Assignment # 1

1-First of all, after opening the excel file, I saw 4 rows in the beginning of the file that should be deleted to have data with specified columns and values.

It can be by hand and also with python code, after reading the CSV file with:

pd.read_csv(df_GDP , skiprows=4)

I took the 4 rows away.

2-When I look at the files, there are 2 things visible to need have a solution, Nan values and some rows data that is not belong to countries like World, IDA total, Africa Eastern and Southern, heavily indebted poor countries (HIPC) and…. should make decision how to consider them.

-In python, for Nan values

df.fillna(0, inplace=True) and

-separating country names

import pycountry

country_names = [country.name for country in pycountry.countries]

Filter the DataFrame based on the list of country names

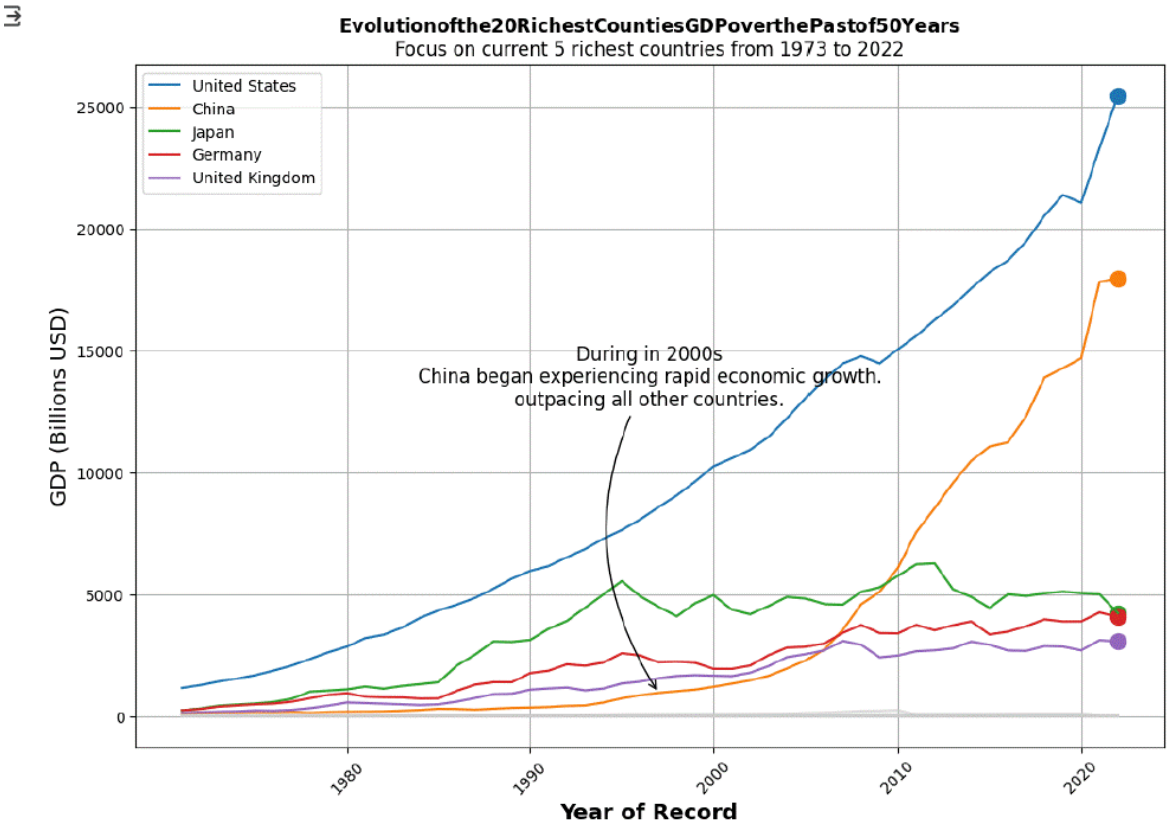
country_df = df[df['Country Name'].isin(country_names)]

3-Considering the image, I can see 5 richest countries in legend, so it needs to calculate mean values for years and select top 5.

Two important items, X-axis and Y-axis, xaxis is bolded from 1980, so for columns of data frame , I set the gdp_columns = df.columns[15:-1] (-1 , not last column 'Unnamed') and in yaxis , it seems the amount are not higher than 25,000 , so I used MinMaxScalaer [0,25000] to put amounts in this range.

4-Now in the plotting,

First of all, plot the 5 countries with for loop and show legend and for other countries with light grey, although I could not reach to your mere image exactly.



https://colab.research.google.com/drive/1Ij6BTNG240C0q5k5Z6_DJgx6Bhn4iBfM?usp=sharing

In Excel

1-Delete 4 rows from beginning

2-Replace all empty cell to Zero, and calculation of mean for every rows. =AVERAGE (Y2:BO2), Y2 from 1980

3-Search and find an Excel on the internet for list of countries to separate non countries and copy the countries and code after all columns as new ones.

4- Put the average in the front of name of real countries, with this formula, New_AVG =IFERROR (INDEX (BP: BP, MATCH (BR2, A: A, 0)), 0)

5-Sort from Z to A

6-Insert - > Pivot table, select the New_AVG and country, sorting on New_AVG and filter just 5 from top.

7-Pivot Chart