## 1  Data Collection

* Download JSON data files containing economic indicators from reputable sources (e.g. such as the World Bank or IMF).
* Ensure the data spans at least 5 years for multiple countries.

First, load the common library required for this assignment.

The data we use is from **the World Bank**. First of all, the country, the indicators (including the code in the database of the International Standards Organization) and the generation time are defined. And a basic query structure "base\_url" is written according to the requirements of the website.

Iterate over the dictionary of indicators collected earlier, creating a new dictionary (data) with the indicator's name as the key and an empty dictionary as the value.

Obtain different APIs by filling different countries, indicators, and years. Obtain the useful information in the json file under the API and load it into the dictionary "data" and save it into the json file "**world\_bank\_data.json**".

*Note that : Because most of the data sets available online are not in json format, we obtained the json file using the API to obtain the transfer file*

## 2  Data Preparation

* If required, load the JSON files into Pandas DataFrames.
* Clean the data by handling missing values, duplicates, and incorrect data types.

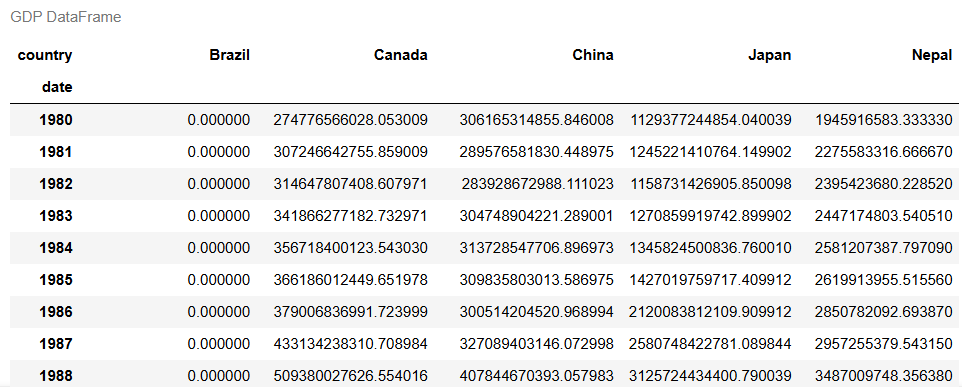
To convert the data in the json file into dataframes and perform data cleaning, we designed three functions.

The first function **generates a dataframe**. First, the corresponding data is extracted from the dictionary, the list is generated in the format of "**[country name, time, index value]**", and the data frame is generated according to the list. The dataframe is then adjusted to a format where the **date** becomes the row index and the **country** becomes the column index, returning the adjusted data box.

The second function is responsible for **the canonical output of the dataframe**. Walking through the generated data boxes converts them to a more observable HTML format.



The third function is responsible for **data cleaning**. Mark the missing value in the data box and replace it with **0**.

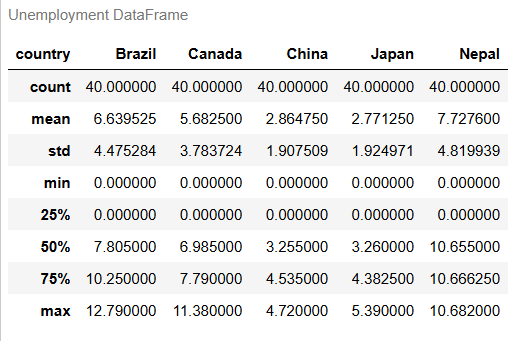
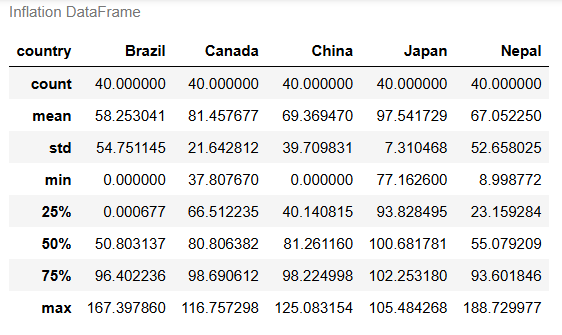


By calling these functions in turn, we realize the **cleaning and rendering** of the metadata

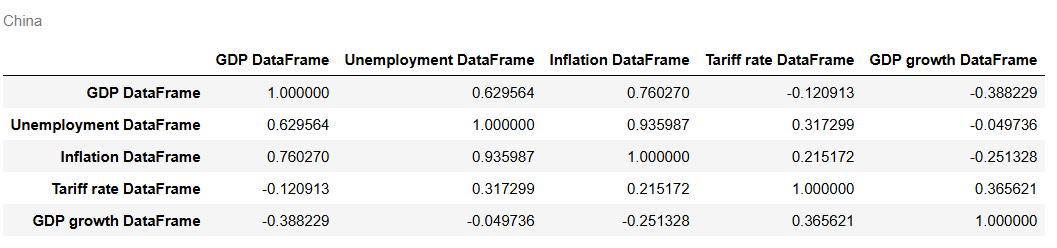
## 3  Data Analysis

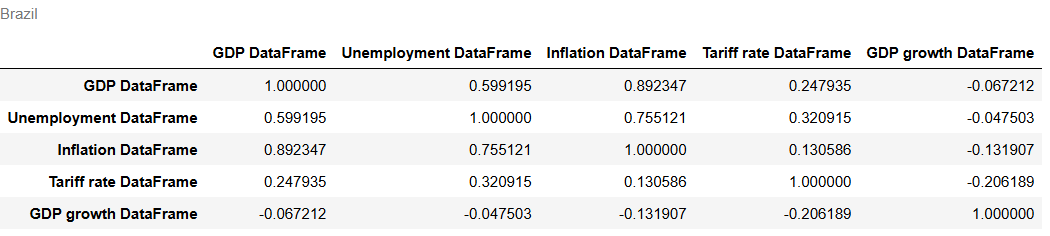
* Perform exploratory data analysis (EDA) to understand the datasets.
* Calculate summary statistics and identify trends over time for each indicator.

First, **descriptive statistical analysis** is carried out. Iterate through the dictionary of dataframes generated in the previous question, print descriptive statistics for each dataframe, and store the results in the dictionary.

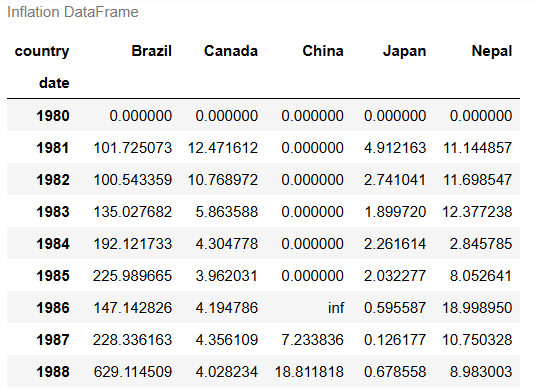
 

Then, **the correlation between each index** is analyzed. Create a function that creates a dataframe for each country. These data boxes are stored in a dictionary (country\_dfs). Traverse these dataframes, using the **df.corr()** function to calculate **the correlation** between the columns in the data box.





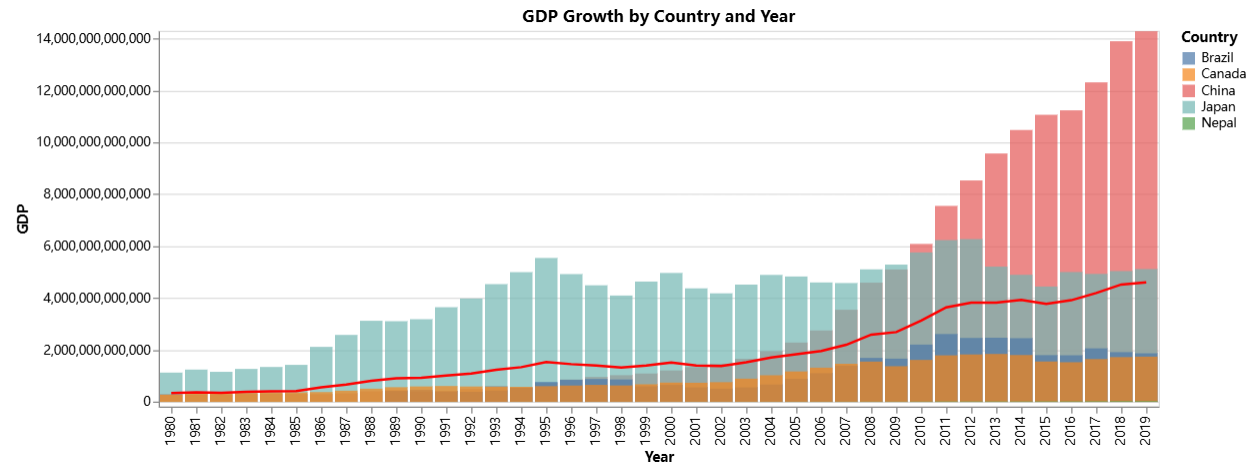
Finally, the change trend of each index over time is calculated. Call the dictionary created above containing each metric, calculate the percentage change between adjacent rows, and then call the created function above to work with the missing values, printing the trend of each metric over time



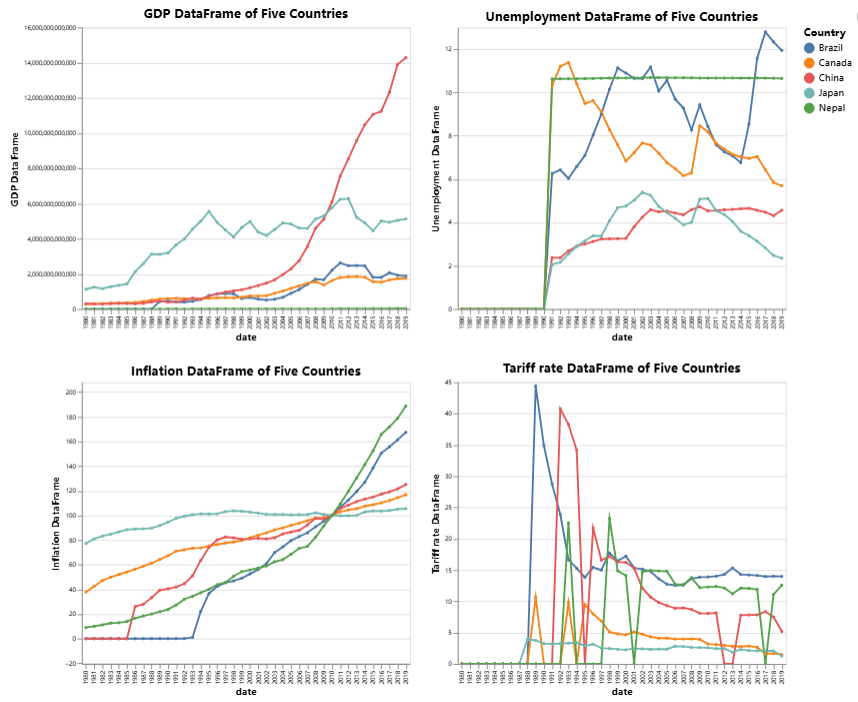
## 4 Data Visualization

* Create visualizations to represent the trends and patterns discovered during EDA.
* Use line plots, bar charts, and scatter plots to compare different economic indicators across countries.
* Employ Seaborn’s advanced features like facet grids or pair plots to visualize multi-dimensional data.

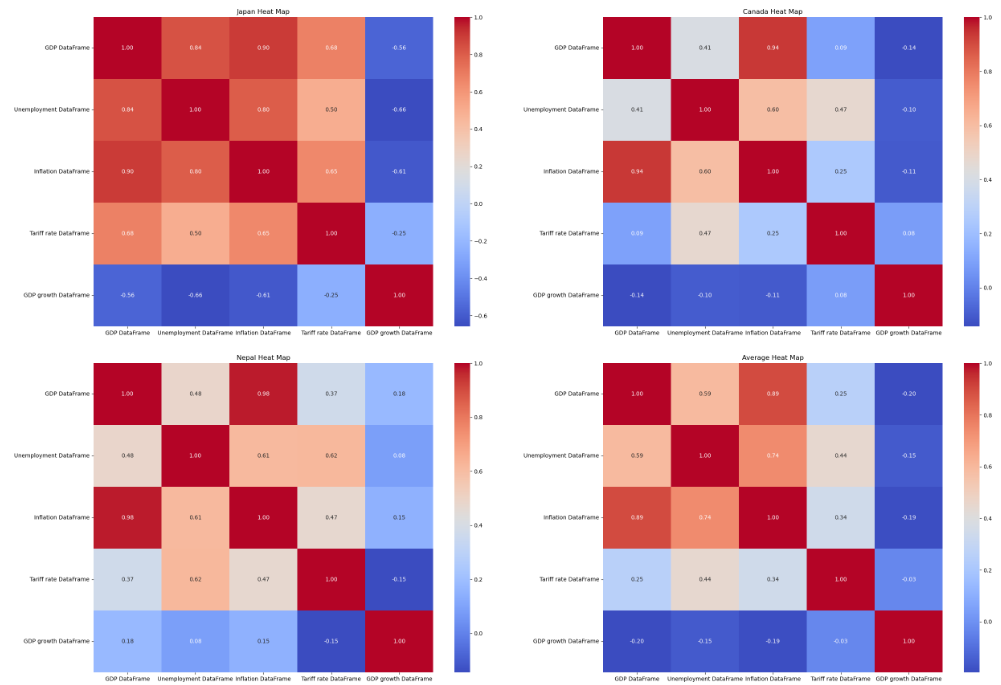
**Bar chart of GDP over time for five countries.** The code creates a bar chart based on the dataframe "df" to visualize GDP data for five countries over time. The chart uses a different color for each country and includes an interactive feature that enhances exploration. In addition, the bar chart is overlaid with a red line representing **the average GDP growth of all countries**, highlighting the overall trend. This comprehensive visualization provides a clear comparison of GDP growth by country and year, as well as an overall trend line of average GDP growth.



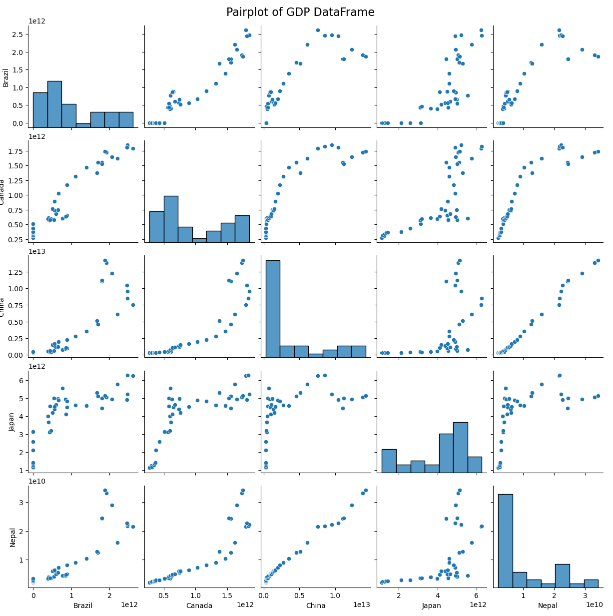
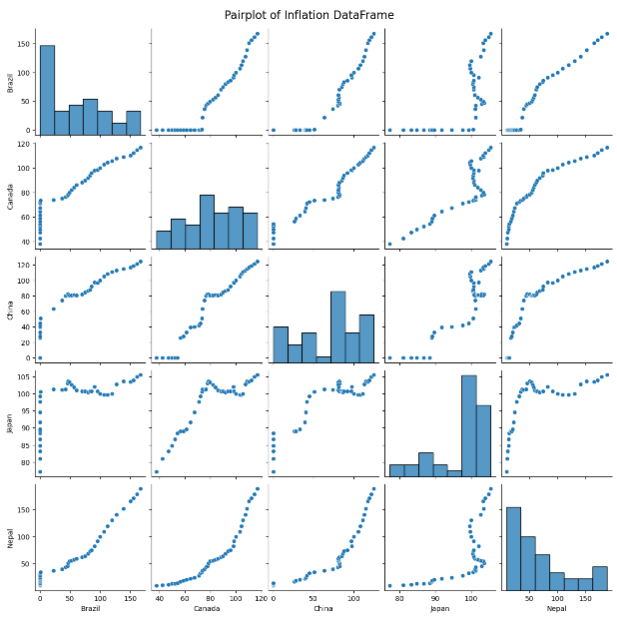
**Comparison of line charts for different indicators in different countries.** A function “**draw\_line\_chart**” is defined here to create an interactive line chart for each indicator in the five countries. Each chart shows trends over time, with different colors representing different countries. The graphs generated by each metric are placed in the list "**charts**", organized into a grid layout with two charts per row, ensuring clear and structured visual comparisons.



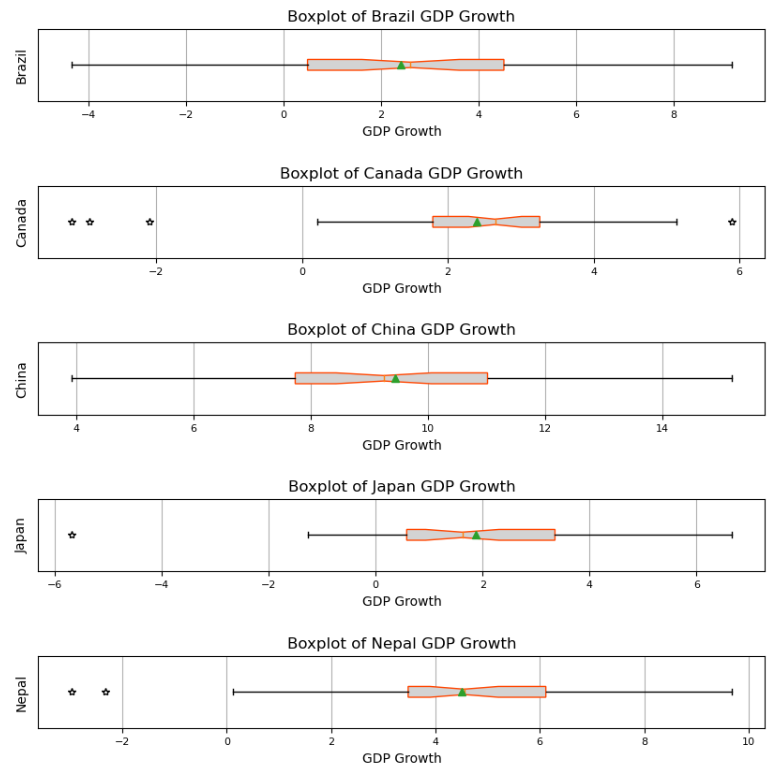
**Heat maps by country.** Using the **Seaborn** library, visualize correlations between individual data for each country in the df\_corrs dictionary to create heat maps. Each heat map uses color gradients to show the strength and direction of the relationship between the indicators. In addition, the average heat map is calculated by taking the average of all the relevant matrices to provide a comprehensive view of the overall trends and patterns of the data for all countries.



**Pair plots.** The code generates pair plots using **Seaborn** to visualize the relationships between GDPs of different countries. Traversing the dictionary "dfs" generates pair plots for each indicator. Each pair plot shows how the indicator of one country correlates with the indicators of other countries, highlighting patterns and potential influences between them. The pair plots include both scatter plots and histograms, providing insights into the distribution and pairwise relationships of GDP data.

****

**Box plots of GDP growth.** Creating box plots for the GDP growth rates of different countries allows for a quick visual comparison of their economic performance. Box plots summarize key statistics like the median, quartiles, and outliers, making it easy to identify trends, variability, and unusual data points across countries. It simplifies complex data, helping to highlight differences and similarities in GDP growth rates effectively.

****

## 5. Interpretation

* Interpret the visualizations to draw meaningful conclusions about theeconomic health of the countries.
* Discuss any correlations or anomalies found between differenteconomic indicators

**From GDP Growth by Country and Year Picture**

GDP growth: The average level of global GDP has risen steadily over the past 40 years, showing the continued improvement of the world economy in recent years.

Japan: The best economy until 2010, and a clear lead until 1991.

China: Surpassed Japan after 2010, and the economic gap widened with each passing year.

**From Line chart of different indicators**

Unemployment rate: High from 1999 to 2004, then declining steadily.

Inflation: Rising since 1980, Nepal has risen most sharply, possibly due to internal unrest and an unstable external environment.

Tariff rates: have declined gradually since 1980, reflecting continued openness and progress in world trade.

GDP growth rate: The GDP growth rate of each country fluctuates around a specific average, especially in China.

**From pair plots**

Cross-country indicator correlation: Most indicators are not correlated.

GDP and tariffs: Strong positive correlation, reflecting trade cooperation and co-growth between countries, such as frequent or co-rapid growth between China and Nepal.

Tariffs and unemployment: Positive correlation, likely reflecting declining domestic economic conditions leading to higher unemployment and higher tariffs.

**From heatmaps**

Correlation between indicators: There is little difference in correlation between indicators among countries, reflecting the same trend.

GDP and inflation rate: Closely correlated, showing that GDP does not directly reflect national prosperity.

High unemployment and inflation: closely related and can occur simultaneously in poor economic conditions.

High unemployment and total GDP: The correlation is about 0.6, indicating that higher unemployment can lead to higher GDP, which is difficult to explain.

**From box-plot picture**

Differences in GDP growth rates: GDP growth rates vary significantly among countries, with China significantly higher than other countries.

Negative growth outliers: Japan, Canada and Nepal have negative growth outliers, which may be related to the economic crisis or pandemic crisis.

**Correlations**

GDP growth typically inversely correlates with unemployment

Inflation and unemployment have a short-term inverse relationship , but it's unstable long-term.

GDP and tariffs may correlate positively in countries heavily reliant on tariff revenues, though high tariffs can long-term hinder trade and GDP growth.

GDP growth can correlate with inflation during demand-pushed price increases, but high inflation can stifle further growth.

**Abnormal**

Nepal's inflation rate is much higher than other countries, which may be due to supply chain problems, political instability or improper monetary policy in the country.

Japan, Canada and Nepal experienced negative growth in a given year, which could be due to economic contractions caused by the global financial crisis, natural disasters or other unexpected events.

China's GDP growth rate continues to outpace that of other countries, which may reflect the success of its strategy of economic reform, industrialization, and export-led growth.

## 6. Presentation

* Compile the analysis and visualizations into a Jupyter Notebook.
* Prepare a presentation summarizing the findings and insights from the project

In this project, we analyzed economic data from five countries - China, Japan, Canada, Nepal and Brazil - over the past 40 years.

We find GDP growth across countries, but it varies significantly across countries, especially with China's economic rise since 2010. Unemployment rose between 1999 and 2004 and has since fallen. Inflation continues to rise and tariffs are falling, reflecting the trend of global trade integration. There is a positive correlation between GDP and tariffs, showing synergies between trade and economic growth.

These findings reveal the economic trajectories of different countries in the context of globalization and provide insights into future economic development trends.