

WhatsApp Network Group Chat Analysis Using Python Programming

Blessing Nwamaka Iduh

Academic Researcher, Computer Science Department, Nnamdi Azikiwe University, Awka, Nigeria

Abstract: - WhatsApp is an instant messaging application which enables users to send and receive messages in real time. It is a platform that has created an enabling environment for users to communicate with friends, groups and business partners at a cost of only a little internet access. This application has created opportunities for users to make free calls internationally in both audio and video modes. It has also provided the means for users to send broadcast messages to as many as 256 contacts at the same time. WhatsApp has helped individuals of like minds to create groups for diverse purposes. This paper, presents WhatsApp network group chat analysis, using Python programming language. The objective of this paper is to present an analysis of the WhatsApp group data to ascertain the level of involvement and participation by members in that group chat. Also, it involves the analysis of the most active date in the group, the number of messages sent on that date, the overall most active user, list of active admins in the group, total number of users, number of posts made by each individuals on the group, and the most used word on the platform. Also an analysis of the top 10 and top 20 users were done. The system was done with Python, and the Python libraries that were applied includes, Numpy, Pandas, Matplotlib and Seaborn. At the end of the work the expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group.

Keywords–WhatsApp, Social Media, Instant Messaging, Mobile phone, Group Chat

I. INTRODUCTION

WhatsApp is an instant messaging application that allows users to send text messages, chat and share media files like images, audio and video files. Users can also share documents and applications. With WhatsApp, users have the opportunity to communicate with several other users at the same time in a group. In addition, a user can send a broadcast message to up to two hundred and fifty six (256) users at a single message stance. This feature makes the message to appear as though it was sent to each individual alone.[1], described the application as a proprietary, instant messaging application for smart-phones, that cuts across all platforms. Apart from text messaging, users can also send images, video, and audio media messages to each other. The application software can be used on different Internet Operating Systems (iOS) such as Android, Apple and Windows iOS. According to [2] WhatsApp is an application that facilitates the exchange of instant messages, pictures, videos and voice calls through an Internet connection. It enables easy communication via text

or voice messages between two or more persons. Basically it helps people to stay connected. WhatsApp can be described as an attractive application. This is because, after installing the application, the sending and receiving of messages seems cost free (in contrast to the original text message function on mobile phones). The fact that WhatsApp can be termed as cost free, clearly explains the success of WhatsApp. Also, its function across different smartphone types like Apple, Android, etc, and its international functionality are also important contributors to this popularity.

According to [3], WhatsApp service had upto 450 million monthly average users as at 2014. Also, a data analysis by [2], showed that the use of WhatsApp accounted for 19.83% of all smartphone in 2015 as compared to Facebook which takes only 9.38%. Also, female folks were reported to be using WhatsApp for significantly longer period of time than the male. Furthermore, they stated that younger people use WhatsApp for a longer duration of time. WhatsApp can be described as free, but since it makes use of data, one can be said to be paying for it since you cannot make use of it without data. It works on iPhone, Android, BlackBerry, Symbian, and Windows devices. The information needed by a user is ones username name and phone number. Messages on WhatsApp are said to be encrypted. The application also uploads all of the users' contacts and require them to individually block users with whom they do not want contact.

To use WhatsApp according to [4]one will need a compatible smartphone or tablet an internet connection, and a phone number. The application uses a user's phone number as its username, and that account is pinned to that phone number, although you can transfer your contacts over to new devices. Users of WhatsApp can share their location in real time over messages. They can also organize lists of contacts so that they can quickly send messages to lots of people in group chats through WhatsApp. One of the very significant features of WhatsApp is the fact that it allows users to make international calls, without incurring the international charges associated with text messages. It ensures that members in a group can follow conversations more easily and respond to specific questions in a group chat or write comments on content in the group. Also on recent feature of the WhatsApp is the ability to reply directly in a private chat, to the sender of a message sent to a group. And in this case, the message can be linked back to the group, from the private page of the sender.

According to [5] “WhatsApp Messenger is a freeware, cross-platform messaging and Voice over IP (VoIP) service owned by Facebook”. WhatsApp has gained lights' speed in terms of a social networking application. It is recorded by [4] that more than 1 billion people in over 180 countries use WhatsApp. The number of downloads for WhatsApp on Android, exceeds one Billion. WhatsApp uses 3G, 4G or Wi-Fi network, to message with friends and family. It is an application that requires you to have at least a memory space of 120 MB for effective installation, this is due to its media files. It is important to note, that there is no restriction on the length and number of messages one can exchange and no carrier IM fees apply. One does not need to install a Sim-card to use WhatsApp; the only requirements are a supported phone, internet connection and storage space on the phone to download the application. Also, the phone number a user uses for WhatsApp, must not necessarily be the number of the SIM card. WhatsApp uses a customized version of the open standard eXtensible Messaging and Presence Protocol (XMPP) to exchange data over the internet. Messages can be in the form of plain text, multimedia message like photos, audio, video, location, address book contact cards and icons.

This seminar paper presents an analysis of the chats and users in certain WhatsApp groups, to ascertain the level of participation of members in the group chat. The objective of this work is to ascertain the level of involvement and participation recorded in a given WhatsApp group. The analysis involves the following; the most active date in the group, the number of messages sent on the most active date, the overall most active user, list of active admins in the group, total number of users, number of posts made by each individuals on the group, and the most used word on the platform. Also an analysis of the top 10 and top 20 users were also done.

II. LITERATURE REVIEW

A survey of WhatsApp group analysis was done by [6], his work was focused on predicting the level of addiction of an individual to the WhatsApp group with respect to the age group and gender. He made use of R statistics software programme. The Research provided the basic idea of Statistical analysis. The work was done on a particular WhatsApp group data, to determine the type of communication medium people prefer the most in WhatsApp group chats, check most active day of week, to find which age group participants are more active. Also to ascertain if the male users are more addicted to the WhatsApp group than females.

Also, [7], in their work, presented a study that explores the classroom communication between teaching staff, faculty staff and high school students using WhatsApp. Their work focused mainly on getting a better understanding of the functions of the WhatsApp groups in the relationship between students and teachers, the activities they carry out with the

application and the way it generally affects educational and academic output.

In the research by [8], an analysis of WhatsApp as a communication medium among emergency surgery teams in a London hospital was conducted. According to their findings, emergency surgery team members that participated in their study, used WhatsApp for a period of 19 weeks. The Initiator and receiver of messages and the communications made were compared for response times and communication types. Safety events were reported. Their research showed that WhatsApp has the ability to consume students study time. WhatsApp takes much of students study time and it can leads to lack of concentration during lectures.

Furthermore, in the review work done by [9], Using evidence from the published literature and case reports indexed in PubMed and other sources, they presented an overview of the various applications of Instagram and WhatsApp in health and healthcare. They also described the main issues surrounding the uses of the WhatsApp and Instagram Applications in health and medicine.

III. ANALYSIS AND METHODS

WhatsApp uses a protocol known as the XMPP (Extensible Messaging and Presence Protocol). This protocol is responsible for handling the message delivery task. According to [10], XMPP is an open eXtensible Markup Language (XML). XML, was designed to store and transport data. XMPP is an XML technology for real-time communication, which powers a wide range of applications including instant messaging, presence and collaboration. The acronym XMPP can be explained as follows: **Protocol** which is a set of standards that allows systems to talk to each other. The P which stands for Presence, tells the servers that you are online, offline or busy. M — Messaging. The ‘messaging’ part of XMPP is ‘what’ you see; which is the Instant Message (IM) sent between clients. XMPP has been designed to send all messages in real-time using a very efficient push mechanism. And then, the X which stands for eXtensible, explains the fact that it is an open standard and using an open systems approach of development and application, XMPP is designed to be extensible. In other words, it has been designed to grow and accommodate changes.

Also, WhatsApp was built using Erlang. Erlang is a programming language used to build massively scalable, soft real-time systems with requirements on high availability. Erlang's runtime system has built-in support for concurrency, distribution and fault tolerance[11]. Erlang capabilities is what enables bug fixes and frequent updates in WhatsApp. Furthermore, WhatsApp also uses a technology known as Mnesia. Mnesia is a multi-user distributed database management system which enables quick response to requests. Mnesia works with the Erlang Runtime System. The Mnesia relational and object hybrid data model is what makes it suitable for developing distributed applications of any

scale[12]. In addition, WhatsApp uses Bootstrap front-end framework.

The data analysis stage involves the Process of cleaning, transforming, inspecting and modelling the data that was used for this system. The goal of this work was to gather useful information about effective and functional users are in a WhatsApp group. Data analysis is a process for acquiring raw data and transforming it into information useful for decision-making by users. In this work, the sample data that was analyzed was gotten from the Unizik Staff Community WhatsApp group chat. The groups were 3 in number as at the time of this research, with 256 members in each group. The objective of this work was to get the most active users in any

WhatsApp group, to discover the most active days of week, to find the top 10 or top 20 most active users or more, as the case maybe, to also the activities of each users and the number of messages sent by the users of the group and also the analyses the word count of users on the platform.

3.1 System Architecture

The system architecture is represented in figure 1. It shows the general architecture of the system. This include, data collection stage, the data input state, data transformation, data exploration and data visualization. The data collection stage is shown in figure 2 while the data input, data transformation, data exploration and data visualization are handled by Python and its libraries.

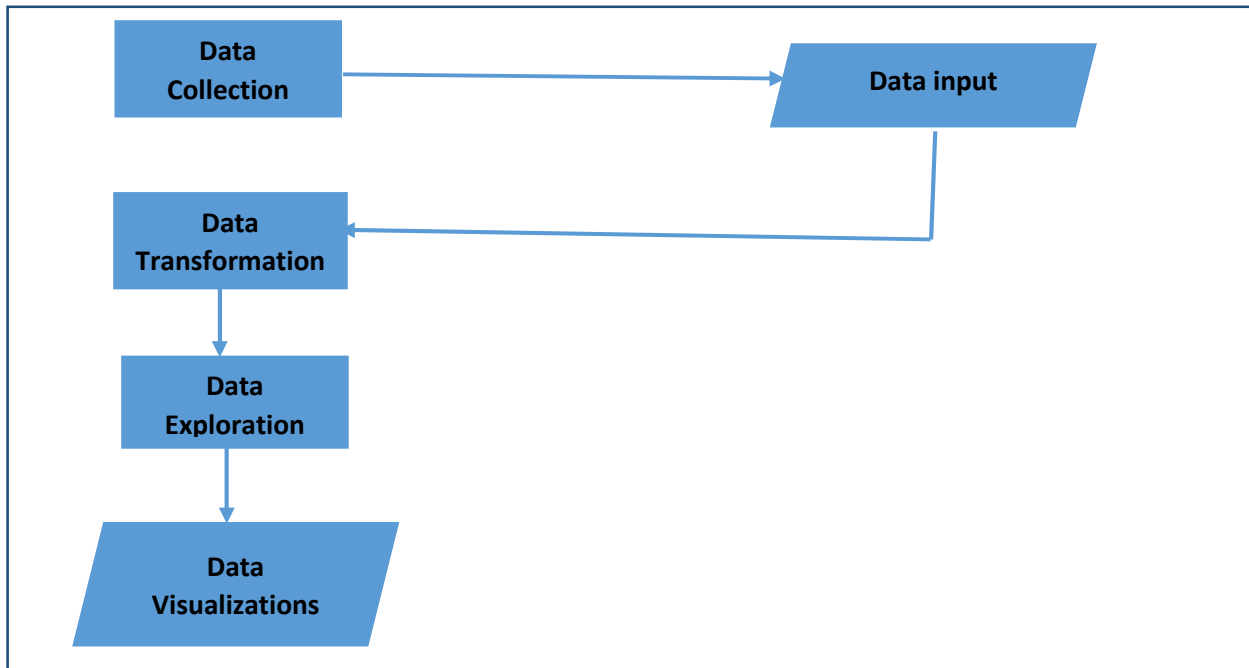


Figure 1 – architectural diagram of the system

3.2. Data Collection stage.

This stage involves the point where the WhatsApp data is collected. This was done by visiting the chat group to be analyzed, to export the WhatsApp data file that was used. The procedures involved, visiting the WhatsApp group page, clicking on the settings, select export data and then select either add media or without media. This simply means to know whether you intend to export the file with the media or not. Note that exporting with the media, will lead to use of larger volume of data and waste of time for data collection. Figure 2 shows a brief illustration of the steps involved in the data collection.

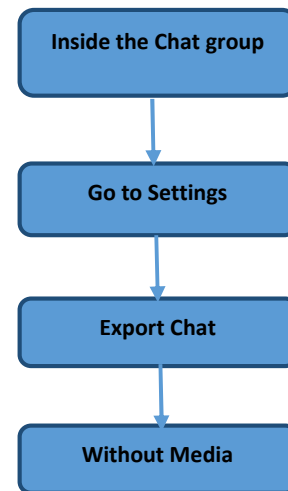


Figure 2. Steps involved in data Collection.

3.3. Implementation Tools

a) Python Programming Language - Python is the programming language used for this work. It is a free open source programming language. It is a High level programming language. It supports object oriented and structured programming fully. Python is Compatible with Major Platforms and Systems. It supports many operating systems. Also, it has a very Robust Standard Library. The data input, data transformation, data exploration and data visualization are handled by Python and its libraries.

b) Pandas for data extraction and preparation - Pandas is a Python library that provides high-level data structures which are simple to use as well as intuitive. It was the tool that enabled the extraction of the data to be analysed. It was used to fetch the dataset in Python from the CSV, Excel, JSON files and manipulated the data to perform operations on it.

c) Numpy – Numerical Python (numpy) was the Python library used to handle the multidimensional arrays and functions that were needed for the classification of the chats into days, hours, minutes and seconds.

d) Matplotlib - Matplotlib was used for the data visualization of this system. It is a standard Python library used for creating 2D plots and graphs. It was imported in this work to create a graph of the dataset.

e) Seaborn - Seaborn data visualization library was also imported in this work. It builds on Matplotlib’s foundations. Being a higher-level library, it was able to expand the plot and better beautify it. It doesn’t work alone hence it works on Matplotlib foundation.

IV. RESULTS AND DISCUSSION

The results of this work, showed several activities on specific dates as specified by the system at a given time. The results showed that the most active date was 15th June, 2018. The number of message sent on that most active date was 190. Also, the overall, most active user was recorded and the user

was shown to have posted over 972 messages to the group. The system also recorded the list of active admins. Furthermore, the total number of users on that group was shown to be 230. In addition, a full List of all users on the platform with their name or their phone number were also outputted plus the number of times each individual on the platform has made a post. And the Most used word was also given as “the” and it was used 43313 times. Figure 3, shows a snap shot of the screen that shows the output of the analysis.

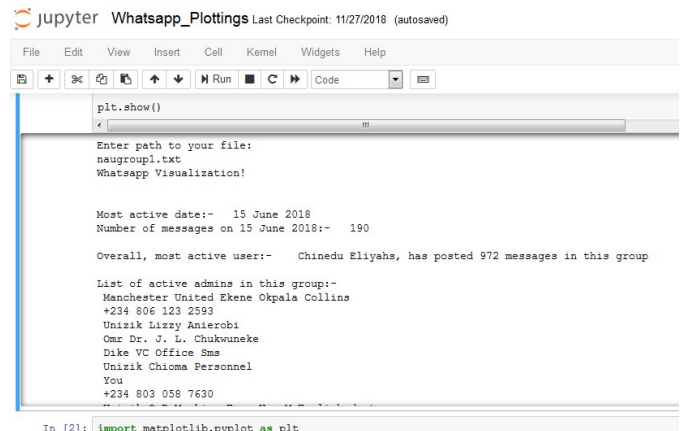


Figure 3- Sample Output of the WhatsApp Plot

Furthermore, an analysis of the top 10 active users and the top 20 active users were also represented. The results are visually represented in figures 4 and figure 5, figure 4 shows the top 20 users of the analyzed WhatsApp group and how active they are in terms of their percentage of participation. They show the name of some users while showing only the phone number of others. This means that the names that appear on the graph plot are among the saved contacts of the developer, while the phones numbers that are displayed indicates those users in the group whose contacts are not stored in the developer’s phone. The following represents the output of the result of the analysis done with Python on the given group chat.

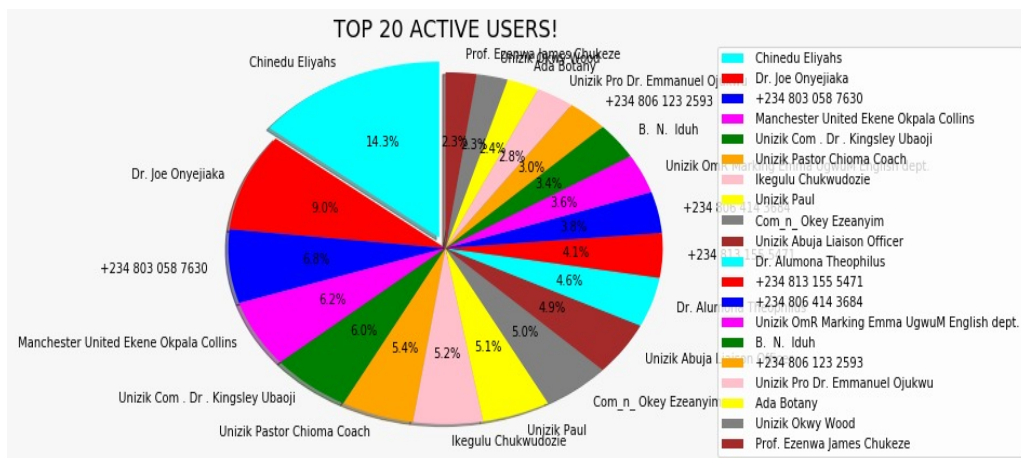


Figure 4. Top 20 users

Figure 5 on the other hand shows the top 10 users alone. This gives a narrower scope of comparison. As explained for figure 4, figure 5 also has some phone numbers and some with

names. This figure clearly gives an analysis of the top 10 users on the group chat, showing them in percentages.

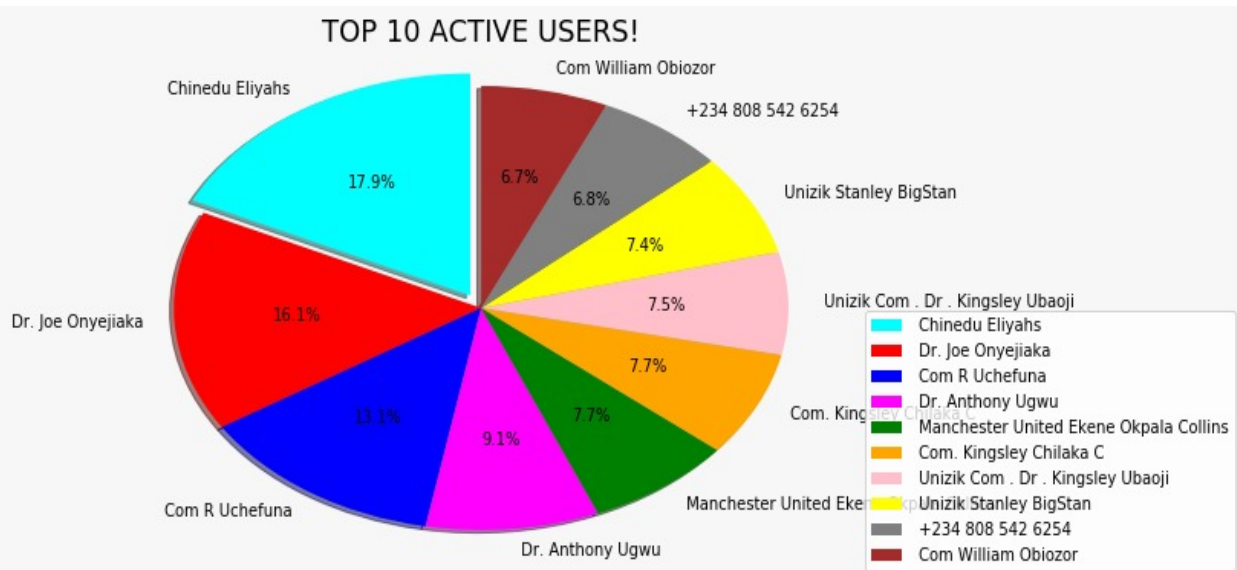


Figure 5. Top 10 users

V. CONCLUSION

In conclusion, it can be said that the capabilities of the WhatsApp application and the power of the Python programming language in implementing whatever network data analysis intended, cannot be overemphasized. This work was able to discuss the WhatsApp application and its capabilities to a great extent, and it was able to use the Python programming language and its libraries, to create an analysis of a WhatsApp group chat and visually represent the top 10 and top 20 users in the chat group. A pseudocode of the plot was given and at the end, a visual representation of the plot was implemented. Also an analysis of the top 10 and top 20 users were done. The system was done with Python, and the Python libraries that were implemented includes, Numpy, Pandas, Matplotlib and Seaborn. At the end of the work the expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group. One thing of note is that this system has the ability to analyze any WhatsApp data input into it.

REFERENCES

- [1] N. Thakur, "Forensic Analysis of WhatsApp on Android Smartphones," University of New Orleans Theses and Dissertations., pp. 1-33, 2013.
- [2] C. Montag, K. Błaszkiwicz, R. Sariyska, B. Lachmann, I. Andone, B. Trendafilov, M. Eibes and A. Markowetz, "Smartphone usage in the 21st century: who is active on WhatsApp?," 4 August 2015. [Online]. Available: <https://bmcsresnotes.biomedcentral.com/articles/10.1186/s13104-015-1280-z>. [Accessed 12 March 2019].
- [3] C. Steele, "what-is-whatsapp-an-explainer," 20 February 2014. [Online]. Available: <https://www.pcmag.com/news/320871/what-is-whatsapp-an-explainer>.
- [4] WhatsApp, "About WhatsApp," 20 April 2019. [Online]. Available: <https://www.whatsapp.com/about/>.
- [5] W. Bani, "Whatsapp," 22 April 2019. [Online]. Available: <https://en.welibani.org/weli/WhatsApp>. [Accessed 24 April 2019].
- [6] S. Patil, "WhatsApp Group Data Analysis with R," International Journal of Computer Applications, vol. Volume 154 , no. 4, p. 0975 – 8887, November 2016.
- [7] D. Bouhnik and M. Deshen, "WhatsApp Goes to School: Mobile Instant Messaging Between Teachers and Students," Journal of Information Technology Education: Research, vol. 13, pp. 217-231, 25 August 2014.
- [8] J. Yeboah and G. D. Ewur, "The Impact of Whatsapp Messenger Usage on Students," Journal of Education and Practice, vol. Vol 5, no. 6, pp. 157-164, 2014.
- [9] M. N. K. Boulos, D. M. Giustini and S. Wheeler, "Instagram and WhatsApp in Health and Healthcare: An Overview," Future Internet Creative common attribution MDPI, vol. 8, no. 37, pp. 01-14, 2016.
- [10] XMPP, "https://xmpp.org/about/," 19 March 2019. [Online]. Available: <https://xmpp.org/about/>.
- [11] Erlang, "https://www.erlang.org/," 21 April 2019. [Online]. Available: <https://www.erlang.org/>.
- [12] Mnesia, "About Mnesia," 12 May 2019. [Online]. Available: <https://elixirschool.com/en/lessons/specifics/mnesia/>.