A Mini Project Report on

WhatsApp Chat Analysis

T.E. - I.T. Engineering

Submitted By

Kiran Suryawanshi - 19104035

Mayuresh Prabhu - 19104051

Tanmay Doshi - 19104024

Under The Guidance Of

Prof. Shafaque Syed

DEPARTMENT OF INFORMATION TECHNOLOGY

A. P. SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year : 2021 - 22

CERTIFICATE

This to certify that the Mini Project report on **WhatsApp Chat Analysis** has been submitted by **Mayuresh Prabhu** (19104051), **Kiran Suryawanshi** (19104035) and **Tanmay Doshi** (19104024) who are a Bonafide students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2021-2022** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Shafaque Syed

Guide

Prof. Kiran Deshpande — Head of the Department - Information Technology

Dr. Uttam D. Kolekar — Principal

External Examiner(s) : 1. _____

2. _____

Place: A. P. Shah Institute of Technology, Thane

Date:

TABLE OF CONTENTS

Chapter 1

Introduction

In the age of internet and socialization, communication and social media applications like “WhatsApp” is common to be used by everybody. Often WhatsApp is the medium where public use it to communicate with someone and also express their opinion on certain topics. In the past it has become a key evidence in certain criminal cases and this application can be use for marketing of product. So there is the necessity to develop an analytical report on data transfers through chats which give answers to common question which is raised during analysis.

Hence we are introducing a web application which can automate the process of analysis of WhatsApp chats and gives the statistical report which will be contain plots and sentimental prediction. In this application user can do analyse chats of a group, member within the group and individual.

1.1 Purpose:

The purpose of this projects is to statistically analyze the whatsapp chat to answer some of the common question which will help to develop the marketing strategy, helps in investigation of the case and also automate the work of exploratory data analysis.

1.2 Objectives:

Following objectives of our project:

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

1.3 Scope:

1. The application can be used by certain investigative officers to analyze skeptical WhatsApp chats for investigation purpose.
2. It can also be used in digital marketing field which can help for making new marketing strategy.
3. 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.

Chapter 2

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

Chapter 3

Proposed System

The system which have been proposed will get the WhatsApp Chat exported text file as input which will than converted into dataframe. After converting into dataframe it will go through some data manipulation and visualization process which will give all the visualization and sentiment prediction as the output.

3.1 Features and functionality

3.2.1 Sidebar:

- It is only feature where user need to do some operations inorder to get output.
- It consist of text file uploader which will upload the data of chat into text file.
- After uploading user can genrate the statistical dashboard by clicking the button “Show Analyses
- It will also contain a dropdown where user can analyse overall or individual within the group chats

3.2.2 Statistical Dashboard:

- It is the most important feature of our web app where user can see the visualization of chats after the analyses.
- Following are the visualisation which will be displayed:
    a) Top Statistics – Displays the total count of messages, words , media files, links shared.
    b) Timeline Graph – It display the line graph which indicates activeness of individual or group throughout year or daily.
    c) Activity Map – It display the bar Graph of most busy day and month.
    d) Weekly Activity Map – It visualize the heat map of days vs time.
    e) Busy User Bar Graph – Displays top 5 most busy users.
    f) Word Cloud
    g) Emoji Analysis – Dataframe and Pie Chart

3.2.3 Sentiment Prediction:

- It is the part of statistical dashboard where it will be displayed at last.
- It will process all chats through natural language processing and give us the prediction whether average nature of all chats are positive , negative or neutral

Chapter 4

Project Outcomes

- User can upload the chats in .txt format as text file.
- User can be able to analyse the chats of group, individual participant within the group and personal chats.
- User can generate the statistical and analytical report on WhatsApp Chat automatically.
- User can also display the prediction sentimental nature of chats as positive, negative and neutral.

Chapter 5

Software Requirements

We are using Python and it's libraries for whole project. Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small- and large-scale projects. It is the most preferred language used for Machine learning And Data Science Projects.

Following are the Python libraries that we will be using:

- Streamlit : It is an open-source web application framework for Machine Learning and Data Science Projects. We are using this library for developing our user interface.

- Pandas: It 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. We are using this library for pre-processing the text file

- Matplotlib: It is a comprehensive library for creating static, animated, and interactive visualizations in Python. We are using this library to visualize the the graphs and pie charts

- Seaborn: It is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. We have used this library to visualize the modern and advanced heat map graph.

- Natural language tool kit : It is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. We have used this library for calculating the polarity scores of each chats and giving the sentimental prediction as positive, negative and neutral.

Chapter 6

Project Design

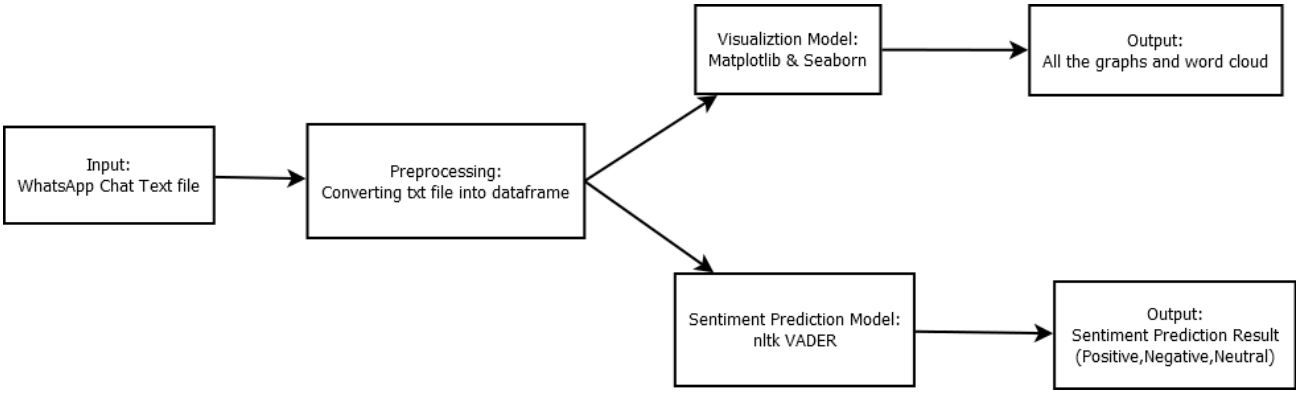

Fig 6.1: Conceptual Block Diagram

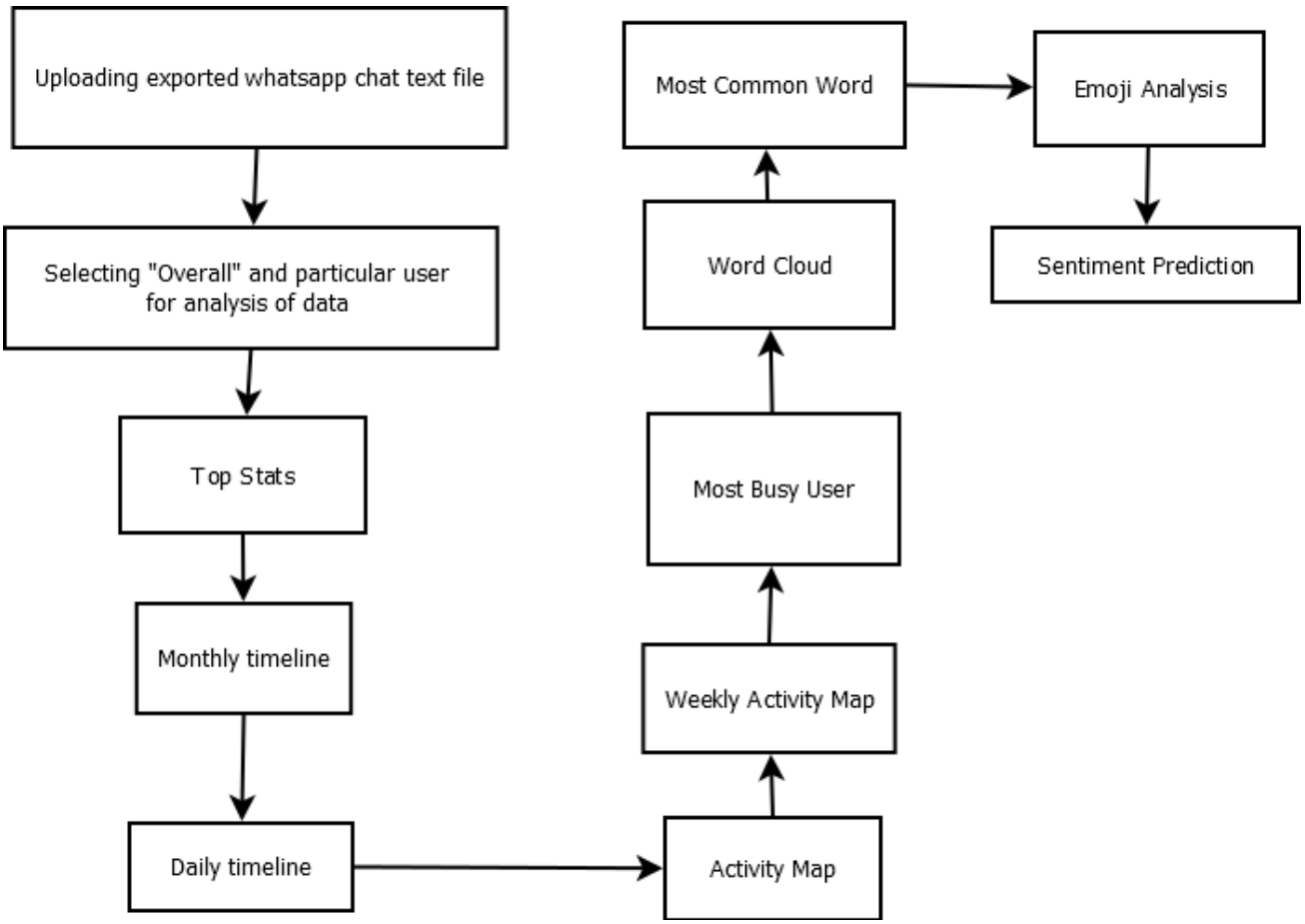

Fig 6.2: Flow of Module

Chapter 7

Project Scheduling Template

Sr. No	Group Member	Time duration	Work to be done
1	Mayuresh Prabhu	1st week of March	Preprocessing part
		2nd week of March	Creating Visualiztion method and model
2	Tanmay Doshi	1st week of April	Creating User Interface
3	Kiran Suryawanshi	2nd week of April	Integrating visualization methods and model into application

Chapter 8

Screenshot of applications

Fig 7.1: Main User Interface

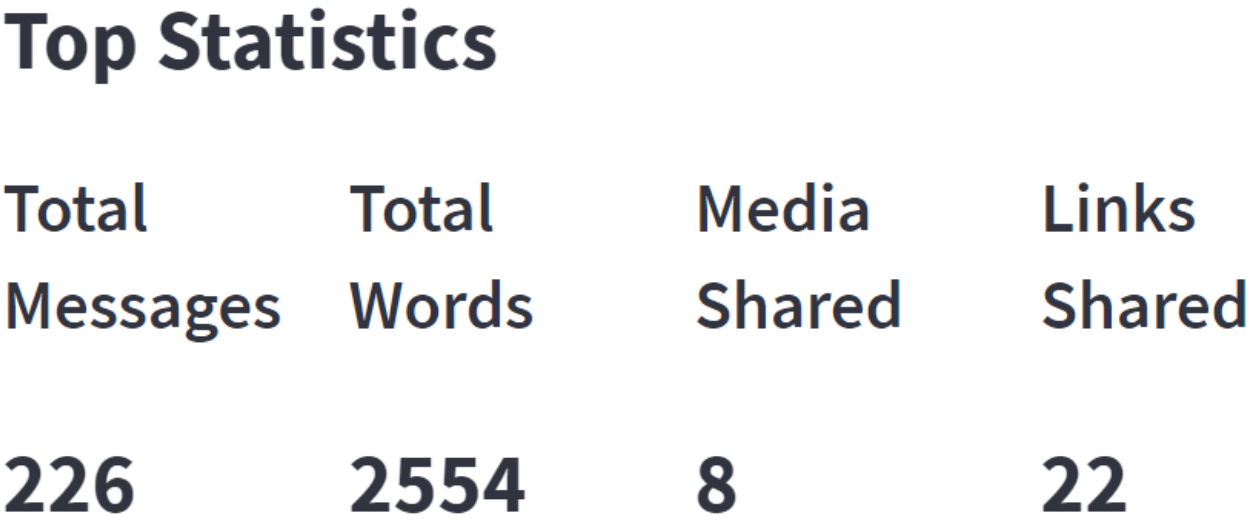

Fig 7.2: Top Statistics

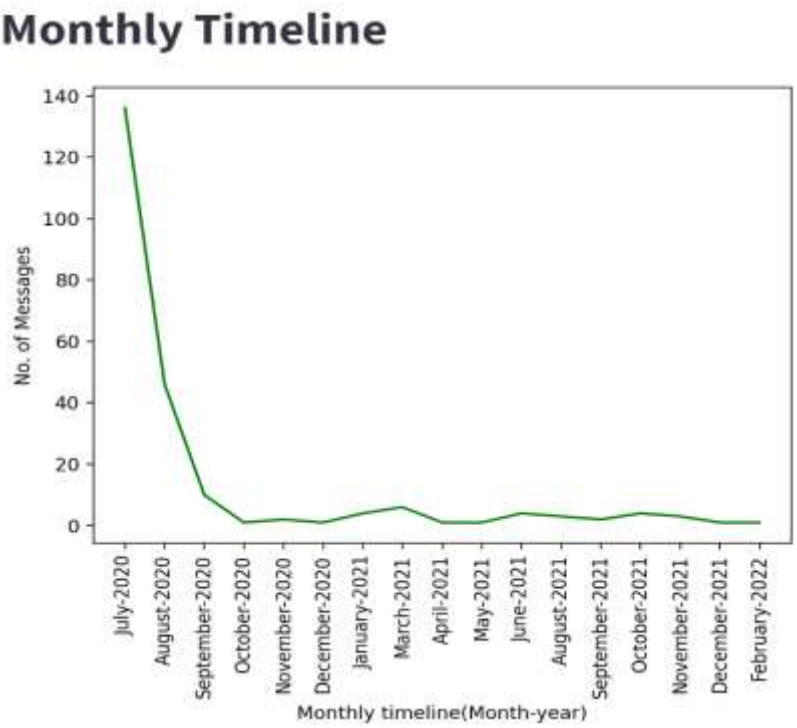

Fig 7.3: Monthly Timeline

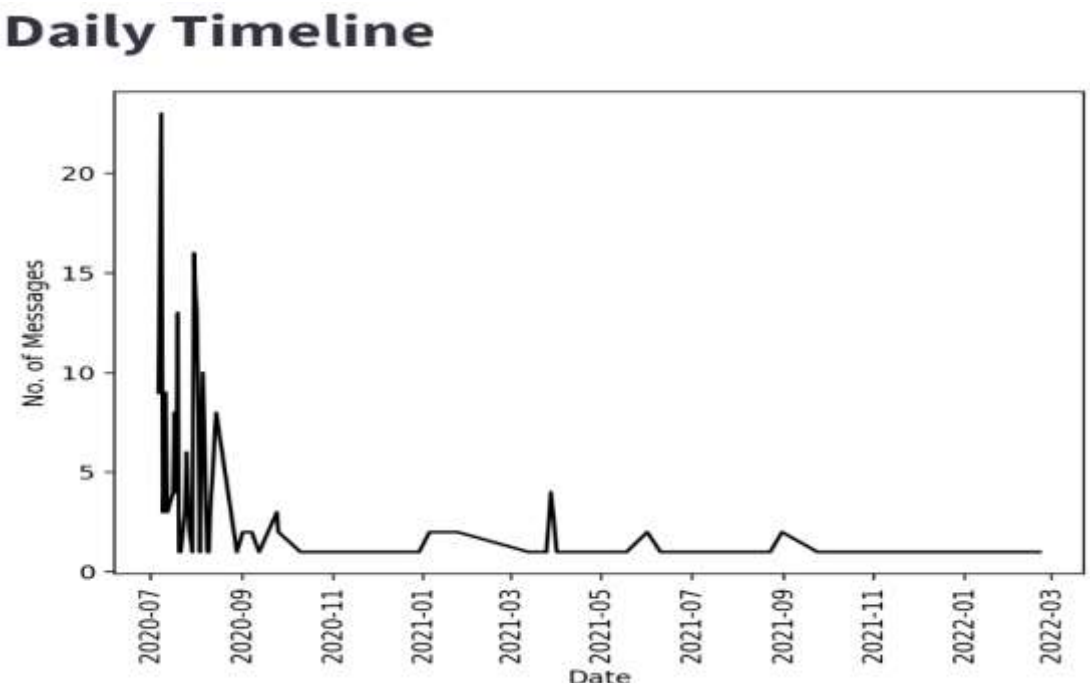

Fig 7.4: Daily Timeline

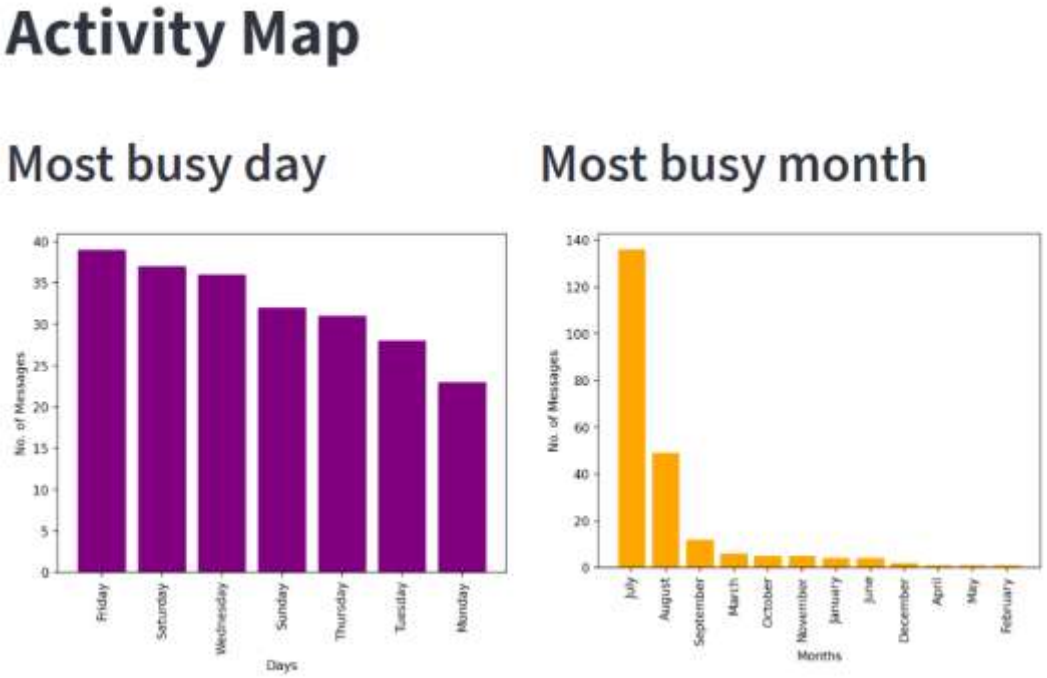

Fig 7.5: Activity Map

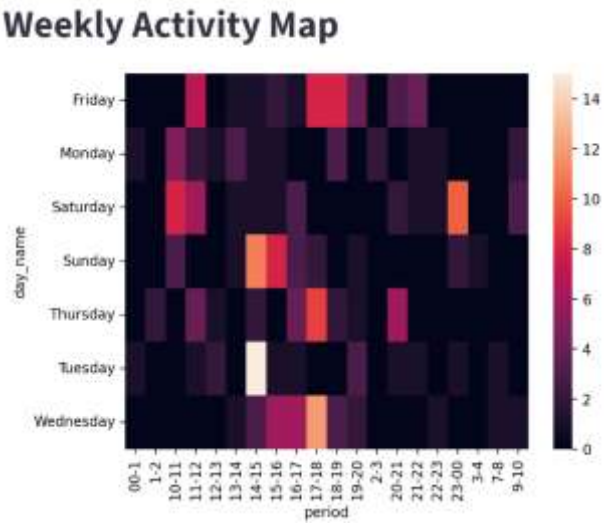

Fig 7.6: Weekly Activity Map

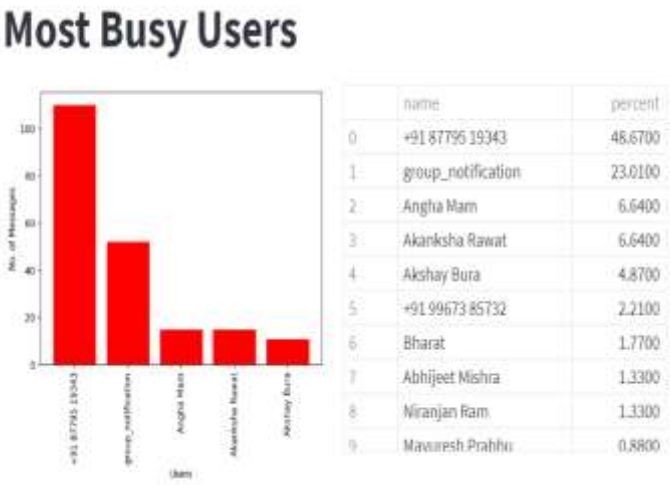

Fig 7.7: Most Busy User

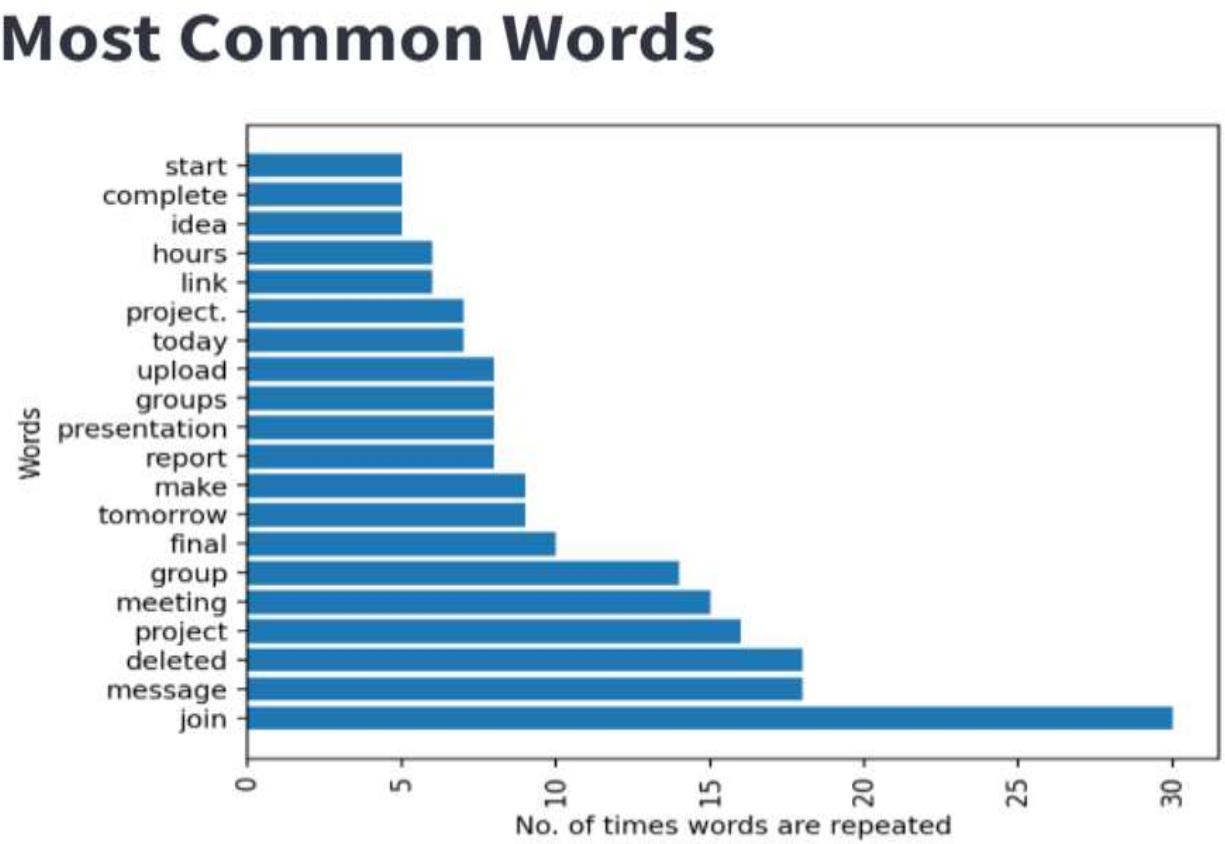

Fig 7.8: Most Common Word

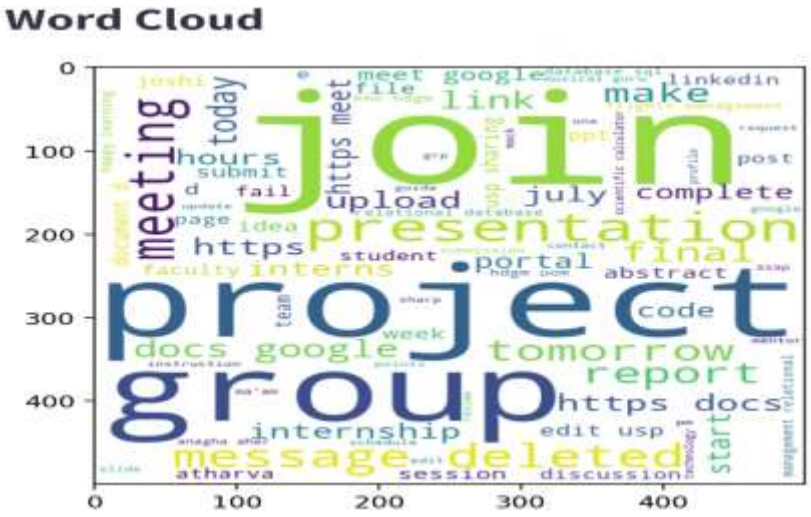

Fig 7.9: Word Cloud

Emoji Analysis

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

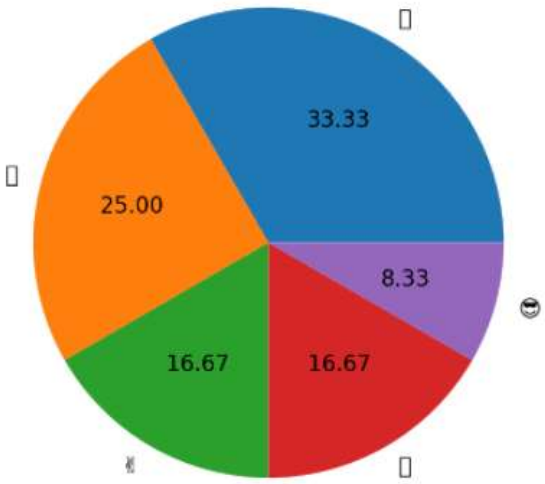

Fig 7.10: Emoji Analysis

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

Fig 7.11: Sentiment Prediction

Chapter 9

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 Serious note, this System has the ability to analyze any WhatsApp group data input into it.

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.

References

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

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof**. **Shafaque Syed** We express our gratitude towards our HOD **Prof. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher **Prof. Nahid Shaikh** who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.