Department of Information Technology
NBA Accredited

A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI
Academic Year 2021-2022

WhatsApp Chat Analysis

Kiran Suryawanshi – 19104035
Mayuresh Prabhu – 19104051
Tanmay Doshi -19104024

Project Guide

Prof. Shafaque Syed

Contents

1. Project Conception and Initiation

2. Project Design

3. Implementation

4. Result

5. Conclusion and future scope

6. References

1. Project Conception and Initiation

1.2. Objectives

- To provide user friendly interface for performing operations.
- To pre-process the data in such a way that can be suitable to the model.
- To automate the process of analysing the WhatsApp chat.
- To have ability to analyse chats of group, individual participant within the group and personal chats.
- To develop a statistical and analytical report on WhatsApp chats.
- To predict the sentiment of uploaded chats as postive, negative and neutral.

1.3. Literature Review

- The inspiration of this project is from the web application name “CHATILYZER”.
- This application is common visualization tool containing statistical features like top stats, most messaged user graph, analysis of emojis and word cloud.
- The drawbacks of this application is that it has only fun stats having funny words (Eg:Yappers) which user may not recognize.
- Also it doesn't contain the major part of sentiment analysis known as positive, negative & neutral chats predictions

1.4. Problem Definition

- WhatsApp Chat Analyzer is a statistical analysis tool for WhatsApp chats.
- Working on the chat files that can be exported from WhatsApp it generates various plots showing, for example: who is the busiest user in the group.
- We propose to employ dataset manipulation techniques to have a better understanding of WhatsApp Chat present in our phones

1.4. Scope

- The application can be used by certain investigative officers to analyze skeptical WhatsApp chats for investigation purpose.

- It can also be used in digital marketing field which can help for making new marketing strategy.

- It can also make the job of exploratory analysis of chats much convenient as it answers most of the common questions which arises during analysis which would be feasible for data analyst and data scientist.

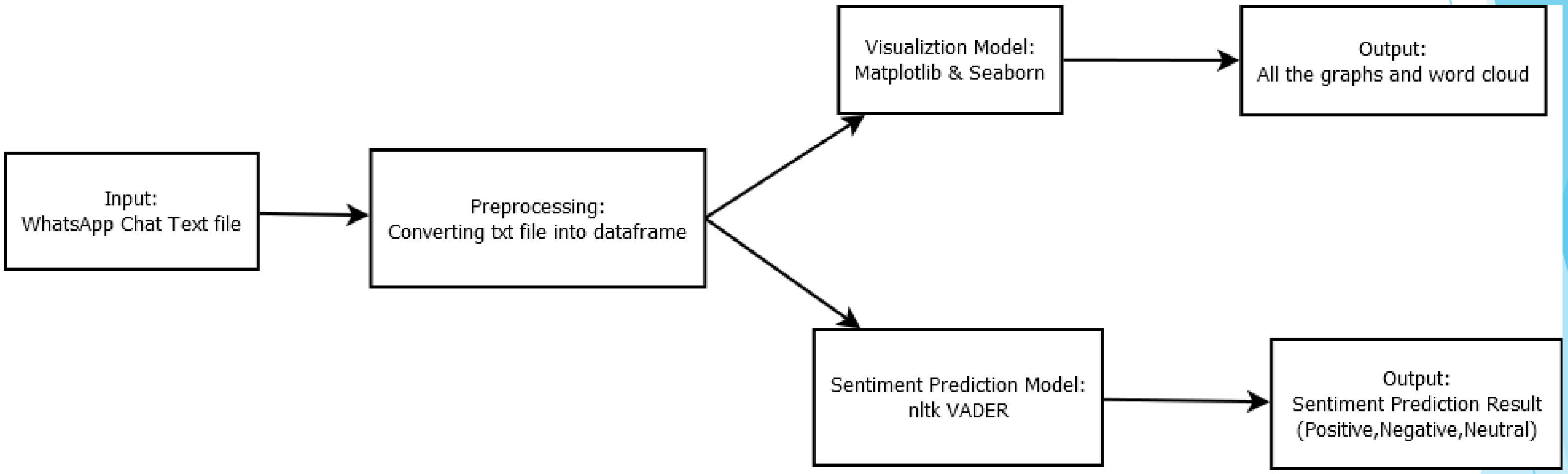
1.5. Technology Stack

1. We are using Core Python and it’s libraries for whole project
2. Following are the Python libraries/pakeages that we will be using:

- Streamlit – For Developing Web UI
- Pandas – For Pre-processing the text file
- Matplotlib, Word Cloud and Seaborn – For Visualizing the Data
- Natural language tool kit – For processing the chats and sentimental prediction.

2. Project Design

2.1 Proposed System



Input:
WhatsApp Chat Text file

Preprocessing:
Converting txt file into dataframe

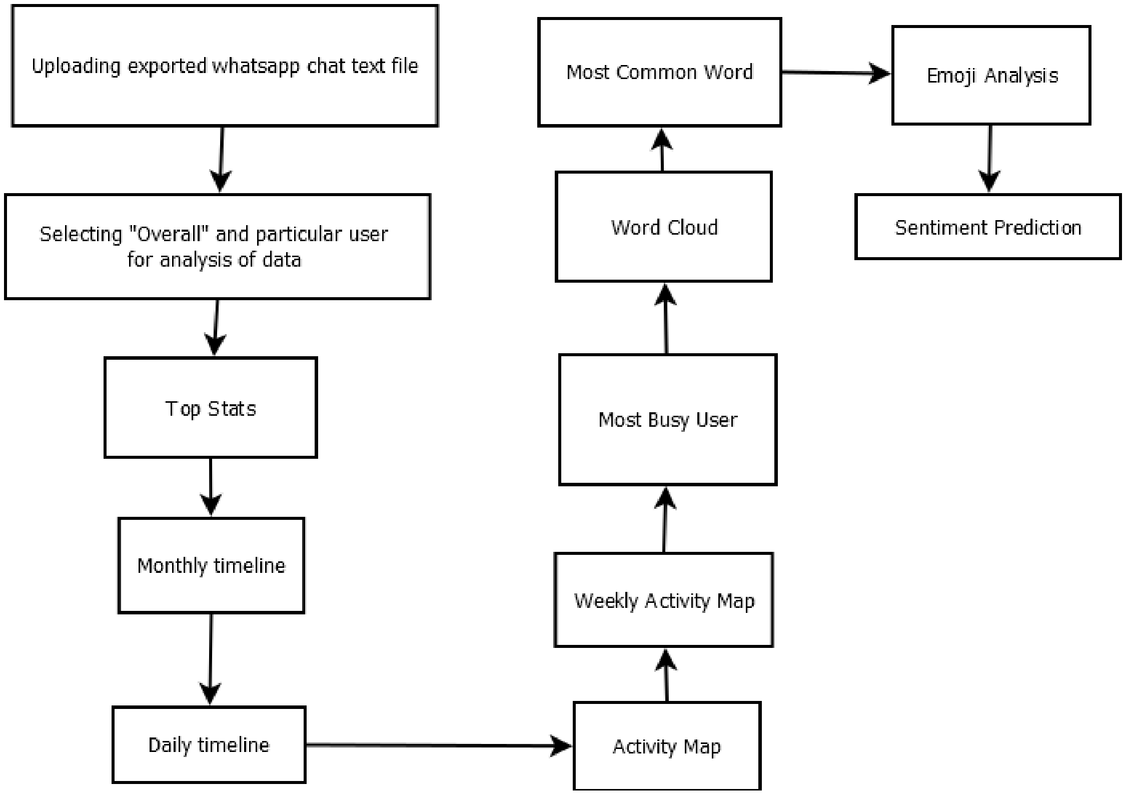
Visualiztion Model:
Matplotlib & Seaborn

Output:
All the graphs and word cloud

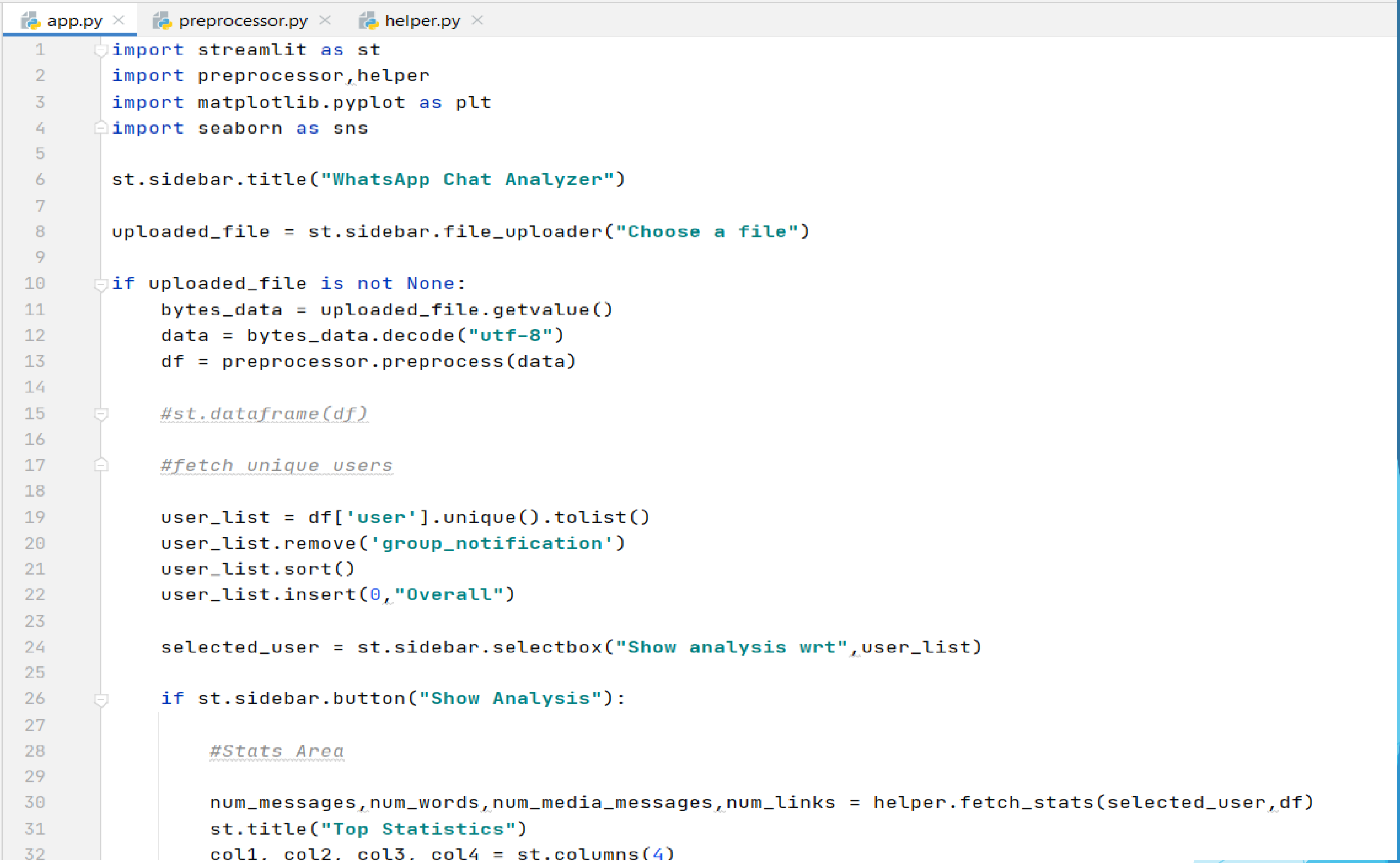
Sentiment Prediction Model:
nltk VADER

Output:
Sentiment Prediction Result
(Positive,Negative,Neutral)

2.2 Design (Flow of Modules)

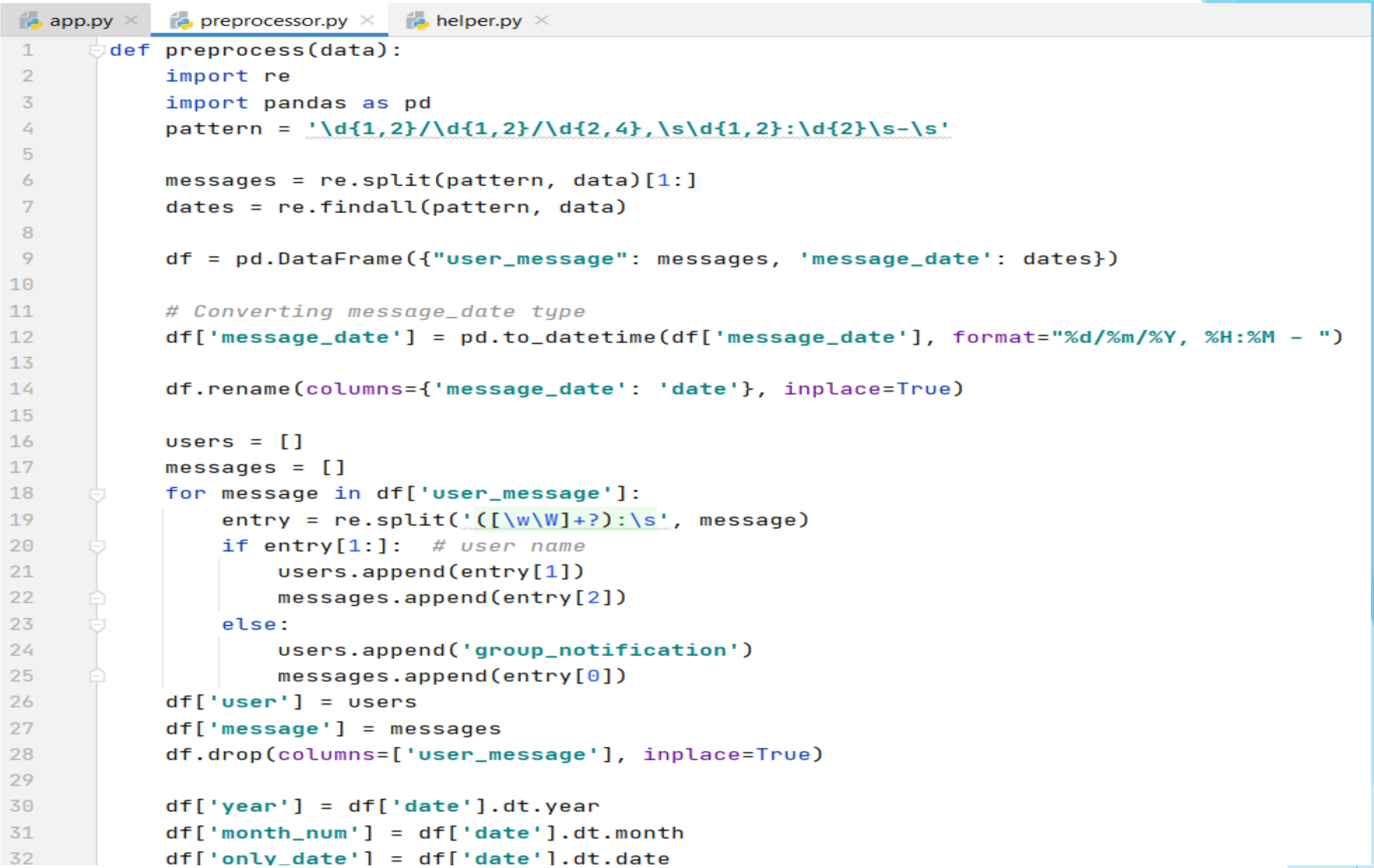


3.Implementation



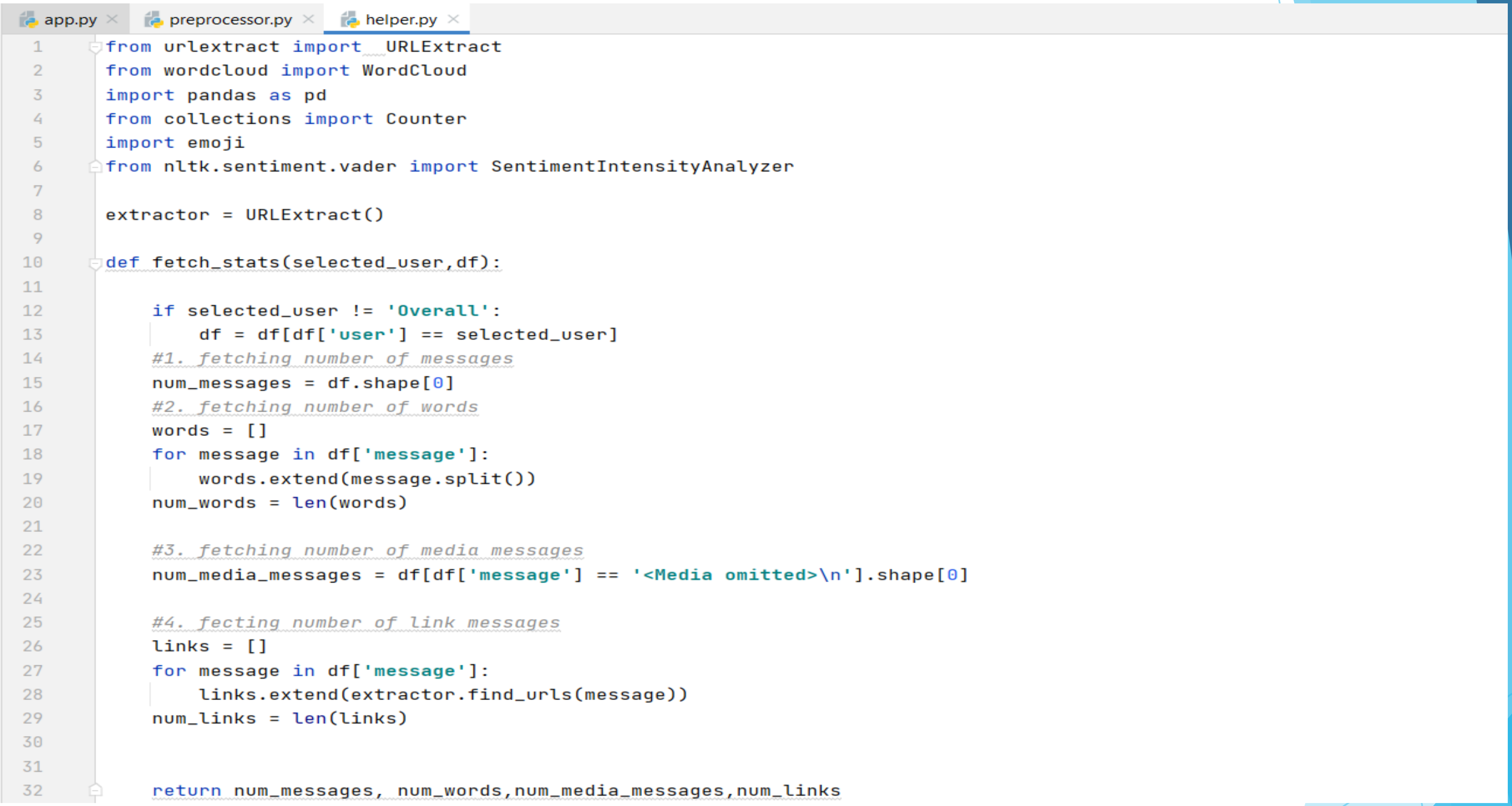
```python
import streamlit as st
import preprocessor,helper
import matplotlib.pyplot as plt
import seaborn as sns

st.sidebar.title("WhatsApp Chat Analyzer")

uploaded_file = st.sidebar.file_uploader("Choose a file")

if uploaded_file is not None:
    bytes_data = uploaded_file.getvalue()
    data = bytes_data.decode("utf-8")
    df = preprocessor.preprocess(data)

    #st.dataframe(df)

    #fetch unique users

    user_list = df['user'].unique().tolist()
    user_list.remove('group_notification')
    user_list.sort()
    user_list.insert(0,"Overall")

    selected_user = st.sidebar.selectbox("Show analysis wrt",user_list)

    if st.sidebar.button("Show Analysis"):

        #Stats Area

        num_messages,num_words,num_media_messages,num_links = helper.fetch_stats(selected_user,df)
        st.title("Top Statistics")
        col1, col2, col3, col4 = st.columns(4)
```

app.py



```python
def preprocess(data):
    import re
    import pandas as pd
    pattern = '\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'

    messages = re.split(pattern, data)[1:]
    dates = re.findall(pattern, data)

    df = pd.DataFrame({"user_message": messages, 'message_date': dates})

    # Converting message_date type
    df['message_date'] = pd.to_datetime(df['message_date'], format="%d/%m/%Y, %H:%M - ")

    df.rename(columns={'message_date': 'date'}, inplace=True)

    users = []
    messages = []
    for message in df['user_message']:
        entry = re.split('([\w\W]+?):\s', message)
        if entry[1:]:  # user name
            users.append(entry[1])
            messages.append(entry[2])
        else:
            users.append('group_notification')
            messages.append(entry[0])
    df['user'] = users
    df['message'] = messages
    df.drop(columns=['user_message'], inplace=True)

    df['year'] = df['date'].dt.year
    df['month_num'] = df['date'].dt.month
    df['only_date'] = df['date'].dt.date
```

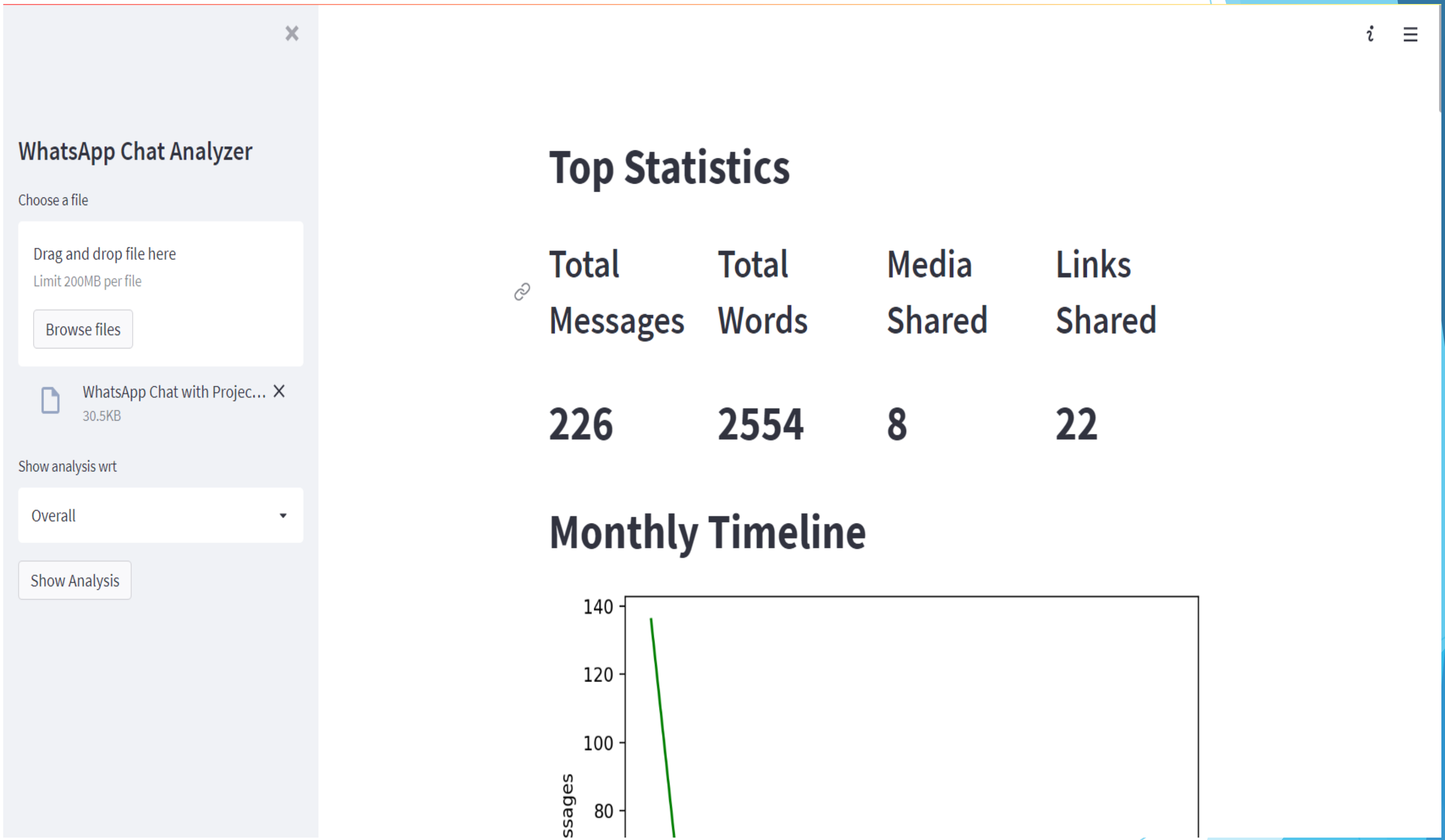
preprocessor.py



```python
from urlextract import  URLExtract
from wordcloud import WordCloud
import pandas as pd
from collections import Counter
import emoji
from nltk.sentiment.vader import SentimentIntensityAnalyzer

extractor = URLExtract()

def fetch_stats(selected_user,df):

    if selected_user != 'Overall':
        df = df[df['user'] == selected_user]
    #1. fetching number of messages
    num_messages = df.shape[0]
    #2. fetching number of words
    words = []
    for message in df['message']:
        words.extend(message.split())
    num_words = len(words)

    #3. fetching number of media messages
    num_media_messages = df[df['message'] == '<Media omitted>\n'].shape[0]

    #4. fecting number of link messages
    links = []
    for message in df['message']:
        links.extend(extractor.find_urls(message))
    num_links = len(links)


    return num_messages, num_words,num_media_messages,num_links
```

helper.py

4.Results



WhatsApp Chat Analyzer

Choose a file

Drag and drop file here

Limit 200MB per file

WhatsApp Chat with Projec…

30.5KB

Show analysis wrt

Overall ▾

Top Statistics

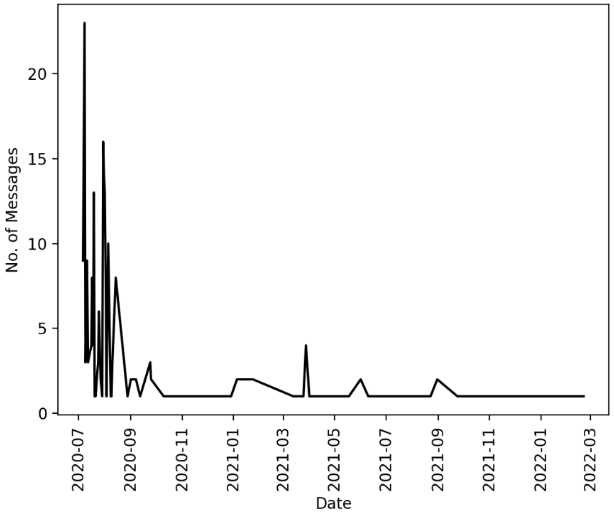
Total Messages	Total Words	Media Shared	Links Shared
226	2554	8	22

Monthly Timeline

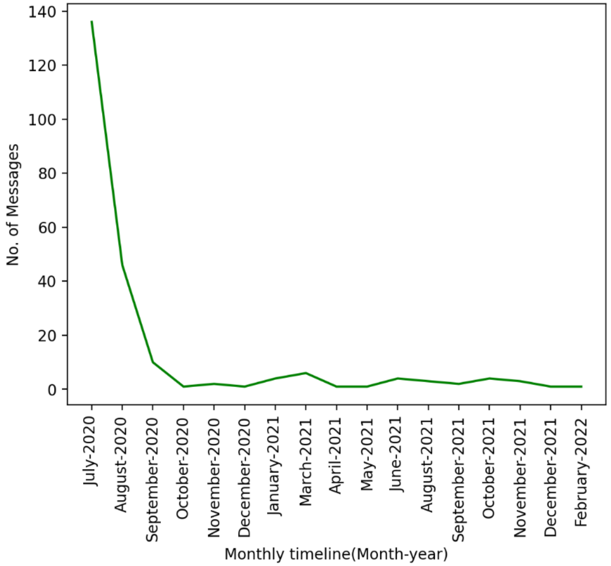
Top Statistics

Total Messages	Total Words	Media Shared	Links Shared
226	2554	8	22

Daily Timeline

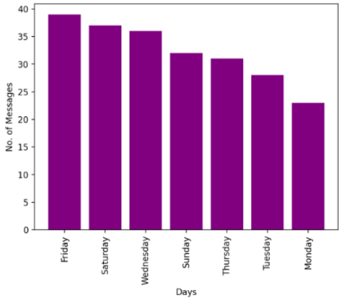


Monthly Timeline

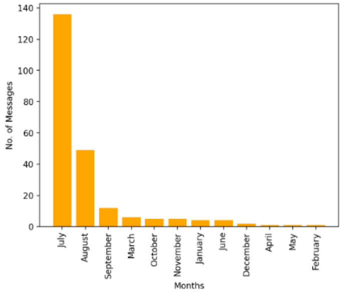


Activity Map

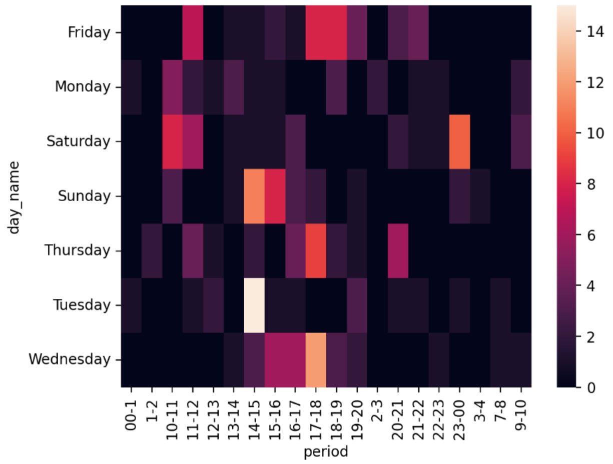
Most busy day



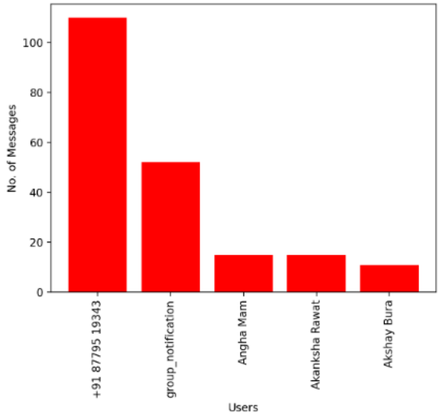
Most busy month

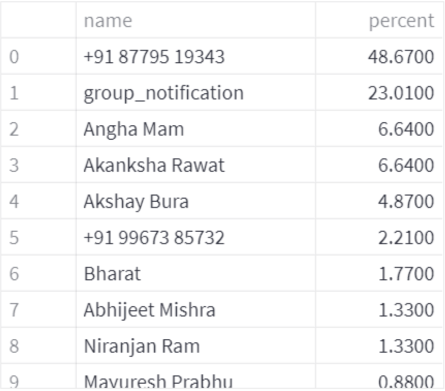


Weekly Activity Map



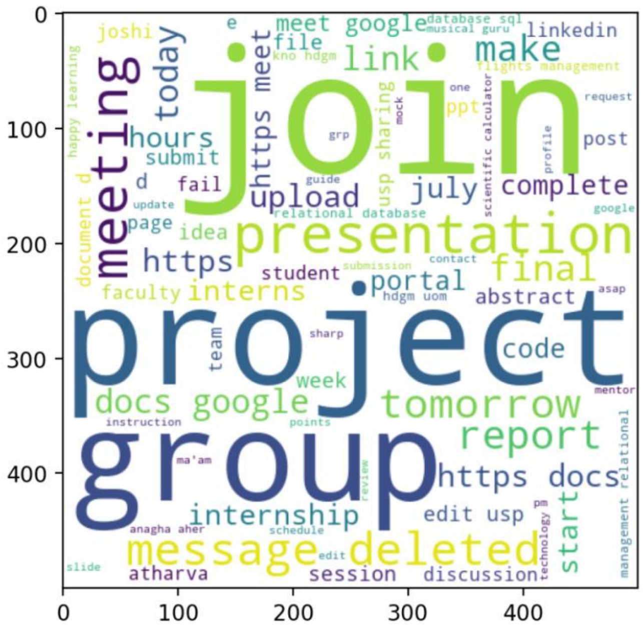
Most Busy Users



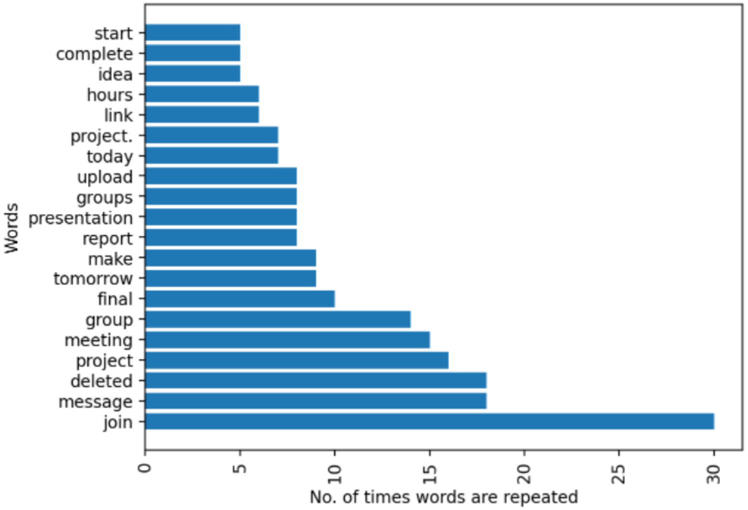


	name	percent
0	+91 87795 19343	48.6700
1	group_notification	23.0100
2	Angha Mam	6.6400
3	Akanksha Rawat	6.6400
4	Akshay Bura	4.8700
5	+91 99673 85732	2.2100
6	Bharat	1.7700
7	Abhijeet Mishra	1.3300
8	Niranjan Ram	1.3300
9	Mayuresh Prabhu	0.8800

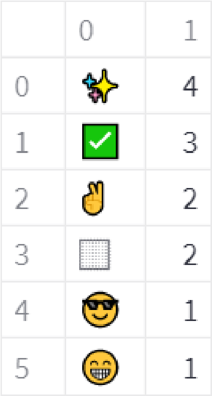
Word Cloud



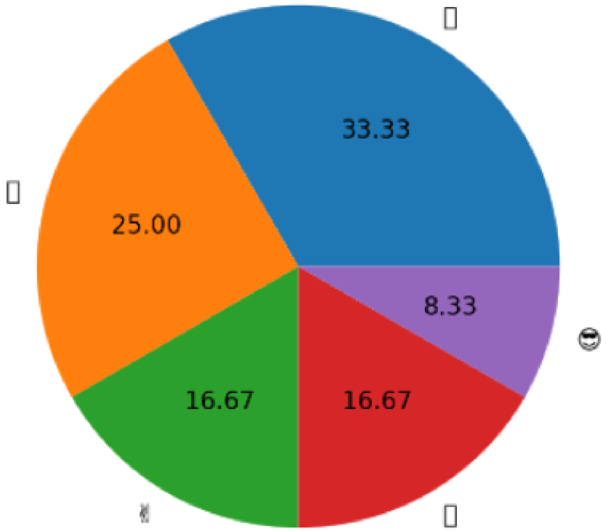
Most Common Words



Emoji Analysis



	0	1
0	✨	4
1	✅	3
2	✌️	2
3	⬚	2
4	😎	1
5	😁	1



Sentiment Prediction

Neutral: The uploaded chats of user/group is neutral which conclude that the chat neither contains enough negative words nor enough of positive words

5.Conclusion and Future Scope

5.1 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 project was able to create an analysis of a WhatsApp group chat and visual representation of chats(i.e which are most active participant,total count of messages, wordcloud of chats).
- On Series note, this System has the ability to analyze any WhatsApp group data input into it.

5.2 Future Scope:

- The Application can be upgraded to perform Topic Modeling(i.e topic of the chat can be decided using contents).
- It can also be upgraded to perform sentiment analysis on images using image processing.
- Since our application is only analyzing english text for sentiment prediction we can further upgrade it for regional languages.

6. References:

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

Thank You...!!