



Assignment #3

part 2

Due Date: 24 October 2021
Sunday, 11.00 am

DSE 309
Advanced Programming
in Python

Working with COVID-19 Data

This assignment is the culmination of all the knowledge you have gained in the earlier classes.

- Many organizations are keeping track of COVID-19 cases worldwide and updating the data on their website and web services periodically. The most prominent are Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>)
- and World-O-Meter (<https://www.worldometers.info/coronavirus/>). These are very reliable sources of data for COVID-19 and they update their statistics very frequently (at least once every 24 hours) so downstream systems get the latest data.

- We can retrieve this data using custom libraries in Python. One such library can be found at <https://ahmednafies.github.io/covid/>. It can retrieve the data from both Johns Hopkins University and World-O-Meter. To install it, create a new notebook for this chapter and run the following command in a code cell: `!pip3 install covid`
- Next, import the library as follows: `from covid import Covid`
- You can fetch the data using this code: `covid = Covid()`
- It fetches the data from Johns Hopkins University by default. You can also explicitly mention the data source: `covid = Covid(source="john_hopkins")`
- To fetch the data from World-O-Meter, change the source value: `covid = Covid(source="worldometers")`
- You can display all the data using the following command: `covid.get_data()`
- This returns a list of dictionaries, as shown in Figure

COVID-19 data

```
In [35]: # get all data
         covid.get_data()
```

```
Out[35]: [{'country': 'North America',
            'confirmed': 3628797,
            'new_cases': 7809,
            'deaths': 178674,
            'recovered': 1648713,
            'active': 1801410,
            'critical': 18601,
            'new_deaths': 945,
            'total_tests': 0,
            'total_tests_per_million': Decimal('0'),
            'total_cases_per_million': Decimal('0'),
            'total_deaths_per_million': Decimal('0'),
```

- We can determine the source of the data as follows:`covid.source`
- The output in this case is shown here: 'worldometers'
- You can also retrieve the status by country name as follows:`covid.get_status_by_country_name("italy")`
- The result is shown in Figure

```
In [37]: covid.get_status_by_country_name("italy")  
Out[37]: {'country': 'Italy',  
          'confirmed': 241956,  
          'new_cases': 0,  
          'deaths': 34899,  
          'recovered': 192815,  
          'active': 14242,  
          'critical': 70,  
          'new_deaths': 0,  
          'total_tests': 5703673,  
          'total_tests_per_million': Decimal('0'),  
          'total_cases_per_million': Decimal('4002'),  
          'total_deaths_per_million': Decimal('577'),  
          'population': Decimal('60459826')}
```

COVID-19 data by
country

- You can retrieve the data by country ID, too (this function is only valid for the Johns Hopkins data source), with this code:
- `covid.get_status_by_country_id(115)`
- To retrieve the list of countries affected by the COVID-19 pandemic, use this syntax:
- `covid.list_countries()`
- It returns the list shown in Figure

```
Out[17]: ['north america',  
          'south america',  
          'asia',  
          'europe',  
          'africa',  
          'oceania',  
          '',  
          'world',  
          'usa',  
          'brazil',  
          'india',  
          'russia',  
          'peru',  
          'chile',  
          'spain',
```

Countries affected by COVID-19

- The total number of active cases can be obtained as follows:
`covid.get_total_active_cases()`
- The total number of confirmed cases can be obtained as follows:
`covid.get_total_confirmed_cases()`
- The total number of recovered cases can be obtained as follows:
`covid.get_total_recovered()`
- The total number of deaths can be obtained as follows:
`covid.get_total_deaths()`
- Run those statements and examine the output

Visualizing the COVID-19 Data

- Now you can convert all this data into a pandas dataframe as follows:

```
import pandas as pd
```
- ```
df = pd.DataFrame(covid.get_data())
```
- ```
print(df)
```

COVID-19 data converted to a dataframe

```
import pandas as pd
df = pd.DataFrame(covid.get_data())
print(df)
```

	country	confirmed	new_cases	deaths	recovered	active	\
0	North America	3628797	7809	178674	1648713	1801410	
1	South America	2614931	1036	96832	1717350	800749	
2	Asia	2700746	20485	64867	1839062	796817	
3	Europe	2513631	8954	194782	1462217	856632	
4	Africa	511949	779	12026	248751	251172	
..	
218	Caribbean Netherlands	7	0	0	7	0	
219	St. Barth	6	0	0	6	0	
220	Anguilla	3	0	0	3	0	
221	Saint Pierre Miquelon	1	0	0	1	0	