Topic: Global technological strength and national development trends: a four-dimensional analysis of infrastructure, economy, population and market

Github link: https://github.com/jianbaii/206Final-Project.git
Name: Jianbai Wu, Jialin Wu, Meilin Wang

Finding:
This report explores the multifaceted relationship between a nation's technological foundation, economic growth, public health, and market dynamics. Drawing on data from the United States, United Kingdom, China, and Japan, we analyze indicators including data center capacity, GDP evolution, life expectancy trends, and stock performance in the technology sector. The goal is to uncover how national strength across these domains may shape broader development trajectories.

The first section examines average data center space in the United States and the United Kingdom. A bar chart reveals that the United States maintains a massive lead, with an average facility size nearing 1.4 million square feet, while the United Kingdom averages just over 120,000 square feet. This disparity reflects the dominant role the U.S. plays in global digital infrastructure, which supports its leadership in cloud computing, artificial intelligence, and high-performance computing industries. A pie chart further emphasizes this contrast: the United States accounts for 91.7% of the combined data center average space between the two countries, while the United Kingdom contributes only 8.3%. This heavy concentration of infrastructure within the U.S. underlines its status as a global technology hub and suggests greater national capacity for handling digital operations at scale.

Moving from infrastructure to economic performance, we present a line chart comparing the GDP growth trajectories of the United States and China from 1980 to projected values in 2030. The United States has experienced steady growth, reaching over $35 trillion, while China's growth, though starting from a lower base, has accelerated rapidly to approach $25 trillion. This convergence illustrates China's emergence as a major economic power and positions both countries at the center of global technological and economic competition.

The next section shifts focus to human development, using a line graph to compare life expectancy in China and Japan between 1975 and 2024. Japan has consistently led with life expectancy nearing 84 years, but China has made remarkable progress, rising from approximately 61 years to around 77. This upward trend in China suggests a correlation between economic growth and improvements in public health, though a significant gap still exists between the two nations.

Finally, we turn to the financial market to examine investor sentiment toward technology, using NVIDIA's stock closing prices over the past 100 days as a case study. The line chart shows notable fluctuations, with prices dropping from a high near $150 to just above $100. This volatility reflects broader uncertainty in the tech sector and highlights how quickly market confidence in innovation leaders can shift.

A. The goals for your project including what APIs/websites you planned to work with and what data you planned to gather

Our goal is to explore global trends by utilizing diverse datasets from the fields of technology, economics, health, and finance across major economies. We will analyze data on data centers, GDP performance, life expectancy, and stock market trends to identify patterns that reflect the economic and social development of key countries.

1. **Data Center Popularity and Capacity**

   Website: https://www.datacenters.com/

   API: https://www.datacenters.com/api/v1/locations/countries/234?

   Data: Number of data centers and their scale in the USA vs other countries; this data will help us understand the density of technological infrastructure in these regions and its potential impact on digital economies.

2. **GDP Growth**

   Website: https://www.imf.org

   API: https://www.imf.org/external/datamapper/api/v1/NGDPD

   Data: The actual and predicted GDP of the USA and China. By comparing the economic growth rates of these two countries, we can examine their respective economic trajectories and the role of their economic policies in global growth.

3. **Life Expectancy (Japan and China)**

   Website: https://www.worldbank.org/

   API: https://api.worldbank.org/v2/country/JP/indicator/SP.DYN.LE00.IN

   https://api.worldbank.org/v2/country/CHN/indicator/SP.DYN.LE00.IN

   Data: Life expectancy data for Japan and China. This will allow us to explore the health outcomes and social policies in these two populous countries, highlighting the importance of healthcare infrastructure in improving life expectancy.

4. **Stock Market Performance of Nvidia**

   Website: https://finance.yahoo.com

   API: Python library finance https://yfinance-python.org/

   Data: Closing stock prices of Nvidia. The analysis of Nvidia's stock will provide insights into the technology sector's performance and how market conditions are influenced by the development of artificial intelligence and advanced hardware.

B. The goals that were achieved including what APIs/websites you actually worked with and what data you did gather

We achieved our goal of exploring global trends by gathering and analyzing diverse datasets across technology, economics, health, and finance to identify patterns in the economic and social development of key countries.

1. Data Center Popularity and Size (USA and UK):
   APIs accessed: https://www.datacenters.com/api/v1/locations/countries/234?
   https://www.datacenters.com/api/v1/locations/countries/77?

   Websites accessed: https://www.datacenters.com/locations/united-states
   https://www.datacenters.com/locations/united-kingdom
   Data gathered:
   - Top 50 most popular data centers in the USA and UK, including their spatial size and other relevant metrics.
   - 60 data points were collected for the USA, and 40 were collected for the UK.

2. GDP Data for USA and China (1980–2029)
   APIs accessed: https://www.imf.org/external/datamapper/api/v1/NGDPD
   Data gathered:
   - Actual and predicted GDP for the USA and China from 1980 to 2029.
   - Data was collected in batches of 20 entries, repeated 5 times.

3. Life Expectancy Data for Japan and China (1975–2024)
   APIs accessed:
   https://api.worldbank.org/v2/country/JP/indicator/SP.DYN.LE00.IN?format=json
   https://api.worldbank.org/v2/country/CHN/indicator/SP.DYN.LE00.IN?format=json
   Data gathered:
   - Life expectancy data for Japan and China from 1975 to 2024.
   - Data was collected in batches of 20 entries, repeated 5 times.

4. Stock Market Data for Nvidia
   APIs accessed: Python library finance https://yfinance-python.org/
   Data gathered:
   - Closing stock prices of Nvidia for the last 100 trading days, collected in batches of 25 entries, repeated 4 times.

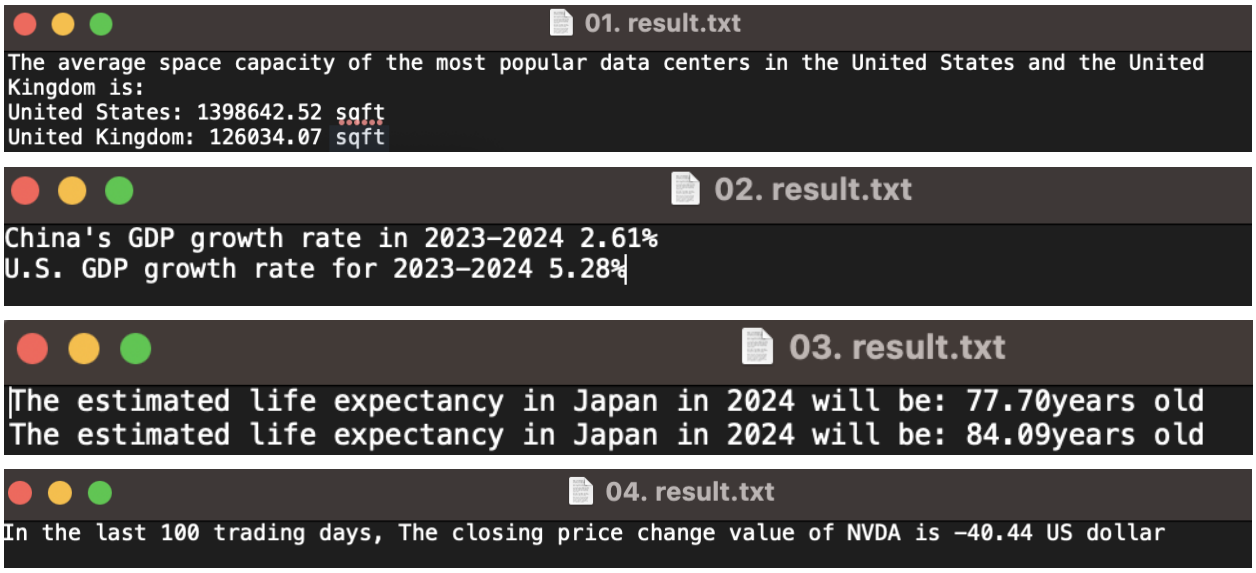
C. The problems that you faced

- It was hard to find a usable API and come up with a good topic for the project.
- When calculating the GDP growth rate for China and the U.S. over the past two years, we encountered errors. After debugging, we found that the issue was due to the year

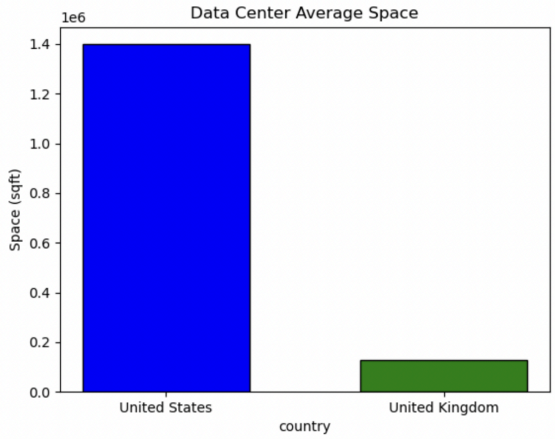
values being saved as integers but being interpreted as strings during the calculation, which caused a format mismatch and led to the failure of the conditional statement.

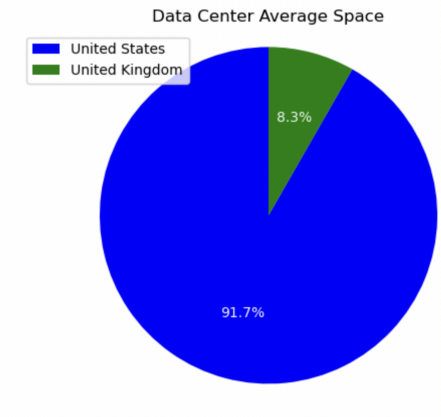
- When plotting the line chart, the date labels on the x-axis were too crowded to read. We solved this by customizing the x-axis ticks—using a list of selected index values and corresponding labels—so that the labels are shown every 5 values. This approach made the x-axis much clearer.
- Regarding NVIDIA stock data, we found that the API had time limits for usage.

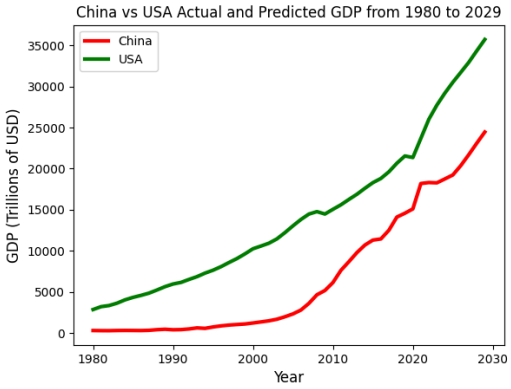
D. The calculations from the data in the database (i.e. a screenshot)

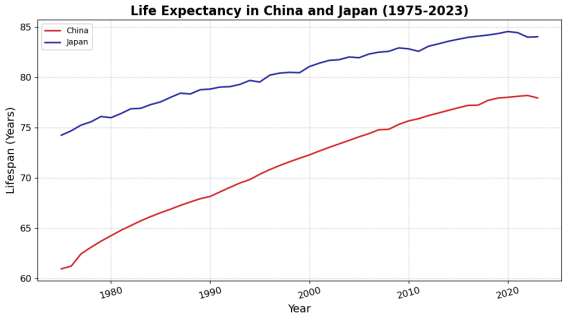


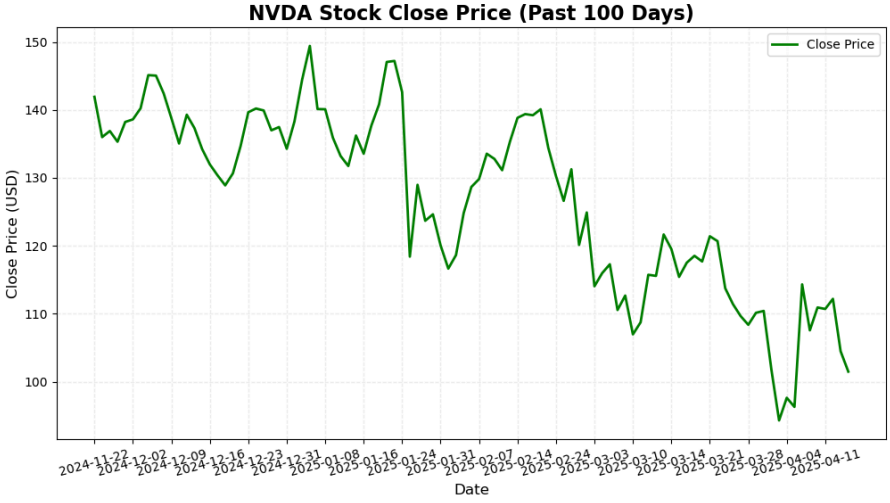
E. The visualization that you created (i.e. screenshot or image file)











F. Instructions for running your code

1. Running 01.DataCenter_crawl.py can collect the names and spaces of countries and corresponding data centers, and save them in the data_center_countries and data_centers tables in the database.
   Running 01.DataCenter_analyse.py can read the data in the data_centers table for calculating results and generating visual graphics.
2. Running 02.GDP_crawl.py can collect the data of countries and corresponding GDP, and save them in the gdp table in the database.
   Running 02.GDP_analyse.py can read the data in the gdp table for calculating results and generating visual graphics.
3. Running 03.LifeSpan_crawl.py can collect the life expectancy data corresponding to all countries, filter out the data of China and Japan, and save them in the lifespan table in the database.
   Running 03.LifeSpan_analyse.py can read the data in the lifespan table for calculating results and generating visual graphics.
4. Run 04.Stock_crawl.py to collect the data of all dates and closing prices of the specified stock within the specified date range, and save them in the stock table in the database.

Run 04.Stock_analyse.py to read the data in the stock table for calculation results and generation of visualization graphics.

G. Documentation for each function that you wrote. This includes describing the input and output for each function

- Data center API
    1. def get_data (url, country_id):
       Parameter:
       url (str): API address
       country_id (int): Country ID
       Return: Center Name and Total Space for USA and UK (the collected data will be saved in the database, and it will be saved once every 20 data collected)
       Output: Data collection log
    2. def analyse_data():
       Parameters: None
       Return: The bar chart and pie chart for Data Center average space for UK and USA will be displayed.
       Output: Calculation results of the average space capacity of the most popular data centers in the United States and the United Kingdom

- GDP API
    1. def get_gdp_data(api_url, country_code):
       Parameter:
       api_url (str): API address
       country_code (str): Country identifier string (CHN/USA)
       Return: Get China and America predict GDP from 1980 to 2029 (the collected data will be saved in the database, once every 20 data collected)
       Output: Data collection log
    2. def analyse_data():
       Parameters: None
       Return: Line chart of projected GDP of the United States and China from 1980 to 2029 will be displayed
       Output: Calculation results of GDP growth rates of the United States and China from 2023 to 2024

- LifeSpan API
    1. def get_life_data(country_code):
       Parameters:
       country_code (str): country identification string
       Return: Life expectancy in China and Japan from 1975 to 2023 (the collected data will be saved in the database, once every 20 data collected)
       Output: Data collection log
    2. def analyse_data():

Parameters: None
Return: Line chart of life expectancy in Japan and China from 1975 to 2023 will be displayed
Output: The estimated life expectancy of Japan and China in 2024

- Stock API
    1. def get_stock_data(stock, start_date, end_date):
       Parameters:
       stock (str): Stock Name identification string
       start_date: start date
       end_date: end date
       Return: Nvidia stock closing price from November 22, 2024 to April 17, 2025 (the collected data will be saved in the database, and it will be saved once every 25 data collected)
       Output: Data collection log
    2. def analyse_data():
       Parameters: None
       Return: Line chart of Nvidia stock closing prices over the past 100 trading days will be displayed
       Output: the closing price change value of the specified stock in the last 100 trading days

H. You must also clearly document all resources you used. The documentation should be of the following form

Date	Issue Description	Location of Resource	Result (did it solve the issue?)
4/15/2025	How to find public APIs	https://www.reddit.com/r/learnprogramming/comments/u5nbek/are_there_any_websites_that_list_public_apis/?rdt=46766 https://github.com/public-apis/public-apis https://rapidapi.com/blog/how-to-find-apis/	Yes
4/17/2025	Access topic-related APIs	https://www.datacenters.com/api/v1/locations/countries/234?	Yes

		https://www.datacenters.com/api/v1/locations/countries/77 https://www.imf.org/external/datamapper/api/v1/NGDPD https://api.worldbank.org/v2/country/JP/indicator/SP.DYN.LE00.IN https://api.worldbank.org/v2/country/CHN/indicator/SP.DYN.LE00.IN https://yfinance-python.org/	
4/17/2025	Need Official API Documentation	https://www.imf.org/external/datamapper/api/help	Yes
4/18/2025	There is no obvious entry for the web API	https://aatt.io/newsletters/how-to-find-and-use-hidden-apis-to-automate-processes	Yes
4/20/2025	Don't know how to use the packaged yfinance	https://piaohua.github.io/post/python/20231231-yfinance/	Yes
4/22/2025	Need to learn matplotlib	https://matplotlib.org/	Yes