Data Analysis Project : Covid 19 Data Analysis

Global COVID-19 Data Analysis Project

1. Introduction

The COVID-19 pandemic has created unprecedented challenges for governments, healthcare systems, and global organizations. Accurate data analysis is essential to understand the spread, severity, and recovery patterns of the disease. This project performs an exploratory and visual analysis of country-wise COVID-19 data to generate actionable insights for policymakers, health organizations, and decision-makers.

---

2. Business Context

Governments, health organizations such as the World Health Organization (WHO), and policymakers require data-driven insights to:

- **Allocate medical resources efficiently**
- **Monitor outbreak severity across countries**
- **Compare regional healthcare performance**
- **Identify high-risk regions and early warning signals**
- **Take preventive and corrective actions**

This project analyses country-wise COVID-19 data to support strategic planning, healthcare management, and policy formulation.

---

3. Dataset Description

The dataset used in this project contains country-wise COVID-19 statistics. Key attributes include:

- **Country/Region**
- **Confirmed Cases**
- **Deaths**
- **Recovered Cases**
- **Active Cases**
- **New Cases, New Deaths, New Recoveries**

- **Weekly Change and Percentage Increase**
- **WHO Region**

The dataset enables both country-level and region-level analysis.

---

4. Tools & Technologies Used

- **Programming Language: Python**
- **Libraries: Pandas, Matplotlib, Seaborn**
- **Techniques: Data Cleaning, Group by Aggregation, Exploratory Data Analysis (EDA)**
- **Visualization: Bar charts, Scatter plots, Histogram, Line charts, Pie charts, KDE, Box plots**

Business Problems & Analysis

◆ Problem 1: How many countries are present in the dataset?

Objective

To identify the total number of unique countries included in the dataset in order to understand the geographical coverage and scale of the analysis.

Approach

The number of unique countries was calculated using the Country/Region column by applying a uniqueness count. This ensures that each country is counted only once, even if multiple records exist.

Result

The dataset contains data for 187 unique countries.

Business Insight

A dataset covering 187 countries provides a comprehensive global view of the COVID-19 pandemic. This wide coverage enables meaningful cross-country and regional comparisons, making the analysis reliable for policy evaluation, healthcare planning, and global trend assessment.

```python
# How many countries are present in the dataset?
country=data["Country/Region"].nunique()
country
```

```
187
```

◆ Problem 2: What is the total confirmed COVID-19 cases worldwide?

Objective

To determine the overall global impact of COVID-19 by calculating the total number of confirmed cases reported across all countries in the dataset.

Approach

The total confirmed cases were computed by applying a summation operation on the Confirmed column. This aggregates country-level confirmed cases into a single global figure.

Result

The total number of confirmed COVID-19 cases worldwide is 16,480,485.

Business Insight

A global total of over 16 million confirmed cases highlights the severity and widespread impact of the COVID-19 pandemic. This metric is critical for:

- **Assessing the global health burden**
- **Planning international medical aid and funding**
- **Supporting global policy decisions and preventive strategies**

```python
# What are the total confirmed COVID-19 cases worldwide?
c=data["Confirmed"].sum()
c
```

```
16480485
```

◆ Problem: What is the total number of deaths and recoveries globally?

Objective

To measure the global impact of COVID-19 by identifying the total number of deaths and recoveries recorded worldwide.

Methodology

- **The dataset columns Deaths and Recovered were selected.**
- **The .sum() function was applied to calculate global totals across all countries.**

Results

Metric	Total Count
Deaths	654,036
Recovered	9,468,087

Key Insights

- **The number of recovered cases (9.46 million) is significantly higher than the number of deaths (0.65 million).**
- **This shows that the majority of infected individuals recovered from the disease.**
- **The recovery rate indicates effective medical treatment and improved healthcare response globally.**

Business / Public Health Interpretation

- **High recovery numbers suggest strong healthcare system performance despite the scale of the pandemic.**
- **The data supports policy decisions focused on early detection, treatment capacity, and healthcare preparedness.**
- **Governments and health organizations can use this insight to plan resources for future outbreaks.**

```
# What is the total number of deaths and recoveries globally?
c=data[["Deaths","Recovered"]].sum()
c
```

```
Deaths        654036
Recovered    9468087
dtype: int64
```

◈ Problem: List all unique WHO regions in the dataset

Objective

To identify how many distinct World Health Organization (WHO) regions are represented in the COVID-19 dataset. This helps in understanding the geographical coverage of the data.

Methodology

- **The WHO Region column was analysed.**
- **The .unique() function was used to count the number of unique WHO regions present in the dataset.**

Result

- **The dataset contains 6 unique WHO regions.**

Key Insight

- **The presence of six WHO regions confirms that the dataset covers global COVID-19 data, spanning multiple continents and health jurisdictions.**
- **This ensures that subsequent regional analysis (such as active cases, deaths, and recoveries by WHO region) is comprehensive and reliable.**

```
# List all unique WHO regions in the dataset.
d=data["WHO Region"].nunique()
d
```

```
6
```

◈ Problem: Which country has the highest number of confirmed COVID-19 cases?

Objective

To identify the country most affected by COVID-19 in terms of total confirmed cases. This helps highlight regions with the highest disease burden and supports global risk assessment.

Methodology

- **The Confirmed cases column was analysed.**
- **The .imam() function was used to locate the row containing the maximum number of confirmed cases.**
- **The corresponding Country/Region and Confirmed values were extracted for reporting.**

Result

- **Country: United States (US)**
- **Total Confirmed Cases: 4,290,259**

Key Insight

- **The United States recorded the highest number of confirmed COVID-19 cases among all countries in the dataset.**
- **This indicates a significant spread of the virus, possibly influenced by factors such as population size, testing volume, and community transmission.**

Business / Analytical Importance

- **Identifying the most affected country helps:**
    - **Understand global hotspots**
    - **Guide international health responses**
    - **Prioritize resource allocation and policy intervention**
- **This insight is crucial for governments, healthcare organizations, and global agencies monitoring pandemic severity.**

```
# Which country has the highest number of confirmed cases?
d=data.loc[data['Confirmed'].idxmax(), ['Country/Region','Confirmed']]
d
```

```
Country/Region         US
Confirmed         4290259
Name: 173, dtype: object
```

◈ Problem: Which country has reported the maximum number of COVID-19 deaths?

Objective

To determine the country with the highest number of COVID-19 related deaths, helping assess the severity and fatal impact of the pandemic across nations.

Methodology

- **The Deaths column was examined for all countries.**
- **The .idxmax() function was used to identify the record with the maximum death count.**
- **The corresponding Country/Region and Deaths values were extracted.**

Result

- **Country: United States (US)**
- **Total Deaths: 148,011**

Key Insight

- **The United States reported the highest number of COVID-19 deaths in the dataset.**
- **This reflects the significant health impact of the pandemic and may be influenced by factors such as population size, healthcare capacity, and timing of interventions.**

Business / Analytical Importance

- **Mortality analysis helps:**

- Measure the severity of outbreaks
- Compare healthcare system effectiveness
- Support public health decision-making

- **High death counts indicate areas requiring policy review and medical support.**

```
# Which country has reported the maximum number of deaths?
c=data.loc[data["Deaths"].idxmax(),['Country/Region','Deaths']]
c
```

```
Country/Region        US
Deaths            148011
Name: 173, dtype: object
```

❖ Problem: Which country has the highest number of recovered COVID-19 cases?

Objective

To identify the country with the highest number of recovered COVID-19 patients, indicating the effectiveness of treatment, recovery capacity, and healthcare response.

Methodology

- **The Recovered column was analyzed for all countries.**
- **The .idxmax() function was used to locate the country with the maximum recovery count.**
- **The corresponding Country/Region and Recovered values were extracted.**

Result

- **Country: Brazil**
- **Total Recovered Cases: 1,846,641**

Key Insight

- **Brazil recorded the highest number of recoveries among all countries in the dataset.**
- **This suggests a large recovered population, which may reflect strong recovery tracking, treatment capacity, or high total case volume.**

Business / Analytical Importance

- **Recovery analysis helps:**
  - **Evaluate healthcare effectiveness**
  - **Understand disease management success**

    - **Compare pandemic control outcomes across countries**

- **High recovery numbers are crucial for assessing health system resilience.**

```
# Which country has the highest number of recovered cases?
data.loc[data["Recovered"].idxmax(),['Country/Region','Recovered']]
```

```
Country/Region     Brazil
Recovered         1846641
Name: 23, dtype: object
```

◈ Problem: Which country currently has the highest number of active COVID-19 cases?

Objective

To identify the country with the highest number of active COVID-19 cases, helping assess the current burden on healthcare systems and the need for immediate medical intervention.

Methodology

- **The Active column was analyzed for all countries.**
- **The .idxmax() function was used to find the country with the maximum active cases.**
- **The corresponding Country/Region and Active values were extracted.**

Result

- **Country: United States (US)**
- **Active Cases: 2,816,444**

Key Insight

- **The United States has the largest number of ongoing active cases, indicating a high current transmission level.**
- **A high active case count suggests sustained pressure on:**
    - **Hospitals**
    - **ICU capacity**
    - **Healthcare workforce**

Business / Policy Importance

- **Active cases represent current demand on healthcare infrastructure.**
- **This analysis helps:**

- **Allocate medical resources**
- **Plan emergency healthcare responses**
- **Monitor real-time outbreak severity**

```
# Which country currently has the highest active cases?
data.loc[data["Active"].idxmax(),["Country/Region","Active"]]
```

```
Country/Region         US
Active            2816444
Name: 173, dtype: object
```

◈ Visualization: Top 10 Countries by Confirmed COVID-19 Cases

Objective

To visually identify and compare the **top 10 countries with the highest confirmed COVID-19 cases**, enabling quick assessment of which countries were most severely affected.

Visualization Description

- A **bar chart** is used to rank countries based on confirmed case counts.
- Countries are sorted in **descending order** of confirmed cases.
- Each bar represents a country, and its height corresponds to the number of confirmed cases.

Why Bar Chart?

- Bar charts are ideal for **ranking and comparison**.
- They clearly show differences in magnitude across countries.
- Easy for policymakers and stakeholders to interpret at a glance.

Design & Styling Features

- **Distinct colors** for each bar to improve visual differentiation.
- **Black edge borders** and increased linewidth for clarity.
- **Transparency (alpha)** for better aesthetics.
- **Rotated x-axis labels** to avoid overlap.
- **Horizontal grid lines** to improve value readability.

- Proper **titles and axis labels** for professional presentation.

Key Insight

- The chart highlights that a small number of countries account for a **significant share of global confirmed cases**.
- Countries like the **US, Brazil, and India** dominate the top positions, indicating high transmission levels and healthcare burden.

Business / Policy Importance

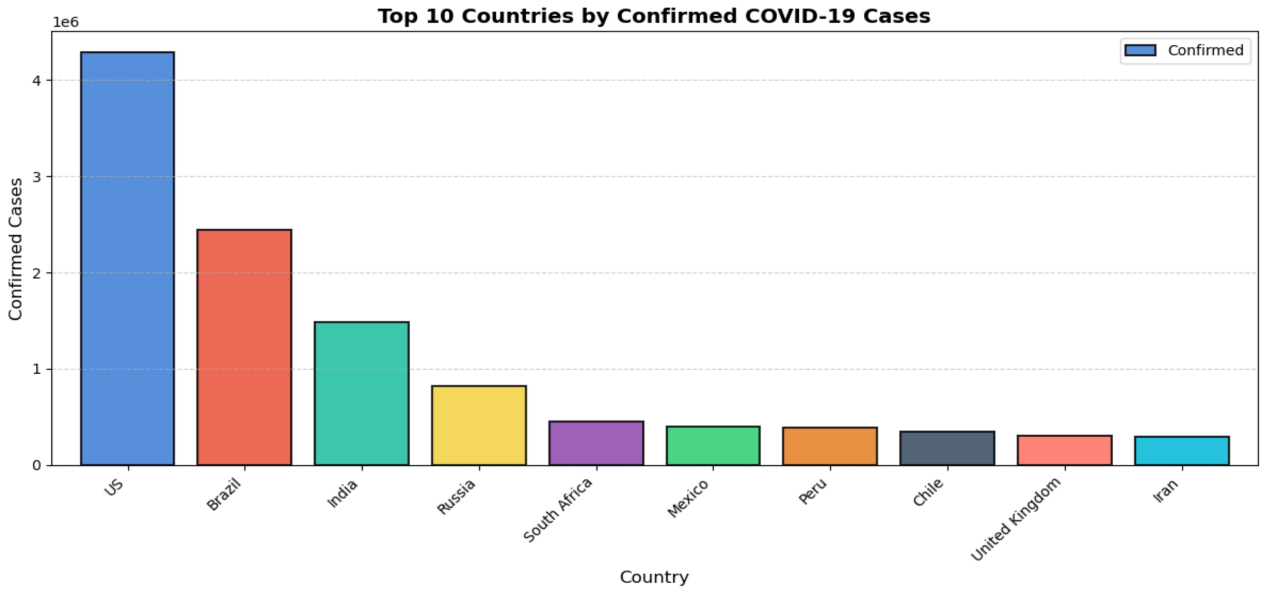
- Helps governments and global health organizations:
    - Prioritize **medical resource allocation**
    - Plan **containment strategies**
    - Compare national outbreak severity
- Useful for **high-level decision-making dashboards**.

Conclusion

This visualization provides a clear and impactful comparison of the most affected countries, making it a crucial component of the COVID-19 analytical dashboard.

```python
import matplotlib.pyplot as plt
c = data.sort_values("Confirmed", ascending=False).head(10)
c.plot.bar(
    x="Country/Region",
    y="Confirmed",
    figsize=(12,6),
    color=["#3A7BD5", "#E94F37", "#1ABC9C", "#F4D03F", "#8E44AD", "#2ECC71",
           "#E67E22", "#34495E", "#FF6F61", "#00B8D9"],
    edgecolor="black",
    linewidth=1.5,
    alpha=0.85,
    width=0.8
)
plt.title("Top 10 Countries by Confirmed COVID-19 Cases", fontsize=14, fontweight="bold")
plt.xlabel("Country", fontsize=12)
plt.ylabel("Confirmed Cases", fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.grid(axis="y", linestyle="--", alpha=0.6)
plt.tight_layout()
plt.show()
```

Output:



◈ Visualization: Deaths vs Recovered COVID-19 Cases

Objective

To analyze the **relationship between total deaths and total recovered COVID-19 cases** across countries, helping evaluate **healthcare effectiveness and recovery outcomes**.

Visualization Description

- A **scatter plot** is used where:
  - Each point represents a **country**.
  - The **x-axis** shows total deaths.
  - The **y-axis** shows total recovered cases.

Why Scatter Plot?

- Scatter plots are ideal for identifying:
  - **Correlation or trends** between two numerical variables
  - **Outliers** (countries with unusually high deaths or recoveries)
- Helps understand whether higher death counts are associated with higher recovery counts.

Design & Styling Features

- **Red markers** with partial transparency (alpha) to reduce overlap.
- **Black edges** around points for better visibility.
- **Grid lines** to enhance value interpretation.
- Clear **titles and axis labels** for professional presentation.

Key Insight

- Countries with higher death counts generally also report higher recoveries, indicating **large outbreak sizes**.
- Some countries appear as **outliers**, reflecting either:
  - Strong recovery systems
  - High fatality pressure
- This variation highlights differences in **healthcare response and reporting practices**.

Business / Policy Importance

- Helps policymakers:
  - Benchmark **healthcare performance**
  - Identify countries with **disproportionately high deaths**
  - Learn from regions with **strong recovery outcomes**
- Useful for global health organizations to **compare response effectiveness**.

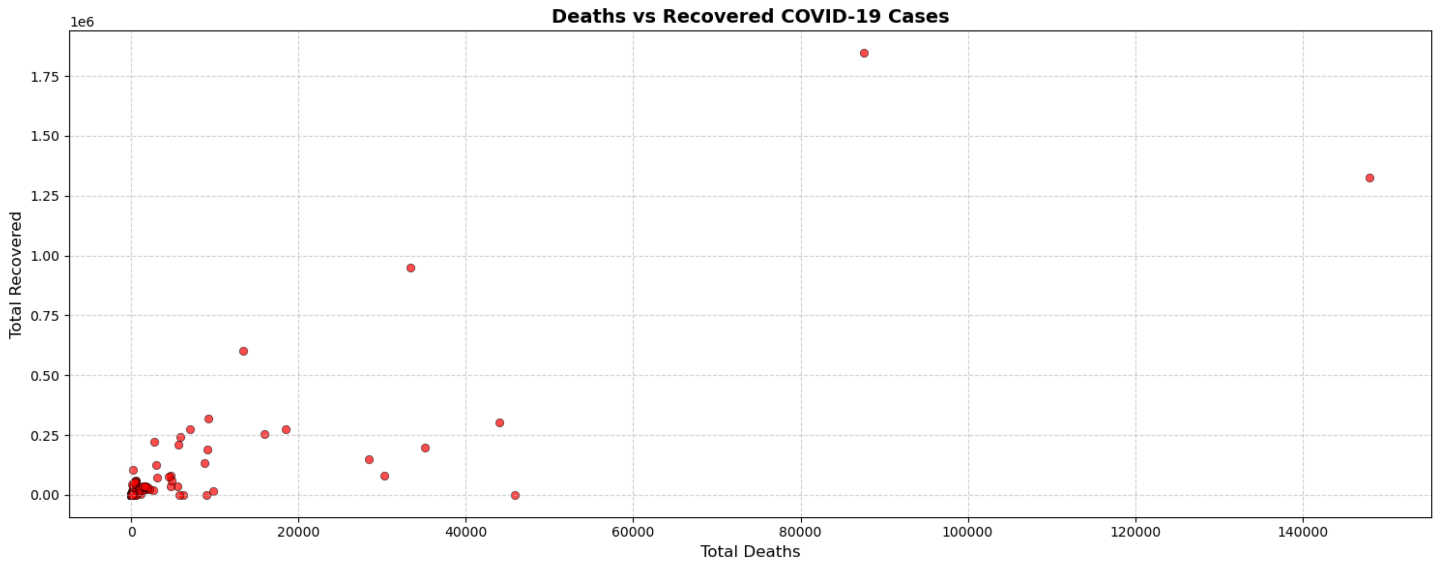
Conclusion

The scatter plot reveals a **positive relationship between deaths and recoveries**, emphasizing that countries with large outbreaks experience both higher recoveries and fatalities, underlining the importance of healthcare capacity and timely intervention.

```
c = data

plt.figure(figsize=(8,6))
plt.scatter(
    c["Deaths"],
    c["Recovered"],
    alpha=0.7,
    edgecolor="black",
    linewidth=0.5,
    c="red"
)
plt.title("Deaths vs Recovered COVID-19 Cases", fontsize=14, fontweight="bold")
plt.xlabel("Total Deaths", fontsize=12)
plt.ylabel("Total Recovered", fontsize=12)
plt.grid(linestyle="--", alpha=0.6)
plt.tight_layout()
plt.show()
```

Output:



🔷 Problem 3: Which WHO regions have the highest active COVID-19 cases?

Objective

To identify **WHO regions with the highest number of active COVID-19 cases**, helping global health authorities understand **regional outbreak severity** and prioritize **medical resource allocation**.

Methodology

- The dataset was grouped by **WHO Region**.
- Total **active cases** were calculated using the sum() function.
- A **bar chart** was created to compare active cases across regions.

Visualization Description

- **X-axis:** WHO Regions
- **Y-axis:** Total Active COVID-19 Cases
- Each bar represents the cumulative active cases for a WHO region.

Why Bar Chart?

- Bar charts are ideal for **comparing aggregate values** across categories.
- They allow quick identification of **regions with the highest healthcare burden**.

Design & Styling Features

- **Distinct colors** for each WHO region to improve clarity.
- **Black edge borders** to clearly separate bars.
- **Grid lines** on the y-axis for easier value comparison.
- **Rotated labels** for better readability.
- Professional **titles and axis labels** for dashboard-ready visuals.

Key Insight

- Certain regions, particularly the **Americas and Europe**, show significantly higher active case counts.
- These regions are experiencing greater **ongoing transmission** and **healthcare pressure**.

Business / Policy Importance

- Helps WHO and governments:
  - Prioritize **funding and emergency response**
  - Deploy **medical staff and equipment**
  - Monitor **regional outbreak intensity**
- Supports data-driven global health planning.

Conclusion

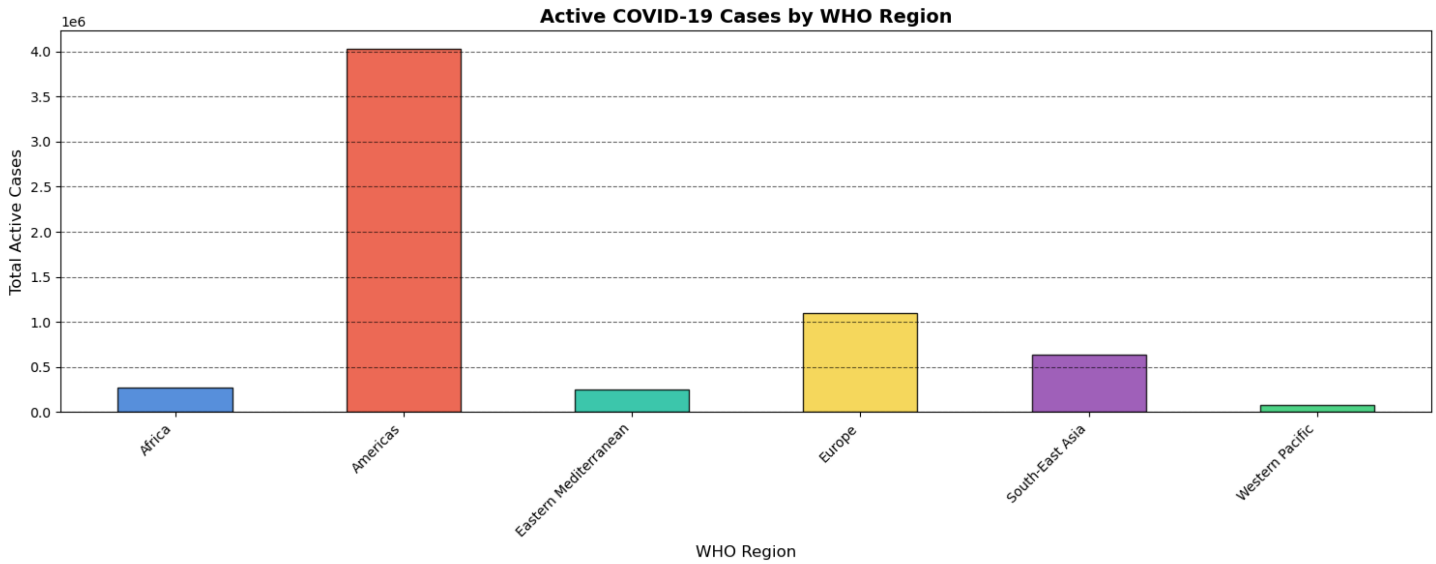
This analysis reveals that **active COVID-19 cases are unevenly distributed across WHO regions**, with some regions facing a substantially higher ongoing burden, emphasizing the need for **targeted intervention strategies**.

```python
# 🔹 Problem 3: Which WHO regions have the highest active cases?

active = data.groupby("WHO Region")["Active"].sum()
active.plot.bar(
    figsize=(10,6),
    color=["#3A7BD5", "#E94F37", "#1ABC9C", "#F4D03F", "#8E44AD", "#2ECC71"],
    edgecolor="black",
    alpha=0.85
)

plt.title("Active COVID-19 Cases by WHO Region", fontsize=14, fontweight="bold")
plt.xlabel("WHO Region", fontsize=12)
plt.ylabel("Total Active Cases", fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.grid(axis="y", linestyle="--", alpha=0.6,color="black")
plt.tight_layout()
plt.show()
```

Output:



◈ Problem 4: Distribution of Recovery Rates Across Countries

Objective

To analyze how COVID-19 recovery rates (%) are distributed across different countries, helping assess overall healthcare effectiveness and patient outcomes globally.

Methodology

- A new metric, Recovery Rate (%), was calculated using the formula:

$$\text{Recovery Rate} = \frac{\text{Recovered}}{\text{Confirmed}} \times 100$$

- This rate was computed for each country.
- A Kernel Density Estimation (KDE) plot was used to visualize the distribution.

Visualization Description

- X-axis: Recovery Rate (%)
- Y-axis: Density (probability distribution)
- The KDE curve represents how recovery rates are spread across countries.

Why KDE Plot?

- KDE plots are effective for:

  - Understanding the overall distribution shape
  - Identifying common recovery rate ranges
  - Detecting skewness or concentration
- Smoother and more informative than a simple histogram for continuous data.

Design & Styling Features

- Filled KDE curve for better visual emphasis.
- Green color to represent recovery and positivity.
- Increased line width for clarity.
- Large figure size for better readability.

Key Insight

- Most countries cluster around a moderate to high recovery rate range, indicating generally effective treatment outcomes.
- A smaller number of countries show lower recovery rates, which may point to:
  - Limited healthcare infrastructure
  - Reporting delays
  - Ongoing active cases not yet resolved

Business / Policy Importance

- Helps global health bodies:
  - Compare healthcare performance across countries
  - Identify regions needing medical support
  - Track progress of pandemic control measures
- Useful for evaluating treatment success trends rather than raw case counts.

Conclusion

The KDE analysis shows that recovery rates vary significantly across countries, but a majority achieve reasonable recovery levels, reflecting global improvements in COVID-19 treatment and management.

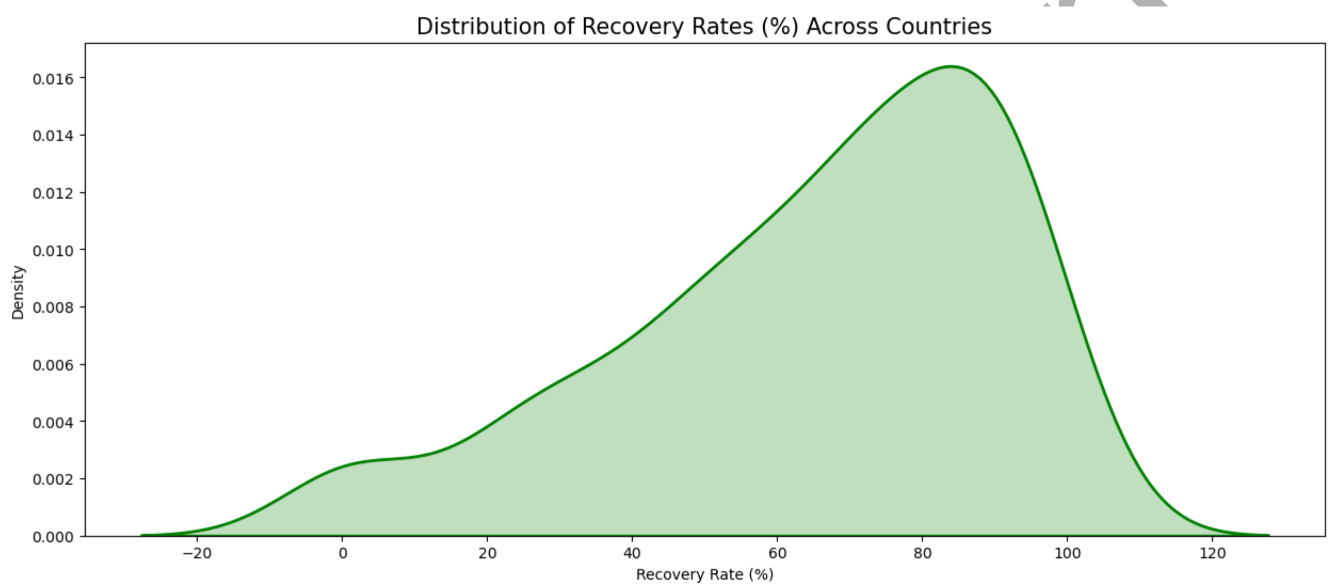
```python
#  ◆  Problem 4: Distribution of Recovery Rates across Countries

data['Recovery Rate'] = (data['Recovered'] / data['Confirmed']) * 100

plt.figure(figsize=(15, 6))
sns.kdeplot(data['Recovery Rate'], fill=True, color="Green",linewidth=2)

plt.title("Distribution of Recovery Rates (%) Across Countries", fontsize=15)
plt.xlabel("Recovery Rate (%)")
plt.ylabel("Density")
plt.show()
```

Output:



❖ Problem 5: Proportional Share of Total COVID-19 Deaths by WHO Region

Objective

To understand how **COVID-19 related deaths are distributed across WHO regions**, highlighting the **regional contribution to global mortality**.

Methodology

- The dataset was grouped by **WHO Region**.
- Total deaths were calculated using the sum() function.
- A **pie chart** was used to visualize each region's proportional contribution to global deaths.

Visualization Description

- Each slice of the pie represents a **WHO region**.

- The **size of the slice** corresponds to the region's share of total deaths.
- Percentage values are displayed directly on the chart.

Why Pie Chart?

- Pie charts are ideal for showing **proportional distribution**.
- They help answer questions about **relative contribution** rather than absolute numbers.
- Easy to interpret for non-technical stakeholders.

Design & Styling Features

- **Distinct color palette (Set3)** for visual clarity.
- **Exploded slices** to emphasize each region's contribution.
- **Percentage labels** displayed using autopct.
- Balanced start angle for better readability.

Key Insight

- A small number of WHO regions account for a **large proportion of total COVID-19 deaths**.
- Regions such as the **Americas and Europe** contribute significantly to global mortality figures.

Business / Policy Importance

- Helps policymakers:
  - Identify regions requiring **urgent healthcare intervention**
  - Prioritize **funding and international support**
  - Understand the **global mortality burden**
- Supports strategic planning by global health organizations like WHO.

Conclusion

The pie chart clearly shows that **COVID-19 deaths are unevenly distributed across WHO regions**, emphasizing the need for **region-specific health strategies and targeted response measures**.

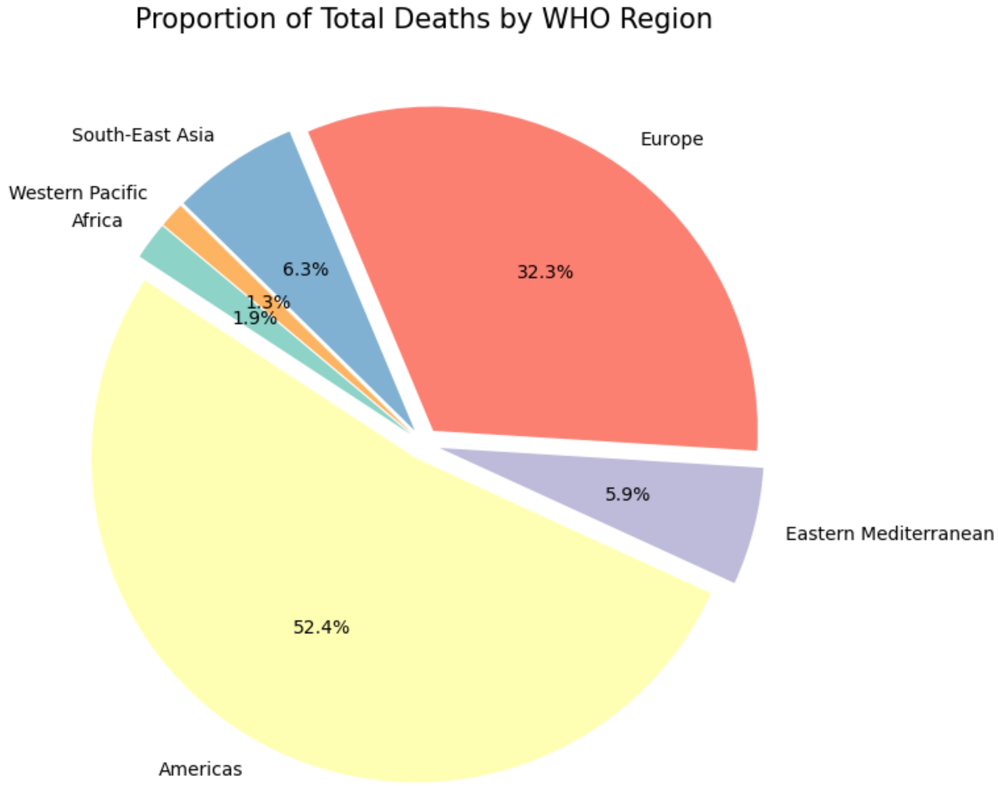
```python
# 🔹 Problem 5: Proportional Share of Total Deaths by Region

death_counts = data.groupby("WHO Region")["Deaths"].sum()

plt.figure(figsize=(8, 8))
plt.pie(death_counts, labels=death_counts.index, autopct='%1.1f%%',
        startangle=140, colors=sns.color_palette("Set3"), explode=[0.05]*6)

plt.title("Proportion of Total Deaths by WHO Region", fontsize=15)
plt.show()
```

Output:



Objective

To analyze the distribution of countries across different WHO regions and understand regional representation within the global health framework.

Methodology

- The dataset was analyzed based on the **WHO Region** column.
- The number of countries in each WHO region was calculated using frequency counts.
- A **count plot** was used to visualize how many countries belong to each region.

Visualization Description

- Each bar represents a WHO region.
- The height of the bar indicates the number of countries in that region.
- Regions are ordered based on the frequency of countries for clear comparison.

Why Count Plot?

- Count plots are effective for visualizing categorical frequency data.
- They allow easy comparison between groups.
- Ideal for showing how data points are distributed across categories.

Design & Styling Features

- Magma color palette used for better contrast and visual appeal.
- Regions sorted in descending order for improved readability.
- Clear axis labels and title enhance interpretability.
- Large figure size ensures labels are clearly visible.

Key Insight

- WHO regions vary significantly in the number of countries they include.
- Some regions have a much higher number of countries, indicating broader geographic coverage.
- Regions with fewer countries may have more homogeneous health data patterns.

Business / Policy Importance

- Helps global health organizations understand regional representation.
- Supports fair comparison of COVID-19 statistics across regions.
- Assists policymakers in allocating resources proportionally based on regional size.
- Important for interpreting regional trends in cases, deaths, and recoveries.

Conclusion

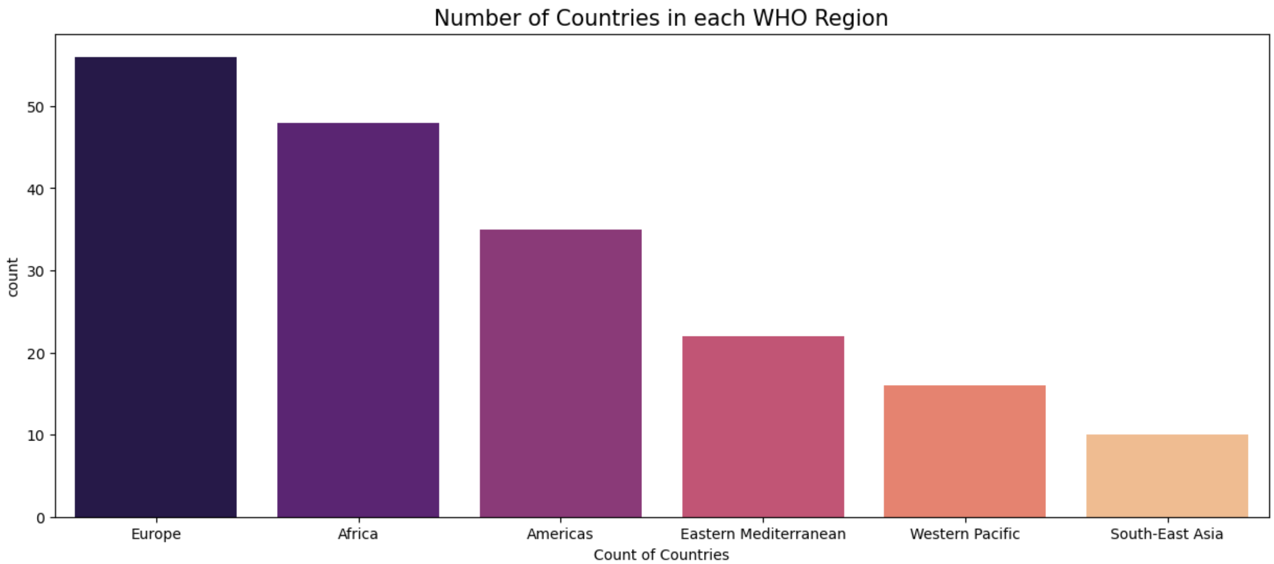
The count plot highlights the uneven distribution of countries across WHO regions. Understanding this regional representation is crucial for accurate global health analysis, fair comparison of pandemic impact, and effective international policy planning.

```
# 🔹 Problem 6: Frequency of Countries per WHO Region

plt.figure(figsize=(15, 6))
sns.countplot(x="WHO Region", data=data, palette="magma", order=data['WHO Region'].value_counts().index)

plt.title("Number of Countries in each WHO Region", fontsize=15)
plt.xlabel("Count of Countries")
plt.show()
```

Output:



Objective

To identify countries with the highest COVID-19 mortality risk by analyzing death rates relative to confirmed cases.

Methodology

- A **Death Rate (%)** was calculated for each country using the formula: *Deaths ÷ Confirmed Cases × 100*.
- Countries were sorted in descending order based on death rate.
- The **top 10 countries** with the highest mortality rates were selected for analysis.
- A bar chart was used to visualize the comparison.

Visualization Description

- Each bar represents a country.
- The height of the bar indicates the death rate (percentage of deaths per 100 confirmed cases).
- Countries are displayed in descending order of mortality risk.

Why Bar Plot?

- Bar plots are effective for comparing values across categories.

- They clearly highlight differences in death rates between countries.
- Ideal for ranking and identifying extreme values (high-risk countries).

Design & Styling Features

- Red color gradient (Reds_r) used to symbolize severity and risk.
- Black edge borders added to bars for clear separation.
- Horizontal layout improves label readability for country names.
- Adequate figure size ensures clarity and visual balance.

Key Insight

- A small group of countries shows significantly higher death rates compared to others.
- High mortality risk may indicate weaker healthcare infrastructure, delayed response, or under-reporting of cases.
- Death rate alone does not reflect total impact but highlights **severity among confirmed cases**.

Business / Policy Importance

- Helps governments identify countries needing urgent healthcare support.
- Supports international organizations in prioritizing medical aid and resources.
- Assists researchers in investigating causes of high mortality (testing rates, demographics, healthcare access).
- Useful for risk assessment and preparedness planning.

Conclusion

The bar chart reveals that COVID-19 mortality risk is unevenly distributed across countries. Identifying high-risk countries is critical for targeted intervention, policy planning, and strengthening global health response systems.

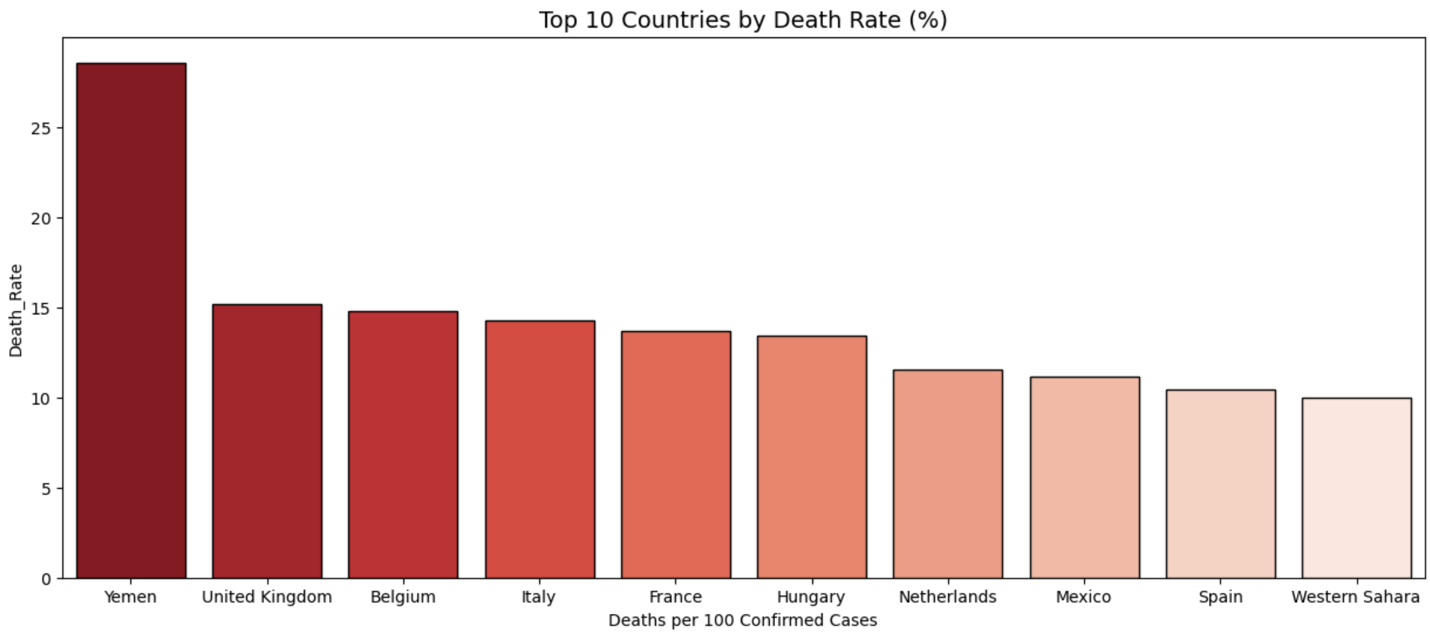
```
# 🔹 Problem 7: Which countries have the highest mortality risk?

data['Death_Rate'] = (data['Deaths'] / data['Confirmed']) * 100
top_death_rate = data.sort_values('Death_Rate', ascending=False).head(10)

plt.figure(figsize=(15, 6))
sns.barplot(y="Death_Rate", x="Country/Region", data=top_death_rate,
            palette="Reds_r", edgecolor="black")

plt.title("Top 10 Countries by Death Rate (%)", fontsize=14)
plt.xlabel("Deaths per 100 Confirmed Cases")
plt.show()
```

Output:



Objective

To analyze the relationship between **recovered cases and deaths** across countries and understand how disease outcomes vary with total confirmed cases.

Methodology

- Recovered cases were plotted on the X-axis and deaths on the Y-axis.
- Bubble size was scaled using total confirmed cases to reflect outbreak magnitude.
- Color intensity also represented confirmed cases, adding an extra data dimension.
- A scatter plot was used to capture correlation and spread.

Visualization Description

- Each point represents a country.
- The horizontal position shows the number of recovered cases.
- The vertical position shows the number of deaths.
- Larger and darker-colored points indicate countries with higher confirmed cases.
- A color bar provides a reference for confirmed case intensity.

Why Scatter Plot?

- Scatter plots are ideal for identifying relationships between two numerical variables.
- They help detect trends, clusters, and outliers.
- Bubble size and color enhance multi-dimensional analysis in a single view.

Design & Styling Features

- Rocket color map used for strong visual contrast and severity indication.
- Semi-transparency (alpha) applied to reduce overplotting.
- Bubble sizes scaled proportionally to confirmed cases for impact emphasis.
- Clear axis labels and title improve interpretability.
- Color bar added to explain the confirmed case scale.

Key Insight

- Countries with higher recoveries generally also report higher deaths, indicating a scale effect of large outbreaks.
- Some countries show high recoveries with comparatively lower deaths, suggesting effective healthcare response.
- A few outliers show disproportionate deaths relative to recoveries, indicating higher fatality risk.

Business / Policy Importance

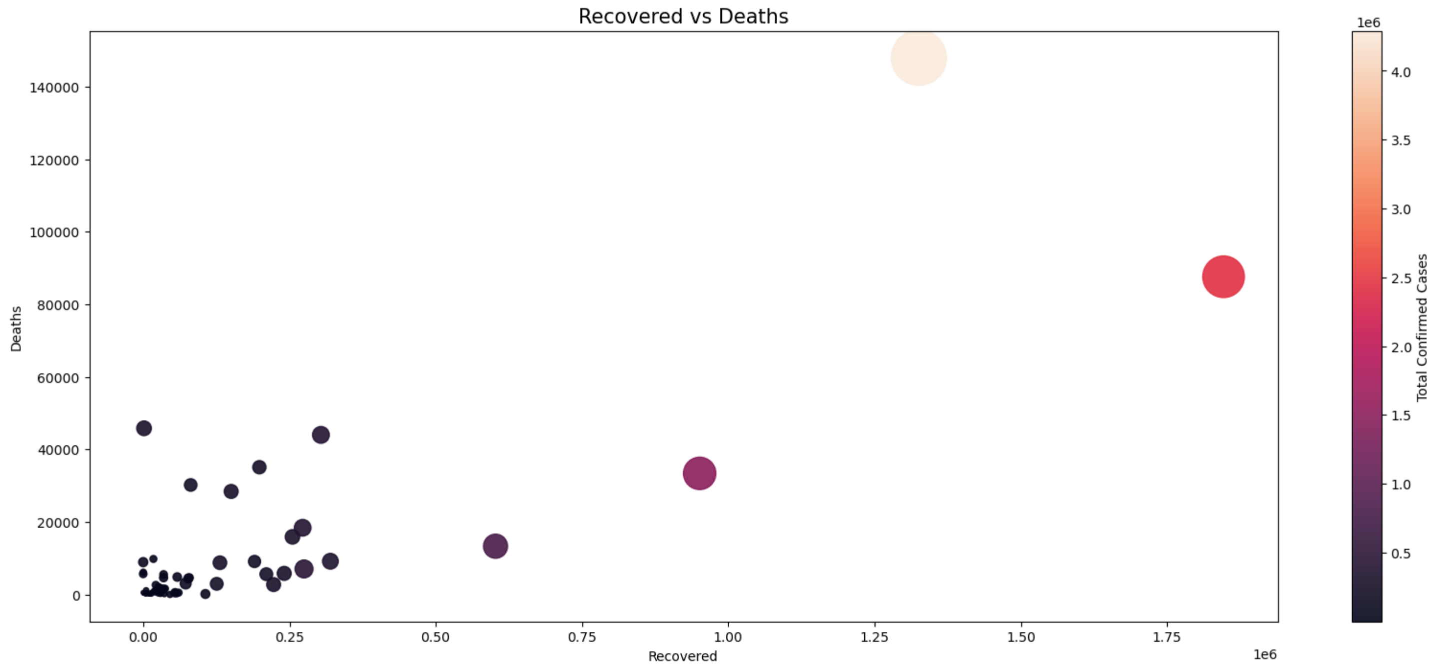
- Helps health authorities evaluate treatment effectiveness across regions.
- Assists policymakers in identifying countries with high fatality inefficiencies.
- Supports healthcare planning by correlating recovery success with outbreak size.
- Useful for global health comparisons and strategic interventions.

Conclusion

The scatter plot demonstrates a strong relationship between recovered cases and deaths, influenced heavily by total confirmed cases. The visualization highlights disparities in health outcomes and underscores the importance of healthcare capacity and response efficiency in managing pandemics.

```
plt.figure(figsize=(12, 8))
plt.scatter(data['Recovered'], data['Deaths'], s=data['Confirmed']*0.0004,
            alpha=0.9, c=data['Confirmed'], cmap='rocket')
plt.title("Recovered vs Deaths", fontsize=15)
plt.xlabel("Recovered")
plt.ylabel("Deaths")
plt.colorbar(label='Total Confirmed Cases')
plt.show()
```

Output:



Objective

To provide a **comprehensive business-level overview of the COVID-19 situation** by combining multiple analytical perspectives into a single dashboard for strategic decision-making.

Methodology

- Multiple visualizations were combined into a **3×2 dashboard layout**.
- Country-level and WHO region-level aggregations were performed.
- Metrics analyzed include:
  - Confirmed cases
  - Deaths
  - Recoveries
  - Recovery rates
- Different chart types were used to highlight trends, proportions, relationships, and distributions.

Visualization-wise Description

1. Top 10 Most Affected Countries

- Displays countries with the highest confirmed cases.
- Bar height represents total confirmed cases.
- Highlights countries facing the maximum outbreak severity.

2. Confirmed Case Share by WHO Region

- Pie chart shows regional contribution to global confirmed cases.
- Each slice represents a WHO region's percentage share.
- Useful for regional comparison at a global level.

3. Relationship: Confirmed vs Deaths

- Scatter plot showing correlation between confirmed cases and deaths.
- Color intensity represents recovered cases.
- Helps identify how mortality scales with outbreak size.

4. Density Distribution of Global Deaths

- KDE plot shows how deaths are distributed across countries.
- Indicates whether deaths are concentrated in a few regions or spread globally.

5. Frequency of Countries per WHO Region

- Count plot showing the number of countries in each WHO region.
- Helps normalize interpretation of regional data.

6. Average Recovery Rate (%) by Region

- Line plot representing average recovery rate per WHO region.
- Shows healthcare effectiveness and patient outcomes across regions.

Why Dashboard Approach?

- Dashboards allow **multi-dimensional insights in a single view**.

- Enables faster decision-making for business leaders and policymakers.
- Ideal for presentations, executive summaries, and reports.
- Reduces the need to analyze individual charts separately.

Design & Styling Features

- Large figure size for clarity and readability.
- Distinct color palettes for visual separation of insights.
- Gridlines added for easier value comparison.
- Rotated axis labels to avoid overlap.
- Clear titles and subtitles for each visualization.

Key Insights

- A small number of countries dominate global confirmed cases.
- Certain WHO regions contribute disproportionately to total cases.
- Higher confirmed cases generally lead to higher deaths, showing scale dependency.
- Deaths are not uniformly distributed; a few countries account for most fatalities.
- Recovery rates vary significantly across WHO regions, reflecting healthcare capacity differences.

Business / Policy Importance

- Helps governments prioritize **high-risk countries and regions**.
- Assists health organizations in **resource allocation and intervention planning**.
- Enables comparative assessment of **healthcare performance** across regions.
- Supports data-driven decision-making for pandemic preparedness.

Conclusion

The dashboard delivers a **holistic, business-oriented view of the COVID-19 pandemic**, integrating severity, spread, mortality, and recovery performance. By combining multiple analytical dimensions, it provides actionable insights that support strategic planning, healthcare policy formulation, and global response coordination.

Code:

```
import matplotlib.pyplot as plt

import seaborn as sns



plt.figure(figsize=(18, 15))

plt.suptitle("Comprehensive COVID-19 Business Analytics Dashboard", fontsize=22, fontweight='bold')



plt.subplot(3, 2, 1)

top_10 = data.sort_values("Confirmed", ascending=False).head(10)

plt.bar(top_10["Country/Region"], top_10["Confirmed"], color=["#3A7BD5", "#E94F37", "#1ABC9C", "#F4D03F", "#8E44AD", "#2ECC71",

        "#E67E22", "#34495E", "#FF6F61", "#00B8D9"], edgecolor="black")

plt.title("1. Top 10 Most Affected Countries", fontweight='bold')

plt.xticks(rotation=45)

plt.grid(axis='y', linestyle='--', alpha=0.7)



plt.subplot(3, 2, 2)

region_data = data.groupby("WHO Region")["Confirmed"].sum()

plt.pie(region_data, labels=region_data.index, autopct="%1.1f%%", startangle=140, colors=sns.color_palette("Set3"))

plt.title("2. Confirmed Case Share by WHO Region", fontweight='bold')



plt.subplot(3, 2, 3)

plt.scatter(data["Confirmed"], data["Deaths"], c=data["Recovered"], cmap="viridis", alpha=0.6)
```

```
plt.title("3. Relationship: Confirmed vs Deaths", fontweight='bold')

plt.xlabel("Confirmed")

plt.ylabel("Deaths")

plt.colorbar(label='Recovered Cases')

plt.grid(True, linestyle=':', alpha=0.6)


plt.subplot(3, 2, 4)

sns.kdeplot(data['Deaths'], fill=True, color="red", bw_adjust=0.5)

plt.title("4. Density Distribution of Global Deaths", fontweight='bold')

plt.xlabel("Number of Deaths")


plt.subplot(3, 2, 5)

sns.countplot(x="WHO Region", data=data, palette="magma", edgecolor="black")

plt.title("5. Frequency of Countries per WHO Region", fontweight='bold')



plt.subplot(3, 2, 6)

data['Recovery_Rate'] = (data['Recovered'] / data['Confirmed']) * 100

avg_recovery = data.groupby("WHO Region")['Recovery_Rate'].mean().sort_values()

plt.plot(avg_recovery.index, avg_recovery.values, marker='o', color='green', mfc='yellow', markersize=10, linewidth=2)

plt.title("6. Avg Recovery Rate % by Region", fontweight='bold')

plt.xticks(rotation=30)

plt.grid(True, color='gray', alpha=0.3)


plt.tight_layout()
```

```python
plt.show()
```

Output