# Exercise I

The task involves reading data from a CSV file, performing some data transformations using Python, and then outputting the cleaned and enhanced data into a JSON file. Follow the steps below to achieve this outcome:

1. Load the provided CSV (sample\_data.csv) file into a Pandas DataFrame.

import pandas as pd

df=pd.read\_csv(r'C:Users/Ahmad K/Documents/ESCWA/data\_sample.csv')

print(df)

1. Replace all nan values in the DataFrame with empty cells.

df.replace(to\_replace= "nan", value="", inplace=True)

1. Remove any trailing spaces from the country names in the GeoAreaName column.

df['GeoAreaName'] = df['GeoAreaName'].str.strip()

1. Add two new columns, ISO and GeoAreaName\_ar, using the information from the attached Excel file (country\_info.xlsx).

country\_info = pd.read\_excel('country\_info.xlsx')

country\_info=country\_info.rename({'ISO3':'ISO','Arabic Name':'GeoAreaName\_ar'})

df = df.merge(country\_info[['ISO', 'GeoAreaName\_ar']], on='GeoAreaName', how='left')

1. Filter out any rows that don't match an ISO code from the Excel file and save these rows in a separate DataFrame, keeping the original row indices.

    filtered\_df = df[df['ISO'].isnull()]  # Rows with no ISO code

df = df[df['ISO'].notnull()]           # Remaining rows with ISO code

1. Drop any columns in the original DataFrame that have no data.

df.dropna(axis=1, how='all', inplace=True)

1. Save the cleaned DataFrame into a JSON file entitled output\_data.json.

df.to\_json('output\_data.json', orient='records', lines=True)

# Exercise II

The task involves reading data from a API, performing some data transformations using Python, and then outputting the cleaned and enhanced data into a csv file. Follow the steps below to achieve this outcome:

1. Load data from from the ‘data ’ object in the API <https://unstats.un.org/SDGAPI/v1/sdg/Series/Data?page=100> for the pages 100 to 105 into a panda DataFrame named df1.

import pandas as pd

import requests

df1 = pd.DataFrame()

for page in range(100, 106):

    url = f'https://unstats.un.org/SDGAPI/v1/sdg/Series/Data?page={page}'

    response = requests.get(url)

    data = response.json().get('data', [])

1. Load data from the ‘data ’ object in the API [https://unstats.un.org/SDGAPI/v1/sdg/Series/Data?page=117430](https://unstats.un.org/SDGAPI/v1/sdg/Series/Data?page=100) from page 117430 to page 117435 into a pandas DataFrame named df2.

import pandas as pd

import requests

df2 = pd.DataFrame()

for page in range(117430, 117436):

    url = f'https://unstats.un.org/SDGAPI/v1/sdg/Series/Data?page={page}'

    response = requests.get(url)

    data = response.json().get('data', [])

1. Merge df1 and df2 into a single DataFrame df3.

df3 = pd.concat([df1, df2], ignore\_index=True)

1. Convert ‘Value’ column in df3 from a string to float.

df3['value']=df3['value'].astype(float)

1. Extract a list of unique sources from the ‘source’ column in df3.

unique\_sources = df3['source'].unique().tolist()

1. Save the final DataFrame as output\_data.csv.

df3.to\_csv('output\_data.csv', index=False)