



# PRESIDENCY UNIVERSITY

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956  
Approved by AICTE, New Delhi



**A MINI PROJECT REPORT ON**

## **WHATSAPP GROUP CHAT ANALYSIS**

*Submitted in partial fulfillment for the award of the  
degree of Bachelor of Engineering*

*In*

**COMPUTER ENGINEERING**

*Submitted by*

**RASHMI KHANDRE [20191COM0173] 7COM3**

**RASHI PANDEY [20191COM0172] 7COM3**

**PRANEETH R [20191COM0162] 7COM3**

**MEGHANA M [20191COM0126] 7COM3**

**PRATHEEK [20191COM0163] 7COM3**

*Reviewed by*

**Dr. PAMELA VINITHA ERIC**

**[Prof, CSE]**



# PRESIDENCY UNIVERSITY

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956  
Approved by AICTE, New Delhi



## *Certificate*

*This is to certify that the mini project work titled*

### **WHATSAPP GROUP CHAT ANALYSIS**

*Submitted in partial fulfillment for the award of the degree  
of Bachelor of Engineering*

**RASHMI KHANDRE [20191COM0173] 7COM3**  
**RASHI PANDEY [20191COM0172] 7COM3**  
**PRANEETH R [20191COM0162] 7COM3**  
**MEGHANA M [20191COM0126] 7COM3**  
**PRATHEEK [20191COM0163] 7COM3**

*During the academic year  
2022-2023*

**Signature of Reviewer**

**Signature of IC**

## **ACKNOWLEDGEMENT**

The satisfaction and euphoria that accompany the successful completion of any task would be, but impossible without the mention of the people who made it possible, whose constant guidance and encouragement crowned my efforts with success.

We thank the management, **Dr. Nissar Ahmed**, Chancellor Presidency University, for providing necessary infrastructure and creating a good environment.

We also record here the constant encouragement and facilities extended to me by **Dr. Mohammed Sameeruddin Khan**, Dean, SOCSE & IS, Presidency University, **Dr. C Kalaiarasan**, Assoc. Dean, SOCSE & IS, Presidency University. I extend my sincere gratitude to them.

We express our gratitude to **Dr. Pamela Vinitha Eric**, Professor, Department of Computer Science and Engineering, Presidency University, our project reviewer for constantly monitoring the development of the project and setting up precise deadlines. Her valuable suggestions were the motivating factors in completing the work.

Finally, a note of thanks to all the teaching and non-teaching staff of School of Computer Science and Engineering and Information Science for their cooperation extended to me and my friends, who helped us directly or indirectly in the course of the project work.

[ STUDENT NAME WITH USN ]

## ABSTRACT

WhatsApp is one of the most used messenger applications today with more than 2 billion users worldwide. It was found that more than 65 billion messages are sent on WhatsApp daily so we can use WhatsApp chats for analyzing our chat with a friend, customer, or a group of people. In this project, we will take you through the task of WhatsApp Chat Analysis with Python.

The WhatsApp Group Chat analysis will show the group members behavior and what is the group motive. We will use python libraries to analyze the chat.  
Find most active users in the group, find time when most users are active.  
Sentiment Analysis on individual member or on Over-all group chat.

We can use your WhatsApp data for many data science tasks like *sentiment analysis*, keyword extraction, named entity reorganization text analysis and several other natural language processing tasks. It also depends on who you are analyzing your WhatsApp messages with because you can find a lot of information from your WhatsApp messages which can also help you to solve business problems.

## CHAPTERS

CHAPTER NO	TITLE	PAGE NO
	<b>Abstract</b>	i
	<b>Acknowledgment</b>	ii
	<b>List of Figures</b>	iii
1	<b>Introduction</b>	1
2	<b>Analysis and Design</b>	2
2.1	Objectives of the Project	3
2.2	Requirement Specification	4
3	<b>Implementation</b>	5
3.1	Python Libraries	6
3.2	Data Processing	7
3.2.1	Creating the Data Frame	8
3.2.2	Dropping Messages with No Authors	9
3.3	Analyzing Number of Messages Sent by the Group Members	
3.4	Finding the Letter and Word Count from Each Message	



## LIST OF FIGURES

Figure No	Figure Name	Page No
<b>Fig.2.3.1</b>	<b>Architectural Diagram</b>	<b>3</b>
<b>Fig.2.3.3.</b>	Steps involved in data Collection	4
<b>Fig.3.1.</b>	Analysis using Time Series Plot	6
<b>Fig.3.2.</b>	No. of message sent by group members	6
<b>Fig.3.3.</b>	Total Message Count	7
<b>Fig.3.4</b>	<b>Active hours of group</b>	<b>7</b>

## CHAPTER 1

# INTRODUCTION

### 1.1 INTRODUCTION TO SOCIAL MEDIA ANALYTICS

Practitioners and analysts alike know social media by its many websites and channels: Facebook, YouTube, Instagram, Twitter, LinkedIn, Reddit and many others.

Social media analytics is broader than metrics such as likes, follows, retweets, previews, clicks, and impressions gathered from individual channels. It also differs from reporting offered by services that support marketing campaigns such as LinkedIn or Google Analytics.

Social media analytics uses specifically designed software platforms that work similarly to web search tools. Data about keywords or topics is retrieved through search queries or web ‘crawlers’ that span channels. Fragments of text are returned, loaded into a database, categorized and analyzed to derive meaningful insights.

Social media analytics includes the concept of social listening. Listening is monitoring social channels for problems and opportunities. Social media analytics tools typically incorporate listening into more comprehensive reporting that involves listening and performance analysis.

### 1.2 INTRODUCTION OF THE PROJECT

In this report I have proposed a WhatsApp Chat Analyzer. WhatsApp chats contains of different types of communications held among groups and personal chats. This chat contains of different topics. This can provide more data for technologies like machine learning. Machine learning models provides right learning experience which is important thing and indirectly affected by the data which provided to that model. This application provides analysis of this data which is WhatsApp provides. The advantage of this application is that it is implemented by simple python libraries like seaborn, pandas, NumPy, streamlit and matplotlib which are commonly use for creating data frames and



different graphs. This is displayed in web using heroku link which can run on all devices which supports browser.

### **1.3 ADVANTAGES OF THE PROJECT**

- Runs on all devices.
- Shows based on WhatsApp chat file.
- Shows different visualizations.
- Total Messages.
- Total words.
- Media shared.
- Link shared.
- Monthly timeline.

## **CHAPTER 2**

### **ANALYSIS AND DESIGN**

#### **2.1 OBJECTIVES OF THE PROJECT**

More than 34 billion texts according to report are exchanged over WhatsApp every day. If we could analyze and get valuable insights from this data and leverage it to not only take better real time decisions but also add value to stake holders at much lower cost and time. We can analyze the sentiment of any group.

The WhatsApp Group Chat analysis will show the group members behavior and what is the group motive. We will use python libraries to analyze the chat. Find most active users in the group, find time when most users are active. Sentiment Analysis on individual member or on Over-all group chat.

#### **2.2 REQUIREMENT SPECIFICATION**

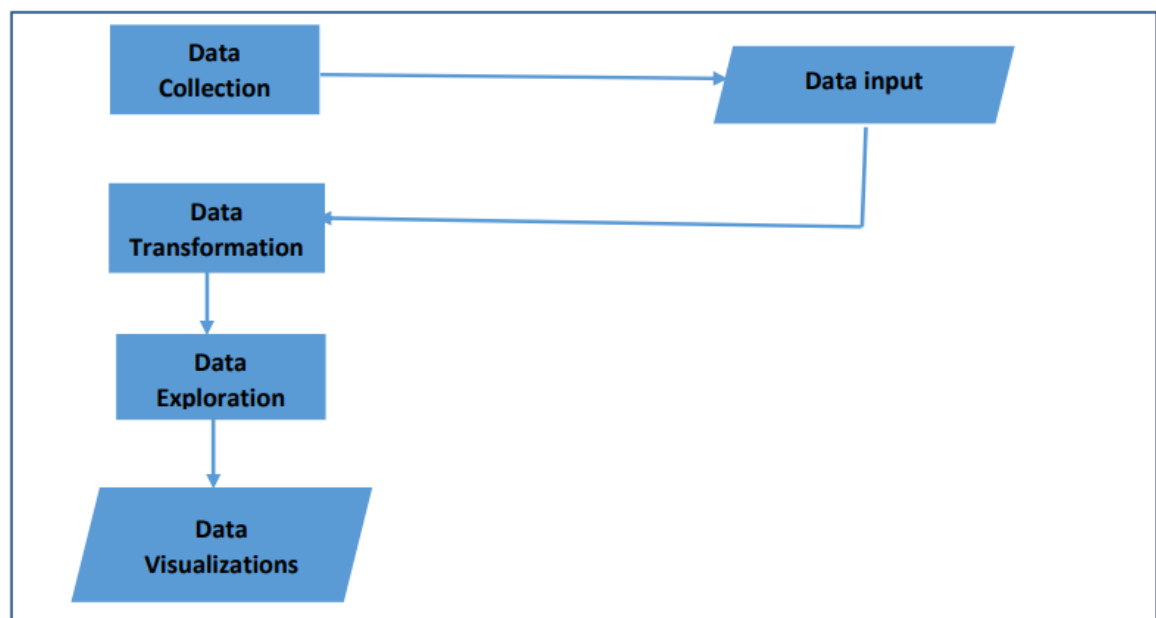
##### **2.2.1. Required Libraries**

- Regex
- Pandas

- Matplotlib
- NumPy
- Seaborn
- Datetime
- Emoji
- Word cloud
- Heatmaps
- NLTK
- Plotly

## 2.3 ALGORITHM / PSUEDO CODE

### 2.3.1 ARCHITECTURAL DESIGN



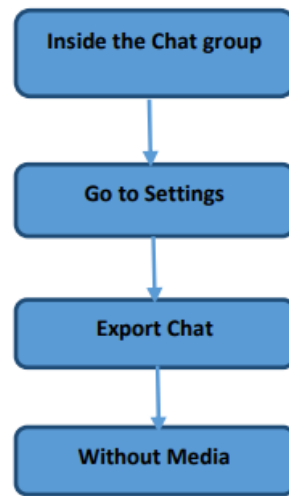
**Fig.2.3.1. Architectural Diagram**

### 2.3.2 ALGORITHM

- Step 1: Import all these libraries
- Step 2: Extract data from WhatsApp
- Step 3: Extract all the useful info. from chat file using regex
- Step 4: Check the basic information of our dataset and clean the dataset
- Step 5: Preprocess our dataset and try to extract useful information from it
- Step 6: Extracting basic statistics from the dataset
- Step 7: Extracting basic statistics of each user

Step 8: create a word cloud of most used words in chat

Step 9: Real time analysis as required



**Fig.2.3.3.** Steps involved in data Collection

## CHAPTER 3

### IMPLEMENTATION

#### 3.1. Python Libraries

We used Jupyter Notebook, which is an open-source web application that allows you to see the intermediate outputs easily. These are the several packages used for this analysis.

#### 3.2. Data Processing

We exported the chat of a WhatsApp group where there are 5 members and it was created way back for this project.

##### 3.2.1. Creating the Data Frame

This plain text file will have to be tokenized and parsed into the above-mentioned attributes in a meaningful manner to be stored in a Pandas Data Frame.

To Detect the {Date} and {Time} tokens from the line of text, I use the RegEx matching.

This will also say if a line of text is a new message or belongs to a multiline message.

Now that we detected the Date, Time and Author token, the remaining portion of the string (Message token) detects automatically.

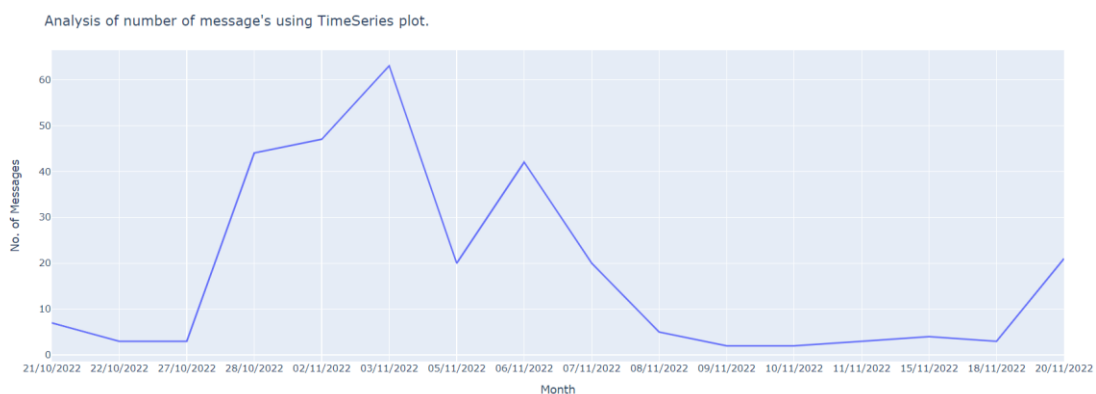
The Last step is that Parsing the entire file line by line and also handling the multiline texts. The below code will check if the line starts with a date, if not then it will be considered as a multiline text. The tokens will be extracted using the methods defined and store it in the list.

Now it's time to create a Data Frame using the Pandas module. The list "parsed data" consist of all the 40,000 messages from the WhatsApp exported file, parsed and ready to be stored in the Pandas Data Frame.

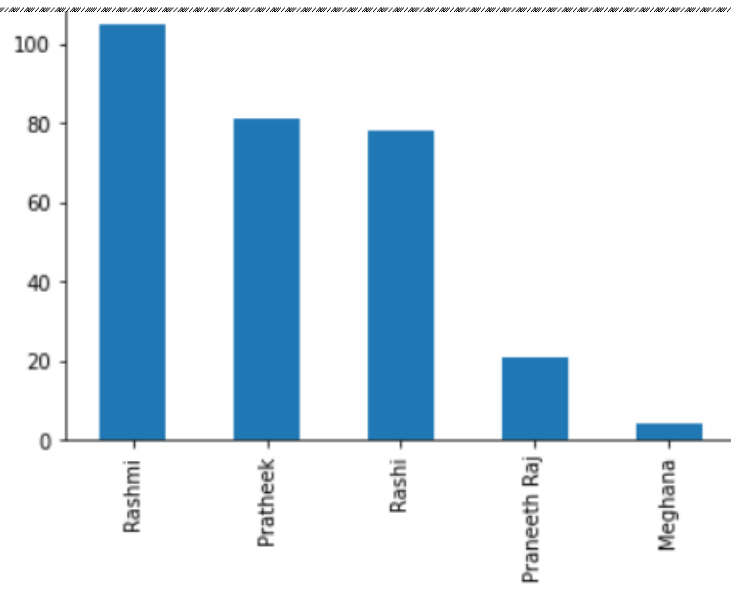
### 3.2.2. Dropping Messages With No Authors

While exporting all the messages from the chat, WhatsApp exports messages related to security changes, the one who left the group, the one who joined, changes in group name etc. These messages will also be exported, but with no authors (Null value). It is necessary to drop all the null value data points to proceed further with the analysis.

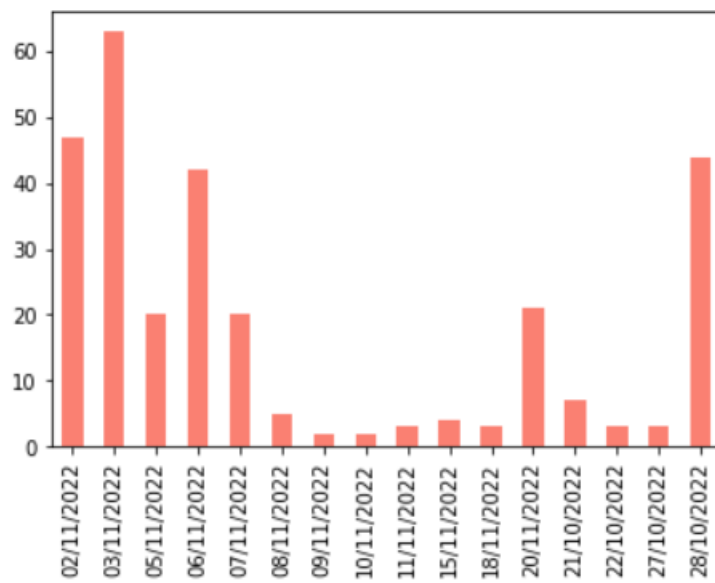
### 3.3. Analyzing Number of Messages Sent by the Group Members



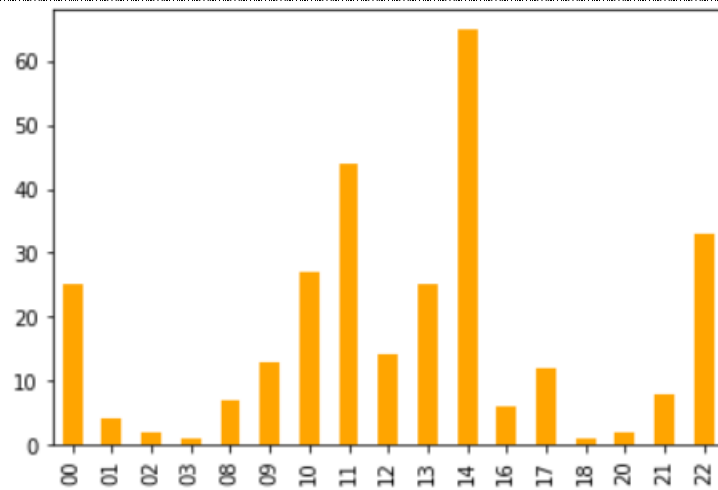
**Fig.3.1.** Analysis using Time Series Plot



**Fig.3.2.** No.of message sent by group members



**Fig.3.3.** Total Message Count



**Fig.3.4Active hours of group**

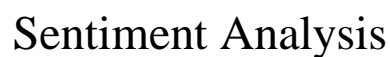
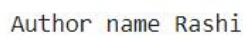
### 3.4. Finding the Letter and Word Count from Each Message

Going deep into the analysis, we extracted the number of letters and words used by a group member. We won't be needing the media values for this analysis, so we dropped it from the Data Frame.

Now it is time to split each line based on the separator tokens like commas (,), hyphens(-), colons(:) and spaces( ) so that the required tokens can be extracted and stored in a Data Frame.

## CHAPTER 4

## SAMPLE OUTPUT



```
[ ] x = sum(data["Positive"])
    y = sum(data["Negative"])
    z = sum(data["Neutral"])

    def sentiment_score(a, b, c):
        if (a>b) and (a>c):
            print("Positive 😊 ")
        elif (b>a) and (b>c):
            print("Negative 😞 ")
        else:
            print("Neutral 😐 ")
    sentiment_score(x, y, z)
```

Neutral 😐

## CHAPTER 5

### CONCLUSION

The major objective that has been decided in the initial phase of the requirement analysis is achieved successfully. After the implementation, the system provides reliable results. The system is totally menu and user friendly, which makes it easy for the users even with limited knowledge of computer environment to operate the developed system. The system avoids the drawbacks of the existing manual system and the validation facility of the system totally eliminates the chances of wrong data entry.

It has following features:

- User friendly.
- Time saving
- Runs on any devices
- Analyzes any WhatsApp imported file.
- Accuracy.
- Reliability.



# BIBLIOGRAPHY

- 1.D.Radha, R. Jayaparvathy, D. Yamini, “Analysis on Social Media Addiction using Data Mining Technique”, International Journal of Computer Applications
- 2.Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance.
- 3.Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic.