[[1]](#footnote-1)

WhatsApp Data Analytics

12010504-Ganesh Karode, 12011084-Omkar Karpe, 12010914-Kartik Rupauliha, 12010397-Kartik Rajput, 12011421-Nakul Kasar

Department of Engineering, Sciences and Humanities (DESH)

*Abstract* — *The most used and efficient method of communication in recent times and one of the most popular social media applications is WhatsApp. WhatsApp chats consists of various kinds of conversations held among group of people. This chat consists of various topics. This information can provide lots of data for data analytics. Data analytics using python aims to provide in depth analysis of this data which is provided by WhatsApp. Irrespective of whichever topic the conversation is based our developed code can be applied to obtain a better understanding of the data. The advantage of this tool is that is implemented using simple python modules such as pandas, matplotlib, regex and tkinter which are used to create data frames, plot different graphs, and make GUI then it is displayed in the GUI application which is efficient and user friendly*

*.*

***Keywords*** *— Python, Data Analytics, WhatsApp, Pandas, Matplotlib, Tkinter GUI*

# INTRODUCTION

T

oday one of the most trendy social media platforms is WhatsApp. It is one of the favorite social media platforms among all of us because of its attractive features. 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 others at a cost of only a little internet access. In the world of social media, the WhatsApp conversation groups are one of the most popular ways to stay in contact with family and friends. WhatsApp claims that nearly 55 billion messages are sent each day. The average user spends 195 minutes per week on WhatsApp, and is a member of plenty of groups. These groups contain a lot of information and data that can be analysed and visualized in forms of plots and graphs.

WhatsApp Data Analyzer means we are analysing our WhatsApp group activities. It tracks our conversation and analyses how much time we are spending on WhatsApp. We can also get the group activity tracked as we can know no. of messages sent, sleep cycle of group, most active person, hours or weekdays people are most active etc. Our aim is developing an GUI based WhatsApp data Analyzer which will have a simple interface and would be very helpful for people to analyse and visualize their WhatsApp activities.

We did refer a few existing technologies for WhatsApp data analytics first of them is AnalayzeTheChat [1] from towardsdatasience.com is an online platform for analyzing WhatsApp group chats where we need to drag and drop our extracted WhatsApp group chat txt file and we get statistics for our chats, limitation to this platform is that someone might not want to share their chats with website and also these chats can be used misused. AnalyticsVidhya.com [2] also provides code in python for WhatsApp data analytics which we need to download and run on any python ide and we can see the visualizations in terminal and plots section, limitations to this code are that it is difficult for people without programming knowledge to run this and also code is too large.

# Implementation details

We have used Python 3.9 which is an interpreted, high-level general-purpose programming language. Created by Guido Van Rossum and first released in 1991. Its language constructs and objects-oriented approach aim to help programmer with clear, logical code for small and large-scale tools. Python is used for web development (server-side), software development, mathematics, it can be used alongside software to create workflows, it can connect to database systems, it can also read and modify files, it can be used to handle big data and perform complex mathematics and can be used for rapid prototyping, or for production-ready software development.

WhatsApp data Analyzer is a statistical analysis tool for WhatsApp chats. Working on the chat files that can be exported from WhatsApp it generates various plots. Data pre-processing, the initial part of the project is to understand implementation and usage of various python built in modules. These various modules provide better code representation and user understandability. The following libraries are used such as pandas, matplotlib, re, tkinter etc. pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. Re or regular expression is a sequence of characters that specifies a search pattern. Usually, such patterns are used by string-searching algorithms for "find" or "find and replace" operations on strings, or for input validation. It is a technique developed in theoretical computer science and formal language theory. Tkinter is a Python binding to the Tk GUI toolkit. It is the standard Python interface to the Tk GUI toolkit, and is Python's de facto standard GUI. Tkinter is included with standard Linux, Microsoft Windows and Mac OS X installs of Python. The name Tkinter comes from Tk interface.

The proposed system is developed by using Spyder software. Spyder is a free and open-source scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep

inspection, and beautiful visualization capabilities of a scientific package. Idea is to implement a data visualisation code using python to make better sense of WhatsApp group chat data.

### Algorithm:-

### We start by selecting the WhatsApp chat data text file(.txt)

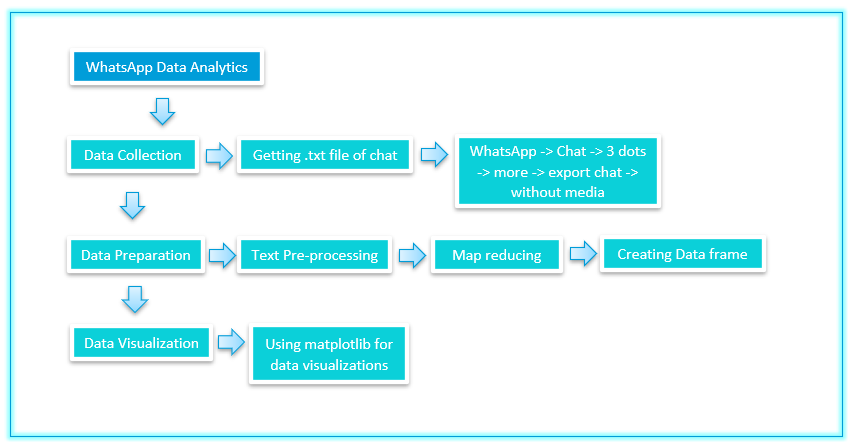
### Further the data from this file is processed and split into required format using regex library

### Data-frame is created from this sorted output using pandas

### Many columns such as message, user, weekday, hour, word count, media messages, group notification are created

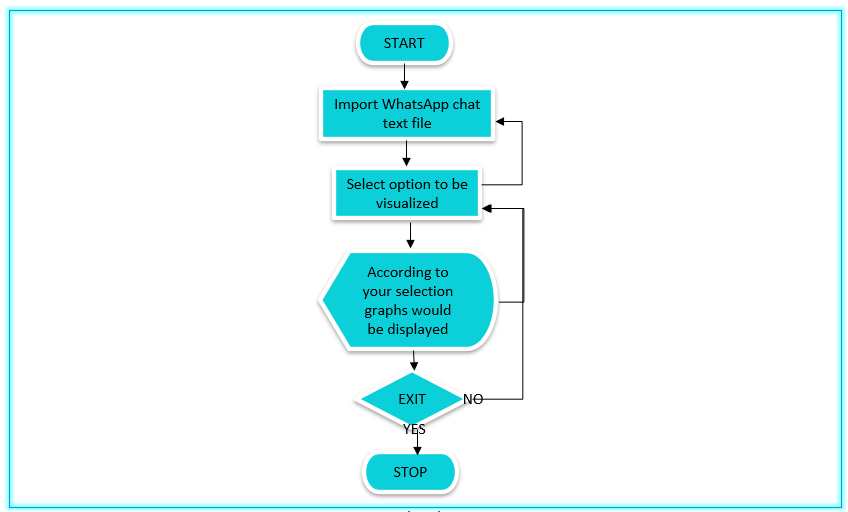
### Data is visualized from this data-frame using matplotlib, pyplot

### We have used tkinter to make a user friendly GUI to use this code easily



1. Block diagram

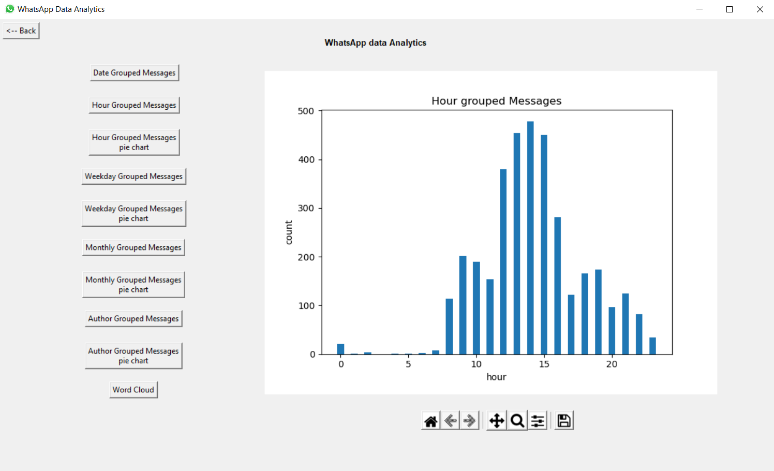
Fig. 1 describes the block diagram of our project, some of the features of our project are it has a very simple, easy to use GUI by using it we would be able to analyze / visualize no. of messages sent by user, letter count, word count, URL count, average no. of messages per day, sleep cycle of group, media messages and emojis, no. of message sent on particular weekday or hour, word cloud of most repeated words.



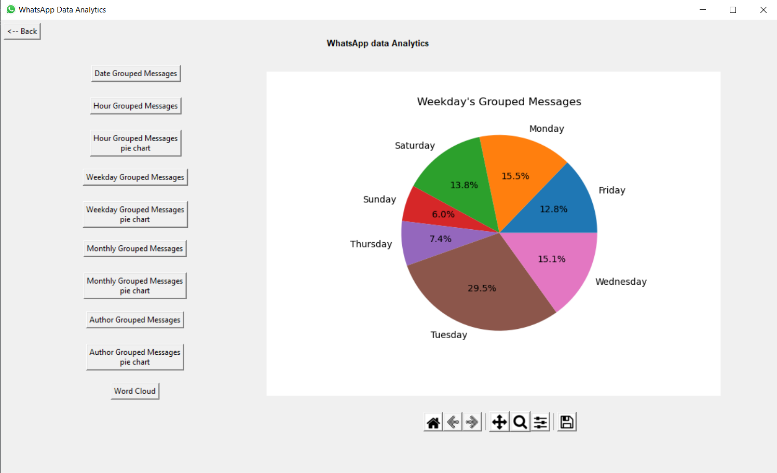
1. Flowchart

# Results and Discussions

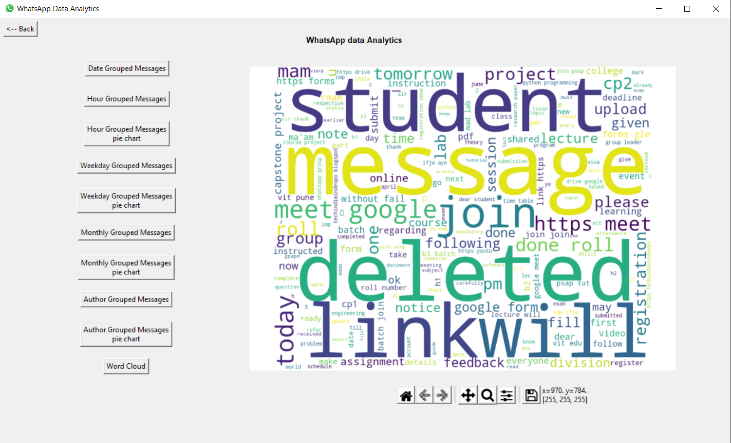
The following represents the output of the result of the analysis done with Python on the given group chat.



1. Output from our code which shows no. of messages per hour



1. Output from our code which shows no. of messages on weekday’s



1. output from our code which shows wordcloud of most repeated words

# Conclusion and Future Scope

This work was able to discuss the WhatsApp application and its libraries, to create an analysis of a WhatsApp group chat and visually represent the data. The system was done with python, and the python libraries that were implemented includes, NumPy, Pandas, Matplotlib and Tkinter. This system can further be upgraded to analyse data from different social media platforms such as twitter, Facebook etc.

# References

1. http://aqeelanwar.pythonanywhere.com/
2. https://www.analyticsvidhya.com/blog/2021/04/whatsapp-group-chat-analyzer-using-python/
3. https://medium.com/mcd-unison/whatsapp-group-chat-analysis-with-python-3f5196280ba
4. GeeksForGeeks.com
5. w3schools.com

1. [↑](#footnote-ref-1)