MAYO INTERNATIONAL SCHOOL

Patparganj, Delhi

INFORMATICS PRACTICES
SUB. CODE - 065
EMAIL DATA VISUALIZATION AND ANALYSIS
AISSCE 2022

python™ matplotlib

Submitted to Ms. Sapna Rai

Submitted by Achal Jain, Kuhoo Gangwar, Aashish Raj

Class 12th E

ACKNOWLEDGEMENT

In performing our assignment, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this assignment gives us much Pleasure. We would like to show our gratitude to Ms. SAPNA RAI, Course Instructor, for giving us a good guideline for assignment throughout numerous consultations. Team member have made valuable comment suggestions on this proposal which gave us an inspiration to improve our project. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in writing this assignment.

With sincere thanks.

CERTIFICATE

This is to certify that Mr. Aashish Raj, Mr. Achal Jain, Ms. Kuhoo Gangwar, student's of Class XII E, completed the Term I – PROJECT SYNOPSIS "Email Data Visualization and Analysis", during the academic year 2021-22 towards partial fulfillment of credit for the Informatics Practices AISSE Project work of CBSE and submitted satisfactory report, as compiled in the following pages, under my supervision.

DATE:

Head of Department Signature:	Principal Seal and Signature:

PROJECT LOGBOOK

PROJECT NAME: Email Data Visualization and Analysis.

YEAR: 2021-22.

CLASS: 12th E

TEACHER NAME: Ms. Sapna Rai.

TEAM MEMBER NAMES

1. Achal Jain.
2. Aashish Raj.
3. Kuhoo Gangwar.

INDEX

INTRODUCTION

Reading, responding to and even organizing emails can oftentimes end up being an enormous time sink. While the numerous email clients out there today go to great lengths to make life easier in this regard, it could be really useful if one could perform email-related tasks in a programmatic manner. Let us consider a scenario where this approach can come in handy — email clutter. Now, it can be argued that the internet is awash with advice on how to reduce email clutter. While all these tips can potentially help you use your email more efficiently going forward, what if currently, your inbox has reached a point of no return, where it is nearly impossible to clean things up manually? A classic case in point — my inbox has over 35000 emails from over 300 different senders. I am positive that most of these are marketing related, and it feels, and rightly so, that it is way too much of an effort to individually unsubscribe to these emails. The first part of this project deals with this issue and explains how to send, retrieve, categorize, delete and unsubscribe to emails using python's imapclient and smtplib libraries.

On the other hand, let us assume you are an individual who is in total control of your inbox and meticulously labels and files every single message. You might be interested in some statistics about your emails like who emails you the most, what does your traffic look like or what your typical email response times are. The second part of this project is meant exactly for this, demonstrating how to analyze email data using Pandas and create visualization tools using Matplotlib.

TEAM ROLES

ROLES OF TEAM MEMBERS:

ROLE	ROLE DESCRIPTION	MEMBER NAME
★ To retrieve data from an Email account after a user entered input providing Mail ID and password ★ To parse the file and create a CSV file of mail data ★ To import the file into python and read it as a Data Frame ★ To view a certain mail on demand based on parameters received by use	To Visualize To Organize To Manipulate	ACHAL JAIN
★ To perform operations on mails and also find common words used ★ To delete mails on demand after receiving parameters from user ★ To run a program on demand to delete certain categories of mails (eg spam mails)	To Analyze To Manipulate	AASHISH RAJ
★ To create various graphs using matplotlib such as weekly email traffic, hourly email traffic, Top email senders, subject word count, common words in subject headers ★ To check, categorize and create sub DataFrames based on their Type e.g., spam, Promotions, Important etc. This will be done via a list of pre-saved words to be search for in mail subjects + taking input from user	To Analyze To Visualize	KUHOO GANGWAR

PROJECT PLAN

PHASE	TASK	ACTUAL START DATE	ACTUAL END DATE	RESPONSIBILITY
PREPARING FOR THE PROJECT	COURSE-WORK READINGS	15 September 2021	20 september 2021	Achal
	SETTING UP TEAM FOLDER	15 September 2021	20 september 2021	Achal
DEFINING THE PROBLEM	BACKGROUND READING	15 September 2021	20 september 2021	Achal
	TEAM MEETING FOR TOPIC SELECTION	15 September 2021	20 september 2021	Kuhoo
BRAINSTORMING	TEAM MEETINGS FOR IDEA GENERATION	15 September 2021	20 september 2021	Kuhoo
DESIGNING YOUR SOLUTION	TEAM MEETINGS TO DESIGN THE SOLUTION	15 September 2021	20 september 2021	Aashish
COLLECTING AND PREPARING DATA	TEAM MEETING TO DISCUSS DATA REQUIREMENTS	15 September 2021	20 september 2021	Kuhoo
PROTOTYPING	DATA COLLECTION	15 September 2021	20 september 2021	Achal
	PREPARATION AND LABELLING	15 September 2021	20 september 2021	Kuhoo, Aashish
PROTOTYPE TESTING	PERFORM DESIRED OPERATIONS	15 September 2021	20 september 2021	Aashish
	INITIATE ACTIONS BASED IN THE RESULT OF YOUR MODEL	15 September 2021	20 september 2021	Aashish

Communications plan

1. Will you meet face-to-face, online or a mixture of each to communicate?
   Ans. We will meet online for all communication

2. How often will you come together to share your progress?
   Ans. We came together once a month

3. Who will set up online documents and ensure that everyone is contributing?
   Ans. We took collective responsibility

4. What tools will you use for communication?
   Ans. Google Meets

#1 Team meeting minutes

Date of meeting: 8/7/2021
Who attended: All Members
Who wasn't able to attend: N/A
Purpose of meeting: To discuss to the topic and project details
Items discussed:

1. What is our Topic ?
2. Work Distribution
3. When to work

Things to do (what, by whom, by when)

1. Research - By All - 1 month time
2. Concept Practice - By All - 1 month time

#2 Team meeting minutes

Date of meeting: 25/10/2021
Who attended: All Members
Who wasn't able to attend: N/A
Purpose of meeting: To discuss to the topic and create synopsis

Items discussed:
1. Topics Details and Subtopics
2. Synopsis Creation

Things to do (what, by whom, by when)
1. Complete Assigned Work - By All - 1 month time.

#3 Team meeting minutes

Date of meeting: 25/10/2021
Who attended: All Members
Who wasn't able to attend: N/A
Purpose of meeting: To compile the data

Items discussed:
1. To submit individual content and compile data
2. To Collect all Data and finishing touches

Things to do (what, by whom, by when)
1. Take print and submit - by Aashish - when asked to submit.

PROBLEM DEFINITION

1.Important issues faced by school/community.

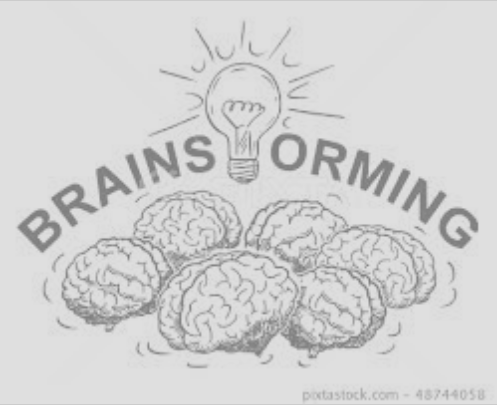
The issue we wish to deal with is the management of mails we receive everyday. We get mails from different people and of different kinds. We wish to find out the frequency and details of such mails to help align and better organize our mails. We often even get spam mails and promotion mails. We need to find which companies/organizations promotional mails the most and consequently deal with them

2.Which issues matter to you and why?

The issues which affect us most are

1. The need to organise our mails
2. The need to detect promotional mails and advertisement
3. The need to segregate between important mails and spam mails

BRAINSTORMING



4.1 Ideas

How might you use the power of Data Science using PANDAS and MATPLOTLIB to solve the users' problem by increasing their knowledge or improving their skills?

Idea #1	By segregating the data into subgroups so that it is easy to access.
Idea #2	Creating a program so that the spam folders will be removed efficiently.
Idea #3	By visualising the data into different graphs to provide a clear picture of what kind of data is possessed by the user.
Idea #4	By performing operations on mails and also find common words used
Idea #5	By deleting the mails on demand after receiving parameters from user

4.2 Briefly summarize the idea for your solution in a few sentences and be sure to identify the tool that you will use.

- Summary - We plan to analyse and visualise the emails in such a way by which the user can have access ,control and most importantly a clear cut image of the source of the data.

- Tools Used - 1. Python Pandas
  2. Matplotlib.pyplot
  3. Numpy
  4. Openpyxl
  5. Getpass

Data

5.1 What data will you need for your project?

CSV of emails to manipulate,analyse, and visualise the data.

5.2 Where or how will you source your data?

Data needed	Where will the data come from?	Who owns the data?	Do you have permission to use the data?	Ethical considerations
Have	Email IDs of group members	Group Members	YES	The group members have given allowance for use
Want/Need	Emails of group members	Group Members	Yes	The group members have given allowance for use
Nice to have	Emails of a random user	User	Depends on User	The user needs to give allowance to use her/her mails

PROTOTYPE

6.1.Tools used to build the prototype.

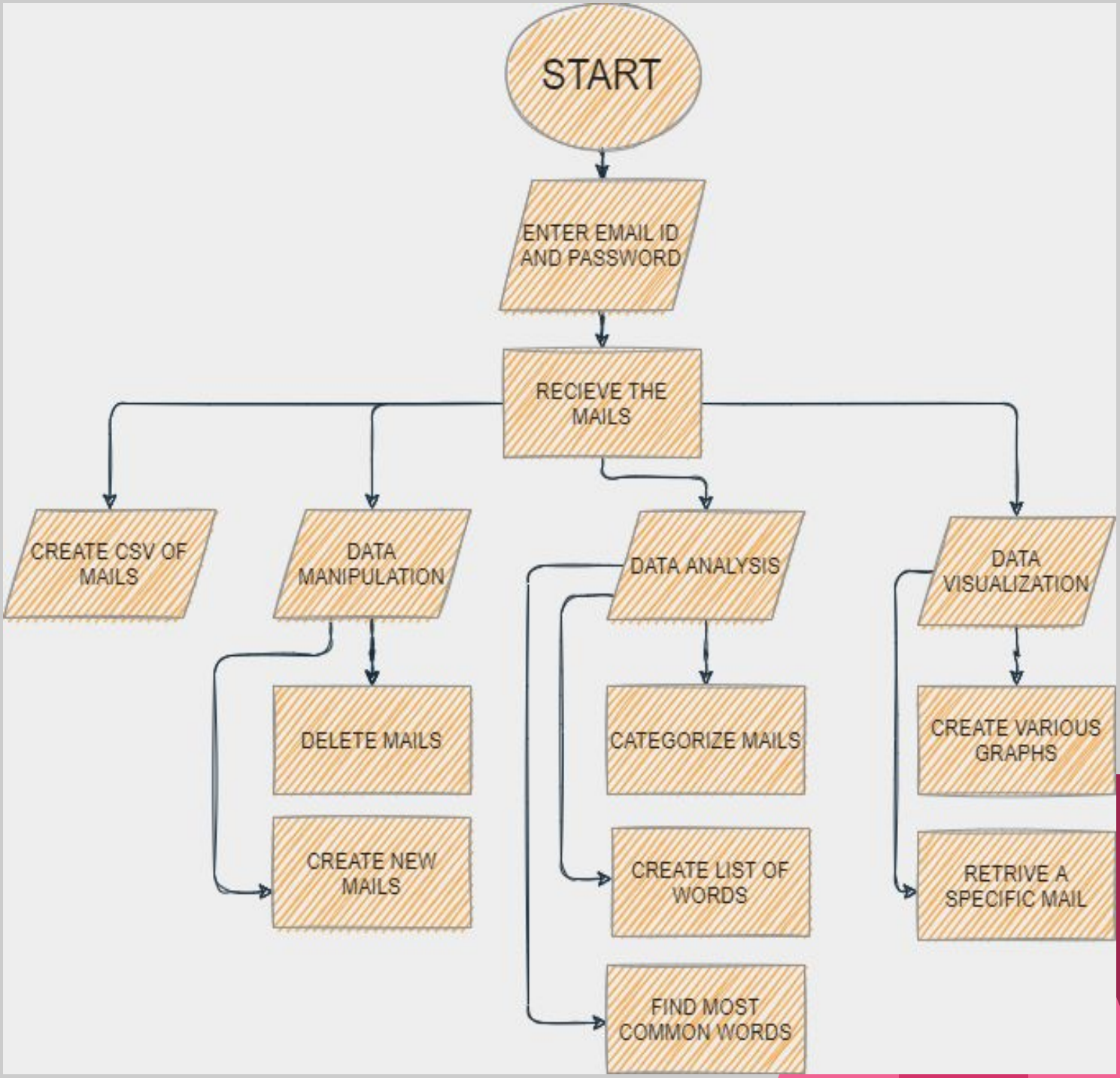
- Python Pandas
- Matplotlib.pyplot
- Numpy
- Openpyxl
- Getpass

6.2 What decisions or outputs will your tool generate and what further action needs to be taken after a decision is made?

- We will be able to divide the data into different sub - groups.
- We will be able to delete any desired mails based on the user's demands.
- We will be able to filter the spam mails and save the memory and storage
- We will be able to visualise the data into different graphs.
- We will be able to run a program on demand to delete certain categories of mails
- We will be able to change data of a mail to insinuate different values and then re-analyze
- We will be able to perform operations on mails and also find common words used
- We will be able to view a certain mail on demand based on parameters received by user.

- FURTHER ACTIONS:: To run the program and check for any gap or error in it.

Methodology/Flow Diagram of the proposed work



Hardware and Software Used

- **PROCESSOR:**

  Intel(R) Core(TM) i5-1035G1

  CPU @ 1.00GHz 1.19 GHz

- **RAM**

  8.00 GB

- **OPERATING SYSTEM**

  Windows 10

- **SYSTEM TYPE**

  64-bit operating system, x64-based processor

- **PYTHON VERSION**

  Python 3.7.8

Python Module Used

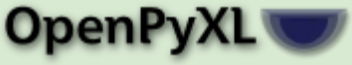
• PANDAS

In computer programming, pandas is a software library written for the Python programming language for Pandas is an open-source library that is made mainly for data manipulation and analysis of data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on the top of the NumPy library.Pandas is fast and it has high-performance & productivity for users.

• MATPLOTLIB

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Python Module Used

- **OPENPYXL** 

Openpyxl is a Python library for reading and writing Excel (with extension xlsx/xlsm/xltx/xltm) files. The openpyxl module allows Python program to read and modify Excel files. Using Openpyxl module, these tasks can be done very efficiently and easily

- **NUMPY** 

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.NumPy is open-source software and has many contributors.

- **GETPASS** 

getpass() prompts the user for a password without echoing. The getpass module provides a secure way to handle the password prompts where programs interact with the users via the terminal.

Analysis/Manipulation/ Visualisation Work Description

To Visualize:

1. To retrieve data from an Email account after a user entered input providing Mail ID and password
2. To parse the file and create a CSV file of mail data
3. To import the file into python and read it as a Data Frame
4. To view a certain mail on demand based on parameters received by user
5. To create various graphs using matplotlib such as weekly email traffic, hourly email traffic, Top email senders, subject word count, common words in subject headers

To Analyze:

1. To check, categorize and create sub Data Frames based on their Type e.g., spam, Promotions, Important etc. This will be done via a list of pre-saved words to be search for in mail subjects + taking input from user
2. To perform operations on mails and also find common words used

To Manipulate:

1. To delete mails on demand after receiving parameters from user
2. To run a program on demand to delete certain categories of mails (eg spam mails)
3. To create mails/new input in the DataFrame to insinuate a received mail
4. To change data of a mail to insinuate different values and then re-analyze
5. Back to STEP 1 TO REANALYZE

SOURCE CODE

MAIN OUTPUT CODE

```
import time
import openpyxl
import matplotlib.pyplot as plt
import pandas as pd
from collections import Counter

#Main Code
def main():
    try:
        df2=pd.DataFrame()
        while True:
            if df2.empty==False:
                df=df2
                received = df[df['Sent/Received'] == 'Received']
            else:
                df = pd.read_excel('Email_Analytics.xlsx')
                received = df[df['Sent/Received'] == 'Received']

            time.sleep(1)
            print('''-------Please select a command from Below-------


    +-------------------------------------------------------------+
    |    *1.View a Mail Category                                  |
    |    *2.View Mails from  a Specific Sender                    |
    |    *3.Read a Mail with subject containg _________           |
    |    *4.Delete a Mail Category                                |
    |    *5.Delete Mails from a Specific Sender                   |
    |    *6.Delete a Mail with subject containg _________         |
    |    *7.View Common Words in Subject Headers                  |
    |    *8.View Weekly Email Traffic                             |
    |    *9.View Top Email Senders                                |
    |    *10.View Subject Word Count                              |
    |    *11.View Monthly Email Traffic                           |
    |    *12.View Hourly Email Traffic                            |
    |    *13.Close Programme                                      |
    +-------------------------------------------------------------+
                        ''')


            ch=int(input('''  ....Select option.... '''))
```

```
if ch==3:
    words=input(''' What Word Sould it Contain : ''')
    alpha=df[df['Subject'].str.contains(words)]
    print(alpha)


elif ch==1:
    print('''    1.Spam Mails
                2.Important Mails
                3.Inbox Mails
                4.Sent Mails
                5.Recieved Mails
                6.Social Mails
                7.All Mails      ''')

    sc_1=int(input('''Enter the Preferred Category :: '''))

    if sc_1==1:
        spam=df[df['Category'].str.contains('Spam')]
        print('\n SPAM MAILS:-')
        print('\n',spam)

    elif sc_1==2:
        important=df[df['Category'].str.contains('Important')]
        print('\n IMPORTANT:-')
        print('\n',important)


    elif sc_1==3:
        inbox=df[df['Category'].str.contains('Inbox')]
        print('\n INBOX:-')
        print('\n',inbox)

    elif sc_1==4:
        sent=df[df['Sent/Received'].str.contains('Sent')]
        print('\n SENT MAILS:-')
        print('\n',sent)

    elif sc_1==5:
        recieved=df[df['Sent/Received'].str.contains('Received')]
        print('\n RECIEVED MAILS:-')
        print('\n',recieved)

    elif sc_1==6:
        a=['Facebook','LinkedIn','CodeChef','Instagram','Twitter',
           'Snapchat','TikTok']
        social=df[df['From (Sender)'].str.contains('|'.join(a))]
        print('\n SOCIAL MAILS:-')
        print('\n',social)
```

```
        elif sc_1==7:
                print('\n ALL MAILS:-')
                print('\n',df)
        else:
                print('NONE')

elif ch==2:
    sender=input(''' Enter the Name/Email ID of sender : ''')
    alpha=df[df['From (Sender)'].str.contains(sender)]
    if alpha.empty==True:
        alpha=df[df['From (Email ID)'].str.contains(sender)]
    print(alpha)

elif ch==5:
    sender=input(''' Whose mails to delete : ''')
    alpha=df[df['From (Sender)'].str.contains(sender)]
    df2=df.drop(alpha.index, axis=0 )
    print(''' The mails have been deleted ''')
    time.sleep(1.5)

elif ch==6:
    words=input('''  What Word Sould it Contain : ''')
    alpha=df[df['Subject'].str.contains(words)]
    if alpha.empty==True:
        alpha=df[df['From (Email ID)'].str.contains(words)]
    df2=df.drop(alpha.index, axis=0)
    print('''   The mails have been deleted ''')
    time.sleep(1.5)

elif ch==4:
    print('''    1.Spam Mails
                2.Important Mails
                3.Inbox Mails
                4.Sent Mails
                5.Recieved Mails
                6.Social Mails
                7.All Mails      ''')

    sc_1=int(input('''Enter the Preferred Category :: '''))

    if sc_1==1:
        alpha=df[df['Category'].str.contains('Spam')]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)

    elif sc_1==2:
        alpha=df[df['Category'].str.contains('Important')]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)
```

```
    elif sc_1==3:
        alpha=df[df['Category'].str.contains('Inbox')]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)

    elif sc_1==4:
        alpha=df[df['Sent/Received'].str.contains('Sent')]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)



    elif sc_1==5:
        alpha=df[df['Sent/Received'].str.contains('Received')]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)

    elif sc_1==6:
        a=['Facebook','LinkedIn','CodeChef',
           'Instagram','Twitter','Snapchat','TikTok']
        alpha=df[df['From (Sender)'].str.contains('|'.join(a))]
        df2=df.drop(alpha.index, axis=0 )
        print(''' The mails have been deleted ''')
        time.sleep(1.5)

    elif sc_1==7:
        df2=pd.DataFrame()
        print(''' The mails have been deleted ''')
        time.sleep(1.5)

    else:
        print('NONE')

elif ch==7:

    word_list_2d=df['Subject'].str.split(' ').fillna('none').tolist()
    word_list_1d=[word for list in word_list_2d for word in list]


    word_list_1d = [word.lower() for word in word_list_1d]

    exclude_list = ['this', 'that', 'your', 'with', 'from']
    word_list_1d = [word for word in word_list_1d if word not in
                    exclude_list and len(word)>3]
```

```
    common_words_map = Counter(word_list_1d).most_common(10)
    common_words = [pair[0] for pair in common_words_map]
    frequency = [pair[1] for pair in common_words_map]
    plt.figure()
    plt.barh(common_words, frequency, color = 'lightcoral', ec = 'black'
    ,linewidth = 1.25)
    plt.gca().invert_yaxis()
    plt.title('Most Common Words in Subjects', fontsize = 14 ,fontweight
    ='bold')
    y = 0.15
    for i in range(len(frequency)):
        if len(str(frequency[i])) == 3:
            x = frequency[i] - 14
        else:
            x = frequency[i] - 10
        plt.text(x,y,frequency[i], fontsize = 10,fontweight = 'bold')
        y = y + 1
    plt.xticks([0,200])
    plt.xlabel('Occurrences', fontweight = 'bold', labelpad=-5)
    plt.show()

elif ch==8:
    df['Day'] = pd.Categorical(df['Day'], categories= ['Mon','Tue','Wed',
    'Thu','Fri','Sat', 'Sun'],ordered=
    count_sorted_by_day = pd.DataFrame(df['Day'].value_counts().
    sort_index())
    count_sorted_by_day.plot(kind='bar', color = 'blueviolet',linewidth
    = 2, ylim = [0,350])
    plt.title('Weekly Email Traffic', fontweight = 'bold' ,fontsize = 14)
    plt.ylabel("Received Email Count", fontweight = 'bold', labelpad = 15)
    plt.grid()
    plt.show()

elif ch==9:
    sender_top_20 =  received['From (Sender)'].value_counts().nlargest(20)
    sender_top_20_count = sender_top_20.values
    sender_top_20_names = sender_top_20.index.tolist()
    plt.figure()
    plt.barh(sender_top_20_names,sender_top_20_count,color='forestgreen'
    , ec = 'black', linewidth = 1.0)
    plt.gca().invert_yaxis()
    plt.title('Top 20 Senders', fontsize = 14 ,fontweight = 'bold')
    plt.xlabel('Received Email Count', fontweight = 'bold')
    plt.tight_layout()
    plt.show()

elif ch==10:
    df['Subject Word Count'] = df['Subject'].str.split(' ').str.len()
    plt.figure()
    plt.hist(df['Subject Word Count'], bins=15, color = 'slategray',
    ec = 'black')
```

```
            plt.axis([0, 30, 0, 1200])
            plt.xlabel('Word Count', fontweight = 'bold')
            plt.ylabel('No. of Emails', fontweight = 'bold')
            plt.title('Subject Word Count Histogram', fontsize = 14,
            Fontweight = 'bold')
            plt.show()

        elif ch==11:
            month = received['Month']
            count_sorted_by_month = month.value_counts()
            count_sorted_by_month.plot(marker = 'o', color='green')
            plt.title('Monthly Email Traffic', fontsize = 14,
            fontweight = 'bold')
            plt.ylabel("Received Email Count", fontweight = 'bold',
            labelpad = 15)
            plt.xlabel("Month of the Year", fontweight = 'bold',
            labelpad = 15)
            plt.xticks(range(len(count_sorted_by_month.index)),
            count_sorted_by_month.index)
            plt.xticks(rotation=90)
            plt.grid()
            plt.show()

        elif ch==12:
            hour = received['Time'].str.split(':').str[0] + ':00'
            count_sorted_by_hour = hour.value_counts().sort_index()
            count_sorted_by_hour.plot(marker = 'o',color = 'green')
            plt.title('Hourly Email Traffic', fontsize = 14,
            fontweight = 'bold')
            plt.ylabel("Received Email Count", fontweight = 'bold',
            labelpad = 15)
            plt.xlabel("Hour of the Day", fontweight = 'bold',
            labelpad = 15)
            plt.xticks(range(len(count_sorted_by_hour.index)),
            count_sorted_by_hour.index)
            plt.xticks(rotation=90)
            plt.grid()
            plt.show()

        elif ch==13:
            print("\n......Thank you for using our service........")
            break
            quit()
    except:
        main()

import recieve_mail
main()
```

TO IMPORT MAILS FROM GIVEN EMAIL ID

```
import imapclient
import smtplib
import getpass


import imaplib
import pprint
import time

imap_server = 'imap.gmail.com'
smtp_server = 'smtp.gmail.com'
time.sleep(1)
username = input('''MAIL ID: ''')
time.sleep(1)
password = getpass.getpass(prompt='''PASSWORD:  ''')
print('''Please Wait a few Minutes... ''')
imapobj = imapclient.IMAPClient(imap_server, ssl=True)
imapobj.login(username, password)
smtpobj = smtplib.SMTP(smtp_server, 587)
smtpobj.ehlo()
smtpobj.starttls()
smtpobj.login(username, password)

pprint.pprint(imapobj.list_folders())
imaplib._MAXLINE = 10000000
imapobj.select_folder('Inbox', readonly=True)
UIDs = imapobj.search(['SINCE', '01-Oct-2021', 'BEFORE','17-Feb-2022'])

category = []
for i in range(len(UIDs)):
    label_dict = imapobj.get_gmail_labels(UIDs[i])
    label = label_dict[UIDs[i]]
    if 'Starred' in str(label):
        category.append('Spam')
    elif 'Important' in str(label):
        category.append('Important')
    elif len(label) == 0:
        category.append('Inbox')
    else:
        category.append('Custom Label')

import pyzmail
from_addresses = []
subjects = []
dates = []
days = []
months = []
years = []
```

```
times = []
    sent_received = []
    unsub_links = []

    for i in range(len(UIDs)):
        raw_message = imapobj.fetch(UIDs[i], ['BODY[]'])
       Message = pyzmail.PyzMessage.factory(raw_message[UIDs[i]][b'BODY[]'])

        if message.get_address('from')[1] == username:
            sent_received.append('Sent')
        else:
            sent_received.append('Received')

        full_date = message.get_decoded_header('date')
        from_addresses.append(message.get_address('from'))
        subjects.append(message.get_subject(''))
        unsub_link = message.get_decoded_header('List-Unsubscribe')
        if len(str(unsub_link)) > 0 and 'mailto' in unsub_link:
            unsub_link = unsub_link.split(',')
            unsub_links.append([unsub_link[idx] for idx, s in
            enumerate(unsub_link) if 'mailto' in s][0])
        else:
            unsub_links.append('No unsubscribe link found')

        day = full_date.split()[0].strip(',')
        date = full_date.split()[1]
        month = full_date.split()[2]
        year = full_date.split()[3]
        time = full_date.split()[4]

        days.append(day)
        dates.append(date)
        months.append(month)
        years.append(year)
        times.append(time)

    from openpyxl import Workbook
    wb = Workbook()
    ws = wb.active
    ws.title = "email_info"
    ws.cell(1,1).value = "Date"
    ws.cell(1,2).value = "Month"
    ws.cell(1,3).value = "Year"
    ws.cell(1,4).value = "Day"
    ws.cell(1,5).value = "Time"
    ws.cell(1,6).value = "From (Sender)"
    ws.cell(1,7).value = "From (Email ID)"
    ws.cell(1,8).value = "Subject"
    ws.cell(1,9).value = "Sent/Received"
    ws.cell(1,10).value = "Category"
```

```
    for i in range(len(UIDs)):
         ws.cell(row=i+2, column=1).value = dates[i]
         ws.cell(row=i+2, column=2).value = months[i]
         ws.cell(row=i+2, column=3).value = years[i]
         ws.cell(row=i+2, column=4).value = days[i]
         ws.cell(row=i+2, column=5).value = times[i]
         ws.cell(row = i+2, column = 6).value = from_addresses[i][0]
         ws.cell(row = i+2, column = 7).value = from_addresses[i][1]
         ws.cell(row = i+2, column = 8).value = str(subjects[i])
         ws.cell(row=i + 2, column=9).value = sent_received[i]
         ws.cell(row=i+2, column=10).value = category[i]

    wb.save('Email_Analytics.xlsx')


except:
     time.sleep(1)
```

OUTPUT

```
                MAIL ID: abc123@gmail.com

                PASSWORD:

              Please Wait a few Minutes...

+-------------------------------------------------------------------+
|           *1.View a Mail Category                                 |
|           *2.View Mails from  a Specific Sender                   |
|           *3.Read a Mail with subject containg _________          |
|           *4.Delete a Mail Category                               |
|           *5.Delete Mails from a Specific Sender                  |
|           *6.Delete a Mail with subject containg _________        |
|           *7.View Common Words in Subject Headers                 |
|           *8.View Weekly Email Traffic                            |
|           *9.View Top Email Senders                               |
|           *10.View Subject Word Count                             |
|           *11.View Monthly Email Traffic                          |
|           *12.View Hourly Email Traffic                           |
|           *13.Close Programme                                     |
+-------------------------------------------------------------------+


....Select option.... 1

                1.Spam Mails
                2.Important Mails
                3.Inbox Mails
                4.Sent Mails
                5.Recieved Mails
                6.Social Mails
                7.All Mails

                Enter the Preferred Category :: 1
```

```
SPAM MAILS:-

Empty DataFrame
Columns: [Date, Month, Year, Day, Time, From (Sender), From (Email ID), Subject, Sent/Received, Category]
Index: []

                                Enter the Preferred Category :: 2

IMPORTANT:-

    Date Month     Year  Day  ...                From (Email ID)                                Subject Sent/Received   Category
5      1   Oct     2021  Fri  ...             noreply@acko.com                  Rapido Trip Insurance     Received  Important
19   Oct  2021 12:34:12    2  ...          noreply@emsecure.in                           Registration     Received  Important
20   Oct  2021 12:35:03    2  ...          noreply@emsecure.in                           Registration     Received  Important
23   Oct  2021 17:12:21    2  ...          noreply@emsecure.in                           Registration     Received  Important
37     4   Oct     2021  Mon  ...             noreply@acko.com                  Rapido Trip Insurance     Received  Important
53     5   Oct     2021  Tue  ...        do-not-reply@amazon.in  Great Indian Festival is here! Redeem Your Rew...     Received  Important

                                Enter the Preferred Category :: 3

INBOX:-

     Date Month  Year  Day      Time  From (Sender)            From (Email ID)                                Subject Sent/Received Category
0      30   Sep  2021  Thu  13:03:27       Facebook    reminders@facebookmail.com  You have 80 notifications about Vijay and others     Received    Inbox
1      30   Sep  2021  Thu  14:32:53       Coursera   Coursera@email.coursera.org          Degree-level learning before you apply?     Received    Inbox
2      01   Oct  2021  Fri  01:01:43        Groupon      noreply@r.grouponmail.ae         At the Top Burj Khalifa Fast-Track Entry     Received    Inbox
3      01   Oct  2021  Fri  01:33:15  Groupon Dubai      noreply@r.grouponmail.ae  5* Classic Treatment of Choice / Lebanon Islan...     Received    Inbox
4      01   Oct  2021  Fri  04:46:42        Groupon      noreply@r.grouponmail.ae  It's official: The brunch and buffet capital o...     Received    Inbox
...   ...   ...   ...  ...       ...            ...                         ...                                            ...          ...       ...
1173   16   Feb  2022  Wed  04:22:59        Groupon      noreply@r.grouponmail.ae       Explore Our Beauty Selection - All Under 30     Received    Inbox
1174   16   Feb  2022  Wed  08:01:41        Groupon      noreply@r.grouponmail.ae            Because The Early Bird Gets The Worm!     Received    Inbox
1175   16   Feb  2022  Wed  12:31:39       LinkedIn  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...     Received    Inbox
1176   16   Feb  2022  Wed  13:31:40       LinkedIn  messages-noreply@linkedin.com                  Achal, you're getting noticed     Received    Inbox
1177   16   Feb  2022  Wed  09:58:16   Adobe Acrobat         mail@mail.adobe.com     Collaborate with everyone, without boundaries     Received    Inbox

[1126 rows x 10 columns]

                                Enter the Preferred Category :: 4

SENT MAILS:-

     Date Month  Year  Day      Time From (Sender)        From (Email ID)                Subject Sent/Received   Category
934    17   Jan  2022  Mon  08:34:48    achal jain  achaljain232@gmail.com  Share images from EzScanner         Sent  Important
1146   13   Feb  2022  Sun  22:37:13    achal jain  achaljain232@gmail.com     Document from Achal Jain         Sent  Important
```

```
                                Enter the Preferred Category :: 5

 RECIEVED MAILS:-

      Date Month  Year  Day      Time  From (Sender)              From (Email ID)                                    Subject Sent/Received Category
0       30   Sep  2021  Thu  13:03:27       Facebook     reminders@facebookmail.com   You have 80 notifications about Vijay and others      Received    Inbox
1       30   Sep  2021  Thu  14:32:53       Coursera    Coursera@email.coursera.org           Degree-level learning before you apply?      Received    Inbox
2       01   Oct  2021  Fri  01:01:43        Groupon       noreply@r.grouponmail.ae          At the Top Burj Khalifa Fast-Track Entry      Received    Inbox
3       01   Oct  2021  Fri  01:33:15  Groupon Dubai       noreply@r.grouponmail.ae  5* Classic Treatment of Choice / Lebanon Islan...      Received    Inbox
4       01   Oct  2021  Fri  04:46:42        Groupon       noreply@r.grouponmail.ae  It's official: The brunch and buffet capital o...      Received    Inbox
...    ...   ...   ...  ...       ...            ...                            ...                                                ...          ...      ...
1173    16   Feb  2022  Wed  04:22:59        Groupon       noreply@r.grouponmail.ae       Explore Our Beauty Selection - All Under 30      Received    Inbox
1174    16   Feb  2022  Wed  08:01:41        Groupon       noreply@r.grouponmail.ae            Because The Early Bird Gets The Worm!      Received    Inbox
1175    16   Feb  2022  Wed  12:31:39       LinkedIn  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox
1176    16   Feb  2022  Wed  13:31:40       LinkedIn  messages-noreply@linkedin.com                  Achal, you're getting noticed      Received    Inbox
1177    16   Feb  2022  Wed  09:58:16  Adobe Acrobat           mail@mail.adobe.com     Collaborate with everyone, without boundaries      Received    Inbox

[1176 rows x 10 columns]

                                Enter the Preferred Category :: 6

 SOCIAL MAILS:-

      Date Month  Year  Day  ...                From (Email ID)                                    Subject Sent/Received Category
0       30   Sep  2021  Thu  ...     reminders@facebookmail.com   You have 80 notifications about Vijay and others      Received    Inbox
10       1   Oct  2021  Fri  ...  messages-noreply@linkedin.com          Achal, add Arjun Jayakumar to your network      Received    Inbox
11       1   Oct  2021  Fri  ...   updates-noreply@linkedin.com                        James Henry recently posted      Received    Inbox
14       1   Oct  2021  Fri  ...  groupupdates@facebookmail.com  Coromon Fanatics and 7 others are new Group su...      Received    Inbox
26       2   Oct  2021  Sat  ...     reminders@facebookmail.com   You have 79 notifications about Anshu and others      Received    Inbox
...    ...   ...   ...  ...  ...                            ...                                                ...          ...      ...
1163    14   Feb  2022  Mon  ...     birthdays@facebookmail.com            It's Jensen Menezes's birthday today      Received    Inbox
1165    14   Feb  2022  Mon  ...    no-reply@mail.instagram.com  achalj.__, see what's been happening on Instagram      Received    Inbox
1171    16   Feb  2022  Wed  ...  messages-noreply@linkedin.com  Jasim Samir Parkar's invitation is waiting for...      Received    Inbox
1175    16   Feb  2022  Wed  ...  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox
1176    16   Feb  2022  Wed  ...  messages-noreply@linkedin.com                  Achal, you're getting noticed      Received    Inbox

[295 rows x 10 columns]

                                Enter the Preferred Category :: 7

 ALL MAILS:-

      Date Month  Year  Day      Time  From (Sender)              From (Email ID)                                    Subject Sent/Received Category
0       30   Sep  2021  Thu  13:03:27       Facebook     reminders@facebookmail.com   You have 80 notifications about Vijay and others      Received    Inbox
1       30   Sep  2021  Thu  14:32:53       Coursera    Coursera@email.coursera.org           Degree-level learning before you apply?      Received    Inbox
2       01   Oct  2021  Fri  01:01:43        Groupon       noreply@r.grouponmail.ae          At the Top Burj Khalifa Fast-Track Entry      Received    Inbox
3       01   Oct  2021  Fri  01:33:15  Groupon Dubai       noreply@r.grouponmail.ae  5* Classic Treatment of Choice / Lebanon Islan...      Received    Inbox
4       01   Oct  2021  Fri  04:46:42        Groupon       noreply@r.grouponmail.ae  It's official: The brunch and buffet capital o...      Received    Inbox
...    ...   ...   ...  ...       ...            ...                            ...                                                ...          ...      ...
1173    16   Feb  2022  Wed  04:22:59        Groupon       noreply@r.grouponmail.ae       Explore Our Beauty Selection - All Under 30      Received    Inbox
1174    16   Feb  2022  Wed  08:01:41        Groupon       noreply@r.grouponmail.ae            Because The Early Bird Gets The Worm!      Received    Inbox
1175    16   Feb  2022  Wed  12:31:39       LinkedIn  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox
1176    16   Feb  2022  Wed  13:31:40       LinkedIn  messages-noreply@linkedin.com                  Achal, you're getting noticed      Received    Inbox
1177    16   Feb  2022  Wed  09:58:16  Adobe Acrobat           mail@mail.adobe.com     Collaborate with everyone, without boundaries      Received    Inbox

[1178 rows x 10 columns]

                                ....Select option.... 2

                                Enter the Name/Email ID of sender : Acko
    Date Month  Year  Day      Time From (Sender)   From (Email ID)                Subject Sent/Received   Category
5      1   Oct  2021  Fri  04:48:26          Acko  noreply@acko.com  Rapido Trip Insurance      Received  Important
37     4   Oct  2021  Mon  04:57:12          Acko  noreply@acko.com  Rapido Trip Insurance      Received  Important
119   12   Oct  2021  Tue  14:32:55          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
130   13   Oct  2021  Wed  14:47:48          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
267   29   Oct  2021  Fri  08:06:34          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
390   11   Nov  2021  Thu  14:34:07          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
436   17   Nov  2021  Wed  06:25:19          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
491   24   Nov  2021  Wed  05:54:05          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
507   26   Nov  2021  Fri  05:14:38          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
647   10   Dec  2021  Fri  07:49:00          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
684   14   Dec  2021  Tue  07:48:17          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox
731   21   Dec  2021  Tue  08:14:13          Acko  noreply@acko.com  Rapido Trip Insurance      Received      Inbox

                                ....Select option.... 3

                                What Word Sould it Contain : Check
     Date Month  Year  Day  ...                From (Email ID)                                    Subject Sent/Received Category
164    18   Oct  2021  Mon  ...       noreply@r.grouponmail.ae  October Sale is on!  Check Out This Week's Dis...      Received    Inbox
425    15   Nov  2021  Mon  ...  messages-noreply@linkedin.com  Achal, Check out what you missed from your pag...      Received    Inbox
486    23   Nov  2021  Tue  ...    no-reply@promotion.noon.com    Check out the craziest home deals of the year       Received    Inbox
497    25   Nov  2021  Thu  ...       noreply@r.grouponmail.ae  Fast Track Ticket: Child (AED 72), Adult (AED ...      Received    Inbox
571     2   Dec  2021  Thu  ...  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox
679    13   Dec  2021  Mon  ...       noreply@r.grouponmail.ae                Dental Check-Up, Scale and Polish      Received    Inbox
690    15   Dec  2021  Wed  ...       noreply@r.grouponmail.ae  5* Grand Festive Brunch or New Year Day Brunch...      Received    Inbox
710    18   Dec  2021  Sat  ...       noreply@r.grouponmail.ae                    Car Service and Health Check      Received    Inbox
776    28   Dec  2021  Tue  ...       noreply@r.grouponmail.ae  Pet - Home Health Assessment / Dental Check-Up...      Received    Inbox
791    30   Dec  2021  Thu  ...       noreply@r.grouponmail.ae  Facial Treatment Packages / Interior and Exter...      Received    Inbox
843    06   Jan  2022  Thu  ...       noreply@r.grouponmail.ae                Dental Check-Up, Scale and Polish      Received    Inbox
909    14   Jan  2022  Fri  ...       noreply@r.grouponmail.ae  Body-Contouring Treatment / 5* Spa Treatment a...      Received    Inbox
944    18   Jan  2022  Tue  ...       noreply@r.grouponmail.ae  Sale is on!  Check Out This Week's Discounted ...      Received    Inbox
991    25   Jan  2022  Tue  ...       noreply@r.grouponmail.ae  4* Choice of 30-Minute Treatment / Dental Chec...      Received    Inbox
1071    4   Feb  2022  Fri  ...  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox
1077   05   Feb  2022  Sat  ...       noreply@r.grouponmail.ae  Up to an extra 20% off & more | Hair Spa, Trim...      Received    Inbox
1175   16   Feb  2022  Wed  ...  messages-noreply@linkedin.com  Check out what you missed from Euclid Labs las...      Received    Inbox

[17 rows x 10 columns]
```

```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 1

The mails have been deleted
```

```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 2

The mails have been deleted
```

```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 3

The mails have been deleted
```

```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 4

The mails have been deleted
```

```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 5

The mails have been deleted
```

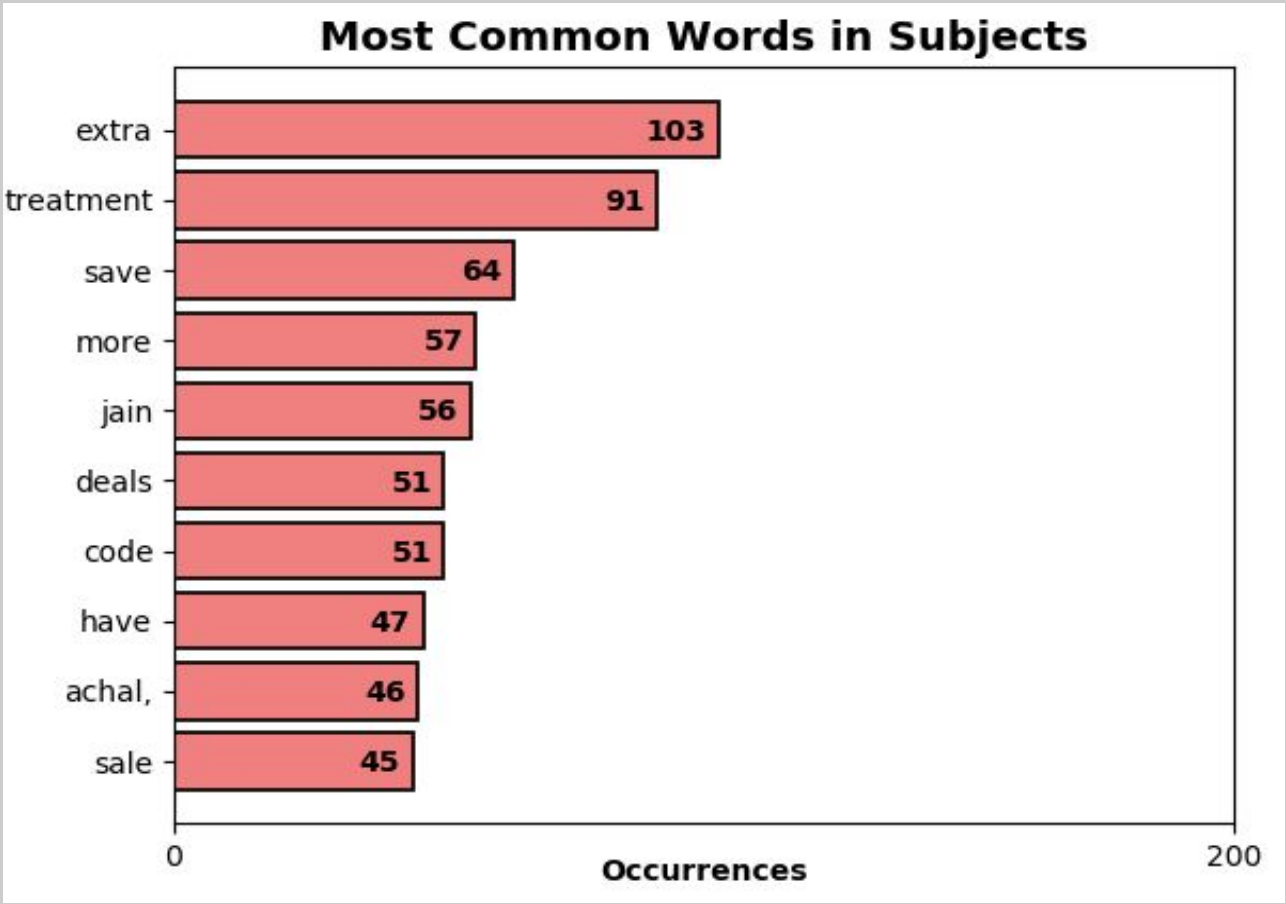
```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 6

The mails have been deleted
```

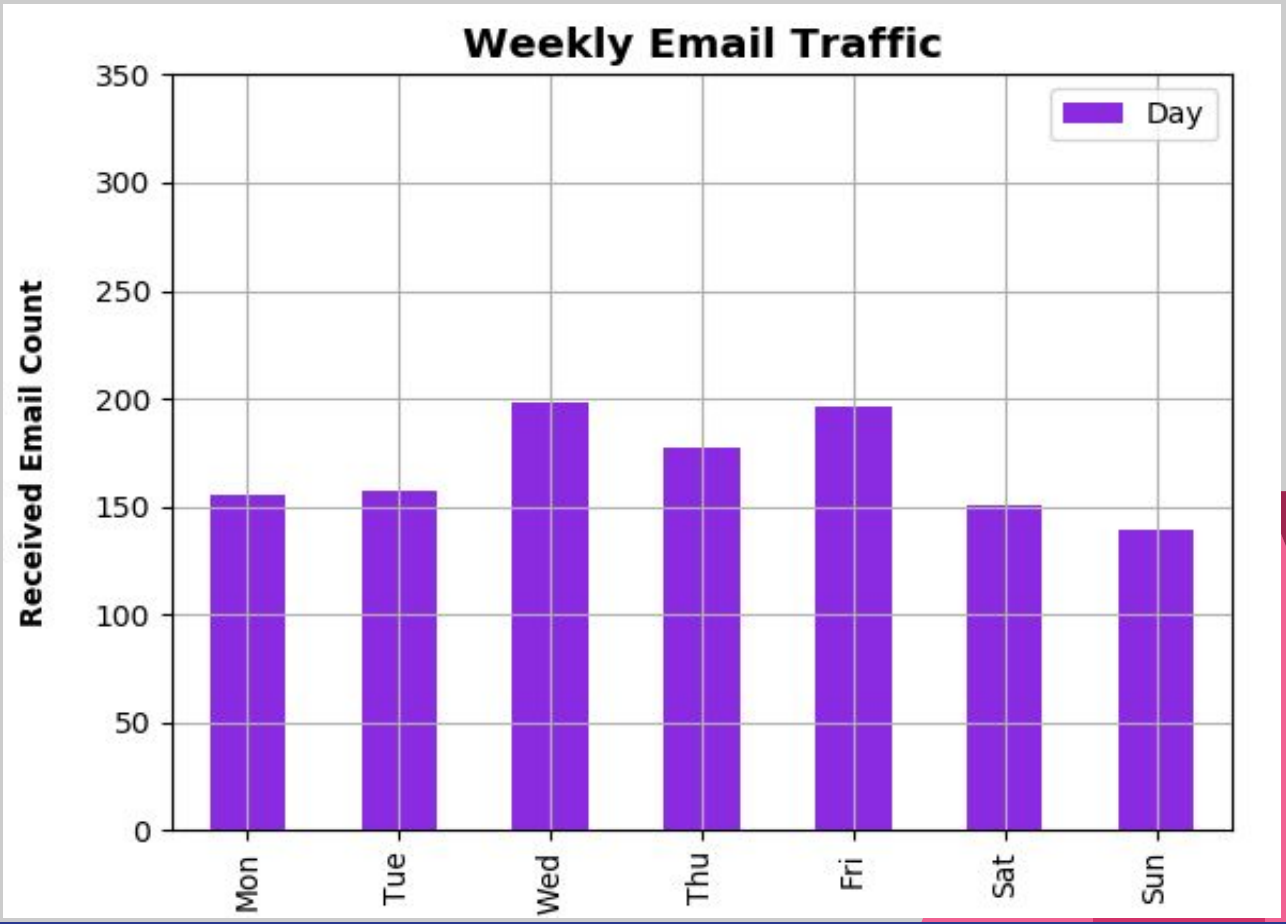
```
....Select option.... 4

 1.Spam Mails
 2.Important Mails
 3.Inbox Mails
 4.Sent Mails
 5.Recieved Mails
 6.Social Mails
 7.All Mails

 Enter the Preferred Category :: 7

The mails have been deleted
```

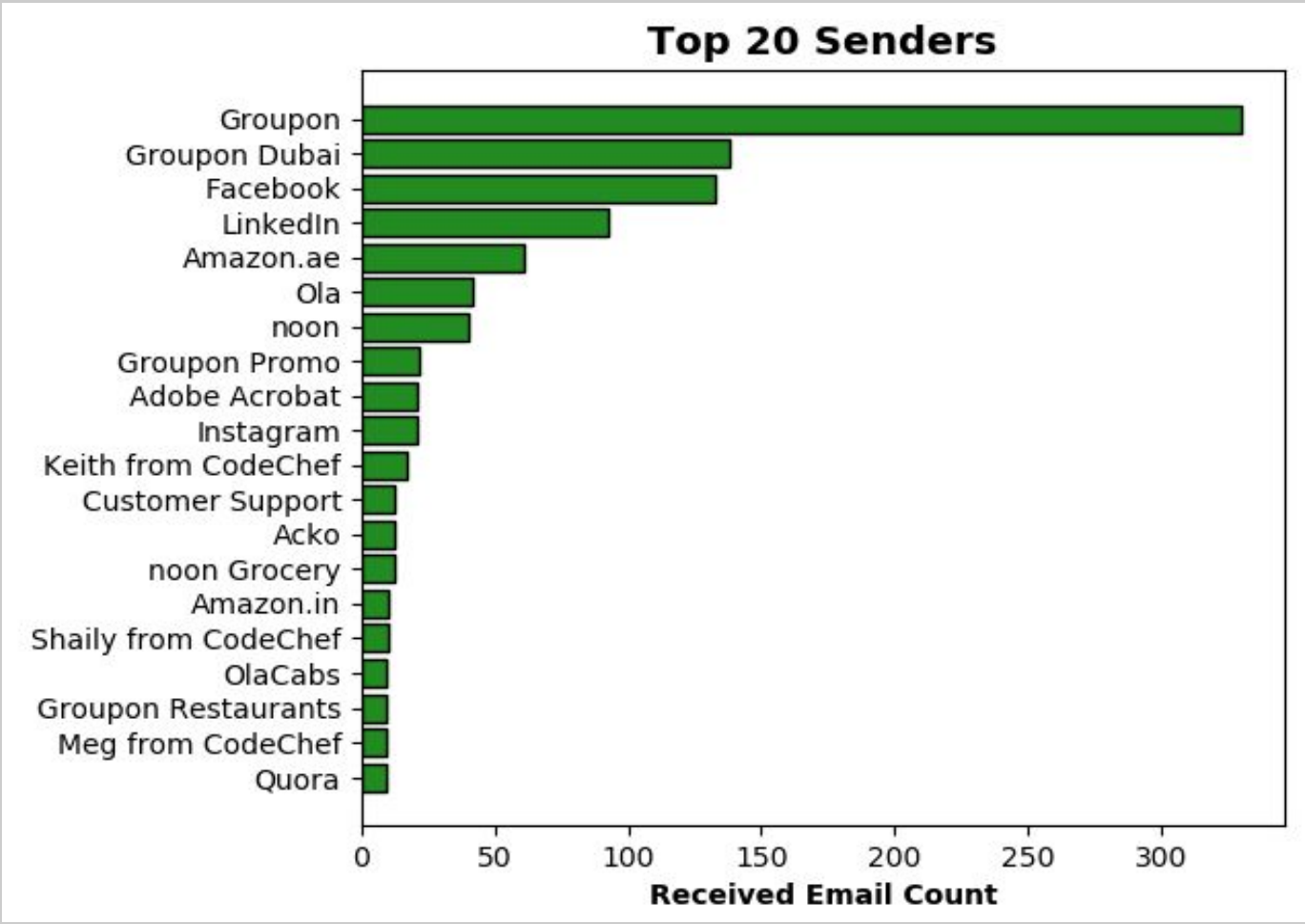
```
....Select option.... 5

Whose mails to delete : Acko

The mails have been deleted
```

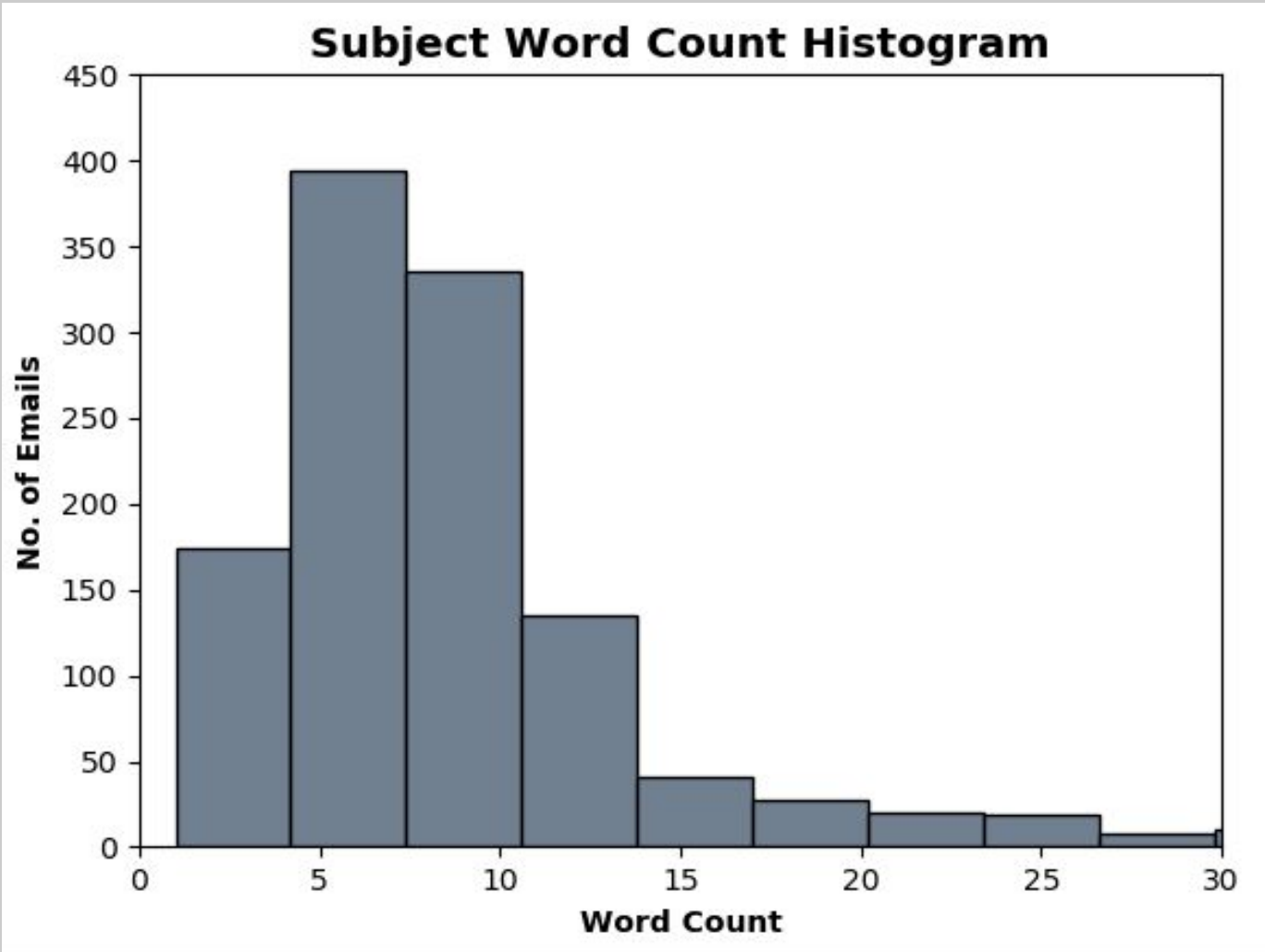
```
....Select option.... 6

What Word Sould it Contain : Hello

The mails have been deleted
```

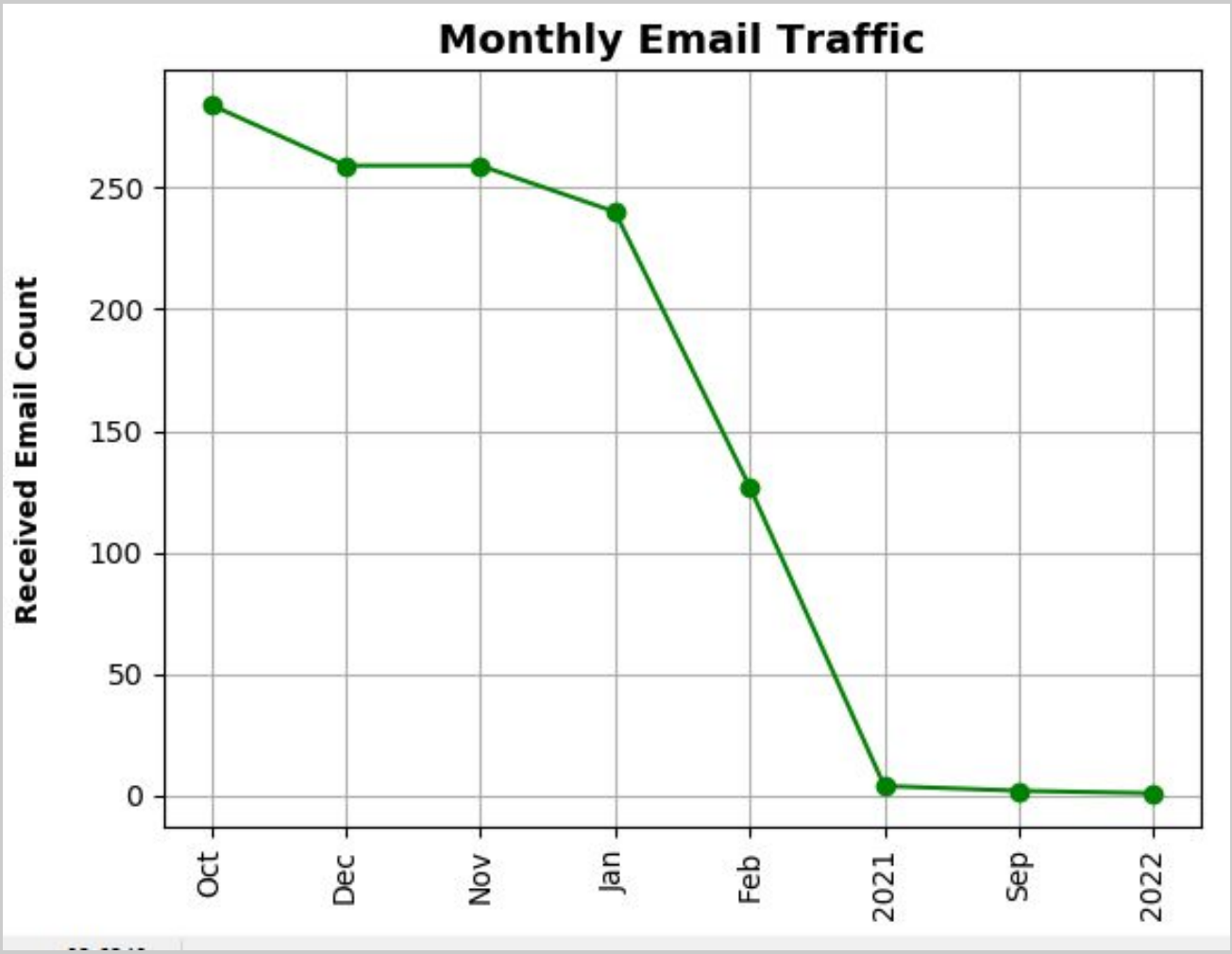
Visualization

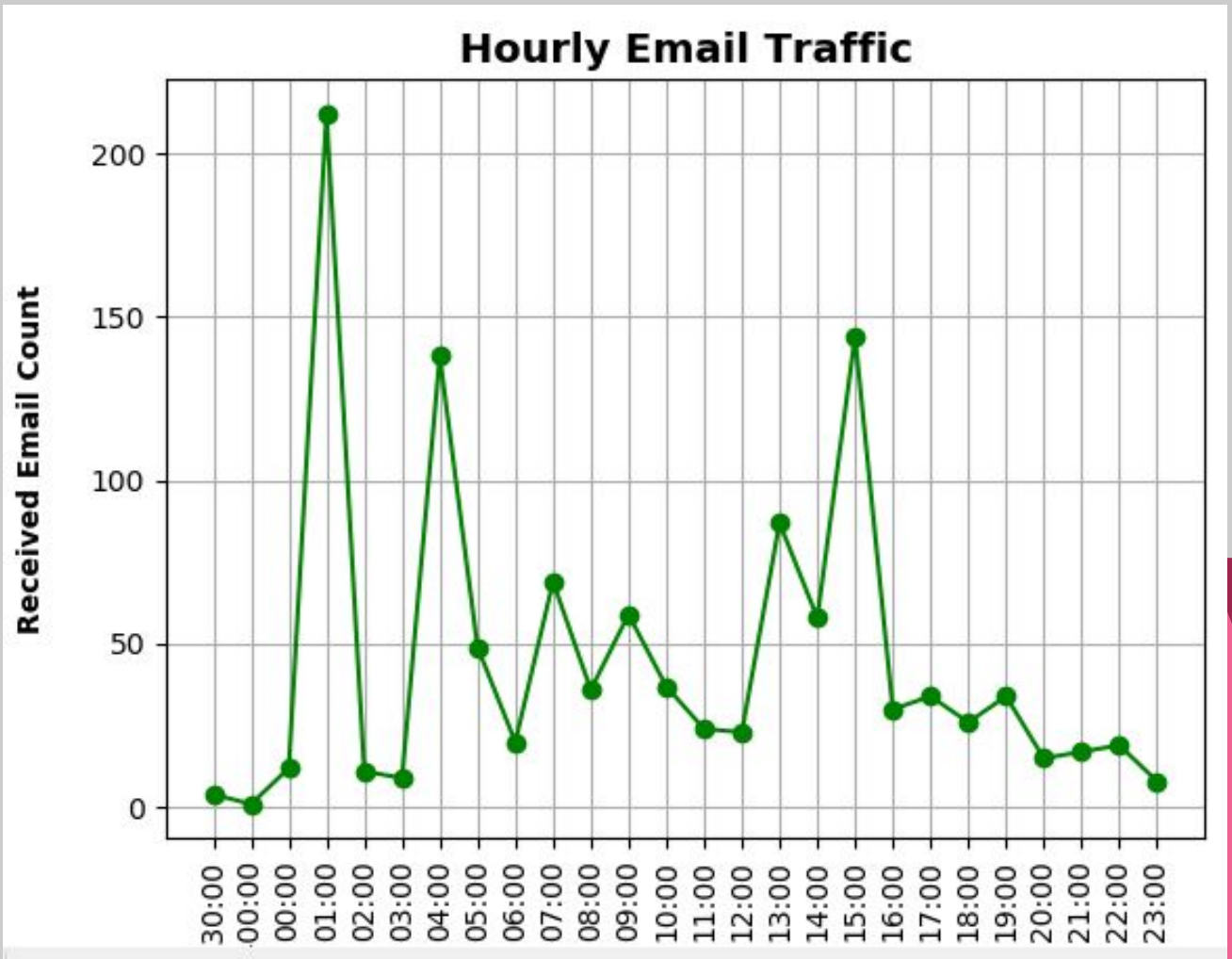





Top 20 Senders
Groupon
Groupon Dubai
Facebook
LinkedIn
Amazon.ae
Ola
noon
Groupon Promo
Adobe Acrobat
Instagram
Keith from CodeChef
Customer Support
Acko
noon Grocery
Amazon.in
Shaily from CodeChef
OlaCabs
Groupon Restaurants
Meg from CodeChef
Quora
0
50
100
150
200
250
300
Received Email Count


Subject Word Count Histogram
No. of Emails
0
50
100
150
200
250
300
350
400
450
0
5
10
15
20
25
30
Word Count


Monthly Email Traffic
Received Email Count
250
200
150
100
50
0
Oct
Dec
Nov
Jan
Feb
2021
Sep
2022


Hourly Email Traffic
Received Email Count
200
150
100
50
0
30:00
00:00
00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00

List of References

- https://medium.com/analytics-vidhya/how-to-read-and-write-data-to-google-spreadsheet-using-python-ebf54d51a72c
- https://towardsdatascience.com/email-automation-analytics-and-visualization-53b022e0f9a0
- http://beneathdata.com/how-to/email-behavior-analysis/