Pandas Task

Downloading the data

Download the covid-data set from https://ourworldindata.org/covid-deaths

Click on the *"Download"* option located at the bottom right. You will get a CSV

Task Description

After downloading the data, we will do the following:

- Import needed packages, note that in case you don't have pandas you can download it simple by typing : pip install pandas
- Then, check how many rows and columns are in the dataframe
- View the first 10 rows in the data and the last 5
- Show the basic summary of data such as count, min, max, unique values, etc
- Drop the following columns: new_deaths_smoothed','new_cases_per_million','total_cases_per_million'
- Rename the following columns:

  'date': 'Date','location':'Country','continent': 'Continent','iso_code':'ISO_code'
- List the continent name:

  Create a list that holds the unique values of continent name
- There is something in sklearn library called Simple imputer helps with missing values in a dataset, read about it here: https://www.geeksforgeeks.org/ml-handle-missing-data-with-simple-imputer/

  What we need is to impute our dataframe just like this:

```
imputer = SimpleImputer(strategy='constant')
df2 = pd.DataFrame(imputer.fit_transform(df),columns=df.columns)
```

- Groupby, `df.groupby()`: The groupby command allows us to divide our data into different groups and perform our data analysis.

  Run this, what is the output from this statment?

```
df3 = df2.groupby(['Date','Country',])[['Date','Country','total_cases','total_deaths','total_vaccinations']].sum().reset_index()
```

- You shoul see there are missing_value because we have used simple imputer to add constant value at all missing places. If you want you can even replace that missing_value with 0

  To solve this we have replace method : https://datatofish.com/replace-values-pandas-dataframe/#:~:text=Suppose that you want to,value'%2C...%5D)

  Replace "missing_value" to 0 for the following columns: ['total_cases','total_deaths','total_vaccinations']

- From df3: Find total countries where total_deaths is greater than 1000000.
- How many dates we have in total where total_deaths is greater than 1000000