Data warehousing and mining

MINI-PROJECT

TE CMPN A

WhatsApp Chat Analyzer

Submitted By

Sr. No	Roll No	Name
1	21102A0014	Deep Salunkhe
2	21102A0003	Omkar Patil
3	21102A0006	Pranav Redij

Under the Guidance of

Dr,. Kavita P Shirsat

Department of Computer Engineering

Vidyalankar Institute of Technology

Wadala(E), Mumbai 400 037

University of Mumbai

2023-24

ABSTRACT

Project Title:

WhatsApp Chat Analyzer

Project Aim:

To Visualize the WhatsApp chats of get insights from them

Project Description:

The project is titled "WhatsApp Chat Analyzer" and is developed using the Streamlit framework. It serves as a tool for analysing and visualizing WhatsApp chat data. The project offers a user-friendly interface to upload a WhatsApp chat data file and then provides various analytical features for users to explore the data.
Key Features of the Project:

1. **File Upload**: Users can upload their WhatsApp chat data in the form of a text file. The project supports the preprocessing and analysis of this data.
2. **Data Preprocessing**: The project preprocesses the uploaded chat data, extracting information such as user names, messages, dates, and other relevant details. It also includes handling group notifications.
3. **User Selection**: Users can choose to analyse the chat data for a specific user or get an overall analysis of the group chat.
4. **Analysis and Visualization**:
    - **Top Statistics**: The project displays key statistics, including the total number of messages, total words, media messages, and links shared.
    - **Most Active Users**: Users can view a list of the most active participants in the chat.
    - **Word Cloud**: A word cloud visualization is provided, highlighting frequently used words in the chat.
    - **Most Common Words**: The project identifies and displays the most common words used in the chat.
    - **Emoji Analysis**: Users can explore emoji usage, with a breakdown of the most frequently used emojis.
    - **Timeline Analysis**: The project generates monthly and daily timelines of messages, offering insights into message activity over time.
    - **Activity Maps**: Users can view activity maps, showing the busiest days and months.
    - **Message Count Over Time**: Users can select different time spans (daily, weekly, or monthly) to visualize message count trends.
    - **Message Length Distribution**: The project provides a histogram of message length distribution in terms of word count.

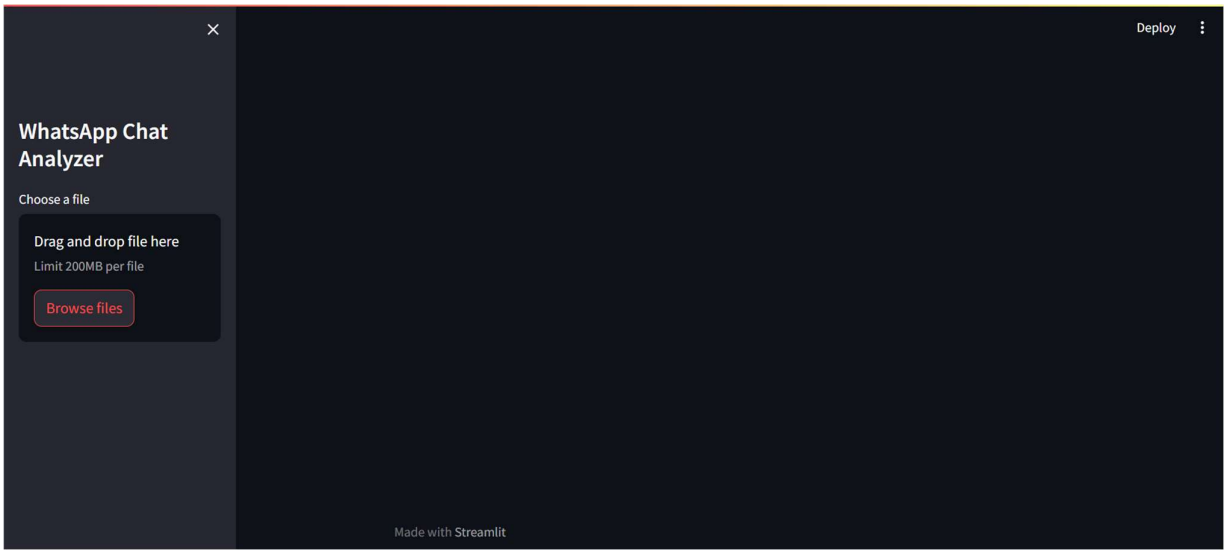
Description of Dataset:

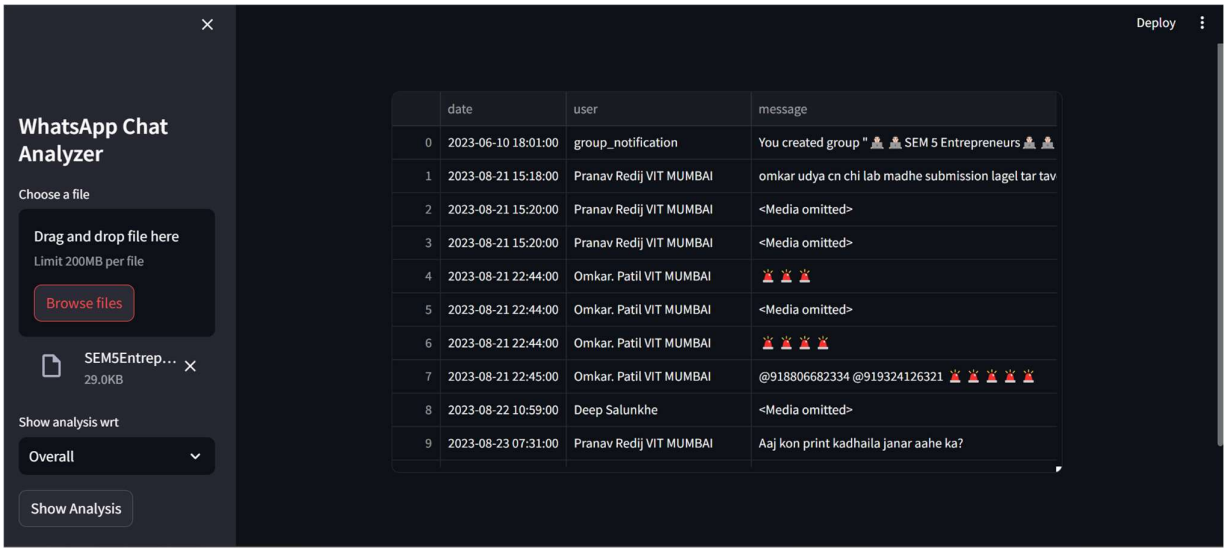
The dataset used for this project is a WhatsApp chat log. It typically consists of messages exchanged in a WhatsApp group or individual conversations. The dataset contains the following types of information:

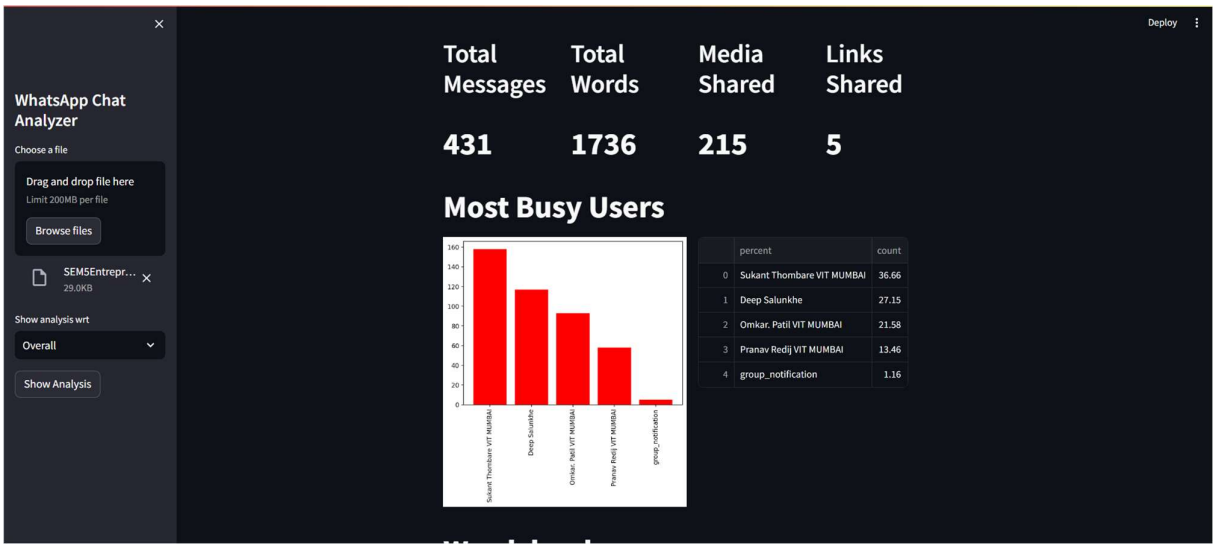
- **User Names**: The names of participants in the chat.
- **Messages**: The text of the messages sent within the chat.
- **Dates and Times**: Timestamps indicating when each message was sent.
- **Additional Date-Related Information**: The project extracts various date-related details such as the year, month, day, day of the week, hour, minute, and message period.

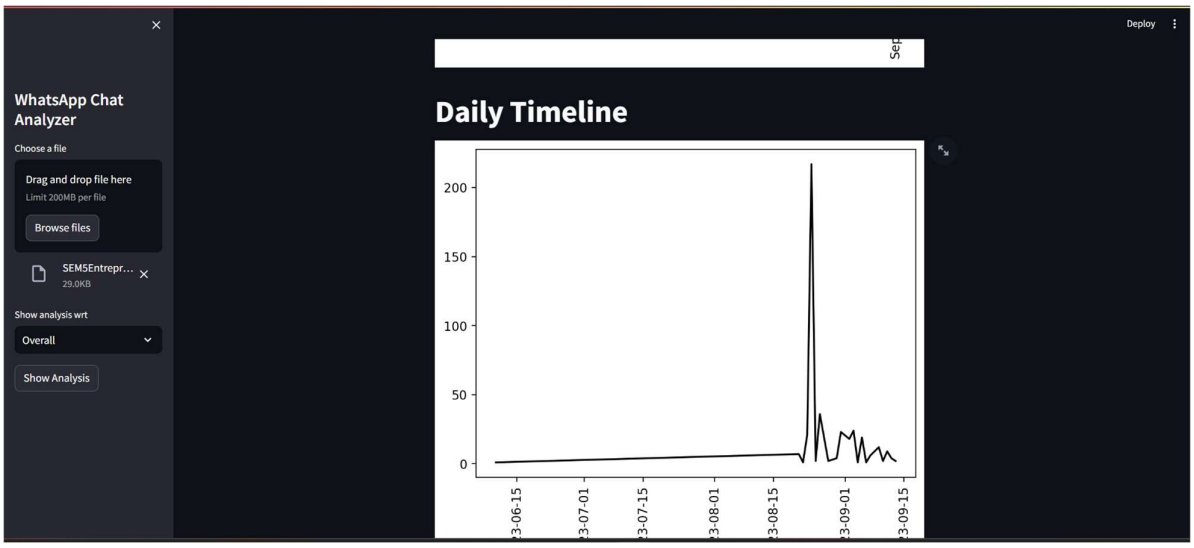
The dataset is preprocessed to extract, organize, and clean the information for analysis and visualization. It also handles group notification messages, ensuring that they are appropriately categorized.

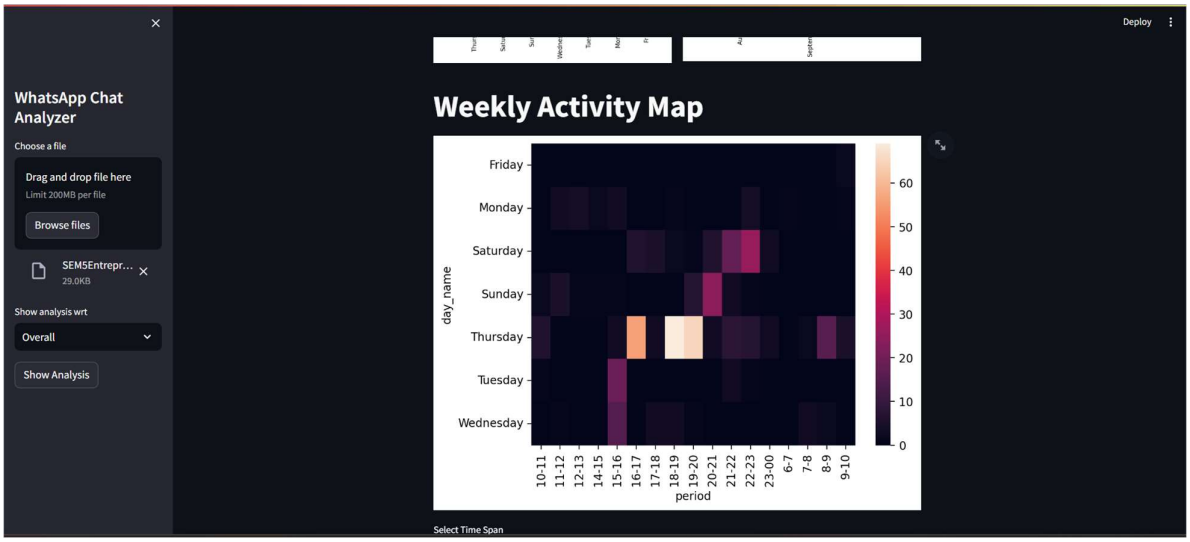
SCREENSHOTS:











CONCLUSION

This project offers a comprehensive way to gain insights into WhatsApp chat data, making it useful for personal chat analysis, group chat monitoring, or any scenario where understanding chat patterns and behavior is important.

The Project has been deployed at: https://dwm-mini-project-c6bhuu5h5xgdp6373hzzmx.streamlit.app/