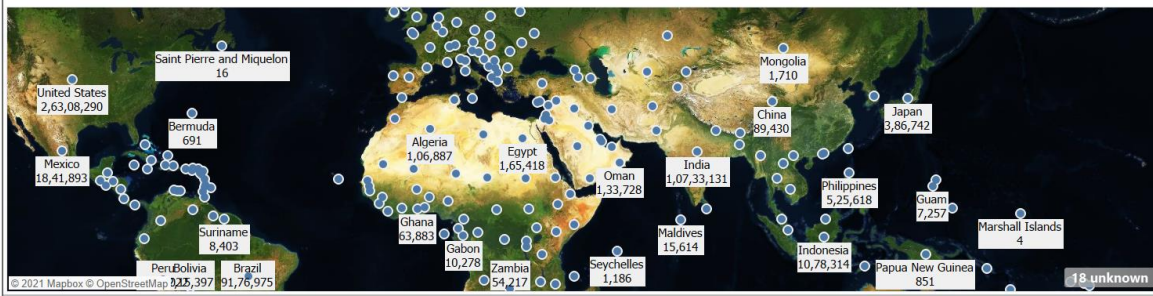
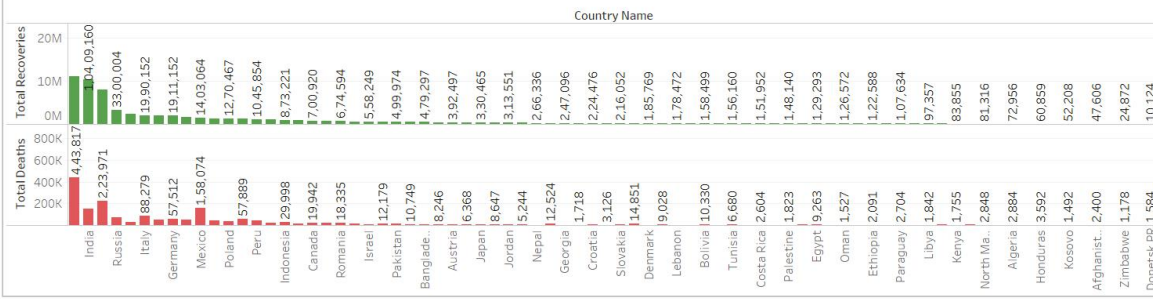


Sheet 1



Sheet 2



```

0: My_First_Projects\main.py
main.py
3
4 #Step 1 Data Collection
5
6 url = 'https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data'
7 req = requests.get(url)
8 data_list = pd.read_html(req.text)
9 target_df = data_list[0]
10
11 # Step 2: Data Cleaning
12
13 #Issue 1: change column names
14 target_df.columns = ['col0', 'Country Name', 'Total Cases', 'Total Deaths', 'Total Recoveries', 'col5']
15
16 #Issue 2: remove 1st and 5th column
17 target_df = target_df[['Country Name', 'Total Cases', 'Total Deaths', 'Total Recoveries']]
18
19 #Issue 3: last and second last rows
20 #target_df = target_df.drop([238, 239]) (not dynamic)
21 last_idx = target_df.index[-1]
22 target_df = target_df.drop([last_idx, last_idx-1])
23
24 #Issue 4: inconsistent country names
25 target_df['Country Name'] = target_df['Country Name'].str.replace('[\.\s]', '')
26
27 #Issue 5: replace "not data value" cell by "0"
28 target_df['Total Recoveries'] = target_df['Total Recoveries'].str.replace('No data', '0')
29 target_df['Total Deaths'] = target_df['Total Deaths'].str.replace('No data', '0')
30 target_df['Total Cases'] = target_df['Total Cases'].str.replace('No data', '0')
31 target_df['Total Deaths'] = target_df['Total Deaths'].str.replace('+', '')
32
33 #Issue 6: Change data type
34 print(target_df.dtypes) #to check data types of entities
35 target_df['Total Cases'] = pd.to_numeric(target_df['Total Cases'])
36 target_df['Total Deaths'] = pd.to_numeric(target_df['Total Deaths'])
37 target_df['Total Recoveries'] = pd.to_numeric(target_df['Total Recoveries'])
38
39 #Step 3: Export data
40 #target_df.to_csv(r'Covid_19_dataset.csv') (use if required)
41 target_df.to_excel(r'Covid_19_dataset.xlsx')
42

```

Variable explorer

Name	Type	Size	Value
data_list	list	25	[Dataframe, Dataframe, Dataframe, Dataframe, Dataframe, Dataframe, Dat ...
last_idx	int	1	239
req	models.Response	1	Response object of requests.models module
target_df	Dataframe	(238, 4)	Column names: Country Name, Total Cases, Total Deaths, Total Recoverie ...
url	str	1	https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data

Python 3.8.5 (default, Sep 3 2020, 21:29:08) [MSC v.1916 64 bit (AMD64)]  
Type "copyright", "credits" or "license()" for more information.

IPython 7.19.0 -- An enhanced Interactive Python.

```

In [1]: runfile('D:/My_First_Projects/main.py', wdir='D:/My_First_Projects')
Country Name      object
Total Cases       object
Total Deaths      object
Total Recoveries  object
dtype: object

In [2]:

```