

---

# Service quality of hospital in Nigeria using Geo-semantic analysis of citizen's social media post

---

**Morolari Boluwatife**

Data Science Nigeria

Lagos, Nigeria.

morolaribolu@gmail.com

## ABSTRACT

Healthcare providers are dealing with life, which is why patient satisfaction is critical. Recently, patients have become more aware and conscious, believing that high-quality care translates to better satisfaction. In Nigeria, hospitals have been faced with a myriad of challenges ranging from inadequate facilities, equipment, medical professional shortages and substandard standards. Social media serves as a platform for citizens to voice their healthcare service experiences and opinions. This study aimed to profile the service quality of hospital in Nigeria by analysing citizens social media data from twitter, using geo-tagging and keyword search techniques. The data was collected and analysed using Natural language processing techniques and sentiment analysis to identify the sentiment and opinions expressed by citizens regarding their experiences with hospitals. After analysing the data the study shows that the overall sentiment towards hospital service quality in Nigeria was negative, with 80% of the post expressing negative sentiment, while only 10% of the post expressed positive sentiment. Geo-semantic analysis of the data showed that the major keywords used by citizens were “poor service delivery,” “inadequate facilities,” “long waiting times,” and “lack of equipment.” These keywords were mostly used in posts originating from major cities in Nigeria, such as Lagos, Abuja, and Oyo. The spatial analysis of the data showed that the majority of the negative sentiment posts were from hospitals located in states like Lagos and Oyo while the positive sentiment were really rare. In conclusion the study used the social media post of citizens to investigate the service quality of hospitals in Nigeria. The study found that citizens were generally dissatisfied with the quality of hospitals in Nigeria. The study identified factors like poor infrastructure, long waiting time, poor access to health care, staffing shortages and inadequate funding as factors which contributed to the negative sentiment towards hospital in Nigeria. The findings provided insights into the service quality of hospital in Nigeria which can be utilised by healthcare policymakers and hospitals for improvement of service quality and distribution of resources effectively.

# 1 Introduction

Healthcare delivery is an essential aspect of any healthcare system, and the quality of care provided is a crucial concern for policymakers, healthcare providers, and the general public. However, in Nigeria, the healthcare system faces numerous challenges. According to the World Health Organization (WHO), Nigeria has one of the lowest life expectancies in the world, at just 54 years for men and 56 years for women. Additionally, the country has a high burden of disease, with an estimated 19% of the population living with malaria, 2.9% living with HIV/AIDS, and 311 cases of tuberculosis per 100,000 people in 2019. Furthermore, access to healthcare services in Nigeria is limited, particularly in rural areas where the majority of the population lives. The doctor-to-patient ratio in Nigeria is one of the lowest in the world, with just 4 doctors per 10,000 people.

Despite these challenges, the quality of healthcare services provided in hospitals remains inadequate, with many citizens expressing dissatisfaction with the care they receive. In recent years, social media usage has continued to rise in Nigeria, with many citizens using platforms such as Twitter, Facebook, and Instagram to share their experiences and express their opinions. Geo-semantics analysis is a technique that combines geospatial and semantic analysis to extract insights from social media data. It involves analyzing social media content, such as tweet or post, and combining it with location data to derive meaningful insights about people's opinions, attitudes, and behaviours related to a particular topic or event.

In this research paper, we aim to assess the service quality of hospitals in Nigeria using geo-semantic analysis of citizens' social media posts. Specifically, we seek to extract and analyze citizens' sentiments towards healthcare services provided in Nigerian hospitals, their perceptions of the quality of care received, and the geographic distribution of these sentiments. By analyzing social media data in this way, we hope to gain a better understanding of the factors contributing to the issues facing healthcare services in Nigeria and to identify strategies to improve the quality of care provided to citizens.

The findings from this study may have important implications for policymakers, healthcare providers, and the general public. For policymakers, the insights gained from geo-semantic analysis can inform policies and interventions aimed at improving the quality of healthcare services in Nigeria. For healthcare providers, the findings may highlight areas where improvements are needed and may help to identify best practices for delivering high-quality care. For the general public, the study may provide insights into the quality of healthcare services in their local area and may help them to make more informed decisions about their healthcare choices.

In conclusion, the quality of healthcare services in Nigeria is a critical issue that affects the health and wellbeing of millions of citizens. By using geo-semantic analysis to analyze citizens' social media posts, we hope to gain a better understanding of the factors contributing to the challenges facing healthcare services in Nigeria and to identify strategies to improve the quality of care provided to citizens.

## 2 Methods

### Data collection

In order to accomplish our research goals, we initiated the process by identifying the prevalent social media platforms utilized in Nigeria for expressing opinions regarding healthcare services. Among these platforms, Twitter stood out as the most widely adopted medium for this purpose. Subsequently, we curated a comprehensive compilation of hospital names throughout Nigeria, which served as our main keywords for extracting Twitter data. To carry out this data extraction task, we employed Twint, a robust Python-based scraping tool, which allowed us to gather tweets without relying on Twitter's API. Each retrieved tweet encompassed a wealth of detailed information, including its unique ID, timestamp, textual content, and the author's handle.

### Data preprocessing

The data pre-processing phase involved utilizing Natural Language Processing techniques in Python programming language. The process consisted of several steps to ensure the quality and suitability of the data for analysis. In Step 1, we focused on removing invalid tweets that were not relevant for topic detection. Tweets that met specific criteria were classified as invalid for this study. These criteria included: (1) tweets containing fewer than three words, and (2) tweets mentioning a hospital but not referring to it directly (e.g., tweets discussing accidents or crimes unrelated to the hospital's activities). These tweets were deemed irrelevant and thus excluded from further analysis. In Step 2, we proceeded to clean the content of the remaining tweets and transform the data for analysis. Various elements that could interfere with topic detection efficiency were removed from the tweet contents. This included eliminating hyperlinks, hashtags, username mentions, emojis, numbers, and punctuation. The NLTK library was employed for tokenization, which involved transforming the text into a list of individual words. Additionally, lemmatization was applied to convert the words to their base or dictionary forms, using NLTK as well. Words that held little meaning or were considered stop words were then removed from the dataset. Following the data pre-processing steps, the dataset was refined, resulting in a collection of 250 tweets that remained suitable for further analysis.

### Data analysis

To automate the process of analyzing the text content of Twitter data, topic modelling was implemented. We utilized a specialized approach called Short Text Topic Modelling (STTM), which is specifically designed for handling short and sparse text data like tweets, chat messages, and product reviews. STTM algorithms employ various techniques like word embedding, clustering, and dimensionality reduction to extract latent topics and subtopics from the limited text data available. In our study, we specifically utilized the Gibbs Sampling Dirichlet Mixture Model (GSDMM) as our STTM model. This choice was motivated by the fact that most of the tweets in our dataset were short, with an average length of approximately 10-14 words after data pre-processing. GSDMM is well-suited for extremely short text with sparse word co-occurrence, as it simplifies the inference process by assuming that each document is associated with one topic. The algorithm hypothesizes a generative process for each document, where the words/terms in the document are drawn from a probability distribution, and a topic is represented by the probability distribution of the words/terms. Configuring the GSDMM model requires setting four parameters: the number of topics ( $K$ ),  $\alpha$ ,  $\beta$ , and the number of iterations.  $\alpha$  represents the probability that a document would be grouped in a cluster, while  $\beta$  is the probability that the words in a document

are similar to the words in other documents. In our study, we used the default values of  $\alpha = 0.1$  and  $\beta = 0.5$  to identify topics from the tweets in our dataset. By performing a fit procedure with different  $K$  values, we were able to identify the suitable maximum number of topics that best described our data. This ensured that the algorithm determined clusters that maintained a good balance between completeness and homogeneity. Once the short text topic modeling method detected distinctive groups of tweets, we employed manual annotation to relate each topic to a subject associated with hospital service quality.

To gain a deeper understanding of the attitudes and emotions expressed by individuals towards the identified topics, we employed sentiment analysis. Sentiment analysis is a computational technique used to determine the subjective sentiment or opinion expressed in textual data. In this study, we utilized Textblob, a widely used Python library for natural language processing, to conduct sentiment analysis. Textblob enabled us to analyze the sentiment polarity, i.e., whether a tweet expressed a positive, negative, or neutral sentiment, within our dataset. The model determines the sentiment polarity of the text and assigns a sentiment score ranging from -1.0 (negative) to 1.0 (positive) to quantify the expressed sentiment. Additionally, we sought to analyze the spatial distribution of the tweets, identify the regions where the hospitals were located, and gain insights into the geographical patterns of opinions expressed on Twitter regarding healthcare services in Nigeria. To achieve this, we utilized a process that involved utilizing the hospital name mentioned in each tweet. Our approach utilized OpenCage, a geocoding service that provides longitude and latitude coordinates based on location names. OpenCage is a powerful geocoding service that leverages various data sources, including open data and proprietary datasets, to convert location descriptions into precise geographic coordinates. It utilizes a combination of algorithms and databases to accurately match location names with corresponding latitude and longitude coordinates. In our study, we took advantage of OpenCage's capabilities by feeding the hospital names extracted from the tweets into the service. OpenCage's geocoding algorithms then processed the hospital names and returned the corresponding longitude and latitude coordinates for each hospital. By associating geographic coordinates with the hospital names, we were able to derive the geographical locations of the hospitals mentioned in the tweets that we were able to use for geo-graphic analysis

### 3 Results

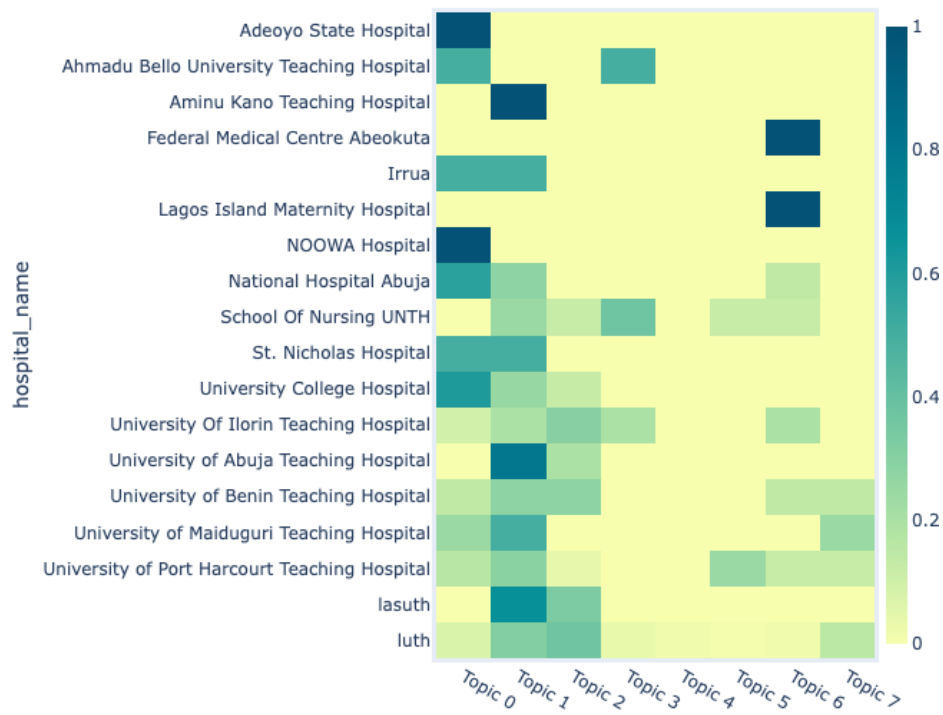
The primary objective of the first question in this study was to ascertain the perspectives of citizens regarding the health system in their local areas. By applying GSDMM, we were able to uncover latent topics and associated keywords from the social media data, providing a comprehensive understanding of the public's perspective on hospital service quality across various regions of Nigeria. The analysis of citizen's social media posts revealed a diverse range of topics related to hospital service quality. These topics included infrastructure and facilities, staff behavior and professionalism, waiting times and appointment management, accessibility and availability of healthcare services and overall patient experience. By identifying these topics, GSDMM enabled us to capture the multi-faceted aspects of service quality as perceived by citizens. The associated keywords for each topic shed light on the specific issues, concerns, and experiences expressed by citizens in their social media posts. For instance, keywords such as "equipment," "maintenance," and "facilities" under the infrastructure and facilities topic highlighted discussions regarding the state of hospital infrastructure and the availability of modern medical equipment.

	cluster	topic_name
0	0	Hospital facilities and services
1	1	Medical professionals and patient outcomes
2	2	Hospital bed space and patient capacity
3	3	General health concerns and issues
4	4	Miscellaneous comments and inquiries
5	5	Assistance and emergency services
6	8	Hospital staff and equipment
7	9	High mortality rates due to inadequate healthc...

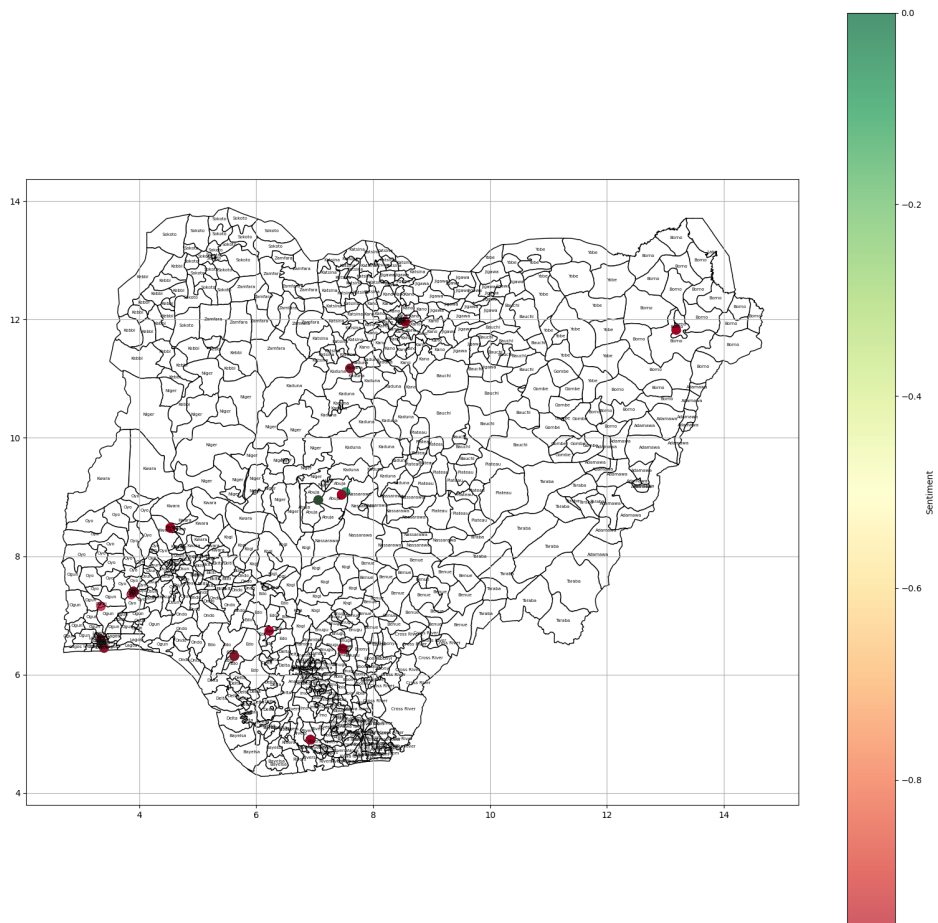
**The diagram shows the list of topic names we defined and the cluster**

The next aim of this study was to investigate the sentiments and geographical distribution on hospitals motioned in tweets. By employing geo-semantic analysis, we were able to explore regional variations in public perception regarding the quality of hospital services. Mapping these sentiments geographically allowed us to identify variations in levels of satisfaction and dissatisfaction across different regions in Nigeria. This analysis served to highlight areas where hospitals demonstrated commendable performance as well as areas requiring improvement. Remarkably, a significant majority of the obtained results indicated negative sentiments.

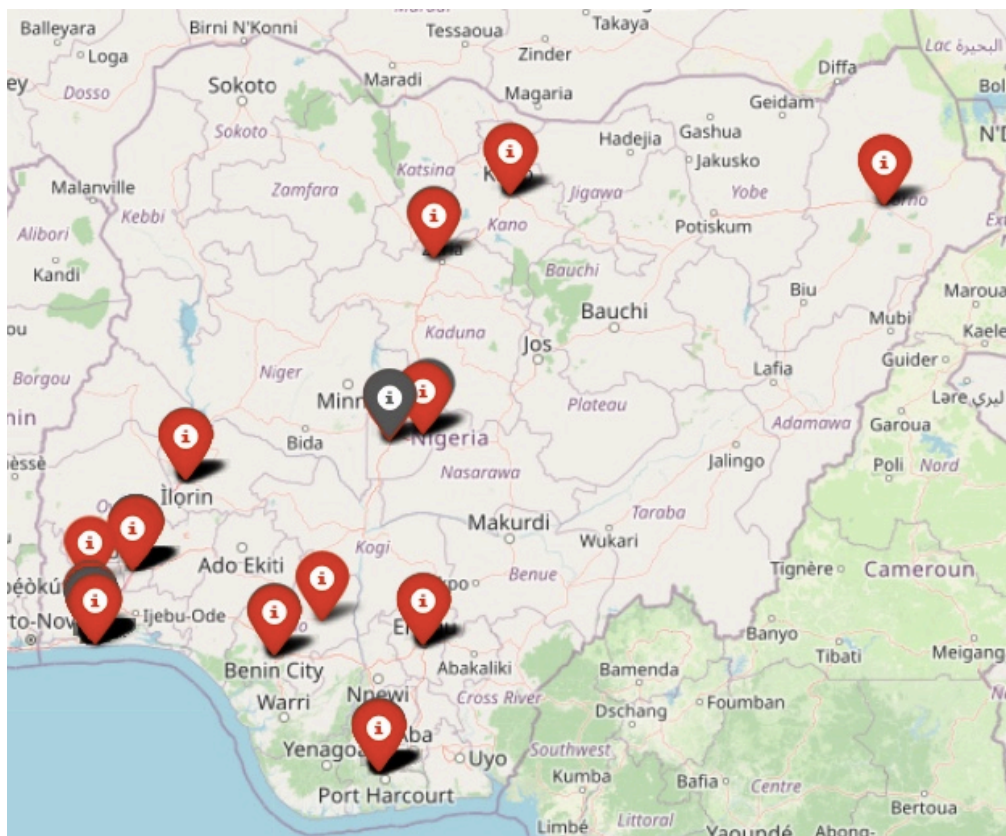
The findings of this study revealed a prominent trend, with approximately 80% of the analyzed tweets expressing negative sentiments. Such a substantial proportion of negative comments underscores the critical issues and challenges faced by citizens concerning the quality of healthcare services provided by hospitals in Nigeria. Leveraging geo-semantic analysis techniques, which amalgamate geolocation data with sentiment analysis, we were able to map the distribution of these negative sentiments across different regions of the country. The resulting map highlighted regions where citizens expressed significant dissatisfaction with hospital services, offering valuable information for policymakers and healthcare providers. The prevalence of negative sentiments in citizens' tweets emphasizes the urgent need for improvements in the quality of healthcare services in Nigeria. Addressing the concerns raised by citizens should be a priority for policymakers and stakeholders, enabling them to concentrate their efforts on enhancing healthcare delivery, ultimately leading to better patient experiences and outcomes. This research paper serves as a call to action, urging healthcare providers and policymakers to pay close attention to the concerns expressed by citizens on social media platforms. Proactively addressing these issues will contribute to achieving better service quality and overall improvements in the Nigerian healthcare system.



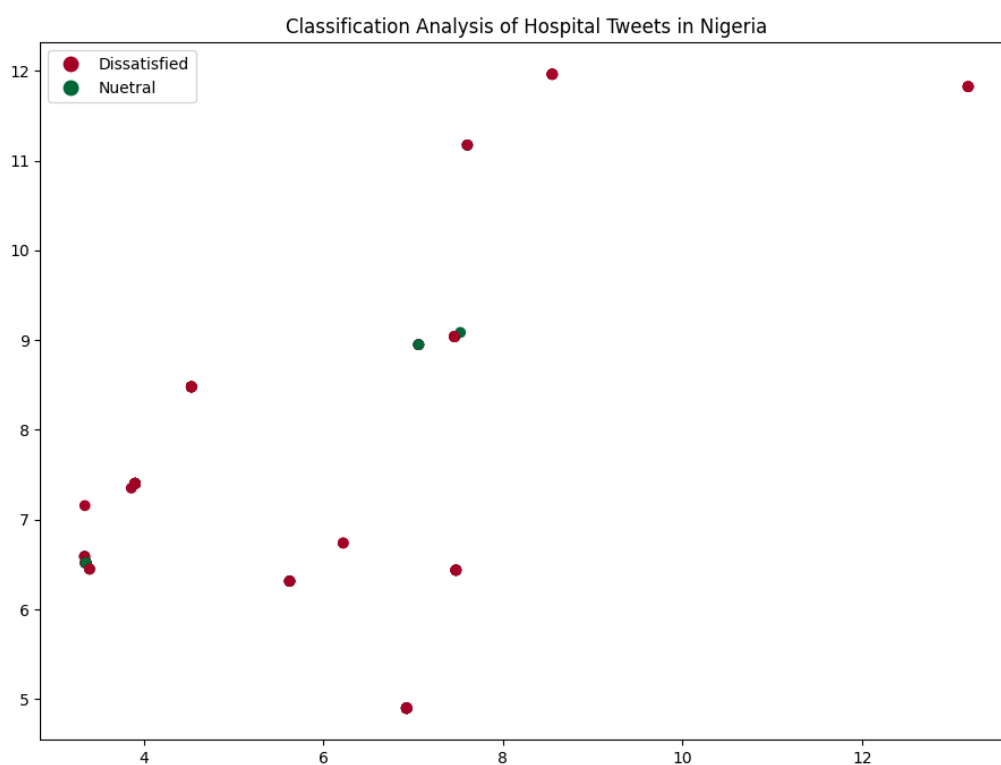
The diagram above is an interactive heat map showing the density of the tweets between the topics and the hospitals. As you can see, Lagos Island Maternity Hospital has a string density of topics related to Topic 6 which is (Hospital Staff and equipment)

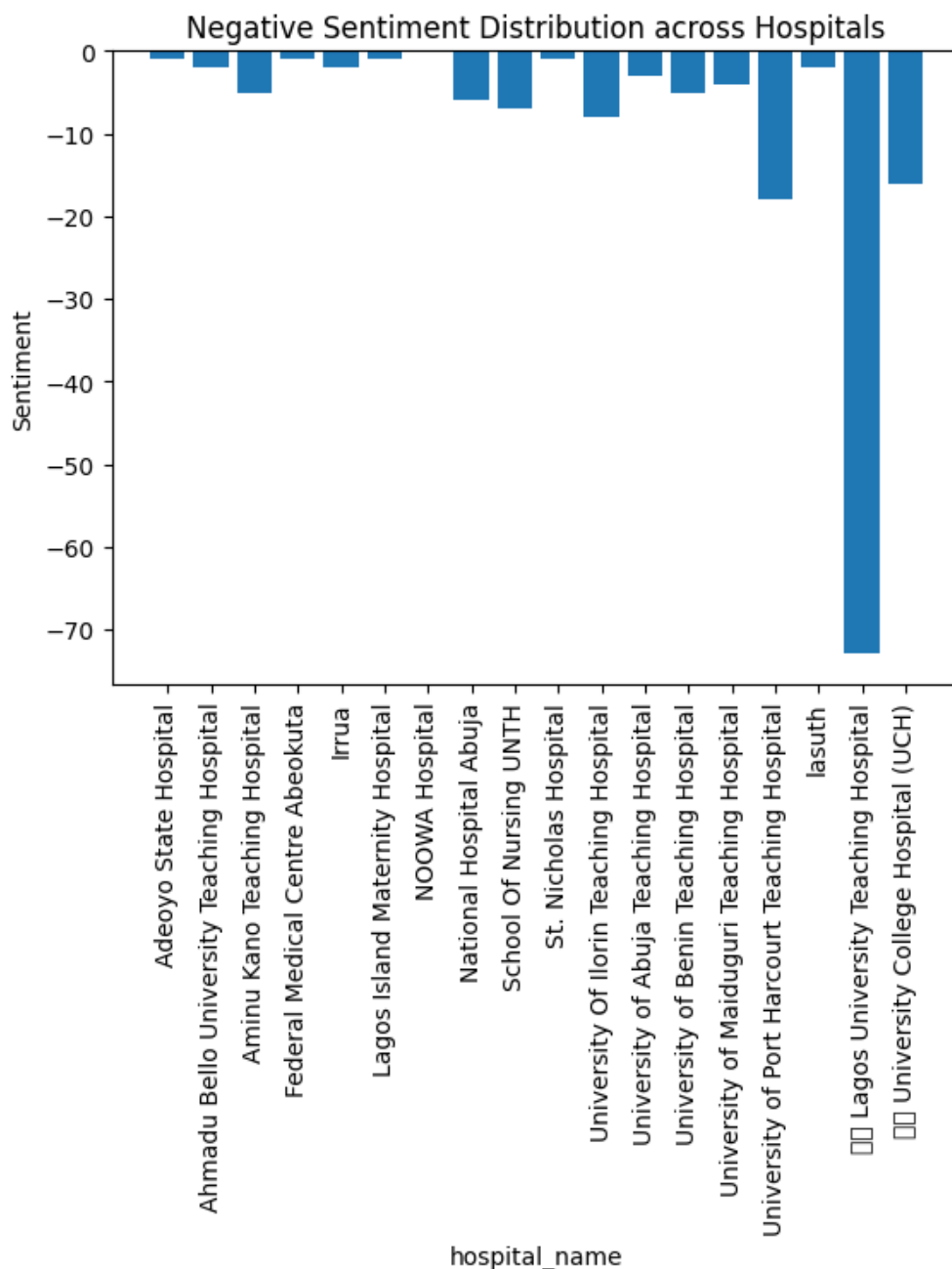


A geographic map displaying the distribution of citizens' tweets based on sentiment.



A map illustrating the geographical distribution of citizens' sentiment in their tweets.





#### 4. Concluding Remarks

The availability of social media platforms, like Twitter, offers a vast amount of data that can be leveraged for diverse applications, such as evaluating the quality of hospital services or predicting political election results. Nevertheless, it is crucial to recognize that the conclusions presented in this research paper are drawn from the examination of social media data, specifically a limited number of tweets, which may not fully represent the perspectives of the entire population. Nonetheless, the implementation of the GSDMM method and the thorough analysis conducted provide valuable insights for healthcare providers, policymakers, and stakeholders who are dedicated to improving the quality of hospital services in Nigeria.

The study's results indicate a pressing need for significant changes within Nigeria's healthcare system across the entire country. Additionally, I believe that every hospital should establish a social media presence as it serves as a means for customers and patients to stay informed about their activities and facilitates communication.



## References

- [1] Greaves F, Lavery AA, Cano DR, *et al* Tweets about hospital quality: a mixed methods study *BMJ Quality & Safety* 2014;23:838-846.
- [2] Hawkins JB, Brownstein JS, Tuli G, *et al* Measuring patient-perceived quality of care in US hospitals using Twitter *BMJ Quality & Safety* 2016;25:404-413.
- [3] Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. *J Med Internet Res*. 2013 Nov 1;15(11):e239. doi: 10.2196/jmir.2721. PMID: 24184993; PMCID: PMC3841376.
- [4] Nghiem, Son & Mehta, Pratik & Tao, Liang. (2013). Twitter for Public Health: An Open-source Data Solution.
- [5] Huang JH, Floyd MF, Tateosian LG, Aaron Hipp J. Exploring public values through Twitter data associated with urban parks pre- and post- COVID-19. *Landsc Urban Plan*. 2022 Nov;227:104517. doi: 10.1016/j.landurbplan.2022.104517. Epub 2022 Jul 26. PMID: 35966883; PMCID: PMC9358034.