Aggregating Covid-19 Data on Vaccination and Mortality Progression for Dashboard Visualization: A Recap

Festus O. Oliha[1, *] and Kelly E. Omobude[2]
[1] Department of Computer Science, University of Benin, Nigeria
[2]School of Built Environment, Engineering & Computing, Leeds Beckett University, UK
[*]Corresponding author: oliha_festus@uniben.edu

Abstract

The visualization of accurate coronavirus disease – covid-19 data and information is vital in a public-related health predicament since the general public's attempt to make individual meaning of its vaccination and mortality trends may lead to a widespread misrepresentation and misinterpretation that could affect the goal of information dissemination and collective actions. The present study frowned towards the variations in covid-19 information across reporting dashboards, and in turn, utilized the efficiency of machine learning techniques in the appropriation of multiple datasets for representing covid-19 information via a dashboard. Using the CRISP-DM methodology, data aggregation and visualization techniques were employed to promote interactive representation of covid-19 pandemic progression across several countries for the general public, government agencies, and health/research organizations who inevitably and dynamically pursue the representation of unified and accurate covid-19 information. The results, using the UK as a case study, visualized the aggregated vaccination and mortality progression rate with the prevalence of accurate covid-19 infected cases and mortality relative to vaccinated persons on a global scale. The study advises from the results and findings that countries with a high incidence rate of covid-19 cases should be more vaccine-oriented with collective actions to reduce the number of deaths associated with the covid-19 plague.

Keywords: covid-19, covid-19 dashboard, mortality trends, dashboard visualization, covid-19 vaccination, machine learning.

1. Introduction

In the veins of a public health crisis, the World Health Organization (WHO), Kaggle, and other several organizations are online platforms aiding the visualization of coronavirus - covid-19 data to corroborate medical reports on the pandemic disease since its inception in December 2019. Accurate data is essential to understand the associated trends in organizational progression via visual dashboards or descriptive analyses while gaining in-depth insight for several purposes that drive decision-making (Mitra *et al.*, 2021). The covid-19 contagion had a distressing effect and has persisted as a universal concern to human well-being and race (Ebubeogu *et al.*, 2022), and the number of illnesses and death have fluctuated unpredictably for almost three years now.

The epidemic trend of the covid-19 pandemic has imposed an unexpected direction toward global prevention of the deadly disease. Monitoring and reporting the epidemic pattern are crucial steps toward establishing preventive measures. Abdelsamad and Karrar (2021) recently, signified that dashboards have become essential tools for representing and visualizing the pandemic progression patterns by providing information in figures regarding countries with associated deaths, vaccinated, hospitalized, etc., cases. In view of the above, dashboards have become powerful in representing covid-19 data for health policymakers with reports of over half a billion (0.5 billion) cases and about 7,000,000 fatalities have been confirmed as of the first quarter of 2022 (Johns Hopkins Coronavirus Resource Center, 2022).

On the other hand, aggregating multiple datasets for understanding the current tendencies and predicting future implications is motivating the present study to avoid a rising trend in casualties from covid-19 or its misrepresentation of vaccination information. On this note, Iqbal *et al* (2022) expressed that "Machine learning models have been found in studies to be particularly effective in predicting more accurate information about the covid-19 pandemic, allowing medical experts to make informed judgments". Engledowl and Karrar, (2021), argued that "the widespread illiteracy led to misinterpretation and misrepresentation in individuals' attempts to interact with data visualization to facilitate decision-making, following the covid-19 pandemic outbreak" – this is a concern that outlines comprehensiveness on covid-19 data represented by various platforms and varying datasets. Thus, the presentation of accurate covid-19 information is sought after by the general public with a dashboard visualizing its progression and trends to guide medical experts in decision-making. However, the problem of inaccurate covid-19 information is consequent on the source of data but the representation in several platforms by health organizations may lead to (if not aggregated) misrepresentation or misinterpretation (panic) of its impact or tendencies for the general public.

The present study is a recap seeking to marry machine learning approaches with data visualization techniques to aggregated datasets with descriptive analysis on real-time covid-19 pandemic progression, particularly on associated deaths with covid-19 vaccinations. Further objectives are to conduct exploratory data analysis to understand the covid-19 pandemic population data; employ a machine learning-based algorithm using Python resources to aggregate and appropriate the covid-19 datasets for dashboard representation, and finally, visualize the aggregated trends and patterns for a selected country. The study hereafter introduces the covid-19 disease with information on its inception, counter-measures, and studies towards prevention and solutions. Section two conceptualizes the theoretical approach and methodology while the third section analyzes the dashboard's visualizations and results. Section four concludes the study with recommendations for further studies.

1.1 Background information on related concepts and studies

Covid-19 information has been represented using several platforms and dashboards for the public to gain insights or understanding of trends and patterns of the illness. However, there was also misleading information and misinterpretation by individual attempts to understand covid-19 information visualized in some reporting platforms (Engledowl and Weiland, 2021). The need for easy comprehension of covid-19 information across dashboards has been emphasized and many research efforts have focused on representing accurate covid-19 information to aid decision and prediction of its trends and progression (Alimohamadi *et al.*, 2020; Jebril *et al.*, 2020; Tarkar, 2020; Asselau *et al.,* 2021; Lamb, 2021; Storlie *et* *al.,* 2021). The goal in most studies has been associated with measures to reduce the epidemic and generate a curable vaccine for the disease. In this vein, WHO, Pfizer, Johnson & Johnson, etc., and the entire healthcare systems have been resilient in the efficacy of covid-19 vaccines and their vaccinations.

Studies on covid-19 pandemic progression have concentrated on predictive machine learning models to identify covid-19 cases with high mortality risk and equip healthcare experts in understanding the epidemic and possible collective actions toward affected patients. Nevertheless, the predictions were not visually available on open or public reporting platforms and dashboards. On this note, Azzam *et al.* (2013) affirmed that cultured research was inspired by the display of relevant information on dashboards aided via visualization methods and thus, aggregating covid-19 data across the globe would aid the culture of epidemiological investigation of the virus and its trends, presentation, analysis of research findings, and methodological reviews to reduce the universal health concerns to human race (Mitra *et al.*, 2021). Relatively, preliminary studies were explored to understand data visualization and covid-19 trends.

In Combra (2020), the technical use of visualization was emphasized to be of immense assistance to the public at large in terms of disseminating information on covid-19 epidemic trends to health professionals, policymakers, and care experts. The intent of Combra (2020), was corroborated in (Biswas *et al*., 2020) adding that, the benefits of data visualization apply to the current health crisis caused by the covid-19 pandemic. The concept of visualization was further explored in Teh et al., (2021), where a multivariate visualization was employed on covid-19 data to identify notable facets in terms of the incidence of the pandemic spreads and progression. It was concluded in their study that visualization remained a dashboard-learning tool about the pandemic spread of covid-19 data across the globe.

In Bowe *et a*l. (2020), machine learning was emphasized as a crucial technique for effectively visualizing covid-19 data towards the representation and interpretation of more feasible and accurate covid-19 information and thereby, dashboarding with ease in the context of understanding a wider array of audiences globally. On the note of representing accurate covid-19 information, there has been evidence of data misrepresentation and interpretations in visualization towards decision-making as a result of poor data quality and widespread illiteracy (Engledowl and Weiland, 2021). Trying to curb panic that may arise from misrepresentation of covid-19 progression, accurate geographical covid-19 pandemic data in terms of its incidences on affected cases, rates of vaccination tendencies, and mortality indices can be visualized with interactive dashboards as maintained in Mitra *et al*. (2021). However, misrepresentation of covid-19 information in terms of geographical data may be consequent on the sources of the data – large multiple datasets and these require a learning model to normalize such datasets for accurate representation purposes.

Machine learning application to covid-19 pandemic research was reported by Kushwaha *et al.* (2020) to be significant for its widespread analytics, determining the relationship between its data and the pandemic mortality rates. In a study where machine learning model was applied in visualizing covid-19 pandemic progression, Mathur *et al.* (2020) explored the United States of America as a sample study to understand disease determinants for covid-19 mortality rates. Machine learning models are of great advantages in creating decision support systems that can assist, given a large dataset, access these data and learn from them based on supervised or unsupervised learning techniques with a range of algorithms to project the representation of accurate covid-19 information and decision making (Kushwaha *et al*., 2020; Singh *et al*., 2016; Natarajan *et al*., 2017).

Consequently, a machine learning-based dashboard that visualizes an aggregated covid-19 dataset in an interactive, simple-to-understand manner has been highlighted as a paramount objective in the present study to curb associated misrepresentation of covid-19 information from multiple data sources.

2. Materials and methods

The study methodically utilized various research techniques and approaches conceptualizing the processes in the field of study. The data aggregation approach was centered on the CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology as corroborated in prior research (Chapman *et al*., 2000; Niaksu, 2015; Schroera *et al*., 2021). The CRISP-DM is phased with different processes aggregating data science and machine learning-related studies in terms of data understanding, preparation, modeling, evaluation, deployment, and visualization as depicted in Figure 1. With relevance to the present study, the CRISP-DM was only adapted to understand and aggregate data for visualization purposes.

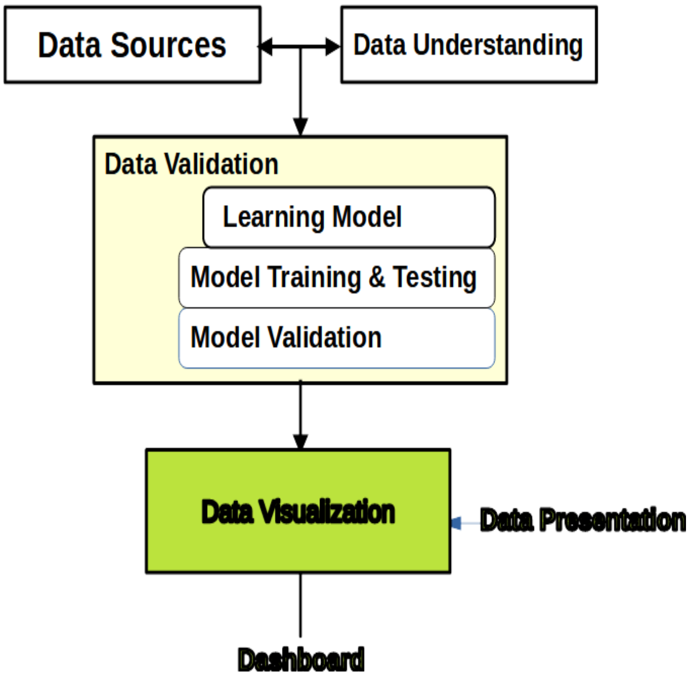

Figure 1 Conceptual framework.

2.1 Data understanding

2.1.1 Data sources

Datasets were mined from different online data sources and repositories for covid-19 vaccinations and death reports. The data sources included the WHO (2022) and Kaggle concerning covid-19 country-wise cases of infection, population, vaccines (authorized), vaccination, etc., and mortality data (https://covid19.who.int/info/; https://www.kaggle.com/gpreda/covid-world-vaccination-progress;

https://www.kaggle.com/rsrishav/world-population).

2.1.2 Data appropriation - preprocessing

The preprocessing of data was paramount to understanding the mined datasets. The preprocessing was explored using Python resources and libraries for selection, preparation, cleaning, and merging. The aggregated data are needed to visualize informed covid-19 data via the proposed dashboard. The dataset contained multiple data on covid-19 cases, population, country, covid-19 vaccination, and WHO vaccines. The properties of each dataset dictionary were described via Python data frames with elements as depicted in Figure 2.

```
covid_df = covid_data[["date", "country","New_deaths"]]
population_df = population[["iso_code","country", "Populatio
vaccine_df = vaccine[["country","iso_code","date","total_vac
                       "daily_vaccinations","vaccines",'total
                        'people_fully_vaccinated_per_hundred'
vaccine_who_df = vaccine_who[['country','number_of_vaccine']
```

Figure 2 Handling multiple datasets.

2.1.3 Data cleaning and merging

As a result of multiple datasets, the Pandas merge functionality was explored via the Python library resources to manipulate and aggregate datasets for further exploration or analysis. Figure 3 is a snapshot of the resultant merged and cleaned dataset containing about 23,195 rows and 14 columns. A series of Python instructions with imported pandas' merge function was employed to finalize the aggregation of the rows and columns from the multiple datasets. The "shape" Python command enables the view of the final merged dataset for analysis to verify a coherent dataset free from duplication and missing values.

```
covid_full.shape
(23195, 14)

covid_full.isna().sum()
country                                0
iso_code                               0
date                                   0
total_vaccinations                     0
people_vaccinated                      0
people_fully_vaccinated                0
daily_vaccinations                     0
vaccines                               0
total_vaccinations_per_hundred         0
people_vaccinated_per_hundred          0
people_fully_vaccinated_per_hundred    0
New_deaths                             0
Population                             0
number_of_vaccine                      0
dtype: int64
```

Figure 3 Cleaned and normalized dataset.

2.2 Data validation

A normalized dataset requires a model to validate its fitness for further analysis. In Figure 3, it was apparent that there was no missing values nor duplicated row or column. Appropriating this, machine learning models were trained to check if the merged dataset is fit for interactive visualization as depicted in Figure 4. Amongst the trained and tested models, the Catboost model was better in training and testing and hence, selected for appropriating the final dataset. Figure 4 depicted the Catboost regression model associating Python sklearn libraries and metrics (test, split, train, R-squared score) for fitness and accuracy in representing the datasets with frames accommodative of x and y respectively. X represented the selected features while Y determined the variable incident on a random split of 80% - 20% train_test to validate the test data as shown in Figure 4.

```
In [ ]: X = covid_full[['total_vaccinations', 'people_vaccinated', 'people_fully_vaccinated', 'daily_vaccinations',
                        'Population','number_of_vaccine','number_of_booster']]

        y = covid_full['New_deaths']

        from sklearn.model_selection import train_test_split

        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=45)
```

Figure 4 Machine learning modeling.

The machine learning model – Catboost was efficient in determining the validity of the test data adjudged with a mean squared error of **97.91,** root mean squared error of **0.99,** mean absolute error of **0.27,** and R squared value of **0.907.** This indicates that the model learns and performs with an accuracy of 91% in validating the merged dataset with appropriateness for both descriptive visualizations and predictive analysis. Thus, the final dataset was deposited in the GitHub repository for further representation and analysis.

2.3 Data visualization - dashboard

With the validation of the dataset via the machine learning model, the exportation of the final merged dataset was paramount for deployment, presentation, and analysis. However, to infer meaning and make an informed decision, a dashboard was developed to visualize the aggregated dataset for ad-hoc analysis, reports, and ongoing decision support.

2.3.1 The proposed visualization dashboard design

Dashboards are a dynamic display of information to support high-quality decision-making. Figure 5 depicts the core steps in designing and developing a dashboard.

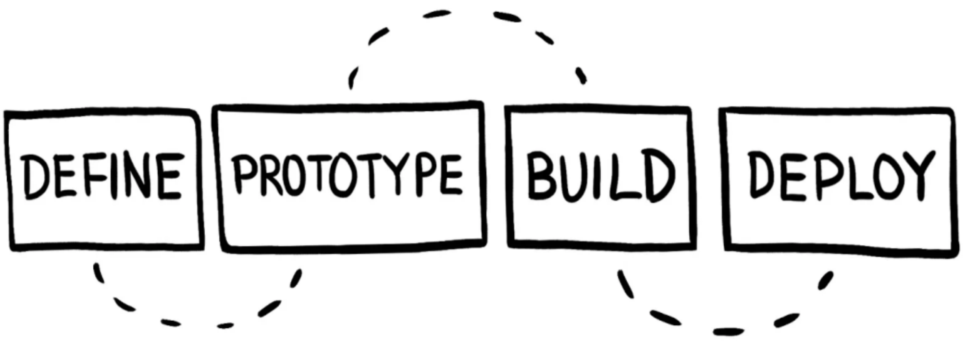

Figure 5 Design framework (David, 2018).

Figure 5 portrays the vital steps in proposing the dashboard with insights on the accurate aggregated covid-19 vaccination data regarding comprehensive visualization metrics, quality, and chart designs. These attributes were all considered to prevent biased interpretation in addition to the application of Python resources for data normalization and preparation. This conception ensured that visualization designs were accurate, clear, empowering, and succinct in terms of fonts, colors, context, views, and layouts with descriptive titles, value, and categorical labels that will encourage and empower viewers and also, factor into the decision-making process for targeted health practitioners and organizations.

The development of the proposed dashboard followed the steps in Figure 5, in which the stakeholders were defined and their responsibilities with relevant metrics to mission-critical data for decision making. The design was based on the defined query-able metrics with sketches, visualization patterns, and iterations composed on a prototype dashboard. The visualization selection was guided by some of the components in the decision tree in Figure 6.

The dashboard was proposed with prototype chart decisions that supported the display of multiple values showing relationships and comparison. Various software components were also employed in the engineering of the dashboard solution to aid synchronization and automatic loading of the aggregated dataset into a repository with an easy find-fetch capability on the dashboard. Thus, a build of the actual dashboard was created using queries to transfer data from the repository into the dashboard charts with defined metrics.

After several iterations with the agile software approach, the final functional dashboard was deployed to share, scale, and maintaining decision support within the designated fields of healthcare.

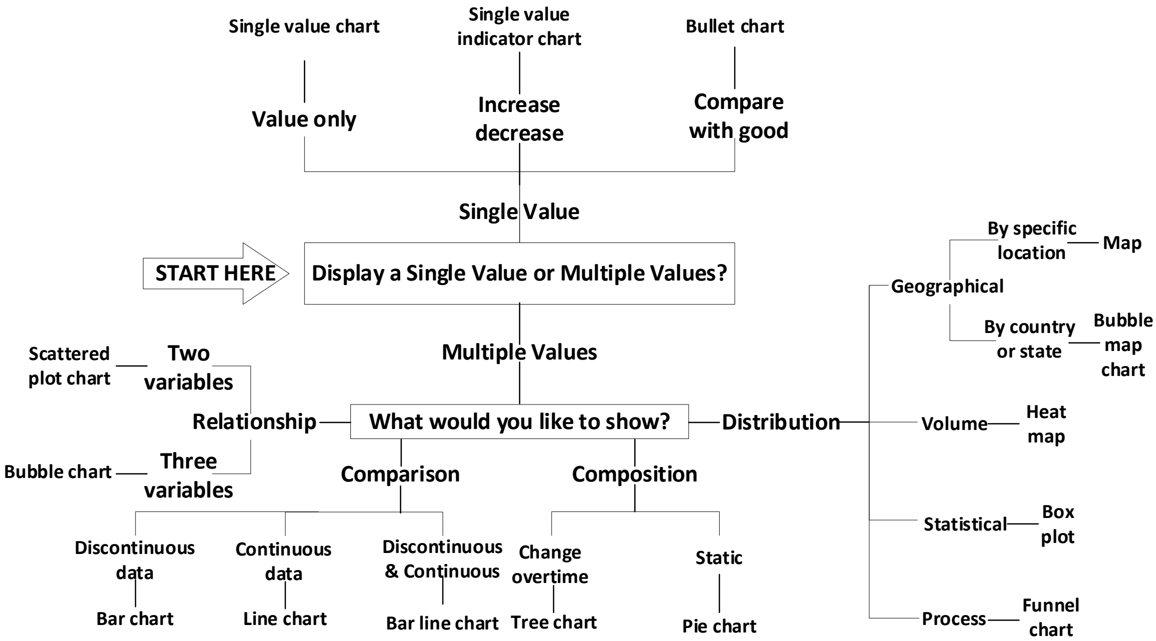

Figure 6 Visualization decision tree.

3. Results

The developed visualization dashboard is capable of representing and displaying covid-19 information with a breakdown for all countries; however, the United Kingdom (UK) covid-19 information was a case study for data representation and analysis for the study. The UK was selected as a result of more readily available and complete covid-19 data within the confinement of the evaluation and visualization metrics. Moreso, the aggregated covid-19 dataset was exported via web API for further representation and visualization with screenshots of valid covid-19 information up to April 2022, and thus, the pandemic progression afterward may vary in trending patterns.

3.1 Dashboard results and visualization analysis

Visualizing the UK covid-19 information, Figure 7 shows the pandemic spread in terms of the population and related covid-19 incidence representing figures and values for each category of covid-19 information. As shown in Figure 7, the UK population totaled 68.3 million with about 7.34 million hospitalized and 689,791 patients reported at intensive care unit. The total number of covid-19 cases amounted to 20.8 million with about 165,848 deaths within the United Kingdom.

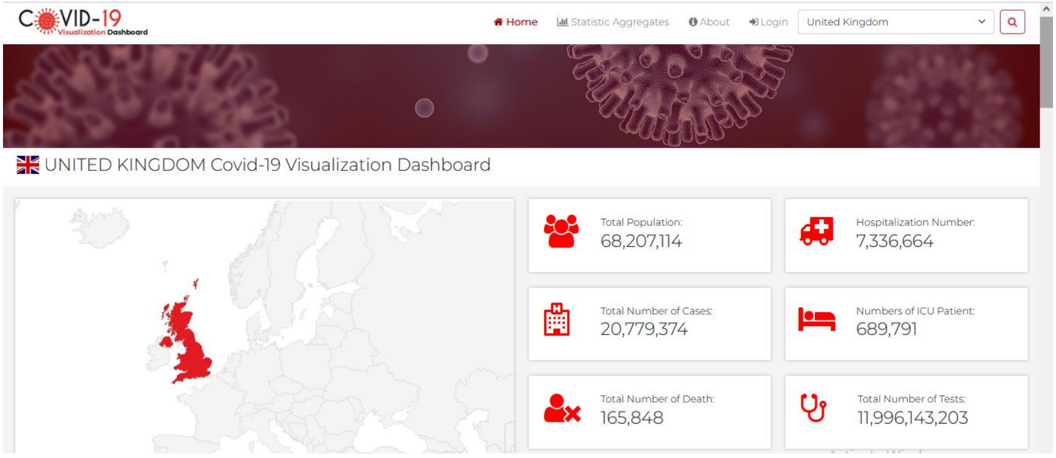

Figure 7 Dashboard visualization of related covid-19 information in the UK.

In Figure 8, the UK pandemic progression was statistically represented in terms of vaccination histories. It also depicted the pattern progression chart for UK covid-19 new cases and deaths.

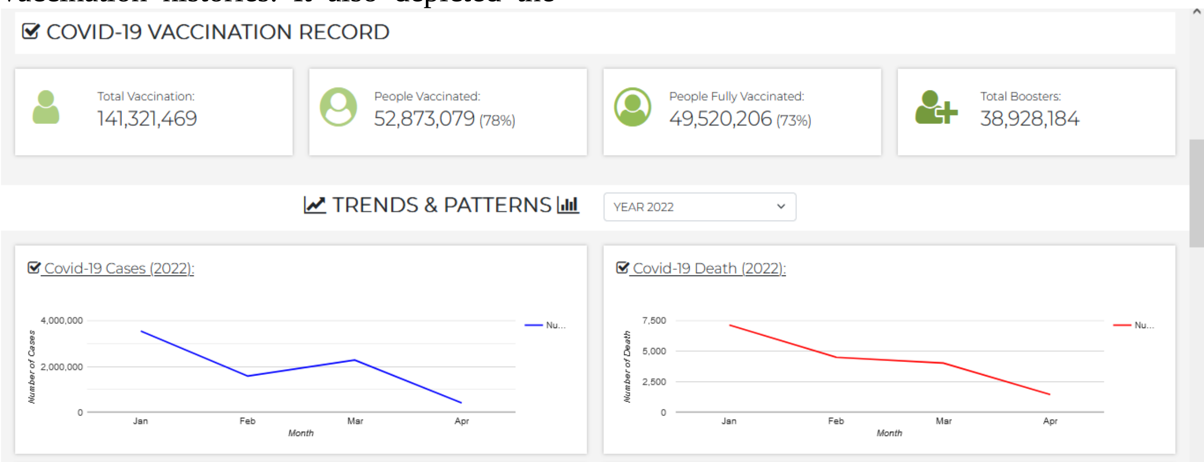

Figure 8 Visualizing covid-19 statistical progression – UK dashboard.

The total number of vaccinations was recorded in terms of people vaccinated and fully vaccinated with the booster doses. From Figure 8, there was an indication of a significant increment in infected cases between February and March for the year 2022. Also noted, was the decrease in mortality rate which was marginally modest within the same months but, higher in the prior month – January.

The covid-19 predictors highlighted in this study were the number of confirmed cases, deaths, hospitalizations, and vaccinations depicted in Figure 8 as variables for line graph comparison of monthly covid-19 information for the UK. The dashboard displays additional information for each predictor on a mouse hovering across the charts. Breaking down the predictors, the chart in Figure 9 represented the associated confirmed new covid cases with vaccination tendencies highlighted with monthly statistics in the UK from January to April for the year 2022. It was evident that January recorded the highest incidence for both predictors with almost 3.6 million cases and 5 million persons vaccinated. This indicates that more persons were vaccinated than infected in January and February except a decrease was reported in vaccination tendencies for March and April of 2022 in the UK as the number of new cases slowly decreased also.

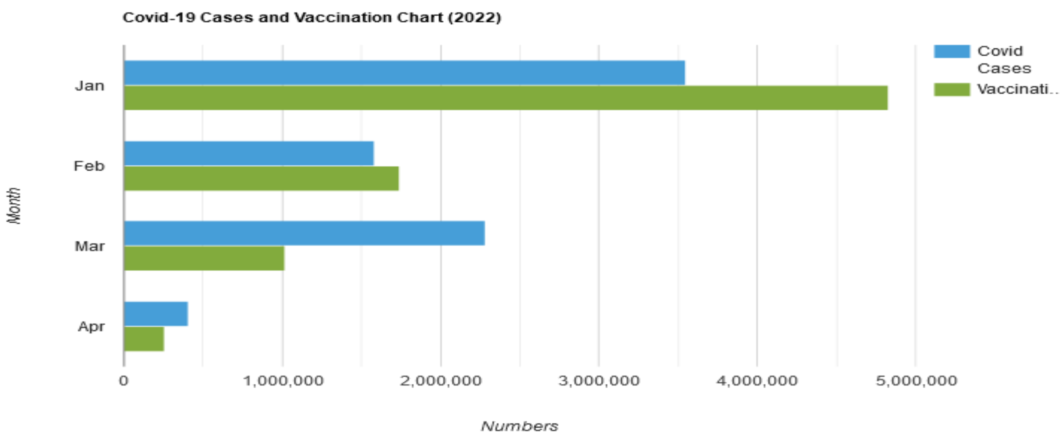

Figure 9 Covid-19 cases and vaccination chart in the UK, 2022.

The predictor of monthly deaths in the UK was represented in Figure 10 with covid-19 progression visualized from January through to April of 2022. Clearly, the month of April was an indicator for the lowest number of deaths with January recording the highest.

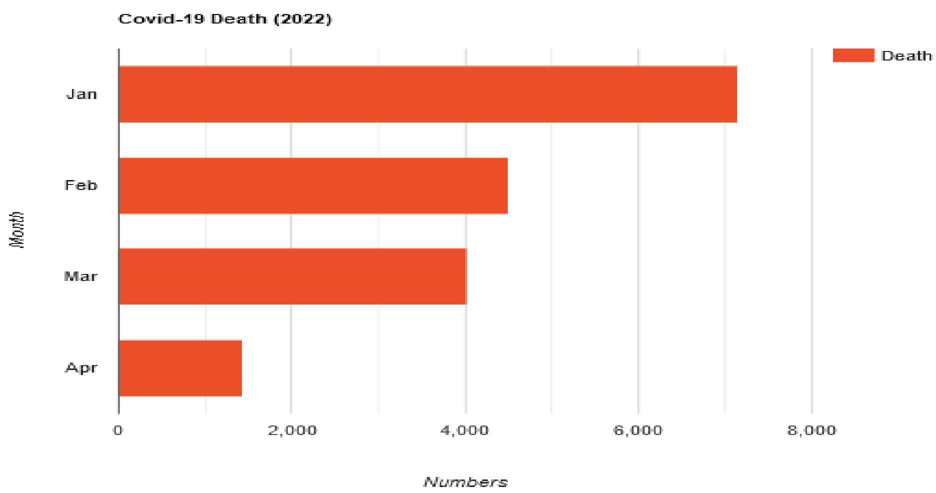

Figure 10 Covid-19 death chart in the UK: January – April 2022.

The death rate from January observed a depleting number through to April in the UK as of 2022. It is safe to visualize this decrease from the chart in Figure 9 where the vaccination rate was at its peak in January causing a sweeping drop or downward covid-19 mortality progression from January through to April in the UK.

Furthermore, the predictors of covid-19 charts for confirmed cases, deaths, and vaccination were visualized for the previous years via the dashboard ease-of-use features. Figure 11 visualized information for the covid-19 predictors: confirmed cases, deaths, vaccinations, hospitalizations, etc., with a side-by-side graphical comparison for each month in the year 2021.

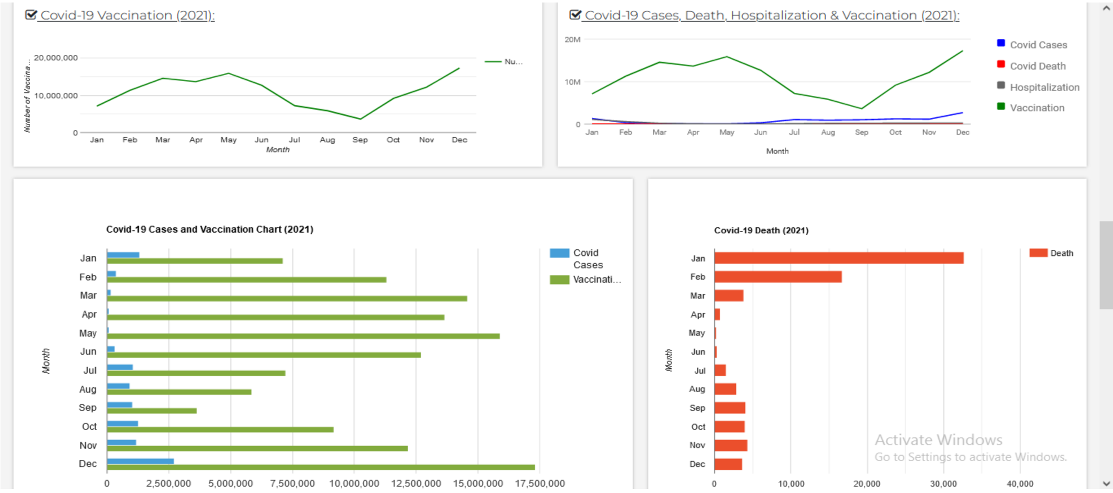

Figure 11 Visualization of UK covid-19 pandemic progression for the year 2021.

The UK covid-19 pandemic trends for the year 2021 reported its highest prevalence in December for the predictors of confirmed cases and vaccination while January highlighted deaths in the highest with May and June recording the lowest. From the graphical information visualized, it is fair to admit that the vaccination tendencies have reduced the impact of the pandemic on the mortality rate for the year 2021 in the UK. However, marginal rise and drop were noted in mortality rate between the same year as observed in Figure 11.

The predictor of vaccination tendencies recorded no covid-19 information in the UK for the year 2020 according to Figure 12. The predictors of infected (confirmed) cases and deaths were represented by the number of infected persons highly recorded in December and April 2020 in the UK. The absence of vaccination records accounted for the increased tendencies in the predictors of confirmed cases and the number of deaths in the UK in 2020.

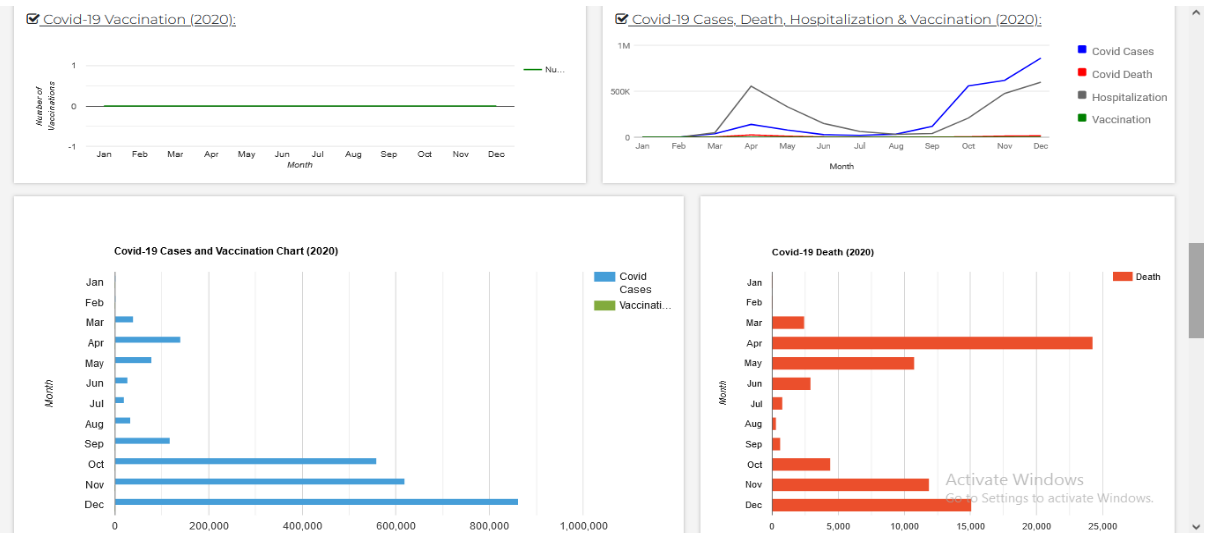

Figure 12 Visualization of UK covid-19 pandemic progression for the year 2020.

Consequently, the progression expedited the research for covid-19 vaccines – that is, the progression from October 2020 for the predictors of infected and death cases was a justifiable cause for the hastiness in the development of the covid-19 vaccines.

3.2 Discussions

In a similar study, (Mathur *et al*., 2020) validated machine learning approaches to explore the pandemic progression using the USA mortality as a case study. Thus, the present study adopted a machine model-based methodology for the appropriation and representation of multiple covid-19 datasets with visualizations of its pandemic progression in consonance with Mitra *et al.* (2021) by displaying a range of charts, line graphs, and bar graphs using data visualization techniques; colors, data patterns, and trends that are easily discernible via a dashboard to evade misrepresentation and misrepresentation that was argued in (Engledowl and Weiland, 2021). It aggregated datasets from covid-19 repositories: world health organization, Kaggle, and Our World in Data using Python resources.

The adoption of the UK as a case study was to evaluate the pattern progression of the covid-19 infected cases, deaths, and vaccination information over three consecutive years using selected predictors on the aggregated dataset with relevance to the country's population. On the pattern progression results for the UK, the analyses of three years were presented in Figures 8 through to Figure 12. Figure 9 and Figure 10 depicted the monthly comparative trending patterns in covid-19 cases, vaccination records, and deaths in the UK, as of April 10, 2022. The month of April was highlighted as a benchmark to comprehend the inclinations of vaccination associated with covid-19 deaths.

A chart of three years (2022, 2021, and 2020) of covid-19 pandemic progression was visualized and the results showed that no vaccinated person was recorded in the trend of 2020 due to the unavailability of covid-19 vaccines but, December 2020 marked the peak month of covid-19 confirmed cases with over 800,000 persons infected and about 25,000 deaths recorded in April 2020 in the UK. Subsequently, it was revealed that the UK had recorded the highest prevalence of covid-19 cases of over 2.5 million in December 2021 while 30,000 or more deaths were documented in January 2021 with about 17.5 million persons vaccinated as of December 2021. More so, the highest incidence recorded from the aggregated dataset for confirmed cases (3.5 million), deaths (slightly above 7,000), and vaccination (about 5 million) were noted in January 2022 as visualized in Figure 9 and Figure 10. Alternatively, the lowest incidence was observed in April 2022 with a modest drop in covid-19 cases below 500,000 and deaths below 2,000. These indications validate that the rate of vaccination in January was relative to the drop in the number of infected cases and deaths, despite the slight rise in infected cases from February to March 2022 as a result of low vaccination tendencies by the UK inhabitants.

Relatively noted in other studies (Biswas *et al*., 2020; Teh *et al*., 2021), dashboard visualization in the present study was deployed with the application of machine learning techniques on data appropriation to avoid misrepresentation of covid-19 pandemic progression across related information platforms. Thus, the aggregated covid-19 dataset also highlighted four predictors for analysis of the pandemic progression for infected cases, fully vaccinated, deaths, and least vaccinated. Though the fear of the vaccine has been masked by the tendencies to vaccination but the findings on its vaccination and mortality rate are still a concern as results implied that the top countries with fully vaccinated status are prevented from the risk of being re-infected and spreading the covid-19 virus due to their dispositions to getting vaccinated.

In the vein of a public health crisis, visualizing aggregated covid-19 information on its vaccination and associated mortality stands to avoid doubts and consequently, increase the tendencies for vaccination towards total eradiation of covid-19 and its associated mortality to help healthcare professionals with surveillance and informed decision-making processes.

The results, in consonance with other related studies (Kushwaha *et al*., 2020; Abdelsamad and Karrar, 2021; Mitra *et al.,* 2021; Ebubeogu *et al*., 2022), have shown the vaccination rate for the UK as a case study and also exposed the prevalence of covid-19 infected cases and mortality for fully and least vaccinated persons on a global scale. The study advises from the results and findings that countries with a high prevalence rate of covid-19 cases should be more vaccine-oriented to reduce the number of deaths from the covid-19 virus in the post-covid-19 era.

The study contributes to the pool of research worthily noting the significance of data appropriation for accurate visualization of the covid-19 pandemic progression via an easy-to-understand dashboard to reduce the pandemic spread that may result from misinterpretation or misrepresentation. However, the results and findings of the present study are not subjective to health practitioners but to potential health platforms where the representation of accurate health information is unavoidable. More so, predicting future implications is also critical to avoid a rising trend in casualties from covid-19 mortality as it has been gathered from news channels (CNN) that re-emerging cases have been mildly noticed around Asia and the USA. Hence, further study is paramount to evaluate the aggregated covid-19 dataset with more machine learning models toward predicting the likelihood of zero death from a potential re-emergence of the covid-19 virus.

4. Conclusion

The covid-19 virus may resurface shortly and the introduction of its vaccination has had a tremendous effect on mortality rate reduction as reported in the present study. The appropriation of multiple datasets towards a better representation of covid-19 information with easy understandability and visualizations that promote more accurate trends and progression, particularly for the United Kingdom and across several countries worldwide have been objectively represented in this paper. The development of the machine learning-based dashboard was in no attempt to condemn existing covid-19 dashboards but to facilitate an interactive platform where accurate covid-19 information is vital for decision support towards eradicating the pandemic spread. Thus, the aggregation and appropriation of covid-19 information is an attempt this study has emphasized to make individual meaning of its vaccination and mortality trends that could better affect the goal of information dissemination and collective actions.

5. Conflict of interest

The authors affirm that the information used in this research exists solely in the context of research and no grant was received for the funding of this research.

References

Abdelsamad, A. M. O., Karrar, A. Z. (2021). An interactive dashboard for monitoring the spread of COVID-19 in Sudan. *International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, pp. 1-6.

Alimohamadi, Y., Sepandi, M., Taghdir, M., Hosamirudsari, H. (2020). Determine the most common clinical symptoms in COVID-19 patients: A systematic review and meta-analysis. *Journal of Preventive Medicine and Hygiene*, vol 61(3), pp. 304-311.

Asselah, T., Durantel, D., Pasmant, E., Lau, G., Schinazi, R. F. (2021). COVID-19: Discovery, diagnostics and drug development. *Journal of Hepatology*, vol 74(1), pp. 168-184.

Azzam, T., Evergreen, S., Germuth, A. A, Kistler, S. J. (2013). Data visualization and evaluation. *New Directions for Evaluation*, vol. 139, pp. 7-32.

Biswas, P., Saluja, K. S., Arjun, S., Murthy, L., Prabhakar, G., Sharma, V. K. (2020).

COVID-19 data visualization through automatic phase detection. *Digital Government: Research and Practice*, vol. 1(4), pp. 1-8.

Bowe, E., Simmons, E., Mattern, S. (2020). Learning from lines: Critical COVID data visualizations and the quarantine quotidian. *Big Data & Society*, vol. 7(2), pp. 1-7.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide, USA: SPSS Inc.

Combra, J. L. (2020). Data visualization for the understanding of COVID-19. *Computing in Science & Engineering*, vol. 22(6), pp. 81-86.

David, M., Miller, T. (2018). How to design a dashboard. *Technical Report*, pp. 1-78.

Ebubeogu, A. F., Ozigbu, C. E., Maswadi, K., Seixas, A., Ofem, P., Conserve, D. F. (2022). Predicting the number of COVID-19 infections and deaths in USA. *Globalization and Health*, vol 18(37), pp. 1-10.

Engledowl, C., Weiland, T. (2021). Data (Mis)representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, vol 29(2), pp. 160-164.

Iqbal, M. I., Leon, M. I., Azim, S. M., Mamun, K. A. (2022). Analysing and predicting coronavirus infections and deaths in Bangladesh using machine learning algorithms. SSRN 3671978, pp. 1-16.

Jebril, N. (2020). World Health Organization declared a pandemic public health menace: a systematic review of the coronavirus disease 2019 "COVID-19". SSRN 3566298, pp. 1-12.

Johns Hopkins Coronavirus Resource Center [Internet]. 2022 [cited 2022 Apr 23]. Available from https://coronavirus.jhu.edu/

Kushwaha, S., Bahl, S., Bagha, A. K., Parmar, K. S. (2020). Significant applications of machine learning for COVID-19 pandemic. *Journal of Industrial Integration and Management*, vol. 5(4), pp. 453-479.

Lamb, Y. N. (2021). BNT162b2 mRNA COVID-19 vaccine: First approval. *Drugs*, vol 81(4), pp. 495-501.

Mathur, P., Sethi, T., Mathur, A., Maheshwari, K., Cywinski, J. B., Khanna, A. K., Dua, S., Papay, F. (2020). Explainable machine learning models to understand determinants of COVID-19 mortality in the United States. *medRxiv*. Available from https://doi.org/10.1101/2020.05.23.20110189

Mitra, A., Soman, B., Singh, G. (2021). An interactive dashboard for real-time analytics and monitoring of COVID-19 outbreak in India: A proof of concept. *In proceedings of the 1st Virtual Conference on Implications of Information and Digital Technologies for Development*, Computers and Society. pp. 335–351.

Natarajan, P., Frenzel, J. C., Smaltz, D. H. (2017). Demystifying big data and machine learning for healthcare. *CRC Press*, Florida.

Niaksu, J. (2015). CRISP data mining methodology extension for medical domain. *Baltic Journal of Modern Computing*, vol. 3(2), pp. 92-109.

Schröera, C., Kruseb, F., Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, vol. 181(1), pp. 526–534.

Singh, A., Thakur, N., Sharma, A. (2016). A review of supervised machine learning algorithms. *In 2016 3rd International Conference on Computing for Sustainable Global Development.* INDIACom, pp. 1310-1315.

Storlie, C. B., Pollock, B. D., Rojas, R. L., Demuth, G. O., Johnson, P. W., Wilson, P. M. (2021). Quantifying the importance of COVID-19 vaccination to our future outlook. *In Mayo Clinic Proceedings*. Elsevier, vol 96(7), pp. 1890-1895.

Tarkar, P. (2020). Impact of COVID-19 pandemic on education system. *International Journal of Advanced Science and Technology*, vol 29(9), pp. 3812-3814.

Teh, J. K., Bradley, D. A.: Chook, J. B., Lai, K. H., Ang, W. T., Teo, K. L. (2021). Multivariate visualization of the global COVID-19 pandemic: A comparison of 161 countries. *PLOS ONE*, vol. 16(5), pp. 252-273.

WHO. Coronavirus (COVID-19) Dashboard [Internet]. 2022[cited 2022 Mar 23]. Available from https://covid19.who.int/.