International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

COVID-19 DATA ANALYSIS USING PANDAS

[1]K. Sunanda,[2]G.Siddarth,[3]C.Sainath,[4]K.Durga prasad, [5]M.Mahindhar Naik

[1]Assistant Professor, [2345]Students

Department of Computer Engineering(Internet of Things)

Siddhartha institute of technology & sciences,narapally

kondapallisunanda.cse@siddhartha.co.in,23tq1a6921@siddhartha.co.in, 23tq1a6904@siddhartha.co.in, 23tq1a6937@siddhartha.co.in, 23tq1a6939@siddhartha.co.in

ABSTRACT

The COVID-19 Data Analysis project focuses on analyzing pandemic data to understand the spread and impact of the coronavirus across different countries. In this project, data analysis techniques are applied to examine important information such as the number of confirmed cases, deaths, and recovered patients. The main aim of the project is to extract meaningful insights from large datasets and present them in an easy-to-understand form using data visualization. The project is implemented using the Python programming language with the help of powerful data analysis and visualization libraries such as pandas, Matplotlib, and Seaborn. The dataset used in the project contains information about COVID-19 cases recorded in different countries over a specific period. Using these tools, the data is cleaned, organized, and analyzed to calculate totals, compare statistics between countries, and identify patterns in the spread of the virus. Through this analysis, the project generates visual representations such as bar charts and graphs that help users easily understand trends in COVID-19 cases. These visualizations make it easier to observe which countries were most affected and how the pandemic evolved over time. The use of data analysis techniques helps transform raw data into useful information for better understanding and decision-making. Overall, this project demonstrates how Python and its data science libraries can be used to perform efficient data analysis on real-world datasets. It also helps beginners learn important concepts of data handling, data visualization, and statistical analysis in a practical way. The project highlights the importance of data analysis in studying global health issues and improving awareness about pandemic trends.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

I INTRODUCTION

The global outbreak of COVID-19 created a significant impact on public health, economies, and daily life around the world. During the pandemic, a huge amount of data was generated every day, including the number of confirmed cases, deaths, and recovered patients across different countries and regions. Analyzing this data is very important to understand how the virus spreads, how it affects different populations, and how governments and health organizations can respond effectively. Data analysis helps convert raw data into meaningful information that can support decision-making and research. In this project, COVID-19 data is analyzed using the Python programming language and the powerful data analysis library pandas. Pandas provides efficient tools for handling large datasets, performing data cleaning, filtering, grouping, and statistical analysis.

II LITERATURE SURVEY

The A literature survey is an important step in understanding the research and studies that have already been conducted in a particular field. For the COVID-19 data analysis project, several researchers and organizations have studied the spread, impact, and trends of the COVID-19 pandemic using different data analysis techniques. These studies help in understanding how data analytics can be used to analyze large healthcare datasets and derive meaningful insights. Many global organizations such as the World Health Organization and research institutions collected and published large datasets related to COVID-19 cases, deaths, and recoveries. These datasets were used by researchers to study infection trends, identify high-risk regions, and understand the impact of the pandemic on different populations. Several studies have applied data science and statistical methods to analyze COVID-19 data. Researchers have used programming languages like Python because of its powerful libraries that support data processing and visualization. Libraries such as pandas are widely used for data cleaning, filtering, grouping, and statistical analysis. These tools allow researchers to efficiently handle large datasets and perform complex data operations. In addition, visualization libraries such as Matplotlib and Seaborn are commonly used to represent COVID-19 data through graphs and charts. These visualizations help researchers easily identify patterns such as daily case growth, regional comparisons, and recovery trends. Some studies also applied machine learning techniques to predict the future spread of COVID-19 using historical data. Forecasting models such as linear regression and time-series analysis were used to estimate future case growth and understand how the pandemic might evolve over time. However, many existing studies focus mainly on statistical reports or theoretical analysis rather than practical implementation using programming tools. Therefore,

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

this project focuses on implementing a practical data analysis system using Python and pandas to analyze COVID-19 datasets and visualize the results effectively.

III SYSTEM ANALYSIS

The COVID-19 Data Analysis Using Pandas project focuses on collecting, processing, and analyzing large-scale COVID-19 datasets to extract meaningful insights about the pandemic. The system is designed to handle structured data such as daily case counts, recoveries, deaths, and vaccination statistics from reliable sources like WHO and government databases. Using the Pandas library in Python, the system performs data cleaning, transformation, and aggregation to prepare the data for analysis. Key functionalities include identifying trends over time, comparing statistics across regions, calculating growth rates, and visualizing data through charts and graphs. The system also supports filtering and querying data based on specific criteria such as country, date range, or case type. This analysis aids researchers, policymakers, and the general public in understanding the spread of COVID-19, evaluating the effectiveness of interventions, and making data-driven decisions. The modular design ensures scalability, allowing the addition of new datasets or features, such as predictive modeling or correlation studies, without affecting existing functionalities.

Existing system

In the existing system, COVID-19 data is primarily accessed through static dashboards, spreadsheets, or government portals, which provide limited interactivity and analysis capabilities. Most of these platforms display raw numbers of cases, recoveries, and deaths without offering the flexibility to perform custom analysis or extract insights according to specific needs. Users often rely on manual calculations or third-party tools to analyze trends, compare regions, or visualize data over time. Additionally, the existing system lacks automation for data cleaning and preprocessing, making it time-consuming to handle large datasets. This limits the ability of researchers, healthcare professionals, and policymakers to quickly interpret data and make timely decisions based on dynamic pandemic trends.

Disadvantages of existing system

- Provides only static data with limited interactivity.
- Manual calculations are often required for trend analysis.
- Time-consuming to clean and preprocess large datasets.
- Lacks customizable analysis options for specific queries or regions.

Proposed System

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

The proposed system aims to automate and streamline COVID-19 data analysis using the **Pandas** library in Python. Unlike existing systems, it provides dynamic data processing, cleaning, and visualization, allowing users to extract meaningful insights efficiently. The system supports custom queries, enabling analysis by country, region, date range, or case type. It can calculate growth rates, compare trends, and generate visual representations such as line graphs, bar charts, and heatmaps for better understanding. By automating data handling and integrating multiple data sources, the system reduces manual effort, minimizes errors, and ensures timely access to updated information. This makes it easier for researchers, policymakers, and healthcare professionals to monitor the pandemic, evaluate intervention strategies, and make informed, data-driven decisions.

Advantages of proposed system

- Automates data cleaning and preprocessing, reducing manual effort.
- Supports dynamic and customizable analysis by country, region, or date range.
- Provides visualizations like charts and graphs for better understanding of trends.
- Enables quick and accurate insights for decision-making.
- Handles large datasets efficiently using Pandas.

IV METHODOLOGY

The methodology of this project involves a systematic approach to collecting, processing, analyzing, and visualizing COVID-19 data using Python and the Pandas library. First, relevant datasets are collected from trusted sources such as WHO, government portals, and COVID-19 APIs. Next, the data undergoes preprocessing, including cleaning missing values, correcting inconsistencies, and formatting dates and numeric fields for analysis. After preprocessing, Pandas functions are used to perform operations like filtering, grouping, aggregation, and statistical calculations to extract meaningful trends and patterns. The processed data is then visualized using libraries such as Matplotlib or Seaborn to generate charts, graphs, and heatmaps that clearly represent the spread of COVID-19 over time and across regions. Finally, the insights derived from the analysis are interpreted to support decision-making, identify trends, and help researchers and policymakers understand the pandemic's impact effectively.

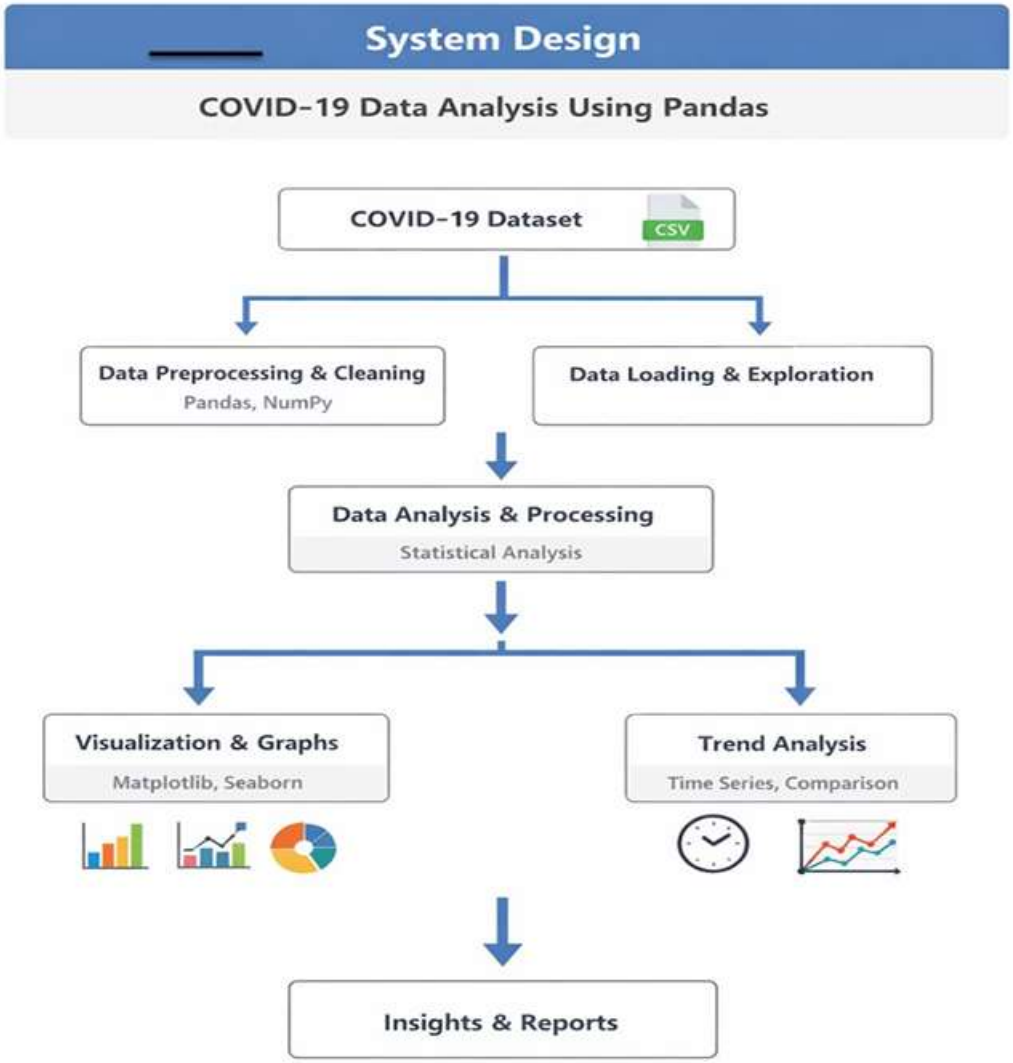
System Architecture

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



The proposed system introduces a modern and efficient approach for analyzing pandemic data related to COVID-19 using data science techniques. Unlike traditional systems that depend on manual data processing and spreadsheet tools, the proposed system uses the programming language Python along with powerful data analysis libraries to process and analyze large datasets effectively. In this system, COVID-19 datasets are collected from reliable public sources such as government portals and global health organizations. The data may include information such as confirmed cases, deaths, recoveries, and dates for different countries or regions. After collecting the dataset, it is processed using the library pandas, which helps in organizing, cleaning, and analyzing the data efficiently.

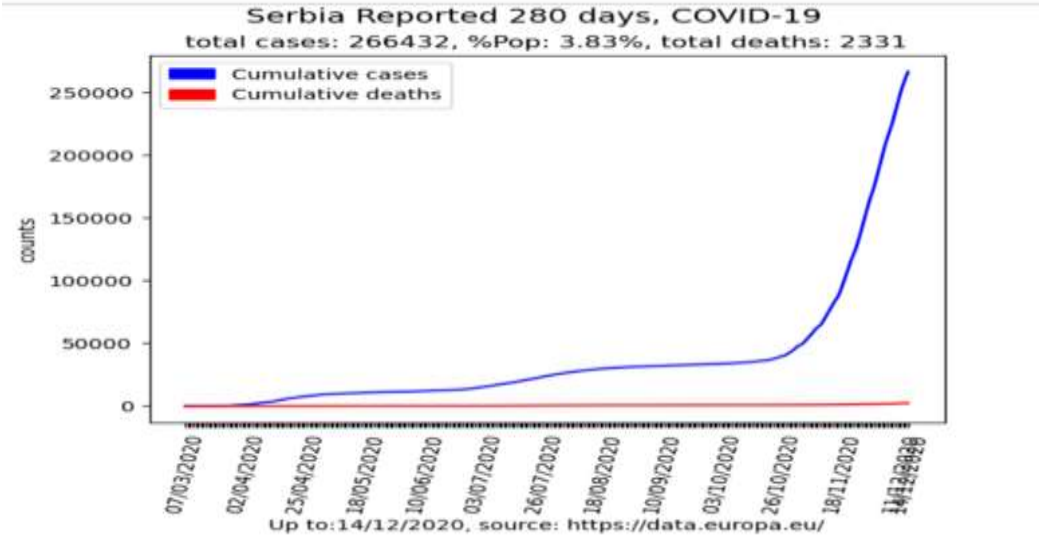
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

V RESULTS&OUTPUT

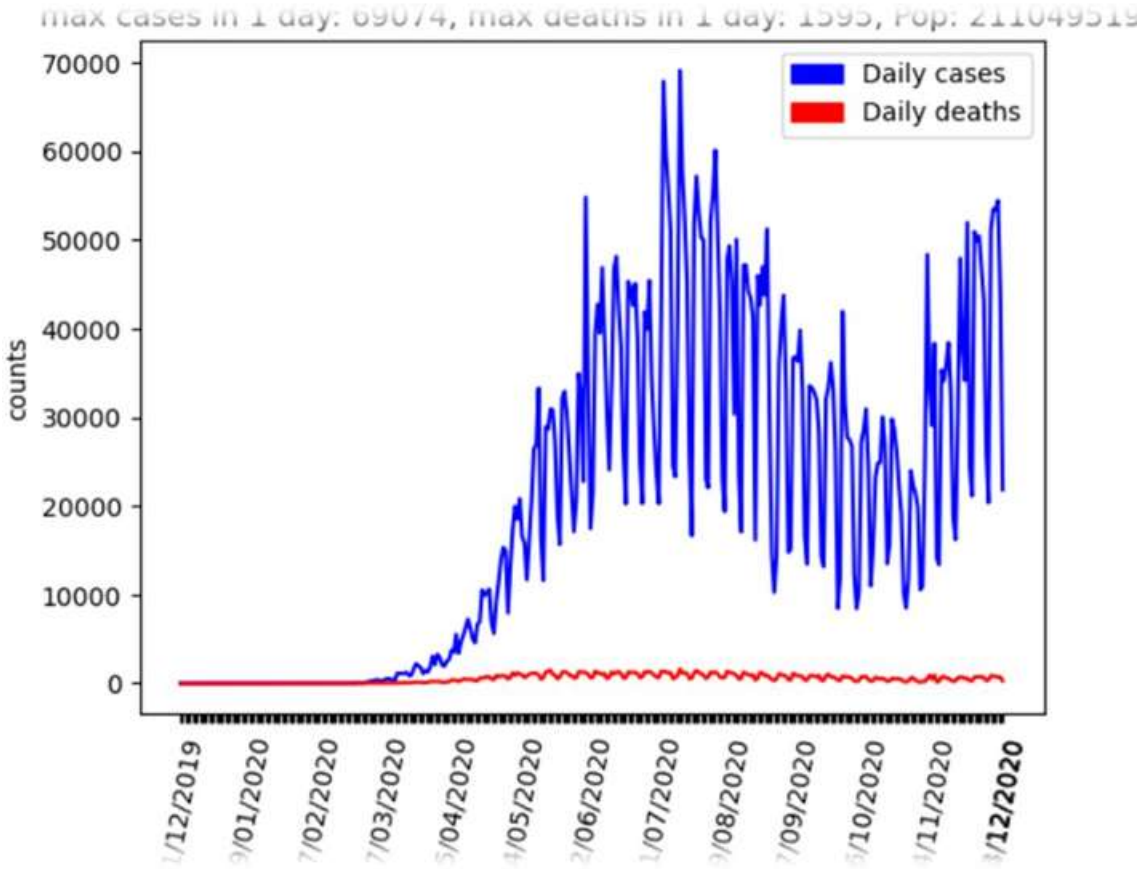


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



VI CONCLUSION

The COVID-19 Data Analysis Using Pandas project demonstrates how modern data analysis techniques can be used to study and understand the spread of COVID-19. During the pandemic, a huge amount of data was generated every day, including information about confirmed cases, deaths, recoveries, and vaccination rates across different countries. Analyzing such large datasets manually is difficult, which highlights the importance of using efficient data analysis tools. In this project, the programming language Python was used along with the powerful data manipulation library pandas to process and analyze the dataset. The project successfully demonstrates how data can be collected, cleaned, organized, and analyzed to obtain meaningful insights. By performing operations such as grouping, filtering, and calculating totals, the system helps in identifying patterns and trends in the COVID-19

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

data. Data visualization is another important part of the project. Using libraries such as Matplotlib and Seaborn, the analyzed data is represented through graphs and charts. These visualizations make it easier to understand the growth of cases, compare statistics between countries, and observe how the pandemic evolved over time. The project also helps students and beginners understand the practical applications of data science. It shows how real-world datasets can be analyzed using programming tools and how valuable insights can be extracted from raw data. Through this project, users gain knowledge about data preprocessing, analysis, and visualization techniques. Overall, the COVID-19 Data Analysis Using Pandas project proves that data analysis tools can play a significant role in understanding global health issues. It provides a simple, efficient, and practical approach for analyzing pandemic data and highlights the importance of data-driven decision making in modern research and public health management.

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.