A PROJECT REPORT

ON

"COVID-19 GLOBAL INSIGHT DASHBOARD"

Submitted By

Surya Kanta Barik

Dibyaprasad Sahu

Monalisha Sahoo

Abhishek Panda

Satyam Sahu

Under the Guidance of

Miss. Gitanjali Digal

CENTRAL TOOL ROOM & TRAINING CENTER

B-36, CHANDAKA INDUSTRIAL AREA, NEAR INFOCITY, BBSR,(O.D) 751024

Department of Data anlytics

CERTIFICATE

Certified that internship project Work entitled "**COVID-19 GLOBAL INSIGHT DASHBOARD**" is certificate course in DATA ANALYTICS from CENTRAL TOOL ROOM & TRAINING CENTER, Bhubaneswar, during the year 2025. It is certified that all the correction/suggestion indicated for internal assessment have been incorporated in the report. CTTC, Bhubaneswar (O.D).

CENTRAL TOOL ROOM & TRAINING CENTER

B-36, CHANDAKA INDUSTRIAL AREA, NEAR INFOCITY, BBSR,(O.D) 751024

Department of Data anlytics

DECLARATION

We hereby declare that the project titled "**COVID-19 Global Insight Dashboard**" using Excel, Tableau and Machine Learning is a result of our independent work carried out as part of our academic curriculum. This project has not been submitted previously to any other institution or university for any degree, diploma, or other qualifications.

The analysis, visualizations, and machine learning models presented in this project are based on the dataset provided and have been prepared using Microsoft Excel, Tableau, Power BI, and Python. Any external sources, references, or tools used in this project have been duly acknowledged.

We affirm that this project represents our honest effort and understanding, and it reflects our learning throughout the data analysis process.

CENTRAL TOOL ROOM & TRAINING CENTER

B-36, CHANDAKA INDUSTRIAL AREA, NEAR INFOCITY, BBSR,(O.D) 751024

Department of Data anlytics

ACKNOWLEDGEMENT

It is our proud privilege and duty to acknowledge the kind of help and guidance received from several people in preparation for this report. It would not have been possible to prepare this report in this form without their valuable help, cooperation, and guidance.

First and foremost, we wish to record our sincere gratitude to **Gitanjali Digal** mam for her constant support and encouragement in preparation of this report Our numerous discussions with her were extremely helpful. The project on "**COVID-19 Global Insight Dashboard**" was very helpful for us in giving the necessary background information and inspiration in choosing this Project. Our sincere thanks to Gitanjali Digal mam, for having supported the work related to this project. Their contributions and technical support in preparing this report are greatly acknowledged. WE appreciate the use of tools such as Microsoft Excel, Tableau and Python for Machine Learning, which were instrumental in drawing meaningful insights from the dataset.

Lastly, We extend our heartfelt thanks to our peers, friends, and family for their motivation, moral support, and patience during this journey.

ABSTRACT

The **COVID-19 Global Insight Dashboard** project is an analytical visualization initiative aimed at uncovering patterns, trends, and critical insights from global pandemic data. As COVID-19 swept across continents, generating a wealth of public health information, the need arose for a centralized, data-driven approach to understand its impact. This project utilizes a comprehensive dataset comprising global confirmed cases, deaths, recoveries, testing statistics, and vaccination records, enabling dynamic exploration and forecasting.

The dashboard integrates derived parameters such as *new confirmed cases*, *test-confirmed ratio*, and *survival ratio*, providing enriched indicators beyond raw data. Using tools like **Microsoft Excel** for preprocessing and **Tableau** for visualization, the project transforms static data into interactive visual stories. Key elements of the analysis include time-series trends of active and recovered cases, identification of pandemic peaks, regional comparisons of daily and cumulative cases, and correlation analysis of vaccine rollouts with declining infection rates.

One of the standout features of the dashboard is the **Multiple Linear Regression (MLR)** model employed to forecast active case trends based on historical data. This statistical modeling provides predictive insight into the trajectory of the pandemic, enabling health authorities to plan resources accordingly. Heatmaps further complement this by visually highlighting global hotspots based on selected criteria.

The insights drawn from the dashboard emphasize the importance of timely vaccination campaigns, robust testing infrastructures, and accurate reporting. Additionally, it underscores disparities between countries in handling the pandemic and offers a framework for future preparedness and response strategies.

Despite data limitations such as missing values or inconsistent country-wise reporting, the dashboard proves to be a powerful analytical tool. It empowers stakeholders—from government bodies to researchers and citizens—with actionable insights that support informed decision-making.

In conclusion, the COVID-19 Global Insight Dashboard is a valuable contribution to public health analytics. By bridging the gap between raw epidemiological data and practical intelligence, it enhances our collective ability to monitor, respond to, and learn from one of the most defining global health crises of the modern era.

INDEX

INTRODUCTION

In the modern educational landscape, data plays a vital role in understanding and improving student learning outcomes. Institutions increasingly rely on data-driven decision-making to enhance academic performance, reduce dropout rates, and personalize the learning experience.

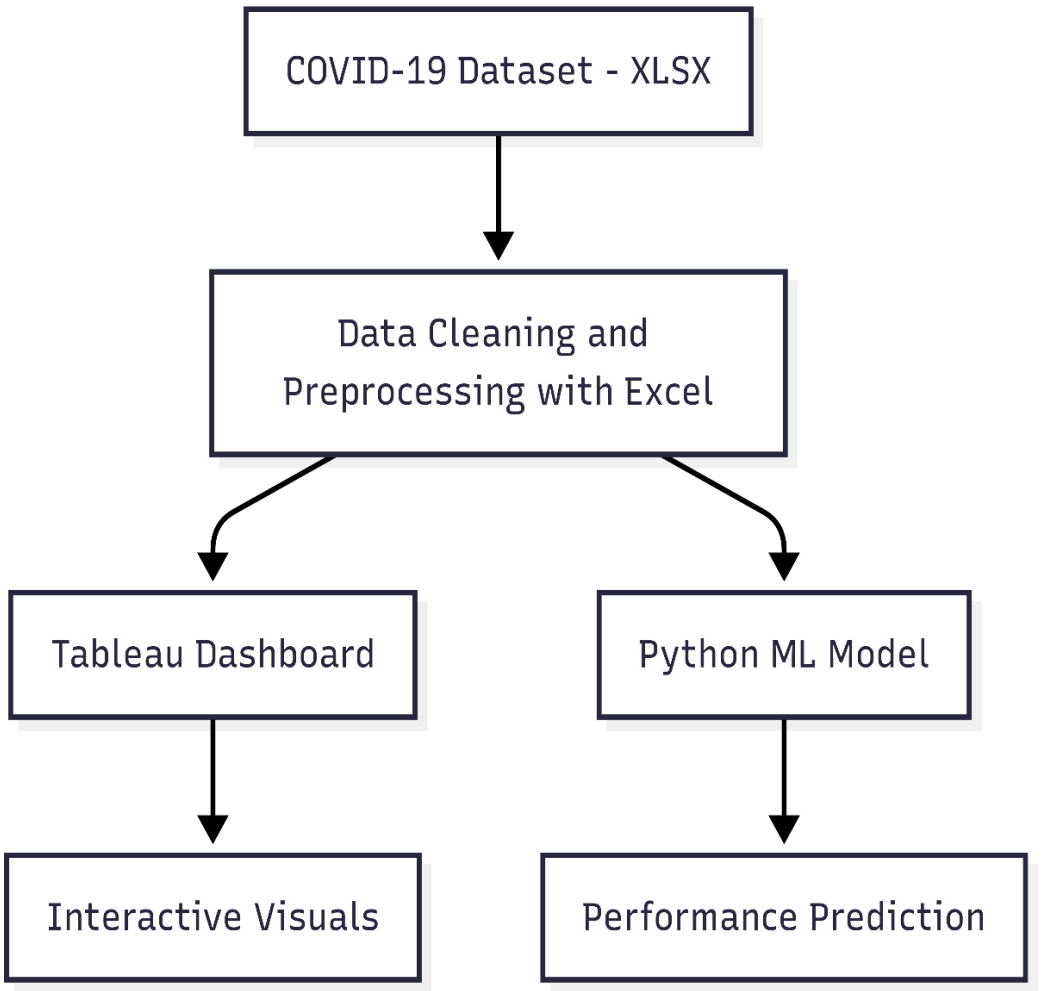
The COVID-19 pandemic has profoundly impacted public health, economies, and daily life across the globe. In the face of this unprecedented crisis, accurate data collection and effective analysis became essential tools for understanding the virus's spread and informing timely interventions. This project, titled **"COVID-19 Global Insight Dashboard,"** focuses on transforming complex global pandemic data into meaningful visual insights to support awareness, policy-making, and preparedness.

Leveraging a comprehensive dataset that includes daily confirmed cases, deaths, recoveries, testing rates, and vaccination statistics, the dashboard employs derived metrics such as *new confirmed cases*, *test-confirmed ratio*, and *survival ratio* to enhance interpretability. By using tools like **Microsoft Excel** for data processing and **Tableau** for visualization, the project provides dynamic charts, heat maps, and trend analyses.

A key highlight is the use of **Multiple Linear Regression (MLR)** to forecast future active cases, helping to anticipate healthcare demands. The dashboard also explores the effect of vaccine rollouts and booster doses on infection trends. Overall, this project offers a practical, data-driven approach to understanding the evolving nature of the COVID-19 pandemic at both national and global levels.

SYSTEM ARCHITECTURE

The Architecture of the proposed system is the design diagram which depicts the scope of the project with the whole system design. In architecture diagram, it highlights the modules with its various functions as a process. It aims to convey the internal design of the proposed system.



HARDWARE REQUIRMENT

To efficiently carry out the **COVID-19 Global Insight Dashboard** , a system with moderate to high performance is recommended. At a minimum, an Intel Core i5 or Ryzen 5 processor, 8 GB RAM, and 256 GB SSD or 500 GB HDD storage are required to handle data processing, visualization tools, and machine learning computations. A standard display with 1080p resolution or higher ensures clear viewing of dashboards and graphs.

For smoother multitasking—especially when working with large datasets in Tableau, or running machine learning models in Python—16 GB RAM and a faster processor are advisable. No dedicated graphics card is necessary unless performing GPU-based model training, making integrated graphics sufficient for this project.

Reliable internet connectivity is also important for accessing online features, downloading Python packages, and publishing dashboards if needed.

💻 Hardware Requirements:

Component	Specification
Processor (CPU)	Intel Core i5 / Ryzen 5 or higher
RAM	Minimum 8 GB (Recommended: 16 GB)
Storage	Minimum 256 GB SSD or 500 GB HDD
Graphics	Integrated graphics (for general use)
Display	13" or larger (1080p resolution or higher)
Internet	Required for Tableau online features and Python package installations

SOFTWARE REQUIRMENT

This project integrates multiple tools across different stages of data analysis, requiring a combination of spreadsheet software, data visualization platforms, and programming environments:

- Microsoft Excel is used for initial data cleaning, sorting, and basic analysis.
- Tableau are employed for building interactive dashboards and visualizing accident trends and severity.
- Python (version 3.8 or higher) is used for implementing machine learning models and data manipulation. Key libraries include pandas, scikit-learn, matplotlib, and seaborn.
- Jupyter Notebook (or any Python IDE like VS Code or PyCharm) is recommended for running ML code interactively.
- Power BI Desktop and Tableau Public/Desktop are required for dashboard creation and publishing.
- Operating System: Compatible with both Windows 10/11 and macOS.

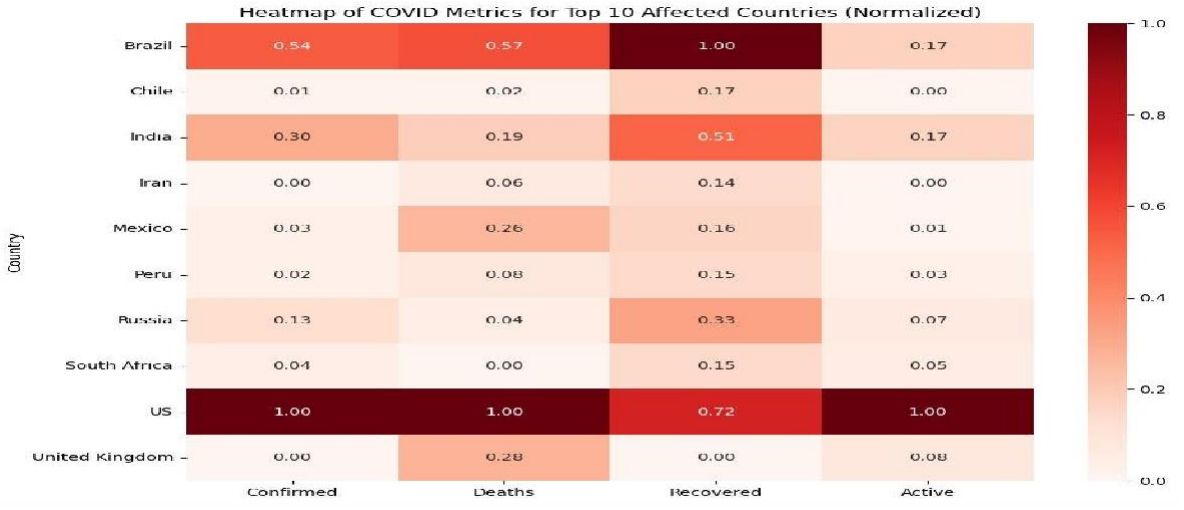
☐💻 Software Requirements

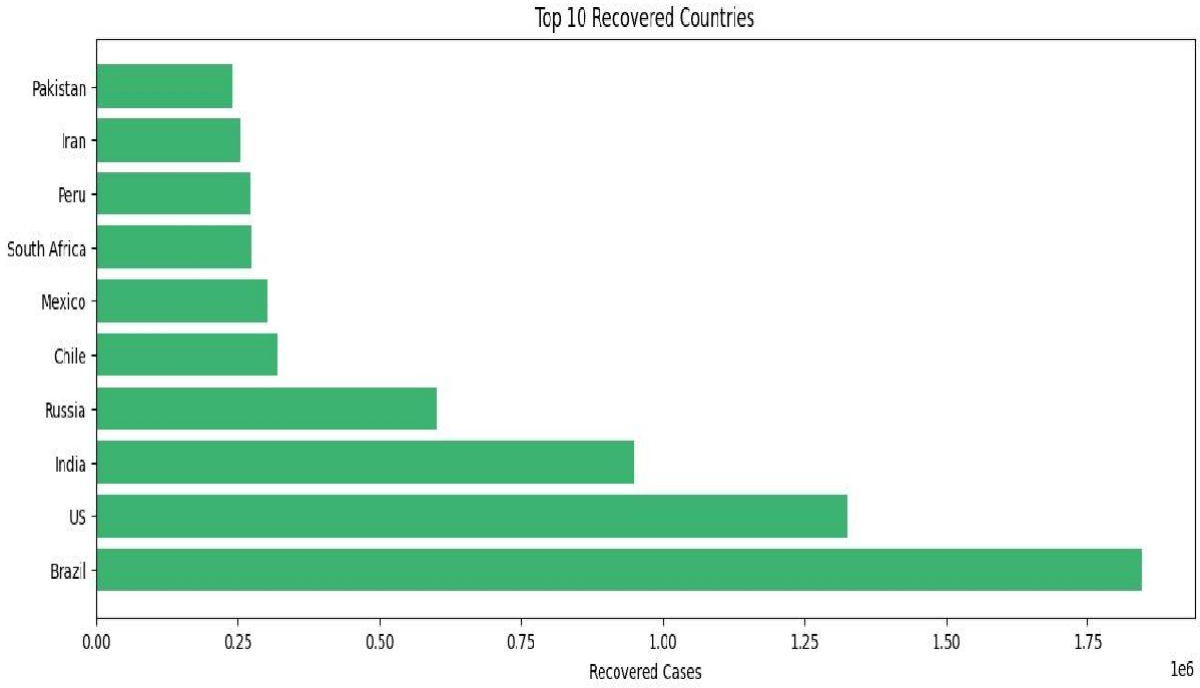
Software Tool	Purpose
Microsoft Excel	Data cleaning, preprocessing, and basic analysis
Tableau	Interactive dashboards and data visualization
Power BI	Dynamic visuals, KPIs, and business analytics dashboards
Python (3.8+)	Machine learning, data manipulation, and modeling
Jupyter Notebook	Interactive environment for writing and running Python code
Libraries	pandas, scikit-learn, matplotlib, seaborn for data analysis and modeling
Operating System	Windows 10/11 or macOS (latest version)

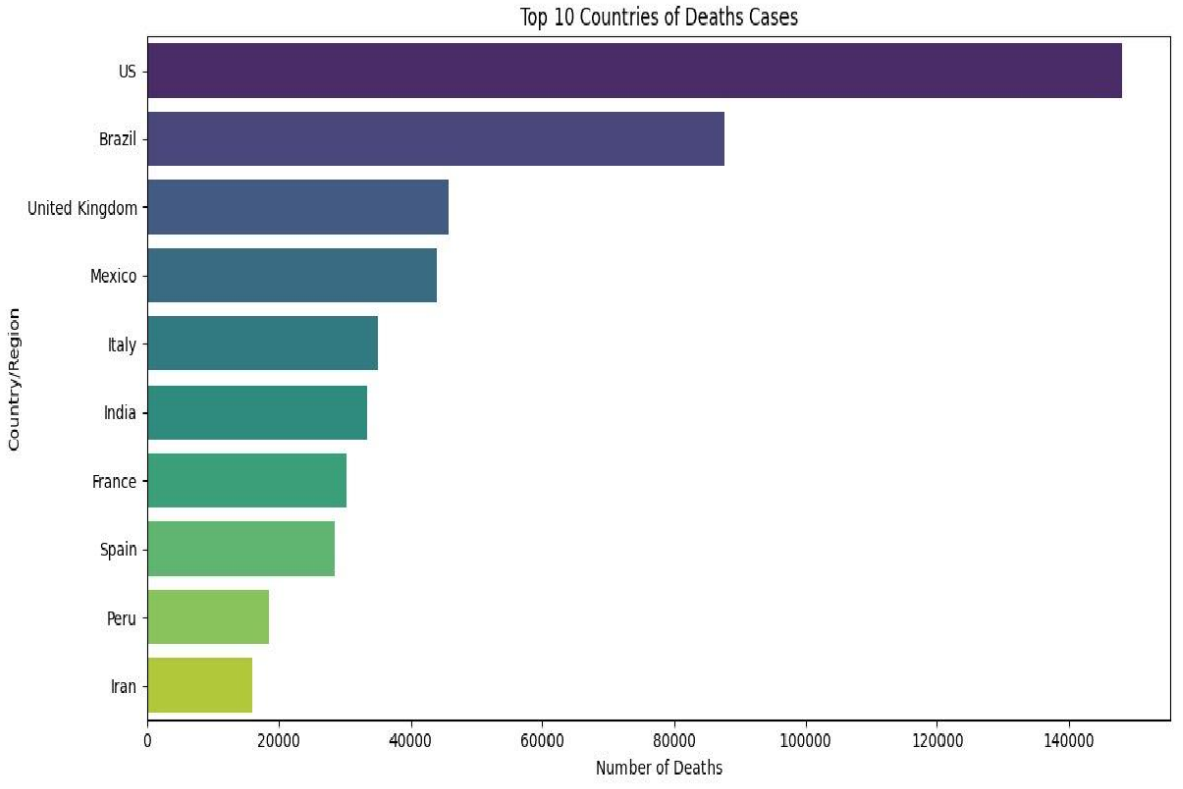
SCREENSHOT

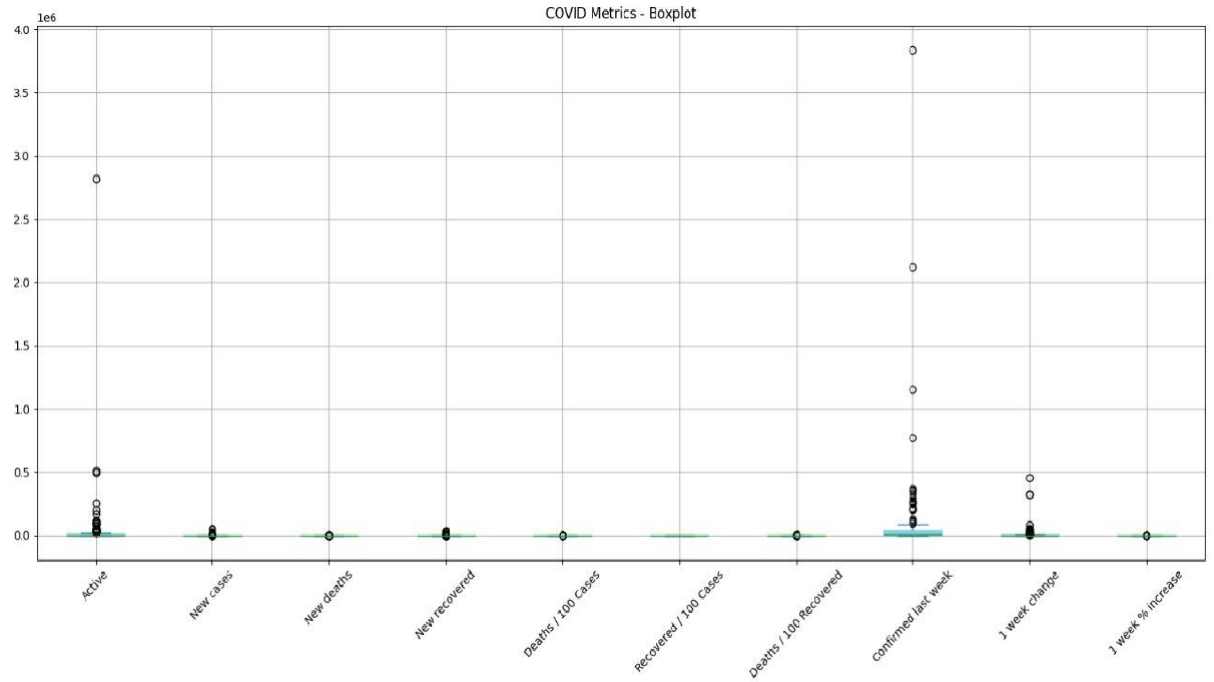
1. DASHBOARD

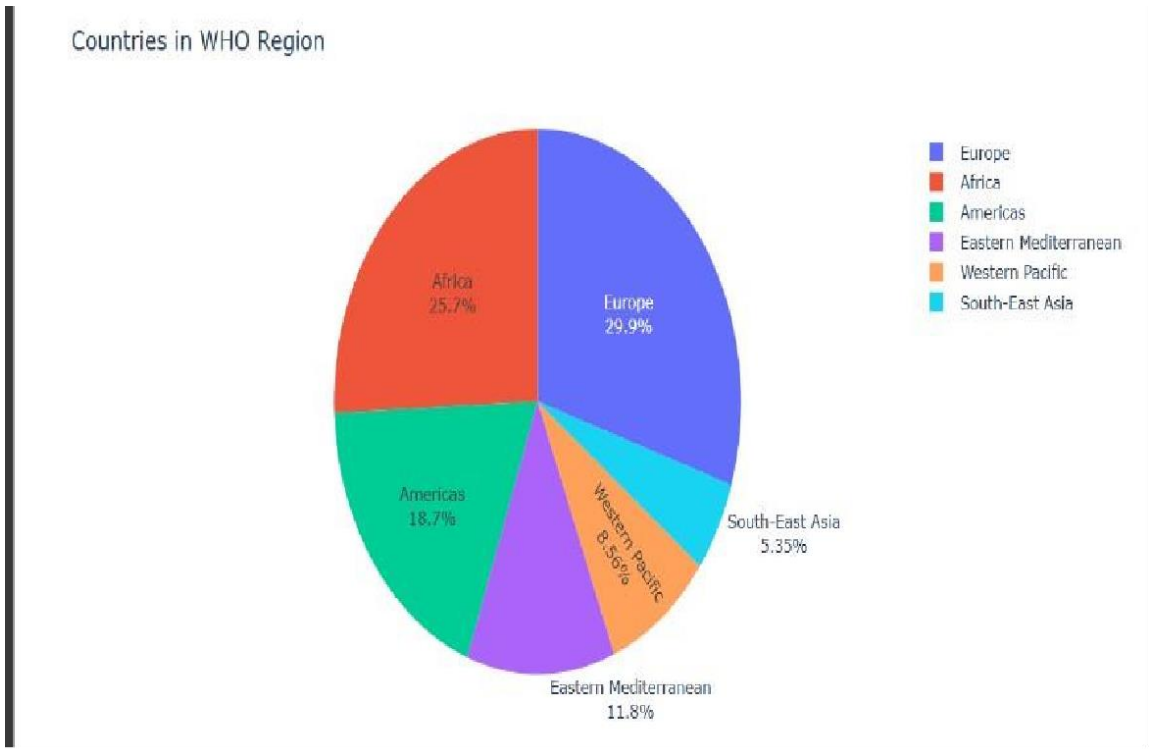
2. Machine Learning Part




Top 10 Recovered Countries
Pakistan
Iran
Peru
South Africa
Mexico
Chile
Russia
India
US
Brazil
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Recovered Cases
1e6


Top 10 Countries of Deaths Cases
US
Brazil
United Kingdom
Mexico
Italy
India
France
Spain
Peru
Iran
Country/Region
0
20000
40000
60000
80000
100000
120000
140000
Number of Deaths


COVID Metrics - Boxplot
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Active
New cases
New deaths
New recovered
Deaths / 100 Cases
Recovered / 100 Cases
Deaths / 100 Recovered
Confirmed last week
1 week change
1 week % increase


Countries in WHO Region
Africa
25.7%
Europe
29.9%
Americas
18.7%
Western Pacific
8.56%
South-East Asia
5.35%
Eastern Mediterranean
11.8%
Europe
Africa
Americas
Eastern Mediterranean
Western Pacific
South-East Asia

EXPLANATION

1. Bar Chart – Top 10 Countries by Deaths

Why use it: A horizontal bar chart clearly compares absolute death counts across countries.

What it tells us: Countries like the USA, India, and Brazil have the highest number of reported COVID-19 deaths. This graph helps identify where the pandemic had the most severe human cost.

Insight: The large gap between top countries and others highlights healthcare strain, possible delays in intervention, or higher population vulnerability.

2. Bar Chart – Top 10 Countries by Recovered Cases

Why use it: Recovery data indicates the efficiency and effectiveness of a country's healthcare and pandemic response.

What it tells us: This graph shows which countries managed to recover the most patients possibly due to strong healthcare systems, mass testing, or early intervention.

Insight: Some countries appear both in high death and high recovery charts, revealing the dual burden of high case volume and effective treatment.

3. Pie Chart – Country Distribution by WHO Region

Why use it: A pie chart provides a quick view of how countries are grouped regionally under WHO.

What it tells us: This chart highlights which WHO regions contain the most member countries, offering perspective on geographical focus and global public health governance.

Insight: Regions like Africa and Europe have large shares of member nations, indicating areas where regional health coordination is crucial.

4. Heatmap – COVID Metrics Across Top 10 Affected Countries

Why use it: A heatmap allows for quick visual comparison across multiple metrics (Confirmed, Deaths, Recovered, Active).

What it tells us: The normalized heatmap shows relative intensity — for example, which countries have the highest ratio of active or recovered cases.

Insight: You can identify countries where deaths are disproportionately high or recovery rates are significantly better, helping evaluate crisis management and public health policies.

WORK

☐ Excel Part

1. Data Cleaning

- Remove duplicate rows
- Handle missing values (e.g., fill using interpolation or leave as blanks)
- Standardize date formats and column headers
- Correct any obvious inconsistencies (e.g., negative values in confirmed cases)

2. Data Preprocessing

- **Sort data by date and country**
- Ensure numerical columns are in proper formats (e.g., integers or floats)
- Filter data by specific countries or date ranges (if needed)

3. Derived Metrics Calculation

Create **new columns** using Excel formulas:

- **New Confirmed Cases**
  =TODAY_TOTAL_CONFIRMED - YESTERDAY_TOTAL_CONFIRMED
- **Test-Confirmed Ratio**
  =TOTAL_TESTS / TOTAL_CONFIRMED
- **Survival Ratio**
  =TOTAL_RECOVERED / TOTAL_DEATHS
- **Active Cases**
  =TOTAL_CONFIRMED - (TOTAL_RECOVERED + TOTAL_DEATHS)

These derived values are essential inputs for further visualizations and modeling in Tableau, Power BI, and Python.

4. Trend Setup for MLR (for Python model)

- Prepare columns for:
  - Date (as numerical values or indices)
  - Active cases
  - Possibly vaccines or boosters for correlation
- Export a cleaned CSV to be used in Python for regression modeling

5. Pivot Tables & Basic Charts (Optional)

- Create pivot tables for country-wise summaries
- Line graphs for confirmed, recovered, and death trends (basic overview before Tableau)

6. Export Cleaned Data

- Final cleaned and enriched dataset is saved as .xlsx or .csv
- This becomes the input to Tableau, Power BI, and Python ML model

Tableau Part

1. Import and Connect Data

- Load the cleaned Excel dataset (.xlsx or .csv) into Tableau.
- Use the **Data Interpreter** if needed for auto-cleaning.
- Verify field types (date, numeric, string) are detected correctly.

2. Build Key Visualizations

A. Time Series Analysis

- **Line charts** to show:
  - Daily confirmed cases
  - New deaths
  - Recoveries over time
- Add filters for country, date, or continent.

B. Global Heat Map (Choropleth)

- **Map of the world** shaded by:
  - Total confirmed cases
  - Deaths per million
  - Active cases
- Use color gradients (e.g., red → green) for intensity.

C. Trend Line with Forecasting

- Visualize **active case trends** over time.
- Overlay forecast using **Tableau's built-in trend line** or MLR result from Excel/Python.

D. KPI Cards

- Show big-number cards for:
    - Total Confirmed Cases
    - Total Deaths
    - Total Recovered
    - Test-Confirmed Ratio
    - Survival Ratio

E. Vaccination Impact

- Line or area chart for:
    - Vaccine doses over time vs active cases
- Dual-axis charts showing booster dose effect on curve flattening

3. Add Filters and Parameters

- Allow interactive filtering by:
    - Country
    - Date range
    - Case type (confirmed, deaths, recovered)
- Add dropdowns, sliders, and checkboxes

4. Dashboard Assembly

- Combine all visualizations into **one or more dashboards**:
    - Use containers to align charts, maps, KPIs
    - Add titles, color legends, and hover tooltips
    - Set layout for desktop and tablet viewing

5. Publish and Share

- Publish to Tableau Public or Tableau Server
- Export to PDF/image if used in presentation
- Optional: Add interactivity like story points or step-by-step narratives

Tableau Dashboard Goals:

- Make insights **easy to understand at a glance**
- Support **drill-down analysis** per country or date
- Correlate **vaccine rollout and public health metrics**

🤖 Machine Learning Part

1. Environment Setup

- **Install Python 3.10+**
- Required libraries:

pip install pandas numpy scikit-learn matplotlib seaborn

2. Load and Prepare Data

- Import cleaned .xlsx or .csv file from Excel.
- Use pandas to load and explore data:

```
import pandas as pd
df = pd.read_csv("covid_cleaned.csv")
```

- Ensure key columns are present:
  - Date (as a numeric index or converted with pd.to_datetime)
  - Total_Confirmed, Total_Recovered, Total_Deaths
  - Vaccinations, Tests

- Calculate Active_Cases:

  df["Active_Cases"] = df["Total_Confirmed"] - (df["Total_Recovered"] + df["Total_Deaths"])

3. Feature Selection

- Choose predictors (X) and target (y):

X = df[["Total_Confirmed", "Vaccinations", "Tests"]]

y = df["Active_Cases"]

4. Build the Multiple Linear Regression (MLR) Model

from sklearn.linear_model import LinearRegression

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()

model.fit(X_train, y_train)

5. Evaluate the Model

from sklearn.metrics import mean_squared_error, r2_score

y_pred = model.predict(X_test)

print("R² Score:", r2_score(y_test, y_pred))

print("MSE:", mean_squared_error(y_test, y_pred))

Visualize:

```
import matplotlib.pyplot as plt
plt.plot(y_test.values, label="Actual")
plt.plot(y_pred, label="Predicted")
plt.legend()
plt.title("Active Case Prediction")
plt.show()
```

6. Forecast Future Active Cases

- Use recent or simulated future values in X to predict future Active_Cases.
- Export predictions as CSV for visualization in Tableau/Power BI:

```
future_df = pd.DataFrame(predicted_data)
future_df.to_csv("predicted_active_cases.csv", index=False)
```

ML Model Output:

- Forecasted values of active cases
- Numerical evaluation (R^2 score, MSE)
- Visualization of actual vs predicted cases

Conclusion :

The COVID-19 Global Insight Dashboard project presents a comprehensive and multidisciplinary approach to understanding the global impact of the COVID-19 pandemic through data-driven insights. By integrating tools such as Microsoft Excel, Tableau, Power BI, and Python-based machine learning, this project bridges the gap between raw public health data and meaningful, actionable information.

The project begins with meticulous data cleaning and preprocessing in Excel, where inconsistencies are resolved, and key derived metrics—such as new confirmed cases, test-confirmed ratio, and survival ratio—are calculated. This structured data serves as the foundation for visualization and predictive analysis.

Interactive dashboards developed in Tableau and Power BI provide intuitive and real-time representations of the pandemic's evolution. These dashboards enable users to explore global trends, identify hotspots, assess vaccine impact, and track daily confirmed and active cases. Key performance indicators (KPIs) and heatmaps allow for comparative analysis between countries and time periods, offering critical insights to policymakers and the general public.

In parallel, the use of machine learning—particularly Multiple Linear Regression—enhances the project by forecasting future active case trends based on historical and correlated features such as vaccinations and testing. These predictive capabilities are essential for proactive public health planning and resource allocation.

Overall, the project exemplifies the power of combining statistical computation, data visualization, and machine learning to extract valuable insights from complex datasets. It reinforces the importance of data literacy and technology in tackling global challenges and highlights the potential of integrated analytical frameworks in driving informed decision-making during public health crises.