

**SENTIMENTAL ANALYSIS ON TWITTER DATA FOR SWACHH BHARAT**

**PROJECT REPORT**

***Submitted by***

**NEERAJ. B - E0320031**

***In partial fulfilment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Artificial Intelligence and Data Analytics)**

**Sri Ramachandra Engineering and Technology**

**Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116**

**MARCH, 2022**



**SENTIMENTAL ANALYSIS ON TWITTER DATA FOR SWACHH BHARAT**

**PROJECT REPORT**

***Submitted by***

**NEERAJ. B - E0320031**

***In partial fulfilment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Artificial Intelligence and Data Analytics)**

**Sri Ramachandra Engineering and Technology**

**Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai -600116**

**MARCH, 2021**



**BONAFIDE CERTIFICATE**

Certified that this project report “**SENTIMENTAL ANALYSIS ON TWITTER DATA FOR SWACHH BHARAT**”is the bonafide work of **NEERAJ. B - E0320031**who carried out the internship work under my supervision.

**Signature of Faculty Mentor Signature of Vice-Principal**

|  |  |
| --- | --- |
| **Prof. Ashokkumar P**  Assistant Professor  Sri Ramachandra Engineering and Technology  Porur  Chennai-600116 | **Prof. M. Prema**  Vice-Principal  Sri Ramachandra Engineering and Technology  Porur  Chennai-600116 |

**Evaluation Date:**

|  |  |
| --- | --- |
| **TABLE OF CONTENTS**  **Title** | **Page** |
| 1. Domain Introduction   * 1. Data Analytics   2. Objective   3. Problem Statement | 7  7  7  7 |
| 2. Project Description  2.1 Sentiment Analysis  2.2 Data Visualization | 8  8  8 |
| 3.Literature Survey | 9 |
| 4. Methodology  4.1 Development Steps[Workflow] | 10  10 |
| 5. Components Used  5.1 Python  5.2 System Requirement | 11  11  11 |
| 6. Technology Used   6.1 Jupyter Notebook | 12  12 |
| 7. Implementation  7.1 Code and Output Screenshots | 13  13 |
| 8. Result  8.1 Conclusion  8.2 Future Scope | 35  35  35 |
| 9. References | 35 |
| 10. Worklog | 36 |
| 11. Offer Letter | 37 |

**ACKNOWLEDGEMENT**

I express my sincere gratitude to our Chancellor, Vice-Chancellor and our sincere gratitude to our Provost **Dr. V.Raju** and our Vice-Principal **Prof. M.Prema** for their support and for providing the required facilities for carrying out this study.

I wish to thank my faculty supervisor(s), **Prof. Ashokkumar P,** Department of Computer Science and Engineering, Sri Ramachandra faculty of Engineering and Technology and **Mr. Abhishek Kumar Jha, Co-founder of** **Utkarshini Edutech** for extending help and encouragement throughout the project. Without his/her continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to all the members of Sri Ramachandra Faculty of Engineering and Technology, my beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

**ABSTRACT**

The project two accomplished during my internship at Utkarshini Edutech Pvt Lmtd from Feb to Mar, is in the field of data science, and it involves choosing a hashtag and collecting data related to Swachh Bharat from Twitter, cleaning the data, and then implementing a concept called Sentimental Analysis on the topic. For the second project , the interpreted solution was used to visualise the output using visualisation tools like as graphs, charts, and various other tools - Sentimental Analysis on Twitter Data for Swachh Bharat.

1. **DOMAIN INTRODUCTION**

**1.1 Data Analytics**

* Data analytics is the process of employing sophisticated computer systems to extract meaning from raw data. To draw inferences and detect trends, these systems transform, organise, and model data.
* The science of data analysis using algorithms, statistics, and technology is known as data analytics. Data analytics focuses on using programmes, data, and computational tools to explore and discover relevant insights in large amounts of data.

**1.2 Objective**

The Main objective is to find and Analyze the sentiment for Swachh Bharat data.

* Plot and Prepare a frequency chart for the hashtags.
* Plot the Donut chart for Sentiment Analysis.
* Sentiment Analysis Plot for period of every six months in Year.
* Word cloud for Sentiment Analysis in the 5 different time period.
* Frequency chart for all the tweets in time period between September and October months for every year.
* Analyzing the Climate based problem in swatch Bharat data.

**1.3 Problem Statement**

The problem is to scrape the Real-Time Tweets from Twitter and implementing the sentiment Analysis on the topic of Swachh Bharat Mission. And analyzing the same using various factors which affected Swachh Bharat Mission and how has it changed over the years.

**2. PROJECT DESCRIPTION**

**2.1 Sentiment Analysis**

* We'll use VADER Sentiment Analysis, which stands for Valence Aware Dictionary and sEntiment Reasoner, for Sentiment Analysis. VADER is a sentiment analytical technique that detects suppositions expressed in web-based media using a lexicon and criteria.
* The method of utilising algorithms to classify numerous samples of linked text into overall positive, negative, and neutral categories is known as sentiment analysis. Sentiment Analysis is used to evaluate whether the general public has a good, negative, or neutral attitude toward the subject of interest.

**2.2 Data Visualization**

* Data visualization is the graphical depiction of data in order to communicate insights to clients, consumers, and stakeholders in general in an interactive and efficient manner.
* Data visualization's major purpose is to make finding patterns, trends, and outliers in massive data sets easier.

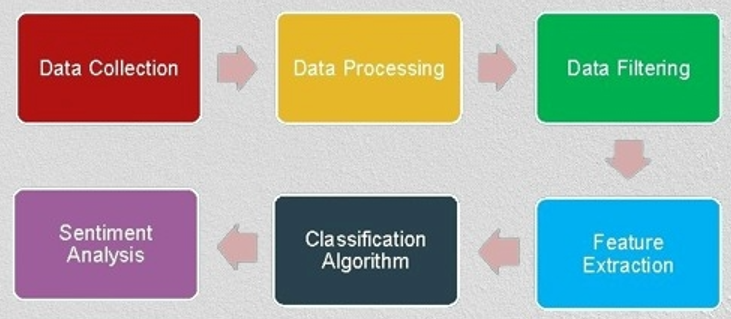
**3. LITERATURE SURVEY**

* **Sentimental Analysis on voice using AWS Comprehend:**
* **Link:** <https://sci-hub.hkvisa.net/10.1109/ICCCI48352.2020.9104105>
* In this project, the author will use Amazon Comprehend to do sentimental analysis. Amazon Comprehend is a natural language processing (NLP) service that extracts document information using machine learning. This service can extract unstructured data such as photos, audio, and so on.
* **An Aspect of Sentiment Analysis: Sentimental Noun with Dual Sentimental Words Analysis:**
* **Link:**<https://sci-hub.hkvisa.net/10.1109/Confluence51648.2021.9377048>
* This research looks at people's emotions to see if they are good or negative in this situation. The polarity technique is used in the paper to identify whether an opinion is positive, negative, or impartial. Three basic keywords are used to assess polarity: "COVID," "Corona virus," and "COVID-19."
* **Real Time Sentimental Analysis on Twitter:**
* **Link:**https://thesai.org/Downloads/Volume10No2/Paper\_48-A\_Study\_on\_Sentiment\_Analysis\_Techniques.pdf
* Sentiment analysis is a typical debate preparation exercise that seeks to uncover the attitudes that underpin ideas in a variety of texts. Opinion research using Twitter data has gotten a lot of attention in the previous decade, and it entails analysing "tweets" and the substance of these statements. As a result, the results of various sentiment analyses applied to Twitter data are examined in this research.

**4. METHODOLOGY**

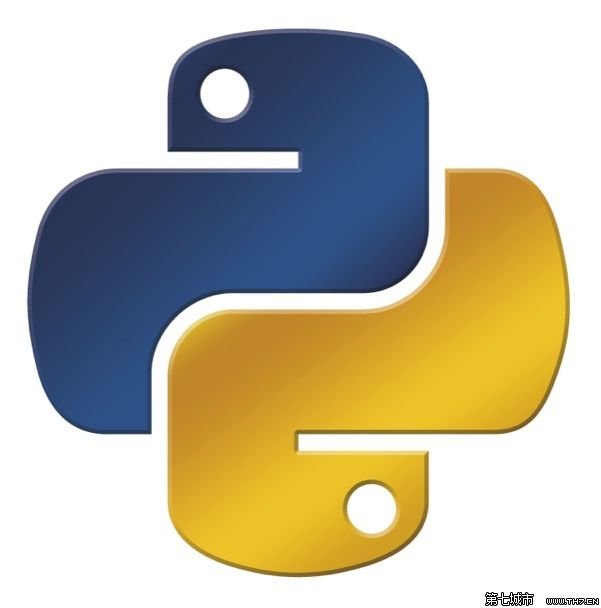
**4.1 Development Steps**

* **Data Collection:** Data relating to the hashtag Swachh Bharat was gathered using Twitter's snscrape.
* **Data Understanding:** Gathering all of the information, processing it, studying the data of column, and applying it to uncover patterns.
* **Data Analysis:** The data was evaluated using several analytical methodologies such as correlation, covariance, mean, and others to find hidden patterns.
* **Sentiment Analysis:** The quantity of positive and negative terms in a particular text is examined and counted.
* **Data Visualization:** The evaluated result was shown to make the hidden links and patterns in the data easier to understand.
* **Derive Conclusions:** Using the visualizations and conclusions, the basic facts of the data set were deduced.

****

**5. COMPONENTS USED**

**5.1 Python**

* Python is a versatile programming language that can be used for a wide range of tasks. A straightforward and intuitive language that performs as well as its primary competitors.
* A simple and natural language that is just as powerful as the major contenders.
* Because it is open source, anyone can contribute to its development. code that is almost as comprehensible as plain English.
* Appropriate for running errands on a regular basis, given the short time it takes to improve.

**5.2 System Requirements**

* Operating System: Windows 7 or later (64 bit).
* Disk Space: 1.32 GB (does not include disk space for IDE or tools).
* Tools: Windows PowerShell 5.0 or newer (it is pre-installed with windows 10)

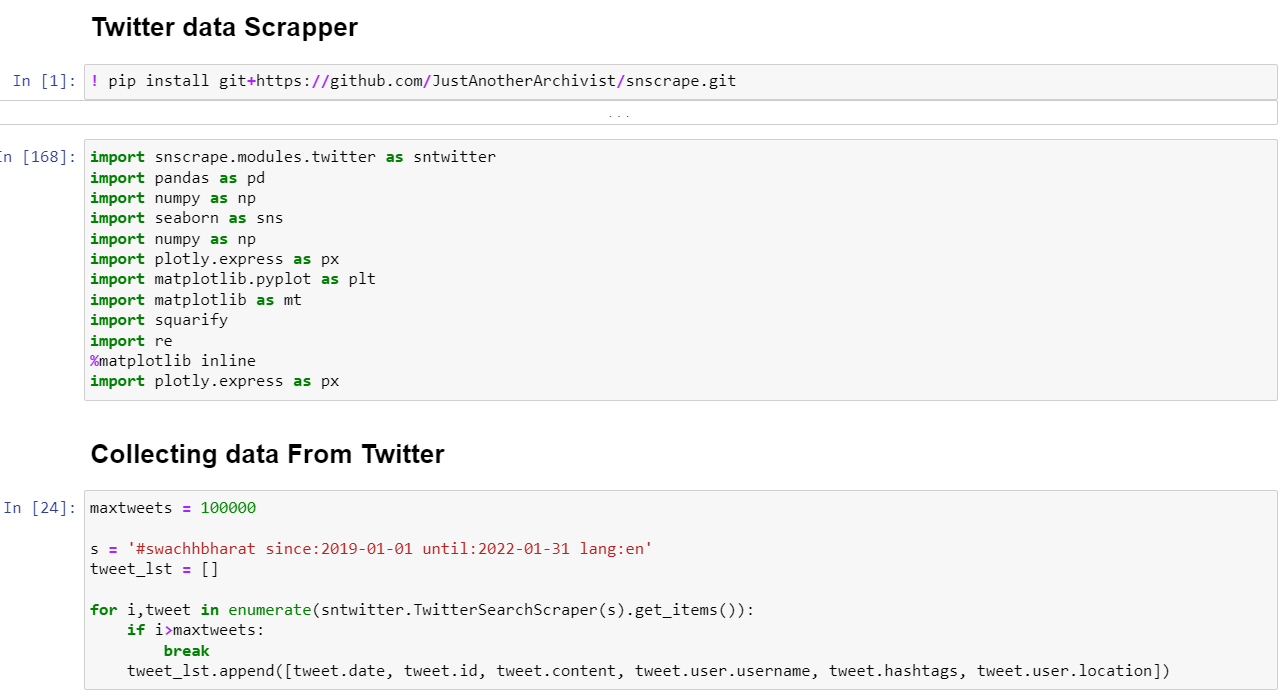
**6.TECHNOLOGY USED**

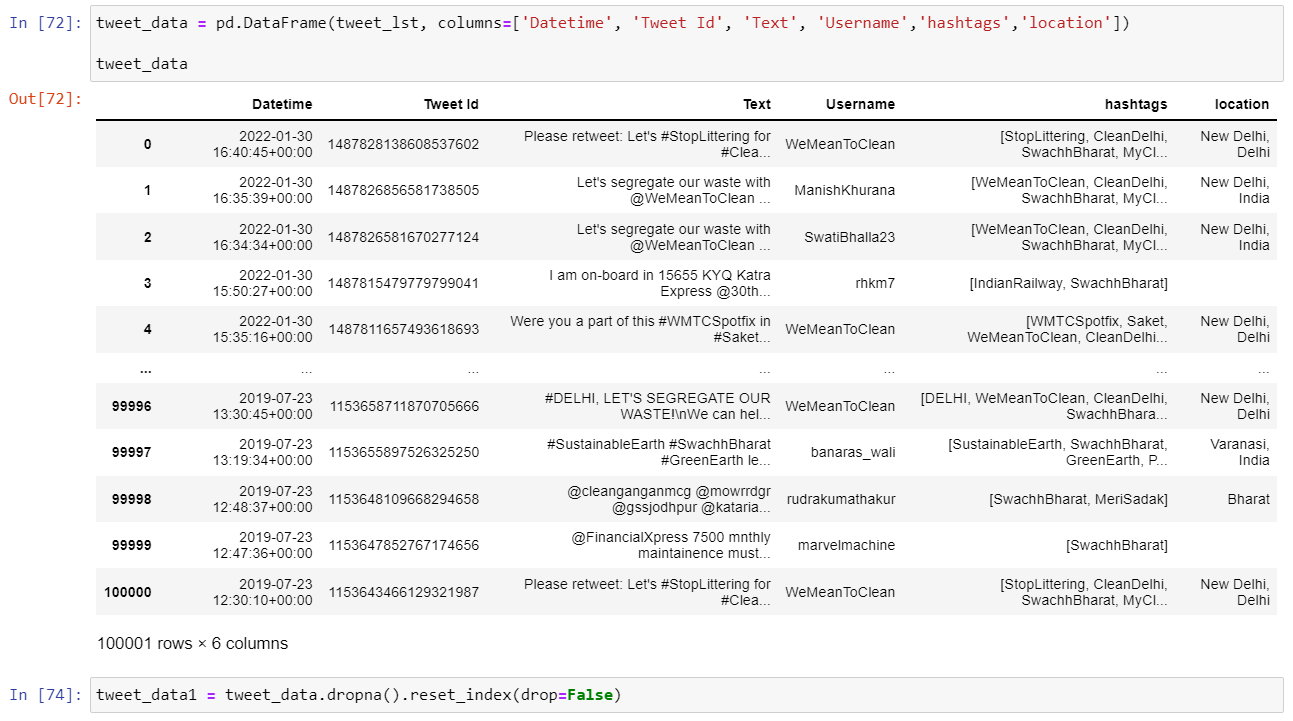
**6.1 Jupyter Notebook**

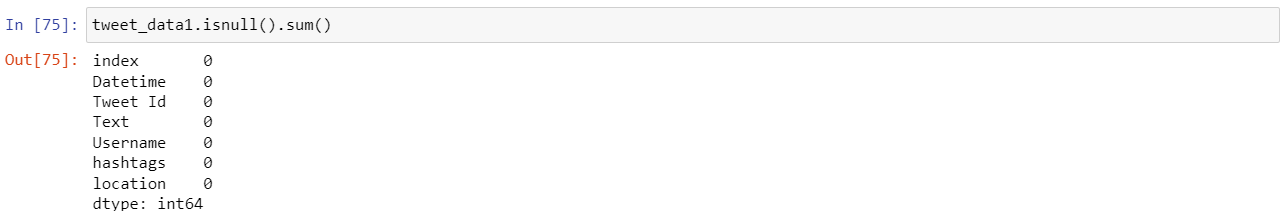
* ****Jupyter Notebook is a free, open-source web software that lets you create and share documents with live code, equations, visualisations, and narrative text. It can be used for a range of activities, including data cleansing and transformation, numerical simulation, statistical modelling, data visualisation, and machine learning.

**7. IMPLEMENTATION**

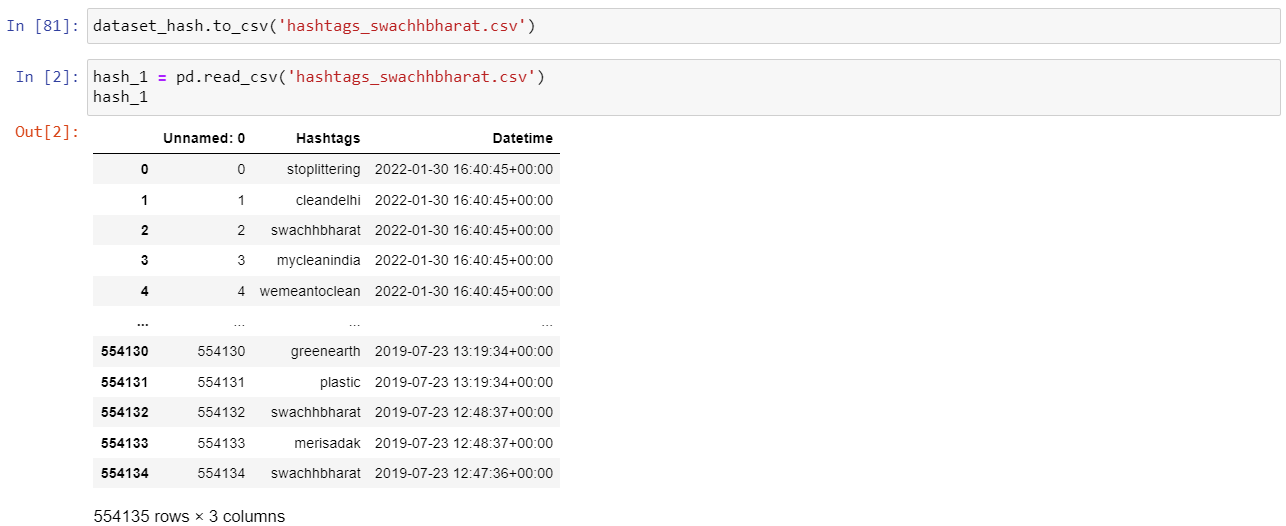
**7.1 Code and Output Screenshot**

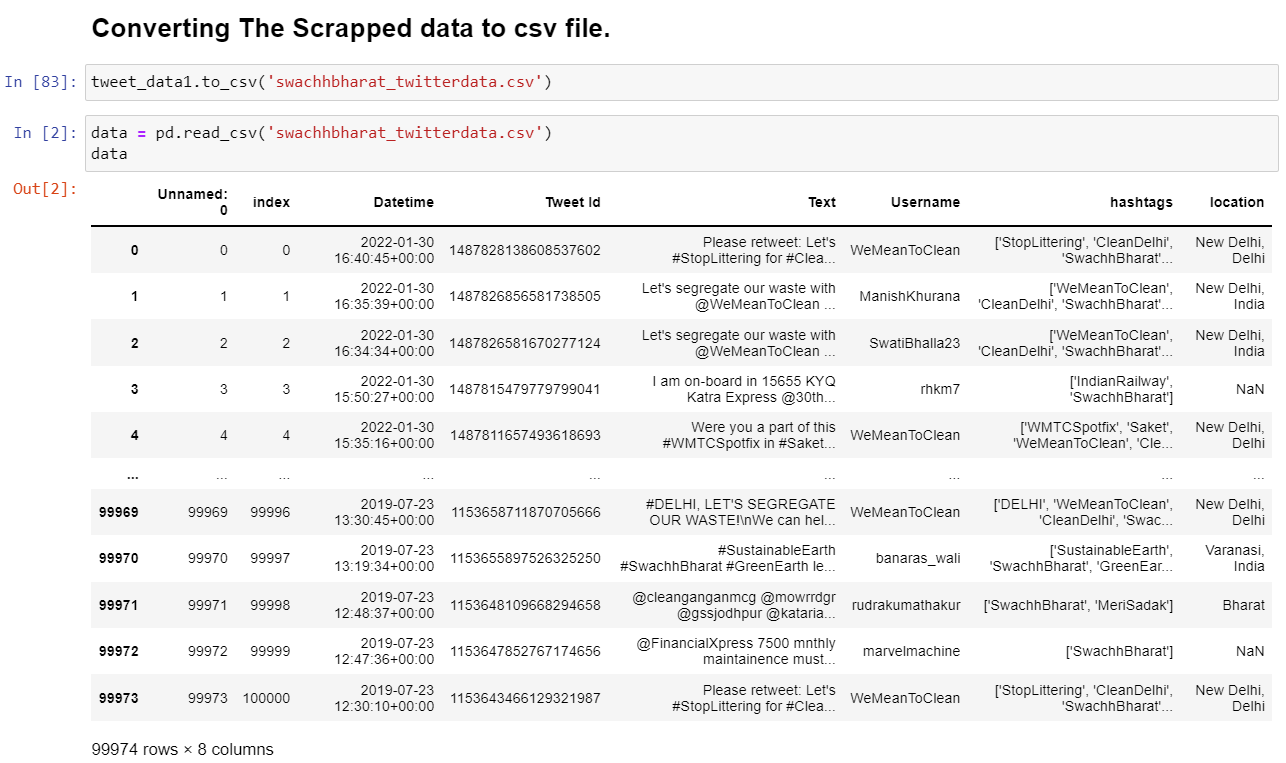
****

****

****

****

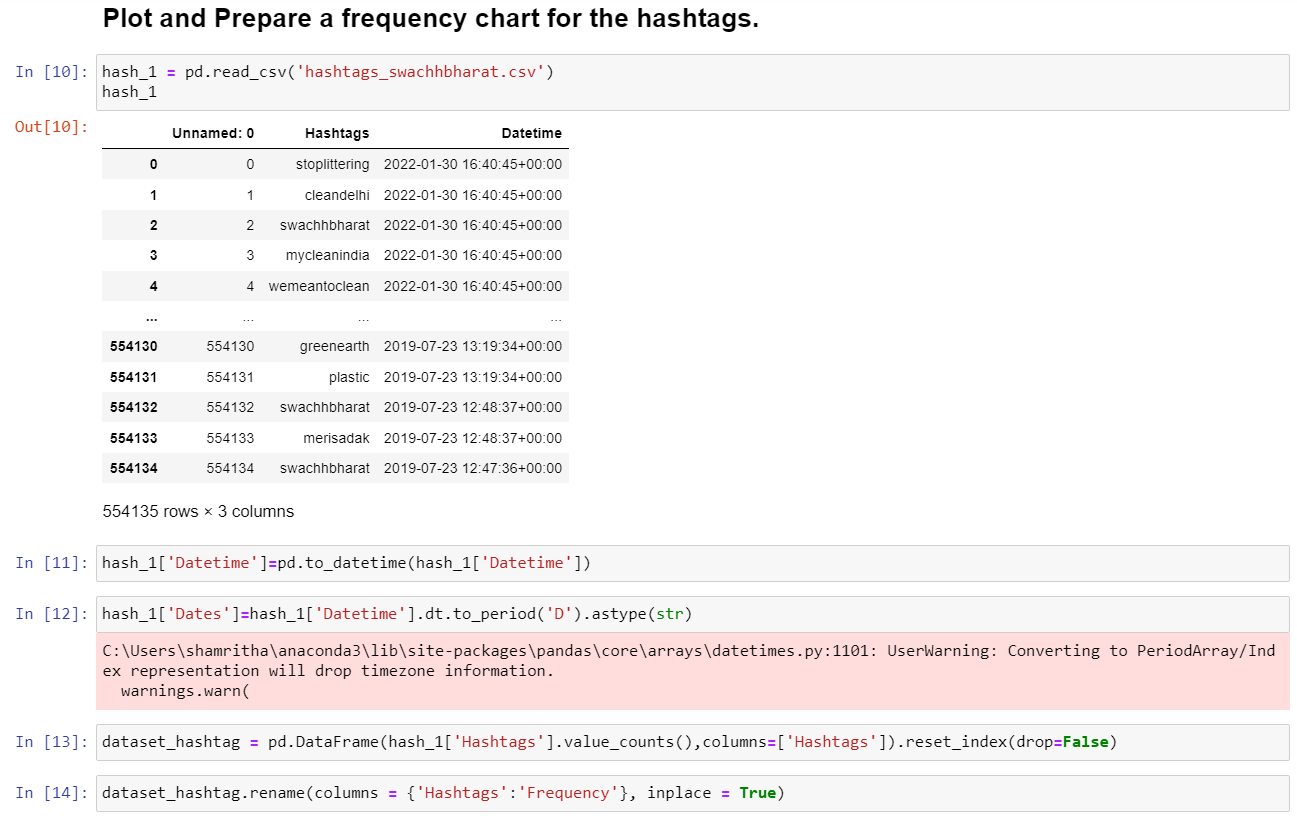
****

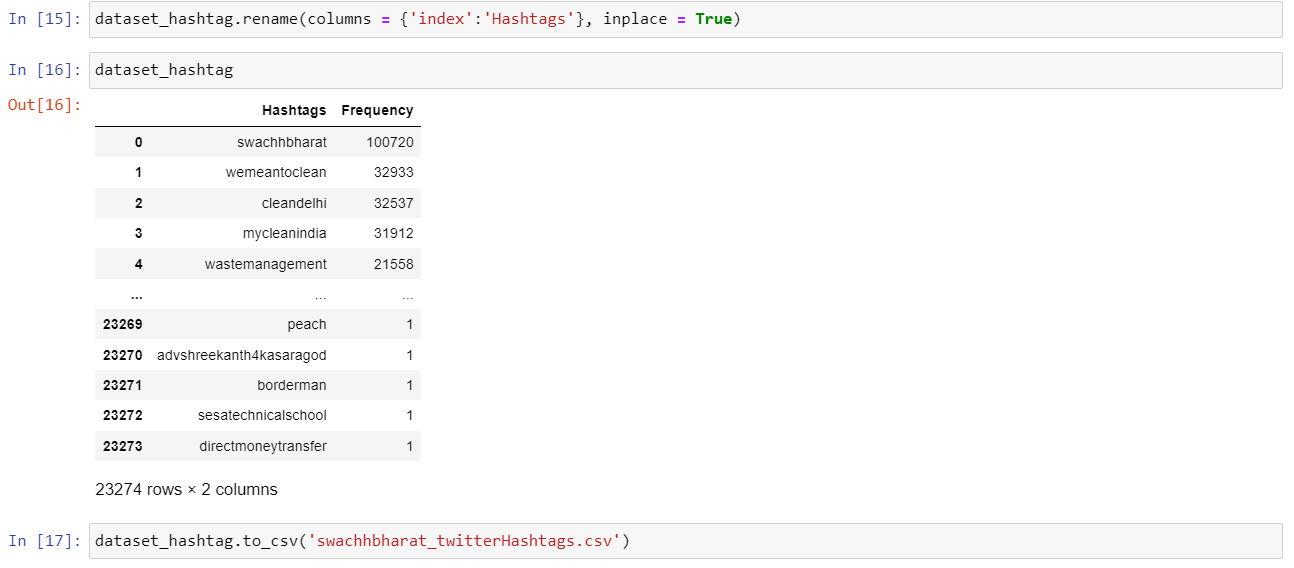
****

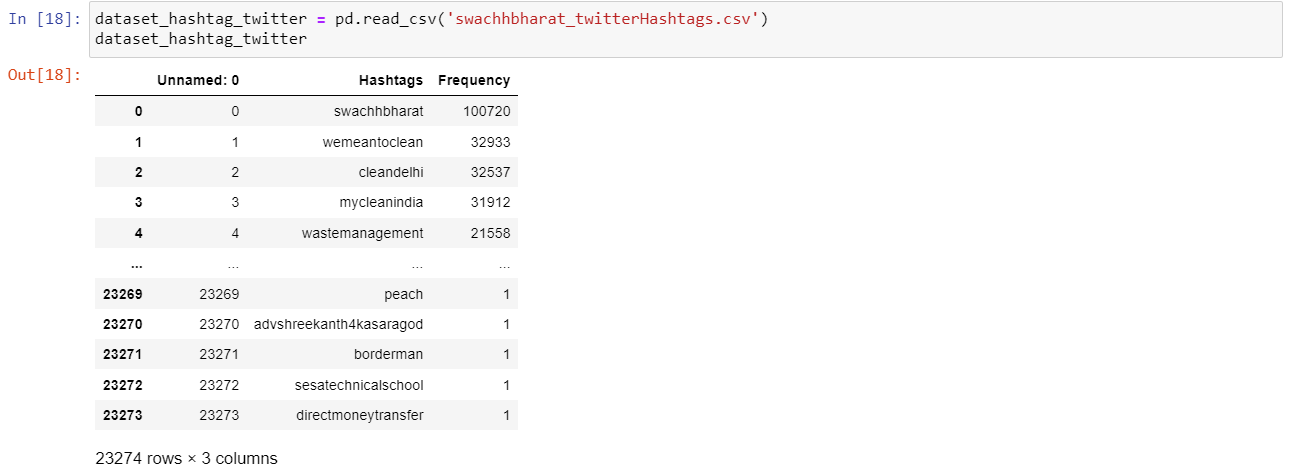
****

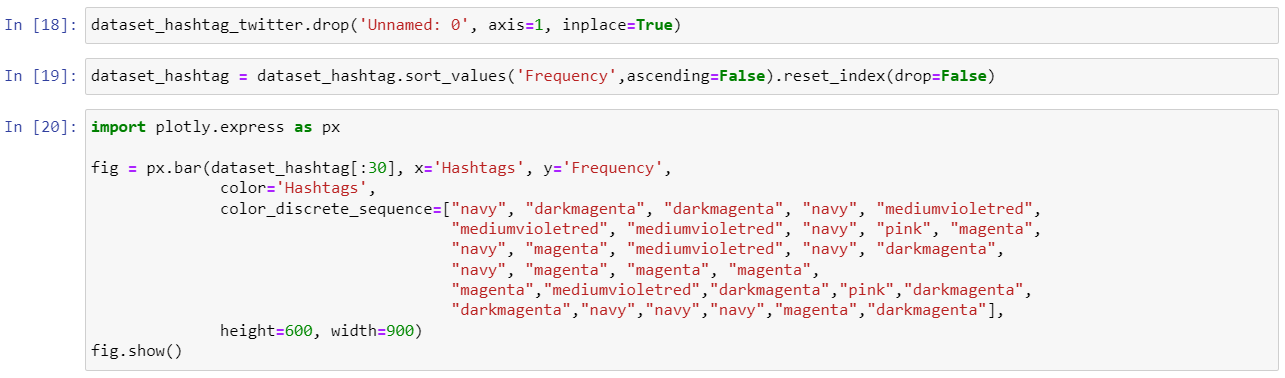
****

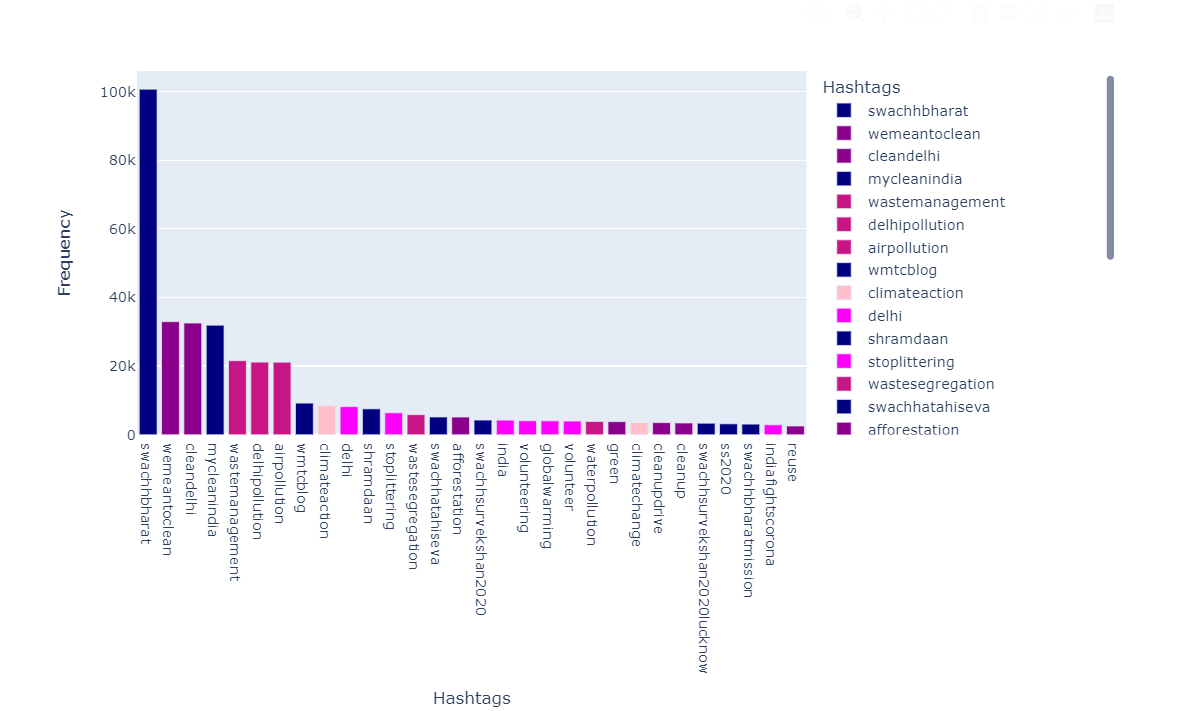
****

****

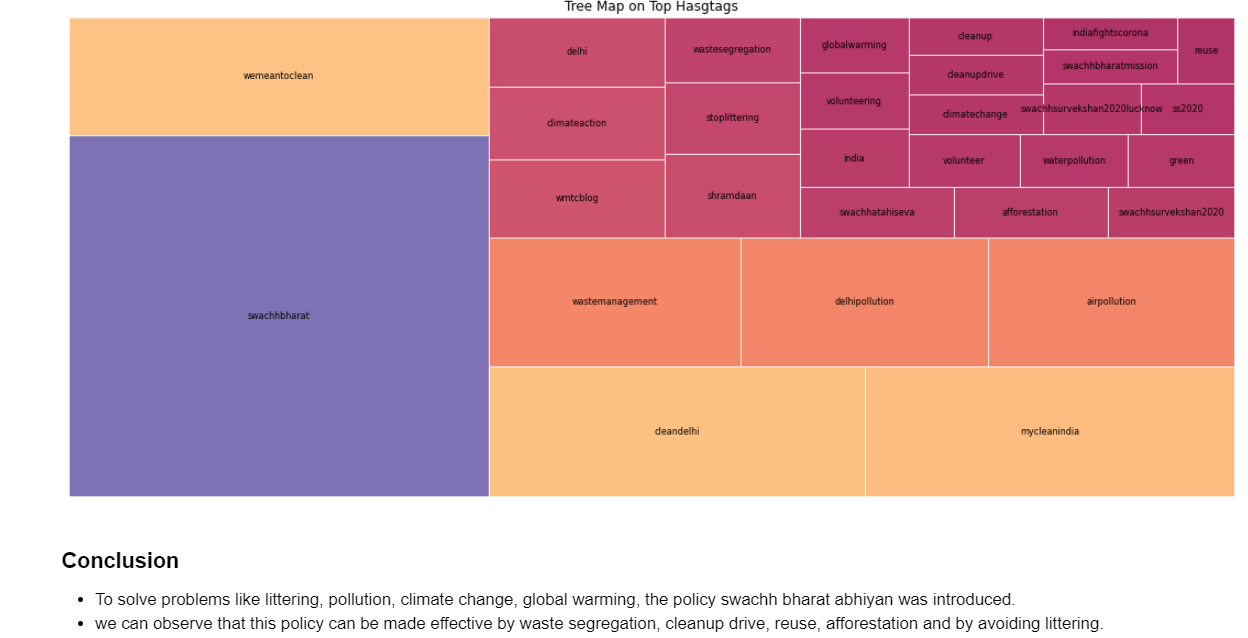
****

****

****

****

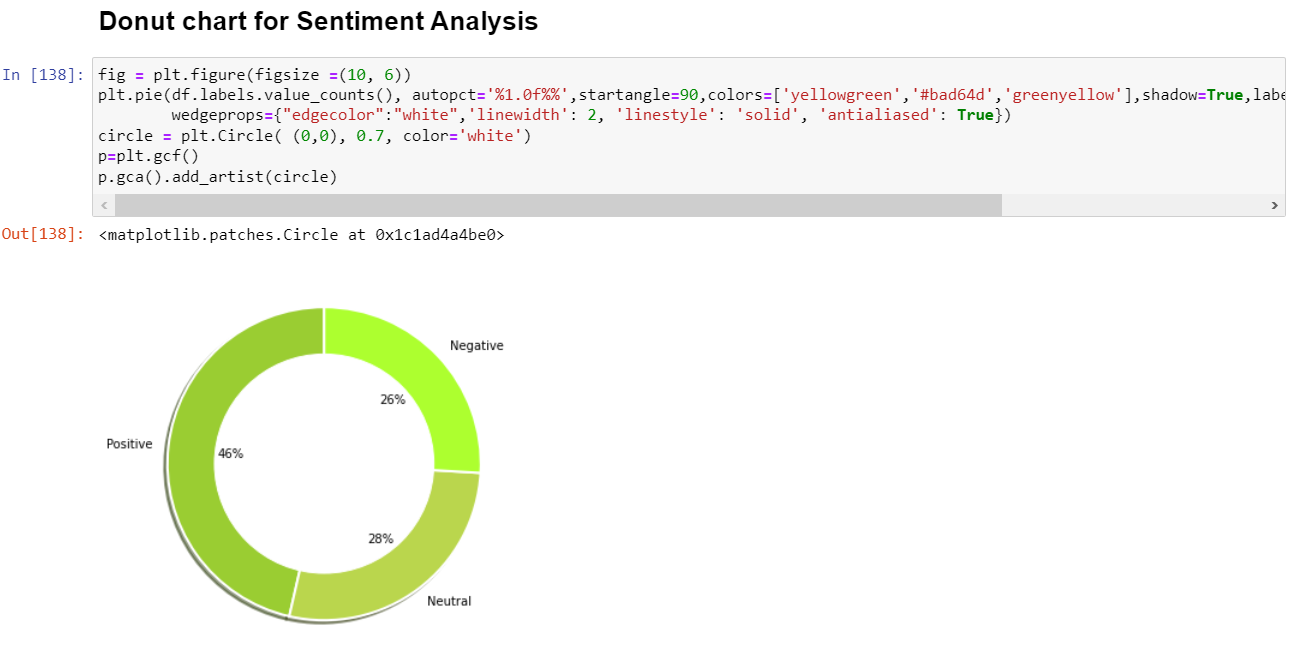
****

****

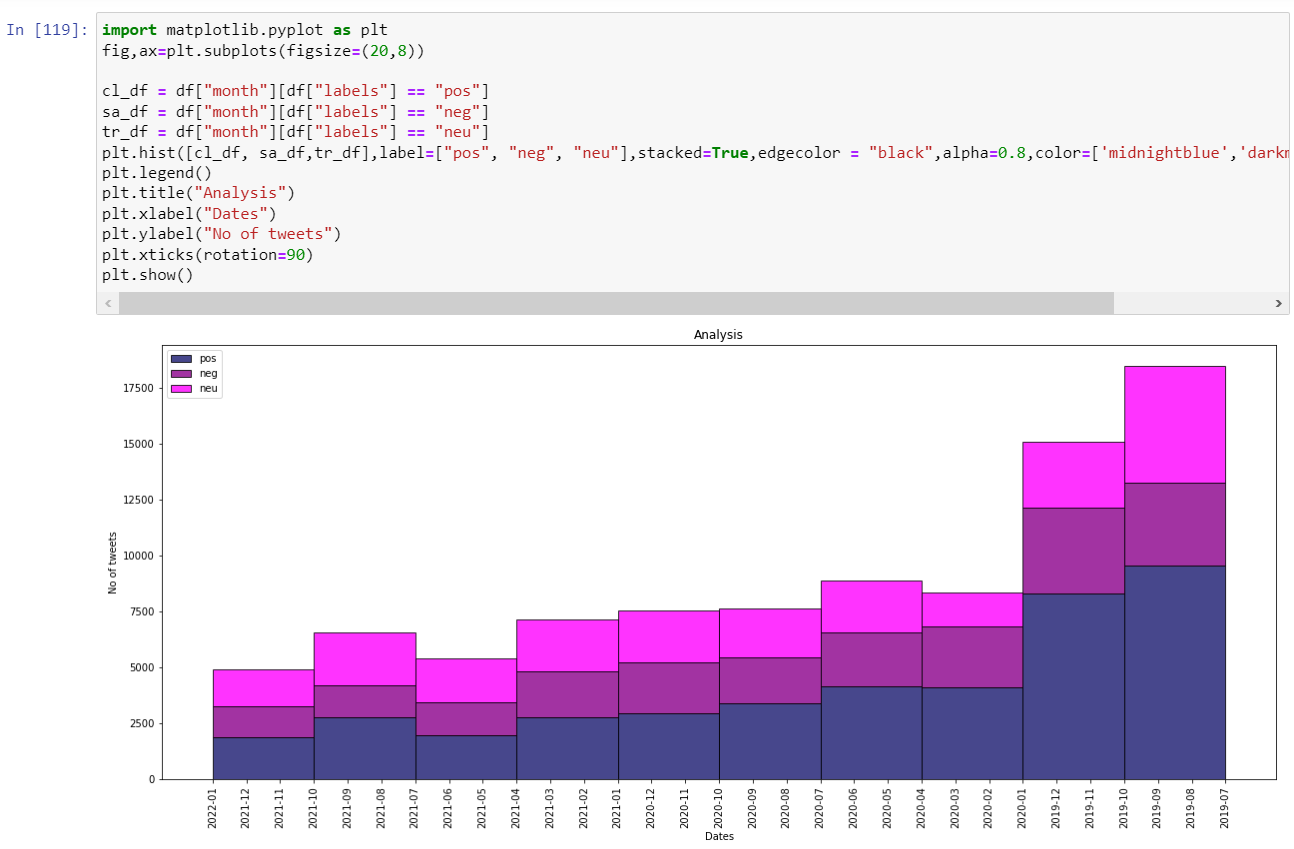
****

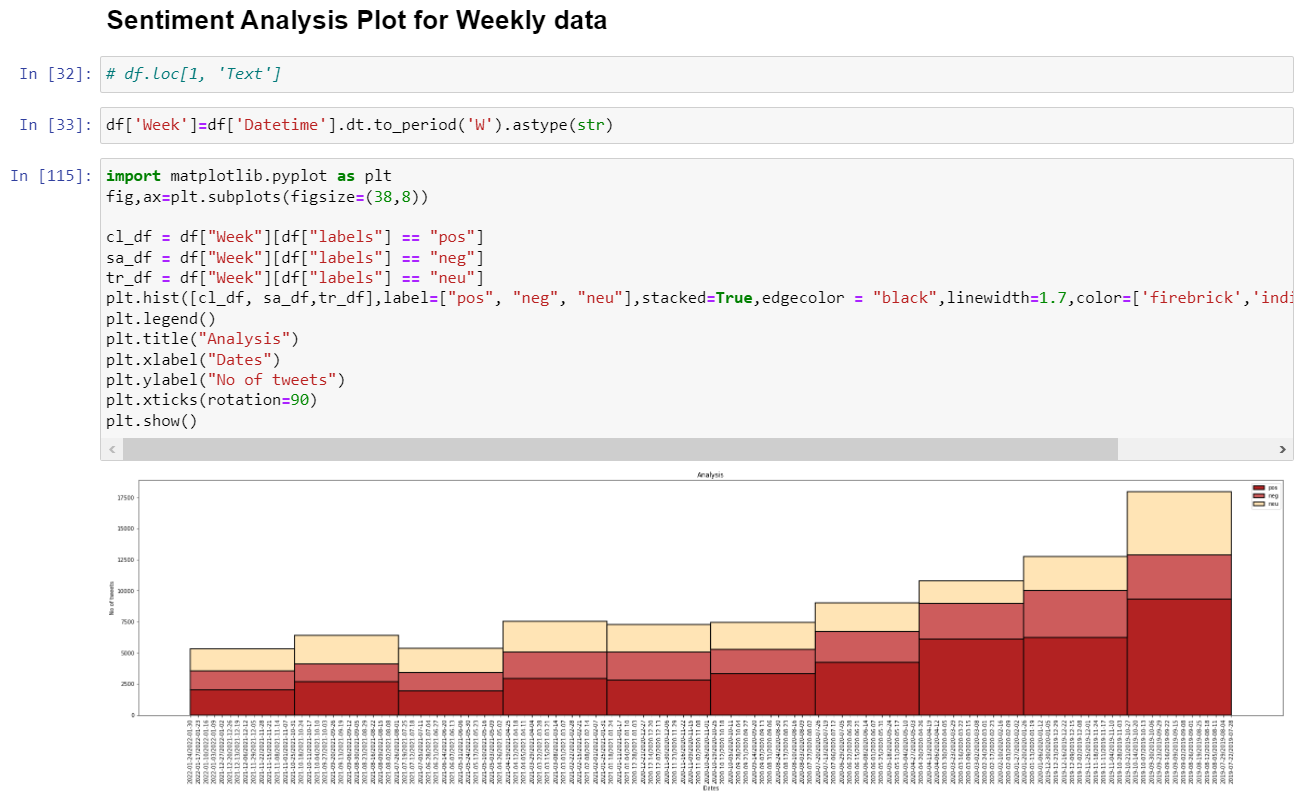
****

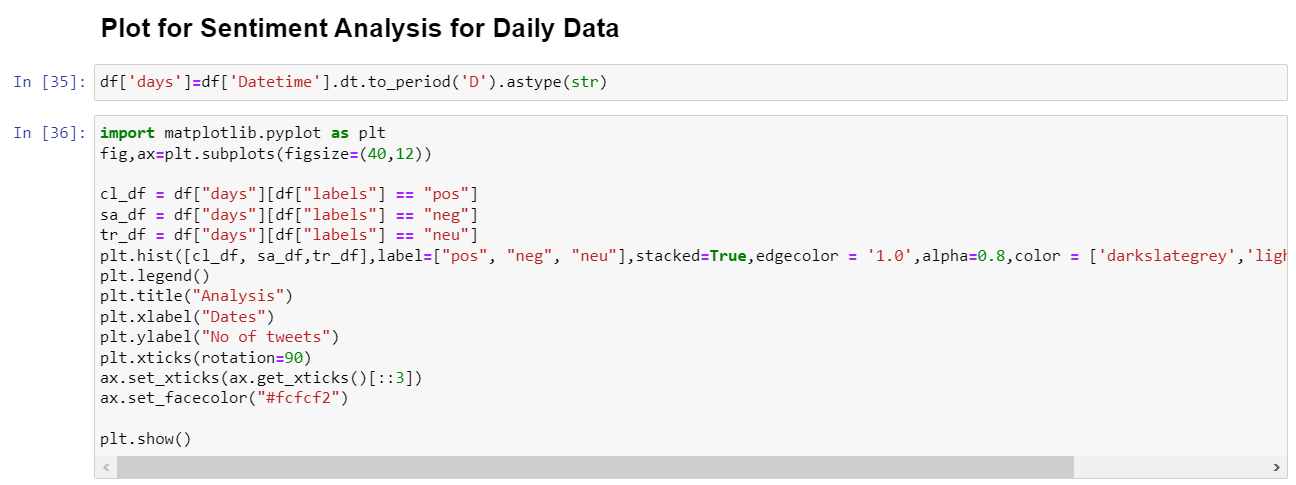
****

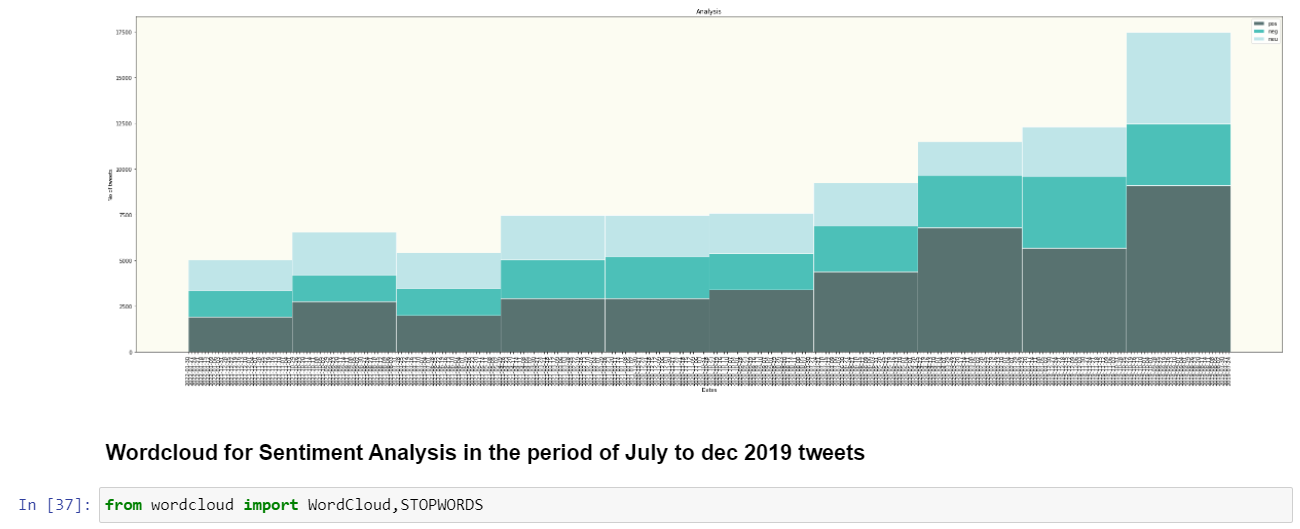
****

****

****

****

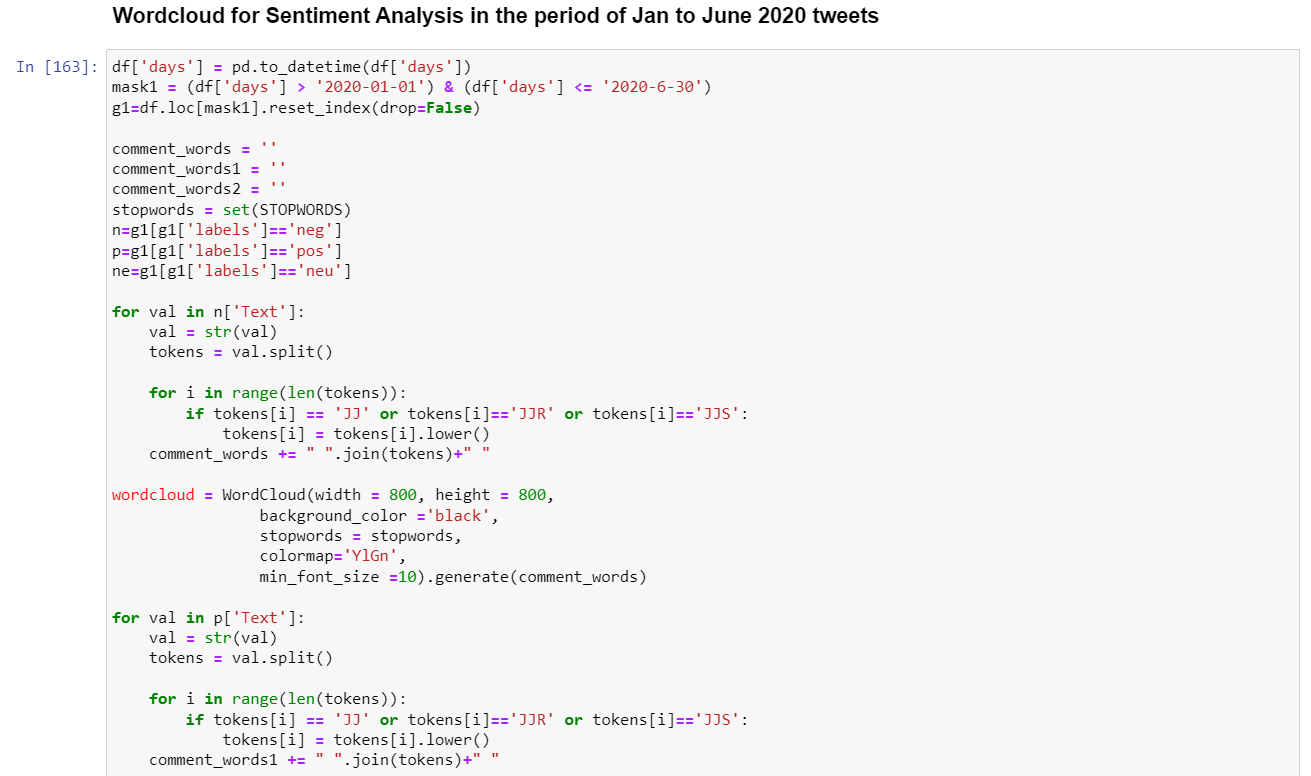
****

****

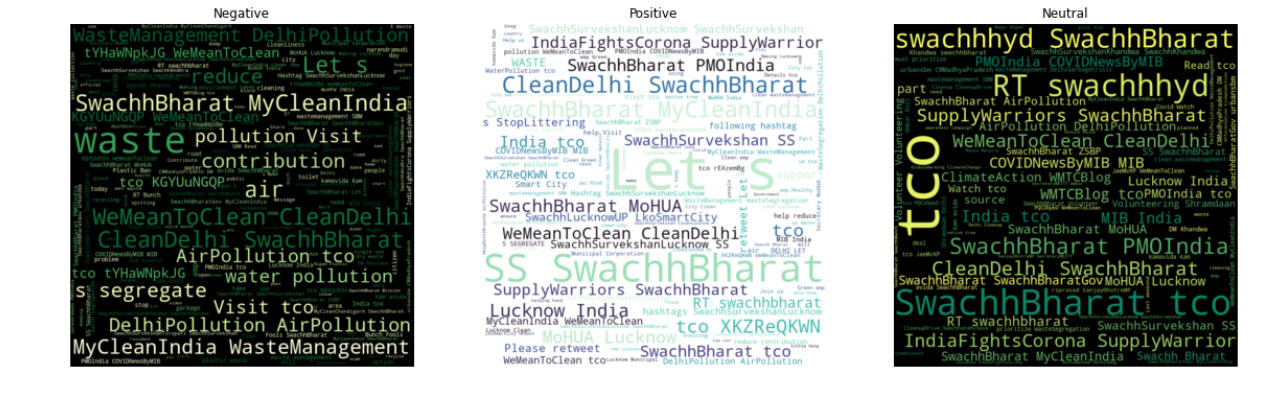
****

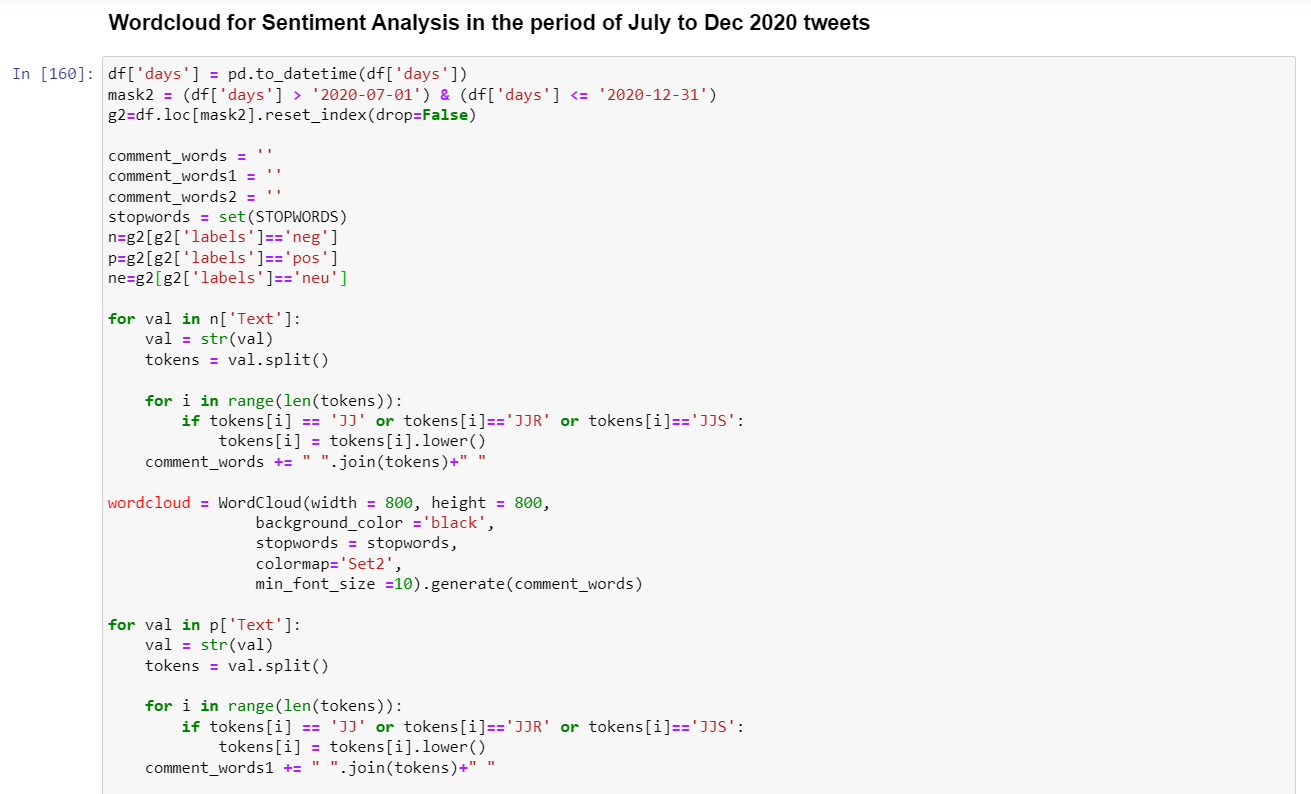
****

****

****

****

****

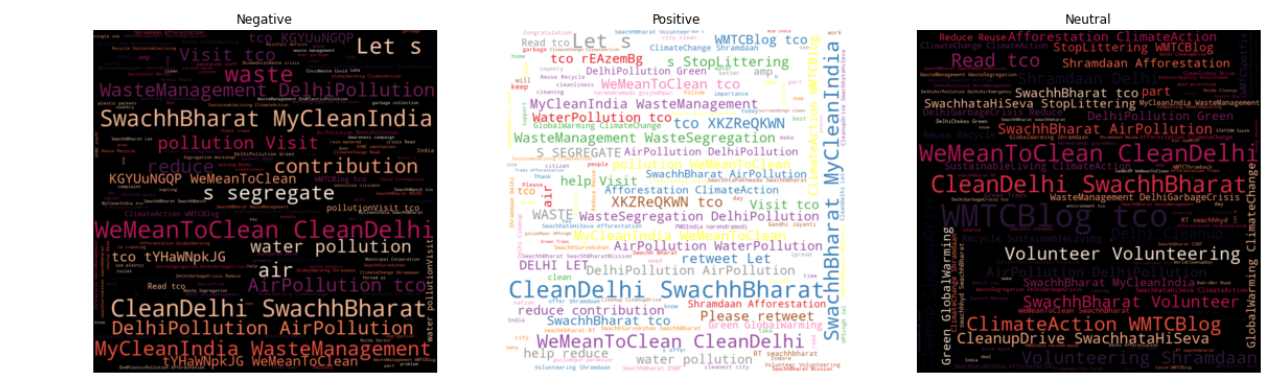
****

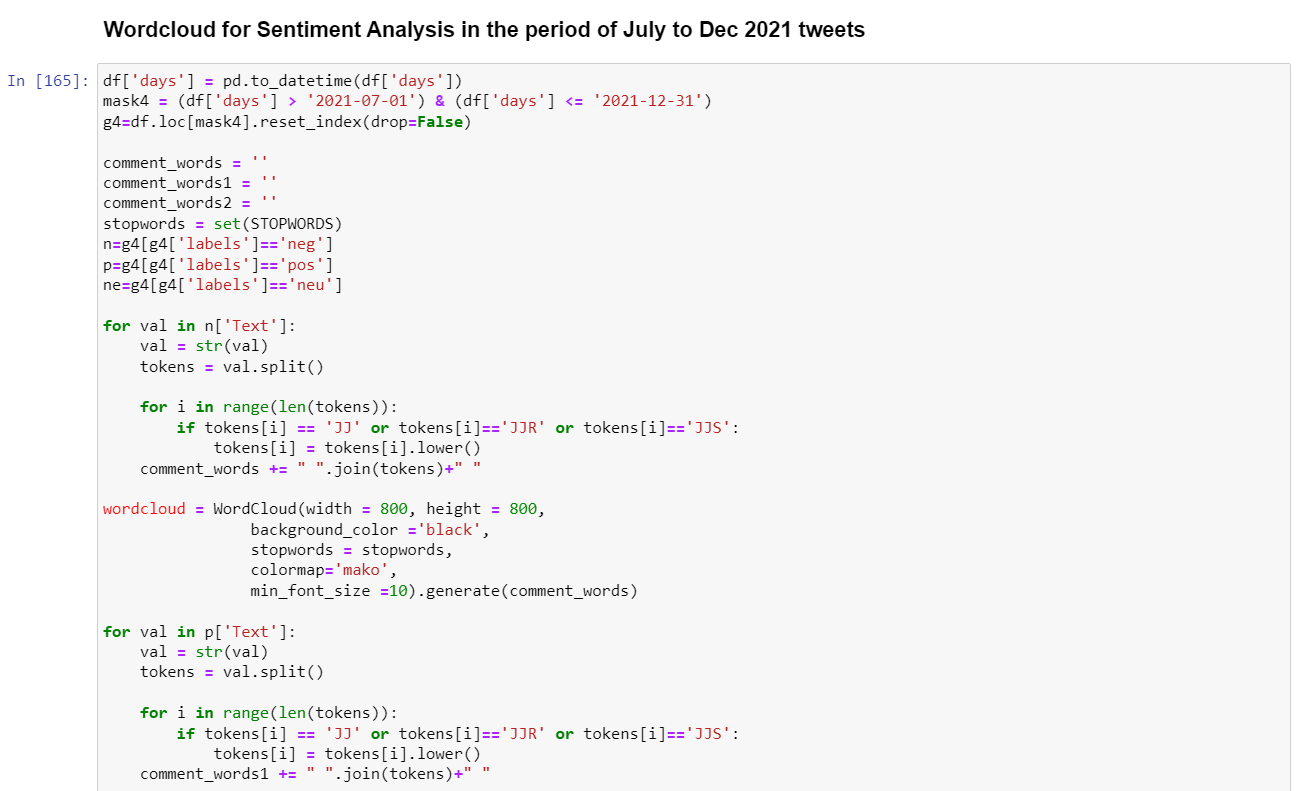
****

****

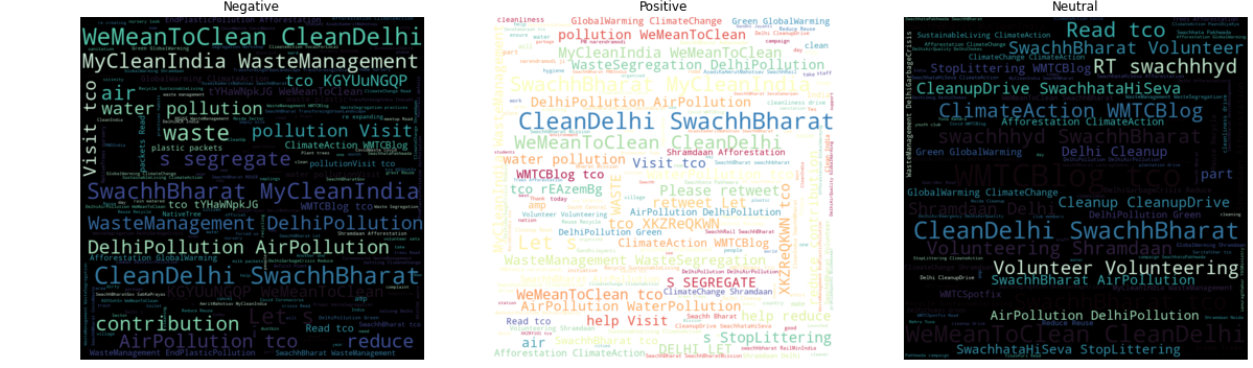
****

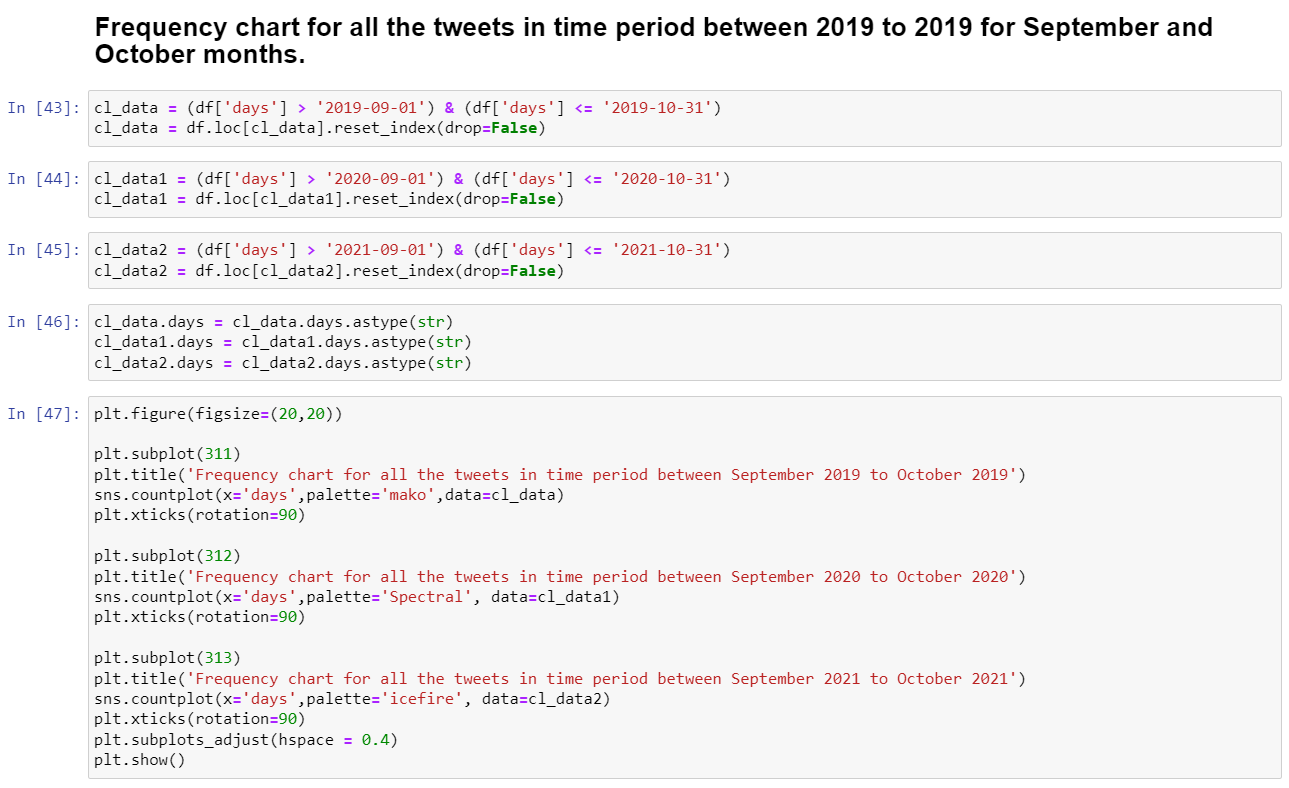
****

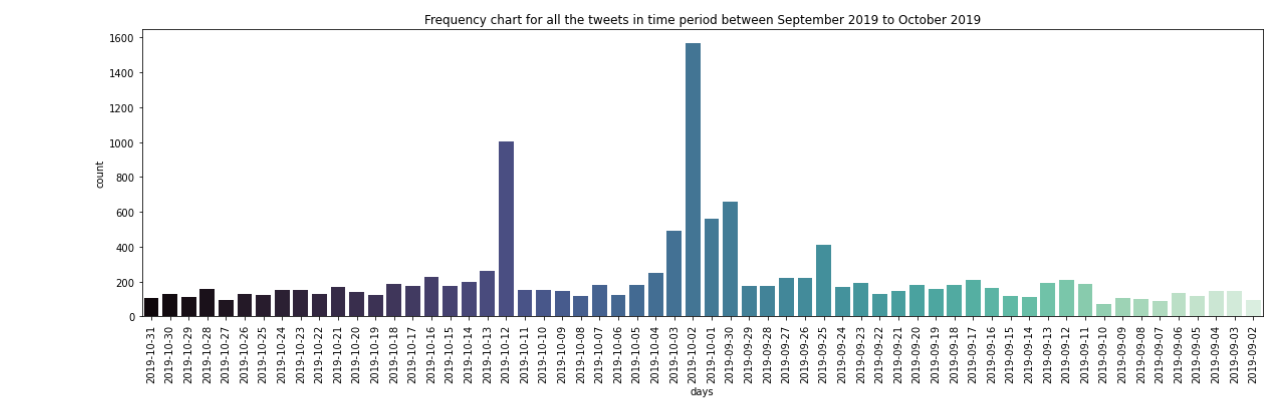
****

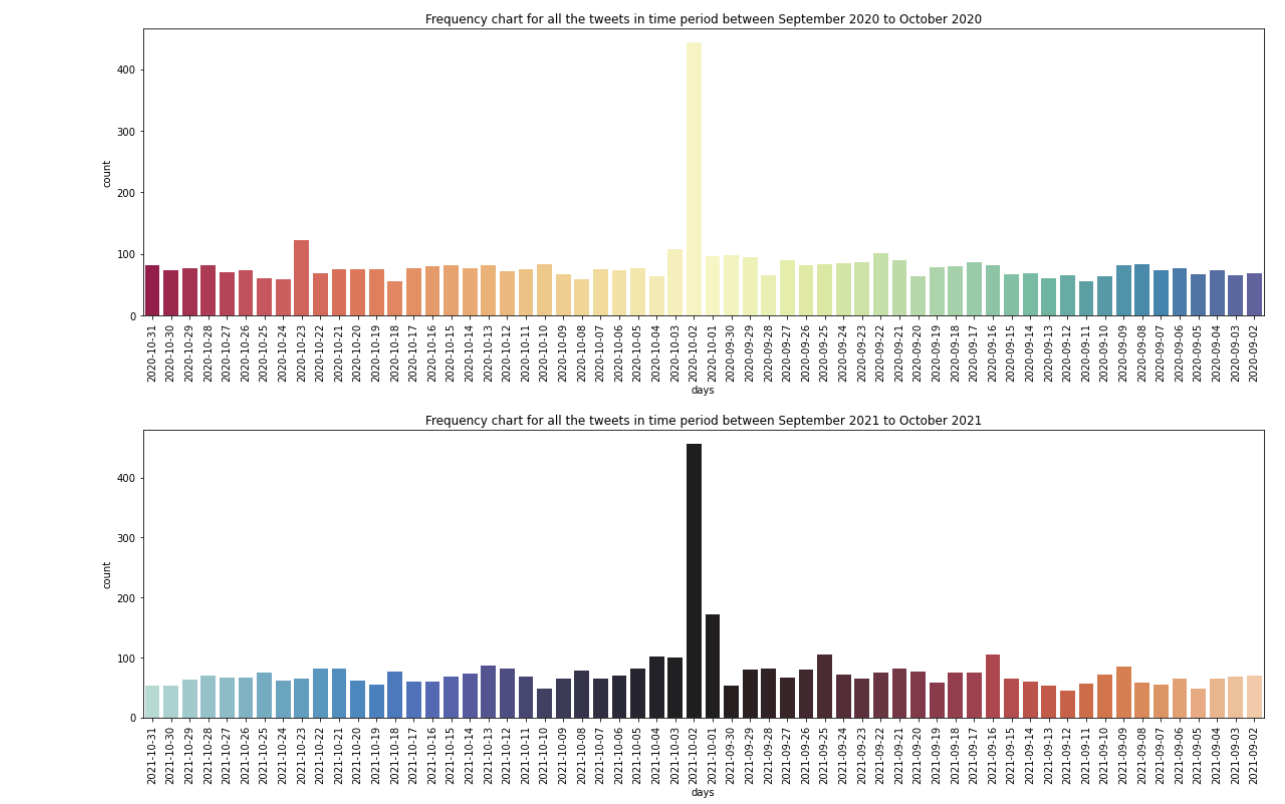
****

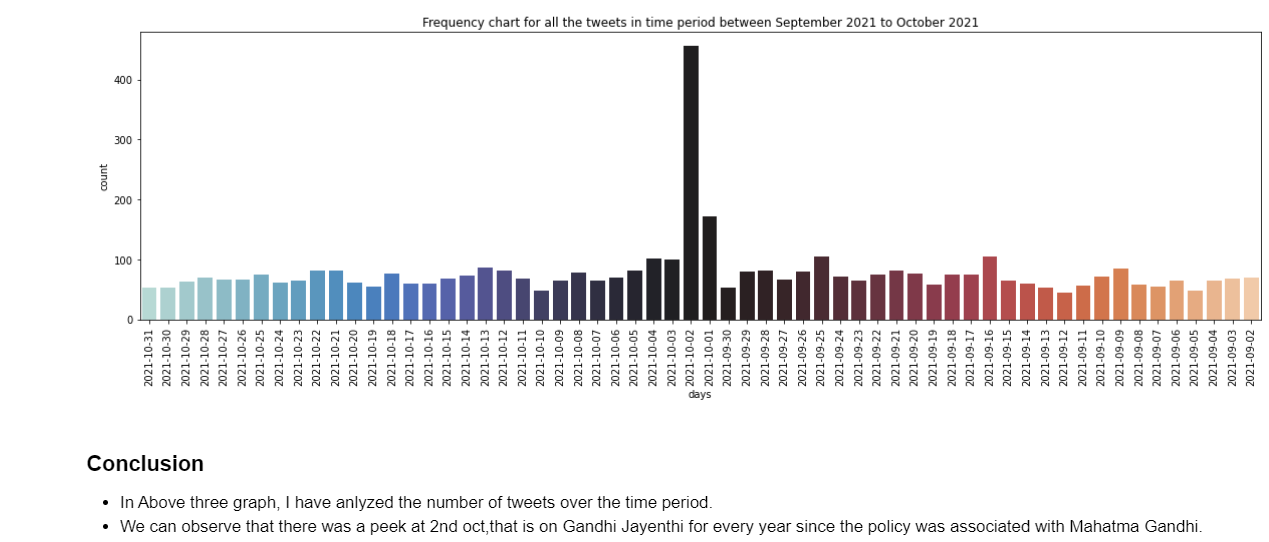
****

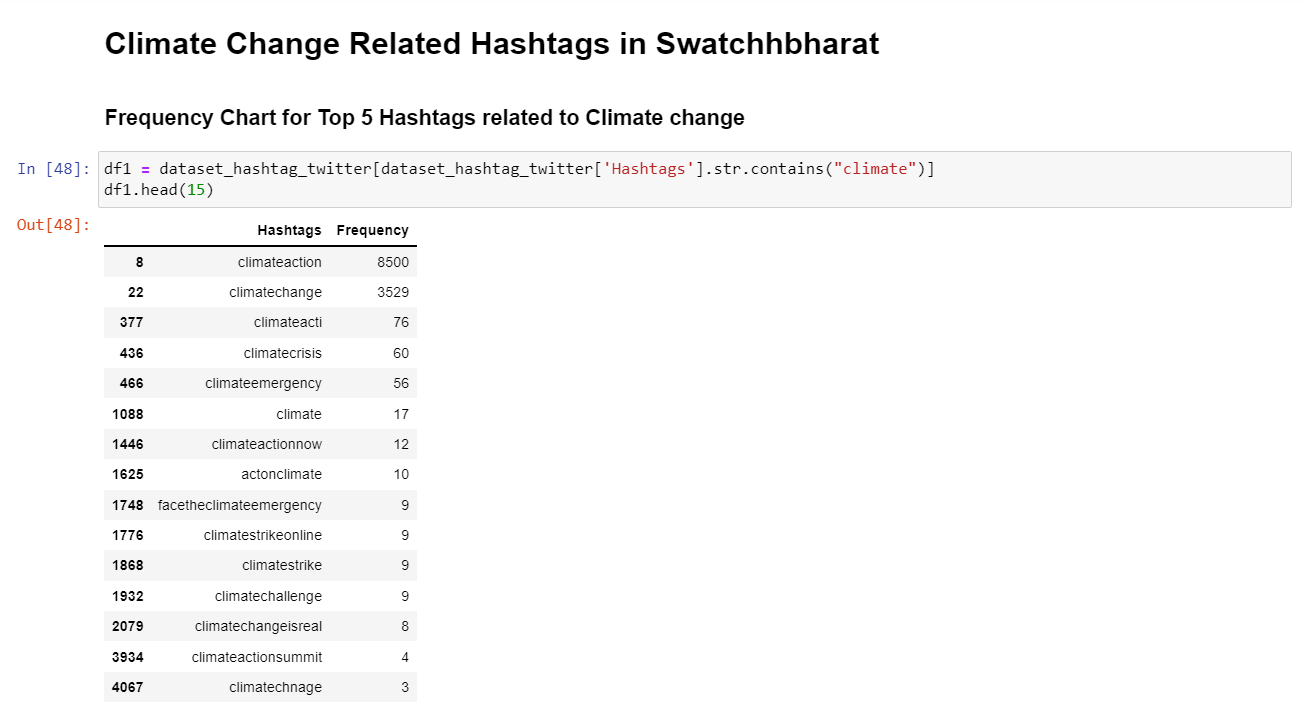
****

****

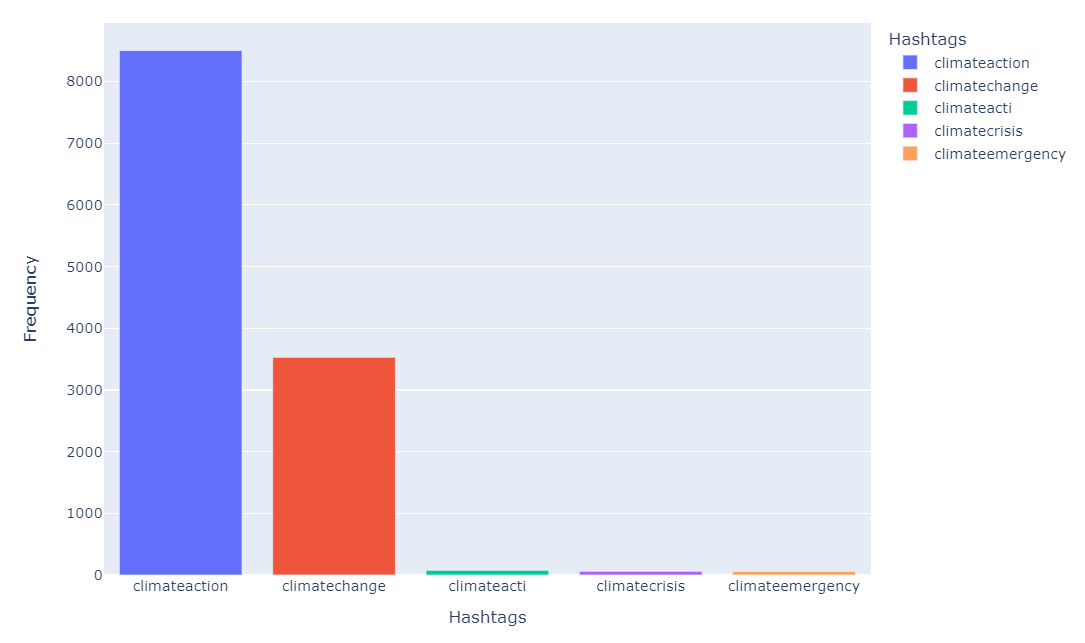
****

****

****

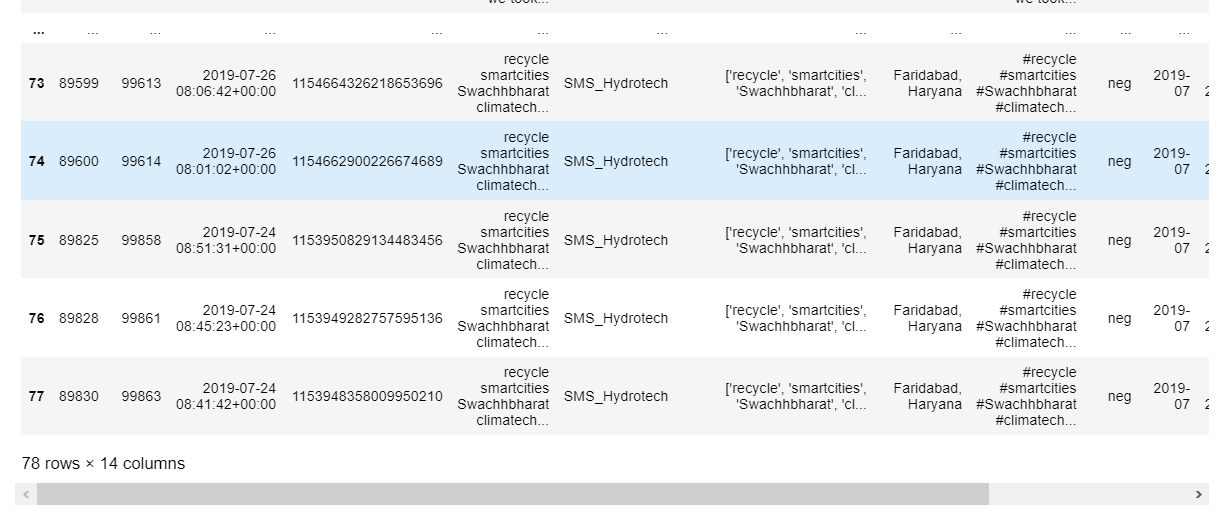
****

****

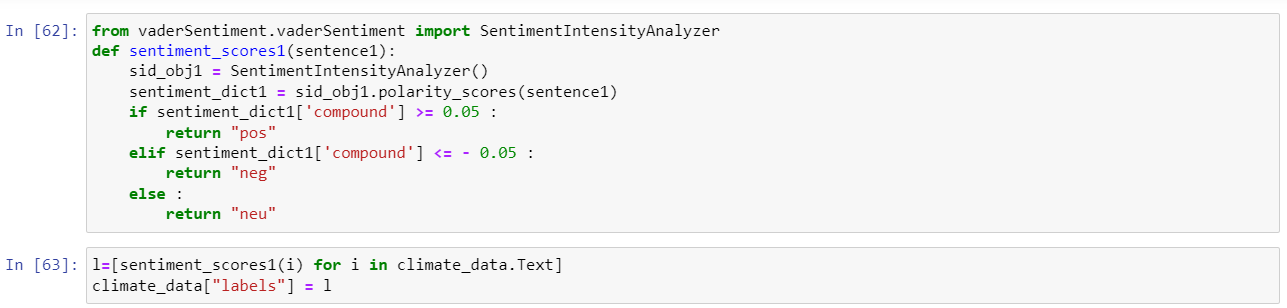
****

****

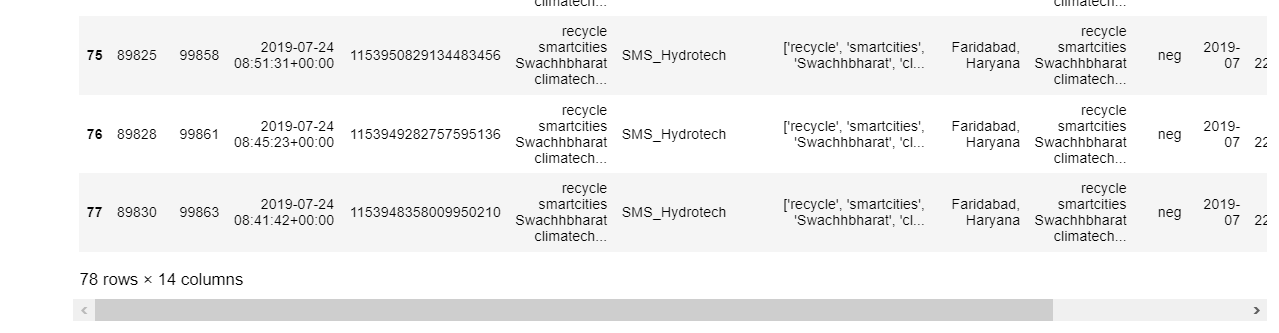
****

****

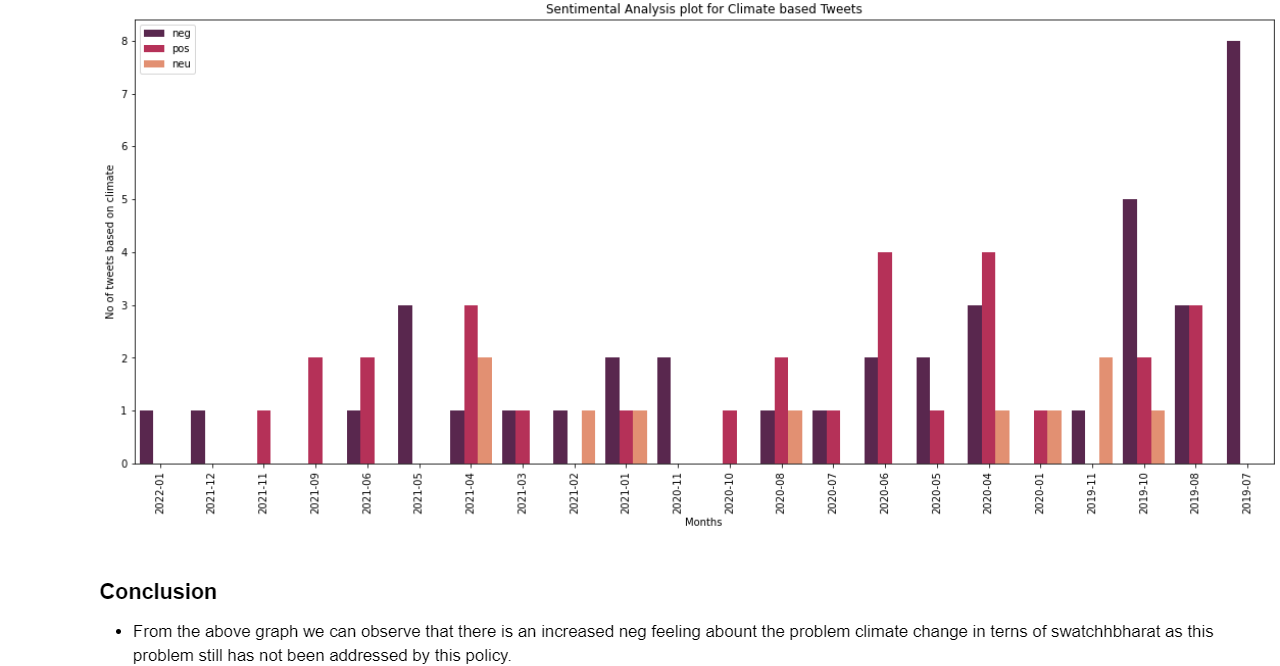
****

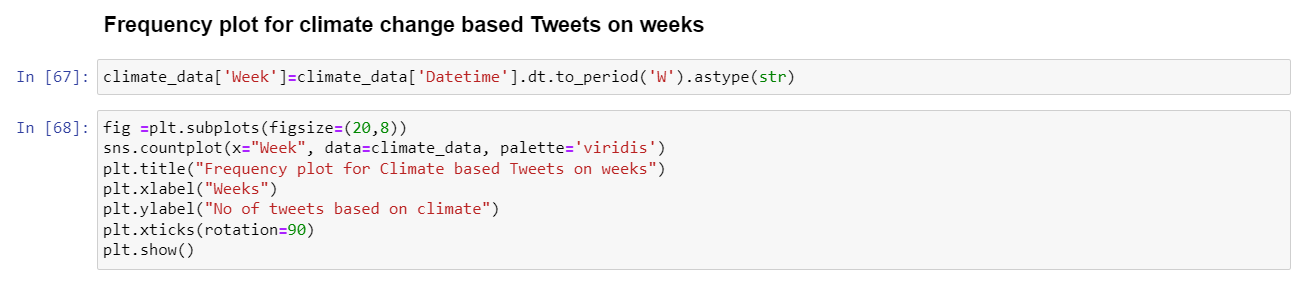
****

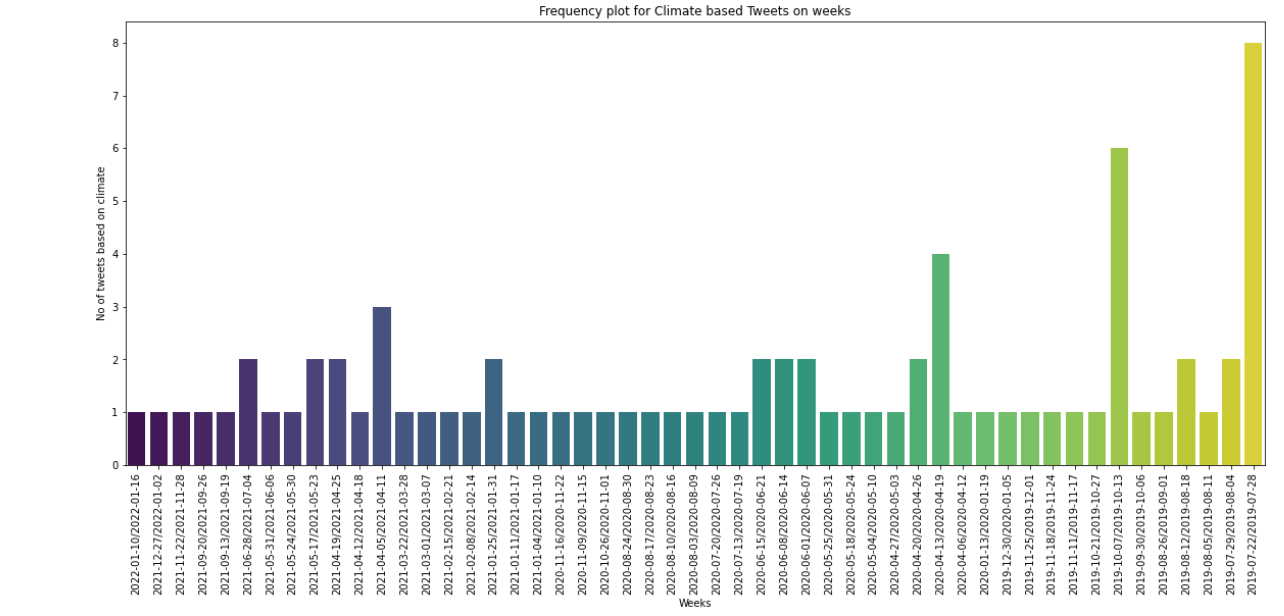
****

****

****

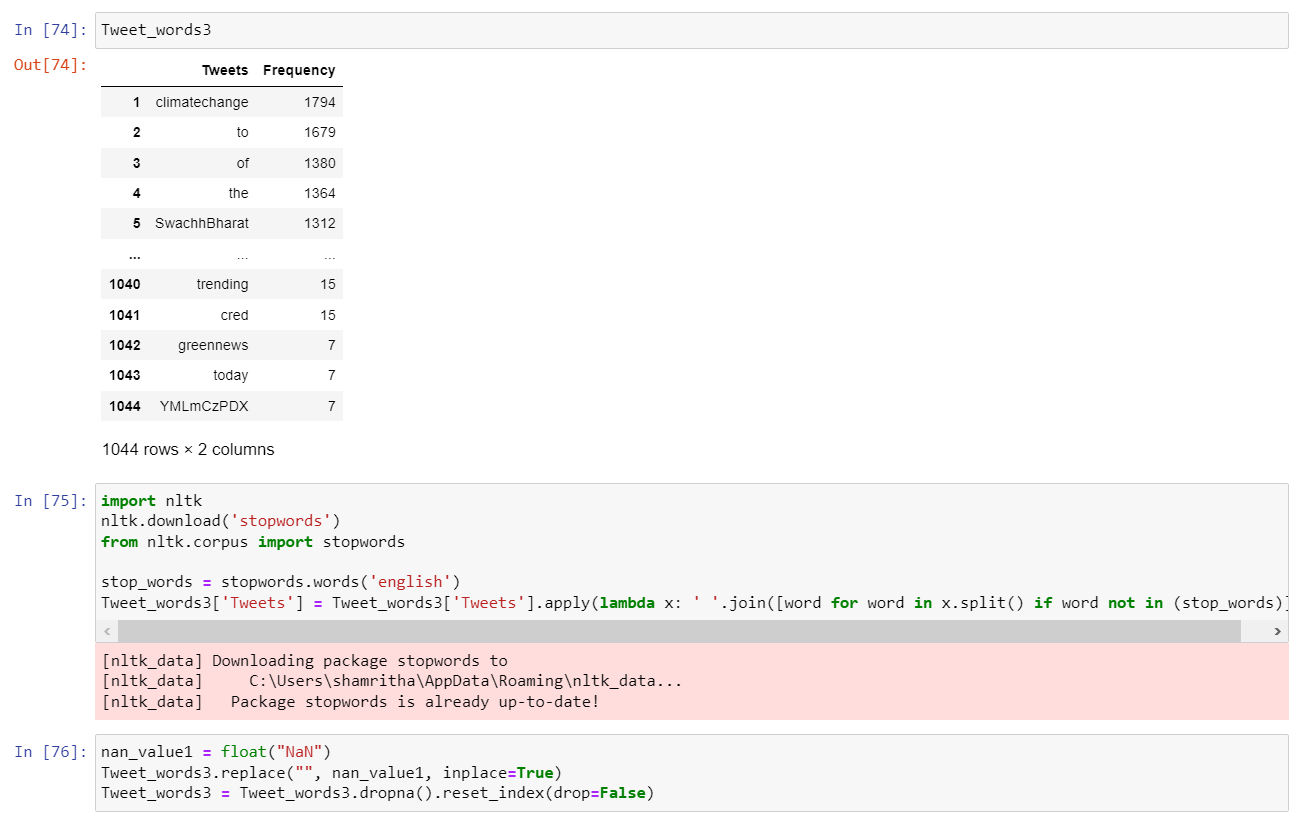
****

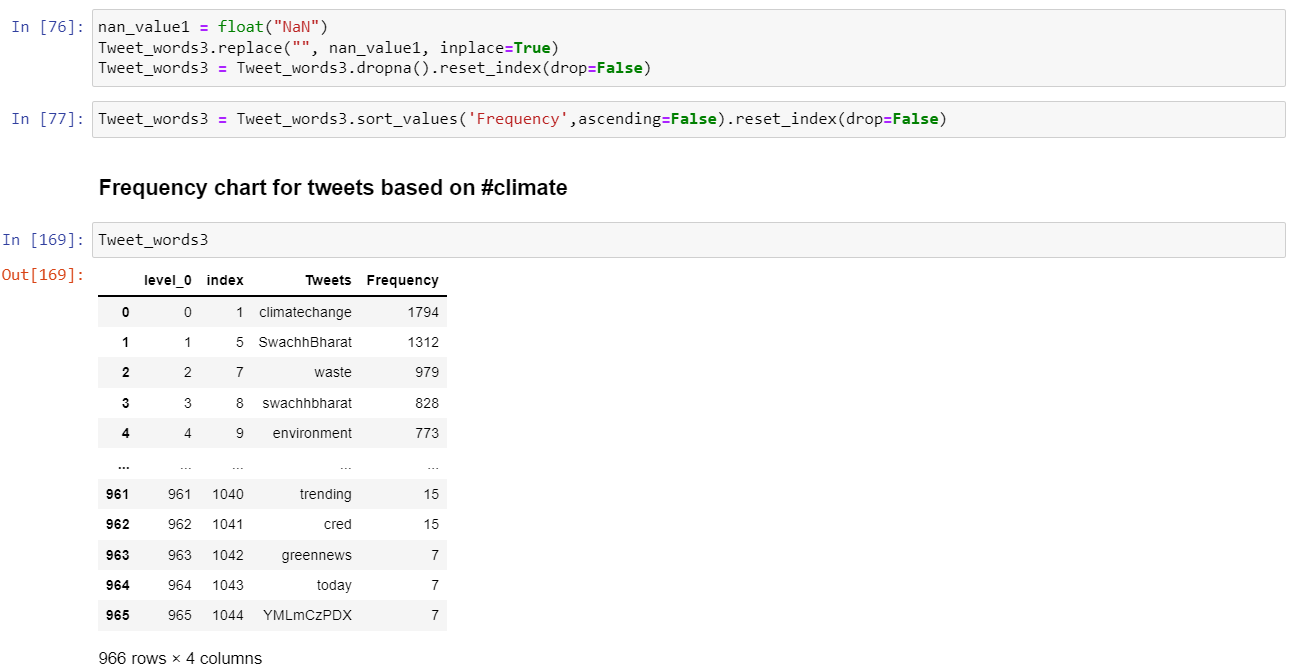
****

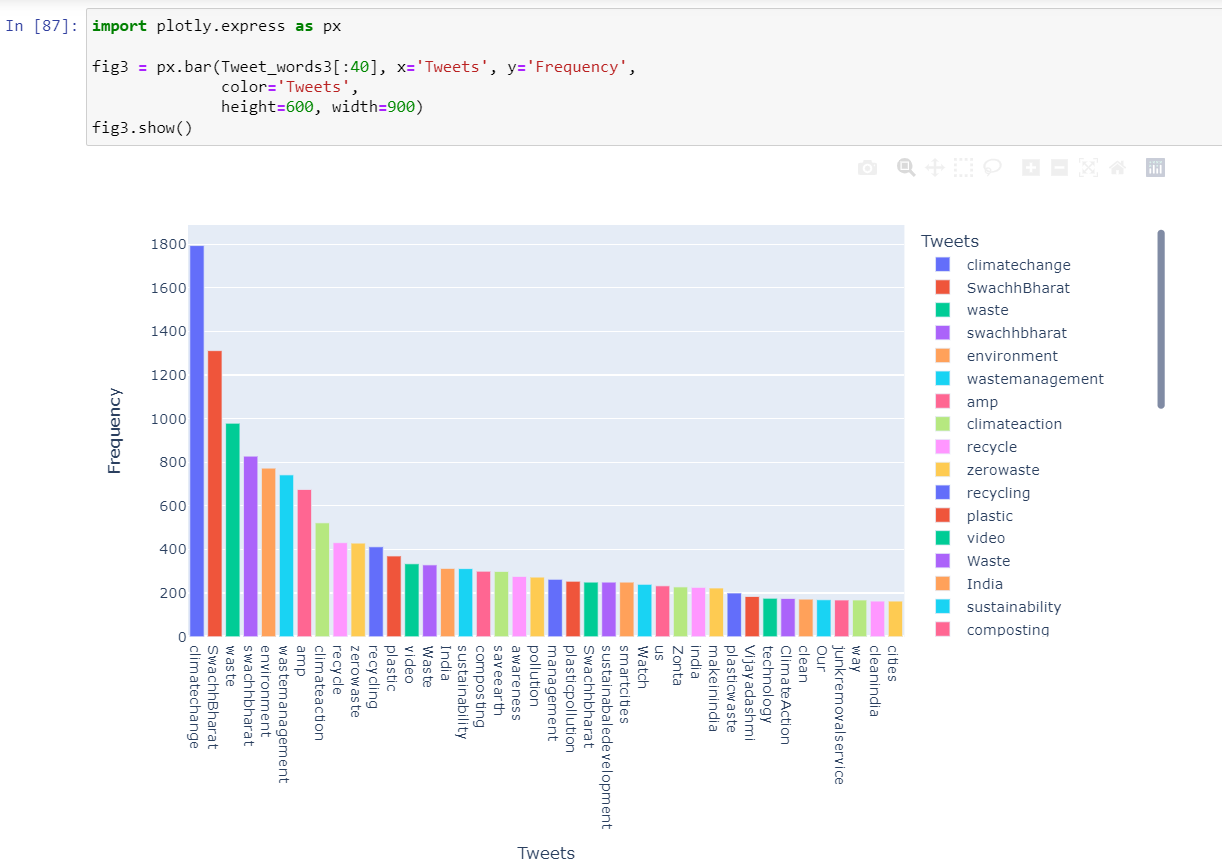
****

****

****

****

****

****

**8. RESULT**

**8.1 Conclusion**

* The fundamental goal of the Swachh Bharat policy was to avoid trash, pollution, and afforestation, and we can see that the policy has a good sentiment among the general public.
* By providing toilets and educating people about basic sanitation and cleanliness, the Swachh Bharat initiative has made a significant difference in the lives of those living in rural regions.

**8.2** **Future Scope**

Future initiatives that I hope to explore:

• To discover whether additional variables have influenced the notion of Swachh Bharat negatively or positively.

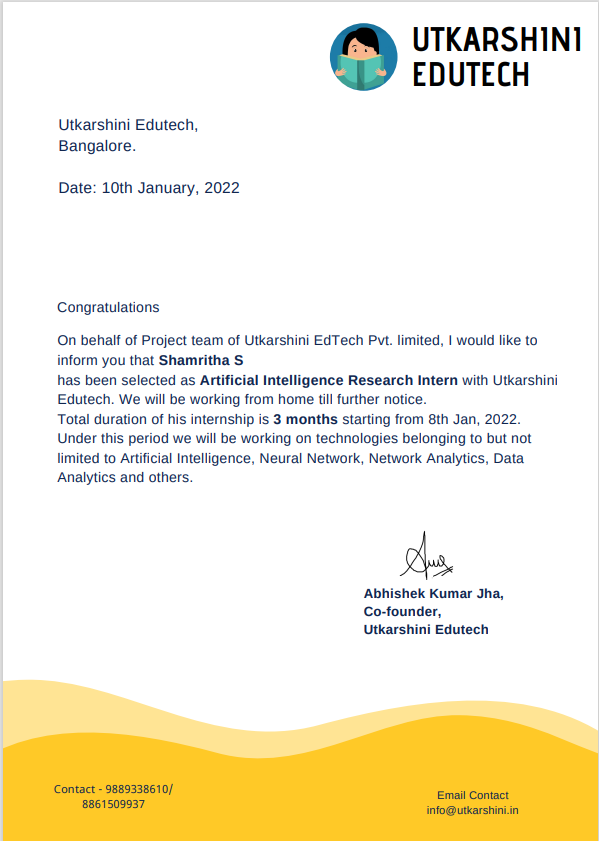
**9. REFERENCE**

* <https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/>
* <https://sci-hub.hkvisa.net/10.1109/ICCCI48352.2020.9104105>
* <https://sci-hub.hkvisa.net/10.1109/Confluence51648.2021.9377048>
* <https://thesai.org/Downloads/Volume10No2/Paper_48-A_Study_on_Sentiment_Analysis_Techniques.pdf>
* <https://thesai.org/Downloads/Volume10No2/Paper_48-A_Study_on_Sentiment_Analysis_Techniques.pdf>
* https://sci-hub.hkvisa.net/10.1109/BIGCOMP.2017.7881754
* <https://sci-hub.hkvisa.net/10.1109/ISCON47742.2019.9036208>
* https://sci-hub.hkvisa.net/10.1109/ISCID.2019.00064

**10. WORKLOG**

|  |  |  |
| --- | --- | --- |
| Day | Date | Task Done |
| Day 1 | 9-Feb-2022 | External: Scraped data from twitter for the topic  #swachhbharat by using snscrape. |
| Day 2 | 10-Feb-2022 | External: Scraped data between 2019 to 2022 for the topic  #swachhbharat by using snscrape. (from the covid started date) |
| Day 3 | 11-Feb-2022 | External: Data cleaning, extraction of individual  hashtag from tweets and plotting graph (Frequency Chart) for the most common hashtag from the scraped data |
| Day 4 | 12-Feb-2022 | External: Learnt basics of sentimental. |
| Day 5 | 14-Feb-2022 | External: Using sentimental analysis i have  analysed the patterns for positive and  negative reviews. |
| Day 6 | 15-Feb-2022 | External: Plotted graph for sentimental analysis  and trying for some advanced plots |
| Day 7 | 16-Feb-2022 | External: Performing word cloud for positive to negative tweets |
| Day 8 | 17-Feb-2022 | External: Performing the simple line chart to  display the no of positive, negative and neutral tweets in the period of time |
| Day 9 | 21-Feb-2022 | External: Performed Word cloud for the period of July to Dec(2019), Jan-Jun(2020) and Jul-Dec (2020) for the Tweets separately for positive, negative and neutral |
| Day 10 | 22-Feb-2022 | External: Performed Word cloud for the period of Jan Jun(2020), Jul-Dec(2020) for the Tweets separately for positive, negative and neutral. |
| Day 11 | 23-Feb-2022 | External: Performed Frequency chart for all the  words present in the word cloud separately for positive negative and neutral |
| Day 12 | 25-Feb-2022 | External: Performed Frequency chart for with the word climate in hashtag. |
| Day 13 | 28-Feb-2022 | External: Using sentimental analysis I have  analysed the patterns for positive and  negative reviews for online. |

**11. OFFER LETTER**

****