Country Clustering Analysis of COVID-19 Pandemic Dynamics: Insights and Implications for Future Preparedness

Putranegara Riauwindu, Boston University, Boston, United States, putrangr@bu.edu

Abstract:

The COVID-19 pandemic has presented diverse challenges to nations worldwide, necessitating a comprehensive understanding of countries' varied dynamics and responses to the crisis. This paper conducts an analysis of multiple features to explore the dynamics observed during the pandemic, including epidemiological statistics, demographic indicators, economic metrics, geographical information, health indicators, mobility patterns, weather, and government responses. Using data sourced from the Google Health COVID-19 Open Data Repository, this study employs clustering techniques to categorize countries into three distinct clusters. By discerning patterns across different clusters, this study aims to aid in formulating effective strategies and policies, with potential implications for bolstering global preparedness in the event of future health emergencies.

1. Research Introduction, Scope, and Objectives

The COVID-19 pandemic has served as a stern reminder of the interconnectedness of nations and the need for collective action in the face of global challenges [1]. Nevertheless, amidst the challenges, the pandemic has also presented unique opportunities for introspection and growth [2], necessitating exploration of the multifaceted nature of the COVID-19 pandemic, examining country dynamics during the pandemics while drawing lessons on the opportunities that have emerged from this crisis to build resilience. This research will focus on clustering analysis of countries, utilizing various features such as epidemiological data, demographics, economy, geography, health indicators, mobility patterns, weather information, and government response metrics. The objective is to retrospectively assess country performance during the pandemic and

categorize them into distinct clusters. The analysis aims to extract valuable insights and lessons learned from each cluster's dynamics, providing knowledge for future pandemic preparedness.

2. Literature Review

Previous research on clustering analysis for the COVID-19 pandemic primarily focused on grouping countries based on active cases. For instance, Zarikas et al. [3] conducted a study clustering countries according to similar case time evolution using Johns Hopkins epidemiological data. Similarly, Sadeghi et al. [4] ranked and clustered 180 countries based on COVID-19 cases and fatality, comparing their results with existing pandemic vulnerability prediction models. This paper research utilizes the same clustering approach but with a different method, using k-means instead of hierarchical clustering, while incorporating various multi-faceted features.

3. Methodology and Process

The first step involved extracting raw data from the Google COVID-19 Open Data Repository [5]. Subsequently, a data preprocessing stage was conducted to ensure the dataset's reliability and integrity. This entailed cleaning, manipulating, and combining the data to create a unified dataset suitable for clustering analysis. The k-means clustering algorithm was employed for clustering analysis, supported by elbow plot analysis and iteration of several number of cluster centers to determine the optimum number of clusters. All analysis were done in Jupyter Notebook using Python. The resulting country cluster model was then explored to extract valuable insights and patterns. Each cluster was named based on its distinctive pandemic response characteristics. To enhance the comprehension of findings, visualizations were employed to summarize the insights.

4. Results and Discussion

The k-means clustering analysis yielded three clusters of country as the optimum number of cluster centers with strong distinction between clusters. The three clusters distribution and member

countries along with some visualizations of correlation between features are illustrated in the Appendix 1 to 5. The resulting Jupyter Notebook Markdown could also be accessed here. The explanation for each of the clusters are as follows:

4.1. Cluster 1: Resilient Urban Health Capitals

The Resilient Urban Health Capitals cluster includes countries with the highest COVID-19 cases, showing significant pandemic challenges. Characterized by high urbanization, robust economies, and strong healthcare systems, they have displayed resilience in combating the outbreak. Their notable health indicators, including well-equipped healthcare and high health expenditure, likely contributed to their effective pandemic response. Despite medium diabetes prevalence, they have the highest life expectancy and lowest pollution and comorbidity mortality rate, indicating better population health conditions. This cluster exhibited the lowest level of government response stringency compared to the other clusters. It suggests that the governments might have faced challenges in implementing robust public health measures, or it could reflect a strategic approach towards achieving herd immunity. The lower government stringency may have influenced mobility patterns and overall pandemic dynamics within this cluster [6]. In terms of mobility, the cluster displayed moderate mobility to essential establishments like retail, recreation, grocery, and pharmacy outlets, as well as transit stations. Notably, workplace mobility was lowest, signifying a significant shift to remote work arrangements during the pandemic. This adaptability showcased their flexibility in response to changing circumstances. Moreover, medium mobility to residential areas indicated adherence to social distancing guidelines, striking a careful balance between sustaining economic activities and curbing virus spread. The cluster's pandemic response reflected an effective blend of economic continuity and prioritization of public health. Their adaptability, robust healthcare system, and resilience, despite the lower government stringency,

have likely played a crucial role in mitigating virus transmission and maintaining an effective response to the COVID-19 outbreak.

4.2. Cluster 2: Vulnerable Rural Health Deficit

This includes countries with the lowest cumulative confirmed, deceased, and tested COVID-19 cases, indicating comparatively lower infection rates. The cluster is characterized by the highest rural population and lowest urban population, presenting challenges in healthcare access and resource allocation. The lowest human development index and economic indicators signify limitations in pandemic response and healthcare infrastructure. Geographically, countries in this cluster have smaller footprints, offering advantages in coordination but challenges in resource allocation. Health indicators highlight vulnerabilities, with lower life expectancy and weaknesses in healthcare systems. Mobility patterns indicate high activity in retail, recreation, grocery, and pharmacy establishments, with limited remote work and stay-at-home compliance which indicates that the people in this cluster still need to go to the store to procure essential needs. Government responses indicate moderate stringency, potentially affecting pandemic containment measures [6]. In summary, the cluster represents countries facing unique pandemic challenges due to rural populations, economic constraints, and healthcare vulnerabilities. Despite lower confirmed cases, demographic and mobility factors may have contributed to their pandemic experiences. Addressing healthcare disparities and understanding rural virus transmission dynamics are crucial for effective pandemic management in these countries.

4.3. Cluster 3: Urban Diabetes Burden & Residential Refuges

This cluster includes countries with moderate COVID-19 cases, indicating a balanced infection rate within their borders. These countries strike a middle-ground in terms of demographics, with both rural and urban populations coexisting. This distribution presents unique

challenges and opportunities for pandemic management, as urban centers may have higher population densities and mobility, while rural areas could have limited healthcare access.

Regarding mobility, the data reveals interesting patterns within this cluster. There is lower mobility to retail, recreation, grocery, and pharmacy establishments, as well as to park and transit stations. This suggests that people in these countries were relatively less engaged in non-essential activities and public spaces during the pandemic, which might have contributed to controlling the spread of the virus to some extent. However, there was moderate mobility to workplaces, indicating a flexible mix of remote work and on-site arrangements. This adaptability showcases the countries' response to pandemic-induced changes in work settings. Moreover, the highest mobility to residential areas within this cluster is particularly noteworthy, indicating a preference for staying at home during the pandemic. This shift towards residential refuges could have played a crucial role in mitigating virus transmission by limiting interactions in public spaces and breaking the chain of infection. The cluster's health indicators showed particularly the highest percentage of people with diabetes, the emphasis on residential refuges gains significance. By reducing mobility to non-essential places and increasing time spent in residential areas, vulnerable populations, such as those with diabetes, could better protect themselves from potential exposure to the virus. This approach likely played a role in controlling the pandemic's impact within the cluster. Emphasizing targeted measures for vulnerable populations, especially those with pre-existing health conditions like diabetes, may be crucial in managing the pandemic's impact in this cluster and beyond.

5. Summary

In conclusion, the country clustering analysis of the COVID-19 pandemic has provided insights into the diverse dynamics observed among different groups of countries. Understanding these distinct clusters can aid in formulating effective strategies and policies for future pandemics, enhancing global preparedness and response efforts.

References

[1] Seshaiyer P, McNeely CL. Challenges and Opportunities From COVID-19 for Global Sustainable Development. World Med Health Policy. 2020 Dec;12(4):443-453. doi: 10.1002/wmh3.380. Epub 2020 Dec 1. PMID: 33362943; PMCID: PMC7753342.

[2] Pradhan, P., Subedi, D. R., Khatiwada, D., Joshi, K. K., Kafle, S., Chhetri, R. P., et al. (2021). The COVID-19 pandemic not only poses challenges, but also opens opportunities for sustainable transformation. Earth's Future, 9, e2021EF001996. https://doi.org/10.1029/2021EF001996

[3] 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. https://doi.org/10.1016/j.dib.2020.105787

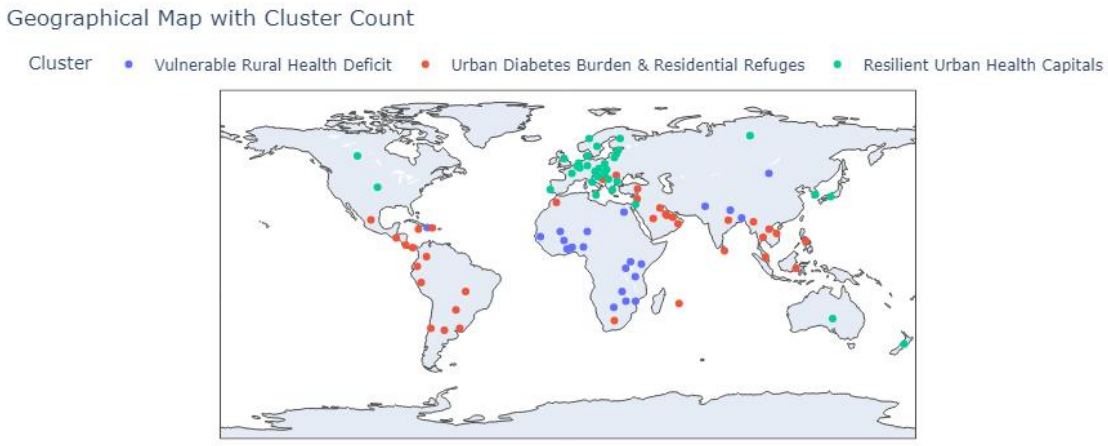
[4] Sadeghi, B., Cheung, R. C., & Hanbury, M. (2021). Using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020. BMJ Open, 11(11). https://doi.org/10.1136/bmjopen-2021-049844

[5] Wahltinez, Cheung, Alcantara, R. *et al.* COVID-19 Open-Data a global-scale spatially granular meta-dataset for *coronavirus* disease. *Sci Data* **9**, 162 (2022). https://doi.org/10.1038/s41597-022-01263-z

[6] Thomas Hale, et.al., "Variation in Government Responses to COVID-19" Version 15. Blavatnik School of Government Working Paper. June 2023. Available: www.bsg.ox.ac.uk/covidtracke

Appendix

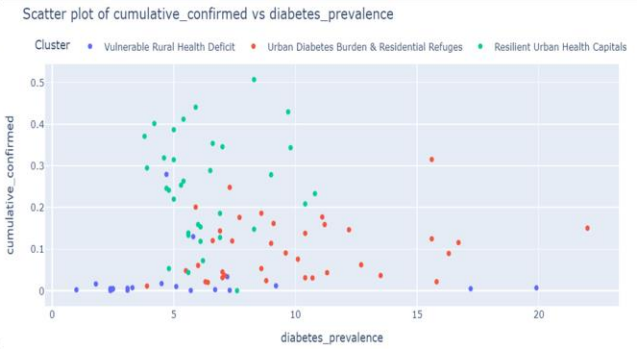
1. Cluster Country Distribution

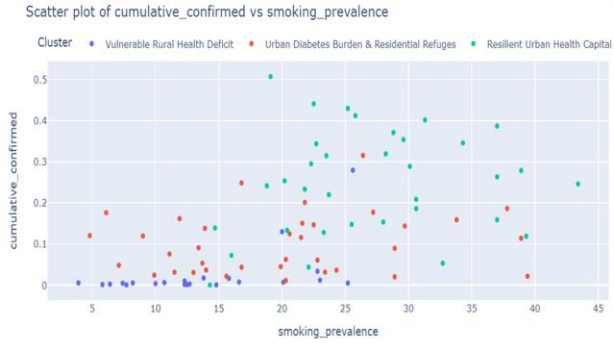


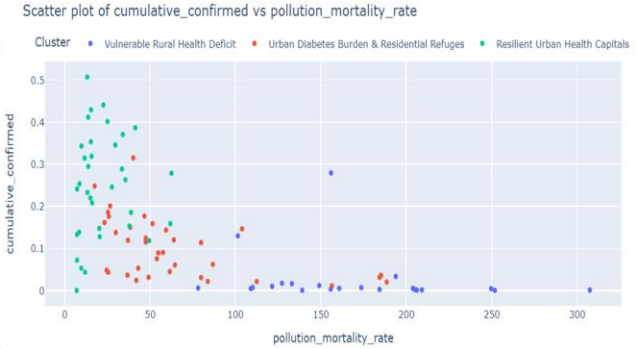
2. Cluster Country Members

Vulnerable Rural Health Deficit	Urban Diabetes Burden & Residential Refuges	Resilient Urban Health Capitals
Bangladesh	Argentina	Australia
Benin	Bahrain	Austria
Botswana	Bosnia and Herzegovina	Belgium
Burkina Faso	Brazil	Bulgaria
Egypt	Chile	Canada
Ghana	Colombia	Croatia
Haiti	Costa Rica	Czech Republic
Kenya	Dominican Republic	Denmark
Mali	Ecuador	Estonia
Mongolia	El Salvador	Finland
Mozambique	India	France
Nepal	Indonesia	Germany
Niger	Jamaica	Greece
Nigeria	Kuwait	Hungary
Pakistan	Laos	Israel
Rwanda	Lebanon	Italy
Senegal	Malaysia	Japan
Tanzania	Mauritius	Latvia
Togo	Mexico	Lithuania
Uganda	Morocco	Luxembourg
Zambia	Myanmar	Malta
Zimbabwe	Oman	Netherlands
	Panama	New Zealand
	Paraguay	Norway
	Peru	Poland
	Philippines	Portugal
	Qatar	Russia
	Romania	Serbia
	Saudi Arabia	Slovakia
	South Africa	Slovenia
	Sri Lanka	South Korea
	Thailand	Sweden
	Turkey	United Kingdom
	United Arab Emirates	United States of America
	Uruguay	
	Vietnam	

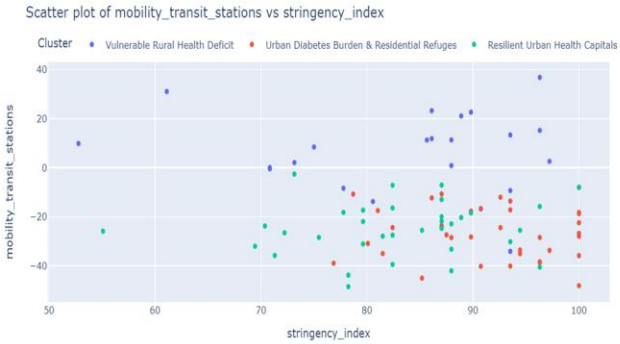
3. Cluster Epidemiology vs Health

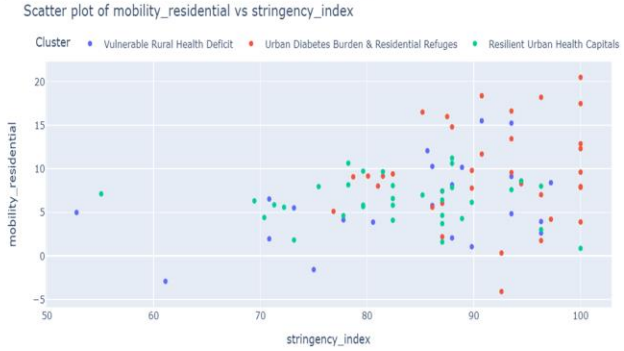






4. Cluster Mobility vs Government Stringency





5. Cluster Epidemiology vs Human Capital and Resources