JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Analysis and Visualization of Data on the Impacts of Covid-19 Globally and Locally

Muhammad Iqbal[1], Julius Chaezar Bernard Buana Yudha[1], Reza Nazilatul Umimah[1], Fadhel Akhmad Hizham[1]

Informatics, Institut Teknologi dan Bisnis Widya Gama Lumajang, Lumajang, Indonesia[1]

Corresponding Author: Julius Chaezar Bernard Buana Yudha (Juliusbernard86@gmail.com)

ARTICLE INFO

Date of entry:
9 Februaryl 2025
Revision Date:
2 April 2025
Date Received:
24 April 2025

ABSTRACT

The COVID-19 pandemic has had a profound impact on multiple aspects of human life, including food supply, mental health, and healthcare service management. This study aims to examine these impacts by applying a combination of data analysis methods such as data preprocessing, exploratory data analysis (EDA), predictive algorithms, and data visualization. The datasets utilized include information related to mental health conditions, food security, and COVID-19-related health statistics. The findings indicate a significant increase in mental health issues, such as anxiety and depression, as well as disruptions in food supply chains that have adversely affected global food security. Moreover, data visualization has proven to be a valuable tool in supporting decision-making processes in healthcare management. However, most implementations remain limited in scope and are often confined to internal agency use. Therefore, this study recommends further development in integrating data sources, enhancing the application of predictive algorithms, and optimizing data visualization for more effective decision-making in managing global health crises.

Keywords: COVID-19, Data Visualization, Predictive Algorithms, Data Preprocessing, Exploratory Data Analysis

Cite this as: Iqbal, M., Yudha, J. C. B. B., Umimah, R. N., & Hizham, F. A. (2025). Analysis and Visualization of Data on the Impacts of Covid-19 Globally and Locally. *Journal of Informatics Development*, *3*(2), 21–32. https://doi.org/10.30741/jid.v3i2.1513

INTRODUCTION

The COVID-19 pandemic has posed unprecedented challenges to global systems, disrupting not only physical health but also food security, mental well-being, and the resilience of healthcare infrastructure (Isbaniah & Susanto, 2020; Khan, 2022). While early discussions focused primarily on the clinical symptoms and mortality of the disease—such as fever, dry cough, fatigue, and respiratory distress—recent studies have highlighted broader implications that extend to psychological health and the socioeconomic fabric of communities (Sinha et al., 2022; Sharma, 2020). Transmission of the virus, even by asymptomatic individuals, and its relatively moderate case fatality rate of approximately 2.3% compared to SARS and MERS, allowed it to spread rapidly and extensively, affecting millions worldwide (Taheri, 2020; Khan, 2022).

JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

One of the critical consequences of the pandemic has been the disruption of food supply chains, leading to reduced accessibility and affordability of essential nutrients such as cereals, fruits, and animal-based protein. These disruptions have worsened existing inequalities and posed a significant threat to global food security (Dhummad, 2025). At the same time, restrictions on mobility and economic insecurity have deepened mental health issues across populations. Various studies have shown increased reports of anxiety, depression, and post-traumatic stress disorder among individuals undergoing isolation or treatment for COVID-19 (Yuanti, 2024; Nasrullah & Sulaiman, 2021). The interplay between physical health, mental health, and nutrition presents a multi-dimensional crisis that demands an integrative, data-informed approach.

In this context, Exploratory Data Analysis (EDA) and data visualization techniques have gained prominence as tools that support understanding and decision-making under uncertainty. EDA promotes an open-minded approach to exploring patterns, trends, and relationships within datasets, providing critical insights that precede model building (Courtney, 2021; Song, 2023). Techniques such as line plots, scatter plots, histograms, and bar charts allow for effective monitoring of dynamic variables such as infection rates, vaccination coverage, and demographic vulnerabilities (Firiza et al., 2024; Khanam et al., 2020). EDA also plays a vital role in detecting missing values, outliers, and inconsistencies—ensuring data reliability for subsequent analysis (Galatro & Dawe, 2023; Sandfeld, 2023). Moreover, advanced tools such as XInsight facilitate interpretable and causal analysis, enhancing users' trust in the analytical process (Ma et al., 2022).

Beyond healthcare, EDA finds application across various domains. In the field of sports analytics, it has been used to study performance trends among Olympic medalists (Sagala & Aryatama, 2022). In education, EDA supports evidence-based school improvement initiatives (Courtney, 2021). Its versatility makes it a valuable framework for analyzing cross-disciplinary issues such as the COVID-19 pandemic, which spans health, nutrition, and behavior.

Given the complexity of these interconnected challenges, this study seeks to address the following research question: How has the COVID-19 pandemic simultaneously affected mental health, food supply, and healthcare services, and how can data analysis and visualization techniques be leveraged to support evidence-based decision-making across these domains? By integrating multiple datasets and analytical tools—including EDA, visualization, and machine learning—this research aims to provide a holistic understanding of the pandemic's impacts and inform future strategies for resilience and recovery.

METHOD

This study uses a quantitative research design with an exploratory and predictive approach to analyze the impact of the COVID-19 pandemic on various aspects, including food supply, mental health, the number of victims, and decision-making in health services. This approach involves data processing using data exploration techniques (EDA), data preprocessing, and predictive algorithms to gain in-depth insights from the available data.

A. Data Collection

The data utilized in this study were obtained from publicly available and relevant secondary sources, including datasets related to food supply, mental health, and COVID-19 case statistics (confirmed, recovered, deceased, and active cases). These datasets were collected from trusted repositories such as Google Scholar, Kaggle, and other relevant platforms that support comprehensive analysis. The research was conducted through both offline and online means, utilizing secondary data collected between December 2024 and January 2025. All data analysis was performed at the researcher's home and campus, using data processing tools and software such as Python and its associated libraries to facilitate preprocessing, visualization, and modeling.

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

B. Data Processing and Analysis Process
The stages of data processing and analysis are carried out as follows:

1. Data Preprocessing
   Data preprocessing is a critical step in data analysis and machine learning, aimed at improving data quality and ensuring datasets are suitable for further analysis. This process includes data cleaning (removing inaccuracies, duplicates, and missing values), data transformation (such as normalization and encoding), and data reduction (through feature selection and dimensionality reduction) to prepare raw data for effective model training (Varma et al., 2023; Y et al., 2022). Although essential, many data science projects fail due to inadequate preprocessing. This highlights the importance of automated tools that streamline these tasks and reduce manual errors (Kureljusic & Karger, 2021; Talayero et al., 2022). Proper preprocessing significantly impacts the reliability and performance of machine learning models.
2. Data Exploration
   Data exploration is a dynamic and iterative process that enables users to interact with large datasets to uncover insights and generate hypotheses. Unlike traditional analysis, this process often occurs without predefined objectives, making it challenging for systems to predict user intent (Saha et al., 2024). As users discover new patterns, they frequently adapt their exploration strategies, highlighting the need for systems that can accommodate and respond to changing user behavior. Understanding these behavioral shifts is essential for designing more effective data exploration tools. Recent studies have proposed various methodologies to automate and enhance the data exploration process. Reinforcement learning has been successfully applied to optimize user interaction with data, while meta-learning approaches aim to reduce the number of iterations required for effective exploration (Amer-Yahia, 2024; Cao et al., 2022). Effective data exploration is crucial across sectors, supporting informed decision-making by helping analysts identify key attributes and relationships prior to applying machine learning models (Battle, 2022; Nagaraju et al., 2024). However, without proper guidance, users may become overwhelmed by the sheer volume of data, which can hinder their ability to make informed decisions.
3. Dimensionality Reduction
   Using algorithms such as t-SNE to reduce the dimensionality of data to facilitate visualization. t-Distributed Stochastic Neighbor Embedding (t-SNE) is a widely used technique for dimensionality reduction and data visualization, particularly effective in preserving local structures within high-dimensional datasets. By leveraging a Student's t-distribution to model pairwise similarities, t-SNE improves upon its predecessor, SNE, in terms of stability and learning speed. It maps high-dimensional data into a lower-dimensional space—typically two or three dimensions—while minimizing a cost function based on Kullback-Leibler divergence, which measures the divergence between the original and reduced-space probability distributions (Kimura, 2021; Kawase et al., 2022). Despite its strengths, t-SNE faces challenges such as computational complexity and sensitivity to random initialization. Several enhancements and variants have been proposed to address these limitations. Generalized t-SNE incorporates information geometry to optimize parameter selection, while parametric t-SNE integrates neural networks to improve mapping efficiency, particularly in quantum datasets (Kimura, 2021; Kawase et al., 2022). Supervised t-SNE introduces class labels and alternative distance metrics to enhance classification accuracy, outperforming conventional approaches (Neto et al., 2024). Furthermore, improvements such as informative initialization strategies and optimized kernel selection have been shown to stabilize the output, making t-SNE a powerful yet parameter-sensitive tool in exploratory data analysis (Chourasia et al., 2022).
4. Prediction and Analysis
   The prediction and analysis stage involves the application of machine learning algorithms to model and quantify the impact of the COVID-19 pandemic on critical sectors such as mental health, food supply, and mortality rates. Predictive models enable the identification of patterns and trends within the data, allowing for the estimation of future outcomes based on historical

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

and current variables. In this study, algorithms such as K-Means clustering and dimensionality reduction techniques (e.g., t-SNE) are utilized to categorize countries based on similarities in pandemic indicators and nutritional availability. This process provides valuable insights into how different regions have been affected and helps guide targeted interventions and policy decisions.

5. Data Visualization
   Data visualization serves as a crucial component of the analytical process, transforming complex datasets into easily interpretable visual formats such as bar charts, scatter plots, line graphs, and heatmaps. These visual tools aid in communicating key findings clearly and concisely, supporting intuitive understanding for both technical and non-technical stakeholders. Effective visualization allows for the identification of anomalies, trends, and relationships that may not be immediately evident through raw numerical data. In the context of this study, visualizations are used to map the spread of the virus, assess food supply fluctuations, and illustrate mental health trends, thereby enhancing the decision-making process and promoting data-driven strategies in pandemic response planning.

RESULTS AND DISCUSSION

The COVID-19 pandemic has had a broad and complex impact on various aspects of people's lives. Through this study, the results obtained include analysis of food supply, mental health, number of victims, and decision-making in health services. Data that has been processed using various visualization methods and data analysis shows significant patterns in these various aspects.

A. Covid-19 data mining visualization

According to research conducted by (Muhammad Muslim, 2023), the use of data visualization in COVID-19 screening on an internal agency scale has succeeded in integrating the digital screening process with data mining using the classification tree algorithm. This study produced various visualization models, such as mapping employee mobility interactions, residential locations around DIY, and mobility patterns during or after work holidays. By using the Orange tool, this study shows that data visualization can be the first step in developing data mining in the health sector, although it is still small-scale and internal to the agency.

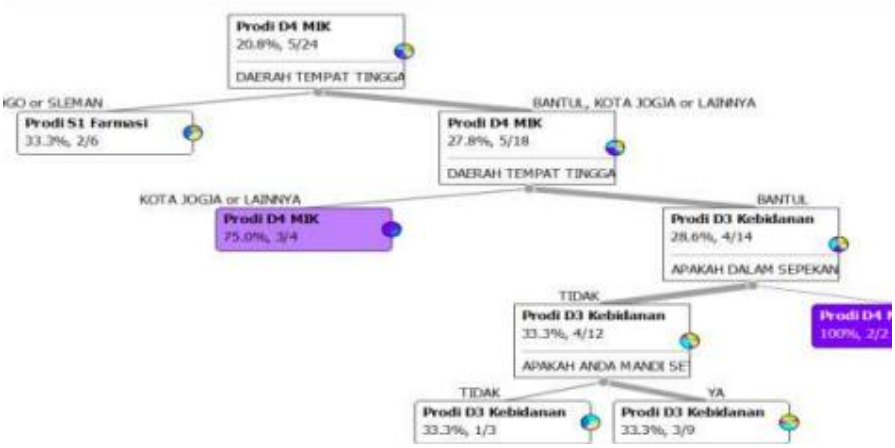

Figure 3.1 Covid-19 mapping tree scheme

The research conducted obtained a mapping tree scheme from the covid-19 screening data that had been mined. The scope of the resulting mapping is data from several units that have been sampled into the classification process. Where this mapping process will affect the classification results on the visualization results of data mining using orange as shown in Figure 3.1. The classification

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

scheme obtained through the orange feature sorts and maps the position of each unit in the agency which is adjusted to the percentage obtained during the covid-19 screening process. The existence of a tree-shaped sorting as in Figure 3.1 will make it easier to recognize knowledge from the characteristics of the clustering.

B. The role of data visualization in improving health service facilities

Data visualization plays a critical role in supporting effective decision-making in healthcare. By presenting data in an intuitive and easy-to-understand manner, visualization helps identify patterns, trends, and anomalies that are relevant to improving the quality of healthcare facilities. In the context of the COVID-19 pandemic, data visualization has become a strategic tool for mapping healthcare facility needs, allocating resources efficiently, and formulating policies that are responsive to emergency situations.

Based on research conducted by (Syifa Nurhaliza, 2024) there are several studies to explore the role of data visualization in improving health services, with the following examples:

1. Literature Review: The results of the literature analysis show that data visualization has become an integral part of modern healthcare management.
2. Survey or Interview: A survey conducted among healthcare professionals showed that the majority of respondents recognized the importance of data visualization in their clinical practice.
3. Secondary Data Analysis: Secondary data analysis of electronic medical records shows an increasing use of data visualization in patient management.
4. Case Studies: Through case studies in several healthcare institutions, it was found that the use of data visualization has resulted in increased efficiency in diagnosing diseases and various other health problems.
5. Tool Development and Evaluation: Development of a data visualization tool specifically for healthcare purposes shows potential in improving healthcare management.

Overall, the results of the study show that data visualization helps improve healthcare services. From the results of literature studies, surveys, secondary data analysis, case studies, to tool development, it was found that data visualization provides smart technology and significant contributions to improving the clinical decision-making process, facilitating more effective patient monitoring, and planning more targeted public health intervention strategies.

C. Mental impact of Covid-19

The presence of this outbreak or virus has an impact not only on a global level but also on Indonesian society, not only in terms of physical, but also psychological (mental) impacts caused by several factors and various problems and anxieties. There are several mental health disorders reported after individuals undergo Covid-19 treatment. These mental health disorders include anxiety disorders, mood disorders such as depression, and post-traumatic stress disorder (Post Traumatic Stress Disorder) (Nasrullah, 2021: 207).

According to the results of a study conducted by Aulia Hera Yuanti (2024), the COVID-19 pandemic has had a significant impact on mental health, triggering various anxieties and psychological problems. The study used the RapidMiner application to visualize data through text processing, producing a word cloud that showed the word "Covid" as the most frequently occurring, and described positive, negative, and neutral sentiments in the data.

The visualization results using worldcloud indicate that the covid issue is an important aspect in the realm of mental health for the community. Sentiments on this issue vary widely and depend on the

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

views of individuals and groups. The word "CORONA" also appears 3 times. This shows that corona is also part of covid-19. If visualized with the wordcloud display in the RapidMiner application as in the following image.

Figure 3.2 COVID as a word with the highest frequency

Based on the image above, the wordcloud displays the words with the highest frequency, the higher the frequency of the word, the greater it is compared to other words. The word with the highest frequency is Covid, and the others show low frequencies. So this indicates that Covid/Corona is one of the factors that has a major impact on people's mental health.

D. The global impact of Covid-19 in terms of mortality, distribution of food supplies and distribution of protein quantities.

The COVID-19 pandemic has had a significant impact globally, both in terms of increased mortality and disruptions to food supplies and nutrient distribution. This health crisis is affecting the availability and distribution of food resources, including protein, which is vital to meeting people's nutritional needs. Strains in food supply chains and mobility restrictions are exacerbating inequalities in food access, resulting in long-term impacts on food security and public health in many parts of the world.

a. Initial Data

This study uses three main datasets taken from Kaggle, namely COVID-19 CountryWise.csv, Food_Supply_Quantity_kg_Data.csv, and Protein_Supply_Quantity_Data.csv. The COVID-19 CountryWise.csv dataset contains information related to the number of confirmed cases, deaths, recoveries, and demographic data such as population in various countries. This dataset is relevant to analyze the global impact of the pandemic on the number of deaths and the distribution of cases in various regions. Meanwhile, Food_Supply_Quantity_kg_Data.csv contains food supply data in kilograms per capita for various food categories in each country, providing insight into food availability during the pandemic. Meanwhile, Protein_Supply_Quantity_Data.csv provides details of the distribution of protein from various food sources, such as meat, milk, and grains, which is important for understanding the impact of the pandemic on people's nutritional adequacy.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

These three datasets complement each other in the study to analyze the relationship between the COVID-19 pandemic and health and food supply. COVID-19 case data provide context for the immediate impact of the pandemic, while food and protein supply data help illustrate the subsequent effects on food security and nutrition distribution. The combination of these three datasets supports a comprehensive exploration of the impact of the pandemic, both globally and locally, and provides a basis for relevant data visualization to support decision-making in the health and food distribution sectors.

b. Data preprocessing

In this study, the data preprocessing process was carried out on the COVID-19 CountryWise.csv dataset to ensure that the data used is relevant and significant to the analysis to be carried out. One of the techniques applied is data reduction using the Chi-Square-based feature selection method. This method aims to select the most relevant features with the target variable, namely Total Deaths. By using SelectKBest from the scikit-learn library, the selection process is carried out with the parameter k = 4, resulting in four (starting from zero) significant main attributes. The initial data without the Total Deaths column is processed together with the target variable to identify the most statistically relevant features.

	Total Cases	Total Cases per 100k pop	New Cases (7 days)	New Cases (24 hours)
0	94152573	28444.658	416281	76462
1	44516479	3225.822	37843	6422
2	34544377	16251.633	53446	10420
3	33766090	51916.394	150781	36147
4	32604993	39204.380	203649	32168

Figure 3.3 Attributes that contribute the most to predicting the number of deaths due to Covid-19
Source: Research, 2025

The results of this feature selection show that the attributes Total Cases, Total Cases per 100k pop, New Cases (7 days), and New Cases (24 hours) are the features that contribute the most to predicting the total number of deaths due to COVID-19 as seen in Figure 3.3. The resulting data output displays the values of these four features for the first few rows. This result narrows the scope of the analysis, allowing for a more efficient data exploration process and predictive model development that focuses on features that have a significant impact.

In the next dataset, data preprocessing is carried out by selecting the SelectKBest method with the f_classif score function because the dataset used is continuous data. This process aims to select the two best features that have the most significant relationship with the target variable (target column). In its implementation, all feature columns (X) and targets (y) are taken from the dataset for analysis. Then, the SelectKBest method is applied to evaluate the score of each feature based on the ANOVA F-value statistical function. After the selection process, the two best features selected were "Cereals - Excluding Beer" and "Fruits - Excluding Wine".

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

	Cereals - Excluding Beer	Fruits - Excluding Wine
0	24.8097	5.3495
1	5.7817	6.7861
2	13.6816	6.3801
3	9.1085	6.0005
4	5.9960	10.7451
...	...	...
165	12.9253	7.6460
166	16.8740	5.9029
167	27.2077	5.1344
168	21.1938	1.0183
169	22.6240	2.2000

Figure 3.4 The two best features resulting from feature selection.
Source: Research, 2025

In Figure 3.4, the results of this process are displayed in the form of a new dataset that only includes the two selected features. This data describes the supply of cereals and fruits in various countries in kilograms per capita. By reducing the number of features, the complexity of the data was successfully minimized without sacrificing significant information, so that further analysis can be more focused and efficient. The output dataset displays the reduced food supply data, which can be used for further analysis stages such as data exploration or data visualization.

Meanwhile, in the Protein_Supply_Quantity_Data.csv preprocessing data, an example of normalization of the dataset is taken using the MinMaxScaler method from the sklearn library. This technique is used to convert values in the dataset to a scale between 0 and 1. This normalization is important to ensure that the data has a uniform scale so that the algorithms used in the analysis, such as predictive models or data visualization, can work optimally without bias towards variables with a larger range of values. This process involves transforming the feature values (columns) in the dataset into a predetermined range.

```
 [0.767366   0.4821016 ]
 [0.46240173 0.24967801]
 [0.21412453 0.32354084]
 [0.58199037 0.15481958]
 [0.44295705 0.81815394]
 [0.50539261 0.20973117]
 [0.78517965 0.13202494]
 [0.65534521 0.23526526]
 [0.08708933 0.04452622]
 [0.44931838 0.31164265]
 [0.32535374 0.14954513]
...
 [0.51039103 0.72748646]
 [0.17734177 0.07803332]
 [0.17352114 0.13100276]
 [0.21986026 0.08839824]]
```

Figure 3.5 Results of applying MinMaxScaler
Source: Research, 2025

The output in Figure 3.5 shows the following application of MinMaxScaler on the prepared dataset, where the transformation results are displayed in the form of a matrix with normalized values. The output shows that each element in the dataset is now in the range of 0 to 1, as seen from the printed matrix. This normalization allows further analysis to be carried out more consistently, such as when using predictive algorithms or comparing data distributions between countries.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

c. EDA

Exploratory Data Analysis (EDA) is performed to identify data that is outside the normal limits, namely data that exceeds the upper limit or is below the lower limit. This analysis helps identify outliers that can provide further insight into patterns or anomalies in the dataset. Based on statistical analysis, it was found that there are several countries that have the number of cases and deaths due to COVID-19 that are outside the upper limit. These countries include North Macedonia, Hungary, Bosnia and Herzegovina, Bulgaria, and Peru. This data includes the total number of cases, deaths, and the rate of spread per 100 thousand population. For example, Bulgaria has a death rate per 100 thousand population of 541.80, while Peru recorded the highest total cases of 4,127,612.

	Country	Region	Total Cases	Total Cases per 100k pop	New Cases (7 days)	New Cases per 100k pop (7 days)	New Cases (24 hours)	Total Deaths	Total Deaths per 100k pop	New Deaths (7 days)	New Deaths per 100k pop (7 days)	New Deaths (24 hours)
96	North Macedonia	Europe	341889	16410.30	628	30.14	141	9516	456.76	8	0.38	1
42	Hungary	Europe	2070443	21192.87	11596	118.70	0	47409	485.27	42	0.43	0
95	Bosnia and Herzegovina	Europe	397602	12119.00	765	23.32	135	16104	490.85	18	0.55	2
55	Bulgaria	Europe	1250250	17985.37	3585	51.57	544	37663	541.80	31	0.45	5
32	Peru	Americas	4127612	12518.59	8820	26.75	1150	216173	655.63	191	0.58	22

Figure 3.6 shows country data as outliers with cases outside the upper limit.
Source: Research, 2025

On the other hand, no data was found below the lower limit. This shows that all countries analyzed have data above the minimum threshold. The presence of outliers above the upper limit is important for further analysis because it can reflect extreme situations that require special attention. Outliers like this can be used to identify specific patterns related to the impact of COVID-19 on a country, both in terms of public health and its impact on food supply and nutrient distribution. Data visualization from the EDA results will be presented to facilitate interpretation and data-based decision making.

Exploratory Data Analysis (EDA) analysis also includes evaluation and comparison of datasets before and after outlier removal. In the initial stage, data was found to be outside the upper limit, which was 456.26, while no data was below the lower limit (-250.63). This is indicated by the presence of outliers in the scatter plot and box plot visualizations, where there are several points that are far outside the normal data range. The presence of these outliers can affect the results of the analysis, especially in the interpretation of distribution and trend data.

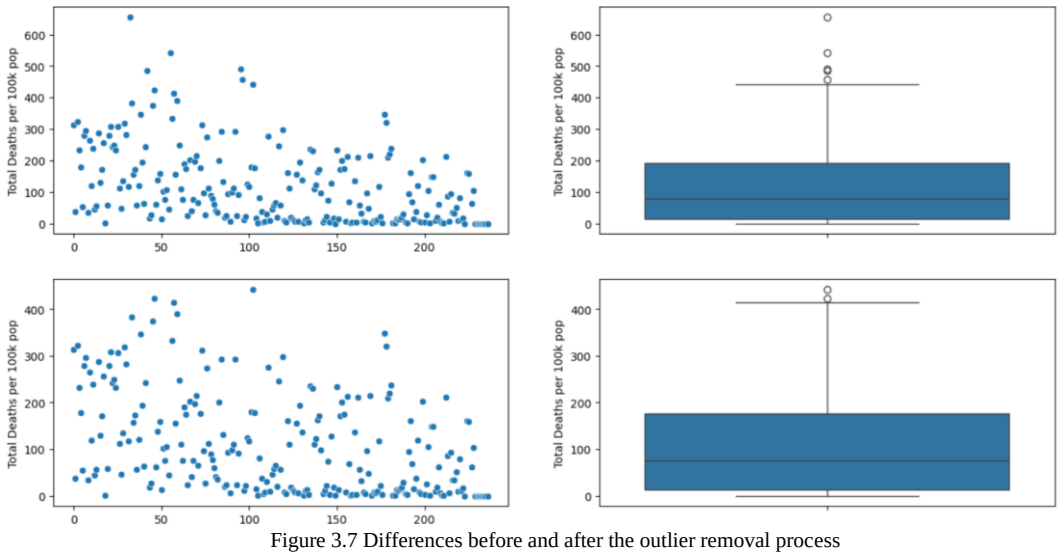

Figure 3.7 Differences before and after the outlier removal process
Source: Research, 2025

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

After the outlier removal process (figure 3.7), the dataset becomes cleaner with all data within the normal range, with some data still in the upper limit but not exceeding the upper limit. Visualization after outlier removal shows a more centered scatter plot and box plot, without any striking anomalies. This process is important to improve the accuracy of further data analysis, especially when creating prediction or inference models. With a dataset free from outliers, the analysis results are better and more representative to describe the actual conditions.

d. Predictive algorithm data

K-Means clustering is an algorithm that is often used to group data into several groups (clusters) based on certain characteristics. In this study, the K-Means algorithm was applied to the Food_Supply_Quantity_kg_Data.csv dataset which includes various attributes, such as food consumption, animal products, vegetable products, obesity rates, malnutrition rates, and pandemic indicators such as the number of confirmed cases, deaths, and recoveries. These attributes provide a comprehensive view of food distribution patterns across countries and their relationship to public health conditions. With the K-Means algorithm, this dataset is clustered based on similarities in food distribution patterns and other indicators, allowing the identification of groups of countries with similar characteristics.

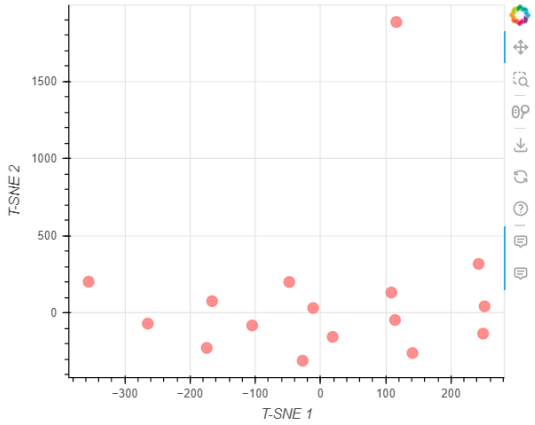

Figure 3.8 Bokeh results from implementing the clustering method
Source: Research, 2025

The selection of the K-Means algorithm itself as a clustering method is the result of an analysis based on its efficiency in handling high-dimensional datasets and its ability to provide well-defined clustering results. K-Means uses a centroid-based approach to minimize variation within clusters, resulting in optimal group division. This is very relevant to understanding complex food distribution patterns among countries. By using the clustering method, this study can more clearly describe the relationship between food supply, health, and population, thus helping to design more effective distribution policies.
Clustering-based predictive algorithms, such as K-Means, are used in this study to group countries based on similarities in food consumption patterns and the impact of the COVID-19 pandemic. For example, two countries, Chad and Ecuador, were selected from the dataset for further analysis. Data related to these two countries include indicators such as food supply (e.g., consumption of meat, fruits, and animal protein), obesity prevalence, and malnutrition rates. In addition, the t-SNE (X, Y) coordinates representing spatial relationships in low-dimensional space are also displayed. This information provides an initial overview of the differences or similarities in the characteristics of the two countries in the context of the pandemic and food supply.

JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

	Country	Alcoholic Beverages	Animal fats	Animal Products	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Obesity	Undernourished	Confirmed	Deaths	Recovered	Active	Population	Unit (all except Population)
0	Chad	0.8297	0.2023	12.2804	0.0	21.6017	0.0495	1.0451	1.0669	3.8543	...	4.8	37.5	0.020578	0.000741	0.016881	0.002957	16877000.0	%

1 rows × 34 columns

[['Chad' 249.2390594482422 -134.38507080078125]]

	Country	Alcoholic Beverages	Animal fats	Animal Products	Aquatic Products, Other	Cereals - Excluding Beer	Eggs	Fish, Seafood	Fruits - Excluding Wine	Meat	...	Obesity	Undernourished	Confirmed	Deaths	Recovered	Active	Population	Unit (all except Population)
1	Ecuador	3.2929	0.1277	15.3469	0.0328	14.9132	0.6902	0.6893	8.039	4.0115	...	19.3	7.9	1.468306	0.085683	1.198184	0.184438	17511000.0	%

1 rows × 34 columns

[['Ecuador' 108.48553466796875 132.05773925781125]]

Figure 3.9 Differences in t-SNE coordinate values between the two countries
Source: Research, 2025

In the program output, the analysis results show that Chad and Ecuador have quite different t-SNE coordinate values, namely (249.239, -134.385) for Chad and (108.486, 132.058) for Ecuador. These results indicate that although both countries may have similar food consumption patterns in some aspects, significant differences remain in the context of other indicators. Through the K-Means algorithm, this data can be grouped into certain clusters that help identify common patterns between countries. This kind of analysis is very relevant to understand how the pandemic affects food security in different regions and helps in better planning of resource distribution.

CONCLUSION

The COVID-19 pandemic has had a wide impact on various sectors of life, including mental health, food supply, and health service management. The results of the study showed that people's mental health experienced an increase in problems, such as anxiety and depression, which can be seen from sentiment analysis in data visualization. The decline in food supplies that have an impact on global food security, as well as disruptions in nutrient distribution, have also worsened the public health situation. In addition, the role of data visualization in decision-making has also proven important for mapping health facility needs and optimizing resource allocation. The use of algorithms such as classification trees in previous studies has shown the potential for data visualization as an effective tool in improving the understanding and management of health data. However, this approach is still limited to a small and internal scale, so it needs to be expanded to provide greater benefits in a global crisis. The success of data visualization in COVID-19 screening shows the potential for further development in the field of data mining in the health sector, but there are limitations in terms of scale and data accessibility.

As a follow-up to this research, it is important to carry out further development in terms of the application of data visualization at a broader and more integrated level. Further research can be focused on the development of more sophisticated predictive analysis methods to map the potential impact of the pandemic on various aspects of life more accurately. In addition, attention should be paid to the development of platforms that can integrate data from various sources to provide a more comprehensive picture of the impact of the pandemic. Increasing collaboration between institutions, both at the local, national, and international levels, will be very important to strengthen food security and mental health in the future. The use of technology and data analytics must be optimized to support policies that are responsive to community needs and reduce negative impacts in the future.

REFERENCES

Amer-Yahia, S. (2024). Intelligent agents for data exploration. Proceedings of the VLDB Endowment, 17, 4521–4530. https://doi.org/10.14778/3685800.3685913

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Battle, L. (2022). Behavior-driven testing of big data exploration tools. Interactions, 29(5), 9–10. https://doi.org/10.1145/3554726

Cao, Y., Xie, X., & Huang, K. (2022). Learn to explore: On bootstrapping interactive data exploration with meta-learning. arXiv, abs/2212.03423. https://doi.org/10.48550/arXiv.2212.03423

Chourasia, P., Ali, S., & Patterson, M. (2022). Informative initialization and kernel selection improves t-SNE for biological sequences. In 2022 IEEE International Conference on Big Data (pp. 101–106). https://doi.org/10.1109/BigData55660.2022.10020217

Dhummad, S. (2025). The imperative of exploratory data analysis in machine learning. Scholars Journal of Engineering and Technology, 13(1), 30–44. https://doi.org/10.36347/sjet.2025.v13i01.005

Firiza, S.-M., Scornea, A.-M., & Topîrceanu, A. (2024). Visualizing the COVID-19 pandemic: Retrospective insights and implications for public health strategies. 2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI), 000461–000466. https://doi.org/10.1109/saci60582.2024.10619727

Galatro, D., & Dawe, S. (2023). Exploratory data analysis. In Exploratory Data Analysis (pp. 13–57). Springer. https://doi.org/10.1007/978-3-031-46866-7_2

Isbaniah, F., & Susanto, A. D. (2020). Pneumonia corona virus infection disease-19 (COVID-19). Journal of the Indonesian Medical Association, 70(4), 87–94. https://doi.org/10.47830/JINMA-VOL.70.4-2020-235

Kawase, Y., Mitarai, K., & Fujii, K. (2022). Parametric t-Stochastic neighbor embedding with quantum neural network. Physical Review Research, 4(043199). https://doi.org/10.1103/physrevresearch.4.043199

Khan, M. I. (2022). COVID-19: A pandemic outbreak. International Journal of Scientific and Management Research, 5(3), 242–260. https://doi.org/10.37502/ijsmr.2022.5319

Khanam, F., Nowrin, I., & Mondal, M. R. H. (2020). Data visualization and analyzation of COVID-19. Journal of Scientific Research and Reports, 26(3), 42–52. https://doi.org/10.9734/JSRR/2020/V26I330234

Kimura, M. (2021). Generalized t-SNE through the lens of information geometry. IEEE Access, 9, 129619–129625. https://doi.org/10.1109/ACCESS.2021.3113397

Ma, P., Ding, R., Wang, S., Han, S., & Zhang, D. (2022). XInsight: Explainable data analysis through the lens of causality. ACM Transactions on Intelligent Systems and Technology, 13(2), 1–27. https://doi.org/10.1145/3589301

Neto, A., Levada, A. L. M., & Haddad, M. F. C. (2024). Supervised t-SNE for metric learning with stochastic and geodesic distances. Canadian Journal of Electrical and Computer Engineering, 47(1), 1–7. https://doi.org/10.1109/icjece.2024.3429273

Nasrullah., & Sulaiman, L. (2021). Analysis of the impact of COVID-19 on mental health of the community in Indonesia. Indonesian Public Health Media, 20(3), 206–211. https://doi.org/10.14710/mkmi.20.3.206-211

Nurhaliza, S. (2024). The role of data visualization in improving health services. SmartTechnology.org, 4(5).

Yuanti, A. H. (2024). Analysis of the impact of COVID-19 on mental health with RapidMiner data visualization. Multidisciplinary Journal of Science Warehouse, 2(1), 183–187. https://doi.org/10.59435/gjmi.v2i1.225

Muslim, M., & Syamsu, W. (2023). Data mining visualization for COVID-19 digital screening in institutions. KAPUTAMA Information Engineering Journal (JTIK), 7(1), 151–156.

Sinha, A., Singh, S., Shekh, M., & Kumari, P. V. (2022). COVID-19 analysis using chest X-rays and CT-scan. I-Manager's Journal on Software Engineering, 16(4), 1–6. https://doi.org/10.26634/jse.16.4.18756

Song, X. (2023). A brief introduction to exploratory data analysis. Advances in Engineering Technology Research, 4(1), 420–430. https://doi.org/10.56028/aetr.4.1.420

Taheri, S. (2020). A review on coronavirus disease (COVID-19) and what is known about it. Journal of Health, 11(1), 87–93. https://doi.org/10.34172/DOH.2020.09

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Sagala, N., & Aryatama, F. Y. (2022). Exploratory data analysis: A study of Olympic medallist. Jurnal Sistem Informasi, 11(3), 578–586. https://doi.org/10.32520/stmsi.v11i3.1857

Sandfeld, S. (2023). Exploratory data analysis. In Data Science Methods (pp. 179–206). Springer. https://doi.org/10.1007/978-3-031-46565-9_9

Y, C., Kiran, P., & B, M. (2022). The novel method for data preprocessing CLI. In Proceedings of the International Conference on Advanced Intelligent Systems and Technology, 117–120. https://doi.org/10.53759/aist/978-9914-9946-1-2_21

Talayero, A., Yürüşen, N. Y., Sánchez Ramos, F. J., & Gastón, R. L. (2022). Machine learning-based met data anomaly labelling. Journal of Physics: Conference Series, 2257(1), 012015. https://doi.org/10.1088/1742-6596/2257/1/012015

Varma, D. R. N., Nehansh, A., & Swathy, P. (2023). Data preprocessing toolkit: An approach to automate data preprocessing. Indian Scientific Journal of Research in Engineering and Management, 7(3). https://doi.org/10.55041/ijsrem18270