

COVID-19 ILLUMINATING DATA WITH STRIKING VISUALS

Usha Rani M

PG, Student

Dept. of MCA

The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
usharanimca2025@gmail.com

Sowmya J

Assistant Professor

Dept. of MCA

The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
sowmyaj@theoxford.edu

ABSTRACT

This study analyzes COVID-19 data to comprehend its global spread, impact, and control measures. World Health Organization (WHO) and Johns datasets are publicly available. At Johns Hopkins University, we analyzed indicators such as confirmed cases, deaths, recovery rates, testing levels, and vaccination coverage across various countries and timeframes. The analysis uses statistical methods and data visualization to identify trends, anomalies, and evaluate intervention effectiveness. Significant variation in case growth and mortality rates across regions is linked to healthcare infrastructure, public health, policies, and vaccination rollout. The study emphasizes the importance of transparent data collection frameworks to address data accuracy and reporting inconsistencies.

KEYWORDS: *covid-19 pandemic Analysis, Understanding pandemic effects on health, Using data to inform health decisions, Assessing vaccine impact, Global Health Trends.*

INTRODUCTION

The COVID-19 Data Analysis includes the collection, processing, and interpretation of pandemic-related data such as confirmed cases, deaths, recoveries, and vaccination rates across countries and over time. It extends to analyzing the relationship between COVID-19 trends and socio-economic indicators from sources like the Worldwide Happiness Report. This analysis supports real-time monitoring, forecasting future outbreaks, evaluating the effectiveness of public health interventions, and identifying at-risk populations. Additionally, it can inform global health policies, resource allocation, and preparedness strategies for future pandemics. The scope is both retrospective and forward-looking, with applications in healthcare, governance, economics, and social sciences.

LITERATURE SURVEY

- Johns Hopkins University COVID-19 Dashboard (Dong et al., 2020): One of the most referenced real-time dashboards, this study presented a comprehensive platform that visualized the global spread of COVID-19 using confirmed cases, deaths, and recoveries, providing valuable data for researchers and policymakers.
- Our world in Data COVID-19 Dataset: This open access project compiled extensive COVID-19 datasets including testing rates, hospitalizations, vaccination data, and government interventions. trend analysis across countries.
- World Happiness Report: This report explored the relationship between COVID-19 and well-being, showing how mental health, trust in government, and social support played a role in coping with the pandemic.
- Impact of COVID-19 On Socioeconomic Conditions: This study analyzed how COVID-19 influenced economics, education, and healthcare systems globally, highlighting the need for data-driven responses and long term recovery planning.
- Machine Learning for COVID-19 Prediction Researches applied machine

learning models such as ARIMA, LSTM, and SVM to forecast COVID-19 trends and hospital demand, demonstrating the effectiveness of AI in pandemic response.

- Geospatial Analysis of COVID-19 : This study used spatial-temporal data analysis to map the spread of COVID-19 and identify regional hotspots, improving the targeting of healthcare resources.
- Data-Driven Policy Making: The paper emphasized how data analytics guided effective policy decisions like lockdowns and mobility restrictions, reducing virus transmission and healthcare strain.

EXISTING SYSTEM

The existing systems for COVID-19 data analysis primarily consist of dashboards, statistical reports, and trend visualizations provided by global health organizations and government agencies. Platforms such as the Johns Hopkins University COVID-19 Dashboard, Our World in Data, and the WHO offer real-time tracking.

However, most of these platforms focus on descriptive analytics presenting raw data and simple metrics without deeper analytical capabilities such as predictive modeling, pattern detection, or correlation analysis across socio-economic factors. Additionally, many of these systems face challenges related to data accuracy, latency, and standardization, as different countries report statistics with varying methodologies and timelines. While they serve as vital tools for information dissemination and policy planning, the existing systems often lack customizable analytical capabilities tailored to specific research needs or regional conditions.

PROPOSED SYSTEM

COVID-19 data analysis systems typically focus on isolated metrics like confirmed cases, deaths, and vaccinations, often presented through dashboards or static reports. These systems usually emphasize short-term trends and overlook the integration of socio-economic indicators, such as happiness levels, to provide a more comprehensive understanding. By not combining public health data with well-being metrics, existing analyses miss opportunities to gain deeper insights into the pandemic's broader societal impact, limiting the understanding of its effects on people's lives beyond just health outcomes. Decision-making is often reactive, with limited predictive capabilities or correlation

analysis across different domains.

Design is multi-step process that focuses on data structure software architecture, procedural details, (algorithms etc.) and interface between modules. The design process also translates the requirements into the presentation of software that can be accessed for quality before coding begins.

METHODOLOGY

- Exploratory Data Analysis: Visualize relationships, outliers, and patterns using charts, heatmaps, and correlations.
- Time Series Analysis: Analyze and forecast trends using models like ARIMA, Prophet, or LSTM.
- Correlation Analysis: Study the relationship between COVID-19 metrics and happiness indicators.
- Comparative Analysis : Compare countries or regions based on case rates, death rates, and happiness scores.
- Clustering: Group countries using unsupervised learning (e.g., K-means) based on similar COVID-19 and happiness profiles.

DATA SET

- **Covid19 Confirmed Dataset:** It includes the daily values of the confirmed covid-19 cases by country and region.
- **Covid19 Deaths Dataset:** The daily published figures of COVID-19 related deaths by country and region.
- **Worldwide Happiness Report:** This report ranks the countries according to GDP, social support, life expectancy, and freedom among others to reveal their well being.

DATA PRE-PROCESSING STEPS

Here are common Data Pre-processing Steps for COVID-19 and happiness datasets:

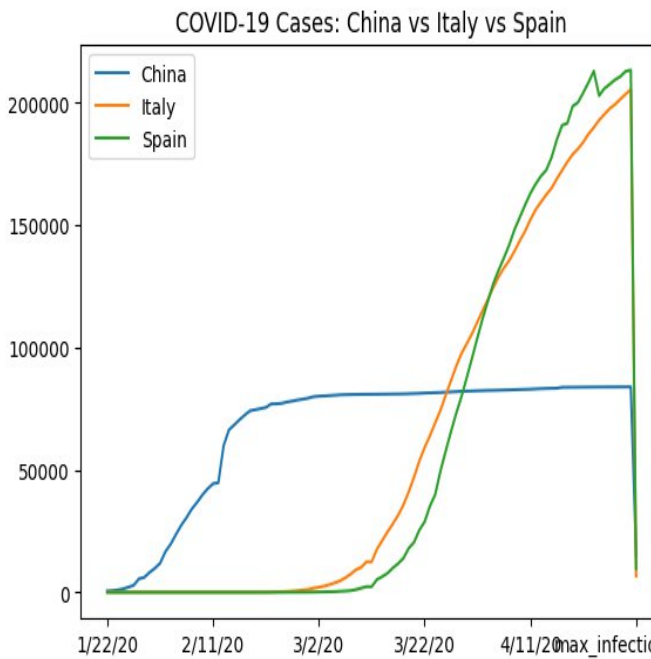
- The pre-processing schedule of data includes the following steps:
- The typical Data Pre-processing Procedures of COVID-19 and happiness datasets are as follows:
- Load Libraries -Load the needed libraries(e.g.: pandas, numpy, matplotlib).
- Loading Datasets-Load all CSV or excel files into dataframes
- Check for Nulls -Identify and manage missing or null data.
- Rename Columns -. Clean up column names to have consistency.
- Data Type Conversions Data Types -

Converting date column values to date type; convert numerical to appropriate format.

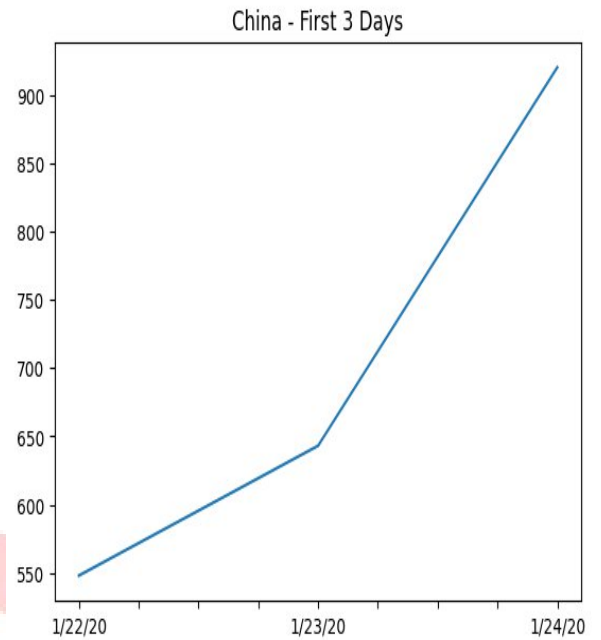
- Merge Datasets Merge using fields such as country name and date.
- Filter Data Filter by time
- Add New Functions to calculate active cases, death rates, recovery rates, etc.
- Achieve Duplicates Cleanup Eliminate duplicate records, in case they exist
- Refresh/Standardize Data Optional, when using a model that needs the feature scaling.
- Drop Redundant Columns Remove columns which are already captured by other columns.
- Check Outliers Recognize and deal with the anomalies or data errors.
- Encode Categorical Data Encode categorical values such as names of countries.

EXPERIMENTAL RESULTS

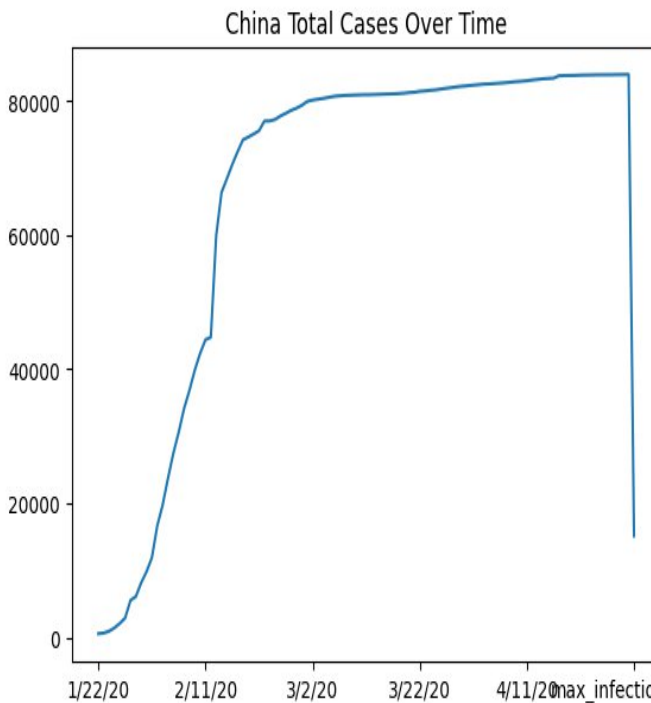
1. COVID-19 Cases: China vs Italy vs Spain:



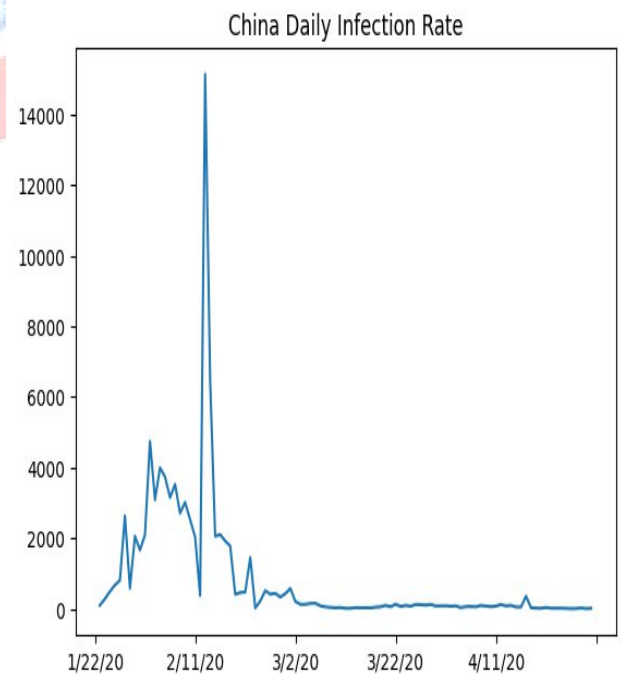
3. China - First 3 Days:



2. China Total Cases Over Time:



4. China Daily Infection Rate:



CONCLUSION

COVID-19 In order to comprehend the pandemic's effects on society, health, and well-being worldwide, data analysis is essential. We can find important insights, forecast future trends, and assist with data-driven decision-making by examining datasets on confirmed cases, deaths, and vaccinations—as well as socioeconomic indicators like those from the Worldwide Happiness Report. This approach not only improves public health responses but also shows how pandemics affect many parts of life. It emphasizes the need for proactive measures and global cooperation in future emergencies.

REFERENCES

- [1] Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5),533534.[https://doi.org/10.1016/S14733099\(20\)30120-1](https://doi.org/10.1016/S14733099(20)30120-1).
- [2] Helliwell, J. F., Layard, R., Sachs, J., & De Neve, J.-E. (2021). *World Happiness Report 2021*. Sustainable Development Solutions Network. <https://worldhappiness.report>.
- [3] Ritchie,H.,Ortiz-Ospina,E.,Beltekian, D.,etal.(2020). Coronavirus Pandemic (COVID-19). *OurWorld in Data*. <https://ourworldindata.org/coronavirs>.
- [4] Pak, A., Adegboye, O. A., Adekunle, A.I.,etal.(2021).EconomicConsequences of the COVID-19 Outbreak: The Need for Epidemic Preparedness. *FrontiersinPublicHealth*,8,241.<https://doi.org/10.3389/fpubh.2020.00241>.
- [5] Arora, P., Kumar, H., & Panigrahi, B. K. (2021).
- [6] Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID- 19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*,118,102202.<https://doi.org/10.1016/j.apgeog.2020.102202>.
- [7] Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228), 931–934.[https://doi.org/10.1016/S01406736\(20\)30567-5](https://doi.org/10.1016/S01406736(20)30567-5).