International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

COVID-19 Data Tracker & Visualizer with Forecasting and Analytics

DANISH SAGEER TAMBOLI

Department of BACHELOR OF VOCATIONAL IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ,

Anjuman -I- Islam's Abdul Razzaq Kalsekar Polytechnic, New Panvel, Maharashtra , India

---

Abstract -The COVID-19 pandemic has triggered a global demand for real-time data analysis, monitoring, and visualization. In response, this paper presents a comprehensive COVID-19 Data Tracker and Visualizer developed using Python, Pandas, Plotly, and Google Colab. This system processes a publicly available Kaggle dataset that contains chronological data on confirmed, recovered, and death cases worldwide. The notebook not only visualizes this data with interactive and informative graphs but also forecasts future trends using simple linear regression models. It presents critical visual summaries, including line plots for daily cases, active case trends, a choropleth world map for spatial distribution, and a pie chart illustrating recovery and death ratios. The results show that the model can effectively communicate evolving COVID-19 patterns and enable meaningful insight for academic, public health, and data science audiences. This project serves as a lightweight, accessible tool to bridge data and decision-making during global health crises.

Key Words: COVID-19, Data Visualization, Forecasting, Linear Regression, Choropleth Map, Pandas, Plotly, Python, Google Colab

1. INTRODUCTION

The novel coronavirus (COVID-19) emerged in late 2019 and rapidly became a global pandemic, affecting economies, healthcare systems, and individuals across continents. One major challenge faced during the pandemic has been the availability and interpretation of real-time data to track the virus's spread and plan appropriate countermeasures. The rise in publicly accessible datasets, particularly those published on platforms like Kaggle, has enabled data scientists to contribute insights and forecasts. Data visualization plays a critical role in interpreting large amounts of pandemic data effectively.

This paper proposes a project-based approach to visualize and analyze COVID-19 data using Python tools. Built entirely in Google Colab for accessibility, the project allows users to interactively explore trends, compare countries, and forecast future outcomes. It leverages key libraries like Pandas for data manipulation and Plotly for interactive graphs. The addition of forecasting using linear regression further adds predictive capability to the dashboard.

s.

2. Literature Survey

Several existing studies have attempted to model and visualize COVID-19 data using AI and ML. A paper by Gupta et al. (2021) applied ARIMA and LSTM to predict future case numbers in India, while Kumar and Singh (2020) developed a web-based dashboard using Dash and D3.js to illustrate global COVID trends. Most such tools focus either on static reporting or on complex machine learning models requiring deep domain knowledge. Our project simplifies this by using intuitive tools like Plotly and linear regression while maintaining informative quality and interactivity. This balances accessibility and functionality, making it ideal for academic use and non-expert user

3 .Dataset Description

The dataset used is titled "COVID-19 Clean Complete Dataset," publicly available on Kaggle [1]. It contains time-series data starting from January 2020, with attributes such as Date, Country, Province/State, Confirmed, Deaths, and Recovered. The data is preprocessed to fill missing values and group statistics at global and country levels. This makes the dataset suitable for daily trend visualization, aggregation, and modeling.

4. Methodology

A. Data Preprocessing

- Convert Date fields to datetime objects
- Group data by date and country
- Calculate active cases using the formula: Active = Confirmed - (Recovered + Deaths)

B. Visualization Modules

- Daily New Cases: Line plots using Plotly showing daily change in Confirmed, Recovered, and Deaths.
- Active Cases Over Time: Visualizes how active cases evolve globally.
- Choropleth Map: Color-coded world map showing intensity of confirmed cases by country.
- Pie Chart: Shows ratio of Recoveries to Deaths based on the most recent date.

C. Forecasting

The forecasting component uses linear regression on the number of days since the first case to predict the next 14 days of confirmed, recovered, and death cases. This simple ML model is sufficient to identify rising or plateauing trends, which are plotted using Matplotlib.

5. Results and Discussion

The results demonstrate successful visualization of COVID-19 data trends. Key findings include:

- A consistent linear rise in confirmed cases globally
- An increasing recovery rate in many regions, visible in pie chart ratios
- Active case load stabilizing over time in some countries
- Forecasted values provide a near-accurate short-term projection of future data trends

6. Conclusion and Future Work

This project illustrates how publicly available data can be transformed into powerful visual tools for understanding pandemic dynamics. It combines descriptive analysis with basic predictive modeling to offer insights into COVID-19 case trends. The system is easy to use, open-source, and lightweight, making it ideal for educational purposes.

Future work could involve using live APIs for real-time data, integrating advanced ML models like LSTM or Prophet for better accuracy, and building a Streamlit-based web dashboard for public access. Multi-country comparison widgets or integration with vaccination datasets are also possible extensions.

7. Acknowledgments

I sincerely thank Anjuman-I-Islam's Abdu Razzaq Kalsekar Polytechnic, New Panvel, and the AI & Data Science department for their support and guidance throughout this project. The author also acknowledges the Kaggle community and open-source contributors for providing cleaned COVID-19 datasets and tools. Gratitude is extended to faculty members and mentors who provided guidance during project development.

8. References

[1] Imdevskp, "Coronavirus Report Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/imdevskp/corona-virus-report

Gupta, A., Kumar, R., and Roy, S. "Forecasting COVID-19 Trends in India Using Time Series Models," Journal of Biomedical Informatics, 2021.

Kumar, D. and Singh, R. "Interactive Dashboard for COVID-19 Using Dash and D3.js," Data Science Journal, 2020.