**Exploratory-Data-Analysis(EDA)-COVID- 19 data**

To analyze and visualize COVID-19 data at a global level, highlighting key statistics such as the number of confirmed cases, recovery rates, death rates, and regional distributions. The goal is to provide insights into the spread and impact of the pandemic.

**--------------------------------------------------------------------------------------------**

**Goles of the Project-**

1. Explore the COVID-19 data set using Pandas
2. Real-World Data Projects
3. Building an Online Portfolio
4. Continuous Learning and Growth
5. Data Analysis and Visualization Mastery

**--------------------------------------------------------------------------------------------**

**Materials and Methods-**

The dataset consist of COVID-19 country-wise data including number of death ,recovered ,active cases and containing information about county ,weekly ratio of new founded cases Dataset also include multiple WHO Country and there ratio. To analysis or aims to understand a recovery rate ,Death rate ,Confirmed cases and there ratio

**Materials**

1. **Dataset:**
   * The dataset used is country\_wise\_latest.csv, containing COVID-19 statistics per country.
   * It includes columns like **confirmed cases, deaths, recovered cases, active cases, new cases, and WHO region**.
2. **Software & Libraries:**
   * **Python**: Used for data processing and analysis.
   * **Pandas**: For data manipulation and cleaning.
   * **NumPy**: For numerical operations.
   * **Matplotlib & Seaborn**: For data visualization.
   * **Warnings Module**: To suppress unnecessary warnings.

**Methods**

**1. Data Preprocessing**

* The dataset is read using pd.read\_csv().
* Basic information (df.info()), summary statistics (df.describe()), missing values (df.isnull().sum()), and unique values (df.nunique()) are checked.
* Duplicate rows are removed (df.drop\_duplicates()).

**2. Data Analysis**

* **New Deaths per Country**: Aggregated using groupby('Country/Region')['New deaths'].sum().
* **Top 10 Countries with Most Cases**: Extracted using df.nlargest(10, 'Confirmed').

**3. Data Visualization**

1. **Bar Chart**: Shows the top 10 countries with the highest confirmed cases.
2. **Scatter Plot**: Visualizes the relationship between death and recovery rates.
3. **Pie Chart**: Displays the distribution of confirmed cases across WHO regions.
4. **Histogram**: Analyzes the distribution of weekly percentage increase in cases.
5. **Stacked Bar Chart**: Compares COVID-19 metrics (confirmed, recovered, deaths) in the top 5 affected countries.

**Summary of COVID-19 Data Analysis**

**Dataset Overview**

* **Total Entries**: 187
* **Total Columns**: 15
* **No Missing Values**
* **No Duplicate Rows After Cleaning**
* **WHO Regions Covered**: 6

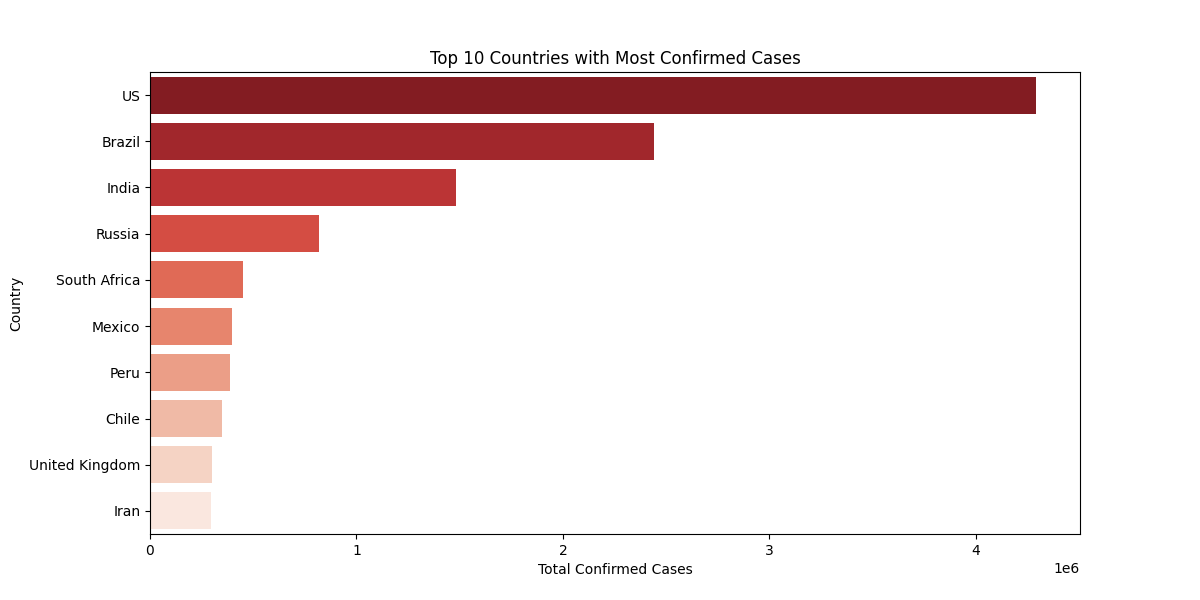
**Key Statistics**

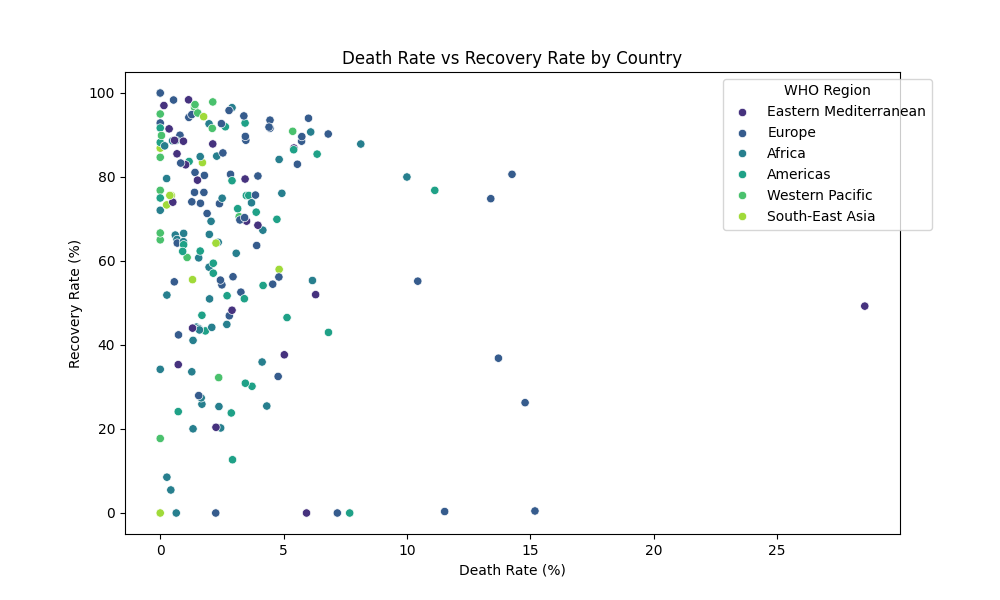
* **Confirmed Cases**: Ranges from **10** to **4.29 million**
* **Deaths**: Maximum **148,011** (US)
* **Recovery Rate**: Averages **64.8%**, with some countries reaching **100%**
* **Death Rate**: Averages **3.02%**, but varies significantly

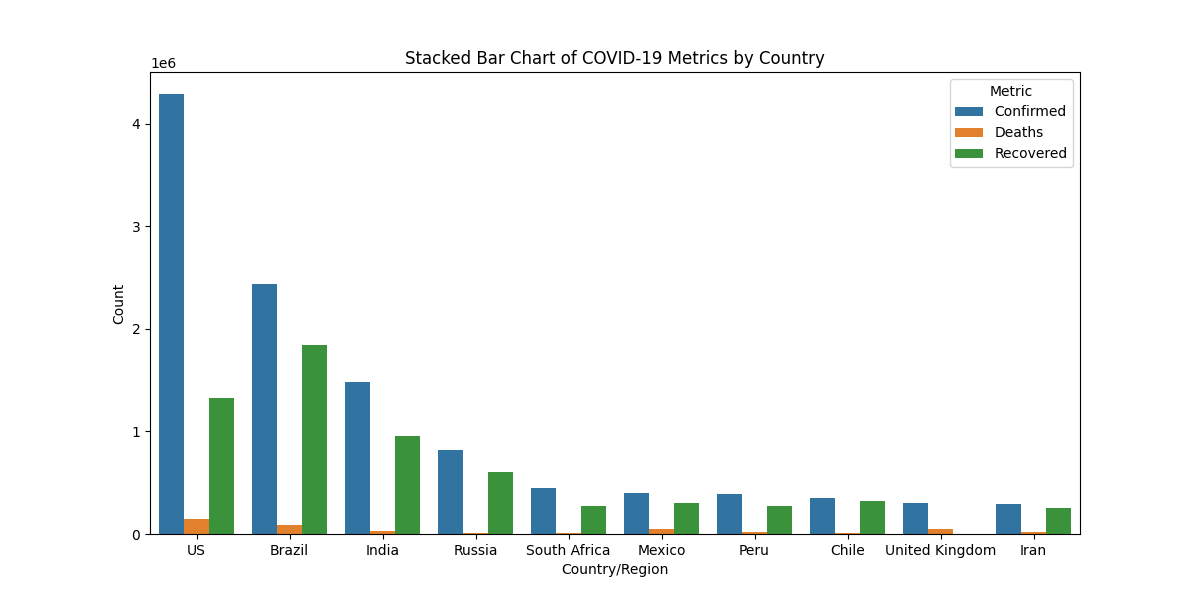
**Top 10 Countries with Most Deaths (New Deaths)**

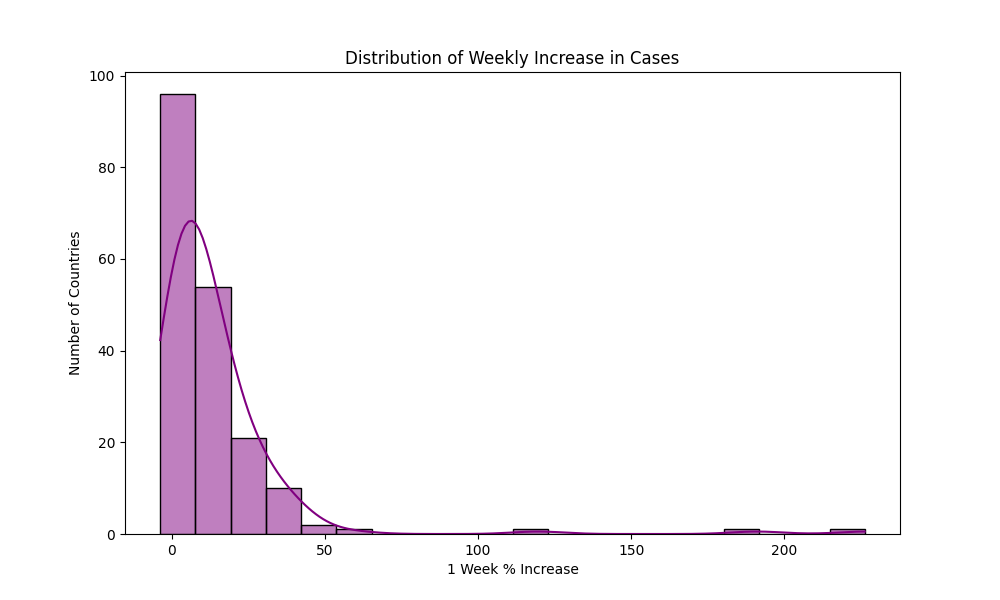
1. **US** - 1076
2. **India** - 637
3. **Brazil** - 614
4. **Peru** - 575
5. **Colombia** - 508
6. **Mexico** - 342
7. **South Africa** - 298
8. **Iran** - 212
9. **Argentina** – 120
10. **Iraq** - 96

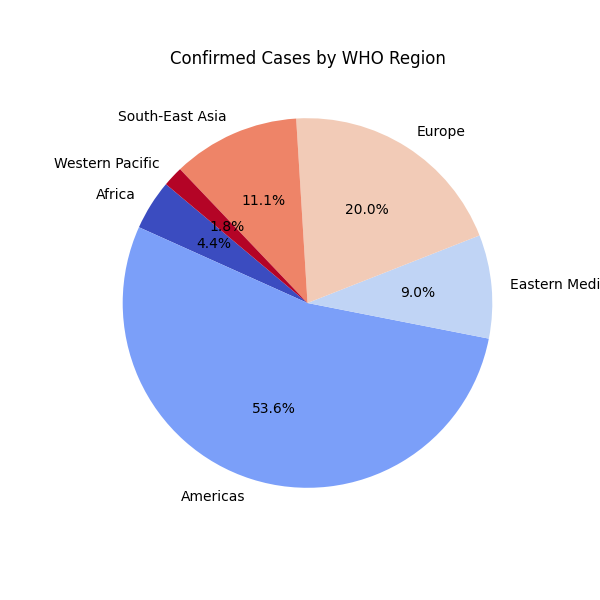
**Visualization-**











**Conclusion-**

Overall, the analysis and visualizations provided a clear picture of the COVID-19 pandemic's trends, impacts, and progression across different countries and regions. These insights can inform public health strategies, resource distribution, and further research to combat the pandemic effectively