VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

"Sentiment Analysis Of Lockdown In India During Covid-19 A Case Study On Twitter using ML"

Submitted in partial fulfilment for the award of degree

BACHELOR OF ENGINEERING IN

"COMPUTER SCIENCE AND ENGG"

Submitted by :

NAME: KUMUDA D K - 4VV19CS072
S.YASHWANTH - 4VV19IS090
LIKHITHA B G - 4VV20CS409

Conducted at

Sain Informatics Private Limited

Vidyavardhaka college of Engineering

Department of CS & IS Engineering

VVCE is an autonomous institute affiliated to Visvesvaraya Technological University (VTU), Belagavi, approved by AICTE & UGC, New Delhi and recognized by Govt. of Karnataka.

CERTIFICATE

This is to certify that the Internship titled **"Sentiment Analysis Of Lockdown In India During Covid-19 A Case Study On Twitter using ML"** carried out **by KUMUDA D K (4VV19CS072), S.YASHWANTH (4VV19IS090) and LIKHITHA.BG (4VV20CS409),** a bonafide student of **Vidyavardhaka college of Engineering**, in partial fulfillment for the award of Bachelor of Engineering, in **Computer Science and Engineering** under **Visvesvaraya Technological University, Belagavi,** during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

Signature of Guide **Signature of HOD** **Signature of Principal**

External Viva:

Name of the Examiner Signature with Date

1)_____

_____

2)_____

_____

DECLARATION

we final year students of CS and IS branch, VVCE - 560 082, declare that the Internship has been successfully completed, in Sain Informatics Private Limited . This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Branch name, during the academic year 2022-2023.

Date : 23/9/2022 :

Place : Mysuru

KUMUDA DK - 4VV19CS072
S.YASHWANTH - 4VV19IS090
LIKITHA B G - 4VV20CS409

OFFER LETTERS :

INTERNSHIP OFFER LETTER

Date: 23rd August, 2022

Name: Kumuda D K
USN: 4VV19CS072

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Sain Informatix Pvt Ltd**, effective Start Date **23rd August, 2022**. All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Sunlin
Product Manager
SAIN INFORMATIX PVT LTD
Bangalore, India

INTERNSHIP OFFER LETTER

Date: 23rd August, 2022

Name: S.Yashwanth
USN: 4VV19IS090

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Sain Informatix Pvt Ltd**, effective Start Date **23rd August, 2022**. All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Sunlin
Product Manager
SAIN INFORMATIX PVT LTD
Bangalore, India

INTERNSHIP OFFER LETTER

Date: 23rd August, 2022

Name: Likhitha B G
USN: 4VV20CS409

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Sain Informatix Pvt Ltd**, effective Start Date **23rd August, 2022**. All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Sunlin
Product Manager
SAIN INFORMATIX PVT LTD
Bangalore, India

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our Principal, Dr.SadashiveGowda for providing facilities to keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, forhelping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

KUMUDA.D.K - 4VV19CS072
S.YASHWANTH - 4VV19IS090
LIKHITHA B G - 4VV20CS409

ABSTRACT

Sentiment Analysis also known as Opinion Mining refers to the use of natural language processing, text analysis to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

In this project, we aim to perform Sentiment Analysis of product based reviews. Data used in this project are online product reviews collected from “amazon.com”. We expect to do review-level categorization of review data with promising outcomes.

Table of Contents

CHAPTER 1

1. COMPANY PROFILE

A Brief History of Sain Informatics Private Limited :

Directors are associated with the organization. Trinetra and Menasinkai Mohan Trilok are presently associated as directors. Sain Informatix Private Limited is an unlisted private company incorporated on 18 August, 2020. It is classified as a private limited company and is located in , Karnataka. It's authorized share capital is INR 10.00 lac and the total paid-up capital is INR 10,000.00 .

The current status of Sain Informatix Private Limited is - Active.

Details of the last annual general meeting of Sain Informatix Private Limited are not available. The company is yet to submit its first full-year financial statements to the registrar.

Sain Informatix Private Limited has two directors - Menasinkai Mohan Trilok and Trinetra.

The Corporate Identification Number (CIN) of Sain Informatix Private Limited is U72900KA2020PTC137383. The registered office of Sain Informatix Private Limited is at #1122, ground floor, service road, kalyan housing society,hampinagar Bangalore , Karnataka. sain informatic private limited is a 2.1 Years old company, incorporated on 18 Aug 2020. It is classified as Private UnListed Indian Non-Government Company. Its authorized share capital is ₹10,00,000.00 (₹10.00 Lakhs) and its paid up capital is ₹10,000.00.

CHAPTER 2

2. ABOUT THE COMPANY

Sain Informatix Private Limited is an unlisted private company incorporated on 18 August, 2020. It is classified as a private limited company and is located in , Karnataka. It's authorized share capital is INR 10.00 lac and the total paid-up capital is INR 10,000.00 .

The current status of Sain Informatix Private Limited is - Active.

Details of the last annual general meeting of Sain Informatix Private Limited are not available. The company is yet to submit its first full-year financial statements to the registrar.

Sain Informatix Private Limited has two directors - Menasinkai Mohan Trilok and Trinetra.

The Corporate Identification Number (CIN) of Sain Informatix Private Limited is U72900KA2020PTC137383. The registered office of Sain Informatix Private Limited is at #1122, GROUND FLOOR, SERVICE ROAD, KALYAN HOUSING SOCIETY,HAMPINAGAR BENGALURU Bangalore , Karnataka.

Products of Sain informatix pvt Limited Technologies.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and zutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but softwaredevelopment is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client- side logic) runs in a web browser. Common web apretail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript,which are supported by a variety of web browsers. Web applications can be considered as a specifific variant of client–server software where the client software is downloaded to the client machine when visiting the relevant web page, using standardprocedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, andacts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specifified usecase. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a “smart” client that performs all the work and queries a “dumb” server, or a “dumb” client that relies on a “smart” server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn‟t allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating mark up then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

Sain informatix pvt Limited plays an essential role as an institute, the level of education, development of student's skills are based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Sain informatix pvt Limited gives you the facility of skilled employees so that you do not feel unsecured aboutthe academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Sain informatix pvt Limited always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Sain informatix pvt Limited.

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Training

CHAPTER 3

3. INTRODUCTION

Introduction to ML

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to "self-learn" from training data and improve over time, without being explicitly programmed. Machine learning algorithms are able to detect patterns in data and learn from them, in order to make their own predictions. In short, machine learning algorithms and models learn through experience.

In traditional programming, a computer engineer writes a series of directions that instruct a computer how to transform input data into a desired output.
Instructions are mostly based on an IF-THEN structure: when certain conditions are met, the program executes a specific action.

Problem Statement

Let's go through the problem statement once as it is very crucial to understand the objective before working on the dataset. The problem statement is as follows:

The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets.

Formally, given a training sample of tweets and labels, where label '1' denotes the tweet is racist/sexist and label '0' denotes the tweet is not racist/sexist, your objective is to predict the labels on the given test dataset.

Note: The evaluation metric from this practice problem is F1-Score.

Take a look at the pictures below depicting two scenarios of an office space – one is untidy and the other is clean and organized.

CHAPTER 4

4. SYSTEM ANALYSIS

1. Existing System

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities. From a user's perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. However, those types of online data have several flaws that potentially hinder the process of sentiment analysis. The first flaw is that since people can freely post their own content, the quality of their opinions cannot be guaranteed. he second flaw is that ground truth of such online data is not always available. A ground truth is more like a tag of a certain opinion, indicating whether the opinion is positive, negative, or neutral.

2. Proposed System

The "Sentiment Analysis Of Lockdown In India During Covid-19 A Case Study On Twitter using ML" executed well without any exceptions . The total time complexity and accuracy of the project could be improved with better searching techniques. The outbreak of COVID-19 caused heavy disruption to the everyday lives of people across the globe. In a country like with a large, diverse population like India, there are bound to be instances of mass hysteria and panic which are further fuelled by unreliable and sometimes inaccurate data. Gauging the feelings/emotions of the citizens would provide insights into the public mindset and would pave the way for the government and many organizations to address these situations by providing them with the right data and information, eradicating fake news, thereby helping in suppressing unnecessary panic among the people. Social media acts as the bridge between the people, the government, and such organizations. The scope of this project lies in the application of sentiment analysis to the views expressed by people on social media, twitter, in this case, to analyze the trends in the dynamic mood of the population. Usually, the terms "fight" and "positive" are used in a negative and positive context respectively, but we observe a role reversal in this situation. The identification of such terms and their usage according to the context would be an essential part of the project. Also, the scope of the project can be found in stopping the spread of fake news related to the pandemic, creating an interactive dashboard that delivers information about the current situation, real-time sentiment analysis of tweets, trend analysis of various COVID-19 related hashtags, engagement on Twitter, overall sector-wise polarity score of the tweets and the public emotion charts.

3. Objective of the System

- The project Sentiment Analysis Of Lockdown In India During Covid-19 A Case Study On Twitter using ML has been made to collect the sentiments of the people during lockdown in the time of lockdown in the database in the most efficient way as possible.
- To allow the user to know the sentiments of the people easily with the help of bar graphs, number of counts of people reviews in a single frame .
- This system will show the total number of counts .
- This system will help admin to add doctors, nurse, ward, receptionist, remove them and search them using their id’s.

CHAPTER 5 : REQUIREMENT ANALYSIS

Hardware Requirement Specification

- Memory and disk space required per user: 1GB RAM + 1GB of disk + . 5 CPUcore.
- Server overhead: 2-4GB or 10% system overhead (whatever is
- larger), . 5 CPUcores, 10GB disk space.
- Port requirements: Port 8000 plus 5 unique, random ports per
- notebook.
-

Software Requirement Specification

- Jupyter notebook
- Python ide
- Command prompt
- browser

CHAPTER 6 : DESIGN & ANALYSIS

Sentiment analysis is a machine learning tool that analyses texts for polarity, from positive to negative. By training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.

To put it simply, machine learning allows computers to learn new tasks without being expressly programmed to perform them. Sentiment analysis models can be trained to read beyond mere definitions, to understand things like, context, sarcasm, and misapplied words. For example:

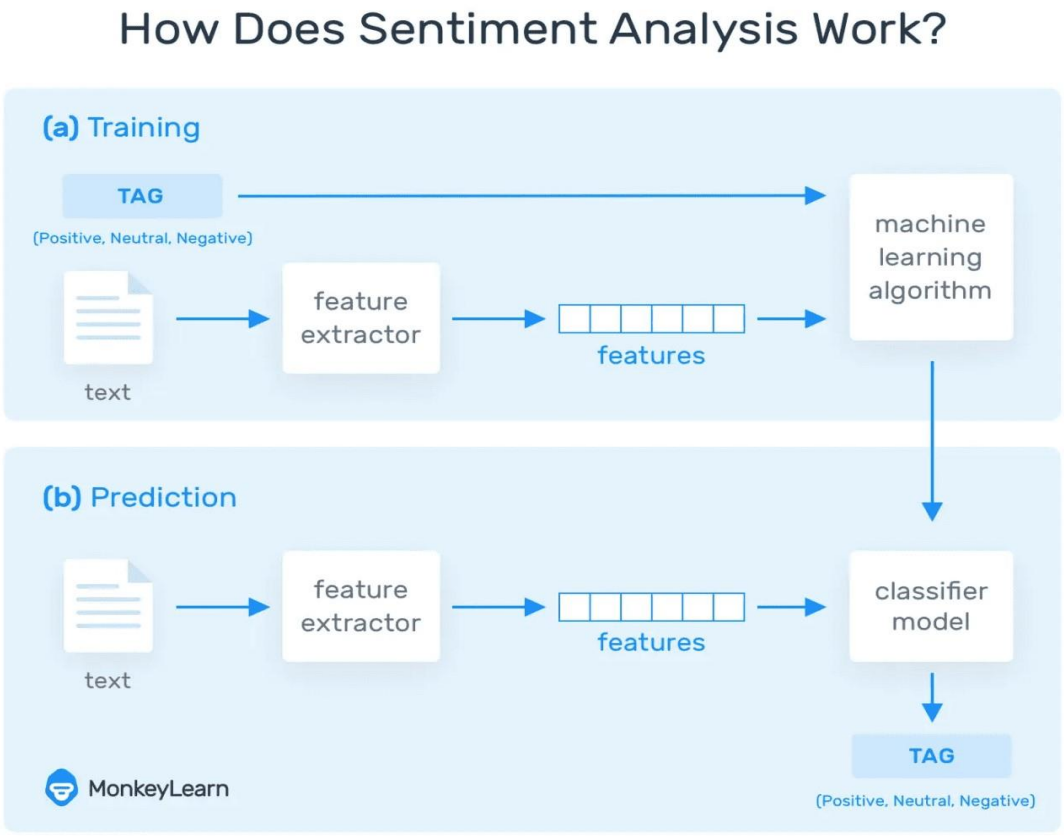
“Super user-friendly interface. Yeah right. An engineering degree would be helpful.”

Out of context, the words ‘super user-friendly’ and ‘helpful’ could be read as positive, but this is clearly a negative comment. Using sentiment analysis, computers can automatically process text data and understand it just as a human would, saving hundreds of employee hours.

Imagine using machine learning to process customer service tickets, categorize them in order of urgency, and automatically route them to the correct department or employee. Or, to analyze thousands of product reviews and social media posts to gauge brand sentiment.

Read on to learn more about how machine learning works and how it can help your business

There are a number of techniques and complex algorithms used to command and train machines to perform sentiment analysis. There are pros and cons to each. But, used together, they can provide exceptional results. Below are some of the most used algorithm.



The Naive Bayes is a fairly simple group of probabilistic algorithms that, for sentiment analysis classification, assigns a probability that a given word or phrase should be considered positive or negative. Essentially, this is how Bayes' theorem works. The probability of A, if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true. But that's a lot of math! Basically, Naive Bayes calculates words against each other. So, with machine learning models trained for word polarity, we can calculate the likelihood that a word, phrase, or text is positive or negative. When techniques like lemmatization, stopword removal, and TF-IDF are implemented, Naive Bayes becomes more and more predictively accurate. The amount of sentimental content available in the Web. Such content is often found in social media web sites in the form of movie or product reviews, user comments, testimonials, messages in discussion forums etc. Timely discovery of the sentimental or opinionated web content has a number of advantages, the most important of all being monetization. Understanding of the sentiments of human masses towards different entities and products enables better services for contextual advertisements, recommendation systems and analysis of market trends. The focus of our project is sentiment focussed web crawling framework to facilitate the quick discovery of sentimental contents of movie reviews and hotel reviews and analysis of the same. We use statistical methods to capture elements of subjective style and the sentence polarity. The paper elaborately discusses two supervised machine learning algorithms: K-Nearest Neighbour(K-NN) and Naive Bayes and compares their overall accuracy, precisions as well as recall values. It was seen that in case of movie reviews Naive Bayes gave far better results than KNN but for hotel reviews these algorithms gave lesser, almost same accuracies.

Types of Sentiment Analysis

Sentiment analysis focuses on the polarity of a text (*positive, negative, neutral*) but it also goes beyond polarity to detect specific feelings and emotions (*angry*, *happy*, *sad*, etc), urgency (*urgent, not urgent*) and even intentions (*interested v. not interested*).

Depending on how you want to interpret customer feedback and queries, you can define and tailor your categories to meet your sentiment analysis needs. In the meantime, here are some of the most popular types of sentiment analysis:

Graded Sentiment Analysis

If polarity precision is important to your business, you might consider expanding your polarity categories to include different levels of positive and negative:

- Positive
- Neutral
- Negative

Emotion detection sentiment analysis allows you to go beyond polarity to detect emotions, like happiness, frustration, anger, and sadness.

Many emotion detection systems use lexicons (i.e. lists of words and the emotions they convey) or complex machine learning algorithms.

One of the downsides of using lexicons is that people express emotions in different ways. Some words that typically express anger, like *bad* or *kill* (e.g. *your product is so bad* or *your customer support is killing me*) might also express happiness (e.g. *this is bad ass* or *you are killing it*).

Aspect-based Sentiment Analysis

Usually, when analyzing sentiments of texts you'll want to know which particular aspects or features people are mentioning in a positive, neutral, or negative way.

That's where aspect-based sentiment analysis can help, for example in this product review: *"The battery life of this camera is too short"*, an aspect-based classifier would be able to determine that the sentence expresses a negative opinion about the battery life of the product in question.

Multilingual sentiment analysis

Multilingual sentiment analysis can be difficult. It involves a lot of preprocessing and resources. Most of these resources are available online (e.g. sentiment lexicons), while others need to be created (e.g. translated corpora or noise detection algorithms), but you'll need to know how to code to use them.

Alternatively, you could detect language in texts automatically with a language classifier, then train a custom sentiment analysis model to classify texts in the language of your choice.

Why Is Sentiment Analysis Important?

Since humans express their thoughts and feelings more openly than ever before, sentiment analysis is fast becoming an essential tool to monitor and understand sentiment in all types of data.

Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

For example, using sentiment analysis to automatically analyze 4,000+ open-ended responses in your customer satisfaction surveys could help you discover why customers are happy or unhappy at each stage of the customer journey.

Maybe you want to track brand sentiment so you can detect disgruntled customers immediately and respond as soon as possible. Maybe you want to compare sentiment from one quarter to the next to see if you need to take action. Then you could dig deeper into your qualitative data to see why sentiment is falling or rising.

CHAPTER 6 : IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

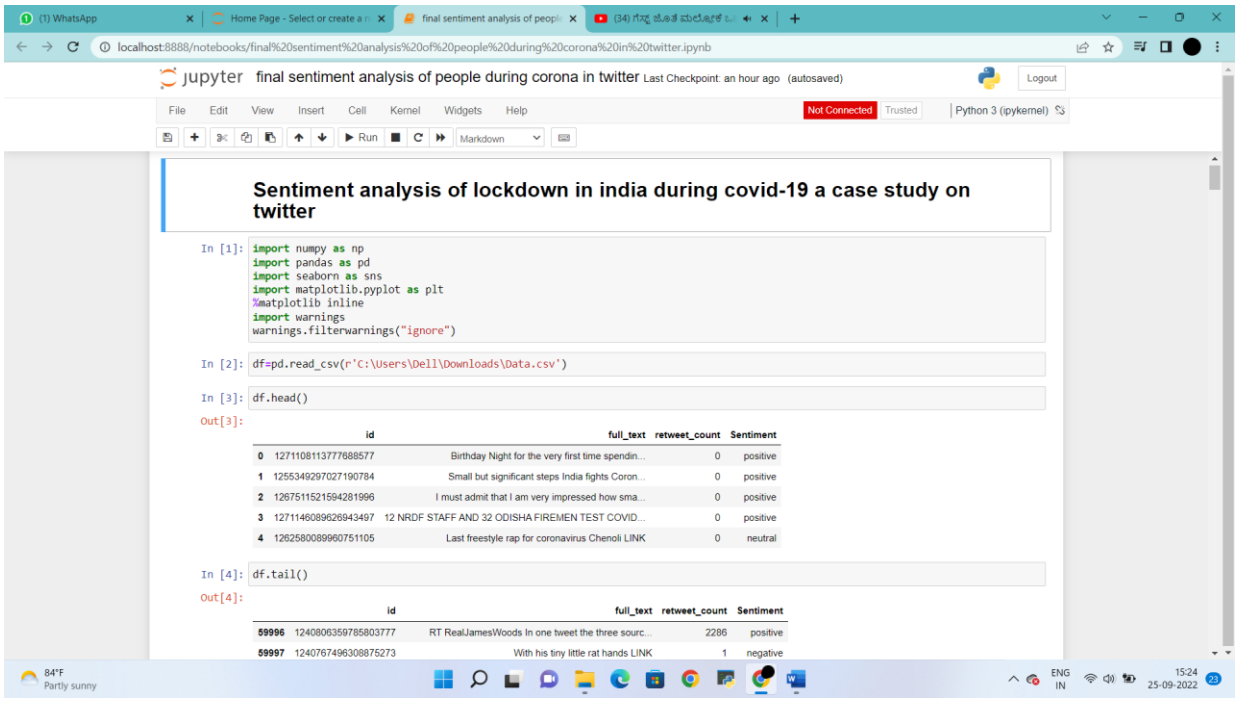
The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

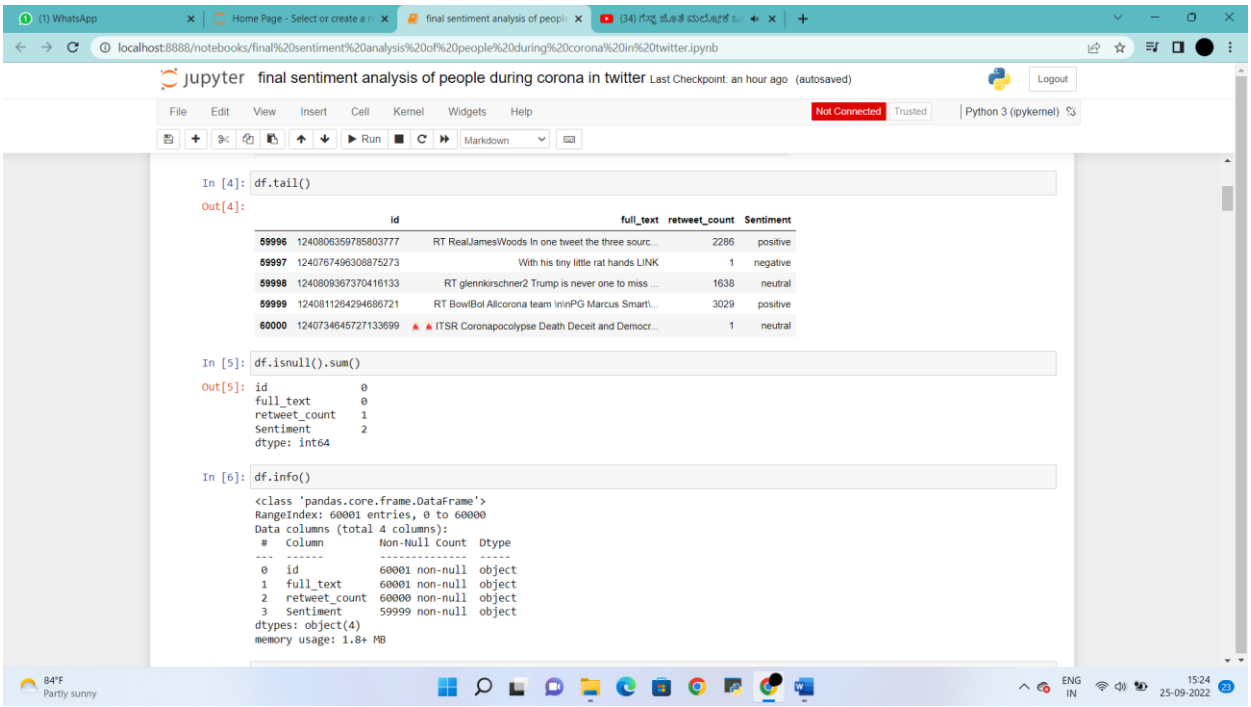
TESTING

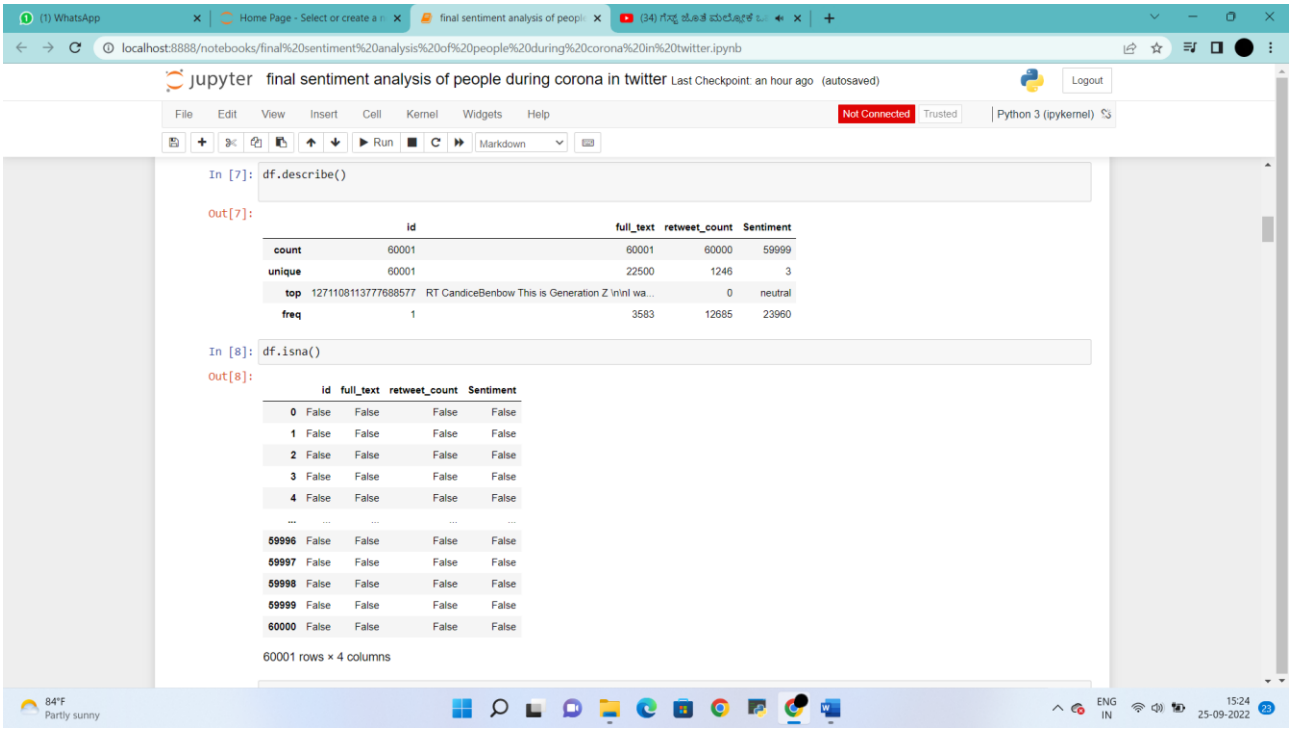
The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

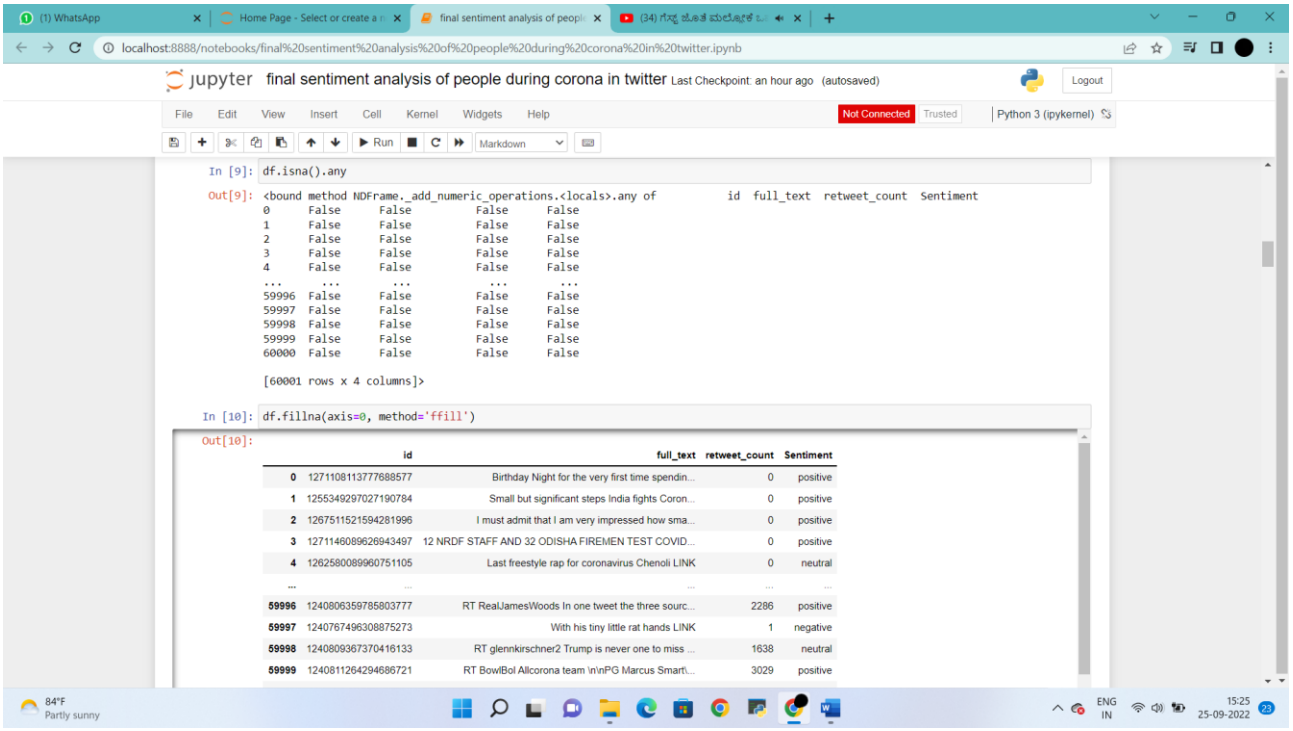
1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether theobjectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

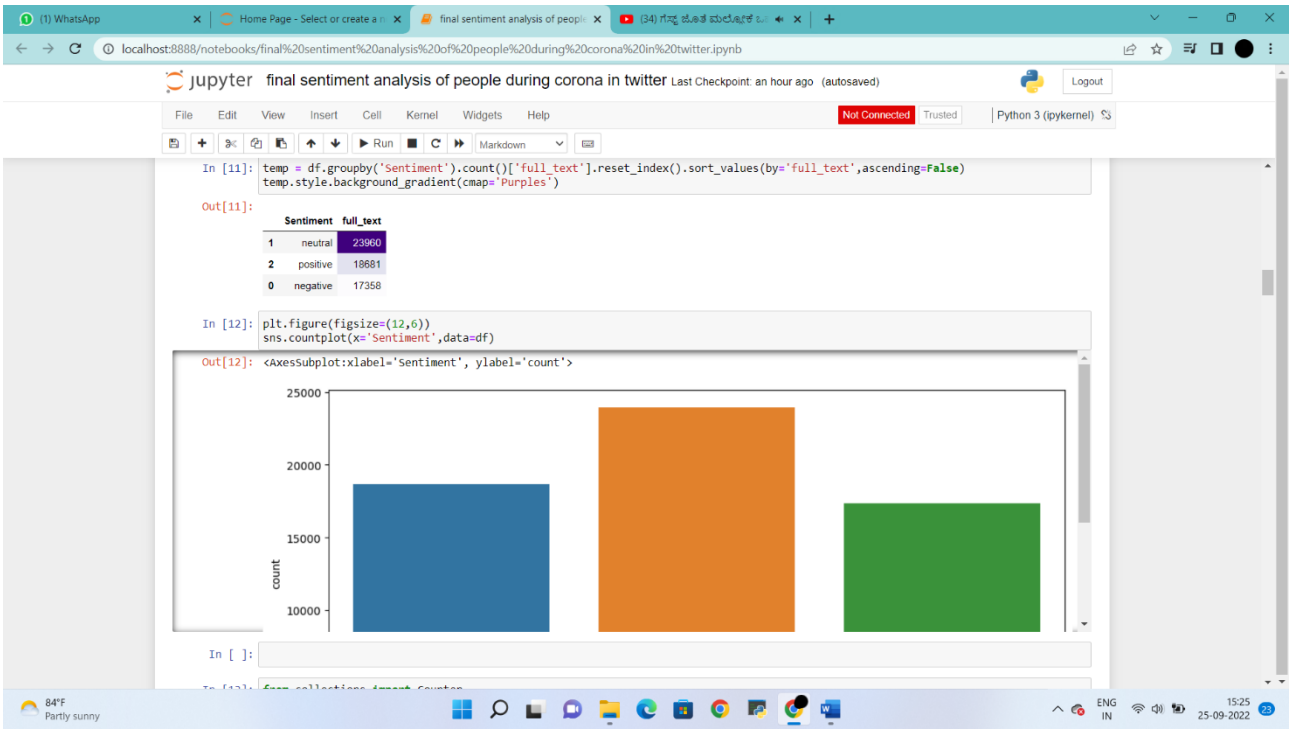
CHAPTER 7 : SNAPSHOTS

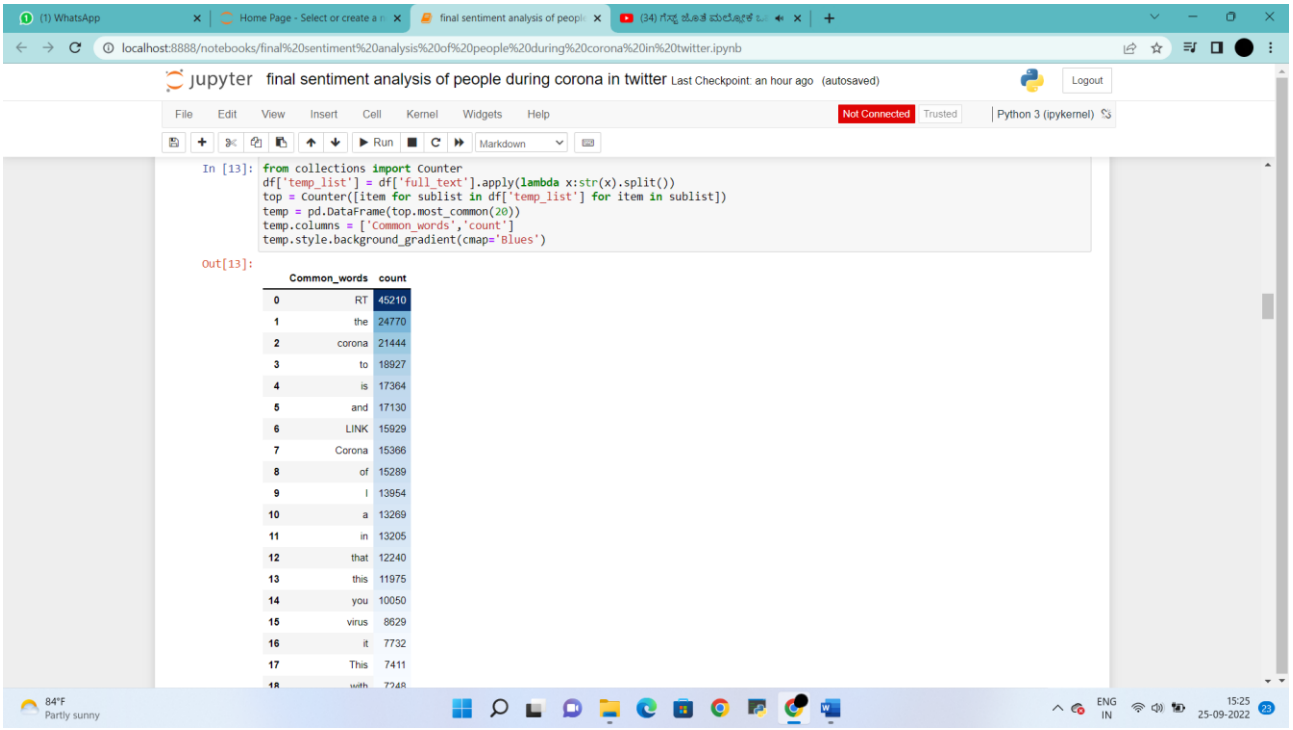


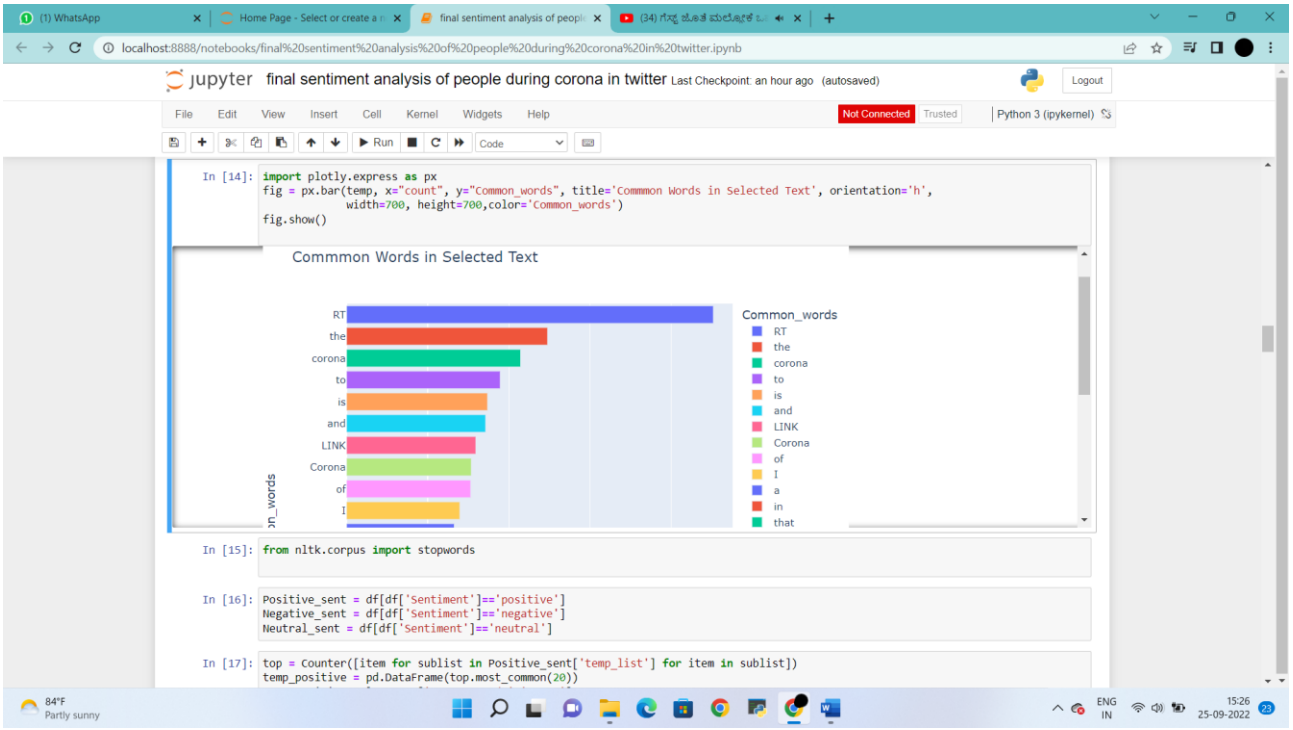


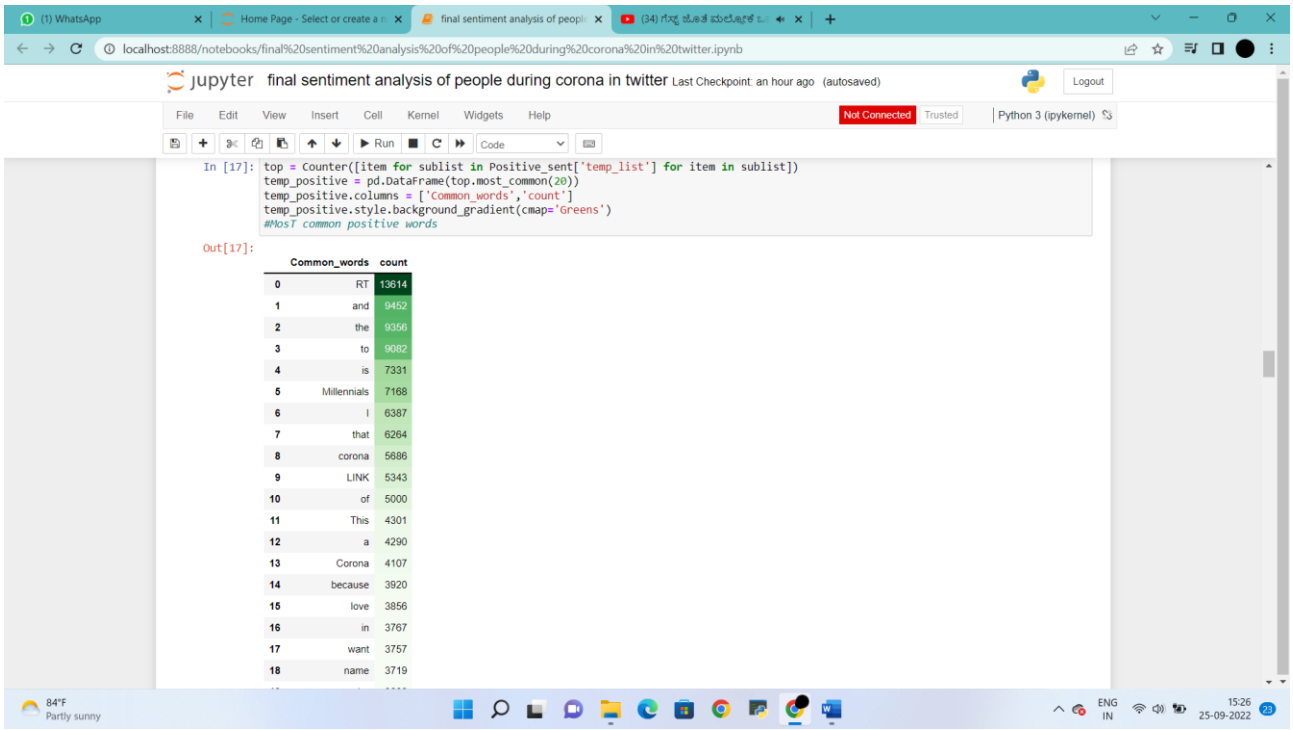

localhost:8888/notebooks/final%20sentiment%20analysis%20of%20people%20during%20corona%20in%20twitter.ipynb
jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected Trusted Python 3 (ipykernel)
Run Markdown
In [7]: df.describe()
Out[7]:
id full_text retweet_count Sentiment
count 60001 60001 60000 59999
unique 60001 22500 1246 3
top 1271108113777688577 RT CandiceBenbow This is Generation Z \n\nI wa... 0 neutral
freq 1 3583 12685 23960
In [8]: df.isna()
Out[8]:
id full_text retweet_count Sentiment
0 False False False False
1 False False False False
2 False False False False
3 False False False False
4 False False False False
59996 False False False False
59997 False False False False
59998 False False False False
59999 False False False False
60000 False False False False
60001 rows × 4 columns
84°F Partly sunny
ENG IN
15:24 25-09-2022

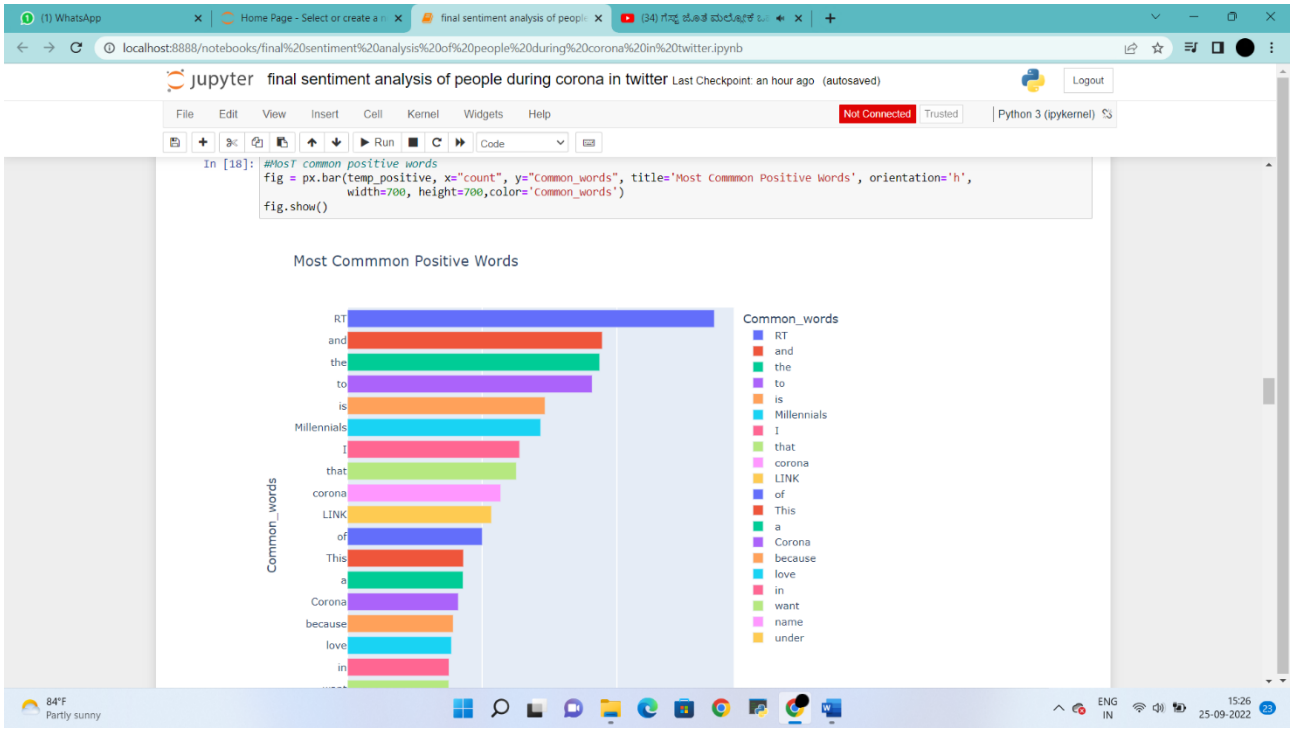

localhost:8888/notebooks/final%20sentiment%20analysis%20of%20people%20during%20corona%20in%20twitter.ipynb
jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected Trusted Python 3 (ipykernel)
Run Markdown
In [9]: df.isna().any
Out[9]: <bound method NDFrame._add_numeric_operations.<locals>.any of id full_text retweet_count Sentiment
0 False False False False
1 False False False False
2 False False False False
3 False False False False
4 False False False False
... ...
59996 False False False False
59997 False False False False
59998 False False False False
59999 False False False False
60000 False False False False
[60001 rows x 4 columns]>
In [10]: df.fillna(axis=0, method='ffill')
Out[10]:
id full_text retweet_count Sentiment
0 1271108113777688577 Birthday Night for the very first time spendin... 0 positive
1 1255349297027190784 Small but significant steps India fights Coron... 0 positive
2 1267511521594281996 I must admit that I am very impressed how sma... 0 positive
3 1271146089626943497 12 NRDF STAFF AND 32 ODISHA FIREMEN TEST COVID... 0 positive
4 1262580089960751105 Last freestyle rap for coronavirus Chenoli LINK 0 neutral
... ...
59996 1240806359785803777 RT RealJamesWoods In one tweet the three sourc... 2286 positive
59997 1240767496308875273 With his tiny little rat hands LINK 1 negative
59998 1240809367370416133 RT glennkirschner2 Trump is never one to miss ... 1638 neutral
59999 1240811264294686721 RT BowlBol Allcorona team \n\nPG Marcus Smart\... 3029 positive
84°F Partly sunny
ENG IN
15:25 25-09-2022

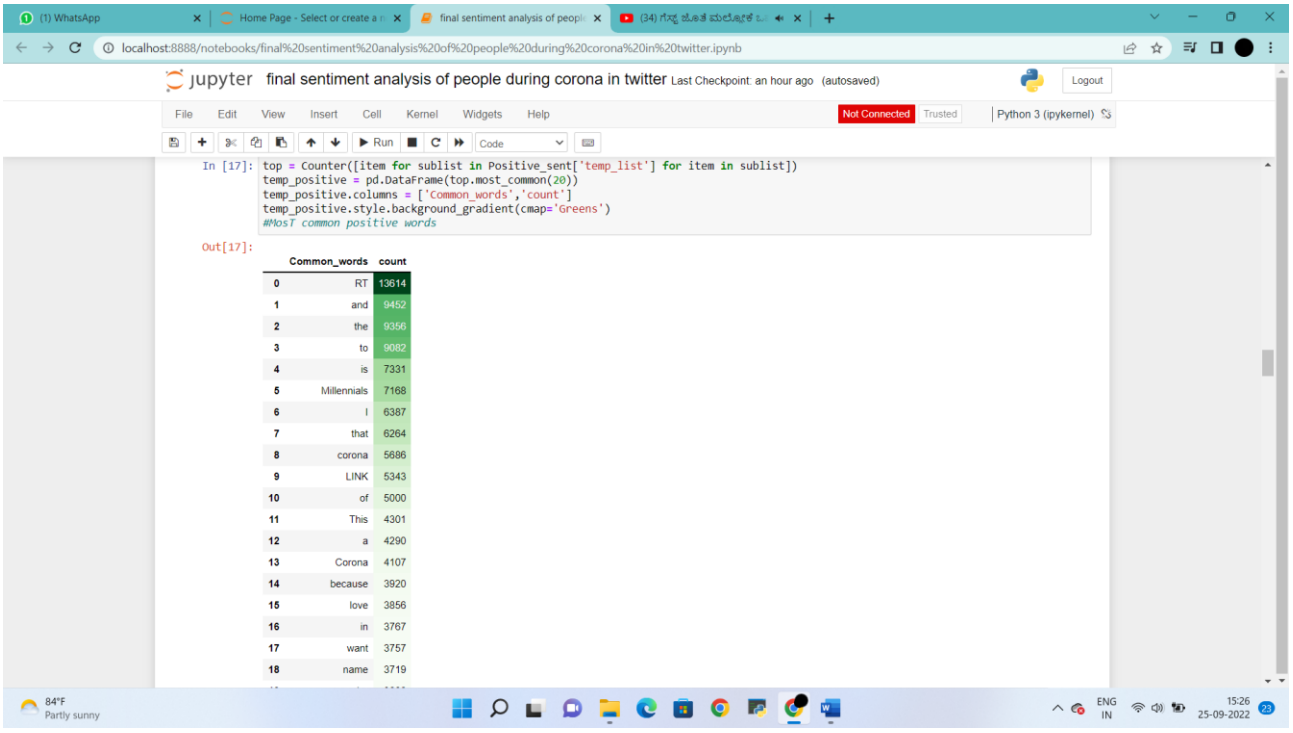

final sentiment analysis of people during corona in twitter
In [11]:
temp = df.groupby('Sentiment').count()['full_text'].reset_index().sort_values(by='full_text',ascending=False)
temp.style.background_gradient(cmap='Purples')
Out[11]:
Sentiment full_text
1 neutral 23960
2 positive 18681
0 negative 17358
In [12]:
plt.figure(figsize=(12,6))
sns.countplot(x='Sentiment',data=df)
Out[12]:
<AxesSubplot:xlabel='Sentiment', ylabel='count'>
25000
20000
15000
10000
count
In [ ]:

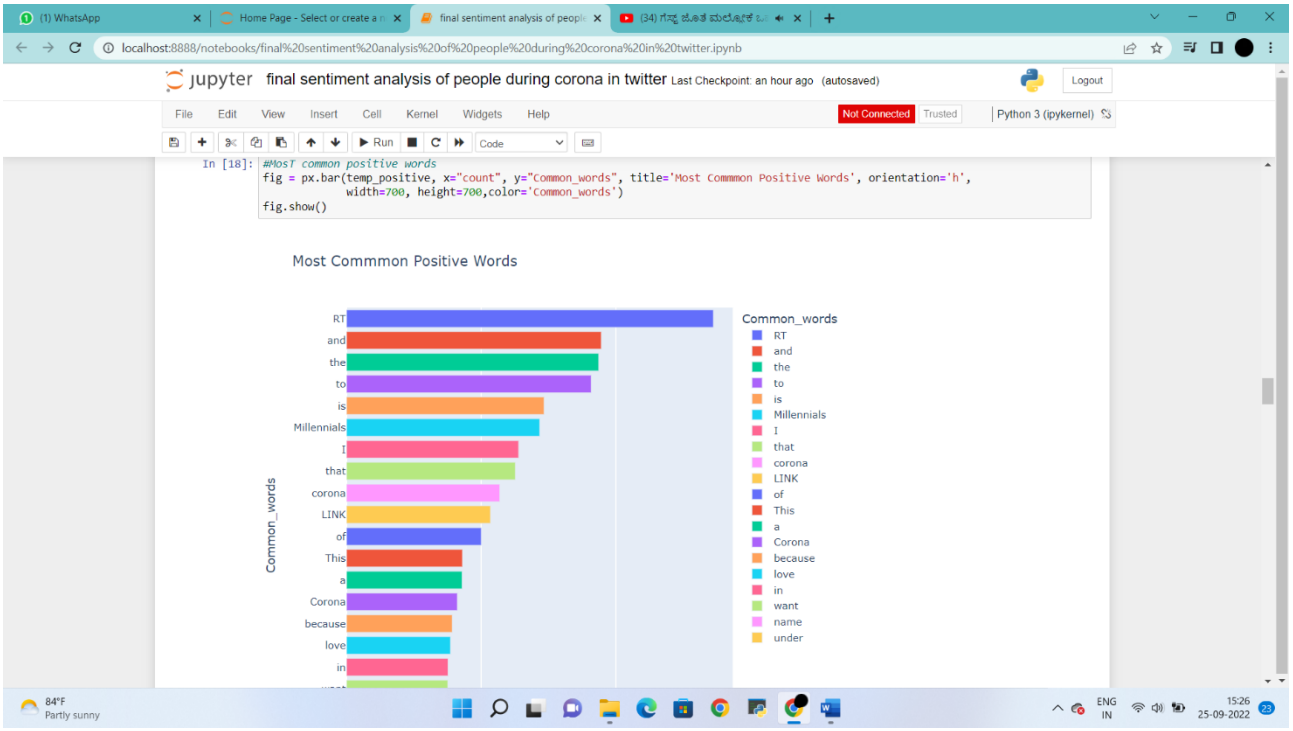

final sentiment analysis of people during corona in twitter
In [13]:
from collections import Counter
df['temp_list'] = df['full_text'].apply(lambda x:str(x).split())
top = Counter([item for sublist in df['temp_list'] for item in sublist])
temp = pd.DataFrame(top.most_common(20))
temp.columns = ['Common_words','count']
temp.style.background_gradient(cmap='Blues')
Out[13]:
Common_words count
0 RT 45210
1 the 24770
2 corona 21444
3 to 18927
4 is 17364
5 and 17130
6 LINK 15929
7 Corona 15366
8 of 15289
9 I 13954
10 a 13269
11 in 13205
12 that 12240
13 this 11975
14 you 10050
15 virus 8629
16 it 7732
17 This 7411

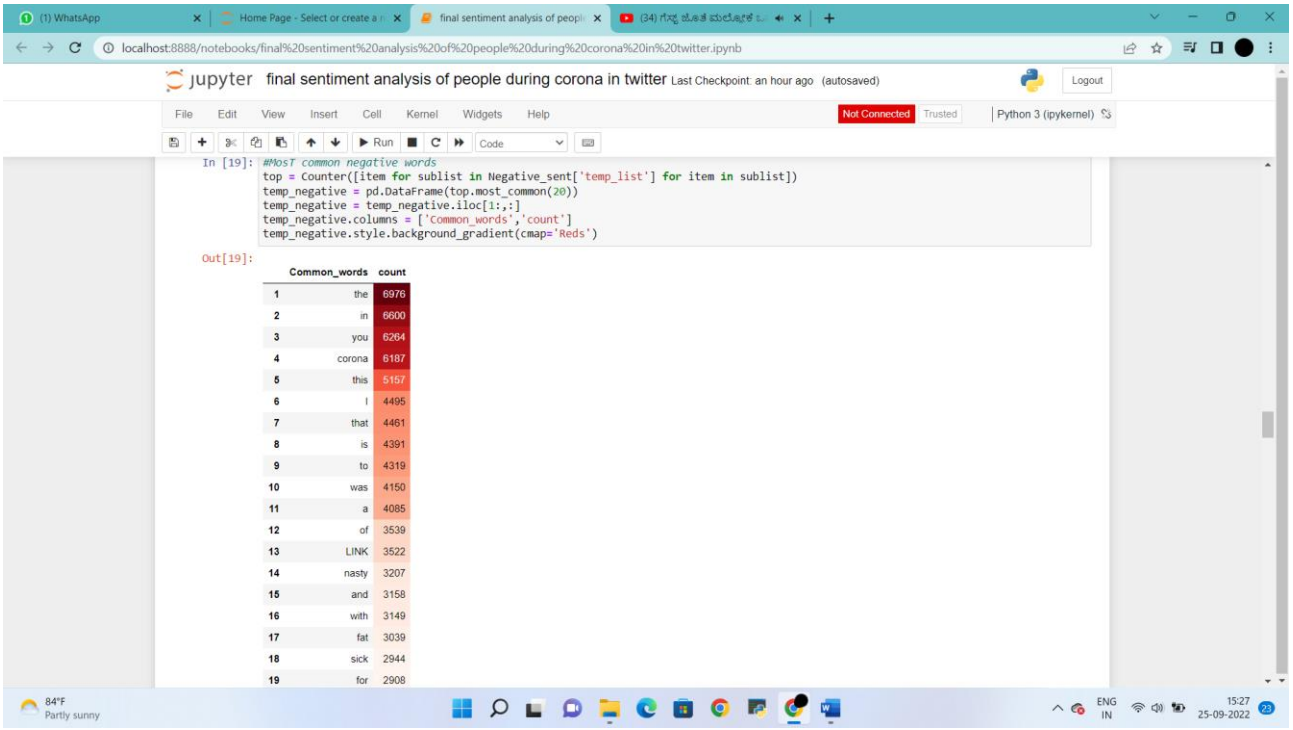

jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected Trusted Python 3 (ipykernel)
Run Code
In [14]:
import plotly.express as px
fig = px.bar(temp, x="count", y="Common_words", title='Commmon Words in Selected Text', orientation='h',
width=700, height=700,color='Common_words')
fig.show()
Commmon Words in Selected Text
RT
the
corona
to
is
and
LINK
Corona
of
I
Common_words
RT
the
corona
to
is
and
LINK
Corona
of
I
a
in
that
In [15]:
from nltk.corpus import stopwords
In [16]:
Positive_sent = df[df['Sentiment']=='positive']
Negative_sent = df[df['Sentiment']=='negative']
Neutral_sent = df[df['Sentiment']=='neutral']
In [17]:
top = Counter([item for sublist in Positive_sent['temp_list'] for item in sublist])
temp_positive = pd.DataFrame(top.most_common(20))
84°F
Partly sunny
ENG
IN
15:26
25-09-2022

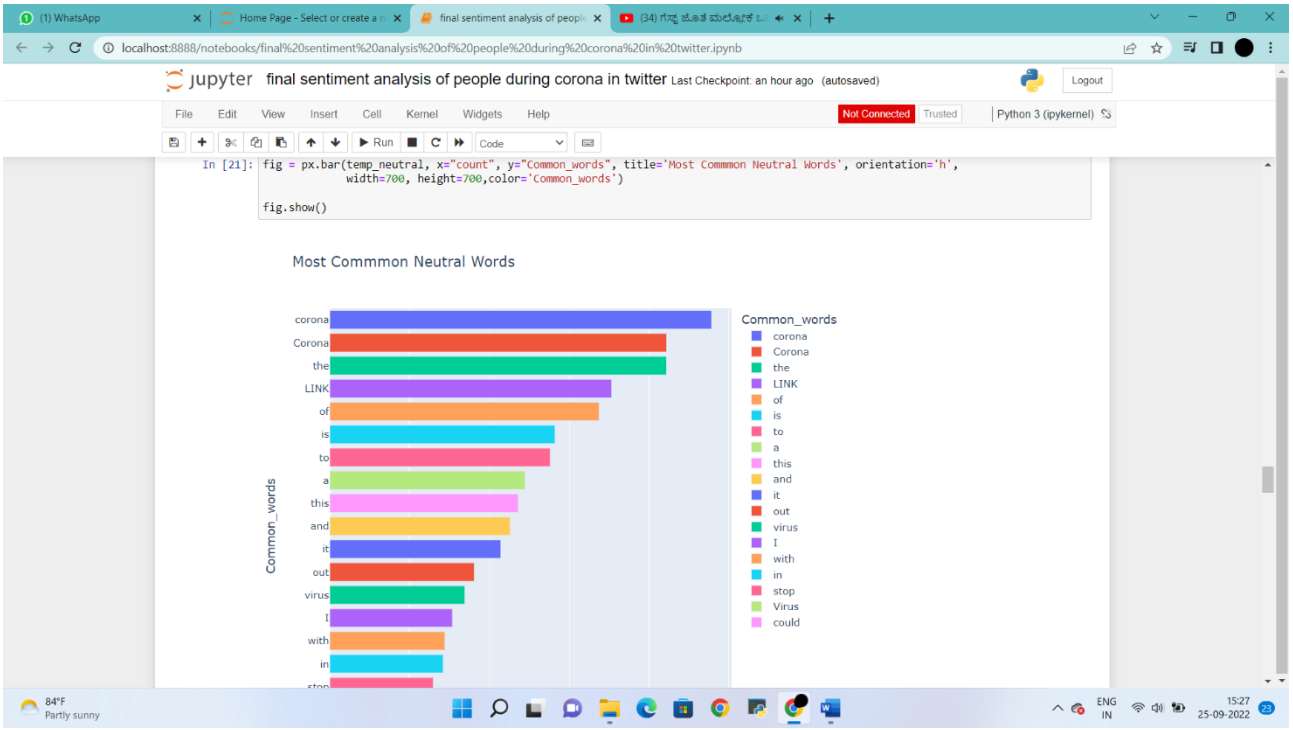

jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected Trusted Python 3 (ipykernel)
Run Code
In [17]:
top = Counter([item for sublist in Positive_sent['temp_list'] for item in sublist])
temp_positive = pd.DataFrame(top.most_common(20))
temp_positive.columns = ['Common_words','count']
temp_positive.style.background_gradient(cmap='Greens')
#Most common positive words
Out[17]:
Common_words count
0 RT 13614
1 and 9452
2 the 9356
3 to 9082
4 is 7331
5 Millennials 7168
6 I 6387
7 that 6264
8 corona 5686
9 LINK 5343
10 of 5000
11 This 4301
12 a 4290
13 Corona 4107
14 because 3920
15 love 3856
16 in 3767
17 want 3757
18 name 3719
84°F
Partly sunny
ENG
IN
15:26
25-09-2022

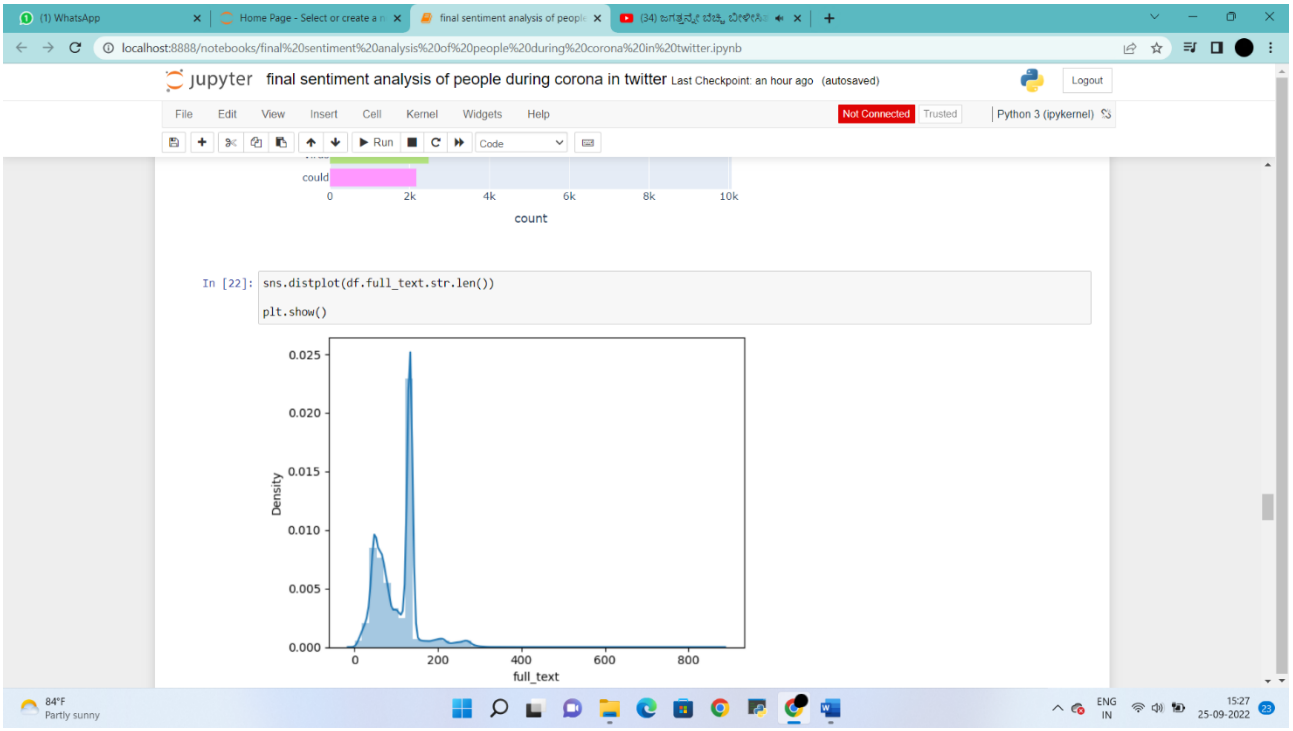

final sentiment analysis of people during corona in twitter
In [18]:
#Most common positive words
fig = px.bar(temp_positive, x="count", y="Common_words", title='Most Commmon Positive Words', orientation='h',
width=700, height=700,color='Common_words')
fig.show()
Most Commmon Positive Words
Common_words
RT
and
the
to
is
Millennials
I
that
corona
LINK
of
This
a
Corona
because
love
in
want
name
under

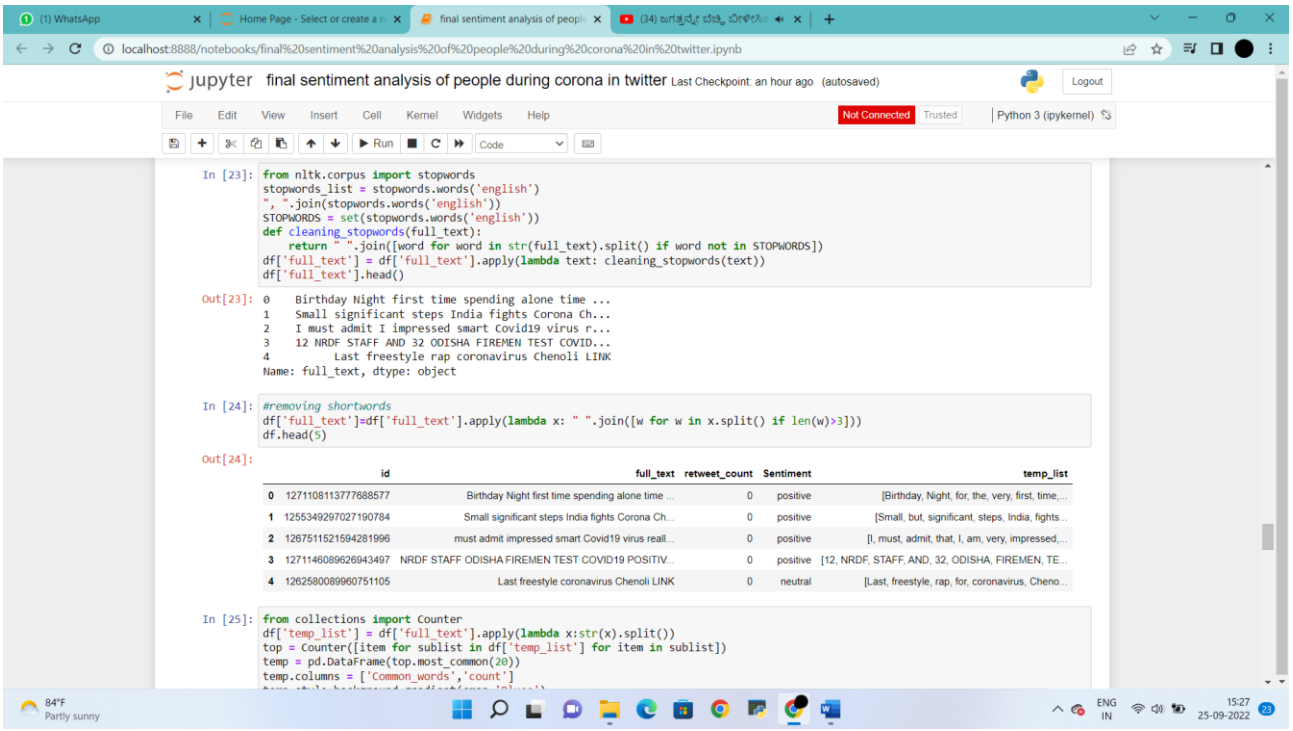

final sentiment analysis of people during corona in twitter
In [17]:
top = Counter([item for sublist in Positive_sent['temp_list'] for item in sublist])
temp_positive = pd.DataFrame(top.most_common(20))
temp_positive.columns = ['Common_words','count']
temp_positive.style.background_gradient(cmap='Greens')
#Most common positive words
Out[17]:
Common_words count
0 RT 13614
1 and 9452
2 the 9356
3 to 9082
4 is 7331
5 Millennials 7168
6 I 6387
7 that 6264
8 corona 5686
9 LINK 5343
10 of 5000
11 This 4301
12 a 4290
13 Corona 4107
14 because 3920
15 love 3856
16 in 3767
17 want 3757
18 name 3719

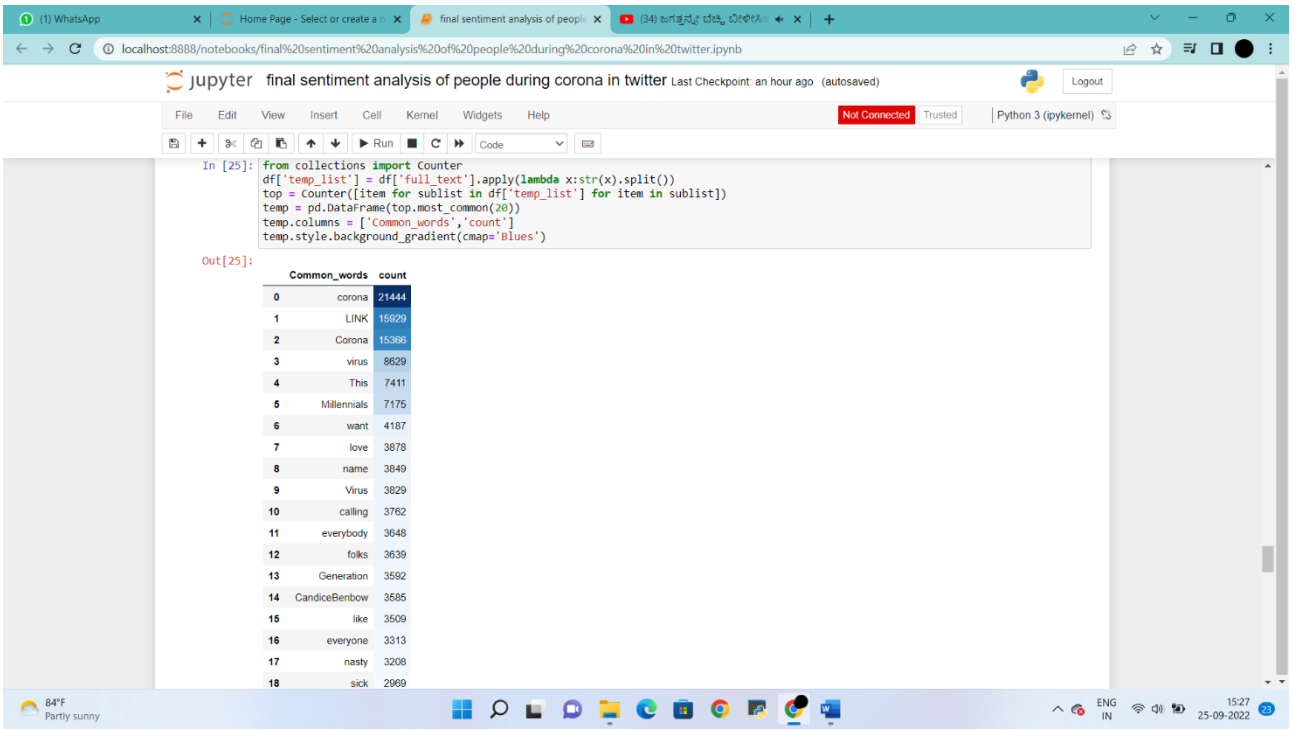

final sentiment analysis of people during corona in twitter
In [18]:
#Most common positive words
fig = px.bar(temp_positive, x="count", y="Common_words", title='Most Commmon Positive Words', orientation='h',
width=700, height=700,color='Common_words')
fig.show()
Most Commmon Positive Words
Common_words
RT
and
the
to
is
Millennials
I
that
corona
LINK
of
This
a
Corona
because
love
in
want
name
under

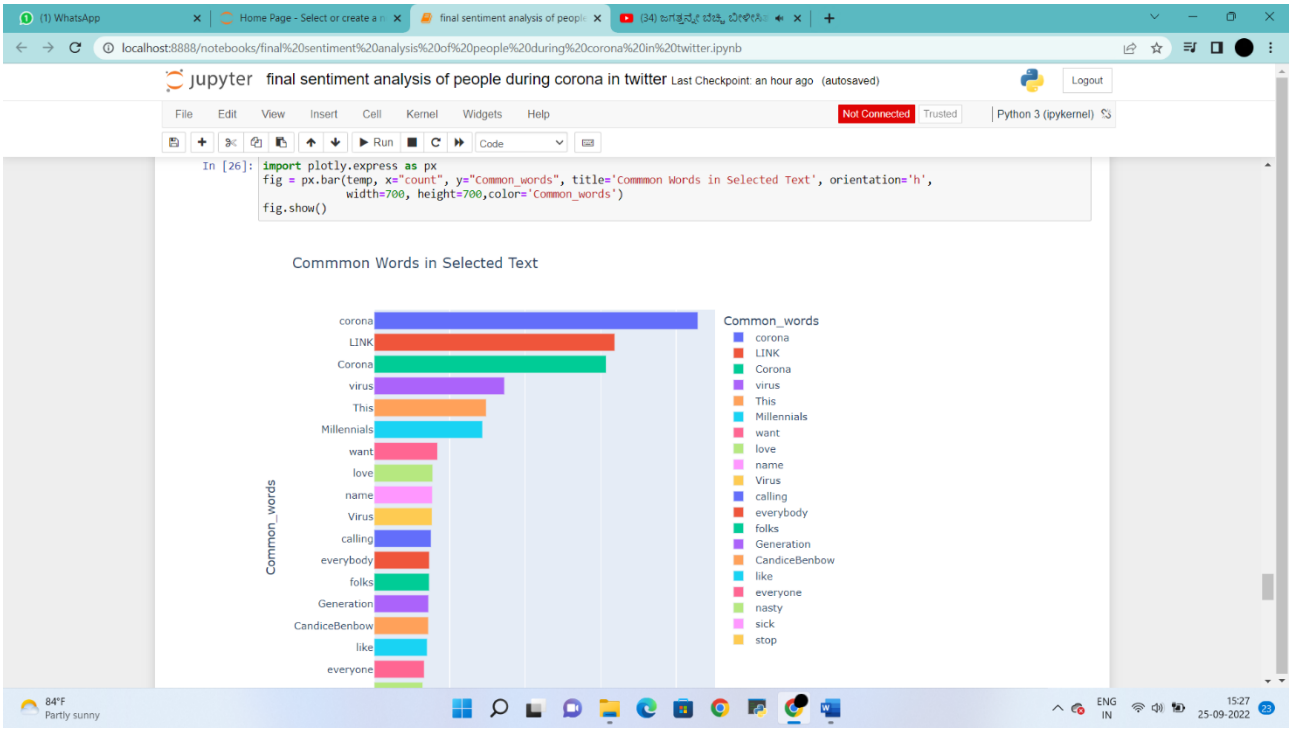

final sentiment analysis of people during corona in twitter
In [19]:
#Most common negative words
top = Counter([item for sublist in Negative_sent['temp_list'] for item in sublist])
temp_negative = pd.DataFrame(top.most_common(20))
temp_negative = temp_negative.iloc[1:,:]
temp_negative.columns = ['Common_words','count']
temp_negative.style.background_gradient(cmap='Reds')
Out[19]:
Common_words count
1 the 6976
2 in 6600
3 you 6264
4 corona 6187
5 this 5157
6 I 4495
7 that 4461
8 is 4391
9 to 4319
10 was 4150
11 a 4085
12 of 3539
13 LINK 3522
14 nasty 3207
15 and 3158
16 with 3149
17 fat 3039
18 sick 2944
19 for 2908

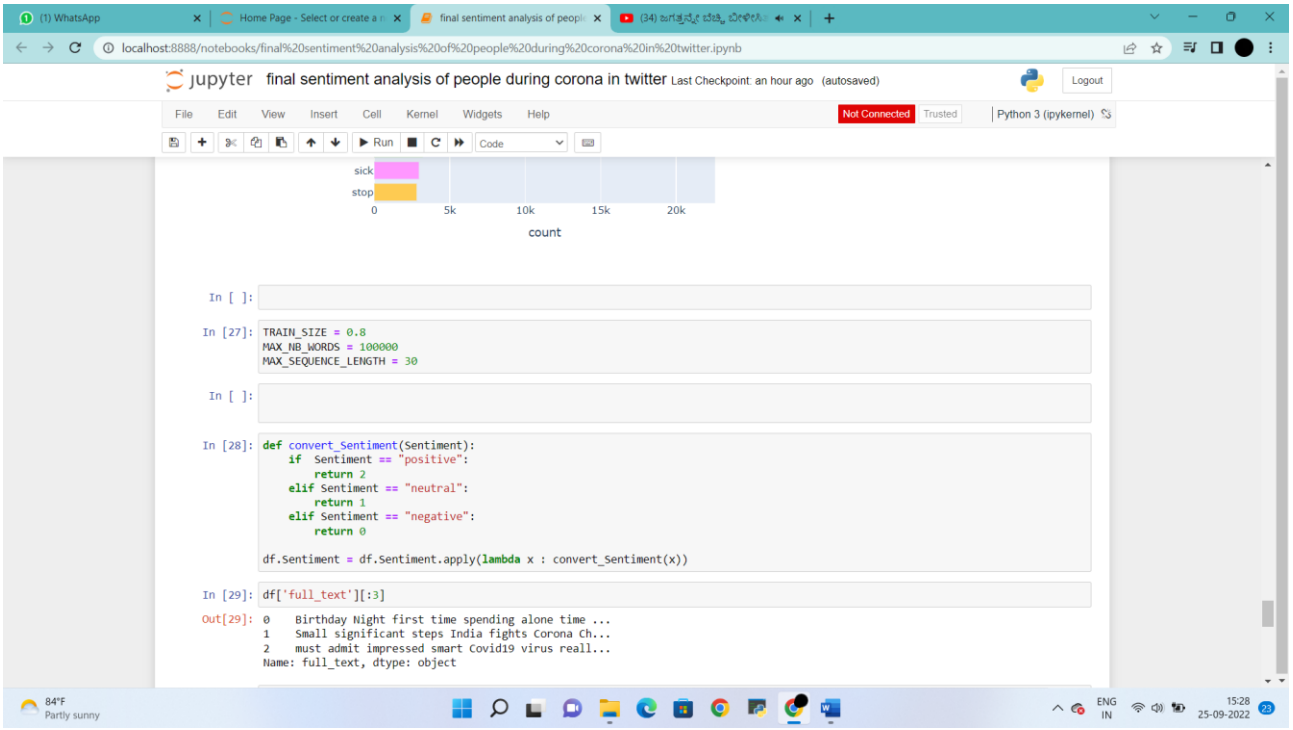

final sentiment analysis of people during corona in twitter
In [21]:
fig = px.bar(temp_neutral, x="count", y="Common_words", title='Most Commmon Neutral Words', orientation='h',
width=700, height=700,color='Common_words')
fig.show()
Most Commmon Neutral Words
Common_words
corona
Corona
the
LINK
of
is
to
a
this
and
it
out
virus
I
with
in
stop
Virus
could

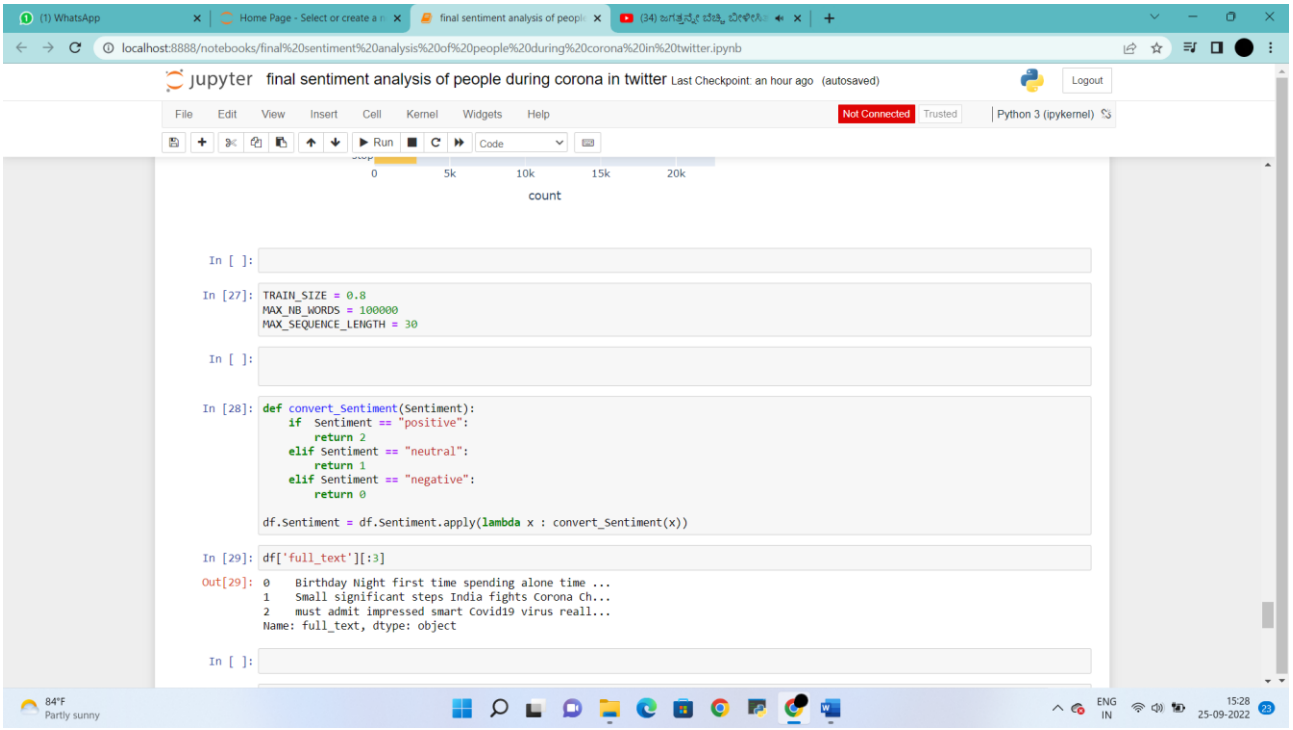

final sentiment analysis of people during corona in twitter
could
0
2k
4k
6k
8k
10k
count
In [22]:
sns.distplot(df.full_text.str.len())
plt.show()
Density
0.025
0.020
0.015
0.010
0.005
0.000
0
200
400
600
800
full_text

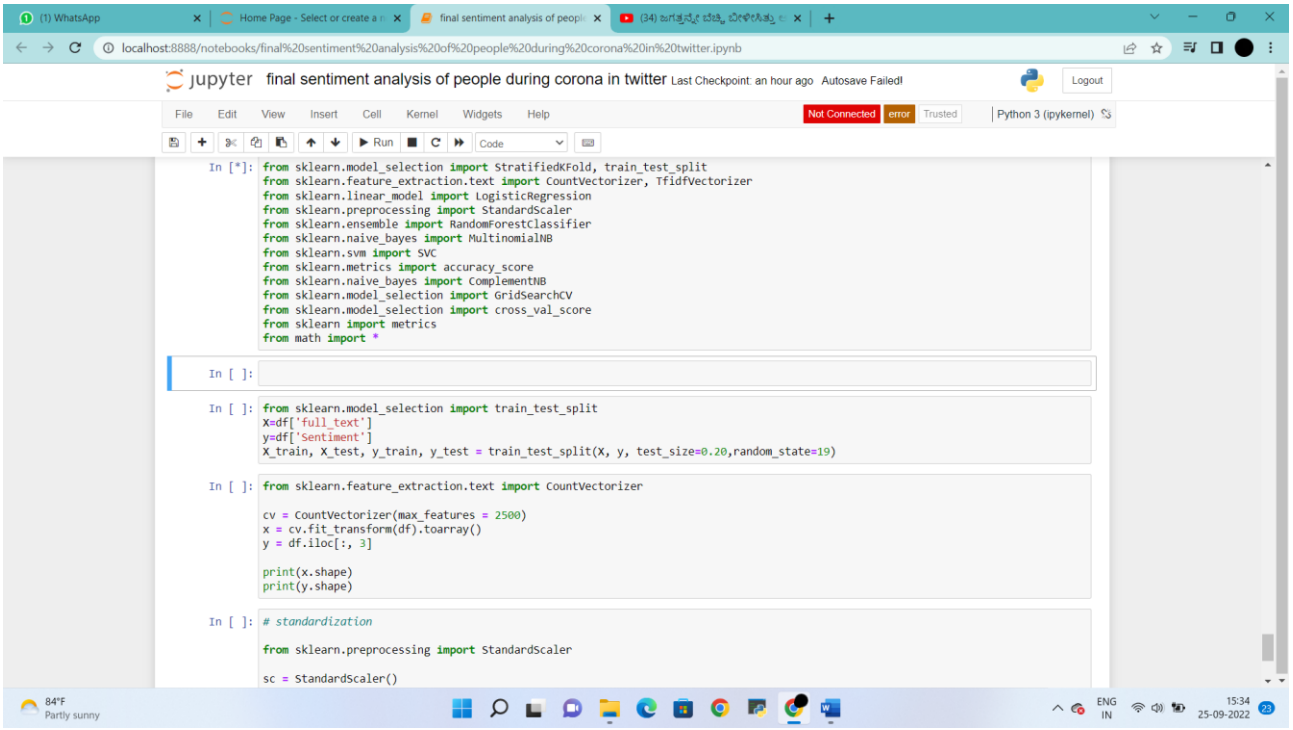

final sentiment analysis of people during corona in twitter
In [23]:
from nltk.corpus import stopwords
stopwords_list = stopwords.words('english')
", ".join(stopwords.words('english'))
STOPWORDS = set(stopwords.words('english'))
def cleaning_stopwords(full_text):
return " ".join([word for word in str(full_text).split() if word not in STOPWORDS])
df['full_text'] = df['full_text'].apply(lambda text: cleaning_stopwords(text))
df['full_text'].head()
Out[23]:
0 Birthday Night first time spending alone time ...
1 Small significant steps India fights Corona Ch...
2 I must admit I impressed smart Covid19 virus r...
3 12 NRDF STAFF AND 32 ODISHA FIREMEN TEST COVID...
4 Last freestyle rap coronavirus Chenoli LINK
Name: full_text, dtype: object
In [24]:
#removing shortwords
df['full_text']=df['full_text'].apply(lambda x: " ".join([w for w in x.split() if len(w)>3]))
df.head(5)
Out[24]:
id full_text retweet_count Sentiment temp_list
0 1271108113777688577 Birthday Night first time spending alone time ... 0 positive [Birthday, Night, for, the, very, first, time,...
1 1255349297027190784 Small significant steps India fights Corona Ch... 0 positive [Small, but, significant, steps, India, fights...
2 1267511521594281996 must admit impressed smart Covid19 virus reall... 0 positive [I, must, admit, that, I, am, very, impressed,...
3 1271146089626943497 NRDF STAFF ODISHA FIREMEN TEST COVID19 POSITIV... 0 positive [12, NRDF, STAFF, AND, 32, ODISHA, FIREMEN, TE...
4 1262580089960751105 Last freestyle coronavirus Chenoli LINK 0 neutral [Last, freestyle, rap, for, coronavirus, Cheno...
In [25]:
from collections import Counter
df['temp_list'] = df['full_text'].apply(lambda x:str(x).split())
top = Counter([item for sublist in df['temp_list'] for item in sublist])
temp = pd.DataFrame(top.most_common(20))
temp.columns = ['Common_words','count']

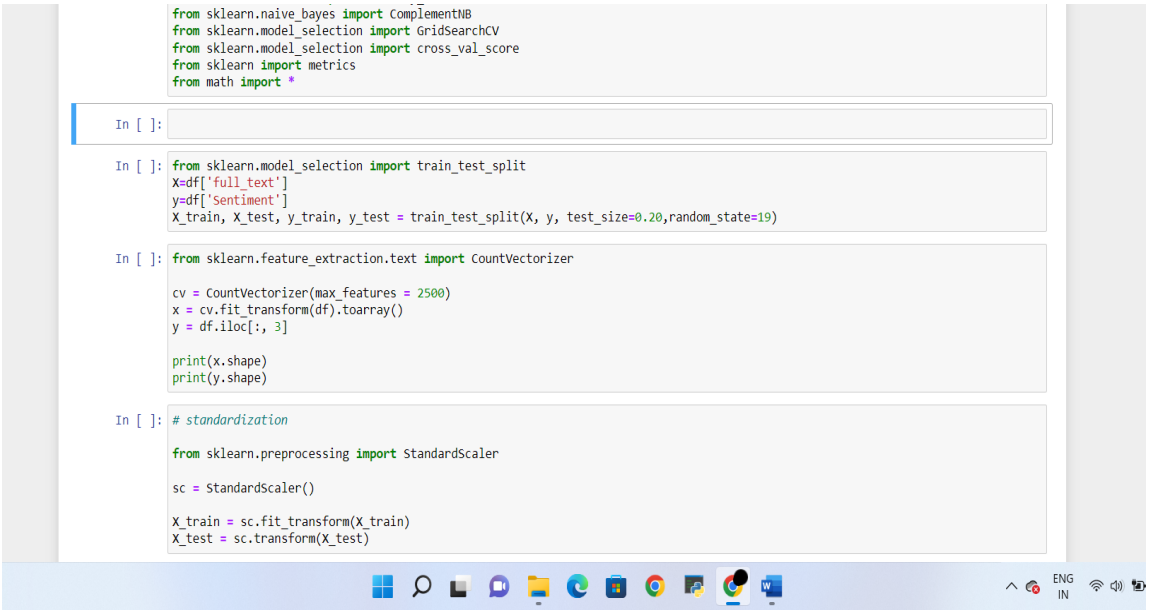

final sentiment analysis of people during corona in twitter
In [25]:
from collections import Counter
df['temp_list'] = df['full_text'].apply(lambda x:str(x).split())
top = Counter([item for sublist in df['temp_list'] for item in sublist])
temp = pd.DataFrame(top.most_common(20))
temp.columns = ['Common_words','count']
temp.style.background_gradient(cmap='Blues')
Out[25]:
Common_words count
0 corona 21444
1 LINK 15929
2 Corona 15366
3 virus 8629
4 This 7411
5 Millennials 7175
6 want 4187
7 love 3878
8 name 3849
9 Virus 3829
10 calling 3762
11 everybody 3648
12 folks 3639
13 Generation 3592
14 CandiceBenbow 3585
15 like 3509
16 everyone 3313
17 nasty 3208
18 sick 2969


final sentiment analysis of people during corona in twitter
In [26]:
import plotly.express as px
fig = px.bar(temp, x="count", y="Common_words", title='Commmon Words in Selected Text', orientation='h',
width=700, height=700,color='Common_words')
fig.show()
Commmon Words in Selected Text
Common_words
corona
LINK
Corona
virus
This
Millennials
want
love
name
Virus
calling
everybody
folks
Generation
CandiceBenbow
like
everyone
nasty
sick
stop


final sentiment analysis of people during corona in twitter
sick
stop
0
5k
10k
15k
20k
count
In [ ]:
In [27]:
TRAIN_SIZE = 0.8
MAX_NB_WORDS = 100000
MAX_SEQUENCE_LENGTH = 30
In [ ]:
In [28]:
def convert_Sentiment(Sentiment):
if Sentiment == "positive":
return 2
elif Sentiment == "neutral":
return 1
elif Sentiment == "negative":
return 0
df.Sentiment = df.Sentiment.apply(lambda x : convert_Sentiment(x))
In [29]:
df['full_text'][:3]
Out[29]:
0 Birthday Night first time spending alone time ...
1 Small significant steps India fights Corona Ch...
2 must admit impressed smart Covid19 virus reall...
Name: full_text, dtype: object


jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected Trusted Python 3 (ipykernel)
Run Code
0 5k 10k 15k 20k
count
In [ ]:
In [27]:
TRAIN_SIZE = 0.8
MAX_NB_WORDS = 100000
MAX_SEQUENCE_LENGTH = 30
In [ ]:
In [28]:
def convert_Sentiment(Sentiment):
if Sentiment == "positive":
return 2
elif Sentiment == "neutral":
return 1
elif Sentiment == "negative":
return 0
df.Sentiment = df.Sentiment.apply(lambda x : convert_Sentiment(x))
In [29]:
df['full_text'][:3]
Out[29]:
0 Birthday Night first time spending alone time ...
1 Small significant steps India fights Corona Ch...
2 must admit impressed smart Covid19 virus reall...
Name: full_text, dtype: object
In [ ]:
84°F Partly sunny
ENG IN 15:28 25-09-2022


jupyter final sentiment analysis of people during corona in twitter Last Checkpoint: an hour ago Autosave Failed!
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Connected error Trusted Python 3 (ipykernel)
Run Code
In [*]:
from sklearn.model_selection import StratifiedKFold, train_test_split
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
from sklearn.naive_bayes import ComplementNB
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_score
from sklearn import metrics
from math import *
In [ ]:
In [ ]:
from sklearn.model_selection import train_test_split
X=df['full_text']
y=df['Sentiment']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20,random_state=19)
In [ ]:
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features = 2500)
x = cv.fit_transform(df).toarray()
y = df.iloc[:, 3]
print(x.shape)
print(y.shape)
In [ ]:
# standardization
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
84°F Partly sunny
ENG IN 15:34 25-09-2022



```
from sklearn.naive_bayes import ComplementNB
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_score
from sklearn import metrics
from math import *
```

In [ ]:

In [ ]:
```
from sklearn.model_selection import train_test_split
X=df['full_text']
y=df['Sentiment']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20,random_state=19)
```

In [ ]:
```
from sklearn.feature_extraction.text import CountVectorizer

cv = CountVectorizer(max_features = 2500)
x = cv.fit_transform(df).toarray()
y = df.iloc[:, 3]

print(x.shape)
print(y.shape)
```

In [ ]:
```
# standardization

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

8. CONCLUSION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- Automation of the entire system improves the efficiency
- It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- It gives appropriate access to the authorized users depending on their permissions.
- It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- The System has adequate scope for modification in future if it is necessary.

8.REFERENCE

- S. ChandraKala1 and C. Sindhu2, “OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY,”.Vol .3(1),Oct 2012,420-427 G.Angulakshmi , Dr.R.ManickaChezian ,”An Analysis on Opinion Mining: Techniques and Tools”. Vol 3(7), 2014 www.iarcce.com
- Padmani P .Tribhuvan,S.G. Bhirud,Amrapali P. Tribhuvan,” A Peer Review of Feature Based Opinion Mining and Summarization”(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 247-250 ,www.ijcsit.com.
- Carenini, G., Ng, R. and Zwart, E. Extracting Knowledge from Evaluative Text. Proceedings of the Third International Conference on Knowledge Capture (K-CAP’05), 2005.
- Dave, D., Lawrence, A., and Pennock, D. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. Proceedings of International World Wide Web Conference (WWW’03), 2003.
- Zhu, Jingbo, et al. "Aspect-based opinion polling from customer reviews." IEEE Transactions on Affective Computing, Volume 2.1,pp.37-49, 2011.Na, Jin-Cheon, Haiyang Sui, Christopher Khoo, Syin Chan, and Yunyun Zhou. "Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews." Advances in Knowledge Organization Volume9, pp. 49-54, 2004.Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability using natural language processing." In Proceedings of the 2nd international conference on Knowledge capture, ACM, pp. 70-77, 2003.
- Li, Shoushan, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang. "Sentiment Classification with Polarity Shifting Detection." In Asian Language Processing (IALP), 2013 International Conference on, pp. 129-132. IEEE, 2013.