

WhatsApp Chat Analysing

Ayush Kedare
Sardar Patel Institute of Technology

Mumbai, India
ayush.kedare@spit.ac.in

Abstract— The WhatsApp Chat Analyzer is a Streamlit-based application for analyzing WhatsApp chat logs. It preprocesses chat data, extracting information like dates, times, users, and messages, then provides interactive visualizations for exploring chat activity. Key features include basic statistics on total messages, words, media, and links shared. Sentiment analysis classifies messages as positive, negative, or neutral. The application also offers K-means clustering for user profiling and an activity map to identify the most active months and days. Additional insights include the most active users, a word cloud of common words, and topic modeling using Latent Dirichlet Allocation (LDA) to discover key topics. This tool helps users understand communication patterns and chat group dynamics.

Keywords— streamlit, analysis, clustering.

I. INTRODUCTION

In the digital age, instant messaging platforms like WhatsApp have transformed how we communicate, allowing people to share messages, images, videos, and other media at a rapid pace. However, as the volume of communication grows, it can become challenging to identify trends and insights within chat data. To address this, the WhatsApp Chat Analyzer project aims to provide a comprehensive tool for analyzing WhatsApp chat logs, enabling users to gain valuable insights into their communication habits and the dynamics within their chat groups.

The WhatsApp Chat Analyzer provides users with a holistic view of their chat activity, highlighting patterns in messaging frequency, identifying sentiment trends, and uncovering the most active users and topics of discussion. Through interactive visualizations and charts, users can explore their chat data to understand communication patterns, track group engagement, and even profile users based on their activity and sentiment.

II. DATASET

The dataset used for the WhatsApp Chat Analyzer project typically consists of chat logs exported from WhatsApp. These logs are structured text files that contain messages exchanged within a WhatsApp chat group or between individual users. The dataset includes a variety of information that can be extracted and analyzed to gain insights into chat activity.

III. RESULT METRIC

1. Basic Statistics: Total messages: Number of messages exchanged in the chat. Total words: Total count of words used in the chat. Media shared: Number of media files (such as images, videos, etc.) shared in the chat. Links shared: Number of links shared in the chat.

2. Monthly and Daily Activity Timeline: Visualization of the number of messages exchanged over different months and days, providing insights into the chat's activity patterns.

3. Sentiment Analysis: Visualization of sentiment analysis results, showing the distribution of positive, negative, and neutral sentiments over time.

4. User Profiling: Clustering users based on their activity patterns, sentiment, and the duration of their participation in the chat.

5. Activity Map: Identification of the most active month and day within the chat, along with visualizations highlighting these periods.

6. Most Active Users: Identification of the users who contribute the most to the chat, both in terms of the number of messages and their percentage contribution.

7. Word Cloud: Visualization of the most common words used in the chat, providing insights into the topics or themes frequently discussed.

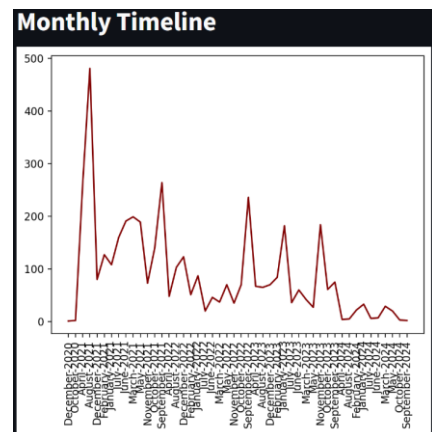
8. Topic Modeling: Identification of topics discussed in the chat through topic modeling, revealing key themes and their associated keywords.

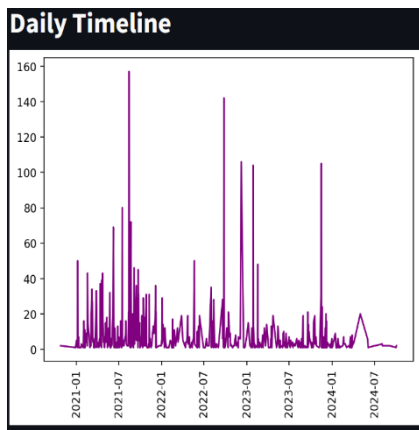
IV. DATA ANALYSIS

1. Overall Statistics

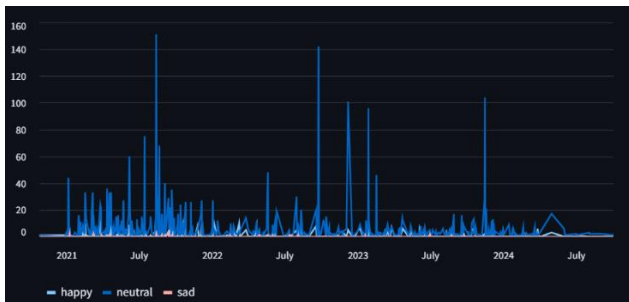
OverAll Basic Statistics			
Total Messages	Total Words	Media Shared	Link Shared
4286	7631	1709	72

2. Timeline



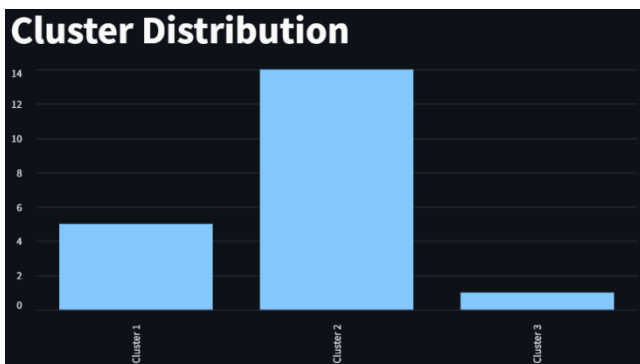


3. Sentiment Analysis

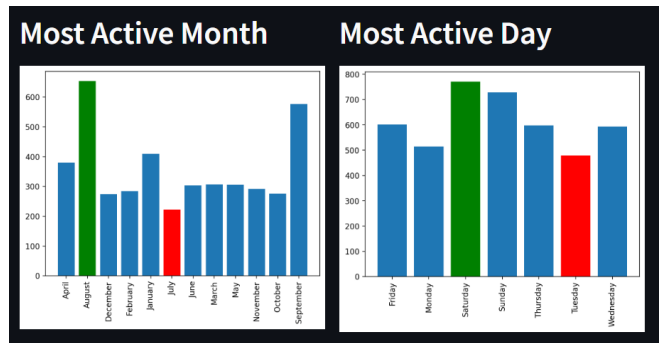


4. User Profiling & Cluster distribution

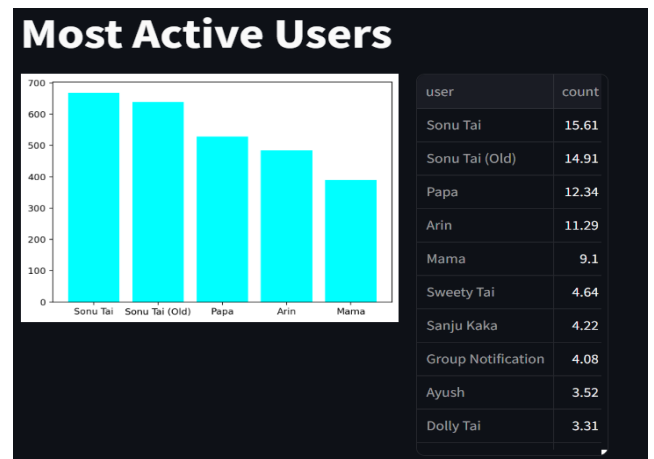
	user	msg	sentiment	date	cluster	sentiment_category
0	+91 77387 86715	53	0.0585	606	Cluster 2	neutral
1	+91 81157 81818	65	0.1754	954	Cluster 2	neutral
2	+91 84229 27985	81	0.0975	577	Cluster 2	neutral
3	+91 87675 00372	36	0.2056	990	Cluster 2	happy
4	+91 90283 40818	1	0.8	1	Cluster 3	happy
5	Aatu	50	0.1392	1,212	Cluster 2	neutral
6	Anushka	96	0.2288	811	Cluster 2	happy
7	Arin	484	0.031	1,212	Cluster 1	neutral
8	Ayush	151	0.2295	1,177	Cluster 2	happy
9	Dolly Tai	142	0.3113	1,248	Cluster 2	happy



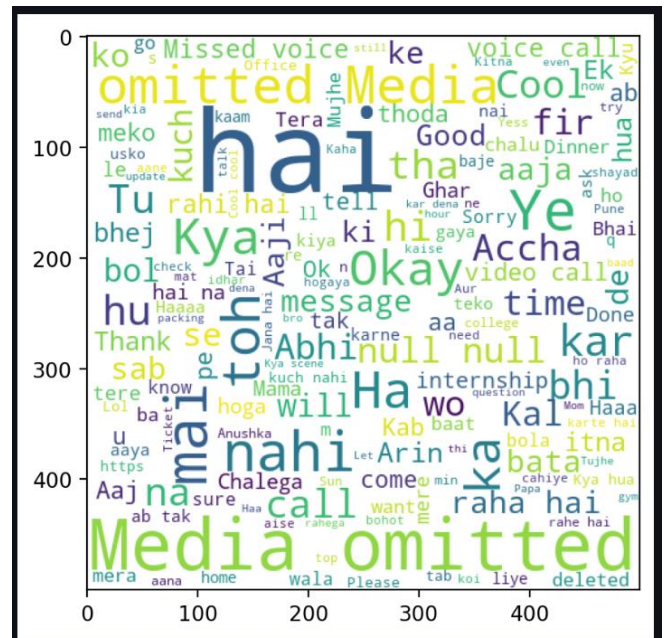
5. Activity Map



6. Active User



7. Most Common Words



8. Topic Modeling



V. CONCLUSION

- [1] The WhatsApp Chat Analyzer project demonstrates the potential of data analysis to unlock insights from everyday communication. By providing a comprehensive tool for analyzing WhatsApp chat logs, the project enables users to understand their chat groups' dynamics and their own communication patterns. From basic statistics to advanced sentiment analysis, user profiling, and topic modeling, the analyzer offers a wide range of features to explore chat data in depth.
- [2] Through this project, users can identify key trends, such as periods of high activity, the most active participants, and common topics of conversation. The sentiment analysis feature adds an additional layer of understanding by categorizing messages based on polarity, allowing users to gauge the general mood of the chat over time. The clustering and profiling features offer insights into the diversity of chat participants, highlighting patterns in behavior and sentiment among different groups of users.
- [3] Overall, the WhatsApp Chat Analyzer provides a powerful and user-friendly platform for extracting valuable information from chat logs. The project's success is driven by its ability to turn unstructured data into meaningful insights, offering interactive visualizations and charts

that make complex analyses accessible to a broader audience. As users explore their chat data, they gain a deeper understanding of their communication habits, which can inform personal reflections or guide improvements in group communication.

REFERENCES

1. R. P. Kumar and N. M. Naik, "Whatsapp Chat Analyzer Using Machine Learning and Natural Language Processing," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-6, doi: 10.1109/INCOFT60753.2023.10425363. keywords: {Freeware;Oral communication;Machine learning;Predictive models;Natural language processing;Internet telephony;Python;WhatsApp;Machine Learning;WhatsApp chat analyzer;Visualizations;Emotions component},
2. S. Yaqub, S. Gochhait, H. A. H. Khalid, S. N. Bukhari, A. Yaqub and M. Abubakr, "WhatsApp Chat Analysis: Unveiling Insights through Data Processing and Visualization Techniques," 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS), Manama, Bahrain, 2024, pp. 862-865, doi: 10.1109/ICETISIS61505.2024.10459604. keywords: {Freeware;Statistical analysis;Data preprocessing;Data visualization;Oral communication;Market research;Digital communication;Python;Data Processing;Visualization;Machine Learning;Statistical Analysis;Pattern Recognition;WhatsApp Chat},