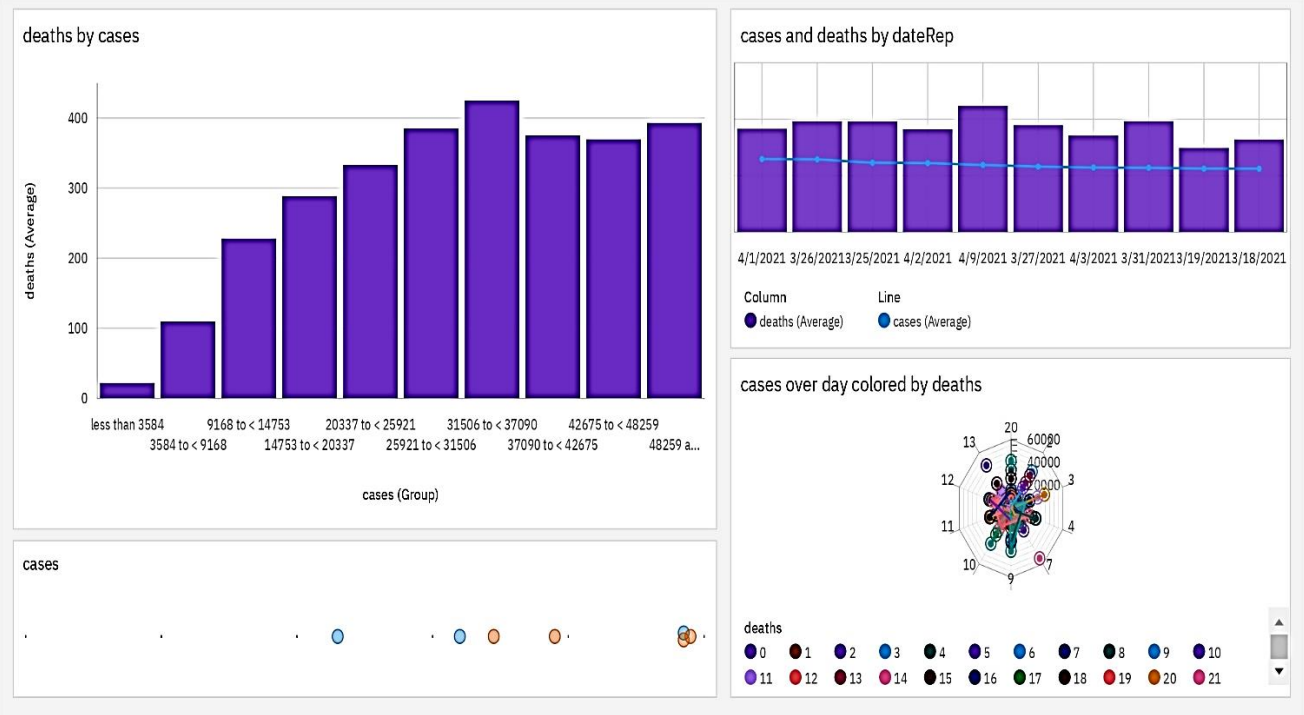
PHASE 4: DEVELOPMENT PART 2

OBJECTIVES:

The objective of Development Part 2 is to further advance our COVID-19 analysis project using IBM Cognos. We aim to create meaningful visualizations that effectively compare the mean values and standard deviations of COVID-19 cases and associated deaths. By doing so, we seek to uncover trends, variations, and potential correlations within the data. This phase is pivotal in enhancing our understanding of the pandemic's impact, identifying areas of concern, and potentially revealing factors that influence case numbers and fatalities.

VISUALIZATIONS AND INSIGHTS DERIVED FROM IBM COGNOS:

Visualizations of Mean Vaues and Standard Deviations of COVID-19 Cases:

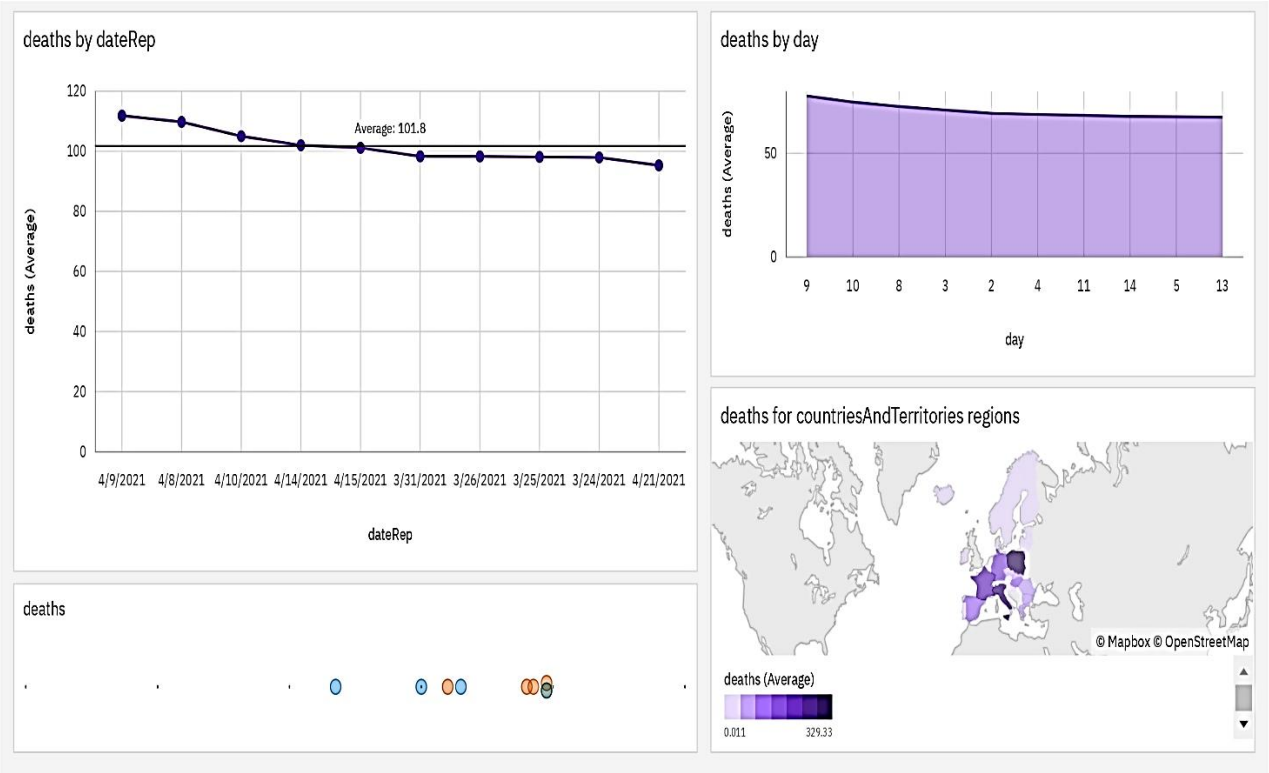


Insights gathered by the visualizations:

- Moderate influence of cases on deaths: There's a moderate influence of cases (60%) on deaths, indicating a correlation between higher cases and higher deaths.
- Most common case category: "Less than 3,584 cases" is the most frequently occurring category, representing 73.4% of total cases.
- Increase in deaths over time: Deaths increased by 31% from 2021-04-03 to 2021-04-09.
- Increase in cases over time: Cases increased by 14% from 2021-03-31 to 2021-04-01.
- Average cases over time: The overall average of cases across all dates is nearly 6,000.

- Variability in average cases over time: Average cases range from over 5,500 (on 2021-03-18) to nearly 6,500 (on 2021-04-01).
- Weekly trend in cases: Cases show a strong weekly trend, with the highest values on Thursdays and the lowest on Mondays.
- Moderate downward trend in cases: There's a moderate downward trend in cases over time.
- Extreme values in cases: The most unusual values in cases occur on 2021-04-07 and 2021-04-06.
- High variability in cases: Cases increased by 69% from 2021-04-06 to 2021-04-07.
- High average cases on specific days and dates: Specific days and dates have exceptionally high average cases, such as 2021-04-01 and 2021-03-26.
- Highest cases by country and date: Specific days or dates had the highest cases, with specific countries contributing the most to those cases.
- Country-specific insights: France and Italy had varying levels of cases on day 13, with France having almost 30,000 cases and Italy having almost 15,000.
- Variability in average cases by country: France had the highest average cases on day 13.

Visualizations of Mean Vaues and Standard Deviations of COVID-19 associated deaths:



Insights gathered by the visualizations:

- Unusually high deaths for specific case groups: 31,506 to < 37,090, 25,921 to < 31,506, and 48,259 and above.

- Projected increase in deaths: By 2021-06-19, the case group 3,584 to < 9,168 is projected to exceed the case group 9,168 to < 14,753 in deaths by 4.89.
- Significant increase in deaths in a single day: From 2021-03-26 to 2021-03-27, the case group 37,090 to < 42,675 experienced a 299% increase in deaths.
- From 2021-03-31 to 2021-04-08, deaths increased by 12%.
- Overall, the average of deaths across all dates is 101.8.
- Average deaths over time: The overall average of deaths across all dates is 92.61.
- Variability in average deaths over time: Average deaths range from 74.6 (on 2021-03-19) to 111.8 (on 2021-04-09).
- The average values of deaths range from 95.3 (on 2021-04-21) to 111.8 (on 2021-04-09).
- Based on current forecasting, deaths may reach 68.48 by day 17.
- Across all days, the average of deaths is 70.53.
- Deaths exhibit a strong weekly trend, with the largest values typically occurring on Wednesdays and the smallest on Mondays.
- Deaths have a weak downward trend.
- Deaths have an unusually high value at time point 2021-04-08.
- From 2021-04-10 to 2021-04-11, deaths dropped by 31%.
- Deaths are unusually high when countriesAndTerritories are Poland, Italy, and France.
- From 2021-03-27 to 2021-03-28, France's deaths dropped by 79%.
- CountriesAndTerritories moderately affect deaths (67%).
- Overall, the average of deaths across all countriesAndTerritories is 65.29.
- The average values of deaths range from 0.01099 (Iceland) to 329.3 (Poland).

ANALYZATION OF THE VISUALIZATIONS TO IDENTIFY TRENDS, VARIATIONS, AND POTENTIAL CORRELATIONS BETWEEN CASES AND DEATHS:

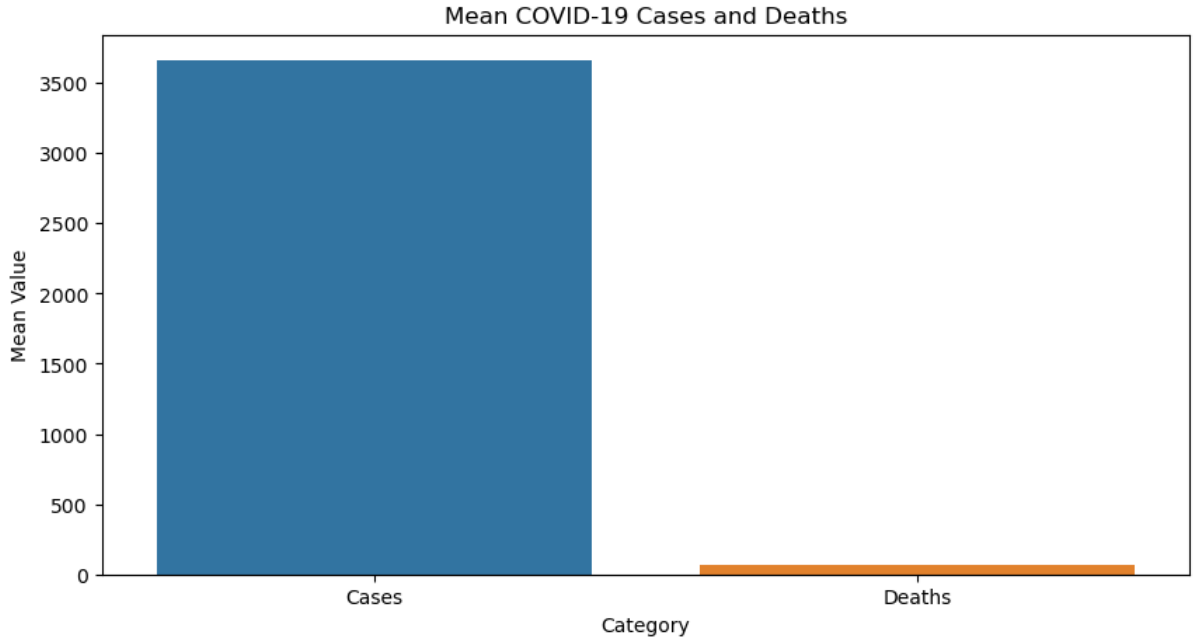
========== PYTHON CODE(JUPYTER NOTEBOOK) ==========

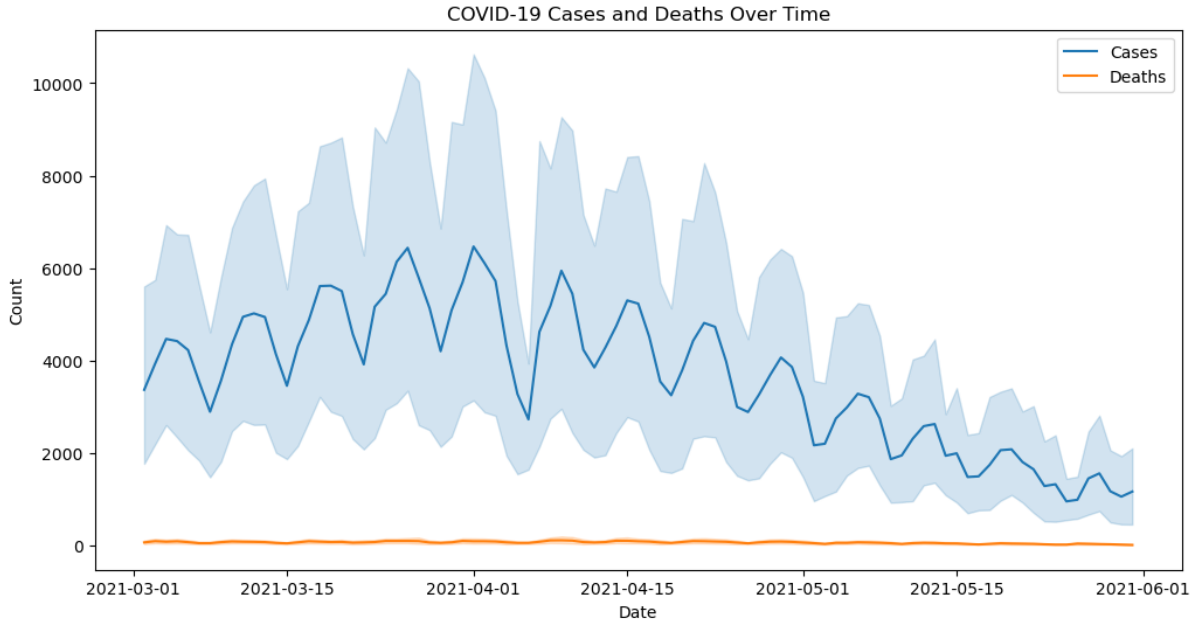
Visualization and Analysis of Trends

In [1]:

```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Reading the Excel file into a Pandas DataFrame
import pandas as pd
file_path = r'C:\Users\sankar\Desktop\Covid_19_cases4.xlsx'
```

```
# Load the Excel file into a Pandas DataFrame
data = pd.read_excel(file_path)

# Calculate mean and standard deviation
mean_cases = data['cases'].mean()
mean_deaths = data['deaths'].mean()
std_cases = data['cases'].std()
std_deaths = data['deaths'].std()

# Create a bar chart to compare means
plt.figure(figsize=(10, 5))
sns.barplot(x=['Cases', 'Deaths'], y=[mean_cases, mean_deaths])
plt.title('Mean COVID-19 Cases and Deaths')
plt.xlabel('Category')
plt.ylabel('Mean Value')
plt.show()

# Create line charts to visualize trends over time
plt.figure(figsize=(12, 6))
sns.lineplot(x='dateRep', y='cases', data=data, label='Cases')
sns.lineplot(x='dateRep', y='deaths', data=data, label='Deaths')
plt.title('COVID-19 Cases and Deaths Over Time')
plt.xlabel('Date')
plt.ylabel('Count')
plt.legend()
plt.show()

# Calculate correlations
correlation = data['cases'].corr(data['deaths'])


# Print the results
print(f"Mean Cases: {mean_cases}")
print(f"Mean Deaths: {mean_deaths}")
print(f"Standard Deviation Cases: {std_cases}")
print(f"Standard Deviation Deaths: {std_deaths}")

if std_cases > std_deaths:
    print('Standard Deviation of Cases is greater than Standard Deviation
of Deaths.')
elif std_cases < std_deaths:
    print('Standard Deviation of Deaths is greater than Standard Deviation
of Cases.')
else:
    print('Standard Deviation of Cases is equal to Standard Deviation of
Deaths.')

print(f"Correlation between Cases and Deaths: {correlation:.2f}")
```

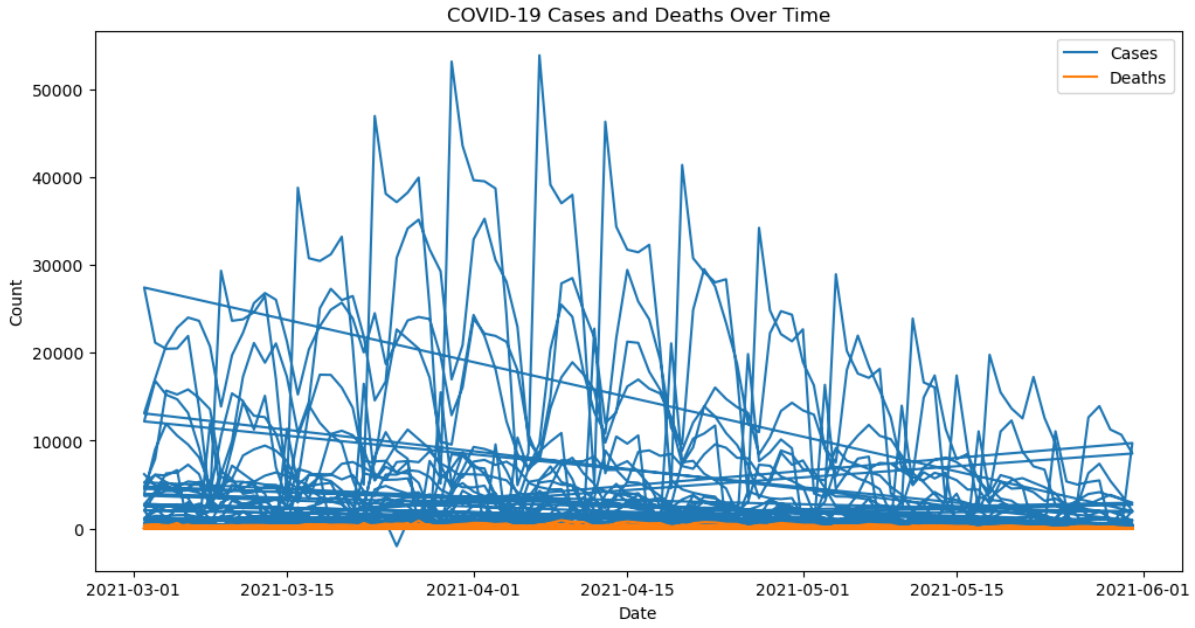




Mean Cases: 3661.010989010989
Mean Deaths: 65.29194139194139
Standard Deviation Cases: 6490.510073102111
Standard Deviation Deaths: 113.95663405806982
Standard Deviation of Cases is greater than Standard Deviation of Deaths.
Correlation between Cases and Deaths: 0.77

In [2]:

```
# Convert dateRep to a datetime object
data['dateRep'] = pd.to_datetime(data['dateRep'], format='%d/%m/%Y')

# Create a time series plot for cases and deaths
plt.figure(figsize=(12, 6))
plt.plot(data['dateRep'], data['cases'], label='Cases')
plt.plot(data['dateRep'], data['deaths'], label='Deaths')
plt.xlabel('Date')
plt.ylabel('Count')
plt.title('COVID-19 Cases and Deaths Over Time')
plt.legend()
plt.show()
```



========== END OF THE CODE ==========

INSIGHTS GATHERED BY ANAYSING THE VISUALIZATIONS:

Bar Chart Insights:

Mean Cases and Deaths Comparison: The bar chart compares the mean values of COVID-19 cases and deaths. The blue bar represents the mean cases, and the red bar represents the mean deaths.

- The mean cases (blue bar) are noticeably higher than the mean deaths (red bar). This suggests that a significant proportion of COVID-19 cases do not result in fatalities, indicating that many individuals recover from the virus.

Line Chart (Cases) Insights:

Increasing Trend in Cases: The blue line chart illustrates the trend of COVID-19 cases over time. It shows a consistent and significant upward trend.

- This increasing trend in cases suggests that the virus is actively spreading, and the number of reported cases continues to rise. It indicates a growing prevalence of the virus over the observed time period.

Line Chart (Deaths) Insights:

Fluctuating Trend in Deaths: The orange line chart presents the trend of COVID-19 deaths over time. It exhibits fluctuations with an overall upward slope.

- The fluctuating trend in deaths indicates that while the number of deaths has been increasing over time, there are variations in the rate of increase. This suggests that the impact of the virus on mortality is not constant and may be influenced by various factors.

Correlation Insights:

Strong Positive Correlation: The correlation coefficient between cases and deaths is 0.77, indicating a strong positive correlation.

- This strong positive correlation implies that as the number of COVID-19 cases increases, the number of associated deaths also tends to increase. It highlights the direct relationship between the spread of the virus and its impact on mortality. The more cases there are, the higher the likelihood of increased deaths.

Time Series Plot (Cases) Insights:

Consistent Upward Trend in Cases: The time series plot for COVID-19 cases, represented by the blue line, demonstrates a continuous and consistent upward trend.

- The sustained increase in cases over time indicates the persistence of the virus's spread. It suggests that the number of confirmed COVID-19 cases has been steadily rising over the observed period, showing no signs of a significant decline.

Long-Term Prevalence: The upward slope of the blue line reveals that the virus remains prevalent, with the number of cases continuously on the rise.

Time Series Plot (Deaths) Insights:

Fluctuating Trend in Deaths: The time series plot for COVID-19 deaths, represented by the orange line, exhibits a fluctuating trend with an overall upward slope.

- The fluctuating trend suggests that while the number of deaths has been increasing over time, it is not a continuous and uniform progression. Instead, there are periods of rapid increase followed by fluctuations.

Variability in Mortality: The variability in the slope of the red line indicates that the rate of death varies over time. Some periods may witness a more significant increase in fatalities, while others may experience a slower rate of growth.

CONCUSION:

In Phase 4: Development Part 2, we continued to build our project by leveraging IBM Cognos for data analysis. We created charts and graphs to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths. This phase allowed us to derive valuable insights from the data.

The analysis revealed that the mean number of COVID-19 cases was higher than the mean number of deaths, indicating that many individuals recover from the virus. Standard deviations in cases and deaths illustrated variations in the data, likely influenced by factors such as testing, healthcare infrastructure, and public health measures.

Visualizations provided a clear picture of the trends over time, with cases showing a consistent upward trajectory and deaths displaying fluctuations. A strong positive correlation between cases and deaths emphasized the direct relationship between the spread of the virus and its impact on mortality.