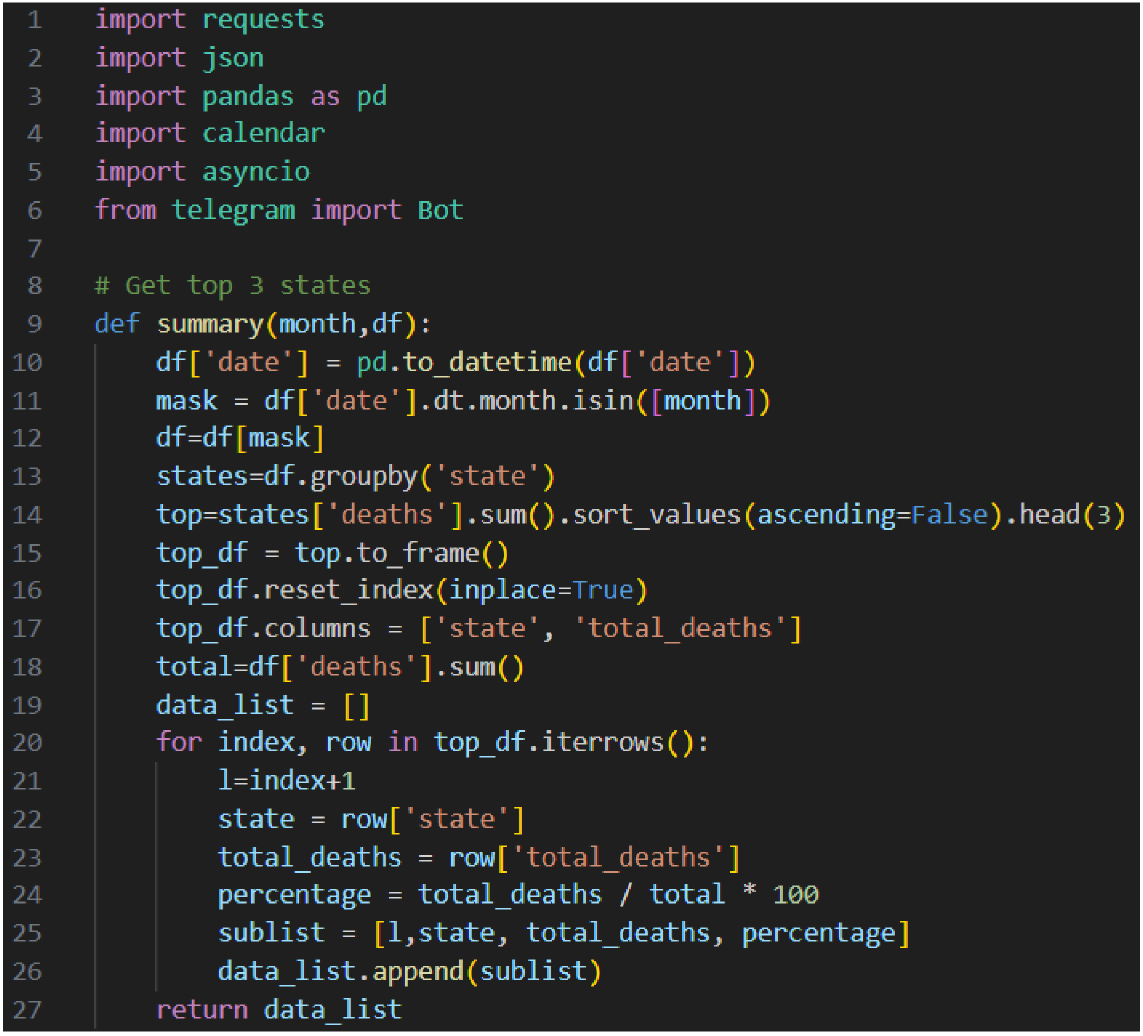
# Data Engineer Intern Assignment from Qure.ai

**Name: Nagulaplli Naga Durga Tulasi**
**Guvi Email: tulasinnd@gmail.com**
**Phone: 9390427349**
**Date: 5/20/2023**

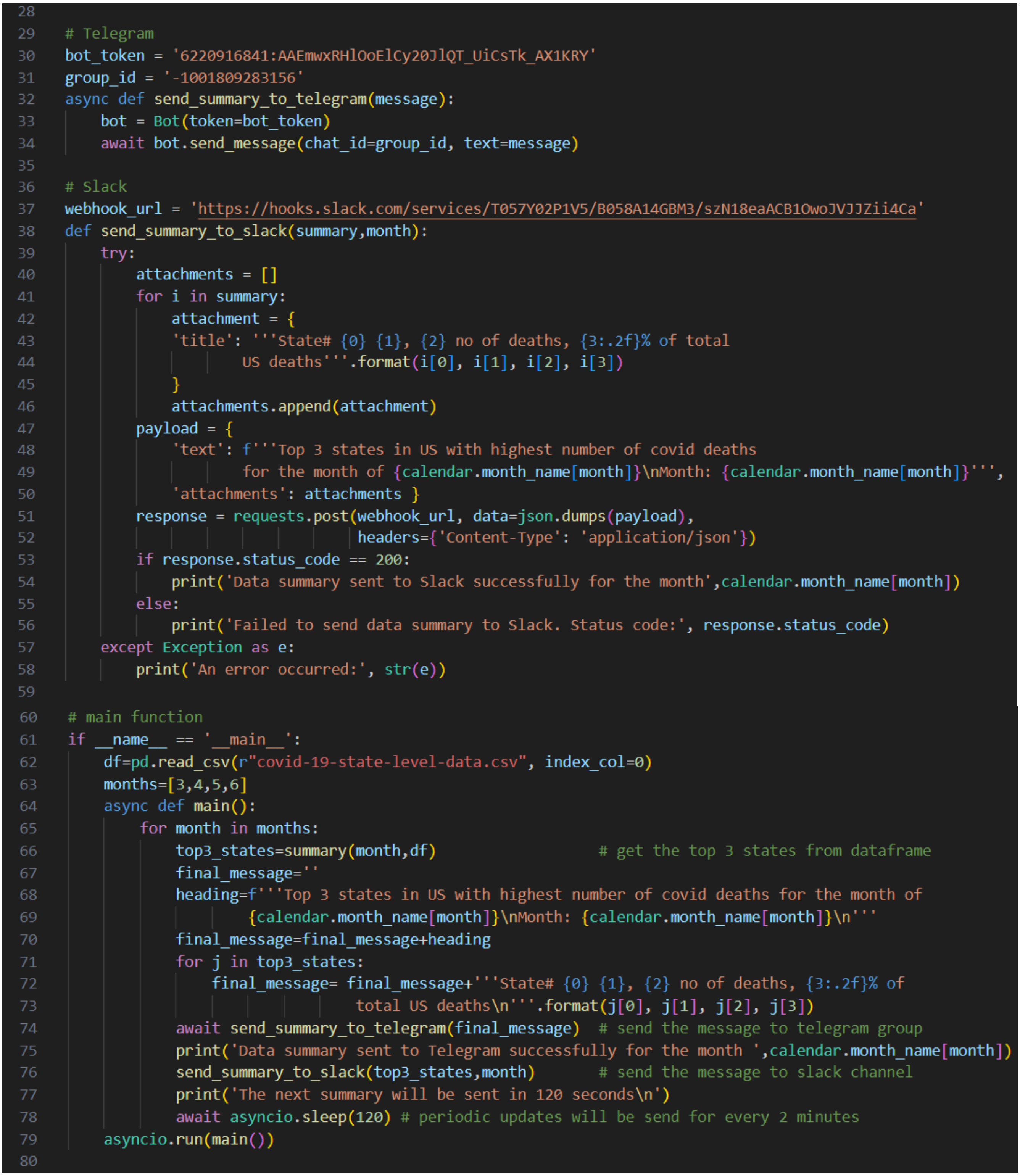
## Assignment Problem Statement:

Write an automated script using Python that sends a periodic data summary to Slack (preferred) or to an Email address or different mechanism (other than Slack/email). The goal is to send updates to a group of users about important metrics periodically. Here using covid19 dataset, post the monthly trend analysis of the number of covid deaths from the top 3 states in the US.

## Python Code:

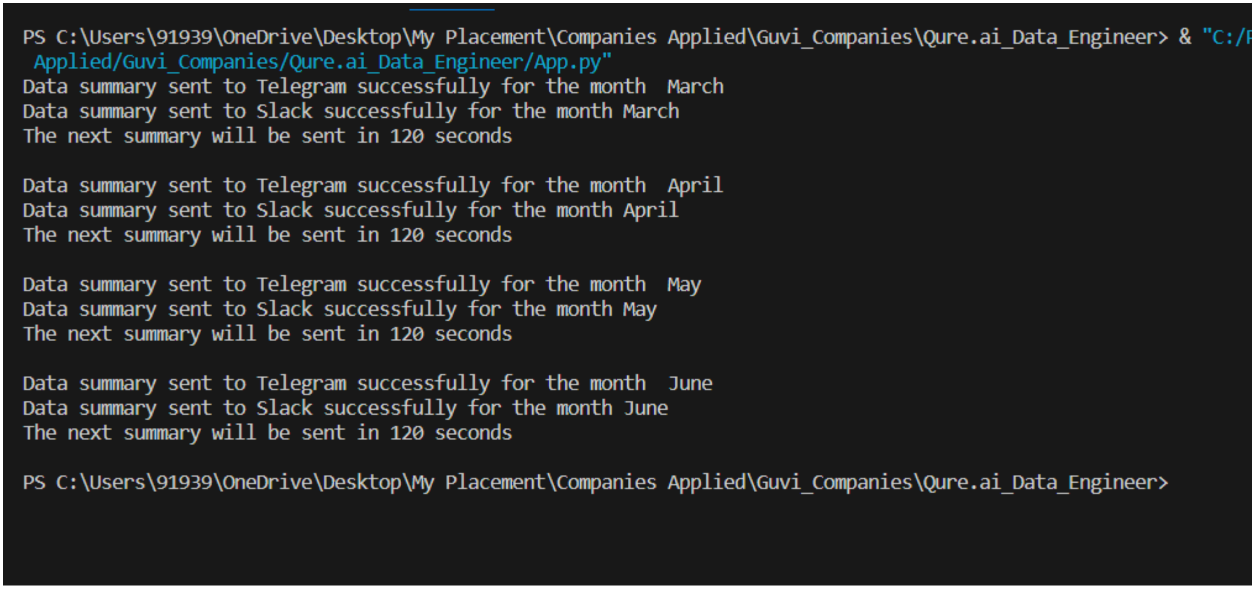


```python
import requests
import json
import pandas as pd
import calendar
import asyncio
from telegram import Bot

# Get top 3 states
def summary(month,df):
    df['date'] = pd.to_datetime(df['date'])
    mask = df['date'].dt.month.isin([month])
    df=df[mask]
    states=df.groupby('state')
    top=states['deaths'].sum().sort_values(ascending=False).head(3)
    top_df = top.to_frame()
    top_df.reset_index(inplace=True)
    top_df.columns = ['state', 'total_deaths']
    total=df['deaths'].sum()
    data_list = []
    for index, row in top_df.iterrows():
        l=index+1
        state = row['state']
        total_deaths = row['total_deaths']
        percentage = total_deaths / total * 100
        sublist = [l,state, total_deaths, percentage]
        data_list.append(sublist)
    return data_list
```



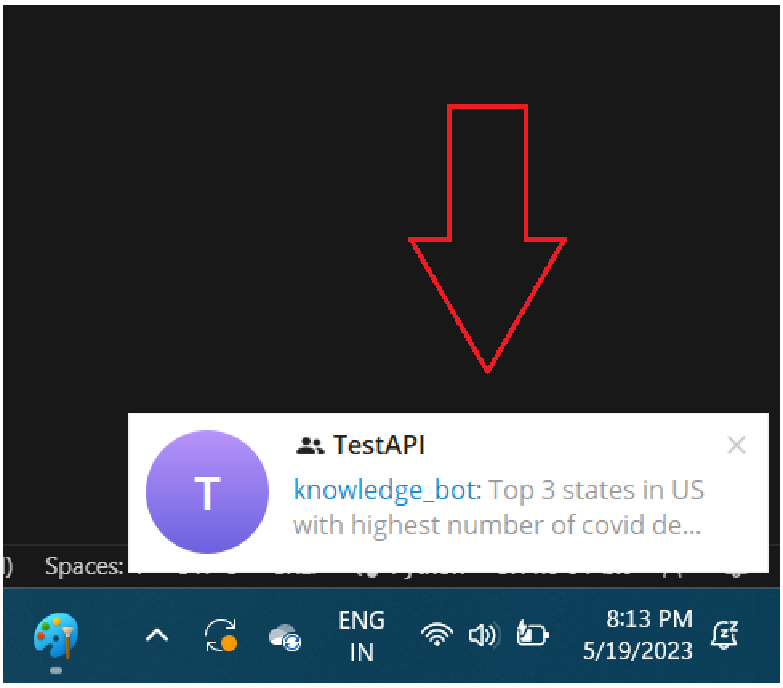
```python

# Telegram
bot_token = '6220916841:AAEmwxRHlOoElCy20JlQT_UiCsTk_AX1KRY'
group_id = '-1001809283156'
async def send_summary_to_telegram(message):
    bot = Bot(token=bot_token)
    await bot.send_message(chat_id=group_id, text=message)

# Slack
webhook_url = 'https://hooks.slack.com/services/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'
def send_summary_to_slack(summary,month):
    try:
        attachments = []
        for i in summary:
            attachment = {
            'title': '''State# {0} {1}, {2} no of deaths, {3:.2f}% of total
                    US deaths'''.format(i[0], i[1], i[2], i[3])
            }
            attachments.append(attachment)
        payload = {
            'text': f'''Top 3 states in US with highest number of covid deaths
                    for the month of {calendar.month_name[month]}\nMonth: {calendar.month_name[month]}''',
            'attachments': attachments }
        response = requests.post(webhook_url, data=json.dumps(payload),
                                 headers={'Content-Type': 'application/json'})
        if response.status_code == 200:
            print('Data summary sent to Slack successfully for the month',calendar.month_name[month])
        else:
            print('Failed to send data summary to Slack. Status code:', response.status_code)
    except Exception as e:
        print('An error occurred:', str(e))

# main function
if __name__ == '__main__':
    df=pd.read_csv(r"covid-19-state-level-data.csv", index_col=0)
    months=[3,4,5,6]
    async def main():
        for month in months:
            top3_states=summary(month,df)                   # get the top 3 states from dataframe
            final_message=''
            heading=f'''Top 3 states in US with highest number of covid deaths for the month of
                    {calendar.month_name[month]}\nMonth: {calendar.month_name[month]}\n'''
            final_message=final_message+heading
            for j in top3_states:
                final_message= final_message+'''State# {0} {1}, {2} no of deaths, {3:.2f}% of
                                total US deaths\n'''.format(j[0], j[1], j[2], j[3])
            await send_summary_to_telegram(final_message)  # send the message to telegram group
            print('Data summary sent to Telegram successfully for the month ',calendar.month_name[month])
            send_summary_to_slack(top3_states,month)        # send the message to slack channel
            print('The next summary will be sent in 120 seconds\n')
            await asyncio.sleep(120) # periodic updates will be send for every 2 minutes
    asyncio.run(main())

```

## Output Terminal:



```
PS C:\Users\91939\OneDrive\Desktop\My Placement\Companies Applied\Guvi_Companies\Qure.ai_Data_Engineer> & "C:/
 Applied/Guvi_Companies/Qure.ai_Data_Engineer/App.py"
Data summary sent to Telegram successfully for the month  March
Data summary sent to Slack successfully for the month March
The next summary will be sent in 120 seconds

Data summary sent to Telegram successfully for the month  April
Data summary sent to Slack successfully for the month April
The next summary will be sent in 120 seconds

Data summary sent to Telegram successfully for the month  May
Data summary sent to Slack successfully for the month May
The next summary will be sent in 120 seconds

Data summary sent to Telegram successfully for the month  June
Data summary sent to Slack successfully for the month June
The next summary will be sent in 120 seconds

PS C:\Users\91939\OneDrive\Desktop\My Placement\Companies Applied\Guvi_Companies\Qure.ai_Data_Engineer>
```

During the execution, I captured a screenshot of the telegram notification, which was sent by this Python script (note that telegram opened in the desktop and notifications allowed)

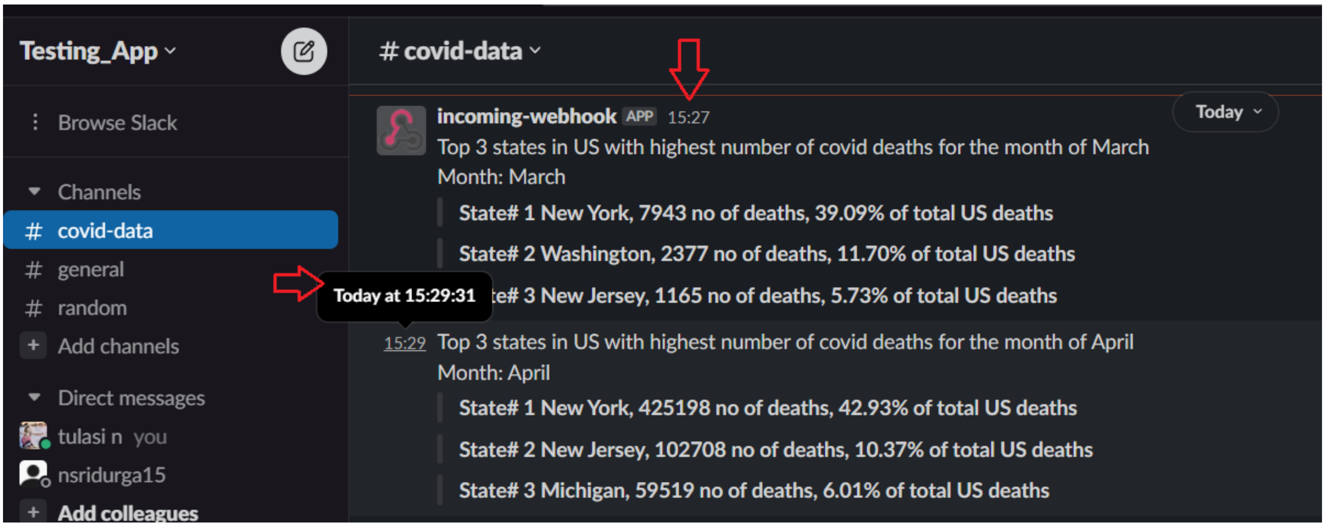


TestAPI

knowledge_bot: Top 3 states in US with highest number of covid de...

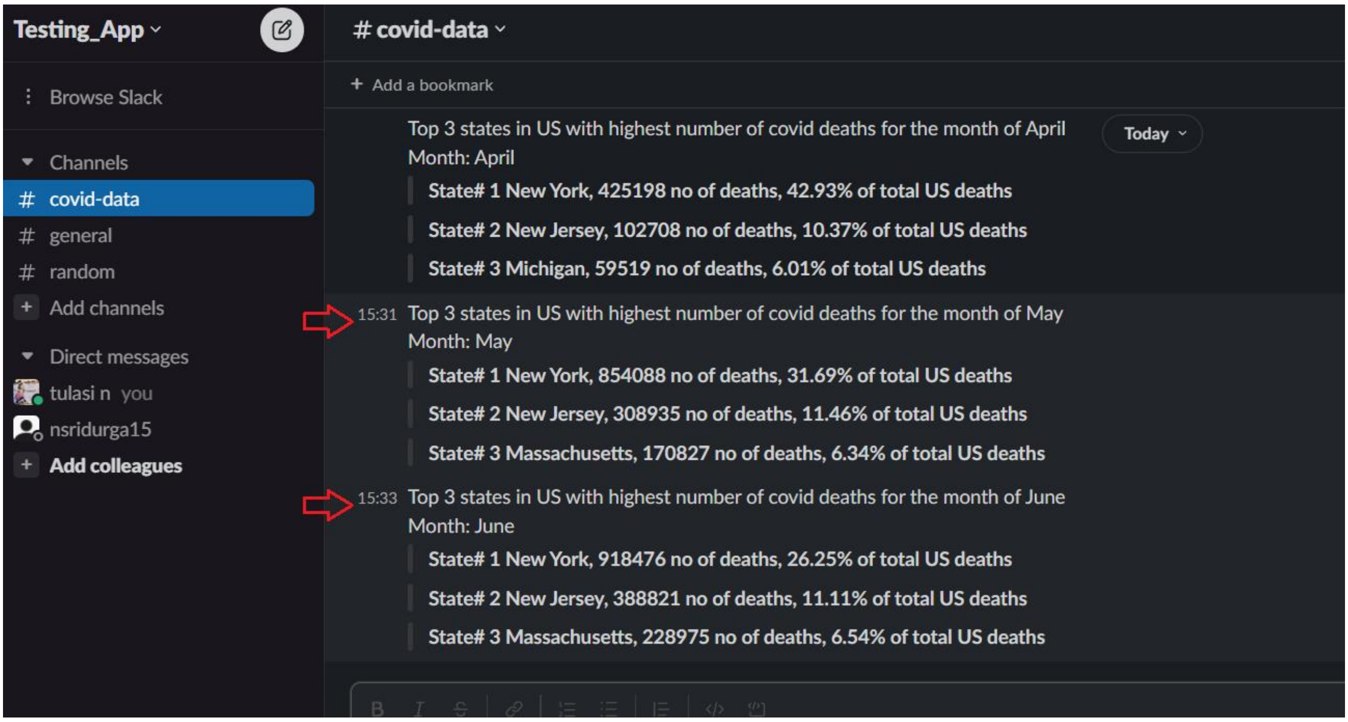
## Slack Channel Message Output:

We can clearly see that the summary related to March, April, May, and June with a fixed time interval of 2 minutes.



incoming-webhook APP 15:27
Top 3 states in US with highest number of covid deaths for the month of March
Month: March
State# 1 New York, 7943 no of deaths, 39.09% of total US deaths
State# 2 Washington, 2377 no of deaths, 11.70% of total US deaths
State# 3 New Jersey, 1165 no of deaths, 5.73% of total US deaths

15:29 Top 3 states in US with highest number of covid deaths for the month of April
Month: April
State# 1 New York, 425198 no of deaths, 42.93% of total US deaths
State# 2 New Jersey, 102708 no of deaths, 10.37% of total US deaths
State# 3 Michigan, 59519 no of deaths, 6.01% of total US deaths



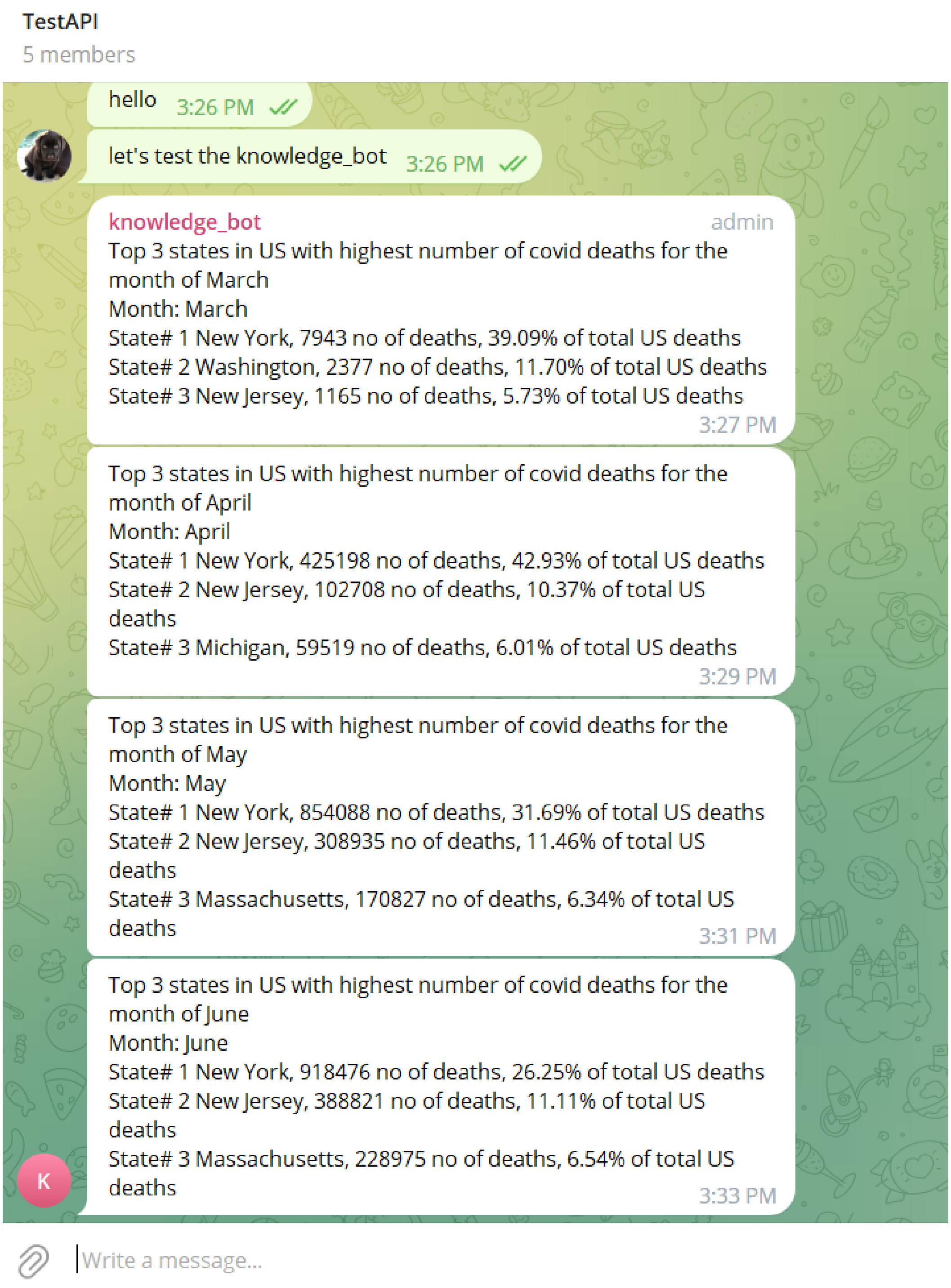
Top 3 states in US with highest number of covid deaths for the month of April
Month: April
State# 1 New York, 425198 no of deaths, 42.93% of total US deaths
State# 2 New Jersey, 102708 no of deaths, 10.37% of total US deaths
State# 3 Michigan, 59519 no of deaths, 6.01% of total US deaths

15:31 Top 3 states in US with highest number of covid deaths for the month of May
Month: May
State# 1 New York, 854088 no of deaths, 31.69% of total US deaths
State# 2 New Jersey, 308935 no of deaths, 11.46% of total US deaths
State# 3 Massachusetts, 170827 no of deaths, 6.34% of total US deaths

15:33 Top 3 states in US with highest number of covid deaths for the month of June
Month: June
State# 1 New York, 918476 no of deaths, 26.25% of total US deaths
State# 2 New Jersey, 388821 no of deaths, 11.11% of total US deaths
State# 3 Massachusetts, 228975 no of deaths, 6.54% of total US deaths

## Telegram Group Message Output:

I have created a telegram bot API called knowledge_bot and added it to the group TestAPI, (with all the necessary permissions) using which the periodic messages are delivered to people present in that group. We can see that there is a fixed time interval of 2 minutes for every new message which is automatically sent by the  Python script



TestAPI
5 members

hello 3:26 PM

let's test the knowledge_bot 3:26 PM

knowledge_bot admin
Top 3 states in US with highest number of covid deaths for the month of March
Month: March
State# 1 New York, 7943 no of deaths, 39.09% of total US deaths
State# 2 Washington, 2377 no of deaths, 11.70% of total US deaths
State# 3 New Jersey, 1165 no of deaths, 5.73% of total US deaths
3:27 PM

Top 3 states in US with highest number of covid deaths for the month of April
Month: April
State# 1 New York, 425198 no of deaths, 42.93% of total US deaths
State# 2 New Jersey, 102708 no of deaths, 10.37% of total US deaths
State# 3 Michigan, 59519 no of deaths, 6.01% of total US deaths
3:29 PM

Top 3 states in US with highest number of covid deaths for the month of May
Month: May
State# 1 New York, 854088 no of deaths, 31.69% of total US deaths
State# 2 New Jersey, 308935 no of deaths, 11.46% of total US deaths
State# 3 Massachusetts, 170827 no of deaths, 6.34% of total US deaths
3:31 PM

Top 3 states in US with highest number of covid deaths for the month of June
Month: June
State# 1 New York, 918476 no of deaths, 26.25% of total US deaths
State# 2 New Jersey, 388821 no of deaths, 11.11% of total US deaths
State# 3 Massachusetts, 228975 no of deaths, 6.54% of total US deaths
3:33 PM

## Python Script Explanation:

First of all, I have utilized Slack and Telegram to send periodic updates or notifications  to a group of users, and in case of Slack the messages will be sent to Slack channels using webhooks, and for Telegram the messages will be sent to Telegram groups using a Telegram bot API

## def summary()

This function is provided with a data frame and month, first I applied a mask to extract the records related to the given month, then I applied group by operation on the states column to get the group of states, I have used aggregated function sum() on deaths column and sorted them to obtain top3 states. The extracted summary will return as a list

## def send_summary_to_telegram()

This function sends messages to the Telegram group of a given group ID, here I have used the bot class which provides methods to interact with the Telegram Bot API, such as sending messages, editing messages, sending files, etc. You need to create an instance of the Bot class by passing your bot token as a parameter. Once the instance is created we can use the send_message() method to send any message to a particular group, the pre-requisite is there should be a bot added in the target group if we want to send periodic updates. By using the bot token and  Telegram group ID, the Telegram group will be uniquely identified and periodic messages will be delivered

## def send_summary_to_slack()

This method sends the messages to the Slack channels using the webhook URL of a particular channel, Slack webhooks are a powerful feature that allows you to send messages and notifications to Slack channels programmatically. A payload dictionary is created, which contains the text and attachments. The payload is then sent as a POST request to the Slack webhook URL using the requests.post method. The payload is converted to JSON format using json.dumps and sent in the request body. Here the webhook needs to be created for the channel inorder to send messages

## Periodic updates or Real-Time Notifications:

The main() function is defined as an async function and executed using asyncio.run() to run the asyncio event loop, it will continuously send updates to Slack and Telegram with a delay of 60 seconds between each update by calling send_summary_to_slack() and send_summary_to_telegram(). Here we can change the time delay according to our requirements. If we want to send the updates every hour we can change the time to 3600 seconds. To make this work in real-time the process needs to run continuously

There is a library called schedule in Python which is more flexible and provides additional time frames like seconds, minutes, hours, days, and even months. It can deal with more complex schedules like a day in a week or a particular time in a day or a particular day in a month. Since the task is to send updates in a fixed time interval, I have used asyncio. sleep() which serves the purpose.