Cluster Analysis of Infections, Recoveries and Deaths Patterns Due to Coronavirus Pandemic in Nigeria

N. O. Adeboye[1]; K. A. Bashiru[2]; O. N. Ogunnusi[3]

[1,2]Department of Statistics,
Osun State University, Osogbo, Nigeria.

[3]Department ofMathematics & Statistics,
Federal Polytechnic, Ilaro, Nigeria.
E-mail: nureni.adeboye@uniosun.edu.ng[1]

***Abstract* - COVID-19 has been recognized globally as a serious pandemic of which its occurrences have posed greater threats to human and economic sustainability. This pandemic has contributed immensely in damaging both the Nigerian as well as the global economy in entirety, thus ongoing modeling and forecasting becomes a priority. This study, on the other hand, examines the pattern of infection, recovery, and deaths at the state level using data derived from Nigerian Center for Disease Control (NCDC) reports over a period of 613 days and explored using K-means and agglomerated hierarchical cluster analysis. The Ward method for distance estimation and the average silhouette methodology for the optimal number of clusters were used as cluster quality indices. The clustering analysis performed on the states in Nigeria revealed homogenous groups of states that can be classified based on their population density and that two clusters may be optimal for grouping the states based on their similarities in the number of confirmed cases, recoveries, and deaths, with accuracy of 88.3%, 89.4%, and 80% respectively. Based on the findings of the investigation, it was revealed that Lagos state has the highest value for the total number of confirmed cases, recovery cases, and death cases compared to all of the other states.**

Keywords: *COVID-19, Data Exploration, Hierarchical Clustering, Pandemic*

I. Introduction

A family of viruses known as COVID-19 is responsible for illnesses such as the common cold, Severe Acute Respiratory Syndrome (SARS), and Middle East Respiratory Syndrome (MERS) (MERS). Coronaviruses are key agents in the development of gastrointestinal diseases in human beings, poultry, and bovines. (Guo *et al.,* 2020; Huang *et al.,* 2019; Oyelola & Adeshina, 2020; Ayinde *et al.,* 2020)

Since the World Health Organization (WHO) proclaimed COVID-19 a pandemic, the entire world has been fighting against this novel coronavirus, and medical professionals have been diligently working on the possibilities of developing a new vaccine that assures its possible therapies. On Thursday, February 27, 2020, Nigeria reported her first confirmed case of the deadly disease after an Italian man tested positive for the virus.

This event occurred on the 27th day of February (Ayinde *et al.,* 2020). A second instance of the virus was discovered in Ewekoro, which is located in Ogun state, on March 9, 2020. A Nigerian citizen who was in contact with an Italian citizen is the second individual to become sick with the disease. Since this event took place, Nigeria has seen an unprecedented explosion in the number of new cases of the virus being reported across the entire country. As of October 31, 2021, it had infected 211,961 people in Nigeria and caused 2,894 deaths since its outbreak in Wuhan, China.

This infection has resulted in the loss of lives, the closure of borders, a total or partial lockdown, a halt in international flights, and a reduction in the price of the fuel pump at the international market, which affected the economy of the country because Nigeria is one of the major oil exporting countries in the world, the loss of jobs, and the closure of all levels of educational institutions.The statistical technique known as cluster analysis is used to analyze data in various ways. It does this by categorizing things into groups, or clusters, according to the degree to which they are connected.

Cluster analysis is an unsupervised learning approach, which means that the number of clusters that already exist in the data before running the model is unknown. This is because the number of clusters is not known before running the model. Cluster analysis, in contrast to a great number of other statistical procedures, is often utilized in situations in which no assumptions are made about the possible relationships contained within the data. It gives information about the locations in the data where associations and patterns can be found.

The process of clustering can therefore be thought of as an optimization problem with several objectives. The individual data set and the purpose for which the results are going to be used both play a role in determining the appropriate clustering algorithm and parameter settings. These settings may include parameters such as the distance function to be used, a density threshold, or the number of expected clusters.

Cluster analysis in and of itself is not a work that can be completed automatically; rather, it is an iterative process of information discovery or interactive multi-objective optimization that requires both success and failure. It is frequently required to make adjustments to the data preprocessing and model parameters to obtain the desired qualities in the final product (Aggarwal, 2013).

Aggarwal and Reddy (2013) used a clustering algorithm and a geographical information system GIS to perform a COVID-19 spread pattern recognition in Iran. This was done to determine the virus spreading possibility from the starting point to the other parts of IranThe primary objective of exploratory data mining in the context of cluster analysis is to carry out pattern identification for infection development monitoring. As is common knowledge, clustering does not refer to a single algorithm but rather to a collection of algorithms that greatly vary from one another in their understanding of what makes a cluster and how to most effectively use those algorithms to complete the task at hand.

This article investigates the spread and clustering pattern of COVID-19 statistics that have been released by the NCDC on daily basis and per state in Nigeria.

II. Methodology

The daily updates of COVID-19 confirmed, recovered, and death cases in Nigeria were plotted using Microsoft Excel 2010 in order to generate an extensible time-series object that can be arranged by time index (NCDC, 2020). This object is depicted in Figures 1 through 6 of the article. R-console version 3.6.2 (R Core Team, 2019) was equally employed in order to further explore the data by employing agglomerated hierarchical cluster analysis. This is one of the machine learning algorithms that is utilized in the process of drawing inferences from unlabeled data, and it is based on the squared Euclidean metric and squared distance approach. This method of data clustering was adopted due to the fact that COVID-19 data is medically oriented and was collected from various testing laboratories in the six (6) geopolitical zones of the federation that made up the thirty-six (36) states plus the FCT, thus looking for a meaningful way from the medical perspective to compare the time series data of COVID-19 as categorized.

2.1 Cluster Analysis

The Euclidean distance metric $d(x, y)$ of elements x and y is represented empirically as

$$d(x, y) = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2} \qquad (1)$$

Where: $x = (x_1, \dots, x_r)$ and $y = (y_1, \dots, y_r)$. Equation (1) for p = 2 and p = 3 is the equivalent to the distance on a plane and in space of two points x and y.

In the case of linkage methods, a wide variety of algorithms can be utilized, most notably single linkage, complete linkage, unweighted pair group average, weighted pair-group average, unweighted pair-group centroid, weighted pair-group centroid, and Ward's minimum variance method. The popular Wards minimal variance approach was used to analyze the COVID-19 clustering pattern, where the criterion for selecting the pair of clusters to merge at each step is based on the optimal value of an objective function, which is the error sum of squares (ESS). The use of that computation approach enables the identification of which states are similar to one another and can be included in the same clusters, as well as to what extent certain clusters are similar to one another and can be combined into structures of larger clusters. The squared Euclidean distance is defined between points as:

$$d_{ij} = d(\{X_i\}, \{X_j\}) = \left|\left|X_i - X_j\right|\right|^2 \qquad (2)$$

The total within sum of squares is calculated based on equation (1) as

$$wws = \sum_{i}^{k} \sum_{x \in C_i}^{k} \left|\left|x - m_i\right|\right|^2 \qquad (3)$$

Where: k - number of clusters; x - element of cluster; C_i- ith data cluster; m_i- centroid of the cluster i, and $\left|\left|x - m_i\right|\right|^2$ Euclidean distance between two vectors. However, the error sum of squares that emerge at each step is given in equation (4) as

$$ESS = \sum_{i}^{N_x} \left| x_i - \frac{1}{N_x} \sum_{x \in C_i}^{N_x} x_j \right|^2 \qquad (4)$$

Where $|\cdot|$ is the absolute value of a scalar value or the norm of a vector.

In addition, average Silhouette method was used for obtaining an optimum number of clusters for further analysis. In achieving this, given that a data has been clustered into k clusters, for data point $i \in C_I$, let

$$a(i) = \frac{1}{|C_I| - 1} \sum_{i \in C_I, i \neq j}^{k} d(i,j) \qquad (5)$$

represents average distance between i and other data points in the same cluster i where $d(i,j)$ is the distance between data points i and j in cluster i; $a(i)$ represents how well i is assigned to its cluster where the smaller the value of i, the better the assignment. In addition, average dissimilarity of point i to some cluster C_J where $C_I \neq C_J$ was used to define $b(i)$ for each data point $i \in C_I$ given as:

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{J \in C_J} d(i,j) \qquad (6)$$

Cluster with the smallest mean dissimilarity is said to be the "neighboring cluster". Hence, we define the Silhouette value for a data point from equation (5) and (6) as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \ , \ if\ |C_I| > 1 \qquad (7)$$

and $s(i) = 0$ if $if\ |C_J| = 1$

The metrics were computed as the sum of the diagonal elements of the confusion matrix, divided by the number of samples to get a value between 0 and 1. The best match between the class and cluster labels was computed using an indicator function. The accuracy is defined as

$$accuracy(y, \hat{y}) = \max_{perm \in P} \ \frac{1}{n} \sum_{i=0}^{n-1} 1(perm(\hat{y}_i) = y_i) \qquad (10)$$

Where P is the set of all permutations in [1, K] where K is the number of clusters.

The datasets were scaled before carrying out the cluster analysis. In achieving the clustering pattern technique, packages used in R include *dplyr*, *ggplot2* for data analysis and visualization. Packages such as *cluster*, *factoextra*, *dendextend* and *tidyverse*, were also used for more advanced data analysis and visualizations. *read.csv* package was also used to import the *csv comma delimited* file to R console.

III APPLICATIONS

The descriptive statistics of the reported pandemic datasets are displayed in Table 1. Values of 379.7, 372.1, and 5.2 indicate that, on the average, hundreds of COVID-19 infection cases were confirmed and recovered on daily basis, with around five deaths (in units). The variance of the average cases from the daily records reveals that most daily confirmed cases, recovery cases, and deaths were high and far away from the mean, as the maximum daily confirmed cases, recovery cases, and deaths were 5908, 8228, and 93, respectively. The large standard deviation value of the daily recovery cases indicates that the data points are more dispersed and far from the mean, whereas the daily death cases were found to be grouped around the mean value. Taking into account the skewness of the dataset, which measures the degree to which the probability distribution deviates from the normal distribution, it was determined that the daily confirmed cases, recovery cases, and death cases were highly positively skewed, indicating that the recorded COVID-19 infection cases were not normally distributed. This is supported by the JB test statistic with a p-value 0.05 level of statistical significance.

Table 1: Descriptive statistics for daily number of Nigeria COVID-19 infections

Statistics	Daily confirmed	Daily recovered	Daily death
Minimum	0.00	0.00	0.00
Maximum	5908.00	8228.00	93.00
Mean	379.7	372.1	5.246
Standard deviation	545.42201	671.19718	7.544834
Median	225.0	178.0	3.00
Q1	84.0	50.0	0.00
Q3	519.0	353.0	8.00
Skewness	5.426698	5.280628	4.318414
Kurtosis	47.408547	44.424424	38.594080
Jarque Bera	53380{0.000}	46678{0.000}	34265{0.000}

Figures in parenthesis represent the P-value

The daily reported laboratory-confirmed cases, recovered cases, and deaths are depicted in Figs. 2, 3, and 4. The plots reveal an exponential increase in the daily number of reported cases from February 2020 with an irregular surged in October 2020, followed by a gradual decline in the number of recorded cases until April 2021 and a gradual increase in May 2021. However, the recovery rate was found to be considerable across the federation, since the majority of survivors reacted daily to therapy. Infected patients in Nigeria appear to have a greater resistance to the disease and have largely adhered to the precautions outlined by the National Center for Disease Control (NCDC) in order to expedite their recovery. Though there were occasional surged the patients' recovery cases between September and November 2020, June 2021, August 2021 and October 2021 while the surged in deaths was recorded in October 2020 and June 2021.

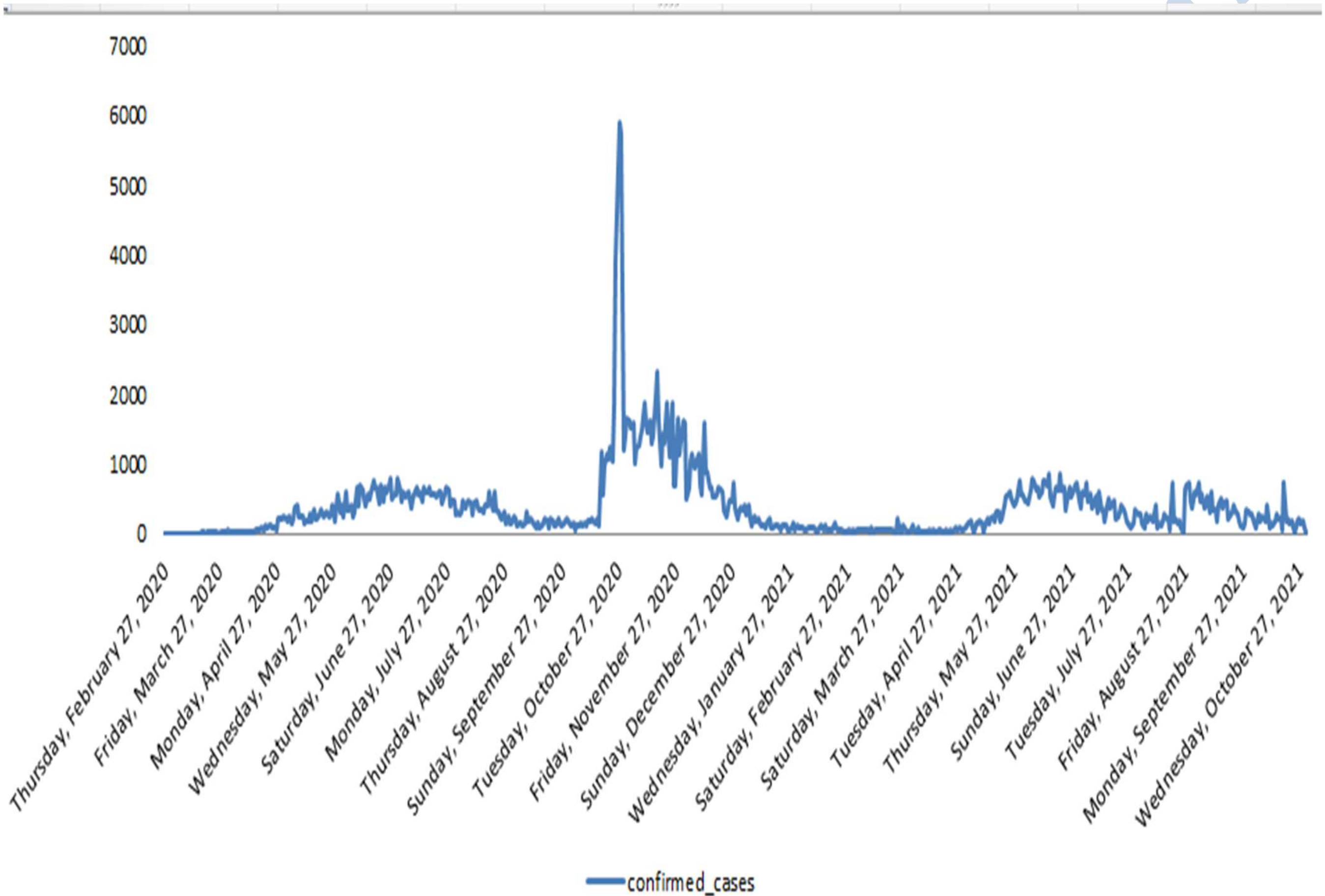


Fig. 1: Daily Confirmed Cases of COVID-19 infection in Nigeria between February 2020 to October, 2021

Spikes in recovered cases of the pandemic within the time frame were due to the proactive response of health workers and the provision of healthcare facilities which facilitate the recovery rate of the infected patients.

From Fig 2, the spike in death cases in the month of June 2021 is attributed to inadequate preparation of member Nigerian states to combat the virus, poor health infrastructure, which lead to poor health surveillance and response system, ill-equipped health facilities for efficient management of reported cases and low patients-health workers' ratio to cater for the infected citizens. Relating the spike between the death and recovered COVID-19 cases in the country might also be attributed to the heterogeneity of the number of reported cases of the virus over time as some states such as Lagos, Rivers and FCT were found to record more of the cases with these states recording surge in fatalities and recoveries over time.

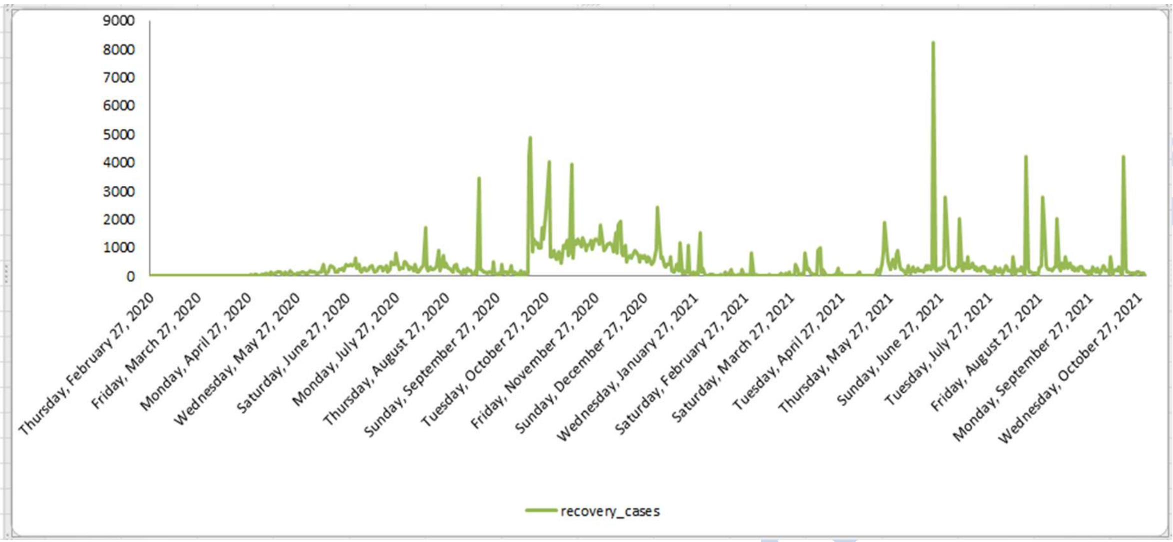


Fig. 2: Daily Recovered Cases of COVID-19 infection in Nigeria between February 2020 to October, 2021

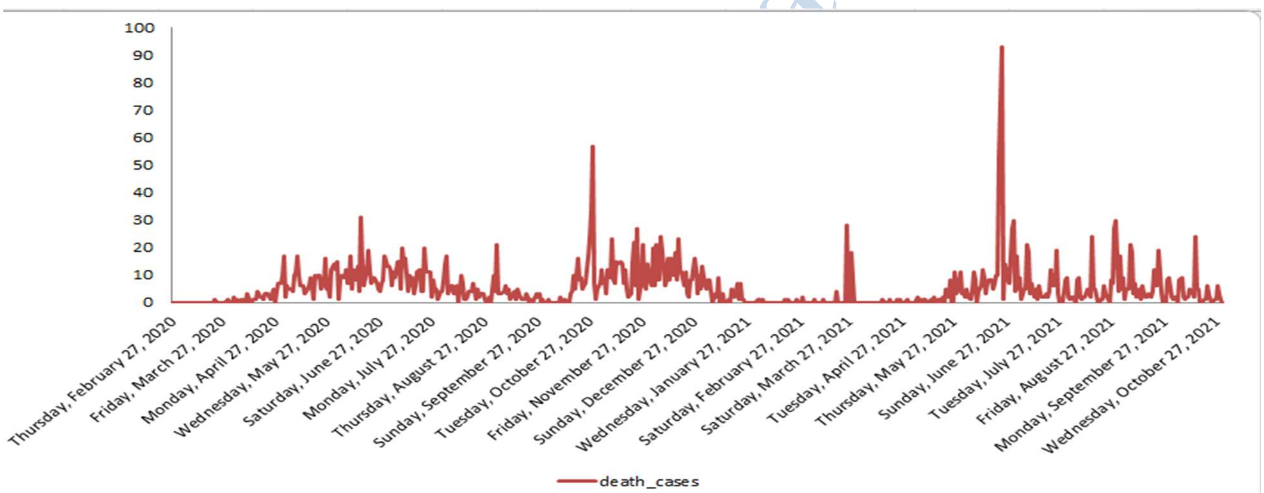


Fig. 3: Daily Death Cases of COVID-19 infection in Nigeria between February 2020 to October, 2021

3.2 Analysis of COVID-19 Cases by States

Fig. 4 depicts the number of infected persons per state. It was evidenced that Lagos state has the highest number of confirmed cases (77,808) followed by FCT (23,261) while Kogi recorded the least (5) number of cases within the last 613 days.

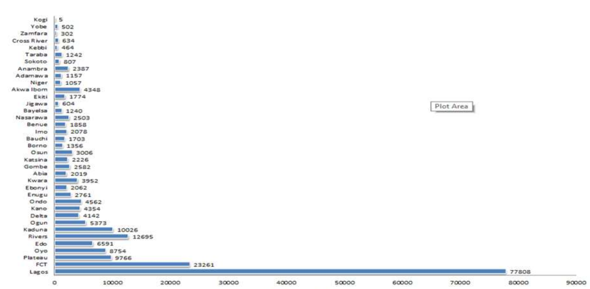


Fig. 4: Evolution of COVID-19 confirmed cases in Nigerian States within the last 613 days

From the evolution of the recovered cases of the pandemic presented in Fig 5, Lagos was found to have the highest recovery rate among the other states of the federation including FCT, justifying its classification as the epic center of the pandemic. The recovery was followed by FCT with the least recovery cases recorded by Kogi. It is pertinent to note that all these recoveries are proportional to the rate of infection per state.

Deaths due to the pandemic per states are presented in Fig 7. The number of deaths in Lagos state compared to the confirmed cases in the state, showed a significant improvement in the management of the pandemic on the part of the state government. Edo state that was ranked 5th in terms of infected cases recorded the second highest deaths due to the pandemic, superseding FCT, Plateau and Oyo states respectively. The mortality rate in Lagos state compared to the confirmed cases is estimated at 1.05% while that of Edo state is estimated at 4.07%; an attestation to the reactiveness of Lagos state in handling the pandemic cases.

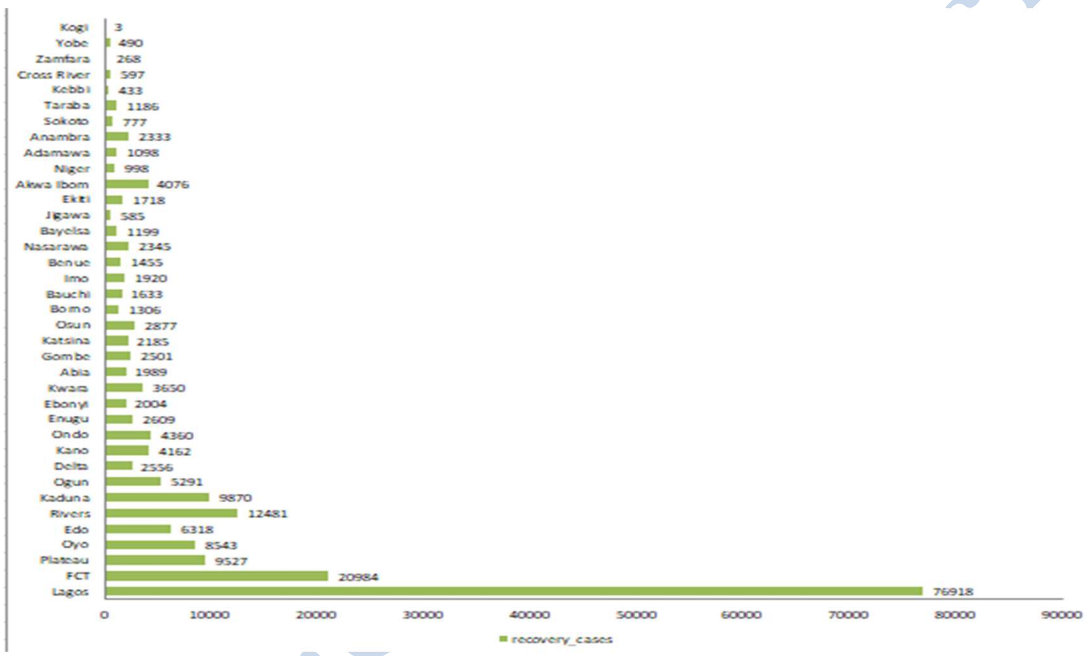


Fig. 5: Evolution of COVID-19 recovered cases in Nigerian States within the last 613 days

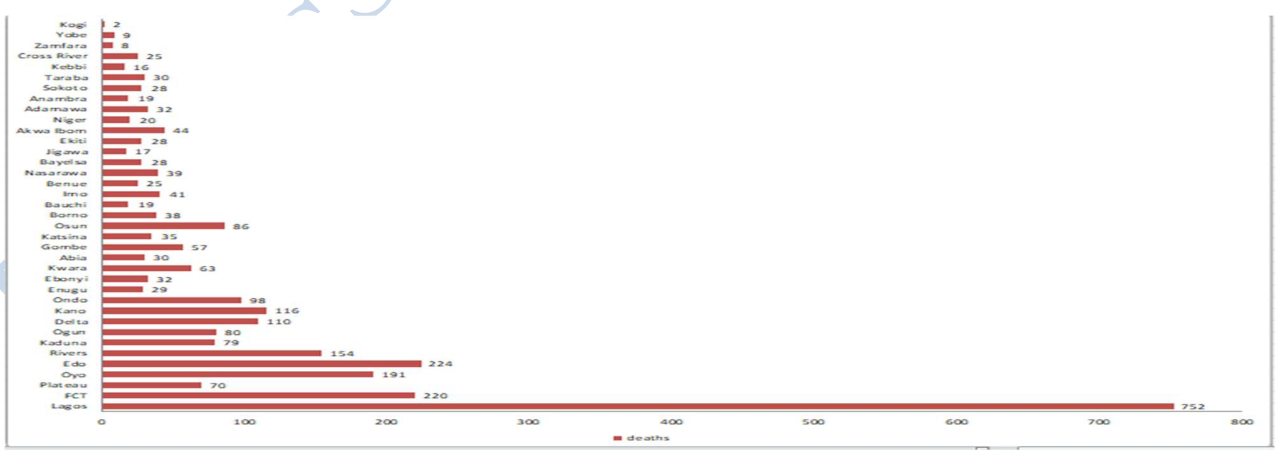


Fig. 6: Evolution of COVID-19 deaths in Nigerian States within the last 613 days

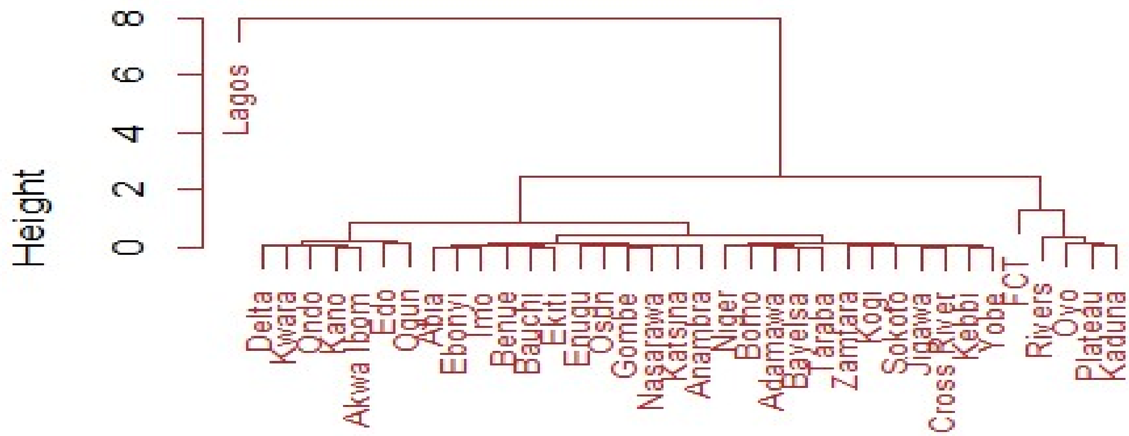


Fig. 7: Cluster Dendrogram for confirmed cases

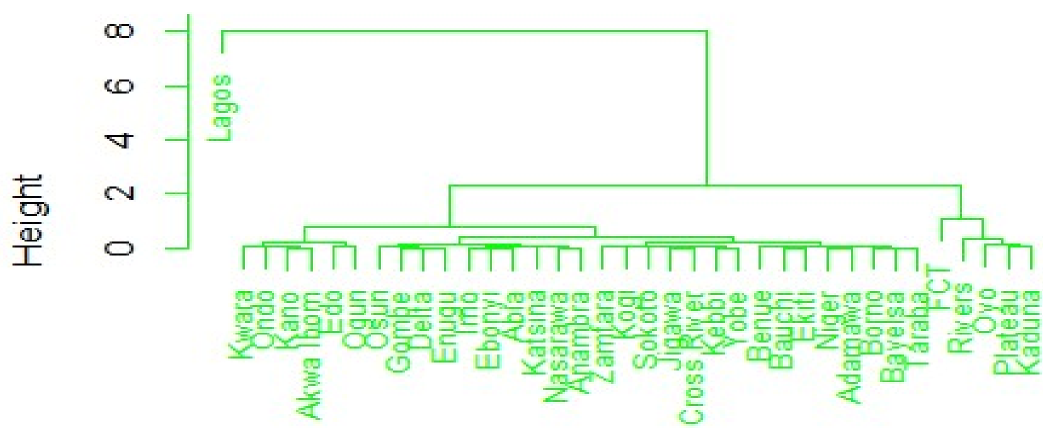


Fig. 8: Cluster Dendrogram for Recovered cases

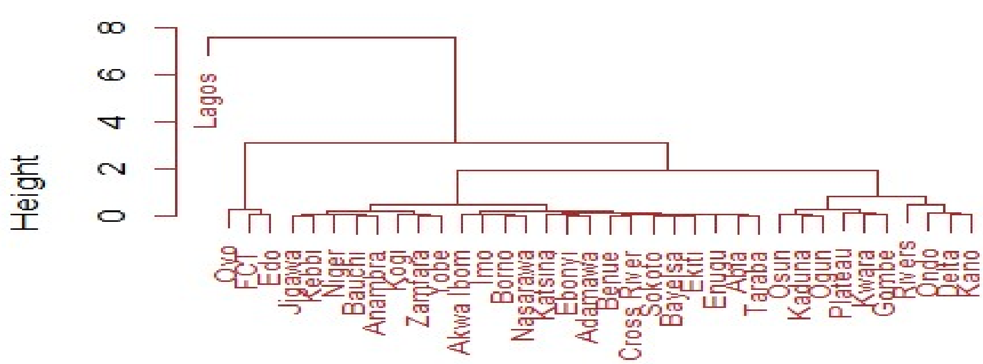
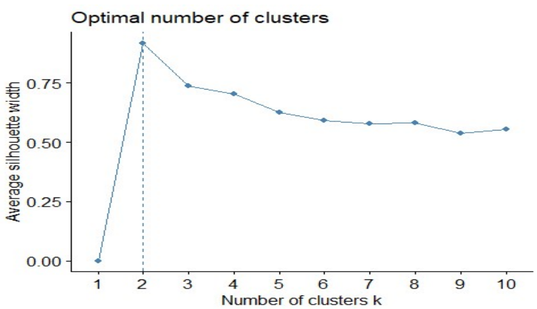


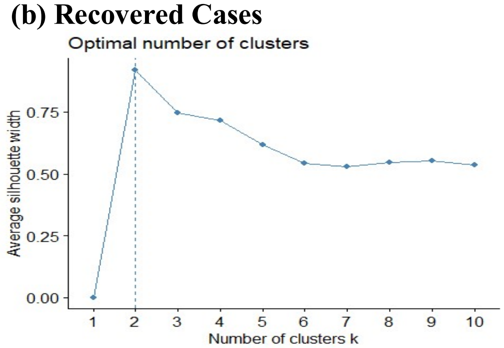
Fig. 9: Cluster Dendrogram for Death cases

Dendrograms for each of the cases with division into measurement states are presented in Figs 7-9. Axis y depicts linkage distance s for particular cases (confirmed, recoveries and deaths) and groups (clusters) of patterns. The trees suggest the occurrence of clusters, i.e., groups of cases patterns with similar histograms. Strong outlier elements (histograms) were also identified, for example, Lagos State for each of the confirmed cases, recoveries and deaths. (Figure 7-9) across the federation. However, it was found that there exist five (5), six (6) and five (5) clusters formed from the confirmed cases, recovered cases and deaths datasets respectively.

(a) Confirmed Cases



(b) Recovered Cases



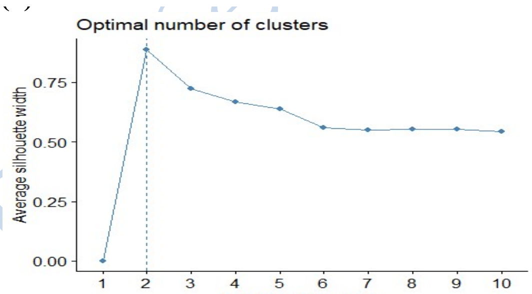


Fig. 10: Average Silhouette Method for obtaining Optimum number of clusters (a-c)

Average Silihouette method for optimal number of clusters for each of the variables can be evidenced in Fig. 10 (a-c), indicating an optimum number of $k = 2$ clusters for the recorded number of covid-19 confirmed cases, recoveries and deaths.

Table 2 shows the memberships of states for covid-19 confirmed cases according to the k-means clustering method. Although, $k = 3$ indicated 96.3% accuracy in classification of the states to their respective clusters compared to the optimal k =2 cluster of 88.3% accuracy. However, FCT, Plateau, Oyo, Rivers and Kaduna states were a single cluster. The distribution by all other states by clusters of confirmed cases can be seen in detail in the table. It cannot be overemphasized that from the confirmatory analysis of result findings, the clustering pattern of Nigerian state confirmed COVID-19 cases were found to be homogenous in nature except for Lagos state, which is heterogeneous due to its dense population which signifies the virus epicentre. The cluster means for clusters 1 and 2 for $k = 2$ and $k = 3$ are negative values of -0.1544466 and -0.2685855 respectively. In the theory of k-means clustering, there is a possibility of having negative cluster means. However, the negative value must not be up to -1. Otherwise, it generally indicates that a sample has been assigned to the wrong cluster. More so, zero (0) mean values indicated an overlapping cluster (See Rousseeuw, 1987).

The within-cluster sum of squares is a measure of the variability of the observations within each cluster. In general, a cluster that has a small sum of squares is more compact than a cluster that has a large sum of squares. Thus in this research, only Lagos state was classified as a group for all the cases, hence it is expected to have zero value for its within-clusters sum of squares.

The zero value of within-cluster sum of squares by the second cluster at $k = 2$ and $k = 3$ as applied to Lagos was due to its single object classified as cluster. As a result of this, the sum of square from the single observation and the cluster centroid will be zero, indicating that the recorded confirmed cases, recoveries and death in Lagos is dissimilar to other states of the Federation.

Table 3 shows the memberships of states COVID-19 recovered cases according to the k-means clustering method. Although, $K = 3$ indicated 97.1% accuracy in classification of the states to their respective clusters compared to the optimal $k = 2$ cluster of 89.4% accuracy. However, FCT, Plateau, Oyo, Rivers and Kaduna states were a single cluster. The distribution by all other states by clusters of recovery cases can be seen in detail in the table. The $K = 2$ clusters were found to be high in precision as depicted from the average silhouette result of Fig. 10(b).

Table 2: Cluster membership of states according to the k-means clustering for daily confirmed cases of Nigeria COVID-19 infections.

Cluster	States	Within Cluster Sum of Squares by Clusters	Cluster Means
***K* = 2**			
1	FCT, Plateau, Oyo, Edo, Rivers, Kaduna, Ogun, Delta, Kano, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	4.226796	-0.1544466
2	Lagos	0.000000	5.5600781
Accuracy: 88.3%			
***K* = 3**			
1	Edo, Ogun, Delta, Kano, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	0.4701208	-0.2685855
2	FCT, Plateau, Oyo, Rivers, Kaduna	0.8488935	0.5532148
3	Lagos	0.0000000	5.5600781
Accuracy: 96.3%			

Source: *Extracted from R Output*

Table 3: Cluster membership of states according to the *K*-means clustering for daily recovery cases of Nigeria COVID-19 infections.

Cluster	States	Within Cluster Sum of Squares by Clusters	Cluster Means
***k* = 2**			
1	FCT, Plateau, Oyo, Edo, Rivers, Kaduna, Ogun, Delta, Kano, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	3.805079	-0.1554682
2	Lagos	0.000000	5.5968552
Accuracy: 89.4%			
***k* =3**			
1	Edo, Ogun, Delta, Kano, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	0.4282963	-0.2663339
2	FCT, Plateau, Oyo, Rivers, Kaduna,	0.6333851	0.5318993
3	Lagos	0.0000000	5.5968552
Accuracy: 97.1%			

Source: *Extracted from R Output*

Table 4: Cluster membership of states according to the k-means clustering for daily death cases of Nigeria COVID-19 infections.

Cluster	States	Within Cluster Sum of Squares by Clusters	Cluster Means
	***k* = 2**		
1	FCT, Plateau, Oyo, Edo, Rivers, Kaduna, Ogun, Delta, Kano, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	7.195268	-0.147055
2	Lagos	0.000000	5.293980
	Accuracy: 80.0%		
	***k* =3**		
1	Plateau, Kaduna, Ogun, Ondo, Enugu, Ebonyi, Kwara, Abia, Gombe, Katsina, Osun, Borno, Bauchi, Imo, Benue, Nasarawa, Bayelsa, Jigawa, Ekiti, Akwa-Ibom, Niger, Adamawa, Anambra, Sokoto, Taraba, Kebbi, Cross River, Zamfara, Yobe, Kogi.	1.0701793	-0.3193872
2	FCT, Oyo, Edo, Rivers, Delta, Kano.	0.7793798	0.7146059
3	Lagos	0.0000000	5.2939802
	Accuracy: 94.9%		

Source: *Extracted from R Output*

Table 4 depicts the memberships of states COVID-19 death cases according to the k-means clustering method. k = 3 indicated 94.9% accuracy in classification of the states to their respective clusters compared to the optimal K =2 cluster of 80.0% accuracy. However, FCT, Oyo, Edo, Rivers, Delta, Kano states were a single cluster taking into consideration k = 3 as the optimal K = 2 cluster was found to be précised as evidenced from the average silhouette test for optimal clusters. Difference in cluster classification and accuracy might be due to the unsupervised nature of the methodology adopted which resulted in allowing the cluster model to discover the data patterns and information. The distribution by all other states by clusters of death cases can be seen in detail in the table. This generally implies that both confirmed, recovered and death cases of COVID-19 infections per states in Nigeria were attributed to the same pattern since same states were found to form clusters taking two optimal clusters into consideration

IV. CONCLUSION

The clustering analysis performed on the states in Nigeria revealed homogenous groups of states that can be classified based on their population density. Based on the findings of the investigation, it was revealed that Lagos State has the highest value for the total number of confirmed cases, recovery cases, and death cases compared to all of the other states. The provision of healthcare facilities by the government, in which patients afflicted with the disease are taken to for treatment, contributed to a higher record of recovery cases among the states. This was the cause of the higher recovery rate among the states. The steadily increasing number of confirmed illnesses and fatalities has been used to categorize states according to their total populations. As a result, it is now clear which states share similar characteristics and which states have distinct characteristics taking the COVID-19 pandemic into consideration. In this regard, the responses that the states that make up the same group have adopted in response to the pandemic are comparable.

When the findings of the survey were analyzed, it was discovered that the number of total cases in Lagos, the Federal Capital Territory (FCT), Rivers, and Plateau was much greater than the average number of cases in the other states. The fact that more precise clustering could be made by including a wider variety of socio-demographic data of the states is one of the limitations of this study. In this particular investigation, socio-demographic aspects of the states were not taken into account. The findings of this research will be extremely significant when it comes to illustrating the disparities that exist between the states in terms of the number of testing devices per one million residents. Because of this information, governments will be able to put appropriate measures into place for future situations that are comparable and the clustering results could be valuable for a variety of policymakers, including data scientists, epidemiologists, academics, finance experts, and statisticians, who are training machine learning models to anticipate future COVID-19 pandemic patterns in Nigeria and around the world.

References

Aggarwal, C. C. & Reddy, C. K. (2013). Data Clustering: Algorithms and Applications. *Chapman and Hall.*

Auguie, B. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics, 2017. *R package version* 2.3. https://CRAN.R project.org/package=gridExtra.

Ayinde, K, Adewale, F. L, Rauf, I. R., Alabi, O. O., Okon, C. E. & Ayinde, E. O. (2020). Modeling Nigerian Covid-19 cases: A comparative analysis of models and estimators. *Chaos, Solitons and Fractals* 138, 109911. https://doi.org/10.1016/j.chaos.2020.109911

Broman, K. W., & Woo, K. H. (2020). Data organization in spreadsheets. *The American Statistician*, 72(1), 2-10.

Galili, T. (2015). dendextend: An R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*. DOI: 10.1093/bioinformatics/btv428

Guo, Y., Cao, Q., Zhong-Si, H. & Yan, Y. (2019). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- an update on the status. *Military medical research,7*1. Doi: 10.1186/s40779-020-00240-0.

Huang, C, Y., Wang, L. X., Ren, L., Zhao. J. & Hu, Y. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 39(5), 497-506. doi: 10.1016/s0140-6736(20)30183-5

Kassambara, A. & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses, 2020. *R package version 1.0.7.* https://CRAN.R-project.org/package=factoextra

Kumar, S. (2020). Use of cluster analysis to monitor novel coronavirus-19 infections in Maharashtra India. Indian *Journal of Medical Sciences,* 72(2): 44-48. doi: 10.25259/IJMS_68_2020

Maechler, .M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K., (2013). Cluster Analysis Basics and Extensions. R package version 1.14.4. NCDC COVID-19 Microsite. https://covid19.ncdc.gov.ng/.

Ogundepo, E., Folorunso, S., Adekanmbi, O., Akinsande, O., Banjo, O. E., Ogbuju, F., Oladipo, O., Abimbola, E. & Oseghale, O. B. (2020). An Exploratory Assessment of a Multidimensional Healthcare and Economic Data on COVID-19 in Nigeria, *Data in Brief* (2020), doi: https://doi.org/10.1016/j.dib.2020.106424

Oyelola, A. A., Adeshina, I. A. & Ezra, G. (2020). Early transmission Dynamics of Novel Coronavirus (COVID-19) in Nigeria (2020). *International Journal of Environmental Research and Public Health.* 1(7)*, 30-54 doi:10.3390/ijerph17093054*

Rousseeuw, P. J. (1987). Silhouettes: aGraphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics 20: 53-65.* Httsp://en/Wikipedia.org/wiki/Silhouette (clustering)

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL https://www.R-project.org/.

Wickham, H. (2020). ggplot2: Elegant graphics for data analysis. *Springer-Verlag, New York*

Wickham, H, Romain, F., Lionel, H. & Kirill, M. (2020). dplyr: A grammar of data manipulation, 2020. *R package version* 1.0.0. https://CRAN.R-project.org/package=dplyr

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino, L., McGowan, François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686- https://doi.org/10.21105/joss.01686

Zarikas, V., Poulopoulos, S. G., Gareiou, Z. & Zervas, E. (2020). Clustering analysis of countries using the COVID-19 cases dataset. *Data in Brief 31:105787*.doi: https://doi.org/10.1016/j.dib.2020.105787