perodical web data scrapper using python and saving data in google sheets with microsoft azure functions

Name: V Adithya

Email: adithyavibakar@gmail.com

m

# OUTLINE: -

1. Scrapping data using python and using google sheets as data base
2. Setting up azure functions with time triggering

I have chosen data from the website <https://covidindia.org/> which updates counts of corona cases, recovered and deaths daily. I have chosen Tamil Nadu state and the daily data of the states corona virus count will be entered to the google sheets daily at 9m automatically with the data and state name as Tamil Nadu and the counts.

1. **Scrapping data using python and using google sheets as data base:-**

I have used lxml and requests python library to extract contents from the webpage and filter the table tags <tr> from it and save it to a variable called table.

Then I have created a new list called data and appended date from the datetime library I have appended UTC time which is 5hrs and 30mins less than us.

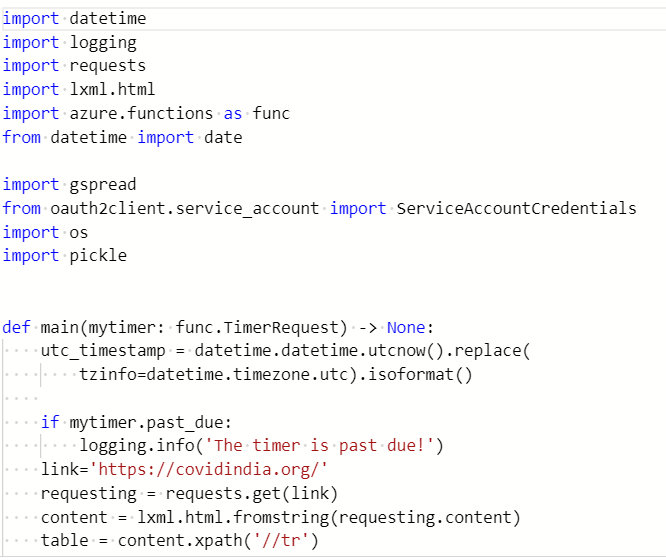
Then a dictionary is created containing credentials of the google sheet which I am using to save my data.

I tried to keep it as a sperate file and extract contents from it but I was not able to open it in azure python console I tried to save the token with the library with pickle so I just converted to credentials.json as a dictionary.

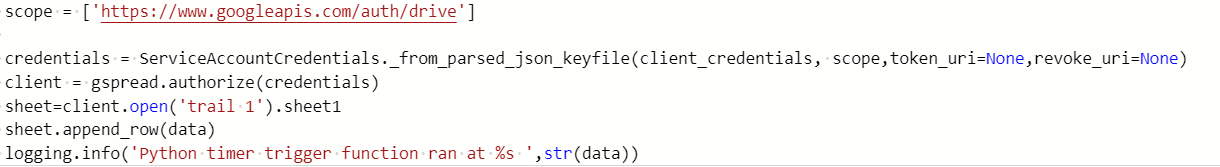
I appended data of table 31 containing data of counts of Tamil Nadu.

Then with use of library gspread and oauth2client library i updated data to the google sheet called trail 1.

I have attached my python code snippets below.



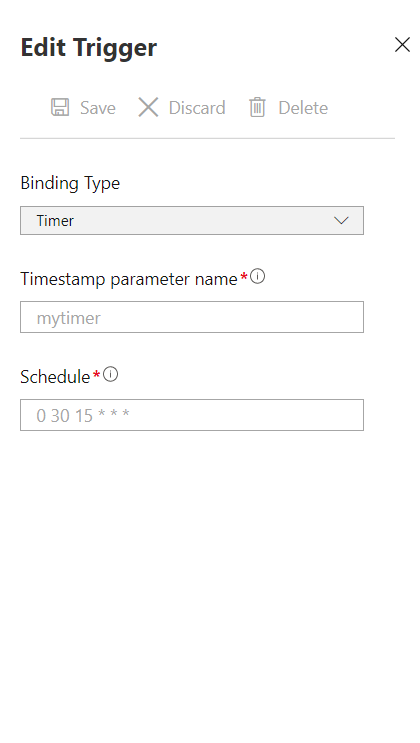




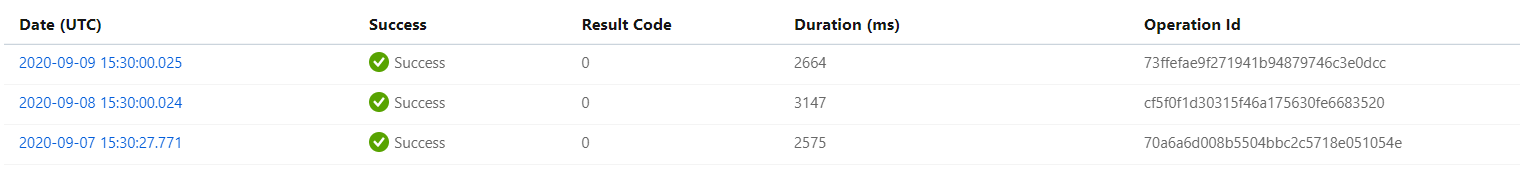
**2.Setting up azure functions with time triggering: -**

The next part is creating a time triggered events that is croning azure functions I totally similar to aws lambda but the only difference is that you can set up a local machine and debug in the local machine and then deploy the function. I used azure because I have free student subscription and it does not ask for credit cards for checking.

The cron expression that I have setup is 0 30 15 \* \* \* according to UTC time since the storage location is at North Europe so that it gets updated at 9pm according to Indian time perfectly.



Below are the screen shots of the logging information: -

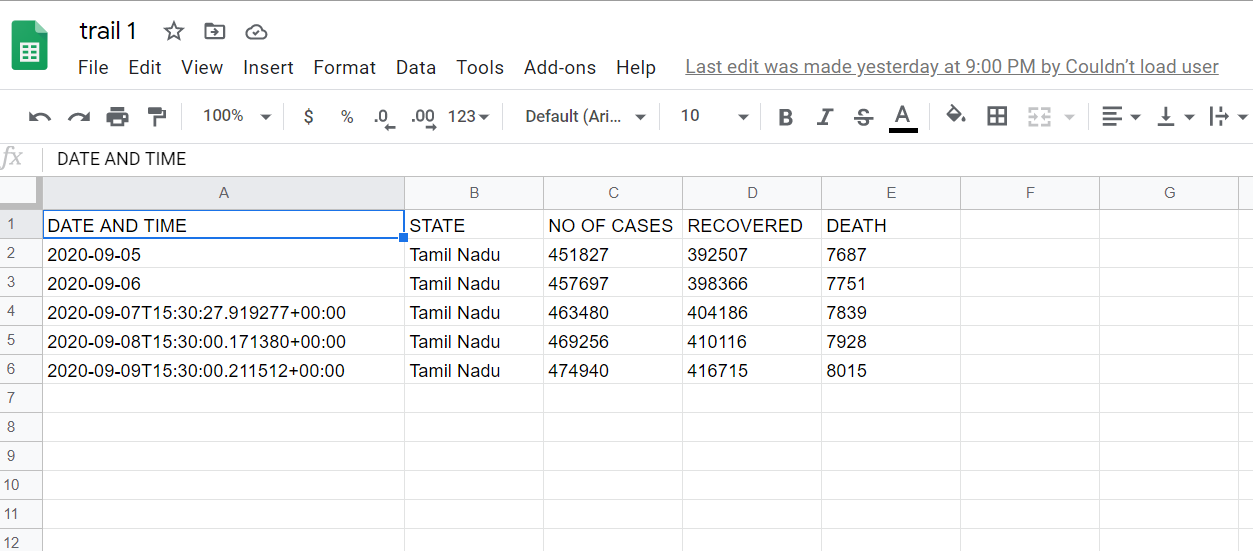


And below is the detailed log of last updating.

The Google spread sheet

<https://docs.google.com/spreadsheets/d/1gZEmXJTijFglb7iOEV3TJgdLA5819CCpYzGwTTkr7J4/edit?usp=sharing>

this is the link of the google spread sheet am attaching a screen shot also



The time is the UTC time the first two entries are historical data and other three are updated by azure functions for three days u can see the edits made was yesterday 9pm so its working perfectly and updating correctly at 9pm.

The link for the sheet I given and the permission is anyone with the link can view .