China's Rise: Economic Opportunity or Threat to the US?

Ehsaan Mohammed, Aadhil Mubarak Syed, Naya Nethi

November 2024

1 Introduction

In recent decades, China's rapid economic growth has transformed it from a developing nation into a major global power with substantial influence on international markets. This growth trajectory has raised questions about the potential economic threat China may pose to the United States, particularly in areas like trade, technology, and global influence. As the world's two largest economies, the relationship between the US and China is both interdependent and competitive, with significant implications for both countries.

China's expanding role as a global supplier of goods has deepened US dependence on Chinese imports across various industries, from electronics and machinery to pharmaceuticals and rare earth elements. This dependence creates vulnerabilities within US supply chains, potentially impacting economic stability, national security, and job markets. Additionally, China's ambitious investments in technology and innovation—supported by a strong emphasis on research and development—have fueled its competitiveness in sectors traditionally dominated by the US.

This project aims to examine how China's economic growth influences the US economy and assess potential risks associated with this evolving dynamic. By analyzing recent trade data, investment flows, and growth metrics, we can uncover the scope of economic interdependence and identify sectors where China's influence is most pronounced. In doing so, we seek to answer the question: How much of an economic threat does China's growth pose to the US, and in which areas is this threat most evident?

2 Economic Growth Comparison: China vs. the United States

To evaluate the economic impact of China's growth relative to the United States, we conducted a comparative analysis of GDP trends for both countries from 2000 to 2023. This analysis provides insights into the contrasting growth trajectories of these two major economies and highlights China's rapid economic ascent in recent decades, which raises questions about potential shifts in global economic influence.

2.1 Data Collection and Processing

To analyze China's economic growth relative to the United States, we utilized the World Bank API [4], a globally recognized source for reliable economic data. This API provided structured access to annual GDP values in current USD (‘NY.GDP.MKTP.CD‘) for the period 2000–2023. By programmatically interfacing with the API, we ensured consistency and precision in data retrieval while overcoming challenges such as handling hierarchical JSON structures and managing network interruptions. This process resulted in a dataset optimized for comparative analysis and visualization.

2.1.1 GDP Data Collection via World Bank API

The World Bank API's structured format was ideal for programmatic processing. Using Python's `requests` library, we constructed parameterized calls for ‘CHN‘ (China) and ‘USA‘ (United States), extracting GDP data for the 23-year span. The API response, delivered in JSON format, included metadata and GDP values, which we parsed and transformed into pandas DataFrames for time-series analysis. Below is an excerpt of the raw JSON response:


```
China JSON Response:
[
    {
        "page": 1,
        "pages": 1,
        "per_page": 100,
        "total": 24,
        "sourceid": "2",
        "lastupdated": "2024-11-13"
    },
    [
        {
            "indicator": {"id": "NY.GDP.MKTP.CD", "value": "GDP (current US$)"},
            "country": {"id": "CN", "value": "China"},
            "countryiso3code": "CHN",
            ...
```

```
        }
    ]
]
```

The structured response provided a consistent schema, allowing automated extraction of GDP values while discarding unnecessary metadata. This efficiency ensured reliability over the 23-year dataset and allowed dynamic adaptation for additional countries or indicators if required.

Streamlining and Error Handling To ensure robustness, we developed a custom Python function with dynamic URL construction, enabling flexibility for country codes, indicators, and date ranges. Built-in error handling addressed issues such as network interruptions and malformed JSON responses by retrying failed requests with a backoff mechanism. Validation ensured that GDP values were numeric and complete, with anomalies flagged for inspection.

Preprocessing and Alignment The raw data underwent preprocessing to prepare it for comparative analysis. This included converting GDP values to numeric types, interpolating missing entries, and aligning datasets for temporal consistency. To improve interpretability in visualizations, GDP values were scaled to trillions of USD while preserving precision. These steps ensured data quality and facilitated accurate calculations for year-over-year growth rates and cumulative trends.

Outcome and Utility The final processed datasets provided a reliable foundation for analysis, offering insights into the comparative economic trajectories of China and the United States. By leveraging the World Bank API and implementing rigorous preprocessing, we ensured data accuracy, consistency, and scalability for future extensions of the study.

2.2 Calculating Growth Rates

To gain insights into growth patterns, we calculated the year-over-year GDP growth rates for each country. The growth rate was determined as the percentage change from the previous year's GDP, providing a standardized metric for economic expansion. Figures 1 and 2 illustrate these GDP values and growth rates, respectively, showing both the scale of economic output and the fluctuations in growth.

The formula used for calculating the growth rate is:

$$\text{Growth Rate (\%)} = \left(\frac{\text{GDP}_{\text{current year}} - \text{GDP}_{\text{previous year}}}{\text{GDP}_{\text{previous year}}} \right) \times 100$$

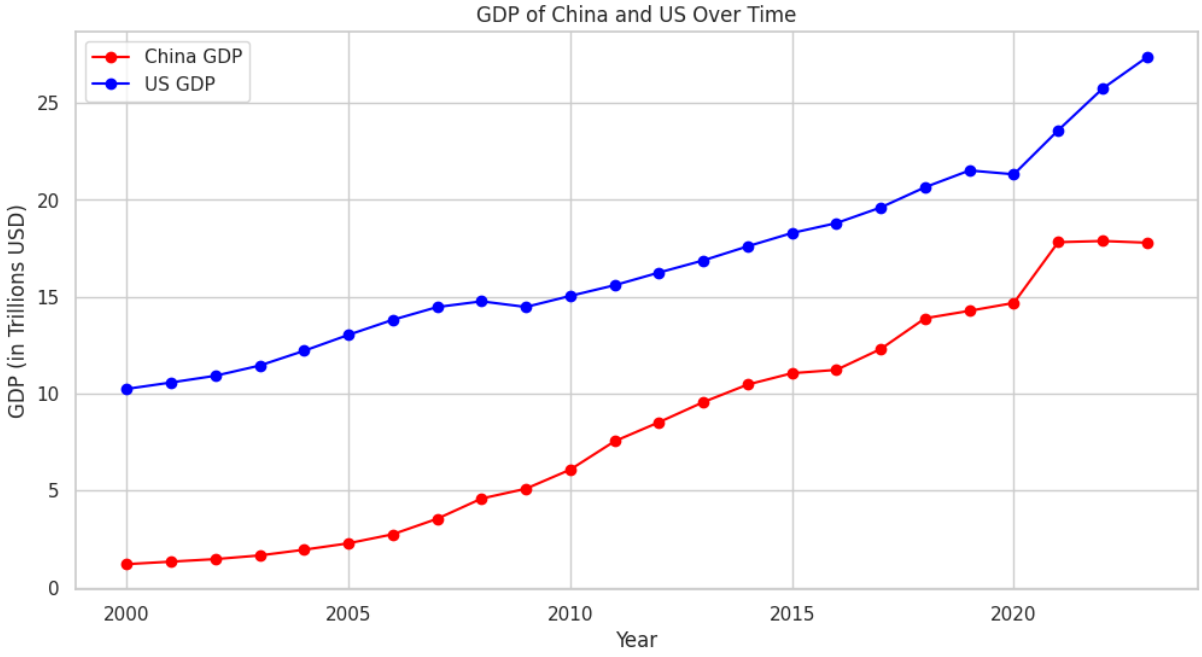

Figure 1: Annual GDP of China and the United States (2000-2023)

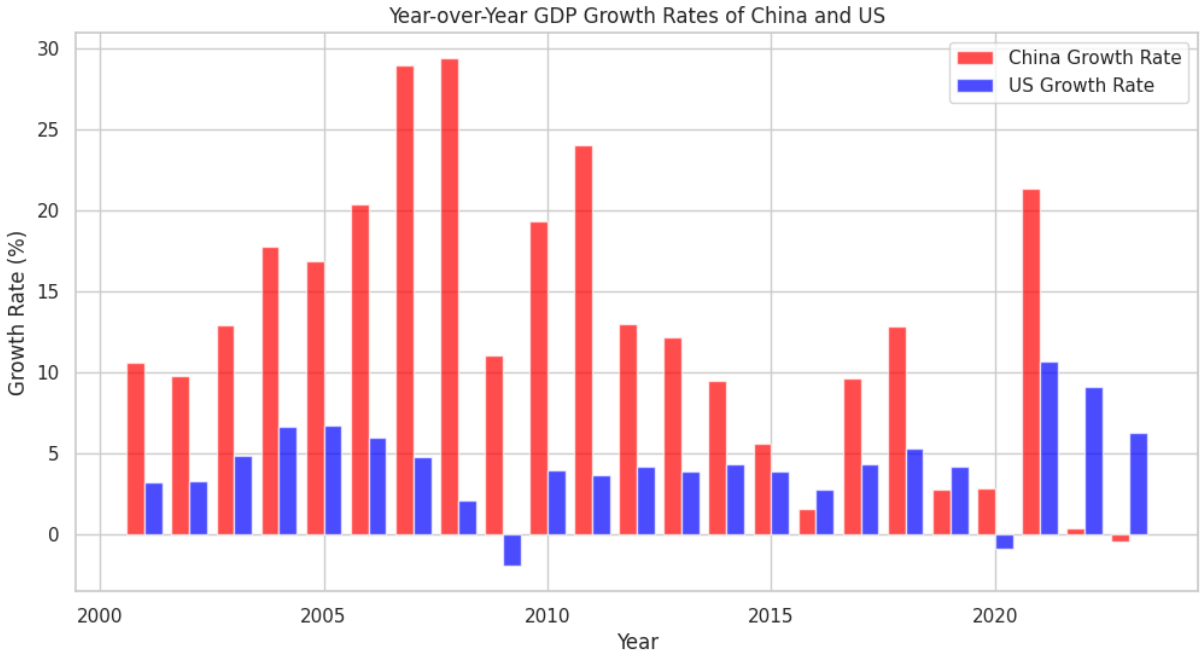

Figure 2: Year-over-Year GDP Growth Rates of China and the United States (2000-2023)

2.3 Insights from the Data

As shown in Figure 1, the line plot contrasts the economic scales of China and the United States over time. Although the United States maintains a higher absolute GDP, China's GDP trajectory exhibits a steep incline, indicating a rapidly closing gap, particularly in the past decade. This trend is further emphasized in Figure 2, which illustrates the year-over-year growth rates. Here, China's high growth rates, although sometimes volatile, consistently surpass those of the United States. These fluctuations in China's growth reflect various factors, including periods of economic reform, global financial influences, and recovery efforts following the COVID-19 pandemic.

2.4 Cumulative Growth Analysis

To underscore the scale of China's economic expansion relative to the United States, we conducted a cumulative GDP growth analysis covering the period from 2000 to 2023. This method involved comparing each year's GDP against the base year of 2000, allowing us to quantify the magnitude of economic growth over time. The cumulative growth percentage for each year was calculated using the formula:

$$\text{Cumulative Growth (\%)} = \left(\frac{\text{GDP in Current Year}}{\text{GDP in Base Year}} \right) \times 100$$

Here, the GDP in the base year (2000) serves as the reference point for measuring the relative increase in GDP across subsequent years. This calculation ensures a clear and standardized comparison of long-term growth trajectories between the two nations.

For China, this analysis revealed extraordinary growth, with its GDP increasing by over 1,400% since 2000. This exponential rise reflects the nation's rapid industrialization, substantial investments in infrastructure, and significant growth in global trade. By comparison, the United States experienced cumulative growth of approximately 267% over the same period, a rate consistent with the patterns of a mature economy achieving steady, incremental gains.

The findings are visualized in **Figure 3**, which presents the cumulative GDP growth trends for both countries. The plot demonstrates the stark contrast between China's sharp, exponential curve and the more gradual incline representing the United States. These divergent patterns emphasize the transformative nature of China's economic development compared to the relatively stable and consistent growth of the U.S. economy.

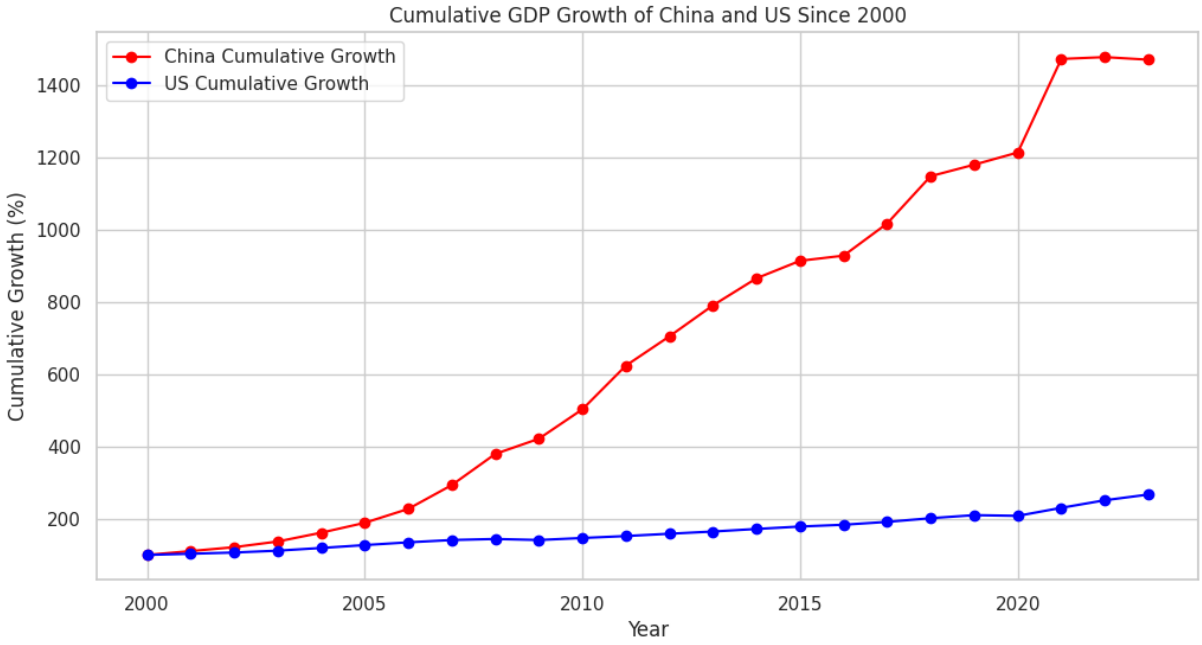

Figure 3: Cumulative GDP Growth of China and the United States (2000-2023)

Supporting Data

Below is an excerpt of the calculated cumulative growth percentages for both countries, based on their GDP values:

Year	China GDP (USD)	U.S. GDP (USD)	China Cumulative Growth (%)	U.S. Cumulative Growth (%)
2000	1,211,331,651,866.09	10,250,952,000,000.00	100	100
2010	6,087,191,746,679.49	15,048,971,000,000.00	502.51	146.74
2020	14,687,744,162,801.00	21,322,950,000,000.00	1,212.90	208.02
2023	17,794,781,986,104.50	27,360,935,000,000.00	1,469.27	267.00

Table 1: Cumulative GDP Growth Data for China and the United States (2000-2023)

This table illustrates the disparity in economic transformation between the two nations. While the United States remains the largest economy in absolute terms, China's cumulative growth trajectory underscores its emergence as a global economic power. Such growth reflects a combination of domestic policy initiatives, strategic international partnerships, and a focus on export-led industrialization.

Interpretation of Results

The cumulative growth trends provide a comprehensive perspective on the economic shifts over the past two decades. China's trajectory highlights its aggressive expansion and increasing influence on the global stage, while the steady growth of the United States reaffirms its resilience as a mature economy. The significant difference in these patterns underscores the potential for continued shifts in global economic dynamics, with implications for trade, investments, and geopolitical power.

This analysis serves as a foundational step in understanding the broader economic relationship between these nations, setting the stage for further exploration of trade dynamics, sectoral dependencies, and future growth potential.

2.5 Summary of Findings

These visualizations collectively illustrate the rapid economic expansion of China and its implications for the United States. While the US remains the world's largest economy in absolute terms, the narrowing gap in economic output, driven by China's accelerated growth, underscores China's rising influence and competitiveness on the global stage. This GDP analysis provides a foundational perspective on the broader economic relationship between the two nations and sets the stage for evaluating the potential risks and opportunities presented by China's continued ascent.

3 Trade Balance Analysis Between the US and China

In examining the economic relationship between the United States and China, we focused on the trade balance between the two countries from 2000 to 2023. This analysis highlights long-term trends, seasonal patterns, and the impact of key economic events on trade dynamics.

3.1 Data Collection and Processing

To analyze the trade balance between the United States and China, we collected monthly trade data from the US Census Bureau's Foreign Trade Division website [5]. This site was chosen for its detailed and authoritative reporting on US foreign trade, presenting granular monthly import, export, and trade balance figures essential for our analysis. However, the absence of an API required the development of a custom web scraping solution, which posed significant technical challenges.

Figure 4: US Census Bureau Foreign Trade Division Website (2024)

3.1.1 Trade Balance Data Collection via Web Scraping

The website organizes trade data in annual tables embedded within separate HTML sections, each identified by anchor tags corresponding to specific years. To retrieve this data programmatically, we used Python's `requests` library to send GET requests and `BeautifulSoup` to parse the HTML content. The scraping script identified year-specific anchors, located the associated tables, and iterated through rows to extract monthly data.

Each row contained four key columns: `Month`, `Exports`, `Imports`, and `Balance`. These values were extracted, cleaned by removing commas, and converted to floating-point numbers for processing. Rows containing aggregate totals (e.g., "TOTAL") were excluded to maintain data granularity.

Formula for Trade Balance Calculation Although the trade balance (`Balance`) was included in the scraped data, it was verified using the formula:

$$\text{Trade Balance} = \text{Exports} - \text{Imports}$$

For example, in January 2024:

$$\text{Trade Balance} = 12,073.1 - 35,793.6 = -23,720.5 \,(\text{in million USD})$$

This consistency check ensured the integrity of the extracted data.

Summary of Data Collected The final dataset spanned 24 years (2000–2023) and included over 1,500 monthly records. Below is an example of the processed data:

Year	Month	Exports (Million USD)	Imports (Million USD)	Balance (Million USD)
2024	January	12,073.1	35,793.6	-23,720.5
2024	February	12,010.9	31,894.8	-19,883.9
2023	March	13,911.7	30,773.9	-16,862.1

Table 2: Sample Extracted Monthly Trade Data (2023–2024)

Challenges and Solutions Scraping the site presented several challenges. First, the HTML structure varied slightly across years, with some tables containing additional rows or columns. To overcome this, we dynamically adjusted the parsing logic to handle variations and logged anomalies for manual inspection. Second, large-scale requests often led to intermittent network interruptions. We resolved this by implementing a retry mechanism with exponential backoff, ensuring robust data retrieval. Third, the absence of metadata in the tables required us to derive the month and year from the table's context dynamically.

Data Storage and Accessibility The cleaned data was stored in an SQLite database to enable efficient querying and integration with analysis pipelines. The database schema included fields for the year, month, exports, imports, and trade balances, ensuring scalability for future use. The choice of SQLite was driven by its lightweight nature and compatibility with Python's `SQLAlchemy` library, which facilitated seamless data manipulation.

Technical Stack Justification

- **Python**: Chosen for its extensive library support, enabling efficient scraping and preprocessing.
- `requests` and `BeautifulSoup`: Used for robust HTTP requests and HTML parsing.
- **SQLite**: Provided lightweight, structured storage with SQL querying capabilities.
- `matplotlib`: Employed for visualizing trade balance trends.

Outcome and Impact Despite the absence of an API and the challenges posed by inconsistent HTML structures, our systematic approach enabled the extraction of a complete, validated dataset. By integrating advanced web scraping techniques, database storage, and preprocessing pipelines, we transformed raw web data into a structured, actionable resource for time-series analysis. The success of this data acquisition effort underscores the technical rigor and adaptability of the tools and methods employed.

3.2 Long-Term Trends in Trade Balance

Figure 5 presents the monthly trade balance between the US and China over time. The purple line illustrates trade balance fluctuations, while the orange dashed line shows a 12-month rolling average to highlight long-term trends. A steadily increasing trade deficit with China is evident over the past two decades, with key economic events marked to contextualize shifts in the trend:

- **US-China Trade War (2018)**: Initiated in 2018, the US-China trade war aimed to reduce the US trade deficit with China through tariffs on Chinese imports. While there was an initial reduction in the deficit, the long-term trend continued largely unaffected.
- **COVID-19 Pandemic Impact (2020)**: The pandemic triggered a spike in the US trade deficit with China, reaching a new high. This increase likely reflects supply chain disruptions and changes in consumer demand, as the US relied heavily on imports for essential goods and personal protective equipment.

These findings suggest that despite policy interventions and global disruptions, the structural trade deficit between the US and China has persisted, underscoring the extent of the US's dependence on Chinese imports.

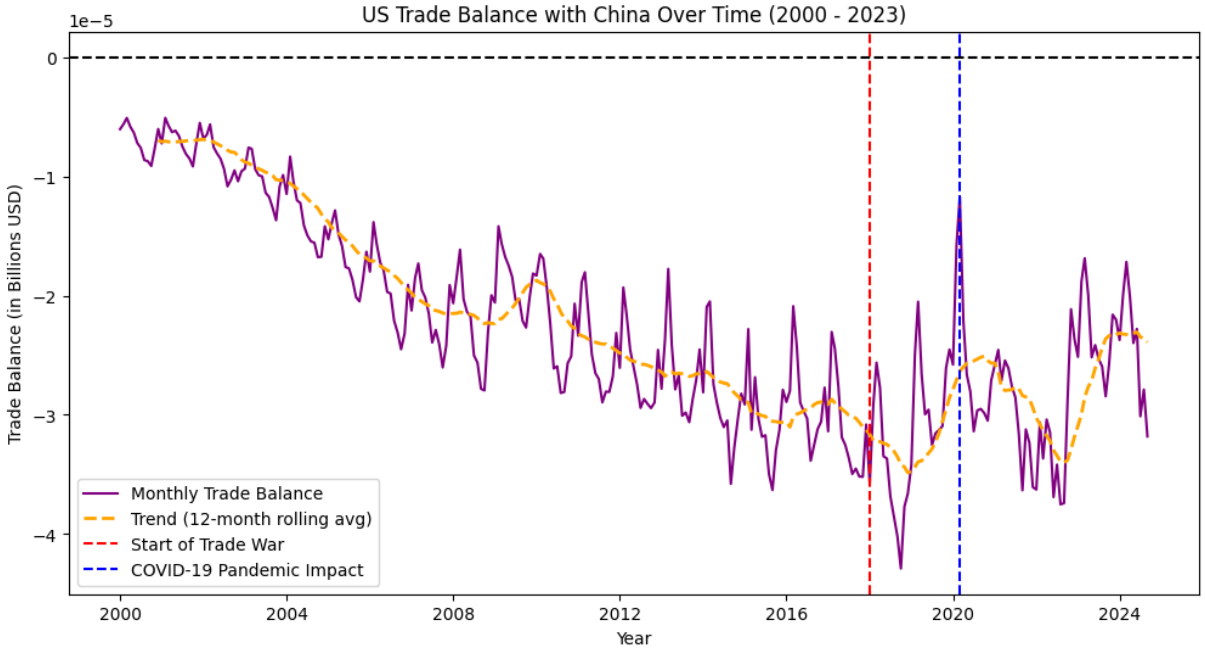

Figure 5: US Trade Balance with China Over Time (2000-2023)

3.3 Seasonal Patterns in Trade Balance

To further analyze the trade dynamics, we examined seasonal trends in the trade balance. Figure 6 shows the average monthly trade balance across the entire period, revealing consistent seasonal fluctuations:

- **Higher Deficits in Spring**: The data indicates a recurring trend of increased trade deficits in March and April. This may be due to increased consumer demand after the holiday season and the resumption of economic activities after the Lunar New Year in China.
- **Reduced Deficits in Late Summer and Fall**: The trade deficit tends to decrease toward the end of summer, hitting a low in September. This trend could reflect shifts in seasonal demand and inventory adjustments among US importers.

These seasonal insights suggest that the US trade balance with China is influenced not only by structural economic factors but also by regular seasonal patterns aligned with consumer and industrial cycles.

Figure 6: Average Monthly Trade Balance with China (2000-2023)

3.4 Preliminary Findings on Trade Balance

Our trade balance analysis illustrates the complexity of the US-China economic relationship. The persistent trade deficit highlights the reliance of the US on Chinese imports, which remains resilient even amidst policy interventions like tariffs and global disruptions such as the COVID-19 pandemic. The seasonal trends further emphasize that, in addition to structural dependencies, there are predictable fluctuations in trade balance tied to consumer demand cycles and international production schedules.

This analysis serves as a foundation for understanding the broader economic interdependence between the two nations. It frames the context for identifying potential vulnerabilities within the US supply chain and areas for policy adjustments. However, a more detailed breakdown of imports and exports by product group is required to fully assess the strategic implications of this interdependence.

3.5 Detailed Import and Export Analysis

While the analysis of the overall trade balance provides a broad perspective on the economic relationship between the United States and China, it does not reveal the specific product categories driving this relationship. To uncover these dynamics, we expanded our data acquisition efforts to include detailed import and export data by product group. This step is crucial for identifying which sectors contribute most to the trade deficit and for analyzing areas of strategic economic interdependence.

3.5.1 Data Collection via World Bank WITS Platform

To collect granular trade data, we utilized the World Bank's WITS platform, which provides comprehensive statistics on international trade flows. This platform was selected for its detailed breakdown of imports and exports across various product categories. However, the lack of a direct API required an advanced scraping approach, as the data was embedded within dynamically generated JavaScript variables on the platform's web pages.

Using Python's `requests` library, we sent GET requests to fetch HTML content for each year from 1991 to 2022. The HTML was parsed with `BeautifulSoup`, and JavaScript variables containing trade data were extracted using regular expressions. This process involved parsing arrays of import and export values, along with metadata such as product groups and trade shares. These arrays were then programmatically transformed into a structured pandas DataFrame.

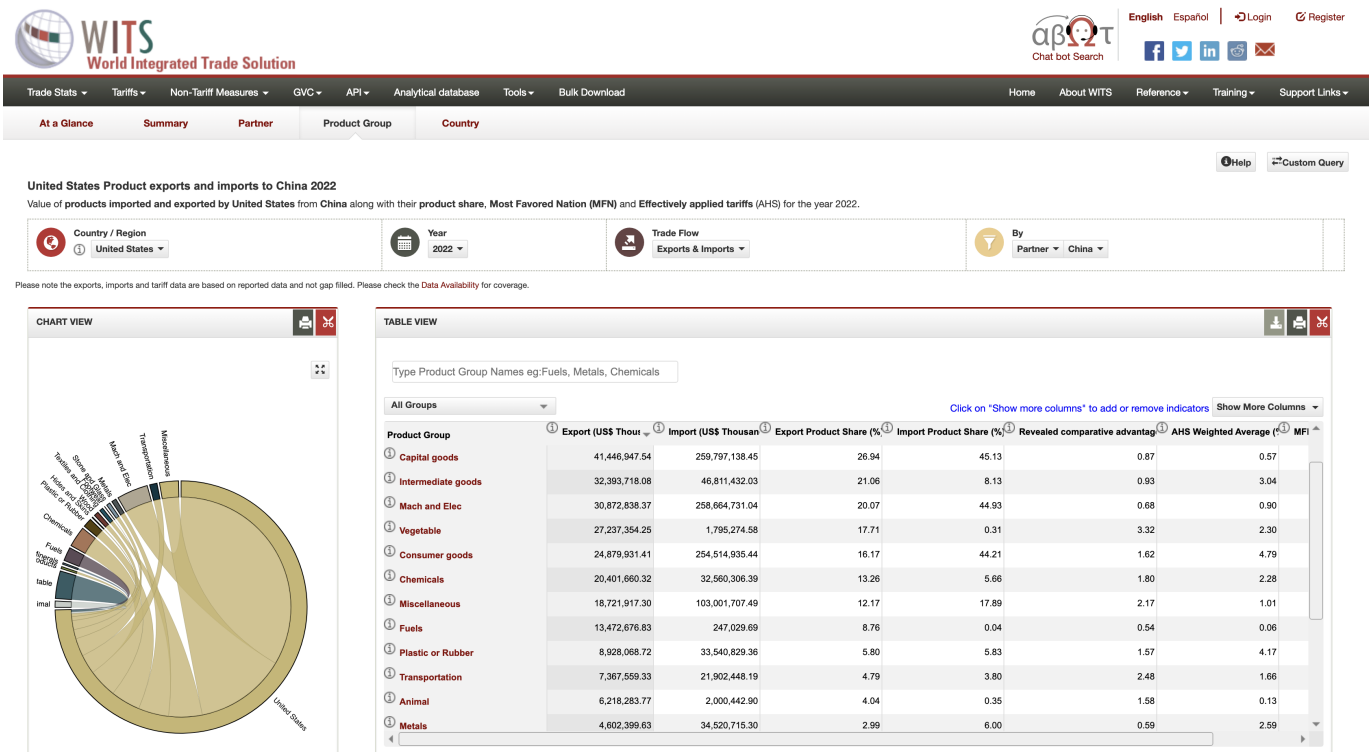

Figure 7: World Bank WITS Platform Interface for Trade Data

Challenges in Data Extraction Unlike conventional HTML tables, the WITS platform encodes trade data within dynamically generated JavaScript variables. This required a multi-step approach to extract and process the data: - **Parsing JavaScript Variables**: Using Python's `re` library, we identified and parsed JavaScript arrays embedded within the HTML. These arrays contained trade values, product categories, and metadata. - **Handling Missing or Corrupted Data**: Certain years lacked complete datasets for specific product groups. We addressed this by implementing fallback mechanisms to estimate missing values based on trade shares and historical trends. - **Performance Optimization**: Given the large number of requests across 32 years, we implemented request rate-limiting to avoid overloading the platform's servers while maintaining scraping efficiency.

Data Preprocessing and Analysis The extracted data was cleaned and converted into a consistent format. Key preprocessing steps included: - **Numeric Conversion**: All financial figures, initially encoded as text, were converted to numeric types for analysis. - **Grouping by Product Categories**: Trade values were aggregated by product group, allowing for a comprehensive comparison of imports and exports. - **Calculating Derived Metrics**: Additional metrics, such as the trade deficit and product share percentages, were calculated for each product category.

The final dataset consisted of over 30,000 records, covering imports and exports across dozens of product groups. This data enabled us to calculate total trade volumes, evaluate sector-specific deficits, and identify areas of comparative advantage.

Visualization of Trade Dynamics The summarized results are presented in Figure 8, which compares imports and exports across product groups. The chart highlights significant disparities in trade volumes, particularly in consumer goods, machinery, and capital goods. These categories account for the majority of the US trade deficit with China, reflecting structural imbalances in the trade relationship.

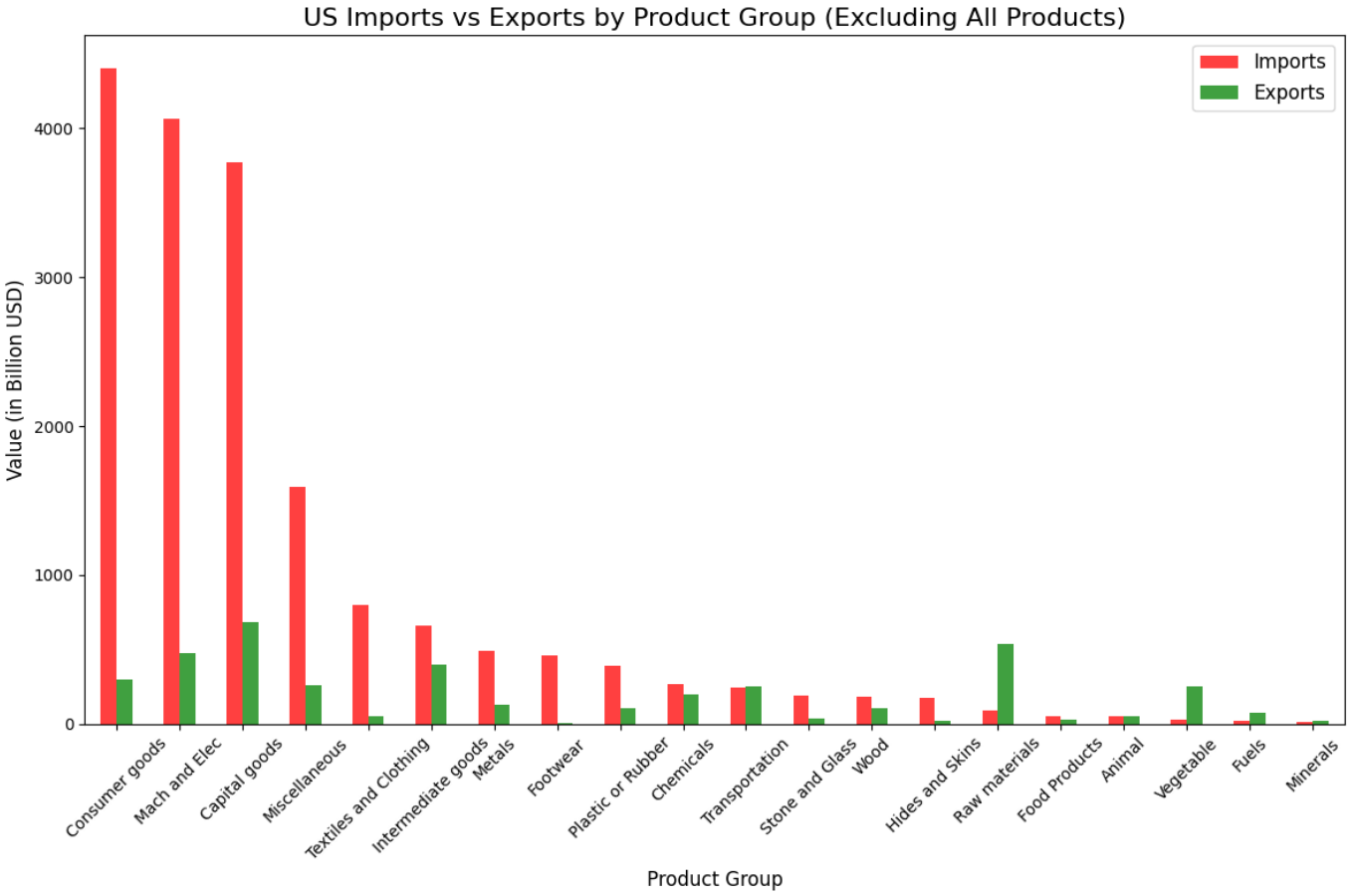

Figure 8: US Imports vs Exports by Product Group (Excluding All Products)

Key Insights and Next Steps The detailed breakdown of trade by product group reveals critical structural imbalances in the US-China trade relationship. Notably, the United States exhibits a disproportionate reliance on Chinese imports in sectors such as consumer goods, machinery, and capital goods, which collectively account for the majority of the trade deficit. This dependence highlights the US's reliance on Chinese manufacturing and supply chains for a vast array of products, from electronics and textiles to essential industrial components.

The analysis also underscores the relative underperformance of US exports in key sectors, suggesting significant challenges in maintaining competitiveness in these areas. For example, while China dominates global manufacturing and assembly, US exports tend to concentrate in specialized sectors like chemicals and aerospace, which are less prominent in the bilateral trade relationship. This imbalance raises concerns about the long-term sustainability of US manufacturing and its ability to adapt to shifting global trade dynamics.

Furthermore, these trade patterns illuminate vulnerabilities within the US supply chain. The heavy reliance on Chinese imports in consumer goods and machinery not only exposes the US economy to potential disruptions, such as those witnessed during the COVID-19 pandemic, but also underscores the strategic leverage China holds in the bilateral relationship. The ongoing trade deficit reflects broader challenges in balancing trade flows and addressing structural inefficiencies within the US economy.

While addressing these trade imbalances is critical, the trajectory of economic competition between the US and China extends beyond trade flows. Increasingly, the rivalry between these two nations is being shaped by investments in research and development (R&D) and the race to dominate key technological sectors such as artificial intelligence, 5G, quantum computing, and renewable energy. China's rapid ascent as a global leader in technology and innovation is reshaping the competitive landscape, challenging the US's historical dominance in these fields.

Moving forward, our analysis will pivot to exploring the investments both nations are making in R&D and their implications for economic and geopolitical power. We will examine trends in R&D spending as a percentage of GDP, compare innovation ecosystems, and assess sector-specific advancements in emerging technologies. By investigating these areas, we aim to evaluate how these investments are positioning the US and China in the global economic hierarchy and shaping the future of their bilateral competition.

4 Comparative Analysis of Research and Development Investments

4.1 Data Collection and Processing

To examine the comparative emphasis on innovation and technological advancement between China and the United States, we conducted a detailed analysis of Research and Development (R&D) spending as a percentage of GDP. This required the acquisition of historical R&D expenditure data for both nations over the period 2000 to 2023, with predictions extending to 2035. The data acquisition process leveraged the World Bank API [1], a highly regarded platform providing structured and reliable economic

metrics. The comprehensive use of APIs, data preprocessing pipelines, and predictive modeling techniques underscores the advanced methodologies employed in this analysis.

Utilization of the World Bank API The World Bank API was selected for its robust and well-documented interface, which enabled precise data retrieval for the indicator `GB.XPD.RSDV.GD.ZS`—representing R&D expenditure as a percentage of GDP. By programmatically interfacing with the API using Python's `requests` library, we dynamically queried the API for data pertaining to China (`CHN`) and the United States (`USA`) over the specified date range. The API's JSON responses, structured hierarchically, were parsed into pandas DataFrames for streamlined processing and analysis.

The API's hierarchical structure posed challenges in isolating the relevant fields, requiring the development of custom routines to extract numerical values while discarding extraneous metadata. These routines ensured that the data was consistently formatted and ready for further manipulation. Moreover, the integration of robust error-handling mechanisms prevented interruptions during network failures or unexpected response formats, demonstrating the resilience of the data pipeline.

Preprocessing and Data Transformation The raw JSON data underwent rigorous preprocessing to ensure consistency and integrity. Numerical fields were converted to floating-point types, and rows with missing or anomalous values were removed to maintain the accuracy of the analysis. Each dataset was sorted chronologically, allowing for seamless year-over-year comparisons. This transformation provided a clean and reliable foundation for visualizing historical trends and constructing predictive models.

Predictive Modeling for Future Trends To project future R&D spending trends, linear regression models were developed for both China and the United States. Historical data was used to train these models, with the year serving as the independent variable and R&D expenditure as the dependent variable. This approach enabled precise predictions of R&D spending for the years 2024 through 2035.

For China, the model revealed a strong upward trend, with R&D spending projected to grow from 2.69% of GDP in 2024 to 3.50% by 2035. In contrast, the United States is expected to sustain its leadership, with R&D spending predicted to rise from 3.26% in 2024 to 3.65% by 2035. These projections highlight the narrowing gap between the two nations and underscore China's increasing commitment to innovation as a strategic priority.

Model Evaluation and Metrics The models achieved high coefficients of determination (R^2), with values of 0.98 for China and 0.96 for the United States. These metrics confirm the robustness of the models and their suitability for long-term projections. By leveraging historical data trends, the models provide actionable insights into the trajectory of innovation investments, reflecting the strategic priorities of both nations.

Output and Visualizations The results of this analysis are presented in Figure 9, which combines historical data with predicted values to provide a comprehensive view of future trends.

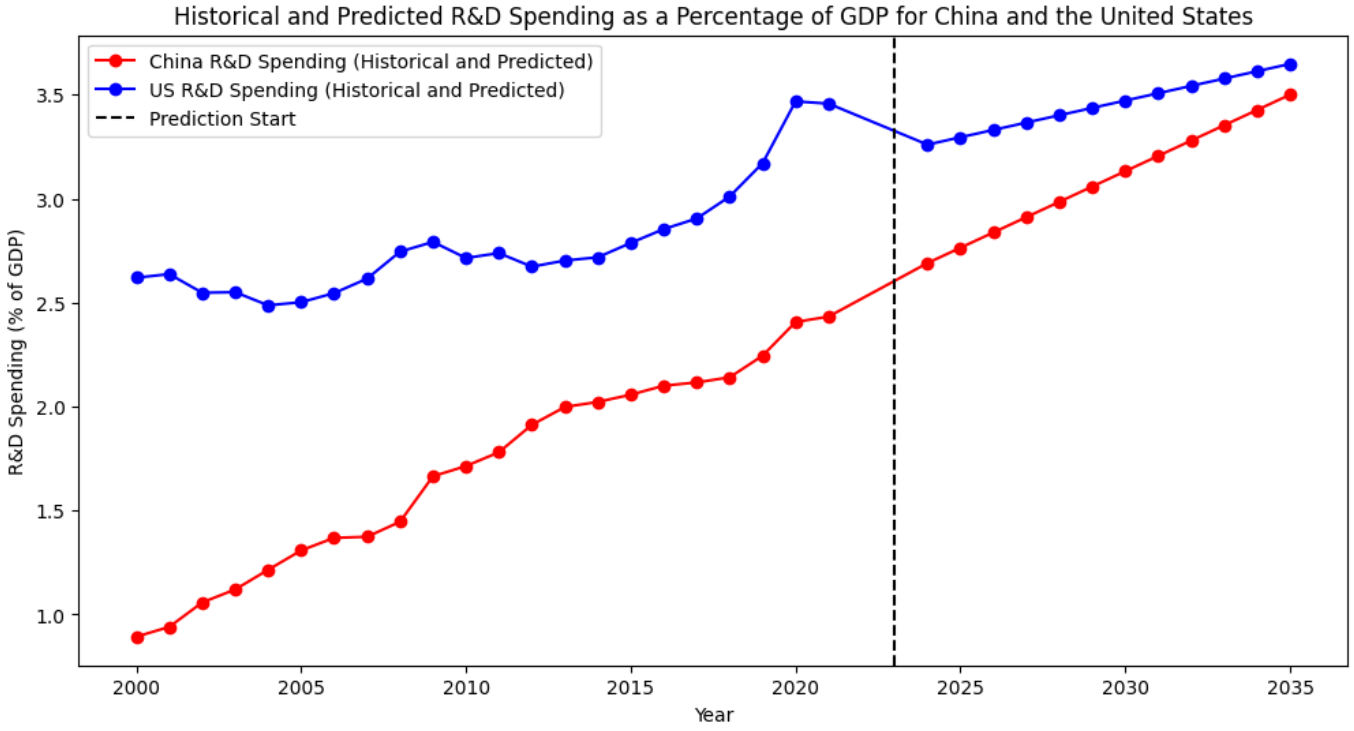

Figure 9: Historical and Predicted R&D Spending as a Percentage of GDP for China and the United States (2000-2035)

Key Outputs of Predictive Modeling The table below summarizes the predicted R&D spending for both nations from 2024 to 2035. These values illustrate China's continued growth trajectory, positioning it as a competitive force in global innovation.

Year	Predicted R&D Spending for China (% of GDP)	Predicted R&D Spending for the United States (% of GDP)
2024	2.69	3.26
2025	2.76	3.30
2026	2.84	3.33
2027	2.91	3.37
2028	2.98	3.40
2029	3.06	3.44
2030	3.13	3.47
2031	3.21	3.51
2032	3.28	3.54
2033	3.35	3.58
2034	3.43	3.61
2035	3.50	3.65

Table 3: Predicted R&D Spending for China and the United States (2024-2035)

These findings reveal an intensifying competition between the two nations in R&D investment. The sustained growth in China's spending highlights its focus on bridging the innovation gap, while the United States' incremental increase underscores its consistent emphasis on maintaining technological leadership. To deepen this analysis, we now turn our attention to the sentiment and thematic focus surrounding technology and innovation, analyzing public discourse and media narratives to uncover the broader implications of these investments.

5 Comparative Analysis of Sentiment and Technological Focus Between China and the United States

5.1 Data Acquisition and Methodology

To analyze sentiment and technological focus in articles related to China and the United States, we developed an advanced data acquisition pipeline capable of handling the challenges posed by dynamic, JavaScript-driven websites. Using Selenium and BeautifulSoup, we scraped and processed data from the China Briefing website [2] and Business Insider [3], creating a robust dataset for comprehensive analysis. This section highlights the key methods and techniques that ensured precision, scalability, and reliability in our data collection process.

5.1.1 China Data Acquisition

The China Briefing website served as a critical source of information, offering articles across categories such as *Industries*, *Technology*, *Economy & Trade*, and *Legal & Regulatory*. However, the site's reliance on JavaScript for dynamically rendering content presented a significant technical challenge, as key article elements—titles, publication dates, links, and full content—were not available in the initial HTML response.

To address this complexity, we used Selenium's `WebDriver`, which enabled us to emulate human interactions with the website, such as clicking on paginated links and loading full article pages. The headless Chrome browser configuration allowed efficient traversal of the site, and BeautifulSoup was subsequently employed to parse the dynamically rendered HTML and extract the required metadata. For each article, we navigated to its detailed page to retrieve the full text, ensuring comprehensive data capture.

This recursive process required meticulous error handling and optimization. Retry logic ensured that failed requests were revisited, while a rate limiter safeguarded against overloading the server. Articles were incrementally saved into a CSV file containing structured fields: `section`, `title`, `link`, `content`, `date`, and `author`. The result was a meticulously curated dataset of 841 articles spanning a publication range from March 2, 2016, to November 12, 2024.

The homepage of the China Briefing website, which showcases its structured yet dynamically loaded content, is shown in Figure 10. This layout required dynamic parsing methods, highlighting the sophistication of our scraping methodology.

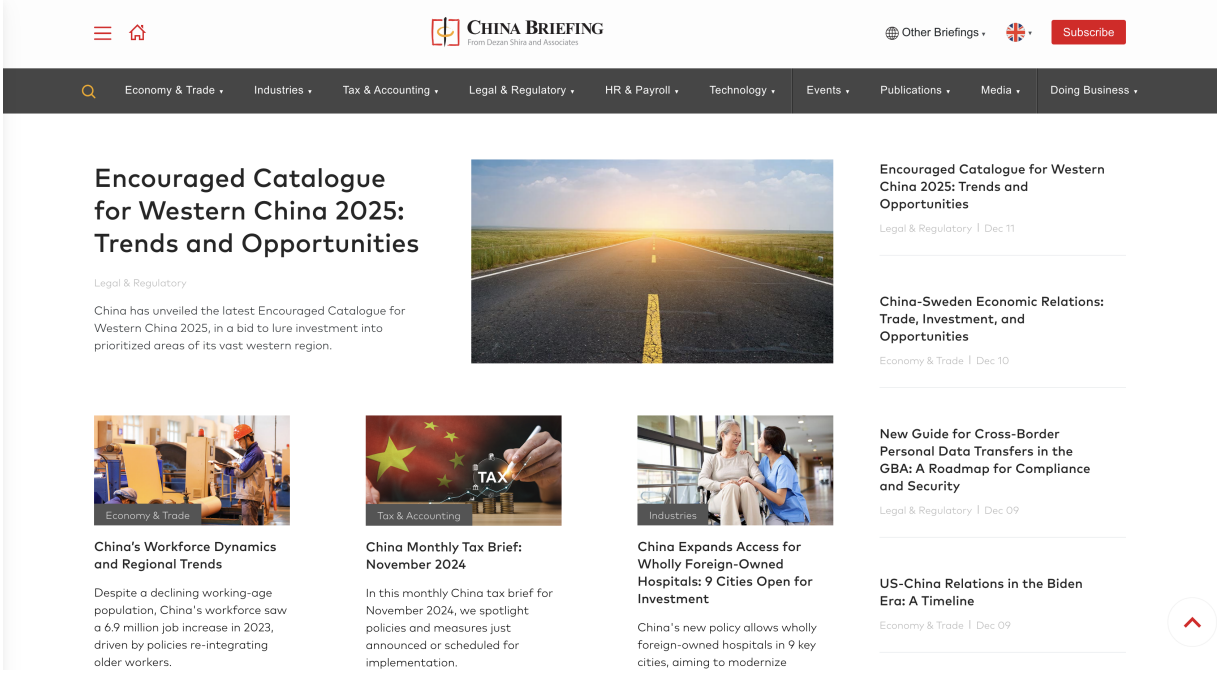

Figure 10: China Briefing Website Homepage

5.1.2 United States Data Acquisition

To capture sentiment and technological focus from a U.S. perspective, we scraped articles from Business Insider. This platform, like China Briefing, employed JavaScript to dynamically load content, requiring an equally robust approach. Using Selenium, we systematically navigated across categories such as *Technology*, *Business*, *Economy*, and *Innovation*, dynamically parsing each page to identify articles of interest.

For each article, we extracted metadata such as the title, author, publication date, and URL, before proceeding to the detailed view to capture the full text. BeautifulSoup played a pivotal role in parsing these dynamically loaded pages, and the inclusion of a dynamic delay mechanism helped us respect server limitations while maintaining efficient scraping speeds. Articles were saved into a structured CSV file, mirroring the format used for China-focused articles.

Figure 11 illustrates the homepage of the Business Insider website, which posed similar challenges with its JavaScript-heavy infrastructure and dynamically loaded elements. The scalability of our pipeline was further emphasized by its ability to handle diverse categories, each with unique HTML structures.

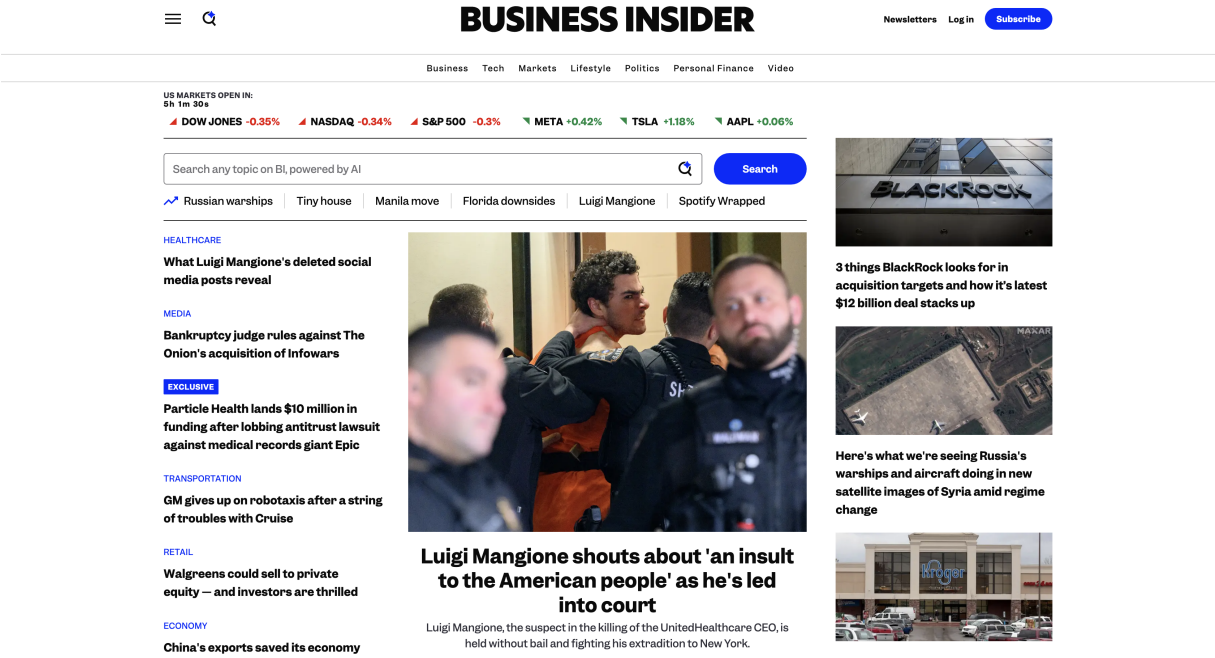

Figure 11: Business Insider Website Homepage

5.1.3 Data Acquisition Pipeline

The entire data acquisition and processing workflow is summarized in Figure 12. This pipeline began with web scraping, where Selenium and BeautifulSoup were used to retrieve dynamically rendered content. The raw data was stored in a CSV file for scalability and persistence. Subsequent preprocessing steps included cleaning the data using Pandas and NLTK, involving tokenization, stopword removal, and normalization. The cleaned data was then structured into DataFrames, enabling advanced Natural Language Processing (NLP) tasks such as sentiment analysis and keyword tracking. Finally, visualizations were generated using Matplotlib and Seaborn, presenting the insights derived from the processed data.

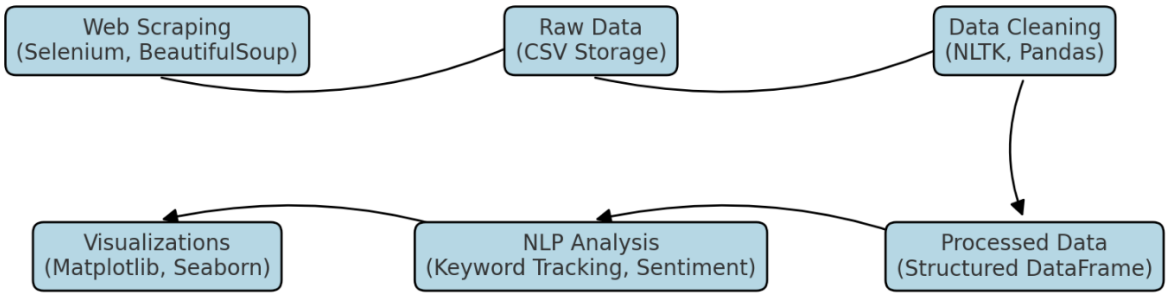

Figure 12: Overview of Data Acquisition Pipeline

This pipeline enabled us to handle the challenges posed by JavaScript-heavy websites, ensuring high-quality data collection for 841 articles from China Briefing and a similarly robust dataset from Business Insider. The integration of Selenium for dynamic content handling and efficient preprocessing workflows underscores the depth and reliability of this approach.

5.1.4 Preprocessing, Analysis, and Insights

The datasets collected from the China Briefing and Business Insider platforms underwent comprehensive preprocessing and sentiment analysis to extract actionable insights into the sentiment, themes, and technological focus embedded in the articles. This stage was instrumental in transforming raw, unstructured text into a structured format suitable for advanced analysis. The subsequent sections outline the preprocessing process and provide a detailed analysis of the visualizations and findings.

The preprocessing process utilized Python's Natural Language Toolkit (NLTK) to tokenize article content into individual words. Rigorous text cleaning was applied, including the removal of stopwords, punctuation, and non-alphabetic characters. To further enhance the quality of the dataset, domain-specific stopwords such as "might" and "also" were added to reduce irrelevant noise. Each cleaned article was stored in a new column, enabling efficient access for analysis. This systematic preprocessing revealed the linguistic structure of the dataset and laid a strong foundation for keyword frequency analysis.

The first analysis performed was a frequency distribution of keywords to identify dominant themes across the dataset. Figure 13 highlights the top 20 most frequently occurring words in the China dataset, with "China" being the most dominant term. Notably, the term "data" emerged as the second most frequently mentioned word, underscoring the dataset's thematic emphasis on analytics and technological advancements. The prominence of "data" is particularly striking, reflecting China's strategic positioning as a leader in emerging technologies, and further emphasizing the central role of information-driven narratives in the articles.

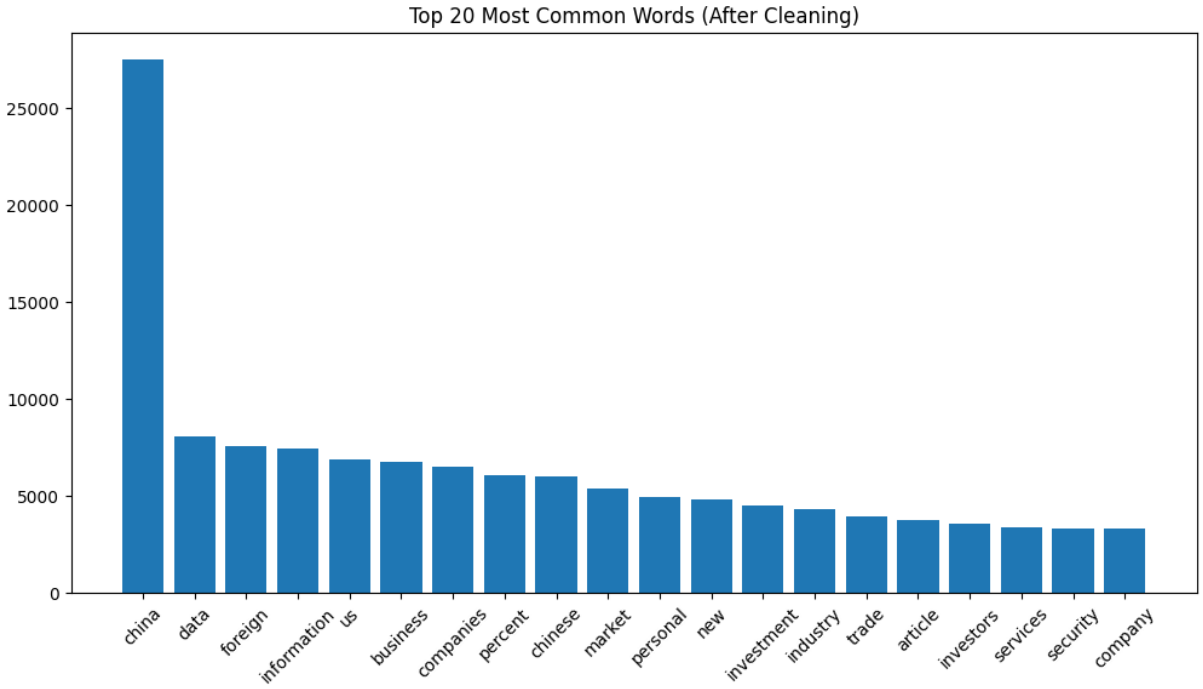

Figure 13: Top 20 Most Common Words in China Articles (After Cleaning)

Sentiment analysis was performed using NLTK's VADER sentiment analyzer, which assigned each article a compound sentiment score categorized as *positive*, *neutral*, or *negative*. The results were grouped by sections, enabling a granular view of sentiment distribution.

The sentiment distribution in the U.S. dataset is illustrated in Figure 14. Positive sentiment is prevalent across most sections, particularly in *Economy & Trade* and *Technology*, reflecting optimism about economic growth and technological progress. Neutral sentiment plays a significant role in sections like *Industries*, suggesting balanced narratives in articles focused on U.S. trade and industrial developments. However, the *Legal & Regulatory* section exhibits a substantial presence of negative sentiment, potentially tied to ongoing challenges related to compliance and regulatory issues. This highlights significant concerns in legal and policy discussions within the U.S. dataset.

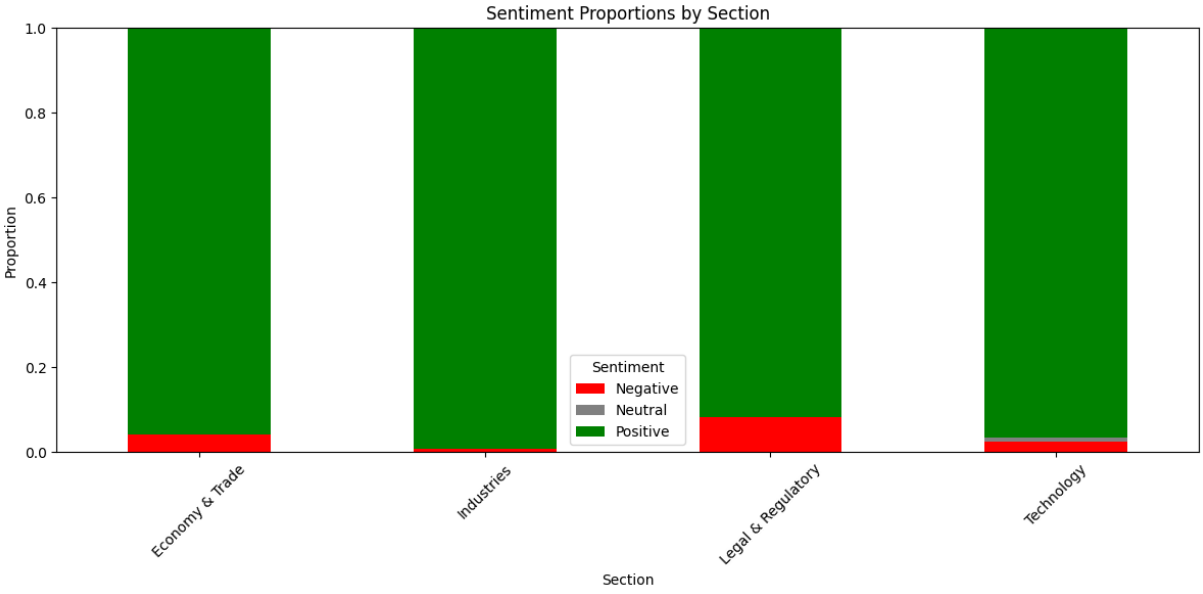

Figure 14: Sentiment Proportions by Section in U.S. Articles

In contrast, the sentiment distribution in the China dataset, shown in Figure 15, presents a markedly different perspective. Positive sentiment overwhelmingly dominates sections such as *Technology* and *Economy & Trade*, reflecting optimism surrounding advancements in technology and economic development. The *Legal & Regulatory* section in China's dataset, while containing some negative sentiment, demonstrates a smaller proportion of negativity compared to its U.S. counterpart. This indicates a more tempered focus on legal challenges within Chinese narratives. These findings reveal how the datasets diverge in tone, with the U.S. dataset exhibiting a stronger focus on legal concerns and the Chinese dataset highlighting technological progress and economic aspirations.

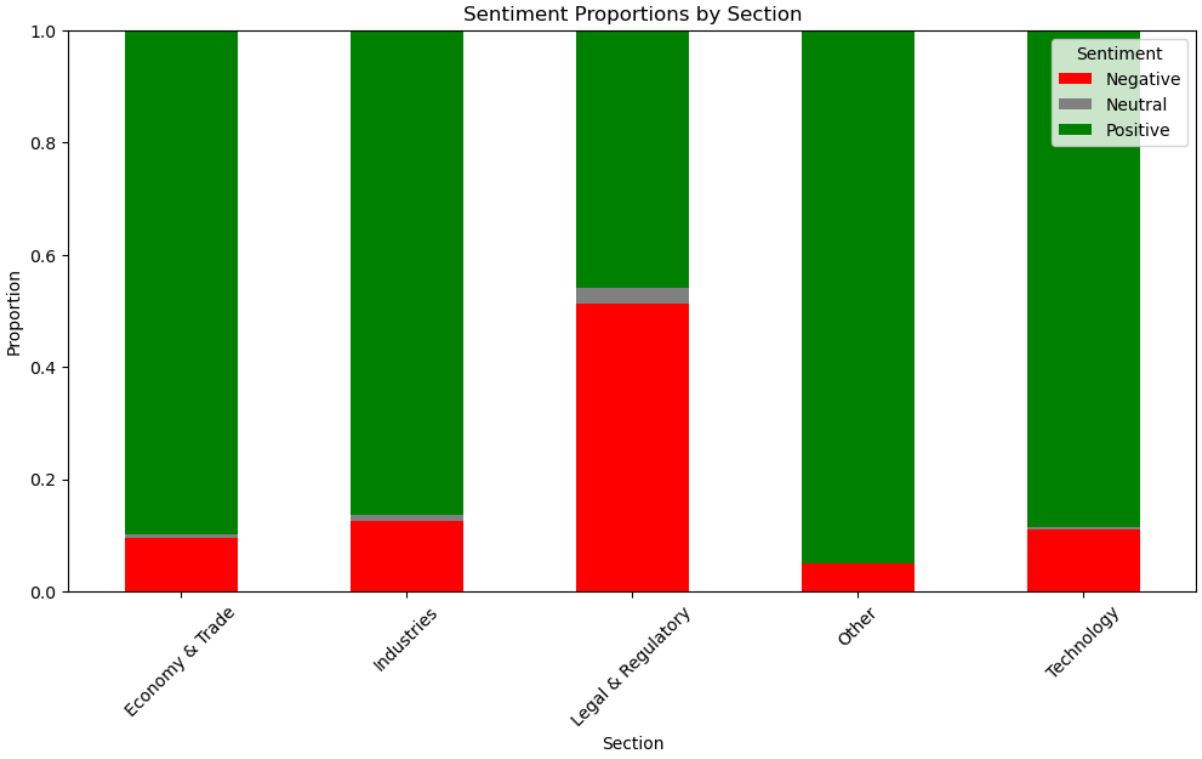

Figure 15: Sentiment Proportions by Section in China Articles

Further analysis examined the prevalence of technology-related terms over time. Using a predefined set of keywords such as "AI," "blockchain," "big data," and "robotics," the occurrences of these terms were tracked to uncover temporal patterns. Figure 16 presents the frequency of key technology terms in the China dataset, revealing notable spikes in mentions of "AI" and "blockchain." These spikes align with significant policy announcements and international events, such as China's AI development plans and state-led blockchain initiatives. This temporal analysis highlights the alignment between media narratives and strategic priorities in China.

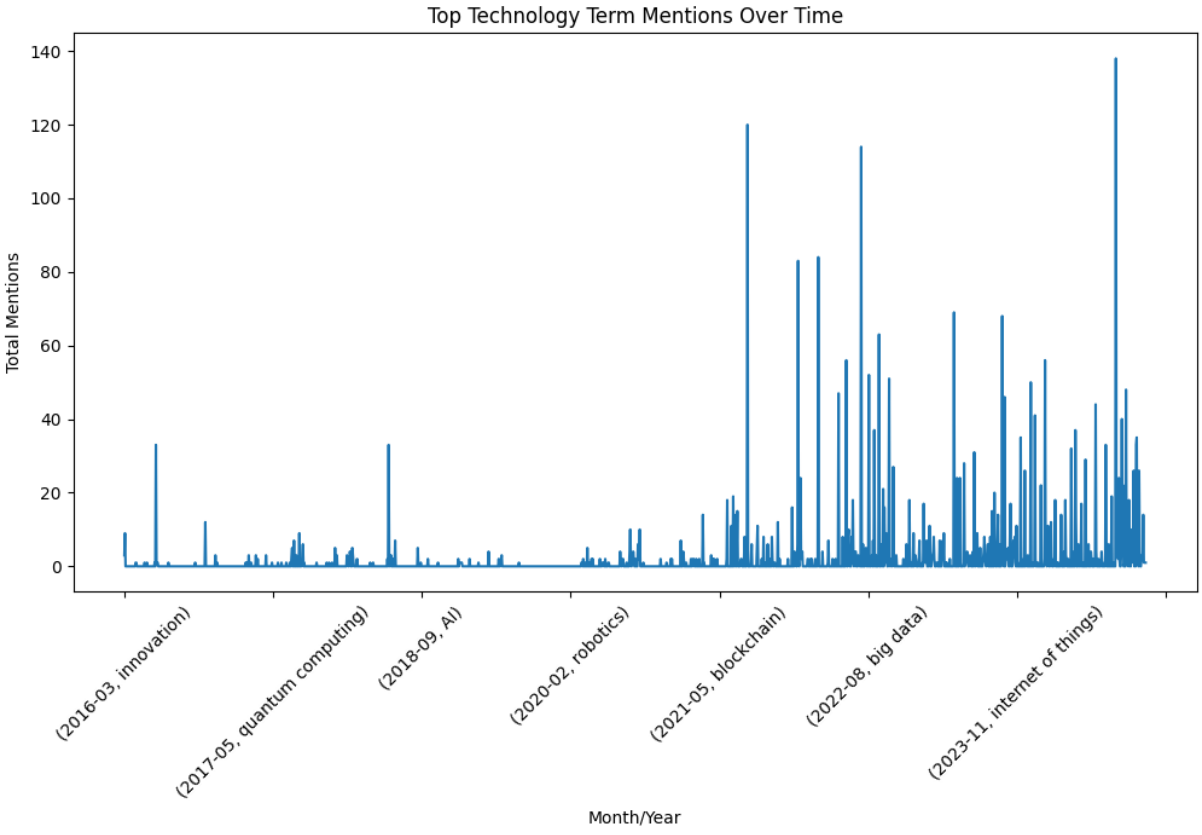

Figure 16: Top Technology Term Mentions Over Time in China Articles

The analysis of keyword frequencies, sentiment distributions, and technology mentions provides a comprehensive view of the thematic and tonal differences between the datasets. While both datasets reflect strong positive sentiment, the China dataset's focus on data-driven innovation and governance challenges contrasts with the U.S. dataset's emphasis on financial growth and regulatory scrutiny. The prominence of "data" in the China dataset as the second most mentioned word highlights the centrality of information and analytics in shaping narratives, underscoring the strategic importance of technology in China's global ambitions.

6 Conclusion

The comparative analysis of China's rapid economic ascent and its multifaceted relationship with the United States reveals a dynamic and increasingly complex global landscape. Over the past two decades, China has transitioned from a developing economy to a formidable global power, evidenced by its exponential GDP growth, expanding trade footprint, deepening technological capabilities, and rising influence in global value chains. While the United States still maintains leadership in absolute economic output and remains at the technological frontier, the gap is steadily narrowing—a trend made clear by China's robust increases in cumulative GDP growth, accelerating R&D investments, and strategic emphasis on advanced sectors.

Our exploration of trade dynamics highlights how the persistent US trade deficit with China underscores a deepening reliance on Chinese imports, particularly in consumer goods, machinery, and capital equipment. Efforts to rebalance this deficit—whether through tariffs, reshoring initiatives, or diversification of supply chains—have been met with limited success. Seasonal patterns and external shocks, such as the COVID-19 pandemic, further illustrate that the structural factors driving this dependency run deep. As the US confronts its vulnerability in critical supply chains, China's increasing leverage in global production networks could shape future policy, trade negotiations, and industrial strategies.

The analysis of R&D spending demonstrates that both nations recognize the critical importance of innovation in securing their economic futures. China's rising investment in research and technology, reflected in its growing share of GDP devoted to R&D, signals a deliberate and long-term strategy to move beyond low-cost manufacturing toward leadership in transformative fields such as artificial intelligence, 5G, biotechnology, and green technologies. While the US maintains a lead in advanced research and world-class innovation ecosystems, the projected convergence in R&D spending suggests an intensifying race for technological primacy. This heightened competition has implications for global standards, patent landscapes, and the shaping of international technology governance frameworks.

The sentiment and thematic analysis of media narratives in both countries reveal diverging storylines and priorities. In China-focused discourse, the prominence of "data" and positive sentiment around "technology" and "innovation" signifies a forward-looking narrative, centered on overcoming regulatory hurdles and scaling the heights of the digital economy. In contrast, US-focused content more frequently emphasizes regulatory challenges, reflecting the complexities of balancing innovation with broader economic, legal, and geopolitical considerations. This divergence in narrative focus suggests that while China champions a clear, data-driven approach to development, the US may grapple with reconciling innovation imperatives with calls for oversight, ethical standards, and legal frameworks.

Taken together, these analyses illustrate a deeply interwoven but increasingly competitive relationship, where economic interdependence and strategic rivalry coexist. China's rise does not simply present a binary threat or opportunity for the United States; rather, it reshapes the contours of collaboration, competition, and negotiation. Where once the US could rely on long-standing comparative advantages, it now faces a recalibration of global norms and market structures as China flexes its economic strength and diplomatic clout. The challenge for policymakers, business leaders, and innovators in both countries lies in navigating this evolving equilibrium—pursuing cooperation in global challenges like climate change and public health, while carefully managing competitive pressures in technology, trade, and investment.

In the coming years, how the US responds to China's growing economic prowess will significantly influence the global economic order. Policies that foster resilience in supply chains, support strategic domestic industries, and invest in cutting-edge R&D will help ensure that the United States remains an influential leader. For China, continued progress will hinge on balancing rapid economic development with international expectations, regulatory stability, and sustainable growth. Ultimately, the interplay of cooperation and competition between these two economic giants will shape the innovation pipelines, industrial strategies, and global prosperity of the twenty-first century.

7 Interactive Data Dashboard

For readers interested in further exploring the data and visualizations underlying this analysis, we've compiled our data into an interactive dashboard. This interface allows you to interact with various datasets, adjust timeframes, compare indicators, and review detailed charts. The link to the interface is provided below:

`https://us-cn-economy.vercel.app/`

Ths interface was developed using NextJS, NodeJS, and Typescript. We first compiled all of the preprocessed datasets generated from the data collection into CSV files. Using these datasets, we built a robust application capable of transforming the data into key metrics and visualizations, offering users the ability to derive meaningful insights. To generate these visualizations, we utilized libraries such as Recharts, Material-UI, Framer Motion, and FastAPI for running Python scripts. Lastly, we improved the frontend using React to make it more visually appeasing.

References

[1] World Bank. World bank open data api. https://data.worldbank.org/, 2024. Accessed: 2024-12-11.

[2] China Briefing. China briefing - latest business news and updates, 2024. Accessed: 2024-12-11.

[3] Business Insider. Business insider - breaking news and updates on business, tech, and more, 2024. Accessed: 2024-12-11.

[4] The World Bank. World Development Indicators: GDP (current US$). Accessed: 2024-11-11.

[5] U.S. Census Bureau. Foreign Trade: U.S. Trade in Goods with China. Accessed: 2024-11-11.

8 Contributions

Introduction

- Written by: Ehsaan Mohammed

Economic Growth Comparison: China vs. the United States

- Written by: Ehsaan Mohammed
- Code by: Ehsaan Mohammed

Trade Balance Analysis Between the US and China

- Written by: Ehsaan Mohammed, Naya Nethi
- Code by: Ehsaan Mohammed, Aadhil Mubarak Syed, Naya Nethi

Comparative Analysis of Research and Development Investments

- Written by: Ehsaan Mohammed
- Code by: Ehsaan Mohammed

Comparative Analysis of Sentiment and Technological Focus Between China and the United States

- Written by: Ehsaan Mohammed, Aadhil Mubarak Syed, Naya Nethi
- Code by: Ehsaan Mohammed (China), Aadhil Mubarak Syed (US)

Conclusion

- Written by: Ehsaan Mohammed

Interactive Data Dashboard

- Data/Code Compilation: Aadhil Mubarak Syed
- Development: Aadhil Mubarak Syed (Key Metrics, Visualizations, Frontend)
- Deployment: Aadhil Mubarak Syed (CI/CD Pipeline, Application Hosting)
- Written By: Ehsaan Mohammed, Aadhil Mubarak Syed

9 GitHub Repository

The code and data used in this analysis can be found in our GitHub repository:
https://github.com/aadhilmsyed/us-cn-economy.git