**Python Project**

I am always fascinated about how data is all pulled together into and maintained, especially when it is being sourced and pulled from different sources and web sites. Looking at the latest covid-19 git hub repositories, it is clear that lots of work has gone into generating the data and providing a time line data set to analyse.

While I have found some great sources where the data is perfect in csv file , I would like to go through my own process of scraping the web data and building a time line of data.

I have found a time machine website that snaps web page that has a table of data.

This is the link: [https://web.archive.org/web/\*/http://www.worldometers.info/coronavirus/](https://web.archive.org/web/*/http:/www.worldometers.info/coronavirus/)

I would like to do the following with python:

1.Build a list of all the snaps on this page, There are multiple snaps per day.

2. Because the Table columns are not consistent over time, I need to do a check on this by counting and each day and snap shots columns and column count

3.Web Scape the HTML each page Table contents , using the first snap of each day

4. Store the data Into a SQL database, and increment it to refresh each day

5. Validate with other sources and and Visualise the data.

**PART 1**

[I](https://web.archive.org/web/*/http:/www.worldometers.info/coronavirus/) would like to use this link to get a list of all the snap shots that this site has , and maybe even a url for each page link , so i can loop through them ;

Parent Page :

[https://web.archive.org/web/\*/http://www.worldometers.info/coronavirus](https://web.archive.org/web/*/http:/www.worldometers.info/coronavirus)

Example of a child Page ;

<https://web.archive.org/web/20200301001656/www.worldometers.info/coronavirus/>

This site stores snaps of a certain web page , and does it multiple times a day, and it isn't at a constant time so i can't pin point exactly the link part '[20200301001656](https://web.archive.org/web/20200301001656/www.worldometers.info/coronavirus/)'