

**Review Paper**

**Whatsapp chat analysis**

**By Rithak**

**Section J**

**Class Roll No- 44**

**B Tech CSE**

**Whatsapp chat analysis**

**Model: Natural Processing Language**

1. **Abstract-**

The most used and efficient way of communication these days is an application called WhatsApp**.** WhatsApp chats consist of differenttypes of conversations between groups of people. This chat consists of various topics. This information can provide **a lot** of data for modern techniques such as machine learning. The most important thing for machine learning models is to provide the right learning experience**.** This is indirectly affected by the data **you** provide to the model. This tool 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, seaborn and sentiment analysis which are used to create data frames and plot different graphs, where then it is displayed in the flutter application which is efficient and less resources consuming algorithm, therefor it can be easily applied to largest dataset.  
Keywords: WhatsApp chat data, Pandas, Seaborn, matplotlib, sentiment analyzer, Flutter application etc

1. **Introduction**

WhatsApp was launched in 2009. It is a free application that can be used on both smartphones and computers. App users can send texts, real-time location, images, voice recordings, documents, and videos (Dodds, 2019; Mefolere, 2016; Boulos et al., 2016). In 2019, WhatsApp was used by 1.5 billion users in 180 countries (Digital Information World 2019). Research on the use of WhatsApp for various purposes (e.g. personal use, medical services, journalism) has increased (Kumar and Sharma, 2017; Awada, 2016; Mefolere, 2016; S´ anchezMoya and CruzMoya, 2015). Curiosity is one of the main things that motivates many people to do great things.The same curiosity kills an insider. emotions will also be based solely on past conversations.  
  
This tool is based on data analysis and processing. The first step in implementing a machine learning algorithm is to understand the good learning experience from which the model begins to improve. Data preprocessing plays an important role in machine learning. To make the model more efficient, we needed a lot of data, so we mainly focused on one of the large-scale data producers owned by Facebook, which is not WhatsApp. WhatsApp claims that nearly 55 billion messages are sent every day. The average user spends 195 minutes per week on WhatsApp and is a member of many groups. With a treasure trove of data right in front of us, we are forced to embark on a mission to better understand the messages our phones are forced to bear.  
  
Besides the normal analysis, we also perform sentiment analysis on group chat data. Sentiment analysis is a method of data mining and natural language processing or sub-process basically used to determine the opinions or feelings of users. In addition, the concept works on the fundamentals of machine learning where a set of data (data set in NLP terms) is used to train and process that data, and a The model was created and used to assess the emotions of the test cat. This process has become a very powerful field in text analysis.

1. **Literature Survey-**

A survey analysis on the usage and impact of WhatsApp Messenger [1] has been conducted and various studies and analysis have been found. These studies include the impact of WhatsApp on the students(youth). In the survey it was found that in the southern part of India, ages 18 to 23 spend around 8 hours using whatsapp and sometimes be online almost 12-16 hours a day. Most of them agreed to be using whatsapp tan any other site. They exchange images, audios and videos. This survey also proved that the whatsapp has been the most widely used app on the smart phones than any other app. This survey was conducted to know the positive and negative impacts of using whatsapp. As we can know that from this survey, whatsapp is most used app by the youth and other generations so, our project can give them the insights of their chats and provide them unknown facts

**3.1 Chat visualizer**

Chat Visualization is a statistical analysis tool for WhatsApp chats. Working with chat files that can be exported from WhatsApp will generate different plots showing, for example, which other participants the user interacted with the most. In order to better understand WhatsApp chats on the phone, we propose to use record manipulation techniques.  
  
The proposed system will be developed using Jupyter software. Jupyter is a nonprofit organization founded to develop open source software, open standards, and services for interactive computing in dozens of programming languages. The idea is to use Python to implement data processing code to better understand WhatsApp group chat data.

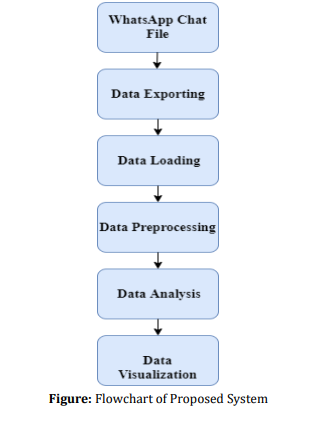
**Method**

Data Analysis-

It is the process of cleaning, transforming, inspecting, and modeling data with the goal of discovering useful information and ultimately drawing some conclusions. Analysis means breaking down the entire component into  
separate components for individual inspection. Data analysis is the process of gathering raw data and transforming it into information that helps users make decisions. This project provides basic  
statistical analysis of WhatsApp chats. The analysis performed was:  
• Find all messages, words, media and links shared in WhatsApp chats.  
• Find the most active people in your group.  
• Find the most used emoji in a group.  
• Find the busiest and least busy days of the month.  
• Find the most common and most used words within a group.  
• Find out the frequency of chats by day and by month.

Working flow

The working of given system is using following flowchart



Implementation

(a) Install and import dependencies: This step will install and import Streamlit, matplotlib, pandas, collections, seaborn, emoji,  
Wordcloud, URLextract and re.  
(b) Preprocessing: In this step the data is preprocessed. Here the data is formatted and delimited in the form of date, time, username and message used.  
(c) Exporting and uploading chat documents from WhatsApp: Here documents are exported from WhatsApp.  
Steps to export chat -> Open individual or group chat -> Options - Miscellaneous - Export chat -> Export without media  
Select -> Document is set. Upload your chat file and click Analyze  
(d) Train chat model and analyze data: Here the collected data is read and processed to train a  
machine learning classification model. increase. Then the model is evaluated and serialized. Analysis complete:  
1. Top stats: These include total messages, total words, media shares, and link shares.  
2nd Monthly Timeline: Frequency of chats for each month.  
3. Daily Timeline: Number of chats per day.  
4. Activity Cards: Shows the busiest and least busy days as well as months Weekly activity card.  
6. Word Cloud: The most common and most used words.  
7. Busy Users: Mostly active users.  
8. Emoji Analysis: The most popular and most used emojis.  
(e) Use the model to perform recognition: When the code runs, it uses the model trained above to make predictions about the user's gesture

**3.2 Sentiment Analysis**

Sentiment analysis is when users consider all kinds of emotional expressions on social media to be valid. Positive and negative sentiments are irrelevant to the overall statement when given opposite sign weights as described in the proposed work section. A relatively high scaling of the appropriateness of positive emotional expressions indicates a high relevance of positive norms, which corresponds to positive expectations of authenticity in social media. But by all means, these results should be considered legitimate and specific to online features. The way individuals adhere to rules of collaboration stems in large part from a natural need to avoid the threat of social approval and rejection. A tool for performing text analysis tasks on WhatsApp data.

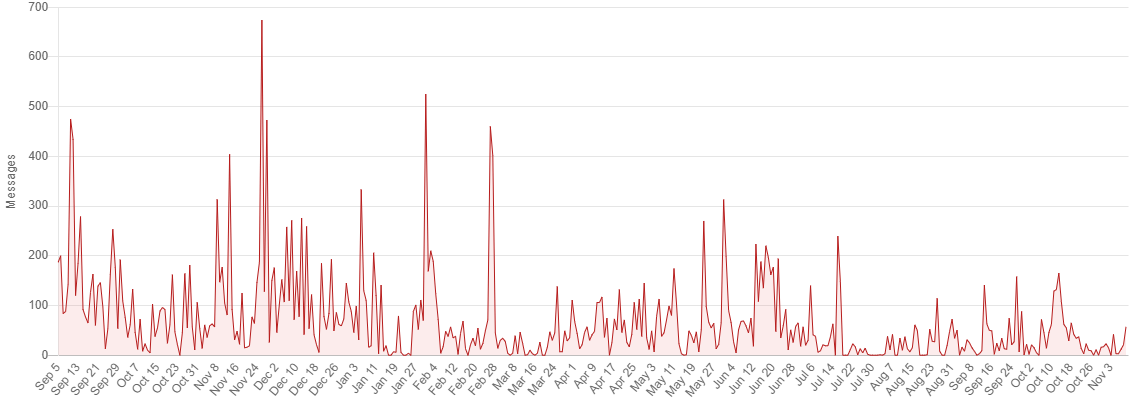
**Method**

1. Best Emotion Selection: In this process, the maximum number of repeated emotes is taken as the total emotes, no other factors are considered. Here, the total score for each emotion is calculated for visualization purposes, so we get the highest  
     
   b) Priority-based model: In this model, all emotions are given a priority number and each A total score is calculated based on the frequency of the emotion and its priority number, and is calculated with its total emotion.  
     
   Model 1- • Anger-2(-ve) • Sad-2(-ve) • Fear-5(-ve) • Happy-1 • Surprised-3 • Frustrated-5(-ve) • Excited-1 • Neutral -1  
   Model 2- 5  
   Many models can be built and tested the same way
2. **Result and Analysis**

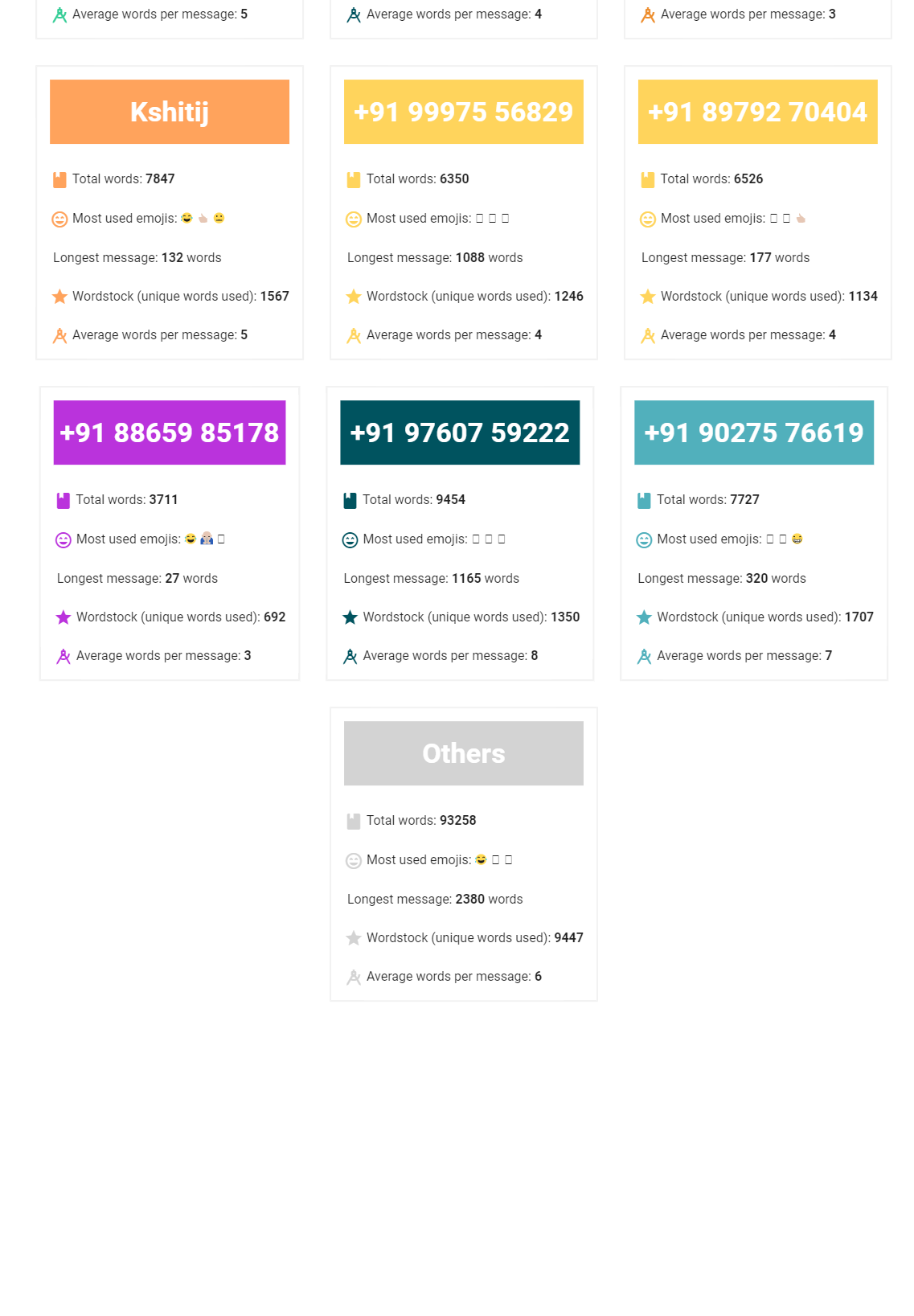
**Result of chat visualization**

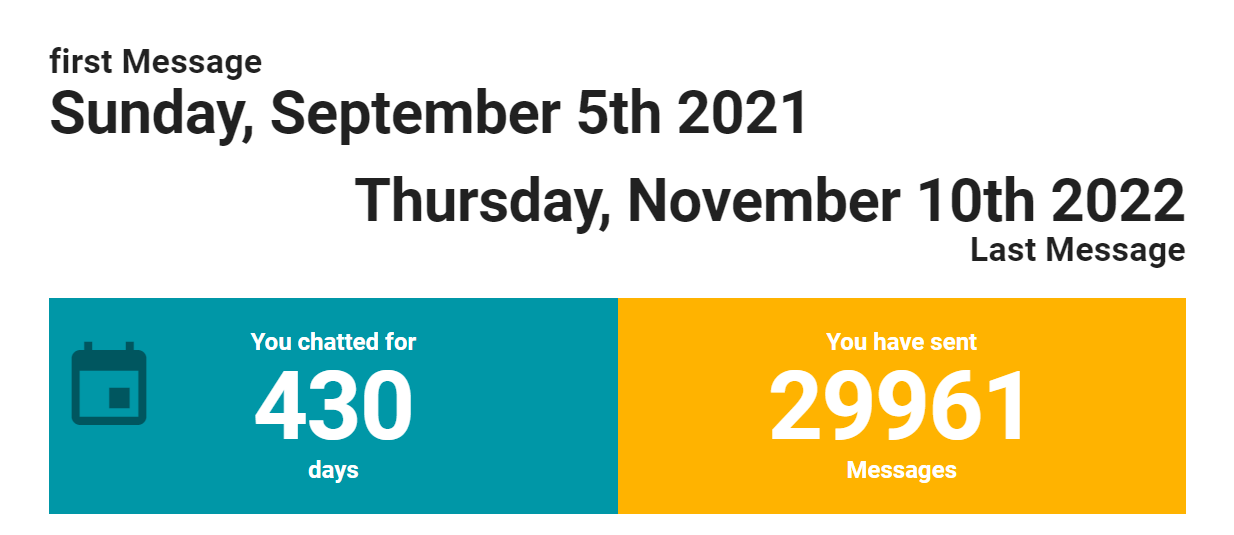
So I have done analysis of what chat of my unofficial group section J , and here are result of those analysis

**Chat timeline output**

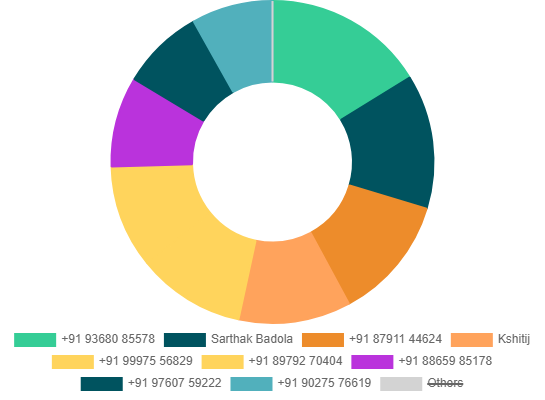
****

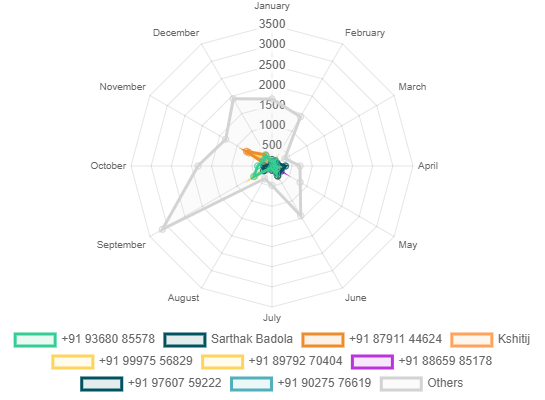
**Chat fun fact output**

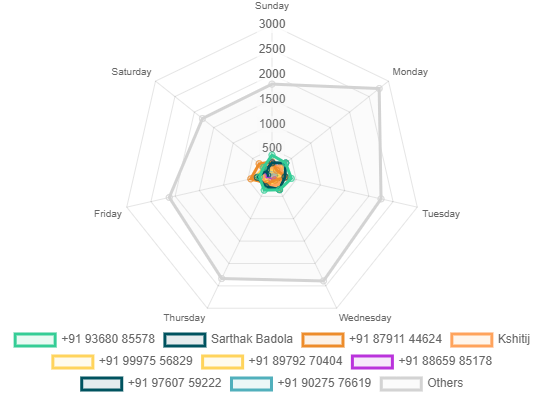
****



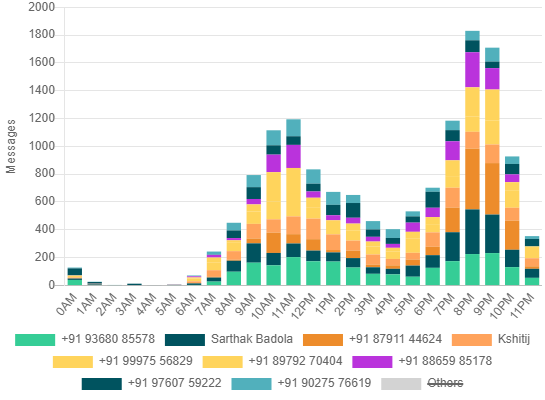
**Message per person**

****

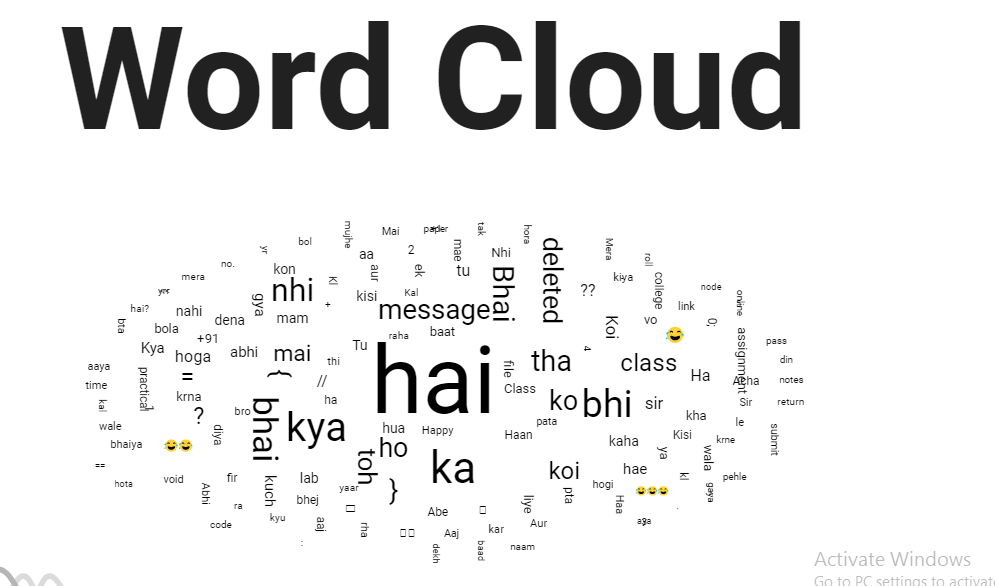
**Message per month**

**Message per weekday**

**Chat time of the day**

****

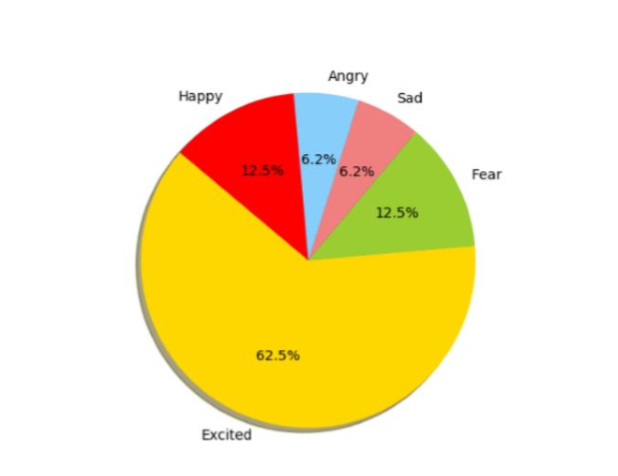
**Word cloud**

****

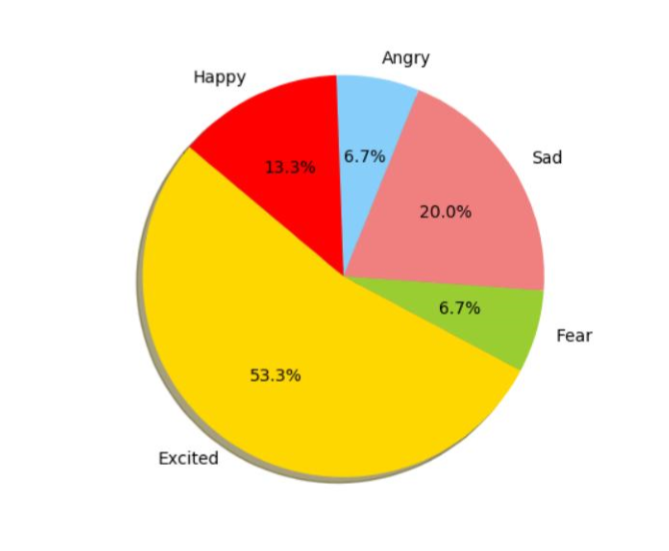
**Result of sentiment analysis**

**Now, using the usual procedure** based on **choosing** the **best emotion, we get the following result  
.** This data **was** processed **based** on chat data collected from **one-on-one chats, preprocessing** and **recognizing that** person A **was in a** particular **chat, with different** emotions **expressed in different colors. Visualized  
.**  
The **following** pie chart **shows what emotions look like**

**The below pie chart shows the variations of the emotions of the person**

****

**The below pie chart shows the variations of the emotions of the person**

****

**References**

https://link.springer.com/chapter/10.1007/978-3-319-26925-2\_17

https://whatsanalyze.com/

https://link.springer.com/chapter/10.1007/978-981-13-6347-4\_5

https://www.analyticsvidhya.com/blog/2021/04/whatsapp-group-chat-analyzer-using-python/

https://assets.researchsquare.com/files/rs-2206392/v1/bd1f509b-1498-4e18-af5a-483fb0b492ff.pdf?c=1667246633

https://www.youtube.com/watch?v=Q0QwvZKG\_6Q

https://www.youtube.com/redirect?event=video\_description&redir\_token=QUFFLUhqbjktbnFuUE1oc3gzdTczYktfRVM2VU1JdkFBQXxBQ3Jtc0ttMTdydWhhNHBxNTlVSEY5cXNXQlA1a0UzRDRlOTZwM2E4YjJmV3hLaTdjQ0ZrS28xZm9RMjkyTHh2MGpQWEFDY1FSOV9FOC1KQ0JMS1Q4RldNN1gtMG9GZXd5bzRReG1qSEN5UFlqajVXRkJYZ2pwaw&q=https%3A%2F%2Fgithub.com%2Fcampusx-official%2Fwhatsapp-chat-analysis&v=Q0QwvZKG\_6Q

https://wca-campusx.herokuapp.com/