International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I5P103

Volume 72 Issue 5, 27-32, May 2024

Original Article

The Importance of Data-Driven Decision-Making in Public Health

Keyur Patel

New Jersey Public Health Association, USA.

[1]*Corresponding Author : keyur.bme@gmail.com*

Received: 03 March 2024 Revised: 07 April 2024 Accepted: 22 April 2024 Published: 10 May 2024

Abstract *- A large portion of the public health frameworks is taking transformations into data-centred techniques as a base for proper decision-making targeting the complicated nature of modern health issues. In recent years, the old-fashioned methods of handling big data related to public health have yet to practically encompass the amount of data emerging from the computerization of medical reporting. This article will look at how cloud-based data analytics can transform people-based decision-making into revolution and modern approaches to the current public health decisions that are being made. The objective of the report was to lay the basis of the literature on the significance of data-driven decision-making for public health, traditional viewpoint obstacles, action of cloud computing, and research studies on cloud computing-based analytics in the public health setting. Therefore, the novella goes along from being data for acquisition and preprocessing to having a model to be developed, trained, and tuned up until it is ready to be deployed and maintain sufficient control of data integrity, security, privacy compliance, including collaborative collection and sharing of information, and continuous monitoring and further update. This highlights that public health organizations are likely to build their capability of using evidence-based data that can be spread globally and in a timely manner, keeping in mind the health outcomes of the whole world's population delicately through cloud-based analytics.*

Keywords *- Big data, Cloud computing, Data engineering, Public health.*

1. Introduction

An unsung hero in public health is data-driven decision-making, which works quietly in common perception in the world today [1]. In environmental circumstances, the capability to obtain and analyze massive data sets and then turn that data into effective strategies is becoming a cardinal factor in the mitigation of present-day health risks. Evidence-based decisions in all circumstances of health problems, such as infectious diseases, chronic conditions, and health crises, help to eliminate the need for guesswork in interventions and also ensure funds do not get misused.

Nevertheless, traditional data analysis techniques may need to be more robust when big data and insights from diverse sources get involved. Consequently, there is an essential urgency to investigate new and more effective ways that can liberate the full potential of these huge data sets so as to improve public health decision-making.

1.1. Problem Statement

Human error, which is characteristic of the old approach of data analysis, becomes unavoidable as one attempts to fight the barrage of complex data in public health today. Normally, the existing system is flexible but only needs a little computational power to do downstream data processing, which is usually from diverse information sources [2].

This will certainly weaken the efficiency of grasping key information on time, which may then trigger a delay in the provision of timely and effective measures during the pandemic. However, the brittle nature they have hampers the flexibility and ability of other programs to expand and dovetail in delivering care.

1.2. Objective

The present scientific article is devoted to studying the prospect of cloud computing being used for analytical data treatment to overcome some of the restrictions that are imposed on traditional methods in public health decision-making. Cloud-based solutions are described as the ones that pave the way to stability, flexibility, and data analytic capabilities advanced by cloud computing.

However, it is also good to remember that this will not end only as a revolution in data processing and use, as it will eventually pass into a public health revolution [2]. This paper debunks the pragmatism, viability, and illustrations of the application of cloud-based data analytics frameworks in diverse public health settings.

1.3. Significance

With the prior achievements of data science and Artificial Intelligence (AI), cloud-based data analytics is foreseen to have a huge capacity to revolutionize how decision-making processes are conducted. Learning from the full capacity of cloud computing as a tool for decision-making, stakeholders might have the capacity to observe trends from high volume and diverse datasets, thus helping in setting up early warning systems and responding to emergency crises [5]. This will, therefore, encompass predictive modeling with machine learning techniques to enhance the availability of interventions based on communal needs and personal health characteristics. Cloud-based systems come along with benefits ranging from enhanced collaboration and data management to even better interoperability. The result is a better public health ecosystem altogether [2]. In brief, not only does this technology revolutionize our disease focal point and response, but it could also be an instrument to increase the efficiency of resource allocation, which would be a great contribution to the vast improvement of health results.

2. Literature Review

2.1. Current State

Generally, traditional approaches to public health decision-making mostly use several methods that can provide data analysis. Such methods involve statistical analysis, epidemiological modeling, and data visualization techniques. Although the tools are beneficial for understanding patterns of diseases, identifying risk factors, and guiding intervention strategies, they have limits in handling the increasing volume, variety, and velocity of public health data [2]. In addition, all traditional analytical tools are static, making real-time decision-making difficult or rendering intervention strategies impossible to deploy due to lateness.

2.2. Challenges in Public Health

There have been several challenges to the effectiveness of public health data analytics [4]. For instance, privacy concerns may limit access to datasets and sometimes only allow collaborative access by stakeholders. This would pose severe technical integration challenges in information coming from a variety of sources, including electronic health records, environmental monitoring systems, and social media, among others. This further increases the scalability challenge of traditional data analysis methods as the dataset sizes and complexities grow, leading to delayed processing and analysis.

2.3. Role of Cloud Computing in Data Analytics

Cloud computing is a different model through which data is analyzed and used to deliver scalable, flexible, and relatively inexpensive computing resources over the Internet [3]. On the flip side, cloud computing is elastic, and thus, the computational resources for public health organizations can now be run more dynamically and resized than before. It thus enables the processing of large datasets in real time [3]. Cloud-based platforms further provide more functionalities for data storage, sharing, and collaboration that become very useful in the integration of different sources of data and interdisciplinary collaboration between organizational units engaged in public health initiatives [5]. With a pay-as-you-go model of cloud services, capital costs are reduced, and cost-effective access to leading analytics tools and techniques is possible.

2.4. Previous Studies on Cloud-Based Analytics in Public Health

The general use of a data-driven manner of work in public health has particularly resulted in the effectiveness of grave health interventions noticeable [6]. Research shows that we are quite good at detecting and presenting data on visualization tools that show an impact on decision-making by the way they help in understanding attitude perception. In addition, the methodology of system dynamic modeling has enabled the prediction of the number of health inputs and care outputs and shown the result of interventions on life expectancy healthy life expectancy, which has led to the establishment of prevention as a necessary investment. Besides this, the Data for Decision-Making (DDM) program improves the health department staff's ability to collect and use high-quality data. The improvement outcome ranges from amending some policies to basing all strategies on data. The holistic formulation of quantitative evidence in public health policy guidelines is also developing, as the use of meta-analysis is increasing to assist policymakers in their decision-making process with more comprehensible information.

3. Materials and Methods

3.1. Data Sources

This method-based study uses the overall data from WHO to analyze data from different authorities such as the CDC, and these information sources are what the world considers injury, crime, and vaccination statistics [5]. Beyond the options related exclusively to information, such as news, sports, and current events, the possibilities expand the realm of entertainment. It has Southeast University's hospitals' electronic health records, CDC's National Notifiable Diseases Surveillance System, which reveals the medical data on the diseases, as well as other ecological and demography data of work, as well as Data World. On top of that, WHO and data.gov datasets are vastly different in terms of global awareness of public health matters, and the latter is centered on the problems within the US only. We set up as criteria the datasets that needed to be highly applicable to the research purpose, and that should allow the data to be analyzed deeply within ethical and privacy limits, which are still central factors. In this case, sensitive data of individual patients would be kept secret to avoid disclosure to the general public. Research rigor might be affected, but it does not matter to the public. The way this knowledge can be acquired may be different, but it does not impact the base knowledge itself. Without a doubt, this information will pinpoint the real names of rape victims, people with HIV also, and others who are not ready to be identified, thus portraying them as vulnerable, hence the decreased chances of them getting a job [7]. These discoveries were inspirational to have an actual live case where the fingerprints are the ones that are not real and if the real person is dead or alive.

3.2. Cloud Analytics Platforms

Each entity has its decision to adopt the cloud's reliance as the cloud analytics platform will be the one to be chosen. One of the considerations will touch on a few factors: scale, execution capacity, working reliability, cost, and machine stability, as well as connecting the machines with analytics tools. There are lots of great open-source cloud platforms, which are worth considering, like PostHog, Matomo, Metabase, Plausible Analytics, Grafana, Umami, OpenReplay, GrowthBook, Redash, Apache Superset, County, and Open Web Analytics, among others. They are not only good on their own, but they can also be easily customized for the analytic needs using Different platforms offering strong ownership styles of the cloud platforms like PowerBI, Tableau, and Looker, provide the ability to data visualization, data integration, and collaboration which is one of the most powerful set of tools [8].

It is worth mentioning that it has a number of benefits, for instance, exemplary customer service, as well as thorough documentation and advanced functions that allow task management on a whole new level – first and foremost, the midsize-to-enterprise level businesses that need lots of commands at hand.

3.3. Analysis tools

PostHog can run on cloud platforms like AWS, GCP, or Azure, which gives a space for capturing and analyzing user interaction on a website while using the application [7]. Integrating the systems leads the company to scale its capabilities with infrastructural development and utilize emerging cloud-based data storage and computing services.
Opening Matomo as a tool gives the power to choose whether to go with cloud services like Cloud and AWS to have it doubled for performance and user behavior tracking [9]. This integration allows us to take advantage of cloud computing resources to deal with user speed and data analysis.

Metabase, which can be set up on cloud platforms like AWS, GCP, and Azure, allows for easy and streamlined queries and data visualization in a user-friendly manner [10]. It can link up to the databases residing on these cloud services by the provision of the platform so as to analyze all cloud data.

They have chosen to do this through a plausible, privacy-centric design that is compatible with cloud infrastructure. This, in turn, enables real-time analytics without having to do away with privacy. The cloud architecture gives this level of structure that data takes care of and unfailing compliance with privacy law [8]. Taking electoral processes approaching even the village level, poll referenda, and presidential campaigns, social media is all around, changing and defining them.

Grafana can help you to compute cloud data sources on AWS, GCP, or Azure and how to build reports and visualizations. This is undoubtedly instrumental in the cloud infrastructure monitoring of the metrics and the application data. In such a way, it eliminates the necessity of spending tedious time scrolling through several systems to get all the needed information [9]. It comes in one convenient dashboard format.

Social networks and cloud platforms can be marketed. Analytical features are available that provide a website analytics overview by tracking visitors' interests [11]. Cloud platforms make integration favorable to meet the scale and need of concurrency as the software grows to match the data size and user demands.

Webb apps, which go hand in hand with cloud-based services platforms, can be integrated with cloud-based services like Open Replay to add a fully functional session-replay feature [13]. This structure provides the resources for people to know their interactions; this system also allows the user experience to be managed, and cloud platforms allow data to be scaled up and saved.

There is no limitation either in terms of cloud platforms or feature flagging; through A/B testing and feature flagging, companies can roll out new functions in an advanced system [14]. The cloud would provide integration platforms around the data and decision-making; users can access any real-time information they wish to consider during the decision-making processes.

Additionally, Cloud platforms, Redash, Apache Superset, and Countly, can be utilized to build out comprehensive visualization and analysis solutions [12]. Therefore, these coordinations have remained extant, serving as the guarantors for the reliability and optimization of cloud resources in executing data processing and analytics of optimal performance.

This service can be an integral component of a cloud-based solution that provides for data processing and storage of web traffic-intensive and massive datasets.

Such functionality may be combined with different cloud analytics tools, like AWS, GCP, or Azure, and it can be done by investigating the scalability, performance, and security questions related to the cloud [13]. Firstly, the tool should be moved to a cloud platform, then connected to the data sources, and finally, we should adjust the analytical instruments settings to fit the organization's evolution [15]. The need for Big Data analytics is that preprocessing will lead to the betterment of architecture planning for the organizations, and then the big data problems will be addressed.

3.4. Implementation Framework

Implementing analytics in a cloud-based framework begins with acquiring and preprocessing the relevant datasets integrated into public health decision-making. This is the stage at which the researchers collect broad-based public health data through the initial assessment, and thus, data protection is regulated. After obtaining the data, the procedure follows a rigorous regime of cleaning and preprocessing, including identifying anomalies and missing values [17]. Some techniques for preparing datasets for further analyses are normalization, standardization, and transformation.

Model development and training follow data preprocessing. Here, the method for analysis and algorithms to be used are decided according to the set objectives and the characteristics of the datasets [3]. Sample data is divided into training, validation, and test sets for building and internal validation of analytical models. The models developed on the training set are optimized for parameter tuning to increase performance [17]. This is followed by validation, whereby the trained model is subjected to validation with the validation dataset to estimate its accuracy, precision, recall, and other performance measures.

The framework proceeds now to model evaluation and validation. In this stage, the framework will evaluate the performance of the models over the testing dataset by validating their predicted outcomes with an actual ground truth value. Additionally, cross-validation techniques and sensitivity analyses continue to build on the robustness and generalizability of the models [15]. The basic level of validation from any of the models includes expert review and domain-specific validation to ensure that the models represent the actual public health phenomena with reasonable accuracy.

4. Results and Discussion

4.1. Meaningful Insights

Through our breakthrough in the data halted from the CDC and data world, we can have a clear understanding of the following information [17]. The data here features some of the major trends and findings in leading causes of death among United States citizens from 1-year-olds to 44 with a time range of 1981 to 2021.

Uncertainty, injury has emerged as the fatal reason, without exceptions, whose upward trend has kept on booming to the zenith now [18]. Hence, if something like this is implemented (probably in replacement of day-to-day establishments), there needs to be easy and convenient access to safety precautions and safety measures.

On the other hand, both large-scale cancer and heart disease, which were always at work to bring the deaths of people, remained the top causes of death during that period [18]. We may also work on identifying an early-detecting diagnostic tool as well as a powerful treatment method.

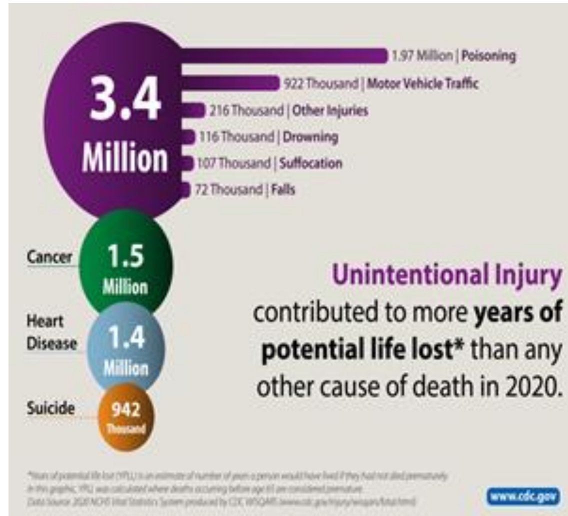

Fig. 1 Causes of death [5]

By the mid-80s in the USA, HIV belonged to the top ten leading infectious diseases with a similar mirror to the impact the epidemic had had at that time [20]. By the time the number of deaths associated with HIV not yet accounting for a mere 0.05% of the WHO report of 1999, the nature of the slogan "taking the menace head-on" took inspiration from the ever use of the semantic formula of the "double-edged sword" and the undying will of the people to overcome the prevention-treatment challenges associated to the HIV.

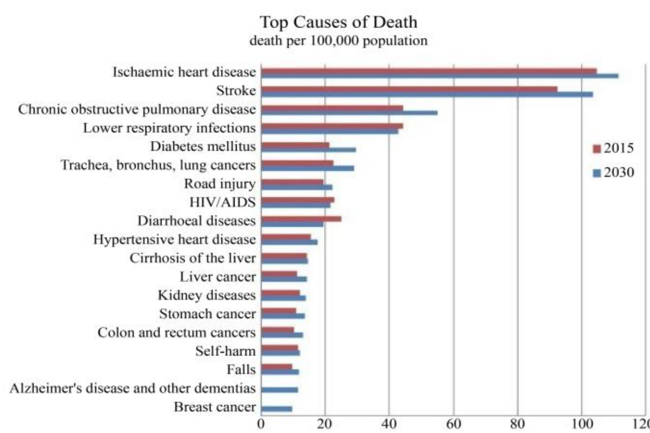

Fig. 2 Causes of death[12]

The murder and suicide rates are always changing. However, these two issues will penetrate society at this large scale, and only social intervention and balance of mental health, as well as violence prevention programs, will work.

Chronic non-communicable diseases are associated with infectious diseases such as cancer, diabetes, and cardiovascular diseases that kill over 50 million around the world. It is through this that the full spectrum of deadliness can be appreciated [19]. This host change happened over time, and it involved Influenza and pneumonia, health care provision, immunization, and public health interventions.

The concern, in like manner, of the infant mortality rate is related to congenital disabilities, too, and they demonstrate a breakaway fall, which can be majorly assigned to prenatal care improvement and the growth of medicine.

Moreover, the randomness of rates of murder and suicides emphasizes the soul-searching social problems that cause mayhem and mentally unstable behavior [20]. Taking into consideration social support programs and mental health initiatives are fundamental measures that promote tackling these causes and influence reduced violence and self-harm rates.

The spread of chronic non-communicable diseases worldwide demonstrates the high incidence of these diseases, such as cancer, diabetes, and cardiovascular ones, which cause worldwide health problems. To tackle these diseases, measures must be multi-dimensional and involve providing healthcare services, immunization, public health, or even individual approaches to prevention or management [17].

The infant mortality rate is lowered mainly because of pregnancy awareness and the medical transport evolution. This shows how medical interventions are the key to keeping the mortality rate low among the vulnerable, underlining the priority of maternal and child health programs.

5. Conclusion

Based on the huge capability of cloud computing, we have made analyses going to our owners and formed a humanoid "three decades" within minutes. This capability suggests that all healthcare institutions are to be well equipped with this technology and that this is not just meant for healthcare delivery but also public health surveillance.

Applying data analysis in cloud computing will significantly impact public health decision-making. Implementing the latest big data and analytic techniques will bring many advantages for stakeholders, such as identifying the tendencies in the health sector, defining risk factors, and using the most suitable strategies for improving the health status of the people.

However, data privacy concerns and technical integration issues, which are pitfalls, should be tackled first to avoid compromising the benefits of cloud-based analytics in public health. Collaboration among various actors, including health providers, policymakers, technology experts, and others, is needed to ensure that the potential of cloud computing in future public health agendas is effectively utilized.

The future must be explored with more research and money on cloud analytics platforms and technologies that can accurately support decision-making at all levels, including the global view. By embracing the innovative approach and using the power of data, public health organizations can make the reforms that will bring new life into healthcare systems and improve health for all of us.

References

[1] Odunayo Josephine Akindote et al., “Comparative Review of Big Data Analytics and GIS in Healthcare Decision-Making,” *World Journal of Advanced Research and Reviews*, vol. 20, no. 3, pp. 1293-1302, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Anna Petrovskis et al., “The DASH Model: Data for Addressing Social Determinants of Health in Local Health Departments,” *Nursing Inquiry*, vol. 30, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Joseph Bamidele Awotunde, Roseline Oluwaseun Ogundokun, and Sanjay Misra, “Cloud and IoMT-Based Big Data Analytics System during the COVID-19 Pandemic,” *Efficient Data Handling for Massive Internet of Medical Things: Healthcare Data Analytics*, pp. 181-201, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Alexy Gene Castillo, Sherwin M. Telan, and Thelma Palaoag, “Cloud-Based Data Mining Framework: A Model to Improve Maternal Healthcare,” *Proceedings of the 2nd International Conference on Cryptography, Security and Privacy*, Guiyang China, pp. 21-28, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nicholas Mirin et al., “Data Science in Public Health: Building Following Generation Capacity,” *Arxiv*, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Louis Ehwerhemuepha et al., “HealtheDataLab is a Cloud Computing Solution for Data Science and Advanced Analytics in Healthcare with Applications to Predict Multi-Center Pediatric Readmissions,” *BMC Medical Informatics and Decision Making*, vol. 20, pp. 1-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mengchun Gong et al., “Cloud-Based System for Effective Surveillance and Control of COVID-19: Useful Experiences from Hubei, China,” *Journal of Medical Internet Research*, vol. 22, no. 4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Anish Jindal et al., “Providing Healthcare-as-a-Service Using Fuzzy Rule Based Big Data Analytics in Cloud Computing,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1605-1618, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] K. Kavitha, and C. Kumuthini, *Cloud-Based Data Analytics for Healthcare 5.0*, Pioneering Smart Healthcare 5.0 with IoT, Federated Learning, and Cloud Security, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fatima Khalique, Shoab A. Khan, and Irum Nosheen, "A Framework for Public Health Monitoring, Analytics, and Research," *IEEE Access*, vol. 7, pp. 101309-101326, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Fernando López-Martínez et al., "A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management," *Algorithms*, vol. 13, no. 4, pp. 1-19, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Laura Mählmann et al., "Big Data for Public Health Policy-Making: Policy Empowerment," *Public Health Genomics*, vol. 20, no. 6, pp. 312-320, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] Patel Keyur, "Lawful and Righteous Considerations for the Use of Artificial Intelligence in Public Health," *International Journal of Computer Trends and Technology*, vol. 72, no. 1, pp. 48-52, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Abraham George et al., "Use of System Dynamics Modelling for Evidence-Based Decision Making in Public Health Practice," *Systems*, vol. 11, no. 5, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ali H. Mhmood, and Murshid Reza, "Urban Health Informatics through Cloud-Based Data Integration," *Journal of Big-Data Analytics and Cloud Computing*, vol. 7, no. 4, pp. 1-17, 2022. [Google Scholar] [Publisher Link]

[16] Shah Jahan Miah, *A Demand-Driven Cloud-Based Business Intelligence for Healthcare Decision-Making*, Health Care Delivery and Clinical Science: Concepts, Methodologies, Tools, and Applications, pp. 964-979, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[17] Innocent Mumba Mwila, "*Geospatial, Cloud and Web-Based Model for Evidence-Based Decision-Making in Tuberculosis Prevention*," Theses and Dissertations, 2020. [Google Scholar] [Publisher Link]

[18] Seungeun Park et al., "Impact of Data Visualization on Decision-Making and its Implications for Public Health Practice: A Systematic Literature Review," *Informatics for Health & Social Care*, vol. 47, pp. 175-193, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Lisa E.M. Hopcroft et al., "Data-Driven Identification of Potentially Successful Intervention Implementations: A Proof of Concept Using Five Years of Opioid Prescribing Data from Over 7000 Practices in England," *medRxiv*, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Sumanth Tatineni, "Cloud-Based Data Analytics for Smart Cities: Enhancing Urban Infrastructure and Services," *International Research Journal of Modernization in Engineering, Technology, and Science*, vol. 5, no. 11, pp. 904-912, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Shuiting Zhang et al., "Anesthesia Decision Analysis Using a Cloud-Based Big Data Platform," *European Journal of Medical Research*, vol. 29, no. 1, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

Appendix

1. Data Sources

This section identifies the various data sources utilized in this study for analyzing public health data:

World Health Organization (WHO): Provides global data on various public health issues, serving as a comprehensive source for analyzing health trends and patterns.

Centers for Disease Control and Prevention (CDC): Offers data on injury, crime, and vaccination statistics in the United States, enabling detailed analysis of health issues at the national level.

Electronic Health Records: Data from hospitals and healthcare facilities, providing insights into patient demographics, medical conditions, and treatment outcomes.

National Notifiable Diseases Surveillance System (NNDSS): Provides medical data on notifiable diseases, aiding in the monitoring and surveillance of infectious diseases.

Data World and Data.gov: Platforms offering a wide range of datasets related to public health, environmental factors, and demographic information, facilitating comprehensive analysis and research.

2. Analysis Tools

This section outlines specific analysis tools that can be integrated with cloud platforms to enhance data processing and analysis in public health.

PostHog: Captures and analyzes user interaction on websites and applications, facilitating scalability and performance improvements on cloud platforms.

Matomo: Provides insights into user behavior and website performance, enhancing data analysis capabilities on cloud services like Cloud and AWS.

Metabase: Enables easy querying and visualization of data on cloud platforms like AWS, GCP, and Azure, streamlining data analysis processes.

Grafana: Monitors and visualizes cloud infrastructure metrics and application data, offering convenient dashboard formats for data analysis on AWS, GCP, or Azure.

Open Replay: Integrates with web apps to add a fully functional session-replay feature, allowing for better user experience management on cloud platforms.

Feature Flagging Tools: Facilitate A/B testing and feature flagging, enabling organizations to roll out new functions and make data-driven decisions on cloud platforms.