**A**

**Project Report**

**On**

**"WHATSAPP CHAT ANALYSIS"**

**Prepared by**

19CE017-Preet Y Chokshi

**Under the guidance of**

Prof. Trusha Patel

Assistant Professor

A Report Submitted to

Charotar University of Science and Technology

for Partial Fulfillment of the Requirements for the

7th Semester Software Group Project-V (CS452)

**Submitted at**

****

**CE**

**CSPIT**

**At: Changa, Dist: Anand – 388421**

**December 2022**



**CERTIFICATE**

This is to certify that the report entitled “**WhatsApp chat analysis**” is a bonafied work carried out by **Preet Chokshi(19CE017)** under the guidance and supervision of **Assistant Prof. Trusha Patel** for the subject CS452 -**Software Group Project VII-** (CE) of 7th Semester of Bachelor of Technology in **CSPIT** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

|  |  |
| --- | --- |
| Prof. Trusha Patel  Assistant Professor  Branch Name(CE)  CSPIT, Changa, Gujarat. |  |
| Prof. Ritesh Patel  Head of Department – U & P.U PATEL DEPARTMENT OF TECHNOLOGY  CHARUSAT, Changa, Gujarat. | Dr. Y.P Kosta  IC. Principal,  DEPSTAR  CHARUSAT, Changa, Gujarat. |

**CHANDUBHAI PATEL INSTITUTE OF SCIENCE AND TECHNOLOGY At: Changa, Ta. Petlad, Dist. Anand, PIN: 388 421. Gujarat**

DECLARATION BY THE CANDIDATE

I hereby declare that the project report entitled “WhatsApp Chat Analytics” submitted by me to

CHANDUBHAI PATEL INSTITUTE OF SCIENCE AND TECHNOLOGY Changa in partial fulfilment of the requirement for the award of the degree of B.Tech in Computer Science Engineering, is a record of bonafide CS452 Software Project Major (project work) carried out by me under the guidance of Prof. Trusha Patel. I further declare that the work carried out and documented in this project report has not been submitted anywhere else either in part or in full and it is the original work, for the award of any other degree or diploma in this institute or any other institute or university.

Chokshi Preet(19CE017)

Prof. Trusha Patel

Assistant Professor,

CSPIT, CHARUSAT-Changa.

ABSTRACT

Data analytics has helped companies optimize and grow their performance for decades. Data analytics and visualization has aided us with several benefits, few of them being identifying emerging trends, studying relationships and patterns in data, analysis in depth and cherry on top are the insights we draw from these patterns. It is requirement of time that we study this concepts in thoroughly for all this benefits it provides.

The most used and efficient method of communication in recent times is an application called WhatsApp. WhatsApp chats consists of various kinds of conversations held among group of people. This chat consists of various topics. This information can provide lots of data for latest technologies such as machine learning. The most important thing for a machine learning models is to provide the right learning experience which is indirectly affected by the data that we provide to the model. This tool aims to provide in depth analysis of this data which is provided by WhatsApp. Irrespective of whichever topic the conversation is based our developed code can be applied to obtain a better understanding of the data.

The advantage of this tool is that is implemented using simple python modules such as pandas, matplotlib, seaborn and sentiment analysis which are used to create data frames and plot different graphs, where then it is displayed in the flutter application which is efficient and less resources consuming algorithm, therefor it can be easily applied to largest dataset.

ACKNOWLEDGEMENT

We, the developers of this project “WHATSAPP CHAT ANALYSIS”, with immense pleasure and commitment would like to present the project assignment. The development of this project has given me wide opportunity to think, implement and interact with various aspects of management skills as well as the new emerging technologies.

Every work that one completes successfully stands on the constant encouragement, good will and support of the people around. We hereby avail this opportunity to express we gratitude to number of people who extended their valuable time, full support and cooperation in developing the project.

We express deep sense of gratitude towards our Head of the CE Department, Prof. Ritesh Patel and project guide Prof. Trusha Patel for the support during the whole session of study and development. It is because of them, that we were prompted to do hard work, adopting new technologies.

We would also like to thank our mentor Mr.Trusha Patel for his guidelines throughout the development phase of the project. He helped me, whenever we were stuck in the program.

We are sincerely thankful to all who helped us complete the project in one way or the other. They altogether provided us favorable environment, and without them it would not have been possible to achieve our goal.

**TABLE OF CONTENTS**

Declaration...........................................................................................................................................ii

Abstract...............................................................................................................................................iii

Acknowledgement..............................................................................................................................iv

Table of contents..................................................................................................................................v

List of figures......................................................................................................................................vi

**Chapter 1 Project defination.............................................................................................................1**

WhatsApp chat analysis.......................................................................................................................2

Definition.............................................................................................................................................3

**Chapter 2 Description........................................................................................................................4**

Description...........................................................................................................................................5

Objective..............................................................................................................................................6

**Chapter 3 Software and Hardware Requirements.........................................................................7**

Tutorials...............................................................................................................................................8 **Chapter 4 Major Functionality….....................................................................................................9**

Functionality......................................................................................................................................10

Limitation...........................................................................................................................................11

**Chapter 5 System flowchart….…...................................................................................................13**

Uber flow chart..................................................................................................................................14

Uber context diagram.........................................................................................................................15

**Chapter 6 Screenshot of output….….............................................................................................16**

Output……........................................................................................................................................17

Output……........................................................................................................................................18

Output……........................................................................................................................................19

**Chapter 7 Conclusion......................................................................................................................20**

Conclusion.........................................................................................................................................21

**Chapter 8**  **References.....................................................................................................................22**

references...........................................................................................................................................23

**LIST OF FIGURES**

1. Chat Analysis(fig1.1) 3
2. Emoji Analysis (fig2.1) 6
3. Chat Analysis Chart(fig5.1) 13
4. Sentiment Analysis Flowchart(fig5.2) 13
5. Monthly Timeline(fig 6.1) 15
6. Statistics(fig6.2) 15
7. Daily Timeline(fig6.3) 16
8. Activity Map(fig6.4) 16
9. Most Busy Users(fig6.5) 17
10. Weekly Activity Map(fig6.6) 17
11. File Selection(fig6.7) 18
12. Wordcloud from Analysis(fig 7.1) 21

**CHAPTER 1**

**PROJECT DEFINATION**

**1.WHATSAPP CHAT ANALYSIS:**

This tool is based on data analysis and processing. The first step in implementing a machine learning algorithm is to understand the right learning experience from which the model starts improving on. Data pre-processing plays a major role when it comes to machine learning. In order to make the model more efficient we need lots of data, so we turned our focus primarily on one of the largescale data producers owned by Facebook which is nothing but WhatsApp.

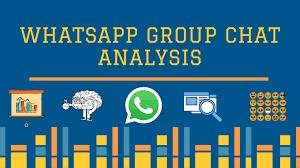
WhatsApp claims that nearly 55 billion messages are sent each day. The average user spends 195 minutes per week on WhatsApp, and is a member of plenty of groups. With this treasure house of data right under our very noses, it is but imperative that we embark on a mission to gain insights on the messages which our phones are forced to bear witness to.

WhatsApp chat Analyzer is an analyzing tool for the WhatsApp chats. The chat files can be exported from WhatsApp and it generates various plots and graphs showing, number of messages or emojis or images sent by a person, most active member in the group etc. It helps us to have a better understanding of our WhatsApp chats. This system is based on data analysis and pre-processing

**2.DEFINITION:**

WhatsApp-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, which another participant a user responds to the most. We propose to employ dataset manipulation techniques to have a better understanding of WhatsApp chat present in our phones.

The dataset of WhatsApp Group chat used for analysis is of 1 year (May, 2015 –May, 2016) which consists of 5,563 records in total and comprises of certain characteristics that define how much a particular person is using WhatsApp Chat Group, such as the years of usage, duration of usage in a day, the response levels, type of messages posted by each individual in the group (Smiley, Text, Multimedia), which age group people are more active and so on



Chat Analysis (fig1.1)

**CHAPTER 2**

**PROJECT DESCRIPTION**

**1.DESCREPTION:**

Data pre-processing, the initial part of the project is to understand implementation and usage of various python-built modules. The above process helps us to understand why different modules are helpful rather than implementing those functions from scratch by the developer. These various modules provide better code representation and user understandability. The following libraries are used such as numpy, scipy pandas, csv, sklearn, matplotlib, sys, re, emoji, nltk seaborn etc. Exploratory data analysis, first step in this to apply a sentiment analysis algorithm which provides positives negative and neutral part of th chat and is used to plot pie chart based on these parameters. To plot a line graph which shows author and message count of each date, to plot a line graph which shows author and message count of each author, Ordered graph of date vs message count, media sent by authors and their count, Display the message which is di not have authors, plot graph of hour vs message count.

**OBJECTIVE:**

1. Define:

Defining a business need is an important part of a business known as business analysis. This includes understanding and identifying the purpose of the organization while defining the direction used. In addition, you should take into account any relevant concerns regarding company success, problems, or challenges.

2. Prototype:

The users can train models from our web UI or from Python using our Data Science Workbench (DSW). At DSW, we support extensive deploying training of in-depth learning models in GPU clusters, tree models, and lines in CPU clusters, and in-level training on a wide variety of models using a wide range of Python tools available.

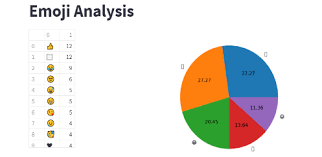
Finding the right combination of data, algorithms, and hyperparameters is a process of testing and self-replication. Going through this process quickly and effectively requires the automation of all tests and results.

3. Production:

Once the working model has been trained, it is important that the model builder is able to move the model to the storage or production area. In Michelangelo, users can submit models through our web UI for convenience or through our integration API with external automation tools. Deployed model is used to make predictions.

4. Measure:

Models are trained and initially tested against historical data. This means that users may not know that the model would work well in the past. But once you have used the model and used it to make predictions on new data, it is often difficult to make sure it is still working properly.

****

Emoji Analysis(fig 2.1)

**CHAPTER 3**

**SOFTWARE**

**AND**

**HARDWARE**

**REQUIREMENTS**

**1.Software and Hardware Requirements:**

* ***Python***
* ***Jupyter Notebook***
* ***Dataset from WhatsApp***
* ***Laptop / PC***
* ***Python:***

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented and functional programming.

* ***Jupyter Notebook:***

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

**CHAPTER 4**

**MAJOR FUNCTIONALITY**

**1.MAJOR FUNCTIONALITIES:**

**Contact-wise statistics:**

* Number of messages per contact
* Number of words per contact
* Number of average words per message per contact
* Number of emojis per contact
* The average number of emojis per message per contact
* Number of Media per contact
* Number of keywords per contact

**Time-wise statistics**

* Number of messages per hour
* Number of messages per weekday
* Number of messages per month
* Number of messages per year

**2.LIMITATION:**

TECHINAL FEASIBILTY

It is the measure of the specific technical solution and the availability of the technical resources and expertise. It is one of the first studies that must be conducted after tool has been identified. A technical study of feasibility is an assessment of the logistical aspects of business operation. This is considered with specifying equipment and software that will successfully satisfy the user requirement. The technical needs of the system may vary considerably but should include the facility to produce outputs in a given time, response time under certain conditions and the ability to process a certain amount of transaction at a certain speed.

The proposed system is developed by using Jupyter software. Jupyter is non-profit organization created to develop open-source software, open standards, and services for interactive computing across dozens of programming languages. The idea is to implement a data processing code using python to make better sense of WhatsApp group chat data.

OPERATIONAL FEASIBILTY

Operational feasibility is mainly concerned with issues like whether the system will be used if it is developed and implemented, whether there will be resistance from the users which will affect the possible application benefits. It is the ability to utilize, support and perform the necessary tasks of a system or program. It includes everyone who creates, operates or uses the system or program. It is the measure of how well a proposed system solves the problem and takes advantages of the opportunities identified during the scope definition and problem analysis phases. This system helps in many ways. It shows the number of users using WhatsApp and gives the data information of their sharing data. Which is organized in Pie-chart and Bar- chart.

ECONOMIC FEASIBILITY

Economic feasibility is the most frequently used method for evaluating the effectiveness of the new system. Economic feasibility is the measure of the cost effectiveness of an information system solution. Without a doubt, this measure is most often and important one of the three. Information systems are often viewed as capital investments for the business, and, as such should be subjected to the same type of investment analysis as other capital investments.

Economic analysis is used for evaluating the effectiveness of the proposed system. In economic feasibility, the most important is cost-benefit analysis. This project is not economical as it mainly depends on the sharing of data between two phones.

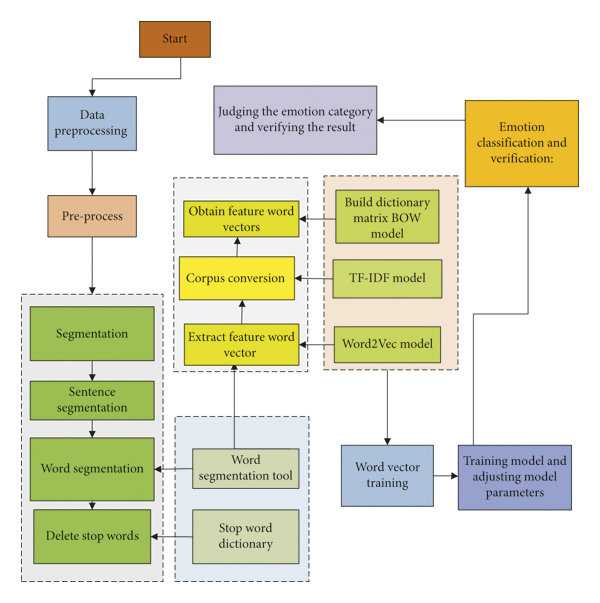
**CHAPTER 5**

**SYSTEM FLOW CHART**

**SYSTEM CHART:**

****

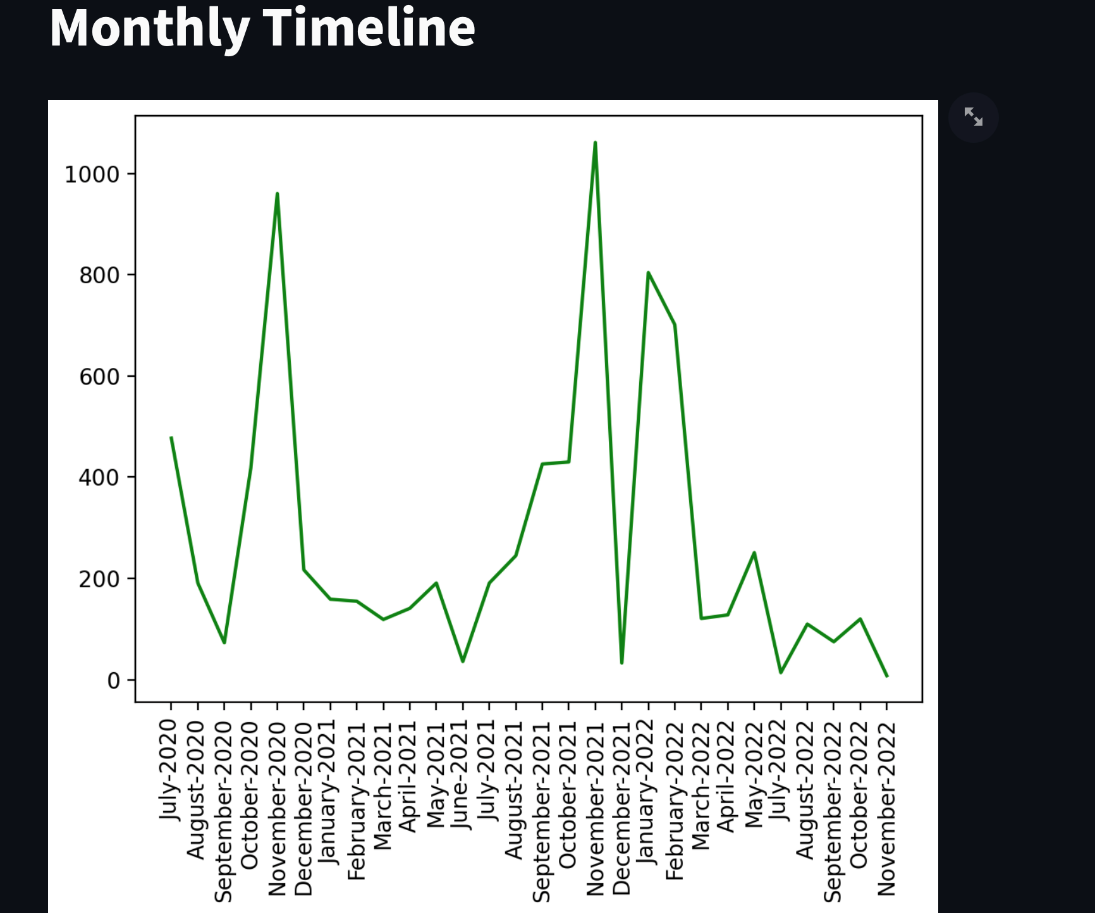
Chat analysis chart(Fig5.1)

****

Sentiment Analysis Flowchart(Fig5.2)

**CHAPTER 6**

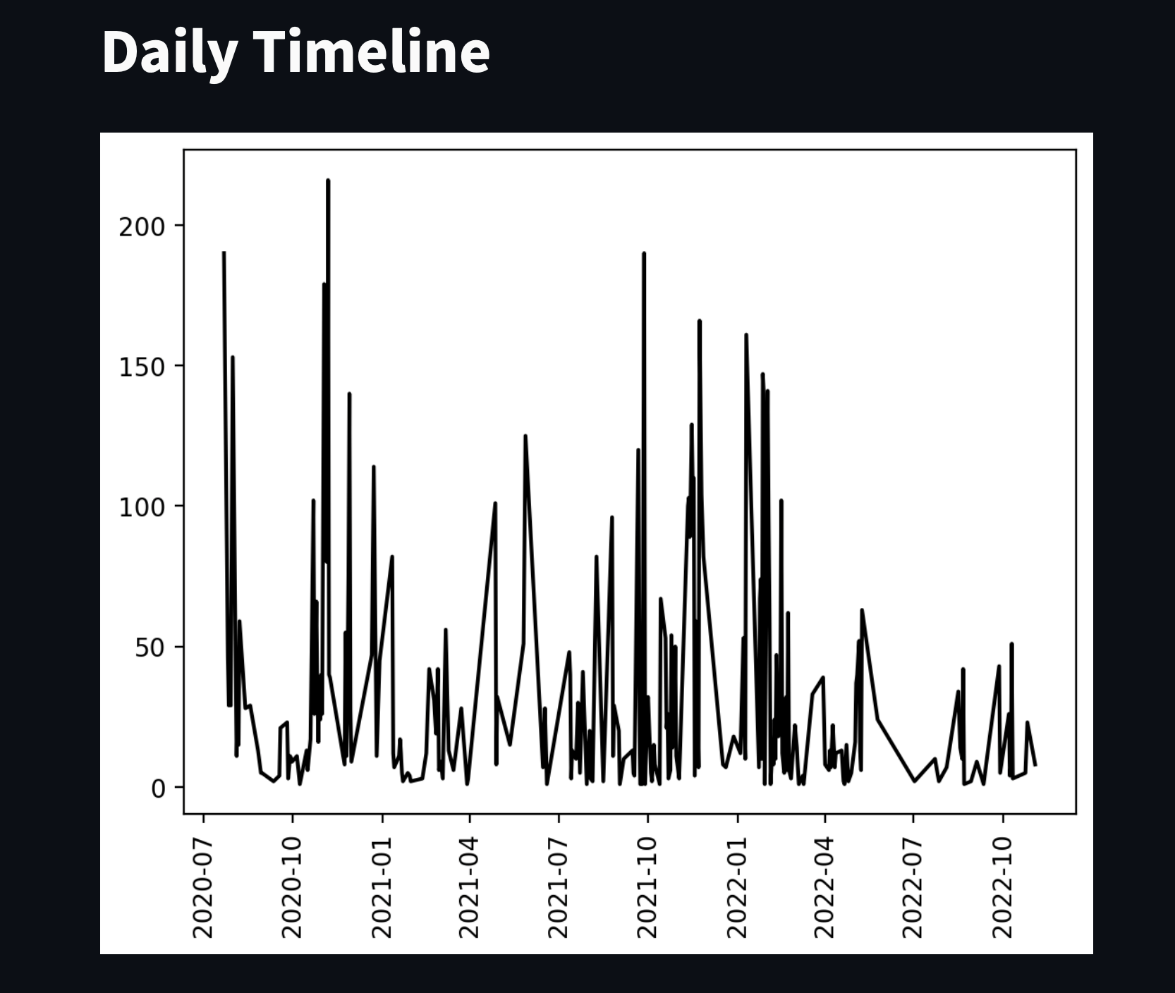
**SCREENSHOTS OF YOUR PROJECT OUTPUT**



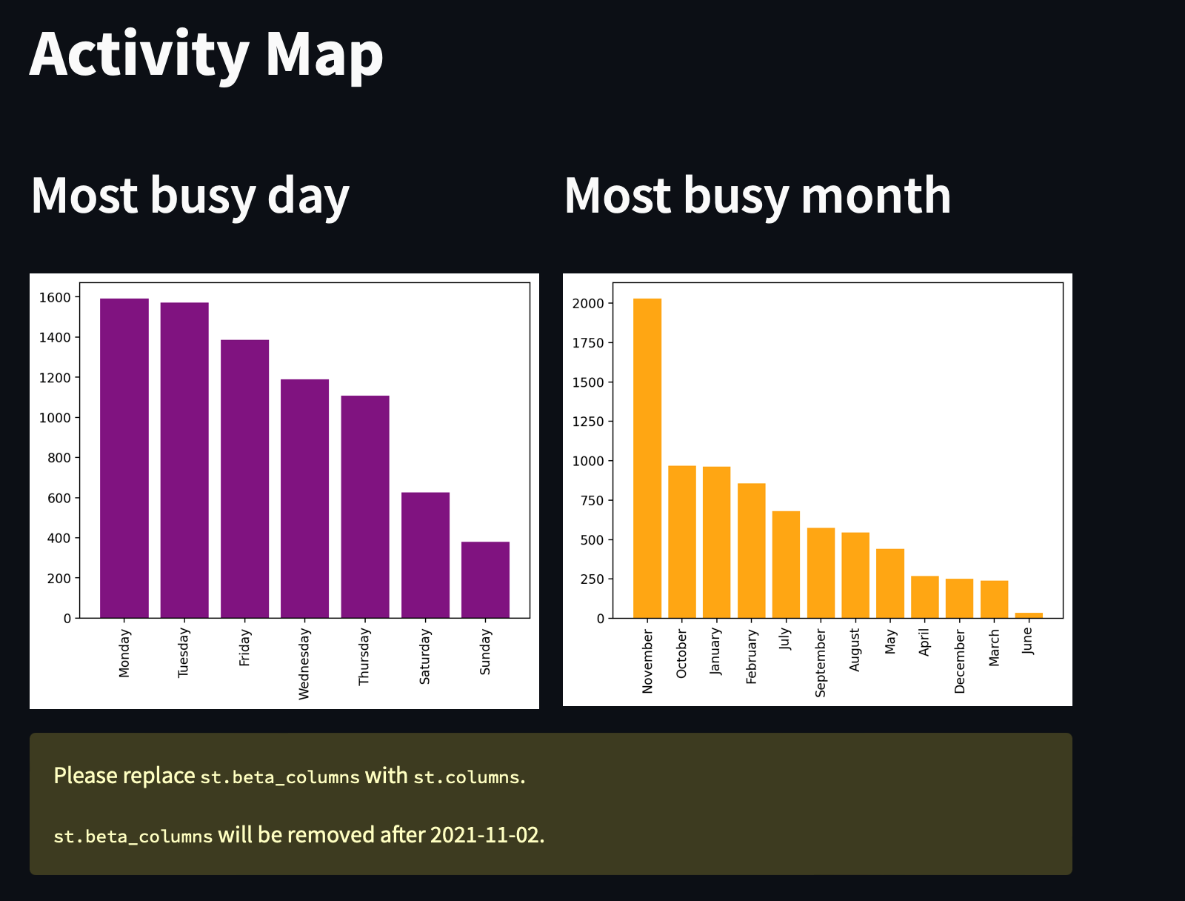
Monthly timeline(fig 6.1)



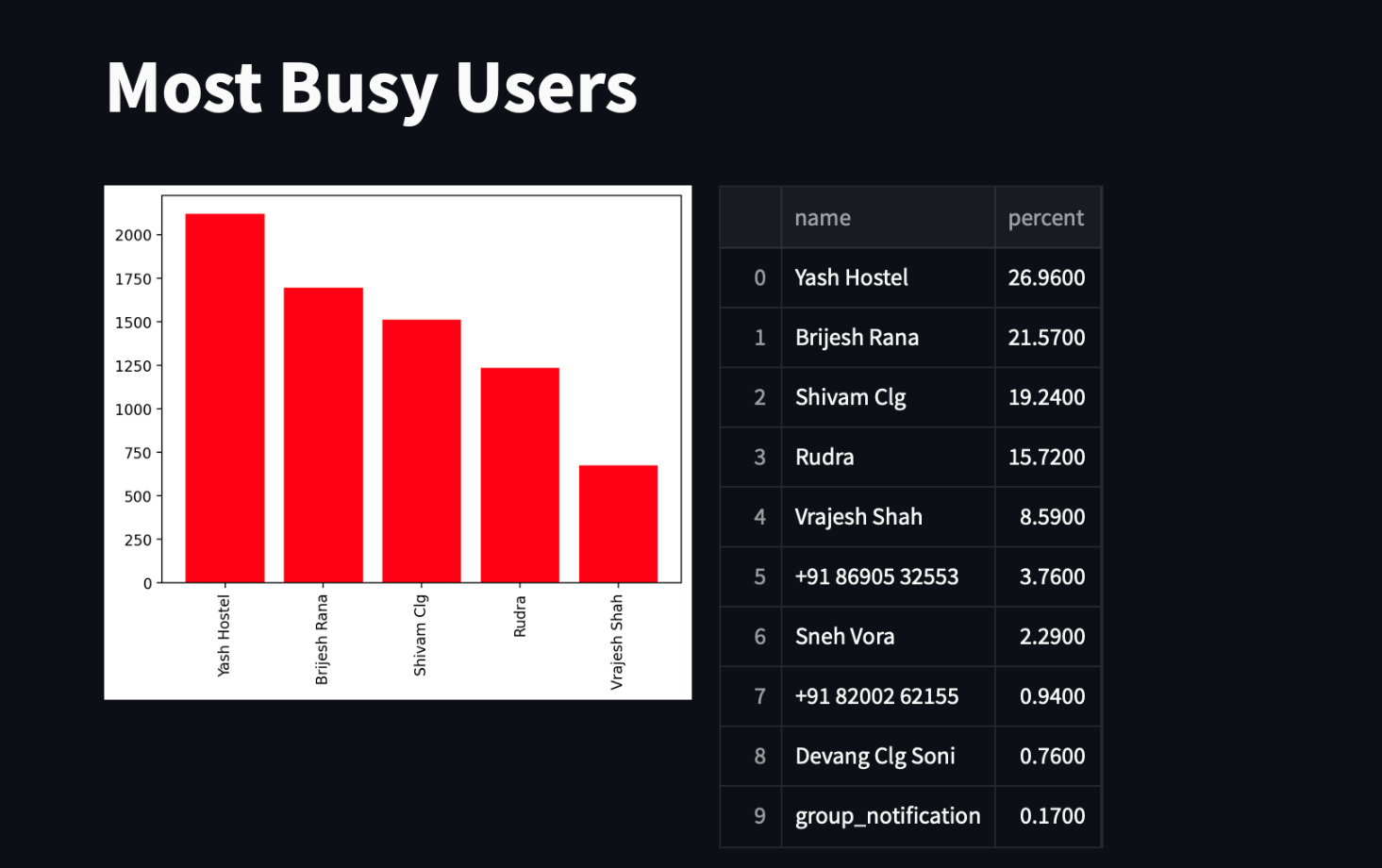
Statistics(fig 6.2)



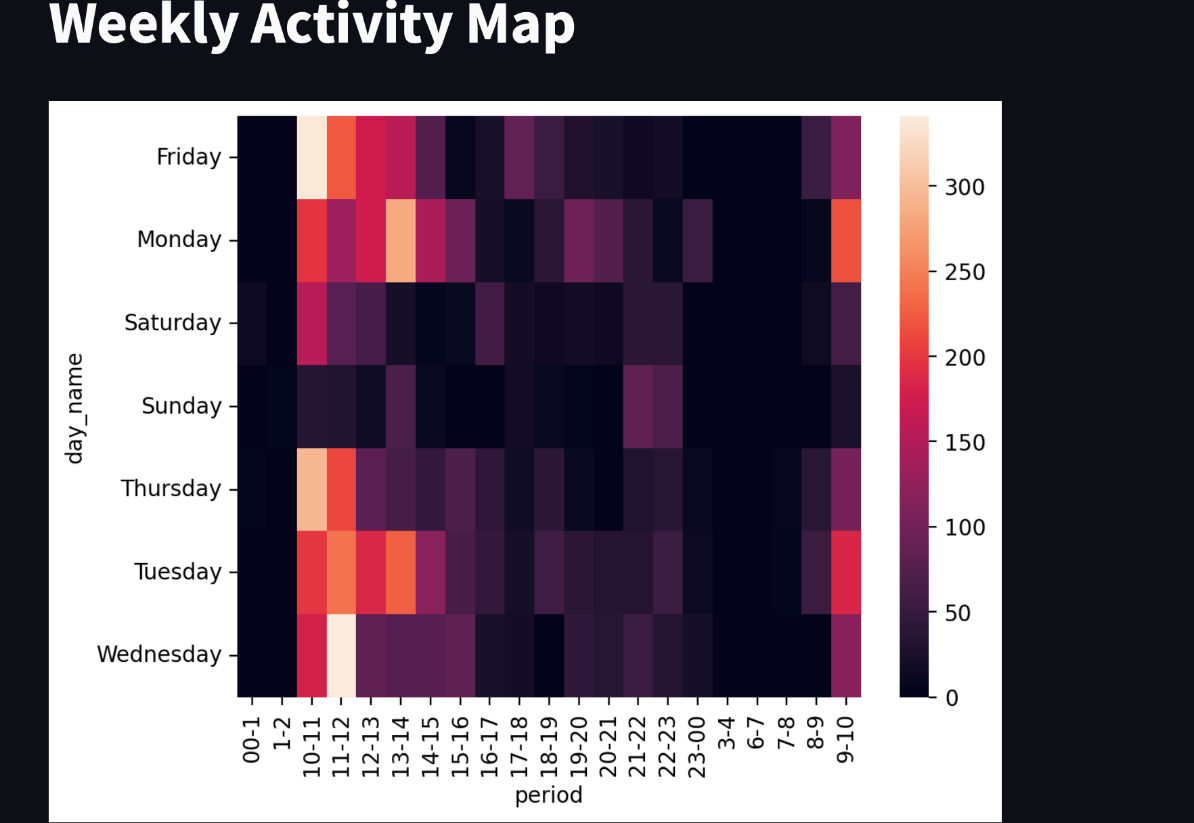
Daily timeline(fig 6.3)



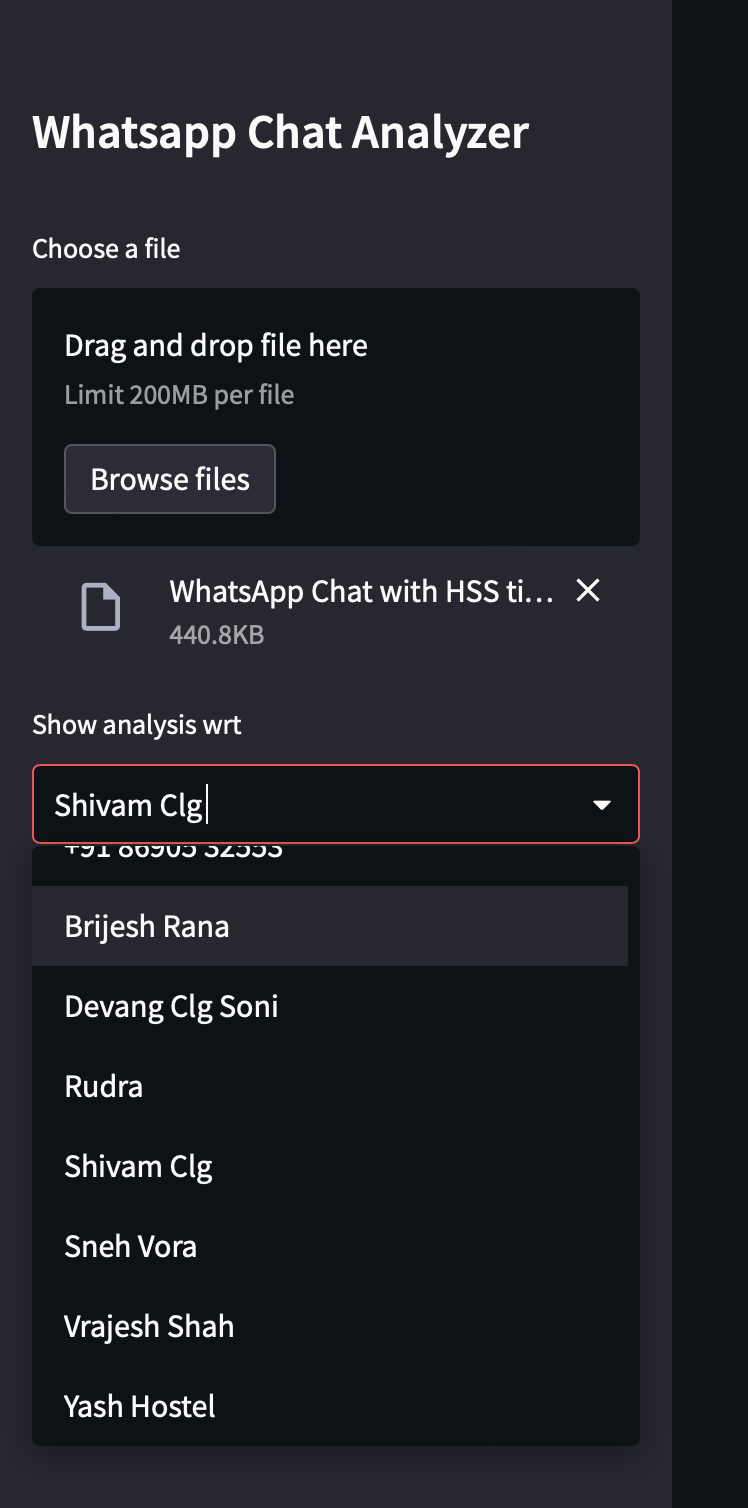
Activity Map (fig 6.4)



Most Busy Users(fig 6.5)



Weekly Activity Map (fig 6.6)



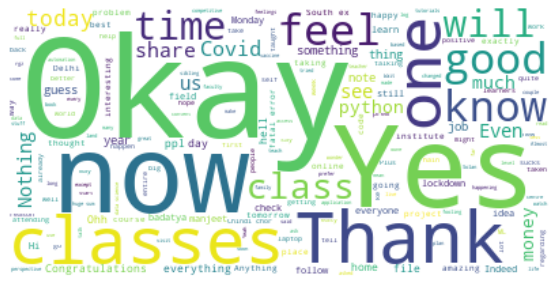
File Selection(fig 6.7)

**CHAPTER 7**

**CONCLUSION**

**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 work was able to discuss the WhatsApp application and its libraries, to create an analysis of a WhatsApp group chat and visually represent the top 10 and top 20 users in the chat groups. A pseudocode of the plot was given and at the end, visual representation of the plot was implemented. Also, an analysis of the top 10 and top 20 users were done. The system was done with python, and the python libraries that were implemented includes, NumPy, Pandas, Matplotlib and Seaborn. At the end of the work expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group. On serious note this system has the ability to analyze any WhatsApp group data input into it.



Wordcloud from Analysis(fig 7.1)

**CHAPTER 7**

**REFERENCES**

**REFERENCES:**

[1] Available from: http://www. statista.com/statistics/260819/number-of-monthly-active-WhatsApp-users. Number of monthly active WhatsApp users worldwide from April 2013 to February 2016(in millions).

[2] Ahmed, I., Fiaz, T., “Mobile phone to youngsters: Necessity or addiction”, African Journal of Business Management Vol.5 (32), pp. 12512-12519, Aijaz, K. (2011).

[3] Aharony, N., T., G., The Importance of the WhatsApp Family Group: An Exploratory Analysis. “Aslib Journal of Information Management, Vol. 68, Issue 2, pp.1-37” (2016).

[4]Access Data Corporation. FTK Imager, 2013. Available at <http://www.accessdata.com/support/product-downloads>.

[5] D.Radha, R. Jayaparvathy, D. Yamini, “Analysis on Social Media Addiction using Data Mining Technique”, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.7, pp. 23-26, April 2016.

[6] Jessica Ho, Ping Ji, Weifang Chen, Raymond Hsieh, “Identifying google talk”, IEEE International Conference on Intelligence and Security Informatics, ISI ‘09, pp. 285-290, 2009.

[7] Mike Dickson, “An examination into AOL instant messenger 5.5 contact identification.”, Digital Investigation, ScienceDirect, vol. 3, issue 4, pp. 227-237, 2006.

[8] Mike Dickson, “An examination into yahoo messenger 7.0 contact identification”, Digital Investigation, ScienceDirect, vol. 3, issue 3, pp. 159-165, 2006.